Frontiers
in
Artificial
Intelligence
and
Applications

EMERGING ARTIFICIAL INTELLIGENCE APPLICATIONS IN COMPUTER ENGINEERING

Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies

Edited by
Ilias Maglogiannis
Kostas Karpouzis
Manolis Wallace
John Soldatos

IOS
Press

EMERGING ARTIFICIAL INTELLIGENCE APPLICATIONS IN COMPUTER ENGINEERING

Frontiers in Artificial Intelligence and Applications

FAIA covers all aspects of theoretical and applied artificial intelligence research in the form of monographs, doctoral dissertations, textbooks, handbooks and proceedings volumes. The FAIA series contains several sub-series, including “Information Modelling and Knowledge Bases” and “Knowledge-Based Intelligent Engineering Systems”. It also includes the biennial ECAI, the European Conference on Artificial Intelligence, proceedings volumes, and other ECCAI – the European Coordinating Committee on Artificial Intelligence – sponsored publications. An editorial panel of internationally well-known scholars is appointed to provide a high quality selection.

Volume 160

Recently published in this series

Vol. 159. E. Tyugu (Ed.), Algorithms and Architectures of Artificial Intelligence
Vol. 158. R. Luckin et al. (Eds.), Artificial Intelligence in Education – Building Technology Rich Learning Contexts That Work
Vol. 157. B. Goertzel and P. Wang (Eds.), Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms – Proceedings of the AGI Workshop 2006
Vol. 156. R.M. Colomb, Ontology and the Semantic Web
Vol. 155. O. Vasilecas et al. (Eds.), Databases and Information Systems IV – Selected Papers from the Seventh International Baltic Conference DB&IS’2006
Vol. 154. M. Duží et al. (Eds.), Information Modelling and Knowledge Bases XVIII
Vol. 153. Y. Vogiazou, Design for Emergence – Collaborative Social Play with Online and Location-Based Media
Vol. 152. T.M. van Engers (Ed.), Legal Knowledge and Information Systems – JURIX 2006: The Nineteenth Annual Conference
Vol. 151. R. Mizoguchi et al. (Eds.), Learning by Effective Utilization of Technologies: Facilitating Intercultural Understanding
Vol. 150. B. Bennett and C. Fellbaum (Eds.), Formal Ontology in Information Systems – Proceedings of the Fourth International Conference (FOIS 2006)
Vol. 149. X.F. Zha and R.J. Howlett (Eds.), Integrated Intelligent Systems for Engineering Design
Vol. 148. K. Kersting, An Inductive Logic Programming Approach to Statistical Relational Learning

ISSN 0922-6389

Emerging Artificial Intelligence Applications in Computer Engineering

Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies

Edited by

Ilias Maglogiannis

Department of Information and Communication Systems Engineering, University of the Aegean, Samos, Greece

Kostas Karpouzis

Institute of Communication and Computer Systems, National Technical University of Athens, Greece

Manolis Wallace

Department of Computer Science, University of Indianapolis, Athens Campus, Greece

and

John Soldatos

Athens Information Technology, Greece

IOS Press

Amsterdam • Berlin • Oxford • Tokyo • Washington, DC

ISBN 978-1-58603-780-2
Library of Congress Control Number: 2007933091

Publisher
IOS Press
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

Distributor in the UK and Ireland
Gazelle Books Services Ltd.
White Cross Mills
Hightown
Lancaster LA1 4XS
United Kingdom
fax: +44 1524 63232
e-mail: sales@gazellebooks.co.uk

Distributor in the USA and Canada
IOS Press, Inc.
4502 Rachael Manor Drive
Fairfax, VA 22032
USA
fax: +1 703 323 3668
e-mail: iosbooks@iospress.com

PRINTED IN THE NETHERLANDS

I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Preface

Since the term "Artificial Intelligence" was first coined in 1955 by John McCarthy in his proposal for the Dartmouth Conference, but also even before that as reflected in works such that of Alan Turing, there has been a fiery philosophical discussion associated with it. Questions such as "what is it?", "can it really exist?", "will it ever surpass human intelligence?", "how should we refer to it?" and so on have troubled us for years and still continue to do so with undiminished intensity.

Regardless of how each one of us chooses to react to the aforementioned philosophical questions, there is one thing that we can all take for granted. The field that is referred to as artificial, computational or machine intelligence, or simply AI, has now begun to mature. Thus, correctly called intelligent or not, there is a vast list of methodologies, tools and applications that have been developed under the general umbrella of artificial intelligence which have provided practical solutions to difficult real life problems. Moreover, it is clear that, as computing progresses, more and more practical problems will find their solution in research performed in the field of artificial intelligence.

In general, intelligent applications build on the existing rich and proven theoretical background, as well as on ongoing basic research, in order to provide solutions for a wide range of real life problems. Nowadays, the ever expanding abundance of information and computing power enables researchers and users to tackle highly interesting issues for the first time, such as applications providing personalized access and interactivity to multimodal information based on user preferences and semantic concepts or human-machine interface systems utilizing information on the affective state of the user.

The purpose of this book is to provide insights on how today's computer engineers can implement AI in real world applications. Overall, the field of artificial intelligence is extremely broad. In essence, AI has found application, in one way or another, in every aspect of computing and in most aspects of modern life. Consequently, it is not possible to provide a complete review of the field in the framework of a single book, unless if the review is broad rather than deep. In this book we have chosen to present selected current and emerging practical applications of AI, thus allowing for a more detailed presentation of topics.

The book is organized in 4 parts. Part I "General Purpose Applications of AI" focuses on the most "conventional" areas of computational intelligence. On one side, we discuss the application of machine learning technologies and on the other we explore emerging applications of structured knowledge representation approaches. Part II "Intelligent Human-Computer Interaction" discusses the way in which progress in the field of AI has allowed for the improvement of the means that humans use to interact with machines and those that machines use, in turn, to analyze semantics and provide meaningful responses in context. Part III "Intelligent Applications in eHealth" focuses on the way that intelligence can be incorporated into medical data processing, thus allowing for the provision of enhanced medical services. Part IV "Real world AI applications in Computer Engineering" concludes the book with references to new and emerging applications of computational intelligence in real life problems.

Finally, all four editors are indebted to the authors who have contributed chapters on their respective fields of expertise and worked hard in order for deadlines to be met and for the overall book to be meaningful and coherent.

Ilias Maglogiannis,
Kostas Karpouzis,
Manolis Wallace,
John Soldatos

May 2007, Athens

Contents

This page intentionally left blank

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Part I: General Purpose Applications of AI

Manolis Wallace
Department of Computer Science
University of Indianapolis Athens
wallace@uindy.gr

This part of the book is devoted to the more "conventional" areas of computational intelligence, i.e. to machine learning and to knowledge representation.

Machine learning, the focus of Section 1, refers to the development of automated systems that are able to process large amounts of data in order to extract meaningful and potentially useful information (data mining) as well as to the exploitation of such information in practical problems (decision support). The combination of both components, i.e. information extraction and application, is called data classification; in this context, the aim is to develop tools that, having studied a large labeled base of available data, are able to automatically label not previously seen data. Chapter 1 starts by presenting the challenges that are associated with the task of classification and continues to provide an extensive review of the field, comprising information on logic based, perceptron based, statistical learning and support vector approaches and also including a comparison of the different techniques and a discussion on the combination of multiple techniques.

One problem that is typically associated with machine learning is humans' inability to monitor and verify the machine's operation. One of the parameters that make the task impossible for humans to tackle is the fact that the large amount and digital format of the data considered by the automated systems. Dimension reduction and data visualization are two closely related research fields that focus on alleviating this difficulty. Chapter 2 focuses on the utilization of neural networks in this direction. Starting with a thorough theoretical foundation, that includes presentation of both standalone and combined approaches, we continue to present applications of dimension reduction and data visualization in a series of real life problems.

Once useful information has been acquired, either through machine learning or even directly from a human expert, a decision support system can be developed that assists humans in making decisions, often in uncertain environments. One problem associated with the penetration and practical application of such systems is humans' justified reluctance to base important and some times critical decisions on the recommendation provided by a system the operation of which is not clear or fully understood. In order to overcome this, argumentation based recommender systems presented in Chapter 3 have been developed that attempt to present and justify the reasoning behind each recommendation they offer. In addition to a detailed presentation of related theory, the chapter also goes on to provide information on practical technologies and applications that are based on argumentation.

Section 2 focuses on knowledge representation in the context of artificial intelligence and its applications. Chapter 4 introduces the topic by discussing knowledge modeling, i.e. the specification of formal structures for the representation of knowledge. Based on profiling, the extensibility mechanism of UML, a UML extension is developed for knowledge modeling. As with most chapters in this book, the theoretical discussion is followed by the practical application of the developed approach in a real life problem.

The remaining of the section focuses on practical applications of ontological representations, which constitute the current trend in knowledge representation. Thus, chapter 5, using DAML+OIL, presents a web resource semantic annotation model. Based on it, a semantic navigation methodology is developed and applied.

Chapter 6 discusses the utilization of ontologies in the management of pervasive systems. Based on the Resource Description Framework, a methodology is proposed to facilitate interoperability between diverse devices and sensors, that enabling futher development of pervasive systems.

Chapter 7 takes us to a more conventional application of ontologies, that of annotation of resources and items for web applications. Specifically, we present a state-of-the-art e-commerce system that, combining ontologies with advanced visualization techniques, is able to allow customers to fully customize products according to their needs.

Chapter 8 introduces the emerging concept of the information market, where marketed goods are organizational knowledge and experience. In this context, we see how ontologies can be used in order to achieve the critical balance between i) making the existence and content overview of offered information items publicly known and searchable and ii) protecting the rights associated with these information items.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Supervised Machine Learning: A Review of Classification Techniques

S.B. KOTSIANTIS
Department of Computer Science & Technology, University of Peloponnese, Greece
sotos@math.upatras.gr

Abstract. The goal of supervised learning is to build a concise model of the distribution of class labels in terms of predictor features. The resulting classifier is then used to assign class labels to the testing instances where the values of the predictor features are known, but the value of the class label is unknown. This paper describes various supervised machine learning classification techniques. Of course, a single chapter cannot be a complete review of all supervised machine learning classification algorithms (also known induction classification algorithms), yet we hope that the references cited will cover the major theoretical issues, guiding the researcher in interesting research directions and suggesting possible bias combinations that have yet to be explored.

Keywords. Classifiers, data mining, intelligent data analysis, learning algorithms

Introduction

There are several applications for *Machine Learning* (ML), the most significant of which is data mining. People are often prone to making mistakes during analyses or, possibly, when trying to establish relationships between multiple features. This makes it difficult for them to find solutions to certain problems. Machine learning can often be successfully applied to these problems, improving the efficiency of systems and the designs of machines.

Every instance in any dataset used by machine learning algorithms is represented using the same set of features. The features may be continuous, categorical or binary. If instances are given with known labels (the corresponding correct outputs) then the learning is called *supervised*, in contrast to *unsupervised learning,* where instances are unlabeled. By applying these unsupervised (clustering) algorithms, researchers hope to discover unknown, but useful, classes of items [1]. Another kind of machine learning is *reinforcement learning* [2]. The training information provided to the learning system by the environment (external trainer) is in the form of a scalar reinforcement signal that constitutes a measure of how well the system operates. The learner is not told which actions to take, but rather must discover which actions yield the best reward, by trying each action in turn.

Numerous ML applications involve tasks that can be set up as supervised. In the present paper, we have concentrated on the techniques necessary to do this. In particular, this work is concerned with classification problems in which the output of instances admits only discrete, unordered values. We have limited our references to recent refereed journals, published books and conferences. In addition, we have added some references regarding the original work that started the particular line of research under discussion. A brief review of what ML includes can be found in [3]. A historical survey of logic and instance based learning classifiers is also presented in [4]. The reader should be cautioned that a single chapter cannot be a comprehensive review of all classification learning algorithms. Instead, our goal has been to provide a representative sample of existing lines of research in each learning technique. In each of our listed areas, there are many other papers that more comprehensively detail relevant work.

Our next section covers wide-ranging issues of supervised machine learning such as data pre-processing and feature selection. Logic-based learning techniques are described in Section 2, whereas perceptron-based techniques are analyzed in Section 3. Statistical techniques for ML are covered in Section 4. Section 5 deals with the newest supervised ML technique—Support Vector Machines (SVMs). In Section 6, learning techniques are compared. Section 7 presents a recent attempt for improving classification accuracy—ensembles of classifiers. Section 8 presents some special classification problems such as learning from imbalanced datasets and learning from multimedia files. Finally, the last section concludes this work.

1. General Issues of Supervised Learning Algorithms

Inductive machine learning is the process of learning a set of rules from instances (examples in a training set), or more generally speaking, creating a classifier that can be used to generalize from new instances. The first step is collecting the dataset. If a requisite expert is available, then s/he could suggest which fields (attributes, features) are the most informative. If not, then the simplest method is that of "brute-force," which means measuring everything available in the hope that the right (informative, relevant) features can be isolated. However, a dataset collected by the "brute-force" method is not directly suitable for induction. It contains in most cases noise and missing feature values, and therefore requires significant pre-processing [5].

What can be wrong with data? There is a hierarchy of problems that are often encountered in data preparation and pre-processing:

- Impossible or unlikely values have been inputted.
- No values have been inputted (missing values).
- Irrelevant input features are present in the data at hand.

Impossible values (noise) should be checked for by the data handling software, ideally at the point of input so that they can be re-entered. If correct values cannot be entered, the problem is converted into missing value category, by simply removing the data. Hodge & Austin [6] have recently introduced a survey of contemporary techniques for outlier (noise) detection.

Incomplete data is an unavoidable problem in dealing with most real world data sources. Generally, there are some important factors to be taken into account when processing unknown feature values. One of the most important ones is the source of "unknown-ness": (i) a value is missing because it was forgotten or lost; (ii) a certain feature is not applicable for a given instance (e.g., it does not exist for a given instance); (iii) for a given observation, the designer of a training set does not care about the value of a certain feature (so-called "don't-care values)." Depending on the circumstances, researchers have a number of methods to choose from to handle missing data [7].

Feature subset selection is the process of identifying and removing as many irrelevant and redundant features as possible [8]. This reduces the dimensionality of the data and enables data mining algorithms to operate faster and more effectively. The fact that many features depend on one another often unduly influences the accuracy of supervised ML classification models. This problem can be addressed by constructing new features from the basic feature set [9].

The choice of which specific learning algorithm we should use is a critical step. Once preliminary testing is judged to be satisfactory, the classifier (mapping from unlabeled instances to classes) is available for routine use. The classifier's evaluation is most often based on prediction accuracy (the percentage of correct prediction divided by the total number of predictions). There are at least three techniques which are used to calculate a classifier's accuracy. One technique is to split the training set by using two-thirds for training and the other third for estimating performance. In another technique, known as cross-validation, the training set is divided into mutually exclusive and equal-sized subsets and for each subset the classifier is trained on the union of all the other subsets. The average of the error rate of each subset is therefore an estimate of the error rate of the classifier. Leave-one-out validation is a special case of cross validation. All test subsets consist of a single instance. This type of validation is, of course, more expensive computationally, but useful when the most accurate estimate of a classifier's error rate is required.

Supervised classification is one of the tasks most frequently carried out by so-called Intelligent Systems. Thus, a large number of techniques have been developed based on Artificial Intelligence (Logic-based techniques, Perceptron-based techniques) and Statistics (Bayesian Networks, Instance-based techniques). In the next section, we will focus on the most important supervised machine learning techniques, starting with logic-based techniques.

2. Logic Based Algorithms

In this section we will concentrate on two groups of logic (symbolic) learning methods: decision trees and rule-based classifiers.

2.1. Decision Trees

Murthy [10] provided an overview of work in decision trees and a sample of their usefulness to newcomers as well as practitioners in the field of machine learning. Thus, in this work, apart from a brief description of decision trees, we will refer to some more recent works than those in Murthy's article as well as few very important articles that were published earlier. Decision trees are trees that classify instances by sorting them based on feature values. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. Figure 1 is an example of a decision tree.

Using the decision tree depicted in Fig. 1 as an example, the instance (at1 = a1, at2 = b2, at3 = a3, at4 = b4) would sort to the nodes: at1, at2, and finally at3, which would classify the instance as being positive (represented by the values "Yes"). The problem of constructing optimal binary decision trees is an NP-complete problem and

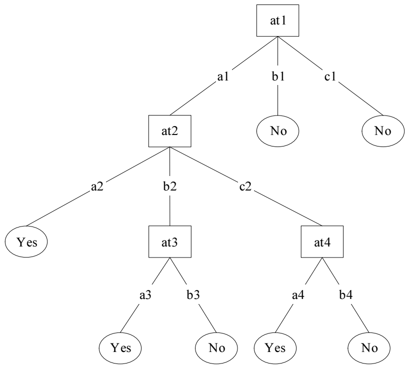


Figure1. A decision tree.

thus theoreticians have searched for efficient heuristics for constructing near-optimal decision trees.

The feature that best divides the training data would be the root node of the tree. There are numerous methods for finding the feature that best divides the training data but a majority of studies have concluded that there is no single best method [10]. Comparison of individual methods may still be important when deciding which metric should be used in a particular dataset. The same procedure is then repeated on each partition of the divided data, creating sub-trees until the training data is divided into subsets of the same class.

A decision tree, or any learned hypothesis h, is said to overfit training data if another hypothesis h′ exists that has a larger error than h when tested on the training data, but a smaller error than h when tested on the entire dataset. There are two common approaches that decision tree induction algorithms can use to avoid overfitting training data: i) Stop the training algorithm before it reaches a point at which it perfectly fits the training data, ii) Prune the induced decision tree. If the two trees employ the same kind of tests and have the same prediction accuracy, the one with fewer leaves is usually preferred. Most algorithms use a pruning method. A comparative study of well-known pruning methods presented in [11].

The most well-know algorithm in the literature for building decision trees is the C4.5 [12]. One of the latest studies that compare decision trees and other learning algorithms has been done by [13]. The study shows that C4.5 has a very good combination of error rate and speed. C4.5 assumes that the training data fits in memory, thus, Gehrke et al. [14] proposed Rainforest, a framework for developing fast and scalable algorithms to construct decision trees that gracefully adapt to the amount of main memory available.

Decision trees are usually univariate since they use splits based on a single feature at each internal node. However, there are a few methods that construct multivariate trees. One example is Zheng's [15], who improved the classification accuracy of the decision trees by constructing new binary features with logical operators such as conjunction, negation, and disjunction. In addition, Zheng [16] created at-least M-of-N features. For a given instance, the value of an at-least M-of-N representation is true if at least M of its conditions is true of the instance, otherwise it is false. Gama and Brazdil [17] combined a decision tree with a linear discriminant for constructing multivariate decision trees. In this model, new features are computed as linear combinations of the previous ones.

Baik and Bala [18] presented preliminary work on an agent-based approach for the distributed learning of decision trees. To sum up, one of the most useful characteristics of decision trees is their comprehensibility. People can easily understand why a decision tree classifies an instance as belonging to a specific class.

2.2. Learning Set of Rules

Decision trees can be translated into a set of rules by creating a separate rule for each path from the root to a leaf in the tree [12]. However, rules can also be directly induced from training data using a variety of rule-based algorithms. Furnkranz [19] provided an excellent overview of existing work in rule-based methods. Classification rules represent each class by disjunctive normal form (DNF). A k-DNF expression is of the form: $(X_1 \wedge X_2 \wedge \ldots \wedge X_n) \vee (X_{n+1} \wedge X_{n+2} \wedge \ldots X_{2n}) \vee \ldots \vee (X_{(k-1)n+1} \wedge X_{(k-1)n+2} \wedge \ldots \wedge X_{kn})$, where k is the number of disjunctions, n is the number of conjunctions in each disjunction, and X_n is defined over the alphabet $X_1, X_2, \ldots, X_j \cup \sim X_1, \sim X_2, \ldots, \sim X_j$. The goal is to construct the smallest rule-set that is consistent with the training data. A large number of learned rules is usually a sign that the learning algorithm is attempting to "remember" the training set, instead of discovering the assumptions that govern it.

A separate-and-conquer algorithm search for a rule that explains a part of its training instances, separates these instances and recursively conquers the remaining instances by learning more rules, until no instances remain. The difference between heuristics for rule learning and heuristics for decision trees is that the latter evaluate the average quality of a number of disjointed sets (one for each value of the feature that is tested), while rule learners only evaluate the quality of the set of instances that is covered by the candidate rule.

It is therefore important for a rule induction system to generate decision rules that have high predictability or reliability. These properties are commonly measured by a function called rule quality. A rule quality measure is needed in both the rule induction and classification processes. In rule induction, a rule quality measure can be used as a criterion in the rule specification and/or generalization process. In classification, a rule quality value can be associated with each rule to resolve conflicts when multiple rules are satisfied by the example to be classified. An and Cercone [20] surveyed a number of statistical and empirical rule quality measures. When using unordered rule sets, conflicts can arise between the rules, i.e., two or more rules cover the same example but predict different classes. Lindgren [21] has recently given a survey of methods used to solve this type of conflict.

To sum up, the most useful characteristic of rule-based classifiers is their comprehensibility. For the task of learning binary problems, rules are more comprehensible than decision trees because typical rule-based approaches learn a set of rules for only the positive class. On the other hand, if definitions for multiple classes are to be learned, the rule-based learner must be run separately for each class separately. For each individual class a separate rule set is obtained and these sets may be inconsistent (a particular instance might be assigned multiple classes) or incomplete (no class might be assigned to a particular instance). These problems can be solved with decision lists (the rules in a rule set are supposed to be ordered, a rule is only applicable when none of the preceding rules are applicable) but with the decision tree approach, they simply do not occur.

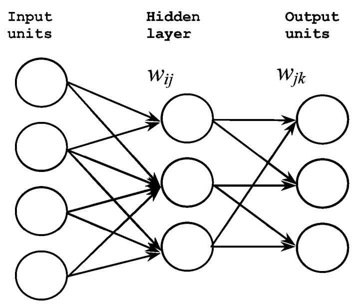


Figure 2. Feed-forward ANN.

3. Perceptron-Based Techniques

Other well-known algorithms are based on the notion of perceptron. Perceptron can be briefly described as: If x_1 through x_n are input feature values and w_1 through w_n are connection weights/prediction vector (typically real numbers in the interval [–1, 1]), then perceptron computes the sum of weighted inputs: $\sum_i x_i w_i$ and output goes through an adjustable threshold: if the sum is above threshold, output is 1; else it is 0.

The most common way the perceptron algorithm is used for learning from a batch of training instances is to run the algorithm repeatedly through the training set until it finds a prediction vector which is correct on all of the training set. This prediction rule is then used for predicting the labels on the test set.

3.1. Neural Networks

Perceptrons can only classify linearly separable sets of instances. If a straight line or plane can be drawn to separate the input instances into their correct categories, input instances are linearly separable and the perceptron will find the solution. If the instances are not linearly separable learning will never reach a point where all instances are classified properly. Artificial Neural Networks have been created to try to solve this problem. Zhang [22] provided an overview of existing work in Artificial Neural Networks (ANNs). Thus, in this study, apart from a brief description of the ANNs we will mainly refer to some more recent articles.

A multi-layer neural network consists of large number of units (neurons) joined together in a pattern of connections (Fig. 2). Units in a net are usually segregated into three classes: input units, which receive information to be processed; output units, where the results of the processing are found; and units in between known as hidden units. Feed-forward ANNs (Fig. 2) allow signals to travel one way only, from input to output. First, the network is trained on a set of paired data to determine input-output mapping. The weights of the connections between neurons are then fixed and the network is used to determine the classifications of a new set of data.

Generally, properly determining the size of the hidden layer is a problem, because an underestimate of the number of neurons can lead to poor approximation and generalization capabilities, while excessive nodes can result in overfitting and eventually make the search for the global optimum more difficult. An excellent argument regarding this topic can be found in [23]. Kon & Plaskota [24] also studied the minimum amount of neurons and the number of instances necessary to program a given task into feed-forward neural networks.

ANN depends upon three fundamental aspects, input and activation functions of the unit, network architecture and the weight of each input connection. Given that the first two aspects are fixed, the behavior of the ANN is defined by the current values of the weights. The weights of the net to be trained are initially set to random values, and then instances of the training set are repeatedly exposed to the net. The values for the input of an instance are placed on the input units and the output of the net is compared with the desired output for this instance. Then, all the weights in the net are adjusted slightly in the direction that would bring the output values of the net closer to the values for the desired output. There are several algorithms with which a network can be trained [25]. However, the most well-known and widely used learning algorithm to estimate the values of the weights is the Back Propagation (BP) algorithm.

Feed-forward neural networks are usually trained by the original back propagation algorithm or by some variant. Their greatest problem is that they are too slow for most applications. One of the approaches to speed up the training rate is to estimate optimal initial weights [26]. Genetic algorithms have been used to train the weights of neural networks [27] and to find the architecture of neural networks [28].

Even though multilayer neural networks and decision trees are two very different techniques for the purpose of classification, some researchers have performed some empirical comparative studies [29,13]. Some of the general conclusions drawn in that work are:

i) neural networks are usually more able to easily provide incremental learning than decision trees.
ii) training time for a neural network is usually much longer than training time for decision trees.
iii) neural networks usually perform as well as decision trees, but seldom better.

To sum up, ANNs have been applied to many real-world problems but still, their most striking disadvantage is their lack of ability to reason about their output in a way that can be effectively communicated. For this reason many researchers have tried to address the issue of improving the comprehensibility of neural networks, where the most attractive solution is to extract symbolic rules from trained neural networks [30].

4. Statistical Learning Algorithms

Conversely to ANNs, statistical approaches are characterized by having an explicit underlying probability model, which provides a probability that an instance belongs in each class, rather than simply a classification. Under this category of classification algorithms, one can find Bayesian networks and instance-based methods. A comprehensive book on Bayesian networks is Jensen's [31]. Thus, in this study, apart from our brief description of Bayesian networks, we mainly refer to more recent works.

4.1. Bayesian Networks

A Bayesian Network (BN) is a graphical model for probability relationships among a set of variables (features) (see Fig. 3). The Bayesian network structure S is a directed

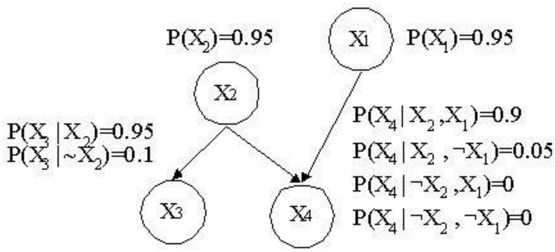


Figure 3. The structure of a Bayes Network.

acyclic graph (DAG) and the nodes in *S* are in one-to-one correspondence with the features *X*. The arcs represent casual influences among the features while the *lack* of possible arcs in *S* encodes conditional independencies. Moreover, a feature (node) is conditionally independent from its non-descendants given its parents (X_1 is conditionally independent from X_2 given X_3 if $P(X_1|X_2,X_3) = P(X_1|X_3)$ for all possible values of X_1, X_2, X_3).

Typically, the task of learning a Bayesian network can be divided into two subtasks: initially, the learning of the DAG structure of the network, and then the determination of its parameters. Probabilistic parameters are encoded into a set of tables, one for each variable, in the form of local conditional distributions of a variable given its parents. Given the independences encoded into the network, the joint distribution can be reconstructed by simply multiplying these tables. Within the general framework of inducing Bayesian networks, there are two scenarios: known structure and unknown structure.

In the first scenario, the structure of the network is given (e.g. by an expert) and assumed to be correct. Once the network structure is fixed, learning the parameters in the Conditional Probability Tables (CPT) is usually solved by estimating a locally exponential number of parameters from the data provided [31]. Each node in the network has an associated CPT that describes the conditional probability distribution of that node given the different values of its parents.

If the structure is unknown, one approach is to introduce a scoring function (or a score) that evaluates the "fitness" of networks with respect to the training data, and then to search for the best network according to this score. Several researchers have shown experimentally that the selection of a single good hypothesis using greedy search often yields accurate predictions [32,33]. Within the score & search paradigm, another approach uses local search methods in the space of directed acyclic graphs, where the usual choices for defining the elementary modifications (local changes) that can be applied are arc addition, arc deletion, and arc reversal. Acid and de Campos [34] proposed a new local search method, restricted acyclic partially directed graphs, which uses a different search space and takes account of the concept of equivalence between network structures. In this way, the number of different configurations of the search space is reduced, thus improving efficiency.

A BN structure can be also found by learning the conditional independence relationships among the features of a dataset. Using a few statistical tests (such as the Chi-squared and mutual information test), one can find the conditional independence relationships among the features and use these relationships as constraints to construct a BN. These algorithms are called *CI-based* algorithms or constraint-based algorithms.

Cowell [35] has shown that for any structure search procedure based on CI tests, an equivalent procedure based on maximizing a score can be specified. Using a suitable version of any of the model types mentioned in this review, one can induce a Bayesian Network from a given training set. A classifier based on the network and on the given set of features $X_1, X_2, \ldots X_n$, returns the label *c,* which maximizes the posterior probability $p(c \mid X_1, X_2, \ldots X_n)$.

The most interesting feature of BNs, compared to decision trees or neural networks, is most certainly the possibility of taking into account prior information about a given problem, in terms of structural relationships among its features. A problem of BN classifiers is that they are not suitable for datasets with many features [36]. The reason for this is that trying to construct a very large network is simply not feasible in terms of time and space. Naive Bayesian networks (NB) are very simple Bayesian networks which are composed of DAGs with only one parent (representing the unobserved node) and several children (corresponding to observed nodes) with a strong assumption of independence among child nodes in the context of their parent. The major advantage of the naive Bayes classifier is its short computational time for training. If a feature is numerical, the usual procedure for all Bayesian algorithms is to discretize it during data pre-processing [37], although a researcher can use the normal distribution to calculate probabilities [38].

4.2. Instance-Based Learning

Another category under the header of statistical methods is Instance-based learning. Instance-based learning algorithms are lazy-learning algorithms [39], as they delay the induction or generalization process until classification is performed. Lazy-learning algorithms require less computation time during the training phase than eager-learning algorithms (such as decision trees, neural and Bayes nets) but more computation time during the classification process. One of the most straightforward instance-based learning algorithms is the nearest neighbour algorithm. Aha [40] and De Mantaras and Armengol [4] presented a review of instance-based learning classifiers. Thus, in this study, apart from a brief description of the nearest neighbour algorithm, we will refer to some more recent works.

k-Nearest Neighbour (kNN) is based on the principle that the instances within a dataset will generally exist in close proximity to other instances that have similar properties. If the instances are tagged with a classification label, then the value of the label of an unclassified instance can be determined by observing the class of its nearest neighbours. The kNN locates the k nearest instances to the query instance and determines its class by identifying the single most frequent class label.

In general, instances can be considered as points within an n-dimensional instance space where each of the n-dimensions corresponds to one of the n-features that are used to describe an instance. The absolute position of the instances within this space is not as significant as the relative distance between instances. This relative distance is determined by using a distance metric. Ideally, the distance metric must minimize the distance between two similarly classified instances, while maximizing the distance between instances of different classes. Many different metrics have been presented [41]. For more accurate results, several algorithms use weighting schemes that alter the distance measurements and voting influence of each instance. A survey of weighting schemes is given by [42].

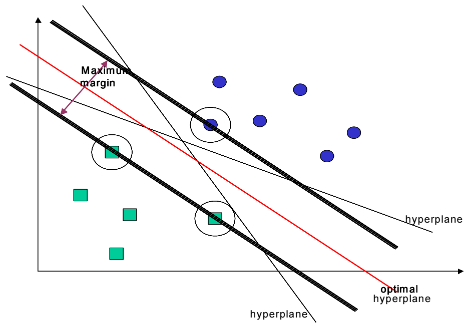


Figure 4. Maximum Margin.

The power of kNN has been demonstrated in a number of real domains, but there are some reservations about the usefulness of kNN, such as: i) they have large storage requirements, ii) they are sensitive to the choice of the similarity function that is used to compare instances, iii) they lack a principled way to choose *k*, except through cross-validation or similar, computationally-expensive technique. Okamoto and Yugami [43] represented the expected classification accuracy of k-NN as a function of domain characteristics including the number of training instances, the number of relevant and irrelevant attributes, the probability of each attribute, the noise rate for each type of noise, and k. They also explored the behavioral implications of the analyses by presenting the effects of domain characteristics on the expected accuracy of k-NN and on the optimal value of k for artificial domains.

As we have already mentioned, the major disadvantage of instance-based classifiers is their large computational time for classification. A key issue in many applications is to determine which of the available input features should be used in modeling via feature selection [8], because it could improve the classification accuracy and scale down the required classification time. Another issue is to determine which of the available instances should be used in modeling via instance selection [44].

5. Support Vector Machines

Support Vector Machines (SVMs) are the newest supervised machine learning technique. An excellent survey of SVMs can be found in [45], and a more recent book is by [46]. Thus, in this study apart from a brief description of SVMs we will refer to some more recent works and the landmark that were published before these works. SVMs revolve around the notion of a "margin"—either side of a hyperplane that separates two data classes. Maximizing the margin and thereby creating the largest possible distance between the separating hyperplane and the instances on either side of it has been proven to reduce an upper bound on the expected generalisation error.

In the case of linearly separable data, once the optimum separating hyperplane is found, data points that lie on its margin are known as support vector points and the solution is represented as a linear combination of only these points (see Fig. 4). Other data points are ignored. Therefore, the model complexity of an SVM is unaffected by the number of features encountered in the training data (the number of support vectors selected by the SVM learning algorithm is usually small). For this reason, SVMs are well suited to deal with learning tasks where the number of features is large with respect to the number of training instances.

Even though the maximum margin allows the SVM to select among multiple candidate hyperplanes, for many datasets, the SVM may not be able to find any separating hyperplane at all because the data contains misclassified instances. The problem can be addressed by using a soft margin that accepts some misclassifications of the training instances [47].

Nevertheless, most real-world problems involve non-separable data for which no hyperplane exists that successfully separates the positive from negative instances in the training set. One solution to the inseparability problem is to map the data onto a higher-dimensional space and define a separating hyperplane there. This higher-dimensional space is called the *feature space*, as opposed to the *input space* occupied by the training instances. With an appropriately chosen feature space of sufficient dimensionality, any consistent training set can be made separable. A linear separation in feature space corresponds to a non-linear separation in the original input space. Mapping the data to some other (possibly infinite dimensional) Hilbert space H as $\Phi : R^d \rightarrow H$. Then the training algorithm would only depend on the data through dot products in H, i.e. on functions of the form $\Phi(x_i) \cdot \Phi(x_j)$. If there were a "kernel function" K such that $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$, we would only need to use K in the training algorithm, and would never need to explicitly determine Φ. Thus, kernels are a special class of functions that allow inner products to be calculated directly in feature space, without performing the mapping described above [48]. Once a hyperplane has been created, the kernel function is used to map new points into the feature space for classification.

The selection of an appropriate kernel function is important, since the kernel function defines the feature space in which the training set instances will be classified. Genton [49] described several classes of kernels; however, he did not address the question of which class is best suited to a given problem. It is common practice to estimate a range of potential settings and use cross-validation over the training set to find the best one. For this reason a limitation of SVMs is the low speed of the training. Selecting kernel settings can be regarded in a similar way to choosing the number of hidden nodes in a neural network. As long as the kernel function is legitimate, a SVM will operate correctly even if the designer does not know exactly what features of the training data are being used in the kernel-induced feature space.

Training the SVM is done by solving N^{th} dimensional QP problem, where N is the number of samples in the training dataset. Solving this problem in standard QP methods involves large matrix operations, as well as time-consuming numerical computations, and is mostly very slow and impractical for large problems. Sequential Minimal Optimization (SMO) is a simple algorithm that can, relatively quickly, solve the SVM QP problem without any extra matrix storage and without using numerical QP optimization steps at all [50]. SMO decomposes the overall QP problem into QP subproblems. Keerthi and Gilbert [51] suggested two modified versions of SMO that are significantly faster than the original SMO in most situations.

Finally, the training optimization problem of the SVM necessarily reaches a global minimum, and avoids ending in a local minimum, which may happen in other search algorithms such as neural networks. However, the SVM methods are binary, thus in the case of multi-class problem one must reduce the problem to a set of multiple binary classification problems [52]. Discrete data presents another problem; although with suitable rescaling good results can be obtained.

6. Comparing Learning Techniques

Generally, SVMs and neural networks tend to perform much better when dealing with multi-dimensions and continuous features. In contrast, logic-based systems (e.g. decision trees, rule learners) tend to perform better when dealing with discrete/categorical features. For neural network models and SVMs, a large sample size is required in order to achieve its maximum prediction accuracy whereas Naive Bayes may need a relatively small dataset. Most decision tree algorithms cannot perform well with problems that require diagonal partitioning. The division of the instance space is orthogonal to the axis of one variable and parallel to all other axes. Therefore, the resulting regions after partitioning are all hyperrectangles. The ANNs and the SVMs perform well when multicollinearity is present and a nonlinear relationship exists between the input and output features.

Although training time varies according to the nature of the application task and dataset, specialists generally agree on a partial ordering of the major classes of learning algorithms. For instance, lazy learning methods require zero training time because the training instance is simply stored. Naive Bayes methods also train very quickly since they require only a single pass on the data either to count frequencies (for discrete variables) or to compute the normal probability density function (for continuous variables under normality assumptions). Univariate decision trees are also reputed to be quite fast—at any rate, several orders of magnitude faster than neural networks and SVMs.

Naive Bayes requires little storage space during both the training and classification stages: the strict minimum is the memory needed to store the prior and conditional probabilities. The basic kNN algorithm uses a great deal of storage space for the training phase, and its execution space is at least as big as its training space. On the contrary, for all non-lazy learners, execution space is usually much smaller than training space, since the resulting classifier is usually a highly condensed summary of the data.

Furthermore, the number of model or runtime parameters to be tuned by the user is an indicator of an algorithm's ease of use. It can help in prior model selection based on the user's priorities and preferences: for a non specialist in data mining, an algorithm with few user-tuned parameters will certainly be more appealing, while a more advanced user might find a large parameter set an opportunity to control the data mining process more closely. As expected, neural networks and SVMs have more parameters than the remaining techniques.

There is general agreement that k-NN is very sensitive to irrelevant features: this characteristic can be explained by the way the algorithm works. In addition, the presence of irrelevant features can make neural network training very inefficient, even impractical.

Logic-based algorithms are all considered very easy to interpret, whereas neural networks and SVMs have notoriously poor interpretability. k-NN is also considered to have very poor interpretability because an unstructured collection of training instances is far from readable, especially if there are many of them.

While interpretability concerns the typical classifier generated by a learning algorithm, transparency refers to whether the principle of the method is easily understood. A particularly eloquent case is that of k-NN; while the resulting classifier is not quite interpretable, the method itself is very transparent because it appeals to the intuition of human users, who spontaneously reason in a similar manner. Similarly, Naive Bayes' is very transparent, as it is easily grasped by users like physicians who find that probabilistic explanations replicate their way of diagnosing. Moreover, decision trees and rules are credited with high transparency.

No single learning algorithm can uniformly outperform other algorithms over all datasets. When faced with the decision "Which algorithm will be most accurate on our classification problem?", the simplest approach is to estimate the accuracy of the candidate algorithms on the problem and select the one that appears to be most accurate. The concept of combining classifiers is proposed as a new direction for the improvement of the performance of individual classifiers. The goal of classification result integration algorithms is to generate more certain, precise and accurate system results. The following section provides a brief survey of methods for constructing ensembles.

7. Combining Classifiers

Numerous methods have been suggested for the creation of ensemble of classifiers [53]. Although or perhaps because many methods of ensemble creation have been proposed, there is as yet no clear picture of which method is best [54]. Thus, an active area of research in supervised learning is the study of methods for the construction of good ensembles of classifiers. Mechanisms that are used to build ensemble of classifiers include: i) using different subsets of training data with a single learning method, ii) using different training parameters with a single training method (e.g., using different initial weights for each neural network in an ensemble) and iii) using different learning methods.

7.1. Different Subsets of Training Data with a Single Learning Method

Bagging is a method for building ensembles that uses different subsets of training data with a single learning method [55]. Given a training set of size t, bagging draws t random instances from the dataset with replacement (i.e. using a uniform distribution). These t instances are learned, and this process is repeated several times. Since the draw is with replacement, usually the instances drawn will contain some duplicates and some omissions, as compared to the original training set. Each cycle through the process results in one classifier. After the construction of several classifiers, taking a vote of the predictions of each classifier produces the final prediction.

Breiman [55] made the important observation that instability (responsiveness to changes in the training data) is a prerequisite for bagging to be effective. A committee of classifiers that all agree in all circumstances will give identical performance to any of its members in isolation. A variance reduction process will have no effect if there is no variance. If there is too little data, the gains achieved via a bagged ensemble cannot compensate for the decrease in accuracy of individual models, each of which now considers an even smaller training set. On the other end, if the dataset is extremely large and computation time is not an issue, even a single flexible classifier can be quite adequate.

Another method that uses different subsets of training data with a single learning method is the boosting approach [56]. Boosting is similar in overall structure to bagging, except that it keeps track of the performance of the learning algorithm and concentrates on instances that have not been correctly learned. Instead of choosing the t training instances randomly using a uniform distribution, it chooses the training instances in such a manner as to favor the instances that have not been accurately learned.

After several cycles, the prediction is performed by taking a weighted vote of the predictions of each classifier, with the weights being proportional to each classifier's accuracy on its training set.

AdaBoost is a practical version of the boosting approach [56]. Adaboost requires less instability than bagging, because Adaboost can make much larger changes in the training set. A number of studies that compare AdaBoost and bagging suggest that AdaBoost and bagging have quite different operational profiles [57,58]. In general, it appears that bagging is more consistent, increasing the error of the base learner less frequently than does AdaBoost. However, AdaBoost appears to have greater average effect, leading to substantially larger error reductions than bagging on average.

A number of recent studies have shown that the decomposition of a classifier's error into bias and variance terms can provide considerable insight into the prediction performance of the classifier [57]. Bias measures the contribution to error of the central tendency of the classifier when trained on different data. Variance is a measure of the contribution to error of deviations from the central tendency. Generally, bagging tends to decrease variance without unduly affecting bias [55,57]. On the contrary, in empirical studies AdaBoost appears to reduce both bias and variance [55,57]. Thus, AdaBoost is more effective at reducing bias than bagging, but bagging is more effective than AdaBoost at reducing variance.

The decision on limiting the number of sub-classifiers is important for practical applications. To be competitive, it is important that the algorithms run in reasonable time. Quinlan [58] used only 10 replications, while Bauer & Kohavi [57] used 25 replications, Breiman [55] used 50 and Freund and Schapire [56] used 100. For both bagging and boosting, much of the reduction in error appears to have occurred after ten to fifteen classifiers. However, Adaboost continues to measurably improve test-set error until around 25 classifiers for decision trees [59].

As mentioned in Bauer and Kohavi [57], the main problem with boosting seems to be robustness to noise. This is expected because noisy instances tend to be misclassified, and the weight will increase for these instances. They presented several cases where the performance of boosted algorithms degraded compared to the original algorithms. On the contrary, they pointed out that bagging improves the accuracy in all datasets used in the experimental evaluation.

MultiBoosting [60] is another method of the same category. It can be conceptualized wagging committees formed by AdaBoost. Wagging is a variant of bagging: bagging uses resampling to get the datasets for training and producing a weak hypothesis, whereas wagging uses reweighting for each training instance, pursuing the effect of bagging in a different way. Webb [60], in a number of experiments, showed that MultiBoost achieved greater mean error reductions than any of AdaBoost or bagging decision trees in both committee sizes that were investigated (10 and 100).

Another meta-learner, DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples), was presented by [61]. This method uses a learner (one that provides high accuracy on the training data) to build a diverse committee. This is accomplished by adding different randomly-constructed examples to the training set when building new committee members. These artificially constructed examples are given category labels that disagree with the current decision of the committee, thereby directly increasing diversity when a new classifier is trained on the augmented data and added to the committee.

7.2. Different Training Parameters with a Single Training Method

There are also methods for creating ensembles, which produce classifiers that disagree on their predictions. Generally, these methods focus on altering the training process in the hope that the resulting classifiers will produce different predictions. For example, neural network techniques that have been employed include methods for training with different topologies, different initial weights and different parameters [62].

Another effective approach for generation of a set of base classifiers is ensemble feature selection. Ensemble feature selection is finding a set of feature subsets for generation of the base classifiers for an ensemble with one learning algorithm. Ho [63] has shown that simple random selection of feature subsets may be an effective technique for ensemble feature selection. This technique is called the random subspace method (RSM). In the RSM, one randomly selects $N^* < N$ features from the N-dimensional dataset. By this, one obtains the N^*-dimensional random subspace of the original N-dimensional feature space. This is repeated S times so as to get S feature subsets for constructing the base classifiers. Then, one constructs classifiers in the random subspaces and aggregates them in the final integration procedure. An experiment with a systematic partition of the feature space, using nine different combination schemes, was performed by [64], showing that there are no "best" combinations for all situations and that there is no assurance that in all cases that a classifier team will outperform the single best individual.

7.3. Different Learning Methods

Voting denotes the simplest method of combining predictions from multiple classifiers [65]. In its simplest form, called plurality or majority voting, each classification model contributes a single vote. The collective prediction is decided by the majority of the votes, i.e., the class with the most votes is the final prediction. In weighted voting, on the other hand, the classifiers have varying degrees of influence on the collective prediction that is relative to their predictive accuracy. Each classifier is associated with a specific weight determined by its performance (e.g., accuracy, cost model) on a validation set. The final prediction is decided by summing up all weighted votes and by choosing the class with the highest aggregate. Kotsiantis and Pintelas [66] combined the advantages of classifier fusion and dynamic selection. The algorithms that are initially used to build the ensemble are tested on a small subset of the training set and, if they have statistically worse accuracy than the most accurate algorithm, do not participate in the final voting.

Except for voting, stacking [67] aims to improve efficiency and scalability by executing a number of learning processes and combining the collective results. The main difference between voting and stacking is that the latter combines base classifiers in a non-linear fashion. The combining task, called a meta-learner, integrates the independently-computed base classifiers into a higher level classifier, a meta-classifier, by relearning the meta-level training set. This meta-level training set is created by using the base classifiers' predictions on the validation set as attribute values and the true class as the target. Ting and Witten [67] have shown that successful stacked generalization requires the use of output class distributions rather than class predictions. In their experiments, only the MLR algorithm (a linear discriminant) was suitable for use as a level-1 classifier.

Cascade Generalization [68] is another algorithm that belongs to the family of stacking algorithms. Cascade Generalization uses the set of classifiers sequentially, at each step performing an extension of the original data by the insertion of new attributes. The new attributes are derived from the probability class distribution given by a base classifier. This constructive step extends the representational language for the high level classifiers, reducing their bias.

Todorovski & Dzeroski [69] introduced meta-decision trees (MDTs). Instead of giving a prediction, MDT leaves specify which classifier should be used to obtain a prediction. Each leaf of the MDT represents a part of the dataset, which is a relative area of expertise of the base-level classifier in that leaf. MDTs can use the diversity of the base-level classifiers better than voting, thus outperforming voting schemes in terms of accuracy, especially in domains with a high diversity of errors made by base-level classifiers.

Another attempt to improve classification accuracy is the use of hybrid techniques. Lazkano and Sierra [70] presented a hybrid classifier that combines Bayesian Network algorithm with the Nearest Neighbor distance based algorithm. The Bayesian Network structure is obtained from the data and the Nearest Neighbor algorithm is used in combination with the Bayesian Network in the deduction phase.

LiMin et al. [71] presented Flexible NBTree: a decision tree learning algorithm in which nodes contain univariate splits as do regular decision trees, but the leaf nodes contain General Naive Bayes, which is a variant of the standard Naive Bayesian classifier. Zhou and Chen [72] generated a binary hybrid decision tree according to the binary information gain ratio criterion. If attributes cannot further distinguish training examples falling into a leaf node whose diversity is beyond the diversity threshold, then the node is marked as a dummy node and a feed-forward neural network named FANNC is then trained in the instance space defined by the used attributes. Zheng and Webb [73] proposed the application of lazy learning techniques to Bayesian induction and presented the resulting lazy Bayesian rule learning algorithm, called LBR. This algorithm can be justified by a variant of the Bayes model, which supports a weaker conditional attribute independence assumption than is required by naive Bayes. For each test example, it builds a most appropriate rule with a local naive Bayesian classifier as its consequent. Zhipeng et al. [74] proposed a similar lazy learning algorithm: Selective Neighborhood based Naive Bayes (SNNB). SNNB computes different distance neighborhoods of the input new object, lazily learns multiple Naive Bayes classifiers, and uses the classifier with the highest estimated accuracy to make the final decision. Domeniconi and Gunopulos [75] combined local learning with SVMs. In this approach an SVM is used to determine the weights of the local neighborhood instances.

8. Special Learning Problems

In this section, we present two special classification problems: a) learning from imbalanced datasets and b) learning from multimedia files.

8.1. Handling Imbalanced Datasets

Learning classifiers from imbalanced or skewed datasets is an important topic, arising very often in practice in classification problems. In such problems, almost all the instances are labelled as one class, while far fewer instances are labelled as the other class, usually the more important class. It is obvious that traditional classifiers seeking an accurate performance over a full range of instances are not suitable to deal with imbalanced learning tasks, since they tend to classify all the data into the majority class, which is usually the less important class. The relationship between training set size and improper classification performance for imbalanced data sets seems to be that on small imbalanced data sets the minority class is poorly represented by an excessively reduced number of examples that might not be sufficient for learning, especially when a large degree of class overlapping exists and the class is further divided into sub-clusters.

The problem of imbalance has got more and more emphasis in recent years. Imbalanced data sets exists in many real-world domains, such as spotting unreliable telecommunication customers, detection of oil spills in satellite radar images, detection of fraudulent telephone calls, information retrieval and filtering tasks, and so on. A number of solutions to the class-imbalance problem are proposed both at the data and algorithmic levels. At the data level [76], these solutions include many different forms of re-sampling such as random over-sampling of minority class with replacement, random under-sampling of majority class, directed over-sampling (in which no new examples are created, but the choice of samples to replace is informed rather than random), directed under-sampling (where, again, the choice of examples to eliminate is informed), over-sampling with informed generation of new samples, and combinations of the above techniques. At the algorithmic level [77], solutions include adjusting the costs of the various classes so as to counter the class imbalance, adjusting the probabilistic estimate at the tree leaf (when working with decision trees), adjusting the decision threshold, and recognition-based (i.e., learning from one class) rather than discrimination-based (two class) learning. Mixture-of-experts approaches [78] (combining methods) have been also used to handle class-imbalance problems. These methods combine the results of many classifiers; each usually induced after over-sampling or under-sampling the data with different over/under-sampling rates. Gary Weiss [79] presents an extensive overview of the field of learning from imbalanced data.

8.2. Multimedia Mining

Generally, multimedia database systems store and manage a large collection of multimedia objects, such as image, video, audio and hypertext data. Thus, in multimedia documents, knowledge discovery deals with non-structured information. For this reason, we need tools for discovering relationships between objects or segments within multimedia document components, such as classifying images based on their content, extracting patterns in sound, categorizing speech and music, and recognizing and tracking objects in video streams. These multimedia files undergo various transformations and features extraction to generate the important features from the multimedia files. With the generated features, mining can be carried out using the well kwon machine learning techniques to discover significant patterns.

Text categorization is a conventional classification problem applied to the textual domain. It solves the problem of assigning text content to predefined categories. In the learning stage, the labeled training data are first pre-processed to remove unwanted details and to "normalize" the data [80]. For example, in text documents punctuation symbols and non-alphanumeric characters are usually discarded, because they do not help in classification. Moreover, all characters are usually converted to lower case to simplify matters. The next step is to compute the features that are useful to distinguish one class from another. For a text document, this usually means identifying the keywords that summarize the contents of the document.

Image categorization classifies images into semantic databases that are manually pre-categorized. In the same semantic databases, images may have large variations with dissimilar visual descriptions (e.g. images of persons, images of industries etc.). In addition, images from different semantic databases might share a common background (some flowers and sunset have similar colors). In [81] the authors distinguish three types of feature vectors for image description: 1) pixel level features, 2) region level features, and 3) tile level features.

Audio data play an important role in multimedia applications. Music information has two main branches: symbolic and audio information. Attack, duration, volume, velocity and instrument type of every single note are available information. Therefore, it is possible to easily access statistical measures such as tempo and mean key for each music item [82].

In video mining, there are three types of videos: a) the produced (e.g. movies, news videos, and dramas), b) the raw (e.g. traffic videos, surveillance videos etc), and c) the medical video (e.g. ultra sound videos including echocardiogram). The first stage for mining raw video data is grouping input frames to a set of basic units, which are relevant to the structure of the video. In produced videos, the most widely used basic unit is a shot, which is defined as a collection of frames recorded from a single camera operation. Shot detection methods can be classified into many categories: pixel based, statistics based, transform based, feature based and histogram based [83]. Color or grayscale histograms (such as in image mining) can also be used [84]. To segment video, color histograms, as well as motion and texture features can be used [85].

Compared with data mining, multimedia mining reaches much higher complexity resulting from: a) the huge volume of data, b) the variability and heterogeneity of the multimedia data (e.g. diversity of sensors, time or conditions of acquisition etc) and c) the multimedia content's meaning is subjective.

9. Conclusions

This paper describes the best-know supervised techniques in relative detail. We should remark that our list of references is not a comprehensive list of papers discussing supervised methods: our aim was to produce a critical review of the key ideas, rather than a simple list of all publications which had discussed or made use of those ideas. Despite this, we hope that the references cited cover the major theoretical issues, and provide access to the main branches of the literature dealing with such methods, guiding the researcher in interesting research directions.

The key question when dealing with ML classification is not whether a learning algorithm is superior to others, but under which conditions a particular method can significantly outperform others on a given application problem. Meta-learning is moving in this direction, trying to find functions that map datasets to algorithm performance [86]. To this end, meta-learning uses a set of attributes, called meta-attributes, to represent the characteristics of learning tasks, and searches for the correlations between these attributes and the performance of learning algorithms. Some characteristics of learning tasks are: the number of instances, the proportion of categorical attributes, the proportion of missing values, the entropy of classes, etc. Brazdil et al. [87] provided an extensive list of information and statistical measures for a dataset.

After a better understanding of the strengths and limitations of each method, the possibility of integrating two or more algorithms together to solve a problem should be investigated. The objective is to utilize the strengths of one method to complement the weaknesses of another. If we are only interested in the best possible classification accuracy, it might be difficult or impossible to find a single classifier that performs as well as a good ensemble of classifiers. Mechanisms that are used to build ensemble of classifiers include: i) using different subsets of training data with a single learning method, ii) using different training parameters with a single training method (e.g., using different initial weights for each neural network in an ensemble) and iii) using different learning methods.

Despite the obvious advantages, ensemble methods have at least three weaknesses. The first weakness is increased storage as a direct consequence of the requirement that all component classifiers, instead of a single classifier, need to be stored after training. The total storage depends on the size of each component classifier itself and the size of the ensemble (number of classifiers in the ensemble). The second weakness is increased computation because in order to classify an input query, all component classifiers (instead of a single classifier) must be processed. The last weakness is decreased comprehensibility. With involvement of multiple classifiers in decision-making, it is more difficult for non-expert users to perceive the underlying reasoning process leading to a decision. A first attempt for extracting meaningful rules from ensembles was presented in [88].

For all these reasons, the application of ensemble methods is suggested only if we are only interested in the best possible classification accuracy. Another time-consuming attempt that tried to increase classification accuracy without decreasing comprehensibility is the wrapper feature selection procedure [89]. Theoretically, having more features should result in more discriminating power. However, practical experience with machine learning algorithms has shown that this is not always the case. Wrapper methods wrap the feature selection around the induction algorithm to be used, using cross-validation to predict the benefits of adding or removing a feature from the feature subset used.

Finally, many researchers in machine learning are accustomed to dealing with flat files and algorithms that run in minutes or seconds on a desktop platform. For these researchers, 100,000 instances with two dozen features is the beginning of the range of "very large" datasets. However, the database community deals with gigabyte databases. Of course, it is unlikely that all the data in a data warehouse would be mined simultaneously. Most of the current learning algorithms are computationally expensive and require all data to be resident in main memory, which is clearly untenable for many realistic problems and databases. An orthogonal approach is to partition the data, avoiding the need to run algorithms on very large datasets. Distributed machine learning involves breaking the dataset up into subsets, learning from these subsets concurrently and combining the results [90]. Distributed agent systems can be used for this parallel execution of machine learning processes [91]. Non-parallel machine learning algorithms can still be applied on local data (relative to the agent) because information about other data sources is not necessary for local operations. It is the responsibility of agents to integrate the information from numerous local sources in collaboration with other agents.

References

[1] S. Kotsiantis, P. Pintelas, Recent Advances in Clustering: A Brief Survey, WSEAS Transactions on Information Science and Applications, Vol. 1, No 1 (73–81), 2004.

[2] Gosavi, A., Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning, Series: Operations Research/Computer Science Interfaces Series, Vol. 25, 2003, Springer.

[3] Dutton, D. & Conroy, G. (1996), A review of machine learning, Knowledge Engineering Review 12: 341–367.

[4] De Mantaras & Armengol E. (1998). Machine learning from examples: Inductive and Lazy methods. Data & Knowledge Engineering 25: 99–123.

[5] Zhang, S., Zhang, C., Yang, Q. (2002). Data Preparation for Data Mining. Applied Artificial Intelligence, Volume 17, pp. 375–381.

[6] Hodge, V., Austin, J. (2004), A Survey of Outlier Detection Methodologies, Artificial Intelligence Review, Volume 22, Issue 2, pp. 85–126.

[7] Batista, G., & Monard, M.C., (2003), An Analysis of Four Missing Data Treatment Methods for Supervised Learning, Applied Artificial Intelligence, vol. 17, pp. 519–533.

[8] Yu, L., Liu, H. (2004), Efficient Feature Selection via Analysis of Relevance and Redundancy, JMLR, 5(Oct):1205–1224.

[9] Markovitch S. & Rosenstein D. (2002), Feature Generation Using General Construction Functions, Machine Learning 49: 59–98.

[10] Murthy, (1998), Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey, Data Mining and Knowledge Discovery 2: 345–389.

[11] Elomaa T. (1999). The biases of decision tree pruning strategies. Lecture Notes in Computer Science 1642. Springer, pp. 63–74.

[12] Quinlan, J.R. (1993). C4.5: Programs for machine learning. Morgan Kaufmann, San Francisco.

[13] Tjen-Sien, L., Wei-Yin, L., Yu-Shan, S. (2000). A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms. Machine Learning 40: 203–228.

[14] Gehrke, J., Ramakrishnan, R. & Ganti, V. (2000), RainForest—A Framework for Fast Decision Tree Construction of Large Datasets, Data Mining and Knowledge Discovery, Volume 4, Issue 2–3, Jul 2000, Pages 127–162.

[15] Zheng, Z. (1998). Constructing conjunctions using systematic search on decision trees. Knowledge Based Systems Journal 10: 421–430.

[16] Zheng, Z. (2000). Constructing X-of-N Attributes for Decision Tree Learning. Machine Learning 40: 35–75.

[17] Gama, J. & Brazdil, P. (1999). Linear Tree. Intelligent Data Analysis 3: 1–22.

[18] Baik, S. Bala, J. (2004), A Decision Tree Algorithm for Distributed Data Mining: Towards Network Intrusion Detection, Lecture Notes in Computer Science, Volume 3046, Pages 206–212.

[19] Furnkranz, J. (1999). Separate-and-Conquer Rule Learning. Artificial Intelligence Review 13: 3–54.

[20] An, A., Cercone, N. (2000), Rule Quality Measures Improve the Accuracy of Rule Induction: An Experimental Approach, Lecture Notes in Computer Science, Volume 1932, Pages 119–129.

[21] Lindgren, T. (2004), Methods for Rule Conflict Resolution, Lecture Notes in Computer Science, Volume 3201, Pages 262–273.

[22] Zhang, G. (2000), Neural networks for classification: a survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C 30(4): 451–462.

[23] Camargo, L.S. & Yoneyama, T. (2001). Specification of Training Sets and the Number of Hidden Neurons for Multilayer Perceptrons. Neural Computation 13: 2673–2680.

[24] Kon, M. & Plaskota, L. (2000), Information complexity of neural networks, Neural Networks 13: 365–375.

[25] Neocleous, C. & Schizas, C., (2002), Artificial Neural Network Learning: A Comparative Review, LNAI 2308, pp. 300–313, Springer-Verlag Berlin Heidelberg.

[26] Yam, J. & Chow, W. (2001). Feedforward Networks Training Speed Enhancement by Optimal Initialization of the Synaptic Coefficients. IEEE Transactions on Neural Networks 12: 430–434.

[27] Siddique, M.N.H. and Tokhi, M.O. (2001), Training Neural Networks: Backpropagation vs. Genetic Algorithms, IEEE International Joint Conference on Neural Networks, Vol. 4, pp. 2673–2678.

[28] Yen, G.G. and Lu, H. (2000), Hierarchical genetic algorithm based neural network design, In: IEEE Symposium on Combinations of Evolutionary Computation and Neural Networks, pp. 168–175.

[29] Eklund, P., Hoang, A. (2002), A Performance Survey of Public Domain Machine Learning Algorithms Technical Report, School of Information Technology, Griffith University.

[30] Zhou, Z. (2004), Rule Extraction: Using Neural Networks or For Neural Networks?, Journal of Computer Science and Technology, Volume 19, Issue 2, Pages: 249–253.

[31] Jensen, F. (1996). An Introduction to Bayesian Networks. Springer.
[32] Madden, M. (2003), The Performance of Bayesian Network Classifiers Constructed using Different Techniques, Proceedings of European Conference on Machine Learning, Workshop on Probabilistic Graphical Models for Classification, pp. 59–70.
[33] Chickering, D.M. (2002). Optimal Structure Identification with Greedy Search. Journal of Machine Learning Research, Vol. 3, pp 507–554.
[34] Acid, S. and de Campos. L.M. (2003). Searching for Bayesian Network Structures in the Space of Restricted Acyclic Partially Directed Graphs. Journal of Artificial Intelligence Research 18: 445–490.
[35] Cowell, R.G. (2001). Conditions Under Which Conditional Independence and Scoring Methods Lead to Identical Selection of Bayesian Network Models. Proc. 17th International Conference on Uncertainty in Artificial Intelligence.
[36] Cheng, J., Greiner, R., Kelly, J., Bell, D., & Liu, W. (2002). Learning Bayesian networks from data: An information-theory based approach. Artificial Intelligence 137: 43–90.
[37] Yang, Y., Webb, G. (2003), On Why Discretization Works for Naive-Bayes Classifiers, Lecture Notes in Computer Science, Volume 2903, Pages 440–452.
[38] Bouckaert, R. (2004), Naive Bayes Classifiers That Perform Well with Continuous Variables, Lecture Notes in Computer Science, Volume 3339, Pages 1089–1094.
[39] Mitchell, T. (1997). Machine Learning. McGraw Hill.
[40] Aha, D. (1997). Lazy Learning. Dordrecht: Kluwer Academic Publishers.
[41] Witten, I. & Frank, E. (2000). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, San Mateo, 2000.
[42] Wettschereck, D., Aha, D.W. & Mohri, T. (1997). A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms. Artificial Intelligence Review 10:1–37.
[43] Okamoto, S., Yugami, N. (2003), Effects of domain characteristics on instance-based learning algorithms. Theoretical Computer Science 298, 207–233.
[44] Sanchez, J., Barandela, R., Ferri, F. (2002), On Filtering the Training Prototypes in Nearest Neighbour Classification, Lecture Notes in Computer Science, Volume 2504, Pages 239–248.
[45] Burges, C. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. 2(2):1–47.
[46] Cristianini, N. & Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge.
[47] Veropoulos, K., Campbell, C. & Cristianini, N. (1999). Controlling the Sensitivity of Support Vector Machines. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI99).
[48] Scholkopf, C., Burges, J.C. & Smola, A.J. (1999). Advances in Kernel Methods. MIT Press.
[49] Genton, M. (2001). Classes of Kernels for Machine Learning: A Statistics Perspective. Journal of Machine Learning Research 2: 299–312.
[50] Platt, J. (1999). Using sparseness and analytic QP to speed training of support vector machines. In Kearns, M., Solla, S. & Cohn, D. (ed.), Advances in neural information processing systems. MIT Press.
[51] Keerthi, S. & Gilbert, E. (2002). Convergence of a Generalized SMO Algorithm for SVM Classifier Design. Machine Learning 46: 351–360.
[52] Crammer, K. & Singer, Y. (2002). On the Learnability and Design of Output Codes for Multiclass Problems. Machine Learning 47: 201–233.
[53] Dietterich, T.G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization, Machine Learning 40: 139–157.
[54] Villada, R. & Drissi, Y. (2002). A Perspective View and Survey of Meta-Learning. Artificial Intelligence Review 18: 77–95.
[55] Breiman, L., Bagging Predictors. Machine Learning, 24 (1996) 123–140.
[56] Freund, Y. & Schapire, R. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. JCSS 55(1): 119–139.
[57] Bauer, E. & Kohavi, R. (1999). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning, Vol. 36, pp. 105–139.
[58] Quinlan, J.R. (1996), Bagging, boosting, and C4.5. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, pp. 725–730, AAAI/MIT Press.
[59] Opitz D. & Maclin R. (1999), Popular Ensemble Methods: An Empirical Study, Artificial Intelligence Research, 11: 169–198, Morgan Kaufmann.
[60] Webb G.I. (2000), MultiBoosting: A Technique for Combining Boosting and Wagging, Machine Learning, 40, 159–196, Kluwer Academic Publishers.
[61] Melville, P., Mooney, R. (2003), Constructing Diverse Classifier Ensembles using Artificial Training Examples, Proc. of the IJCAI-2003, pp. 505–510, Acapulco, Mexico.
[62] Maclin & Shavlik (1995), Combining the Prediction of multiple classifiers: Using competitive learning to initialize ANNs. In the Proc of the 14th International Joint Conference on AI, 524–530.

[63] Ho, T.K. (1998): The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20: 832–844.
[64] Kuncheva, L., Whitaker, C. (2001). Feature Subsets for Classifier Combination: An Enumerative Experiment. volume 2096 of Lecture Notes in Computer Science, pages 228–237. Springer-Verlag.
[65] Roli, F., Giacinto, G., Vernazza, G. (2001). Methods for Designing Multiple Classifier Systems. Volume 2096 of Lecture Notes in Computer Science, pages 78–87. Springer-Verlag.
[66] Kotsiantis, S., Pintelas, P. (2004), Selective Voting, In Proc. of the 4th International Conference on Intelligent Systems Design and Applications (ISDA 2004), August 26–28, Budapest, pp. 397–402.
[67] Ting, K. & Witten, I. (1999). Issues in Stacked Generalization. Artificial Intelligence Research 10: 271–289.
[68] Gama, J. & Brazdil, P. (2000). Cascade Generalization. Machine Learning 41: 315–343.
[69] Todorovski, L., Dzeroski, S. (2003). Combining Classifiers with Meta Decision Trees. Machine Learning 50: 223–249.
[70] Lazkano, E., Sierra, B. (2003), BAYES-NEAREST: A New Hybrid Classifier Combining Bayesian Network and Distance Based Algorithms, Lecture Notes in Computer Science, Volume 2902, Pages 171–183.
[71] LiMin Wang, SenMiao Yuan, Ling Li, HaiJun Li (2004), Improving the Performance of Decision Tree: A Hybrid Approach, Lecture Notes in Computer Science, Volume 3288, Pages 327–335.
[72] Zhou, Z., Chen, Z. (2002), Hybrid decision tree. Knowl.-Based Syst. 15(8): 515–528.
[73] Zheng, Z., Webb, G. (2000), Lazy Learning of Bayesian Rules, Machine Learning, Volume 41, Issue 1, Pages 53–84.
[74] Zhipeng Xie, Wynne Hsu, Zongtian Liu, Mong Li Lee (2002), SNNB: A Selective Neighborhood Based Naive Bayes for Lazy Learning, Lecture Notes in Computer Science, Volume 2336, Pages 104–115.
[75] Domeniconi, C., Gunopulos, D. (2001), Adaptive Nearest Neighbor Classification using Support Vector Machines, Advances in Neural Information Processing Systems 14, 665–672.
[76] Chawla, N.V., Hall, L.O., Bowyer, K.W., and Kegelmeyer W.P. (2002), SMOTE: Synthetic Minority Oversampling TEchnique. Journal of Artificial Intelligence Research, 16:321–357.
[77] Provost, F., & Fawcett, T. (2001). Robust classification for imprecise environments. Machine Learning, 42, 203–231.
[78] Joshi, M.V., Kumar, V., and Agarwal, R.C., (2001). Evaluating boosting algorithms to classify rare cases: comparison and improvements. In First IEEE International Conference on Data Mining, pages 257–264, November.
[79] Weiss, G. (2004), Mining with rarity: A unifying framework.SIGKDD Explorations, 6(1):7–19, 2004.
[80] Sebastiani F. (2002), Machine learning in automated text categorization. ACM Computing Surveys, Vol. 34, pp. 1–47.
[81] Zhang Ji, Wynne Hsu, Mong Li Lee (2001). Image Mining: Issues, Frameworks and Techniques, in Proc. of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), San Francisco, CA, USA, pp. 13–20.
[82] Liu Z., J. Huang, Y. Wang, and T. Chen (1998), Audio Feature Extraction and Analysis for Scene Segmentation and Classification, Journal of VLSI Signal Processing, Vol. 20, pp. 61–79.
[83] Boresczky J.S. and L.A. Rowe (1996), A comparison of video shot boundary detection techniques, Storage & Retrieval for Image and Video Databases IV, Proc. SPIE 2670, pp. 170–179.
[84] Ardizzone E. and M. Cascia (1997). Automatic video database indexing and retrieval. Multimedia Tools and Applications, Vol. 4, pp. 29–56.
[85] Yu H. and W. Wolf (1997). A visual search system for video and image databases. In Proc. IEEE Int'l Conf. On Multimedia Computing and Systems, Ottawa, Canada, pp. 517–524.
[86] Kalousis A., Gama, G. (2004), On Data and Algorithms: Understanding Inductive Performance, Machine Learning 54: 275–312.
[87] Brazdil P., Soares C. and Da Costa J. (2003), Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results, Machine Learning, 50: 251–277.
[88] Wall, R., Cunningham, P., Walsh, P., Byrne, S. (2003), Explaining the output of ensembles in medical decision support on a case by case basis, Artificial Intelligence in Medicine, Vol. 28(2) 191–206.
[89] Guyon, I., Elissee, A. (2003), An introduction to variable and feature selection. Journal of Machine Learning Research, 3:1157 1182.
[90] Basak, J., Kothari, R. (2004), A Classification Paradigm for Distributed Vertically Partitioned Data. Neural Computation, 16(7): 1525–1544.
[91] Klusch, M., Lodi, S., Moro, G. (2003), Agent-Based Distributed Data Mining: The KDEC Scheme. In Intelligent Information Agents: The AgentLink Perspective, LNAI 2586, pages 104–122. Springer.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Dimension Reduction and Data Visualization Using Neural Networks

Gintautas DZEMYDA [1], Olga KURASOVA, Viktor MEDVEDEV
Institute of Mathematics and Informatics, Vilnius, Lithuania

Abstract. The problem of visual presentation of multidimensional data is discussed. The projection methods for dimension reduction are reviewed. The chapter deals with the artificial neural networks that may be used for reducing dimension and data visualization, too. The stress is put on combining the self-organizing map (SOM) and Sammon mapping and on the neural network for Sammon's mapping SAMANN. Large scale applications are discussed: environmental data analysis, statistical analysis of curricula, comparison of schools, analysis of the economic and social conditions of countries, analysis of data on the fundus of eyes and analysis of physiological data on men's health.

Keywords. Visualization, dimension reduction, neural networks, SOM, SAMANN, data mining.

Introduction

For the effective data analysis, it is important to include a human in the data exploration process and combine flexibility, creativity, and general knowledge of the human with the enormous storage capacity and computational power of today's computer. Visual data mining aims at integrating the human in the data analysis process, applying human perceptual abilities to the analysis of large data sets available in today's computer systems [1].

Objects from the real world are frequently described by an array of parameters (variables, features) $x_1, x_2, ..., x_n$. The term "object" can cover various things: people, equipment, produce of manufacturing, etc. Any parameter may take some numerical values. A combination of values of all parameters characterizes a particular object $X_j = (x_{j1}, x_{j2}, ..., x_{jn})$ from the whole set $X_1, X_2, ..., X_m$, where n is the number of parameters and m is the number of analysed objects. $X_1, X_2, ..., X_m$ are the n-dimensional vectors (data). Often they are interpreted as points in the n-dimensional space R^n, where n defines the dimensionality of the space. In fact, we have a table of numerical data for the analysis: $\{x_{ji}, j = 1, ..., m, i = 1, ..., n\}$. A natural idea arises to present multidimensional data, stored in such a table, in some visual form. It is a complicated problem followed by extensive researches, but its solution allows the human to gain a deeper insight into the data, draw conclusions, and directly interact with the data. In Figure 1, we present an example of visual presentation of the data table (n=6, m=20) using multidimensional scaling method, discussed below. The dimensionality of data is reduced from 6 to 2. Here vectors X_4, X_6, X_8, and X_{19}

---

[1] Corresponding Author: Institute of Mathematics and Informatics, Akademijos Str. 4, LT 08663, Vilnius, Lithuania; E-mail: Dzemyda@ktl.mii.lt.

form a separate cluster that can be clearly observed visually on a plane and that cannot be recognized directly from the table without a special analysis.

This chapter is organized as follows. In Section 1, an overview of methods for multidimensional data visualization via dimension reduction is presented. The methods based on neural networks are discussed, too. Section 2 deals with a problem of combining the SOM and Sammon mapping. Investigations of the neural network for Sammon's mapping SAMANN are discussed in Section 3. Applications of the dimension reduction and data visualization using neural networks are presented in Section 4. They cover the visual analysis of correlations: environmental data analysis and statistical analysis of curricula; multidimensional data visualization applications: comparison of schools and analysis of the economic and social conditions of Central European countries; visual analysis of medical data: analysis of data on the fundus of eyes and analysis of physiological data on men's health. Conclusions generalize the results.

	x_1	x_2	x_3	x_4	x_5	x_6
1	0,57	0,11	0,47	0,54	0,84	0,53
2	0,65	0,14	0,53	0,55	0,89	0,54
3	0,61	0,13	0,49	0,52	0,92	0,57
4	0,44	0,16	0,45	0,42	0,99	0,45
5	0,55	0,11	0,53	0,52	0,92	0,55
6	0,44	0,16	0,43	0,41	0,98	0,43
7	0,49	0,12	0,56	0,46	0,82	0,48
8	0,41	0,17	0,42	0,43	0,98	0,42
9	0,54	0,12	0,59	0,46	0,85	0,48
10	0,69	0,16	0,63	0,56	0,91	0,49
11	0,66	0,16	0,51	0,48	0,87	0,55
12	0,70	0,13	0,52	0,48	0,81	0,52
13	0,65	0,14	0,63	0,52	0,89	0,49
14	0,59	0,11	0,51	0,57	0,89	0,52
15	0,53	0,11	0,52	0,57	0,89	0,50
16	0,52	0,13	0,51	0,61	0,92	0,51
17	0,58	0,12	0,55	0,61	0,95	0,52
18	0,45	0,10	0,52	0,53	0,92	0,51
19	0,42	0,17	0,41	0,42	0,98	0,40
20	0,53	0,11	0,42	0,57	0,91	0,57

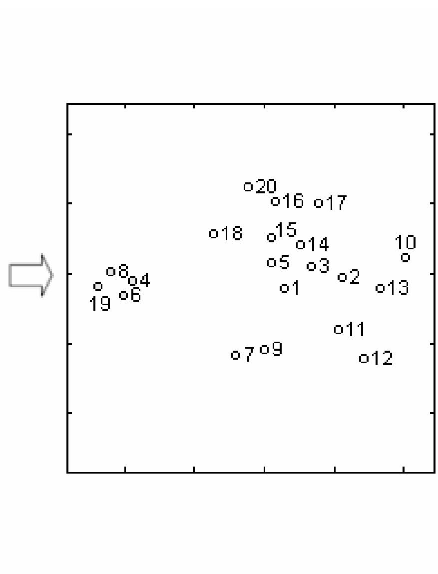


Figure 1. Visualization power

1. Overview of the Dimension Reduction Methods

We discuss possible approaches in visualizing the vectors $X_1,...,X_m \in R^n$ below. A large class of methods has been developed for the direct data visualization. It is a graphical presentation of the data set providing a quality understanding of the information contents in a natural and direct way: parallel coordinates, scatterplots, Chernoff faces, dimensional stacking, etc. (see [2], [3]). Another way is to reduce the dimensionality of data. There exist a lot of so-called projection methods that can be used for reducing the dimensionality, and, particularly, for visualizing the n-dimensional vectors $X_1,...,X_m \in R^n$. A deep review of the methods is performed e.g. by Kaski [4], Kohonen [5], and Kurasova [6]. The discussion below is based mostly on these reviews. The discussion shows the place of neural networks in the general context of methods for reducing the dimensionality of data.

The goal of projection method is to represent the input data items in a lower-dimensional space so that certain properties of the structure of the data set were preserved as faithfully as possible. The projection can be used to visualize the data set if a sufficiently small output dimensionality is chosen. One of these methods is a principal component analysis (PCA). The well-known principal component analysis [7] can be used to display the data as a linear projection on such a subspace of the original data space that best preserves the variance in the data. Effective algorithms exist for computing the projection, even neural algorithms (see e.g. [8], [9]). The PCA cannot embrace nonlinear structures, consisting of arbitrarily shaped clusters or curved manifolds, since it describes the data in terms of a linear subspace. Projection pursuit [10] tries to express some nonlinearities, but if the data set is high-dimensional and highly nonlinear, it may be difficult to visualize it with linear projections onto a low-dimensional display even if the "projection angle" is chosen carefully.

Several approaches have been proposed for reproducing nonlinear higher-dimensional structures on a lower-dimensional display. The most common methods allocate a representation for each data point in a lower-dimensional space and try to optimize these representations so that the distances between them are as similar as possible to the original distances of the corresponding data items. The methods differ in that how the different distances are weighted and how the representations are optimized. Multidimensional scaling (MDS) refers to a group of methods that is widely used [11]. The starting point of MDS is a matrix consisting of pairwise dissimilarities of the data vectors. In general, the dissimilarities need not be distances in the mathematically strict sense. There exists a multitude of variants of MDS with slightly different cost functions and optimization algorithms. The first MDS for metric data was developed in the 1930s: historical treatments and introductions to MDS have been provided by, for example, [12], [13], and later on generalized for analysing nonmetric data. The MDS algorithms can be roughly divided into two basic types: metric and nonmetric MDS. The goal of projection in the metric MDS is to optimize the representations so that the distances between the items in the lower-dimensional space would be as close to the original distances as possible. Denote the distance between the vectors X_i and X_j in the feature space R^n by d_{ij}^*, and the distance between the same vectors in the projected space R^d by d_{ij}. In our case, the initial dimensionality is n, and the resulting one (denote it by d) is 2. The metric MDS tries to approximate d_{ij} by d_{ij}^*. If a square-error cost is used, the objective function to be minimized can be written as

$$E_{MDS} = \sum_{\substack{i,j=1\\ i<j}}^{m} w_{ij}(d_{ij}^* - d_{ij})^2 . \tag{1}$$

The weights are frequently used: $w_{ij} = \dfrac{1}{\sum\limits_{\substack{k,l=1\\ k<l}}^{m} (d_{kl}^*)^2}$, $w_{ij} = \dfrac{1}{d_{ij}^* \sum\limits_{\substack{k,l=1\\ k<l}}^{m} d_{kl}^*}$, $w_{ij} = \dfrac{1}{m d_{ij}^*}$.

Usually the Euclidean distances are used for d_{ij} and d_{ij}^*. However, a perfect reproduction of Euclidean distances may not always be the best possible goal, especially if the components of the data vectors are expressed on an ordinal scale. Then, only the rank order of the distances between the vectors is meaningful, not the exact values. For solving this problem, the nonmetric MDS [11] may be used. Although the nonmetric MDS was motivated by the need of treating ordinal-scale data, it can also be used, of course, if the inputs are presented as pattern vectors in an Euclidean space. The projection then only tries to preserve the order of distances between the data vectors, not their actual values.

A particular case of the metric MDS is Sammon's mapping [14]. It tries to optimize a cost function that describes how well the pairwise distances in a data set are preserved:

$$E_S = \frac{1}{\sum\limits_{\substack{i,j=1\\ i<j}}^{m} d_{ij}^*} \sum_{\substack{i,j=1\\ i<j}}^{m} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*}. \tag{2}$$

Due to the normalization (division by d_{ij}^*), the preservation of small distances is emphasized. The coordinates y_{ik}, $i = 1,...,m,\ k = 1,2$ of the two-dimensional vectors $Y_i = (y_{i1}, y_{i2})$ are computed by the iteration formula [14]:

$$y_{ik}(t+1) = y_{ik}(t) - \alpha \frac{\partial E_S(t)}{\partial y_{ik}(t)} \Bigg/ \left| \frac{\partial^2 E_S(t)}{\partial y_{ik}^2(t)} \right|, \tag{3}$$

where t denotes the iteration order number, and α is a parameter, which influences the optimization step.

The analysis of a relative performance of the different algorithms in reducing the dimensionality of multidimensional vectors starting from the paper [15] indicates Sammon's projection to be still one of the best methods of this class (see also [16]).

There are some other nonlinear methods for data visualization: principal curves, generative topographic mapping, locally linear embedding, etc. Such a variety of methods is determined by the human desire to see the data from a different perspective.

The PCA can be generalized to form nonlinear curves. While in PCA a good projection of a data set onto a linear manifold was constructed, the goal in constructing a principal curve is to project the set onto a nonlinear manifold. The principal curves [17] are smooth curves defined by the property that each point of the curve is the average of all the data points projected to it, i.e. to which that point is the closest one on the curve. Principal curves are generalizations of principal components, extracted using PCA, in the sense that a linear principal curve is a principal component.

Generative topographic mapping (GTM) is a technique in which a topographic mapping function between the input data and the visualization space is found [18]. The idea is to use a function, which maps a density distribution in the visualization space in combination with a Gaussian noise model into the original data space.

Locally linear embedding (LLE) is an unsupervised learning algorithm that computes low-dimensional, neighbourhood preserving embeddings of high-dimensional data. LLE assumes that the high-dimensional data lie on (or near to) a smooth low-dimensional (nonlinear) manifold [19].

1.1. Dimension Reduction Methods Based on Neural Networks

Artificial neural networks may be used for reducing dimension and data visualization. A feed-forward neural network is utilised to effect a topographic, structure-preserving, dimension-reducing transformation of the data, with an additional facility to incorporate different degrees of associated subjective information. The MDS got some attention from neural network researchers [20], [21], [22].

There also exists a neural network architecture designed specifically for topographic mapping, and that is the self-organising map (SOM) [5], which exploits implicit lateral connectivity in the output layer of neurons. The SOM is a method used for both clustering and visualization (dimension reduction) of multidimensional data. Especially two algorithms, the *k*-means clustering [23] and the principal curves, are very closely related to the SOM. An important practical difference between SOM and MDS (in terms of data visualization) is that SOM offers a possibility to visualize new points (that were not used during learning), whereas MDS and Sammon's mapping do not.

In [24], a visualization method, called ViSOM, is proposed. It constrains the lateral contraction force between the neurons in the SOM and hence regularises the inter-neuron distances with respect to a scaleable parameter that defines and controls the resolution of the map.

A curvilinear components analysis (CCA) [25] is proposed as an improvement to self-organizing maps, the output space of which is continuous and takes automatically the relevant shape. The structure of the CCA network consists of two layers: the first one of which performs vector quantization on the dataset, and the second layer, called the projection layer, performs a topographic mapping of the structure obtained by the vector quantization layer. The projection layer is a free space, which takes the shape of a submanifold of the data. The CCA is also claimed to allow an inverse projection, that is, from the two-dimensional space to the n-dimensional space, by a permutation of the input and output layers.

An autoassociative feed-forward neural network [26], [27] allows dimension reduction by extracting the activity of d neurons of the internal "bottleneck" layer (containing fewer nodes than input or output layers), where d is the dimensionality of the visualization space. The network is trained to reproduce the data space, i.e. training data are presented to both input and output layers while obtaining a reduced representation in the inner layer.

The specific neural network model NeuroScale [28] uses a radial basis function neural network (RBF) [29] to transform the n-dimensional input vector to the d-dimensional output vector, where, in general, $n>d$. The other capacity of NeuroScale is to exploit additional knowledge available on the data and to allow this knowledge to influence the mapping. This allows the incorporation of supervisory information to a totally unsupervised technique.

The autoencoder network for dimension reduction is proposed in [30]. It is the nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network to transform the high-dimensional data into a low-dimensional code and a similar "decoder" network to recover the data from the code. It is discovered in [30] that the nonlinear autoencoders work considerably better compared to widely used methods such as PCA or LLE.

The combination and integrated use of data visualization methods of a different nature are under a rapid development. The combination of different methods can be applied to make a data analysis, while minimizing the shortcomings of individual methods. We develop and examine some possibilities: combination of the SOM with Sammon's mapping (see Section 2), and SAMANN algorithm: feed-forward neural network for Sammon's mapping [20] (see Section 3). As an alternative to SAMANN unsupervised learning rule, one could also train a standard feed-forward neural network, using supervised backpropagation on previously calculated Sammon's mapping. Despite that it requires much more computation, as it involves two learning phases (one for Sammon's mapping, one for the neural network), it should perform at least as well as SAMANN [31].

2. Combinations of the SOM and Sammon Mapping

The self-organizing map (SOM) [5] is a class of neural networks that are trained in an unsupervised manner using competitive learning. It is a well-known method for mapping a high-dimensional space onto a low-dimensional one. We consider here the mapping onto a two-dimensional grid of neurons. Let $X_1,...,X_m \in R^n$ be a set of n-dimensional vectors for mapping. Usually, the neurons are connected to each other via a rectangular or hexagonal topology. The rectangular SOM is a two-dimensional array of neurons $M = \{m_{ij},\ i = 1,...,k_x,\ \ j = 1,...,k_y\}$. Here k_x is the number of rows, and k_y is the number of columns. Each component of the input vector is connected to every individual neuron. Any neuron is entirely defined by its location on the grid (the number of row i and column j) and by the codebook vector, i.e. we can consider a neuron as an n-dimensional vector $m_{ij} = (m_{ij}^1, m_{ij}^2, ..., m_{ij}^n) \in R^n$.

The learning starts from the vectors m_{ij} initialized at random. At each learning step, the input vector X is passed to the neural network. The Euclidean distance from this input vector to each vector m_{ij} is calculated and the vector (neuron) m_c with the minimal Euclidean distance to X is designated as a winner, $c = \arg\min_{i,j}\{\| X - m_{ij} \|\}$. The components of the vector m_{ij} are adapted according to the general rule: $m_{ij}(t+1) = m_{ij}(t) + h_{ij}^c(t)(X - m_{ij}(t))$, where t is the number of iteration (learning step), h_{ij}^c is the so-called neighbourhood function, $h_{ij}^c(t) \to 0$, as $t \to \infty$.

There are various neighbourhood functions and learning rules. In [32], $h_{ij}^c = \dfrac{\beta}{\beta \eta_{ij}^c + 1}$; $\beta = \max\left(\dfrac{e+1-\hat{e}}{e}, 0.01\right)$; η_{ij}^c is the neighbourhood order between the neurons m_{ij} and m_c (all neurons adjacent to a given neuron can be defined as its neighbours of a first order, then the neurons adjacent to a first-order neighbour, excluding those already considered, as neighbours of a second order, etc.); e is the number of training epochs; the vector m_{ij} is recalculated if $\eta_{ij}^c \leq \max[\beta \max(k_x, k_y), 1]$.

Let us describe the term "training epoch" more in detail. An epoch consists of m steps: the input vectors from X_1 to X_m are passed to the neural network in a consecutive or random order. The consecutive order was used in [32]. Both the orders were examined in [33]. In this chapter, we use the random order, because we try to eliminate the influence of numeration of the input vectors on the training process.

After a large number of training steps, the network has been organized and n-dimensional input vectors $X_1,...,X_m$ have been mapped – each input vector is related to the nearest neuron, i.e. the vectors are distributed among the elements of the map during training. Some elements of the map may remain unrelated with any vector of the set $\{X_1,...,X_m\}$, but there may occur elements related with some input vectors. The neurons related with the input vectors are called neurons-winners. In the case of the rectangular topology, we can draw a table with cells corresponding to the neurons (see Table 2). However, the table does not answer the question, how much the vectors of the neighbouring cells are close in the n-dimensional space. A natural idea arises to apply the distance-preserving projection method to additional mapping of the neurons-winners in the SOM. Sammon's mapping or other MDS-type methods may be used for these purposes.

Two scenarios of combinations of the SOM and Sammon mapping are discussed and examined in this chapter: a consecutive combination and the integrated one.

The way of consecutive combination of Sammon's mapping and the SOM has been investigated in [32]. The algorithm is as follows: all the input vectors $X_1,...,X_m$ first are processed using the SOM; then the vectors-winners, whose total number r is smaller than or equal to m, are displayed using Sammon's mapping.

The idea of the integrated combination algorithm is as follows: multidimensional vectors are analysed by using Sammon's mapping, taking into account the process of SOM training. This algorithm is presented below. The experiments below and in [6], [34] have shown that namely this combination of the SOM and Sammon mapping is very good in search for a more precise projection of multidimensional vectors in the sense of criterion E_S (Eq. (2)), when the vectors, corresponding to the neurons-winners of the SOM, are analysed.

We have suggested the following way of integrating the SOM and Sammon algorithm (see Figure 2):

1. The training set consists of m n-dimensional vectors $X_1, X_2,..., X_m$. The neural network will be trained using e training epochs.

2. All the e epochs are divided into equal training parts – blocks. Before starting the training of the neural network, we choose the number of blocks γ into which the training process will be divided. Each block contains v training epochs ($e = v\gamma$). Denote by q a block of the training process consisting of v epochs ($q = 1,...,\gamma$).

3. Denote vectors-winners, obtained by the q-th block of the training process, by $Z_1^q, Z_2^q,..., Z_{r_q}^q$ and two-dimensional projections of these vectors-winners, calculated

using Sammon's algorithm, by $Y_1^q, Y_2^q, ..., Y_{r_q}^q$ ($Y_i^q = (y_{i1}^q, y_{i2}^q)$, $i = 1, ..., r_q$). Note that the number of vectors-winners r_q will be smaller than or equal to m. The vectors-winners $Z_1^1, Z_2^1, ..., Z_{r_1}^1$, obtained after the first block of the training process ($q = 1$), are analysed by using Sammon's algorithm. There is a unique relation between a vector-winner and the corresponding vector (or several vectors) from the training set $\{X_1, X_2, ..., X_m\}$. The initial coordinates of two-dimensional vectors $Y_i^0 = (y_{i1}^0, y_{i2}^0)$, $i = 1, ..., r_1$, for Sammon's algorithm are set as follows: $y_{i1}^0 = i + \frac{1}{3}$, $y_{i2}^0 = i + \frac{2}{3}$. Two-dimensional projections $Y_1^1, Y_2^1, ..., Y_{r_1}^1$ of vectors-winners are calculated using Sammon's algorithm.

4. The vectors-winners obtained after the q-th block of the training process are analysed by using Sammon's algorithm. The initial coordinates of two-dimensional vectors $Y_1^q, Y_2^q, ..., Y_{r_q}^q$ for Sammon's algorithm are selected taking into account the result of the (q-1)-st block. Note that $r_q \neq r_{q-1}$ in general. The way of selecting the initial coordinates is presented below. We must determine the initial coordinates of each two-dimensional vector Y_i^q correspondent to the neuron-winner Z_i^q, $i = 1, ..., r_q$. The sequence of steps is as follows. Determine vectors from $\{X_1, X_2, ..., X_m\}$ that are related with Z_i^q. Denote these vectors by $X_{i_1}, X_{i_2}, ...$ ($X_{i_1}, X_{i_2}, ... \in \{X_1, X_2, ..., X_m\}$). Determine neurons-winners of the (q-1)-st block that were related with $X_{i_1}, X_{i_2}, ...$. Denote these neurons-winners by $Z_{j_1}^{q-1}, Z_{j_2}^{q-1}, ...$ ($Z_{j_1}^{q-1}, Z_{j_2}^{q-1}, ... \in \{Z_1^{q-1}, Z_2^{q-1}, ..., Z_{r_{q-1}}^{q-1}\}$), and their two-dimensional projections, obtained as a result of Sammon's algorithm, by $Y_{j_1}^{q-1}, Y_{j_2}^{q-1}, ...$ ($Y_{j_1}^{q-1}, Y_{j_2}^{q-1}, ... \in \{Y_1^{q-1}, Y_2^{q-1}, ..., Y_{r_{q-1}}^{q-1}\}$). The initial coordinates of Y_i^q are set to be equal to the mean value of the set of vectors $\{ Y_{j_1}^{q-1}, Y_{j_2}^{q-1}, ... \}$. Afterwards, the two-dimensional projections $Y_1^q, Y_2^q, ..., Y_{r_q}^q$ ($Y_i^q = (y_{i1}^q, y_{i2}^q)$, $i = 1, ..., r_q$) of the vectors-winners are calculated using Sammon's algorithm.

5. The training of the neural network is continued until $q = \gamma$. After the γ-th block, we get two-dimensional projections $Y_1^\gamma, Y_2^\gamma, ..., Y_{r_\gamma}^\gamma$ of the n-dimensional vectors-winners $Z_1^\gamma, Z_2^\gamma, ..., Z_{r_\gamma}^\gamma$ that are uniquely related with the vectors $X_1, X_2, ..., X_m$.

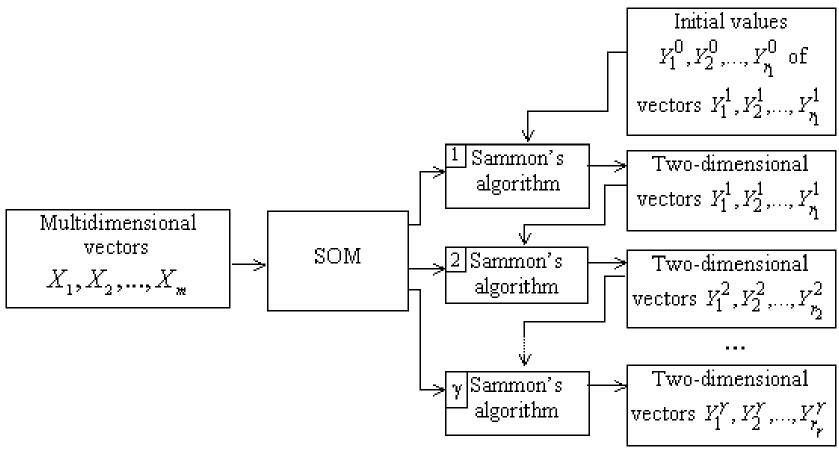


Figure 2. Scheme of the integrated combination of the SOM and Sammon mapping

Denote a consecutive combination of the SOM and Sammon mapping by *SOMSam(a)*, and the integrated one by *SOMSam(b)*. The combined mapping have been examined analysing the data on coastal dunes and their vegetation in Finland [35] (see Section 4.1.1).

Cases with various parameters of the algorithms and their constituent parts have been analysed:

- size of SOM (2x2, 3x3, 4x4, 5x5, 6x6);
- number of training epochs *e* (100, 200, 300);
- number of training blocks γ and the number of epochs v per each training block ($e = v\gamma$);
- values of the parameter α in Sammon's mapping (0.1; 0.11;…; 0.99; 1) (Eq. (3)).

Under the same initial conditions, the errors of projection E_S (Eq. (2)) have been calculated. The experiments have been repeated for 200 times with different (random) initial values of the components of the neurons-vectors $m_{ij} = (m_{ij}^1, m_{ij}^2, \dots, m_{ij}^n) \in R^n$.

Table 1. The ratios between the projection errors of *SOMSam(a)* and *SOMSam(b)*.

e	100					200						300				
v	50	25	20	10	5	50	40	25	20	10	5	50	25	20	10	5
γ	2	4	5	10	20	4	5	8	10	20	40	6	12	15	30	60
2x2	2.30	2.85	2.84	2.85	2.86	3.07	3.09	3.12	3.11	3.10	3.15	3.25	3.30	3.29	3.23	3.34
3x3	1.11	1.14	1.14	1.20	1.26	1.12	1.14	1.18	1.20	1.25	1.31	1.15	1.20	1.22	1.27	1.30
4x4	1.49	1.66	1.67	1.75	1.79	2.85	2.96	3.06	3.10	3.18	3.23	4.60	4.83	4.92	4.95	5.05
5x5	1.03	1.04	1.05	1.06	1.06	1.04	1.05	1.06	1.06	1.07	1.07	1.06	1.07	1.07	1.07	1.08
6x6	1.07	1.08	1.11	1.13	1.13	1.11	1.20	1.20	1.22	1.23	1.24	1.23	1.24	1.25	1.26	1.27

The ratios between the mean projection errors, obtained by *SOMSam(a)* and *SOMSam(b)* (see Table 1) have been calculated. It is apparent from Table 1 and Figure 3 that these ratios are always greater than one, i.e. the projection error, obtained by

SOMSam(b) is smaller than that, obtained by *SOMSam(a)*. The experiments in the papers [6], [34] show that the slight differences in the error value cause essential differences in the distribution of points projected on a plane.

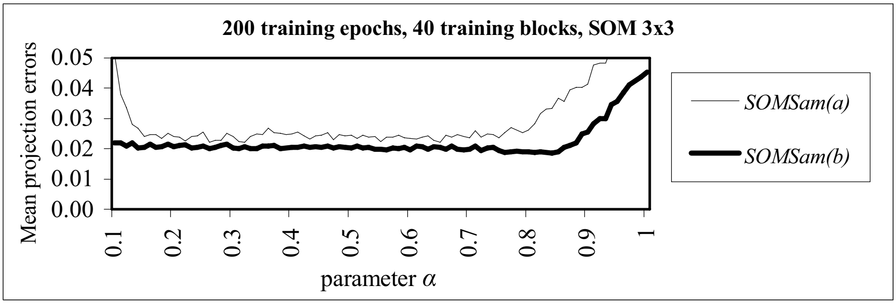


Figure 3. Dependence of the projection error on the parameter α

An advantage of the combined mapping over Sammon's projection used individually is the fact that the vectors are almost uniformly distributed after a direct compression of high dimensional points on a plane by Sammon's projection (Figure 4a), meanwhile the points compose some groups after combined mapping (Figure 4b) through the SOM possibility of clustering. Twenty four 24-dimensional vectors $X_1, X_2, ..., X_{24}$ are used in this example (see [36]). We do not present legends and units for both axes in all figures with visualization results, because we are interested in observing the interlocation of points on a plane only. Numbers in the figures are order numbers (indices) of vectors $X_1, X_2, ..., X_{24}$.

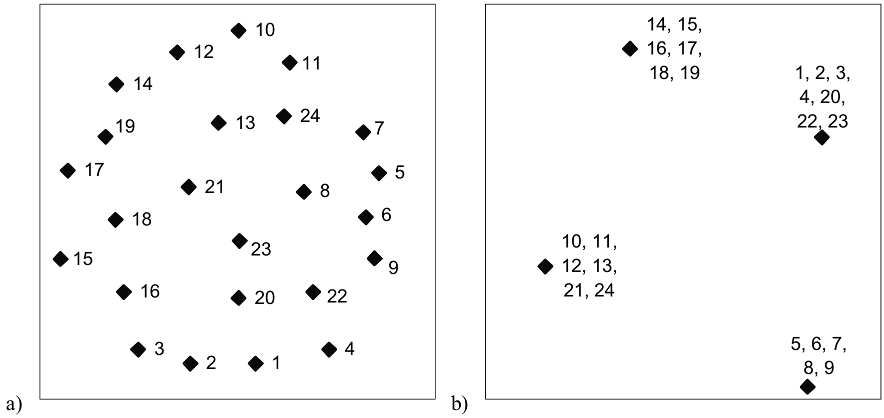


Figure 4. Distribution of points, obtained: (a) by Sammon's projection; (b) combined mapping

3. SAMANN – A Neural Network for Sammon's Projection

There is no mechanism to project one or more new data points without expensively regenerating the entire configuration from the augmented data set in Sammon's algorithm. To project new data, one has to run the program again on all the data (old data and new data). To solve this problem, some methods (triangulation algorithm [37], [15], standard feed-forward neural network [31]) have been proposed in the literature.

A specific backpropagation-like learning rule has been developed to allow a normal feed-forward artificial neural network to learn Sammon's mapping in an unsupervised way. The neural network training rule of this type was called SAMANN [20]. The network is able to project new multidimensional vectors (points) after training.

Mao and Jain [20], who evaluated different network types on eight different data sets, three numerical criteria and visual inspection, conclude that the SAMANN network preserves the data structure, cluster shape, and interpattern distances better than a number of other feature extraction or data projection methods.

The architecture of the SAMANN network is a multilayer perceptron where the number of input vectors is set to be the input space dimension, *n*, and the number of output vectors is specified as the projected space dimension, *d* (Figure 5).

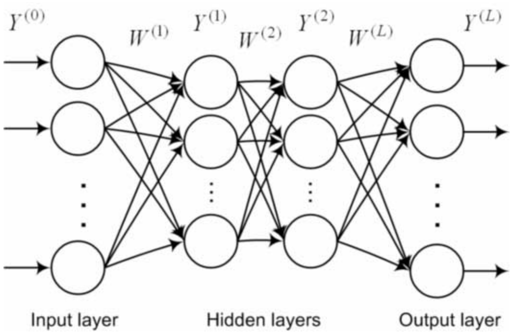


Figure 5. SAMANN network for *d*-dimensional projection

Mao and Jain have derived a weight updating rule for the multilayer perceptron [20] that minimizes Sammon's error (projection error E_S) (Eq. (2)), based on the gradient descent method. The general updating rule for all the hidden layers, $l = 1,\ldots,L-1$ and for the output layer ($l = L$) is:

$$\Delta\omega_{jk}^{(l)} = -\eta \frac{\partial E_S(\mu,\nu)}{\partial \omega_{jk}^{(l)}} = -\eta(\Delta_{jk}^{(l)}(\mu) y_j^{(l-1)}(\mu) - \Delta_{jk}^{(l)}(\nu) y_j^{(l-1)}(\nu)), \tag{4}$$

where ω_{jk} is the weight between the unit *j* in the layer $l-1$ and the unit *k* in the layer *l*, η is the learning rate, $y_j^{(l)}$ is the output of the *j*th unit in the layer *l,* and μ and ν are two vectors from the analysed data set $\{X_1, X_2, ..., X_m\}$. The $\Delta_{jk}^{(l)}$ are the errors accumulated in each layer and backpropagated to a preceding layer, similarly to the standard backpropagation. The sigmoid activation function with the range $(0,1)$ is used for each unit.

The network takes a pair of input vectors each time in the training. The outputs of each neuron are stored for both points. The distance between the neural network output vectors can be calculated and an error measure can be defined in terms of this distance and the distance between the points in the input space. From this error measure, a weight update rule has been derived in [20]. After, training the network, one is able to use it to generalise on previously unseen data.

The SAMANN unsupervised backpropagation algorithm is as follows:

1. Initialize the weights in the SAMANN network randomly.
2. Select a pair of vectors randomly, present them to the network one at a time, and evaluate the network in a feed-forward fashion.
3. Update the weights in the backpropagation fashion starting from the output layer.
4. Repeat steps 2-3 a number of times.
5. Present all the vectors and evaluate the outputs of the network; compute the projection error E_S; if the value of Sammon's error is below a predefined threshold or the number of iterations (from steps 2-5) exceeds the predefined maximum number, then stop; otherwise, go to step 2.

In our realization of the SAMANN training, one iteration means showing all possible pairs of vectors $X_1, X_2, ..., X_m$ to the neural network once.

A drawback of using SAMANN is that the original dataset has to be scaled for the artificial neural network to be able to find a correct mapping, since the neural network can only map to points in the sigmoid's output interval $(0,1)$. This scaling is dependent on the maximum distance in the original dataset. It is therefore possible that a new vector, shown to the neural network, will be mapped incorrectly, when its distance to a vector in the original dataset is larger than any of the original interpattern distances. Another drawback of using SAMANN is that it is rather difficult to train and it is extremely slow.

Control Parameters of the SAMANN Network Training. The rate, at which artificial neural networks learns, depends upon several controllable factors. Obviously, a slower rate means that a lot more time is spent in accomplishing the learning to produce an adequately trained system. At the faster learning rates, however, the network may not be able to make the fine discriminations possible with a system that learns more slowly. Thus, if η is small when the algorithm is initialized, the network will probably take an unacceptably long time to converge. The experiments, done in [38], have shown in what way the SAMANN network training depends on the learning rate.

The training of the SAMANN network is a very time-consuming operation. In the consideration of the SAMANN network, it has been observed that the projection error depends on different parameters. The latest investigations have revealed that, in order to achieve good results, one needs to select the learning rate η properly. It has been stated so far that projection yields the best results if the η value is taken from the interval $(0,1)$. In that case, however, the network training is very slow. One of the possible reasons is that, in the case of SAMANN network, the interval $(0,1)$ is not the best one. Thus, it is reasonable to look for the optimal value of the learning parameter that may not necessarily be within the interval $(0,1)$.

The experiments [38] demonstrate that with an increase in the learning rate value, a better projection error is obtained. That is why the experiments have been done with higher values of the learning rate beyond the limits of the intervals $(0,1)$. In all the experiments a two-layer (one hidden layer) network with 20 hidden units is used, the projected space dimension d=2. The results are presented in Figure 6. The following datasets were used in the experiments: Iris dataset (Fisher's iris dataset) [39], Salinity dataset [40], HBK dataset (artificial dataset) [41]. It has been noticed that the best results are at $\eta > 1$.

The experiments allow us to conclude that the optimal value of the learning rate for the datasets considered is within the interval $[10, 30]$. By selecting such values of the learning rate, a significant economy of computing time is possible. For the fixed number of iterations, good projection results are obtained in a shorter time interval than that taking the learning rate values from the interval $(0,1)$. However, with an increase in the value of the learning rate, the error variations also increase, which can cause certain network training problems. Lower values of the learning rate within the interval $(0,1)$ guarantee a more stable convergence to the minimum of the projection error.

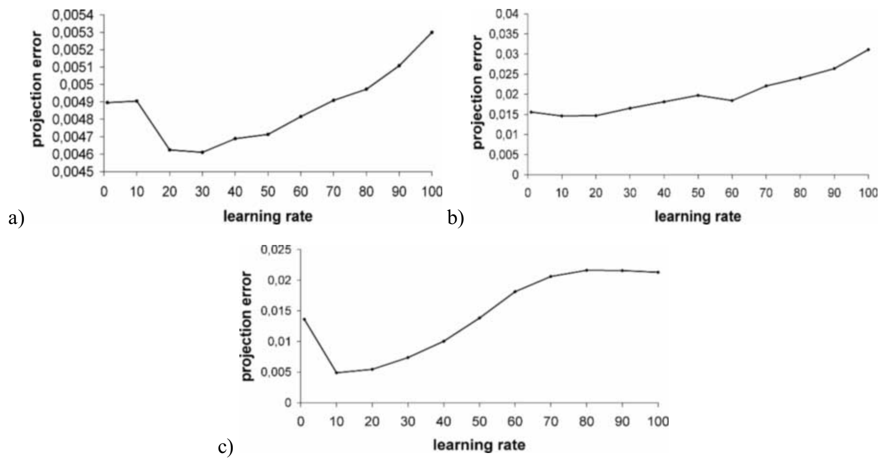


Figure 6. The dependence of the data projection accuracy on the learning rate η, $\eta \in [1, 100]$: (a) Iris dataset; (b) Salinity dataset; (c) HBK dataset

Retraining of the SAMANN Network. After training the SAMANN network, a set of weights of the neural network is fixed. A new vector shown to the network is mapped into the plane very fast and quite exactly without any additional calculations. However, while working with large data amounts there may appear a lot of new vectors, which entails retraining of the SAMANN network after some time.

Let us name the set of vectors that have been used to train the network by the primary set, and the set of the new vectors, that have not been used for training yet, by the new set. Denote the number of vectors in the primary dataset by N_1, and the number of vectors in the new dataset by N_2.

Two strategies for retraining the neural network that visualizes multidimensional data have been proposed and investigated [38]. The first strategy uses all possible pairs of data vectors (both from primary and new datasets) for retraining. The second strategy uses a restricted number of pairs of the vectors (a vector from the primary set – a vector from the new set).

The strategies of the neural network retraining data are as follows:

1. The SAMANN network is trained by N_1 vectors from the primary dataset, weights W_1 are obtained, then the projection error $E_S(N_1)$ is calculated, and vector projections are localized on the plane. After the emergence of N_2 new

vectors, the neural network is retrained with all the $N_1 + N_2$ vectors, using the set of weights W_1 as the initial one. The new weights W_2 of SAMANN network is found.

2. The SAMANN network is trained by N_1 vectors from the primary dataset, weights W_1 are obtained, and the projection error $E_S(N_1)$ is calculated. Since in order to renew the weights W_1, a pair of vectors is simultaneously provided for the neural network – one vector from the primary dataset and one from the new one. The neural network is retrained with $2N_2$ vectors at each iteration. Here one iteration involves computations, where all vectors from the new dataset are provided to the neural network once. The new set of network weights W_2 is found.

The experiments [38] have shown that the first strategy yield good results, however retraining of the network is slow. In the analysis of the strategies for the network retraining, a particular case of the SAMANN network was considered: a feed-forward artificial neural network with one hidden layer (20 hidden units) and two outputs (d=2). The set of initial weights was fixed in advance. To visualize the primary dataset, the following parameters were employed: the number of iterations M=10000, the learning rate was $\eta = 10$; to visualize the set of new vectors: the learning rate was $\eta = 1$, and the number of iterations depended on the strategy chosen.

The best visualization results are obtained by taking points for network retraining from the primary dataset and the new dataset (second strategy). Figures 7 and 8 demonstrate the results of calculation. The results of retraining the SAMANN network only with the new vectors (points) are indicated in the figures. After each iteration the projection error $E_S(N_1 + N_2)$ is calculated and the computing time is measured. The Iris dataset and 300 randomly generated vectors (three spherical clusters with 100 5-dimenional vectors each) have been used in the experiments. The second strategy enables us to attain good visualization results in a very short time as well as to get smaller projection errors and to improve the accuracy of projection.

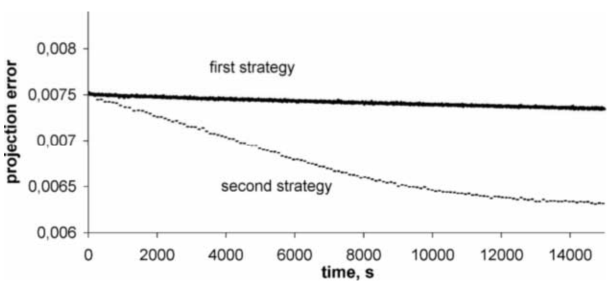


Figure 7. Dependence of the projection error on the computing time for the Iris dataset

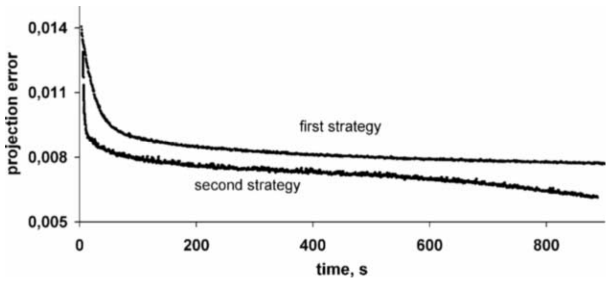


Figure 8. Dependence of the projection error on the computing time for randomly generated vectors

In some experiments we observe the fast stabilization of the projection error in the case of the first strategy, while the projection error by the second strategy decreases. This is an advantage of the second strategy. The experiments lead to the idea of a possibility to minimize the SAMANN neural network training time by dividing the training process into two subprocesses: (1) training of the network by a part of the data vectors; (2) retraining of the network by the remaining part of the dataset. In this case, the training set will consist of two subsets of $\{X_1, X_2, ..., X_m\}$. A smaller number of the pairs of vectors will be used when training the network by vectors of the subsets. This allows us to obtain a similar visualization quality much faster.

While working with high-dimensional datasets, it is important to have a way to speed up the network training process. In [42], the algorithm (parallel realization) which allows us to realize this requirement has been proposed. The algorithm divides the SAMANN training dataset into two (or more) parts. Two (or more) identical neural networks are created and weight sets for each network are calculated. Of all the weight sets it remains the one with smaller projection error. Using the obtained weight set, the neural network is trained with all the data set vectors.

4. Applications

4.1. Applications of the Combined Mapping to the Visual Analysis of Correlations

The values obtained by any parameter x_i can depend on the values of the other parameters $x_j,\ j = 1,...,n,\ i \neq j$, i.e. the parameters can be correlated. The problem is to discover knowledge about the interlocation of parameters, and about groups (clusters) of parameters by the values of elements of the correlation matrix. The approach, reviewed in this section, is oriented to the analysis of correlation matrices via the visual presentation of a set of variables on a plane. More details on the approach can be found in [32].

The correlation matrix $R = \{r_{x_i x_j}, i, j = 1,...,n\}$ of parameters can be calculated by analysing the objects that compose the set. Here $r_{x_i x_j}$ is a correlation coefficient of the parameters x_i and x_j. The specific character of the correlation matrix analysis problem lies in the fact that the parameters x_i and x_j are related more strongly if the absolute value of the correlation coefficient $|r_{x_i x_j}|$ is higher, and less strongly if the value of $|r_{x_i x_j}|$ is lower. The minimal relationship between the parameters is equal to 0. The maximal relationship is equal to 1 or –1.

Let S^n be a subset of an n-dimensional Euclidean space R^n containing vectors of unit length, i.e. S^n is a unit sphere, $\|Y\|=1$ if $Y \in S^n$. It is necessary to determine a system of vectors $Y_1,...,Y_n \in S^n$ corresponding to the system of parameters $x_1,...,x_n$ so that $\cos(Y_i,Y_j)=|r_{x_i x_j}|$ or $\cos(Y_i,Y_j)=r_{x_i x_j}^2$. If only the matrix of cosines $K=\{\cos(Y_i,Y_j), i,j=1,...n\}$ is known, it is possible to restore the system of vectors $Y_s=(y_{s1},...,y_{sn}) \in S^n$, $s=1,...,n$, as follows: $y_{sk}=\sqrt{\lambda_k}\alpha_{sk}$, $k=1,...,n$, where λ_k is the k-th eigenvalue of the matrix K, $(\alpha_{1k},...,\alpha_{nk})$ is a normalized eigenvector corresponding to the eigenvalue λ_k. The system of vectors $Y_1,...,Y_n \in S^n$ exists, if the matrix of their scalar products is non-negative definite. The correlation matrix $R=\{r_{x_i x_j}, i,j=1,...,n\}$ is non-negative definite. It has been proven in [32] that the matrix $R^2=\{r_{x_i x_j}^2, i,j=1,...,n\}$ is non-negative definite as well. Therefore, the system of vectors $Y_1,...,Y_n \in S^n$ may be restored for the matrices R and R^2.

The set of vectors $Y_1,...,Y_n \in S^n$, which corresponds to the set of parameters $x_1,...,x_n$, can be mapped on a plane trying to preserve a relative distance between $Y_1,...,Y_n \in S^n$. This leads to the possible visual observation of the layout of parameters $x_1,...,x_n$ on the plane.

4.1.1. Environmental Data Analysis

One of the most developing fields of late has been environmetrics that covers the development and application of quantitative methods in the environmental sciences and provides essential tools for understanding, predicting, and controlling the impacts of agents, both man-made and natural, which affect the environment [43]. Correlations of environmental parameters and their analysis appear in various research studies (see e.g. [44], [35], [45], [46], [47], [48]). The references cover air pollution, vegetation of coastal dunes, groundwater chemistry, minimum temperature trends, zoobenthic species-environmental relationships, development and analysis of large environmental and taxonomic databases. We present two applications more in detail: correlation matrix of 10 meteorological parameters that describe the air pollution in Vilnius and correlation matrix of 16 environmental parameters that describe the development of coastal dunes and their vegetation in Finland. These two problems are very urgent because they are of an ecological nature: a visual presentation of data stored in the correlation matrices makes it possible for ecologists to discover additional knowledge hidden in the matrices and to make proper decisions. The conclusions on the similarity of the measured environmental parameters (air pollution and development and vegetation of coastal dunes) as well as on the possible number of clusters of similar parameters are drawn analysing the visual data.

Meteorological parameters are as follows: x_1, x_2, and x_3 are the concentrations of carbon monoxide CO, nitrogen oxides NO_x, and ozone O_3, respectively; x_4 is the vertical temperature gradient measured at a 2 - 8 m height; x_5 is the intensity of solar radiation; x_6 is the boundary layer depth; x_7 is the amount of precipitation; x_8 is the temperature; x_9 is the wind speed; x_{10} is the stability class of atmosphere. The correlation matrix is presented in [44]. The results of combined mapping (4×4 SOM and Sammon's mapping) are presented in Figure 9a.

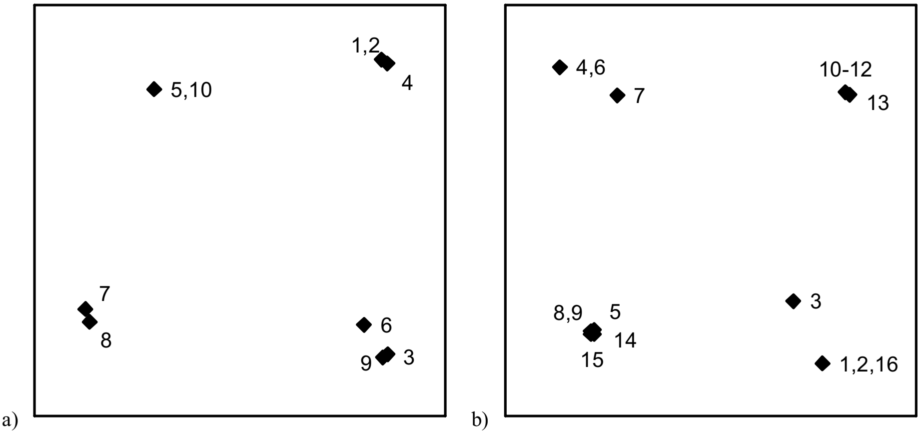


Figure 9. Combined mapping of: (a) the meteorological parameter set; (b) environmental parameter set

Environmental parameters are as follows: x_1 is the distance from the water line; x_2 is the height above the sea level; x_3 is the soil pH; x_4, x_5, x_6, and x_7 are the contents of calcium (Ca), phosphorous (P), potassium (K), and magnesium (Mg); x_8 and x_9 are the mean diameter and sorting of sand; x_{10} is the northernness in the Finnish coordinate system; x_{11} is the rate of land uplift; x_{12} is the sea level fluctuation; x_{13} is the soil moisture content; x_{14} is the slope tangent; x_{15} is the proportion of bare sand surface; x_{16} is the tree cover. The correlation matrix is presented in [44]. In Table 2, the distribution of the environmental parameters on the 4x4 SOM is presented. Here the cells corresponding to the neurons-winners are filled with the numbers of vectors $Y_1, Y_2, \ldots, Y_n \in S^n$, $n = 16$, some cells remain empty. One can visually decide on the distribution of vectors in the n-dimensional space according to their distribution among the cells of the table. However, the table does not answer the question, how much the vectors of the neighbouring cells are close in the n-dimensional space. It is expedient to apply Sammon's projection method to an additional mapping of the neurons-winners in the SOM. The results of combined mapping (4×4 SOM and Sammon's mapping) are presented in Figure 9b.

Table 2. Distribution of the environmental parameters on the 4x4 SOM

4, 6		5	8, 9
7		15	14
3			
1, 2, 16		13	10, 11, 12

4.1.2. Statistical Analysis of Curricula

The main proposition in the analysis [49] is that, in most cases, a student gets similar marks in the related subjects. If a student is gifted for the humanities, he will be successful in most of the humanities. Mathematical aptitude yields good marks in mathematical subjects. When analysing the curriculum of studies, we know in advance which subjects are mathematical and which of the humanities. However, there are subjects that cannot be assigned to any well-defined class of subjects. Computer science subjects may serve as an example of such subjects. The analysis made it possible to evaluate the level of mathematization of different computer science subjects or their proximity to the humanities. This level is never quantified, but it may be estimated considering the whole complex of subjects that compose the curriculum. The necessary data were the results of examination, only. The correlation matrix of 25 subjects obtained on the basis of examination results at the Faculty of Mathematics and Informatics of Vilnius Pedagogical University has been analysed.

The list of subjects: x_1 – Probability theory, x_2 – Methods of teaching mathematics, x_3 – Geometry 1, x_4 – Pedagogy and psychology, x_5 – Geometry 2, x_6 – Mathematical analysis 1, x_7 – Pedagogy, x_8 – Geometry 3, x_9 – Psychology 1, x_{10} – Algebra, x_{11} – Mathematical analysis 2, x_{12} – Foreign language, x_{13} – Geometry 4, x_{14} – Psychology 2, x_{15} – Algebra and number theory 1, x_{16} – Mathematical analysis 3, x_{17} – Informatics 1, x_{18} – Algebra and number theory 2, x_{19} – Mathematical analysis 4, x_{20} – Informatics 2, x_{21} – Development of training computer programs, x_{22} – Methods of teaching informatics, x_{23} – Packages for solving mathematical problems, x_{24} – Algorithm theory, x_{25} – Programming methods. The correlation matrix is presented in [49].

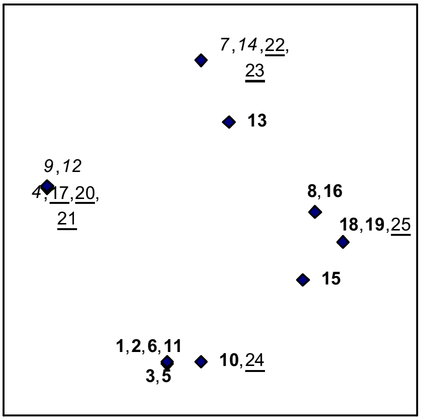


Figure 10. Combined mapping of subjects

The results of combined mapping (4×4 SOM and Sammon's mapping) are presented in Figure 10. Mathematical subjects are bold formatted in Figure 10, computer science subjects are underlined, and the humanities are presented in italics. This kind of notation allows us to perceive the results easier. The visual analysis of the distribution of points in Figure 10 leads to a conclusion that computer science subjects range from those of a purely mathematical nature to that of the humanities. However, these subjects do not part from the groups of mathematical subjects or the humanities.

4.2. Applications of the Multidimensional Data Visualization

4.2.1. Comparison of Schools

A qualitative comparison of schools from the standpoint of "city – rural district" or "gymnasium – secondary school" is performed in [50], [51]. In most cases, education in gymnasia is of a higher quality as compared with that in the usual secondary schools. This may often be concluded when comparing education in city schools with the rural district ones. However, how great are these differences? In addition, it would be useful to get some knowledge of the influence of the qualification, age, and number of teachers on the state of school. The research is based on 19 schools from the Panevėžys town and 9 schools from the Panevėžys district (Panevėžys is the northern Lithuanian town). In the tables and figures of this section, the schools from the city are labelled by the numbers from 1 to 19, and that of the district are labelled by the numbers from 20 to 28. There are two gymnasia (numbers 1 and 2). The remaining schools are the secondary ones. The following indicators were selected to describe a school: the percent of teachers of the highest qualification (that have a degree of a methodologist or expert); the percent of teachers that do not have a desired qualification (i.e. who don't do the job they were trained for); the percent of teachers whose age is over 55 years; the percent of teachers who are younger than 35 years; the percent of the annual increase in the number of teachers. The results of combined mapping are presented in Figure 11. We observe two clusters of schools. On the upper side of the figure, we observe a cluster of city schools where there are both gymnasia. This cluster contains only one district school (26). The second cluster (lower side of Figure 11) contains both city and district schools, but the district schools dominate here.

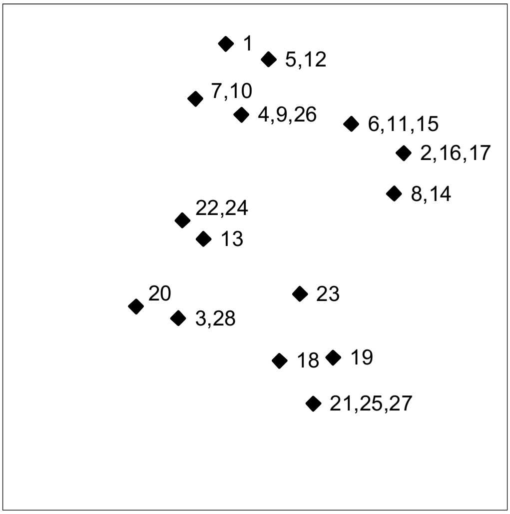


Figure 11. Distribution of the Panevėžys town and district schools in the 1999/2000 school year

4.2.2. Analysis of the Economic and Social Conditions of Central European Countries

Ten countries were compared in [33]: (1) Hungary, (2) Czech Republic, (3) Lithuania, (4) Latvia, (5) Slovakia, (6) Poland, (7) Romania, (8) Estonia, (9) Bulgaria, (10) Slovenia. The data on the economic and social conditions of Central European countries are taken from the USA CIA *The World Factbook 1999* database [52] and World Bank *Country Data* database [53]. The wide used essential economic and social parameters are selected: the infant mortality rate (deaths / 1000 live births); the Gross Domestic Product (GDP) per capita in US dollars; the percentage of GDP developed in the industry and services (not in the agriculture); the export per capita in thousands of US dollars; the number of telephones per capita. The results of combined mapping (4×4 SOM and Sammon's mapping) are presented in Figure 12. The average values of parameters compose point AVE in Figure 12, the worst values of parameters compose MIN, and the best values of parameters compose MAX. From the visually presented results in Figure 12, it is easy to draw the comparative conclusions on the economic and social conditions of the countries analysed.

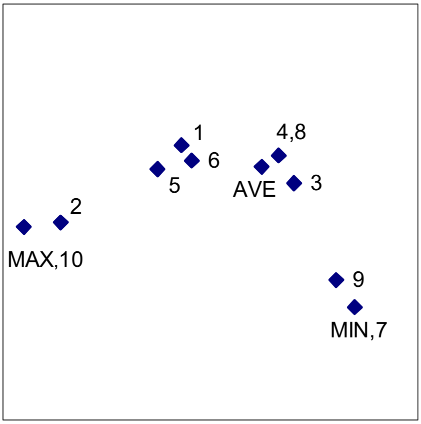


Figure 12. Distribution of countries

4.3. Visual Analysis of Medical Data

4.3.1. Analysis of Data on the Fundus of Eyes

The images of fundus of eyes are analysed. The fundus of eye is presented in Figure 13. 27 numerical parameters of fundus of eyes have been measured on 42 patients. They are evaluated from the photos of the fundus. The parameters can characterise some diseases of eyes. There are four groups of parameters: (1) parameters of optic nerve discs (OND): x_1 – major axis of OND ellipse, x_2 – minor axis, x_3 – semimajor axis, x_4 – semiminor axis, x_5 – horizontal diameter, x_6 – vertical diameter, x_7 – area, x_8 – eccentricity, x_9 – perimeter; (2) parameters of excavation (EKS) (excavation is a degenerated part of OND): x_{10} – major axis of EKS ellipse, x_{11} – minor axis, x_{12} – semimajor axis, x_{13} – semiminor axis, x_{14} – horizontal diameter, x_{15} – vertical diameter, x_{16} – area, x_{17} – eccentricity, x_{18} – perimeter; (3) ratios between various OND, EKS, NRK parameters (neuroretinal rim (NRK) is an OND part that is not degenerated): x_{19} – ratio of EKS and OND horizontal diameters, x_{20} – ratio of EKS and OND vertical diameters, x_{21} – NRK area, x_{22} – ratio of NRK and OND areas, x_{23} – ratio of EKS and OND; (4) thickness of NRK parts: x_{24} is inferior disc sector, x_{25} – superior disc sector, x_{26} – nasal disc sector, x_{27} – temporal disc sector [54] [55].

In our experiments, the eyes of patients nos. 1-18 are healthy, and the eyes of patients nos. 19-42 are damaged by glaucoma. The target of the analysis is to evaluate how the vectors, consisting of the parameters of eye fundus, are distributed on the plane, whether they form specific groups. Is it possible to identify glaucoma using this system of parameters?

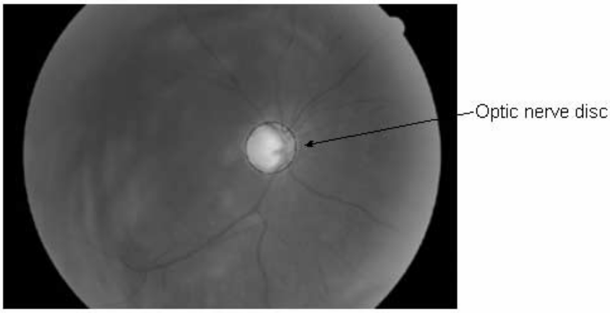


Figure 13. Fundus of eye

The results of combined mapping are presented in Figures 14 and 15. The data for analysis are obtained in two ways: (1) one data vector contains the parameters of a patient – all 27 parameters or several of them (Figure 14); (2) data vectors are restored from the correlation matrix of 27 parameters using the method proposed in [32] (see Section 4.1) – it allows a visual observation of dependencies of 27 parameters (Figure 15).

When analysing the patient data vectors, comprised of all the 27 parameters, it is impossible to differentiate the diseased eyes from the healthy ones (Figure 14a). The probable reason is due to too many parameters, most of which are not essential for the problem – they cause some noise. When analysing individual groups of the parameters, it is possible to notice some regularities. Most of the points, corresponding to the healthy eyes (nos. 7, 8, 10, 13-15), are distributed in one corner (see Figure 14b, EKS parameters are analysed). The points, corresponding to the eyes, damaged by glaucoma, are in the opposite corner. If we view the points starting from the top left corner of the picture, moving to the bottom right corner, it is possible to notice that at first there are only the points, corresponding to damaged eyes, later the number of such points decreases, and finally there are only points, corresponding to the healthy eyes (Figure 14b). The problem is to extract the proper set of parameters.

When analysing the system of parameters, it is possible to evaluate the similarities of parameters and to simplify the system by eliminating some parameters or introducing the new ones that represent a subset of parameters from the initial group. In Figure 15, the possible subsets of similar parameters are rimmed by a dotted line.

The experiments above illustrate the new possibilities for the ophthalmologic data analysis. They illustrate the common strategy that may be applied in the development and optimization of the set of parameters.

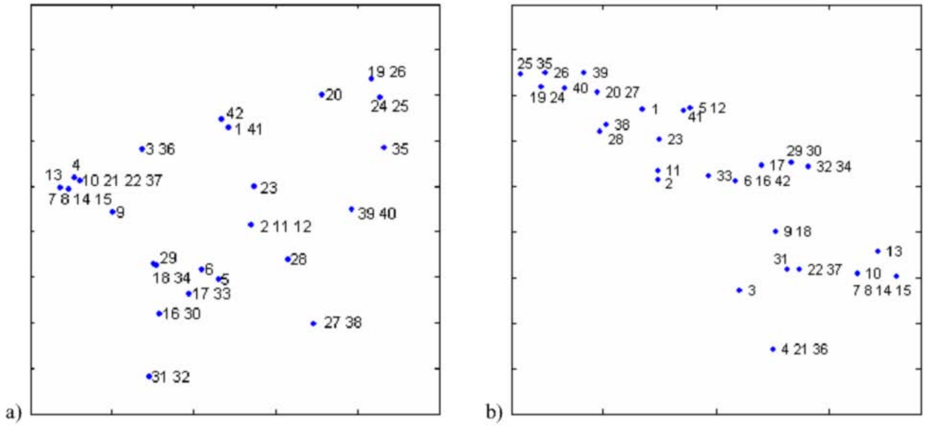


Figure 14. Projections of the vectors, corresponding to the ophthalmologic data, on the plane, when analysing: (a) all the 27 parameters; (b) only 7 parameters of EKS

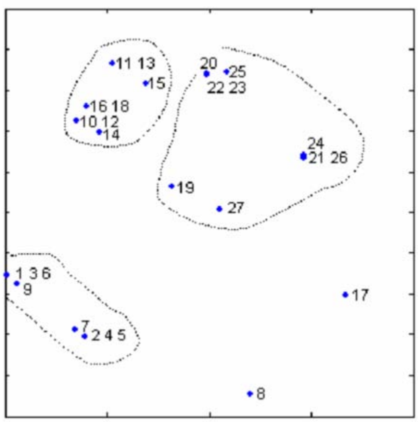


Figure 15. Distribution the ophthalmologic parameters on the plane

4.3.2. Analysis of Physiological Data

The purpose of analysis is to evaluate men's health state and their possibility of going in for sports. The analysed physiological data set consists of three groups: (1) ischemic heart-diseased men (61 items), (2) healthy persons (not going in for sports) (110 items), and (3) sportsmen (161 items). Non-specific physiological features that are frequently used in clinical medicine and that describe the human functional state are as follows: heart rate (HR), interval in the electrocardiogram from point J to the end T of the wave (JT interval), systolic blood pressure (SBP), diastolic blood pressure (DBP), and the ratios between some parameters (SBP-DBP)/SBP, JT/RR (RR=60/HR). Multidimensional vectors are formed by the method, presented in [56]. The method is based on the evaluation of some fractal dimensions.

In the analysis of the multidimensional data visualized, a particular case of the SAMANN network has been considered: a feed-forward artificial neural network with one hidden layer and two outputs. To visualize the initial dataset, the following parameters were employed: the number of iterations M=100000, the learning parameter

$\eta = 1$ (Eq. (4)). The SAMANN network is trained by two groups (ischemics and sportsmen), using the standard backpropagation algorithm with a learning rate of 1.0 and the momentum value of 0.3 for 100000 iterations. A new set of SAMANN network outputs has been found for these two groups (ischemics and sportsmen) (Figure 16a), where the line is the decision surface of two classes, obtained by support vector machine (SVM) [57]. The set of SAMANN network weights W has been calculated. Now, the points of the third group data (healthy persons) are shown to the network and mapped among the previously projected points very fast and quite exactly without any additional calculations (Figure 16b). In Figure 16b, only the points of the third group are visible. According to the positions of the points in respect of the decision surface, we can draw preliminary conclusions about the health state of persons.

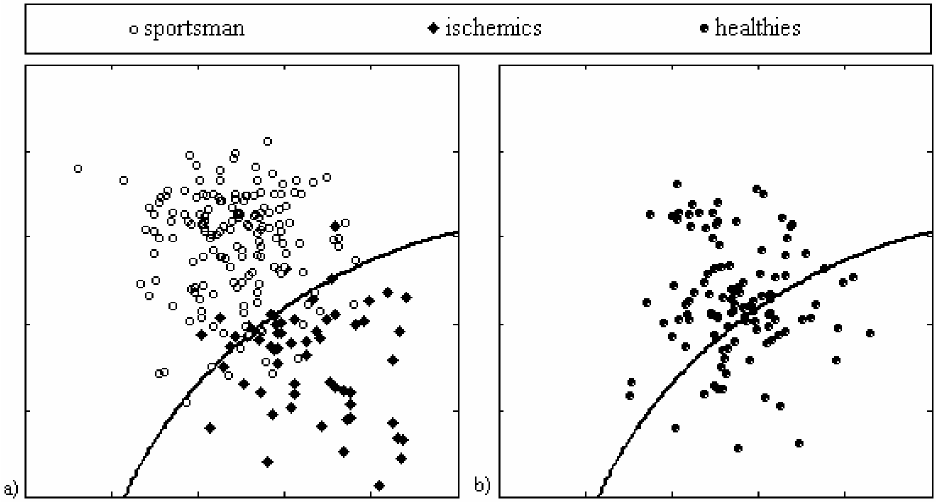


Figure 16. Projection of the physiological data, obtained by SAMMAN: (a) Groups 1 and 3; (b) all the 3 groups

5. Conclusions

With massive data sets existing in reality, we have conceived the difficulty and complexity of constructing computational tools extending human analysis capabilities to higher dimensions. A well performed visualization extends the human perceptual capabilities providing a deeper visual insight into data. One of the effective ways of visualizing multidimensional data is the reduction of dimension. The dimensionality reduction methods based on neural networks have been discussed here. The stress is put on combining the SOM and Sammon mapping and on the neural network for Sammon's mapping SAMANN.

Large scale applications are discussed: environmental data analysis, statistical analysis of curricula, comparison of schools, analysis of the economic and social conditions of countries, analysis of data on the fundus of eyes and that of physiological data on men's health. In all the cases, the visualization enabled us to display the properties of data with complex relations.

References

[1] D.A. Keim, M. Ward, *Visualization*. In: M. Berthold, D. J. Hand (eds.) *Intelligent Data Analysis: an Introduction*, Springer-Verlag, (2003), 403-427.

[2] P.E. Hoffman, G.G. Grinstein. A Survey of Visualizations for High-Dimensional Data Mining. In: U.Fayyad, G.G. Grinstein, A. Wierse (eds.) *Information Visualization in Data Mining and Knowledge Discovery*, Morgan Kaufmann Publishers, San Francisco, 2002.

[3] G. Grinstein, M. Trutschl, U. Cvek, High-Dimensional Visualizations, *Proceedings of Workshop on Visual data Mining, ACM Conference on Knowledge Discovery and data Mining* (2001), 1-14.

[4] S. Kaski, (1997). Data Exploration Using Self-Organizing Maps. In *Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series* **82**, Espoo: Finish Academy of Technology, 1997, http://www.cis.hut.fi/~sami/thesis/thesis_tohtml.html.

[5] T. Kohonen, Self-Organizing Maps (3rd ed.). Springer Series in Inform. Sciences **30**, Springer, 2001.

[6] O. Kurasova, Visual Analysis of Multidimensional Data Using Self-Organizing Maps (SOM), Dissertation ,Vilnius: MII, 2005.

[7] D.N. Lawley, A.E. Maxwell, Factor Analysis as a Statistical Method. London: Butterworths, 1963.

[8] E. Oja, Principal Components, Minor Components, and Linear Neural Networks, *Neural Networks* **5** (1992), 927-935.

[9] J. Rubner, P. Tavan, A Self-Organizing Network for Principal Component Analysis, *Europhysics Letters* **10** (1989), 693-698.

[10] C. Brunsdon, A.S. Fotheringham, M.E. Charlton, An Investigation of Methods for Visualising Highly Multivariate Datasets in Case Studies of Visualization in the Social Sciences, In: D. Unwin and P. Fisher (eds.) *Joint Information Systems Committee, ESRC, Technical Report Series* **43** (1998), 55-80.

[11] I. Borg, P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, Springer, 1997.

[12] J. de Leeuw, W. Heiser, Theory of Multidimensional Scaling. In: P.R.Krishnaiah and L.N.Kanal (eds.), *Handbook of Statistics* **2**, Amsterdam: North-Holland (1982), 285-316.

[13] M. Wish, J.D. Carroll, Multidimensional Scaling and Its Applications. In: P.R.Krishnaiah and L.N.Kanal (eds.), *Handbook of Statistics* **2,** Amsterdam: North-Holland (1982), 317-345.

[14] J.W. Sammon, A Nonlinear Napping for Data Structure Analysis, *IEEE Transactions on Computers* **18** (1969), 401-409.

[15] G. Biswas, A.K. Jain, R.C. Dubes, Evaluation of Projection Algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **3(6)** (1981), 701-708.

[16] A. Flexer, Limitations of Self-Organizing Maps for Vector Quantization and Multidimensional Scaling. In: M.C.Mozer, M.I.Jordan and T.Petsche (eds.), *Advances in Neural Information Processing Systems* **9**. Cambridge, MA: MIT Press/Bradford Books (1997), 445-451.

[17] T. Hastie, W. Stuetzle, Principal Curves, *Journal of the American Statist. Associat.* **84** (1989), 502-516.

[18] C.M. Bishop, M. Svensen, C.K.I. Wiliams, GMT: A Principles Aalternative to the Self-Organizing Map, *Advances in Neural Information Processing Systems* **9** (1997), 354-363.

[19] S. T. Roweis, L. Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science* **290** (2000), 2323-2326.

[20] J. Mao, A.K. Jain, Artificial Neural Networks for Feature Extraction and Multivariate Data Projection, *IEEE Trans. Neural Networks* **6** (1995), 296-317.

[21] D. Lowe, M.E. Tipping, Feed-forward Neural Networks and Topographic Mappings for Exploratory Data Analysis, *Neural Computing and Applications* **4** (1996), 83-95.

[22] M.C. van Wezel, W.A. Kosters, Nonmetric Multidimensional Scaling: Neural networks Versus Traditional Techniques. Intelligent Data Analysis **8(6)** (2004), 601-613.

[23] J. Hartigan, Clustering Algorithms. New York: Wiley, 1975

[24] H. Yin, H, ViSOM-A Novel Method for Multivariate Data Projection and Structure Visualisation, *IEEE Trans. on Neural Network*s **13** (2002), 237–243.

[25] P. Demartines, J. Hérault, Curvilinear Component Analysis: A Self-Organizing Neural Aetwork for Nonlinear Mapping of Data Sets, *IEEE Transaction on Neural Networks* **8(1)** (1997), 148–154.

[26] B. Baldi and K. Hornik. Neural Networks and Principal Component Analysis: Learning from examples without local minima, *Neural Networks* **2** (1989), 53–58.

[27] M.A. Kramer. Nonlinear Principal Component Analysis Using Autoassociative Neural Networks, *AIChE Journal* **37(2)** (1991), 233–243.

[28] M.E. Tipping, Topographic Mappings And Feed-Forward Neural Networks, Dissertation, Aston University, Birmingham, 1996.

[29] D. Lowe, Radial Basis Function Networks, In: M. A. Arbib (Ed.), The Handbook of Brain Theory and Neural Networks, Cambridge, Mass: MIT Press, (1995) 779–782.

[30] G.E. Hinton, R.R. Salakhutdinov, Reducing The Dimensionality Of Data With Neural Networks, *Science* **313** (2006), 504 – 507.

[31] D. de Ridder, R.P.W. Duin, Sammon's Mapping Using Neural Networks: A comparison, *Pattern Recognition Letters* **18** (1997), 1307-1316.
[32] G. Dzemyda, Visualization of a Set of Parameters Characterized by Their Correlation Matrix, *Computational Statistics and Data Analysis* **36(1)** (2001), 15-30.
[33] G. Dzemyda, V. Tiešis, Visualization of Multidimensional Objects and the Socio-Economical Impact to Activity in EC RTD Databases, *Informatica* **12(2)** (2001), 239-262.
[34] G. Dzemyda, O. Kurasova, Heuristic Approach for Minimizing the Projection Error in the Integrated Mapping, *European Journal of Operational Research* **171(3)** (2006), 859-878.
[35] P. Hellemaa, The Development of Coastal Dunes and their Vegetation in Finland, Dissertation, Fenia 176:1, Helsinki, 1998 (downloadable from website http://ethesis.helsinki.fi/julkaisut/mat/maant/vk/hellemaa/index.html).
[36] G. Dzemyda, O. Kurasova, Dimensionality Problem in the Visualization of Correlation-Based Data, *Proceedings of International Conference on Adaptive and Natural Computing Algorithms – ICANNGA 2007, Lecture Notes in Computer Science* **4432** (2007), 544-553.
[37] R.C.T. Lee, J.R. Slagle H. Blum, A Triangulation Method for Sequential Mapping of Points from *n*-Space to Two-Space, *IEEE Transactions on Computers* **27** (1977), 288-299.
[38] V. Medvedev, G. Dzemyda, Optimization of the Local Search in the Training for SAMANN Neural Network, *Journal of Global Optimization*, Springer **35** (2006), 607-623.
[39] R.A. Fisher, The use of Multiple Measurements in Taxonomic Problem, *Annual Eugenics* **7(II)** (1936), 179-188.
[40] D. Ruppert, R.J. Carroll, Trimmed Least Squares Estimation in the Linear Model, *Journal of the American Statistical Association* **75** (1980), 828-838.
[41] D.M. Hawkins, D. Bradu, G.V. Kass, Location of Several Outliers in Multiple Regression Data Using Elemental Sets, *Technometrics* **26**, (1984), 197–208.
[42] S. Ivanikovas, V. Medvedev, G. Dzemyda, Parallel Realizations of the SAMANN Algorithm, *Proceedings of International Conference on Adaptive and Natural Computing Algorithms – ICANNGA 2007, Lecture Notes in Computer Science* **4432** (2007), 179-188.
[43] A. El-Shaarawi, W. Piegorsch, Encyclopedia of Environmetrics, Wiley, 2001.
[44] G. Dzemyda, Visualization of the Correlation-Based Environmental Data, *Environmetrics* **15(8)** (2004), 827-836.
[45] Water Research Commission, Linking Groundwater Chemistry to atypical Lymphocytes. SA Waterbulletin **25(2)**, 1999, http://www.wrc.org.za/wrcpublications/wrcbulletin/v25_2/lympho.htm
[46] R.L. Smith, CBMS course in environmental statistics, 2001, http://www.stat.unc.edu/postscript/rs/envstat/env.html
[47] D.G. Fautin, R.W. Buddemeier, Biogeoinformatics of Hexacorallia (Corals, Sea Anemones, and Their Allies): Interfacing Geospatial, Taxonomic, and Environmental Data for a Group of Marine Invertebrates, 2001, http://www.kgs.ukans.edu/Hexacoral/Envirodata/Correlations/correl1.htm
[48] E.N. Ieno, Las Comunidades Bentónicas de Fondos Blandos del Norte de la Provincia de Buenos Aires: Su rol Ecológico en el Ecosistema Costero. Tesis Doctoral Universidad Nacional de Mar del Plata, Zoobenthic species measured in an intertidal area in Argentina, 2000, http://www.brodgar.com/benthos.htm
[49] G. Dzemyda, Application of Mapping Methods to the Analysis of Curricula of Studies, *Computational Statistics & Data Analysis* **49** (2005), 265-281.
[50] V. Šaltenis, G. Dzemyda, V. Tiešis, Quantitative Forecasting and Assessment Models in the State Education System, *Informatica* **13(4)** (2002), 485-500.
[51] G. Dzemyda, V. Šaltenis, V. Tiešis, Forecasting Models in the State Education System, *Informatics in Education* **2(1)** (2003) 3-14.
[52] The World Factbook, 1999, http://www.odci.gov/cia/publications/factbook/country.html.
[53] The World Bank Group-Development Data, 1999, http://www.worldbank.org/data/.
[54] A. Paunksnis, V. Barzdžiukas, D. Jegelevičius, S. Kurapkienė, G. Dzemyda, The Use of Information Technologies for Diagnosis in Ophthalmology, *Journal of Telemedicine and Telecare* **12** (2006), 37-40.
[55] J. Bernataviciene, G. Dzemyda, O. Kurasova, V. Marcinkevicius, The Problem of Visual Analysis of Multidimensional Medical Data. In: Torn A, Žilinskas J (eds) *Models and Algorithms for Global Optimization. Springer Optimization and its Applications* **4** (2007), 277-298.
[56] J. Bernataviciene, G. Dzemyda, O. Kurasova, A. Vainoras, Integration of classification and visualization for diagnosis decisions, *International Journal of Information Technology and Intelligent Computing* **1(1)** (2006), 57-68.
[57] J. Bernataviciene, G. Dzemyda, O. Kurasova, V. Marcinkevicius, Decision Support for Preliminary Medical Diagnosis Integrating the Data Mining Methods. In: Pranevičius H, Vaarmann O, Zavadskas E (eds.) *Proceedings of 5th International Conference on Simulation and Optimisation in Business and Industry* (2006), 155-160.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Recommender System Technologies based on Argumentation [1]

Carlos Iván CHESÑEVAR [a,b,c,2], Ana Gabriela MAGUITMAN [a,b] and Guillermo Ricardo SIMARI [a]

[a] *Department of Computer Science and Engineering*
Universidad Nacional del Sur
Av. Alem 1253
B8000CPB Bahía Blanca, Argentina
E-mail: {cic,agm,grs}@cs.uns.edu.ar.
[b] *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina*
[c] *Artificial Intelligence Research Institute (IIIA-CSIC)*
Campus UAB, Bellaterra, Spain[3]

Abstract. In recent years there has been a wide-spread evolution of support tools that help users to accomplish a range of computer-mediated tasks. In this context, recommender systems have emerged as powerful user-support tools which provide assistance to users by facilitating access to relevant items. Nevertheless, recommender system technologies suffer from a number of limitations, mainly due to the lack of underlying elements for performing qualitative reasoning appropriately. Over the last few years, argumentation has been gaining increasing importance in several AI-related areas, mainly as a vehicle for facilitating rationally justifiable decision making when handling incomplete and potentially inconsistent information. In this setting, recommender systems can rely on argumentation techniques by providing reasoned guidelines or hints supported by a rationally justified procedure. This chapter presents a generic argument-based approach to characterize recommender system technologies, in which knowledge representation and inference are captured in terms of Defeasible Logic Programming, a general-purpose defeasible argumentation formalism based on logic programming. As a particular instance of our approach we analyze an argument-based search engine called ArgueNet, an application oriented towards providing recommendations on the web scenario.

Keywords. argumentation, logic programming, user support systems, knowledge engineering, recommender systems

---

[3]The first author is external researcher at IIIA-CSIC within the research project MULOG (TIN 2004-07933-C03-01/03). Ministry of Science and Technology, Spain.

[1]This chapter is based on the paper entitled *"Argument-based User Support Systems using Defeasible Logic Programming"* presented by the authors at the 3rd IFIP Conference on Artificial Intelligence Applications & Innovations (AIAI), June 7-9, 2006, Athens, Greece [11].

[2]Corresponding Author: Carlos Iván Chesñevar – Department of Computer Science and Engineering, Universidad Nacional del Sur, B8000CPB Bahía Blanca, Argentina; E-mail: cic@cs.uns.edu.ar.

1. Introduction

User support systems have evolved in the last years as specialized tools to assist users in a plethora of computer-mediated tasks by providing guidelines or hints [23]. Recommender systems are a special class of user support tools and are mostly based on machine learning and information retrieval algorithms. The resulting systems typically provide suggestions based on *quantitative* evidence (i.e. measures of similarity between objects or users), whereas the inference process which led to these suggestions is commonly unknown (i.e. 'black-box' metaphor). Although the effectiveness of existing recommenders is remarkable, they still have serious limitations. On the one hand, they are incapable of dealing formally with the defeasible nature of users' preferences in complex environments. Decisions about user preferences are mostly based on heuristics which rely on ranking previous user choices or gathering information from other users with similar interests. On the other hand, they are not equipped with inference capabilities. As a consequence, much of the implicit information remains undiscovered.

Quantitative approaches in AI, as opposed to qualitative approaches, have often been criticized for their inability to obtain conclusions supported by a rationally justified procedure. Indeed, the quantitative techniques adopted by most existing user support systems suffer from this limitation. The absence of an underlying formal model makes it hard to provide users with a clear explanation of the factors and procedures that led the system to come up with some particular recommendations. As a result, serious trustworthiness issues may arise, especially in those cases when business interests are involved, or when external manipulation is possible. Logic-based approaches could help to overcome these issues, enhancing recommendation technology by providing a means to formally express constrains and to draw inferences. However, as has been discussed by a number of sources (e.g., [41]), traditional logic-based systems are limited as they are unable to handle rules with exceptions, which are recurrent in recommendation scenarios. We contend that a solution for this problem can be provided by integrating existing user support technologies with appropriate inferential mechanisms for qualitative reasoning.

In this context, *defeasible argumentation* frameworks [9,36] constitute an interesting alternative, as they have matured in the last decade to become a sound setting to formalize commonsense, qualitative reasoning. Recent research has shown that argumentation can be integrated in a growing number of real-world applications such as multiagent systems [1], legal reasoning [37,36], and analysis of news reports [22], among many others. In the last few years, particular attention has been given to extensions of logic programming as a suitable framework for formalizing argumentation in a computationally attractive way. One of such approaches that has been considerably successful is *Defeasible Logic Programming* (DeLP) [18], a general-purpose argumentation formalism based on logic programming. An important feature of the DeLP approach is that, by performing defeasible reasoning dialectically, it can handle information that is in principle contradictory. The process of deciding if a conclusion is supported, or *warranted*, begins by analyzing if there exists an undefeated argument (a *warrant*) supporting that conclusion, i.e, an argument for which every possible attacking argument has been defeated. The notion of *attack* and *defeat* is formally introduced below but can be intuitively described as the consideration of arguments that are in conflict (attack) with the supporting argument. The attack becomes a defeat when the attacking argument is better, in some specific sense, than the supporting argument. As a consequence, the use of argumentation will allow the system to present reasoned suggestions, which the user will be able to further investigate and accept only if a convincing case can be made by the recommendation tool.

This chapter presents a generic approach to characterize *argument-based recommender systems*, *i.e.* user support tools in which recommendations are provided on the basis of arguments. The proposed approach is based on modelling user preference criteria are modelled by means of facts, strict rules and defeasible rules encoded as part of a DeLP program. These preference criteria are combined with additional background information and used by the argumentation framework to prioritize potential suggestions, thus enhancing the final results provided to the user. The rest of the chapter is structured as follows. Section 2 introduces the fundamentals of recommender systems. Section 3 presents the main concepts which characterize argumentation frameworks in general, and DeLP in particular. Section 4 discusses our proposal for integrating defeasible argumentation and recommendation technologies. Section 5 describes a particular real-world application which emerged as an instance of this approach oriented towards providing suitable decision support in the context of web search. Finally, Section 6 discusses related work and Section 7 closes with some conclusions.

2. Recommender Systems: A Brief Overview

User support systems operate in association with the user to effectively accomplish a range of tasks. Some of these systems serve the purpose of expanding the user's natural capabilities, for example by acting as intelligence or memory augmentation mechanisms. Some of these systems reduce the user's work by carrying out the routinizable tasks on the user's behalf. Others offer tips on how to refine or complete human generated products (such as electronic documents) by highlighting potential inaccuracies and proposing alternative solutions. Some aides "think ahead" to anticipate the next steps in a user's task providing the capability for the user to confirm the prediction and ask the system to complete the steps automatically.

Recommender systems are a special class of user support tools that act in cooperation with users, complementing their abilities and augmenting their performance by offering proactive or on demand context-sensitive support. They usually operate by creating a model of the user's preferences or the user's task with the purpose of facilitating access to items (e.g., news, web pages, books, etc.) that the user may find useful [23]. While in many situations the user explicitly posts a request for recommendations in the form of a query, many recommender systems attempt to anticipate the user's needs and are capable of proactively providing assistance [39,8]. In order to come up with recommendations for user queries, conventional recommender systems rely on *similarity measures* between users or contents, computed on the basis of methods coming either from the information retrieval or the machine learning communities. Recommender systems adopt mainly two different views to help predict information needs. The first approach is known as *user modelling* and relies on the use of a profile or model of the users, which is created by observing users' behavior (e.g., [29,15]). The second approach is based on *task modelling*, and recommendations are based on the context in which the user is immersed. The context may consist of an electronic document the user is editing, web pages the user has recently visited, etc. An example of context-aware recommender system is Watson [8], which is part of a family of programs known as Information Management

Assistants (IMAs) developed at the InfoLab of Northwestern University in Chicago. The purpose of the IMAs is to anticipate the user's needs and to provide proactive and on demand support for the user current activity. In order to achieve this goal, the IMAs generate a model of the user's task, access information retrieval systems on the user's behalf, and unobtrusively deliver useful material. The IMAs provide an environment in which resources are retrieved proactively as well as mechanisms that augment users' explicit queries with keywords extracted from the current task. Another context-aware system, CALVIN [26], monitors the user's web browsing activity and applies case-based reasoning techniques to proactively and unobtrusively suggest stored material when the user context is similar to the one associated with the stored cases.

Two main techniques have been used to compute recommendations: *content-based* and *collaborative filtering*. Content-based recommenders are driven by the premise that user's preferences tend to persist through time. These recommenders frequently use machine-learning techniques to generate a profile of the active user. Typically, a model of the active user is stored as a list of rated items. In order to determine if a new item is a potentially good recommendation, content-based recommender systems rely on *similarity measures* between the new items and the rated items stored as part of the user model.

Recommender systems based on collaborative filtering are based on the assumption that users' preferences are correlated. These systems maintain a pool of users' profiles associated with items that the users rated in the past. For a given active user, collaborative recommender systems find other similar users whose ratings strongly correlate with the current user. New items not rated by the active user can be presented as suggestions if similar users have rated them highly. Tapestry [20] is usually referred to as the first collaborative filtering system. It provided a mechanism for filtering email and news messages based both on the content of the messages and on implicit or explicit feedback from users. Feedback included manual annotations and automatically observed reactions (e.g., some user sent a reply to a message). Following Tapestry's initiative, a large number of recommender systems were developed and applied to diverse domains, providing different levels of personalization. Given the huge amount of information existing on the web it is not surprising that the great majority of the recommender systems have been built around content and resources available online. WebWatcher [2] is an early attempt to assists users to locate information on the web by highlighting hyperlinks in a page based on the declared preferences and browsing history of a user as well as information gathered from other users with similar interests. Letizia [27] is a user interface agent that unobtrusively assists web browsing. As the user navigates web pages, Letizia performs a breadth-first search augmented by several heuristics to anticipate what items may be of interest to the user. Syskill & Webert [33] uses information retrieval techniques to process the content of a page rated by a user, and machine learning to acquire a model that is utilized to predict which links on a web page a user will find useful.

A combination of collaborative-filtering and content-based recommendation gives rise to hybrid recommender systems. Fab [4] is a hybrid content-based, collaborative web page recommender system that learns to browse the web on behalf of a user. Fab generates recommendations by the use of a set of collection agents (that find pages for a particular topic) and selection agents (that find pages for a particular user). Users' explicit ratings of the recommended pages combined with several heuristics are used to update personal-agents' profiles, remove unsuccessful agents, and duplicate successful

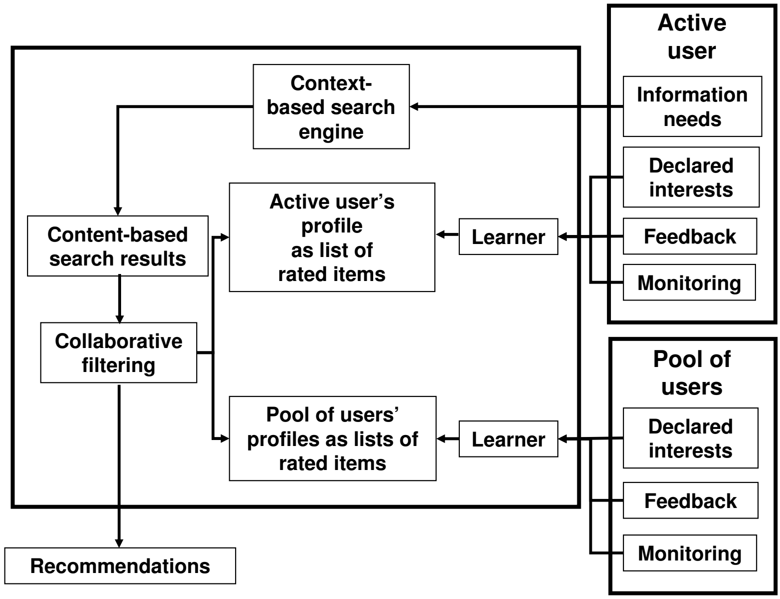


Figure 1. A schematic view of a hybrid recommender system

ones. Examples of hybrid news filtering systems are NewsDude [7], a learning agent that is trained by the user with a set of interesting news articles, and Butterfly [44], a system that uses keywords to find interesting conversations in Usenet newsgroups. Collaborative news recommender systems include GroupLens [38] and PHOAKS [42]. P-Tango [14] is an instance of hybrid recommender for personalized news filtering.

Figure 1 depicts a generic architecture for a hybrid recommender systems, where the main elements associated with collaborative filtering and content-based techniques can be identified. Additional dimensions of analysis for recommender systems are the content of the suggestion (e.g., news, URLs, people, articles, text, products), the purpose of the suggestion (sales or information), the event that triggers the search for suggestions (user's demand or proactive), and the level of intrusiveness (none, low, moderate or high).

One of the main issues faced by recommender systems is the "sparsity problem". This problem is due to the reluctance of users to rate items. Therefore, much research has focused on the development of strategies for filling in incomplete users' models. The system can learn the user's interests by receiving feedback from the user, engaging in conversation with the user, or it can be programmed. However, monitoring the users' behavior is one of the most common and successful approaches to deal with the sparsity problem.[4] While this approach minimizes the user's effort and generates large amounts of data, it has the disadvantage of resulting in noisy descriptions of the user's interests. Another issue faced by recommendation systems is that of trustworthiness. This is some-

[4]See for example the Amazon recommendation system for books (http://amazon.com).

times due to the absence of any rationale supporting the presented suggestions. Therefore, the user is unable to evaluate the reasons that led the system to present certain recommendations. In certain domains (e.g., e-commerce), this lack of justification can be associated with ulterior motives on the recommendation provider's side, leading to lack of confidence or reliability. Typical approaches to recommendation (especially those based on IR and machine learning techniques) are usually limited in this sense.

As we will later see in Section 5, the use of argumentation will allow to enhance recommender systems with inference abilities to present reasoned suggestions, which the user will be able to further investigate and accept only if a convincing case can be made by the recommendation tool.

3. Argumentation in AI: Background

Artificial Intelligence (AI) has long dealt with the enormous challenge of modelling commonsense reasoning, which almost always occurs in the face of incomplete and potentially inconsistent information [30]. A logical model of commonsense reasoning demands the formalization of principles and criteria that characterize valid patterns of inference. In this respect, classical logic has proven to be inadequate, since it behaves *monotonically* and cannot deal with inconsistencies at object level.

When a rule supporting a conclusion may be defeated by new information, it is said that such reasoning is *defeasible* [34,32,41]. When we chain defeasible reasons or rules to reach a conclusion, we have *arguments* instead of proofs. Arguments may compete, rebutting each other, so a *process* of argumentation is a natural result of the search for arguments. Adjudication of competing arguments must be performed, comparing arguments in order to determine what beliefs are ultimately accepted as *warranted* or *justified*. Preference among conflicting arguments is defined in terms of a *preference criterion* which establishes a relation " $\preceq$ " among possible arguments; thus, for two arguments A and B in conflict, it may be the case that A is strictly preferred over B ($A \succ B$), that A and B are equally preferable ($A \succeq B$ and $A \preceq B$) or that A and B are not comparable with each other. In the above setting, since we arrive at conclusions by building defeasible arguments, and since *logical argumentation* is usually called *argumentation*, we sometimes call this kind of reasoning *defeasible argumentation*.

For the sake of example, let us consider a well-known problem of nonmonotonic reasoning in AI about the flying abilities of birds, recast in argumentative terms. Consider the following sentences:

1. *Birds usually fly.*
2. *Penguins usually do not fly.*
3. *Penguins are birds.*

The first two sentences correspond to *defeasible rules* (rules which are subject to possible exceptions). The third sentence is a *strict rule*, where no exceptions are possible. Now, given the fact that *Tweety is a penguin* two different arguments can be constructed:

1. Argument A (based on rules 1 & 3): Tweety is a penguin. Penguins are birds. Birds usually fly. So Tweety flies.

2. Argument B (based on rule 2): Tweety is a penguin. Penguins usually do not fly. So Tweety does not fly.

In this particular situation, two arguments arise that cannot be accepted simultaneously (as they reach contradictory conclusions). Note that argument B seems rationally preferable over argument A, as it is based on more *specific* information. As a matter of fact, specificity is commonly adopted as a syntax-based criterion among conflicting arguments, preferring those arguments which are *more informed* or *more direct* [35]. In this particular case, if we adopt specificity as a preference criterion, argument B is justified, whereas A is not (as it is defeated by B). The above situation can easily become much more complex, as an argument may be defeated by a second argument, which in turn can be defeated by a third argument, *reinstating* the first one.

Several defeasible argumentation frameworks have been developed on the basis of extensions to logic programming (see [9,36]). *Defeasible logic programming* (DeLP) [18] is one of such formalisms, combining results from defeasible argumentation theory [41] and logic programming. DeLP is a suitable framework for building real-world applications that deal with incomplete and contradictory information in dynamic domains. In what follows we will present a brief overview of the DeLP framework. A more in-depth treatment can be found elsewhere [18].

3.1. Defeasible Logic Programming: A Short Introduction

Defeasible logic programming (DeLP) [18] is a general-purpose defeasible argumentation formalism based on logic programming, intended to model reasoning with inconsistent and potentially contradictory knowledge. This formalism provides a knowledge representation language which gives the possibility of representing tentative information in a declarative manner, and a reasoning mechanism with considers all ways a conclusion could be supported and decides which one has the best support. In the following paragraph we will give a brief introduction.[5]

A defeasible logic program is a set (Π, Δ) of rules, where Π and Δ stand for sets of *strict* and *defeasible* knowledge, resp. The set Π involves *strict rules* of the form $P \leftarrow Q_1, \ldots, Q_k$ and *facts* (strict rules with empty body), and it is assumed to be *non-contradictory*. The set Δ involves *defeasible rules* of the form $P \prec Q_1, \ldots, Q_k$. The underlying logical language is that of extended logic programming [19], enriched with a special symbol "$\prec$" to denote defeasible rules. Both default and classical negation are allowed (denoted not and $\sim$, resp.). Literals preceded by not are called *extended literals* [19]. DeLP rules are to be thought of as *inference rules* rather than implications in the object language.

Example 1 (Adapted from [10]) *Consider an intelligent agent controlling an engine with an oil pump, a fuel pump and an engine, as well as three switches* $sw1$*,* $sw2$ *and* $sw3$*. These switches regulate different features of the engine, such as the pumping system, speed, etc. This agent may have the following defeasible knowledge base for diagnosing possible problems with the engine:*

- *If the pump is clogged, then the engine gets no fuel.*

[5]For space reasons, we will restrict ourselves to a basic set of definitions and concepts which make this chapter self-contained. For more details, see [18,9].

- *When $sw1$ is on, normally fuel is pumped properly.*
- *When fuel is pumped properly, fuel seems to work ok.*
- *When $sw2$ is on, usually oil is pumped.*
- *When oil is pumped, usually it works ok.*
- *When there is oil and fuel, usually the engine works ok.*
- *When there is heat, then the engine is usually not ok.*
- *When there is heat, normally there are oil problems.*
- *When fuel is pumped and speed is low, then there are reasons to believe that the pump is clogged.*
- *When $sw2$ is on, usually speed is low.*
- *When $sw2$ and $sw3$ are on, usually speed is not low.*
- *When $sw3$ is on, usually fuel is ok.*

Suppose also that the agent knows some particular facts: $sw1$, $sw2$ and $sw3$ are on, and there is heat. The knowledge of such an agent can be modelled by the DeLP program $\mathcal{P}_{eng} = (\Pi_{eng}, \Delta_{eng})$ shown in Figure 2, where the set Π_{eng} of strict knowledge corresponds to clauses $(1) - (4)$, and the set Δ_{eng} of defeasible knowledge corresponds to clauses $(5) - (16)$.

Deriving literals in DeLP results in the construction of *arguments*. Formally:

Definition 1 (Argument) *Given a DeLP program $\mathcal{P}$, and let Q be a literal in $\mathcal{P}$. An* argument *$\mathcal{A}$ for the query Q, denoted $\langle \mathcal{A}, Q \rangle$, is a subset of ground instances of defeasible rules in $\mathcal{P}$ such that:*

1. *there exists a* defeasible derivation *for Q from $\Pi \cup \mathcal{A}$;*
2. *$\Pi \cup \mathcal{A}$ is non-contradictory (i.e, $\Pi \cup \mathcal{A}$ does not entail two complementary literals P and $\sim P$, nor does $\mathcal{A}$ contain literals S and* not *S)*[6] *and*
3. *$\mathcal{A}$ is the minimal set (with respect to set inclusion) satisfying (1) and (2).*

An argument $\langle \mathcal{A}_1, Q_1 \rangle$ is a sub-argument *of another argument $\langle \mathcal{A}_2, Q_2 \rangle$ if $\mathcal{A}_1 \subseteq \mathcal{A}_2$. Given a DeLP program $\mathcal{P}$, $Args(\mathcal{P})$ denotes the set of all possible arguments that can be derived from $\mathcal{P}$.*

The notion of defeasible derivation corresponds to the usual query-driven SLD derivation used in logic programming, performed by backward chaining on both strict and defeasible rules, but only "collecting" defeasible rules as part of the argument. In this context a negated literal $\sim P$ is treated just as a new predicate name no_P. Minimality imposes the 'Occam's razor principle' [41] on arguments. Any superset $\mathcal{A}'$ of $\mathcal{A}$ can be proven to be 'weaker' than $\mathcal{A}$ itself, as the former relies on more defeasible information. The non-contradiction requirement forbids the use of (ground instances of) defeasible

[6]We use the term "contradictory" instead of "inconsistent" to avoid confusion, as the latter is commonly used in the context of classical logic.

$$
\begin{array}{ll}
(1) & \sim fuel\_ok \;\leftarrow\; pump\_clog \\
(2) & sw1 \\
(3) & sw2 \\
(4) & sw3 \\
(5) & heat \\
(6) & pump\_fuel \;-\!\!\!\prec\; sw1 \\
(7) & fuel\_ok \;-\!\!\!\prec\; pump\_fuel \\
(8) & pump\_oil \;-\!\!\!\prec\; sw2 \\
(9) & oil\_ok \;-\!\!\!\prec\; pump\_oil \\
(10) & engine\_ok \;-\!\!\!\prec\; fuel\_ok, oil\_ok \\
(11) & \sim engine\_ok \;-\!\!\!\prec\; heat \\
(12) & \sim oil\_ok \;-\!\!\!\prec\; heat \\
(13) & pump\_clog \;-\!\!\!\prec\; pump\_fuel, low\_speed \\
(14) & low\_speed \;-\!\!\!\prec\; sw2 \\
(15) & \sim low\_speed \;-\!\!\!\prec\; sw2, sw3 \\
(16) & fuel\_ok \;-\!\!\!\prec\; sw3
\end{array}
$$

Figure 2. DeLP program $\mathcal{P}_{eng}$ (example 1)

rules in an argument $\mathcal{A}$ whenever $\Pi \cup \mathcal{A}$ entails complementary literals. It should be noted that non-contradiction captures the two usual approaches to negation in logic programming (viz. default negation and classic negation), both of which are present in DeLP and are related to the notion of counterargument, as we will see in the next Definition.

Example 2 *Consider the program $\mathcal{P}_{eng}$ in Example 1. Arguments $\langle \mathcal{B}, fuel\_ok \rangle$ and $\langle \mathcal{C}, oil\_ok \rangle$ can be derived from $\mathcal{P}_{eng}$, with*

$$
\begin{array}{c}
\mathcal{B} = \{ pump\_fuel \;-\!\!\!\prec\; sw1;\, fuel\_ok \;-\!\!\!\prec\; pump\_fuel \} \\
\mathcal{C} = \{ pump\_oil \;-\!\!\!\prec\; sw2; oil\_ok \;-\!\!\!\prec\; pump\_oil \}.^{7}
\end{array}
$$

Similarly, an argument $\langle \mathcal{A}_1, engine\_ok \rangle$ can be derived from $\mathcal{P}_{eng}$, where

$$\mathcal{A}_1 = \{ engine\_ok \;-\!\!\!\prec\; fuel\_ok, oil\_ok \} \cup \mathcal{B} \cup \mathcal{C}$$

Note that in this last case the arguments $\langle \mathcal{C}, oil\_ok \rangle$ and $\langle \mathcal{B}, fuel\_ok \rangle$ are subarguments *of $\langle \mathcal{A}_1, engine\_ok \rangle$.*

Definition 2 (Counterargument – Defeat) *An argument $\langle \mathcal{A}_1, Q_1 \rangle$ is a* counterargument *for an argument $\langle \mathcal{A}_2, Q_2 \rangle$ (or equivalently $\langle \mathcal{A}_1, Q_1 \rangle$ counterargues $\langle \mathcal{A}_2, Q_2 \rangle$) if and only if*

[7] For the sake of clarity, we use semicolons to separate elements in an argument $\mathcal{A} = \{ e_1 \,;\, e_2 \,;\, \ldots ; e_k \,\}$.

1. *There is an subargument* $\langle \mathcal{A}, Q\rangle$ *of* $\langle \mathcal{A}_2, Q_2\rangle$ *such that the set* $\Pi \cup \{Q_1, Q\}$ *is contradictory, or*
2. *An extended literal* not Q_1 *is present in some rule in* $\mathcal{A}_2$.[8]

A preference criterion $\preceq\ \subseteq Args(\mathcal{P}) \times Args(\mathcal{P})$ *will be used to decide among conflicting arguments. An argument* $\langle \mathcal{A}_1, Q_1\rangle$ *is a* defeater *for an argument* $\langle \mathcal{A}_2, Q_2\rangle$ *if* $\langle \mathcal{A}_1, Q_1\rangle$ *counterargues* $\langle \mathcal{A}_2, Q_2\rangle$*, and* $\langle \mathcal{A}_1, Q_1\rangle$ *is preferred over* $\langle \mathcal{A}_2, Q_2\rangle$ *with respect to* $\preceq$*. For cases (1) and (2) above, we distinguish between* proper *and* blocking defeaters *as follows:*

- *In case (1), the argument* $\langle \mathcal{A}_1, Q_1\rangle$ *will be called a* proper defeater *for* $\langle \mathcal{A}_2, Q_2\rangle$ *if and only if* $\langle \mathcal{A}_1, Q_1\rangle$ *is strictly preferred over* $\langle \mathcal{A}, Q\rangle$ *with respect to* $\preceq$*.*
- *In case (1), if* $\langle \mathcal{A}_1, Q_1\rangle$ *and* $\langle \mathcal{A}, Q\rangle$ *are unrelated to each other with respect to* $\preceq$*, or in case (2),* $\langle \mathcal{A}_1, Q_1\rangle$ *will be called a* blocking defeater *for* $\langle \mathcal{A}_2, Q_2\rangle$*.*

It must be noted that in DeLP there is no explicit distinction between undercutting and rebutting defeat [34], as both of them are subsumed by the notion of counterargument. The notions of proper and blocking defeat in Def. 2 distinguish between defeaters which are "strictly better" and defeaters which are "as good as" the argument under attack, respectively.

Example 3 *Consider the argument* $\langle \mathcal{A}_1, engine\_ok\rangle$ *given in Example 2 wrt the program* $\mathcal{P}_{eng}$*. A counteragument for* $\langle \mathcal{A}_1, engine\_ok\rangle$ *can be found, namely the argument* $\langle \mathcal{A}_2, \sim fuel\_ok\rangle$*, with*

$$\begin{array}{rcl} \mathcal{A}_2 & = & \{\ pump\_fuel \prec sw1,\ low\_speed \prec sw2, \\ & & pump\_clog \prec pump\_fuel, low\_speed\}\ \} \end{array}$$

Argument $\langle \mathcal{A}_2, \sim fuel\_ok\rangle$ *is a counterargument for* $\langle \mathcal{A}_1, engine\_ok\rangle$ *as there exists a subargument* $\langle \mathcal{B}, fuel\_ok\rangle$ *in* $\langle \mathcal{A}_1, engine\_ok\rangle$ *(see Example 2) such that* $\Pi_{eng} \cup \{fuel\_ok, \sim fuel\_ok\}$ *is contradictory.*

Specificity [41] is used in DeLP as a syntactic preference criterion among conflicting arguments, favoring those arguments that are *more informed* or *more direct* [41]. However, other alternative preference criteria could also be used [18].

Example 4 *Consider the arguments* $\langle \mathcal{A}_1, engine\_ok\rangle$ *and* $\langle \mathcal{A}_2, \sim fuel\_ok\rangle$ *in Example 3. Then* $\langle \mathcal{A}_2, \sim fuel\_ok\rangle$ *is a proper defeater for* $\langle \mathcal{A}_1, engine\_ok\rangle$*, as the argument* $\langle \mathcal{A}_2, \sim fuel\_ok\rangle$ *counterargues* $\langle \mathcal{A}_1, engine\_ok\rangle$ *with disagreement subargument* $\langle \mathcal{B}, fuel\_ok\rangle$*, and* $\langle \mathcal{A}_2, \sim fuel\_ok\rangle$ *is strictly more specific than* $\langle \mathcal{B}, fuel\_ok\rangle$*.*

Given an argument $\langle \mathcal{A}, Q\rangle$, the definitions of counterargument and defeat allows to detect whether other possible arguments $\langle \mathcal{B}_1, Q_1\rangle, \ldots, \langle \mathcal{B}_k, Q_k\rangle$ are defeaters for $\langle \mathcal{A}, Q\rangle$. Should the argument $\langle \mathcal{A}, Q\rangle$ be defeated, then it would be no longer supporting its conclusion Q. However, since defeaters are arguments, they may on their turn be defeated. That prompts for a complete recursive dialectical analysis to determine which arguments

[8]The first notion of attack is borrowed from the Simari-Loui framework [41]; the second one is related to Dung's argumentative approach to logic programming [17] as well as to other formalizations, such as [37,24]. For an in-depth discussion see [18].

are ultimately defeated. To characterize this process we will introduce some auxiliary notions.

An *argumentation line* starting in $\langle \mathcal{A}_0, Q_0 \rangle$ (denoted $\lambda^{\langle \mathcal{A}_0, Q_0 \rangle}$) is a sequence $[\langle \mathcal{A}_0, Q_0 \rangle, \langle \mathcal{A}_1, Q_1 \rangle, \langle \mathcal{A}_2, Q_2 \rangle, \ldots, \langle \mathcal{A}_n, Q_n \rangle, \ldots]$ that can be thought of as an exchange of arguments between two parties, a *proponent* (evenly-indexed arguments) and an *opponent* (oddly-indexed arguments). Each $\langle \mathcal{A}_i, Q_i \rangle$ is a defeater for the previous argument $\langle \mathcal{A}_{i-1}, Q_{i-1} \rangle$ in the sequence, $i > 0$. In order to avoid *fallacious* reasoning, dialectical constraints are imposed on such an argument exchange to be considered rationally acceptable in light of a given program $\mathcal{P}$, namely:

1. **Non-contradiction:** given an argumentation line λ, the set of arguments of the proponent (resp. opponent) should be *non-contradictory* wrt $\mathcal{P}$.
2. **No circular argumentation:** given an argumentation line λ, no argument $\langle \mathcal{A}_j, Q_j \rangle$ is allowed to appear as a sub-argument of another argument $\langle \mathcal{A}_i, Q_i \rangle$, for $i < j$.
3. **Progressive argumentation:** every blocking defeater $\langle \mathcal{A}_i, Q_i \rangle$ in λ is defeated by a proper defeater $\langle \mathcal{A}_{i+1}, Q_{i+1} \rangle$ in λ.

The first condition disallows the use of contradictory information on either side (proponent or opponent). The second condition eliminates the "circular reasoning" fallacy. The last condition enforces the use of a stronger argument to defeat an argument which acts as a blocking defeater. An argumentation line satisfying the above restrictions is called *acceptable*, and can be proven to be finite. It must be noted that other argumentative frameworks introduce similar constraints on argumentation lines to avoid infinite "chains" of defeaters, reciprocal defeaters, etc. A more detailed analysis on such situations can be found in [36].

Example 5 *Consider the argument $\langle \mathcal{A}_1, engine\_ok \rangle$ and the defeater $\langle \mathcal{A}_2, \sim fuel\_ok \rangle$ in Example 4. Note that the argument $\langle \mathcal{A}_2, \sim fuel\_ok \rangle$ has the associated subargument $\langle \mathcal{A}_2', low\_speed \rangle$, with $\mathcal{A}_2' = \{ low\_speed \prec sw2 \}$. From the program $\mathcal{P}_{eng}$ (Figure 2) a blocking defeater for $\langle \mathcal{A}_2, \sim fuel\_ok \rangle$ can be derived, namely $\langle \mathcal{A}_3, \sim low\_speed \rangle$. Note that this third defeater can be thought of as an answer of the proponent to the opponent,* reinstating *the first argument $\langle \mathcal{A}_1, engine\_ok \rangle$, as it defeats the opponent's defeater $\langle \mathcal{A}_2, \sim fuel\_ok \rangle$. The above situation can be expressed in the following argumentation line:*

$$[\langle \mathcal{A}_1, engine\_ok \rangle, \langle \mathcal{A}_2, \sim fuel\_ok \rangle, \langle \mathcal{A}_3, \sim low\_speed \rangle].$$

Note that the proponent's last defeater in the above sequence could be on its turn defeated by a blocking defeater $\langle \mathcal{A}_2', low\_speed \rangle$, resulting in

$$[\langle \mathcal{A}_1, engine\_ok \rangle, \langle \mathcal{A}_2, \sim fuel\_ok \rangle, \langle \mathcal{A}_3, \sim low\_speed \rangle, \langle \mathcal{A}_2', low\_speed \rangle \ldots]$$

However, such line is not acceptable*, as it violates the condition of non-circular argumentation.*

Given a program $\mathcal{P}$ and an initial argument $\langle \mathcal{A}_0, Q_0 \rangle$, the set of all acceptable argumentation lines starting in $\langle \mathcal{A}_0, Q_0 \rangle$ accounts for a whole dialectical analysis for

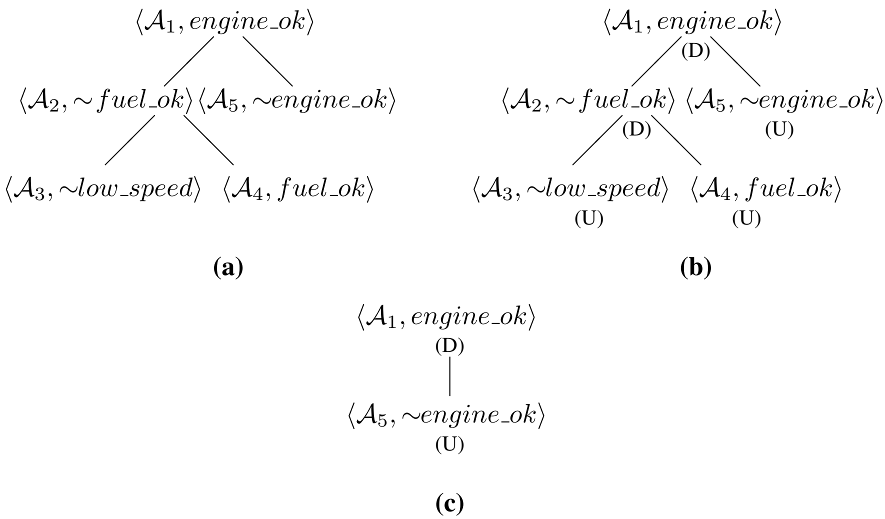


Figure 3. (a) Dialectical tree for $\langle \mathcal{A}_1, engine\_ok \rangle$; (b) marked dialectical tree for $\langle \mathcal{A}_1, engine\_ok \rangle$; (c) associated pruned dialectical tree

$\langle \mathcal{A}_0, Q_0 \rangle$ (i.e., all possible dialogues rooted in $\langle \mathcal{A}_0, Q_0 \rangle$), formalized as a *dialectical tree* $\mathcal{T}_{\langle \mathcal{A}_0, Q_0 \rangle}$.

Example 6 *Consider $\langle \mathcal{A}_1, engine\_ok \rangle$ from Example 2, and the argumentation line shown in Example 5. Note that the argument $\langle \mathcal{A}_2, \sim fuel\_ok \rangle$ has a second (blocking) defeater $\langle \mathcal{A}_4, fuel\_ok \rangle$. The argument $\langle \mathcal{A}_1, engine\_ok \rangle$ has also a second defeater $\langle \mathcal{A}_5, \sim engine\_ok \rangle$. There are no more arguments to consider.*

There are three acceptable argumentation lines rooted in $\langle \mathcal{A}_1, engine\_ok \rangle$, namely:

$$\begin{aligned}
\lambda_1^{\langle \mathcal{A}_1, engine_ok \rangle} &= [\, \langle \mathcal{A}_1, engine_ok \rangle, \langle \mathcal{A}_2, \sim fuel_ok \rangle, \langle \mathcal{A}_3, \sim low_speed \rangle \,] \\
\lambda_2^{\langle \mathcal{A}_1, engine_ok \rangle} &= [\, \langle \mathcal{A}_1, engine_ok \rangle, \langle \mathcal{A}_2, \sim fuel_ok \rangle, \langle \mathcal{A}_4, fuel_ok \rangle \,] \\
\lambda_3^{\langle \mathcal{A}_1, engine_ok \rangle} &= [\, \langle \mathcal{A}_1, engine_ok \rangle, \langle \mathcal{A}_5, \sim engine_ok \rangle \,]
\end{aligned}$$

The corresponding dialectical tree $\mathcal{T}_{\langle \mathcal{A}_1, engine\_ok \rangle}$ rooted in the argument $\langle \mathcal{A}_1, engine\_ok \rangle$ is shown in Figure 3.

Nodes in a dialectical tree $\mathcal{T}_{\langle \mathcal{A}_0, Q_0 \rangle}$ can be marked as *undefeated* and *defeated* nodes (U-nodes and D-nodes, resp.): all leaves in $\mathcal{T}_{\langle \mathcal{A}_0, Q_0 \rangle}$ will be marked U-nodes (as they have no defeaters), and every inner node is to be marked as *D-node* iff it has at least one U-node as a child, and as *U-node* otherwise. An argument $\langle \mathcal{A}_0, Q_0 \rangle$ is ultimately accepted as valid (or *warranted*) with respect to a DeLP program $\mathcal{P}$ iff the root of its associated dialectical tree $\mathcal{T}_{\langle \mathcal{A}_0, Q_0 \rangle}$ is labeled as *U-node*.

Example 7 *Consider the dialectical tree* $\mathcal{T}_{\langle \mathcal{A}_1, engine\_ok \rangle}$ *from Example 6. The marking procedure results in the nodes of* $\mathcal{T}_{\langle \mathcal{A}_1, engine\_ok \rangle}$ *marked as* U*-nodes and* D*-nodes as shown in Figure 3.*[9]

Solving a query Q with respect to a given program $\mathcal{P}$ accounts for determining whether Q is supported by a warranted argument. Different doxastic attitudes are distinguished when answering query q according to the associated status of warrant, in particular:

1. Yes: accounts for believing Q iff there is at least one warranted argument supporting Q on the basis of $\mathcal{P}$.
2. No: accounts for believing $\sim Q$ iff there is at least one warranted argument supporting $\sim Q$ on the basis of $\mathcal{P}$.
3. $Undecided$: neither Q nor $\sim Q$ are warranted wrt $\mathcal{P}$.
4. $Unknown$: Q does not belong to the signature of $\mathcal{P}$.

Thus, according to DeLP semantics, given a program $\mathcal{P}$, solving a query Q will result in a value belonging to the set Ans= $\{Yes, No, Undecided, Unknown\}$.

Example 8 *Consider program* $\mathcal{P}_{eng}$*, and the goal* $engine\_ok$*. The only argument supporting* $engine\_ok$ *is not warranted (as shown in Figure 3). On the contrary, there exists an argument* $\langle \mathcal{A}_5, \sim engine\_ok \rangle$ *supporting* $\sim engine\_ok$*, and such argument has no defeaters, and therefore it is warranted. The answer to goal* $engine\_ok$ *will therefore be* NO.

Consider now the same program $\mathcal{P}_{eng}$*, and the goal* $fuel\_ok$*. The only argument supporting* $fuel\_ok$ *is* $\langle \mathcal{A}_4, fuel\_ok \rangle$*, which is defeated by a blocking defeater* $\langle \mathcal{A}_2, \sim fuel\_ok \rangle$*. The analysis for* $\langle \mathcal{A}_2, \sim fuel\_ok \rangle$ *is analogous, as this argument is defeated by* $\langle \mathcal{A}_4, fuel\_ok \rangle$*. Thus both arguments 'block' each other, neither of them being warranted. The resulting answer is* UNDECIDED.

The computation of dialectical trees is performed automatically by the DeLP interpreter on the basis of the program available. As shown in Figure 4, this process is based on an abstract machine which extends Warren's abstract machine for PROLOG [18]. The emerging semantics is skeptical, computed by DeLP on the basis of the goal-directed construction and marking of dialectical trees, which is performed in a depth-first fashion. Additional facilities (such as visualization of dialectical trees, zoom-in/zoom-out view of arguments, etc.) are integrated in the DeLP environment to facilitate user interaction when solving queries. The DeLP environment is available online to test at `http://lidia.cs.uns.edu.ar/delp_client`.

4. Argument-based Recommender Systems using DeLP

Argument-based reasoning can be integrated into recommender systems in order to provide a qualitative perspective in decision making. This can be achieved by integrating

[9]The search space associated with dialectical trees is reduced by applying $\alpha - \beta$ pruning [12] (e.g., in Figure 3, if the right branch is computed first, then the left branch of the tree does not need to be computed).

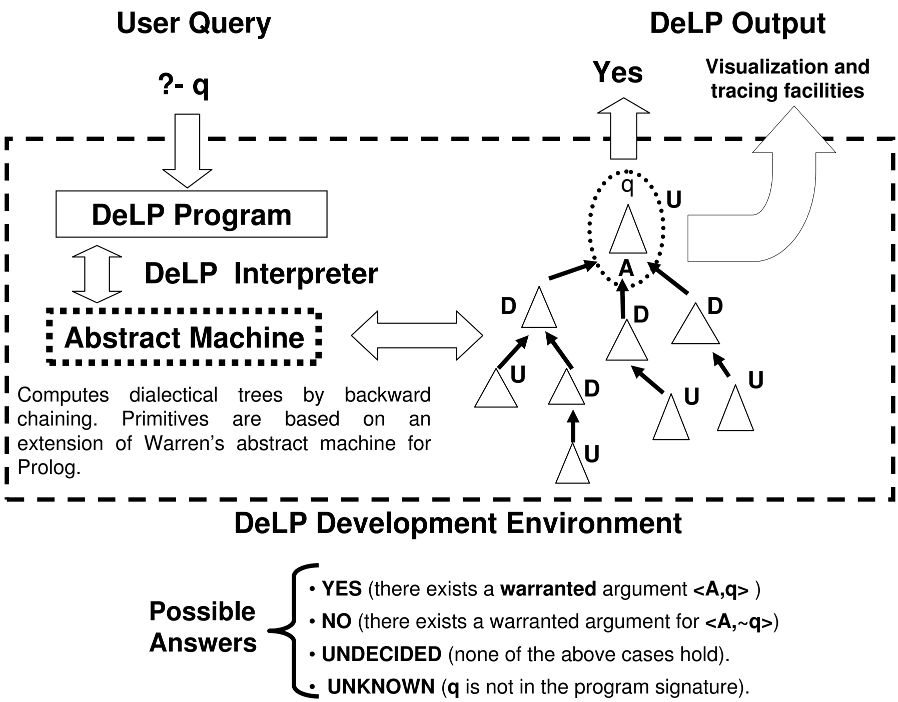


Figure 4. The DeLP Framework

inference abilities to offer reasoned suggestions modelled in terms of arguments in favor and against a particular decision. This approach complements existing qualitative techniques by enriching the user's mental model of such computer systems in a natural way: suggestions are statements which are backed by arguments supporting them. Clearly, conflicting suggestions may arise, and it will be necessary to determine which suggestions can be considered as valid according to some rationally justified procedure. The role of argumentation is to provide a sound formal framework as a basis for such analysis.

In this context, our proposal is based on modelling users' preference criteria in terms of a DeLP program built on top of a traditional content-based search engine. Figure 4 depicts the basic architecture of a generic argument-based user support system based on DeLP. In such a setting users' preferences and background knowledge can be codified as facts, strict rules and defeasible rules in a DeLP program. These facts and rules can come from different sources. For example, user's preferences could be entered explicitly by the user or could be inferred by the system (e.g., by monitoring the user's behavior). Additional facts and rules could be obtained from other repositories of structured (e.g., databases) and semistructured data (e.g., the web.).

We will distinguish particular subsets in a DeLP program, representing different elements in a user support system. For example, a DeLP program could take the form $\mathcal{P} = \mathcal{P}_{user} \cup \mathcal{P}_{pool} \cup \mathcal{P}_{domain}$, where sets $\mathcal{P}_{user}$ and $\mathcal{P}_{pool}$ represent preferences and behavior of the active user and the pool of users, respectively. In the case of the active user, his/her profile can be encoded as facts and rules in DeLP. In the case of the pool of

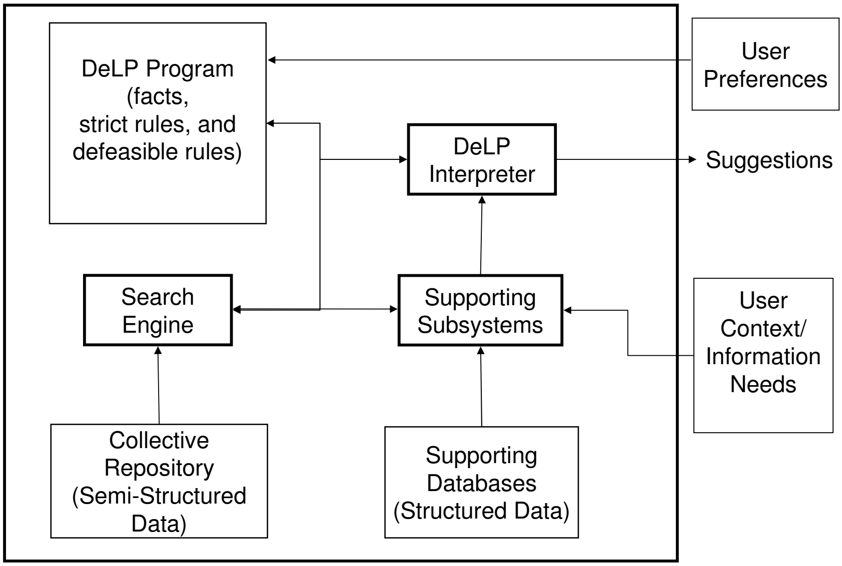


Figure 5. A Generic Argument-Based User Support System based on DeLP

users, rule induction techniques are in order[10] resulting in defeasible rules characterizing trends and general preference criteria (e.g., *normally if a given user likes X then she also likes Y*). The set $\mathcal{P}_{domain}$ represents the domain (background) knowledge, encoded using facts and rules in DeLP. Either proactively or upon a user's request, an argument-based user support system triggers the search for suggestions. If needed, the collected results could be codified as facts and added to the DeLP program. Finally, a DeLP interpreter is in charge of performing the qualitative analysis on the program and to provide the final suggestions to the user.

Given the program $\mathcal{P}$, a user's request is transformed into suitable DeLP queries, from which different *suggestions* are obtained. For the sake of simplicity, we will assume in our analysis that user suggestions will be DeLP terms associated with a distinguished predicate name rel (which stands for *relevant* or *acceptable as a valid suggestion*). Using this formalization, suggestions will be classified into three sets, namely:

- S^w (warranted suggestions): those suggestions s_i for which there exists at least one warranted argument supporting $rel(s_i)$ based on $\mathcal{P}$;
- S^u (undecided suggestions): those suggestions s_i for which there is no warranted argument for $rel(s_i)$, neither there is a warranted argument for $\sim rel(s_i)$ on the basis of $\mathcal{P}$, and
- S^d (defeated suggestions): those suggestions s_i such that there is a warranted argument supporting $\sim rel(s_i)$ on the basis of $\mathcal{P}$.

Given a potential suggestion s_i, the existence of a warranted argument $\langle \mathcal{A}_1, rel(s_i) \rangle$ built on the basis of the DeLP program $\mathcal{P}$ will allow to conclude that s_i should be presented as a final suggestion to the user. If results are presented as a ranked list of sug-

[10] An approach for inducing defeasible rules from association rules can be found in [21].

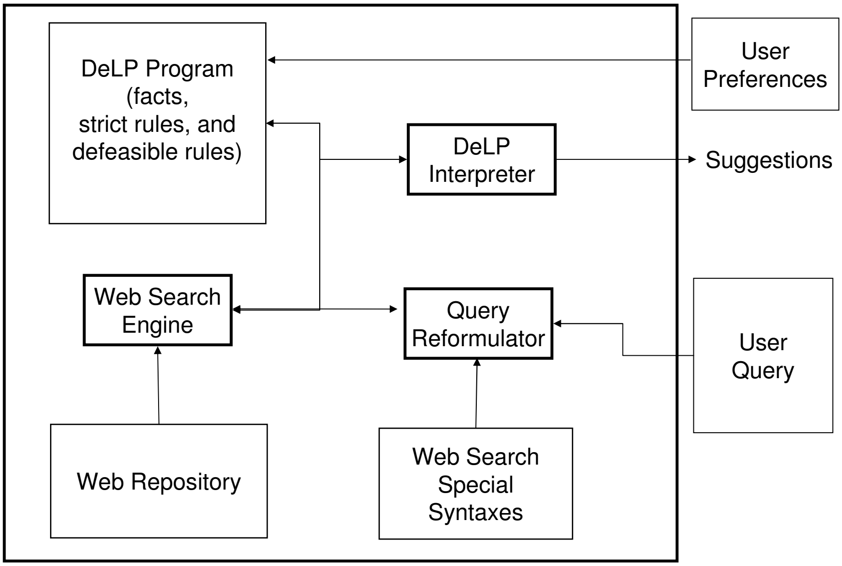


Figure 6. The ARGUENET Framework as a particular instance of the Generic Argument-Based User Support System for argument-based web search

gestions, then warranted suggestions will be more relevant than those which are undecided or defeated. Note that the above classification has a direct correspondence with the doxastic attitudes associated with answers to DeLP queries.

5. ARGUENET: Argument-based User Support for Web Search

Next, we will present a concrete instantiation of an argument-based user support system: a recommendation tool for web search queries called ARGUENET [10]. In this context, the intended user support system aims at providing an enriched web search engine which categorizes results, and where the user's needs correspond to strings to be searched on the web. The search engine is a conventional search engine (e.g., GOOGLE). Final recommendation results for a query q are prioritized according to domain background knowledge and the user's declared preferences. Figure 4 illustrates the architecture of an argument-based news recommender system.

Given a user query q, it will be given as an input to a traditional content-based web search engine, returning a list of search results L. If required, the original query q could be suitably re-formulated in order to improve the quality of the search results to be obtained. In the list L we can assume that s_i is a unique name characterizing a piece of information $info(s_i)$, in which a number of associated features (meta-tags, filename, URL, etc.) can be identified. We assume that such features can be identified and extracted from $info(s_i)$ by some specialized tool, as suggested by Hunter [22] in his approach to dealing with structured news reports. Such features will be encoded as a set $\mathcal{P}_{search}$ of new DeLP facts, extending thus the original program $\mathcal{P}$ into a new program $\mathcal{P}'$. A special operator $\mathbf{Revise}$ deals with possible inconsistencies found in $\mathcal{P}_{search}$ with

ALGORITHM Recommend_on_Query
INPUT: Query q, DeLP program $\mathcal{P} = \mathcal{P}_{user} \cup \mathcal{P}_{pool} \cup \mathcal{P}_{domain}$
OUTPUT: List L_{new} {*recommendation results wrt* $\mathcal{P}'$}
Let $L = [s_1, s_2, \dots s_k]$ be the output of solving q
wrt content-based search engine SE
{L *is the list of (the first* k*) results obtained from query* q *via* SE }
$\mathcal{P}_{search}$ = {facts encoding $info(s_1), info(s_2) \dots info(s_k)$}
{$info(s_i)$ *stands for features associated with result* s_i }
$\mathcal{P}'$:= **Revise** ($\mathcal{P} \cup \mathcal{P}_{search}$).
{**Revise** *stands for a belief revision operator to ensure consistency in* $\mathcal{P}'$ }
Initialize S^w, S^u, and S^d as empty sets.
{S^w*,* S^u*, and* S^d *stand for the set of results* s_i*'s which are warranted as relevant, undecided and warranted as non-relevant, respectively* }
FOR EVERY $s_i \in L$
DO
Solve query $rel(s_i)$ using DeLP program $\mathcal{P}'$
IF $rel(s_i)$ is warranted **THEN** add s_i to S^w
ELSE
IF $\sim rel(s_i)$ is warranted **THEN** add s_i to S^d
ELSE add s_i to S^d
Return Recommendation $L_{new} = [s^w_1, s^w_2, \dots, s^w_{j1}, s^u_1, s^u_2, \dots, s^u_{j2}, s^d_1, \dots, s^d_{j3}]$

Figure 7. Algorithm for solving queries ARGUENET

respect to $\mathcal{P}'$, ensuring $\mathcal{P} \cup \mathcal{P}_{search}$ is not contradictory.[11] Following the algorithm shown in Figure 7 we can now analyze L in the context of a new DeLP program $\mathcal{P}'$=$\mathcal{P} \cup Facts$, where $Facts$ denotes the set corresponding to the collection discussed above and $\mathcal{P}$ corresponds to domain knowledge and the user's preferences about the search domain.[12] For each s_i, the query $rel(s_i)$ will be analyzed in light of the new program $\mathcal{P}'$. Elements in the original list L of content-based search results will be classified into three sets of warranted, undecided, and defeated results. The final output presented to the user will be a sorted list L' in which the elements of L are ordered according to their epistemic status with respect to $\mathcal{P}'$. Figure 7 outlines a high level algorithm, which will be exemplified in the case study shown next.

Example 9 *Consider a journalist who wants to search for news articles about recent climate changes due to global warming. A query q containing the terms* climate, changes, global *and* warming *will return thousands of search results. Our journalist may have some implicit knowledge to guide the search, such as:*

1. *she always considers relevant the newspaper reports written by Bob Beak;*
2. *she usually considers relevant the reports written by trustworthy journalists;*
3. *Reports written by trustworthy journalists which are out of date are usually not relevant;*

[11] For example, contradictory facts may be found on the web. A simple belief revision criterion is to prefer the facts with a newer timestamp over the older ones.

[12] In this particular context, note that $\mathcal{P} = \mathcal{P}_{domain} \cup \mathcal{P}_{user}$.

$$
\begin{array}{rcl}
rel(X) & -\!\!\prec & author(X,A), trust(A). \\
\sim rel(X) & -\!\!\prec & author(X,A), trust(A), outdated(X). \\
trust(A) & -\!\!\prec & \mathsf{not}\ faked\_news(A). \\
\sim rel(X) & -\!\!\prec & address(X,Url), biased(Url). \\
biased(Url) & -\!\!\prec & industrialized\_country(Url). \\
\sim biased(Url) & -\!\!\prec & domain(Url,D), D = \text{“}env.org\text{”}. \\
rel(X) & \leftarrow & author(X, bob\_beak). \\
faked\_news(tim\_greenhouse) & \leftarrow & \\
oudated(X) & \leftarrow & date(X,D), getdate(Today), \\
 & & (Today - D) > 100. \\
 & & \\
industrialized\_country(X) & \leftarrow & \text{[Computed elsewhere]} \\
domain(Url,D) & \leftarrow & \text{[Computed elsewhere]} \\
getdate(T) & \leftarrow & \text{[Computed elsewhere]}
\end{array}
$$

Figure 8. DeLP program modeling preferences of a journalist

4. *Knowing that a journalist has not faked reports provides a tentative reason to believe he or she is trustworthy. By default, every journalist is assumed to be trustworthy.*
5. *Newspapers from industrialized countries usually offer a biased viewpoint on global warming;*
6. *The* “The Environmentalist” *(http://env.org) is a newspaper from an industrialized country which she usually considers non biased;*
7. *Tim Greenhouse is known to have faked a report.*

Such rules and facts can be modelled in terms of a DeLP program $\mathcal{P}$ as shown in Figure 8. Note that some rules in $\mathcal{P}$ rely on “built in” predicates computed elsewhere and not provided by the user.

For the sake of example, suppose that the above query returns a list of search results L=$[s_1,\ s_2,\ s_3,\ s_4]$. Most of these results will be associated with XML or HTML pages, containing a number of features (e.g. author, date, URL, etc.). Such features can be encoded as discussed before in a collection of DeLP facts as shown in Figure 9. We can now analyze s_1, s_2, s_3 and s_4 in the context of the user's preference theory about the search domain by considering the DeLP program $\mathcal{P}'$=$\mathcal{P}\cup Facts$, where $Facts$ denotes the set corresponding to the collection of facts in Figure 9. For each s_i, the query $rel(s_i)$ will be analyzed wrt this new program $\mathcal{P}'$.

Consider the case for s_1. The search for an argument for $rel(s_1)$ returns the argument $\langle \mathcal{A}_1, rel(s_1)\rangle$: s_1 should be considered relevant since it corresponds to a newspaper article written by Tim Greenhouse who is considered a trustworthy author (note

$author(s_1, tim\_greenhouse).$ $address(s_1, \text{“}env.org/...\text{”}).$ $date(s_1, 20070203).$
$author(s_2, jen\_doe).$ $address(s_2, \text{“}news.net/...\text{”}).$ $date(s_1, 20001003).$
$author(s_3, jane\_truth).$ $address(s_3, \text{“}env.org/...\text{”}).$ $date(s_3, 20070203).$
$author(s_4, bob\_beak).$ $address(s_4, \text{“}mynews.net/...\text{”}).$ $date(s_4, 20070203).$

Figure 9. Facts encoded from original web search results

that every journalist is considered to be trustworthy by default.) In this case we have the argument[13]

$$\mathcal{A}_1 = \{\ rel(s_1) \prec author(c_1, tim\_greenhouse), trust(tim\_greenhouse)\ ;\ trust(tim\_greenhouse) \prec \mathsf{not}\ \ faked\_news(tim\_greenhouse)\ \}.$$

Search for defeaters for argument $\langle \mathcal{A}_1, rel(s_1)\rangle$ will result in a defeater $\langle \mathcal{A}_2, \sim rel(s_1)\rangle$: s_1 is not relevant as it comes from a newspaper from an industrialized country, which is assumed to be biased about global warming. In this case we have the argument

$$\mathcal{A}_2 = \{\ \sim rel(c_1) \prec address(c_1, \text{“}env.org...\text{”}), biased\ (\text{“}env.org...\text{”})\ ;\ biased(\text{“}env.org...\text{”}) \prec industrialized\_country\ (\text{“}env.org...\text{”})\ \}.$$

Note that we also have an argument $\langle \mathcal{A}_3, \sim biased(\text{“}env.org...\text{”})\rangle$ which defeats $\langle \mathcal{A}_2, \sim rel(s_1)\rangle$: Usually articles from the “The Environmentalist” are not biased. In this case we have the argument

$$\mathcal{A}_3 = \{\ \sim biased(\text{“}env.org...\text{”}) \prec domain(\text{“}env.org...\text{”}, \text{“}env.org\text{”}),\ (\text{“}env.org\text{”} = \text{“}env.org\text{”})\ \}.$$

Finally, another defeater for the argument $\langle \mathcal{A}_1, rel(s_1)\rangle$ is found, namely the argument $\langle \mathcal{A}_4, faked\_news(tim\_greenhouse)\rangle$, with $\mathcal{A}_4 = \emptyset$. No other arguments need to be considered. The resulting dialectical tree rooted $\langle \mathcal{A}_1, rel(s_1)\rangle$ is shown in Figure 10a

[13] For the sake of clarity, semicolons separate elements in an argument $\mathcal{A} = \{e_1\ ;\ e_2\ ; \ldots ;\ e_k\ \}$.

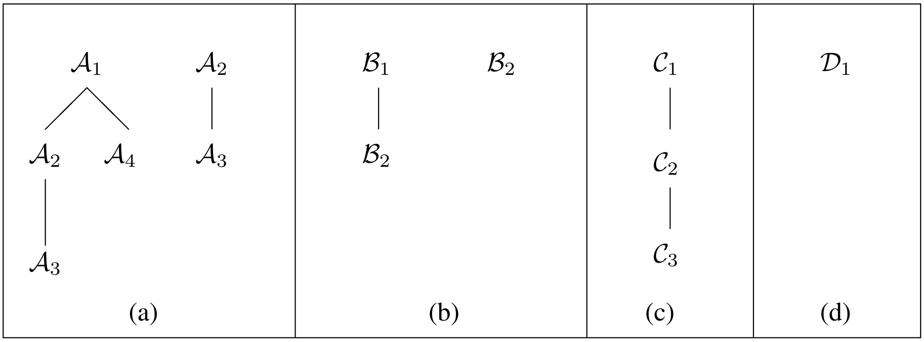


Figure 10. Dialectical trees associated with (a) $\langle \mathcal{A}_1, rel(s_1) \rangle$ and $\langle \mathcal{A}_2, \sim rel(s_1) \rangle$; (b) $\langle \mathcal{B}_1, rel(s_2) \rangle$ and $\langle \mathcal{B}_2, \sim rel(s_2) \rangle$; (c) $\langle \mathcal{C}_1, rel(s_3) \rangle$ and (d) $\langle \mathcal{D}_1, rel(s_4) \rangle$

(left). Not all paths have odd length, and hence $\langle \mathcal{A}_1, rel(s_1) \rangle$ is not warranted. Carrying out a similar analysis for $\sim rel(s_1)$ results in the dialectical tree shown in Figure 10a (right). A similar situation results. There are no other candidate arguments to consider; hence s_1 is deemed as undecided.

The case of s_2 is analogous. The argument $\langle \mathcal{B}_1, rel(s_2) \rangle$ can be built, with

$$\mathcal{B}_1 = \quad \{\, rel(s_2) \prec author(s_2,), trust(jen\_doe)\ ; \\ trust(jen\_oldie) \prec \text{not } faked\_\ news\ (jen\_doe)\ \}.$$

This argument is defeated by $\langle \mathcal{B}_2, \sim rel(s_2) \rangle$, with

$$\mathcal{B}_2 = \quad \{\sim rel(s_2) \prec author(s_2, jen\_doe), trust(jen\_doe), outdated(s_2)\ ; \\ trust(jen\_doe) \prec \text{not } faked\_news(jen\_doe)\}.$$

There are no more arguments to consider, and $\langle \mathcal{B}_1, rel(s_2) \rangle$ is deemed as non warranted ((Figure 10b (left)). The analysis of $\sim rel(s_2)$ results in a single argument. Thus, its associated dialectical tree has a single node $\langle \mathcal{B}_2, \sim rel(s_2) \rangle$, the only possible path has an odd length, and it is warranted.

Following the same line of reasoning used in the case of s_1 we can analyze the case of s_3. An argument $\langle \mathcal{C}_1, rel(s_3) \rangle$ can be built supporting the conclusion $rel(s_3)$ (a newspaper article written by Jane Truth is relevant as she can be assumed to be a trustworthy author). A defeater $\langle \mathcal{C}_2, \sim rel(s_3) \rangle$ will be found: s_1 is not relevant as it comes from an industrialized country's newspaper, which by default is assumed to be biased about bird flu. But this defeater in its turn is defeated by a third argument $\langle \mathcal{C}_3,\ biased(s_3) \rangle$. The resulting dialectical tree for $\langle \mathcal{C}_1, rel(s_3) \rangle$ is shown in Figure 10c (left)). The original argument $\langle \mathcal{C}_1, rel(s_3) \rangle$ can be thus deemed as warranted. *Finally let us consider the case of s_4. There is an argument $\langle \mathcal{D}_1, rel(s_4) \rangle$ with $\mathcal{D}_1 = \emptyset$, as $rel(s_4)$ follows directly from the strict knowledge in $\mathcal{P}$. Clearly, there is no defeater for an empty argument (as no defeasible knowledge is involved). Hence $rel(s_4)$ is* warranted *(see dialectical tree in Figure 10d).*

Applying the criterion given in the algorithm shown in Figure 7, the initial list of search results $[s_1,\ s_2,\ s_3,\ s_4]$ will be shown as $[s_3,\ s_4,\ s_1,\ s_2]$ (as $\langle \mathcal{C}_1, rel(s_3) \rangle$ and $\langle \mathcal{D}_1, rel(s_4) \rangle$ are warranted, $\langle \mathcal{A}_1, rel(s_3) \rangle$ is undecided and $\langle \mathcal{B}_2, \sim rel(s_2) \rangle$ is warranted (i.e., s_2 is warranted to be a non-relevant result).

6. Related work

Several kinds of recommender systems that operate on top of Internet services have been proposed over the past years (e.g., [2,27]). In the case of web-based recommender systems, the usual approach involves taking into account the user's interests –either declared by the user or conjectured by the system– to rank or filter web pages.

Some web recommender systems include LIRA [4], BASAR [43], ifWeb [3], Let's Browse [28], Margin Notes [40], and Quickstep [31]. Examples of hybrid news filtering system include NewsDude [7], GroupLens [38] and PHOAKS [42]. These recommender systems, however, differ from our proposal in that they do not attempt to perform a qualitative analysis to warrant recommendations.

Recently there have been other efforts oriented towards integrating argumentation in generic decision making systems. In [46], an argument-based approach to modelling group decision making is presented, in which argumentation is used to support group decision task generation and identification. In contrast with our approach, the argumentation process is not automated, and the authors use argumentation for agreement among multiple users in a team, whereas we focus on argumentation for eliciting conclusions for a particular user on the basis of available information. In [45] a number of interesting *argument assistance tools* are presented. Even though there is a sound logical framework underlying this approach, the focus is rather restricted to legal reasoning, viewing the application of law as dialectical theory construction and evaluating alternative ways of representing argumentative data. In contrast, our analysis is oriented towards characterizing more generic argument-based user support systems.

Finally, it must be remarked that there has been a growing attention to the development of a Semantic web [6], and a number of semantically enhanced recommendation techniques have been proposed (e.g., [47]). Although the vision of the Semantic Web is still at its beginning, the use of defeasible reasoning for qualitative analysis can also be naturally integrated into such approaches, as recently shown in [5].

7. Conclusions. Future work

In this chapter we have introduced a novel approach towards the development of user support systems by enhancing recommendation technologies through the use of qualitative, argument-based analysis. In particular, we have shown that DeLP is a suitable computational tool for carrying on such analysis in a real-world application for intelligent web search, providing thus a tool for higher abstraction when dealing with users' information needs. Preliminary experiments on the use of ARGUENET were performed on the basis of a prototype. However, it must be remarked that these initial experiments only serve as a "proof of concept" prototype, as thorough evaluations are still being carried out. As performing defeasible argumentation is a computationally complex task, an abstract machine called JAM (Justification Abstract Machine) has been specially developed for an efficient implementation of DeLP [18], allowing to solve queries and computing dialectical trees very efficiently. The JAM provides an argument-based extension of the traditional WAM (Warren's Abstract Machine) for PROLOG. A full-fledged implementation of DeLP is available online,[14] including facilities for visualizing arguments and

[14]See `http://lidia.cs.uns.edu.ar/delp_client`

dialectical trees. Several other features leading to efficient DeLP implementations have also been recently studied, in particular those related to comparing conflicting arguments by *specificity* as a syntax-based preference criterion, pruning dialectical trees to speed up the argumentative inference procedure and extending DeLP to incorporate possibilistic reasoning [13]. Equivalence results with other extensions of logic programming have also been established.

One important web search results, encoding them as part of a DeLP program. Although HTML tags associated with web documents are not intended to convey a formal semantics, these tags can be usefully exploited to extract meaningful content [16,25]. The emergence of XML as a standard for data representation on the web contributes to further simplify the above problem. In this context, the approach proposed by Hunter [22] to represent semi-structured text through logical formulas is particularly relevant for enhancing the capabilities of argument-based user support systems as presented in this article.

Current trends in user support system technologies show clearly that the combination of quantitative and qualitative analysis of user preferences will play a major role in the future. In this context, we think that defeasible argumentation techniques will constitute a powerful tool to make inference in recommender systems more reliable and user-friendly. The approach presented in this chapter intends to be a first step to reach this long-term goal.

Acknowledgements

This research work was supported by Projects TIN 2004-07933-C03-03 and TIN2006-15662-C02-02 (Spain), Projects 24/N016 and 24/ZN10 (Secretaría de Ciencia y Tecnología, Universidad Nacional del Sur, Argentina), by CONICET (Argentina) and by Agencia Nacional de Promoción Científica y Tecnológica, (Projects PICT 2002 No. 13.096, PICT 2003 No. 15043, PICT 2005 No. 32373 and PAV 076).

References

[1] L. Amgoud, N. Maudet, and S. Parsons. An argumentation-based semantics for agent communication languages. In *Proc. of the 15th. European Conference in Artificial (ECAI), Lyon, France*, 2002.
[2] R. Armstrong, D. Freitag, T. Joachims, and T. Mitchell. Webwatcher: A learning apprentice for the world wide web. In *AAAI Spring Symp. on Information Gathering*, pages 6–12, 1995.
[3] F. Asnicar and C. Tasso. ifWeb: a prototype of user models based intelligent agent for document filtering and navigation in the World Wide Web. In *Sixth International Conference on User Modeling*, Chia Laguna, Sardinia, Italy, June 1997.
[4] M. Balabanovic, Y. Shoham, and Y. Yun. An adaptive agent for automated web browsing. *Journal of Visual Communication and Image Representation*, 6(4), 1995.
[5] N. Bassiliades, G. Antoniou, and I. Vlahavas. A defeasible logic reasoner for the semantic web. In *Proc. of the Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 49–64, 2004.
[6] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, May 2001.
[7] D. Billsus and M. J. Pazzani. A hybrid user model for news classification. In *In Kay J. (ed.), UM99 User Modeling - Proceedings of the Seventh International Conference*, pages 99–108. Springer-Verlag, 1999.
[8] J. Budzik, K. Hammond, and L. Birnbaum. Information access in context. *Knowledge based systems*, 14(1–2):37–53, 2001.

[9] C. Chesñevar, A. Maguitman, and R. Loui. Logical Models of Argument. *ACM Computing Surveys*, 32(4):337–383, December 2000.

[10] C. Chesñevar, A. Maguitman, and G. Simari. Argument-Based Critics and Recommenders: A Qualitative Perspective on User Support Systems. *Journal of Data and Knowledge Engineering*, 59(2):293–319, 2006.

[11] C. Chesñevar, A. Maguitman, and G. Simari. *Artificial Intelligence Applications and Innovations*, chapter Argument-based User Support Systems using Defeasible Logic Programming, pages 61–69. Springer Verlag (IFIP Series), 2006.

[12] C. Chesñevar, G. Simari, and L. Godo. Computing dialectical trees efficiently in possibilistic defeasible logic programming. *LNAI/LNCS Springer Series (Proc. of the 8th Intl. Conference on Logic Programming and Nonmonotonic Reasoning LPNMR 2005)*, pages 158–171, September 2005.

[13] C. Chesñevar, G. Simari, L. Godo, and T. Alsinet. Argument-based expansion operators in possibilistic defeasible logic programming: Characterization and logical properties. *Lecture Notes in Artificial Intelligence (LNAI) 3571 (Proc. of the 8th European Conference on Symbolic and Qualitative Aspects of Reasoning with Uncertainty – ECSQARU 2005, Barcelona, Spain)*, pages 353–365, July 2005.

[14] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin. Combining content-based and collaborative filters in an online newspaper. In *Proceedings of ACM SIGIR Workshop on Recommender Systems*, August 1999.

[15] Mukund Deshpande and George Karypis. Item-based top-n recommendation algorithms. *ACM Trans. Inf. Syst.*, 22(1):143–177, 2004.

[16] R. B. Doorenbos, O. Etzioni, and D. Weld. A scalable comparison-shopping agent for the world-wide web. In W. Lewis Johnson and Barbara Hayes-Roth, editors, *Procs. of the First Intl. Conf. on Autonomous Agents (Agents'97)*, pages 39–48, Marina del Rey, CA, USA, 1997. ACM Press.

[17] Phan M. Dung. On the Acceptability of Arguments and its Fundamental Role in Nomonotonic Reasoning and Logic Programming. In *Proc. of the 13th. International Joint Conference in Artificial Intelligence (IJCAI), Chambéry, Francia*, pages 321–357, 1993.

[18] A. García and G. Simari. Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, 4(1):95–138, 2004.

[19] Michael Gelfond and Vladimir Lifschitz. Logic programs with classical negation. In *Proceedings of the 7th International Conference on Logic Programming. Jerusalem*, June 1990.

[20] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.

[21] G. Governatori and A. Stranieri. Towards the application of association rules for defeasible rules discovery. In *Legal Know. & Inf. Sys.*, pages 63–75. JURIX, IOS Press, 2001.

[22] A. Hunter. Hybrid argumentation systems for structured news reports. *Knowledge Engineering Review*, pages 295–329, 2001.

[23] J. Konstan. Introduction to recommender systems: Algorithms and evaluation. *ACM Trans. Inf. Syst.*, 22(1):1–4, 2004.

[24] Robert Kowalski and Francesca Toni. Abstract Argumentation. *Artificial Intelligence and Law*, 4(3-4):275–296, 1996.

[25] N. Kushmerick, D. S. Weld, and R. B. Doorenbos. Wrapper induction for information extraction. In *IJCAI'97*, pages 729–737, 1997.

[26] D. Leake, T. Bauer, A. Maguitman, and D. Wilson. Capture, storage and reuse of lessons about information resources: Supporting task-based information search. In *Proceedings of the AAAI-00 Workshop on Intelligent Lessons Learned Systems. Austin, Texas*, pages 33–37. AAAI Press, 2000.

[27] H. Lieberman. Letizia: An agent that assists web browsing. In *IJCAI'95*, pages 924–929. Morgan Kaufmann, 1995.

[28] H. Lieberman, N. Van Dyke, and A. Vivacqua. Let's browse: a collaborative Web browsing agent. In *Proceedings of the 1999 International Conference on Intelligent User Interfaces (IUI'99)*, pages 65–68, Los Angeles, CA, USA, 1999. ACM Press.

[29] F. Linton, D. Joy, and H. Schaefer. Building user and expert models by long-term observation of application usage. In *Proceedings of the seventh international conference on User modeling*, pages 129–138. Springer-Verlag New York, Inc., 1999.

[30] J. McCarthy and P. Hayes. Some Philosophical Problems from the Standpoint of Artificial Intelligence. In B. Meltzer and D. Mitchie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, 1969.

[31] S. Middleton, D. DeRoure, and N. Shadbolt. Capturing knowledge of user preferences: Ontologies in recommender systems. In *Proc. ACM K-CAP'01*, Canada, 2001. ACM Press.

[32] D. Nute. Defeasible Reasoning. In James H. Fetzer, editor, *Aspects of Artificial Intelligence*, pages 251–288. Kluwer Academic Publishers, Norwell, MA, 1988.

[33] M. J. Pazzani, J. Muramatsu, and D. Billsus. Syskill & webert: Identifying interesting web sites. In *AAAI/IAAI'96*, pages 54–61, 1996.

[34] J. Pollock. *Knowledge and Justification*. Princeton, 1974.

[35] D. Poole. On the Comparison of Theories: Preferring the Most Specific Explanation. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 144–147. IJCAI, 1985.

[36] H. Prakken and G. Vreeswijk. Logical Systems for Defeasible Argumentation. In D. Gabbay and F.Guenther, editors, *Handbook of Phil. Logic*, pages 219–318. Kluwer, 2002.

[37] Henry Prakken and Giovani Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7:25–75, 1997.

[38] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*, pages 175–186, Chapel Hill, North Carolina, 1994. ACM.

[39] B. Rhodes and P. Maes. Just-in-time information retrieval agents. *IBM Systems Journal special issue on the MIT Media Laboratory*, 39(3-4):685–704, 2000.

[40] B. J. Rhodes. Margin notes: Building a contextually aware associative memory. In *The Proceedings of the International Conference on Intelligent User Interfaces (IUI '00)*, January 2000.

[41] G. Simari and R. Loui. A Mathematical Treatment of Defeasible Reasoning and its Implementation. *Artificial Intelligence*, 53:125–157, 1992.

[42] L. Terveen, W. Hill, B. Amento, D. McDonald, and J. Creter. PHOAKS: a system for sharing recommendations. *Communications of the ACM*, 40(3):59–62, 1997.

[43] Christoph G. Thomas and Gerhard Fischer. Using agents to personalize the Web. In *Proceedings of the 2nd international conference on Intelligent user interfaces*, pages 53–60. ACM Press, 1997.

[44] N. Van Dyke, H. Lieberman, and P. Maes. Butterfly: a conversation-finding agent for internet relay chat. In *Proceedings of the 4th international conference on Intelligent user interfaces*, pages 39–41. ACM Press, 1999.

[45] B. Verheij. Artificial argument assistants for defeasible argumentation. *Artif. Intell.*, 150(1-2):291–324, 2003.

[46] P. Zhang, J. Sun, and H. Chen. Frame-based argumentation for group decision task generation and identification. *Decision Support Systems*, 39:643–659, 2005.

[47] C. Ziegler. Semantic web recommender systems. In *Proceedings of the Joint ICDE/EDBT Ph.D. Workshop*, 2004.

Knowledge Modelling Using UML Profile for Knowledge-Based Systems Development

Mohd Syazwan Abdullah[1a], Richard Paige[2] , Ian Benest [2] and Chris Kimble[2]
[1] *Faculty of Information Technology,*
syazwan@uum.edu.my
Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia
[2] *Department of Computer Science,*
University of York, Heslington, York YO10 5DD, UK
{paige, idb, kimble}@cs.york.ac.uk

Abstract. Knowledge modelling techniques are widely adopted for designing knowledge-based systems (KBS) used for managing knowledge. This chapter discusses conceptual modelling of KBS in the context of model-driven engineering using standardised conceptual modeling language. An extension to the Unified Modeling Language (UML) for knowledge modelling is presented based on the profiling extension mechanism of UML. The UML profile discussed in this chapter has been successfully captured in a Meta-Object-Facility (MOF) based UML tool – the eXecutable Modelling Framework (XMF) Mosaic. The Ulcer Clinical Practical Guidelines (CPG) Recommendations case study demonstrates the use of the profile, with the prototype system implemented in the Java Expert System Shell (JESS).

1. Introduction

The use and management of knowledge in enterprises has become a commercial necessity for many enterprises, in order that they manage their corporate intellectual assets and gain competitive advantage [1]. Most knowledge resides in human memories and managing it is often seen as a human-oriented process rather than a technology-based solution. Nevertheless, technology can be utilised as a knowledge management (KM) enabler with automated tools, including the , knowledge-based systems (KBS), internet and groupware systems [2]. Knowledge-based systems (KBS) were developed for managing codified knowledge in the field of Artificial Intelligence (AI). Widely known as expert systems, these were originally created to emulate the human expert reasoning process [3] and is one of the successful inventions that has been derived from AI technologies. KBS are developed using knowledge engineering (KE) techniques, which are similar to those used in software engineering (SE), but have an emphasis on knowledge rather than on data or information processing [4].

[a] Corresponding Author

Central is the conceptual modelling of the system during the analysis and design stages of the development process; this is widely known as knowledge modelling. Many knowledge engineering methodologies have been developed with an emphasis on the use of models, for example: CommonKADS [5]. KBS continue to evolve as the need to have a stable technology for managing knowledge grows. Their current role as an enabler for knowledge management (KM) initiatives has led to greater appreciation of this technology [6,7]. It has matured from a non-scalable technology to one that can be adopted for managing the knowledge used in demanding commercial applications; it is a tool that is widely accepted by industry [8,9]. Because it is a maturing technology, the Object Management Group (OMG), which governs object-oriented software modelling standards, has started a standardisation process [10] for knowledge-based engineering services and production rule representation (PRR).

This chapter is organized as follows: Section 2 describes and discusses knowledge modelling issues in designing KBS. Section 3 explains the UML extension mechanism. Section 4 discusses in detail the knowledge modeling profile, while section 5 describes a case study that illustrates how the profile can be used to develop a KBS. Finally, section 6 concludes and indicates future directions for the work.

2. Knowledge Modelling

Traditional KE techniques were widely used to construct ES – systems built from the knowledge of one or more experts – essentially, a process of knowledge transfer [4]. This is the development process of the first generation of ES, in which the knowledge of the expert is directly transferred into the knowledge base in the form of rules. The disadvantage of this approach is that the knowledge of the expert is captured in the form of hard coded rules within the system with little understanding of how these rules are linked or connected with each other [5]. KE is no longer simply a means of mining the knowledge from the expert's head, it now encompasses *"methods and techniques for knowledge acquisition, modelling, representation and use of knowledge"* [5] and KBS development is viewed as a modeling activity [4] in the analysis and design stages of the systems development.

The foundation for the modelling process is based on the *knowledge-level* principle, popularised by [11] for KE purposes, requires that knowledge be modelled at a conceptual level independent of the implementation formalism. Knowledge modelling is similar to that of conceptual modelling, which is widely used to refer to implementation-independent models in SE; both the terms are used inter-changeably in the KE domain. The *knowledge-level* principle is fundamental to the process of conceptualisation for problem solving [12] and is used in KE for the explicit representation of the real world problem that is to be solved by the proposed system [13].

While knowledge about the domain is usually addressed through the use of ontologies, the independent reasoning process is specified with Problem Solving Methods (PSM) [7]. Both ontologies and PSM provide components that are reusable across domains and tasks [14] enabling KBS to be designed, built and deployed quickly. Ontologies are formal declarative representations of the domain knowledge; that is, they are sets of objects with describable relationships [15]. Thus an ontology used for knowledge modelling defines the content-specific knowledge representation elements such as domain-dependent classes, relations, functions and object constants [16]. PSM however describe the reasoning-process (generic inference patterns) at an abstract level, which is independent of the representation formalism (e.g. rules, frames, etc.) [14]. PSM have influenced the leading KE frameworks such as Task Structures, Rôle-Limiting Methods, CommonKADS, Protégé, MIKE, VITAL and others [14]. PSM can be considered to be design patterns in KE for KBS development [5].

It is commonly agreed by researchers [17] that conceptual modelling is an important stage in any software system construction. However, both SE and KE communities have developed different modelling techniques that are now almost unrelated [18] as a result of the fundamental computational difference between them in solving the same problem [19]. As a result, although both field's ultimate goal is to build software systems, the different experiences are difficult to interchange [19]. Nevertheless, most KE modelling notations are derived from the SE field as these are better established.

Research has shown that the successful adoption of KBS technology does not depend on technical or economic reasons. The main culprits are mainly related to organisational and managerial issues as reported by Gill [20, 21]. According to Gill, KBS project are specialised in nature that requires the team members to have knowledge of both the problem domain and the development tools. As a result, the team members are skilful individuals and there is a great threat to the overall project if they leave the team in the course of the project or during the maintenance period of the system. KBS that are well-designed through the use of appropriate, well-understood standard modelling language and representation technique based on sound KBS development methodologies could still be more easily comprehended by new team members even in the event of human turnover. The major setback with knowledge modelling is that there is no standard language available [22] to model the knowledge for developing a KBS. As the SE community has adopted UML as the *defacto* standard for modelling OO system and the KE community could do the same. After all, KBS are integrated into other enterprise system [23] and system designs based on a standardised modelling language would help facilitate communication and sharing of blue prints among developers.

KBS development processes are usually associated with structured techniques for modelling and coding the system. Nevertheless, OO features such as abstraction, inheritance, aggregation, polymorphism, modularity and encapsulation can be utilised in developing KBS as shown in [24]. This is possible as knowledge can be viewed as 'information about information' [5]. These features are supported by UML and popular object oriented programming languages such as JAVA [25], which can be exploited in developing these knowledge intensive applications. Even OO-based inference engines have been developed to harness the OO features such as Object Prolog [26], and Java Expert System Shell (JESS) [27].

Another important factor to consider is that most system analysis and design courses these days are teaching UML as a tool for systems modelling and development [25, 28]. The main influence is the growing importance of OO programming languages like Java in systems development. Researchers such as [29], have proved that the extent to which diagrammatic notations to use are usually based on personal preference, convention or technical ease of use. Due to the formal training received and the adoption of OO programming by this generation of system analyst, most will have knowledge of UML and will be able to use them for modelling purposes.

In addition to this, modern enterprise systems are an integration of various systems built on different platforms with the ability to communicate with each other [23]. Knowledge-based systems are no longer stand-alone systems, but are part of the enterprise group of systems [18, 23, 30] as it attracts higher rates of usage if its' embedded with conventional systems [24]. KBS have been integrated with other systems such as Computer Aided Design (CAD) systems for managing engineering product design knowledge, intelligent SCADA alarm interpretation and Geographical Information System (GIS) as intelligent advisor. Having a standardised modelling language promotes the use of a common modelling language, so that the vision of integration, reusability and interoperability within an enterprise's system will be achieved [23,30]. Therefore, we proposed to model design knowledge about KBS using an extension to UML.

3. UML Extensibility Mechanism

The OMG's Model Driven Architecture (MDA) – a model-driven engineering framework – provides integration with, and interoperability between, different models developed using its standards [31] (such as UML, Meta-Object Facility (MOF), and others). The growth of MDA will fuel the demand for more meta-models to cater for domain specific modelling requirements [31,32]. Profiles have defined semantics and syntax, which enables them to be formally integrated into UML, though of course they must adhere to the profile requirements proposed by OMG. Previous profile development for knowledge modelling has concentrated only on certain task types such as product design and product configuration [33]. In contrast, the work described here emphasises the development of a generic profile. Developing a meta-model for knowledge modelling will enable it to be integrated into the MDA space allowing the relation between the knowledge models and other language models to be understood. It provides for seamless integration of different models in different applications within an enterprise.

UML is a general-purpose modelling language [31] that may be used in a wide range of application domains. It can be extended to model domains that it does not currently support, by extending the modelling features of the language in a controlled and systematic fashion. The OMG [34, 35] defines two mechanisms for extending UML: profiles and meta-model extensions. Both extensions have (unfortunately) been called profiles [31]. The "lightweight" extension mechanism of UML [35] is profiles. It contains a pre-defined set of Stereotypes, TaggedValues, Constraints, and notation icons that collectively specialize and tailor the existing UML meta-model. The main construct in the profile is the stereotype that is purely an extension mechanism. In the model, it is marked as «stereotype» and has the same structure (attributes, associations, operations) as that defined by the meta-model. However, the usage of stereotypes is restricted; changes in the semantics, structure, and the introduction of new concepts to the meta-model are not permitted [36]. The "heavyweight" extension mechanism for UML (known as the meta-model extension) is defined through the MOF specification [24] which involves the process of defining a new meta-model [36]. This approach should be favoured if the semantic gap between the core modelling elements of UML and the newly defined modelling elements is significant [31].

4. Knowledge Modelling Profile

The work presented in this chapter adopts the XMF Mosaic approach [33] in designing the knowledge modelling profile as the OMG only specifies how profiles should be constituted and not how to design them. By adopting the XMF Mosaic approach, the profile development is structured into well-defined stages that are easy to follow and methodologically sound. The XMF Mosaic is a newly developed object-oriented meta-modelling language, and is an extension to existing standards defined by OMG. The XMF Mosaic approach to creating a profile can be divided into three steps: the derivation of an abstract syntax model of the profile concepts, a description of the profile's semantics, and the presentation of the profile's concrete syntax (not discussed here) if this is different from UML diagrams. Details of the XMF Mosaic approach and the profile development stages can be found in [33]. XMF Mosaic was adopted in the original design [34], but since XMF is not MOF compliant, UML tools were not able to support the resulting profile.

Profiles are sometimes referred to as the "lightweight" extension mechanism of UML [35]. A profile contains a predefined set of Stereotypes, TaggedValues, Constraints, and notation icons that collectively specialize and tailor the UML to a specific domain or process. The main construct in the profile is the stereotype that is purely an extension mechanism. In the model, it is marked as <<stereotype>> and has the same structure (attributes, associations, operations) as defined by the meta-model that is used for its description. Nevertheless, the usage of stereotypes is restricted, as changes in the semantics, structure, and the introduction of new concepts to the meta-model are not permitted [36]. In the case of knowledge modelling, the existing constructs of UML are sufficient in representing the KBS concepts.

The Profile Concept

The concepts that underpin the profile are those taken from the existing BNF definition of the CommonKADS Conceptual Modelling Language (CML) [5], providing a well-defined and well-established set of domain concepts. Most of these elements are generally those adopted in the KBS literature and are widely used for representing the concepts of KBS in the KE domain. These knowledge modelling concepts are itemised in Table 1 and the abstract syntax model of the profile is shown in Figure 1.

Table 1. Main Knowledge Modelling Concepts

Modelling Concept	Description
Concept (class)	Class that represents the category of things
FactBase	Collection of information/fact that will be matched against the rule
Inference	The lowest level of functional decomposition consisting of primitive reasoning steps
Transfer Function	Transfers information between the reasoning agent and external entities (system, user)
Task	Defines the reasoning function
Task Method	Describes the realization of the task through subfunction decomposition
Static Role	Specifies the collection of domain knowledge that is used to make the inference
Dynamic Role	Run-time inputs and outputs of inferences
Rule	Expressions that involve an attribute value of a concept
Decision Table	Rules that are formalized into a table-like representation
Production Rule	Rules that consist of the premises and the action of the rule
Knowledge Base	Collection of data stores that contains instances of domain knowledge types

Model Extension

In the profile, all the modeling concepts are stereotyped and extended from the UML meta-class Class as these elements are not defined in the standard UML. A Concept class is used to represent structural things and these have attributes contained in them; it is similar to class in the UML meta-model. When the attributes are used in rules they are known as knowledge elements. A Concept is linked to the Rule class in the model. Concepts are diagrammatically associated with FactBase; as the values of the attributes are stored here and are extracted during the reasoning process of the inference. The instances of each attribute, contained in the FactBase class, are accessed by the dynamic role, which passes them to the inference process that matches the premise with the consequent part of an implication rule.

Task class defines the reasoning function and specifies the overall input and output of the task. Each task will have an associated task method that executes the task. The structure of the task, its task method, and the set of associated inference processes can be defined with the knowledge model from the problem-solving method library. The task-type, knowledge model, will help in identifying the inference structure needed to perform the desired task. Task method can be decomposed into sub-tasks for certain task-types. Task method class will specify the type of inference that is to be performed. The control structure of the method captures the inference reasoning strategy, which is described using an activity diagram. If the inference process requires additional input, either from the user or from an external entity, the task method will invoke a transfer function. Such functions are used to transfer additional information between the reasoning processes.

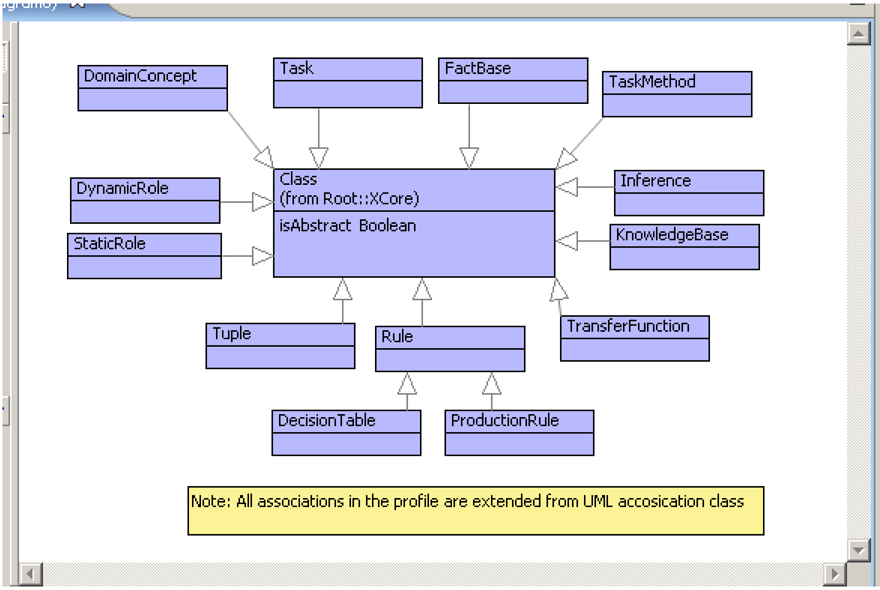


Fig.1. Extension of the UML with stereotypes for the Profile defined in XMF Mosaic tool

The Dynamic Role class specifies the ‘information’ flow of attribute instances from the concepts. It also specifies the outputs that arise from executing the inference sets. The output of this inference process is the ‘result’ of matching the antecedent of the rule with the consequent part. Depending on what the KBS is reasoning about, if it is not the final output of the system, then the output can be used in another inference. The Static Role class is the function responsible for fetching the collection of domain knowledge (rules) from the knowledge base prior to an active inference. Inferences do not access the knowledge base directly, but request the necessary rules related to the particular inference from the static roles. In some KBS shells this is similar to posting the rules to the inference process or similar to setting which rule should be fired. This allows the inference process to handle a specific reasoning task and invoke those rules that are appropriate.

An Inference class executes a set of algorithms for determining the order in which a series of non-procedural, declarative statements are to be executed. The inference process infers new knowledge from information/facts that are already known. The Task Method invokes this. The input (information/fact) used by this process is provided by the dynamic role. The result of the inference process is then passed to the dynamic role. The knowledge element used in the inference is accessed through the Static Role, which fetches the group of rules from the knowledge base. There are several different inference processes for a given task, most of which are run in the background by the inference engine. The knowledge base class contains domain knowledge, represented as rules, which are used by the inference process. The contents of the knowledge base are organized in tuples (records). A tuple is used to group rules according to their features. This allows the partitioning of the knowledge base into modules that enables the inference process to access the rules faster. The maintainability of the rules is enhanced when it is organised in this manner.

The Rule class of the profile describes the modelling of rules within the domain concept. Rule class is used to represent knowledge elements in KBS and is viewed as ‘information about information’. Rule class allows for rules to be in different formats. There are two types of rule: production rule implication rule, and decision table. A production rule is of the form: ‘if-then’ premise followed by an action. This type of representation is widely used in KBS; they are known as implication rules. A decision table is an addition to the rule class. It is introduced here because certain rules are best expressed in the form of a decision table, even though they are usually converted to flattened production rules.

5. Case Study – Clinical Practice Guideline Recommendations

The CPG recommendation was implemented as a KBS application for educational purposes to list the recommendations based on evidence strength using the following classification (a) evidence strength only; (b) evidence strength and category; (c) category only; and (d) factors, evidence and category. The rules for the KBS was defined based on these classifications (in the actual recommendation, each recommendation has a brief explanation rather than ID as I1, II2, III4, etc which are much more convenient for discussions.).

The KBS domain concept ‘CPG’ is composed of the five category of recommendations which are represented as domain concept ‘CPGManagement’, ‘CPGCleansing’, ‘CPGQualityAssurance’, ‘CPGAssessment’ and ‘CPGEducation’ shown at the top section of figure 2. Each of the domain concepts has three attributes (name, factors and evidence strength) upon which four types of rules for the system were defined based on their values. The instances of these attribute are stored in the fact base of the system which are accessed by dynamic role to get the facts for the inference reasoning process. The inference executes the reasoning task based on the task method specification which only specifies a single inference execution for the CPG system. The production rules of the system are stored in the knowledge base which are organised into tuples.

KBS design is very much different to that of a conventional system, as the overall aim of the KBS is to gather the needed facts to fire the rules. In doing so, completing the whole reasoning cycle involves activation of different processes and message passing between objects. As a result, it is difficult to capture these vital information using object diagram due to the fact that several snapshots are needed to gather the whole picture. However, this limitation was solved with the aid of another type of UML diagram, namely the sequence diagram. Using sequence diagrams, the processing elements of the KBS gathered from the profile are listed as objects with an additional Interface object to model the flow of logic that captures the dynamic behaviour of the KBS as shown on figure 3. The input from the user is entered through the interface which becomes the fact for the system when the recommendation type selection question has been answered. These facts are gathered by dynamic role and the inference engine gets these facts and matches them with the rule gathered from the knowledge base to provide the recommendation.

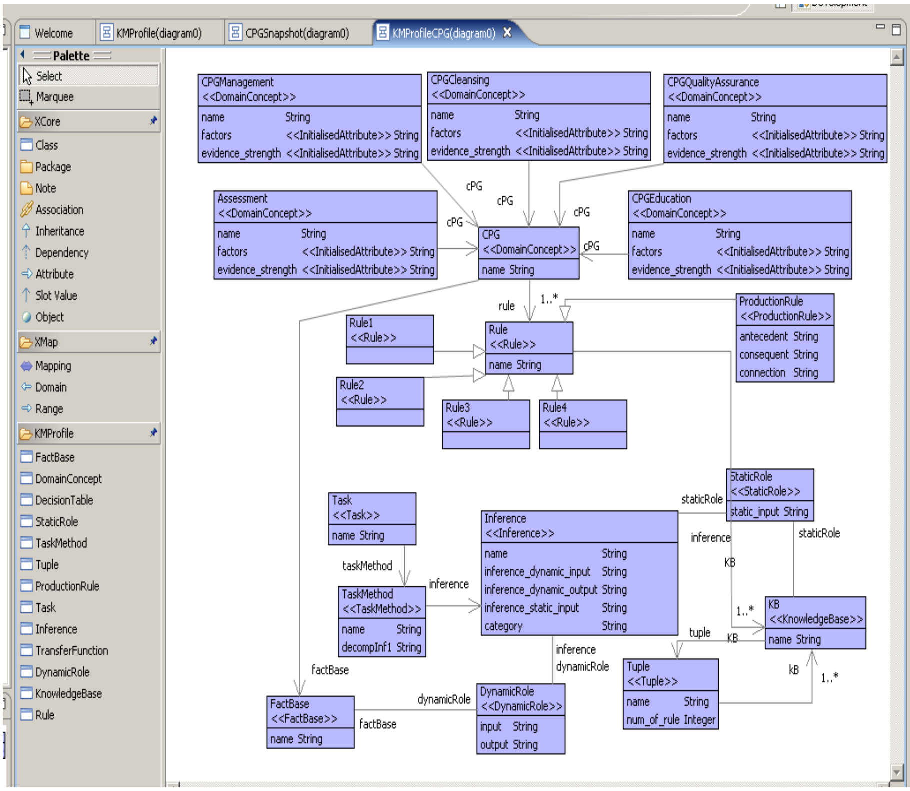


Fig. 2. CPG knowledge model

The CPG prototype recommendation system was implemented using Java Expert System Shell (Jess) [27] rule engine, which is a popular variation of the CLIPS rule engine developed in Java. Jess was chosen as the implementation platform as it is the reference implementation of the JSR 94 Java Rule Engine API that defines standard API for Java developer to interact with a Java rule engine widely used in commercial products and open source software projects.

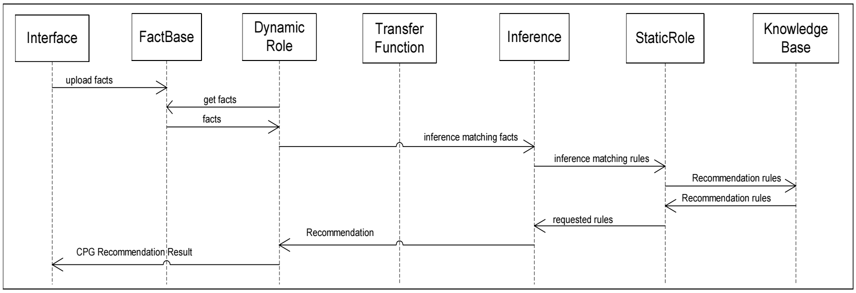


Fig. 3. Sequence Diagram of CPG system

The system receives the user input value for the strength, category and factor which are the facts for the system to fire the rules through the interview module based on the questions from the question module and the ask module performing error checking on the answers. In the recommendation module, the CPG rules are defined (evidence strength only; category only; evidence strength and category; and factors, evidence and category) and these rules are matched against the facts to fire the activated recommendation rule. The report module produces the recommendation report of the system which contains the explanation and the recommendation value. Table 2 presents the Jess program summary for CPG system and the sample screenshot is shown in figure 4.

Table 2. Jess Program Summary for CPG System

```
;; Module MAIN
(deftemplate CPG)
(deftemplate S-C-F)
(deftemplate question)(deftemplate answer)
(deftemplate recommendation)
;;Module Question
(deffacts question-data)
(defglobal ?*crlf* = "")
;; Module ask
(defmodule ask)
(deffunction ask-user (?question ?type))
(defmodule startup)
;; Module interview
(defmodule interview)
(defrule request-strength => assert ask strength)))
(defrule assert-user-fact
  (answer (ident strength)     (text ?i))
  (answer (ident cate_gory)    (text ?d))
  (answer (ident factors_type) (text ?j))
  =>  (assert (user (strength ?i) (cate_gory ?d)
(factors_type ?j))))
;; Module recommend
```

```
(defmodule recommend)
( defrule S-C-F-1-0-0
 user (strength ?i&:(= ?i 1))(cate_gory ?d&:
( = ?d 0))factors_type ?j&:(= ?j 0))) => assert
recommendation (S-C-F STR1) (explanation "Strength
equals 1 Recommendation ( I1 , I2 , I3 , I4 )") )))

;; Module report
(defmodule report)
(deffunction run-system ()
  (reset)(focus startup interview recommend report)
  (run))
(while TRUE (run-system))
```

Table 3 lists the possible mapping of the profile elements to the Jess. The domain concept elements of the profile can be mapped to `deftemplate`, `defclass` or `definstance` of Jess. However, for the CPG system, only `deftemplate` was used to represent the CPG domain concept which has three different slots for strength, factor type and category. The factbase element of the profile can be mapped to `deffacts` and for the CPG system; the question-data were used to gather the needed facts for the application.

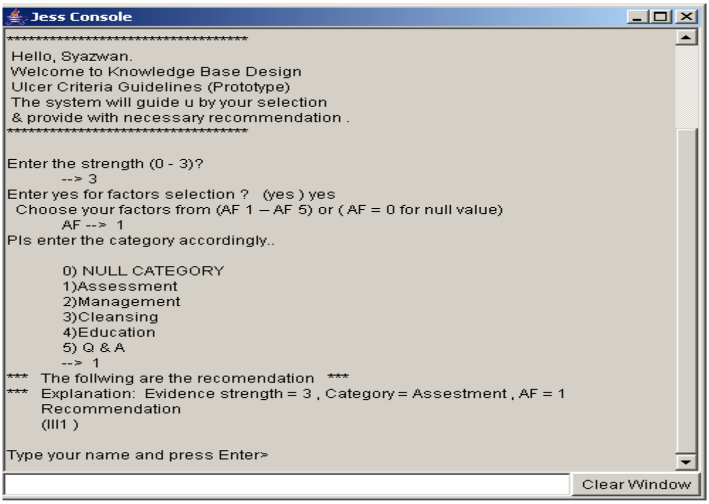


Fig. 4. Sample screenshot of the CPG system

There are no direct mapping for task and task method to Jess but `defmodule` can be used to divide the application into structured modules. To perform the reasoning process, inference is activated through the function '`run`', which is a Jess function that starts the pattern matching process. The dynamic role can be mapped to the Jess function '`assert`' which asserts all facts into the working memory of the inference engine. In the CPG system, this can be seen in the interview module in getting the facts to the working memory and asserting the recommendations.

There is no direct mapping for knowledge base and tuple, but the `defmodule` constructs of Jess allows large number of rules to be physically organised into logical groups. Modules also provide a control mechanism that only allows the module that has the *focus* to fire the rule in it, and only one module can be in focus at a time. In the CPG system, the recommend module is used to organise the rules into knowledge base and static role can be mapped to the focus function of Jess since all the CPG rules for the inference engine are contained here. The role of transfer function in obtaining additional information can be mapped to the `defmodule` construct that implements the appropriate functions to get this information.

The rule element of the profile can be mapped directly to the `defrule` construct of Jess in which the antecedent part corresponds to the left-hand side (LHS) of the rule and the consequent part corresponds to the right-hand side (RHS) of the rule. The following example of manual mapping the CPG system rule '`S-C-F-1-0-0`' shown in table 4 would help demonstrate this better.

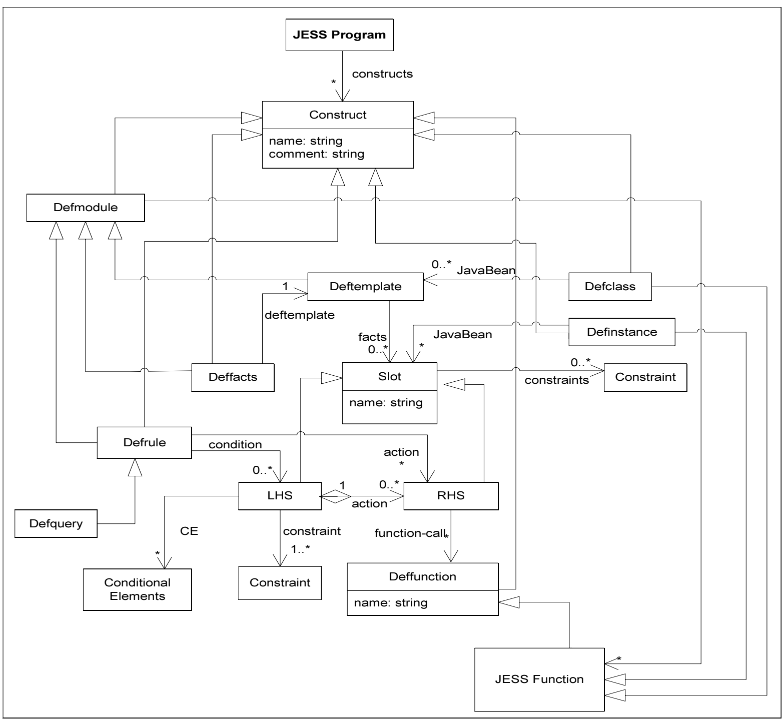


Fig. 5. Jess Meta-model

Table 3. Possible mapping of the Knowledge Modelling

Profile Concepts	mapping	JESS Concepts
DomainConcept	=	Deftemplate (Frame) Slot Defclass Definstance
FactBase	=	Deffacts
Task	≈	Defmodule
Task Method	≈	Defmodule
Inference	≈	Deffunction – run ( )
Dynamic role	≈	Deffunction – assert ( )
Static Role	≈	Defmodule - focus
Transfer function	≈	Defunction
Knowledge base	≈	Defmodule - focus
Tuple	≈	Defmodule – focus (partition the rules)
Rule	=	Defrule
• Implication Rule o Antecedent o Consequent	= = =	Defrule • LHS, RHS Deffunction, Conditional Elements, Defquery

In line 1, we define the rule using defrule which states the name of the rule – in this case strength = 1, category = null and factor = null S-C-F-1-0-0 which will list all recommendation of strength values of 1. Line 2, 3 and 4 is the LHS of the rule which consists of facts matching patterns and line 5 and 6 contains the function call (RHS) which asserts the recommendations values.

Table 4. CPG `'S-C-F-1-0-0'` rule

```
1 defrule S-C-F-1-0-0
2 user (strength ?i&:(= ?i 1))
3 cate_gory  ?d&:(= ?d 0))
4 factors_type ?j&:(= ?j 0)))
5 => assert recommendation S-C-F STR1) (explanation
6 "Strength equals 1 Recommendation I1,I2,I3,I4)"))))
```

6. Conclusion and Future Work

KBS development is similar to that of SE where they both rely on conceptual modelling of the problem domain to provide an orientation on how the system should address the problem. UML has been adopted in the SE domain as a standard for modelling, but there is still no consensus in the field of KE. This chapter describes an extension to UML using the profile mechanism for knowledge modelling that allows KBS to be designed using an object-oriented approach. The profile has been successfully tested on several case studies involving KBS design and development from scratch and in re-engineering an existing KBS.

The following discusses KBS modelling in the context of the OMG Production Rule Representation standardisation work. The PPR work mainly requires the use of activity diagrams to model the relationship between rulesets to action states. However, in this work we have identified that the use of activity diagram is limited to model a particular process of the system. Furthermore, class diagram can only provide partial snaphots of the system at a particular point in time which is less meaningful in complex inference cycles. To overcome this limitation, we have used the sequence diagram which clearly helps to understand the flow of logic in the system as shown in section 4. The profile described in this chapter would help in understanding how rules are related to the domain concept elements in the KBS and the processes that are involved in activating the rule to fire with the help of activity and sequence diagram. Furthermore, the profile only shows the categories of rule which can be modelled in a single diagram with the other model elements. Thus the profile would help overcome the current problem of omitting rules from the model.

Mapping the profile to PSM is only limited to domain concept, factbase and implication rule. The rest of the profile elements are useful to describe the KBS and usually implemented differently as runtime concepts in various rule engines. Nevertheless, this proves that the most important work in designing and developing KBS is writing the rules based on the domain concepts which attribute values stored in the fact base will activate the rules. As such, the standardisation work in PRR should first emphasise on agreeing standard representation of rule elements in writing rules, which are portable across different inference engines.

Currently work has concentrated on building an Eclipse plug-in to support the profile as it is a popular implementation tool for UML profiles. The plug-in allows profile-compliant diagrams to be drawn and validated, and XML or XMI representations produced. The infrastructure in the Eclipse makes this mapping straightforward to implement. The future work in this area involves studying how to automate the generation of Jess code from the profile elements that can be mapped to Jess meta-model. The work in automating the generation of Jess code from models is still in progress [37].

References

[1] Davenport, T.H. and Prusak, L. *Working Knowledge: How Organizations Manage What They Know*. 2000, Boston: Harvard Business School Press.

[2] Devedzic, V., *Knowledge Modelling - State of the Art.* Integrated Computer-Aided Engineering, 2001. **8**(3): p. 257-281.

[3] Giarratano, J.C. and G.D. Riley, *Expert Systems: Principles and Programming*. 4 ed. 2004, Boston, Massachusetts: Course Technology - Thomson.

[4] Studer, R., Benjamins, R.V. and Fensel, D. *Knowledge Engineering: Principles and Methods.* Data & Knowledge Engineering, 1998. **25**: p. 161-197.

[5] Schreiber, G., Akkermans, H., Anjewierden, A., deHoog, R., Shadbolt, N. de Velde, W.V., and Wielinga, B. *Knowledge Engineering and Management: The CommonKADS Methodology*. 1999, Massachusetts: MIT Press.
[6] Ergazakis, K., Metaxiotis, K., and Psarras, I. *Knowledge Management in Enterprises: A Research Agenda.* Intelligent Systems in Accounting, Finance and Management, 2005. **13** (1): p. 17-26.
[7] Studer, R., Decker, S., Fensel, D., and Staab, S. *Situation and Perspective of Knowledge Engineering*, in *Knowledge Engineering and Agent Technology. IOS Series on Frontiers in Artificial Intelligence and Applications.*, J. Cuena, et al., Editors. 2000, IOS Press: Amsterdam.
[8] Liebowtiz, J., *If You Are A Dog Lover, Build Expert System; If You Are A Cat Lover, Build Neural Networks.* Expert Systems With Applications, 2001. **21**: p. 63.
[9] Preece, A., *Evaluating Verification and Validation Methods in Knowledge Engineering*, in *Micro-Level Knowledge Management*, R. Roy, Editor. 2001, Morgan-Kaufman: San Francisco. p. 123-145.
[10] Tabet, S., Wagner, G., Spreeuwenberg, S., Vincent, P., Jacques, G., de Sainte Marie, C., Pellant, J., Frank, J., and Durand, J . *OMG Production Rule Representation - Context and Current Status*, in *W3C Workshop on Rule Languages for Interoperability.* 2005. Washington, D.C., USA.
[11] Newell, A., *The Knowledge Level.* Artificial Intelligence, 1982. **18**: p. 87-127.
[12] Gómez, A., Moreno, A., Pazos, J., and Sierra-Alonso, A. *Knowledge maps: An essential technique for conceptualisation.* Data & Knowledge Engineering, 2000. **33**(2): p. 169-190.
[13] Juristo, N. and A.M. Moreno, *Introductory paper: Reflections on Conceptual Modelling.* Data & Knowledge Engineering, 2000. **33**(2): p. 103-117.
[14] Gomez-Perez, A. and V.R. Benjamins. *Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods.* in *IJCAI-99 Workshop on Ontologies and Problem-Solving Methods (KRR5)*. 1999. Stockholm, Sweden, p. 1-1 - 1-15.
[15] Gruber, T.R., *Toward Principles For The Design Of Ontologies Used For Knowledge Sharing*. 1993, Stanford University.
[16] Kende, R., *Knowledge Modelling in Support of Knowledge Management.* Lecture Notes in Artificial Intelligence, 2001. **2070**: p. 107-112.
[17] Naumenko, A. and A. Wegmann. *A Metamodel for the Unified Modeling Language*. in *UML 2002*. 2002. Dresden, Germany.: Springer, Berlin, p. 2-17.
[18] Cuena, J. and M. Molina, *The Role Of Knowledge Modelling Techniques In Software Development: A General Approach Based On A Knowledge Management Tool.* Int. Journal of Human-Computer Studies, 2000. **52**: p. 385-421.
[19] Juristo, N., *Guest editor'.* Knowledge Based System, 1998. **11**(2): p. 77-85.
[20] Gill, G.T., *Early Expert Systems: Where Are They Now?* MIS Quarterly, 1995. **19**(1): p. 51-81.
[21] Tsui, E. (2005) The Role Of It In KM: Where Are We Now And Where Are We Heading. *Knowledge Management,* 9, 3-6.
[22] Chan, C. W. (2004) Knowledge And Software Modeling Using UML. *Software And Systems Modelling,* 3**,** 294-302.
[23] Krovvidy, S., Bhogaraju, P. & Mae, F. (2005) Interoperability And Rule Languages. *W3C Workshop On Rule Languages For Interoperability.* Washington, D.C., USA.
[24] Parpola, P., *Inference in the SOOKAT object-oriented knowledge acquisition tool.* Knowledge and Information Systems, **8**(3) : p.
[25] Cowling, A.J. *The role of modelling in the software engineering curriculum*. Journal of Systems and Software. **75**(1-2): p. 41-53.
[26] Bratko, I. PROLOG Programming for Artificial Intelligence - 3rd edition, 2000, Boston, Addison Wesley.
[27] Friedman-Hill, W. Jess in Action: Rule-Based System in Java, 2003, Greenwich, US, Manning Publications Co.
[28] Uhl, A.., *Point: Model Driven Architecture IS Ready for Prime Time*. IEEE Software, 2003. 20(5): p.70-71.
[29] Purchase, H.C., Welland, R., McGill, M., and Colpoys, L., *Comprehension of diagram syntax: an empirical study of entity relationship notations.* International Journal of Human-Computers Studies, 2004. **61**: p. 187-203.
[30] McClintock, C. (2005) Ilog's Position On Rule Languages For Interoperability. *W3C Workshop On Rule Languages For Interoperability.* Washington, D.C., USA.
[31] Muller, P.-A., Studer, P., and Bezivin. J. *Platform Independent Web Application Modeling*. In The Sixth International Conference On The Unified Modeling Language (UML 2003). 2003: Springer, p. 220-233.
[32] OMG, *Unified Modeling Language specification (version 1.4)*. 2001.

[33] Clark, T., Evans, A., Sammut, P., and Willians, J. *Metamodelling for Model-Driven Development, 2005; Accessible at* http://albini.xactium.com.

[34] Abdullah, M.S., Evans, A., Paige, R., Benest, I., and Kimble, C. Modelling Knowledge Based Systems Using the eXecuta ble Modelling Framework (XMF), in Cybernetic and Intelligent Systems (CIS) 2004. IEEE Press, p 1053-1059.

[35] OMG, *Requirements for UML Profile*. 1999, Object Management Group: Framingham, MA, U.S.A. p. 8.

[36] Perez-Martinez, J.E., *Heavyweight Extensions To The UML Metamodel To Describe The C3 Architectural Style.* ACM SIGSOFT Software Engineering Notes, 2003. **28**(3), p. 5-5.

[37] Wu, C.G. (2004) Modelling Rule-Based Systems with EMF. Accessed at http://www.eclipse.org/articles.

A Semantic-Based Navigation Approach for Information Retrieval in the Semantic Web

Mourad OUZIRI
Centre de Recherche en Informatique de Paris 5, Paris Descartes University, France

Abstract. Information retrieval in the Web needs more efficient way than traditional approaches. That is, traditional information retrieval approaches are not accurate because they consider only the syntactical level when processing Web resources. So, more conceptual approaches are needed to increase accuracy of information retrieval. In the next generation of the Web, namely the semantic Web, the challenge of information retrieval approaches is to design automatic programs that are able to understand and process semantics of Web resources. Therefore, semantics needs to be clearly formalized and processed by programs. In this chapter, we use a knowledge representation formalism to represent the semantics of Web resources and then to perform semantic navigation. The knowledge representation formalism is used to design a semantic index of the Web resources. Generally, it references a huge amount of Web resources. So we present an efficient navigation strategy of this index to make more accurate information retrieval.

Keywords. Knowledge Representation, Subject-oriented navigation, Semantic Web, Topic Maps

Introduction

Web-based applications require access to multiple datasources to supply relevant information. Unfortunately, these datasources represent a huge and heterogeneous collection of data. So, searching the good, relevant, up-to-date information in this huge volume of data is a complex task.

In the first generation of the Web, datasources provide data in HTML documents. Information retrieval is made in two steps:

- Step 1: Indexing HTML documents. It consists to extract significant terms from HTML documents to construct the index. The index is an efficient data structure. It contains for each indexing term the set of relevant HTML documents.
- Step 2: Keyword search. The user submits a keyword, or a boolean expression of keywords. All the HTML documents referenced by the user keyword in the index are returned. This keyword search is syntactic. That is, only syntactic comparison between user keyword and keywords describing HTML documents is performed.
- Step 2: Navigation. Starting from returned HTML documents, the user gets information from these HTML documents or navigates by following contained hyperlinks.

This traditional information retrieval is not efficient because it is based on syntactic indexing and search. That is, many non-relevant HTML documents may be returned for a keyword search. This would bring user to make navigation starting from non-relevant HTML documents. Then, the user makes a less accurate information retrieval even when the interface adapts its content to user profile.

In the second generation of the Web, namely the semantic Web, quality of Web research is improved by considering semantics of Web resources. The vision of the semantic Web is to process semantics of Web resources automatically by machine programs. For this aim, Web resources are annotated using, as in traditional Web, terms. But these terms have common semantics defined as concepts of an ontology. Semantic Web technologies intend to manage Web resources, annotation terms and ontology in order to perform efficient and accurate information retrieval based on semantics.

The rest of this chapter is presented as follows. In the next section, we give some related works that aim to represent and search information in the Web. Then, we present in the second section two knowledge representation formalisms adapted to manage the semantics of Web resources. We use the second formalism in this chapter, namely the Topic Maps. The core of this chapter is presented in the third section. We present the design a semantic index that describes Web resources and we define an efficient strategy to make more accurate navigation using the designed index.

1. Related Works

In order to access the Web resources that belong to multiple datasources through a unique interface, content of these datasources is represented and indexed in a knowledge structure. The efficiency of information retrieval depends of this index. We present in this section some existing approaches that aim to represent content of Web datasources and make information retrieval.

1.1. Semistructured Document-Based Data Integration

Data integration represents an important task to access and query multiple Web datasources in a coherent way. In the Web, data are represented with semi-structured HTML or XML models. For HTML documents, data integration consists to link HTML (or XML) documents with each other by hyperlinks. This is a static and rigid approach because semantic relationships are not considered. XML-based data integration is realized with a query language for querying multiple XML documents using one query [1] or by providing a uniform view of multiple XML documents [2]. To integrate XML documents, a mechanism to identify multiple instances of a same real object is proposed in [3]. Semi-structured data models, OEM [4] and XML, are used in data integration process [5,6]. This type of integration is not useful for expressing semantics. However, XML does not give any semantics about taggs.

1.2. Ontology-Based Approach

Ontology represents a pertinent way to resolve problems related to semantic heterogeneity. Ontology plays an important role in the knowledge representation. It allows sharing semantics of structural units. Ontology is mainly used like a global schema of Web datasources and a query interface. Ontology concepts are linked to datasources through a global meta-model. Theses links are used to identify the relevant datasources and to transform queries into sub-queries. In the literature, three main ontology-based architectures are proposed [7,8]: simple, multiple and hybrid architectures. Buster [9] uses hybrid ontology. In this system, the ontology is seen as a knowledge base on which semantic integration is based. The terms of a datasource are defined with a local ontology. The integration in Buster consists to define a common ontology which is used to annotate the terms resulted from the first annotation. Multiple ontologies are used in Observer [10]. It associates an ontology to each datasource. This association is formalized by links between the concepts of the ontology and the terms of the datasource. The datasource integration is done with semantic links (synonymy, hyponymy, disjunction, etc.) between the concepts of the ontologies.

1.3. Adaptive Navigation

Integration of multiple Web datasources provides a huge amount of data to be navigated. To make more accurate navigation, Web navigation interfaces become more and more adaptive. In the context of the Web, these navigation interfaces implement the so-called adaptive hypermedia navigation. Many adaptive hypermedia conception techniques are proposed [11,12]. These modeling techniques are based on data modeling, navigation structure modeling and abstract interface modeling.

The idea of adaptive hypermedia navigation consists to couple together the domain model and the user model using an adaptivity engine [13]. The domain model represents concepts and their associations of the application domain. The user model represents user objectives, preferences, goals and backgrounds. Adaptation is made by the adaptivity engine using rules [14]. These rules have the form: If event Then action.

Some adaptation rules are executed before presenting the document and others are executed after presenting the document. The rules of the first class are useful to select the document fragments to be shown or hidden and to define the order of these fragments and their visualization specificities. The rules of the second class are mainly used to update the user profile according to this action.

According to [14], hypermedia adaptation is performed at two levels: content adaptation and navigation adaptation. The content adaptation consists to include only pertinent fragments in the presented document, which are organized according to user preferences and goals. In AVANTI system [15], a document fragment can be described using a visual description or acoustic description. For visually handicapped persons, the fragment is presented using acoustic description. As in ELM-ART II [16], the navigation adaptation consist to reorder, annotate, color according to pertinence of referenced document, disabled or removed if the referenced document is not pertinent, the hyperlinks of the presented document.

In [17], adaptation is made into two steps: query preview and query refinement. In the first step, the interface presents only some attributes, which are used by the user to specify the so-called *query preview* in order to select only parts on universe of data containing pertinent information. In the second step, the interface presents more attributes, which are used to formulate a *refinement query* evaluated over the data resulted from the first step.

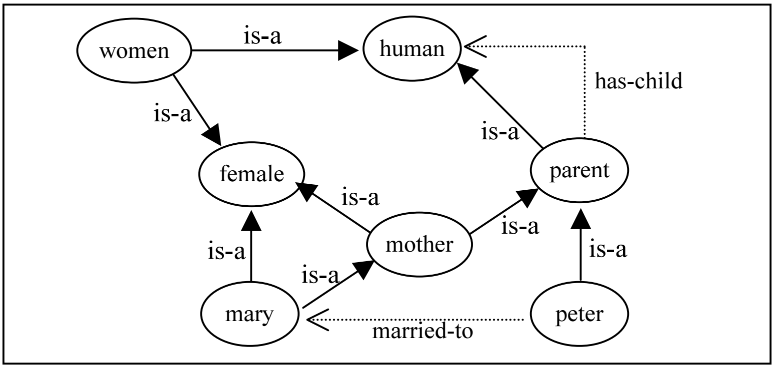


Figure 1. A semantic network representing knowledge about a family.

2. Knowledge Representation Formalisms for the Semantic Web

Knowledge representation formalisms aims to express and organize knowledge about the world in computers. So, automatic program may retrieve information and make reasoning to infer implicit knowledge. In this section, we present two useful knowledge representation formalisms for the semantic Web, namely semantic networks and Topic Maps.

2.1. Semantic Networks

A semantic network is a graphic notation for representing knowledge in patterns of interconnected nodes and arcs [18]. The basic form of semantic networks consists of a directed graph where vertices represent concepts or individual objects and edges represent semantic relations between the concepts and objects. What is common to all semantic networks [18] is a declarative graphic representation that can be used either to represent knowledge or to support automated systems for reasoning about knowledge.

The structure of the network defines its meaning. The meanings are merely which node has a pointer to which other node. The network defines a set of binary relations on a set of nodes. In the Fig. 1, *peter* is a *parent* and is *married to mary*. *mary* is a *female parent*. *parent* is *human* having a child, which is *human*.

Creations and uses of semantic networks have led to some problems. That is, there is no formal semantics, no agreed-upon notion of what a given representational structure means. For example, the *is-a* relationship, as typically used in semantic networks, appears to hide the important distinction between an individual object (such as Mary) or concept and a term specifying a class of objects or entities (such as *woman*, or *human*). The statements '*mary* is a *mother*' and '*woman* is a *human*' would have quite different analyses in predicate logic, the latter being the disguised universal statement 'All *womens* are *human*'. So semantic networks can be criticized for obscuring such important logical distinctions.

Of course, the success of logic in this respect is debatable, but semantic networks do tend to rely upon the procedures that manipulate them, especially for inference. From the example of Fig. 1, using statements '*mother* is a *female* and *parent*' and

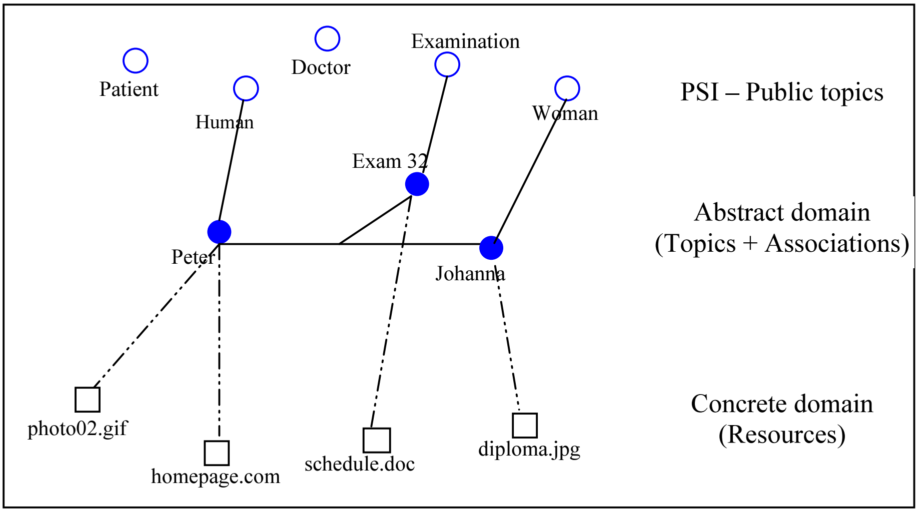


Figure 2. An example of a topic map representing part of a knowledge base in medicine.

‘*woman* is a *female* and *human*’ we can intuitively infer that ‘*mother* is a *women*’. However, in semantic networks there is no well-defined reasoning algorithms that allow to make automatiquelly this inference.

2.2. Topic Maps

Like semantic networks, Topic Maps [19] is a graph-based formalism for knowledge representation. It is also used as a mechanism for representing and optimizing access to resource [20]. Topic Maps are adapted to express distributed knowledge on the Web.

A topic map[1] is built with topics in a networked form. A topic can be anything that is a subject, regardless of whether it exists or not. It is the formal representation of any subject, abstract or real, in a computer system, such as a person, John, the earth, the planet, etc.

Topics can be linked together by associations expressing some given semantics. Topic Maps applications define the nature of the associations and the roles played by the topics in these associations. Associations are used to express knowledge between topics and not between occurrences. Topics and associations represent the abstract part of a topic map (see Fig. 2). The concrete part is represented by occurrences. Occurrences are resources linked to topics. A topic occurrence can be any information that is represented/indexed by a given topic. For example, personal homepage, photos and diploma may be occurrences of a topics representing persons. In the Fig. 2, *doctor Peter examines the woman Johanna, the exam is saved in the word document ‘schedule.doc’*, *Peter has a photo and a homepage, Johanna has a diploma scanned in the file diploma.jpg*.

In order to process and exchange topic maps on the web, a standard called XTM [21] – XML for Topic Maps – defining a Topic Maps model and its syntax was

[1] Topic map (t, m in tiny) references a knowledge base structured with respect to the Topic Maps formalism (T, M in capital letters).

```
<topicMap  xmlns:xlink="http://www.w3.org/1999/xlink">
  <topic id="peter-id">                                   /* topics of the topic map */
    <instanceOf><subjectIndicatorRef xlink:href="http://www.xx.com/onto.daml#human"/>
    </instanceOf>
    <topname><basename>Dr. Peter</basename></topname>
  </topic>
  <topic id="Johanna-id">
    <instanceOf><subjectIndicatorRef xlink:href="http://www.xx.com/onto.daml#woman"/>
    </instanceOf>
    <topname><basename>Ms. Johanna</basename></topname>
  </topic>
  <topic id="exam32-id">
    <topname><basename>examination-032</basename></topname>
    <instanceOf><subjectIndicatorRef xlink:href="http://www.xx.com/onto.daml#examination"/>
    </instanceOf>
  </topic>
  <association>                                           /* association between topics*/
     <instanceOf> <topicRef="# exam32 -id"/></instanceOf>
     <member>
          <topicRef="#peter-id"/>
          <roleSpec>
            <subjectIndicatorRef xlink:href="http://www.xx.com/onto.daml#doctor"/>
          </roleSpec>
     </member>
     <member>
          <topicRef="#Johanna-id"/>
          <roleSpec>
              <subjectIndicatorRef xlink:href="http://www.xx.com/onto.daml#patient"/>
          </roleSpec>
     </member>
  </association>
</topicMap>
```

Figure 3. An XTM document representing topic map of Fig. 2.

edited by ISO/IEC 13250 in 1999. XTM is a XML type definition which defines a format for Topic Maps specifications. It allows to manage and process topic maps by a large amount of applications. In the Fig. 3, we give a XTM representation of the topic map of the Fig. 2.

In the topic map of Fig. 3, the individuals *Peter* and *Johanna* are represented by the topics *peter-id* and *johanna-id*. These two topics are associated by the association reified by the topic *exam-32*, which is instance of the external topic *examination*. Each topic of an association plays a role. The topic *peter-id* plays the role of doctor in the association *examination* and *johanna-id* plays the role of *patient*. We note that, generally, associations expresses the most important knowledge.

The use of Topic Maps in our design of semantic Web is motivated by two important features:

– First, anything is topic in Topic Maps. Subjects, objects, associations and roles are represented by topics. In the previous example, the individuals *Peter* and *Johanna* are represented by topics, their association is an instance of a topic, and roles are topics too. This characteristic allows Topic Maps to smooth knowledge whatever its structure of origin.

– Second, Public Subject Indicator – PSI [21]. PSI is a URI to a public topic maintained (in an ontology) apart from the topic map and referenced in the subjectIndicatorRef element of XTM. In Fig. 3, *peter-id* is instance of topic *human*, which is maintained as public subject in an ontology.

However, in Topic Maps, as a kind of semantic networks, semantics of topics and associations is not well-defined. Therefore, there is no standard reasoning algorithms, which allow to make inferences on topic maps. More defined semantics in representation formalisms are defined for logics. In the semantic Web, description logics [22] are the most used logics. Description logics are not presented in this chapter.

3. Semantic Navigation on the Web Resources

As motivated in the introduction, traditional information retrieval in the Web is not efficient because it does not consider semantics of HTML documents. In this section, we present more efficient information retrieval by navigating on the Web resources. We use on the design of the system an approach that captures semantics of Web resources and use it to make navigation. Web resources belong to multiple datasources. Semantics of Web resources is common to all datasource. So, Web resources are interpreted with respect to this shared semantics.

To be processed by automatic programs, semantics must be formalized. In the framework of the semantic Web, semantics is formalized by ontology. Ontology is defined as a formal specification of a shared conceptualization [23]. So, to make semantic navigation of Web resources, we represent content of Web resources using concepts of an ontology. This produces a semantic index. Then, we design a navigation strategy over this semantic index to access Web resources more efficiently.

3.1. Representing Web Resources Using Topic Maps

Representing Web resources consists to describe them using computer readable formalism in order to be retrieved automatically by computer programs. In the presented approach, we represent Web resources belonging to multiple Web datasources to be indexed by a homogeneous and semantic index. This resulting index will be used to design a powerful strategy to explore the Web resources.

In the presented apporoach, Web resources are represented and indexed using Topic Maps [19]. Topic Maps is an expressive formalism. It can express any knowledge whatever its complexity. Knowledge expressed using Topic Maps, a topic map, is serialized in a XTM [21] document.

Considering the e-learning application domain. Then, Web resources are learning objects, learner profiles, etc. We consider only learning objects in this study. So, resources are courses, illustrations, exercises, etc. In Fig. 4, resources are a course and an example in pdf files. They are represented by the topics *course12-id* and *example209-id* respectively. These topics are related using an association reified by the topic *relevance-id*. This association represents the following knowledge '*the example given by the learning resource represented by the topic example209-id is relevant to the course described by the learning resource represented by the topic course12-id*'.

```
<topicMap  xmlns:xlink="http://www.w3.org/1999/xlink">
  <topic id="learning-unit"> <topname> learning unit </topname> </topic>
  <topic id="example-unit">
     <topname>example illustrating some idea</topname>
  </topic>
  <topic id="relevance">
    <topname>example relevance to course</topname>
  </topic>
  <topic id="title"> <topname> title of a learning resource</topname>
  </topic>
  <topic id="course12-id">                    /* topics of the topic map */
    <instanceOf><topicRef xlink:href="#learning-unit"/> </instanceOf>
    <occurrence>
       <instanceOf> <topicRef xlink:href="#title"/> </instanceOf>
       <resourceData> introduction to relational databases </resourceData>
    </occurrence>
    <occurrence>
       <resourceRef> ftp://site.com/file65.pdf </resourceRef>
    </occurrence>
  </topic>
  <topic id="example209-id">
   <instanceOf><topicRef xlink:href="#example-unit"/> </instanceOf>
     <occurrence>
        <instanceOf> <topicRef xlink:href="#title"/> </instanceOf>
        <resourceData> An example of a relational schema </resourceData>
     </occurrence>
    <occurrence>
       <resourceRef> ftp://site.com/file106.pdf </resourceRef>
    </occurrence>
  </topic>
  <association>                                  /* topic associations*/
      <instanceOf> <topicRef="#relevance"/></instanceOf>
      <member>
         <topicRef xlink:href="#example209-id"/>
         <roleSpec> <topicRef xlink:href="#illustration"/> </roleSpec>
      </member>
      <member>
         <topicRef xlink:href="#course12-id"/>
         <roleSpec> <topicRef xlink:href="#illustrated"/> </roleSpec>
      </member>
   </association>
</topicMap>
```

Figure 4. An example of an e-learning knowledge in Topic Maps.

Each topic plays a role in an association. In this example, the topic *example209-id* plays the role of an *illustration* and the topic *course12-id* plays the role of the *illustrated*.

So, each Web datasource is represented and indexed by a topic map as in Fig. 4. Now we integrate the generated topic maps representing the Web datasources into a unique global topic map-based index. Before that, we enrich topic maps with semantics in order to make more efficiennt integration of Web datasources.

3.2. Representing Semantics of Web Resources

Semantics is defined as a shared knowledge, which provides common meaning to data. In this chapter, semantics is given as a DAML+OIL [24] ontology. DAML+OIL is compatible with Topic Maps. We show how DAML+OIL and Topic Maps can be easily used jointly to represent semantics about Web resources. Let us see an example of a DAML+OIL ontology given in the following Figure:

```
<daml:Class rdf:ID="Course">
        <rdfs:subClassOf rdf:resource="#Learning Object"/>
        <daml:ObjectProperty rdf:ID="require">
            <rdfs:range rdf:resource="#Course"/>
            <rdfs:domain rdf:resource="#illustration"/>
        </daml:ObjectProperty>
</daml:Class>
<daml:ObjectProperty rdf:ID="relevance">
      <daml:inverseOf rdf:resource="#require"/>
</daml:ObjectProperty>
<daml:Class rdf:ID="Illustration">
        <rdfs:subClassOf rdf:resource="#Learning Object"/>
</daml:Class>
```

Figure 5. Part of a DAML+OIL ontology in e-learning application domain at http://onto.org/ontology.daml.

Technically, data enriching with semantics is very appropriate using Topic Maps and DAML+OIL. It is simply done by adding, for each topic of a topic map, the subjectIdentity element [21] of Topic Maps. In Topic Maps, the subjectIdentity element should reference a PSI (Published Subject Indicator [21]). As defined in Topic Maps, PSIs are a set of unambiguous and well-defined subjects or concepts. They are public. It means that they are accessed by anyone. In our works, we consider ontology as a set of PSIs. That is, each concept of a DAML+OIL ontology can be referenced as a PSI.

So, adding semantics to learning resources consists to reference concepts of an ontology from the topics that index/describe the Web resources using the subjectIdentity element. In the Fig. 6, we enrich the Web/learning resources represented in the Fig. 5 by semantics given by the DAML+OIL ontology of Fig. 5.

In this example, the topic learning-unit is semantically defined as a course. Technically, this is made by linking the topic learning-unit to its corresponding concept Course in the ontology http://onto.org/ontology.daml using the subjectIdentity element.

3.3. Semantic Integration of Distributed Web Resources

Web resources belong to different Web datasources. Integratiing several Web datasources allows accessing to more complete resources through a unique index.

As shown previously, Web resources are represented in Topic Maps. Semantics of Web resources is incorporated in topic maps as links to corresponding concepts in a DAML+OIL ontology. This is made for each Web datasource leading to one topic map for each datasource. So, semantic integration consists to merge these topic maps into a single one.

We note that we use a single ontology architecture. Multiple ontologies architecture [7] may also be used.

Integrating the Web datasources consists to merge all the resulted topic maps into a single one. Topic Maps is an adequate formalism for integration. That is, everything is topic in Topic Maps. Even associations are reified by topics. Thus, integration is simply made by merging topics referencing a same concept of the common ontology.

```
 ...
<topic id="learning-unit">
    <subjectIdentity>
        <subjectIndicatorRef xlink:href="http://onto.org/ontology.daml#Course"/>
    <subjectIdentity>
    <topname> learning unit </topname>
</topic>
<topic id="example-unit">
    <subjectIdentity>
      <subjectIndicatorRef  xlink:href="http://onto.org/ontology.daml#Illustration"/>
    <subjectIdentity>
    <topname>example illustrating some idea</topname>
</topic>
<topic id="relevance">
   <subjectIdentity>
       <subjectIndicatorRef xlink:href="http://onto.org/ontology.daml#relevance"/>
  <subjectIdentity>
  <topname>relevance of an example to a course</topname>
</topic>
 ...
</topicmap>
```

Figure 6. Representing semantics of Web-based learning resources in Topic Maps (completes the Fig. 4).

```
<course ID="c123">
   <description> databases for beginners <description/>
   <require>
       <illustration id="i98"/>
   <require>
</course>
```

Figure 7. A resource course represented in XML.

Let us consider, in the following figure, another learning datasource giving an XML-based learning resources.

The corresponding topic map describing content and semantics of this XML-based learning is given in Fig. 8.

Now, we consider that we have to integrate the two Web learning datasources represented in the topic maps of Figs 4&6 (the Fig. 4 and Fig. 6 constitute one complete topic map) and Fig. 8, respectively.

Semantic integration of topic maps consists to determine the topics representing the same concept of the ontology. This is done using the subjectIdentity sub-element. In Figs 4&6, the topic learning-unit references the concept Course in the common ontology. So, the learning object course12-id described by the topic learning-unit is a Course. Then, the topics learning-unit of Figs 4&6 and the topic course of Fig. 8, referencing the same concept of the common ontology, are merged into the single topic course, in the global topic map. In the same way, the topics example-unit and illustration reference the same concept in the common ontology. Thus, they are merged into a single topic that we call illustration. This is done for all topics of the topic maps that reference same concepts of the ontology.

```
<topicMap  xmlns:xlink="http://www.w3.org/1999/xlink">
   <topic id="c123">
     <instanceOf><topicRef xlink:href="#course"/> </instanceOf>
     <occurrence>
       <instanceOf> <topicRef xlink:href="#description"/> </instanceOf>
       <resourceData> databases for beginners </resourceData>
      </occurrence>
   </topic>
   <topic id="i98-id">
      <instanceOf><topicRef xlink:href="#illustration"/></instanceOf>
   </topic>
   <topic id="course">
       <subjectIdentity>
           <subjectIndicatorRef xlink:href="http://onto.org/ontology.daml#Course"/>
        <subjectIdentity>
   </topic>
   <topic id="illustration"> … </topic>
   <topic id="require">  …  </topic>
   <topic id="description">
      <subjectIdentity>
          <subjectIndicatorRef xlink:href="http://www.xx.com/ontology.daml#title"/>
      <subjectIdentity>
     <topname> description/title of learning resource </topname>
   </topic>
   <association>
       <instanceOf> <topicRef="# require"/></instanceOf>
       <member>
          <topicRef xlink:href="#c123"/>
          <roleSpec><topicRef xlink:href="#course"/></roleSpec>
       </member>
       <member>
          <topicRef xlink:href="#i98"/>
          <roleSpec><topicRef xlink:href="#illustration"/></roleSpec>
       </member>
   </association>
</topicMap>
```

Figure 8. Topic map representing the XML-based learning resource of Fig. 7.

Another integration form concerns topics that reference two different concepts of the ontology but these concepts are linked by a semantic relation in the ontology. For example, the topics relevance (Figs 4&6) and require (Fig. 8) reference the concepts relevance and require, respectively, which are related in the common ontology with the relation daml:inverseOf (Fig. 5). Therefore, the system adds two implicit knowledge (not existing in the datasources) to the global topic map.

That is, for all associations t_1 is relevant to t_2 in the first datasource, the systems adds the knowledge t_2 requires t_1 in the global index and for all associations t_3 requires to t_4 in the Second datasource, the systems adds the knowledge t_4 is relevant to t_3 in the global index.

In the next section, we show an efficient strategy for the navigation of the Web resources using the Topic Maps-based index. This navigation strategy is based on the Web resources semantics to make more accurate information retrieval.

3.4. Semantic Navigation of Web Resources

As they are defined by ISO [19], Topic Maps are suitable to be used as representation formalism and navigation structure of Web resources. In the previous section, we show how Web resources are represented and indexed using Topic Maps. The Topic Maps-based index describes content of Web resources using concepts of an ontology. So, it represents Web resources at the semantic level. However, indexed Web resources result from multiple Web datasources. So, the Topic Maps-based index may represent a huge amount of Web resources. Therefore, it is important to define an efficient strategy to explore, by navigation, the huge amount of Web resources.

One of the proposed solutions to make efficient navigation in the Web consists to design navigation interfaces that adapt their content to the user profile [11,12]. In this section, we present a navigation strategy called subject-oriented navigation, which adapts navigation according to user search instead of user profile. Characteristics of the so-called subject-oriented navigation are given below.

Navigation is the main search mode on the Web. It is a natural and intuitive search mode. However, it has the *lost-in-hyperspace* drawback especially when exploring huge amounts of Web resources. The subject-oriented navigation improves traditional navigation into two directions: semantics-based navigation and subject-oriented navigation.

Semantics-Based Navigation

Topic Maps-based index describes Web resources using topics. Topics are linked by associations reified, themselves, by topics. Each topic is defined by a concept of an ontology. So, the Topic Maps-based index gives a semantic description of Web resources content. This semantics-based index is visualized as a concept map, which helps considerably user to comprehend the content of the Web resources since human mental organizes knowledge as concept map too.

Subject-Oriented Navigation

The proposed subject-oriented navigation is a strategy that aims to make more accurate navigation. It is based on the notion of subject. A subject is a user-defined object on which a navigation session is oriented. That is, all information visualized in the navigation interface is relevant to the user-defined subject. The interface that implements the subject-oriented navigation must be adaptive, dynamic and progressive. It does not present the entire concept map at once because it is not efficient to search information by exploring a huge volume of data. The interface progressively presents the concept map following the user navigation in order to adapt the content of the visualized concept map to the user needs expressed during navigation. When the data is selected by user, the visualized concept map adapts its content consequently.

At the beginning of the navigation session, the user chooses a subject, '*relational databases courses*' for example. Then, the interface visualizes only part of the concept map that is directly related to the user-defined subject. Let us consider the topics *exercise* and *illustration*. When user navigates to the concept *exercise*, respectively *illustration*, only exercises, respectively illustrations, about relational databases are visualized. Let us consider that the concept *exercise* is related to the concepts *solution* and *video* and that the user selects exercises about *SQL language*. So, only exercises about *SQL language* are given in the concept map and the concept *video* is removed from the visualized concept map if there is no video recorded for these exercises. Let us suppose that the user in interested only by *SQL language exercises* designed by the *professor Peter*. When user navigates to the concept *solutions*, only solutions of *Peter's SQL language exercises* are returned in the visualized concept map. The navigation so described is illustrated in the following Figure:

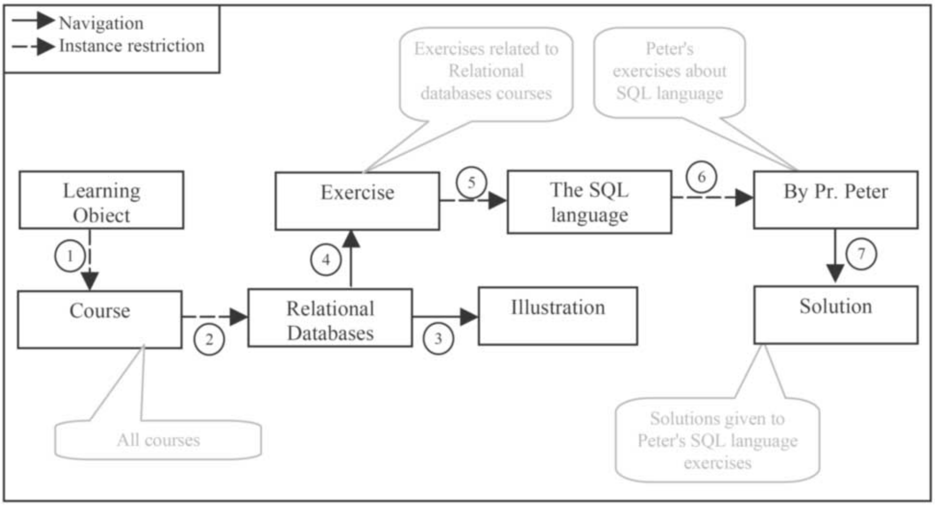


Figure 9. An example of the subject-oriented navigation in e-learning.

In the medical domain, this navigation strategy is very interesting particularly for accessing electronic patient record. In the Fig. 10, we reproduce an adaptive navigation to get *Peter's* medical record. From the concept *Patient*, we select the *patient Peter* then we navigate to the concept *Disease*. At this step, the system includes in the visualized concept map only diseases that concern the patient *Peter*. When we go to the *Examination* concept (step 3), the system gives all *Peter's examinations*. But if we select at first the *Cardiac* diseases (step 4) and we continue navigation to the concept *Examination* (step 5) then the visualized concept map gives only *Peter's examinations* about *cardiac diseases*.

The subject-oriented navigation allows to make more accurate navigation and limits non-relevant information to be returned. At each step of the navigation, the presented information is relevant to the previously visited one. In the example of Fig. 9, at the step 4 the interface presents only exercises that are relevant to *Relational Databases courses*. In Fig. 10, only *Peter's examinations* about *cardiac* diseases are presented at the step 5.

3.5. Comparison to the Traditional Navigation

Now we make the same navigation of the one given in Fig. 10 using traditional navigational approach (see Fig. 11). From the *patients.html* page, the user follows navigation from *Peter* anchor. An *html* document containing all *Peter's diseases* is returned. Then, the user follows navigation from *Cardiac disease*. Often, the return *html* page contains all *Peter's diseases* and not only *cardiac diseases*. From the *cardiac html* page, the user

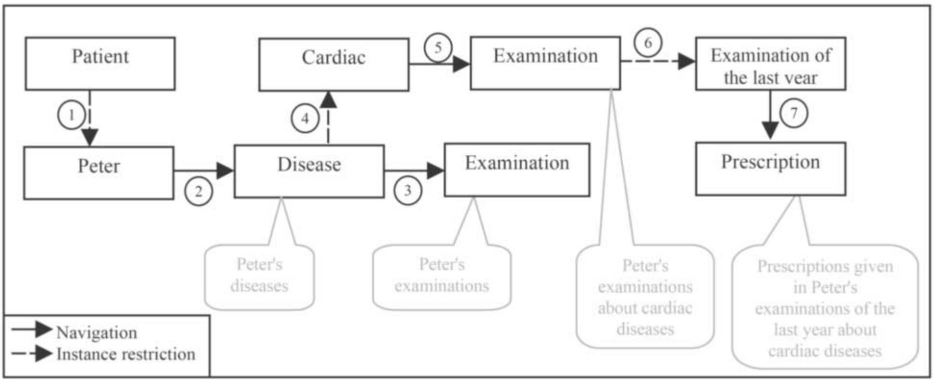


Figure 10. An example of the subject-oriented navigation to access a patient record.

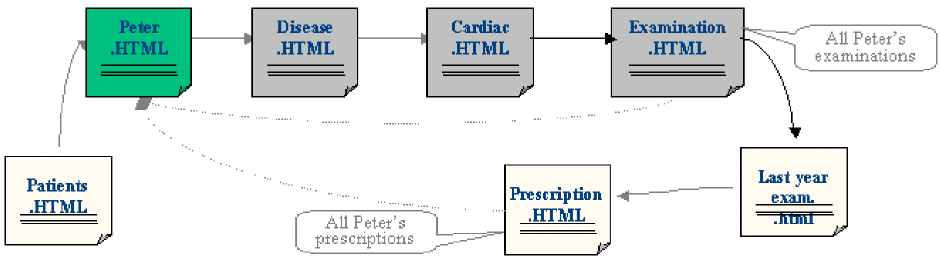


Figure 11. A simulation of the navigation made in Fig. 10 using traditional navigation.

follows navigation by the anchor *examination*. The returned *html* page would contain all *Peter's examinations* in which the user searches the *examinations* that concern *cardiac disease*.

At the end of the navigation, the *prescription html* page would contain all prescriptions of *Peter* and not only those concerning examinations of the last year. So, the user must search prescriptions of last year in this returned html page.

As conclusion, in traditional navigation the user makes another search from the returned pages. However, human search often contains errors and is time-consuming, which leads user to a non-relevant information space known by the *lost-in-hyperspace* syndrome.

4. Conclusion

Compared to the traditional Web navigation, the subject-oriented navigation presented in this chapter is more accurate. It presents only pertinent information with respect to a user-defined subject at each step of the navigation.

Web resources are represented using Topic Maps. Topic Maps formalism is convenient to make semantic navigation. In this chapter, we use Topic Maps to design a semantic index of Web resources. That is, an index is a topic map that describes Web resources using concepts of an ontology.

The structure of a topic map is simply visualized as a concept map. This visualization mode is very efficient and intuitive to user as it looks like mental knowledge. Semantic index describes a large amount of Web resources. Thus, we need for an efficient navigation strategy to explore this index. So, we propose in this chapter the subject-oriented navigation strategy. This navigation strategy restricts the visualized information space at each step of the navigation, with respect to the user-defined subject, which leads to a more accurate navigation.

The presented work may be extended according several directions. One direction is to consider constraints expressed over Web resources. Constraints brings useful semantics that allow to make more accurate navigation. For example, the constraint '*all exercises are of a big difficulty level*' allows to eliminate all exercises for beginners. Unfortunately Topic Maps do not support such constraints and there is no reasoning facilities for this formalism. Therefore, designers are brought to use other formalisms, especially those based on logics. Logics languages, such as description logics [24], have well-defined semantics and reasoning algorithms built on constraints but they are not able to manage distributed knowledge. Thus, we can use jointly Topic Maps and logic-based language as proposed in [25].

References

[1] S. Cohen, et al.: Xsearch: a semantic search engine for XML. *VLDB* (2003), 45–56.
[2] S.D. Camillo, C.A. Heuser, R.S. Mello: Querying heterogeneous XML sources through a conceptual scheman. *Proc. of ER* (2003), 186–199.
[3] D. De Brum Saccol, C.A. Heuser: Integration of XML data. *Proc. Of EEXTT* (2002), 68–80.
[4] R. Goldman, S. Chawathe, A. Crespo, J. McHugh: A standard tectual interchange format for the object exchange model (OEM). *Department of computer science, Standford University*, California, USA, 1996 5 p.
[5] Y. Papakonstantinou, H. Garcia-Moulina, J. Widom: Object exchange across heterogeneous information sources. *Proc. Of IEEE int. Conf. On Data Engineering*, (1995), 251–260.
[6] G. Gardarin, A. Mensch, A. Tomasic: An introduction to the e-XML data integraiton suite. *Proc. Of EDBT* (2002), 297–306.
[7] H. Wache, et al.: Ontology-based integration of information-a survey of existing approaches. *Workshop ontologies and information sharing. IJCAI 2001* (2001), 10 p.
[8] E. Mena, V. Kashyap, A. Illarramendi, A. Sheth: Domain specific ontologies for semantic information brockering on the global information infrastructure. *Int. Conf. On Formal ontologies in information systems. FOIS'98* (1998), Italy, 15 p.
[9] R. Meyer, et al.: Intelligent brockering of environmental information with Buster System. In Informaiton Age Economy: proc. *Of the 5th Int. Conf. 'Wirtschaftsinformatik'* (2001), Physica-Verlag, 8p.
[10] E. Mena, A. Illarramendi, V. Kashyap, A. Sheth: Observer: an approach for query processing in global information systems based on interoperation across pre-existing ontologies. *In the Int. Journal Distributed and Parallel Databases* (2000), vol. 8, n°2, 223–271.
[11] D. Schwabe, G. Rossi, S.D.J. Barbosa: Systematic Hypermedia Application Design with OOHDM. *Proc. of the 7th ACM Conference on HyperText* (1996), 116–129.
[12] F. Garzotto, P. Paolini, D. Schwabe: HDM – a model-based approach to hypermedia application design. *ACM Transactions on Information Systems* 11 (1993), 1–26.
[13] P. Brusilovsky: Methods and techniques of adaptive hypermedia. *In User Modeling and User Adapted Interaction* (1996), vol. 6, n° 2–3, 87–129.
[14] P. De Bra, P. Brusilovsky, G. Houben: Adaptive hypermedia: From Systems to Framework. *ACM Computing Surveys*, (1999) vol. 31, n° 4.
[15] C. Stephanidis, A. Paramythis, M. Sfyrakis, A. Stergiou, N. Maou, A. Leventis, G. Paparoulis, C. Karagianidis: Adaptable and Adaptive User Interfaces for Disabled Users in AVANTI Project. *In 5th International Conference on Intelligence in Services and Networks, Technology for Ubiquitous Telecom Services* (1998), *Antwerp, Belgium*, 153–166.

[16] G. Weber, M. Specht: User modeling and adaptive navigation support in WWW-based tutoring systems. *In proceedings of the 6th Int. Conf. on User Modeling, UM97*, (1997) 289–300.

[17] K. Doan, C. Plaisant, B. Shneiderman, T. Bruns: Interface and Data Architecture for Query Preview *in Networked Information Systems. ACM Transactions on Information Systems* (1999) vol. 17, n° 3, 320–341.

[18] *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, Sowa, John F. (ed.), Morgan Kaufmann Pub, 1991. 582p.

[19] ISO/IEC 13250, 1999. *Topic Maps*, Dec. ISO/IEC FCD.

[20] A. Sigel: Towards knowledge organization with Topic Maps. *XML Europe 2000,* Palais des congrès Paris (2000), 12–16 June 2000.

[21] TopicMap.org Authoring Group, (2000). XML Topic Map (XTM) 1.0.

[22] A. Borgida: Description Logics in data management. *IEEE Trans. on Knowledge and Data Engineering* (1995), vol. 7, n° 5, 671–682.

[23] T.R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2) (1993), 199–220.

[24] I. Horrocks, P.F. Patel-Schneider, F. van Harmelen: Reviewing the Design of DAML+OIL: An Ontology Language for the Semantic Web. *Proc. of 18th National Conference on Artificial Intelligence, AAAI-0* (2002) 792–797.

[25] M. Ouziri: Accessing data in the semantic web: An intelligent data integration and navigation approaches. *In Artificial Intelligence Applications and Innovations, 3rd IFIP Conference on Artificial Intelligence Applications and Innovations* (2006), 119–128.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Ontology-based Management of Pervasive Systems

Nikolaos Dimakis [a], John Soldatos [a,1], and Lazaros Polymenakos [a]
[a] *Athens Information Technology, 19.5 Markopoulou Avenue, 19002 Peania, Greece*

Abstract. Ubiquitous and pervasive systems rely on a highly dynamic and heterogeneous software / hardware infrastructure. A key factor in dealing with this diverse and sophisticated environment is the ability to 'map' all entities into a robust, scalable and flexible directory mechanism, able to maintain and manage a large number of heterogeneous components while acting not only as an information repository but also to be able to associate individual and component-specific information. This mechanism should enable the ubiquitous service designer to use a global directory service which can answer queries in an intelligent manner and relieve both the user and developer from tedious procedures of querying databases. In this paper we elaborate on the benefits gained by incorporating semantic technologies in the area of ubiquitous and pervasive computing as a means of meetings these needs. We base our experience on a case study of software component exchange among 3 different vendors which demonstrated the flexibility of adopting such a mechanism of ontology registry.

Keywords. Semantic web, reasoning, pervasive computing, ontologies

Introduction

Modern computing systems move towards pervasive environments, providing ubiquitous services to users. These services require that all devices, users, and software entities interact with each other and are integrated into an 'intelligent environment', in order to service the end-user in an "*anywhere, at any-time, to any device*" fashion [21]. An important action to realize this vision is to equip all component-members of this ubiquitous infrastructure with knowledge and reasoning capabilities so that they can understand the local and infer the global context of the event which takes place in the 'intelligent environment'. Moreover, this knowledge should be able to be distributed (shared) with new users/devices/software agents, as the operating environment of these services is by nature highly dynamic.

A key catalyst for realizing this infrastructure is the use of the Semantic Web technology, which includes languages for building ontologies and tools for processing and reasoning over information described using these ontologies. These emerging technologies will highly facilitate the expansion of pervasive context-aware systems for the following reasons:

[1]Corresponding Author: John Soldatos, Athens Information Technology; e-mail: jsol@ait.edu.gr

- Semantic Web technologies can help devices and agents that were not designed to work together, to interoperate, achieving “serendipitous interoperability” [8],
- Ontologies provide a means for distributed agents to share context knowledge and to reason about contexts [5],
- Ontology languages can be used to define policy languages for building security and privacy management systems in a pervasive computing environment.

OWL is one of the emerging Semantic Web languages that are endorsed by the W3C for building ontologies [11]. In the Semantic Web vision, OWL can help web services and agents to share information and interoperate. Using OWL, one can:

- Formalize a domain by defining classes and properties of those classes,
- Define individuals and assert properties about them, and
- Reason about these classes and individuals.

With the recent emergence of Semantic Web technologies such as the RDF [9], [3] and OWL [10], [13] numerous ontologies have been developed to conceptualize a number of pervasive computing systems such as [4], [17]. This expansion of such technologies is assisted by the use of innovative tools which facilitate the design and implementation of sophisticated ontology structures, as for example the Protégé tool for ontology management [20]. Semantic reasoning [2] has been applied in numerous occasions in the area of pervasive and ubiquitous computing, examples of which are [7] and [19]. In this paper we demonstrate how this technology can be used to support a fully automated ubiquitous computing service comprising numerous sensorial and actuating devices, context acquisition components as well as a number of distributed components such as databases and filesystems which are used as content repositories.

This work has been partly sponsored by the General Secretariat of Research and Development of Greece, under the project PRIAMOS (GSRT Measure 3.3, on Audio-Video Processing, Project Code: 26) [16] to support the implementation of human-centric context-aware applications for intelligent in-door environments. The ontological reasoning plays a significant role as the global registry service for all entities in the ‘intelligent environments’ such as users, devices and processing components.

1. Architecture

Pervasive and ubiquitous applications rely heavily on a highly dynamic and heterogeneous infrastructure. The software architectural design plays a significant role in the overall behaviour of the resulting service development and the user experience. Pervasive systems are by nature highly dynamic systems which require taking into consideration that users, software components and devices could join and/or leave dynamically. In this section we will briefly present the architectural tiers which describe our pervasive system as implemented in the scope of the PRIAMOS project supporting a rich set of ubiquitous applications. We illustrate the use of the ontology registry and how it has facilitated the development of these applications.

In figure 1 we illustrate the architecture modeled in a 3-tier stack which demonstrates the separation of each of the distinct operating levels. For the purpose of this research we will focus our description on the lower levels of the architecture which deal with the sensor manipulation and context extraction. Such layers are key to implementing

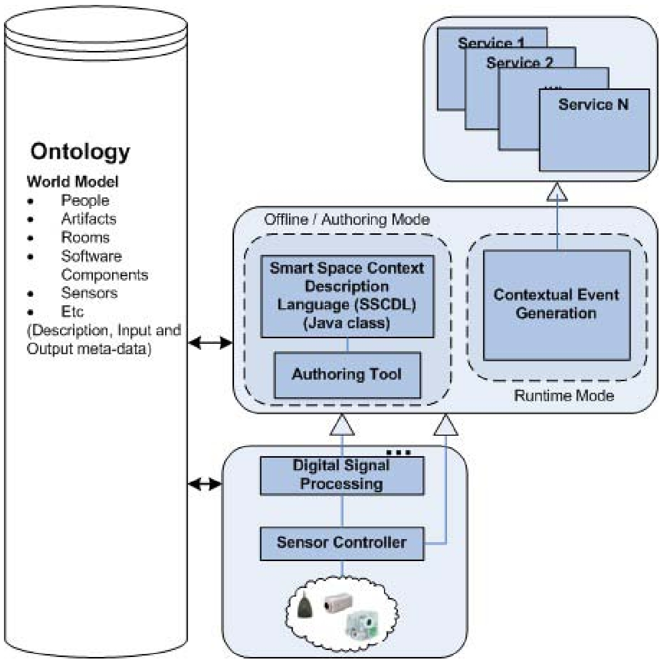


Figure 1. The development framework following a 3-tier structure

the ubiquitous service logic as they will provide the means of human instrumenting and enable situation understanding. Similar approaches were followed in the cases of higher layered software components.

1.1. Sensor and context-acquisition virtualization

At the bottom tier, the sensor control layer controls the sensor devices and facilitates the streaming of the captured data in the form of a flow, rather than raw data, which encapsulate additional information such as the nature of the sensor and the resolution. These continuous streams of sensor data are forwarded to the next layer of processing which acts as the first-order perceptive layer of the pervasive application, acquiring information pertaining to people's location and identification. This context is extracted by the use of sophisticated, real time software components which implement complicated signal processing algorithm, able of processing the received sensor stream. Examples of such algorithms implemented in our prototype laboratory are [15], [14].

1.2. Semantics Mapping

As seen in figure 1, the ontology plays a vital role in our architectural framework. Designed as a "vertical" component instead of a "horizontal" one, we aim in covering all other layers by a reference ontological representation. For each one of the previously described components, a corresponding entry has been implemented which maps the entity's characteristics to the ontology repository. Each entry contains, in addition to the technical characteristics of the specific component, additional information which can be

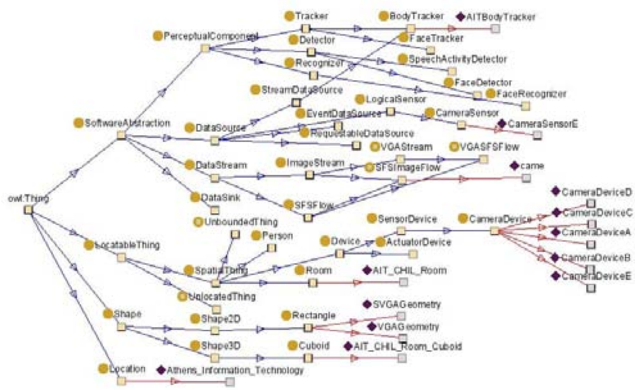


Figure 2. A fragment of the AIT CHIL Ontology demonstrating the relationship among entities

used by a higher level component. Moreover, employing the multi-inheritance features of the RDF technology, we are able to interconnect classes and properties which assist in the querying processes, as they "bind" together separate entries, for example the *Camera Device* which extends the *Sensor Device* class, which in turn extends the general *Device* class.

1.3. High-level reasoning

To manage this complicated ontology, we have developed a sophisticated tool, which we call Knowledge Base Server (KBS) [12]. The KBS is pending on incoming requests by any component-member of the pervasive service, processes the query (which can request information such as the IP and port of another component which it needs to connect to), searches the ontology for such entries, and adequately responds. The KBS maintains all information as all components, during startup, register themselves to the KBS.

To further boost the benefits of using semantic reasoning, we have developed mechanisms which can 'infer' logic from context. Using this high level reasoning approach we can more intelligently answer questions which can arise during the evolution of an actual event in the smart room. Examples of such intelligent queries can be 'Which camera is facing the door?', 'Which face tracking component is active in this session?' and 'Which perceptual components are available?'. The ability of the RDF technology which enables the inheritance of multiple entries was used in the case study we present in section 2.1. For example, a query "Which camera is mounted on the ceiling" would require the searching for all camera devices which have as a property "panoramic camera". Figure 2 shows a snapshot of a fragment of our ontology structure.

2. Applications

Using this framework for information management, we have developed and evaluated a series of applications which make heavy use of this ontological reasoning. In the following paragraphs we elaborate on our applications and illustrate how this mapping has assisted in their management, maintenance and deployment.

2.1. Perceptual Interfaces exchange

One of the attempts which proved the component collaboration and the benefits of using Semantic Web technologies to manage and coordinate our pervasive services, was an attempt to ‘exchange’ three perceptual components (i.e. a body tracker perceptual component which extracts peoples’ locations inside a smart room [15]). This exchange required the parallel operation of 3 components all developed at different laboratories using a variety of technologies, platforms and algorithms, but following the predefined architecture, mentioned in [18]. The sequence of operation was the following:

1. All sensor processing components register themselves as content providers to the ontology management system, using an XML-over-TCP schema, which contains an RDF interpretation of the entity itself, including information about the sensors.
2. All processing components register themselves to the ontology management system using an XML-over-TCP schema, using again an RDF representation.
3. All perceptual components query the ontology management system about the presence of sensor, which fulfill some criteria (e.g. are mounted on the ceiling, are facing the door, are at the North-West corner). The ontology management system responds by providing the requested information.
4. All perceptual components connect to the output stream of the sensor(s) they require.

It has to be mentioned that this approach would have been much more challenging without the assistance of the ontological reasoning as these perceptual components by nature, require the presence of specific type of sensor equipment. RDF enables this association between entries so that each component which requires a specific type of information can query the ontology for a more general entry type having a given property. In our case, the Body Trackers which were evaluated were:

- 2 2D Body Trackers which operated on a ceiling mounted camera,
- a 3D Body Tracker which operates on 4 corner cameras placed on the 4 corners of the smart room.

During this evaluation, the sensor processing components registered themselves in the ontology management system and ‘advertised’ the nature of the sensor which they control, e.g. Panoramic Camera, along with additional technical characteristics such as resolution, frame rate, etc. This information is crucial for the fine tuning of the context-acquisition components as they need to calibrate their internal parameters according to the configuration of the sensors. This information was provided from the ontology management system in the form of an RDF-formatted message which was decoded from the requesting components.

2.2. The Memory Jog

The Memory Jog service has been thoroughly described in our previous work [6], which demonstrated the multi-agent implementation providing a powerful framework for developing human-centric context-aware services. In the scope of this paper we will emphasize on the Semantic Web aspect of the application which covers the information management.

The Memory Jog service is a sophisticated service that encapsulates a set of features which are distributed and operate on the whole range of sensor and actuator infrastructure. This information is registered in the ontology, covering all information which can be used in the scope of the Memory Jog (people, software components, hardware components, sensors etc) leading to a large amount of information, eventually intractable. The RDF technology comes to provide a means of referencing each entry (class) and feature (property) with another entry of the ontology, easing the searching procedure.

The Memory Jog is dependent on the ontology description logic to determine which software components and hardware sensors are available during startup. It requires a set of information streams to be available, so that it can operate seamlessly and not cause operational gaps leading to user distraction. These queries are generated and request the location of components with specific capabilities, e.g. the presence of a 'Body Tracker', or the presence of a 'Face Identification' perceptual component. Moreover, other sub-services of the Memory Jog may as well request the presence of software components such as video recording [1] and thus the operation of camera sensor controllers. The use of the RDF technology has facilitated such queries as it enables the developer to make much more high level queries which could have been tedious to be organized in the cases of RDBMSs. For example, the 'Body Tracker' can be requested as a Software Component, or a Perceptual Component with the capabilities of a body tracker. Either way, the corresponding ontology management software can return a list of the currently available context providers, device controllers etc.

2.3. High-level context post querying

As the ubiquitous service is generating large amounts of contextual information, these are stored for later post processing. The services can extract high level information such as the time when a specific individual entered the room, etc. The RDF technology can significantly facilitate and speed up the querying processes as it can intelligently associate entries and properties. The user can by making a request retrieve all related information about the subject in question. For example, given a 'smart room', the user can at any point in time query 'Who entered the room and when?' or 'Did John enter the room after 10:00?'.

3. Evaluations

Our system has been evaluated in numerous occasions both at individual component evaluations but also in full system runs. Our attempt to exchange perceptual components, demonstrated the design and implementation independence that can be achieved when following such intelligent knowledge base directory services. Though the time-to-reply from the initiation of a request to the ontology management system until the receipt of the response was slightly higher than in the case of an RDBMS, the benefits of using semantic technologies for servicing intelligent requests overcome these issues which are partially due to the incorporation of the techniques for "inferring" context, techniques that are expected to introduce additional overhead.

4. Conclusions

In this paper we demonstrated the use of ontologies and high level reasoning in designing human-centric ubiquitous computing applications which operate on a highly dynamic and heterogeneous environment. This approach introduces a variety of benefits, not only for the system and service developer, but also for the service end-user as the later needs not to have detailed knowledge of the system specifications, only to provide the location of the knowledge repository and the reasoning interpreter to the end service.

Acknowledgements

This work has been partly sponsored by the General Secretariat of Research and Development of Greece, under the project PRIAMOS (GSRT Measure 3.3, on Audio-Video Processing, Project Code: 26).

References

[1] Azodolmolky, S., Dimakis, N., Mylonakis, V., Souretis, G., Soldatos, J., Pnevmatikakis, A., Polymenakos, L.: Middleware for in-door ambient intelligence: The polyomaton system. In: 4th IFIP TC-6 Networking Conference, 2nd Next Generation Networking Workshop(NGNM). Waterloo, Canada (2005)

[2] Berners-Lee, T., Hendler, J., Lassilla, O.: The semantic web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American **284**, 34–43 (2001)

[3] Brickley, D., Guha, R.V.: RDF vocabulary description language 1.0: RDF schema. W3c recommendation, RDF Core Working Group, World Wide Web Consortium (2004). URL http://www.w3.org/TR/rdf-schema

[4] Chen, H., Finin, T., Joshi, A.: An ontology for context-aware pervasive computing environments. Knowledge Engineering Review **18**(3), 197–207 (2003). DOI http://dx.doi.org/DOI:10.1017/S0269888904000025

[5] Chen, H., Tolia, S., Sayers, C., Finin, T., Joshi, A.: Creating context-aware software agents. In: First GSFC/JPL Workshop on Radical Agent Concepts (2001)

[6] Dimakis, N., Polymenakos, L., Soldatos, J.: Enhancing learning experiences through context-aware collaborative services: Software architecture and prototype system. In: Fourth IEEE International Workshop on Wireless, Mobile and Ubiquitous Technology in Education - (WMTE'06), vol. 0, pp. 50–57. IEEE Computer Society, Los Alamitos, CA, USA (2006). DOI http://doi.ieeecomputersociety.org/10.1109/WMTE.2006.12

[7] Ejigu, D., Scuturici, M., Brunie, L.: An ontology-based approach to context modeling and reasoning in pervasive computing. In: Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerComW'07), vol. 0, pp. 14–19. IEEE Computer Society, Los Alamitos, CA, USA (2007). DOI http://doi.ieeecomputersociety.org/10.1109/PERCOMW.2007.22

[8] Heflin, J.: Web ontology language (owl) use cases and requirements. Tech. rep., World Wide Web Consortium (2004). Recommendation REC-webont-req-20040210

[9] Manola, F., Miller, E.: RDF primer. Tech. rep., World Wide Web Consortium (W3C) (2004). URL http://www.w3.org/TR/rdf-primer/

[10] Mcguinness, D.L., van Harmelen, F.: OWL web ontology language overview (2004). URL http://www.w3.org/TR/2004/REC-owl-features-20040210

[11] OWL: W3C recommendation: OWL web ontology language overview. at http://www.w3.org/TR/owl-features/

[12] Paar, A., Reuter, J., Soldatos, J., Stamatis, K., Polymenakos, L.: A formally specified ontology management api as a registry for ubiquitous computing systems. In: 3rd IFIP Conference in Artificial Intelligence Applications and Innovations, pp. 137–147 (2006)

[13] Patel-Schneider, P.F., Hayes, P., Horrocks, I.: OWL web ontology language semantics and abstract syntax. W3C recommendation, World Wide Web Consortium (2004). URL http://www.w3.org/TR/2004/REC-owl-semantics-20040210/
[14] Pnevmatikakis, A., Polymenakos, L.: An automatic face detection and recognition system for video streams. In: 2nd Joint Workshop on Multi-Modal Interaction and Related Machine Learning Algorithms, Edinburgh (2005)
[15] Pnevmatikakis, A., Polymenakos, L.: 2d person tracking using kalman filtering and adaptive background learning in a feedback loop. In: CLEAR Workshop (2006)
[16] PRIAMOS: PRIAMOS project. at http://pyrros.cs.ntua.gr/priamos/
[17] Qasem, A., Heflin, J., Munoz-Avila, H.: Efficient source discovery and service composition for ubiquitous computing environments. In: Workshop on Semantic Web Technology for Mobile and Ubiquitous Applications (ISWC) (2004)
[18] Soldatos, J., Dimakis, N., Stamatis, K., Polymenakos, L.: A breadboard architecture for pervasive context-aware services in smart spaces: Middleware components and prototype applications. Personal and Ubiquitous Computing Journal **11**(2), 193–212 (2007). DOI http://dx.doi.org/10.1007/s00779-006-0102-7. URL http://www.springerlink.com/content/j14821834364128w/
[19] Song, Z., Masuoka, R., Agre, J., Labrou, Y.: Task computing for ubiquitous multimedia services. In: MUM '04: Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia, pp. 257–262. ACM Press, New York, NY, USA (2004). DOI http://doi.acm.org/10.1145/1052380.1052416
[20] Stanford Medical Informatics: The protege ontology editor and knowledge acquisition system website (2006). URL http://protege.stanford.edu/. Stanford Medical Informatics
[21] Weiser, M.: The computer for the 21st century. Scientific American **9**(9), 94–104 (1991)

A DIYD (Do It Yourself Design) e-commerce system for vehicle design based on Ontologies and 3D Visualization

Lambros MAKRIS, Nikolaos KARATZOULIS and Dimitrios TZOVARAS
Informatics & Telematics Institute, Greece[1]

Abstract. The customization level of vehicles is growing in order to deal with increasing user needs. Web browsers are becoming the focal point of vehicle customization, forming personalized market places where users can select and preview various setups. However the state of the art for the completion of the transaction is still very much characterized by a face-to-face sales situation. Direct sales over the internet, without sales person contacts, are still a small segment of the market, of only a few percent, for European manufacturers. This chapter presents an Intelligent DIY e-commerce system for vehicle design, based on Ontologies and 3D Visualization that aims at enabling a suitable representation of products with the most realistic possible visualization outcome in order to help prospective customers in their decision. The platform, designed for the vehicle sector, includes all the practicable electronic commerce variants and its on-line product configuration process is controlled by an ontology that was created using the OWL Web Ontology Language.

Keywords. DIYD system, Configurator, Ontologies, 3D Visualization

Introduction

Automotive enterprises are becoming more customer-centric to meet today's challenging market demands. This calls for restructured B2C relations and related new technologies. The automotive industry has furthermore become highly networked and requires improved communication on products and components in relation to its B2B relations.

Information and communications technology (ICT) can be used to support business and design activities. ICT does not change the fundamental goals of any organisation but makes it possible to optimize and coordinate design, manufacturing and marketing. In the automotive industry, ICT can:

- Improve design procedures;
- Allow optimization of design, manufacturing and marketing;
- Fine tune manufacturing processes;

[1] *Informatics & Telematics Institute, 1st Km Thermi-Panorama Road, PO Box 361, GR-57001, Thermi-Thessaloniki, Greece; Email: {lmak, nkaratz, tzovaras}@iti.gr*

- Provide the best product or service to current and potential customers; and
- Respond rapidly to customers' needs.

It has been envisioned that e-commerce and mass customization will emerge as a primary style of manufacturing in the coming decade and beyond. The integration of design, manufacturing, and logistics over the Internet will be the trend for the factory of the future. Effective supply chain management for mass customization will enhance profitability through a synergy of increasing customer-perceived value while reducing the costs of design, production and distribution. Companies successfully adapting to this new style of manufacturing will be able to reduce reliance on the traditional marketing channels, to gain more market shares globally, and to achieve high-efficiency product realization. This technology can enhance the established strengths of nowadays' industries in global manufacturing. It will benefit a wide variety of industries such as electronics, machinery, appliances, and logistics [1].

1. DIYD systems and the automotive business

DIYD (Do It Yourself Design) systems enable companies to extend their markets anywhere, anytime via the Internet. BMW, for example, sells six out of ten cars to order. Although the order-to-delivery time is very long, up to two months, much longer than that for regular cars, customers are prepared to wait [2]. However current systems usually only allow for a simple selection of options and a visual presentation of the result in the format of data sheets, tables and photographs of the vehicle exteriors and/or interior. The configuration system is not driven by customer requirements. The user has to interpret to what extent the various technical features will support his/her functional requirements and needs.

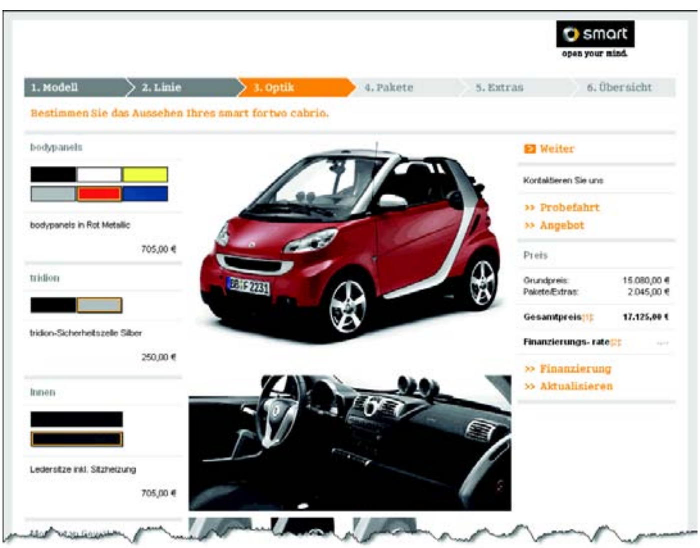


Figure 1. The online configuration system of the SMART fortwo car is considered state of the art in the field (http://www.smart.com).

In order to support the user in his/her choices, the information about the product should be presented in the most comprehensible way. This can be accomplished by offering the user detailed access to information using reconfigurable electronic catalogues and presenting the resulting configuration using 3D virtual prototypes.

1.1. e-Catalogues

Electronic catalogues are online and permanently up-to-date, and enable integration of content (i.e., access to product information) and direct communication between suppliers and customers. The most important characteristic of electronic catalogues lies in that they can be potentially integrated with other functions of the company and of its business partners, such as synchronization with product databases, communication with suppliers' ordering systems, and electronic payments of customer orders [5].

Like other e-commerce applications, electronic catalogues are still in their infancy. An electronic catalogue is not an electronic replica of a hardcopy catalogue. Rather, it involves characteristics of both the technology and related business practices. A common problem of current online sales is that searching for products on the Internet is always a cumbersome process. This is because the product information is unstructured and the sites are overloaded with information, with no navigation support [6]. Instead of translating directly the usage and representation patterns of hardcopy catalogues to the Internet, we should adopt a comprehensive and structured approach in organization and delivery of product information to enable customized and structured retrieval of information as well as navigation support.

In addition, prevailing solutions to online marketing of products usually focus on the ordering process and skip the negotiation process by providing only a direct link to a shopping basket. Under the online sales paradigm, electronic catalogues must support online configuration of products, resulting in large data volumes related to combinations of subassemblies and parts as well as configuration constraints. This raises the pressure on the managing capability of electronic catalogues. Furthermore, customization leads to a wide variety of configured products, in which a wide range of combinations of product features and design parameters may yield millions of variants for a single product. The traditional approach to variant handling is to treat every variant as a separate product by specifying a unique Bill-of-Materials (BOM) for each variant [7]. This works with a low number of variants but not when customers are granted a high degree of freedom for specifying products. Design and maintenance of such a large number of complex data structures with minimum data redundancy are difficult, if not impossible. Therefore, it is necessary to understand the implications of variety and to be able to deal with a large number of variants effectively.

1.2. Virtual Environments

The use of 3D virtual prototypes in a virtual environment can enhance visualisation and perspective viewing of the designed car. Unlike 2D graphics, users can interact via the web browser to navigate around an object and to move and rotate it. This type of Virtual Reality is much more flexible than a static image and allows for an apparently infinite number of different views on the vehicle. It is characterized by the use of 3D computer models presented on a 2D computer screen using 2D interaction devices like a mouse. The use of such interactive visuals has already undergone two cycles of hype in the internet business. But it was not successful due to overloaded solutions and bandwidth problems of the internet. Recently the technologies behind have gained momentum in the European automotive industry in the field of Digital Mock-ups [3], which are used very successfully in the development process employing again the internet for both intranet and B2B communication. Thus the time has come to transfer this success to field of mass customization and DIY design.

But Virtual Environments go beyond. Immersive systems allow for a dynamic stereoscopic view on the vehicle exterior and interior and add intuitive 3D interaction by tracking technology to the user interface. These sophisticated VR solutions are currently used in vehicle design and development. Internal projects in the automotive industry showed that VR technology was still too expensive to be used in customer communication and vehicle configuration around 5 years ago. This was mainly due to the use of expensive hardware. In the meantime inexpensive VR systems on commodity hardware basis have been developed in Europe, e.g. within the VIEW project [4]. These results can be further exploited to develop appropriate VR systems for mass customization and visualisation.

2. The CATER System

As already stated, current systems usually only allow for simple selections and 2D visual presentation. We present an intelligent and user-friendly e-commerce solution, namely CATER, by adopting additional technologies such as a configuration engine supported by ontologies, advanced search mechanisms, and 3D visualisation in a virtual reality environment. The focus of the system is on the vehicle industrial sector; however the intention is that the system will be suitable for suppliers, and wholesalers, from other sectors, such as furnishing, clothing etc.

In our use scenario a customer is connected to the CATER system using a traditional web browser. He searches in the 3D object database, by example, to find particular components that are of interest to him. The system, using an ontology, prevents him from selecting components which are incompatible. At the same time the user can pose particular constraints, such as maximum cost, which are honored by the system. He can then use a VR interface to connect the components together and form a design that suits him. The final selection can then be saved or forwarded to the factory for realization. Figure 2 presents the basic modules of the CATER system architecture.

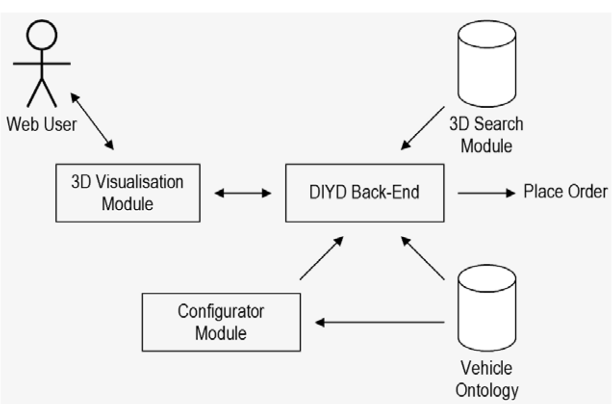


Figure 2. The main modules of the DIYD system

3. Intelligent Configurator module and Ontology

The Intelligent Configurator module in CATER is a web based application that allows the user to assembly vehicles based on the available vehicle parts that are being stored in the systems repository maintained by the vehicle manufacturer. Figure 3 displays the Units of the Configurator module.

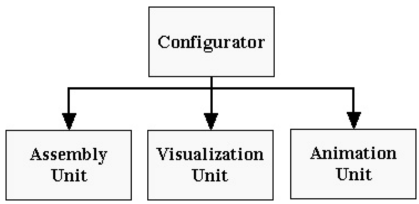


Figure 3. The Units of the Configurator Module

The Assembly Unit allows the user to insert individual 3D objects to the scene that can consist of a fully functional vehicle. The user can compose the desired vehicle according to his/her needs by selecting the vehicle's parts. The vehicle part and the texture selection processes are being controlled by the restriction mechanisms that are generated from the system Ontology [8][9][10]. The main functionalities of the Assembly Unit are the following: (i) Insertion of 3D object parts, (ii) Selection of texture and (iii) Assembly process based on rules (e.g. weight).

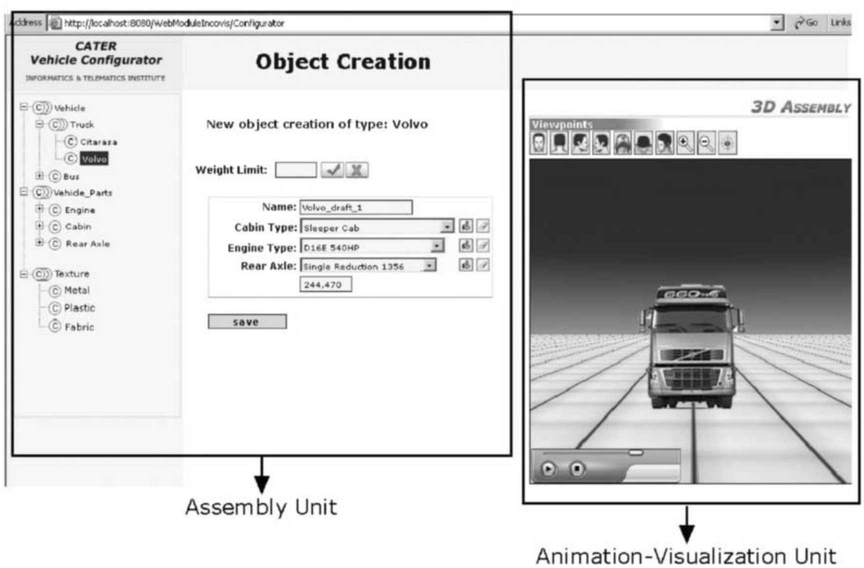


Figure 4. The web interface of the Configurator

Once the user has selected the preferred vehicle parts then the selected parts are loaded to the 3D scene and the user is allowed to modify the colouring scheme of each selected part by activating the textures menu option (Figure 4). The allowed colouring scheme is defined in the CATER ontology.

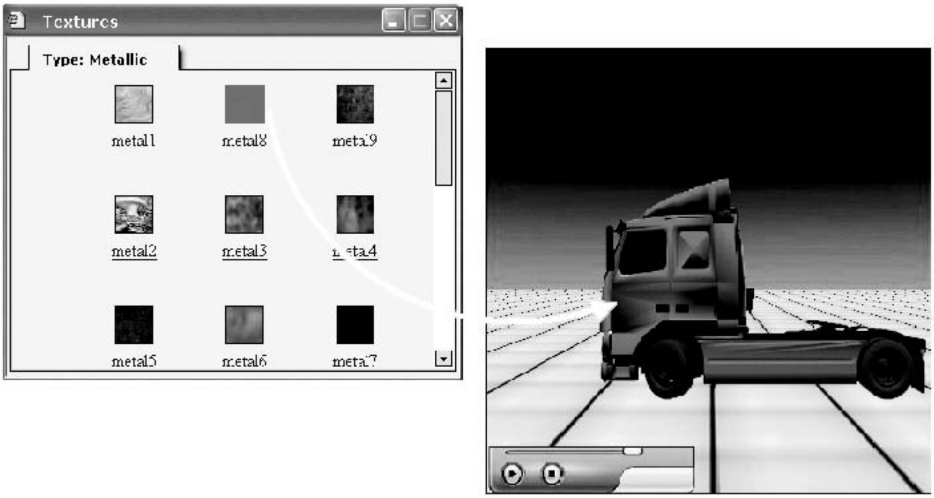


Figure 5. Modification of the colouring scheme of the inserted 3D parts based on the CATER ontology.

The purpose of the Visualization Unit is to record and store the 3D object assembly steps in real-time. The assembly sequence is being stored in the 3D animation repository for future reproduction. The Visualization Unit allows the user to select various viewpoints in order to preview the assembly process from various viewing angles (Figure 5).

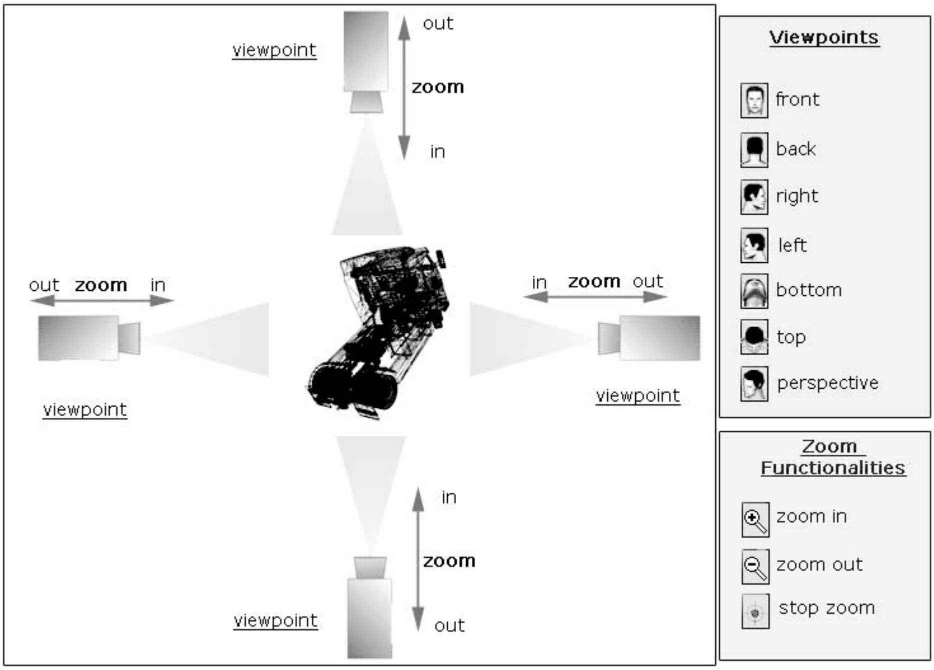


Figure 6. The viewpoints supported by the Visualization Unit.

The Animation Unit allows the reproduction of the vehicle parts assembly processes that are stored in the animation system repository. In the Animation Unit the user can control the viewpoints and the playback of the loaded vehicle assembly process. The web interface of the Configurator Module is depicted in Figure 4.

3.1. Configurator Architecture

The Configurator is implemented (Figure 7) using the Java programming language. The system runs on Apache Jakarta Tomcat [11] as a Java Servlet and it is based on the Jena framework [12], which is a Java framework for building Semantic Web applications.

The ontology is created using Protégé [13], which is an open source knowledge-base framework. The persistent store of the ontology is achieved using the persistence subsystem of Jena, while the 3D visualization was developed using the VRML [14] standard and External Authoring Interface (EAI) mechanisms.

3.2. Ontology Development

The ontology was created using the OWL Web Ontology Language [15], and the Protégé OWL-Plugin [16], which is an extension of Protégé with support for the Web Ontology Language (OWL).

The OWL-DL profile which was used in order to create the ontology, is based on Description Logics. Description Logics are a decidable fragment of First Order Logic2 and are therefore amenable to automated reasoning. It is therefore possible to automatically compute the classification hierarchy and check for inconsistencies in an ontology that conforms to OWL-DL [17].

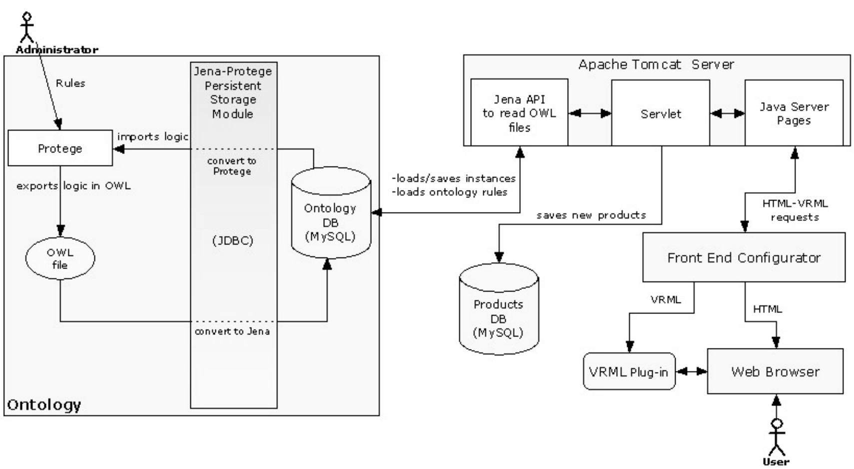
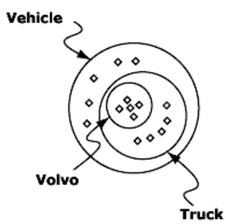
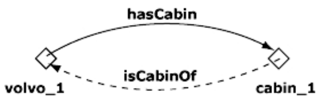


Figure 7. Configurator Architecture

3.2.1. Specification of Classes

The classes in the Ontology are interpreted as sets that contain individuals. They are described using formal descriptions that state precisely the requirements for membership of the class. For example, the class "Vehicle" contains all the individuals that are of type Vehicle in the CATER domain. The taxonomy of the classes is being achieved using the superclass-subclass model hierarchy.

Table 1. Example of the class hierarchy of the class "Vehicle" and an example of an Object Type Property for the individual "vehicle_1".

Class Model	Object Type Property
	

3.2.2. Specification of Properties

There are two types of properties supported by our ontology a) Data Type Properties and b) Object Type Properties. These OWL Properties represent relationships between two individuals.

In OWL, properties are used to create restrictions. In our ontology the restrictions were used to restrict the individuals that belong to a class. We used the universal quantifier ∀ restrictions to constrain the relationships along a given property to individuals that are members of a specific class. For example, the universal restriction ∀ hasCabin cabin_1 describes the individuals all of whose hasCabin relationships are members of the class Cabin.

Table 2. List of the property restrictions applied to the example class Volvo

Class: Volvo
NECESSARILY
Truck
hasCabinEngineType
hasEngineType
hasRearAxle
accessType write
INHERITED
Root
owl:Thing
Vehicle

Cardinality restrictions were used to define the order in which the individual object parts should appear during the 3D assembly process (i.e. real-time animation). The cardinality restrictions provided the way to describe the class of individuals that have at least, at most or exactly a specified number of relationships with other individuals or datatype values.

The hasValue restrictions, denoted by the symbol ∍, were used to describe the set of individuals that have at least one relationship along a specified property to a specific individual. For example, when we wanted to predefine the dimensions of an individual object part we used a hasValue restriction (dimensions ∍ "40-50-80").

3.3. Ontology Reasoning

In Jena, inference engines are structured as Graph combinators called Reasoners. An instance of a Reasoner combines one or more RDF Graphs and exposes the entailments from them as another RDF Graph in which some of the retrievable triples are virtual entailments rather than materialized data. The input Graphs contain both the ontology and instance data, with optional separation between the two. In particular, it is possible to partially-bind a Reasoner to an ontology and then use the resulting specialized Reasoner to access multiple different data Graphs – reusing the ontology inferences [18].

Ontology reasoning was achieved using the Jena OWL reasoner (Figure 8). The Jena OWL reasoner could be described as instance-based reasoner that works by using rules to propagate the if- and only-if- implications of the OWL constructs on instance data. Reasoning about classes is done indirectly - for each declared class a prototypical instance is created and elaborated. The sub-class and sub-property lattices are cached using the embedded OWL reasoner. Each domain, range, sub-property and sub-class declaration is eagerly translated into a single query rewrite rule. The result of a query to the graph will be the union of the results from applying the query plus all the rewritten versions of the query to the underlying graph [18].

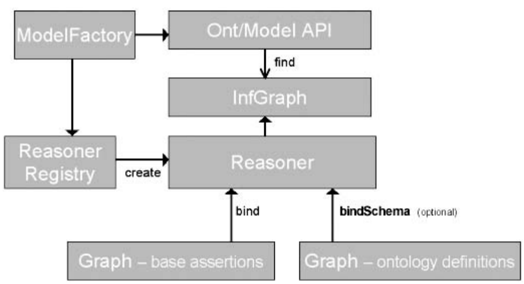


Figure 8. The Jena Inference API layering [19]

The ontology model for handling the OWL ontology was created using the Jena API:

OntModelSpec.OWL_MEM_RDFS_INF

This choice adds a rule-based reasoner that will add the entailments from the source data using the semantic rules of RDFS only. This includes entailments from subclass and sub-property hierarchies, and domain and range constraints, but not, for example, entailments from the disjoint-ness of classes.

3.3.1. Extending the reasoning process by using rules

We are planning to extend the current inference model by adopting a Hybrid rule engine that is supported by the latest version of the Jena API [19]. The rule reasoner has the option of employing the Forward (RETE) individual rule engine and the Backward (LP) engine in conjunction. In the hybrid mode the data flows as presented in Figure 9.

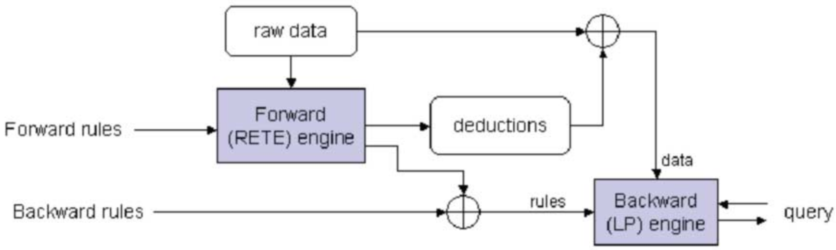


Figure 9. The Hybrid Rule engine supported by the Jena API.

The forward engine runs, and maintains a set of inferred statements in the deductions store. Any forward rules which assert new backward rules will instantiate those rules according to the forward variable bindings and pass the instantiated rules on to the backward engine.

Queries are answered by using the backward chaining LP engine, employing the merge of the supplied and generated rules applied to the merge of the raw and deduced data.

This split allows us to achieve greater performance by only including backward rules which are relevant to the dataset at hand. In particular, the forward rules can be used to compile a set of backward rules from the ontology information in the dataset.

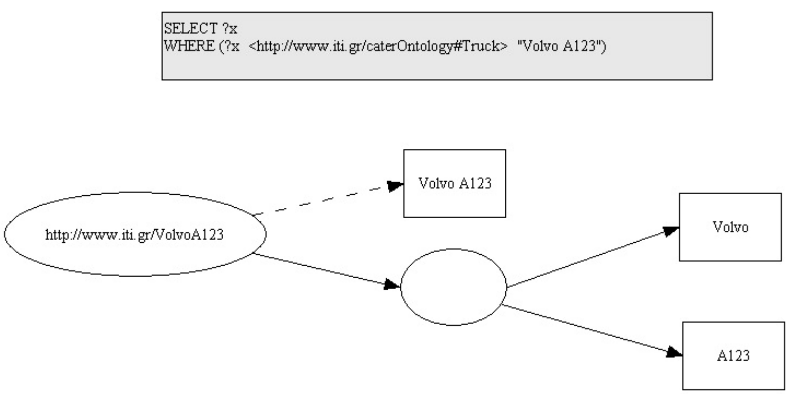


Figure 10. Graphic representation of an example RDQL query used for the CATER ontology.

3.4. RDQL – Data oriented query model for the CATER Ontology

The CATER ontology uses an implementation of the RDQL query language for quering RDF models using the Jena API (Figure 10). RDQL provides a data-oriented query model so that there is a more declarative approach to complement the fine-grained, procedural Jena API.

RDQL queries only the information held in the models; there is no inference being done. The RDQL system receives the description of what the application requests, in the form of a query, and returns that information, in the form of a set of bindings.

RDQL is an implementation of the SquishQL RDF query language, which itself is derived from rdfDB. This class of query languages regards RDF as triple data, without schema or ontology information unless explicitly included in the RDF source [12].

RDF provides a graph with directed edges - the nodes are resources or literals. RDQL provides a way of specifying a graph pattern that is matched against the graph to yield a set of matches. It returns a list of bindings - each binding is a set of name-value pairs for the values of the variables. All variables are bound (there is no disjunction in the query).

4. 3D Visualisation Module

Virtual Reality (VR) interfaces can provide the most realistic presentation of a configuration for end users. It combines high quality visualization with the correct perception of depth and scale, which enables a feeling for the roominess of the interior of a car. Additionally, a highly intuitive interface allows easy manipulations of models, assemblies and parts. Combined with simulations (e.g. physics or a man-model), the models can be evaluated in terms of packaging (fit of the components) of ergonomics.

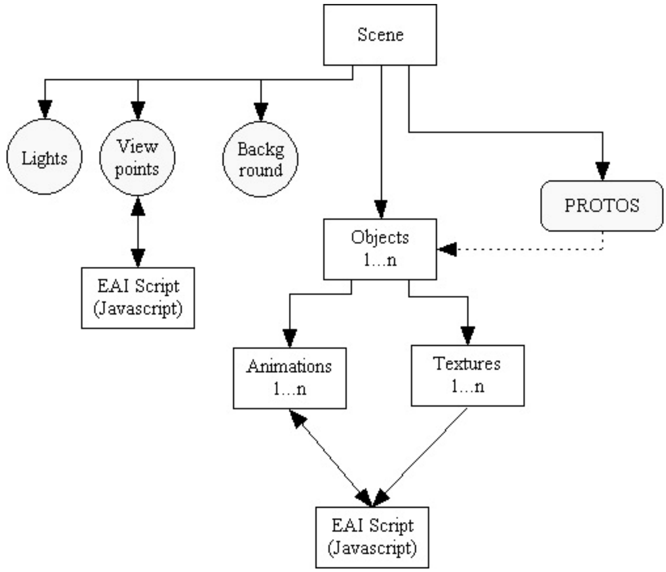


Figure 11. The 3D scene structure of the 3D Visualization Module

The 3D Visualization Module of the CATER system is realized on the Visualization and Animation Units. The structure of the individual 3D scenes supported by the 3D Visualization Module can be seen in Figure 11.

The 3D scene contains the viewpoints, the lighting of the 3D world, the background, the 3D objects and the object functionalities (interactions) that are created dynamically according to the ontology specifications. The user interaction with the 3D scene is achieved by the use of predefined VRML Protos. For every 3D object that is inserted in the 3D scene an animation representing its assembly process is dynamically generated.

The playback functionalities of the assembly process are controlled by a panel (Figure 12) that was developed using several VRML sensors (TouchSensor & PlaneSensor).

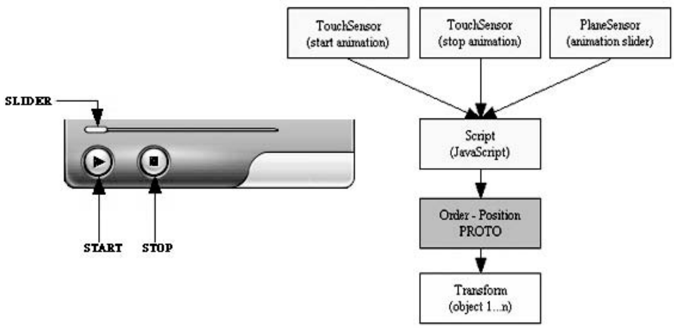


Figure 12. The assembly process control

5. 3D Geometry Search Module

The 3D geometry search tool utilizes novel algorithms for low-level feature extraction from 3D objects, based on geometric characteristics. The algorithms are robust to affine transformations (rotation, translation, scaling) and are applied to 3D objects regardless of their degeneracies, formats and levels of detail. This results in a more compact representation of the objects, which uniquely characterizes them. The 3D search tool can be used whenever the user wants to provide a specific part of a vehicle as a query and retrieve similar objects from the repository.

Every 3D object is described with a rotation, scaling and translation invariant descriptor vector, which is formed according to the Spherical Trace Transform (STT) [20]. Initially, every object is translated and scaled so as all objects are expressed in the same coordinate system. To achieve the latter, a local coordinate system has been defined centred to the mass centre of the object and scaled so as the object fits to the unit sphere. Then, the object's binary volumetric function is computed

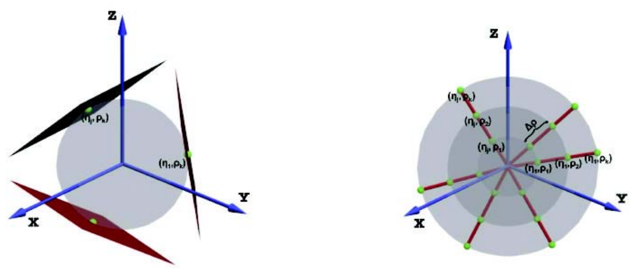
$$f(\mathbf{x}) = \begin{cases} 1 & \text{when } \mathbf{x} \text{ lies within the 3D model's volume,} \\ 0 & \text{otherwise.} \end{cases}$$

and the STT (Figure 13) is performed as follows:

- A set of radius segments is defined. Every radius segment Λ_i is formed by the intersection of a radius (η_i, ρ_j) with the object.
- A set of spheres, concentric to the unit sphere is defined.
- A set of plane segments is defined for every sphere S_r. Every plane segment Π_i^r is formed by the intersection of a plane tangent to the sphere at point P_i with the object.

The points P_i and radii r_i are uniformly distributed on the sphere's surface exploiting the icosahedric-based tessellation.

Every Λ_i segment is treated as a one-dimensional signal and descriptors based on classical 1D Discrete Fourier Transform and an integration transform are computed. Every plane segment Π_i^r is treated as a 2D signal and descriptors based on the Krawtchouk, the Zernike and the Hu moments, the polar wavelet transform and the 3D Radial Integration Transform [21] are computed. Then, the spherical Fourier Transform is applied on the extracted separately on every descriptor, so as the final descriptors are invariant under rotation and, thus, appropriate for 3D object matching.



(a) (b)

Figure 13. The Spherical Trace Transform

The matching procedure is based on a mixture of the weighted Minkowski L1 distance and the normalized distance.

$$L1_n^i = \sqrt{\sum_{k=1}^{S} | W_{nk} D^Q(k) - W_{Ck} D^{A^i}(k) |},$$

Equation 1. Weighted Minkowski L1 distance, where DT (k) is the k-th element of object T descriptor vector and Wi,T is the assigned weight

The computation of the weights for every single descriptor is based on the statistical behaviour of every descriptor for every class (e.g. mean value, standard deviation, etc). Two different methods for weight assignment have been proposed. The method has been tested on the Princeton Shape Benchmark. Figure 14 depicts the results in terms of precision and recall diagrams.

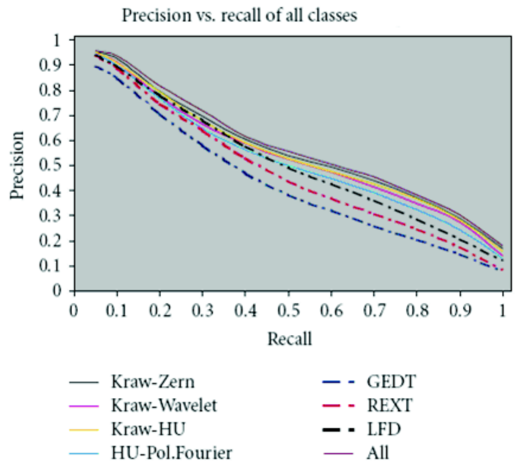


Figure 14. Efficiency of the STT using combination of different descriptors (Krawtchouk, Zernike, Hu etc).

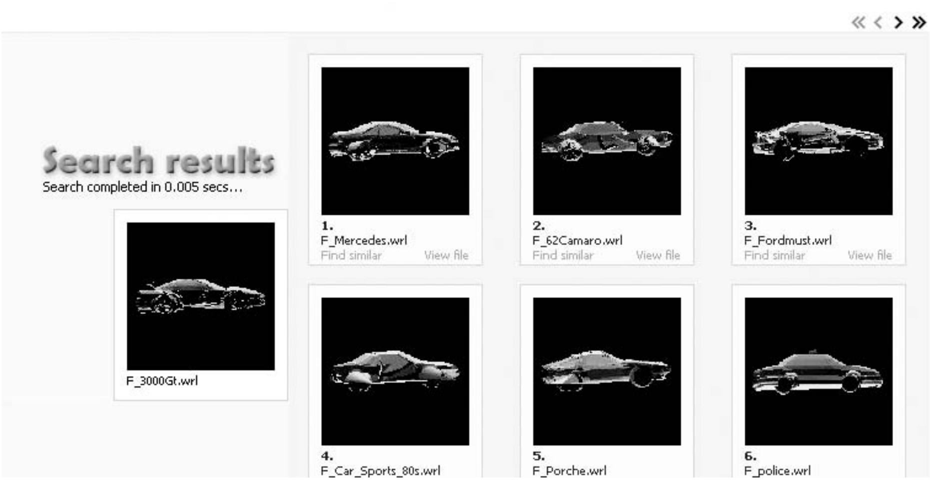


Figure 15. Example 3d search using sample data.

6. The e-shopping Platform in Practice

The use of the CATER platform "brings" advantages for both suppliers and buyers regarding (i) the cutback of transaction costs, (ii) the use of automated supply procedures, (iii) economy of scale, (iv) wide access on both local and international markets, (v) dynamic real-time price mechanisms/modules and (vi) the use of compatible/expandable technologies.

The requirements of the described CATER platform for vehicle products that together with the Intelligent Configurator Module and the 3D Visualisation Module comprises the advanced 3D Shop system are:

- search and present all the available products, based on multi-criteria search engines
- group products into multilevel categories (set by the e-shop administrator)
- make offers/ sales and promote them
- update both the product catalogue and all items' availability (set by the e-shop administrator)
- create/use shop baskets (by the end buyers)
- provide several convenient pay/receive methods
- provide a secure e-payment credit card transaction (with the use of HTTPS and SSL protocols).

However, the efficiency and overall quality of an e-commerce service depends "heavily" on its automatic connection with the existing ERP (Enterprise Resource Planning) system for the catalogue, prices, stock and product update. In order to integrate all the available ERP data with the e-shop database, a powerful staging mechanism is developed and securely transfers all necessary data. This staging process uses a smart "track changes" algorithm, to enhance the update speed.

There are two staging processes, Real Time Staging and Off Line Staging (that uses an automated batch process). The characteristics of the two staging "methods" are compared in the following table.

Table 3. Staging Procedures Comparison

	Real Time Staging	**Off Line Staging**
Data Update	(+) All data are updated at all times	(-) All data are updated at specifically defined time periods
Infrastructure	(-) Reliable, high speed, technical infrastructure is necessary, available on a 24x7x365 basis	(+) Not so advanced technical infrastructure is necessary
Security	(-) The system can be secure but certain "protective" actions must be taken	(+) Security is obvious

The previous table shows that a real time staging procedure should be followed only if the nature of the commodity traded imposes the constant database update. In our case an every day off line procedure is chosen for both security and convenience reasons.

Yet, if we try to deduct a general case example we must notice that each company's and product's needs, concerning the use of an e-market, are different; therefore the connectivity solutions (between an e-shop and an ERP) provided vary depending on: (i) the ERP used (it can be a widely used international ERP such as SAP, Oracle Applications, etc. or it can be a custom made system that fits to specific needs), (ii) the transaction volume and the form of the data transferred, (iii) the importance of the information transferred (regarding time, safety etc. aspects), (iv) the use of unilateral or bilateral communication and (v) whether it is an on-line or a batch transfer of data.

7. Conclusions

E-commerce services offered through a B2C (business to consumer) or B2B (business to business) system, provide the necessary infrastructure for real time e-business and an added value package of services that guarantee faster and more efficient buy and sell transactions, access to a broadened database of buyers/suppliers and business opportunities through the development of new partnerships.

In conclusion, in this chapter we presented an interactive and user-friendly e-commerce solution for the vehicle sector, but appropriate for other sectors as well. Volvo Technology Corporation (VTEC) has been the end-user responsible for using and testing the CATER platform, so a number of its vehicles were integrated in the platform for evaluation and testing purposes.

Finally, the main contribution is that our approach adopts additional technologies such as a configuration engine supported by ontologies, advanced search mechanisms, and 3D visualisation in a virtual reality environment aiming at enabling a suitable representation of products in order to achieve the most realistic possible visualization and simulate an up to close shopping procedure.

Acknowledgements. We wish to acknowledge the CATER project Consortium for their valuable contributions to this work. The CATER project is partially funded by the EC.

References

[1] J. Jiao and M. G. Helander. Development of an electronic configure-to-order platform for customized product development. Computers in Industry 57 (3), 231-244. Apr. 2006.
[2] Research Tools Survey: Fighting Back, The Economist. September 2, 2004.
[3] D. Döllner and G. Kellner. Digital Mock-up in Rapid Prototyping in Automotive Product Development. Journal of Integrated Design and Process Science 4 (1), 56-66. Mar. 2000.
[4] J. Wilson and M. D'Cruz. Virtual and interactive environments for work of the future. International Journal of Human-Computer Studies 64 (3), 158-169. Mar. 2006.
[5] A. Segev, D. Wan, and C. Beam. Designing Electronic Catalogs for Business Value: Results of the CommerceNet Pilot. CITM Working Paper CITM-WP-1005. Oct. 1995.
[6] K. Stanoevska-Slabeva and B. Schmid. Internet Electronic Product Catalogs: An Approach Beyond Simple Keywords and Multimedia. Computer Networks. 32 (6), 701-715. May 2000.
[7] J. Jiao, M. M. Tseng, Q. Ma, and Y. Zou. Generic Bill of Materials and Operations for High-Variety Production Management, Concurrent Engineering: Research and Application, 8 (4), 297-322. Dec. 2000.
[8] I. Kompatsiaris, V. Mezaris, and M. G. Strintzis, Multimedia content indexing and retrieval using an object ontology, Multimedia Content and the Semantic Web: Methods, Standards and Tools, G. Stamou and S. Kollias (Eds), ISBN 0-470-85753-6, Wiley, pp. 339-371. May 2005.
[9] I. Tsampoulatidis, G. Nikolakis, D. Tzovaras and M. G. Strintzis, Ontology Based Interactive Graphic Environment for Product Presentation, in Proc. CGI 2004, pp.644 - 647, Heraklion, Crete, Greece. June 2004.
[10] V. Mezaris, I. Kompatsiaris, M. G. Strintzis, An Ontology Approach to Object-Based Image Retrieval, Proc. IEEE International Conference on Image Processing (ICIP 2003), Barcelona, Spain, vol. II, pp. 511-514. Sep. 2003.
[11] Apache Jakarta; http://jakarta.apache.org/.
[12] Jena, A Semantic Web Framework for Java; http://jena.sourceforge.net/.
[13] Protégé, An Ontology Editor and Knowledge-base Framework; http://protege.stanford.edu.
[14] Web 3D Consortium, VRML Standard; http://www.web3d.org/x3d/vrml/.
[15] The OWL Web Ontology Language; http://www.w3.org/TR/owl-features/.
[16] Protégé OWL Plugin; http://protege.stanford.edu/plugins/owl/.
[17] M. Horridge, H. Knublauch, A. Rector, R. Stevens, and C. Wroe, A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools, Edition 1.0, The University of Manchester. Aug. 2004.

[18] J. Carroll, I. Dickinson, C. Dollin, D. Reynolds, A. Seaborne, and K. Wilkinson, Jena: Implementing the Semantic Web Recommendations, Digital Media Systems Laboratory, HP Laboratories Bristol. 2003.
[19] Jena 2 Inference Support; http://jena.sourceforge.net/inference/.
[20] D. Zarpalas, P. Daras, A. Axenopoulos, D. Tzovaras, and M. G. Strintzis. 3D Model Search and Retrieval Using the Spherical Trace Transform. EURASIP Journal on Advances in Signal Processing, vol. 2007, Article ID 23912, 14 pages, 2007. doi:10.1155/2007/23912
[21] P. Daras, D. Zarpalas, D. Tzovaras, and M. G. Strintzis. Efficient 3-D model search and retrieval using generalized 3-D radon transforms. IEEE Transactions on Multimedia, Volume 8, No 1, pp. 101-114, February 2006.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Semantics Enabled Problem Based Brokering of Organizational Knowledge

K. Kafentzis[1], M. Wallace[2], P. Georgolios[1], P. Alexopoulos[1], G. Mentzas[3]
[1]*IMC Research, Athens, Greece, {kkafentzis , pgeorgolios, palexopoulos}@imc.com.gr.*
[2]*University of Indianapolis, Athens, Greece, wallace@uindy.gr*
[3]*National Technical University of Athens, Athens, Greece, gmentzas@mail.ntua.gr*

Abstract. Knowledge deriving from owned information and experience is an asset that has begun to be recognised by organizations of various scales as a marketable product. In previous work we have developed a system that facilitated the formal description of knowledge possessed by an organization, so that external entities could search in it and request it. In this chapter we extend on that work in a couple of ways: i) we develop a mediator system enabling the search to concurrently consider multiple possible sources of information and ii) we allow for the query to posed on a more natural way by expressing the information needs of the requestor rather than by describing the information items that may satisfy these needs. Such a system opens the way for fully automated problem based online information brokering, which is expected to be the next trend in the knowledge market.

Keywords. Knowledge commerce, problem domain, mediator service, information brokering.

Introduction

Managers specialize on the analysis of data in order to reach strategic decisions, thus affecting the future of the companies they run. Similarly, decision support systems focus on the intelligent processing of information in order to provide optimized recommendations regarding the required steps to achieve a sequence of set goals. In both cases, the amount and quality of available data sets the upper bound for the effectiveness of the reached decisions or recommendations.

Having realized this, companies are now starting to also realize the need to efficiently manage the knowledge they may possess in different forms, so that it may be readily and easily accessible when needed. In this framework, theory and systems have been developed to facilitate the formalization of organizational knowledge and information resources (Tiwana, 2000; Davies et al., 2002; Mentzas et al., 2002). Using those, knowledge gethered in an organization, either via the recruiting of domain experts or through practical experience, can be packaged in a system as a searchable and retrievable knowledge item.

Taking a step further, such knowledge items may also be shared between different cooperating organizations, or provided in the form of information products in the emmerging knowledge commerce marketplaces, enabling organizations to share or commercially exploit their knowledge outside narrow organizational borders (Skyrme, 2001; Muller et al., 2002).

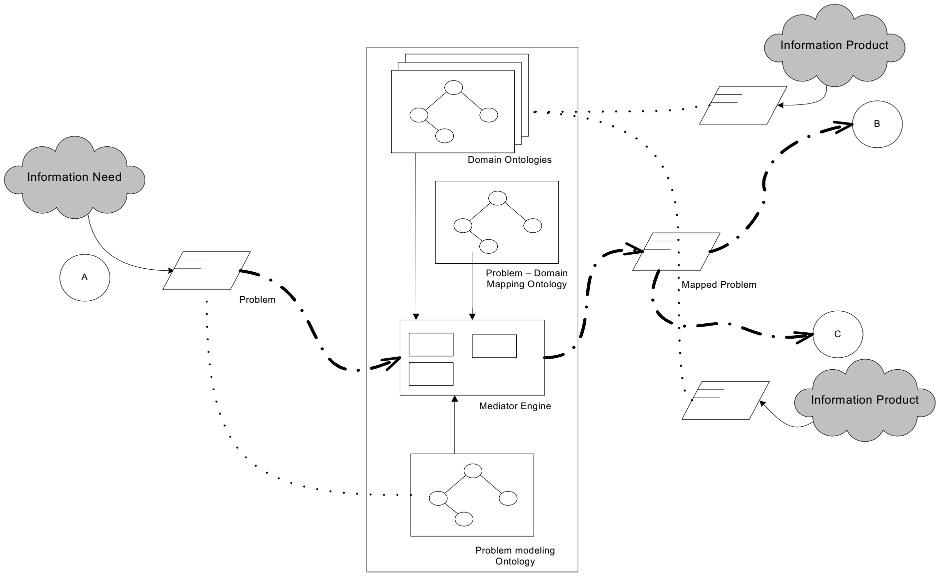


Figure 1. The overall architecture

Of course, one may easily realize how a knowledge market place differs fundamentally from any other kind of market: where in the traditional case goods are best marketed through detailed presentations and explanations, in the case of knowledge a mere display of the product renders its value to null; no one would care to pay for access to information they have already seen. In previous work we have provided a solution to this problem, by utilizing domain and service ontologies to describe the content of different knowledge items (IKASS IST Project). In this manner knowledge items can be searched and evaluated by potential buyers without being exposed and thus rendered commercially worthless.

In this chapter we attempt to make a two-fold extension to this framework. Firstly, we utilize a mediated architecture that allows searches to be concurrently contacted on repositories of multiple information providers, in order to alleviate the un-natural requirement that the buyers know beforehand which organization holds the information they need. Secondly, we perform problem to domain mapping in order to alleviate the also un-natural requirement that buyers know beforehand the complete list of information bits that are useful for their task at hand.

In the figure we outline the main components of the proposed architecture. In the section that follows we briefly present the modelling approach followed for the different abstract information components involved, so that in section 3 we can continue to explain the ways in which our mediator engine offers novel and enhanced knowledge item retrieval services. In section 4 we identify areas for further research and in section 5 we list our concluding remarks.

1. Information Modeling

Seen from a broader point of view, the overall purpose of our system is to provide a meaningful match between information needs of organization A and information products offered by organizations B, C and so on. Clearly, the very first step required in order for such a system to operate in an automated manner is the formalization of the representation of these two abstract types of information. In the following subsections we outline the formal models used for this purpose.

1.1. Knowledge product modelling

Information owned and intended for commercial exploitation by an organization can only be searched by an external party if its representation follows a known and agreed standard. Following the current trend, we utilize ontologies to represent the semantics of available information products. Our system utilizes domain-specific ontologies that specify domain dependent semantics.

These allow for discrimination and classification of the knowledge objects that our knowledge service will provide access to. Domain ontologies are typically internal elements of the mediator, i.e. they reside within the mediator's knowledge base, but the option to reference external ontological sources is also available.

A strictly internal ontological representation is utilised for the description of attributes such as the kind of content available in a knowledge repository and its physical manifestation; this is the content type ontology. In the current version of the system this information is used strictly for browsing purposes, i.e. it is displayed to the user for knowledge products selected for the posed query, but in planned extensions, as also mentioned in section 4, this kind of information will also be considered in the retrieval, ranking and contracting procedures.

Further information on the ontological modeling of available knowledge products is available via the INKASS project, as ontologies were used to model offered information even in the non-mediated version of the system.

1.2. Problem modeling

An issue that arises when trying to practically apply the above is that managers within organization A cannot possibly know beforehand what kind of information may be available within organizations B and C. Therefore, they are unable to form the proper queries that will retrieve all the potentially useful knowledge items for their task in hand. In other words, our system's end users do not know the answer beforehand, but they do know the problem.

For this reason we introduce querying with the application context. In simple words, instead of forcing end users to utilize domain modelling ontologies to formulate queries, we allow them to utilize a context/problem modelling ontology for this task in figure two we see a snapshot of a section of the contextual modelling information in the mediator system. In the following we see the formalized query posed by a manager interested in starting up a new construction company in China in face of the 2008 Olympics.

We can see that the manager had only to specify the application context for which information is needed, not the information attributes themselves. This is a feature that is not available in INKASS or in other knowledge services in the literature.

```
<problem>
<problem_description>
    Start up a construction company in China and looking for legal advice
```

```
</problem_description>
<required_action>
    Establish company
</required_action>
<application_context>
    Civil and Structural Engineering
</application_context>
<application_context>
    Legislation
</application_context>
<geographic_area>
    China
</geographic_area>
</problem>
```

Code segment 1: A sample problem formulation based on the problem modeling ontology.

Mediation

One of the main tasks of mediators is to fuse information from heterogeneous information sources (Papakonstantinou et al., 1996); in our work, the fusion of the different kinds of information available in the different participating knowledge repositories is achieved via the utilization of domain ontologies for the modelling of their content. But this is not the only role that the mediator has in our system.

Although mediators were originally applicable mainly in the context of information and multimedia retrieval, their applicability in the context of E-commerce soon became apparent (Pan et al., 2002). In this field of application, additionally to their ability to fuse heterogeneous information sources, mediators also demonstrate their ability to operate as MLSs (Multiple Listing Systems), offering unified access to multiple sources of goods. This is a feature that is also desired in the emerging field of the knowledge market; rather than keeping a record of all organizations that possess and offer information services, information requestors will only have to contact a single centralized system, which will be in charge of contacting sources of information, retrieving relevant items from them and repackaging them as a single response.

The main issue to consider here is that the information need of organization A is expressed using the problem modelling ontology whereas the content of the information products offered by organizations B and C are expressed using domain and content type modelling ontologies. Thus, one of the operations of the mediator engine of the proposed system, and also one of the main contributions of this work, is the automated transformation of the context query to a content query.

```
<mapped_problem>
<domain>
    Chinese Investment Legislation
<subdomain>
    Chinese company foundation procedures
</subdomain>
<subdomain>
    Compliance with chinese regulations
</subdomain>
```

```
<subdomain>
    Construction Sector
</subdomain>
<subdomain>
    Industrial Activities
</subdomain>
</domain>
<content_type>
    Expert
 <expertise>
    Law
</expertise>
 </content_type>
 <content_type>
    Guidelines
</content_type>
 </mapped_problem>
```

Code segment 2: The problem's query expressed based on the domain modelling ontology.

For this purpose a problem to domain mapping ontology has been developed. In this ontology, elements of the contextual ontology are mapped to elements of the domain ontologies. The problem mapping subsystem of the mediator engine is a reasoning component that starts by independently looking up the various components of the problem in the mapping ontology and then continues by combining findings so that only the truly relevant entries are kept. In this manner, for example, the problem formulation of Code segment 1 is transformed to the query to information repositories presented in Code segment 2.

As we explain in the following, all this processing takes place in an online mode of operation and thus time complexity of the involved procedures is important. On the other hand, as experience has indicated, having thorough ontological representations with many levels of detail makes all reasoning attempts, even those of the simpler kinds, extremely time consuming due to the recursive nature of the procedures involved. Of course, due to the financial worth and vast variety of organizational information, thorough ontological representations are a requirement. Therefore, a recursion alleviation methodology is needed.

Towards this goal, we follow a transitive closure methodology (Wallace et al., 2006b). This methodology allows us to transform our ontological descriptions in a way that they can be processed with one pass algorithms rather than with recursive searches; such algorithms have also been developed within previous work and have been adopted to the work presented herein (Wallace et al., 2006a).

From a systems' point of view, the mediator would also have to incorporate distinct components for the interaction with each one of the information sharing organizations. Again, following the current trend, we alleviate this need by utilizing OWL-S (former DAML-S) in order to employ semantic web technology for service description (Turner et al., 2004). This allows the mediator to contain a single general purpose subsystem that is able to interface to semantically annotated web services.

As far as stability is concerned, the mediator is designed to send parallel queries to all known knowledge repositories and collect responses received within a predefined period of time. After that time the response is compiled and forwarded to the information requestor. In this way, if a specific source of information is unavailable or overloaded this will not result in the overall system also being unable to provide results or overly slow.

An alternative approach would be to allow the mediator system to perform offline processing if information available in repositories (crawling) so that contextual queries can be answered on the fly without contacting the information providers, much like Google. Our strategic decision was not to follow this approach as the value and quality of knowledge is greatly affected by timing. Thus, it is essential that our system always processes information in an online mode and prepares its results based on current content and availability.

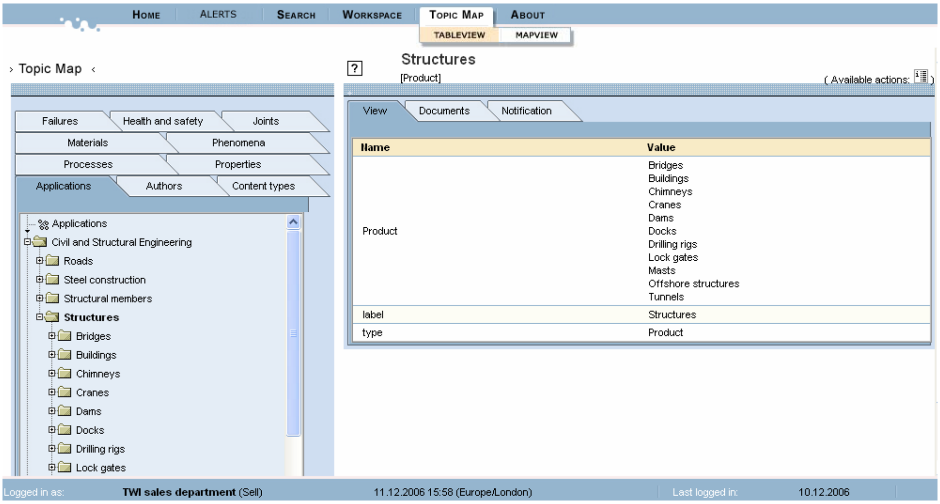


Figure 2. The query interface

More trivial subsystems (in the sense that similar systems have already been described in the literature in other works) include the result ranking and ordering system and the web service offered as interface to information requesting entities such as organization A of our running example.

2. Further extensions

From a more pragmatic point of view, although a concrete mediation service has been added to the original INKASS system and contextual semantics allow for the queries to be posed by managers who are not entirely familiar with the internal workings of the information offering entities, the proposed system is still not ready to become a standalone and fully automated organizational knowledge brokering system. It is rather more suitable as a tool for internal organizational information structuring, management and retrieval, for organizations whose large scale and distributed nature demand it.

In order for such a system to be utilized in an automated manner for inter-organizational information exchange, some additional issues will need to be considered and tackled:

First of all, in the context of a knowledge market, information items will be associated with a financial (or other) cost. Therefore, a mediator that merely locates and retrieves all information items that are relevant to a presented problem is far from adequate. What is needed is a component that is able to automatically assess which combinations of knowledge products provide adequate information for the information requestor to be able to address the problem in question. Thus, it is not enough to have an ontology that maps application contexts to relevant application domains; we also need an ontology that describes which kinds of information items are hard requirements as well as which kinds of information items may be interchangeable.

Automated evaluation of pricing with regards to the application context is also a required component. It is expected that knowledge products will be priced based on how rare they are, how hard they are to get, how sensitive and potentially damaging information they incorporate and how important their possible application domains are. On the other hand, information requestors are only interested in what a specific knowledge product can do for them and for their specific problem in hand. Thus, the mediator needs to have a subsystem that is able to assess not only the whether a knowledge product is relevant for the problem in hand, but also the degree to which the specific knowledge product would be useful to tackle the given problem, as this would provide the basis for the evaluation of the product with regards to its price.

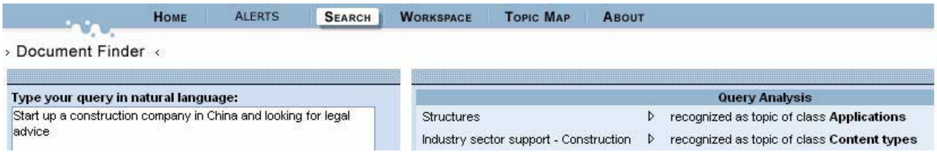


Figure 3. The results interface

So far we have made the silent assumption that cost is the sole criterion for the selection of knowledge items. In a more realistic setting, additional criteria, such as physical manifestation and time of delivery can also be important. The consideration of such additional criteria generates the need for a multiple criteria decision making module that will select among the different Pareto efficient sets of knowledge products the ones that best match the requestors contracting preferences.

Finally, the management of digital rights is certainly an important issue to consider (Delgado et al., 2002). Similarly to the case of the extensive annotation of multimedia documents, the semantic description of the knowledge products is a valuable asset itself that cannot be made available to everyone without any control. In this respect there are two distinct approaches that may be followed: that of the trusted entity mediator, where the mediator is given full access to all marketable information and the mediator is in charge of enforcing digital right management agreements in the way this information is exposed and that of the incremental exposure where information exposure and contracting take place in a sequential and incremental manner. Digital signing and automated contracting is also a relevant field (Langendorfer, 2003; Angelov and Grefen, 2004).

3. Conclusions

In this chapter we have extended on our previous work in knowledge sharing by developing a system able to tackle important usability issues observed in the original system. In the INKASS prototype, knowledge exchange was facilitated with the utilization of domain ontologies. In this way, queries could be posed at a semantic level and links to relevant knowledge products could be retrieved.

In this work we have added a mediation step, thus making the overall system an MLS that is able to provide concurrent, unified access to multiple knowledge repositories. What is more important, we have developed a contextual ontology, using which information requestors can query the mediator by describing the context of application of the information they are requesting, rather than the information itself. This alleviates problems related to the open world structure of the system which makes it impossible to information requestors to manually specify the characteristics of all available information items that may be potentially useful to them.

The problem based query is then processed by a subsystem of the mediator so that it can be mapped to the domain ontologies. Through this mapping, a new query is formulated which can be used to directly query the linked information offering organizations. A decision has been made not to keep indexed records of information repository content centrally, so that knowledge product selection is always based on current information.

Finally, a lot more needs to be done for this technology to reach the level required for a global knowledge market to operate in a fully automated manner. We have outlined what needs to be done in an as concise and complete way as possible at this stage and we have started to work towards this goal.

4. Acknowledgements

The authors are grateful to the end-user partners of INKASS for the constructive feedback they provided throughout the project. The INKASS project was funded by the European]Community under the “Information Society Technology” Programme (1998-2002).

References

[1] Angelov S., Grefen, P., 2004, The business case for B2B e-contracting, Proceedings of the 6th international conference on Electronic commerce.

[2] Hung, P., Li H., Jeng J., 2004, WS-Negotiation: An Overview of Research Issues, Proceedings of the 37th Hawaii International Conference on Systems Sciences

[3] Bui, T., Gachet A., 2005, Web Services for Negotiation and Bargaining in Electronic Markets: Design Requirements and Implementation Framework, Proceedings of the 38th Hawaii International Conference on Systems Sciences.

[4] Davies, J., Fensel, D., Harmelen, F. van (eds.), 2002, Towards the Semantic Web: Ontology-driven Knowledge Management. John Wiley & Sons.

[5] Delgado, J., Gallego, I., Garcia, R., Gil, R., 2002, An Ontology for Intellectual Property Rights: IPROnto. In: 1st Int. Semantic Web Conference (ISWC2002)

[6] Kafentzis, K., Apostolou, D, Mentzas, G. and Georgolios, P., 2004, Knowledge marketplaces: strategic issues and business models, Journal of Knowledge Management, Vol 8, No. 2, 2004.

[7] Langendorfer, B.M., 2003, Towards automatic negotiation of privacy contracts for Internet services, The 11th IEEE International Conference on Networks.

[8] Mentzas, G., Apostolou, D., Abecker, A., Young, R., 2002, Knowledge Asset Management: Beyond the Process-centred and Product-centred Approaches, Series: Advanced Information and Knowledge Processing, Springer London.

[9] Muller, R., Spiliopoulou, M., Lenz, H-J., 2002, Electronic Marketplaces of Knowledge: Characteristics and Sharing of Knowledge, In Proceedings of the International Conference on Advances in Infrastructure for e-Business, e-Education and e-Medicine on the Internet.
[10] Pan, A., Montoto, P., Molano, A., Álvarez, M., Raposo, J., Orjales, V., Viña, Á., 2002, Mediator Systems in E-Commerce Applications, Proceedings of the Fourth IEEE International Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems
[11] Papakonstantinou, Y., Abiteboul, S., Garcia-Molina, H., 1996, Object Fusion in Mediator Systems, Proceedings of the Twenty-second International Conference on Very Large Databases.
[12] Skyrme, D. J., 2001, Capitalizing on Knowledge: From e-business to k-business, Butterworth-Heinemann, London, 2001.
[13] Tiwana, A., 2000, The Knowledge Management Toolkit: Practical Techniques for Building a Knowledge Management System, Prentice Hall PTR Upper Saddle River, NJ, USA
[14] Turner, M., Zhu, F., Kotsiopoulos, I., Russell, M., Budgen, D., Bennett, K., Brereton, P., Keane, J., Layzell, P., Rigby, M., 2004, Using Web Service Technologies to create an Information Broker: An Experience Report, Proceedings of the 26th International Conference on Software Engineering
[15] Wallace, M., Athanasiadis, Th.., Avrithis, Y., Delopoulos, A., Kollias, S., 2006, Integrating Multimedia Archives: The Architecture and the Content Layer, IEEE Trans. Systems, Man, and Cybernetics, Part A: Systems and Humans, 36(1), pp. 34 – 52.
[16] Wallace, M., Avrithis, Y., Kollias, S., 2006, Computationally efficient sup-t transitive closure for sparse fuzzy binary relations, Fuzzy Sets and Systems 157(3), pp. 341-372.
[17] INKASS. Intelligent Knowledge Asset Sharing and Trading, IST – 2001 – 33373
[18] OWL-S. http://www.daml.org/services/owl-s/

This page intentionally left blank

Part II: Intelligent Human-Computer Interaction

Kostas Karpouzis
Institute of Communication and Computer Systems,
National Technical University of Athens (ICCS/NTUA), Greece
kkarpou@cs.ntua.gr

Wikipedia defines Human-Computer Interaction (HCI) as being targeted "…to improve the interaction between users and computers by making computers more usable and receptive to the user's needs". During the recent decades, especially since the advent of the term 'affective computing' by R. Picard, computing is no longer considered a 'number crunching' discipline, but should be thought of as an interfacing means between humans and machines and sometimes even between humans alone. To achieve this, application design must take into account the ability of humans to provide multimodal and semantic input to computers, moving away from the monolithic window-mouse-pointer interface paradigm and utilizing more intuitive concepts, closer to human niches and understanding and capitalizing situation and user context.

Until recently, the ability of computers to recognize user input or semantic concepts in digital media, such as video and audio, was restricted to mouse clicks and statistical processing (e.g. colour histograms or pauses in speech). However, these features are simply not sufficient to describe the actual *conceptual* contents of a video file or the *meaning* of a gesture or utterance from the user. In addition to this, handling of user input and presentation of the results from an interactive session is usually monolithic and inflexible in the sense that information is received, processed and fed to the users irrespectively of their preferences, technology savvy and affective state. As a result, users are often scared away from using automated or interactive systems, since the response they get hardly ever matches their needs.

This part of the book presents innovative concepts related to extracting high-level, semantic information from videos and adapting retrieval and presentation according to user preferences, need and other characteristics, all in a standardized and, therefore, reusable manner. Emerging multimedia standards, such as MPEG-7 and MPEG-21, cater for the flexible description of multimedia content, both with respect to its low-level characteristics (duration, encoding, etc.), as well as regarding the concepts conveyed by its producer (actual objects and their behaviour, for example). In addition to this, they provide support for content adaptation with respect to the machine that plays the content back or to the specific needs and preferences of the user/consumer. As a result, they are constantly being introduced into the content design, production and consumption processes, hence providing new and improved production lines and interactive experiences for the users.

Chapter 9 describes how simple image processing algorithms can be used on still images and, as an extension, video files to extract high-level concepts such as relative spatial placement. Landscape images are used a hands-on example: in this case, the colour information of the different objects, as well as their relative size and positioning are combined with *a priori* knowledge embedded in an ontology. Results from this verification process are introduced back to the content extraction module to help refine results and also succeed in producing confidence values and symbolic labels.

Chapter 10 utilizes profiling and content adaptation procedures to reshape content in an e-learning framework. Here, standardized concepts are put to user so as to refine the presentation and actual content of educational material, taking into account factors such as user savvy and feedback or the semantic content of each lesson.

Finally, chapter 11 builds on the concept of Digital Items (DIs) introduced in the framework of the MPEG-21 multimedia standard to cater for the adaptation of the presentation of heterogeneous multimedia material, coming from diverse providers, with respect to viewer preferences. A full example of the production cycle in the context of movie broadcasts is described here, taking into account the preferences of the viewing audience to adapt content retrieval and presentation.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

High-Level Concept Detection in Video Using a Region Thesaurus

Evaggelos SPYROU [a,1] and Yannis AVRITHIS [a]

[a] *Image, Video and Multimedia Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens*

Abstract. This work presents an approach on high-level semantic feature detection in video sequences. Keyframes are selected to represent the visual content of the shots. Then, low-level feature extraction is performed on the keyframes and a feature vector including color and texture features is formed. A region thesaurus that contains all the high-level features is constructed using a subtractive clustering method where each feature results as the centroid of a cluster. Then, a model vector that contains the distances from each region type is formed and a SVM detector is trained for each semantic concept. The presented approach is also extended using Latent Semantic Analysis as a further step to exploit co-occurrences of the region-types. High-level concepts detected are desert, vegetation, mountain, road, sky and snow within TV news bulletins. Experiments were performed with TRECVID 2005 development data.

Keywords. High-level feature detection, MPEG-7, TRECVID, Latent Semantic Analysis, Region Thesaurus, Region Types, Model Vectors

Introduction

The recent advances in telecommunication technologies, along with the World Wide Web proliferation, have boosted the wide-scale creation and dissemination of digital visual content. More specifically, during the last years, a tremendous increase on the number of video documents has been observed and many users tend to share their personal collections via many popular websites. However, this rate of growth has not been matched by a concurrent emergence of technologies to support efficient video retrieval and analysis. Thus, it appears very common for the average internet user to possess a large amount of digital information, without the ability to effectively browse and retrieve it. Moreover, the number of diverse, recently emerging application areas, which rely increasingly on image and video understanding systems, has further revealed the tremendous potential of the effective use of visual content through semantic analysis.

However high-level concept detection in video documents still remains an unsolved problem. As almost all of the typical recognition problems, this one also has two aspects. The first is the extraction of the various features of a video sequence, such as color,

[1]Corresponding Author: Evaggelos Spyrou, Image, Video and Multimedia Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, 9 Iroon Polytechniou Str., 157 80 Athens, Greece; E-mail: espyrou@image.ece.ntua.gr.

texture, motion and audio a process commonly called low-level feature extraction and then form a description by combining them. The other aspect is the method used for assigning these low-level descriptions to high-level concepts, a problem that is often referred to as the Semantic Gap. Many approaches have been proposed that share the goal of bridging the semantic gap, thus allowing the proper extraction of high-level concepts of multimedia documents using and combining heterogeneous audiovisual features.

In [7], a prototype multimedia analysis and retrieval system is presented, that uses multi-modal machine learning techniques in order to model semantic concepts in video, from automatically extracted multimedia content. A region-based approach in content retrieval that uses Latent Semantic Analysis (LSA) techniques is presented in [15].The choice of global local visual features also appears crucial for good analysis results. In order to exploit the spatial content of a keyframe, the extraction of low-level concepts is performed after the image is modeled using grids, thus color and texture features are selected locally [2]. A similar approach is presented in [13]. Here, the features are extracted by regions of an image that resulted using a mean-shift algorithm. Finally in [18], a region-based approach is presented, that uses knowledge encoded in the form of an ontology. MPEG-7 visual features are extracted and combined and high-level concepts are detected.

In the context of TV news bulletins, a hybrid thesaurus approach is presented in [10]. There, semantic object recognition and identification for video news archives is achieved, with emphasis to face detection and TV channel logos. A lexicon-driven approach for an interactive video retrieval system is presented in [3].The core of this solution is the automatic detection of an unprecedented lexicon of 101 concepts. A lexicon design for semantic indexing in media databases is also presented in [1].

In this work, the problem of concept detection in video is approached in the following way: Each shot is represented by a keyframe, thus, the first step is keyframe extraction from shots. Then a clustering algorithm is applied on the RGB color values of every keyframe and splits the keyframes in homogeneous regions. The centroids of the clusters denote the color description of the image. For each cluster, a texture descriptor is extracted, and this way, the texture description of the image is formed. Fusing both low-level descriptions, a feature vector for each image is formed. Using a significantly large number of keyframes and applying a subtractive clustering method, we construct a region thesaurus, containing all the region types, which may or may not represent the concepts that are chosen to be detected. This thesaurus acts as the knowledge base and facilitates the association of low to high-level features. Each region type of the region thesaurus contains the appropriate merged color and texture description. By measuring the distances of the regions of an image to the region types, a model vector is formed that captures the semantics of an image. A support vector machine is trained to detect each high-level semantic concept, based on the values of the model vectors.

During the last years, there has been an effort to effectively evaluate and benchmark various approaches in the field of information retrieval, by the TREC conference series. Within this series the TRECVID [16] evaluation attracts many organizations and research interested in comparing their research in tasks such as automatic segmentation, indexing, and content-based retrieval of digital video. For its first years, the interest of TRECVID has been the TV news domain. Within the context of TRECVID, the presented work is applied on TV news bulletins from TRECVID 2005 development data, to

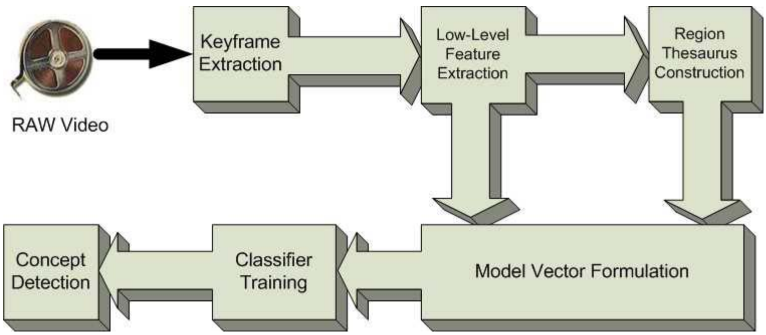


Figure 1. Presented Framework

detect specific high-level features: *desert*, *vegetation*, *mountain*, *road*, *sky* and *snow*. The presented framework is depicted in figure 1.

This paper is organized as follows: Section 1 presents the method used for the extraction of the color and texture features of a given keyframe. The method for the construction of the thesaurus containing all the region types derived from the training set is presented in section 2, followed by the construction of the model vectors that include the semantic image features in section 3. Then, section 4 presents the application of the Latent Semantic Analysis technique to the presented problem. Section 5 presents SVM-based high-level feature (concept) detectors, followed by experimental results in section 6. Finally, conclusions are drawn in section 7 accompanied by plans for future work.

1. Low-Level Feature Extraction

For the representation of the low-level features of a given keyframe, this work considers only color and texture features. For the color properties, a simple description similar to the approach of the MPEG-7 Dominant Color Descriptor and for the texture properties the actual MPEG-7 Homogeneous Texture Descriptor [8] have been applied, respectively.

1.1. Color Features

It is shown in many image retrieval applications that a set of dominant colors in an image or a region of interest is usually capable of efficiently capturing its color properties. The standardized MPEG-7 Dominant Color Descriptor [8] is formed after the clustering of the present colors within an image or a region of interest. This way, the representative colors of each keyframe are calculated. The selected low-level visual features of the image consist of the representative (dominant) colors, their percentages in the region, and optionally their spatial coherencies and their variances. What discriminates the use of the dominant color description of an image, instead of i.e. a color histogram is that the representative colors are computed each time, based on the features of the given image rather than being fixed in the color space.

Figure 2. An input image used for the extraction of four dominant colors

In our approach, the well-known K-means clustering method is applied on the RGB values of a given keyframe. As opposed to the MPEG-7 Dominant Color descriptor, where the number of the extracted representative colors varies from image to image allowing a maximum of eight colors that are allowed to be extracted from each image, a fixed number of colors is each time preselected in our approach. Using this predefined number, an image is then represented by the features of a fixed number of regions. Since the colors are clustered and the average color of each image region (cluster of the color space) is considered, our approach describes the color properties in a similar way to the Dominant Color Descriptor and there is no need to further extract more colors from its region, since the regions already occur from color clustering and share similar color properties.

The color description is then formed as follows:

$$DCD_N = [\{C_1, P_1\}, \{C_2, P_2\}, \ldots, \{C_N, P_N\}]$$

An example of an input image is depicted in figure 2. For the case of four dominant colors the four images each one containing one of the four regions are depicted in figure 3.

1.2. Texture Features

To efficiently capture the texture features of an image, the MPEG-7 Homogeneous Texture Descriptor (HTD) [8] is applied, since it provides a quintative characterization of texture and comprises a robust and easy to compute descriptor. The image is first filtered with orientation and scale sensitive filters. The mean and standard deviation of the fil-

Figure 3. The four extracted regions of image depicted in figure 2 by the K-means clustering described in section 1.1

tered outputs are computed in the frequency domain. The frequency space is divided in 30 channels, as described in [9], and the energy and energy deviation of each channel are computed and logarithmically scaled.

The texture description for a region of an image is then formed as follows:

$$HTD = [f_{DC}, f_{SD}, e_1, e_2, ..., e_{30}, d_1, d_2, ..., d_{30}]$$

Where f_{DC} and f_{SD} denote the mean and the standard deviation of the image texture respectively. Since each image is divided into four regions based on its color features, its texture properties are described by four homogeneous texture descriptors. The energy deviation of each channel is discarded, in order to simplify the description and moreover to prevent biasing towards the texture features, since they already have a significantly higher dimensionality than the color ones.

1.3. Fusion of color and texture features

All the low-level visual descriptions of a keyframe are merged into a unique vector. If DCD, HTD_1, HTD_2, HTD_3, $\ldots HTD_N$ are respectively the N dominant (representative) colors and the homogeneous texture descriptors of each region of the keyframe referenced before, then the merged keyframe description D_{KF} is defined as:

$$D_{KF} = [DCD|HTD_1|HTD_2|HTD_3|\ldots|HTD_4] \quad (1)$$

It is necessary that all color and texture features should have more or less the same numerical values to avoid scale effects. To achieve that, the dominant color and the homogeneous texture descriptions are normalized before their fusion into this unique vector which will be referred to as *feature vector*.

2. Region Thesaurus Construction

By observing the set of the keyframes extracted from the entire video collection and their low-level visual features extracted as described in section 1, becomes obvious that keyframes with similar semantic features should have similar low-level descriptions. Apart from that, there are many keyframes that share almost identical color and texture features since in our case they are taken in a studio with a still camera and have only a small time difference. Moreover, many shots within the same TV news bulletin contain only the anchorman and the same artificial background. To exploit this, clustering is performed on all the descriptions of the training set. Since we cannot have a priori knowledge for the exact number of the required classes, a K-means or a Fuzzy C-means clustering approach does not appear useful enough. To overrun this problem *Subtractive clustering* [4] is the method we choose to apply on the low-level description set. This method assumes each data point is a potential cluster center and calculates a measure of the likelihood that each data point would define the cluster center, based on the density of surrounding data points. In other words, this algorithm defines the number of the clusters and their corresponding centroids.

After the application of the clustering technique on the training data set, the following observations become obvious: First, each cluster may or may not represent a high-level feature. There are some clusters that contain region types belonging to the same high-level concept. Apart from those, most of the clusters contain region types that do not belong to the same high-level feature and are mixed up because they share similar color and texture features. Second, some concepts can be found in more than one clusters, since they cannot be described uniquely by their visual characteristics. For example, the concept *desert* can have more than one instances differing in i.e. the color of the sand, each represented by the centroid of a cluster. Moreover, in a cluster that may contain instances from the semantic entity i.e. *sea*, these instances could be mixed up with parts from i.e. *sky*, if present in the data set.

Generally, a *thesaurus* combines a list of every term in a given domain of knowledge and a set of related terms for each term in the list. In our approach, the constructed Region Thesaurus contains all the Region Types that are encountered in the training set. These region types are the centroids of the clusters and all the other feature vectors of a cluster are their synonyms. It is important to mention that when two region types are considered to be synonyms, they belong to same cluster, thus share similar visual features, but do not necessarily share the same semantics. By using a significantly large training set of keyframes, our thesaurus is constructed and enriched. As it will be presented in section 3, the use of the thesaurus is to provide a means of association of the low-level features of the image with the high-level concepts.

Since the number of the region types can be very large depending on the selected thresholds for the potential above or below from which a data point will be selected or rejected respectively as a cluster center, the dimensionality of the model vector may become very high. It is then possible that the extracted region types may carry redundant information. However, since two region types may sometimes be strongly correlated although they may appear visually different. To avoid this, principal component analysis (PCA) is applied in order to reduce the dimensionality and facilitate the performance of the high-level feature detectors which are presented in section 5.

3. Model Vectors for keyframe representation

After the construction of the region thesaurus, a model vector is formed for each keyframe. Its dimensionality is equal to the number of concepts that constitute the thesaurus. The distance of a region to a region type is calculated as a linear combination of the dominant color and homogeneous texture distances respectively. The MPEG-7 standardized distances are used for each case and a linear combination is used to fuse the distances as in [6].

3.1. Similarity Measures

We begin with the similarity measure used for the case of the color descriptor. Since the color representation is rather simple, the well known Euclidean distance is used since it works effectively in many applications and retrieval systems.

$$D(F_1, F_2) = \sqrt{(R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2} \tag{2}$$

where $F_i = (R_i, G_i, B_i), i = 1, 2$ are the two RGB values. The distance between 2 *Homogeneous Texture Descriptors* is computed as:

$$D(HTD_1, HTD_2) = \sum_k \left| \frac{HTD_1(k) - HTD_2(k)}{a(k)} \right| \tag{3}$$

where $a(k)$ is the standard deviation of the Homogeneous Texture Descriptors for a given database. In this approach the standard deviation of the database is ignored, because it does not affect the result since the values are normalized afterwards.

However we should notice that the MPEG-7 standard does not strictly define the distance functions to be used, thus leaving the developers the flexibility to develop their own dissimilarity/distance functions and to exploit other well-known similarity functions such as i.e. the Minkowski distance.

3.2. Model Vector Formulation

Having calculated the distance of each region (cluster) of the image to all the words of the constructed thesaurus, the model vector that semantically describes the visual content of the image is formed by keeping the smaller distance for each high-level concept. More specifically, let: $d_i^1, d_i^2, ..., d_i^j, i = 1, 2, 3, 4$ and $j = N_C$, where N_C denotes the number of words of the lexicon and d_i^j is the distance of the i-th region of the clustered image to the j-th region type. Then, the model vector D_m is the one depicted in equation 4.

$$D_m = [min\{d_i^1\}, min\{d_i^2\}, ..., min\{d_i^{N_C}\}], i = 1, 2, 3, 4 \tag{4}$$

4. Latent Semantic Analysis

Apart from the obvious next step of simply training classifiers using the aforementioned model vectors as the means of representing the extracted features of the given keyframe, we also perform some experiments using a Latent Semantic Analysis [5](LSA) approach. LSA is a technique in natural language processing, which exploits the relationships between a set of documents and the terms they contain more often by producing a set of concepts related to the documents and terms. In our approach, since a keyframe is described as a set of region types, it appears obvious that LSA can easily be applied with the keyframe and the region types corresponding to a document and the terms it contains, respectively.

We first construct the co-occurrence matrix of region types in given keyframes of the training set in contexts (region types in the thesaurus). The distance function we use to compare a given region type with one of the thesaurus, in order to assign each region of the image to the correct prototype region is a linear combination of a Euclidean distance for the dominant color and the MPEG-7 standardized distance for the HTD. After of the construction of the co-occurrence matrix, we solve the SVD problem and transform all the model vectors to the semantic space. For each semantic concept, a separate SVM is then trained having as input the model vector in the semantic space.

5. SVM Feature Detector Training

Support Vector Machines [17] are feed-forward networks that can be used for pattern classification and nonlinear regression. Their main idea is to construct a hyperplane that acts as a decision space in such a way that the margin of separation between positive and negative examples is maximized. This hyperplane is not constructed in the input space, where the problem may not be linearly solvable, but in the feature space where the problem is driven. This is generally referred as the Optimal Hyperplane, a property that is achieved as the support vector machines are an approximate implementation of the method of structural risk minimization. Despite the fact that a support vector machine does not incorporate domain-specific knowledge, it provides a good generalization performance, a unique property among the various different types of neural networks. Support vector machines have been used for image classification based on their histogram as in [12] and for the detection of semantic concepts such as goal, yellow card and substitution in the soccer domain [14].

An inner-product kernel between an input vector $\mathbf{x}$ and a *support vector* $\mathbf{x}_i$ is the main characteristic on the support vector machines. The support vectors consist of a small subset of the training set vectors and are extracted by the optimization algorithm. The kernel can be implemented in various ways, thus leading to different types of nonlinear learning machines. The most important are *Polynomial* learning machines, *Radial-Basis Function* networks and Single-hidden layer *Perceptrons*, where the kernel function is polynomial, exponential or a hyperbolic tangent function, respectively.

The nonlinear mapping may be denoted by a set of nonlinear transformations as $\{\phi_j(\mathbf{x})\}_{j=1}^{m_1}$. Then, a hyperplane in the feature space is defined as:

Table 1. Classification rate using both visual descriptors for various numbers of the region types

Concept	35 Region Types	62 Region Types	125 Region Types
Desert	82.5%	77.5%	70.1%
Vegetation	80.5%	71.3%	67.2%
Mountain	83.6%	77.7%	67.0%
Road	72.0%	67.0%	65.9%
Sky	80.1%	77.4%	70.0%
Snow	70.5 %	62.1%	55.2%

$$\sum_{j=1}^{m_1} w_j \phi_j(\mathbf{x}) + b = 0 \tag{5}$$

If we denote the inner-product kernel of the support vector machine as $K(\mathbf{x}, \mathbf{x}_i)$, it is defined by:

$$K(\mathbf{x}, \mathbf{x}_i) = \boldsymbol{\phi}^T(\mathbf{x})\boldsymbol{\phi}(\mathbf{x}_i) \tag{6}$$

For each semantic concept, a separate support vector machine is trained, thus solving a binary problem, of the existence or not of the concept in question. The input of the SVM is the model vector D_m described in section 3. The well known polynomial support vector machine, described in equation 7 is selected in our framework.

$$K(\mathbf{x}, \mathbf{x}_i) = (\mathbf{x}^T\mathbf{x}_i + 1)^p \tag{7}$$

6. Experimental Results

For the evaluation of the presented framework, part of the development data of TRECVID 2005 were used. This set consists of approximately 65000 keyframes, captured from TV news bulletins. For the following experiments, a set of 5000 keyframes was selected in order to include examples of all the selected features and also some keyframes not containing any of them. The high-level features for which feature detectors were implemented are: *desert, vegetation, mountain, road, sky and snow*. The annotation was provided by the LSCOM Lexicon Definitions and Annotations [11]. The color visual features were extracted using a standard K-means clustering algorithm and the texture visual features using the MPEG-7 eXperimentation Model (XM).

Experiments were performed on the size of the region thesaurus, the number of dominant colors and the presence or not of both visual descriptors. Results are shown in table 1 for different sizes of the region thesaurus, in table 2 for fixed size of the region thesaurus and the use of different numbers of dominant colors and finally in table 3 for each descriptor and their combination. Also, experiments were performed for the case of fixed size of the region thesaurus and for each one or both of the descriptors, using LSA as an intermediate step between model vector formulation and SVM training and classification. The performance of these experiments is presented in table 4

Table 2. Classification rate using both visual descriptors for various numbers of the dominant colors, thesaurus size = 35

Concept	2 DC + HT	3 DC + HT	4 DC + HT	5 DC + HT
Desert	77.5%	80.5%	82.5%	79.0%
Vegetation	70.5%	77.5%	80.5%	81.2%
Mountain	70.3%	82.0%	83.6%	78.6%
Road	68.0%	70.0%	72.0%	70.0%
Sky	77.5%	80.1%	80.1%	79.0%
Snow	57.2%	62.0%	70.5%	72.2%

Table 3. Classification rate using only color, only texture and both visual descriptors, thesaurus size = 35

Concept	DC	HT	DC+HT
Desert	80.2%	77.2%	82.5%
Vegetation	72.5%	75.0%	80.5%
Mountain	72.1%	77.5%	83.6%
Road	71.5%	70.2%	72.0%
Sky	85.0%	70.1%	80.1%
Snow	75.0%	60.1%	70.5%

Table 4. Classification rate using only color, only texture and both visual descriptors and LSA in all cases, thesaurus size = 35

Concept	DC	HT	DC+HT
Desert	83.2%	75.2%	87.2%
Vegetation	74.5%	75.2%	82.5%
Mountain	77.5%	77.5%	80.6%
Road	78.2%	73.7%	76.7%
Sky	88.2%	72.5%	82.2%
Snow	79.0%	65.0%	72.5%

7. Conclusions - Future Work

The experimental results indicate that the extraction of the aforementioned low-level features is appropriate for semantic indexing. The selected concepts can be successfully detected when a keyframe is represented by a model vector that contains the distances to all the semantic entities of a constructed lexicon containing unlabeled semantic features. Latent Semantic Analysis was also successfully applied in the given problem and led to an improvement of the results. Plans for future work include the extraction of more visual features, exploitation of the spatial context of a keyframe and extension of this method for applications such as shot/image classification. Finally, integration of the presented framework to the one of [18] and fusion of their results is also intended.

8. Acknowledgements

The work presented in this paper was partially supported by the European Commission under contracts FP6-027026 K-Space and FP6-027685 MESH. Evaggelos Spyrou is funded by PENED 2003 Project Ontomedia 03ED475.

References

[1] M. Naphade A. Natsev and J. Smith. Lexicon design for semantic indexing in media databases. In *International Conference on Communication Technologies and Programming*, 2003.

[2] S. Aksoy, A. Avci, E. Balcuk, O. Cavus, P. Duygulu, Z. Karaman, P. Kavak, C. Kaynak, E. Kucukayvaz, C. Ocalan, and P. Yildiz. Bilkent university at trecvid 2005. 2005.

[3] Dennis C. Koelma Cees G.M. Snoek, Marcel Worring and Arnold W.M. Smeulders. Learned lexicon-driven interactive video retrieval. 2006.

[4] S.L. Chiu. *Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification*. John Wiley and Sons, 1997.

[5] S. Deerwester, Susan Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the Society for Information Science*, 41(6):391–407, 1990.

[6] E.Spyrou, H.LeBorgne, T.Mailis, E.Cooke, Y.Avrithis, and N.O'Connor. Fusing mpeg-7 visual descriptors for image classification. In *International Conference on Artificial Neural Networks (ICANN)*, 2005.

[7] IBM. Marvel: Multimedia analysis and retrieval system.

[8] B.S. Manjunath, J.R. Ohm, V.V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE trans. on Circuits and Systems for Video Technology*, 11(6):703–715, 2001.

[9] MPEG-7. Visual experimentation model (xm) version 10.0. ISO/IEC/ JTC1/SC29/WG11, Doc. N4062, 2001.

[10] V. Gouet N. Boujemaa, F. Fleuret and H. Sahbi. Visual content extraction for automatic semantic annotation of video news. In *IS&T/SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia, part of Electronic Imaging symposium*, January 2004.

[11] M. R. Naphade, L. Kennedy, J. R. Kender, S.-F. Chang, J. R. Smith, P. Over, and A. Hauptmann. A light scale concept ontology for multimedia understanding for trecvid 2005. IBM Research Technical Report, 2005.

[12] O.Chapelle, P.Haffner, and V.N.Vapnik. Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10(5):1055–1064, 1999.

[13] B.Le Saux and G.Amato. Image classifiers for scene analysis. In *International Conference on Computer Vision and Graphics*, 2004.

[14] C. G. M. Snoek and M. Worring. Time interval based modelling and classification of events in soccer video. In *Proceedings of the 9th Annual Conference of the advanced School for Computing and Imaging (ASCI)*, 2003.

[15] F. Souvannavong, B. Mérialdo, and B. Huet. Region-based video content indexing and retrieval. In *CBMI 2005, Fourth International Workshop on Content-Based Multimedia Indexing, June 21-23, 2005, Riga, Latvia*, Jun 2005.

[16] TREC. - video retrieval evaluation. http://www-nlpir.nist.gov/projects/t01v/.

[17] V. Vapnik. *Statistical Learning Theory*. John Wiley and Sons, 1998.

[18] N. Voisine, S. Dasiopoulou, V. Mezaris, E. Spyrou, Th. Athanasiadis, I. Kompatsiaris, Y. Avrithis, and M. G. Strintzis. Knowledge-assisted video analysis using a genetic algorithm. In *6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005)*, April 13-15, 2005.

An Integrated Approach Towards Intelligent Educational Content Adaptation

Phivos MYLONAS, Paraskevi TZOUVELI and Stefanos KOLLIAS
National Technical University of Athens
School of Electrical and Computer Engineering, Department of Computer Science
Image, Video and Multimedia Laboratory,
Zographou Campus, Athens, Greece, 157 73

Abstract. One of the major shortcomings of modern e-learning schemes is the fact that they significantly lack on user personalization and educational content representation issues. Semi- or fully automated extraction of user profiles based on users' usage history records forms a challenging problem, especially when used under the e-learning perspective. In this chapter we present the design and implementation of such a user profile-based framework, where educational content is matched against its environmental context, in order to be adapted to the end users' needs and qualifications. Our effort applies clustering techniques on an integrated e-learning system to provide efficient user profile extraction and results are promising.

Keywords. Online education, Personalized education, E-learning, Clustering-based user profiling

1. Introduction

In the current Internet-based world, new trends concerning online education and e-learning evolve. Users are constantly confronted with a series of technological improvements and developments; however, most of the proposed approaches do not tackle sufficiently the raised user personalization and educational content representation issues. Currently, traditional teaching techniques are finding themselves under revision and re-evaluation and new or sometimes radical ones do come into play. Above all, the Internet plays a significant role in all fields of education, contributing the most to the educational procedures [1]. As a result, Internet-oriented applications arise in the aid of educational needs, trying to close the gap between traditional educational techniques and technology-oriented education.

One of the most innovative ways to empower such a workforce with the skills and knowledge it needs is the utilization of e-learning; e-learning is currently propagating at quick rates and its impact on teaching and learning raises many questions with no clear answers. The impact of Information and Communications Technology (ICT) is a task that has become more and more apparent in learning and teaching at all levels of education [1]. Towards that goal, during the last years, e-learning systems were developed in terms of rather static software applications, lacking on educational multimedia environments and personalized capabilities and without any interest given to the real users input and feedback [2–4].

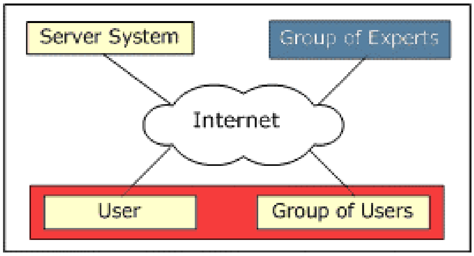


Figure 1. SPERO System Architecture.

In the framework of the presented Leonardo SPERO project [5], we introduce a novel method for gathering information and estimating the ICT level of learners in all fields of education. This is achieved through a web-based interface that takes into account personalized, profile-based schemes. It has been designed to enable learners to gracefully increase their ICT knowledge and provide them with credible information and feedback, such as suitably selected e-courses and multimedia educational content.

The structure of this chapter is as follows: in Section 2, the overall architecture design of the SPERO system is introduced, including its basic corresponding groups and components as well as its e-questionnaires. Afterwards, a short reference to the IEEE e-learning model is presented, current approach's adaptation is analyzed and the additional features provided by it are explained. In Section 3, we begin by tackling the problem of the learner profile creation, followed by issues concerning the initial static profile extraction procedures. All of the above are used as the main feedback source for the forthcoming intelligent clustering profiling procedure, which is presented in Section 4. In the same section a description of the utilized clustering algorithm is provided, together with experimental results on the clustering-based profiling scheme. Next Section 5 describes the general context for this work, dealing with the educational content offering of the system and briefly presenting its categorized e-courses and operational examples of use. Finally, in Section 6, we present our concluding remarks and some ideas for potential future efforts on the subject.

2. Framework's Overview

The first step to consider towards the establishment of an efficient, integrated e-learning framework is the definition of its basic architecture, which in our case is shown in Fig 1.

Three main networked components can be identified which are:

1. the group of the *system's users*
2. the group of *system's experts*, who play a key role in the initialization of the personalization process; this group includes teachers, experts in e-learning, data analysts, psychologists and software engineers
3. the actual *server system*, which includes all hardware and software needed to establish a 2-tier system core [6].

The first group includes every kind of teacher working either for general education or for the Special Education sector. Eventually, expanding the system's architecture, a

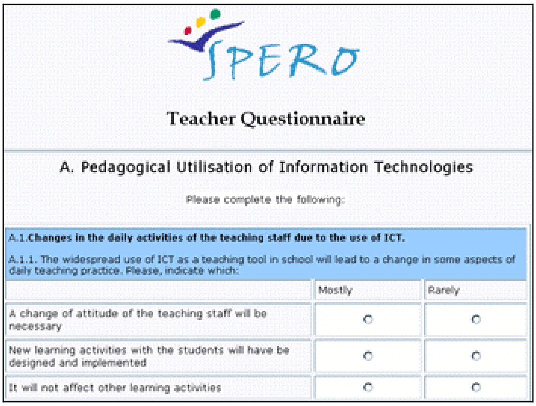


Figure 2. Part of Teachers' Questionnaire.

user could be identified as any learner, whether a student, teacher or employee. The second group's role is crucial in the personalization process of the SPERO system, since it defines the initial set of specifications and limitations of the end-users' profiles, which justifies the variety of people comprising it. The third group includes all hardware and software that enables a web-server to be active, as well as efficient and robust. All distinct SPERO web applications together with an underlying Relational Database Management System to support profiling and user information are included in this configuration.

Moreover, one of the basic interaction components of SPERO and the mean of communication with its end-users is formed by the so called *e-questionnaire*. SPERO experts have designed and illustrated two groups of e-questionnaires: The first group contains questions about school units in order to collect general details about them. The second group contains questions about teachers' ICT background. The questions, which are addressed to the teachers, are intended to collect information about teachers' educational background, as well as their background in ICT. In addition, information concerning teachers' opinions about pedagogical utilization of ICT and the amount of using ICT in teaching procedure is also extracted.

As already mentioned, learners could either be teachers or students, however both of them are in great need of ICT: on the one hand, teachers mainly because their role is continuously evolving and demanding new formation and on the other hand students, because of their need to have distance e-courses in the field of ICT. Each questionnaire is divided into several subsections, a portion of which is depicted in Fig. 2. The presented sub-questionnaire collects information about general teachers' educational background, as well as their background in ICT.

SPERO e-questionnaires are developed in the framework of conducting a European survey. Consequently, they are translated in eight European languages and these translations are stored within SPERO's Relational Database Management System. Software has been developed for automatic presentation of e-questionnaires in every one of these eight languages. Moreover, the e-questionnaire is used to estimate the ICT level of individual users, by using the calculated user's profile categorizations that are automatically extracted in the following by the SPERO software. More than one learning resources (e-courses and educational material) are selected by experts, up to one for

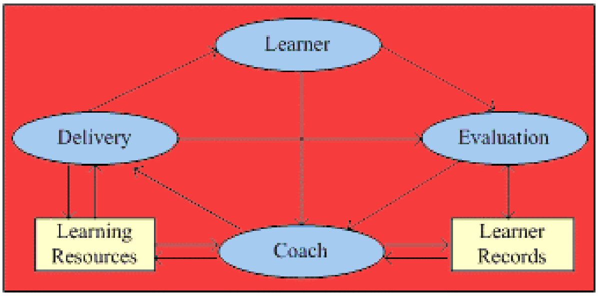


Figure 3. IEEE learning system entities.

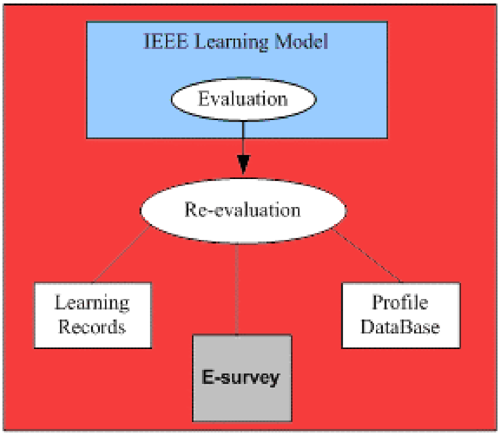


Figure 4. Proposed replacement of *Evaluation* entity.

each of the distinct collaborative user profiles categories. The set of e-questionnaires is used for ICT level estimation in the framework of the distance-learning architecture that is presented in Fig. 3.

In this work we attempt to extract learner profiles through the evaluation entity of the above architecture, proposed by the IEEE Reference Model (WG1 LTSA) of the Learning Technology Standards Committee [7,8]. This standard covers a wide range of systems, such as learning technology, education and training technology, computer-based training, computer assisted instruction, intelligent tutoring, and is pedagogically neutral, content-neutral, culturally neutral, and platform-neutral.

However, in this generic approach to e-learning systems, a system's ability to adapt its operation to the user is not defined, although an evaluation process exists. Aiming at extracting learner profiles through this entity, we are proposing the replacement of the IEEE standard *Evaluation* entity, with the novel *Re-evaluation* entity, which, additionally, is strongly related to two new entities: the *E-survey* entity for gathering statistical information and the *Profile Database* entity dealing with all learners' profiles. A schematic diagram of the proposed replacement is presented in Fig. 4, the replaced learning components are shown in Fig. 5, whereas a detailed description of the above concepts can be found in [9].

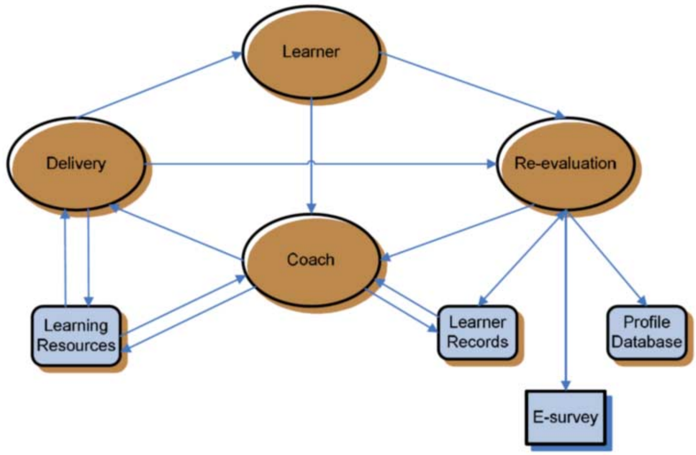


Figure 5. Proposed e-learning system after replacement of *Evaluation* entity.

In order to assist the profiling process, the need for this re-evaluation step is essential; the usage of an appropriate e-questionnaire is considered necessary in order to collect user input data and build a large, reliable ground truth, basing profiling information on top of them. For this purpose, experts designed and illustrated an e-questionnaire which collects information about learners' ICT background, learners' opinions about pedagogical utilization of ICT and the amount of using ICT in teaching procedure. Additionally, software has been developed to allow e-surveys to be conducted based on users' answers.

As a result, the core of the system relies on this replaced *Evaluation* entity. The latter forms an independent personalization subsystem, where user profiling information is extracted, according to statistics gathered from the e-questionnaire database and the e-survey. Delivery of educational content is then possible, based on the results of the profiling procedure, providing personalized views to the system's end-users and taking into consideration their particular ICT levels of education and needs.

3. Initialization of User Profiles

At this point, a brief presentation of the system's personalization subsystem is essential. It is the intended nature of SPERO that dictates utilization of two profiling approaches; both of them look very different in the beginning, but they are combined at a later phase. One approach is followed at the initial stage of constructing the profiling ground truth and is characterized by a static profiling mechanism, whereas the second approach exploits results provided by the previous one towards dynamic extraction of current and future SPERO users. First step is materialized through a fixed mapping of obtained user input to specific types of user profiles explained in the following within this section. Second step is acting on top of the first one and is founded on application of a clustering technique [10]. The latter is presented in the next section. Accordingly, the corresponding dataset was divided into two parts, one primary utilized by experts during the first step and another utilized by the automated clustering methodology.

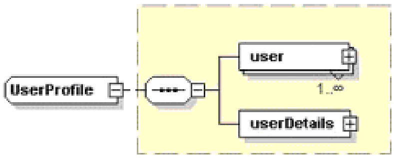


Figure 6. Structure of the *UserProfile*.

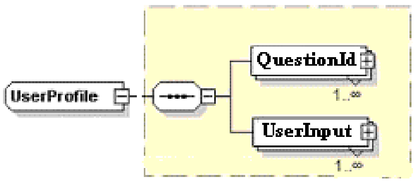


Figure 7. Mapping structure of *UserProfiles*.

Let us illustrate the functionality of the initial profile representation, which is based on the static profiling mechanism. Experts, based on experience and intuition, define a set of three user characterizations, forming a static profiling representation. This step forms a pre-processing task performed by hand, combined with statistical analysis and later cross-correlation selection of features, which indicates the most important ones that were considered in the following. It is considered to be extremely reliable and accurate, as it was the result of three years of intensive collaboration among experts from 8 European countries that examined this through questionnaires in national and cross-national analysis. These initial characterizations are also utilized at a later stage, during the intelligent profiling process, providing a rock-solid point of reference and although they are thought to be static, they are actually generated automatically from the system. Their basis is information provided by the users' input data, obtained from the e-questionnaires. In this case, personalization was needed in order to aid with the ultimate educational content offered by the system; this was successful, based on the electronic mining of knowledge gathered from the system's questionnaires subsystem.

The profiling mechanism creates updates and uses system's user profiles, matching specific e-questionnaires question triggers, to particular identified patterns. The profile model's design facilitates both the process of using user preferences in profile creation, as well as the process of preference tracking throughout the whole profiling procedure. Furthermore, it is designed in a way that allows for the automated extraction of user profiles, based on these preferences and the users' input history. This model forms an initial static version of the user profile denoted by ***UserProfile*** (Fig. 6).

As seen in the figure, the main abstract structure of ***UserProfile*** compound type contains two elements. The first one (i.e. *user*) stores information about the user's history, while the second (i.e. ***userDetails***) stores the user preferences. As the initial profiling process instantiates, all user profiles are stored within a single, central mapping structure, whose abstract model is presented in Fig. 7. The ***UserProfile*** is mapped against information retrieved either from the e-questionnaire itself, or directly from the input of the users. The first element, ***QuestionId*** holds all the information required for identifying the underlying e-questionnaire question, as well as its type, aiming at better

B.1.4.Teaching Experience	
Under 5 years	Beginner
5 to 10 years	Advanced
11 to 20 years	Advanced / Expert
Over 20 years	Expert

Figure 8. Static profile mapping example (1st part of e-questionnaire).

B.2.5.Which of the following tasks have you performed at least once ?	
Installation of software	a
Installation of a printer	b
Greation of backup	b
Installation of applications (e.g. MS Office)	b
Problem solving related to software	e
Problem solving related to the printer or Internet card	f
None of the above	g
Other, please indicate:	text

Checked (a…f)>4 => “Expert”

2<Checked (a…f) <4 => “Advanced”

2 < Checked (a…f) OR Checked (g) => “Beginner”

Figure 9. Static profile mapping example (2nd part of e-questionnaire).

B.3.4.How many hours a week, on average, do you use the Internet or educational software with your students?	
General	if (General or/and SEN teaching hours)>5 “=> Expert"
SEN	else if 2 < t. h.) <5=>"Advanced" else if (t. h.)<2=>"Beginner"

Figure 10. Static profile mapping example (3rd part of e-questionnaire).

understanding and fitting of the currently generated profile. The second element, ***UserInput*** contains user data related information, such as the user’s answers. Both, the sequences of ***QuestionId*** and ***UserInput*** elements denote the existence of large amount of different system’s e-questionnaires, questions and users’ input data.

The core of this methodology is summarized in the following step of weights association, performed according to the following guidelines: Once a user answers a question of the input e-questionnaire, a relevance degree is associated to it and adjusted to her/his specific QuestionId element, and thus also propagated to the UserProfile element. As more and more answers from the end user enter the UserProfile structure, additional relevance degrees are registered to the corresponding QuestionID elements. Depending on the particular question, as well as the part of the e-questionnaire that this question belongs, different degrees are propagated. The latter is based on comparison of the provided numerical values with the range of values a-priori associated with the profiles. In order to better understand the underlying mapping structures, examples are presented in Fig. 8, Fig. 9 and Fig. 10, derived from the group of experts directly from the e-questionnaires.

	User ID	Professional Development	Personal ICT Background	Teaching use of ICT
1.	509	beginner	unspecified	beginner
2.	659	beginner	unspecified	beginner
3.	807	expert	expert	beginner
4.	808	unspecified	beginner	beginner
5.	809	expert	expert	advanced
6.	811	expert	expert	advanced
7.	813	expert	expert	advanced
8.	814	expert	expert	advanced

Figure 11. Initial static profiling mapping.

The e-questionnaire acts as an intermediate towards the information gathering process, and as the amount of the answered questions increases, the more entries are summed up in the "UserProfile" structure. Thus, the overall process results in an aggregated weighted mapping of the end user to the specified profile, which is different for each user's answers and depends on their particular nature. This mapping is temporarily preserved and as the completion of the e-questionnaire is carried out, the above mentioned weighted mappings are aggregated. In that manner, they continuously and dynamically change every user's profiling, until a final equilibrium profile state is achieved. Test bed experimental results within the SPERO project indicate that after answering an approximately 50% of questions, it is possible for the system to balance to a solid, static, initial user profile with great confidence. As a result, the entire preprocessing task is governed by great accuracy and reliability, although in general it is difficult to handle in such cases, where data is characterized by numerous measurable features like answers to e-questionnaires. This is the case when multiple independent features characterize data, and thus more than one meaningful similarity or dissimilarity measures can be defined. A common approach to the problem is the lowering of input dimensions [10], which may be accomplished by ignoring some of the available features/answers, and is the one followed herein. Additionally, a statistical cross-correlation analysis of the importance of features was also performed. For the sake of space we omit detailed presentation of the feature selection process, however the basic principle to be followed throughout this chapter is that while we expect elements of a given meaningful set of e-questionnaires to have random distances from one another according to most features, we expect them to have small distances according to the features that relate them. We rely on this difference in distribution of distance values in order to identify the context of a set of elements, i.e. the subspace in which the set is best defined and provides the most meaningful results in terms of semantic clarity. The result of this process is a set of 44 meaningful e-questionnaire questions. The final output of this process, following the application of the weights, is the extraction of a "1-1" profile–end user relation. In that manner, each end user is classified to an initial, static profile that characterizes his behaviour, his interests and his further treatment from the system. This particular profile characterization forms the basis of the following intelligent clustering procedure, which includes the notion of profile extraction and integration within this system. In Fig. 11 we present an indicative sample of the end-users' static profiling, extracted by previously analyzed procedure within the SPERO system, according to the users' answers collected by the e-questionnaires.

4. Clustering-Based Profiling

At this point, experts are considered to have successfully labelled the first part of the dataset, provided initially as the ground truth of the profiling mechanism, by assigning profiles to each user according to the responses given to e-questionnaires. The next step to follow consists of applying intelligent clustering techniques to group similar profiles in the current as well as second part of the dataset. Traditional classification schemes are not applicable in this case and an intelligent clustering methodology is favoured. This step is necessary for the unsupervised operation of SPERO, where no manual labelling is available. Of course, at this phase various predefined profile schemes may arise and results are not bound to the so far manual characterization of system's experts. In this manner, more than three profiles may be obtained and categorization to the predominant ones needs to be applied. The latter is necessary for us, in order to be able to compare results between the two phases and derive meaningful conclusions. After a small theoretical analysis of the proposed clustering technique, we present in detail the steps of the clustering algorithm for the problem at hand.

The core of the clustering data concept is to identify homogeneous groups of objects based on the values of their attributes. It is in general a difficult problem to tackle and is unquestionably related to various scientific and applied fields, especially when clustering is applied to user modelling [11,12]. The problem gets more and more challenging, as input space dimensions become larger and feature scales are different from each other, as is the case in our system. In particular, a consideration of the original set of questions of the e-questionnaires as input space, results into a large number of 176 unique features to be taken into consideration when performing clustering on the user answers. The best way to go in this direction is to use a hierarchical clustering algorithm, which is able to tackle such a large scale of features [10,13]. Although such a method does not demand the number of clusters as input, still it does not provide a satisfactory framework for extracting meaningful results. This is mainly due to the "*curse of dimensionality*" that dominates such an approach, as well as the inevitable initial error propagation and complexity along with data set size issues.

In order to increase the robustness and reliability of the whole clustering step of our system, the use of an unsupervised extension to hierarchical clustering in the means of feature selection was evident [10]. Using the results of the application of this clustering to a portion of the system's dataset in question are then refined and extended to the whole dataset. The performance of the proposed methodology is finally compared to the previous step of fixed clustering, using the predefined profile characterizations as a priori label information.

The general structure of such hierarchical clustering algorithms, which forms the structure of SPERO's clustering approach as well, is summarized in the following steps and presented analytically in [10]:

1. Turn each input element into a singleton, i.e. into a cluster of a single element.
2. For each pair of clusters c1, c2 calculate their distance d(c1, c2).
3. Merge the pair of clusters that have the smallest distance.
4. Continue at step 2, until the termination criterion is satisfied. The termination criterion most commonly used is thresholding of the value of the distance.

It is worth noticing, though, that in our case, where the input space dimensions are large, the Euclidean distance is thought to be the best distance measure used [14]. Still, this is not always the case, due to the nature of the individual features; consequently a selection of meaningful features needs to be performed, prior to calculating the distance [14]. Moreover, one feature might be more important than others, while all of the features are useful, each one to its own degree. In this work we tackle weighting of features based on the following principles:

a) we expect elements of a given meaningful set to have random distances from one another according to most features, but we expect them to have small distances according to the features that relate them,
b) we select meaningful features based on the nature of the specific questions of the e-questionnaires. In particular, system experts perform an initial selection of meaningful questions, restricting the input space dimensions and
c) we further perform a second level filtering of the input data, based on the type of the input, leaving out answers – and thus questions – of arbitrary dimensions, such as free text input boxes of the e-questionnaires. Information collected from such answers fails out of the scope of clustering data and identifying user profiling information, being more useful for plain statistical approaches.

More formally, let c_1 and c_2 be two clusters of elements. Let also r_i, $i \in N_F$ be the metric that compares the *i*-th feature, and F the overall count of features (the dimension of the input space). A distance measure between the two clusters, when considering just the *i*-th feature, is given by:

$$f_i(c_1, c_2) = \sqrt[\kappa]{\frac{\sum_{\alpha \in c_1, b \in c_2} r_i(a_i, b_i)^\kappa}{|c_1||c_2|}} \tag{1}$$

where e_i is the *i*-th feature of element e, $|c|$ is the cardinality of cluster c and κ is a constant. The overall distance between c_1 and c_2 is calculated as:

$$d(c_1, c_2) = \sum_{i \in N_F} x_i(c_1, c_2)^\lambda f_i(c_1, c_2) \tag{2}$$

where x_i is the degree to which i, and therefore f_i, is included in the soft selection of features, $i \in N_F$ and λ is a constant. Based on the principle presented above, values of vector x are selected through the minimization of distance d, i.e.:

$$x_1(c_1, c_2) = \frac{1}{\sum_i \left[\frac{f_1(c_1, c_2)}{f_i(c_1, c_2)}\right]^{\frac{1}{\lambda - 1}}} \tag{3}$$

$$x_i(c_1, c_2) = x_1 \left[\frac{f_i(c_1, c_2)}{f_1(c_1, c_2)}\right]^{\frac{1}{\lambda - 1}} \tag{4}$$

Of course, when $\lambda = 1$ the solution is trivial and the feature that produces the smallest distance is the only one selected. The degree to which it is selected is 1. A more detailed approach on the issue can be found in [15].

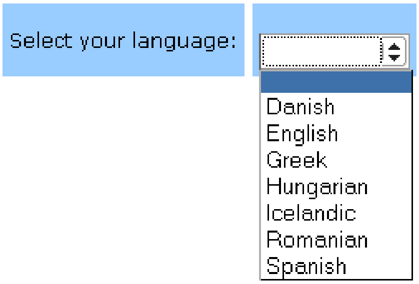


Figure 12. SPERO dataset languages.

In the following, we present the proposed algorithm implementation with our system's data set, using the Euclidean distance as the distance measure. The clustering algorithm has been applied to a small portion of the dataset, namely a 10% of the overall system's users; it contained 100 elements/users, characterized by 44 meaningful features/questions. The overall dataset consisted of answers provided by 1000 users from 8 European countries and in 7 European languages, shown in Fig. 12, on the same set of 44 selected features.

Statistical analysis and cross-correlation selection of features indicated the most important ones that were considered in the following. Although meaningful features formed a merely 25%, this proved to be accurate and efficient in the process. They correspond to a set of questionnaire questions together with their possible answering options that are summarized in the following Table 1 and have been considered indicative of the profiling extraction process. Features (column 1) are grouped by the corresponding question id (column 2) of the e-questionnaire. For each verbal question description presented in the third column of Table 1, the number of associated question ids and features used in the clustering procedure are shown in the first and second columns of Table 1 respectively.

The above elements belonged to three fixed profile classes, but this labeling information was not used during clustering; the labels were used, though, for the evaluation of the quality of the clustering procedure, prior to projecting the results to the whole data set. More specifically, each detected cluster was then assigned to the experts' provided class that dominated it. In the general case, identified clusters define specific interests and profiles, which do not necessarily correspond to the a priori known classes that are utilized during the first phase. These clusters are useful in producing collaborative recommendations of the e-learning content to the end users at a later stage. Results are shown in Table 2, Table 3 and Table 4, whereas the numbers inside parenthesis separated by commas denote the elements belonging to its one of the three profile classes in each step.

Performing the initial clustering on a mere 10% subset is not only more efficient computationally wise, it is also better in the means of quality and performance, when compared to the approach of applying the hierarchical process to the whole data set. Although clustering over this 10% of the data set resulted in different possible identifiable clusters, optimal results have been obtained for a number of nine clusters, as indicated in the following Table 2, Table 3 and Table4, where clustering results are presented for three variations of output clusters (3, 5 and 9).

Table 1. Feature selection

Feature	*QuestionID*	*Description*
1	125	Are you a teacher dedicated to or working in Special Education Needs?
2, 3, 4	126, 127, 128	Qualification/training in Special Education Needs?
5	130	Teaching Experience
6	133	Do you have a computer at home?
7	134	Do you have access to the Internet from your home?
8	139	How often do you personally use your Internet connection at home?
9, 10, 11, 12, 13	141, 142, 143, 144, 145	For which of the following did you use the computer at least once in the past month?
14, 15, 16, 17, 18, 19, 20	147, 148, 149, 150, 151, 152, 153	Which of the following tasks have you performed at least once, without any help?
21	155	Are there any computers in your work environment?
22, 23	300, 301	How often did you use the computer last week at the school?
24, 25	168, 169	Do you have access to the Internet or educational software in your work environment?
26, 27	174, 175	In your teaching, how many hours a week, on average, do you use the Internet or educational software with your students?
28, 29	176, 177	Do you use the Internet for search and retrieval of information relating to the needs and problems faced by SEN students?
30, 31	182, 183	Do you use the Internet from the school in order to find additional sources of educational material?
32, 33	400, 401	Do you use the Internet to connect with other schools?
34, 35, 36, 37	261, 262, 264, 265	Is your post permanent – temporary?
38, 39, 40, 41, 42, 43	267, 268, 270, 271, 273, 274	Age of your students
44	275	Area served by your school

Table 2. 100 users clustering results – 3 clusters

Clusters	*Elements*	*%*
1^{st}	(2, 6, **9**)	(11.77%, 35.29%, **52.94%**)
2^{nd}	(11, 2, **25**)	(28.95%, 5.26%, **65.79%**)
3^{rd}	(14, 1, **30**)	(31.11%, 2.22%, **66.67%**)

More specifically, Table 2 presents the clustering results of 100 users. The hierarchical clustering algorithm terminated by the time it reached a threshold of 3 clusters.

Table 3. 100 users clustering results – 5 clusters

Clusters	*Elements*	*%*
1st	(3, 1, 7)	(27.27%, 9.09%, 63.64%)
2nd	(5, 1, 8)	(35.72%, 7.14%, 57.14%)
3rd	(5, 1, 13)	(26.32%, 5.26%, 68.42%)
4th	(5, 9, 11)	(20.00%, 36.00%, 44.00%)
5th	(11, 1, 19)	(35.48%, 3.23%, 61.29%)

Table 4. 100 users clustering results – 9 clusters

Clusters	*Elements*	*%*
1st	(1, 1, 4)	(16.66%, 16.66%, 66.66%)
2nd	(0, 1, 6)	(0.00%, 14.28%, 85.71%)
3rd	(4, 2, 5)	(36.36%, 18.18%, 45.45%)
4th	(3, 3, 6)	(25.00%, 25.00%, 50.00%)
5th	(4, 2, 5)	(36.36%, 18.18%, 45.45%)
6th	(8, 4, 5)	(47.05%, 23.52%, 29.41%)
7th	(4, 1, 4)	(44.44%, 11.11%, 44.44%)
8th	(3, 10, 6)	(15.78%, 52.63%, 31.57%)
9th	(1, 4, 3)	(12.50%, 50.00%, 37.50%)

The first cluster comprises of 17 users, namely 2 belonging to the *Experts* class, 6 to the *Beginners* class and 9 to the *Advanced* class. The corresponding percentage distribution clearly indicates that the 9 *Advanced* dominate the first cluster. The second cluster consists of 38 users: 11 *Experts*, 2 *Beginners* and 25 *Advanced*. *Advanced* are dominant in this cluster as well, whereas in terms of percentages their domination is confident (i.e. 65,79%). Finally, the third cluster contains 45 users, 14 of whom are *Experts*, 1 is *Beginner* and 30 are *Advanced*. Domination of *Advanced* is more indicative, since a 66,67% gives them a clear advantage.

Table 3 presents the clustering results of the same 100 users, however a new threshold of 5 clusters terminates the clustering algorithm earlier. In this case, all clusters contain lesser users, in comparison to the prior case. The first cluster consists of 11 users, 3 *Experts, 1 Beginner* and 7 *Advanced*. The corresponding percentage distribution clearly indicates that the 7 *Advanced* users dominate the first cluster. The same applies to the second cluster with 14 users as well, where 5 users belong to the *Experts*, 1 user is *Beginner* and 8 users are in the intermediate state, i.e. *Advanced*. Third, fourth and fifth clusters are all resulting in the supremacy of *Advanced*; third cluster containing 19 users (distributed accordingly to 5 *Experts*, 1 *Beginner* and 13 *Advanced*), fourth cluster containing 25 users (5 *Experts*, 9 *Beginners* and 11 *Advanced*) and fifth cluster containing 31 users (11 *Experts*, 1 *Beginner* and 19 *Advanced*).

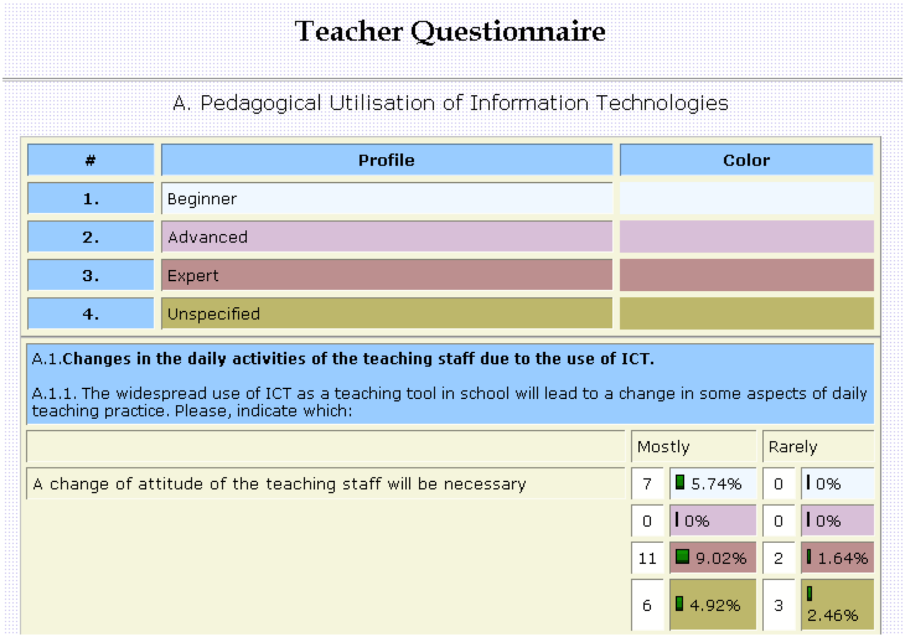
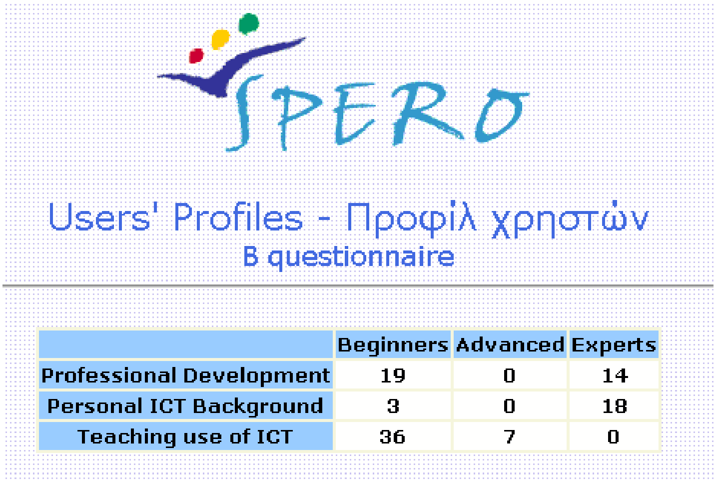


Figure 13. Learning resources linked to profile categories.

Continuing in Table 4, the results of the clustering step demonstrate the clear trend underlying in the system's input data: users are characterized by intermediate ICT skills and expertise. This observation is extremely evident in the third column of Table 4, which indicates clearly that most users of the system belong to the static, intermediate *Advanced* profile. The first two clusters identified by our algorithm are unambiguously dominated by the third profile class, i.e. *Advanced*. Additionally, clusters 3, 4 and 5 indicate a clear majority of the same third class in their elements as well. Consequently, 5 out of 9 clusters (55.55%) are indicating a clear advantage of the *Advanced*. Moreover, cluster 7 acts as an intermediary between *Advanced* and *Experts*, as it illustrates a draw in the elements between those two profile classes. Clusters 8 and 9 are dominated by the *Beginner* profile class, whereas cluster 6 forms a solid representative of the *Experts*.

The above illustrated clustering approach forms the basic procedure, with the aid of which each SPERO end user is automatically categorized to a specific profile class that characterizes his behaviour and his future interests and choices within the system. Additionally, each identified cluster is related to replies and specific comments of the e-questionnaires. We used clustering results to classify the responses of users to specific parts of the e-questionnaires, deriving information based on the users' profiles. The automatic extraction of clusters during the SPERO project provided experts with the ability to distinguish the responses of the clustered users with respect to more specific profiles. More specifically, we combined profiling information from different parts with respect to the users input, a statistical analysis of which is provided in Fig. 13, as well as analyzed results belonging to different parts of the e-questionnaires, as illustrated in Fig. 14.



	Beginners	Advanced	Experts
Professional Development	19	0	14
Personal ICT Background	3	0	18
Teaching use of ICT	36	7	0

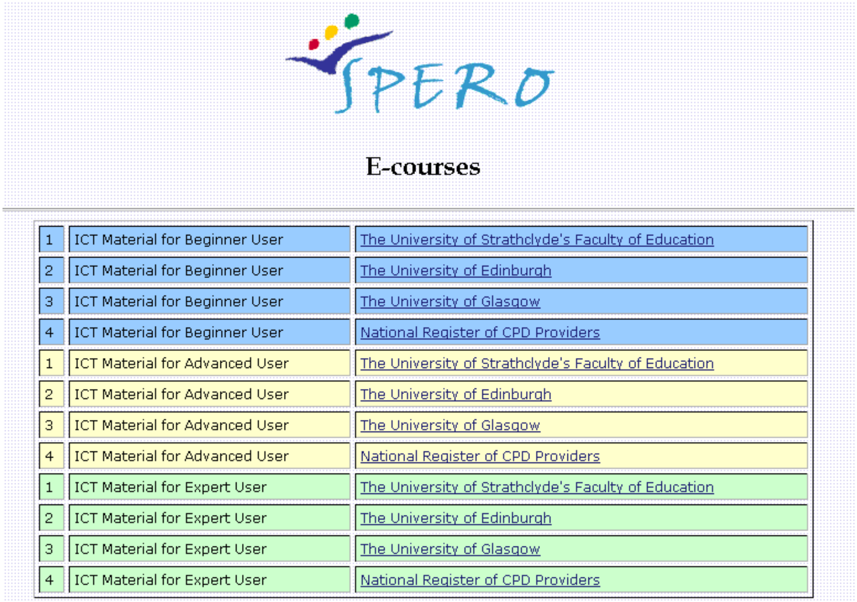
Figure 14. Combined profiling information based on clustering results.

This research approach could also be combined with neural network machine learning techniques. For instance, another work of ours described in [15] uses clustering to initialize a three-layer neural network classifier that contains personalization information for extracting the local interest of users. However, this is not considered to be the main focus of the current chapter, which focuses on utilization of clustering towards providing content collaborative recommendations to the end users. According to the cluster to which each user belongs, educational content, appropriately selected by the system's experts, is offered to him. Because of flexibility and protection of crucial personal data reasons, the step of user characterization is only provided as an added value characteristic to the users that are willing to use it. Suitable verification procedures ensure that content offering filtering features are only enabled according to each end user's will.

5. Content Adaptation and User Tracking

The SPERO system software forms an integrated, web-based learning portal, designed and implemented according to well-known learner-friendly solutions and flexible e-learning software applications [16]. When system's users visit the SPERO portal, validation against the system user database is performed. Subsequently, they are called to answer the e-questionnaires in order to automatically establish their user profile based to the intelligent clustering techniques presented in the previous sections. This automatic profile extraction provides the extremely useful and fully personalized information needed. Learning resources have been linked up to each profile category that has been defined during the profile extraction process and are illustrated in Fig. 15. The set of e-courses appropriate for each of the identified groups of user profiles is selected by the group of system's experts.

SPERO 's content offering contains links to educational content, separated into various sectors and providing services, like: Courses Catalogue, Announcement Ser-



1	ICT Material for Beginner User	The University of Strathclyde's Faculty of Education
2	ICT Material for Beginner User	The University of Edinburgh
3	ICT Material for Beginner User	The University of Glasgow
4	ICT Material for Beginner User	National Register of CPD Providers
1	ICT Material for Advanced User	The University of Strathclyde's Faculty of Education
2	ICT Material for Advanced User	The University of Edinburgh
3	ICT Material for Advanced User	The University of Glasgow
4	ICT Material for Advanced User	National Register of CPD Providers
1	ICT Material for Expert User	The University of Strathclyde's Faculty of Education
2	ICT Material for Expert User	The University of Edinburgh
3	ICT Material for Expert User	The University of Glasgow
4	ICT Material for Expert User	National Register of CPD Providers

Figure 15. Learning resources linked to profile categories.

Figure 16. Personalized e-course listing sample.

vice, Search Service, E-mail Service, Upload Files and Help Service. In particular, the main menu of the SPERO portal contains links to the following sectors/services:

- **Courses Catalogue**: It contains the titles, as well as a small textual description of one or more e-learning courses, that learners may take. An intelligent module takes over the selection of e-courses, according to user preferences and profiles, as well as their usage history. A small overview for each e-course is provided, demonstrating its main topics and concepts. A small notion of a selected e-course listing is presented in Fig. 16.

- **Announcement Service**: This service provides a bulletin board where topics about e-courses or other educational subjects are published. Relative documents, regarding e-courses outlines and requirements are posted herein. Students' and teachers' messages are presented in a threaded view layout.
- **Search Service**: It provides a search environment to facilitate information and educational materials retrievals from SPERO site, e.g. members, school units, e-lessons, e-books, e-lectures, exercises, "live" educational content broadcasts, etc.
- **E-mail Service**: SPERO users are able to send and receive e-mails through the SPERO system.
- **Upload Files**: Learners have their own personal space where they can store their own material to which other learners may or may not have access to. Several levels of authorization access are implemented.
- **Help Service**: Analytical description of the usage and tasks of SPERO menu choices. It provides information about library links and online resources outside the SPERO system and answers general Frequently Asked Questions.

In order to improve the ICT level of learners, different e-courses are also designed and implemented. Indicatively, groups of e-courses characterized by increasing difficulty and strong topic relativity are possible, such as the following chain of e-courses: *Introduction to Information's Technologies, Introduction to Operating Systems, Presentation and usage of Office and Educational software, Introduction and Usage of Internet*. Each group is characterized by the following aspects:

- **Group 1**: Introduction to Information Technologies (definition of data, bit, byte, presentation of hardware components, presentation of type of software).
- **Group 2**: Presentation and usage of operating systems.
- **Group 3**: Usage of text editor, software for work sheets, software for creation a presentation, educational software.
- **Group 4**: Usage of Internet (explorer in a browser, search machines, sending and receiving e-mails, access to a news group, access to a chat room).

The offered e-courses correspond to the available SPERO user profiles obtained at the previous step. For instance, e-courses for a learner, identified by the system as an "Expert", are depicted in Fig. 17. This is ensured by an intelligent user tracking mechanism. This mechanism is based on each user individual session, the starts and stops of which are signaled by the time the user enters and leaves the SPERO portal respectively. Session information, along with validation and user access rights is stored in the "userDetails" part of the static user profile. It provides a robust and reliable method to ensure independency amongst the system's users and efficiency of the whole user profiling-based content offering. As an illustrative example, consider that students who receive an e-course are tracked throughout the SPERO system and their behavior is observed and tracked internally. The overall procedure is transparent and provides the main source of feedback for the users' future system (e-courses and material) selections. Additionally, learners' interaction is periodically monitored to allow dynamically change of their user profiles, according to their improvement of ICT skills and skillfulness. For instance, if a user studies, then it is foreseeable that he/she will improve his/her profile. Towards that scope, it is possible for the learner to re-enter the SPERO portal and provide new data input for the e-questionnaires. Then and based on his/hers unique id, the system monitors the user's progress and changes his/hers ratings and

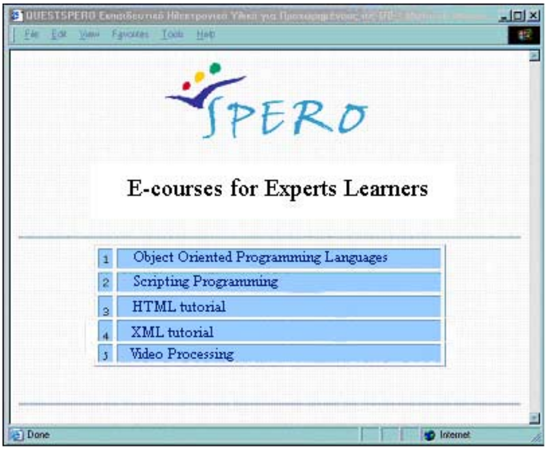


Figure 17. E-courses list for Experts Learners.

corresponding content offering. This leaves an option for the learner that the system can provide and sets him/her on a specific personalized track.

6. Conclusions and Future Work

The methodology presented in this book chapter can be exploited towards the development of more intelligent, efficient and personalized educational content offering systems, thus enhancing the learning experience of their end-users. The proposed approach forms an integrated, state-of-the-art system that is able to identify its individual users. During this process, it extends work performed on precise, high level personalization algorithms, as it utilizes (transparently to the end-user) personalization techniques towards profiling extraction, introducing a novel conjunction of static and dynamic profiling mechanisms. It was also within our intensions in this book chapter to focus on acquired representative results of this work. For more e-learning type of results the reader is encouraged to visit SPERO's online applications (http://www.image.ntua.gr/spero). Moreover, the main research effort of this chapter was to successfully introduce and apply clustering techniques in the process of user profiling and provide collaborative recommendations of e-learning content to the end users.

A major area of future research for this work is the utilization of a fuzzy relational knowledge representation model in the learners' profile weight estimation process. Our findings so far indicate, that such a combination between semantic and statistical information is possible and will have very interesting results, regarding the personalization of the educational content offered to the end-users. Additionally, the work proposed herein could also be enriched by combining it with neural network machine learning techniques. This work is part of our ongoing efforts in the field of designing and implementing an integrated, fully automated e-learning portal system. The main focus is given to the personalization aspects of the system's user handling and educational content offering. Possible future work includes better selection of the clustering algorithm threshold criteria and possible increase of the static profiles categories. Moreover, the system's e-questionnaires are susceptible to evaluation and improvements, as well as an increase in the number of participants in the e-surveys is viable. The overall proposed architecture of SPERO could be easily adapted to other e-learning schemes, mainly due to its robustness and entities clarity.

References

[1] Rosenberg J.M., 2002. E-Learning: Strategies for Delivering Knowledge in the Digital Age. McGraw-Hill, Inc, ISBN:007137809X.
[2] Karagiannidis, C., Sampson, D.G., and Cardinali, F., 2002. An architecture for Web-based e-Learning promoting reusable adaptive educational e-content. Educational Technology & Society, 5(4), pp. 27–37.
[3] Linser R., Waniganayake M. and Wilks S., 2004. Where is the Teacher? e-Learning Technology, Authority and Authorship in Teaching and Learning, In Proc. of "EDMEDIA-2004", Lugano, Swizterland.
[4] Varenne H., 2001. On pedagogical authority: One teacher's choice. [online]. Available from: http://varenne.tc.columbia.edu/hv/pol/pedagog_author.html.
[5] Linser R., 2004. Suppose you were someone else: The learning environment of a web-based role-play simulation. In SITE 2004 Conference Proceedings, Atlanta GA, USA.
[6] SPERO, 2001-2004. Leonardo Da Vinci programme, Themes 5: E-Learning. [online]. Available from: http://www.image.ntua.gr/spero.
[7] Skordalakis E., 1991. Introduction to Software Engineering. Athens.
[8] IEEE LTSC, 2002. P1484.1 Architecture and Reference Model. WG1 LTSA. Available from: http://ieeeltsc.org/.
[9] IEEE LTSC, WG11 Computer Managed Instruction, 2007. Available from: http://www.ieeeltsc.org/working-groups/wg11CMI/.
[10] Mylonas Ph., Tzouveli P. and Kollias S., 2004. Towards a personalized e-learning scheme for teachers. 4th IEEE International Conference on Advanced Learning Technologies, Joensuu, Finland.
[11] Ph. Mylonas, M. Wallace, S. Kollias, 2004. Using k-nearest neighbour and feature selection as an improvement to hierarchical clustering. Proceedings of 3rd Hellenic Conference on Artificial Intelligence, Samos, Greece.
[12] Godoy D., Amandi A., 2005. User profiling in personal information agents: a survey. The Knowledge Engineering Review, Vol. 20 (4), pp. 329–361, Cambridge University Press.
[13] Smith A.S., 2001. Application of Machine Learning Algorithms in Adaptive Web-based Information Systems. CS-00-01, School of Computing Science Technical Report Series, ISSN 1462-0871.
[14] Theodoridis S. and K. Koutroumbas, 1999. Pattern Recognition. Academic Press. San DiegoYager R.R., 2000. Intelligent control of the hierarchical agglomerative clustering process. IEEE Transactions on Systems, Man and Cybernetics, Part B 30(6): 835–845.
[15] Wallace M. and Stamou G., 2002. Towards a Context Aware Mining of User Interests for Consumption of Multimedia Documents. IEEE International Conference on Multimedia and Expo (ICME), Lausanne, Switzerland.
[16] Wallace M. and Kollias S., 2003. Soft Attribute Selection for Hierarchical Clustering in High Dimensions. International Fuzzy Systems Association World Congress (IFSA), Istanbul, Turkey.
[17] Blackboard Academic Suite™: Course Management System. Available from: http://www.blackboard.com/products/Academic_Suite/index.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

A Collaborative Filtering Approach to Personalized Interactive Entertainment using MPEG-21

Phivos Mylonas, Giorgos Andreou and Kostas Karpouzis
National Technical University of Athens
School of Electrical and Computer Engineering
Department of Computer Science
Image, Video and Multimedia Laboratory,
Zographou Campus, Athens, Greece, 157 80

Abstract. In this chapter we present an integrated framework for personalized access to interactive entertainment content, using characteristics from the emerging MPEG-21 standard. Our research efforts focus on multimedia content presented within the framework set by today's movie content broadcasting over a variety of networks and terminals, i.e. analogue and digital television broadcasts, video on mobile devices, personal digital assistants and more. This work contributes to the bridging of the gap between the content and the user, providing end-users with a wide range of real-time interactive services, ranging from plain personalized statistics and optional enhanced in-play visual enhancements to a fully user- and content-adaptive platform. The proposed approach implements and extends in a novel way a well-known collaborative filtering approach; it applies a hierarchical clustering algorithm on the data towards the scope of group modelling implementation. It illustrates also the benefits from the MPEG-21 components utilization in the process and analyzes the importance of the Digital Item concept, containing both the (binary) multimedia content, as well as a structured representation of the different entities that handle the item, together with the set of possible actions on the item. Finally, a use case scenario is presented to illustrate the entire procedure. The core of this work is the novel group modelling approach, on top of the hybrid collaborative filtering algorithm, employing principles of taxonomic knowledge representation and hierarchical clustering theory. The outcome of this framework design is the fact that end-users are presented with personalized forms of multimedia content, thus enhancing their viewing experience and creating more revenue opportunities to content providers.

Keywords. Personalization, Collaborative Filtering, MPEG-21, Digital Item, Network Management Adaptation.

Introduction

In the new era of interactive public and home entertainment, a new generation of content consumers has been born and is currently confronted with a series of technological developments and improvements in the digital multimedia content realm. Their expectations are high and the need for a high quality service problem handling is more stressful than ever. At the same time, digital video is the most demanding and complex data structure, due to its large amounts of spatiotemporal interrelations;

efficient manipulation of visual media is currently not considered a trivial task. Multimedia standards such as MPEG-4 [1], [3], [4] and MPEG-7 [2], provide important functionalities for coding, manipulation and description of objects and associated metadata; however, personalized filtering of the content, provided it is accompanied by corresponding metadata, is out of the scope of these standards, motivating heavy research efforts and the emerge of a new standard, i.e. MPEG-21 [5], [6].

Domains characterized by inherent dynamics, such as movie collections and broadcasting, make the above expectations even higher; thus, broadcasting corporations and organizations need to preserve or build up their competitive advantage, seeking new ways of creating and presenting enhanced content to their new, demanding content consumers. From the emerging mobile devices point of view, third-generation (3G) services provide the ability to transfer simultaneously both voice data (i.e. a telephone call) and non-voice data, such as downloading of a movie. In marketing 3G services, video telephony has often been used as the killer application for 3G. Both aspects will greatly enhance the multimedia content transmission and consuming potential of mobile devices. Consequently, the need for innovative services over 3G networks is large, in order to facilitate wide take up of the new technology by their end-users.

Multimedia content retrieval and filtering in the last decade has been influenced by the important progress in numerous fields such as digital content production, archiving, multimedia signal processing and analysis, as well as information retrieval. One major obstacle, though, such systems still need to overcome in order to gain widespread acceptance, is the semantic gap [7]. This refers to the extraction of the semantics of multimedia content, the interpretation of user information needs and requests, as well as to the matching between the two. This obstacle becomes even harder when attempting to access vast amounts of multimedia information and metadata contained within a movie.

In current research activities it is becoming apparent that offering of integrated personalized interactive services upon diverse and possibly heterogeneous – pre-existing – multimedia content will only be feasible through novel techniques and methodologies. In [30] for instance, a personalized content preparation and delivery framework for universal multimedia access is introduced. On the other hand, [31] focuses on a novel approach to support adaptive services for multimedia delivery in heterogeneous wireless networks.

Moreover, motion pictures (movies) continue to attract interest and are among the most popular media attractions in the world today. Consequently, multimedia applications developed for them have a huge potential market impact. Among these applications are the provision of enhanced content, statistics, dynamic interactive content and optionally advertisements. In this framework, our work targets the provision of enhanced content and statistics to provide a friendly, easily assimilated interface, as well as dynamic, personalized interactive content and advertisements to enable the user to interact with the content, thus enhancing the user experience by providing further information on the specific movie.

Our efforts resulted in an integrated framework, offering transparent, personalized access to heterogeneous multimedia content, using characteristics from the emerging MPEG-21 standard. Although recently applied in the sports domain [38], focusing on a network, device and user independent solution, this approach contributes towards bridging the gap between the semantic nature of user needs and raw multimedia documents - as expressed by movies, serving as a management mediator between end-users and movie repositories. Its core contribution relies on the fact that it provides a personalized delivery of content over heterogeneous networks and terminals, using the core functionality of the MPEG-21 standard and providing the missing link for an integrated personalized interactive experience. The latter is achieved by utilizing the notion of an MPEG-21 Digital Item [39], using it to encapsulate personalization-useful information at the multimedia content level and not at the level of terminal or system.

In this context, a user is any entity that interacts with or makes use of a Digital Item. A hybrid collaborative filtering method is then applied, based on this unified knowledge model and multimedia documents (i.e. movies) are clustered according to their ratings through clustering on their features. Future user requests are then analyzed and processed to retrieve movies from the framework's repository, according to the underlying user preferences. This chapter presents an integrated approach in the framework of the MPEG-21 standard to establish the necessary infrastructure to support the virtual value chain for personalized interactive entertainment events broadcasting over wireless, cable and digital networks, offering valuable and revenue-building services.

It should have been obvious by now that watching multimedia entertainment content at home or in public tends clearly to be a social activity. So, adaptive content providers and consumers need to adapt content to groups of users rather than to individual users. In this chapter, we discuss a hybrid strategy for combining individual user models to adapt to groups, which is basically inspired by the Social Choice Theory [37], i.e. how humans select a sequence of items (e.g. movies) for a group to watch, based on data about the individuals' preferences. The latter offers the possibility of personalized viewing experiences, based on features that pre-exist in the information accompanying each multimedia item/movie. In our framework, information on movie characteristics is derived from the Internet Movie Database (IMDB) [35].

The IMDB consists of the largest known single accumulation of data on a vast amount of multimedia content, including individual films (together with their complete cast and crew listings), television programs (including complete cast and crew listings), direct-to-video product and videogames reaching back to their respective beginnings, and worldwide in scope. Wherever possible, the information goes beyond simple screen or press credits to include uncredited personnel involved, either artistically or technically, in the production and distribution, thus aiming at completeness of detail. Furthermore, a collateral database of all persons identified in the product database exists, including biographical details and information, such as theatrical appearances, commercial advertising appearances, etc. Information is largely provided by a cadre of volunteer contributors and is considered to be the most accurate and up-to-date multimedia content database at the time of the writing of this book chapter.

Adapting this kind of multimedia content to individual viewers is a topic in itself, and a lot of research has already been done. Moreover, different domains have been identified in which a personalization process would have a great impact, such as education [32], advertising [33], and electronic program guides [34]. This research tends to build on decades of work on content-based and social filtering. As already discussed, herein we focus our efforts on exploring an even more difficult issue: adaptation of multimedia content to a group of viewers. We believe this to be essential for interactive multimedia content viewing as, in contrast to the plain use of personal computers or televisions, multimedia content viewing is largely a family or social activity. In this context, recommender systems are a special class of personalized systems that aim at predicting a user's interest on available products and services by relying on previously rated items or item features. Human factors associated with a user's personality or lifestyle, although potential determinants of user behaviour are rarely considered in the personalization process. It is a fat, that, the concept of lifestyle can be incorporated in the recommendation process to improve the prediction accuracy by efficiently managing the problem of limited data availability.

The structure of this chapter is as follows: Section 1 provides a high level overview of the proposed framework, focusing on its structure and data models. It also describes the notions of Collaborative Filtering (CF) and Hierarchical Clustering, together with a brief introduction to the MPEG-21 standard. Subsection 1.5 provides a detailed use case scenario, defining the scope of this work. Section 2 describes in detail the proposed hybrid collaborative filtering approach, based on hierarchical clustering applied on the movies' features. Continuing, section 3 discusses the basics of the MPEG-21 Digital Item utilization, followed by the corresponding resource adaptation within the proposed framework. In section 4 conclusions are drawn and some future work aspects of this work are also discussed.

1. Overview of the proposed framework

1.1. Framework architecture

The proposed framework is illustrated in Figure 1 and involves a variety of user terminals and networks, such as Personal Digital Assistants (PDA) and Personal Computers (PCs) over TCP/IP networks, Set-Top-Boxes (STB) over Local Cable TV networks, High Definition Digital Television Sets (HDTV) and networks, as well as Mobile Devices over UMTS, GPRS, or GSM networks. Given the diversity and singularity that characterize each type's multimedia content receiving, processing and displaying capabilities, specific care must be taken for its adaptability and presentation to the end-user, as depicted by the Content Preparation and Adaptation component, as well as intelligent content customization (e.g. subtitling and/or dubbing), as depicted by the Information Merging Unit. Prior Visual Enhancements Engine includes preparation of (optional) advertising content, preparation of non-standard events to be offered during a broadcast (e.g. real-time movie trivia) and designing descriptive templates for the display of optional enhanced content at transmission time. Content adaptation to any kind of end-user terminal is performed according to the so called "create once publish everywhere" principle, adapted to the targeted network and terminal prior to transmission, to allow for efficient display and manipulation on the end-user side.

On top of that, reusability of the content is considered a prerequisite in modern content management applications, although still an open and difficult issue to tackle; creating reusable multimedia content demands well-structured and efficient data models, together with highly-refined content repositories on a massive scale and is an expensive process that requires careful planning and design. Besides this, since the integrated framework handles, encodes and presents multimedia content coming from different vendors, the respective intellectual property rights (IPR) must also be retained throughout the complete process. These two additional requirements can be dealt with successfully via the inclusion of concepts presented within the emerging MPEG 21 framework, as discussed in the following sections of this chapter. Figure 1 shows only an overview of the higher-level information flow, between the different framework components. The components that undertake the task of collecting, packaging and delivering of multimedia data have been collectively enhanced to cater for the aforementioned provisions in the final content, which is transmitted to the end-user via the Video Content Transmission component. The reader is encouraged to find a detailed description of the complete process level architecture framework in [8]

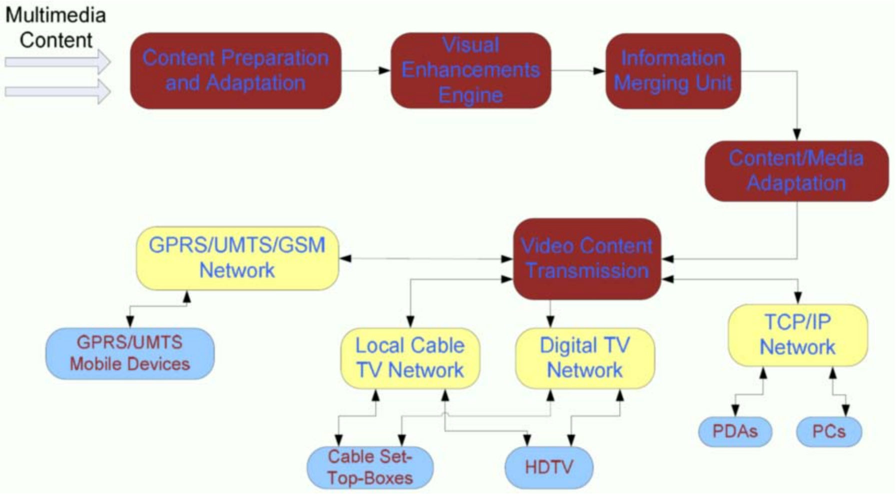


Figure 1. Overview of the proposed framework

The basic idea is that content is adapted to the different terminals and transmission networks targeted by the proposed framework and then delivered via the respective transmission channels. In the case of TCP/IP and GPRS/UMTs/GSM broadcast, the video is streamed in MPEG-4 over an MPEG-2 Transport Stream. The video resolution is then reduced to fit the lower transmission and playback capabilities of mobile terminals. Since the targeted receiver architectures offer different degrees of media delivery, interactivity and responsiveness, it is essential to break down both the captured and synthesized material to match the relevant device. As a result, different versions of the content are prepared for delivery. In the next subsections we briefly present our proposed methodology guidelines, in order to enable personalization aspects, based on the content preparation principles discussed previously.

1.2. Collaborative Filtering

In the context of bridging the gap between the content and the user and providing personalized interactive services, we implement and extend a widely-known Collaborative Filtering (CF) technique. Collaborative Filtering is the method of making automatic predictions or filtering about the interests of a user by collecting preference information from a larger pool of users [21]. The underlying assumption in all CF approaches is that users who agreed in the past, tend to agree again in the future. In the case of a collaborative filtering system for multimedia content preferences one could make predictions about which movie a user should like given a partial list of that user's preferences. These predictions are specific to the user, but use information gleaned from many users. This differs from the more simplistic approach of giving an average score for each movie of interest, for example based on its number of favouring votes.

Many variations of collaborative filtering algorithms and systems exist; however, most of them usually take two steps:

1. Look for users who share the same rating patterns with the active user (the user who the prediction is for).
2. Use the ratings from those like-minded users found in step 1 to calculate a prediction for the active user.

Alternatively, item-based collaborative filtering [16], [17] popularized by Amazon.com (i.e. "users who bought x also bought y") proceeds in an item-centric manner:

1. Build an item-item matrix determining relationships between pairs of items
2. Using the matrix, and the data on the current user, infer his/her preference

In the age of information explosion such techniques can prove very useful as the number of items in multimedia content (such as music, movies, news, web pages, etc.) have become so large that a single person cannot possibly view them all in order to select relevant ones. On the other hand, relying solely on a scoring or rating system which is averaged across all users ignores specific demands of a user, and its outcome may be particularly poor in tasks where there is large variation in interest, like movies or music recommendation. Consequently, other methods to combat information explosion must aid in the process and in the scope of this work we focused on one of them, i.e. hierarchical data clustering.

1.3. Hierarchical Clustering

The essence of clustering data is the classification of similar objects into different homogeneous groups, based on the values of their attributes. More precisely, data clustering is the partitioning of a given data set into subsets (clusters), so that the data in each subset share some common trait according to some defined distance measure. It is a problem that is related to various scientific and applied fields and has been used in science and in the field of data mining for a long time, with applications of techniques ranging from artificial intelligence, machine learning, data mining and pattern recognition, to image analysis, bioinformatics, databases and statistics [18].

There are different types of clustering algorithms for different types of applications and a common distinction is between hierarchical and partitioning clustering algorithms. Hierarchical algorithms find successive clusters using previously established clusters, whereas partitioning algorithms determine all clusters at once. Although hierarchical clustering methods are more flexible than their partitioning counterparts, in that they do not need the number of clusters as an input, they are less robust in initial error propagation and computational complexity issues and thus must be used with caution and under specific circumstances, as depicted in the following. In general, clustering of data is still considered an open issue, basically because it is difficult to handle in the cases that data is characterized by numerous measurable features, as in the case of movie features.

1.4. MPEG-21 Digital Items

The basic architectural concept in MPEG-21 is the Digital Item. Digital Items are structured digital objects, including a standard representation, identification and metadata. They are the basic unit of transaction and distribution in the MPEG-21 framework [22]. More concretely, a Digital Item is a combination of resources (such as videos, audio tracks, images, etc), metadata (such as descriptors, identifiers, etc), and structure (describing the relationships between resources). The second part of MPEG-

21 (ISO/IEC 21000-2:2003) specifies a uniform and flexible abstraction and interoperable schema for declaring the structure and makeup of Digital Items. Digital Items are declared using the Digital Item Declaration Language (DIDL) and declaring a Digital Item involves specifying its resources, metadata and their interrelationships. In this context, complex digital objects, as the ones containing multimedia content feature information used in the presented hybrid CF approach, may be declared using the notion and language of a Digital Item. DIDL language defines the relevant data model via a set of abstract concepts, which form the basis for an XML [23] schema that provides broad flexibility and extensibility for the actual representation of compliant data streams.

The usage of the MPEG-21 Digital Item Declaration Language to represent such complex digital objects, has introduced benefits to the proposed framework in two major areas: The management of the initial content presentation and the management and distribution of multimedia content, such as video, images and metadata. The platform allows the creation of predefined templates during the planning process before broadcasting; these templates form the Initial Scene that is used to generate the initial MPEG-4 scene. During the broadcasting phase, templates are used in order to control the real time updates of the MPEG-4 content. Having all the information packaged in one entity, i.e. initial scene, customization points, etc. brings the benefit of reduced complexity data management.

Furthermore, the benefit from the adoption of MPEG-21 is that every Digital Item can contain a specific version of the content for each supported platform. The dynamic association between entities reduces any ambiguity over the target platform and the content. Having all the necessary information packaged in one entity enables the compilation and subsequent adaptation of a Digital Item to be performed only once (during its creation) and not on a per-usage basis, thereby effectively eliminating the need for storage redundancy and bringing the benefit of reduced management and performance complexity in the Information Repository. The adopted MPEG 21 concepts and their structure are described in detail in subsequent sections of this chapter.

1.5. A Use Case Scenario

At this point let us assume that the multimedia content offered to the end-users of the proposed standalone system contains a set of movies to choose from. These can be movies whose main genre is comedy, drama, science fiction, etc. For the sake of simplicity, we utilize only a part of the IMDB movie information. More specifically, we take into consideration only the subset of 14 movie attributes presented in Table 1.

Let the end-users have preference ratings over the set of movies, either for a specific movie or for a group of movies (i.e. cluster of movies), in the following 1-10 scale-based manner: 1 is used to denote a really negative preference, i.e. “really hate”, whereas 10 denotes a “really like” preference. The basic problem is which movies the content provider should offer to new users, given the preferences of existing users over the set of offered movies.

A simplified example of this situation may be given as follows: three end-users, John, George and Mary are already watching their favourite movies: John has invited his friends at home to watch a comedy film in his new home cinema, George is travelling by train and watches his favourite drama movie in his PDA and Mary is waiting for her turn in the doctor’s office watching a drama movie on a Set-Top-Box.

All three of them have established their user preferences for a set of ten (randomly selected) movies (A to J), that include the three that are currently viewing. As expected, each end-user has a different view on the quality of the 10 selected movies and rates them according to his/her subjective criteria. A new user, Tom, opens his personal computer and requests from the content provider the top movies, according to system's user ratings, to select from.

Table 1. List of the property restrictions applied to the example class Volvo

#	Feature
1.	Actor
2.	Actress
3.	Director
4.	Genre
5.	Language
6.	Location
7.	MPAA Rating
8.	Plot summary
9.	Producer
10.	Rating
11.	Release Date
12.	Running Time
13.	Title
14.	Writer

2. Personalization and Filtering: a Hybrid Approach

One of the technical novelties introduced in the proposed framework is the handling of its users in a personalized manner, by building different profiles according to their preferences. The system is able to provide each user personalized multimedia content according to his/her specific user profile; a functionality provided considering a hybrid collaborative filtering methodology, based on hierarchical clustering on content information acquired by all participating content material. Current section of this work presents the design and implementation of the profile-based framework, which matches content to the context of interaction, so that the latter can be adapted to the user's needs and capabilities.

In this context, we introduce a novel hybrid collaborative filtering approach, based on multimedia content clustering. More specifically, we apply traditional data mining techniques, such hierarchical clustering on the multimedia content itself (i.e. movies), according to a predefined set of features. This set includes movies' characteristics, such as movie genre, filming date, movie type, etc and is distinctive of the content. All this information is encapsulated within the Digital Item concept of MPEG-21, to ensure interoperability and robustness of the overall approach, as well as network and terminal independency. The latter is achieved through the adoption of the MPEG-21 standard and the lack of a single centralized system database; quite on the contrary, all necessary information is content- and user-centric, decentralized to all participating user terminals.

In the remaining of this section, we present a brief overview of our hybrid collaborative filtering clustering technique, together with a detailed description of the proposed algorithm.

2.1. Hierarchical Clustering Algorithm

In order for the proposed framework to provide the discussed kind of personalized access to interactive entertainment, a number of steps need to take place. The first step in identifying the suitable set of top ranked movies in the system is to cluster them according to the set of features under consideration. This step is necessary in order to identify homogeneous patterns in the movie data set that will aid in the personalization process in terms of selection speed and quality. As already discussed, the main problem a clustering technique is asked to solve is the identification of homogeneous groups of objects based on the values of their attributes. In general, this is a difficult problem to tackle, related to various scientific fields, especially when clustering is applied to user modelling [9], [10]. The problem gets more and more challenging, as input space dimensions become larger and feature scales are different from each other, as is the case of our framework. In particular, a consideration of the original set of movie characteristics, as described in the Internet Movie Database [20] and Movielens [19] as input space, results into a large number of unique features to be taken into consideration when performing clustering on this kind of multimedia content, i.e. movies.

The best way to go in this direction is to use a hierarchical clustering algorithm, which is able to tackle such a large scale of features [11], [12]. Although such a method does not demand the number of clusters as input, still it does not provide a satisfactory framework for extracting meaningful results. This is mainly due to the “curse of dimensionality” that dominates such an approach, as well as the inevitable initial error propagation and complexity along with data set size issues. The “curse of dimensionality” is a term applied to the problem caused by the rapid increase in volume associated with adding extra dimensions to a mathematical space. As an example consider that having 100 observations covering the one-dimensional unit interval [0,1] on the real line provides a quite well performance; however considering the corresponding 10-dimensional unit hypersquare, the 100 observations would be isolated points in a huge empty space. To get similar coverage to the one-dimensional space would now require 10^{20} observations, which is at least a massive undertaking.

In this context and in order to increase the robustness and reliability of the whole clustering step of our system, the use of an unsupervised extension to hierarchical clustering in the means of feature selection is evident [11]. The results of the application of this clustering to only a portion of the movie dataset in question are then refined and extended to the whole dataset. In this approach we follow the standard hierarchical clustering algorithm structure:

1. We start from turning each movie into a singleton m_i (i.e. a cluster containing a single movie).
2. Then, for each pair of clusters m_1, m_2 we calculate their distance $d(m_1, m_2)$.
3. Identifying the smallest distance amongst all possible pairs of clusters results into the merging of this pair.

4. The above described process is repeated from step 2, until a meaningful clustering termination criterion is satisfied; the termination criterion most commonly used is a meaningful threshold on the clusters' distance value $d(m_1, m_2)$.

As it is typical in cases where the input space dimensions are large, the Euclidean distance is considered to be the best distance measure used [13]. Still, this is not always the case, due to the nature of the individual features; consequently a selection of meaningful features needs to be performed, prior to calculating the distance $d(m_1, m_2)$ [14]. On the one hand, one feature might be more important than others, while at the same time all features are useful to some degree. As a result and according to these principles, an additional meaningful weighting of features is followed within our approach. The key element of the above algorithm is the ability to define a unique distance among any pair of clusters, given the input space and the clustering features. More formally, letting m_1 and m_2 be two clusters of movies, we propose the following distance measure when considering just the i-th feature:

$$f_i(m_1, m_2) = \sum_{i \in \mathbb{N}_F} \sqrt[\mu]{\frac{\sum_{x \in m_1, y \in m_2} r_i(x_i, y_i)^{\mu}}{|m_1||m_2|}} \tag{1}$$

where $r_i, i \in \mathbb{N}_F$ is the metric that compares the i-th feature, F the overall count of features, $|m_1|$ the cardinality of cluster m_1 and μ a constant. Obviously, $\mu = 1$ approaches the mean value and $\mu = 2$ yields the Euclidean distance. The overall distance between m_1 and m_2 is calculated as:

$$d(m_1, m_2) = \sum_{i \in N_F} x_i(m_1, m_2)^{\lambda} f_i(m_1, m_2) \tag{2}$$

where x_i is the degree to which i, and therefore f_i, is included in the soft selection of features, $i \in N_F$ and λ is a constant. When $\lambda = 1$ the solution is trivial and the feature that produces the smallest distance is the only one selected with degree equal to 1. For the sake of simplicity, the interested reader is encouraged to read the detailed approach on the issue that can be found in [15]. Finally, the above described clustering approach creates crisp clusters of movies and does not allow for overlapping among the detected clusters. Thus, it forms the basic procedure, with the aid of which movies are automatically categorized to a distinct group class that will be used during the collaborative filtering step of this approach. The movies' clustering characterizes the users' behaviour and future interests and choices of multimedia content.

2.1.1. Hierarchical Clustering Algorithm Implementation

In this section, we examine the implementation of the proposed hierarchical clustering algorithm using system's movie data set and the Euclidean distance measure. The clustering algorithm has been applied to a small portion of the data set, namely a 10% of the overall movies; it contained 100 elements (movies), characterized by 14 meaningful features. These features have been considered appropriate for the personalization process and were selected a priori by a group of experts. Identified clusters define specific interests and preference information. These clusters are useful in producing collaborative recommendations of the multimedia content to the end-users at the later request stage, as described in section 3. Results are shown in Table 2, Table 3 and Table 4, whereas the letters inside parenthesis separated by punctuation marks denote the movies belonging to its cluster in each step.

Performing the initial clustering on a mere 10% subset is not only more efficient computationally wise, it is also better in the means of quality and performance, when compared to the approach of applying the hierarchical process to the whole data set. Although clustering over this 10% of the data set resulted in different possible identifiable clusters, optimal results have been obtained for a number of nine clusters, as indicated in the following tables, where clustering results are presented for three variations of output clusters (3, 5 and 9):

More specifically, Table 2 presents the clustering results of 100 learners. The hierarchical clustering algorithm terminated by the time it reached a threshold of 3 clusters. The first cluster comprises of 17 learners, namely 2 belonging to the *Experts* class, 6 to the *Beginners* class and 9 to the *Advanced* class. The corresponding percentage distribution clearly indicates that the 9 *Advanced* dominate the first cluster. The second cluster consists of 38 learners: 11 *Experts*, 2 *Beginners* and 25 *Advanced*. *Advanced* are dominant in this cluster as well, whereas in terms of percentages their domination is confident (i.e. 65,79%). Finally, the third cluster contains 45 learners, 14 of whom are *Experts*, 1 is *Beginner* and 30 are *Advanced*. Domination of *Advanced* is more indicative, since a 66,67% gives them a clear advantage.

Table 2. 100 movies clustering results – 3 clusters

Clusters	Elements
1^{st}	17
2^{nd}	38
3^{rd}	45

Table 3 presents the clustering results of the same 100 movies; however a new threshold of 5 clusters terminates the clustering algorithm earlier. In this case, all clusters contain lesser learners, in comparison to the prior case. The first cluster consists of 11 movie items. The corresponding percentage distribution clearly indicates that the 7 movies dominate the first cluster. The same applies to the second cluster with 14 movies as well. Third cluster contains 19 movies, fourth cluster contains 25 movies and fifth cluster contains 31 movies.

Table 3. 100 movies clustering results – 5 clusters

Clusters	Elements
1^{st}	11
2^{nd}	14
3^{rd}	19
4^{th}	25
5^{th}	31

Continuing, in Table 4 are presented the results of the clustering step terminating in 9 movie clusters:

Table 4. 100 movies clustering results – 5 clusters

Clusters	Elements
1st	6
2nd	7
3rd	11
4th	12
5th	11
6th	17
7th	9
8th	19
9th	8

2.2. Collaborative Filtering

Our CF algorithm recommends movies to the active user based on the ratings to the previously clustered movie titles of *n* other users. It it summarized in the following principles:

a) Let the set of all movie titles be M and the rating of user i for title j as $r_i(j)$. The function $r_i(j): M \to \Re \cup \{\perp\}$ maps titles to real numbers or to $\perp$, the symbol for "no rating."
b) Denote the vector of all of user i's ratings for all titles as $r_i(M)$
c) Denote the vector of all of the active user's ratings as $r_a(M)$.
d) Define $NR \subset M$ to be the subset of titles that the active user has not rated, and thus for which we would like to provide predictions. That is, title j is in the set NR if and only if $r_a(j) = \perp$.
e) Then the subset of titles that the active user has rated is M-NR.
f) Define the vector $r_i(S)$ to be all of user i's ratings for any subset of titles $S \subseteq M$, and $r_a(S)$ analogously.
g) Finally, denote the matrix of all users' ratings for all titles simply as r. In general terms, a collaborative filter is a function *f* that takes as input all ratings for all users, and outputs the predicted ratings for the active user:

$$r_a(NR) = f(r_1(M),\ r_2(M),\ \dots,\ r_n(M)) = f(r) \tag{3}$$

where the $r_i(M)$'s include the ratings of the active user.

2.2.1. Collaborative Filtering Implementation

End-users have preference ratings over the set of clustered movies in the following 1-10 scale-based manner: 1 is used to denote a really negative preference, i.e. “really hate”, whereas 10 denotes a “really like” preference. The basic problem is which movies should the content provider offer to new users, based on the ratings of existing users. Following this principle, we provide an example of 3 individual user ratings over the identified 9 clusters on the subset of 100 movies, as depicted in Table 5:

Table 5. Example ratings for a group of viewers – MC: Movie Cluster

	MC1	MC2	MC3	MC4	MC5	MC6	MC7	MC8	MC9
User 1	10	4	3	6	10	9	6	8	8
User 2	1	9	8	9	7	9	6	9	3
User 3	10	5	2	7	9	8	5	6	7

Many strategies, also called “social choice rules” or “group decision rules” have been devised for reaching group decisions given individual opinions. The one followed herein originates from the Social Choice Theory and will be illustrated with the example introduced above. Table 6 shows the “group preference ranking/rating” resulting from the strategy, a sequence indicating in which order movie clusters would be chosen, when a new end-user requests a movie rating. In this approach, utility values for each alternative are used, instead of just using ranking information as in other approaches (e.g. in the “plurality voting” approach). More specifically, ratings are added, and the larger the sum the earlier the alternative appears in the final movie rating sequence. This strategy is widely spread and used also in a variety of systems and approaches, such as multi-agent systems [36].

Table 6. Example group ratings for a new user – MC: Movie Cluster

	MC1	MC2	MC3	MC4	MC5	MC6	MC7	MC8	MC9
User 1	10	4	3	6	10	9	6	8	10
User 2	1	9	8	9	7	9	6	9	3
User 3	10	5	2	7	9	8	5	6	7
Group	21	18	13	22	26	26	17	23	20
Group Rating	(MC5, MC6), MC8, MC4, MC1, MC9, MC2, MC7, MC3								

2.3. Personalization using MPEG-21 concepts

According to the previously analyzed methodology, the system provides end-users with the possibility to see only movies and information about movies that they are interested in. One flexible way to perform content personalization is to filter the content that is streamed to the client. In the case of a STB display, since the same content is broadcast to all clients, filtering should occur at the client side, i.e. on the STB Mary is watching a drama movie before its presentation. The MPEG-21 framework is used for personalization and content filtering in the following way: Mary’s STB contains an MPEG-21 DIA Description that specifies her user preferences on content. When her user terminal receives multimedia content, this is filtered according to its genre and

Mary's user preferences indicated in the DIA Description. The main issue is to find a way to transport synchronously multimedia content and its associated metadata indicating its genre, in order to make sure that the multimedia content is not received before its description. One way to achieve this is by grouping the multimedia content and its genre within a DID, and to stream the complete DID to the clients. In the case of Mary, it safe to assume that this DID indicates that the multimedia content belongs to the genre "Drama". Obviously, according to the user preferences of a "Comedy"-based DIA Description, the multimedia content will be filtered out by the client terminal and therefore not displayed, whereas in the case of a "Drama"-based DIA Description preferences, it will be promoted and presented to Mary. The same applies to John and George at home and on a train, respectively.

3. MPEG-21 Digital Item Utilization

3.1. Digital Item Declaration

The task of creating a robust architecture framework for creating and delivering of diverse multimedia content has been in the past and currently continues to be an ambitious mission. MPEG-21 introduced the Digital Item (DI), a new interoperable unit for multimedia delivery and transaction. As in any environment that proposes to facilitate a wide range of actions involving "Digital Items", there is a need for a very precise description for defining exactly what constitutes such an "item". The basic concept of a DI is essentially a container for all kinds of metadata and content and at the first glance might look partially similar to work undertaken in other fields, such as e-learning [24]. However, in MPEG-21, a complete, flexible and rich delivery framework based around a more versatile DI specification has been standardized at a higher level framework.

The general structure of a DI is provided by a Digital Item Declaration (DID) [25]. A DID is a document that specifies the makeup, structure and organization of a DI. The DID formally expresses and identifies the content and the metadata, that are considered to be the components of the DI. Further, the DID binds together groups of resources and metadata, i.e. movies together with their characteristic features. In the case of the proposed framework, the DIs follow the standardized MPEG-21 principle elements, where items like genre and/or user ratings are grouped together into components that are grouped into a container; an example of a DI declaration code is depicted in Figure 2.

The DID opens with the XML namespace declarations familiar to users of XML and the root DIDL element. Within this DIDL element we have a single Item which has an id attribute, allowing external or internal referencing of the Item. Item identifiers of any level (i.e. DII Identifier elements) are not included here for brevity. Human readable text Descriptor/Statement combinations (such as the title, date, genre, plot outline and user rating of the movie) follow in order to provide interoperability.

3.2. Digital Item Identification

Up to now, digital identification in multimedia documents came in usual ways, such us nested digital sign in file headers and declared types in file extensions. The lack of homogeneity as it was developed by the diversity of media formats and the richness of multimedia content itself provoked as much confusion as opacity in the multimedia chain. On the contrary, the MPEG-21 vision of interoperability and transparency, leads the way to fill the gap between the different technologies. As already discussed, MPEG-21 objects, like "item" and "user," become the main entities, in order to describe the general idea in a more transparent and less complicated manner.

```
<?XML VERSION="1.0" ENCODING="utf-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="urn:mpeg:mpeg21:2002:01-DIDL-NS didl.xsd"
xmlns:dii="urn:mpeg:mpeg21:2002:01-DII-NS">
<ITEM ID="Movie #1">
  <DESCRIPTOR>
    <STATEMENT TYPE="text/plain">Title: The Devil Wears Prada</STATEMENT>
  </DESCRIPTOR>
  <ITEM>
    <DESCRIPTOR>
      <STATEMENT TYPE="text/plain">
        Date: 2006
      </STATEMENT>
    </DESCRIPTOR>
  </ITEM>
  <ITEM>
    <DESCRIPTOR>
      <STATEMENT TYPE="text/plain">
         Plot Outline: A naive young woman comes to New York and scores a job as the
     assistant to one of the city's biggest magazine editors, the ruthless and cynical
     Miranda Priestly.
      </STATEMENT>
    </DESCRIPTOR>
  </ITEM>
  <ITEM>
    <DESCRIPTOR>
      <STATEMENT TYPE="text/plain">
        User rating: 7.0/10.0
      </STATEMENT>
    </DESCRIPTOR>
  </ITEM>
</ITEM>
</DIDL>
```

Figure 2. Example of a Digital Item Description.

The role of Digital Item Identification (DII) is not only to propose the way to identify DIs in a unique manner, but also to distinguish different types of them. These Identifiers are placed in a specific part of the DID, which is the statement element, and they are associated with DIs: DIs are identified by encapsulating uniform resource identifiers, which are a compact string of characters for identifying an abstract or physical resource. The elements of a DID can have zero, one or more descriptors; each descriptor may contain a statement which can contain an identifier relating to the parent element of the statement. Besides the references to the resources, a DID can include information about the item or its parts. An example about the metadata that a movie could have in MPEG-21 within our framework is visualized in Figure 3. As obvious from the figure, it is necessary for DII to allow differentiating between the different schemas that users can use to describe content in general. Consequently, MPEG-21 DII uses the XML [23] mechanism to achieve this.

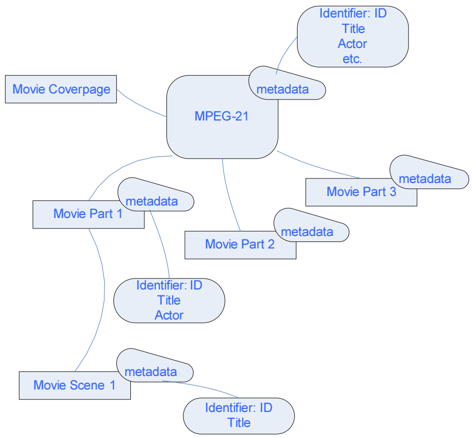


Figure 3. Visualization example of a movie Digital Item implementation**.**

3.3. Multimedia resource adaptation

The focus of resource adaptation is the framework of DIA (Digital Item Adaptation), where messages between servers and end-users are in the form of XML documents with URL links to resources or encoded binary data. In the case of linked resources, a Digital Resource Provider decides which variation of the resource is best suited for the particular user, based on the user's terminal capabilities, the environment in which the user is operating and the available resource variations. In our use case scenario, for example, where George views his favourite drama movie travelling on a train, i.e. a streaming media resource, adaptation will depend on the available bandwidth, screen size, audio capabilities and available viewer software in his PDA terminal, all part of an automated process, as capabilities and preferences should be automatically extracted and enforced.

DIA is the key element in order to achieve transparent access to distributed advanced multimedia content, by shielding end-users like George from network and terminal installation, management and implementation issues. The latter enables the provision of network and terminal resources on demand to form user communities where multimedia content can be created and shared, always with the agreed/contracted quality, reliability and flexibility, allowing the multimedia applications to connect diverse sets of users, such that the quality of the user experience will be guaranteed.

Towards this goal the adaptation of DIs is required. In the described platform dynamic media resource adaptation and network capability negotiation is especially important for the mobile paradigm (the George/PDA paradigm) where users (George) can change their environment (i.e. locations, devices etc) dynamically (e.g. get off the speeding train or request the same content for his mobile phone as well). MPEG-21 addresses the specific requirements by providing the DIA framework, for scalable video streaming [25]. The DIA framework, specifies the necessary mechanisms related to the usage environment and to media resource adaptability. This approach was the one adopted for the proposed platform. Alternative approaches to this issue may be the

HTTP and RTSP based [27], [26], [29] or the agent based content negotiation mechanisms [28].

The DIA framework, regarding resource adaptation, includes the Usage Environment Description Tools (i.e. User Characteristics, Terminal Capabilities, Network Characteristics, Natural Environment Characteristics) and the Digital Item Resource Adaptation Tools (i.e. Bitstream Syntax Description - BSD, Adaptation QoS, Metadata Adaptability). The latter are the main tools, which enable resources adaptation. BSD language provides information about the high level structure of bitstreams so that streaming can be modified accordingly to this information. Adaptation QoS schema provides the relationship between QoS parameters (e.g., the current network interface bandwidth in the case of Mary or George's PDA computational power) and the necessary adaptation operations needed to be executed for satisfying these parameters. The associated video or media quality, which is the outcome of the adaptation procedure, is also included as parameter in the adaptation schema.

4. Conclusions

The core contribution of this work has been the provision of an integrated framework for personalized access to heterogeneous interactive entertainment multimedia content, using characteristics from the emerging MPEG-21 standard. It contributed to the bridging of the gap between the raw content and the end-user over a variety of networks and terminals. This is accomplished by implementing a novel collaborative filtering approach. It utilized a hierarchical clustering algorithm towards the scope of group modeling implementation, illustrating at the same time the benefits from the use of MPEG-21 standard components, such as the Digital Items. Finally, a real-life use case scenario is presented to illustrate the entire procedure.

The methodology presented in this book chapter can be exploited towards the development of more intelligent, efficient and personalized multimedia content access systems, thus enhancing the viewing experience of end-users. In order to verify its efficiency, we plan to have it thoroughly tested in the framework of a future multimedia retrieval, personalization and filtering Digital TV application. Another interesting perspective for future work is the utilization of end-users' usage history. Additionally, design and efficiency issues improvement of knowledge representation and hierarchical clustering process presented herein lies within our first research priorities.

References

[1] Battista S., Casalino F., Lande C., "MPEG-4: A Multimedia Standard for the Third Millenium, Part 2", IEEE Multimedia, Volume 7 (1), pages 76-84, 2000.
[2] Sikora T., "The MPEG-7 Visual standard for content description - an overview", IEEE Trans. on Circuits and Systems for Video Technology, special issue on MPEG-7, 11(6):696-702, June 2001.
[3] Coding of Audio-Visual Objects (MPEG-4) – Part 1: Systems, ISO/IEC 14496-1:2001.
[4] ISO/IEC JTC1/SC29/WG11/N3382 14496-1:2001 PDAM2 (MPEG-4 Systems), Singapore, March 2001.
[5] MPEG-21 Overview v.5, ISO/IEC JTC1/SC29/WG11/N5231, Shanghai, October 2002, http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm.

[6] MPEG-4 Overview v.21, ISO/IEC JTC1/SC29/WG11 N4668, March 2002, http://www.chiariglione.org/mpeg/standards/MPEG-4/MPEG-4.htm.
[7] 13. Zhao R. and W.I. Grosky, Narrowing the Semantic Gap-Improved Text-Based Web Document Retrieval Using Visual Features, IEEE Trans. on Multimedia, Special Issue on Multimedia Database, Vol. 4, No 2, June 2002.
[8] E. Papaioannou, K. Karpouzis, P. de Cuetos, V. Karagianis, H. Guillemot, A. Demiris, N. Ioannidis, “MELISA – A Distributed Multimedia System for Multi-Platform Interactive Sports Content Broadcasting”, Proceedings of EUROMICRO Conference, pp. 222-229, 2004.
[9] Paliouras G., Papatheodorou C., Karkaletsis V. and Spyropoulos C.D., “Discovering User Communities on the Internet Using Unsupervised Machine Learning Techniques. Interacting with Computers’, v. 14, n. 6, pp. 761-791, 2002.
[10] Smith A. S. “Application of Machine Learning Algorithms in Adaptive Web-based Information Systems”,CS-00-01, School of Computing Science Technical Report Series. ISSN 1462-0871, 2001
[11] Miyamoto, S. “Fuzzy Sets in Information Retrieval and Cluster Analysis”. Kluwer Academic Publishers,1990
[12] Duda, R. O., Hart, P. E., & Stork, D. G., “Pattern Classification”, Second Edition, Wiley,2001
[13] Yager R.R., “Intelligent control of the hierarchical agglomerative clustering process”, IEEE Transactions on Systems, Man and Cybernetics, Part B 30(6): 835–845,2000
[14] Wallace, M., Stamou, G. “Towards a Context Aware Mining of Learner Interests for Consumption of Multimedia Documents”, Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Lausanne, Switzerland,2002
[15] Wallace, M. and Kollias, S., "Soft Attribute Selection for Hierarchical Clustering in High Dimensions", Proceedings of the International Fuzzy Systems Association World Congress (IFSA), Istanbul, Turkey, June-July 2003
[16] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In Proc. of the 10th International World Wide Web Conference (WWW10), Hong Kong, May 2001
[17] Vucetic S. and Obradovic Z., “A Regression-Based Approach for Scaling-Up Personalized Recommender Systems in E-Commerce," Web Mining for E-Commerce Workshop at the Sixth ACM SIGKDD Inl'l Conf. on Knowledge Discovery and Data Mining, Boston, MA., 2000
[18] Hirota, K., Pedrycz, W. (1999) Fuzzy computing for data mining. Proceedings of the IEEE 87:1575–1600
[19] MovieLens, Movie recommendation website, http://movielens.umn.edu
[20] The Internet Movie Database, http://www.imdb.com/
[21] D.M. Pennock and E. Horvitz. Analysis of the axiomatic foundations of collaborative filtering. In AAAI Workshop on Artificial Intelligence for Electronic Commerce, Orlando, Florida, July 1999.
[22] MPEG-21 Overview v.5, ISO/IEC JTC1/SC29/WG11/N5231, Shanghai, October 2002, http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm.
[23] http://www.w3.org/XML/
[24] K. Verbert and E. Duval, “Toward a global architecture for learning objects: A comparative analysis of learning object content models,” in Proc. ED-MEDIA 2004 World Conf. Educational Multimedia, Hypermedia and Telecommunications, 2004, pp. 202–209.
[25] Information Technology-Multimedia Framework (MPEG-21)-Part 2: Digital Item Declaration, ISO/IEC 21000-2:2003, Mar. 2003.
[26] Apache Group, “Apache HTTP Server: Content Negotiation”, http://httpd.apache.org/docs/content-negotiation.html.
[27] K. Holtman and A. Mutz, “Transparent Content Negotiation in HTTP”, IETF RFC 2295, 1998.
[28] K. Munchurl, L. Jeongyeon, K. Kyeongok, and K. Jinwoong, “Agent-Based Intelligent Multimedia Broadcasting within MPEG-21 Multimedia Framework”, ETRI Journal, Volume 26, Number 2, April 2004, pp 136-148.
[29] S. Seshan, M. Stemm, R. H. Katz, “Benefits of transparent content negotiation in HTTP”, Global Internet Mini Conference, Globecom 98, November 1998
[30] O. Steiger, M. Schneider Fontan, D. Marimon Sanjuan, Y. Abdeljaoued, T. Ebrahimi, S. Dominguez, J. San Pedro Wandelmer, N. Denis, F. Granelli and F. De Natale, “Personalized Content Preparation and Delivery for Universal Multimedia Access”, No 2005.010, April 2005
[31] S. Kalasapur, M. Kumar, B. Shirazi, “Personalized Service Composition for Ubiquitous Multimedia Delivery”, In Proceedings of the Sixth IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM'05), 2005
[32] J. Masthoff and R. Luckin, “Preface: Workshop Future TV: Adaptive instruction in your living room”, In Proceedings of the Future TV: Adaptive Instruction In Your Living Room (A workshop for ITS), 2002

[33] G. Lekakos, K. Chorianopoulos, and D. Spinellis, "Information systems in the living room: A case study of personalized interactive TV design", In Proceedings of the 9th European Conference on Information Systems, June 2001.
[34] Ardissono, L., Kobsa, A., Maybury, M. T. (Eds.), 2004, Personalized Digital Television, Targeting Programs to Individual Viewers, Vol. 6, p. 331
[35] http://www.imdb.com
[36] L. M.J. Hogg, N. R. Jennings, "Socially Intelligent Reasoning for Autonomous Agents", IEEE Transactions on Systems, Man and Cybernetics – Part A , 31(5):381--399, 2001.
[37] Kenneth J. Arrow, "Social Choice and Individual Values", 1951, 2nd ed., 1963, ISBN 0-300-01364-7
[38] M. Papaioannou, K. Pardalis, K. Karpouzis, G. Andreou, N. Ioannidis, "Possible usages of MPEG-21 to improve the MELISA sports broadcasting system", ISO/IEC JTC1/SC29/WG11/M11675, 71st MPEG meeting, Hong Kong, China, January 2005
[39] K. Karpouzis, I. Maglogiannis, E. Papaioannou, D. Vergados, A. Rouskas, "MPEG-21 digital items to support integration of heterogeneous multimedia content", Special issue on Emerging Middleware for Next Generation Networks, Computer Communications, Elsevier, 2006.

This page intentionally left blank

Part III: Intelligent Applications in eHealth

Ilias Maglogiannis
Department of Information and Communication Systems Engineering
University of the Aegean Greece
imaglo@aegean.gr

During the last decades, there has been a significant increase in the level of interest regarding the use of artificial intelligence tools in medicine and healthcare provision. As the development of electronic health systems (the field known as e-health) expands to support more clinical activities, healthcare organizations are asking physicians and nurses to interact increasingly with computer systems to perform their duties. In this context the field of computational analysis and artificial intelligence in medicine attracts a lot of researchers working in the AI domain. The majority of research efforts in this area involve the development of diagnostic tools, designed to support the work of medical professionals. Expert systems and machine learning algorithms are used to provide second opinions in diagnosis, while intelligent tele-medicine and tele-health applications cover the need of constant medical supervision of habitants in remote, isolated and underserved locations. Medical and biomedical informatics utilizing data mining, computational analysis, visualization, and simulation of data and experiments advance the discovery of new knowledge regarding the biological and therapeutical procedures.

This part of the book is devoted to this emerging field of computational intelligence in medicine and biology, attempting to provide surveys and practical examples of artificial intelligent applications in medical image analysis, fuzzy systems in biomedicine and computational bioinformatics. The included in this part chapters, while of course not comprehensive in addressing all the possible aspects of emerging artificial intelligence applications in medicine, are indicative of the explosive nature of interdisciplinary research going on in this area.

A significant methodology related to computer vision systems in medical imaging involves wavelet analysis. The first chapter reviews the use of wavelets in medical image analysis. Initially, key concepts of wavelet decomposition theory are defined, focusing on the overcomplete discrete dyadic wavelet transform, suitable for image quality preserving analysis. Basic principles underlying methods such as (i) wavelet coefficient manipulations involved in image denoising and enhancement, and (ii) wavelet feature extraction involved in image segmentation and classification tasks are highlighted. The main contribution of this chapter is the provision of application examples for various medical imaging modalities, while emphasis is given on mammographic imaging.

The timely diagnosis and treatment of pressure ulcers is a critical task and constitutes a challenge in patient rehabilitation. This important subject is discussed in chapter 13, which presents a preliminary study for automated pressure ulcer stage classification using standard image processing techniques and an SVM classifier. The deployment requirements, the internal architecture as well as the employed techniques are outlined within the chapter. Furthermore, the preliminary processing results are provided to demonstrate the feasibility of automated classification of pressure ulcer regions in various grades. The methodology can be applied to segmentation–based image classification tasks, provided that colour and texture can give meaningful information.

Chapter 14 reviews the state of the art techniques used in the development of computation vision systems for the detection of skin cancer. The chapter presents the discrete modules of such systems, beginning from the installation and image acquisition setup, continues on the visual features used for skin lesion classification and image processing methods for defining them and results in how to use the extracted features for skin lesion classification by employing artificial intelligence methods, i.e., Discriminant Analysis, Neural Networks, Support Vector Machines and Wavelet Analysis. The chapter summarizes the characteristics of all the existing systems found in literature that deal with the specific problem.

Chapter 15 presents recent advances of fuzzy systems in biomedicine. A short introduction is made on the main concepts of fuzzy sets theory. Then, a survey of recent research reports (2000 and beyond) is performed, in order to map existing theoretical trends in fuzzy systems in biomedicine, as well as important real-world biomedical applications using fuzzy sets theory. The surveyed research reports are divided into different categories either (a) according to the medical practice (diagnosis, therapy and imaging - including signal processing) or (b) according to the kind of problem faced (device control, biological control, classification and pattern analysis, and prediction-association). Recently emerging biological topics related to gene expression data, molecular - cellular analysis and bioinformatics, using fuzzy sets theory, are also reported in the chapter.

Finally, chapter 16, which concludes this part of the volume, deals with bioinformatics and provides a brief overview of the basic processing concepts involved in microarray experiments, which are the main source of such data. Microarray technology monitors the emitting fluorescence reflecting the expression levels of thousands of genes simultaneously, which are bound to the oligonucleotide probes specific for each of the putative gene sequences comprising the total genome of the investigated organism. In this context, the chapter gives a feeling for what the genomic data actually represent, and information on the various computational methods that one can employ to derive meaningful results from the corresponding experiments.

I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Intelligent Processing of Medical Images in the Wavelet Domain

Lena COSTARIDOU, Spyros SKIADOPOULOS,
Philippos SAKELLAROPOULOS and George PANAYIOTAKIS
Department of Medical Physics, School of Medicine,
University of Patras, 265 00 Patras, Greece

Abstract. Some of the major aspects of computer vision systems in medical imaging involving wavelet analysis are reviewed in this chapter. Initially, key concepts of wavelet decomposition theory are defined, focusing on the overcomplete discrete dyadic wavelet transform, suitable for image quality preserving analysis. Next, basic principles underlying methods such as (i) wavelet coefficient manipulations involved in image denoising and enhancement, and (ii) wavelet feature extraction involved in image segmentation and classification tasks are highlighted. Finally, application examples corresponding to the above mentioned methods are provided for various medical imaging modalities with emphasis on mammographic imaging.

Keywords: Wavelets, Wavelet Image Denoising and Enhancement, Wavelet Feature Analysis, Medical Computer Vision Systems

Introduction

Wavelet image analysis methods have been the subject of numerous publications in medical imaging, including excellent review articles [1], [2], with state-of-the-art methods reported in a recently published special issue of a major journal in medical imaging [3]. The most promising of these multiresolution analysis methods are the ones involving image, volume and time series medical data, utilized in critical clinical tasks such as disease screening, detection and diagnosis. Recently, such methods have started to be exploited in the field of bioinformatics, aiming at molecular profiling of disease [4].

Wavelet image analysis methods have found wide acceptance in many industrial applications fields [5], involving signal and image analysis, such as acoustics, non destructive testing, satellite imagery, to mention a few, with its most known application in still image and video compression. The contribution of wavelets to data compression is represented by their adoption in the JPEG and MPEG standards [6]-[10]. Wavelet compression methods in medical imaging applications, especially the lossy ones, are expected to provide significantly higher compression ratios with less image degradation. Preservation of diagnostic image information, as dictated by the American College of Radiology (ACR) and the National Electrical Manufacturing Association (NEMA), has lead to the investigation of "visually lossless" compression ratios. However, diagnostic information preservation is subjectively defined. Currently, clinical evaluation of JPEG 2000 image compression algorithms is an active area of research [11], [12].

Focusing on wavelet medical image processing and analysis methods, these can be categorized in methods targeted to improving the signal to noise ratio of early and occult signs of disease and methods capturing salient tissue features, such as texture, encoded in wavelet coefficients capable of tissue characterization. The former are especially important for medical image soft-copy display and refer to image denoising and enhancement methods, based on wavelet coefficient quantization, while the latter utilize wavelet feature vectors combined with classifiers for medical image segmentation and detection/characterization tasks [13].

In the following sections, key concepts of wavelet decomposition theory are defined, basic principles of wavelet image denoising and enhancement methods are highlighted and wavelet features utilized in classification tasks are reviewed. Finally, application examples corresponding to the above mentioned methods are provided for mammographic imaging, including also examples from Magnetic Resonance Imaging (MRI) and Computed Tomography (CT).

1. Some Key Concepts of Wavelet Analysis

The wavelet decomposition [14], [15] of a 1D function of space $f(x)$ according to the dyadic wavelet transform series is defined by the convolution:

$$W_s f(x) = f(x) * \psi_s(x) \tag{1}$$

where $\psi_s(x)$ are dilated versions of a wavelet function $\psi(x)$ by a factor $s=2^j$. Function $\psi(x)$ must be well localized both in space and frequency and be an oscillatory function. The scaling property (zooming) of wavelet is capable of capturing signal transients (singularities) superimposed on more slowly varying signal components. In Figure 1, three scaled versions of a Difference of Gaussians wavelet function are shown.

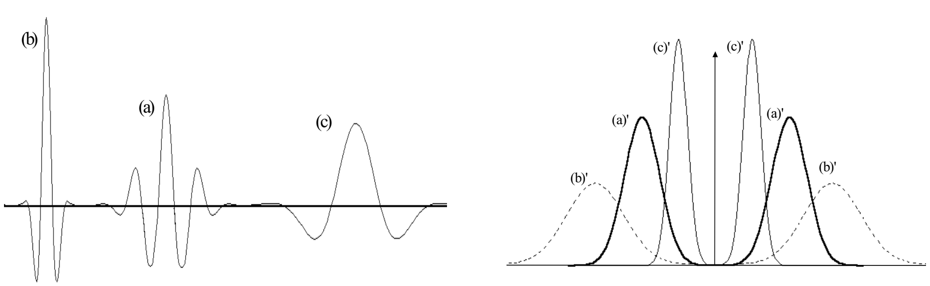


Figure 1. (a),(b),(c) Plot of a wavelet $\psi(x)$ for three different scales (a'),(b'),(c') Fourier transforms of (a), (b) and (c), respectively. The high frequency wavelet is narrow, packing the oscillations of the wavelet in a small space extent, while the low frequency wavelet, obtained by stretching, is much wider.

In general, $f(x)$ can be recovered by its wavelet transform:

$$f(x) = \sum_{j=-\infty}^{+\infty} W_s f(x) * \chi_s(x) \tag{2}$$

where $\chi(x)$ is the reconstruction wavelet [2], [3].

The approximation of $f(x)$ at scale s is defined as:

$$S_s f(x) = f(x) * \phi_s(x) \tag{3}$$

where $\varphi(x)$ is a smoothing function called *scaling* function [2], [3].

As scale s increases, more details are removed by the S_s operator. Dyadic wavelet transform series $W_s f(x)$ between scales 2^1 and 2^J contain the details existing in the $S_I f(x)$ representation that have disappeared in $S_J f(x)$.

1.1. Overcomplete Wavelet Analysis

As in computer vision systems related to diagnostic tasks, image quality preservation is a must, overcomplete wavelet analysis is preferable to the standard subsampled wavelet decomposition, avoiding aliasing artifacts and ensuring translation invariance, by improving localization in space and frequency [16]-[18]. Fast, biorthogonal, overcomplete wavelet transforms are based on families of wavelet functions $\psi(x)$ with compact support, which are derivatives of Gaussian-like spline functions $\theta(x)$ [1].

A two dimensional (2D) biorthogonal, overcomplete wavelet transform of $f(x,y)$, based on the first-order derivative of a smoothing function, has two components (H: horizontal, V: vertical):

$$\begin{Bmatrix} W_s^H f(x,y) \\ W_s^V f(x,y) \end{Bmatrix} = s \begin{Bmatrix} f(x,y) * \psi_s^H(x,y) \\ f(x,y) * \psi_s^V(x,y) \end{Bmatrix} = s \begin{Bmatrix} \dfrac{\partial}{\partial x}\left[f(x,y) * \theta_s(x,y)\right] \\ \dfrac{\partial}{\partial y}\left[f(x,y) * \theta_s(x,y)\right] \end{Bmatrix} = s \cdot \vec{\nabla}(f * \theta_s)(x,y) \tag{4}$$

resulting in a multiresolution hierarchy of approximation images $S_{2^j} f(m,n)$ and detail images $\left(W_{2^j}^H f(m,n), W_{2^j}^V f(m,n)\right)\Big)_{1 \le j \le J}$, with (m,n) discrete samples.

Eq. (4) computes the sampled horizontal and vertical components of the multiscale gradient vector, related to local contrast [19], from which magnitude-orientation representation of the gradient vector is derived by (Figure 2):

$$M_s(m,n) = \sqrt{\left|W_s^H(m,n)\right|^2 + \left|W_s^V(m,n)\right|^2} \;,\; O_s(m,n) = \arctan\left(\frac{W_s^V(m,n)}{W_s^H(m,n)}\right) \tag{5}$$

---

[1] Fourier transforms of these functions are defined as follows:
$\hat{\psi}(\omega) = (j\omega)\left(\dfrac{\sin(\omega/4)}{\omega/4}\right)^{2n+2}$ $\hat{\theta}(\omega) = \left(\dfrac{\sin(\omega/4)}{\omega/4}\right)^{2n+2}$, where n expresses the derivative order

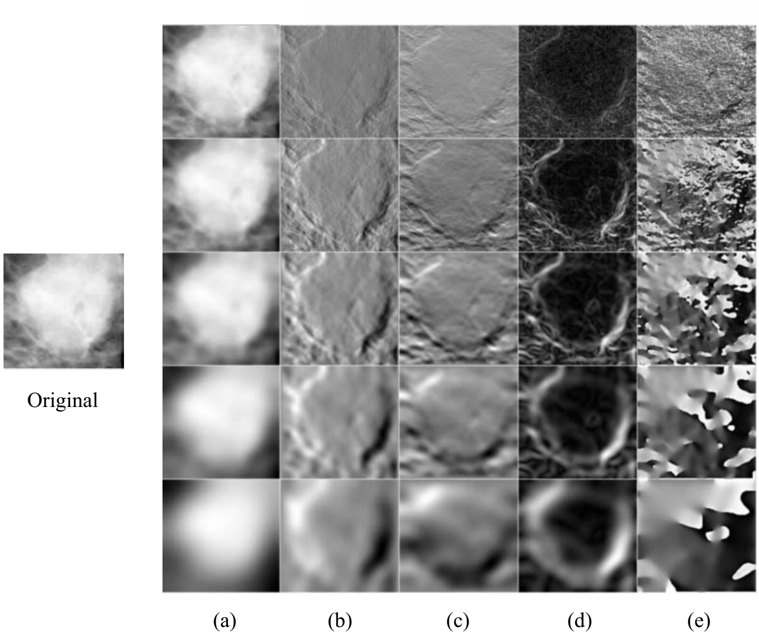


Figure 2. Five scale overcomplete wavelet decomposition of an original region of interest (ROI) containing a circumscribed mass in a mammogram. (a) Approximation, (b) horizontal, (c) vertical, (d) gradient magnitude and (e) gradient orientation subimages from dyadic scale 2^1 (top) to scale 2^5 (bottom).

1.2. Wavelet Coefficients Decay across Scales

Besides zooming on signal transients (image singularities), wavelet transform can be used to distinguish different entities in an image by the decay order of the amplitude of their corresponding wavelet coefficients over scales, assessed by local Lipschitz regularity [16], [20]. Image singularities (edges) are characterized by positive Lipschitz exponents and their amplitude values increase with increasing scale, while noise singularities are characterized by negative Lipschitz exponents whose amplitudes decrease with increasing scale. Figure 3 provides an example of differentiation of a film artifact (FA) and a microcalcification (MC) in a mammogram, using regularity.

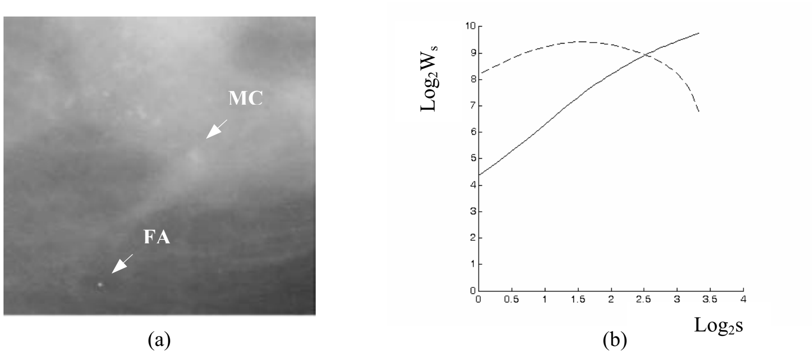
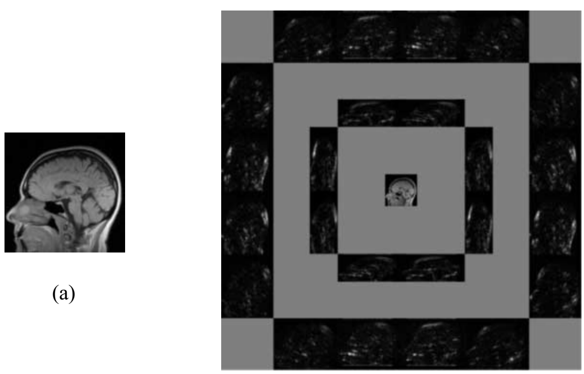


Figure 3. (a) Original ROI containing a microcalcification (MC) cluster and film artifacts (FA). (b) Amplitude of wavelet coefficients (W_s) of a MC (solid) and a FA (dashed) across dyadic scales (s).

1.3. Orientation Selectivity

As mentioned above, wavelets can "zoom in" on point singularities by focusing mainly on horizontal, vertical and diagonal orientations [1], [2], [16], thus characterized by limited orientation selectivity. To deal with singularities at various radial directions, the ridgelet and curvelet wavelet transforms have been introduced [21]-[25]. The ridgelet transform is optimal for representation of linear radial structures at various resolutions, while the curvelet transform is optimal for detecting curves. The curvelet transform has found many applications in denoising and texture characterization of medical images [23], [25]. It produces a multiscale pyramid with many directional subimages at each scale, since curvelets are produced by dilations, rotations and translations of the "mother" curvelet. Each subimage corresponds to a specific "wedge" in the frequency domain.

A set of 2D and 3D discrete implementations of the continuous wavelet transform are implemented in the "CurveLab" Toolbox (see Appendix) [24]. Figure 4b displays the curvelet pyramid of the original brain MRI image (Figure 4a). The parameters of the algorithm are the number of resolutions and the number of angles at the second coarsest level (a multiple of four and at least eight). Three levels of resolutions and eight angles were selected. The coarsest resolution (approximation) coefficients are displayed in the center, while the curvelet coefficients of the higher resolutions are located at rectangular "coronae". Each corona contains the directional subimages with angle starting at the top-left corner and increasing clockwise. The number of angles is doubled from one scale to the next finer scale.



(a) (b)

Figure 4. (a) Original brain MRI image. (b) Two directional subimages at the 2nd and 3rd resolution levels are highlighted.

2. Wavelet Denoising and Enhancement

Enhancement of medical images in the wavelet domain can be accomplished by manipulating wavelet coefficients at selected scales and then reconstructing from the modified coefficients. This type of selective enhancement relies on the fact that features of interest in medical images mainly reside at a particular scale, whereas noise and less significant structures reside at other scales of wavelet analysis. In the advent of medical imaging modalities utilizing digital detectors capable of at least 1,024 distinct levels of grey, whereas radiologists can detect at most 128 of those, there is motivation for development of data mining contrast enhancement methods for soft-copy display of medical images [26].

However, complete separation of signal from noise is difficult and any technique that simply amplifies wavelet coefficients will increase the presence of noise. For this reason, enhancement techniques always combine a denoising step, except for the case of images with very high signal-to-noise ratio.

2.1. Denoising by Shrinkage

Conventional noise filtering techniques, such as Wiener filtering [27], reduce noise by suppressing the high-frequency image components. The drawback of these methods is that cause edge blurring.

Wavelet-based denoising methods can effectively reduce noise while preserving edges. The most widely used wavelet denoising method is coefficient shrinkage [28]. This method consists of comparing wavelet coefficients against a threshold, which distinguishes signal coefficients from noise coefficients (Figure 5). This approach is justified by the decorrelation and energy compaction properties of the wavelet transform. Signal (or image) energy in the wavelet domain is mostly concentrated in a few large coefficients. Therefore, coefficients below a threshold are attributed to noise and are set to zero. Coefficients above the threshold are either kept unmodified (hard-thresholding) or modified by subtracting the threshold (soft-thresholding).

Use of a overcomplete wavelet transform as a basis for wavelet shrinkage is beneficial, as thresholding in a shift-invariant transform outperforms thresholding in an orthogonal transform by reducing pseudo-Gibbs phenomena [29]. Applying thresholding on gradient magnitude subimages (Figure 2d) is an efficient strategy for avoiding orientation distortions [30]-[32].

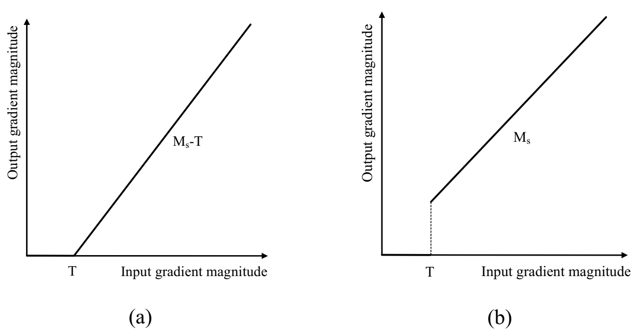


Figure 5. Soft-thresholding (a), and hard-thresholding (b) denoising functions.

2.2. Global/Local Denoising

A global threshold value *T* is usually set at a fixed percentile of coefficients values or using a threshold rule $T=\sigma\cdot[2log(MxN)]^{1/2}$, where σ is an estimation of Gaussian noise variance [29] and MxN the dimensions of gradient magnitude subimages. A global threshold cannot however accommodate for varying image characteristics. In smooth regions the coefficients are dominated by noise, thus most of these coefficients should be removed. In regions with large variations, the coefficients carry signal information, thus they should be slightly modified to preserve signal details. For these reasons, spatially adaptive thresholding strategies have been proposed [33], [34]. Specifically, soft-thresholding using a local threshold is applied on wavelet coefficient magnitudes. The threshold is calculated at each (dyadic) scale and position by applying a local window and using the formula $T=\sigma^2/\sigma_x$, where σ_x the local variance of the true image signal, which can be estimated using a local window moving across the image or, more accurately, by a context-based clustering algorithm [33].

In both methods, an estimation of the noise variance is needed. A robust estimation proposed in [29] uses the median value of the wavelet coefficients at the finer scale: $\sigma=median[(M_1 f(m,n))]/0.6745$. An alternative approach is to use a region of the image without signal to measure the noisy background [34].

Other methods of wavelet-based image denoising work by analyzing the evolution of multiscale coefficients across scales. In the work of Mallat and Hwang [35] Lipschitz exponents from scales 2^2 and 2^3 are used to eliminate edge points with negative exponents and reconstruct the maxima at the finest scale 2^1, which is mostly affected by noise. The image is then reconstructed from the denoised set of multiscale edges.

2.3. Global Enhancement

Similarly to denoising, contrast enhancement in the wavelet domain is also accomplished by global or spatially adaptive wavelet enhancement functions. Linear enhancement functions ensure that all regions of the image are enhanced in the same way and are equivalent to multiscale unsharp masking [31]. A drawback of linear enhancement is that leads to inefficient usage of the dynamic range available since it emphasizes high-contrast and low-contrast edges with the same gain. In order to accentuate the visibility of low contrast regions, a nonlinear transformation function, proposed by Laine *et al.* [31] has been widely used (Figure 6a):

$$M_s^e(m,n) = \begin{cases} k_s M_s(m,n), & \text{if } M_s(m,n) \le T_s \\ M_s(m,n) + (k_s - 1)T_s\,, & \text{if } M_s(m,n) \ge T_s \end{cases} \tag{6}$$

Another global wavelet mapping function proposed by Lu *et al.* [36] is the hyperbolic function (Figure 6b):

$$M_s^e(m,n) = \frac{\tanh(aM_s(m,n) - b) + \tanh(b)}{\tanh(a-b) + \tanh(b)} \tag{7}$$

where the parameters a and b satisfy $a > 0$ and $0 \le b \le 0.5a$.

The hyperbolic mapping function has some desired properties. It is saturated at large gradient magnitude values and thus avoids over-enhancement of high-contrast features. It also suppresses small gradient magnitudes in order to control noise amplification. The maximum slope (gain) occurs at *b/a*.

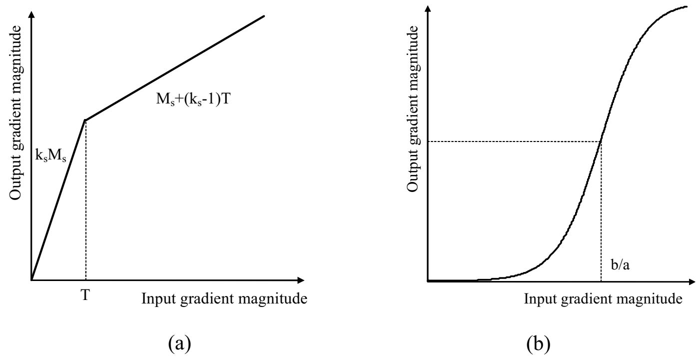


Figure 6. Global piecewise linear (a), and hyperbolic (b) contrast enhancement transformation functions.

Parameters of the global enhancement functions can be the same or vary across wavelet scales to selectively enhance features of different sizes. In Figure 7, the effect of wavelet enhancement as pre-processing step to 3D lung segmentation, using multislice CT data is demonstrated. Enhancement parameters are automatically selected using a minimum area overlap criterion based on Gaussian modeling of the lung volume histogram [37]. In the example, low contrast small size edges, such as the thin anterior junction line shown, is highlighted.

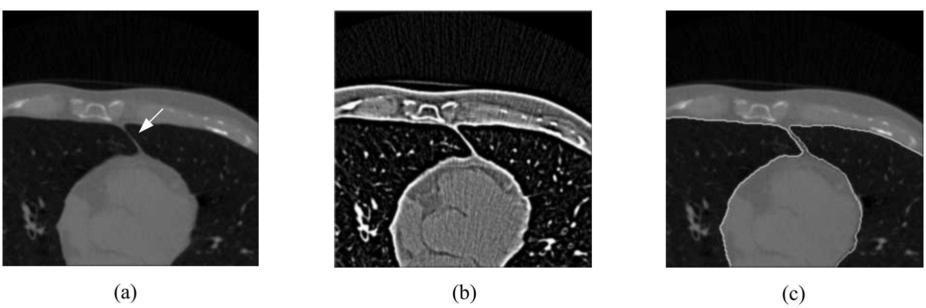


Figure 7. (a) Part of an original CT slice image depicting a thin anterior junction line (arrow). (b) Wavelet edge highlighted image and (c) the corresponding lung border delineation.

2.4. Locally Adapted Enhancement

To adapt enhancement functions to varying local contrast characteristics, a locally adaptive enhancement strategy has been recently proposed [34]. It is based on local range modification of gradient magnitude values in a local window, using a spatially adapted gain factor $G_{L,s}(m,n)$ as:

$$G_{L,s}(m,n) = \begin{cases} G_s(m,n) \equiv M_{1,\max} / M_{s,\max}(m,n), & if \quad G_s(m,n) < L \\ L, & otherwise \end{cases} \tag{8}$$

where $M_{1,max}$ the maximum value of the magnitude subband image at scale 2^1, $M_{s,max}(m,n)$ the local maximum value in a local window at the magnitude subband image at scale s and position (m,n) and L a local gain limit parameter used for controlling over-enhancement.

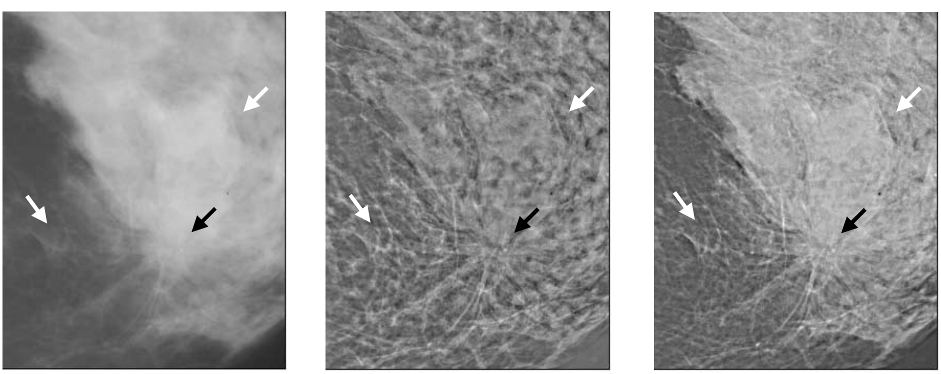
2.5. Regionally Adapted Enhancement

A different approach to enhancement adaptation has also been recently proposed taking into account anatomically defined regions, using Gaussian mixture modeling [38]. Coefficient mapping of each breast component region C and scale s is controlled by a gain factor $G_{C,s}(m,n)$ defined by:

$$G_{C,s}(m,n) = \frac{GL_{max} - GL_{min}}{FWHM_C} \tag{9}$$

where GL_{max}-GL_{min} is the original image grey level range and $FWHM$ is the full width at half maximum of the Gaussian distribution of the C^{th} breast component.

In Figure 8, processing examples with locally and regionally adapted wavelet enhancement methods are provided. Both methods enhance the center and the spicules of the mass (black arrows). Glandular and connective tissue of dense and fatty regions (white arrows) are better visualized with the use of the regionally adapted method, preserving the original anatomical structure.



(a) (b) (c)

Figure 8. Region of an original mammogram (a). Results of processing with locally adapted (b) and regionally adapted (c) wavelet enhancement.

3. Wavelet Computer Vision Systems

3.1. Wavelet Feature Extraction

Various wavelet-based image features in combination with classification methods have been proposed for design and implementation of intelligent systems targeted to automated tissue segmentation, detection/characterization tasks, the latter referred as CAD or CAD_x systems [39], [40]. The motivation of using multiresolution image features is due to the spatial-frequency localization property of wavelet functions, providing an ideal representation for texture analysis, by means of quantitative metrics relevant to the status of biological tissues [39], [41], [42]. The following major categories of features derived from the wavelet coefficients of each subband image have been reported in medical image analysis and content-based image retrieval literature: (a) histogram features [17], [41], [43]-[60], (b) co-occurrence matrices features [47], [48], [55], [57] and (c) fractal dimension features [55], [57], [61].

3.1.1. Histogram Features

Wavelet coefficient histogram analysis of subband images *k* at various scales *s,* such as $W_s^H f(m,n), W_s^V f(m,n)$, can provide a number of first order statistics features, such as mean, standard deviation, extrema density, energy, energy standard deviation and Shannon entropy. In the case of 2D biorthogonal, overcomplete wavelet transform as defined in Eq. 4, features based on the magnitude-orientation multiresolution representation $M_s f(m,n), O_s f(m,n)$ can be extracted, such as the folded gradient orientation and its standard deviation, coherence and entropy of orientations. Representative features utilized in wavelet-based medical image analysis systems are summarized in Table 1.

Another approach of extracting features is based on wavelet coefficient histogram distribution modeling. Two statistical models have been defined; the Generalized Gaussian Density (GGD) model [46] and the Hidden Markov Tree (HMT) model [58]. The first model, proposed by Mallat [14], assumes that all subband wavelet coefficient histograms appear to be similar in shape, symmetrical and with a sharp peak. In the formula of GGD model (Table 1), $\alpha>0$ is a parameter describing the standard deviation (spread) of the distribution, $\beta>0$ is a (shape) parameter, which is inversely proportional to the decreasing rate of the peak and $\Gamma(.)$ is the Gamma function. The couples ($\alpha(j,k)$, $\beta(j,k)$) corresponding to scale 2^j and subband *k* are estimated by the maximum-likelihood algorithm [46] and used as features of an image. The aforementioned model makes the assumption of independence of the wavelet coefficients distribution across different subbands and scales.

To overcome this limitation a second model was suggested, treating the wavelet coefficients as being in one of two (hidden) states: “high” corresponding to wavelet coefficients containing significant contributions of signal energy, or “low” representing coefficients with small signal energy. Associating a high-variance, zero-mean Gaussian probability density for the “high” state and a low-variance, zero-mean Gaussian density for the “low” state, the result is a two-state Gaussian mixture model. In the formula of HMT model (Table 1), one Gaussian distribution has a small standard deviation (σ_L) capturing the peak of the real distribution, the other one has larger standard deviation (σ_H) capturing the heavy tails, whereas λ^L expresses “low” probability state.

Table 1. Wavelet features used for medical image texture analysis either for tissue segmentation or characterization.

Feature	**Formula**
Mean of Amplitude [17], [43], [47]-[52], [54]-[57]	$\frac{1}{M \times N}\sum_{m=1}^{M}\sum_{n=1}^{N}\left\lvert W_s^k(m,n)\right\rvert$
Normalized Energy [27], [44], [45], [47], [53]-[55], [57]	$\frac{1}{M \times N}\sum_{m=1}^{M}\sum_{n=1}^{N}\left\lvert W_s^k(m,n)\right\rvert^2$
Normalized Shannon Entropy [16], [50], [51], [53], [56]	$-\sum_{m=1}^{M}\sum_{n=1}^{N}\left[\frac{\left\lvert W_s^k(m,n)\right\rvert^2}{\sum_{m=1}^{M}\sum_{n=1}^{N}\left\lvert W_s^k(m,n)\right\rvert^2}\right]\log_2\left[\frac{\left\lvert W_s^k(m,n)\right\rvert^2}{\sum_{m=1}^{M}\sum_{n=1}^{N}\left\lvert W_s^k(m,n)\right\rvert^2}\right]$
Folded Gradient - Orientation [49], [50]	$\begin{cases} O_s(m,n)+\pi, & if\ \frac{1}{M \times N}\sum_{m=1}^{M}\sum_{n=1}^{N}\left\lvert O_s^+(m,n)\right\rvert - \left\lvert O_s(m,n)\right\rvert > \frac{\pi}{2} \\ O_s(m,n)-\pi, & if\ \left\lvert O_s(m,n)\right\rvert - \frac{1}{M \times N}\sum_{m=1}^{M}\sum_{n=1}^{N}\left\lvert O_s^-(m,n)\right\rvert > \frac{\pi}{2} \\ O_s(m,n), & otherwise \end{cases}$
Coherence [50], [51]	$M_s(m,n)\frac{\sum_{(x,y)\in W}\left\lvert M_s(x,y)cos[O_s(m,n)-O_s(x,y)]\right\rvert}{\sum_{(x,y)\in W} M_s(x,y)}$
GGD model: Spread and Shape [45]-[47]	$\frac{\beta}{2\alpha\Gamma(1/\beta)}e^{-\left(\frac{\left\lvert W_s^k(m,n)\right\rvert}{\alpha}\right)^{\beta}}$
HMT model: Low probability state and "low" – "high" standard deviations [58]-[60], [62]	$\lambda^L\left(2\pi\sigma_L^2\right)^{-\frac{1}{2}}e^{-\left(\frac{\left\lvert W_s^k(m,n)\right\rvert^2}{2\sigma_L^2}\right)}+(1-\lambda^L)\left(2\pi\sigma_H^2\right)^{-\frac{1}{2}}e^{-\left(\frac{\left\lvert W_s^k(m,n)\right\rvert^2}{2\sigma_H^2}\right)}$

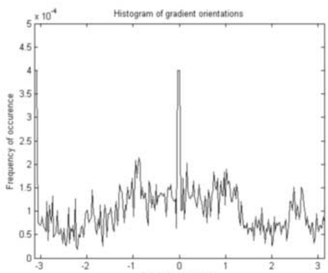
O^+ *(m,n)* and O^- *(m,n)* are the mean values of positive and negative gradient orientations, respectively, and *(x,y)* pixels belonging in a local window W centered in the position *(m,n)*.

For capturing propagation of wavelet coefficient amplitudes across scales, called persistence, the hidden states are connected across scales constructing a (hidden Markov) tree [58]-[60], [62]. This is expressed by the state transition probability matrix for the wavelet coefficients of the coarse level (parent) to the coefficients at the next intermediate level that correspond to the same location (children), defined by $\lambda_{2^j}^{\xi} = \sum_{\xi'} \lambda_{2^{j-1}}^{\xi'} \lambda_{2^j}^{\xi' \to \xi}$ for j=2,3, …, J and $\xi, \xi' \in$ {"low", "high"}, where $\lambda_{2^j}^{\xi' \to \xi}$ is the probability that a child wavelet coefficient at scale 2^j is in the state ξ given its parent wavelet coefficient is in the state ξ'. The HMT model is completely defined by "low" and "high" probability states of the first scale, the state transition probability matrices from 2nd to J^{th} scale and "low"-"high" standard deviations of all scales.

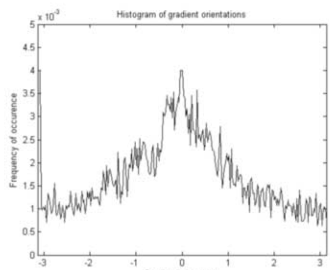
In Figure 9, an example of gradient orientation and coherence features, extracted from the biorthogonal overcomplete wavelet transform subimages, is provided for a mass containing and normal dense tissue [50] ROIs.

(a)

(b)

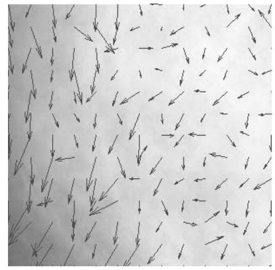


(c)



(d)

(e)



(f)

Figure 9. Regions of a spiculated mass (a) and a normal dense tissue (b). Their gradient orientation histograms (c and d) and coherence images (e and f) of dyadic scale 2^2, respectively. The degree of flow anisotropy is depicted by means of arrow length and orientation.

The directions of spicules of the oriented histogram of the mass (Figure 9c) differs from the directions of the corresponding histogram of normal tissue (Figure 9d). Specifically, pixels of normal dense mammographic tissue have gradient orientations pointing to a certain direction (e.g. oriented towards the nipple), while pixels of spicules of the mass tend to have gradient orientations distributed in more directions. Thus, the spread of the orientation histogram is much wider in case of the spiculated mass with respect to the normal dense tissue one.

Coherence is a local measure of the degree of anisotropy of flow [51], depicted by means of arrow length and orientation. Specifically, pixels near mass edges and connective tissue line structures tent to have high gradient magnitudes and similar gradient orientations in their neighborhood region (Figure 9e), while pixels of homogeneous regions are characterized by decreased coherence due to small gradient magnitude scattered in a wide range of orientations (Figure 9f).

3.1.2. Co-Occurrence Matrices Features

Co-occurrence matrix is a well-established robust statistical tool for extracting second order texture information from images [63]. It characterizes the spatial distribution of wavelet coefficients in the selected subband image. An element at location (i,j) of the co-occurrence matrix signifies the joint probability density of the occurrence of wavelet coefficients with i and j in a specified orientation θ and specified distance d from each other. Working on different resolution scales of subband images corresponds to different d values, zooming in and out on the underlying texture structure analyzing dependencies existing between adjacent or more distant wavelet coefficients.

A number of features can be derived from each co-occurrence matrix such as Angular Second Moment, Contrast, Correlation, Variance, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Entropy, Information Measure of Correlation 1, Information Measure of Correlation 2, Shade and Prominence. These statistical measures make up the co-occurrence signature of a subband image [47], [48], [55], [57].

3.1.3. Fractal Dimension Features

Fractal analysis provides a means of accessing geometrical complexity of 2D shapes, by considering image pixel intensities as heights above a plane, defining an intensity surface. Such intensity patterns are considered as being composed of repeated sequences of a basic pattern at multiple scales, assessed by fractal dimension. The property of repeated occurrence of a basic pattern is related to image texture. Multiresolution fractal feature vectors, derived from wavelet image decomposition at various scales, are targeted to encode significant texture information captured by wavelet coefficients of various subimages [55], [57], [61].

3.1.4. Rotation Invariant Features

Wavelet-based texture analysis should be invariant to rotation. However, most of the proposed techniques assume that texture has the same orientation [64]. Recently, several attempts have been made to use the wavelet transform for rotation-invariant texture analysis. Two approaches have been described in the literature. The first approach is based on preprocessing steps to make the feature analysis invariant to rotation [64], while the second approach is based on rotated wavelets exploiting steerability to calculate the wavelet transform for different orientations achieving isotropic rotationally invariant features for each subband image [61], [62], [65].

3.2. Classification

Computer vision applications in medical imaging aim to assist radiologists in increasing their confidence level in tissue detection and/or characterization tasks, especially in case of subtle disease signs.

A variety of wavelet multiresolution texture features, extracted from wavelet frequency subbands, have been reported in medical applications targeted to tissue segmentation, detection and characterization, aiming to capture salient tissue properties by exploiting the wavelet image representation scheme [39]. Successful paradigms of such systems refer to mammographic image-based medical diagnosis, i.e. lesion detection (CAD) [48]-[50] and characterization (CAD_x) [52], [53] systems. Recently, wavelet-based computer vision systems have been proposed for ultrasound-based characterization of focal liver lesions [54]-[56] and discrimination of viable from nonviable myocardium [57].

A generic scheme for a wavelet-based medical image texture segmentation/characterization system is provided in Figure 10. To our knowledge, classifiers used in systems reported in the literature, include regression algorithms, k-nearest neighbor and artificial neural networks. Usually, a feature dimensionality reduction step is required to determine the optimal subset of features, followed by classifier training and testing using different tissue samples [66]. The area under Receiver Operating Characteristic (ROC) curve (A_z) of the classifier is a widely acceptable performance evaluation metric [67].

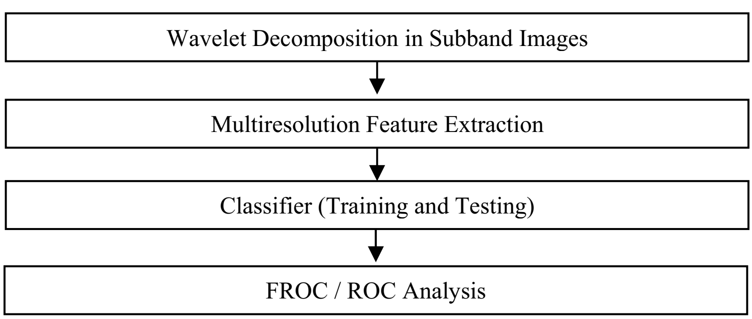


Figure 10. Stages of a medical image wavelet-based feature classification system used in tissue segmentation and characterization tasks.

An example of exploiting wavelet texture analysis in tissue surrounding microcalcification clusters in mammography [68], is provided in Figure 11. The feasibility of grey level and wavelet texture, based on first order and co-occurrence matrices statistics, in discriminating malignant from benign tissue is investigated using a probabilistic neural network (PNN). Classification outputs of the most discriminating feature sets are combined using a majority voting rule and compared to radiologist' assessment of malignancy, in a case sample of 100 subtle MC clusters (46 benign and

54 malignant) originating from the Digital Database for Screening Mammography (DDSM) database.

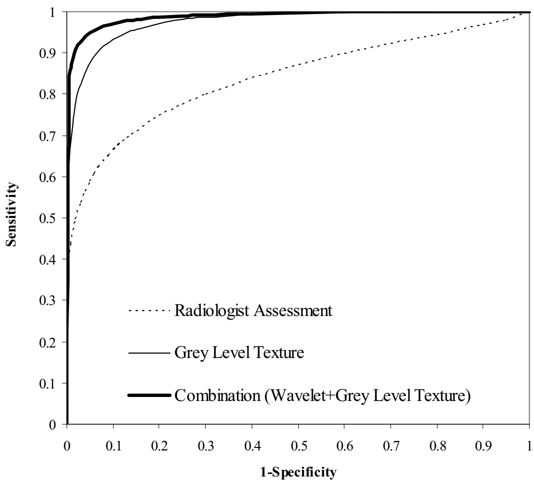


Figure 11. ROC curves corresponding to radiologist' assessment of malignancy, grey level texture and combined classification scheme. The higher the area under ROC curves the higher the discrimination ability.

4. Discussion-Conclusion

Multiscale image representation in the wavelet domain provides a means of separation of image signals from noise, while the amplitude of wavelet coefficients at subbands measures the correlation of image salient features to the wavelet basis functions.

Focusing on medical image analysis, overcomplete wavelets constructed from derivatives of spline functions, have been particularly successful in image denoising-enhancement, by exploiting the magnitude of image gradient vector for isotropic processing.

Wavelet-based denoising-enhancement methods are expected to contribute in efficient dynamic range utilization in soft-copy display of digital imaging systems or as pre-processing steps of intelligent wavelet-based feature classification system.

Finally, regarding intelligent wavelet-based feature classification systems, wavelet domain feature extraction in combination to recently proposed directional wavelets is expected to provide efficient means of capturing textural image patterns independent of neighborhood size and orientation [25]. To optimize analysis for a specific classification task the best wavelet basis has to be selected [39].

Acknowledgement

Part of this work has been supported by the European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II), and particularly the Program PYTHAGORAS I (B.365.011).

Appendix

Following, a list of implementations of wavelet transformations and corresponding scientific tools is provided.

- *"WaveLab", Stanford University:* A library of Matlab routines for wavelet analysis, wavelet-packet analysis, cosine-packet analysis and matching pursuit (http://www-stat.stanford.edu/~wavelab/).
- *"The MATLAB Wavelet Toolbox", The MathWorks:* Continuous wavelet transforms, discrete wavelet transforms, wavelet packet transforms, denoising by thresholding (http://www.mathworks.com/products/wavelet/index.shtml).
- *"The Rice Wavelet Toolbox", Rice University:* Matlab toolbox with orthogonal and biorthogonal transforms and applications to denoising (http://www-dsp.rice.edu/software/rwt.shtml).
- *"WavBox" Software, WavBox ToolSmiths:* Matlab toolbox (GUI and command line) with wavelet transforms and adaptive wavelet packet decompositions (http://www.toolsmiths.com/p/wavbox.aspx).
- *"PiefLab", Maarten Jansen:* Library of Matlab algorithms for wavelet noise reduction (http://www.cs.kuleuven.ac.be/~maarten/software/pieflab.html).
- *Wavelet denoising Matlab software, Antoniadis, A. et al.:* Wavelet shrinkage and wavelet thresholding estimation (http://www-lmc.imag.fr/SMS/software/GaussianWaveDen/).
- *"MegaWave2", CMLA of Ecole Normale Superieure de Cachan:* Contains 1) a C library of modules, that contains original algorithms written by researchers; and 2) a Unix/Linux package designed for the fast development of new image processing algorithms (http://www.cmla.ens-cachan.fr/~megawave).
- *"Lastwave", Centre de Mathematiques Appliquees:* Software in C that deals with high-level structures such as signals, images, wavelet transforms, extrema representation, short time Fourier transform, etc (http://www.cmap.polytechnique.fr/~bacry/LastWave/index.html).
- *"WAILI" C++ Wavelet Transform Library, Uytterhoeven, Wulpen, Jansen:* It includes some basic image processing operations based on the use of wavelets. It provides integer wavelet transforms based on the Lifting Scheme, noise reduction based on wavelet thresholding and edge enhancement (http://www.cs.kuleuven.ac.be/~wavelets/).
- *"CurveLab Toolbox", California Institute of Technology:* A collection of Matlab and C++ programs for the Fast Discrete Curvelet Transform in two and three dimensions (http://www.curvelet.org/software.html).

References

[1] M. Unser and A. Aldroubi, A review of wavelets in biomedical applications, *Proc of IEEE* **84** (1996), 626-638.
[2] A. Laine, Wavelets in temporal and spatial processing of biomedical images, *Annu Rev Biomed Eng* **2** (2000), 511-550.
[3] M. Unser, A. Aldroubi and A. Laine, Wavelets in Medical Imaging (Special Issue on Wavelets in Medical Imaging), *IEEE Trans Med Imaging* **22** (2003), 285-288.
[4] X.H. Wang, R.S.H. Istepanian and Y.H. Song, Microarray image enhancement by denoising using stationary wavelet transform, *IEEE Trans Nanobioscience* **2** (2003), 184-189.
[5] F. Truchetet and O. Laligant, Wavelets in industrial applications: A review in wavelet applications in industrial processing II. In: F. Truchetet, O. Laligant (eds), *Proc of SPIE* **5607** (2204), 1-14.
[6] M. Antonini, M. Barlaud, P. Mathieu and I. Daubechies, Image coding using the wavelet transform, *IEEE Trans Image Process*, **2** (1992), 205-220.
[7] J.M. Shapiro, Embedded image coding using zerotrees of wavelet coefficients, *IEEE Trans Signal Process,* **41** (1993), 3445-3462.
[8] A. Said and W.A. Pearlman**,** An image multiresolution representation for lossless and lossy image compression, *IEEE Trans Image Process* **5** (1996), 1303-1310.
[9] M.W. Marcellin, M. Gormish, A. Bilgin and M.P. Boliek, An overview of JPEG2000, *Proc of Data Compression Conference*, Snowbird, Utah, (2000), 523-544.
[10] M. Unser, Mathematical properties of the JPEG2000 wavelet filters, *IEEE Trans Image Process* **12** (2003), 1080-1090.
[11] M-M. Sung, H-J. Kim, E-K. Kim, J-Y. Kwak, J-K. Yoo, H-S. Yoo, Clinical evaluation of JPEG2000 compression algorithm for digital mammography, *IEEE Trans Nucl Sci* **49** (2002), 827-832.

[12] O. Kocsis, L. Costaridou, L. Varaki, E. Likaki, C. Kalogeropoulou, S. Skiadopoulos and G. Panayiotakis, Visually lossless threshold determination for microcalcifications detection in wavelet compressed mammograms, *Eur Radiol* **13** 2003, 2390-2396.

[13] Medical Image Analysis Methods, *In L. Costaridou (Ed), CRC Press, Taylor & Francis Group, Boca Raton, FL, US,* (2005).

[14] S. Mallat, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans Pattern Ana. Machine Intell* **11** (1989), 674–693.

[15] I. Daubechies, Orthogonal bases of compactly supported wavelets, *Commun Pure Appl Math* **41** (1988), 909–996.

[16] S. Mallat and S. Zhong, Characterisation of signals from multiscale edges, *IEEE Trans Pat Anal Machine Intell* **14** (1992), 710-732

[17] M. Unser, Texture classification and segmentation using wavelet frames, *IEEE Trans Image Process* **4** (1995), 1549–1560.

[18] I. Daubechies, The wavelet transform time-frequency localization and signal analysis, *IEEE Trans Inform Theory* **36** (1990), 961–1005.

[19] L. Costaridou, P. Sakellaropoulos, A. Stefanoyiannis, E. Ungureanu and G. Panayiotakis, Quantifying image quality at breast periphery vs. mammary gland in mammography using wavelet analysis, *Br J Radiol* **74** (2001), 913-919.

[20] P. Heinlein, J. Drexl and W. Schneider, Integrated wavelets for enhancement of microcalcifications in digital mammography, *IEEE Trans Med Imaging* **22** (2003), 402-413.

[21] E.J Candès and D.L. Donoho, Ridgelets: A key to higher-dimensional intermittency?, *Phil Trans R Soc Lond A*, **357** (1999), 2495-2509.

[22] E.J. Candès and D.L. Donoho, Curvelets - A surprisingly effective non adaptive representation for objects with edges, *In Curves and Surfaces, A. Cohen, C. Rabut and L. Schumaker (Eds), Vanderbilt University Press, Nashville, TN,* 1999.

[23] J-L Starck, E.J. Candès and D.L. Donoho, The curvelet transform for image denoising, *IEEE Trans Image Process* **11** (2002), 670-684.

[24] E. Candès, L. Demanet, D. Donoho and L. Ying**,** Fast discrete curvelet transforms, *Multiscale Modeling and Simulation,* **5** (2006), 861-899.

[25] L. Dettori and L. Semler, A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography. *Comp Biol Med* **37** (2007), 486 – 498.

[26] E.B. Cole, E.D. Pisano, D. Zeng, K. Muller, S.R. Aylward, S. Park, C. Kuzmiak, M. Koomen, D. Pavic, R. Walsh, J. Baker, E. Gimenez and R. Freimanis, The effects of gray scale image processing on digital mammography interpretation performance, *Acad Radiol* **12** (2005), 585-595.

[27] N. Wiener, The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications, *New York: John Wiley and Sons* (1949).

[28] D.L. Donoho, De-noising by soft-thresholding, *IEEE Trans Inform Theory* **41** (1995), 613–627.

[29] R.R. Coifman and D.L. Donoho, Translation-Invariant De-noising Wavelets and Statistics, *In A. Antoniadis and G. Oppenheim (Eds), Berlin, Germany: Springer-Verlag,* 1995.

[30] A. Laine, S. Schuler, J. Fan and W. Huda, Mammographic feature enhancement by multiscale analysis, *IEEE Trans Med Imaging* **13** (1994), 725-740.

[31] A. Laine, J. Fan and W. Yang, Wavelets for contrast enhancement of digital mammography, *IEEE Eng Med Biol* **14** (1995), 536-550.

[32] L. Costaridou, P. Sakellaropoulos, S. Skiadopoulos and G. Panayiotakis, Locally Adaptive Wavelet Contrast Enhancement, *In Medical Image Analysis Methods*, *Costaridou L. (Ed), CRC Press, Taylor & Francis Group, Boca Raton, FL, US,* 2005, 225-270.

[33] S. Chang, B. Yu and M. Vetterli, Spatially adaptive wavelet thresholding with context modeling for image denoising, *IEEE Trans Image Process* **9** (2000), 1522-1531.

[34] P. Sakellaropoulos, L. Costaridou and G. Panayiotakis, A wavelet based spatially adaptive method for mammographic contrast enhancement, *Phys Med Biol* **48** (2003), 787-803.

[35] S. Mallat and W. Hwang, Singularity detection and processing with wavelets, *IEEE Trans Inform Theory* **38** (1992), 617-643.

[36] J. Lu and D.M. Heally, Contrast enhancement of medical images using multiscale edge representation *Opt Eng* **33** (1994), 2151-2161.

[37] P. Korfiatis, S. Skiadopoulos, P. Sakellaropoulos, C. Kalogeropoulou and L. Costaridou, Combining 2D wavelet edge highlighting and 3D thresholding for lung segmentation in thin slice CT, *Br J Radiol* (2007), (in Press).

[38] S. Skiadopoulos, A. Karahaliou, P. Sakellaropoulos, G. Panayiotakis and L. Costaridou, Breast component adaptive wavelet enhancement for soft-copy display of mammograms, *Proc of 8th International Workshop in Digital Mammography (IWDM),* **4046** (2006), 549-556.

[39] Y. Jin, E. Angelini and A. Laine, Wavelets in Medical Image Processing: De-noising, Segmentation, and Registration, *In: J.S. Suri, D.L. Wilson, S. Laxminarayan (Eds), Handbook of Biomedical Image Analysis, Vol 1, Segmentation Models Part A, Springer US,* Chapter 6 (2005).

[40] P.M. Sampat, M.K. Markey and A.C. Bovik, Computer-Aided Detection and Diagnosis in Mammography. *In: A.C. Bovik (ed), Handbook of Image and Video Processing 2nd ed. Academic Press* (2005), 1195-1217.

[41] T. Chang and C. Kuo, Texture analysis and classification with tree-structured wavelet transform, *IEEE Trans Image Process* **2** (1993), 429–441.

[42] L. Bocchi, G. Coppini, R. De Dominicis and G. Valli, Tissue characterization from X-ray images, *Med Eng Phys* **19** (1997), 336-342.

[43] J. Wang, Multiwavelet packet transform with application to texture segmentation, *Electron Lett* **38** (2002) 1021-1023.

[44] A. Laine and J. Fan, Texture classification by wavelet packet signatures, *IEEE Trans Pattern Anal Machine Intell* **15** (1993), 1186–1191.

[45] G.V. Wouwer, P. Scheunders and D.V. Dyck, Statistical texture characterization from discrete wavelet representations, *IEEE Trans Image Process* **8** (1999), 592–598.

[46] M.N. Do and M. Vetterli, Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance, *IEEE Trans Image Process* **11** (2002), 146–158.

[47] M.H. Pi, C.S. Tong, S.K. Choy and H. Zhang, A fast and effective model for wavelet subband histograms and its application in texture image retrieval, *IEEE Trans Image Process* **15** (2006), 3078-3088.

[48] D. Wei, H-P. Chan, M.A. Helvie, B. Sahiner, N. Petrick, D.D. Adler and M.M. Goodsitt, Classification of mass and normal breast tissue on digital mammograms: Multiresolution texture analysis, *Med Phys* **22** (1995), 1501-1513.

[49] S. Liu, C.F. Babbs and E.J. Delp, Multiresolution detection of spiculated lesions in digital mammograms, *IEEE Trans Image Process* **10** (2001), 874-884.

[50] P. Sakellaropoulos, S. Skiadopoulos, A. Karahaliou, L. Costaridou and G. Panayiotakis, Using wavelet-based features to identify masses in dense breast parenchyma, *Proc of 8th International Workshop in Digital Mammography (IWDM)*, **4046** (2006), 557-564.

[51] C-M.Chang, A. Laine, Coherence of multiscale features for enhancement of digital mammograms, *IEEE Trans Med Imaging* **3** (1999), 32-46.

[52] C.M. Kocur, S.K. Rogers, L.R. Myers, T. Burns, M. Kabrisky, J.W. Hoffmeister, K.W. Bauer and J.M. Steppe, Using neural networks to select wavelet features for breast cancer diagnosis, *IEEE Eng Med Biol* **15** (1996), 95-102.

[53] D. Kramer and F. Aghdasi, Texture analysis techniques for the classification of microcalcifications in digitised mammograms. *IEEE 5th Africon Conf,* **1** (1999), 395-400.

[54] Mojsilovic′, M. Popovic′, S. Markovic′ and M. Krstic′, Characterization of visually similar diffuse diseases from B-scan liver images using non separable wavelet transform, *IEEE Trans Med Imaging* **17** (1998), 541–549.

[55] W-L. Lee, Y-C. Chen and K-S. Hsieh. Ultrasonic liver tissues classification by fractal feature vector based on M-band wavelet transform, *IEEE Trans Med Imaging* **22** (2003), 382-392.

[56] H. Yoshida, D.D. Casalino, B. Keserci, A. Coskun, O. Ozturk and A. Savranlar, Wavelet-packet-based texture analysis for differentiation between benign and malignant liver tumours in ultrasound images *Phys Med Biol* **48** (2003), 3735–3753.

[57] E.D. Kerut, M. Given and T.D. Giles, Review of methods for texture analysis of myocardium from echocardiographic images: A means of tissue characterization, *Echocardiography* **20** (2003), 727-736.

[58] M. Crouse, R.D. Nowak, and R.G. Baraniuk, Wavelet-based signal processing using hidden Markov models (Special Issue on Wavelets and Filterbanks), *IEEE Trans Signal Process* **46** (1998), 886-902.

[59] J.K. Romberg, H. Choi and R.G. Baraniuk, Bayesian tree-structured image modeling using wavelet-domain hidden Markov models, *IEEE Trans Image Process* **10** (2001), 1056-1068.

[60] G. Fan and X-G Xia, Wavelet-based texture analysis and synthesis using hidden Markov models, *IEEE Trans Circuits Systems—I: Fundamental Theory Applications* **50** (2003), 106-120.

[61] D. Charalampidis and T. Kasparis, Wavelet-based rotational invariant roughness features for texture classification and segmentation, *IEEE Trans Image Process* **11** (2002), 825-837.

[62] M.N. Do and M. Vetterli, Rotation-invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models, *IEEE Trans Multimedia* **4** (2002), 517-527.

[63] R.M. Haralick, K. Shanmugam and I. Dinstein, Textural features for image classification, *IEEE Trans Syst Man Cybern* **SMC-3** (1973), 610–621.

[64] K. Jafari-Khouzani and H. Soltanian-Zadeh, Rotation-invariant multiresolution texture analysis using Radon and wavelet transforms, *IEEE Trans Image Process* **14** (2005), 783-795.

[65] M. Kokare, P.K. Biswas and B.N. Chatterji, Rotation-invariant texture image retrieval using rotated complex wavelet filters*, IEEE Trans Syst, Man, Cybern - Part B: Cybern* **36** (2006), 1273-1282.

[66] V.F. Ruiz and S.J. Nasuto, Biomedical-Image Classification Methods and Techniques, *In Medical Image Analysis Methods*, *L. Costaridou (Ed), CRC Press, Taylor & Francis Group, Boca Raton, FL, US,* 2005, 137-183.

[67] A.R. Erkel, P.M. Pattynama, Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology, *Eur J Radiol* **27** (1998), 88-94.

[68] A. Karahaliou, S. Skiadopoulos, I. Boniatis, P. Sakellaropoulos, E. Likaki, G. Panayiotakis, L. Costaridou, Texture analysis of tissue surrounding microcalcifications on mammograms for breast cancer diagnosis, *Br J Radiol* (2007), (in Press).

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Automated Pressure Ulcer Lesion Diagnosis: An Initial Study

Dimitrios I. KOSMOPOULOS [a,*] and Fotini L. TZEVELEKOU [b]
[a] *National Centre for Scientific Research "Demokritos", Institute of Informatics and Telecommunications, Greece*
[b] *Intensive Care Unit, 251 General Air Force Hospital, Greece*

Abstract. The timely diagnosis and treatment of pressure ulcers is a critical task and constitutes a challenge in patient rehabilitation. In this chapter we present a preliminary study for automated pressure ulcer stage classification using standard image processing techniques and an SVM classifier. The deployment requirements, the internal architecture as well as the employed techniques are outlined. Furthermore, the preliminary processing results are provided to demonstrate the feasibility of automated classification of pressure ulcer regions in various grades. The methodology can be applied to segmentation–based image classification tasks, provided that colour and texture can give meaningful information.

Keywords. Pressure ulcer classification, segmentation, Gabor filter, Support Vector Machines

1. Introduction

A pressure ulcer is a lesion caused by unrelieved pressure resulting in damage of underlying skin tissue when the body stays in one position for too long without shifting the weight. Pressure ulcers occur in acute and chronic health care settings. Even short time bedridden patients (for example, after surgery or an injury) can be found at high risk of developing pressure ulcers. Long term care facilities nursing the elderly and wheelchair dependent patients face high rates of pressure ulcer incidence. In the United States, pressure ulcer prevalence in acute care hospitals is as high as 17 percent, in long term care facilities is 28 percent and at home care at 29 percent. Moreover, there are more than one million new pressure ulcer cases annually and more than 60,000 deaths every year are associated with pressure ulcers. The treatment of pressure ulcers requires nursing care by especially educated personnel, the use of special support surfaces, nutrition supplements and a variety of products that enhance the healing process of the pressure ulcer area. In the United States the estimated cost involved in management of pressure ulcers is estimated to be $6.4 billion annually and the healing cost of a single pressure ulcer can be as much as $50,000 [1].

The accurate diagnosis and the appropriate treatment are crucial because starting the treatment too late may lead to the development of more severe lesions which may be life-threatening; on the other hand starting early results in the unnecessary deploy-

* Corresponding Author: Dimitrios I. Kosmopoulos, National Centre for Scientific Research "Demokritos", Institute of Informatics and Telecommunications, 15310 Agia Paraskevi, Greece, Email: dkosmo@iit.demokritos.gr.

ment of expensive means and measures. In most cases the digital images are adequate for recognizing the case severity by experienced personnel. This indicates that digital images of pressure ulcers can be used for telemedicine purposes as pointed out by works such as [2]. Furthermore, the two-dimensional information contained in the images, which are analyzed by humans for diagnosis, could be analyzed by machines for the same purpose, provided that proper processing techniques exist and can be employed.

Many tests have been performed in the past, where experts were given images and were requested to identify the lesion severity based only on them, e.g., [3,2]. Furthermore, many researchers have been occupied with the analysis of pressure ulcer images especially for follow up monitoring. However, none of them has dealt with the problem of automated lesion classification. For example [4] proposes a tool using images for classifying the pressure ulcer and for healing monitoring but the related data are entered manually and the classification is done manually as well. In [5] the lesions are measured from photos using a 3D acquisition system employing a pattern-projecting device, so that the spatial data can be extracted, but the automated lesion classification problem is not handled. The idea of using pressure ulcer images for telemedicine purposes has been examined mainly for synchronous applications, where there is always a human who diagnoses the pressure ulcer stage e.g., in [6].

The major priority of the pressure ulcer prevention can best be accomplished by identifying individuals who are at risk for the development of pressure ulcers (conforming to commonly used scales such as the Norton [7] and Braden [8]) and by the initiation of early preventive measures. However, the timely diagnosis and treatment of pressure ulcers constitutes a challenge in patient rehabilitation, since there is a variety of classification systems and because highly skilled and experienced personnel is required. Consistency in diagnosis from various experts can not be easily achieved as mentioned, e.g., in [3], because the human factor itself implies subjective clinical judgement and bias. Moreover, the patients who develop pressure ulcers have serious mobility problems and therefore experts are required, who will be able to monitor them on-site at regular intervals. However, this is not always possible due to the worldwide nursing shortage (see e.g., [9]) since pressure ulcer management services are mainly run by nurses. Of course the problem is even more acute in areas with difficult access. An additional factor that has to be considered is that frequently the pressure ulcers are identified at later stages and in such cases immediate actions are required. Therefore, systems that would be able to perform automated diagnosis in an *objective* and *consistent* manner are of obvious utility.

In this work we examine the potential of using digital images to classify regions appearing in pressure ulcer images and thus being able to have a diagnosis about the patient's current state. According to our knowledge no other similar study has been conducted in the past. So far we have used already available images from anonymous medical archives. The images were produced without following any strict procedures thus several uncertainties were introduced as we will explain in the following. However, we believe that this choice is reasonable for this preliminary study, given the fact that the creation of a consistent and adequate dataset depends on patient availability and willingness to be photographed; therefore the process may last for several months or years.

In this chapter we describe a methodology for characterising pressure ulcer lesions by image segmentation using colour and texture information and then by classifying

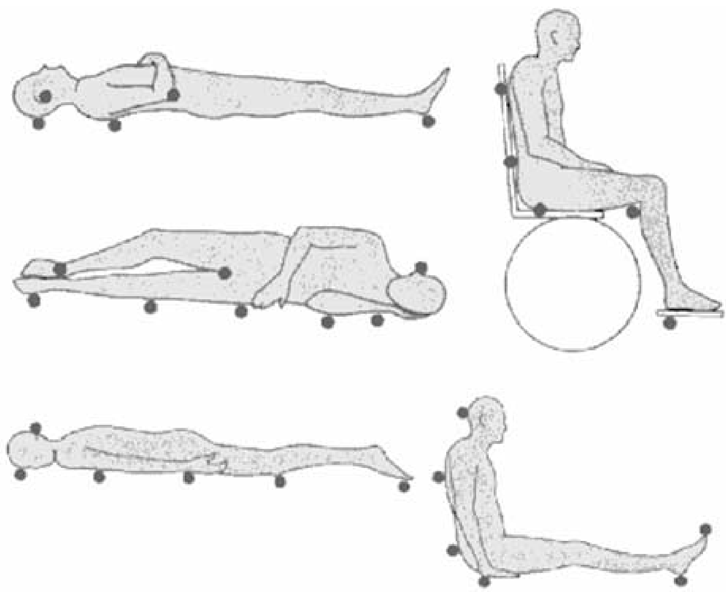


Figure 1. Common locations of pressure ulcer development (source: [10]).

the segments. The methodology can be applied to segmentation–based image classification tasks, provided that colour and texture can give meaningful information.

The rest chapter is structured as follows: in Section 2 we describe the pressure ulcer characteristics and its various grades; in Section 3 we describe the processing method requirements; in Section 4 we describe the method giving details about the segmentation, feature definition and extraction and the classification; in Section 5 we describe the experimental results; Section 6 concludes this chapter.

2. Pressure Ulcer Characteristics

The National Pressure Ulcer Advisory Panel (NPUAP) define as pressure ulcer *every area of localised damage to the skin and underlying tissue caused by pressure, shear, friction and/or a combination of them* [1].

Pressure ulcers are usually located over bony prominences and are staged to classify the degree of tissue damage observed. The pressure ulcers occur if a wheelchair is used or if the person is bedridden, even for a short period of time (for example, after surgery or an injury). The constant pressure against the skin reduces the blood supply to that area, and the affected tissue dies. The most common places for pressure ulcers are over bony prominences (bones close to the skin) like the heels, back, elbow, hips, ankles, shoulders, and the back of the head (see Fig. 1). A pressure ulcer starts as reddened skin but gets progressively worse, forming a blister, then an open sore, and finally a crater.

Various classification grading systems have been developed to assist clinical personnel with gathering consistent information and to define and describe skin damage and to improve diagnosis, treatment and allocation of resources. The classification system proposed by European Pressure Ulcer Advisory Panel (EPUAP) defines the following grades of severity [11]:

- Grade 1: non-blanchable erythema of intact skin. Discolouration of the skin, warmth, oedema, induration or hardness may also be used as indicators, particularly on individuals with darker skin.
- Grade 2: partial thickness skin loss involving epidermis, dermis, or both. The ulcer is superficial and presents clinically as an abrasion or blister.
- Grade 3: full thickness skin loss involving damage to or necrosis of subcutaneous tissue that may extend down to, but not through underlying fascia.
- Grade 4: extensive destruction, tissue necrosis, or damage to muscle, bone, or supporting structures with or without full thickness skin loss.

Additional classes that may coexist with the ones above in pressure ulcers are:

- Black necrosis, which is a sign of dead tissue in arterial insufficiency (of dark colour), which can also be seen in parts of an ulcer after an infection.
- White necrosis, which indicates dead tissues and is a preliminary stage of black necrosis (of light colour).

The aforementioned classes are depicted in Fig. 2a.

3. Requirements

A classification tool for skin lesions has to fulfill, apart from ease of use, several other requirements so that it is of practical use to the related personnel. We have identified the following:

- *Robustness to acquisition conditions*. The tool has to provide results under a variety of conditions. It can not be assumed that the image acquisition conditions will be uniform or ideal. Therefore, the noise and reflections in the images have to be treated accordingly. Furthermore, the distance of the image acquiring camera, as well as the image orientation, are not known in advance.
- *Discrimination of regions based on colour*, since colour is one of the most discriminative features for lesion classification into the different stages.
- *Discrimination of regions based on texture*, since texture can provide useful information about the pressure ulcer stage.
- *Consideration of patient's skin colour*, since the skin colour in case of stage 1 ulcer is quite similar to skin. Different colour for stage 1 ulcers is expected when comparing dark skinned patients with light skinned ones.
- *Classification accuracy*, due to the criticality of the task.
- *Ability to classify correctly coexisting different lesion stages*. Different pressure ulcer stages may coexist around the same lesion. It is not correct to classify the whole ulcer to a single class since the separate regions have different properties and consequently need different treatment.
- *Acceptable performance* to enable use for high volumes of data and to ensure response within reasonable time. The performance issue becomes more important as the scale of use grows.

These requirements have been used as a guide for developing the proposed method.

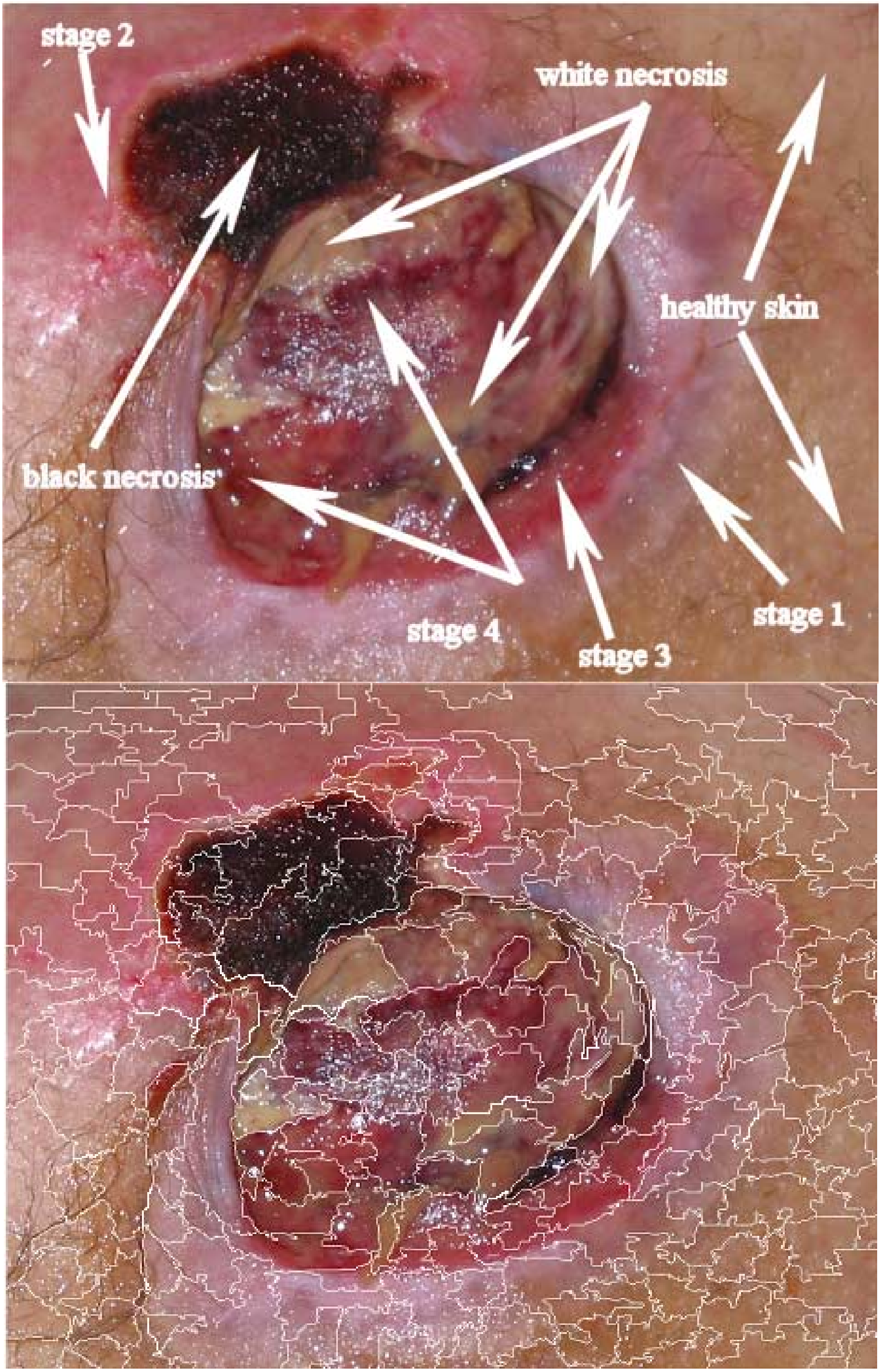


Figure 2. (a) Example of coexisting classes in the same pressure ulcer lesion (b) The same image after segmentation.

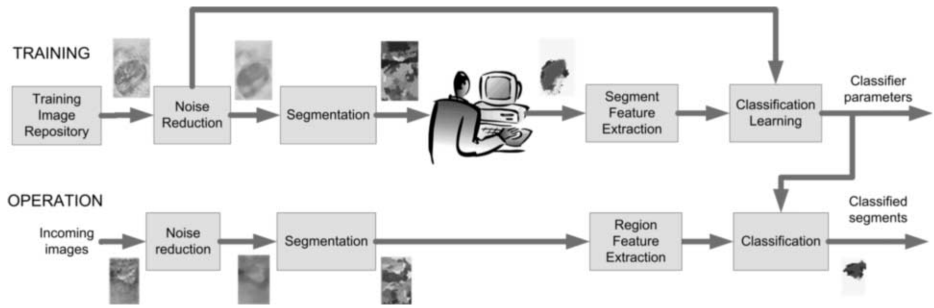


Figure 3. Processing stages for training and operation modes.

4. Classification Method

In Fig. 3 the classification process is depicted. The prototype system that we created works in two modes: *training*, during which it learns the characteristics of the pressure ulcer classes and *operation*, during which image segments are classified. A brief description of the process is given in the following, while the image processing and the classification are described in more detail in the next section.

The *training* process starts by retrieving selected images from a repository, which has to include as many representative cases as possible. After noise filtering the images are automatically segmented to find regions with homogeneous characteristics. The image segments are presented to the user, who then labels them according to the class she/he believes that they belong to. After labeling, for each region colour and texture features are extracted. Then the features are input to a learning module, which determines the classifier parameters, which will be used during the operation mode.

In the *operation* mode the tool performs classification using the classifier parameters defined in training. The images coming from remote locations are processed in a fashion similar to training (noise reduction, segmentation and feature extraction) without user intervention. The calculated features per region are input to the trained classifier, which then provides a classification result (pressure ulcer stage) for each segment separately.

The segmentation, the segment representation and the classification steps are described in the following.

4.1. Segmentation

As stated above, one of the primary goals of the tool is to segment the regions according to their colour and texture. It is noticed that the lesions have discriminative colour and texture that differentiate them from the surrounding tissues, so colour and texture can provide the primary features for their detection. Furthermore, the colours regarding each stage are distinct too, thus enabling us to use them for classification.

Due to non-uniform illumination we expect that reflections will be present mainly in a form similar to “salt and pepper” noise due to the uneven nature of the skin and the ulcer (however the noise can be dense in some regions). This noise can be minimised by applying median filtering. The median filtering is undesired in the general case, due to the fact that it eliminates ulcer detail, which may be useful for texture feature extraction; therefore we limit its application to regions where we have unusually low saturation values.

In our case the segmentation algorithm has to consider both colour and colour variability within the examined region; texture represents exactly this variability in our images. This means that regions in the image having either similar colour or similar type of colour variability have to be considered as separate segments.

Graph-based segmentation algorithms are good candidates for this purpose due to the fact that they can easily capture the variability induced by texture. The image is represented as an undirected graph $G = (V,E)$, where each pixel p_i has a corresponding vertex $v_i \in V$ and $(v_i,v_j) \in E$ are the edges corresponding to neighbouring vertices. The weight of the edges represents the dissimilarity measure between connected vertices (high dissimilarity is expected between different segments). A segmentation S is a partition of V into components such that each component (or region) $C \in S$ corresponds to a connected component in a graph $G' = (V,E')$, where $E' \in E$. In other words, any segmentation is induced by a subset of the edges in E.

Intuitively the intensity differences across the boundary of two regions are perceptually important if they are large relative to the intensity differences inside at least one of the regions. Therefore, edges between two vertices in the same component should have relatively low weights, and edges between vertices in different components should have higher weights.

Several graph-based segmentation algorithms exist such as [12–16]. We use the method proposed by [16] due to its effectiveness and efficiency. It is a characteristic example of graph-based methods and thus worthy to examine further. It is able to capture perceptually distinct regions, even though their interior is characterized by large variability, by considering global image characteristics. It follows the general concept of graph-based methods and measures the evidence of a boundary between two regions by computing (a) intensity differences across the boundary and (b) intensity differences between neighbouring pixels within each region.

The internal difference $Int(C)$ of a component $C \in V$ is defined to be the largest weight in the minimum spanning tree of the component, $MST(C,E)$ (the minimum spanning tree is a spanning tree with the smallest weight among all spanning trees connecting the nodes of a graph and can be calculated by Kruskal's algorithm, see e.g. [17]). The difference $Dif(C_1,C_2)$ between two components $C_1,C_2 \in V$ is defined to be the minimum weight edge connecting the two components. If there is no edge connecting C_1 and C_2 then the difference measure takes an infinite value. Adopting as a measure for difference the median weight instead of the minimum distance would offer robustness to outliers, but this would imply solving an NP-hard problem and the authors claim that using the former option still gives good results.

The region comparison predicate evaluates if there is evidence for a boundary between a pair or components by checking if the difference between the components, $Dif(C_1,C_2)$, is large relative to the internal difference within at least one of the components, $Int(C_1)$ and $Int(C_2)$. A threshold function is used to control the degree to which the difference between components must be larger than minimum internal difference. In [16] the authors define a pairwise comparison predicate $D(C_1,C_2)$, as:

$$\begin{aligned} D(C_1,C_2) &= true \text{ if } Dif(C_1,C_2) > MInt(C_1,C_2) \\ &\quad else\ D(C_1,C_2) = false \end{aligned} \tag{1}$$

where the minimum internal difference, *MInt*, is defined as:

$$MInt(C_1,C_2) = \min(Int(C_1) + \tau(C_1), Int(C_2) + \tau(C_2)) \tag{2}$$

The threshold function τ controls the degree to which the difference between two components must be bigger than their internal differences in order for there to be evidence of a boundary between them. To ensure scalability, in [16] a threshold function based on the size of the component is used:

$$\tau(C) = k / |C| \tag{3}$$

where $|C|$ denotes the size of C, and k is some constant parameter. Of course the selection of k is crucial for the results returned by the algorithm and there is no way of defining this parameter automatically. The selection of k is related to the specific type of images that have to be segmented and is defined by the user through observation.

The algorithm uses a greedy approach and is described in the following. The input is a graph $G = (V,E)$, with n vertices and m edges. The output is a segmentation of V into components $S = (C_1, \ldots, C_r)$.

1. Sort E into $\pi = (o_1, \ldots, o_m)$, by non-decreasing edge weight.
2. Start with a segmentation S_0, where each vertex v_i is in its own component.
3. Repeat step 4 for $q = 1, \ldots, m$.
4. Construct S_q given S_{q-1} as follows. Let v_i and v_j denote the vertices connected by the q-th edge in the ordering, i.e., $o_q = (v_i, v_j)$. If v_i and v_j are in disjoint components of S_{q-1} and the weight $w(o_q)$ is small compared to the internal difference of both those components, then merge the two components otherwise do nothing.
5. Return S_m

The above algorithm is implemented to run in $O(m \log m)$ time, where m is the number of edges in the graph. For the colour images that we examine the intensity difference is calculated for each channel separately. We have used the 8-neigborhood to define the edges for each pixel-vertice.

4.2. Feature Definition and Extraction

The lesions are composed of segments and each segment is represented by a feature vector, which includes features related to colour and texture.

The calculated colour features are based on colour histograms in the HSV space. From these values we exclude the ones with very low saturation because they belong most probably to reflections. The vector is given by

$$\mathbf{f}_c = [h_1 \ldots h_b\ s_1 \ldots s_b\ v_1 \ldots v_b] \tag{4}$$

where b is the bin number. Due to varying illumination conditions and different skin colours we use as additional feature the Bhattacharya distance of each region from the colour of healthy skin given by the following relation (for each channel):

$$D = \sum_{i=1..b} \sqrt{p_i(x) q_i(x)} \tag{5}$$

So the related feature vector is given by:

$$\mathbf{f}_{\mathrm{d}} = [h_d \; s_d \; v_d] \tag{6}$$

Since the human skin varies within certain limits in the colour space this feature is quite discriminative and less sensitive to illumination changes.

For texture processing we used for each region the Gabor functions, which form a complete but non-orthogonal basis set, which can be used for expanding the signal to provide a localized frequency description. A class of self-similar functions, which compose the "filter dictionary", referred to as Gabor wavelets are obtained by appropriate dilations and rotations of a "mother" Gabor function. The redundancy of information due to non-orthogonality is reduced by appropriate definition of scale factors and filter parameters as shown in [18].

A 2D Gabor function g(x,y) and its Fourier transformation G(u,v) are defined by:

$$g(x,y) = \frac{1}{2\pi\sigma_{\mathrm{x}}\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_x^2} \right) + 2\pi jWx \right] \tag{7}$$

$$G(u,v) = \exp\left[-\frac{1}{2}\left(\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right) \right] \tag{8}$$

where $\sigma_u = 1/2\pi\sigma_x$ and $\sigma_v = 1/2\pi\sigma_y$. The texture features are calculated using Gabor wavelets for each region. Due to the unknown pose of the camera relatively to the target ulcer we calculate the features in several scales (achieved by reducing resolution) and orientations (achieved by rotating the filters).

If $g(x,y)$ is the mother Gabor wavelet

$$g_{mn}(x,y) = \alpha^{-m} G(x',y') \quad \alpha > 1,\ m,\ n \text{ are integer} \tag{9}$$

$$x' = a^{-m}(x\cos\theta + \mathrm{y}\sin\theta) \text{ and } y' = a^{-m}(-x\sin\theta + \mathrm{y}\cos\theta) \tag{10}$$

where $\theta = n\pi/K$ and K is the total number of orientations. The values of σ_u, σ_v are calculated as shown in [18] to minimize the redundancy of the non-orthogonality.

Given an image I(x,y) its Gabor wavelet transform is defined by:

$$W_{mn}(x,y) = \int \mathrm{I}(x,y) g^*_{mn}(x - x_1, y - y_1) dx_1 dy_1 \tag{11}$$

where * indicates the complex conjugate.

We use as features the mean μ_{mn} and the standard deviation σ_{mn} of the transform coefficients magnitude $|W_{\mathrm{mn}}(x,y)|$ as in the following vector:

$$\mathbf{f}_{\mathrm{t}} = [\mu_{00} \;\; \sigma_{00} \;\; \ldots \;\; \mu_{mn} \;\; \sigma_{mn}] \tag{12}$$

The size and range of histogram bins in the above vectors are defined using a vector quantization scheme to achieve optimal representation and to limit the number of bins due to performance considerations [19]. Here we have used the Lloyd algorithm [20], which finds the partition and the codebook with the minimum error based on the probability density function calculated by a training set. This is achieved iteratively, giving as input a random initial value for the partition and codebook vectors. The error is calculated as the mean square error of the quantized training data. In the following the partition and codebook vectors are redefined for the next iteration until a convergence criterion is satisfied for the vectors.

Finally the overall feature vector is given by:

$$\mathbf{f}_t = [\mathbf{f}_c\ \mathbf{f}_d\ \mathbf{f}_t] \tag{13}$$

4.3. Classification

For the segment classification task we employ the Support Vector Machine (SVM) classifier, which is known for its computational efficiency and effectiveness, even for high dimensional spaces for classes that are not linearly separable [21]. Alternative methods could have been classifiers based on Neural Networks, Naïve Bayesian methods, Decision Trees etc. However, since the SVM methods are considered to be the state of the art in classification methods we don not expect significant improvement by testing the alternatives.

We have placed more emphasis on selecting the kernel and the parameters that are most appropriate for this problem. Such common kernels are the linear, the radial basis, the polynomial, the sigmoid function each of which has its own parameters. Another parameter that has to be defined is the *C* (trade-off between training error and margin). All these can not be defined generally, but after experimentation with the specific problem.

At the training phase we need to provide the extracted feature vector **f** along with labels indicating if the features correspond to a class or not for each of the seven classes. The SVM gives as output a number of support vectors, which are then used in operation mode for separating the classes. The output of the classification is a map of the image, where the different classes are displayed. Neighboring regions belonging to the same class appear as single regions.

5. Experimental Results

The pressure ulcer data used in this initial study have been acquired in various hospitals under non uniform conditions by the nursing personnel. The images belong to medical archives and they were not planned to be used for our research at the time they were acquired. Of course using non-standard acquisition procedures made the classification task quite difficult due to the uncertainties that were introduced in colour and texture analysis.

The image acquisition has been performed using mainstream high resolution CCD cameras with a single colour sensor. A flash module was enabled providing non-uniform illumination, thus affecting adversely the acquisition process. The non-uniform flash illumination resulted in several reflections, which affected adversely the classification.

Table 1. Experimental classification results

Class	False positive rate (%)	False Negative rate (%)
Stage 1	82.75	15.65
Stage 2	76.56	19.32
Stage 3	81.68	22.33
Stage 4	82.51	19.23
White necrosis	88.78	14.33
Black necrosis	91.77	7.15

The images were acquired from a variety of distances using several lenses of different length. No standard distance or focal length was used making texture analysis less reliable. The images of lower resolution were used as reference for the texture features calculation, i.e., only the scales that matched these lower resolution images were used for calculating texture so that all images could be comparable.

The images were in RGB format, with low or no compression at all, to ensure high quality, using 8 bits per colour channel. The colour was not normalized during processing.

The images at our disposal were eighty five and each of them could provide as many as fifty segments, providing samples of different classes. The rectangular regions of interest, which included the pressure ulcer, were manually defined by the operator and then processed by the system for segmentation and classification. Each region included several hundred to several thousand pixels. Approximately one hundred samples were used for training each class. For each patient we also included regions representing healthy skin tissues to compare with pressure ulcer regions in stage 1, which looks quite similar to skin. During the segmentation process we favoured the over-segmentation in order to have as compact regions as possible, but we also defined a minimum segment size, so that we avoided very small regions. The segmentation result was satisfactory in most cases (see Fig. 2b).

For classification we selected the Gaussian radial basis function after experimenting with the *C* parameter and the exponential coefficient gamma. Other kernels and their parameters were also tested with rather inferior results. The training and test samples were of equal size. We performed 100 experiments and in each of them the test and training images were randomly selected. The overall task for each image was completed in less than 30 seconds in a standard PC for images of about 5Mpixels.

The classification error false positive and false negative rates at the region level are presented in Table 1 for the case of Gaussian radial basis kernel, which provided the best results. Most misclassifications occurred between close stages (difference one).

6. Conclusions

An initial study for characterization of pressure ulcers through digital imaging has been presented. According to our knowledge this is the first time that such as study is performed. The images are segmented and the segments are separately classified using colour and texture features.

The experimental results can justify that the first approach for this challenging problem is promising despite the difficulties posed by the illumination and scale variability (due to unconstrained environmental conditions).

The error factors that are inherent in the proposed processing are stemming from (a) the variability of colour due to non-uniform illumination, reflections and image sensor properties (b) to scale variations, which affect texture (c) to segmentation errors, and (d) to separation errors during training. The errors stemming from (a) are partially compensated by using the healthy skin colour as reference. Errors stemming from (b) can be reduced if a more controlled acquisition process can be executed, i.e., if approximate target distance and camera focal length can be determined, so that the appropriate processing scale can be determined. Errors of type (c) were not that significant to influence the system outcome. Errors stemming from (d) can be reduced by using more specific image features and more complex kernel functions. Our goal is to further automate the feature calculation by applying additional criteria, e.g., class adjacency, to reduce classification errors and to simplify the acquisition process. The scale problem will be able to be automatically calculated by including in the image a simple pattern of known dimensions (marker), placed next to the target.

Acknowledgement

The authors would like to thank the Greek subsidiary of Coloplast S.A. for offering the images for this research.

References

[1] National Pressure Ulcer Advisory Panel, www.npuap.org.
[2] Houghton P.E., Kincaid C.B., Campbell K.E., Woodbury M.G., Keast D.H. Photographic assessment of the appearance of chronic pressure and leg ulcers. *Ostomy Wound Management.* 46(4): 20–6, 28–30, 2000.
[3] T. Defloor, L. Schoonhoven, “Inter-rater reliability of the EPUAP pressure ulcer classification system using photographs”, Journal of Clinical Nursing, 13, 952–959, 2004.
[4] H. Sanada et al., Reliability and validity of DESIGN, a tool that classifies pressure ulcer severity and monitors healing, Journal of Wound Care, 13(1), 13–18, 2004.
[5] C. Roques, L. Theot, N. Frasson, S. Meaume, “PRIMOS: an optical system that produces three-dimensional measurements of skin surfaces”, Journal of Wound Care, 12(9), 362–364, 2003.
[6] C. Mathewson, V.K. Adkins, M.L. Jones, “Initial Experiences With Telerehabilitation, and Contingency Management Programs for the Prevention and Management of Pressure Ulceration in Patients With Spinal Cord Injuries”, Journal of Wound Ostomy and Continence Nurses, 27:269–271, 2000.
[7] Waterlow J. Pressure sores: a risk assessment card. Nurs Times 1985; 81:49–55.
[8] Bergstrom N., Braden B.J., Laguzza A., Holman V. The Braden scale for predicting pressure sore risk. Nurs Res, 36:205–10, 1987.
[9] S.J. Ross, D. Polsky, J. Sochalski, “Nursing shortages and international nurse migration”, International Nursing Review, 52 (4), 253–262, 2005.
[10] NSW Health, http://www.health.nsw.gov.au.
[11] European Pressure Ulcer Advisory Panel, www.epuap.org.
[12] Shi, J. and Malik, J. Normalized cuts and image segmentation, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 731–737, 1997.
[13] Weiss,Y., Segmentation using eigenvectors: A unifying view. Proceedings of the International Conference on Computer Vision, 2:975–982. 1999.
[14] Jermyn, I. and Ishikawa, H. Globally optimal regions and boundaries as minimum ratio weight cycles. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23:1075–1088, 2001.

[15] Boykov, Y., Funka-Lea, G., Graph cuts and efficient N-D image segmentation, *International Journal of Computer Vision* 70 (2), pp. 109–131, 2006.
[16] P.F. Felzenszwalb, D.P. Huttenlocher, "Efficient graph-based image segmentation", International Journal of Computer Vision, 59(2), 167–181, 2004.
[17] T. Cormen, C. Leiserson, R. Rivest, Introduction to Algorithms, MIT Press, Cambridge MA, 2000.
[18] B.S. Manjunath, W.Y. Ma, "Texture features for Browsing and Retrieval of Image Data, IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(8), 837–842, 1996.
[19] S. Theodoridis, K. Koutroumbas, Pattern recognition, 3rd Edition, Academic Press, 2006.
[20] S.P. Lloyd, "Least squares quantisation in PCM, IEEE Transactions on Information Theory", 28(2), 129–137, 1982.
[21] N. Cristianini and J. Shawe-Taylor, "Support Vector Machines and other kernel-based learning methods", Cambridge University Press, 2000.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Reviewing State of the Art AI Systems for Skin Cancer Diagnosis

Ilias MAGLOGIANNIS* and Charalampos DOUKAS
Univ. of Aegean, Dept. of Information and Communication Systems Engineering
83200 Karlovasi, Greece

Abstract. In the recent years artificial intelligence and vision-based diagnostic systems for dermatology have demonstrated significant progress. In this chapter, we review these systems by firstly presenting the installation, the visual features used for skin lesion classification and methods for defining them. Then we describe how to extract these features through digital image processing methods, i.e., segmentation, registration, border detection, color and texture processing and then we present how to use the extracted features for skin lesion classification by employing artificial intelligence methods, i.e., Discriminant Analysis, Neural Networks, Support Vector Machines, Wavelets. We finally list all the existing systems found in literature that deal with the specific problem.

Keywords. AI medical systems, Skin Cancer, Pattern Analysis, Melanoma, Dermoscopy, Classification Methods

1. Introduction

The malignant melanoma is among the most frequent types of cancer and one of the most malignant tumors. Its incidence has increased faster than that of almost all other cancers and the annual incidence rates have increased on the order of 3–7% in fair-skinned populations in recent decades [1]. The advanced cutaneous melanoma is still incurable, but when diagnosed at early stages it can be treated without complications. However, the differentiation of early melanoma from other benign pigmented skin lesions (e.g., benign neoplasms or dysplastic naevi that simulate melanoma) is not trivial even for experienced dermatologists. The issue has attracted the interest of many researchers, who have developed systems for automated detection of malignant melanoma in skin lesions, which will be surveyed here.

The main design issues for a machine vision system for melanoma detection concern the image acquisition set up, the image processing and the classification methodology. More specifically the following issues have to be addressed:

1. How can we acquire good images?
2. How are the image features defined, i.e., what are we looking for?
3. How are these features detected in the image? (usually trivial for humans but non-trivial for machines).

* Corresponding Author: Dr. Ilias Maglogiannis, University of the Aegean, Dept. of Information and Communication Systems Engineering, 83200 Karlovasi, Greece; E-mail: imaglo@aegean.gr.

4. How many of the defined features should be used for optimal results? (feature selection).
5. Which classifiers are used and how is determined the "importance" of each feature to classification?
6. How can we assess the performance of a classifier?

In this chapter we provide an updated state of the art review regarding the latest developments in methodologies that answer the above questions and we describe the most important implementations concerning all the above issues. More specifically in Section 2 we provide a small introduction about the skin cancer detection problem, while in Section 3 we survey the existing techniques dealing with image acquisition and analysis of dermatological images. In Section 4 we discuss the computational intelligence aspect of the image classification problem and in Section 5 we summarize the characteristics of the most important implementations found in literature. Finally, Section 6 discuss the findings and concludes the chapter.

2. Introduction to Skin Cancer and Malignant Melanoma

The skin consists of a number of layers with distinct function and distinct optical properties. White light shone onto the skin penetrates superficial skin layers and whilst some of it is absorbed, much is remitted back and can be registered by a camera. The stratum corneum is a protective layer consisting of keratin-impregnated cells and it varies considerably in thickness. Apart from scattering the light, it is optically neutral. The epidermis is largely composed of connective tissue. It also contains the melanin producing cells, the melanocytes, and their product, melanin. Melanin is a pigment which strongly absorbs light in the blue part of the visible spectrum and in the ultraviolet. In this way it acts as a filter which protects the deeper layers of the skin from harmful effects of UV radiation. Within the epidermal layer there is very little scattering, with the small amount that occurs being forward directed. The result is that all light not absorbed by melanin can be considered to pass into the dermis. The dermis is made of collagen fibres and, in contrast to the epidermis, it contains sensors, receptors, blood vessels and nerve ends (see Fig. 1).

Pigmented skin lesions appear as patches of darker color on the skin. In most cases the cause is excessive melanin concentration in the skin. In benign lesions (e.g. common naevi) melanin deposits are normally found in the epidermis. Malignant melanoma is a skin cancer. It occurs when melanocytes reproduce at a high, abnormal rate (see Fig. 2). Whilst they and their associated melanin remain in the epidermis, melanoma is termed 'in situ'. At this stage it is not life-threatening and its optical properties make it conform to those of the normal, highly pigmented skin. When malignant melanocytes have penetrated into the dermis, they leave melanin deposits there, changing the nature of skin coloration.

The presence of melanin in the dermis is the most significant sign of melanoma. However, it cannot be used as a sole diagnostic criterion because in situ melanomas do not have dermal melanin. Moreover, some benign naevi have dermal deposits, although their spatial patterns tend to be more regular than in melanoma. Other signs, some of which can be indicative of melanoma in situ, are thickening of the collagen fibres in the papillary dermis (fibrosis); increased blood supply at the lesion periphery (erythemat

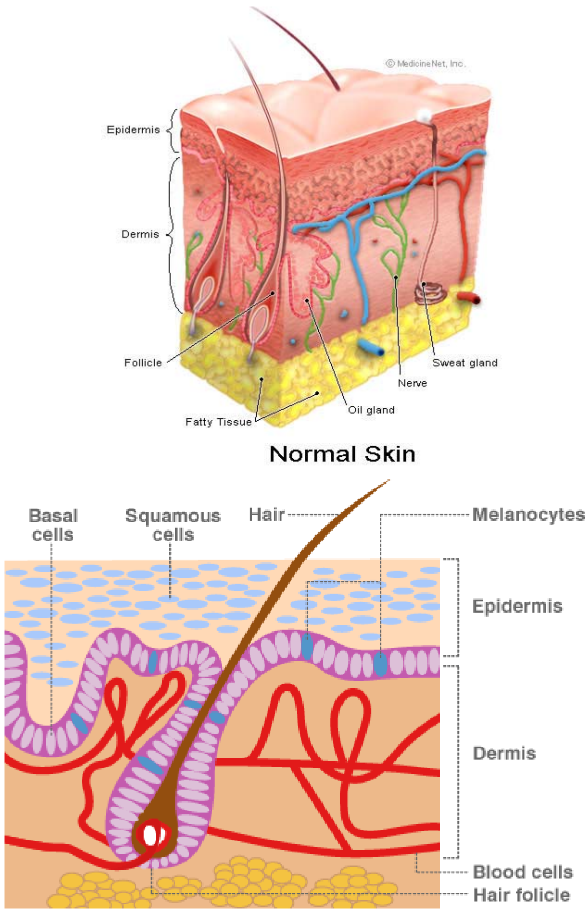


Figure 1. Normal skin lesions and main components (source: MediceNet).

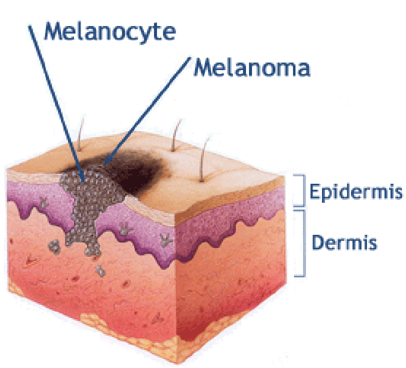
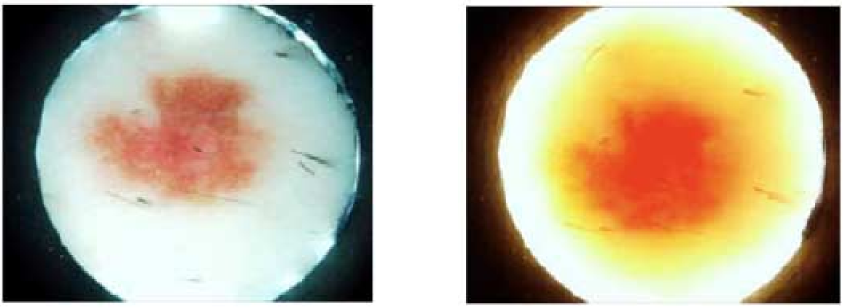


Figure 2. Illustration of Melanocytes and Melanoma on skin (source: MediceNet).



(a) (b)

Figure 3. (a) Typical skin lesion image obtained using ELM, (b) translumination image of the same skin lesion: Increased blood volume and vasculature are visible surrounding the area of pigmentation (source: [4]).

reaction); and lack of blood within the lesion, in the areas destroyed by cancer. The colors associated with skin which has melanin deposits in the dermis normally show characteristic hues not found in any other skin conditions. This provides an important diagnostic cue for a clinician. If the visual approach corroborates a suspicion of skin cancer, histology [2] is needed to make explicit diagnosis.

3. Image Acquisition and Analysis

3.1. Image Acquisition

The first step in machine vision-based expert systems involves the acquisition of the tissue digital image, which answers question 1 of the introductory section. The main techniques used for this purpose are the epiluminence microscopy (ELM or dermoscopy), transmission electron microscopy (TEM) and the image acquisition using still or video cameras. By placing a thin layer of oil on a lesion and then pressing a special hand-held microscope against the oil field on the patient's skin, ELM provides for a more detailed inspection of the surface of pigmented skin lesions and renders the epidermis translucent, making many features become visible. TEM on the other hand can reveal the typical structure of organization of elastic networks in the dermis and thus is mostly used for studying growth and inhibition of melanoma through its liposomes [3]. A recent method of EML imaging is Side-transillumination (transillumination). In this approach, light is directed from a ring around the periphery of a lesion towards its center at an angle of 45 degrees, forming a virtual light source at a focal point about 1 cm below the surface of the skin, thus making the surface and subsurface of the skin translucent. The main advantage of transillumination is its sensitivity to imaging increased blood flow and vascularization and also to viewing the subsurface pigmentation in a nevus. This technique is used by a prototype device, called Nevoscope, which can produce images that have variable amount of transillumination and cross-polarized surface light [4,5].

Recently new techniques have been presented, that use multispectral images. The chosen wavelengths interact preferentially with constituents of the skin and are able to reveal the structure of the skin lesion. An example is the work presented in [6]. In [7] in vivo confocal microscopy is described as a high-resolution imaging tool that allows noninvasive optical sectioning of live human skin and other accessible tissues in real time. A novel acquisition method presented in [8] that utilizes functional infrared imaging for skin cancer screening. Based on the fact that malign melanoma have higher metabolism as well as increased blood flow, it has been conjectured that the latter also have slightly higher temperature compared to the healthy skin that can be measured by high resolution functional infrared imaging.

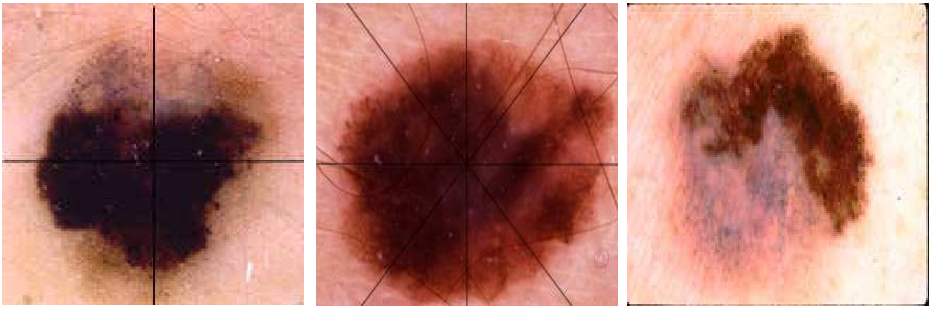
The construction of systems with the ability to capture reliable and reproducible images of skin is rather challenging due to equipment and environmental constraints, such as image resolution, image noise, illumination, skin reflectivity and poses uncertainty. The use of commercially available photographic cameras is quite common in skin lesion inspection systems, particularly for telemedicine purposes [9]. However, the poor resolution in very small skin lesions, i.e., lesions with diameter of less than 0.5 cm and the variable illumination conditions are not easily handled and therefore high-resolution devices with low-distortion lenses have to be used. However, the requirement for constant image colors, (necessary for image reproducibility) remain unsatisfied, as it requires real time, automated color calibration of the camera, i.e., adjustments and corrections to operate within the dynamic range of the camera and to measure always the same color regardless of the lighting conditions. The problem can be addressed by using video cameras [10] that are parameterizable online and can be controlled through software [11,12] at the price of higher complexity and costs. In addition to the latter, improper amount of immersion oil or misalignment of the video fields in the captured video frame due camera movement capture can cause either loss or quality degradation of the skin image. Acquisition-time error detection techniques have been developed [13] in an effort to overcome such issues. Finally, computed tomography (CT) images have also been used [14] in order to detect melanomas and track both progress of the disease and response to treatment.

3.2. Definition of Features for Detection of Malignant Melanoma

In this section we will examine the features, i.e., the visual cues that are used for melanoma detection, providing answers to question 2. Similarly to the traditional diagnosis procedure, the computer-based systems look for features and combine them to characterize the lesion as malignant melanoma or dysplastic nevus. The features employed have to be measurable and of high sensitivity, i.e., high correlation of the feature with malignant melanoma and high probability of true positive response. Furthermore, the features should have high specificity, i.e., high probability of true negative response. Although in the typical classification paradigm both factors are considered important (a trade-off expressed by maximizing the area under the Receiver-Operating-Characteristic curve), in the case of malignant melanoma the suppression of false negatives (i.e., increase of true positives) is obviously more important.

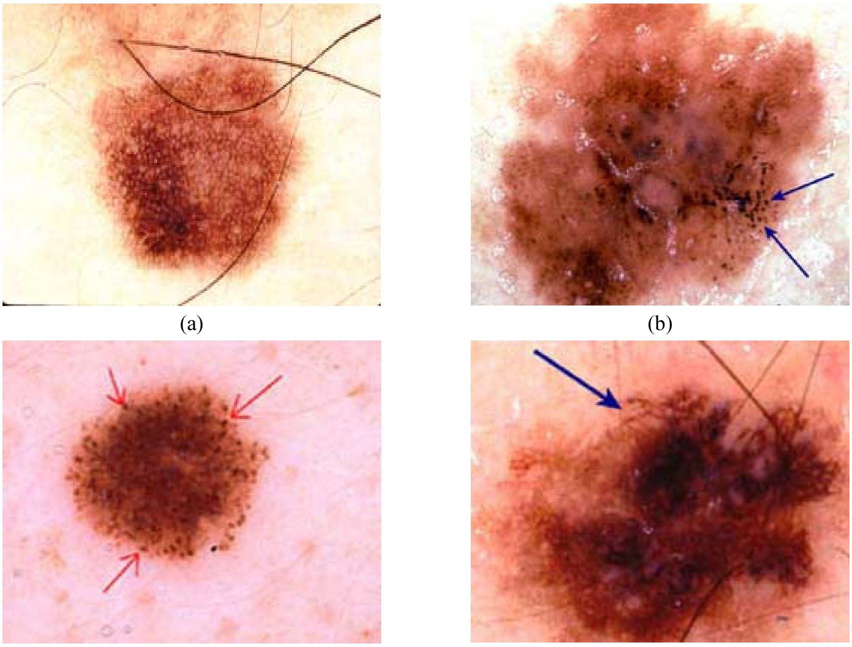
In the conventional procedure, the following diagnostic methods are mainly used [24]: (i) *ABCD rule* of dermoscopy; (ii) *Pattern Analysis*; (iii) *Menzies* method; (iv) *7-Point Checklist*; and (v) *Texture Analysis*. The features used for these methods are presented in the following.

The *ABCD rule* investigates the *asymmetry* (A), *border* (B), *color* (C) (Fig. 4), and *differential structures* (D) (Fig. 5) of the lesion and defines the basis for a diagnosis by a dermatologist. More specifically:



(a) (b) (c)

Figure 4. Asymmetry Border Color Features; (a) Asymmetry Test, (b) Border Test, (c) Color variegation (source: [25]).



(a) (b) (c) (d)

Figure 5. Differential Structures; (a) Pigmented network, (b) Dots, (c) Brown globules, (d) Branched streaks (source: [25]).

- Asymmetry: The lesion is bisected by two axes that are positioned to produce the highest symmetry possible, in terms of borders, colors, and dermoscopic structures.
- Border: Then skin lesion border is examined if it is irregular or if there is a sharp, abrupt cut-off of pigment pattern at the periphery of the lesion or a gradual, indistinct cut-off.
- Color: The color variation is determined. The chromatic values (average and standard deviation) are calculated for several channels. In addition, color tex-

ture features (correlation, entropy, etc) are used for determining the nature of melanocytic skin lesions [15].

- Differential structures: The number of structural components present is determined, i.e., Pigment Network, Dots (scored if three or more are present), Globules (scored if two or more are present), Structureless Areas (counted if larger than 10% of lesion), Streaks (scored if three or more are present).

The *Pattern Analysis* method seeks to identify specific patterns, which may be global (Reticular, Globular, Cobblestone, Homogeneous, Starburst, Parallel, Multicomponent, Nonspecific) or local (Pigment network, Dots/globules/moles [16], Streaks, Blue-whitish veil, Regression structures, Hypopigmentation, Blotches, Vascular structures).

The *Menzies* method looks for negative features (Symmetry of pattern, Presence of a single color) and positive (Blue-white veil, Multiple brown dots, Pseudopods, Radial streaming, Scar-like depigmentation, Peripheral black dots/globules, Multiple (5–6) colors, Multiple blue/gray dots, Broadened network).

The 7-point checklist [17,18] refers to seven criteria that assess both the chromatic characteristics and to the shape and/or texture of the lesion. These criteria are Atypical pigment network, Blue-whitish veil, Atypical vascular pattern, Irregular streaks, Irregular dots/globules, Irregular blotches, and Regression structures. Each one is considered to affect the final assessment with a different weight. The dermoscopic image of a melanocytic skin lesion is analyzed in order to evidence the presence of these standard criteria; finally, a score is calculated from this analysis, and, if a total score of 3 or more is given, the lesion is classified as melanoma otherwise is classified as nevus.

Texture Analysis is the attempt to quantify texture notions such as ‘fine’, ‘rough’ and ‘irregular’ and to identify, measure and utilize the differences between them. Textural features and texture analysis methods can be loosely divided into two categories: statistical and structural. Statistical methods define texture in terms of local gray-level statistics that are constant or slowly varying over a textured region. Different textures can be discriminated by comparing the statistics computed over different sub-regions. Neighboring gray-level dependence matrix (NGLDM) and lattice aperture waveform set (LAWS) are two textural approaches used for analyzing and detection the pigmented network on skin lesions [19].

The researchers that seek to identify automatically malignant melanoma exploit the available computational capabilities by searching for many of the above, as well as, for additional features. The main features used for skin lesion image analysis are summarized below, as well as their calculation method, answering question 3:

Asymmetry Features

The *asymmetry* is examined with respect to a point, one or more axes. The asymmetry index is computed by first finding the principal axes of inertia of the tumor shape in the image and it is obtained by overlapping the two halves of the tumor along the principal axes of inertia and dividing the non-overlapping area differences of the two halves by the total area of the tumor.

Border Information Extracture and Border Features

In order to extract border information, image segmentation is performed. It is considered to be a very critical step in the whole process of melanoma detection and involves the extraction of the region of interest (ROI), which is the lesion. Most usual methods are based on thresholding, region growing and color transformation (e.g., principal components transform, CIELAB color space and spherical coordinates, etc. [26]). Additional methods involving artificial intelligence techniques, like fuzzy borders [20] and declarative knowledge (melanocytic lesion images segmentation enforcing by spatial relations based declarative knowledge) are used for determining skin lesion features. The latter methods are characterized as region approaches, because they are based on different colorization among the melanoma regions and the main border. Another category of segmentation techniques are contour approaches using classical edge detectors (e.g., soble, canny, etc.) that produce a collection of edges leaving the selection of the boundary up to the human observer. Hybrid approaches [21] use both color transformation and edge detection techniques, whereas Snakes or active contours [22] are considered the most state-of-the art technique for border detection. More information regarding border detection as well as a performance comparison of the aforementioned methods can be found in [69].

The most popular *border* features are the Greatest Diameter, the Area, the Border Irregularity, the Thinness Ratio [53], the Circularity index [55], the variance of the distance of the border lesion points from the centroid location [55], and the Symmetry Distance [20]. The Circularity index (CIRC) is mathematically defined by the following equation:

$$\mathrm{CIRC} = \frac{4A\pi}{p2}, \tag{1}$$

where A is the surface of the examined area and p its perimeter. Symmetry Distance (SD) calculates the average displacement among a number of vertexes as the original shape is transformed in to a symmetric shape. The symmetric shape closest to the original shape P is called the symmetry transform (ST) of P.The SD of an object is determined by the amount of effort required to transform the original shape into a symmetrical shape, and can be calculated as follows:

$$\mathrm{SD} = \frac{1}{n}\sum_{i=0}^{n-1} \| Pi - \widehat{P}i \| \tag{2}$$

Apart from regarding the border as a contour, emphasis is also placed on the features that quantify the transition (swiftness) from the lesion to the skin [23]. Such features are the minimum, maximum, average and variance responses of the gradient operator, applied on the intesity image along the lesion border.

Color Features

Typical color images consist of the three-color channels RGB (red, green and blue). The color features are based on measurements on these color channels or other color channels such as CMY (Cyan, Magenta, Yellow), HSV (Hue, Saturation, Value), YUV (Y-luminance, U-V chrominance components) or various combinations of them, linear or not. The most typical preprocessing for color information extraction is (Feature of malignant melanoma based on color information): (a) image acquisition, (b) conversion to grey scale, (c) contrast emphasis, (d) thresholding, (e) noise reduction.

Color variegation may be calculated by measuring minimum, maximum, average and standard deviations of the selected channel values and by measuring chromatic differences inside the lesion [25]. Another method for computing skin colors based on normal skin structure model is presented in [2].

Differential Structures

The differential structures as described in the ABCD method, as well as most of the patterns that are used by the pattern analysis, the Menzies method and the 7-points checklist are very rarely used for automated skin lesion classification, obviously due to their complexity. A novel method presented in [15] uses three-dimensional pseudoelevated images of skin lesions that reveal additional information regarding the irregularity and inhomogeneity of the examined surface.

Skin Lesion Kinetics

Several efforts concern the kinetics of skin lesions e.g., [47,48]. In [49] the ratio of variances RV has been defined as

$$RV = \frac{SD_{B^2}}{SD_{B^2} + SD_{I^2} + SD_{A^2}} \tag{3}$$

SD_B^2 (Standard Deviation Between Days) is between day variance of the color variable computed using the mean values at each day of all wound sites and subjects.

SD_I^2 (Standard Deviation Intra Day) is the intra day variance of the color variable estimated from the computations at each day of all wound sites and subjects.

SD_A^2 (Standard Deviation Analytical) is the variance of the color variable computed using normal skin sites of all subjects and times.

3.3. Feature Selection

The success of image recognition depends on the correct selection of the features used for the classification, which answers question 4. This is a typical optimization problem, which may be resolved with heuristic strategies, greedy or genetic algorithms, other computational intelligence methods [27] or special strategies from statistical pattern recognition, (e.g., cross-validation (XVAL), leave-one-out method (LOO), Sequential forward floating selection (SFFS) and Sequential backward floating selection (SBFS)) [54]. The use of feature selection algorithms is motivated by the need for highly precise results, by computational reasons and by a peaking phenomenon often observed when classifiers are trained with a limited set of training samples. If the number of features is increased the classification rate of the classifiers decreases after a peak [28,29].

4. Computational Intelligence in Skin Lesion Classification

In this section we will answer questions 5 and 6 by examining the most popular methods for skin lesion classification. The task involves mainly two phases after feature selection, learning and testing [25], which are analyzed in the following.

During the learning phase typical feature values are extracted from a sequence of digital images representing classified skin lesions. The most classical recognition paradigm is statistical [30]. Covariance matrices are computed for the discriminative measures, usually under the multivariate Gaussian assumption. Parametric discriminant functions are then determined, allowing classification of unknown lesions (discriminant analysis). The major problem of this approach is the need for large training samples.

Neural networks are networks of interconnected nodes composed of various stages that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning. Learning occurs through training over a large set of data where the training algorithm iteratively adjusts the connection weights (synapses), by minimizing a given error function [33,34]. The weights of the features are thus automatically calculated. Popular choices for the error function in skin lesion image classification are the Euclidean Distance or the ratio deviation defined as follows:

$$E = \sqrt{\sum_i (x_i - \bar{x})^2}\ ,\ E = \sum_i \left|1 - \frac{x_i}{\bar{x}}\right| \tag{4}$$

where x_i is the ith sample and $\bar{x}$ is the population mean [31].

The Support Vector Machines (SVMs) is a popular algorithm for data classification into two classes [32,35,36]. SVMs allow the expansion of the information provided by a training data set as a linear combination of a subset of the data in the training set (support vectors). These vectors locate a hypersurface that separates the input data with a very good degree of generalization. The SVM algorithm is based on training, testing and performance evaluation, which are common steps in every learning procedure. Training involves optimization of a convex cost function where there are no local minima to complicate the learning process. Testing is based on the model evaluation using the support vectors to classify a test data set. Performance evaluation is based on error rate determination as test set data size tends to infinity.

The Adaptive Wavelet transform-based tree-structure classification (ADWAT) method [37] is a specific melanoma image classification technique that uses statistical analysis of the feature data to find the threshold values that optimally partitions the image-feature space for classification. A known set of images is decomposed using two-dimensional wavelet transform and the channel energies and energy ratios are used as features in the statistical analysis. The mean energy, e, of a sub-image, or channel, *f(m,n)* is calculated as

$$e = \tfrac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left|f(m,n)\right| \tag{5}$$

where M and N are the pixel dimensions of the image *f(m,n)* in the x and y directions respectively. The mean, variance, and the histogram of the feature values for each of the target classes are used to determine if a feature generates a unimodel distribution or segregates into a bi-modal distribution between the image classes. All pooled feature values that generate uni-modal distributions are rejected. For all features that segregate into bi-modal distributions, thresholds are calculated for optimal separation of the image classes. The latter are used for the creation of a threshold tree-structure. During the classification phase, the tree-structure of the candidate image obtained using the same decomposition algorithm is semantically compared with the tree-structure models of melanoma and dysplastic nevus. A classification variable (CV) is used to rate the tree-structure of the candidate image. CV is set to a value of 1 when the main image is de-

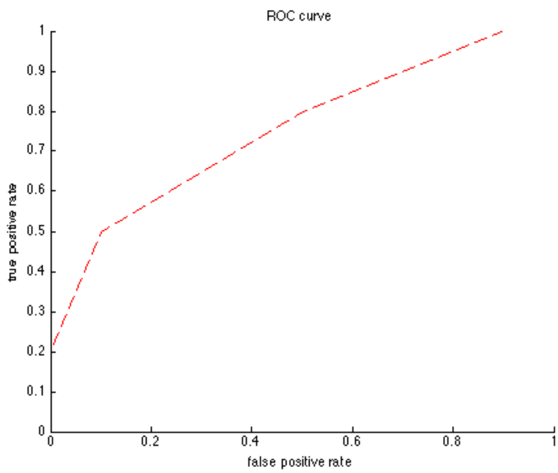


Figure 6. Example of ROC curve. X-axis represents the false positive rate (1-Sp, where Sp is the specificity) and the Y-axis the true positive rate (or Sensitivity, Se).

composed. The value of CV is incremented by one for every additional channel decomposed. When the algorithm decomposes a dysplastic nevus image, only one level of decomposition should occur (channel 0). Thus, for values of CV equal to 1, a candidate image is assigned to the dysplastic nevus class. A value of CV greater than 1 indicates further decomposition of the candidate image, and the image is accordingly assigned to the melanoma class.

Answering question 6 the performance of each classifier is tested using an ideally large set of manually classified images. A subset of them, e.g., 80% of the images is used as training set and the rest 20% of the samples is used for testing using the trained classifier. The training and test images are exchanged for all possible combinations to avoid bias in the solution. In small datasets an alternative approach is the stratified 10-fold cross validation procedure. In this case the image dataset is divided randomly into 10 parts in which the class is represented in approximately the same proportions as in the full dataset. Each part is held out in turn and the learning scheme trained on the remaining nine-tenths; then its error rate is calculated on the holdout set. Thus the learning procedure is executed a total of 10 times on different training sets (each of which have a lot in common). Finally, the 10 error estimates are averaged to yield an overall error estimate.

Most usual classification performance assessment in the context of melanoma detection is the True Positive Fraction (TPF) indicating the fraction of melanoma lesions correctly classified as melanoma and the True Negative Fraction (TNF) indicating the fraction of dysplastic or non-melanoma lesions correctly classified as non-melanoma, respectively [37,5]. A graphical representation of classification performance is the Receiver Operating Characteristic (ROC) curve (see Fig. 6), which displays the "tradeoff" between sensitivity (i.e. TPF) and specificity (i.e. TNF) that results from the overlap between the distribution of lesion scores for melanoma and nevi [38,39,10]. A good classifier is one with close to 100% sensitivity at a threshold such that high specificity is also obtained. The ROC for such a classifier will plot as a steeply rising curve. When different classifiers are compared, the one whose curve rises fastest should be best. If sensitivity and specificity were weighted equally, the greater the area under the ROC curve (AUC), the better the classifier [40]. An extension of ROC analysis found in literature [41] is the Three-way ROC analysis that applies to trichotomous tests. It summarizes the discriminatory power of a trichotomous test in a single value, called the volume under surface (VUS) in analogy to the AUC value for dichotomous tests. Just as the AUC value for dichotomous tests is equivalent to the probability of correctly ranking a given pair of normal and abnormal cases, the VUS value for trichotomous tests is equivalent to the probability of correctly distinguishing three cases, where each case is from a different class (e.g., malignant, benign, neutral skin spot).

5. Summary of the Most Important Implementations

The development of automated systems for the melanoma classification task preoccupies many biomedical laboratories, e.g., [45,46,50,52,54], which will be further examined in the following. It is also interesting to include in our survey works that deal with the general problem of skin lesion images characterization, because they face quite similar problems. The lesions include among others tumor, crust, hair, scale, shiny and ulcer [33,42], erythema [43], burn scars [44] and wounds [47,48].

The most common installation type seems to be the video camera, obviously due to the control features that it provides [33,42–45]. The still camera is of use in some installations, e.g., [47,48], while infra red or ultraviolet illumination (in situ or in vivo) using appropriate cameras is a popular choice, e.g., [50–52] correspondingly. Microscopy (or epiluminence microscopy) installations are applied in the works of [46,54] and digital videomicroscopy in [23] and [10].

The most common features that are used for automated lesion characterization are the ones that are associated with color in various color spaces (RGB, HIS, CIELab), e.g, color values in [33,42,43,56] and Colorbin (i.e., the percentage of the malignant melanoma colored foreground pixels) [56]. Some of them combine features in more than one color spaces for better results, e.g., HIS and RGB in [45–48,54], or RBG and colors peculiar to malignant melanomas [57]. The intensity characteristics are also used in works such as [44] and ratios of maximum to minimum intensity value [52]. Asymmetry and border features are quite common e.g., [54,56,23], while features based on differential structures are very rare. Some works [5,58,59] rely also on the whole ABCD rule for lesion characterization. Shape and color features, like Area and Elevation, calculated mannually by dermatologists have also been used [56].

The most common classification methods are the statistical and rule-based ones, e.g., [43,45,46,51,52,23,59]. More advanced techniques such as neural networks are presented in works like [33,42,56], while the k-nearest neighborhood classification scheme is applied in [54]. Evidence Theory (Upper and lower probabilities induced by multivalued mapping) based on the concept of lower and upper bounds for a set of compatible probability distributions is used in [62] for melanoma detection. Classification and Regression Trees (CART) [60] analysis has been used in [61]. Finally, [5,37] use ADWAT method for lesion classification.

The success rates for the methods presented in the literature indicate that the work towards automated classification of lesions and melanoma in particular may provide good results. These rates along with the other system features are summarized in Table 1. We should note here that the results are not comparable but rather indicative, mainly due to the fact that different images from different cases are used. Moreover, the classification succcess rates are not applicable to the methods calculating healing indexes.

Table 1. Computer-based systems for the characterization of digital skin images

Reference	Detection goal	Installation type	Visual Features	Classification method	Success rates
[33,42]	Tumor, crust, hair, scale, shiny ulcer of skin lesions	Video RGB Camera	Color (chromaticity) coordinates (more)	Neural networks	85–89% in average
[43]	Skin erythema	Video RGB Camera	Color – CIE L*a*b* color space	Statistical	Monitoring indexes for Follow ups
[44]	Burn scars	Video RGB Camera	Image Intensity, Skin Elasticity	Finite element analysis,	Monitoring indexes for Follow ups
[45]	Melanoma Recognition	Video RGB Camera	Color in RGB and HIS (more)	Statistical	5% deviation from manual diagnosis
[46]	Melanoma Recognition	Tissue microscopy	Epidermal and dermal features (epidermis volume, thickness, dermal epidermal junction ratio, cellular and collagen densities)	Statistical	Difference in epidermal features was 5.33%, for dermal features it was 2.76%
[48]	Wound Healing	Still CCD Camera	Ratio of variances, in HIS and RGB	Healing indexes measuring, the wound area and the wound color.	Monitoring indexes for Follow ups
[51]	Melanoma Recognition	In situ, ultraviolet illumination	Auto fluorescence of skin tissues	Statistical	77% (81% manual diagnoses)
[52]	Melanoma Recognition	Ultraviolet illumination	Imax/Imin, (fluorescence intensity)	Statistical	Sensitivity of 82.5%, specificity of 78.6% positive predictive value of 58.9% (Average values 14.3 for melanoma, 5.7 for naevi and 6.1 for other skin lesions)
[54]	Melanoma Recognition	Epilumines-cence microscopy (ELM)	RGB/HIS/Border	Statistical (k-nearest-neighbor)	Sensitivity of 87% and a specificity of 92%
[62]	Melanoma Recognition	ELM	Color	Statistical (Evidence Theory)	Sensitivity of 85.2% and a specificity of 72.22%

Table 1. (Continued.)

Reference	Detection goal	Installation type	Visual Features	Classification method	Success rates
[57]	Melanoma Recognition	ELM	Color (RGB space including color information peculiar to malignant melanomas)	Ruled-based classification	Detection rate of 26%
[63]	Melanocytic Lesion Segmentation	ELM	HSV Color model and Histogram information	Declarative Knowledge (Using segmentation rules on Histogram information)	See reference for more information on results
[56]	Melanoma Recognition	No information provided	Irregularity, Asymmetry Index, Average RBG inside the tumor, Colorbin, Variance of Local Average Color for RGB, Area and Elevation	Neural Network (back-propagation learning)	Sensitivity of 81% and specificity of 86.7%
[23]	Melanoma Recognition (focused on distinguishing between nevi and melanoma)	Video microscopy	Lesion boundary using luminance values.	Statistical (Multivariate discriminant analysis)	Sensitivity of 85.9% and a specificity of 74.1 %
[10]	Melanoma Recognition	Video microscopy	Boundary shape, mass and color distribution	Support Vector Machines (SVM)	Sensitivity of 100% and specificity of 63.65%
[4]	Melanoma Recognition	Side-transillumination (using Nevoscope)	Correlation coefficient, edge strength and Lesion size.	Scoring system based on segmentation results	Up to 95% success ration (using images with correlation > 0.90)
[58]	Melanoma Recognition	Digital ELM	ABCD rule	Neural Networks	90% correlation between manual and automated assessment
[37]	Melanoma Recognition	Epi-illumination (using Nevoscope)	ABCD rule	Adaptive Wavelet Transform-based Tree-structure Classification (ADWAT)	Sensitivity of 93.33% and a specificity of 91.12%
[59]	Melanoma Recognition	See reference for more information	ABCD rule	Rule-based classification	Sensitivity between 74.2%–86% and specificity between 83.2%–86.3%

Table 1. (Continued.)

Reference	Detection goal	Installation type	Visual Features	Classification method	Success rates
[61]	Melanoma Recognition	ELM	Histogram features (mean value, standard deviation, skewness, kurtosis and entropy of the grey level distribution)	Classification and Regression Trees (CART)	Sensitivity of 92.09% and a specificity of 92.734%
[64]	Melanoma Recognition	Side-transillumination (using Nevoscope)	Texture analysis	SVM (using 4^{th} – degree polynomial kernel)	70% average accuracy

6. Discussion and Conclusions

The most remarkable systems for the automated detection of malignant melanoma have been surveyed. These systems employ a variety of methods for the image acquisition, the feature definition and extraction as well as the lesion classification from features.

The most promising image acquisition techniques appear to be those that reveal the skin structure through selected spectral images. However, the problem of repeatability of the measurements for follow-up studies has not been satisfactorily resolved.

Regarding the features, it is clear that the emphasis has been on assessment of lesion size, shape, color, and texture. These statistical parameters were chosen primarily for computational convenience; they can be acquired with well-established analytic techniques at a manageable computational cost. However, they do not correspond to known biological phenomena and do not model human interpretation of dermoscopic imagery. On the contrary, the structural patterns that are considered essential for manual lesion categorization seem to have been neglected by the computational intelligence community, due to their complexity, although their exploitation could provide crucial information.

As far as the classification method is concerned, the SVM seems to perform better. However, it is actually the selected features that are critical for the performance of the classifier and the training procedure as well, which has to include the biggest possible variety of cases.

The results presented so far, from the research community are promising for the future. It is now necessary to examine more patients in order to increase the number of cases, particularly during the classification phase. This will clarify the issue of selecting the most powerful variables for classification and may also enable even better classification if examination of the differences in results between the two methods casts light on why misclassifications can arise.

Further Reading

More information on skin melanoma, dermatological features, medical detection and inhibition can be found in [65]. Related work in the context of machine learning methods comparison for diagnosis of pigmented skin lesions is presented in [41,66,67]. Finally, for an overview of machine learning methods, techniques and tools, [68] is recommended.

References

[1] Marks R. Epidemiology of melanoma. Clin Exp Dermatol 2000; 25:459–63.

[2] E. Claridge, S. Cotton, P. Hall, M. Moncrieff, From colour to tissue histology: Physics-based interpretation of images of pigmented skin lesions, Medical Image Analysis 7 (2003), 489–502.

[3] T.L. Hwang, W.R. Lee, S.C. Hua, J.Y Fang, Cisplatin encapsulated in phosphatidylethanolamine liposome enhances the in vitro cytotoxicity and in vivo intratumor drug accumulation against melanomas, Journal of Dermatological Science (2007) 46, 11–20.

[4] G. Zouridakis, M. Doshi, N. Mullani, Early Diagnosis of Skin Cancer Based on Segmentation and Measurement of Vascularization and Pigmentation in Nevoscope Images, Proc. Of the 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA, Sept. 1–5, 2004.

[5] S.V. Patwardhan, S. Dai, A.P. Dhawan, Multi-spectral image analysis and classification of melanoma using fuzzy membership based partitions, Computerized Medical Imaging and Graphics 29 (2005) 287–296.

[6] Moncrieff M., Cotton S., Claridge E., Hall P. (2002) Spectrophotometric intracutaneous analysis – a new technique for imaging pigmented skin lesions. British Journal of Dermatology 146(3), 448–457.

[7] S. Singh, J.H. Stevenson, D. McGurty, An evaluation of Polaroid photographic imaging for cutaneous-lesion referrals to an outpatient clinic: a pilot study, British Journal of Plastic Surgery (2001), 54, 140–143.

[8] T.M. Buzug, S. Schumann, L. Pfaffmann, U. Reinhold, J. Ruhlmann, Functional Infrared Imaging for Skin-Cancer Screening, Proc. Of the 28th IEEE EMBS Annual International Conference, New York City, USA, Aug. 30–Sept. 3, 2006.

[9] Loane M., Gore H., Corbet R., Steele K., "Effect of Camera performance on diagnostic accuracy", Journal of Telemedicine and Telecare, 3, pp. 83–88, 1997.

[10] M. Amico, M. Ferri, I. Stanganelli, Qualitative Asymmetry Measure for Melanoma Detection, IEEE International Symposium on Biomedical Imaging: Nano to Macro, April 2004, vol. 2, 155–1158.

[11] I. Maglogiannis, D. Kosmopoulos: "A System for the Acquisition of Reproducible Digital Skin Lesion Images" Technology and Healthcare, IOS Press 11 (2003) 425–441.

[12] Gutenev A., Skladnev V.N., Varvel D., "Acquisition-time image quality control in digital dermatoscopy of skin lesions", Computerized Medical Imaging and Graphics, 25, pp. 495–499, 2001.

[13] A. Gutenev, V.N. Skladnev, D. Varvel, Acquisition-time image quality control in digital dermatoscopy of skin lesions, Computerized Medical Imaging and Graphics 25 (2001) 495–499.

[14] Jeffrey Solomon, Sara Mavinkurve, Derrick Cox, Ronald M. Summers, Computer-assisted Detection of Subcutaneous Melanomas, Acard. Radiol. 2004, June, 11 (6), 678–685.

[15] Aglaia G. Manousaki, Andreas G. Manios, Evgenia I. Tsompanaki, Androniki D. Tosca, Use of color texture in determining the nature of melanocytic skin lesions – a qualitative and quantitative approach, Computers in Biology and Medicine 36 (2006) 419–427.

[16] Tim K. Lee, M. Stella Atkins, Michael A. King, Savio Lau, David I. McLean, Counting Moles Automatically From Back Images, IEEE Transactions on Biomedical Engineering, vol. 52, no. 11, November 2005, 1966–1969.

[17] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, M. Delfino, Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis, Arch. Dermatol. 1998, 134 (12). 1563–1570.

[18] G. Betta, G. Di Leo, G. Fabbrocini, A. Paolillo, M. Scalvenzi, Automated Application of the "7-point checklist" Diagnosis Method for Skin Lesions: Estimation of Chromatic and Shape Parameters, IMTC 2005 – Instrumentation and Measurement Technology Conference, May 2005.

[19] Murali Anantha, Randy H. Moss, William V. Stoecker, Detection of pigmented network in dermatoscopy images using texture analysis, Computerized Medical Imaging and Graphics, 28 (2004) 225–234.

[20] Vincent T.Y. Ng, Benny Y.M. Fung, Tim K. Lee, Determining the asymetry of skin lesion with fuzzy borders, Computers in Biology and Medicine, 35 (2005), 103–120.

[21] S.E. Umbaugh, R.H. Moss, W.V. Stoecker, An automatic color segmentation algorithm with application to identification of skin tumor borders, Computerized Medical Imaging and Graphics, 16 (1992) 227–235.

[22] Khaled Taouil, Nadra Ben Romdhane, Automatic Segmentation and Classification of Skin Lesion Images, Distr. Frameworks for Multimedia Applications, 2006, 1–12.
[23] Costantino Grana, Giovanni Pellacani, Rita Cucchiara, Stefania Seidenari, A New Algorithm for Border Description of Polarized Light Surface Microscopic Images of Pigmented Skin Lesions, IEEE Transactions on Medical Imaging, vol. 22, no. 8, August 2003, 959–964.
[24] Giuseppe Argenziano et al., Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet, J. Am. Acad. Dermatol., Volume 48, Number 5, p. 680–693, 2003.
[25] Electronic available information at Skin Oncology Teaching Center http://www.dermoncology.com/.
[26] I. Maglogiannis, C. Caroni, S. Pavlopoulos, V. Karioti, "Utilizing Artificial Intelligence for the Characterization of Dermatological Images", 4th International Conference "Neural Networks and Expert Systems in Medicine and Healthcare", NNESMED, pp. 362–368, Greece 2001.
[27] H. Handels, Th. Roß, J. Kreusch, H.H. Wolff and S.J. Pöppl: "Feature selection for optimized skin tumor recognition using genetic algorithms" Artificial Inteligence in Medicine 16 283–297 1999.
[28] Jain A.K. "Advances in statistical pattern recognition. In: P.A. Devijer and J. Kittler Editors", Pattern recognition, theory and applications Springer, Berlin (1986).
[29] A.K. Jain and W.G. Waller "On the optimal number of features in the classification of multivariate gaussian data" Pattern Recogn. 10 (1978), pp. 365–374.
[30] Duda R.O., Hart P.E. "Pattern classification and skin analysis" New York, John Willey 1973.
[31] A. Jain, R. Dubes: "Algorithms for clustering data" Prentice Hall Englewood Cliffs NJ 1988.
[32] Burges C: A tutorial on support vector machines for pattern recognition [http://www.kernel-machines.org/].
[33] Scott E. Umbaugh, Randy H. Moss, William V. Stoecker: "Applying Artificial Intelligence to the identification of Variegated Coloring in Skin Tumors", IEEE Engineering in Medicine and Biology Magazine, December 1991, pp. 57–62.
[34] Ajaya. Durg, William V. Stoecker, John P. Vookson, Scott E Umbaugh, Randy H. Moss: "Identification of Variegated Coloring in Skin Tumors", IEEE Engineering in Medicine and Biology Magazine, September 1993, pp. 71–75.
[35] Christianini N, Shawe-Taylor J: An introduction to support vector machines. Cambridge University Press, 2000.
[36] Schölkopf B.: Statistical learning and kernel methods. [http://research.Microsoft.com/~bsc].
[37] Sachin V. Patwardhan, Atam P. Dhawan, Patricia A. Relue, Classification of melanoma using tree structured wavelet transforms, Computer Methods and Programs in Biomedicine. 72 (2003), 223–239.
[38] J.A. Nimunkar, P.A. Dhawan, P.A. Relue, S.V. Patwardhan, Wavelet and Satistical Analysis for melanoma classification, Progress in biomedical optics and imaging, 2002, 3(3), 1346–1353.
[39] M. Elbaum, A.W. Kopf, H.S. Rabinovitz, R. Langley, H. Kamino, M. Mihm, A. Sober, G. Peck, A. Bogdan, D. Krusin, M. Greenebaum, S. Keem, M. Oliviero, S. Wang, Automatic differentiation of melanoma from melanocytic nevi with multispectral digital dermoscopy: a feasibility study, J. Am. Acad. Dermatology, 2001, 44 (2), 207–218.
[40] Hagen D. Test characteristics: how good is that test? Prim Care 1995, 22, 213–33.
[41] Stephan Dreiseitl, Lucila Ohno-Machado, Harald Kittler, Stal Vinterbo, Holger Billhardt, Michael Binder, A Comparison of Machine Learning Methods for the Diagnosis of Pigmented Skin Lesions, Journal of Biomedical Informatics, 34, 2001, 28–36.
[42] S. Umbaugh, Y. Wei, M. Zuke: "Feature Extraction in Image Analysis" IEEE Engineering in Medicine and Biology pp. 62–73 Jul./Aug. 1997.
[43] M. Nischic and C. Forster: "Analysis of Skin Erythema using true color images" IEEE Transactions on Medical Imaging Vol. 16 No. 6 December 1997.
[44] L. Tsap, D. Goldgof, S. Sarkar, P. Powers: "Vision-based tecnique for objective assesment of burn scars" IEEE Transactions on Medical Imaging vol. 17, pp. 620–633 1998.
[45] S. Tomatis, C Bartol, G. Tragni, B. Farina, R. Marchesini: "Image analysis in the RGB and HS colour planes for a computer assisted diagnosis of cutaneous pigmented lesions" Tumori vol. 84 pp. 29–32 1998.
[46] J. Sanders B. Goldstein, D. Leotta K. Richards: "Image proccesing tecniques for quantitative analysis of skin structures" Computer Methods and Programs in Biomedicine 59 pp. 167–180 1999.
[47] G. Hansen, E. Sparrow, J. Kokate, K. Leland, P. Iaizzo: "Wound Status Evaluation Using Color Image Processing" IEEE Transactions on Medical Imaging, vol. 16, no. 1 pp. 78–86 Feb. 1997.
[48] M. Herbin, F. Bon, A. Venot, F. Jeanlouis, M. Dubertret, L. Dubertret, G. Strauch: "Assessment of Healing Kinetics Through True Color Image Processing" IEEE Transactions on Medical Imaging, vol 12, no. 1 pp. 39–43 Mar. 1993.
[49] S. Chin: "The assessment of methods of measurements" Stat. Med. Vol. 9 pp. 351–362, 1990.
[50] W. Lohman, E. Paul: "In situ detection of melanomas by fluorescence measurements" Naturewissenschaften 1988, 75 201–202.

[51] Bono, S. Tomatis, C. Bartoli: "The invisible colors of melanoma. A telespectrophotometric diagnostic approach on pigmented skin lesions" European Journal of Cancer 1996 32A, 727–729.
[52] B.W. Chwirot, S. Chwirot, J. Redziski and Z. Michniewicz: "Detection of melanomas by digital imaging of spectrally resolved ultraviolet light-induced autofluorescence of human skin" European Journal of Cancer October 1998 34, 1730–1734.
[53] I. Maglogiannis, E. Zafiropoulos: "Utilizing Support Vector Machines For The Characterization of Digital Medical Images" BMC Medical Informatics and Decision Making 2004, 4:4.
[54] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, H. Kittler, "Automated melanoma recognition", IEEE Transactions on Medical Imaging, 20 (3), Mar. 2001 233–239.
[55] A. Bono, S. Tomatis, C. Bartoli, G. Tragni, G. Radaelli, A. Maurichi, and R. Marchesini. The ABCD system of melanoma detection: A spectrophotometric analysis of the asymmetry, border, color, and dimension. Cancer 85(1), 1999, p. 72–77.
[56] Zhao Zhang, Randy H. Moss, William V. Stoecker, Neural Networks Skin Tumor Diagnostic System, IEEE Int. Conf. Neural Networks & Signal Processing, 2003.
[57] Hisao Motoyama, Toshiyuki Tanaka, Masaru Tanaka, Hiroshi Oka, Feature of Malignant Melanoma based on Color Information, SICE Annual Conference in Sapporo, August 2004.
[58] J. Boldrick, C. Layton, J. Ngyuen, S. Swtter, Evaluation of digital dermoscopy in a pigmented lesion clinic: Clinician versus computer assessment of malignancy risk, Journal of the American Academy of Dermatology, vol. 56, no. 3, 417–421.
[59] R. Joe Stanley, Randy Hays Moss, William Van Stoecker, Chetna Aggarwal, A fuzzy-based histogram analysis technique for skin lesion discrimination in dermatology clinical images, Computerized Medical Imaging and Graphics, 27 (2003), 387–396.
[60] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Chapman & Hall, New York, London, 1993.
[61] M. Wiltgen, A. Gerger, J. Smolle, Tissue counter analysis of benign common nevi and malignant melanoma, International Journal of Medical Informatics, 69 (2003) 17–28.
[62] Lefevre E., Colot O., Vannoorenberghe P., de Brucq D., Knowledge modeling methods in the framework of evidence theory: an experimental comparison for melanoma detection, 2000 IEEE International Conference on Systems, Man and Cybernetics, vol. 4, 2806–2811.
[63] Kwasnicka H., Paradowski M., Melanocytic lesion images segmentation enforcing by spatial relations based declarative knowledge, Proc. of 5th International Conference on Intelligent Systems Design and Applications, 2005, 286–291.
[64] Xiaojing Yuan, Zhenyu Yang, George Zouridakis, Nizar Mullani, SVM-based Texture Classification and Application to Early Melanoma Detection, Proc. of the 28th IEEE EMBS Annual International Conference, New York City, USA, Aug. 30–Sept. 3, 2006.
[65] The Melanoma Book by Howard L. Kaufman, M.D., FACS.
[66] G. Argenziano, Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet, Journal of the American Academy of Dermatology, vol. 48 (5), 679–693.
[67] Andrea Sboner, Claudio Eccher, Enrico Blanzieri, Paolo Bauer, Mario Cristofolini, Giuseppe Zumiani, Stefano Forti, A multiple classifier system for early melanoma diagnosis, Artificial Intelligence in Medicine, 27 (2003), 29–44.
[68] Ian H. Witten, Eibe Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd Ed., 2005, Elsevier.
[69] K. Taouil, B. Romdhane, M. S. Bouhlel, A New Automatic Approach for Edge Detection of Skin Lesion Image, Information and Communication Technologies, ICTTA 2006, 212–220.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Fuzzy Systems in Biomedicine

Georgios DOUNIAS[1]

University of the Aegean, Business School,
Department of Financial and Management Engineering,
31 Fostini Street, 82100 Chios, Greece
Phone: +30-22710-35454, Fax: +30-22710-35409
e-mail: g.dounias@aegean.gr

Abstract: The chapter presents recent advances of fuzzy systems in biomedicine. A short introduction is made on the main concepts of fuzzy sets theory. Then, a survey of recent research reports (2000 and beyond) is performed, in order to map existing theoretical trends in fuzzy systems in biomedicine, as well as important real-world biomedical applications using fuzzy sets theory. The surveyed research reports are divided into different categories either (a) according to the medical practice (diagnosis, therapy and imaging - including signal processing) or (b) according to the kind of problem faced (device control, biological control, classification and pattern analysis, and prediction-association). Recently emerging biological topics related to gene expression data, molecular - cellular analysis and bioinformatics, using fuzzy sets theory, are also reported in the chapter.

Keywords: Fuzzy systems, biomedicine, survey

1. Introduction and basic concepts

Biomedicine is the branch of medical science that applies biological and physiological principles to clinical practice[2]. Another definition correlates biomedicine with the branch of medical science that deals with the ability of humans to tolerate environmental stresses and variations, as in space travel. Also biomedicine is sometimes defined as the application of the principles of the natural sciences, especially biology and physiology, to clinical medicine[3]. For the National Cancer Institute, US National Institute of Health[4], biomedicine (also called mainstream medicine) corresponds to a system in which medical doctors and other healthcare professionals (such as nurses and therapists) treat symptoms and diseases using drugs, radiation, or surgery. Similarly, by the term biomedical engineering we refer to a variety of tools of the physical sciences which are implemented and used in order to advance and understand problems in the biological and medical sciences.

---

[1] Corresponding author: Georgios Dounias, Associate Professor, Department of Financial and Management Engineering, Business School, University of the Aegean, 31 Fostini Street, 82100 Chios, Greece, fax: +30-22710-35409, e-mail: g.dounias@aegean.gr
[2] www.thefreedictionary.com/biomedicine
[3] *The American Heritage Dictionary of the English Language*, 4th Edition, 2000, Houghton Miffin Company Publ.
[4] www.cancer.gov

Fuzzy logic deals with fuzzy sets and logical connectives for modelling the human-like reasoning problems of the real world. A fuzzy set, unlike conventional sets, includes all elements of the universal set of the domain, but with varying membership values in the interval [0,1]. The logic of fuzzy sets was introduced in the systems theory by L.A. Zadeh and later was extended for approximate reasoning in expert systems.[5] Fuzzy systems are synonymous with the effective modelling and handling of vagueness, uncertainty and imprecision contained in expert tasks related to judgement, choice and decision of everyday practice[6]. The difference between probability theory and fuzzy sets theory is that the first one is concerned with occurrence of well-defined events -subjective or objective ones- while the latter deals with graduality of concepts and describes their boundaries, having nothing to do with frequencies and repetitions of an event.[7]

Fuzzy sets theory is particularly capable of mapping properly the experience of specialised staff, such as medical doctors. Thus, fuzzy sets are quite popular for their successful application in various biomedical problems. Decision variables are modelled as fuzzy sets corresponding to membership functions defined by domain experts, while readable decision rules following the practice and principles of formal mathematical logic can represent higher forms of expert knowledge in the so-called fuzzy rule based systems, an advance of the well-known expert systems. Fuzzy sets theory, when compared to other competitive intelligent tools and techniques, is particularly characterized from simplicity in use and concept representation, lacking accurate mathematical models during the formulation of a problem solving strategy or methodology. Referring to fuzzy sets and systems, we generally mean (a) either fuzzy logic principles and theoretical advances, or (b) applied fuzzy rule-based systems by which we mean decision making tools based in fuzzy logic principles for representing expert knowledge in order to effectively handle real world decision problems, or finally (c) hybrid intelligent models combining in a superior way, one fuzzy decision component with one or more other components –intelligent or not- for efficient problem solving performance in complex domains.

The beginning of applying intelligent techniques in biomedicine was closely related to image analysis and signal processing. Fuzzy systems were used for effective medical image processing, monitoring and control of time-varying biomedical processes. Recent literature contains a variety of real-world examples for different implementations and applications of nonlinear signal processing technologies to biomedical problems. Also, structures of Kohonen, Hopfield, and multiple-layer "designer" networks with other approaches to produce hybrid systems can be found. Even newer approaches include uncertainty management, analysis of biomedical signals, application of algorithms to EEG and heart rate variability signals, event detection and sample stratification in genomic sequences, applications of multivariate analysis methods to measure glucose concentration, etc. In addition, the reader can search for emerging applications in data mining, biomedical textual mining, and knowledge discovery research, including topics such as semantic parsing and analysis for patient records, biological relationships, gene pathways, and metabolic networks, exploratory genomic data analysis, joint learning using data and text mining, and disease informatics and outbreak detection. Similar approaches propose the design of

[5] Konar A., *AI and Soft Computing*, CRC Pr., 2000, p.14

[6] Dubois D. and Prade H., The three semantics of fuzzy sets, *Fuzzy Sets and Systems* 90(2), 1 September 1997, pp. 141-150

[7] Chen Z., *Computational Intelligence for Decision Support*, CRC Pr., 1999, p. 314

multivariate rule bases through self-learning by mapping fuzzy systems onto neural network structures. A number of applications also concern the bridging of the gap between low-level sensor measurements and intermediate or high-level data representations.

Regarding the application of fuzzy sets and systems in biomedicine, fuzzy control is considered exceptionally practical and cost-effective due to its unique ability to accomplish tasks without knowing the mathematical model of the system, even if it is nonlinear, time varying and complex. Nevertheless, compared with the conventional control technology, most fuzzy control applications are developed in an ad-hoc manner with little analytical understanding and without rigorous system analysis and design. Usual topics cover (a) structures of fuzzy controllers/models with respect to conventional fuzzy controllers/models, (b) analysis of fuzzy control and modelling in relation to their classical counterparts, (c) stability analysis of fuzzy systems and design of fuzzy control systems, (d) sufficient and necessary conditions on fuzzy systems as universal approximators, (e) real-time fuzzy control systems for treatment of life-critical problems in biomedicine.

Bioinformatics has emerged as an interdisciplinary research area that brings together experimental and computational approaches to biology and biomedicine and aims at integrating various sources of information to perform inference tasks on the data. Such fields of research interest within bioinformatics include (a) interpretation of complex data generated by analytic processes in molecular genetics and cell biology, (b) processing of micro-array data, (c) analysis of gene-expression data, (d) applications of intelligent technologies in molecular biology and molecular genetics, (e) management of medical data including bioinformatics databases and distributed computing, (f) modelling and simulation of cellular systems, (g) biomedical computer vision using geometrical and dynamical models, (h) microarray image processing, (i) intelligent optimization for computational chemistry and molecular biology problems, (j) applications of intelligent technologies in proteomics, (k) visualization of bioinformatics data, etc.

Some indicative topics of medical interest related to bioinformatics include:

- Intelligent technologies for tumor classification, gene function analysis and prediction, protein modelling and prediction, pathway analysis.
- Intelligent technologies for complex clinical data analysis, processing and visualization.
- Intelligent technologies for medical and biomedical information extraction, knowledge discovery and management.
- Intelligent retrieval and integration of biological and medical information.
- Intelligent technologies for processing, analysis and interpretation of medical and bioinformatics images.

2. Recent Research Advances of Fuzzy Systems in Biomedicine

In a recent general search in PubMed / Medline around fuzzy systems and their application to medicine, biomedicine and bioinformatics [1], were found more than 600 related research reports in journals, books and conferences within the period 2000-2006. This makes obvious that the application of fuzzy sets in biomedicine constitutes a rather popular approach for researchers. The trend is slightly increasing

over time for the number of related publications found annually in bibliographic resources.
In this chapter, attempting to describe the current trend in research around fuzzy systems in biomedicine, a brief bibliographic survey was conducted, for research reports published from the year 2000 and beyond. Search was performed in PubMed / Medline[8] and ISI / Web of Science[9], only by using the combination of the words:

{fuzzy} AND [{biomedical} OR {biomedicine}]

More than 240 papers were initially found with this search, later reduced to 93 after a more careful additional "filtering" of their content. The main idea of these papers is briefly presented in the following chapters, together with a rough grouping of them into major categories, in respect to their main advantage and the special function that fuzzy systems perform in specific biomedical applications.
Related surveys and special issues or special session proceedings organised around the use of intelligent methodologies (including approaches related to fuzzy sets and systems) in medicine, biomedicine and bioinformatics can be found in [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12] and [13]. Indicative book publications containing methodological aspects of fuzzy systems and their application to medicine, biomedicine and bioengineering are [14], [15], [16], [17], [18].
From the medical viewpoint, a rough division of research reports appeared in biomedical literature draws three main areas of medical interest, (a) diagnosis, (b) therapy, and (c) imaging and signal processing. According to this analysis, most papers (47 out of 93, i.e. about half of them) primarily attempt to perform medical or biomedical diagnosis. Then, a considerable part of the related research (36 out of 93, or approximately 39% of the total) deals primarily with image analysis and biomedical image processing issues. Finally, only a small part of the reviewed papers (12 out of 93, or about 15%) concerns issues mainly dealing with medical therapy.
Another categorization of the reviewed papers, related to the methodological type of problem under examination, could contain (a) device control, (b) biological control, (c) classification and pattern recognition and (d) prediction and association. Only five (5) papers deal clearly with control of devices, whereas four (4) papers deal primarily with biological control and eight (8) are mostly related to prediction and association. The vast majority (almost 75%) of the research reviewed is closely related to classification and pattern recognition. Below we briefly describe some of the approaches found, according to the abovementioned categorization. Note that most of the papers might be classified in more than one of the abovementioned categories. In that sense below is only given an indicative list of research reports for each category, in order to picture out the current trends in each biomedical domain.

3. Diagnosis

As it was mentioned already, most fuzzy systems are implemented around problems related to diagnosis. From the papers that are presented below, nine (9) approaches mostly use basic principles of fuzzy sets and systems for representing human expertise (e.g. variable representation in membership functions), while five (5) papers implement more sophisticated fuzzy rule-based systems and ten (10) of them combine

[8] http://medline.cos.com/
[9] http://wos.ekt.gr

fuzzy approaches with other intelligent approaches in a hybrid computational intelligence scheme. This last observation shows the current trend in fuzzy systems in biomedicine, which is to combine fuzzy approaches with genetic algorithms or neural networks in an attempt to (a) improve classification accuracy, (b) speed up optimization tasks, or (c) produce more meaningful outcomes during knowledge discovery processes. The applications related to diagnosis that were found in this paper, can be classified as follows:

a. Nuclear medicine: bone scintigraphy for tumor identification
b. Pneumonology: sleep apnea diagnosis
c. Cardiology: (I) identification of left ventricle dysfunctions, (II) identification of systolic dysfunction, (III) EEG analysis, (IV) ventricular premature contraction in Holter systems
d. Neurology: (I) aphasia diagnosis, (II) epilepsy risk identification in diabetic neuropathy, (III) diagnosis of multiple sclerosis, (IV) capturing of epileptic EEG spikes, (V) analysis of epileptic crises
e. Radiology: (I) MR imaging for identification of carotid atherosclerosis, (II) identification and analysis of mammography microcalcifications (III) identification of tumors in breast sonograms (IV) general US image based analysis (V) brain analysis for sulcal landmarks identification
f. Bio-pathology / Biochemistry: (I) analysis of the metabolic profiling of urine, (II) continuous biomedical parameters monitoring, (III) blood cell analysis through image segmentation, (IV) study of intracellular kinetics of thiamine during intestine cell analysis.
g. Paediatrics: (I) respiratory analysis in neonates for identification of problems, (II) Doppler ultrasound of neonatal cerebral hemodynamics
h. Orthopaedics: EMG-based analysis of multifunctional prosthesis systems

Moving to a more detailed description of the abovementioned applications of fuzzy systems in biomedicine, the work in [19] deals with bone scintigraphy, which is an effective method to diagnose bone diseases such as bone tumors. A fuzzy system called characteristic-point-based fuzzy inference system is employed to implement the diagnosis system. The resulting computer-aided diagnosis system is of a small-sized rule base such that the resulting fuzzy rules can be not only easily understood by radiologists, but also matched to and compared with their expert knowledge. In [20] are reported the results of the clinical evaluation of a fuzzy system suitable for detection and classification of sleep apnea syndromes. The system uses breathing signals: nasal flow, thorax movement, and abdomen movement. In [21] is reported an adaptive diagnostic system for the classification of breathing events for the purpose of detecting sleep apnea syndromes. The system employs two classification engines used in series, one fuzzy logic-based and another based on a centre of gravity engine, designed to work in collaboration to the first one for sorting out the not-sure events.

The work in [22] presents a new technique for identification of regional dysfunctions in the left ventricle from 2-D echocardiography. The work uses a novel left ventricular border tracking algorithm based on fuzzy inference system. The main advantage of this proposed approach is the smaller date handling in regional dysfunction identifications unlike other existing methods. In [23] the authors deal with aphasia diagnosis, a particularly challenging medical diagnostic task due to the linguistic uncertainty and vagueness, inconsistencies in the definition of aphasic syndromes, large number of measurements with imprecision, natural diversity and subjectivity in test objects as well as in opinions of experts who diagnose the disease.

A hierarchical fuzzy rule-based structure is proposed here that considers the effect of different features of aphasia by statistical analysis in its construction. Also, in [24] emphasis is placed on a new VLSI design for fuzzy processor in a biomedical application. The parallel computing architecture incorporated in the design decides epilepsy risk level in diabetic neuropathy.

In [25] methods are developed to characterize atherosclerotic disease in human carotid arteries using multiple MR images having different contrast mechanisms (T1W, T2W, PDW). To enable the use of voxel gray values for interpretation of disease, they have created a new method, local entropy minimization with a bicubic spline model (LEMS), to correct the severe (approximately 80%) intensity inhomogeneity that arises from the surface coil array. LEMS is compared to a modified fuzzy c-means segmentation based method and a linear filtering method. The work in [26] presents the development of a general and fast method for metabolic profiling of urine, using capillary electrophoresis-electrospray ionisation mass spectrometry and multivariate data analysis. Analysis of the entire resulting data set, with no prior knowledge of the target compounds, using pair-wise 'fuzzy' correlation and eigenvalue analysis enable the samples to be discriminated between on the basis of blank urine and urine collected after drug intake. In [27] is given an overview of the currently available literature on characterization of malignant and benign microcalcifications in mammography. The work compares and evaluates some of the classification algorithms on microcalcifications in mammograms used in various computer aided design systems, which are separated into categories according to the method in use.

The authors in [28] focus on the 2D echocardiograms of the left ventricle. After pre-processing of related images, segmentation of the left ventricle is performed by fuzzy systems. Then the volumes are measured by single and biplane methods along with the perimeter, short axis length and long axis length in each frame, from which the two indices Sphericity Index and Normalized Eccentricity Index are determined. It has been found that the diastolic phase is short in the case of systolic dysfunction, and its volume variation is not uniform as in the normal case. In [29] the authors aim to recover transient trial-varying evoked potentials, in particular the movement-related potentials embedded within the background cerebral activity at very low signal-to-noise ratios. A new adaptive neuro-fuzzy technique attempts to estimate movement-related potentials within multi-channel EEG recordings, enabling this method to completely adapt to each input sweep without system training procedures. The proposed framework is tested with simulations to validate the analytical results before applying them to the real biological data. The method is likely to complement other similar estimation techniques. Also, in [30] is proposed a novel method for detecting ventricular premature contraction (VPC) from the Holter system using wavelet transform and fuzzy neural network.

In [31] the authors evaluate various spiculate and jagged margin shape features, known to be malignant characteristics in breast sonograms. To determine the boundary of lesions, Markov random field segmentation is used. The goal is the classification of benign and malignant lesions on the breast sonogram. Our algorithm consists of two steps, segmentation and classification. In the first step, a breast sonogram is segmented using low resolution and Gaussian-Markov random field. The fuzzy clustering method algorithm is then applied to the preprocessed image to initialize the segmentation. To discriminate benign and malignant tumors three types of lesion characteristics are investigated: jag count, compactness, and acutance.

The work in [32] deals with pulse oximetry which has become an essential technology in respiratory monitoring of neonates and paediatric patients, still fraught with artefacts causing false alarms resulting from patient or probe movement. A technique is developed for classifying plethysmogram pulses into two categories, valid and artefact, via implementations of fuzzy inference systems, which were tuned using an adaptive-network-based fuzzy inference system (ANFIS) and receiver operating characteristics curves analysis. The paper in [33] proposes compensatory fuzzy neural networks (CFNN) without normalization, which can be trained with a backpropagation learning algorithm, as a pattern recognition technique for intelligent detection of Doppler ultrasound waveforms of abnormal neonatal cerebral hemodynamics. Doppler ultrasound signals are recorded from the anterior cerebral arteries of normal full-term babies and mature babies with intracranial pathology. The features of normal and abnormal groups as inputs to pattern recognition algorithms are extracted from the maximum velocity waveforms by using principal component analysis. The proposed technique is compared with the CFNN with normalization and other pattern recognition techniques applied to Doppler ultrasound signals from various arteries.

Research reported in [34], deals with continuous biomedical parameters corresponding to normally hybrid signals, because they contain both sub-symbolic and symbolic information. The paper describes a methodology to design adequate processing systems for the automatic analysis of this kind of signals, supported by a fuzzy system. In [35] the work deals with a method for the identification of the dynamics of nonlinear (patho-) physiological systems by learning from data. The key idea consists in the integration of qualitative modelling methods with fuzzy logic systems. The major advantage is its capability both to represent the structural knowledge of the system at study and to determine, by exploiting the available experimental data, a functional approximation of the system dynamics that can be used as a reasonable predictor of the patient's future state. The method has been successfully applied in the identification of the intracellular kinetics of thiamine from data collected in the intestine cells.

In [36] the authors describe the application of a novel unsupervised pattern recognition system to the classification of the Visual Evoked Potentials (VEP's) of normal and multiple sclerosis patients. The method combines a traditional statistical feature extractor with a fuzzy clustering method, all implemented in a parallel neural network architecture. The clustering module uses a modification to the Fuzzy c-Means (FCM) clustering algorithms, where an optimization routine adjusts a set of cluster centers to minimize an objective error function. The paper in [37] investigates algorithms for clustering of epileptic electroencephalogram (EEG) spikes. They compare the fuzzy C-means (FCM) algorithm and a graph-theoretic algorithm, and they give criteria for determination of the correct level of outlier contamination. The performance is then studied by aid of simulations, which show good results for a range of circumstances, for both algorithms. The graph-theoretic method gave better results than FCM for simulated signals. Also, when evaluating the methods on seven real-life data sets, the graph-theoretic method was the better method, in terms of closeness to the manual assessment by a neurophysiologist.

The work in [38] presents a new algorithm for segmenting general US images that is composed of two major techniques; namely, the early-vision model and the discrete-snake model. By simulating human early vision, the early-vision model can capture both gray-scale and textural edges while the speckle noise is suppressed. By performing deformation only on the peaks of the distance map, the discrete-snake model promises better noise immunity and more accurate convergence. Moreover, the constraint for most conventional snake models that the initial contour needs to be located very close to the actual boundary has been relaxed substantially.

In [39] is introduced a new method for color blood cell image segmentation based on a Fuzzy c-means (FCM) algorithm. By transforming the original blood microscopic image to indexed image, and by doing the colormap, a fuzzy approach to obviating the direct clustering of image pixel values, the quantity of data processing and analysis is enormously compressed. The approach proposed overcomes the usual problem of difficult convergence of the FCM algorithm and thus the iteration time of iterative convergence is reduced, the execution time of algorithm is decreased, and the correct segmentation of the components of color blood cell image is implemented. Then, the work in [40] presents a heuristic fuzzy logic approach to multiple electromyogram (EMG) pattern recognition for multifunctional prosthesis control. Basic signal statistics (mean and standard deviation) are used for membership function construction, and fuzzy c-means (FCMs) data clustering is used to automate the construction of a simple amplitude-driven inference rule base. The result is a system that is transparent to, and easily "tweaked" by, the prosthetist/clinician.

The authors in [41] propose a modified parcellation method, one of several brain analysis methods, a procedure popular for subdividing the regions identified by segmentation into smaller topographically defined units. The method provides the reliable and reproducible regions of interest using successive fuzzy c-means (sFCM) and boundary-detection algorithms. The method displays simultaneously both original brain image for identifying the sulcal landmarks and its tissue-classified image for referring to patterns of sulci. The whole cerebral region is extracted by the semiautomated region growing method and then classified to gray matter, white matter, and cerebrospinal fluid by sFCM. The volume ratio of the whole cerebrum to the parceled object can be used to investigate structural abnormalities for the pathological detection of the various mental diseases such as schizophrenia, obsessive-compulsive disorder. In addition, [42] deals also with parcellation. The fuzzy clustering algorithm is mainly used to preprocess parcellation into several segmentation methods, because it is very appropriate for the characteristics of magnetic resonance imaging (MRI), such as partial volume effect and intensity inhomogeneity. However, some gray matter, such as basal ganglia and thalamus, may be misclassified into the white matter class using the conventional fuzzy C-Means (FCM) algorithm. Parcellation has been nearly achieved through manual drawing, but it is a tedious and time-consuming process. Improved classification is proposed using successive fuzzy clustering and implementing the parcellation module with a modified graphic user interface (GUI). Finally, the work in [43] presents a hybrid expert system (HES) intended to minimise some complex problems pervasive to knowledge engineering such as, the knowledge elicitation process (known as the bottleneck of expert systems) the choice of a model for knowledge representation to codify human reasoning, the number of neurons in the hidden layer and the topology used in the connectionist approach, the difficulty to extract an explanation from the network. Two algorithms are applied to developing of HES, one for the training of the fuzzy neural network and another for obtaining explanations on how the fuzzy neural network attains a conclusion. A case study is presented related to epileptic crisis.

4. Therapy

Regarding fuzzy systems related to therapy, fuzzy rule-based approaches are applied, as well as classical expert systems including fuzzy rules, reflecting medical expertise related to the therapeutic task. Regarding the applications described in this section, they can be classified as follows:

a. Endocrinology: two applications regarding glucose regulation
b. Intensive care medicine: mechanical ventilation monitoring after serious head injury
c. Orthopaedics: (I) ankle arthrodesis, (II) myoelectric prostheses functionality control
d. Cardiology: arterial pressure control

More specifically, research reported in [44] presents a detailed glucose regulation model using fuzzy inference system (FIS) descriptions of hormonal control action and the familiar Michaelis-Menten (M-M) kinetic description for glucose transport. The fuzzy M-M model is compared and contrasted with a well-known comprehensive glucose model. The two models give similar results for glucose response, endogenous glucose production, and total uptake. The fuzzy M-M model features a renal subsystem that provides 25% of the endogenous glucose production. The work demonstrates the successful application of fuzzy logic and fuzzy inference to biological modelling. In [45] the authors investigate different fuzzy logic controllers for the regulation of blood glucose level in diabetic patients. A fuzzy logic control (FLC) recently proposed for maintaining blood glucose level in diabetics within acceptable limits, was shown to be more effective with better transient characteristics than conventional techniques. In fact, FLC is based on human expertise and on desired output characteristics, and hence does not require precise mathematical models. This observation makes fuzzy rule-based technique very suitable for biomedical systems where models are, in general, either very complicated or over-simplistic. Another attractive feature of fuzzy techniques is their insensitivity to system parameter variations, as numerical values of physiological parameters are often not precise and usually vary from patient to another.

In [46] is provided automatically continuous propofol sedation for patients with severe head injury, unconsciousness, and mechanical ventilation in order to reduce the effect of agitation on intracranial pressure (ICP) using fuzzy logic control (FLC) in a neurosurgical intensive care unit (NICU). Results indicate that FLC can easily mimic the rule-base of human experts (i.e., neurosurgeons) to achieve stable sedation similar to the RBC group. Furthermore, the results also show that a self-organising FLC can provide more stable sedation of ICP pattern because it can modify the fuzzy rule-base to compensate for inter-patient variations. The work in [47] deals with Kinematic parameters for normal subjects and patients with ankle arthrodesis, grouped using the fuzzy cluster paradigm. The clinical utility of the fuzzy clustering approach is demonstrated with data for subjects with ankle arthrodesis, where changes in membership of the clusters provide an objective technique for measuring changes of gait pattern after ankle arthrodesis.

In [48] a rule-based system was designed to control the mean arterial pressure and the cardiac output of a patient with congestive heart failure, using two vasoactive drugs, sodium nitroprusside and dopamine. The controller has three different modes that engage according to the hemodynamic state. After extensive testing and tuning on a hemodynamics nonlinear model, the control system has been applied in dog experiments, which led to further enhancements. In [49] the work concerns the use of a supervisory expert system based on fuzzy logic for the parameter adjusting of myoelectric prostheses. The prosthesis system is an artificial arm, which is equipped with an on-board actuation system. The expert system guides patients through an interactive session whose aim is to test the prosthesis functionality and, when necessary, to self-adjust the parameters.

5. Imaging and signal processing

Most applications of fuzzy systems in image and signal processing, primarily concern radiology, and secondarily other medical specialties for which the analysis of signal or images of patients with the aid of a specialist in radiology is absolutely necessary. Regarding the nature of the fuzzy methods used for imaging and signal processing, the vast majority of them are modern hybrid intelligent schemes, mostly based on fuzzy C-means approaches combined to wavelet transforms, neural networks, Bayesian classifiers or other adaptive clustering techniques. There are also approaches which make use of simple principles of fuzzy logic like fuzzy logic control and fuzzy representation techniques. In some cases there exist research reports which mainly stress on the power of new methods for signal processing, based on artificial or benchmark data and simulations, rather than focusing primarily on specific biomedical applications. In this section the following medical applications of signal and image processing are briefly mentioned:

a. cerebral lobe segmentation in MR images
b. dynamic PET image segmentation,
c. automatic segmentation of MR images,
d. brain tissue analysis in MR images
e. automatic segmentation of abdominal organs
f. spectra analysis of colorectal adenocarcinoma
g. biomedical MR image analysis
h. positron emission tomography
i. gray and white matter identification
j. tongue carcinoma identification from MR images
k. analysis of cytological and histological characteristics in tumors
l. automatic segmentation of MR images
m. multicontrast MRI for coronary atherosclerotic plaque characterization
n. content-based retrieval of dynamic PET-images
o. regional and global brain volume identification from serial MR images
p. micro-tomography for measuring trabecular thickness of animal femur bones
q. frontal lobe identification
r. analysis of color images in breast biopsy
s. tumor detection
t. identification of microcalcification clusters in digitized mammograms
u. fuzzy segmentation of MR images for artifact decrease
v. discrimination between swallow acceleration signals and artefacts in imaging
w. ischemic brain injury prediction by MR imaging
x. ultrasound image segmentation (crackles and squawks in lung sound signals)

Analysing further the above categorization, the authors in [50] present a novel computer-aided system for automatically segmenting the cerebral lobes from 3T human brain MR images. Due to the fact that the anatomical definition of cerebral lobes on the cerebral cortex is somewhat vague (fuzzy) for use in automatic delineation of boundary lines, and there is no definition of cerebral lobes in the interior of the cerebrum, a new method is proposed for defining cerebral lobes on the cerebral cortex and in the interior of the cerebrum. The proposed method determines the boundaries between the lobes by deforming initial surfaces. Research described in [51] investigates an effective processing method for biomedical images based on the fuzzy C-means (FCM) algorithm and wavelet transforms.

The authors in [52] work on medical microscopy. Image analysis proves useful to pathologists as it can be applied to several problems in cancerology, like quantification of DNA content, quantification of immunostaining, nuclear mitosis counting, characterization of tumor tissue architecture etc. An automatic segmentation method combining fuzzy clustering and multiple active contour models is presented. The method is illustrated through two representative problems in cytology and histology. In [53] is proposed a VOI segmentation of dynamic PET images by utilizing both the three-dimensional (3-D) spatial and temporal domain information in a hybrid technique that integrates two independent segmentation techniques of cluster analysis and region growing. The proposed technique starts with a cluster analysis that partitions the image based on temporal similarities. The resulting temporal partitions, together with the 3-D spatial information are utilized in the region growing segmentation. The technique is compared with the k-means and fuzzy c-means cluster analysis segmentation methods. In [54] is presented an algorithm that automatically segments and classifies the brain structures in a set of magnetic resonance (MR) brain images using expert information contained in a small subset of the image set. The algorithm is intended to do the segmentation and classification tasks mimicking the way a human expert would reason. The algorithm uses a knowledge base taken from a small subset of semiautomatically classified images that is combined with a set of fuzzy indexes capturing the experience and expectation a human expert uses during recognition tasks.

In [55] is proposed an algorithm which simultaneously assigns to each element in an image a grade of membership in each one of a number of objects (which are believed to be contained in the image). Then they prove the existence of a fuzzy segmentation that is uniquely specified by a desirable mathematical property, and illustrate that on several biomedical examples a new implementation of the algorithm that produces the segmentation is approximately seven times faster than the previously used implementation. In [56] is presented a method to segment brain tissue from T1-weighted Magnetic Resonance (MR) images. A modified Bayes-Shrink method is utilized to filter the image in wavelet transform domain before segmentation and then, the fuzzy c-means clustering is applied to segment brain tissue into cerebrospinal fluid, gray matter and white matter.

The work in [57] proposes a framework combining the atlas registration and the fuzzy connectedness for the automatic segmentation of abdominal organs. The performance of the proposed method is being qualitatively validated with success on CT and MRI images with manual segmentation as ground truth. The work in [58] tackles the problem of the in situ extraction of specific geometrical primitives from a three-dimensional (3D) biomedical data set. The task involves two main problems, the segmentation of major structures and the extraction of the features of interest. In [59] are presented three different clustering algorithms, applied to assemble infrared (IR) spectral maps from IR microspectra of tissues. Using spectra from a colorectal

adenocarcinoma section, is showed how IR images can be assembled by agglomerative hierarchical clustering (Ward's technique), fuzzy C-means clustering, and k-means clustering. Among the cluster imaging methods, Ward's clustering algorithm proves to be the best method in terms of tissue structure differentiation.
The authors in [60] focus on local 2D image processing using Lukasiewicz algebra with the square root function. The methodology deals with fuzzy logic based function analysis and decomposition of Lukasiewicz networks, and prove to increase the quality of image processing as demonstrated on the biomedical MR image. The paper in [61] presents a fuzzy fusion approach for combining cell-phase identification results obtained from multiple classifiers. The proposed method has been used to combine the results from three classifiers, and the combined result is superior to any of the results obtained from a single classifier. In [62] is proposed the Positron emission tomographic map reconstruction using fuzzy-median filter. Positron emission tomography is widely used in medical physics for the reconstruction of the distribution of radionuclei molecules for analyzing regional physiological functions. A potential function is proposed, based on fuzzy-median filter for noise-free image reconstruction. The reconstruction methodology is useful for obtaining artefact-free reconstruction of biomedical specimens.
In [63] is proposed an algorithm to determine the human brain (gray matter and white matter) from computed tomography head volumes with large slice thickness is proposed based on thresholding and brain mask propagation. They combine a 2D reference image for the intensity characteristics of the original 3D data set. Secondly, the region of interest of the reference image is determined as the space enclosed by the skull. Fuzzy C-means clustering is employed to determine the threshold for head mask and the low threshold for brain segmentation. The algorithm has been validated against one non-enhanced CT and one enhanced CT volume with pathology. In [64] is proposed novel hierarchical image segmentation or the extraction of tongue carcinoma from magnetic resonance (MR) images. A genetic algorithm-induced fuzzy clustering is used for initial segmentation of MR images of head and neck. Then these segmented masses are refined to reduce the false-positives using an artificial neural network-based symmetry detection and image analysis procedure.
The authors in [65] propose a framework that combines atlas registration, fuzzy connectedness segmentation, and parametric bias field correction (PABIC) for the automatic segmentation of brain magnetic resonance imaging (MRI). Fuzzy clustering techniques are applied on the PABIC corrected MRI to get the final segmentation. Expert human intervention is avoided a fully automatic method for brain MRI segmentation is provided. The work described in [66] compares coronary atherosclerotic plaque characterization using multicontrast MRI on (a) freshly excised vessels under simulated in vivo conditions, and (b) preserved vessels. A three-dimensional spatially penalized fuzzy C-means technique was applied to classify different plaque constituents. In [67] is presented a volume of interest (VOI) based content-based retrieval of four-dimensional (three spatial and one temporal) dynamic PET images. By segmenting the images into VOIs consisting of functionally similar voxels (e.g., a tumor structure), multidimensional visual and functional features were extracted and used as region-based query features. A prototype VOI-based functional image retrieval system has been designed to demonstrate the proposed multidimensional feature extraction and retrieval.
The work in [68] proposes a temporally consistent and spatially adaptive longitudinal MR brain image segmentation algorithm, which aims at obtaining accurate

measurements of rates of change of regional and global brain volumes from serial MR images. The algorithm incorporates image-adaptive clustering, spatiotemporal smoothness constraints, and image warping to jointly segment a series of 3-D MR brain images of the same subject that might be undergoing changes due to development, aging, or disease. In [69] is described an automated pattern recognition (APR) method based on the fuzzy C-means cluster adaptive wavelet algorithm, which consists of two parts. One is a fuzzy C-means clustering (FCMC) using the features from an M-band feature extractor adopting the adaptive wavelet algorithm and the second is a Bayesian classifier using the membership matrix generated by the FCMC. A FCMC-cross-validated quadratic probability measure criterion is used under the assumption that the class probability density is equal to the value of the membership matrix.

The work in [70] measures trabecular thicknesses of femur bones in post mortem rats, using the cross-sectional images taken with the zoom-in micro-tomography technique. To compensate for the limited spatial resolution in the zoom-in micro-tomography images, the fuzzy distance transform is used, for calculation of the trabecular thickness. In [71] is described a new knowledge-based automated method designed to identify several major brain sulci and then to define the frontal lobes by using the identified sulci as landmarks. To identify brain sulci, sulcal images are generated by morphologic operations and then separated into different components based on connectivity analysis. Subsequently, the individual anatomic features are evaluated by using fuzzy membership functions. Results show relatively high accuracy of using this novel method for human frontal lobe identification and segmentation.

In [72] is developed an automated, reproducible epithelial cell nuclear segmentation method to quantify cytologic features quickly and accurately from breast biopsy. The method, based on fuzzy c-mean clustering of the hue-band of color images and the watershed transform, was applied to images from three histologic types (typical hyperplasia, atypical hyperplasia, and ductal carcinoma in situ, cribriform and solid). In [73], the basic properties and generic features of biomedical signals are examined using a wide range of examples. Algorithmic results are presented to show not only the potential performance but also the limitations of the processing resources. Signal matching, scenario recognition, and data fusion are also discussed. The authors in [74] survey fuzzy logic (FL) applications in brain researches generally are related to pattern recognition for localization in brain structures or tumor detection, image segmentation, and simulations.

In [75] is developed a parameter optimization technique for the segmentation of suspicious microcalcification clusters in digitized mammograms. A fuzzy rule-based classifier is used to segment individual microcalcifications, and then clustering analysis is applied for reducing the number of false positive clusters. For the segmentation of individual microcalcifications, the new algorithm uses a neural network with fuzzy-scaled inputs. The fuzzy-scaled inputs are created by processing the histogram features with a family of membership functions, the parameters of which are automatically extracted from the distribution of the feature values. The paper in [76] presents a novel algorithm for fuzzy segmentation of magnetic resonance imaging (MRI) data and estimation of intensity inhomogeneities using fuzzy logic. MRI intensity inhomogeneities can be attributed to imperfections in the radio-frequency coils or to problems associated with the acquisition sequences. The result is a slowly varying shading artefact over the image that can produce errors with conventional intensity-based classification. The proposed algorithm is formulated by modifying the objective function of the standard fuzzy c-means (FCM) algorithm to compensate for such inhomogeneities and to allow the labeling of a pixel (voxel) to be influenced by the labels in its immediate neighborhood. The work in [77] presents a hybrid intelligent system consisting of fuzzy logic and committee networks, which proves a reliable tool for recognition and classification of acceleration signals due to swallowing. In order to improve the reliability of the recognition or automated diagnostic systems, hybrid fuzzy logic committee neural networks were developed and the system was used for recognition of swallow acceleration signals from artefacts. Two sets of fuzzy logic-committee networks (FCN) each consisting of seven member networks are developed, trained and evaluated.

In [78] a fuzzy color segmentation approach is developed for the analysis of colonoscopic images. The segmentation is made up of two phases: segmentation through histogram space filtering and region merging using fuzzy rule-based reasoning. The first phase involves using a scale-space filter to analyze the hue, saturation, and intensity (HSI) histograms to determine the number of classes and construct a 3-D class grid. The color image is then segmented based on the class grid. In the second phase, region merging based on applying the fuzzy rule-base is employed to guide the combining process of the segmented regions. For fuzzy reasoning, three criteria are evaluated, namely, the edge strength along the boundary, color similarity, and spatial connectivity of adjoining regions.

According to [79] a major difficulty in staging and predicting ischemic brain injury by magnetic resonance (MR) imaging is the time-varying nature of the MR parameters within the ischemic lesion. A new multispectral (MS) approach is described to characterize cerebral ischemia in a time-independent fashion. MS analysis of five MR parameters (mean diffusivity, diffusion anisotropy, T2, proton density, and perfusion) has been employed to characterize the progression of ischemic lesion in the rat brain following 60 minutes of transient focal ischemia. K-means and fuzzy c-means classification methods were employed to define the acute and subacute ischemic lesion. The authors in [80] develop a multistage computer-aided diagnosis (CAD) scheme for the automated segmentation of suspicious microcalcification clusters in digital mammograms. The scheme consists of three main processing steps. First, the breast region is segmented and its high-frequency content is enhanced using unsharp masking. In the second step, individual microcalcifications are segmented using local histogram analysis on overlapping subimages. For this step, eight histogram features are extracted for each subimage and are used as input to a fuzzy rule-based classifier that identified subimages containing microcalcifications and assigned the appropriate thresholds to segment any microcalcifications within them. The final step clusters the segmented microcalcifications and extracts for each cluster the number of microcalcifications, the average distance between microcalcifications, and the average number of times pixels in the cluster. Fuzzy logic rules incorporating the cluster features have been designed to remove nonsuspicious clusters, defined as those with typically benign characteristics. In [81] is proposed a new adaptive snake model for ultrasound image segmentation. The proposed snake model is composed of three major techniques, namely, the modified trimmed mean (MTM) filtering, ramp integration and adaptive weighting parameters. With the advantages of the mean and median filters, the MTM filter is employed to alleviate the speckle interference in the segmentation process. The weak edge enhancement by ramp integration attempts to capture the slowly varying edges, which are hard to capture by conventional snake models. The adaptive weighting parameter allows weighting of each energy term to

change adaptively during the deformation process. The proposed snake model has been verified on the phantom and clinical ultrasound images.
In [82] an automated signal processing approach is presented which uses two fuzzy inference systems, operating in parallel, to perform the task of adaptive separation. The method results in an orthogonal least squares-based fuzzy filter which is applied to fine/coarse crackles and squawks, selected from three lung sound databases. In [83] an adaptive noise cancellation with neural-network-based fuzzy inference system (NNIFS) is used. The NNFIS was carefully designed to model the visually evoked potential (VEP) signal. An advantage of the method in this paper is that no reference signal is required. The NNFIS based on Takagi and Sugeno's fuzzy model has the advantage of being linear-in-parameter, which is able to closely fit any function mapping and can track the dynamic behavior of VEP in a real-time fashion.
The paper in [84] applies fuzzy logic to diffusion tensor images of the human spinal cord to discriminate between gray and white matter. The technique uses common anisotropy indices and newly developed indices based on properties of the diffusion ellipsoid. Preliminary applications to subjects with varying levels of spinal cord injury are also presented in this study. Results indicate larger contrast between gray and white matter compared to the traditional fractional anisotropy index.
In [85] the authors investigate hemispheric asymmetry using the fractal dimension (FD) of the skeletonized cerebral surface. Sixty-two T1-weighted magnetic resonance imaging volumes from normal Korean adults have been used. The skeletonization of binary volume data, which correspond to the union of the gray matter and cerebrospinal flow classified by fuzzy clustering, have been performed slice by slice in the sagittal direction, and then skeletonized slices have ben integrated into the three-dimensional (3-D) hemisphere. Finally, the FD of the 3-D skeletonized cerebral surface are calculated using the box-counting method. The FD of the skeletonized cerebral surface seems to stn as a novel measure of cerebral asymmetry.

6. Device control

Most applications of fuzzy systems related to device control, belong either to orthopaedics (kinetic control of robotic exoskeletons, motion assistance of paraplegics, etc.) or to anaesthesiology (control of the depth of anaesthesia, supply of anaesthetic and analgesic drugs, etc). The fuzzy logic based methods used for device control, are either simple or more complex and hybrid, but as expected they are typical fuzzy control approaches (neuro-fuzzy control, adaptive fuzzy control, etc).
Analysing further the above, in [86] the effect concentrations of anaesthetic and analgesic drugs are used to model the pharmacodynamic interactions of the two drugs on the cardiovascular parameters, and on the auditory evoked potentials. An adaptive network-based fuzzy inference system is used to model the different signals. A stimulus model is used to establish the effects of surgical stimulus on the cardiovascular parameters. This model is constructed into a Mamdani type of fuzzy model, using anaesthetist's knowledge described by fuzzy IF-THEN rules. Clinical data are used to construct the patient model.
The work in [87] attempts neuro-fuzzy control of a robotic exoskeleton with EMG signals. The researchers propose a robotic exoskeleton for human upper-limb motion assist, a hierarchical neuro-fuzzy controller for the robotic exoskeleton, and its adaptation method.

The authors in [88] develop a novel control system for functional electrical stimulation (FES) locomotion, which aims to generate normal locomotion for paraplegics via FES. Artificial control techniques such as neural network control, fuzzy logic, control and impedance control are incorporated to refine the control performance. A musculoskeletal model with 7 segments and 18 muscles is constructed for the simulation study.

In [89] is proposed an indoor personal rowing machine which has been modified for functional electrical stimulation assisted rowing exercise in paraplegia. To successfully perform the rowing manoeuvre, the voluntarily controlled upper body movements must be coordinated with the movements of the electrically stimulated paralyzed legs. To achieve such coordination, an automatic controller is developed that employs two levels of hierarchy. A high level finite state controller identifies the state or phase of the rowing motion and activates a low-level state-dedicated fuzzy logic controller (FLC) to deliver the electrical stimulation to the paralyzed leg muscles.

In [90] the authors deal with the non-invasive monitoring of the depth of anaesthesia (DOA). Based on adaptive network-based fuzzy inference system (ANFIS) modeling, a derived fuzzy knowledge model is proposed for quantitatively estimating the DOA and validation is performed by experiments using dogs undergoing anaesthesia with three different anaesthetic regimens (propofol, isoflurane, and halothane). By eliciting fuzzy if-then rules, the model provides a way to address the DOA estimation problem by using electroencephalogram-derived parameters. The parameters include two new measures (complexity and regularity) extracted by nonlinear quantitative analyses, as well as spectral entropy. The model demonstrates good performance in discriminating awake and asleep states for three common anaesthetic regimens.

7. Biological control

Similarly as above, the fuzzy approaches used for biological control, are again usually based in standard fuzzy control methodologies (neuro-fuzzy control or adaptive fuzzy control) but other fuzzy methodologies are applied such as fuzzy rule-based expert systems, fuzzy relation composition methods, etc.

Related applications found and presented within this survey include:

a. sound analysis of human organs with an electronic stethoscope
b. superoxide dismutase in oral diseases treatment
c. study of systolic pressure variability
d. adaptive heart rate control by motion and respiratory rate

Specifically, in [91] the effect of superoxide dismutase (SOD) in treating certain types of oral diseases is evaluated with a fuzzy relation composition method and a fuzzy integral method. Results demonstrate that SOD is effective in treating oral diseases.

The authors in [92] present a solution to reduce Ambulatory Systolic Blood Pressure (ASBP) variability during the analysis of a 24-h profile. A database has been collected, and two models linking ASBP variations with body acceleration and heart rate measurements are developed. A regression one, based on a priori knowledge of ASBP variations, and a fuzzy one, automatically built from experimental data.

The work in [93] develops a fuzzy expert system (FES) for different sounds produced by different organs in the human body. A unique electronic stethoscope has also been constructed. Data and relation between variables chosen for each organ sound are inserted in a fuzzy rule-based expert system was built. The work in [94] proposes a pacemaker algorithm which controls heart rate adaptively by motion and respiratory rate. After chronotropic assessment exercise protocol (CAEP) tests are performed to collect activity and respiratory rate signals, the intrinsic heart rate was inferred from these two signals by a neuro-fuzzy method. The neuro-fuzzy method has been applied to a real pacemaker by reduced mapping of the neuro-fuzzy look-up table.

8. Classification and pattern recognition

Classification and pattern recognition tasks performed with the aid of fuzzy systems mainly refer to hybrid intelligent methodologies such as neuro-fuzzy, fuzzy sets and wavelet transformations, genetic-fuzzy systems, combination of fuzzy principles with support vector machines, and also variations of fuzzy clustering techniques. Usually fuzzy sets principles are used for the modelling of the decision variables and the combined methodology is used for the best possible classification or discrimination of the data set under investigation. Classification and pattern recognition tasks include (a) biomedical image classification, (b) high-dimensional biomedical data classification, (c) uncertain biomedical data classification, (d) image segmentation, (e) heartbeat data classification, and (f) electrocardiographic beat recognition and classification.

In detail, the work in [95] proposes a neuro-fuzzy technique for classification of biomedical images on the basis of a combined feature vector merging colour and texture features into a single feature vector. The system uses concept based on pixel descriptors, which combines the human perception of color and texture into a single vector, for region extraction. In the proposed method efforts have been made to model the imprecision using fuzzy interpretation. In [96] is developed an efficient fuzzy wavelet packet (WP) based feature extraction method for the classification of high-dimensional biomedical data such as magnetic resonance spectra. The key design phases involve wavelet transformations, feature extraction via fuzzy clustering and finally, signal classification. Application is made in magnetic resonance spectra data.

In [97] is presented a genetic fuzzy feature transformation method for support vector machines (SVMs) for performing more accurate data classification. Genetic algorithms are used to optimize the fuzzy feature transformation so as to use the newly generated features to help SVMs do more accurate biomedical data classification under uncertainty. The authors in [98] claim that SVMs are suitable for classifying biological and biomedical data, which characteristically have a few data samples with relatively large number of input features. A multi-SVM fuzzy classification and fusion model (MSFCF) is proposed to combine multiple SVMs to enhance the generalization ability of SVMs. The approach constructs a fuzzy fusion system to combine output values from several SVMs in respect to the accuracy information of each SVM. The final decision is determined based on all SVMs.

The paper given in [99] describes the overall role of soft computing in signal processing and pattern recognition with specific applications to biomedical engineering, image processing and other engineering applications. The application of fuzzy logic to image segmentation is presented in the work. The work in [100] proposes a heartbeat classification algorithm based on linear discriminant analysis and artificial neural network. To evaluate the performance of the proposed algorithm, the result of the proposed algorithm is compared with that of a fuzzy inference system classifier. The authors in [101] present and evaluate an adaptive fuzzy k-nearest neighbour classifier for EMG signal decomposition. The developed classifier uses an adaptive assertion-based classification approach for setting a minimum classification threshold. The similarity criterion used for grouping motor unit potentials (MUPs) is based on a combination of MUP shapes and two modes of use of motor unit firing pattern information: passive and active. The performance of the developed classifier is evaluated using synthetic signals with specific properties and experimental signals and compared with the performance of an adaptive template matching classifier.

In [102] is presented the application of the fuzzy neural network for electrocardiographic (ECG) beat recognition and classification. The new classification algorithm of the ECG beats, applying the fuzzy hybrid neural network and the features drawn from the higher order statistics has been proposed in the paper. The cumulants of the second, third, and fourth orders have been used for the feature selection. The hybrid fuzzy neural network applied in the solution consists of the fuzzy self-organizing sub-network connected in cascade with the multilayer perceptron, working as the final classifier. The c-means and Gustafson-Kessel algorithms for the self-organization of the neural network have been applied.

9. Prediction and association

Various approaches are used for prediction and association tasks, from simple or adaptive neuro-fuzzy schemes, to combinations of fuzzy principles and methodological components with other data analysis tools and techniques, such as wavelets, fuzzy C-means, Markov models, k-nearest neighbour, feed forward neural networks, Kohonen networks, etc.

The applications included in this section are related to surgery, cardiology, gynaecology / urology, and rehabilitation medicine. Specifically there are references to:

a. prognosis tasks in rehabilitation
b. advanced statistical inference support in semen studies of infertile men
c. tonsillectomy and adenoidectomy with excessive bleeding risks
d. prognosis of heart valve diseases
e. design and application of wearable intelligent agents / assistants
f. breast and prostate data analysis
g. prostate prognosis from I-PSS data
h. arrhythmia detection from cardiac rhythm related data

Analysing further the above, the paper in [103] presents an input classification scheme used in an evidence-based dynamic recurrent neuro-fuzzy system for prognosis in rehabilitation. All external variables which may have an effect on the outcome of the rehabilitative process are classified into facts, contexts and interventions. Their effects on patients' physical and/or physiological states are represented by fuzzy rules and/or non-linear models of physiologic processes. In [104] is presented how fuzzy logic can be used as an alternative or supplement to statistics in biomedical analysis. It shows an adaptive neuro-fuzzy inference computing in comparison with linear and curvilinear regression. The application domain is the semen of infertile man, with the independent variable concentration of spermatozoa and the dependent variable number of spermatozoa. In [105], a fuzzy set theoretic methodology is described that serves as a classification pre-processing strategy for supervised feed-forward neural networks. The proposed "fuzzy interquartile encoding methodology", determines the respective degrees to which a feature belongs to a collection of fuzzy sets that overlap at the respective quartile boundaries of the feature. The methodology is applied to two biomedical data sets relating to tonsillectomy and/or adenoidectomy patients who may or may not have had a predisposition to excessive bleeding during their operation.

In [106] a biomedical system is proposed, based on hidden Markov model for diagnosis of the heart valve diseases. First, feature extraction processing takes place by using the Doppler Ultrasound. Then wavelet entropy is applied to these features and in the classification stage, a hidden Markov model (HMM) is used. Comparisons are made with fuzzy C-means (FCM)/K-means algorithms. The authors in [107] sought to quantify the left ventricle systolic dysfunction by a geometric index from two-dimensional (2D) echocardiography by implementing an automated fuzzy logic edge detection algorithm for the segmentation. 2D echocardiogram and M-mode recordings are performed over the control group and those with the dysfunctions. From 2D recordings, individual frames are extracted for at least five cardiac cycles and then segmentation of left ventricle is done by automated fuzzy systems.

In [108] is presented a dynamic recurrent neuro-fuzzy system that implements expert- and evidence-based reasoning. It is intended to provide context-awareness for wearable intelligent agents/assistants (WIAs). A neuro-fuzzy modelling framework is developed for estimating rehabilitative change that can be applied in any field of rehabilitation if sufficient evidence and/or expert knowledge are available. It is intended to provide context-awareness of changing status through state estimation, which is critical information for WIA's to be effective. In [109] is investigated the fuzzy k-nearest neighbor (FK-NN) classifier as a fuzzy logic method that provides a certainty degree for prognostic decision and assessment of the markers, and to compare it with logistic regression as a statistical method and also with multilayer feed forward back propagation neural networks. In order to achieve this aim, breast and prostate cancer data sets are considered as benchmarks for this analysis.

The work in [110] deals with the International Prostate Symptom Score (I-PSS) which is used exclusively for evaluating patients with a prostate condition and following various treatment modalities. The paper suggests the application of an artificial neural network model to assess patients with lower urinary tract symptoms. The authors in [111] propose the notion of short-time multifractality and use it to develop a novel approach for arrhythmia detection. Cardiac rhythms are characterized by short-time generalized dimensions (STGDs), and different kinds of arrhythmias are discriminated using a neural network. To advance the accuracy of classification, a new fuzzy Kohonen network is presented.

10. Bioinformatics

Recent advances found in literature with application of fuzzy systems in bioinformatics, refer among others to (a) comparison of protein structures, (b) interpretation of gene-expression data classification outcomes and (c) clinical genomics and patient protein activity studies. Basic fuzzy sets and systems principles are used from methodological viewpoint, such as fuzzy thresholds for membership function construction, or simple rule-based classifiers.

To become more specific on the above, the work in [112] deals with the comparison of protein structures which corresponds to an important problem in bioinformatics. The work proposes a generalization of the maximum contact map overlap problem (MAX-CMO) by means of fuzzy sets and systems. A contact map is defined by means of one (or more) fuzzy thresholds and one (or more) membership functions. In addition it is demonstrated how a fuzzy sets-based metaheuristic can be used to compute protein similarities based on the new model. In [113] the authors deal with the interpretation of classification models derived from gene-expression data, which usually corresponds to a difficult task. The performance of small rule-based classifiers based on fuzzy logic is investigated in five datasets that are different in size, laboratory origin and biomedical domain. In [114] is described a complex intelligent methodology involving also pieces of fuzzy sets theory and apply it on a clinical-genomic example as well as in real, extensive patient record data and on molecular design data originally studied in part to test the ability to deduce the effects of simple natural patient sequence variations ("SNPs") on patient protein activity.
It should be noted that the reference to bioinformatics applications is only indicative in this paper, as a variety of similar works and advances (paper ,books, conferences) can be found on the application of different intelligent techniques in bioinformatics, including fuzzy systems usually as parts of hybrid intelligent structures, see for example [1], [12], [13].

11. Conclusions and Further Directions

Fuzzy systems have become attractive and popular to medical diagnosis, therapy and imaging, due to their simplicity in meaning, ease in application and effectiveness in use by domain experts. The most attractive ideas around the use of fuzzy systems are (a) fuzzy rule based systems representing medical expertise, (b) neuro-fuzzy systems - the most popular type of hybrid intelligent systems found in literature- for adaptive control problems of dynamic processes and (c) simple properties of fuzzy logic theory for determining areas, borders or findings in image analysis and signal processing applications. In respect to biomedical applications, almost all areas make use of fuzzy systems, from cardiology, haematology, surgery, radiology, oncology and so on, to bioinformatics, medical device control, patient monitoring and forecasting of evolving malfunctions. According to the analysis performed within this chapter, the most interesting conclusions drawn are the following:

a. Fuzzy systems seem to be generally one of the methods of choice when (a) control tasks are to be performed, or (b) a combination of intelligent techniques for complex data analysis needs to take place. When data are numerical, fuzzy sets are more suitable for variable definition and representation, while other intelligent algorithmic applications such as neural networks or genetic programming seem to be more capable for classification tasks and for building decision models. Finally, when there is a need to incorporate human expertise or to interpret a diagnostic model and discover meaningful knowledge from data, then fuzzy systems should be the first option.
b. In medical diagnosis tasks, neurology, radiology, biopathology and cardiology seem to be in the front line of interest for researchers. The

majority of the approaches currently used, combine fuzzy approaches with other intelligent techniques.

c. In tasks related to medical therapy, simpler methodological approaches seem to work adequately, like fuzzy rule-based approaches and classical expert systems including fuzzy rules, for their ease to represent medical expertise. Applications in diabetology and orthopaedics seem to be more popular in therapeutic tasks.
d. In image and signal processing tasks, all applications primarily concern radiology, and secondarily other medical specialties for which the signal or images analysis is necessary. Modern hybrid intelligent schemes, mostly based on fuzzy C-means are used in literature. Similar conclusions exist both for (a) prediction and association tasks, and (b) classification and pattern recognition studies, regarding the methodologies applied. No special focus on a specific medical area exists for these two domains.
e. In device control and biological control applications, simple fuzzy control applications are usually the methods of choice. Orthopaedics, anaesthesiology, cardiology and wearable medical devices seem to be in the prime line of interest for these areas.

Future trends in the area might include (a) the implementation of more advanced techniques for automated diagnosis using radiological devices related to MR spectroscopy, PET images, Doppler, etc., by embedding to them intelligent components for automated signal processing and image analysis, as well as (b) the increase of applications related to intelligent wearable technologies for routine patient monitoring. At the methodological level, type-II fuzzy sets, second order neural networks, neural ensemble technologies and finally nature-inspired intelligence might come into play as parts of hybrid intelligent methodological schemes for complex biomedical tasks.

References

1. Magoulas G, and Dounias G. (Eds.), (2007), Special Issue: Computational Intelligence in Medicine and Biology. *Applied Intelligence: The International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies*, Springer, forthcoming, Summer 2007.
2. Luo ZP, An KN, Fuzzy systems in biomedical science, *INT J GEN SYST* 30 (2): 209-217 2001
3. Mohr MT, Redl H. Data analysis now and then: significant changes in approaches and results. *Stud Health Technol Inform*. 2003; 97:73-77.
4. Hanai T, Honda H. Application of knowledge information processing methods to biochemical engineering, biomedical and bioinformatics fields. *Adv Biochem Eng Biotechnol*. 2004;91:51-73
5. Dounias G., Linkens D.A. (Eds), (2001), *Adaptive Systems & Hybrid Computational Intelligence in Medicine*, Joint Public. of the Univ. of the Aegean and EUNITE, The European Network on Intelligent Technologies for Smart Adaptive Systems, ISBN: 960-7475-19-4
6. Abbod M. F., Linkens D.A., Mahfouf M. and Dounias G., (2002), Survey on the use of Smart and Adaptive Engineering Systems in Medicine, *Artificial Intelligence in Medicine*, Vol. 26, pp. 179-209.
7. Dounias G., Roudsari A.V. (Eds), (2002), *Intelligent E-Health Applications in Medicine*, Joint Publication of the Univ. of the Aegean and EUNITE, The European Network on Intelligent Technologies for Smart Adaptive Systems, ISBN: 960-7475-20-8
8. Dounias G. (Ed.), (2003), *Hybrid and Adaptive Computational Intelligence in Medicine and Bio-informatics*, Joint Publication of the Univ. of the Aegean and EUNITE, The European Network on Intelligent Technologies for Smart Adaptive Systems, ISBN: 960-7475-23-2

9. Dounias G., Magoulas G. and Linkens D. (Eds.), (2004), *Intelligent Technologies in Bioinformatics and Medicine*, Joint Publication of the University of the Aegean and EUNITE, The European Network on Intelligent Technologies for Smart Adaptive Systems, ISBN: 960-7475-28-3
10. Dounias G. and Linkens D.A., (Eds.), (2004), "Special Issue: Adaptive Systems and hybrid Computational Intelligence in Medicine", *Int. Journal of Artificial Intelligence in Medicine*, AIIM 32(3), 151-216, Nov. 2004, Elsevier.
11. Mahfouf M. and Dounias G. (Eds.), (2004), Special Issue: «Intelligent E-Health Applications in Medicine», *Transactions of the Institute on Measurement & Control* 26(3), pp. 167-260, Arnold Journals (UK).
12. Magoulas G., Dounias G. Linkens D. (Eds.), (2006), Special Issue: Intelligent Tools for Problem Solving in Bioinformatics and Medicine, *International Journal of Artificial Intelligence Tools*, 15(3), pp. 331-411 , World Scientific.
13. Magoulas G., Dounias G. (Eds.), (2006), Special Issue on Intelligent Technologies in Bioinformatics and Medicine, (2006), *Computers in Biology and Medicine*, 36(10), October 2006, pp. 1045-1184, Elsevier S.P.
14. Akay M., *Nonlinear Biomedical Signal Processing, Fuzzy Logic, Neural Networks, and New Algorithms* (IEEE Press Series on Biomedical Engineering), 2000, Wiley-IEEE Press
15. Ying H. *Fuzzy Control and Modeling: Analytical Foundations and Applications* (IEEE Press Series on Biomedical Engineering), 2000, Wiley-IEEE Press
16. Szczepaniak P., Lisboa J., Kacprzyk J. *Fuzzy Systems in Medicine (Studies in Fuzziness and Soft Computing)*, 2000, Physica-Verlag Heidelberg
17. Barro S. and Marin R. *Fuzzy Logic in Medicine (Studies in Fuzziness and Soft Computing)*, 2002, Physica-Verlag Heidelberg
18. Chen CM, Lu HH, Lin YC. An early vision-based snake model for ultrasound image segmentation. *Ultrasound Med Biol.* 2000 Feb;26(2):273-285.
19. Yin TK, Chiu NT, A computer-aided diagnosis for locating abnormalities in bone scintigraphy by a fuzzy system with a three-step minimization approach, *IEEE TRANS. MED IMAGING* 23 (5): 639-654 MAY 2004
20. Al-Ashmouny K, Morsy A, Loza S. Sleep apnea detection and classification using fuzzy logic: clinical evaluation. *Conf Proc IEEE Eng Med Biol Soc.* 2005;6:6132-6135.
21. Morsy A, Al-Ashmouny K. Sleep apnea detection using an adaptive fuzzy logic based screening system. *Conf Proc IEEE Eng Med Biol Soc.* 2005;6:6124-6127.
22. Manivannan J, Ramasubba Reddy M, Thanikachalam S, Ajay Kumar R. Left ventricular systolic dysfunction identification by motion analysis. *Conf Proc IEEE Eng Med Biol Soc.* 2005;2:1582-1583.
23. Akbarzadeh-T MR, Moshtagh-Khorasani M.A hierarchical fuzzy rule-based approach to aphasia diagnosis. *J Biomed Inform.* 2006 Dec 24;
24. Paramasivam K, Harikumar R, Sundararajan R, Simulation of VLSI design using parallel architecture for epilepsy risk level diagnosis in diabetic neuropathy, *IETE J RES* 50 (4): 297-304 JUL-AUG 2004
25. Salvado O, Hillenbrand C, Zhang S, Wilson DL. Method to correct intensity inhomogeneity in MR images for atherosclerosis characterization. *IEEE Trans Med Imaging.* 2006 May;25(5):539-552.
26. Ullsten S, Danielsson R, Backstrom D, Sjoberg P, Bergquist J. Urine profiling using capillary electrophoresis-mass spectrometry and multivariate data analysis. *J Chromatogr A.* 2006 Jun 2;1117(1):87-93.
27. Sakka E, Prentza A, Koutsouris D. Classification algorithms for microcalcifications in mammograms. *Oncol Rep.* 2006;15; 1049-1055.
28. Manivannan J, Reddy MR, Thanikachalam S, Kumar RA. Detection of left ventricle systolic dysfunction from shape deformity. *Australas Phys Eng Sci Med.* 2005 Mar;28(1):51-55.
29. Ben Dayan Rubin DD, Baselli G, Inbar GF, Cerutti S. An adaptive neuro-fuzzy method (ANFIS) for estimating single-trial movement-related potentials. *Biol Cybern.* 2004 Aug;91(2):63-75.
30. Shyu LY, Wu YH, Hu W. Using wavelet transform and fuzzy neural network for VPC detection from the Holter ECG. *IEEE Trans Biomed Eng.* 2004 Jul;51(7):1269-1273.
31. Kim KG, Kim JH, Min BG. Classification of malignant and benign tumors using boundary characteristics in breast ultrasonograms. *J Digit Imaging.* 2002;15 Suppl 1:224-227.
32. Belal SY, Taktak AF, Nevill AJ, Spencer SA, Roden D, Bevan S. Automatic detection of distorted plethysmogram pulses in neonates and paediatric patients using an adaptive-network-based fuzzy inference system. *Artif Intell Med.* 2002 Feb;24(2):149-165.

33. Seker H, Evans DH, Aydin N, Yazgan E. Compensatory fuzzy neural networks-based intelligent detection of abnormal neonatal cerebral Doppler ultrasound waveforms. *IEEE Trans Inf Technol Biomed.* 2001;5(3):187-194.
34. Moreno L, Estevez JI, Aguilar RM, Sigut J, Gonzalez C. Exploiting the advantages of symbolically interpretable continuous biomedical parameters with a fuzzyfied symbolic model. *Artif Intell Med.* 2001 Vol.21(1-3):253-262.
35. Bellazzi R, Guglielmann R, Ironi L. Learning from biomedical time series through the integration of qualitative models and fuzzy systems. *Artif Intell Med.* 2001 Jan-Mar;21(1-3):215-220.
36. Dasey TJ, Micheli-Tzanakou E. Detection of multiple sclerosis with visual evoked potentials--an unsupervised computational intelligence system. *IEEE Trans. Inf Technol Biomed.* 2000 Sep;4(3):216-224.
37. Wahlberg P, Lantz G. Methods for robust clustering of epileptic EEG spikes. *IEEE Trans Biomed Eng.* 2000 Jul;47(7):857-868.
38. Chen H., Fuller S., Friedman C., Hersh W. *Medical Informatics: Knowledge Management and Data Mining in Biomedicine (Integrated Series in Information Systems)*, Springer, 2005
39. Wang G, Wang Z, Chen W, Zhuang J. Classification of surface EMG signals using optimal wavelet packet method based on Davies-Bouldin criterion. *Med Biol Eng Comput.* 2006 Oct;44(10):865-872.
40. Ajiboye AB, Weir RF. A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control. *IEEE Trans Neural Syst Rehabil Eng.* 2005 Sep;13(3):280-291.
41. Yoon UC, Kim JS, Kim JS, Kim IY, Kim SI. Adaptable fuzzy C-Means for improved classification as a preprocessing procedure of brain parcellation. *J Digit Imaging.* 2001 Jun;14(2 Suppl 1):238-240.
42. Yoon U, Lee JM, Kim JJ, Lee SM, Kim IY, Kwon JS, Kim SI. Modified magnetic resonance image based parcellation method for cerebral cortex using successive fuzzy clustering and boundary detection. *Ann Biomed Eng.* 2003, Apr. 31(4):441-447.
43. Brasil LM, de Azevedo FM, Barreto JM. A hybrid expert system for the diagnosis of epileptic crisis. *Artif Intell Med.* 2001 Jan-Mar;21(1-3):227-233.
44. Ward E, Martin T. A fuzzy model of glucose regulation. *J Med Syst.* 2006 Jun;30(3):187-203.
45. Ibbini M. A PI-fuzzy logic controller for the regulation of blood glucose level in diabetic patients. *J Med Eng Technol.* 2006 Mar-Apr;30(2):83-92.
46. Huang SJ, Shieh JS, Fu M, Kao MC. Fuzzy logic control for intracranial pressure via continuous propofol sedation in a neurosurgical intensive care unit. *Med Eng Phys.* 2006 Sep;28(7):639-647.
47. Su F, Wu W, Cheng Y, Chou Y. Fuzzy clustering of gait patterns of patients after ankle arthrodesis based on kinematic parameters. *Med Eng Phys.* 2001 Mar;23(2):83-90.
48. Held CM, Roy RJ. Hemodynamic management of congestive heart failure by means of a multiple mode rule-based control system using fuzzy logic. *IEEE Trans Biomed Eng.* 2000 Jan;47(1):115-123.
49. Bonivento C, Davalli A, Fantuzzi C. Tuning of myoelectric prostheses using fuzzy logic. *Artif Intell Med.* 2001 Jan-Mar;21(1-3):221-225.
50. Kobashi S, Fujiki Y, Matsui M, et al., Interactive segmentation of the cerebral lobes with fuzzy inference in 3T MR images, *IEEE T SYST MAN CYB.* 36 (1): 74-86 FEB 2006
51. Yan YH, Wang HN, Li SP, Biomedical image processing using FCM algorithm based on the wavelet transform, *J WUHAN UNIV TECHNOL* 19 (3): 18-20 SEP 2004
52. Schupp S, Elmoataz A, Fadili MJ, et al., Fast statistical level sets image segmentation for biomedical applications, *LECT NOTES COMPUT SC* 2106: 380-388 2001
53. Kim J, Cai W, Feng D, Eberl S. Segmentation of VOI from multidimensional dynamic PET images by integrating spatial and temporal features. *IEEE Trans Inf Technol Biomed.* 2006 Oct;10(4):637-646.
54. Algorri ME, Flores-Mangas F, Classification of anatomical structures in MR brain images using fuzzy parameters, *IEEE T BIO-MED ENG* 51 (9): 1595-1608 SEP 2004
55. Carvalho BM, Herman GT, Kong TY, Simultaneous fuzzy segmentation of multiple objects, *DISCRETE APPL MATH* 151 (1-3): 55-77 OCT 1 2005
56. Hou Z, San Koh T. Wavelet shrinkage prefiltering for brain tissue segmentation. *Conf Proc IEEE Eng Med Biol Soc.* 2005;2:1604-1606.
57. Zhou Y, Bai J. Atlas-based fuzzy connectedness segmentation and intensity nonuniformity correction applied to brain MRI. *IEEE Trans Biomed Eng.* 2007 Jan;54(1):122-129.

58. Jacq JJ, Roux C, Stindel E, et al., Analytical surface recognition in three-dimensional (3D) medical images using genetic matching: Application to the extraction of spheroidal articular surfaces in 3D computed tomography data sets, *INT J IMAG SYST TECH* 11 (1): 30-43 2000
59. Lasch P, Haensch W, Naumann D, et al., Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis, *BBA-MOL BASIS DIS* 1688 (2): 176-186 MAR 2 2004
60. Majerova D, Kukal J, Lukasiewicz ANN for local image processing, *NEURAL NETW WORLD* 15 (6): 535-551 2005
61. Pham TD, Ran DT, Image classification by fusion for high-content cell-cycle screening, *LECT NOTES ARTIF INT* 4251: 524-531 2006
62. Mondal PP, Positron emission tomographic map reconstruction using fuzzy-median filter, *APPL PHYS LETT* 89 (15): art. no. 153903 OCT 9 2006
63. Hu Q, Qian G, Aziz A, Nowinski W. Segmentation of brain from computed tomography head images. *Conf Proc IEEE Eng Med Biol Soc.* 2005;4:3375-8.
64. Zhou J, Krishnan S, Chong V, Huang J. Extraction of tongue carcinoma using genetic algorithm-induced fuzzy clustering and artificial neural network from MR images. *Conf Proc IEEE Eng Med Biol Soc.* 2004;3:1790-1793.
65. Zhou Y, Bai J. Organ segmentation using atlas registration and fuzzy connectedness. *Conf Proc IEEE Eng Med Biol Soc.* 2005;3:3241-3244.
66. Sun B, Giddens DP, Long R Jr, Taylor WR, Weiss D, Joseph G, Vega D, Oshinski JN. Characterization of coronary atherosclerotic plaque using multicontrast MRI acquired under simulated in vivo conditions. *J Magn Reson Imaging.* 2006 Oct;24(4):833-841.
67. Kim J, Cai W, Feng D, Wu H. A new way for multidimensional medical data management: volume of interest (VOI)-based retrieval of medical images with visual and functional features. *IEEE Trans Inf Technol Biomed.* 2006 Jul;10(3):598-607.
68. Xue Z, Shen D, Davatzikos C. CLASSIC: consistent longitudinal alignment and segmentation for serial image computing. *Neuroimage*. 2006 Apr 1;30(2):388-399.
69. Cui J, Loewy J, Kendall EJ. Automated search for arthritic patterns in infrared spectra of synovial fluid using adaptive wavelets and fuzzy C-means analysis. *IEEE Trans Biomed Eng.* 2006 May;53(5):800-809.
70. Cho MH, Chun IK, Lee SC, Cho MH, Lee SY. Trabecular thickness measurement in cancellous bones: postmortem rat studies with the zoom-in micro-tomography technique. *Physiol Meas.* 2005 Oct;26(5):667-676.
71. Shan ZY, Liu JZ, Yue GH. Automated human frontal lobe identification in MR images based on fuzzy-logic encoded expert anatomic knowledge. *Magn Reson Imaging.* 2004 Jun;22(5):607-617.
72. Latson L, Sebek B, Powell KA. Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy. *Anal Quant Cytol Histol.* 2003 Dec;25(6):321-331.
73. Coatrieux JL. Signal processing and physiological modeling--part 1: Surface analysis. *Crit Rev Biomed Eng.* 2002;30(1-3):9-35
74. Bay OF, Usakli AB. Survey of fuzzy logic applications in brain-related researches. *J Med Syst.* 2003 Apr;27(2):215-223.
75. Gavrielides MA, Lo JY, Floyd CE Jr. Parameter optimization of a computer-aided diagnosis scheme for the segmentation of microcalcification clusters in mammograms. *Med Phys.* 2002 Apr;29(4):475-483.
76. Ahmed MN, Yamany SM, Mohamed N, Farag AA, Moriarty T. A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Trans Med Imaging.* 2002 Mar;21(3):193-199.
77. Das A, Reddy NP, Narayanan J. Hybrid fuzzy logic committee neural networks for recognition of swallow acceleration signals. *Comput Methods Programs Biomed.* 2001 Feb;64(2):87-99.
78. Yang X, Krishnan SM, Chan KL. Color image segmentation based on fuzzy rule-based reasoning applied to colonoscopic images. *Crit Rev Biomed Eng*. 2000; 28(3 - 4):355-361.
79. Carano RA, Li F, Irie K, Helmer KG, Silva MD, Fisher M, Sotak CH. Multispectral analysis of the temporal evolution of cerebral ischemia in the rat brain. *J Magn Reson Imaging.* 2000 Dec;12(6):842-858.
80. Gavrielides MA, Lo JY, Vargas-Voracek R, Floyd CE Jr. Segmentation of suspicious clustered microcalcifications in mammograms. *Med Phys.* 2000 Jan; 27(1):13-22.
81. Chen CM, Lu HH. An adaptive snake model for ultrasound image segmentation: modified trimmed mean filter, ramp integration and adaptive weighting parameters. *Ultrason Imaging.* 2000 Oct;22(4):214-236.
82. Mastorocostas PA, Tolias YA, Theocharis JB, Hadjileontiadis LJ, Panas SM. An orthogonal least squares-based fuzzy filter for real-time analysis of lung sounds. *IEEE Trans Biomed Eng.* 2000 Sep;47(9):1165-1176.

83. Du CJ, Yin HE, Wu SC, Ren XY, Zeng YJ, Pan YF. Visual evoked potentials estimation by adaptive noise cancellation with neural-network-based fuzzy inference system. *Conf Proc IEEE Eng Med Biol Soc.* 2004;1:624-627
84. Ellingson BM, Ulmer JL, Schmit BD. A new technique for imaging the human spinal cord in vivo. *Biomed Sci Instrum.* 2006;42:255-260.
85. Lee JM, Yoon U, Kim JJ, Kim IY, Lee DS, Kwon JS, Kim SI. Analysis of the hemispheric asymmetry using fractal dimension of a skeletonized cerebral surface. *IEEE Trans Biomed Eng.* 2004 Aug;51(8):1494-1498.
86. Nunes CS, Mahfouf M, Linkens DA, Fuzzy modelling for controlled anaesthesia in hospital operating theatres, *CONTROL ENG PRACT* 14 (5): 563-572, 2006
87. Kiguchi K, Tanaka T, Fukuda T, Neuro-fuzzy control of a robotic exoskeleton with EMG signals. *IEEE T FUZZY SYST* 12 (4): 481-490 AUG 2004
88. Zhang D, Zhu K. Modeling biological motor control for human locomotion with functional electrical stimulation. *Biol Cybern.* 2007 Jan;96(1):79-97
89. Davoodi R, Andrews BJ. Fuzzy logic control of FES rowing exercise in paraplegia. *IEEE Trans Biomed Eng.* 2004 Mar;51(3):541-543.
90. Zhang XS, Roy RJ. Derived fuzzy knowledge model for estimating the depth of anesthesia. *IEEE Trans Biomed Eng.* 2001 Mar;48(3):312-323.
91. Feng S, Li L, Wang P, et al., Fuzzy modeling of the medical treatment effects of superoxide dismutase. *EXPERT SYST* 23 (5): 323-329 NOV 2006
92. Charbonnier S, Galichet S, Mauris G, et al., Statistical and fuzzy models of ambulatory systolic blood pressure for hypertension diagnosis, *IEEE T INSTRUM MEAS* 49 (5): 998-1003 OCT 2000
93. Kumar BH. A fuzzy expert system design for analysis of body sounds and design of an unique electronic stethoscope (development of HILSA kit). *Biosens Bioelectron.* 2007 Jan 15;22(6):1121-1125.
94. Shin JW, Yoon JH, Yoon YR. Rate-adaptive pacemaker controlled by motion and respiratory rate using neuro-fuzzy algorithm. *Med Biol Eng Comput.* 2001 Nov;39(6):694-699.
95. Tapaswi S, Joshi RC, Classification of bio-medical images using neuro-fuzzy approach, *LECT NOTES COMPUT SC* 2973: 568-581 2004
96. Li D, Pedrycz W, Pizzi NJ. Fuzzy wavelet packet based feature extraction method and its application to biomedical signal classification. *IEEE Trans Biomed Eng.* 2005 Jun;52(6):1132-1139.
97. Jin B, Tang YC, Zhang YQ, Support vector machines with genetic fuzzy feature transformation for biomedical data classification, *INFORM SCIENCES* 177 (2): 476-489 JAN 15 2007
98. Chen XJ, Harrison R, Zhang YQ, Multi-SVM fuzzy classification and fusion method and applications in bioinformatics, *J COMPUT THEOR NANOS* 2 (4): 534-542 DEC 2005
99. Suzuki Y, Itakura KI, Saga S, et al., Signal processing and pattern recognition with soft computing, P IEEE 89 (9): 1297-1317 SEP 2001
100. Song M, Lee J, Park H, Lee K. Classification of Heartbeats based on Linear Discriminant Analysis and Artificial Neural Network. *Conf Proc IEEE Eng Med Biol Soc.* 2005;2:1151-1153.
101. Rasheed S, Stashuk D, Kamel M. Adaptive fuzzy k-NN classifier for EMG signal decomposition. *Med Eng Phys.* 2006 Sep;28(7):694-709.
102. Osowski S, Linh TH. ECG beat recognition using fuzzy hybrid neural network. *IEEE Trans Biomed Eng.* 2001 Nov;48(11):1265-1271.
103. 103 Wang Y, Winters J. An input classification scheme for use in evidence-based dynamic recurrent neuro-fuzzy prognosis. *Conf Proc IEEE Eng Med Biol Soc.* 2004;5:3198-3201.
104. Virant-Klun I, Virant J, Fuzzy logic alternative for analysis in the biomedical sciences, *COMPUT BIOMED RES* 32 (4): 305-321 AUG 1999
105. Pizzi N. Bleeding predisposition assessments in tonsillectomy/adenoidectomy patients using fuzzy interquartile encoded neural networks. *Artif Intell Med.* 2001 Jan-Mar;21(1-3):65-90.
106. Uguz H, Arslan A, Turkoglu I, A biomedical system based on hidden Markov model for diagnosis of the heart valve diseases, *PATTERN RECOGN LETT* 28 (4): 395-404 MAR 1 2007
107. Manivannan J, Reddy MR, Thanikachalam S, Kumar RA. Quantitative evaluation of left ventricle performance from two dimensional echo images. *Echocardiography*. 2006 Feb;23(2):87-92.
108. Wang Y, Winters JM. A dynamic neuro-fuzzy model providing bio-state estimation and prognosis prediction for wearable intelligent assistants. *J Neuroengineering Rehabil.* 2005 Jun 28;2(1):15.

109. Seker H, Odetayo MO, Petrovic D, Naguib RN. A fuzzy logic based-method for prognostic decision making in breast and prostate cancers. *IEEE Trans Inf Technol Biomed.* 2003 Jun;7(2):114-122.
110. Wadie BS, Badawi AM, Ghoneim MA. The relationship of the International Prostate Symptom Score and objective parameters for diagnosing bladder outlet obstruction. Part II: the potential usefulness of artificial neural networks. *J Urol.* 2001 Jan;165(1):35-37.
111. Wang Y, Zhu YS, Thakor NV, Xu YH. A short-time multifractal approach for arrhythmia detection based on fuzzy neural network. *IEEE Trans Biomed Eng.* 2001 Sep;48(9):989-995.
112. Pelta D, Krasnogor N, Bousono-Calzon C, et al., A fuzzy sets based generalization of contact maps for the overlap of protein structures, *FUZZY SET SYST* 152 (1): 103-123 MAY 16 2005
113. Vinterbo SA, Kim EY, Ohno-Machado L. Small, fuzzy and interpretable gene expression based classifiers. *Bioinformatics*. 2005 May 1;21(9):1964-1970.
114. Robson B, Mushlin R. Clinical and pharmacogenomic data mining: 2. A simple method for the combination of information from associations and multivariances to facilitate analysis, decision, and design in clinical research and practice. *J Proteome Res.* 2004 Jul-Aug;3(4):697-711.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Interpretation of gene expression microarray experiments

Aristotelis Chatziioannou, Panagiotis Moulos
Insititute of Biological Research and Biotechnology, National Hellenic Research Foundation, 48 Vassileos Constantinou ave., 11635 Athens, Greece

Abstract. Microarrays nowadays have an almost ubiquitous presence in modern biological research The extent and versatility of the techniques that are available for analysis and interpretation of microarray experiments can be somehow bewildering to the interested biologists. Functional genomics involves the high-throughput analysis of large datasets of information derived from various biological experiments. Microarray technology makes this possible by monitoring the emitting fluorescence reflecting the expression levels of thousands of genes simultaneously, which are bound to the oligonucleotide probes specific for each of the putative gene sequences comprising the total genome of the investigated organism, under a particular condition.. This chapter is a brief overview of the basic concepts involved in a microarray experiment; and it aspires to provide a concise overview of key issues regarding the various steps of implementation of this promising experimental methodology. In this sense, the chapter gives a feeling for what the data actually represent, and will provide information on the various computational methods that one can employ to derive meaningful results from such experiments.

Keywords. microarrays, bioinformatics, gene expression, data normalization, genomics, metaanalysis.

Introduction

Recently, genomics have gained crucial importance in modern biological research. This is also corroborated by the deluge of papers referring to various aspects of functional genomics, as microarray gene profiling experiments, the investigation of transcriptional regulatory networks, the correlation of the total expression of the genome with specific biological processes through the formulation of appropriate ontologies, the study of Single Nucleotide Polymorphisms of specific genes, the comparative analysis of clusters of genes based on their sequence, etc. The analysis of large datasets of information derived from various biological experiments, was greatly enabled by the completion of the Human Genome Project a milestone which purported the complete sequencing of the human genome as well as that of many other biological species, a number constantly augmenting ever since, thus yielding huge amounts of information concerning the genomic content of species. Besides that, recent technological developments in the field of computer architecture, which speeded up the computational capacity of computer processors, just as the adoption of state-of the art techniques of photolithography rendered microarrays a novel promising platform for high-throughput biological research.

Microarrays nowadays have an almost ubiquitous presence in biological research [1], [2], [3], [4]. At the same time, the statistical methodology for microarray analysis has progressed so as to shift the limits of the application from the simple visual assessments of results of the first days to a weekly deluge of papers that describe purportedly novel algorithms for analyzing changes in gene expression [5]. The extent and versatility of the techniques that are available can be somehow bewildering to the interested biologists. Statistical geneticists though, are recognizing commonalities among the different methods. Many are special cases of more general models, and points of consensus are emerging about the general approaches that warrant use and elaboration.

Functional genomics involves the high-throughput analysis of large datasets of information derived from various biological experiments. Microarray technology makes this possible by monitoring the emitting fluorescence reflecting the expression levels of thousands of genes simultaneously, which are bound to the oligonucleotide probes specific for each of the putative gene sequences comprising the total genome of the investigated organism, under a particular condition.

This chapter is a brief overview of the basic concepts involved in a microarray experiment; and as such it aspires to provide an introductory yet concise overview of key issues regarding the application of this experimental methodology. In this sense, the chapter gives a feeling for what the data actually represent, and will provide information on the various computational methods that one can employ to derive meaningful results from such experiments.

1. Key features of a microarray experiment

A key feature of the gene profiling microarray experiments is their capability to process large datasets of gene expression signals derived from various biological experiments, thus constituting a promising high-throughput experimental methodology and an important means for functional genomics studies [6]. The study of gene expression profiling of cells and tissue has recently become a major discovery goal in biological research. Therefore microarrays, constitute an indispensable tool in the disposal of modern molecular biological research, for the inspection of genome-wide changes in gene expression for a given organism. Two of the frequent goals of genome-wide gene expression experiments are to identify significant alterations in transcript levels resulting from the exposure of a living system to a test agent at a given dose and time [7] and develop genetic signatures in order to distinguish between health and disease states. Additionally, such high–throughput expression profiling can be used to compare the level of gene transcription in clinical studies conditions in order to: i) identify and categorize diagnostic or prognostic biomarkers ii) classify diseases, e.g. tumours with different prognosis status that are indistinguishable by microscopic histological examination iii) monitor the response to therapy and iv) understand the mechanisms involved in the genesis of disease processes [3], [4], [8].

Generally speaking, the conduct of a microarray experiment entails a number of steps, which are depicted in Figure 1. According to this a microarray experiment can be processed in a series of steps, namely: ***image acquisition and processing, signal quantitation and background correction, data normalization, gene filtering and missing values imputation, statistical selection of differentially expressed probes and meta-analysis***. Each of these steps seeks to resolve different problems, varying from the more correct estimation of the signal intensity relating to the expression of a specific gene, the limitation of the influence of the noise, the mitigation of the undesired technical variability among samples, the more precise estimation of the differential expression of a set of genes, the more robust statistical selection in terms of specificity and sensitivity for a number of significant genes and finally the derivation of a pattern of putative correlation among them. The many current methods applied in each step can be reduced to a modest number of categories, but there is often little evidence to support one method within a category over others. In such situations, it is more a question of which of all possible methods, follows the same assumptions that are implicit in the implementation and the interpretation of the experiment by the experts. This entails concrete knowledge of the effect of the various mathematical transformations, applied to the population of values of the experiment.

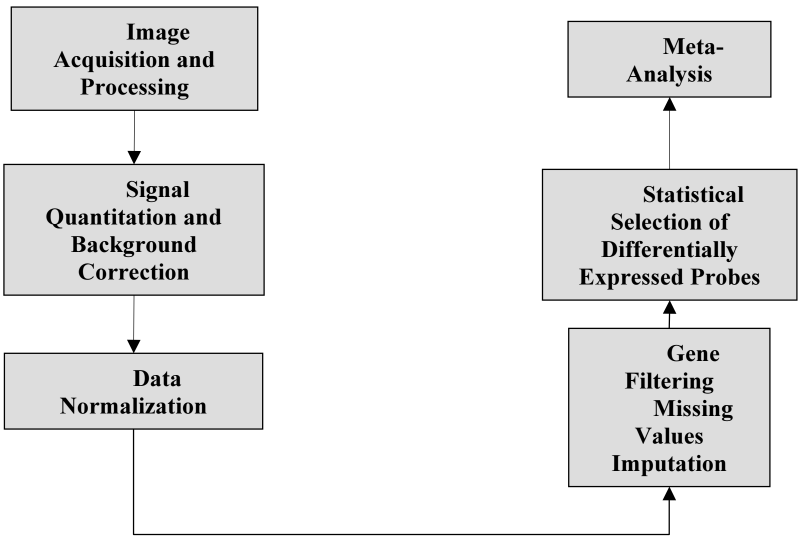


Fig. 1. General steps included in typical microarray experiment data analysis.

The performance of microarray experiments is based in the concept of DNA hybridization that is, the physicochemical interaction among complementary nucleic acids strands resulting in their binding [9]. According to the length of the probes, arrays can be classified into ‘‘complementary DNA (cDNA) arrays,’’ which use long probes of hundreds or thousands of base pairs (bps), and ‘‘oligonucleotide arrays,’’ which use short probes (usually 50 bps or less). Gene expression microarrays, can be considered as large parallel Northern blot analyses (instead of a gel, the probes are attached to an inert surface, namely the microarray slide), where oligonucleotides are used to hybridize to messenger RNA (mRNA) [10]. mRNA is extracted from tissues or cells, reversed-transcribed, and according to the microarray technology applied labelling is following, either with one (single channel technology - Affymetrix, Nimblegen) or multiple (usually two, Cy3 (green)) for the reference sample and Cy5 for the treated sample - Sanger, Agilent, and other custom made proprietary cDNA microarrays from various institutes) fluorescent dyes, and hybridized on the array. In the case of two-channel cDNA microarrays, an experimental technique is adopted, enabling expression to be assayed and compared between appropriate sample pairs. Hybridization and washes are carried out under high stringency conditions to diminish the possibility of cross-hybridization between similar genes. The next step is to acquire a snapshot of the whole genomic expression, as this is given by the emitted fluorescence from the labelled hybridized nucleic acid strands on to the probes, corresponding consistently to the thousands of genes comprising the genome of an organism. This is done by exploiting technological innovations in the field of laser-induced fluorescent imaging such as array scanners are (i.e. Affymetrix, Packard biosciences, Agilent), and by storing the derived digitized image in a format preserving the information content of the image. The aforementioned process is presented in Figure 2.

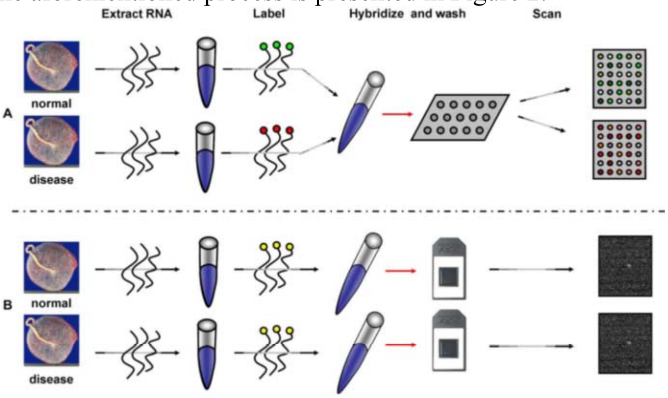


Fig. 2. Schematic representation of the steps involved in microarrays. The upper panel (A) illustrates the two channel technology while the lower panel (B) illustrates the single channel technology. The experiment is designed to compare the mRNA expression profile of the control (usually physiologically normal condition) with that of the disease. mRNA is extracted from the biological source tissue. In panel (A), the normal and disease mRNA are labeled with two different dyes, mixed and then hybridized on the same array. After washing, the array is scanned at two different wavelengths to yield two images one for the control case and one for the disease which following can be projected to yield one composite image colored according to a consensus pseudocolor scale. In panel (B) (single channel), each sample is labeled with the same fluorescent dye, but independently hybridized on different arrays (taken from [11]).

2. Processing of the signals of the microarray images

Once the microarrays have been hybridized, the resulting images are used to generate a dataset. This dataset needs pre-processing prior to the analysis and interpretation of the results. Pre-processing is a step that entails useful transformations and assessment of the signal quality of the gene probes, in order to extract or enhance meaningful data characteristics which render the dataset amenable to the application of various data analysis methods.

The principle behind the quantification of expression levels is that the amount of fluorescence measured at each sequence specific location is proportional to the amount of mRNA hybridized onto the gene probes fastened on the array slide. These images must be then processed in order to track down the arrayed spots and quantify the relative fluorescence intensities for each element. Microarray experiments do not directly provide insight on the absolute level of expression of a particular gene; nevertheless, they are useful to compare the expression level among conditions and genes (e.g. health vs. disease, treated vs. untreated) [1], [11].

Typical pre-processing steps are:

- background noise correction

- filtering procedures to eliminate non-informative genes
- logarithmic transformation of the signal channel values and calculation of their ratio in the case of two channel experiments

2.1. Background Correction

The background correction is intended to adjust for non-specific hybridization, such as hybridization of sample transcripts whose sequences do not perfectly match those of the probes on the array [12] and for other systematic or technical issues such as the presence of artefacts on the arrays, scanner technical setbacks, washing issues or quantum fluctuations. Depending on the type of microarray technology adopted (cDNA or high density oligonoucleotide probe arrays), and the specifics of the design of the slide which include various types of spots for estimation of the background (empty spots, exogenous negative control spots (e.g., Arabidopsis DNA probes, a plant, for a human array), different methods for the background correction task are in disposal. In the case of Affymetrix arrays, the background can be estimated through the 'mismatch probes' which cover all surface of the slide. Spot background correction, when applied, can be effected either by subtracting the background value from the signal value for each probe, or by dividing the background value from the signal value for each probe that is, subtraction of the log transformed values of the background from the signal respectively or by providing model based corrected signal estimations. For the estimation of the signal and the background, either the mean or the median of the spot signal and background can be utilized. In the case of the division of the signal value by the background value, the signal-to-noise content of a signal, an established notion in systems theory, is taking into account the perception of the experimentalist about the quality of a signal, focusing in the fact that the signal quality is considered good when its strength is some orders of magnitude greater than that of the background, that is it ratio compared to noise, is big enough, according to criteria either set by the expert or inferred statistically. This might prove critical especially when dealing with weak signal datasets whereas a majority of spot signals is close or even below the background contamination levels. [13].

Other options for background correction constitute for example of using computational techniques that model the distribution of the observed intensity values and estimate the background noise based on mathematical models. The filtering procedures aim at excluding problematic or low in information content array spots and are usually based in processes that use the amount of spot background contamination or outlier detection to locate candidate gene to be removed from further analysis. The ratio between channels of treated and reference samples is a simple and intuitive measure which can investigate relationships between related biological samples based on expression patterns.

2.2. Logarithmic Transformation

A better transformation procedure is to take the logarithm base 2 value of the expression ratio (i.e. log2 (expression ratio)). This has the major advantage that it treats differential up-regulation and down-regulation equally [1] and also projects data on a continuous mapping space. For example, if the expression ratio is 1, then $\log_2(1)$ equals 0 and represents no change in expression. If the expression ratio is 4, then $\log_2(4)$ equals +2 and for expression ratio of $\log_2(1/4)$ equals -2. Thus, in this transformation the mapping space is continuous and the same level of up-regulation and down-regulation are treated equally. Having explained the advantages of using expression ratios as a metric for gene expression, it should also be understood that there are disadvantages of using expression ratios or transformations of the ratios for data analysis. Even though expression ratios can reveal patterns inherent in the data, they remove all information about absolute expression levels of the genes. For example, genes that have R/G ratios of 400/100 and 4/1 will end up having the same expression ratio of 4, and associated problems will emerge when one tries to reliably validate genes computationally identified as differentially expressed by, taking into account the absolute amount of their expression too, namely if they are expressed at high or low absolute levels with respect to the total RNA content.

2.3. Data Filtering

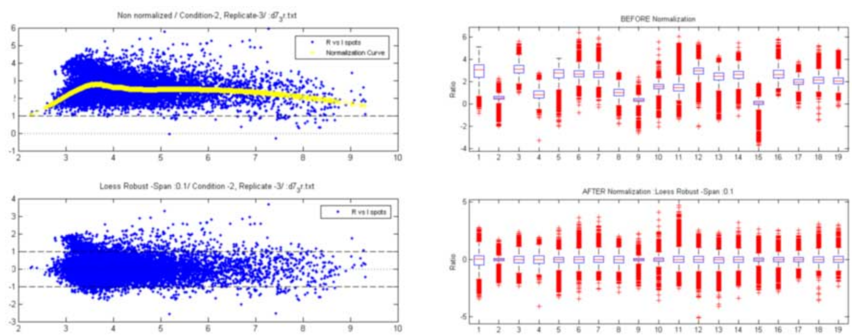
The filtering procedures aim at excluding problematic or low, regarding their information content, array spots. They are usually based in processes that use the amount of spot background contamination or outlier detection to locate candidate genes to be removed from further analysis. Such problems in spot quality might be the result of hybridization of sample transcripts whose sequences do not perfectly match those of the probes on the array [11] and for other systematic or technical issues such as the presence of artefacts on the arrays, scanner technical setbacks or washing issues. Poor quality spots that should not participate in data normalization or further analysis can be identified with a variety of methods. Some of them follow [13]:

- A signal-to-noise threshold filter based on the formula $\left(\overline{S}/\overline{B}\right) < T$, where T is a threshold below which noisy spots are filtered out.
- A filter based on the signal and background noise distributions distance: a spot is robust against this filter if its signal and noise distributions diverge from each other a distance determined by the respective standard deviations. Spots sensitive to this filter are determined by the inequality $\overline{S} - x\sigma_S^2 < \overline{B} + y\sigma_B^2$, where *x* and *y* parameters.
- A test filter for the reproducibility of measurements: such a test could be a statistical hypothesis test like t-test (parametric) or Wilcoxon (non–parametric) test to verify that for each spot, the derived ratios from the channel measurements of all condition replicates follow a normal (or more generally a continuous symmetrical) distribution with mean (median) equal to the average ratio for this spot among all replicates. This test should track and exclude outliers among the replicate slides of a condition.

Generally, array spots sensitive to any filtering procedure applied should be excluded from the estimation of the normalization curve to alleviate the normalization procedure from the impact of systematic measurement errors [13].

2.4. Data Normalization

As another preprocessing step, normalization is considered critical to compensate for systematic differences among genes or arrays and provide appropriate balances in order to derive meaningful biological comparisons. The need for data normalization stems from a variety of reasons which among others attempts to tackle problems such as unequal quantities of RNA, differences in labelling or the fluorescent dyes and systematic biases in the measured expression levels coming either from acquiring microarray images with different PMT or laser gains or from flaws in the washing/dilution procedure[1], [80]. Typical normalization methods are global mean or median normalization [14], rank invariant normalization [15] and LOWESS/LOESS methods [16]. While global methods take into account the expression of all genes in a microarray slide, based on the assumption that the total level of RNA expression through varying conditions is roughly constant, insignuating that the level of expression of the housekeeping genes a) does not exhibit significant variation and b) holds safely by far the greatest percentage of the total RNA expression, the plausibility of this assumption might prove refutable according to the number the type and the number of replicates for the genes printed on the array. On the other hand, rank invariant methods are based exclusively on signal intensities and they do not use all genes during the normalization procedure. Instead, a statistical algorithm determines a subset of genes which are found to be non differentially expressed across different slides. However, they do not account for systematic dependencies of the log ratio statistics to signal intensity [1]. Due to the aforementioned reasons, it seems that the most widely used normalization methods in microarray experiments are the LOWESS/LOESS methods, mostly because of their local processing nature. The effect of normalization is depicted in Figure 3.



(a) (b)

Fig. 3. (a) An MA plot for a cDNA microarray slide. It displays the $\log_2(R/G)$ for each element of the array as a function of the $1/2\log_2(R*G)$ product intensities and can reveal systematic intensity dependent effects on the $\log_2$ ratio values [1]. R and G correspond to corrected Cy5 and Cy3 signal intensities respectively. The upper panel of (a) depicts the MA plot before normalization while the bottom panel, after LOESS normalization. The bright line on the upper panel draws the normalization curve. After normalization, the cloud is centered around zero (b) A boxplot for 19 cDNA arrays. The upper panel of (b) presents the systematic trends in microarray slides briefly mentioned in sections 2.1-2.4. Direct comparisons in ratios cannot be made since data are not centered around a common reference value. The trends are removed after normalization, as depicted in the bottom panel of (b).

Given the impact of normalization methods on subsequent analysis steps [17], the decision on which normalization method is proper may depend on the biological nature of the dataset interrogated. There exist several methods for normalizing cDNA as well as oligonucleotide microarray data and abundant literature is available on the subject [18], [19], [20], [21], [22], [23], [24], [25], [26]. Because of this impact of normalization methods on the final interpretation of a gene expression microarray experiment [81], claiming superiority-equivalency over alternative normalization methods is more than justified. Though the topic of comparing / assessing different normalization methods in microarray experiments is not adequately explored as in general the comparison among results of microarray experiments [5], there has been a seminal effort [80], in the field of comparison among normalization methods to propose a framework in order to aid the comparison of normalization algorithms and the selection of the most appropriate one; with the scope of providing a checklist for the researcher decide which algorithm to use for his dataset. In brief, the algorithm that is associated with the less over-normalization potential (smaller reduction of the total signal entropy), smaller bias (greater accuracy) and variance (greater precision) is the optimal one for the dataset at hand.

2.5. Statistical Inference

2.5.1. Imputation of Missing Values

As a result of several unpredictable factors (e.g. array scratches, scanner improper configuration, spot light saturation etc.) during a microarray experiment, expression measurements may become unavailable for several probes. One possible solution could be to exclude whole slides that appear problematic. However, this solution is impractical since usually none of the slides is perfect and modern arrays contain tens of thousands of probes making measurements more sensitive to artefacts. Since many statistical algorithms require that all values are present in the population under analysis, proper algorithms are required to calculate replacements for the missing values. A naïve solution is to estimate missing data for a probe by the mean expression of the remaining probes over the rest of the slides. However, this strategy demonstrates poor performance [27] and does not account for the correlation among genes. More sophisticated methods have been proposed using matrix Singular Value Decomposition and k-Nearest Neighbors methods [27], [28].

2.5.2. Statistical Selection

Historically, the log ratio fold-change statistic was the first measurement used to determine if a gene is differentially expressed or not. However, even if in biology it is believed that “the greater the magnitude of change, the higher the likelihood of physiologic or pathologic significance” [11] it was found inadequate since it does not incorporate the variance of gene expression measurements [5] and the fold-change thresholds set by investigators were arbitrary. Consequently, the application of more sophisticated statistical techniques involving hypothesis testing was incorporated in microarray data analysis.

The most common goal in a microarray experiment is to identify genes that are differentially expressed between two or more states. Thus, for each gene on the array, the “null” hypothesis that this gene is not differentially expressed among different condition is usually compared against the “alternative” that there is difference in the gene’s expression among the different states. There exist several widely used statistics which are appropriate for such hypothesis testing, parametric (t-test, ANOVA) [12], [29], [30], [31] or non-parametric (Wilcoxon sign-rank, Kruskal-Wallis) [32] which can be applied to cDNA or oligonucleotide arrays [33]. A comparison of several statistical selection methods can be found in [34]. These procedures usually result in gene lists ranked with a statistical significance score (p-value). A threshold is then applied to the score to determine the differentially from the non-differentially expressed genes. A common problem with this approach is that while a strict p-value threshold would provide assurance on the statistical significance, many genes may not reach that threshold and end up with very limited gene lists with poor biological relevance.

As in general there is not a golden standard method for microarray data statistical analysis and the results of several methods often demonstrate small overlap [35], a fact which does not appear to depend on different microarray platform technologies [36], [37], a recent study [38] suggests that the overlap between results derived by different methods, platforms and also different laboratories investigating similar biological issues is amplified if a less stringent threshold on statistical scoring is used in combination with a reasonable fold-change cut-off.

2.5.3. Multiple Testing Issues

Hypothesis testing in statistics involves two kinds of errors: the Type I error occurs when the null hypothesis is incorrectly rejected. In the context of microarray experiments, type I errors are committed when a gene is declared differentially expressed while it is not. On the other hand, Type II errors occur when the null hypothesis is not rejected while it is false, that is, in the context of microarrays, when the test fails to identify a differentially expressed gene [39]. When conducting multiple testing, the probability of Type I errors is increased proportionally to the number of tests. This is allowable when the number of tests is small, but in the case of microarrays where thousands of tests are performed, a large number of false positives is undesirable. For example, if an experiment involves testing over 10000 genes and the p-value threshold to determine differential expression is set to 0.01, then at least 100 false positives are expected. Such unwelcome results necessitate the correction of statistical scores to adjust for multiple hypothesis testing.

There exist two main categories of multiple testing correction methods: the Family Wise Error Rate (FWER) and the False Discovery Rate (FDR) methods. FWER procedures correct for multiple testing by adjusting p-values to account for multiple testing. For example the Bonferroni procedure adjusts p-values by dividing with the number of hypotheses n to be tested ($p_{adj}=p/n$), but this type of correction proves to be extremely stringent practically for the vast majority of microarray experiments. Moreover it does not take into account at all the distribution of signal values of the experiment and therefore the possible reasons for the inference of false positives. See [39] for further details on FWER procedures. FWER methods are unsuitable for microarray data mostly because they are too conservative: after correcting for multiple testing, no single gene may meet the threshold for statistical significance. In contrast, FDR methods [40], [41] instead of adjusting p-values, they seek to minimize the proportion of errors committed by falsely rejecting null hypotheses. As they are less stringent than FWER methods they are considered more suitable for microarray data [5], [11]. However, a common drawback with both of them is that they do not assume general variable dependence which is usually the case for microarrays because genes are involved in complicated interaction networks and pathways [11].

2.6. Software

Several microarray data analysis software packages, commercial (e.g. GeneSight™, BioDiscovery Inc., Los Angeles, USA) or open source (e.g. Bioconductor, [42]), are currently available. A major drawback with most of the latter is the absence of a predefined analysis protocol that starts from raw data and results in differentially expressed gene lists. Furthermore, many analysis packages are provided as sets of routines (e.g. Bioconductor) which can often be of little use to biologists with small experience in programming and other packages which provide a graphical user interface are specialized in data visualization or statistical selection. An exception to the latter rules is MIDAS [43] and ANDROMEDA [44] where the user can pre-program the desired analysis steps in a batch process. Out of the latter, only ANDROMEDA supports input from several microarray image analysis programs without prior manual transformation.

3. Meta Analysis

No matter how important the role of microarray statistical analysis of microarray data might be, the extraction of significantly differentially expressed gene lists still remains a first step lying at most times far away from gaining real insight on the biological subject interrogated. As a result, researchers usually find themselves bewildered by a myriad of findings [5], with no evident pattern of correlation among them. More specifically, an attempt to understand isolated genes in a list would prove particularly demanding and laborious for the following reasons [45]:

- Multiple hypothesis testing correction may leave no statistically significant individual genes because of modest relevant biological differences relatively to the noise inherent to the microarray technology.
- A long list of significant genes may not be connected to any unifying biological theme that would provide space for ad hoc and biologist's area of expertise dependent interpretation.
- Single-gene analysis may miss important information on biological pathways.
- In the gene-list level, studies from different laboratories on the same subject show dramatically little overlap [46].

Transforming information into knowledge and obtaining further insight over specific biological questions which the researcher targets to investigate through a microarray experiment, are by all means self-evident tasks. Additional analysis is required regarding the exploitation of available knowledge databases, in order to reveal biological pathways and mechanisms which underlie the biological issue under investigation. Moreover, the use of sophisticated statistical learning and data mining techniques to identify important genes which are able to distinguish among different experimental conditions (e.g. healthy tissue from cancer tissue or two different tumour types) and create molecular genetic signatures for specific diseases would lead to specific targets for drug design research. This exhaustive search process can be termed as 'meta-analysis' and its steps usually include pathway analysis to uncover genes with a certain expression profile which are mapped to or regulated by elements of the same pathway, exploration for common regulatory elements among groups of genes and gene functional analysis based on biological databases or ontologies. It also includes the utilization of statistical algorithms which are able to isolate the informative from the non-informative genes whose function characterizes a biological problem under investigation and of statistical classification techniques which serve to the discovery of genetic signatures consisting of several genes which appear to act as "switches" between two or more biological states.

3.1. Classification

3.1.1. Unsupervised and Supervised Learning

In many situations there is concern on assigning biological samples to classes on the basis of gene expression measurements acquired through a microarray experiment [12], aiming at knowledge improvement on matters of health issues such as disease prognosis and prediction. Such problems are mainly classified into two categories: clustering and discrimination or unsupervised and *supervised learning*. In unsupervised learning, the algorithm is given a set of objects and attempts to group them into classes without any prior knowledge of these classes or any labelled output [47]. The learning task is to gain some understanding of the process that is captured in the generated data and also to estimate the number of classes. This task is accomplished through proper metrics that estimate the relative distance between objects in a proper defined variable space (the rationale supporting the distance estimation is that objects which are closer, present an increased probability to hold common functional grounds for their similarities and so should be assembled together) and algorithms that determine how the closest objects will be grouped [48]. Examples of unsupervised learning techniques are *k-means clustering*, *hierarchical clustering [12], gene-shaving* [49], self-organizing maps [50], [51] and FLAME [52]. As there are no gold standards on which clustering algorithm is best [12] the solution to this dubiety depends on the nature of the biological dataset under interrogation and the experimental design [53]. A short comparative study of cluster analysis algorithms can also be found in [52].

On the other hand, supervised learning algorithms make use of a set of classified examples and given this sample of input-output pairs, they estimate a function that maps any input to any output such that disagreement with future input-output pairs is minimized. Supervised learning is usually subdivided in *classification problems*, where given a set of samples and their class label (training set), the task of the learning algorithm is to correctly predict the class of new unlabelled observations, and *regression problems* where the output is a set of real numbers instead of class labels. Examples of supervised learning algorithms are: *Linear and Logistic Regression*, *Regression Trees*, *k-Nearest Neighbor* algorithms (k-NN) and *Support Vector Machines* (SVM) [54].

3.1.2. Feature Selection

Feature selection is one of the most important topics in classification which aims at reducing the dimensionality of the variable space by filtering out non-informative variables whose presence produces noise in the dataset. It is particularly relevant in the context of microarray data analysis, where most microarray studies explore datasets with hundreds to tens of thousand of variables given only a small number of samples. Such an experimental configuration entails the well known *curse of dimensionality* *problem* which is the term given to the situation of having too many features in a model but not enough instances to completely describe the target concept [54]. In the context of microarray experiments, the investigator faces the challenge of extracting informative features out of thousands of variables (genes) given only a small number of samples (arrays). Further discussion on the curse of dimensionality in microarray data can be found in [*55*].

There exist two main categories of dimensionality reduction methods: *filtering* algorithms and *wrapper* algorithms [*56*]. Filtering methods are preprocessing algorithms that attempt to assess the merits of the features from the data, usually by directly correlating individual features with the class label vector or by statistically comparing feature distributions corresponding to different classes, ignoring the effects of the selected subset on the performance of the learning algorithm. These methods usually result in ranked lists of features based on statistical scoring or information theoretic criteria. Further discussion and a comparison of the performance of several feature selection methods for microarray data can be found in [57].

Alternatively, wrapper methods assess subsets of variables according to their usefulness to a given predictor. They conduct a search for best subsets of features in the variable space using the learning algorithm itself as a part of the evaluation function. If a wrapper method performs exhaustive search then it has to explore 2^n cases, n being the number of features. Such a procedure is extremely computationally intensive and is known to be *NP-hard* [*58*]. Nevertheless, there exist certain strategies that can reduce the computational burden, two popular among them being the *Forward Selection* and *Backward Elimination*: in Forward Selection, the search starts from an empty set and consequently, features are added in the set, based on the minimization of the generalization error produced by the combination of these features. The search continues until no possible improvement can be done. In Backward Elimination the search is performed in the opposite direction of the forward approach. The initial set contains all the n features and the first feature to be removed is the one that allows the highest reduction of the generalization error. The process is continued until no possible improvement is observed. The most important advantage of the wrapper methods is that they take into account and offset the bias of the classifier but with a heavy computational cost. A review of wrapper methods can be found in [*59*].

There is not a clear distinction about which feature selection methods perform better. On one hand, filtering methods do not take into account the bias of the classifier; as a result, different classifiers may demonstrate considerable differences in performance depending on the filter applied for dimensionality reduction [60]. On the other hand, wrappers are usually much more computationally intensive. A good strategy [61] would be to carry out the feature selection process in two steps: firstly, a filtering method should be applied to reduce the number of features; secondly, a wrapper method could be employed to isolate the “best of the best” features.

3.1.3. Examples of Feature Selection and Classification methods in microarray studies

Although the detailed mathematical description of feature selection and classification algorithms is pretty wide and complex, nevertheless it would be of benefit for the reader to illustrate some examples of these techniques, consicely. Thus, the next three subsections provide a brief description of the Principal Component Analysis method for feature selection, the Gene Shaving clustering and the k-Nearest Neighbours supervised classification algorithms.

Principal Component Analysis

Principal components analysis (PCA) [62] is a statistical technique for determining the key variables in a multidimensional data set that explain the differences in the observations, and can be used to simplify the analysis and visualization of multidimensional data sets. In the case of microarray data analysis, where it is often not clear whether a set of experimental conditions are measuring fundamentally different gene expression states or similar states created through different mechanisms [63], PCA can provide useful information on establishing a core set of genes, out of the thousands whose expression is measured, which is able to distinguish among different experimental conditions.

Given m observations on n variables on an m by n data matrix, the goal of PCA is to reduce the dimensionality of the data matrix by finding r new variables, where r is less than n. These r new variables are termed Principal Components and together account for a fixed significant according to specific criteria percentage with respect to the total variance yielded by the original n variables, while remaining mutually uncorrelated and orthogonal. In microarray data analysis, the genes can be considered as variables and the conditions as observations, the opposite or even both [64]. When genes are variables, the analysis creates a set of "principal gene components" that indicate the features of genes that best explain the experimental responses they produce. When conditions are the variables, the analysis creates a set of "principal condition components" that indicates the features of the experimental conditions that best explain the gene behaviours they elicit. When both experiments and genes are analyzed together, there is a combination of these effects, the utility of which remains to be explored [63].

PCA has been used in a wide range of biomedical problems, including the analysis of microarray data in search of outlier genes [65] as well as in the optimization of clustering techniques for microarray gene expression data [66] and classification of microarray data [67]

Gene Shaving

The statistical method called "gene shaving" [49] identifies subsets of genes with coherent expression patterns and large variation across conditions. It differs from hierarchical clustering and other widely used gene expression analysis methods in that genes may belong to more than one cluster and the clustering may be supervised by an outcome measure. The technique can be 'unsupervised', that is, the genes and samples are treated as unlabeled, or partially or fully supervised by using known properties of the genes or samples to assist in finding meaningful groupings. Gene shaving uses a standardized expression matrix where genes are represented as rows and samples (arrays) as columns. It performs an iterative PCA process on genes (rows) to obtain nested gene clusters and then uses a statistic to determine the optimum cluster size. The gene shaving algorithm can be summarized as follows:

1. The leading principal component g_c is obtained out of the expression matrix of the experiment and the remaining genes are sorted according to their absolute correlation to g_c in ascending order.
2. A proportion α of genes with the lowest correlation to g_c is removed (shaved off). Steps 1 and 2 are repeated until only one gene remains.

3. The process in steps 1 and 2 produces a nested sequence of gene clusters with different cluster sizes. The optimal cluster size is estimated according to an appropriate statistic (*gap* statistic, [49]).
4. The above steps are repeated to find *M* maximum optimal clusters where *M* is chosen *a priori*.

k-Nearest Neighbours (k-NN) classification

The k-NN classification is a very simple, yet powerful classification method. The key idea behind k-NN is that similar observations belong to similar classes. Thus, one simply has to look for the class designators of a certain number of the nearest neighbors and weight their class numbers to assign a class number to the unknown instance (Figure 4). The weighting scheme of the class numbers is often a majority rule, but other schemes are conceivable. The class assessment is usually performed according to the distance of the new instant from the majority of the instances that belong to different classes. The distance estimation is carried out using an appropriate metric. The number of the nearest neighbors, *k*, should be odd in order to avoid ties, and it should be kept small, since a large k tends to create misclassifications unless the individual classes are well-separated. One of the major drawbacks of k-NN classifiers is that the classifier needs all available data. This may lead to considerable overhead, if the training data set is large.

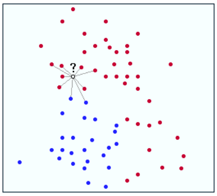


Fig. 4. Two distinct clouds of data. Spot with question mark needs to be classified

The k-NN algorithm, unlike other classification algorithm that are considered more powerful (e.g. SVMs), can be naturally expanded to treat multiclass problems since the only difference is that there are more than two data clouds in the data space and the k-NN is challenged to classify new data between more than 2 classes.

3.2. Pathway Analysis

There are currently several algorithms and software tools which aspire to conduct meaningful microarray experiment based pathway analysis. Some of these tools offer a simple aid by visualizing pathways and correlating the respective genes of these pathways with annotations referring to their assumed functional role while other use sophisticated statistical or graph-theoretical techniques, to mine important biological functions at a hierarchical fashion.

Pathway analysis tools could be subdivided into three main categories: (i) visualization tools (ii) statistical analysis tools which use a short gene list containing the top ranked genes by a statistical score and (iii) statistical analysis tools which consider the distribution of the differentially expressed genes in the entire set of genes. The tools belonging to category (i) offer specialized visualization and pathway editing flexibility (features usually not available in categories (ii) and (iii)) without going further in statistical ranking with respect to important functional features, leaving this decision up to the user of the tool. The tools belonging to category (ii) utilize a short list of specified fixed percentage of the total selected gene list, containing the most expressed genes which are examined against predefined sets of genes representing different pathways, to determine whether any set is overrepresented in the short list compared to the whole list [68]. Several analysis packages exist for this category. A comparison of 14 such tools can be found in [69]. The tools belonging to category (iii) consider all genes in an experiment regardless their statistical score, along with predefined gene sets that correspond to specific biological processes; such sets are usually derived from knowledge sources such as Gene Ontology repository [70] for instance, or can even be manually curated. For each gene set, the distribution of gene ranks of the gene set is compared against the distribution of the rest of the genes by using proper statistical hypothesis tests like Kolmogorov-Smirnov or Anderson-Darling [32] whereas thereinafter resampling (permutation/bootstrap) methods are applied to check the significance of the derived statistical score estimation of null distributions.

The methods utilized by the tools in category (ii) are rational but present at least three drawbacks [68]:

- Only the top significant part of the statistically scored gene list is used, treating the less relevant genes as irrelevant leading to possible information loss.
- The order of genes on the significant gene list is not taken into consideration.
- The correlation structure of gene sets which is an important aspect to consider in assessing statistical significance is not considered at all.

However, using the whole gene list for the discovering of same functionality sharing among genes may conceal some danger from the point that the nature of microarray technology is subjective to noise and the measurements are based on light emissions which depend on many fluctuated factors.

Some of the tools that implement analysis algorithms belonging to the three general categories described above are cited next with a brief description of their functionality.

3.2.1. Visualization Tools (category (i))

GennMAPP

GennMAPP [71] is a biological pathway visualization tool. It offers aid by mapping gene expression data on pathways and colour-coding the genes according to investigator's criteria. It also provides annotation data and tools for computational design and modification of biological pathways.

PathwayAssist

PathwayAssist [72] is a software application for navigation and analysis of biological pathways, gene regulation networks and protein interaction maps. It comes with a built-in database describing more than 100000 events of regulation, interaction and modification between proteins, cell processes and small molecules. It mostly offers visualization and pathway edit and construction features.

3.2.2. Class Over-representation Analysis Tools (category (ii))

MAPPFinder

MAPPFinder [73] is a visualization and analysis tool that dynamically links gene expression data to the GO hierarchy. For each GO entity it calculates the percentage of the genes measured that meet a user-defined criterion (e.g. fold change cutoffs, p-value thresholds). Then, by using this percentage, the association of groups of genes to GO parental nodes and a simple statistical score it ranks the GO terms by their relative amount of gene expression changes generating a profile in functional levels. The results of MAPPFinder can be integrated to GennMAPP for further analysis.

GOstat

GOstat [74] is a web-based tool for the statistical analysis of GO terms, utilizing a group of genes of interest (usually the DE gene list) and a control set of genes which is used to obtain a total count of occurrences for each GO term (usually the gene list of the whole array). The significance of GO terms is assessed through the X^2 test or the Fisher's exact test. GOstat offers visualization abilities through AmiGO, a tool for the GO structure visualization in the Gene Ontology database.

ErmineJ

ErmineJ [75] is an application which, apart from the gene class over-representation analysis implemented in previous tools, supports also gene expression data pathway analysis using statistical learning methods [54] and proper scoring for the association of genes to functional annotation classes (usually using GO). The main rational supporting this method is that the genes must not only cluster together in space (i.e., be co-expressed), but also be sufficiently distinct from other genes in the data set to be distinguishable as a class [76].

GOALIE

GOALIE [77] is an application that uses clustered numerical microarray gene expression value measurements mostly from time-course data to reconstruct the cluster analysis based on time windows, proper statistical measures and temporal logic methods. It results in a temporal logic model connecting gene functionalities across different time points and annotating this structure using GO.

3.2.3. Predefined Gene Set Analysis Tools (category (iii))

GSEA

GSEA [45] is an application which uses a ranked list of genes (usually the result of a microarray experiment) depicting the whole array instead of the top scoring portion of it, together with predefined gene sets (based on prior biological knowledge, published information, GO etc.) and sample phenotypic labels to determine whether members of a gene set tend to occur toward the top (or bottom) of the gene list, in which case the gene set is correlated with the phenotypic class distinction. The goal of the algorithm behind GSEA is to determine whether the members of the gene set tend to be randomly distributed in the gene list or primarily found at the top or bottom. This is achieved through the use of an enrichment score similar to a Kolmogorov-Smirnov statistic which is calculated for each gene set comparing its distribution to that of the gene list. The significance of this score is assessed through the estimation of a null distribution using resampling (non-parametric) methods across samples in order to calculate empirical p-values. GSEA is considered one of the most successful meta-analysis methods so far [5], [78].

PAGE

PAGE [79] is an algorithm very similar to GSEA but it uses parametric instead of non-parametric statistics: the enrichment score is replaced by a z-score from a parameter such as fold change value between two experimental groups and the empirical null distribution is replaced by the normal distribution. Because of the use of parametric statistics, PAGE turns to be less computationally intensive than GSEA while sometimes giving better results.

References

[1] Quackenbush, J., Microarray Data Normalization and Transformation, *Nat. Genetics* **32** (2002) 496-501.

[2] Quackenbush J., Microarray Analysis and Tumor Classification, *N. Engl. J. Med.* **354** (2006) 2463-2472.

[3] Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lonning, P.E., Brown, P.O., Borresen-Dale, A.L., Botstein, D., Repeated Observation of Breast Tumor Subtypes in Independent Gene Expression Data Sets, *Proc. Nat. Acad. Sci.* **100** (2003) 8418-8423.

[4] Mariadason, J.M., Arango, D., Shi, Q., Wilson, A.J., Corner, G.A., Nicholas, C., Aranes, M.J., Lesser, M., Schwartz, E.L., Augenlicht, L.H., Gene Expression Profiling-Based Prediction of Response of Colon Carcinoma Cells to 5-Fluorouracil and Camptothecin, *Cancer Res.* **63** (2003) 8791-8812.

[5] Allison, D.B., Cui, X., Page, G.P., Sabripour, M., Microarray Data Analysis: from Disarray to Consolidation and Consensus, *Nat. Rev. Genet.* **7** (2006) 55-65.

[6] Babu, M., Introduction to Microarray Data Analysis, *Computational Genomics: Theory and Applications*, (Ed: R. Grant) Horizon Press (2004).

[7] Rosenweig, B.A., Pine, P.S., Domon, E.O., Morris, S.M., Chen, J.J., Sistare, F., Dye-Bias Correction in Dual–Labeled cDNA Microarray Gene Expression Measurements, *Environ. Health Perspect.* **112** (2004) 480-487.

[8] Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lonning, P.E., Borresen-Dale, A.L., Brown, P.O., Botstein, D., Molecular Portraits of Human Breast Tumours. *Nature* **406** (2000) 747-752.

[9] Gibson, G., Muse, S.V., A Primer of Genome Science, Sinauer Associates, Inc. Sunderland, MA (2002).

[10] Knudsen, S., Guide to Analysis of DNA Microarray Data. (Ed: N.J. Hoboken) John Wiley & Sons, Inc., (2004).

[11] Tarca, A.L., Romero, R., Draghici, S., Analysis of Microarray Experiments of Gene Expression Profiling, *American Journal of Obstetrics and Gynecology* 195 (2006) 373-388.

[12] Speed, T., Statistical Analysis of Gene Expression Microarray Data, Chapman & Hall/CRC, (2003).

[13] Chatziioannou, A., Moulos, P., Kolisis, F., Aidinis, V., ANDROMEDA: a Pipeline for Versatile 2-Colour cDNA Microarray Data Analysis Implemented in MATLAB, (2007) submitted.

[14] Bilban, M., Buehler, L.K., Head, S., Desoye, G., Quaranta, V., Normalizing DNA Microarray Data, *Curr. Issues Mol. Biol.* **4** (2002) 57-64.

[15] Tseng, G.C., Oh, M.K., Rohlin, L., Liao, J.C., Wong, W.H., Issues in cDNA Microarray Analysis: Quality Filtering, Channel Normalization, Models of Variations and Assessment of Gene Effects, *Nucleic Acids Research* **29** (2001) 2549-2557.

[16] Cleveland, W.S., Grosse, E., Shyu, W.M.: Local Regression Models, *Statistical Models in S*, (Eds: J.M. Chambers, T.J. Hastie), Wadsworth & Brooks/Cole Dormand, J.R. (1992).

[17] Hoffmann, R., Seidl, T., Dugas, M., Profound Effect of Normalization on Detection of Differentially Expressed Genes in Oligonucleotide Microarray Data Analysis, *Genome Biol.* **3** (2002) RESEARCH0033.
[18] Quackenbush, J., Computational Analysis of Microarray Data, *Nat. Rev. Genet.* **2** (2001) 418-427.
[19] Finkelstein, D.B., Ewing, R., Gollub, J., Sterky, F., Somerville, S., Cherry, J.M., Iterative Linear Regression by Sector, *Methods of Microarray Data Analysis*, (Eds. S.M. Lin, K.F. Johnson), Cambridge, MA: Kluwer Academic (2002) 57-68.
[20] Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Earle– Hughes, J., Snesrud, E., Lee, N., Quackenbush, J., A Concise Guide to cDNA Microarray Analysis, *Biotechniques* **29** (2000) 548-562.
[21] Durbin, B.P., Hardin, J.S., Hawkins, D.M., Rocke, D.M.: A Variance Stabilizing Transformation for Gene-Expression Microarray Data, *Bioinformatics* **18** (2002) S105-S110.
[22] Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., Speed, T.P., Normalization for cDNA Microarray Data: a Robust Composite Method Addressing Single and Multiple Slide Systematic Variation, *Nucleic Acids Research* **30** (2002) e15.
[23] Cui, X., Kerr, M.K., Churchill, G.A., Transformations for cDNA Microarray Data, *Stat. Appl. Genet. Mol. Biol.* **2** (2003) Article4.
[24] Wu, Z., Irizarry, R., Gentleman, R.C., Murillo, F.M., Spencer, F., A Model Based Background Adjustment for Oligonucleotide Expression Arrays, *Collection of Biostatistics Research Archive* (2004) Article1.
[25] Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., Speed, T.P., Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data, *Biostatistics* **4** (2003) 249-264.
[26] Affymetrix, Statistical Algorithms Description Document, Affymetrix, Inc., (2002).
[27] Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B., Missing Value Estimation Methods for DNA Microarrays, *Bioinformatics* **17** (2001) 520-525.
[28] Nguyen, D.V., Wang, N., Carrol, R.J., Evaluation of Missing Value Estimation for Microarray Data, *Journal of Data Science* **2** (2004) 347-370.
[29] Dudoit, S., Yang, Y.H., Speed, T., Callow, M.J., Statistical Methods for Identifying Differentially Expressed Genes in Replicated cDNA Microarray Experiments, *Statistica Sinica* **12** (2002) 111-139.
[30] Kerr, M.K., Martin, M., Churchill, G.A., Analysis of Variance for Gene Expression Microarray Data, *J. Computational Biol.* **7** (2000), 819-837.
[31] Ideker, T., Thorsson, V., Siehel, A.F., Hood, L.E.: Testing for Differentially Expressed Genes by Maximum Likelihood Analysis of Microarray Data. J. Comput. Biol. 7 (2000) 805-817.
[32] Conover, W.J., Practical Nonparametric Statistics, Wiley (1980).
[33] Tusher, V.G., Tibshirani, R., Chum G.: Significance Analysis of Microarrays Applied to the Ionizing Radiation Response. Proc. Nat. Acad. Sci. 98 (2001) 5116-5121.
[34] Pan, W., A Comparative Review of Statistical Methods for Discovering Differentially Expressed Genes in Replicated Microarray Experiments. Bioinformatics 18 (2002) 546-554.
[35] Kim, S.Y., Lee, J.W., Sohn, I.S., Comparison of Various Statistical Methods for Identifying Differential Gene Expression in Replicated Microarray Data, Statistical Methods in Medical Research 15 (2006) 3-20.
[36] Yauk, C.L., Berndt, M.L., Williams, A., Douglas, G.R., Comprehensive Comparison of Six Microarray Technologies, *Nucleic Acids Research* 32 (2004) e124.
[37] Canales, R.D., Luo, Y., Willey, J.C., Austermiller, B., Barbacioru, C.C., Boysen, C., Hunkapiller, K., Jensen, R.D., Knight, C.R., Lee, K.Y., Ma, Y., Maqsodi, B., Papallo, A., Peters, E.H., Poulter, K., Ruppel, P.L., Samaha, R.R., Shi, L., Yang, W., Zhang, L., Goodsaid, F.M., Evaluation of cDNA Microarray Results with Quantitative Gene Expression Platforms, *Nat. Biotech.* 24 (2006) 1115-1122.
[38] Guo, L., Lobenhofer, E.K., Wang, C., Shippy, R., Harris, S.C., Zhang, L., Mei, N.,Chen, T., Herman, D., Goodsaid, F.M., Hurban, P., Phillips, K.L., Xu, J.,Deng, X., Sun, Y.A., Tong, W., Dragan, Y.P., Shi, L., Rat Toxigonecomic Study Reveals Analytical Consistency Across Microarray Platforms, Nat. Biotech. 24 (2006) 1162-1169.
[39] Dudoit, S., Shaffer, J.P., Boldrick, J.C., Multiple Hypothesis Testing in Microarray Experiments, *Statistical Science* **18** (2003) 71-103.
[40] Benjamini, Y., Hochberg, Y., Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing, *J. R. Statist. Soc.* **57** (1995) 289-300.
[41] Storey, J.D., Tibshirani, R., Statistical significance for genomewide studies, *Proc. Nat. Acad. Sci.* **100** (2003) 9440-9445.
[42] Gentleman, R., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B.Gautier, L., Ge, Y., Gentry, J., Hornik, K.,Hothorn, T., Huber, W., Iacus, S., Irizarry,R.,Leisch, F., Li, C., Maechler, M., Rossini, A.J.,Sawitzki, G., Smith, C., Smyth, G., Tierney, L.,Yang, J.Y.H., Zhang, J., Bioconductor:

Open Software Development for Computational Biology and Bioinformatics, *Genome Biology* **5** (2004) R80.
[43] Saeed A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., Quackenbush, J., TM4: a free, open-source system for microarray data management and analysis, *Biotechniques* **34** (2003) 374-378.
[44] Chatziioannou, A., Moulos, P., Kolisis, F., Aidinis, V., ANDROMEDA: a Pipeline for Versatile 2-colour cDNA Microarray Data Analysis Implemented in MATLAB, submitted (2007).
[45] Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P.: Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proc. Nat. Acad. Sci.* 102 (2005) 15545-15550.
[46] Fortunel, N.O., Otu, H.H., Ng, H.H., Chen, J., Mu, X., Chevassut, T., Li, X., Joseph, M., Bailey, C., Hatzfeld, J. A., et al., Comment on "'Stemness': Transcriptional Profiling of Embryonic and Adult Stem Cells" and "A Stem Cell Molecular Signature", *Science* **302** (2003) 393.
[47] Newton, J., Analysis of Microarray Gene Expression Data Using Machine Learning Techniques, Technical Report, University of Alberta, Canada, (2002).
[48] Kaufman, L., Rousseeuw, P.J., Finding Groups in Data: An Introduction to Cluster Analysis, New York, Wiley (1990).
[49] Hastie, T., Tibshirani, R., Eisen, M.B., Alizadeh, A., Levy, R., Staudt, L., Chan, W.C., Botstein, D., Brown, P.: 'Gene Shaving' as a Method for Identifying Distinct Sets of Genes with Similar Expression Patterns, *Genome Biology* **1** (2002) research0003.1-0003.21.
[50] Kohonen, T., Self-Organizing Maps, Springer, Berlin (1995).
[51] Torkkola, K., Gardner, R.M., Kaysser-Kranich, T., Ma, C., Self-Organizing Maps in Mining Gene Expression Data, *Information Sciences* **139** (2001) 79-96.
[52] Fu, L., Medico, E., FLAME, a Novel Fuzzy Clustering Method for the Analysis of DNA Microarray Data, BMC Bioinformatics 8 (2007) 3.
[53] Dobbin, K., Simon, R., Comparison of Microarray Designs for Class Comparison and Class Discovery, *Bioinformatics* **18** (2002) 1438-1445.
[54] Hastie, T., Tibshirani, R., Friedman, J., Elements of Statistical Learning, Springer-Verlag (2001).
[55] Simon, R., Supervised Analysis When the Number of Candidate Features (p) Greatly Exceeds The Number of Cases (n), *ACM SIGKDD Explorations Newsletter* **5** (2003) 3-36.
[56] Guyon, I., Elisseeff, A.: An Introduction to Variable and Feature Selection, *Journal of Machine Learning Research* **3** (2003) 1157-1182.
[57] Li, T., Zhang, C., Ogihara, M.: A Comparative Study of Feature Selection and Multiclass Classification Methods for Tissue Classification Based on Gene Expression, *Bioinformatics* **20** (2005) 2429-2437.
[58] Amaldi, E., Kann, V., On the Approximation of Minimizing non-zero Variables or Unsatisfied Relations in Linear Systems, *Theoretical Computer Science* **209** (1998) 237-260.
[59] Kohavi, R., John, G.H.: Wrappers for Feature Subset Selection, *Artificial Intelligence* **97** (1997) 273-324.
[60] Liu, H., Jinyan, L., Limsoon, W., A Comparative Study on Feature Selection and Classification Methods Using Gene Expression Profiles and Proteomic Patterns, *Genome Informatics* **13** (2002) 51-60.
[61] Xing, E.P., Jordan, M.I., Karp, R.M., Feature Selection for High-Dimensional Genomic Microarray Data, *Proceedings of the 18th International Conference on Machine Learning* (2001) 601-608.
[62] Jolliffe, I.T.: Principal Component Analysis, Springer (2002).
[63] Raychaudhuri, S., Stuart, J.M., Altman, R.B.: Principal Component Analysis to Summarize Microarray Experiments: Application to Sporulation Time Series, *Pac. Symp. Biocomput.* **5** (2000) 452-463.
[64] Parmigiani, G., Garett, E.S., Irizarry, R.A., Zeger, S.L., The Analysis of Gene Expression Data: Methods and Software, Springer (2003).
[65] Hilsenbeck, S.G., Friedrichs, W.E., Schiff, R., O'Connell, P., Hansen, R.K., Osborne, C.K., Fuqua, S.A.W.: Statistical Analysis of Array Expression Data as Applied to the Problem of Tamoxifen Resistance, *J. Natl. Cancer Institute* **91** (1999) 453-459.
[66] Yeung, K.W., Ruzzo, W.L.: Principal Component Analysis for Clustering Gene Expression Data, *Bioinformatics* **17** (2001) 763-774.
[67] Liu, A., Zhang, Y., Gehan, E., Clarke, R: Block Principal Component Analysis with Application to Gene Microarray Data Classification, *Stat. Med.* **21** (2002) 3465-3474.
[68] Tian, L., Greenberg, S.A., Kong, S.K., Altschuler, J., Kohane, S.A., Park, P.J.: Discovering Statistically Significant Pathways in Expression Profiling Studies, *Proc. Nat. Acad. Sci.* **102** (2005) 13544-13549.
[69] Khatri, P., Drăghici, S.: Ontological Analysis of Gene Expression Data: Current Tools, Limitations and Open Problems, *Bioinformatics* **21** (2005) 3587-3595.

[70] The Gene Ontology Consortium, Gene Ontology: Tool for the Unification of Biology, *Nature Genet*. **25** (2000) 25-29.

[71] Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C., Conklin, B.R.: GenMAPP, a New Tool for Viewing and Analyzing Microarray Data on Biological Pathways, *Nat. Genet.* **31** (2002) 19-20.

[72] Nikitin, A., Egorov, S., Daraselia, N., Mazo, I.: Pathway Studio-The Analysis and Navigation of Molecular Networks, *Bioinformatics* **19** (2003) 2155-2157.

[73] Doniger, S.W., Salomonis, N., Dahlquist, K.M.,Vranizan, K., Lawlor, S.C., Conklin, B.R., MAPPFinder: using Gene Ontology and GenMAPP to Create a Global Gene Expression Profile from Microarray Data, *Genome Biology* **4** (2003) R:7.

[74] Beißbarth, T., Speed, T.P., GOstat: Find Statistically Overrepresented Gene Ontologies Within a Group of Genes, *Bioinformatics* **20** (2004) 1464-1465.

[75] Lee, H.K., Braynen, W., Keshav, K., Pavlidis, P., ErmineJ: Tool for Functional Analysis of Gene Expression Data Sets, *BMC Bioinformatics* **6** (2005) 269.

[76] Pavlidis, P., Lewis, D.P., Noble, W.S.: Exploring Gene Expression Data with Class Score, *Pac. Symp. Biocomput.* **7** (2002) 474-485.

[77] Ramakrishnan, N., Antoniotti, M., Mishra, B.: Reconstructing Formal Temporal Models of Cellular Events using the GO Process Ontology, *Bio-Ontologies SIG Meeting, ISMB 2005* Detroit, U.S.A. (2005).

[78] Bild, A., Potti, A., Nevins, J.R.: Linking Oncogenic Pathways with Therapeutic Opportunities, *Nat. Rev. Cancer* **6** (2006) 735-741.

[79] Kim, S.Y., Volsky, D.J.: PAGE: Parametric Analysis of Gene Set Enrichment, *BMC Bioinformatics* **6** (2005) 144.

[80] Argyropoulos, C.,, Chatziioannou, A.A, Nikiforidis G.,, Moustakas A.,, Kollias, G., and Aidinis, V. Operational criteria for selecting a cDNA microarray data normalization algorithm. *Oncology reports.* ***15*** *Spec no.**4***, (2006) 983-996.

[81] Hoffmann R, Seidl T, Dugas M (2002). Profound effect of normalization on detection of differentially expressed genes in oligonucleotide microarray data analysis. *Genome Biol **3***:RESEARCH0033.

I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Part IV: Real world AI applications in Computer Engineering

John Soldatos
Athens Information Technology
jsol@ait.edu.gr

The fourth part of the book includes contributed chapters emphasizing on real world AI applications spanning a variety of domains. Specifically, the following chapters present applications falling under the domains of telecommunications and networking, multimedia applications, pervasive computing and surveillance applications, as well as web engineering applications. Along with different application domains, the presented applications manifest also the wealth of AI approaches to building applications, which span the areas of machine learning, pattern recognition, intelligent data analysis, probabilistic reasoning, neural networks and genetic algorithms.

In the application field of telecommunications and networking, three distinct chapters are included spanning the areas of intrusion detection and wireless networking. Two chapters emphasize on approaches to intrusion detection, while another deals with smart antennas for in-door environments. In their contributed chapter 'Using Artificial Intelligence for Intrusion Detection' François Gagnon and Babak Esfandiari emphasize on knowledge representation, machine learning, and multi-agent architectures for intrusion detection in computer networks. The chapter discusses different AI approaches for intrusion detection, namely knowledge representation and machine learning for signature-based and anomaly-based intrusion detection systems (IDS) respectively. They also illustrate how a multi-agent architecture can be used to implement IDS systems. Furthermore, they provide pointers to other AI based network security projects and related reading material.

The second chapter on AI for Intrusion Detection is co-authored by Stefanos Koutsoutos, Ioannis Christou and Sofoklis Efremidis and titled 'A Classifier Ensemble Approach to Intrusion Detection for Network-Initiated Attacks'. This chapter illustrates a different AI approach for Intrusion Detection, which relies on a classifier ensemble system that uses a combination of Neural Networks and rule-based systems as base classifiers. This system is capable of detecting network-initiated intrusion attacks on web servers, while at the same time being able to recognize novel attacks (i.e., abnormal situations and attacks never seen before). The authors evaluate their approach and conclude that it is very efficient for identifying new attacks, yet it suffers from non-negligible rates of false alarms.

The collection of AI applications in networking is complemented by the chapter 'Prediction Models of an Indoor Smart Antenna System using Artificial Neural Networks', co-authored by Nektarios Moraitis and Demosthenes Vouyioukas. This chapter focuses on a method for predicting propagation paths of angle of arrivals (AoAs) of a Smart Antenna System in indoor environments. The method makes use of Artificial Neural Networks (ANN), based on a Multilayer Perceptron and a Generalized Regression Neural Network trained with measurements. In evaluating their approach the authors compare their approach with the theoretical Gaussian scatter density model for the derivation of the power angle profile. The results of the method are evaluated in terms of average error, standard deviation and mean square error and show decent accuracy. Specifically, it is shown that the Gaussian model provides greater errors because it takes into account a smaller range of azimuth angle. This is due to the fact that NNs are trained with measurements inside buildings and include realistic propagation effects, which cannot be taken into account within analytic equations.

Following the presentation of AI applications in the area of networking, there is a collection of three chapters emphasizing on multimedia applications. The first of these chapters presents techniques devised within the European Commission co-funded AXMEDIS (Automating Production of Cross Media Content for Multi-channel Distribution) project for interoperable cross media content and digital rights management (DRM) for content distribution over multiple channels. In particular, the chapter is titled 'Interoperable cross media content and DRM for multichannel distribution' and co-authored by Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, Davide Rogai and Paolo Vaccari. This chapter's contributions are developed and presented across the following axes: a generic Digital Rights Management Scenario for content distribution, challenges and techniques for the interoperability of the cross media protected content, the AXMEDIS content processing GRID platform, as well as the development of solutions for interoperable multi-channel distribution.

The second chapter on multimedia applications, co-authored by Sofia Tsekeridou and titled 'Video Watermarking and Benchmarking', constitutes a comprehensive overview of techniques for video watermarking and benchmarking including AI techniques. Following an in-depth review of watermarking and benchmarking techniques, emphasis is put on illustrating how AI techniques can tackle with the existing challenges through overcoming the limitations of conventional methodologies (e.g., their inability to tackle with real-time and three dimensional constraints).

The last chapter of the multimedia applications and image processing series is titled 'Portrait identification in digitized paintings on the basis of a face detection system '. The authors, Christos – Nikolaos Anagnostopoulos, Ioannis Anagnostopoulos, I. Maglogiannis and D. Vergados, propose the use of a Probabilistic Neural Network (PNN) approach for automatic identification of portraits in paintings collections. The PNN is trained for face identification and used for detecting skin areas in color images. The evaluation of the system based on digitized paintings from the State Heritage Museum reveals success rates exceeding 88%.

Three chapters of the fourth part of the book are devoted to AI applications in the more specific field of pervasive computing. These chapters emphasize on applications that make use of real sensors and AI techniques for ubiquitous and autonomous processing of sensor streams. In the first of these chapters titled 'Where and Who? Person Tracking and Recognition System', Aristodemos Pnevmatikakis introduces a novel system for far-field unconstrained person tracking and video-to-video recognition, which can be used for non-obtrusive identification of humans and their location in smart spaces. This system goes far beyond conventional AI techniques, which are used in typical security and access control applications. Specifically, the presented pattern matching and classification techniques are appropriate for dealing with far-field recordings where people are recorded from far-away and act naturally (e.g., they do not have to pose).

The second chapter of the pervasive computing collection by Josep R. Casas and Joachim Neumann is titled 'Context Awareness triggered by Multiple Perceptual

Analyzers' and presents work conducted in the context of the European Commission co-funded CHIL (Computers in the Human Interaction Loop) project. The chapter discusses techniques for integrating a multitude of technologies from computer vision, acoustic signal analysis and natural language processing towards implementing multi-modal perceptual components. Therefore this work lies at a higher integration layer comparing to the recognition and tracking system presented in the previous chapter. The chapter ends-up presented a context-aware application (i.e. the 'Memory Jog'), which relies on the identification of a multitude of contextual-states.

Moving beyond smart spaces and perceptual processing algorithms, the third chapter of the pervasive computing collection presents a specific robotic sensor network devoted to monitoring electro-magnetic fields (EMF). The chapter is titled 'Robotic Sensor Networks: An Application to Monitoring Electro-Magnetic Fields' and is co-authored by Francesco Amigoni, Giulio Fontana, and Stefano Mazzuca. The chapter focuses on robotic sensor networks, in which mobile robots carry sensors around an environment to detect phenomena and produce detailed environmental assessments. An implementation of a specific robotic sensor network is presented. This implementation hinges on the following three-step cycle: using explorers to measure the EMF in their current positions, integrating the measurements and building an hypothesis about the location of the EMF sources, and move explorers in the environment to reach the new interesting measurement positions.

The last part of the book's section on Real world AI applications in Computer Engineering, presents an AI application in Web engineering. Specifically, the chapter 'Assembling Composite Web Services from Autonomous Components', written by Jyotishman Pathak, Samik Basu and Vasant Honavar focuses on methodologies and tools that enable (semi-) automatic composition of services by taking into account the functional, non-functional and behavioral requirements of the service developer. The chapter focuses on key concepts and issues associated with service composition. Furthermore it provides a representative collection of AI planning techniques, as well as formal methods addressing the challenges of service composition. The chapter introduces also an iterative and incremental technique for modeling composite Web services.

Overall the presented applications are expected to provide readers with a 360 degree view of the use and potential of AI techniques for non-trivial realistic computer engineering applications. Last but not least, it should be noted that the chapters achieve a healthy balance between novel research approaches, review works, applications and case studies. This mix ensures a multidisciplinary coverage of the field, which at the same time providing ideas and opportunities for further reading and research.

This page intentionally left blank

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Using Artificial Intelligence for Intrusion Detection

François Gagnon and Babak Esfandiari
Network Management and Artificial Intelligence Laboratory
Carleton University, Canada

Abstract. Artificial intelligence is playing an increasingly important role in network management. In particular, research in the area of intrusion detection relies extensively on AI techniques to design, implement, and enhance security monitoring systems. This chapter discusses ways in which intrusion detection uses or could use AI. Some general ideas are presented and some actual systems are discussed. The focus is mainly on knowledge representation, machine learning, and multi-agent architectures.

Keywords. Intrusion Detection, Context, Knowledge Representation, Machine Learning, Multi-Agents

Introduction

Intrusion detection is the subarea of network management concerned with monitoring computer networks for possible malicious intrusions. An intrusion can be defined as any set of actions that attempts to compromise the integrity, confidentiality or availability of a resource [8]. To achieve their goal, intrusion detection systems (IDS) may rely on many different AI concepts, the most prominent being knowledge representation, machine learning and multi-agent architectures.

In this chapter, we describe intrusion detection systems that rely on different aspects of AI. The chapter is structured as follows: Section 1 provides a quick introduction to intrusion detection; Sections 2 and 3 discuss knowledge representation and machine learning in the context of signature-based and anomaly-based IDS, respectively; Section 4 presents IDS built with a multi-agent architecture; finally, a glimpse at some other network security-related projects relying on AI and pointers for further reading will close this chapter.

1. Intrusion Detection Background

An intrusion detection system is a program that aims to detect malicious activities in a specified environment. IDS are not meant to prevent such intrusions, but merely to report them to a human security officer, who will take the necessary measures (often to recover as quickly as possible from an attack that has already occurred). Of course, preventing such attacks would be a lot more useful than simply detecting them, but false alarms would then disrupt the service and inconvenience users. Some systems are able carry out prevention; mostly they are IDS with a module that can be disabled, allowing the malicious content to be stopped (or slowed) from reaching the destination.

IDS are typically built to monitor two different environments: computer networks and individual computers. When an IDS is designed to monitor a network of computers, it is called a network-based IDS, while an IDS monitoring an individual computer is called host-based. Usually, a network-based IDS monitors the environment by analyzing the network traffic between the hosts to see if any malicious content is present in the data exchanged, or if the communication patterns have suddenly become irregular. Most host-based IDS search a computer's audit files to find evidence of malicious behavior (multiple failed login attempts, recent access to critical files, etc) or monitor the execution flow of different user programs to see if their behavior is malicious.

Another important classification of IDS (which is probably the most important design choice) is the strategy it employs to monitor its environment. There are two well-known and widely used strategies: signature-based and anomaly-based IDS. A signature-based IDS (also known as "misuse" IDS) uses a representation of what is considered malicious (usually a list of known attack signatures) and raises an alarm whenever it sees an event matching one of its signatures. An anomaly-based IDS uses the opposite approach; it represents what is considered normal activity for a given environment and raises an alarm whenever an event deviates from the normal model. More details will be provided on signature-based and anomaly-based IDS in Sections 2 and 3 respectively.

2. AI in Signature-based IDS

This section provides a more detailed study of signature-based IDS and examines the AI components on which they rely; mainly, the emphasis is on knowledge representation and machine learning. As will soon become clear, AI is not fundamentally embedded in signature-based IDS, but is instead used in many research projects to enhance their performance.

2.1. Introduction

As mentioned earlier, signature-based IDS aim to detect known attacks by carrying a database of attack signatures. When an event matches a signature, the event is tagged as malicious and an alarm is reported to the security officer. Since most signature-based IDS are network-based, this section concentrates on network-based IDS and, unless stated otherwise, the term "IDS" stands for signature-network-based IDS.

Among the best-known IDS, Snort [21] is signature- and network-based. Snort rules utilize attribute-value requirements for protocol header fields and pattern matching for the string contained in the payload. For instance, a rule could look as follows:

```
msg:"DOS Real Server template.html"; reference:bugtraq,1288;   (1)
alert tcp External any -> Internal 8080                        (2)
content:"/viewsource/template.html?"                           (3)
```

where line 1 specifies the message to give to the security officer and a reference for this attack un the Security Focus vulnerability database [23]. Lines 2 and 3 indicate when the alarm should be raised: the packet uses the TCP protocol, comes from outside the monitoring network from any port to inside the monitored network on port 8080, and its payload contains the string "/viewsource/template.html?" (this is the actual attack signature).

With the discovery of squealing [18], it became evident that context was crucial when detecting attacks. Squealing means sending individual packets crafted to trigger an alarm. Since these packets are not part of a valid TCP communication session, the target will not interpret the packets, and thus the attack will fail. However, since IDS look at each packet independently (without context), an alarm will be triggered, and it is possible to flood the security officer with false alarms (maybe hiding the real attack among 10,000 false attacks). Soon after the publication of squealing, IDS evolved to support the very restricted notion of context associated with valid TCP sessions. Nowadays, Snort rules (like the one presented above) come with the extra condition "flow:to_server,established;" meaning that a valid TCP connection must be established between the attacker and the target.

2.2. *Limitations*

The principal limitation of signature-based IDS is the large amount of false alarms they generate. Even though IDS can achieve pretty low false-positive rates, a 1% FP rate per packet still generates a huge amount of false alarms when the network carries large amounts of traffic; see [1]. Sections 2.3 and 2.4 illustrate how AI can help IDS do a much better job with respect to false positives.

Another major limitation of signature-based IDS is their inability to detect new (never seen before) attacks. As long as the signature of an attack has not been added to the rule database of an IDS, the IDS will not be able to detect it. Thus when a new attack comes out, security analysts must study it and manually extract a pattern representing this attack (the new signature). Then signature databases must be updated. The anomaly-based approach to intrusion detection, discussed in Section 3, was designed especially to address this shortcoming.

2.3. *Knowledge Representation*

This section illustrates how knowledge representation, up to now present in a very primitive form, if not completely absent, can improve the ability of signature-based IDS. Mainly, we will study knowledge representation in two closely related aspects of intrusion detection: context representation and context gathering. As previously mentioned, context enables a reduction in IDS false alarms.

2.3.1. *Context Representation*

A good context representation enables an IDS to use information surrounding an attack attempt to determine the importance (or the likelihood of success) of this attempt. As a case study, we will look in detail at how $chronicles$ [4] can be used to implement context representation in the world of intrusion detection. Basically, a chronicle is a set of events linked together by time constraints, and whose occurrence may depend on the context. Predicate "event(X, t_1)" states that event X must occur at time t_1, while "hold($A : v, t_2, t_3$)" means that the value of attribute A must remain v between t_2 and t_3.

[17] uses chronicles to reduce the number of alarms delivered to the security officer, by clustering many related events into one, and enhance the quality of the alarm messages, by correlating different events to get a better idea of what is going on.

We present an example in which context representation in chronicles is used to eliminate false alarms. Table 1 describes a chronicle that determines if an attack has some chance of succeeding by correlating an alarm with the vulnerability of the target. This chronicle states that if Snort raises an alarm concerning the vulnerability described by bid3335 [23], and if the target of this attack is running the operating system (OS) Linux RH 7.0, then an alarm should be raised. Indeed, vulnerability bid3335 is only present in this particular version of Linux. Thus if the target is running Windows XP, then the chronicle will not match, and no (false) alarm will be given to the security officer.

With a similar chronicle, it would be possible to specify the valid TCP session context discussed previously: replace line 3 with a statement saying that the predicate validTCPSession between the source and the target is true from some point up to time t_2.

A different kind of context representable with chronicles is reaction context. For instance, if Snort raises an alarm of type webAttack, and less than 2 seconds after that, the target of this attack responds with an HTTP error message, then the attack is likely[1] to have succeeded and an alarm should be raised for the security officer.

1	chronicle bid3335 {
2	event(snortAlarm(bid3335,?source,?target), t_2)
3	hold(OS(?target):Linux RH 7.0 (t_1, t_2))
4	$t_1 < t_2$
5	
6	When recognized {
7	emit event(alarm(bid3335,?source,?target),t_2)
8	}
9	}

Table 1. Vulnerability Context Chronicle

2.3.2. Context Gathering

Paramount to context representation is the task of context gathering. Once we are able to represent the context, it is essential to be able to gather it automatically, especially in an environment as dynamic as a computer network. Recall that in the chronicle example of Table 1 we used an attribute-value pair $\langle$target, OS$\rangle$. However, such information is not always available a priori and is volatile. Many tools exist to automatically retrieve the operating system of distant computers, either passively or actively. Unfortunately, passive methods are not sufficiently accurate (mainly because they try to match every

[1]It has been observed that against some successful attacks, some web servers would normally respond with an http error message. However, the rule is not always true, and may depend on many other factors. It is given here as a simplified example.

%ARP Request
os(X,macOS) ∨ os(X,sunOS) :- arp(X,_,1,macFF_FF_FF_FF_FF_FF).
os(X,freeBSD5_0) :- arp(X,_,1,Y), Y != mac00_00_00_00_00_00, Y != macFF_FF_FF_FF_FF_FF.
os(X,windows2000) ∨ os(X,windowsXP) ∨ os(X,linuxRedHat7_1) ∨ os(X,linuxRedHat5_2) ∨ os(X,linuxRedHat8_0) :- arp(X,_,1,mac00_00_00_00_00_00).
%TCP Syn
os(X,linuxRedHat5_2) :- tcp(X,_,_,_,no,syn,64).
os(X,windows2000) ∨ os(X,windowsXP) :- tcp(X,_,_,_,yes,syn,128).
os(X,freeBSD5_0) ∨ os(X,linuxRedHat7_1) ∨ ∨ os(X,linuxRedHat8_0) ∨ os(X,macOS) ∨ os(X,sunOS) :- tcp(X,_,_,_,yes,syn,64).

Table 2. Some Passive Rules for Hybrid OS discovery

single packet to a signature and they are memoryless) and active methods are too noisy (most of them use abnormal packets that will themselves trigger IDS alarms) to provide context for IDS. [26] presents another case where the knowledge of the target operating system is important for the context of an IDS.

As a second example, we look at a new hybrid approach for operating system discovery (OSD) specially designed for context gathering for IDS. This hybrid approach [3] relies on a new knowledge representation formalism, Answer Set Programming (ASP) [2]. In short, ASP is an extension to Prolog with disjunction in the head, both weak and strong negation, non-monotonic reasoning, and a clear semantic. ASP rules can be defined for OSD, such as those in Table 2, meaning that if we see a packet originating from machine Ip and matching the right side of a rule, then the OS running on Ip is one of the OS defined by the left side of this rule.

If we see an ARP packet from $10.0.0.2$ with the destination MAC field filled with 0s, then we can deduce (using the third ARP Request rule from Table 2) that $10.0.0.2$ is running Windows 2000, Windows XP, RH 5.2, RH 7.1, or RH 8.0. Now if we see a TCP packet originating from $10.0.0.2$ with the syn flag set, the don't fragment bit set and a time to live of 64, then we could deduce (using the third TCP Syn rule from Table 2) that $10.0.0.2$ is running either RH 7.1 or RH 8.0, since the choices are currently limited by the previous deduction triggered by the ARP packet (multi-packet reasoning is possible here, as well as the use of memory). From there, if a user (possibly an IDS) queries the system to know if $10.0.0.2$ is running Windows XP, the knowledge base is sufficient to answer no. However, if the user wants to know if $10.0.0.2$ is running RH 7.1, then we can't tell for sure. That is when the active part comes in. A plan will be created to actively fetch the missing information in order to answer the query. In this case, to distinguish between RH 7.1 and RH 8.0, it should be sufficient to execute only one active test (much better than active tools that always perform all tests).

2.4. Machine Learning

Once it has been decided to use and represent context to enhance IDS capability, an interesting question is "Can we automatically learn new contextual detection rules that will outperform the classical rules?". For instance, it would be interesting to automatically generate chronicles that augment Snort rules with some contextual information. Chronicle learning has been explored as a component for an Interface Agent for network supervision [5]. The on-line learning algorithm used was inspired by Mitchell's version spaces [16]. Vulnerability context (as in the example of Table 1) can be harvested from databases like Security Focus [23], but it is much harder to automatically generate reaction contextual rules. An effort to automatically learn reaction patterns to assess the likelihood of success of an intrusion attempt is currently underway at Canada's Communication Research Centre. The training is done on a large data set of attack attempts against many different systems, where some attempts were successful and some failed; see [15]. Another interesting avenue would be to learn environment-specific context rules (i.e., the network structure, which machine is a server, etc). This would allow an IDS to be automatically tuned for the specific environment it will monitor.

Another interesting machine learning project, named ALAC (Adaptive Learner for Alert Classification), is presented in [20]. ALAC learns how the security officer classifies the alarms (false positive vs true positive) and can then automatically pre-classify new alarms. It can also drop high probability false positives in order to reduce the workload of the security officer. ALAC is implemented using the RIPPER rule learner. Equation 4 provides a rule learned by ALAC; this rule means that if the number of alarms classified as intrusion within the last minute is 0 and there have been other alarms raised by a given signature and targeted at the same IP address as the current alarm in the last 3 minutes, then the alarm should be classified as a false positive.

$$\#intr(1) \leq 0 \wedge \#sign(3) \geq 1 \wedge \#dstIP(3) \geq 1 \Rightarrow \text{Class = FALSE} \qquad (4)$$

3. AI in Anomaly-based IDS

This section provides a study of anomaly-based IDS and examines the AI components they rely on; the emphasis is again on knowledge representation and machine learning. Unlike signature-based IDS, for which AI is used only to enhance their capability, anomaly-based IDS are fundamentally built using AI components.

3.1. Introduction

Anomaly-based IDS aim to address the main drawback of signature-based IDS; that is, to detect new, previously unseen attacks. To do this, they use the opposite strategy. Instead of modeling known attacks and raising alarms when an event matches the model, they model normal activity and raise an alarm when an event deviates from the model. This idea is based on the assumption that the normal behavior of a given environment (network or program) remains mostly the same and changes only gradually over a long period of time (so it should be possible to gradually update the model).

There are three main design issues in an anomaly-based IDS: How should normal behavior be represented? How should normal behavior be learned? When should an event be considered malicious? Here we present two different anomaly-based IDS: one is host-based (Section 3.2) and the other is network-based (Section 3.3). For each of them, we discuss the knowledge representation it uses to describe normal behavior, the learning technique it uses to populate the normal model, and the mechanism to detect malicious events.

3.2. Sense of Self for Unix Process

Here we present a host-based IDS monitoring the system calls of Unix-based systems inspired by biological immune systems [6].

3.2.1. Knowledge Representation

The model is built with short sequences of system calls. There is one model per program (since different programs behave differently) and possibly per user (since two different users could use a given program very differently). So the model is n-grams over a sequence of system calls (typically $n = 6$ is used, see [27]).

3.2.2. Learning

Consider the training set to be one long sequence of recorded system calls. For each system call s seen in the training set, compute s_i (for $1 \leq i \leq n$), the set of system calls that are seen at position i in any subsequence starting with s. For instance, from the training sequence "open, read, mmap, mmap, open, getrlimit, mmap, close", the sets modeling the possible 4-grams starting with open are $\text{open}_1 = \{\text{read,getrlimit}\}$, $\text{open}_2 = \{\text{mmap}\}$, $\text{open}_3 = \{\text{mmap, close}\}$.

- **Inductive Bias:** The sequence $\rho = s, s^1, s^2, s^3$ is normal for every $s^i \in s_i$, even if ρ is not a proper subsequence of the training set. For instance, using, once again, the training sequence "open, read, mmap, mmap, open, getrlimit, mmap, close", then the sequence "open,read,mmap,close" is totally normal, even if it is not a proper subsequence of the training sequence. It is normal because read $\in \text{open}_1$, mmap $\in \text{open}_2$, and close $\in \text{open}_3$.
- **On-line Learning:** It is possible to update the model by adding new valid subsequences of system calls. The challenge when expanding the model with new sequences is to make sure these sequences are normal (not malicious). However, it is not clear how malicious sequences could be removed from the model.
- **Noise:** This learning method is not tolerant to noise and thus requires an attack-free data set (which is very hard to provide) or supervised learning (which is very tedious, if not impossible).

3.2.3. Detection

Simply put, this anomaly-based IDS uses the following procedure to detect abnormal behavior of a program: for a given sequence ρ of length n, compute how much ρ deviates from the normal model. That is how many wrong system calls come to be present in ρ. If this value is greater than some threshold, an alarm is raised.

3.3. Payload Anomaly Detection

Here we present a network-based IDS monitoring payload data of network packets [30]. The payload data is the message content of a packet (the html code of a web page, the binary content of a file download, etc).

3.3.1. Knowledge Representation

The model is composed of byte frequency distribution in the packet payload (each of the 256 possible ASCII characters). There is one model per port, since data on each port is expected to be different, and also per payload length, since data for a small payload (http error messages) is significantly different from data in a big payload (video streaming). So the model is the frequency of every possible 1-grams over bytes in the payloads.

3.3.2. Learning

Consider the training set to be a set of packets sent on the same port. For each payload length, compute the average relative frequency of each possible ASCII character and its standard deviation. Use some form of clustering to merge two neighboring models (with respect to payload length) that are similar enough (using the Mahalanobis distance over their frequency distribution). The clustering is important to eliminate redundant models and thus reduce the total number of models needed to represent normal traffic.

- **Inductive Bias:** The main assumption here is that the ordering of characters is not important in the payload. This led to a mimicry attack (see Section 3.4) against this IDS, which pads the payload with dummy characters to make the payload fit the normal character distribution (the normal character distribution can be computed when needed by an attacker); see [12]. However, an update was proposed to the model (using Bloom filters instead of frequency analysis) to address this mimicry attack, but it is out of scope for this chapter, see [29].
- **On-line Learning:** It is possible to update the model with new normal instances without going through the training phase again, simply by adjusting the average values.
- **Noise:** It is tolerant to some small amount of noise, since the presence of a few malicious packets will not be sufficient to dramatically change the overall average distribution of characters. But typically it is used in supervised mode (since clean data sets are very difficult to get). [30] claims it is possible to automatically purify the training set by training on it, then removing the instances that are far enough from the learned model and, finally, retraining on the instances that were not removed. However, such an approach could easily lead to overfitting.

3.3.3. Detection

For a given packet p, simply compute the frequency distribution of its payload. Then check how far this frequency distribution deviates from the normal model, using the Mahalanobis distance. If this distance is greater than some threshold, then the packet is deemed malicious and an alarm is raised.

3.4. Limitations

The main problem of anomaly-based IDS is definitely to create a training data set that contains no malicious traffic, that is representative of the usual traffic for the targeted network, and that contains no artifacts (artifacts are often present in emulated traffic; see [14]). If the training set contains malicious traffic, then this traffic will be considered normal by the IDS and it will be possible for someone to compromise the network without being detected. If the training set is not representative of the usual traffic, then the model will not be accurate and will generate many false alarms. This is very important, as anomaly-based IDS are extremely environment-dependent; they cannot easily be trained in one environment and then installed in another, since the definition of normal behavior is often different between environments. If the training set contains artifacts, then the model will probably overfit the training examples.

Another issue with anomaly-based IDS is the vulnerability to mimicry attacks [28]. A mimicry attack allows a sophisticated attacker to cloak its intrusion to avoid detection by an anomaly-based IDS. The idea, in the context of a host-based IDS, is to transform a malicious sequence of system calls into one that is considered normal by the IDS but will still result in exploiting the vulnerability. This is mainly done by adding dummy system calls in between the malicious ones, such that each subsequence considered by the IDS will look normal.

4. Multi-Agent Architectures for IDS

Using a multi-agent architecture for intrusion detection offers several advantages: multiple points of view of a network (useful for detecting distributed attacks); autonomy of the agents (to make overall maintenance easier); independence of the agents (allows for cross-monitoring to prevent the agents themselves from being compromised); adaptability (add/remove agents to start/stop monitoring events of current high/low interest); scalability (since the detection process is not centralized). However, a distributed architecture raises two major challenges: secure communication and data fusion from many distinct sensors. Here, we give a high-level description of two multi-agent-based IDS architectures. Similar work includes the use of mobile agents (see [9]) and the FIPA-OS environment (see [19]).

4.1. Autonomous Agents for Intrusion Detection (AAFID)

AAFID [25] is an architecture using multiple types of agents (Figure 1). Agents of the first type are called filters. There are many filters on each machines; each filter having a very specific task (monitoring authentication attempts, foraging information from log files, monitoring network activity, etc). The filters report their events to a transceiver, in some standard format. There is one transceiver per host and it correlates events from its filters and controls those filters (creation, deletion, etc). A transceiver will then decide if the situation is worth being reported, and if so, will notify its superior, called a monitor (a transceiver may be controlled by more than one monitor). Monitors correlate events from (and control) many transceivers. Monitors report (when they choose to do so) directly to the agent interfacing with the security officer. This architecture uses very domain-specific agents (filters), general agents (transceivers), and agents with a global network view (monitors).

4.2. A Suite for Multi-Agent Intrusion Detection

Another interesting multi-agent architecture is presented in [7]. This architecture is a suite for intrusion detection, not just an IDS. It has three main components: an attack simulator, an intrusion detection module, and a learning module.

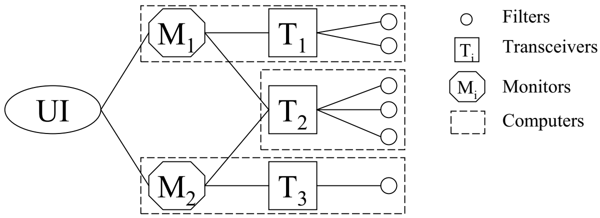


Figure 1. AAFID Architecture

The agent-based simulator of attacks against computer networks (ASACN) is a module to simulate attacks. Attacks are specified using a formal language. A coordinator agent requests the participation of some attack agents and coordinates the attack execution. The script language is used to reproduce multi-phase attacks (e.g. create a buffer overflow, then gain root access, and finally execute malicious code) while distributed attacks (e.g. distributed denial of service) can be done using many agents at different places in the network.

The multi-agent intrusion detection system (MIDS) uses multiple levels of agents. Low-level agents perform data pre-processing (e.g. flow reconstruction). Intermediate-level agents perform task-specific analysis (access control, buffer overflow, etc) on the formatted data. High-level agents correlate events from the intermediate-level agents.

The multi-agent intrusion detection learning system (MIDLS) uses the attack simulator to automatically generate the agents used by the MIDS module. Two classes of learning agents are used: searching for patterns and frequent episodes in ordered sequences of events and extracting rules from attribute-value structured data.

Conclusion and Further Reading

Throughout this chapter, we have seen some examples of intrusion detection research projects that rely on AI concepts to win the arms race against hackers. We mainly focused on knowledge representation, machine learning and multi-agent architectures. A lot of work still remains to be done to integrate AI concepts into intrusion detection. Reducing false-positives, emulating normal traffic, avoiding evasion by sophisticated attacks, and enhancing the quality of the alarm messages are the problems researchers are currently focusing on, and chances are some solutions to these problems can be found within AI.

The projects discussed so far in the present chapter are just a few among many network security-related projects and ideas relying on AI. Below we provide some pointers for further reading on other AI-related topics in network security.

[13] proposes learning how computers react when infected by a virus and emulating this reaction in order to be able to record and study the full infection process of new viruses, using honeypots. [11] and [24] study how it would be possible to learn some general worm propagation traffic patterns (by releasing worm-spreading viruses in a virtual environment) and use this to detect when new worms start to propagate. Another effort to detect the spreading of worms is done by mining DNS traffic data to discover new worm propagation payload content; see [10]. [22] proposes using a naive Bayes classifier to filter malicious e-mail attachments.

References

[1] Stefan Axelsson. The Base-Rate Fallacy and the Difficulty of Intrusion Detection. *ACM Transactions on Information and System Security*, 3(3):186–205, 2000.

[2] Chitta Baral. *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press, 2003.

[3] François Gagnon, Babak Esfandiari, and Leopoldo Bertossi. A Hybrid Approach to Operating System Discovery Using Answer Set Programming. *To appear in Proceedings of the 10th IFIP/IEEE Symposium on Integrated Management (IM'07)*, 2007.

[4] Christophe Dousson. *Suivi d'évolutions et reconnaissance de chroniques*. PhD thesis, LAAS, Toulouse, France, 1994.

[5] Babak Esfandiari, G. Deflandre, and J. Quinqueton. An Interface Agent for Network Supervision. *Proceedings of the 1st Intelligent Agents for Telecom Applications (IATA'96)*, pages 21–28, 1996.

[6] Stephanie Forrest, Steven A. Hofmeyr, Anil Somayaji, and Thomas A. Longstaff. A Sense of Self for Unix Processes. *Proceedings of the 1996 IEEE Symposium on Security and Privacy (S&P'96)*, pages 120–128, 1996.

[7] Vladimir Gorodetski and Igor Kotenko. The Multi-Agent Systems for Computer Network Security Assurance: Frameworks and Case Studies. *Proceedings of the 2002 IEEE International Conference on Artificial Intelligence Systems (ICAIS'02)*, pages 297–302, 2002.

[8] Richard Heady, George Luger, Arthur Maccabe, and Mark Servilla. The Architecture of a Network Level Intrusion Detection System. Technical report, Computer Science Department, University of New Mexico, August 1990.

[9] Guy Helmer, Johnny S.K. Wong, Vasant Honayar, Les Miller, and Yanxin Wang. Lightweight Agents for Intrusion Detection. *Journal of Systems and Software*, 67(2):109–122, 2003.

[10] Keisuke Ishibashi, Tsuyoshi Toyono, Katsuyasu Toyama, Mashiro Ishino, Haruhiko Ohsima, and Ichiro Mizukoshi. Detecting Mass-Mailing Worm Infected Hosts by Mining DNS Traffic Data. *Proceeding of the 2005 Annual Conference of the Special Interest Group on Data Communication (SIGCOMM'05)*, 2005.

[11] Hyang-Ah Kim and Brad Karp. Autograph: Toward Automated, Distributed Worm Signature Detection. *Proccedings of the 13th USENIX Security Symposium*, 2004.

[12] Oleg Kolesnikov and Wenke Lee. Advanced Polymorphic Worms: Evading IDS by Blending in with Normal Traffic. *USENIX Security Symposium*, 2006.

[13] Corrado Leita, Ken Mermoud, and Marc Dacier. ScriptGen: an Automated Script Generation Tool for Honeyd. *Proceedings of the 21st Annual Computer Security Applications Conference (ASCAC'05)*, 2005.

[14] Matthew V. Mahoney and Philip K. Chan. An Analysis of the 1999 DARPA/Lincoln Laboratory Evaluation Data for Network Anomaly Detection. *Proceedings of the 6th International Symposium on Recent Advances in Intrusion Detection (RAID'03) - LNCS*, 2820, 2003.

[15] Frédéric Massicotte, François Gagnon, Mathieu Couture, Yvan Labiche, and Lionel Briand. Automatic Evaluation of Intrusion Detection Systems. *Proceedings of the 2006 Annual Computer Security Applications Conference (ACSAC'06)*, 2006.

[16] Tom Mitchell. *Machine Learning*. McGraw Hill, 1997.

[17] Benjamin Morin and Hervé Debar. Correlation of Intrusion Symptoms: An Application of Chronicles. *Proceedings of the 6th Symposium on Recent Advances in Intrusion Detection (RAID'03) - LNCS*, 2820:94–112, 2003.

[18] Samuel Patton, William Yurcik, and David Doss. An Achilles' Heel in Signature-Based IDS: Squealing False Positives in SNORT. *Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection (RAID'01)*, 2001.

[19] Taraka D. Pedireddy and Jose M. Vidal. Multiagent Network Security System using FIPA-OS. *Proceedings of the 2002 IEEE SoutheastCon*, pages 229–233, 2002.

[20] Tadeusz Pietraszek. Using Adaptive Alert Classification to Reduce False Positives in Intrusion Detection. *Proceedings of the 7th International Symposium on Recent Advances in Intrusion Detection (RAID'04) - LNCS*, 3224:102–124, 2004.

[21] Martin Roesch. SNORT - Lightweight Intrusion Detection for Networks. *Proceedings of the 13th System Administration Conference (LISA'99)*, 1999.

[22] Matthew G. Schultz, Eleazar Eskin, Erez Zadok, Manasi Bhattacharyya, and Salvatore J. Stolfo. MEF: Malicious Email Filter - A UNIX Mail Filter that Detects Malicious Windows Executables. *USENIX*

Annual Technical Conference, 2001.

[23] Security Focus. Security Focus HomePage. www.securityfocus.org.

[24] Sumeet Singh, Cristian Estan, George Varghese, and Stefan Savage. Automated Worm Fingerprinting. *Proceedings of the 6th USENIX Symposium on Operating System Design and Implementation (OSDI'04)*, 2004.

[25] Eugene H. Spafford and Diego Zamboni. Intrusion Detection Using Autonomous Agents. *Computer Networks*, 34(4):547–570, 2000.

[26] Greg Taleck. Ambiguity Resolution via Passive OS Fingerprinting. *Proceedings of the 6th International Symposium on Recent Advances in Intrusion Detection (RAID'03) - LNCS*, 2820:192–206, 2003.

[27] Kymie M.C. Tan and Roy A. Maxion. Why 6? Defining the Operational Limits of stide, an Anomaly-Based Intrusion Detector. *Proceedings of the 2002 IEEE Symposium on Security and Privacy (S&P'02)*, pages 188–201, 2002.

[28] David Wagner and Paolo Soto. Mimicry Attacks on Host-Based Intrusion Detection Systems. *Proceedings of the 9th ACM Conference on Computer and Communication Security (CCS'02)*, pages 255–264, 2002.

[29] Ke Wang, Janak J. Par, and Salvatore J. Stolfo. Anagram: A Content Anomaly Detector Resistant to Mimicry Attack. *Proceedings of the 9th International Symposium on Recent Advances in Intrusion Detection (RAID'06)*, pages 226–248, 2006.

[30] Ke Wang and Salvatore J. Stolfo. Anomalous Payload-based Network Intrusion Detection. *Proceedings of the 7th International Symposium on Recent Advances in Intrusion Detection (RAID'04)*, 2004.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

A Classifier Ensemble Approach to Intrusion Detection for Network-Initiated Attacks

Stefanos KOUTSOUTOS

Athens Information Technology, 19.5km Markopoulou Ave. Paiania 19002 Greece
skou@ait.edu.gr

Ioannis T. CHRISTOU

Athens Information Technology, 19.5km Markopoulou Ave. Paiania 19002 Greece
ichr@ait.edu.gr

Sofoklis EFREMIDIS

INTRACOM S.A. Telecom Solutions, 19,7 km Markopoulou Ave. Peania 19002, Greece
sefr@intracom.gr

Abstract. We present a classifier ensemble system using a combination of Neural Networks and rule-based systems as base classifiers that is capable of detecting network-initiated intrusion attacks on web servers. The system can recognize novel attacks (i.e., attacks it has never seen before) and categorize them as such. The performance of the Neural Network in detecting attacks from network data alone is very good with success rates of more than 78% in recognizing new attacks but suffers from high false alarms rates. An ensemble combining the original ANN with a second component that monitors the server's system calls for detecting unusual activity results in high prediction accuracy with very small false alarm rates. We experiment with a variety of ensemble classifiers and decision making schemes for final classification. We report on the results we got from our approach and future directions for this research

Introduction

Intrusion Detection is a major issue that every administrator has to deal with effectively so as to maintain proper operation of his/her Internet-connected servers. Network security and Intrusion Detection was studied as early as 40 years ago [6], but with the current growth of the Internet and the number of attackers, it has taken on a very prominent role in the fields of computer and communications security ([7, 12, 18, 21]). For this reason, the problem has been widely studied from a number of different perspectives. Data Mining approaches have been used to detect unusual activities on the servers ([8, 10]). System calls are monitored in [9] to discover patterns that are characteristic of an intrusion. Combination of unsupervised clustering and supervised decision tree learning has been recently proposed as an excellent means to improve attack detection with practically zero false alarms while simultaneously improving system robustness and stability [24]. Actual intrusions (when a person actually takes over someone else's workstation in an unauthorized way) have also been studied in [16, 17, 20] and dealt with a number of different approaches, by employing utilizing Artificial Intelligence techniques in several cases.

The idea of using Artificial Neural Networks or ANNs ([3, 22]) for finding novel attacks is not new ([1, 2, 4, 13, 14]). The usual approach is to train a neural network with the behavior of the system and let it recognize any behavior which substantially differs from the one considered as normal. The behavior is described to a neural network with a number of features, like system calls invocation, rates of system call invocation, etc. In this paper *we explore the possibility of using a pattern classifier ensemble [25] based on neural networks and decision trees which will be trained and accept data directly from the data network as well as the Server's OS kernel and try to identify attacks using this data, in real time, by continually monitoring the network as well as the server state.*

The ultimate goal is to be able to identify both known and unseen attacks, on the fly, mainly using their binary signature, as they travel through the network. There are numerous kinds of attacks, each of which has a completely different structure of signature. Because different applications may use different protocols, network traffic may substantially differ from one case to another. This variety renders the task of building a neural network, which will be able to respond to every possible attack, essentially infeasible. In order to overcome this kind of problem, *we chose to fix the target application/protocol used as well as the kind of attacks we are targeting. The application chosen is an Apache web server (Linux/i386), and the kind of attacks we are targeting is remote code injection attacks, exploiting stack or heap overflows, etc.*

IDS Ensemble Architecture

Using a single trained ANN alone as an IDS leads to serious performance problems with normal traffic that is characterized as malicious. Because of the very high false alarms rates [26], such a system would be useless as any system administrator would quickly decide to ignore all alarms raised by the ANN, including of course the real ones. For this reason we combine a properly trained ANN as network data classifier with another rule-based classifier trained on different input data so as to maximize classifier diversity and exploit different input sources available. For the second classifier we have experimented with ANNs as well as with rule-based classifiers. The System Call classifier accepts input from O/S system calls and helps identify real attacks on the server –thus eliminating false alarms raised by the network data classifier. Monitoring system call invocations for unusual patterns has been studied before for detecting intrusion attempts with reasonable success [9].

The rule-based system call classifier used is C4.5 [27]. Decision tree based classifiers are inherently unstable classifier methods and for this reason may be well suited for classifier ensembles [25]. A decision maker coordinates the actions of the two weak classifiers (network-traffic and system-call), and taking their combined outputs into account, makes the final decision about whether to issue an alarm or not.

For this purpose, we built a tool that monitors the system calls invoked by the Apache web server that is our target server application. Figure 1 shows our architecture. In the following we describe each weak classifier in more detail.

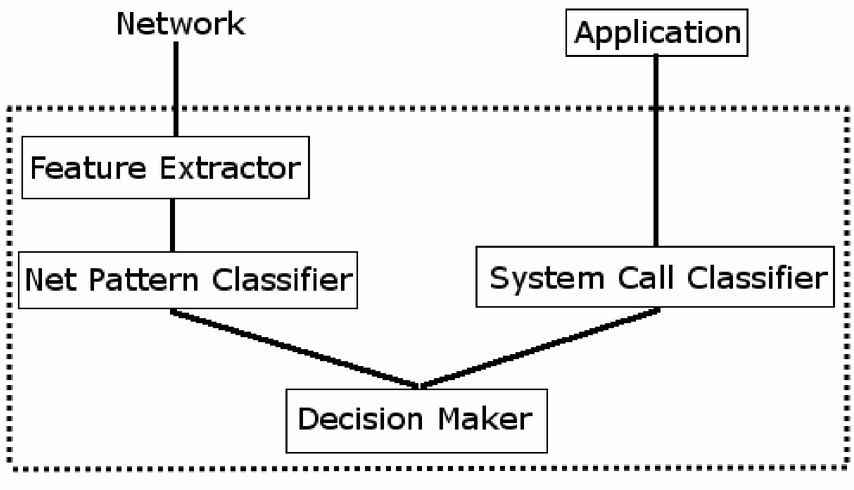


Figure 1. Architecture of the Classifier Ensemble for network-initiated intrusion attacks.

An ANN Classifier for Detecting Attacks in the Network Traffic of a Web Server

One of the first issues in the design of a Neural Network is the type of inputs it will accept, or the features to be used in the representation. Having an "aggregator function" that does not essentially loose any important features in our network data for classification purposes (operating on the packets right before they enter the ANN) may both minimize their size and keep their "structure" intact. The aggregator function we chose is [26]:

$$f(m) = \frac{\sum_{i=0}^{w} m_i \cdot m_i}{w \cdot 255^2}$$

In the above formula, the m is the incoming packet whereas m_i is its ith byte. The window w is a system parameter, which specifies the amount of bytes "grouped" together in one number. The denominator normalizes the output in the interval [0, 1]. The main goal is to minimize the data load passed on to the ANN.

Training the ANN

A component that listens to inbound web server traffic, will face HTTP protocol page requests (HTTP GET requests), but some of them may well be file upload or other form-handling requests (HTTP POST, for example). These are also included in the traffic that can be seen. Using the feature extractor introduced above, we can verify that pure text (ASCII) won't ever cause the ANN to give an output larger that 0.22. This of course means that the task of differentiating between text and binary data is an easy task. The HTTP protocol is a text based one, but of course not all binary traffic is malicious. File uploading (such as the uploading of multimedia material), supported by many sites, may result in the ANN seeing images or other kinds of binary files. So, in order to differentiate between this category and malicious code, we trained an ANN with jpeg image file parts for one class and actual code for the other. Libc was used for creating the code exemplars since it's the library closer to the system than any other, so actual exploits should have strong similarities to the exemplars created thus.

In the general case of measuring the ability of the neural network to distinguish between the two classes, we found out that the ANN can indeed classify code as such, but it's difficult for it to correctly identify image binary data from potentially harmful code. In our tests, we trained several neural networks, using standard BackPropagation [22]. The input layer sizes selected were 10 and 20. The hidden layer sizes ranged from half the input layer size to its double. The epochs each ANN had gone through were 1000, 5000 and 10000. In total we trained 30 ANNs with 800 exemplars per group. Figure 2 shows the performance graph of all those ANNs. The line marked *t/set-1* refers to the train-set data which should be classified as malicious traffic, while the *t/set-0* refers to normal traffic. The line *class-1* refers to the NN performance on real malicious traffic, which has not been seen by the NN during its training. Accordingly, *class-0* is normal traffic, which the NN has not seen during its training.

The correct classification of real malicious traffic (class-1) reaches 80%, for the trained NNs. This means that 80% of completely novel attacks are correctly identified as such. We can also see that the percentage of correct classification of real normal traffic (class-0) is always below that of class-1, for all the cases where the performance for class-1 is above 50%.

By augmenting the training set with more code (t/set-1) and less image data (t/set-0), we got an interesting outcome for the same NN architectures. The train set had 300 exemplars of (artificial) normal traffic (t/set-0) and 500 ones of code (t/set-1). The results are shown in Figure 3.

What is important to notice in these two graphs is that although the correct classification percentage for malicious traffic (class-1) has improved substantially, the one for normal traffic remained at the same levels. Nevertheless, although the ANN can correctly classify potentially malicious code as such, it is also often mislead when presented with jpeg images, giving false alarms at unacceptably high rate.

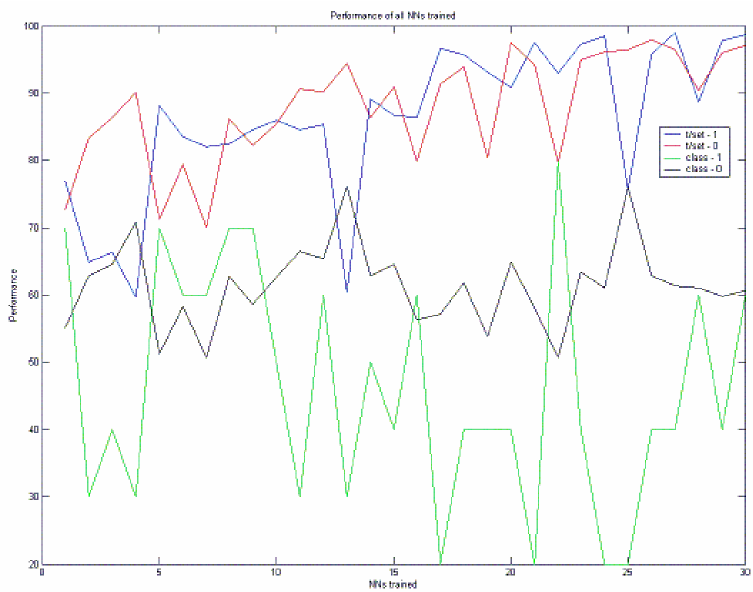


Figure 2. ANN Performance using 800 exemplars per category.

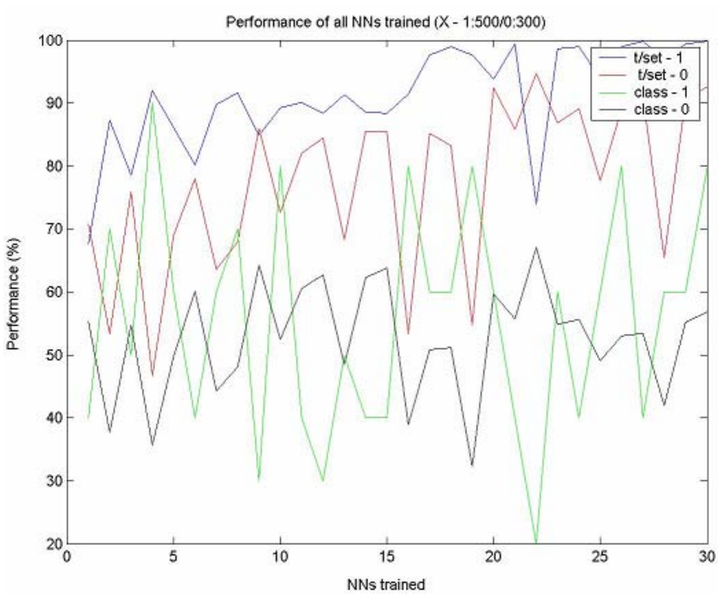


Figure 3. ANN Performance using 500 and 300 exemplars per category.

System Call Classifier

In order to successfully create and test a system-call classifier, we needed a training set containing some patterns of misbehavior as well as the usual patterns of normal system call invocations under normal circumstances. For this reason, we created a web application as a CGI program in C with two exploitable buffer overflows and installed it on our Apache test server running under Linux. Afterwards, a number of attacks, together with normal requests were created. All the requests combined (normal and attack-carrying requests) were issued to the server.

The first System Call classifier we experimented with was an ANN-based one. The results shown in the following table are for unseen cases during training. The column labeled "threshold" indicates the threshold value above which, if the output of the ANN is raised, an alarm is issued. They indicate that the System-Call classifier by itself when properly trained can recognize 50% of new attacks while *never raising a false alarm.*

Table 1: ANN-based System Call Classifier.

epochs	hidden layer size	Threshold	% true alarms	% false Positives	% overall accuracy
300	3	0.50	70.00	85.00	33.33
300	3	0.70	60.00	12.50	78.33
300	3	0.90	0.00	0.00	66.67
300	15	0.50	70.00	85.00	33.33
300	15	0.70	65.00	5.00	85.00
300	15	0.90	55.00	0.00	85.00
1000	5	0.50	65.00	12.50	80.00
1000	5	0.70	40.00	0.00	80.00
1000	5	0.90	0.00	0.00	66.67
1500	3	0.50	0.00	5.00	63.33
1500	3	0.70	0.00	5.00	63.33
1500	3	0.90	0.00	0.00	66.67
2000	3	0.50	70.00	87.50	31.67
2000	3	0.70	50.00	0.00	83.33
2000	3	0.90	20.00	0.00	73.33
2000	8	0.50	70.00	87.50	31.67
2000	8	0.70	50.00	5.00	80.00
2000	8	0.90	50.00	0.00	83.33

In the second experiment, Quinlan's C4.5 [27] was used as the classifier learning technique.

Table 2: C4.5-based System Call Classifier.

Request Type	Number of Requests	Number Alarms	%Accuracy
POST normal	20	0	100
POST attack	10	6	60
Post Multipart normal	20	1	95
Post Multipart attack	10	7	70

A single decision tree was grown using default program parameters for learning. The results of Table 2 correspond to system performance on unseen test instances during training. Using C4.5 the overall system accuracy is 86.6% with a false alarm rate of less than 2.5%. The ability of the system alone to recognize novel attacks reaches 65%.

Ensemble Implementation Results

Regarding the Decision Maker component of the classifier ensemble, we experimented with a variety of strategies. The first strategy works by continuously listening to the output of the network classifier, and when the output exceeds a certain threshold the Decision Maker asks the opinion of the System-Call Classifier. If the output of the second component also exceeds a certain threshold, then an alarm is raised.

The ensemble consisting of an ANN for classifying network traffic and a C4.5 grown decision tree for classifying system call sequences and combining them with a Decision Maker following the first strategy is capable of detecting 65% of novel attacks with a false alarm rate of only 2.5%. The detailed results are shown in Table 3.

Table 3: Ensemble Performance with C4.5 used for system call monitoring.

net(raw) epochs	**net(raw) hid. layer size**	**net(raw) thres.**	**dm thres.**	**% true attacks detected**	**% false alarms**	**% overall accuracy**
300	35	0.43	8.75	65.00	10.00	81.67
300	20	0.57	7.88	65.00	2.50	86.67
300	50	0.52	8.75	65.00	10.00	81.67
1000	20	0.55	8.25	65.00	5.00	85.00
1500	50	0.47	8.63	65.00	5.00	85.00
2000	35	0.51	8.38	65.00	2.50	86.67

In Table 3, the first three columns indicate the characteristics of the ANN trained for detecting attacks from monitoring network traffic. The fourth column describes the threshold value that the decision maker uses to raise an alarm for a particular request. The last three columns are the system performance in terms of accuracy in detecting novel (unseen) attacks, false alarm rates, and overall classification accuracy as a percentage of requests arriving at the server.

Raw results from the combination of a single ANN-based classifier for monitoring network traffic and a single ANN-based classifier for monitoring system-calls are shown in Figure 4.

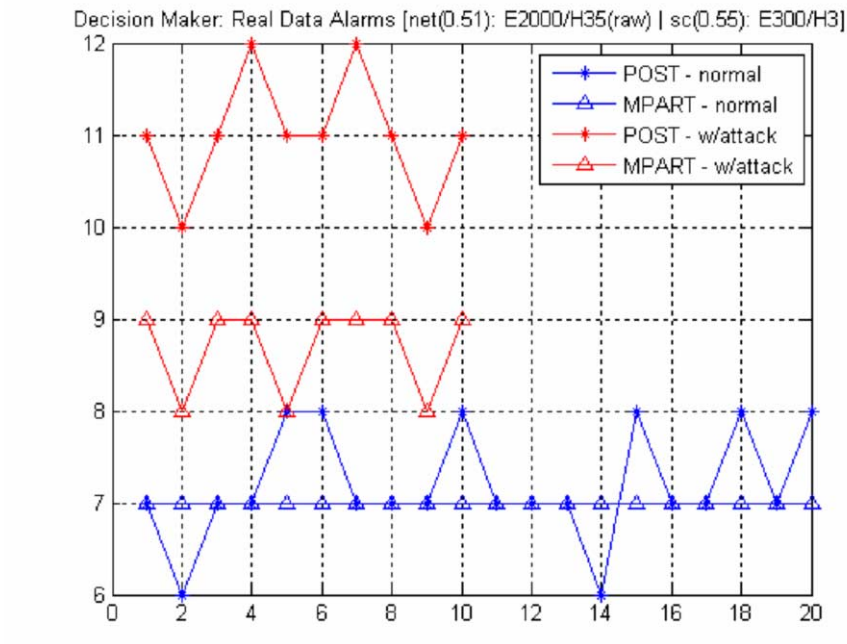


Figure 4: Number of Alarms raised per request from base ANN classifiers combination. Decision maker bases decision on total number of alarms raised from single request.

The overall results from applying this strategy are shown in the table below.

Table 4: ANN-Ensemble Performance.

sc epochs	sc hid. layer size	sc thres.	Net (raw) epochs	Net (raw) hidden layer size	Net (raw) thres.	dm thres.	% true alarms accuracy	% false alarms	% over-all accu-racy
300	3	0.55	300	35	0.43	8.50	85.00	25.00	78.33
300	3	0.55	300	20	0.57	7.63	100.00	10.00	93.33
300	3	0.55	300	50	0.52	8.50	85.00	25.00	78.33
300	3	0.55	1000	20	0.55	8.00	85.00	10.00	88.33
300	3	0.55	1500	50	0.47	8.38	85.00	12.50	86.67
300	3	0.55	2000	35	0.51	8.13	85.00	0.00	95.00
300	15	0.49	300	35	0.43	8.63	45.00	45.00	51.67
300	15	0.49	300	20	0.57	7.75	85.00	82.50	40.00
300	15	0.49	300	50	0.52	8.63	45.00	45.00	51.67
300	15	0.49	1000	20	0.55	8.13	45.00	27.50	63.33

300	15	0.49	1500	50	0.47	8.50	45.00	35.00	58.33
300	15	0.49	2000	35	0.51	8.25	45.00	15.00	71.67
1000	5	0.41	300	35	0.43	9.00	50.00	25.00	66.67
1000	5	0.41	300	20	0.57	8.13	85.00	10.00	88.33
1000	5	0.41	300	50	0.52	9.00	50.00	25.00	66.67
1000	5	0.41	1000	20	0.55	8.50	85.00	27.50	76.67
1000	5	0.41	1500	50	0.47	8.88	85.00	35.00	71.67
1000	5	0.41	2000	35	0.51	8.63	85.00	15.00	85.00
1500	3	0.24	300	35	0.43	9.13	50.00	25.00	66.67
1500	3	0.24	300	20	0.57	8.25	50.00	60.00	43.33
1500	3	0.24	300	50	0.52	9.13	50.00	25.00	66.67
1500	3	0.24	1000	20	0.55	8.63	50.00	77.50	31.67
1500	3	0.24	1500	50	0.47	9.00	50.00	12.50	75.00
1500	3	0.24	2000	35	0.51	8.75	50.00	65.00	40.00
2000	3	0.69	300	35	0.43	6.50	85.00	25.00	78.33
2000	3	0.69	300	20	0.57	5.63	100.00	10.00	93.33
2000	3	0.69	300	50	0.52	6.50	85.00	25.00	78.33
2000	3	0.69	1000	20	0.55	6.00	85.00	10.00	88.33
2000	3	0.69	1500	50	0.47	6.38	85.00	12.50	86.67
2000	3	0.69	2000	35	0.51	6.13	85.00	0.00	95.00
2000	8	0.68	300	35	0.43	6.38	85.00	25.00	78.33
2000	8	0.68	300	20	0.57	5.50	100.00	10.00	93.33
2000	8	0.68	300	50	0.52	6.38	85.00	25.00	78.33
2000	8	0.68	1000	20	0.55	5.88	100.00	27.50	81.67
2000	8	0.68	1500	50	0.47	6.25	85.00	12.50	86.67
2000	8	0.68	2000	35	0.51	6.00	85.00	0.00	95.00

In Table 4, ANN-based classifiers were used both for network-traffic classification and for system-call sequence classification. Decision maker bases decision on output values exceeding threshold shown in column labeled “dm thres.”. The first three columns describe the characteristics of the ANN trained to recognize attacks using system call sequences. Columns 4-6 describe the characteristics of the ANN trained to learn attacks carried on network traffic data. The columns “sc thres.” and “net thres.” denote the threshold value in the output above which the classifier raises an alarm. It is the value that minimizes the error in the training phase. As can be seen from the table, *the ensemble of two ANNs combined with a simple decision making approach based on* *threshold values can recognize 85% of new (unseen during training) attacks without raising any false alarms.*

The second strategy involves using a meta-classifier (using an ANN again) to learn how to detect whether an attack is taking place by combining the outputs of the network-traffic and system-call classifiers. Using two (differently trained) ANNs for net-traffic and another two ANNs for system-call classification, a meta-classifier ANN was trained on the output of these 4 ANNs to learn when a real attack is occurring. Raw data are shown in the following figure.

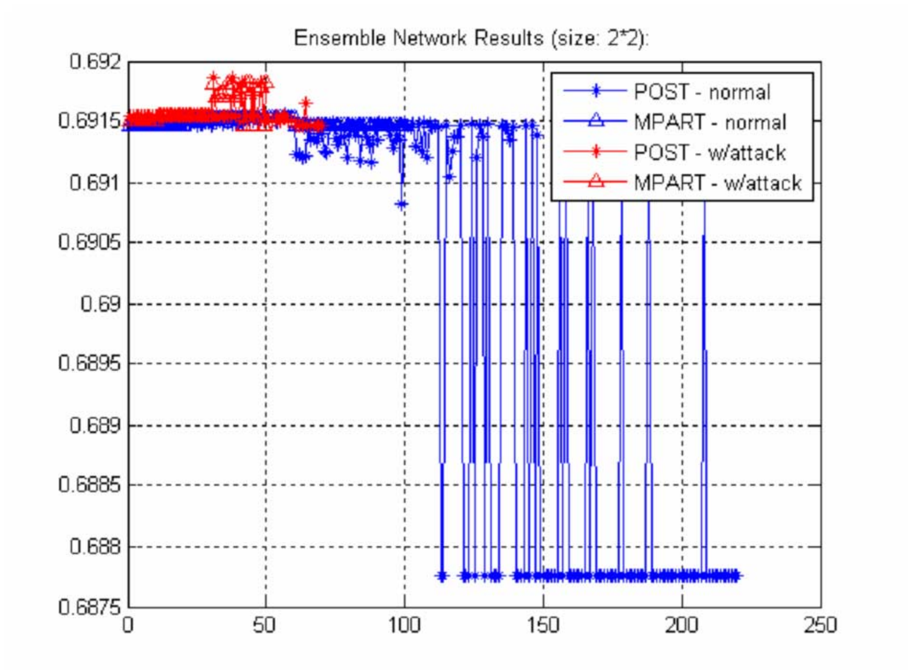


Figure 5: ANN-based Classifier Ensemble Performance.

Table 5: ANN-based Classifier Ensemble Performance.

1 sc epochs	2 sc epochs	1 net epochs	2 net epochs	dm thres.	% true attacks detected	% false alarms raised	% overall accuracy
300	300	1000	1500	0.59	42.50	18.75	70.68
300	300	1000	1500	0.91	66.67	12.50	81.82
300	300	1000	1500	0.97	82.50	12.50	86.14
300	300	1000	1500	0.99	85.00	14.06	85.68
300	300	1000	1500	0.93	65.83	12.81	81.36
300	300	1000	1500	0.90	63.33	17.50	77.27

300	300	1000	1500	0.94	100.00	36.88	73.18
300	300	1000	1500	0.95	100.00	28.75	79.09
300	300	1000	1500	0.86	48.33	6.25	81.36
300	300	1000	1500	0.95	100.00	26.25	80.91
300	300	1000	1500	0.98	85.83	15.00	85.23
300	300	1000	1500	0.98	100.00	15.31	88.86
1000	1500	1000	1500	0.38	0.00	34.69	47.50
1000	1500	1000	1500	0.97	71.67	14.38	81.82
1000	1500	1000	1500	0.95	100.00	30.63	77.73
1000	1500	1000	1500	0.99	100.00	22.50	83.64
1000	1500	1000	1500	0.61	52.50	56.88	45.68
1000	1500	1000	1500	0.98	82.50	12.50	86.14
1000	1500	1000	1500	0.68	31.67	19.38	67.27
1000	1500	1000	1500	0.66	65.00	93.75	22.27
1000	1500	1000	1500	0.37	55.00	90.63	21.82
1000	1500	1000	1500	0.59	22.50	31.56	55.91
1000	1500	1000	1500	0.98	85.00	14.69	85.23
1000	1500	1000	1500	0.99	100.00	17.81	87.05
2000	2000	1000	1500	0.87	100.00	32.81	76.14
2000	2000	1000	1500	0.95	82.50	12.50	86.14
2000	2000	1000	1500	0.73	19.17	32.19	54.55
2000	2000	1000	1500	0.99	100.00	31.88	76.82
2000	2000	1000	1500	0.92	97.50	18.13	86.14
2000	2000	1000	1500	0.96	80.83	12.50	85.68
2000	2000	1000	1500	0.94	88.33	37.19	69.77
2000	2000	1000	1500	0.99	69.17	12.50	82.50
2000	2000	1000	1500	0.95	86.67	30.63	74.09
2000	2000	1000	1500	0.59	79.17	93.75	26.14
2000	2000	1000	1500	0.96	82.50	6.25	90.68
2000	2000	1000	1500	0.99	82.50	12.50	86.14

The horizontal axis shows request-id whereas the vertical axis shows the output value of the meta-classifier that is the decision maker in the overall system. The performance of the system for various ANN architectures for the base and decision maker classifiers is depicted in Table 5.

The above 4-base classifier ANN ensemble can reach 100% accuracy in detecting novel attacks but with 15% false alarm rate which is for most practical purposes too high. Differently trained ensembles can reach detection accuracy of 82.5% while raising false alarms only 6.25% of the time.

Conclusions and Future Directions

We have shown that classifier ensembles are useful tools for detecting novel network initiated attacks against computer systems, and in particular web servers. We have combined input sources from raw network traffic together with system call monitoring and trained a number of classifiers on each data source to learn what constitutes a threat against an information system. We have experimented with a variety of decision making procedures for combining the output of the base classifiers and we found that a simple decision making procedure based on raising alarms when threshold values are exceeded from two ANNs each trained to work with network traffic and system call sequences can learn to recognize up to 85% of novel (unseen during training) attacks while never raising a false alarm (practically 0% false alarm rate). Overall system classification accuracy reaches 95%.

A line of research we are actively pursuing is that of improving system-calls monitoring-based attack detection with the application of other machine learning techniques; in particular, having a larger time window to work with, a rule-based induction system such as SLIPPER2 [23] can be used to derive useful rules regarding what constitutes an attack. Further, we are in the process of investigating the effectiveness of alternatives to BackPropagation as the learning-from-errors mechanism such as the GeneRec algorithm, that also happens to have a more plausible biological basis [11], even though it seems to require more computing power.

References

[1] W.Lee, S.Stolfo, K.Mok: “A Data Mining Framework for Building Intrusion Detection Models” - Proceedings of the 1999 IEEE Symposium on Security and Privacy (May 1999)

[2] H.Debar, M. Becker, D. Siboni: “A Neural Network Component for an Intrusion Detection System” - Proceedings of the 1998 National Information Systems Security Conference (NISSC'98) October 5-8 1998. Arlington, VA.

[3] LiMin Fu: “A Neural Network Model for Learning Rule-Based Systems” - Proceedings of the 1992 International Joint Conference on Neural Networks, I-343:I- 348. [R, L].

[4] H. Teng, K. Chen, S. Lu: “Adaptive Real-time Anomaly Detection Using Inductively Generated Sequential Patterns” - Proceedings of the IEEE Symposium on Research in Security and Privacy, pages 278-284, Oakland CA, May 1990.

[5] T. Lane, C. Brodley: “Approaches to Online Learning and Concept Drift for User Identification in Computer Security” - Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, pp. 259-263 (1998).

[6] J.Anderson: “Computer Security Threat Monitoring and Surveillance” - Tech. Rep., James P Anderson Co., Fort Washington, PA, Apr. 1980.

[7] A. Lazarevic, P. Dokas, L. Ertoz, V. Kumar, J. Srivastava, P. Tan: “Cyber Threat Analysis – A Key Enabling Technology for the Objective Force (A case study in Network Intrusion Detection)” - Proceedings 23rd Army Science Conference, Orlando, FL, December 2002.

[8] P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava, P. Tan: "Data Mining for Network Intrusion Detection" - Tutorial at the Pacific-Asia Conference on Knowledge Discovery in Databases, Seoul, April 30, 2003.

[9] C. Warrender, S. Forrest, B. Pearlmutter: "Detecting Intrusions using System Calls – Alternative Data Models" - IEEE Symposium on Security and Privacy (1998).

[10] L. Ertoz, E. Eilertson, A. Lazarevic, P. Tan, P. Dokas, V. Kumar, J. Srivastava: "Detection and Summarization of Novel Network Attacks Using Data Mining" – Minnesota INtrusion Detection System (MINDS) Technical Report, 2003.

[11] R. C. O'Reilly, and Y. Munakata: "Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain", MIT Press, Boston, MA, 2000.

[12] T. Lunt: "Real-Time Intrusion Detection" - Technical report, Computer Science Laboratory, SRI international, Menlo Park, CA, February 1992.

[13] J. Frank: "Artificial Intelligence and Intrusion Detection – Current and Future Directions" - Technical Report, Division of Comp. Science, University of California at Davis, 1994.

[14] J. Ryan, M. Lin, R. Miikkulainen: "Intrusion Detection with Neural Networks" - *AI* Approaches to Fraud Detection and Risk Management: Papers from the 1997 AAAI Workshop (Providence, Rhode Island), pp. 72-79. Menlo Park, CA: AAAI.

[15] M. Mahoney, P. Chan: "Learning Rules for Anomaly Detection of Hostile Network Traffic" - Proceedings of the Third IEEE International Conference on Data Mining, p.601, November 19-22, 2003

[16] A. Ghosh, A. Schwartzbard, M. Schatz: "Learning Program Behavior Profiles for Intrusion Detection" - Reliable Software Technologies Corporation, 1999.

[17] J. Cannady: "Artificial Neural Networks for Misuse Detection" - National Information Systems Security Conference (1998).

[18] L. Ertoz, A. Lazarevic, E. Eilertson, P. Tan, P. Dokas, V. Kumar, J. Srivastava: "Protecting Against Cyber Threats in Networked Information Systems" - SPIE Annual Symposium on AeroSense, Battlespace Digitization and Network Centric Systems III, Orlando, FL (2003)

[19] L. Lankewicz, M. Benard: "Real-time Anomaly Detection Using a Nonparametric Pattern Recognition Approach" - Proceedings of the of 7th Computer Security Applications conf., San Antonio, TX, 1991.

[20] J. Shavlik, M. Shavlik: "Selection, Combination and Evaluation of Effective Software Sensors for Detecting Abnormal Usage on Computers Running Windows NT/2000" – Shavlik Technologies Apr. 2002.

[21] K. Ilgun: "USTAT: A Real-time Intrusion Detection System for UNIX" Proceedings of the IEEE Symposium on Security and Privacy,Oak-land, CA, May 1993.

[22] N. Bose, P. Liang: "Neural Network Fundamentals with Graphs, Algorithms, and Applications" – McGraw-Hill, 1996.

[23] W. W. Cohen, and Y. Singer, "Simple Fast & Effective Rule Learner", AAAI/IAAI 1999, pp. 335-342.

[24] S. R. Gaddam, V. V. Phoha, and K. S. Balagani, "K-Means+ID3: A Novel Method for Supervised Anomaly Detection by Cascading K-Means Clustering and ID3 Decision Tree Learning Methods", IEEE Transactions on Knowledge and Data Engineering, 19(3), 2007, pp 345-354.

[25] L. KUNCHEVA, "Combining Pattern Classifiers: Methods and Algorithms". John Wiley & Sons, Hoboken, NJ, 2004.

[26] S. Koutsoutos, I. T. Christou, and S. Efremidis, "A Hybrid ANN-based Intrusion Detection System for Network-Initiated Attacks", Proceedings of the 3rd IFIP Conference on Artificial Intelligence Applications and Innovations, Athens Greece, June, 2006.

[27] R. QUINLAN, "C4.5 Programs for Machine Learning". Morgan Kaufmann Publishers, San Francisco, CA, 1993.

Prediction Models of an Indoor Smart Antenna System Using Artificial Neural Networks

Nektarios MORAITIS [a] and Demosthenes VOUYIOUKAS [b]
[a] *Mobile Radiocommunications Laboratory, National Technical University of Athens, 9 Heroon Polytechniou str. 15773, Zografou, Athens, Greece, morai@mobile.ntua.gr*
[b] *Dept. of Information and Communication Systems Engineering, University of the Aegean, Karlovasi 83200 Samos, Greece, dvouyiou@aegean.gr*

Abstract. This study presents the prediction propagation paths of angle of arrivals (AoAs) of a Smart Antenna System in an indoor environment utilizing Artificial Neural Networks (ANN). The proposed models consist of a Multilayer Perceptron and a Generalized Regression Neural Network trained with measurements. For comparison purposes the theoretical Gaussian scatter density model was investigated for the derivation of the power angle profile. The antenna system consisted of a Single Input Multiple Output (SIMO) system with two or four antenna elements at the receiver site and the realized antenna configuration comprised of Uniform Linear Arrays (ULAs). The proposed models utilize the characteristics of the environment, the antenna elements and their spacing for prediction of the angle of arrivals of each one of the propagation paths. The results are presented towards the average error, standard deviation and mean square error compared with the measurements and they are capable for the derivation of accurate prediction models for the case of AoA in an indoor millimeter wave propagation environment.

Keywords. ANN, SIMO, Millimeter band, Smart Antenna

Introduction

Smart Antenna Systems [1] and especially MISO (Multiple Input Single Output) [2] or SIMO (Single Input Multiple Output) [3] systems have already been evaluated for the optimization of wireless system performance. In millimeter wave frequencies the propagation modeling, apart from the known empirical models, can be realized based on geometrical optics using ray-tracing theory. In the 60 GHz region the diffraction phenomenon can be neglected, and the sum of the direct ray and the reflected rays is enough to describe the behavior of the propagation channel with great accuracy. The prediction of the field strength is a very complex and difficult task. In most cases, there are no clear line-of-sight (LOS) conditions between the transmitter and the receiver. Generally, the prediction models are classified as empirical [4] or theoretical [5], or a combination of these two [6]. However, the main problem of the classical empirical models is the unsatisfactory accuracy, while the theoretical models lack in computational efficiency.

During last years, Artificial Neural Networks (ANN) have experienced a great development. ANN applications are already very numerous. Classificators, signal processors, optimizers and controllers have already been realized. Although there are several types of ANN's all of them share the following features [7]: exact analytical formula impossible; required accuracy around some percent; medium quantity of data to process; environment adaptation that allows them to learn from a changing environment and parallel structure that allows them to achieve high computation speed. All these characteristics of ANN's make them suitable for predicting field strength in different environments and furthermore angle of arrivals (AoA).

The prediction of field strength and AoA can be described as the transformation of an input vector containing topographical and morphographical information (e.g. path profile) to the desired output value. The unknown transformation is a scalar function of many variables (several inputs and a single output), because a huge amount of input data has to be processed. The inputs contain information about the transmitter and receiver locations, surrounding buildings, frequency, etc while the output gives the propagation loss for those inputs. From this point of view, research in propagation loss modeling consists in finding both the inputs and the function that best approximate the propagation loss. Given that ANN's are capable of function approximation, they are useful for the propagation loss and angle of arrival modeling. The feedforward neural networks are very well suited for prediction purposes because do not allow any feedback from the output (field strength or path loss) to the input (topographical and morphographical data).

In this paper, the presented studies develop a number of Multilayer Perceptron Neural Networks (MLP-NN) and Generalized Radial Basis Function Neural Networks (RBF-NN) based models trained on extended data set of propagation path loss measurements taken in an indoor environment. The smart antenna measurement system was a SIMO one where a continuous wave (CW) signal at 60 GHz was transmitted from a fixed base station to a fixed receiver, comprised of two or four antenna elements. The signal envelope as a function of time was recorded. The performance of the neural network based models is evaluated by comparing their prediction, standard deviation and mean square error (MSE) between their predicted values and measurements data. Also, a comparison with the results is obtained by applying the Gaussian model.

The remainder of this paper is organized as follows. Section 1 deals with the ANN overview describing and explaining the behavior of the two NN utilized models. In Section 2, an analytically description of the geometry of the measurement environment under consideration is presented along with the measurement procedure. In Section 3, the NN prediction models are implemented analytically describing the implementation method and the prediction results are presented in terms of measured Power Angle Profile (PAP), taking also into consideration the Gaussian model. Finally, Section 4 is devoted to conclusions derived by the prediction procedure.

1. The ANN Overview

1.1. Multilayer Perceptron Neural Network (MLP-NN)

Figure 1 shows the configuration of a multilayer perceptron with one hidden layer and one output layer. The network shown here is fully interconnected. This means that each neuron of a layer is connected to each neuron of the next layer so that only forward

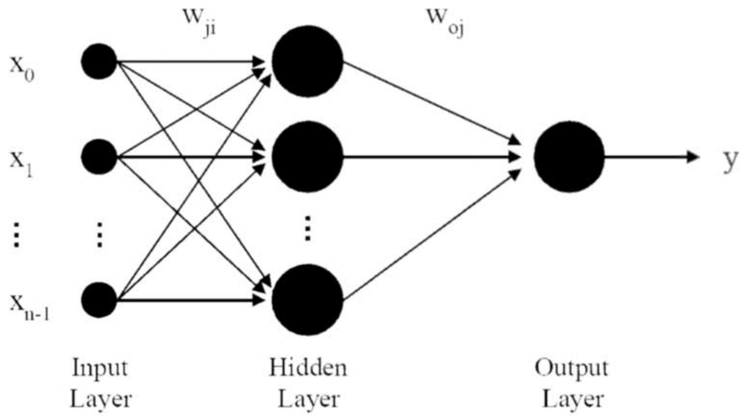


Figure 1. MLP-NN configuration.

transmission through the network is possible, from the input layer to the output layer through the hidden layers. Two kinds of signals are identified in this network:

- The function signals (also called input signals) that come in at the input of the network, propagate forward (neuron by neuron) through the network and reach the output end of the network as output signals;
- The error signals that originate at the output neuron of the network and propagate backward (layer by layer) through the network.

The output of the neural network is described by the following equation:

$$y = F_o\left(\sum_{j=0}^{M} w_{0j}\left(F_h\left(\sum_{i=0}^{N} w_{ji}x_i\right)\right)\right) \tag{1}$$

where

- w_{0j} represents the synaptic weights from neuron j in the hidden layer to the single output neuron,
- x_i represents the *i-th* element of the input vector,
- F_h and F_o are the activation function of the neurons from the hidden layer and output layer, respectively,
- w_{ji} are the connection weights between the neurons of the hidden layer and the inputs.

The learning phase of the network proceeds by adaptively adjusting the free parameters of the system based on the mean square error E, described by Eq. (2), between predicted and measured path loss for a set of appropriately selected training examples:

$$E = \frac{1}{2}\sum_{i=1}^{m}\left(y_i - d_i\right)^2 \tag{2}$$

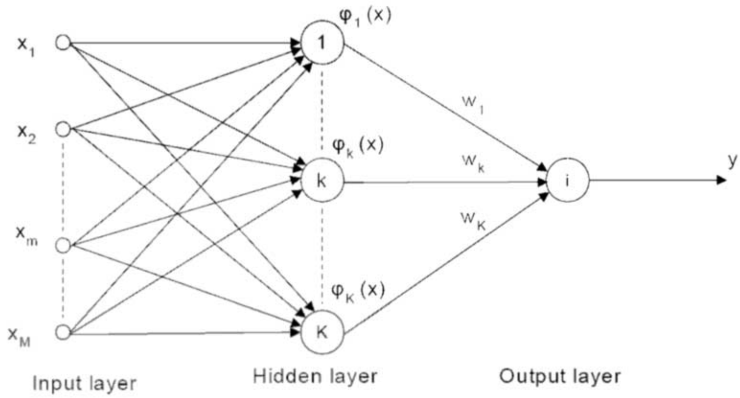


Figure 2. RBF-NN architecture.

where y_i is the output value calculated by the network and d_i represents the expected output.

When the error between network output and the desired output is minimized, the learning process is terminated and the network can be used in a testing phase with test vectors. At this stage, the neural network is described by the optimal weight configuration, which means that theoretically ensures the output error minimization.

1.2. Generalized Radial Basis Function Neural Network (RBF-NN)

The Generalized Radial Basis Function Neural Network (RBF-NN) is a neural network architecture that can solve any function approximation problem. The learning process is equivalent to finding a surface in a multidimensional space that provides a best fit to the training data, with the criterion for the "best fit" being measured in some statistical sense. The generalization is equivalent to the use of this multidimensional surface to interpolate the test data.

As it can be seen from Fig. 2, the Generalized Radial Basis Function Neural Network (RBF–NN) consists of three layers of nodes with entirely different roles:

- The input layer, where the inputs are applied.
- The hidden layer, where a nonlinear transformation is applied on the data from the input space to the hidden space; in most applications the hidden space is of high dimensionality.
- The linear output layer, where the outputs are produced.

The most popular choice for the function φ is a multivariate Gaussian function with an appropriate mean and autocovariance matrix. The outputs of the hidden layer units are of the form:

$$\varphi_k\left[x\right] = \exp\left[-\frac{\left(x-v_k^x\right)^T\left(x-v_k^x\right)}{2\sigma^2}\right] \tag{3}$$

when v_k^x are the corresponding clusters for the inputs and v_k^y are the corresponding clusters for the outputs obtained by applying a clustering technique of the input/output data that produces K cluster centers [8]. v_k^x and v_k^y are defined as:

$$v_k^x = \sum_{x(p) \in \text{cluster k}} x(p) \tag{4}$$

$$v_k^y = \sum_{y(p) \in \text{cluster k}} y(p) \tag{5}$$

The outputs of the hidden layer nodes are multiplied with appropriate interconnection weights to produce the output of the GRNN. The weight for the hidden node k (i.e., w_k) is equal to:

$$w_k = \frac{v_k^x}{\sum_{k=1}^{K} N_k \exp\left[-\frac{d\left(x, v_k^x\right)^2}{2\sigma^2}\right]} \tag{6}$$

where N_k is the number of input data in the cluster centre k, and

$$d\left(x, v_k^x\right) = \left(x - v_k^x\right)^T \left(x - v_k^x\right) \tag{7}$$

2. Measurement Environment and Procedure

The measurement took place in a typical office environment as indicated in Fig. 3. As it is evident, with dark shaded colours are presented the furniture (wooden or metal closets and workbenches) that are above the direct path between the transmitter (height 1.6 m) and receiver (height 1.6 m) and would potentially block the signal propagation (apart from the partitions). With lighter colours are depicted furniture surfaces (e.g., desks) that do not obstruct the direct signal propagation. The closets have a height of 2 m, the workbenches 1.4 m and the desks (including the computers) 1.15 m.

Surface A is an external wall with windows in consecutive order separated by concrete pillars. Each window has 5 mm glass with aluminium frame. The windows have metallic window shades in front which during the measurements were down. Surface B is an internal thick wall made of brick and covered with plaster and paint on both sides. The total wall thickness is 23 cm.

The floor is made of concrete and covered with marble and a thin antistatic plastic layer. The ceiling is made of concrete with a total height of 3.4 m. Approximately 60 cm below the ceiling a metal frame structure suspends, holding the fluorescent light tubes. The Partition is made of 5 mm glass with aluminium studs every 1.5 m. The internal doors consist of 5 mm glass with aluminium frame. The internal doors during the

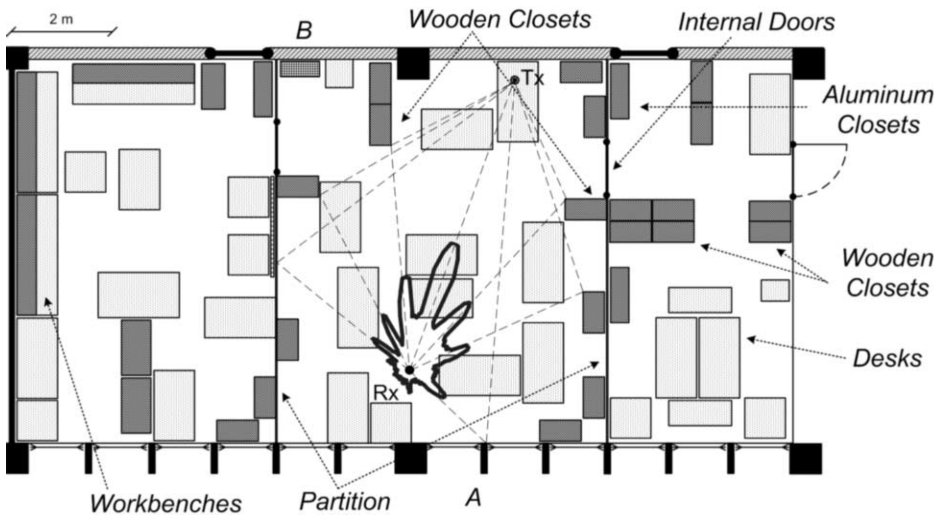


Figure 3. Measurement environment and superimposed the derived Power Angle Profile.

measurements were closed. The wooden closets have a thickness of 42 cm and made of 1.5 cm wooden chipboard covered with melamine and 5 mm glass as a front cover. Similarly the metal closets are 36 cm thick and consist of 3 mm galvanized steel with 5 mm glass as a front cover. We should mention though, that the total true material thickness, which the signal penetrates, is 2 cm for the wooden and 8 mm for the metal closet.

The measurement was accomplished by transmitting a continuous wave (CW) signal at 60 GHz, from a fixed base station to a fixed receiver, and recording the signal envelope as a function of time. Details for the measurement setup can be found in [9]. The transmitter was fixed at 1.6 m above the floor, at position Tx shown in Fig. 3, whereas the output power was +10 dBm. The receiver hardware is located on a trolley, which was stationary at the measurement position. The distance between the transmitter and receiver was 6 m. After amplification, the received signal is down-converted to 300 MHz IF and fed to a commercial receiver. The input to the automatic gain control (AGC) of the receiver is then sampled at 2 kHz and the data values were stored to a portable PC. The receiver had a noise floor of –90 dBm. For this measurement, an omnidirectional with 0 dBi gain was used as the transmitter antenna, and a horn antenna with 35 dBi gain was used as the receiver antenna. Both antennas are vertically polarized. The half power beamwidth of the horn antenna was 4^{o} in azimuth and 3^{o} in elevation. The directional receive antenna was fixed at 1.6 m above the floor. When a highly directional antenna is used, the system provides high spatial resolution to resolve multipath components with different AoAs.

During the measurements, a mechanically steered directional antenna was used to resolve multipath components. An automated system was used to precisely position the receiver antenna along a linear track and then rotate the antenna in the azimuthal direction. At each position, the receiver antenna is rotated in azimuth from 0 to 360^{o} with a step size of 5^{o} and power was recorded at each of the 72 angular steps. Then, a local average is calculated from the measurement results at four different positions along the linear track being $\lambda/2$ apart. The local average helps to remove any residual small-scale or time-varying fading that may occur at individual positions. The precisions of the track and spin positions are better than 1 mm and 1°, respectively.

From the aforementioned measurement procedure, the PAP of a SISO channel can be derived. Consequently, if we know the Power Angle Profile (PAP) of a SISO channel, we can calculate the channel matrix of a SIMO channel multiplying the array response vector at the receiver. The PAP of a SISO channel can be yielded by either PAP measurements between fixed transmit and receive terminals, a properly trained NN model and, a theoretical model (e.g. Gaussian model). Furthermore, the aforementioned PAP extraction methods can be used in order to calculate the channel matrix for the antenna element configuration (ULA) and then we can easily estimate the channel capacity of the system.

3. Prediction Models' Implementation

The goal of the prediction is not only to produce small errors for the set of training examples but also to be able to perform well with examples not used in the training process. This generalization property is very important in practical prediction situation where the intention is to use the propagation prediction model to determine the angle of arrival of potential transmitter locations for which no or limited measured data are available.

The selection of the set of training examples is very important in order to achieve good generalization properties [7,10]. The set of all available data is separated in two disjoint sets that are training set and test set. The test set is not involved in the learning phase of the networks and it is used to evaluate the performance of the neural model. An important problem that occurs during the neural network training is the overadaptation that is the network memorizes the training examples and it does not learn to generalize the new situations. In order to avoid overadaptation and to achieve good generalization performances, the training set is separated in the actual training subset and the validation subset, typically 10–20% of the full training set [7]. In order to make the neural network training process more efficient, the input and desired output values are normalized so that they will have zero mean and unity standard deviation. With the intention of establishing the optimum configuration of the MLP neural network, networks with different architectures and different training algorithms were investigated. The results presented here refer to the optimum MLP-NN for each prediction case.

Since the purpose is to train the neural networks to perform well for all the routes, we should build the training set including points from the entire set of measurements data. For training and test purpose we have used the same number of patterns as in the prediction models for the indoor environment. Various inputs for the neural network were taken into consideration, consisting of position, gain and height of the transmitter site, the sector where the receiver antenna is located, the type of interior where the receiver is located, distances between the transmitter and receiver, received multipaths rays and penetration parameters such as number of penetrated walls and windows and accumulated losses.

The input parameters that describe the transmitter and receiver site are quantized so the effect of each parameter is more obvious for the neural network. For example, in order to describe the type of interior where the receiver is located, parameters like size of the corridors are quantized as follows: 1 for the large corridor and 0.3 for the medium corridor. The attenuation factors for different types of walls intervening between

Table 1. Prediction results of the ANN models' implementation

Model	Average Error [dB]	Standard Deviation [dB]	Mean Square Error [dB]
RBF-NN	3.8	2.0	4.3
MLP-NN	5.2	3.7	6.5
Gaussian Model	6.4	3.7	8.1

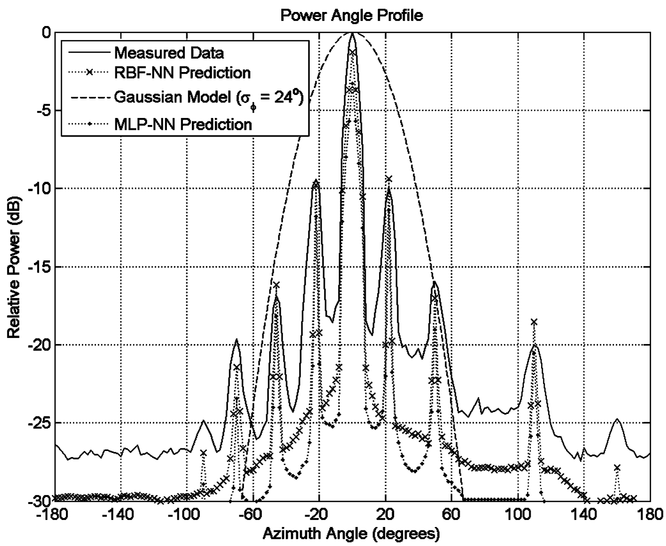


Figure 4. Comparison between the measured PAP, RBF-NN, MLP-NN prediction, and theoretical Gaussian model.

transmitter and receiver, as well as the loss for glasses were used as reported in [11] for this particular type of building. All parameters are normalized to the range [–1, +1]. The output layer of the Artificial Neural Network consists of one neuron that provides the received power.

A data set of 298 patterns, that represents 20% from all available patterns, was used for training purpose. A set of 1186 patterns was used to test the model. In order to train the NN model the measured PAP was used. In Table 1, the average error, the standard deviation and the mean square error are presented, obtained from the training set by the proposed Multilayer Perceptron Neural Network and the Generalized Regression Neural Network. Figure 4 presents the measured Power Angle Profile (PAP) together with the results derived by the MLP-NN and the RBF-NN predictions. As it is evident the results between the measured and the predicted PAP are very good with the Mean Square Error (MSE) equals to 6.5 dB for the MLP-NN model and 4.3 dB for the RBG-NN model. Furthermore, the theoretical Gaussian model for angular profile prediction is utilized for comparison reasons and presented also in Table 1 and Fig. 4. The Gaussian model is given by [12]:

$$PAP(\phi) = \frac{1}{2\pi\sigma_\phi^2}\exp\left[-\frac{\phi^2}{2\sigma_\phi^2}\right]. \tag{8}$$

The measured angular spread σ_ϕ was calculated 24° hence the same value will be used in Eq. (8). The measured angular spread is calculated by [13]:

$$\sigma_\phi = \sqrt{1-\frac{\|F_1\|^2}{\|F_2\|^2}}\ ,\ F_n = \int_0^{2\pi} p(\theta)\exp(jn\theta)d\theta \tag{9}$$

where F_n (n =1 or 2) is given by [13], and $p(\theta)$ is the measured PAP. The MSE between the measured PAP and the Gaussian model was found equal to 8.1 dB. All the results are summarized in Table 1.

From Fig. 4 it is clear that the prediction of the trained NN models is very good, whereas the best results are yielded by the RBF-NN model. On the other hand the Gaussian model provides greater errors than the other two cases because it is not so accurate, and takes into account a smaller range of azimuth angle.

4. Conclusions

This study examined the applicability of the neural networks for the prediction of angle of arrivals in an indoor smart antenna system. The data measurement of an indoor environment using multiple element antennas comprising of linear and uniform array antennas at the millimeter wave band of 60 GHz, were taken into consideration for training purposes of the NN. Two NN models (RBF and MLP) were considered for the derivation of the prediction models as well as the Gaussian theoretical model is evaluated for comparison purposes. The main advantage of the proposed NN models is that the models should be easily adjusted to some specific environments and complex propagation conditions. The results are depicted in terms of average error, standard deviation and mean square error compared with the measurements and showed very good accuracy. The MSE between the measurements and the NN-models was found 6.5 dB for the MLP-NN model and 4.3 dB for the RBG-NN model. The Gaussian model provides greater errors because it takes into account a smaller range of azimuth angle. High accuracy can be obtained, because the NNs are trained with measurements inside buildings and thus include realistic propagation effects considering parameters which are difficult to include in analytic equations. In more specific local cases, the accuracy can be improved by using additional NNs training. Results are always connected with some uncertainty but accuracy is sufficient for prediction purposes.

References

[1] J.H. Winters, Smart antennas for wireless systems, IEEE Personal Commun., 5(1) (1998), 23–27.
[2] K.C. Zangi, L.G. Krasny, Capacity-achieving transmitter and receiver pairs for dispersive MISO channels, IEEE Trans. on Wireless Communications, 2(6) (2003), 1204–1216.

[3] J.K. Tugnait, A multidelay whitening approach to blind identification and equalization of SIMO channels, IEEE Trans. on Wireless Communications, 1(3) (2002), 456–467.
[4] M. Hata, Empirical formula for propagation loss in land mobile radio services, IEEE Trans. on Vehicular Technology, 29(3) (1980), 317–325.
[5] J. Walfisch, H.L. Bertoni, A theoretical model of UHF propagation in urban environments, IEEE Trans. On Antennas and Propagation, 36(12) (1988), 1788–1796.
[6] G.K. Chan, Propagation and coverage prediction for cellular radio systems, IEEE Trans. on Vehicular Technology, 40(4) (1991), 665–670.
[7] S. Haykin, Neural Networks: A Comprehensive Foundation, IEEE Press, McMillan College Publishing Co., 1994.
[8] W. Honcharenko, H.L. Bertoni, J.L. Dailing, J. Qian, and H.D. Yee, Mechanisms Governing UHF Propagation on Single Floors in Modern Office Buildings, IEEE Trans. on Vehicular Technology, 41(4) (1992), 496–504.
[9] N. Moraitis, P. Constantinou, Indoor channel measurements and characterization at 60 GHz for wireless local area network applications, IEEE Trans. on Antennas and Propagation, 52(12) (2004), 3180–3189.
[10] C. Christodoulou, and M. Georgiopoulos, Applications of Neural Networks in Electromagnetics, Artech House, 2001.
[11] N. Moraitis, P. Constantinou, Millimeter Wave Propagation Measurements and Characterization in an Indoor Environment for Wireless 4G Systems, Proc. PIMRC'05 (2005).
[12] R. Janaswamy, Angle and time of arrival statistics for the Gaussian scatter density model, IEEE Trans. Wireless Commun., 1(3) (2002), 488–497.
[13] H. Hu, V. Kukshya, and T.S. Rappaport, Spatial and temporal characteristics of 60-GHz indoor channels, IEEE J. Select. Areas Commun., 20(3) (2002), 620–630.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Interoperable cross media content and DRM for multichannel distribution

Pierfrancesco Bellini, Ivan Bruno, Paolo Nesi, Davide Rogai, Paolo Vaccari
Distributed Systems and Internet Technology Lab,
Department of Systems and Informatics, University of Florence, Florence, Italy
nesi@dsi.unifi.it, http://www.axmedis.org

Abstract. Business and final users are becoming more and more interested in using complex interactive cross media digital content that can be used on different devices from different distribution channels. Furthermore, the present situation provides the usage of different digital rights management, DRM, solutions on different devices and channels. The proposed solution allows to provide interoperable content that can be both used on different devices and managed by different DRM solutions. The paper presents the related results produced by AXMEDIS IST FP6, a research and development integrated project (Automating Production of Cross Media Content for Multi-channel Distribution) partially funded by the European Commission. The focus is on the automated content production of interoperable cross media content with multiple DRM. This allows the management of multichannel solutions: PC (on the internet), PDA, kiosk, mobile phones and interactive TV.

Keywords: cross media content interoperability, DRM interoperability, MPEG-21, OMA, multichannel distribution, production on demand, GRID, content processing profiling.

1. Introduction

The evolution of the digital content market is rapidly changing. Every day users are asking for more functionalities to content and content distributors. New models of content usage based on new forms of content exploiting fully digital content distribution are opening the paths for a larger set of new applications and markets largely beyond the limitations of physical media, especially for business to business (B2B) applications. An evolving set of business models and solutions is proposed on the market for the acceptance by the users. These recent distribution models have been enabled by a large set of new technologies grounded on content formats, high speed connections and digital transmission, content processing and adaptation algorithms, content protection models and solutions, hardware capabilities, and finally the new solutions for Digital Rights Management, DRM. See for a general overview [1].

In terms of content format and integrated DRM, a set of solutions is available on the market such as Apple i-Tunes, Microsoft Windows Media DRM, Adobe DRM (http://WWW.Adobe.COM , Intertrust http://www.intertrust.com/, and OMA (Open Mobile Alliance, [2], http://www.openmobilealliance.org); and many others. For a review see [3], [4]. A part of the above mentioned solutions can be used only to control audio visual content distribution, for example Apple I-Tunes. Others are more focused on controlling document distribution; some of them are more suitable to control data access in decoders and STB (set top box) rather than to manage rights (they are frequently called Conditional Access Systems, CAS). A large part of the above mentioned solutions is proposed for the PC environment (e.g., Microsoft Windows DRM, Adobe DRM), other solutions are mainly focused on mobiles (such as OMA). Most of them present relevant limitations on the content formats; since they support only a limited number of media formats and devices. Others present DRM mechanisms which allow exploiting/controlling only a limited number of rights on the acquired digital content and thus allowing to establish only a limited number of business models. For instance, they can control the content usage/playing in the selected platform the content has been bought for and they have no or limited flexibility in controlling rights like (i) burning; (ii) content copying to other devices, mobiles, media centers, (iii) reusing of content for personal audiovisual composition/production, etc.

At present, there is a large number of content formats ranging from basic digital resources (documents, video, images, audio, multimedia, etc.) to integrated *content packages* such as: MPEG-21 [5], [6], WEDELMUSIC [7] (http://www.wedelmusic.org), SCORM [8] (http://www.adlnet.gov/scorm/index.cfm), OMA (http://www.openmobilealliance.org/), etc. These content formats can wrap digital resources in a package with other related pieces of information (e.g., metadata, descriptors, identification codes), making them ready for delivery (streaming and/or downloading) according to the protection/DRM model used. These solutions are much more flexible with respect to proprietary solutions in which the DRM can be applied only to single resources (e.g., audio-visual files or documents). The definitions and usages of content packages allow wrapping different digital resources, whereas some limitations are imposed by the players that have to guarantee a secure environment for all of them. Business and final users are becoming more and more interested in using complex digital content (e.g., interactive content with several kinds of related resources: audio, video, games, images, text, styles, documents, etc., organized in XML, HTML and/or SMIL) [9]. Besides, they expect to receive this kind of content from several different distribution channels and to use it on different devices/tools, according to packaging tools and innovative business models (e.g., renting for a month, pay per play, reselling rights, etc.), thus overcoming the limitations of the traditional physical media and models.

The present state of the art is dominated by the lack of interoperability among the different content formats (that may range from simple resources to cross media), DRM solutions, distribution channels, devices and tools, accounting information, licenses, protection models, certificates, etc. The solution to the interoperability problems may be solved by the creation and adoption of unique international standards. More recently, several additional issues have to be taken into account (i) the proliferation of standardization bodies has also brought along the production of several competitive solutions that are no longer conquering the market with the same effectiveness as it occurred in the past, (ii) the presence of different device capabilities compelling to make changes in the content format, (iii) the complexity of cross media content which includes in a unique content package many kinds of digital resources and information, and (iv) the needs of using different DRM solutions on different content and devices. The combination of these aspects increased enormously the complexity of solutions for content interoperability.

The focus of this paper is on the problems lying behind any enabling of content interoperability among multiple channels of content distribution. It implies mainly interoperability among content tools, formats, and DRM solutions. The study and the solutions have been elaborated and integrated in AXMEDIS (Automating Production of Cross Media Content for Multi-channel Distribution) project and framework. AXMEDIS is a large research and development Integrated Project FP6 of the European Commission (http://www.axmedis.org), [10], [11], [12]. The AXMEDIS consortium consists of leading European digital content producers, integrators, aggregators, and distributors, together with information technology companies and large research groups. One of the main objectives of AXMEDIS is to create a framework for the automated production and distribution of cross-media contents over a number of different distribution channels (e.g., P2P, networked PC, PDA, kiosk, mobile phone, i-TV, etc.) with DRM. In order to satisfy these needs, several new enabling technologies have been developed such as the definition of content models integrated with a flexible DRM, dynamic content adaptation, content production on demand, content licensing, license adaptation, content interoperability, DRM interoperability, license processing, languages for multimedia content processing and distribution, algorithms for automated content production and formatting, etc. In the AXMEDIS architecture a number of different standards for content formats and DRM are supported and many other models can be easily integrated.

The paper is organized as follows. Section 2 presents a generic Digital Rights Management Scenario for content distribution. In Section 3, the main problems related to the interoperability of the cross media protected content are presented and discussed. Section 4 shows the AXMEDIS content processing GRID platform. In Section 5 there is an example of the usage of the AXMEDIS solution so as to realize interoperable multi-channel distribution. Conclusions are drawn in Section 6.

2 Content distribution and DRM overview

Before presenting *cross media content interoperability* it is better to have a general overview of the state of the art as to the content distribution with its integration with DRM solutions.

Figure 1 depicts the typical content production and distribution scenario, which summarizes the most relevant needed phases and components from content packaging to content distribution. The distributor has established a contract/agreement with the content Producer. The content distributor may do business by granting the access to content to consumers by means of specific licenses. These produced licenses describe the set of rights granted to consumers (rights are actions that can be performed on the content, such as play, print, copy, show, etc., with some constraints [13]). Therefore, the final user may purchase for some specific content some rights with some conditions (period of use, price, etc.), according to the agreed business model: renting, pay per play, flat rate, etc.

The digital resources (single or cross media) are packaged so as to produce a distribution package to be distributed via portal. The content in the package may be protected by using some protection tools. The corresponding information to unprotect/open each digital resource and/or the package has to reach the final user device when the user is authorized to perform the action. This piece of information is typically called Protection Information or IPMP, Intellectually Property Management and Protection information, like in MPEG-21 [14], [3], [15], AXMEDIS protection information [16], etc. The Protection Information is represented in Figure 1 and below with a circle containing a "P". The Protection Information may be directly transferred to the License Server providing this information to the player/device when needed. In alternative, it may be included into the License and to this end it has to be transferred to the License Production module/server.

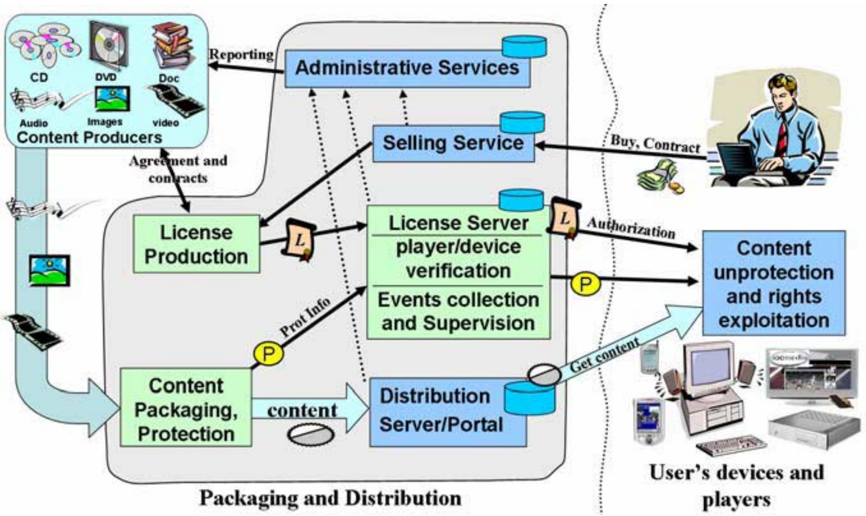


Figure 1 – A simplification of a DRM based Content Production and Distribution model and architecture

Once the contract with the Final User is signed, the Distributor has to issue a corresponding digital license "L" to allow the user to exploit the acquired rights. The license formalizes the digital contract between the entity granting the access to the content rights (namely by providing access to some specific sets of rights) and the other party who is supposed to use/exploit them. It codes in some way the business model and it includes the rights which can be exploited by the users on a specific content. It is formalized in some formal and consistent language such as OMA DRM, XrML (http://www.xrml.org/) [17], Microsoft Windows Media DRM MPEG-21 REL [18], as in the AXMEDIS native DRM. The rights' expression language is based on a dictionary of words for defining the semantics of rights and their related constraints, or those proposed by MI3P. According to the license model, different kinds of business models can be implemented: pay per view/play/print, monthly rate subscription, all you can eat, renting, pay per download, pay per burning the CD, transcoding for migrating content on different devices or distributing them on different channels, etc. Those business models may support different constraints (number of plays, temporal windows where the rights can be exploited in pay per play, expiration date, countries, etc.) and additional features such as allowing to make copies, building a collection, previewing with-out paying for limited time, try and buy, etc.

Once the content package has reached the consumer's player, he/she may perform some actions to exploit the acquired rights, for example a play. In order to allow only authorized actions, the player/device has to verify if that action can be authorized on the basis of the license, for example contacting the License Server. The License Server has to process the license database to verify if the authorization can be granted. When the player/device makes a request to the License Server, it has to be authenticated (to prevent access to unregistered customers) and verified (namely, by controlling if it has come under some attacks). Therefore a protected channel is open from the Player to License Server. Then, once the grant authorization (according to what is specified in the license) has been received, the player/device has to get/receive the Protection Information to get access to the content in order to exploit eventually the related rights.

Each time a right is exploited, the involved actors may need to have written evidence of it in their Administrative Services. This piece of information can be easily recovered in architectures and solutions forcing the player/device to get into contact with the License Server for each authorization. As a first step, a sort of Action Log record (Event Report in MPEG-21) can be produced by the player/device and it can be communicated via the so called return channel (for example, a record of a play action).

In some cases, the producers and/or the distributors may be interested in revoking some of the licenses they have produced, and/or in revoking the usage of some objects if their safeness has been compromised. A verification process is activated to each connection to detect possible attacks.

According to the above discussion, what has been highlighted is that formats for content, license, protection information, DRM, and action log, do strongly affect the architecture and protocols of a multichannel distribution and can create differences that may limit or make impossible any cross media content interoperability among the different channels and DRM models.

3. Interoperability of Cross Media Protected Content

Content interoperability has a precise meaning for end users: *to be given the possibility of using the same content on different devices and distribution channels*. This implies the possibility of purchasing a given content (e.g., a video of Batman, the last song of Madonna, a complex cross media content like a DVD or an educational course) and, on the basis of the contract/license, to have the possibility of enjoying the same content on the different devices owned by the user: TV, DVD player, mobile phone, MP3 player, car player, game station, etc. From this perspective the interoperability among devices, content format and DRM represents a very relevant feature. As to content format, what is meant is cross media content model including: content identification, metadata, descriptors and digital resources, glued data such as XML, SMIL or HTML.

Let us try to keep separate the content from DRM interoperability, even if a complete separation is not possible, because when the content is protected, additional limitations are imposed.

Several different approaches to cope with the content interoperability problem from a source and a destination device are possible and they may consist in:

1. allowing the content to move from one final user's device to another's. This is not a problem if the device/player source and destination can read/play/execute the same content format; otherwise the content has to be adapted/transcoded. To this end, the source tool/device has to adapt the content for the destination, or the destination has to adapt the received content -- taking into account at least the device and the user's profiles. This implies that the content is produced in its final version on the source or destination devices and not on the distributor/producer servers;
2. having the distributor producing the content in different formats *before* the protection/distribution of the content for any possible different channels and

devices. This activity can be very expensive, since there are several different combinations of device and network profiles and users' preferences;

3. having the distributor producing the content in different formats according to the device, user and network profiles on the basis of the request of content with a suitable production/adaptation of the content, protection and then distributing it. This solution fits for the production and distribution on demand – e.g., VOD.

In all the above cases, the main enabling technology is the capability of performing the adaptation on the basis of the user, the device and the network profiles. They are defined in AXMEDIS according to MPEG-21 DIA [19]. Many content distributors are supporting the content distribution/production on demand.

The above scenario is complicated by the needs of manipulating protected/DRMed content. The presence of protected content involves working with certified and authenticated players/tools, registered users and licenses which provide the authorizations.

- For case (1), the license for the target/destination player/device should be produced by the source player/device or prepared in advance, by means of production or adaptation, for a range of possible receiving players/devices.
- For case (2), the protection increases the complexity, since each produced content could be protected with the available different protection models, thus increasing the number of different pre-produced content files. In this case, the license has to be produced in advance or provided by the source device/player, for production or adaptation.
- In case (3), the protection is applied on demand and therefore the adaptation can be performed before content protection and the license can be also produced on demand.

Furthermore, the device performing the adaptation and producing the license has to be authorized. Once the DRMed content is used on the target player, the log of the actions performed on license rights can be tracked by the license server. Different license servers may provide collecting societies and content producers with interoperable information in order to monitor the exploitation of rights. See for example DDEX (Digital Data Exchange) (http://www.ddex.net), which is a consortium of music labels which is defining standard for interchanging administrative data among different DRM platforms, mainly for reporting and providing the producing licenses with the formal authorization.

Another relevant complication is due to the usage of cross media content. The adaptation of cross media content has to take into account (i) the format/package to keep together the different information: metadata, identification, license, descriptors, digital resources (images, video, animation, audio, etc.), (ii) the formal model to define the synchronization and relationships among other digital resources. in an AXMEDIS object the cross media content is organized as a set of SMIL and/or HTML files or with one or more MPEG-4 defining the layout of the cross media content rendering. In AXMEDIS, as well as in other platforms, not all the devices can support the same format: MPEG-4 is supported on PC and PDA, SMIL on PC, PDA and mobile devices, whereas HTML only on PC in a protected manner. The production of content on demand implies also to carry out a reasoning on both profiles and content structure and features, so as to decide how to identify the needed steps and parameters for producing the adapted content for the target device/player.

Therefore, as to cross media protected content (e.g., an educational documentary on birds with several SMILs, video, audio, documents, HTML pages, etc.) the adaptation algorithm and its processing have to:

- consider:
  - device, user and network profiles (see [19]);
  - general metadata, descriptors and their related ontology ;
  - digital resources high level descriptors in MPEG-7 and other formats -- e.g., rhythm, genre, topic, keywords, etc.;
  - digital resources features – e.g., size, duration, type, position in the screen (such as header, footer, body, background, etc.), color distribution,
  - digital resources metadata – e.g., description of content contained in the picture, video, etc.;
  - content layout and its related descriptors -- e.g., in SMIL and XSLT;
  - protection models and protection information - -e.g., [15], [16];
  - license and formalized rights for each digital resource.
- produce:
  - metadata, while filtering out those not needed and mapping the present to those that are mandatory according to the destination metadata mapper;
  - content layout, may be shifting from a SMIL template to a different SMIL by using style sheet, or from a HTML to SMIL, or from SMIL to MPEG-4, etc.;
  - digital resources and descriptors; reshaping of the whole cross media content (e.g., elimination of some digital resources) if some functional aspects cannot be converted to the formats the device supports . For example converting video in images, animations in videos, etc.;
  - license which filters out rights that cannot be exploited on the target device.

4. AXMEDIS Content Processing GRID platform

The AXMEDIS Content Processing (AXCP) GRID platform provides an infrastructure with a suitable set of tools to automate the above described adaptation process and harmonize it with the other backoffice activities, such as access to servers with databases and web services in general. The AXCP architecture consists in a GRID Engine where processes can be executed on GRID nodes and can be specified through the AXMEDIS Content Processing Script Language which is an extension of ECMA Script language ([9]). It is also supported by specific tools for editing, debugging and script processing and a scheduler for the allocation and control of processes executed on the multiprocessor GRID architecture.

The tools and scripts allow automating all the phases of content production, protection and distribution and they cope with:

- processing of Metadata, mapping/transcoding and adaptation with XSLT profiles associated devices and/or user preferences;
- Reasoning about user, device and network profiles (that are called in short UP, DP and NP, the UP can be recovered from the UID, User ID) so as to identify

parameters for the adaptation according to the standard MPEG-21 DIA Digital Item Adaptation [19],

- processing of digital resources for the
    - extraction of fingerprint, synchronization;
    - adaptation (change in resolution and format, transcoding, etc.
    - estimation of descriptors – e.g. MPEG-7 descriptors for audio, video and documents;
- processing of cross media content structure for
    - reusing other packages for content composition;
    - packaging and protection with MPEG-21 IPMP, OMA DRM;
    - production of SMIL, HTML;
    - Transcoding from SMIL to MPEG-4 [20];
    - layouting with automated selection of suitable SMIL template, and suitable style sheets XSLT;
    - layout optimization based on genetic algorithm;
- processing and production of DRM licenses
    - according to MPEG-21 REL and OMA DRM;
    - verification of consistency among different licenses;
    - transcoding among different license models/formats;
- content protection: application of Protection Information according to MPEG-21 IPMP, AXMEDIS and OMA, allowing the dynamic selection of protection tools and algorithms.
- content and information access, gathering/crawling via AXCP tools and language; from/to content management systems such as ODBC, ORACLE, MSSQL, MySQL, XML databases; from/to operating system files; via communication protocols such as web services, WEBDAV, HTTP, FTP.
- multichannel distribution of content with license and protection information posting according to different business models: internet distribution; satellite data broadcast distribution (towards kiosk, PCs and STBs); mobiles distribution (to smart phones and PDAs); P2P networks (automating both publication and download of digital content from and to a P2P network for B2B distribution).

All the mentioned activities and flows can be managed by the external workflow management systems such as Open Flow, towards single tools and the AXMEDIS content processing platform.

5. Multichannel production and distribution on demand

AXMEDIS framework and tools allow to set up a large range of different architectures and configurations in order to manage and harmonize several distribution channels for production and distribution with on demand response capabilities.

Figure 2 shows an example of a distributor managing three different interoperable distribution channels for content on demand delivery by a unified back-office. The first channel is based on AXMEDIS/MPEG-21 content format (e.g., for PC, and STBs) and it exploits functionalities of the AXMEDIS license server [10]. The second is based on OMA distribution according to the OMA Separate Delivery. The last channel is based on Microsoft Windows Media DRM. These channels may have a separate and/or a shared unified portal with services for content promotion and distribution in streaming and/or download. In Figure 2, the user profile is recovered from the UID, User ID, while DID is the distributor ID; OID is the Object ID, etc.

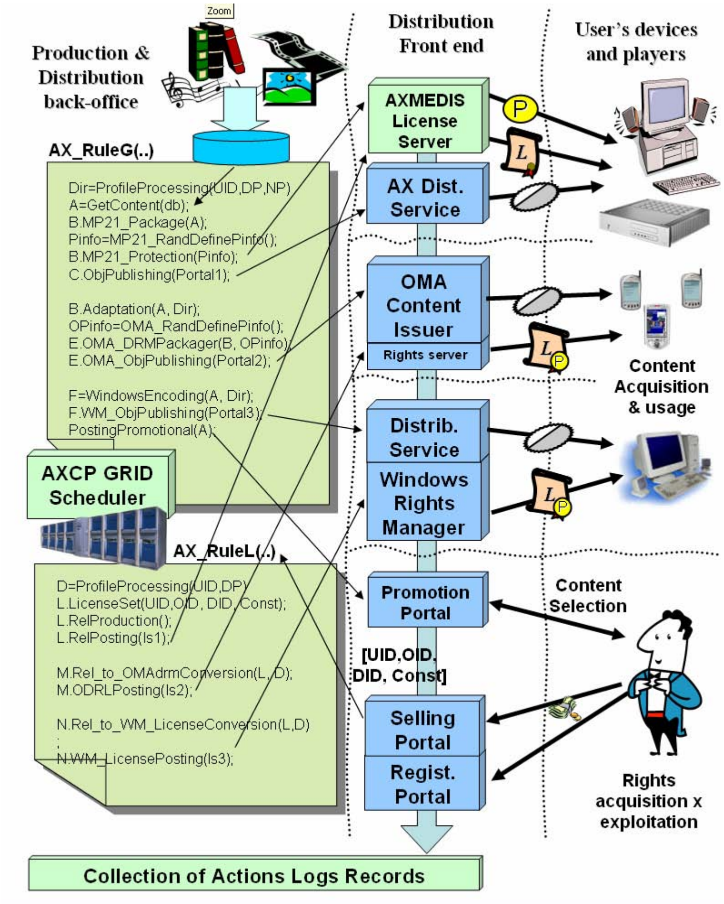


Figure 2 – Multi-channel content distribution on demand, back-office details

The content portals for the different channels are fed with the corresponding content format in an automatic way by the AXCP GRID that manages the back-office. To this end, a set of AXCP Script Rules (see AX_RuleG() in Figure 2) could be used to automate the gathering of content, integration of metadata, adaptation, packaging, protection and posting on front end portals. The rules reported in Figure 2 are a simplification and do not report phases of content authentication, object registration, object ID assignment, etc. It should be remarked that the script has performed the adaptation of content *A* originally used for PC distribution, so as to have the content distributed towards mobiles.

Once the user has decided to buy certain content, the player/device and/or the content itself may redirect the user to get access to the Selling Portal where he/she can define a contract for multiple channel consumption and usage of the selected content (multiple playing). The contract definition may lead to the production of one license up to the full set of them with the corresponding posting on the specific license servers.

For example, the user could be interested in buying the access to a single content on a single device, or he/she may be interested in having the same content on all its own devices (in this case devices are supported by different DRM models). In this latter case, if the distributor has C channels and M objects, C*M licenses have to be produced and the same service has to be accessible to a large number of users per second. To this end, the license production is demanded to AX_RuleL(). The activation is performed by transfering a set of parameters: the object unique ID, the user or device unique ID, the distributor ID (if the selling portal is making this work for more than one distributor) and a set of constraints which define the license parameters: duration, cost, geographic constraints, etc. (mentioned as “Const” in AX_RuleL()).

Firstly, AX_RuleL() creates and posts an MPEG-21 REL license on the License Server, then OMA and Windows Media licenses and information are produced by transcoding the MPEG-21 REL format in the other formats (see AXMEDIS web portal and [21]). Then, the produced licenses are posted on their respective license servers: ls2 and ls3. This approach allows to manage in a unified manner channels based on different content and/or DRM models. If the considered channels are all based on the same DRM model and content format (MPEG-21 or OMA, etc.), as it becomes possible using the AXMEDIS technology, interoperability can be provided as well.

6. Conclusions

This paper reported the interoperable capabilities for content, channel and DRM of the AXMEDIS solution. The solution supports interoperability by allowing the production and adaptation of content, formats/layout, licenses, protection information and processing profiles for user, device and network, among different channels and from different DRM formats. Such different channels may have their specific devices and business models and their specific DRM solutions. This paper described only a part of the whole AXMEDIS framework and its related architecture which is dealing with many other problems and critical points. AXMEDIS framework can support data gathering from accessible Content Management Systems, transform legacy digital content in AXMEDIS objects, content authoring, and it can process them in the production, preserving security level along the whole value chain and therefore creating an integrated and open environment for content production, protection and distribution at both B2B and B2C levels. AXMEDIS is mainly based on MPEG-21 and OMA models and it provides and stimulates the usage and the exploitation of the developed features for creating other AXMEDIS compliant tools and solutions, while making the core aspects and solution accessible in the form of AXMEDIS Framework. More technical information and/or how to make registration or affiliation to the AXMEDIS can be found on www.axmedis.org

Acknowledgements

The authors would like to thank all the AXMEDIS project partners (ANSC, AFI, EUTELSAT, Giunti ILABS, HP Italy, FHGIGD, DIPITA, TISCALI, XIM, ACIT, CPR, EXITECH, University of Leeds, University of Cataluña, University of Reading, etc.), the Expert-User-Group and all the affiliated members for their supporting contributions and collaboration efforts . The authors would like also to express their thanks to the European Commission IST DG-INFSO FP6 for partially funding AXMEDIS.

References

[1] Koushanfar, F., Inki Hong, Miodrag Potkonjak, "Behavioral synthesis techniques for intellectual property protection", ACM Transactions on Design Automation of Electronic Systems (TODAES), Vol.10, Issue 3, ACM Press, July 2005.

[2] Iannella, R., "Open Digital Rights Language (ODRL)", Version 1.1 W3C Note, 19 September 2002, http://www.w3.org/TR/odrl

[3] Lin, E.T., Eskicioglu, A.M., Lagendijk, R.L., Delp, E.J., "Advances in Digital Video Content Protection", Proceedings of the IEEE, Vol.93, N.1, pp.171-183, January 2005,

[4] IEEE Signal Processing Mag.—Special Issue on Digital Rights Management, vol. 21, no. 2, Mar. 2004.

[5] Wang, X., De Martini, T., Wragg, B., Paramasivam M., Barlas C., "The MPEG-21 rights expression language and rights data dictionary", IEEE Transactions on Multimedia, Vol.7, N.3, pp.408-417, 2005.

[6] Burnett, I.S., Davis, S.J., Drury, G. M., "MPEG-21 digital item declaration and Identification-principles and compression", IEEE Transactions on Multimedia, Vol.7, N.3, pp.400-407, 2005.

[7] Bellini, P., Barthelemy, J., Bruno, I., Nesi, P., Spinu, M., "Multimedia Music Sharing among Mediatheques, Archives and Distribution to their attendees", Journal on Applied Artificial Intelligence, Vol.17, N.8-9, pp.773-796, 2003.

[8] Mourad, M., Hnaley, G.L., Sperling, B.B., Gunther, J., "Toward an Electronic Marketplace for Higher Education", Computer of IEEE, pp.58-67, June 2005.

[9] Bellini P., Bruno I., Nesi P., ``A language and architecture for automating multimedia content production on grid", Proc. of the IEEE International Conference on Multimedia & Expo (ICME 2006), IEEE Press, Toronto, Canada, 9-12 July, 2006.

[10] AXMEDIS, "Framework and Tools Specifications", http://www.axmedis.org

[11] Bellini P., Nesi P., "An architecture of Automating Production of Cross Media Content for Multi-channel Distribution", Proc. of the first International Conference on Automated Production of Cross Media Content for Multi-channel Distribution, AXMEDIS 2005, 30 November - 2 December 2005, Florence, Italy, IEEE Computer Society press.

[12] Bellini P., Nesi P., Ortimini L., Rogai D., Vallotti A., ``Model and usage of a core module for AXMEDIS/MPEG21 content manipulation tools", Proc. of the IEEE International Conference on Multimedia & Expo (ICME 2006), IEEE Press, Toronto, Canada, 9-12 July, 2006.

[13] Chellini S., Martini T., Nesi P., ``AXMEDIS: an MPEG-21 based solution for protected cross media content production and distribution", Proc. of the IEEE Joint Conference on E-Commerce Technology (CEC '06) and Ent. Comp., E-Commerce and E-Services (EEE '06) (CEC/EEE 2006) June 2006 Palo Alto, California, USA.

[14] Prados, J., Rodriguez, E., Delgado, J., "Interoperability between different rights expression languages and protection mechanisms", Proc. of the 1st International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution, AXMEDIS 2005, Florence, Italy, 30 Nov.-2 Dec. 2005, IEEE Computer Society press.

[15] MPEG Group MPEG-21 IPMP, "Introducing MPEG-21 IPMP Components", www.chiariglione.org/mpeg/technologies/mp21-ipmp/

[16] Nesi P., Rogai D., Vallotti A., ``A Protection processor for MPEG-21 Players", Proc. of the IEEE International Conference on Multimedia & Expo (ICME 2006), Toronto, Canada, IEEE Press, 9-12 July, 2006.

[17] Rosenblatt B., Dykstra G., "Integrating Content Management with Digital Rights Management" - White Paper, 2003. http://www.xrml.org/reference/CM-DRMwhitepaper.pdf

[18] MPEG Group MPEG-21 REL, "Introducing MPEG-21 REL Components", www.chiariglione.org/mpeg/technologies/mp21-rel/

[19] MPEG Group MPEG-21 DIA, "Introducing MPEG-21 DIA, Digital Item Adaptation", www.chiariglione.org/mpeg/technologies/mp21-dia/

[20] Shao, B., Moro Velazquez, L., Scaringella, N., Singh, N., and Mattavelli, M., ``SMIL to MPEG-4 bifs conversion," Proc. of AXMEDIS'06, IEEE Comp. Soc. Press, 2006, pp.77-84, Dec, 2006.

[21] Delgado J., Prados J., Rodriguez E., "A new Approach for Interoperability between ODRL and MPEG-21 REL," Proc of Second International ODRL Workshop A new Approach for Interoperability between ODRL and MPEG-21 REL, Lisbon, Portugal, 2005.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Video Watermarking and Benchmarking

Sofia TSEKERIDOU
Athens Information Technology (AIT), 0,8km Markopoulou Ave., Peania, 19002 Athens, Greece

Abstract. The rapid expand of Internet and the drastic turn to the digital era have led to the investigation of digital watermarking as a complementary technology to traditional protection mechanisms to ensure digital content protection. Significant research efforts have been reported in the fields of audio and image watermarking, with video and 3D data watermarking following. This chapter provides an in-depth overview of different video watermarking techniques in order to single out the particularities of that field, in relation with artificial intelligence algorithms that may be used as embedding and/or detection counterparts. Furthermore, to complete the picture, the chapter proceeds in presenting benchmarking requirements for objective video watermarking performance evaluation. There is still potential for novel techniques to be explored in conjunction with artificial intelligence approaches to target to optimal video watermarking performance under a variety of attacks. The current chapter will present the essential starting point for the novice researcher in the field.

Keywords. Copyright protection, video watermarking, attack, benchmarking.

Introduction

The rapid growth of Internet and networked multimedia systems in the past decade has raised concerns from the content designers, since multimedia data nowadays can be flawlessly copied and rapidly disseminated at large scale. Encryption and steganography were proved to be insufficient for digital media protection and thus digital watermarking emerged, aiming at embedding auxiliary information into a host digital signal by imposing secure, imperceptible signal changes (with the employment of a special constructed signal, called watermark that is embedded into original content such as image, video, or audio, producing a watermarked signal). Digital watermarking allows the user to manipulate the content.

The focus in this chapter is on digital video watermarking, where the time dimension, and associated content redundancy, enhances the flexibility of the solution space. Available data are greater than image data, a fact that during watermark design is useful both for the designer and the attacker as it supports reliable embedding of auxiliary data using sophisticated temporal masking, but also, allows the attacker to make greater use of correlation that leads to more effective watermark estimation and removal attacks.

There is a great academic and industrial interest on the design of a copyright protection system for MPEG-2 coded video distributed on Digital Versatile Disk (DVDs), employing the digital video watermarking technology [25]. A video watermarking system has also been designed by the Galaxy Group to complement the existing content scrambling system (CSS) that is part of the DVD standard; the technology is now called WaterCast and is being applied in the automatic monitoring of digital video broadcasts [9].

In Figure 1 a general model is provided presenting the end-to-end generic video watermarking process.

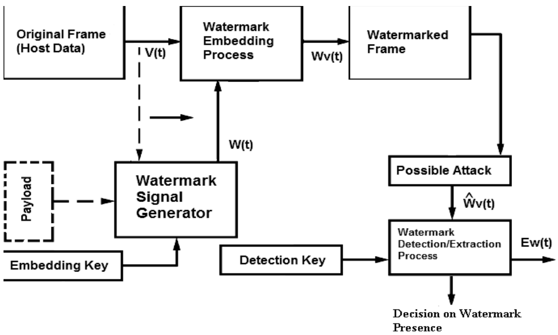


Figure 1. Video Watermarking Process Model

The video watermarking process consists of two main stages:

1. Watermark Generation and Embedding, and
2. Watermark Detection and/or Extraction.

At a first step, the watermark signal generator creates the watermark signal and is provided with an embedding key (the use of a secret such key, to create and embed the watermark is often required for security reasons) and possibly a payload (auxiliary information), and produces $W(t)$, the watermark, to be inserted into the video data. Some watermarking techniques further deploy the original video frame sequence $V(t)$ to achieve more effective watermark embedding (video-dependent watermark generation and embedding). Once the watermark is constructed, it is inserted into the original video frame to produce the watermarked video frame. The specific methods by which the watermark is constructed and embedded is dependent on the watermarking technique. The output of embedding is the watermarked video $W_v(t)$.

In Figure 1, $\widehat{W}_v(t)$ denotes the watermarked video that is possibly attacked and is provided to the detector. If the video has not been attacked, then $W_v(t)$ is identical to $\widehat{W}_v(t)$ for all t. The watermark detector examines the received video and determines if the watermark is present. In Figure 1, $E_w(t)$ denotes the extracted watermark. The detector is also provided with a detection key necessary for the detection of the watermark. A symmetric (private key) watermark uses identical embedding and detection keys, whereas asymmetric (public key) watermarks use distinct but related such keys, similar in concept to public key cryptography.

In the sequel, a review of existing video watermarking techniques is presented in Section 2, after defining potential attack categories targeted at watermarked video data aiming at either removing the watermark or destroying it in Section1. Following the survey and classification of different embedding methodologies in Section 2 and detection/extraction methodologies in Section 3 and performance evaluation measures definition in Section 4, a necessary complement to this chapter is to tackle benchmarking requirements, necessary for the objective performance evaluation of different video watermarking methodologies. Such descriptions are included in Section 5, and conclusions as well as future directions are outlined in Section 6.

1. Types of Attacks

An attack is any processing that aims at impairing watermark detection or communication of information conveyed by it [5]. An attack causes the watermarked video to be altered, intending to remove the embedded watermark or make detection more difficult (*intentional attacks*).

Watermarked data on the other hand is often naturally processed in some way prior to detection. This may include compression, signal enhancement, or digital-to-analog (D-A) and analog-to-digital (A-D) conversion. Thus, we should take into account the case that an embedded watermark is unintentionally impaired by such processing (*non-intentional attacks*).

In this section, we concentrate mainly on presenting *intentional attacks* which are more difficult to compensate for:

- *Simple or noise/waveform attacks*: attempt to modify both host data and watermark without intending to trace and remove the watermark. Linear/non-linear, temporal/spatiotemporal filtering, waveform-based compression, noise addition are included in this category.
- *Geometric attacks* (or *synchronization attacks*): are accomplished by geometrically transforming the data. For video data, this means frame spatial shift, frame rotating and temporal filtering attacks. The watermark is not ultimately removed by the data (as the goal of these attacks is to force the detector to confront a more difficult synchronization problem), so it is possible to successfully detect and recover it. Temporal synchronization attacks in video include frame dropping, insertion, transposition, averaging (temporal interpolation or scaling).
- *Removal attacks*: are focused on detecting the watermark, isolating it from the host data and eventually removing it, without breaking the security of the watermarking algorithm (e.g., without the key used during watermark embedding, as in [4]). This category includes *de-noising*, *quantization* (e.g., for compression), *re-modulation*, and *collusion attacks* (these occur when an attacker obtains collections of video frames that are analyzed or combined with the purpose of producing a non watermarked copy of the original).
- *Forging attacks*: attempt to sabotage the owner's watermark, that is, the attacker wants to forge the original watermark.
- *Statistical attacks:* try to detect the embedded watermark by comparing and finding similarities among a number of watermarked signals that belong to the same owner (whereas collusion attacks involve many copies of a given data set, each signed with a different key).
- *Protocol attacks*: attempt to subvert the security of the watermark; hence attack the entire concept of the watermarking application. They do not directly impact watermark detection.
- *Ambiguity attacks* are based on the concept of invertible watermarks. The malicious forger knows that the data are watermarked. He tries to subtract his own watermark from the watermarked data to later claim to own them and therefore cause uncertainty regarding their true owner. It is essential for copyright protection applications to employ non-invertible watermarks to eliminate the possibility of ambiguity attacks.

Another attack in this category is the *copy attack*: it aims at estimating a watermark from the watermarked data and copies it to some other "target" data without ultimately destroying the watermark or hindering its detection [5].

2. Video Watermarking Methodologies

The embedding process of a watermark into multimedia signals is divided in three categories regarding the entry domain:

1. The watermarks that are embedded in the *spatial/temporal domain*, commonly named as *spatial/temporal/spatiotemporal watermarks*.

Method [1] models a multi-stage watermarking process. The amount of watermarking imposed on a specific stage counterbalances the quality of the final result. Each selected stage is watermarked by selecting a set of "constraints" (that indicate the presence of the author's signature), then using preprocessing of the stage's input and post processing of the stage's output to ensure that a disproportionate number of these constraints are satisfied.

In [2], the embedding process employs meaningful information bits in the luminance mean values of each frame. To deal frame removal attacks, synchronization bits are also integrated alternating with the watermark information bits (in both cases a pseudo random sequence (PRS) generator of different length is used). The watermark PRS values are per frame embedded by modifying the mean luminance value of individual frames.

In [10], a state machine key generator is used to produce time-invariant, time-independent, and time-periodic key schedules, to support temporal synchronization for blind video watermarking. The design of the watermark and its key schedule affect the ease of synchronization. The use of a feature vector allows the key sequence produced to be video-dependent. A video-dependent key schedule can increase the difficulty of inverting the watermark and make it more robust against ownership [26] and copy [27] attacks but may cause temporal synchronization loss due to attacks changing the feature vectors [10].

2. The watermarks incorporated into the *frequency/transform domain*, commonly named as *spectral (or transform-based) watermarks*.

They are integrated within the related transform coefficients. In particular, they involve use of DCT, DWT and DFT or FFT within the embedding process. In video watermarking, significant research efforts are reported to employ 3D DCT, 3D DWT, 3D DFT [24], 3D TWT [12]. The *Temporal Wavelet Transform* (TWT) has scalable temporal resolution.

In [4] the Integer-to-Integer DWT (IIDWT) is used so that both the input and output data to DWT are characterized in integer values. The watermark data is embedded in high frequency regions [4], [6] to improve the watermark effectiveness. Embedding is done in only those coefficients whose norm is greater than a specified threshold in order to achieve perceptual invisibility and robustness against MPEG encoding and re-encoding.

3. The watermarks inserted at the *compressed domain*.

The process of partial or full decompression of video files is skipped thus avoiding quality loss and extra computational cost. The watermarking process can be executed in real time.

In [8] the watermark is embedded directly in an MPEG-2 compressed bit stream by intentionally forcing bit errors. Thus, the error recovery option of a bi-directionally decodable packet (initially used to handle channel errors in [22]) is exploited so as to embed and retrieve the watermark. Reversible VLCs (RVLC) exhibiting error resiliency are implemented due to their two-way decoding directions capabilities. The watermark is encrypted prior to insertion to make it indistinguishable from randomly extracted bits.

3. Video Watermark Detection/Extraction Methodology

A prevalent classification of watermark detection is based on whether the original data are used or not. Specifically, if the watermark detector does not require access to the original signal, the watermarking technique is called *blind*. Otherwise, it is known as *non-blind.*

A cryptographic system used in [1] based on public key encryption prevents the forger from discovering a set of constraints that match the original signature. A single metric, P_c is used, showing the probability of how many of the selected constraints (used to map an author's signature) are satisfied. Basically, P_c is the probability of a non-watermarked solution carrying the watermark. If the value of P_c is very low, the more effective the watermark scheme is. P_c is calculated as a sum of binomials, as shown in [1].

Table 1. Performance of method discussed in [1] under various attacks

Attack	**Performance**
Ambiguity Attacks	Brute-force attacks become computationally infeasible if the proof of authorship threshold is set sufficiently low (e.g., $P_c \leq 2^{-56}$)
Removal Attacks	Possible for an attacker to use tampering methods to remove a signature known to him, or to add an entirely new signature.
Forging Attacks	Successfully prevented when using a key encryption system

In [2], the detection process is based on the cross correlation of the embedded PRSs and the video frame mean luminance sequence. Each video frame has different luminance, so the use of an amplitude limiting filter followed by a whitening filter prior to correlation is proposed, in order to improve the detector performance.

Table 2. Performance of method discussed in [2] towards different kinds of attacks

Attack	**Performance**
Geometric Attacks: (frame spatial shift and frame rotating attacks)	Good performance against frame spatial shift attacks and frame rotating attacks.

Frame Removal Attacks and Temporal Filtering Attacks	Successfully prevented. Detection based on application of a low pass-filter on the luminance values and observation of the corresponding cross-correlation

In [4], the whole watermark extraction process occurs in the decoded video per frame based on a detection key. The averaged watermark obtained is compared to the embedded original watermark in order to ensure that it is exactly the same. Each frame is randomized prior to embedding of the watermark according to the value of a pair of keys, derived from the detection key, to successfully deal collusion and statistical attacks. In [4], it is proved that low values of PSNR (<34dB) are obtained with quite big watermarks (> 25 bits).

Table 3. Performance of method in [4] under MPEG encoding and re-encoding

Attack	**Performance**
MPEG Encoding	Average BER ranges from 20-23% and increases steadily when the watermark length is 24 bits and over. Compared to the common DWT technique, the IIDWT one has a 4-8% better performance whatever the watermark length.
MPEG Re-encoding (2 MPEG encoding iterations)	Overall BER increased about 2-3% compared to single MPEG encoding. More robust than DWT, by an 8% difference in BER rate

In [8] the watermarked VLC must be identified the moment it is decoded. Therefore, the inserted watermark must immediately cause decoding failure in order to trigger reverse decoding that begins from the end-of-packet. To ensure forward detection failure right at the edge of a watermarked VLC, the decoded watermarked bit stream must begin with a sequence of so called flag bits, guaranteeing detection failure. If the packet length is known to the decoder, the last VLC to be recovered on reverse decoding is the same VLC that failed detection on forward decoding. At the end of this process, the watermark bits are extracted, whereas the stream is restored to its initial state. If the packet length is unknown to the decoder, another flag is used, a reverse flag, that causes detection failure on reverse decoding. The watermarking process of compressed media in the VLC domain is inherently fragile since the watermark is vulnerable to re-compression or transcoding. Errors during the detection process-when an incorrect watermark was decoded- are significantly low (they range from 0 -0.15%) and they are not proportionally affected by the file size.

In [10] a model for symmetric blind video watermark detection is described using a detection key. The watermark detector applies a spatial de-correlating filter to reduce the host-signal interference, followed by a correlation detector and comparison with a threshold.

Table 4. Performance of method in [10] towards various synchronization attacks

Attack	**Performance**
Frame Dropping	Poor performance in cases of little temporal redundancy.
Frame Transposition	Performance similar to the frame dropping one.
Frame Insertion	Does not affect watermark detection. Method achieves a detection rate of 100%.

4. Performance Evaluation

One of the metrics widely used to evaluate watermarking schemes is the *False acceptance rate (FAR)*. FAR states the probability that an unknown individual will be falsely 'recognized' as the rightful owner of the reference video data upon presentation of his or her verification data. FAR is dependent on the selected tolerance limit within which the verification and reference data must match for there to be a successful authentication: the lower the tolerance limit, the lower the FAR and the higher the probability of FRR errors.

The *False rejection rate (FRR)* metric states the probability that the rightful owner of the reference data will be wrongly rejected. FRR is dependent on the tolerance limit within which the verification and reference data must match for there to be a successful authentication: the higher the tolerance limit, the lower the FRR and the higher the probability of FAR errors.

A quite simple metric used for evaluation is the *Bit Error Rate (BER)* that denotes the number of error bits divided by the watermark length (BER is calculated per frame).

5. Benchmarking Guidelines for Video Watermarking

Since the complete theoretical analysis of a watermarking algorithm performance with respect to different attacks is rather complicated, the developers of watermarking algorithms refer to the results of experimental testing performed in the scope of some benchmark. The benchmark combines the possible attacks into a common framework and weights the resulted performances depending on the possible application of the watermarking technology.

In image watermarking, several benchmarking tools have been developed to evaluate different methodologies, such as **Stirmark**, **Checkmark** and **Optimark**. *Stirmark* is a generic tool provided with a watermarked input image that generates a number of modified images used to verify watermark existence after a number of attacks. Stirmark proposes combination of different detection results and computation of an overall score. Stirmark has limited potentials for sophisticated image watermarking schemes as it does not properly model the watermarking process. *Optimark* is a benchmarking tool [32] that supports various attacks and employs differentiated performance metrics depending on the type of the detector used (and the output it produces) as well as on the characteristics of the watermarking algorithm.

Likewise, the main design challenges for a video watermarking benchmark framework are listed below:

- Detection performance evaluation using multiple trials employing different sets of data
    - For watermarking schemes that follow frame-by-frame approaches where a different watermark is inserted in each video frame, the chosen set of data must include all the different watermarks used in order to evaluate their robustness,
    - For watermarking schemes that also follow frame-by-frame approaches but embed the same watermark in all video frames, a much smaller set of frames need to be chosen,
    - For more sophisticated watermarking methodologies based on a compression standard or embed a watermark in a three-dimensional (3-D)

transform, the chosen set of data must be carefully chosen in order to ensure that all the possible watermarks used are evaluated and also the range of the testing data excludes the possibility of estimation errors (i.e. the case where the entire set of data is non-watermarked),

- Evaluation of the following detection/decoding performance metrics:,
  - o Bit error rate,
  - o Signal to Noise Ratio (SNR), and Peak Signal to Noise Ratio (PSNR) (as indirect measures for watermarked video quality estimation),
  - o False acceptance rate,
  - o False rejection rate,
- Evaluation of the mean embedding and detection time,
- Interface to input watermarking schemes and deploy watermark embedding and detection processes, thus weighing the outputs for certain attacks based on the target application
- Option for the user to choose any combination of attacks based on the target application.

Types of attacks that could be included in such a benchmarking tool are: Copy attacks, Geometric attacks, Simple Waveform attacks (i.e. MPEG compression), Removal attacks (such as de-noising, frame removal, frame linear transformations).

6. Conclusions and Future Directions

Video watermarking is a recent area of exploration of digital watermarking, that may deploy artificial intelligence algorithms both for efficient and content-dependent watermarking aiming at watermark robustness but also for optimal watermark detection and identification. The increasing concern of multimedia owners for copyright protection motivates further research in this field. In this chapter, we have reviewed a number of existing video watermarking schemes that cover a wide range of applications, varying from frame-based watermarking to more sophisticated video specific watermarking in a three-dimensional space. Furthermore, we compared a number of existing video watermarking techniques performance against attacks, and found that there is indeed room for improvement since all attacks cannot be completely dealt with. We further need to define detailed constraints based on the targeted application at hand. Finally, benchmarking design requirements have been presented for video watermarking in order to objectively evaluate and rate a wide range of video watermarking methodologies.

Only a few video watermarking algorithms have been proposed that meet the real-time or the three-dimensional constraint. There are thus technical challenges that are still unexplored. Future research on these areas will play a decisive role in digital video watermarking, and subsequently content protection paradigms.

References

[1] B. Kahng et al., Watermarking Techniques for Intellectual Property Protection, *35th IEEE Design Automation Conf*erence., San Francisco, USA, 1998, 776 - 781

[2] Yao Zhao, Reginald L. Lagendijk, Video Watermarking Scheme Resistant to Geometric Attacks, *IEEE International Conference on Image Processing (ICIP 2002*), 2002, Vol..2, 145 - 148

[3] A. Kejariwal, Watermarking, *Potentials IEEE*, 22(4), 2003, 37 - 40
[4] S.N. Merchant et al., Watermarking of Video Data Using Integer-to-Integer Discrete Wavelet Transform, *Conference on Convergent Technologies for Asia-Pacific Region (TENCON 2003*), 2003, 939 - 943
[5] S. Voloshynovskiy et al., Attacks on Digital Watermarks: Classification, Estimation-Based Attacks, and Benchmarks, *IEEE Communications Magazine*, August, 2001, 119 - 126
[6] M. Kutter, Digital Image Watermarking: Hiding Information in Images, *PhD thesis*, EFPL, Lausanne, Switzerland, 1999
[7] C. I. Podilchuk, E. J. Delp, Digital Watermarking: Algorithms and Applications, *IEEE Signal Processing Magazine*, July, 2001, 33 - 46
[8] B. G. Mobasseri, D. Cinalli, Lossless Watermarking of Compressed Media using Reversibly Decodable Packets, *Elsevier Signal Processing*, September, 2005
[9] K. Su, D. Kundur, D. Hatzinakos, Statistical Invisibility for Collusion-resistant Digital Video Watermarking, *IEEE Trans. on Multimedia*, 7(1), 2005, 43 - 51
[10] E.T. Lin, E.J. Delp, Temporal Synchronization in Video Watermarking, *IEEE Trans. on Signal Processing*, 52(10), Part 2, 2004, 3007 - 3022
[11] F. Deguillaume, G. Csurka, and T. Pun, Countermeasures for Unintentional and Intentional Video Watermarking Attacks, Proc. SPIE, Vol. 3971, Jan. 2000, 346 - 357
[12] M. D. Swanson, B. Zhu, and A. T. Tewfik, Multiresolution Scene-based Video Watermarking using Perceptual Models, *IEEE Journal of Selected Areas in Communications*, 16(4), 1998, 540 - 550
[13] C.-S. Lu, J.-R. Chen and K.-C. Fan, Real-time Frame-dependent Video Watermarking in VLC Domain, *Elsevier Signal Processing: Image Communication*, 20(7), 2005, 624 - 642
[14] S. H. Kwok, C. C. Yang, K. Y. Tam and Jason S. W. Wong, SDMI-based Rights Management Systems, *Elsevier Decision Support Systems*, 38(1), 2004, 33 - 46
[15] X. Kong, Y. Liu, H. Liu and D. Yang, Object Watermarks for Digital Images and Video, *Elsevier Image and Vision Computing*, 22(8), 2004, 583 - 595
[16] G. Doërr and J.-L. Dugelay, A Guide Tour of Video Watermarking, *Elsevier Signal Processing: Image Communication*, 18(4), 2003, 263 - 282
[17] P. Judge and M. Ammar, WHIM: Watermarking Multicast Video with a Hierarchy of Intermediaries, *Elsevier Computer Networks*, 39(6), 2002, 699 - 712
[18] M.P. Queluz, Authentication of Digital Images and Video: Generic Models and a New Contribution, *Elsevier Signal Processing: Image Communication*, 16(5), 2001, 461 - 475
[19] F. Hartung and B. Girod, Watermarking of Uncompressed and Compressed Video, *Elsevier Signal Processing*, 66(3), 1998, 283 - 301
[20] B.G. Mobasseri, M.P. Marcinak, Watermarking of MPEG-2 Video in Compressed Domain using VLC Mapping, *7th ACM Workshop on Multimedia and Security (MM&Sec '05)*, Aug. 2005
[21] X. Zhang, S. Wang, A New Watermarking Scheme against Inserter-based Attacks Suitable for Digital Media with Moderate Size, *3rd Int. ACM Conference on Information Security (InfoSecu '04)*, Nov. 2004
[22] M. Kutter and F. A. P. Petitcolas, A Fair Benchmark for Image Watermarking Systems, *Security and Watermarking of Multimedia Contents*, http://citeseer.ist.psu.edu/kutter99fair.html, 1 - 14
[23] B. Girod, Bi-directionally Decodable Streams of Prefix Code-words, IEEE Communication Letters, 3(8), 1999, 245 - 247
[24] F. Deguillaume et al., Robust 3D DFT Video Watermarking,, *SPIE Conference on Security and Watermarking of Multimedia Contents I*, Vol. 3657, San Jose, USA, Jan. 25–27, 1999, 113 - 124
[25] J. A. Bloom et al., Copy Protection for DVD Video, *Proceedings of IEEE*, Vol. 87, Jul. 1999, 1267 - 1276
[26] S. Craver et al., Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks, and Implications, IEEE Journal of Selected Areas in Communications, Vol. 16, May 1998, 573 - 586
[27] M. Kutter, S. Voloshynovskiy, and A. Herrigel, The Watermark Copy Attack, *SPIE Security and Watermarking of Multimedia Contents II*, Vol. 3971, San Jose, CA, Jan. 24–26, 2000, 371 - 380
[28] G. Doërr and J.-L. Dugelay, Security Pitfalls of Frame-by-Frame Approaches to Video Watermarking, *IEEE Trans. on Signal Processing*, 52(10), 2004, 2955 - 2964
[29] S. Pereira et al., Second Generation Benchmarking and Application Oriented Evaluation, *Information Hiding Workshop III*, Pittsburgh, USA, April 2001
[30] S. Voloshynovskiye et al., Attack Modeling: Towards a Second Generation Benchmark, *Signal Processing, Special Issue: Information Theoretic Issues in Digital Watermarking*, May 2001
[31] F. Petitcolas, R. Anderson, M. Kuhn, Attacks on Copyright Marking Systems, *2nd Int. Workshop on Information Hiding (IH'98*), Portland, U.S.A., April 15-17, 1998

[32] V. Solachidis, A. Tefas, N. Nikolaidis, S. Tsekeridou, A. Nikolaidis, I.Pitas, A Benchmarking Protocol for Watermarking Methods, *IEEE Int. Conference on Image Processing*, Thessaloniki, Greece, 7-10 Oct. 2001, 1023 - 1026

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Portrait Identification in Digitized Paintings on the Basis of a Face Detection System

Christos-Nikolaos ANAGNOSTOPOULOS [a], Ioannis ANAGNOSTOPOULOS [b], I. MAGLOGIANNIS [b] and D. VERGADOS [b]
[a] *Cultural Technology & Communication Dpt., University of the Aegean*
[b] *Information & Communication Systems Engineering Dpt., University of the Aegean*

Abstract. In this paper, the problem of automatic identification of portraits in paintings collections is addressed. A face detection approach in digital images is now implemented n digitized paintings, which is based on fuzzy logic rules especially set for detecting possible skin areas in color images. The candidate regions are then forwarded in a Probabilistic Neural Network (PNN), which is properly trained for face identification. The test sample for assessing the proposed method consists of 200 digitized paintings downloaded from the website of the State Hermitage Museum. The overall performance of the system reached 88.8%.

Keywords. Portrait identification, digital collections, face detection

Introduction

A lot of image processing techniques for face recognition in digital images have been presented and assessed in the literature [1]. However, the applications were mostly restricted in image and video processing for real world images and it is quite tempting to investigate the possibility of implementing similar techniques in a retrieval interface for a cultural application.

Museums and Cultural Institutions are becoming increasingly aware of how vital the emerging technologies are for reaching and engaging today's new audiences. However, organizations must be capable of offering rich content to fully benefit from new media technologies. This obviously includes high-quality digital images, which have already proven to be tremendously useful in all aspects of museums and other cultural institutions activities.

The method described in this paper belongs to Content-Based Image Retrieval (CBIR) methods. This is the process of retrieving images from a collection on the basis of image features and appearance (such as colour, texture and shape) that have been automatically extracted from the images themselves. The original method has been developed for the identification of faces in real world digital images [2] and intelligent image retrieval from the World Wide Web [3,4]. In the present paper, this method is implemented in digital paintings for automatic identification of portraits. In [5] automatic portrait identification in databases of art images is also addressed, but the method includes only image processing routines based on color and shape information. In contrast, the proposed method implements artificial intelligence techniques such as fuzzy logic rules for extracting color information and a neural network as a classifier.

The main scope of this research is to classify digital paintings into portraits and non – portraits using the assumptions that a portrait is a realistic representation of the sitter (the person in the portrait), showing the subject in mainly frontal view with plain backgrounds or ornate ones with curtains, architectural fragments, landscapes, etc. Image showing the subject standing or sitting and the face is in the focus (i.e. it is a foreground object) of the image.

1. Face Detection Method for Real World Scenes

1.1. Fuzzy Logic (FL) System

The color of human skin is distinctive from the color of many other objects and therefore the statistical measurements of this attribute are of great importance for face detection [6,7]. It is expected that the face color tones are distributed over a discriminate space in the color planes. So, the first step of the face detection system is the location of potential skin areas in the image, using color information of specific color models. Many approaches in the literature used similar detect procedures either based on the RGB, chrominance (CbCr) [8] or Hue and Saturation (HSV) space [9]. In the face detection method described in [2–4], a combination of the RGB model and YCbCr model was used for human skin discrimination. In addition, YCbCr was identified as the most representative color model for modelling human skin [10–12,14].

The basic concept in FL, which plays a central role in most of its applications, is that of fuzzy "if-then" rules or, simply, the fuzzy rules. In this work, the skin-masking algorithm presented in [13] is partially adapted along with RGB cluster groups that represent skin color extracted from experimental tests in a large database of human face images and an additional rule using the YCbCr model [4]. The above measurements define the fuzzy logic rules, which are used to formulate the conditional statements that comprise the fuzzy logic-based skin color detector. Applying fuzzy logic rules the proposed system decides whether a pixel in the inspected image represents or not a potential skin region. However, a skin region does not represent always a face, and therefore the candidate area should be normalized and checked whether it corresponds to a face or not. The fuzzy logic rules applied for skin discrimination are the following:

If $R/G > 1.3$ and $R/B > 1.4$ then possible skin

If $R/G < 1.3$ or $R/B < 1.4$ then no skin

If $R/G > 1.3$ and $G/B > 1.5$ then possible skin

If $77 < Cb < 127$ and $133 < Cr < 173$ then possible skin

where: (R=Red, G=Green, B=Blue in the RGB color model and Cb=Chromaticity blue, Cr=Chromaticity red in the YCbCr color model). The transformation from RGB to YCbCr color space is given by the following equations:

$$Y = 0.299R + 0.587G + 0.114B$$

$$Cb = -0.1687R - 0.3313G + 0.5B + 128$$

$$Cr = 0.5R - 0.4187G - 0.0813B + 128$$

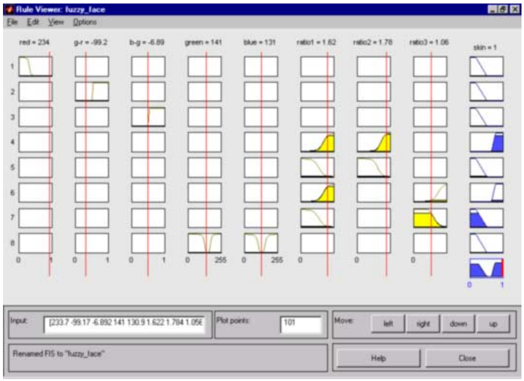


Figure 1. A screenshot of the fuzzy logic rules in Matlab rule viewer.

Figure 2. Human skin detection.

The first step is to take the inputs and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions. Once the inputs have been fuzzified, the fuzzy logical operations must be implemented. For this application the OR operator (max) was used and the weights in every rule were set equal to one. The aggregation method for the rules is the maximum value, while the defuzzification method is the middle of maximum (the average of the max value of the output set). The fuzzy rules were successfully applied to a Fuzzy Inference System (FIS), using the Fuzzy Logic Toolbox of Matlab 7.0 by MathWorks Inc. A screenshot of the FL in Matlab rule viewer is shown in Fig. 1. The inputs of the FIS program are the RGB and CrCb values of the input image. In a Pentium IV at 3.2 GHz, the required time for skin area detection varied from 1 to 2 seconds according the image size. An example of input and output images is presented in Fig. 2 respectively.

1.2. The Artificial Neural Network (ANN) for Image Classification

After the collection of images with possible skin areas, the next step involves the correct identification of images with human faces. This requires image processing techniques in order to properly feed the image classifier. The image-processing operations consist of four distinct parts.

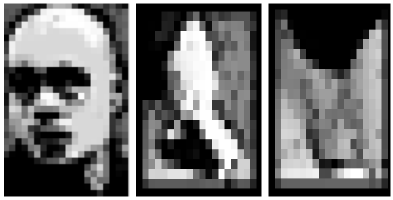


Figure 3. Results from image processing part 1 (face, left hand, right hand).

Firstly, potential skin areas are clustered to form the Region of Interest (RoI), roughly describing its shape, on the basis of the FL output. Every image is transformed in gray scale and resized to the specific size of 100×100 pixels. Then two morphological operations, which help to eliminate some of the noise in the tested image, follow. Specifically, simple erosion with a 10×10 matrix of ones is performed followed by dilation. Further on, the created image is parsed through a skeletonisation technique, removing simultaneously all the areas that are considered as 'holes'. As a result, the RoIs of all the possible skin areas are identified (e.g. Fig. 2).

After the RoI identification, the next step merges objects that belong to the same area, performing a simple dilation once again, with a structural element, which is now a 5×5 matrix of ones. With this technique, segmented pixels in the same neighbourhood, are merged in one region. All the pixels that are included in the defined RoIs, are then transformed to gray scale. In the following step, all the segmented images are resized to a 225×225 pixels. Finally, the latter images are divided into non-overlapping subimages of 15×15 pixels and the mean value for each of them is calculated, followed by histogram equalization, which expands the range of intensities in the window [15]. During this procedure, a lower resolution image is created, forming in parallel a descriptor vector that consists of 225 gray scale values from 0 to 255. Figure 3 presents an example of the input for the Artificial Neural Network (ANN). The proposed ANN is trained to identify the skin regions that represent faces. The training set of the ANN consists of a large group of images of the size 15×15, representing face regions or other skin areas. The idea of this approach was motivated by the observation that human faces present a high degree of resemblance when they are sampled in low-resolution as also proposed in [16]. This is due to the fact that all faces have dark areas in the eyes and the mouth. Therefore, it is easier for an ANN to recognize the presence of a face, judging from a low quality image. Additionally, the numbers of the computational units are significantly smaller for a low quality image. Artificial Neural Networks have been successfully applied for face detection in images as shown in [16–18] and [19].

The ANN is a two layer Probabilistic Neural Network (PNN) with biases and Radial Basis Neurons in the first layer and Competitive Neurons in the second one [20]. Training a neural network for the face detection task is quite challenging due to the difficulty in characterizing prototypical "non-face" images. Unlike in face recognition, where the classes to be discriminated are different faces, in face detection, the two classes to be discriminated are "face" and "non-face/other". Figure 4 depicts the topology of the proposed PNN as well as the transformation of a face image in the appropriate input vector form, which consists of 225 gray scale values.

A sample of 129 frontal view face images was used as training set for the class 'Face', as well as a large sample of 296 images corresponding to other correct or erroneously detected skin areas, such as hands, legs and other objects. Table 1 presents the

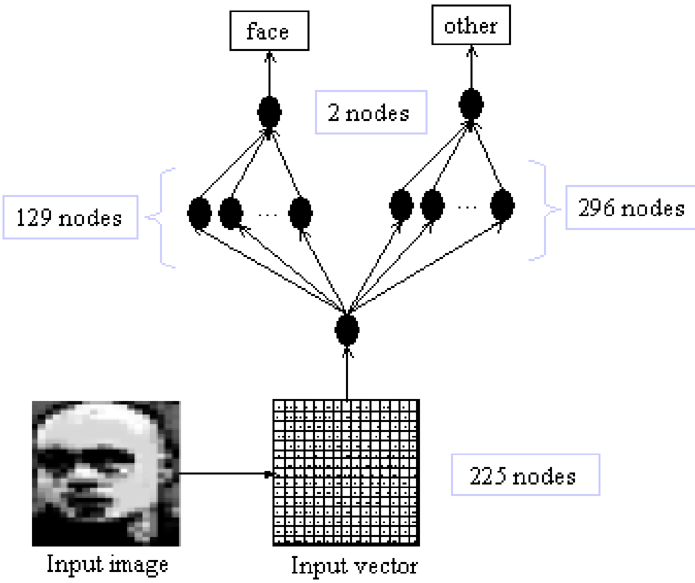


Figure 4. The architecture of the Probabilistic Neural Network.

Table 1. Training confusion matrix

	Face	**Other skin area – object**
Face	99.22% (128/129)	0.88% (1/129)
Other skin area – Object	1.01% (3/296)	98.99% (293/296)

confusion matrix percentages in terms of the learning ability during the training epoch. The training set consisted of 425 sub-images of size 15×15 in a vector form, as these were extracted from 103 color images according the proposed image processing steps. The neural network ‘learned’ to identify 128 from the 129 sub-images corresponding to human faces as well as 293 from the 296 sub-images corresponding to other skin areas and objects. The time needed for the completion of one training epoch in a Pentium IV at 3.2 GHz with 1024 MB RAM, was 10 seconds. The topology of the proposed neural network was 225-425-2. This means that the PNN had a 225-input vector (the 15×15 input image) and a 2-output vector corresponding to the decision of the system (whether it is a face or not). Finally, the system had 425 (129+296) nodes in the middle layer corresponding to the total training set.

1.3. Image Processing Performance

The performance of the system for face detection was tested using 317 color images of various formats, types and size containing human faces. More specifically, the sample of 317 color images contained 482 faces. The system implementing the fuzzy logic rules segmented totally 841 skin areas. However, 30 faces were not identified and therefore the performance of this system was 93.77% (452/482). Following the fuzzy logic system, the ANN received the 841 skin areas and decided that 397 of them repre-

Table 2. System's Performance in the testing set

	Testing set	
Total images	**Number of faces**	
318	482	
	Fuzzy Logic rules	
Segmented areas	841	452 faces + 389 possible skin areas
FL performance	452/482	93.77%
	Artificial Neural Network (ANN)	
Faces	397	
No faces	444	
ANN Performance	397/452	87.83%
Total Performance	397/482	82.36%

sent faces. Thus, the performance of the ANN was 87.83% (397/452). Finally, the overall system performance reached 82.36%, since 397 from a total of 482 faces were identified. All the results are shown analytically in Table 2.

2. Face Detection System for Portrait Identification

The approach, which was described above, is now used implemented for face detection in digitized paintings. The method is based on the fuzzy logic rules especially set for detecting possible skin areas in the paintings on the basis of color information. The candidate regions are then forwarded in the Probabilistic Neural Network (PNN) that is properly trained for the identification of faces from skin areas. Images containing face regions should be classified as portraits. However, it should be emphasized that there are some restrictions imposed due to the nature of the application. Firstly, the system handles only color and not gray scale images. Without the color information, the FL system cannot be activated. Moreover, the system can recognize portraits among realistic representative paintings and not modern artworks (symbolic, abstract or expressionistic portraits). Finally, this implementation is limited to the detection of human faces in frontal view.

2.1. Testing Set

The testing set of the portrait identification system consist of 200 digitized paintings of various types and size, containing individual portraits (1 face in the artwork), group portraits (more than one face in the painting), landscape and still life that were downloaded from the website of the State Hermitage Museum [21]. More specifically, the sample of 200 digital images included 120 non portrait images and 80 portraits (individual or group portraits) that totally contained 104 faces.

The system implementing the fuzzy logic rules segmented totally 228 skin areas. However, 5 faces were not selected and therefore the performance of this system is 95.2% (99/104). Following the fuzzy logic system, the ANN received the 228 skin areas and decided that 93 of them represent faces (93.9% success, i.e. 93/99 faces). Those 93 faces belong to 71 paintings that were successfully classified as portraits. The re-

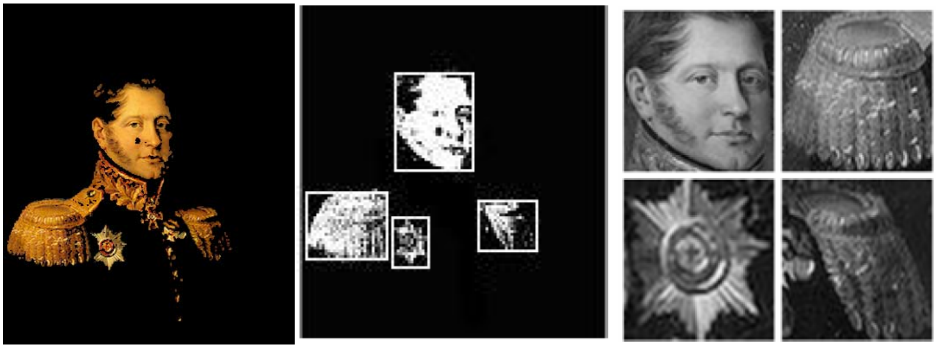


Figure 5. Left image: results of FL rules in the original painting (Dawe, George, Portrait of Alexander I. Gressor, Hermitage Collection). Middle image: Regions of Interest. Right image: the segmented potential face regions in grey scale. Each region will be downsized to 15×15 pixels and forwarded to the PNN.

Table 3. Portrait identification recognition rates

Number of paintings	200	
Number of portraits	80	
	FL rules	
Segmented areas	228	99 faces + 129 possible skin areas
	Neural Network	
Identified faces in paintings	93	Belonging in 71 paintings
	Total Performance	
Portraits correctly identified	71	71/80 (88.8%)

maining 9 portraits (containing 11 faces) were missed due to the FL and PNN misclassifications (5 and 6 errors respectively). Thus, the overall performance as shown in Table 3 is 88.8%, since 71 from a total of 80 portraits were identified.

Some screenshots are presented below highlighting the performance of the system in paintings. In Fig. 5, the segmented potential face regions in grey scale are featured. In a group portrait shown in Fig. 6, the system has identified correctly three of the four faces. The neural network failed to identify the boy in the left, probably due to the fact that it is not a frontal view of his face. However, the painting was classified as a portrait.

3. Conclusions

This paper presented in details the Fuzzy Logic rules joined with a Probabilistic Neural Network for face detection in natural scenes and its application for portrait identification in a collection of digitized paintings. The architecture of both systems applied directly to colour images fir identification of portraits in art images database is described. The method consists of a multistage algorithm, which segments the images based on colour, intensity and edge information. After the image segmentation, the resulting regions are forwarded to a neural network, which is trained to identify face and nonface regions. This classification is the basis for the classification of a painting as a portrait or non-portrait.

Figure 6. Left image: results of FL rules in the original group portrait (Bruyn, Bartholomaus I., Portrait of a Man and His Three Sons, late 1530s – early 1540s, Hermitage Collection). Middle image: Regions of Interest. Right image: PNN results. One face was not identified. However, the painting was classified as portrait.

The results are very encouraging for further development of the method, but a lot of limitations still exist. This implementation is limited to the detection of human faces in frontal view. A possible and interesting extension should include identification of sided-view faces as well.

Reviewing the literature, the only relevant work is described in [22] with comparable results to our approach. Specifically, the features were extracted from overall 188 images, divided into two disjoint subsets. The training set contained of 88 images (38 portraits, 50 non-portrait pictures), and the testing set contained 100 images (50 portraits, 50 non-portraits). The ground truth location of the faces in images was manually determined and the extracted regions were labelled with target values 1 (face) and 0 (non-face). The segmentation of the images resulted in 38 face regions and 322 non-face regions in the training set and 50 face regions and 280 non-face regions in the testing set. A three-layer back-propagation artificial neural network (ANN) was trained to classify the candidate regions as a face or non-face using as input the values of the extracted ellipse and bounding box features. The ANN consisted of the input layer, one hidden layer containing up to 18 nodes and the output layer with one node. The respective success rate reached 88% for face detecting, with a false positive rate of 6.4%.

The method proposed herein, can be used for retrospective documentation, for retrieving images through raw digitized material without proper documentation (images that were created at different periods of time without any registration number), for classification of digital images and for publishing them into the category of portraits on websites. Actually, the above categories enable users' search in a digital collection and they are very helpful for browsing a large collection of images on the web, too. Furthermore, the automatic identification of portraits can be combined with the creation of meaningful teaching and learning resources on the web. It could forward the design of a resource pack or an educative web game that introduce students to the vocabulary, history and major themes of portrait and to explore through some examples history, art, stylistic features of an epoch, or an artist, execution techniques, type of portraits, details for the sitter like his position, prestige, profession, etc. Portraiture gives many opportunities for developing rich content education resources and activities and we assume that the proposed application can contribute in exploring different meanings in a portrait image. A portrait does not have to be painted on canvas. It can be many things; it can be tiny (a miniature) or life-size, painted on wood, a sculpture, a drawing or a photograph and all these types in a digital collection can be treated as digital images. In a future application the sample can be expanded in order to include all these different types of portraits.

References

[1] Yang M.H., Kriegman D.J. and Ahuja N.: "Detecting Faces in Images: A Survey", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, no. 1, Jan. 2002, pp. 34–58.

[2] Anagnostopoulos, C., Anagnostopoulos, J., Vergados, D., Kayafas, E., Loumos, V. and Stassinopoulos, G.: "A Neural Network and Fuzzy Logic System for face detection on RGB images", 16th ISCA International Conference, 28–30 March 2001, Seattle, Washington, USA, ISBN: 1-880843-37-4, pp. 233–236.

[3] Anagnostopoulos, I., Anagnostopoulos, C., Kouzas, G., Vergados, D.D.: Precise photo retrieval on the web with a fuzzy logic\neural network-based metasearch engine, LectureNotes in Computer Science, LNCS, Methods and Applications of Artificial Intelligence, Volume 3025, April, 2004, pp. 43–53.

[4] Anagnostopoulos, I., Anagnostopoulos, C., Psoroulas, I., Loumos, V., Kayafas, E.: Information fusion meta-search interface for precise photo acquisition on the Web; Information Technology Interfaces, 2003. ITI 2003. Proceedings of the 25th International Conference on 16–19 June 2003, Page(s): 375–381.

[5] Sikudova, E., M. Gavrielides, Pitas, I.: Automatic Identification of Portraits in Art Images Databases, Proceedings of the DELOS workshop in audiovisual content, Crete, Greece, June 2–3, 2003.

[6] Belongie, S., Carson, C., Greenspan, H. Malik, J.: Color- and texture-based image segmentation using EM and its application to content-based image retrieval, Proceedings of the 6th IEEE International Conference in Computer Vision, January 1998, pp. 675–682.

[7] Grecu, H., Buzuloiu, V., Beuran, R., Podarou, E.: Face recognition using 3D distance maps and principal component analysis, H., Proceedings of Optim 2000, Brasov, May 2000, vol. 3, pp. 799–804.

[8] Garcia, C., Tziritas, G.: Face detection using quantized skin color regions merging and wavelet packet analysis, IEEE Trans. On Multimedia, vol. 1, no. 3, 1999, pp. 264–277.

[9] Tsekeridou, S., Pitas, I.: Facial features extraction in frontal views using biometric analogies, Proceedings of 9th European Signal Processing Conference, Rhodes, Greece, 8–11 September 1998,vol. 1, pp. 315–318.

[10] Bernd, M., Brünig, M.: Locating human faces in color images with complex background. In Proc. IEEE Int. Symposium on Intelligent Signal Processing and Communication Systems ISPACS'99, Phuket, Thailand, December 1999, pp. 533–536.

[11] Saber, A., Tekalp, M.: Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions, Pattern Recognition Letters, vol. 19, June 1998, pp. 669–680.

[12] Sobottka, K., Pitas, I.: A novel method for automatic face segmentation, facial feature extraction and tracking, Signal processing: Image communication, 12, pp. 263–281, 1998.

[13] Umbaugh, S: Computer Vision and Image Processing, Prentice Hall International, NJ, 1998.

[14] Chai, D., Ngan, K.N.: Face segmentation using skin color map in videophone applications, IEEE Trans. on Circuits and Systems for Video Technology, vol. 9, June 1999, pp. 551–564.

[15] Sung, K.K., Poggio, T.: Example-based learning for view-based human face detection, A.I. Memo 1521, CBCL Paper 112, MIT, December 1994.

[16] Dai, Y., Nakano, Y.: Recognition of facial images with low resolution using a Hopfield memory model, Pattern Recognition, vol. 31, no. 2, 1998, pp. 159–167.

[17] Rowley, H., Baluja, S., Kanade, T.: Neural Network based Face Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, 1998, pp. 23–28.

[18] Lin, S.-H., Kung, S.-Y., Lin, I.-J.: Face recognition/detection by probabilistic decision-based neural network, IEEE Trans. on Neural Networks, vol. 8, January 1997, pp. 114–131.

[19] Moghaddam, B., Pentland, A.P.: Probabilistic visual learning for object detection, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 17, July 1997, pp. 696–710.

[20] D.F. Sprecht, "Probabilistic Neural Networks", Neural Networks, vol. 3, pp.109–118, 1990.
[21] The State Hermitage Museum Digital Collection, online. http://www.hermitagemuseum.org/, last date of visit: 1 September 2006.
[22] Sikudova, E., Gavrielides M., Pitas, I.: Extracting semantic information from art images, Computer Vision and Graphics, K. Wojciechowski et al. (eds.), Springer Netherlands, 2006, pp. 394–399.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Where and Who? Person Tracking and Recognition System

Aristodemos PNEVMATIKAKIS [1]
Autonomic and Grid Computing, Athens Information Technology, Greece

Abstract. Many artificial intelligence applications rely on the ability to locate and recognize the people that are using the provided service. The involved tracking and recognition tasks are grossly different from those of the typical security application. The people are recorded from far away and act naturally. On the other hand, there is typically more than one sensor recording them and the decisions are based on some video stream, not on a single frame. This chapter addresses far-field unconstrained person tracking and video-to-video recognition, effectively answering the questions where are the humans, and who are they.

Keywords. Person tracking, face detection, video-to-video face recognition, fusion

Introduction

There are artificial intelligence applications aiming to providing their users with services in an unobtrusive way. Such applications include modern human-machine interfaces, computer-assisted living for groups that need assistance like the elderly or the very young, and computer-assisted working, like meeting support, to name a few.

A common requirement of such applications is the need for the system to keep answering two fundamental questions: Where are the human users of the service? Who are they? Answering the 'where?' and 'who?' questions requires a tracker to follow the position of humans around the monitored space, a face detector to extract faces from the tracked human bodies, and a face recognizer to provide the identity of the detected faces.

The needed system is quite different from typical deployments of face recognition systems for security and access control. It needs to operate with users that are unaware of its presence. This constraints the deployment of the cameras, which are positioned far away from the humans being recorded, typically on room corners near ceilings, tilted downwards, resulting to far-field recording conditions. The use of inexpensive, sub-megapixel cameras in typical living rooms or meeting rooms result to face sizes that are challengingly small. The unobtrusive operation also results to unconstrained video streams, where the humans being recorded go about their business, only occasionally facing towards the camera. In such unconstrained conditions, illumination, pose and expression vary widely.

What makes systems that answer the 'where?' and 'who?' questions possible is the use of temporal information. Such systems work with video streams, not with isolated

[1] Corresponding author: Aristodemos Pnevmatikakis, Autonomic and Grid Computing, Athens Information Technology, 0.8 km Markopoulou Ave., PO Box 68, 19002 Peania, Athens, Greece; E-mail: apne@ait.edu.gr.

frames. The use of temporal information helps the system in two important ways: Firstly, it allows the use of a human body tracker that narrows down the search region for human faces. Secondly, the face recognizer needs not derive a decision from a single image of the face, but rather from a sequence of faces collected over time, using some sort of fusion. This way, the effect of adverse pose, illumination and expression that can render a face undetectable or unrecognizable is alleviated. A face more suitable for detection and recognition is likely to appear occasionally. This is especially true when multiple cameras are used.

In the rest of this chapter the body tracker, face detector and face recognizer are detailed. A publicly available database, collected by the partners of the CHIL project [1] is then presented, followed by the results of the proposed system on it. Finally, the conclusions are drawn, together with some directions for extending the system.

1. Face Detection and Recognition System

The proposed system comprises three modules: A tracker follows the bodies of humans around the monitored space. A face detector detects faces in the upper part of the tracked bodies and validates them using color criteria. Finally, a face recognizer operates on the faces segmented from the video streams; the resulting identities are fused to provide a single identity over the video duration of interest.

1.1. Body Tracker

The goal of the body tracker is to provide the frame regions occupied by human bodies. It is based on a dynamic foreground segmentation algorithm [2,3] that utilizes adaptive background modeling with learning rates spatiotemporally controlled by the states of a Kalman filter. The block diagram of the tracker is shown in Figure 1. It comprises three modules in a feedback configuration: adaptive background modeling that is based on Stauffer's algorithm [4] provides the pixels that are considered foreground to the evidence formation module. The later combines the pixels into body evidence blobs, used for the measurement update state of the Kalman filter module. The states of the Kalman filter are used to obtain an indication of the mobility of each target, as a combination of translation motion and its size change. Also the position and size of the targets are contained in the states of the Kalman filter. This information is fed back to the adaptive background modeling module to adapt the learning rate in the vicinity of the targets: frame regions that at a specific time have a slow-moving target have smaller learning rates.

The proposed spatiotemporal adaptation of the learning rate of the adaptive background modeling module solves the problem of Stauffer's algorithm when foreground objects stop moving. Without it, targets that have stopped moving are learnt into the background. With the proposed feedback configuration this process is halted long enough for the intended application, i.e. tracking people in a meeting. For the details of the body tracing algorithm and its application both for in-doors and out-doors tracking, see [2,3].

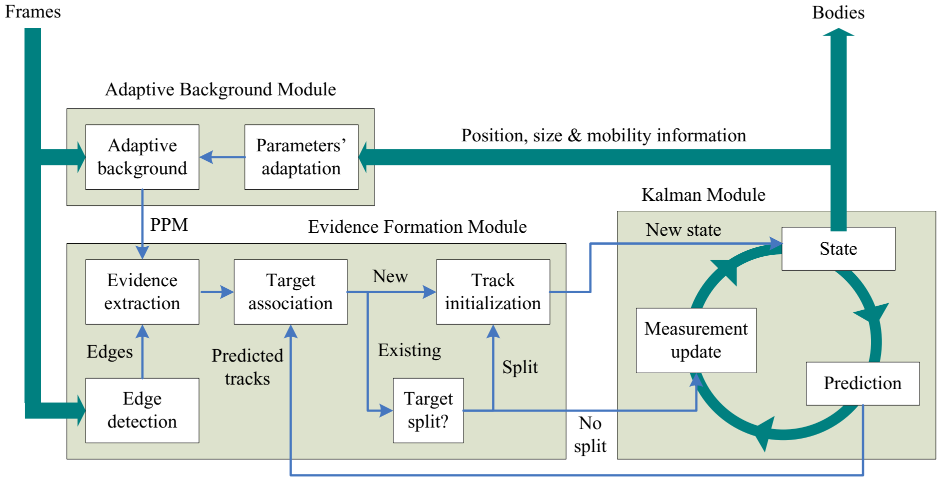


Figure 1. Block diagram of the complete feedback tracker architecture. Frames are input to the adaptive background and evidence formation modules, and targets are output from the Kalman module.

1.2. Face Detector

Many face detectors can be found in the literature [5-7]. They are split into those that attempt to find face futures, and from them find the faces themselves, and those that search directly for faces. The former are usually slower, and require adequate resolution for accurate feature detection. As the faces are tiny compared to the frame size, the natural choice for a detector is the boosted cascade of simple features [7]. Its implementation in OpenCV [8] is chosen, as this is publicly available. Although a trained cascade of simple classifiers is already provided with OpenCV, it is not suitable for our needs as the faces in our far-field recordings are too small. That detector has very high miss rate. A more suitable detector is thus trained. To do so we use 6,000 positive samples (images with marked faces), 20,000 negative samples (images with no human or animal face present), an aspect ratio of 3/4, minimum feature size 0, 99.9% hit rate, 50% false alarm, tilted features, non-symmetric faces and gentle AdaBoost learning [8].

The detector thus trained is applied on the grayscale portions of the frames that are designated as human bodies by the body tracker. Only the upper part of the bodies that has height equal to the body width is examined by the detector to speed up the process. Unfortunately the face detector suffers from false detections. These are constrained by the fact that the detector is applied on a limited portion of the frames, but nevertheless the detections need to be validated prior to propagating them to the face recognizer. The system tolerates more some misses than false detections. The face validation is done using the color and brightness of the detected region. The human skin and non-skin color models of Jones and Rehg [9] are used to build the likelihood ratio of human skin versus non-skin. At least a portion of the pixels of the detected region has to exhibit high skin color likelihood; hence the median of the skin likelihood needs to be high. Also, human faces exhibit a lot of brightness variation due to self-shadowing. Hence also the standard deviation of the brightness of the detected region needs to be high. Detected regions that pass the skin color and brightness thresholds are considered faces for recognition.

1.3. Face Recognizer

Training (gallery) faces are needed for recognition. These faces can be manually selected, but this can cause face registration difference from the automatically detected faces used by the system when it is in operation (probe faces). Hence the gallery faces are generated using the same automatic body tracking and face detection and validation procedure described above; only it is applied on the gallery videos. Many face recognition algorithms exist [6]; due to the limited face resolution and the real-time constraint of the intended application, the family of linear subspace projection algorithms is used by the system. Both Principal Components Analysis (PCA) [10] and Linear Discriminant Analysis (LDA) [11] are employed to build unsupervised and supervised projection matrices respectively. PCA aims at transforming the training vectors so that their projections in lower-dimensional spaces has maximum scatter. This guarantees optimality in terms of minimum squared error of the representation of the original vectors in any lower-dimensional space. The determination of the transformation matrix does not require any class information, hence it is unsupervised. Although the optimality in representation does not offer any guarantee for optimality in classification, the use of PCA has led to the successful Eigenface face recognition method [10]. The dimension D of the recognition subspace onto which the training vectors are projected is a parameter of the method, to be determined empirically. Suppressing some of the dimensions along which the scatter of the projected vectors is smallest not only increases the speed of the classification, but also seems to be suppressing variability that is irrelevant to the recognition, leading to increased performance. LDA on the other hand aims at maximizing the between-class scatter under the constraint of minimum within-class scatter of the training vectors, effectively minimizing the volume of each class in the recognition space, while maximizing the distance between the classes. The dimensions of the LDA subspace are K-1, where K is the number of classes. The determination o the LDA projection matrix requires class information, hence it is supervised. LDA suffers from ill-training [12], when the training vectors do not represent well the scatter of the various classes. Nevertheless, given sufficient training, its use in the Fisherfaces method [11] has led to very good results.

LDA is better for large faces with accurate eye labels [13], but PCA is more robust as face size and eye labeling accuracy drop. LDA is robust to illumination changes [11]. PCA can be made more robust to illumination changes if some of the eigenvectors corresponding to the largest eigenvalues are excluded from the projection matrix, but this reduces the robustness of PCA under eye misalignment errors. At far-field viewing conditions, resolution is low. This reduces recognition performance even when the position of the eyes is correctly determined. But at such resolutions this is very difficult, even for human annotators. For these reasons, both methods are used, and their results are fused. A note is due at this point for the application of LDA. Contrary to the Fisherfaces method [11], in this case the small sample size problem [14] does not apply. The number of pixels of the faces is smaller than the available gallery stills, no matter the gallery duration or the face cropping method employed. Hence no PCA step is used, without the need for a direct LDA algorithm [14].

According to the Eigenfaces [10] or Fisherfaces [11] methods, the gallery images are represented by their class means after projection to the recognition space. Recognition is based on the distance of a projected gallery face from those means. This is not effective in the case of unconstrained movement of the person, since then the intra-personal variations of the face manifold due to pose variations can be far more pronounced than the extra-personal variations [15]. In this case it is better to use a nearest neighbor classifier. The implication is that all the projected gallery faces have to be kept and compared against every probe projected face. Different distance metrics can be used for classification. In this chapter, the Euclidian distance is used.

A two-stage fusion scheme is employed, based on the sum rule [16]. The first stage performs fusion jointly across time and camera views, while the second stage fuses the results of the two classifiers. According to the sum rule, each of the decision ID_i of the probe faces in a testing segment casts a vote that carries a weight w_i. The weights w_i of every decision such as $ID_i = k$ are summed to yield the weights W_k of each class:

$$W_k = \sum_{i:ID_i=k} w_i \tag{1}$$

where $k = 1,\dots,K$ and K is the number of classes. Then the fused decision based on the N individual identities is:

$$ID^{(N)} = \arg\max_k \left(W_k\right) \tag{2}$$

The weight w_i in the sum rule for the i-th decision is the sixth power of the ratio of the second-minimum distance $d_i^{(2)}$ over the minimum distance $d_i^{(1)}$:

$$w_i = \left[\frac{d_i^{(2)}}{d_i^{(1)}}\right]^6 \tag{3}$$

This choice for weights reflects classification confidence [17]: If the smallest distances from two probes are approximately equal, then identification based on smallest distance is unreliable. In this case the weight is close to unity, weighting down the particular decision. If on the other hand the minimum distance is much smaller than the second-minimum, identification is reliable and the decision is heavily weighted.

The decisions $ID^{(\mathrm{PCA})}$ and $ID^{(\mathrm{LDA})}$ of the PCA and the LDA classifiers are again fused using the sum rule to yield the reported identity. For this fusion, the class weights W_k of equation (1) are used instead of the distances in equation (3). Setting:

$$\begin{aligned} k_1 &\equiv \left[\text{best matching class}\right] = ID^{(N)} \\ k_2 &\equiv \left[\text{second-best matching class}\right] \end{aligned} \tag{4}$$

the weights of the PCA and LDA decisions become:

$$w_i = \frac{W_{k_1}^{(i)}}{W_{k_2}^{(i)}},\ i \in \left\{\text{PCA, LDA}\right\} \tag{5}$$

Then the final decision fused across the PCA and LDA classifiers is:

$$ID = \begin{cases} ID^{(\mathrm{PCA})} & \text{if } w_{\mathrm{PCA}} \geq w_{\mathrm{LDA}} \\ ID^{(\mathrm{LDA})} & \text{if } w_{\mathrm{PCA}} < w_{\mathrm{LDA}} \end{cases} \quad (6)$$

2. Video database

The only truly far-field video recordings known to the author are those collected by the partners of the CHIL project [1] and already used in the Classification of Events, Activities and Relationships (CLEAR 2006) evaluation [18]. Unfortunately, many algorithms are only tested on custom built video databases, which are not publicly available, or for which not all the necessary data are reported. Unlike still-to-still face recognition, there have been no evaluations for video-to-video face recognition. The only exceptions are the CLEAR 2006 and the upcoming CLEAR 2007 evaluations [18], which include a video-to-video face recognition task. Table 1 summarizes the CLEAR 2006 database.

Note that the pose variations in the CLEAR 2006 database are extreme: some of the shorter videos do not contain any face with both eyes visible. This is alleviated by the use of 4 different camera views: one of the views is bound to capture some frames with faces having both eyes visible. The durations reported in Table 1 for this database are per camera view; there are actually four times as many frames to extract faces from. Some statistics of the gallery and probe faces extracted from the CLEAR 2006 videos using the proposed body tracker and face detector are given in Table 2. Finally, Figure 2 shows examples of the cropped gallery faces for one individual. The severity of pose variation and the problems of ill-framing the faces by the detector are apparent.

Table 1. Summary of the CLEAR 2006 video database used for video-to-video face recognition. The frame rate is 30 fps

No. of people	26
Camera views	4, at room corners
Gallery duration	15 and 30 sec
Probe duration	1, 5, 10 and 20 sec
No. of probe videos	613 (1 sec), 411 (5 sec), 289 (10 sec) and 178 (20 sec)
Scenario	Moving freely: meeting with presentation
Pose, expression	Any pose, natural talking expression
Illumination	Changes due to projector beam, overhead lights
Recording conditions	Far field, median eye distance 9 pixels

Table 2. Summary of the gallery and probe still sets generated from the CLEAR 2006 videos using the body tracker and the cascaded detector.

Gallery stills per person	Length (sec)	15		30	
	Min	118		251	
	Average	428		886	
	Max	890		1696	
Probe stills per video	Length (sec)	1	5	10	20
	Min	1	2	19	81
	Average	25	127	226	515
	Max	90	348	793	1406
	Empty videos	0	0	0	0

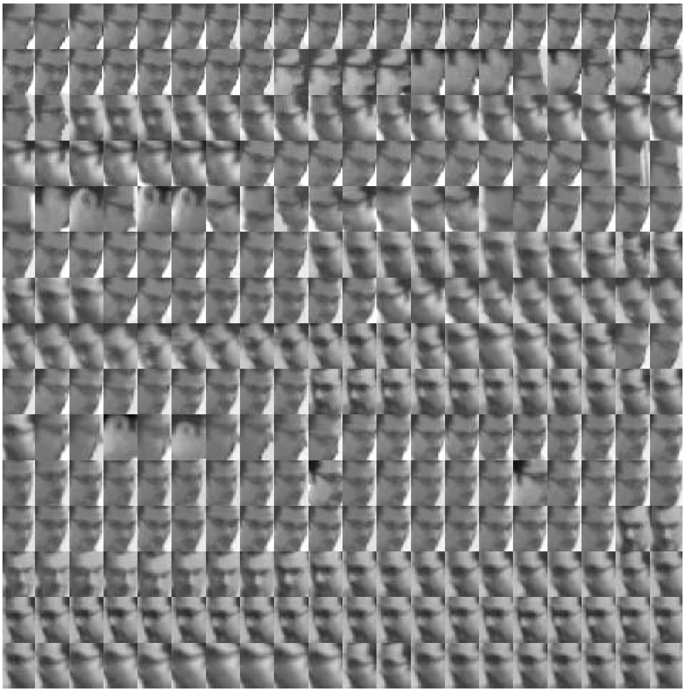


Figure 2. Temporally sub-sampled gallery faces automatically cropped from the 15 sec gallery videos, using all four cameras, for one person

3. Performance

The performance of the video-to-video face recognition system described in section 1 on the database presented in section 2 is analyzed next. The recognition rate in the probe videos is presented in Figure 3. For comparison, also the best performance achieved in the CLEAR 2006 evaluations is included.

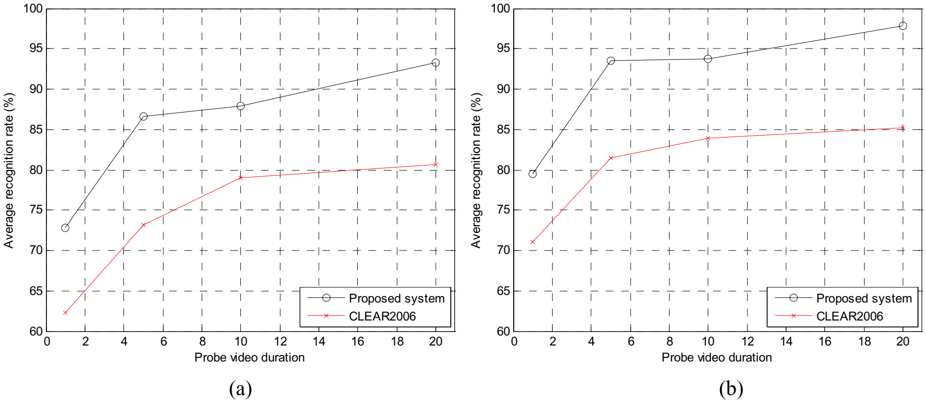
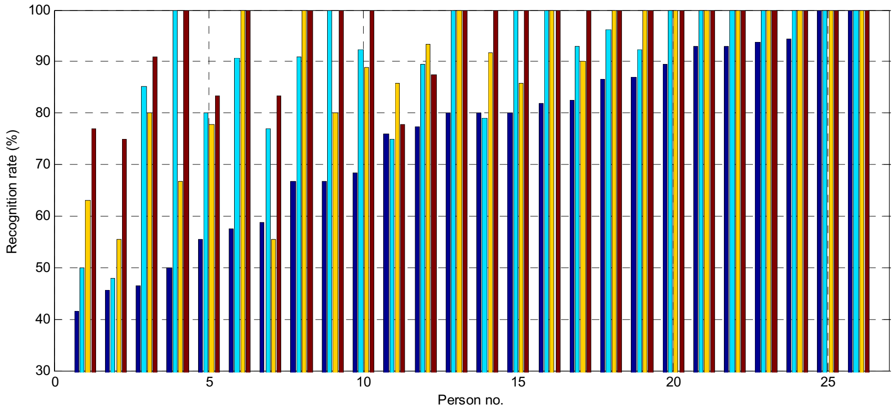
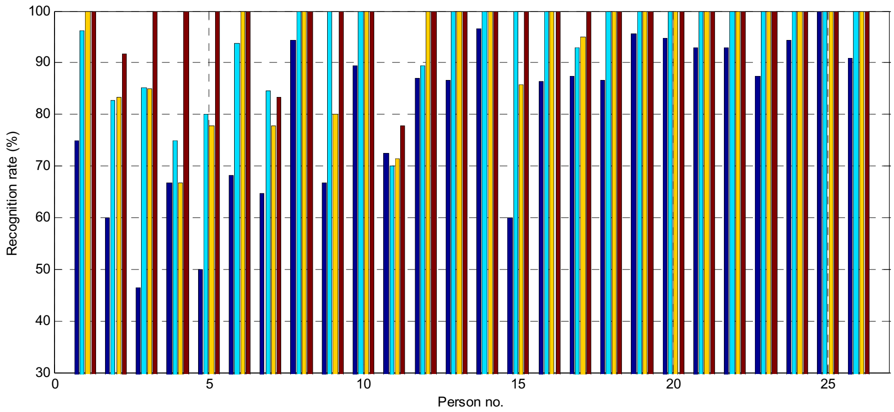


Figure 3. Average recognition rates for the various probe video durations, for (a) 15 sec gallery videos duration and (b) 30 sec gallery videos duration



(a)



(b)

Figure 4. Per person recognition rates for the different durations of the probe videos (grouped) and for the 15 sec (a) or 30 sec (b) long gallery videos. The people are sorted by ascending recognition rate for the 1 sec long probe and the 15 sec long gallery videos.

Finally, it is interesting to investigate if some people are harder to recognize than others. The bar graphs of Figure 4 depict the recognition rates for the 26 different people, under the two training and four testing conditions. Evidently, not always people that are very difficult to recognize in one of the eight training and testing conditions remain difficult in other conditions. This is because the actions of a person in the probe and gallery videos can be matched in different degrees as those videos change. For example, the most difficult person in the 15 sec gallery video and 1 sec probe videos is easier than people 2 and 7 in the 10 sec probe videos, and easier than people 2-6, 8, 10 and 14 in the 30 sec gallery video.

4. Conclusion and possible extensions

In this chapter we have presented a fully-automatic system for face detection and recognition, applied on far-field, unconstrained recordings. It is demonstrated that given long probe video durations, the performance of a system based on a body tracker, a frontal face detector, linear subspace projection and nearest neighbour classifier more or less solves the problem, with average recognition rates above 95%. In applications where long probe videos are impractical, performance is still low (recognition rates of 74% or 80% for 1 sec probe and 15 sec or 30 sec gallery video durations), especially given that the number of people are limited to the modest number of 26. To further enhance performance, there are some possible system improvements:

- Multiple face detectors can be trained, including poses other than frontal. Also, face detection can be coupled with a probabilistic tracker based on particle filtering [19] or a deterministic tracker based on colour histograms using CAMShift [20]. This will provide more stills, capturing more pose variations.
- Other distance metrics (weighted Euclidian, cosine) can be used for nearest neighbour classification.
- Modelling of face sequences, similar to the exemplar approach of [21], to automatically detect outliers that cannot be justified as smooth pose transitions, but rather are face detector errors. The cleaner face sequences thus obtained can be used to model pose transitions, allowing more efficient utilization of temporal information than weighted voting [15, 22-25].

Acknowledgements

This work has been partly sponsored by the General Secretariat of Research and Development of Greece, under the project PRIAMOS (GSRT Measure 3.3, on Audio-Video Processing, Project Code: 26). The author wishes to acknowledge the invaluable help and support of the members of the Autonomic and Grid Computing Group of Athens Information Technology and in particular of Andreas Stergiou and Nikos Katsarakis, in the development of some of the algorithms that comprise the presented system.

References

[1] A. Waibel, H. Steusloff and R. Stiefelhagen, CHIL: Computers in the Human Interaction Loop, *5th International Workshop on Image Analysis for Multimedia Interactive Services*, Lisboa, Portugal, Apr. 2004.

[2] A. Pnevmatikakis and L. Polymenakos, Kalman Tracking with Target Feedback on Adaptive Background Learning, in S. Renals, S. Bengio and J. Fiscus (eds.), *MLMI 2006, Lecture Notes in Computer Science*, Springer-Verlag, Berlin Heidelberg, 2006.

[3] A. Pnevmatikakis and L. Polymenakos, Robust Estimation of Background for Fixed Cameras, *15th International Conference on Computing (CIC2006)*, Mexico City, Mexico, 2006.

[4] C. Stauffer and W. E. L. Grimson, Learning patterns of activity using real-time tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22** 8 (2000), 747–757.

[5] R.-L. Hsu, M. Abdel-Mottaleb and A. K. Jain, Face Detection in Color Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24** 5 (2002), 696-706.

[6] S. Z. Li and J. Lu, Face Detection, Alignment and Recognition, in G. Medioni and S. Kang (eds.), *Emerging Topics in Computer Vision*, 2004.

[7] P. Viola, and M. Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, Dec. 2001, p. 511

[8] G. Bradski, A. Kaehler and V. Pisarevsky, Learning-Based Computer Vision with Intel's Open Source Computer Vision Library, *Intel Technology Journal* **9** (2005).

[9] M. Jones and J. Rehg, Statistical color models with application to skin detection, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1999, 274–280.

[10] M. Turk and A. Pentland, Eigenfaces for Recognition, *J. Cognitive Neuroscience*, **3** (1991), 71-86.
[11] P. Belhumeur, J. Hespanha and D. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19** 7 (1997), 711-720.
[12] A. Martínez and A. Kak, PCA versus LDA, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23** 2 (2001), 228-233.
[13] E. Rentzeperis, A. Stergiou, A. Pnevmatikakis, and L. Polymenakos, Impact of Face Registration Errors on Recognition, *Artificial Intelligence Applications and Innovations*, Peania, Greece, June 2006.
[14] H. Yu and J. Yang, A direct LDA algorithm for high-dimensional data with application to face recognition, *Pattern Recognition* **34** (2001), 2067–2070.
[15] Y. Li, S. Gong and H. Liddell, Video-Based Online Face Recognition Using Identity Surfaces. *Proceedings of IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Vancouver, Canada, July 2001, 40-46.
[16] J. Kittler, M. Hatef, R.P.W. Duin and J. Matas, On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20** 3 (1998), 226–239.
[17] A. Stergiou, A. Pnevmatikakis and L. Polymenakos, A Decision Fusion System across Time and Classifiers for Audio-visual Person Identification, in R. Stiefelhagen and J. Garofolo, (eds.), *CLEAR 2006, Lecture Notes in Computer Science 4122*, Springer-Verlag, Berlin, Heidelberg, 2007, 218-229.
[18] R. Stiefelhagen, K. Bernardin, R. Bowers, J. Garofolo, D. Mostefa and P. Soundararajan, The CLEAR 2006 Evaluation, in R. Stiefelhagen and J. Garofolo, (eds.), *CLEAR 2006, Lecture Notes in Computer Science 4122*, Springer-Verlag, Berlin, Heidelberg, 2007.
[19] S. Zhou, R. Chellappa and B. Moghaddam, Visual tracking and recognition using appearance-adaptive models in particle filters, *IEEE Transactions on Image Processing* **13** 11 (2004), 1491-1506.
[20] G. Bradski, Computer Vision Face Tracking for Use in a Perceptual User Interface, *Intel Technology Journal* **2** (1998).
[21] S. Zhou, V. Krueger and R. Chellappa, Probabilistic recognition of human faces from video, *Computer Vision and Image Understanding* **91** 7 (2003), 214-245.
[22] J. Weng, C.H. Evans and W.-S. Hwang, An Incremental Learning Method for Face Recognition under Continuous Video Stream, *Proceedings of IEEE Conference on Automatic Face and Gesture Recognition*, Grenoble, France, March 2000, 251-256.
[23] K.-C. Lee, J. Ho, M.-H. Yang and D. Kriegman, Video-based face recognition using probabilistic appearance manifolds, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, USA, June 2003, 313-320.
[24] X. Liu and T. Chen, Video-based face recognition using adaptive hidden markov models, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Madison, Wisconsin, USA, June 2003, 340-345.
[25] G. Aggarwal, A.K. Roy-Chowdhury and R. Chellappa, A System Identification Approach for Video-based Face Recognition, *Proceedings of International Conference on Pattern Recognition*, Cambridge, UK, Aug. 2004.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Context Awareness Triggered by Multiple Perceptual Analyzers*

Josep R. CASAS and Joachim NEUMANN
UPC – Technical University of Catalonia

Abstract. A multitude of technologies from computer vision, acoustic signal analysis and natural language processing are used to implement multi-modal perceptual components. The output of this analysis is used to gain context awareness – a necessity when designing a computer-based service that interacts reactively and proactively with humans. This article describes the integration process and our experience in implementing one such information service, the "Memory Jog", in a particular scenario where the computer system supports a group of journalists in their daily work.

Keywords. Multi-modal analysis, Multi-sensor network, Perceptual component, Memory Jog service

Introduction

Computers are tools we use for multiple purposes as providers of information, communication and entertainment. They are great for information management, useful tools for educational tasks and valid helpers for creative design processes. As far as the functional interaction is designed around a well-defined situation, with the computer as the centre of the task and the human as the controller, computers perform tedious tasks for us and have proven to operate efficiently providing help in working scenarios.

One long-term goal in Human Computer Interfaces (HCI) has been to migrate the "natural" means that humans employ to communicate with each other into HCI [1]. Advanced HCI set the user as the centre of the interaction instead of the computer. Instead of the human controller being aware of the task performed by the computer [2], it is the computer which should "be aware" of what the humans are doing around it. This demands equipping computer systems with audio-visual sensors conveying signals from the scene into analysis modules. These analysis modules, also called "perceptual components", are computer programs that analyse the signals from the sensors in a smart environment.

The work presented in this article was developed within the CHIL project [3] and has been inspired by the underlying vision of "Computers in the Human Interaction Loop". CHIL aims to change the way we use computers by yet taking modern HCI one step further. Instead of focussing on the direct Human-Computer interaction, which might keep the Human in the Computer Interaction Loop, a CHIL-like computer service stays in the background and helps the users by understanding their needs and providing support in a natural human-to-human interaction scenario [4]. In the vision of

* This work has been partially supported by the European Union, IP 506909 (CHIL).

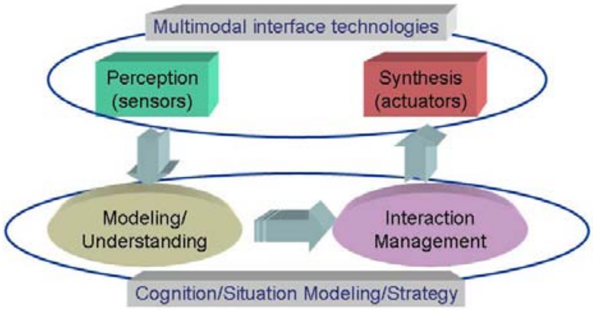


Figure 1. The chain Perception-Understanding-Interaction-Synthesis organized in the two levels of interfaces (sensors/actuators in the upper row) and the higher level of cognition (understanding/interaction management in the lower row).

CHIL, the humans don't need make any unnecessary effort to "understand" the computer and its processes, because it is the computer which is forced to get into the loop of humans in order to understand their interaction.

The first pre-requisite for such an ambitious aim is to gain context awareness, i.e., detection of people, objects, events and situations in the interaction scene. The information needed to build the relevant context awareness stems from the analysis of the signals acquired in real-time from a collection of sensors. Smart scenarios for HCI base their interaction with the users on "Multimodal interface technologies" for the detailed analysis of the environment, which require the design of flexible and reconfigurable sensor networks feeding data to the perceptual analysis components, as well as actuators providing natural signals to address human users.

However, it is not sufficient to simply gather the analysis result as a growing list of detections coming from the perceptual components. The overall process of Human-Computer interaction should be managed at a higher cognition level, where ontologies of objects, situations and events must be defined. This strategy provides the necessary knowledge complementing the perceptual analysis for adequate understanding of the human-to-human interaction, in order to take the most appropriate action at any time.

In the following sections we first explain how multimodal interface technologies and a simple situation/event model complement each other to build a CHIL-like service. In Section 2, we describe in detail the functionality of the Memory Jog service we have implemented. Section 3 reviews the perceptual technologies involved in the service and Section 4 presents the overall software architecture allowing the integration in the service prototype.

1. Multimodal Interface Technologies and Situation Models for CHIL-Like Services

Multimodal interface technologies comprise both the perceptual analysis and the display/synthesis front-end of a CHIL environment, as shown in Fig. 1. The perceptual analysis front-end consists of a collection of sensors and perceptual components detecting and classifying low-level features which can be later interpreted at a higher semantic level. At the synthesis side, actuators such as an Embodied Conversational Agent or

ECA [5] might address the user providing (and requesting) information in a natural, human-like way.

The smart room at UPC is equipped with multiple cameras and microphones [6]. Continuous room video monitoring is achieved by several calibrated cameras connected to dedicated computers, whose fields of view aim to cover completely the scene of interest, usually with a certain amount of overlap allowing for triangulation and 3D data capture for visual tracking, face localization, object detection, person identification, gesture classification and overall scene analysis. A multi-microphone system for aural room analysis deploys a flexible microphone network comprising microphone arrays, microphone clusters, table top microphones and close-talking microphones, targeting the detection of multiple acoustic events, voice activity detection, ASR and speaker location and tracking. Also for acoustic sensors a calibration step is defined, according to the purpose of having a jointly consistent description of the audio-video sensor geometry. Timestamps are added to all the acquired data for temporal synchronization.

Perceptual components are computing modules analyzing the signals provided by the network of sensors in order to detect and classify objects of interest, persons and events adding information to context awareness. In the Service implemented at the UPC smart room, context awareness consists of knowledge about the number of persons in the room, their identification, position in the room and their orientation. Objects in the room and acoustic events also add to the context awareness.

Context awareness stems from the set of detections coming from the perceptual components, which needs to be properly organized to get a correct understanding of the situation. The model of the possible situations and a well-defined strategy for the interaction should be defined to react properly (i.e. providing the needed information to the user in a timely manner) in the different situations identified by the system. A simple state-model was designed at UPC for the service at hand and the detected events were processed according to the current state of the state-model representing the knowledge introduced in the computer about the events and situations of interest. Some combinations of events were allowed to trigger a change of the state. Based on the current state, some of the detected events trigger a reaction of the Memory Jog service.

2. The Memory Jog as a CHIL-Like Service

The Memory Jog Service focuses at providing information. In the specific implementation of the Memory Jog at UPC, information is provided to a group of newspaper journalists gathered together in the CHIL smart room. They have to decide within ten minutes on the front page of tomorrow's edition of their newspaper. In addition to the information provided to the journalists in the smart room, a field journalist and a latecomer propose an additional news story. In this case, the Memory Jog service makes news available that have been created elsewhere (information-shift). The Memory Jog service is also capable of providing background information about the news (information-pull) and in some occasions may decide to jump in human-to-human communication to provide a pro-active service (information-push). The design paradigm behind these three ways to provide information was to enhance human-to-human communication, i.e., the journalists are helped to freely interact with each other instead of being forced to focus on how to interact with the Memory Jog Service. The two most outstanding means of the Memory Jog to interact with the journalists are:

- A Talking Head that not only informs the journalists about available resources, points out events such as the arrival of a latecomer or news being contributed by remote colleagues, but also facilitates information requests from the journalists in a human-like interface based on automatic speech recognition technologies.
- A remote field journalist is enabled to easily communicate with the journalists in the smart room. A Skype-based bi-directional audio communication is supported by real-time video managed by an automatic cameraman. The video is further enhanced by text annotations that reflect the context awareness of the Memory Jog.

The task of the journalists is to decide on two most important pieces of news to appear on the front page of tomorrow's newspaper. The news from which the journalists have to select from are imported by an up-to-date RSS feed. The Fig. 2 shows an example of the resulting front page.

3. Perceptual Components and Technologies

This chapter portrays hardware setup of the UPC smart room and requirements set by the integration of technologies in the Memory Jog Service. Thereafter, the perceptual components that perform the perceptual analysis (first step in Fig. 1) are introduced. Subsequent sections describe additional technologies that enable the Memory Jog to actuate and thus to provide its service.

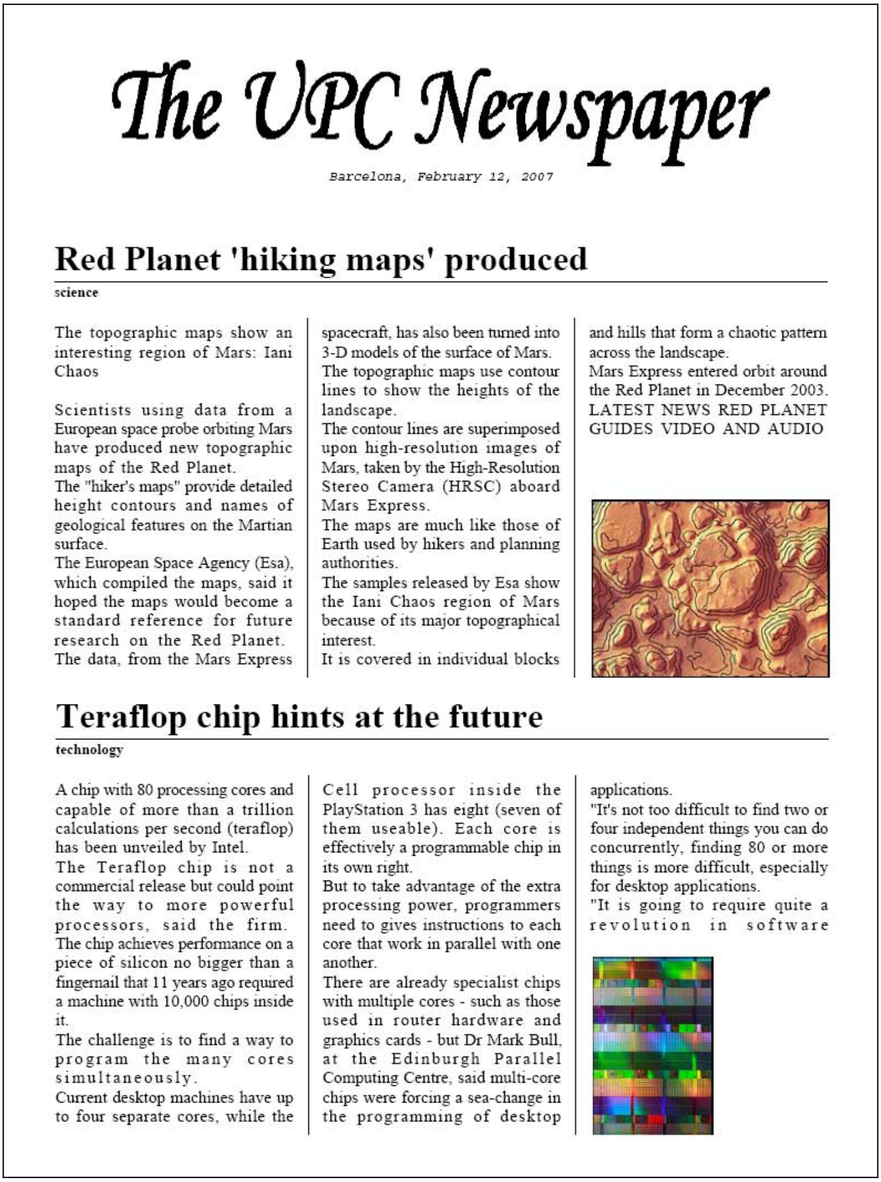
3.1. Hardware Setup of the UPC Smart Room

The unobtrusiveness desired for a natural interaction between humans and computers sets limitations on the positioning of the sensors in the room: the acoustic technologies applied in the UPC smart room limit themselves to far-field wall-mounted microphones that allow the participants to freely move around in the room without being concerned about how and where their voices and other sounds are picked up. However, the signals that these microphones deliver show an unfavourable signal-to-noise ratio and contain a large amount of reverberation due the scarce furniture and the acoustically hard floor and walls.

The consumer-type video sensors were as well chosen and mounted to yield an unobtrusive observation of the whole room. For example, the angle of the corner-cameras is wide enough to cover each point in the room by at least two cameras. Consequently, the quality and details obtained from these lenses is limited. Even the zoom camera that points at the entrance of the room and a webcam on the console shows close-ups of faces which are not more than 60×80 pixels in size.

3.2. Integrating Technologies to a Service

The two multi-modal perceptual components described in the following receive the raw audio and video streams as input data. The analysis of these data is often inspired by our knowledge about human perception in vision and hearing. Most of the technologies on which both perceptual components are based have their typical applications with well established evaluation methods. However, the criteria that determine their useful-



The UPC Newspaper

Barcelona, February 12, 2007

Red Planet 'hiking maps' produced

science

The topographic maps show an interesting region of Mars: Iani Chaos

Scientists using data from a European space probe orbiting Mars have produced new topographic maps of the Red Planet.
The "hiker's maps" provide detailed height contours and names of geological features on the Martian surface.
The European Space Agency (Esa), which compiled the maps, said it hoped the maps would become a standard reference for future research on the Red Planet.
The data, from the Mars Express spacecraft, has also been turned into 3-D models of the surface of Mars.
The topographic maps use contour lines to show the heights of the landscape.
The contour lines are superimposed upon high-resolution images of Mars, taken by the High-Resolution Stereo Camera (HRSC) aboard Mars Express.
The maps are much like those of Earth used by hikers and planning authorities.
The samples released by Esa show the Iani Chaos region of Mars because of its major topographical interest.
It is covered in individual blocks and hills that form a chaotic pattern across the landscape.
Mars Express entered orbit around the Red Planet in December 2003.
LATEST NEWS RED PLANET GUIDES VIDEO AND AUDIO

Teraflop chip hints at the future

technology

A chip with 80 processing cores and capable of more than a trillion calculations per second (teraflop) has been unveiled by Intel.
The Teraflop chip is not a commercial release but could point the way to more powerful processors, said the firm.
The chip achieves performance on a piece of silicon no bigger than a fingernail that 11 years ago required a machine with 10,000 chips inside it.
The challenge is to find a way to program the many cores simultaneously.
Current desktop machines have up to four separate cores, while the Cell processor inside the PlayStation 3 has eight (seven of them useable). Each core is effectively a programmable chip in its own right.
But to take advantage of the extra processing power, programmers need to gives instructions to each core that work in parallel with one another.
There are already specialist chips with multiple cores - such as those used in router hardware and graphics cards - but Dr Mark Bull, at the Edinburgh Parallel Computing Centre, said multi-core chips were forcing a sea-change in the programming of desktop applications.
"It's not too difficult to find two or four independent things you can do concurrently, finding 80 or more things is more difficult, especially for desktop applications.
"It is going to require quite a revolution in software

Figure 2. The final result of the work in the journalist scenario: this example shows the front page generated by a group of journalists on February 12, 2007.

ness can unexpectedly change when they are integrated in a multi-modal system that aims at acquiring context awareness for providing services in real-time. In some cases, astonishing large error rates can be tolerable if a technology is backed up by a similar one that uses a different modality. In other cases, strict criteria of synchronisation and low delay can arise as a consequence of the integration.

When humans experience the computer-driven service like the Memory Jog, another subjective bias naturally arises: unexpected actions of the service triggered by a false-positive detection of one of the technologies turn out to be far more annoying than a service not provided due to false-negative detection.

The following sections look at each of the applied technologies and describe briefly our experience with their role in the Memory Jog Service. Special focus is given to the usability of the output of their analysis rather than to implementation details or to their individual performance.

3.3. Multi-Modal Perceptual Component: Person and Object Tracking

Person Tracking is based on an Acoustic Localizer and a multi-camera 3D Person and Object tracker. The latter detects regions of interest (e.g., persons, chairs or laptops) via foreground segmentation in each of the five cameras. A three-dimensional representation of these regions of interest is obtained by a ShapeFromSilhouette algorithm [7] that receives the binary foreground masks from all five cameras. These 3D regions of interest are consequently labelled and tracked over time. To resolve ambiguities (people crossing, someone sitting on a chair, etc.), a colour histogram is acquired from each person and object. The output of the multi-camera 3D Person and Object Tracker is enriched by (a) an algorithm that is able to distinguish between an object and a person - assuming an average range of physical properties of adult humans- and (b) an algorithm that analyses human body posture [8] (standing, sitting, etc.) with a standard model of the human body that is aligned to the 3D regions of interest earlier classified as 'person'.

The real-time multi-microphone acoustic localization and tracking system [9] is based on the cross-power spectrum phase from three T-shaped microphone arrays. It is robust to the speaker head orientation and provides one or more 3D localizations with detected acoustic activity. The output of the multi-microphone acoustic localization is enriched by an Acoustic Event Classifier [10] that is based on a combination of ASR features and acoustic features. Typical events such as door opening/closing, phone ringing, chair moving, speech, cough, laugh, etc. can be detected.

The combination of video-based and audio-based tracking systems allows the system to gain a basic understanding of what happens in the smart room:[1]

a) A person of interest (e.g. the latecomer) can be tracked in the room. This location is used to direct the talking head and an automatic cameraman (cf. Section 3.5) to his current position.
b) The position of all participants can be used to guesstimate changes of the state of the session, e.g. between the states "people enter", "meeting starts" or "coffee break".
c) The position of sudden acoustic event can be determined. The automatic cameraman has been configured to capture these events by choosing the camera that is positioned furthest from the location of the acoustic event.
d) In the current implementation of the context awareness, the detection of a latecomer is based on a multitude of criteria amongst which the first two depend on the Person and Object Tracking: increase of the number of person, appearing of a new object close to the door, detection of the acoustic signal of a door-knock, a door-slam or steps close to the door.

[1] This requires proper handling of timestamps. We chose to synchronize the clocks of the involved computers in a network time protocol cluster (NTP peers) and to always forward the timestamp corresponding to the acquisition of the signal (audio or video frame) to the next processing stage.

The delay requirements of the Person Tracker are very relaxed in the application described here. Even a delay of 500 ms in the reaction of the system might be unnoticed. Since the precision of the multi-camera tracker is in the order of a few centimetres, a less precise acoustic localisation of the present speaker is tolerable since the origin of the acoustic energy can be re-mapped to the position of the nearest person detected by the visual tracker.

3.4. Multi-Modal Perceptual Component: Person Identification

Person Identification is based on Face ID [11] and Speaker ID technologies [12]. The Face ID algorithm is applied to faces that are captured either close to the door or images that are captured by the webcam that is mounted on top of the monitor at the console.

In a pre-processing step, a Face Detector is applied on those parts of the image that have previously been classified as 2-dimensional regions of interest in a binary foreground mask in the pre-processing stage of the Person and Object Tracker. Multiple face-like regions collected at different time instants are then analysed to select a frontal view of the face for further processing by the Face Identification technology. Face ID matches these frontal views against faces stored in a database. If no frontal view is available, the algorithm is capable to base the Face ID on side and profile views, although the identification is less reliable.

The second technology, Speaker ID, provides real-time information about the identity of an active speaker. The SpeakerID algorithm is based on the comparison of Gaussian mixture models. Apart from the poor signal-to-noise ratio and the reverberation in the signals obtained from the far-field microphones, this technology is challenged by the necessity to separate the microphone signal into segments uttered by a single speaker. In an unscripted scenario, this would require the detection of speech activity, the detection of a shift of speakers and the detection of the number of persons speaking simultaneously. In order to circumvent this segmentation problem, the ID of the active speaker is only determined during the usage of the dialogue system (cf. Section 3.4), because in this situation the signal naturally stems from a single speaker talking in a quiet background.

Instead of simply listing all recognized detections of the multi-modal Person Identification system, the ID output is assigned to the corresponding person in the Person Tracker. This allows for accumulating the IDs obtained for a person from both Speaker ID and Face ID in the course of the session. The Person Identification technology also provides a valuable feedback to the 3D Person Tracker whether it is still tracking the same person.

Since Person Identification was also used to allow the Talking Head to address the session participants with their name, reducing false positive IDs was emphasised during debugging and error minimization. For the same purpose, both the audio and the video based ID technologies have incorporated a model-class for unknown IDs.

3.5. Dialogue System

The Dialogue System allows a human-like verbal interaction with the Memory Jog System. It is based on two components: A commercially available 2D animation of a talking head (PeoplePutty® by Haptec [13]) and an ASR based dialogue system that utilizes the Cambridge University Engineering Department's HTK [14] recognizer.

Figure 3. Face of the PeoplePutty® Talking Head: The parameters of the talking head software allow adapting its voice, emotions and look to be appropriate for the Dialogue system according to the context.

Trigger sentences		
[PLEASE]	**CHIL ROOM**	**[PLEASE]**
[PLEASE]	**CHIL SERVICE**	**[PLEASE]**
[PLEASE]	**CHIL SYSTEM**	**[PLEASE]**
[PLEASE]	**CHIL ASSISTANT**	**[PLEASE]**
Commands		
	NOTHING	
	NOTHING NOTHING	
[PLEASE]	**SHOW ME THE FRONT PAGE OF OUR COMPETITOR**	**[PLEASE]**
[PLEASE]	**SHOW ME THE FRONT PAGE OF OUR COMPETITORS**	**[PLEASE]**
[PLEASE]	**SHOW ME YESTERDAYS FRONT PAGE**	**[PLEASE]**
	WHO IS OUR FIELD JOURNALIST	**[PLEASE]**
[PLEASE]	**SHOW ME YESTERDAYS BUSINESS NEWS**	**[PLEASE]**
[PLEASE]	**SHOW ME YESTERDAYS ENTERTAINMENT NEWS**	**[PLEASE]**
[PLEASE]	**SHOW ME YESTERDAYS POLITICS NEWS**	**[PLEASE]**
[PLEASE]	**SHOW ME YESTERDAYS SCIENCE NEWS**	**[PLEASE]**
[PLEASE]	**SHOW ME YESTERDAYS TECHNOLOGY NEWS**	**[PLEASE]**

Figure 4. The trigger sentences and the commands understood by the dialogue system. Most of the sentences can be accompanied by an optional “please”.

The speech synthesis part of the talking head was enhanced by a politeness delay unit that acquired a speech activity flag and obliged the speech synthesis engine not to interrupt a human-to-human conversation.[2]

The visual representation of the talking head was projected on one of the walls of the smart room – next to the graphical user interfaces (GUI) used by the journalists to publish the front page (cf. Figs 2 and 7).

The talking head also serves as the voice of the Memory Jog, e.g. giving indications about how to use the GUIs, commenting on acoustic events, welcoming the participants or a latecomer, pointing out when the participants run out of time, congratulating upon a successful contribution, saying good bye, etc.

The KTH based speech recognition was trained for two different grammars: the trigger sentences and the commands. The possible phrases of the two grammars are listed in Fig. 4.

[2] However, we implemented a message-specific timeout after which the talking head was allowed to interrupt human-to-human communication. The adequate duration of this timeout depends on the amount and duration of pauses at speaker change and thus might depend on the cultural background of the meeting participants.

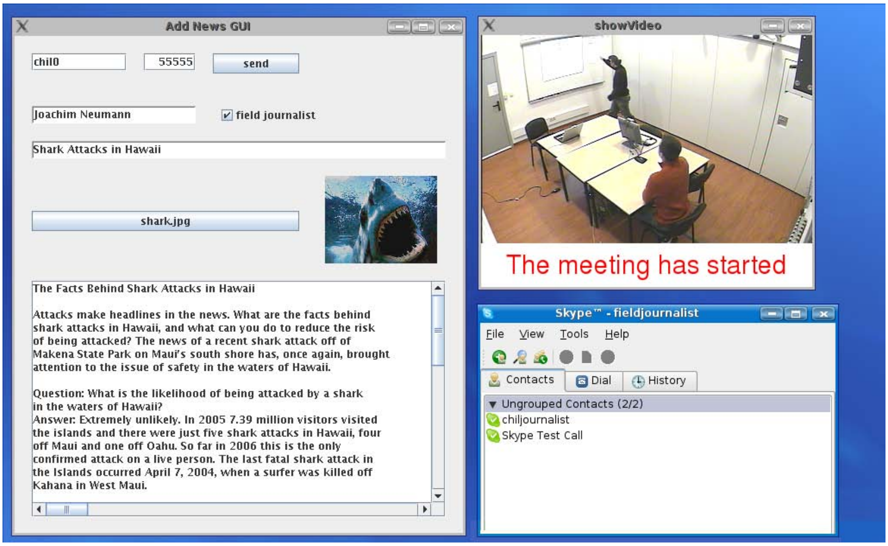


Figure 5. Screenshot of the Field Journalist's laptop: in the lower right, the Skype-based bi-directional audio communication allows talking to the journalists in the room. The upper right shows the real-time video stream from one of the cameras of the meeting room. An automatic cameraman is choosing the optimal camera from five possible angles. This decision is based on the location of the last acoustic event and smoothed by a hysteresis to avoid rapid camera-changes. The real-time video streaming also displays annotations in the form of subtitles that explain the situation, e.g., "people enter", "interaction with ASR", "sound of keys", "front page published", etc or as shown here: "The meeting has started". On the left side of the screen, a graphical user interface allows the field journalist of add a piece of news (a test and an image) to the decision GUI of the journalists in the room.

When not active, the Dialogue System is listening in the human-to-human conversation to detect one of the trigger sentences. If a trigger sentence is detected with a high enough confidence level, the Talking Head utters "Yes, please?" as positive feedback to inform the person in the room who raised the question. When the following command is understood, the corresponding action is initiated. If the Dialogue System wrongly detects a trigger sentence, the user can re-set the Dialogue System by uttering a simple "nothing".

For the Field Journalist who connects from a remote location, an annotated videoconferencing system was developed. The screen of the field journalist's laptop is shown in Fig. 5.

3.6. Personalized Question-Answering System

The Journalists in the smart room have an advanced Question-Answering System [15] at hand that allows then to ask questions related to the news of previous weeks. In comparison to a Google based search engine, this system is capable of giving the exact answer for factual questions instead of merely listing promising text-snippets.

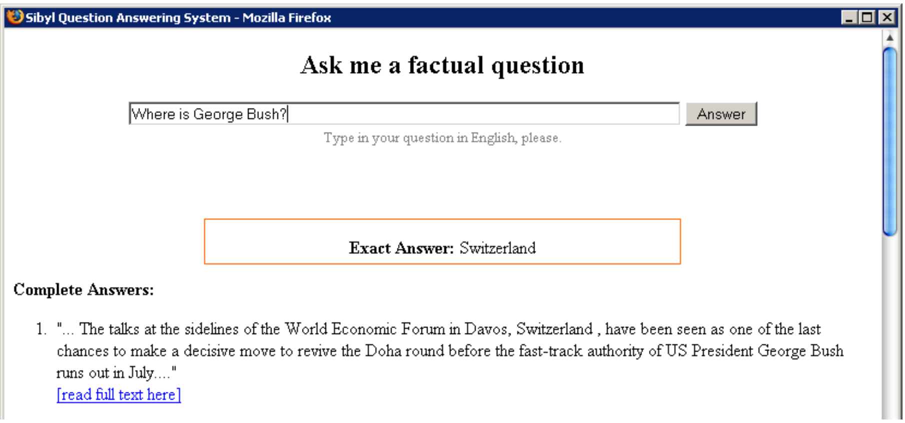


Figure 6. The front end of the personalized Question-Answering system runs in any browser: questions like "Where is George Bush" are answered with a specific location if such a location has been mentioned in the news of the last weeks.

To personalize the output of the Question-Answering System, the result of the Face ID obtained from the webcam at the console is utilised. For this purpose, the field of expertise of each Journalist has been pre-configured. For example, a Journalist working on business news would receive a different answer to the question "Who is meeting in New York?" than a journalist responsible for entertainment news. In order to allow a journalist to access the news from all fields, the automatic selection of specific areas from which the Question-Answering System gains its knowledge can be manually overridden in a small GUI.

3.7. Decision GUI

With this graphical user interface, the journalists decide on the front page. It allows editing the automatically downloaded news text, pre-viewing the resulting front page and also initiates the final publishing stage (cf. Fig. 2).

3.8. RSS Feed of BBC News

A Bash script is executed every five minutes on one of the Linux servers to download the latest RSS feeds[3] from the BBC world news. The downloaded web pages are processed with the text-based Lynx internet browser (lynx.browser.org). The text of the news and a corresponding image are extracted with awk and made available via Samba to the computers that display the graphical user interfaces for the journalists.

4. Software Architecture

Most of the technologies described in the previous section require a high-bandwidth data stream of several hundreds of Megabytes per second and the results of some dozens of analysis algorithms need to be collected in a thread-safe manner be a single ap-

[3] For example: newsrss.bbc.co.uk/rss/newsonline_world_edition/europe/rss.xml.

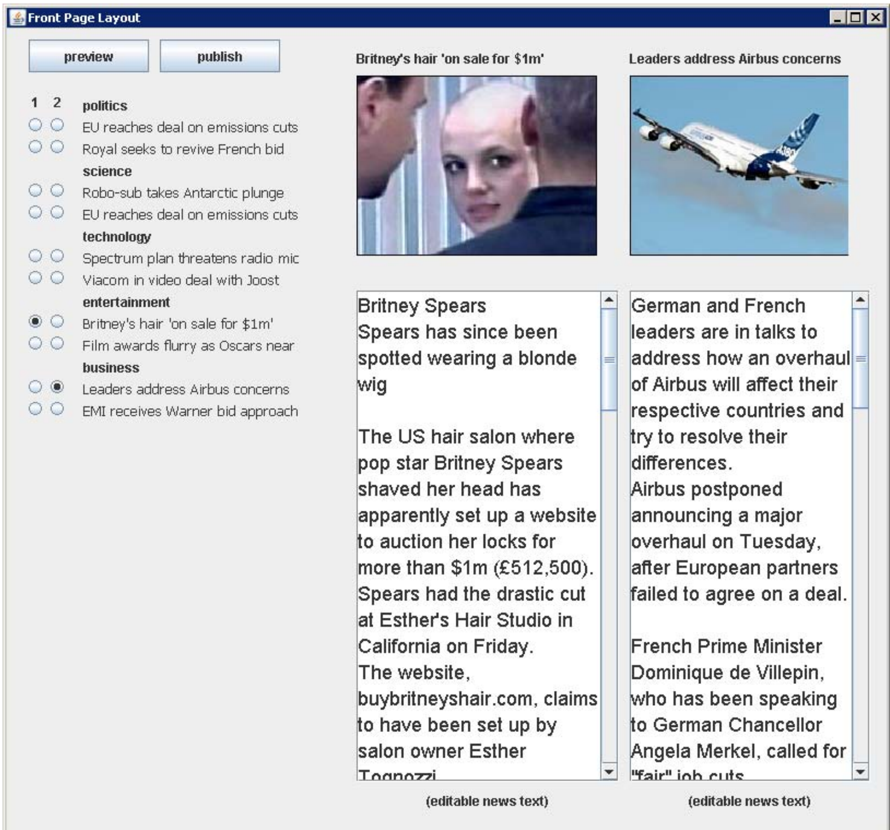


Figure 7. Screenshot of the Decision graphical user interface: this GUI is the main tool of the field journalist. If shows and allows editing of two selected top news. Clicking on the related image directs the journalists to the internet page with the source of the information. The GUI is also projected on one of the walls of the smart room to allow the journalists to move freely in the room while discussing the news.

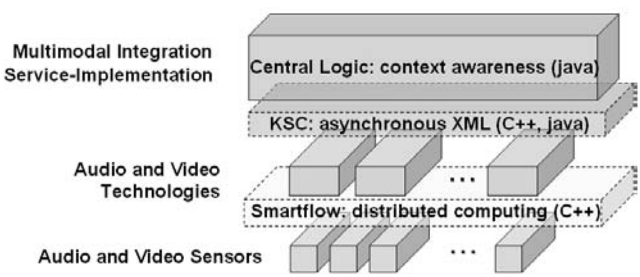


Figure 8. Software architecture in the UPC smart room.

plication which we call the Central Logic. The software architecture chosen in the UPC smart room is based on NIST smartflow system [16] and KSC Socket messaging system. This is illustrated in Fig. 8.

The lower level of the software architecture consists of the video and audio sensors. The signal capture software is implemented as smartflow clients in the computers with the corresponding acquisition hardware. The resulting data streams are transferred as smartflow flows into other computers that can either pre-process the data streams (as in the case of the foreground segmentation which is part of the multi-camera Person Tracker) or directly analyze the raw data streams (as in the case of the Speech Activity Detection audio technology). [17]

Smartflow also provides a mechanism to dynamically decide on which computer in the local area network a specific technology should run – while assuring the correct handling of the data streams between the involved technologies.

The KSC message server and the KSC client library allow sending results of data analysis asynchronously to other technologies or to the Central Logic in a convenient fashion. In order to gain context awareness in the Central Logic framework, some hundreds of KSC messages have to be processed asynchronously in the multi-treaded Central Logic. A multitude of simple voting processes and if-then rules determine shifts of the underlying stage model that comprises the states “people enter”, “instruction”, “meeting”, “dialogue system”, “decision” and “end”. Depending on the present system stage, actions of the Memory Jog Service are triggered by the incoming events: the corresponding visualisations are updated, subtitles are added to the annotated videoconferencing system and the talking head informs the journalist about an event or turns its head towards the location of interest.

5. Conclusion

Our experience with the integration of multiple technologies into two multi-modal perceptual components and an integrated Memory Jog Service that interacts with humans in our smart room was very positive. The technology developers were challenged by the computational demands of a real-time implementation of their technology, signals from unobtrusive sensors and “noise” from a real-world scenario. These constrains tested the technologies at their limits of performance, processing delay and robustness. Still, the system has been successfully put to test with real-users visiting the Smart Room at UPC. A demonstration of the Memory Jog Service is available at the UPC smartroom.

Additionally, the implementation of the multi-modal perceptual components and the integrated service sparked numerous ideas of how to combine technologies in a new way to introduce a higher level of robustness in real-world applications. Subjects that have experienced the Memory Jog Service agree that the service was helpful and some added ideas to our wish list of future services.

References

[1] R. Sharma, V.I. Pavlovic, T.S. Huang, Toward multimodal human-computer interface, *Proceedings of the IEEE* **66** (1998), 853–869.

[2] K. Vredenburg, S. Isensee, C. Righi, *User-centered design: An integrated approach*, Prentice Hall, Englewood Cliffs, NJ, 2001.

[3] CHIL Project website, http://chil.server.de.

[4] R. Stiefelhagen, H. Steusloff, A. Waibel, CHIL – Computers in the Human Interaction Loop, *Proceedings of 5th International Workshop on Image Analysis for Multimedia Interactive Services*, Lisbon, Portugal, April 2004.

[5] *Embodied Conversational Agents*. Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill Editors, MIT Press, Cambridge/London, 2000.
[6] J.R. Casas, R. Stiefelhagen et al, Multi-camera/multi-microphone system design for continuous room monitoring, CHIL-WP4-D4.1-V2.1-2004-07-08-CO, CHIL Consortium Deliverable D4.1, July 2004.
[7] J.L. Landabaso, M. Pardas, Foreground Regions Extraction and Characterization Towards Real-Time Object Tracking, *Machine Learning for Multimodal Interaction*, Springer LNCS **3869** (2006), 241–249.
[8] C. Canton-Ferrer, J.R. Casas, M. Pardàs, M. Human Model and Motion Based 3D Action Recognition in Multiple View Scenarios, European Signal Processing Conference (EUSIPCO), Florence, Italy, September 2006.
[9] A. Abad et al., UPC Audio, Video and Multimodal Person Tracking Systems in the CLEAR Evaluation Campaign, *Multimodal Technologies for Perception of Humans* (CLEAR), Springer LNCS **4122** (2007), 93–104.
[10] A. Temko et al., Evaluation of Acoustic Event Detection and Classification Systems, *Multimodal Technologies for Perception of Humans* (CLEAR), Springer LNCS **4122** (2007), 311–322.
[11] V. Vilaplana, C. Martinez, J. Cruz, F. Marques, Face Recognition Using Groups of Images in Smart Room Scenarios, IEEE International Conference on Image Processing (ICIP), Atlanta (GA), USA, October 2006.
[12] J. Luque, et al., Audio, video and multimodal person identification in a smart room, *Multimodal Technologies for Perception of Humans* (CLEAR), Springer LNCS **4122** (2007), 258–269.
[13] Haptek's PeoplePutty website, http://www.haptek.com.
[14] S.J. Young, The HTK Hidden Markov Model Toolkit: Design and Philosophy, TR 152, Cambridge University Engineering Dept, Speech Group, 1993.
[15] M. Surdeanu, D. Dominguez-Sal, P. Comas, Performance Analysis of a Factoid Question Answering System for Spontaneous Speech Transcriptions, Proceedings of the 10th International Conference on Speech Communication and Technology (Interspeech), Pittsburgh (PA), USA September 2006.
[16] NIST Smart Flow System: http://www.nist.gov/smartspace/nsfs.html.
[17] J. Neumann, Multimodal Integration of Sensor Network, 3rd IFIP Conference on Artificial Intelligence Applications & Innovations (AIAI), Athens, Greece, June 2006.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Robotic Sensor Networks: An Application to Monitoring Electro-Magnetic Fields [1]

Francesco AMIGONI [a,2], Giulio FONTANA [a] and Stefano MAZZUCA [a]
[a] *Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy*

Abstract. Robotic sensor networks are distributed systems in which mobile robots carry sensors around an environment to detect phenomena and produce detailed environmental assessments. In this paper we present a specific robotic sensor network devoted to monitoring electro-magnetic fields.

Keywords. Robotic sensor networks, Multirobot systems, Electro-magnetic fields.

Introduction

Many applications require to monitor an environment in order to detect phenomena and produce detailed environmental assessments [14,22]. These applications, including surveillance, plant and environmental monitoring, and damage assessment, are usually addressed using *sensor networks*, in which sensing nodes are set in fixed locations [26]. The use of multirobot systems for carrying sensors around the environment represents a solution that has recently received considerable attention [1,21] and that can provide some remarkable advantages. For example, these *robotic sensor networks* can dynamically concentrate the sensing robots around a place of interest.

In this paper we present a specific robotic sensor network oriented to monitoring Electro-Magnetic Fields (EMFs). The monitoring of EMF phenomena is extremely important in practice, especially to guarantee the safety of people living and working where these phenomena are significant. It is thus important to localize and monitor the EMF sources in an environment and to assure that the EMF levels are compatible with human exposure [8]. These activities can be naturally carried out by robotic sensor networks, such as the one described in this paper. A robotic sensor network does not require any fixed infrastructure and thus its use is appropriate when a single campaign of EMF measurements is required (e.g., before the opening of a new hospital). The robotic sensor network presented in this paper has been developed as part of a larger project, illustrated in [1].

The architecture of our system is hierarchical. A *coordinator* (a computer) supervises the activities of the system, while a number of *explorers* (mobile robots equipped

[1]Preliminary versions of this work appeared in [3,5].

[2]Corresponding author: Francesco Amigoni, Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy; E-mail:amigoni@elet.polimi.it.

with EMF sensors) navigate in the environment and perform EMF measurement tasks. In our implemented system there are two explorers. The main cycle of activities of the robotic sensor network we implemented can be summarized as follows:

(a) the explorers (independently) measure the EMF in their current positions,
(b) the coordinator integrates these measurements and builds an hypothesis about the location of the EMF sources,
(c) the explorers move in the environment to reach the new interesting measurement positions determined by the coordinator; then the cycle starts again from (a).

We explicitly note that the explorers are autonomous mobile robots that carry both "navigation" sensors (for example, cameras and sonars) and "monitoring" sensors (such as the EMF sensors).

This paper is structured as follows. The next section reviews the relevant literature on robotic sensor networks. Section 2 describes the robotic sensor network we developed to monitor EMF sources. Section 3 presents the experimental validation of our system. Finally, Section 4 concludes the paper.

1. Robotic Sensor Networks

A *sensor network* (*SN*) is a collection of sensory elements or *nodes* that gather, process, and communicate data about environmental phenomena [26]. Practical applications of SNs are countless: the need for sophisticated monitoring systems is growing in several fields, including scientific research, traffic control, medical applications, environmental protection, disaster management, industrial plants, and military applications. A rough classification distinguishes between *fixed sensor networks*, in which nodes are spatially distributed in fixed locations, and *mobile sensor networks*, in which some or all of the nodes can move around the environment [15]. The interest about mobile SNs is motivated by the need to have SNs that are not part of the system to be monitored (e.g., an industrial area), but that can be deployed on location in a short time to cope with fast-evolving and partially-unknown phenomena. To monitor such phenomena with a fixed SN, a large number of nodes scattered over a broad region is required. This implies using very simple, low-cost, low-power, disposable sensory elements with limited processing and communication capabilities [12]. An alternative could be the use of a mobile SN composed of a smaller number of nodes, initially deployed randomly within the area, but able to reach interesting locations to optimize observations and communications. Mobile SNs typically rely on wireless communication systems [22], because wired links are incompatible with their design goals, in particular with a fast deployment phase.

Mobile nodes can be implemented as *autonomous perceptive mobile robots* [21] or as *perceptive robots* [2], whose sensor systems address both navigation and environmental monitoring tasks. In this sense, *robotic sensor networks* are particular mobile sensor networks. The sensing robots can be more complex and expensive than sensing nodes typically used in a fixed SN, enabling them to perform complex processing and communication tasks. Such tasks can include the efficient and autonomous navigation through the environment, the prediction of the evolution of the observed phenomena, and the planning of the actions to be undertaken in order to optimally track the phenomena and to perform sophisticated monitoring, reconfiguration, and self-repairing actions. The main drawback of the multirobot approach to SNs is represented by the intrinsic complexity and the high cost of the robot platforms. A number of applications have been addressed by robotic sensor networks, including environmental monitoring [23] and search and rescue [24].

In the sequel of this section, we review some robotic sensor networks that have been presented in literature. Some of the proposed systems are only simulated (see, for example, [11,17]) or are composed of a single robot (see, for example, [18]). Some works addressed the definition of an architecture for a robotic sensor network. For example, the hierarchical architecture presented in this paper (and similar to that discussed in [4,6]) differs from the architecture proposed in [21] since, in our case, the coordinator is responsible for data integration and for building the model that represents the perceived phenomenon (essentially, the locations of EMF sources), while, in [21], all the sensing robots are involved in the integration of the perceived data. Although hierarchical architectures have some known drawbacks (e.g., scalability), they are easy to implement and adequate for a small number of sensing robots, as in our case (see [19] for another example related to odor maps). Another architecture for robotic sensor networks, based on the decentralized data fusion approach, is proposed in [13].

Many works in literature address a single aspect of robot sensor networks: the deployment of sensing robots [15,16,27,28]. Usually these approaches use a *local* distributed strategy to drive the spatial configuration of the robots: each robot adjusts its position according to those of its neighbors, for example using optimization algorithms based on potential fields. Our architecture adopts a *global* centralized strategy to deploy the sensing robots in the environment: the explorers move in the positions that the coordinator globally evaluates as interesting. In this sense, our approach shares some similarities with that in [7]. The advantage of our approach is that, usually, local approaches are able to deploy sensing robots only to cover an environment [10,20]. There have been some attempts to define local distributed strategies to let sensing robots converge toward events of interest. For example, [9] presents some control strategies similar to potential fields that allow robots to concentrate around events happening in the environment; however, the methods are tested in simulated environments without any reference to a specific application.

2. A Robotic Sensor Network for Monitoring EMFs

In this section, we illustrate a specific robotic sensor network we have developed in our Artificial Intelligence and Robotics Lab (AIRLab). The system we present here is oriented to localize and characterize a (possibly moving) EMF source. We first describe the general architecture of the system, then the localization algorithm it uses.

2.1. General Architecture

As said, the architecture of our system is hierarchical with a coordinator supervising two explorers (more explorers can be easily added). We assume that the three agents can communicate with each other by exchanging messages. The system maintains a grid map of the environment, in which each cell can be either free or occupied by an obstacle (or by a robot). The map is supposed to be known by the coordinator and the explorers. The

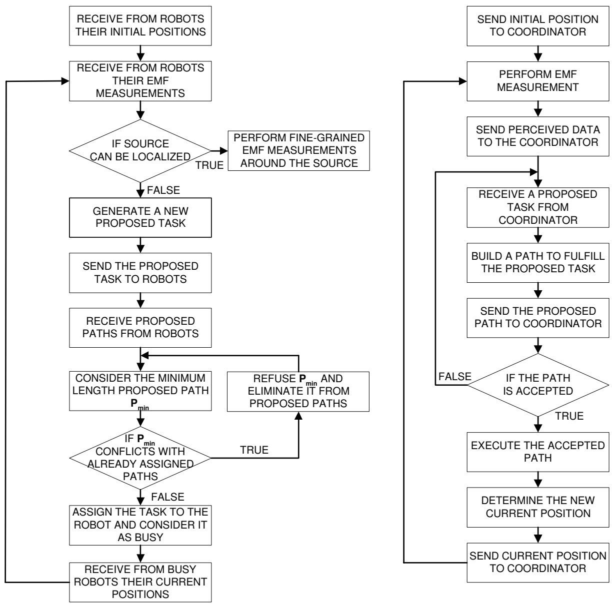


Figure 1. The flows of activities of coordinator (left) and of explorers (right)

environment is assumed to be static (i.e., the positions of the obstacles do not change while the system is working). The map is used by the explorers to navigate in the environment and by the coordinator to localize the EMF sources. For simplicity, we consider a single EMF source. The position of the EMF source is thus given by the cell of the map in which the source is located.

The activities of the coordinator and of the explorers during the EMF monitoring of an environment are schematically reported in Figure 1. EMF monitoring involves two activities: localizing the source of the field and characterizing the source, namely checking if the emissions of the source are compatible with human exposure. The localization of the source is performed according to the algorithm described in the next section. Once a source has been localized, it has to be reached by an explorer both for confirming the position and for more detailed EMF measurements in the region around the source (characterization). In this paper, we focus on the localization of the source.

We describe one of the most significant features of our system: the task allocation. We use the *contract net* paradigm for allocating tasks to the robots [25]. The allocated tasks are EMF measurements to be performed in given positions. The coordinator proposes a cell for visit if it needs an EMF measurement from that cell, either for estimating the position of the source (see next section) or for determining whether the EMF is within the limits established by the law for human exposure. The coordinator sends the proposed task to free explorers, waits for their proposed paths (or until a timeout expires), and assigns the task to an explorer. A proposed path is a sequence of movements that brings an explorer from its current position to the cell to be visited. The coordinator considers the proposed path of minimum length and assigns the task to the proposing explorer. The other explorers are notified that their proposed paths have been rejected.

For more details on the architecture, please refer to [5].

2.2. The Localization Algorithm

In this section, we discuss the algorithm for the localization of the EMF sources that represents the “core” algorithm of our robotic sensor network. In general, a localization algorithm strongly depends from the available sensors. The sensors we used have been developed by the Dipartimento di Elettrotecnica of the Politecnico di Milano and are extensively described in [8]. They can detect only the magnetic field, but they still allow to localize a EMF source. Each sensor is connected through an RS232 serial link to the computer that governs the explorer. We used two sensors, mounted on the two explorers at (approximately) the same height. For each spatial component of the magnetic field, the sensor outputs the modulus of its DFT (Discrete Fourier Transform) evaluated at 10000 frequency points separated by nearly 1 Hz intervals in the range 10 Hz $-$ 10 kHz. Note that, since the sensors we use are isotropic, the coordinator requests a measurement from a position, without specifying the heading. Note also that, when the coordinator requests a measurement from a position, the measured data will be the same if this measurement is performed by one explorer or by the other one.

Our localization algorithm is based on the following general assumptions:

- The environment contains a single, possibly moving, source of magnetic field. Other less powerful sources can be present, but we require the magnetic field generated by the main source to be (as measured by the sensor) significantly stronger than the background field in the environment.
- In each of the measurement positions the magnetic field vector has a non-null radial component and a null tangential component, as evaluated in the two-dimensional polar coordinate system defined in the horizontal plane $\mathcal{H}$ containing the sensor and centered on the projection of the source on $\mathcal{H}$. In our experimental setting, we enforced this by using a solenoid with its axis orthogonal to the floor as source, by performing near-field measurements, and by avoiding to take measurements on the horizontal plane containing the source.
- No *a priori* assumption is made on the spectrum of the magnetic field produced by the source. However, since the sensor we employed ignores the frequency components outside the 10 Hz $-$ 10 kHz range, it can only detect sources which have significant magnetic emissions in that frequency band.

The localization is based on triangulation, which in turn is based on the fact that, under the above assumptions, the *line of propagation* (*LOP*) of the magnetic field can be extracted from the sensor data. A LOP is the line connecting the (center of the) magnetic field source to the measurement position, as projected on $\mathcal{H}$. In principle, with a fixed source, if we can extract from the measured data two different LOPs, their intersection gives the source position. In order to obtain two different LOPs, at least two magnetic field measurements, taken in two different positions in the environment, are needed. Let us consider a generic spatial position P in the environment. If the three spatial components of the magnetic field in P were available, the calculation of the LOP passing through P would be trivial. Unfortunately, the sensors we used do *not* output these spatial components: they output instead the modulus of their DFTs, but not the associate phase information. This implies that for each position P two *possible LOPs* (*pLOPs*) are generated. One of them is the actual LOP, while the other one is an artifact of the generation process. With the data from a single measurement it is not possible to choose between them. The algorithm for pLOPs extraction is described in detail in [3], along with a description of the possible special cases.

To determine the actual source position our localization algorithm uses the data returned by three measurements. These must be taken at three different positions in the environment because, if the source is stationary, successive measurements performed in the same position would lead to the same pLOPs. After the first two measurements, we obtain a set of 4 *possible source locations* (*pSLs*). These pSLs are the intersection points between the 4 pLOPs returned by the two measurements (Figure 2). Then, a third measurement is used. This also gives two pLOPs and, since the source must belong to one of them, one of these new pLOPs must pass through one of the 4 pSLs. So, checking the distance between the new pLOPs and the 4 pSLs, we can identify the actual source position among the pSLs. If none of the new pLOPs pass through one of the pSLs, it means that the source has moved and the three measurements refer to different source positions. There are degenerate cases where the relative locations of the three measurement positions do not allow source localization (for details see [3]): in these cases more than three measurements are required to complete the localization process.

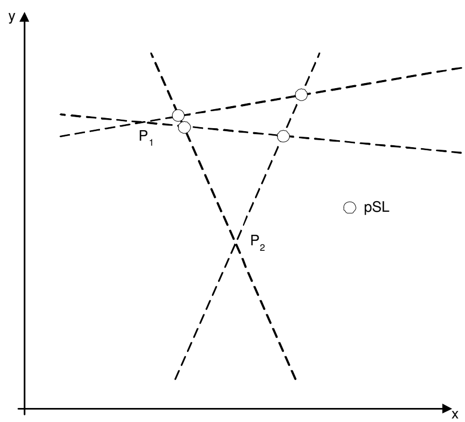


Figure 2. Possible source locations after two measurements in positions P_1 and P_2

The system is presently configured to use the last three available measurements to locate the source, in order to quickly react to source movements. However, it is possible to use more measurements to improve the accuracy of source localization. Once the position of the source is known, it is trivial to find the cell corresponding to the source in the grid map of the environment.

3. Some Experimental Results

In this section, we present some of the experiments we performed to validate our robotic sensor network. In particular, first we discuss an example of operations performed by the system and, second, we illustrate some performance results.

In the first experiment, we show a test with a moving EMF source. We used simulated EMF sensors that return data in accordance with the position of a simulated EMF

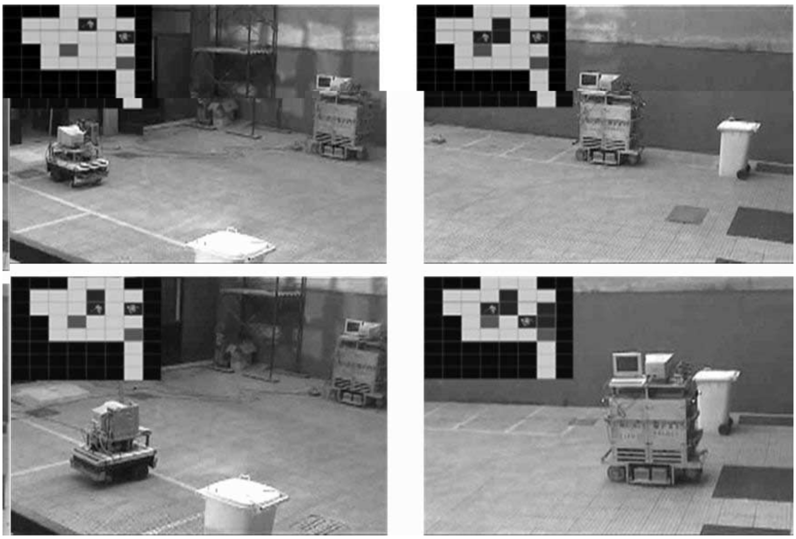


Figure 3. The explorers move toward a EMF source changing its position

source placed in the environment via a graphical interface. As our simulated sensors give complete information about the magnetic induction field vector, only two measurements are needed to determine the source location, instead of the three needed with the real sensors. In detail, the following activities have been performed (see Figure 3):

- the explorers performed an EMF measurement in their initial locations;
- the coordinator used the data gathered by the explorers to determine the estimated position of the EMF source (represented by the white trash bin, left top of Figure 3);
- the coordinator generated a task that required to perform EMF measurements around the source position;
- after having negotiated the assignment of the task, one of the explorers (the closest one) moved toward the cell where the EMF source was estimated to be (left bottom of Figure 3): each time the robot reached the center of a cell along the path, it executed an EMF measurement (these data were used by the coordinator to confirm the hypothesis about source position);
- we manually changed the position of the EMF source in the the environment (right top of Figure 3);
- after one of these measurements the coordinator reacted to the change of the EMF source by changing the hypothesis and identifying a new task;
- after having negotiated the assignment of the new task, the second explorer (closer than the first one) moved toward the cell where the EMF source was now estimated to be (right bottom of Figure 3).

We also performed a number of experiments to evaluate the performances of our robotic sensor network with real magnetic field sensors, using a simple single-frequency magnetic field generator. We built this test source using $200\ \mathrm{m}$ of electrician wire, tightly wound in a cylinder approximately $20\ \mathrm{cm}$ in diameter and $15\ \mathrm{cm}$ in height, with a central

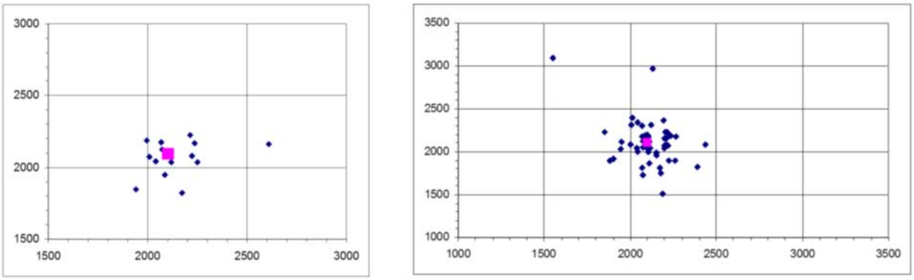


Figure 4. Source location (the square at $(2100, 2100)$ represents the center of the source, units are mm) and estimated source positions (diamonds)

hollow with a diameter of approximately $8\ \mathrm{cm}$. The axis of the cylinder was orthogonal to the floor plane. We connected this device to the $230\ \mathrm{V} - 50\ \mathrm{Hz}$ power line, in series with a $500\ \mathrm{W}$ halogen lamp used as resistive load.

We tested the effective localization of the EMF source provided by the system. We put the center of the EMF source at coordinates $(2100, 2100)$ (units are mm) and performed its localization by randomly generating measurement positions until localization was possible. Recalling previous section, this usually requires three measurement positions; but when the relative location of source and measurement positions leads to a degenerate case or signals are too weak, additional measurements are required to complete the localization process. Measurement positions for the explorers have been generated within a distance of $2\ \mathrm{m}$ from the source. We repeated the test 15 times. Each test was stopped when the system produced an estimated source position. Results are shown in Figure 4 (left). The position of the source has been determined with good precision in almost all tests. Only in one test the source has been localized with a precision worse than $40\ \mathrm{cm}$ (recall that the diameter of our source is $20\ \mathrm{cm}$). In 13 tests, the source has been localized after with three measurements, according to what discussed in Section 2.2. In the other 2 tests, four measurements have been required to disambiguate source position.

We report another experiment we have conducted in order to validate the reliability of the localization of the EMF source provided by the system. Also in this case, we put the EMF source at coordinates $(2100, 2100)$ but we performed its localization by randomly generating 9 measurement positions. In this way, by using a number of measurements larger than the minimum required for the localization, we could validate the robustness of the localization. Measurement positions for the explorers have been generated within a distance of $2\ \mathrm{m}$ from the source. We repeated the test 8 times. In principle, with 9 measurements, each test should produce 7 estimated source positions (because localization cannot be done when only the first two measurements are available), and, over all the tests, we should have 56 estimated source positions. However, we could get only 53 estimated source positions, because in two cases weak and disturbed signals prevented the localization of the source and in another case a further measurement was required to disambiguate source position. Results are shown in Figure 4 (right). The position of the source has been determined with good precision; only in a dozen of cases the source has been localized with a precision worse than $40\ \mathrm{cm}$ (recall that the diameter of our source is $20\ \mathrm{cm}$). We averaged the estimated source positions obtained in the single tests, obtaining 8 “average” estimated source positions that are shown in Figure 5.

A final remark is on our localization algorithm. Since it works by considering the last three measurements, the algorithm can easily track moving sources (at least when this movement is slow compared to that of the robots); however, the algorithm does not increase the precision of the estimated source position over time because it does not take into account all the previous measurements.

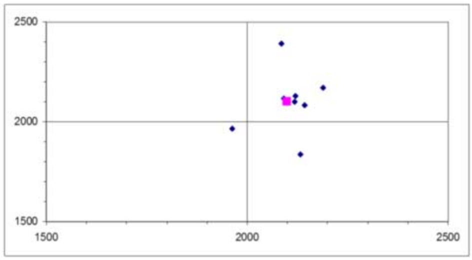


Figure 5. Source location (the square at $(2100, 2100)$ represents the center of the source, units are mm) and "average" estimated source positions (diamonds)

4. Conclusions

In this paper we have presented a multirobot system devoted to EMF monitoring. The system is a robotic sensor network that can autonomously deploy its explorers in an environment to cope with dynamic events, like a moving EMF source. Despite some features of our system have been specifically designed for the application we tackled, some architectural solutions (such as the contract net for assigning tasks to explorers) are quite general and could be easily reused in other application domains.

The importance of robotic sensor networks is expected to increase in the near future, due to the technological advances in miniaturization and communication. One of the most interesting challenges for the next generation robotic sensor networks will be the capability to address more tasks at the same time: for instance, the simultaneous monitoring of multiple environmental phenomena and map building.

Acknowledgements

This work is part of the APE Project funded by the MIUR (Italian University and Research Ministry) in 2002-2004.

References

[1] F. Amigoni, A. Brandolini, V. Caglioti, V. Di Lecce, A. Guerriero, M. Lazzaroni, F. Lombardi, R. Ottoboni, E. Pasero, V. Piuri, O. Scotti, and D. Somenzi. Agencies for perception in environmental monitoring. *IEEE Transactions on Instrumentation and Measurements*, 55(4):1038–1050, 2006.

[2] F. Amigoni, A. Brandolini, G. D'Antona, R. Ottoboni, and M. Somalvico. Artificial intelligence in science of measurements: From measurement instruments to perceptive agencies. *IEEE Transactions on Instrumentation and Measurement*, 52(3):716–723, 2003.

[3] F. Amigoni, S. Cadonici, V. Caglioti, and G. Fontana. Experimenting with a robotic system for localizing magnetic field sources. In *Proc. IEEE VECIMS2005*, pages 44–49, 2005.

[4] F. Amigoni and V. Caglioti. A multirobot architecture for environmental perception. In *Proc. Second Int'l Workshop on Advanced Environmental Sensing and Monitoring Technologies*, pages 83–88, 2003.

[5] F. Amigoni, V. Caglioti, and G. Fontana. A perceptive multirobot system for monitoring electromagnetic fields. In *Proc. IEEE VECIMS2004*, pages 95–100, 2004.

[6] F. Amigoni and M. Somalvico. Multiagent systems for environmental perception. In *Proc. Third AMS Conference on Artificial Intelligence Applications to Environmental Science*, 2003.

[7] M. Batalin and G. Sukhatme. Using a sensor network for distributed multi-robot task allocation. In *Proc. IEEE ICRA2004*, pages 158–164, 2004.

[8] A. Brandolini, G. D'Antona, M. Faifer, M. Lazzaroni, and R. Ottoboni. Low frequency magnetic flux density measurements based on navigation agents. In *Proc. SICON04 Conference*, pages 86–90, 2004.

[9] Z. Butler and D. Rus. Event-based motion control for mobile sensor networks. *IEEE Pervasive Computing*, 2(4):34–42, 2003.

[10] J. Cortés, S. Martínez, and T. Karatas ans F. Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on Robotics and Automation*, 20(2):243–255, 2004.

[11] J. Costa Seca, C. Pinto-Ferreira, and L. Correia. A society of agents in environmental monitoring. In *From Animals to Animats 5*, pages 447–452. Bradford Book, 1998.

[12] J. Feng, F. Koushanfar, and M. Potkonjak. System-architectures for sensor networks issues, alternatives, and directions. In *Proc. IEEE ICCD2002*, pages 226–231, 2002.

[13] B. Grocholsky, V. Kumar, and H. Durrant-White. Anonymous cooperation in robotic sensor networks. In *Proc. AAAI-04 Workshop on Sensor Networks*, 2004.

[14] M. F. Guard. A status report on environmental monitoring. *IEEE Transactions on Instrumentation and Measurement*, 51(4):782–785, 2002.

[15] A. Howard, M. Mataric, and G. Sukhatme. An incremental self-deployment algorithm for mobile sensor networks. *Autonomous Robots*, 13(2):113–126, 2002.

[16] A. Howard, M. Mataric, and G. Sukhatme. Mobile sensor network deployment using potential fields: A distributed, scalable solution to the area coverage problem. In *Proc. DARS2002*, pages 299–308, 2002.

[17] J. P. Jamont, M. Occello, and A. Lagreze. A multiagent architecture for the instrumentation of an underground hydrographic system. In *Proc. IEEE VIMS2002*, pages 20–25, 2002.

[18] A. Lilienthal and T. Duckett. Building gas concentration gridmaps with a mobile robot. *Robotics and Autonomous Systems*, 48(1):3–16, 2004.

[19] L. Marques, A. Martins, and A. T. de Almeida. Environmental monitoring with mobile robots. In *Proc. IEEE/RSJ IROS2005*, pages 3624– 3629, 2005.

[20] Y. Mei, Y.-H. Lu, C. Hu, and C. S. G. Lee. Deployment strategy for mobile robots with energy and timing constraints. In *Proc. IEEE ICRA2005*, pages 2816–2821, 2005.

[21] E. M. Petriu, T. E. Whalen, R. Abielmona, and A. Stewart. Robotic sensor agents: A new generation of intelligent agents for complex environment monitoring. *IEEE Instrumentation and Measurement Magazine*, 7(3):46–51, 2004.

[22] G. J. Pottie and W. J. Kaiser. Wireless integrated network sensors. *Communications of the ACM*, 43(5):51–58, 2000.

[23] M. Rahimi, R. Pon, W. Kaiser, G. Sukhatme, D. Estrin, and M. Srivastava. Adaptive sampling for environmental robotics. In *Proc. IEEE ICRA2004*, 3537-3544, 2004.

[24] J. Reich and E. Sklar. Robot-sensor networks for search and rescue. In *Proc. IEEE Int'l Workshop on Safety, Security and Rescue Robotics*, 2006.

[25] R. G. Smith. The contract net protocol: High-level communication and control in a distributed problem solver. *IEEE Transactions on Computers*, 29(12):1104– 1113, 1980.

[26] M. Tubaishat and S. Madria. Sensor networks: An overview. *IEEE Potentials*, 22(2):20–23, 2003.

[27] A. Winfield. Distributed sensing and data collection via broken ad hoc wireless connected networks of mobile robots. In *Proc. DARS2000*, pages 273–282, 2000.

[28] Y. Zou and K. Chakrabarty. Sensor deployment and target localization based on virtual forces. In *Proc.IEEE InfoCom2003*, pages 1293–1303, 2003.

Emerging Artificial Intelligence Applications in Computer Engineering
I. Maglogiannis et al. (Eds.)
IOS Press, 2007

Assembling Composite Web Services from Autonomous Components

Jyotishman PATHAK, Samik BASU and Vasant HONAVAR
Department of Computer Science
Iowa State University
Ames, IA 50011-1040, USA
{*jpathak, sbasu, honavar*}@*cs.iastate.edu*

Abstract.
Web services are fast emerging as the technology of choice to build distributed information systems in multiple domains including e-Business and e-Science. An important challenge is to develop methodologies and tools that enable (semi-) automatic composition of services by taking into account the functional, non-functional and behavioral requirements of the service developer. This paper presents the fundamental concepts and issues related to service composition and provides a representative sample of existing work proposed by the AI planning and formal methods communities to address some of the challenges in service composition. It also provides a brief introduction to an iterative and incremental technique for modeling composite Web services proposed by the authors.

Keywords. Web Services, Composition, AI Planning, Formal Methods

1. Introduction

Recent advances in networks, information and computation grids, and the World Wide Web has led to the emergence of a new approach to build highly distributed information systems using Web-based services[1] [1]. They hold the promise to enable development of rich and flexible applications in multiple domains including e-Business and e-Science owing to their loosely coupled and interoperable nature. Consequently, there has been a significant interest in recent years to build Service-Oriented Architectures [2,3] that support the creation and deployment of complex Web services to accomplish different tasks. In particular, the ability to discover and integrate existing services into a composite service (a.k.a. *Web Service Composition*) has received a lot of attention from both academia and industry, and many techniques based on formal methods, AI planning and logic theory have been proposed. The main objective of these approaches is to allow service developers to flexibly locate the required component services, compose them and orchestrate their execution to achieve the desired requirements, which otherwise cannot be fulfilled by a single (available) service.

[1]In this paper, we use the terms "Web services" and "services" interchangeably.

However, developing tools and techniques for service composition is non-trivial due to many inherent challenges. These include:

- How to search for the most suitable set of services that when composed appropriately will satisfy the desired requirements?
- How to model expressive description languages for representing services and service compositions?
- How to validate and verify the behavioral and executional properties of the composite service?
- How to enable execution monitoring, repair and adaptation of service compositions?
- How to build user-friendly and intuitive tools for modeling real-world complex services?

In other words, automatic composition of Web services require addressing various challenges related to Web service discovery, integration, orchestration, verification and execution monitoring. In addition, to ensure that the proposed techniques can be applied in practical settings, their efficiency, scalability and usability are of great significance.

Against this background, in this paper we discuss few representative approaches to Web service composition that attempt to address some of the challenges outlined above. In particular, we focus on approaches that rely on techniques developed primarily by the AI planning and formal methods communities (Section 2). Even though the specific research topics addressed by these two areas may vary, we have noticed a strong parallelism between the Web service composition techniques based on them. Consequently, we have explored such a synergy to propose a novel framework for iterative and incremental modeling of composite Web services which we discuss briefly in this paper (Section 3). Finally, we conclude the paper by outlining some of the open research problems (Section 4) that we believe have to be addressed effectively to develop robust, efficient and practical tools and techniques for Web service composition.

2. State-of-the-art in Web Service Composition: A Short Survey

2.1. *What is Web Service Composition?*

Web services are software system designed to support interoperable machine-to-machine interaction over a network [4]. Typically, they have an interface described in a machine-processable format which specify their functionalities that can be invoked by other systems through message-based interactions. However, in certain cases a desired functional (and/or non-functional) requirement cannot be met by a single Web service in its entirety, but could be possibly met by appropriately integrating and composing a set of available services. Informally, given a user (goal) requirement G and a set of available Web services $W = \{W_1, W_2, \ldots, W_n\}$, Web service composition amounts to: (i) generating a new composite Web service W_G (Figure 1(a)) by suitably combining a subset of available services, $W_i \ldots \otimes \ldots W_j \ldots \otimes \ldots W_k$, where $\otimes$ is the composition operator, or (ii) establishing a linkage structure $L = \{L_{ij}, L_{ik}, \ldots\}$ (Figure 1(b)) between the participating services, $W_i, \ldots W_j, \ldots, W_k$, that will allow them to communicate directly via message exchanges. The former approach, commonly referred as the *mediator-*

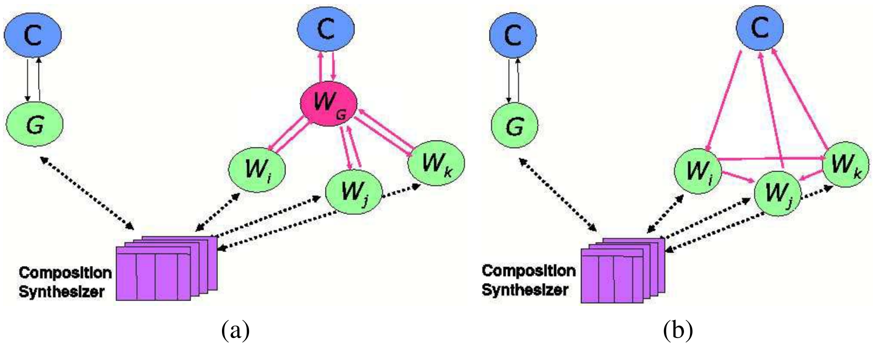


Figure 1. Two different types of composition [8]: (a) Mediator-based (b) Choreography-based

based composition, results in developing a mediator which will enable communication between the client (either a human being or a software agent) and the participating services. Each and every message exchange between the client and the participating services is channeled through the mediator. Whereas, in the latter approach, commonly referred as the *choreography-based composition*, the message exchange channels or links are established between the participating services themselves. Thus, the client communicates directly with the respective services.

Many techniques have been proposed in the literature (see [5,6,7] for surveys) for both the approaches in the recent years. In particular, there has been a significant interest from researchers in the AI planning and formal methods communities to develop techniques for service composition that leverage and build on the existing work. We discuss a representative sample of such approaches in the following sections.

2.2. AI Planning for Web Service Composition

In general, AI planning can be regarded as an area of study that is concerned with automatic generation of plans that will be able to solve a problem within a particular domain. Typically, a plan consists of sequence of actions, such that given an initial state or a condition, a planner will suitably select a set of actions which, when executed according to the generated plan, will satisfy certain goal conditions. In the context of Web services, a planning domain can be represented by a sextuplet $(W, S, A, \longrightarrow, s_0, s_G)$, where W is the set of available Web services, S is the set of all possible states of these services (world), A is the set of actions/functions provided by the services that the planner can perform in attempting to change the state from one to another in the world, $\longrightarrow \subseteq S \times A \times S$ is the set of state transitions which denote the precondition and effects for execution of each action, and finally $s_0 \in S$ and $s_G \in S$ are the initial and goal states, respectively, specified in the requirement of the Web service requesters to indicate that the plan initiates its execution starting from state s_0 and terminates at state s_G. Given this domain, many approaches have been by applying a variety of planning techniques that will generate a plan for realizing the goal requirements.

PDDL [9] (Planning Domain Definition Language) is one of the very widely known description languages in the planning domain and has influenced the development of Web service description languages such as OWL-S [10] (Web Ontology Language for Services). McDermott [11] extended PDDL by introducing the notion of "value of the action", essentially representing certain information that is created or learned as a consequence of executing a particular action. The main intention of introducing this extension was to have the ability to capture the information and the content of messages that are exchanged between the services. The work demonstrates how this extended language can be used with estimated regression planners to create conditional plans that achieve the desired goal. Medjahed et al. [12] applied a rule-based planning technique for finding feasible compositions and introduced a declarative language for describing the goal requirements. The core of the approach comprised of developing composability rules that consider and analyze syntactic and semantic properties of the Web services to devise a plan. Such rules, for example, might specify that two Web services W_1 and W_2 are composable only if the output messages of W_1 are compatible with the input messages of W_2. Sycara et al. [13] proposed an approach for automatic discovery, interaction and composition of semantic Web services using HTN (Hierarchical Task Network) planning. Their technique represents services using DAML-S [14] (Darpa Agent Markup Language for Services, the predecessor of OWL-S), and provides multiple "degree of match" criteria to determine whether the service provider capabilities conform to the requirements of the requester. Another approach which relied on using the HTN planner for automatic composition of services described in OWL-S was proposed in [15]. The authors provide an algorithm for translating OWL-S service descriptions into SHOP2 (an HTN planner) domain and prove the correctness of their approach by showing the correspondence to the situation calculus semantics of OWL-S. SEMAPLAN [16] attempts to leverage traditional AI planning and information retrieval techniques for building a semi-automated service composition tool. The technique relies on domain-dependent/independent ontologies [17] for calculating semantic similarity scores between the concepts/terms in service descriptions, and applies this score to guide the searching process of the planning algorithm. The experimental results demonstrate that SEMAPLAN performs superior compared to the traditional planning based techniques. A similar approach is also proposed in [18] which attempts at combining traditional AI planning techniques with semantics-based approaches to build an end-to-end solution for service composition.

2.3. *Formal Methods for Web Service Composition*

Formal Methods is an area of study that provides a language for describing a software artifact (e.g., specifications, design, source code) such that formal proofs are possible, in principle, about properties of the artifact so expressed. In the context of Web service composition, typically the property proved is that an implementation is functionally correct, that is, it fulfills a particular specification. In the recent past, many research efforts for service composition have adopted formal methods techniques to leverage its mathematically-precise foundation for providing theoretically sound and correct formalisms. We discuss few of those approaches in the following paragraphs.

Pistore et al. [19,20] represent Web services using transition systems [21] that communicate via exchanging messages. Their approach relies on symbolic model checking techniques to determine a parallel composition of all the available services, and then generate a controller to control the composed services such that it satisfies the user-specified requirements. Informally, if $W = \{W_1, W_2, \ldots W_n\}$ is the set of available services, ρ is the user-specified requirement (i.e., ρ describes the goal G), and $\|$ is the composition operator, the aim is to determine a "controller" W_c, such that: $W_c \triangleright (W_1 \| \ldots \| W_n) \models \rho$. Similarly the Colombo framework [8] models Web services using labeled transition systems. However, this approach relies on specifying linkages to establish communication channels between services that have identical signatures and exploits them to determine a feasible composition that satisfies the goal requirements by reducing the composition problem to satisfiability of a suitable deterministic propositional dynamic logic formula. On a slightly different note, Gwen Salaün et al. [22] apply Process Algebra [23] (PA) to model Web services in at least two different ways: (i) at *design time*, PA can be used to describe an abstract specification of the system to be developed, which can be validated and used as a reference for implementation; (ii) by applying *reverse engineering*, existing Web service interface descriptions can be translated to PAs. Specifically, this work adopted CCS [21] as the PA and demonstrated techniques for translating BPEL [24] (Business Process Execution Language) processes into CCS, which can then be verified to reason about properties specified in temporal logic. Hamadi and Benatallah [25] apply a petri net-based algebra to model the control flow and capture semantics of complex Web service compositions. Their framework provides various control flow constructs such as sequence, alternative, iterative and arbitrary, and the authors show how these constructs can be used to determine and verify a composition. However, it is unclear whether the composition is done (semi-) automatically or manually. SELF-SERVE [26] extended this work by providing the ability for dynamically composing and executing Web services represented as state charts. One of the key features of SELF-SERVE is to adopt a peer-to-peer (P2P) computing environment for executing the (composite) services, which in practice has multiple advantages (in terms of scalability, fault-tolerance etc.) compared to centralized architectures.

3. The MoSCoE Approach

3.1. Problems with Existing Web Service Composition Techniques

In the previous sections, we have outlined some of the representative approaches that apply AI planning and formal methods techniques for building (semi-) automated solutions to Web service composition. Nevertheless, these approaches have several drawbacks that limit their practical viability and wide-scale adoption. These limitations include:

- *Complexity of Modeling Composite Services:* For specifying functional requirements, the current techniques for service composition require the service developer to provide a specification of the desired behavior of the composite service (goal) in its entirety. Consequently, the developer has to deal with the cognitive burden of handling the entire composition graph (comprising appropriate data and control flows) which becomes hard to manage with the increasing complexity of the goal service. Instead, it will be more practical to allow developers to begin with an abstract, and possibly incomplete, specification that can be incrementally modified and updated until a feasible composition is realized.
- *Inability to Analyze Failure of Composition:* The existing techniques for service composition adopt a ‘single-step request-response’ paradigm for modeling composite services. That is, if the goal specification provided by the service developer cannot be realized by the composition analyzer (using the set of available component services), the entire process fails. As opposed to this, there is a requirement for developing approaches that will help identify the cause(s) for failure of com-

position and guide the developer in applying that information for appropriate reformulation of the goal specification in an iterative manner. This requirement is of particular importance in light of the previous limitation because in many cases the failure to realize a goal service using a set of component services can be attributed to incompleteness of the goal specification.

- *Inability to Analyze Non-Functional Characteristics:* Barring a few approaches, most of the techniques for service composition focus only on the functional aspects of the composition. In practice, since there might be multiple (available) services that can provide the same functionality, it is of interest to explore the non-functional properties of the components to reduce the search space for determining compositions efficiently.
- *Inability to Handle Differences in Service Semantics:* Individual Web services needed for realizing a desired functionality are often developed by autonomous groups or organizations. Consequently, semantic gaps, arising from different choices of vocabulary or ontologies for specifying the behavior of the services, are inevitable. This requires frameworks for assembling complex Web services from independently developed component services to provide support for bridging the semantic gaps.

To overcome some of these limitations, we are working towards developing a novel Web service composition and execution framework called MoSCoE[2] [27,28,29] (*Modeling Web Service Composition and Execution*) that is based on three basic principles namely *abstraction, composition* and *reformulation*. By abstraction, we refer to the ability of MoSCoE that allows the users (i.e., service developers) to specify an abstract and possibly incomplete specification of the (goal) service. This specification is used to select a set of suitable component services such that their composition realizes the desired goal in terms of both functional and non-functional requirements. In the event that such a composition is unrealizable, the cause for the failure of composition is determined and is communicated to the user thereby enabling further reformulation of the goal specification. This process can be iterated until a feasible composition is identified or the user decides to abort. We discuss further details about the MoSCoE framework and the system architecture in the following section.

3.2. MoSCoE Framework

Figure 2 shows the architectural diagram of the MoSCoE framework illustrated above. As mentioned, the system accepts from the user, an abstract (high-level and possibly incomplete) specification of the goal service. In our current implementation, the goal service specification takes the form of an state machine that provides a formal, yet intuitive specification of the desired goal functionality. This goal service and the available component services (published by multiple service providers) are represented using labeled transition systems augmented with state variables, guards and functions on transitions, namely, Symbolic Transition Systems[3] (STS), where states are abstraction of the service configuration and transitions represent the way in which such configurations are

[2] `http://www.moscoe.org`

[3] The STS specifi cations for component services can be obtained from service descriptions provided in high-level languages such as BPEL or OWL-S by applying translators similar to those proposed in [30,31].

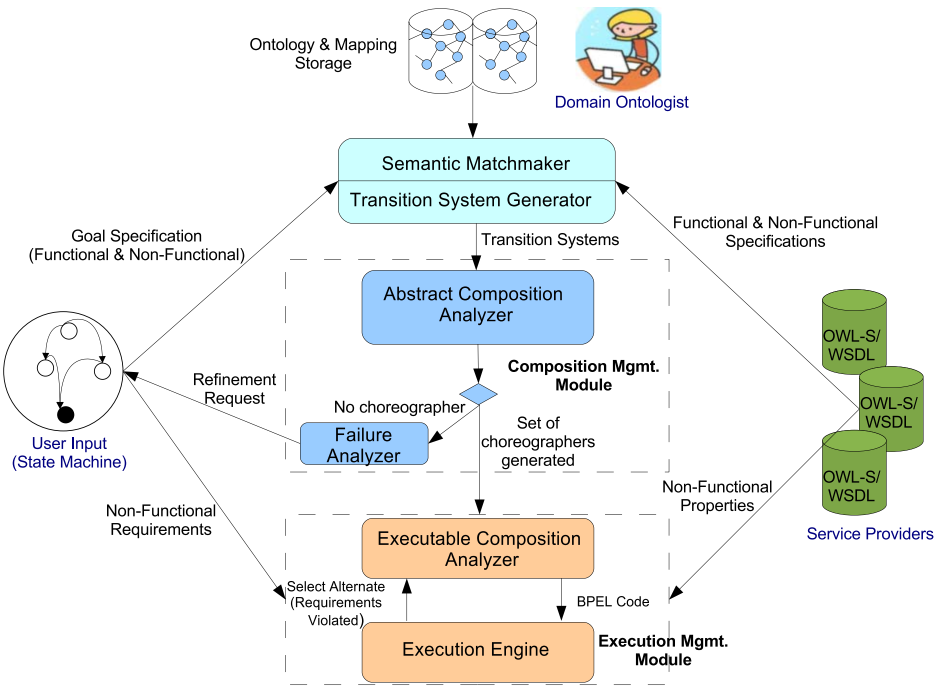


Figure 2. MoSCoE Architectural Diagram

updated. In addition, the STSs are semantically annotated using appropriate domain ontologies from a repository by importing OWL ontologies into the state machine model [32]. MoSCoE assumes that these ontologies (and mappings between them) are specified by a domain expert using existing tools such as INDUS [33]. The user also provides non-functional requirements (e.g., `cost,reliability`) that need to be satisfied by the goal service.

MoSCoE manipulates these input data (user-provided service specification and published component service descriptions) and automatically identifies a composition that realizes the goal service. However, in the event that a composition cannot be realized, the system identifies the cause(s) for the failure of composition and provides that information to the developer for appropriate reformulation of the goal specification. The system architecture comprises of two main modules: *composition management module* and *execution management module*. The former identifies feasible compositions (if any) that realize the goal, while the latter deals with the execution of the composite service. We describe these modules in the following paragraphs.

Composition Management Module. Given the STS representations of a set of N component services $\{\mathtt{W}_{STS_1}, \mathtt{W}_{STS_2}, \ldots, \mathtt{W}_{STS_N}\}$ and a desired goal $\mathtt{W}_{STS_G}$, service composition in MoSCoE amounts to identifying a subset of component services, which when composed with a mediator (to be generated) $\mathtt{W}_{STS_M}$, realize the goal service $\mathtt{W}_{STS_G}$. The role of the mediator is to replicate input/output actions of the user as specified by the goal and to act as a message-passing interface between the components and between the component(s) and the client. It is not capable of providing any functionality (e.g., credit card processing) on its own; these are provided only by the component services. The algorithms for generating such a mediator are discussed in [27,28] and the techniques essentially identify whether $\mathtt{W}_{STS_M}$ realizes $\mathtt{W}_{STS_G}$ using the notion of *simulation* and *bisimulation* equivalence. Informally, simulation equivalence ensures that every behavioral pattern in the goal is present in the composed mediator, whereas bisimulation equivalence is a symmetric relation which ensures that the composition offers exactly the same behavior as specified in the goal, and nothing more.

However, the algorithms proposed in [27,28] suffer from the state-space explosion problem since the number of ways the component services can be composed is exponential to the number of component service states. This becomes a challenge with the increasing size of the search space of available component services. Hence, to address this limitation, we consider non-functional aspects (e.g., Quality of Service) to winnow components (thereby reducing the search space) and compositions that are functionally equivalent to the goal, but violate the non-functional requirements desired by the user [29,34]. The non-functional requirements are quantified using *thresholds*, where a composition is said to conform to a non-functional requirement if it is below or above the corresponding threshold, as the case may be. For example, for a non-functional requirement involving the `cost` of a service composition, the threshold may provide an *upper-bound* (maximum allowable `cost`) while for requirements involving `reliability`, the threshold usually describes a *lower-bound* (minimum tolerable `reliability`). If more than one "feasible composition" meets the goal specification (both functional and non-functional requirements), our algorithm generates all such compositions and ranks them. It is then left to the user's discretion to select the best composition according to the requirements.

In the event that a composition as outlined above cannot be realized using the available component services, the composition management module provides feedback to the user regarding the cause(s) of the failure [27,28,29]. The feedback may contain information about the function names and/or pre-/post-conditions required by the desired service that are not supplied by any of the component services. Such information can help to identify specific states in the state machine description of the goal service. In essence, the module identifies all un-matched transitions along with the corresponding goal STS states. Additionally, the failure of composition could be also due to non-compliance of non-functional requirements specified by the user. When such a situation arises, the system identifies those requirements that cannot be satisfied using the available components, and provides this information to the service developer for appropriate reformulation of the goal specification. This process can be iterated until a realizable composition is obtained or the developer decides to abort.

Execution Management Module. The result from the composition management module is a set of feasible compositions each defining a mediator that will enable interaction between the client and the component services. The execution management module considers non-functional requirements (e.g., `performance, cost`) of the goal (provided by the user) and analyzes each feasible composition. It selects a composition that meets all the non-functional requirements of the goal, generates executable BPEL code, and invokes the MoSCoE execution engine. This engine is also responsible for monitoring the execution and recording violation of any requirement of the goal service at runtime. In the event a violation occurs, the engine tries to select an alternate feasible composition.

Furthermore, during execution, the engine leverages a pre-defined set of inter-ontology mappings to carry out various data and control flow transformations [35].

4. Open Research Directions

In the following, we outline several research challenges that we believe have to be addressed by the research community to make existing solutions for automatic Web service composition (including MoSCoE) better and useful in practice. These challenges include:

Composition Efficiency: The practical feasibility of approaches to automated service composition is ultimately limited by the computational complexity of the service composition algorithms. However, the existing composition techniques run into exponential complexities and become impractical in real-world situations comprising of hundreds, if not thousands, of services. Hence, intelligent approaches and heuristics for reducing the number of candidate compositions that need to be examined are urgently needed in order to scale up service composition techniques sufficiently to make them useful in practice.

Execution Models: Most of the existing implementations for composite Web service execution adopt a centralized architecture, that is, there exists an orchestrator (representing the composite service) in a centralized location responsible for coordinating and forwarding the intermediate results during the execution. Such a design has its limitation in terms of scalability, failure resiliency, and network bottlenecks. Towards this end, we believe that decentralized [36] or Peer-to-Peer (P2P) based architectures such as SELF-SERVE [26] will prove to be more beneficial in practical settings.

Failure Handling and Fault Tolerance: Web services are by nature autonomous and have an unpredictable behavior. For example, it is possible for a particular Web service W_i that is part of a composition to become inaccessible or updated (furnishing additional functions and/or removing existing ones, thereby altering its original behavior). Consequently, an existing integrated system (or a composite service) which comprises of multiple services including W_i, will require appropriate update in the form of replacing W_i. However, very limited research [37,38] has been carried out to address this issue which needs further investigation. The problem becomes even more non-trivial when the replacement of the faulty service has to be carried out, while the composite service is being executed, in such a way that it is transparent to the client.

Security: Addressing security concerns is important for any Web-based system and various researchers have proposed mechanisms for ensuring security in Web services (see [39] for a survey). However, most of techniques build a trust-based framework or assume the existence of an environment, where once a service is identified to be "good" (loosely speaking) based on its security policy etc., it is considered to be trustworthy. However, in certain cases, even though a particular service is trustworthy, it might delegate a part of its functionality to another service which cannot be trusted. For example, an online air ticket reservation service W_x might delegate the process of verifying authenticity of payment methods (e.g., credit card) required to purchase the air tickets to a third-party service provider W_y (in a manner transparent to the client), which may not follow the same security policy as W_x causing a potential security threat. Unfortunately, it is hard to detect such vulnerabilities. Furthermore, even if W_x claims to be "good", it may not strictly adhere to its own security policy, which makes it even harder to detect whether the integrity of client information has been compromised. We believe that addressing these two issues is a significant and important research challenge for the Web services community.

Tool Support: An important component of making techniques for automatic Web service composition useful for masses is to develop user-friendly tools and platforms that will allow non-experts to model complex services. Towards this end, model-driven based approaches [40] has shown some promise, although a lot of research has to be carried out, in particular by leveraging techniques from human-computer interaction and cognitive science.

Experimental Benchmark: At present, due to lack of a benchmark (dataset) of Web services, there is no uniform way of comparing, for example, an existing service composition algorithm with another. We believe that developing a comprehensive benchmark and testbed of Web services will act as a quick aid for testing and ease of prototyping to evaluate different techniques. Such a benchmark should comprise of various hardware platforms and a variety of synthetic and real-world Web services. To the best of our knowledge, WSBen [41] is one of the preliminary efforts in this direction.

5. Concluding Remarks

We have briefly surveyed some of the existing techniques for (semi-) automatic composition of Web services that are based on AI planning and formal methods and discussed some of their drawbacks. The paper demonstrates how our framework (called MoSCoE) attempts to address some of these limitations and envisions to provide a user-friendly technique for incrementally and iteratively developing complex Web services. Furthermore, we have outlined potential research issues that need to be addressed by blending techniques from artificial intelligence, software engineering, networks and distributed systems in order to develop solutions for Web service composition that are of practical significance and value.

Acknowledgments

This research is supported in part by the ISU Center for Computational Intelligence, Learning & Discovery (`http://www.cild.iastate.edu`), NSF-ITR grant 0219699 to Vasant Honavar and NSF grant 0509340 to Samik Basu.

References

[1] G. Alonso, F. Casati, H. Kuna, and V. Machiraju. *Web Services: Concepts, Architectures and Applications*. Springer-Verlag, 2004.

[2] T. Erl. *Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services*. Prentice Hall, New Jersey, 2004.

[3] D. Ferguson and M. Stockton. Service-Oriented Architecture: Programming Model and Product Architecture. *IBM Systems Journal* 44(4):753–780, 2005.

[4] D. Booth, H. Haas, F. McCabe, and et al. Web Services Architecture, W3C Working Group Note 11. *http://www.w3.org/TR/ws-arch/*, 2004.

[5] S. Dustdar and W. Schreiner. A Survey on Web Services Composition. *International Journal on Web and Grid Services* 1(1):1–30, 2005.

[6] R. Hull and J. Su. Tools for Composite Web Services: A Short Overview. *SIGMOD Record* 34(2):86–95, 2005.

[7] J. Rao and X. Su. A Survey of Automated Web Service Composition Methods. *1st Intl. Workshop on Semantic Web Services and Web Process Composition*, pp. 43–54, 2004.

[8] D. Berardi, D. Calvanese, D. G. Giuseppe, R. Hull, and M. Mecella. Automatic Composition of Transition-based Semantic Web Services with Messaging. *31st International Conference on Very Large Databases*, pp. 613–624, 2005.

[9] A. Gerevini and D. Long. Preferences and Soft Constraints in PDDL3. *ICAPS Workshop on Preferences and Soft Constraints in Planning*, 2006.

[10] D. Martin, M. Burstein, J. Hobbs, and et al. OWL-S: Semantic Markup for Web Services, Version 1.1. *http://www.daml.org/services/owl-s*, 2004.

[11] D. V. McDermott. Estimated-Regression Planning for Interactions with Web Services. *6th Intl. Conference on Artificial Intelligence Planning Systems*, pp. 204–211, 2002.

[12] B. Medjahed, A. Bouguettaya, and A. K. Elmagarmid. Composing Web services on the Semantic Web. *The Very Large Databases Journal* 12(4):333–351, 2003.

[13] K. Sycara, M. Paolucci, A. Ankolekar, and N. Srinivasan. Automated Discovery, Interaction and Composition of Semantic Web Services. *Journal of Web Semantics* 1(1):27–46, 2003.

[14] A. Ankolekar, M. Burstein, J. R. Hobbs, and et al. DAML-S: Semantic Markup for Web Services. *International Semantic Web Workshop*, 2001.

[15] E. Sirin, B. Parsia, D. Wu, J. Hendler, and D. Nau. HTN Planning for Web Service Composition using SHOP. *Journal of Web Semantics* 1(4):377–396, 2004.

[16] R. Akkiraju, B. Srivastava, A.-A. Ivan, R. Goodwin, and T. F. Syeda-Mahmood. SEMAPLAN: Combining Planning with Semantic Matching to Achieve Web Service Composition. *4th IEEE International Conference on Web Services*, pp. 37–44. IEEE CS Press, 2006.

[17] T. Gruber. Ontolingua: A Mechanism to Support Portable Ontologies. *Technical Report, KSL-91-66, Stanford University, Knowledge Systems Laboratory*, 1992.

[18] V. Agarwal, K. Dasgupta, and et al. A Service Creation Environment Based on End to End Composition of Web Services. *14th International Conference on World Wide Web*, pp. 128–137. ACM Press, 2005.

[19] M. Pistore, A. Marconi, P. Bertoli, and P. Traverso. Automated Composition of Web Services by Planning at the Knowledge Level. *19th Intl. Joint Conferences on Artificial Intelligence*, pp. 1252–1259, 2005.

[20] M. Pistore, P. Traverso, P. Bertoli, and A. Marconi. Automated Synthesis of Composite BPEL4WS Web Services. *3rd Intl. Conference on Web Services*, pp. 293–301. IEEE Press, 2005.

[21] R. Milner. *A Calculus of Communicating Systems*. Springer-Verlag New York, Inc., 1982.

[22] G. Salaün, L. Bordeaux, and M. Schaerf. Describing and Reasoning on Web Services using Process Algebra. *2nd IEEE International Conference on Web Services*, pp. 43–50. IEEE Computer Society, 2004.

[23] M. Hennessy. *Algebraic Theory of Processes*. MIT Press, 1988.

[24] T. Andrews, F. Curbera, and et al. Business Process Execution Language for Web Services, Version 1.1. *http://www.ibm.com/developerworks/library/ws-bpel/*, 2003.

[25] R. Hamadi and B. Benatallah. A Petri Net-based Model for Web Service Composition. *14th Australasian Database Conference*, pp. 191–200. Australian Computer Society, Inc., 2003.

[26] B. Benatallah, Q. Sheng, and M. Dumas. The Self-Serv Environment for Web Services Composition. *IEEE Internet Computing* 7(1):40–48, 2003.

[27] J. Pathak, S. Basu, R. Lutz, and V. Honavar. Selecting and Composing Web Services through Iterative Reformulation of Functional Specifi cations. *18th IEEE International Conference on Tools with Artificial Intelligence*, pp. 445–454. IEEE CS Press, 2006.

[28] J. Pathak, S. Basu, R. Lutz, and V. Honavar. Parallel Web Service Composition in MoSCoE: A Choreography-based Approach. *4th IEEE European Conference on Web Services*, pp. 3–12. IEEE CS Press, 2006.

[29] J. Pathak, S. Basu, and V. Honavar. Modeling Web Services by Iterative Reformulation of Functional and Non-Functional Requirements. *4th International Conference on Service Oriented Computing*, pp.

314–326. LNCS 4294, Springer-Verlag, 2006.

[30] M. Pistore, P. Traverso, and P. Bertoli. Automated Composition of Web Services by Planning in Asynchronous Domains. *15th Intl. Conference on Automated Planning and Scheduling*, pp. 2–11, 2005.

[31] P. Traverso and M. Pistore. Automated Composition of Semantic Web Services into Executable Processes. *3rd International Semantic Web Conference*, pp. 380–394. Springer-Verlag, 2004.

[32] D. Duric. MDA-based Ontology Infrastructure. *Computer Science and Information Systems* 1(1):91–116, 2004.

[33] D. Caragea, J. Pathak, J. Bao, A. Silvescu, C. Andorf, D. Dobbs, and V. Honavar. Information Integration and Knowledge Acquisition from Semantically Heterogeneous Biological Data Sources. *2nd International Workshop on Data Integration in Life Sciences*, pp. 175–190. Springer-Verlag, 2005.

[34] J. Pathak, N. Koul, D. Caragea, and V. Honavar. A Framework for Semantic Web Services Discovery. *7th ACM Intl. Workshop on Web Information and Data Management*, pp. 45–50. ACM press, 2005.

[35] J. Pathak, D. Caragea, and V. Honavar. Ontology-Extended Component-Based Workflows-A Framework for Constructing Complex Workflows from Semantically Heterogeneous Software Components. *2nd International Workshop on Semantic Web and Databases*, pp. 41–56. LNCS 3372, Springer-Verlag, 2004.

[36] W. Binder, I. Constantinescu, and B. Faltings. Decentralized Orchestration of Composite Web Services. *4th IEEE International Conference on Web Services*, pp. 869–876. IEEE Computer Society, 2006.

[37] B. Benatallah, F. Casati, and F. Toumani. Representing, Analysing and Managing Web Service Protocols. *Data and Knowledge Engineering* 58(3):327–357, 2006.

[38] L. Bordeaux, G. Salaün, D. Berardi, and M. Mecella. When are Two Web Services Compatible? *5th International Workshop on Technologies for E-Services*, pp. 15–28. LNCS 3324, Springer-Verlag, 2004.

[39] C. Gutiérrez, E. Fernández-Medina, and M. Piattini. A Survey of Web Services Security. *International Conference on Computational Science and Its Applications*, pp. 968–977. LNCS 3043, Springer-Verlag, 2004.

[40] K. Pfadenhauer, S. Dustdar, and B. Kittl. Challenges and Solutions for Model Driven Web Service Composition. *14th IEEE Intl. Workshop on Enabling Technologies: Infrastructures for Collaborative Enterprises*, pp. 126–131. IEEE Press, 2005.

[41] S.-C. Oh, H. Kil, D. Lee, and S. R. T. Kumara. WSBen: A Web Services Discovery and Composition Benchmark. *4th International Conference on Web Services*, pp. 239–246. IEEE Press, 2006.

This page intentionally left blank

Author Index

This page intentionally left blank

This page intentionally left blank