Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005

A HYBRID OF CONCEPTUAL CLUSTERS, ROUGH SETS AND ATTRIBUTE ORIENTED INDUCTION FOR INDUCING SYMBOLIC RULES

QING-SHUANG JIANG, SYED SIBTE RAZA ABIDI

Faculty of Computer Science, Dalhousie University, Halifax B3H 1W5, Canada
E-MAIL: sraza@cs.dal.ca

Abstract:

Rule induction is a data mining process for acquiring knowledge in terms of symbolic decision rules from a number of specific 'examples' to explain the inherent causal relationship between conditional factors and a given decision/outcome. We present a Decision Rule Acquisition Workbench (DRAW) that discovers conjunctive normal form decision rules from un-annotated data-sets. Our rule-induction strategy uses (i) conceptual clustering to cluster and generate a conceptual hierarchy of the data-set; (ii) rough sets based rule induction algorithm to generate decision rules from the emergent data clusters; and (iii) attribute oriented induction to generalize the derived decision rules to yield high-level decision rules and a minimal rule-set size.

Keywords:

Rule Induction; Rough Sets; Conceptual Clusters; Attribute Oriented Induction

1. Introduction

Rule induction is the process of acquiring knowledge (i.e. symbolic decision rules) from a number of specific 'examples' (i.e. the data-set), to explain the cause-and-effect relationship between conditional factors and a given decision/outcome. However, rule induction algorithms are supervised in nature and typically work on annotated data-sets, yet there is a case for interpreting un-annotated data-sets in terms of Conjunctive Normal Form (CNF) symbolic rules. This can be achieved by inductively clustering the data-sets and then explaining the relationships between the attributes, that manifest as data clusters, in terms of symbolic decision rules [1, 2, 3].

In this paper we present a rule induction framework that discovers CNF decision rules from un-annotated data-sets. We use rough sets as the base rule-induction method, which has been successfully applied for this task [4]. Additionally, we present techniques to (a) reduce the size of the rule-set; and (b) to generalize the rules to high-level concepts. We present a hybrid rule-induction strategy that uses (i) conceptual clustering to cluster the un-annotated data-set (to acquire the underlying class information) and to generate a conceptual hierarchy that describes the data at different levels of abstraction [5]; (ii) rough sets based rule induction algorithm to generate decision rules from the emergent data clusters [6]; and (iii) attribute oriented induction in conjunction with the data's conceptual hierarchy to generalize the derived decision rules to high-level decision rules [7], and in the process minimizing the rule-set size without compromising classification accuracy. The rule induction framework is implemented in terms of *DRAW* (Decision Rule Acquisition Workbench), as shown in figure 1, and is evaluated with standard machine learning datasets.

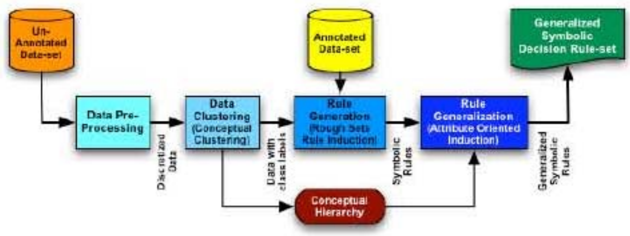


Figure 1. The Functional Architecture of DRAW

We applied DRAW to derive decision rules for classifying *Confocal Scanning Laser Tomography* (CSLT) images of the optic disk for the diagnosis of glaucoma. The decision rules are expected to provide: (a) insight into the relationships between optic disk's shape features in relation to glaucoma damage; and (b) a description of different classes (healthy, glaucoma, focal, senile, myopic damage) in terms of the features. The derived decision rules can be used for the classification of a new CSLT image—i.e. whether the patient is a normal or has a specific type of glaucoma.

2. Rule Induction with Rough Sets

Rule induction is an active machine learning research area with a wide variety of existing algorithms—to construct, classify and prune the rules—grounded in different theoretical principles [8, 9, 10]. Statistical methods, classified as classical and modern (such as Alloc 80, CASTLE, Naïve Bayes) methods, are based on linear discrimination and estimation of the joint distribution of the features within each class, respectively. Decision tree based rule induction methods, such as ID3, CART, C4.5 and CAL5, partitions the data points in terms of a tree structure such that the nodes are labeled as features and the edges as attribute values [11].

Rough sets provide an alternative method for rule induction due to their unique approach to handle inconsistencies in the input data. Rough set based data analysis involves the formulation of approximate concepts about the data based on available classification information [4, 12]. Functionally, rough sets based rule induction involves the formation of approximate concepts based on a minimum number of features—the approximate concepts, also called *dynamic reducts*, are capable of differentiating the various classes within the data. Rule induction involves the representation of dynamic reducts as succinct symbolic *if then* decision rules that potentially can inter-attribute dependencies and attribute significance and class information. There is no standard approach to rough set based rule induction, hence rough sets can be used at different stages a rule induction process [13, 14]. A number of rough set based rule induction approaches has been suggested such as *System LERS* [13, 15]; *Discernibility matrix* 16] leading to a computational kernel of the system *Rosetta* [17]; *RoughDAS* [18, 19] that allows users to specify conditions; rule induction framework, involving rule filtering methods, to generate cluster-defining rules [2, 3]; and so on..

3. A Hybrid Rule Induction Strategy

The featured work is an extension of the rule induction framework proposed by Abidi [3], in that we have introduced methods to generate generalized decision rules and to minimize the rule-set size. We have postulated a hybrid rule induction approach that dictates the systematic transformation of un-annotated data-sets to deductive symbolic rule-sets via a sequence of phases, as described below:

3.1. Phase 1 - Data Clustering and Conceptual Hierarchy Generation

Given an un-annotated dataset, the first step involves the unsupervised clustering of the data into k clusters in order to derive classification information that is required for subsequent rule induction. We have adapted *a Similarity Based Agglomerative Clustering* (SBAC) algorithm that is based on a similarity measure from a biological taxonomy [5]. The clustering process is driven by two factors 1) the criterion function for evaluating the goodness of partition structure and 2) the control algorithm for generating the partition structure. The similarity measure used allows us to process both numeric and nominal valued attributes within a common framework, and can be conveniently coupled with a simple agglomerative control strategy that constructs a conceptual hierarchy.

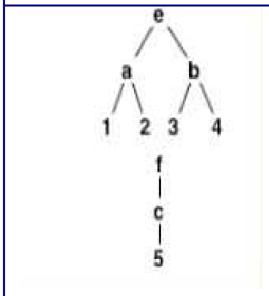
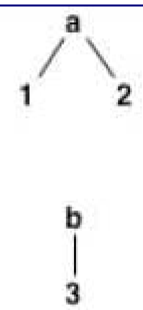
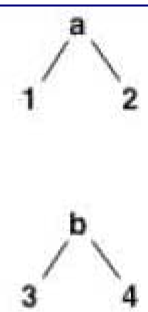
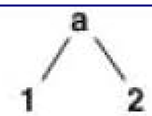
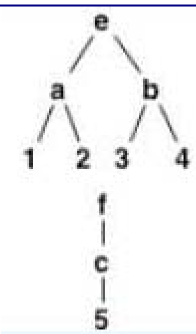
To derive the conceptual hierarchy, the original algorithm [5] is modified as follows: (a) Each attribute of the data-set is applied the SBAC individually such that the similarity matrix is calculated based on the similarity measure among the values of each attribute instead of the combination of all the attributes; and (b) based on the similarity matrix of each attribute, the distinct values of an attribute are clustered agglomeratively such that a tree-like structure is formed, where the lowest level is made up of the distinct attribute values and the higher level is made up of clustered values which contain the values from lower level. The root of the conceptual hierarchy recursively contains all the distinct values of the attribute, which provides the most general attribute information (as shown in Table 1).

The outcome of phase I is two fold: (i) the data is partitioned into k number of data clusters; and (ii) a hierarchical concept tree is constructed to form the basis for attribute-oriented induction and the generalization of the cluster-defining symbolic rules.

3.2. Phase 2 - Symbolic Rule Discovery

Given an annotated and discretized data-set together with the conceptual hierarchy that describes the characteristics of the data-set, the task in phase II is to generate a set of symbolic CNF rules that model the data-set in terms of class-membership principles and complex inter-relationships between the data attributes. In our work, we implemented a symbolic rule generation approach based on the Rough Set approximation [4, 12], as it provides a sound and interesting alternative to statistical and decision tree based rule induction methods

Table 1. Conceptual hierarchy derived from the TGD data set. Attribute values are discretized into intervals. Intervals labeled by letters are at the higher level. Intervals labeled by numbers are on the lower level of the hierarchy and can be merged to form a more generalozed interval.

T3r (10)	Sthy (5)	Stri (6)	bTSH (3)	mTSH (10)
1 = [_ , 97) 2 = [97, 100) 3 = [100, 118) 4 = [118, 125) 5 = [125, _) a = [_ , 100) b = [100, 125) c = [125, _) e = [_ , 125) f = [125, _)	1 = [_ , 5.7) 2 = [5.7, 14.2) 3 = [14.2, _) a = [_ , 14.2) b = [14.2, _)	1 = [_ , 1.25) 2 = [1.25, 3.65) 3 = [3.65, 3.95) 4 = [3.95, _) a = [_ , 3.65) b = [3.65, _)	1 = [_ , 4.30) 2 = [4.30, _) a = [_ , _)	1 = [_ , 0.35) 2 = [0.35, 0.65) 3 = [0.65, 7.25) 4 = [7.25, 10.5) **5 = [10.5, _)** a = [_ , 0.65) b = [0.65, 10.5) **C = [10.5, _)** e = [_ , 10.5) f = [10.5, _]
				

Our rule induction methodology is as follows:

Step 1: Dynamic Reducts Computation: We use *k-fold cross validation* to split the data-set into training and test sets. From the training data (for each fold), we compute multiple dynamic reduct sets, such that each reduct set is found through the identification of minimum attribute sets that are able to distinguish a data point from the rest of the data set. This is achieved via (a) *vertical* reduction whereby the redundant data objects are eliminated and (b) *horizontal* reduction whereby the redundant attributes are eliminated since logically duplicated attributes cannot help distinguish a data point from the rest of the data set. At the completion of the reduction process we end up with a minimum set of attribute values that can distinguish any data point from rest of the data-set—i.e. a reduct set. Dynamic reducts are the reducts that have a high frequency of occurrence across all the available reduct-sets, and are of particular interest for rule generation purposes. The search for dynamic reducts is an NP-hard problem—the time complexity for finding the minimum attribute set increases exponentially with respect to the linear increase of the data set—and a genetic algorithm based reducts approximation method is used to compute the dynamic reducts [20, 21].

Step II: Symbolic Rule Generation via Dynamic Reducts: After computing the dynamic reducts we generate symbolic rules from them. Instead of using all the dynamic reducts to generate a large set of symbolic rules, we attempt to generate symbolic rules from the shortest possible length dynamic reducts; the rationale being that shorter length dynamic reducts have been shown to yield concise rule-sets that exhibit higher classification accuracy and generalization capabilities [2, 3]. Our rule generation strategy therefore involves: (1) the selection of dynamic reducts that have a short length and (2) the generation of rules that satisfy a user-defined accuracy level. Our strategy for generating symbolic rules is as follows [3]:

Step 1 : Specify an acceptable minimum accuracy level for the rule set.

Step 2 : Find dynamic reducts from the sub-samples and place in set *DR*. Note that *DR* will comprise reducts with varying lengths.

Step 3 : From the reducts in *DR* determine the shortest reduct length (*SRL*).

Step 4 : From *DR*, collect all reducts that have a length equal to *SRL* and store them as set *SHRED*.

Step 5 : Generate symbolic rules from the reducts placed in *SHRED*.

Step 6 : Determine the overall accuracy of the generated rules with respect to the test data.

Step 7 : IF Overall accuracy of the generated rules is lower than the minimum accuracy level AND there exist reducts in the *DR* set with length > *SRL*

THEN Empty *SHRED* AND Update the value of *SRL* to the next highest reduct length in *DR* AND Repeat from step 6.
ELSE Symbolic rules with the desired accuracy level cannot be generated.

3.3. Phase 3 – Rule-Set Generalization

Typically, The rule-set generated in phase II is quite large and might contain low quality rules, thereby compromising the classification efficiency of the classifier. In phase III, we attempt to minimize the rule-set size by generalizing the induced rules—i.e. a large number of low level concepts represented in the rules can be generalized to fewer higher-level concepts. For rule generalization we have adapted a set-oriented induction method—called Attribute-Oriented Induction (AOI) [7]—that employs the concept hierarchy generated in phase I.

Input: (i) A CNF rule set, (ii) the learning task, (iii) the (optional) preferred concept hierarchies, and (iv) the (optional) preferred form to express learning results (e.g., generalization *threshold*).

Output. A characteristic rule set generalized from the input rule set.

Method. Basic attribute-oriented induction consists of the following four steps:

Step 1. Collect the task-relevant data,
Step 2. Perform basic attribute-oriented induction
  begin {basic attribute-oriented induction}
    for each attribute Ai (1 ≤ i ≤ n, n = # of attributes) in the generalized relation GR *do*
    while #_of_distinct_values_in_Ai > threshold *do* {
      if no higher level concept in the concept hierarchy table for Ai
      then remove Ai
      else substitute Ai 's by its corresponding minimal generalized concept;
      merge identical rules }
        while #_of_rules in GR > threshold *do* {
          selectively generalize attributes; merge identical rules }
    end.
Step 3. Simplify the generalized relation, and
Step 4. Transform the final relation into a logical rule.

4. Experimental Results

The experiments reported evaluate the final classification accuracy of rules discovered by DRAW, and also demonstrate the performance of the individual modules. Three experimental scenarios were performed, with standard data-sets, to demonstrate the performance of the various combinations of modules (as shown in Figure 1):

a) Rough set based rule induction (with class information) --> Decision rules
b) Conceptual Clustering (without class information) + Rough set based rule induction --> Decision rules
c) Conceptual Clustering (without class information) + Rough set based rule induction + Attribute-Oriented Induction --> Generalized decision rules

4.1 Classification Accuracy for Standard Data-Sets

Standard machine learning data-set—i.e. Thyroid gland data (TGD); Wisconsin Breast Cancer (WBC); Balance Scale Weight & Distance (BaS); Iris Plants Database (IRIS); Pima Indians Diabetes Database (Pima) were used for our evaluation. Table 2 shows the classification accuracy for the above experimental scenarios. Also, we compared the overall accuracy of our rules with rules derived from C4.5.

From the classification results the following was observed: (a) The overall classification accuracy offered by DRAW, for both annotated and un-annotated data-sets, is quite high and is comparable (in fact better in three cases) with the C4.5 method. The classification accuracy for scenario A is largely maintained throughout the subsequent stages of the process indicating the robustness of the rough set based rule-induction method. We conclude that our rough set based rule induction method amicably derives the underlying class structure from the data; (b) The classification accuracy for un-annotated data using SBAC conceptual clustering (scenario B) is comparable to both scenario A and C4.5, indicating the effectiveness of the conceptual clustering approach for rule induction for un-annotated data; (c) The classification accuracy for generalized rules (scenario C) is comparable to both scenario A and C4.5, indicating the effectiveness of the AOI method for rule generalization. Although no significant gain in the accuracy is noted for scenario C, yet the real impact of the AOI approach is noted in the minimization of the rule-set size without a discernable loss of classification accuracy; (d) Comparison of scenarios B & C indicate that the conceptual hierarchy derived in phase I is effective for AOI based rule generalization, and the generalized rules do not compromise the classification accuracy. This vindicates the role of the conceptual hierarchy and the AOI method for rule generalization.

4.1.2 Rule-Set Generalization

The application of the AOI based rule generalization method has reduced the rule-set size without compromising the classification (see table 3). Furthermore, different degrees of rule generalization takes place (as shown in table 4):

Case 1 shows single-level generalization, where a single attribute mTSH is generalized at level 1 of the conceptual hierarchy as [1,2]-> a;

Case 2, 3, 4 show multi-level generalization, where attribute mTSH is generalized at level 1 of the conceptual hierarchy as [1,2]-> a and [3,4]->b, the generalization goes on at the level 2 as [a,b] -> e, and continues until the root of the conceptual hierarchy tree is reached;

Case 5 shows that generalization can be performed on more than 2 rules, the number of the rules can be generalized is determined by the number of children that belong to the same parent.

Table 2: Experimental results for scenarios A-C and the baseline C4.5 method

Data Sets	Accuracy				Sensitivity				Specificity			
	A	B	C	C4.5	A	B	C	C4.5	A	B	C	C4.5
TGD	0.887	0.892	0.836	0.931	0.968	0.927	0.875	0.973	0.916	0.962	0.955	0.973
WBC	0.930	0.904	0.932	0.925	0.972	0.915	0.963	0.982	0.957	0.988	0.968	0.982
BaS	0.620	0.573	0.645	0.643	0.759	0.749	0.728	0.794	0.817	0.765	0.886	0.794
Pima	0.872	0.889	0.849	0.923	0.941	0.955	0.859	0.963	0.927	0.931	0.988	0.963
IRIS	0.586	0.589	0.631	0.619	0.727	0.744	0.711	0.739	0.806	0.792	0.882	0.739

Table3: Rule-set size comparison for the five different data-sets.

	Clustered Rough-Set Based Rules								
Data Sets	Before AOI				After AOI				Percentage Reduction
	Acc	Sens	Spec	# of Rules	Acc	Sens	Spec	# of Rules	
TGD	**0.892**	0.927	0.962	**67**	**0.836**	0.875	0.955	**52**	**22.39%**
WBC	**0.904**	0.915	0.988	**103**	**0.932**	0.963	0.968	**72**	**30.10%**
BaS	**0.573**	0.749	0.765	**217**	**0.645**	0.728	0.886	**194**	**10.60%**
Pima	**0.889**	0.955	0.931	**455**	**0.849**	0.859	0.988	**357**	**21.52%**
IRIS	**0.589**	0.744	0.792	**16**	**0.631**	0.711	0.882	**15**	**6.25%**

Table 4: Different degrees of rule generalization achieved via AOI

Case	Decision Rule Before AOI	Decision Rule After AOI	Explanation
1	sthy(3) mTSH(1) -> 2 sthy(3) mTSH(2) -> 2	sthy(3) mTSH(a) -> 2	mTSH(1) and mTSH(2) are generalized to mTSH(a)
2	t3r(3) mTSH(1) -> 1 t3r(3) mTSH(2) -> 1	t3r(3) mTSH(a) -> 1	mTSH(1) and mTSH(2) are generalized to mTSH(a)
3	t3r(3) mTSH(3) -> 1 t3r(3) mTSH(4) -> 1	t3r(3) mTSH(b) -> 1	mTSH(3) and mTSH(4) are generalized to mTSH(b)
4	t3r(3) mTSH(a) -> 1 t3r(3) mTSH(b) -> 1	t3r(3) mTSH(e) -> 1	mTSH(a) and mTSH(b) are generalized to mTSH(e)
5	ucz(2) bn(1) ucp(4) -> 3 ucz(2) bn(2) ucp(4) -> 3 ucz(2) bn(4) ucp(4) -> 3	ucz(2) bn(a) ucp(4) -> 3	bn(1), bn(2)and bn(4) are generalized to bn(a)

5. Concluding Remarks

We have presented an interesting and efficient rule induction strategy that ensures the generation of complex and high-level decision rules for un-annotated data-sets. Rule granularity has been regarded as being an important factor in the comprehensibility of the discovered knowledge in terms of rules—longer and more specific rules do not necessarily provide better classification accuracy as compared to shorter and generalized rules. Hence, the need for more concise rules and smaller rule-sets. The rule induction strategy presented here allows for the generalization of rules whilst maintaining classification accuracy via the incorporation of attribute oriented induction—which integrates machine learning methodology with relational database operations—with rough sets based rule induction methods. The use of a conceptual hierarchy to describe the data is an interesting idea as it allows for viewing the data at different levels of abstraction, and enables the users to derive rules at a desired level of abstraction. We also will like to point out that the use of the base rough-sets method for rule induction does not impose any static statistical parameters or models upon the data, hence minimizing assumptions and allowing the data to represent itself.

References

[1] Fayyad, U.M., Piatetsky-Shapiro, G. & Smyth, P.: Knowledge Discovery and Data Mining: Towards a Unifying Framework. International Conference on Knowledge Discovery in Databases (1996), pp. 82-88.

[2] Abidi, S.S.R., Hoe, K. M., Goh, A.: Healthcare Simulation Model Specification Featuring Multi-Stage Neural Network Rule Extraction. Proc. 4th Int. Eurosim Congress, Netherlands (2001),

[3] Abidi, S.S.R., Hoe, K. M., Goh, A.: Analyzing Data Clusters: A Rough Set Approach to Extract Cluster Defining Symbolic Rules. In Hoffman F, Hand DJ, Adams N, Fisher D & Guimaraes G (Eds.) Lecture Notes in Computer Science 2189: Advances in Intelligent Data Analysis, 4th Int. Conf. of Intelligent Data Analysis, Cascais, Springer Verlag: Berlin, 2001. pp. 248-257.

[4] Pawlak, Z.: Rough Sets. In: Lin T.Y., Cercone N. (eds.): Rough Sets and Data Mining: Analysis of Imprecise Data. Kluwer Academic Publishers, Dordrecht (1997) pp. 3-7.

[5] Biswas, G., Weinberg, J., Li, C.: ITERATE: A Conceptual Clustering Method for Knowledge Discovery in Database. In Artificial Intelligence in the Petroleum Industry, B. Braunshweig and R. Day, Editors. Editions Techniq, chapter 3, pp. 111-139 (1995).

[6] Slowinski K., Slowinski R., Stefanowski J.: Rough sets approach to analysis of data from peritoneal lavage in acute pancreatitis. Medical Informatics, vol. 13, pp. 143-159 (1988).

[7] Han, J., Cai, Y. and Cercone, N.: Knowledge discovery in databases: an attribute-oriented approach. In Proceedings of the 18th VLDB Conference, pp. 547-559, Vancouver, British Columbia, Canada (1992).

[8] Michalski R. S.: A theory and methodology of inductive learning. In R.S. Michalski, J.G. Carbonell and T.M. Mitchell, Machine Learning: An Artificial Intelligence Approach, Morgan Kaufman, pp. 83-134 (1983)

[9] Shavlik J. W., Dietterich T.G.: Readings in Machine learning, Morgan Kaufman (1990).

[10] Weiss S. M., Kulikowski C. A.:, Computer systems That Learn: Classification and Predication Methods from statistics, Neural Nets, Machine Learning and export systems, Morgan Kaufman, (1991).

[11] Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, California (1993).

[12] Hu, X., Cercone, N.: Learning Maximal Generalized Decision Rules via Discretization, Generalization and Rough Set Feature Selection. Proc. 9th Int. Conf. on Tools with Artificial Intelligence (1997).

[13] Grzymala Busse J. W.: LERS-a system for learning from examples based on rough sets. In R. Slowinski, (ed.) Intelligent Decision Support, Kluwer Publishers, (1992).

[14] GrzymalaBusse J. W., Stefanowski J., Ziarko W.: Rough Sets: Facts versus Misconceptions. Informatica, 20, (1996).

[15] Chan C.C., GrzymalaBusse J.W.: On the two local inductive algorithms: PRISM and LEM2. Foundations of Computing and Decision Sciences, vol. 19 (4), (1994).

[16] Skowron A., Polkowski L.: Synthesis of decision systems from data tables. In Lin T.Y., Cecrone N.(ed.), Rough Sets and Data Mining, Kluwer Publishers, (1997).

[17] Ohrn A., Komorowski J.: Diagnosing Acute Appendicitis with Very Simple Classification Rules. 3rd European Symposium on Principles and Practice of Knowledge Discovery in Databases, Prague, Czech Republic, LNAI 1704, Springer-Verlag, pp. 462-467.

[18] Mienko R., Slowinski R., Stefanowski J., Susmaga R.: RoughFamily software implementation of rough set based data analysis and rule discovery techniques. In Tsumoto S. (ed.) Proceedings of the Fourth International Workshop on Rough Sets, Fuzzy Sets and Machine Discovery, Tokyo Nov. 6-8, pp. 437-440 (1996).

[19] Slowinski R. (ed.).: Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publishers (1996).

[20] Bazan J. G.: Dynamic reducts and statistical inference. Proc. 6th Int. Conf. on Information Procesing and Management of Uncertainty in Knowledge-Based Systems. Vol. 3, Granada, (1996) pp. 1147-1152.

[21] Wróblewski, J.: Finding Minimal Reducts using Genetic Algorithms. Proc. 2nd Ann Joint Conf. on Information Sciences, Wrightsville Beach, NC. (1995) pp.186-189.