

修士論文

Automatic Detection of Repressed Anger from Text Messages

20〇〇年〇月〇〇日

(Replace with publication date)

学籍番号 16906291

ヘスス・マリア・セスマ・ソランセ

指導教員 湯川 高志

長岡技術科学大学大学院工学研究科

経営情報システム工学専攻

平成 年 月 日

専攻名	経営情報システム工学	学籍番号	16906291
申請者氏名	ヘスス・マリア・セスマ・ソランセ		
指導教員氏名	湯川高志 教授		

審査委員主査 〇〇 〇〇 教授

審査委員 ○○ ○○ 准教授

審査委員 ○○ ○○ 准教授

審查委員

審査委員

専攻主任印

論 文 要 旨

論文題目	Automatic Detection of Repressed Anger from Text Messages
------	---

[illegible]

修士論文

Automatic Detection of Repressed Anger from Text Messages

20〇〇年〇月〇〇日

(Replace with publication date)

学籍番号 16906291

ヘスス・マリア・セスマ・ソランセ

指導教員 湯川 高志

長岡技術科学大学大学院工学研究科

経営情報システム工学専攻

Contents

1	Classification Techniques	1
1.1	Fundamentals of Classification	1
1.2	General classification problem solving	3
1.3	Support Vector Machine	4
1.4	K-Nearest Neighbor	6
1.5	Neural Networks	6
	References	7
	Acronyms	8

List of Figures

1.1	Classification of animals. The image is extracted from Exploring Nature.	1
1.2	Classification as a task of mapping a set attributes x into its fitting class label y	2
1.3	General approach for classification model building and new instance category prediction.	3
1.4	hyperplane	6

List of Tables

1.1	Animal kingdom dataset.	2
-----	---------------------------------	---

Chapter 1

Classification Techniques

The aim of this chapter has two purposes. The first one, is to make an introduction of the basic concepts of classification, which is essential for the detection of repressed anger. The second one, is to explain how the algorithms used in this study work.

1.1 Fundamentals of Classification

According to [10], classification can be defined as the task of predicting an outcome from a given input. This outcome is produced by the process of mapping a group of characteristics present in the input to a certain category. In other words, it consists in assigning objects (the input) to one of several predefined classes (the outcome) [9]. Examples of classification can be found in everyday life, such as e-mail spam detection, news classifiers, Optical Character Recognition (OCR), animal kingdom classification (see Figure 1.1), among many others.

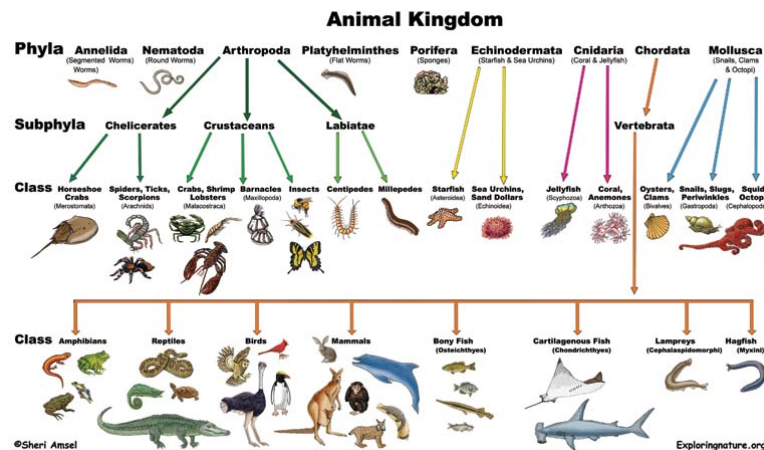


Figure 1.1: Classification of animals. The image is extracted from Exploring Nature.

The input data for a classification task is composed by a collection of records, the dataset. In the same time, each record, also known as an instance, is composed by a set attributes. From all these attributes there is one considered special, which

is called the target attribute or the class label. Regular attributes can be both discrete or continuous values. For the values signed for the class label, however, they must be discrete. This characteristic is what distinguishes classification from regression. Table 1.1 shows a sample dataset for animal classification into the following categories: amphibian, bird, fish or mammal.

Common Name	Hair	Feathers	Eggs	Milk	Aquatic	Legs	Class Label
antelope	Yes	No	No	No	No	4	mammal
catfish	No	No	Yes	No	Yes	0	fish
dolphin	No	No	No	Yes	Yes	0	mammal
dove	No	Yes	Yes	No	No	2	bird
duck	No	Yes	Yes	No	Yes	2	bird
elephant	Yes	Yes	No	Yes	No	4	mammal
flamingo	Yes	Yes	Yes	No	No	2	bird
frog	No	No	Yes	No	Yes	4	amphibian
fruit bat	Yes	No	No	Yes	No	2	mammal
gull	No	Yes	Yes	No	Yes	2	bird
herring	No	No	Yes	No	Yes	0	fish
kiwi	No	No	Yes	No	No	2	bird
lark	No	Yes	Yes	No	No	2	bird
lynx	Yes	No	No	Yes	No	4	mammal
mole	Yes	No	No	Yes	No	4	mammal
mongoose	Yes	No	No	Yes	No	4	mammal
newt	No	No	Yes	No	Yes	4	amphibian

Table 1.1: Animal kingdom dataset.

A more mathematical definition of classification states that classification is the process of learning a target function f , also known as classification model, that maps each attribute set x to one of the predefined class labels y .

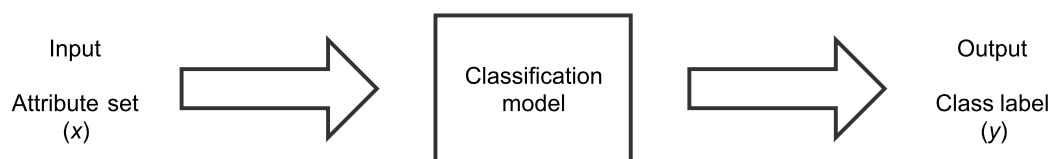


Figure 1.2: Classification as a task of mapping a set attributes x into its fitting class label y .

A classification model is useful for the following purposes:

- **Descriptive Modeling:** Since a classification model presents the main features of the data, it can serve as an explanatory tool to distinguish between instances of different categories [8].

- **Predictive Modeling:** A classification model can also be used to predict the class label of an unknown new instance. As shown in Figure 1.2, a classification model can be represented as a black box that automatically assigns a class label to an instance by providing its attribute set.

It is important to remark that classification techniques perform their best when used for predicting or describing datasets which its class label is binary or nominal. Since they do not consider properties such as ordinality or the implicit order among the categories, they become ineffective with ordinal class labels [4].

1.2 General classification problem solving

For general classification problems solving, popular techniques consist of a process that starts with building classification models from a sample dataset. Each technique depends on a learning algorithm which is in charge of generating the classification model. A good model should define the relationship between the input attribute set and its belonging category that suits the best. Therefore the model should be valid for both, the sample data used to generate the model and also for new unknown instances. Among popular classification techniques Support Vector Machine (SVM), Neural Networks (NNs), Naive Bayes or Decision Trees (DTs) can be found [5].

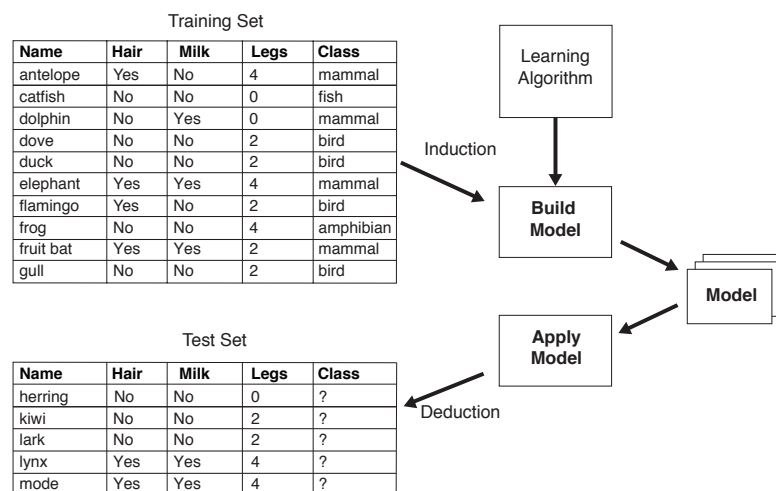


Figure 1.3: General approach for classification model building and new instance category prediction.

[STARTS COPY PASTE]

Figure 1.3 shows a general approach for solving classification problems. First, a training set consisting of records whose class labels are known must be provided. The training set is used to build a classification model, which is subsequently applied to the test set, which consists of records with unknown class labels. Evaluation of the performance of a classification model is based on the counts of test records correctly and incorrectly predicted by the model. These counts are tabulated in a table known as a confusion matrix. Table 4.2 depicts the confusion matrix for a binary

classification problem. Each entry f_{ij} in this table denotes the number of records from class i predicted to be of class j . For instance, f_{01} is the number of records from class 0 incorrectly predicted as class 1. Based on the entries in the confusion matrix, the total number of correct predictions made by the model is $(f_{11} + f_{00})$ and the total number of incorrect predictions is $(f_{10} + f_{01})$. Although a confusion matrix provides the information needed to determine how well a classification model performs, summarizing this information with a single number would make it more convenient to compare the performance of different models. This can be done using a performance metric such as accuracy, which is defined as follows:

Accuracy = Number of correct predictions = $f_{11} + f_{00}$. (4.1) Total number of predictions $f_{11} + f_{10} + f_{01} + f_{00}$

Equivalently, the performance of a model can be expressed in terms of its error rate, which is given by the following equation:

Error rate = Number of wrong predictions = $f_{10} + f_{01}$. (4.2) Total number of predictions $f_{11} + f_{10} + f_{01} + f_{00}$

Most classification algorithms seek models that attain the highest accuracy, or equivalently, the lowest error rate when applied to the test set. We will revisit the topic of model evaluation in Section 4.5.

1.3 Support Vector Machine

There are four main advantages: Firstly it has a regularisation parameter, which makes the user think about avoiding over-fitting. Secondly it uses the kernel trick, so you can build in expert knowledge about the problem via engineering the kernel. Thirdly an SVM is defined by a convex optimisation problem (no local minima) for which there are efficient methods (e.g. SMO). Lastly, it is an approximation to a bound on the test error rate, and there is a substantial body of theory behind it which suggests it should be a good idea. The disadvantages are that the theory only really covers the determination of the parameters for a given value of the regularisation and kernel parameters and choice of kernel. In a way the SVM moves the problem of over-fitting from optimising the parameters to model selection. Sadly kernel models can be quite sensitive to over-fitting the model selection criterion [3]

SVMs are a new promising non-linear, non-parametric classification technique, which already showed good results in the medical diagnostics, optical character recognition, electric load forecasting and other fields.

Suitable for binary classification tasks.

The advantages of the SVM technique can be summarised as follows [1]:

1. By introducing the kernel, SVMs gain flexibility in the choice of the form of the threshold separating solvent from insolvent companies, which needs not be linear and even needs not have the same functional form for all data, since its function is non-parametric and operates locally. As a consequence they can work with financial ratios, which show a non-monotone relation to the score and to the probability of default, or which are non-linearly dependent, and this without needing any specific work on each non-monotone variable.

2. Since the kernel implicitly contains a non-linear transformation, no assumptions about the functional form of the transformation, which makes data linearly separable, is necessary. The transformation occurs implicitly on a robust theoretical basis and human expertise judgement beforehand is not needed.
3. SVMs provide a good out-of-sample generalization, if the parameters C and r (in the case of a Gaussian kernel) are appropriately chosen. This means that, by choosing an appropriate generalization grade, SVMs can be robust, even when the training sample has some bias
4. SVMs deliver a unique solution, since the optimality problem is convex. This is an advantage compared to Neural Networks, which have multiple solutions associated with local minima and for this reason may not be robust over different samples.
5. With the choice of an appropriate kernel, such as the Gaussian kernel, one can put more stress on the similarity between companies, because the more similar the financial structure of two companies is, the higher is the value of the kernel. Thus when classifying a new company, the values of its financial ratios are compared with the ones of the support vectors of the training sample which are more similar to this new company. This company is then classified according to with which group it has the greatest similarity.

Furthermore, $K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ is called the kernel function. four basic kernels [7]:

- **Linear:** $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$.
- **Polynomial:** $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \gamma > 0$.
- **Radial Basis Function (RBF):** $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0$.
- **Sigmoid:** $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$.

Here, γ , r , and d are kernel parameters.

Support vector machines (SVM) were originally designed for binary classification [6].

solving multi-class SVM in one step: “all-together” methods: [25], [27] and [7]. We then compare their performance with three methods based on binary classifications: “one-against-all,” “one-against-one,” and DAGSVM [23]. Our experiments indicate that the “one-against-one” and DAG methods are more suitable for practical use than the other methods.

asdasdasd [2]

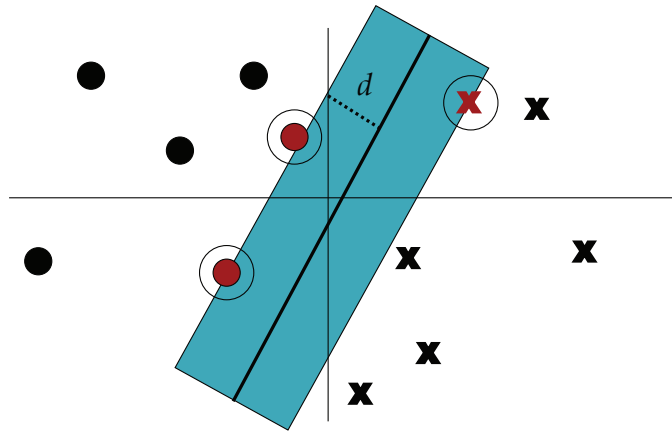


Figure 1.4: hyperplane

1.4 K-Nearest Neighbor

1.5 Neural Networks

References

- [1] Laura Auria and Rouslan A Moro. “Support vector machines (SVM) as a technique for solvency analysis”. In: (2008).
- [2] Robert Berwick. “An Idiot’s guide to Support vector machines (SVMs)”. In: *Retrieved on October 21* (2003), p. 2011.
- [3] Gavin C Cawley and Nicola LC Talbot. “On over-fitting in model selection and subsequent selection bias in performance evaluation”. In: *Journal of Machine Learning Research* 11.Jul (2010), pp. 2079–2107.
- [4] Eibe Frank and Mark Hall. “A simple approach to ordinal classification”. In: *European Conference on Machine Learning*. Springer. 2001, pp. 145–156.
- [5] GV Garje et al. “SENTIMENT ANALYSIS: CLASSIFICATION AND SEARCHING TECHNIQUES”. In: (2016).
- [6] Chih-Wei Hsu and Chih-Jen Lin. “A comparison of methods for multiclass support vector machines”. In: *IEEE transactions on Neural Networks* 13.2 (2002), pp. 415–425.
- [7] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. “A practical guide to support vector classification”. In: (2003).
- [8] David Madigan. *Descriptive Modeling*. 2002.
- [9] Tan Pang-Ning, Michael Steinbach, Vipin Kumar, et al. “Introduction to data mining”. In: *Library of congress*. Vol. 74. 2006.
- [10] Fabricio Voznika and Leonardo Viana. *Data Mining Classification*. 2007.

Acronyms

DT Decision Tree. 3

NN Neural Network. 3

OCR Optical Character Recognition. 1

SVM Support Vector Machine. 3