A Semi-supervised Approach to Multi-dimensional Classification with Application to Sentiment Analysis

Jonathan Ortigosa-Hernández[1], Juan Diego Rodríguez [1], Leandro Alzate[2], Iñaki Inza[1], and José A. Lozano[1]

[1]Intelligent Systems Group, Department of Computer Science and Artificial Intelligence
The university of the Basque Country, San Sebastián, Spain
[2]Socialware Company, Bilbao, Spain
✉{jonathan.ortigosa,juandiego.rodriguez,inaki.inza,ja.lozano}@ehu.es, leandro.alzate@asomo.net

Abstract

A classical supervised classification task tries to predict a single class variable based on a dataset composed of a set of labelled examples. However, in many real domains more than one variable could be considered as a class variable, so a generalisation of the single-class classification problem to the simultaneous prediction of a set of class variables should be developed. This is referred to as multi-dimensional supervised classification.

In addition, when performing classification tasks one can be concerned about the fact that, in some practical circumstances, obtaining enough labelled examples for a classifier may be costly and time consuming. Thus, it is very desirable to have learning algorithms that are able to incorporate a large number of unlabelled data with a small number of labelled data when learning classifiers. This is referred to as semi-supervised learning.

In this paper, we integrate multi-dimensional classification and semi-supervised learning by means of the multi-dimensional Bayesian network classifiers and the EM Algorithm. Results are presented on a real Sentiment Analysis dataset demonstrating that this integration can be beneficial to improve the recognition rates.

1 Introduction

Supervised classification [2] is one of the most important tasks in the pattern recognition field and it is widely used to deal with many real-life problems known as classification problems. In these tasks, a training set of instances is available, and each instance is described by a set of features and a unique known class variable. With the help of a training set, the classification process attempts to construct a description of the class variable, which in turn helps to classify a new unlabelled instance.

However, many application domains naturally consider more than one class variable. For instance, a text document or a semantic scene can be assigned to multiple topics, a gene can have multiple biological functions or a patient may suffer from multiple diseases. This kind of problems is known as multi-dimensional classification problems [23].

Due to the fact that they only consider a unique class variable, classical supervised classification approaches cannot be straightforwardly applied to the multi-dimensional problems scenario. Several attempts have been made to adapt single-class classifiers to multi-dimensional classifiers, but none of them exactly captures the problem characteristics. The most simple approach is to divide the multi-dimensional problem into several one-class problems (one for each class variable) and tackle them as if they were independent. However, this approach does not capture the real characteristics of the problem, because it does not explicitly model the correlations between the different class variables and hence, it does not take advantage of the information that they may provide [19]. So, within this context, multi-dimensional supervised classification [1][18][19][23], which is an approach that tries to take advantage of this correlation, appears in order to not only capture the underlying nature of the multi-dimensional problems, but also to use the relationships between the class variables to improve the recognition rates.

In addition, we are also concerned about the fact that several papers in the recent past have addressed the multi-dimensional classification task by building classifiers that rely exclusively on labelled examples [1]. However, some practical circumstances, obtaining enough labelled examples for a classifier may be costly and time consuming, and this problem is accentuated when using a large number of target variables. Thus, the scarcity of labelled data also motivates us to deal with unlabelled examples in a semi-supervised framework when working with the exposed multi-dimensional problem.

Motivated by the previous comments on the state-of-the-art of the multi-dimensional classification, in this paper, the following proposals are presented: (i) the extension of multi-dimensional classification to the semi-supervised learning framework by proposing a set of semi-supervised algorithms which make use of the EM algorithm [6][15], (ii) a supervised filter learning algorithm for Multi-dimensional J/K dependences Bayesian classifiers [19], and (iii) the demonstration of the fact that using both multi-dimensional classification and semi-supervised learning can be beneficial to improve recognition rates in a real application of Sentiment Analysis.

The rest of the paper is organised as follows. Section 2 defines either the multi-dimensional supervised classification paradigm and the multi-dimensional class Bayesian network classifiers. A group of algorithms to learn different types of multi-dimensional Bayesian classifiers in a supervised framework is introduced in Section 3. Section 4 extends the supervised algorithms presented in the previous section to the semi-supervised framework. Section 5 reviews the work related to Sentiment Analysis and shows the experimental results of applying the proposed semi-supervised classification algorithms to a real database of opinion analysis. Finally, Section 6 sums up the paper with some conclusions.

2 Multi-dimensional Classification

A typical supervised classification problem consists of building a classifier from a labelled training dataset $D = \{(\mathbf{x}^{(1)}, c^{(1)})...(\mathbf{x}^{(N)}, c^{(N)})\}$ in order to predict the value of a class variable C given a set of features $\mathbf{X} = (X_1, ..., X_n)$ of an unseen unlabelled instance $\mathbf{x} = (x_1, ..., x_n)$. A generalisation of this problem to the joint prediction of several class variables has recently been proposed in the research community [1][18][19][23]. This generalisation is known as multi-dimensional supervised classification. Our purpose is to predict the value of each class variable in the class variable vector $\mathbf{C} = (C_1, ..., C_m)$ given the feature vector of an unseen unlabelled instance.

2.1 Multi-dimensional Class Bayesian Network Classifiers

In order to deal with these multi-dimensional problems, we propose the use of multi-dimensional class Bayesian network classifiers (MDBNC) [18][23], which are a recent generalisation of the classical Bayesian network classifiers [13] to deal with multiple class variables.

MDBNC represent the underlying joint probability of the data by making use of directed acyclic graphs (DAG) over the class variables and over the feature variables separately, and then, by connecting the two sets of variables by means of a bi-partite directed graph. So, the DAG structure $S = (\mathbf{V}, \mathbf{A})$ has the set $\mathbf{V}$ of random variables partitioned into the sets $\mathbf{V}_C = \{C_1, \ldots, C_m\}$, $m > 1$, of class variables and the set $\mathbf{V}_F = \{X_1, \ldots, X_n\}$, $n \geq 1$, of features. Moreover, the set of arcs $\mathbf{A}$ can be partitioned into three sets: $\mathbf{A}_{CF}$, $\mathbf{A}_C$ and $\mathbf{A}_F$ with the following properties:

- $\mathbf{A}_{CF} \subseteq \mathbf{V}_C \times \mathbf{V}_F$ is composed of the arcs between the class variables and the feature variables, so we can define the feature selection subgraph of S as $S_{CF} =$

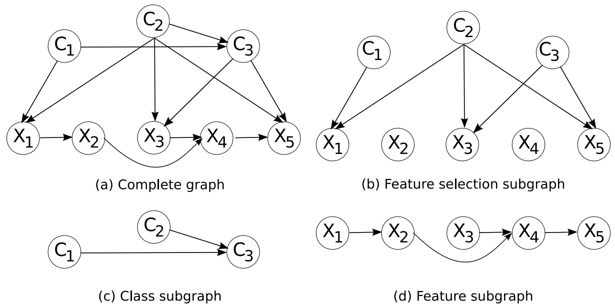

Figure 1: A multi-dimensional Bayesian classifier and its division [18].

$(\mathbf{V}, \mathbf{A}_{CF})$. This subgraph represents the selection of features that seems relevant for classification given the class variables.

- $\mathbf{A}_C \subseteq \mathbf{V}_C \times \mathbf{V}_C$ is composed of the arcs between the class variables, so we can define the class subgraph of S induced by $\mathbf{V}_C$ as $S_C = (\mathbf{V}_C, \mathbf{A}_C)$.

- $\mathbf{A}_F \subseteq \mathbf{V}_F \times \mathbf{V}_F$ is composed of the arcs between the feature variables, so we can define the feature subgraph of S induced by $\mathbf{V}_F$ as $S_F = (\mathbf{V}_F, \mathbf{A}_F)$.

Figure 1 shows a multi-dimensional class Bayesian network classifier with 3 class variables and 5 features, and its partition into the three subgraphs.

Depending on the structure of the three subgraphs, several sub-families[1] of MDBNC have been proposed in the state-of-the-art literature. These are the Multi-dimensional naive Bayes classifier (MDnB), the Multi-dimensional tree-augmented classifier (MD-TAN) and the Multi-dimensional J/K dependences Bayesian classifier (MD J/K).

In addition to its structure, a MDBNC (as happens in the classical Bayesian networks) also consists of a set of parameters $\mathbf{\Theta}$ that codify the local probability distribution.

[1] In [23] and [18], instead of *multi-dimensional*, the term *fully* is used in order to name the classifiers.

3 Supervised Learning of MDBNC

Each previously introduced sub-family has different restriction in its structure, so a different structure learning algorithm is needed for each one. In this section, we define the MDnB, MDTAN and MD J/K sub-families and provide a structure learning algorithm for each sub-family.

3.1 MDnB

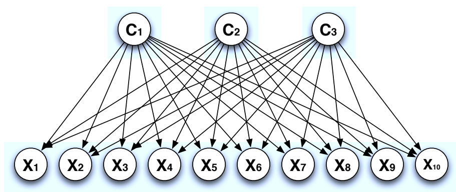

Figure 2: An example of a MDnB structure.

In the MDnB [23] (see Figure 2), both the class subgraph and the feature subgraph are empty, and the feature selection subgraph is complete. As happens in uni-dimensional naive Bayes, this classifier assumes conditional independence between each pair of features given the entire subset of class variables. When this classifier is provided with a training dataset with a determined number of class variables and features, it has a fixed structure and so, there is no need for a structure learning. Therefore, learning a MDnB classifier consists of just estimating the parameters $\mathbf{\Theta}$ of the fixed structure by using a training dataset D.

3.2 MDTAN

The MDTAN (see Figure 3), as proposed in [23], is the generalisation of the uni-dimensional TAN classifier [9]. In this multi-dimensional classifier, both the class subgraph and the feature subgraph are directed trees.

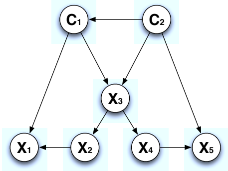

Figure 3: An example of a MDTAN structure.

A structure learning algorithm that learns a MDTAN from a given dataset is proposed in [23]. This algorithm follows a wrapper approach by performing a local search over the A_{CF} structure. Its aim is to produce the MDTAN structure that maximises the accuracy from a given dataset. In order to obtain a MDTAN structure in each iteration, it generates a set of different A_{CF} from a current A_{CF} and uses mutual information to create the class subgraph and feature subgraph by building two maximum spanning trees. When there is no improvement when generating a new set of structures, the algorithm stops.

3.3 MD J/K

The MD J/K (see Figure 4) appears in [19] and it is the multi-dimensional generalisation of the well-known K-DB [20]. It allows each class variable C_i to have a maximum of J dependences with other class variables, and each predictive variable X_i to have, apart from the class variables, a maximum of K dependences with other predictive variables.

As happens in the uni-dimensional K-DB version proposed by Sahami [20], this structure is also able to move through the spectrum of allowable dependence in the multi-dimensional framework, from the MDnB to the full multi-dimensional Bayesian classifier. Note that setting $J = K = 0$ we can learn a MDnB, setting $J = K = 1$ a MDTAN is learned and so on. The full multi-dimensional Bayesian classifier, which is the classifier with has the three subgraphs complete, can be learnt by setting $J = (m-1)$ and $K = (n-1)$, where m and n are the number of class variables and predictive features respectively.

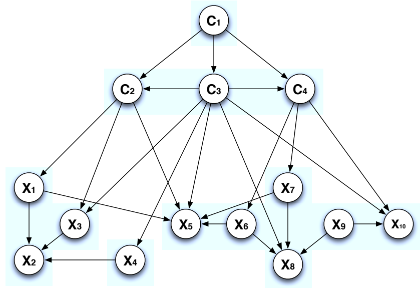

Figure 4: An example of a MD 2/3 structure.

Although the MD J/K structure has been proposed in the state-of-the-art literature, to the best of our knowledge, a specific MD J/K learning algorithm has not been defined by the research community. For that reason, we propose, in this section, a filter algorithm in a supervised learning framework able to learn this type of structure (see Algorithm 1), i.e. the learning algorithm uses mutual information to measure the dependency between the variables, instead of performing a search for the structure that maximises the accuracy.

In this algorithm, we do not directly use the mutual information as measured in the previous MDTAN learning algorithm [23]. This is due to the fact that the mutual information is not normalised when the cardinalities of the variables are different, so we use an independence test to determine if a dependence between two variables is strong enough to be part of the model.

Algorithm 1 MD J/K structure learning algorithm using a filter approach

1: **Learn the A_C structure**

1. Calculate the α-value (significance of the mutual information) using the independence test for each pair of class variables, and rank them.
2. Remove the α-values lower than the threshold $s_\alpha = 0.90$.
3. Use the ranking to add arcs between the class variables fulfilling the conditions of no cycles between the class variables and no more than J-parents per class.

2: **Learn the A_{CF} structure**

1. Calculate the α-value (significance of the mutual information) using the independence test for each pair C_i and X_j and rank them.
2. Remove the α-values lower than the threshold $s_\alpha = 0.90$.
3. Use the ranking to add arcs from the class variables to the features.

3: **Learn the A_F structure**

1. Calculate the α-value (significance of the conditional mutual information) using the conditional independence test for each pair X_i and X_j given $\mathbf{Pa_c}(X_j)$ and rank them.
2. Remove the α-values lower than the threshold $s_\alpha = 0.90$.
3. Use the ranking to add arcs between the class variables fulfilling the conditions of no cycles between the features and no more than K-parents per feature.

The A_C and A_{CF} structures are constructed as follows: we use the mutual information between each pair of random variables to create an independence test based on [12] (see pp. 155–158). This test is used to calculate the significance of each relationship. In the case of A_C each pair of class variables (C_i, C_j) is tested, while in the case of A_{CF} each pair of a class variable and a feature (C_i, X_j) is tested. If the significance of the relationship between the two variables surpasses a determined threshold, then an arc is included between the random variables. Note that, in the case of the A_C structure, we have to consider some restrictions when adding a new arc. These restriction are the allowable maximum number of parents and no cycles in the structure.

In order to construct the A_F structure, we use the conditional mutual information between two random variables for the independence test [12] (see pp. 166–167). In this case, we perform the test in each pair of features X_i and X_j given the class parents of X_j, $\mathbf{Pa_c}(X_j)$. After that, the modus operandi of the algorithm is the same as in the case of the A_C structure.

Before the end of this section, we want to make some comments about two designing decisions we have made:

- In order to make the algorithm more robust [20], we have decided to introduce the threshold s_α in steps 1.2, 2.2 and 3.2 of Algorithm 1. This decision allows more flexibility by not forcing the inclusion of dependencies that do not appear to exist when the values of K and J are set too high.
- From the previous paragraphs and step 3 of Algorithm 1, one can easily deduce that the class parents of a feature do not count for the restriction of maximum K parents for each predictive feature. As in the case of the uni-dimensional KDB, the class parents can either count for the restriction or be left out of it.

4 Semi-supervised Learning in Multi-dimensional Classification

In the semi-supervised learning framework [4][26], the dataset D is divided into two parts: the instances $D_L = \{(\mathbf{x}^{(1)}, \mathbf{c}^{(1)}), ..., (\mathbf{x}^{(L)}, \mathbf{c}^{(L)})\}$ for which labels are provided, and the instances

$D_U = \{(\mathbf{x}^{(L+1)}, \mathbf{?}), ..., (\mathbf{x}^{(N)}, \mathbf{?})\}$, where the labels are not known. Therefore, we have a dataset of N instances, where there are L labelled examples and $(N - L)$ unlabelled examples. The aim of semi-supervised learning is to build more accurate classifiers using both labeled and unlabelled data, rather than using exclusively labeled examples as happens in supervised learning.

In order to deal with these kind of problems, the EM algorithm [6][15] has been widely used in the semi-supervised learning framework [5][16]. The aim of the EM algorithm as typically used in semi-supervised learning [16] is to find the parameters of the model that maximise the likelihood of the data, using both labelled and unlabelled instances. The iterative process works as follows: in the i-th iteration the algorithm alternates between completing the unlabelled instances by using the parameters $\mathbf{\Theta}^{(i)}$ (E-step) and updating the parameters of the model $\mathbf{\Theta}^{(i+1)}$ calculating the maximum likelihood estimator (MLE) with the whole dataset (M-step), i.e. the labelled data and the unlabelled instances that have been previously classified in the E-Step. Note that the structure remains fixed in the whole iterative process.

Although good results have been achieved with the EM algorithm in uni-dimensional classification [5][16], we are concerned about the restriction of maximising just the parameters of a fixed structure in the multi-dimensional framework. Due to the fact that several class variables have to be simultaneously predicted, the structures of the MDBNC tend to be more complex than the structures of the uni-dimensional Bayesian network classifiers. For that reason, it seems more appropriate to perform a structural search. Thus, we perform several changes to the EM algorithm in order to avoid fixing the structure of the model during the iterative process. The proposal is shown in Algorithm 2.

Algorithm 2 Our version of the EM Algorithm

Input: A training dataset with both labelled and unlabelled data and an initial model $\psi^{(i=0)}$ with a fixed structure and with an initial set of parameters $\mathbf{\Theta}^{(i=0)}$.

```
while the model ψ^(i) does not converge do
    E-STEP Use the current model ψ^(i) to estimate the probability of each configuration of class variables for each unlabelled instance.
    M-STEP Learn a new model ψ^(i+1) with structure and parameters, given the estimated probabilities in the E-STEP.
end while
```

Output: A classifier ψ, that takes an unlabelled instance and predicts the class variables.

In this version of the EM algorithm we want to find the model, both structure and parameters, that maximises the likelihood of the whole data. So, in this version, the i-th iteration of the algorithm is performed as follows: it alternates between completing the unlabelled instance by the previously learnt model $\psi^{(i)}$ (E-step) and learning a new model $\psi^{(i+1)}$ by using a learning algorithm with the whole dataset, both labelled and completed instances (M-step). In the semi-supervised learning research community, the input parameter $\psi^{(i=0)}$ of the EM Algorithm is usually learnt from the labelled subset D_L. Hence, we will continue using this modus operandi in this version of the algorithm. Note that our version of the EM algorithm is closer to the Bayesian structural EM algorithm proposed in [8] rather than the original [6].

Using Algorithm 2, all the supervised learning approaches proposed in the previous section can be straightforwardly used. The learning algorithm is used in the M-STEP, where it learns a model using labelled and unlabelled data that have been previously labelled in the E-STEP. So, applying our EM Algorithm we have extended the MDBNC to the semi-supervised learning framework generating a new set of semi-supervised multi-dimensional classifiers.

5 Sentiment Analysis

In this section, we review several works proposed for approaching Sentiment Analysis classification. After that, our proposed algorithms are applied to a real dataset obtained from the company Socialware, one of the most important companies of mobilised opinion analysis in Europe, with 3 different class variables and 14 features.

5.1 State-of-the-art literature

Sentiment Analysis (SA), which is also known as Opinion Mining, is defined as the computational study of opinions, sentiments and emotions expressed in text [14]. It mainly originated to meet the need for organisations to automatically find on the Internet the opinions or sentiments of the general public about their products and services. Although SA has been treated as multiple uni-dimensional different problems, it has indeed a multi-dimensional nature, as we show in this paper.

Treating SA as a text classification problem is the area that has been most researched in the academia [14]. This approach is mainly divided into two sub-problems that have been widely studied in the last few years: sentiment classification and subjectivity classification. Both can be formulated as learning problems:

1. The aim of *Sentiment classification* [11] is to learn classifiers able to classify an opinionated text, defined as a set of features, as expressing a positive, neutral or negative opinion. Some authors do not consider the class label neutral [17], while others consider the 1-5 stars rating [21].

2. *Subjectity classification* [24] can be formulated as the learning problem of determining whether a text is objective or subjective.

Due to the fact that both can be expressed as learning problems, the existing methods can be readily applied for sentiment and subjectivity classification, e.g. Bayesian network classifiers or support vector machines. However, in order to apply straightforwardly those learning methods, the text has to be transformed into a set of features $\mathbf{X} = \{X_1, X_2, ...X_n\}$. Liu, in [14], presents several approaches that have been used to represent a text in a set of features such as "terms and their frequency" or "part-of-speech tags".

Although a large amount of work have been proposed not only in engineering a suitable set of features, but also in studying both sub-problems, most existing researches only focus on one of the two sub-problems, and to the best of our knowledge neither of them performs a simultaneous sentiment-subjectivity classification framework. Even though, several papers have noticed the need to predict both the sentiment and the subjectivity values of a given text [7][25]. For instance, [25] reported a study which tries to classify subjective sentences and also determine their opinion orientation. However, they perform both classifications in series, not simultaneously, without using the correlation between the sentiment and the subjectivity as is explicitly modelled in multi-dimensional classification.

Finally, despite the fact that SA is broadly applied to extract information from the web where a huge amount of unlabelled data can be found, little work [10] had been carried out in learning SA problems in a semi-supervised framework.

5.2 The ASOMO dataset

Since typical benchmark data repositories in Sentiment Classification do not provide datasets with multiple class variables, we do not test our algorithms in the common collections used in the SA papers. Our experimentation is performed using a real dataset extracted from *ASOMO service* of mobilised opinion analysis.

The ASOMO dataset has been collected by Socialware and it consists of $2,542$ reviews written in Spanish. 150 of these documents have been labelled, in isolation, by an expert in Socialware company and $2,392$ are left as unlabelled instances.

Each document is preprocessed using an open source morphological analyser [3]. This analyser provides information related to part-of-speech, which is helpful in detecting a list of 14 morphological features, expressed as a ratio in the range $[0, 1]$, which characterise each analysed document.

The dataset has three different class variables, which are naturally related in SA: *Sentiment* and *Subjectivity* (as mentioned before), and a third one called *Will to Influence*. This class variable, frequently used in the framework of ASOMO, is defined as the dimension that rates the desire of the opinion holder to provoke a certain reaction in the potential readers of the text. It has four possible values: declarative text, soft, medium and strong will to influence. Sentiment, in this dataset, has 5 different labels as occurs in the 1-5 stars ratings. So, in addition to the three classic values, it has the values "very negative" and "very positive".

5.3 Experimentation on the ASOMO dataset

In order to evaluate our semi-supervised multi-dimensional set of algorithms, the following experiment is performed: The ASOMO dataset has been used to learn three different (uni-dimensional) Bayesian network classifiers and three different sub-families of MDBNC. For uni-dimensional classification, naive Bayes classifier (*nB*), tree-augmented network classifier (*TAN*) and a 2 dependence Bayesian classifier have been chosen. In the multi-dimensional side, *MDnB*, *MDTAN*, *MD* $2/K$, with $K = 2, 3, ...6$ structures have been selected. In order to compare supervised and semi-supervised frameworks, all the structures are learnt in both scenarios. As stated before, the uni-dimensional approach cannot be straightforwardly apply to deal with the multi-dimensional problems, so, when learning uni-dimensional classifiers we divide the dataset into three one-class variable tasks and tackle them as independent.

The supervised learning procedure only uses the labelled dataset (consisting of 150 documents), whilst the semi-supervised approach uses the $2,532$ reviews. Our multi-dimensional extension of the EM algorithm is used in the latter approach and it terminates after finding a local likelihood maxima or 250 iteractions. Due to the fact that the features of the dataset are continuous, they are discretised into three values using equal frequency discretisation. The parameters of the models are calculated by MLE corrected with Laplace smoothing. Finally, the performance of each model has been estimated via 5 runs of 5-fold non-stratified cross validation [22].

Table 8 shows the results of the algorithms over the ASOMO dataset in both supervised and semi-supervised approaches. In addition to the joint accuracy[2], the accuracies per each class variable are shown. The accuracies in bold correspond to the best accuracies obtained in both supervised and semi-supervised frameworks for each class variable and for the joint accuracy metric. Based on the results, several conclusions can be extracted:

1. The multi-dimensional classification approach statistically outperforms the uni-dimensional classification in terms of joint accuracy (Student's t-test with $\alpha = 0.95$).
2. The semi-supervised learning framework obtains better joint accuracies than supervised learning.
3. The class variable Will to Influence tends to degrade its performance in semi-supervised learning, whilst the others tend to achieve better single accuracies.
4. The MDTAN approach [23] tends to behave more similar to the uni-dimensional approaches rather than to the approaches it belongs to, the multi-dimensional ones.

In conclusion, we show in this experiment that the proposed semi-supervised multi-dimensional formulation proposes a novel perspective for this kind of problems, opening new ways to deal with these problems. In addition, it can also be seen that the explicit use and the representation of the relationships between different class variables, as well as the

[2] This measure estimates the values of all class variables simultaneously, that is, it only counts a success if all the classes are correctly predicted, otherwise it counts an error [19].

Classif.	LABELLED DATA				LABELLED AND UNLABELLED DATA			
	Will to Influence	Sentiment	Subjectivity	JOINT Acc	Will to Influence	Sentiment	Subjectivity	JOINT Acc
nB	56.80 ± 2.18	28.00 ± 3.42	82.80 ± 1.19	11.47 ± 2.76	$\mathbf{58.67 \pm 2.21}$	27.33 ± 3.86	54.13 ± 4.15	09.33 ± 2.75
TAN	51.33 ± 1.76	27.20 ± 3.75	79.47 ± 1.19	09.47 ± 1.79	48.13 ± 3.60	28.40 ± 3.79	70.40 ± 2.69	10.53 ± 3.63
2DB	56.53 ± 1.10	29.33 ± 3.40	82.80 ± 0.56	11.73 ± 2.24	35.20 ± 6.23	29.47 ± 1.85	83.60 ± 0.89	09.20 ± 2.51
MDnB	53.60 ± 0.89	29.33 ± 1.63	83.07 ± 1.21	14.80 ± 1.66	53.87 ± 2.60	$\mathbf{32.13 \pm 0.99}$	$\mathbf{84.00 \pm 0.67}$	$\mathbf{16.80 \pm 2.28}$
MDTAN	52.13 ± 2.23	26.27 ± 1.61	$\mathbf{83.47 \pm 0.73}$	13.33 ± 2.05	34.93 ± 6.30	28.27 ± 2.69	83.07 ± 1.12	08.57 ± 2.88
MD 2/2	52.00 ± 1.70	28.80 ± 3.03	78.53 ± 2.51	15.07 ± 3.64	52.40 ± 2.61	26.27 ± 4.70	74.80 ± 2.33	12.52 ± 1.45
MD 2/3	56.67 ± 2.26	29.07 ± 1.92	76.00 ± 1.05	$\mathbf{15.87 \pm 1.91}$	50.67 ± 4.42	27.33 ± 1.42	75.60 ± 4.53	14.40 ± 1.30
MD 2/4	56.93 ± 1.67	27.74 ± 3.22	77.73 ± 2.65	14.80 ± 1.73	51.60 ± 3.25	29.73 ± 2.77	77.47 ± 1.97	16.00 ± 1.56
MD 2/5	$\mathbf{56.94 \pm 2.69}$	27.86 ± 2.47	75.73 ± 3.18	13.73 ± 2.24	53.20 ± 3.21	26.93 ± 5.00	78.27 ± 3.96	15.47 ± 1.59
MD 2/6	56.07 ± 3.56	$\mathbf{31.05 \pm 3.41}$	76.57 ± 2.10	15.47 ± 2.18	53.60 ± 1.80	28.67 ± 1.05	77.73 ± 1.21	16.53 ± 0.55

Table 1: Accuracies on ASOMO dataset

use of unlabelled data, can be beneficial to improve the recognition rates in SA problems, demonstrating that the SA problem has indeed a multi-dimensional underlying nature.

6 Conclusion

Multi-dimensional classification and semi-supervised learning are two different branches of machine learning. While multi-dimensional supervised classification is the generalisation of the single-class supervised classification problem to the simultaneous prediction of a set of class variables, semi-supervised learning is the learning paradigm concerned with the study of how more accurate classifiers can be learnt by adding unlabelled examples to the labelled ones. In this paper, we establish a bridge between them.

At first instance, a supervised filter learning algorithm for MD J/K classifiers has been proposed. After that, we have extended, in addition to the MD J/K learning algorithm, the state-of-the-art of supervised MDBNC learning algorithms by means of a multi-dimensional extension of the EM Algorithm. Then, we have applied the proposed battery of semi-supervised multi-dimensional learning algorithms to a real dataset of SA showing that they are competitive with many existing supervised learning algorithms. Besides, we have also shown that multi-dimensional learning algorithms outperform the uni-dimensional techniques when dealing with semi-supervised multi-dimensional problems. This demonstrates that SA classification is, in fact, a multi-dimensional problem.

In short, we have proposed a novel perspective for this kind of problems and demonstrated that the use of multi-dimensional classification, as well as the use of unlabelled data, can be beneficial to improve the recognition rates.

Acknowledgments

This work has been partially supported by the Saiotek, Etortek and Research Groups 2007-2012 (IT-242-07) programs (Basque Government), TIN2008-06815-C02-01, Consolider Ingenio 2010-CSD2007-00018 projects and MEC-FPU grant AP2008-00766, and COMBIOMED network in computational biomedicine (Carlos III Health Institute).

References

[1] C. Bielza, G. Li, and P. Larrañaga. Multi-dimensional classification with Bayesian networks. Technical Report UPM-FI/DIA/2010-1, Departmento de Inteligencia Artificial, Universidad Politécnica de Madrid, Madrid, Spain, 2010.

[2] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2006.

[3] X. Carreras, I. Chao, L. Padro, and M. Padro. An open-source suite of language analyzers. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, volume 10, pages 239–242, 2006.

[4] O. Chapelle, B. Scholkopf, and A. Zien. *Semi-supervised learning.* The MIT Press, 2006.

[5] I. Cohen, F. G. Cozman, N. Sebe, M. C. Cirelo, and T. S. Huang. Semisupervised learning of classifiers: Theory, algorithms and their application to human-computer interaction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(12):1553–1567, 2004.

[6] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[7] A. Esuli and F. Sebastiani. Sentiwordnet: A publicly available lexical resource for opinion mining. In *In Proceedings of the 5th Conference on Language Resources and Evaluation (LRECÕ06)*, pages 417–422, 2006.

[8] N. Friedman. The bayesian structural EM algorithm. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 129–138. Morgan Kaufmann Publishers, 1998.

[9] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2-3):131–163, 1997.

[10] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger. Pulse: Mining customer opinions from free text. In *Proceedings of the 6th International Symposium on Intelligent Data Analysis*, pages 121–132, 2005.

[11] M. Koppel and J. Schler. Using neutral examples for learning polarity. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2005)*, 2005.

[12] S. Kullback. *Information Theory and Statistics.* Wiley, 1959.

[13] P. Larrañaga, J. A. Lozano, J. M. Peña, and I. Inza. *Special Issue on Probabilistic Graphical Models for Classification*, volume 59(3). Machine Learning, 2005.

[14] B. Liu. Sentiment analysis and subjectivity. In N. Indurkhya and F. J. Damerau, editors, *Handbook of Natural Language Processing.* Chapmand & Hall, second edition, 2010.

[15] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions.* Wiley-Interscience, 1997.

[16] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell. Text classification from labeled and unlabeled documents using em. *Machine Learning*, 39(2):103–134, 2000.

[17] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'08)*, pages 79–86, 2002.

[18] J. D. Rodríguez and J. A. Lozano. Multi-objective learning of multi-dimensional bayesian classifiers. In *Eighth Conference on Hybrid Intelligent Systems (HIS'08)*, pages 501–506, September 2008.

[19] J. D. Rodríguez and J. A. Lozano. Learning Bayesian network classifiers for multi-dimensional supervised classification problems by means of a multi-objective approach. Technical Report EHU-KZAA-TR-3-2010, Department of Computer Science and Artificial Intelligence, University of the Basque Country, San Sebastián, Spain, 2010.

[20] M. Sahami. Learning limited dependence bayesian classifiers. In AAAI Press, editor, *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 335–338, 1996.

[21] B. Snyder and R. Barzilay. Multiple aspect ranking using the good grief algorithm. In *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL)*, pages 300–307, 2007.

[22] M. Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B*, 36(1):111–147, 1974.

[23] L. C. van der Gaag and P. R. de Waal. Multi-dimensional Bayesian network classifiers. In *Proceedings of the Third european workshop in probabilistic graphical models*, pages 107–114, 2006.

[24] J. M. Wiebe, R. F. Bruce, and T. P. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 246–253, 1999.

[25] H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2003.

[26] X. Zhu and A. B. Goldberg. *Introduction to Semi-Supervised Learning.* Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2009.