Approaching Sentiment Analysis by Using Semi-supervised Learning of Multi-dimensional Classifiers

Artificial Experiments Report

1. Introduction

In this report, we execute the proposed battery of semi-supervised multi-dimensional learning algorithms over a set of designed artificial datasets as commonly done in the machine learning research community. These experiments are performed in order to demonstrate that the proposed algorithms are able to take advantage of the underlying nature of the multi-dimensional problems even in the presence of a small set of labelled data and a huge set of unlabelled data. By means of these experiments, we would like to shed some light on the following questions:

1. Are there significant differences between the uni-dimensional and the multi-dimensional supervised learning algorithms when there are scarcity of labelled examples? Are there significant differences between the uni-dimensional and the multi-dimensional semi-supervised learning algorithms?

2. If the correct structure of the generative model is obtained, do unlabelled data improve the classifier?

3. If the learning algorithm can lead to the correct structure of the generative model, do unlabelled data improve the classifier?

4. Can adding unlabelled data contribute to an increase in the classification performance (in terms of joint accuracy) when there is a small amount of labelled data in a multi-dimensional framework?

5. Do multi-dimensional classifiers performs better than the uni-dimensional in a multi-dimensional semi-supervised framework?

2. Simulation of datasets

In order to carry out the experimentation process required to evaluate our proposals, we use a set of artificial multi-dimensional datasets. These datasets are sampled from multi-dimensional feature-class variable probability distributions $p(\mathbf{x}, \mathbf{c})$ represented as multi-dimensional Bayesian network classifiers. These classifiers have been created in four steps: First, the structure of the multi-dimensional Bayesian network classifier was created. Second, the parameters of the classifiers were obtained by sampling a Dirichlet distribution. Third, a dataset from each classifier was sampled. Finally, the semi-supervised nature of the sampled dataset, i.e. both subsets of labelled and unlabelled data, was generated by choosing instances at random for both types of data subsets. The entire process of creating the artificial datasets has been performed by means of the free software ICLAB library [1].

The following sub-families of multi-dimensional Bayesian network structures have been chosen for this experimentation: *MDnB*, *MDTAN*, *MD* 2/2 and *MD* 2/3. A number of 5 different structures per each subfamily with a different number of features (from 5 to 20) and a different number of class variables (from 2 to 4) have been created. The cardinality of the features ranges from 2 to 4 and the cardinality of the class variables from 2 to 3. By means of these sub-families and structures we are trying to cover a broad range of structures of different complexity in order to check the statement "performance degradation may occur whenever the modelling assumptions adopted for a particular classifier do not match the characteristics of the distribution generating the data [5]". In order to generate the parameters of the classifiers which are defined by the previous structures, a different Dirichlet distribution with all its parameters equal to one is sampled per each classifier. The associated Bayes errors of the resulting classifiers can be seen in Table 1.

Once the multi-dimensional Bayesian network classifiers have been constructed, a dataset from each of them is sampled. Specifically, we sample 20 artificial datasets (5 for each different sub-family) of $15,100$ instances. Then, we divide each dataset into two different parts: a training set of $10,100$ instances used to learn the classifiers, and a test set of $5,000$ samples used to estimate the prediction error of the classifiers. In order to simulate the semi-supervised nature of the training dataset, 100 instances were chosen at random to form the subset D_L, in which labelled instances are i.i.d. as in $p(\mathbf{x}, \mathbf{c})$. The labels of the class variables of the remaining $10,000$ instances are removed.

All the datasets. as well as the structures and the parameters of the designed classifiers can be found in the following website[1].

3. Experimental Setup

Once the semi-supervised multi-dimensional datasets are created, we use them to perform the following set of experiments. In the evaluation phase, we proposed nine different algorithms to learn the classifiers from the sampled datasets. The first four algorithms are the approaches explicitly designed for multi-dimensional classification proposed in the paper: *MDnB*, *MDTAN*, *MD* 1/1. *MD* 2/2 and *MD* 2/3. Due to the fact that MDTAN learning algorithm [13] follows a wrapper approach, the MD 1/1 is included in this experiments as a filter approach in order to establish a comparison between both techniques. The others are the well-known uni-dimensional approaches: naive Bayes classifier (*nB*), tree-augmented network classifier (*TAN*) [8] and two K dependence Bayesian classifiers [12] (one setting $K = 2$, 2-*DB*, and the other setting $K = 3$, 3-*DB*). As stated in the article, the uni-dimensional approach cannot be straightforwardly applied to deal with the multi-dimensional problems, so, we divided the multi-dimensional problem into several one-class variable tasks and tackled them as independent.

In order to compare the supervised and the semi-supervised frameworks, all the algorithms are learnt in both scenarios. In the case of supervised learning the algorithms are straightforwardly applied to the 100 instances of the labelled subset. When learning in the semi-supervised framework, on the contrary, the algorithms are used, as proposed in this work, in conjunction with our extension of the EM algorithm

[1] http://www.sc.ehu.es/ccwbayes/members/jonathan/home/ISG/News_and_Notables/Entries/2010/11/30_IMACS_2011.html

(Algorithm 2 of the article) and applied to the whole training dataset (100 labelled + $10,000$ unlabelled). The EM algorithm terminates after finding a local likelihood maxima or after 250 unsuccessful trials. The parameters of the model are calculated by maximum likelihood estimation (MLE) [4], corrected with Laplace smoothing. The estimation method for performance evaluation metrics is hold-out [10]. In hold-out, a subset of instances is chosen randomly from the initial dataset to form a training set used to learn a classifier, and the remaining instances are retained as the testing data, used to estimate the error of the classifier. This method has been chosen in order to evaluate all the algorithms in the same testing set, avoiding the variance of the error estimation given by the cross-validation methods [11]. Finally, the performance evaluation is performed by the joint evaluation criteria.

4. Results

Table 1 shows the results of the nine algorithms over the 20 datasets when they are applied in the supervised scenario. Table 2 shows, instead, their results in the semi-supervised framework. Each value in both tables corresponds to the joint accuracy obtained for each algorithm when the learnt classifier is evaluated in the testing set of $5,000$ samples. The accuracies in **bold** (per row) correspond to the technique with the best accuracy in just the labelled dataset (Table 1), and the best in the whole dataset (Table 2). Moreover, the best joint accuracy per dataset (per row in Tables 1 and 2) is highlighted in ***bold italics***.

In order to answer the questions proposed in the introduction of this report, we have performed an exhaustive analysis of the results of Tables 1 and 2. The following sections summarise the studies made to shed some light on the questions. In each section, the conclusions that answer these questions has been underlined.

LABELLED DATA											
Family	Dataset	$(1-e_B)$	nB	TAN	2DB	3DB	MDnB	MDTAN	MD 1/1	MD2/2	MD2/3
MDnB	NB01	*93.93*	77.06	80.20	75.68	74.38	**88.48**	66.22	77.38	76.40	75.96
	NB02	*93.25*	76.96	80.80	79.10	78.40	**89.90**	71.58	82.14	76.62	75.24
	NB03	*91.65*	76.98	79.94	76.86	75.54	**88.29**	68.43	79.72	75.82	73.86
	NB04	*74.36*	43.71	48.81	40.35	42.83	**65.57**	39.59	40.85	43.43	47.38
	NB05	*66.45*	40.20	43.92	37.19	38.55	**59.44**	44.64	40.71	37.99	34.82
MDTAN	TA01	*66.86*	57.94	***61.01***	56.14	47.86	53.53	57.86	58.18	57.98	54.36
	TA02	*47.61*	31.13	33.61	28.97	29.29	29.75	34.77	**37.46**	36.36	36.04
	TA03	*56.80*	40.59	46.36	39.35	37.74	37.56	39.12	**46.60**	44.26	44.38
	TA04	*64.21*	43.33	46.43	43.87	36.40	43.31	41.85	**49.26**	48.14	44.01
	TA05	*58.75*	31.78	41.70	28.81	30.85	27.85	42.03	45.87	45.31	**45.89**
MD2/2	2201	*66.26*	56.04	**58.64**	53.48	46.46	54.48	56.50	57.46	55.32	53.84
	2202	*65.44*	45.01	42.29	41.55	36.39	39.19	**46.11**	38.41	44.89	43.01
	2203	*61.72*	44.92	40.52	39.80	39.96	44.82	41.80	**48.06**	47.36	47.90
	2204	*84.85*	80.86	79.98	80.02	81.22	81.47	**83.96**	81.22	82.91	83.45
	2205	*50.28*	37.98	34.37	33.41	29.65	33.83	***43.80***	40.72	39.62	39.34
MD2/3	2301	*57.56*	**43.38**	40.93	39.40	37.80	39.16	38.82	41.84	41.36	42.90
	2302	*69.11*	61.38	57.10	60.52	55.14	56.24	55.70	60.54	**61.44**	59.88
	2303	*50.31*	29.10	34.35	28.98	27.04	28.58	31.06	***34.41***	31.22	33.41
	2304	*78.45*	67.13	68.41	67.27	61.63	64.09	***73.85***	71.91	72.15	71.33
	2305	*61.34*	48.82	47.72	46.40	49.34	47.48	**56.12**	54.24	51.92	54.48

Table 1: Estimated accuracies of the proposed algorithms in the supervised scenario.

LABELLED AND UNLABELLED DATA											
Family	Dataset	$(1-e_B)$	nB	TAN	2DB	3DB	MDnB	MDTAN	MD 1/1	MD2/2	MD2/3
MDnB	NB01	*93.93*	49.24	50.12	49.88	49.98	***92.16***	50.86	83.52	86.16	81.22
	NB02	*93.25*	41.30	42.26	41.34	41.58	***91.06***	41.94	87.32	84.06	83.78
	NB03	*91.65*	41.69	64.05	42.65	43.09	***89.09***	55.68	81.31	83.75	84.41
	NB04	*74.36*	3.62	30.66	14.17	24.08	***67.63***	26.72	48.94	50.24	48.64
	NB05	*66.45*	3.79	22.03	13.19	5.95	***60.43***	15.81	51.57	47.29	46.36
MDTAN	TA01	*66.86*	43.28	50.45	39.86	36.81	48.61	49.33	59.60	**60.02**	48.44
	TA02	*47.61*	5.43	19.08	6.27	21.46	25.15	28.71	***39.82***	37.68	38.30
	TA03	*56.80*	20.42	47.08	20.57	19.60	35.08	41.45	***47.82***	46.74	47.36
	TA04	*64.21*	28.39	39.45	19.52	22.49	43.35	38.55	52.56	53.06	***53.28***
	TA05	*58.75*	28.83	***48.01***	28.44	32.05	26.27	42.59	45.99	45.57	46.33
MD2/2	2201	*66.26*	14.10	36.20	27.12	28.30	53.24	35.46	57.80	***59.65***	59.06
	2202	*65.44*	31.31	33.59	31.71	24.72	38.93	15.62	40.89	***47.11***	46.27
	2203	*61.72*	27.66	35.12	28.76	35.96	42.72	31.40	45.94	48.56	***49.52***
	2204	*84.85*	73.94	72.26	73.86	73.74	78.18	***83.97***	79.14	80.28	81.33
	2205	*50.28*	20.96	11.01	12.77	12.41	28.95	30.61	42.18	**42.62**	41.62
MD2/3	2301	*57.56*	23.48	24.16	24.08	26.18	37.08	25.96	43.53	45.08	***46.22***
	2302	*69.11*	51.10	50.98	50.98	48.94	50.00	42.34	62.74	62.42	***64.12***
	2303	*50.31*	22.78	27.86	22.34	20.10	27.22	26.06	**33.20**	29.50	29.80
	2304	*78.45*	36.35	29.75	24.50	23.76	61.47	39.21	70.95	72.33	**72.81**
	2305	*61.34*	21.08	31.24	24.74	39.66	43.78	***57.72***	50.94	52.70	51.68

Table 2: Estimated accuracies of the proposed algorithms in the semi-supervised scenario.

4.1 Uni-dimensional and multi-dimensional learning algorithms comparison

Firstly, we want to determine if there are significant differences between the uni-dimensional and the multi-dimensional learning algorithms in both learning frameworks, i.e. in the supervised and semi-supervised learning frameworks. In order to do so, for each table, we compare the 20 results obtained by each uni-dimensional algorithm with the ones obtained by its multi-dimensional generalisation, i.e. nB with MDnB, TAN with MDTAN and so on. This comparison is made per columns (in Tables 1 and 2) by means of the Wilcoxon signed-rank test with $\alpha = 0.05$, a non-parametric statistical hypothesis test. The null hypothesis is that "Both classifiers has the same distribution, i.e. there is not statistical difference in the behaviour of both learning algorithms". The use of this non-parametric test is justified: the Kolmogorov-Smirnov test ($\alpha = 0.05$) reject the Gaussian assumption of the results.

The results of the Wilcoxon test can be found in Table 3. The first 5 rows correspond to the supervised framework (Table 1) and the last 5 to the semi-supervised (Table 2). Per row in Table 3, the learning algorithms involved in the comparison, the means (per column) of both algorithms, the p-value of the null hypothesis and the result of the Wilcoxon test (if the null hypothesis is accepted or rejected) are shown. Moreover, the greatest mean per row is highlighted in **bold**.

Framework	Comparison	Uni-dimensional Mean	Multi-dimensional Mean	p-value	H_0
Supervised	nB vs MDnB	51.72	**53.65**	0.34	Accepted
	TAN vs MDTAN	**53.35**	51.69	0.15	Accepted
	TAN vs MD1/1	53.35	**54.35**	0.10	Accepted
	2DB vs MD2/2	49.86	**53.53**	> 0.01	**Rejected**
	3DB vs MD2/3	47.82	**53.07**	> 0.01	**Rejected**
Semi-supervised	nB vs MDnB	29.44	**52.02**	> 0.01	**Rejected**
	TAN vs MDTAN	38.27	**38.99**	0.37	Accepted
	TAN vs MD1/1	38.27	**56.29**	> 0.01	**Rejected**
	2DB vs MD2/2	29.53	**56.74**	> 0.01	**Rejected**
	3DB vs MD2/3	31.54	**56.48**	> 0.01	**Rejected**

Table 3: Results of the Wilcoxon signed-rank test ($\alpha = 0.05$) of comparing the results in the 20 datasets of each uni-dimensional algorithm with its generalisation to the multi-dimensional framework.

From Table 3, the following comments can be extracted: In most of the cases, the multi-dimensional approach obtains the best results in terms of mean accuracy. Although there are small differences between the uni-dimensional and the multi-dimensional approaches in the supervised framework (only the MD 2/3 and MD 2/3 report statistical differences), in the semi-supervised these differences grow larger (except for the case of the MDTAN learning algorithm, the rest of the multi-dimensional approaches report statistical differences). In this framework, the multi-dimensional approaches lead to better results, while performance degradation occurs in the uni-dimensional ones, where the mean accuracies drop dramatically. A reason to explain this could be that the uni-dimensional approaches cannot match the underlying multi-dimensional nature of the generative structures. Moreover, the good results obtained by the MD J/K highlight the flexibility of this kind of learning algorithms to capture different types of structures.

4.2 Learning the generative structure

We are concerned about the fact that there are situations in which the addition of unlabelled data causes degradation of the performance of the classifier [3], in contrast to the improvement of performance when adding unlabelled data, as happens with the uni-dimensional approaches in the previous section.

Many researchers, in order to prevent these situations, have proposed in the literature certain assumptions [2] [14] that must be held when learning in a semi-supervised framework. One of the most important assumptions is the hypothesis presented in [3] which states that “If the correct structure of the generative model is obtained, unlabelled data improve the classifier, otherwise, unlabelled data can actually degrade performance”. So therefore, we need to test if this hypothesis is verified in the proposed multi-dimensional domains. Hence, we fix, per each dataset, the structure that generates the data and learn the parameters of the model in the supervised (MLE) and semi-supervised (using the EM algorithm as defined in [6]) frameworks.

The results of this learning process in the 20 databases are shown in Table 4. The “LABELLED” column shows the accuracies obtained in all the datasets by using just supervised learning whilst the “ALL DATA” column shows the accuracies of the semi-supervised learning process. The column “$(1 - e_B)$” shows the optimal Bayes accuracy (the opposite to the Bayes error), measured as a percentage.

Based on these results, we can conclude that, as happens in the uni-dimensional framework [3], when using the real structure in the multi-dimensional framework, the unlabelled data always helps.

Family	Dataset	$(1 - e_B)$	LABELLED	ALL DATA	Helps?
MDnB	NB01	*93.93*	88.48	92.16	Yes
	NB02	*93.25*	89.90	91.06	Yes
	NB03	*91.65*	88.29	89.09	Yes
	NB04	*74.36*	65.57	67.33	Yes
	NB05	*66.45*	59.44	60.43	Yes
MDTAN	TA01	*66.86*	66.69	66.79	Yes
	TA02	*47.61*	43.56	43.80	Yes
	TA03	*56.80*	50.53	50.89	Yes
	TA04	*64.21*	58.57	60.77	Yes
	TA05	*58.75*	49.99	50.71	Yes
MD2/2	2201	*66.26*	62.04	63.86	Yes
	2202	*65.44*	60.10	60.77	Yes
	2203	*61.72*	50.22	51.24	Yes
	2204	*84.85*	83.46	83.65	Yes
	2205	*50.28*	47.82	47.82	Equal
MD2/3	2301	*57.56*	53.32	54.30	Yes
	2302	*69.11*	67.26	67.62	Yes
	2303	*50.31*	43.45	45.22	No
	2304	*78.45*	75.36	75.78	Yes
	2305	*61.34*	57.34	57.40	Yes

Table 4: Accuracies obtained by supplying the EM Algorithm (as it is usually used in semi-supervised learning) with the structure that generates the data.

4.3 Reaching the generative structure

In almost all problems that we face in the machine learning field, there is no clue of the generative structure of the dataset. For that reason, we want to check if we can reach the generative structure and, therefore, obtain better results in terms of accuracy using the specific semi-supervised learning algorithms in each multi-dimensional Bayesian network family.

In order to do so, we check the 5 results obtained by each multi-dimensional algorithm in the family where it can lead to the generative structured, i.e. the MDnB learning algorithm in the MDnB structure family, etc. To answer the main question of this section, we compare the 5 results obtained in the supervised framework (Table 1) with the results obtained in the semi-supervised framework (Table 2). This comparison is made with a Wilcoxon signed-rank test with $\alpha = 0.05$ (the Kolmogorov-Smirnov test ($\alpha = 0.05$) reject the Gaussian assumption).

Table 5 sums up the results of the statistical test. Per row, the family of multi-dimensional Bayesian network in which the learning algorithm is applied, the learning algorithm used, the means (over just 5 results) of the learning algorithm in both supervised (Table 1) and semi-supervised frameworks (Table

2), the p-value of the null hypothesis and the result of the Wilcoxon test are shown. Moreover, the greatest mean per row is highlighted in **bold**.

Famiiy	Algorithm	Supervised Mean	Semi-supervised Mean	p-value	H_0
MDnB	MDnB	78.34	**80.07**	0.02	**Rejected**
MDTAN	MDTAN	**43.13**	40.13	0.11	Accepted
	MD 1/1	47.47	**49.16**	0.02	**Rejected**
MD 2/2	MD 2/2	54.02	**55.64**	0.11	Accepted
MD 2/3	MD 2/3	52.40	**52.93**	0.34	Accepted

Table 5: Results of the Wilcoxon signed-rank test ($\alpha = 0.05$) of comparing the results of both supervised and semi-supervised frameworks using the multi-dimensional algorithms that can lead to the generative structure in the five datasets of each family.

From this results, the following comments can be made:

- With the exception of the MDTAN learning algorithm, in the semi-supervised framework the learning algorithms lead to better results.
- Although it is difficult to obtain significant differences with only 5 results per each framework, in two cases (MDnB and MD 1/1) the differences are significant. So, we can claim that there is a tendency to better results in the semi-supervised framework when the used algorithm can lead to the generative structure.
- From the results obtained in this section and Section 4.1, it seems that the MDTAN algorithm proposed in [13] leads to very suboptimal solutions.

4.4 Supervised and semi-supervised learning frameworks comparison

Once tested the algorithms that can reach the generative structure, a wider comparison has to be made. In this section, we compare the behaviour of each learning algorithm (both uni-dimensional and multi-dimensional ones) in both supervised and semi-supervised scenarios.

To achieve this mission, we compare by means of a Wilcoxon signed-rank test ($\alpha = 0.05$) the results obtained in all the datasets in both supervised (Table 1) and semi-supervised learning (Table 2) frameworks, i.e. comparing the accuracies per columns, e.g. the nB column in Table 1 with the nB column in Table 2, etc.

Scenario	Algorithm	Supervised Mean	Semi-supervised Mean	p-value	H_0
Uni-dimensional	nB	$\mathbf{51.72}$	29.43	> 0.01	**Rejected**
	TAN	$\mathbf{53.35}$	38.27	> 0.01	**Rejected**
	2DB	$\mathbf{49.86}$	29.54	> 0.01	**Rejected**
	3DB	$\mathbf{47.83}$	31.54	> 0.01	**Rejected**
Multi-dimensional	MDnB	$\mathbf{53.65}$	52.02	> 0.01	**Rejected**
	MDTAN	$\mathbf{51.69}$	39.00	> 0.01	**Rejected**
	MD1/1	54.35	$\mathbf{56.29}$	0.01	**Rejected**
	MD2/2	53.53	$\mathbf{56.74}$	> 0.01	**Rejected**
	MD2/3	53.07	$\mathbf{56.48}$	> 0.01	**Rejected**

Table 6: Results of the Wilcoxon signed-rank test ($\alpha = 0.05$) of comparing the 20 results of each learning algorithm in both supervised and semi-supervised frameworks

Table 6 shows the results of the statistical test. Per row, the learning algorithm, the means of the learning algorithm in both supervised (Table 1) and semi-supervised frameworks (Table 2), the p-value of the null hypothesis and the result of the Wilcoxon test are shown. Moreover, the greatest mean per row is highlighted in **bold**.

From the results, the following conclusions are extracted:

- All the learning algorithms behave different in both learning frameworks.

- In the uni-dimensional approaches, performance degradation occurs in the semi-supervised framework. This is probably because "If the correct structure of the generative model is obtained, unlabelled data improve the classifier, otherwise, unlabelled data can actually degrade performance" [3].

- With respect to the MD J/K learning algorithms, an improvement in terms of accuracy in the semi-supervised framework is observed. As stated before, this highlights the flexibility of this kind of learning algorithms to capture different types of complex structures.

- The MDnB learning algorithm reports significant performance degradation in the semi-supervised, while in the previous section, it reports significant improvement when learning MDnB structures. It denotes that the MDnB learning algorithm is very specific, it obtains very good results while dealing with problems with an underlying MDnB structure, but when the generative models are a bit complex, its rigid structure makes the algorithm to lead to very suboptimal solutions.

- In this comparison, the MDTAN algorithm also shows very poor performances in the semi-supervised framework. In addition, from the numerical results (means), it seems to be closer to the uni-dimensional approaches rather than to the multi-dimensional ones.

4.5 Behaviour of the learning algorithms in the semi-supervised framework

After showing the potential of semi-supervised learning, we are going to check whether statistical differences exist among the semi-supervised classifiers: not only between the multi-dimensional approaches, but also among the uni-dimensional ones. Specifically, we use Friedman test [7] with a Shaffer's static post-hoc test with $\alpha = 0.1$ as recommended in [9]. The test results are represented by means of critical difference diagrams [7], which show the mean ranks of each algorithm across all the domains in a numbered line. If there is no statistically significant difference between two algorithms, they are connected in the diagram by a straight line.

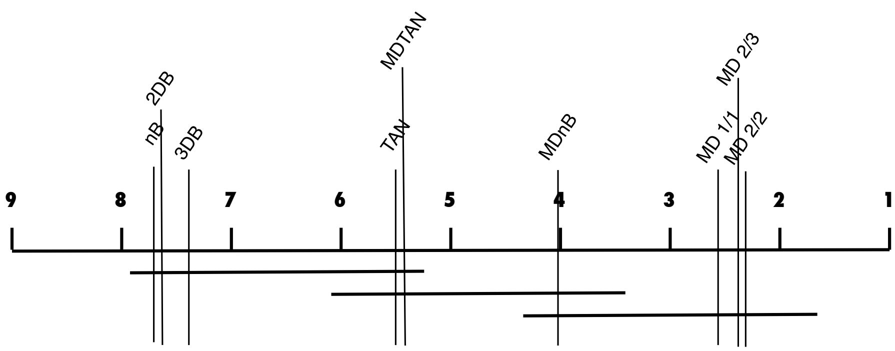

Figure 1: Accuracy ranking using both labelled and unlabelled data for the different algorithms on the 20 artificial datasets, $\alpha = 0.05$.

From the critical difference diagram (see Figure 1), we confirm the sensations extracted from the previous sections and deduce that, in the semi-supervised framework, clearly the multi-dimensional classifiers outperform the uni-dimensional techniques, with the exception of the MDTAN classifier.

5. General conclusions

From the experimental results shown in this report, the following major conclusions, that answer the experimental questions, can be extracted:

1. As happens in the uni-dimensional framework [3], when using the real structure to semi-supervisely learnt multi-dimensional classifiers, the unlabelled data always helps.

2. There is a tendency to achieve better classifiers in terms of joint accuracy in the semi-supervised framework when the used multi-dimensional algorithm can reach to the generative structure.

3. In the uni-dimensional approaches, performance degradation occurs in the semi-supervised framework. This is probably due to the fact that the uni-dimensional approaches are not able to match the actual multi-dimensional structure of the problems.

4. Although there are small differences between the uni-dimensional and the multi-dimensional approaches in the supervised framework (only the MD 2/2 and MD 2/3 report statistical differences), in the semi-supervised these differences grow larger (except for the case of the MDTAN learning algorithm, the rest of the multi-dimensional approaches report statistical differences).

5. In the semi-supervised framework, clearly the multi-dimensional classifiers outperform the uni-dimensional techniques, with the exception of the MDTAN classifier.

6. The MDnB learning algorithm [13] is very specific, it obtains very good results when dealing with problems with an underlying MDnB structure, but when the generative models are more complex, its rigid structure makes the algorithm to lead to very suboptimal solutions.

7. The MDTAN algorithm [13] also shows very poor performances in the semi-supervised framework.

8. The MD J/K learning algorithms have great flexibility to capture different types of complex structures that results in an improvement in terms of joint accuracy in the semi-supervised framework.

References

[1] B. Calvo and J. L. Flores. Iclab intelligent computing laboratory: A library of data mining and optimization algorithms, 2009. `http://iclab.sourceforge.net/`.

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

[3] I. Cohen. *Semisupervised Learning of Classifiers with Application to Human-Computer Interaction*. PhD thesis, University of Illinois at Urbana-Champaign, 2003.

[4] G. Cooper and E. Herskovits. A bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.

[5] F. Cozman and I. Cohen. Risk of semi-supervised learning. In O. Chapelle, B. Scholkopf, and A. Zien, editors, *Semi-supervised learning*. The MIT Press, 2006.

[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] J. Demsar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7:1–30, 2006.

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

[9] S. García and F. Herrera. An extension on "statistical comparisons of classifiers over multiple data sets". *Journal of Machine Learning Research*, 9:2677–2694, 2008.

[10] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, 1992.

[11] J. D. Rodriguez, A. Perez, and Lozano J. A. Sensitivity analysis of k-fold cross-validation in prediction error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(3):569–574, 2010.

[12] 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.

[13] 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.

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