Barycentric Coordinates for Ordinal Sentiment Classification

Brian Keith
Departamento de Ingeniería de Sistemas y Computación,
Universidad Católica del Norte
Antofagasta, Chile
brian.keith@ucn.cl

Claudio Meneses
Departamento de Ingeniería de Sistemas y Computación,
Universidad Católica del Norte
Antofagasta, Chile
cmeneses@ucn.cl

ABSTRACT

Sentiment analysis and opinion mining is an area that has experienced considerable growth over the last decade. This area of research attempts to determine the feelings, opinions, emotions, among other things, of people on something or someone. To do this, techniques from natural language processing and machine learning algorithms are mainly used. This article discusses the problem of determining the polarity of reviews using a novel ordinal classification technique called Barycentric Coordinates for Ordinal Classification (BCOC). The aim of this analysis is to explore the viability of application of BCOC on the field of sentiment analysis. This new method is based on the hypothesis that the ordinal classes can be represented geometrically inside a convex polygon on the real plane by using barycentric coordinates. A set of experiments were conducted to evaluate the capability and performance of the proposed approach relative to a baseline, using accuracy as the general measure of performance. The experiments include testing on generic ordinal classification data sets and on multi-class sentiment analysis data sets. In general the method is competitive with the baseline. The results show no significant difference over the baseline in the case of generic ordinal classification and sentiment analysis with three classes. However, in the case of sentiment analysis with four classes the results show improvements in the overall accuracy.

CCS CONCEPTS

•Computing methodologies →Supervised learning by classification; Machine learning algorithms; •Information systems →Data mining; *Sentiment analysis;*

KEYWORDS

Opinion Mining, Ordinal Classification, Sentiment Analysis, Barycentric Coordinates

1 INTRODUCTION

Opinions are central to almost all human activities due to their crucial influence on people's behavior. Every time that there is a need to make a decision, human beings seek to know other people's opinions. In the real world, companies and organizations want to know the public's opinion about their products and services. Moreover, shoppers want to know what other customers think about a product before purchasing it. In the past, people turned to their friends and family for opinions, whereas companies relied on surveys or focus groups. However, nowadays the explosive growth of social media and the increase in the available sources of data has made individuals and organizations use the information provided by these to support their decision-making process. The field of sentiment analysis, also known as opinion mining [10] has been developing in this context.

One of the main tasks in sentiment analysis is determining the polarity. This can be seen as a classification problem in general, as most of the literature does. The most common approaches to deal with this task are support vector machines and the naive Bayes classifier [15]. A literature review shows that most works focus on binary classification between positive and negative opinions; the multi-class case has not been exhaustively studied [15].

Most works on determining polarity use a binary classification approach [15]. Other studies that use multi-class classification with five levels apply regression methods and then a transformation into the corresponding class [10]. However, recent works such as [18] illustrate that multi-class classification of polarity at the full document level remains elusive, even when using the deep learning approach proposed in that article.

In most cases, determining the polarity can be seen as a task of ordinal classification (in the multi-class case, because in the binary case order has no relevance), where the ordering of the different classes corresponds to the natural order provided by their different labels (i.e., the order of the classes is the following: very negative, negative, neutral, positive and very positive). In spite of the ordinal nature of the problem and that ordinal classification methods have been widely studied, upon reviewing the literature it can be noted that these characteristics of the data have not yet been exploited exhaustively to obtain better classifiers.

The term ordinal classification makes reference to the supervised learning problem of classification where classes have a natural order imposed on them due to the characteristics of the concept studied. When the problem has in fact an ordinal nature, it would be expected that this order would also be present in the input space [18]. In contrast to nominal classification, there is an ordinal relationship among the categories and it is different from regression in that the possible number of ranks is finite and the exact differences between each rank are not defined. In this way, ordinal classification lies somewhere between nominal classification and regression [16].

2 RELATED WORK

The effective development of opinion mining systems has a great number of challenges. First, it is necessary to identify the contents in a text. This is a non-trivial task due to the nature of language, which contains countless semantic nuances that are not present in other types of data. Second, sentiments must be classified in some way and thus determine their orientation. There are diverse ways of approaching this issue [12].

Most of the methods in the field of sentiment analysis correspond to techniques for nominal data, namely, data in which the class labels belong to a set with no natural order. In contrast to this approach, the problem can be tackled with ordinal classification methods (sometimes called ordinal regression), which lies in a middle ground between classic classification and regression [20].

In a problem of authentic ordinal nature, this order is also expected to be present in the input space [7]. Defining a space where the ordinal nature of the data is evident could prove useful for the functioning of a classifier. The hypothesis behind the proposed method is based on the intuition that exploiting the ordinal nature of the classes should bring about a positive effect in a classifier's performance, in particular in problems of multi-class classification.

In the literature there can be found three approaches to deal with the problems of ordinal classification [9]:

(1) Working with the ordinal scale as if it were an ordinary multi-class classification problem.
(2) Using regression to estimate a continuous value and then discretize it.
(3) Using specialized algorithms for ordinal classification.

The first approach does not take advantage of the inner structure of the data, as it ignores the existing natural order of the classes. Although there are several examples of methods that work successfully following this approach, it would be expected that taking into account the information of the order followed by each class would lead to more accurate classifications than those obtained through traditional methods that do not exploit this structure.

The problem of the second approach is that it is an *ad hoc* solution to an ordinal classification problem, since regression models have been thought for continuous data. Although it is possible to convert the ordinal data into continuous data and then execute a post-processing step to obtain the classes, this does not assure that the ordinality of the classes of the problem is being taken into consideration.

The third approach takes advantage of the structure of ordinal classification. This can be done through the use of machine learning algorithms modified to exploit this structure. However, some of them present some complexities in terms of implementation and training. There are other simpler approaches that offer promising results without incurring greater computational complexity. These consist of applying decompositions of the problem in a specific way, modifying the objective function during the training of the methods or using a threshold based model [6].

Finally, it is worth noting that the difference between ordinal and nominal classification is not remarkable in the case of binary classification, due to the fact that there is always an implicit order in "positive class" and "negative class". Since most works of determination of polarity are still centered on the binary case, it is natural that ordinal classification methods have not been explored thoroughly. The previous discussion is summarized in the diagram of Figure 1.

Approaches of ordinal classification can be applied complementarily with nominal classification models to obtain improvements (whether by the separation of the problem into sub-problems or by modifying the objective function to be optimized). Regarding the approaches that use specialized machine learning algorithms, these are more complex and require non-trivial changes in the training methods.

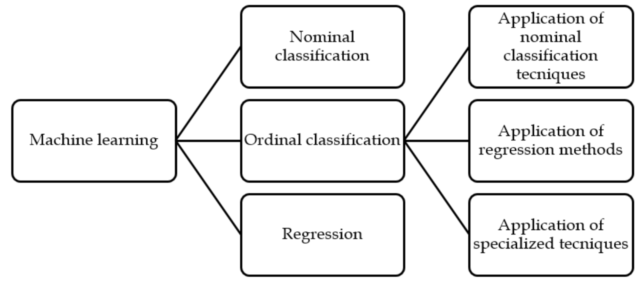

Figure 1: Ordinal classification techniques.

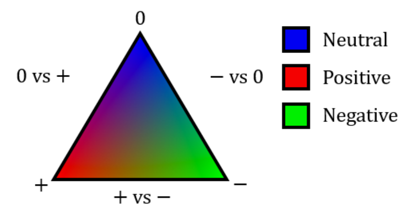

Figure 2: Expected classification regions for BCOC.

In the field of sentiment analysis, the strategies used penalize in function of the distance of the assigned class with respect to the true class during training. The quality of the model will depend on the distance function used [11].

Regarding ordinal classification in general, there are several methods to accomplish this task. Some can be found in the work of Sánchez-Monedero [17]. Among the utilized methods, one of the competitive approaches is the use of support vector machines (SVMs). Although they do not always present the best performance, they obtain good results in general and in some cases better than other approaches applied.

3 PROPOSED METHOD

In this section the proposed method to solve the problem of ordinal classification for sentiment analysis is described. The central idea of this method arises from the triangular representation used by SentiWordNet 3.0 to model the different terms [2]. This representation is generalized and adapted to form the classifier.

Ideally, the model would learn similar classification regions to the ones in the diagram shown in Figure 2 (an exemplification of the three-class case). It should be noted that the neutral region stretches over two zones. First, in the upper part of the triangle, where texts that are neutral and objective are ideally represented. Second, in the lower central part that ideally represents subjective but neutral texts, that is, those that express opinion but do not display a strong polarity.

It is proposed to use a barycentric coordinates system [5] or other similar variation not only as a visualization tool, but also as input to carry out the classification. It should be noted that the coordinate transformation would be a simple mathematical function. Intuitively, this barycentric coordinates representation is much closer to the structure of opinions. In a certain way, this is implicit in the representation used by SentiWordNet 3.0. The use of

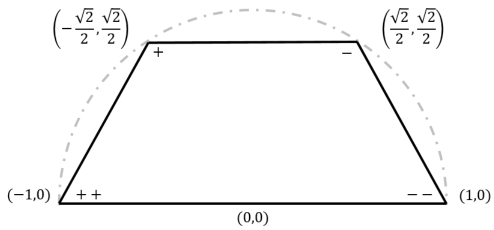

Figure 3: Representation of four ordinal classes.

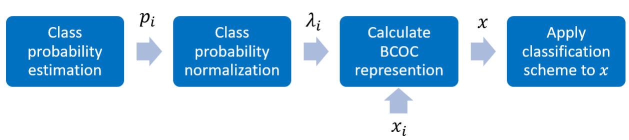

Figure 4: Process flow of the BCOC method.

this representation would situate the documents within the space enclosed by the equilateral triangle shown in Figure 2.

It should be noted that in the said equilateral triangle the classes are situated at equal distance from each other. This contradicts the ordinal nature of the problem, therefore a representation must be constructed so that the separation between the most distant classes (according to their natural order) is reflected geometrically. In particular, in the three-class case, an isosceles triangle with a greater separation between positive class and negative class must be used.

Based on the previous discussion, the proposal made in this work corresponds to a generalization of this representation to a problem with n classes, taking as a base the case of $n = 3$. While the idea originates from the representation of terms in a sentiment analysis lexicon, the notion is general enough to be used in other classification contexts, given that the labels have an ordinal structure.

3.1 Barycentric Coordinates for Ordinal Classification (BCOC)

The proposed method is based on the use of the vertices of a convex polygon inscribed in a semicircumference to represent the different classes. An example of this concept can be seen in Figure 3. Note that the order of the classes is preserved by the relation of distance between the points that represent them. It is clear that the class "--" is more distant from "++" than from "-" or "+". This is easily confirmed by the geometric intuition of the representation and it is easy to demonstrate applying the triangle inequality from classical linear algebra and elementary geometry.

The general process flow of the BCOC method is represented in Figure 4 and is explained in detail in the next paragraphs.

Specifically, barycentric coordinates will be used. The method is called BCOC and it is based on the construction of n classifiers using a one-vs-all approach. Using these classifiers, an estimate of the probability of belonging to each class must be obtained. The results of these classifiers are transformed into a geometric representation and are then classified using a new classifier that is trained in function of the new representation[1].

Each class $C_i (i = 1, \ldots, n)$ is associated with a point x_i which corresponds to a vertex in a convex polygon of n sides (as seen in the example of Figure 3), then the system of barycentric coordinates is utilized to obtain the position of the example (taking into account the result of all classifiers) within the polygon. Each point x_i can be determined using the formula in Equation 1. Which generates the vertices of a n-sided polygon inscribed in a semicircumference. The point $(1, 0)$ is associated with the lowest class and the point $(-1, 0)$ with the highest class with respect to the order of the labels.

$$x_i = \left(\cos\left[\frac{\pi(i-1)}{n-1}\right], \sin\left[\frac{\pi(i-1)}{n-1}\right]\right), \quad i = 1, \ldots, n \tag{1}$$

Given an example from the data set, a probability p_i is generated for each class i from the lower level classifiers. This probability represents the chance that the current instance belongs to class i (obtained in a one-vs-all fashion). Due to the fact that each probability is obtained using independent classifiers, the sum of the obtained probabilities will not necessarily add up to 1. However, the definition of barycentric coordinates requires that the sum of the ponderations of each point add up to 1 [5]. Therefore, it is necessary to normalize the resulting probabilities using Equation 2.

$$\lambda_i = \frac{p_i}{\sum_{j=0}^{n} p_j}, \quad i = 1, \ldots, n \tag{2}$$

This normalized coefficient is denoted by λ_i and represents the weight of the class i for the current instance. These coefficients in turn correspond with the barycentric coordinates themselves and can be turned directly into a point in the plane inside the convex polygon. According to the generalized version of barycentric coordinates, the formula to obtain the final representation of each point inside a convex polygon is shown in Equation 3.

$$x = \sum_{i=0}^{n} \lambda_i \cdot x_i = \frac{\sum_{i=0}^{n} p_i \cdot x_i}{\sum_{i=0}^{n} p_i} \tag{3}$$

The point x is finally fed into an additional classifier that is in charge of determining the class. In the exceptional case when $\sum_{i=0}^{n} p_i = 0$, in order to avoid division by zero, x is assigned to the centroid of the polygon. It should be noted that this is equivalent to the case where all $p_i = 1$, but with the assigned classes reversed, in both cases the result will be assigned to the centroid.

The training required before applying BCOC can be summarized as follows:

(1) For each class train a one-vs-all classifier that can provide probability estimates.
(2) For each instance in the data set estimate the class probabilities using the previous classifiers.
(3) For each instance in the data set normalize the class probabilities according to Equation 2.
(4) For each instance obtain the new representation in the BCOC two-dimensional space using Equation 3.

[1] The source code for the base class of this method can be found at: http://mii.ucn.cl/files/8214/8712/7619/BCOC.py. The code provides a skeleton that can be easily modified.

(5) Train a multi-class classifier using the new feature space as input.

The application of the BCOC method on a test data set is similar:

(1) For each instance in the data set estimate the class probabilities using the previously trained one-vs-all classifiers.
(2) For each instance in the data set normalize the class probabilities according to Equation 2.
(3) For each instance obtain the new representation in the BCOC two-dimensional space using Equation 3.
(4) Apply the final multi-class classifier on the new feature space.

This completes the description of the BCOC method.

3.2 Observations

The intuition behind the proposed method is that this representation allows directly taking advantage of the ordinal nature of the classes. Moreover, the use of multiple classifiers allows obtaining a more accurate classification, because the construction of different classifiers would allow increasing noise tolerance by having several independent estimators.

Note that the type of classifier that must be used in each level has not been indicated. This provides a flexible two-level architecture for classifying ordinal data and allows the combination of different classifiers. The proposed method only indicates a general architecture and the way of combining the classifiers, the only requirements are that the upper level classifier must be capable of handling multi-class labels and that the lower level classifiers must provide an estimate of the probability of belonging to each class.

It is important to note that the BCOC method works by creating a new two-dimensional feature space. Once this space is constructed, the data is projected onto it using the described formulas. Given this, the method may be regarded as a supervised dimensionality reduction technique for ordinal data. The n-dimensional input space is always reduced to a two-dimensional space.

Figure 5 shows how some examples would be located in the new system of coordinates and using the regions of Figure 2. The final point x is represented by the small dot, while the probabilities generated for the other classifiers are represented by the crosses on each line segment. From this representation, it can be inferred that the classification of each example would be (approximately): "neutral (objective)", "negative", "neutral (subjective)" and "neutral (subjective)", respectively.

Finally, while the election of the points in Equation 1 to represent the classes may seem arbitrary, the current setup has the advantage that it leverages the ordinal structure of the data through the application of basic geometry (the triangle inequality). In principle, any polygon could be used to build this representation and it might even be reasonable to use a polyhedron or a higher dimensional structure. However, the interpretation of these more complex spaces would be considerably difficult.

4 METHODOLOGY

The present section describes the data sets used, the evaluation metrics utilized and the preprocessing of the data sets. Also, the parameters of the evaluated models are detailed.

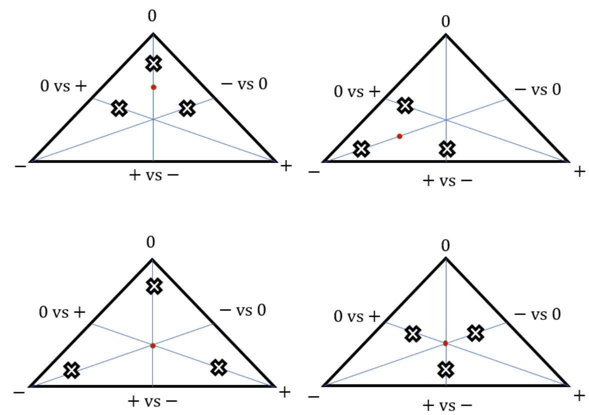

Figure 5: Examples of the application of BCOC.

Table 1: The ordinal data sets used in the experiments (N = number of examples, K = number of classes).

Data set	N	K	Distribution of the ordinal classes
newthyroid	215	3	(30, 150, 35)
eucaliptus	736	5	(180, 107, 130, 214, 105)
bondrate	57	5	(6, 33, 12, 5, 1)
winequality-red	1599	6	(10, 53, 681, 638, 199, 18)

First, in order to evaluate the performance in the task of ordinal classification, a selection of the data sets used by Sánchez-Monedero [17] in his work on ordinal classification is used. These sets are publicly available at the UCI [1] and at the repositories of mldata.org [19]. Details are shown in Table 1.

Secondly, the method is evaluated in the task of determining sentiment polarity on a data set of film reviews. The data utilized corresponds to the data set developed by Pang & Lee [11] that contains film reviews labeled according to the positivity and negativity of the review. Specifically, the data set has a version with three labels (1, 2, 3) and another with four labels (1, 2, 3, 4) to differentiate the possible intensity and polarity of the opinion. The original data set also considers authors as a factor, but for the sake of simplicity the film reviews of the four authors have been grouped in one data set. Differences among authors of each review are not considered.

With the purpose of evaluating the performance on other domains of application of opinion mining and not only film reviews, a selection from the data set of Blitzer et al. [4] has been used. This multi-domain data set contains different kinds of reviews, the selection of the domains used is shown in Table 2. All the data sets have four classes (very negative, negative, positive and very positive). Note that there is no neutral class in this data set.

On the other hand, the method has been evaluated on a data set of reviews in Spanish and English of scientific articles in the context of an international conference [8]. The data set contains a total of 405 reviews, reviews written in English (17 instances) and empty reviews (6 instances) were discarded, leaving a total of

Table 2: Selection of multi-domain data sets (sentiment analysis, $K = 4$)

domain	Size (kb)	Number of reviews
apparel	7098	9246
automotive	654	736
video	57192	36180

$N = 382$ reviews [2]. The instances are evaluated according to two scales.

- The first one expresses the perception of the review (i.e. how positive or negative the review is perceived by the reader). This scale is called "orientation", making reference to the semantic orientation of the opinion.
- The second one expresses the evaluation of the article emitted by the reviewer. So this scale is called "evaluation".

The method has been evaluated in the binary case (both orientation or evaluation with positive or negative value) and the ternary case (both orientation or evaluation with positive, neutral or negative value). For this data set, the binary case was evaluated in order to ascertain the fact that in binary classification the ordinal structure of the data should make no major difference in the results.

It is necessary to pre-process the data sets for sentiment analysis before directly using the algorithm. Due to the fact that the data sets correspond to the texts of the reviews and the associated classes, it is necessary to obtain a vectorial representation of the text. In order to accomplish this, the input is first tokenized to obtain the words that compose it. Afterwards, a stopwords filter [3] is applied without eliminating some of the important words for this particular domain, such as "no". Depending on the language, the stopword dictionary is either in English or Spanish (only for the paper reviews data set). Then, stemming is applied by means of Porter's algorithm [14] to simplify the terms to their roots and thus reduce the final dimensionality of the data. After finishing the pre-processing, a representation is obtained using TF-IDF. The final representation is obtained by applying LSA [10] (using the $n = 100$ most important components).

Regarding the evaluation of the methods, a 10-fold cross validation approach is utilized. In particular, the overall accuracy obtained in each of the classifiers generated by the 10-fold cross validation is used as main metric. Even though accuracy is a simple metric and does not take into account all the aspects of the classification, it allows to obtain a useful estimate to evaluate the performance attained. Also, the corresponding standard deviation and the best accuracy achieved are reported in each case.

The implementation of the methods was carried out in Python using the sklearn library [13]. The Nave Bayes and SVM methods are used as a comparison baseline. In spite of the fact that these algorithms are not specialized in ordinal classification, their results on ordinal data sets are competitive with those obtained using more specialized methods. These two methods have been selected because of their wide use in the literature of sentiment analysis [15], which would correspond finally to the main application of the proposed BCOC method.

[2]The data set used in this study can be found at http://mii.ucn.cl/files/2814/8570/2080/reviews.json

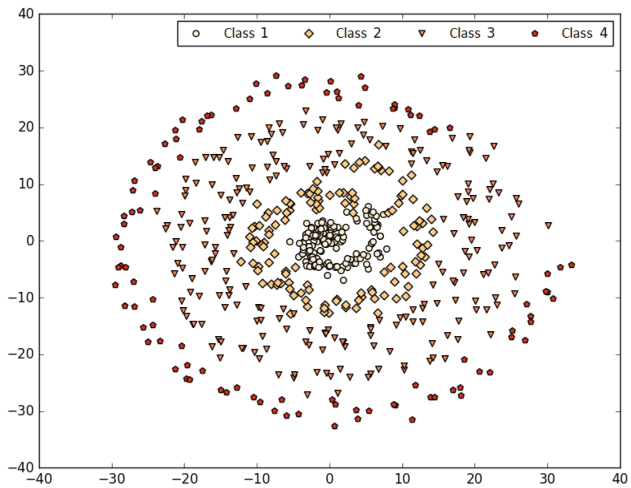

Figure 6: Non-linear synthetic data set.

Regarding the BCOC method, two arquitectures are used: NB-SVM and SVM-SVM. The first uses Bayes simple classifiers in the inferior level and a SVM with linear kernel in the superior level. The second will use SVM with Gaussian/linear kernel (depending on which one yields better performance) in the inferior level and a SVM with linear kernel in the superior level. The inferior level classifiers are parameterized in the same way as individual classifiers used in the baseline. The superior level classifier is parameterized with the default values of the sklearn library.

It is important to note that while SVM in its basic form is neither probabilistic nor multi-class. However, with appropriate modifications it can be used for multi-class problems and to obtain probability estimates. To do this, the built-in implementations from the sklearn library are used.

5 RESULTS AND DISCUSSION

5.1 Synthetic Data Set

A visualization of the synthetic data set [17] used is shown in Figure 6. The set has four classes and as it can be seen, it presents a non-linear structure.

In Table 3 the classification results for the non-linear synthetic data set are shown. The results obtained by NB and SVM for the same data set are taken as the baseline.

It is important to note the increase in overall accuracy in the case of BCOC-NB with respect to NB without applying BCOC. Although it does not exceed the results of applying SVM, this indicates that BCOC could improve the classification performance under certain conditions.

Based on the results obtained on the synthetic data set, more experiments were carried out with real ordinal data sets to determine whether the proposal presents an improvement in general.

Table 3: Results for the non-linear synthetic data set.

Accuracy (%) for ordinal classification			
Data set	Method	Average (+- SD)	Best
non-linear-synthetic	NB	67.70+-6.05	79.66
	BCOC-NB	90.77+-3.62	95.00
	SVM	96.31+-3.03	100.00
	BCOC-SVM	96.49+-2.87	100.00

Table 4: Results for ordinal classification data sets. The best result for each case is highlighted in boldface.

Accuracy (%) for ordinal classification			
Data set	Method	Average (+- SD)	Best
newthyroid	NB	96.25+-5.84	100.00
	BCOC-NB	95.86+-4.53	100.00
	SVM	95.77+-3.51	100.00
	BCOC-SVM	96.26+-4.36	100.00
eucalyptus	NB	37.65+-4.33	43.84
	BCOC-NB	38.43+-6.67	51.35
	SVM	55.18+-7.36	68.49
	BCOC-SVM	56.26+-8.17	71.23
bondrate	NB	52.66+-23.29	83.33
	BCOC-NB	58.00+-19.76	100.00
	SVM	57.67+-14.66	83.33
	BCOC-SVM	58.00+-15.49	83.33
winequality-red	NB	54.10+-3.83	59.36
	BCOC-NB	55.65+-3.75	64.37
	SVM	60.46+-4.81	66.88
	BCOC-SVM	60.04+-3.21	64.77

5.2 General Ordinal Classification

The results obtained are presented in this section. Table 4 shows the results of classification for each data set.

The results obtained for each data set are briefly discussed:

- Note that in the three-class case (newthyroid) there is no significant difference between any of the methods, reaching even a 100% accuracy in the best case on the test set.
- In the case of bondarate there are five classes. No significant difference can be observed. While in terms of average and best performance one could speak of an improvement, the high standard deviation renders these results not significant.
- In the case of winequality-red there are 6 classes. Although there is a slight improvement in the case of NB, in general no significant improvement with respect to the baseline can be observed.

Based on these results, it can be observed that although the proposed method is competitive, it does not offer a significant improvement on the non-synthetic ordinal data sets.

Table 5: Results for the data set of Pang & Lee. The best result for each case is highlighted in boldface.

Accuracy (%) for sentiment analysis in Pang & Lee data set			
Number of classes	Method	Average (+- SD)	Best
3 classes	NB	59.30+-2.13	63.00
	BCOC-NB	58.68+-2.91	63.87
	SVM	67.47+-2.76	70.86
	BCOC-SVM	67.28+-1.90	70.20
4 classes	NB	51.99+-2.04	54.29
	BCOC-NB	52.05+-3.09	56.29
	SVM	58.07+-2.40	61.00
	BCOC-SVM	58.68+-2.52	61.40

Table 6: Results for the data set of multi-domain sentiment analysis. The best result for each case is highlighted in boldface.

Accuracy (%) for sentiment analysis in multidomain data set			
Domain	Method	Average (+- SD)	Best
apparel	NB	67.33+-1.12	69.37
	BCOC-NB	66.65+-1.60	69.37
	SVM	67.17+-1.98	69.08
	BCOC-SVM	69.94+-1.11	71.86
automotive	NB	60.45+-5.04	68.92
	BCOC-NB	60.18+-6.34	71.62
	SVM	62.09+-5.25	72.97
	BCOC-SVM	63.32+-5.61	73.97
video	NB	60.89+-0.70	62.19
	BCOC-NB	61.62+-0.69	62.85
	SVM	66.16+-0.46	66.86
	BCOC-SVM	N/A	N/A

5.3 Sentiment Analysis

Results for 3 and 4 classes on the set of Pang & Lee [7] are shown in Table 5.

While in the case of 3 classes there is no significant difference observed for the proposed methods, it can be observed that the performance of these methods is superior in the 4 classes case. In particular, in the case of BCOC-SVM the difference in the 3 classes case is not significant. However, in the 4 classes case, the BCOC-SVM method shows a clear improvement in performance with respect to the baseline formed by NB and SVM. Table 6 shows the results obtained on a selection of domains from the data set of Blitzer et al. [4].

In the case of the "apparel" domain there is a statistically significant difference between the BCOC-SVM method and the SVM original method, with a *p-value* < 0.01. In the case of the "video" domain, BCOC-NB has a statistically significant difference with respect to the NB original method, with a *p-value* < 0.05. It has not been possible to test the BCOC-SVM method on the data set of reviews of videos. This is exclusively due to a problem of available

Table 7: Results for the paper reviews data set, the best result in each case is highlighted in boldface.

Accuracy (%) for sentiment analysis in paper reviews data set			
Data set	Method	Average (+- SD)	Best
Binary orientation	NB	67.62 +- 5.08	76.19
	BCOC-NB	67.50 +- 5.30	76.19
	SVM	70.24 +- 5.11	77.38
	BCOC-SVM	69.64 +- 5.67	77.38
Ternary orientation	NB	45.65 +- 3.44	50.43
	BCOC-NB	46.78 +- 5.78	56.52
	SVM	48.17 +- 5.47	55.65
	BCOC-SVM	49.31 +- 2.98	55.65
Binary evaluation	NB	56.08 +- 4.40	63.92
	BCOC-NB	56.02 +- 4.30	63.92
	SVM	67.22 +- 3.53	74.23
	BCOC-SVM	66.91 +- 3.45	73.20
Ternary evaluation	NB	46.09 +- 4.00	52.17
	BCOC-NB	42.34 +- 4.46	51.13
	SVM	55.82 +- 3.62	61.74
	BCOC-SVM	57.39 +- 3.04	60.87

computing power. In particular, because of the high quantity of data in the said data set and the number of SVMs that must be trained for each of the required iterations. As it has not been possible to test BCOC-SVM, it is not possible to determine if it performs better than its counterpart without applying BCOC.

Finally, the results obtained on the data set of paper reviews [8] are shown in Table 7. The binary case has been evaluated to illustrate that the BCOC method does not have a significant influence when applied on a task of non-multiclass classification, because there is no distinction between the nominal and ordinal case in the case of binary classification.

It can be seen that in the case of classifying orientation, the approach based on BCOC is competitive with the baseline both in the binary and ternary case. In the binary case, it is expected that there is no difference, in fact in the case of NB the results were practically identical. In the ternary case, a slight improvement is observed, although it is not statistically significant. In the case of classifying evaluation, the results are similar, even though it should be noted that the classification of evaluation is a more difficult task than classifying orientation, because the semantic orientation of the text does not always coincide with evaluation [8].

Note that, as it was expected, the method does not yield accuracy improvements in the case of binary classification. This is due to the fact that the method has been specifically designed for multi-class ordinal classification, which in the binary case is indistinguishable from nominal classification.

5.4 Geometric Visualization

Figures 7 and 8 show the intermediate representation obtained in the training sets for the case of three classes and four classes. These images were obtained on Pang & Lee's data set.

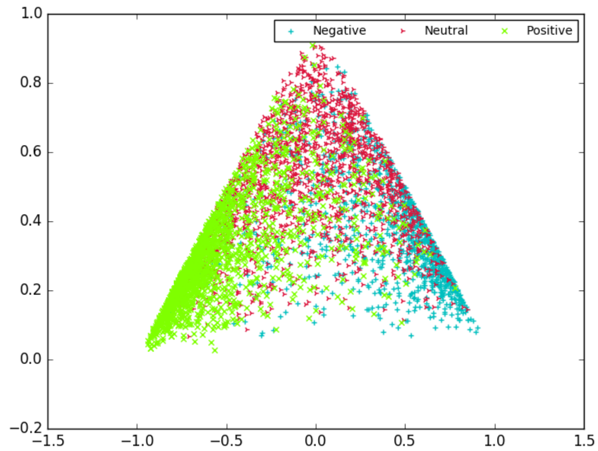

Figure 7: Intermediate representation for 3 classes (Pang & Lee data set).

The distribution of the training examples in this intermediate representation can be observed in Figure 7. The upper corner would represent the neutral class. Note that in the region between positives and negatives in the lower part there are not many elements. This is expected because naturally the elements that fall between positive and negative should tend to neutrality, that is, they should accumulate in the upper corner of the triangle. It can also be observed that there is a higher concentration in the central zones of the triangle. This agrees with the fact that, in general, ordinal classification is biased towards neutral values.

The triangle in Figure 7 is isosceles (the base goes from -1 to 1) and the distances among the classes would be $d(-,+) = 2$, $d(-,0) = \sqrt{2}$, $d(0,+) = \sqrt{2}$, this agrees with the method's geometric intuition. The further apart the classes are according to the order, the more distant they are in this representation.

In the four-class case in Figure 8, a similar phenomenon can be observed. There is a bias towards the two central classes rather than to the extreme classes. It can also be observed that the zone between extreme classes presents fewer elements. Note that for the four classes case there is no neutral class.

The geometric representation obtained agrees with the intuition behind the method, as it does in the case of the triangle. In addition, a bias towards neutrality can be observed in the data distribution as it is expected in an ordinal classification problem. This bias is also natural considering the class distribution of the data set.

5.5 BCOC Variations

It is possible to construct a version of BCOC that is based on the same geometric intuition, but using a one-vs-one approach instead of a one-vs-all approach. In this case, classifiers would be represented on the sides of the polygons. The final point to be used in the classification would be obtained by averaging, just as in the original version.

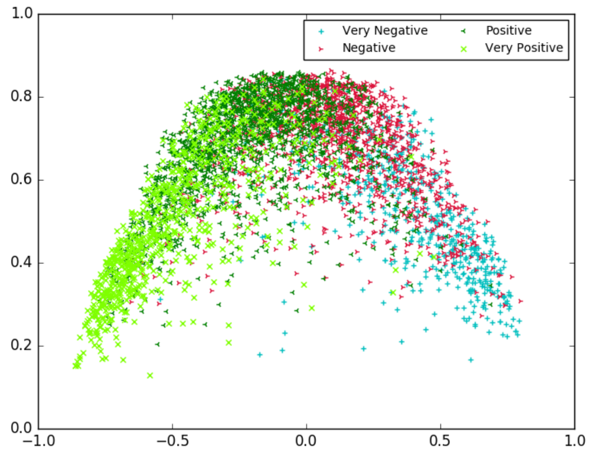

Figure 8: Intermediate representation for 4 classes (Pang & Lee data set).

Table 8: Comparison between BCOC versions one-vs-one (alternative) and one-vs-all (original).

Accuracy (%) for sentiment analysis			
Data set	Method	Average (+- SD)	Best
Pang & Lee (3 classes)	One-vs-one	68.29+-2.13	71.80
	One-vs-all	67.28+-1.90	70.20
Pang & Lee (4 classes)	One-vs-one	58.47+-1.95	61.07
	One-vs-all	58.68+-2.52	61.40

Experiments were carried out with this variation on the sentiment analysis data set. In both cases SVMs were utilized both in the inferior and superior levels. Results are shown in Table 8.

No significant difference has been found between both variations (see Table 8). In general it would be expected that the one-vs-all version would present a better performance because it would have more data than its counterpart based on one-vs-one, and therefore it would allow to generate more accurate estimators. However, this implies that training time is longer and requires more computing resources.

5.6 Discussion

In general, these results indicate that BCOC works well as a strategy to group classifiers for multi-class ordinal case in the field of sentiment analysis. Moreover, given the method's nature, it would be possible to combine it with other classifiers. In this case SVMs has been used in the superior level and NB/SVM in the inferior level, but other classifiers could be nested as well. The only requirement is that the first level delivers as a result an estimate of the probability of the corresponding class.

In general, the developed method is competitive in the case of ordinal classification. Even though in some cases it provides good results, these are not statistically significant because in most cases the performance is similar to the baseline (within a standard deviation). It is estimated that the cases where the method fails are due to lack of sufficient data to be able to obtain adequate classifiers for each class.

The results obtained in the study case for sentiment analysis show that the proposed approach is effective in more complex data sets, in an application domain that is particularly difficult due to the particularities of the data in film reviews (high level of noise, use of sarcasm, non-uniform structure, grammatical and orthographical errors). Based on the results, BCOC would present a more robust performance with respect to the number of classes of the problem than classic approaches of SVM and NB.

In general, the method behaves in a competitive way with the baseline and in some cases it presents improvements. In the three-class case there was no significant improvement in any of the data sets used. This can be due to the fact that the geometrical intuition of the method cannot be fully leveraged with such small number of classes, and thus as the quantity of classes increases, the use of the method would be expected to be more beneficial. In the case of four classes or more, it is observed that the method offers an enhanced performance in some of the cases presented.

6 CONCLUSIONS

In this work, a new method of ordinal multi-class classification based on the mathematical concept of barycentric coordinates has been proposed. Experiments have been carried out on data sets from the fields of ordinal classification and sentiment analysis. Through these, it has been shown that this proposal yields competitive results in multiple domains and in some cases superior results. Other ways of implementing the algorithm and the geometrical interpretation of this representation in barycentric coordinates have also been considered.

Considering the results obtained, it can be affirmed that the determination of multi-class polarity could benefit from the use of ordinal classification methods that complement the traditional classification approaches. In particular, it can be affirmed that the proposed method is competitive and in some domains it offers an improvement for problems with at least 4 ordinal classes.

The intermediate representation used by the method to feed the final classifier has a clear geometrical interpretation, as it could be observed in the discussion section, showing clearly the behavior of the different classes on the training set. The combination of the probabilities from multiple one-vs-all classifiers to form the intermediate representation is a key point in this method. Taking this into consideration, different class distributions may affect the classification probabilities and in turn, might produce different final classification results.

As future work, it is proposed to explore different combinations of classifiers. Depending on the specific problem, a specific combination of classifiers could deliver improved results. In particular, it is proposed to experiment with more advanced methods such as those obtained through deep learning techniques. Although SVM and NB are the most widely used techniques in literature, applying other methods together with BCOC might offer better results.

REFERENCES

[1] Arthur Asuncion and David Newman. 2007. UCI machine learning repository. (2007).

[2] Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.. In *LREC*, Vol. 10. 2200–2204.

[3] Ricardo Baeza-Yates, Alistair Moffat, and Gonzalo Navarro. 2002. Searching large text collections. In *Handbook of massive data sets*. Springer, 195–243.

[4] John Blitzer, Mark Dredze, Fernando Pereira, and others. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, Vol. 7. 440–447.

[5] Harold Scott Macdonald Coxeter, Harold Scott Macdonald Coxeter, Harold Scott Macdonald Coxeter, and Harold Scott Macdonald Coxeter. 1969. *Introduction to geometry*. Vol. 136. Wiley New York. 216–221 pages.

[6] Pedro Antonio Gutiérrez, Maria Perez-Ortiz, Javier Sanchez-Monedero, Francisco Fernandez-Navarro, and Cesar Hervas-Martinez. 2016. Ordinal regression methods: survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering* 28, 1 (2016), 127–146.

[7] Jens C Huhn and Eyke Hullermeier. 2008. Is an ordinal class structure useful in classifier learning? *International Journal of Data Mining, Modelling and Management* 1, 1 (2008), 45–67.

[8] Brian Keith, Exequiel Fuentes, and Claudio Meneses. 2017. Sentiment Analysis and Opinion Mining Applied to Paper Reviews. (2017).

[9] Sotiris B Kotsiantis and Panagiotis E Pintelas. 2004. A cost sensitive technique for ordinal classification problems. In *Hellenic Conference on Artificial Intelligence*. Springer, 220–229.

[10] Bing Liu. 2011. Opinion mining and sentiment analysis. In *Web Data Mining*. Springer, 459–526.

[11] Bo Pang and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics, 115–124.

[12] Bo Pang, Lillian Lee, and others. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval* 2, 1–2 (2008), 1–135.

[13] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, and others. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, Oct (2011), 2825–2830.

[14] Martin F Porter. 1980. An algorithm for suffix stripping. *Program* 14, 3 (1980), 130–137.

[15] Kumar Ravi and Vadlamani Ravi. 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems* 89 (2015), 14–46.

[16] Javier Sánchez Monedero. 2014. Challenges in ordinal classification: artificial neural networks and projection-based methods. (2014).

[17] Javier Sánchez-Monedero, Pedro A Gutiérrez, Peter Tiňo, and César Hervás-Martínez. 2013. Exploitation of pairwise class distances for ordinal classification. *Neural computation* 25, 9 (2013), 2450–2485.

[18] Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, Christopher Potts, and others. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, Vol. 1631. Citeseer, 1642.

[19] S Sonnenburg, CS Ong, S Henschel, M Braun, and PO HOYER. 2011. Machine Learning Data Set Repository. *URL http://mldata. org/.(Cited on pages 115 and 131.)* (2011).

[20] Donghui Wang, Junhai Zhai, Hong Zhu, and Xizhao Wang. 2014. An Improved Approach to Ordinal Classification. In *International Conference on Machine Learning and Cybernetics*. Springer, 33–42.