A Sentiment Analysis Model to Analyze Students Reviews of Teacher Performance Using Support Vector Machines

Guadalupe Gutiérrez Esparza[1], Alejandro de-Luna[1], Alberto Ochoa Zezzatti[2], Alberto Hernandez[3], Julio Ponce[4], Marco Álvarez[1], Edgar Cossio[5] & Jose de Jesus Nava[6]

[1]Universidad Politécnica de Aguascalientes, Mexico.
[2]Universidad Autónoma de Ciudad Juárez, Mexico.
[3]Universidad Autónoma del Estado de Morelos, Mexico
[4]Universidad Autónoma de Aguascalientes, Mexico
[5]Universidad Enrique Diaz de Leon, Mexico
[6]Pinnacle Aerospace, Mexico

guadalupe.gutierrez@upa.edu.mx

Abstract. Teacher evaluation is considered an important process in higher education institutions to know about teacher performance and implement constructive strategies in order to benefit students in their education. The opinion of students has become one of the main factors to consider when evaluating teachers. In this study we present a Model called SocialMining using a corpus of real comments in Spanish about teacher performance assessment. We applied Support Vector Machines algorithm with three kernels: linear, radial and polynomial, to predict a classification of comments in positive, negative or neutral. We calculated sensibility, specificity and predictive values as evaluation measures. The results of this work may help other experiments to improve the classification process of comments and suggest teacher improvement courses for teachers.

Keywords. Support vector machines, teacher performance assessment, performance evaluation.

1 Introduction

In the teaching and learning process it is crucial to evaluate the teaching performance. This evaluation is one of the most complex processes in any university, since various factors and criteria should be met to be concentrated in order to provide a final assessment to the professional. Teacher evaluation can be performed by an observation guide or a rubric with different evaluation criteria. However, when teacher performance is evaluated by students, varied opinions are collected from the same established criteria. It is at this time when emerges the need to use sentiment analysis methods to the analysis these comments.

Sentiment Analysis [1] is an application of natural language processing, text mining and computational linguistics, to identify information from the text. Education is

S. Omatu et al. (eds.), *Distributed Computing and Artificial Intelligence, 14th International Conference*, Advances in Intelligent Systems and Computing 620, DOI 10.1007/978-3-319-62410-5_19

one of the areas that currently have seen favorable sentiment analysis application, in order to improve attended sessions or distance education. In his research, Binali [2] ensures that students represents their emotions in comments, so it is a way to learn about various aspects of the student. Other research [3] use sentiment analysis on teacher's feedback from students enrolled in online courses in order to know their opinion and determine whether there is a connection between emotions and dropout rates. Student feedback on quality and standards of learning is considered as a strategy to improve the teaching process [3] and can be collected through a variety of social network, blogs and surveys.

In this paper, we presented a Model called SocialMining to support the Teacher Performance Assessment applying Support Vector Machines (SVMs). We selected SVMs as a classifier due to its high performance in classification applications. Further experiments with other machine learning algorithms will follow.

This paper is organized as follows. Section 2 presents related work. Section 3 shows the SocialMining Model architecture. Section 4 describes data and methods used, besides the experimental design. Section 5 includes the results obtained and a discussion. Finally, the conclusions of the work done, and the future work, are presented in Section 6.

2 Related work

The table 1 shows an overview of some related work. All these works have been successful in their different combinations of methods and algorithms. This table is not exhaustive.

Table 1. List of features to analize

Reference	Description	Algorithms used	Dataset used
[4]	Development of an application to know the student emotional state.	SVM, *Naïve Bayes, Complement Naïve Bayes.*	Students reviews at the University of Portsmouth
[5]	Development of an application, called SentBuk to retrieve identify users' sentiment polarity and emotional changes.	Lexicon based, machine-learning and hybrid approaches.	Around 1,000 comments in Spanish
[6]	Construction of a teaching evaluation sentiment lexicon, to analyzes the terms and phrases from student opinions	Lexicon in Thai, SVM, ID3 and Naïve Bayes.	Reviews by students at Loei Raja hat University.
[7]	Design of an experimental study to predict teacher performance.	Lexicon with 167 keywords positive and 108 keywords negative.	1,148 feedbacks bye students, obtained from RateMyProfessors.com

[8]	Proposed a method to detect the feeling of students on some topics and support the teacher to improve their teaching process.	Latent Dirichlet Allocation, SVM, Naïve Baye and, Maximum Entropy.	Movie reviews dataset [9] and comments by 75 students collected from Moodle.
[10]	Proposed a system to help the developer and educator to identify the most concentrated pages in E-learning portals.	Bayesian classification, Naïve Bayesian and SVM.	Use a collection of 100 users review from the website Functionspace.org.
[11]	Implement sentiment analysis in M-learning System, to know the user opinions of the M-learning system.	*Naïve Bayes*, KNN and *Random Forest.*	300 Reviews from www.market.android .com

From table 1 we can see that most of previous research has focused on particular aspects of Education. In this work we proposed a model to evaluate teacher performance consider Spanish reviews from students and applying machine learning algorithms to classify them as positive, negative and neutral. The results of this work may help to other experiments to improve the classification process of comments and suggest teacher improvement courses.

3 SocialMining model architecture

The SocialMining model is composed of three phases, where there a comments extraction process (teacher's feedback from students), a feature selection process and a comments classification into positive, negative and neutral, applying SVMs.

Phase 1. Comments extraction and cleaning process. At this phase we extracted teacher's feedback from students to generate a comments corpus. Then we do a labeling process to classify the comments into positive and negative considering a numeric range. The numeric range varied from: -2 to -0.2 is used for negative comments, -2 value express very negative comments. Values between +0.2 to +2 apply to positive comments, +2 is used as a positive comments. Likewise, those comments labeled with the number 1 are considered as neutral.

At cleaning process, the stop words and nouns that appear in most of the comments are deleted (e.g. teacher, university, class, subject, school and others). In addition, punctuation marks are removed and capitalized words are converted to lowercase.

Phase 2. Feature selection process. Once finished the cleaning process, we performed a feature selection process, removing repetitive terms and applying some functions to select the required terms or features, this process is like a filtering.

A feature in sentiment analysis is a term or phrase that helps to express an opinion positive or negative. There are several methods used in features selection, where some of them are based on the syntactic word position, based in information gain, using genetic algorithms and using trees like random forest importance variable.

At this phase it's necessary to know the importance of each feature, by their weight. So it's applied Term Frequency Inverse Document Frequency (TF-IDF).

Phase 3. Classification comments. At this phase the corpus of comments and features (matrixCF) is partitioned into two independent datasets. The first dataset is dedicated to training process (train) and is used in classification algorithms to find patterns or relationships among data; the second dataset is considered for the testing process (test) in order to adjust the model performance. In this work two thirds of the matrixCF are used for train dataset and one third for test dataset. Then the cross-validation method of k iterations is applied to control the tuning and training of SVMs algorithm. In this method matrixCF is divided into K subsets. One of the subsets is used as test data and the remaining (K-1) as training data. The cross-validation process is repeated for k iterations, with each of the possible subsets of test data, resulting in a confusion matrix with average values. Once the k iterations have completed cross validation accuracy is obtained. In this research, k is equal to 10.

The tuning process in SVMs allows adjusting the parameters of each kernel (linear, radial basis and polynomial). Then is performed a training process, through which identified whether the value of the parameters vary or remain constant.

Finally, the implementation of SVMs is performed presenting as a result the confusion matrix and accuracy values as well as the metric ROC.

4 Material and methods

4.1 Data

The dataset used in this work comprises 1040 comments in Spanish of three groups of systems engineering students at Polytechnic University of Aguascalientes. They evaluated 21 teachers in the first school grade (2016). For this study we considered only those comments free from noise or spam (characterized in this study as texts with strange characters, empty spaces no opinion or comments unrelated to teacher evaluation). In this work we identified a set of 99 features. An extract of the features are listed in table 2.

Table 2. List of features

Feature	Polarity feature	Feature	Polarity feature
Atenta	Positive	Debería	Negative
Agrada	Positive	Debe	Negative
Apoya	Positive	Impuntual	Negative
Aprender	Positive	Elitista	Negative
Bien	Positive	Impaciente	Negative
Bipolar	Positive	Problemático	Negative

4.2 Performance measures

We used typical performance measures in machine learning such as: accuracy, balanced accuracy, sensitivity, specificity, and metric ROC curve[12, 13].

4.3 Classifiers

SVMs is an algorithm introduced by Vapnik [14] for the classification of both linear and nonlinear data, it has been known for its quality in text classification [15]. There are kernels that can be used in SVMs, such as: linear, polynomial, radial basis function (RBF) and sigmoid. Each of these kernels has particular parameters and they must be tuned in order to achieve the best performance. In this work we selected the first three kernels to classify comments; this is mainly because of their good performance. The kernel used in this study, are: linear, radial basis and polynomial.

4.4 Experimental design

We created a dataset containing 1040 comments and 99 features associated with teacher performance assessment. We used train-test evaluation, two-thirds (2/3) for training and (1/3) one-third for testing, then performed 30 runs applying SVMs with polynomial, radial basis function (RBF) and linear kernel. For each run are computed performance measures. In each run we set a different seed to ensure different splits of training and testing sets, all kernels use the same seed at the same run.

Each kernel requires tuning different parameters (see table 3). A simple and effective method of tuning parameters of SVMs has been proposed by Hsu [16], the grid search. The C values used for the kernels, range from 0.1 to 2, the value of sigma (σ) varied from 0.01 to 2, the degree value parameter range from 2 to 10, and values between 0 and 1 are assigned for coef parameter. We performed 30 train-test runs using different seeds and calculated the accuracy and balanced accuracy of each run.

5 Discussion and Results

In this section, we present the experiments results with three kernels in SVMs. The first step is to determine the parameters of each kernel of SVMs, so we first load the data and create a partition of corpus of comments, then divided it into training and testing datasets, then use a traincontrol in R [17] to set the training method. We use the Hsu [16] methodology to specify the search space in each kernel parameter. ROC is the performance criterion used to select the optimal kernels parameters of SVMs.

Setting the seed to 1 in the process of optimization parameters we generated paired samples according to Hothorn [18] and compare models using a resampling technique. The table 3 shows the summary of resampling results using resamples in R [17], the performance metrics are: ROC, sensibility and specificity.

Table 3. Summary resampling results of parameters optimization

ROC						
Kernel	**Min.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
Linear	0.7872	0.8315	0.8616	0.8579	0.8714	0.9080
Radial	0.7881	0.8218	0.8462	0.8462	0.8635	0.9176
Poly	0.8350	0.8608	0.8773	0.8755	0.8918	0.9321
Sensibility						
Kernel	**Min.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
Linear	0.5857	0.6571	0.7	0.7006	0.7571	0.8143
Radial	0.6571	0.7286	0.7571	0.7663	0.8	0.9
Poly	0.3429	0.7	0.7571	0.7131	0.7714	0.8143
Specificity						
Kernel	**Min.**	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**
Linear	0.7971	0.8696	0.8971	0.889	0.9118	0.9565
Radial	0.6812	0.7681	0.7971	0.7925	0.8261	0.8696
Poly	0.8088	0.8529	0.8841	0.8849	0.913	0.9855

Once obtained optimized parameters for each kernel, the execution of each SVMs model is performed. The table 4 shows the average results of each kernel of SVMs across 30 runs. Also the standard deviation of each metric is presented.

Table 4. Average results across 30 runs in three kernels of SVM

	Accuracy	**balanced Accuracy**	**Sensitivity**	**Specificity**
SVM Linear	**0.8038**	**0.8149**	**0.8936**	**0.7160**
	0.0153	0.0146	0.0277	0.0364
SVM Radial	**0.7850**	**0.7893**	**0.8242**	**0.7467**
	0.0190	0.0159	0.0422	0.0561
SVM Poly	**0.6779**	**0.7014**	**0.8661**	**0.4941**
	0.0363	0.1336	0.1649	0.1009

The linear kernel obtained a balanced accuracy above 0.80. Values obtained in Sensitivity were much higher than those obtained in specificity in all kernels. The kernel polynomial (SVM Poly) had the lowest performance in all metrics except in sensitivity. The three kernels resulted more sensitive than specific.

6 Conclusions

In computer science its attractive the use of this type of machine learnings models to automate processes, save time and contribute to decision making. The SocialMining model supports the analysis of the behavior from unstructured data provided by students. The sentiment analysis is based on the analysis of texts and the SocialMining can provide a feasible solution to the problem of analysis of teacher evaluation comments. Further experiments will be conducted in this ongoing research project.

It is important to point out that is necessary reduce the number of features through a depth analysis to identify the most relevant features of teacher performance assessment, in order to improve the results of comments classification process. Also we considered important have a corpus of comments balanced (positive and negative comments in equal quantity) for testing and training process.

In addition to conduct a deeper analysis for relevant features selection, we considered necessary implement other machine learning algorithms in order to measure the performance of each algorithm in the classification of comments and select the optimal with high accuracy results.

Based on the adequate results that have been obtained Model SocialMining applying Naïve Bayes and a corpus of subjectivity [19], it is considered that with the implementation of other algorithms of machine learning known for their good performance in classification process, the model proposed will support teachers to improve their classes. Another contribution SocialMining Model is that it could support the department of teacher improvement, to create and suggest different courses to teacher training skills, according to the possible opportunities that are detected in the comments of students.

References

1. Bing, L., *Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies).* Morgan & Claypool Publishers., 2012.
2. Binali, H.H., Chen Wu, and Vidyasagar Potdar., *A new significant area: Emotion detection in E-learning using opinion mining techniques.* Digital Ecosystems and Technologies, 2009. DEST'09. 3rd IEEE International Conference on. IEEE,, 2009.
3. Wen, M., Yang, D., & Penstein Rosé, C., *Sentiment Analysis in MOOC Discussion Forums: What does it tell us?* . Proceedings of Educational Data Mining, 1 - 8. (http://goo.gl/fViyBH), 2014.
4. Altrabsheh, N., Cocea, M., & Fallahkhair, S., *Learning Sentiment from Students' Feedback for Real-Time Interventions in Classrooms.* Adaptive and Intelligent Systems. Volume 8779 of the series Lecture Notes in Computer Science, 40-49. doi: http://dx.doi.org/10.1007/978-3-319-11298-5_5, 2014.
5. Ortigosa, A., Martín, J. & Carro, R., *Sentiment analysis in Facebook and its application to e-learning.* Computers in Human Behavior. Volume 31, 527 - 541, http://dx.doi.org/10.1016/j.chb.2013.05.024, 2014.
6. Pong-inwong, C.R., W., *TeachingSenti-Lexicon for Automated Sentiment Polarity Definition in Teaching Evaluation.* In Semantics, Knowledge and Grids (SKG), 10th International Conference, 84 – 91, Beijing: IEEE., 2014.
7. Kaewyong, P., Sukprasert, A., Salim, N., and Phang, A., *The possibility of students' comments automatic interpret using lexicon based sentiment*

analysis to teacher evaluation. Conference: The 3rd International Conference on Artificial Intelligence and Computer Science 2015, At Penang,MALAYSIA, 2015.

8. F., C., M., de Santo and L., Greco, *SAFE: A Sentiment Analysis Framework for E-Learning.* International Journal of Emerging Technologies in Learning, Vol. 9, Issue 6, pp. 37, 2014.
9. Pang, B., & Lee, L., *Thumbs up? Sentiment Classification using Machine Learning Techniques.* EMNLP '02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10. Pages 79-86. doi>10.3115/1118693.1118704, 2002.
10. Bharathisindhu, P., & Brunda, S., *Identifying e-learner's opinion using automated sentiment analysis in e-learning.* IJRET: International Journal of Research in Engineering and Technology. Volume: 03 Issue: 01. doi: http://dx.doi.org/10.15623/ijret.2014.0319086, 2014.
11. Kirubakaran, A.J.a.E., *M-Learning Sentiment Analysis with Data Mining Techiniques.* International Journal of Computer Science and Telecommunications. Trichy-India., 2012.
12. I.H. Witten, E.F., and M. A. Hall., *Data Mining: Practical Machine Learning Tools and Techniques.* Morgan Kaufmann, third edition., 2011.
13. J. Han, M.K., and J. Pei, *Data Mining: concepts and techniques.* Morgan Kaufmann, San Francisco, USA., 2012.
14. Vapnick, V., *Statistical Learning Theory.* Wiley, New York, 1998.
15. Esuli, A.S., F., *SENTIWORDNET: a publicly available lexical resource for opinion mining.* . In Proceedings of the 5th Conference on Languaje Resources and Evaluation (LREC 2006). pages 417-422, Genoa, Italy., 2006.
16. C.C. Hsu, C.W.C.a.C.J.L., *A practical guide to support vector classification.* Technical report, Department of Computer Science, National Taiwan University, 2003.
17. R-Core-Team, *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. (http://goo.gl/e40yiU), 2013.
18. Hothorn, T., Leisch, F., Hornik, K. and Zeileis A., *The design and analysis of benchmark experiments.* Journal of Computational and Graphical Statistics. Vol 14 (3) pp. 675-699., 2005.
19. Gutiérrez, G., Padilla, A., Canul-Reich, J., De-Luna, A. and Ponce, J., *Proposal of a Sentiment Analysis Model in Tweets for Improvement of the Teaching - Learning Process in the Classroom Using a Corpus of Subjectivity.* International Journal of Combinatorial Optimization Problems and Informatics, Vol. 7, No. 2, May-Aug 2016, pp. 22-34., 2016.