Analyzing Students Reviews of Teacher Performance Using Support Vector Machines by a Proposed Model

G. Gutiérrez[1(✉)], J. Ponce[2], A. Ochoa[3], and M. Álvarez[1]

[1] Universidad Politécnica de Aguascalientes, Aguascalientes, Mexico
`guadalupe.gutierrez@upa.edu.mx`
[2] Universidad Autónoma de Aguascalientes, Aguascalientes, Mexico
[3] Universidad Autónoma de Ciudad Juárez, Ciudad Juárez, Mexico

Abstract. Sentiment Analysis arises from areas such as natural language processing and data mining, it has become a key area for society because it is possible to identify emotions expressed in texts. Teacher evaluation is considered an important process in higher education institutions to measure teacher performance and implement constructive strategies to benefit students in their education. This paper describes the design and development of a Model with the purpose of analyzing the students reviews of teacher performance. The collection of comments was done in two ways, the first is that comments were collected from a teacher evaluation conducted in 2016 and the second was made on Twitter through the participation of students of the Universidad Politécnica de Aguascalientes. In this work we applied Support Vector Machines with three kernels: linear, radial and polynomial, to predict a classification of comments in positive, negative or neutral and we calculated evaluation measures.

Keywords: Reviews of teacher performance · Support Vector Machine
Evaluation measures

1 Introduction

In the teaching process is crucial to evaluate the teaching performance [1]. This evaluation is one of the most complex processes in any university, since various factors and criteria like: planning of classes, schedules, delivery of evaluation evidence, attendance to courses for the improvement of teaching and teaching styles, among others, should be met to be concentrated in order to provide a final assessment for the professor. Teacher evaluation can be performed by an observation guide or a rubric. However, when teacher performance is evaluated by students, varied opinions are collected from the same established criteria. Therefore Education is one of the areas that in recent years has shown interest in analyzing the comments of students in order that teachers improve their teaching techniques promoting appropriate learning in students. This is possible by Sentiment Analysis [2] an application of natural language processing, text mining and computational linguistics, to identify information from the text.

In his research, Binali [3] ensures that students represents their emotions in comments, so it is a way to learn about various aspects of the student. Wen [4] applies Sentiment Analysis on feedback from students about their teachers, enrolled in online courses in order to know their opinion and determine whether there is a connection between emotions and dropout rates. Students feedback on quality and standards of learning is considered as a strategy to improve the teaching process [4] and can be collected through a variety of social networks, blogs and surveys.

In this paper, we presented a model called SocialMining to support the Teacher Performance Assessment applying Support Vector Machines (SVM). We selected SVM as a classifier due to its high performance in classification applications [5, 6]. 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 and experimental design. Section 5 includes the results. Finally, the conclusions of the work are presented in Sect. 6.

2   Related Work

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

From Table 1 we can see that most of previous research has focused on particular aspects like: know the student emotional state, analyze the terms and phrases from opinions of students, detect the feelings of students on some topics and know the user opinions of the E-learning systems. In this work we proposed a model to evaluate teacher performance considering 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 improve the classification process of comments and suggest courses to teachers.

Table 1. List of features to analyze

Reference	Description	Algorithms used	Dataset used
[7]	Development of an application to know the student emotional state	SVM, *Naïve Bayes, Complement Naïve Bayes*	Students reviews at the University of Portsmouth
[8]	Development of an application, called SentBuk to retrieve and identify sentiment polarity and emotional changes of users	Lexicon based, machine-learning and hybrid approaches	Around 1,000 comments in Spanish
[9]	Construction of a teaching evaluation sentiment lexicon, to analyze the terms and phrases from opinions of students	Lexicon in Thai, SVM, ID3 and Naïve Bayes	Reviews by students at Loei Raja hat University

Table 1. (*continued*)

Reference	Description	Algorithms used	Dataset used
[10]	Design of an experimental study to predict teacher performance	Lexicon with 167 keywords positive and 108 keywords negative	1,148 feedbacks by students, obtained from RateMyProfessors.com
[11]	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 Bayes and, Maximum Entropy	Movie reviews dataset [12] and comments by students collected from Moodle
[13]	Proposed a system to help the developer and the educator to identify the most concentrated pages in E-learning portals	Bayesian classification, Naïve Bayesian and SVM	Use a collection of 100 reviews of users from the website Functionspace.org
[14]	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

3   SocialMining Model Architecture

The SocialMining model is composed of three phases: a comments extraction process (feedback from students about their teachers) and cleaning, a feature selection process, and classification of comments into positive, negative and neutral, applying SVM. The last phase includes an evaluation process of SVM results in kernels.

Phase 1: Comments Extraction and Cleaning Process. In this phase we extracted feedback from students about their teachers to generate a corpus of comments. 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 (Fig. 1).

In this 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. The output in this phase is the corpus of comments.
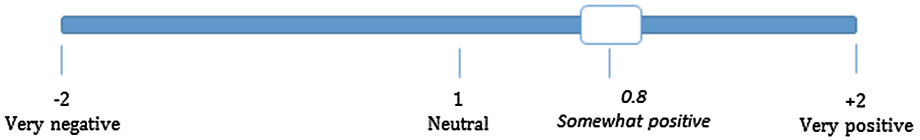
Fig. 1. Numeric range used to tagging process in comments

Phase 2: Feature Selection. 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. So the input in this phase is the corpus of comments and the output are the features.

A feature in Sentiment Analysis is a term or phrase that helps to express a positive or negative opinion. There are several methods used in feature selection, where some of them are based on the syntactic word position, based in information gain, using a importance variable calculated by genetic algorithms [15] and trees like the variable importance measures for random forest [16]. In this phase is necessary to know the importance of each feature, by their weight. So the Term Frequency - Inverse Document Frequency (TF-IDF) is applied (Fig. 2).
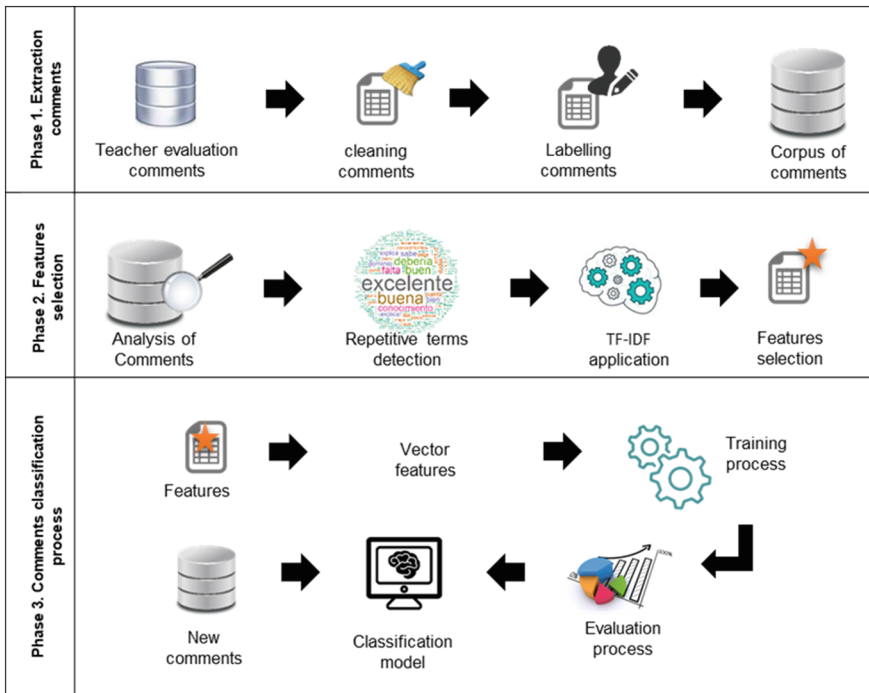

Fig. 2. SocialMining model architecture

Phase 3: Comments Classification Process. In 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 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 training dataset and one-third for test dataset. Then the cross-validation method of k iterations is applied to control the tuning and training of SVM. 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 been completed, cross validation accuracy is obtained. In this research, K is equal to 10.

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

Finally, the implementation of SVM is performed presenting as a result the confusion matrix and accuracy values as well as the metric Receiver operating characteristic (ROC) curve.

4 Materials and Methods

4.1 Data

The dataset used in this work comprises 1040 comments in Spanish of three groups of systems engineering students at Universidad Politécnica de Aguascalientes. They evaluated 21 teachers in the first scholar 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 (Spanish)	Feature	Polarity feature	Feature	Feature (Spanish)	Polarity feature
Atenta	Attentive	Positive	Debería	Should	Negative
Agrada	Likes	Positive	Debe	Must	Negative
Apoya	Supports	Positive	Impuntual	Impuntual	Negative
Aprender	Learn	Positive	Elitista	Elitist	Negative
Bien	Fine	Positive	Impaciente	Impatient	Negative
Imparcial	Impatial	Positive	Problemático	Problematic	Negative

4.2   Performance Measures

We used typical performance measures in machine learning such as:

- Accuracy, primary measure to evaluate the performance of a predictive model.
- Balanced accuracy, a better estimate of a classifier performance when a unequal distribution.
- Sensitivity, which measures the proportion of true positives.
- Specificity, measures the proportion of true negatives.
- ROC curve, measures the performance of a classifier through graphical representation [17, 18].

4.3   Classifiers

SVM is an algorithm introduced by Vapnik [19] for the classification of both linear and nonlinear data, it has been known for its quality in text classification [20]. There are kernels that can be used in SVM, 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 in text classification [5, 6]. Table 3 shows the parameters of each kernel used in this study.

Table 3. Kernel parameters.

Kernel	Parameter
Linear	C
Radial basis	C, sigma
Polynomial	C, degree

- C is the parameter for the soft margin cost function, it determines a tradeoff between a wide margin and classifier error. A very small value of C cause a larger margin separating hyperplane and the model get fit tighter to data, however a large value of C reduce the margin and this may cause more error on the training set.
- Sigma determines the width for Gaussian distribution in Radial basis kernel.
- Degree control the flexibility of the resulting classifier in Polynomial Kernel.

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 there were performed 30 runs applying SVM with polynomial, radial basis function (RBF) and linear kernel. For each run performance measures are computed. 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 SVM has been proposed by Hsu [21], 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 for each run.

5   Discussion and Results

In this section, we present the results with three kernels in SVM. The first step is to determine the parameters of each kernel of SVM, 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 train control in R [22] to set the training method. We use the Hsu [21] methodology to specify the search space in each kernel parameter. ROC is the performance criterion used to select the optimal kernels parameters of SVM.

Setting the seed to 1 in the process of optimization parameters, we generated paired samples according to Hothorn [23] and compare models using a resampling technique. Table 4 shows the summary of resampling results using R [22], the performance metrics are: ROC, sensibility and specificity. In the Fig. 3 we can see the plot of summary resampling results, in this case, the polynomial kernel apparently has a better performance than linear and RBF (radial).

Table 4. 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
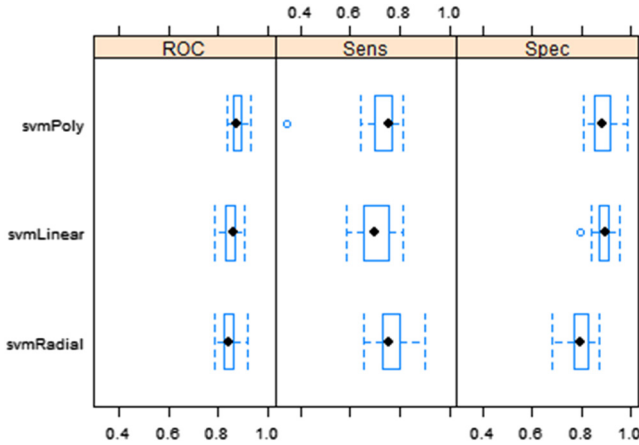
Fig. 3. Summary resampling results of kernels performance

Once obtained optimized parameters for each kernel, the execution of each SVM model is performed.

The Table 5 shows the average results of each kernel of SVM across 30 runs. Also the standard deviation of each metric is presented.

Table 5. Average results across 30 runs in three kernels of SVM

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

The linear kernel obtained a balanced accuracy above 0.80, this is an indicator that the classifier is feasible to use in comments classification. Values obtained in Sensitivity were much higher than those obtained in specificity in all kernels, which indicates that the classifier can detect the negative comments of the teachers. 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 is 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 having a corpus of balanced comments (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 to 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 by the SocialMining model applying Naïve Bayes and a corpus of subjectivity [24], we considered that with the implementation of other algorithms of machine learning well-known for their good performance in classification process.

About how SocialMining model support the improvement of teaching in the first instance it allows a quicker analysis of student comments, identifying which teachers have mostly negative comments which allows interventions with the teacher in order to support it through teacher improvement courses. Each school period, courses are offered to teachers, however the comments of students are not considered among the criteria to recommend a certain course to the teacher. For this reason it is believed that the Model presented in this work will support the improvement of teaching.

References

1. Careaga, A.: La evaluación como herramienta de transformación de la práctica docente. Educere **5**(15), 345–352 (2001)
2. Bing, L.: Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies). Morgan & Claypool Publishers, San Rafael (2012)
3. Binali, H.H., Wu, C., Potdar, V.: A new significant area: emotion detection in E-learning using opinion mining techniques. In: 3rd IEEE International Conference on Digital Ecosystems and Technologies (DEST 2009). IEEE (2009)
4. Wen, M., Yang, D., Penstein Rosé, C.: Sentiment analysis in MOOC discussion forums: what does it tell us? In: Proceedings of Educational Data Mining, pp. 1–8 (2014). http://goo.gl/fViyBH
5. Altrabsheh, N., Cocea, M., Fallahkhair, S.: Learning sentiment from students' feedback for real-time interventions in classrooms. In: Bouchachia, A. (ed.) Adaptive and Intelligent Systems. LNCS, vol. 8779, pp. 40–49. Springer, Cham (2014). http://dx.doi.org/10.1007/978-3-319-11298-5_5
6. Manning, C.D., Raghavan, P., Schütze, H.: Support vector machines and machine learning on documents. In: Introduction to Information Retrieval, pp. 319–348 (1998)
7. Sarkar, A., et al.: Text Classification using Support Vector Machine (2015)
8. Ortigosa, A., Martín, J., Carro, R.: Sentiment analysis in Facebook and its application to e-learning. Comput. Hum. Behav. **31**, 527–541 (2014). http://dx.doi.org/10.1016/j.chb.2013.05.024

9. Pong-Inwong, C.R., Rungworawut, W.S.: Teaching senti-lexicon for automated sentiment polarity definition in teaching evaluation. In: 10th International Conference on Semantics, Knowledge and Grids (SKG), Beijing, pp. 84–91. IEEE (2014)

10. Kaewyong, P., Sukprasert, A., Salim, N., Phang, A.: The possibility of students' comments automatic interpret using lexicon based sentiment analysis to teacher evaluation. In: The 3rd International Conference on Artificial Intelligence and Computer Science 2015, Penang, Malaysia (2015)

11. Francesco, C., de Santo, M., Greco, L.: SAFE: a sentiment analysis framework for e-learning. Int. J. Emerg. Technol. Learn. **9**(6), 37 (2014)

12. Pang, B., Lee, L.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10, pp. 79–86 (2002). https://doi.org/10.3115/1118693.1118704

13. Bharathisindhu, P., Brunda, S.: Identifying e-learner's opinion using automated sentiment analysis in e-learning. Int. J. Res. Eng. Technol. **3**(1) (2014). http://dx.doi.org/10.15623/ijret.2014.0319086

14. Kirubakaran, E.: M-learning sentiment analysis with data mining techniques. Int. J. Comput. Sci. Telecommun. (2012)

15. Witten, I.H., Frank, E., Hall, M.A.: Data Mining: Practical Machine Learning Tools and Techniques, 3rd edn. Morgan Kaufmann (2011)

16. Punch III, W.F., et al.: Further research on feature selection and classification using genetic algorithms. In: ICGA, pp. 557–564 (1993)

17. Strobl, C., et al.: Conditional variable importance for random forests. BMC Bioinform. **9**(1), 307 (2008)

18. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco (2012)

19. Vapnick, V.: Statistical Learning Theory. Wiley, New York (1998)

20. Esuli, A., Sebastiani, F., SENTIWORDNET: a publicly available lexical resource for opinion mining. In: Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy, pp. 417–422 (2006)

21. Hsu, C.-W., Chang, C.-C., Lin, C.-J.: A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University (2003)

22. R-Core-Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2013). http://goo.gl/e40yiU

23. Hothorn, T., Leisch, F., Hornik, K., Zeileis, A.: The design and analysis of benchmark experiments. J. Comput. Graph. Stat. **14**(3), 675–699 (2005)

24. Gutiérrez, G., Padilla, A., Canul-Reich, J., De-Luna, A., 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. Int. J. Comb. Optim. Probl. Inform. **7**(2), 22–34 (2016)