ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/313964921

Sentiment analysis of students feedback: A study towards optimal tools

Conference Paper · December 2016

DOI: 10.1109/IWCI.2016.7860361

CITATIONS
3

READS
132

1 author:

Mohammad Aman Ullah
International Islamic University Chittagong

5 PUBLICATIONS **4** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project M.Sc. Thesis View project

All content following this page was uploaded by Mohammad Aman Ullah on 06 November 2018.

The user has requested enhancement of the downloaded file.

2016 International Workshop on Computational Intelligence (IWCI)
12-13 December 2016, Dhaka, Bangladesh

Sentiment Analysis of Students Feedback: A Study towards Optimal Tools

Mohammad Aman Ullah
Department of Computer Science and Engineering,
International Islamic University Chittagong, Chittagong-4203, Bangladesh,
ullah047@yahoo.com

Abstract—**Educational Institutions attempts to gather feedback from students' to study their sentiments towards courses and instructors and to enhance the performance of the instructors. Basically, such feedbacks are gathered at the end of the semester with the use of survey forms. However, this technique is very tedious, slow and time consuming. With the advent of social media, especially Facebook, the collection of feedback become easier through Facebook pages and groups. But, analyzing those feedbacks is equally challenging. This paper addresses those problems and uncovers the best model for analyzing those feedbacks with the use of machine learning techniques such as Support Vector Machines (SVM), Maximum Entropy (ME), Naive Bayes (NB), and Complement Naive Bayes (CNB) and applying neutral class. And, found SVM as the highest performer with an accuracy of 97% by applying different preprocessing and feature extraction techniques and avoiding neutral class, which outperform state-of-art work by 2%.**

Keywords—Student; Feedback; Analysis; Machine; Learning

I. INTRODUCTION

With an advent of the social network, feedback system became very common and useful in every domain of interest. One of such domain is educational Institution, where, instructors used to give the lecture in the class and students give the opinions in the social media such as Facebook or twitter etc., which provides the chances and equal methodological difficulties in understanding students' behavior, but need to address those difficulties and process them carefully to derive useful information to support instructors in improvement for further lectures. One of the study supports this is known as Sentiment Analysis (SA). SA is a field of Computational Study of people's like/dislike about an entity or objects [1]. Also, it reveals the polarity of natural language texts, such as positive, negative and in some cases neutral is considered [2], [3].

As research shows, sentiment analysis could be easily done and highly useful, if it could be applied in a particular domain, such as in this case Education domain [2]. So, the purpose of this research is to find best suited tools for each step of sentiment analysis, such as preprocessing, feature extraction and Machine learning techniques in educational settings in order to improve teaching methods, faculty members, student's related facility and infrastructure. Also, this research will find the best combinations of sentiment analysis tools as a whole.

The remaining work is organized as follows: Section II includes Literature Review, Section III represents details methodology. Whereas, Section IV describes the results and Discussions and section V includes Conclusions.

II. LITERATURE REVIEW

An optimal combination of pre-processing tools was searched for performance enhancement of the sentiment analysis from student feedback by Ortigosa et al. [4], Martin et al.[6]. They found that, if the following pre-processing tools such as stop words removal, case normalization, repeated letter removal, Removing Exclamation, Punctuation and question marks, negation, and spelling check being used in combination than the performance improve to a large extent. Troussas et al. [5] used emoticons along with above tools and also found the improved performance. With the combination of pre-processing tools such as stop words removal, case normalization, repeated letter removal, Removing Exclamation, Punctuation and question mark Altrabsheh et al. [2] got improved accuracy of 95%, which is highest among all others in the same sort of analysis.

An optimal feature extractor is also necessary for improving performing of the analysis of student feedback. Among all of the available features, Term co-occurrence is widely used. It may be unigram, bigram, N-gram etc. Agarwal et al. [7] and Wang et al.[8] found that, unigrams is better in performance than bigrams and trigrams in most of the cases. For instance, Go et al. [10], found decreased performance in case of bigrams. On the other hand, a twitter analysis done by Pak and Paroubek [9] shows that bigrams gives a higher performance than unigrams. But among all, Altrabsheh et al. [2] got the highest performance with accuracy 95%, with the use of the combinations of Term co-occurrences. Also, among different machine learning algorithms Support Vector Machines (SVM) were found to be the best by Song et al. [11] in the case of student feedback analysis. Altrabsheh et al. [2] also found SVM as the best classifier tools in analyzing student feedback. Altrabsheh et al. [13] in one of their research on real time student feedback found

978-1-5090-5769-6/16/$31.00 ©2016 IEEE

Complement Naive Bayes (CNB) as the highest performer among all the techniques with 84% accuracy. Fallakhair et al. [14] conduct an in-depth analysis of real time student feedback in class room and found SVM RB as the best performer among all the algorithms with accuracy 94%.

III. METHODOLOGY

The methodology of this research is shown in Fig. 1, and the details descriptions of each step of the methodology such as corpus collection, Data preprocessing, Feature Extraction, sentiment classification, polarity detection and performance analysis are given below:

A. Corpus Collection

One of the important tasks of any research is the collection of data. Similarly important is to identify sources of the data. This research considered Facebook as a source of data, as Facebook messages are not limited in size like tweets and become more eligible for research in our said domain. We have used the same data set of the University of Portsmouth as used in state-of-art work by Nabeela et al. [2], where they have collected data from different universities and the students are allowed to express their opinions in free of context about the lectures. The data set consists of 1036 data, among them 641 are positive, 292 are negative, and remaining are neutrally distributed. The amount of data for analysis was fixed up using trial and error method. The data are labeled into different classes as positive, negative, according to the intensity of expression, if the intensity is hard to determine, the neutral class is assigned. We have measured percent agreement (80.7%), the Fleiss kappa (0.665) and Krippendor's alpha (0.666) in order to verify reliability. The first measure is optimistic and the later measure is conservative.

B. Preprocessing

As the research aim is to find the optimal tools that help in improving the accuracy of the study. So, firstly, we have tried to find the optimal combination of pre-processing tools, as we know, the more the data set is error free, the more the accuracy is. For this, we have firstly conducted a test without pre-processing, as the result was not promising, we have done few low level pre-processing, as the result was improved, so, we have conducted different levels of pre-processing till the result exceeds the state-of-art research. The total pre-processing steps are as described below and shown in Table II, III, and IV.

1. Preprocessing W/O: At this level, the data set is tested without any preprocessing except case normalization. And the result of the test is used as the baseline for the different level of preprocessing being applied.
2. Preprocessing P1: At this level, unnecessary characters, numbers, exclamations, punctuations, question marks, and stop words that do not contain any sentiments are removed to reduce noise from the data set.
3. Preprocessing P2: At this level, the data set is the previously preprocessed one in P1. The pre-processing included at this level are, removing repeated words and replacing "n't" with negation "not" in order to increase the possibility of matching sentiment words, thus, improve the performance of analysis.
4. Preprocessing P3: This level includes the data set that being preprocessed in P2. The pre-processing done at this level is stemming, where, the short words are returned to its original words, thus improve accuracy analysis.
5. Preprocessing P4: This level includes the data set that being preprocessed in P3. The pre-processing done at this level is the removal of emoticons in order to ease and improve the results of the analysis.
6. Preprocessing P5: This level includes the data set that being preprocessed in P4. Additional preprocessing done at this level is the removal of the URL in order to reduce the complexity of the analysis.

It is to be noted that, the combination we have used in our study was little different than that was used by state-of-art research. As our study outperform the previous research with this combination of pre-processing, so, this combination could be considered as optimal pre-processing tools.

C. Feature Extraction

As the research shows, proper extraction and selection of features help improve the accuracy of sentiment analysis to a great extent. So, secondly, we have tried to find optimal feature extraction tools from among Term Co-occurrence and Opinion words or phrase. From among different Term Co-occurrence, we have used Uni-gram, Bigrams and Trigrams. But the results of precision and recall shows that, trigrams outperform the results of Uni-gram, Bigrams, and Opinion words as in Table I, so, trigram could be considered as optimal feature extraction tool.

TABLE I

PRECISION AND RECALL VALUES FOR FEATURE EXTRACTION

Features	**Features extraction**	
	Precision	***Recall***
Uni-gram	60	65
Bi-gram	55	70
Tri-gram	80	75
Opinion words	65	70

D. Sentiment classification Techniques

Thirdly, From among different sentiment classification techniques, we have tested our experiments by only Support Vector Machines (SVM), Naive Bayes (NB), Complement Naive Bayes (CNB), and Maximum Entropy (ME) due to their wide acceptance and better performance in many similar and state-of-art researches. SVM also outperform other models in this case.

E. Polarity Detection

Polarities were detected after the sentiment being classified using different classifier.

F. Performance Analysis

This subsection analyzes the performance on the different combination of preprocessing, features, and machine learning algorithms and their relative performance comparison, which are discussed in details in Result and Discussion section.

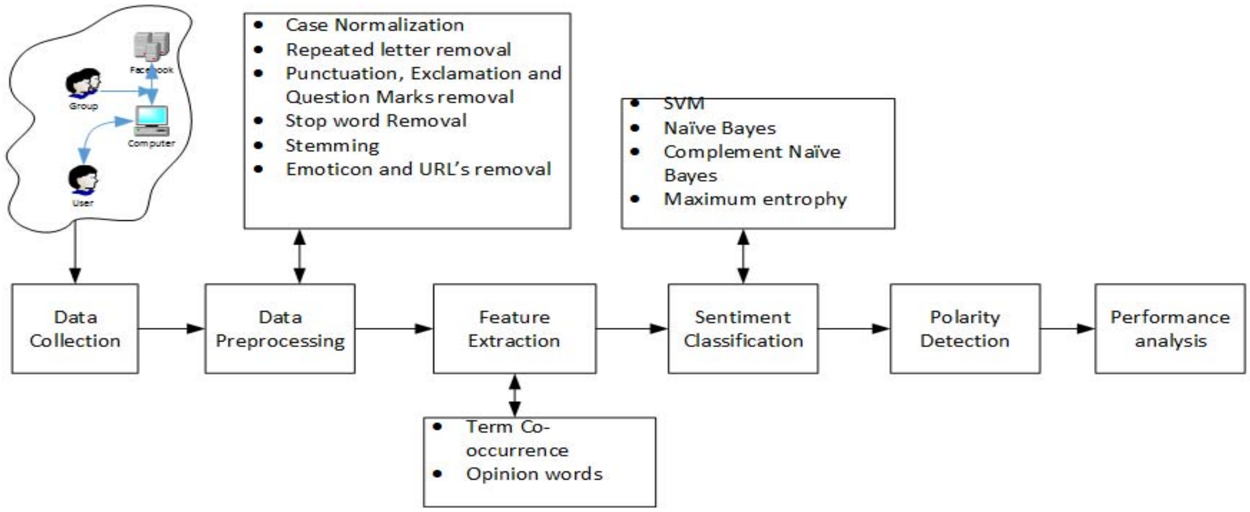

Fig. 1. Methodology of this Research

IV. RESULTS AND DISCUSSIONS

In this research, models were validated by measuring different matrix of the confusion matrix such as accuracy, recall, precision, and F-score and tested using 10-fold cross-validation. Table II represent the comparison between highest outcome, shown by the different models such as SVM, NB, CNB, and ME in the two state of art research and our study after pre-processing. From the Table II, Fig. 2 and Fig. 3, it is clear that, as per expectation, almost all the models shows better performance after preprocessing in both the study, but our study outperforms the stat-of-art researches by 2%, where our optimal model was SVM and their Optimal model was SVM with 95% accuracy [2] and CNB with 84% accuracy [13], this was due to the proper combination of pre-processing tools in our study. It was also observed that, CNB also shows good accuracy in all the research studied in this work. But, NB and ME suddenly shows improved accuracy in our work, than the previous research, which is 89% and 87% respectively.

TABLE II

COMPARISON IN HIGHEST RESULTS FOR EACH MODEL IN OUR STUDY AND STATE-OF-ART RESEARCH

Verification Parameters	**Result of State-of-art Research1**				**Result of State-of-art Research2**				**Result of Our Study**			
	Machine Learning Techniques				***Machine Learning Techniques***				***Machine Learning Techniques***			
	SVM	NB	CNB	ME	SVM	NB	CNB	ME	SVM	NB	CNB	ME
Accuracy	0.690	0.500	0.840	0.570	0.945	0.517	0.842	0.717	0.966	0.891	0.931	0.871
Recall	0.690	0.490	0.840	0.300	0.947	0.526	0.878	0.342	0.883	0.701	0.867	0.821
Precision	0.740	0.490	0.870	0.330	0.945	0.521	0.842	0.407	0.910	0.695	0.780	0.920
F-Score	0.570	0.480	0.840	0.310	0.944	0.499	0.848	0.372	0.910	0.705	0.885	0.920

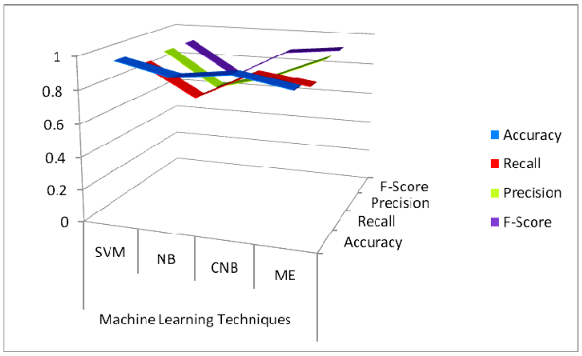

Fig. 2. Performance analysis of different Machine Learning Techniques in our Study

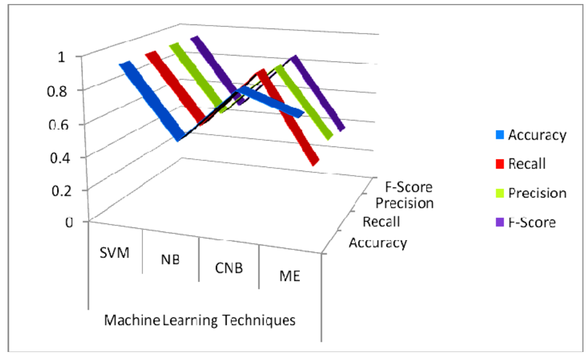

Fig. 3. Performance analysis of different Machine Learning Techniques in State-of-art Research2

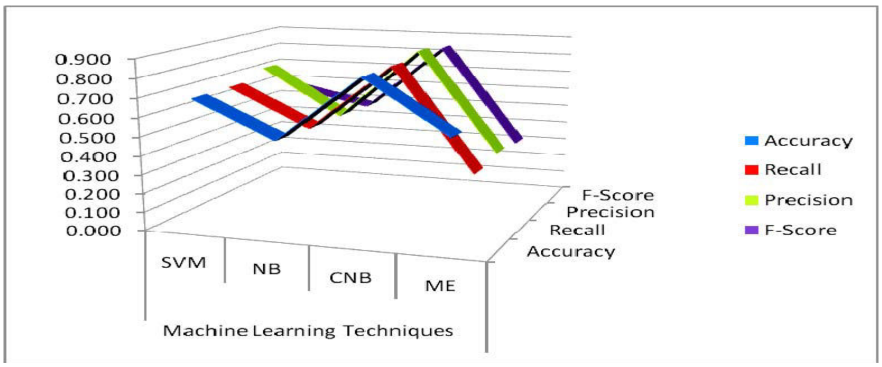

Fig. 4. Performance analysis of different Machine Learning Techniques in State-of-art Research

From the Table II, it is also seen that, SVM has a relatively high performance not only as because of the highest level of accuracy i.e.; 97% but also highest recall of SVM, which is 88%. But, in case the of precision and F-score, ME shows the highest value, which was 92% for both the cases respectively. From the analysis in our case, it was also seen that, CNB present nearby accurate result as SVM i.e.; 93%, this is because it dealt with the uneven class problem of this research. From Fig. 2, Fig. 3, and Fig. 4. Shows the performance analysis of different machine learning techniques in our study, state-of-art reserach1, and research2. In Fig 2. the peak of the curve shows nearby and very high for all the techniques, but in Fig. 3 and Fig. 4 the techniques are shown to fluctuates and differ in a great extent in their all of performance parameters.

TABLE III

RESULT OF DIFFERENT MODELS AFTER DIFFERENT LEVEL OF PREPROCESSING

Pre-processing	**Machine learning Techniques**							
	SVM				***ME***			
	Accuracy	***Recall***	***Precision***	***F-score***	***Accuracy***	***Recall***	***Precision***	***F-score***
P1	0.754	0.721	0.693	0.701	0.756	0.652	0.673	0.621
P2	0.823	0.772	0.792	0.810	0.773	0.755	0.736	0.711
P3	0.854	0.801	0.821	0.822	0.824	0.712	0.736	0.810
P4	0.893	0.859	0.831	0.840	0.843	0.782	0.801	0.815
P5	0.962	0.883	0.910	0.910	0.871	0.821	0.811	0.844

TABLE IV

RESULT OF DIFFERENT MODELS AFTER DIFFERENT LEVEL OF PREPROCESSING (CONTN.)

Pre-processing	**Machine learning Techniques**							
	NB				***CNB***			
	Accuracy	***Recall***	***Precision***	***F-score***	***Accuracy***	***Recall***	***Precision***	***F-score***
P1	0.698	0.444	0.432	0.435	0.701	0.650	0.555	0.624
P2	0.753	0.454	0.443	0.541	0.763	0.654	0.643	0.741
P3	0.784	0.652	0.564	0.555	0.840	0.662	0.764	0.755
P4	0.853	0.687	0.665	0.645	0.893	0.758	0.777	0.864
P5	0.891	0.701	0.695	0.705	0.931	0.867	0.780	0.885

Table III and Table IV shows the changes of performance, happens due to the different levels of preprocessing and model used. Here, we have included only the results that are remarkable. Almost every model show improved performance as the level of preprocessing proceeds. The change is slight in some cases but in most cases the preprocessing and model used, bring a great improvements of the analysis, which may be 20 to 30 percent and thus reflect that, the pre-processing are immense necessary in order to carry on feedback analysis of the students as the students in most of the cases doesn't express their opinion in plain or good English.

V. CONCLUSIONS AND FUTURE WORK

In this research, we have investigated different levels of preprocessing in combination with different feature extraction methods and machine learning algorithms on student feedback data to find out optimal tools among them for this sort of analysis. We found that, proper use of preprocessing tools, feature extraction tools and machine learning algorithms increased accuracy, which in turn shows improved performance of our analysis up to 20 to 30 percent, which we desired before analysis. It is also observed that, some level of preprocessing such as emoticons removal, in some cases loses valuable information, thus decrease the performance of the models. From the study we suggest that, the combination of preprocessing tools and feature extraction tools we have used in this study may be considered as optimal tools for this sort of analysis.

We also observed that, NB and CNB show highest performance in the case of uneven class problems, but in most of the cases it doesn't show the considerable result. But SVM and ME shows the acceptable results in most of the combinations of preprocessing, features and machine learning techniques. So, from the analysis we could say SVM and ME could be considered as best model and we suggest this could be used further for these sorts of feedback analysis.

In the future, there would be a possibility of including more preprocessing such as removing substring, feature extraction, such as syntactic dependency as well as sentiment word list extraction, and finally, more machine learning techniques such as an Artificial Neural Network (ANN) in this domain.

REFERENCES

[1] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Engineering Journal, vol. 5, pp. 1093-1113, 2014.

[2] N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Sentiment analysis: towards a tool for analysing real-time students feedback," in 2014 IEEE 26th international conference on tools with artificial intelligence, 2014, pp. 419-423.

[3] A. Hogenboom, D. Bal, F. Frasincar, M. Bal, F. de Jong, and U. Kaymak, "Exploiting emoticons in sentiment analysis," in Proceedings of the 28th Annual ACM Symposium on Applied Computing, 2013, pp. 703-710.

[4] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," Computers in Human Behavior, vol. 31, pp. 527-541, 2014.

[5] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno, and J. Caro, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning," in Information, Intelligence, Systems and Applications (IISA), 2013 Fourth International Conference on, 2013, pp. 1-6.

[6] J. M. Martín, A. Ortigosa, and R. M. Carro, "SentBuk: Sentiment analysis for e-learning environments," in Computers in Education (SIIE), 2012 International Symposium on, 2012, pp. 1-6.

[7] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in Proceedings of the workshop on languages in social media, 2011, pp. 30-38.

[8] W. Wang and J. Wu, "Notice of Retraction Emotion recognition based on CSO&SVM in e-learning," in Natural Computation (ICNC), 2011 Seventh International Conference on, 2011, pp. 566-570.

[9] A. Pak and P. Paroubek, "Twitter based system: Using Twitter for disambiguating sentiment ambiguous adjectives," in Proceedings of the 5th International Workshop on Semantic Evaluation, 2010, pp. 436-439.

[10] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford, vol. 1, p. 12, 2009.

[11] D. Song, H. Lin, and Z. Yang, "Opinion mining in e-learning system," in Network and Parallel Computing Workshops, 2007. NPC Workshops. IFIP International Conference on, 2007, pp. 788-792.

[12] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in Proceedings of the 42nd annual meeting on Association for Computational Linguistics, 2004, p. 271.

[13] N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Learning sentiment from students' feedback for real-time interventions in classrooms," in Adaptive and Intelligent Systems, ed: Springer, 2014, pp. 40-49.

[14] S. Fallahkhair, M. Cocea, and N. Altrabsheh, "Learning sentiment from students' feedback," in research gate, 2014, pp. 44-48.

View publication stats