Application of Language Modelling in Sentiment Analysis for Faculty Comment Evaluation

Ria A. Sagum, Jessiree Grace M. de Vera, Prince John S. Lansang, Danica Sharon R. Narciso, Jahren K. Respeto

Abstract—Sentiment Analysis is a field of Natural Language Processing that deals with the study of people's comments or opinions expressed in written text. In this study, the researchers utilized sentiment analysis to determine the rating of positivity or negativity of the comments given by the students in a Faculty Evaluation Form. This is in accordance with the objective of the research, which is to develop an application that will be used in assessing a faculty member using the comments in the faculty evaluation survey with the aid of sentiment analysis. The comments, stored and retrieved from the repository, were initially subjected to preprocessing. SentiWordNet then served as the source of the positivity scores of the comments. To increase the accuracy of the sentiment value, the comments undergone polarity identification, negation tagging, and intensity multiplication, taking into consideration the terms surrounding a given word.

***Index Terms*—Language Model, Polarity, Sentiment Analysis**

I. Introduction

Employers normally would want to ensure that each employee functions according to standards. To reach this goal, various ways are being practiced. One way is to observe the employees during work time. And to organize the observations gathered, employers often use an evaluation form, a kind of document an employer uses to measure how well an employee's performance at work. Many of the companies, business establishments and even schools use evaluation forms for this certain objective. And for schools and other learning institutions, one way of evaluating their faculty members is by using the Faculty Evaluation Form which is the main subject of this study.

Faculty evaluation is the process of gathering and processing data to measure the effectiveness of teaching. There are different areas to be considered for evaluating a faculty [1] such as teaching, advising and research and scholarly activities. The most important benefit of this evaluation is the feedback the forms provide directly to instructors, so that they can refine the courses and teaching practices to provide students with better learning experiences.

The great advancement of technology literally changed the way our world goes. Technology made it possible for people to communicate from the different sides of the globe. Most things that are manually done ages ago are replaced now by automated technology.

For the past years, the use of a computer for evaluation purposes has substantially increased. The aims of using automated evaluation include achieving the advantages of a system with the following characteristics [2]: first to increase the accuracy of evaluating the faculty; second, to reduce the school administrators' workload by automating part of the student evaluation task and finally to provide summary of the assessment.

This research shows the use of sentiment analysis to evaluate the student's narrative comment in the evaluation of their respective Faculty-in-charge.

II. Related Works

Sentiment Analysis nowadays is a very useful tool in different areas like politics, finance, business, media, and others because of the rapid growth of technology wherein people uses this branch of Natural Language Processing (NLP) to know the pulse and reviews of the people for a certain area.

There are different studies of Sentiment Analysis wherein the researchers used reviews to identify the sentiment of user where companies can use this so they can easily determine the sentiment of a review, whether it's positive or negative.

A study by Jebaseeli and Kirubakaran of India [3] conducted a survey on sentiment analysis of product reviews wherein they have discussed different researches that has different methods of sentiment analysis. And they came up with a conclusion that sentiment analysis or opinion mining do really helps in making decision about product or services, not only in getting concepts on text mining but also in information retrieval.

Manuscript received December 7, 2014; revised January 24, 2015.

R. A. Sagum is with the Department of Computer Science, Polytechnic University of the Philippines (e-mail: rasagum@pup.edu.ph)

J. G. M. de Vera is with the Department of Computer Science, Polytechnic University of the Philippines (e-mail: jessireedevera@gmail.com)

P. J. S. Lansang is with the Department of Computer Science, Polytechnic University of the Philippines (e-mail: enprixd@gmail.com)

D. S. R. Narciso is with the Department of Computer Science, Polytechnic University of the Philippines (e-mail: danicasharon_05@yahoo.com)

J. K. Respeto is with the Department of Computer Science, Polytechnic University of the Philippines (e-mail: jko_respeto@yahoo.com)

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Meanwhile, the study of Yessenov, et. al., [4] is about using sentiment analysis on movie review comments wherein the researchers used different machine-learning methods: Naive Bayes, Decision Trees, Maximum-Entropy and K-Means clustering. For the text corpora, they used the comments on articles from Digg. The results show that simple bag-of-words model can perform relatively well, and it can be further refined by the choice of features based on syntactic and semantic information from the text.

One interesting approach of Sentiment Analysis was presented in the work of Nasukawa and Yi [5] wherein they extract sentiments with polarities of positive and negative for specific subjects from a document, instead of classifying if the document is positive or negative. In this paper, they applied semantic analysis with a syntactic parser and semantic lexicon which gave them a high precision of 75% to 95% in finding the sentiments within web pages and news articles. They have also tested the system into an open test corpus related to camera reviews which is also from web pages wherein the system got 241 correct sentiments out of 255 cases.

Another application of Sentiment Analysis was used for Hotel Reviews and was implemented by Kasper and Vela [6] of Germany. In this paper, their system called BESAHOT is an interactive web application based on the GWT framework. The core system handles data acquisition, analysis and storage. The analysis part of the system has the sentiment analysis application wherein they analyzed the segments through a statistical polarity classifier and followed by linguistic information extraction components for finer grained analysis of the polarity and topic of the review. These polarity values are always assigned to text segments, and not the whole review. Polarities are classified into three: positive, negative and neutral. The system got an accuracy of 67%.

But as we know, not all sentiments or opinions can be only found on reviews. Most of the time, reviews or opinions can be found on Social Media like Facebook and Twitter where millions of people can see the user's opinion about something. NLP researchers thought that this would be a big advantage for them to extract more data and information by using Sentiment Analysis.

A paper by Skyes in 2012 [7] was about predicting the electoral outcomes using a sentiment analysis tool wherein they extracted a set of tweets from Twitter about the candidates of the user's interest. Using this data set, they have analyze the tweets and categorize each tweet whether it's positive or negative and predict the contest winners based on the aggregate sentiment scores for the candidate of interest over a period of time.

Twitter was an interesting website to researchers since the tweets are submitted in real-time and Sureka et. al., [8] used sentiment analysis in Twitter feeds to determine the reaction of the Twitter users about a certain search keyword. The system successfully explains why a particular keyword has positive or negative reactions through the supporting news section and also shows the variety of opinion of users over a single topic.

Also in 2012, Kucuktunc et al. [9], used a sentiment extraction tool to investigate the influence of factors such as gender, age, education level, the topic at hand, or even the time of the day on sentiments in the context of a large online question answering site. And after using the sentiment analysis tool, they came up with the conclusion that the best answers, which are preferred over the best answers, have lower sentimentality.

Lastly, sentiment analysis can also be used in the financial studies like what Atreya, et. al., [10] did. They conducted a study wherein they used sentiment analysis for financial signal prediction. For the testing and training data, they have gathered articles from New York Times about finance and tried two approaches to produce sentiments. One is the manual approach, where a person manually tagged the articles, and the automatic approach using the market movements. The researchers found out that the prediction accuracy using this approach is 70%.

III. DISCUSSION OF THE METHODOLOGY

The process started with the pre-processing of the input texts were the comments for the faculty, which were composed of one or many sentences connected to a certain person, specifically a professor. For this project, the comments were assumed to be correct in terms of spelling and grammar. No correction efforts were made with the comments inputted by the user. Example of words not accepted by the program are contractions such as doesn't or wasn't. Instead, the semantic analyzer accepts *does not* and *was not*.

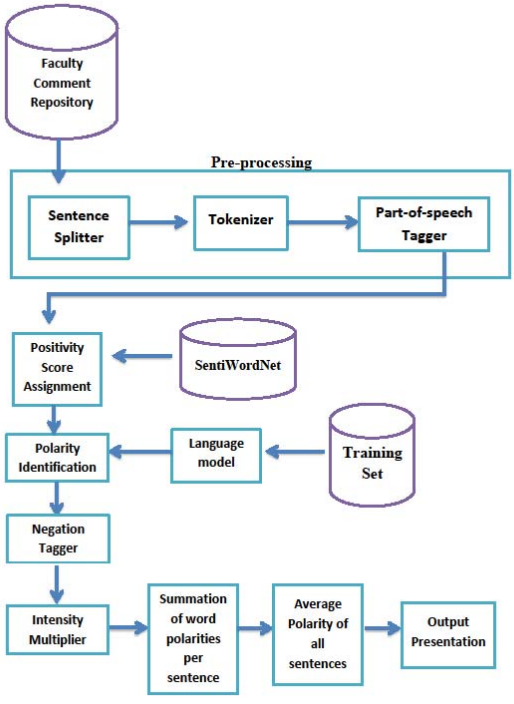

Fig 1. System Architecture

The first part of the preprocessing module is the sentence splitter. It is the process where the comment was broken down into smaller chunks, specifically in sentence level

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

under the process. Afterwards, these sentences broken down from the sentence splitter were further broken down into words. This process is called Tokenization. The output from tokenization will be used on the next process which is the Part-of-Speech tagging.

Part-of-Speech Tagging – This is a process where the words are tagged in their respective part of speech. There are tools already available that accomplish the said task, and one of which is the Stanford Universities' Part-of-Speech Tagger, which the researchers used. The Stanford Universities' Part-of-Speech Tagger is a piece of software, written in java language that reads texts in certain languages, and assigns the corresponding part-of-speech to each word. It tags, not just the 8 basic part of speeches, but as well as the sub-classifications. The tags were represented by certain acronyms or codes, which are contacted to each word, separated by an underscore '_'. The tagging was an essential part of the process, because only the parts of speech that may carry subjectivity will be in focus.

The next process does not utilize the tagged text, but rather the raw comment. In getting the polarity of the words, SentiWordNet was utilized. It is a lexical source for opinion mining, in which a database of words is maintained, and each word is assigned three sentiment scores: positivity, negativity, objectivity.

Positivity Score Assignment – A process where each word in the comment, is searched in the SentiWordNet, and assigned corresponding scores. The assigned value serves as the numerical data in determining the overall sentiment of the comments.

Language Model – The language model assigns probability to a set of string based on its occurrence in texts prior processed: in this case, the probability of positive, negative, or neutral occurrence is stored based on data that was fed beforehand. Each comment produces bigrams and trigrams, upon which the probabilities are constructed upon. To get the probability, the system counts the number of occurrence of the word divided by the total number of words. The language model dictated whether a certain word is positive or negative.

Then, the output of the Positivity Score Assignment and the output of the language model met in the process called Polarity Identification.

Polarity Identifaction – The process where the positivity score assignment is set. It is dictated by the numeric score or degree of positivity of the words.

Afterwards, the stream of texts were then processed by the negation tagger.

Negation Tagger – This tagger searched for keywords that suggest negativity in the next word, or set of words. This words were "not", "neither", and "nor". However, not all of the proceeding words were affected by these keywords, especially if the words were objective. The parts of speech affected by negation keywords are: Verb (gerund), Verb (3rd person singular), Verb (in form), Adverb, and Adjective. To identify if a word is negated, the substring "_1" is attached to the word. Otherwise, the substring "_0" is attached to the word. In the next step, the polarity of the words with substring "_1" were inverted. This means that if a word was tagged as positive before the negation tagger, it was then tagged as negative by the Negation Tagger, and vice versa.

Intensity Multiplier – This process searched for keywords that denote intensity or degree, and therefore increase/decrease the positivity or negativity of a certain word/s. Example of these words were absolutely, fully, completely, extremely, fairly, entirely, greatly, incredibly, indeed, perfectly, much, quite, really, strongly, terribly, totally, and very. The said words, however denoted different intensities. For example, the word quite denoted uncertainty, or partiality in the truth of a statement. In that case, it was assigned a multiplier of 0.5 to the polarity of the word. For words, on the other hand, that expressed superlative degree or high intensity, a multiplier of 2 or 3 was assigned to the word/ set of words following the intensity keywords.

Next, the polarity of the words in each sentences were summed up, and divided by the total number of subjective words. Then, it was classified as Strongly Positive, Positive, Negative or Strongly Negative. Then, the polarities of all the sentences were averaged. After the averaging, the final output was classified based on three classifications: Positive, Neutral or Negative.

IV. TEST RESULTS

Initially, the researchers gathered comments from the Faculty Evaluation Survey for computer science students. Seventy percent (70%) were used as training data and the remaining 30% used as the test data.

The test data used were manually tagged as positive or negative by the researchers using the school management's perspective. And by using the sentiment analyzer, the results shows that the *Sensitivity* (or Recall), which is the *True Positive Rate* shows a rating of 57.78%. However, the *True Negative Rate* or the Specificity rate is 77.78%, and from this, we can conclude that the system can accurately determine if the comment is negative than determining if the comment is positive. The researchers got the recall rate by counting the true positives divided by the sum of true positives and false negatives.

As mentioned above, the system's true positive rate should be higher than the results shown but because the researchers did not apply any smoothing methods on the language model, thus, a lot of false positives turned up.

The results of the initial testing is quite below the standards of an accurate system, especially on the True Positive Rate, thus, the researchers have done some revisions on the sentiment analyzer to improve its accuracy.

On the final testing, the researchers collected additional test data and noticed that there are words that weren't accepted by the analyzer like *It's, it's, she's* and *he's*. After further testing on this limitation, the researcher's have

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

discovered that the reason behind this is because the preprocessing of the system does not match with the training data. The system normalizes first the words, unlike in the training data where the comments collected were inputted as it is.

Another concern that was discovered by the researchers is that the system can't accept intensifier words like *much, more, fully, extremely,* etc., if there is no word after them. In the module, Intensity Multiplier, the intensifier always checks the word after it. In the comment: *He is a good professor. He knows his lessons so much.*, since *much* is an intensifier, the system checks the next word and finds nothing, then, the system didn't accept the comment.

Again, after testing these comments, here are the results:

TABLE I: FINAL TESTING RESULTS

	Results
True Positive Rate	61.22%
True Negative Rate	81.25%
Precision	83.33%
F1 Score	70.58%

On the table above, the researchers included the precision and the F1 score of the system. This precision rate is the positive predictive value of the system. High precision means that an algorithm returned substantially more relevant results than irrelevant.[11] On the other hand, the F1 Score is a measure of a test's accuracy. It is widely used on different Natural Language Processing systems to determine the accuracy of the system. [12] And as seen on the table, the system is 70.58% accurate but still has a lot to go for improvement. The research is still ongoing and the researchers are finding ways to improve the accuracy of the sentiment analyzer.

Still, the revision didn't include any smoothing yet, so the system can still be improved. But in this revision, the false positives were quite lower than the initial test results.

V. CONCLUSIONS

The sentiment analyzer was able to serve its purpose, in determining the sentiment of the students in the faculty comments. However, there are numerous words that the program was not able to tag correctly. There are also bigrams that have varying results like *doesn't teach*, *only boasts*, *good prof, very intelligent* and *can't understand* which are tagged differently in two different test runs.

According to Ogneva of Biz360 [13] , human raters typically agree 79% of the time when it comes to sentiment analyzers. Thus, a sentiment analyzer that has 70.85% accuracy rate is quite doing well as humans do in analyzing sentiments.

VI. RECOMMENDATIONS

To reduce the possibility of the said limitation(s), the researchers recommend the use of a larger set of training data, from a broader category of students. This will give the sentiment analyzer, a better basis in determining the positivity of a word or set of words. In addition to this, there is a high likelihood that the students would use their native language, like in the Philippines – the Filipino language in the comments,

a situation in which the program can't handle, because it is solely designed for English language. Therefore, the researchers also recommend an improved model that can process both English with another language in a sentence. This will lead to a bilingual sentiment analysis for this kind of data. There's also a possibility to enhance the accuracy of this research if noise words like hashtags (#), internet slangs (e.g. LOL, LMAO) and emoticons will be considered (due to the influence of social media). The researchers also recommend to include a smoothing method like Kneeser-Kney [14] to the language model for better results.

ACKNOWLEDGMENT

Our sincere thanks to our family, friends, CCIS family and batch CCIS/DCS batch 2015, who is always there to support us. And most especially, to our God Almighty who gave us the knowledge and guidance in our everyday lives.

REFERENCES

[1] Barrentine, Shelby, Chalmers, Lynne and Zidon, Margaret, *Faculty Evaluation Guidelines and Handbook*, 2010

[2] Brown, S., Race, R., and Bull, J. *Computer-assisted assessment in higher education*: Kogan Page, 1993.

[3] Jebaseeli, A. Nisha; Kirubakaran, Phd., E., *"A Survey on Sentiment Analysis of (Product) Reviews"*, International Journal of Computer Application(0975-888), pp 36-39, vol. 47: June, 2012.

[4] Yessenov, Kuat and Misailovic, Sasa, *"Sentiment Analysis of Movie Review Comments"*, May 2009.

[5] Nasukawa, Tetsuya and Yi, Jeonghee, *"Sentiment Analysis: Capturing Favorability using Natural Language Processing"*, K-CAP-03, 2nd International Conference on Knowledge Capture, pp 70-44: 2003.

[6] Kasper, Walter and Vela, Mihaela, *"Sentiment Analysis for Hotel Reviews"*, Proceedings of the Computational Linguistics--Applications Conference, pp 45-52: 2011.

[7] Skyes, Jenn. *"Predicting Electoral Outcomes with SAS Forecast Studio"*, SAS Global Forum 2012: 2012.

[8] G., Kaushik, S., Patira, and Sureka, R., *"Twitter Sentiment Analysis"*, Georgia Tech.

[9] Kucuktunc, Onur, Cambazoglu, B. Barla, Weber, Ingmar, and Ferhatosmanoglu, Hakan, *"A Large-Scale Sentiment Analysis for Yahoo! Answers"*.

[10] Zhai, Jinjian, Cohen, Nicholas, and Atreya, Anand, *"Sentiment analysis of news articles for financial signal prediction"*.

[11]Precision and Recall [Online] Available: http://en.wikipedia.org/wiki/Precision_and_recall

[12] F1 Score [Online] Available: http://en.wikipedia.org/wiki/F1_score

[13] M. Ogneva. (April 2010). How Companies Can Use Sentiment Analysis to Improve Their Business. [Online] Available: http://mashable.com/2010/04/19/sentiment-analysis/

[14] MacCartney, Bill, *NLP Lunch Tutorial: Smoothing*. : 2005.

Jessiree Grace M. de Vera. Born in Sta. Monica Hospital, Tanong, Marikina City on January 14, 1995. She finished her primary education at Barangka Elementary School in Marikina City and her secondary education at Gen. Licerio Geronimo Memorial National High School in Rizal. She is currently enrolled at the Polytechnic University of the Philippines, Sta. Mesa, Manila taking up Computer Science. She is currently on her senior year.

She took up her summer internship at Novare Technologies located at Makati City as a Software Engineer Intern and served as a project manager on her team. Her current research interests include Sentiment Analysis and Information Extraction in Natural Language Processing.

Prince John S. Lansang. Born in Pasay City on October 12, 1993. He graduated elementary in Notre Dame of Greater Manila and finished her secondary education in Ramon Magsaysay (Cubao) High School in Quezon City. He is currently enrolled at the Polytechnic University of the Philippines, Sta. Mesa, Manila taking up Computer Science. He is currently on his senior year.

He took his summer internship at Deltek Inc. as a member of the Automation Team. He is also a member of the special interest group of the Department of Computer Science at the Polytechnic University of the Philippines that specializes in Game Development. His current research interests are Natural Language Processing and Game Development.

Danica Sharon R. Narciso. Born in Dr. Jose Fabella Memorial Hospital, Sta.Cruz, Manila on December 29, 1994. She graduated elementary in Cabaritan West Elementary School in Ballesteros, Cagayan and finished her secondary education in Ballesteros National High School in the same town. She is enrolled at the Polytechnic University of the Philippines, Sta. Mesa, Manila and took BS in Computer Science. She is currently on her senior year.

She took her summer internship in DevKinetics Inc. as a JAVA EE / ANGULARJS DEVELOPER. Her current research interests are applications in Natural Language Processing.

Jahren K. Respeto. He was born in Chinese General Hospital, Sta. Cruz Manila on November 29, 1994. He attended Northern Rizal Yorklin School for his basic education and Caloocan City Science High School for his secondary education. He is **currently a senior student of the Polytechnic** University of the Philippines (Manila Campus), taking up Bachelor of Science in Computer Science.

He took his summer internship at DevKinetics Inc. as the Lead R & D (Read and Develop). His current research interest is Sentiment Analysis in Natural Langauge Processing.

Ria A. Sagum. She was born in Laguna, Philippines on August 31, 1969. She took up Bachelor of Computer Data Processing Management from the Polytechnic University of the Philippines and Professional Education at the Eulogio Amang Rodriguez Institute of Science and Technology. She received her Master's degree in Computer Science from the De La Salle University in 2012. She is pursuing her postgraduate studies and is taking Doctorate in Information Technology at De La Salle University. She is currently teaching at the Department of Computer Science, College of Computer and Information Sciences, Polytechnic University of the Philippines in Sta. Mesa, Manila and a lecturer at the Institute of Information and Computing Studies, University of Santo Tomas in Manila. Ms. Sagum has been a presenter at different conferences, including the 6th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management 2013 (HNICEM), 2012 International Conference on e- Commerce, e-Administration, e-Society, e-Education, and e-Technology and National Natural Language Processing Research Symposium. She is a member of different professional associations including ACM-CSTA and a board member of the Computing Society of the Philippines- Natural Language Processing Special Interest Group.

ISBN: 978-988-19253-2-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)