BLOG COMMENTS SENTIMENT ANALYSIS FOR ESTIMATING FILIPINO ISP CUSTOMER SATISFACTION

FREDERICK F. PATACSL, PROCESO L. FERNANDEZ

Pangasinan State University, Ateneo de Manila University
E-mail: Frederick_patacsil@yahoo.co.uk, pfernandez@ateneo.edu

Abstract- Blog comments have become one of the most common means for people to express and exchange their personal opinions. The study proposed and developed a system for automated opinion retrieval from blog comments to estimate the customer satisfaction for three main Filipino Internet Service Providers (ISPs). Data were first gathered from comments located in some of the most popular blog sites discussing the Filipino ISPs. Automatic word seeding, N-gram tokenization, stemming and other Sentiment Analysis (SA) techniques were applied to extract useful information from the textual data. The data collected were manually labeled in order to establish ground truth. Furthermore, the study experimented bag of words and Ruled-based method of identifying the polarity of the blog comments sentences. In addition, Naïve Bayes (NB) and Support Vector Machine (SVM) language, N-gram and preprocessing features such as stopwords and stemmer were applied and experimented. The results of the experimentation showed N-gram has a significant effect on the increase of performance of the proposed automated classifier. In contrast, preprocessing features such as stop words and stemmer did little help to increase the performance of the said proposed automated classifier. The application of SVM learning machine and bi-gram feature on the proposed model were resulted in higher classification performance and far above the baseline classification performance of the previous researches. The impact of adding ruled based can indeed be a big help to automatically determine the sentiments of blog sentences. These good results indicate that it can be possible for some interested parties to have a sense of the sentiments of their customers by applying some automated sentiment analysis on blog comments.

Index Terms- Bag of Words, Blogs, Naïve Bayes, Rule-based, Sentiment Analysis, Support Vector Machines, Stopwords, Stemmer, Translator Machine.

I. INTRODUCTION

A blog (a truncation of the expression weblog) is a discussion or informational site published on the World Wide Web (WWW) and consisting of discrete entries ("posts") typically displayed in reverse chronological order (in some cases, the most recent post appears first) [1]. Beneath the blog posts is an area for customers/readers to make their own comments about the topic of the post. These comments make a blog very interesting, interactive and interrelated. Blog comments are the lifeblood of a blog and the most popular blogs have a very active community of people who voice their opinions on the posts.

Considering this social aspect, it makes blogs such a powerful component of the social web. Zhou et. al. viewed a blog post as a summary of online news articles it linked to, with added personal opinions [2]. Universal McCann released a report stating that social media is a global phenomenon. The report found that social media, in particular blogs, are “becoming more important part of global media consumption for Internet users than some traditional media channels.”

The Internet customer comments on online review are an influential factor that may affect decision of other customers about the product and the services of a certain entity. Weblog comments serve as “a simple and effective way for webbloggers to interact with their readership” [88] and they are one of the defining set of weblog characteristics [3]. Therefore, these sentiments are very important sources of information that should be taken into account by companies in improving their services and development of their products. However, existing research largely ignores comments by focusing on blog posts only. This is due to the sheer volume of online comments that makes it difficult for any human to process and extract all meaningful information. As a result, there has been a trend toward the development of systems that can automatically summarize opinions from a set of reviews and display them in an easy to process manner has been emerged.

The process of analyzing and summarizing opinions is known as Sentiment Analysis (SA), a type of natural language processing for tracking the moods and sentiments of the public about a particular service, product or topic and the application of Natural Language Processing (NLP), computational linguistics, and text analytics to identify and extract subjective information in source materials such as those discussions about certain products and services [4]. However, automated sentiment analysis of web data can be a hard endeavor due to the large volumes of noisy information it contained. Consequently, previous studies have predominantly incorporated manual or semi-automated methods [5]. Manual examination of thousands of information can be an extremely tedious effort when applied across thousands of postings. With the increasing web usage, the need for automated text classification and analysis techniques has grown in recent years.

Blogs are one of the platforms that express personal opinions about a specific topic, and the comments
section of the blogs typically contains the reaction and opinions of the readers of a blog. Naturally, such comment section contains various kinds of expressions that are either positive or negative in nature, and these can be indicative of the customer satisfaction. In this study, SA was used to analyze the data gathered from the comments section of selected blogs.

II. RELATED LITERATURE

A. Sentiment Analysis using Micro Blogging

As the blogging phenomenon continues its exponential growth, its increasingly influential role in the global marketplace of ideas and opinions is widely being acknowledged. Harvard Business Review of February 2005 notes that “the blogosphere is gaining the power to influence what people think and do”. Bloggers are driven by a desire to share their ideas and opinions with anyone who cares to tune in” [6]. Web Blog Mining which is the efficient and effective way of analyzing the sentiments of consumer reviews pertaining to specific products becomes desirable and essential.

B. Machine Translation

The primary focus of this study is to use sentiment analysis in the context of tracking of blog comment reactions of Filipino customer satisfaction with regard of the services provided by their respective ISPs. However, the convenience of mixing Tagalog and English in the Philippines is very prominent and it has been observed by linguists that today writings in “Taglish” are commonly encountered in tabloids, comics, and the Internet [29].

Attempts of using machine translation in different natural language processing tasks have not been widely used due to poor quality of translated texts, but previous studies show that machine translation systems are reaching a reasonable level of maturity when employed for multilingual sentiment analysis [7, 8, 9, 10].

In reference this, the study utilized online Google machine translator to convert Filipino blog comments to its English form.

C. Seed of Words

Several studies adopted the idea of semantic orientation to choose the initial set of seeds and picked sets of reference seeds, single word, one positive and one negative or multiple set of seeds. Then, expanding seed lists through searching for and retrieving the synonyms and antonyms of the initial seeds from several sentiment dictionary tools such as Wordnet, SentiNet, and dictionary glosses such as online dictionaries. Zagibalov and Carroll use the word “Good” as their initial word seed [11], Turney [8] utilized two arbitrary seed words (“Poor” and “Excellent”) to calculate the semantic orientation of phrases. Rothfels and Tibshirani hand also adopted this approach where they picked two sets of reference seeds, one positive and one negative[12].

D. Bag of Words

Bag of words is a model that takes individual words in a sentence as features, assuming their conditional independence. The text is represented as an unordered collection of words. Each feature of the vector represents the existence of one word.

Pang et al. [13] show that manual keyword model is outperformed by statistical models, where a good set of words that represent features are selected by their occurrence in the existent training corpus.

E. Part of Speech Tagging

Annotating corpora was done manually prior to the availability of automated methods. Tagging texts by hand is a very slow and laborious process. A Part of Speech Tagger (POS) takes a word in a phase and classifies it as a noun, adjective, verb, and so on. The categories into which the items are grouped are called “labels.” POS was first explored in 1962-1963 using hand-written rules. Thereafter, stochastic Markov models proved to be more accurate as large corpora became available [14]. In some study, POS features have been used effectively in sentiment classification [15].

F. Ruled-Based Approaches

Rule-based approaches targeting the analysis of contextual sentiment were proposed by several researchers [16, 17]. They developed a method to extract and classify local sentiment expressions (as positive or negative) for giving specific subjects using various grammatical levels (adjectives, adverbs, and verbs). Naturally, a document with higher number of adjectives and adverbs are more likely to be more subjective than a document with little amount of these parts of speech. Similarly, the fundamental function of adjectives and adverbs is to denote the qualities of entities and events [18, 19].

G. Machine Learning Approach

Machine Learning techniques for classification use a training set and a test set. Training set contains input feature vectors and their corresponding class labels. Using this training set, a classification model is developed, which tries to classify the input feature vectors into corresponding class labels. Subsequently, a test set is used to validate the model by predicting the class labels of unseen feature vectors. Furthermore, the machine learning approach involves constructing a model from a training corpus, which is basically an electronically stored set of texts [20]. In order to train an algorithm to identify features that associate with positive, negative and neutral categories, such set of
texts, annotated for polarity by human coders, are used [21].

H. N-gram and Preprocessing Features.
H.1 N-gram
In Natural Language Processing (NLP) classification tasks used bag of words approach applying N-gram as their special feature [22, 21, 23]. In this way, the size of a feature space is big because it depends on the vocabulary size of the corpus. Therefore, choosing the features that are directly related to sentiment analysis is important, because it can improve performance and time and space efficiency.

H.2 Stemming
Stemming is a linguistic technique that can be used to reduce the number of words. The process reduces a word into a base form or stem. Some studies found that stemming has mixed success in both information retrieval and text mining [24, 25] and significantly increases the retrieval results for both rule based and statistical approach [26].

H.3 Stopwords Removal
Stopwords removal is the process of removing words that have high frequency which are not important to the sentiment of the sentence. Words such as; ‘a’, ‘the’, ‘or’ are likely to be considered as stopwords which have been listed in as non-sentimental word. Several studies showed that it is beneficial not to remove classical pre-compiled stopwords from micro blogging in the sentiment classification setting [27, 28].

III. GENERAL METHODOLOGY
A. Information Extraction
Web 2.0 with its emphasis on sites that promote information sharing, communication with the customers, and user collaboration. The utilization of Web 2.0 Internet applications such as blogs, social networking and forums is quickly increasing and it empowers Internet users/customers through the formation of communities and the mass publication of user-generated content.

It’s that social aspect that makes blogs such a powerful component of the social web. People like to feel involved. Leaving blog comments allows readers to join in the conversation about a topic that interests them [29]. The study considered only the comments from the blog as an input for the training and testing of the proposed model.

B. Information Selection and Collection
In this study, the first task was to search for blog articles using a search engine that discuss and compare the services of the three main ISPs. Blog articles were manually evaluated and filtered such that blog articles that contain many comments from their customers/readers were highly considered as part of the dataset.

For extraction and collection of blog comments, an application system was developed to extract comments from the blog article web sites and automatically stored into a Mysql database. The URLs address of the blog articles was the input in the PHP application program for blog comments extraction.

C.Blog Comments Cleansing and Tokenization
C.1 Data Cleansing and Filtering.
Data obtained from blog comments usually contain syntactic features, html code and entities like &lt; &gt; &amp which get embedded in the original comments. Thus, it is necessary to remove those contents from the data because they may affect the result of sentiment classification and are not useful for the machine learning for sentiment analysis.

C.2 Tokenization
For English, an uncomplicated and effective tokenization technique is the used of white space and punctuation as sentence delimiters [30]. Blog sentence tokenization is useful for obtaining an array of sentences from the given blog comments. The basic and simple sentence delimeter used is (".") plus white space. But it is the limited way of dividing the sentences of a paragraph of a text. Therefore, to handle cases correctly, the extended splitters used are (".?") ("!?") ("?").

D.Language Translation
Language Translation. The comments were not all written in English. There were some Filipino words, and in some instances even sentences, that were observed from the collected data. A machine translation using Google Translate was employed. Specifically, a modified application in PHP was used to automatically convert Filipino sentences into their
Blog Comments Sentiment Analysis For Estimating Filipino Isp Customer Satisfaction

English equivalent using the Google Translate API. Filipino words that were not recognized by the machine were manually corrected and converted using the same tool.

E. Seed of Words

Before carrying out a polarity classification on the dataset, there is a need to create a word dictionary to be used as the base for sentiment analysis. This research utilized automatic identification of seed of words or dictionary based on the data set. The system counts the number of adjectives occurrences (one negative and one positive) and these were used as the initial seeds of the seed of words.

The initial seed set consisted of the keyword {"Good" and "Slow"}. An application program automatically searches for and retrieves the synonyms and antonyms of the initial seeds from an online thesaurus dictionary using http://thesaurus.altervista.org/thesaurus/v1 as part of the word of seeds of this research. After the first process, the application retrieved the first synonym word found and repeats the process of searching and retrieving of synonyms and antonyms from the online thesaurus dictionary. The process is repeated until no more new words have been added to the word collection.

F. Generation of Blog Comments Corpus through Automatic Polarity Detection

F.1 Bag of Words

The bag of words focused completely on the opinionated words in the blog comment sentence. These words have their own polarity value (positive 1, negative -1) in the seed of words. When these words are found in the sentences, values of the opinionated words are typically all added up and the result is a sentiment estimation of the blog comment sentence.

F.2 Part-of-Speech Tagger and Ruled-based

The adverbs, adjectives, nouns and verbs are the most significant set of words that may have an effect on the sentences and Part of Speech (POS). Tagger identifies and explains the significance of the words in the sentences. POS tagging was used to extract adverbs, adjectives, verbs, nouns and negation terms in the sentences. Extracted phrases of the sentence containing and combining adjectives, adverbs, verbs, and nouns are good indicators of sentiments. In addition, retrieve sentiment words that were included in the seed of words (dictionary) were identified in the sentence and treated with ruled based analysis in sentiment extraction to ensure that the blog sentence contains only meaningful sentiment words. At the phrase level the general order of sentiment computing adjectives, adverbs, verbs, and nouns combinational rules are the following:

- Adjective and Adjective combination
- Adjective and Verb combination
- Adjective and Noun combination
- Adjective and Adverb combination
- Adverb and Adverb combination
Blog Comments Sentiment Analysis For Estimating Filipino ISP Customer Satisfaction

- Adverb and Noun combination.
- Verb and Noun combination.
- Negative

It is a feature to present if the opinionated clue is preceded by a negative tool. This rule is used to identify whether a word is negated in the sentence.

1. Spatt(l)= "JUI"
2. Spatt(3) = "NNN"
3. Spatt(6) = "VBN"
4. Spatt(7) = "VBN"
5. Spatt(10) = "JJS"
6. Spatt(12) = "VBR"
7. Spatt(13) = "VBR"
8. for(Sx2 = 0; Sx2 < count(Sxdummy1)-1; Sx2++) {
9. Spatial = X.trim(Sxdummy1[Sx2]);
10. $pattern1(Sx2) = $x;
11. $sentiment[Sx2] = $sentiment1(Sx2) - 1;
12. if(preg_match($pattern1,trim(strtolower($Stagdummy[$Sx1])))) {
13. if (preg_match($pattern1,trim($Stagdummy[$Sx1]))) {
14. else
15. $sentiment1[Sx2] = $sentiment1[Sx2] + 1;

Figure V. Sample code Automated Sentiment Polarity Identification Using Bag of Words+Rule-Based Approach

Table I. Distribution of Automatically Labeled Dataset using Bag of Words and Rule-Based Approach.

<table>
<thead>
<tr>
<th>Sentences</th>
<th>No. of Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>208</td>
</tr>
<tr>
<td>Negative</td>
<td>2352</td>
</tr>
<tr>
<td>Positive</td>
<td>1794</td>
</tr>
<tr>
<td>No Sentiment</td>
<td>10163</td>
</tr>
<tr>
<td>Total</td>
<td>14517</td>
</tr>
</tbody>
</table>

G. Manual Annotation of Blog Comments

G.1 Sentence Level Annotation

The collected blog comments were manually annotated by four groups with three members per group. The task was to identify and select the polarity of the blog comments with the aid of an application program. The blog comment sentences were displayed on the screen for the annotators to read the whole sentence, then decide and select the polarity.

Table IV. Fleiss’ Kappa Agreement Results of Three Group Sentiment Polarity Annotator

<table>
<thead>
<tr>
<th>Annotators</th>
<th>Expected Agreement</th>
<th>Observation Agreement</th>
<th>Fleiss’ Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>0.965</td>
<td>0.754</td>
<td>0.857</td>
</tr>
<tr>
<td>Group 2</td>
<td>0.892</td>
<td>0.547</td>
<td>0.76</td>
</tr>
<tr>
<td>Group 3</td>
<td>0.914</td>
<td>0.681</td>
<td>0.73</td>
</tr>
<tr>
<td>Group 4</td>
<td>0.815</td>
<td>0.552</td>
<td>0.586</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td>0.74</td>
</tr>
</tbody>
</table>

Legend: Landis and Koch [31] suggest the following interpretations:

Kappa Agreement

< 0   Less than chance agreement
0.01-0.20  Slight agreement
0.21-0.40  Fair agreement
0.41-0.60  Moderate agreement
0.61-0.80  Substantial agreement
0.81-0.99  Almost perfect agreement

The results reveal that group 1 has an “almost perfect agreement” with a kappa value of $\kappa = 0.857$. Groups 2 and 3 have “substantial agreement” with kappa values of $\kappa = 0.76$ and $\kappa = 0.73$, respectively. $\kappa = 0.586$ is the lowest kappa value and was obtained by group 4. The average inter-rater reliability for the sentence polarity annotators was found to be $\kappa = 0.74$. According to the definition of the Fleiss’ Kappa statistic, the accuracy of the inter-rater reliability is considered to be “Substantial agreement”.

Table II. Distribution of Manually Labeled Negative Sentences

<table>
<thead>
<tr>
<th>Sentences</th>
<th>No. of Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>2793</td>
</tr>
<tr>
<td>Positive</td>
<td>539</td>
</tr>
<tr>
<td>Total</td>
<td>3332</td>
</tr>
</tbody>
</table>

Blog comment sentences that have no sentiment and neutral where bypass by the annotators and disregarded by the application system.

G.2 Phrase Level Annotation

The results of sentence level manual annotation will be further subjected to phrase level polarity by three previous annotations. They were instructed to identify the polarity as positive or negative using the customized PHP application to label the sentiment polarity of the sentences.

However, the application will display sentiment words as a clue for the annotators to identify the polarity of the sentences. In the phrase level, inter-rater agreement on manual classification is $K = 0.905$ and is considered to be “Almost agreement”. Further, the results of phrase level annotation is the gold standard dataset of the study.

Table III. Distribution of Manually Labeled Positive and Negative Blog Comment sentences using a Phrase Level.

<table>
<thead>
<tr>
<th>Sentences</th>
<th>No. of Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>1370</td>
</tr>
<tr>
<td>Positive</td>
<td>527</td>
</tr>
<tr>
<td>Total</td>
<td>1897</td>
</tr>
</tbody>
</table>

H. Training and Testing the Automated Comment Sentence Datasets

The training and testing of the proposed sentiment analysis model were carried out using 10-fold cross validation utilizing Rapid Miner 5.3. The experimentation was conducted to train the automated bag of words and ruled-based dataset applying N-gram and preprocessing features (Stopwords and Stemming). Furthermore, the sentiment analysis models that were built during the testing phase were tested against the manually annotated (Gold Standard) dataset.
IV. EXPERIMENTATION AND RESULT

To gain more insight into the proposed model performances, the summary of the experimental results are shown in Table V and Table VI.

Table V. Summary F-measure Performance Measures of the Proposed Sentiment Classifier Tested on Manually Labeled Classified Dataset

<table>
<thead>
<tr>
<th>Preprocessing Features</th>
<th>Bag of Word</th>
<th>Bag of Word-Ruled Based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NB</td>
<td>SVM</td>
</tr>
<tr>
<td></td>
<td>NB</td>
<td>SVM</td>
</tr>
<tr>
<td>U</td>
<td>77%</td>
<td>60%</td>
</tr>
<tr>
<td>B</td>
<td>84%</td>
<td>65%</td>
</tr>
<tr>
<td>T</td>
<td>84%</td>
<td>66%</td>
</tr>
<tr>
<td>U+S</td>
<td>76%</td>
<td>60%</td>
</tr>
<tr>
<td>B+S</td>
<td>83%</td>
<td>62%</td>
</tr>
<tr>
<td>T+S</td>
<td>82%</td>
<td>63%</td>
</tr>
<tr>
<td>U+PS</td>
<td>76%</td>
<td>59%</td>
</tr>
<tr>
<td>B+PS</td>
<td>84%</td>
<td>66%</td>
</tr>
<tr>
<td>T+PS</td>
<td>84%</td>
<td>66%</td>
</tr>
<tr>
<td>U+5+S+PS</td>
<td>74%</td>
<td>59%</td>
</tr>
<tr>
<td>B+5+S+PS</td>
<td>84%</td>
<td>66%</td>
</tr>
<tr>
<td>T+5+S+PS</td>
<td>82%</td>
<td>63%</td>
</tr>
</tbody>
</table>

It is noticeable in Table V that there was a significant increase in classification F-measure performance of bi-gram and tri-gram features as compared to uni-gram in all preprocessing features. Furthermore, combining N-gram to other preprocessing features (stemmer and removing stopwords) produce almost similar F-measure performance results as compared with N-gram results as shown in the same table. The implication of these results suggests that, incorporating stemmer and stopwors removal preprocessing features did little help to increase the F-measure performance of the proposed automatic classifiers.

In terms of method used, the obtained F-measure under bag of words + ruled-based was slightly higher than the bag of word method in almost all features both NB and SVM. These results indicated that adding ruled based on the bag of words improved the performance of the proposed automatic classifier.

Interestingly, the results revealed that N-gram and pre-processing features have almost the same pattern of increase and decrease in terms of F-measure performance in all measure areas. The same pattern indicated that pre-processing features such as stemming and removing stopwords did not help to improve the classifying performance of the proposed automatic classifiers.

Table VI shows the summary of classification accuracy results which indicates that removing stopwords and stemming processing features made little help to increase classification performance of both models.

However, adding of N-gram seem helpful to enhance the classification accuracy performance specially NBs for bag of words and SVMs for bag of words+ruled based model. Bag of words+ruled based model outperformed bag of words in all measure areas as shown in the above table.

In terms of method used, the obtained classification accuracy obtained under bag of words + ruled-based was slightly higher than bag of word method in almost all features both NB and SVM. However, SVMs bi-gram, tri-gram + stopwords removal and tri-gram+stemming under bag of words+ruled-based registered an average increase of 8 percent. This result shows the impact of adding ruled-based on bag of words and indeed improves the classification performance of the proposed model. But, opposite results were seen in the preprocessing features, where in stemming and removing stopwords contributed nothing to the increase of classification performance of the proposed methods.
CONCLUSION

Classification performance results showed that SVMs outperformed NB in almost all performance measures. This confirmed other research results that SVM tends to be the best and obtained higher classification performance. This proved that SVM is effective in classifying shallow sentences like blog comments. Furthermore, F-measure and accuracy results showed that bi-gram and tri-gram of N-gram features obtained the best classification performance gaining higher classification performance on both negative and positive classes in all measured areas.

In addition, adding ruled-based on the bag of words increases accuracy performance in all measured areas. Bag of words model does not capture the effect of negation and the contextual polarity efficiency, hence, incorporating context of adjectives, adverbs, nouns, and verbs together with the effect of negation to the features of the bag-of-words model proved to be highly efficient with increased classification accuracy by an average of 8 percent. Bi-gram and tri-gram models have been shown to be extremely effective, since it captures more contextual meaning by considering the neighboring words of a given phrase, resulting to a higher classification accuracy.

The impact of adding ruled based can indeed be of great help to automatically determine the sentiments of blog sentences. The application of SVM learning machine and bi-gram feature on the proposed model were resulted in higher classification performance and far above the baseline classification performance of the previous researches.

REFERENCES

Blog Comments Sentiment Analysis For Estimating Filipino Isp Customer Satisfaction


