

Estimating Filipino ISPs Customer Satisfaction Using Sentiment Analysis

Frederick F. Patacsil^{1,*}, Alvin R. Malicdem², Proceso L. Fernandez³

¹Information Communications Technology Department, Pangasinan State University, Philippines

²College of Information Technology, Don Mariano Marcos Memorial State University, Philippines

³Ateneo de Manila University, Loyola Heights, Philippines

Copyright © 2015 Horizon Research Publishing All rights reserved.

Abstract Sentiment Analysis (SA) combines Natural Language Processing (NLP) techniques and text analytics to extract useful information from textual data. This study uses SA to estimate the Filipino internet customers' satisfaction related to the quality of the service provided by the Internet Service Providers (ISPs). Data were collected from Blog comments shared with online social media. Automatic word seed selection was applied using the word pair set {"Good" and "Slow"} as initial seed for the word dictionary. The Naïve Bayes method was used as a classifying tool to identify the dominant words used to express customers' sentiments and to determine the sentiment polarity of their opinions. The proposed automatic classifier successfully identifies positive and negative polarity of the blog sentences with a 91.50% accuracy in the training set. However, the results of the actual evaluation of the manually labelled test set show a drop in accuracy rate of 60.27%. Some of the reasons for this drop in accuracy are investigated in this paper.

Keywords Internet Service Providers, Sentiment Analysis, Blogs, Translator Machine, Naïve Bayes

1. Introduction

Sentiment Analysis (SA) refers to 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 (Roebuck, 2012)[6]. In this study, we use SA to the data gathered from the comments section of selected blogs. Blogs are one of the platforms that express personal opinions about a specific topic, and the comments section such as 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 the Philippines, there are three major ISP, namely Smart Communication, Globe Telecom and the Sun Cellular.

As of 2012, Smart has 1.73 million total subscribers followed by Globe with 1.7 million subscribers and Sun with 650,000 total subscribers (Miniwatts Marketing Group, 2014) [4].

With the continuous growth of internet access, online user reviews are increasingly becoming the de-facto standard for measuring the quality of products and services. Many Filipino internet customers have expressed their sentiments about the quality of services and the speed of the internet access provided by an ISP through social networks such as Facebook, Twitter and blog sites. The Internet customer comments on online review is an influential factor that may affect decision of other customers about the product and the services of a certain entity. Therefore, these sentiments are very important source of information that should be taken into account by ISPs in improving their services and development of their products. However, the sheer volume of online comments makes it difficult for any human to process and extract all meaningful information. As a result, there has been a trend toward systems that can automatically summarize opinions from a set of reviews and display them in an easy to process manner. 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. Furthermore, SA may involve building a system or automated method of collecting and examining opinions about the product made through blog posts, comments, reviews or tweets (Chesley, Srihari, Vincent and Xu, 2005)[1].

In this study, we proposed a method for the automatic estimation of customer sentiment polarity using a simple Naïve Bayes classifier and word seeds.

2. Related Literature

There have been several previous efforts research on sentiment analysis utilizing unsupervised automatic classification using word seeds to classify the polarity of a document which can be an entire article, a paragraph or even a single sentence. Prior researches who conducted sentiment classification used an unsupervised approach and seed of

words as stated in the foregoing discussions.

Zagibalov and Carroll (2008)[8] describe and evaluate a method of automatic seed word selection for unsupervised sentiment classification of product reviews in Chinese. Their aim was to investigate means to improve the classifier by automatically finding a better seed word. They based the selection of their initial seed based on the following observation (1) The initial seed should always be more often used without negation in positive texts, while in negative texts it is more often used with negation and (2) the seed occurs more often in positive texts than negative, and more frequently without negation than with it. They decided "Good" as their initial word seed. The results obtained are close to those of supervised classifiers and sometimes better, up to an F1 score of 92%.

Turney (2002)[7] presented in his paper a simple unsupervised learning algorithm that utilized two arbitrary seed words ("Poor" and "Excellent") to calculate the semantic orientation of phrases. The algorithm has three steps: (1) extract phrases containing adjectives or adverbs, (2) estimate the semantic orientation of each phrase, and (3) classify the review based on the average semantic orientation of the phrases. The core of the algorithm is the second step, which uses Pointwise Mutual Information and Information Retrieval (PMI-IR) to calculate the semantic orientation. The sentiment of a document is calculated as the average semantic orientation of all such phrases. This approach was able to achieve 66% accuracy for the movie review domain at the document level. He found movie reviews to be the most difficult because of this argument "the whole review is not necessarily the sum of the parts (Turney, 2002)[7].

Another study that made use of unsupervised system and word of seed was conducted by Rothfels and Tibshirani(2010). They examined an unsupervised system of iteratively extracting positive and negative sentiment items which can be used to classify documents. They adopted the idea of semantic orientation to choose the initial set of seeds and they hand picked two sets of reference seeds, one positive and one negative. They used arbitrarily positive words such as "Good", "Excellent", "Amazing", "Incredible", and "Great" and for our negative set, "Bad", "Poor", and "Terrible". The results of their study achieved an accuracy of 65.5%, a modest but significantly better than a near-baseline accuracy of 50.3% with the original approach.

Zhang, Xia, Meng and Yu (2009)[9], pursued the analysis of product reviews using a bootstrapping method to find the product features and opinion words in iterative steps. Furthermore, a method was presented to get the initial seeds of product features and opinion words automatically.

The task of selecting seed words includes the following steps:

- (i) In the product reviews, choose a small set of features and opinions as "seed words"
- (ii) Count co-occurrence of candidates and seed words in the product reviews
- (iii) Use a figure of merit based upon these counts to select new seed words
- (iv) Return to step (ii) and iterate n times

The experimental results of that study were encouraging, which indicate that the proposed method and techniques were effective in performing this task of feature level opinion mining.

Our work presents a new approach based on the use of English translator tools to translate online comments in Filipino into English form, automatic word seed selection using a pair of set words and a simple unsupervised automated classification technique that determines the sentiment polarity of the sentences

3. General Methodology

3.1. Proposed Automatic Classifier Algorithm

3.1.1. Information Extraction

In this study we first searched for blog articles from Google that discuss and compare the services of the three main ISP. Blog articles that contain many comments from their customers were highly considered. The following are examples of blog articles that were included in the data: "Comparing Globe Tattoo, SmartBro and Sun Broadband Wireless » tonyocruz.com", "Globe Tattoo vs. Smart Bro vs. Sun Broadband Wireless Which is the best Reader Comments TechPinas Philippines' Technology News, Tips and Reviews Blog" and "The Best ISP (Internet Service Provider) in the Philippines Jehzlau Concepts".

Blog comments from the selected blog sites that feature the services of the major internet providers were extracted using a customized PHP web scraping application which retrieves the customer comments and other important information, and then stores these in a database automatically. The unnecessary parts such as the title, articles, and other contents were not included in the mining process.

3.1.2. Machine 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. Thus, after cleansing, a machine translation using Google Translate was employed. Specifically, a modified application in PHP was used to automatically convert Filipino sentences into their 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.

3.1.3. Building the training and test datasets

We utilized two datasets for our proposed study. The actual data for this model contains of 14000 sentences derived from 5280 blog comments. These data were used to identify the sentiments of the customers regarding their satisfaction about services provided by Globe, Smart, and Sun. To obtain the training data set, we randomly selected 10000 sentences from the original dataset using the following MYSQL command.

```
"SELECT * FROM tablename ORDER BY RAND()"
```

```

for($x = 0; $x < count($comment); $x++){
    $seeddummy = "";
    echo $comment[$x];
    $sdummy = explode(" ", $comment[$x]);
    $ssdummy = array();
    for($x1 = 0; $x1 < count($sdummy); $x1++){
        $ssdummy[$x1] = Stemmer::Stem(strtolower($sdummy[$x1]));
    }
    foreach($seeds as $key => $seed) {
        if(in_array(trim($seed), $ssdummy, true)){
            $seeddummy = $seeddummy . " " . $seed;
            $count = $count + $seedid[$key];} // count sentence polarity
    }
}
    
```

Figure 1. Sample PHP Code that identify sentiment polarity of the sentences

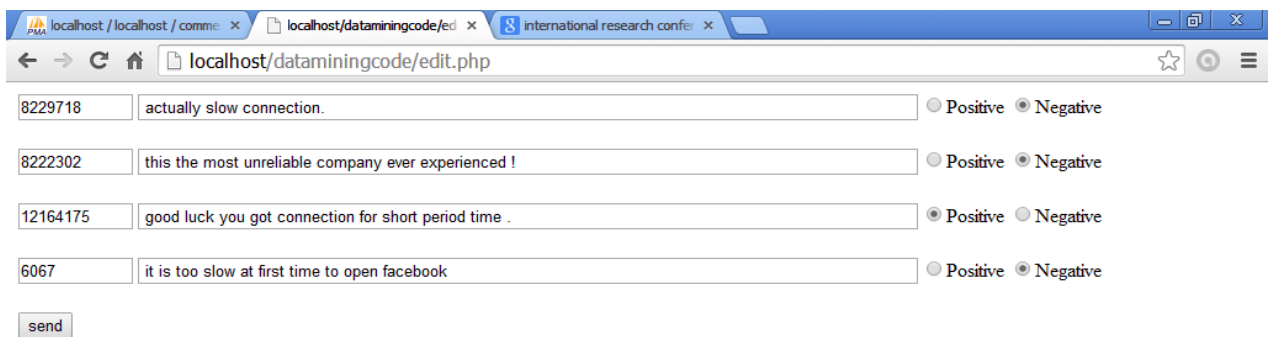


Figure 2. Sample screenshot of the application used for manually labeling thepolarity of the sentences.

3.1.3.1. Sentiment Polarity Identification of the training datasets

The identification of sentiment polarity was done using a modified PHP application specifically for sentiment analysis. The application counted the number of positive and negative words, then computed the total score. If the total score (positive score-negative score) was greater than 0, it would be considered as a positive sentiment and if it was smaller than 0, it would be considered it as a negative sentiment. Sample PHP code is shown in Fig. 1.

3.1.3.2. Training Dataset

The 10000 sentences were further processed to remove stop words (Common words that have a little value in the process of identifying sentiment analysis e.g. “a”, “the”, etc.) and stemming (words that carry similar meanings, but in different grammatical forms such as “loved”, “Loves” and “Loving” was combine into one word “love”) was also applied in the sentence preprocessing. In this way, the sentences can show a better representation (with stronger correlations) of these terms, and even the size of the dataset can be reduced for achieving faster processing time. This preprocessed sentences were fed into the proposed sentiment classifier for polarity identification and labeling. To have a balance training dataset, we further randomly selected 1000 positive and 1000 negative sentences for the final training dataset.

3.1.3.3 Test Dataset

The remaining sentences from the actual dataset was labeled manually by three sentiment polarity raters. The raters used a customized PHP application to label the sentiment polarity of the sentence. A screenshot of the application is shown in Fig. 2.

An interrater reliability analysis using the Fleiss' - Kappa statistic was performed to determine the consistency among the raters in the identification of sentence polarity. Table 1 shows the result of interrater reliability among the three sentence polarity raters.

Table 1. Fleiss' Kappa agreement results of three sentiment polarity rater

Expected Agreement	Observed Agreement	Fleiss' Kappa
0.683	0.965	0.888

Legend: Landis and Koch 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 interrater reliability for the sentence polarity raters was found to be K=0.88.

According to the definition of the Fleiss' Kappa statistic, the accuracy of the interrater reliability is considered to be "almost perfect agreement" (Landis, J.R., and Koch, 1977)[3].

After the Fleiss' Kappa interrater reliability was determined, the three raters agreed on the sentiment polarity of other sentences that were labeled differently to come up with a common polarity for each sentence of the testing data.

Furthermore, only sentences with positive and negative polarity were included as part of the test dataset.

3.1.4. Determination of Word Seeds

Before carrying out with the classification of the training data set, there was a need to create a word dictionary to be used as the base for sentiment analysis. Several research works utilized automatic seed selection using a set of words as initial seed for the dictionary. The pioneer work was initiated by Turney's (2002)[7] which classifies a document using two human-selected seed words (the word "Poor" as negative and "Excellent" as positive). Zagibalov and

Carroll(2008)[8] also utilized this approach and they described it as 'almost-unsupervised' system that starts with only a single, human-selected seed ("Good"). They also claimed that the 'almost-unsupervised' system produces a better result.

This research also utilized automatic identification of seeds of the word for the dictionary based on the data set. The adjective that had the most number of occurrences as highlighted in Fig. 3 were used as the initial seeds.

The initial seed set consisted of the keyword {"Good" and "Slow"}. An application program that automatically searches for and retrieves the synonyms and antonyms of the initial seeds from an online thesaurus dictionary using <http://thesaurus.altavista.org/thesaurus/v1> as part of the word seeds of this research. After the first process, the application retrieves 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.

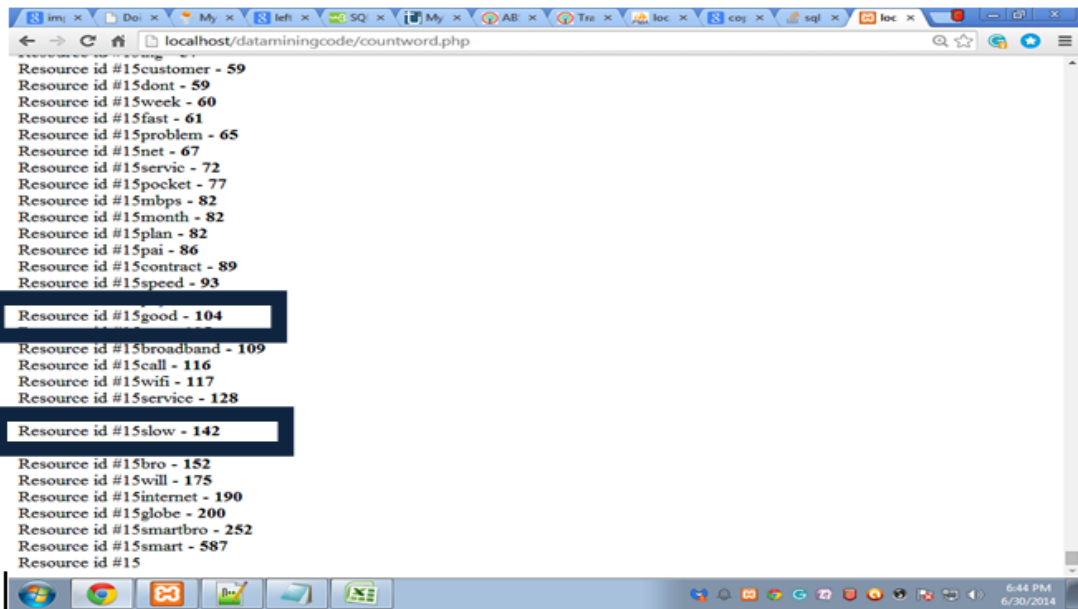
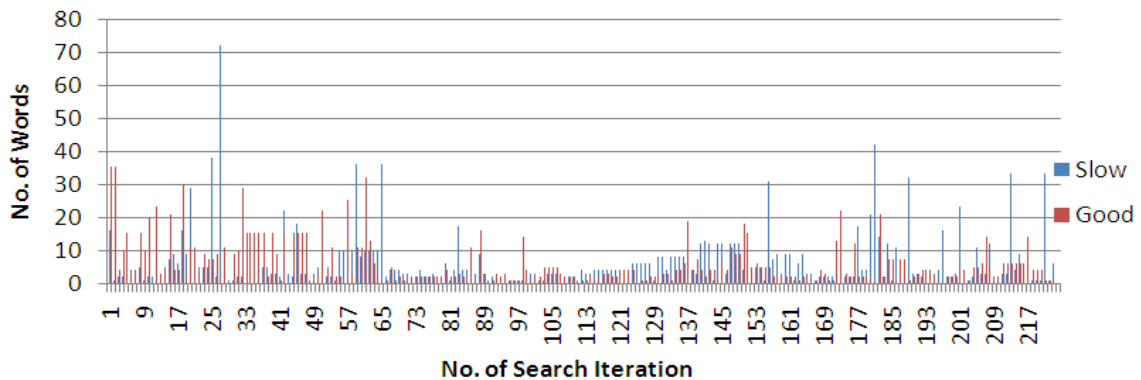


Figure 3. Word count based on the training data set



*** No. of words searched (repeated word was automatically disregarded by the application system)
 Positive - 930
 Negative - 1351

Figure 4. Result of word seeds searching using an online thesaurus dictionary

4. Experimental and Result

4.1. The Training of the Proposed Automatic Sentiment Classifier

In this research, we proposed an automated Naïve Bayes sentiment classifier to classify sentences polarity. The proposed sentiment classifier was trained to identify the polarity of unlabeled sentences. Furthermore, it uses its own predictions to teach itself to classify unlabeled sentences using a positive and negative bank word seeds.

The training of the proposed automated sentiment algorithm was carried out using Rapid Miner 5.3. The presented training results were applied using 10 fold cross validation. In addition, results were presented using confusion matrix that contains information about the actual and predicted classification done by a Naïve Bayes classifier (Hamilton,2009)[2].

Table 2. Confusion matrix for two-class classifier

Predicted Class	True Negative	True Positive	Class Precision	F-Measure
Negative	tn	fp	Negative	F-Measure Negative
Positive	fn	tp	Positive	F-Measure Positive
Class Recall	Negative	Positive		

Accuracy:

The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the (1) :

$$AC = \frac{tn+tp}{tn+fp+fn+tp} \quad (1)$$

The accuracy (AC) is the correct classifications divided by all classifications and considered as simplest and most intuitive assessment.

The recall (in the case of positive cases) is the proportion of positive cases that were correctly identified, as calculated using the (2):

$$Recall = \frac{tp}{tp+fp} \quad (2)$$

The Recall shows the fraction of relevant positive instances that are retrieved during the classification process.

The precision (in the case of positive cases) is defined as the proportion of positive cases that were classified correctly, as calculated using the (3):

$$Precision = \frac{tp}{tp+fn} \quad (3)$$

The Precision shows the fraction of retrieved positive instances that were relevant during the classification process.

The F-measure can be interpreted as a weighted average of the precision and recall, where F-measure values can reach its best value at 1 and worst score at 0 and as calculated using (4)

$$F - Measure = 2 \frac{(\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (4)$$

The F-measure is used in the field of information retrieval for measuring search, document classification, and query classification performance (Beitzel., Steven M., 2006)[10].

We experimented Naïve Bayes classifier for our automated sentiment polarity classifier. For a preliminary experiment of the proposed classifier, a 10-fold cross validation was performed to estimate the generalizability. Table 3 reports the average precision, recall, F-Measure and accuracy for all measures.

Table 3. Evaluation results of the proposed sentiment classifier using training dataset

Predicted Class	True Negative	True Positive	Class Precision	F-Measure
Negative	940	110	89.52%	92%
Positive	60	890	93.68%	91%
Class Recall	94%	89%		

Accuracy : 91.50 %

The table shows that the proposed sentiment polarity classifier correctly classified 940 sentences out of 1000 negative instances and 890 positive cases out of 1000 positive sentences. The results revealed that negative sentences are classified with better recall and precision than positive sentences. Considering F-Measure, negative sentences had the weighted average value of 92% and positive as 91%, which shows that proposed classifier performed slightly better in the classifying negative sentences than those for positive sentences. A total of 1830 cases out of 2000 test data instances were correctly classified with an accuracy of 91.50%. Thus, the proposed system efficiently performed sentiment polarity classification on the sentences using word seeds and an automated sentiment classifier.

4.2. Testing the Proposed Automatic Sentiment Classifier

We utilized the Naïve Bayes model generated during the training to test the performance of the proposed automated sentiment polarity classifier. Table 4 reports the performance of the proposed classifier in terms of average precision, recall, F-Measure and accuracy for all measures.

Table 4. Classifying results of the proposed sentiment classifier using test dataset

Predicted Class	True Negative	True Positive	Class Precision	F-Measure
Negative	683	269	72%	72%
Positive	259	118	31.30%	30.89%
Class Recall	73%	30.49%		

Accuracy : 60.27 %

The above table shows the classifying results of the proposed automated sentiment polarity classifier. In view of

precision, recall and F-Measure, the sentiment classifier performed slightly better in classifying negative sentences than those for positive ones. Moreover, the results showed that the accuracy is 60.27%, which is much lower as compared to the results during the training of the data set of the automated sentence polarity classifier.

The result of the accuracy test suggests that the proposed automatic sentiment polarity classifier does not perform well. The following are contributors in the lower performance:

1. There are errors in the translation of sentences or some Filipino words have no equal terms in English that may cause to change the polarity of the sentence.
e.g. (Filipino) “*Globe Tattoo has BAD service internetmga 2 hrslangang nagagamit mo.*”
(English Translation) – “*Globe Tattoo internet service has BAD 2 hrs only usable to.*”
2. Users employed special characters to write their comments and some sentences are erroneously written.
e.g. (*mbps download speed bul @ # \$ % hit !*).
3. Ambiguity in synonyms and antonyms of a word in dictionary. There are some word seeds that carry polarity whose meaning cannot be definitively resolved.
e.g. “*there sun cellular cell sites near our house will improve connection with that being said thanks !*” - will positive polarity.
4. Finally, the use of urban words (e.g. shit !, Bullshit, yucks, badtrip), colloquial language, and Short Message (SMS) like sentences in which user express their sentiments but, the proposed automatic polarity classifier cannot establish the sentiment of sentences because these words are not part of the word seed.

5. Conclusions

This research paper presents a method of evaluating Filipino internet customers’ sentiments using an automatic approach utilizing the seed word, language translation machine and an application that utilized Naïve Bayes method to classify the sentiment polarity of the sentences.

The proposed automatic sentiment polarity classifier performs well in classifying negative sentences both in training and testing evaluation periods. Furthermore, the proposed sentiment classifier produces a result of 91.50% accuracy in the classification of sentence polarity during training period. Although during the actual testing period of the test data set, the proposed system was able to correctly classify 60.27% sentiment of the sentences. There are

several factors that are contributory to low performance, such as language translation tools, uncertainty of synonyms and antonyms of words in the online dictionary, and use of urban words, colloquial language, Short Message (SMS), and long words.

For future work, we will employ more advanced tools and techniques like other online dictionary (Bhing and Mosses), better language translation machine and Part-of-Speech (POS) tagging to improve the performance of our proposed automated sentiment classifier.

REFERENCES

- [1] Chesley, P. Vincent, B., Xu, L., and Srihari, R. (2006). "Using verbs and adjectives to automatically classify blog sentiment", *Proc. AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, Stanford University, CA, March 2006, AAAI Press, TR SS-06-03, pp.27-29.
- [2] Hamilton, H. (2009). Computer Science 831: Knowledge discovery in databases, Retrieved from http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html
- [3] Landis, J.R., and Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*.
- [4] Miniwatts Marketing Group. (2014). Philippines Internet Usage Stats and Marketing Report. Retrieved from <http://www.internetworldstats.com/asia/ph.htm>.
- [5] Rothfels, J. and Tibshirani, J. (2010). Unsupervised sentiment classification of English movie reviews using automatic selection of positive and negative sentiment items, <http://nlp.stanford.edu/courses/cs224n>.
- [6] Roebuck, Kevin. (2012). *Sentiment Analysis: high-impact strategies - what you need to now: definitions, adoptions, impact, benefits, maturity, vendors*, Emereo Publishing.
- [7] Turney, D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, Philadelphia, pp. 417-424.
- [8] Zagibalov, T. and Carroll, J. (2008). Automatic seed word selection for unsupervised sentiment classification of Chinese text, *Proceedings of the 22nd International Conference on Computational Linguistics*, pages 1073–1080.
- [9] Zhang, S., Xia, Y., Meng, Y., and Yu, H. (2009). A Bootstrapping method for finer-grained opinion mining using Graph model, *23rd Pacific Asia Conference on Language, Information and Computation*, pages 589–595.
- [10] Beitzel, Steven M. (2006). *On Understanding and Classifying Web Queries* (Ph.D. thesis). IIT. CiteSeerX: 10.1.1.127.634