World Applied Sciences Journal 12 (Special Issue on Computer Applications & Knowledge Management): 32-38, 2011
ISSN 1818-4952

User-Centered Sentiment Analysis on Customer Product Review

Yoke Yie Chen and Ken Vinn Lee

Department of Computer and Information Sciences Universiti Teknologi PETRONAS,
Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia

Abstract: E-commerce has been growing rapidly and is still expanding worldwide. More and more people are purchasing their products online and giving reviews and comments through the Internet. These reviews are valuable information if extracted and summarized properly. Using this information, users will be able to buy a product that suits their needs; suppliers will know what the customers' likes and dislikes and will be able to order their supplies accordingly. With the advancement in technology, it is possible to extract the information from user reviews which are natural human language. Sentimental analysis or opinion mining is used to mine the information and summarize into useful graphs and charts. The objective of this paper is to develop a prototype that extract and analyze the sentiments of customer's product reviews. After the prototype is developed, the steps completed in order to program a fully functional system are discussed in the system documentation section. With a thorough testing being carried out on the system, the system manages to obtain a precision percentage of 87% and recall of 93%. With such percentage, it is concluded that the steps involved have a high accuracy rate. However, it could be better improved by taking into consideration more grammar orientation of a sentence and the output of the system can be made into more interactive to users.

Key words: Sentiment Analysis · Opinion Mining · POS Tagging · Customer Review

INTRODUCTION

Social media on the Internet has changed the way people express their feelings and opinions. Internet users can post product reviews on blogs and forums and express their opinion towards the product. This information is valuable to both customers and manufacturer. For consumer, one could search on these social media to find the opinion of existing users. They do not rely solely on opinions from friends and family when the search for a product review is simple and convenient. Consumers are able to gather information anytime, anywhere. For manufacturer, apart from using traditional way such as customer surveys, they can gather consumers' reviews from these social media to make product improvements by understanding the likes and dislikes of consumers.

Sentiment analysis, or opinion mining, is a discipline at the crossroads of Information Retrieval and of Computational Linguistics. Mining opinion as the name stated, is not concerned with the topic of a document, but with the opinion it expresses [1]. Sentiment analysis has been used for many purposes; including predicting the outcome of an election [2], provide a summary of search result from a search engine [3], recognizing spam emails [4] and many more. The a pproach for each study is different; each approach is to cater for its own purpose. These approaches can be categorized as statistical or linguistic. Statistical approach is more quantitative whereby a decision is made by comparing the occurrence of a word associated with another. There is granularity, meaning that every word is somehow related to one another; and co-occurrence, meaning that different words with same meaning are missed out on this approach. On the other hand, a linguistic approach will weigh its decision more on the reviewer's choice of words. Besides that, words are analyzed with large corpus of words to extract their lexical meaning. Apart from that, there are approaches that are based on the grammar from a sentence, like the usage of the word "but" will mean a contradicting polarity between two parts of a sentence [5].

This research aims to use both statistical and linguistic approach to categorize sentiments as positive and negative. The resulted prototype is a general sentiment analyzer that allows users to intervene with the system. Hence, the system does not only analyze opinions at a specific type of product.

Corresponding Author: Chen Yoke Yie, Department of Computer and Information Sciences Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia, E-mail: chenyokeyie@petronas.com.my.

Related Works: To gather information on 'what other people think' on the internet is a simple but challenging tasks. Many product review sites are easily accessible on the Internet and it has become a main tool that assists users in decision making process. However, when users are overloaded with information, they can have issue in difficulty to make decision. To overcome this, we can provide a summary of all the opinions regarding a specific item or product. In general, we can divide an opinionated sentence into 3 main parts: object, feature and opinion. To understand what others think about the features of an object, the review document will need to be separated into the 3 main components to perform further analysis. There are a total of 8 problems that need to resolve as mentioned in [6, 7]:

- Identify the object of the review.
- More than one feature is mentioned in a sentence.
- Identify the explicit and implicit feature.
- Identify the opinion holder.
- Identify the opinion.
- Analyze the polarity of the opinion.
- Deal with emotions.
- Explicit and implicit opinion.

Previous research has focus on solving some of the abovementioned problems. The most commonly used techniques for sentiment classification is machine learning algorithms such as Support Vector Machine (SVM) and naïve Bayesian. Research done by Boiy and Moens [8] focused on the feelings that people express in English, Dutch and French with regard to certain consumption products. This work was done by combining three different classifiers, SVM, Multinominal Naïve Bayes and Maximum Entropy, in a pipelined cascaded way. It was found that by using unigrams to determine the features yield high accuracy for the three different languages. Their work has also promoted the use of active learning when labeling training examples provides noticeable improvements in the overall results. Zahran and Kanaan [9] have used Particle Swarm Optimization (PSO) for feature selection to improve the performance of Arabic text categorization.

Work by Tan and Zhang [10] has compared different feature selection (MI, IG, CHI and DF) and learning methods (centroid classifier, K-nearest neighbor, window classifier, Naïve Bayes and SVM) in mining opinion for Chinese documents. Their work has showed that IG performs the best for sentiment terms selection and SVM exhibits the best performance for sentiment classification. One of the important findings from their work is that sentiment classifiers are heavily dependent on domains and topics.

Tian *et al*. [11] mining opinions of Chinese review sentences to obtain comprehensive evaluation of product and ranking product in all features by using lexicon and ontology. Product and features that users show interest in will be extracted by searching in ontology. Polarity strength for the remaining words of the sentence will be determined using HowNet "Sentiment Mining Lexicon". The polarity of each word is then mapped to the product and features by using syntactic parser. It was found that, the precision of product ranking can be improved by introducing syntactic parsing into product ranking.

Another approach that has been proposed by Won *et al*. [12] is the use of association rules in mining product reviews. Features and opinions of a review sentence were extracted in the form of transaction data. The association rules is then discovered from the transaction data and then summarized the polarity of the opinion using PMI-IR algorithm.

MATERIAL AND METHODS

There are four main parts involved in the Sentiment Analysis Platform (SAP) which consists of text preprocessing, feature classification, polarity classification and, summarizing and decision making. The system architecture is illustrated in Fig. 1.

Text Preprocessing: Text preprocessing is the process of trimming out unwanted words, correcting misspelled words and finding out important words in a sentence. In our proposed test preprocessor, there will be three steps involved which are, misspellings correction, Part-of-Speech (POS) tagging and word stemming. These steps have to be done in order to avoid any error. If POS tagging is done after correcting misspelling word, the tagging process will not be accurate, on the other hand, if stemming is done before POS tagging, words like "playing" might be tagged as a noun after it was stemmed into the word "play".

Fig. 2 shows the overall process flow of text preprocessing with a sample customer review. At the end of the process, the sentence has been check for spelling errors, tagged with their respective part of speech and stemmed to its original form. This final form of sentence will be passed to the next process, feature classification.

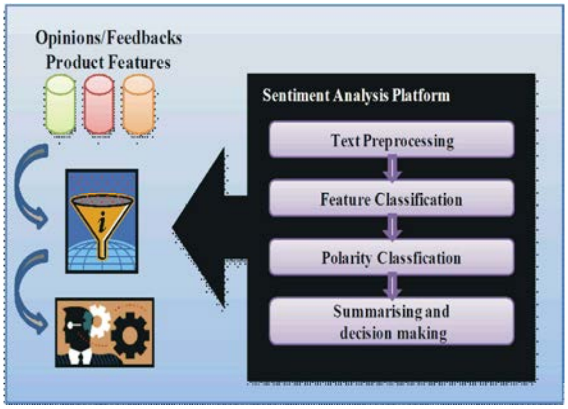

Fig. 1: System Architecture

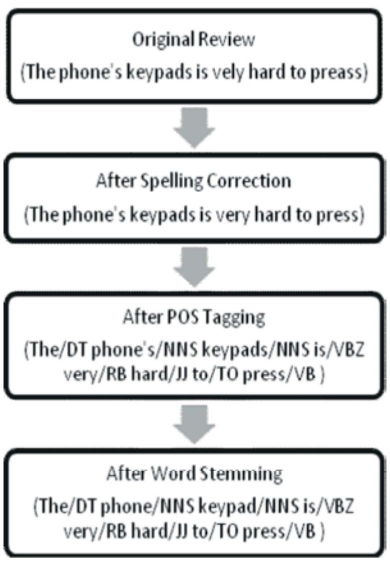

Fig. 2: Process Flow of Text Preprocessing

Feature Classification: Feature classifying is the process of determining which feature a sentence is referring to. To demonstrate the basic idea of feature classifying process in our proposed work, the product features of a mobile phone are given here as an example. In this example, we have to decide if a sentence is referring to the product features of a mobile phone. When a sentence explicitly mentions the feature like, "the camera's picture is very clear", we can extract the feature easily after preprocessing has been done, which will leave us with only two nouns: "camera" and "picture". In order to identify the main feature of this sentence, we could compare it with existing reviews to find out which word is more frequently used or compare it with a set of training data that is already preprogrammed in the system. Apart from that, if there is no existing data, we can use the grammatical orientation of the sentence to determine which feature is the main feature. For example, in this scenario, "camera's picture" has obviously state that picture is a part of the camera function, hence using the "'s" we can determine that camera is the feature expressed in this sentence.

However, with a sentence like "this phone is too heavy". This process will have to figure out that this short sentence is referring to the phone's weight, even when it is not specifically mentioned in the sentence. There are a few methods to identify the feature from such sentences. Firstly, we can use the adjective referring to the feature, which is "heavy"; "heavy" can be related to something's weight and as a result we can conclude the feature is weight. Another method is similar to the method mentioned before, which is to compare with existing reviews. By looking at each features' descriptions, we can find one feature that most frequently use the word "heavy", or even compare it with the word "phone". Finally, when all methods failed to yield any results, we allow the user to classify this sentence for the system. Once the user had chosen a feature for such sentence, future sentences with such ambiguity can be classified using this information as existing reviews.

There is another challenge to overcome when it comes to feature classifying. There may be some features of a phone that can have different names but are referring to the same feature. For example, volume and loudness would refer to the same feature, but we should not categorize it into two separate features. In order to group similar features, WordNet can be used to find out the synonymous words of a feature. WordNet® is chosen database to be referred as it has a large lexical database of English and its use has been widely accepted and cited in many research works. To create a promising system, all methods above are to be combined and the best methods are to be used first followed by other methods if the first method fails. Fig. 3 shows the overall process flow of classifying a noun into a feature that belongs to a mobile phone.

In general, there are two possible patterns for each chunk of words. One possible pattern is that it does not contain a noun; these chunks can be ignored for any further processing. A chunk of words without any nouns cannot possibly contain any positive or negative review about the phone. The other possibility is that there is one noun in the chunks of words. It should be known that the system splits the sentence so that a noun will be the last word inside the chunks of words. Next, the system has to identify which adjective is linked with which noun.

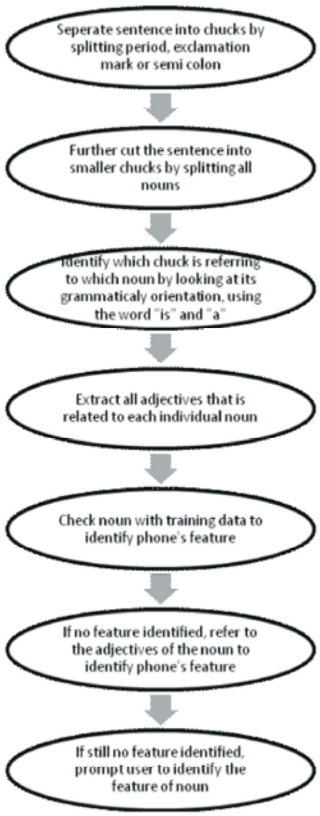

Fig. 3: Process Flow of Feature Classifying

Again, there are two possibilities. Adjectives of a noun can be located behind the noun; e.g. "beautiful design" or behind the noun; e.g. "design that is beautiful". The keyword to distinguish between these two phrases is the word "is". Therefore, the system detects if the chunks of words contains the word "is" or the word "are" and attached the adjectives accordingly. For example if a chunk of words does not contain the word "is" or "are", all adjectives from that chunk are attached with the noun behind. If it contains such words, all adjectives are attached with the noun from previous chunk. However there is an exception to this rule, which occurs in sentences with the word "a". For example, "this is a nice phone". Although there is the word "is", the adjective nice is still describing phone as there is a word "a" there to refer the adjective to the noun. Therefore, the system also has to figure out if such exceptions are in the user reviews.

The next step is to check the list of features that has been preset in the database as the features that are wanted for the system to identify. Each noun is checked with this list of features and if they are matched; the feature of the noun will be identified. However, if it is not identified; the noun is matched with another list that contains the preset features and words that are related to them. For instance, the word "sound" is related to the feature of audio and this has been stored in the database. If the system detects the noun "sound", the system will be able identified the feature the word "sound" belongs to. This secondary list of words will keep increasing as more and more reviews are saved and its feature identified. If the feature is still unidentified in this process, the system will use the adjectives related to the noun to identify the feature.

Similarly, there is a list of words that relate the adjectives with a feature in the database. When the amount of a feature is linked with an adjective for more than two times, the system will automatically assign the feature to the noun based on its adjectives. For example, a comment: "this phone's price is cheap" is saved into the system for three times. The adjectives cheap and the feature price are saved inside this list. If the comment "this phone is cheap" is input into the system; although there are no feature mentioned and the only noun available is the word "phone" that can refer to anything the system will be able to identify the feature "price" using its adjective. Finally, if it still fails to yield any results, the system will prompt the user to select the feature themselves. After the user selected a feature for a noun, it will be saved into the system and therefore if the noun is to appear another time, the system will be able to identify a feature for that noun. By doing so, the system will continuously learn from each usage. This system will be able to grow and become more reliable as it receives more inputs. Apart from storing which noun is related to which features, it also stores the adjectives to identify the feature.

Polarity Classification: Polarity Classifying is the process of determining whether the expressed opinion in a sentence is positive or negative (neutral). Firstly, lists of globally accepted positive words like "great, good, amazing, exciting, wonderful" will be stored as a training data set. Likewise a set of negative words will also be stored. An adjective of a sentence will be considered as positive or negative if it exists in the synonymous set of words from the respective polarity from WordNet®. After that, this word will be added into the training data set. As the system is being utilized, the training data set will grow, thus increases the capability of the system to distinguish between a good comment and a bad comment.

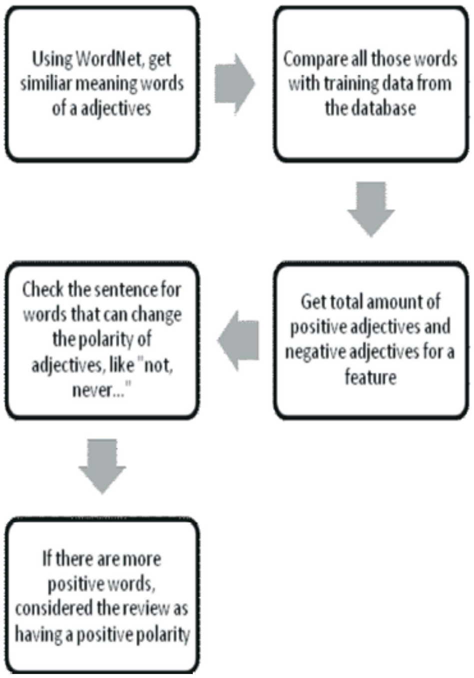

Fig. 4: Process Flow of Polarity Classifying

Similar to feature classification, polarity classifying has its own sets of challenges. The first is when words like “not”, or “never” is used. Such words will alter the polarity of the sentence completely and therefore we must ensure that such words are not overlooked. Apart from that, there is the problem of ambiguous adjective like “long battery life”. The word “long” could be positive or negative expression to a non-trained system, although it is obviously a positive statement to human beings. There are a number of ways to deal with this problem. The first is to make use of the grammatical words surrounding this adjective. For example the sentence “this phone is ugly, but it is quite light”, “light” used here is ambiguous, but by comparing it with the first comment which can be decided as a negative statement using the word “ugly” and using the word “but”, we can conclude that the statement after the word “but” is the opposite of the polarity of the first statement which is negative. After classifying “light” as a positive adjective, we can classify future reviews containing the word “light” as positive if it falls under the feature of weight. Other words that can determine the polarity of a review are; “too, yet and, although”, etc. If the proposed method still fails to classify the polarity, the system will prompt the user to classify the statement so that the system is able to distinguish itself whether the adjective is positive or negative.

Fig. 4 shows how the system is employed to identify if a feature is being reviewed positively or otherwise. First of all, the synonym set of words for all adjectives is obtained from WordNet®. This process enlarges the possibility that the word is identified in the training data set. Secondly, these synonyms along with the original adjectives are compared with the set of training data that is preset in the database. This training data is categorized into positive adjectives and negative adjectives. If there is a match, the system will identify whether the adjective is positive or negative. Next, the amount of positive adjectives and negative adjectives are counted to decide whether the comment is actually giving a positive review or not. If there are more positive adjectives compared to negatives ones, the review is considered to be positive. Words like “not”, “never”, that can change the polarity of an adjectives is detected so that the system will correctly identify the polarity.

Summarizing and Decision Making: Finally, with all the processes completed the results of a comment’s polarity and what the comment is referring to will be obtained. Storing this information, the system will be able to generate an output of all collected comments in graphical formats. From these graphs, user can make decisions on whether a mobile phones is positively accepted or not and which features are lacking in performance.

RESULTS AND DISCUSSION

The quality of the sentiment analysis of customers’ reviews with regards to certain product needs to be assessed in order to ensure the proposed approach is effective in determining the comments are either positive or negative. Hence, an intrinsic method has been used to evaluate the system generated analysis by comparing it with human generated surveys. The human generated surveys were obtained from a group of university undergraduate and postgraduate students with age range of 18-25. The surveys were done by randomly picked the students in three universities who were using a certain mobile phone model regardless what mobile service operators that they were subscribed to fill up a questionnaire. The questionnaire was divided into two parts. In the first part, the students were asked to write the product reviews in words. In the second part, the students were asked to circle a round number between 1 to 10 scale (rated as the poorest to the best) to reflect the overall opinion about the mobile phone model.

In our study, those rated 1 to 5 are considered as negative reviews while those rated 6 to 10 will be categorized as negative reviews. The number of positive reviews and negative reviews will be counted and this formed the human generated surveys.

On the other hand, in order to make use our proposed system, the Part 1 of each of the questionnaire was fed into the system for further processing. The system would then generate an output by assigning the questionnaire (product review) as a positive or negative review. Each of the output was then compared with the result from human generated survey to check if the analysis was done correctly (correctly assigned) or incorrectly (incorrectly assigned).

The performance measures used to evaluate the quality of the analysis are precision and recall ratio which are shown accordingly below. Precision is defined as the probability that the review is relevant given that it is returned by the system whereby recall is the probability the relevant object is returned [13]. For precision, the higher the values, the better the system is in analyzing the opinions given by the users. On the other hand, the higher the recall values the more effective the system would be in retrieving the relevancy of the product reviews. The value of recall ratio is 1.0 when all relevant comments are retrieved by the system and analyzed correctly.

$$Precision(\rho) = \frac{\#of\ Correctly\ Assigned}{\#\ of\ Correctly\ Assigned + \#\ of\ Incorrectly\ Assigned}$$

$$Recall(r) = \frac{\#of\ Correctly\ Assigned}{\#\ of\ Correctly\ Assigned + \#\ of\ Incorrectly\ placed}$$

The evaluation results of the proposed system are summarized in Table 1. A total of 50 students have given response to our surveys. Out of these 50 students, 29 of them gave positive reviews (rate 6-10) while the remaining gave a rate of 1-5. As shown in the table, the precision of the proposed system nearly 90% accuracy in both positive and negative reviews. This means that given a pool of product reviews, the proposed system can assign each of the review whether it is positive or negative opinion with a probability of about 0.9. However, when it comes to recall ratio, the ratio is about 10% higher for positive reviews as compared to negative reviews. It shows that the proposed system can interpret the positive words better than the negative words. These could be due to the definition in the human generated surveys in which review with rating 5 is considered as negative review but the reviewer may interpret as neutral in their opinion.

Table 1: Evaluation of sentiment analysis of product review for a mobile phone model

	Positive Reviews	Negative Reviews
Correctly Assigned	27	17
Incorrectly Assigned	4	2
Precision	0.87	0.89
Recall	0.93	0.81

CONCLUSION

The proposed work has introduced us the benefits of having a sentiment analysis system that can categorize positive and negative (neutral) comments. There are many steps involved in opinion mining as natural human language requires a lot of processing and analyzing before it can become meaningful to a computer. However, by using the appropriate process, it can be done with an acceptable accuracy rate. There are still many ways that this system's method can be improved and the processes can be done more intensively once technology allows an even faster processing speed. As a conclusion, this system will be able to mine valuable information from the huge pool of resources in the internet. Recommendation of the system is that it will be able to take into account more rules regarding the polarity of a sentence. This will increase the accuracy of the system and eventually become much more reliable.

REFERENCES

1. Esuli A., 2008. Automatic Generation of Lexical Resources for Opinion Mining: Models, Algorithms and Applications. Newsletter ACM SIGIR Forum, 42(2).
2. Wanner, F., C. Rohrdantz, F. Mansmann, D. Oelke and D.A. Keim, 2009. Visual Sentiment Analysis of RSS News Feeds Featuring the US Presidential Election in 2008. Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW), Florida, USA, 8 February 2009, pp: 1-8.
3. Turney, P.D., 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. 40th Annual Meeting Association for Computational Linguistic, Philadelphia, USA, 6-12 July 2002. pp: 417-424.
4. Jindal, N. and B. Liu, 2007. Review Spam Detection. 16th International Conference on World Wide Web, New York, USA, 8 – 12 May 2007, pp: 1189-1190.

5. Hu, M. and B. Liu, Mining Opinion Features in Customer Reviews, Proceedings of the 19th National Conference on Artificial Intelligence, Italy, 25-29 July 2004, pp: 775-760.
6. Lie, B., 2010. Sentiment Analysis and Subjectivity, Handbook of Natural Language Processing, Second Edition.
7. Pang, B. and L. Lee, 2008. Opinion Mining and Sentiment Analysis. J. Foundations and Trends in Information Retrieval, 2: 1-2.
8. Boiy, E. and M.F. Moens, 2009. A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts. Information Retrieval, 12(5): 526-558.
9. Zahran, B.M. and G. Kanaan, 2009. Text Feature Selection using Particle Swarm Optimization Algorithm. World Applied Sciences J. 7(Special issue of Computer and IT): 69-74.
10. Tan, S. and J. Zhang, 2008. An Empirical Study of Sentiment Analysis for Chinese Documents. International J. Expert System with Applications, 34(4): 159-165.
11. Tian, P., Y. Liu, M. Liu and S. Zhu, 2009. Research of Product Ranking Technology on Opinion Mining. 2nd International Conference on Intelligent Computation Technology and Automation, Hunan, China, 10-11 October 2009, pp: 239-243.
12. Won, Y.K., S.R. Joon, I.K. Kyu and M.K. Ung, 2009. A Method for Opinion Mining of Product Reviews using Association Rules. 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, Seoul, Korea, 24-26 November 2009, pp: 270-274.
13. Goutte, C. and E. Gaussier, 2005. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. Advances in Computer Science LNCS, 3408/2005, pp: 345-359.