Opinion Mining on Twitter Data for Airline Services

Pakawan Pugsee[1,a], Tuangrath Chongvisuit[1,b] and Kananant Na Nakorn[1,c]

[1] Innovative Network and Software Engineering Technology Laboratory, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok, 10330, Thailand

[a]pakawan.p@chula.ac.th, [b]tuang-cat_a_kingz@hotmail.com, [c]kananant@yahoo.com

Abstract. Opinion mining on Twitter data for airline services is a valuable application which assists user in collecting messages expressing the positive opinion or the negative comment. The objective of the application is to analyze messages about airline services on Twitter data by sentiment analysis. The proposed technique uses information about syntax and semantics of words in the message to define features before learning by Naïve Bayes and generate the classification model. Then, this application can classify opinions or comments about the airline services on the variety of Twitter messages. Therefore, customers of airline services can know how good the service quality is. In addition, the airline service providers can improve services by customer satisfaction and also meet customer requirements. Furthermore, the application can help users to collect the information to make a decision for airline services. The result of application performance testing shows that the accuracy and the precision are about 70%.

Introduction

Sharing data is currently available on Social Networks. Some of them are the facts and reliable information, while the rest of them are opinion or attitude. In addition, there are many kinds of data for opinion expression such as reviews, blogs, discussions, and comments. If there is an application which can automatically analyze information and opinions, the user can gain new knowledge and the summary of opinions may be the trustworthy source for making a decision. The great benefit of this application based on the idea is revealed on studying consumer behavior by sentiment analysis. [1] Opinion or comments are the key factors of the consuming manner and actions because whenever we need to make a decision, we often seek out the opinions of others. In addition, consumers always post reviews of services which express their views. Furthermore, they exchange personal experiences about almost anything on the internet.

Sentiment analysis in another word, opinion mining involves natural language processing, computational linguistics, and text analytics to identify sentiment polarity. One basic objective of opinion mining is to classify the polarity of a given text on the document, sentence, phrase, or entity feature level depending on summaries expressed opinions or the attitude on documents, in sentences, or at entities. Therefore, sentiment classification is used to identify whether the opinion expressed in a document is positive or negative. Opinion or sentiment classification techniques can be classified into two main categories [2]: machine learning approach and the semantic orientation approach. For the machine learning, a collection of representative data labelled manually with sentiment values is necessary for training classification model. In contrast, the semantic orientation approach does not require prior training because of calculating directly by positive and negative sentiment scores, such as lexicon-based sentiment analysis. [3] presented a lexicon-based approach to extracting sentiment from text called the Semantic Orientation CALculator (SO-CAL). It is applied to the polarity classification, assigning a positive or negative label to a text and capturing the text's opinion towards its main subject matter. However, each approach has its own advantages and disadvantages, so both methods will be applied to develop the alternative method for opinion mining in this paper.

The detail of this paper will be described in the following sections. The paper is organized as follows. The related works of opinion mining are described in Section 2. The proposed method for opinion mining is elaborated in Section 3. Then, the results of the experiment on using the application

to classify opinion are demonstrated in Section 4. Finally, Section 5 summarizes the paper and presents the conclusion of this study.

Related works

There are some studies [4] [5] [6] combine the machine learning approach with the lexicon-based approach. The results of these findings are that combining both approaches will improve sentiment classification performance. [4] implemented a lexicon-enhanced method to generate a set of sentiment words based on a sentiment lexicon as features. These sentiment features produced using five sets of online product reviews were trained and learned in the machine learning approach. In [5], the Micro-blog Sentiment Analysis System (MSAS) was proposed as a system to automatically analyze customer opinions from the Twitter micro-blog service. This system used a lexicon to perform entity-level sentiment analysis and the machine learning-based approach was applied to identify opinions automatically. [6] focused on phrase-level sentiment analysis and the objective is to automatically distinguish between prior and contextual polarity, which a focus on understanding what features are important for this task. They also presented an annotation scheme for marking sentiment expressions and their contextual polarity.

According to all works in the previous paragraph, opinion mining or sentiment analysis by combining both methods are useful for customers and product providers in different domains. Sentiment classification in [4] helps both service and product providers and users to understand factors in order to influence consumer purchasing decisions from online product reviews. In addition, the system in [5] can signify the customers' opinions by product features. Furthermore, the experiments of [6] showed that the combination of all features together in the best performance of learning algorithms. Therefore, the development of opinion mining application from Twitter for an individual domain will be customers' and providers' benefit in specific domains. In this research, Twitter messages focus on airline services because there are different customers and various airline service providers. The suitable application can help customers to make decisions for airline service provider selection.

Proposed Method for Opinion Mining

The proposed method in the application uses existing tools and lexicons under the terms of the GNU General Public License. The overview of the opinion mining application is shown in Fig. 1. This technique analyzes messages by semantic information from lexicons to generate features for sentiment analysis. Sentiment classification by the machine learning technique is applied to identify positive or negative opinions and comments.

Step 1: Preprocessing. According to Fig. 1, these messages about airline services from Twitter are collected in text files before the messages are analyzed and subjective messages are filtered. In preprocessing, the special characters, punctuation, or determiners, including links of photos, and links of URLs are deleted.

Step 2: Tag words. Then, the remaining words in the message are separated into individual words and tagged the part of speech by Stanford part of speech (POS) Tagger [7]. The tagger technique is a maximum-entropy or conditional Markov model (CMM) with the University of Pennsylvania (Penn) Treebank English POS tag set. For the result, the content words are tagged into syntactic categories such as nouns, verbs, adverbs, and adjectives.

Step 3: Detect subjective words. After tagging words, all words in the message are analyzed by Wordnet Stemmer [8]. This tool is API (Application Programming Interface) for stemming which is one part of the Massachusetts Institute of Technology (MIT) Java Wordnet Interface (JWI). Stemming is to reduce words to their base forms or stems. For example, 'agree' is the stem or the base form of the words 'agrees', 'agreed', and 'agreeable'.
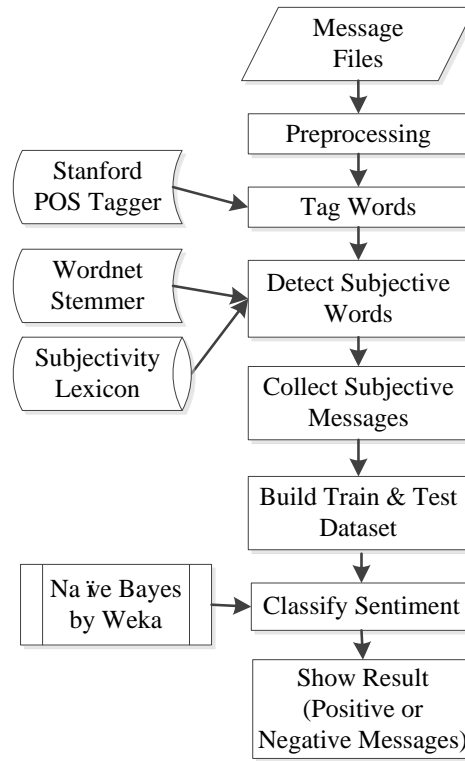
Fig. 1: The overview of the opinion mining application

This process is necessary to focus about the contents or meaning of words. Therefore, more correctness of the word detection in the next process can be achieved. To detect subjectivity words in messages, the modified subjectivity lexicon is implemented. Multi-Perspective Question Answering (MPQA) Subjectivity Lexicon [9] was the first version of the subjectivity lexicon in this research. Nevertheless, after many Twitter messages about airline services were considered manually, some words expressing sentiments are added into this lexicon. Examples of word sentiment information in the subjectivity lexicon are shown in Fig. 2.

type=strongsubj word=absolute pos=adj stemmed=n priorpolarity=neutral

type=strongsubj word=worst pos=adj stemmed=n priorpolarity=negative

type=strongsubj word=yearning pos=noun stemmed=n priorpolarity=positive

Fig. 2: Examples of word sentiment information in the subjectivity lexicon

Step 4: Collect subjective messages. Whenever messages have contained words found in the subjectivity lexicon at least one word, these messages will be identified as subjective messages and those words are marked scores by sentiment information in the lexicon. On the other hand, if there are no subjective words found in the message, these messages will be recognized as neutral statements describing the information or knowledge without opinions. Only subjective messages will be collected to be the data set for learning sentiment classification model.

Step 5: Build train and test dataset. In learning process for sentiment classification, the features of each message for classifying statements are generated using all information from previous processes, such as word parts of speech and word sentiment scores.

Step 6: Classify sentiment. Naïve Bayes classifier of Weka [10] which is the mining tool of the University of Waikato is employed to simulate the classification model. The objective of this process is to separate subjective messages into two groups that are positive statements or negative statements. Naïve Bayes is one technique of machine learning using the probability of the data. In addition, the performance of classifier depends on suitable features of data sets. Therefore, feature selection on this proposed method is based on Twitter messages about airline services.

Step **7: Show result.** Finally, the output of the opinion mining method is three categories of messages about airline services which are composed of positive, negative or neutral statements. The detail of the proposed application is described in section 4 Experiment and Result.

Experiment and Result

The experimental research was conducted on collecting Twitter Message about airline services composed of American Airlines, Asiana Airlines, EasyJet, Emirates Airlines, Etihad Airways, Garuda Airlines, Gata Airlines, Ryanair, Singapore Airlines, Southwest Airlines, Thai Airways, Turkish Airlines, and United Airlines. There are 1,529 messages that are separated into 1,022 subjective messages and 507 information messages. For subjective messages, they are identified manually by expert views as 610 positive messages and 412 negative messages.

The experiment was designed to analyze sentiments automatically using the propose application. The output from the application is separated into three classes of messages, which are positive, negative or neutral.

In leaning experiment, the feature set is composed of words contained in messages, word parts of speech and word sentiment information. Only appropriate features are inputted to the Naïve Bayes classifier for training and learning. All data messages were separated randomly into ten data sets. One fold was used as the test data set, as well as the remaining nine folds were used as the training data sets. Therefore, this leaning classification model applied 10-fold cross validation for concluding results. The results of sentiment classification are figured by values in the confusion matrix comparing the actual classes with predicted classes by the opinion mining application. A confusion matrix displays the number of correct and incorrect classifications for sentimental messages. Table 1 and Table 2 indicate the results of opinion mining on Twitter messages for airline services which are composed of the positive and the negative classification model.

Table 1: The resulting confusion matrix of positive classification

Actual Class	Truth Overall	Predicted Class	
		Positive	Not Positive
Positive	610	475	135
Not Positive	919	130	789
Total	1,529	605	924

Table 2: The resulting confusion matrix of negative classification

Actual Class	Truth Overall	Predicted Class	
		Negative	Not Negative
Negative	412	284	128
Not Negative	1,117	82	1,035
Total	1,529	366	1,163

According to the confusion matrix of positive classification in Table 1 above, values in the second column are the number of actual positive messages and not positive messages mean the summary of the number of negative and neutral messages. For sentiment analysis, 475 messages of 610 positive messages were correctly classified as the positive statement, while 135 messages were incorrectly classified as the not positive statement. On the other hand, there are 130 messages of 919 messages were incorrectly categorized as the positive statement, whereas 789 messages were correctly categorized to the not positive group.

Referring to the confusion matrix of negative predication in Table 2, the summary of the number of positive and neutral messages is the number of the not negative. In negative classification, 284 messages of 412 positive messages were actual negative statements, while 82 messages were the false negative statements. However, there are 1,035 correctly classified as the not negative statement, but 128 messages incorrectly classified as the not negative statement.

Both sentiment classification results were evaluated by the accuracy and the precision rate. Percent of the accuracy of positive and negative classification is calculated by (1). To ensure the performance of the application, Percent of the precision is demonstrated by (2). Summary of results from all evaluations are shown in Table 3.

$$Accuracy = \frac{The\ number\ of\ correct\ messages * 100}{The\ number\ of\ actual\ messages}$$ ……………………………..(1)

$$Precision = \frac{The\ number\ of\ correct\ messages * 100}{The\ number\ of\ predicted\ messages}$$ ……………………………..(2)

Table 3: Percent of accuracy and precision rates

Actual Class	Truth Overall	Correct Prediction	Predicted Class	Percent of	
				Accuracy	Precision
Positive	610	475	605	77.869%	78.512%
Negative	412	284	366	68.932%	77.596%

Referring to evaluated accuracy and calculated precision equations, the results of experiments by the application were shown in Table 3. The accuracy and precision of classifying positive opinions are more than 75%. Although the accuracy rate of negative comment classification is less than 70%, the precision rate is approximately 77%. This can be interpreted that these features of airline service messages can determine opinion or analyze sentiments effectively.

As a result, the proposed opinion mining method on Twitter messages is accurate and precise. Consequently, a lot of messages about airline services on Twitter can be analyzed for summarizing the sentiments automatically. Furthermore, this application is an advantage in the decision making in the airline business.

Conclusions

At the present time, a large volume of information is available over Social Network like Twitter. Most of contents may be opinions or comments, except those are the information or knowledge. In addition, opinions or comments are valuable in the decision making. Therefore, the application analyzing data about interesting services such as airline business on Twitter will be an actual useful assistance to airline service customers and providers.

For the previous reason, this research proposed opinion mining using subjective words' information from the modified subjectivity lexicon and applied Naïve Bayes classifier to leaning the sentiment. The outcome of the proposed mechanism is the application that can analyze many contents on Twitter messages about airline services easier and faster. Moreover, this application can help customers to make a decision about airline service selection of different airline brands. Furthermore, the airline service providers can gain information to improve the service qualities and set the marketing plans.

References

[1] B. Liu: *Sentiment Analysis and Subjectivity in Handbook of Natural Language Processing 2nd Edition*, edited by N. Indurkhya and F.J. Damerau, Chapman & Hall/CRC Press, Taylor & Franics Group (2010).

[2] B. Liu: *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data 2nd Edition*, Springer (2011)

[3] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede: *Lexicon-Based Methods for Sentiment Analysis,* Computational Linguistics Vol. 37 No. 2 (2011), p. 267-307.

[4] Y. Dang, Y. Zhang and H. Chen: *A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews,* IEEE Intelligent Systems Vol. 25 No. 4 (2010), p. 46-53.

[5] W. Chamlertwat, P. Bhattarakosol, T. Rungkasiri and C. Haruechaiyasak: *Discovering Consumer Insight from Twitter via Sentiment Analysis,* Journal of Universal Computer Science Vol. 18 No. 8 (2012), p. 973-992.

[6] T. Wilson, J. Wiebe and P. Hoffmann: *Recognizing Contextual Polarity: An Exploration of Features for Phrase-level Sentiment Analysis,* Computational Linguistics Vol. 35 No. 3 (2009), p. 399-433.

[7] K. Toutanova, D. Klein, C. Manning, and Y. Singer: *Stanford Part-Of-Speech (POS) Tagger,* The Stanford Natural Language Processing Group, Stanford University, Information on http://nlp.stanford.edu/software/ tagger.shtml.

[8] The Computer Science and Artificial Intelligence Laboratory (CSAIL): *Wordnet Stemmer (JWI 2.3.3),* Massachusetts Institute of Technology, Information on http:// projects.csail.mit.edu/jwi/api/edu/mit/jwi/morph/WordnetStemmer.html.

[9] T. Wilson, J. Wiebe and P. Hoffmann: *Multi-Perspective Question Answering (MPQA) Subjectivity Lexicon,* Multi-Perspective Question Answering Resources, University of Pittsburgh, Information on http://mpqa.cs.pitt.edu/ lexicons/subj_lexicon/..

[10] Machine Learning Group: *Weka 3: Data Mining Software in Java,* University of Waikato, Information on http://www.cs.waikato.ac.nz/ml/weka/ downloading.html.