Applying Opinion Mining Technique on Tourism Study Case:
Lake Toba

Verawaty Situmorang*, Tia Elyani, Roberto Tambunan and Yohana Gulto
Faculty of Informatics and Electrical Engineering, Del Institute of Technology,
Sitoluama, Indonesia

*verawaty@del.ac.id

Abstract. Lake Toba is the largest lake in Indonesia and the largest volcanic lake in the world. Lake Toba has unique cultures and area which makes the area as tourism destination either for local tourists or overseas tourists. Because of that, information related to Lake Toba is important for the development of Lake Toba tourism. These information are the tourists’ opinion about hotel service, accommodation, transportation and the other things related to Lake Toba tourism. These information can be obtained from social media used by the society. Both society and tourists usually use social media for telling their experiences or for assessing something that is or has been done. As a result, it can be the most potential sources of information. By applying opinion mining technique and using social media twitter as a data source, the information can be gathered, processed and classified into three categories: positive, negative and neutral in order to be used by various stakeholders. This research gathers data from tweets in social media Twitter using Indonesian language and combines Lexicon Based and Naïve Bayes. Lexicon Based method is processed for labelling data training which will be used for classification in data testing. Naïve Bayes method is processed for calculating and classifying data testing into three categories: positive, negative and neutral. The result of this research is 82% the accuracy of data and 0.85 F-measure in classifying data correctly.

1. Introduction

Information is important in the development of Lake Toba tourism. “Someone provides information on a product or a service based on different view point and 50subjectively.” (Aurchana P, et al., 2014). Furthermore, it will affect the information values from a large number of populations. The information values are tourists’ opinion about hotel service, accommodation, transportation and the other things which influence Lake Toba tourism. If these information cannot be processed properly then these information will give wrong conclusion or become useless for the development of Lake Toba tourism until get the best information.

In this era, “word-of-mouth” is the important power of electronic business (Zheng & Ye, 2009), because the public information obtained from electronic business is subjective. A subjective information can be grouped and developed until it can be useful for decision making. The technique that can be used for processing information is Sentimen Analysis. Sentiment Analysis is also called opinion mining which is a product, service, organization, individual, problem, event, topic and its attributes (Liu, 2012).
The applying Sentiment Analysis in tourism has an important impact on tourism development. Because, Sentiment Analysis identifies and analyzes opinion and emotion containing review delivered by tourists or tourism service provider (Gao, et al., 2015). In the similar research concerning opinion mining, tweets about tourism can be used to automate identification on tourism object in Filipina (Julia & Junshean, 2015). Because of that references, this research uses twitter as data source.

The tweets obtained from social media Twitter are opinion in Indonesian Language about tourism in Lake Toba area. Furthermore, the tweets can be gained, quantified and analyzed by researchers in order to be used for stakeholder as a decision making for increasing tourism management and development in Lake Toba area. Besides, many researchers applied the extension rule to the model counting problem and many amended it so as to applied it into the TP of modal logic. Still some researchers improved the extension rule, and put forward series of algorithms such as NER, RIER, etc.

Extension-rule based TP method has commended considerable respect from many related researchers. For example, Murray has applied the extension rule into the generation of the target language based on the knowledge compilation, and achieved good results. Besides, many researchers applied the extension rule to the model counting problem, and many amended it so as to applied it into the TP of modal logic. Still some researchers improved the extension rule, and put forward series of algorithms such as NER, RIER, etc.

This paper is organized as follows. In section 2, the related extension-rule based TP methods are given. In section 3, the parallel TP method based on the Semi-extension rule is presented. The experimental results of comparing the algorithm proposed in this paper with other algorithms are also presented in section 4. Finally, our work of this paper is summarized in the last section.

2. Opinion Mining Technique

Opinion mining, also called sentiment analysis, is a feeling analysis (attitudes, emotion and opinion) using natural language processing. It can be seen from the number of likes, shares or comments gained from advertising, new products, blog post and video for understanding how people give respond to cases. Whether the respond for positive thing or negative thing or sarcasm or ideology (Afzaal & Usman, 2015).

Sentiment Analysis technique separated into two categories: machine learning and lexicon-based (Gao, et al., 2015 dikutip dari A.H, et al., 2014). In this research, researchers use lexicon based approach and naive bayes classifier.

Lexicon based approach used for determining whether opinion express positive or negative thing using dictionary so that, Lexicon based approach can also be called as dictionary approach (Ding, 2008). If the opinion is positive category, the score is more than 0.2 and if the opinion is negative opinion, the score is less that 0.2 and if the opinion is neutral category, the score is between -0.2 and 0.2 (Bucur, 2014). According to (Ding, 2008), positive words are given semantic orientation score +1 and negative words are given semantic orientation score -1. The overall score will be calculated with the following formula (Ding, 2008).

\[
\text{Score}(f) = \sum_{\text{wi} \in \text{V}} \frac{\text{wi}.SO}{\text{dis}(\text{wi},f)^{\prime}}
\]

Notes:
\(\text{wi}.SO\) = positive of negative score each i sentiment
\((\text{wi}, f)^{\prime}\) = the number of sentiment word that can be found

Naïve Bayes Classifier is a method for classifying using probability formula (Ling, et al., 2014). In the basic term, Naïve Bayes classification assumes that the existence of a future is not correlated with other future (Vijayarani & Dhayanand, 2015 cited from Breheny, 2013). The advantage of Naïve...
Bayes classification is it requires a small amount in training data for estimating means and variances of the required variable for classification (Vijayarani & Dhayanand, 2015).

In the Naïve Bayes Classifier algorithm, every data training represented by attribute pair “\(x_1, x_2, x_3, \ldots, x_n\)”, where \(x_1\) is first word, \(x_2\) is second word and so forth while \(V\) is the set of data training category. In classification, the algorithm will search the highest probability from all of the tested documents category \((V_{MAP})\), where the equation is as follow (Ling, et al., 2014).

\[
V_{MAP} = \arg \max_{V \in V} \left( \frac{P(x_1, x_2, x_3, \ldots, x_n | V_j)P(V_j)}{P(x_1, x_2, x_3, \ldots, x_n)} \right)
\]  

(2)

For \(P(x_1, x_2, x_3, \ldots, x_n)\) the value is constant for all categories \((V_j)\), then the equation is as follow (Ling, et al., 2014).

\[
V_{MAP} = \arg \max_{V \in V} \left( P(x_1, x_2, x_3, \ldots, x_n | V_j)P(V_j) \right)
\]  

(3)

The equation can be simplified as follows (Ling, et al., 2014).

\[
V_{MAP} = \arg \max_{V \in V} \prod_{i=1}^{n} (P(x_i | V_j)P(V_j))
\]  

(4)

Notes:

\(V_j\) = Data category

\(P(x_i | V_j)\) = Probability \(x_i\) in \(V_j\) category

\(P(V_j)\) = Probability of \(V_j\)

For \(P(V_j)\) and \(P(x_i | V_j)\) are calculated when creating training data and the equation as follow.

\[
P(V_j) = \frac{|\text{docs } j|}{|\text{Contoh}|}
\]  

(5)

\[
P(x_i | V_j) = \frac{nk + 1}{n + |\text{kosakata}|}
\]  

(6)

|docs j| = the number of documents in j category

|contoh| = the number of documents from all categories

\(n_k\) = the number of words occurrences \(x_i\) in \(V_j\) category

\(n\) = the number of words in every category

|kosakata| = the number of all words in all categories

3. Experimental Design

After doing literature study about Sentiment Analysis in Tourism (Aurchana P, et al., 2014), researchers make process model for processing opinion with referring to the previous researches. Theoretical structure in this research can be observed in Fig. 1.
The first phase process is to take tweets using Indonesian language from Twitter using Twitter API. The tweets contain specified keyword. In acquiring tourism-related tweets, there are no certain focus for identifying tweets (Julia & Junshean, 2015). In the previous researchs, the tweets are obtained by keywords “Bangkok” and “Phuket” and other keywords using basic queries containing name of facilities and events (Julia & Junshean, 2015). In defining the keywords, this research refers to keyword “Danau Toba”, tourism object and hotel in the Lake toba tourism area.

The second phase process is preprocessing. Every tweet need to be cleaned first before processed. The purpose of this phase is to clean the tweet from words that do not contain sentiment. The preprocessing stage in this research is remove URL tweets, case folding, remove unimportant symbol, convert word, remove stopword, and stemming. This phase refers to previous research, where preprocessing includes remove URL tweets, remove punctuation, tokenization, conversion to lower case, and lemmatization (Julia & Junshean, 2015). Preprocessing that is used will be adapted based on the needs of this research.

The third phase process is analyzing sentiment using two method: Lexicon Based approach and Naïve Bayes Classifier. The combination of both methods is done sequentially by using Lexicon based method at first for determining sentiment value then it can be label for data training. Furthermore, the result data from Lexicon base will be used as data training for Naïve Bayes Classifier. That process
makes labeling data automatically. On the other words, this combination makes Lexicon based method as model of leaning on Naïve Bayes classification. Data testing will be used for measuring the accuracy from model generated by training data.

Before Naïve Bayes classification, the preprocessing tweets will be split into a number of n using N-gram. In this research, researchers use uni-gram for splitting tweets into each word. N-gram is chased because it will be assumed every word is meaningful and it influences the assessment of tweet classification. Moreover, uni-gram gives the best result rather that the other n-gram (Anggraeni, 2008) cited from (Pang, et al., 2002). The other research by Ahmad Fathan Hidayatullah and Azhari SN (2014), shows that the selection of features using uni-gram and Naïve Bayes Classifier has 73.81% accuracy.

4. Experimental Result
Performance evaluation of learning model is important to understand the quality of model, to improve model and to ensure whether the model is appropriate for classification (Oprea, 2014). The measures that most frequently used are precision, recall, and accuracy (Elgamal, 2016). Accuracy is adalah probability of labeling is given correctly (Elgamal, 2016).

In a classification that has more that two class (not a binary classification), precision will be replaced by accuracy and recall will be replaced by decision. Decision is the number of classifier classifies existing document appropriately. For the best analysis of accuracy value is more than 50% (Jayanti & Noeryanti, 2014). The formula for calculating accuracy is as follow (Pak & Paroubek, 2010).

\[
accuracy = \frac{N(\text{correct classification})}{N(\text{all classification})}
\] (7)

The formula for calculating decision is as follow (Pak & Paroubek, 2010).

\[
decision = \frac{N(\text{retrieved documents})}{N(\text{all documents})}
\] (8)

Notes:
\(N(\text{correct classifications})\) = the number of documents classified appropriately
\(N(\text{all classifications})\) = the number of all documents classified
\(N(\text{retrieved documents})\) = the number of document classified
\(N(\text{all documents})\) = the number of all testing documents

After getting the value of accuracy and decision, the value of F-Measure can be defined. F-measure serves for evaluating the performance of method in predicting polarity correctly (Padmaja, et al., 2014). The formula for calculating F-measure is as follow (Pak & Paroubek, 2010).

\[
F = (1 + \beta^2) \frac{accuracy \cdot decision}{\beta^2 \cdot accuracy + decision}
\] (9)

Notes:
\(\beta = 0.5\)
For F-measure value, if the score is 1 then the F-measure is ideal and if the score is 0 then F-measure value is bad (Padmaja, et al., 2014). In conclusion, if the F-measure value is closer to score 1 then the value is better and if the F-measure value is closer to score 0 then the value is getting worst.

The performance evaluation includes evaluation of preprocessing phase and classification phase. Input of the evaluation is raw data tweets which have not been through the preprocessing phase and classification phase. In this research, the evaluation involves not only researchers but also expert reviewer and average reviewer for evaluating the result in every phase. After evaluating preprocessing phase and classification phase then the conclusion is in Table 1.

| No | Reviewer            | Data Training |      |      |      |
|----|---------------------|----------------|----|----|----|
|    |                     | Positive       | Negative | Neutral | Total |
| 1  | Researchers         | 3667           | 998      | 6646   | 11311 |
| 2  | Researchers         | 1000           | 1000     | 1000   | 3000  |
|    | Expert & Average    | 1000           | 1000     | 1000   | 3000  |

Table 2. EVALUATION RESULT DATA TESTING

| No | Reviewer            | Data Testing | Accuracy | Decision | F-Measure |
|----|---------------------|--------------|----------|----------|-----------|
| 1  | Researchers         | 3294         | 761%     | 1        | 0.79      |
| 2  | Researchers         | 750          | 82%      | 1        | 0.85      |
|    | Expert & Average    | 750          | 80%      | 1        | 0.83      |

In Naïve Bayes classification process, the researchers find an imbalance in the number of learning document. According to Saraswati (2011), the experimental result, in the data testing obtained the understanding related to tendency to follow the largest amount of training data category. If the positive training data is greater than other categories, the classification result will tend to indicate as positive opinion, if the negative training data is greater than other categories, the classification result will tend to indicate as negative opinion and if the neutral training data is greater than other categories, the classification result will tend to indicate as negative opinion. Therefore, the number of training data need to be balanced and the classification results are only determined by the number of words in testing data without influence probability priory (Saraswati, 2011). Based on the result of testing and exploratory analysis about the effects of unbalance training data to improve the accuracy value of testing data, researchers conduct further testing with balanced training data.

The tables below indicate the example comparison between balanced and unbalanced training data.

Table 3. Unbalanced Data Calculation

| No | Word | Positive | Negative | Neutral |
|----|------|----------|----------|---------|
| 1  | danau| 0.69     | 0.42     | 0.77    |
| 2  | toba | 0.7      | 0.415    | 0.796   |
| 3  | kotor| 0.0002015| 0.00261 | 0.0008  |
|    | VMAP | 0.00002919 | 0.0001364 | 0.0001471 |
After making balanced training data, the accuracy value increases into 82% and F-measure is 0.85. Based on the result of second evaluation, the accuracy of testing data is already good because the accuracy is more than 50%. As a result, the conclusion of the second evaluation result is the opportunity to label correctly is 82% and F-measure is 0.85 where the number is closer into score 1. Based on the F-measure, it can be concluded that the performance evaluation of method in predicting polarity correctly is good. This testing is supported by expert and average reviewer. According to the survey from expert and average reviewer, the accuracy value achieve 80% and the F-measure is 0.83. This accuracy and F-measure value is close on to the test that has done by the researchers. As the result, the second test is validated by the expert and average reviewer.

After doing a research about sentiment analysis in tourism with study case Lake Toba, according to 750 testing data, the researchers find that the data tend to neutral category. In the classification of 750 testing data, there are 304 positive tweets, 37 negative tweets and 409 neutral tweets.

5. Conclusion
The combination of Lexicon Based Approach and classification can produce data training and do the classification to new data tweets. The accuracy of the data is 82% and F-measure is 0.85.

The result of the raising accuracy value is influenced by some factors that are total number of balance data training and the increase new words like negation word, sentiment word and non-standard word. The best accuracy value is gained when the training data is balanced.

In order to gain the balance data training, the researchers need to eliminate the tweets manually from learning documents that have made using Lexicon Based Approach.

The result of research is represented in a web based application that shows the classification of tweets in three categories: positive, negative and neutral.

The used of preprocessing is one of stages that is very important for cleaning the data without breaking of the tweets meaning before the classification stages. There are 6 phase in preprocessing: remove url tweets, case folding, remove unimportant symbol, convert word, remove stopword, dan stemming.

References
[1] A.H, Y., W. M., & H.K, M., 2014. Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams Engineering Journal, Volume V, pp. 1093-1113.
[2] Afzaal, M. & Usman, M., 2015. A novel Framework for Aspect-based Opinion Classification for Tourist Places. The Tenth International Conference on Digital Information Management.
[3] Anggraeni, D., 2008. Klasifikasi Topik Menggunakan Naive Bayes. Depok: Universitas Indonesia.
[4] Aurchana P, Iyyapan R & Periyasamy P, 2014. Sentiment Analysis in Tourism. International Journal of Innovative Science, Engineering, and Technology, I(9), pp. 443-450.
[5] Elgamal, M., 2016. Sentiment Analysis Methodology of Twitter Data with an application on Hajj season. International Journal of Engineering Research & Science (IJOER), II(1), pp. 82-87.
[6] Gao, S., Hao, J. & Fu, Y., 2015. The Application and Comparison of Web Services for Sentiment Analysis in Tourism. Service System and Service Management (ICSSSM).
[7] Hidayatullah, A. F., 2016. Pengaruh Stopword Terhadap Performa Klasifikasi Tweet Berbahasa Indonesia. JISKa, I(1), pp. 1-4.
[8] Jayanti, R. D. & Noeryanti, 2014. Aplikasi Metode K-Nearest Neighbor dan Analisis Diskriminasi Untuk Analisis Resiko Kredit Pada Koperasi Simpan Pinjam di Kopintra Sumber Rejeki. Yogyakarta, Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST).
[9] Julia, C. & Junshean, K., 2015. Fun in The Philippines: Automatic Identification and Sentiment Analysis of Tourism-related Tweets. IEEE 15th International Conference on Data Mining Workshops, pp. 660-667.
[10] Ling, J., Kencana, P. E. N. & Oka, T. B., 2014. Analisis Sentimen Menggunakan Metode Naive
Bayes Classifier Dengan Seleksi Fitur Chi Square. *E-Jurnal Matematika*, III(3), pp. 92-99.

[11] Liu, B., 2012. *Sentiment Analysis and Opinion Mining*. San Rafael: Morgan & Claypool Publishers

[12] Oprea, C., 2014. Performance Evaluation of The Data Mining Classification Methods. *Annals of the Constantin Brancusi University of Targu Jiu*, pp. 249-253.

[13] Padmaja, S., Fatima, P. S. S. & Bandu, S., 2014. Evaluating Sentiment Analysis Methods and Identifying Scope of Negation in Newspaper Articles. *International Journal of Advanced Research in Artificial Intelligence*, III(11), pp. 1-6

[14] Pang, B., Lee, L. & Vaithyanathan, S., 2002. *Thumbs up? Sentiment Classification using Machine Learning Techniques*. New York: Cornell University

[15] Saraswati, N. W. S., 2013. *Naive Bayes Classifier dan Support Vector Machines Untuk Sentiment Analysis*. Denpasar, Seminar Nasional Sistem Informasi Indonesia