AIP Conference Proceedings

Aspect-based sentiment analysis to review products using Naïve Bayes

Cite as: AIP Conference Proceedings **1867**, 020060 (2017); https://doi.org/10.1063/1.4994463
Published Online: 01 August 2017

Mohamad Syahrul Mubarok, Adiwijaya and Muhammad Dwi Aldhi

View Online

Export Citation

ARTICLES YOU MAY BE INTERESTED IN

The effects of pre-processing strategies in sentiment analysis of online movie reviews
AIP Conference Proceedings **1891**, 020089 (2017); https://doi.org/10.1063/1.5005422

Opinion mining feature-level using Naive Bayes and feature extraction based analysis dependencies
AIP Conference Proceedings **1692**, 020020 (2015); https://doi.org/10.1063/1.4936448

A classification of marked hijaiyah letters' pronunciation using hidden Markov model
AIP Conference Proceedings **1867**, 020036 (2017); https://doi.org/10.1063/1.4994439

© 2017 Author(s).

Aspect-based Sentiment Analysis to Review Products Using Naïve Bayes

Mohamad Syahrul Mubarok[a)], Adiwijaya[b)], Muhammad Dwi Aldhi[c)]

School of Computing, Telkom University
Jl. Telekomunikasi no. 1 Terusan Buah Batu Bandung 40257 Indonesia

[a)]msyahrulmubarok@telkomuniversity.ac.id
[b)]adiwijaya@telkomuniversity.ac.id
[c)] muhammaddwialdhi@telkomuniversity.ac.id

Abstract. Product reviews can provide great benefits for consumers and producers. Number of reviews could be ranging from hundreds to thousands and containing various opinions. These make the process of analyzing and extracting information on existing reviews become increasingly difficult. In this research, sentiment analysis was used to analyze and extract sentiment polarity on product reviews based on a specific aspect of the product. This research was conducted in three phases, such as data preprocessing which involves part-of-speech (POS) tagging, feature selection using Chi Square, and classification of sentiment polarity of aspects using Naïve Bayes. Based on evaluation results, it is known that the system is able to perform aspect-based sentiment analysis with its highest F1-Measure of 78.12%.

Keywords: Sentiment Analysis, Preprocessing, POS Tagging, Chi Square, Naïve Bayes

INTRODUCTION

In its development, online marketing publications media generally provide features to consumers for giving opinions of product being marketed. Extracting information, especially about sentiment, from product reviews are demanded for both consumers and producers to know market response, thus each of both could take necessary actions on the product based on extracted information. However, problems occur since reviews could contain incomplete information, biased information, and also diverse information. Extracting and generating information from existing reviews could be done by using sentiment analysis. It tries to find sentiment polarity of a sentence and classify it into positive or negative class. Furthermore, sentiment analysis could be used to conclude the factors or aspects that are often discussed in those opinions. In other words, it is the process of opinions mining and opinions summarization [1].

As mentioned before, several problems occurs on extracting information of product reviews. We could see those problems as a problem of uncertainty. One of machine learning methods that could be used for those problems is Naïve Bayes classifier. Naïve Bayes is one of uncertainty reasoning methods that uses probabilistic model and Bayes' rule for inference. Naïve Bayes forces a naïve assumption that is among attributes are assumed conditionally independent given the class. Nevertheless, it had been proved of having good performances for many classification problems, one of them is carried out by Xhemali et al. [2] which concentrates on comparison of three methods of Naïve Bayes, Decision Tree, and Neural Networks for classifying training course web pages. The results showed that Naïve Bayes yields best performance for their study domain. Naïve Bayes is also known to have a high level of performance with a simple calculation [3].

International Conference on Mathematics: Pure, Applied and Computation
AIP Conf. Proc. 1867, 020060-1–020060-8; doi: 10.1063/1.4994463
Published by AIP Publishing. 978-0-7354-1547-8/$30.00

In this research, we focused on aspect-based sentiment analysis which tries to find an aspect that is being discussed in an opinion and its sentiment polarity. There are three processes involving in the system. These are data preprocessing, feature selection, and classification of aspect and its sentiment. As the first step, data preprocessing is aimed to clean and prepare data for next step. The second step is feature selection in which we employed Chi Square to select a subset of relevant terms to be used in the construction of Naïve Bayes model. The last step is classification of aspect and its sentiment using Naïve Bayes. We used Naïve Bayes because it is one of uncertainty reasoning methods which we believe it is suitable approach for solving uncertainty problem as found on opinion mining task.

In addition, this research was also conducted to provide final result in the form of a summary of overall existing reviews. Data set used in this paper is about product reviews with domain of restaurants obtained from SemEval-2014 Task 4 which focused on aspect-based sentiment analysis [4]. Each review data owned label of some aspects (food, service, price, ambience, and miscellaneous) and sentiments (positive, negative, conflict, and neutral).

RELATED STUDIES

National Research Council of Canada [5] applied Multi Class Support Vector Machine (SVM) and dictionary-based approach as an additional feature in the classification process. Its classification performance in the form of F1-Measure is 88.57%. In addition, research conducted by the Xerox Research Centre Europe [6] using symbolic parser designed with special lexicon and combined with SVM obtained F1-Measure of 82.28%. Another study by the University of West Boheemia [7] using Maximum Entropy classifier with 12 features such as words, LDA, bigrams, word clusters, tf-idf, and other features provided F1-Measure of 81.04%.

A research conducted by Citius [8] proved that Naïve Bayes gave high performances in the process of sentiment analysis on English Tweets. A similar study conducted by Prasad [9] showed that Naïve Bayes was able to give high performances in the process of sentiment analysis on micro-blogging. In addition, the research on sentiment analysis conducted by Xhemali et al. [2] in which comparing three methods (e.g., Naïve Bayes, Decision Tree, and Neural Networks) showed that Naïve Bayes was outstanding against the two other methods. However, the three mentioned researches applied Naïve Bayes classifiers for classifying sentiment polarity only, and the classification of sentiment polarity was done at sentence level not at aspect level. While, this research proposed the usage of Naïve Bayes to identify an aspect on product reviews and also to classify the sentiment polarity of the aspect - i.e., sentiment analysis on aspect level. Naïve Bayes is combined with Chi Square method as well as POS tagging as a feature selectors.

SYSTEM DESIGN

Data Preprocessing

Preprocessing is an early stage in processing data to make data easier or suitable for use in the mining process [10]. Preprocessing is done for the purpose of uniformity and readability as well as the classification process. In this research, data preprocessing consisted of case folding, tokenization, stop word removal, and stemming.

The case folding is aimed to make every word in a sentence is in the form of lowercase. While tokenization is used to cut a sequence of characters from a given set of documents into pieces of word or token according to the requirements system. In the process of stop word removal, every word of the previous results is selected again. The deleted words are words which are included in a stoplist. The stoplist used in this research is the stoplist from Stanford CoreNLP [11] consisting of 219 words. Part-of-speech tagging (POS tagging) is done to provide a tag or a marker of every word in a sentence. POS tagging is typically used to analyze the linguistic text. In this research, POS tagging was done by using a library of Stanford CoreNLP [11].

The final stage of data preprocessing is stemming that is converting words to their word stem or root form. We used Porter Stemming algorithm [12] which is the most common algorithm for stemming words in English language.

Feature Selection

The purpose of this process is to extract all terms used in the classification process. At this stage, two bag-of-words were created. The first one is for a group of words that contain aspects, and the second one is for a group of words that contain tendency of sentiment polarity. The words in each group were selected based on the results of POS tagging. The words having tags of JJ, JJR, JJS, RB, RBR, RBS were grouped into a group of words containing sentiments, while the words having tags of NN, NNS, NP, and NPS are grouped into a group of words containing aspects. All words in both bag-of-words were then selected using Chi Square [10] to choose words that have high relevance to each opinion.

Naïve Bayes Classifier Model

The classification was performed for two variables called aspects and sentiments. Describing the system in generative model, we may say that both variables influence the use of words in sentences. Therefore, probability distribution of words depends upon the value of each variable. The conditional generative model of Naive Bayes in the system is described by using plate notation in Fig. 1.

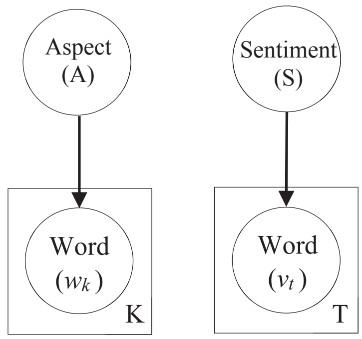

FIGURE 1. Conditional generative model using plate notation of constructed Naive Bayes.

Let $A = \{a_1, a_2, ..., a_i\}$ and $S = \{s_1, s_2, ..., s_j\}$ denote aspects and sentiments, respectively. Given a document d containing terms w_k from group of aspect and terms v_t from group of sentiment, $d = \{w_1, ..., w_k, v_1, ..., v_k\}$, the probability of document d to be categorized in aspect a_i and sentiment s_j is calculated using equation 1.

$$P(a_i, s_j \mid d) \propto \left(P(a_i) \prod_{1 \le k \le n_d} P(w_k \mid a_i) \right) \left(P(s_j) \prod_{1 \le t \le n_d} P(v_t \mid s_j) \right) \quad (1)$$

Maximum a posteriori decision rule (MAP) is used to define the final aspect and its sentiment as shown in equation 2.

$$c_{MAP} = \underset{a_i \in A, s_j \in S}{\arg\max} P(a_i, s_j \mid d) \quad (2)$$

Summary of Classification Result

The last process was making a conclusion based on sentiment polarity of certain aspects. It is done by counting the number of positive, negative, conflict, and neutral polarity. The percentage of each amount of data in a predefined aspect was calculated. The conclusion is then shown in a rating chart form. Figure 2 shows an example of rating chart as summary of classification result.

FIGURE 2. An example of rating chart as summary of classification result

SYSTEM EVALUATION

The aims of examination

The purposes of examining this system are as follows,

1) Analyzing the effect of the distribution of the number of training set against the results of the aspects and sentiments classifications.
2) Analyzing the effect of feature selection methods and the significance of Chi Square value against the results of the aspects and sentiments classifications.

Dataset

The data used in this research is dataset of SemEval 2014 Task 4 about product reviews by consumers on two entities or domain. These are laptops and restaurants. However, in this research we used restaurants domain. The data was divided into training set and test set.

1) Training set

The training set contains 3618 reviews that are divided into five aspects. These are food, service, price, ambience, and miscellaneous. Each aspect may have one of four sentiments, such as positive, negative, neutral, or conflict. The distribution of aspects and its sentiments on training set can be seen in Table 1.

TABLE 1. Data distribution of aspects and sentiments on training set.

	Positive	Negative	Neutral	Conflict	Total
Food	846	201	89	66	1202
Price	175	113	10	16	314
Service	319	217	19	34	589
Ambience	261	97	23	46	427
Miscellaneous	515	193	349	29	1086

2) Test set

Test set consists of 96 reviews containing five aspects and four sentiments as well as on training set. The distribution of aspects and its sentiments on test set can be seen in Table 2.

TABLE 2. Data distribution of aspects and sentiments on test set.

	Positive	Negative	Neutral	Conflict	Total
Food	21	8	1	1	31
Price	2	1	1	1	5
Service	5	1	1	1	8
Ambience	2	1	1	1	5
Miscellaneous	32	6	8	1	47

Results and Analysis

The topic in this research is one of subtasks of SemEval-2014 [4]. It is Task 4 focusing on aspect-based sentiment analysis. It has two subtasks. These are subtasks 3 concerning aspect detection and subtasks 4 concerning polarity category of the aspect. This study focused on the process of classifying opinion into a particular aspect and knowing the tendency of the sentiment.

Tests were carried out to examine the effect of training-test distribution to the classification results, and also to examine the effect of significant value of Chi Square features to the classification results. The results of this system were compared to the baseline and the results of other participants of SemEval-2014 at the same subtasks.

The Effect of Training Set Distribution

Data distribution on training set were examined trough two schemes. The first scheme is classification using default data distributions of SemEval-2014 that have imbalance distribution among the aspects and also among the sentiments, as can be seen in Table 1. The second scheme is classification using dataset that have been through a sampling process to get slightly equal data distributions. We combined undersampling and oversampling technique for the second scheme. We did oversampling process for the data of less than thirty so they amount to thirty. The number of thirty was chosen because according to Roscoe [13] if the sample is broken down into subsample, then the number of thirty is considered as an appropriate minimum size of subsample. The sampling process was not applied for data amount to between thirty and one hundred. While, for data amount to over one hundred we did undersampling to obtain one hundred data. Data distribution on training set after the sampling process can be seen in Table 3.

TABLE 3. Data distribution on training set resulted by sampling.

	Positive	Negative	Neutral	Conflict	Total
Food	100	100	89	66	355
Price	100	100	30	30	260
Service	100	100	30	34	264
Ambience	100	97	30	46	273
Miscellaneous	100	100	100	30	330

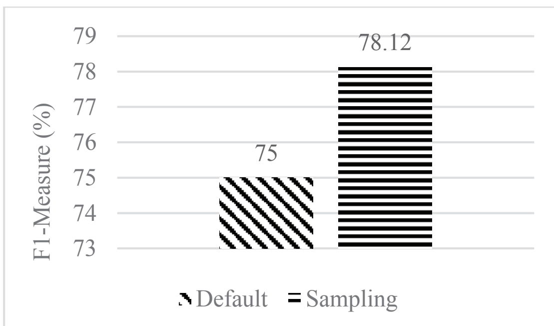

FIGURE 3. Comparison of classification between default training set and the result of sampling.

The results of overall classifications between default training set and sampling result can be seen in Fig. 3. From Fig. 3, it can be seen that the results of overall classifications using sampling training set yielded better performance than default training set. This is because the distributions of classes in the sampling training set have no significant differences compared to the default training set. As the results, the differences of prior probabilities for each class are also not very significant. The performance of aspects and its sentiments classification using the data of sampling results can be seen in Table 4 and Table 5.

TABLE 4. Performance on aspects classification.

Aspect	Accuracy	Precision	Recall	F1-Measure
Food	92.13	92.85	83.87	88.13
Price	95.34	55.56	100	71.42
Service	96.47	77.78	87.5	82.35
Ambience	97.61	80	80	80
Miscellaneous	87.23	88.89	85.1	86.95

TABLE 5. Performance on sentiment polarity classification.

Polarity	Accuracy	Precision	Recall	F1-Measure
Positive	69.76	92.85	62.9	75
Negative	77.92	50	64.7	56.41
Conflict	92.3	50	40	44.44
Neutral	71.42	28.57	66.67	40

Effect of Chi Square Significant Values to Classification Performance and Run Time

Chi Square was used to reduce the amount of data to be used in constructing Naïve Bayes models. The smaller value of significant Chi Square made higher critical value, and it produced smaller number of selected features. At the end, the number of features will influence the classification performances. Several significant values of Chi Square were observed. These are 0.2, 0.1, 0.075, 0.05, and 0.01. These values were selected because of their commonly used in the features selection process by Chi Square. In addition, we also compared the classification performances that were obtained without applying Chi Square. The results can be seen in Table 6.

TABLE 6. Classification performances on several significant values.

α	Accuracy	Precision	Recall	F1-Measure
Without Chi Square	95.87	78.12	78.12	78.12
0.2	92.86	66.67	66.67	66.67
0.1	89.76	57.41	57.41	57.41
0.075	89.76	57.41	57.41	57.41
0.05	87.88	52.73	52.73	52.73
0.01	82.48	42.00	42.00	42.00

From Table 6, we can see that the data that were not selected by Chi Square method gives best results compared to other data obtained through the selection process. This is because of the data obtained without Chi Square have more number of features that may enrich information to be learned by Naïve Bayes. Meanwhile, the data selected by Chi Square method provide less features. Hence, the learned information become less adequate. However, in term of speed of classification process, feature selection by Chi Square can speed up the computation time since it reduced the features used in the classification. The average run times of classifications without Chi Square and also with several significance values of Chi Square can be seen in Fig. 4.

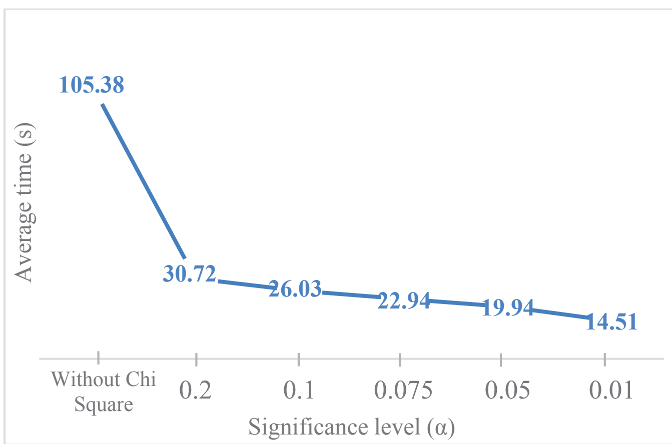

FIGURE 4. Comparison of classification run time among Chi Square significant levels and also without Chi Square.

From Fig. 4, it is obviously that Chi Square speeded up the computation time of classification. Run time of classification held by using Chi Square is three to seven times faster than without using Chi Square. However, Chi Square degraded the system performance. This comparison can be seen in Table 6. From Fig. 4, it is also obviously that each significance level of Chi Square influenced the speed of run time. The lower value of significance level made faster the run time.

Misclassified Data

In the process of classification, there were documents that were failed to be classified by system. Misclassification can be caused by any of the following reasons,

a. The number of words that make up the document tends to be slightly to the range of one to nineteen words.
b. There is an essential feature that is missing resulted by previous processes.

c. There is a dominant conditional probability of a term given of a particular class, so that when the term appears in a document with a different class, the document likely to be identified as the documents that are in a more dominant class.

Comparison of Results against Baseline and Other Studies

Table 7 displays comparison of our result against the baseline and other studies for the same subtasks and domain of SemEval-2014. It can be seen from Table 7 that the system we built ranked seventh in the subtasks and domain, and also it is above the baseline performance.

TABLE 7. Comparison of the system performance against baseline and other studies on SemEval-2014.

Team	F1-Measure
NRC-Can.	88.57
XRCE	82.28
UWB	81.04
UNITOR	80.76
SAP_RI	79.04
SNAP	78.22
Our Research	**78.12**
UBham	74.24
SeemGo	73.75
SINAI	73.67
JU_CSE.	70.46
lsis_lif	68.27
ECNU	67.29
UFAL	64.51
Baseline	**63.89**
COMMIT.	59.3

Summarization of Classification Results

To facilitate the users in viewing results of aspect-based sentiment analysis on the overall reviews, the summarization of restaurants domain was carried out in the form of rating for each aspect. The rating of each aspect was calculated based on the number of each polarity (positive, negative, conflict, and neutral) on the corresponding aspect. The summarization of restaurants domain in the form of rating chart can be seen in Fig. 5.

FIGURE 5. Summary of aspect-based sentiment analysis results.

CONCLUSION

Based on several conducted tests, it can be concluded that Naïve Bayes classifier performed well for aspect-based sentiment analysis with the best F1-Measure of 78.12%. The best F1-Measure for aspect classification is 88.13%, and the best F1-Measure for sentiment classification is 75%.

The POS tagging approach and the Chi Square method can be involved for features selection which are further used for classification process in Naïve Bayes classifier. The Chi Square also has been proven to speed up the computation time in the classification process of Naïve Bayes although it degraded the system performance.

In addition, visualizing the classification results in the form of rating chart based on specific aspects of the product is helping viewers to capture a general conclusion regarding to the assessment on the product.

ACKNOWLEDGMENTS

The authors would like to thank Telkom University for funding this research.

REFERENCES

1. M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04 (ACM, New York, NY, USA, 2004), pp. 168–177.
2. D. Xhemali, C. J. Hinde and R. G. Stone, "Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *International Journal of Computer Science Issues,* vol. 4, no. 1, pp. 16-23, 2009.
3. C. C. Aggarwal and C. Zhai, Mining Text Data, London: Springer, 2012.
4. International Workshop on Semantic Evaluation (SemEval-2014), Dublin, Ireland, 2014. [Online] http://alt.qcri.org/semeval2014/task4/. [Accessed 16 June 2016].
5. S. Kiritchenko, X. Zhu, C. Cherry, and S. Mohammad, "NRC-Canada-2014: Detecting aspects and sentiment in customer reviews," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014), pp. 437–442.
6. C. Brun, D. N. Popa, and C. Roux, "XRCE: Hybrid classification for aspect-based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014), pp. 838–842.
7. T. Brychcin, M. Konkol, and J. Steinberger, "UWB: Machine learning approach to aspect-based sentiment analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014), pp. 817–822.
8. P. Gamallo and M. Garcia, "Citius: A naïve-bayes strategy for sentiment analysis on english tweets," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 2014), pp. 171–175.
9. S. Prasad, "Micro-blogging sentiment analysis using bayesian classification methods," Tech. Rep. (Stanford University, 2010).
10. S. Garca, J. Luengo, and F. Herrera, Data Preprocessing in Data Mining (Springer Publishing Company, Incorporated, 2014).
11. C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky, "The stanford corenlp natural language processing toolkit," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014*, June 22-27, 2014, Baltimore, MD, USA, System Demonstrations (The Association for Computer Linguistics, 2014), pp. 55–60.
12. M. Porter, "The Porter Stemming Algorithm," January 2006. [Online]. Available: http://tartarus.org/martin/PorterStemmer/. [Accessed 17 June 2016].
13. J. Roscoe, *Fundamental Research Statistics for the behavioral sciences*, 2nd ed. (Holt, Rinehart & Winston, New York, 1975).