A Performance Comparison of Feature Selection Methods for Sentiment Classification

Lai Po Hung[(✉)], Rayner Alfred , and Mohd Hanafi Ahmad Hijazi

Knowledge Technology Research Unit (KTRU), Faculty of Computing and Informatics,
Universiti Malaysia Sabah (UMS), Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia
janelai627@gmail.com, {ralfred,hanafi}@ums.edu.my

Abstract. Document sentiment analysis is the task of determining whether a document has a positive, negative or neutral sentiment. It is made up of subtasks including feature extraction, feature selection and sentiment classification. Feature selection is the task of selecting relevant features that can aid the classifier to produce better results. This paper focuses on comparing the classification performances based on several feature selection methods used to select relevant features and also to minimize the document-term matrix representation of the documents. The purpose of applying feature selection besides selecting relevant features is also to reduce the number of features to preserve the efficiency of the whole system. In this work, the experiment setup is designed in order to investigate the effectiveness of several selected feature selection methods in improving the sentiment analysis results. Based on the findings from the experiment, although common feature selection methods such as Document Frequency (DF), Information Gain (IG) and Chi-Squared Statistics (CSS) are found to be able to produce high sentiment analysis accuracy, the Categorical Probability Proportional Difference (CPPD) method is found to be more effective as it produces higher performance accuracy in classifying the documents based on the sentiments. Although, the Categorical Proportional Difference (CPD) method produces acceptable classification results, it is weak in reducing the number of features. In short, the CPPD method enables the sentiment analysis task to be conducted with higher accuracy rate couples with high feature reduction rate too.

Keywords: Sentiment analysis · Feature selection · Feature extraction
Information Gain · Chi-Squared Statistics · Categorical Proportional Difference

1 Introduction

Public opinion contributes a significant impact on the operations of business and product perception. Sentiment analysis is the task of classifying a document into positive, negative or neutral sentiment groups. This task is important in opinion mining. Opinions are useful resources as highlighted above, and the internet has enabled an enormous amount of opinion to be circulated. Harvesting and summarizing large amount of data is better done with technology than manual labor. Sentiment analysis plays an important role in which it analyses the opinion so that a summary of general opinion polarity can be produced. Sentiment analysis is normally done by applying a rule based system or a

machine learning based system. Machine learning based systems are more popular because they are easier to apply and more versatile. A machine learning algorithm is trained to recognize the underlying patterns in documents in order to classify documents. There are three essential steps involved in machine learning based sentiment analysis and they are feature extraction, feature selection and the machine learning classifier [1]. Features represent a single unit used in the analysis process when classifying documents into the corresponding polarities. Huge amount of features will burden the overall system with heavy processing load. Irrelevant features will cause over fitting or under fitting models of classifiers to be trained and produced. It is ideal for the system if the feature set is considerably small but informative and accurate. Feature extraction and selection is responsible for producing the desired feature set. Feature extraction involves extracting features from the text that can be recognized and analyzed by the classifier. These features include N-Grams, Part of Speech, Named-Entity Recognition, Senti-WordNet [1, 2, 25].

Feature selection performs the tasks of filtering the extracted features for useful features and removing irrelevant features. Feature selection also helps reduce the feature set size to preserve the efficiency of the system. Feature Selection brings two benefits to the system. Firstly, using smaller feature set allows the system to perform efficiently as it takes less time to finish its run but maintains its accuracy. Complex classifying algorithms that involve many processes with large amounts of features, will require feature selection to ensure that the algorithm does not become impractical due to the large amount of processing load and time required [3]. Secondly, irrelevant features can negatively impact the performance of the system as they affect the quality of model produced during the modelling process [4]. This paper focuses on comparing the classification performances based on several feature selection methods used to select relevant features and also to minimize the document-term matrix representation of the documents. These feature selection methods include some of the common methods used such as Document Frequency (DF), Information Gain (IG) and Chi-Squared Statistics (CSS), and also less common methods such as Total Term-Frequency Inverse Document Frequency (TF-IDF), Categorical Proportional Difference (CPD), Probability Proportional Difference (PPD) and Categorical Probability Proportional Difference (CPPD). There are minimal to none performance comparison performed on the less common methods as they are relatively new, and TF-IDF is normally used for feature weighting, however we chose to test the total TF-IDF as a feature selection method instead.

This paper contains several parts. Section 2 discusses previous work in feature selection for sentiment analysis. Section 3 explains the experimental setup and Sect. 4 provides the results obtained and discusses the findings based on the experimental results. Section 5 concludes this paper by discussing future works to be explored which can bring improvement the process of sentiment analysis.

2   Related Works

There are several feature selection methods that can be used to perform the feature scoring [5]. These feature selection methods are described next.

Document Frequency (DF) is a simple method that is scalable to the size of the dataset. This method selects based on the counts of documents in a training set where the feature term occurs. It removes feature terms with document frequencies which are under or over given thresholds. The assumption applied here considers feature terms that are too rare and too common occurring in training documents to be less helpful in improving accuracy of classification in sentiment analysis [6].

Information Gain (IG) is a method that values the usefulness of a term for classification based on the information it can provide to help the classifier discern between classes. It measures the amount of information a term carries by comparing the entropy of the term absence and term presence in a document [5]. The formula is shown below,

$$
\begin{aligned}
IG(f) = & \left[-\sum_{i=1}^{|c|} P(c_i) \log P(c_i)\right] - \left[-P(f) \sum_{i=1}^{|c|} P(c_i|f) \log P(c_i|f)\right] \\
& - \left[-P(f') \sum_{i=1}^{|c|} P(c_i|f') \log P(c_i|f')\right]
\end{aligned}
\tag{1}
$$

where $P(c_i)$ is the probability of the class occurring, $P(f)$ is the probability of the feature term presence and $P(f')$ is the probability of the feature term being absent [7].

Mutual Information (MI) is differs from IG as it does not consider the absence of term [8]. MI measures the dependencies between two variables. When the MI score is higher, it means that the variables are more dependent on each other as shown below, where $P(f,c)$ is the probability of the feature term and class occurring together, $P(f)$ is the probability of feature term occurrence, $P(c)$ is the probability of class occurrence.

$$
MI(f,c) = \log \frac{P(f,c)}{P(f) \times P(c)}
\tag{2}
$$

The drawback of MI is that rare terms with equal conditional probabilities will have higher scores than those of common terms. Considering the drawback, MI scores of features that have great difference in frequencies cannot be compared [7].

Chi-Squared Statistics (CSS) can measure the association strength between a feature term and a class by calculating how dependent the term and class is on each other. If a term and class is dependent on each other, then the term is useful in identifying the class it belongs to. The formula and explanation [1] is as below,

$$
\chi^2(f,c) = \frac{N \times (AD - CB)^2}{(A+C) \times (B+D) \times (A+B) \times (C+D)}
\tag{3}
$$

where A is the frequency that f and c occurs together, B is the frequency that f occurs without c, C is the frequency that c occurs without f, D is the frequency where neither c nor f occurs and lastly N is the total number of documents.

Simeon and Hilderman **proposed the Categorical Proportional Differences (CPD)** method to measure the usefulness of a term to differentiate between different categories [9]. The measurement is based on the ratio of the frequencies of a word occurring across different categories of documents. The calculation is as shown,

$$CPD(f, c) = \frac{A - B}{A + B} \qquad (4)$$

where A is the frequency of the word and category appearing together and B is the frequency of the word occuring without the category. When CPD score approaches maximum score of 1, this shows that the feature occurs more frequently in documents of a particular category only and is helpful to distinguish between categories. CPD is quite a recent method and has been used in a few experiments only [9, 10]. The benefit of CPD is that it can eliminate common terms with high document frequency but are not important such as stop-words, based on their equal occurrence in all classes of documents [11]. On the other hand, the weakness of CPD is that it might retain irrelevant rare terms due to their low document frequencies.

A variation of the CPD method that combines CPD and another method called the Probability Proportional Difference (PPD) method has also been proposed, in which the authors proposed as a measure of the feature terms' relation with a particular category [11]. A high PPD value shows that the term belongs to a certain class, whereas, a low PPD value shows that it appears in multiple classes and is not very helpful for classification. The PPD calculation for sentiment analysis is computed as follows,

$$PPD(f) = \frac{nPos(f)}{sPos + S} - \frac{nNeg(f)}{sNeg + S} \qquad (5)$$

where $nPos(f)$ represents the number of positive documents where the feature is present, $nNeg(f)$ represents the number of negative documents where the feature is present, $sPos$ represents the number of features in the positive class, $sNeg$ represents the number of features in the negative class and S represents the total number of features. The combined method proposed is the Categorical Probability Proportional Difference (CPPD) method which selects feature terms that achieves both thresholds set for CPD and PPD. The advantage of this method is that CPD and PPD compensate for the weaknesses of each other. When CPD selects irrelevant terms with low frequency, PPD will eliminate the term, whereas when PPD selects irrelevant terms with high frequency, CPD will eliminate the term.

Table 1 summarizes several works performed which focus on performing a comparison of the performance of various feature selection methods. The popular methods that have always been used are DF, IG and CSS, these methods are methods that are matured and frequently used by researchers for text classification related works.

DF method is the simplest method of selection but it is weaker in terms of performance. IG and CSS are more commonly used among classification tasks because they have better performances. Three of the mentioned methods measure the usefulness of a feature through consideration of the presence and absence of the feature among documents in a feature set. The IG and CSS methods are also aggressive selectors adding to their favor, the two of them are commonly applied in works of text classification [12–16].

Table 1. Previous works that performed comparison on feature selection methods

Authors	Feature selection methods	Classifiers	Dataset	Best performing method
Yang and Pedersen [7]	DF, IG, MI, CSS, Term Strength	k-Nearest Neighbor(k-NN), Linear Least Squares Fit	Reuters-22173 collection, OHSUMED collection of medical journals	IG, CSS
Tan and Zhang [17]	DF, MI, CSS, IG	Centroid classifier, k-NN, Naïve-Bayes, Winnow classifier, Support-Vector-Machine(SVM)	ChnSentiCorp-Chinese sentiment documents	IG
Sharma and Dey [6]	DF, IG, Gain Ratio, CSS, Relief-F algorithm	Naïve Bayes, Max Entropy, SVM, Decision Tree, k-NN, Winnow classifier, AdaBoost classifier	Cornell Movie Review Dataset	Gain Ratio
O'Keefe and Koprinska 10]	CPD, SentiWordNet Subjectivity Scores, SentiWordNet Proportional Difference	Naïve Bayes, SVM	Cornell Movie Review Dataset	CPD
Sarkar and Goswami [8]	IG, MI, CSS, Symmetrical Uncertainty	Naïve Bayes, SVM,Decision Tree, k-NN	CNAE-9, SMS Spam Collection, Preview of hotel data	IG
Zhao et al. [18]	DF,IG, CSS, MI	Naïve Bayes, Max Entropy, SVM, Winnow classifier, C4.5 Decision Tree Classifier	Chinese news comments	CSS

3   Experimental Setup

In this paper, the experimental setup is designed to investigate the effectiveness of feature selection methods in influencing the classification accuracy of the sentiment analysis and also the ability to reduce the feature count. Based on the best performing methods listed in Table 1, seven feature selection methods are implemented and evaluated in this work, and these methods include the DF, IG, CSS, Total TF-IDF, CPD and its variation, PPD and CPPD.

The dataset used here is obtained from the Multi-Domain Sentiment Dataset [19]. The dataset consist of English language product reviews from different domains. The reviews chosen in this work are obtained from the domains of Books, DVD and Electronics. For each domain, 1000 reviews are obtained and the reviews are preprocessed and represented as features consisting of a combination of unigrams and bigrams.

In this work, the TF-IDF method is used as a feature weight in the term document matrix for classification and the Total TF-IDF value for a feature in all reviews is used as a feature selection metric. TF-IDF value rare terms higher based on the observation that rare terms can be more selective. The TF-IDF is defined as follows,

$$TFIDF = TF \times IDF = (1 + \log_2 tf) \times (\log_2 \frac{N}{n_i}) \tag{6}$$

where, *tf* is term frequency in a document, N is the total number of documents in the dataset and n_i is the number of documents where the term occurs.

Naïve Bayes classifier is used in this experiment because its application is simple and fast and its performance in text classification is quite highly regarded by many researchers [6, 20–23]. Naïve Bayes classifiers assume that variables have conditional independence and calculate the probability of a document or feature being assigned to a particular class (positive or negative). This classifier is built based on the Bayes theorem [1]. In this experiment, all classification tasks will be conducted by using the classifiers obtained from the WEKA (Waikato Environment for Knowledge Analysis) system developed by the University of Waikato. Weka is a data mining tool that contains a group of machine learners and other processing applications. The features can be run in WEKA after converting the file format to ARFF format. The training and testing of classifiers based on cross fold validation is carried out within WEKA [24] (Fig . 1).
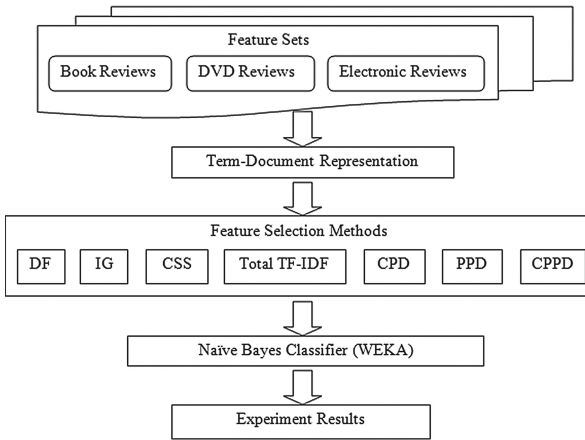

Fig. 1. The experimental setup designed to investigate the effectiveness of feature selection methods on the classification accuracy of sentiment analysis

The feature selection methods outlined previously assigns each feature a value that reflects their usefulness for classification. Prior to classification, only features with values that exceed certain thresholds are chosen and used. In the experiments conducted, 10 threshold values are generated for each feature selection method in each dataset experiment where the values are in increasing order. The 10 thresholds are generated by arranging all unique values calculated by a feature selection method in increasing order, and selecting the threshold value found within 10 equal intervals in position. Results for the threshold that has the best sentiment analysis accuracy of each method will be selected and shown as the results in Sect. 4 in order to compare the performances of the feature selection methods.

4 Results and Discussion

Table 2 outlines the results obtained from running sentiment analysis on three different dataset domains with different feature selection methods. The best threshold is obtained by taking the threshold scores of the feature selection methods that produce the best classification accuracy. The feature count represents the number of features selected after applying the threshold and the reduction rate provides the percentage of features reduced by the threshold and finally the accuracy rate presents the classification accuracy of sentiment analysis performed with the features selected.

Table 2. A performance comparison of feature selection methods results on document sentiment analysis for three topics: Book, DVD and Electronic reviews datasets

Feature selection method	Best threshold	Feature count	Reduction rate (%)	Accuracy rate (%)
Book				
No feature selection	–	113273	0.0	62.1
Document frequency	35	1308	98.8	63.1
Information gain	0.00331	1116	99.0	76.2
Chi-Square	7.23037	199	99.8	75.7
Total TFIDF	186.12464	570	99.5	63.4
CPD	1.0	98262	13.3	**98.6**
PPD	0.00021	27	**99.9**	70.7
CPPD	0.00002	2002	98.2	97
DVD				
No feature selection	–	110715	0.0	72.7
Document frequency	18	1313	98.8	72.6
Information gain	0.00331	1150	99.0	84.9
Chi-Square	4.96159	620	99.4	84.5
Total TFIDF	167.88627	678	99.4	72.9
CPD	1.0	96042	13.3	**99.1**
PPD	0.00006	477	**99.6**	79.9
CPPD	0.00002	2032	98.2	96.3
Electronics				
No feature selection	–	66524	0.0	74.1
Document frequency	15	1247	98.1	75
Information gain	0.00293	1929	97.1	82.6
Chi-Square	6.49176	300	99.5	82.5
Total TFIDF	119.00907	950	98.6	74.7
CPD	1.0	56816	14.6	**90.8**
PPD	0.00012	285	**99.6**	81.8
CPPD	0.00002	5237	92.1	**90.8**

Based on the comparison results shown in Table 2 above, it can be concluded for this experiment that CPD is the best feature selection method out of the others since it produces the highest accuracy for the sentiment analysis classification results. However CPD's reduction rate is the lowest of all. While others achieve reduction rate of 90% and above, CPD is the only one that achieved reduction rate of less than 20%. The PPD managed to reduce the most number of features from 100% to only less than 1% of the

features compared to the other methods. The weaker methods that produced lower accuracies are DF and Total TF-IDF. Based on the observations of the results, it can be summarized that the pair of methods (DF and Total TF-IDF), (IG and CSS) and (CPD and CPPD) methods produce similar results to each other. This observation means that these pairs of methods have similar behaviors as shown in their formulas.

The DF and total TF-IDF methods rely on the presence of features as the basis to their measurement and their results are lower compared to the others. DF counts how many document does the feature appear in, but it does not discern whether the appearance of the feature is equally distributed among documents of different classes or more skewed towards document of a certain class. As a result, it cannot find features that are discriminative. Since the TF-IDF method weights features more by their value in a document, the poor result obtained suggests that the importance of a feature in a single document does not emulate the importance of the feature in the whole dataset. The two methods can be used as good selectors if the focus of the task is on the reduction of features. This is because they are quite aggressive in feature reduction, in which more than 90% of the features in datasets are reduced and the classification accuracies are acceptable although they are weaker when compared to the other feature selection methods.

The IG and CSS methods also produce similar results and calculation parameters. In order to compute the score of feature by using the IG and CSS methods, the frequency of feature appearance and absence in a certain class need to be taken into consideration. In other words, when measuring the importance of a feature in the dataset, the two methods compare whether the feature is dispersed equally among classes or not. This is probably why they performed better than the two previous methods (e.g., DF and TF-IDF). IG and CSS methods are considered as very good feature selection techniques because they are aggressive in feature reduction and at the same time with the reduced features they can produce acceptable classification accuracy that are quite high. For these reasons they are very popular methods used in classification tasks.

The CPD and CPPD methods yield the best results in the experiments. CPPD is basically a method that is designed based on two previously defined methods, CPD and PPD. Both of these methods compare the presence and absence of particular features. However, the CPD method tends to eliminate common terms, whereas PPD method tends to remove rare terms. Based on the results obtained, it is found that keeping rare terms yield better classification accuracies compared to keeping common terms as the CPD method outperforms the PPD method. The CPD method is able to produce high accuracies because it can select terms that are more exclusive to a certain class only, however the downfall of it is that it is very poor in feature reduction since its reduction rate is less than 20% in all the three datasets. This is probably due to fact that CPD retains the extremely rare features that appear in only one document out of the whole datasets. These features are either extremely informative or useless. The PPD method on the other hand, produces lower accuracies compared to the CPD method but it is very aggressive in feature reduction. The PPD method was created to behave the opposite of CPD as it eliminates those rare features that occur in only one document. Thus, the CPPD method, which is the result of combining both the CPD and PPD methods, enjoys the best of both worlds as it manages to reduce the features by more than 90%, and at the same time

retains quite high performance too. So, when considering the traits of a good feature selector which is increased performance and high feature reduction, the CPPD method is considered to be a better feature selection method for classifying document into classes of sentiment polarities.

5   Conclusion

This paper presents an assessment of several feature selection methods that can be applied to select relevant features for the purpose of classifying documents based on sentiment analysis. An experiment has been conducted to compare seven feature selection methods that include DF, IG, CSS, Total TF-IDF, CPD and its variation, PPD and CPPD. Three datasets have been used to perform the assessment which are the reviews taken from three main topics (e.g., Book, DVD and Electronics). The reviews are represented as combinations of unigram and bigram and a Naïve Bayes classifier is used to perform the sentiment classification. Based on the results obtained, the best sentiment analysis result can be obtained when the CPD method is applied to reduce the number of features presented to the Naïve Bayes classifier in classifying documents based on the sentiment polarities (e.g., positive, negative and neutral polarities). However, the number of features reduced is minimal and the large set of features selected is not efficient to be used in the classification task. On the other hand, the PPD method yields lower classification accuracies with higher percentage of features reduced. The CPDD method is found to be very effective in balancing the classification accuracy and the percentage of features reduced. CPPD combines CPD and PPD elements and this method is able to create a balanced performance between CPD and PPD methods since it can achieve high sentiment classification accuracy and also high percentage of feature reduction rate in achieving that accuracy. Since, feature extraction process may also influence the classification accuracy of classifying document into sentiment polarities, future works may include the task of classifying documents into positive, negative and neutral sentiments based on types of methods used for feature extraction. These feature extraction methods include N-gram, stemmed, Part of Speech, and Named-Entity Recognition based methods.

References

1. Hung, L.P., Alfred, R., Hijazi, M.H.A., Ibrahim, A., Asri, A.: A review on the ensemble framework for sentiment analysis. Adv. Sci. Lett. **21**(10), 2957–2962 (2015). http://dx.doi.org/10.1166/asl.2015.6494
2. Alfred, R., Leow, C.L., Chin, K.O., Anthony, P.: Malay named entity recognition based on rule-based approach. IJMLC **3**(4), 300–306 (2014)
3. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval the Concepts and Technology Behind Search, 2nd edn. Pearson Education Limited, England (2011)
4. Vanaja, S., Ramesh, K.K.: Analysis of feature selection algorithms on classification: a survey. Int. J. Comput. Appl. **96**(17) (2014). 0975–8887
5. Hung, L.P., Alfred, R., Hijazi, M.H.A.: A review on feature selection methods for sentiment analysis. Adv. Sci. Lett. **21**(10), 2952–2956 (2015). http://dx.doi.org/10.1166/asl.2015.6475

6. Sharma, A., Dey, S.: A performance investigation of feature selection methods and sentiment lexicons for sentiment analysis. IJCA Special Issue Adv. Comput. Commun. Technol. HPC Appl. **3**, 15–20 (2012)

7. Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: Proceedings of the Fourteenth International Conference on Machine Learning, ICML 1997, pp. 412–420 (1997)

8. Sarkar, S.D., Goswami, S.: Empirical study on filter based feature selection methods for text classification. Int. J. Comput. Appl. **81**(6), 38–43 (2013)

9. Simeon, M., Hilderman, R.: Categorical proportional difference: a feature selection method for text categorization. In: Proceedings of the 7th Australasian Data Mining Conference, vol. 87, pp. 201–208 (2008)

10. O'Keefe, T., Koprinska, I.: Feature selection and weighting methods in sentiment analysis. In: Proceedings of the 14th Australasian Document Computing Symposium, Australia (2009)

11. Agarwal, B., Mittal, N.: Categorical probability proportion difference (CPPD): a feature selection method for sentiment classification. In: Proceedings of the 2nd Workshop on Sentiment Analysis where AI Meets Psychology (SAAIP 2012), Mumbai, pp. 17–26 (2012)

12. Azhagusundari, B., Thanamani, A.S.: Feature selection based on information gain. Int. J. Innovative Technol. Explor. Eng. (IJITEE) **2**(2), 18–21 (2013). ISSN 2278–3075

13. Hassan, A., Abbasi, A., Zeng, D.: Twitter sentiment analysis: a bootstrap ensemble framework. In: International Conference of Social Computing, pp. 357–364 (2013)

14. Koncz, P., Paralic, J.: An approach to feature selection for sentiment analysis. In: 15th International Conference on Intelligent Engineering Systems, Poprad, Slovakia (2011)

15. Shang, L.: A feature selection method based on information gain and genetic algorithm. In: International Conference on Computer Science and Electronics Engineering (2012)

16. Uchyigit, G.: Experimental evaluation of feature selection methods for text classification. In: 9th International Conference on Fuzzy Systems and Knowledge Discovery (2012)

17. Tan, S., Zhang, J.: An empirical study of sentiment analysis for Chinese documents. Expert Syst. Appl. **34**, 2622–2629 (2008)

18. Zhao, Y., Dong, S., Lee, L.: Sentiment analysis on news comments based on supervised learning method. Int. J. Multimedia Ubiquit. Eng. **9**(7), 333–346 (2014)

19. Blitzer, J., Dredze, M., Pereira, F.: Biographies, bollywood, boom-boxes and blenders: domain adaptation for sentiment classification. Association of Computational Linguistics (ACL) (2007)

20. Wang, G., Sun, J., Ma, J., Xu, K., Gu, J.: Sentiment classification: the contribution of ensemble learning. Decis. Support Syst. **57**, 77–93 (2014)

21. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, Philadelphia, Association of Computational Linguistics, pp. 79–86 (2002)

22. Su, Y., Zhang, Y., Ji, D., Wang, Y., Wu, H.: Ensemble learning for sentiment classification. In: Ji, D., Xiao, G. (eds.) Proceedings of 13th Chinese Conference on Chinese Lexical Semantics (CLSW 2012), pp. 84–93. Springer, Heidelberg (2012)

23. Xia, R., Zong, C., Li, S.: Ensemble of feature sets and classification algorithms for sentiment classification. Inf. Sci. **181**, 1138–1152 (2011)

24. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explor. **11**(1), 10–18 (2009)

25. Lai, P.H., Alfred, R.: A performance comparison of feature extraction methods for sentiment analysis. In: Król, D., Nguyen, N., Shirai, K. (eds.) Advanced Topics in Intelligent Information and Database Systems. ACIIDS 2017. SCI Book Series vol. 710 (2017)