

An Ensemble Sentiment Classification on Multidomain Dataset



S. Madhusudhanan, N. M. Jyothi, A. Vishnukumar, S. Sathya,
and N. Arun Vignesh

Abstract With the introduction of web 2.0, individuals now have a perfect platform to communicate their ideas, thoughts, and feelings. Web Opinion Mining / Sentiment Analysis is a text mining job aimed at developing a system that automatically extracts, recognizes, and categorizes user opinions from natural language text, user provided content, or user generated media. We have employed classifiers to analyse data sets from a wide variety of domains. The ensemble algorithm is applied to boost its performance. With an delicacy of 80.93%, the Logical Regression using Ensemble Classifier exceeds the others.

Keywords Text classification · Sentiment classification · Machine learning · Bagging

1 Introduction

Blogs, forums, and Internet community are allowing users to share their opinion on any issue. For example, express their dissatisfaction with a product they purchased, discuss current events, or express their political opinions [5]. This form of user data analysis is required for recommender systems and customization. “Everyone cares

S. Madhusudhanan (✉) · N. M. Jyothi
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur (DT), Andhrapradesh, India
e-mail: ssmadhu@kluniversity.in

A. Vishnukumar
Department of Information Technology, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala
Engineering College, Vel Nagar, Avadi,Chennai, Tamilnadu, India

S. Sathya
Department of Electronics & Communication Engineering, Gojan School of Business and
Technology, Edapalayam, Redhills, Chennai, Tamilnadu, India

N. Arun Vignesh
Department of Electronics and Communication Engineering, Gokaraju Rangaraju Institute of
Engineering and Technology, Hyderabad, India

about what other people think.” A persuasion or a judgement about anything that has a substantial impact on a person’s decision-making process is referred to as an opinion. The practise of determining people’s feelings or views is known as sentiment analysis or opinion mining. Circumstances affect people’s thoughts, feelings, and sentiments. For example, “The film was filled with fun elements” is a positive opinion about the movie, “he made fun with Ram’s appearance” is a negative opinion [2, 6].

Sentiment analysis is an understanding of natural language exertion that attempts to categorize the texts predicated on their expressed sentiments on a decisive content [1, 4]. It is a technique for determining a speaker’s or writer’s attitude towards a topic or the overall impression of a text. Because of its multiple and potential uses, such as automatic product review classification, it has three types of categorization: document, phrase, and feature. At the document classification stage, the authors discuss the entire document and decide whether it is favourable or unfavourable. Sentiment categorization evaluates each sentence separately to determine its polarity at the phrase level. At the feature level, we may categorize sentiment in terms of many different features of things. Aspect level emotion categorization demands a more comprehensive analysis because most features are provided implicitly.

1.1 Motivation

Prior sentiment analysis research resulted in the development of a classification model for a specific domain. So, a multidomain sentiment analysis is offered in order to align the classifier across many domains simultaneously [9, 10]. The main contribution of the proposed work are as follows:

- We employed the TF-IDF feature selection with Chi square (χ^2) method for best feature selection.
- Sentiment classification based on selected feature are implemented with multidomain dataset.
- To increase classification performance, a classifier with bagging is employed.
- Sentiment classification using different regression analysis methods are also employed in this work.
- Finally, based on evaluation measures, the comparisons are made.

Section 2 discussed about the existing work, Sect. 3 implementation of proposed method, Sect. 4 contains result and discussion, and Sect. 5 about conclusion and future ideas.

2 Related Work

The authors [20] attempt to address the issue by developing a sentiment aware dictionary utilizing data from several domains. By using dictionary, they have assorted the

target domain's unlabeled reviews. The task was performed in Hindi with a 76% accuracy rate. As an approach for cross-domain sentiment classification, the authors [3] propose a sentiment sensitive distributional thesaurus where sentiment sensitivity in thesaurus is achieved by including sentiment labels in the context vector at the document level. Using a benchmark data set that comprises Amazon user reviews for a variety of product categories, the proposed approach outperforms alternative baselines and achieves results equivalent to previously reported cross-domain sentiment classification algorithms.

The researchers [21] combined sentiment data from four different sources. Sentiment lexicons are a good place to start because they include sentiment polarity for widely used sentiment terms. Sentiment classifiers from diverse source domains are the second source. As the third source is to construct domain-specific sentiment correlations between words.

The target domain's tagged data is the fourth source. They offer a unified architecture for gathering all four forms of sentiment data and training a domain-specific sentiment classifier in the target domain.

The authors [13] suggested a fuzzy technique to describe the polarity learnt from training sets or from a training set. This newly acquired knowledge is combined with additional conceptual knowledge collected from two widely known sentiment analysis resources, SenticNet and the General Inquirer vocabulary. The advised strategy yielded the best results.

Researchers [11] developed a multidomain sentiment classification technique that reduced domain reliance whilst increasing overall performance. The suggested approach employs a way of combining several classifiers. They employ a method that involves training domain classifiers individually with domain-specific data before merging the classifiers to get the desired output.

The authors [8] investigated the capability of four different machine learning classification algorithms utilizing frequently used feature selection methods with three fame datasets are used to evaluate the proposed approaches: IMDb movie reviews, electronics reviews, and kitchen items reviews. The first step is to choose feature subsets from one of three available feature selection methods. After that, set theory concept such as union and intersection are applied to evoke the top ranking features. The combined approach is achieved the greatest accuracy of 92.31% on the SMO classifier.

The authors [12] investigated the performance of a few feature selection techniques for sentiment analysis. A feature extraction method called as Term Frequency Inverse Document Frequency is utilized to create feature vocabulary. To choose the best set of word vectors, a variety of Feature Selection techniques are employed. Machine learning classifiers are used to teach the required attributes. To improve sentiment analysis performance, classifiers use bagging and random subspace. They proved the effectiveness of feature selection strategies trained with ensemble classifiers outperform neural networks with far less training time and parameters, removing the requirement for hyper-parameter modification.

3 Proposed Approach

Sentiment Classification steps:

Step1: In this work, a multidomain product review database(Books, DVD, Electronics, Kitchen and Housewares) is collected and used to tackle this problem.

Step2: Preprocessing methodology is required to eliminate noisy, inconsistent, and incomplete data by considering tokenization, stop words removal, and stemming approaches.

Step3: Feature Extraction and Selection: To begin, we utilize the TF-IDF and Bag of Words (BoW) methods to generate a feature vector in a document. it will receive a score of 1 if it is present, otherwise the score is 0. Following that the CHI feature selection approach is used to pick distinct feature subsets.

Step4: Classification: To train the given feature, Support Vector Machines (SVM-RBF), SVM-(Linear Kernel), MNB, Decision Tree (DT), and Linear Regression (LR) are employed.

Step5: Bagging: Finally, the ensemble process helps to escalate the classification accuracy.

3.1 Methodology

The proposed architecture is depicted in Fig. 1, with further information on each preliminary function provided in the subsections that follow. The sentiment categorization job in this study is performed using multidomain product review data. In the dataset, there are 1000 positive and 1000 negative tagged reviews in each domain [14, 15] are chosen for implementation. Table 1 summarizes the statistics for this data set.

3.1.1 Preprocessing Task

Tokenization: The tokenizer may separate review into distinct tokens like as words, numbers, special characters, and so forth, making it ready for further processing.

Stop word Removal: To improve the effectiveness of the feature selection strategy,

Table 1 Multidomain sentiment dataset

Dataset	Total reviews	Positive review	Negative review
Books	2000	1000	1000
DVD	2000	1000	1000
Electronics	2000	1000	1000
Kitchen and housewares	2000	1000	1000

this stage entails deleting commonly used stop words such as prepositions, unnecessary words, special characters, ASCII code, new lines, and excessive white spaces.

Stemming: it entails transforming each token to its stem, or root, form.

3.1.2 Feature Extraction and Selection

It plays vital role to increase the sentiment categorization process accuracy. TF-IDF and BoW are used to obtain its traits. TF-IDF is known technique to deal the text into feature vocabulary. TF-IDFs are calculated using Eqs. 1, 2 and 3 [7, 16].

$$TF(t) = \frac{\text{(Number of times term } t \text{ appears in a document)}}{\text{(Total number of terms in the document)}} \quad (1)$$

$$IDF(t) = \log_e \frac{\text{Total amount documents}}{\text{Number documents with term } t \text{ in it}} \quad (2)$$

To find the TF-IDF score:

$$TF-IDF(w) = TF(w) \times IDF(w) \quad (3)$$

By calculating the frequency of the entire document, BoW translates text input into numeric numbers. By ignoring word order and focusing on word frequency, it generates feature vocabulary across all pages. Selecting the appropriate feature from the feature lexicon is an important task [13]. A typical statistical test for detecting the relation between a term and the linked class is the Chi square (χ^2) statistic. If there is no association between the feature set and class, then it is said to be a null hypothesis. The χ^2 value is calculated by using Eq. (4) [6, 17].

$$\chi^2 = \frac{\sum(\text{observed value} - \text{Expected value})^2}{\text{Expected value}} \quad (4)$$

3.2 Classification

3.2.1 Multinomial Naive Bayes (MNB)

It is a popular method for classifying documents based on statistical analysis of their contents. To classify the documents by assessing the probability that a document belongs to the same class as on the same topic. Vectors $\theta_y = (\theta_{y1}, \dots, \theta_{ym})$ are distribution parameters, and θ_{yi} is the probability $P(x_i|y)$ of feature i appearing on the same class y [14]. The parameter θ_y is estimated by Eq. (5),

$$\hat{\theta}_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n} \quad (5)$$

where $N_i = \sum_{\{x \in T\}} xi$ is the number of times feature i appears in a class y in the training set T , and $N_y = \sum_{i=1}^n N_{yi}$ is total count of all features for class y and $\alpha \geq 0$.

3.2.2 Support Vector Machine (SVM)

SVM converts the data points into a higher-dimensional space, allowing them to be separated linearly. By determining the optimum hyperplane for dividing the group of data. The main aim is to shorten the distance amongst all data group and the hyperplane. The hyperplanes that should be used are specified by kernel functions. The linear kernel is used if the data can be linearly separated. The Radial Basis Function (RBF) kernel is used for non-linear data. The training data is labelled with (x_i, y_i) , $i = 1, 2, 3$. where $x_i \in R^n$ and $y \in \{1, -1\}^l$. The SVM optimization problem can be solved in the following Eq. 6:

$$\min_{w, b, \zeta} \frac{1}{2} W^T W + C \sum_{i=1}^n \zeta_i \text{ subject to } y_i (w^T \phi(x_i) + b) \geq 1 - \zeta_i, \\ \zeta_i \geq 0, i = 1, 2, 3, \dots, n \quad (6)$$

3.2.3 Logistic Regression (LR)

A supervised classifier, the logistic regression model is a rule set deal with multiclass problems. The logistic function determines the relation between two class labels. If the likelihood is greater than 0.5, label “1” is assigned; otherwise, label “0” is assigned [16, 17]. It operates by reducing the loss function by determining the optimal set of weight parameters.

3.2.4 Decision Tree (DT)

The Decision Tree is a binary tree with conditions at the core and class labels at the decedent nodes. Attribute selection method uses information gain or Gini index method to determine the value of each attribute. The highest information gain of attribute is chosen since it gives the most information as a split node [18, 19]. This process is repeated until the last node.

3.3 Bagging Ensemble Techniques

It is a bootstrap ensemble that uses re-placement to construct subsets of data from the original data. On each data subset, base classifiers are trained, and the individual predictions are concatenated to get the final prediction [12]. By training many weak classifiers on subsets of the original data, bagging enhances the classifier performance.

4 Experiments

The various classification algorithms is examined in this section. The experiments employed the Multidomain Product Review Dataset [15]. In the Review Dataset, TF-IDF and BoW models are used for feature extraction, and Chi square feature selection models are used to choose top features. Logistic Regression (LR), Support Vector Machines with RBF, Linear Kernels and Grid Search, Decision Tree, Multinomial Naive Bayes (MNB) and Random Forest (RF) are used to train selected features. These classifiers are subsequently trained using bagging techniques.

As performance measures for the classifiers mentioned above, accuracy and F-score are used. Equation 7 depicts the accuracy metric, which is defined as the ratio of correctly predicted numbers to total predicted numbers. The model is perfect if the F-score number is 1. The F-score is calculated using the Eqs. 8, 9 and 10.

$$\text{Accuracy} = \frac{\text{Number of Correctly Predicted}}{\text{Total Number of Predicted}} \quad (7)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

where TP is True Positive and FP is False Positive.

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

where TP is True Positive, and FN is False Negative.

$$F1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (10)$$

5 Result and Discussion

The experiment results depict how each classifier responds to feature selection. The proposed technique's efficacy is shown in Tables 2 and 3. Tables 4 and 5 represent the F1-score of the classifiers.

From the above tables, it shows that LR technique provides better accuracy than other base classifiers. In Bagging, LR classifiers trained on BoW with Chi square feature selection greatly outperformed the other basic classifiers, as shown in Table 3. On multidomain datasets, the classifiers SVM (RBF) and DT achieved the same results for TF-IDF feature extraction and Chi square feature selection regardless of any classification methods.

Table 2 Base classifiers accuracy

Feature selection	Classifiers				
	LR	SVM	SVM (RBF)	DT	MNB
TF-IDF	76.5	75.5	51.37	68.25	76.87
TF-IDF + CHI	80.06	80.08	51.37	68.25	79.81
BoW	77.93	75.5	51.37	68.93	77.12
BoW + CHI	80.06	77.18	51.43	68.93	78.87

Table 3 Ensemble classifiers accuracy

Feature Selection	Bagging + Classifiers					
	LR	SVM	SVM (RBF)	DT	MNB	KNN
TF-IDF	79	78.31	51.37	75.25	75.93	51.56
TF-IDF + CHI	79	79.03	51.37	75.25	75.93	51.56
BoW	79.17	77.06	51.37	73.37	76.31	55.68
BoW + CHI	80.93	80.06	51.43	73.37	78.18	60.12

Table 4 F1-score of base classifier

Feature selection	Classifiers					
	LR	SVM	SVM (RBF)	DT	MNB	KNN
TF-IDF	76.80	75.77	67.28	68.25	76.84	66.77
TF-IDF + CHI	76.80	75.77	67.28	68.25	76.84	66.77
BoW	78.06	75.65	67.28	69.82	77.29	67.68
BoW + CHI	80.17	77.39	67.31	69.82	79.05	67.46

Table 5 F1-score of ensemble classifiers

Feature selection	Bagging + Classifiers					
	LR	SVM	SVM (RBF)	DT	MNB	KNN
TF-IDF	79.02	75.77	67.28	68.25	76.84	67.11
TF-IDF + CHI	79.02	75.77	67.28	68.25	76.84	67.11
BoW	78.06	77.33	67.28	73.89	76.38	65.83
BoW + CHI	80.17	80.12	67.31	73.92	78.62	68.38

6 Conclusion

Sentiment analysis on a multidomain dataset is performed using basic classifiers and ensemble classifiers. According to the experiment results, the LR algorithm outperformed the TF-IDF and BoW models, along with the Chi square feature selection technique, the accuracy of weak classifiers is also improved by using ensemble classifiers. The performance of neural network methods on multidomain datasets will be studied in the future. Other ensemble methods will be investigated in the future, and its performance will also be compared.

References

1. Anjali Devi, S., Sapkota, P., Rohit Kumar, K., Pooja, S., & Sandeep, M. S. (2020). Comparison of classification algorithms on twitter data using sentiment analysis. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(5), 8170–8173.
2. Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of 42th Annual Meeting on Association for Computational Linguistics (ACL)* (pp. 271–278).
3. Bollegala, D., Weir, D., & Carroll, J. (2013). Cross-domain sentiment classification using a sentiment sensitive thesaurus. *IEEE Transactions on Knowledge and Data Engineering*, 25(8), 1719–1731. ISSN 1041-4347.
4. Pang, b., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of EMNLP*.
5. Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th International Conference on World Wide Web* (pp. 342–351).
6. Bulusu, A., & Sucharita, V. (2019). Research on machine learning techniques for POS tagging in NLP. *International Journal of Recent Technology and Engineering*, 8(1, Special Issue- 4).
7. Manning, C. D., *Introduction to Information Retrieval by Christ Manning*.
8. Ghosh, M., & Sanyal, G. (2018). An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning. *Journal of Big Data* 5, 44. <https://doi.org/10.1186/s40537-018-0152-5>
9. Kotturu, P., Sasank, V. V. S., Supriya, G., Manoj, C. S., & Maheshwarredy, M. V. (2019). Prediction of chronic kidney disease using machine learning techniques. *International Journal of Advanced Science and Technology*, 28(16).
10. Chakraborty, K., Bhattacharyya, S., & Bag, R. (2020). A survey of sentiment analysis from social media data. *IEEE Transactions on Computational Social Systems*, 7(2).

11. Li, S. S., Huang, C. R., & Zong, C. Q. (2011). Multi-domain sentiment classification with classifier combination. *Journal of Computer Science and Technology*, 26, 25–33. <https://doi.org/10.1007/s11390-011-9412-y>
12. Madasu, A., & Elango, S. (2020). Efficient feature selection techniques for sentiment analysis. *Multimedia Tools Applications*, 79, 6313–6335. <https://doi.org/10.1007/s11042-019-08409-z>
13. Magoni, M., & Petrucci, G. (2017). A fuzzy-based strategy for multi-domain sentiment analysis. *International Journal of Approximate Reasoning*, 93, 59–73.
14. Nafis, N. S. M., & Awang, S. (2021). Enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. *IEEE Access*, 9, 52171–52192.
15. Ni, J., Li, J., & McAuley, J. (2019). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 188–197).
16. Hou, Q., Han, M., & Cai, Z. (2020). Survey on data analysis in social media: A practical application aspect. *Big Data Mining and Analytics*, 3(4)
17. Sajana, T., & Narasingarao, M. R. (2018). Classification of imbalanced malaria disease using Naïve Bayesian Algorithm. *International Journal of Engineering & Technology*, 7(2.7), 786–790.
18. Sanjay Bhargav, P., Nagarjuna Reddy, G., Ravi Chand, R. V., Pujitha, K., & Mathur, A. (2019). Sentiment analysis for hotel rating using machine learning algorithms. *International Journal of Innovative Technology and Exploring Engineering*, 8(6), 1225–1228.
19. Srinivas, P. V. V. S., Pavan Sai Sujith, L. V. N., Sarvani, P.M., Kumar, D. S., & Parasa, D. (2019). Prediction of hospital admission using machine learning. *International Journal of Scientific and Technology Research*, 8(12).
20. Jha, V., Savitha, R., Shenoy, P. D., Venugopal, K. R., & Sangaiah, A. K. (2018). A novel sentiment aware dictionary for multi-domain sentiment classification. *Computers & Electrical Engineering*, 69, 585–597.
21. Wu, F., Huang, Y., & Yuan, Z. (2017). Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources. *Information Fusion*, 35, 26–37. ISSN 1566–2535. <https://doi.org/10.1016/j.inffus.2016.09.001>