International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Aspect Category Extraction for Sentiment Analysis using Multivariate Filter Method of Feature Selection

Bhavana R. Bhamare, P. Jeyanthi, R. Subhashini

Abstract: ***Aspect-oriented sentiment analysis is done in two phases like aspect term identification from review and determining related opinion. To carry out this analysis, features play an important role to determine the accuracy of the model. Feature extraction and feature selection techniques contribute to increase the classification accuracy. Feature selection strategies reduce computation time, improve prediction performance, and provides a higher understanding of the information in machine learning and pattern recognition applications etc. This work specifically focuses on aspect extraction from restaurant review dataset but can also be used for other datasets. In this system, we proposed a multivariate filter strategy of feature selection which works on lemma features. This method helps to select relevant features and avoid redundant ones. Initially, the extracted features undergo preprocessing and then the "term-frequency matrix" is generated which contains the occurrence count of features with respect to aspect category. In the next phase, different feature selection strategies are applied which includes selecting features based on correlation, weighted term frequency and weighted term frequency with the correlation coefficient. The performance of weighted term frequency with correlation coefficient approach is compared with the existing system and shows significant improvement in F1 score.***

Keywords: ***Aspect-Based Sentiment Analysis (ABSA), Natural Language Processing(NLP), Machine Learning (ML), feature selection, correlation coefficient, Term Frequency-Inverse Document Frequency (TF-IDF).***

I. INTRODUCTION

Due to the quick expansion of the social networking sites, people post their opinions freely. The growth of internet technologies led to increase in online shopping and posting reviews about the products. This helps customers to compare multiple products and gives them further options to choose from. It is a difficult task to analyze products by overall comparison and hence the need to compare products. Comparison can be done on the basis of aspects. ABSA has become a research interest and a challenging task for the researchers. ABSA includes different subtasks namely aspect term identification, opinion target extraction and corresponding sentiment determination. The sentiment classification is done at three levels like aspect level, sentence level and document level.

Following is the example from a restaurant review dataset. Restaurant reviews can have major aspect categories as price, ambiance, food, service, etc. So instead of determining overall review sentiment, it is useful to extract the aspect from review and then determine sentiment for that aspect. In the following example, sentence 1 denotes food aspect category and sentence 2 shows price and food aspect categories.

- "The food was great."
- "The food was pricey and not too tasty."

The aspect categories may be explicit or implicit. In sentence 2, price aspect is explicit but the food aspect is implicit.

The focus of this work is to extract aspect categories from review sentences. Hence, this is a text categorization problem. This system is trained and tested using SemEval 2014 restaurant review dataset. The reviews in the given dataset had 5 aspect categories like food, ambiance, price, service and miscellaneous. When enough review data is available and aspect categories are defined, then supervised algorithms can be used to forecast the aspect categories. The accuracy of the supervised algorithms is reliant on the quality of the features extracted and selected. We proposed a multivariate filter method of feature assortment to reduce the dimensionality of feature space.

Existing feature selection methods are classified to major classes like wrapper, filter, and hybrid. In the wrapper approach, at first different feature subsets are selected and then the feature sets are evaluated using the selected classifier. In a filter-based approach, the selection of features is not reliant on any machine learning algorithm. In this, features are preferred on the basis of their numerical weight [9], [12]. The hybrid approach is the union of the above two approaches. Filter method is further divided into two parts namely univariate and multivariate approach. In univariate filter method, features are evaluated with respect to relevance and in the multivariate approach, the correlation between features is calculated and redundant features are avoided.

Manuscript published on 30 September 2019
* Correspondence Author
Bhavana R. Bhamare*, Research Scholar, Department of Computer Science & Engineering, Sathyabama Institute of Science & Technology, Chennai, India. Email: brkanawade@gmail.com
P. Jeyanthi, Department of Information Technology, Sathyabama Institute of Science & Technology, Chennai, India.
Email: jeyanthi.it@sathyabama.ac.in
R. Subhashini, Department of Information Technology, Sathyabama Institute of Science & Technology, Chennai, India.
Email: subhaagopi@gmail.com

We are proposing a multivariate filter approach which selects relevant features and avoids redundant ones. The paper catalogueis as below. Section II is related work, the proposed system is described in section III, section IV shows the results of experimentation and section V contains the conclusion followed by the future scope.

II. LITERATURE SURVEY

A. Related work

Sentiment analysis (SA) is a vastly used term to classify user's opinion using NLP and ML Approaches. Various researchers have used multiple methods for aspect based classification, polarity based classification [1], [3], [8], etc. Product review based sentiment analysis is similar to the proposed sentiment analysis approach. Fig. 1 summaries the basic model of the sentiment classification task.

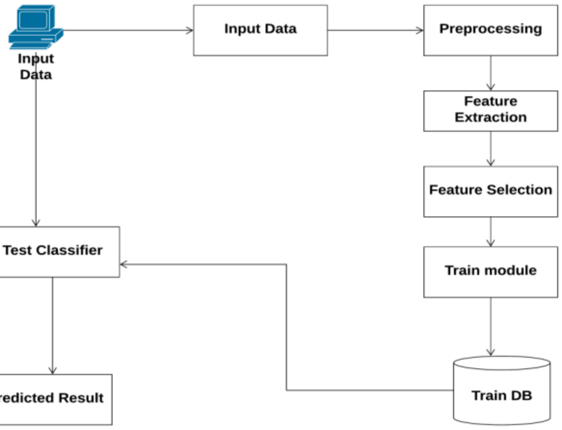

Fig. 1: Supervised learning approach for SA: a basic model view

The supervised approach of classification requires a training set and a test set for classification. The classifier is trained on a pre-labeled training set. Then a test set is utilized to verify the model by deriving the class labels of unknown features. Some feature components that are used for feature categorization are unigrams, bigrams NLP based [1], [7] and ontology-based features [13]. Now days, many systems use word dependency-based and ontology-based features [16] to train the classifier. These feature components can be utilized to discover the semantically related words, expressions& sentences. A variety of machine learning algorithms like Simple Naïve Bayes (NB), Random Forest and Support Vector Machines (SVM) [6], [16] are used for ABSA of reviews. The NB produces good results if word features are to be considered independent or unrelated to one another. The model depends on Bayesian calculation. Here, lemma features are extracted, and for feature selection different strategies are applied, like feature selection based on term frequency, weighted term frequency, term frequency with correlation, and weighted term frequency with correlation. In this approach, relevant features are selected, redundancy is avoided and unique features are procured. By using these weighted features, a training model is derived. For test sentences, the probability of every aspect class is calculated and the aspect class with the highest probability is the actual aspect label.

B. Existing Methodologies

Mahdieh Labani et. al. [2] projected a feature selection strategy based on multivariate filter method that is used for different text classification approaches. This method emphasizes on selecting relevant features and minimizing redundant features to improve the result of classification. The proposed methodology takes into consideration, document frequencies for each term while estimating their usefulness. It selects the features with maximum relevancy and the redundancy between them is taken into consideration using a correlation factor. Comparative analysis with other filter methods show that the results are improved using this approach.

Asriyanti Indah Pratiwi and Adiwijaya [4] proposed feature selection and classification based on information gain for document sentiment analysis. Information Gain Classifier (IGC) is utilized to extract the different features from movie review dataset. IG-DF-FS proposed by authors, based on hybrid method is called a combination of Information Gain + Document Frequency Feature Selection etc.

Haoyue Liu et. al. [5] projected a system of feature selection for imbalanced data. Imbalanced dataset has a problem of bias-to-majority. This problem is addressed using Weighted Gini Index (WGI) approach. In WGI approach, the impurity reduction score is calculated for each feature & the higher scoring features are assorted.

Asha S Manek et. al. [6] proposed a strategy for aspect term mining for SA of movie review dataset. The work is accomplished using the Gini Index (GI) algorithm for feature selection after NLP processing. It uses SVM classifier to classify the test data. This study describes a statistical approach for calculation of weight by GI method which is further used for feature selection in SA. This framework for SA which utilizes SVM classifier is compared with various other methods of feature selection on movie reviews. Using this approach there is significant improvement in accuracy.

Muhammad Zubair Asghar et. al. [7] proposed a heuristic pattern based framework for aspect-based opinion mining. Authors advised a framework containing modules for aspect extraction, summary generation and hybrid sentiment classifier dealing with intensifiers and negations. These modules are implemented using features based on heuristic patterns extracted from POS tagged sentences. The system obtained improved classification results compared to other systems with a precision of 0.85. This generalized methodology can classify aspect-based opinions on various domains. Rule based approach for feature extraction is also utilized in [10], [14]. The focus of [14] was to extract aspects from product review dataset.

Kim Schouten et. al. [8] proposed a system aspect category detection for SA using supervised as well as unsupervised algorithms. In this paper, the first method provided is an unsupervised method that applies association rule mining on co-occurrence frequency data procured from a corpus to search aspect categories. The second, supervised, method calculates the co-occurrence between aspect categories and both lemma features and dependency-based features.

At the time of testing, if the maximum conditional probability is greater than the associated trained threshold, then the corresponding aspect category is assigned to that sentence.

The accuracy of the system is around 83% for a supervised method.

Laith Mohammad Abualigah et. al. [11] proposed a feature selection methodology which is a union of Genetic operators (GA) and particle swarm optimization algorithm for text clustering. The hybrid approach improved the accuracy of text clustering. The GA is used to solve the unsupervised feature selection problem, called Feature Selection based Genetic Algorithm for Text Classification (FSGATC). In order to obtain more accurate clusters on multiple review text datasets, creation of a new subset of informative features is done using FSGATC. The text clustering results based on various common benchmark datasets, were significantly improved when using FSGATC in comparison with other methods.

V. K. Singh et. al. [15] proposed a method for SA of movie review dataset. In this work, a domain-specific feature-based heuristic is described for aspect-level sentiment analysis of movie reviews. An aspect-oriented system is devised that analyses the textual reviews of a movie and assigns it a sentiment label on each aspect. Document-level classification done by authors, used different linguistic feature combinations such as Adverb + Adjective to Adverb + Adjective + Verb etc. This System used SentiWordNet dictionary to compute the sentiment class label for document-level sentiment analysis as well as for ABSA.

Basant Agarwal et. al. [13] proposed a system named concept-level sentiment analysis with Dependency-Based Semantic Parsing. This system illustrates a fundamental problem of the SA task and uses concepts as features. To extract semantic features, a concept extraction algorithm on the basis of a novel concept parser scheme is provided that exploits the semantic relationship between words in natural language text. The system further provides a feature to extract the actual concept using ConceptNet ontology like RDF framework. The Minimum Redundancy and Maximum Relevance feature selection technique helps in the selection of important concepts and elimination of redundant concepts. The selected concepts are then used to build a machine learning model that classifies a given document as positive or negative

III. RESEARCH METHODOLOGY

The proposed system follows the following steps; preprocessing, feature extraction, feature selection, generation of training matrix and evaluation of system performance using test dataset in a sequential manner. Fig. 2 visualizes the proposed system architecture with overall execution.

A. Preprocessing and Features Extraction

Each sentence in the training dataset undergoes preprocessing like tokenization, transformation from uppercase to lowercase, stop word filtering and stemming. The standard stop word dictionary is used that is available on https://gist.github.com/larsyencken/1440509. Stemming as well as Lemmatization, these are the two crucial feature normalization methods that are used in the preprocessing stage. The restoration of all the affected words in the text into root form, defined as stem words, is done using stemming method. For example; ‘studying,’ ‘studies,’ are each converted into the stem ‘study’ and ‘studi’ respectively. Principally, in lemmatization, the conversion of all the forms of words to their basic lemma takes place. For example, the terms “studying”, “studies” will be converted to the lemma “study”. So the accuracy of lemma features is considered to be more than that of stemmed features. In this experimentation, lemmas are extracted as features and then they undergo feature selection methods.

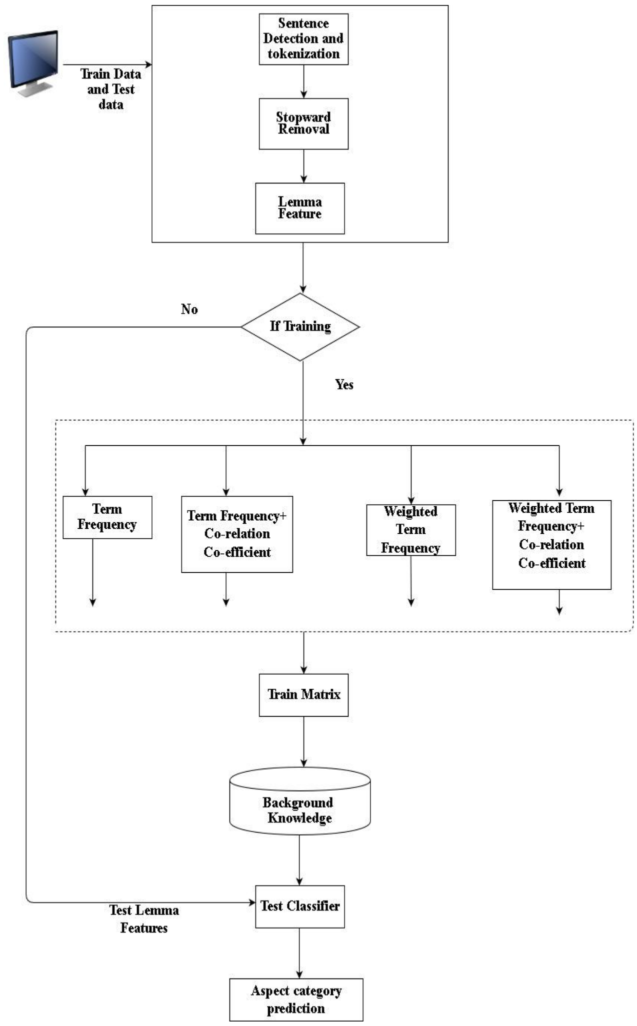

Fig. 2: Supervised system architecture for aspect category extraction

B. Feature Selection

In the system, various feature selection approaches are analyzed and hybrid approach for feature selection is proposed. The feature selection strategies analyzed are:

i. Term frequency (TF)

In this approach, feature selection is done on the basis of term frequency count. Term frequency of each feature with respect to each aspect category is calculated. A threshold is set for feature selection. Features having term frequency greater than ‘2’ are selected in each aspect category. The result of this is, for each aspect category a respective term frequency matrix is generated.

Further, a compound matrix is generated containing terms with their occurrence count in all aspect categories. From this matrix, a binary train matrix is generated where ‘1’ is considered for non-zero term frequency.

ii. Weighted Term Frequency (WTF)

In this approach, the weight of each term is calculated using equation (1). The weight of a term is the conditional probability where $X_{t,k}$ is the occurrence count of term t in aspect category k and X_t is the total occurrence count of a term t in all aspect categories. If the proportion of occurrence of a term t in aspect category k is more with respect to the other aspect categories, then the weight of t increases. A threshold is set on weight for each aspect category. Terms (features) having a weight greater than threshold are selected to generate a binary train matrix. Weight calculation of term is also done in Kim Schouten et. al. [8]. This work follows a similar approach as [8] and proposes a hybrid approach for feature selection using correlation to avoid redundancy in features. In [8], weights are used to determine aspect category of test sentence and in this approach, weight is used for feature selection and for generation of a binary train matrix.

$$\text{weight}(t) = \frac{X_{t,k}}{X_t} \qquad (1)$$

iii. Term Frequency with Correlation Coefficient (TF+CC)

In classification, features must be relevant but not redundant to increase the accuracy of the classifier. In this strategy, the term frequency matrix obtained in (i) is used. Features obtained using this matrix are relevant but also redundant. So to avoid redundancy, correlation of each feature is calculated with respect to other features in that aspect category. Pearson correlation coefficient is used to calculate correlation.

$$COweight\,[t_i] = \frac{n\,(\sum X[]Y[]) - (\sum X[]) * (\sum Y[])}{\sqrt{[n \sum X^2 - (\sum X)^2]}\sqrt{[n \sum Y^2 - (\sum Y)^2]}} \qquad (2)$$

Equation (2) is used to calculate the correlation of each term with respect to other terms where $x[]$ and $y[]$ are vectors of term t_i and t_{i+1} respectively, containing term frequency with respect to each aspect category. Correlation of a term t with respect to other terms in that aspect category is averaged. Terms having correlation value less than or equal to 0.85 are selected to generate a binary train matrix.

iv. Weighted Term Frequency with Correlation Coefficient (WTF+CC)

In this approach, weighted matrix obtained in (ii) is used to generate a new matrix which contains the weight of a term with respect to each aspect category. Equation (2) is used to calculate the correlation of each term with respect to other terms where $x[]$ and $y[]$ are vectors of term t_i and t_{i+1} respectively, containing the weight of term $t$in each aspect category. Finally, a binary train matrix is generated as mentioned in (iii).

Contribution of this paper is to propose a supervised approach for aspect category extraction which selects relevant features and avoids redundancy by calculating correlation among features. Acquired results show that weighted term frequency with correlation approach has comparatively more F-score. In this work, it is observed that features selected using weighted term frequency are more relevant but also redundant. Redundancy among features in an aspect category is avoided by calculating the correlation. In Algorithms 1, detailed steps of the proposed approach (iii) are given.

Algorithm 1: Feature selection using Weighted Term Frequency (WTF) + Co-relation Coefficient (CC) method.

Input: $TF[i]_k$ is the set of selected features containing term frequency of term t_i in aspect category k, Threshold on term frequency $Th_f = 3$, Threshold on correlation Th_c.
Output: A binary train matrix.

Step 1: for each feature/term f in TF
Frequenct(f) ← $TF[f]_k$
if (Frequenct(f) ≥ Th_f
weightedList.add(f)
end for

Step 2: for each term t in weightedList calculate weight

$$\text{Weight}(t) = \frac{X_{t,k}}{X_t}$$

if weight(t) >Threshold_k
add t in weight[t][w], where t is the term and w
is its weight in corresponding aspect category
k_w.
end for

Here, threshold on weight is different for each aspect category.

Step 3: A matrix X[i][j] is generated, where i = {t_1, . . ,t_n}
and j={k_1, . . ,k_5} for 5 aspect categories. Each row in matrix is weight of X[t] in k_1 . . k_5 aspect categories.

Step 4: Calculate correlation of each t with all terms in X
as:
for each t_i in X, calculate correlation with t_{i+1}..$t_n$in aspect
category k

$$COweight\,[t] = \frac{n\,(\sum X[]Y[]) - (\sum X[]) * (\sum Y[])}{\sqrt{[n \sum X^2 - (\sum X)^2]}\,\sqrt{[n \sum Y^2 - (\sum Y)^2}}$$

end for
Calculate average of correlation for each term.

Step 5: To avoid redundant features, select features having
averaged correlation less than equal to 0.85 from each aspect category.
if (COweight[term] ≤ Th_c)
FinalFeatureset← {Term, Aspect}

Step 6: Generate a matrix Mat[i][j] where Mat[i] is a vector of term t_i containing correlation value for each aspect category j where j={k_1..k_5}.

Step 7: Generate a binary train matrix MatB[i][j].

```
for each term Mat[t_i] in Mat[i][j]
  for j from k_1 to k_5

    if Mat[t_i][j] >0.0 then
      MatB[t_i][j]=1
    else
      MatB[t_i][j]=0
    end if
  end for
end for
```

Retrieval Number: C4566098319/19©BEIESP
DOI:10.35940/ijrte.C4566.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Algorithm 2: Classification Algorithm

Preprocessing: Input test sentences are preprocessed and lemma features are extracted.

Input: Binary train matrixMatB[i][j], Lemma[i]={lemmas for each sentence i in test dataset}.

Output: Aspect label prediction for each test sentence.
Step 1:
for each sentence i from test dataset
  for each lemma L in lemma [i]
    for each term t_i in MatB
      if lemma L in lemma[i] = = term in MatB
        test[L][1..5]=MatB[t][1..5]
        copy binary vector from MatB to test vector
        for lemma L.
      end if
    end for
  end for
end for
Step 2: Calculate conditional probability of each aspect category.
Step 3: Return aspect category corresponding to highest probability score.

IV. RESULTS AND DISCUSSIONS

The proposed system is implemented in java with windows environment, some inbuilt functions are used during the feature selection as well as extraction. Experimental analysis is explained below.

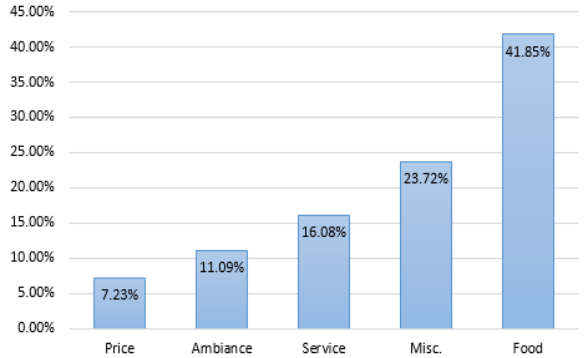

Fig. 3: Aspect category distribution in test dataset

Fig. 3 shows the percentage distribution of the number of sentences in each aspect category in test data. For this experimentation, SemEval 2014 dataset is used which contains3000training instances and 800 test instances. According to the above% distribution, "food" aspect category has the maximum instances being exactly at 41.85% while the "price" is having the lowest of them all i. e. 7.23%.

Fig. 4 shows comparative F-score using different approaches of feature selection like term frequency (TF), term frequency with correlation coefficient (TF+CC), weighted term frequency (WTF) and weighted term frequency with correlation coefficient approach (WTF+CC). Results show that there is a steady increment in F-score using TF+CC for all aspect categories than using TF only. F-score obtained using WTF is better compared to TF+CC approach, for aspect categories such as food, service, and ambiance. F-score results obtained using WTF+CC approach are better than WTF in all aspect categories except service. This experiment proves two things; first that uneven distribution of aspect categories in dataset impacts the results, and second, using correlation coefficient, redundancy in features is removed and helps to increase the F-score.

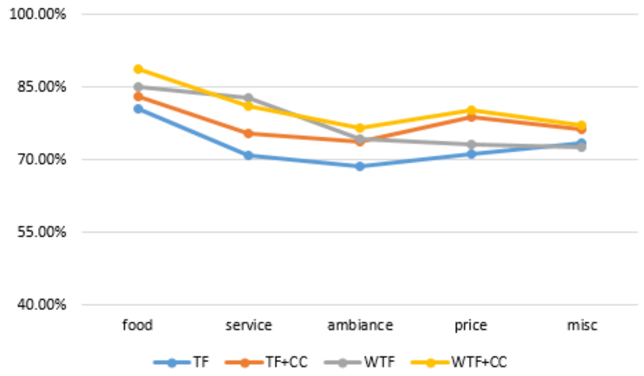

Fig. 4: F-score prediction using different approaches

Fig. 5 demonstrates the comparative analysis of the proposed system with the existing approach [8]. The existing system has worked on different feature sets like only lemmas, the only dependency-based features, and both lemmas with dependency-based features. Here, the results of lemma-based implementations (existing systems) are compared with the proposed system. Authors of [8] have calculated the weight of each selected term and that matrix is used to determine the aspect category of test sentence. In the proposed approach, weight is used for feature selection. Figure 5 shows, comparative F-score obtained for different aspect categories using the proposed approach WTF+CC and existing approach. The proposed system displays comparatively increased F-score for all aspect categories compared to existing unsupervised approach. An increase is noticed in the F-score for food, ambiance and miscellaneous aspect categories as compared to existing supervised approach.

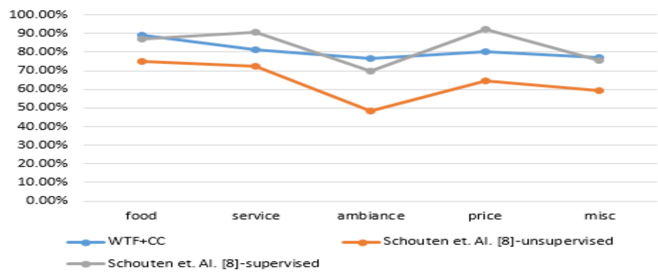

Fig. 5: F-Score for each aspect category with proposed vs existing approach

V. CONCLUSION

The proposed ABSA approach illustrates various feature selection approaches. The purpose of this experimentation is to prove that the results of classification can be improved using the relevant feature set.

Results prove that not only relevant feature sets are important but also redundancy among features should be reduced. The proposed system calculates the correlation of each feature with respect to other features in the corresponding aspect category. Features having average correlation less than or equal to 0.85 are selected which further reduces redundancy. Among the various methods used for feature selection, WTF+CC has obtained comparatively better results than TF, WTF, TF+CC and the existing approaches.

VI. FUTURE SCOPE

Proposed work uses a supervised approach in which a binary train matrix is used to determine an aspect category of the test sentence. The aspect category yielding higher probability score is selected as an aspect label for test sentence. This work can be executed on different feature sets using machine learning algorithms. Till date, very few aspect-labeled datasets exist in the English language, so generation of more aspect-labeled datasets for reviews can also be considered a significant contribution to this field.

REFERENCES

1. Shafie AS, Sharef NM, Murad MA, Azman AShafie AS, Sharef NM, Murad MA, Azman A. Aspect extraction performance with pos tag pattern of dependency relation in aspect-based sentiment analysis. In2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP) 2018 Mar 26 (pp. 1-6). IEEE.
2. Labani M, Moradi P, Ahmadizar F, Jalili M. A novel multivariate filter method for feature selection in text classification problems. Engineering Applications of Artificial Intelligence. 2018 Apr 1; 70:25-37.
3. Kim K. An improved semi-supervised dimensionality reduction using feature weighting: Application to sentiment analysis. Expert Systems with Applications. 2018 Nov 1; 109:49-65.
4. Pratiwi AI. On the feature selection and classification based on information gain for document sentiment analysis. Applied Computational Intelligence and Soft Computing. 2018;2018.
5. Liu H, Zhou M, Lu XS, Yao C. Weighted Gini index feature selection method for imbalanced data. In2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC) 2018 Mar 27 (pp. 1-6). IEEE.
6. Manek AS, Shenoy PD, Mohan MC, Venugopal KR. Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. World wide web. 2017 Mar 1;20(2):135-54.
7. Asghar MZ, Khan A, Zahra SR, Ahmad S, Kundi FM. Aspect-based opinion mining framework using heuristic patterns. Cluster Computing. 2017:1-9.
8. Schouten K, Van Der Weijde O, Frasincar F, Dekker R. Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. IEEE transactions on cybernetics. 2017 Apr 14;48(4):1263-75.
9. Abualigah LM, KhaderAT, Al-Betar MA, Alomari OA. Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Systems with Applications. 2017 Oct 30; 84:24-36.
10. Kang Y, Zhou L. RubE: Rule-based methods for extracting product features from online consumer reviews. Information & Management. 2017 Mar 1;54(2):166-76.
11. Abualigah LM, KhaderAT, Al-Betar MA. Unsupervised feature selection technique based on genetic algorithm for improving the text clustering. In2016 7th international conference on computer science and information technology (CSIT) 2016 Jul 13 (pp. 1-6). IEEE.
12. Fattah MA. New term weighting schemes with combination of multiple classifiers for sentiment analysis. Neurocomputing. 2015 Nov 1; 167:434-42.
13. Agarwal B, Poria S, Mittal N, Gelbukh A, Hussain A. Concept-level sentiment analysis with dependency-based semantic parsing: a novel approach. Cognitive Computation. 2015 Aug 1;7(4):487-99.
14. Poria S, Cambria E, Ku LW, Gui C, Gelbukh A. Poria S, Cambria E, Ku LW, Gui C, Gelbukh A. A rule-based approach to aspect extraction from product reviews. InProceedings of the second workshop on natural language processing for social media (SocialNLP) 2014 Aug (pp. 28-37).
15. Singh VK, Piryani R, Uddin A, Waila P. Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification. In2013 International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s) 2013 Mar 22 (pp. 712-717). IEEE.
16. Schouten K, Frasincar F, de Jong F. Ontology-enhanced aspect-based sentiment analysis. InInternational Conference on Web Engineering 2017 Jun 5 (pp. 302-320). Springer, Cham.

AUTHORS PROFILE

Ms. Bhavana R. Bhamare, Research Scholar, Department of Computer Science & Engineering, Sathyabama Institute of Science & Technology, Chennai. Her area of research is aspect based sentiment analysis. Her research interests are Machine Learning, Data Mining and Natural Language Processing.

Dr. P. Jeyanthi, Associate Professor, Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai. Her research Interests are: Machine Learning, Data Mining, Image retrieval, Internet of Things, Big Data Analytics. She has published more than 35 research publications in journals.

Dr. R. Subhashini, Head of the Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai. Her research Interests are: Machine Learning, Cloud Computing, Deep Learning, Data Analytics and Big Data. She is the Project coordinator for DST-FIST Cloud Computing lab. She has published more than 40 research publications in journals.