An Improved Probabilistic Method for Product Review Classification

Meenakshi Madan
Research Scholar, Dept. of Computer Science & Engineering, Sat Priya Group of Institutions, Rohtak, Haryana, India.

Rupali Malhotra
HOD, Dept. of Computer Science & Engineering, Sat Priya Group of Institutions, Rohtak, Haryana, India.

Abstract – Opinion mining is having the significance to process the textual data and to identify the hidden sentiments. The product reviews collected by a website can also be processed under this classifier to identify the positive and negative reviews. In this work, a statistical transformation based Bayesian network model is provided for review classification. The work is divided in two main stages. In first stage, the input textual review was processed under series of operations to extract the keywords. Later on the text is converted to the numerical form. This numerical formed dataset is processed under Bayesian network model to perform the classification. The results show that the method has improved the accuracy of sentiment classification.

Index Terms – Sentiment Classification, Opinion Mining, Statistical, Bayesian Network.

1. INTRODUCTION

Sentiments or the expressions are the expressive knowledge that is generated itself when a human presents the knowledge in different form. The facial expression, human gesture, speech utterances are the some of the common features of human knowledge expression. Today the user is very much involved with web using the social media. While posting the blogs or the tweets user also includes his expressions and expressions in the form of text. These expressions or expressions are based on the word or the phrase identification done by the web user. Some of the web service providers or the companies also accept the user message or the review to analyze the popularity of the site, product or the service. Once this kind of textual review is accepted by the service provider, an analysis can be performed under different dimensions based on the type of knowledge incorporated in it. These features and the feature dimensions can be different according to the application areas. Some of the common forms of expressions represented by different researchers along with impact classes are listed in table 1.

Emotion	Impact Class
Excited, Awesome, Elated, Enthusiastic, Strong, Wonderful Wow, Warmhearted	Strongly Positive Affect
Aroused, Astonished Surprised	Strongly Engaged
Content, Happy, Kindly Satisfied, Pleased	Pleasantness
Dull, Drowsy, Sleepy Fine, Better, Ok	Low Positive Affect
at rest, calm, relaxed placid	Low Negative Affect
Quiescent, Quiet Still	Disengagement
Blue, grouchy, lonely Sad, sorry, unhappy not good, bad	Unpleasantness
Worst, scornful, litter, hostile, fearful, distressed	High Negative Affect

Table 1 : Emotion Classes

These emotion or expression classes are used to identify the user interest in the product or the service so that the service or the product provider can take the relative action to increase his sales.

The identification of the emotion class from the review, post or the message is called sentiment analysis. Sentiment analysis is one of the most emerging research areas that include the integrated concepts of text mining and the natural language processing. The sentiments or the expressions are here observed under different features shown in figure 1. The high level classification of the sentiment analysis is considered in two main features called the positive and the negative sentiments. There are number of existing methods to perform the topic oriented classification.

In this paper, a numerical transformed and Bayesian network based method is provided for sentiment classification. In this section, the sentiment processing along with relative featured method is defined. In section II, the work defined by earlier researchers is presented. In section III, the proposed methodology is defined. In section IV, the results related to work are presented. In section V, the conclusion of work is presented.

2. RELATED WORK

Lot of work is already defined by different researchers to process the reviews and to identify the sentiments from it. Different statistical and analytical methods are provided by the researchers under different aspects and with relative measures. In this section, some of the work defined by earlier researchers is provided. Author[1] has provided a study work on aspect based analysis on software project and defined categorized solution. The method is based on the aspect identification and relative concept integration to recognize the hidden sentiment. The concept centric analysis can be defined to generate the promising classification results. Author[2] has defined compression specific sentiment analysis for news processing and generating the aspect driven observation. The special potential feature based semantic analysis is provided for identification of review class. The discriminative conditional analysis with process modeling was provided for class identification. Author[3] has provided a work on stock price prediction using regression modeling and method under the sentiment class analysis. Author integrated the Naïve Bays algorithm along with Random Forest Algorithm to predict the prices. The textual quote process was provided for analysis of the movement in the price. Author[4] has defined the work on social issues so that the verb specific option analysis will be applied. The opinion processing and the relative request processing and performance analysis were provided by the author. Author[5] has defined an improved Meta heuristic feature function to perform the sentiment analysis. Author implied the Genetic Algorithm along with Rough Set Theory for identification of the review analysis. The feature specific analysis and relative observations were taken to remove the irrelevant features and improving the classification process.

Author[6] has defined a clustering based score generation method to improve the sentiment classification. Author utilized the bog of word along with aspect level clustering to generate the new feature set and to provide more accurate clustering and sentiment analysis. The score specific observation to the learned method is provided. Author[7] used the neural network approach for classification of reviews in more effective form. Author defined three pairs of convolution layers for architectural pooling and normalized the input. The work is to improve the accuracy and robustness of the classification. Author[8] has provided a work on weight driven analysis for sentiment analysis and the modeling for message classification. Author used the metric based Bays algorithm for this classification. The work was defined specifically for finance community. Author[9] has provided an entity based analysis for sentiment analysis. The framework is here defined to evaluate the domain specific features to improve the classification results. Author[10] has defined a textual model for review processing under different aspects and modeling. The profile specific analysis and domain specific analysis was provided.

Author[11] has defined the learning methods for sentiment processing categorize the in the relative classes. Author worked on mobile, laptop and other online products. The feature vector based classification was generated under the positive and negative aspects. Author[12] has provided a work on issue processing and relative solution generation for sentiment classification. The analytical modeling was provided to recover the limitation of the processing and generating the capability driven analysis to identify the sentiments over the product review. Author[13] has defined a work on step driven analysis on different subjects and the semantics along with association, polarity and grammatical relationship. The sentiment lexicon and subject observation was also provided. Author[14] has defined a work on interaction chain analysis to generate the clustered topic chain to compare the work under similarity observation and provided the polarity specific analysis. Author provided the heuristic measure map along with micro blogging to identify the hidden sentiments.

3. RESEARCH METHODOLOGY

Opinion Mining is the knowledge extraction process applied on textual reviews to identify the hidden or expressed emotions or sentiments. It a computational form of natural language processing and having scope in all kind of web communication. This communication can be in the form of public reviews, user chat and the twitter posts. The observations can be formed to identify the sentences, phrases and words of these web posts. Some deterministic measures can be applied to take the featured decisions based on intended meaning of words or sentences. In this proposed work, a multi featured probabilistic method is defined to improve the classification method. In the preprocessing stage of review processing, the filtration and stemming will be applied to obtain the effective knowledge from these textual reviews. In second stage, weight specific feature extraction method is applied on positive and negative words separately. Frequency, Entropy and the ratio map is applied to generate the quantified features. These features are finally processed under Bayesian net approach to perform the classification.

Here figure 2 has showed the basic processing model relative to the presented work. The proposed work model has accepted the textual messages as input and then transformed to the numerical form. For this, the keyword extraction, stop word removal and the adjective analysis is provided. Once the filtered keyword list is generated along

With adjectives, the adjective type is identified. The keyword frequency, number of objectives collectively identified the weight of the sentiments. This work is applied on training and testing sets. Once the dataset is transformed, the final work is to apply the bays classifier to perform the classification. The results relative to the work are shown in next section.

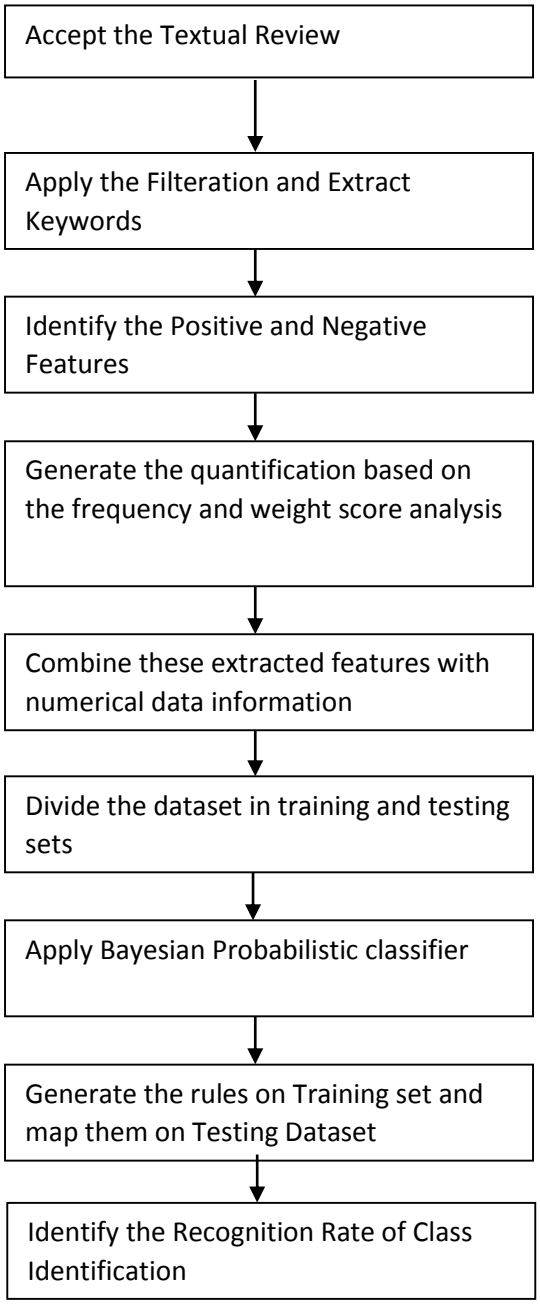

Figure 2: Flow of Work

4. RESULTS

The presented work is implemented in weka integrated java environment. The graphical interface is designed to apply the query and to process the training and testing set. The bays usage is here done using weka integration. The work is applied on three different datasets taken from external web sources. The accuracy results on three different datasets are shown here in figure 2.

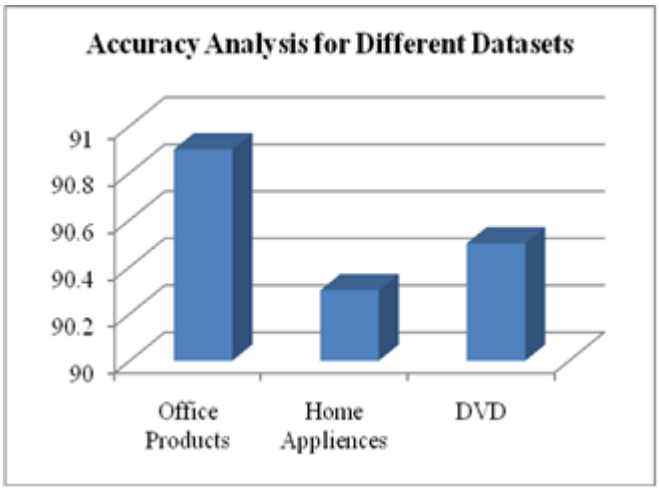

Figure 3 : Accuracy Analysis Results

Here figure 3 is showing the comparative analysis on 3 different data sets in terms of accuracy. The results show that the method has provided over 90% accuracy for each method. The comparative results against some existing methods are shown here in figure 4.

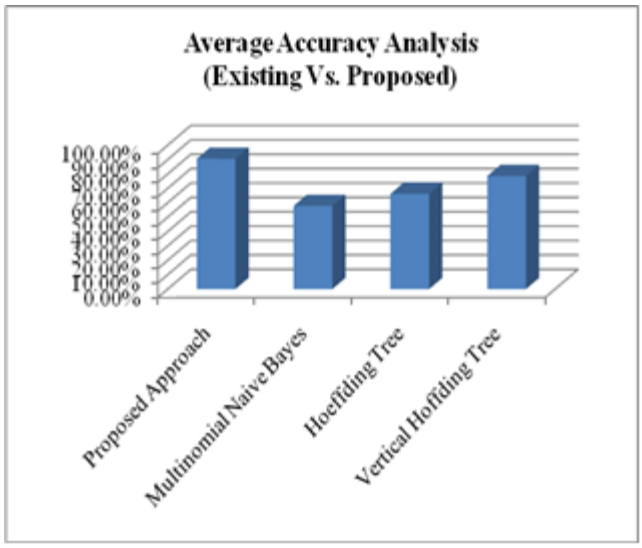

Figure 4 : Comparative Analysis

Here figure 4 is showing the comparative accuracy results of proposed approach against previous method. The results show that the defined method has improved the classification accuracy. The existing method provided the maximum of 79% accuracy whereas the proposed work provided accuracy over 90%.

5. CONCLUSION

In this paper, a method is defined to process the product reviews and convert it to the numerical form. The keyword based analysis and filtration is applied to transform the textual

set to numerical form. Finally, the Bayesian network is applied to perform the product review classification. The comparative results show that the proposed method has improved the accuracy and classification results.

REFERENCES

[1] K. Schouten and F. Frasincar, "Survey on Aspect-Level Sentiment Analysis," in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 3, pp. 813-830, March 1 2016.

[2] W. Che, Y. Zhao, H. Guo, Z. Su and T. Liu, "Sentence Compression for Aspect-Based Sentiment Analysis," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 23, no. 12, pp. 2111-2124, Dec. 2015.

[3] Y. E. Cakra and B. Distiawan Trisedya, "Stock price prediction using linear regression based on sentiment analysis," 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, 2015, pp. 147-154.

[4] M. Karamibekr and A. A. Ghorbani, "Sentiment Analysis of Social Issues," Social Informatics (SocialInformatics), 2012 International Conference on, Lausanne, 2012, pp. 215-221.

[5] S. R. Ahmad, A. A. Bakar and M. R. Yaakub, "Metaheuristic algorithms for feature selection in sentiment analysis," Science and Information Conference (SAI), 2015, London, 2015, pp. 222-226.

[6] M. Farhadloo and E. Rolland, "Multi-Class Sentiment Analysis with Clustering and Score Representation," Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on, Dallas, TX, 2013, pp. 904-912.

[7] X. Ouyang, P. Zhou, C. H. Li and L. Liu, "Sentiment Analysis Using Convolutional Neural Network," Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on, Liverpool, 2015, pp. 2359-2364.

[8] A. B. Eliaçik and E. Erdoğan, "User-weighted sentiment analysis for financial community on Twitter," Innovations in Information Technology (IIT), 2015 11th International Conference on, Dubai, 2015, pp. 46-51.

[9] C. B. Ward, Y. Choi, S. Skiena and E. C. Xavier, "Empath: A framework for evaluating entity-level sentiment analysis," Emerging Technologies for a Smarter World (CEWIT), 2011 8th International Conference & Expo on, New York, NY, 2011, pp. 1-6.

[10] V. K. Singh, R. Piryani, A. Uddin and P. Waila, "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification," Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on, Kottayam, 2013, pp. 712-717.

[11] M. S. Neethu and R. Rajasree, "Sentiment analysis in twitter using machine learning techniques," Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on, Tiruchengode, 2013, pp. 1-5.

[12] Z. Wang, V. Joo, C. Tong and D. Chan, "Issues of Social Data Analytics with a New Method for Sentiment Analysis of Social Media Data," Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on, Singapore, 2014, pp. 899-904.

[13] W. Y. Chong, B. Selvaretnam and L. K. Soon, "Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets," Artificial Intelligence with Applications in Engineering and Technology (ICAIET), 2014 4th International Conference on, Kota Kinabalu, 2014, pp. 212-217.

[14] N. Gu, D. y. Sun, B. Li and Z. Li, "Sentiment Analysis for Topics based on Interaction Chain Model," Intelligence and Security Informatics Conference (EISIC), 2015 European, Manchester, 2015, pp. 133-136.