Asian Journal of Information Technology 15 (5): 887-890, 2016
ISSN: 1682-3915
© Medwell Journals, 2016

Fuzzy Classification Techniques for Effective Sentiment Analysis Using Twitter Data

V. Soundarya and D. Manjula
Department of Computer Science and Engineering, Anna University, CEG Campus, Chennai, India

Abstract: In this study, we propose new techniques for feature selection and sentiment analysis using classification algorithms. For this purpose, we collected tweets from 5000 users for a period of one month. We considered the sentiments such as happy, joy, sadness, anger, fear, surprise, distress and disgust and identified the words used to express these features. Based on synonym analysis, we select features for positive sentiments and negative sentiments by proposing a new feature selection algorithm using keyword frequency and semantic analysis. Moreover, we propose a new classification algorithm based on a new type of support vector machine called Group Support Vector Machines (GSVM) to perform major and sub classification of sentiments and to form groups based on the sentiments of people with respect to change in time and location. Finally, the groups are used to form discussion forums on various topics including business, tour, e-learning, religion and sports. The main advantage of the proposed research is to identify people with similar interest based on the sentiments identified from tweets and to form interest groups for discussion on interesting topics. From the experiments conducted in this research, it is observed that the groups formed by sentiment analysis provided >95% accuracy in identifying members for forming interest groups on twitter and hence, is more accurate than the existing systems.

Key words: Sentiment classification, sentiment analysis, feature selection, group SVM, twitter

INTRODUCTION

Sentiment classification is a recent sub area of natural language processing which is concerned not only with the topic of a document but also about the opinion it expresses. Therefore, sentiment analysis is becoming popular in many applications including social networks, business reviews, medical reports and group discussions which predict the sentiments such as opinion and emotion from text documents available in social networks particularly twitter. Sentiment classification can be performed either in word/phrase level or in sentence and document level. Moreover, sentiment analysis using social networks analysis has now become an important approach for extracting sentiments on tweets. A sub area in sentiment classification is subjectivity analysis in which the language is divided into sentiments namely negative, positive and neutral. Positive sentiments are expressed using words such as happy, joyful, merry, likes, loves, enjoys and laughs. Similarly, negative sentiments are identified using the words sad, sorry, regret, cry, pain and hate. There are some words which are neither positive nor negative with respect to sentiment classification. For example, walk, meet, go, swim and call. All these sentiments can be identified using classification, frequency counting and semantic analysis (Shelke *et al*., 2012; Liu, 2012).

In business information retrieval systems, opinion mining is useful to know the feedback from customers. If the sentiments of customers are positive then the customers like the product. Similarly, if the sentiments are identified as negative then the customers are not interested in the product. If the customer has neither positive nor negative sentiment then they will have neutral sentiment. Therefore, it is necessary to identify the features for positive, negative and neutral sentiments. By applying these features, the opinions expressed by customers through social networks are used as feedback and the new customers are classified based on the classification of old customers. In such a scenario, the combination of syntax analysis, semantic analysis, feature selection and classification can be used to make effective decisions about the customer interests on products.

Corresponding Author: V. Soundarya, Department of Computer Science and Engineering, Anna University, CEG Campus, Chennai, India

In this research, we propose a new sentiment classification system which uses natural language processing techniques, frequency count of words and Group Support Vector Machine (GSVM) classification algorithm for classifying the customers based on their opinions and sentiments. This is helpful to find the most promising customers, most useful areas and relevant timings for carrying out the business activities. Moreover, the proposed research focuses on classifying the customers based on their interest and opinion in the usage of cosmetic items namely soap and powder. The main advantage of the proposed research is that it helps to provide focused advertisements on selected people alone.

Literature review: There are many works have been done in this direction by various researchers in the past (Pang and Lee (2008). Esuli and Sebastiani (2006a), Neviarouskaya *et al.* (2009) and Li *et al.* (2010). Among them, Ortigosa *et al.* (2014) proposed a new machine learning algorithm which is used for sentiment analysis that involves training of the classifier on benchmark datasets and also the use of the trained model for new sentences classification in documents. Pang and Lee (2008) provided a survey of related research on sentiment analysis and opinion mining. Moreover, a dictionary based sentiment analysis has been discussed (Balahur *et al.*, 2012; Lloret *et al.*, 2012). Ganapathy *et al.* (2013) proposed a new feature selection algorithm along with a classifier for effective classification of intrusion dataset. An opinion summary is generated based on opinion sentences by considering frequent features are explored by Balahur *et al.* (2012). The researcher also uses query based information retrieval techniques for analyzing the reviews.

Esuli and Sebastiani (2006b) formalizes that sentiment is an affective part of opinion or simply used as synonyms for each other without any true definition of their own. Various studies on sentiment classification have been conducted and evolved at different levels word, sentence level and document level Li *et al.* (2010). Devitt and Ahmed (2007) research work compares the performance of stemming and non-stemming algorithms and a general normalization work is done using information retrieval techniques. Ganapathy *et al.* (2013) discusses about the various feature selection and classification algorithms.

Trilla and Alias (2012) also analyzed the machine learning method for sentiment analysis which involves the trained model for new document classification. To evaluate a possible improvement to search methods Lindholm (2011) extracts the news studies and describes the implementation techniques which is generic and easily adopted to new data sources. Wiebe and Riloff (2011) identifies that information extraction techniques may be used to learn informative clues of subjectivity. Ohana and Tierne (2009) proposed a technique of sentiment classification by using features built from the SentiWordNet database of term polarity scores. Their approach consisted of counting positive and negative term scores to determine sentiment orientation.

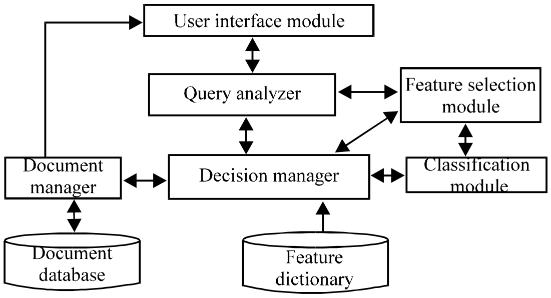

Fig. 1: System Architecture

System architecture: Figure 1 shows the architecture of the system proposed in this study. It consists of eight components namely user interface, query and document analyzer, decision manager, feature selection module, classification module, document manager, document database and lookup database. The user can interact with the system through the user interface.

The query given by the user is given to the query analyzer. The query analyzer searches the twitter for user discussions based on the query. The discussions are analyzed and they are stored in the document database. The decision manager is responsible for controlling the entire system. The feature selection module is responsible for forming features. The classification module classifies the documents and submits them to the decision manager. The decision manager uses the classified results and lookup table to make decisions on formation of groups. After forming groups the user interface is informed by the decision manager through document manager. The document manager is responsible for managing the discussions made by group members and their profiles.

MATERIALS AND METHODS

Proposed work: In this research, a new sentiment classification system is proposed which uses group support vector machine for effective classification. For this purpose, a new feature selection method based on syntax analysis, frequency count and semantic analysis is proposed. The steps of the algorithm are as follows:

- Input: Tweets from twitter
- Output: Sentiment features and classified documents
- Step 1: Read one document from twitter
- Step 2: Apply the rules for regular expressions to check words
- Step 3: Perform stemming
- Step 4: Apply parts of speech tagging
- Step 5: Select features based on dictionary and documents for negative, neutral and positive sentiments
- Step 6: Perform classification of the current document into one of the groups with negative, neutral and positive sentiments
- Step 7: Read query from users
- Step 8: Select features
- Step 9: Perform classification and identify the sentiments
- Step 10: Put the users in proper group.

RESULTS AND DISCUSSION

We have implemented this paper using JAVA programming language. The experiments in this research have been carried out using real dataset collected from twitter related to products based on discussions by users.

Figure 2 shows the sentiment analysis made from five documents obtained from twitter. The positive, negative and neutral sentiments are considered for analysis. From Fig. 2, it is observed that document1 and 2 have more positive tweets, documents 3 and 4 have more negative tweets and document 5 has more neutral tweets. Therefore, the users belonging to documents 1, 2 and 5 can be focused to perform advertisement on the given product.

Figure 3 shows the accuracy analysis for five documents obtained from twitter. In this experiment, classification was carried out with and without feature selection. Here, we have used SVM for classification. From Fig. 3, it is observed that feature selection helps to improve the accuarcy by >4%. This is due to the fact that the uses of feature selection method.

Figure 4 shows the classification accuracy analysis between three classifiers namely neural classifier, SVM and GSVM. From Fig. 4, it is observed that GSVM provides more accuracy than other two classifiers. This is due to the fact that GSVM makes group discussion before making decision.

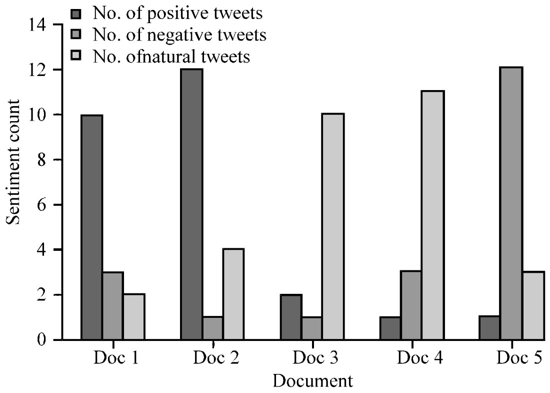

Fig. 2: Sentiment analysis using twitter

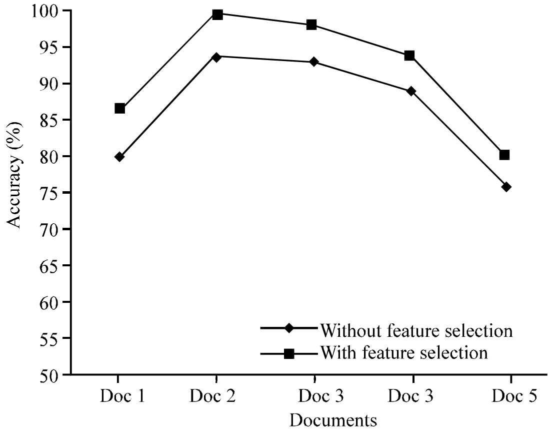

Fig. 3: Accuracy analysis

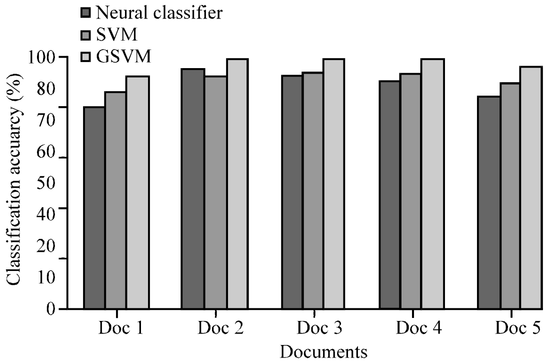

Fig. 4: Classification accuracy analysis

CONCLUSION

In this study, a new sentiment classification algorithm is proposed and implemented for analyzing the documents obtained from twitter. Here, the sentiments considered are negative sentiments, positive sentiments and neutral sentiments. From the experiments carried out in this research, it is observed that feature selection helps to improve the classification accuracy. Moreover, the use of newly proposed classifier called GSVM provides more accuracy than the existing classifiers in sentiment classification.

RECOMMENDATIONS

Future research in this direction can be the use of fuzzy logic to classify low, medium and high sentiments in each of the categories considered in this research.

REFERENCES

Balahur, A., M. Kabadjov, J. Steinberger, R. Steinberger and A. Montoyo, 2012. Challenges and solutions in the opinion summarization of user-generated content. J. Intell. Inf. Syst., 39: 375-398.

Devitt, A. and K. Ahmad, 2007. Sentiment polarity identification in financial news: A cohesion-based approach. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, June 23-30, 2007, Prague, Czech Republic, pp: 984-991.

Esuli, A. and F. Sebastiani, 2006a. Determining term subjectivity and term orientation for opinion mining. Proceedings of the 11th Meeting on European Chapter of the Association for Computational Linguistics, April 3-7, 2006b. Trento, pp: 193-200.

Esuli, A. and F. Sebastiani, 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. Proceedings of the 5th International Conference on Language Resources and Evaluation, May 22-28, 2006, Genoa, Italy, pp: 417-422.

Ganapathy, S., K. Kulothungan, S. Muthurajkumar, M. Vijayalakshmi, P. Yogesh and A. Kannan, 2013. Intelligent feature selection and classification techniques for intrusion detection in networks: A survey. EURASIP J. Wireless Commun. Network. 10.1186/1687-1499-2013-271.

Li, B., L. Zhou, S. Feng and K.F. Wong, 2010. A unified graph model for sentence-based opinion retrieval. Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Association for Computational Linguistics, Stroudsburg, PA, USA, pp: 1367-1375.

Lindholm, S., 2011. Extracting content from online news sites. Master's Thesis, Computing Science, UMEA University, Sweden.

Liu, B., 2012. Sentiment Analysis and Opinion Mining. Morgan and Claypool Publishers, Florida, ISBN: 9781608458844, Pages: 167.

Lloret, E., A. Balahur, J.M. Gomez, A. Montoyo and M. Palomar, 2012. Towards a unified framework for opinion retrieval, mining and summarization. J. Intell. Inf. Syst., 39: 711-747.

Neviarouskaya, A., H. Prendinger and M. Ishizuka, 2009. Sentiful: Generating a reliable lexicon for sentiment analysis. Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, September 10-12, 2009, Amsterdam, pp: 1-6.

Ohana, B. and B. Tierne, 2009. Sentiment classification of reviews using SentiWordNet. Proceedings of the IT&T 9th International Conference on Information Technology and Telecommunication, October 22-23, 2009, Dublin, Ireland -.

Ortigosa, A., J.M. Martin and R.M. Carro, 2014. Sentiment analysis in Facebook and its application to e-learning. Comput. Human Behav., 31: 527-541.

Pang, B. and L. Lee, 2008. Opinion mining and sentiment analysis. Found. Trends Inform. Retrieval, 2: 1-135.

Shelke, N.M., S. Deshpande and V. Thakre, 2012. Survey of techniques for opinion mining. Int. J. Comput. Applic., 57: 30-35.

Trilla, A. and F. Alias, 2012. Sentence-based sentiment analysis for expressive text-to-speech. IEEE Trans. Audio Speech Language Process., 21: 223-233.

Wiebe, J. and E. Riloff, 2011. Finding mutual benefit between subjectivity analysis and information extraction. Trans. Affect. Comput., 2: 175-191.