

Neuro-Fuzzy Classification Techniques for Sentiment Analysis using Intelligent Agents on Twitter Data

V. Soundarya and D. Manjula

Department of Computer Science and Engineering,
College of Engineering Guindy, Anna University, Chennai, India

Copyright © 2016 ISSR Journals. This is an open access article distributed under the *Creative Commons Attribution License*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: In this paper, we propose a new classification algorithm called Intelligent Agent and Neuro-Fuzzy Rule based Group Support Vector Machines (IGSVM) to perform major classification of sentiments and to form groups based on the sentiments of people with respect to change in time and place. Finally, the groups are used to form discussion forums on various topics including business, e-learning, tour and sports. The main advantage of the proposed work is to identify the user interest based on the sentiments identified from tweets and to form similar interest users groups for discussion on specific topics. From the experiments conducted in this work, it is proved that the user groups formed by sentiment analysis provided more than 94% accuracy in identifying members for forming interest groups on twitter and hence is more accurate than the existing systems.

KEYWORDS: Sentiment classification, sentiment analysis, feature selection, Intelligent Group SVM, twitter.

1 INTRODUCTION

Recently, sentiment analysis is playing major role to predict the user opinion / interest on social networks in many applications such as business reviews, medical reports and group discussions. Sentiment classification is also major and latest sub area of natural language processing which is not only considers the topic of a document but also expresses about the users opinions. Sentiment classification is performed either in word or sentence and document levels. Moreover, social network analysis is becoming important method for extracting sentiments on tweets in sentiment analysis process. In this sentiment classification process, it is divided into sentiments such as positive, negative and neutral. Positive sentiments are indicates using words namely joyful, happy, likes, loves and laughs. Negative sentiments are using the words sorry, sad, regret, cry, pain and hate. All these sentiments words are identified using classification, frequency and semantic analysis [1] [2].

In business information retrieval systems, opinion mining is useful to understand the comments from users. If the sentiments of customers are positive then, the customers just like the product. Further, if the sentiments are diagnosed as bad then the clients are not inquisitive about the product. If the consumer has neither high-quality nor negative sentiment then they may have neutral sentiment. Therefore, it is miles important to identify the functions for nice, poor and impartial sentiments. Through applying those capabilities, the reviews expressed by way of clients through social networks are used as remarks and the new customers are labeled based totally on the category of vintage clients. In this kind of scenario, the aggregate of syntax analysis, semantic analysis, characteristic selection and classification may be used to make effective choices about the consumer interests on merchandise [2].

In this paper, we propose a new sentiment classification technique which makes use of natural language processing techniques, frequency of phrases and Intelligent Agent and Neuro-Fuzzy Rule based Group Support Vector Machines (IGSVM) classification algorithm for classifying the customers based totally on their reviews and sentiments. This is useful to discover the most promising customers, maximum useful regions and relevant timings for wearing out the business activities. Moreover, the proposed paintings focuses on classifying the customers based on their interest and opinion in the utilization

of cosmetic items namely soap and powder. The major advantage of the proposed works is that it allows providing centered commercials on decided on humans by myself. The remainder of this paper is organized as follows: section 2 provides a detailed survey of related work. Section 3 depicts the structure of the proposed system architecture. Section 4 offers the information about the proposed feature selection and classification algorithms. Section 5 shows the results obtained from this paper with suitable discussions. Section 6 gives conclusion on this work and shows a few suitable future works.

2 RELATED WORKS

There are many works have been done by various researchers in the past in this direction of classification, sentiment analysis and users behavior prediction analysis. Among them, Ortigosa et al [3] proposed a new machine learning technique that is used for sentiment evaluation that involves schooling of the classifier on benchmark datasets and also the usage of the trained model for brand new sentences type in files. Pang and Lee et al [4] supplied a survey of associated work on sentiment evaluation and opinion mining. An Opinion summary is generated based totally on opinion sentences via thinking about common features are explored through Balahur et al [5]. The writer also uses query based records retrieval techniques for studying the evaluations.

Esuli et al [6] formalizes that sentiment is an affective part of opinion, or truly used as synonyms for every other without any actual definition of their own. Devitt et al [7] studies work compares the performance of stemming and non-stemming algorithms and a well-known normalization work is finished using information retrieval strategies. Various researches on sentiment classification had been performed and advanced at special tiers word, sentence level and file stage [8]. Ganapathy et al [9] proposed a brand new feature choice algorithm along with a classifier for effective category of intrusion dataset.

Trilla et al [10] additionally analyzed the device learning approach for sentiment analysis which includes the educated model for new record category to evaluate a likely development to go looking methods Lindholm et al [11], extracts the news articles and describes the implementation techniques that's time-honored and effortlessly followed to new information assets. Wiebe et al [12] identifies that statistics extraction strategies may be used to analyze informative clues of subjectivity. Ohana et al [13] proposed a way of sentiment type by means of the use of features built from the SentiWordNet database of time period polarity scores. Their method consisted of counting high quality and poor term scores to determine sentiment orientation.

3 SYSTEM ARCHITECTURE

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

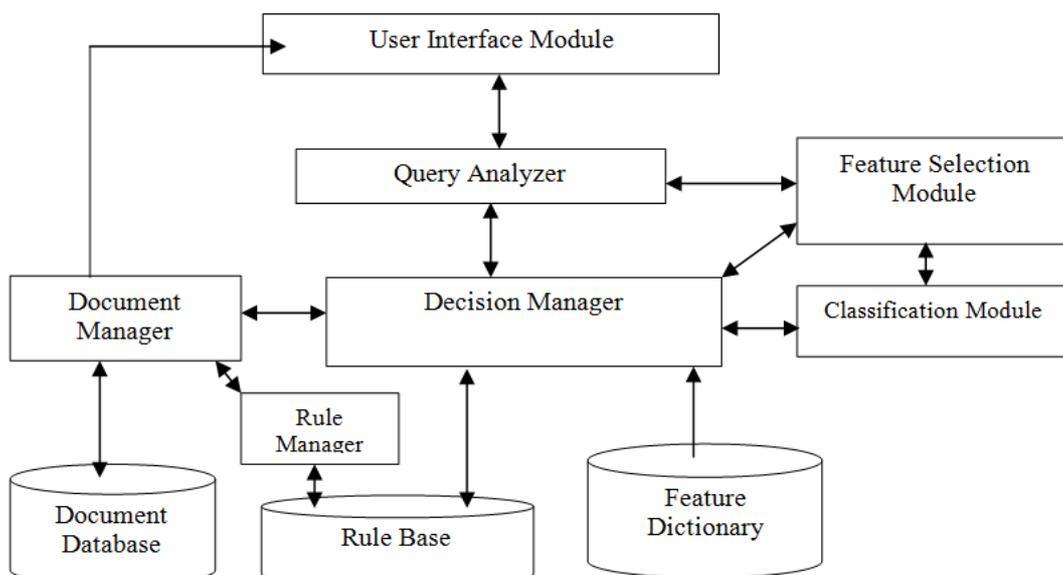


Figure 1 System Architecture

The query given through the consumer is given to the question analyzer. The question analyzer searches the twitter for user discussions based totally at the question. The discussions are analyzed and they're saved inside the file database. The selection supervisor is chargeable for controlling the whole system. The feature selection module is liable for forming features. The classification module is performed the classification of the files and also forwarded them to the decision manager. The decision manager makes use of the categorized consequences and research desk to make decisions on formation of groups. After forming groups the user interface is informed through the decision manager via document manager. The document manager is responsible for handling the discussions made by way of group participants and their profiles. The document manager handles the documents based on the rule manager suggestions. Rule manager handling the rules which are stored in rule base. The decision manager makes decisions based on the neuro-fuzzy rules which are stored in rule base. The rule base contains number of neuro-fuzzy rules which are helpful to make decisions on documents.

4 PROPOSED WORK

A new intelligent agent and neuro-fuzzy rule based sentiment classification system is proposed by using group support vector machine algorithm for effective classification. For this purpose, a new intelligent feature selection algorithm based on syntax analysis, frequency count and semantic analysis is proposed. The steps of the proposed algorithm are as follows:

Intelligent Agent and Neuro-Fuzzy Rule based Group Support Vector Machine Algorithm

Input: Tweets from twitter

Output: Sentiment capabilities and categorized documents

- Step 1: Read one file from twitter
- Step 2: Intelligent agent apply the rules for regular expressions to check phrases
- Step 3: Agent carry out stemming process.
- Step 4: Agent applies the suitable parts of Speech tagging.
- Step 5: Agent Select features based totally on dictionary and documents for terrible, neutral and advantageous sentiments.
- Step 6: Intelligent agent carry out the classification process by applying neuro-fuzzy rules on present day record into one of the organizations with negative, impartial and fantastic sentiments.
- Step 7: Read query from users by agent.
- Step 8: Pick suitable features by intelligent agent.
- Step 9: Agent carry out the category and identify the suitable sentiments with the help of knowledge base.
- Step 10: Intelligent agent decides the proper place for the particular place and place it.

5 RESULTS AND DISCUSSION

The proposed system is implemented using JAVA programming language. The experimental results of the proposed work had been completed the usage of standard dataset accumulated from twitter related to merchandise primarily based on discussions by means of users.

Table 1 shows the sentiment analysis using twitter. Determine 2 suggest the sentiment analysis made from 5 files obtained from twitter. The positive, negative and neutral sentiments are taken into consideration for analysis.

Table 1 Sentiment Analysis using Twitter

Documents	Sentiment Count		
	No. of Positive Tweets	No. of Negative Tweets	No. of Neutral Tweets
Doc1	9	3	2
Doc2	11	1	5
Doc3	4	1	9
Doc4	3	3	12
Doc5	2	9	4

The table 1 shows that doc1 and doc2 has more positive counts than other three documents doc3, doc4 and doc5. Based on the tweet counts, we classify the users interest is on doc1, doc2 and also neutral with doc5.

The accuracy analysis of the five different twitter documents is shown in Figure 2. Experiments are carried out for document classification with feature selection using GSVM classification. Results show that the accuracy of classification with feature selection is better than classification without feature selection. It also shows that the classification accuracy is more than 5% accuracy increased with feature selection when it is compared with without feature selection.

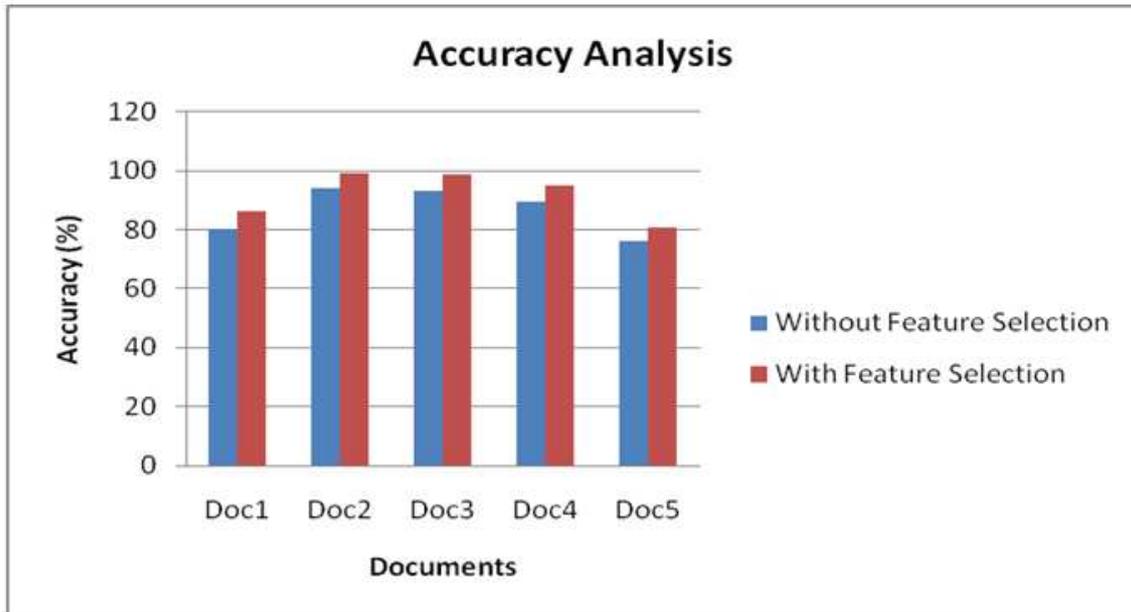


Figure 2 Accuracy analysis

And also, the experiments are carried out to show the classification accuracy of classifiers SVM, IGSMV and Naïve Bayes classifier which is shown in Table 2.

Table 2 Classification Accuracy Analyses

Documents	Naïve Bayes	SVM	IGSMV
Doc1	81.62	86.45	93.54
Doc2	92.12	90.28	99.23
Doc3	90.60	92.54	99.59
Doc4	90.13	91.33	97.28
Doc5	85.21	90.36	95.17

From table 2, it is observed that IGSMV provides more accuracy than other two classifiers. This is due to the fact that GSVM makes group discussion and applies neuro-fuzzy rules before making decision.

6 CONCLUSION AND FUTURE WORKS

In this paper, an intelligent agent based Neuro-Fuzzy rule based Group Support Vector Machine (IGSMV) classification algorithm is proposed and implemented for analyzing the documents received from twitter. Here, the sentiments considered are negative sentiments, effective sentiments and neutral sentiments. From the experiments achieved in this work, it's far located that function choice facilitates to enhance the class accuracy. Furthermore, the proposed IGSMV provides more accuracy than the present classifiers in sentiment classification. Future works on this route may be the use of rough set theory to select most suitable sentiment for making suitable decisions.

REFERENCES

- [1] Nilesh M. Shelke, Shriniwas Deshpande, Vilas Takre, "Survey of techniques for opinion mining," *International Journal of Computer Applications*, 57,13, pp 0975-8887, 2012.
- [2] Bing Liu. *Sentiment Analysis and Opinion Mining*, Morgan & Claypool Publishers, May 2012.
- [3] Alvaro Ortigosa, Jose M. Martín, Rosa M. Carro, "Sentiment Analysis in Face book and its application to e-learning," *Computers in Human Behavior Journal Elsevier* 2013.
- [4] Bo Pang, Lillian Lee, "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval* Vol. 2, Nos. 1-2, 2008.
- [5] Alexandra Balahur, Mijail Kabadjov, Josef Steinberger, Ralf Steinberger, Andres Montoyo, "Challenges and solutions in the opinion summarization", *Journal of Intelligent Information Systems*, Springer, pp.375-398, 2012.
- [6] Esuli A, Sebastiani F, "Determining term subjectivity and term orientation for opinion mining," *Proceedings the 11th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pp. 193-200, 2006.
- [7] Devitt A, Ahmad K, "Sentiment Polarity Identification in Financial News: A Cohesion-based Approach," *In Proceedings of the Association for Computational Linguistics (ACL)*, pp. 984-991, 2007.
- [8] Li B, Zhou L, Feng S, Wong K, "A Unified Graph Model for Sentence-based Opinion Retrieval," *In Proceedings of ACL 2010*, pp. 1367-1375, 2010.
- [9] Sannasi Ganapathy, Kanagasabai Kulothungan, Sannasy Muthurajkumar, Muthuswamy Vijayalakshmi, Yogesh Palanichamy, Arputharaj, Kannan, "Intelligent feature selection and classification techniques for intrusion detection in networks: a survey. *EURASIP Journal on Wireless Communication. and Networking*, Vol. 271, pp. 1-16, 2013.
- [10] Alexandra Trilla, Francesc Alias "Sentence-Based Sentiment Analysis for Expressive Text-to-Speech", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 2, pp.223-233, February 2013,.
- [11] Sigrid Lindholm, "Extracting Content from Online News Sites," *Master's Thesis in Computing Science*, UMEA University, Sweden 2011.
- [12] Janyce Wiebe and Ellen Riloff, "Finding Mutual Benefit between Subjectivity Analysis and Information Extraction," *IEEE Transactions on Affective Computing*, Vol.2, No.4, pp. 45-56, 2011.
- [13] Bruno Ohana and Brendan Tierne, "Sentiment Classification of reviews using SentiWordNet," *In the Proceeding of IT&T Conference*, Dublin Institute of Technology, Dublin, Ireland, 22-23 October, 2009.