Opinion Analysis For Educational Field

Sanjay Singh Bhadoria

Abstract: Opinion mining is an important area of research in the recent years which combines web mining with computational intelligence to collect opinions through websites, social media, company data analysis and customers. Opinion mining algorithms collect opinions from websites and classify them using the mining process such as Support Vector Machines (SVM), Neural Networks, Decision Tree, Naïve Bayes and other classifiers. Moreover, opinion mining is useful in business since it highlights the positive or negative attitude of their students as well as the products and services. This helps the business managers to improve their method of services and to modify the products which will suit the student interests. Sentiment analysis is a type of opinion mining technique which uses natural language processing and other computational intelligence techniques to make effective decisions.

Index Terms: Support Vector Machines (SVM), Neural Networks, Decision Tree, Naïve Bayes, Sentiment analysis

1. INTRODUCTION

The recent developments in the areas of Internet and e-commerce have become responsible for using social networks to provide reviews for students. Recently, people form online groups discuss not only about the positive qualities of a service but also about complaints. If the negative feedbacks are not addressed properly by the service providers, their negative feedback will increase and the popularity of the product/service will come down. Online feedback using computers and internet are also possible with the advent of online discussion forums. Many techniques are available in the literature, based on visualization tools for obtaining student feedback. An analysis of student feedback can be used not only to recommend the student services but also helps to improve the quality of student services. An analysis based on previous data needs techniques from Artificial Intelligence (AI), datamining, soft computing and statistics. Hence, intelligent systems which can analyse the comments from web are necessary for effective opinion analysis prediction. Moreover, questionnaires, user feedbacks and feedbacks from newspapers and general public are also important to enhance the business intelligence in decision making. In recent years, technological improvements in computing have led to the development of sophisticated decision support systems in educational field to provide support to the students who are using web for fetching educational services. In the past, certain researchers grouped service/product reviews into positive and negative slots, which were used to make decisions to select suitable products and services for customers and to provide guidelines to the business personalities. A recommendation system which addresses all these issues can provide students better options in their choice of educational services. Sentiments for a student feedback system can be either positive sentiment which may be use keywords such as happy, sweet, nice, good and tasty or negative sentiments which may use words with the keywords like not and worst. Sentiment analysis performs decision making using sentiments such as emotions and finds the words indicating sentiments. Sentiment analysis can also be used to determine the attitude of students through online feedbacks given by them on products, students services and on educational institutes.

2 LITERATURE REVIEW

Sentiment analysis and opinion mining (Hsinchun Chen & David Zimbra 2010) are subareas of machine learning. They are techniques used to determine attitudes, emotions and opinions expressed by individuals on particular topics of interest. To identify relevant information and extract subjective information from source documents, sentiment analysis techniques based on opinion mining can use natural language processing (NLP). In recent years, computer users have begun using social media for communication. Therefore, user interest in topics can be identified by carrying out sentiment analysis. Document level sentiment analysis (Danushka Bollegala et al. 2013) techniques described whether the whole document defines a positive or negative opinion. The main purpose of document-level sentiment analysis is to classify a document to identify its overall subjectivity and/or sentiments expressed in it. In such a case, it is possible to apply both supervised andunsupervised learning methods to perform a document-level analysis. From the literature review performed in this work, it is seen that the major advantage of performing opinion mining using sentiment analysis. Fine-grained analysis (Niphat Claypo & Saichon Jaiyen 2014) is carried out in aspect-level sentiment analysis, where opinions about entities and aspects present in the said entities are produced. Hsinchun Chen & David Zimbra 2010, Rui Xia et al. 2015 based on the conditional probability used in the Bayesian theorem, with similar assumptions on the variables usedas features. The advantage of using the Naive Bayes classifier is that it only takes a little training time and provides the required accuracy based on the rules formed. Moreover, it is highly scalable and requires only simple parameters. The only limitation is that it needs a large dataset to perform well, since the accuracy in this model is high only when large datasets with more features are used and, consequently, is not suited to small datasets with minimal features. For classification and regression analysis (Hsinchun Chen & David Zimbra 2010), support vector machines are used to perform pattern recognition on datasets. The SVM algorithm works by developing a model assigning the variables to suitable categories between two types of categories. Moreover, it performs a linear classification and provides a two wayclassification. Compared with existing classification algorithms for binary classification, the SVM algorithm provides enhanced classification accuracy and hence can be used as an effective algorithm for sentiment classification (Yingcai Wu et al. 2010). In many applications, however, it is possible to apply the maximum entropy (Hsinchun Chen & David Zimbra 2010) classifier since it does not consider relationships among features and executes classification

- Sanjay Singh Bhadoria is currently pursuing PhD degree program in Computer Application Development from Dr. APJ Abdul Kalam University, Indore, India, PH +91 993941402, E-mail: sanjay.bhadoria@gmail.com
- Dr. Dhanraj Verma, Computer Application Development from Dr. APJ Abdul Kalam University, Indore, India, PH +91 9755301831, E-mail: Dhanrajimtech@gmail.com
independent of features. The advantage of the maximum entropy method is that it suits the text mining domain very well, and is particularly suited to sentiment analysis from documents. With respect to text mining, the application of the maximum entropy algorithm has provided better results, on many occasions, than other classification algorithms such as the Naive Bayes and SVM. Next to supervised is unsupervised which involves rule based method and Lexicon based method.

In natural language processing, semantic analysis plays an important role. The rules applied for semantic analysis (Danushka Bollegala et al. 2013) when natural language processing is used for performing classification of text documents developed from reviews will provide more accurate results due to the use of semantic analysis. Point-wise mutual information (PMI) (Zhongwu Zhai et al. 2012) is a method used in text mining which analyzes statistical dependence among different words in a document. It considers adjective and adverb tags for feature space construction. Considering the hybrid technique it involves, lexicon into SVM, concept level and augmented lexicon method. Turning a lexicon-based application into an SVM-based one is a combined approach that completes learning vectors consisting of words (Ioannis Katakas et al. 2014).

3 OPINION MINING

In recent times, Opinion mining has been playing a major role in predicting user opinions / interests on social networks in a range of applications such as business reviews, medical reports and group discussions. Sentiment classification is a major, as well as the latest subarea of natural language processing, consider the topic of a document but also expresses users’ opinions. Sentiment classification is performed at the word, sentence or document levels. Social network analysis is becoming crucial in extracting sentiments from tweets in the sentiment analysis process. In this sentiment classification process, it is divided into sentiments such as positive, negative and neutral. Positive sentiments are indicated with words like joyful, happy, likes, loves and laughs. Negative sentiments are expressed using words like sorry, sad, regret, cry, pain and hate. All sentiment words are identified using classification, frequency and semantic analysis (Nilesh M Shelke et al. 2012; Bing Liu 2012). In business information retrieval systems, opinion mining is useful to understand comments from users. If customers' sentiments are positive, it is presumed that they just like the product. Likewise, if the sentiments are diagnosed as negative, it is plain they are unlikely to be interested in the product. If consumers express neither high-quality nor negative sentiments, they may be sentiment-neutral. Therefore, it is essential 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 totally labelled based on the category of vintage clients. In a set-up of this sort, the aggregate of syntax analysis, semantic analysis, characteristic selection and classification may be used to make effective choices about consumer interest in merchandise (Bing Liu 2012).

The Domain-based Opinion Word Service System (DOSS) approach is proposed to deal with the problems above. To understand the features of DOSS, it is essential to first define an opinion word as “A word which gives an identity to a domain”. Opinion words are also called opinion features (Zhenhai Zhai et al. 2014) or sometimes referred to as keywords of an entity. For example, consider the sentences “In educare Swiz, the students services are first-rate. Here the domain considered is a education services and the opinion words (keywords) extracted are “service” and “first-rate”, which confer the domain an identity. Opinion word extraction and its associated terms have been studied by researchers working in the realms of data mining, information retrieval and recommendation systems. Research in the DOSS approach faces issues such as:

1. Extracting the domain-based opinion word
2. Computing a weighting factor (rating) for the opinion word-based user expectations, and
3. Drawing a similarity comparison and providing service recommendations.

When considering extraction of opinion word, the process involves pre-processing, extraction and identification of domain dependent and independent opinion word (Zhenhai Zhai et al. 2014). The pre-processing step includes the collection of the dataset from the trip advisor, tagging and chunking. Opinion words are extracted through natural language processing strategies and dependent parsing techniques. Identification of opinion words as domain-dependent and independent is determined by the intrinsic domain based and extrinsic domain-based rating of specific aspects. WordNet is used to classify opinion-based words.

When analysing the computation of weighting factor, it depends on TF-IDF approach (Zhenhai Zhai et al. 2014). Term Frequency and Inverse Document Frequency involve dispersion and deviation techniques. The dispersion value is calculated by determining how significantly the opinion word is mentioned across all documents, and the deviation is calculated by determining how significantly the opinion word is mentioned across a particular document. Due to this dispersion and deviation value the aspect rating is given for each and every extracted aspect. Finally, the similarity between users is computed through a collaborative filtering approach. If a similarity exists, the recommender system provides recommendations to users.

4 CONCLUSIONS

Opinion mining is carried out in this work by combining web mining techniques with computational intelligence for collecting opinions through websites, social media, company data analysis and customers. New opinion mining algorithms are proposed in this research to collect customer opinions from websites and classify those using existing data mining techniques. New algorithms are proposed for sentiment classification and sentiment analysis by extending classification algorithms. Clustering is carried out in this work to effect grouping and group analysis. The K-means clustering algorithm is used as a base for forming clusters. In addition, rules are applied to refine groups based on interests. In this work, a criminal behaviour analysis model has been developed by applying clustering techniques. Sentiment analysis for the proposed system considering words that express positive sentiments. In this work, sentiment analysis has been carried out to provide suitable recommendations for successful decision making, using sentiment-based emotional words as well as words implying sentiments. Moreover, sentiment analysis has been used in this work to determine the attitude of customers through online feedbacks given by them on products & educational institutes.
REFERENCES


