A novel machine learning approach for sentiment analysis

Dakshata Patl, Pranali Mhatre, Manaswi Patil, Kiranjeet Kaur

Abstract— There is tremendous rise in the volume of textual data especially for the unstructured data generated from people who express opinion through various web and social media platform. We have choosen textual data in the form of hotel reviews for sentiment analysis with opinion mining from customer perspective. We are adding a pictures of hotel based on customer reviews .natural language processing and computation at linguistics are been used to automate the classification of sentiments generated from reviews in this paper. we proposed a regional pie chart where we can get reviews of a particular hotel from a particular region. We are using naive bayes algorithms a deep learning algorithm. After analysis of the algorithm we get a pie chart of regional reviews of the hotel our project helps people and travellers to find the best hotels focusing on positive, negative, sarcastic and duplicate reviews. So the main future work of our project is to detect fake reviews, which will have a positive impact for booking of hotels with high rating. And hence the results of hotel reviews will be represented in the form of star rating as well as pie chart.

1 Introduction

The world has experienced a tremendous rise in the volume of textual data especially for the unstructured data generated from people who express opinions through various web and social media platforms for different reasons such as shopping,bookings,and various buying and selling of goods or products. Mountains of these textual data, initially could be equated to garbage which would need to be disposed from time to time. However, with the advancement in storage capacity accompanied by the increasing sophistication in data mining tools, opportunities and challenges have been created for analysing and deriving useful insights from these mountains of data. In our project, we have chosen textual data in the form of hotel reviews for sentiment analysis with opinion mining from customer perspectives and regional reviews by the customer. Sentiment analysis uses the techniques of natural language processing. Mining hotel reviews is desirable to gain deeper knowledge of customer expectations and support effective management of customer relationships. It would enable the hotel managers to have a good understanding of customer needs, discover areas for further improvement and improve service quality. The hotel reviews are provided exclusively by customers who have made reservations at a particular hotel. Customers post feedback about hotels which include hygiene, quality of food, location, customer service quality and hospitality exhibited by hotel staff.

2 Related Work

Sentiment analysis and opinion mining are terms that refer to the field of study that analyses opinions, evaluations, appraisals, attitudes and emotions of people towards entities such as products, services, organizations, individuals, issues, events, topics and their attributes. These terms were used interchangeably to define opinions that entail positive or negative sentiments. Cambria et al. Disputed the interchange of these concepts by classifying opinion mining as polarity detection and sentiment analysis as focusing on emotion recognition. The opinion mining system only needs to understand polarity that can be positive, negative or neutral sentiments depending on the nature of sentences expressed in a review. The process of detecting polarity is strongly linked to analysing sentiments on a particular subject. Most researches on sentiment analysis are focused on descriptive data.

Manke and Shivale explored the significance of social networks as preferred environments for opinion mining and sentiment analysis. They introduced the original method of opinion classification and tested their algorithm on real social network datasets. They concluded from their findings that social networks exhibit properties that make them suitable for opinion mining activities. Comprehensive surveys have been presented on various methods used in opinion mining with limited focus on aspect oriented analysis.

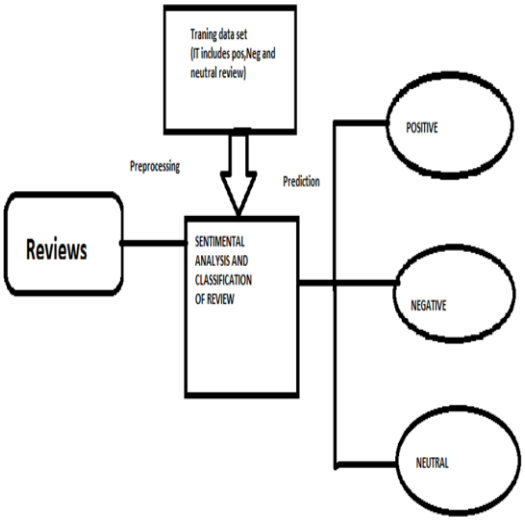
opinion mining using supervised learning algorithms to find the polarity of the student feedback based on pre-defined features of teaching and learning. The study conducted involves the application of a combination of machine learning and natural language processing techniques on student feedback data gathered from module evaluation survey results of Middle East College, Oman. In addition to providing a step by step explanation of the process of implementation of opinion mining from student comments using the open source data analytics tool Rapid Miner, this paper also presents a comparative performance study of the algorithms like SVM, Naïve Bayes, K Nearest Neighbor and Neural Network classifier.

3 Proposed System

The proposed system preprocesses the reviews and then classifies the customer reviews. This extracted information is classified using machine learning algorithms like Naïve Bayes. The reviews are classified into positive, negative, duplicate and sarcastic reviews. The proposed system can overcome all the limitations of the existing system. The system provides accurate results helps to extract information efficiently reducing manual work. The existing system has many more disadvantages and many more difficulties to work well. The proposed system tries to eliminate or reduce these difficulties up to some extent. Using this system we can not only detect positive and negative sentiments but also duplicate and sarcastic sentiments and also neutral reviews. The proposed system helps is user friendly and time efficient.

Whereas, our system is working for sentiment analysis and opinion mining of hotel reviews where the customer reviews are going to be classified, whether the reviews are positive , negative and neutral . Based on this our system gives a regional structure of the hotels which makes new customers to

make choices based on the reviews and ratings



4 Experimental analysis and result

The conceptual view of this system, begins with the feedback collection. Customers respond to questionnaires concerning their feelings about services received from the selected hotels. This can be done in a number of ways, for example opening a web portal through which customers can drop comments. The next step will be to label the comments based on sentiment analysis.
This will be done by human agents who simply read the comments and assign labels based on perceptions. Once data are transformed to a desire format, the next step will be to convert the labelled text to feature vectors through the use of filters. This will make it easier to implement a classification algorithm for training and testing of data. The next step involves the selection of an appropriate classification algorithm while the last step is the training and testing of the selected algorithm on dataset and capturing of results.

5 CONCLUSION

The sentiment analysis with opinion mining framework reported in this paper can be incorporatedinto a hotel technology system that can help improve customer relationship management. From the sentiment polarity exercise that we did, we found out that some comments may be wrongly viewed as neutral while they will be either positive or negative. This system provides a lot of good features for a customer to choose their rooms n hotels.

6 References

1. Feng, S., Xing, L., Gogar, A., & Choi, Y., “Distributional Footprints of Deceptive Product Reviews”, In J. G. Breslin, N. B. Ellison, J. G. Shanahan, & Z. Tufekci (Eds.), ICWSM, 2012.

3. Peter D. Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In Proceedings of the Annual Meeting of the Association of Computational Linguistics (ACL), 2002.

4. Samuel Brody and Nicholas Using Word Lengthening to Detect Sentiment in Microblogs. In Proceedings of Empirical Methods on Natural Language Processing (EMNLP), 2011.

5. Alexander Pak and Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In Proceedings of international conference on Language Resources and Evaluation (LREC), 2010

6. Luciano Barbosa and Junlan Feng. Robust Sentiment Detection on Twitter from Biased and Noisy Data. In Proceedings of the international conference on Computational Linguistics (COLING), 2010.

7. Ankit Pradeep Patel, Ankit Vithalbhai Patel, Sanjaykumar Ghanshyambhai Butani ,Prashant B. Sawant, Literature Survey on Sentiment Analysis of Twitter Data using Machine Learning Approaches International Journal for Innovative Research in Science & Technology,2017