A Sentiment Polarity Categorization Technique for Online Product Reviews

First Author: K Ramesh, Associate professor in Department of CSE, Ramireddy Subbarami Reddy (Rsr) Engineering College, Kadanuthala, Andhra Pradesh.

Second Author: Kotapoori Lilly Grace Mary, Assistant Professor in Department of CSE, Ramireddy Subbarami Reddy (Rsr) Engineering College, Kadanuthala, Andhra Pradesh.

Abstract:

Sentiment analysis is also known as opinion mining which shows the people's opinions and emotions about certain products or services. online data presents useful information to customers for buying a product and for manufacturers to improve the quality of product. When an individual wants to make a decision about buying a product or using a service, they have access to a huge number of user reviews, but reading and analyzing all of them is a tedious task. Reading all of them is generally inefficient.

The main problem in sentiment analysis is the sentiment polarity categorization that determines whether a review is positive, negative or neutral. There is a need for summarization in product reviews. Sentimental analysis helps customer visualize satisfaction while purchasing by simple summarization of these reviews into positive or negative two broader classified classes. The study aims to tackle the problem of sentiment polarity categorization.

1. INTRODUCTION

The importance of users' sentiments has been realized by the business sector in the last decade. Since then, social media platforms and other websites are used to extract users' opinions about products. Such phenomena is called sentiment analysis or opinion mining. Opinion mining is identifying, extracting and understanding the user's attitude or opinion by analyzing the text. This process usually involves natural language processing, statistical analysis and machine learning techniques for sentiment analysis.

In past days, buying of products was more based on getting product review from nearby neighbors, relatives etc. as products were purchased directly from merchants. But with change in technology, development of E-commerce industry with sites flooded by products from different brands made available to customers at the touch of one click.[1] The availability of product based sites with doorway delivery has made it convenient for customers to shop online. With so much change in shopping pattern, merchants providing customers with feedback option about the product. Customers write reviews from all parts of the world. A lot of reviews are very long, making it hard for a potential customer to review them to make an informed decision on whether the customer should purchase the product or not. A huge number of reviews also make it difficult for product manufacturers to keep

log of customer opinions and sentiments expressed on their products and services. It thus becomes necessity to produce a summary of reviews. Summarization of reviews is done using sentiment analysis.[2]

Sentiment analysis tends to extract subjective information required for source materials by applying natural concept of natural language processing. The main task lies in identifying whether the opinion stated is excellent, good or bad. Since customers usually do not express opinions in simple manner, sometimes it becomes tedious task to judge an opinion stated. Some opinions are comparative ones while others are direct.[3] Sentimental analysis helps customer visualize satisfaction while purchasing by simple summarization of these reviews into positive or negative two broader classified classes. Comments are mainly used for helping customers purchase online and for knowing current market trends about products which is helpful for developing market strategies by merchants

2. LITERATURE SURVEY

The techniques available for carrying out sentiment analysis can be classified into three main categories. Such as

(a) Knowledge-based approaches, (b) Statistical based techniques and (c) Hybrid methods, the hybrid techniques or methods are the combination of the previous two approaches. Lexical Knowledge-based approaches normally focus on univocal words like happy, sad, afraid, etc., while statis-tical methods use the automated techniques to judge the sentiments based on machine learning analysis and hybrid approaches use both techniques collectively to analyze the

results on reviews which are not clearly stated but have some link with the product.

Some studies are very much linked with our approach such as Fang and Zhan [9], proposed a process which is used to categorize the polarity based on parts of speech (POS). Another approach presented by Hu and Liu [10] provided a list of different words (i.e. both Positive and Negative words).The proposed list of words consisted of 2006 positive and 4783 negative words respectively. These words are based on online reviews which are used to extract the subjective information for this research. Moreover, in a proposed text categorization technique, Pang and Lee [11], proposed how to remove objective sentences by extracting the subjective ones as mainly we should focus on subjective contents and should not waste time for irrelevant material.

In another technique proposed by Gann and Day [12], the authors applied token based approach on twitter data as they assigned certain sentimental scores to every token which is being used to analyze that if a certain opinion is positive, negative or neutral. Some other techniques are also useful like topic modeling [13] in which the author proposed a process of automatically identifying the features or aspects of a product. Narrowing down the opinion, in the research community several approaches have been proposed on the sentiment analysis of micro blogging sites like twitter. Das and Chen [14], presented an approach for extracting the sentiment from stock message board where the authors suggested that market activities can have an effect on the sentiments of median and small range investors. Another study conducted by Nasukawa and Yi [15], focuses on subject centric aspect of sentiment analysis. The proposed technique

develops a mechanism that determines the polarity scores (i.e. negative and positive sentiment scores) associated with a speci c subject instead of determining or calculating the sentiment for complete documents. The proposed techniques have been evaluated using datasets of different domains, such as news and other web pages. The proposed technique achieves an overall precision score of 75 to 76 percent, depending on the different types of datasets used in the eval-uation. Natural language processing techniques have been employed for carrying out the sentiment analysis. Conven-tionally, sentiment analysis can be performed using three main types of approaches, these approaches are a) Machine learning approach, b) Lexicon based approach and c) Hybrid approach [16].

R. Xia. et al. [17], developed a hybrid technique for sen-timent analysis. The proposed technique combines both lex-icon and machine learning based approaches for sentiment analysis. POS along with their associated and word-related features are selected from lexicon and then machine learn-ing classi ers (i.e. Naiv Bayes, ME and SVM) are applied to determine the sentiment of words. In order to achieve better classi cation results, experiments were performed on the dataset using different combinations, such as xed weighted, meta classi ers and ensemble combination tech-niques. Couple of variations of Naive Bayes were presented by Gamallo et al. [18], various variations of Naive Bayes classi er were applied to classify opinions into different classes. Features like Valence Shifters, Polarity Lexicon, Lemmas and Multiword were used in the experimentation. Nandi and Agrawal [19], presented a layered hybrid tech-nique for sentiment analysis. The proposed approach has two layers; the rst layer is a lexicon based approach while the second layer is machine learning approach. Machine learning classi ers are used to classify the sentiment of opinions into

different classes, such as positive, negative and neutral classes.Rajganesh et al. [20], presented a hybrid approach for sentiment analysis. The approach is a feedback based recommendation system that uses sentiment analysis.

3. PROPOSED WORK
i. DATA PRE PROCESSING

In the pre-processing step, in the first phase the boundary of sentence is to be determined and after verifying the sentence boundary the next phase is to tokenize the text into single words. Pre-processing step also includes removal of stop words, white spaces, new line tags, html tags, emotions and special symbols.

ii. RATING AND CLASSIFICATION

For every review there is always a star rating which is assigned by a user on the basis of his/her experience for a particular product. Thus, Amazon also contains star ratings whenever customer shares opinions. To evaluate 5-starratings of the review, the rest step is to and out the range which is from the highest to the lowest ratings. To calculate these star rates, range from 0 to 1, different researchers contributed

Each review is a variable sequence of words and the sentiment of each review must be classified into above mentioned star rating classes. The Large Amazon Review Dataset contains 308,420 highly-polar reviews (good or bad) for training and testing. The problem is to determine whether a given review has a different sentiment depending on polarity of adverb features. Various methodologies have been practiced by different studies over the years starting from tree based classifier to neural network based approaches. We have chosen Naive Bayes, Decision Tree, Random Forest, Sup-port Vector

Machine, Gradient Boosting and Sequence to Sequence Recurrent Neural Network.

iii. FEATURE EXTRACTION

Features in reviews are extracted so that it helps customer to know which feature has positive comment and which one has negative. Since, overall conclusion about product is much needed but there is also situation where customer requirements come into the scenario. Use of adjectives is done to classify opinions as positive or negative using unigram model. For example, "the Samsung camera I bought was good; it has got great touch screen, awesome flashlight." The feature extracted out of it would be like: Domain: Mobile; Product: Samsung; Feature: Camera; Adjective: Good.
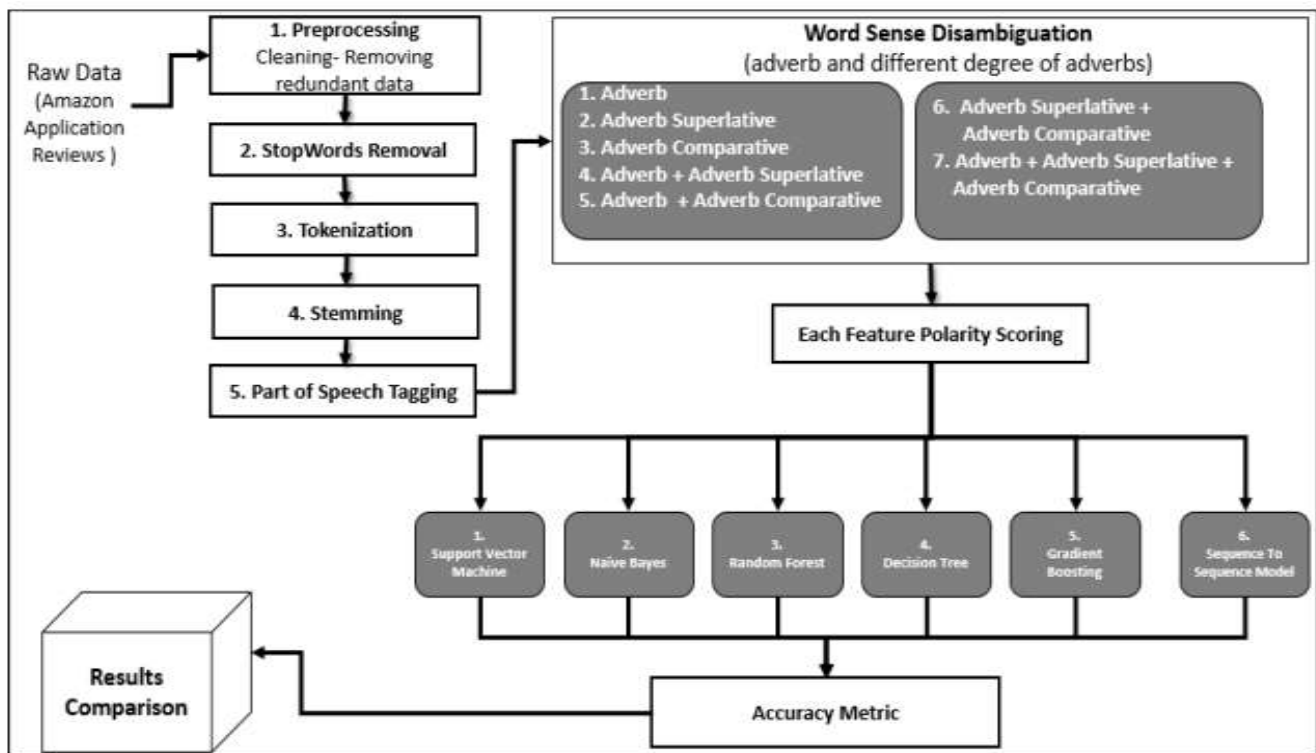

Fig 1: Proposed work model

4. CONCLUSION

Sentiment analysis deals with identifying and aggregating the sentiment or opinions expressed by the users. Sentiment analysis is to classify the polarity of text in document or sentence whether the opinion expressed is positive, negative, or neutral. The approach have been compared and a result for the product review dataset has been done.

5. REFERENCES

[1]   B. Liu, "Sentiment analysis and opinion mining," Synth. Lectures Hum. Lang. Technol., vol. 5, no. 1, pp. 1 167, 2012.

[2]   P. Smith, "Sentiment analysis: Beyond polarity thesis proposal," School Comput. Sci. Univ. Birmingham, Birmingham, U.K., Tech. Rep., Oct. 2011, pp. 1 42.

[3]   R. S. Jagdale, V. S. Shirsat, and S. N. Deshmukh, "Sentiment analysis on product reviews using machine learning techniques," in Cognitive Informatics and Soft Computing. Singapore: Springer, 2018, pp. 639 647, doi: 10.1007/978-981-13-0617-4_61.

[4]   M. Devika, C. Sunitha, and A. Ganesh, "Sentiment analysis: A com-parative study on different approaches," Procedia Comput. Sci., vol. 87, pp. 44 49, Jan. 2016.

[5]   K. Schouten and F. Frasincar, "Survey on aspect level sentiment analysis," IEEE Trans. Knowl. Data Eng., vol. 28, no. 3, pp. 813 830, Mar. 2016.

[6]   R. Arulmurugan, K. R. Sabarmathi, and H. Anandakumar, "Classi cation of sentence level sentiment analysis using cloud machine learning tech-niques," Cluster Comput, vol. 22, no. S1, pp. 1199 1209, Jan. 2019, doi: 10.1007/s10586-017-1200-1.

[7]   P. Burnap, R. Gibson, L. Sloan, R. Southern, and M. Williams, "140 characters to victory?: Using Twitter to predict the UK 2015 gen-eral election," Electoral Stud., vol. 41, pp. 230 233, Mar. 2016, doi: 10.1016/j.electstud.2015.11.017.

[8]   A. S. Manek, P. D. Shenoy, M. C. Mohan, and V. K. R, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classi er," World Wide Web, vol. 20, no. 2, pp. 135 154, Mar. 2017, doi: 10.1007/s11280-015-0381-x.

[9]   X. Fang and J. Zhan, "Sentiment analysis using product review data," J. Big Data, vol. 2, no. 1, p. 5, Dec. 2015, doi: 10.1186/s40537-015-0015-2.

[10]   M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2004.

[11]   B. Pang and L. Lee, "Opinion mining and sentiment analysis," Found. Trends Inf. Retr., vol. 1, nos. 1 2, pp. 1 135, 2008, doi: 10.1561/1500000011.

[12]   W.-J. K. Gann, J. Day, and S. Zhou, "Twitter analytics for insider trading fraud detection system," in Proc. 2nd ASE Int. Conf. Big Data, 2014.

[13]   Y. Liu, "Social media tools as a learning resource," J. Educ. Technol. Develop. Exchange, vol. 3, no. 1, pp. 101 114, Mar. 2017.

[14]   S. R. Das and M. Y. Chen, "Yahoo! For Amazon: Sentiment parsing from small talk on the Web," Inst. Oper. Res. Manage. Sci., Catonsville, MD, USA, Tech. Rep., Sep. 2007, vol. 53, no. 9, pp. 1 16.

[15]   T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in Proc. Int. Conf. Knowl. Capture (K-CAP), 2003, pp. 70 77.

[16]   A. Cambero, "A comparative study of Twitter sentiment analysis methods for live applications," B. Thomas Golisano College Comput. Inf. Sci., Rochester Inst. Technol., Rochester, NY, USA, Tech. Rep. 8, 2016.

[17]   R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classi ca-tion algorithms for sentiment classi cation," Inf. Sci., vol. 181, no. 6, pp.   1138   1152,   Mar.   2011,   doi: 10.1016/j.ins.2010.11.023.

[18]   P. Gamallo and M. Garcia, "Citius: A naive Bayes strategy for sentiment analysis on english tweets," in Proc. 8th Int. Workshop Semantic Eval. (SemEval), 2014, pp. 171 175.

[19]   V. Nandi and S. Agrawal, "Political sentiment analysis using hybrid approach," Int. Res. J. Eng. Technol., vol. 3, no. 5, pp. 1621 1627, 2016.

[20]   N. Rajganesh, C. Asha, A. T. Keerthana, and K. Suriya, "A hybrid feedback based book recommendation system using sentiment analysis," Int. J. Sci. Res. Comput. Sci., Eng. Inf. Technol., vol. 3, no. 3, pp. 2456 3307, 2018.

[21]   M. Salehan and D. J. Kim, "Predicting the performance of online consumer reviews: A sentiment mining approach," in Proc. ICIS, 2014.

[22]   Y.-H. Hu, K. Chen, and P.-J. Lee, "The effect of user-controllable lters on the prediction of online hotel reviews," Inf. Manage., vol. 54, no. 6, pp.   728 744, Sep. 2017.

[23]   N. Pappas and A. Popescu-Belis, "Explicit document modeling through weighted multiple instance learning," J. Artif. Intell. Res., vol. 58, pp.   591 626, Jul. 2018.

[24]   P. Lak and O. Turetken, "Star ratings versus sentiment analysis A comparison of explicit and implicit measures of opinions," in Proc. 47th Hawaii Int. Conf. Syst. Sci., Jan. 2014, pp. 796 805, doi: 10.1109/HICSS.2014.106.

[25]   C. Boon, C. Hawkins, K. Bisht, S. J. Coombes, B. Bakrania, K.-H. Wagner, and A. Bulmer, "The king's speech metalanguage of nation, man and class in anecdotes about George III," Boon, vol. 16, no. 2, pp. 281 299, Jul. 2012.

[26]   B. Pang and L. Lee, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales," in Proc. 43rd Annu. Meeting Assoc. Comput. Linguistics (ACL), 2005, pp. 115 124.