

SENTIMENT ANALYSIS AND RUMOUR DETECTION IN ONLINE PRODUCT REVIEWS

E. Poonguzhali¹, K. Priya², P. Janani³

¹Assistant Professor, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry

^{2,3}B.Tech, Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry

Abstract - Since the start of 2000, sentiment analysis has become one in every of the foremost active research areas by researchers acting on natural language processing and social networking analysis. Additionally data mining, web mining, and text mining also are studied extensively. Moreover, the strategy of sentiment analysis has touched many fields, from applied science to management science, from social science to economics, because of the importance given to the business world as an entire and therefore the collectivity. The sentiment analysis study, which is in fact a classification study, was conducted using machine learning algorithms on Facebook, Twitter, and Amazon and Flipkart datasets. The results of those studies were interpreted by completing an accuracy analysis. Here the training datasets are collected and the sentiments and emotions are categorized accordingly. The comments or the reviews will be classified in to positive and negative sentences and later it'll be tagged to the actual 8 emotions like happy, sad or anger, etc after which the reinforcement algorithm DQN will be used. Also the Quantitative reviews like thumbs up, thumbs down star rating and emoji's are classified as satisfaction or dissatisfaction. Thus our system will summarize the feedbacks, extracting the opinions from all this information, giving an overall view of the merchandise and provides the best products as recommendation, that might save time and ease the choice process of the customer thus producing a far better efficiency when above compared to the present systems and predicting the false feedbacks by which the rumours will be identified. Along with which the USSD system is implemented to update the stock details in the product database in the offline mode.

Key Words: Emotions, DQN, Product recommendation, Rumour, Quantitative reviews, USSD.

1. INTRODUCTION

In Today's world the event of internet technologies has given us the chance to get the views and experiences that we've both personal likewise as professional critics. Such studies show that more and more people are commencing to present their opinions on the net. Sentiment analysis could be a rapidly emerging domain in the area of research in the field of Natural Language Processing (NLP) and Machine Learning (ML). It increases the worth of interest in recent years. Sentiment classification is used to verify and analyze the comments or datasets collected from users regarding their opinion, reviews, etc. Sentiment analysis could be a machine learning approach within which machines classify

and analyze the human's sentiments, emotions, opinions etc. about the products using comments, ratings, stars, and thumbs up and down. It is an identifier with which the most effective products is easily obtained from the opinions given by the purchasers. Due to rapid climb of knowledge in E-commerce, it's accustomed to reveal the standard of product. Here the dataset is pre-processed and classified and also the features are extracted. Then the reinforcement algorithm is employed to label the emotions like anger, disgust, fear, guilt, joy, sadness, shame, positive and negatives. After which the false reviews are identified and a notification or a prompt message is send to them. This is often drained in order to prevent the fake comments and also the false predictions for any valuable products. USSD also known as the Unstructured Supplementary Service Data code system is used in both android and non-android phones. This system is implemented here to update the inventory and the price of the goods of our e-commerce site or any e-commerce websites which is done via the USSD codes in the offline mode.

2. SENTIMENT CLASSIFICATION AND RUMOUR DETECTION TECHNIQUES

In the section below, different sentiment classification and rumour detection techniques are discussed in detail.

[1] The purpose of sentiment classification is to work out whether a particular document incorporates a positive or negative nuance. Sentiment classification is extensively employed in many business domains to enhance products or services by understanding the opinions of consumers regarding these products. Deep learning achieves state-of-the-art ends up in various challenging domains. With the growth of deep learning, many studies have proposed deep-learning-based sentiment classification models and achieved better performances compared with conventional machine learning models. However, one practical issue occurring in deep learning based sentiment classification is that the foremost effective model structure depends on the characteristics of the dataset on which the deep learning model is trained; moreover, it's manually determined supported the domain knowledge of an expert or selected from a grid search of possible candidates. Herein, we present a comparative study of varied deep-learning-based sentiment classification model structures to derive meaningful implications for building sentiment classification models. Specially, eight deep learning models, three

supported convolution neural networks supported recurrent neural networks, with two forms of input structures, i.e., word level and character level, are compared with datasets, and thus the classification performances are discussed under different perspectives.

[2] Sentiment analysis is additionally called as opinion mining which shows the people's opinions and emotions about certain products or services. The most problem in sentiment analysis is that the sentiment polarity categorization that determines whether a review is positive, negative or neutral. Past studies proposed different methods, but still there are some research gaps, i) some studies include only three sentiment classes such as positive, neutral and negative, but none of them considered quite 3 classes ii) sentiment polarity features were considered on individual basis but none of them considered on both individual and on combined basis ii) No previous technique considered give sentiment classes with 3 sentiment polarity features like a verb, adverb, adjective and their combinations. During study, we propose a sentiment polarity categorization technique for an oversized data set of online reviews of Instant Videos. A comprehensive data set of 5 hundred thousand online reviews is employed in our research. There are five different classes namely Strongly Negative, Negative, Neutral, Positive and Strongly Positive. Here we also consider three polarity features such as Verb, Adverb, Adjective and their different combinations with their different senses in review-level categorization are included.

[3] In recent years, with the rapid development of Internet technology, online shopping has become a mainstream way for users to buy. Sentiment analysis of outsized number of user reviews on e-commerce platforms can effectively improve user satisfaction. This paper proposes a replacement sentiment analysis model-SLCABG, which is predicted on the sentiment lexicon and combines Convolutional Neural Network (CNN) and attention-based Bidirectional Gated Recurrent Unit (BiGRU). In terms of methods, the SLCABG model combines the benefits of sentiment lexicon and deep learning technology, and overcomes the shortcomings of existing sentiment analysis model of product reviews. The SLCABG model combines the benefits of the sentiment lexicon and deep learning techniques. First, the sentiment lexicon is employed to boost the sentiment features within the reviews. Then the CNN and also the Gated Recurrent Unit (GRU) network are applied to extract the most sentiment features and context features within the reviews and use the attention mechanism to weight. This paper crawls and cleans the important dangdang.com, a famous Chinese e-commerce website, for training and testing, all of which are supported on Chinese. The experimental results proved that the system can effectively improve the performance of text sentiment analysis.

[4] A review expresses the concerned aspects and corresponding assessments a customer has towards a specific item. Extracting the user's interests and product's

features from their aggregated reviews and matching them together to predict the rating could be a common paradigm during a review-based recommender. However, such a paradigm train's model on the aggregated historical review which grows with time and should have much conflicting semantics, thus the scalability and accuracy is also compromised. During this paper, a novel review semantics based model (RSBM) is proposed to boost the performance of the review-based recommender. It consists of three parts: the review semantics extractor, the review semantics generator and also the rating regressor. Firstly, the review semantics extractor uses a convolutional neural network (CNN) to extract the semantic features of a specific review text. Secondly, the semantics generator uses a memory-network liked structure and a focus mechanism to simulate the decision-making process which assesses each concerned aspect of an item to get the review semantics. Within the training phase, the generated semantics is compared with the semantics extracted by the review semantics extractor. Within the last, the generated semantic features are fed into the rating regressor to predict the rating. Experiments on a series of reality datasets show that the proposed model gains better performances than several state of the art recommendation approaches in terms of accuracy and scalability. Within the future work, we we'll attempt to extract the aspect keywords and mix them with users and items to make a knowledge graph. The graph-based techniques like Graph Convolution Network (GCN) are applied there to to create more accurate recommendations.

[5] With the event of e-commerce, more and more users begin to post reviews or comments about the standard of products on the web. Meanwhile, people usually read and understand the previous reviews posted in the site before purchasing online products. However, people are frequently deceived by deceptive opinion spam, which is typically used for promoting the products or damaging their reputations because of economic benefit. Deceptive opinion spam can mislead people's purchase behavior, that the techniques of detecting deceptive opinion spam have extensively been researched in past ten years. To handle the issue in previous techniques, this paper first introduces the task of deceptive opinion spam detection. Then, we summarize the prevailing dataset resources and their construction methods. Then the prevailing methods are analyzed from two different perspectives: the traditional statistical methods and neural network models.

[6] In recent years, recommendation systems have seen significant evolution within the field of information engineering. Usually, the advice systems match users' preferences supported the star ratings provided by the users for various products. However, just relying on users' ratings about a product can produce biased opinions, as a user's textual feedback may differ from the item rating provided by the user. Here, we propose Social Rec, a hybrid context-aware recommendation framework that utilizes a rating inference approach to include users' textual reviews into traditional collaborative filtering methods for personalized

recommendation of varied items. Text-mining algorithms are applied on a large-scale user item feedback dataset to compute the sentiment scores and results. We propose a greedy heuristic to supply ranking of things supported user's social similarities and matching preferences. To deal with challenges resulting from cold start and data sparseness, Social Rec introduces pre-computation models supported Hub-Average (HA) inference.

[7] With the appearance of web technology, user-generated textual reviews have become increasingly accumulated on many e-commerce websites. These reviews contain not only the user comments on different aspects of the products but also the user sentiments related to the aspects. Although these user sentiments function as vital side information for improving the performance of recommender systems, most existing approaches ignore to completely exploit them in modeling the ne-grained user-item interaction for improving recommender system performance. Thus this paper proposes a sentiment aware deep recommender system with neural attention network (SDRA), which may capture both the aspects of products and therefore the underlying user sentiments related to the aspects for improving the advice system performance. Particularly, a semi-supervised topic model is intended to extract the aspects of the merchandise and therefore the associated sentiment lexicons from the user textual reviews, which are then incorporated into long short term memory (LSTM) encoder via an interactive neural attention mechanism for better learning of the user and item sentiment-aware representation. The extensive experiments on different datasets showed that our proposed SDRA model is able to do better performance over the baseline approaches.

[8] In order to get evaluation information about the assorted aspects of products or services, the Fine-grained Topic Sentiment Unification (FG-TSU) model is proposed based on the development of LDA (Latent Dirichlet Allocation) model. Firstly, the topics are divided into local and global topic and also the sliding window is added to lower co-occurrence information from document to sentence level, to implement fine-grained extraction of local topics. Then the indicator variables are used to differentiate aspects and opinions. Lastly, we incorporate the sentiment layer into LDA model to get the sentiment polarity of the full review and specific aspects of the products and reviews. The datasets of hotel and mobiles are selected to verify the domain adaptability of this model. The experimental results verified the feasibility of FG-TSU model within the realization of opinion mining.

[9] There are many blogs that recommend places and foods and on the internet. Additionally, there are various fake news that provides false information. They both are written by a blogger; bloggers can write on any topics of their own choice. Web visitors read these blogs and judge if place or food item is satisfactory. This suggests that the choice is predicated on the blogger's prejudice. This can be not objective because all the selections depend upon the

blogger's disposition. Other visitors, who had followed the bloggers recommendation, may have dis trust with the blogger. To avoid this convict, all the words and sentences within the posts must be analyzed objectively. All entries like direction, address, excessive compliments, and monophonic are analyzed. This study also analyzed the entries to work out their correlation and, it can make the choice if a blog is trustable with an anomaly sign and also found out the fake reviews in the blogs.

[10] Fake news on major social media platforms has planet consequences on the feelings of the citizens. The research methodology is to implement current block chain technology with advanced computer science in social media platform to forestall fake news. This study aims to supply a considerable review on implementing block chain on social media so as to make trust on credible news and forestall spread of fake news via social media. Particularly, this paper provides the research problem and discusses state-of-the-art block chain solutions and technical constraints also as points out the long term research direction in tackling the challenges.

[11] Unstructured Supplementary Services Data (USSD) is also a real-time session oriented technology, accustomed to provide mobile based network and banking services with/without Internet using USSD codes over GSM channel. For banking services using USSD, the Mobile Network Operator (MNO) is employed as an interface between the customer and his respective bank. USSD may be a user interactive menu driven, cheaper and faster solution and is much better than SMS with regard to cost, security and channel usage. It's platform independent, available in multiple regional languages and doesn't require any software download. USSD is taken into account as an improved choice for non-smartphone users of rural community to avail network services and bank payment solutions. To make security enhancements in USSD, it's necessary to know its architecture, its benefits, threats and issues. Thus this paper presents the small print on USSD architecture and USSD based mobile banking application information security. The long run activity shall include design and evaluation of a multifactor multimodal system to deal with all the identified security requirements gaps i.e. Authentication, Confidentiality, Authorization and Data Integrity in mobile banking application.

[12] Here it examines the employment of Network-based location determination as an alternate to symbolic locations within the provision of location-based mobile advertising services using SMS and USSD. Two alternatives are explored one is that the use of reverse geo-coding and also the other is that the use of coordinates to spot the locations of both users and providers. Both approaches are found to be feasible. However, the shortage of a close geo-code database to be used in reverse geo-coding moreover because the lack of mobile network cooperation within the provision of user coordinates emerge because the main challenges to implementing the proposed model. The proposed LBMA using network-based positioning can be applied in delivering

location-specific information from an SMS and USSD based system. The employment of network based positioning eliminates the requirement for users and service / product providers to explicitly specify their location when making inquiries moreover as when on the move. A user trying to find information during a locality would be able to receive contacts of providers currently in their vicinity.

2.1 PROPOSED METHOD

Sentiment analysis and Rumour detection is done on dataset collected by creating an own website. Data Collection is that the process of collecting information set from a specific source and analyzing the data's so as to produce a desired output. The initial data are collected from the own e-commerce website named T4Tigo. The info is collected in various different forms like textual format and also ratings such as star rating, thumbs-up and thumbs-down and Emoji. The dataset consists of the attributes such as: Reviewer ID, Product ID, Review Text, Rating, Contact number and Location. The primary stage of study involves preprocessing of the reviews. Preprocessing involves the subsequent operations: stemming, stop word removal and part-of-speech tagging. The data pre-processing includes data cleaning, missing data, noisy data, data transformation, data reduction, etc. The input data is pre-processed to boost the results. Parts of Speech Tagging (POST) are accustomed to tag the elements for positive phrases. This processing is especially done to eliminate the incomplete, noisy and inconsistent data and stop words in any input data or sentence. Term Frequency-Inverse Document Frequency (TF-IDF) is used to transform text into a meaningful representation of numbers. This process is widely used to label the features across various natural language process applications. Then the reinforcement algorithm DQN and prediction techniques are employed to label the sentiments and emotions like negative, positive, shame, anger, joy, sadness, fear, happy, guilt and disgust. Common parameters for sentence categorization evaluation include accuracy, precision and recall. A Comparative analysis of SVM and DQN will also be obtained.

Our system will summarize the feedbacks, extracting the opinions from all this information, giving an overall view of the merchandise along with product recommendation, that might save time and ease the choice process of the customer. Fake review is predicted and notification is sent to the particular customer in order to avoid it. Finally the USSD code system (Unstructured Supplementary Service Data) is used to update the details in the product database. The details such as inventory or stock update and price update can be done by employing the USSD system. This messaging system can be used in both the android mobiles and the non-android mobiles i.e., Basic key pad mobiles. By using this system the customers having the key pad phones can easily update their goods details in any e-commerce website by their own. Thus it can be done in offline mode where no e-commerce site has this feature. This serves to be one of the most unique features of our system.

The future works will be done using the IVR(Interactive voice call) system by using voice commands for goods update and Offline payment from one bank to another bank through the offline mode.

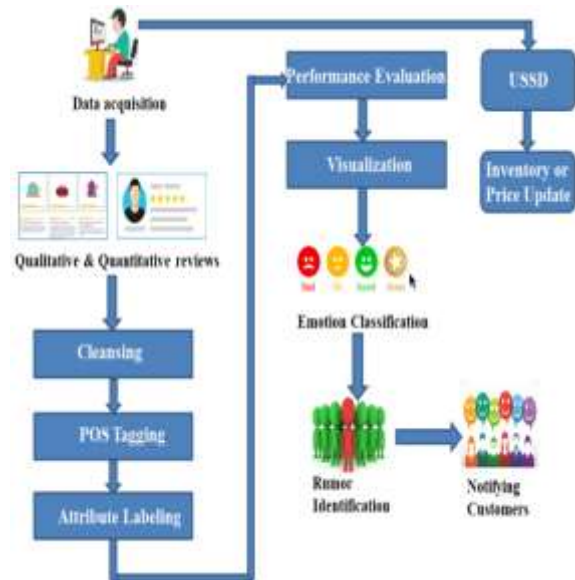


Fig -2.2: Work Flow of Proposed Method

3. CONCLUSIONS

The increase in the use of computers and the internet has caused a serious increase in the methods of information extraction from social media and different E-commerce sites. The sentiment analysis work was conducted on T4Tigo data's. Obtaining those reviews, clearing data, transforming data into numerical form, extracting meaningful results and interpreting them are performed. Sentiment analysis can be used to filter natural language explanations into advice of what to do and warnings of what not to do. Negative classifications should not be immediately trusted since there is a high likelihood of false negatives. Hence we implemented reinforcement learning algorithm particularly DQN with functional approximations, to train a ML agent to develop an optimal strategy to perform automated trading. Thus any one of the eight different emotions will be tagged to the proper review. This is extended to include the content of news about product which will help the model to classify the news more accurately, thereby increasing the accuracy. Thus this system provides a better efficiency and adaptability to the end users to view the best results of the products and also detects the fake reviews posted by certain customers and analysis the results. In the future studies, it is aimed to carry out sentiment analysis studies on different sets using different machine learning and intelligent optimization algorithms. In order to increase value of accuracy, it is foreseen to prepare more suitable data sets to increase the accuracy rate of the studies. Using intelligent search and optimization algorithms with optimized parameters may also be used with integrated feature selection methods to increase the sentiment analysis

performances. Finally it helps even the people with less use to smart phones also to update their goods details in any e-commerce site on their own thus provides ease of use and increasing their economy.

REFERENCES

- [1] Seungwan Seo, Czangyeob Kim and Haedong Kim, "Comparative study of Deep Learning-Based Sentiment Classification", in IEEE Access, vol. 8, pp. 6861-6875, 2020.
- [2] Yafeng Ren and Donghong Ji, "Learning to detect deceptive opinion spam", in IEEE Access, vol. 7, pp. 42934-42945, 2019.
- [3] Rizwana Irfan, Osman Khalid and Faisal Rehman, "A Context-Aware recommendation framework with explicit sentiment analysis", in IEEE Access, vol. 7, pp. 116295-116308, 2019.
- [4] Liping Yu, Liming Wang and Dongsheng Liu, "Research on intelligence computing models of fine-grained opinion mining in online reviews", in IEEE Access, vol. 7, pp. 116900-116910, 2019.
- [5] Samina Kausar, Xu Huahu and Waqas Ahmad, "A sentiment polarity categorization technique for online product reviews", in IEEE Access, vol. 8, pp. 3594-3605, 2020.
- [6] Li Yang, Ying Li and Jin Wang, "Sentiment analysis for e-commerce product reviews in chinese based on sentiment lexicon and deep learning", in IEEE Access, vol. 8, pp. 23522-23530, 2020.
- [7] Aminu da'u and Naomie Salim, "Sentiment-Aware deep recommender system with neural attention networks", in IEEE Access, vol. 7, pp. 45472-45484, 2019.
- [8] Renhua Cao, Xingming Zhang and Haoxiang Wang, "A review semantics based model for rating prediction", in IEEE Access, vol. 8, pp. 4714-4723, 2020.
- [9] Wee Jing Tee and Raja kumar Murugesan, "Trust network, blockchain and evolution in social media to build trust and prevent fake news", in Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), Subang Jaya, Malaysia, 2018, pp. 1-6.
- [10] Hoon Ko, Libor Mesicek and Jong Youl Hong, "Blog reliability analysis with conflicting interests of contexts in the extended branch of cyber-security", in IEEE Access, vol. 7, pp. 143693-143698, 2019.
- [11] Kirthiga Lakshmi, Himanshu Gupta and Jayanthi Ranjan, "USSD-Architecture, analysis, security threats, issues and enhancements", 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), Dubai, 2017, pp. 798-802.
- [12] Moses Thiga, "A model for the delivery of sms and ussd location based mobile advertising network based positioning", 2016 IST-Africa Week Conference, Durban, 2016, pp. 1-10.