Sentiment Analysis of Review Data with Blockchain Security

Barathkumar L[1], Dr. Lekha J[2], Gowtham Prasad S[3], Naveen B[4]

[1,2,3,4]Department Of Computer Science/Sri Krishna Arts and Science College/Affiliated to Bharathiar University/India

--*****************--

ABSTRACT

The analysis of a large number of reviews data generated in the business process can be used to facilitate regulation: since the review data is short text and it is easy to extract the features through deep learning methods. Through these features, the sentiment analysis of the review data can be carried out to obtain the users' emotional tendency for a specific product or a business. Regulators can formulate reasonable regulation strategies based on the analysis results. However, the data has many issues such as poor reliability and easy tampering at present, which greatly affects the outcome and can lead regulators to make some unreasonable regulatory decisions according to the results. Blockchain provides the possibility of solving these problems due to its transparency and non-modifiable features. Based on these, the blockchain technology can be applied for data storage, and the Long short-term memory (LSTM) network can be employed to mine reviews data for emotion analysis. In order to improve the accuracy of the results, we designed a method to make LSTM better understand text data such as reviews which can also contain idioms. In order to prove the effectiveness of the proposed method, different experiments were used for verification, with all results showing that the proposed method can achieve a good outcome in the emotion analysis, thus leading to businesses making better decisions.

Keywords: Blockchain, Decentralized, LSTM, Sentiment analysis

I. INTRODUCTION

With the development of e-commerce, a large number of products and its corresponding reviews have been generated. The analysis of reviews data can provide a basis for regulation. In addition, it solves many problems such as descriptions about the product on the website when it does not match the actual object. Due to the fact that the reviews contain emotional information, the sentiment analysis of reviews not only provides references for consumers, but also enables business people to objectively recognize the advantages and disadvantages of their products. Thus, the emotion analysis of reviews has good commercial value as well as playing an important role in many researches. [3]

Sentiment analysis is also called review mining or opinion mining, which aims at identifying, extracting and organizing the emotions contained in text data collected from social applications, blogs, tweets, reviews and others. Most traditional sentiment analysis methods are based on sentiment knowledge, which uses some existing sentiment dictionaries and language technologies can meet these challenges well. It has powerful computational models that improve the many tasks of sentiment analysis including sentiment classification of sentences, sentiment extraction and lexicon learning. However, it still cannot solve some problems that currently exist in data analysis, such as weak data source reliability, data being easily tampered with, and asymmetric permissions for data access. These problems will greatly affect the accuracy of the analysis results.

Blockchain provides a way to solve these problems. The distributed feature of the blockchain network means each node has equal possibility permission and can share the data. This means that, information of transaction can be recorded in the block-chain after the transaction is finished, where it cannot be tampered with and it is open to all nodes in the entire network. The data recorded on it can be considered as a reliable source of reference information, because of the transparent feature. In addition, the blockchain network can also record the information of every link involved in the whole transaction process, which provides an effective basis for the implementation of regulation.[7]

Motivated by these, we propose a sentiment analysis method for review text combining blockchain and a deep learning model to provide regulatory basis and strategy. Blockchain is used to record transactions information and review data after the transactions have finished. Review data like some containing idioms may cause analysis errors can also be well stored

in the blockchain. Its features such as complete, non-tamper-able and fully shared can provide reliable data for sentiment analysis.

Here, sentiment analysis is conducted by a Long Short-Term Memory (LSTM) network since it has great performance in text analysis, and has been verified in the experiments. The highlights of this research can be divided into two parts.

- In order to ensure the authenticity and validity of the data, a platform based on blockchain has been developed for data storage. Users can make transactions and post related review information through this platform.
- According to the results of sentiment analysis, the proportion of negative reviews can provide a basis and strategy for regulation. The case study proves the effectiveness of the method used for market regulation.

The remaining structure of the paper: The Literature review is discussed in Section 2. Section 3 presents the Existing system/method, proposed system is described in section 4, Section 5 demonstrates the Framework of the system,Implementation and experiment in section 6 and a Conclusion in Section 7.

II. LITERATURE REVIEW

Table I: Literature Review

S.No.	Author(s)	Year of Publication	Title	Review
1	Dylan Yaga, Peter Mell, Nik Roby, Karen Scarfone	2018	Blockchain Technology Overview	This paper justifies that the Blockchain offers a secured way to transact data using decentralized network.
2	Wang L, Niu J, Yu, S. Sentidiff	2019	Combining textual information and sentiment diffusion patterns for twitter sentiment analysis.	Focuses on the fusion of textual information and sentiment propagation patterns using Twitter messages.
3	Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P	2011	Natural language processing (almost) from scratch.	This paper proves that a simple deep learning framework outperforms better approaches in multiple NLP tasks.
4	Tang B, He H,Baggenstoss P.M, Kay S	2016	A Bayesian classification approach using class-specific features for text categorization.	Traditional ML models like Naïve-Bayes were used along with the deep learning algorithms.
5	MahajanD, Chaudhary D.K	2018	Sentiment analysis using RNN and google translator	Authors used Recurrent Neural Network(RNN) to analyze the sentiment tendencies in text data and Google Translate to improveaccuracy.
6	Pelosi s	2020	Semantically Oriented Idioms for Sentiment Analysis	Author discusses about a semantic oriented model to catch idioms for sentiment analysis.

III. EXISTING SYSTEM

Existing systems include forming sentiment propagation patterns using traditional Machine Learning algorithms like Naïve-Bayes (NB) and Support Vector Machine (SVM). But these systems provide very less accuracy. Later some authors have included a neural network called Recurrent Neural Network (RNN) to analyze the sentiment tendencies in text data and to improve accuracy.And also, the NLP sometimes cannot understand and identify idioms in the data. However, the data has many issues such as poor reliability and easy tampering at present, which greatly affects the outcome and can lead regulators to make some unreasonable decisions[1][14][15]. These flaws have to be corrected and the proposed system is given below.

A. Limitations

- When working with large sequence data, RNN or other machine learning models consume more memory and in turn takes more time to train the data.
- RNN uses only one unit gate for input thus it cannot keep the data in memory for long time.

- In reviews, idioms are used to express some emotions. Normal NLP cannot identify these idioms thus gives an improper analysis of data.
- In results of sentiment analysis, negative review provides a strategy for regulation but NLP cannot understand negative idioms given in the reviews.
- Transaction of sentiment data is not authentic and it is vulnerable to tampering. It results mixed analysis of the review data.

IV. PROPOSED SYSTEM

We propose a sentiment analysis method for review text combining blockchain and a deep learning model to provide regulatory basis and strategy. Blockchain is used to record transactions information and review data after the transactions have been finished.The decentralized nature of the blockchain network means each node has equal possibility permission and can share the data [11]. Sentiment analysis is conducted by a Long Short-Term Memory (LSTM) network since it has good performance in text analysis.

A. Primary Objective

Primary Objective of this research is to increase the accuracy of sentiment analysis of review data and to secure the analysis using Blockchain.

B. Secondary Objective

The improved model is based on LSTM, which has four neural network layers, and each network layer interacts with others in a special way. RNNs have a disadvantage that it cannot be used in large sequence data thus resulting in gradient vanishing. LSTM overcomes this problem by filling the gap length of relative insensitivity. We developed a method combined with LSTM to better understand idioms and other phrases in the sentiment analysis process. An external memory component matrix is combined with LSTM to store idioms and many phrases can be synthesized from the words composed of it when modeling the sentence. Each row in the matrix corresponds according to the input from the input gate.Since there is a possibility of poor reliability and easy tampering at present in analysis, Blockchain is used in this process. The data is separated into blocks and stored securely. Blockchain provides transparency and non-modifiable features.[9]

V. THE FRAMEWORK

As a data hosting platform, blockchain can ensure the reliability of data due to its distributed and tamper proof features. Therefore, these trustworthy data can be used to feed the improved LSTM for data analysis to obtain creditable results, which is very important for regulation. In essence, the blockchain relies on modern P2P technology to achieve decentralized data sharing and storage. This feature enables any node in the network to view and access the data in the blockchain. [7]
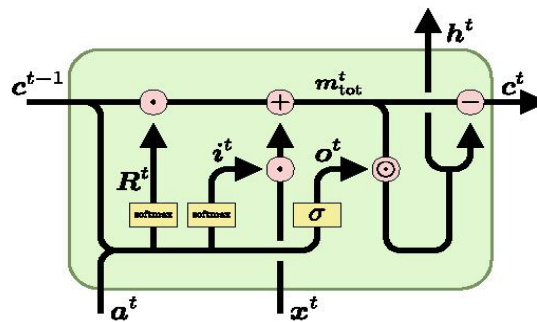

Figure 1: LSTM algorithm flow [4]

The proposed method mainly involves three types of entities: users/stakeholders, processors and regulators. Based on these three types of entities, the method mainly includes three modules, which are transaction execution and review completion, sentiment analysis and regulation based on the results. Based on these modules, the details of these entities can be described as follows.

Every transaction between nodes consists of a consensus among stakeholders. This function provides a more flexible and easier framework for the system we want to develop. The decentralized system runs on a blockchain-based virtual machine, allowing users to independently evaluate transactions and receive feedback about transactions through smart contracts triggered by the transactions. It can meet the needs of the users more quickly, and can integrate the work of regulatory agencies into existing systems at the lowest cost, which is more effective. [3]
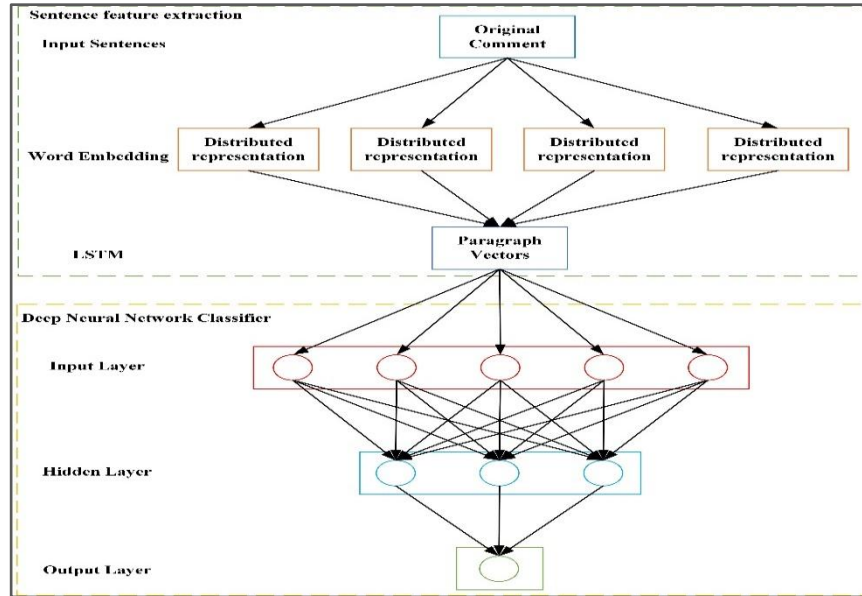
Figure 2: Sentiment analysis LSTM architecture [5]

A. *Pseudocode for LSTM implementation*

```
-- Initialize the size of the review data, Sn = {Sᵢ, Sᵢ, Sᵢ,…..Sₙ₋₁,
Sₙ}, where (i = 1,2,3,…)

-- Initialize matrix for idioms and phrases R = {aₙ, bₙ}

-- function LSTMCELL(prev_st, curr_st, input)

--      combine = prev_st + input

--      ft = forget_layer(combine)

--      ct = candidate_layer(combine)

--      it = input_layer(combine)

--      ot = output_layer(combine)

--      A = (prev_st * ft) + (ct + it)

--      B = ot * tanh(A)

--      return A,B

-- Feature engineer data with LSTM_classifier

-- Initialize batch size Bn = {N < Sn}

-- Train – Test split evaluation

--      Train the data {np_array(x_train)}

-- Initialize raw data

--      Test the data {np_array(x_test)}
```

where, combine stores the sum of previous state and input. ft, it, ot, ct are the gates of LSTM cell with combine value. 'A' stores the sum of previous state multiplied by ft and sum of ct and it. B stores tanh values of A multiplied by 'ot' gate.

VI. IMPLEMENTATION AND RESULTS

This method includes two components – implementing review analysis with LSTM and transaction of reviews through blockchain. First pre-processing of review data is required as it contains irrelevant words and symbols.

Sentiment analysis of review data is combined with LSTM. It has four neural network layers and three gates – input, Forget, output. These three gates control the state of the LSTM cell. LSTM supports a dataset consisting of formats such as text and image. Here only text is used format is used. Now the features of the data are extracted for the model training. This is called as Sentence Feature Extraction. Then the extracted data flows through the input gate as input. Output and Forget gate control the extent of the LSTM cell. This is called Deep Neural Network Classifier.[2]
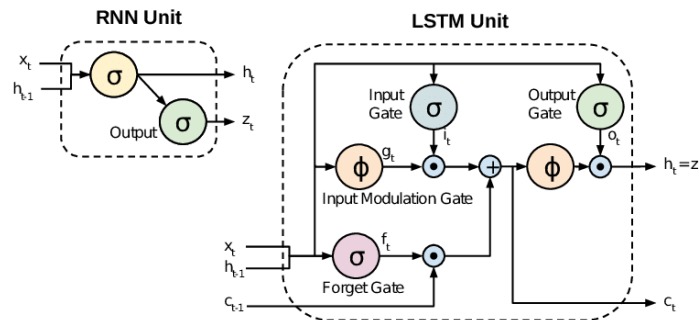

Figure 3: LSTM's input, output and forget gates [6]

To analyze the evaluation results, Accuracy and F1-score are used. Accuracy is calculated by dividing the predicted outcomes by total number of observations.Through this method we have achieved an accuracy of 87.54%. Below chart indicates the difference in accuracy between SVM, Naïve-Bayes and LSTM.
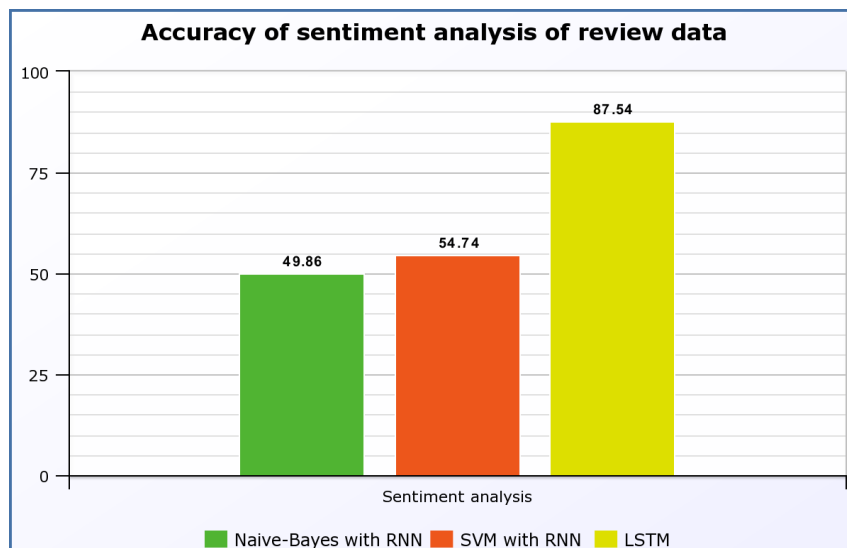

Figure 4: Comparison of accuracy of analysis with different methods

The F1-score combines the precision and recall of a classifier into a single metric by taking their harmonic mean. F1-score of this model is found to be 83.31%.

After the input data is analyzed by LSTM, then it is stored as blocks and becomes part of the ledger.Blockchain simplifies the process of transaction through its distributed network. First the review data is stored as blocks and fed as input to

Sentiment Analysis module. After the transaction is complete, users can view the analysis. The process is implemented through Ethereum smart contracts. Since data are stored in blockchain, it is more secure and it is not liable to tampering.

A. *System flow of review data*
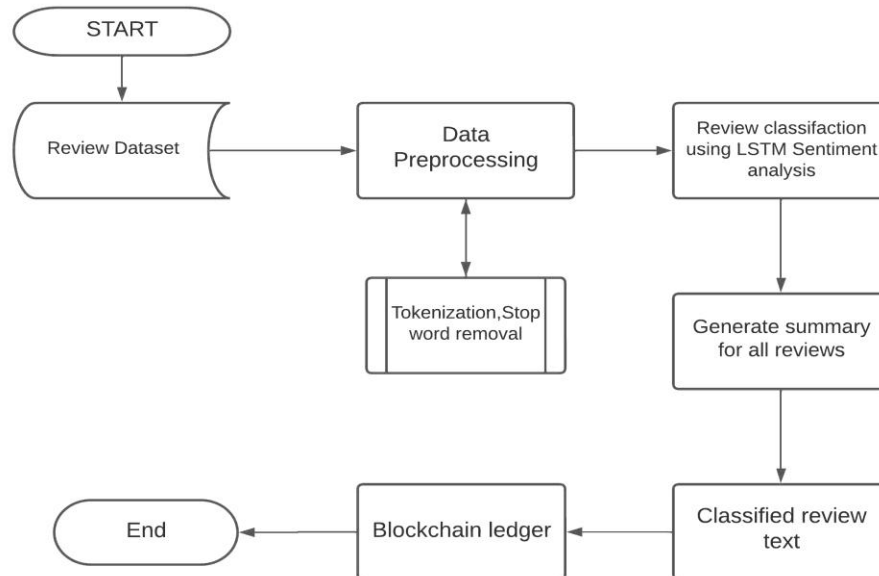
CONCLUSION

This research paper showcases the idea of improving the accuracy sentiment analysis of review data. As mentioned above in Section 2 and 3, previous methods for analyzing review data are not very promising and not intended for deployment. And also, it does not recognize idioms and other symbols. To overcome this problem LSTM classifier is used. It contains three special gates with external memory matrix component to analyze idioms. Blockchain is used for transaction of data and storage as it is more secure and cannot be modified. Further study is being done to increase the speed of analysis and transaction.

REFERENCES

[1] Ms. Gaurangi Patil, Ms. Varsha Galande, Mr. VedantKekan, Ms. Kalpana Dange, Sentiment Analysis Using Support Vector Machine, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1, January (2014).
[2] Sentiment Analysis using LSTM. Tech Stack: Python, Scikit-Learn… | by Abishek PSS | MLearning.ai | Medium
[3] Zhihua Zhao, Zhihao Hao, Guancheng Wang, Dianhui Mao, Bob Zhang , Min Zuo, Jerome YenandGuangjian Tu, Sentiment Analysis of Review Data Using Blockchain and LSTM to Improve Regulation for a Sustainable Market, Journal ofTheoretical and Applied Electronic Commerce Research, December (2021). https://www.mdpi.com/0718-1876/17/1/1/htm
[4] Image from the paper MC-LSTM: Mass-Conserving LSTM, Pieter-Jan Hoedt, Frederik Kratzert, Daniel Klotz, Christina Halmich, Markus Holzleitner, Grey Nearing, Sepp Hochreiter, and Günter Klambauer, (2021).
[5] Image from the paper Cross-Country Skiing Gears Classification using Deep Learning, Aliaa Rassem, Mohammed El-Beltagy, Mohamed Saleh, (2017).
[6] Image from the paper Sentiment Analysis of Review Data Using Blockchain and LSTM to Improve Regulation for a Sustainable Market, Zhihua Zhao, Zhihao Hao, Guancheng Wang, Dianhui Mao, Bob Zhang , Min Zuo, Jerome YenandGuangjian Tu, Journal of Theoretical and Applied Electronic Commerce Research, December (2021). https://www.mdpi.com/0718-1876/17/1/1/htm
[7] Estuar, Ma. Regina Justina E., Ph.D., Towards the development of a blockchain-enabled voice-to-text transcriber plugin in an electronic medical record for doctor-patient conversations, Ateneo de Manila University, (2019). https://archium.ateneo.edu/theses-dissertations/119/

[8] Hoang Tam Vo IBM Research – Australia, Ashish Kundu IBM Research – Yorktown Heights, USA, Mukesh Mohania IBM Research – Australia, Research Directions in BlockchainData Management and Analytics, March (2018). http://www.dke.jku.at/general/news/res/N000026/Mohania%20EDBT%20paper-227.pdf

[9] Sun, S.; Luo, C.; Chen, J, A review of natural language processing techniques for opinion mining systems. Inf. Fusion (2017), Elsevier, Volume 36, Pages 10–25.

[10] Jiawei Yao, Automated sentiment analysis of text data with NLTK. Journal of Physics Conference Series (2019), Issue 5, Volume 1187.

[11] Patel NikunjkumarSureshbhai, Pronaya Bhattacharya, Sudeep Tanwar, A Blockchain-Based Sentiment Analysis Framework for Fraud Cryptocurrency Schemes, IEEE International Conference on Communications Workshops (ICC Workshops), June (2020).

[12] Li, D.; Qian, J. Text sentiment analysis based on long short-term memory. In Proceedings of the 2016 First IEEE International Conference on Computer Communication and the Internet (ICCCI), Wuhan, China, 13–15 October 2016; IEEE: iscataway, NJ, USA, 2016.

[13] Hochreiter, S.; Schmidhuber, J. Long short-term memory, Neural Computation, 1997, Volume 9, Issue 8, IEEE.

[14] Goudjil, M.; Koudil, M.; Bedda, M.; Ghoggali, N. A novel active learning method using SVM for text classification. International Journal Automatic and Computation, Volume 15, Pages 290-298, July 2016.

[15] Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent Trends in Deep Learning Based Natural Language Processing [Review Article, IEEE Computational Intelligence, Volume 13, Issue 3, 2018.