SENTIMENT ANALYSIS OF REVIEW DATA WITH BLOCKCHAIN SECURITY

## Barathkumar L1, Gowtham Prasad S2, Naveen B3, Mrs. Lekha J4

*1 2 3 4 Department of Computer Science*

***Abstract:*** *E-commerce has developed hugely in the past years, as such, its regulations have become one of the most important research areas in order to implement a sustainable market. 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 unmodifiable features. Based on these, the blockchain can be applied for data storage, and the Long short-term memory (LSTM) network can be employed to mine reviews data for emotional tendencies analysis. In order to improve the accuracy of the results, we designed a method to make LSTM better understand text data such as reviews containing 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 sentiment analysis leading to regulators making better decisions.*

***Keywords:* Sentiment analysis, LSTM, Blockchain, Decentralized, Non-Tamper**

1. **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.

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 [7], sentiment extraction and lexicon learning [8]. 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 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 every transaction information 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.

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 related work 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 followed by Future works in Section 7.

# Literature Review

# Table 1. 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 | Mahajan D, 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 improve accuracy. |
| 6 | Pelosi S | 2020 | Semantically Oriented Idioms for Sentiment Analysis | Author discusses about a semantic oriented model to catch idioms for sentiment analysis. |

# EXISITING 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. These flaws have to be corrected and the proposed systems is given below.

# PROPOSED SYSTEM

Author names and affiliations are to be centered beneath the title and printed in Times New Roman 12-point, non-boldface type. (See example below)

### Affiliations

*Affiliations are centered, italicized, not bold. Include e-mail addresses if possible.*

For example:

Author1, Author2 and Author3

*1Affiliation 2Affiliation 3Affiliation*

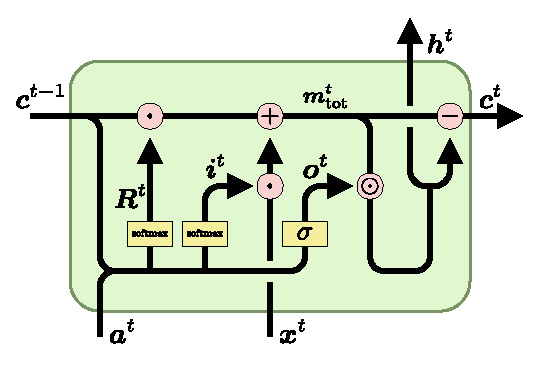
*1Email, 2Email, 3Email*

### Corresponding Author

Corresponding author should have an asterisk sign (\*) if possible, after the corresponding author’s name. The Corresponding author (e.g., \*Corresponding Author) label should be appeared at the footnote section of the first page of the paper, Times New Roman in style and 10 in font size.

# THE FRAMEWORK

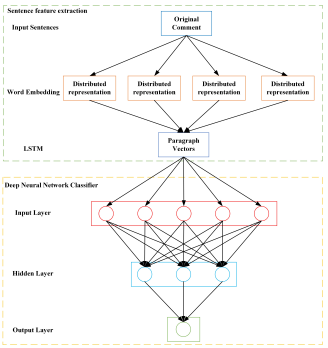
As a data hosting platform, blockchain can ensure the reliability of data due to its distributed and tamper proof features. Therefore, these trustworthy data can be used to feed the improved LSTM for data analysis to obtain creditable results, which is very important for regulation. In essence, the blockchain 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.



**Figure 1. LSTM algorithm flow**

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.



**Figure 2. Sentiment analysis LSTM architecture**

# Type-style and Fonts

Wherever Times New Roman is specified, Times New Roman may be used. If not available in your word processor, please use a font closest to Times New Roman that you have access to. Please avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

# Main Text

Type your main text in 11-point Times New Roman, single-spaced. Do not use double- spacing. All paragraphs should be indented 1 pica (approximately 1/6- or 0.17-inch or

0.43 cm). Be sure your text is fully justified, flush left and flush right. Please do not place any additional blank lines between paragraphs.

## Tables

Place tables as close as possible to the text they refer to and aligned center. A table is labeled *Table* and given a number (*e.g.*, **Table 1. Sample Datasheet with Attributes in Linguistic Term**) it should be numbered consecutively. The table label and caption or title appears 9pt space above the table, 6pt space after the text or paragraph if any; it should be uniforms fonts and font size, and use 11pt font size and Helvetica style, capitalized similar to paper title, aligned center and bold face. Sources and notes appear below the table, aligned left. All tables must be in portrait orientation.

For Example:

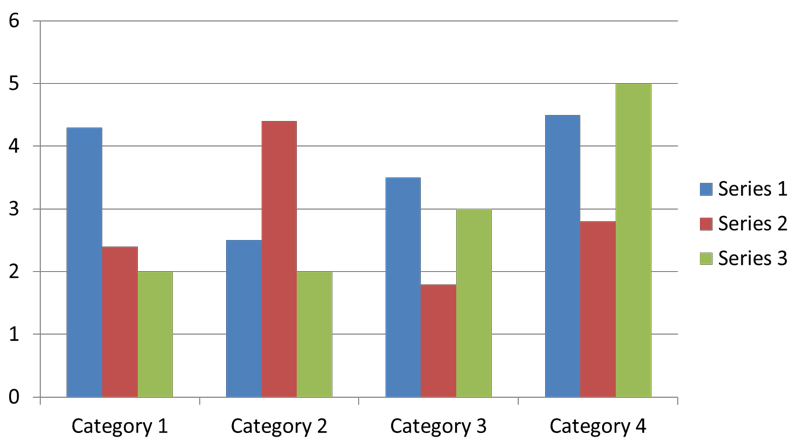
**Table 1. Table Label**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## Figures

Place figures as close as possible to the text they refer to and aligned center. Photos, graphs, charts or diagram should be labeled *Figure* (do not abbreviate) and appear 6pt space below the figure, 12pt space before the next text or paragraph, and assigned a number consecutively. The label and title should be in line with the figure number (*e.g.*, **Figure 1. Location Error Rate of Three Schemes**), it should be uniforms fonts and font size; use 11pt font size and Helvetica style, capitalized similar to paper title, aligned center and bold face. Source (if any) appear underneath, flush left. Figures should be at good enough quality. Minimum image dimensions are 6 cm (2.3622 in) wide by 6 cm (2.3622 in) high.

For Example:



## Equations

**Figure 1. Figure Label**

Including symbols and equations in the text, the variable name and style must be consistent with those in the equations. Equations should be indented at the left margin and numbered at the right margin, equation number is enclosed with open and close parenthesis () Time New Roman in style and 11pt font size. Define all symbols the first time they are used. All equation symbols must be defined in a clear and understandable way.

For Example:

*k*

2



(*z*)   2

*e*  2

 2

[*eik* *z*  *e* 2 ]

(1)

*k*

2

*z*

2

# First-order Headings

For example, “**1. Introduction**”, should be Times New Roman 13-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

* 1. **Second-order Headings (Sub-heading)**

As in this heading, they should be Times New Roman 11-point boldface, initially capitalized, flush left, with one blank line before, and one after.

* + 1. **Third-order Headings:** Third-order headings, as in this paragraph, are discouraged. However, if you must use them, use 11-point Times New Roman, boldface, initially capitalized, flush left, and proceeded by one blank line, followed by a colon and your text on the same line.

# Footnotes

Use footnotes sparingly (or not at all) and place them at the bottom of the column of the page on which they are referenced to. Use Times New Roman 9-point type, single- spaced. To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

# Appendix

An appendix, if needed, should appear before the acknowledgments.

# Acknowledgments

These should be brief and placed at the end of the text before the references.

# REFERENCES

*List and number all bibliographical references that has important contribution on the paper, (if possible, limit to 30, which only are necessary citations are recommended). 9- point Times New Roman, fully justified, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [1]. Do not abbreviate the months. Don’t forget to put period (.) at the end of each reference. (See examples below)*

#### Journal Article

1. *C. D. Scott and R. E. Smalley, “Diagnostic Ultrasound: Principles and Instruments”, Journal of Nanosci. Nanotechnology., vol. 3, no. 2,* ***(2003)****, pp. 75-80.*

#### Book

1. *H. S. Nalwa, Editor, “Magnetic Nanostructures”, American Scientific Publishers, Los Angeles,* ***(2003)****.*

#### Chapter in a Book

1. *H. V. Jansen, N. R. Tas and J. W. Berenschot, “Encyclopedia of Nanoscience and Nanotechnology”, Edited H. S. Nalwa, American Scientific Publishers, Los Angeles, vol. 5,* ***(2004)****, pp. 163-275.*

#### Conference Proceedings

1. *J. Kimura and H. Shibasaki, “Recent Advances in Clinical Neurophysiology”, Proceedings of the 10th International Congress of EMG and Clinical Neurophysiology, Kyoto, Japan,* ***(1995)*** *October 15-19.*

#### Patent

1. *C. E. Larsen, R. Trip and C. R. Johnson, “Methods for procedures related to the electrophysiology of the heart”, U.S. Patent 5,529,067,* ***(1995)*** *June 25.*