# A Novel Parallelized LSTM For Detecting Internet Food Safety

line 1: 1st Qingchao Huang
line 2: *IoT school*
line 3: *Jiangnan University*
line 4: Wuxi, China
line 5: la_lala199@163.com

line 1: 2nd Jun Sun
line 2: *IoT school*
line 3: *Jiangnan University*
line 4: Wuxi, China
line 5: sunjun_wx@hotmail.com

line 1: 3rd Jianhua Wang
line 2: *Business school*
line 3: *Jiangnan University*
line 4: Wuxi, China
line 5: jianhua.w@jiangnan.edu.cn

***Abstract*—Food safety is a major problem concerning people's livelihood. With the advent of the era of Internet, many people choose to order food online, while the regulation of online food safety is faced with enormous challenges. Through the analysis of the comments data from third-party platform, a food safety evaluation dataset of violation and risk is constructed. In order to find the relationship between the comment data and risks level of online food, a novel parallelized distributed long and short term memory network model is proposed to predict the risk value of merchants, and an early warning system for network takeout merchants is established.**

***Keywords—food safety, dataset, distributed, risk index***

## I. INTRODUCTION

A serious problem of offline food are sold on third-party internet platform without certain standards and regulations, only a food safety certificate and business license are required for online merchants, making it difficult to detect some illegal businesses. In addition, there are many kinds of food result in that it is very difficult to collect and investigate the information, and impossible to inspect the production environment and hygiene status of food merchants on the spot. Therefore, food safety problems become serious.

From the perspective of consumers, it is possible to predict whether there exist illegal behaviors by analyzing online comments of merchants. However, there are no enough comments data on internet catering and without a scientific system to evaluate the risk level of merchants.

There are also many datasets in the field of natural language processing, including THUCnews, the Chinese news classification dataset collected by tsinghua university, the amazon product evaluation dataset, the movie rating dataset. However, those datasets are poorly constructed without using the legal basis and the scientific system to evaluate the risk score of comments.

With the development of deep learning, a number of new networks and models appear in neural networks. For the time series data, the recurrent neural network (RNN) has a good training effect. However, for the sequential data, the recurrent neural network has the problem of gradient disappearance and gradient explosion. Therefore, S. Horeiter et al. proposed the long and short-term memory network (LSTM) to solve the dependence problem of RNN. For the existing time series data, the model was built by using the single-machine processing method. For large-scale time series data, the current popular single-machine platform could not meet the requirements.

Considering the situation above, this paper build a more scientific datasets for food safety analysis based on legal basis. Besides, a distributed parallelization LSTM network model is implemented on Spark platform to deal with large-scale time-series data, and obtain the probability of violation behavior of merchants.

The rest of the paper are organized as follows: Section II introduces the basic RNN and LSTM model simply. Section III describes the food safety dataset we had create. Experiment results and analysis are demonstrated in Section IV. Finally, we conclude this paper in Section V.

## II. MODEL BUILDING

### A. Recurrent Neural Network

Recurrent Neural Network (RNN) is proposed for temporal sequence data, It is a special network structure allows the output of the neuron to act directly on itself as input at the next time point The output of the implementation network is the result of the input at that moment acting together with all moments of history.

### B. Long Short-Term Memory

The goal of RNN is to learn the long-term dependency of timing data, But RNN has a hard time learning and retaining long-term information. It has no effect on the historical moment beyond a certain time, which results in the poor performance of RNN on long sequence data. Therefore, LSTM is proposed to solve this problem through a unique gating unit. LSTM is widely used in the processing of time series data, and has achieved good experimental results in the traditional language processing methods. The model of the LSTM network is shown in Fig. 1.

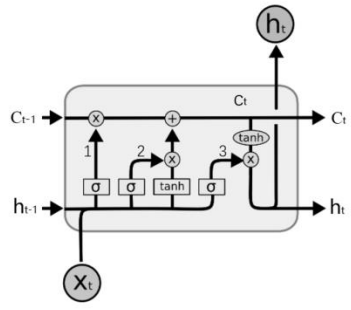


Fig. 1. Long Short-Term Memory model

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00072

As shown in Fig.1, LSTM is made up of cells that are also known as nodes. Each node consists of a Forget Gate, an Input Gate, and an Output Gate, as well as various input-output connections that they control. The LSTM unit processes the information of the previous memory state through the forgetting gate to determine the information to be forgotten from the memory state. The input gate determines the information to be updated for each node. The memory state that needs to be output is controlled through the output gate.

The calculation of each state transfer process in LSTM network is shown in the following formula:

$$f^t = \sigma\left(W_f \cdot \left[h^{(t-1)}, x^t\right] + b_f\right) \quad (1)$$

$$i^t = \sigma(W_i \cdot [h^{t-1}, x^t] + b_i) \quad (2)$$

$$\tilde{C}^t = \tanh(W_g \cdot [h^{t-1}, x^t] + b_g) \quad (3)$$

$$o^t = \sigma(W_o \cdot [h^{t-1}, x^t] + b_o) \quad (4)$$

$$C^t = f_t * C^{t-1} + i^t * \tilde{C}^t \quad (5)$$

$$h^t = o^t \cdot \tanh(C^t) \quad (6)$$

The output of the LSTM network model will consist of three gates: forget gate $f^t$, input gate $i^t$, out gate $o^t$. Where $x^t$ represents the input at time t, $h^t$ represents the hidden output at time t, $W$ represents the weight parameter of the connection, $b$ is the bias parameter, $\tilde{C}^t$ updates the status after multiplying the input gate.

### C. Distributed Long Short-Term Memory

For very large-scale problem, the processing of traditional stand-alone mode is very slow. In order to speed up the processing, a distributed architecture is required. In this paper, the parallelization of LSTM algorithm is designed. The structure of the parallelized LSTM model is shown in Fig.2.

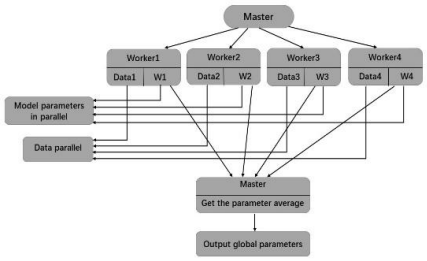


Fig. 2. Parallel Long Short-Term Memory Architecture

The parallelized LSTM model is established as follows:

First, the configuration and parameters of the global LSTM network model are initialed on the master node. It mainly includes LSTM network structure, network layers, learning rate, activation function, and network weight.

Then, data are read by worker nodes, configuration and parameters from the master node are broadcasted to each worker slave node. Subsequently, each worker node trains the data they received and finally the parameters are averaged and returned to the master.

The main parameters that need to be configured for training are as follows:

(1) The batch size of each worker.

(2) The average frequency of the parameters. The period cannot be set too long or too short.

(3) Configure the time for data repartitioning

During training, parameters can easily fall into a local minimum, Therefore, the optimization method can make the connection weight coefficients of some hidden nodes in the LSTM network model ineffective by adding random noise and using the Dropout method. Get all your neurons fully trained. In order to speed up the training rate, the learning rate is adaptively changed with the training process to obtain the best LSTM model.

## III. DATASET SUMMARY

### A. Collected data and data preprocessing

A crawler program was used to capture the basic information data set of a third-party food and beverage sales platform for takeaway businesses. From a consumer perspective, a review of a store can directly reflect the probability of violations. In this dataset, the comment text data are cleaned firstly, then we obtained about 1 million comments. After handling users nicknames as anonymous users, the dataset is obtained. The format of the data is shown in the Table 1.

TABLE I. COMMENTS

| *Nickname* | *Comment time* | *Comments* |
|---|---|---|
| Anonymous User | 2019-07-05 | It's disgusting to eat a hair |
| Anonymous User | 2018-07-22 | A weird smell, really bad |
| Anonymous User | 2018-08-13 | There seems to be no sand |
| Anonymous User | 2019-05-01 | Very tasty, will buy next time |

### B. Building a Scientific DataSet

In order to strengthen the food safety supervision and management of online catering services, standardize the operation behavior of online catering services and ensure food safety, China Food and Drug Administration reviewed and approved the measures for the supervision and administration of food safety in online catering services in 2017. Based on this law and other food safety regulations such as the Food Safety Law, the content of comments that may exist in violation of laws and regulations corresponds to laws and regulations, and 16 types of comment data on possible violations of law and regulations are constructed as shown in Table 2.

TABLE II. COMMENT CLASSIFICATION

| *Different stage* | *example* |
|---|---|
| produce | Use of spoiled ingredients:<br>Odor of meat |
| | Improper food processing procedures:<br>Unripe food, raw vegetables |
| | Foreign matter in the food:<br>have hair |
| | Unhealthy business environment:<br>Eat out flies |
| Delivery | Use of unqualified packaging materials:<br>Smelly packaging |
| | Provide substandard tableware:<br>Moldy tableware |
| | Irregular packaging:<br>Leaked. |
| | Special food quality guarantee:<br>Requires refrigerating, is already warm at hand. |
| | Distribution container hygiene status:<br>Dirty shipping container. |
| Sales | Product label failed:<br>No product label. |
| | Sale of expired food:<br>Leftovers, drinks expire |
| | Inconsistent online and offline sales:<br>Inconsistent online and physical store sales. |
| | Merchant does not process complaints:<br>Merchant does not answer customer calls |
| | The merchant does not provide a sales voucher:<br>Asking successors not to invoice |
| | Merchant provides false information:<br>The name of the dish is different from the actual food |
| | Slip orders and false comments:<br>Merchant sales are brushed out |

Finally, the dataset for food safety analysis is created, it contains 500,000 training sets and 5000 test sets. Based on this dataset, a risk factor for violating laws and regulations of catering platform merchants is constructed.
And store the processed data set in HDFS.

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

### A. Introduction to the experimental environment

Based on Hadoop's distributed file system and Spark big data processing framework, this article implements a parallelized LSTM to predict the risk of illegal and illegal Internet catering businesses.

The platform we used has one master node and three slave nodes. The detail environment configuration is shown in Table 3.

TABLE III. LAB ENVIRONMENT

| *Configuration* | *Parameter Value* |
|---|---|
| Hadoop | Hadoop 2.8 |
| Spark | Spark 2.3.2 |
| Programming Language | Scala 2.11.8 |
| RAM | 32G |
| System | Centos |
| Hard Disk | 16T |

### B. Analysis of Results

Considering that some restaurant merchants' store reviews are not updated frequently, the risk value is calculated every 7 days. The risk value of a merchant during one month is shown in TABLE4

TABLE IV. MERCHANT RISK VALUE

| *Comment time* | *score* |
|---|---|
| 2019-04-01/2019-04-07 | 0.02565545707812748 |
| 2019-04-08/2019-04-14 | 0.02157355142379308 |
| 2019-04-15/2019-04-21 | 0.02134782930654703 |
| 2019-04-22/2019-04-28 | 0.02281103271600721 |
| 2019-04-29/2019-05-05 | 0.02290133275593217 |

For the convenience of display, the risk value is normalized by a linear function:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Subsequently, a line chart of the risk value of the merchant is drawn in Fig. 3, we can see the risk value of this merchant stays in a low position and is decreased with time varying.

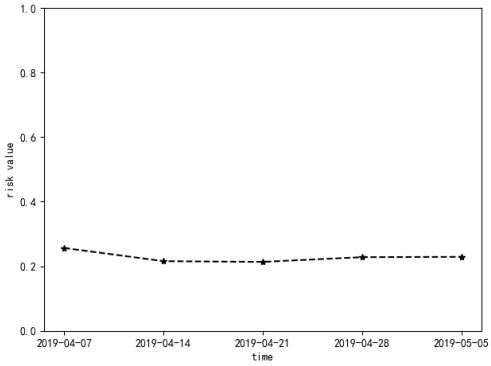


Fig. 3. Change in risk value of a certain merchant

In order to analyze the change of risk value in detail, an average risk value is calculated every four weeks to reflect the change in risk value for this month. For example, the risk value of a merchant over the past 5 months is shown in Figure 4 below:

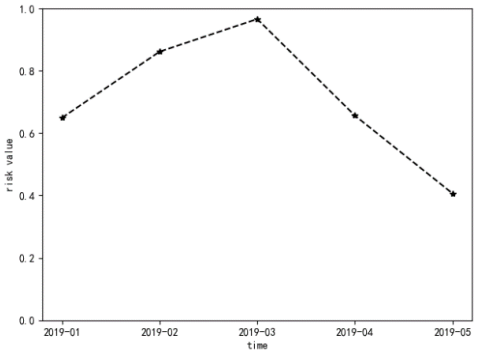


Fig. 4. Risk value of a merchant during 5 months

It can be seen that this merchant has a relatively high risk value in March, and then decreased in the following months.

TABLE V. A SHOP REVIEW IN MARCH

| *Comment time* | *Comments* |
|---|---|
| 2019-03-02 | You can see the order, but today some bullfrogs are red and a little bit stinky |
| 2019-03-03 | The weight of string beans and shredded potatoes is not at the same level. There are too few string beans. |
| 2019-03-06 | Both dishes are so oily, I do n't feel well after eating it, |
| 2019-03-29 | The potatoes are not rotten, stiff |

In order to find the reason that infulences the risk value of this merchant, we search the merchant's store reviews in March are shown in Table5. It is clearly that the comments of this merchant are mostly illegal.

Based on this model and data set, we compare and analyze the risk value of all districts in Wuxi, China.

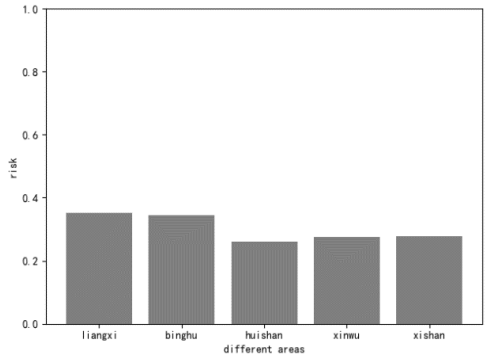


Fig. 5. Comparison of risk value in different regions

By dividing the takeaway businesses in each districts of Wuxi City by region, it can be seen from Fig.5 that the overall risk value of Internet catering businesses in each area of Wuxi City is at a relatively low level. Among them, the risk value in Liangxi District is higher than other areas.

## V. CONCLUSION

Aiming at the massive data of internet catering merchants, we build a scientific dataset based on laws and regulations, and a parallel LSMT network model is established on a distributed computing platform. Through analyzing the massive time series data under this model to obtain the illegal risk value of the restaurant business, so as to achieve the effect of early warning to businesses.

## ACKNOWLEDGMENT

The work described in this paper was fully supported by a grant from the National Key R&D Program of China (No. 2018YFC1603303).

## REFERENCES

[1] Kim Dang Anh,Xuan Tran Bach,Tat Nguyen Cuong,Thi Le Huong,Thi Do Hoa,Duc Nguyen Hinh,Hoang Nguyen Long,Huu Nguyen Tu,Thi Mai Hue,Dinh Tran Tho,Ngo Chau,Thi Minh Vu Thuc,Latkin Carl A,Zhang Melvyn W B,Ho Roger C M. Consumer Preference and Attitude Regarding Online Food Products in Hanoi, Vietnam.[J]. International journal of environmental research and public health,2018,15(5).

[2] Maosong Sun, Jingyang Li, Zhipeng Guo, Yu Zhao, Yabin Zheng, Xiance Si, Zhiyuan Liu. THUCTC: An Efficient Chinese Text Classifier. 2016.

[3] John Blitzer, Mark Dredze, Fernando Pereira. Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification. Association of Computational Linguistics (ACL), 2007

[4] Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011, June). Learning word vectors for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1 (pp. 142-150). Association for Computational Linguistics.

[5] Gers F A, Eck D, Schmidhuber J. Applying LSTM to Time Series Predictable through Time-Window Approaches[M]. Artificial Neural Networks-ICANN 2001. Springer Berlin Heidelberg, 2001:669-676

[6] Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM

[7] Xinxin Wang,Depeng Dang,Zixian Guo. Evaluating the crowd quality for subjective questions based on a Spark computing environment[J]. Future Generation Computer Systems,2020,106.

[8] Yinan Xu,Hui Liu,Zhihao Long. A distributed computing framework for wind speed big data forecasting on Apache Spark[J]. Sustainable Energy Technologies and Assessments,2020,37.

[9] Eftim Zdravevski,Petre Lameski,Cas Apanowicz,Dominik Ślęzak. From Big Data to business analytics: The case study of churn prediction[J]. Applied Soft Computing Journal,2020,90.

[10] Chow V. Predicting auction price of vehicle license plate with deep recurrent neural network[J]. Expert Systems with Applications, 2020, 142: 113008.

[11] Zaharia M, Chowdhury M, Franklin M J, et al. Spark: Cluster computing with working sets[J]. HotCloud, 2010, 10(10-10): 95.

[12] Yeung R M W, Morris J. Food safety risk[J]. British food journal, 2001.