

The Application of Hadoop in Natural Risk Prevention and Control of Rural Microcredit

Huaqing Mao, Li Zhu*

Wenzhou University Oujian College, Wenzhou, China
Email: Mr.maohuaqing@qq.com, 38995791@qq.com

Received 12 February 2015; accepted 15 March 2015; published 17 March 2015

Copyright © 2015 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Rural microcredit means that the loan institutions extend the small amount of loans to the farmers. The purpose of rural microcredit is to meet the increasing needs of agriculture, animal husbandry, aquaculture, and the other business activities associated with the rural economic development. However, the rural microcredit is currently facing severe problems, such as operation risk, business risk and natural risk. Of those risks, the natural risk of rural microcredit has the most different forms and complex relationships, and the effective coping strategies lack of controllability. In the event that we can't control and make up the losses from natural risks, it will cause the rural incomes and productions stepping down; and there is no way to get any compensation from the other capital, and this will cause the farmers can't pay the principal and interest. As a result, natural risk prevention and control become a very important issue in rural microcredit. This paper analyzed the original cause of formation and characteristic of natural risk, and discussed how to predict the natural risk in rural microcredit. Finally, we gave the result and performance evaluation, and provided various methods to defend against the natural risk.

Keywords

Cloud Computing, Hadoop, Rural Microcredit, Natural Risk

1. Introduction

Rural microloan corporation is a new type of financial organization suitable for rural areas and intensify efforts. It can establish perfect countryside social security system, reduce the threshold of rural lending, and ameliorate the rural financial service system [1] [2]. Although the rural microcredit has a history of more than ten years in development, it is still at the initial stage of exploration in China. During the course of maintaining the sustained,

*Corresponding author.

rapid and sound development of the rural finance, there are many risk factors from both external forces of social environment and internal forces of financial system [3]. Those risk factors create an unfavorable impression for the rural microloan corporation in future development. **Figure 1** shows that the probability and damage degree of different risk factors during the loan period. The range of probability and damage degree is 1 to 10. As a result, analyzing and studying the risk management of rural micro loan corporation can effectively prevent and reduce the risk, and it will improve their profit-earning capability which is important to social and economic development in China.

The rural microcredit corporations loan their money to farmer, and let them engage in farming, planting, animal breeding. Those activities of traditional agricultures are strong reliance on natural condition, and their capability to withstand natural disasters is low. Whenever serious natural disasters befall, lots of customers will break the contract at same time, and the loan corporation may go into bankruptcy. But the factors that affect natural risks are many-sided and complex, and it suppress the development of rural microcredit [4]-[6].

To control the natural risk of rural microcredit, a lot of researches have been done home and abroad. In the area of risk-control mechanism, Anyu Li and Man Zhang described the action mechanism of joint loans and dynamic excitation in the risk-dodging, and they systematically expounded the risk of rural microcredit [7]. Qing Ye found that the state shall supported the popularization of agricultural techniques to promote the prompt application of advanced agricultural techniques to agricultural production; the state not only helped with materials and money, but also bonded the monetary policy to balanced allocation of risks [8]. Smith studied different types of risk indexes which would cause the farmer to break the loan contract, and he started a new research method of bank loan risk management [9]. In the area of the risk assessment, Liangwei Chen set up the assess model of farmer credit rating by using the decision tree algorithm [10]. Tao Sun and Gongjing Zhang studied the examining and approving credit card loans by combining the genetic algorithm and rough set theory [11]. Jingxian Zhao and Ziping Du used the credit risk assess model based on neural network algorithm to generate the decision tree and type of breach of contract [12]. Jacobson took the loans business of commercial bank as research subjects, and he used the qualitative and quantitative analysis to identify the risk in loan business, and he constructed the evaluation indexes for future refund ability of loan customer [13].

2. The Cause and Feature of Natural Risk

2.1. Causation Analysis of Natural Risk

The main body of rural microcredit is farmers, who engage in agricultural production, but in China agriculture is weak, it is restricted and influenced by natural condition, so the risk is unavoidable. China is a large agricultural country with numerous mountains, with mountainous areas accounting for 69.2 percent of the country's total land area and the number of people engaging in agricultural production making up 80 percent of the country's total population. Most rural areas agricultural infrastructure is weak, means of production lag behind and it does not have sufficient ability to fight natural calamities. Some farmers are suffered from serious natural calamities

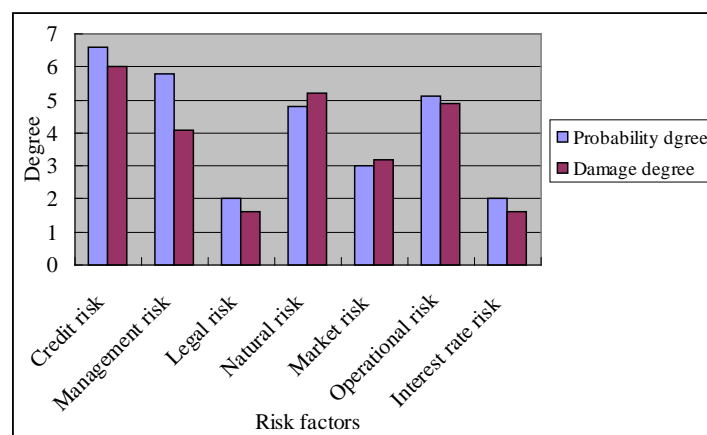


Figure 1. Probability and damage degree of different risk factors during the loan period.

and animal epidemics, those disasters can't predictable and effective prevention, and insurance companies do not cover any accident which is caused by nature. Once farmer sustain a loss cause by nature, and can't get the alternative risk transfer and adequate compensation. For example, a farmer get 60,000 RMB agricultural loan from bank, he used those money to plastic sheeting of vegetables. His annual income was assessed at 120,000 RMB. But the low temperature and rainy days during the growing period resulted in a considerable decrease in production, and he didn't have the wherewithal to repay the loan.

To sum up, the natural disasters have something to do with the rural microcredit fall into two areas: paroxysmal disaster and chronic disaster. The state has set up an emergency system and a social relief system to deal with abrupt natural disasters including: earthquake, mud-rock flow, typhoon, and flood. The chronic disaster is connected with environmental harm including: desertification and drought.

2.2. Feature Analysis of Natural Risk

The characters of natural disaster are mainly manifested in following aspects:

1) Natural disaster has extensiveness and regional. The distribution range of natural disaster is particularly wide, no matter land or sea, no matter urban or rural area, no matter mountain or plain, natural disaster may take place in all the area. The regional of natural environment decided the regional of natural disaster.

2) Natural disaster has frequency and uncertainty. There are many natural disaster take place all over the world, and the occurrence of disaster present the gradual increment trend. However, the uncertainty of time, place and scale are very much increasing the difficulties of resisting natural disaster.

3) Natural disaster has periodicity and unrepeatability. The main natural disasters, whether drought or flood, are periodic occur, but the process and result of disaster are different.

4) Natural disaster has connection. For the bulk of its energy China relies on coal, a leading source of acid rain and pollution. Earthquake may cause the mud-rock flow or air pollution.

5) Natural disaster is unavoidable and alleviated. Since the contradiction always exist between human and nature, as long as the earth is moving and human is developing, the natural disaster will never disappear. However, we can turn the harmful into the beneficiate, and ultimate alleviate the loss of disaster.

3. Cloud Computing Environment of Hadoop

3.1. Feature of Hadoop

If we want to calculate the natural risk of microcredit, we need to process and analysis the history data, and get the reliable conclusion. History data is normally large, but the reading speed of disk can't keep pace with the times. Parallel process those history data by using cloud computing can highly efficient calculate and analysis the big data. Compared with the traditional RDBMS, cloud computing is suitable for the problem which can batch processing, and RDBMS is suitable for point query and point update. Strengths and weaknesses of cloud computing and RDBMS is shown in [Table 1](#).

As the most popular distribute framework of cloud computing, Hadoop is software framework compiled by Java, it can run distribute computation of big data by using cluster composed with lots of computer. Hadoop has two advantages, one is that Hadoop is simple and easy programming the distribute program, the other is that Hadoop has good robustness and extensibility which make it adequate to complex job. The cluster of Hadoop is shown in [Figure 2](#).

3.2. The MapReduce Data Process Model

The MapReduce data process model is a core component in Hadoop, and its great strength is that it is easy to extend the data to the multiple compute nodes. In the MapReduce model, data processing is divided into two stages: mapping and reducing, each stage has one data process function named mapper and reducer. In the mapping stage, MapReduce get the input data and load them into mapper; in the reducing stage, reducer process the output data come from mapper, and give the final result. In short, mapper filter and transfer the original data, and reducer accomplish the aggregation of data. The working processes are as follows:

1) Format the input data into the list with KEY/VALUE like list (<k1, v1>). If you need to deal with multiple files, then the input format should be list (<String filename, String file_content>), and the format of log file should be list (<Integer line_number, String log_event>).

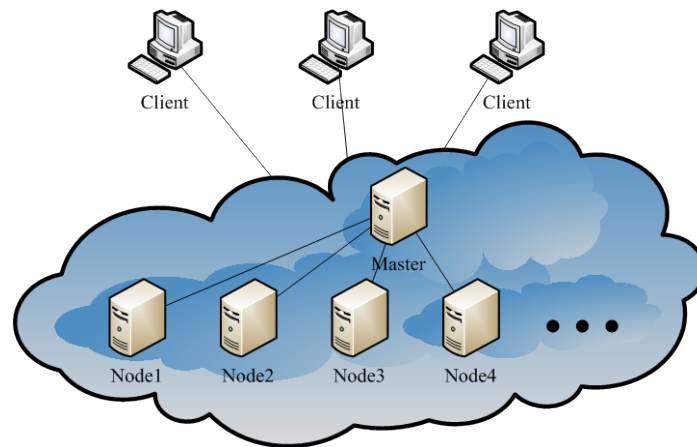


Figure 2. Hadoop cluster.

Table 1. The comparison of RDBMS and cloud computing.

Feature	RDBMS	Cloud computing
Data size	GB	TB
Access mode	Interactive	Batch process
Update mode	Multiple read and write	Write once and multiple read
Structure	Static mode	Dynamic mode
Integrity	High	Low
Extensibility	Nonlinear	Linear

2) Split the list contains the KEY/VALUE, and then calls the map function in mapper to deal with each KEY/VALUE $\langle k1, v1 \rangle$. Mappers process each $\langle k1, v1 \rangle$ and put the result $\langle k2, v2 \rangle$.

3) Shuffle all the output of mapper and output into another list contains $\langle k2, v2 \rangle$, the new list put the VALUE which has same KEY together and like $\langle k2, \text{list}(v2) \rangle$. Reducer will process the $\langle k2, \text{list}(v2) \rangle$ and output the final result. The work flow of whole process is shown in **Figure 3**.

4. Natural Risk Calculation Based on Hadoop

Among the natural risk of rural microcredit, weather is the most important or influential factor. We use weather data as data mining example and process the natural risk calculation. There are lots of meteorological sensors spread around all over China, and they gather the weather data at regular intervals. Those massive weather data is semi-structured data, they store the data as record mode which is very suitable for processing by MapReduce of Hadoop cluster.

4.1. Data Pretreatment

The data we used are derived from China meteorological data sharing service system (<http://cdc.cma.gov.cn/home.do>). The data set named China earth international switching stage day weather value dataset. The dataset is stored with ASCII code, each line represent one record. There are lots of meteorological elements in the dataset, for the sake of calculation convenience; we focus on the basic elements. Such as: Precipitation, pressure, wind speed, temperature, and relative humidity. The unit and precision of sampled-data are shown in **Table 2**.

The above table shows that one line of sampled-data, which is separated into multiple lines to display the meaning of each field. In the sampled-data file those fields are integrated into one line and used the semicolon as delimiter. The weather station identifier 58,457 means Hangzhou which has latitude 30.14° North, longitude 120.1° East, and average altitude 40 meters. Because there are many meteorological observatories, the entire

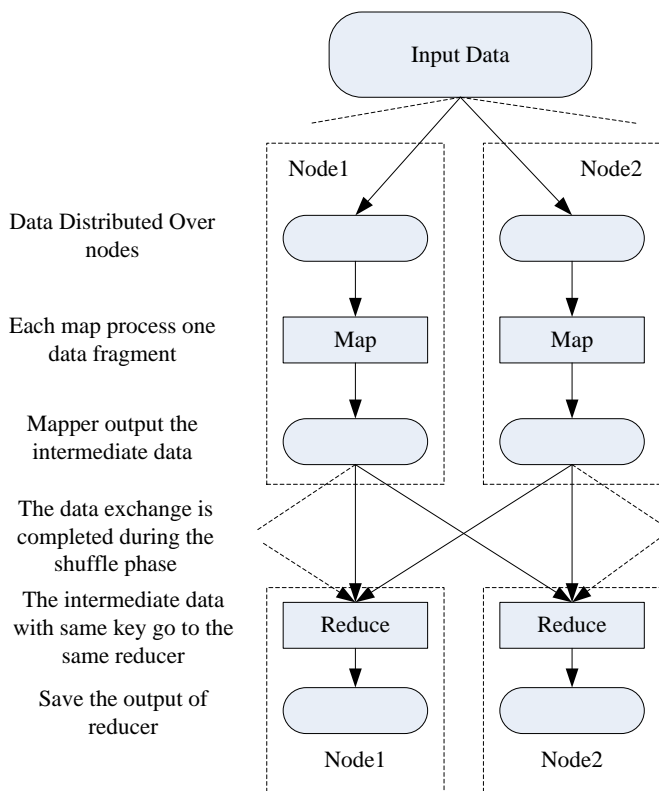


Figure 3. The data stream of MapReduce.

Table 2. Description of data format.

Element name	Unit/precision	Example data
Weather station identifier		58457
Year		2000
Month		1
Date		1
Precipitation	0.1 mm	0
Average pressure	0.1 hPa	10142
Average wind speed	0.1 m/s	18
Average temperature	0.1 °C	78
Average relative humidity	1%	85

dataset is made up with lots of small file. In most cases, Hadoop is more efficient in dealing with small number of large files, so we preprocess the data file by year, and put the data file with same year into a separate file.

4.2. Data Analysis and Computation

In this section, we take the temperature as example, and calculate the data by using MapReduce, and get the maximum/minimum Average temperature. In map stage, we use the text file as input data, which is easy to process the dataset line by line. The KEY value is equal to the offset address of current line's starting position. By using map function, we can pull out the year and average temperature as the input data for reduce function, in the meantime, the missing, false temperature data will be filtered. The part of input data of map stage is as

follows:

```
58457;2000;1;1;0;10142;18;78;85
58457;2000;1;2;0;10178;13;71;85
58457;2000;1;3;0;10155;8;58;88
58457;2000;1;4;0;10119;10;97;80
58457;2000;1;5;32700;10092;13;130;77
```

.....

After we convert the data into KEY/VALUE list, it should be:

```
(0, 58457;2000;1;1;0;10142;18;78;85)
(31, 58457;2000;1;2;0;10178;13;71;85)
(62, 58457;2001;1;1;0;10158;10;66;69)
(93, 58457;2001;1;2;32700;10164;10;87;74)
(128, 58457;2001;1;3;9;10187;13;90;83)
```

.....

In the KEY/VALUE list, the KEY value means offset address which is not belong to statistical information, and we can ignore it. Map function only need to extract the year and temperature, we use boldface type for them.

The results of extracting are:

```
(2000, 78)
(2000, 71)
(2001, 66)
(2001, 87)
(2001, 90)
```

After the output of map function is sent to reduce function, reduce function need to sort and group the output result KEY/VALUE by KEY value. The input data of reduce function should be:

```
(2000, [78,71])
(2001, [66,87,90])
```

From the input data of reduce function we can see that there are many temperature data belong to the year it's in. Reduce function only need to traversal the VALUE dataset and find out the maximum value as the final result.

```
(2000, 78)
(2001, 90)
```

If we need to calculate the other property like minimum temperature, we only need to modify the reduce function and find the minimum value in VALUE dataset.

4.3. Conclusion and Performance Evaluation

In order to test the efficiency of calculation, we use different size of dataset to execute the experiment in single PC and Hadoop cluster respectively. The sizes of five datasets are list in [Table 3](#).

The operation result is shown in [Figure 4](#), which we can see that when the size of dataset is small, the runtime of Hadoop will be longer than single PC. The cause lies in the fact that Hadoop need to initialize the dataset which is cost lot of time. On the other hand, the file system of Hadoop is fragmented; the default size of data block is 64 M. If data file is less than 64 M, Hadoop will process it as one block, which means small file will waste the storage space, and the query speed is slower than big file. As a result, when the dataset is bigger than default data block size, the advantages of Hadoop cluster will be more obvious.

Table 3. The test suit of big data.

Dataset name	Dataset size
1.txt	1.5 MB
2.txt	536.2 MB
3.txt	4.7 GB
4.txt	10.3 GB
5.txt	21.5 GB

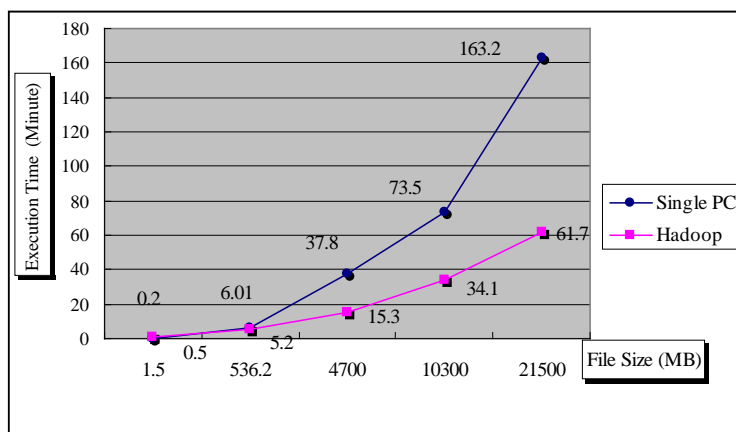


Figure 4. Execution time of calculations.

5. Precautions of Natural Risk

Natural environment is an important factor of agricultural production, so we can't eliminate the risk of rural microcredit by keep away the natural environment, but we can use it to evade and response the risk.

Weather is an external factor without effective control; we can use scientific predictions and forecast to reduce the negative influences. If farmer learn the real time information of weather in advance, then they can take special precautions to minimize the losses. When the losses cause by weather is unavoidable, we can compensate the losses in two ways. On the one hand, we can use the advanced agricultural techniques to improve soil fertility and increase the output, so we can use the maximum production under the favorable conditions to make up for the shortage under the unfavorable conditions. On the other hand, we can introduce advanced technologies and tools to improve the survival rate of crops, such as plastic sheeting for vegetables and digital equipment.

The techniques described above can't reduce the natural risk of rural microcredit fully guarantee, there is still some chance that we can't predict the natural disaster. The most appropriate way is agricultural insurance which use the advantage of large sample to distract large risk of personal. Agricultural insurance has two types of factors, the external factors (the lack of necessary legislative, economic, and administrative supports) and the internal factor (the backward operating technique in agricultural insurance).

Beyond that, when loan facility chooses the farmer as borrower, they should divide the farmers with same properties into different group, and associate them to the farmer with different properties. This operation can minimize the total risk cause by natural environment.

6. Conclusions

By analyzing the present situation of rural microcredit, we find out that there are many defects in China, and the most crucial and most sensitive issue is risk and risk management, especially natural risk. At the same time, risk management will directly affect the other factor impact on rural microcredit. Risk control and management should be the first and foremost of all things to be considered.

In order to prevent natural environment risk, we recommend that we should make full use of new high level technology and agricultural insurance. Since the natural risk has so many influencing factors, we propose an efficient method to analysis the natural factor. Compared with the single PC, the performance of calculation is improved by Hadoop cluster. Other than that, we can use the other derivative financial instrument to hedge and reduce the losses cause by natural risk.

Acknowledgements

This work was supported by Soft Science Research Project of Zhejiang Province (2014C35060), Department of Education of Zhejiang Province's Research Project (Y201430369), Zhejiang Provincial Natural Science Foundation of China (LQ13D010001).

References

- [1] Barry, J.J. (2012) Microfinance, the Market and Political Development in the Interact Age. *Third World Quarterly*, **33**.
- [2] Arora, S. and Meenu (2012) The Banking Sector Intervention in the Microfinance World: A Study of Bankers' Perception and Outreach to Rural Microfinance in India with Special Reference to the State of Punjab. *Development in Practice*, **22**, 991-1005. <http://dx.doi.org/10.1080/09614524.2012.696092>
- [3] Wagner, W. (2010) Loan Market Competition and Bank Risk-Taking. *Journal of Financial Services Research*, **37**, 71-81. <http://dx.doi.org/10.1007/s10693-009-0073-8>
- [4] Altunbas, Y., Gambacorta, L. and Marques-Ibanez, D. (2010) Bank Risk and Monetary Policy. *Journal of Financial Stability*, **6**, 121-129. <http://dx.doi.org/10.1016/j.jfs.2009.07.001>
- [5] Kauffman, R.J. and Riggins, F.J. (2012) Information and Communication Technology and the Sustainability of Microfinance. *Electronic Commerce Research and Applications*, **11**, 450-468. <http://dx.doi.org/10.1016/j.elerap.2012.03.001>
- [6] Wang, R.Y. (2012) Reliability Improvement of Fluorescent Lamp Using Grey Forecasting Model. *Microelectronics Reliability*, **42**, 127-134.
- [7] Li, A.Y. and Zhang, M. (2011) Research about Loan for Small and Middle-Sized Enterprises Based on Perspective of Information Economics. *China Information Times*, **9**, 43-48.
- [8] Ye, Q. (2011) Policy Selection and Development Path of Current Rural Microcredit Based on the View of Rural Financial Extension. *Foreign Investment in China*, **20**.
- [9] Smith, B.L. (2011) Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting Transportation Research. *Emerging Technologies*, **10**, 303-312. [http://dx.doi.org/10.1016/S0968-090X\(02\)00009-8](http://dx.doi.org/10.1016/S0968-090X(02)00009-8)
- [10] Chen, L.W. (2008) Applied Research for Decision Tree Algorithm in Rural Microcredit. *Computer Engineering and Applications*, **31**.
- [11] Sun, T. and Zhang, G.J. (2010) Risk Evaluation for Commercial Bank Based on Genetic Algorithm. *Journal of Qingdao University (Natural Science Edition)*, **2**.
- [12] Zhao, J.X. and Du, Z.P. (2009) Research on Credit Risk Assessment Model Based on Hybrid Neural Network and Decision Tree Algorithm. *Journal of Beijing Institute of Technology (Social Sciences Edition)*, **1**.
- [13] Jacobson, T. and Roszbach, K. (2012) Bank Lending Policy, Credit Scoring and Value-at-Risk. *Journal of Banking & Finance*, **27**, 615-633. [http://dx.doi.org/10.1016/S0378-4266\(01\)00254-0](http://dx.doi.org/10.1016/S0378-4266(01)00254-0)