

Port Customer Credit Risk Prediction Based on Internal and External Information Fusion

Guo Yi^{1,*}, Huang Lei¹ and Liu Ziqiang²

¹School of Economics and Management, Beijing Jiaotong University, Beijing, 100044, P.R. China; ²Electronic Engineering Institution of PLA, Anhui, Hefei, 230037, P.R. China

Abstract: Concerning the information asymmetry, subjectivity and other problems in the prediction of port customer credit risk level, a port customer credit risk evaluation system was designed, which provided the quantification base for port customer credit risk prediction. External-internal information fusion model was proposed for port customer credit risk prediction. Customers' internal history information and relevant external dynamic information were used as information sources. The prediction model fuses external and internal information to predict customers' credit risk level, therefore assisted port enterprises to estimate customers' credit risks level more accurately. Top 100 coal industry customers have been selected in experiment, and their operation data were extracted from Guangzhou Port group operation and management system. Relevant external information has also been crawled in experiments correspondingly. Experiment results show that port customer credit risk prediction model based on external-internal information fusion has a better prediction accuracy than non-information-fusion model.

Keywords: Back propagation neural network, credit risk prediction, information fusion, port.

1. INTRODUCTION

Port is an important hub of national comprehensive transportation system and international logistics. With the enhancing of the trend of global trading and the rapid development of Chinese economy, ports service more customers correspondingly. For maintaining major customers and long-term customers, ports often take the postpaid way as convenience. Although this credit-dependence settlement provides convenience for business activities, it increase the potential risk in daily operation. As a result of the lack of honesty and imperfect credit system, customers will default operation charges when they are suffering poor financial status. This situation often involves hundreds of thousands Yuan or even more debts, which seriously affects the normal operation of the port.

Owing to the serious information asymmetry between ports and customers, it's hard to find potential customers credit risk from internal data and information. Therefore, predicting the customer credit risk more accurate and earlier become ports' great challenges. The traditional method depends on personal judgments with manager's experience and social sources. They judge customers' management status and then estimate the potential risks of customers.

Moreover, some scholars also explored in port customer credit risk prediction. Mei [1] proposed a port customer credit rating evaluation model, which established dynamic customer credit files by handling customers' finance information. Chu [2] chose willingness to pay, payment history,

financial statements, financial strength and collateral as credit evaluation factors to evaluate customers' credit status. Wang [3] proposed a port customer classification system based on the actual port operation and classified port customers by using BP neural network, which made progress in quantitative research than before.

However, in the research of port customer credit risk prediction, most of the current literature have focused on building credit evaluation system or risk control methods selection, only a few literatures explored in quantitative research. Existing quantitative researches mainly refer to customers risk prediction methods in bank or insurance industries, which used customers' financial information to predict their credit risks. But unlike the banking and insurance industries, ports are at the disadvantageous position in the information asymmetry. It's difficult to ports to obtain comprehensive information of customers. So there has possibility that customers use unilateral information to mislead ports. Once the financial trouble occurred, for their own benefits, malicious information concealing or manipulation can hardly be avoided. Thus ports' predicting results also been misled by false information, which increases the potential operation risks. Moreover, port industry has its own characteristics, other non-financial information also has important impacts. Ports can't only take financial information to predict customers' credit risks, in order to avoid getting into another kind of unilateral prediction.

Using external information can reduce asymmetry of customers' information for ports, meanwhile, increase the accuracy and robustness of prediction of port customer credit risk. Accordingly, this paper firstly designed a port customer credit risk evaluation system, which quantitatively evaluated

the level of port customer credit. On this basis, a multisource external-internal information fusion model has been proposed, which predict port customer credit risk level by fusing internal business and operation data with external related news articles and industries indexes. The actual business and operation data extracted from Guangzhou Port group operation and management system, corresponding external news and industry indexes have been crawled and used in experiments as data source.

2. PORT CUSTOMER CREDIT EVALUATION SYSTEM

Different from the bank and insurance industries, port industry doesn't have a complete and systematic customer credit risk evaluation system to evaluate customer's credit risk level. It is often based on personal subjective judgment, which is not conducive to the risk control. For evaluating the credit level of port customers, establishing the basis of port customer credit risk prediction, the port customer credit evaluation system should be designed firstly.

The establishment of port customer credit evaluation system should base on actual situation of the port industry. Based on objective quantification of port customer credit, port can find credit risk customers earlier. Thus port can reduce or eliminate the potential risk and possible loss. In order to ensure that the port customer credit evaluation system which can provide the objective and scientific basis for the evaluation, the evaluation indicators selection should follow the principles below:

1. **Scientific Principle:** Scientific is the most fundamental principles of evaluation system. The port customer credit evaluation system must follow the general law of port operation, which objectively reflects actual level of port customer credit.
2. **Purposeful Principle:** The establishment of evaluation system should be purposeful. The purposes of this evaluation system are evaluating the level of port customers' risk, finding potential or existing credit risk, thus helping port to avoid operation risk and possible loss.
3. **Systematic Principle:** All influential factors should be taken into account in an effective evaluation system. And all these factors should be integrated systematically. During designing the port customer credit evaluation system, constructor needs a comprehensive analysis of both customer and port operation's characteristics, then extracts key indicators that can affect customer credit evaluation effectively.
4. **Feasibility Principle:** Feasibility principle guarantees that the evaluation system can be applied. It should find balance between comprehensiveness and the objective conditions in practical application. Following this principle will ensure that all the data the evaluation system needs can be collected and processed.
5. **Pertinence Principle:** Pertinence is momentous principle in construction of industry evaluation system. There are extensive and significant differences between different industry and different enterprises. Excessive generalized

evaluation system is difficult to evaluate port customer credit accurately. Therefore, the port customer credit evaluation system needs select the representative factor according to the characteristics of port customer and operation.

Based on the above principles, financial situation, credit history, cooperative relationship and customer characteristics are selected as first grade indicators of port customer credit evaluation system. These 4 factors can cover the main concern of port customer credit risk evaluation. Furthermore, these 4 factors can be divided into 13 second grade indicators. The structure of port customer credit evaluation system is shown in Fig. (1).

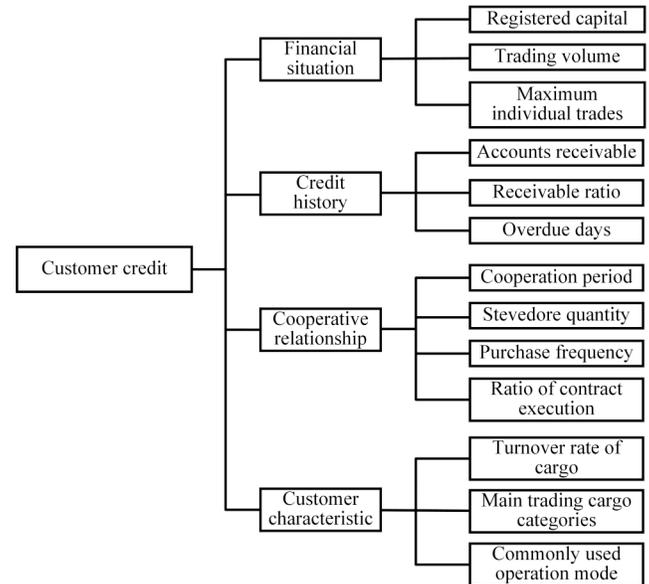


Fig. (1). Port customer credit risk evaluation system structure.

The main purposes of this evaluation system is evaluating the level of port customers' credit risk, so the target layer of port customer credit evaluation system is customer credit. The First grade indicators include financial situation, credit history, cooperative relationship and customer characteristic. Financial situation mainly describes customers' financial ability, which contains 3 second grade indicators: registered capital, trading volume and maximum individual trades. Credit history represents past credit situation of customer, which has been divided into accounts receivable, receivable ratio and overdue days. Cooperative relationship consists of cooperation period, stevedore quantity, purchase frequency and ratio of contract execution. Customer characteristic is subdivided into turnover rate of cargo, main trading cargo categories and commonly used operation mode.

For combining both quantity and quality methods thus helping decision-making, weighting each indicator's effect on port customer credit evaluation, an 'Analysis Hierarchy Process' (AHP) method has been used. AHP method proposed in 1970s [4], which is a multiple objective decision method with combination of qualitative and quantitative methods. By dividing the complex problem into some key factors in different hierarchies and constructing corresponding judgment matrix, AHP method can calculate the weights of indicators [5]. In this paper, 7 relevant domain experts and

Table 1. Indicators weight of port customer credit risk evaluation system.

First grade Indicator	Weight	Second Grade Indicator	Weight
Financial situation	0.3359	Registered capital	0.1233
		Trading volume	0.6394
		Maximum individual trades	0.2373
Credit history	0.4378	Accounts receivable	0.4068
		Receivable ratio	0.4012
		Overdue days	0.192
Cooperative relationship	0.1328	Cooperation period	0.0791
		Stevedore quantity	0.5734
		Purchase frequency	0.1208
		Ratio of contract execution	0.2268
Customer characteristic	0.0935	Turnover rate of cargo	0.5996
		Main trading cargo categories	0.2238
		Commonly used operation mode	0.1766

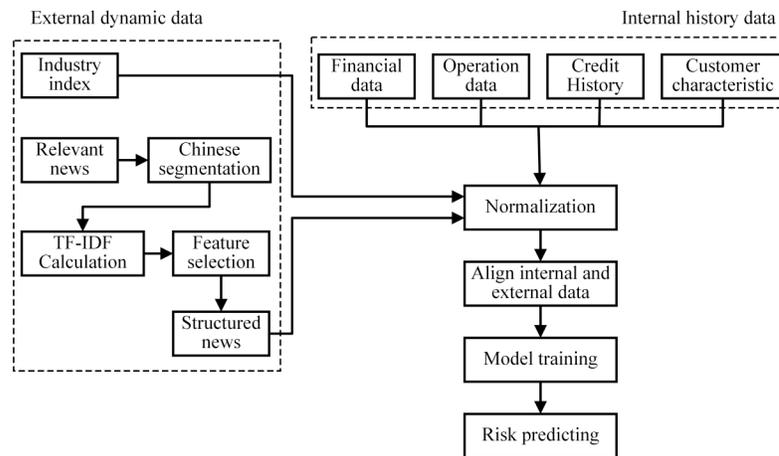


Fig. (2). Multisource external-internal information fusion procedure.

port managers were invited to give their quantitative opinion on each indicator’s weight. Finally, indicators’ weight in port customer credit evaluation system was calculated with the AHP algorithms, which is listed in Table 1. Both designing port customer credit evaluation system and weighting each indicator are bases of port customer credit evaluation.

3. MULTISOURCE EXTERNAL-INTERNAL INFORMATION FUSION MODEL

3.1. Multisource External-Internal Information Fusion

The procedure of multisource external-internal information fusion is shown in Fig. (2). All information are collected from two main sources: internal history data and external dynamic data. Internal history data include financial data, operation data, credit history data and customer characteristic data stored in management or operation systems. External

dynamic data consist of both relevant industries indexes and news articles.

3.1.1. Internal History Data Preprocessing

Internal history data are mainly extracted from port management system and operation system. In preprocessing, the data of customer’s registered capital, trading volume, accounts receivable, overdue days, cooperation period, stevedore quantity, purchase frequency, contract requirements, cargo balance, trading cargo categories and operation mode need to be extracted. Receivable ratio, ratio of contract execution and turnover rate of cargo are calculated, sorted chronologically and merged with former basic data correspondingly.

3.1.2. External Dynamic Data Preprocessing

External dynamic data consist of relevant industries indexes and news articles. Most relevant industries indexes are

structured data, which can be preprocessed as same as internal data.

But news articles are usually unstructured or semi-structured. It is difficult to fuse structured data and unstructured data directly. News articles should be transformed to structured data in preprocessing, thus structured news can fuse with structured data. The main steps of News articles processing are listed below:

1. Chinese segmentation. It's a common way that using vector space model (VSM) in text structuring, and VSM method is based on word or phrase. Unlike Latin has 'Space' as the natural separator, Chinese article needs to use segmentation method to divide text into a combination of words and phrases. Then the article can be converted to a vector form. Besides, there also has proper nouns in port and relevant industries domains that should be considered. Thus a domain relevant dictionary has been employed for segmentation.
2. Weighting. Term Frequency-Inverse Document Frequency (TF-IDF) proposed by Salton in 1973 [6]. Because of its simple and both high accuracy rate and recall rate, it's often used in text weighting [7]. After the weight calculation, each news can be transformed into a vector form.
3. Feature selection. Too many and dispersed features have no help for further predictions. On the contrary, it will increase computing cost and reduce generalization ability. Feature selection can effectively reduce the dimension of vector of news article, thus focusing on significant factors in prediction model. Selecting top 10% or fixed number of the highest frequency words and phrases [8, 9] are common methods. In this paper, top 100 of the highest frequency words or phrases have been selected as features.
4. Structured news. After feature selection, the original high dimensional vector has been transformed into a 100-dimension-vector by dimension reduction. Till then, the news text processing has completed, these news text vectors can fuse with other structured data.

3.1.3. Normalization

Both internal history data and external dynamic data are ready for training model and predicting after preprocessing. But considering the strong effects of incomparability of different dimensional and great disparity of magnitude in training and predicting, these data need to be normalized furthermore. We use linear algorithm to normalize all data by Equation (1). We denote the i -th feature as f_i , then f_{ik} represents the k -th value in i -th feature. $n(f_{ik})$ indicates the normalization value. The range of $n(f_{ik})$ after the normalization is [0, 1].

$$n(f_{ik}) = \frac{f_{ik} - \min(f_i)}{\max(f_i) - \min(f_i)} \quad (1)$$

3.1.4. Align Internal And External Data

Aligning internal and external data is based on the assumption that every emerging news article means that rele-

vant customers or industries have something changed. And these changes gain their influences in sequence. So it's necessary to align internal and external data before fusion that helping eliminating the effects from irrelevant information. Thus both model training and predicting are simplified, and which helps the model to meet the needs of practical application.

3.2. Port Customer Credit Risk Prediction Model

Estimation theory, uncertainty reasoning and intelligent fusion are basic methods in information fusion. Because of intelligent fusion has better adaptive ability and robustness, it has received extensive concern in both academic circles and industrial community. The BP neural network is often used in intelligent fusion for its better performance on non-linear fitting, parallel processing and stability.

In the traditional intelligent fusion process, structured data can be fused by BP neural network [10, 11]. Due to vectorization, news articles are transformed to structured data in vector form, so these news articles can fuse with structured data as the inputs of BP neural network.

The port customer credit risk prediction model is shown in Fig. (3), which N indicates the number of neurons. According to the design of port customer risk evaluation system and external data structure, there are 114 neurons in input layer, and 1 neurons in output layer in this model, which outputs the prediction result.

Hidden layer is the key to map inputs to outputs. Hidden layer neurons fit the mapping pattern by constantly adjusting their own weights. The number of hidden layer neurons will affect the result accuracy: less hidden layer neurons can't fit the mapping pattern completely, but more hidden layer neurons can also lead to over fitting. There is no complete appropriate method in hidden layer neurons design. On the basis of *Kolmogorov* theorem, any continuous function $f : [0, 1]^n \rightarrow R^m, y = f(x)$ can be simulated with a three layer BP neural networks. It determined that only on hidden layer need to be built. The number of hidden layer neurons can be calculated by empirical formula which Equation (2) shows. n_h indicates the number of hidden layer neurons, n_i indicates the number of input layer neurons, m indicates the number of output layer neurons and a is a integer constant between 1 and 10. We found that there has minimum error in our experiments when $a = 8$. So the number of hidden layer neurons is $n_h = \sqrt{n_i + m} + a = \sqrt{114 + 1} + 8 \approx 19$.

$$n_h = \sqrt{n_i + m} + a \quad (2)$$

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Data Sets

Both internal history data and external dynamic data serve as the data sets of the experiments.

1. Internal history data. The internal history data have been extracted from Guangzhou Port group operation and management system. The internal data set contains top 100 coal industry customers' history trading and opera-

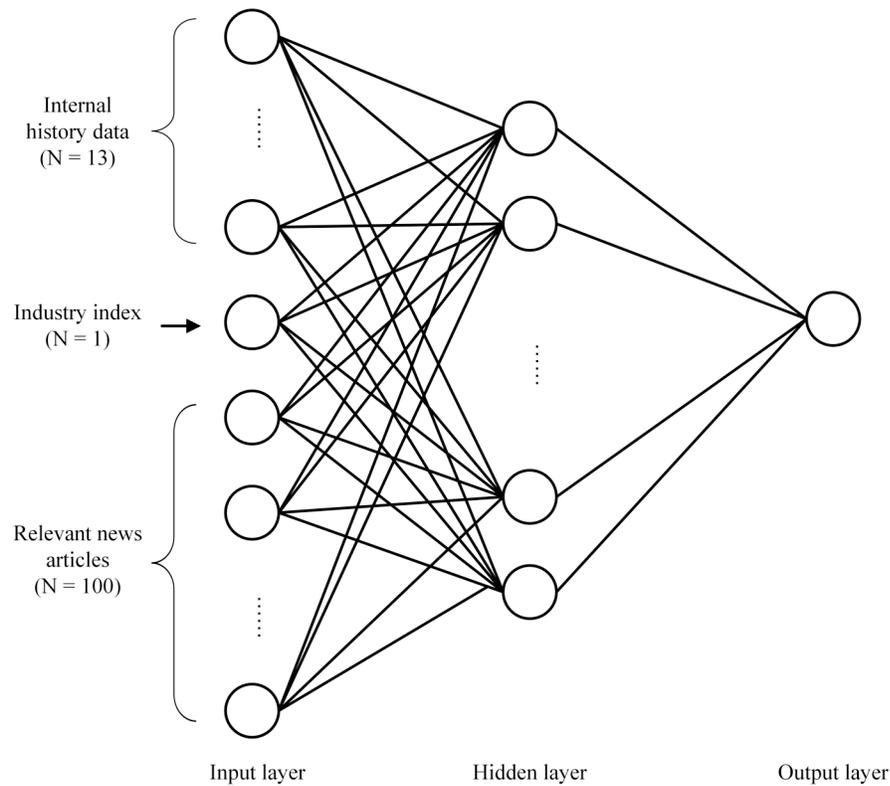


Fig. (3). Port customer credit risk prediction model based on external-internal information fusion.

tion data in Guangzhou Port between 2011 and 2012, which has 54,315 records and 2,500 merging records after preprocessing.

- External dynamic data. The coal industry indexes between 2011 and 2012 have been extracted from China Energy Net¹. More than 200 relevant news articles have been crawled and preprocessed as vectors. There are 23,800 records in total after aligning internal and external data.

In order to evaluate the performance of our system, the whole data set is split into June, 2012 and September 2012, respectively, forming two training/testing sets: 1) 18/6 months and 2) 21/3 months. The numbers of records in each set are listed in Table 2.

Table 2. Scale of training set/test set in data sets.

	Scale of Training Set	Scale of Test Set
Data set 1	16,600	7,200
Data set 2	20,400	3,400

4.2. Results and Finding

We measured the accuracy of model by Mean Squared Error with Equation (3) in our experiments.

$$MSE = \frac{1}{n} \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2 \quad (3)$$

¹ <http://www.china5e.com/energy-index/> (2015/01/10)

The MSE results are listed in Table 3. Compared with non-information fusion model which only use internal data or external data, the results show that internal-external information fusion model have smaller MSE in all data sets and better fitting performance. It is proved that internal-external information fusion model is more accurate in port customer credit risk prediction.

The prediction results of data set 2 have better accuracy than data set 1 correspondingly among all three models, which indicate more adequate data is helpful to train model more fitting thus improve the prediction accuracy. That means collecting more relevant information is contribute to predict customers' credit risk level in practical.

Besides, the experiment results also show that the model only used external data has worst prediction accuracy in both data sets. It indicates that relying on external information and neglecting internal information also has one-sidedness cause of the effects of different characteristics between industries. It also proved that port customer credit evaluation system is essential.

5. CONCLUSION

It is insufficient in quantitative research of port customer credit risk prediction. In this paper, we have designed port customer credit evaluation system as the basis of port customer credit risk prediction. For getting rid of subjectivity and one-sidedness in traditional manual way, we have proposed a port customer credit risk prediction model based on external-internal information fusion. We extracted actual operation and business data from Guangzhou Port group operation and management system, crawled relevant news

Table 3. Mse result of port customer credit risk prediction from different models.

	Information Fusion Based Model	External Information Based Model	Internal Information Based Model
Data set 1	7.46E-05	5.37E-04	2.39E-04
Data set 2	7.73E-07	4.23E-04	4.36E-06

articles and coal industry indexes in the same period. Results of experiments show that internal-external information fusion model has better accuracy in port customer credit prediction, which support port decision-making for perceiving and avoiding potential customer credit risk earlier.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This research is supported by the Natural Science Foundation of China under Grant Nos. 71132008

REFERENCES

- [1] Y. Mei, "Port income risk control method based on accounting management", *Finance and Accounting for Communications*, vol. 12, pp. 34-39, 2013.
- [2] S. X. Chu, "Prevention and control of financial risk in port enterprises", *China Ports*, vol. 3, pp. 34-35, 2006.
- [3] R. Wang, "Research on Port Freight Operational Risk Control Based on BP Neural Network", M.S. thesis, Beijing Jiaotong University, China, 2014.
- [4] W. Yoram, and T. L. Saaty, "Marketing applications of the analytic hierarchy process," *Management Sciences*, vol. 26, no. 7, pp. 641-658, 1980.
- [5] J. Y. Guo, Z. B. Zhang, and Q. Y. Sun, "Study and applications of analytic hierarchy process," *China Safety Science Journal*, vol. 18, no. 5, pp. 148-153, 2008.
- [6] G. Salton, and C. T. Yu, "On the construction of effective vocabularies for information retrieval", In: *Proceedings of the 1973 Meeting on Programming Languages and Information Retrieval*, New York, 1973, pp. 48-60.
- [7] C. Y. Shi, C. J. Xu, and X. J. Yang, "Research of TF-IDF algorithm", *Journal of Computer Applications*, vol. 29, no. B06, pp. 167-170, 2009.
- [8] R. Feldman, and S. James, "The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data," Cambridge: UK, 2007, pp. 57-62.
- [9] X. D. Li, X. D. Huang, and X. T. Deng, and S. F. Zhu, "Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information", *Neurocomputing*, vol. 142, pp. 228-238, 2014.
- [10] B. Xie, Z. C. Cai, and W. B. Song, "Study on the application of information fusion in fault diagnosis of aero engine", *Science and Technology of West China*, vol. 11, no. 2, pp. 34-35, 2012.
- [11] E. F. Nakamura, A. A. F. Loureiro, and A. C. Frery, "Information fusion for wireless sensor networks: methods, models, and classifications", *ACM Computing Surveys*, vol. 39, no. 3, p. 9, 2007.

Received: June 10, 2015

Revised: July 29, 2015

Accepted: August 15, 2015

© Yi et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.