Research on Credit Evaluation of Construction Market Entities Based on Conditional Random Field

Zou Zhenyu¹, Feng Jingchun¹,², Huang Zhecheng¹, Zhang Ke¹,², & Zou Lei¹

¹Business School, Hohai University, Nanjing 211100;
²Jiangsu Provincial Collaborative Innovation Center of World Water Valley and Water Ecological Civilization, Nanjing 211100

Abstract. This paper firstly judges the change of credit level of the credit subject from the dynamic perspective of "state-behavior". Then, according to the conditional random field theory, the data of credit behavior and credit status are taken as random variables in the conditional random field to construct multiple characteristic functions representing the relationship between the credit behavior and the credit status. And finally, the credit evaluation system based on the conditional random field is constructed. In summary, this paper constructs a credit evaluation framework of market entities from the credit evaluation system and conditional random field credit labeling, which can be used in the construction of market credit dynamic supervision for reference.

1. Introduction

Over the years, the state has further promoted the establishment of a credit system and issued a series of policies for the establishment of a credit system, thereby establishing a new type of credit-based supervision model. Strong credit supervision is an important means to improve the government’s modern governance capability and level, an important method of the socialist market economic system, an important measure to further promote the reform of "delegation, regulation, and service", and a key factor in optimizing the business market. In recent years, new technologies such as big data and artificial intelligence have developed rapidly. These new technologies and new methods can be used in the credit evaluation of construction market entities. The combination of new technologies and traditional credit evaluation methods of the construction market entity will have broad applications in the future prospect.

Domestic and foreign researches on credit management are mainly based on the following starting points. They study the construction of market credit evaluation index systems, the establishment of traditional theory-based construction market credit evaluation models, and the construction market credit evaluation models using machine learning methods. Owens JB[1] divided a credit subject into multiple modules according to the relevant theory of the control system, conducted a separate credit evaluation for each module, and then sorted the evaluation results according to the importance of the modules, and finally obtained a credit evaluation sequence which provides decision support for decision makers. Jungtae Mun[2] analyzed the establishment of fuzzy trust evaluation model by enterprises in the manufacturing industry through goal orientation in cooperation. Mehrdad Ashtiani[3] considered the hesitation of decision makers and the ambiguity of trust issues and proposed a new trust calculation model. Mehrdad Ashtiani[4] selected the corresponding index system by analyzing the fuzziness of known information, the background of decision-maker and the priority of target, and finally constructed a credit model of fuzzy multi-standard decision. Niklis[5] built a credit rating model.
through linear and nonlinear support vector agencies, and verified the validity of the model through case data. The above literature analysis and summary show that there are few evaluation methods that take into account the main credit components and behavior in the existing research, so the dynamic nature of credit evaluation is insufficient; besides, there are few data-driven methods, resulting in low accuracy in the credit evaluation of construction market entities. In view of this, the paper considers the characteristics of the credit data of construction market entities in the real situation, and from a data-driven perspective, with the help of machine learning methods, puts forward a conditional random field-based credit evaluation model framework for construction market entities.

2. The credit structure and measurement of market entities
This paper establishes the credit structure of the market entities, with a purpose to serve the credit evaluation. Credit itself is an unpredictable concept, but it has both internal basis (credit status) and external performance (credit behavior), both of which are measurable. Therefore, this paper evaluates the subject's credit level by means of credit status and credit behavior. When a credit subject enters the market, the government institutions concerned often evaluate the credit of the subject as per its basic credit elements, obviously this evaluation method is not accurate, but after a period of time, as the economic activity occurs in the construction market, the evaluation can gradually correct the previous credit evaluation through the credit behaviors, making the credit evaluation of the credit subject more and more accurate.

Figure 1 shows the change process of the credit level of the credit subject during a certain period. \( t_0 \) moment represents the initial credit level of a credit subject. When there is no credit behavior, its credit level is determined by its own credit components. \( t_1 \) to \( t_n \) indicate that, with the occurrence of economic activities of the credit subject in the construction market, the credit status will change accordingly with the occurrence of various credit behaviors of the credit subject based on the previous credit status.

3. Conditional random field
Conditional random field model (CRF) is a Markov random field with random variable \( Y \) output given a series of random variables \( X \). As an undirected graph model, it has the advantage of overcoming Label Bias Problem\(^{[6]}\). When solving the conditional probability \( P(Y|X) \), model parameters are learned by means of the input variable \( X \) set, namely the observation sequence, and the output variable \( Y \) set, namely the marker sequence. And the corresponding parameters of the model are obtained by using the maximum likelihood estimation or regularized maximum likelihood estimation. After the corresponding parameters are obtained, the \( X \) set is observed according to the input data, and the \( Y \) sequence set with the highest conditional probability is output.
Figure 2. Conditional random field

As shown in Figure 2, define $X$ set ($X_1, X_2, ..., X_n$), $Y$ set ($Y_1, Y_2, ..., Y_n$) to represent chain structure variables in pairs, in other words, in the case of the random variable $X$ set, the conditional probability distribution $P(Y|X)$ of the random variable $Y$ set can form a chain conditional random field. In the sequence labeling problem, the $X$ set represents the corresponding observation sequence, while the $Y$ set represents the corresponding state sequence.

When the random variable $y$ is under the condition of the random variable $x$, the parameterized form of the conditional probability of $P(y|x)$ is as follows:

$$P(y|x) = \frac{1}{Z(x)} \exp \left( \sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_{i+1}, x, i) \right)$$

The normalization factor $Z(x)$ in the formula is as follows:

$$Z(x) = \sum_y \exp \left( \sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_{i+1}, x, i) \right)$$

In order to further simplify the formula, the sum of transfer eigenfunctions and state eigenfunctions in the same position is considered to be written as a global eigenfunction, so as to simplify the parameterized form of conditional random field and turn it into the product form of a weight vector and an eigenvector.

Suppose there are $K$ feature functions, including $K_1$ transition feature functions and $K_2$ state feature functions, the transition feature $t_k$, state feature $s_l$ and their weights $\lambda_k$ and $\mu_l$ are represented by a unified symbol:

$$\omega_k = \begin{cases} \lambda_k, & k = 1, 2, ..., K_1 \\ \mu_l, & k = K_1 + l; l = 1, 2, ..., K_2 \end{cases}$$

Simplify the sequence $Y$ formula to the following formula:

$$P(y|x) = \frac{1}{Z(x)} \exp \left( \sum_{i,k} \omega_k f_k(y, x) \right)$$

$$Z(x) = \sum_y \exp \left( \sum_{k=1}^{K} w_k f_k(y, x) \right)$$
4. The framework design of the credit evaluation of construction market entities

The evaluation model framework is designed according to the credit evaluation objectives of the construction market subject discussed above, the credit structure and measurement of the construction market subject, and the key technologies involved in the credit evaluation model of the construction market subject (see Figure 3). The evaluation model framework mainly consists of the following three parts:

First, collect data of both the credit status and credit behavior of the construction market entities. On the website of the data source that builds the credit behavior and status data of market subjects, the data that can be used to build the model are searched, and the available data are collected. After that, the data are analyzed and processed according to the training requirements of the model. Finally, the data of credit behavior and status are quantified as required to facilitate the training of the model.

Next, use the data of credit status and behavior to train the conditional random field model. From the perspective of behavior state, this paper constructs a credit evaluation model based on conditional random field, that is, the process of calculating the weight vector $\omega$ under the condition of inputting random variable credit behavior data $X$ set and outputting random variable credit state data $Y$ set.

Finally, the accuracy of the model is verified by experiments. The credit behavior data $X$ as the test set is imported into the established credit evaluation model of the conditional random field, and the prediction accuracy is analyzed by comparing the credit status data $Y$ output after the model evaluation with the actual credit status data $Y$ of the training set, and the corresponding conclusions are drawn by comparing with other methods.

5. Conclusions and prospects

This paper studies and analyzes the theory of credit evaluation of construction market entities. Firstly, it constructs the credit structure of construction market entities; secondly, the theory of conditional random field is studied, and a credit evaluation method based on conditional random field is proposed, aiming at the nature of state transfer of credit data time series. Finally, the paper designs the
framework of the credit evaluation model of construction market entities, which lays a foundation for the follow-up research.

In future studies, this model can be extended from the following aspects:

First, consider whether it is possible to dig out other behavioral characteristics from the credit subject in addition to the credit behavior, and further improve the accuracy of the model. For example, from the perspective of business logic, finance, financing, divestment, and other behaviors, this paper will analyze the possibility of applying these behavior characteristics to the model.

Second, consider whether the text recognition technology can be integrated with the model. By using text recognition technology, credit behavior can be classified and judged better, and the time-consuming and labor-consuming labeling stage can be eliminated.

References

[1] Owens J B, Firestone J, Hawthorne E T, et al. Operational risk assessment and control: U.S. Patent 8,260,654[P]. 2012-9-4.

[2] Jungtae Mun, Moonsoo Shin, Kyunghuy Lee, Mooyoung Jung. Manufacturing enterprise collaboration based on a goal-oriented fuzzy trust evaluation model in a virtual enterprise[J]. Computers & Industrial Engineering, 2008, 56(3).

[3] Mehrdad Ashtiani, Mohammad Abdollahi Azgomi. A hesitant fuzzy model of computational trust considering hesitancy, vagueness and uncertainty[J]. Applied Soft Computing, 2016, 42.

[4] Mehrdad Ashtiani, Mohammad Abdollahi Azgomi. Trust modeling based on a combination of fuzzy analytic hierarchy process and fuzzy VIKOR[J]. Soft Computing, 2016, 20(1).

[5] Dimitrios Niklis, Michael Doumpos, Constantin Zopounidis. Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines[J]. Applied Mathematics and Computation, 2014, 234.

[6] Lafferty J, Mccallum A, Pereira F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[C]. 2001.