2016 3rd International Conference on Economics and Management (ICEM 2016)
ISBN: 978-1-60595-368-7

Research on Demand Forecast of Project Materials Based on Classifier

Jun-jun GAO[1,a], Peng-peng TIAN[1,b,*] Fang-hua HONG[2,*]

[1]Sydney Institute of Language and Commerce, Shanghai University, Shanghai, China

[2]State Grid Corporation of China SGCC Shanghai branch, Shanghai, China

[a]gaojunjun@shu.edu.cn, [b]Peter_8912@163.com

*Corresponding author

Keywords: Multi-project Management, Project Materials Demand, Classifier, Demand Forecasting.

Abstract. Based on the phenomenon that companies often encounter difficult multi-project management with a variety of project materials, this paper put forward a project material demand forecasting model on the basis of classifier. First, C4.5 decision tree classification algorithm was applied to training set data, which generated classifier. Then, the classifier was used to sort out test set data. Finally, demand forecasting about the post-classified project materials was conducted. The empirical results showed that this classification method has improved the accuracy of project materials demand plan.

1. Introduction

Projects are important boosters of company wealth and social development. Multi-project management becomes universal in the process of company operation. However, parallel project implementations often trigger competition and conflicts of money, time, manpower and other resources.

Domestic and foreign project-related research mainly focuses on project management. Some concepts and methods has been put forwarded, such as project management concept [1], concurrent engineering [2], critical chain method [3,4]. Existed studies about engineering materials procurement concentrated on the supply chain management for engineering project procurement [5,6] and procurement optimization decision [7].

But there is a sorely lack of literature on project material demand forecast management. One of the significant reasons is that project material demand is not as stable as ordinary material demand. Large scale engineering project has complexity characteristic, due to large scale, long project cycle, many construction units, great influenced by natural environment and so on. The complexity results in fluctuating, discontinuous and disorder demand. Quantitative forecasting method will turn out poor effect when applied to unstable demand data. Therefore, this paper put forward demand forecast of project materials based on classifier (DFPMBC).

2. Theory Introduction

2.1 Project materials classification

For the traditional material classification, there are different classification criteria based on different demand requirements. In this paper, project materials are classified according to two demand characteristics including project relevance and demand forecast feasibility. All materials are divided into 4 kinds: project weak correlation and feasible prediction, project weak correlation and infeasible prediction, project strong correlation and feasible prediction, project strong correlation and infeasible prediction.

2.2 Demand forecasting methods

1) Quantitative demand forecasting method

It is also named mathematical analysis method, including Holt Winter methods, exponential smoothing and curve fitting model. The quantitative methods can be divided into continuous demand forecasting and the intermittent demand forecasting. They can be applied to materials with steady and regular demand, refer to project weak correlation and feasible prediction type.

2) Qualitative demand forecasting method

It will combine all views of experts as the main forecasting basis. These experts, who are familiar with business knowledge, make judgments about the future development based on rich personal experience and comprehensive analysis ability. Delphi method, leading indicator method and situation prediction method are common qualitative demand forecasting methods.

3) Combining method of quantitative and qualitative

Quantitative and qualitative demand forecasting methods are applied at the same time, one is to get basic predictive value, and the other is for adjusting predictive value, in order to achieve better prediction results and improve prediction reliability. It can be applied to project materials of project strong correlation and feasible prediction type.

3. Demand Forecast Model of Project Materials

This paper put forward a demand forecast of project materials based on classifier (DFPMBC). DFPMBC model consists of four stages: (1) found a combined forecasting model; (2) classify project materials based on prediction results and the relevance to projects; (3) train classifier and apply it to classify test set; (4) demand forecasting adjustment according to classification.

3.1 The combined demand forecasting model foundation

The combined forecasting model means applying a variety of forecasting methods at the same time. Existing studies have proved that the prediction accuracy of combined forecasting is much better than single forecasting, especially for project materials. This paper used the following three forecasting models: curve fitting model, Holt-Winter model and Croston's model. Firstly, each material is forecasted by the above methods. Secondly, the forecasting results of forecasting methods are weighted. Finally, calculate the predictive value of the combined model based on the weighting coefficient. The calculation process of the weights involves two steps:

First step: preliminary selection

Firstly, Curve Fitting model, Holt-Winter model and Croston's model are used to forecast one year's amount, then absolute percentage errors of observation and forecast values can be obtained. If the absolute percentage errors are equal to or less than 50%, the method can be accepted, or it will be taken as rejection.

Second step: the forecasting methods are weighted

After the first step, it will turn out to be 4 kinds of results:

(1) There is no suitable method; then Holt-Winter model will be selected and the weighting factor is 1.

(2) There is only one feasible method; then the weighting factor of this method is 1.

(3) Two methods are feasible; then apply the least squares method to estimate the weighting factor of these two methods.

(4) All of them are feasible; weighting factor confirmation is same with the third result.

3.2 Project materials classification

In this paper, project materials are divided into 4 kinds based on project relevance and feasibility of demand forecasting.

(1) Project weak correlation and feasible prediction (C1): materials demands are hardly affected by the project changes and demand characteristics are continuous, stable and regular. Using the quantitative forecast model can get satisfactory results.

(2) Project weak correlation and infeasible prediction (C2): the project changes have only modest effects on the demand of this kind of materials, but the demand is intermittent and unregularly. Not suitable for the demand forecast.

(3) Project strong correlation and feasible prediction (C3): the project changes have a great of effects on the demand of this kind of materials. Even though the demand of these materials are continuously, stably and not always regularity, the qualitative and quantitative method of combining forecasts should be used to get satisfactory result.

(4) Project strong correlation and feasible prediction (C4): the demand of this kind of materials that vary by the changes of projects, so the demand characteristics of this kind of project materials are variable, with big fluctuation. Quantificational forecast model doesn't make sense, the qualitative method could be used to get the quantity demanded of a good per period of time.

This paper put forward a project-type demand classification method and combined forecasting model: Firstly, we calculate the proportion α_i of materials using on projects, $\alpha_i = \frac{S_i'}{S_i}$, in which S_i' represent the amount of material x_i used in the projects, S_i represent the full amount of material x_i. Secondly, the sample is analyzed by the combined forecasting model. Thirdly, MAPE criteria and α_i are used in order to make all samples are divided into 4 kinds of project material.

3.3 Training classifiers and classify testing set

There are three steps in this process. Firstly, project materials are randomly divided into training set and test set. Secondly, Select demand characteristics as variable properties to train classifiers. Thirdly, classify test samples with the classifier.

This paper applied the C4.5 algorithm decision tree, which takes information gain ratio $(\mathrm{ratio}(f_n))$ as attribute choice criterion. Suppose that the sample T_q is divided into p different classes T_{qx} (x = 1, 2,.., p).

$$\mathrm{ratio}(f_n) = \frac{\mathrm{gain}(f_n)}{\mathrm{infox}(f_n)} \tag{1}$$

$$\mathrm{gain}(f_n) = \mathrm{info}(T_q) - \mathrm{infof}_n(T_q) \tag{2}$$

$$info(T_q) = -\sum_{p=1}^{k} \frac{|(C_p,\ T_q)|}{|T_q|} log_2 \frac{|P(C_p,\ T_q)|}{|T_q|} \tag{3}$$

$$infof_n(T_q) = -\sum_{x=1}^{P} \frac{|T_{qx}|}{|T_q|} \cdot info(T_q) \tag{4}$$

$$infox(f_n) = -\sum_{x=1}^{P} ((\frac{|T_{qx}|}{|T_q|}) \cdot log(\frac{|T_{qx}|}{|T_q|})) \tag{5}$$

$\mathrm{info}(T_q)$ is the average volume of information of the sample T_q, f_n is the property set of the sample

T_q, $gain(f_n)$ is the gain information of property set f_n, $infofn(T_q)$ is the average volume of information of the property set f_n.

The basic idea is that, the higher $ratio(f_n)$, the higher the weight of the f_n. After selecting the highest one as the root of the decision tree, a branching node will then be generated if the property set provides the highest $ratio(f_n)$. In this paper, test samples were classified with the classifier. To test and evaluate the performance of C4.5 Classifier, an experiment tests the classification effect of this method by a confusion matrix. A confusion matrix describes how many results were correctly classified and how many were incorrectly classified for each of the categories.

Table 1. Forecast Effect Evaluation Matrix.

Samples	No	Yes	Total
No	TN	FP	$TN+FP$
Yes	FN	TP	$FN+TP$
Total	$TN+FN$	$FP+TP$	$TN+FP+FN+TP$

This paper introduces 3 commonly used classification evaluation criteria.

1) $\text{Total accuracy} = \frac{TP+TN}{TN+FP+FN+TP}$;
2) $\text{Sensitive} = \frac{TP}{TP+FN}$;
3) $\text{Specificity} = \frac{TN}{TN+FP}$.

3.4 Demand forecasting adjustment based on classification

According to the classification, the predicted results can be revised to further improve their accuracy. Four types project materials.

Project weak correlation and feasible prediction. The predictive value of the combined model is as the final forecast value.

$$\check{y}_i = \omega_1 \cdot winters + \omega_2 \cdot ARIMA + \omega_3 \cdot Croston's \tag{6}$$

In the formula, $\omega_r (r = 1,2,3)$ are the weighted values of the forecasting model. $\check{y}_i$ is the predictive value of the combined model. According to the method of least squares (OLS), the optimal calculation formula of the weighted values was obtained.

$$\omega_r^{'} = \frac{\sum(\check{y} - \bar{y})^2}{\sum(y - \bar{y})^2} \tag{7}$$

$\bar{y}$ is the mean value of the observations. ω_r is generated after standardizing $\omega_r^{'}$.

Project weak correlation and infeasible prediction. This type of project materials are not suitable for demand forecast, only can be predicted by experience.

Project strong correlation and feasible prediction. Firstly, initial value is generated by using the combined model. Secondly, calculate ratio of the planned costs and forecasts θ_i, the final forecasts are $\check{y}_i$.

$$\check{y}_i = (\omega_1 \cdot winters + \omega_2 \cdot ARIMA + \omega_3 \cdot Croston) \cdot (\frac{1}{\theta_i} \cdot \alpha_i + (1 - \alpha_i)) \tag{8}$$

Project strong correlation and infeasible prediction. Calculate the annual demand of this type of project materials according the planned costs.

4. Empirical Analysis

In order to verify the validity of the forecasting model, this paper conducted an empirical analysis

based on demand data collected from ST Company. The data consists of actual consumption data and project data from 2011 to 2014. The performance of this model was analyzed by means of confusion matrix and the forecasting accuracy was evaluated by MAPE.

4.1 Description of data set

ST Company is a typical project management enterprise, so multi-projects' management is the core of production. Project material data was collected from this company. After that, firstly, data cleaning is vital to improve data quality of information system. Secondly, organize data into monthly data. Finally, analyzed demand characteristics, including standard deviation, mean value, variance coefficient, number of clients, the number of demanding months and time interval of requirements.

4.2 Experiment setup

Given many different measurement scales are employed to analyze the data set, there is a big difference among standard deviation, mean value, coefficient of variation, number of clients, the number of demanding months and time interval of requirements for different materials. So it is not suitable apply these demand characteristics directly. The paper focuses in the following four aspects. (1) Demand fluctuation: variance coefficient; (2) Demand dispersion: number of clients; (3) Demand continuity: time interval of requirements; (4) Demand frequency: the proportion of the number of demanding months

In order to improve the reliability and stability of the experiments, need to ensure the balance of the samples distribution. In this part, the random sampling method is used to divide to training set and testing set. The classification experiments have been carried out by using SPSS-clementer.

4.3 Results Analysis

As shown in Table 2: (1) The number of clients has more influence on the predictive accuracy of the project strong correlation materials than other demand characteristics, and the influence is positive, the more the number of clients, the easier to predict; (2) The predictive accuracy of the project weak correlation materials was influenced most by the variance coefficient, followed by time interval of requirements. The demand uncertainty is described by variance coefficient, so the accuracy of the project weak correlation materials becomes worse with the increase in demand uncertainty.

Table 2. The Weight of Characteristic Vectors.

Weight	Variance coefficient	Number of clients	Time interval of requirements	Demand frequency
Project weak correlation	0.618	0.033	0.273	0.075
Project strong correlation	0.159	0.618	0.218	0.007

Table 3 illustrate the classification results: (1) The results are satisfied, with high accuracy more than 80%; (2) The accuracy of the project weak correlation materials is slightly higher than the other one.

Table 3. Effect Evaluation of Forecasting Model.

Effect evaluation	Total accuracy	Sensitive	Specificity
Project weak correlation	85.3%	82.6%	88.9%
Project strong correlation	85.5%	88.1%	85.2%

Table 4 illustrate the demand forecasting results: (1) The quantitative forecasting MAPE of C1 is less than 10%, which means forecast accuracy are more than 90%; (2) Both quantitative forecasting MAPE and qualitative forecasting MAPE of C2 are very high, so the forecast accuracy are very low. (3) The

adjusting forecasting MAPE of C3 is lower than the other two MAPE. (4) The qualitative forecasting MAPE of C4 is lower than the quantitative forecasting MAPE. In a word, the empirical results are aligned with the model analysis in this paper

Table 4. Forecasting Precision Analysis Matrix.

SKU	Classification	Quantitative forecasting MAPE	Qualitative forecasting MAPE	Adjusting forecasting MAPE
10kV AC transformers, 400kVA	C1	5%	-	5%
10kV AC transformers, 630kVA	C1	7%	-	7%
10kV cubicle-type substation, 630kVA	C2	78%	92%	-
20kV AC transformers, 800kVA	C3	52%	32%	21%
20kV AC transformers, 1000kVA	C3	63%	26%	18%
20kV cubicle-type substation, 400kVA	C4	121%	41%	41%

5 Conclusion

This paper put forward a project-type demand forecasting research model on the basis of classifier. This method is simple, convenient, available, feasible and also applicable to multi-project management. Empirical analyzing results had proven that not only applying C4.5 to classify project materials is feasible, but also demand forecasting based on classification is also feasible for project materials. Besides, as the results showed, the model sharply raised accuracy of project-type demand forecasting, which has strong practical application value and theory referential value. In next research stage, multi-classifiers fusion technology will be applied to project materials classification to improve classification performance.

Reference

[1] Tan Y. A research on Sort-Order Problems of enterprise Multi-Project Management [J]. Journal of Industrial Engineering & Engineering Management, 2005.

[2] Fine C H, Golany B, Naseraldin H. Modeling tradeoffs in three-dimensional concurrent engineering: a goal programming approach [J]. Journal of Operations Management, 2011, 106 (12) 2156-66.

[3] Yeo K T, Ning J H. Integrating supply chain and critical chain concepts in engineer-procure-construct (EPC) projects [J]. International Journal of Project Management, 2002, 20 (4) 253-262.

[4] Yeo K T, Ning J H. Managing uncertainty in major equipment procurement in engineering projects [J]. European Journal of Operational Research, 2006, 171 (1) 123-134.

[5] Yildirim H. Piecewise procurement of a large-scale project [J]. International Journal of Industrial Organization, 2004, 22 (s8-9) 1349-1375.

[6] Li F X, Xu C X. Research on optimal strategy of engineering materials procurement based on supply chain financing [C] / Management Science and Engineering (ICMSE), 2010 International Conference on. IEEE, 2010: 1688-1693.

[7] An Z Y, Liu Y C. The Price Decision for Large-Scale Project's Procurement with Stage-Demand Uncertainty [J]. Chinese Journal of Management Science, 2013: 1085-1094.