http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (21): 6164-6168, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.6164.6168

Projects Bidding Decision Risk Analysis Based on Multi-factor Clustering Analysis

Chao Yang
School of Economics and Management, Beihang University, Beijing, 100191, China

Abstract: Project risk management is a very complex work which the pre-bid decision-making risk is one of the important aspects among. This study is based on the consideration the facts that the general contractor risk management of project is a strong empirical work as well as the indicators described the project characteristic are mostly qualitative indicators. It adjusts the usual quantification mode of qualitative indicators did in the past. It introduced cluster analysis technique, established the analysis model based on the multi-objectives function optimum factors and hybrid scale stepwise clustering. Additionally, through the approach of established projects sample risk factor data mining, this study proposed scientific theory evidence for the decision making of general contractor construction project bidding.

Key words: Project, cluster analysis techniques, sample, bidding decision

RISK ANALYSIS METHODS RESEARCH

International project risk management is a very complex work and risk analysis before bidding decision is one of the important links which provides a direct reference in bidding for the project. Currently used methods of risk analysis include: investigation and expert scoring method, Monte Carlo simulation, PERT (program evaluation and review technique), sensitivity analysis, analytic hierarchy process, fuzzy mathematics, multi-objective decision tree model, influence diagram, utility theory. Some of these methods focus on qualitative analysis, such as the investigation and expert scoring method and AHP; some focus on quantitative analysis, such as the fuzzy mathematics method and sensitivity analysis; also some focus on comprehensive analysis, such as Monte Carlo simulation method.

Currently the methods on the project risk assessment is focus on qualitative research, for there are still many difficulties on the risk of quantitative research. With the rapid development of computer technology and extensive use of calculation methods and some mature, people have paid more attention on the combination methods of qualitative and quantitative of risk estimates and has been showing good prospects. Scholars have also done a lot of research work in theoretical circles of project risk estimation methods, such as the project schedule risk research is more mature and has been in successful used of projects in the example, but it is generally accepted and used in projects.

From an engineering point of view of practical experience, according to the characteristics of the project characteristic factors, seek statistical law in project which has been built and then use analogies, to dig until construction project risk management approach is an effective method and resulting conclusions have reference value. With the increasingly fierce market competition, contractors continue to accumulate already built project risk management information and as a follow-up decision-making reference for project bids are becoming contractors to improve their management level and enhance the competitiveness of an important measure .Scientific analysis on project risk management information which has been built, has become the key to the implementation of this measure. Therefore, using the method of clustering analysis is a viable risk estimates and scientific approach.

The cluster analysis techniques to project risk analysis has been reported in the literature. Literature (Chi *et al*., 2001; Jiang, 2003)applied clustering algorithm to analyze and evaluate the risk of various projects, including the literature (Wang and Lu, 2005; Zhang *et al*., 2004). Application equivalence relation clustering research project to determine the risk factors weighting coefficients, calculated to provide parameters for the analytic hierarchy process; literature (Wang and Lu, 2005) is the application of equivalence relation clustering (commercial credit) risk analysis subjective opinions of experts, in order to achieve the classification of evaluation experts; literature (Jia and Liu, 2001; Wang and Lu, 2005; Ye *et al*., 2005; Fang and Xiong, 2005) use the hierarchical clustering method and the equivalence relation clustering method Perform variable clustering analysis, in order to achieve the purpose of object classification (Jia and Pan, 2003).

As can be seen from these research literature, cluster analysis in risk analysis research has made some achievements, but only as an auxiliary tool, the main purpose is to provide quantitative calculation of qualitative analysis indicators, but more for commercial banks, credit risk analysis and market prospects problems in the construction general contracting project risk analysis of precedents on the issue yet.

CONSTRUCTION RISK CLUSTERING MODEL

Combination of the above risk factors identified, the use of cluster analysis techniques, build a clustering model. Specific steps are as follows:

defines the measure of the degree of similarity between samples: Flip is for nominal scale samples from a similarity measure defined, it uses the number of samples to define the attributes of the flip nominal scale sample "convergence" price, that is, two samples of nominal scale how many times the corresponding properties flipped into the same , defined as:

$$|x_i\text{-}x_j| = D(x_i\text{-}x_j) = \text{Times of turn until } x_I = x_j \quad (1)$$

where, x_I, x_j is the nominal scale sample, its properties x_{ik}, x_{jk} $\in$N, N is a natural number set, k = 1, 2, ..., n. Flip distance defines the measure of the degree of similarity between the samples, it can be seen from the above analysis, similarity measure is determined, can be based on the properties of the sample matrix M (X) to calculate the distance of the sample to be classified (or matching coefficient of similarity coefficient) matrix D (X)

$$D(X)=\begin{bmatrix} d_{11} & d_{11} & d_{1n} \\ & d_{22} & d_{2n} \\ & \cdots & \cdots \\ & & d_{mm} \end{bmatrix} \quad (2)$$

where in, dij (i, j = 1, 2, ..., m) sample xi and xj represents the distance between; This is a real symmetric matrix of m×m, only the calculation or lower triangular portion can be.

To establish the sample project database: Sample database creation project has been built, the system has been built to analyze and summarize the success of project risk management, risk analysis experience and effectively transplanted to be built in the project's risk management practices to be realistic. Build the sample project database has been built projects accumulated data for subsequent projects to be built to support risk management is a foundation work, is also a risk clustering analysis. To this end, as a contractor you first need to strengthen data management, the establishment of enterprise information system itself and promote the construction of enterprise information.

Sample data extraction project properties: Describe all the characteristics of a project involves a lot of data needs, if not all of these data analysis to extract the characteristics used to describe the sample project, is bound to increase the difficulty of data processing and even lead to adverse consequences blinders. Therefore, select some typical data, the characteristics as described sample project is necessary.

The basic characteristics of the project data refers to projects that reflect the basic situation, the most typical and most essential characteristics of data; these data can depict the whole picture of the whole process of construction projects, irreplaceable, using these data clustering, can clearly reflects the differences between the categories. Reflecting the general contracting projects characterization data can be divided into two categories, one that can be measured with a scale interval data, such as building area, the estimated amount, etc. Another measure for the use of nominal scale data, such as type of structure, the height of the type , risk level, etc.

Q represents a project with a set of attributes, then $Q = Q_1 \cup Q_2$ where Q_1 is measured with a granularity attribute data: Q_2 for the nominal scale measurement data, so the properties of the project sample xi rearrange components of the vector and divided into two parts, $P(x_i) = (x_{i1}\ x_{i2} \ldots x_{ir}\ c_{ir+1} \ldots x_{in})$, where the r-th component $x_{ik} \in Q_1$ (k = 1, …, r)or the interval scale property, after the n-r-th component $x_{ik} \in Q_2$ (k=r+1,…,n)for the nominal scale properties.

$$x_{ik}^r = \frac{x_{ik} - \min_i x_{ik}}{\max_i x_{ik} - \min_i x_{ik}} \quad (3)$$
$$i = 1, 2, \ldots, m; k = 1, 2, \ldots, r$$

APPLICATION EXAMPLES

In order to verify the effectiveness of clustering analysis algorithm, combined with their actual work, the Company collected between 2007 and 2012 by way of the construction of the part of the general contracting civil works project archives and extract the characteristic data of these projects are presented in Table 1. Table 1 in the "Risk Level" one of the data is the company's decision-making expert in the project tender bids data analysis based on risk assessment is given and in the process of building the project is completed according to the actual situation has been corrected, the purpose of doing so is to accumulate data to the company for future reference for similar projects bidding decision.

Our aim is to apply clustering algorithm presented in the previous section, the data in Table 1 on cluster analysis and assuming that the last item to be tender, according to the clustering results determine its level of risk.

In Table 1, the project's "estimated amount", "building area", "story" is a description of the properties of interval scale, nominal scale with the rest of the properties described, is a typical mixed-scale clustering problem. First, the data in Table 1 pretreatment, namely, standardization and scale. The scale of nominal scale data are shown in Table 2; preprocessed data in Table 3.

To highlight the scale properties of a class effect cluster analysis, two types of tables can be calculated from the attributes given different weights. Determine the weight of a basic approach is based on two types of attributes in the total number of properties in the number of proportion, that is the number of attributes granularity metrics

- W1 = measured with a granularity attribute/total number of attributes

With a nominal scale measures the number of attributes

- W2 = attribute with a nominal scale measure/total number of attributes

Based on the actual situation in the appropriate adjustments. In this case w1 = 0.33, w2 = 0.67; with this cluster analysis to develop their own software, the data in Table 3 Clustering. Initial classification as Category 2:class I = {1#, 3#, 5#, 7#, 9#}; class II = {2#, 4#, 6#, 8#, 10#}, then, poly class total consideration G0 = 19.2480; through iteration, the final clustering results are obtained: class I ={1#, 3#, 4#, 8#, 10#} (10# project is assumed to be tender): class II = {7,# 9,#};class III = {2#, 5#, 6#}; clustering total consideration G0 = 9.2038.

Analyze the clustering results can be seen, class I projects have complex structures and foundation type, duration to seeking tight construction environment in general and construction estimates for the characteristics of large area, so a high level of risk of such projects.

Class II project is characterized by a more complex structure and foundation type, time is tight, the construction environment is poor, compared with estimates for big characteristic, so the level of risk of such projects is moderate; class III projects are brick structure, the height is small, based on simple, time in general, the construction environment and construction estimates for the small size, so the risk level is also low. This result is entirely consistent with the field experience.

As a test of the correctness of this clustering algorithm, the authors used statistical analysis package provides Matlab Toolbox cluster analysis function of the data in Fig. 1 cluster analysis. First, the need for granularity in Fig. 1. Conversion qualitative variables, the upcoming "estimate the amount of", "gross floor area" and "story" three interval scale variables into nominal scale variables: the conversion is based on expert experience "estimated amount", "building area" and the "story" on behalf of the three scale variables, the results are as follows:

Figure 2 shows the clustering result is the Eq. 1 defines the distance as the sample flip clustering similarity measure, this study presents the results of the clustering algorithm with clustering results exactly. First, the results of Fig. 2 clustering method using the system need to convert non-nominal scale properties, making each attribute clustering samples are nominal scale variables, can be used as a similarity measure from flipping; And

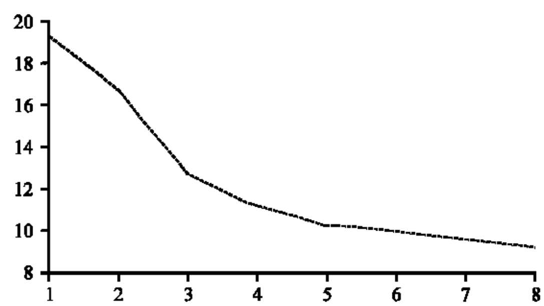

Fig. 1: Clustering total consideration of the process of change

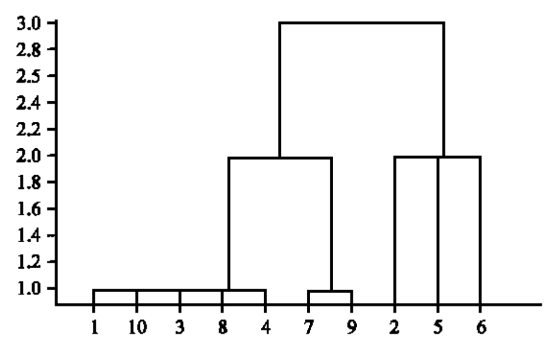

Fig. 2: Adoption(2-1)measure clustering pedigree chart

Table 1: Project characteristic data

Project name	Characteristic data									
	Estimatethe amount(yuan)	GFA	Structure type	Storey	Basic type	Decorative type	Schedule requirements	Ownersreu tation	Construction environment	Risk level
A high-rise residential	22223633	23960	Superstructure	22	Pile	Ordinary	Ordinary	Good	Ordinary	High
A residential	1754645	2437	Brick	6	Bar	Ordinary	Ordinary	Good	Good	Low
Teaching Building in a university	38034648	21072	Frame	10	Pile	Advanced	Urgent	Good	Ordinary	High
laboratory building in a university	9136662	6969	Frame	7	Pile	Advanced	Urgent	Good	Better	High
An office building	2220949	2994	Brick	7	Precision type	Advanced	Ordinary	Good	Good	Low
A student dormitory	3247128	5011	Brick	8	Raft	Ordinary	Urgent	Good	Good	Low
A residential building	2874475	3875	Reinforced concrete	7	Box	Ordinary	Ordinary	Good	Poor	Middle
A university library	19260973	13000	Frame	6	Pile	Advanced	Urgent	Good	Ordinary	Middle
A university teaching building	16703350	13102	Reinforced concrete	7	Pile	Ordinary	Ordinary	Good	Poor	Middle
A high-rise residential	22310641	19867	Superstructure	28	Pile	Ordinary	Urgent	Good	Ordinary	High

Table 2: The scale of nominal scale data

	Structure type				Basic type				Decorative type		Schedule requirements		Owner sreputation		Construction environment				Risk level		
Characteristic data	Brick	Reinforced concrete	Frame	Superstructure	Bar	Raft	Box	Pile	Ordinary	Advanced	Advanced	Urgent	Poor	Good	Poor	Ordinary	Better	Good	Low	Middle	High
Scale data	1	2	3	4	1	2	3	4	1	2	1	2	1	2	1	2	3	4	1	2	3

Table 3: Characteristic data preprocessing results

Sample No.	x_1 Estimates	x_2 GFA	x_3 Storey	x_4 Structure type	x_5 Basic type	x_6 Decorative type	x_7 Schedule	x_8 Owners reputation	x_9 Construction environment	x_{10} Risk level
1#	0.5642	1.0000	0.7273	4	4	1	1	2	2	2
2#	0.0000	0.0000	0.0000	1	1	1	1	2	3	1
3#	1.0000	0.8658	0.1818	3	4	1	2	2	2	3
4#	0.2035	0.2106	0.0455	3	4	2	2	2	3	3
5#	0.0129	0.0259	0.0455	1	3	2	1	2	4	1
6#	0.0411	0.1196	0.0909	1	2	1	2	2	4	1
7#	0.0309	0.0668	0.0455	2	3	1	1	2	1	2
8#	0.4825	0.4908	0.0000	3	4	2	2	2	2	2
9#	0.4120	0.4955	0.0455	2	4	1	1	2	1	3
10#	0.5666	0.8098	1.0000	4	4	1	2	2	2	3

Table 4: Standard name of the variable scaling

Estimates (M) yuan		GFA S		Storey	
Actual value	Scaling value	Actual value	Scaling value	Actual value	Scaling value
M≤5million	1	S≤5thousand square meters	1	Multilayer	1
5millions<M≤1ten million	2	5 thousand square meters<S≤1ten thousandsquare meters	2		
1ten millions<M≤2ten million	3	ten thousand square meters<S≤2ten thousand square meters	3	Small high	2
2ten millions<M≤3ten million	4	S>22ten thousand square meters	4	High level	3
M>3ten million	5				

this conversion right by using the scaling of the results (Table 4) is greatly dependent on the different scaling results will directly affect the clustering result which is very subjective. The author in the study, try to change the scale results in Table 4, to put coarse scale value, the resulting clustering pedigree chart on a different Fig. 2. This shows that the interval scale deliberately converted to nominal scale may sacrifice features implicit in the sample properties.

ARTICLE INNOVATION AND PENDING IMPROVEMENTS

Considering project general contracting empirical strong risk management, project characteristics describe indicators mostly qualitative indicators feature will change the past, the idea of quantifying qualitative indicators, the introduction of cluster analysis technique was established based on the multi-objective optimization factors, gradually mixing scale clustering model, mining projects have risk factors sample data for the general contractor for projects bidding to provide a scientific basis for decision making.

The proposed clustering algorithm in two types of attributes increases the weight on the scale, you can better reveal the law of the sample data. In addition, when the cluster sample size is large, the proposed algorithm clustering algorithm to solve the other problems of slow system.

REFERENCES

Chi, G.T., J. Hao, C. Xu and Z.Y. Zhu, 2001. Credit risk evaluation index weight clustering analysis. Syst. Eng. Theory Methodol. Appl., 10: 64-67.

Fang, X.M. and P. Xiong, 2005. Credit risk of commercial banks, monitoring and evaluation of the thinking of the Central. Univ. Finance Econ., 7: 14-18.

Jia, X.D. and D.H. Pan, 2003. Tech enterprise venture capital and investment risk evaluation. Inform. Control, 32: 203-208.

Jia, Y. and L.M. Liu, 2001. Current national debt risk and size analysis of Beijing. Univ. Technol. (Soc. Sci. Edn.), 1: 27-31.

Jiang, L.M., 2003. Based on fuzzy clustering credit risk evaluation and decision-making of Yunnan. Univ. Nationalities (Nat. Sci. Edn.), 12: 245-248.

Wang, M.Y. and Y.B. Lu, 2005. Software project outsourcing risk management contractor evaluation and selection. Stat. Decision, 4: 59-62.

Ye, J.P., Y.S. Zhang and X.F. Wang, 2005. Based on fuzzy clustering analysis comparative study of the operational phase of the real estate of Shenyang. Jian Zhu Univ. (Nat. Sci. Edn.), 21: 170-172.

Zhang, Z.Y., G.Q. Yuan and Z. Li, 2004. Technical innovation risk factors fuzzy clustering analysis and research. Hebei Acad. Sci., 21: 56-61.