Design of risk assessment and classification strategy using concentrated numeric statistic matrix and entropy method

Ai Nurhayati[1]* and Aditya Gautama[2]
[1]Department of Industrial Engineering, STT Bandung, Indonesia
[2]Department of Electrical Engineering, Batam State Polytechnic, Batam, Indonesia

*ain38375@gmail.com

Abstract. The construction project is a dynamic and risky field. Risks can have an effect on productivity, quality and cost limits of the project. Risk management is crucial for any construction project to avoid loss of cost, quality and time. This research was conducted at PT. XYZ. Risk identification is done through field survey and stakeholder interviews. There are 73 identified risks that are classified into 7 types of risks. Questionnaires were distributed to experts for probability and impact data. Calculation of probability average value and the risk impact are being done by concentrated numeric statistic matrix (CNSM) method. The final step in CNSM is the data consistency test using G reliability test. While the division of risk category uses entropy method. The benefits of combining these two methods are the assessment and classification of risk can be more accurate. CNSM will calculate the highest risk from all of the 73 identified risks. While entropy method will find the highest risk from the 7 risks type. From the result of the 2 methods, we can crosscheck to find the highest identified risk in the highest risk group type. The result shows that the risk of weather variability is the highest one.

Keywords: Concentrated Numeric Statistic Matrix, Entropy, Risk

1. Introduction

1.1 Background

The construction service industry is an industry sector that has a high enough risk. Risks and uncertainties exist in all aspects of construction work regardless of the size, complexity, location, resources, or speed of construction of a project. Perceptions of risk are key factors in making decisions that must be taken into account in managed risk management procedures. Risk management is an approach to risk by understanding, identifying and evaluating a project's risks. If the risk occurs, it will has an impact on the disruption of overall project performance that leads to cost loss, time and quality of work. In this research we will use CNSM method and entropy method. Concentrated numerical statistic matrix is used because it can help the expert in making more reasonable risk assessment (1). While entropy method is used because it is not as complicated as monte carlo simulation and can give risk probability rather than just rank them as in AHP (2). We used these approach in hoping that it will give a reasonable, accurate and less complicated risk assessment. This research was conducted at PT. XYZ.

1.2 Research questions
The problem formulation in this research is how to assess and classify the risk of construction project.

1.3 Research purposes
The purpose of this study is to assess and classify the risks of construction projects.

1.4 Benefits of research
The results of this study can be utilized for the parties related to the construction services industry.

2. Literature review

2.1 Concentrated Numeric Statistic Matrix
The following formula can be used to calculate the average value of probability or impact:

$$\bar{p} = \frac{\frac{1}{2}\sum_{k=1}^{n}\left[(p_2^k)^2 - (p_1^k)^2\right]}{\sum_{k=1}^{n}[p_2^k - p_1^k]} \tag{1}$$

Each expert on risk probability or impact judgment is an interval estimate, recorded as $[p_1^k, p_2^k]$ (k denote the kth expert). The reason CNSM is used because stakeholder can answer probability with ranged value. For example the risk of worker demonstration strike is 30-40 %, so we can write it [30,40]. The other reason for using CNSM is risk has characteristic to be uncertain thus a ranged value can describe it better, especially in construction industry.

Reliability test (1):

$$G = \frac{\frac{1}{3}\sum_{k=1}^{n}\left[(p_2^k - \bar{p})^3 - (p_1^k - \bar{p})^3\right]}{\sum_{k=1}^{n}[p_2^k - p_1^k]} \tag{2}$$

G reliability test is used to test data consistency. Smaller G value indicate more consistency in the data.

Table 1 Probability.

Probability (%)	Explanation
0 – 10	Very unlikely
11 – 30	Not likely to happen
31 – 60	Occur in the middle of project
61 – 90	Is likely to happen
91 – 100	Will come

Table 2 Impact.

Impact	Explanation
1	Most functions can still be realized
2	Small function cannot be achieved
3	Risk factors will make some function cannot be achieved
4	Risk factors will make the mostly function cannot be
5	Risk factors will influence the aim of the project greatly

$$R_i = \sum_{i=1}^{N} I_i \times P_i \tag{3}$$

I is the symbol of risk impact. P is the probability value of risk. R is the value of risk.

2.2 Entropy
In this research entropy method is used to calculate risk value for each risk group. Entropy method based on the multi-indexes assessment, at various risk index value (2).

$$R' = \begin{bmatrix} r'_{11} & \cdots & r'_{1n} \\ \vdots & \ddots & \vdots \\ r'_{m1} & \cdots & r'_{mn} \end{bmatrix} \tag{4}$$

R' is standardized to $R = (r_{ij})_{m \times n}$

$$r_{ij} = \frac{r'_{ij} - min(r'_{ij})}{max(r'_{ij}) - min(r'_{ij})} \tag{5}$$

If there are m risks and n experts, the entropy of index i is as follow :

$$H_i = -k \sum_{j=1}^{n} f_{ij} \ln f_{ij} \qquad\qquad i = 1,2,\ldots,m \qquad (6)$$

$$k = \frac{1}{\ln n} \qquad (7)$$

$$f_{ij} = \frac{r_{ij}}{\sum_{j=1}^{n} r_{ij}} \qquad (8)$$

The entropy eight of index i as follow :

$$\omega_i = \frac{1-H_i}{m-\sum_{i=1}^{m} H_i} \qquad (9)$$

$$R = \sum_{i=1}^{m} \omega_i H_i \qquad (10)$$

Gradation of risk level :

Table 3 Risk Level.

Risk level	Risk size	Risk value	Risk decision-making criteria
1	1~ 4	0 ~ 0,16	Acceptable and do not have to look at the management
2	5 ~ 8	0,16 ~ 0,32	Acceptable and management review
3	9 ~ 12	0,32 ~ 0,48	Accept or refuse to avoid the senior management decision
4	13 ~ 16	0,48 ~ 0,64	Unacceptable
5	17~25	0,61 ~ 1	Denied

3. Research methodology

This research was conducted at PT. XYZ. Risk identification is done by field survey and interviews with stakeholders. There are 73 risks detected then classified into seven types of risk. The questionnaires were distributed to six experts to determine the opportunity value and impact of 73 risks. Assessment of probabilities and risk impacts by the concentrated numeric statistic matrix method. Risk classification is determined by entropy method.
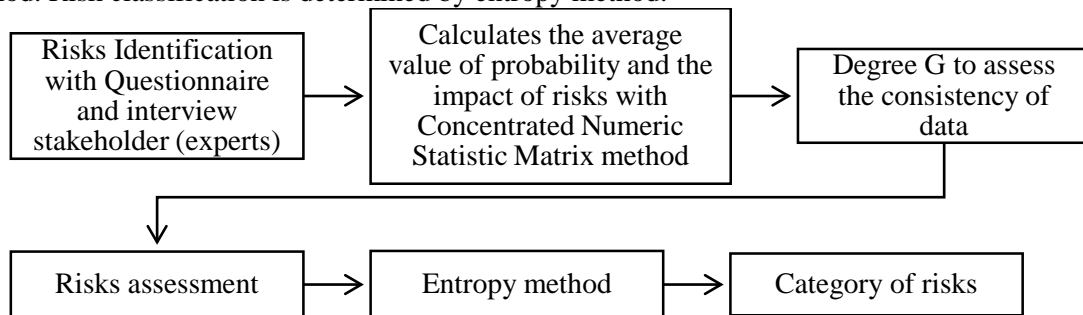

Figure 1 Flowchart of research.

4. Results

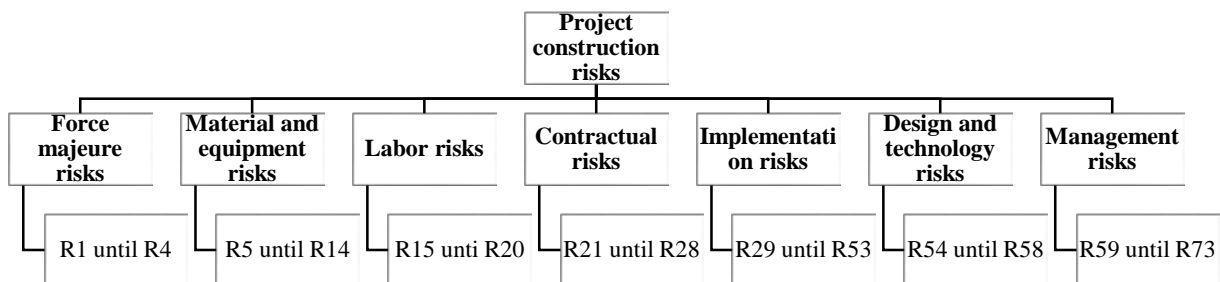The result of the study is described in the diagram at Figure 2 below:

Figure 2 Risks.

From the diagram above, it can be seen that the project construction risk is divided into 7 risk types. Each type consists of risk numbers obtained from the experts. The probability of the risk can be seen in the table 4 as following:

Table 4 The Probability of Risks.

Risk Factor	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	The average value of probability (%)	The data consistency test (Degree of G)
R3	[50,60]	[63,64]	[89.9,90.1]	[30.9,31.1]	[40,61]	[12,13]	51.22	0.9107
R36	[11,15]	[31,60]	[69.9,70.1]	[9.9,10.1]	[61,90]	[30,35]	55.48	4.2888
R54	[11,15]	[31,60]	[99.9,100]	[29.9,30.1]	[61,90]	[30,35]	55.56	4.2761
R56	[5,10]	[31,60]	[89.9,90.1]	[4.9,5.1]	[61,90]	[20,21]	55.68	4.9792
R58	[10,15]	[61,90]	[89.9,90.1]	[34.9,35.1]	[31,60]	[20,25]	54.22	4.8957

The risk probability values in table 4 is obtained from the experts. The integration of probability is calculated from the formula (1). The order of risk with the highest probability :
1. R56 is cracking and leakage risk
2. R54 is redesign risk
3. R36 is unsafe project workspace risk
4. R58 is safety rules disobedient risk
5. R3 is weather variability risk

From the table above we can see that R3 has the lowest G value. Thus R3 is more reliable to be assumed as the risk with highest probability.

Table 5 Impact of Risks.

Risk of factor	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	The average value of Impact	G	Class of risk
R5	[2,3]	[4,5]	[4.9,5.1]	[2.9,3.1]	[2,3]	[3.9,4.1]	3.5	1.06	4
R10	[4,5]	[4,5]	[4.9,5]	[4.9,5]	[3,4]	[4.9,5]	4.33	0.35	4
R21	[3,4]	[4,5]	[4.9,5]	[2.9,3.1]	[2,3]	[4.9,5]	3.64	0.87	4
R24	[2,3]	[3,5]	[4.9,5]	[3.9,4.1]	[3,4]	[3.9,4.1]	3.61	0.62	4
R25	[2,3]	[4,5]	[4.9,5]	[4.9,5]	[3,4]	[3.9,4.1]	3.69	0.85	4
R62	[2,3]	[3,4]	[4.9,5]	[3.9,4.1]	[4,5]	[3.9,4.1]	3.64	0.76	4

The order of risk with the highest impact :
1. R10 is material price increase risk
2. R25 is owner late payment risk

From the table 5, we can see that R10 has the lowest G value. Thus R10 is more reliable to be assumed as the risk with highest impact.

Table 6 Ranking of risks

Risk	The average value of Impact (I)	The average value of Probability (P)	R = I x P	Proportion = $\dfrac{I_i \times P_i \times 100\%}{\sum_1^{73} I \times P}$	Ranking of risk	Level of risk
R3	3.38	51.22	173.12	2.36	1	High
R54	3.03	55.56	168.35	2.29	2	High
R62	3.64	44.8	163.07	2.22	3	High

Classification of types of risk with probability and high impact using concentrated numeric statistic matrix method are:
1. Weather variability risk (R3)

2. Redesign risk (R54)

3. Subcontractor low performance risk (R62)

　　The classification of risk category in entropy method is done by using calculation from equations 4 to 10 in chapter 2. One risk group is displayed as one matrix. Inside the matrix, one row is one type of risk. First row from the top is R1, second row is R2, and so on. While the column describe expert opinion. The first column from the left is the answer from expert 1, the second row is from expert 2, and so on.

$$R'_{\text{Force Majeure}} = \begin{bmatrix} 3.5 & 4.5 & 3 & 3 & 1.5 & 3 \\ 2.5 & 3.5 & 4 & 3 & 1.5 & 4 \\ 3.5 & 3.5 & 5 & 3 & 3.5 & 3 \\ 2.5 & 4.5 & 4 & 3 & 1.5 & 3 \end{bmatrix}$$

$$R_{\text{Force Majeure}} = \begin{bmatrix} 0.67 & 1 & 0.5 & 0.5 & 0 & 0.5 \\ 0.4 & 0.8 & 1 & 0.6 & 0 & 1 \\ 0.25 & 0.25 & 1 & 0 & 0.25 & 0 \\ 0.33 & 1 & 0.83 & 0.5 & 0 & 0.5 \end{bmatrix}$$

$$k = \frac{1}{\ln 6} = 0.558$$

$$H_{\text{Force Majeure}} = (-0.558) \begin{bmatrix} -1.566 \\ -1.558 \\ -1.154 \\ -1.535 \end{bmatrix} = \begin{bmatrix} 0.874 \\ 0.869 \\ 0.644 \\ 0.856 \end{bmatrix}$$

$$\omega_{\text{Force Majeure}} = \begin{bmatrix} \dfrac{1 - 0.874}{4 - 3.243} \\ \dfrac{1 - 0.869}{0.757} \\ \dfrac{1 - 0.644}{0.757} \\ \dfrac{1 - 0.856}{0.757} \end{bmatrix} = \begin{bmatrix} 0.166 \\ 0.173 \\ 0.47 \\ 0.19 \end{bmatrix}$$

$$R_{\text{Force Majeure}} = 0.145 + 0.15 + 0.3 + 0.163 = 0.758$$

　　Force Majeure risk is highest risk.

5. Conclusion

CNSM method will calculate the highest risk from all of the 73 identified risk. While entropy method will find the highest risk from the 7 risk type. From the result of the 2 methods, we can crosscheck to find the highest identified risk in the highest risk group type.

The ones included in the highest risk class according to concentrated numeric statistic matrix (CNSM) method is weather variability risk (R3).

Risks that fall into the high risk category with entropy method is Force majeure risk (R1 until R4).

Result from data analysis shows that weather variability risk, from the force majeure risk group is identified as the highest risk class.

6. Reference

1. Gao C-J, Zhao D-Y, Ye W-M, Chen X. Risk assesment method based on concentrated numeric statistic matrix. In: 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE). Chengdu, China: IEEE; 2013. p. 384–7.

2. Xiao-mei G, Xiao-jun L. Application of Entropy Measurement in Risk Assessment of the Engineering Project of Construction-agent System. Syst Eng Procedia. 2011 Jan 1;1:244–9.