Journal of Applied Sciences

ISSN 1812-5654

Journal of Applied Sciences 13 (20): 4188-4194, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.4188.4194

A Risk Matrix Approach Based on Clustering Algorithm

[1]Lin He, [1]Yan Chen and [2]LanYing Liu
[1]College of Transportation Management, Dalian Maritime University, Dalian, China
[2]School of Continuing Education, Jing Gang Shan University, Jianxi, China

Abstract: The traditional risk matrix is insufficient in some cases due to the static, non-meticulous classification of probability and severity defined without considering of the context. In order to deal with this issue, this study proposes an extended RMA method (eRIA), by adding probability level and severity level steps in classification based on the basic steps of RMA. With the help of a classic clustering algorithm in data mining, which is K-means, through the iterative learning algorithm, dynamic division of probability and severity is realized. At last, the eRIA is applied to evaluate the traffic block risk and three ranks of risk sources are founded by the risk index.

Key words: Risk matix, risk index, clustering algorithm, k-means, traffic block risk

INTRODUCTION

Risk Matrix Approach (RMA) is a typical semi-quantitative assessment tool to evaluate various kinds of risks based on historically statistic data. It evaluate the risk factor through the research of the probability distribution and influences of each uncertainty factor. Risk matrix approach was first developed by Electronic System Center, US Airforce in April, 1995 to assess the risk in the life cycle of purchase project (Paul and Lansdowne, 1998).

In recent years, the research on risk matrix focused on two directions: (1) Ranking the risk index in RMA (2) Introducing the RMA to audience and discussing its application in certain fields. Markowski and Mannan (2008) introduced fuzzy set to RMA and proposed a fuzzy risk matrix (FRM) used for emerging fuzzy logic applications in different safety analysis. Ni *et al*. (2010) developed some arithmetic extensions on risk matrix approach and establishes a risk-matrix-style assessment framework. Burns and Kichak (2001) introduced RMA into NASA risk assessment and management roadmap. Zhu *et al*. (2003) illustrated the process of RMA and proposed the method by applying it to technical project risk management. Ruge (2004) applied the risk matrix as tool for risk assessment in the Chemical Process Industries. Zhao and Gao (2005) proposed some operable risk assessment procedures based on RMA after systematically discerning and analyzing the risk of ERP and some relevant factors. Zeng and Hou (2007) developed a customized risk matrix to assess the operation risk of metro system. Li *et al*. (2010) applied Risk Matrix in Classification of Dangerous and Hazardous Factors. Zhang *et al*. (2010) proposed a risk assessment model of information security based on risk matrix. Liu and Gao (2012) assessed the fire risk of hospital building based on improved risk matrix method.

The core of the risk matrix is the calculation of risk index. Risk index is the product of probability index and severity index, so the classfication of probability and severity directly affect the risk index levels. In the current study, probability and severity is divided into 5 levels or 7 levels. However, In the actual application, if the distribution of probability or severity is concentrated, there might be some blank area and some dense area in the traditional predefined classification of probability and severity. Then the RMA can not distinguish the relative risk of each factor.

In order to solve this problem, this paper improves the risk matrix calculation approach by adding a dynamic division step of probability and severity which based on the clusting method. It can dynamic classify the probability and severrity accroding to the certain sample data and the risk index can better reflect relative risk between various factors.

RMA (RISK MATRIX APPROACH)

ISO 2002 provided the standard definition of risk as the combination of the probability of an event and its consequence. Risk matrix method good expressed the definition. The risk index is the core of RMA, whick is defined as $R = P*S$, R is the risk index of a factor, P is the probability index of a factor, S is the severity index of a factor. The risk is proportional to the probability and severity.

Corresponding Author: Lin He, College of Transportation Management, Dalian Maritime University, Dalian, China

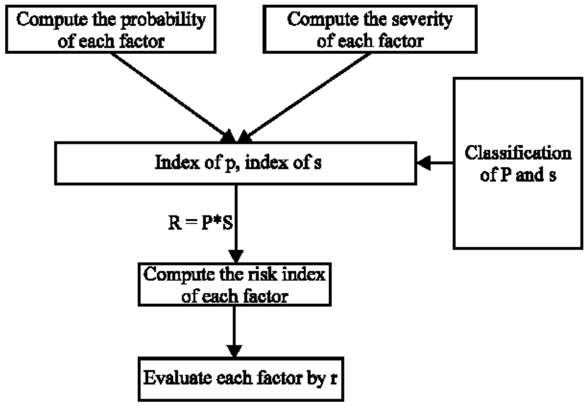

Fig. 1: Basic flow of RMA

Critical	M	H	H	H	H
Serious	M	M	M	H	H
Moderatec	L	M	M	M	H
Minor	L	L	M	M	H
Negligible	L	L	L	M	M
Origin	0.00~0.10	0.10~0.40	0.40~0.60	0.60~0.90	0.90~1.00

Fig. 2: Traditional risk matrix

The constructing the risk matrix as shown in Fig. 1 contains four basic steps: (1) Calculate the probability and severity of each factor (2) Categorization and scaling of the severity of consequences and probability (3) Categorization and scaling of output risk index based on the formula R = P*S (4) Build-up risk-based rules knowledge and graphical edition of the risk matrix.

However, from the above steps, it can be found that the classification of probability and the severity is independence with the sample data. According to the MIL-STD-882D (1984), the probability is divided into five level (remote (0.00~0.10), low (0.10~0.40), medium (0.40~0.60), high (0.60~0.90), very high (0.90~1.00) and the severity is divided into five level (1): Negligible, (2): Low, (3): Moderate, (4): High, (5): Catastrophic). Then a risk matrix with 25 cell, three level of risk (show as Fig. 2) is set up. The risk matrix builded in this way has one limitation that cannot be neglected,namely the static, non-meticulous classification of probability and severity. The traditional RMA is insufficient in some cases due to the classification defined without considering of the context. Especially when the distribution of probability or severity concentrated in a numerical interval, such as Probability distribution concentrated in between 0.1 to 0.4, the traditional RMA can not effectively rank the realitive risk of each factor.

ERIA(EXTEND RISK INDEX ASSESSMENT APPROACH)

According to the above issues, this study added probability level and severity level steps in classification based on the basic steps of RMA. With the help of a classic clustering algorithm in data mining, which is K-means, through the iterative learning algorithm, dynamic division of probability and severity is realized.

K-means is one of non supervised real-time clustering algorithms that proposed by Mac Queen, the assumption of clustering issues is that there is a set that consists of a group of data X = {xi, i = 1, 2, ..., n} is to be clustered. K means clustering issues is to find a division of X, P_k = {C_i, i = 1, 2, ..., n} ,which can minimize the target function:

$$f(Pk) = \sum_{i=1}^{k} \sum_{xi\in ci}^{n} d(xi, mi)$$

Which, mi = 1/ni, $\sum_{xi\in ci} xi$ represents the central position of cluster I, I = 1, 2, ..., k; ni is the number of the data items of cluster Ci; d (xi, mi) represents the distance between xi and mi,this study adopted Euclidean distance, that is:

$$d(xi, mi) = \left(\sum_{k=1}^{p} |xik - mik|^2\right)^{1/2}$$

After adding k-means cluster algorithm, we obtained the eRIA evaluation method, just as the Graph 3. Firstly, compute each factor's probability and severity, respectively by using K-means algorithm, aggregate probability and severity into cluster C (C = 2, 3, ..., m), determine each cluster's number range; on this basis, we can obtain the probability index and severity index of each factor; then, based on R = P*S, compute the risk index that correspond to the causes; quantitative risk evaluation can obtain from the risk index, cluster the risk index by using K-means algorithm, then we can obtain the category of dangerous source.

Specific steps are as follows:

Definition 1: The probability of each risk factor p.

p = {p (i), i = 1, 2, ..., m}, which p (i) represents the probability of risk factor i, which is F (i), may happen.

Assume one factor F (i) happens p (i) times in T, the total number of events that happen in T is n, the probability of F (i) is p (i), then according to the Bernoulli's law, for any positive integer ε>0, lim n→∞{|n(i)/n-p (I)|<ε} = 1. That is to say, if number n is

Table 1: Probability index

Probability Index	Level description
1	Impossible
2	Slightly possible
3	possible
4	Probable
5	Frequent

Table 2: Probability index

Probability Index	Level description
1	Severity: Low
2	Severity: Middle
3	Severity: High

very large, the probability n (i)/n of reason F (i) that happen converge to the probability p (i) that this factor happens. Thus, n (i)/n can be seen as p (i) approximately.

Definition 2: Each factor's probability index P.

Assume that the probability is divided into C (C = 2, 3, ..., m) partitions, for example, if c = 5, the probability of the event that initiated by the factor is clustered into 5 levels, just as the Table 1. The value of C needs to measure according to the samples data.

Determine each level range through K-means clustering algorithm based on the Euclidean distance, cluster p which is the probability of each risk factor happens by using K-means algorithm, then determine the data range of each probability index describes in the Table 1.

The clustering steps are as follows:

- Assume choose c (c = 2, 3 (default), ..., m) objects from n data objects, which are the initial clustering center
- Cycle 3 to 4, until each clustering never change
- According to the means of each clustering object (central object), compute the distance between each object and these central objects and re-divise corresponding objects according to the minimum distance
- Re compute the means (central objects) of each changeable clustering

After the iteration of above 4 steps, we can obtain final value range of the probability index in each level.

Definition 3: The severity of the event initiated by each risk factor S.

S represents the set of severity, S = {S (i), I = 1, 2, ..., 36} which S (i) represents the severity of the event initiated by factor i, that is F (i).

Definition 4: The severity index of each risk factor S.

Assume that the severity index is divided into C (C = 2, 3, ..., m) categories, for example, if c = 3, the severity of the factor is divided into 3 levels, just as Table 2 as follows. The value of C needs to measure according to the samples data.

The value range of each level can be obtained by using K-means algorithm to cluster each factor's severity.

Definition 5: Each factor's Risk index R_B (P, S) R_B (P, S) = P*S.

Risk evaluation: According to the value of each factor's risk index, evaluate the risk level among the factors, whose purpose is to distinguish relative risk level.

Risk level recognition: By using k-means algorithm, cluster the risk index, base on which, we can obtain the risk source structure diagram.

APPLICATION OF ERIA

Taking 276 pieces of traffic block information in a regional highway network of the north China in 2010 (MOC, 2011) as sample data,using the eRIA to evaluate the traffic block risk factor of the regional road net. The column 1, 2, 3, 4 of Table 3 list the classification of causes of traffic blocking.

Definition 1: S, the severity of traffic block, s = the time of traffic block * the miles of traffic block (miles.days).

Two hundred and seventy six block events occurred in this road net in 2010. Actually 201 of all caused by sudden reasons and other 75 caused by planned reasons. Probability and severity of traffic blockcaused by bad weather ranking first, were 42.03% and 73232.2 km day.

Calculate 2: The probability of traffic block caused by variety of reasons (Table 3).

Calculate 3: The severity of traffic block caused by variety of reasons (Table 4).

According to Table 3. we can see that ,most of the distribution of probability concentrated in 0 to 30%.In the traditional RMA, they share the same probabolity index as low. So, it's difficult to distinguish the realitive risk between them by traditional RMA. We will use the eRIA to assess the realitive risk.

Cluster 4: The probability and severity of traffic block caused by variety of reason. According to the samples data, the probability is divided into three categories, Table 5 lists the clustering ceter and content of each category.

Table 3: Probability of traffic block caused by variety of reasons

Factor Level I	Factor Level II	Code of factor	Description for factor	Probability (%)
1 Burstiness	11 Geological disasters	112	Landslide	0.0000
		113	Flood	0.0000
		115	Ground collapse, settlement or cracking	0.0000
		119	Other goelogical reason	0.0000
	13 Bad weather	133	Snow	10.5072
		134	Fog	28.6232
		135	Dust	0.3623
		139	Other weather reason	2.5362
	14 Accident disasters	145	Vehicle traffic accidents	23.6467
		149	Other	0.7246
	15 Others	159	Other unexpected reasons	9.4203
2 Planning	21 Construction and maintenance	211	Highway construction and maintenance	25.7246
		212	Construction and maintenance of bridge and tunnel	0.0000
		219	Other planning	0.0000
	22 Social activities	22	Major social activities	0.0000
		221	Carry out guard tasks	0.0000
	23 Others	239	Other	1.4493

Table 4: Severity of traffic block caused by variety of reasons

Factor Level I	Factor Level II	Code of factor	Description for factor	Severity (km day)
1 Burstiness	11 Geological disasters	112	Landslide	0.000
		113	Flood	0.000
		115	Ground collapse, settlement or cracking	0.000
		119	Other goelogical reason	0.000
	13 Bad weather	133	Snow	21855.019
		134	Fog	48873.657
		135	Dust	3548.970
		139	Other weather reason	2503.524
	14 Accident disasters	145	Vehicle traffic accidents	7407.405
		149	Other	802.000
	15 Others	159	Other unexpected reasons	2133.000
2 Planning	21 Construction and maintenance	211	Highway construction and maintenance	39426.296
		212	Construction and maintenance of bridge and tunnel	0.000
		219	Other planning	0.000
	22 Social activities	22	Major social activities	0.000
		221	Carry out guard tasks	0.000
	23 Others	239	Other	559.120

Table 5: Clustering results of probability index value

Level	Description	Clustering center	Content
1	Impossible	0.543	135, 149
2	Slightly possible	1.993	139, 239
3	possible	9.964	159, 133
4	Probable	24.686	145, 211
5	Frequent	28.623	134

Table 6: Clustering results of severity index value

Level	Description	Clustering center	Content
1	Low	1199.529	135, 139, 145, 149, 159, 239
2	Middle	7407.405	133
3	High	21855.020	134, 211

The Severity is divided into three categories. Table 6 lists the clustering ceter and content of each category.

Calculate 5: The risk index of traffic block caused by sub-factors R_B (P, S) = P*S and evaluation sub-factors of traffic block by risk index.

Bad weather: Bad weather includes fog , snow, dust, rain, typhoon, strong wind, high temperature and so on, according to Table 7. it can be found out that, in demonstration area, causes of block are mainly concentrated in some special weather like fog, snow and dust. The foggy days are the most important causes of block, it has a large probability of occurrence and high severity. Snowy days followed by. Foggy days mainly occurred from September to December and snowy days from November to February. In other words, in autumn and winter, we need to strengthen the supervision and prediction of fog and snow weather in this area, in order to prevent the occurrence of secondary transport event. Meanwhile, research from some Highway Meteorological Research Institute in china find out that, before heavy fog, there will be a small increase in visibility and then decline rapidly. Therefore, the detection of visibility can be used as auxiliary. If there is a visibility mutation in visibility curve, the heavy fog will probable appear in the next few hours.

Table 7: Sub-factors index of bad weather

Factor	Probability index	Severity index	Risk index
Fog	5	3	15
Snow	3	2	6
Other weather reason	2	1	2
Dust	1	1	1

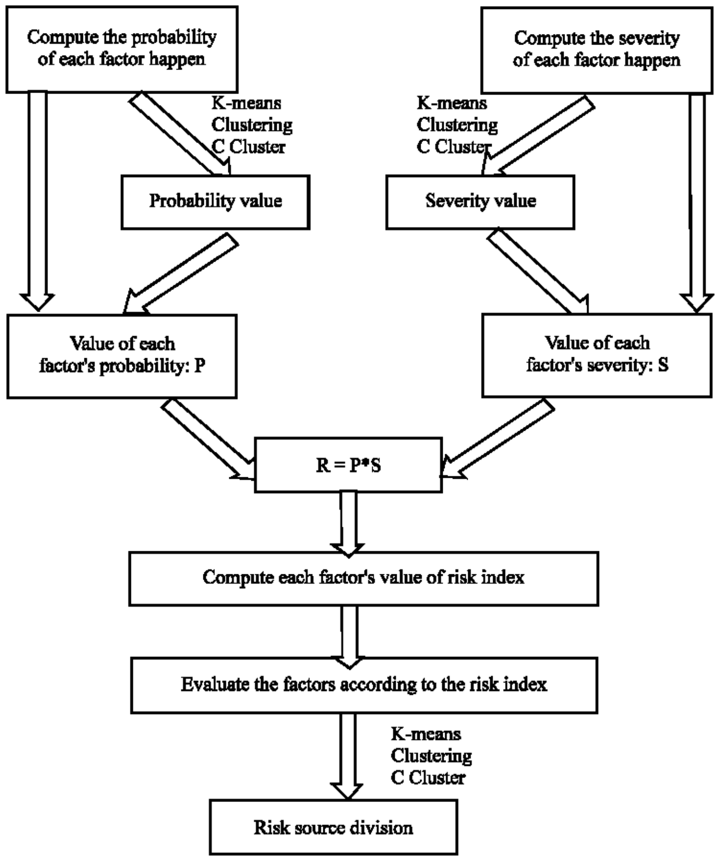

Fig. 3: Basic flow of extended risk index evluation approach

Table 8: Sub-factor index in accident disaster

Factor	Probability index	Severity index	Risk index
Vehicle traffic accidents	4	1	4
Other accident disasters	1	1	1

Table 9: Sub-factor index in other sudden events

Factor	Probability index	Severity index	Risk index
Other sudden reasons	3	1	3

Accident disaster: In accident disasters, includes vehicle traffic accidents, leakage of dangerous materials, vehicle breakdown, accidents involved in bridge and tunnel and so on. According to Table 8 it can be found out that, in demonstration area, the main accident reasons that caused traffic block are vehicle traffic accidents, it has a high probability, but the severity of block is relatively low. Therefore, in this area, this project will analysis the reasons of traffic accidents in detail in next step, mining the potential relationship among various of traffic accidents causes, find the potential regularity, reduce and avoid similar accidents.

Other sudden events: According to Table 9 we can see that, in demonstration area, the probability of traffic block impacted by other unexpected reasons is in possible level, the severity of block is relatively lower,the value of risk index is 3. But, there is no detailed description for other reasons in historical data. So it is suggested that decomposing other unexpected events into large vehicle flow caused by passing flow other sections, fees dispute, law enforcement event, service area events, performing garrison mission and so on. In this way, it helps to analysis the traffic block and determine the monitoring point.

Table 10: Sub-factor index in construction and maintenance events

Factor	Probability index	Severity index	Risk index
Construction and maintenance of highway	4	3	12

Construction and maintenance: Construction and maintenance includes highway, bridge and other sub-factors. Table 10 shows that, the probability of highway construction and maintenance in demonstration area is high, so is the severity of traffic block. Therefore, the advanced ork plan of highway

Table 11: Sub-factor index in other planning events

Factor	Probability index	Severity index	Risk index
Others	2	1	2

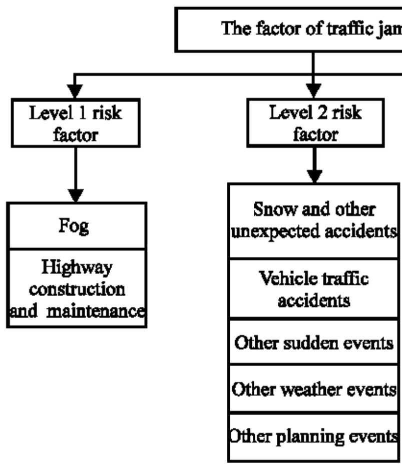

Fig. 4: Risk source of traffic block

construction and maintenance in this area and constructions based on the plan could effectively reduce road traffic blocking events.

Other planning events: The probability of other planning events is low, the severity is lighter. Historical data shows that the traffic block is mainly because that traffic police control and shunt vehicles. These factors can be planned in advance to avoid large-scale congestion.

Classify 6: The risk source based on the blocking risk index.

Based on the Analysis results above, this study builds the structure diagram of risk source of road traffic block (Fig. 4).

The risk source in the first class includes fog, highway construction and maintenance. The risk source in the second class includes snow, vehicle traffic accidents and other unexpected accidents (including large vehicle flow caused by passing flow other sections, fees dispute, law enforcement event, service area events, performing garrison mission and so on). The risk source in the third class includes dust, other accident reasons and other planning reasons.

CONCLUSION

This study introduces the clustering algorithm to risk matrix and establishes an extended risk index assessment approach, which adding probability level and severity level steps in classification based on the basic steps of RMA. With the help of a classic clustering algorithm in data mining, which is K-means, through the iterative learning algorithm, dynamic division of probability and severity is realized. Finally, the case study of eRIA used to evaluate the risk factor in traffic block showes the feasibility and validity of the extension. To develop the approach further, the following issues are worth considering in the future work. First, the decrease of risk tie should be considerate. Then, the fuzzy set could be introduced into the classification to solve the problem of high boundary division of k-means.

ACKNOWLEDGMENTS

Supported by NSFC (Natural Science Foundation of China) “Konwledge discovery and inference in road traffic safety analysis based on cooperation work” (No. 70940008). Supported by“ the Fundamental Research Funds for the Central Universities” (No. 3132013074; DUT13LAB05). Supported by “The Ministry of Education of Humanities and Social Science Fund for young” (No.13YJCZH052).

REFERENCES

Burns, J. and L. Kichak, 2001. NASA risk assessment and management roadmap. Proceedings of the International Conference on Systems Engineering Capstone, June 2001, Hampton, VA., pp: 183-188.

Li, S.Q., Z. Yan and Y. Duan, 2010. Application of risk matrix in classification of dangerous and hazardous factors. China Saf. Sci. J., 4: 83-87.

Liu, J. and X.N. Gao, 2012. Fire risk assessment of hospital building based on improved risk matrix method. Math. Practice Theory, 42: 22-22.

MIL-STD-882D, 1984. Military standard system safety program requirements. Department of Defense, USA. http://www.system-safety.org/Documents/MIL-STD-882B.pdf

MOC, 2011. ICP record No. 05046837. The Ministry of Communications of the People's Republic of China. http://www.moc.gov.cn/zhuzhan/gongluchuxing/lukuangxinxi/zuduanxinxi/.

Markowski, A.S. and M.S. Mannan, 2008. Fuzzy risk matrix. J. Hazard. Mat., 159: 152-157.

Ni, H., A. Chen and N. Chen, 2010. Some extensions on risk matrix approach. Safety Sci., 48: 1269-1278.

Paul, R. and Z.F. Lansdowne, 1998. Risk matrix: An approach for identifying, assessing and ranking program risks. Air Force J. Logist., 25: 16-19.

Ruge, B., 2004. Risk matrix as tool for risk assessment in the chemical process industries: Paper No. 0192. ESREL, Berlin, Germany.

Zeng, T.M. and J.G. Hou, 2007. Risk management in metro operation. Eng. J. Wuhan Univ., 40: 84-93.

Zhang, T., D.J. Mu, S. Ren and L. Yao, 2010. Risk assessment model of information security based on risk matrix. Comput. Eng. Appl., 5: 242-245.

Zhao, P. and X.M. Gao, 2005. Risk assessment method of ERP based on risk matrix. Manufact. Technol. Mach. Tool, 3: 87-90.

Zhu, Q.C., X.H. Kuang and Y.P. Shen, 2003. Risk matrix method and its application in the field of technical project risk management. Eng. Sci., 5: 89-94.