Study on the model of safety assessment based on Collaborative filtering algorithm

Guoqing Xiao [1], Hengbo Zhao [1,*] ,Junhao Feng [2], Renjie Xiao [1], Qionghui Wang [1]

[1]Oil & Gas Fire Protection Key Laboratory of Sichuan Province, College of Chemistry and Chemical Engineering, Southwest Petroleum University, Chengdu, 610500,China

[2] Jiayuan Gas Ltd, Co. Chengdu, Sichuan, 610213, China

Abstract

In safety evaluation, operating conditions and equipment facilities cannot be quantitatively evaluated due to various reasons. In this paper, based on a collaborative filtering algorithm of feature vector, a safety evaluation model was established with LEC evaluation method. Through the comparison of similar projects, like safety evaluation and prediction, this model can evaluate non-quantifiable project and a number of unknown project, so it is important to eliminate the hidden dangers. This model has been verified by MATLAB programming to have a good application value.

Keywords

Collaborative filtering algorithm; LEC; Prediction; Evaluation model; MATLAB

1. Introduction

With industrial development, technological advances, people's living standard had a qualitative leap, but the development of the security issues are also increasingly prominent, major accidents still occurs. To prevent accidents, it is necessary to strengthen the security technology research, so it is intrinsically safe, as well as strengthen the security management. Thus the use of advanced scientific management methods safety is very important. Safety evaluation is the technical foundation of safety management, there are qualitative and quantitative evaluation methods that are commonly used for safety assessment methods, qualitative evaluation methods, expert evaluation method [1], the safety check list [2], fault tree [3]; quantitative evaluation methods are probabilistic risk assessment methods, risk index evaluation method. For more complex systems, you need to establish a quantitative safety assessment model, but in the safety evaluation model [4] [5] also has the following disadvantages: Due to operating conditions, health of equipment and facilities' uncertainty, variability, and high complexity, as well as the risk of accidents statistical data is insufficient, plus the subjective differences, the situation often can not be quantified. Therefore, this article will apply collaborative filtering algorithm introduced cloud model safety evaluation model to better solve the problem of uncertainty evaluation model due to various factors.

2. Collaborative filtering algorithm

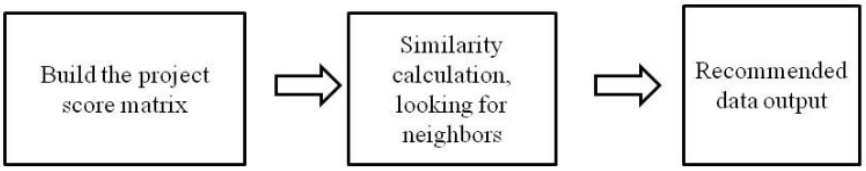
In 1992, Gothenburg, who proposed social filter, also known as collaborative filtering, was first used in research mail recommendation systems [6]. Model-based and the memory-based methods are the two aspects constitute a collaborative filtering technology. Learning a scoring system from it self's existing user information to get an model, and then use this model to predict needed information and recommendations, are called model-based; First, calculating the similarity between the project and user, and then find the target project or the user by select the most similar existing projects or users, that is to find the nearest neighbor, finally scoring or rating information to produce the project's recommendation by the previous user, this method is called memory-based method [7].

The similarity of the target project or projects with the highest user or user screened collaborative filtering algorithm occupies an extremely important position. Because, based on various methods to calculate a target selected item or similarity with existing user or users on the project, according to

the similarity level obtained target project or user finding the most similar projects or users, is the basic principle for collaborative filtering algorithm. The available similarity calculation methods are: the modified cosine similarity, cosine similarity and correlation similarity etc. When there are missing items on users information data in their daily lives, default data accuracy result in certain similarity calculation method declined, affecting the quality of recommendation. In this paper, the cloud model is introduced, in order to improve the quality recommended in the evaluation of the data is sparse.

From a statistical point of view, the Ex (Expected value), variance Va (Variance), kurtosis Kb (Peakedness) all three figures can express the overall characteristics of cloud model. Ex expectation is desirable to quantify the most typical sample, showing qualitative point; variance Va used to measure the degree of deviation, showing the relationship between random variables and its mathematical expectation between; kurtosis Kb is to describe the distribution aspect of the degree of steep relief.

Collaborative filtering algorithm basic steps, as shown in Figure 1.



Collaborative filtering algorithm step in Figure 1

3. Model building

3.1 LEC evaluation method and evaluation

LEC evaluation method is a method for safety evaluation by consider L, E, C three factors[8] [9]: L represents the probability of an accident, E represents the frequency of personnel in hazardous environments jobs, C represents consequences of the accident. D = L ×E ×C, D representing the risk, the greater its value, the more dangerous hazards and working conditions, L, E, C were assigned in the following table 1,2,3,4.

The probability of an accident in Table 1 assignment L

Assignment L	The probability of an accident
0.1	Practically impossible
0.2	Almost impossible
0.5	Conceivable
1	Totally unexpected
3	Possible
6	Likely
10	Totally expected

Table 2 how often in hazardous environments job assignment E

Assignment E	How frequently people in hazardous environments jobs
0.5	Very rare
1	Several times a year
2	Monthly
3	Accidental
6	Daily working hours
10	Frequently

Table 3 accident consequences assignment C

Assignment C	Accident Consequences
1	Minor injuries
3	Disabled
7	Severely disabled
15	1 death
40	Number of deaths
100	More than 10 death

Table 4 standard of risk

Assignment D	More than 320	160 to 320	70 to 160	20 to 70	Less than 20
The degree of risk	Extremely dangerous Stop job	Highly dangerous Needs rectification	Significant dangerous Needs rectification	Common danger Pay attention	Slightly dangerous Acceptable

Since the needs assessment model, the five levels of D assignment: 5,4,3,2,1 see Table 5.

Table 5 Risk level assignments

Assignment D	More than 320	160 to 320	70 to 160	20 to 70	Less than 20
Assignment	5	4	3	2	1

3.2 Introducing cloud model

Characterized by three numbers represent the overall characteristics of qualitative concept, to establish the feature vector cloud model:

$$C(Ex，Va，Kb)$$

Among them,

$$Ex = x_1p_1 + x_2p_2 + x_3p_3 + \cdots + x_np_n$$
$$Va = (x_1 - m)^2 + (x_2 - m)^2 + \cdots + (x_n - m)^2/n$$
$$m = （x_1 + x_2 + x_3 + \cdots + x_n）/n$$
$$kb = \frac{\sum_{i=1}^{n} （x_i - Ex）^4}{（n - 1）Va^2}$$

3.3 Building projects accidents Matrix

Assign numbers according to the evaluation of the severity of the project, were recorded from mild to severe from {1,2,3,4,5}. For evaluation of multiple projects, then recorded as U (u1, u2, u3, u4, u5), where u1 to u5 for the given number of times corresponding to the five levels of the accident, U is the project accident frequency vector.

Establish project LTA model:

$$A = \begin{bmatrix} R & u_1 & u_2 & u_3 & u_4 & u_5 \\ R_1 & u_{11} & u_{21} & u_{31} & u_{41} & u_{51} \\ R_2 & u_{12} & u_{22} & u_{32} & u_{42} & u_{52} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ R_n & u_{1n} & u_{2n} & u_{3n} & u_{4n} & u_{5n} \end{bmatrix}$$

R1, R2 ... Rn is project name;

u_{11}, u_{12} ... u_{5n} as the number of accidents occur.

3.4 Construction of the project similarity matrix

Firstly cloud model matrix:

$$A = \begin{bmatrix} \mathrm{R} & u_1 & u_2 & u_3 & u_4 & u_5 \\ R_1 & u_{11} & u_{21} & u_{31} & u_{41} & u_{51} \\ R_2 & u_{12} & u_{22} & u_{32} & u_{42} & u_{52} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ R_n & u_{1n} & u_{2n} & u_{3n} & u_{4n} & u_{5n} \end{bmatrix} \rightarrow$$

$$E = \begin{bmatrix} \mathrm{C}_1 & \mathrm{C}_2 & \cdots & \mathrm{C}_n \end{bmatrix}$$

C_i and C_j is the two cloud droplets of i and j denote the vector, whose cosine angle called cloud droplets similarity between i and j:

$$Sim(i,j) = cos(C_i, C_j) = \frac{\mathrm{C}_i \bullet \mathrm{C}_j}{\|C_i\| \|C_j\|}$$

On this basis, the project to establish similarity matrix:

$$Sim = \begin{bmatrix} & R_1 & R_2 & \cdots & R_n \\ R_1 & Sim(1,1) & Sim(1,2) & \cdots & Sim(1,n) \\ R_2 & Sim(2,1) & Sim(2,2) & \cdots & Sim(2,n) \\ \vdots & \vdots & \vdots & \ddots & \\ R_n & Sim(n,1) & Sim(n,2) & \cdots & Sim(n,n) \end{bmatrix}$$

3.5 Evaluation Model

In MATLAB, project evaluation model based on the similarity established as shown in Figure 2 :

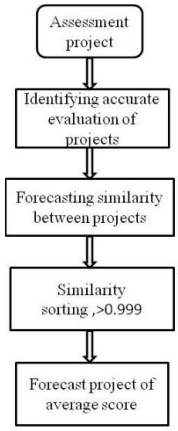


Figure 2 Evaluation Model

4. 3. Implementation and application of the model

A LEC evaluation data collection sheet partially shown as in Table 6. Project average is the arithmetic mean of each project in accordance with a plurality of LEC evaluation results are assigned. Table 6 contains no items 7245481,7625225,4156658,5997834,9214078,2515537 etc., the need for the project to evaluate and forecast item 7245481, 7625225, 4156658, 5997834, 9214078, 2515537.

Table 6 LEC evaluation data table

Item Number	Average project
702699	4.079044

219560	4.173949
860262	4.217676
235338	4.109977
495027	3.820144
⋮	⋮

Excel on a project with the accident frequency data collation, partly as shown in Table 7.

Table 7 Project accident frequency table

Item Number	Accident frequency and severity of accidents under assignment				
	1	2	3	4	5
117711	0	2	2	9	0
183976	0	0	6	41	20
217101	2	1	98	191	36
269836	0	1	4	19	7
285794	0	0	0	6	1
⋮					⋮
7245481	1	7	131	359	19
7625225	0	1	170	102	0
4156658	10	19	50	139	84
5997834	0	0	1	10	6
9214078	1	3	46	162	59
2515537	0	11	194	81	18
⋮					⋮

According to the project data to calculate accident frequency eigenvalues, some results in Table 8 below.

Table 8 eigenvectors project

Project	Ex	Variance	Kurtosis
117711	3.538462	0.954142	1.078352
183976	4.208955	0.26224	4.566525
217101	3.786585	0.040042	470.6123
269836	4.032258	0.510926	3.542002
285794	4.142857	2.755102	0.011907
⋮			⋮
7245481	3.750484	0.024837	666.6854
7625225	3.369963	0.039281	54.27566
4156658	3.887417	0.046304	1698.169
5997834	4.294118	1.148356	0.207077
9214078	4.01476	0.056106	278.1481
2515537	3.348684	0.03501	505.3625
⋮			⋮

Using MATLAB to achieve the above steps 4 and 5, some predictions are shown in Table 9:

Table 9 model predictions

Item Number	According to the similarity forecast
7245481	4.055408971
7625225	3.765625
4156658	3.76984127
5997834	4
9214078	4.118257261
2515537	3.857142857

5. Conclusion

In this paper, we have done a lot of improvements based on the feature vector of traditional collaborative filtering algorithm, embodied in:

To adopt a consistent approach to improved ratings for the item rating scale problems inconsistencies can provide a solution.

Recommendation algorithm based on feature vectors proposed the introduction of a cloud model, calculated on the basis of the similarity of feature vectors to optimize the sparcity of data, it simplifies the computational complexity, improve the quality of recommendation.

After the model calculated the similarity of the project, not in isolation or select an item 1 or n to the nearest neighbors are compared to predict, but the threshold is set to select the nearest neighbor project. Thus the number of nearest neighbor project may be unstable, but the similarity does not have large errors.

Combining LEC evaluation method used as common means, establishing safety evaluation model based on collaborative filtering algorithm, considering various reasons for not be able to give quantitative evaluation of operating conditions, equipment and facilities as well as numerous other projects were predicted and evaluated. Therefore, the evaluation results provide significant guidance to the elimination of hazard working environment.

Acknowledgement

This work was supported by National Natural Science Foundation of China, ratified no 51474187, and Research Fund for the Doctoral Program of Higher Education of China, ratified no 20135121110003.

References

[1] Tian Jian, Li Zhiqiang, Zhang Bin transportation construction engineering safety evaluation of technical status and trend research [J] China Safety Science Journal, 2008, 18 (6): 171-175.

[2] Wang Qinghui, Liu Peng, Wang Danfeng study safety checklist for the operating conditions of risk analysis methods correction [J] China Safety Science and Technology, 2013, (8): 125-129

[3] Buquan Min, Wangyong Tao, Wonder light a fire Application of fault tree analysis method [J] Southwest Petroleum University ISTIC PKU, 2007,29 (4):141-144.

[4] Zhang Yongping, Zhang Yongwen, Sun Feng safety assessment methods used in China status quo and thinking [J] Science Technology and Engineering, 2010, 10 (27): 6828-6831.

[5] Zhang Zhaoning, gray Bing risk assessment process control system operation [J]. Science Technology and Engineering, 2013, 13 (17): 5071-5077.

[6] Deng Ailin Zhu Yang Yong, Shi Bole recommendation algorithm [J] project-based collaborative filtering score prediction software Sinica, 2003,14 (9): 1621-1628.

[7] Cong, Liang Changyong, based on domain nearest neighbor Mary collaborative filtering recommendation algorithm [J] Computer Research and Development, 2008,45 (9): 1532-1538.

[8] FENG Y, ZHANG B, WU W, et al. Power System Operation Risk Assessment Based on Credibility Theory Part One Propound and Development of Operation Risk Assessment [J][J]. Automation of Electric Power Systems, 2006, 1: 003.
[9] Zhu Yuan Yue, Fu Xuehua, Li Kerong etc. Application of LEC Method to improve water conservancy and hydropower project construction period Hazard Evaluation [J] China Safety Science and Technology, 2009, 5 (4): 51-54.