Journal of Computational Science & Engineering 21 (2015) 666-670

Available at http://www.asocse.org ISSN 1710-4068

Analysis of Dangerous Chemical Accident Cause based on Apriori Algorithm

Hao Yu[1,2,*]

[1]School of Emergency Management, Henan Polytechnic University, Jiaozuo, Henan 454000, China

[2]China University of Geosciences, Beijing, 100083, China

Abstract: **In order to identify the association between causes of accidents due to hazardous chemicals, clear and definite the key points of safety supervision, and improve the efficiency of security supervision, this study researches 150 accident investigation report as data source by data mining technology, Apriori algorithm, and finds out the relationship between factors such as time, step, type, severity and cause. By analyzing the 8 valuable rules, it proposes 4 suggestions for improving key safety supervision, thus to promote the optimization of safety supervision strategy of hazardous chemicals.**

Keywords: Safety supervision; Dangerous chemicals; Data mining; Apriori algorithm

Received 29 October 2015, Revised 23 November 2015, Accepted 25 November 2015

*Corresponding Author: Hao Yu, haoyu@163.com

1. Introduction

Hazardous Chemicals (hereinafter referred to as "HC") are hypertoxic or other chemicals with the properties of poison, explosion, corrosion, combustion, combustion that may cause severe harms to human beings, facilities, and the environment. Accidents due to HC usually happen in a sudden, with diverse forms, deep severity and difficult disposal [1-2]. According to incomplete statistics, from 2010 to 2014, accidents due to HC came up to 326 in total where 2237 people died. They've caused serious damage to people's life and property. So measures are urgently needed to put forward to prevent or reduce accidents in order to promote the optimization of safety supervision strategies.

At present, related domestic researches on the safety of hazardous chemicals mainly focus on statistics analysis, information system establishment and credit system establishment and etc. Zhao Laijun and Wu Ping found the characteristics and regularity of such accidents by analyzing collected statistics [3]; Wu Zongzhi and Sun Meng made accident analysis of HC accidents on China's highway transportation based on statistics [4-5]. Ma Liangjun and Wei Lijun established a software platform for emergency support based on the analysis of general characteristics and disposals [6]. Bao Aiguang and Wang Mingxian established enterprise security credit system to assist the government's supervision and management of relevant enterprises [7].

The above researches seldom studies on the association between the factors involved, especially the relation between type, consequence and cause, let alone the countermeasures for government safety regulation. Based on large data thinking, data mining technology and Apriori algorithms, the author is to dig out the association rules between cause factors, and then put forward some suggestions to optimize the safety supervision of hazardous chemicals.

2. Theoretical background

Data Mining generally refers to searching hidden information by algorithm in large data. Mining by association rule refers to finding interesting and frequent pattern, association and correlation based on item database, relational database and other information stored in a large number of data items [8-9] to make them be useful knowledge.

Association rule is defined as: Suppose I= { I1, I2……, Im } is a collection of items and a database D is given with the fact that each item T is a not loophole set of I, that is to say each item has a unique identifier ---TID. Supportability in association rule is both X and Y in database D; Confidence means the the probability of X under the condition of having X.

Under the condition that minimum supportability meets with minimum confidence, the association rule is considered as interesting, where the thresholds are identified according to the needs of mining. There are two steps in this process. The first step is to find all frequent item sets, the frequency of each is as half as what is supposed to be. The second step is that strong association with minimum supportability and confidence emerges among the frequent item sets.

Apriori Algorithm is commonly used in association rule proposed by Agrawal and R.Srikant in 1994. It was an original algorithm for Boolean association rules mining frequent item sets. Use the iterative method by searching step by step where K item set is to explore (k + 1) ones. First, by scanning the database, the time of the appearance of each item is counted, and items that satisfy minimum supportability is collected and marked L1. Then use L1 in the frequent item set NO.1 to find L2 in the frequent item set NO.2, then L3 and so on until K item set is found.

2015 © The American Computational Science Society. All rights reserved.

3. Ciation rule mining for accidents due to hazardous chemicals

3.1. Data preparation

Data preparation includes four parts: raw data selection, data extraction, data reduction and data transformation.

(1) Raw data selection: In order to ensure the timeliness and authoritativeness of data, accident investigation reports for the last 1 decade in the website of National Administration of Production Safety Supervision and Management and the sub websites of provinces (autonomous regions and municipalities) are chosen. And accident databases established by professionals are also selected as the raw data. Factors such as time, place, and type are not deliberately considered to ensure the randomness.

Table 1 Direct reason code

Category No.	
Unsafe states	
A11	Lack of or defective devices of protection, insurance, and signal
A12	Defective equipment, facilities, tools and accessories
A13	Poor production (construction) environment
Unsafe behaviors	
A21	Error operation, safety neglecting and warning ignoring
A22	Safety device failure
A23	Use unsafe equipment
A24	Venture into dangerous places
A25	Distraction
A26	Ignore the usage of personal protective equipment or utensils when urged
A27	Error handling of inflammable, explosive and other dangerous items

(2) Data Extraction: An information form is made after the analysis of selected 150 reports. In order to ensure the integrity and clear expression of the information, the form is divided into three parts: basic information, accident information, and accident cause. Basic information includes date, time, location (provinces, districts and counties). Accident information includes accident link, category, severity, the number of death, injuries and direct economic loss. According to the accident investigation analysis of workers (GB6442-86) and the casualty classification standard of workers (GB6441-86) in combination with the practical situations, accident causes are divided into direct and indirect reasons. And direct reasons are divided into 2 types: unsafe conditions and unsafe behaviors with 10 items. Indirect reasons are divided into 7 items in which unreasonable labor organization is one of them.

Table 2 Indirect reason code

Category No.	Indirect reasons
B1	Flaws in the design of process, operation method, or in construction and material usage
B2	Lack of education and training, or no training, or lack of safety operation knowledge
B3	Unreasonable labor organization
B4	Lack of inspection for field work or error guidance
B5	No or inadequate operation of safety procedures
B6	No or not serious implementation of preventive measures against accidents, or lack of rectification of hidden dangers
B7	Others

(3) Data Reduction: Based on the analysis of principal components and experts' opinions, the author extracts the information and makes reduction, where dirty, incomplete and inconsistent data is abandoned to improve the quality, accuracy and efficiency of data mining. For example, the number of death, injured and direct economic loss can be summarized in severity. Merge repeated data input by reason of different expressions of accident sites. Finally, 102 pieces of data is ended up with.

(4) Data Transformation: Make the reduced data standardized, discredited and layered to form machine-readable information. For example, accident time is divided into A, B, C, D four periods starting with zero. Direct accident cause is divided into A11 A12, A21, A22, while indirect accident cause is divided into B1, B2, B3, with T (true) representing the reason exists. Through the above four steps, a standardized database is eventually formed, as table 3 shows.

3.2. Association rules mining

Apriori Algorithm is used to analyze standardized database, setting the minimum confidence as 60%, the minimum supportability as 10%, and.

2015 © The American Computational Science Society. All rights reserved.

Table 3 The standardized database

ID	Time	link	Severity	Category	A11	..	A13	A21	..	A27	B1	..	B6	B7
1	B	Use	Major	Explosion			T						T	
2	C	Storage	Major	Fire			T						T	T
3	B	Product	Major	Explosion							T		T	T
4	B	Trans	Major	Fire	T						T		T	T
5	A	Product	Larger	Pois-choke				T						T
6	C	Product	Larger	Leak	T	..			..			..		T
7	C	Storage	Larger	Explosion				T		T				T
8	C	Use	Larger	Pois-choke			T				T			T
9	A	Trans	Extra	Fire						T	T		T	T
10	A	Product	Ordinary	Explosion			T						T	
…	…	…	…	…				…				…		
102	A	Product	Ordinary	Burns	T		T				T			T

First of all, find all frequent item sets. Calculate with 5% of iterative diminishing unit according to the formula:

$$\text{support}(A \Rightarrow B) = \frac{Count(A \cup B)}{Count(D)},$$

and the results show that the iterative diminishing unit becomes 10% after 18 times interation, 1 item sets with 23 items, 2 item sets with 117 items, 3 item sets with 111 items and 4 item sets with 23 items.

From those item sets, tight association can be found. According to the formula:

$$confidence(A \Rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)},$$

292 pieces of rules are found with the confidence of no less than 60%.

Table 4 The strong association rules of dangerous chemical accidents

Rules ID	Rules consequent	Rules antecedent	Confidence	Support
1	B7=T	Severity=larger B3=T	100%	22.5%
2	B7=T	Time=C category=explosion	100%	19.6%
3	B7=T	Category=explosion A21=T	100%	17.6%
4	B7=T	B1=T B6=T	100%	17.6%
5	Severity=larger	Time=C A27=T B7=T	100%	10.8%
6	B7=T	Time=B B3=T	100%	11.8%
7	Category=pois-choke	A26=T B2=T	100%	10.8%
8	Severity=major	Time=B B6=T B7=T	91%	10.8%
9	Severity=major	A21=T B6=T B7=T	83%	10.8%
10	Category=explosion	Project=production A21=T B7=T	71%	10.8%
……	……	……	……	……
291	Category=explosion	Time=C severity=larger B7=T	60%	11.8%
292	Severity=larger B7=T	Time=C category=explosion 20	60%	11.8%

Choose valuable rules. According to the formula, rules with back item of severity or category, and front item of time, project or reason are correct. Combining experts' opinions and actual cases, correct rules are selected; repeated or meaningless rules are deleted; finally, 8 valuable rules are found as table five shows in the following.

A large number of diverse data is the basis of data mining by algorithm. Less data might restrain the obtaining of value rules. For example, only one valuable rule obtained from a fire accident is usually not the real cause. And because the main sources of data are from reports of the administration of national security supervision and management, there are more serious accidents or special serious accidents than ordinary ones. So consequent valuable rules with severity = ordinary as its back item are not found. In the later study, the author should combine with accident database of local enterprises to establish accident data cloud platform, data analysis based on greater number of accidents, thus to get more valuable rules and carry out the optimization of regulation strategy.

2015 © The American Computational Science Society. All rights reserved.

Table 5 Valuable rules of dangerous chemical accidents

Rules ID	Rules consequent	Rules antecedent	Confidence	Support
1	Category=pois-choke	A23=T B2=T	100%	10.8%
2	Category=pois-choke	Project=use B1=T B2=T	83%	10.8%
3	Category=explosion	A27=T B7=T	73%	11.8%
4	Category=explosion	Project=production A21=T B7=T	71%	10.8%
5	Category=fire	A12=T B6=T	63%	10.8%
6	Severity=larger	Time=C A27=T B7=T	100%	10.8%
7	Severity=major	Time=B B2=T B6=T	100%	12.7%
8	Severity=major	Time=B B6=T B7=T	91%	10.8%

4. The optimization of safety regulation on hazardous chemicals

On the basis of the authoritativeness and timeliness of data, association rule mining for accidents due to hazardous chemicals has scientifically and effectively found the relation between time, link, cause, type and severity, and made make focuses clear for relevant safety regulators to increase efficiency. Based on valuable rules in table 4, the following advice is concluded.

As for those enterprises that might accumulate poisonous and harmful gases, the focus should be on security training to prevent employees from suffocation caused by lack of safety consciousness into dangerous places. The equipment maintenance should be checked to prevent leakage accidents caused by the defects in usage, maintenance, and transformation of equipment.

As for the enterprises that use explosive chemicals, screening efforts for the hidden perils should be strengthened. In particular, behaviors like illegal production, operation, storage, sales of fireworks should be cracked down to prevent explosion accident caused by improper treatment for hazardous chemicals. As for those enterprises that produce flammable and explosive chemicals, the inspection of personnel quality should also be strengthened to prevent explosion caused by operation mistakes, security and warning ignorance.

As for the enterprises that produce, use, store, or transport flammable chemicals, the key is to check hidden perils, particularly equipment, facilities, accessories and tools related with flammable chemicals to prevent fire accident caused by bad repair, adjustment or abnormal operation.

According to the statistics, accidents due to hazardous chemicals mainly happen from 6:00 am to 6pm. It might be related to high labor in production, loading and unloading, and transportation during the day. From ID7 and ID8, lack of or no personnel training might cause improper loading and unloading, and transportation; enterprises might not seriously implement preventive measures against accidents; and relevant departments lack of supervision and inspection of the above two. All these might easily lead to larger or major accidents. In daily regulation, therefore, we should focus on two items which are quality safety supervision and the inspection of enterprises' implementation of preventative measures.

5. Conclusion

From the thinking of big data, using data mining and Apriori algorithms, with 150 accident investigation reports as data source, the essay finds out the association between time, link, cause type, and severity. It ends up with eight valuable rules with confidence more than 60%, supportability more than 10%.

After the analysis of the 8 valuable rules, it is concluded that daily regulation should focus on the quality safety supervision and the inspection of implementation preventive measures. It puts forward the corresponding regulatory focuses for enterprises that use flammable, explosive, toxic chemicals respectively, in order to improve regulatory efficiency and optimize the the regulatory strategy.

The essay carried on the preliminary exploration of the application of data mining technology in safety regulation. However, there are not enough accident data with imbalanced types. Further study will use the thinking of big data based on internet cloud platform to dig for huge amounts of data to better promote the optimization of safety regulation strategy.

References

[1] GB13690-2009, The general principles of the chemical classification and risk of the public.
[2] L.J. Zhao, P. Wu, X. Ke, China Safety Science Journal, 07 (2009) 165-170.
[3] J.Q. Ren, Y.X. Mu, Chemical Enterprise Management, 16 (2015) 28-31.
[4] Z.Z. Wu, M. Sun, Journal of Safety Science and Technology, 02 (2006) 3-8.
[5] M. Sun, Z.Z. Wu, H.Y. Zhang, China Safety Science Journal, 08 (2003) 25-27.
[6] L.J. Ma, L.J. Wei, G.X. Yi, Journal of Safety Science and Technology, 05 (2009) 23-26.
[7] A.G. Bao, M.X. Wang, China Safety Science Journal, 12 (2008) 75-79, 177.

2015 © The American Computational Science Society. All rights reserved.

[8] S.Y. Liu, L. Yang, Coal Technology, 04 (2015) 318-320.
[9] R. Agrawal, T. Imielinski, A.S. Wami, Mining association rules between sets of items in large databases. In:Proc of the ACM.

2015 © The American Computational Science Society. All rights reserved.