Using Sequence Mining in Cloud Service to Design the Warning System for Construction Safety

Chun-Ling Ho, Tung-Chiung Chang

ABSTRACT: In the era of knowledge economy, to establish adaptive management and warning system can enhance performance depends on the foundation with using technology for decision makers and management team. In view of the fact that the labor force of the middle-aged and elderly people in Taiwan is increasing year by year, the manpower of the construction industry is gradually becoming middle-aged, which will be associated with the aging of the body of the phenomenon. The elderly workers in the construction safety education and its relative safety experience in warning services should be taken seriously. Therefore, this study takes the construction safety as the core and focuses on the warning design of the dynamic working safety for the warning index and pattern by the cloud service. In order to enhance the safety of pre-working and working safety, the study will collect the initial data of safety training history in the iWork cloud service, and carry out real-time data analysis and sequence mining to find out the warning mode of affecting safety behavior. Sequence mining will use the concept of the phenomenon in monitoring, and extract the events in order to establish the relationship between the time sequence and then establish the exact rule of sequence calculation. It is in order to adapt to the middle-aged workers in the construction warning service with security model.

KEYWORDS: Sequence Mining, safety warning, working training in construction, cloud service, mobile service.

I. INTRODUCTION

Considering that Taiwan's construction workers will gradually become more advanced and aging, it is an urgent task to enhance the safety working. The research focused on the concept of construction safety and early warning design on dynamic application of safety. Under the service of safety training system in cloud platform, the accurate rules of sequence score calculation were established by using the sequence mining to extraction events and data. To adapt to middle aged and older workers, it builds the safety warning service model.

According to the report of the major occupational disaster in the Labor Safety and Health Institute in recent three years, the proportion of workers suffering from the disaster at the age of 50 to 59 is the highest in terms of age structure in the employment construction industry for job

losses. That is accounting for 40.4% of the total number of people affected by the disaster. If the number of workers aged 40-49 is added, the proportion is as high as 70%.That is, the number of middle-aged workers over 45 who are employed in the industry is higher than 1 for their annual Accident Rate. This shows that there is a higher incidence of standardized accidents in middle-aged workers than other age groups. It highlights the importance of building early warning services for construction safety and their relative safety records.

In this study, the warning services is based on iWork that a construction safety cloud App service system with operation management function, under the cloud platform service and managers could easily and quickly access to information. Its functions include "Immediate Worker Management", "Safety Training Knowledge Management" and "Resilience Safe Worker Management". The warning system allows managers to deal with the safety situation of workers for the first time before the construction. So how to establish accurate automatic warning services will improve the construction safety indicators for the value.

II. RELATED WORKS

2.1 The safety construction of middle-age workers

The rapid aging of population in Taiwan, it means that the situation of "aging society" has become increasingly evident. It will also bring about changes in labor productivity in society as a whole. The labor force participation rate for the elderly, the aging of the labor force will be derived from the creativity and productivity and other negative effects (Cheng, J., Chiang, C.H., 2012 & Laflamme L., 1996) . In Taiwan, the number of middle-aged and elderly working people has been increasing year by year. However, the increase in the elderly working population is not sufficient to solve the problem of labor shortage. Therefore, regarding tertiary employment in the business service for 57.38% of the highest, industry is 34.66%, agriculture is 7.95%, of which 9.93% in the construction of labor force. This shows that the substantial increase in national income has driven the change of people's employment concepts. The young working people are reluctant to work in a more labor-intensive construction, and workers are gradually moving towards an aging population with high mobility in the construction. The middle-aged (over 45 years old) workers will be the major employment groups in the labor market in the next few years.

This will be accompanied by the aging body proposed by relevant phenomena such as decreased musculoskeletal motility, decreased acuity, deteriorated cardiopulmonary function, decreased sleep quality, and other negative changes(Rosenthal, A., et al., 2009).

According to the report of the major occupational disaster in the Labor Safety and Health Institute in the recent three years, the top three in the construction for job losses and disasters were crashed (20.1%), sting abrasions (17.3%) and objects flying (5.9%).As far as the age structure is concerned, the highest incidence among workers aged 50 to 59 is 40.4% of all those affected by the disaster. The number of middle-aged and over 45-age workers with Accident Rate is more than 1. It highlights the importance of educating the construction safety to older workers.

2.2 Data Mining

Data mining is an emerging technology developed for artificial intelligence and databases, and is also a widely used intelligent technology. Data mining is an important step in the process of knowledge discovery. Its purpose is to obtain meaningful relationships and differentiation of data. Its evolutionary process includes classical statistics, artificial intelligence and machine learning (Kleissner, C., 1998). Therefore, knowledge discovery uses a variety of techniques to explore valuable information or decision-making knowledge in the database. With the knowledge extracted, it is applied to the field of decision-making, prediction and estimation(Holsheimer, M. & Siebes, A., 1994). It shows that data mining is a largely automated process that uses statistical analysis to mine potential, non-obvious, unknown, potentially useful information from a large number of databases (Fayyad, U., G. Piatetsky-Shapiro, et al., 1996). The architecture of data mining system consists of five basic parts, includes database, knowledge base, data mining engine, attribute evaluation and interpretation, graphical user interface. Data mining is a technique for applying databases to discover hidden patterns and relationships in data and to make inferences to predict future results.

Data mining can generate many different analysis methods according to its commercial or application purposes and can extract knowledge from a large number of data bases through calculation rules. Therefore, the analysis methods can be divided into five categories, including Classification, Forecasting, Cluster, Association Rule and Sequential Pattern. Data mining also uses special algorithms to extract knowledge and discover unknown, potentially useful and ultimately understandable data patterns from the data base (Chen et al., 2003). The extracted knowledge can be used for decision-making or other purposes. Therefore, the main data mining techniques include Standard Statistics, Market Basket Analysis, Memory-based Reasoning, Cluster Detection, Link Analysis, Decision Tree, Artificial Neural Networks, Genetic Algorithms(Michael Berry, Gordon Linoff, 1997). In data mining process, SPSS and NCR companies gradually developed CRISP-DM (Cross Industry Standard Process for Data Mining) in 1990. The process includes defining Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Fayyad points to complete data mining or

knowledge mining with data selection, preprocessing, data transformation, data mining, interpretation and evaluation.

Therefore, this study integrates SPSS and Fayyad scholars to put forward the data mining steps and procedures for the safety training with warning module design. The structure of the content is in addition to pre-job "problem definition", the continuation is for the data selection (cloud safety training data), data pre-processing and transformation, data mining and early warning module.

III.   RESEARCH METHODS & RESULTS

In this study, it takes the safety education of construction site as the theme, and the middle-aged workers are as the object for construction safety and education. In the existing "iWork-mobile and real-time information integration services ", it applies data sequence exploration to design warning module from Safety history events.

Therefore, the study also assumes that workers are technology acceptance model and can conduct safety education through the popular mobile device smart phone. Therefore, the research in the data mining application is described as follows:

(1) With data visualization techniques, the raw data is presented in a variety of graphs to see the characteristics and trends of the overall database. There are four general ways of visualizing information: clustering, hierarchy, self-organizing network, and geographical landscape(Agrawal and Srikant, 1995). This article uses cluster visualization techniques to classify data according to their different attributes and present scatter plots in two or three dimensions.

(2) Use cluster analysis algorithm to find the cluster and characteristics of the safety training record database, and find out the relevance of each cluster.

(3) Find out the association rules hidden in the database by using the association rules algorithm. Apriori is the most representative algorithm for association rules. It was proposed by Agrawal et al in 1994 and derived from other algorithms for different situations or purposes. The study will use the Apriori algorithm to find frequent item sets and association rules in the database.

(4) The use of classification algorithm to produce the decision tree, and find the rules of each node, as the rules of the sequence model basis.

(5) To further identify the order in which things occur, the most powerful function of timing rules is usually used to identify the next action after some activities.

To sum up, this study first collected safety training information on the cloud platform, including the training module, security Incident module and fitness training recommendation module. With the use of cloud-based security history data, it continues to follow-up knowledge mining. That is to be two major stages as narrative exploration and forecasting explorations.

Therefore, the exploration process of overall research includes data purification, data integration, data selection and conversion, data mining, sequence model mining, knowledge prediction, as shown in Figure 1. So the required tasks for the following key points:

(1) Regularization: the values of average, minimum, maximum, total, percentage and others through excavated data.
(2) Related Mining: mining the data to meet certain conditions.

(3) Classification rules: to apply characteristics to classify information and to use these categories and rules for the new information.
(4) Cluster Rules: to find a series of rules for discriminating data to separate different categories and to group similar data of similar characteristics into one category.
(5) Predictive analysis: to find the laws and trends from the time series analysis and to be used as the judgment value to predict the future phenomenon.
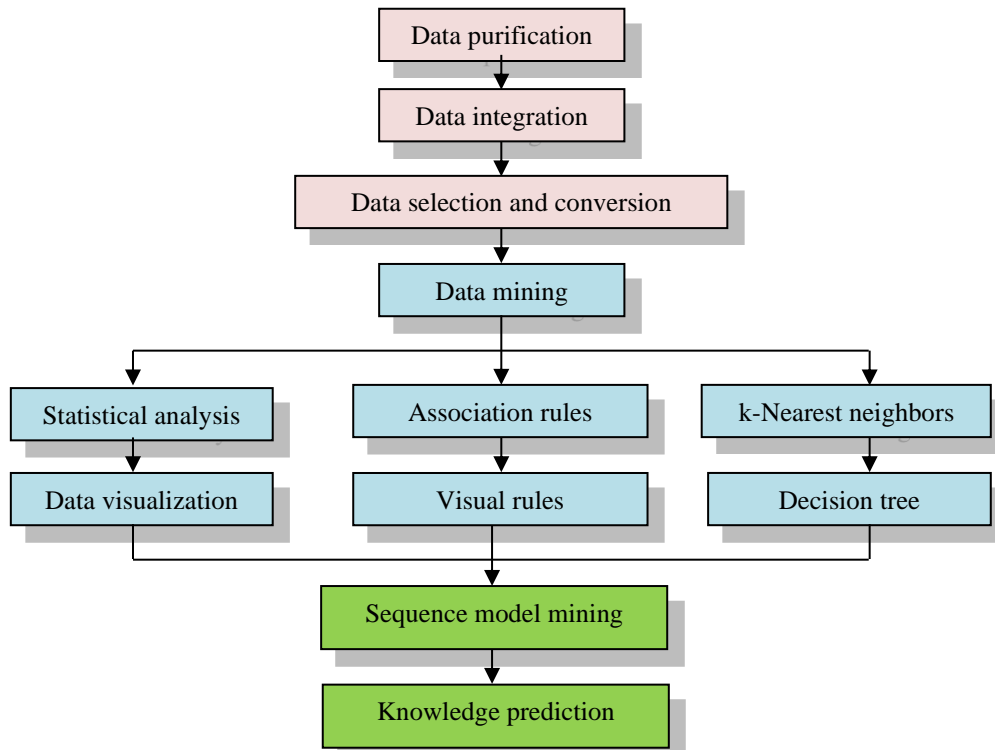

Figure 1: The mining process of the research

IV. DESIGN THE WARNING SYSTEM BY USING SEQUENCE MINING

This study will implement zero-burden of users and managers' proximity and mobility services through the appropriate cloud safety training of workers under the "information literacy" and "information application capability" in construction. In order to enhance the safety of pre-working and working, the research will collect the initial data of safety training resumes for real-time data analysis and sequence mining under the cloud service and find out the warning modes that will affect the safety behaviors.

The research focuses on implementing the safety warning module and investigating the data series carried out. Adopting the concept of "phenomenon monitoring", the key classification rules related to security training resumes by using the sequence classification model as the basis. Therefore, the contents in the research include:

(1) To develop "iWork App" with a historical records in the cloud and as a basis for data analysis to design safety warning system.

(2) To acquire valuable knowledge and characteristics related to safety training through data analysis as a parameter of safety warning system.
(3) To provide a sequential classifier model from safety history events and to integrate sequence styles and classification models.
(4) To propose a screening mechanism of classification rules that close to the characteristics of data set on the safety warning.
(5) To establish the rules of sequence score for improving the accuracy on automatic classification.
(6) To establish a safe warning service mode that is suitable for the middle-aged workers in construction.

This study uses sequence mining method, which is a kind of technology in data mining and can not only analyze the correlation between items, but also find out the order of the items rules. Sequence Mining was first proposed by Agrawal(1995). The main task of sequence mining is to find sequence patterns that occur in a specific order and frequently occur.

In order to understand the timing relationship of the events after the classification and grouping of the association rules, the study will use the sequence mining to find out the order relationship of the safety training and build the basis of automatic safety warning. Based on the sequential pattern from classification model, this study applies the mechanism of classification rule by screening and estimating the classification data through the scores of calculation pattern. In addition to the data mining mentioned above, the forecasting mining stage focuses on sequence mining. The detail work required is the collection of sequence databases, including the identification of all sequence patterns in the data set. The training data set is mainly used to build the model of the whole classifier, which will be used as a follow-up test data set to verify the comparison of the classification results of the established classifiers. That is as shown in Figure 2.
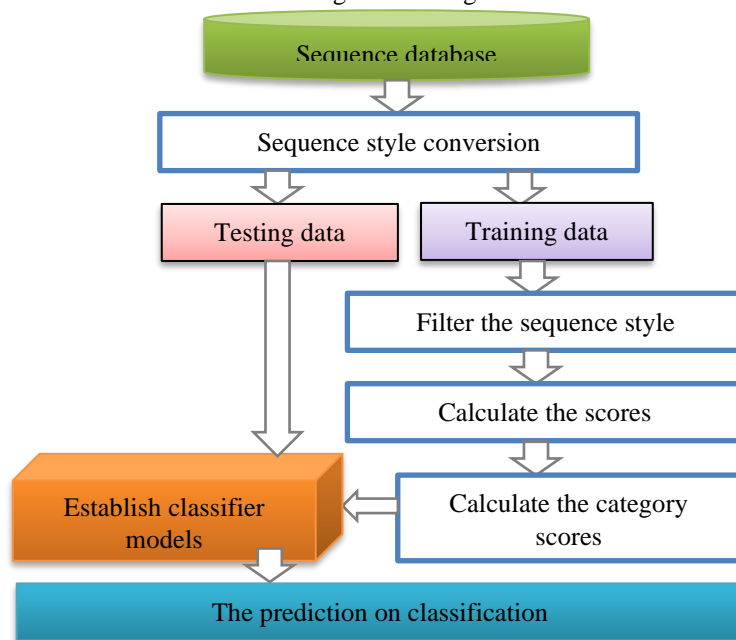

Figure 2: The prediction on classification by using sequence mining

V. CONCLUSIONS

The sequence mining in cloud is designed to express all the characteristic and correlations from safety training records through the association rules, including the basic background of workers, training and learning paths, safety records and incident records. The first phase of the correlative analysis was to identify high-frequency teams and the second phase was related to the high-frequency teams. In order to achieve the warning function, the sequential pattern analysis is applied to generate the data prediction function with time series through data relation ranking, high frequency data set setting, data conversion and maximum sequence calculation.

Therefore, this study takes the construction safety as the core and focuses on the warning design of the working safety for the warning index and pattern by the cloud service. It is in order to adapt to the middle-aged workers in the construction warning service with security model.

REFERENCES

1. Agrawal and Srikant (1995), Mining sequential patterns, In Proceedings of the International Conference on Data Engineering, 3-14.
2. Chen, Chiang and Kao(2003), Discovering time-interval sequential patterns in sequence databases, Expert Systems with Applications, 25(3), 343-354.
3. Cheng, J., Chiang, C.H.(2012), Phased Retirement for Older Workers in Taiwan, Journal of Family Economics, 3, 328–337.
4. Fayyad, U., G. Piatetsky-Shapiro, et al. (1996), From Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence, 17, 37-54.
5. Holsheimer, M. & Siebes, A., (1994), Data Mining: The Search for Knowledge in Databases, Technical Report CS-R9406, CWI.
6. Kleissner, C., (1998), Data Mining for the Enterprise, Thirty-First Annual Hawaii International Conference on System Sciences, 7, 295-304.
7. Laflamme L.(1996), Age-related accident ratios in assembly work: A study of female assembly workers in the Swedish automobile industry, Safety Science, 23, 27-37.
8. Michael Berry, Gordon Linoff (1997), Mastering data mining : the art and science of customer relationship management, Wiley Computer Pub.
9. Rosenthal, A., Mork, P., Li, M. H., Stanford, J., Koester, D., & Reynolds, P.(2009), loud computing: A new business paradigm for biomedical information sharing, Journal of Biomedical Informatics, 43(2), 342-353.