

# Application of Association Rules Big Data Analysis in Building Safety Production Supervision System

Ganglong Fan<sup>1</sup> and Hongsheng Xu<sup>1, a</sup>

<sup>1</sup>Luoyang Normal University, Henan Luoyang, 471934, China

<sup>a</sup>85660190@qq.com

**Keywords:** Big data; Safety production supervision; Association rule; Cloud platform; Pattern analysis

**Abstract.** This paper first discusses building of safety production supervision system in smart city by cloud platform and points out the existing problems. Then, this paper analyzes pattern of association rules data mining for big data background. The task of big data mining is to discover patterns hidden in the safety production supervision data, and its patterns are divided into two categories: descriptive model and predictive model. Finally, the paper presents application of association rules big data analysis in building safety production supervision system.

## Introduction

Information on the one hand to accelerate the safety of production accident information dissemination speed, resulting in an unprecedented upsurge of safety in production, on the other hand, but also to solve the problem of safety production has brought big data. Currently, the big data is at an alarming rate of penetration into more and more areas, such as electricity suppliers, retailers, IT companies, such as the successful case of large data is not uncommon. Big data applications in production safety, the basic function is from the safe production of vast amounts of data for the rules of the accident, to predict the future, so as to suit the remedy to the case, to effectively curb the accident occurred.

Enterprises safety standardization information library and enterprise storage safety standards of various aspects of enterprises safety standardization operation, enterprise safety standards of compliance, system operation state regulation, the enterprise daily safety standardization of basic information.

The task of data mining is to discover patterns hidden in the data, and its patterns are divided into two categories: descriptive model and predictive model [1]. The description pattern is a specification description of the facts that exist in the current data, and the general characteristics of the current data. The forecast model is based on the time as the main key parameter, for time series data, according to its history and current value to predict its future value.

With the extensive application of information technology, a large number of data generated and stored in various fields of information systems, the data showed an explosive growth. Data mining is in this "data explosion, lack of knowledge" in the case of the emergence of it. Data mining (mining Data) is a multi-disciplinary research field, which combines the database technology, machine learning, artificial intelligence, knowledge engineering and statistics and other disciplines. Data mining has been widely used in many fields, especially in Telecom, banking, transportation, insurance, retail and other commercial areas. Data mining, also known as knowledge discovery from data, specifically is from large amounts of data extraction people are interested in a non trivial, implicit, prior unknown and has potential uses of the model or knowledge. The paper presents application of association rules big data analysis in building safety production supervision system.

## Building of Safety Production Supervision System in Smart City by Cloud Platform

Business object full coverage, mainly for the daily work of the supervision and administration of production safety and business processes related to the business object phase information associated with a full coverage. Such as: enterprise basic information collection and all the information related to

the collection of all the information. Construction goal of the project is the unification of four, realize the data center, unified platform, electronic map and application system support, integrate and improve relevant existing business systems, in unified supporting platform, system and data migration and integration, realize the centralized operation and maintenance management, improve the system's ease of use.

Based on unified supporting platform, departments at all levels and enterprises in various business systems architecture and unified operation, system interoperability and data sharing, avoid because information system separate from the construction caused the information island and repeated investment [2]. The project in the city's unified deployment of the supporting platform and the construction of public application system. City Safety Supervision Bureau and the district administration of work safety, chemical industry park, and public, as is shown by equation(1), common function module unified built in supporting platform, to avoid duplication, the incompatibility between system and data exchange.

$$z(m, s) = \Psi(m, s)X(m) + \bar{v}(m, s) \quad (1)$$

In order to guarantee the high availability, high reliability and economy, large-scale cloud computing platform usually are stored in a distributed way to store data, distributed storage system with redundant storage to guarantee the reliability of data storage, is a copy of the same data stored in multiple copies. Distributed storage and traditional storage devices compared not only is a hardware, but a network equipment, storage equipment, server, software applications, common access interface, access network, and client program multiple parts of complex system. Each part of the storage device as the core, through the application of software to provide data storage and business access to external services.

The administrative license information database of the safety production, the storage of the license application for approval of the approval of the approval of the examination and approval of the preliminary examination, approval, and licensing, licensing and licensing and other information related to the information database, as is shown by equation(2) [3].

$$p(s(k) | a(k)) = \frac{p(a(k) | s(k))p(s(k))}{p(a(k))} \quad (2)$$

In this paper, the establishment of the security regulatory information service exchange center as the entire framework of the data and service center, is a single level of security regulatory information services to share the central hub of the exchange framework. And data of safety production and service users and producers is Exchange Center for the core nodes of the radiation so that the whole framework formed a data service center as the center, as is shown by equation(3), and production users for the nodes of a star topology structure. Safety supervision information service exchange center to manage and maintain the daily operation of the system platform through the safety supervision information service sharing exchange platform.

$$f_i(k) = N[\hat{v}_i(k) : 0, S_i(k)] = [(2\pi)^m | S ]^{-1/2} \exp(-\frac{1}{2} \hat{v}_i^T S_i^{-1} \hat{v}_i) \quad (3)$$

Unity based on cloud computing, networking, intelligent terminal technology support platform, based on software component reuse technology for the construction of management information system with high efficiency and high quality has important significance, has provided the basic framework of support for the realization of e-government [4]. Provide data modeling, report engine, workflow engine, software interface customization, system permissions role configuration, data coding metadata management, system security and performance management, and composite information visualization management. For the application of the system evolution, flexibility, high availability, safety, stability and other performance indicators can provide the basic protection; reduce

the cost of normalization construction, development and maintenance, realizing the highly reusable software components target.

Traditional government data update methods, generally pushed by the higher authorities to the division of grain size is more detailed, the power and responsibility of the lower level of the completion of a more explicit [5]. Therefore, the sharing level is only divided into the industry level, can not meet the demand, and need a shared exchange framework to meet the needs of distributed updates, as is shown by equation(4).

$$u_i = \frac{1}{\sqrt{\lambda_i}} Av_i \quad i = 1, 2, \dots, r \quad (4)$$

Comprehensive safety supervision system interface module development of micro channel reporting module, micro channel platform to increase micro report page to receive micro camera to upload pictures, and be able to query and reporting the results. Increase WeChat query information disclosure, agency publicity, credit information publicity module, data and communication safety supervision system.

### Pattern Analysis of Association Rules Data Mining for Big Data Background

With the wide application of network information and the rapid development of the, data mining technology as a kind of new science and technology, not only in the commercial field production and management are fully utilized, but also can be integrated into the program requirements, irregular and complex information environment. Its purpose is to use technical means, the remaining, gathering data information on the network to dig out, and then processed, edited into a collection of information for people to view, capture and capture applications.

SDN control plane and data plane separation design philosophy based on, this paper describes the data driven adaptive routing service customization framework, control plane as custom routing service logic decision center, each program operation of each components within the output results as a component of the input information, finally get the current state of the network can best meet the needs of users and ISP interests routing service customization project. Data plane does not have independent decision-making ability, can only be based on the control plane of the decision to make the routing function assembly.

The rapid expansion of the size of the data to data mining, especially for the Internet has brought great challenges to the data mining [6]. In order to make huge amounts of data can be calculated and some new high performance calculation method has gradually emerged. In 2004, Google launched the first to the MapReduce for non relational data management technology, as for data analysis and processing large parallel computing model, soon caused widespread concern in the academia and industry. In data mining technology oriented, scholars at home and abroad were also some preliminary exploration [7]. For example, according to traditional analysis software expansion difference and analysis of Hadoop weak characteristics, IBM is committed to integration of R and Hadoop. R is open source statistical analysis software, through the depth integration of R and Hadoop, as is shown by equation (5), the parallel framework of the calculation to the data.

$$F = \sup \{ \min [E(n_1), h(E(1))] \quad , \quad \min [E(n_2), h(E(2))] \} \quad (5)$$

Association rule mining is one of the most extensive and active research methods in the field of data mining. The original research motivation is to solve the problem of shopping basket analysis; the purpose is to solve the problem of finding the link between different commodities in the database rules. Association Rules refers to the interesting association or correlation relationships among a large set of data items. Commonly used algorithms are Apriori algorithm [8].

For large data, the data is massive, and the whole is more significant, so how to split the data while maintaining the overall characteristics of the data is a very important problem. Simple random

sampling is not suitable for solving this kind of problems. From the perspective of data granulation, according to a certain strategy, large data sets split for a number of small data sets were inferred through each small data set, and then the small data set that is formed by the fusion of a whole, which could reflect the nature of the whole large data sets. The essence is how to split the difficult decomposition in some data reflect the (a, a) the overall features. The preliminary study shows that for large-scale data granulation may be required to comply with the 3 criteria: approximation, transitivity.

The common clustering algorithms will be greatly challenged in the time complexity. Therefore, it is needed to satisfy some conditions of high dimension and large sample data clustering. The complexity is low, and the best is linear [9]. In this paper, by using the clustering algorithm, the algorithm proposed a LSC algorithm based on clustering, the algorithm is mainly choose  $p$  ( $n$ ) has the representative points as a landmark, and the  $P$  landmark of linear combination to replace the original data. Note that the common spectral clustering algorithm is usually used to express each data in all samples. Therefore, this method greatly reduces the complexity of the similarity matrix, namely from the secondary dropped to  $a$ , which incidentally also the solution of characteristic roots of complexity is reduced to linear.

## Experiments and Analysis

Data storage by using the popular big data storage technology for different data were stored, for example, you can use the ETL Technology to form data warehouse, storage in HDFS cluster medium for analyzing and mining the data; data analysis mainly provides analysis and mining tools, data of different take different data analysis techniques, can effectively provide information service, also in order to intelligent information service for people, but also need to provide interactive query service, the predictive warning services and report generation services. Finally is the data release, on the one hand can provide the corresponding inquiry service for the information user's different request, and on the other hand also may release the similar to the BI system visual data release report.

OLAP (online analytical processing) is fit to the timeliness requirements of this kind of data analysis has been proposed [10]. However, in the era of big data, the growth of data not only reflected in the fast, but also necessary in the analysis time period of incremental data on a large scale, which makes the traditional incremental single or small batch incremental machine learning highlights the limitations of technology, big data call more efficient online data analysis technology. The main challenge lies in: Based on large number of incremental data update mechanism and efficient algorithm.

Principles of the construction of the project is three full coverage, namely business process, to cover all the departments, business objects, are as follows: business process full coverage, full coverage, seamless, fine management mode is a scientific and efficient management, as is shown by equation (6),  $X$  is not only a kind of management idea and flexible use of the management methods of work.

$$X_{i+1,V}(m) = H_i^* X_{i,V}(m) + \sum_{r=1}^{q-1} G_{r,i}^* X_{i,r,D}(m) \quad (6)$$

Cloud computing center network resources sharing, a variety of applications will be hosted on the same network. In the fusion of the physical network to better the business of logic partition, network administrator how reasonable allocation of physical resources according to changing application requirements; in the equipment and different network environment how to reduce operation and maintenance workload and cost; now stage is the most important tool is the switch virtualization technology. Switch virtualization technology can create a logical entity with multiple physical entities, which can be used to compute, store, network, or application resources. The specific way is to integrate a number of network nodes (called horizontal integration), the virtual into a logical device, enhance the cloud computing center network availability, the node performance will greatly simplify the network architecture.

## Summary

Data center platform for the construction of a unified data reporting and registration system, a unified database and a unified data management and mining platform. The paper presents application of association rules big data analysis in building safety production supervision system, safety in production hidden danger investigation information database, intermediary agency information database, enterprise safety standard library, library of occupational health supervision, safety knowledge base, emergency resource data centralized management, unified data standards are formulated, and integrated independent of each sub system, realize the sharing of data and information.

## Acknowledgements

This paper is supported by the science and technology research major project of Henan province Education Department (13B520155) and Henan Province basic and frontier technology research project (142300410303).

## References

- [1] Sirivimol Thanchalatudom, Namfon Assawamekin, "Using Big Data Technology for Information Management in Hybrid Learning System", RNIS, Volume 12, pp. 179 ~ 182, 2013
- [2] Yulong Wang, Fengfeng Wang, Jing Guo, "A rapid detection algorithm for malformed H-DoS in the cloud platform", IJACT, Vol. 5, No. 9, pp. 474 ~ 481, 2013.
- [3] Yoon-Su Jeong, Kun-Hee Han, "Group Management Scheme for Security Assurance of Big Data in Cloud Environment", JNIT, Vol. 4, No. 8, pp. 561 ~ 568, 2013.
- [4] Niu Huiqing, Chen Hong, "Knowledge Recommendation of Industrial Clusters Based on Association Rules", AISS, Vol. 5, No. 5, pp. 603 ~ 611, 2013.
- [5] Kyoo-Sung Noh, Seung-Hee Lee, "A Study on Innovative Model for Communication System of Political Parties in Korea by using Big data", IJACT, Vol. 5, No. 13, pp. 342 ~ 348, 2013.
- [6] Rodrigo Rocha Silva, Celso Massaki Hirata, Joubert de Castro Lima, "Computing BIG data cubes with hybrid memory", JCIT, Vol. 11, No. 1, pp. 13 ~ 30, 2016.
- [7] Ill Woong Jang, Mu Wook Pyeon, Yang Dam Eo, Min Cheol Jeon, Sang Beom Lim, "Trend Analysis of Spatial Information using Big Data", RNIS, Volume 14, pp. 396 ~ 399, 2013.
- [8] Ping Dong, "Study Based on SPSS of the Organizational Factors of the Supervision Staffs", AISS, Vol. 5, No. 9, pp. 98 ~ 104, 2013.
- [9] Dawei Sun, Ge Fu, Xinran Liu, Hong Zhang, "Optimizing Data Stream Graph for Big Data Stream Computing in Cloud Datacenter Environments", IJACT, Vol. 6, No. 5, pp. 53 ~ 65, 2014.
- [10] Hassan Ansari, "Combining Association Rules and Clustering in Data Mining", RNIS, Volume 9, pp. 30 ~ 36, 2012.