

# DATA MINING AND CRITICAL SUCCESS FACTORS IN DATA MINING PROJECTS

Yun Chen<sup>1</sup>, Dengfeng Hu<sup>1,2</sup>, Guozheng Zhang<sup>1</sup>

<sup>1</sup>*Shanghai University of Finance & Economics, Shanghai, China. Email: [hdengfeng0212@sohu.com](mailto:hdengfeng0212@sohu.com).* <sup>2</sup>*Anhui University of Finance & Economics, Bengbu, China, 230135*

**Abstract:** Utilization of historical data to support evidence-based decision making, is leading many organizations to recognize that the effective use of data is the key element in the next generation of client-server enterprise information technology. To deploy data mining projects successfully, organizations need to know the key factors for successful data mining. This article concentrates on studying the data mining and the key factors for successful data mining, pointing out the progress of the data mining and crucial success factors (CSFs) currently exist in China.

**Key words:** CSFs, progress of the data mining, organization data mining.

## 1. INTRODUCTION

The major reason that data mining has attracted a great deal of attention in the formation industry in recent years is due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge, and can be used for applications ranging from business management, production control, and market analysis, to engineering design and science exploration. Some applications include marketing, financial investment, fraud detection, manufacturing and production, and network management. Presently, the majority of ODM

This project is supported by the Shanghai Shuguang Project of China under the grant No.05SG38.

---

*Please use the following format when citing this chapter:*

Chen, Yun, Hu, Dengfeng, Zhang, Guozheng, 2006, in International Federation For Information Processing (IFIP), Volume 207, Knowledge Enterprise: Intelligent Strategies In Product Design, Manufacturing, and Management, eds. K. Wang, Kovacs G., Wozny M., Fang M., (Boston: Springer), pp. 281-287.

projects focus on customers and profitability, customer retention and attrition management, risk forecasting and profitability analysis.

What is attracting many analysts to data mining is the relative ease with which new insights can be gained in comparison to traditional statistical approaches. Data mining is always a hypothesis-free approach, whereas most popular statistical techniques require the development of a hypothesis in advance. Data mining algorithms typically produce a much wider set of data types and make fewer assumptions about their distribution – or no assumptions at all.

Data mining outcomes include classification, clustering, prediction, estimation, and affinity grouping. Berry and Linoff [1] used the term techniques for market based analysis, memory-based reasoning, cluster detection, link analysis, decision trees and rule induction, artificial neural networks, genetic algorithms, and on-line analytic processing (OLAP).

This paper is conducted via Chinese periodical and organization survey to find the progress in China, and this text into five sections. The first section introduces the concept, content and the progress of data mining. The second section discusses the research of data mining in China, focusing on published papers. The third section overviews previous literatures, pointing out the CSFs in organizations. The fourth section concentrates in the application in organization. In this text, we use a questionnaire indirectly, for it is difficult in collecting data in organizations. Finally, the concluding section highlights the important of data mining to organizations.

## 2. RESEARCH OF DATA MINING IN CHINA

We can pass and search Chinese relevant periodicals, to understand China excavate the research situation of the respect in the data mining.

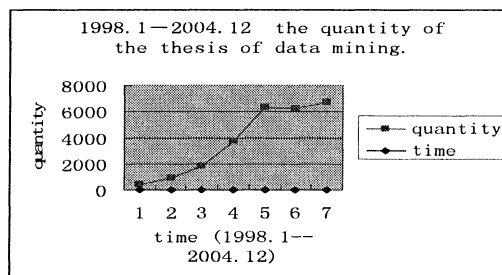


Figure 1. the quantity of relevant theses of research published.

Source of the materials: Chinese periodical network Chinese periodical full text date base

Now, we searched it from 1998 to 2004, figure 1 showing the quantity of relevant theses of research in the respect of the data mining.

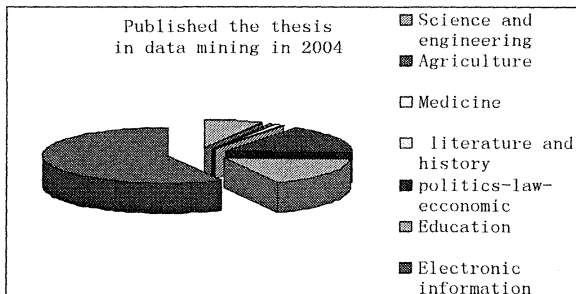


Figure 2. the quantity of relevant theses of research in each field Source of the materials: Chinese periodical network Chinese periodical full text data base

Based on “key words” and “abstract”, We have carried on sort research .In 1998, 430 papers on data mining published, and from 1998 to 2004 the thesis published has been increasing progressively, 6448 published in 2004.In China, the research in this respect has been developed in each field, especially in the field of electronic information, secondly in political economy. About the theoretical research of this, have not represented and carry on relevant actual application in China.

We continue carrying on search to the articles published in journals in management from 1999 to 2003. Based on “key words” and “abstract”, we have carried on sort research on data mining (table 1).

Table 1. The article published in data mining from 1998 to 2004

Journal name	quality
Chinese management science	35
Management engineering journal	39
System engineering theory and practice	33
Theory application of the system engineering	32
Theory application of the system engineering	29
Total	168

The research approach is a problem with most important research field, the mistake of the research approach will result in studying the mistake of the question directly. Based on Enrique Claver’s achievement, the research approach is divided into two kinds: theory study and empirical study.

The theoretical study is mainly to the concept, structure and describing, but not to the realistic systemic observation, namely focal point studied on concept research, but not datum nor observe. The theoretical study can be

divided into: the concept type studying, the proving type studying and the concept applied type studying .The concept type is mainly some definitions of frames, models or theories and provides relevant explanation. The proving type studying offers relevant action or in what kind of situation finishes a certain behavior mainly to study, and mainly as guidance of practice, what is focal point, how to do, but not why; the concept applied type studying is the combination that is above-mentioned two kinds of methods. The core of empirical study draws the meaningful conclusion of studying to observation of reality, or put forward new assumption or proposition to the new phenomenon, either examines or revises the existing theory. According to study course it may divide into positive research: case study, field study, field experiments study and laboratory experiments study.

We classify the study approach judging by 168 articles, get results shown table 2.

Table 2.the quantity of article of the research approach

Study approach	Theory study		Empirical study				
	Concept type	Proving type studying	Concept applied type studying	Case study	Field study	Field experim-ents study	Laboratory experiments study
Article quantity	102	48	16	2	0	0	0

From the above table2, the domestic study approach overweight on the research of the theory mainly, but empirical study little. The concept type study is generally introduction of new concept, theory and discussion, accounted for 60% of the total sample; At the same time some study overweight on the application in different fields of some theories, method and technology, accounted for 28%, which propose enterprise should do in data mining, and what is a key factor causing success of enterprise, and why should be done in this way, few people to study.

### 3. CRITICAL SUCCESS FACTORS (CSFS) IN DATA MINING PROJECTS IN ORGANIZATIONS

To deploy data mining projects successfully, organizations need to know the key factors for successful data mining [2]. The key factors can be critical success factors (CSFs), which is “the limited number of areas in which results, and if they are satisfactory, will ensure successful competitive performance for the organization [3]”. “ If these factors are crucial for the success of the system, then their failure would lead to the failure of the

information system [4]". Pinto and Slevin [5] wrote of the diversity of reported project successes in the information technology area.

Now we review which data mining projects are planned and implemented. Chung and Gray (1999) suggest utilizing 9 steps in data mining. Fayyad, Piatetsky -Shapiro, and Smyth [6,7] suggested 7 steps for successful data mining projects. Han and Kamber [8] arranged data mining as 7 steps

Based on this previews successful research, the architecture of a typical data mining process involves of three major elements: 1) Data mining: data mining engine, techniques, applications, and interpretation and using discovered knowledge.2) Data management: database, knowledge base, data warehouse, or other information repository, data cleaning and data integration steps.3) Information system (IS) project: dealing with data mining projects based on IS environment.

While no strong consensus exists in the project management literature on how to define success, a number of models and techniques have been developed to aid in the definition and measurement process. There is also a link between choosing appropriate CSFs to engage management's attention, and their support of the project activity. Slevin and Pinto [9] developed supporting criteria for ten factors of project success; budget of these is more important in project.

They perceive this critical path layout to be important and relevant to successful implementation. A clear mission, top management support, a defined schedule or plan, client consultation, right personnel, appropriate technology, and client acceptance are defined as CSFs. Communication, monitoring, and feedback are simultaneous process steps, with troubleshooting occurring as needed throughout the implementation.

Outsourcing must confirm a creditable organization to implement. Because systematic maintenance time is longer, so how take precautions against agreement fulfilled, it determines the success of the project [18].

A number of principles emerged from the case study that may be relevant to applications of the CSF process in other contexts: (1) Top management support is essential; (2) Substantial group process is required; (3) Data collection is a significant problem; (4) Managers need to be continuously encouraged to consider the relevance of CSF for them; (5) The use of the factors must be pushed down through the organization; (6) Survey development will most likely be needed [10]; (7) Develop realistic project schedules; (8) Match skills to needs at the proper time; (9) Establish and control the performance of contractors.

#### **4. QUESTIONS IN ORGANIZATIONS IN CHINA**

We once did a questionnaire to 120 enterprises (annual sales revenue more than 100million), designing following 8 questions of various fields according to the principle of the above. In the end, we gained 86 shares of effective questionnaire.

Enterprises have already built information system to a certain degree, accounted for 91%, such as handling official business, online information obtained. Among these enterprises, 63% plans to set up the data warehouse, and only 20% has already been set up, but others are being looked around. In enterprises having already set up, only 40% of the enterprises think that there is some benefit brought to enterprises, 30% of the enterprises believe it is very useful, other 20% of the enterprises think it as useless.

Question and answer about the function of enterprise's historical data, 85% of the enterprises think they can find out the reason of the achievement in the past according to the report forms; 10% of the enterprises think that can find the question that existed in the past. 5% enterprise think they can find future chance from it, as for how to find the chance, how to use tools to analyses, but the answers are not very clear. So, we can draw the conclusion, a lot of enterprises do not know the function of the historical data at all in China, and do not know how to find these real values of historical data.

Who must advocate building database or warehouse? 80% enterprise think it is required by leaders, and 15% was promoted by other enterprises, only 5% by markets. This conclusion is similar with "Computer Sciences Corporation" analysis. But referring to data mining, 85% enterprise do not know the concept and 95% not how to use for their business, though the story of the relative between peer and nappy has been for a long time.

In our another questionnaire to graduates from science and engineering and IT, we find 60% knowing little about it, and 20% knowing principle, but no man use it in practice.

From the up questionnaires, it is easy for us to think the reason that the application level of IT is lower, not soul of touching enterprises in China.

#### **5. CONCLUSION**

Except crucial successful factor mentioned above, promoting factors of application of data mining has several factors in our country. Firstly, the domestic study on the data mining is at the introduction stage in each field at present, there is still relatively few practical application; secondly, the scarce of talents in the data mining has influenced the application of the data mining in enterprises directly, Especially the backwardness of our course

offered has influenced the application with the data mining in enterprises directly; thirdly, how to build a data warehouse is key factor to the data mining in enterprise, but few of enterprises are engaged in the accumulation in this respect consciously. This will require organizations to further educate their employees if they want to narrow this knowledge gap, and to innovate not only by ways and means, but also through mechanism.

## 6. REFERENCES

1. Berry, M. and Linoff, G., *Data Mining Techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc., New York, NY, 1997
2. Jaesung Sim, B.P.A., M.P.A., M.S. Critical Success Factors in Data Mining Projects , Dissertation Prepared for the Degree of Doctor of Philosophy.
3. Rockart, J.F., Chief executives define their own needs, *Harvard Business Review*,57(2), March-April 1979, pp.81 - 93.
4. Zahedi, F., Reliability of information systems based on the critical success factors formulation, *MIS Quarterly* 11(2), June 1987, pp. 187 - 203.
5. Pinto, J. K. and Slevin, D. P. ,Project success: definitions and measurement techniques, *Project Management Journal* 19(1), February 1988, pp. 67-72.
6. Fayyad, U., Uthurusamy, R., Data mining and knowledge discovery in databases, *Communications of the ACM*. 39(11), November 1996, pp. 24-26.
7. Fayyad, U., Piatetsky Shapiro, G., and Smyth, P., The KDD process for extracting useful knowledge from volumes of data, *Communications of the ACM* 39(11), November 1996, pp. 27-34.
8. Han, J., and Kamber, M., *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, San Francisco, California, 2001
9. F.Warren McFarlan, Richard L. Nolan and Chen Guoqing, Seizing Strategic *IT Advantage in China*, 2003(4), Higher Education Press, P306-309.
10. Slevin, D. P., Stiemann, P. A. and Boone, L.W. , Critical Success Factor Analysis for Information Systems Performance Measurement and Enhancement, *Information Management* 21(3), October 1991, pp. 161-174.