BEST: International Journal of Management Information Technology and Engineering (BEST: IJMITE)
ISSN (P): 2348–0513, ISSN (E): 2454–471X
Vol. 7, Issue 12, Dec 2019, 15–22

APPLICATION OF WBS-RBS AND APRIORI ALGORITHM IN SOFTWARE PROJECT RISK MANAGEMENT

WANG FANDI

Department of Management Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China

ABSTRACT

This paper takes the software development project as the research goal, and uses the work-risk decomposition method: Work Breakdown System-Risk Breakdown System (WBS-RBS) to identify the risk of the software development project. The Apriori algorithm is used to mine the risk correlation and provide the countermeasure strategy for the risk management of the software development project. Firstly, the software development project WBS is constructed, and then the RBS is constructed. Based on this, the WBS-RBS matrix is established, and the association between WBS and RBS is constructed. Using risk identification data as the data source, the association rules mining is implemented by Apriori in SPSS Clementine12.0 software. The mining results are analyzed and sorted, and valuable association rules are obtained, and rationalization suggestions for software project risk management are proposed accordingly.

KEYWORDS: Software Project Risk Management, WBS-RBS, Apriori, & Data Mining

1. INTRODUCTION

With the development of information technology and the popularity of computers, people's demand and dependence on software is getting stronger and stronger. A good computer software can greatly help its users. Many companies and universities develop professional software based on their needs.

Unlike other projects, the final product of a software project is not a tangible item, but an intangible code. The development process does not involve the consumption of tangible materials, but the process of intellectual creation. The rapid update of technology requires developers to constantly learn new technologies to avoid the elimination of development results. These factors are the embodiment of high risk and high uncertainty in software development projects.

For the software development project risk management research, there is no system in China at present. In the preliminary discussion of risk management of software projects, Wen (2002) explored the main content and methods of risk management, and established a plan to avoid risks. In terms of management methods of software development projects, Zhang &Li (2003) conducted a comparative study on various methods. Hu (2004) applied the Bayesian network to the research of software risk management and provided new ideas. Zhang & Wang (2010) improved the WBS-RBS method to identify the risks of software projects. The rapid development of software development technology has made its application field more and more extensive, and the degree of specialization is getting higher and higher. Customer's functional requirements and expectations for software continue to increase, and the complexity of software is rapidly increasing, all of which pose greater risks to software developers. Song (2010) proposed to introduce data mining technology into the software project risk management decision system. After that, the application of data mining in software project risk management has gradually increased.

However, in the existing research, scholars generally ignore the relationship between risk factors of software projects. The many influencing factors for software development projects are independent of each other in existing research, but this is not the case. In order to solve this problem, this paper introduces the association mining algorithm in data mining method into the software project risk management process, identifies the risk factors of software project, and then uses Apriori algorithm to mine the relationship between risk factors in software development projects. It is not only necessary to analyze the risk factors that have a greater impact on software development projects, but also to analyze the relationship between the various influencing factors and to control the factors that have a greater impact on sensitive factors.

2. METHODOLOGY

The Apriori algorithm is a representative algorithm for association rule mining, which ranks among the top ten data mining algorithms. Association rule mining is a very important research direction in data mining, and it is also a long-standing topic. Its main task is to find out the inner relationship between things. In the software development project risk management, the Apriori algorithm is used to mine the relationship between multiple risk factors.

The association mode is the association rule existing between data items, which is the correlation between different items appearing in the same event, such as the correlation between different goods purchased by the customer in the same purchase activity. The item set $I=\{i_1,i_2,\cdots,i_n\}$, where i_n is called item; the transaction set $D=\{T_1, T_2, \cdots, T_p\}$, where T_p is called transaction, and also item collection, and $T\subseteq I$. An association rule is an implication relation of the form $X\Rightarrow Y$, where $X\subset Y$, $Y\subset I$, and $X\cap Y=\varnothing$. The support of rule $X\Rightarrow Y$ in transaction set D is the ratio of the number of transactions containing X and Y in the transaction set to the number of all transactions, denoted as support($X\Rightarrow Y$), i.e.:Support($X\Rightarrow Y$)=$|\{T:X\cup Y\subseteq T,T\in D\}|/|D|$.

The confidence of the rule $X\Rightarrow Y$ in the transaction set D is the ratio of the number of transactions containing X in the transaction set D, which is denoted as confidence ($X\Rightarrow Y$), i.e.: Confidence($X\Rightarrow Y$)=$|\{T:X\cup Y\subseteq T,T\in D\}|/|\{T:X\subseteq T,T\in D\}|$.

The core idea of the Apriori algorithm is that for a given transaction set D, the mining association mode problem is to generate association rules with support and credibility greater than the minimum support (minsupport) and minimum confidence (minconfidence) given by the user respectively.

The Apriori algorithm has two laws:If a collection is a frequent item set, all its subsets are frequent itemsets; If a collection is not a frequent item set, then all of its supersets are not frequent itemsets. The basic steps of the Apriori algorithm are as follows:

- Scan the database, calculate the support degree of each item set, compare it with the minimum support degree threshold, and go to the set of frequent 1 item set.
- Starting from the 2 item set, the frequent k item sets are generated from the frequent k-1 item sets by using the connection step and the pruning step.
- The loop ends when there is only one item set in the generated frequent k item set, or the frequent item set cannot be generated.

Work Breakdown System-Risk Breakdown System (WBS-RBS). The establishment of the WBS-RBS matrix is mainly divided into the following three steps: firstly, construct the WBS; then construct the RBS; finally, use the lowest-level job package set of the WBS as the matrix column, and use the lowest risk factor set of the RBS as the matrix row to establish

the WBS-RBS matrix, constructing the association between WBS and RBS.

3. SOFTWARE DEVELOPMENT PROJECT RISK IDENTIFICTION AND RISK CORRELATION MINING

3.1 Software Development Project WBS

For software development projects, based on the usual project work tasks and processes, the work breakdown structure WBS is constructed, as shown in Figure 1.

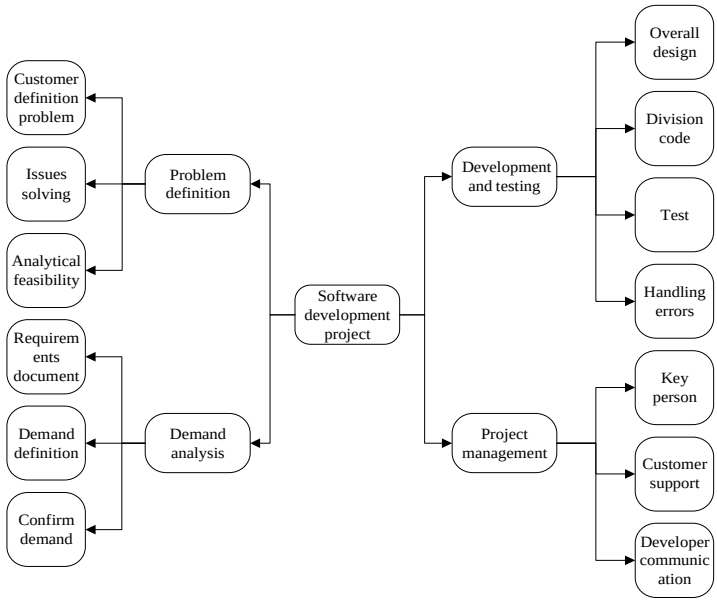

Figure 1: WBS Matrix of Software Development Project.

The explanation of software development project WBS is as follows:

- Problem definition and planning. The customer needs to have a specific definition of the problem to be solved by the software developed this time. After the customer explains its goal to the developer, the developer needs to understand and analyze its feasibility.
- Demand analysis. Developers should analyze and organize the problems according to user needs, form a requirements document, and send it to the user for confirmation.
- Software development and testing. Developers create code architectures based on existing code based on business needs and design, including some common foundation code. The programmer will analyze the specific design of the code according to the needs and find out where to change and add. At the same time, the project manager will discuss with the developer the criteria for defining the entire development process, defining the development and project management tools and guidelines used. Testers write test cases as needed, test the software and process the errors after the code is complete.
- Project team management. Project management exists throughout the software development process, including personnel selection, task assignment, task tracking, and status tracking. The formulation and implementation of the project plan etc. Project managers and developers work together to develop development standards and select the right development tools. It also includes human resource management for team members, motivating teams to complete tasks efficiently.

3.2 Software Development Project Disk Classification

According to the WBS in 3.1, the risk factors are summarized and classified by the following research on software project risk management:

Table 1. Software Project Risk Classification

Risk Classification	Description
Defining risk	Involved in the customer's definition of the problem is not accurate, the developer cannot understand or understand the risk of errors
Demand risk	Risks related to changes in the scope of software requirements, unclear customer expression, and incorrect analysis
Technical risk	Risks related to technology, code writing, testing, and modification errors
Management risk	Involving project team management, human resources, coordination and communication, task assignment and other risks

On the basis of Table 1, the various risks of the project are subdivided and each refined risk is numbered as shown in the following table:

Table 2: Project Risk Identification List

Risks	Number	Description
Defining Risk	DR-1	Customers cannot define their issues, making it impossible for developers to perform next steps
	DR-2	The definition of the customer's problem is not clear enough, the developer understands the error or cannot understand
	DR-3	Due to lack of relevant knowledge, developers cannot understand the problem defined by the customer.
Demand Risk	RR-1	Unclear or unresolved requirements lead to errors or different understandings, resulting in wasted time and resources
	RR-2	User needs change frequently, demand documents need to be updated continuously, and subsequent development cannot be performed based on documents.
	RR-3	There is no standard definition of requirements, resulting in a lack of standard specifications for different project requirements documents.
	RR-4	User demand is too high, resulting in a lack of feasibility or excessive resources, resulting in excessive risk
Technical Risk	TR-1	Lack of key technologies, making project functions difficult to achieve
	TR-2	Lack of key technical talents to solve bottlenecks in software development
	TR-3	By the self-owned attributes of the software development project, the knowledge and technology update is too fast, and the final delivery of the project is eliminated.
	TR-4	Programs coded by developers who lack technology often have a large number of errors, burdening later tests and bug fixes.
Management Risk	MR-1	The mobility of personnel has led to a lack of stable development environment for the project, resulting in wasted time and increased costs.
	MR-2	Lack of communication between members of the organization leads to inefficiency
	MR-3	The organizational structure is irrational, the task assignment is uneven, and the ability of the members of the organization cannot be fully utilized.

In order to be able to conduct a comprehensive assessment of risks, combined with the development of the company's software projects, a combination of software development tasks and risk identification, use the management tool WBS to build WBS according to the risk identification list (RBS) in Table1 and Table 2. The criterion is: 1 means there is risk, 0 means risk does not exist or the impact is small. At this point, the data in Table 3 provides the source data using the Apriori algorithm. Organize it into a Boolean matrix that the algorithm can handle and build a database.

Table 3: WBS-RBS Risk Matrix

		Problem Definition			Requirements Analysis			Development and Testing				Project Management		
		Customer Definition Problem	Issues Solving	Analytical Feasibility	Requirements Document	Demand Definition	Confirm Demand	Overall Design	Division Code	Test	Handling Errors	Key Person	Customer Support	Developer Communication
Defining Risk	DR-1	1	1	1	1	1	1	1	1	1	0	0	0	0
	DR-2	1	1	1	1	1	1	1	1	0	0	0	0	1
	DR-3	0	1	1	1	1	1	1	1	1	0	0	0	0
Demand Risk	RR-1	0	0	0	1	1	1	1	1	0	0	0	1	1
	RR-2	0	0	0	1	1	1	1	1	1	0	0	1	0
	RR-3	0	0	0	1	0	0	0	0	0	0	0	1	0
	RR-4	0	0	1	0	1	0	1	1	1	1	1	1	0
Technical Risk	TR-1	0	0	0	0	0	0	0	1	0	1	1	0	0
	TR-2	0	0	0	0	0	0	0	1	0	1	1	0	0
	TR-3	0	0	0	1	1	1	0	0	0	0	0	1	1
	TR-4	0	0	0	0	0	0	0	1	0	1	0	0	0
Management Risk	MR-1	0	0	0	0	0	0	1	1	0	1	1	0	1
	MR-2	0	0	0	0	0	0	0	1	1	1	1	0	1
	MR-3	0	0	0	0	0	0	1	1	1	1	1	0	1

3.3 Risk Knowledge Mining with Apriori

Save the data in Table 3 to the txt file with a comma or space as the data separator.

- Open SPSS Clementine 12.0 and drag the data source, data type, algorithm, and output into the work panel and link them in order. The "variable file" is the data source, the "type" is to filter and process the data in the data source, and finally input into the Apriori algorithm, and the mining result is output from the "table". Double-click the variable file, select the “Risk Data” file on the desktop, click Apply and confirm, and “Variable File” becomes “Risk Data.txt”.

- Define the "type" to determine the input and output data and data type. Since the data source Boolean matrix in the Apriori algorithm contains only 0 or 1, all data types are defined as flag types. Since the associations between fields are mutual, all fields (risk factors) can be either input or output.

- Set the Apriori algorithm, set all fields to the first and last items, and set the minimum rule confidence and the minimum condition support. Click Apply and OK when setup is complete. Run the algorithm to get the association.

From the mining results, we can find that not all association rules are meaningful. Rules for detailed risk factors belonging to the same type of risk are not considered. For example, the confidence and minimum support between TR-1 and TR-2 are satisfactory, but in reality, the two are independent of each other. Only consider the relationship between different types of risks. After eliminating the relationship between similar risks, the following valuable correlation rules are obtained with a minimum rule confidence of 80% and a minimum conditional support of 20%. The following are valuable association rules and explanations:

Table 4: Association Rules

Number	Last Term	First Term	Confidence (%)	Support (%)
1	MR-2 = 1	TR-1 = 1	23.077	100
2	RR-4 = 1	TR-1 = 1	23.077	100
3	MR-1 = 1	TR-2 = 1	23.077	100
4	RR-4 = 1	TR-1 = 1 and MR-2 = 1	23.077	100
5	MR-2 = 1	TR-1 = 1 and RR-4 = 1	23.077	100
6	RR-4 = 1	TR-1 = 1 and MR-1 = 1	23.077	100
7	MR-1 = 1	TR-1 = 1 and RR-4 = 1	23.077	100
8	RR-4 = 1	TR-1 = 1 and MR-3 = 1	23.077	100
9	RR-4 = 1	TR-2 = 1 and MR-1 = 1	23.077	100
10	MR-3 = 1	TR-2 = 1 and RR-4 = 1	23.077	100

Table 4: Contd.,				
11	RR-1 = 1	TR-3 = 1 and DR-2 = 1	30.769	100
12	RR-2 = 1	TR-3 = 1 and DR-1 = 1	23.077	100
13	RR-2 = 1	MR-3 = 1 and DR-1 = 1	23.077	100
14	RR-4 = 1	MR-3 = 1 and DR-1 = 1	23.077	100
15	RR-4 = 1	MR-3 = 1	46.154	83.333

- By rules 1, 2, and 3: When the technical risk of the project developer is large, the management risk and the demand risk will increase.
- By rules 4, 6, 8, and 9: When technical risks and management risks coexist, the probability of occurrence of demand risks increases.
- Management work becomes difficult by rules 5, 7, and 10: lack of key technologies and high user demand.
- By rules 11, 12: The definition of the problem is the premise of the demand analysis, and the increase of the risk and the technical risk increases the demand risk.
- Rule 13, 14, 15: Unreasonable organizational structure and uneven management of tasks will result in different levels of demand.

From the above association rules mining, we know that in the risk management practice of software development projects, we should:

Focus on the precise definition of the problem to be solved. As a first step in a software development project, the definition of risk caused by inaccurate problem definitions can lead to an increase in demand risk after that. When both the defined risk and the technical risk increase, the demand risk will increase greatly, causing an increase in the risk of activities after that.

Focusing on the level of demand risk, it has a great impact on the smooth progress of management activities. Rapid changes in requirements will require re-allocation of tasks and adjustment of organizational structure. Further analysis of the risk factors affecting the quality of the demand description key technologies, key personnel and the speed of knowledge renewal are important management and supervision objects, and have a profound impact on the success or failure of the entire software project.

Focus on solving practical problems associated with risk management during the software project requirements phase. Provide sufficient resources, master sufficient key technologies, and rationally arrange work tasks. Improve software development efficiency through full communication between developers and greater customer support.

4. CONCLUSIONS

This paper combines WBS-RBS with Apriori algorithm in data mining, identifies and analyzes the risk factors of software development projects, and mines the association rules between risk factors. Policy recommendations for future risk management of software development projects through analysis of association rules. Due to the nature of the software development project itself, it is necessary to accumulate past experience and collect relevant data. Data mining based on this will have a broader application space.

After the research and analysis, in order to do a good job in knowledge discovery and management, we must make breakthroughs in technology and innovate. As the premise of data mining, risk data is the basis of the whole process, so this requires enterprises to continuously accumulate development experience, summarize failure lessons, and constantly innovate knowledge discovery and management methods. At the same time, the software development process requires a large number of key technologies, relying on professional and technical personnel, so companies must pay attention to the training of relevant talents. Through the comprehensive application of WBS-RBS and Apriori algorithm, this research is a new idea to innovate the risk management of software projects.

REFERENCES

1. Wen, Y. (2002), "Risk Management of Software Projects", Journal of Computer Systems, Vol.:11, No.:2, pp. 72–74.
2. Zhang, Y., Li, S. (2003), "Comparison and Research on Risk Management Methods of Software Projects", Computer Engineering, Vol.:29, No.:3, pp. 91–94.
3. Kaur, P., & Vashisht, S. (2013). Evaluation of intrusion detection techniques and algorithms in terms of performance and efficiency through data mining. *Evaluation of Intrusion Detection Techniques and Algorithms in Terms of Performance and Efficiency through Data Mining*.
4. Hu, X. (2004). Research on software risk management based on Bayesian network. M.Sc. Thesis, China. Nanjing University of Science and Technology.
5. Ejaz, A. M. N. A., Khalid, A. S. F. A. N. D. Y. A. R., Ahmed, S., & Cheema, M. (2016). Effectiveness of Requirements Elicitation Techniques in Software Engineering Process: A Comparative Study Based on Time, Cost, Performance, Usability and Scalability of Various Techniques. *International Journal of Management, Information Technology and Engineering*, *4*(5), 23–28.
6. Zhang, Z., Wang, W. (2010),"Project Risk Identification Method Based on WBS-RBS Matrix", Project Management Technology, Vol.:8, No.:4, pp. 74–78.
7. Jain, V., Narula, G. S., & Singh, M. (2013). Implementation of data mining in online shopping system using tanagra tool. *International Journal of Computer Science and Engineering*.
8. Song, C. (2010),"Research on Risk Management Decision System of Software Project Based on Data Mining Technology", China New Technology and New Products, Vol.:2, pp. 24–24.