Multiple Association Rule Mining Algorithms for Supply Chain Finance Risk Prediction

Yuezheng Yang(✉)

Henan Agricultural Investment Financial Holding Co., Ltd., Henan, China
yeungyuehcheng@gmail.com

Abstract. As China's economy enters a new stage of development, the financial dilemmas faced by small and medium-sized enterprises, which are a major component of the national economy, are attracting more and more attention. In this paper, while reviewing a large amount of literature, the causes and correlations of supply chain financial risks are analyzed in a relevant way, and the corresponding data are generated by data mining, and the Apriori correlation algorithm and its related improvement algorithm are examined so as to obtain the relationships between risks, and corresponding risk prevention countermeasures are given based on the results of data mining.

Keywords: supply chain · financial risk · Apriori association method

1 Introduction

In data mining, Apriori association rule algorithm has been favored by many scholars due to its wide range of applications and easy implementation. Jin Xia [10] used the correlation rules in data mining to construct a supply chain risk identification model based on association rules, and evaluated it. This approach can effectively reduce the occurrence of risks and take effective measures to reduce the losses of enterprises after the occurrence of risks. With the advent of the era of big data, enterprises have applied data mining technology to the risk prediction of supply chain, and mined relevant data from all aspects of the supply chain and analyzed them for risk, so as to enhance the resilience and stability of enterprises.

Currently, many types of Ali and Jingdong are conducting their own supply chain financial services, and these banks can get some financial help by breaking their inherent financing deficiency through these channels. Supply chain financing has been used more and more in human production and life, thus promoting the efficient use of corporate funds and thus the sustainable development of the whole enterprise. Although the development of supply chain finance in China has made significant progress in recent years, in practice, the large number of its members and relatively cumbersome operation procedures have led to the gradual exposure of many problems in its development. For example, in the mode of equity financing, due to the relatively low disclosure of information among banks, the policyholder can create false information through cooperation

J. Yen et al. (Eds.): ICBIS 2023, AHCS 14, pp. 888–893, 2023.
https://doi.org/10.2991/978-94-6463-198-2_92

with third parties and obtain funds in different financial organizations through various ways. How to strengthen and improve the risk of supply chain finance is an issue that must be paid attention to when conducting financial activities in China at present.

2 Algorithm Process

Application association rule algorithm. The core of this method is to find a plausible and efficient rule by searching the entire data in the database and discovering events at the same time. The algorithm is a typical association rule that generates new candidates from successive joins and pruning, and computes them for plausibility and support [19].

(1) Minimum support and minimum plausibility

The degree of support is the probability that items A and B appear in the database at the same time: A + B = P(A + B), i.e., the probability that the set of B items appears in the database when the set of items A appears: Confidence (A + B) = P(B + A). The degree of support is an important indicator to judge the validity of data mining results. The higher the degree of support, the more frequently it can be proved; if the degree of support is low, it indicates that this conclusion is random and has no meaning. From a practical point of view, there is no reason not to pay attention to data mining results that lack support, and there is no need to adjust for this rare phenomenon. Therefore, the goal of minimal support is to exclude irrelevant correlation rules.

The degree of credibility is derived from the correlation rule, so it has a certain degree of confidence. Assuming that the result of data mining is A→B (80%), that is, event A infers event B with 80% confidence, it can be assumed that the probability of event B is 80% when event A occurs. As the confidence level increases, the confidence level of the rule also increases. The minimum support is a user-defined threshold of the association rule, which is the minimum confidence level of the association rule. In the association rule algorithm, if the conclusion obtained meets the minimum and minimal threshold, it is called a strong correlation criterion [20].

(2) Item set

An item set is composed of multiple items. In an item set, K items are called K-item set.

(3) Support calculation

The number of occurrences of all item collections in the repository is called "support degree". When the support degree of item set K exceeds the threshold value initially set, K is called a frequent item set and is called LK. If the support degree of item set, the support degree of item set A, and the support degree of item set A are known, the support and confidence degree of the related rule A→B can be found. The main codes are as follows.

```
def create_C1(data_set):
Create frequent candidate 1-itemset C1 by scaning data set.
Args:
data_set: A list of transactions. Each transaction contains several items. Returns:
C1: A set which contains all frequent candidate 1-itemsets
C1 = set()
for t in data_set:
for item in t:
item_set = frozenset([item]) C1.add(item_set)
return C1
big_rule_list = [] sub_set_list = []
for i in range(0, len(L)):
for freq_set in L[i]:
for sub_set in sub_set_list:
if sub_set.issubset(freq_set):
conf = support_data[freq_set] / support_data[freq_set - sub_set]
big_rule = (freq_set - sub_set, sub_set, conf)
if conf >= min_conf and big_rule not in big_rule_list:
data_set = load_data_set()
L, support_data = generate_L(data_set, k=3, min_support=0.2) big_rules_list =
generate_big_rules(L, support_data, min_conf=0.7) for Lk in L:
print "="*50
print "frequent " + str(len(list(Lk)[0])) + "-itemsets\t\tsupport" print "="*50
for freq_set in Lk:
print freq_set, support_data[freq_set] print
print "Big Rules"
for item in big_rules_list:
print item[0], "=>", item[1], "conf: ", item[2]
```

3 Supply Chain Financial Risk Prediction Example and Experiment

3.1 Experimental Parameters and Their Settings

In this paper, we obtain the risk events of 100 supply chain node enterprises in the past year, eliminate invalid and redundant data from the collected data, and calculate the missing data based on the industry average. If it comes from multiple databases, these databases need to be merged. The parameters contain S1 information risk S2 management risk S3 credit risk. Among them, S1 financial information risk contains: I11 distorted information of partners, I12 improper selection of IT system and software, I13 low level of information sharing, I14 computer failure; S2 financial management risk contains: I21 corporate culture difference, I22 improper selection of suppliers, I23 poor service quality, I24 poor customer management relationship; S3 credit risk contains I31 high debt ratio, I32 negative information, I33 uneven profit distribution of partners, I34 self-interested behavior of partners.

A professional supply chain risk assessment team is set up to sort out the supply chain risk events of the enterprise through the method of simulation experiment, that is, the experimental data shown in Table 1 can be obtained.

Table 1. Supply chain finance risk simulation data

Enterprise	Financial Risk Events
C1	I12,I13,I22,I24,I32
C2	I13,I14,I24,I33
C3	I12,I13,I14
C4	I23,I31,I32,I33
C5	I13,I24,I31,I32
C6	I14,I22,I24,I33
C7	I12,I14,I21,I22,I34
C8	I13,I24,I31,I32
…	…

Table 2. Apriori algorithm results

Rule	Support	Confidence
I2-> I3	50%	66.6667%
I3-> I2	50%	80%
I1-> I2	37.5%	75%
I1-> I3	37.5%	75%
I3-> I1	37.5%	60%
I2I1-> I3	37.5%	100%

3.2 Prediction Process

Assuming a minimum support of 0.3 (count degree of 3), then by scanning the above 8 transaction datasets. The frequent item set can be used to know the correlation between each Table 2.

For example, when event I2 occurs, then the support for event I3 to occur is 50% and the confidence level is 66.667%, so when event I2 is encountered, one should also be prepared for the occurrence of event I3.

3.3 Data Mining Implementation Methods

With the increasing maturity of data mining technology, data mining tools are also developing, and various data mining tools are emerging. Generally speaking, data mining software is divided into two main categories: professional data mining tools designed and used by professional companies for specific purposes, and general data mining tools used by general data mining researchers and ordinary enterprises. Specialized domain mining tools are mainly used to quickly solve problems in a specified domain and perform

relevant computational processes based on actual business results and data structures. However, the cost of such targeted development tools is relatively high. For example, IBM has developed an advanced scouting system for NBA players to help analyze data about each player or game. Coaches can decide which plays to use based on the data analysis results. Generic data mining tools have good compatibility, are used in a wide range of market sectors, and have more routine data processing. Users of generic data mining tools can usually write logical business algorithms according to their actual needs and then deploy them to the tool platform to implement data processing.

4 Conclusions

The confidence level for I14 (computer failure) to I13 (poor information sharing) is 67%, which means that if there is a (computer failure), there is a 67% probability (poor information sharing). Based on this correlation rule, we found that there is a high probability that there is a valid rule that there is a low degree of information sharing if there is a computer failure. If the information sharing is not high and the customer is not managed properly, then the limitation of the industry can be inferred, and the information sharing will be lower when the industry is constrained, but the customer does not fully understand the constraint of the industry, which will lead to the deterioration of the relationship between customers. So this rule is valid. That is, an appropriate credit crisis can be inferred from this association when the members of a firm appear to be self-interested. Therefore, it is necessary for the selection of collaborators, once they find their own advantages, to take precautions to avoid suffering more harm again and to find a suitable partner in the shortest possible time.

References

1. Jong-yoon Chang. Reading supply chain finance [M]. People's University of China Press. 2019.
2. Xueling Zhong, Fang Hou, Hongxia Zhang. Python financial data mining [M]. Higher Education Press. 2020.
3. Zhu Meimei. Analysis of supply chain finance model innovation and risk control based on blockchain technology [J]. Business News, 2021(35): 92-94.
4. Demica. Supply chain finance: a third report from Demica [R]. London, UK, 2009.
5. Bi Jiaxin. Supply Chain Finance: Emergence Motivation, Operation Model and Risk Prevention [J]. North China Finance, 2010(03): 20-23.
6. Xuejian, Li Rifei. Research on the prevention and control mechanism of credit risk diffusion in supply chain finance [J]. Financial Development Research, 2017(2): 63–67.
7. Tomkalil. Big data is a big deal, 2012, https://www.whitehouse.gov/blog/2021/03/29/big-data-big-deal.
8. Wang Jiao. The application of data mining methods in supply chain finance risk prevention [J]. Logistics Technology, 2013, 32(21): 380-382.
9. Xie Jianglin, He Yiqing, Chen Tao. Application of data mining in supply chain finance risk control [J]. Journal of Nanchang University (Science Edition), 2008(03): 278-281.
10. Jin Xia. Research on supply chain risk identification and assessment [D]. Lanzhou Jiaotong University, 2015.