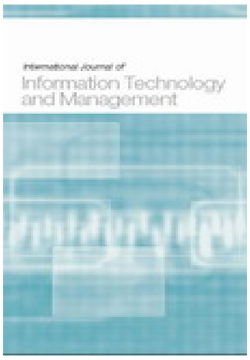


International Journal of Information Technology and Management

ISSN online: 1741-5179 - ISSN print: 1461-4111
https://www.inderscience.com/ijitm

Research on evaluation of network payment security

Xiang Xie, Yingte Wang, Ying Cui

DOI: 10.1504/IJITM.2023.10055154

Article History:

Received:	09 January 2018
Last revised:	05 May 2018
Accepted:	30 March 2019
Published online:	05 April 2023

Research on evaluation of network payment security

Xiang Xie*, Yingte Wang and Ying Cui

School of Economics and Management,
Beijing Jiaotong University,
No. 3 ShangYuanCun, Beijing 100044, China
Email: xxie@bjtu.edu.cn
Email: 18120623@bjtu.edu.cn
Email: 16125351@bjtu.edu.cn
*Corresponding author

Abstract: This paper collects data about network payment fraud through a questionnaire and then analyses the data. We propose a prediction model based on the hybrid support degree apriori algorithm. First, is to find out the factors that are closely related to the success of cheating. Then, we find for the nodes of the decision tree and their importance by using the ID3 algorithm to construct the decision tree. We then construct and verify the evaluation index system of the network payment from the user's point of view while considering the results of association rules and decision tree, and combined with principal component analysis and text mining results. Finally, we determine the weight using entropy method. With the proposed model, we can determine the probability of being deceived according to the situation of different people. This not only reminds people to be vigilant but also provides a reference for the community.

Keywords: apriori algorithm; ID3 algorithm; internet payment security; styling; evaluation system; principal component analysis; entropy method.

Reference to this paper should be made as follows: Xie, X., Wang, Y. and Cui, Y. (2023) 'Research on evaluation of network payment security', *Int. J. Int. J. Information Technology and Management*, Vol. 22, Nos. 1/2, pp.110–126.

Biographical notes: Xiang Xie is an Associate Professor of Information Management and Information System at the Beijing Jiaotong University, China. He received his BA in Economic Information Management from the Beifang Jiaotong University in 1999, and PhD in Management Science and Engineering from the Beijing Jiaotong University in 2009. His research interests fall into the area of data science and scientific evaluation.

Yingte Wang is a postgraduate student in the School of Economics and Management at Beijing Jiaotong University, China. She has completed her Bachelor of Art in Information Management and Information System at Qingdao University, Qingdao in 2018. Her current research interest: information security and information system.

Ying Cui is a postgraduate student in the School of Economics and Management at Beijing Jiaotong University, China. She has completed her Bachelor of Art in Information Management and Information System at Hebei University of Technology, Shijiazhuang in 2017. Her research interests include network payment and text mining.

This paper is a revised and expanded version of a paper entitled 'The construction of network payment security evaluation index system based on users' presented at International Conference on Logistics, Informatics and Service Sciences, Japan, 24–27 July 2017.

1 Introduction

According to the 39th 'China Internet Development Statistics Report' published recently by China Internet Network Information Center, which shows that as of December 2016, the scale of Chinese internet users reached 731 million, the equivalent of the total population of Europe. The internet penetration rate reached 53.2%, which goes beyond the global average of 3.1 percentage points, the Asian average of 7.6 percentage points. There are 42.99 million new internet users in total, the growth rate is 6.2%. The rapid development of the internet brings vigorous development of online payment, especially the habit of the mobile payment offline. According to the report of December 2016, China's mobile phone users reached the scale of 695 million, the growth rate of more than 10% for three consecutive years. Mobile internet users grow rapidly, reaching 469 million, the annual growth rate of which is 31.2%, the proportion of internet users of online payment grows from 57.7% to 67.5%. The rapid penetration of mobile payment to offline payment, greatly enriched the payment scene. There are 50.3% of internet users using mobile phone payment.

The development of online payment, gradually changing people's habits, not only brought great convenience for people, but also brought a lot of payment security issues, network fraud (Li and Tang, 2013). According to 2016 mobile payment security survey issued by China Unicom, telecommunications fraud cases continued high and the incidence of consumer damage continued to rise in 2016. About one-quarter of respondents said they had suffered a phishing network fraud, which increased 11 percent over 2015.The frequent occurrence of network payment security issues brought troubles to people at the same time, reduced the degree of trust in the payment platform. Therefore, it is necessary to study the payment of security. It is not only good for users, but also good for the payment platform.

Because the main research object of this paper is user, we need to collect user data. We mainly use questionnaires to collect user data. The experimental data of this article is mainly from the questionnaires issued. In this experiment, a total of 500 questionnaires were distributed and 400 valid questionnaires were recovered. Of these, 350 were used for analysis and 50 were used for verification of experimental results. We can see the related indicators of the questionnaire in the following analysis process. Firstly, we clean the collected data, mine the association rules of data using apriori algorithm and ID3 algorithm. Then we develop evaluation index system on the basis of the association rules, to better judge the user behaviour. Finally we use principal component analysis to determine the evaluation index, and the entropy method or expert scoring method to determine the weight of the index.

2 Literature review

Since Tim Berners Lee, a computer scientist in the UK, first opened the World Wide Web server for commercial in 1991, and the first successful online retail transaction in 1994 to ensure customer safety. Worldwide, online trading and payments develop rapidly. A typical e-commerce transaction consists of three phases: the information search phase, the ordering and payment phase, and the logistics and distribution phase. Online payment is the most crucial in three stages because once the online payment is completed, the logistics distribution is a matter of course, which means that the completion of a complete online transaction. If online payment is not conducted, e-commerce will not be able to enter the real trading phase. Therefore, online payment has become the top priority of the transaction. However, the attendant is also a very serious security issues. Especially with the popularisation of the smart phones and 3G, 4G network, the rapid development of mobile payment process becomes more convenient, but the payment security is still not optimistic.

At present, on the fraud payment in the network, many sites, companies have done a related study on the fraud data from the perspective of large data. According to 2016 mobile payment security survey report released by China Union Pay, common fraudulent practices include social account theft, SMS Trojan links, cheat SMS verification code, phishing scams, sweep unknown two-dimensional code, unidentified wireless network, etc.

2.1 Internet payment security

In the theoretical research, Xu (2014) mainly pointed out some problems about the network payment security and studied electronic payment security technology. Yuan (2014) mainly analysed the importance of payment security in the life cycle of an account and the development of payment security technology and threats, studied payment security regulation policies and the status, and finally put forward the problems and suggestions for the development of payment security in our country. Li's (2014) research included mobile payment security protocols, such as passwords, digital certificates, authentication and so on. Xu (2018) analysed the relationship between the three party payment platform and the online shopping consumer from the perspective of game theory. Third-party payment is an effective way to solve the security problem of e-commerce payment, which can well promote the healthy development of e-commerce.

In terms of technology, Yang (2015) put forward the mobile payment security application solution based on the characteristics of NFC technology system and secure payment through a detailed analysis of the existing mobile payment system. Liu et al. (2018) proposed a mobile payment security solution based on NFC by analysing the existing mobile payment system and combining the features of NFC and secure payment.

On the aspects of cybercrime, Yang (2015) analyses the current situation, characteristics and causes of the telecommunications fraud and puts forward corresponding countermeasures. Yuan (2014) analyses the payment security technologies and threats driven by the development of internet finance, studies the domestic and international payment security regulatory policies and the status quo, and finally puts forward the problems and suggestions for the development of payment security in our country. Jin (2014) mainly studies the case from the perspective of criminal law in regard to the current popular card-less crime.

2.2 Correlation algorithm

Apriori algorithm is a frequent itemsets algorithm for mining association rules. Its core idea is to mine frequent item sets through two stages: candidate generation and closed down detection. And the algorithm has been widely applied to various fields such as commerce, network security. In the study of apriori algorithm, Wei et al. (2015) proposed improvements to apriori algorithm based on Bigtable and MapReduce. Zhong and Liu (2016) proposed an improved apriori algorithm combining the compression matrix with apriori algorithm based on compressed matrix.

The ID3 algorithm is a greedy algorithm used to construct a decision tree. On the research of ID3 algorithm, Du (2014) has done research on the basic concept, process and application of ID3 algorithm. The ID3 algorithm and C4.5 algorithm were analysed and compared in detail by Miao and Yu (2014).

Principal component analysis mainly uses the idea of dimensionality reduction to convert multiple indicators into a few comprehensive indicators, each of which can reflect most of the information of the original variables, and the information contained therein does not overlap. For the principal component analysis, the relevant scholars also made a lot of research and application. Ruan (2014) proposed a quantum PCA algorithm based on the related theories and methods of quantum information and inspired by the basic idea of arithmetic coding. A face feature-coding scheme is designed to further reduce the feature space after dimensionality reduction.

Entropy method mainly uses the theory of entropy to determine the weight of rating index. Wang and Guo (2017) thought it was better to determine the index weights by entropy method with the original index data. ‘The construction and evaluation of the index system for the fusion of producing cities in national high and new technology zones’ is the application of entropy method in practice by Wang et al. (2014).

2.3 Evaluation index system

In terms of the construction of evaluation index system, Chi and Li (2014) established the evaluation index system of the final construction by combining principal component analysis and information entropy. Xu et al. (2015) reviewed the status of enterprise innovation capability evaluation and evaluation methods at home and abroad, and build a technical evaluation of enterprise technology innovation index system and carried out Empirical Research according to the actual situation of enterprises. Huo (2016) designed a comprehensive low carbon grid quantitative evaluation method, and based on the statistical data of some domestic power grid enterprises index system and evaluation methods are analysed.

3 Algorithm

3.1 Apriori algorithm

Related concepts:

1. Support: probability that a transactions contains $X \cup Y$, called $X \Rightarrow Y$

   $$Support(X \Rightarrow Y) = \left|\{T : X \cup Y \subseteq T, T \in D\}\right| / |D|$$

   X, Y is the thing, T is the thing that X, Y happen at the same time, D is the total thing.

2. Confidence: the confidence of the rule $X \Rightarrow Y$ in the transaction set is the ratio of the number of transactions containing X and Y to the number of transactions containing X, called $P\,(Y \mid X)$

   $$Confidence(X \Rightarrow Y) = \left|\{T : X \cup Y \subseteq T, T \in D\}\right| / \left|\{T : X \subseteq T, T \in D\}\right|$$

3. Item: the collection of items.
4. K-item set: a collection contains k items.
5. Frequent item: item which meet the minimum support.
6. Self-joining Lk – 1: generate length k candidate item from length (k – 1) frequent items, and test the candidates against DB.
7. Pruning: if the (k – 1)-item subset of a candidate k-item set is not in Lk – 1, the candidate cannot be frequent and can be deleted.

The basic idea of the apriori algorithm is to use a iterative method of layer-by-layer search. The k-item set is used to search for (k + 1)-item sets. The keys of it are the connection and pruning. The following is the process:

$$D \to C1 \to L1 \to C2 \to L2 \to C3 \to L3 \to \ldots \to Ck \to Lk \to \ldots \geq Cm-1 \to Lm-1$$

So go on until you cannot find frequent m items.

The basic framework of the apriori algorithm is described as below:

```
Input: Database D of transactions, minimum support threshold min-support method
Output: L1 frequent items in D
L1 = {Large 1-itemsets};
For (k = 2; Lk-1 ≠ ∅; k++) do begin
Ck = apriori_gen(Lk – 1, minsupport)
Forall transaction t ∈ D do begin
C1 = subset(Ck,t);
Forall Candidate C ∈ C1 do
C.Count++
End
Lk = {C ∈ Ck|C.Count >= minsupport}
End
Answer = UkLk
```

3.2 ID3 algorithm

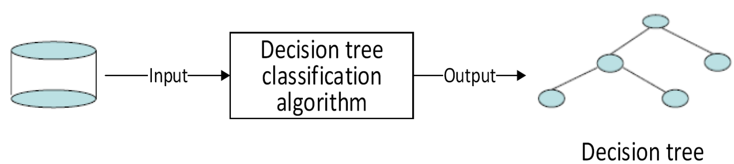
Related concepts:

1. Attributes: category field (continuous value attribute need to be discretised in advance).
2. Information gain: all attribute assumptions are category fields can be applied to numeric fields after modification.
3. IBM intelligent miner: can be applied to categories and numeric fields.
4. Class label: classification logo: each sample is a tuple, and a property of a tuple is used to determine the class of training samples.

The ID3 algorithm is a top-down recursive method of constructing a hierarchical tree, all training samples are in the roots. The information gain theory in information theory is used to find the fields with the largest information quantity in the data set, establish a node of the decision tree and the branches of the tree according to the different values of the fields. A decision tree can be created in the process of repeating the lower nodes and branches of each branch subset. Divide the samples recursively based on the selected attributes and stop the division when the following three cases are encountered:

1. all samples of a given node belong to the same class
2. no attributes can be used to segment the data
3. there is no sample of a branch of the attribute.

Figure 1 Calculation process (see online version for colours)



Calculation process as follows:

1. Calculate the expected information

   Let S be the set of training samples, where the class labels for each sample are known. Let S contain a sample of C_i (the number of samples belonging to C_i). The probability that an arbitrary sample belongs to class C_i is / s. s is the total number of objects in set S. The desired information for a given sample classification m is:

   $$I(S_1, S_2, \ldots, S_m) = -\sum_{i=1}^{m} \left(\frac{si}{s}\right) * \log_2 \left(\frac{si}{s}\right)$$

2. Computes the entropy of any attribute A in a tuple

   $$E(A) = \sum_{j=1}^{v} \left(S_{1j} + \cdots + S_{mj/s}\right) I\left(S_{1j}, \ldots, S_{mj}\right)$$

The value of attribute A: $\{a_1, a_2, \ldots, a_v\}$. A can divide S into subsets: $\{S_1, S_2, \ldots, S_v\}$, where S_j contains those samples with A values a_j in S. Let S_j contain S_{ij} samples of class C_i, The expected information according to this division of A is called the entropy of A.

3 The information gain of the division on attribute A

$$Gain(A) = I(S_1, S_2, \ldots, S_m) - E(A)$$

$$I(S_1, S_2, \ldots, S_m) = -\sum_{i=1}^{m} \left(\frac{si}{s}\right) * \log_2\left(\frac{si}{s}\right)$$

$$E(A) = \sum_{j=1}^{v} (S_{1j} + \cdots + S_{mj/s}) I(S_{ij}, \ldots, S_{mj})$$

4 Create a decision tree

Calculate the desired information for a given sample classification and entropy of each attribute. Create a decision tree based on these.

3.3 *Principal component analysis*

The principal component analysis method is the analysis and statistical method of transforming a plurality of related elements into several unrelated comprehensive indicators Tang et al. (2016). The composite index may contain many overlapping information. The principal component analysis reduces the dimension of the original index under the principle of least loss of information, and saves some irrelevant indexes. It turns the original more indexes into the index that can reflect the research phenomenon of less comprehensive index, this can simplify the complex research, to ensure the accuracy of the premise to improve the efficiency of the study.

1 Find the matrix of standardised data metrics. Suppose the number of evaluation indicators is n, the number of person is m, the data matrix is X, x_{ij} is the j^{th} number of the i^{th} individuals.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$$

2 Correlation coefficient matrix

$$r_{ij} = \frac{\sum_{k=1}^{n} |x_{ki} - \bar{x}_i||x_{kj} - \bar{x}_j|}{\sum_{k=1}^{n} (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^{n} (x_{kj} - \bar{x}_j)^2}$$

r_{ij} is the correlation coefficient between the i^{th} indicator and the j^{th} indicator of standardised data.

3 Find the eigenvalues and eigenvectors of the correlation coefficient matrix, as well as the variance contribution rate and cumulative contribution rate corresponding to the eigenvalues.

$$c = (c_1, c_2, \ldots, c_q)$$

4 Take the corresponding q eigenvectors and normalise them.

$$A = \left[e_f - \min(e_f)\right] \big/ \left[\max(e_f) - \min(e_f)\right]$$

e is the eigenvector, $f \in [1, q]$.

5 The overall contribution rate of each indicator

$$P = C * A^Q \Big/ \sum_{i=1}^{q} C_i = (p_1, p_2, \ldots, p_n)$$

6 Do the normalisation to P

$$W_j = \frac{p_j}{\sum_1^n p_j} = (w_1, w_2, \ldots, w_n)$$

w_j is the weight of the j^{th} indicator.

3.4 Entropy method

The concept of entropy stems from thermodynamics and is a measure of the uncertainty of the state of the system. In information theory, information is a measure of the degree of systematic order. Entropy, on the other hand, is a measure of the degree of disorder in the system. The absolute value of the two is equal, but the sign is opposite. According to this property, we can make use of the inherent information of each scheme in evaluation to obtain the entropy of each index through entropy method. The smaller the entropy, the lower the degree of disorder of information, the greater the utility of information, the bigger of the weight of the index.

Entropy is a measure of uncertainty. If we use p_j to represent j information uncertainties (that is, probabilities of occurrence), then the uncertainty information of the entire information (provided with n) can also be expressed as follows.

$$S = -k \sum_{j=1}^{n} p_j \ln(p_j)$$

This is entropy. Where K is a positive number, when the probability of occurrence of each information is equal, namely $p_j = 1/n$, the value of S is the largest, and entropy is the largest.

Through the factor analysis of 250 respondents, we set $m = 250$ and $n = 13$, x_{ij} is the factor score of the i^{th} respondent's j^{th} factor, where $i = 1, 2, 3, 4, 5, 6, \ldots, 250$, $j = 1, 2, 3, 4, 5, 6$.

1 Standardise the raw data

Forward indicators: $x'_{ij} = (x_{ij} - \bar{x}) \big/ S_j$

Reverse indicators: $x'_{ij} = (\bar{x} - x_{ij}) \big/ S_j$

x_{ij} is the i^{th} sample and j is the original value of the indicator, x'_{ij} is the normalised indicator value, and $\overline{x}$ and S_j are the average and standard deviation of the j^{th} indicator, respectively. Due to the use of logarithm in the entropy method, the normalised value cannot be used directly. To reasonably solve the effect of negative numbers, translate the normalised values:

$$Z_{ij} = x'_{ij} + A$$

Z_{ij} is the translated value, A is the translation amplitude.

2 n is the number of samples (cities), m is the number of indicators and P_{ij} is the weight.

$$P_{ij} = Z_{ij} \Big/ \sum_{i=1}^{n} Z_{ij} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m)$$

3 The entropy of the j^{th} indicator

$$e_j = k \sum_{i=1}^{n} P_{ij} \ln P_{ij} \quad \left(k = \frac{1}{\ln n}, e_j \geq 0 \right)$$

4 The difference coefficient of the j^{th} indicator

$$g_j = 1 - e_j$$

5 Calculate the weight of the j^{th} indicator

$$W_j = g_j \Big/ \sum_{j=1}^{m} g_j \quad (j = 1, 2, \ldots, m)$$

6 Evaluation index

$$F_i = \sum_{j=1}^{m} W_j P_{ij}$$

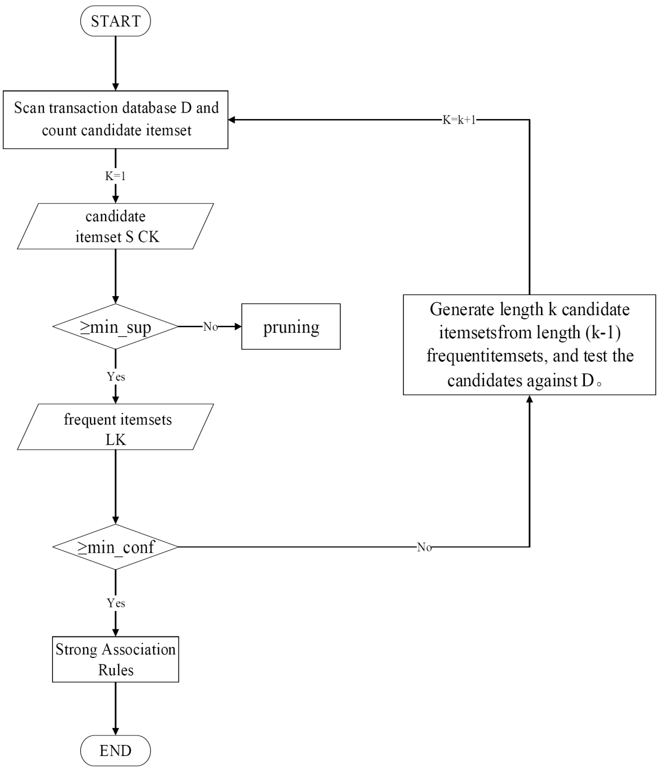
4 Data processing and results

Questionnaire 500 points were distributed in this experiment, 400 parts by effective recovery. In the survey of 400 questionnaires, we need to filter out invalid information first, such as lack of information. Because our purpose is to investigate what characteristics of the crowd are easily deceived, so we need to filter out the experience of not being cheated and then sort the data. A total of 350 valid questionnaires were left, 300 of which were used for analysis and the remaining 50 were validated for the evaluation of the index system.

4.1 Apriori algorithm

The process of algorithm is as shown in Figure 2.

Figure 2 The process of apriori algorithm



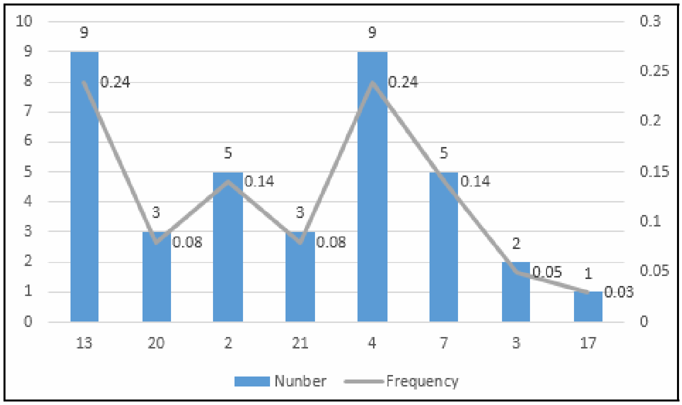
Finally, the minimum support we set is 15%, the minimum confidence is 85%. There are 14 as shown in Table 1.

Table 1 Association rules

Rule	*Support (%)*	*Confidence (%)*
13. 4. 20.	16.153846153846153	90.47619047619048
13. 4. 20. 2	15.384615384615385	90.0
13. 4. 2	20.0	88.46153846153845
13. 21. 4.	17.692307692307693	86.95652173913044
13. 21. 4. 2	16.923076923076923	86.36363636363636
13. 4.	22.30769230769231	86.20689655172413
7. 4.	22.30769230769231	86.20689655172413
7. 3. 2	16.153846153846153	85.71428571428571
13. 5. 4. 2	16.153846153846153	85.71428571428571
13. 8. 4.	15.384615384615385	85.0
13. 3. 21.	15.384615384615385	85.0
7. 3. 21.	15.384615384615385	85.0
7. 3. 17.	15.384615384615385	85.0
7. 6. 20.	15.384615384615385	85.0

The statistics of the association rules are as shown in Figure 3.

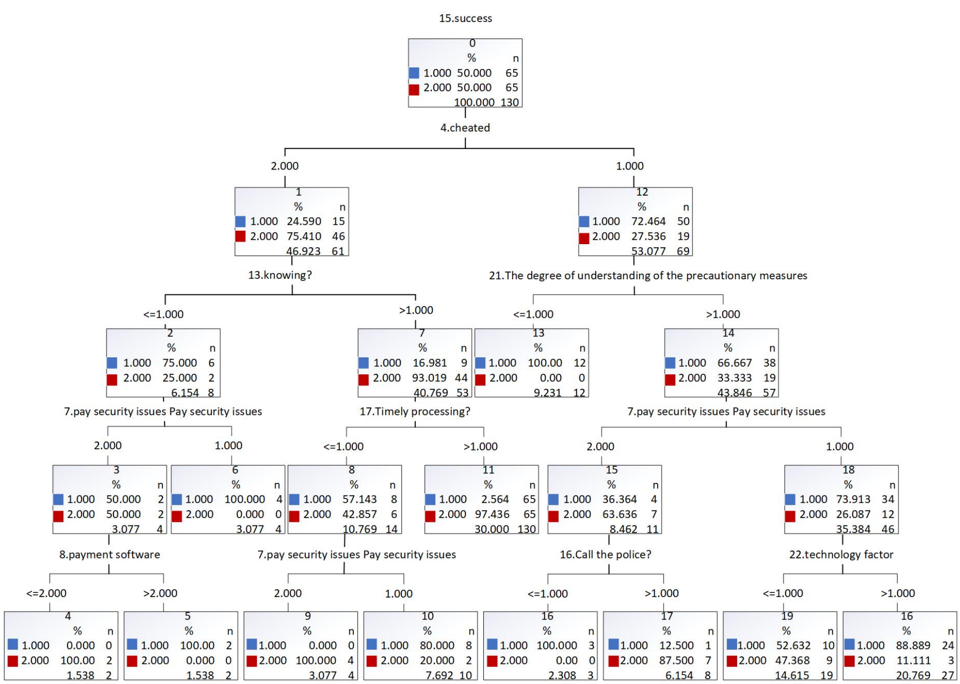
Figure 3 The statistics of the association rules (see online version for colours)



4.2 *ID3 algorithm*

Parts of the decision tree results from ID3 algorithm mining are as shown in Figure 4.

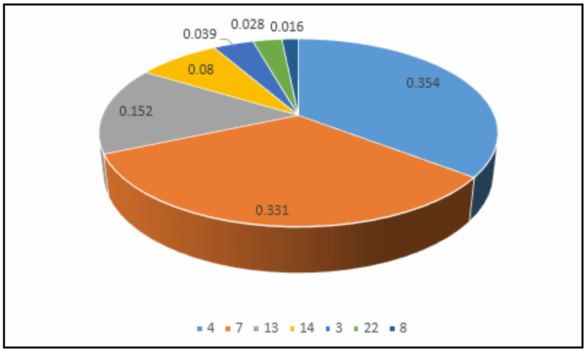
Figure 4 Part of the decision tree results (see online version for colours)



From the decision tree, we can see the final analysis results and make decisions by the characteristics which people show.

The importance of each node is as shown in Figure 5.

Figure 5 The importance of each node (see online version for colours)



4.3 Principal component analysis

The principal components analysis using the top several indicators to determine the matrix is as shown in Table 2.

Table 2 Principal component analysis

	Ingredient
Doubt?	.270
Age	–.624
Knowing?	.006
The degree of understanding of network transactions	.012
The degree of understanding of the precautionary measures	–.187
Weak legal consciousness	–.082
Education	.712
Technology factor	.423
Third party media	-.073

4.4 The evaluation of questionnaire

The reliability of the mining is analysed as shown in Table 3.

Table 3 Alpha coefficient

Rule	α
Age	0.832
Doubt?	0.814
Cheated?	0.815
Third party media	0.804
The degree of understanding of the precautionary measures	0.805
Knowing?	0.802
Technology factor	0.834

Table 3 Alpha coefficient (continued)

Rule	α
The degree of understanding of network transactions	0.805
Weak legal consciousness	0.803
Education	0.801
Lack of legal construction	0.807
Pay security issues	0.806
Payment software	0.801

The α of the items which has a significant relationship with the questionnaire are greater than 0.8, indicating that reliability of the scale is well. The validity analyse of the questionnaire is as shown in Figure 6.

Figure 6 The validity

KMO Bartlett

Kaiser-Meyer-Olkin		.816
Bartlett		419.460
	df	10
	Sig.	.000

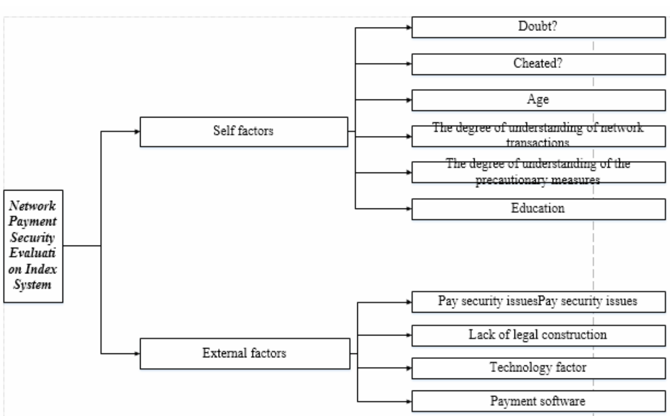
The KMO value is 0.816, so the questionnaire is good.

4.5 The construction of evaluation index system

First, we establish the indicators combined with the results of apriori algorithm and IDS algorithm, as well as the combination of relevant knowledge. Then according to the importance in the ID3 algorithm analysis and the frequency in the apriori algorithm association rules appear in, as well as expert scoring method, we determine the weight of each indicator.

The final evaluation index system is as shown in Figure 7.

Figure 7 Evaluation system



4.6 Weight

The evaluation index system is mainly composed of internal factors, that is, users and external factors. The internal factors are mainly reflected by the questionnaire, and the entropy method can be used to determine the weight of internal factors. The external factors are determined by expert scoring method Weights. The part of the code is as follows:

```
>> [n,m]=size(x);
[X,ps]=mapminmax(x′,0,1);
[X,ps]=mapminmax(x′);
ps.ymin=0.002;
ps.ymax=0.996;
ps.yrange=ps.ymax-ps.ymin;
X=mapminmax(x′,ps);
% mapminmax(‘reverse’,xx,ps);
X=X′;
for i=1:n
    for j=1:m
        p(i,j)=X(i,j)/sum(X(:,j));
    end
end
k=1/log(n);
for j=1:m
    e(j)=-k*sum(p(:,j).*log(p(:,j)));
end
d=ones(1,m)-e;
w=d./sum(d);
s=w*p’;
```

The result is as shown in Table 4.

Table 4 Internal factor weight

Age	0.0327
Education	0.1637
Cheated?	0.2573
Doubt?	0.3744
The degree of understanding of network transactions	0.0758
The degree of understanding of the precautionary measures	0.0961

For external factors, experts scored as shown in Table 5.

Table 5 External factor weight

Technology factor	0.15
Pay security issues pay security issues	0.1
Lack of legal construction	0.1
Payment software	0.05

Internal factors and external factors accounted for the ratio of 6:4. Finally, according to the importance in the ID3 algorithm analysis and the frequency in the apriori algorithm association rules appear in, as well as this ratio, we determine the weight of the index.

The weight formula is:

$$Weight = (The\ \ frequency\ in\ the\ apriori\ algorithm + The\ importance\ in\ the\ ID3\ algorithm + Entropy\ method\ weight / Expert\ score) / Overall\ score$$

Experts score the results as shown in Table 6.

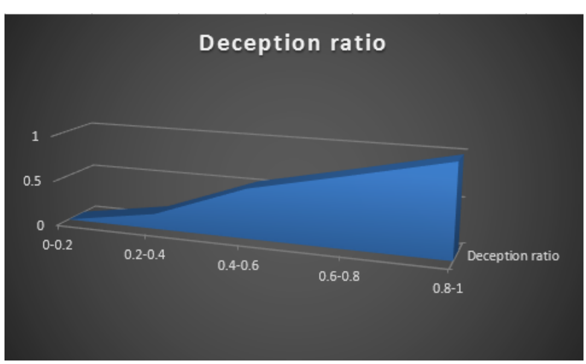
Table 6 Factor weight

Age	*0.05759*
Education	*0.067548*
Cheated?	*0.270011*
Doubt?	*0.109912*
The degree of understanding of network transactions	*0.045272*
The degree of understanding of the precautionary measures	*0.049667*
Technology factor	0.15
Pay security issuesPay security issues	0.1
Lack of legal construction	0.1
Payment software	0.05

4.7 The verification of evaluation index system

The evaluation index system was validated by 50 additional sets of data and the results are as shown in Figure 8.

Figure 8 Validated data (see online version for colours)



According to the evaluation index system, the higher the value, the higher the probability of being cheated. From the chart, we can see that the probability of being cheated is proportional to the evaluation index. The validation data conforms this rule, so the correctness of the evaluation index system was verified.

5 Conclusions

In this paper, we analysed the factors of cheated users from the customer point of view with the means of the questionnaire survey, apriori algorithm, ID3 algorithm, text mining. The factors we analysed that lead to the user cheating are significant, like Whether it has been cheated, age, education, weak legal consciousness, pay security issues, third party media, lack of legal construction, technology factor, The degree of understanding of the precautionary measures and so on. We also develop rating index system and the relevant weight combined with the results of algorithm mining and expert scoring to evaluate the ease of the cheated users. Moreover, the correctness of the evaluation index system was verified by data.

For users, we should strengthen awareness of prevention; enhance the degree of understanding of network affairs and legal awareness. For the payment platform, they should improve the level of security and anti-virus capabilities for users to build a more safe payment environment. The government should speed up the race of relevant legislation; increase the crackdown of the network fraud payment. Only when we are involved, can we build a safer network payment environment.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (Grant No. 71573011) and Foundation of China Railway Research Projects (No. N2018Z008).

References

Chi, G.T. and Li, Z.J. (2014) 'Information contribution model based on principal component – entropy evaluation index system', *Scientific Research Management*, Vol. 35, No. 12, pp.137–144.

Du, L.Y. (2014) 'Application of decision tree algorithm', *Light Industry Technology*, Vol. 30, No. 10, pp.67–68.

Huo, Z. (2016) 'Study on the evaluation system of urban low carbon communities in Guangdong Province', *Environmental Science and Management*.

Jin, C. (2014) *Research on the Criminal Law Affirmation of Cardless Credit Card Fraud*, East China University of Political Science and Law, Shanghai.

Li, H.F. (2014) *Research on Security Protocol in Mobile Payment*, MS thesis, Xidian University, Xian.

Li, M. and Tang, M. (2013) 'Information security engineering: a framework for research and practices', *International Journal of Computers Communications & Control*, Vol. 8, No. 4, pp.578–587.

Liu, J.W., Xu, W., Lv, B.X. et al. (2018) 'Design of mobile payment system based on NFC technology', *Telecommunications Science*, Vol. 34, No. 2, pp.131–138.

Miao, H.X. and Yu, J.K. (2014) 'Comparison of ID3 and C4.5 algorithms based on decision tree', *Modern Computer (Professional Edition)*, No. 15, pp.7–14.

Ruan, Y. (2014) 'Quantum principal component analysis algorithm', *Journal of Computer*, Vol. 37, No. 3, pp.666–676.

Tang, M., Li, M. and Zhang, T. (2016) 'The impacts of organizational culture on information security culture: a case study', *Information Technology and Management*, Vol. 7, No. 2, pp.179–186.

Wang, H. and Guo, C.Y. (2017) 'Effect analysis of linear dimensionless methods on the index weights by entropy method', *China Population, Resources and Environment*, Vol. 27, No. S2, pp.95–98.

Wang, X., Wang, Y.H., Su, L., Guo, B. and Wang, S.W. (2014) 'Index evaluation system on the degree of production-city integration in high-tech zones in China: based on factor analysis and entropy-based weight', *Science of Science and Management of S. & T.*, Vol. 35, No. 7, pp.79–88.

Wei, L., Wei, Y.J. and Gao, C.Y. (2015) 'Apriori algorithm based on Bigtable and MapReduce', *Computer Science*, Vol. 42, No. 10, pp.208–243.

Xie, Z., Gao, X., Feng, C. and He, J. (2017) 'Study on the evaluation system of urban low carbon communities in Guangdong Province', *Ecological Indicators*, Vol. 74, No. 3, pp.500–515.

Xu, D. (2018) 'Evolutionary game research on the three party payment from the perspective of e -commerce platform', *Business Research*, No. 2, pp.161–166.

Xu, K.S. (2014) 'Research on electronic payment security technology', *Electronic Test*, Vol. 36, No. S1, pp.91–93.

Xu, L.P., Jiang, X.R. and Yin, C. (2015) 'A research on the evaluation index system about innovation ability of enterprises', *Scientific Research Management*, pp.122–126.

Yang, J.H. (2015) *Research on the Causes and Countermeasures of Telecom Fraud Crime*, MS thesis, Chinese Academy of Social Sciences, Beijing.

Yang, Y.L. (2015) 'Research on the security application of mobile payment based on NFC technology', *Research and Development*, No. 20, pp.56–60.

Yuan, Q. (2014) 'Internet financial development facilitates payment security', *Network and Information Security*, pp.9–11, 2014.

Yuan, Q. (2014) 'Payment security driven by the development of internet finance', *Mobile Communication*, Vol. 38, No. 23, pp.9–11.

Zhong, Y.R. and Liu, D. (2016) 'An efficient association mining method via matrix compression', *Proceedings of 2016 9th International Symposium on Computational Intelligence and Design (ISCID)*, pp.188–192.