

# Design of Enterprise Crisis Predicting System Based on Cluster and Outlier Data Mining

Song Yan

School of Economics and Management Harbin Engineering University, China 150001  
songyan@hrbeu.edu.cn

## ABSTRACT

In order to solve such problems as half-structured and non-structured data analysis in enterprise crisis predicting system, a predicting system based on cluster and outlier data mining is put forward. The system organization, frame construction, function and working principles are illustrated. And the working process is showed by an example of cheat predicting. The experimental results show that this method is efficient and it is a new way to solve such problems.

## Keywords

Cluster, outlier data mining, crisis predicting system, enterprises.

## 1 INTRODUCTION

The main task of a crisis predicting system is to judge if there are some kinds of targets and factors over the warning limit and decide if give a warning, what level warning should be given and what way should be used to give this warning. The reasons of crisis are very complicated and so many, some factors are qualitative and some are quantitative. Because some of these reasons can be measured and the others can't, they are half-structured or non-structured problems with many uncertain factors. Many factors seldom can be calculated or predicted even with scientific methods because of lack of historical data and statistical data. There are many limitations in traditional analyzing methods and predicting results are not so ideal. So, the new methods to solve the above problems are urgent demands. Clustering principles and outlier data mining are introduced into enterprise crisis predicting system, an enterprise crisis predicting system based on cluster and outlier data mining is put forward in this paper to help enterprises to judge, evaluate and predict possible crisis. It is a new idea and method to solve crisis predicting for enterprises.

## 2 BASIC PRINCIPLE OF CLUSTER AND OUTLIER DATA MINING

Cluster is a kind of multi-factors analyzing method to classify. Actually, there are many fuzzy factors within classifying. So, using fuzzy cluster analyzing method to determine relation among objects is more reasonable.

There are often some objects which are different from regular data model or data. Such irregular data objects are named outliers. Outliers come into being perhaps because of measure errors, input errors or running errors. So, many data mining algorithms try to decrease outlier effects or kill them. But, this may cause loss of important information because one's noise may be one's signal. In other words, outliers may have special meaning sometimes, such as in cheat detecting, outliers will mean there are cheats. So, outliers detecting and analysis are very significant data mining task. In clustering algorithm, outliers are often abandoned as noise data. In my paper, noise data will used to realize crisis predicting.

## 3 DESIGN OF ENTERPRISE CRISIS PREDICTING SYSTEM BASED ON CLUSTER AND OUTLIER DATA MINING

The basic idea of this crisis predicting system is: first, cluster is used to make data elements into groups which have the same trend and model and group's feature is used to describe every group to decrease the number of information. Then, define outliers and discover outliers according to detection. Third, data matching scheme is used to deal with nonlinear problems in system. Such scheme must have abilities of fault-tolerance and noise-counteract. At the same time, feature-describing for clustering results make the results meaningful. So, the combination of cluster and outlier data mining have many characters and advantages and they can deal with and solve half-structured and non-structured problems such as analysis in crisis predicting.

### 3.1 Organization and function of crisis predicting system

This system is organized by system interface model, system maintain model, cluster model, outlier data mining model and warning model. Interface model is to communicate with users and collect relevant background knowledge about enterprise crisis and understand and describe users' need to provide a personal interface. Maintain model is to accomplish users' management, system organization and maintain of knowledge database. Knowledge database is mainly used to store kinds of reasons, procedures, characters, lessons, dealing methods of after-crisis to provide users inquiry and consultancy about enterprises and at the same time to provide related schemes and games of preventing and controlling crisis. Cluster model is

used to select suitable data according to users' demands and make data into groups of same trend and model. And then, selecting features to describe those groups is to reduce the amount of expressing information. Outlier data mining model can detect outliers according definition to mine. It can discover outlier-groups and collect data features and store into knowledge database. Warning model use data matching scheme to analyze data and give warning information for those matched as outliers and point out it maybe a crisis.

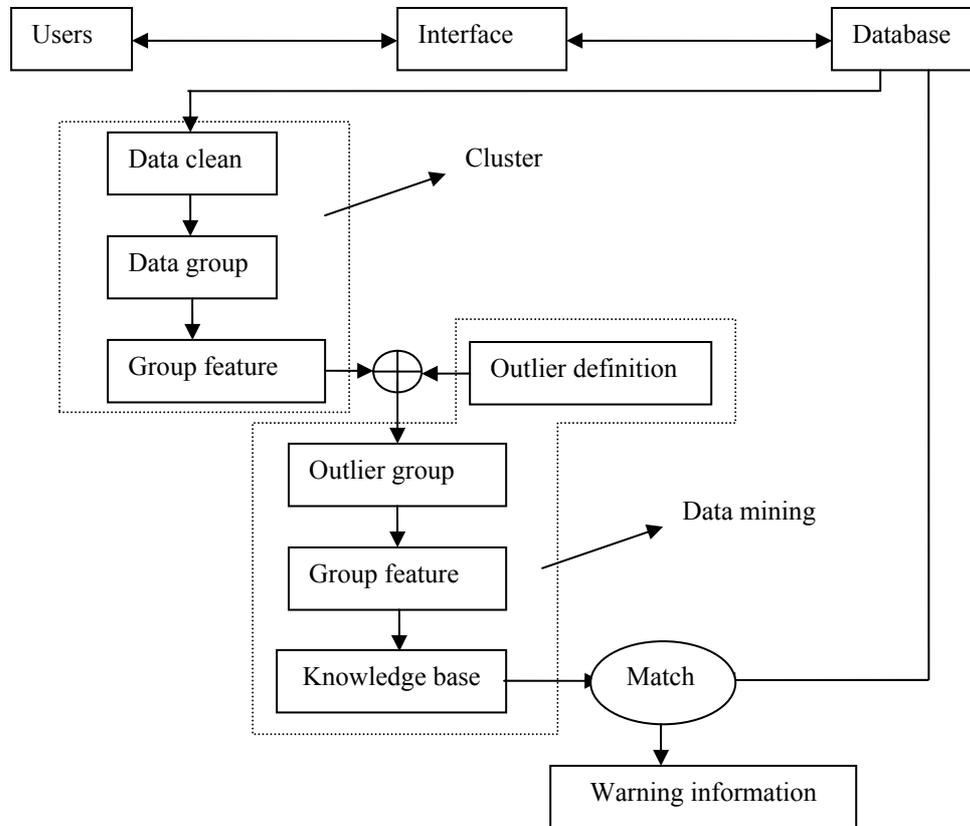


Figure1. Frame construction of crisis predicting system based on cluster and outlier data mining

3.2 Frame construction of crisis predicting system

The frame construction of crisis predicting system based on cluster and outlier data mining is as figure1.

In this system, set principle is used to describe objects and their properties are two-valued. Such expression can solve how to express half-structured and non-structured data. We can only record the data which property-value is 1 and only describing them is enough. So, the amount of expressing information can be reduced. According to the number of objects in a clustering set, outlier can be defined. As to matching scheme, numerical order is used to define boundaries of some classification and then judge if data is within this boundary fields, it can be put into this classification. Or, it won't.

4 AN EXAMPLE OF CHEAT PREDICTING

There are 6customers who purchased products in an enterprise. Ten properties are used to describe every customer and the first seven properties are used to describe the amount of buying products and the other four are describe customer's paying method, paying cycle, the amount in arrears and arrear-time. According to data transferring rules, two-value is to express objects properties and 1-value is recorded only. Calculate different degree and cluster to get results in table 1.

In this example, outlier is defined as a classification which has only two objects at most. So, classification 2 and classification 3 are outlier-sets. In table 2, set boundary is used to express clustering results. New data is added in this database as in table 3. And the matching results are in table 4.

So, the group of customer 11 and customer 13, the group of customer 12 and customer 15 are two outliers, while customer 14 is a new outlier because it can't be matched as any classification. Then, identify outlier-customer with warning and give out warning information and store it into knowledge database as a model. As information is increasing constantly, cluster and outlier data mining will go on and on. So, knowledge database can be refreshed continuously and their model-function will be more useful.

| Classification | Customer Number | Customer amount | Number of same properties | Number of different properties | Different degree |
|----------------|-----------------|-----------------|---------------------------|--------------------------------|------------------|
| X1(1)          | 1,3,6           | 3               | 1,3,4,5,6,8               | 7,10                           | 0.25             |
| X2(1)          | 2,4             | 2               | 1,9                       | 8,10                           | 0.5              |
| X3(1)          | 5               | 1               | 1,2,3,6,7                 |                                | 0                |

Table 1 Clustering Results

| Classification | Set of infimum | Set of superior  |
|----------------|----------------|------------------|
| B1             | 1,3,4,5,6,8    | 1,3,4,5,6,8,7,10 |
| B2             | 1,9            | 1,8,9,10         |
| B3             | 1,2,3,6,7      | 1,2,3,6,7        |

Table 2 Cluster Boundary

| Customer Number | Set of property-numbers which value is equal to 1 |
|-----------------|---|
| 10              | 1,3,4,5,6,7,8,10                                  |
| 11              | 1,9,10  |
| 12              | 1,2,3,6,7   |
| 13              | 1,8,9   |
| 14              | 1,2,3,4,5   |
| 15              | 1,2,3   |

Table 3 New Added Customers Data

| Classification Number | Customer Number |
|-----------------------|-----------------|
| 1                     | 10              |
| 2                     | 11,13           |
| 3                     | 12,15           |

Table 4 Matching Results

**5 CONCLUSION**

A method about a combination of cluster and outlier data mining is put forward to solve enterprise’s crisis predicting. Using two-value expression to describe data objects can reduce complexity and calculating time of system. Outlier data mining based on outliers can easily discover some concealed relation and procedure among data. Numerical order is used to match data and it can deal with nonlinear problems simply and efficiently. And it has strong abilities of fault-tolerance and noise-counteract. Feature-expression about clustering results makes mining results meaningful. So, combination of cluster and outlier data mining has many characters and advantages and it is very suitable for half-structured and non-structured problems in crisis predicting system. The example of cheat detecting can explain this method is practical and efficient. The experimental results show that the above method is a new idea and method for enterprises to predict crisis.

**REFERENCES**

1. He Zhengchu.(2003) Discussion about building enterprises crisis management system, *System Engineering, China*
2. LiuBingxiang.(2003) A design of enterprise crisis predicting system based on rough neural network, *Information and Control, China*
3. Li Yongjiang.(2003) Cluster analysis about factors of crisis predicting system in real estate enterprises, *Economist, China*
4. Ren Hua.(2003) A crisis predicting system for enterprises, *Journal of Wuhan Science and engineering university, China*