更多数学建模资料获取

B 站数学建模试看课程：

https://www.bilibili.com/video/av20238704

清风数学建模视频课程的课表：

http://note.youdao.com/noteshare?id=2e1cc342928466bb632b33304a56e368

视频录制了 45 个小时，累计 800 多页 PPT，讲解的了 20 多个常用模型，内容非常详细，并配有作业讲解。

关注微信公众号：《数学建模学习交流》可获得更多数学建模技巧和比赛信息。

怎么在手机上购买资料：在公众号后台发送“买”这个字，就可以获得一个链接，点进去后就可以购买了（购买方式和淘宝一样）

第四届“认证杯”数学中国

数学建模国际赛

承　诺　书

我们仔细阅读了第四届'认证杯'数学中国数学建模国际赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：1032

我们选择的题目是：A

参赛队员　(签名) ：

队员 1：李烁

队员 2：刘秋敏

队员 3：高明成

参赛队教练员 (签名)：王新赠

第四届“认证杯”数学中国

数学建模国际赛

编 号 专 用 页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：1032

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

Network Behavior Evaluation Model of Terrorism

Summary

Abstract: Terrorism has been existence for a long time in the world, but terrorist incidents frequently happened, which endangered the world peace seriously. The text will evaluate users' psychology by their behavior on the Internet, and gives reasonable anti-terrorism advice to Obama. Generally speaking, the text conducts a profound research from the perspective of internet, **big data**, machine learning, **analog simulation** and **cluster analysis**.

Firstly, we classify the internet users according to the different Internet activities.The activities include mailing, browsing the web, playing games,uploading and watching video in YouTube, online Transaction(Bank) and fixing GPS position on the Internet. Besides, we determine corresponding **risk index** and evaluate their tendency of danger one by one. Thus, we use the index to analyze from the perspective of time, number and frequency, establish **Risk Evaluation Model** and calculate the risk index for different users.

Secondly, we use **Monte Carlo** to simulate the data which indicates users' Internet activities. By Risk Evaluation Model, We can calculate all users' the risk index which is a comprehensive result which reflects users' network activities. Every users' risk index can indicate the possible of whether they are dangerous. So we regard index and every value of index as variable, adopting modified **K-means and FCM clustering** algorithm to analyze. In order to lower the program complexity, we use **Binary Enumeration Subset Algorithm** in C++ and get reasonable results.

Finally, according to the Risk Evaluation Model and Clustering Analysis Model which has established before, we hope to take scientific measures to oppose terrorism. We think the relevant agencies should carry out the **abnormal activities detection** on the Internet, and we hope president Obama accept our suggestion to take measures of monitoring technology to solve the problem of terrorism in the era of big data.

Key words: Terrorism; Risk Index; FCM Cluster algorithm; K-means Cluster

Contents

1 Introduction....6
1.1 Problem Description....6
1.2 Terminology and Definitions....6
1.3 Our work....7
2 Problem Analysis....7
2.1 Analysis of Risk Evaluation....7
2.2 Analysis of Terrorism....7
2.2.1 Characteristics of Terrorism....7
2.2.2 Behavior of Terrorism....7
2.2.3 Terrorist Activities....8
2.3 Analysis of Behavioral Psychology....8
2.4 Analysis of Risk Index....8
2.4.1 The threat sources' motivations....8
2.4.2 The ability of the threatening behavior....9
2.4.3 The vulnerability....9
2.4.4 The index of global terrorism....9
2.5 Analysis of Push-and-Pull Factor....9
2.5.1 The origin of the ‘push-and-pull’ theory....9
2.5.2 The power of push and pull....9
2.6 Analysis of Using Internet....10
3 Models....10
3.1 Risk Evaluation Model....10
3.1.1 Terms, Definitions and Symbols....10
3.1.2 Assumptions....11
3.1.3 Risk Index....11
3.1.4 Risk Score....16
3.1.5 Solution of Risk Index....17
3.1.6 Model Strength and Weakness....18
3.2 K-means Clustering Analysis Model....18
3.2.1 Extra Terms,Definitions and Symbols....18
3.2.2 Monte Carlo Method....19
3.2.3 FCM Cluster Algorithm Analysis....20
3.2.4 Fuzzy Set Establishment....21
3.2.5 Fuzzy K-means Cluster....21
3.2.6 FCM Algorithm Application and Solution....23
3.2.7 Model Strength and Weakness....23
4 Conclusions....23
4.1 Proposal....23
4.2 Model Application....24
5 Future Work....25
5.1 Model Evaluation....25
5.1.1 Model Limitation....25

5.1.2 Sensitivity Analysis.....25
5.2 Model Improvement.....25
6 References.....26
Appendix.....27

1 Introduction

1.1 Problem Description

The tasks as follow:

Task 1: Build a mathematical model to obtain a risk index, so we can evaluate the situation of each monitored person use it.

Task 2: Experts use the expression big data to indicate huge amounts of information. We'll get a lot of monitoring data, Please develop a series of statistical techniques to categorize them in an effective, fast and automatic manner.

Task 3: If President Obama asked for your advice on fighting terrorism, what would you tell him? What should he do about ISIS?

1.2 Terminology and Definitions

- **Terrorist:**Terrorism is an act which may be by force or violence, or threat of force or violence[1], or by any other form of intimidation to civilian to reach political object.
- **Risk:**Risk is the possibility of loss in the special situation or at certain times.Risk consists of risk factors,risk incidents,risk loss and so on[2].In other words,risk is the distance between people's exceptions and results.
- **Statistical techniques:**The technique is concrete ways of statistical technique. For instance control chart, histogram, scatter diagram and so on.
- **ISIS:**Islamic State of Iraq and al Shams is an extremist terrorist organization claiming to found a country and to be active in Iraq and Syria.Currently,the organization's control area is a large area of northwest Iraq and northeast Syria.The picture is ISIS Organization's map,and their ambitions are found the Islam world.

Figure 1: The Islam World

1.3 Our work

This paper tries to combine the user's Internet behavior psychological analysis with terrorism.**In section 1**, we analyze the cause of terrorism. **In section 2**, we establish risk index to evaluate the Internet user's online behavior. **In section 3**, we use statistical techniques to design a classification method. **In section 4**, we make a proposal to President Obama. **In section 5,** we analyze all of our model and get our strength and weakness.

2 Problem Analysis

2.1 Analysis of Risk Evaluation

According to the conceptual model of risk, the preparatory work before evaluation is including the actions and ways of terrorism activity, plans of risk evaluation, procedures of risk evaluation, ways and tools of risk evaluation, facing threat and vulnerability. Based of making sure security measures, we should analyze the factors and abilities of threat, being used because of vulnerability, the value and impact of terrorist incidents.

We evaluate the results of the above five aspects respectively, grade them, calculate the evaluation result of five aspects, and obtain the risk level finally.

2.2 Analysis of Terrorism

2.2.1 Characteristics of Terrorism

Terrorism is based on the violence which is purposive, systemic and abnormal[1].The violence of terrorism takes civilian or symbol of social standard as target to make extreme fear.Terrorism’s final purpose is political religious or social.

2.2.2 Behavior of Terrorism

The act of terrorism includes violent activity in based of terrorism and assistant behaviors. The act is a behavioral system which can be divided into the two levels:

- The principle act is carrying out the direct violent attack action to elected targets in order to realize their ambition. The act can be explained to conform to objective elements of terrorism. In past, the act was assassination, killing in public and so on. After the invention of dynamite, various explosive ways are emerging. Terrorist activities in the modern life are exploring, assassination, hijack, assault, kidnap, poisoning and so on; exploring and kidnap among them account for mostly[3].
- Assistant behaviors is an act in order to carry out the principle act[2]. Assistant behaviors includes the next aspects. Terrorism raises money and make illegal money transactions to support the terrorism activities. They make and buy weapons to enhance horror effect. They make various false certificates for transnational terrorism to flee. They communicate with each other by all kinds of information media. Only in this way, can they transmit various terrorist crimes

orders and coordinate the terrorist activities. They collect information of target who they will attack to make terrorist attack succeed. They recruit new members and control their behaviors. They contact other extremist organizations and so on.

2.2.3 Terrorist Activities

- After the cold war, ideology makes conflict which breeds hotbed of terrorism worse.
- American becomes the only empire in the world, whose Unilateral Policy has brought about dissatisfaction of other countries in the world.
- The reason why Islamic doctrine and International Terrorism have more conflict is the different between Eastern and Western Culture Value.
- American pro-Israel policy stimulate anti American sentiment which find an extreme reflection in fundamentalism.
- The education of some Muslim countries is underdeveloped, and most of the education for young people are provided free of charge by religious schools. However religious schools are bases which drive young Muslims to pursuit of religious.

2.3 Analysis of Behavioral Psychology

Analysis of the causes of terrorism is from two levels of individual and group. At the individual level, psychological research has explored the factors such as personality, motivation and cognition of terrorists and so on. At the group level, deindividuation theory in the field of social psychology analysis efficacy of group of terrorist organization. Now analysis from personal and group level[4].

2.4 Analysis of Risk Index

The index has to show the prediction of risks, it has to contain the protected object, the threatens it faced with and the vulnerability of existence. After confirmed that they have took safety precautions, we can analyze the threat sources' motivation, its ability, the possibility of its vulnerability be used, the value of the terrorism event and how it influently.

2.4.1 The threat sources' motivations

Even traditional American terrorists' intension is to harm American's interests, their ability of using computers and the Internet is not well than the other opponents, so, what they can make is limited Internet threaten. As bomb 's influence still so influential when compared to characters'[5], the terrorists may concentrate on traditional attract method in the short run. We can predict that a bigger scale of the Internet threaten will appear in the future's powerful generation. Their aim is to intersperse the tension mood among the American common people. This time, their aims are attract 50 thousands or more people , attract American economic and weaken it ,transfer the direction of counterterrorism struggles.

2.4.2 The ability of the threatening behavior

Assessment the effectiveness of the threat concentrated to determine the potential successful attack on the system capacity and effectiveness of the hostile forces of mankind. The ability refers to the attackers' knowledge of hostile(eg they have knowledge and are trained). Effection is the possibility of a capable enemy who can attack[6](eg they have a means of attack). Assessing the threat is to assessment the possibility of how threat event happened. There are a lot of factors to consider towards the probability of occurrence of natural events to failure behavior of the individual or assess the accident event. Consider many factors and make limited calculation or measurement towards them[7], it is needed that the standard of the consistency measure of the report.

2.4.3 The vulnerability

Vulnerability refers to the condition decided by natural, social, economic and environmental and other factors or activities, which makes someone suffer from crisis violation more easily[8]. The more 'vulnerability' people is, the more easily people affected by crisis.

2.4.4 The index of global terrorism

According to the American economy and US Institute of Peace (IEP) published the "global terrorism index" (GTI) [9] which is pointed out that in the list of 150 countries which are influenced by terrorist activities, India's fourth place after Iraq, Pakistan and Afghanistan. Yemen, Somalia and Nigeria is after the Indian. Thailand, Russia, the Philippines and the above mentioned countries occupy "global terrorism" index ranked the top ten. China that is also one of the biggest victimized country of terrorism is the 23rd.

2.5 Analysis of Push-and-Pull Factor

2.5.1 The origin of the 'push-and-pull' theory

The birth time of the 'push-and-pull' theory can date back to 19th century. The first one who does research on the movement of population is E.Ravenstien, British. He published a paper named the Law of the Movement of Population, which majored in the factors of the movement of population. The most important macroscopic theory of demography is 'push-and-pull', which was first come up with by D.J.Bagne[7]. Bagne thinks that the purpose of the movement of population is to improve the living conditions. The factors which are good to improve living conditions of in-flow areas becomes the 'pull' while the bad conditions of out-flow areas become the 'push'[10]. Similarly, people who turn to terrorism may be related to this push and pull two aspects, so we can apply the 'push-and-pull' theory to the research of it.

2.5.2 The power of push and pull

'Push and pull' refers to both internal and external impetus which people turn to

terrorists. The push is considered by in internal perspective, when people has certain violent tendency in their mind, maybe there are some internal driving forces that makes it develop into terrorist. While the pull is considered by external perspective. The recruiters of terrorist can choose certain groups as target object. They prompt some people who do not develop into terrorists to be drawed in and then turn into terrorists.

2.6 Analysis of Using Internet

With the WWW in the fourth media sweeping the global, browsing internet information has become main component of the Internet user act. But, we only can know how they use the Internet by obtaining web data. Obviously, no data classification can clearly indicate what website they browse, what they chat, what message they email or what files they download. The data can only indicate how people use different types of Internet resources and the difference in tendency of using various resources at the same time.

3 Models

3.1 Risk Evaluation Model

In the era of internet, Data Mining technology is widely paid attention to gradually; at the same time the use of Internet spread all over our life. We obtain the monitoring data by Internet technology. Browsing is always used to explore new or unknown information space. In process of browsing, user can always express their psychological model accordingly. As result, we should build corresponding risk index in based of users' online activities, and we can monitor users abnormal behavior by their risky index so that we can judge dangerous tendency level of users.

3.1.1 Terms, Definitions and Symbols

- Definitions State

1) Risk:In the process of decision-making,risk is the possibility of decision schemes causing adverse outcomes or degree of loss because of the effect of various uncertainties.In this problem,risk indicates the possibility of Internet activities causing user to be dangerous person.
2) R-score:Risk score measures the degree of whether user can become dangerous person. The greater the index is, the more likely that the risk would be a risk.

- Symbols State

Here, list the general form of symbol. Individual symbols will be described when they are first quoted.

Symbols	State
P_{ij}^{k}	The proportion of User k, Internet activity i, index j-th different levels of risk index in total;
P_{i}^{k}	The proportion of User k, Internet activity i-th risk index in total;

S_{ij}^{k}	Risk index of User k, Internet activity i,index j-th different levels;
S_j	Total risk index on the j index;
S^k	Risk index of user K.

3.1.2 Assumptions

- Assuming the data of simulation can reflect the reality;
- Assuming we can monitor people's Internet usage situation;
- Assuming our classification include all possible way of Internet using.

3.1.3 Risk Index

We classify the **risk index** according to Internet activities and we will analyze the index from four aspects of User/Address/Type,Time,Number,Frequency.
Firstly,we classify the Internet activities which include email,browsing website,playing games,uploading or watching video,online transaction and smart mobile phone position.We should determine the index according to the activities one by one and mark the index.

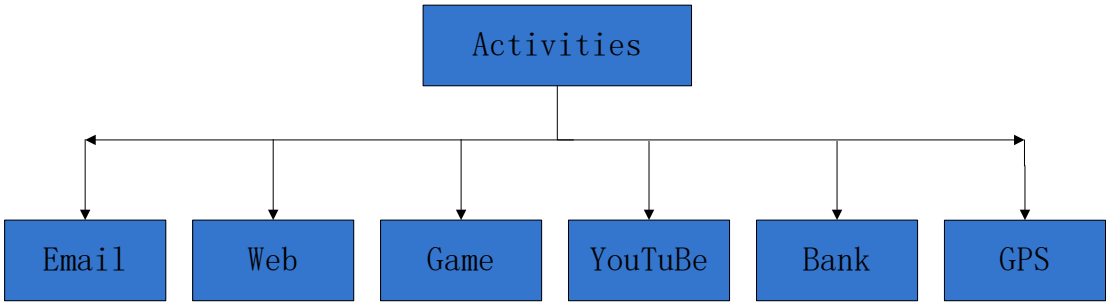

Figure 2:Network Activities Classification

3.1.3.1 Email

We make the activities of emailing specification in each index,which will be shown in the following table:

Table 1:Email

Contact user	P_{11}(%)	Time	P_{12}(%)	Number	P_{13}(%)	Frequency	P_{14}(%)
Frequent Contact	20	23:00-3:00	80	all emails	90	Frequently	60
		4:00-7:00	80	The number of 75%	80		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 25%	60		
Regular Contact	60	23:00-3:00	80	all emails	90	Frequently	60
		4:00-7:00	80	The number of 75%	80		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 25%	60		
Irregular	90	23:00-3:00	80	all emails	90	Frequently	60

Contact		4:00-7:00	80	The number of 75%	80		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 25%	60		

It should be noted that although our technology can't monitor the content of email,we can monitor the frequency and time of emailing between two IP and the email's number of the two same IP and so on.When it comes to emailing between two user,we classify them as frequent contact,regular contact and irregular contact.The users who contact frequently maybe due to business contacting or friendship,so the risk of them may be little. The users who contact regularly maybe due to business contacting or being threatened by terrorism,so the risk of them may be higher.The users who contact irregularly maybe due to being threatened by terrorism to carry out illegal transaction,so the risk of them may be the highest.Obviously,the classification is reasonable.

As a result, the possibility of users who will be dangerous is set 20%、60%、90% in three types email communication respectively.At the same time,the time of emailing also reflects that whether the user are dangerous person.Generally,emailing correspondence during the day mostly is normal communication.But emailing correspondence during the night mostly is abnormal communication which indicates the user tends to be dangerous. We divided a day into four time quantum which are 23:00-3:00, 4:00-7:00, 8:00-12:00 and 13:00-22:00. The possibility of users who will be dangerous is set 80%, 80%, 20%, 20% in every stage respectively.

We can monitor the email's number of the two users.If two users communicate with each other frequently,and we can monitor that one of the user's mailbox is almost all emails which are between them. So the risk index of them is high.Therefore, we can divide the number of email into all email, the number of 25%, the number of 50% and the number of 75%, corresponding dangerous possibility respectively is 90%, 80%, 70% and 60%. The same,we can also obtain the frequency of communicating between two users and we divide them into frequent communicating and sudden communicating and their dangerous possibility is 60% and 90%.

3.1.3.2 Browsing Website

We make the activities of browsing website specification in each index,which will be shown in the following table:

Table 2:Browsing Website

Types of websites	P_{21}(%)	Time	P_{22}(%)	Number	P_{23}(%)	Frequency	P_{24}(%)
Ammunition and Guns Website	95	23:00-3:00	80	All Websites	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Political News Website	40	23:00-3:00	80	All Websites	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		

Others	20	23:00-3:00	80	All Websites	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		

Similarly,We can't monitor that users browse the site's specific content,but we are able to monitor that different kinds of websites accounted for the quality,frequency,time and number of the user browsing the website.For users who browse the site,we will classify websites by purpose ,such as sites sell firearms and ammunition, news sites, other sites.

Users will be more likely to have risk who often browse the guns and ammunition website and the political news website. However, users who browse other websites will have less possibility to be dangerous. So the possibility of browsing different type websites respectively is set at 95%, 40%, 20%. Similarly with e-mail in time, browsing the site also reflects the possibility of user becoming dangerous people, so the four time periods are given the same probability v alue as e-mail.

Similarly,we can monitor the number of websites which users browse in common with the number of mailing probability value in the internet.Frequency of browsing the websites is also divided into frequent and sudden and it is in common with the frequency of mailing's probability value.

3.1.3.3 Playing Games

We make the activities of playing games specification in each index,which will be shown in the following table:

Table 3:Playing Games

Types of Games	P_{31} (%)	Time	P_{32} (%)	Number	P_{33} (%)	Frequency	P_{34} (%)
Adventure	80	23:00-3:00	80	All Games	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Role Playing	70	23:00-3:00	80	All Games	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Strategic Simulation	80	23:00-3:00	80	All Games	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Others	10	23:00-3:00	80	All Games	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		

We are able to monitor the types of games that users played, their playing time and

the proportion of different kinds of games in total games users played.Therefore, aiming at different players, we divide the games into action adventure, role-playing, strategy simulation, and others.

Players will be more likely to have risk who often play action adventures and strategy simulation games corresponding probability value of 80%. For players who often play role-playing games have higher possibility to be dangerous and the corresponding probability value is 70%.Players who play other games have less probability of risk.The time of playing games is similar to mailing time,and also reflects the possibility whether the user is dangerous and give them same probability value.

Similarly, we can also monitor the number of games which users play account for all games and they are in common with the number of mailing. So we give the same probability value as mailing. Frequency of playing games is also given the same probability value as mailing.

3.1.3.4 YouTube

We make the activities of uploading and watching videos in YouTube specification in each index,which will be shown in the following table:

Table 4:Video

Types of video	P_{41} (%)	Time	P_{42} (%)	Number	P_{43} (%)	Frequency	P_{44} (%)
Bloody	90	23:00-3:00	80	All videos	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Political News	80	23:00-3:00	80	All videos	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Others	10	23:00-3:00	80	All videos	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		

The same as above,we can also monitor the type of friends users have and the time of people upload and watch videos.We divided the videos into bloody,political information and others.User is more likely a dangerous person who always watch violent videos and political information videos,and corresponding probability values are 90% and 80%.The classification time of uploading and watching videos is similar with mailing.The time also reflects whether users are dangerous,so we give the same probability values to the four time periods as mailing.

Similarly,we can also monitor the number of videos which users upload and watch account for all videos and they are in common with the number of mailing in the internet.So we give the same probability value as mailing in different numbers.Frequency of playing games is also the same probability value as mailing.

3.1.3.5 Online Transaction

We make the activities of online transaction specification in each index,which will be shown in the following table:

Table 5:Online Transaction

Types of Transaction	P_{51} (%)	Time	P_{52} (%)	Number	P_{53} (%)	Frequency	P_{54} (%)
Transfer Accounts	80	23:00-3:00	80	All Transaction	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Consumption	20	23:00-3:00	80	All Transaction	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Others	10	23:00-3:00	80	All Transaction	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		

We can also monitor the frequency of transaction,the time of transaction and the type of transaction. So, We divided the transaction into transferring account,consuming and others. The transferring of money in terrorist is more than other activities. Therefore, User is more likely a dangerous person who has a sudden transaction. The possibility of users who consume by online transaction is less. The possibility of users who carry out the other activities is the least.

Similarly with the time of mailing, online transaction reflects the possibility of whether user is a dangerous person. Therefore, different time period of online transaction has the same probability value. And the other index is the same with mailing and given the same value.

3.1.3.6 GPS

We make the activities of smart phone location specification in each index,which will be shown in the following table:

Table 6:Smart Phone Location

Address	P_{61} (%)	Time	P_{62} (%)	Number	P_{63} (%)	Frequency	P_{64} (%)
Frequent Travel	20	23:00-3:00	80	All Travels	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Regular Travel	60	23:00-3:00	80	All Travels	90	Frequently	60
		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		
Irregular	90	23:00-3:00	80	All Travels	90	Frequently	60

Travel		4:00-7:00	80	The number of 25%	60		
		8:00-12:00	20	The number of 50%	70	Suddenly	90
		13:00-22:00	20	The number of 75%	80		

As positioning of smart phone is more and more popular,Internet can monitor the change of users' location and the time of users' location changing.Therefor,we can divide the data which GPS obtain about users' location into frequent travel place,regular travel place and irregular travel place.The risk index is increased in turn.The time of travel is similar with mailing time,and reflects the possibility whether user is dangerous.So we give the same probable value as mailing.

Similarly,we can also monitor the proportion of different types travels in total travels,and they will be given the same probability value.The frequency of location's change is the same as mailing and given the same probability value.

In **summary**,we can use the fishbone diagram to indicate risk index,as following:

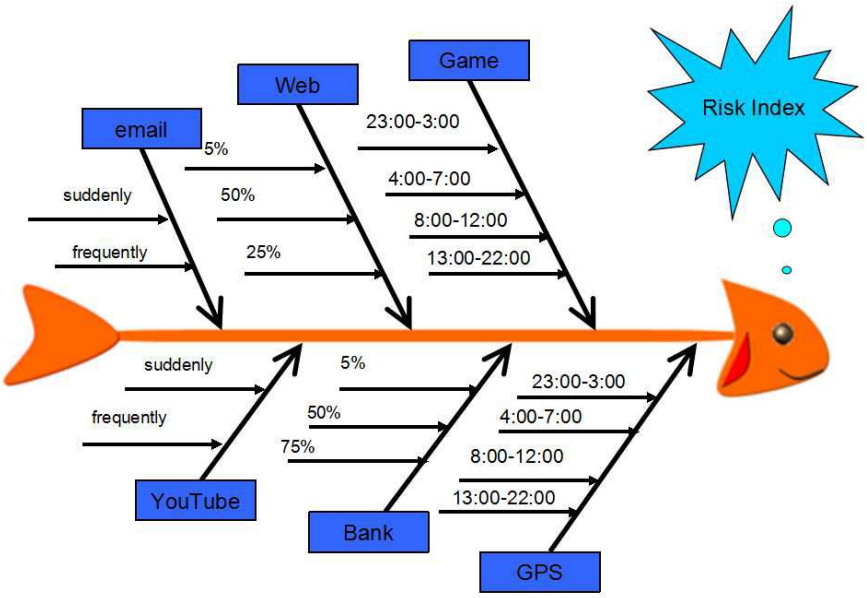

Figure 3:Fish Bone Diagram Analysis

3.1.4 Risk Score

We use points system to determine the value of risk index according to Monte Carol Theory.The higher the score is,the bigger risk index is.Various index S_j is following:

Table 7:Value of Various Risk Index

R-Index	User/Address/Type	Time	Number	Frequency
R-Score(S_j)	20	30	20	30

Generally,one user maybe do many Internet activities.We weigh the user's risk index

in every Internet activity to obtain users' final risk index.The way of determining the weight is the same as the above ways of determining index by point system.The level of score represent the degree of activities' impact on risk index.Proportion of various activities' index in total risk index P_i^k is following:

Table 8:Risk Index of VariousActivities

Internet activities	Email	Web	Game	YouTube	Cyber Bank	GPS
P_i^k	20	30	5	10	10	25

According to the above index,we can calculate every user's risk index by monitoring.Different users carry on different types of activities, and the risk index of them is different,computational formula is following:

$$S_{ij}^{\ k} = S_j \times P_{ij}^{\ k}. \quad (1)$$

Among Formula (1), $S_{ij}^{\ k}$ is on behalf of the user k risk index in the i-th and j-th of network activity; S_j is on behalf of in the j-th in the total risk index; $P_{ij}^{\ k}$ is on behalf of the proportion of network activity of total risk index by user k in the i-th and j-th type; Then,we can get the risk index by the formula as follow:

$$S^k = \sum_i \sum_j (S_{ij}^{\ k} \times P_i^k). \quad (2)$$

Among Formula (2), P_i^k is on behalf of the proportion of network activity of total risk index by user k in the i-th type; $S_{ij}^{\ k}$ is on behalf of the user k risk index in the i-th and j-th of network activity; S^k is on behalf of the user k risk index.So we can get the risk index, and the larger the risk index, the larger the degree of terrorism.

3.1.5 Solution of Risk Index

We use C++ to write program(Reference Appendix 1), and get the result as follow(other result reference Appendix 3):

Table 9:Data Form

1	Email	Irregular	8:00-12:00	100%	Suddenly	69
1	Web	News	8:00-12:00	50%	frequently	46
1	Game	challenge	4:00-7:00	25%	Suddenly	79
1	YouTube	Violence	4:00-7:00	75%	Suddenly	85
1	Bank	others	13:00-22:00	100%	Suddenly	53
1	Gps	often	8:00-12:00	25%	frequently	52
2	YouTube	Violence	23:00-3:00	50%	Suddenly	83
2	Bank	others	8:00-12:00	75%	frequently	42
2	Gps	often	23:00-3:00	75%	Suddenly	83

3.1.6 Model Strength and Weakness

3.1.6.1 Strength

- According to the different network activity, the model that establishes indicators of risk from details, web activity is divided into six categories, and is respectively from the User/location/Type, Time, Number, Frequency was analyzed, and comprehensive, objective and clear;
- The model by using the Monte Carlo Method has carried on the simulation of risk indicators, and its rationality was verified;
- The model calculates the risk index at the different user, different network activity, reasonable and clear.

3.1.6.2 Weakness

- While considering the details of the model is more, the model also has significant limitations;
- The weights of the model has a certain subjectivity.

3.2 K-means Clustering Analysis Model

According to risk indicators built in the Risk Evaluation Model as above,we can calculate every user's Risk Index.And this risk index is the results that synthesize every user's all network activity corresponding all indexes[11]. In the same time,every index can indicate risk index that the user become subversive or terrorism.So we use the improved K-means to classify for each index and risk index as a variable.Our algorithm uses big data as analysis object,which mainly embody the strength of machine learning.

3.2.1 Extra Terms,Definitions and Symbols

- Definitions State

1) Monte Carlo method:The theoretical basis is law of large numbers.This method is a kind of statistical simulation method, which is based on probability and statistics theory as the guidance of a class of important numerical method, this method is to use a random number (or pseudo random number) to solve the problem of a lot of calculation method.
2) Law of large numbers: the laws of describing the results of quite a few times repeated trials.According to this laws,the more samples we know, the more real value the average tend to be.

- Symbols State

Special individual symbols used first are explained here.

Symbols	States
$\mu_A(x)$	The function based on the degree of the object x attach to set A.
J	Objective function or value function
$\underset{\sim}{A}$	Fuzzy subset

3.2.2 Monte Carlo Method

Due to the randomness of generating random Numbers, when we use N random points by Monte Carlo Method to simulate the Internet user's real situation, the more the value of N access,the more realistic Internet users situation.We randomly get 5000 users on the Internet by using the Monte Carlo Method to simulate Internet monitoring data.

❖ Binary Enumeration Subset Based on the Monte Carlo Algorithm

Firstly,we assume that there is 6 kinds of activities, which is respectively denoted by Email,Web,Game,YouTube,Bank,Gps (labels of 0,1,2,3,4,5).

Table 10:Algorithm of BES

Input	**(Id,Score)**
Step1	As for every act,we set five attributes that stored in struct Pro,which contains name,user,data,number,rate,score.Define q [I] container which is on behalf of the Numbers for the i-th person activities.
Step2	In the form of binary enumeration subset, we must get the enumeration of all these activities.And we can ensure that each combination contains at least three kinds of activities, at most 6 kinds of activities. The binary number must be put in bin container.
Step3	We get a binary number from Bin Container.Firstly,we query the acts base on the binary number.Assume the act is email,we call do_email(int id) function.Then randomly output five attributes of corresponding value of email and calculate the risk in turn.Finally, put Pro into q [id].
Step4	According to the acts of 5000 people,we can get the total risk factor and output to file.
Output	**Table n**

According to the description of algorithm,we can get the flow chart of algorithm as follow:

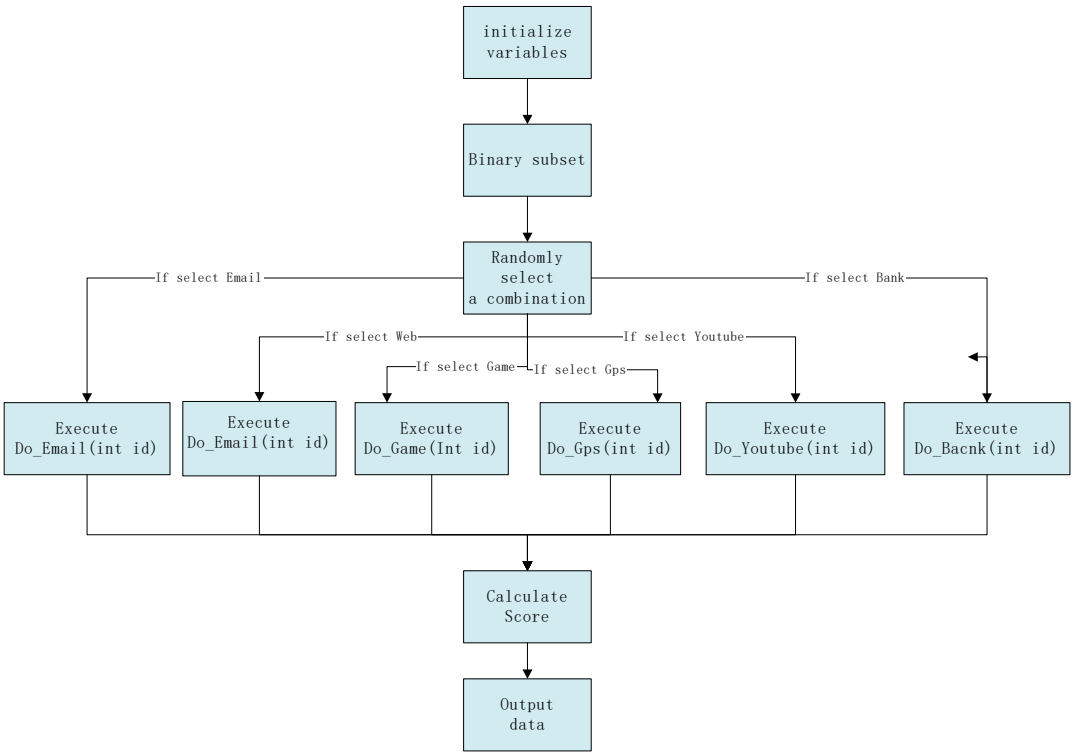

Figure 4:Flow Chart

We use C++ language to program(Reference Appendix 2),then we get every user's data.The form of all data as follow:

Table 11:Data Form

id:67					
	Web	Entertainment site	8:00-12:00	Web 25%	frequently
	YouTube	Friend for a long time	8:00-12:00	Video 50%	frequently
	Cyber Bank	Current account on a regular basis	13:00-22:00	Deal 25%	frequently
Score:33					

The table shows the user whose id is 67 has a variety of network activity,and all activities are to be carried out during the day, be regular network activity, and the number of network activity is small.

As as result,the risk index we have got is low,which declares the user's risk tends to be smaller.And all data reference appendix 3.

3.2.3 FCM Cluster Algorithm Analysis

According to the analysis of data as above,we know that the more similar between users of network activity, the more similar between the size of the dangerous tendency that is corresponding to the activity.So we use FCM algorithm.

The object of the algorithm is that make the same cluster similarity between objects reach maximum,but the similarity between different varieties will reach minimum.So according to the user's activity,the risk index are classified.

FCM cluster algorithm analysis is a improved K-means algorithm.In ordinary K-means algorithm,data division is hard,but FCM algorithm is flexible fuzzy division[10].

3.2.4 Fuzzy Set Establishment

The scope of the variables of $\mu_A(x)$ are all belonging to object A which set A places all points in space. Apparently, the data range is between 0 and 1 that $0 \le \mu_A(x) \le 1$. $\mu_A(x) = 1$ is on behalf of x completely belonging to set A, and x∈A.Fuzzy subset $\underset{\sim}{A}$ is a definition in the space X = {X} on the membership functions.Fuzzy subset $\underset{\sim}{A}$ can be expressed as follow:

$$\underset{\sim}{A} = \{(\mu_A(x_i), x_i) | x_i \in X\}. \tag{3}$$

In the problem of clustering, we can regard clustering as a fuzzy set that generated by the cluster.A element that is subordinated to the fuzzy set is not hard.As a result, each sample point belongs to the cluster membership whose value is between 0 and 1.

3.2.5 Fuzzy K-means Cluster

The FCM algorithm should divide n vectors x_i (i=1,2,…,n) into c fuzzy groups firstly.Then we must calculate the central cluster of every group in order to make the similarity index value function reach minimum.FCM algorithm use obscure division to determine the degree of each group by membership which every data point between 0 and 1[11].With the introduction of fuzzy partition,the data in member matrix U can allow it between 0 and one.But the membership in a data set must add up to 1 under the Normalization rules:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1,...,n. \tag{4}$$

Then, the form of the value function or object function in FCM algorithm:

$$J(U, c_1, ..., c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^m d_{ij}^2 \tag{5}$$

Among Formula (4), $u_{ij} \in [0,1]$, c_i is on behalf of the central cluster in fuzzy group I.

$$d_{ij} = \| c_i - x_j \|. \tag{6}$$

Among Formula (4), d_{ij} is on behalf of the Euclidean distance between the i-th central cluster and j-th central cluster data point, $m \in [1, \infty)$ is weighted index number.

We construct the new object function as follow,and get the requirement that makes Formula (5) reach the minimum:

$$\begin{aligned}\overline{J}(U,c_1,\ldots,c_c,\lambda_1,\ldots,\lambda_n) &= J(U,c_1,\ldots,c_c)+\sum_{j=1}^{n}\lambda_j\left(\sum_{i=1}^{c}u_{ij}-1\right),\\ &=\sum_{i=1}^{c}\sum_{j}^{n}u_{ij}^{m}d_{ij}^{2}+\sum_{j=1}^{n}\lambda_j\left(\sum_{i=1}^{c}u_{ij}-1\right).\end{aligned} \tag{7}$$

Among Formula (7), λ_j, $j=1\cdots n$ is on behalf of n Constraints of Lagrange multiplier in Formula (4).

All the input parameters can obtain derivation.Then we can get the requirement that reaches the minimum in Formula (5)

$$c_i=\frac{\sum_{j=1}^{n}u_{ij}^{m}x_j}{\sum_{j=1}^{n}u_{ij}^{m}}. \tag{9}$$

Among Formula (9), $u_{ij}\in[0,1]$, c_i is on behalf of the central cluster of fuzzy group I.

We get the computing method of membership function as follow:

$$u_{ij}=\frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}. \tag{10}$$

Among Formula (10), d_{ij} is on behalf of the Euclidean distance between the i-th central cluster and j-th central cluster data point.

With the above two necessary conditions,Fuzzy C-Means Clustering Algorithm is a simple iterative process.Then we get FCM algorithm to make sure the center of clustering c_i and Membership matrix U as follow steps:

Table 12:FCM Algorithm

Step1	Use the random number between 0 and one to initialize subject matrix U,and make it meet the constraints in Formula (4);
Step2	Calculate the central cluster c_i ,i=1-c by Formula(8);
Step3	Calculate the value function by Formula (5).If it's less than a certain threshold, or the variation is less than the certain threshold,the algorithm will break;
Step4	Calculate new Matrix U by Formula (9),then go back to **Step2**。

The algorithm as above can also initialize the central of cluster firstly.Then it will run the iterative process.Because we can't make sure that FCM algorithm is convergence into the best optimal solution,the algorithm performance must rely on initialization clustering center.As a result,we need to use different initialization clustering centers every time to start the algorithm.And after multi-iterations,all classification will reach

a stable value.

3.2.6 FCM Algorithm Application and Solution

By the means of the discussion as above,we can easily see that FCM algorithm needs two parameters.One is cluster number *C*,and another is parameter *m*.In general,*C* is far less than the total number of cluster sample.In the same time,we must make sure that the *C* >1.As for *m*,it's a flexible parameter in control algorithm.If *m* is too big,the effect of cluster will be bad.But to the contrary, if *m* is too small,the algorithm will be close to the traditional K-means cluster algorithm.So when we sort the simulation data,it's necessary to adjust the parameter *m* and the sorted number so as to make the classification more reasonable.

We write the code of the algorithm in the use of C++ language(Reference Appendix 2).And we get the results as follow(All simulation results reference Appendix 4):

Table 13:Data Form

738	Email	often	13:00-22:00	25%	frequently
738	Web	Others	13:00-22:00	50%	frequently
738	Game	others	8:00-12:00	50%	Suddenly
738	Bank	consume	8:00-12:00	100%	frequently
738	Gps	Irregular	13:00-22:00	50%	frequently
40					

3.2.7 Model Strength and Weakness

3.2.7.1 Strength

- The model can apply C++ to program,and use the Binary Enumeration Subset Algorithm whose time complexity is lower than others;
- The model use membership function which needs a small amount of data to get the final results.So it's effective and simple.
- FCM cluster algorithm is a kind of process that is unsupervised and self-adaptive learning,which can be easily applied.Also,the speed of the algorithm is rapid.
- In the objective function, Euclidean Distance is available. As a conclusion,the model has more intuitive geometric meaning.

3.2.7.2 Weakness

- K-means cluster algorithm must rely on the original central values,so the model has a certain limitation;
- FCM cluster algorithm only uses central class to express the class.In this way, the model may be sensitive to the noise data in many cases.

4 Conclusions

4.1 Proposal

If President Obama asked for our advice on fighting terrorism, we would tell him the proposal against ISIS as follow.

Dear Honorable President Obama:

Firstly, we are grateful to you for reading our recommendations even if you are in your busy schedule.

The principle aim we write the recommendations is anti-terrorism. Recently a terrorism incident happened in Paris. Beautiful Paris was the place numerous people yearned for in the past, but severe terrorism incident make it become gloomy. Now, peace and development are the themes of world.You, as an honor president of the most developed country and Nobelist, have a long way to go to maintain world peace and development.

Since terror attacks 9/11 incident,terrorism still spread all over the world and carry out terrorist incidents abruptly,which makes citizens' personal security in danger.So our purpose of research is to know the factors of terrorism and take measures to oppose them.

With the coming of Internet era, it narrows the distance between people.But many terrorism are more likely to obtain users' data of web surfing buy some illegal activities.By this way,they give the strength of 'pulling and pushing' to recruit new members.We can't let it happen,so we can take technique of anomaly Detection to grab the information of Internet users.Though we can't obtain the specific Internet data,we can obtain their psychological feature and dangerous tendency by analyzing type of their activities,time of surfing,frequency of surfing,number of surfing and so on.

Our specific recommendation is following:

1.Build effective detection and supervision system on Internet.Internet is a public platform,we should detect users' Internet action in real time to judge whether the action is abnormal according to risk index.

2.Build humanized heart guidance system.We should transform the strength of 'pulling and pushing' into active heart guidance.And we can apply behavioristic psychology to all aspects of our life.

3.Improve anti-terrorism agency.We should strengthen to monitor terrorism's Internet activities and crack down every terrorism severely.We must make corresponding policy to the abnormal data they obtain from Internet.

Thank you again.

Wish you success in work,happiness in family!

A college student
November 30th, 2015

4.2 Model Application

The model must input the data in the base that we can monitor some people's Internet use.Therefore,there is a certain error in the result from our program.The results will tell us the user whether is a subversive(terrorism).As a result,our model will play a positive role in the case of terrorism against.

5 Future Work

5.1 Model Evaluation

5.1.1 Model Limitation

- K-means cluster algorithm:The dependency of the original central values in this algorithm is very large.We should update central values repeatedly,and examine their validities.Thus,the results will be more reliable and real.
- FCM algorithm:First, FCM cluster algorithm only apply central class to express class. As a consequence,this algorithm is sensitive to noise data in many case. As well as, Bezdek has proved that FCM algorithm only makes sure that it can be convergent to the fixed point[10]. So, we can't insure that the result is convergent to minimal values of objective function. Second, this algorithm has two parameters. And we must give appropriate values to them in advance before cluster analysis(fuzzy weighted index and cluster number). So it will affect FCM algorithm analysis effects. Last, this algorithm is sensitive to the initialized sorted parameters, so different initialized values may get different partitioning results. In the same time, it will be more sensitive to noise data. Essentially speaking, FCM algorithm is a local searching hill climbing methodsin practice. So it will easily get error local extreme point, but it can't get the minimum values of objective function.

5.1.2 Sensitivity Analysis

In consideration of K-means cluster algorithm that must rely on the original central values,we should adjust the values repeatedly,and compare the new results with before.Therefore we can get better central value.

5.2 Model Improvement

So as to weaken the influence of human factor to the result,we use Hadoop model and improve Hadoop model.It makes MapReduce computing model combine open source distributed and parallel framework.Programmer can write a program in virtue of Hadoop model,and it can be running in computer cluster.Thus,we can make it easy to handle mass data[12].

In addition, Hadoop also provides a HDFS and HBase in order to attach data to every computing node. So,you can think generally that Hadoop = HDFS(file system related to data storage) + HBase(database) + MapReduce(data processing). The Structure of Hadoop Framwork as follow:

Table 14:Hadoop Framework

Cloud Computing and Hadoop Framework	
HDFS	MapReduce API(Map,Reduce)

We can calculate and store large data with the help of Hadoop framework and cloud computing core technology(MapReduce). Also HDFS and HBase can be blended in cloud computing framework.Consequently,distributed cloud computing,parallel

computing and storing,managing large data ability can come true commendably.

6 References

[1] Baidu. Terrorism[EB/OL]. http://baike.baidu.com/link?url=9eFwQVOOjxunnHDEoCWS1GRJUJsUiLm1kUaeHXXkD3orf8C43deOs3oRMsHjvs-nUB3xSHbQ4gIOa6fHUQO-JK. 2015-11-30.

[2] HE Bing-song, LIAO Bin. Comparative Study on Terrorism Concept[J]. Study of Comparative Method,2003,04:44-54.

[3] Baidu. ISIS[EB/OL]. http://baike.baidu.com/link?LG1f5v_HTzmy3JIEdwV8GUvtEEIxXFlLbdZWNeRc00PluYtCKztpQdSsXwJGi. 2015-11-30.

[4] JIA Feng-xiang, SHI Wei. Psychological Review Based on the Terrorists Terrorism [J]. Advances in Psychological Science, 2010, 18(10): 1660-1667.

[5] LIU Heng-chao. Psychological Analysis of the roots of Terrorism[J]. Dongyue Tribune, 2009, 30(3): 64-67.

[6] YANG Guo-liang, REN Jin-xia, WANG Zhi-liang. Affective Modeling Based on Emotional Psychology[J]. Computer Engineering, 2007, 33(22): 209-211.

[7] Personal Blog of HUA Rong. Summary of Network Threat Source[EB/OL]. http://blog.cechina.cn/huarong/442792/message.aspx. 2015-11-30.

[8] CHEN Bin,CHEN Song-can,PAN Zhi-song,LI Bin. Summary of Anomaly Detection [J].Shandong University Journal(liberal arts ed.),2009,06:13-23.

[9] PHBang. The Rink List of the Countries Around the World by the Terrorist Threat[EB/OL]. http://www.phbang.cn/plus/view.php?aid=941. 2015-11-30.

[10] LI Ling-ling, XIN Hao. FCM Algorigthm and Its Effective Measures[J]. Anhui technical college of electronic information Journal, 2011, 10(5): 38-40.

[11] WANG Shi-qing, QIU Lin, WANG Zhi-liang, HAN Xiao-jun. Statistical Analysis Method of Determining the Membership Function[J]. North China University of Water Conservancy and Electric Power Journal, 2002, 23(1): 68-71.

[12] ZHANG De-ran. Test Method of Abnormal Value in Statistical Data[J]. Statistical Research, 2003, (5): 53-55.

Appendix

1、Code of Problem one

Running Environment is Code::Blocks 13.12 under Operating System win10

```
#include <cstdio>
#include <algorithm>
#include <string>
#include <map>
#include <vector>
using namespace std;
struct Pro {
	string name;
	string user;
	string data;
	string number;
	string rate;
	int score;
	Pro(string a = "", string b = "", string c = "", string d = "", string e = "", int _score = 0) {
		name = a, user = b, data = c, number = d, rate = e, score = _score;
	}
};
vector <Pro> q[10000];
void do_email(int id) {
	string user[10] = {"often", "regular", "Irregular"};
	int user_num = rand() % 3;
	int score = 0;
	if (user_num == 0) score += 20 * 0.2;
	if (user_num == 1) score += 20 * 0.6;
	if (user_num == 2) score += 20 * 0.9;
	string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
	int data_num = rand() % 4;
	if (data_num == 0) score += 30 * 0.8;
	if (data_num == 1) score += 30 * 0.8;
	if (data_num == 2) score += 30 * 0.2;
	if (data_num == 3) score += 30 * 0.2;
	string number[10] = {"100%", "25%", "50%", "75%"};
	int number_num = rand() % 4;
	if (number_num == 0) score += 20 * 0.9;
	if (number_num == 1) score += 20 * 0.6;
	if (number_num == 2) score += 20 * 0.7;
	if (number_num == 3) score += 20 * 0.8;
	string Possible[10] = {"frequently", "Suddenly"};
	int poss_num = rand() % 2;
	if (poss_num == 0) score += 30* 0.6;
```

```
        if (poss_num == 1) score += 30* 0.9;
        q[id].push_back(Pro("Email", user[user_num], data[data_num],
                                    number[number_num], Possible[poss_num], score));
}
void do_Web(int id) {
        string user[10] = {"Guns", "News", "Others"};
        int user_num = rand() % 3;
        int score = 0;
        if (user_num == 0) score += 20 * 0.95;
        if (user_num == 1) score += 20 * 0.4;
        if (user_num == 2) score += 20 * 0.2;
        string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
        int data_num = rand() % 4;
        if (data_num == 0) score += 30 * 0.8;
        if (data_num == 1) score += 30 * 0.8;
        if (data_num == 2) score += 30 * 0.2;
        if (data_num == 3) score += 30 * 0.2;
        string number[10] = {"100%", "25%", "50%", "75%"};
        int number_num = rand() % 4;
        if (number_num == 0) score += 20 * 0.9;
        if (number_num == 1) score += 20 * 0.6;
        if (number_num == 2) score += 20 * 0.7;
        if (number_num == 3) score += 20 * 0.8;
        string Possible[10] = {"frequently", "Suddenly"};
        int poss_num = rand() % 2;
        if (poss_num == 0) score += 30* 0.6;
        if (poss_num == 1) score += 30* 0.9;
        q[id].push_back(Pro("Web", user[user_num], data[data_num],
                                    number[number_num], Possible[poss_num], score));
}
void do_Game(int id) {
        string user[10] = {"challenge", "character", "Strategy", "others"};
        int user_num = rand() % 4;
        int score = 0;
        if (user_num == 0) score += 20 * 0.8;
        if (user_num == 1) score += 20 * 0.7;
        if (user_num == 2) score += 20 * 0.8;
        if (user_num == 3) score += 20 * 0.1;
        string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
        int data_num = rand() % 4;
        if (data_num == 0) score += 30 * 0.8;
        if (data_num == 1) score += 30 * 0.8;
        if (data_num == 2) score += 30 * 0.2;
        if (data_num == 3) score += 30 * 0.2;
```

```
    string number[10] = {"100%", "25%", "50%", "75%"};
    int number_num = rand() % 4;
    if (number_num == 0) score += 20 * 0.9;
    if (number_num == 1) score += 20 * 0.6;
    if (number_num == 2) score += 20 * 0.7;
    if (number_num == 3) score += 20 * 0.8;
    string Possible[10] = {"frequently", "Suddenly"};
    int poss_num = rand() % 2;
    if (poss_num == 0) score += 30* 0.6;
    if (poss_num == 1) score += 30* 0.9;
    q[id].push_back(Pro("Game", user[user_num], data[data_num],
                        number[number_num], Possible[poss_num], score));
}
void do_YouTube(int id) {
    string user[10] = {"Violence", "information", "others"};
    int user_num = rand() % 3;
    int score = 0;
    if (user_num == 0) score += 20 * 0.9;
    if (user_num == 1) score += 20 * 0.8;
    if (user_num == 2) score += 20 * 0.1;
    string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
    int data_num = rand() % 4;
    if (data_num == 0) score += 30 * 0.8;
    if (data_num == 1) score += 30 * 0.8;
    if (data_num == 2) score += 30 * 0.2;
    if (data_num == 3) score += 30 * 0.2;
    string number[10] = {"100%", "25%", "50%", "75%"};
    int number_num = rand() % 4;
    if (number_num == 0) score += 20 * 0.9;
    if (number_num == 1) score += 20 * 0.6;
    if (number_num == 2) score += 20 * 0.7;
    if (number_num == 3) score += 20 * 0.8;
    string Possible[10] = {"frequently", "Suddenly"};
    int poss_num = rand() % 2;
    if (poss_num == 0) score += 30* 0.6;
    if (poss_num == 1) score += 30* 0.9;
    q[id].push_back(Pro("YouTube", user[user_num], data[data_num],
                        number[number_num], Possible[poss_num], score));
}
void do_Bank(int id) {
    string user[10] = {"transfer", "consume", "others"};
    int user_num = rand() % 3;
    int score = 0;
    if (user_num == 0) score += 20 * 0.8;
```

```
    if (user_num == 1) score += 20 * 0.2;
    if (user_num == 2) score += 20 * 0.1;
    string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
    int data_num = rand() % 4;
    if (data_num == 0) score += 30 * 0.8;
    if (data_num == 1) score += 30 * 0.8;
    if (data_num == 2) score += 30 * 0.2;
    if (data_num == 3) score += 30 * 0.2;
    string number[10] = {"100%", "25%", "50%", "75%"};
    int number_num = rand() % 4;
    if (number_num == 0) score += 20 * 0.9;
    if (number_num == 1) score += 20 * 0.6;
    if (number_num == 2) score += 20 * 0.7;
    if (number_num == 3) score += 20 * 0.8;
    string Possible[10] = {"frequently", "Suddenly"};
    int poss_num = rand() % 2;
    if (poss_num == 0) score += 30* 0.6;
    if (poss_num == 1) score += 30* 0.9;
    q[id].push_back(Pro("Bank", user[user_num], data[data_num],
                                    number[number_num], Possible[poss_num], score));

}
void do_Gps(int id) {
    string user[10] = {"often", "regular", "Irregular"};
    int user_num = rand() % 3;
    int score = 0;
    if (user_num == 0) score += 20 * 0.8;
    if (user_num == 1) score += 20 * 0.2;
    if (user_num == 2) score += 20 * 0.2;
    string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
    int data_num = rand() % 4;
    if (data_num == 0) score += 30 * 0.8;
    if (data_num == 1) score += 30 * 0.8;
    if (data_num == 2) score += 30 * 0.2;
    if (data_num == 3) score += 30 * 0.2;
    string number[10] = {"100%", "25%", "50%", "75%"};
    int number_num = rand() % 4;
    if (number_num == 0) score += 20 * 0.9;
    if (number_num == 1) score += 20 * 0.6;
    if (number_num == 2) score += 20 * 0.7;
    if (number_num == 3) score += 20 * 0.8;
    string Possible[10] = {"frequently", "Suddenly"};
    int poss_num = rand() % 2;
    if (poss_num == 0) score += 30* 0.6;
```

```
    if (poss_num == 1) score += 30* 0.9;
    q[id].push_back(Pro("Gps", user[user_num], data[data_num],
                        number[number_num], Possible[poss_num], score));
}
int main () {
    freopen("out.txt", "w", stdout);
    string name[10];
    name[0] = "Email"; name[1] = "Web"; name[2] = "Game";
    name[3] = "YouTube"; name[4] = "Bank"; name[5] = "Gps";
    map <string, int> id;
    id["Email"] = 0; id["Web"] = 1; id["Game"] = 2;
    id["YouTube"] = 3; id["Bank"] = 4; id["Gps"] = 5;
    vector <int> bin;
    for (int i = 1; i < (1 << 6); i++) {
        int sum = 0;
        for (int j = 0; j < 6; j++) {
            if (i & (1 << j)) {
                sum ++;
            }
        }
        if (sum >= 3 && sum <= 6)
            bin.push_back(i);
    }
    int n = bin.size();
    for (int i = 1; i <= 5000; i++) {
        int num = bin[rand() % n];
        int sum = 0;
        for (int j = 0; j < 6; j++) {
            if (num & (1 << j)) {
                int what = j;
                if (what == 0) do_email(i);
                if (what == 1) do_Web(i);
                if (what == 2) do_Game(i);
                if (what == 3) do_YouTube(i);
                if (what == 4) do_Bank(i);
                if (what == 5) do_Gps(i);
            }
        }
    }
    map <string, int> _score;
    _score["Email"] = 20, _score["Web"] = 30, _score["Game"] = 5;
    _score["YouTube"] = 10, _score["Bank"] = 10, _score["Gps"] = 25;
    for (int i = 1; i <= 5000; i++) {
        int sum = 0;
```

```
            vector <int> all;
            for (int j = 0; j < q[i].size(); j++) {
                sum += _score[q[i][j].name];
                all.push_back(_score[q[i][j].name]);
                printf("%-15d%-15s%-15s%-15s%-15s%-15s%-15d\n", i,
                                                        q[i][j].name.c_str(),
                                                        q[i][j].user.c_str(),
                                                        q[i][j].data.c_str(),
                                                        q[i][j].number.c_str(),
                                                        q[i][j].rate.c_str(),
                                                        q[i][j].score);
            }
            int score = 0;
            for (int j = 0; j < all.size(); j++) {
                score += q[i][j].score * all[j] / sum;
            }
        }
        return 0;
}
```

2、Code of Problem two

Running Environment is Code::Blocks 13.12 under Operating System win10

```
#include <cstdio>
#include <algorithm>
#include <string>
#include <map>
#include <vector>
using namespace std;
struct Pro {
        string name;
        string user;
        string data;
        string number;
        string rate;
        int score;
        Pro(string a = "", string b = "", string c = "", string d = "", string e = "", int _score = 0) {
            name = a, user = b, data = c, number = d, rate = e, score = _score;
        }
};
vector <Pro> q[10000];
void do_email(int id) {
        string user[10] = {"often", "regular", "Irregular"};
        int user_num = rand() % 3;
        int score = 0;
        if (user_num == 0) score += 20 * 0.2;
```

```
        if (user_num == 1) score += 20 * 0.6;
        if (user_num == 2) score += 20 * 0.9;
        string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
        int data_num = rand() % 4;
        if (data_num == 0) score += 30 * 0.8;
        if (data_num == 1) score += 30 * 0.8;
        if (data_num == 2) score += 30 * 0.2;
        if (data_num == 3) score += 30 * 0.2;
        string number[10] = {"100%", "25%", "50%", "75%"};
        int number_num = rand() % 4;
        if (number_num == 0) score += 20 * 0.9;
        if (number_num == 1) score += 20 * 0.6;
        if (number_num == 2) score += 20 * 0.7;
        if (number_num == 3) score += 20 * 0.8;
        string Possible[10] = {"frequently", "Suddenly"};
        int poss_num = rand() % 2;
        if (poss_num == 0) score += 30* 0.6;
        if (poss_num == 1) score += 30* 0.9;
        q[id].push_back(Pro("Email", user[user_num], data[data_num],
                                    number[number_num], Possible[poss_num], score));
}
void do_Web(int id) {
        string user[10] = {"Guns", "News", "Others"};
        int user_num = rand() % 3;
        int score = 0;
        if (user_num == 0) score += 20 * 0.95;
        if (user_num == 1) score += 20 * 0.4;
        if (user_num == 2) score += 20 * 0.2;
        string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
        int data_num = rand() % 4;
        if (data_num == 0) score += 30 * 0.8;
        if (data_num == 1) score += 30 * 0.8;
        if (data_num == 2) score += 30 * 0.2;
        if (data_num == 3) score += 30 * 0.2;
        string number[10] = {"100%", "25%", "50%", "75%"};
        int number_num = rand() % 4;
        if (number_num == 0) score += 20 * 0.9;
        if (number_num == 1) score += 20 * 0.6;
        if (number_num == 2) score += 20 * 0.7;
        if (number_num == 3) score += 20 * 0.8;
        string Possible[10] = {"frequently", "Suddenly"};
        int poss_num = rand() % 2;
        if (poss_num == 0) score += 30* 0.6;
        if (poss_num == 1) score += 30* 0.9;
```

```
        q[id].push_back(Pro("Web", user[user_num], data[data_num],
                                    number[number_num], Possible[poss_num], score));
}
void do_Game(int id) {
        string user[10] = {"challenge", "character", "Strategy", "others"};
        int user_num = rand() % 4;
        int score = 0;
        if (user_num == 0) score += 20 * 0.8;
        if (user_num == 1) score += 20 * 0.7;
        if (user_num == 2) score += 20 * 0.8;
        if (user_num == 3) score += 20 * 0.1;
        string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
        int data_num = rand() % 4;
        if (data_num == 0) score += 30 * 0.8;
        if (data_num == 1) score += 30 * 0.8;
        if (data_num == 2) score += 30 * 0.2;
        if (data_num == 3) score += 30 * 0.2;
        string number[10] = {"100%", "25%", "50%", "75%"};
        int number_num = rand() % 4;
        if (number_num == 0) score += 20 * 0.9;
        if (number_num == 1) score += 20 * 0.6;
        if (number_num == 2) score += 20 * 0.7;
        if (number_num == 3) score += 20 * 0.8;
        string Possible[10] = {"frequently", "Suddenly"};
        int poss_num = rand() % 2;
        if (poss_num == 0) score += 30* 0.6;
        if (poss_num == 1) score += 30* 0.9;
        q[id].push_back(Pro("Game", user[user_num], data[data_num],
                                    number[number_num], Possible[poss_num], score));
}
void do_YouTube(int id) {
        string user[10] = {"Violence", "information", "others"};
        int user_num = rand() % 3;
        int score = 0;
        if (user_num == 0) score += 20 * 0.9;
        if (user_num == 1) score += 20 * 0.8;
        if (user_num == 2) score += 20 * 0.1;
        string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
        int data_num = rand() % 4;
        if (data_num == 0) score += 30 * 0.8;
        if (data_num == 1) score += 30 * 0.8;
        if (data_num == 2) score += 30 * 0.2;
        if (data_num == 3) score += 30 * 0.2;
        string number[10] = {"100%", "25%", "50%", "75%"};
```

```
    int number_num = rand() % 4;
    if (number_num == 0) score += 20 * 0.9;
    if (number_num == 1) score += 20 * 0.6;
    if (number_num == 2) score += 20 * 0.7;
    if (number_num == 3) score += 20 * 0.8;
    string Possible[10] = {"frequently", "Suddenly"};
    int poss_num = rand() % 2;
    if (poss_num == 0) score += 30* 0.6;
    if (poss_num == 1) score += 30* 0.9;
    q[id].push_back(Pro("YouTube", user[user_num], data[data_num],
                        number[number_num], Possible[poss_num], score));
}
void do_Bank(int id) {
    string user[10] = {"transfer", "consume", "others"};
    int user_num = rand() % 3;
    int score = 0;
    if (user_num == 0) score += 20 * 0.8;
    if (user_num == 1) score += 20 * 0.2;
    if (user_num == 2) score += 20 * 0.1;
    string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
    int data_num = rand() % 4;
    if (data_num == 0) score += 30 * 0.8;
    if (data_num == 1) score += 30 * 0.8;
    if (data_num == 2) score += 30 * 0.2;
    if (data_num == 3) score += 30 * 0.2;
    string number[10] = {"100%", "25%", "50%", "75%"};
    int number_num = rand() % 4;
    if (number_num == 0) score += 20 * 0.9;
    if (number_num == 1) score += 20 * 0.6;
    if (number_num == 2) score += 20 * 0.7;
    if (number_num == 3) score += 20 * 0.8;
    string Possible[10] = {"frequently", "Suddenly"};
    int poss_num = rand() % 2;
    if (poss_num == 0) score += 30* 0.6;
    if (poss_num == 1) score += 30* 0.9;
    q[id].push_back(Pro("Bank", user[user_num], data[data_num],
                        number[number_num], Possible[poss_num], score));
}
void do_Gps(int id) {
    string user[10] = {"often", "regular", "Irregular"};
    int user_num = rand() % 3;
    int score = 0;
    if (user_num == 0) score += 20 * 0.8;
    if (user_num == 1) score += 20 * 0.2;
```

```
        if (user_num == 2) score += 20 * 0.2;
        string data[10] = {"23:00-3:00", "4:00-7:00", "8:00-12:00", "13:00-22:00"};
        int data_num = rand() % 4;
        if (data_num == 0) score += 30 * 0.8;
        if (data_num == 1) score += 30 * 0.8;
        if (data_num == 2) score += 30 * 0.2;
        if (data_num == 3) score += 30 * 0.2;
        string number[10] = {"100%", "25%", "50%", "75%"};
        int number_num = rand() % 4;
        if (number_num == 0) score += 20 * 0.9;
        if (number_num == 1) score += 20 * 0.6;
        if (number_num == 2) score += 20 * 0.7;
        if (number_num == 3) score += 20 * 0.8;
        string Possible[10] = {"frequently", "Suddenly"};
        int poss_num = rand() % 2;
        if (poss_num == 0) score += 30* 0.6;
        if (poss_num == 1) score += 30* 0.9;
        q[id].push_back(Pro("Gps", user[user_num], data[data_num],
                                    number[number_num], Possible[poss_num], score));
}
int main () {
        freopen("two.txt", "w", stdout);
        string name[10];
        name[0] = "Email"; name[1] = "Web"; name[2] = "Game";
        name[3] = "YouTube"; name[4] = "Bank"; name[5] = "Gps";
        map <string, int> id;
        id["Email"] = 0; id["Web"] = 1; id["Game"] = 2;
        id["YouTube"] = 3; id["Bank"] = 4; id["Gps"] = 5;
        vector <int> bin;
        for (int i = 1; i < (1 << 6); i++) {
                int sum = 0;
                for (int j = 0; j < 6; j++) {
                        if (i & (1 << j)) {
                                sum ++;
                        }
                }
                if (sum >= 3 && sum <= 6)
                        bin.push_back(i);
        }
        int n = bin.size();
        for (int i = 1; i <= 5000; i++) {
                int num = bin[rand() % n];
                int sum = 0;
                for (int j = 0; j < 6; j++) {
```

```
            if (num & (1 << j)) {
                int what = j;
                if (what == 0) do_email(i);
                if (what == 1) do_Web(i);
                if (what == 2) do_Game(i);
                if (what == 3) do_YouTube(i);
                if (what == 4) do_Bank(i);
                if (what == 5) do_Gps(i);
            }
        }
    }
    map <string, int> _score;
    _score["Email"] = 20, _score["Web"] = 30, _score["Game"] = 5;
    _score["YouTube"] = 10, _score["Bank"] = 10, _score["Gps"] = 25;
    vector <pair <int, int> > out;
    for (int i = 1; i <= 5000; i++) {
        int sum = 0;
        vector <int> all;
        for (int j = 0; j < q[i].size(); j++) {
            sum += _score[q[i][j].name];
            all.push_back(_score[q[i][j].name]);
        }
        int score = 0;
        for (int j = 0; j < all.size(); j++) {
            score += q[i][j].score * all[j] / sum;
        }
        out.push_back(make_pair(score, i));
    }
    sort(out.begin(), out.end());
    printf("########################\n");
    for (int i = 0; i < 5; i++) {
        int id = out[i].second;
        int score = out[i].first;
        for (int j = 0; j < q[id].size(); j++) {
            printf("%-15d%-15s%-15s%-15s%-15s%-15s\n", id,
                                                    q[id][j].name.c_str(),
                                                    q[id][j].user.c_str(),
                                                    q[id][j].data.c_str(),
                                                    q[id][j].number.c_str(),
                                                    q[id][j].rate.c_str());
        }
        printf("%d\n", score);
    }
    printf("########################\n");
```

```
    for (int i = 1250; i < 1255; i++) {
        int id = out[i].second;
        int score = out[i].first;
        for (int j = 0; j < q[id].size(); j++) {
            printf("%-15d%-15s%-15s%-15s%-15s%-15s\n", id,
                                                    q[id][j].name.c_str(),
                                                    q[id][j].user.c_str(),
                                                    q[id][j].data.c_str(),
                                                    q[id][j].number.c_str(),
                                                    q[id][j].rate.c_str());
        }
        printf("%d\n", score);
    }
    printf("########################\n");
    for (int i = 2500; i < 2505; i++) {
        int id = out[i].second;
        int score = out[i].first;
        for (int j = 0; j < q[id].size(); j++) {
            printf("%-15d%-15s%-15s%-15s%-15s%-15s\n", id,
                                                    q[id][j].name.c_str(),
                                                    q[id][j].user.c_str(),
                                                    q[id][j].data.c_str(),
                                                    q[id][j].number.c_str(),
                                                    q[id][j].rate.c_str());
        }
        printf("%d\n", score);
    }
    printf("########################\n");
    for (int i = 4994; i < 5000; i++) {
        int id = out[i].second;
        int score = out[i].first;
        for (int j = 0; j < q[id].size(); j++) {
            printf("%-15d%-15s%-15s%-15s%-15s%-15s\n", id,
                                                    q[id][j].name.c_str(),
                                                    q[id][j].user.c_str(),
                                                    q[id][j].data.c_str(),
                                                    q[id][j].number.c_str(),
                                                    q[id][j].rate.c_str());
        }
        printf("%d\n", score);
    }
    return 0;
}
```

3、Result of Problem one

Run the Code of Problem one

1	Email	Irregular	8:00-12:00	100%	Suddenly	69
1	Web	News	8:00-12:00	50%	frequently	46
1	Game	challenge	4:00-7:00	25%	Suddenly	79
1	YouTube	Violence	4:00-7:00	75%	Suddenly	85
1	Bank	others	13:00-22:00	100%	Suddenly	53
1	Gps	often	8:00-12:00	25%	frequently	52
2	YouTube	Violence	23:00-3:00	50%	Suddenly	83
2	Bank	others	8:00-12:00	75%	frequently	42
2	Gps	often	23:00-3:00	75%	Suddenly	83
3	Web	Guns	13:00-22:00	75%	frequently	59
3	Game	character	4:00-7:00	100%	Suddenly	83
3	YouTube	information	4:00-7:00	100%	Suddenly	85
3	Gps	Irregular	8:00-12:00	25%	Suddenly	49
4	Web	Others	4:00-7:00	25%	frequently	58
4	Game	challenge	13:00-22:00	50%	frequently	54
4	Bank	transfer	8:00-12:00	100%	frequently	58
5	Web	News	8:00-12:00	25%	frequently	44
5	Bank	transfer	8:00-12:00	50%	frequently	54
5	Gps	often	4:00-7:00	100%	Suddenly	85
6	Game	challenge	8:00-12:00	100%	frequently	58
6	YouTube	information	23:00-3:00	75%	frequently	74
6	Gps	often	13:00-22:00	50%	Suddenly	63
7	Email	regular	8:00-12:00	50%	Suddenly	59
7	YouTube	others	4:00-7:00	75%	Suddenly	69
7	Bank	transfer	23:00-3:00	50%	Suddenly	81
7	Gps	Irregular	4:00-7:00	50%	Suddenly	69
8	Email	Irregular	23:00-3:00	50%	Suddenly	83
8	Game	character	4:00-7:00	25%	frequently	68
8	Gps	regular	8:00-12:00	25%	frequently	40
9	Email	Irregular	13:00-22:00	50%	Suddenly	65
9	Web	News	8:00-12:00	50%	Suddenly	55
9	YouTube	information	8:00-12:00	50%	Suddenly	63
9	Bank	others	13:00-22:00	25%	Suddenly	47
10	Game	character	13:00-22:00	25%	Suddenly	59
10	YouTube	information	8:00-12:00	25%	frequently	52
10	Gps	Irregular	8:00-12:00	50%	frequently	42
11	Email	Irregular	23:00-3:00	75%	frequently	76
11	Web	Others	8:00-12:00	75%	frequently	44
11	Game	character	23:00-3:00	75%	Suddenly	81
11	Bank	transfer	8:00-12:00	50%	Suddenly	63
11	Gps	Irregular	8:00-12:00	100%	frequently	46

4、Result of Problem two

Run the Code of Problem two

```
#########################
738             Email           often           13:00-22:00     25%             frequently
738             Web             Others          13:00-22:00     50%             frequently
738             Game            others          8:00-12:00      50%             Suddenly
738             Bank            consume         8:00-12:00      100%            frequently
738             Gps             Irregular       13:00-22:00     50%             frequently
40
3265            YouTube         others          8:00-12:00      50%             frequently
3265            Bank            others          13:00-22:00     25%             Suddenly
3265            Gps             Irregular       8:00-12:00      25%             frequently
40
1850            Game            character       8:00-12:00      50%             frequently
1850            YouTube         others          13:00-22:00     100%            frequently
1850            Bank            consume         13:00-22:00     25%             frequently
1850            Gps             regular         13:00-22:00     25%             frequently
41
3496            Web             News            13:00-22:00     25%             frequently
3496            Game            others          13:00-22:00     100%            frequently
3496            Bank            others          13:00-22:00     25%             frequently
41
67              Web             Others          8:00-12:00      25%             frequently
67              YouTube         Violence        8:00-12:00      25%             frequently
67              Bank            consume         8:00-12:00      25%             frequently
42
#########################
1549            Email           regular         13:00-22:00     75%             Suddenly
1549            Web             Others          8:00-12:00      25%             Suddenly
1549            Bank            others          4:00-7:00       50%             frequently
1549            Gps             regular         4:00-7:00       25%             Suddenly
56
1599            Email           Irregular       4:00-7:00       25%             frequently
1599            Web             Others          8:00-12:00      25%             Suddenly
1599            YouTube         others          8:00-12:00      75%             Suddenly
1599            Bank            transfer        13:00-22:00     100%            frequently
56
1630            Email           regular         13:00-22:00     75%             frequently
1630            Game            challenge       13:00-22:00     75%             frequently
1630            YouTube         Violence        13:00-22:00     100%            Suddenly
56
1652            Game            Strategy        23:00-3:00      25%             Suddenly
1652            YouTube         Violence        13:00-22:00     50%             Suddenly
```