ATM Transaction Status Anomaly Detection Based on Unsupervised Learning

Yuan-ming ZHANG\textsuperscript{1,*}, Dan-ni LU\textsuperscript{2}, Jing ZHOU\textsuperscript{3} and Guo-liang YUAN\textsuperscript{1}

\textsuperscript{1}College of Information Engineering, Shanghai Maritime University, Shanghai 201306, China
\textsuperscript{2}College of Transport & Communications, Shanghai Maritime University, Shanghai 201306, China
\textsuperscript{3}School of Software Engineering, East China Normal University, Shanghai 200062, China

*Corresponding author

Keywords: ATM transaction status, Unsupervised learning, K-means clustering, SOM neural network.

Abstract. This paper uses information technology to monitor the data obtained by ATM equipment in real-time (three indicators of traffic volume, transaction success rate and transaction response time) and constructs an anomaly detection scheme for unsupervised learning (K-means clustering and SOM neural network). After that, the data in a day was simulated to consider the above schemes. Both schemes can make decisions quickly and have sound anomaly detection effects.

Introduction

With the advancement of technological innovation, we are entering an era of consumer-to-bank relationship transformation, and we have embraced the era of digital banking. ATM equipment are also becoming more and more popular, which brings great convenience for us to withdraw and deposit money. However, occasionally ATM equipment will fail, not only give us a lot of trouble, but also give some criminals loopholes to do something illegal during the failure. Therefore, the bank must be able to monitor the fault in real time and send out a warning.

Glossary

In order to grasp the business status of the whole bank in real-time, the data center monitoring system of the head office of the commercial bank summarizes the transaction information of each branch every minute. The summary information includes three indicators: business volume, transaction success rate, and transaction response time. The indicators are explained as follows:

1. Business volume: the total number of transactions that occurred per minute;
2. Transaction success rate: the ratio of the number of successful transactions per minute to the volume of business;
3. Transaction response time: The average time (in milliseconds) that each transaction is processed in the back-end within one minute.

Model Hypothesis

1. The daily business volume obeys a similar distribution, the business volume during the working time is large, and the non-working business volume is small. The fitting situation applies to the daily trading volume (after standardization);
2. The abnormality of the three indicators of traffic volume, transaction success rate and transaction response time may lead to the failure of the transaction status;
3. ATM failure should account for a small part of total business transaction, and it should only be a small probability event.
ATM Anomaly Detection Model

Unsupervised Anomaly Detection Scheme Based on K-means Clustering

We first selected the K-means clustering method to classify the samples into normal and abnormal. The K-means algorithm is the most classical partition-based clustering method. The basic idea is to cluster the k points in the space and classify the objects closest to them. The iterative method is used to update the values of each cluster center one by one until the best clustering result is obtained.

Suppose to divide the sample set into c categories. The algorithm is described as follows:

1. Appropriately select the initial centers from c categories;
2. In the kth iteration, for any sample, find the distance to the c centers, and classify the sample into the class with the shortest center;
3. Update the center value of the class using methods such as mean;
4. For all c cluster centers, if the value remains unchanged after the iterative method of (2)(3) is updated, the iteration ends. Otherwise, the iteration is continued.

The most significant advantage of this algorithm is fast and straightforward. The key to the algorithm is the initial center selection and distance formula. We use the kmeans () function in R to perform cluster analysis on samples. We selected different cluster centers and selected 2, 4, 8, 16, 30, and 50 cluster centers, respectively. The clustering results are as follows:

![Figure 1. Clustering effect of different cluster centers.](image)

When a category is divided into fewer samples, and the variance of these samples is small, we consider the class to be a valid exception class. Define abnormal validity $\rho$ and intraclass coefficient of variation $CV$:

$$\rho_i = \frac{x_i}{x}$$

$$CV_i = \frac{\sum_{j=1}^{M} sd(a_{ij})}{mean(a_{ij})}$$ (2)

In the formula, $x$ is the total sample size, $x_i$ is the sample size of class $i$, $\rho_i$ is the abnormal validity of class $i$, $M$ is the number of features, $sd(a_{ij})$ is the standard deviation of class $i$ in the $j$th sample parameters, and $mean(a_{ij})$ is the mean of the $j$th parameter of the sample in the $i$th class.
In general, we believe that the clustering results are meaningful when the abnormal validity is above 99%, and the intra-class variation should not be too high. The following table lists the exceptions for different cluster centers:

| Classification | Significant abnormal validity | The most significant indicator of a sample within the class and the variation within the class |
|----------------|-------------------------------|---------------------------------------------------------------------------------------------|
| 2 types        | 0.8267271                    | Clustering effect is meaningless                                                            |
| 4 types        | 0.9743995                    | Clustering effect is not good CV=3.256                                                     |
| 8 types        | 0.9888103, 0.9864136         | Clustering is general CV=2.821                                                            |
| 16 types       | 0.999855, 0.9990077          | (0.22382810.36733627.39)CV=0.543                                                           |
| 30 types       | 0.9990077, 0.999855          | (0.71689620.600016896.00)CV=0.543                                                           |
| 50 types       | 0.9998626, 0.9986414         | (0.58544030.000057174.78)CV=0.444                                                           |

When we select two cluster centers, we can see that there is no meaning of classification. So we added clustering center, it can be seen that with the increase of clustering center, the clustering effect has been significantly improved, and more and more abnormal classes have appeared. When testing 16 or more cluster centers, 19 samples were always in a small group, and it was considered that these 19 samples had transaction failures. These 19 samples are as follows:

| Date    | Time | Business volume | Transaction success | Response time |
|---------|------|-----------------|---------------------|---------------|
| 3.23    | 00:48| 11              | 0.1818              | 46256         |
| 3.23    | 00:49| 13              | 0.3077              | 39376         |
| 3.23    | 00:50| 8               | 0                   | 50624         |
| 3.23    | 00:51| 17              | 0.0588              | 53543         |
| 3.23    | 00:52| 22              | 0.1364              | 49018         |
| 3.23    | 00:53| 15              | 0.1333              | 49593         |
| 3.23    | 00:54| 14              | 0                   | 57211         |
| 3.23    | 00:55| 18              | 0.0556              | 53889         |
| 3.23    | 00:56| 12              | 0.1667              | 47397         |
| 3.23    | 00:57| 11              | 0.0909              | 51697         |
| 3.23    | 00:58| 10              | 0                   | 56758         |
| 3.23    | 00:59| 16              | 0.1875              | 45991         |
| 3.23    | 1:00 | 19              | 0.2632              | 44476         |
| 3.23    | 1:01 | 24              | 0.7917              | 28820         |
| 4.16    | 4:01 | 17              | 0.1765              | 46529.53      |
| 4.16    | 6:00 | 49              | 0.3673              | 33627.39      |
| 4.16    | 6:01 | 18              | 0                   | 57174.78      |
| 4.16    | 6:02 | 18              | 0                   | 57177.44      |
| 4.16    | 6:03 | 17              | 0                   | 57067.59      |

When we select the two cluster centers, we can see that there is no meaning of classification. So we added clustering center, it can be seen that with the increase of clustering center, the clustering effect has been significantly improved, and more and more abnormal classes have appeared. When testing 16 or more cluster centers, 19 samples were always in a small group, and it was considered that these 19 samples had transaction failures. These 19 samples are as follows:

| Date    | Time | Business volume | Transaction success | Response time |
|---------|------|-----------------|---------------------|---------------|
| 3.23    | 00:48| 11              | 0.1818              | 46256         |
| 3.23    | 00:49| 13              | 0.3077              | 39376         |
| 3.23    | 00:50| 8               | 0                   | 50624         |
| 3.23    | 00:51| 17              | 0.0588              | 53543         |
| 3.23    | 00:52| 22              | 0.1364              | 49018         |
| 3.23    | 00:53| 15              | 0.1333              | 49593         |
| 3.23    | 00:54| 14              | 0                   | 57211         |
| 3.23    | 00:55| 18              | 0.0556              | 53889         |
| 3.23    | 00:56| 12              | 0.1667              | 47397         |
| 3.23    | 00:57| 11              | 0.0909              | 51697         |
| 3.23    | 00:58| 10              | 0                   | 56758         |
| 3.23    | 00:59| 16              | 0.1875              | 45991         |
| 3.23    | 1:00 | 19              | 0.2632              | 44476         |
| 3.23    | 1:01 | 24              | 0.7917              | 28820         |
| 4.16    | 4:01 | 17              | 0.1765              | 46529.53      |
| 4.16    | 6:00 | 49              | 0.3673              | 33627.39      |
| 4.16    | 6:01 | 18              | 0                   | 57174.78      |
| 4.16    | 6:02 | 18              | 0                   | 57177.44      |
| 4.16    | 6:03 | 17              | 0                   | 57067.59      |

It can be seen from the table that there is a significant transaction failure phenomenon on March 23 and April 16, and the clustering scheme can significantly identify the fault group. When the clustering center is selected to be large, it can separate the significant fault samples, but there may be a high risk of missed judgment. Therefore, we can select multiple minimum classes, but the false positive rate will also rise accordingly.

**Unsupervised Anomaly Detection Scheme Based on SOM**

We tried a self-organizing (competitive) neural network (SOM) to cluster the samples.

SOM and the current popular neural network model are structurally similar and consist of elementary neuron structures, but SOM is a type of "unsupervised learning" model. In the network structure, it is generally a two-layer network composed of an input layer and a competition layer; the neurons between the two layers implement bidirectional connection, and the network has no hidden layer. Sometimes there are horizontal connections between neurons in the competition layer. In learning algorithms, it simulates the dynamics of information processing between excitability,
coordination, and inhibition, and competition between biological neurons to guide the learning and work of the network, unlike the network of multi-layer neural networks (MLP) which the error is used as a criterion for the algorithm. The basic idea of competing for neural networks is that the neurons in the competition layer of the network compete for the opportunity to respond to the input mode. Finally, only one neuron becomes the winner of the competition. This winning neuron represents the classification of the input pattern. Therefore, it is easy to associate such results with clustering.

Therefore, we use the selforgmap() function in the neural network toolbox in MATLAB to construct the SOM network and then substituting the samples into the training to obtain the clustering results. The MATLAB code for the training process is shown in the appendix. We observe the clustering results obtained when setting different size competition layers:

From the clustering results, we can see that when the competition layer is selected 2*2 and 3*3, it does not achieve a good clustering effect. When the competition layer selects 3*5, there are apparent abnormal classes, and 18 samples are classified into one class. We take these 18 samples and find the same results with K-means clustering, and with the expansion of the competition layer, these 18 samples have been separated into a single category (marked by stars in Fig.2). Moreover, due to the characteristics of SOM clustering, similar two types will be together. We can see that the large-scale competition layer is still an abnormal subclass around the significant anomaly class, so we can consider that the neighborhood sample of the significant exception class is also a regional anomaly.

Model Checking
We randomly simulated the data for one day to test the performance of the above two schemes. In the simulation data of the day, we specially added a significant fault in the early morning and added some fuzzy samples in the evening. We selected the normal data of a specific day and randomly changed 23 abnormal data.

We take the samples one by one, in order to make timely alarms when the fault is detected, and use reasonable setting parameters in each scheme, so that the running time of the scheme should not be too long, and the maximum time is less than 1 minute (the time period of the sample), in this way, to prevent the accumulation of samples to be tested, and to make timely and accurate warnings of faults. In the unsupervised scheme, we use some auxiliary samples as the background to see the clustering of the samples to be tested.

Results Based on K-means Clustering Scheme
When we adopted this scheme, we chose a cluster center of 30. If the sample to be tested is clustered into the smallest class, we think that it has an abnormality. If it is gathered into the subclass, we think
that it may cause parameter fluctuation due to the abnormal system, which may cause anomalies. The specific results are as follows:

```
The ATM system may be busy in 51
The ATM system may be out of order in 158
The ATM system may be out of order in 255
The ATM system may be busy in 303
The ATM system may be busy in 312
The ATM system may be busy in 345
The ATM system may be busy in 400
The ATM system may be busy in 420
The ATM system may be busy in 454
The ATM system may be busy in 614
The ATM system may be out of order in 1006
The ATM system may be busy in 1007
The ATM system may be busy in 1008
The ATM system may be busy in 1010
The ATM system may be busy in 1011
The ATM system may be busy in 1012
The ATM system may be out of order in 1013
The ATM system may be out of order in 1014
The ATM system may be out of order in 1015
The ATM system may be busy in 1016
The ATM system may be out of order in 1017
The ATM system may be out of order in 1018
The ATM system may be out of order in 1759
The ATM system may be out of order in 1800
The ATM system may be busy in 1801
The ATM system may be out of order in 1802
The ATM system may be out of order in 1803
The ATM system may be out of order in 2208
```

Figure 3. Verifying the results of the K-means clustering scheme.

It can be seen that the program screened 13 abnormal points and 15 possible abnormal points. Among them, 23 points that we artificially set are all screened out. We found the remaining abnormal points that the transaction success rate is <80%. It has a particular meaning to be screened out, so we can prevent abnormalities here. The program has a single run time of 0.52s. It can be seen that the scheme has a strong screening ability and reliable usability.

**Results Based on SOM Clustering Scheme**

In this scenario, we choose a 5*5 competition layer. If the sample to be tested is clustered into the smallest class, we think that it has an exception. If it is clustered into the neighborhood of the smallest class, we think it may be abnormal and it should be filtered out.

The results of this program are similar to the K-means scheme. They filtered out 14 abnormal points and 18 possible abnormal points. Among them, 23 points that we artificially set are all screened out, and the remaining possible abnormal points have guiding significance for the status of the ATM trading system, that is, the system may be busy and alert for abnormality. The program is slow to train, with a single run time of 8.8s. The program runs slower while the quality of the training is higher than that of the K-means program, which can reduce the false positives while making predictions immediately.

**Conclusion**

The two unsupervised schemes (K-means clustering and SOM (self-organizing map) neural network) constructed in this paper cluster abnormal samples without prior information, and can accurately identify the anomalies and give them timely alerts. In the end, we got 19 significant abnormal periods from the four months of data to determine the ATM failure. In this scheme, we can choose a smaller
class to make a significant abnormal decision and can warn the second subclass as an approximate exception class, which has significance in practical applications.

The evaluation method system for ATM system anomaly detection can also be widely applied to anomaly detection for power grids, networks, and similar anomaly detection based on measured data, and the number of anomaly detection indicators can be limitedly increased.

Acknowledgement

This research was financially supported by the China National University Students Innovation and Entrepreneurship Training Program Foundation.

References

[1] Gecheng Chen, Yue Liu, Zhiqiang Ge. K-means Bayes algorithm for imbalanced fault classification and big data application[J]. Journal of Process Control, 2019, 81.

[2] Feng Nan, Yang Li, Xue Yong Jia, Li Yan Dong, Yong Jie Chen. Application of improved SOM network in gene data cluster analysis[J]. Measurement, 2019, 145.

[3] Zaffar Haider Janjua, Massimo Vecchio, Mattia Antonini, Fabio Antonelli. IRESE: An intelligent rare-event detection system using unsupervised learning on the IoT edge[J]. Engineering Applications of Artificial Intelligence, 2019, 84.

[4] Zhe Li, Jingyue Li, Yi Wang, Kesheng Wang. A deep learning approach for anomaly detection based on SAE and LSTM in mechanical equipment[J]. The International Journal of Advanced Manufacturing Technology, 2019, 103(1-4).

[5] Maurizio Aiello, Maurizio Mongelli, Marco Muselli, Damiano Verda. Unsupervised learning and rule extraction for Domain Name Server tunneling detection[J]. Internet Technology Letters, 2019, 2(2).

[6] Christopher Chung Neo, Mirza Bilal, Choi Howard, Wang Jie, Wang Ding, Ping Peipei, Wang Wei. Unsupervised Classification of Multi-Omics Data during Cardiac Remodeling using Deep Learning.[J]. Methods (San Diego, Calif.), 2019.

[7] Zhang Shibin. Mathematical modeling ideas and methods. Shanghai: SJTU Press, 2015.