Identification of Unexpected Behavior of an Automatic Teller Machine Using Principal Component Analysis Models

Rimvydas Simutis[1], Darius Dilijonas[1], and Lidija Bastina[2]

[1] Vilnius University, Kaunas Faculty of Humanities, Muitines 8, Kaunas, Lithuania
Rimvydas.Simutis@ktu.lt, Darius.Dilijonas@vukhf.lt
[2] JSC Penkių kontinentų bankinės technologijos, Kalvarijų 142, Vilnius, Lithuania
Lidija.B@5ci.lt

Abstract. Early detection of the unexpected behavior of the automatic teller machine (ATM) is crucial for efficient functioning of ATM networks. Because of the high service costs it is very expensive to employ human operators to supervise all ATMs in an ATM network. This paper proposes an automatic identification procedure based on PCA models to supervise continually the ATM networks. This automatic procedure allows detecting the unexpected behavior of the specific automatic teller machine in an ATM network. The proposed procedure has been tested using simulations studies and real experimental data. The simulation results and the first real tests show the efficiency of the proposed procedure. Currently the proposed identification procedure is being implemented in professional software for supervision and control of ATM networks.

Keywords: Automatic teller machine, principal component analysis, ATM network supervision, unexpected behavior.

1 Introduction

Automatic teller machines (ATMs) are computerized telecommunication devices which provide a financial institution's customers a method of financial transactions in a public space without the need for a human clerk. According the estimates developed by ATMIA (ATM Industry Association) the number of ATMs worldwide in 2007 was over 1.6 million. As the ATM networks expand it is very important the proper monitoring, supervision and cash management of the ATM networks [1, 2].

The crucial elements in development of efficient ATM network supervision and management system are creation of the cash demand forecasting models for every ATM and identification of unexpected behaviour of the ATMs in ATM network. The forecasting models have to be created based on historical cash demand data. The historical cash demand for every ATM varies with time and is often overlaid with non stationary behaviour of users and with additional factors, such as paydays, holidays, and seasonal demand of cash in a specific area. Cash drawings are subject to trends and generally follow weekly, monthly and annual cycles. The development of

W. Abramowicz and D. Flejter (Eds.): BIS 2009 Workshop, LNBIP 37, pp. 53–61, 2009.

efficient cash demand forecasting models for ATMs we have introduced in earlier papers [3, 4]. Although these models generally can be used for detection of the outliers in ATMs' cash demand behaviour, they can't state the reason of these outliers. E.g., the wetter conditions can influence the cash demand of a specific ATM significantly, but this behaviour isn't anyhow connected with malfunctions of ATM or clients' illegal actions.

In this paper we propose a new computational procedure for identification of unexpected behaviour of an ATM in ATM network. The procedure is based on application of principal component analysis methods. The unexpected behaviour of an ATM can emerge from different reasons, e.g., it can be bundled with some rising obstacles in the ATM environment, with the operational problems of the ATM, or with clients' illegal actions. It is important to note, that for the identification of the unexpected behaviour of a specific ATM it is necessary to compare the ATM's behaviour with the behaviour of similar ATMs in the neighbourhood. If for some reasons (whether conditions, events in the region, etc.) disturbances are common for all ATMs in neighbourhood, then the changed behaviour of the specific ATM hasn't to be interpreted as unexpected. For the banking institutions it is crucial to identify the unexpected behaviour of an ATM as quick as possible and then act adequately to solve these problems timely. Because of the size of the ATM networks (some service institutions maintain ATM networks with over 1000 ATMs in network) human operators can't supervise efficiently the functioning of all ATMs. Therefore automatic procedures for detection of the unexpected behaviour of the ATMs have to be employed. This paper proposes a new solution for this task.

The paper is structured as follows. After short introduction of the problem in this section, the proposed identification procedure is introduced in section 2. In section 3, simulation studies using the proposed identification procedure are depicted and in section 4 practical tests are presented. Finally, the main results of this work are discussed in section 5.

2 Identification Procedure

To identify whether an ATM in ATM network shows an unexpected behaviour it is important to evaluate carefully the transactions prosecuted on the specific ATM together with the transactions prosecuted on the other ATMs with similar transactions' patterns. Based on this information the conclusions about the disturbances in behaviour of partial ATMs can be made. The proposed identification procedure consists of following steps:

a) Historical data of transactions (cash withdrawal) in ATM network have to be analyzed and clusters of the ATMs which similar behaviour must be formed. Each cluster includes specific number of ATMs. This number j can be defined by the user and in this applications was $j = 4 \div 5$;
b) For every ATM cluster a group of principal component analysis (PCA) models must be build. By development of the PCA models the historical data of ATM transactions are used. Inputs for PCA models are transactions data collected from ATMs cluster. Number of inputs for every model is j-1 and the

total number of PCA models for one ATM cluster is *j*. Each model uses combination of inputs which differs from inputs of other models;

c) If the new data point comes, PCA models should be used to estimate the squared prediction error (*SPE*) between the new sample and its projection into the *k* principal components. These estimations are carried out for all PCA models in ATM cluster. The *SPE* indicates how the transactions data of each ATM group conform to the designed PCA model for that ATM group;

d) If the *SPE* for the analysed group of ATM is bigger than the threshold value, then the conclusion about unexpected behaviour of ATM group is made. Advance analysis of information about the *SPE* in the other groups of the ATM cluster allows to identify the specific ATM showing the unexpected behaviour. This information is provided then to the ATM network operators.

The schema of the proposed identification procedure is presented in the Figure 1.

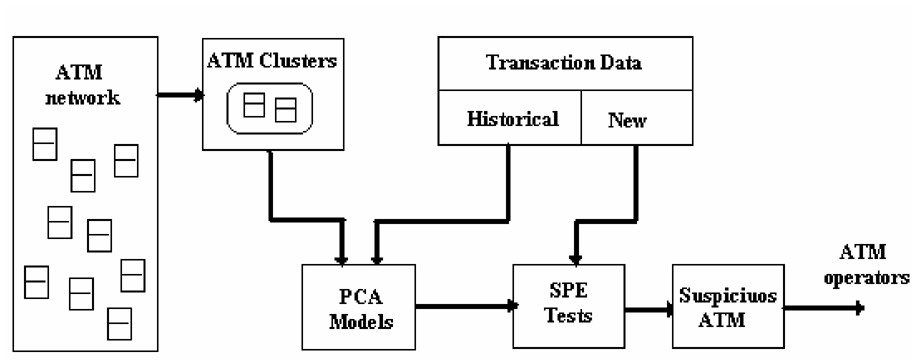

Fig. 1. Schema of the identification procedure for detection of the ATM with unexpected behaviour

In the first step of the procedure the ATM clusters have to been formed. Each cluster typically includes 4÷5 ATMs. The forming of the ATM clusters is based on correlation analysis of historical data. The ATMs with largest correlation coefficients join together in one cluster. In the second step one develops a group of principal component models for every ATM cluster. Principal component analysis is a technique for mapping multidimensional data into lower dimension with minimal loss of information and finding linear combination of the original variables with maximum variability [5,6]. PCA analysis has been extensively applied in various technical applications. Mathematically, PCA relies upon eigenvector decomposition of the covariance matrix of the original process variables. For a given data matrix $\boldsymbol{X}$ with m rows (data points, in our case - daily ATM's transactions) and n columns (variables, number of ATMs in ATM group) the covariance matrix of $\boldsymbol{X}$ is defined as:

$$\boldsymbol{R} = cov(\boldsymbol{X}) = \frac{\boldsymbol{X}^T\boldsymbol{X}}{m-1} \;, \tag{1}$$

where the columns of $\boldsymbol{X}$ have been scaled, i.e. the mean subtracted from each column and divided by the standard deviation. PCA decomposes the data matrix $\boldsymbol{X}$ into the sum of the outer product of so-called score vector t_i and so-called loading vector p_i with a residual error $\boldsymbol{E}$:

$$\boldsymbol{X} = t_1 p_1^T + t_2 p_2^T \ldots. + t_k p_k^T + \boldsymbol{E} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E}\,, \tag{2}$$

where k<m. The first principal component is that linear combination of the columns of $\boldsymbol{X}$ which describes the greatest amount of variability. In the m-dimensional space, p_1 defines the direction of the greatest variability, and t_1 represents the projection of each observation vector onto p_1. The second principal component explains the greatest amount of variability of the residual data. One can proceed in this manner until k principal components are obtained. If the variables in $\boldsymbol{X}$ are correlated, after calculating k $(k<m)$ principal components most of the variation in the data set $\boldsymbol{X}$ has been explained. The score vector t_i contains information on how data points relate to each other. The loading vector p_i contains information how variables relate to each other. The columns of the loading matrix $\boldsymbol{P}$ are the eigenvectors corresponding to the n largest eigenvalues of the covariance matrix $\boldsymbol{R}$.

There are a number of methods that can be used to transform the input data matrix in score and loading vectors. In this case we used Non-linear Iterative Least Squares (NIPALS) method available within Mathworks's MATLAB software package [7].

In the proposed identification procedure the moving window historical data of ATMs transactions (cash withdrawal under normal operation conditions) were used to form the ATM clusters and the ATM groups. After that, the ATMs group matrix $\boldsymbol{X}$, covariance matrix $\boldsymbol{R}$, score matrix $\boldsymbol{T}$ and loading matrix $\boldsymbol{P}$ where determined. When an ATM cluster has j ATMs, then the number of inputs for every PCA model is j-1 and total number of PCA models for one ATM cluster is j. Each PCA model in the ATM cluster uses combination of inputs (cash withdrawal from ATM) which differs from inputs of other models. Once the PCA models for ATM cluster are developed, new observation samples can be projected to the principal component space und the new ATMs data can be tested for possible disturbances and unexpected behaviour. For this purpose in the third step of the identification procedure the PCA models are used to estimate squared prediction error (SPE) between the new sample and its projection into the k principal components, also referred to as the Q statistic [8]. For the new observation vector x_{new} the SPE of the PCA model is estimated using equation

$$SPE = x_{new}\,(\boldsymbol{I} - \boldsymbol{P}_k \boldsymbol{P}_k^T) x_{new}^T\,, \tag{3}$$

where $\boldsymbol{P}_k$ is the matrix of the k loading vectors retained in the partial PCA model and $\boldsymbol{I}$ is the (nxn) identity matrix. The SPE indicates how well the new sample conforms to the PCA model, obtained with historical data. If the SPE of the analysed PCA model exceeds some threshold value (typical value - six squared values of standard deviation of the PCA model, developed with normal operation data), than the forth step of the identification procedure is activated. In this step it is necessary to make an advance analysis of all PCA models developed for one ATM cluster. For the reason that each PCA model has combination of inputs which differs from the inputs of other PCA models in the ATM cluster, it is possible to identify uniquely which input (ATM number) is responsible for the increased SPE value in PCA models.

Consequently the behaviour of this ATM is declared as unexpected and the ATM network operators are informed about this event.

3 Simulation Tests

To test the possibilities of the identification procedure to detect the unexpected behavior of the ATMs (unexpected changes in daily money withdrawal) a special simulation environment was created. An artificial ATM network with 100 ATMs was created and the daily money withdrawals from ATMs were simulated using weekly and monthly seasonality along with long term trends and special events (holiday effects). The simulation environment has imitated the daily money withdrawal from ATMs in typical ATM network in Lithuania. The simulation time was 500 days. Simulation tests and development of PCA models (subroutine *princomp*) were carried out in MATLAB programming environment. The ATM networks' simulation data were processed with correlation technique and according to the correlation coefficients the ATM clusters were formed. Typical money withdrawal patterns for one ATM cluster (4 ATMs) are presented in the Figure 2.

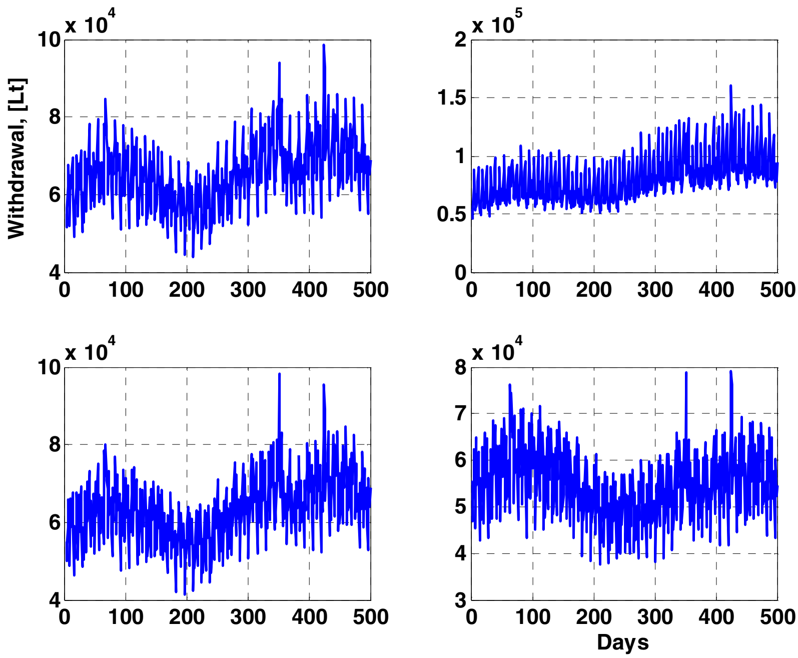

Fig. 2. Daily money withdrawal patterns for one ATM cluster (4 ATMs)

Then the identification procedure depicted above was carried out every day, based on the last 50-days moving window data. For every ATM cluster the four PCA models were build. Each PCA model has three inputs. They are scaled daily money withdrawal data of each ATM. Two principal components are used to describe the variability of the process. After that, the developed PCA models were used to project the new next day observations to the principal component subspace and to estimate the squared prediction error (*SPE*) of the PCA model. The squared prediction errors of PCA models for one ATM cluster with normal operation conditions (without unexpected disturbances) are presented in the Figure 3.

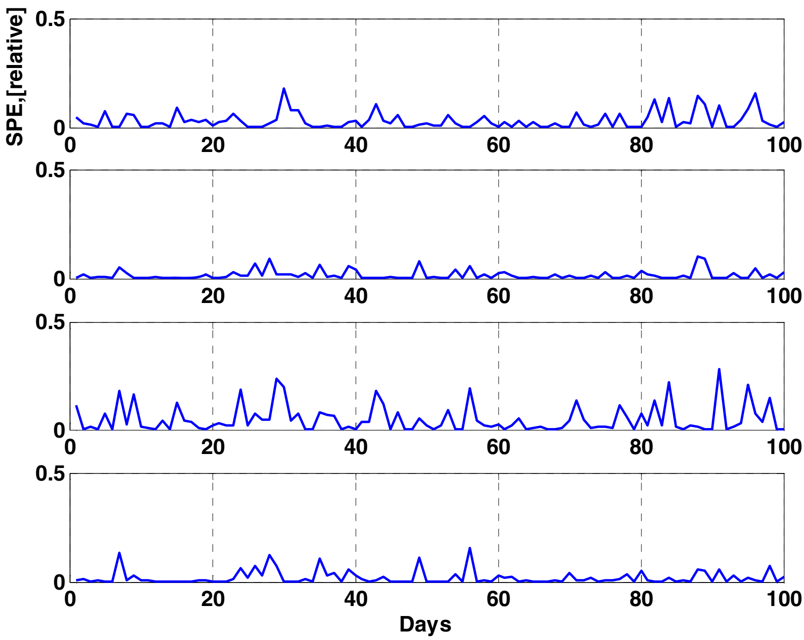

Fig. 3. Squared prediction errors of PCA models for one ATM cluster (4 ATMs) with normal operation conditions

For the same ATM cluster artificial disturbances (money withdrawal disturbances) were imitated. The work of every ATM was disturbed with additional money withdrawal equal to the average daily cash withdrawal. The first ATM was disturbed at *t=15*, second at *t=25*, third at *t=35* and forth at t= 45 days. The squared prediction errors of PCA models for this ATM cluster are presented in the Figure 4. If the *SPE* of the analysed PCA model exceeds the fixed threshold value (six squared values of standard deviation of developed PCA model in normal operation conditions) one can state that unexpected behaviour in the ATM group is observed. The next step is to identity the specific ATM responsible for this behaviour. Since every PCA model in ATM cluster has combination of inputs which differs from inputs of other PCA models it is easy to determine the ATM with unexpected behaviour. For example, in Figure 4, the *SPE* of the PCA models exceeds the threshold value for three ATM groups at time t=15 day. Only for the first ATM group *SPE* is normal at this time. It let to conclude that the unexpected behaviour shows the ATM which isn't included in this group. In this case it is the ATM with Number 1.

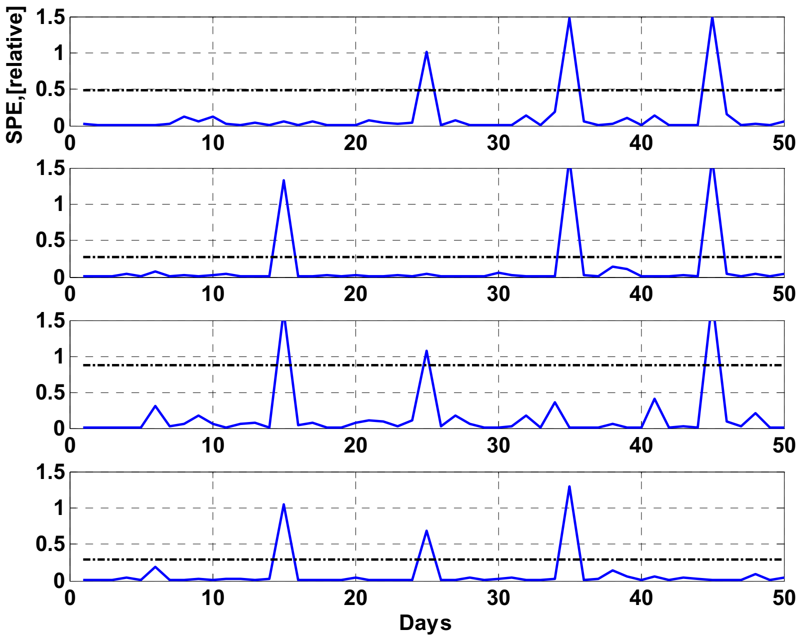

Fig. 4. Squared prediction errors of PCA models for ATM cluster (4 ATMs) with disturbed operation conditions. Dashed lines show the threshold values which are used to define the suspicious ATM groups.

In similar way we can identify easy the ATMs with unexpected behaviour at time t=25, t=35 and t=45 days. In these cases they are the ATMs with Numbers 2, 3 and 4. The simulation results shown in Figure 4 illustrate the efficiency of the proposed identification procedure. Similar results are obtained for the ATM clusters with other transaction patterns. These results confirm that the proposed procedure can be valuable in supervision of real ATM networks.

4 Test in Real ATM Network

The proposed identification procedure was partially tested with operating data from real ATM network. Because of the confidentiality of the problem only the basic information about the test is given here. Daily money withdrawals from 1900 ATMs in time range between 2-3 years have been analyzed and ATM clusters were build. Then the PCA models were developed using the ATM data collected in normal operation conditions for all ATM clusters. Later the PCA models were tested for operation conditions where big disturbances in the functioning of the ATMs had been occurred. The proposed identification procedure allowed detects the unexpected behavior of ATMs approximately in 80% of the real disturbed ATM cases. Figure 5 presents the typical *SPE* patterns by detection of unexpected behavior of

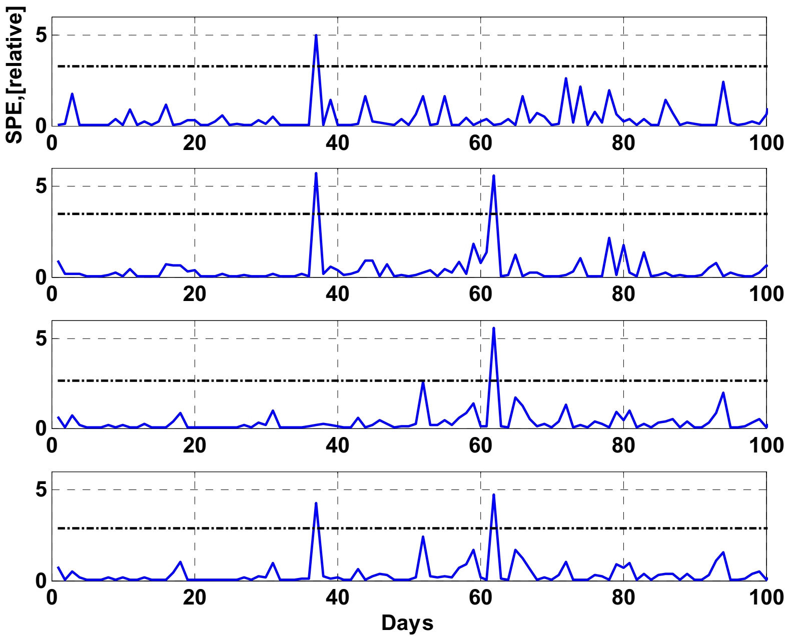

Fig. 5. Typical *SPE* pattern by detecting the unexpected behavior of ATM in ATM cluster. The data comes from a real ATM network. Unexpected behavior is detected for ATM Nr.3 (t=37) and for ATM Nr.1 (t=62).

ATM using PCA models for one ATM cluster. On the day t=37 the identification procedure detected unexpected ATM behavior for ATM Nr.3 and on the day t=62 unexpected behavior for ATM=1 (ATM Nr. 3 isn't included in ATM group 3, and ATM Nr.1 isn't included in ATM group 1). In both cases the detected behaviors match with real functioning disturbances at these ATMs. The performed real tests confirmed the efficiency of the proposed identification procedure. In the further research we will compare these tests with the results obtained using emerging data analysis technique - Exploratory Projection Pursuit (EPP) algorithms [9]. Currently the proposed procedure is being implemented in professional software for supervision and control of ATM networks.

5 Conclusions

Principal component analysis finds and eliminates linear correlation in the data. Here we analyze the possibilities of the application of the PCA models for supervision of ATM network. Early detection of the unexpected behavior of the ATM machines is crucial for efficient functioning of ATM networks. Because of the service costs it is very expensive to employ human operators to supervise continually the ATM network. This paper proposes an automatic identification procedure which is based on PCA models. This procedure allows detecting the unexpected behavior of the specific automatic teller machine in an ATM network. The proposed procedure was tested using simulations tests and real experimental data. The simulation results and the first real tests showed that supervision of ATM network using PCA models is an efficient approach for identification of the unexpected behavior of the specific ATM. Currently the proposed identification procedure is being implemented in professional software for supervision and control of ATM networks.

Acknowledgments. This research work has been supported by means of EU structural funds (Project ASOMIS).

References

1. Snellman, H., Viren, M.: ATM networks and cash usage. Research Discussion Papers. Bank of Finland Nr. 21, 1–33 (2006)
2. Bounie, D., Francois, A.: Cash, Check or Bank Card: The effect of transaction characteristics on the sue of payment instruments. Telecom Paris, Working papers in Economics and Social Sciences, ESS-06-05 (2006)
3. Simutis, R., Dilijonas, D., Bastina, L., Friman, J., Drobinov, P.: Optimization of cash management for ATM network. Informacinės technologijos ir valdymas = Information technology and control 36(1A), 117–121 (2007)
4. Simutis, R., Dilijonas, D., Bastina, L.: Cash demand forecasting for ATM using neural networks and support vector regression algorithms. In: EurOPT 2008: the 20th International Conference Euro Mini Conference on Continuous Optimization and Knowledge-Based Technologies, Neringa, Lithuania, Vilnius, May 20-23, pp. 416–421 (2008) (selected papers), ISBN 978-9955-28-283-9
5. Roffel, B., Betlem, B.: Process Dynamics and Control. John Wiley & Sons, Ltd, Chichester (2006)
6. Jackson, J.E.: A User's Guide to Principal Components. John Wiley, Chichester (2003)
7. Nomikos, P., MacGregor, J.: Multivariate SPC Charts for Monitoring Batch Processes. Technometrics 37, 41–59 (1995)
8. Qin, S.J.: Statistical Process Monitoring: Basics and Beyond. Journal of Chemometrics 17, 480–502 (2003)
9. Friedman, J.H.: Exploratory projection pursuit. J. Am. Statist. Assoc. 82, 249–266 (1987)