Expert Systems with Applications 37 (2010) 6070–6076

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A hybrid model for plastic card fraud detection systems

M. Krivko

Department of Mathematics, University of Leicester, Leicester, LE1 7RH, UK

ARTICLE INFO

Keywords:
Fraud detection
Hybrid model
Plastic card fraud
One-class classification

ABSTRACT

In this paper we present the framework for a hybrid model for plastic card fraud detection systems. The proposed data-customised approach combines elements of supervised and unsupervised methodologies aiming to compensate for the individual deficiencies of the methods. We demonstrate the ability of the hybrid model to identify fraudulent activity on the real debit card transaction data. We also explore the model's efficiency against that of the existing monitoring system of the collaborating bank, using appropriate performance assessment criteria.

1. Introduction

Plastic cards have successfully become an essential part of the modern payment system, providing a broad range of services to the users of the system. Despite being one of the most advanced forms of payment, modern plastic cards still suffer from the same fraud related problems that cash does, namely being counterfeited and stolen. In the present context, we consider plastic card fraud as an unauthorized account activity committed by means of the debit/credit facilities of a legitimate account. In this paper we consider the problem of detecting potentially fraudulent activity on a debit card account and describe the results of our collaborated work with a bank in tackling this issue.

Plastic card fraud is growing along with an increasing volume of payment traffic, advancement and expansion of modern technology, and sophistication of fraudulent tactics. This causes significant losses and great inconvenience to issuing companies, merchants and customers world-wide. In 2007, total card fraud losses on UK issued cards increased by 25% from the previous year and amounted to £535 million (APACS, 2008). The range of fraud tactics observed in the industry can be broadly described within the following categories: lost and stolen card fraud, counterfeit card fraud, card not present fraud, mail non-receipt card fraud, account takeover fraud and application fraud. This list evolves over time as fraudsters adapt new strategies in response to practices of issuing companies and merchants to protect against identified tactics in the future. Currently the largest type of plastic card fraud in the UK is Card-not-Present (CNP) fraud, where the physical card is not present at the point-of-sale (POS). This includes fraud conducted over the Internet, by telephone, fax and mail order and amounts to 54% of all fraud on UK cards. It is expected that the volume of CNP fraud will continue to grow as face-to-face fraudulent transactions become increasingly difficult.

The nature of transaction data and some particular operational issues present a number of challenges for designing a fraud detection system:

- The volume of transactions processed by plastic card issuers daily is high, furthermore each transaction includes more than 70 fields of coded information. Transaction data is heterogeneous and time-varying within and between accounts. Patterns and trends vary significantly for different groups of merchants, holiday seasons and geographical regions.
- The generally accepted fraud rate within the plastic card industry is 0.1–0.2%, i.e. the occurrence of fraud is relatively rare. Frequently this leads to the problem that the majority of cases flagged by the fraud detection system as being potentially fraudulent are in fact legitimate. This type of error is referred to as false positive (FP). As the number of FPs increase so do the associated costs and customer inconvenience.
- Alerts arising from the fraud detection system are usually passed on to the fraud department for further investigation. The suspected cases are followed up with a call to a cardholder for verification of the transactions, where it is required by the bank policy. As a result of this, the number of alerts should be kept at a level such that it can be handled by the available number of investigators and fraud analysts.
- Fraudulent cases missed by the fraud detection system are reported to the issuing company when the cardholder identifies that their account has been compromised. This can take up to several months, resulting in a delay in correctly labelling each case. Some fraudulent cases remain unidentified and therefore mislabelled. Thus, a fraud detection model is almost certainly trained on noisy data.

In order to discourage fraud and to decrease the losses suffered due to fraud, the industry and their member banks employ various technologies to detect and prevent plastic card fraud. Some of the

E-mail address: mk211@le.ac.uk

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.02.119

preventive measures on the cards are consistency checks based on chip and pin, 3-D Secure for online transactions, card reader security and security questions for internet banking, etc. Fraud detection comes into play once prevention has failed and aims to stop the abuse in progress as quickly as possible after it's first occurrence.

Fraud detection systems can be based on various approaches (Bolton & Hand, 2002; Fawcett & Provost, 2002; Phua, Lee, Smith, & Gayler, 2005). The emphasis on fraud detection methodology is usually put upon supervised classification at transaction level that constructs an assignment procedure for new cases from the given training samples of fraudulent and non-fraudulent transactions (Maes et al., 2002; Brause, Langsdorf, & Hepp, 1999). An example of such a system in the banking industry is a rule-based system that consist of rules of the form: If {*assertions*}, Then {*consequence*}. Typically, the in-use set of rules combines the results of a non-statistical expert analysis by a fraud team, findings of investigators, and rules derived from a tree-based algorithm. The strategy is to monitor individual transactions and combinations of the short-term history of transactions. This approach is proven to reliably detect patterns of fraudulent activity which have previously been observed. To extract a rule with confidence there should be an adequate number of cases perpetrated in the same fashion. Time is required to collect the cases, extract an appropriate rule and put it into operation. By the time this circle is complete fraudsters may have changed their tactics. Fraud is an organized criminal enterprise which evolves over time; furthermore there are over 20 main plastic card issuers in the UK and there is no centralized system that collects all identified fraudulent cases. This dissemination of information may prevent a clear and accurate understanding of the incidence of plastic card fraud.

In contrast to the supervised approach, fraud detection systems based on an unsupervised methodology monitor account activity and flag transactions inconsistent with an account's usual behaviour observed over a period of time (Bolton & Hand, 2001; Juszczak, Adams, Hand, Whitrow, & Weston, 2008). Some banks deploy the unsupervised methodology in the form of so-called "behavioural models" which build an individual profile for each account. This includes characteristics of account typical transaction activity, such as merchant types, time of day, monetary values, geographic locations, etc. With a vast number of variations of behaviour and even larger number of opportunities of adapting new patterns it is difficult to cover all possible scenario of legitimate transaction activity. Very often an unusual transaction is in fact legitimate. For instance, purchase of airline tickets or transfer of a lump sum to credit card can be an event which has not previously occurred or has been observed with different parameters. As a result, alerts created by the system in many situations are for incorrectly implicated legitimate cases.

In this paper, we approach the fraud detection problem with a hybrid model that incorporates one-class classification and rule-based approaches at account level. This model has arisen gradually over the implementation stage of our collaborated work with the bank and tackles the issues which supervised and unsupervised approaches may lack individually, by their combination. Given that the use of "behavioural models" might be accompanied by a high number of incorrectly alerted cases whereas the use of rule-based system might result in a poor performance, for instance, when fraud tactics had changed, the logical extension is to apply a combined methodology.

In the data pre-processing step we adapt methodology proposed in (Whitrow, Hand, Juszczak, Weston, & Adams, 2008) for transaction aggregation over a period of time. Since the transaction aggregation yields a smoothed data representation, it is expected to result in a more consistent and stable model than a system built at transaction level. This framework along with a data-customised methodology is deployed in order to build a model of aggregated spending behaviour of an account in a time window. The data-customised methodology is based on separating accounts into several behavioural groups. A model is fitted to each group of accounts, rather than handling each account individually. This results in a reduction of the number of parameters while not adversely affecting the matching of account behaviour to models.

The proposed fraud detection system operates on two levels. At the first level the system monitors any deviation from the account model of aggregated spending behaviour in the time window and assigns a score according to the level of suspicion of fraud. The aggregated sequence of transactions scored above a prescribed threshold is passed on for further refinement to the second level of hybrid model – rule-based filters. A case that contravenes any of the rules is flagged as suspected to be fraudulent. The rules extracted from the transaction records are aimed to enhance the output of the system.

The paper is organized as follows. Section 2 presents the framework for aggregating transaction information at the data pre-processing step. The hybrid model is described in Section 3. Criteria for performance assessment of fraud detection systems are discussed in Section 4. Section 5 contains results of experiments with the real debit card transaction data and performance comparison of the implemented hybrid system with the rule-based fraud detection system of the collaborating bank. Concluding remarks are given in Section 6.

2. Debit card transaction data pre-processing

Let us assume that each transaction $x_i(t)$ of an account i at time t is an object described by a d-dimensional vector of features containing a set of real-valued measurements and categorical indicators such as account number (integer), transaction amount $m_i(t)$ (real non-negative number), transaction type (categorical indicator), etc. Transaction type is an important indicator that defines whether the transaction was conducted at an automatic teller machine (ATM) or at a point-of-sale (POS) terminal. The latter category is subdivided into the point-of-sale type when card is present (POS (CP)) and the point-of-sale type when card is not present (POS (CNP)), i.e. when the transaction is made through Internet, mail or telephone orders.

Consider a time period $[t_1, T]$. For an account i, suppose that there have been n_i number of transactions over the time period $[t_1, T]$ and denote the time-ordered series of transactions as $x_i(t_1), x_i(t_2), \ldots, x_i(t_{n_i})$ such that jth transaction occurred at time t_j. Then we introduce the time-ordered series of transactions over a time window, Δt, of k-day length as

$$X_i(t) = \{x_i(t_j),\ \textit{where } t_j : t - \Delta t \leqslant t_j \leqslant t\} \textit{ for any } t \in [t_1 + \Delta t, T].$$

The transformation from the transaction level to the account level requires an account level summary of the transaction data, i.e. $Y_i(t) = \Phi(X_i(t))$, where Φ is a pre-processing transformation.

To this end, we introduce the notation. Consider the set of transaction types $\mathcal{S} = \{POS(CNP), ATM, POS(CP)\}$. For a transaction $x_i(t_j)$, introduce a three-dimensional column vector $z_{ij} = \left(z_{ij}^1, z_{ij}^2, z_{ij}^3\right)^T$ such that

$$z_{ij}^1 = \begin{cases} 1, & \text{if the transaction } x_i(t) \text{ is of type POS(CNP)}, \\ 0, & \text{otherwise}, \end{cases}$$

and z_{ij}^2 and z_{ij}^3 are defined analogously for the types *ATM* and *POS*(*CP*), respectively.

We choose the pre-processing transformation to be the total value and count of particular type of transactions in the time window Δt. Then the account summary of transaction data over the time window Δt of k-day width is,

$$Y_i(t) = \Phi(X_i(t)) = \left(\sum_{j=1}^{l_i} z_{ij}, \sum_{j=1}^{l_i} m_i(t_j) z_{ij} \right),$$

where l_i is the number of transactions falling into the time window Δt.

The data pre-processing can be summarised as follows; the dimensionality of original feature space is reduced by selection of specific features and then is mapped by the transformation to another feature space, where the object $Y_i(t)$ is a 3 by 2 matrix. Each row of $Y_i(t)$ is a real-valued vector of features, namely the count and total value in the time window of the corresponding type of transactions from $\mathcal{S}$.

The aggregation period is chosen to be a rolling window that ends at the time of the current transaction and starts k-days ago. This means that for each transaction a time window is constructed so that it includes the transaction itself and all other transactions conducted over last n-days. As a new transaction is made, the window shifts so that it starts at the time of the latest transaction and goes back in time up to the window width. An account summary of transaction data is thus created with every new transaction made, hence the account level fraud detection system gives an immediate response.

To illustrate a relative superiority of this approach over a methodology which creates a new account summary every fixed period of time, let us consider, for example, a partition of a time interval into non-overlapping windows of three day width. As account summaries are updated only every three days, the fraud detection system is incapable of detecting suspicious activity at a higher time resolution than this. Timeliness is crucial to the success of the fraud detection system, consequently information about the status of each account has to be continually updated as new transactions occur.

3. Methodology

One of the main dilemmas when designing fraud detection system is what statistics or data mining approach to use. In particular, for one-class classification methods a vast number of tools can be applied ranging from density based classifiers such as Gaussian density (Bishop, 1995) and Parzen window models (Parzen, 1962) through to the domain-based classifiers including modern developments such as support vector data description (Tax & Duin, 1999) and minimum spanning tree data description (Juszczak, 2006).

It should be remarked that the recent methodologies for fraud detection proposed in academic publications are sophisticated and typically require special expertise in data mining techniques. As a result, the cost of implementation of these methods may exceed the resources and needs which are currently available to retail banks. Furthermore, as it was emphasized in (Hand, 2006), the alleged superiority of these high level methods must be measured relative to the costs of implementing them as compared with simpler systems. A practical application has to fulfill the designated task taking into account business related issues. In our case, the goal is to detect fraudulent activity on a debit card with low costs attached such as time and effort assigned for bank practitioner training to achieve a required level of expertise in classification methods, computational power required to put the system into practice, etc. We use a relatively simple strategy for one-class classification, with the enhancement that the data description boundary is modified according to the account spending behaviour type. To improve the outcome of the model, we introduce the post-processing operation, where flagged accounts are passed through the rule-based filters. This identifies the most likely compromised amongst them. This section explains our methodology of the hybrid model in more detail.

The pre-processing of the data for the hybrid model is performed as outlined in Section 2. We limit our further analysis here to the POS(CNP) class of transactions; the other two types, POS(CP) and ATM, can be approached similarly.

In the data pre-processing step for each account the following values are calculated:

- Counts of the number of CNP transactions falling into the rolling time window. Following the previous notation, it is given by $\sum_{j=1}^{l_i} z_{ij}^1$. For convenience, for an account i over the time period $[t-\Delta t, t]$ let us now denote this sum as $count_i(t)$.
- The total value of CNP transactions falling into the rolling time window. Following the previous notation, it is given by $\sum_{j=1}^{l_i} m_i(t_j) z_{ij}^1$. For convenience, for an account i over the time period $[t-\Delta t, t]$ let us now denote this sum as $amount_i(t)$.

The model of account aggregated spending behaviour is constructed on the data set containing only legitimate transactions, whereas a test set consist of both fraudulent and legitimate transactions.

The model of an account i consists of a set of descriptors that quantify the time-ordered series:

$$\{count_i(t_1), \ldots, count_i(t_{S_i})\}$$

and

$$\{amount_i(t_1), \ldots, amount_i(t_{S_i})\},$$

where S_i is the number of transactions of an account i in the data set that consist of only legitimate transactions.

The numerical descriptors selected contain the information necessary to answer several specific questions. Namely,

- *How much on average does an account spend in total in the time window?*

  Denote this descriptor for an account i as $amount\_on\_average_i$:

  $$amount\_on\_average_i = \frac{1}{S_i} \sum_{j=1}^{S_i} amount_i(t_j);$$

- *How far on average does an account total spend in the time window deviates from it's amount_on_average?*

  Denote this descriptor for an account i as $amount\_spread_i$:

  $$amount\_spread_i = \sqrt{\frac{1}{S_i - 1} \sum_{j=1}^{S_i} \left(amount_i(t_j) - amount\_on\_average_i\right)^2};$$

- *How many transactions on average does an account tend to make in the time window?*

  Denote this descriptor for an account i as $count\_on\_average_i$:

  $$count\_on\_average_i = \frac{1}{S_i} \sum_{j=1}^{S_i} count_i(t_j);$$

- *How far on average does the number of transactions of an account made in the time window deviates from it's count_on_average?*

  Denote this descriptor for an account i as $count\_spread_i$:

  $$count\_spread_i = \sqrt{\frac{1}{S_i - 1} \sum_{j=1}^{S_i} \left(count_i(t_j) - count\_on\_average_i\right)^2}.$$

As a result, the model of the aggregated spending behaviour of an account i in the time window is specified by the following descriptors: $amount\_on\_average_i$, $amount\_spread_i$, $count\_on\_average_i$, $count\_spread_i$.

Fig. 1 visualizes these considerations in the plane count-amount for four accounts. The stars have the coordinates $(count(t_j), amount(t_j))$, $j = 1,\ldots,S_i$, $i = 1,\ldots,4$. The dashed line represents the data description boundary of the model of the aggregated spending behaviour of an account i in the 3 day time window and has vertices with the coordinates $(0,0)$, $(0, amount\_spread_i)$, $(count\_spread_i, amount\_spread_i)$ and $(count\_spread_i, 0)$.

In the formulation of a one-class classification problem the objects that are outside of the data description domain should be marked as outliers (Tax, 2001), or in terms of the hybrid model considered here the aggregated sequence of transactions outside of the rectangle is an indicator that account has been compromised. However, this approach would obviously lead to a huge number of legitimate cases mislabelled, resulting in a poor performance. To enhance the system outcome, the data description domain should be extended. This is accomplished through separating all accounts into 10 groups based on a set of different characteristics and for each of them a different procedure for the data description boundary transformation is applied. For the first 7 groups the indicators that guide the boundary modification are the magnitudes of account's *amount_on_average* and *amount_spread*. For instance, Group 2 has the following characteristics.

An account with $amount\_on\_average \leqslant 100$ and $50 < amount\_spread \leqslant 100$ has the coordinates of the upper boundary modified to $(0, 5\ amount\_spread)$ and $(count\_spread, 5\ amount\_spread)$. This means that the account boundary of the aggregated legitimate spend in the time window is equal to 5 *amount_spread*, let us denote it as *boundary_amount*. In other words, if this account spend exceeds its usual spend in the time window by more than five of *amount_spread* then it is marked as "suspected to be compromised".

Groups 8 and 9 contains accounts which frequently use their debit cards for CNP type of transactions. Group 9 is the compliment of Group 8 in this respect. This is motivated by the observation that accounts with a high transaction frequency more often exceed their average rate of transactions in the time window than accounts with a moderate transaction frequency. For Groups 8 and 9 the right boundary of the data description domain is extended according to their group parameters. Let us denote the extended boundary of *count_spread* as *boundary_count*. Finally, accounts with frequent gambling or gaming transactions form Group 10. Spending behaviour of these accounts often significantly deviates from their previously observed patterns. For Group 10 the upper and right boundary of the data description domain is extended.

Note that this structure of groups and, in particular, their parameters were guided by exploratory analysis of the collaborating bank data and may differ for another application. However, the idea of an approach tuned to the cardholder behaviour generally seems to suit well the heterogeneous nature of transaction data.

Fig. 2 visualizes the foregoing modifications of the data description boundary. The dashed line represents the original data description boundary, the solid line gives the modified data description boundary.

Finally, after modification the data description domain that represents the model of the aggregated spending behaviour of an account i in the time window is completely specified by the following descriptors: $amount\_on\_average_i$, $boundary\_amount_i$, $count\_on\_average_i$, $boundary\_count_i$.

Once the model is constructed we want to assess the "status" of each account with every new transaction made. The hybrid model produces a suspiciousness score, a real number between 0 and 1 associated with each account which is updated as a new transaction occurs. This is then compared with a threshold ϑ in order to assign account "status" to one of two classes: "suspected to be compromised" and "assumed to be legitimate". The threshold is set up during model training such that it delivers user-specified values of performance measures.

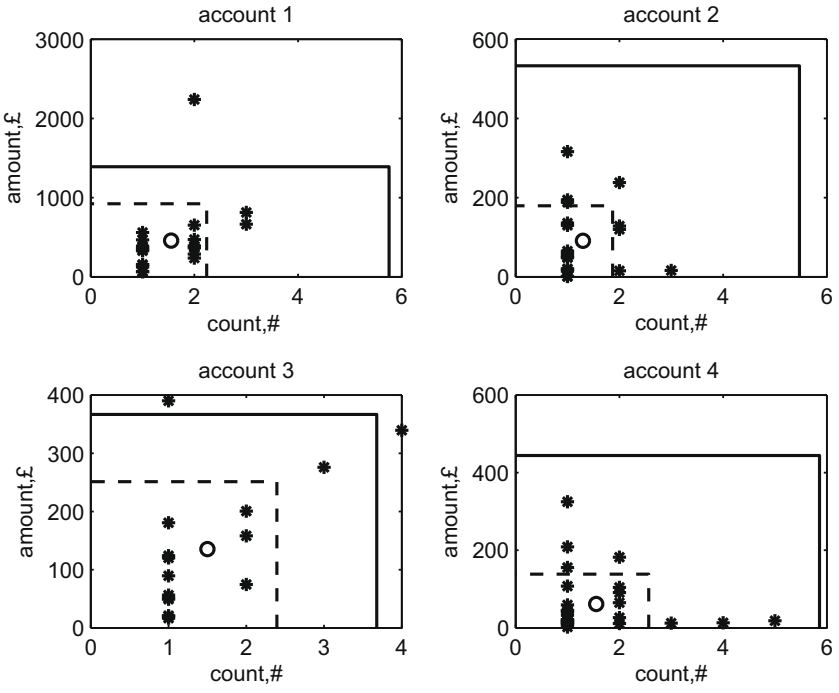

Fig. 1. The data description boundary (dashed line) of four accounts' models of aggregated spending behaviour in the 3 day time window. The set of the aggregated transactions are stars. The round point has coordinates $(count\_on\_average_i, amount\_on\_average_i)$.

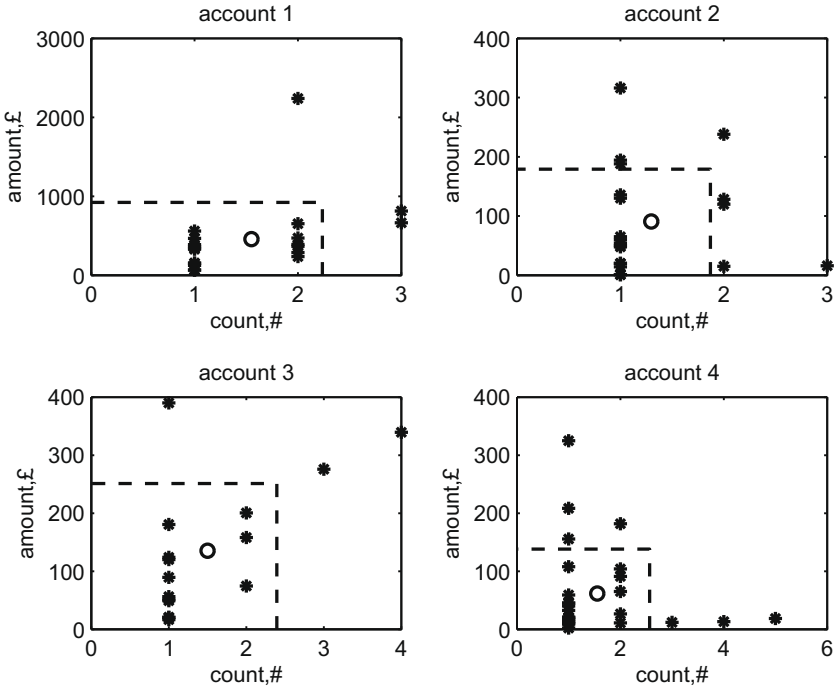

Fig. 2. The data description boundaries of four accounts' models of aggregated spending behaviour in the 3 day time window before and after their modification (dashed and solid line respectively).

Consider an account i that makes a new transaction at time t_{new}. For model training, t_{new} is the time of transaction from the test set whereas for operational stage t_{new} is the real time of current transaction. The time window shifts such that it ends on the transaction and covers the last k-day period. The total amount spent, $amount_i(t_{new})$ and the number of transactions made, $count_i(t_{new})$ in the period covered by the time window is calculated. We wish to determine whether an account i at time t_{new} might be compromised. That is we wish to evaluate account "status" and assigns it a suspiciousness score. To this end, the standardized differences between $amount_i(t_{new})$ and $amount\_on\_average_i$, and between $count_i(t_{new})$ and $count\_on\_average_i$ are calculated, i.e.

$$\frac{amount_i(t_{new}) - amount\_on\_average_i}{boundary\_amount_i}$$

and

$$\frac{count_i(t_{new}) - count\_on\_average_i}{boundary\_count_i}.$$

In order to produce the suspiciousness score the logistic transform is applied:

$$score_{amount} = \frac{1}{1 + \exp\left(-\frac{|amount_i(t_{new}) - amount\_on\_average_i|}{boundary\_amount_i}\right)}$$

and

$$score_{count} = \frac{1}{1 + \exp\left(-\frac{|count_i(t_{new}) - count\_on\_average_i|}{boundary\_count_i}\right)}.$$

We combine both scores using a product rule, such that the total score is the product of score for the amount and count, i.e.

$$score = score_{amount} \times score_{count}.$$

Thus, at the end of the level one of the hybrid model we have the suspiciousness score for the account. If the score is below the prescribed threshold then the account is assumed to be legitimate otherwise it is passed to the level two.

At level two the account is processed through a set of rule-based filters to increase the confidence that it has indeed been compromised. The filters are designed to reflect the fact that not all sudden changes in behaviour are actually due to fraud. For instance, a legitimate event of purchase of airline tickets or transfer to credit card can be mislabelled by the fraud detection system as it is typically not an every day event and can be of a large amount. Since the first level models the account aggregated spending behaviour using measures of central tendency, it is clear that the scores for such rare, lump sum events are in the class "suspected to be compromised". To avoid false alerts in such legitimate situations a set of rule-based filters is embedded into the level two of the hybrid model. In particular, one of the filters concentrates on airline merchants: if an account receives a high suspiciousness score then it is checked on how many airline transactions have been made in the time window; if it is greater than a specified constant, the account retains "suspected to be compromised" status. Otherwise, the account passes through the system marked as "assumed to be legitimate" or, alternatively, redirected for some special investigation by a fraud analyst.

Finally, alerted accounts are ranked according to their suspiciousness score in descending order and this forms the output of the hybrid model.

4. Performance criteria

The stream of accounts flagged by the fraud detection system includes the compromised accounts, true positives (TP), and incorrectly implicated cases, false positives (FP). Complementarily, accounts that pass through the fraud detection system with no alert created is a mixture of legitimate accounts, true negatives (TN) and missed fraudulent cases, false negatives (FN).

Clearly, we wish to construct a fraud detection system that satisfies at least the following requirements:

- minimum number of accounts that are false positive and false negative
- maximum number of accounts that are true positive and true negative

Various performance criteria can be used in an application to the fraud detection problem. These include the measures such as misclassification rates, the area under receiver-operating curve (ROC) and more recently proposed criteria such as the area under the modified ROC curve called (the performance curve) (Hand, Whitrow, Adams, Juszczak, & Weston, 2008). Generally, an appropriate performance measure has to take into account a number of important issues related to the fraud detection problem including timeliness of fraud alert raised and different costs associated to different types of errors, namely FN and FP. It is typically considered that the error committed in assessing a fraudulent case as legitimate (FN) is more serious than the complementary type of error (FP).

In this work we use four performance measures. The first of these is the proportion of legitimate accounts mislabelled as fraudulent to fraudulent accounts identified correctly, namely FP:TP. This measure shows how many alerted accounts should be investigated to find one fraudulent. This measure is commonly used in practise, for example by the collaborating bank. Rapid fraud detection after its occurrence is an important requirement that a fraud detection system should satisfy. Thus, the number of fraudulent transactions that are missed by the fraud detection system before an alarm has been risen should be kept at minimum. We quantify performance of fraud detection system in this respect by timeliness ratio. It is proportion of fraudulent transactions which have escaped detection to all fraud transactions which have occurred on account. The other two measures we use are the percentage of fraud identified correctly on all accounts compromised and the savings assigned to the identified fraud.

5. Experiments

This section presents the experimental results of the hybrid model achieved on the debit card transaction data provided by the collaborating bank. We also outline the results of comparison of the hybrid model with the in-use rule-based fraud detection system of the collaborating bank.

The available data set consisted of about 189 million transactions occurred in the period between 01/10/07 and 30/04/08 and each transaction contained 76 fields of coded information. The data required substantial pre-processing to extract the relevant information and to transform it into a suitable representation for the classifier. In particular, the accounts with less than 5 transactions in total were removed from the data set.

We sub-sample two data sets of a manageable size from the whole available data set. These data sets, A and B, were split into two subsets. The transactions occurred up to 31 March 2008 (subsets A_1 and B_1) are used to construct the model and the transactions dated 1 April 2008 onwards (subsets A_2 and B_2) used to evaluate the model performance. The data sets A and B have 11,555 accounts, of which 1555 experienced fraud at POS (CNP) terminals in the period between 01/04/08 and 30/04/08 and amounts to all compromised accounts in April through the Internet, telephone and mail orders. The compromised accounts (1555 accounts) are the same for both data sets whereas the legitimate accounts (10,000 accounts) are different. One of the data sets is used for training and another one for validation procedures. Each of them contains only POS(CNP) transactions.

We retain all fraudulent transactions in the data sets in order to quantify the system's potential ability for fraud detection. The sets of legitimate accounts are of a relatively small size in order that a larger number of experiments is tractable. The fraud rates in the training and validation data sets do not correspond to the rates of the whole data set. To establish the correct ratio of FP:TP, we multiply the number of FP achieved on the sub-sampled data sets in proportion to the size of the whole data set. We take the number of all accounts (618,712 accounts) that have POS (CNP) in the period between 01/04/08 and 30/04/08, and use the proportion of false positive accounts to the legitimate accounts in the training and validation sets in order to estimate the number of accounts that potentially can be incorrectly alerted by the fraud detection system.

The model of the aggregated spending behaviour in the time window is built for each individual account from the data set A_1 of legitimate transactions. The test data set A_2 of legitimate and fraudulent transactions is used to quantify the performance of the model. For model validation the data set B_1 of only legitimate transactions and B_2 of mixed transactions are used. The characteristics of the data sets A_1, A_2 and B_1, B_2 are presented in Table 1.

We build the fraud detection system based on three different widths of the rolling time window to analyze how it reflects the overall system performance. We use 1, 3 and 7 day aggregation periods. The rule-based fraud detection system of the collaborating bank operates at the transaction level, therefore the aggregation period is of zero width and contains a single transaction. The performances of the hybrid (HM) and rule-based (RM) models are compared based on the FP:TP, the total percentage of fraudulent accounts identified and the average timeliness ratio discussed in the previous section. The results are shown in Table 2.

From Table 2 we see that the rule-based model (RM) is superior in the number of fraudulent accounts identified whereas the hybrid model delivers better FP:TP rate. Clearly, in the hybrid model the threshold can be set such that it delivers a higher number of identified fraudulent accounts, this increase however would be inevitably accompanied by a deteriorated FP:TP rate. The parameters of the hybrid model are set such that it delivers the maximum possible number of compromised accounts while keeping the number of alerts manageable.

Table 1
Characteristics of data sets.

Data set	Time period	# of accounts, fraud/legitimate	# of transactions, fraud/legitimate
A_1	01/01/08-31/03/08	0/11,555	0/144,298
A_2	01/04/08-30/04/08	1555/10,000	5242/51,593
B_1	01/01/08-31/03/08	0/11,555	0/153,018
B_2	01/04/08-30/04/08	1555/10,000	5242/54,598

Table 2
Performance assessment of rule-based (RM) and hybrid (HM) models.

Model	Aggregation period, days	FP:TP, training/validation	Compromised accounts identified, %	Timeliness ratio
RM	0	10.15:1/16.46:1	29	N/A
HM	1	9.17:1/12.46:1	16.9	0.7091
HM	3	9.09:1/11.32:1	19.7	0.7265
HM	7	7.36:1/11.4:1	27.6	0.7432

Table 3
The sets of compromised accounts detected by the rule-based and hybrid models.

Aggregation period, days	Overlapping set of compromised accounts, #			Non-overlapping set, hybrid model, #
	HM quicker	RM quicker	The same	
1	9	36	18	200
3	8	35	21	242
7	9	35	21	364

Table 4
Potential monetary savings of the hybrid model.

Time window, days	Fraud identified by HM and missed by RM, %	Savings, £
1	12.8	92,845.00
3	15.6	92,380.00
7	23.4	92,376.20

We note that the rule-based system is of the supervised classification type of methods, in contrast the hybrid model is based on the unsupervised approach. It is interesting to take a closer look at the identified fraudulent accounts of both models. We are first interested in the overlapping set of compromised accounts identified by both models and in particular which of the models detects fraud more quickly on this set. The non-overlapping set of compromised accounts detected by the hybrid model will give us an indication of the model's ability to identify additional types of fraudulent behaviour that are not embedded into the rule-based model. The results are shown in Table 3.

Table 3 shows that the rule-based system detects fraud faster in the overlapping set in the most cases. However, the hybrid model mainly detects compromised accounts that are missed by the rule-based system. This indicates that the hybrid model should be implemented alongside the existing rule-based system, together they provide detection of up to 56.6% of all compromised cases. We can also conclude that the concept of the hybrid model conforms to the theoretical background of the unsupervised approach and detects novel types of fraudulent behaviour.

Table 4 shows the percentage of the new fraud cases identified on all accounts compromised that is detected by the hybrid model and missed by the rule-based model. We also provide the monetary values of the potential savings in British pounds attached to the new fraud cases detected. By savings we understand the value of fraudulent transactions that occur on an account after the hybrid model raises fraud alarm, as these would not have occurred if the alert raised by the hybrid model had been processed.

Table 4 shows that the model based on the 7 day time window identifies the largest percentage of fraud, however the total savings are in the same range for all three aggregation periods. This and timeliness ratio results indicate that the model with the larger aggregation periods are generally able to detect a greater percentage of fraud at the cost of worse timing in raising fraud alarm, this results in more missed fraudulent transactions. Currently the hybrid model is in operation with the 3 day aggregation, this meets the collaborating bank targets. For simultaneous realization of a greater percentage of identified fraud (achieved with larger time windows) and better timing (achieved with smaller time windows), the system can be run with several aggregation periods, if business issues allow.

6. Conclusions

In this paper, we have presented the framework for a hybrid model based on the combination of supervised and unsupervised techniques as a component of the plastic card fraud detection system. The methodology has arisen through our experience of collaborated work with the bank in developing an automated system that extends the bank's existing monitoring system. The relative simplicity of the chosen one-class classification approach is well suited to the large-scale and complex transaction data and has enabled us to build an intuitively appealing group structure of accounts that provides customised models for different types of cardholder behaviour. We have observed that this approach along with the post-processing level of the rule-based filters results in a significant advantage against the generic model. Indeed, the heterogeneous nature of transaction data suggests a flexible approach tuned to the cardholder behaviour is more appropriate than a generic technique applying the same procedure for all accounts/transactions.

We have compared performance of the collaborating bank's existing rule-based system and our hybrid model. Overall, the hybrid model was capable of identifying fraudulent activity in a timely manner resulting in substantial monetary savings. The experimental results show that the majority of the fraudulent cases identified with the use of hybrid technique are not detected by the bank's rule-based system and vise versa. This suggests that both systems should be run in parallel to achieve the maximum results.

Acknowledgment

The authors would like to thank all those who contributed to this paper, in particular to Michael Tretyakov and David Packwood at Leicester University, Department of Mathematics. Further to this, we gratefully acknowledge those in the Fraud Analysis Team at the collaborating bank for their time and input. Also, special thanks go to Claire Brunt and Debbie Smith.

References

APACS. (2008). The UK payments association. Fraud the facts 2008 <http://www.apacs.org.uk/resources_publications/documentsFraudtheFacts2008.pdf>.
Bishop, C. (1995). *Neural networks for pattern recognition*. Oxford University Press.
Bolton, R. J., & Hand, D. J. (2001). Unsupervised profiling methods for fraud detection. In *Proceedings of the conference on credit scoring and credit control*, (Vol. 7). Edinburgh.
Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science, 17*(3), 235–255.
Brause, R., Langsdorf, T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. In *Proceedings of the 11th IEEE international conference on tools with artificial intelligence* (pp. 103–106).
Fawcett, T., & Provost, F. (2002). Fraud detection. In W. Klosgen & J. Zytkow (Eds.), *Handbook of Knowledge Discovery and Data Mining* (pp. 726–731). Oxford University Press.
Hand, D. J. (2006). Classifier technology and the illusion of progress. *Statistical Science, 21*(1), 1–14.
Hand, D. J., Whitrow, C., Adams, N., Juszczak, P., & Weston, D. (2008). Performance criteria for plastic card fraud detection tools. *Journal of the European Operational Society, 58*, 956–962.
Juszczak, P. (2006). Learning to recognise. A study on one-class classification and active learning. Ph.D. thesis, Delft University of Technology.
Juszczak, P., Adams, N., Hand, D. J., Whitrow, C., & Weston, D. (2008). Off-the peg and bespoke classifiers for fraud detection. *Computational Statistics and Data Analysis, 52*(9), 4521–4532.
Maes S., Tuyls K., Vanschoenwinkel B., & Manderick B. (2002). Credit card fraud detection using Bayesian and neural networks. In *Proceedings of first international NAISO congress on neuro fuzzy technologies: NF2002* (pp. 16–19).
Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics, 33*(3), 1065–1076.
Phua, C., Lee, V., Smith, K., & Gayler, R. (2005). *A comprehensive survey of data mining based fraud detection research*. Clayton School of Information Technology, Monash University.
Tax, D. (2001). One-class classification. Ph.D. thesis, Delft University of Technology.
Tax, D., & Duin, R. (1999). Support vector domain description. *Pattern Recognition Letters, 20*(11-13), 1191–1199.
Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. (2008). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery, 18*(1), 30–55.