Hybrid model for plastic card fraud detection systems

M. Krivko
Department of Mathematics, University of Leicester

Women in Mathematics, Isaac Newton Institute, 15th April 2010
Plan of the talk

- Background (fraud definition, types, challenges)
- Fraud detection approaches
- Debit card transaction data pre-processing
- Hybrid model methodology
- Performance criteria
- Experimental results/Conclusions
Plastic card fraud

Fraud implies unauthorised and illegal use of the debit/credit facilities of a legitimate account.

Plastic card fraud losses in the UK in 2008 increased by 14% from 2007 to £ 609.9mn (APACS). These losses are absorbed by customers, merchants, lenders.

Methods of compromise:
- Theft of card details (Card Not Present)
- Counterfeit card
- Lost/stolen card
- Mail non-receipt
- 3rd party application fraud /Account takeover

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.
Challenges

- Large volume of transactions – millions
- Large number of variables – 76 per transaction
- Unbalanced class sizes – 0.1% transactions fraudulent
- The asynchronous and heterogeneous nature of transactions between and within accounts
- Fraudsters don’t give up! but change strategies
- Delay in learning class labels
- Mislabeled classes
Supervised classification

Constructs an assignment procedure for new cases from the given training samples of fraudulent and non-fraudulent transactions.

- Need examples of past fraudulent and legitimate activities.
- Highly effective at detecting known fraud types, however to extract a rule with confidence there should be an adequate number of cases perpetrated in the same fashion.
- Ineffective at novel types.

Example

A rule-based system that consist of rules of the form: If {a certain condition}, Then {a 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.
Unsupervised classification

Given a set of legitimate transactions, build a model for the “norm” for this customer and detect when it deviates

- Need examples of past legitimate activity
- Highly effective at detecting novel fraud types

The “norm” can be based on
- this customer compared with self at previous times
- this customer compared with other customers
- segmentation into customer types

Example

A “behavioural model” 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.
Data pre-processing

Each transaction $x_i(t)$ of an account $i$ at time $t$ is a $d$-dimensional vector of features. Consider a time period $[t_1, T]$. The time-ordered series of transactions over a time window, $\Delta t$, of $k$-day length

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

Features: when? how much? what?

- Time $t$, transaction time
- Amount $m_i(t_j)$, the amount in pounds of the transaction at time $t$
- Transaction type indicator for POS(CP), POS (CNP) or ATM
- Merchant code, categorical indicator for merchant type
Data pre-processing

For a transaction $x_i(t_j)$, introduce a three-dimensional column vector $z_{ij} = (z_{ij}^1, z_{ij}^2, z_{ij}^3)^T$ 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.

The account summary of transaction data over the time window $\Delta t$ 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 $\Phi$ is a pre-processing transform, chosen to be the count and total value of particular type of transactions in the time window.

We choose a rolling type window that ends on the transaction and goes back in time for $k$ days.
Data pre-processing

Training set, for each account only legitimate transactions.

Testing set, both fraudulent and legitimate transactions.

- Counts of the number of CNP transactions falling into the rolling time window is $\sum_{j=1}^{l_i} z_{ij}^1$. For convenience, for an account $i$ over the time period $[t - \Delta t, t]$ denote this sum as $\text{count}_i(t)$.

- The total value of CNP transactions falling into the rolling time window is $\sum_{j=1}^{l_i} m_i(t_j)z_{ij}^1$. For convenience, for an account $i$ over the time period $[t - \Delta t, t]$ denote this sum as $\text{amount}_i(t)$. 

The numerical descriptors selected are

\[
\text{amount\_on\_average}_i = \frac{1}{S_i} \sum_{j=1}^{S_i} \text{amount}_i(t_j);
\]

\[
\text{amount\_spread}_i = \sqrt{\frac{1}{S_i - 1} \sum_{j=1}^{S_i} \left( \text{amount}_i(t_j) - \text{amount\_on\_average}_i \right)^2};
\]

\[
\text{count\_on\_average}_i = \frac{1}{S_i} \sum_{j=1}^{S_i} \text{count}_i(t_j);
\]

\[
\text{count\_spread}_i = \sqrt{\frac{1}{S_i - 1} \sum_{j=1}^{S_i} \left( \text{count}_i(t_j) - \text{count\_on\_average}_i \right)^2}.
\]

The model of aggregated spending behaviour of an account \(i\) in the time window is: \(\text{amount\_on\_average}_i, \text{amount\_spread}_i, \text{count\_on\_average}_i, \text{count\_spread}_i\).
Figure: The data description boundary (dashed line) of four accounts’ models of aggregated spending behaviour in the 3 day time window.
Methodology

All accounts are separated into 10 groups based on a set of different characteristics, such as

- **Group 1-7**: magnitudes of account’s `amount_on_average` and `amount_spread`.
- **Group 8**: high frequency of CNP type of transactions. Group 9 is the complement of Group 8 in this respect.
- **Group 10**: high frequency of gambling or gaming transactions.

**Example**

Group 2 has the following characteristics: `amount_on_average \leq 100` and `50 < amount_spread \leq 100`. The upper boundary modified to `(0, 5 amount_spread)` and `(count_spread, 5 amount_spread)`.

After modification the data description domain that represents the model of aggregated spending behaviour of an account $i$ in the time window is: `amount_on_average_i`, `boundary_amount_i`, `count_on_average_i`, `boundary_count_i`.
Figure: 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).
Methodology

An account $i$ makes a new transaction at time $t_{new}$. The total amount spent, $\text{amount}_i(t_{new})$ and the number of transactions made, $\text{count}_i(t_{new})$ in the period covered by the time window is calculated.

We wish to evaluate account “status” and assigns it a suspiciousness score.

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

and

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

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

**Example**

\[
\text{If } \text{score} > \text{threshold and } (\#\text{of airline transactions} > 2 \text{ or } \#\text{of airline transactions} = 0) \text{ then retain else discard.}
\]

A similar set of rules is developed for financial services and travel agencies.

Finally, accounts marked as “suspected fraudulent” are ranked according to their score in descending order and this list is the output of the model.
Performance criteria

<table>
<thead>
<tr>
<th>Predicted class</th>
<th>True class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fraud</td>
</tr>
<tr>
<td>Fraud</td>
<td>TP</td>
</tr>
<tr>
<td>Legitimate</td>
<td>TN</td>
</tr>
</tbody>
</table>

We use four performance measures:

- **FP:** TP - Proportion of legitimate accounts mislabelled as fraudulent to fraudulent accounts identified correctly
- **TP/(TP+FP)** - Percentage of fraud identified correctly
- **Timeliness ratio** - Proportion of fraudulent transactions escaped detection to all fraud transactions occurred on account
- **Savings assigned to identified fraud**
Results

Data set available - 189mn transactions between 01/01/08 and 30/04/08 with 76 fields per transaction.

Table: Characteristics of data sets.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Time period</th>
<th># of accounts, fraud/legitimate</th>
<th># of transactions, fraud/legitimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>01/01/08-31/03/08</td>
<td>0/11,555</td>
<td>0/144,298</td>
</tr>
<tr>
<td>A2</td>
<td>01/04/08-30/04/08</td>
<td>1,555/10,000</td>
<td>5,242/51,593</td>
</tr>
<tr>
<td>B1</td>
<td>01/01/08-31/03/08</td>
<td>0/11,555</td>
<td>0/153,018</td>
</tr>
<tr>
<td>B2</td>
<td>01/04/08-30/04/08</td>
<td>1,555/10,000</td>
<td>5,242/54,598</td>
</tr>
</tbody>
</table>
Table: Performance assessment of rule-based (RM) and hybrid (HM) models.

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

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

<table>
<thead>
<tr>
<th>Aggregation period, days</th>
<th>Overlapping set of compromised accounts, #</th>
<th>Non-overlapping set, hybrid model, #</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HM quicker</td>
<td>RM quicker</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>35</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
<td>35</td>
</tr>
</tbody>
</table>

**Table:** Potential monetary savings of the hybrid model.

<table>
<thead>
<tr>
<th>Time window, days</th>
<th>Fraud identified by HM and missed by RM, %</th>
<th>Savings, £</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12.8</td>
<td>92,845.00</td>
</tr>
<tr>
<td>3</td>
<td>15.6</td>
<td>92,380.00</td>
</tr>
<tr>
<td>7</td>
<td>23.4</td>
<td>92,376.20</td>
</tr>
</tbody>
</table>
Conclusions

- 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 vice versa. The hybrid and the rule-based systems should be run in parallel to achieve detection of 56.6% of all compromised cases.

- The model can be extended with the different features. For example, time between transactions.

- The hybrid model with 3 day aggregation period is in operation at the collaborating bank since February, 2009. The results delivered by the model meet banks targets.