Computational Statistics and Data Analysis 52 (2008) 4521–4532

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Off-the-peg and bespoke classifiers for fraud detection

Piotr Juszczak [a], Niall M. Adams [b], David J. Hand [a,b,*], Christopher Whitrow [a], David J. Weston [a]

[a] *Institute for Mathematical Sciences, Imperial College London, United Kingdom*
[b] *Department of Mathematics, Imperial College London, SW7 2AZ London, United Kingdom*

ARTICLE INFO

Article history:
Received 11 June 2007
Received in revised form 8 March 2008
Accepted 11 March 2008
Available online 15 March 2008

ABSTRACT

Detecting fraudulent plastic card transactions is an important and challenging problem. The challenges arise from a number of factors including the sheer volume of transactions financial institutions have to process, the asynchronous and heterogeneous nature of transactions, and the adaptive behaviour of fraudsters. In this fraud detection problem the performance of a supervised two-class classification approach is compared with performance of an unsupervised one-class classification approach. Attention is focussed primarily on one-class classification approaches. Useful representations of transaction records, and ways of combining different one-class classifiers are described. Assessment of performance for such problems is complicated by the need for timely decision making. Performance assessment measures are discussed, and the performance of a number of one- and two-class classification methods is assessed using two large, real world personal banking data sets.

1. Introduction

Retail banks have successfully deployed plastic cards to provide a broad range of products and banking opportunities to consumers. This provision has been accompanied by the very serious problem of plastic card fraud. Our interest in this paper is detecting fraudulent transactions. Loosely, fraud implies unauthorised and illegal use of the credit facilities of a legitimate account. It is estimated that losses attributed to such fraud in the UK in 2004 amounted to £505 million. The most recent figures suggest a slight decrease for the first half of the year 2006, down by 5% from previous years, to £209 million (APACS, 2006). It is speculated that this drop can be attributed to the introduction of PIN authentication by the chip and PIN scheme, launched on 14th February 2006. Responsibility for the financial burden of fraud is absorbed by lenders, merchants and legitimate customers. Lenders and merchants expend significant resource to secure their systems and procedures in an attempt to limit their liability for such costs.

Tackling fraud in the context of plastic card finance is a daunting problem. The effort can be divided into *fraud prevention*, that attempts to block fraudulent transactions at source, and *fraud detection*, where successful fraud transactions are subsequently identified. For prevention purposes, financial institutions challenge all transactions with rule based filters and methods based on neural networks; e.g. FALCON. For fraud detection it is obviously desirable to detect fraud as rapidly as possible. In both cases, prevention and detection, the problem is magnified by a number of domain constraints and characteristics. First, care must be taken not to prevent, or incorrectly implicate, too many legitimate transactions. Customer irritation is to be avoided. Second, most banks process vast numbers of transactions, of which only a small fraction is fraudulent, often less than 0.1%.

* Corresponding author at: Department of Mathematics, Imperial College London, SW7 2AZ London, United Kingdom. Tel.: +44 207 594 8521; fax: +44 207 594 8547.
E-mail address: d.j.hand@imperial.ac.uk (D.J. Hand).

0167-9473/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2008.03.014

Many approaches to fraud problems have been considered. Fawcett and Provost (2002) and Kou et al. (2004), provide a general discussion. Statistical views are explored by Bolton and Hand (2002), while data mining perspectives are discussed by Phua et al. (submitted for publication). In the context of plastic card fraud, various authors (e.g. Brause et al. (1999) and Maes et al. (2002)) have approached fraud detection as a classification problem. To use such approaches, a number of problems have to be solved. First, extensive processing of the irregularly timed transaction sequences is required, to convert the data into a representation suitable for classification algorithms. Furthermore, fraudsters change tactics – supervised approaches may only find existing tactics. The marked heterogeneity of transaction behaviour within and between accounts, along with the highly imbalanced classes, might indicate that supervised classification is not the most natural or appropriate tool for this problem.

In this paper, we consider plastic card fraud detection approaches based on a one-class classification (e.g., Tax (2001) and Juszczak (2006)). The idea is to monitor each account separately and using suitable descriptors, attempt to identify and flag transactions that are abnormal. Abnormality will be defined in comparison to a model related to the estimated probability density of the account's legitimate transaction descriptors. We propose a two stage process. In the estimation stage, we obtain a model of the distribution of the legitimate class. In the subsequent operational phase, transactions are designated legitimate or abnormal. We see the estimation phase as estimating the distribution of the normal class, followed by assignment to this class or some "other" class. One application of one-class classifiers is outlier detection (Barnett and Lewis, 1994; Hodge and Austin, 2004; Ferdousi and Maeda, 2006). Important steps in the approach include judicious selection of descriptors, estimation of the distribution of the normal class, and the specification of an *alert threshold* for the contours of estimated probability, such that any transaction lying outside distribution is regarded as abnormal.

A particularly attractive feature of one-class classification methods is that they have the capacity to respond to new types of fraud, since no explicit model is constructed for fraudulent behaviour — the models are based on legitimate behaviour only. One potential problem with this approach is that not all fraudulent transactions are abnormalities. The model of normal behaviour will in fact be based on a mixture of legitimate transactions and fraudulent transactions that appear legitimate. However, since the prevalence of fraudulent transactions is generally very low, we expect this to have negligible impact. A second, and complementary potential problem is that not all abnormalities will be fraudulent transactions. The proportion of flagged abnormalities which are in fact legitimate will be the false positive rate, and this will be a component of the performance measure.

2. Credit card transaction data

Credit card transaction data is rich and highly structured. The fundamental entity is an *account*, identified by a unique number. Customers are associated with accounts, but complications can arise, since for example, more than one card, and more than one person, can be associated with the same account. Each account is associated with a collection of static information (personal details and so on), and a sequence of transactions. The focus of this paper is the transaction sequence. For account number i, denote the transaction sequence as $X_i = \{\mathbf{x}_t | \mathbf{x}_t \in \mathbb{R}^N, \ t = 1, \ldots, n_i\}$ where $\mathbf{x}_t$ refers to an N-dimensional vector extracted from the tth transaction record in the sequence.

An archetypal transaction record is a rich data structure. Our commercial collaborators have provided transaction records with as many as 77 fields. These refer to a diverse set of information, including fields as seemingly mundane as card reader response status codes. In fact, such fields can be used to identify important details of the transaction precisely. The `service ID` is an important field that indicates transaction type, determining whether the transaction was conducted at an automatic teller machine (`ATM`) or at a point-of-sale (`POS`) terminal. As mentioned in Section 1, the introduction of plastic cards with a chip requiring `PIN` authentication in UK appears to have had a positive impact at `POS`. Thus, the `service ID` indicator partitions the transaction record and provides a fundamental distinction. Note that it is possible to complete a number of transaction types at an `ATM` in addition to cash withdrawals. For example, topping up mobile phones is a recent innovation.

Other important information included in a transaction record includes;

- Merchant code; categorical, up to 2000 levels. Larger merchants often have a single code while smaller merchants are grouped together. Some unusual examples include "carpentry contractors" and "massage parlours".
- Time and date of transaction.
- Amount and currency of transaction.
- Category of transaction; for example payment, refund and so on.
- In addition, a large set of reader and card response codes are recorded.

The systems and infrastructure that provide almost instantaneous processing of credit card transactions are complicated and we make no attempt to describe them here. It is sufficient to note that, during processing, all transactions face a number of fraud prevention filters. These filters often include simple and rapidly tested rules, such as consistency checks. On occasion, a filter will block a transaction, and perhaps even possibly resulting in the suspension of the card.

The vast majority of transactions make it past the fraud prevention filters, and later action may be required. Fraudulent transactions are often discovered when a customer challenges one or more transactions on their account statement. An investigation follows to determine whether the transactions were fraudulent. Once fraud is confirmed, the account is labelled as such, and usually closed. In this sense we can talk about a fraudulent account, though not necessarily a fraudulent customer. On completion of this process the bank is able to label some transactions as fraudulent. The nature of the process,