Mining Information from Plastic Card Transaction Streams

D.K. Tasoulis[1], N.M. Adams[2], D.J. Weston[1], and D.J. Hand[1,2]

[1] Institute for Mathematical Sciences, Imperial College London
[2] Department of Mathematics, Imperial College London South Kensington Campus, SW7 2AZ, UK
{*d.tasoulis,d.weston,n.adams,d.j.hand*}*@imperial.ac.uk*

Abstract. Detecting fraudulent plastic card transactions is an important problem in retail financial services. The problem is challenging because the data are streaming, heterogeneous, dynamic, and there is a large imbalance between fraudulent and legitimate transaction classes. Due to the complexity of the problem, a variety of knowledge discovery approaches can be fruitfully deployed. However, existing information systems' infrastructure constrains the information available to analyse each transaction. To provide a richer representation we therefore consider extending the practitioner's toolkit to extract feature information using recently proposed stream clustering algorithms. Such algorithms have to account for the temporal structure of the data. In this paper, we explore the utility of on-line density-based stream clustering methods for plastic card transaction fraud detection. Experiments with real data suggest that such methods have merit for both fraud detection and for and for revealing aspects of temporal structure of the transaction stream. This suggests that new features can be discovered to enhance existing detection algorithms.

Keywords: Fraud Detection, Data clustering, Data Streams

1 Introduction

Plastic cards have been a major advance in the banking and credit services offered by lenders. However, accompanying such advances, plastic card providers are challenged with the serious problem of fraudulent card transactions. To illustrate the magnitude of the problem, it is estimated that losses attributed to such fraud in the UK in the first six months of 2006 amounted to £209 million (APACS, 2008). Note that during this period, UK lenders introduced a scheme requiring PIN authentication for the majority of transactions. Since then, fraudsters have adopted new tactics to circumvent the scheme. Fraud detection methods must therefore have the potential to adapt to shifting fraudulent behaviour. In this arms-race the data available to lenders is constrained by existing information processing infrastructure. To enhance the data available to lenders, in this paper we explore the utility of extracting new features from the stream of transactions. Large and sophisticated infrastructure exists to rapidly process plastic card transactions.

Loosely, fraud implies unauthorized and illegal use of the facilities of a legitimate account. Tackling fraud in the context of plastic card finance is a daunting

problem. A number of complicating factors are involved including the sheer volume of transactions to process, the asynchronous and heterogeneous nature of transactions, and the adaptive behaviour of fraudsters. The effort to handle fraud can be broadly divided into *fraud prevention*, that attempts to block fraudulent transactions as they occur, and *fraud detection* where successful fraud transactions are subsequently implicated. For fraud prevention purposes, lenders typically challenge all transactions with rule based and other filters, often based on third party software such as, for example, VISA's VISOR fraud detection tool (VISA, 2003). Fraud detection should find fraudulent transactions as rapidly as possible after they occur. In the case of both fraud prevention and detection, the problem is magnified by specific characteristics of plastic card finance. First, to avoid customer irritation, the number of incorrectly implicated transactions needs to be kept to a minimum, Second, most lenders routinely process vast numbers of transactions, more than 20000 per day is not uncommon, of which only a small fraction is fraudulent, often less than 0.1%.

Many approaches to fraud problems have been considered (for example (Kou *et al.* , 2004), provide general discussion). Statistical views are explored by (Bolton & Hand, 2002). In the context of plastic card fraud, various authors (Brause *et al.* , 1999; Maes *et al.* , 2002) have approached fraud detection as a classification problem. There are a number of difficulties with this approach, including the extensive processing requirement associated with irregularly timed transaction sequences, and the conversion of the data into a representation suitable for classification algorithms. Moreover, the approaches may ignore important temporal aspects of fraud, particularly that fraudsters change tactics – classification approaches can only find existing tactics. We propose tools that complement standard detection methods by providing new features.

There are many other ways to approach this problem, such as peer group analysis (Weston *et al.* , 2008) or outlier detection (Juszczak *et al.* , 2008). It is unlikely that just one approach will successfully detect all types of fraud. Practical systems adopt more than one approach. We are attempting to provide new features that should enhance these hybrid systems. To this end, we consider the application of streaming clustering algorithms (Cao *et al.* , 2006; Tasoulis *et al.* , 2006), that to the best of our knowledge, have not yet been applied to this problem.

Streaming data, consisting of multiple indefinitely long and time-evolving sequences, is becoming ubiquitous. Such data presents new challenges to data mining algorithms. These challenges arise primarily from the dynamically changing nature of the streams, thus clustering algorithms must have the capacity to adapt rapidly to changing dynamics of the sequences. Additionally, timely results and scalability in the number of sequences is becoming increasingly desirable, as data collection technology develops. Plastic card transaction data, has precisely these characteristics. In this contribution we utilize density-based clustering methods, that are an important category of clustering algorithms (Jain *et al.* , 1999). These methods partition data into clusters of high density, surrounded by regions of low density. To address the issues of stream dynamics, timeliness and scalability, recent developments (Cao *et al.* , 2006; Tasoulis *et al.* , 2006) have extended the most successful density clustering

algorithms to the streaming data model. An analysis of these methods is presented in Section 2.

Among the collection of methods used in plastic card fraud detection, stream clustering algorithms are potentially advantageous for two particular reasons. Such algorithms can operate asynchronously at the transaction level in real time. Also as exploratory tools, these algorithms have the potential to identify different types of fraud, and characterize temporal components of fraud transactions.

In the next section we briefly review the literature on data stream clustering, and describe the algorithms used in this work. In Section 3 we describe the plastic card transaction data stream. Next, in Section 4, we present experimental results using streaming clustering. We conclude with a discussion in Section 5.

2 Data stream clustering

Traditional clustering methods are not able to accommodate the needs of the streaming data model, since they rely on the assumption that the data are available in a permanent memory structure, from which global information can be obtained at any time. Recently however new algorithms (Aggarwal *et al.* , 2004; Cao *et al.* , 2006; Tasoulis *et al.* , 2006), have been developed that embrace the need of clustering in streaming applications. Here we focus on density based clustering methods and in particular on the DenStream and WStream algorithms, for three reasons (Cao *et al.* , 2006; Tasoulis *et al.* , 2006). First, both of them are based on successful static clustering algorithms. Second, they have the desirable characteristic of providing approximations to the cluster number without prior knowledge, and at the same time they are able to detect non-convex clusters without any particular data transformation. Finally, they are both computationally efficient, which is a fundamental requirement for the application to credit card transactions.

The WStream algorithm The WStream algorithm (Tasoulis *et al.* , 2006) uses containers (*windows*) in the form of hyper-rectangles that are adjusted though time to discover and track the evolution of the underlying clusters. This is achieved using two procedures, "*movement*" and "*enlargement-contraction*". The "*movement*" of windows incrementally recenters windows every time a new streaming data point arrives. Windows are recentered to the mean of the points they include at each time point in a manner that also depends on each point's timestamp. A fading function, that decreases with time, associates a weight with each timestamp. This function depends on a parameter called the *forgetting factor*, that provides a compromise between the ability to track changes and the need to suppress the uninformative stochastic behaviour of the data. The larger the value of this parameter the faster the algorithm forgets previous information. On the other hand when it attains small values the algorithm is strongly influenced by historic information. The "*enlargement-contraction*" procedure aims to iteratively adapt the window widths, that relate to the scales of the different variables. The width of each co-ordinate of a window is enlarged or contracted depending on rules that also depend on user defined parameters.

The algorithm maintains a list of windows that are iteratively adjusted each time a new streaming data point arrives. New windows are created when the new data are not included in any of the windows. For a detailed description of WStream and a sensitivity analysis of its parameters see (Tasoulis *et al.* , 2006).

The DenStream algorithm DenStream (Cao *et al.* , 2006) was developed from its static counterpart (DBSCAN (Sander *et al.* , 1998)), which dictates that in a neighbourhood of a given radius, for each point in a cluster at least a minimum number of points should be contained. DenStream utilizes micro-clusters to extend this concept to the spatio-temporal setting. Micro-clusters are defined as quantities that capture the weight of points that reside in an area of a specific size, called the *diameter* of the micro-cluster. The *weight* is computed by again utilizing a fading function applied on the timestamps of the points inside each such area. Similar to WStream, the fading function is based on a parameter called forgetting factor.

Two types of micro-clusters are considered based on user defined parameters. We have a *core-micro-cluster*, if the cluster weight is large enough, and its diameter is small enough. These account for a "dense" region of the data. Otherwise the micro-cluster is called an *outlier-micro-cluster*. DenStream maintains two lists; one for the core-micro-clusters, and the other for the outlier-micro-clusters. These lists are updated each time new data arrives. Initially, an attempt is made to merge the new data into its nearest core-micro-clusters. If the resultant micro-cluster has a significantly large diameter (based on user defined rules) the merge is omitted. In this case, a new attempt is made to merge the new data into its nearest outlier-micro-cluster, using a similar rule. If this merge also fails a new outlier-micro-cluster is created, centered at the new data. If however the merge changed the outlier-micro-cluster such that it can be considered a core-micro-cluster it is moved to the appropriate list.

In order to derive the clustering result a variant of the DBSCAN algorithm is applied on the list of core-micro-clusters. Each core-micro-cluster is regarded as a virtual point located at its center, having its respective weight. The concept of density-connectivity (Sander *et al.* , 1998), is used to derive the final clustering result.

3 The transaction stream

A credit card transaction record is a complex entity. A fundamental identifier is the particular *account* associated with the transaction. The focus of this paper is the sequence of transaction as they are processed by the lender. We are less concerned with account level structure, as with exploring the spatio-temporal structure of all transactions.

One of our commercial collaborators provided transactions records with 77 fields. These refer to a diverse set of information, including process-oriented fields like card reader response status codes. Such fields can be used to precisely identify important details of the transaction. A fundamental distinction is provided by the service ID, that indicates transaction type, determining whether the transaction was

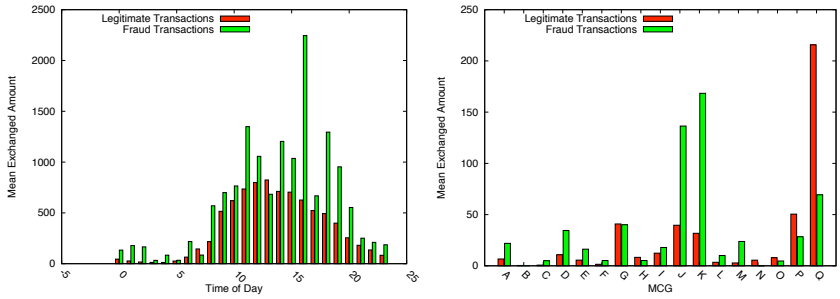

Fig. 1. Time of Day – Amount Plot(left), MCG – Amount Plot (right).

conducted at an automatic teller machine (ATM) or at a point-of-sale (POS) terminal. Note that it is possible to complete a variety of transaction types at an ATM in addition to cash withdrawals. For example, transferring money between accounts.

We extract a variety of data from transaction records. All POS transactions have a merchant category code marker. The merchant category code (MCC) is used to identify in which market segment the transaction was performed. For example "Book Stores". The MCC is a four digit number, though there are far fewer than 10000 codes currently in use. A merchant category group (MCG) is a grouping of MCCs into a broader market segment using business criteria. More details of MCCs and MCGs are given in (Weston *et al.* , 2008). We use MCGs to reduce the dimensionality of the merchant categories to 17. We introduce one further merchant category group to label ATM transactions. It is clear that streams consisting of transactions from heterogeneous accounts, require extra processing to manage this extra structure. We handle this by comparing each transaction with the profile of the respective account (Juszczak *et al.* , 2008) as explained later.

The data we extract from each transaction is the extended MCG and two other features, the time of day the transaction occurred and the amount of money exchanged. For illustration the left plot of Fig 1 shows the relation between time of day and mean amount exchanged, for 20000 transactions, by fraud status. Similarly the right plot of Fig. 1 illustrates the relation between the mean amount and the MCG (in arbitrary order). We include these figures to illustrate the complex structure of the data. However easy it may seem to identify fraudulent behaviour, we must note first that the figure refers to a single day's data and second that some structure in the graph may be an artefact of imbalanced classes. For such reasons we are interested in methods that view the data differently.

To handle heterogeneity induced by accounts, for each account holder we construct a simple updating profile and measure a transaction's deviation from its profile. This profile consists of the mean amount exchanged and the number of times each MCG was used in a transaction, as a proportion of the total number of transactions the account holder performed. Thus each transaction in the stream is composed of 20 features, the first of which is the time it arrives. The second is the deviation

from the profile mean for the amount associated with the transaction. The remaining 18 are the MCG counts as described above.

4 Streaming clustering

In this section we explore the behaviour of streaming clustering on fraud transaction data. Our objective is not to provide a comparative analysis of this method for fraud detection, but rather to characterise the structures that the method reveals.

To evaluate the stream clustering algorithm's performance we deployed WStream and DenStream to adapt their clustering results on the first 100000 transactions. Subsequently we examined the relationship between the label of the last 2000 transactions in this sequence to the cluster structure discovered by the algorithm. The results are tabulated in Table 1, where we label clusters based on the absence or present of fraudulent transactions. A cluster is labeled as fraudulent if it includes at least one fraudulent transaction, otherwise it is labeled as legitimate. Note that WStream discovers larger clusters than DenStream, while maintaining the fraud detection potential. Importantly, both methods place the same 3 of the 6 fraudulent transactions in similarly dense clusters.

	L.T. Clusters	F.T. Clusters	L.T. Outliers	F.T. Outliers
WStream	100;	3	386 trans.	3 trans
	Biggest Cluster: 88 trans.	60 trans.		
	Smallest Cluster: 2 trans.	30 trans.		
DenStream	160	4	193 trans.	2 trans
	Biggest Cluster: 30 trans.	47 trans.		
	Smallest Cluster: 1 trans.	9 trans.		

Table 1. Clustering results for the WStream and DenStream algorithms (L.T: Legitimate transactions, F.T.: Fraudulent Transactions).

To evaluate the stream clustering performance over the complete dataset we used the following methodology. We deployed WStream to continuously adapt to the complete data stream. Each time a fraud transaction occurred, we examined the correspondence of the most recent 2000 transactions, on the clustering adapted so far. Each transaction that did not belong to a window was regarded an outlier. We define as the False Positive (FP) ratio the proportion of legitimate transactions flagged as outliers. Respectively, the proportion of fraudulent transactions appearing in those 2000 transactions that are flagged as outliers is defined as the True Positive (TP) ratio. The results exhibited in Fig. 2 report the False Positive ratio (solid line), and the True Positive ratio (dotted line), for two runs of the algorithm with different values for the forgetting factor. In the top graph of Fig. 2, the forgetting factor was set to emphasize recent transaction activity while in the bottom graph of Fig. 2, the forgetting factor was set to emphasize historical data. Regardless of the choice of forgetting factor, certain fraudulent transactions can be detected. These transactions are dissimilar from other transactions regardless of the time they appear. The choice

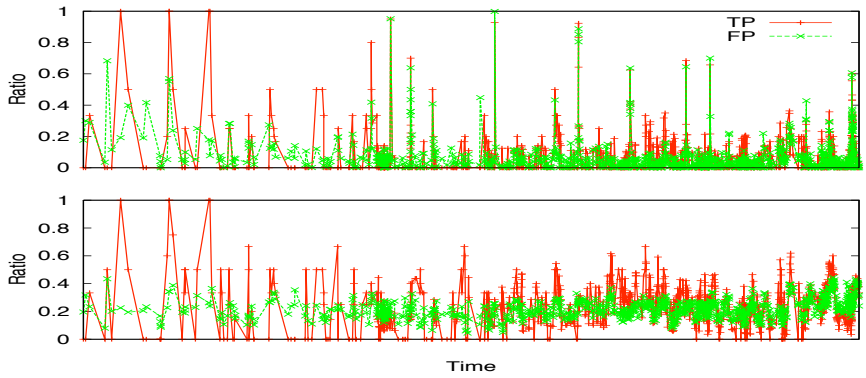

Fig. 2. Top: Fast forgetting, Bottom: Slow forgetting.

of forgetting factor affects which fraudulent transactions we detect. Choices of forgetting factor that rapidly adapt to immediate changes in the stream lead to a lower mean FP rate. Rapid adaptation occasionally produces very high FP rate, which we interpret as evidence for temporally local transactions far apart in the feature space. Slower adaptation induces a more steady FP rate, indicative of a slowly changing structure of the streams. Slower adaptation also appears to enhance fraud detection in the sense that TP rate is higher than the FP rate.

5 Discussion

In this paper we propose stream clustering as an innovative feature extraction method to enhance fraud detection methodology. Such innovation is essential in battling the adaptive behaviour of fraudsters. The streaming tools we propose attempt to adapt to both the measurement and temporal structure of the data. In this way, we hope to reveal different aspects of the character of fraud.

Our experimental results, based on real data, suggest that it is possible to deploy these streaming clustering methods for knowledge discovery. Strikingly, different rates of forgetting appear to reveal different types of fraud structure. This mined information can be utilized to produce novel features for fraud detection. For example, determining if a transaction is an outlier based on the clustering structure for the rest of the stream, provides a binary variable which enriches the description of the transaction, as as the experimental results show. This enriched representation could be passed to standard fraud detection technology.

6 Acknowledgements

The work of Dimitris Tasoulis work was undertaken as part of the ALADDIN (Autonomous Learning Agents for Decentralised Data and Information Systems) project and is jointly funded by a BAE Systems and the EPSRC (Engineering and

Physical Research Council) strategic partnership, under EPSRC grant EP/C548 051/1. The work of David Weston was supported by grant number EP/C5 32589/1 from the UK Engineering and Physical Sciences Research Council. The work of David Hand was partially supported by a Royal Society Wolfson Research Merit Award. We would like to express appreciation to the bank that provided the fraud data.

Bibliography

APACS (2008) *Card fraud facts and figures* `http://www.apacs.org.uk/resources_publications/card_fraud_facts_and_figures.html`.

VISA (2003) *VISA EU launches new advanced fraud detection tool,* `http://www.visaeurope.com/pressandmedia/newsreleases/press178_pressreleases.jsp`.

AGGARWAL, C., HAN, J., WANG, J., & YU, P. (2004). A framework for projected clustering high dimensional data streams. *Pages 852–863 of: 13th Int. Conference on Very Large Data Bases*.

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. *Page 103 of: Int. Conference on Tools With Artificial Intelligence*.

CAO, F., ESTER, M., QIAN, W., & ZHOU, A. (2006). Density-based clustering over an evolving data stream with noise. *Pages 326–337 of: SIAM Conference on Data Mining*.

JAIN, A.K., MURTY, M.N., & FLYNN, P.J. (1999). Data Clustering: A review. *ACM Computing Surveys*, **31**(3), 264–323.

JUSZCZAK, P., ADAMS, N.M., HAND, D.J., WHITROW, C., & WESTON, D.J. (2008). *Off-the-peg or bespoke classifiers for fraud detection? Computational Statistics and Data Analysis*, in press.

KOU, Y., LU, C.-T., SIRWONGWATTANA, S., & HUANG, Y.-P. (2004). Survey of fraud detection techniques. *Pages 749–754 of: IEEE Int. Conference on Networking, Sensing and Control*, vol. 2.

MAES, S., TUYLS, K., VANSCHOENWINKEL, B., & MANDERICK, B. (2002). Credit card fraud detection using Bayesian and neural networks. *In: International NAISO Congress on Neuro Fuzzy Technologies*.

SANDER, J., ESTER, M., KRIEGEL, H.-P., & XU, X. (1998). Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Data Mining and Knowledge Discovery*, **2**(2), 169–194.

TASOULIS, D.K., ADAMS, N.M., & HAND, D.J. (2006). Unsupervised clustering in streaming data. *Pages 638–642 of: 6th IEEE Int. Conference on Data Mining*.

WESTON, D.J., HAND, D.J., ADAMS, N.M., WHITROW, C., & JUSZCZAK, P. (2008). Plastic card fraud detection using peer group analysis. *Advances in Data Analysis and Classification*, **2**(1), 45–62.