Suspicious Behavior Detection in Debit Card Transactions using Data Mining: A Comparative Study using Hybrid Models

Ehsan Saghehei, Department of Industrial Engineering, Islamic Azad University Malayer Branch, Malayer, Iran
Azizollah Memariani, Department of Mathematics and Computer Science, University of Economic Sciences, Tehran, Iran

1. INTRODUCTION

Plastic cards are indispensable parts of the modern payment system. They provide users with a wide variety of services (Krivko, 2010). While these systems enjoy support from advanced software and hardware, the payment process still suffers from inadequate security (Panigrahi, Kundu, Sural, & Majumdar, 2009). This situation has made abuse and fraud possible and has imposed severe harm and much aggravation to the companies issuing the cards. To grasp the importance of the issue, statistics on the quantity of plastic card frauds must be reviewed. Actually, determining the amount of plastic card fraud is not an easy task, in part due to the companies’ reluctance to reveal fraud numbers publicly, partly to temper their custom-
ers’ dismay, and in part due to the volatility of fraud numbers reported during any one time period. Nevertheless, numerous estimates exist. For example, Aleskerov et al. cited estimates of $700 million in the United States each year for Visa/Mastercard fraud and $10 billion worldwide in 1996 (Bolton & Hand, 2002). In 2013, total amount of card fraud losses for cards issued in UK was reported to be £450.4 million (APACS, 2014).

There are various types of fraud in the plastic card industry. For a detailed classification one can refer to Kou, Lu, Sirwongwattana, & Huang (2004). As better efforts are made to detect and defeat fraud, the perpetrators become more devious and sophisticated, and the whole system becomes more complex. There is a need for continuous innovation in approaches to fraud prevention, detection and control. Financial fraud detection methods have been divided into two broad categories: supervised and unsupervised (Bolton & Hand, 2001). Phua et al. (2005) categorized fraud detection techniques in four groups: supervised, hybrid (supervised and unsupervised), semi supervised and unsupervised. In general hybrid models are comprised of two or more techniques of either supervised or unsupervised or combination of both.

In the supervised group, artificial neural network techniques have been used frequently (Aleskerov, Freisleben, & Rao, 1997; Dorronsoro, Ginell, Sngchez, & Cruz, 1997; Ghosh & Reilly, 1994; Maes, Tuylls, Vanschoenwinkel, & Manderick, 2002; Syeda, Zhang, & Pan, 2002). Also recently Akhilomen (2013) used this technique for cyber credit card fraud detection. Another widely used technique in the field of supervised models is the decision tree approach. Examples are Wang, Fan, Yu, & Han (2003) and Wei (2004). Bhattacharyya et al. (2011) applied support vector machines and random forests techniques as a comparative study. Their article was based on real-life data obtained from transactions from an international credit card operation.

In the field of unsupervised models, Zaslavsky and Strizhak (2006) designed a model Self-Organizing Map (SOM) for suspicious behavior detection in banking payment systems. Quah and Sriganesh (2008) generalized earlier SOM research to include fraud detection in real-time with minimum cost. They designed a model which involved three layers. The first layer was the initial authentication and screening layer and the third layer was used for reviewing and decision making. The second layer (core layer) was the main layer. It consisted of two sub-layers. The purpose of the core layer was risk scoring and customer behavioral analysis. In that research, SOM used the second layer for clustering input data and detecting hidden patterns in the input data. The second layer acted as a filtering and organizing mechanism for the other layers. Ogwueleka (2011) has also used SOM to detect fraud in credit cards. His system consists of two parts. The first section, the outer layer, acts as the interface with the database and ensures the completeness of the data. The internal layer is Kohonen network which clusters unlabeled data into four groups. The grouping is used for fraud risk. The Receiver Operating Characteristic (ROC) diagram used to tell abnormal from normal cases showed that the reliability of the system on real data is above 95%. This indicates that improvement is significant when the detection algorithms are joined with other detection software.

With regard to the other unsupervised techniques, Sanchez et al. (2009) used the association rules model to extract fuzzy rules from a dataset containing genuine and fraudulent data. They applied special methodology on data of credit card obtained from some of the most important retail companies in Chile. Srivastava et al. (2008) did a survey of unsupervised models in which unlabeled data is passed on into a hidden Markov model for fraud detection.

On hybrid models, some researchers carried out a variety of studies on designed meta classifier models and cost-based models for credit fraud detection (P. Chan & Stolfo, 1998; Lee & Stolfo, 1998; S. Stolfo, Fan, Lee, Prodromidis, & Chan, 1997; S. J. Stolfo, Fan, Lee, Prodromidis, & Chan, 2000). Also, P. K. Chan et al. (1999) utilized naive Bayes,
Object-Oriented Software Reuse in Business Systems
www.igi-global.com/chapter/object-oriented-software-reuse-business/13994?camid=4v1a