Journal of Network and Computer Applications 68 (2016) 90–113

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

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

Review

Fraud detection system: A survey

Aisha Abdallah *, Mohd Aizaini Maarof, Anazida Zainal

Information Assurance and Security Research Group, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia

ARTICLE INFO

Article history:
Received 18 August 2015
Received in revised form
4 January 2016
Accepted 11 April 2016
Available online 13 April 2016

Keywords:
Fraud
Fraud detection systems (FDSs)
Areas of fraud
E-commerce systems
Credit card system
Telecommunication system
Issues and challenges
Concept drift
Skewed distribution
Large amount of data
Support real time detection

ABSTRACT

The increment of computer technology use and the continued growth of companies have enabled most financial transactions to be performed through the electronic commerce systems, such as using the credit card system, telecommunication system, healthcare insurance system, etc. Unfortunately, these systems are used by both legitimate users and fraudsters. In addition, fraudsters utilized different approaches to breach the electronic commerce systems. Fraud prevention systems (FPSs) are insufficient to provide adequate security to the electronic commerce systems. However, the collaboration of FDSs with FPSs might be effective to secure electronic commerce systems. Nevertheless, there are issues and challenges that hinder the performance of FDSs, such as concept drift, supports real time detection, skewed distribution, large amount of data etc. This survey paper aims to provide a systematic and comprehensive overview of these issues and challenges that obstruct the performance of FDSs. We have selected five electronic commerce systems; which are credit card, telecommunication, healthcare insurance, automobile insurance and online auction. The prevalent fraud types in those E-commerce systems are introduced closely. Further, state-of-the-art FDSs approaches in selected E-commerce systems are systematically introduced. Then a brief discussion on potential research trends in the near future and conclusion are presented.

Contents

* Corresponding author.

http://dx.doi.org/10.1016/j.jnca.2016.04.007
1084-8045/

1. Introduction

Nowadays, most organizations, companies and government agencies have adopted electronic commerce to increase their productivity or efficiency in trading products or services; in areas such as credit card, telecommunication, healthcare insurance, automobile insurance, online auction, etc. (Bolton and Hand, 2002; Allan et al., 2010; Pejic-Bach, 2010). Electronic commerce systems are used by both legitimate users and fraudsters; hence they become more vulnerable to large scale and systematic fraud. Fraud is a crime where the purpose is to appropriate money by illegal means. The Association of Certified Fraud Examiners (ACFE) defines "fraud" as: the use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organization's resources or assets (ACFE, 2002). Internet Crime Complaint Centre (IC3) is a valuable resource for both victims of Internet crime and law enforcement agencies in identifying, investigating and prosecuting these crimes. In 2014, the IC3 received 269,422 complaints with an adjusted dollar loss of $800,492,073; which is a 2.39 percent increase in reported losses since 2013 ($781,841,611) (IC3, 2014). Table 1 summarizes the number of complaints received by the IC3 between 2011 and 2014 and the corresponding dollar losses. From this table, amount of loss steadily increase while number of complaints decrease; this is because, fraud is causing more loss now compared to the past. These huge number of losses have increased the importance of fraud fighting Kou et al., 2004). The purpose of fraud prevention mechanism is to protect the technological systems against fraud by stopping fraud from occurring in the first place. Nevertheless, this mechanism alone is not enough to halt fraud. Fraud detection is also proposed to improve the technological systems security. Fraud detection detects and recognizes fraudulent activities as they enter the systems and reports them to a system administrator (Behdad et al., 2012). Similar to detection approaches in Intrusion detection system (IDS), FDS also uses misuse and anomaly based approaches to detect fraud (Fawcett and Provost, 1997; Sasirekha et al., 2012). Both misuse based FDSs and anomaly based FDSs utilize data mining techniques to determine fraud from large amount of incoming data stream (Ngai et al., 2011). However, there are issues and challenges that hinder the development of an ideal FDS for E-commerce system; such as concept drift, supports real time detection, earliness of detection, skewed distribution, large amount of data, misclassification cost, etc. The presence of any one of these challenges will lead to high false alerts, low detection accuracy and slow detection. These are the parameters used to characterize the performance of FDS. In this paper, we will survey fraud detection systems in five areas that frauds usually occur which are credit card, telecommunication, healthcare insurance, automobile insurance and online auction.

Table 1
IC3 Report on Internet crime.

Year	Complaints received	Dollar loss
2011	314,246	$485,253,871 Million
2012	289,874	$581,441,110 Million
2013	262,813	$781,841,611 Million
2014	269,422	$800,492,073 Million

The remainder of this paper is outlined as follows. Section 2 presents the definition of fraud. Section 3 contains the related review and survey papers in the fraud detection system. Section 4 addresses approaches and mechanisms used to protect against fraud. Section 5 introduces the challenges and difficulties faced by fraud detection systems. Section 6 further defines the types of fraud, fraud detection system approaches and techniques and introduces the challenges and difficulties faced by the fraud detection system in each area. Section 7 discusses the challenges of fraud detection systems and their impact. Finally, Section 8 concludes the paper.

2. Fraud

There are many definitions of fraud and fraudulent activities. The Association of Certified Fraud Examiners (ACFE) defines "fraud" as: the use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organization's resources or assets (ACFE, 2002). The main reason behind the commitment of fraud is to achieve gain on false ground by an illegal means. This has a dramatic impact on the economy, law and even the human moral values (Alexopoulos et al., 2007). Almost all technological system that involves money and services can be compromised by fraudulent acts; for example the credit card, telecommunication, health care insurance, automobile insurance and online auction system (Almeida, 2009). Therefore, frauds in these systems are considered as cyber-crime, causing huge amount of financial losses.

According to the Basel Committee on Banking Supervision, there are different kinds of fraud: internal/occupational frauds or external frauds. Internal frauds happen when an employee commits frauds against his or her organization. In Phua et al. (2005), internal fraud is layered into two levels, it is a high level fraud if the employee is from the management division, and it is considered low level if the employee is not part of the management division. In contrast, external frauds involve a wide range of schemes, including vendors, customers or thefts by other third parties (Chen and Gangopadhyay, 2013). There are three types of external fraudster: 1) the average offender is called soft fraud, 2) criminal offender, and 3) organised crime offender is called hard fraud (Bhowmik, 2011).

3. Related works

Fraud detection system is important in several significant and sensitive sectors or areas. Therefore, fraud detection has been the topic of various surveys and review articles; that may be based on topics such as fraud areas, fraud types, fraud detection approaches and techniques. Bolton and Hand (2002), Kou et al. (2004), Phua et al. (2005), Allan et al. (2010) and Pejic-Bach (2010) surveyed

fraud detection done on different areas based on data mining and statistical techniques. Behdad et al. (2012) reviewed fraud detection utilizing nature inspired techniques. Nature inspired techniques, as the name implies, are artificial intelligence techniques which are inspired by how natural systems work. For instance, neural network is inspired by an animal's central nervous system (particularly the brain) which is capable of learning and recognizing. In their study, Behdad et al. (2012) also covered on the challenges that can be faced by FDS. Li et al. (2008), Travaille et al. (2011) and Liu and Vasarhelyi (2013) surveyed and analyzed fraud detection statistical methods for health care fraud detection. From another aspect, Delamaire et al. (2009) presented the different types of credit card frauds, such as bankruptcy fraud, counterfeit fraud, theft fraud, application fraud and behavioral fraud, and discussed on the appropriate techniques to fight them; such as a pair wise matching, decision trees, clustering techniques, neural networks, and genetic algorithms. In the same area, Raj et al. (2011) analyzed different kind of methods that are used to detect credit card fraud. Rebahi et al. (2011) presented the VoIP fraud problem and surveyed the fraud detection systems proposed in various areas, and their usability in the VoIP context. These detection systems are categorized into rule based supervised and unsupervised methods. Richhariya (2012), Ngai et al. (2011) and Wang (2010) provided a comprehensive survey and review for different data mining techniques used to detect financial fraud. Lookman Sithic and Balasubramanian (2013) presented an extensive survey for fraud types in medical and motor insurance systems and many types of data mining techniques are used to detect fraud in these insurance sectors. As we can see, there are numerous articles that surveyed FDS techniques, although almost all existing surveys do not highlight challenges and issues faced by FDS. Thus, this survey attempts to make a structured and comprehensive overview of the research on fraud detection. This is done by covering the fraud types, fraud detection approaches, fraud detection techniques, as well as fraud detection issues and challenges in the five identified areas: credit card, telecommunications, healthcare insurance, automobile insurance and online auction. Table 2 summarizes the content of this survey and other existing surveys and review articles in terms of the techniques used and fraud areas studied. This survey aimed to improve the understanding of fraud detection directions in which research has been done on this topic, also we intend to identify what issues and challenges should be considered for an efficient fraud detection system. Fig. 1 shows the outline of this survey.

4. Protection against fraud

Fraud is increasing dramatically with the progression of modern technology and global communication. As a result, fighting fraud has become an important issue to be explored (Kou et al., 2004; Magalla, 2013). As presented in Fig. 2, the detection and prevention mechanisms are used mostly to combat fraud. Next subsections will explain further on fraud protection mechanisms.

4.1. Fraud prevention systems (FPS)

Fraud prevention system is the first layer of protection to secure the technological systems against fraud. The purpose of this phase to stop fraud from occurring in the first place. Mechanisms in this phase restrict, suppress, destruct, destroy, control, remove, or prevent the occurrence of cyber-attacks, in computer systems (hardware and software systems), networks, or data. Example of such mechanism includes using encryption algorithm that is applied to scramble data. Another mechanism is firewall where it forms a blockade between the internal privately owned network and external networks. It does not only help to secure systems from unauthorized access but also to allow an organization to enforce a network security policy on traffic flowing between its network and the Internet (Oppliger, 1997; Magalla, 2013). However, this layer is not always efficient and strong (Belo and Vieira, 2011). There are, in some occasions, where prevention layer could be breached by fraudsters.

4.2. Fraud detection System (FDS)

Fraud detection system is the next layer of protection; which is also the concern of this paper. Fraud detection tries to discover and identify fraudulent activities as they enter the systems and report them to a system administrator (Behdad et al., 2012). In previous years, manual fraud audit techniques such as discovery sampling have been used to detect fraud, such as in Tennyson and Forn (2002). These complicated and time consuming techniques transact with various areas of knowledge like economics, finance, law and business practices. Therefore, to raise the effectiveness of detection, computerized and automated FDS was invented. However, FDS capabilities were limited because the detection fundamentally depends on predefined rules that are stated by experts (Li et al., 2008). More complex FDSs integrating a wide range of data mining methods are required and are being developed for effective fraud detection (Akhilomen, 2013; Koh and Tan, 2005; Guo and Li, 2008; Ogwueleka, 2011; Desai and Deshmukh, 2013;

Table 2
Fraud detection system survey and review articles.

Reference	Technique	Fraud area
Bolton and Hand, 2002; Kou et al., 2004; Phua et al., 2005; Allan et al., 2010; Pejic-Bach, 2010	The intelligent systems: neural networks, fuzzy intelligence, genetic algorithms, evolutionary programming, genetic programming, evolution strategies, and particle swarm optimization.	Telecommunications, insurance, auditing, medical care, credit card transactions, e-business, bid pricing and identity verification.
Behdad et al., 2012	Nature inspired techniques.	Email, spam, phishing and network intrusion.
Li et al., 2008; Travaille et al., 2011; Liu and Vasarhelyi, 2013	Spatial temporal data mining techniques.	Healthcare insurance.
Rebahi et al., 2011	Rule-based, supervised and unsupervised methods.	Voice over IP (VoIP) networks.
Wang, 2010; Richhariya, 2012; Ngai et al., 2011	Data mining and statistical techniques.	Financial fraud detection.
Lookman Sithic and Balasubramanian, 2013		Home insurance, life insurance, motor insurance and medical insurance.
Delamaire et al., 2009; Chaudhary and Yadav, 2012; Zareapoor and Alam, 2012; Singh and Narayan, 2012; Tripathi and Pavaskar, 2012; Sethi and Gera, 2014		Credit card fraud.
This survey	Fraud types, fraud detection approaches, fraud detection techniques and fraud detection issues and challenges.	Credit card fraud, telecommunications fraud, healthcare insurance fraud, automobile insurance fraud, online auction fraud.

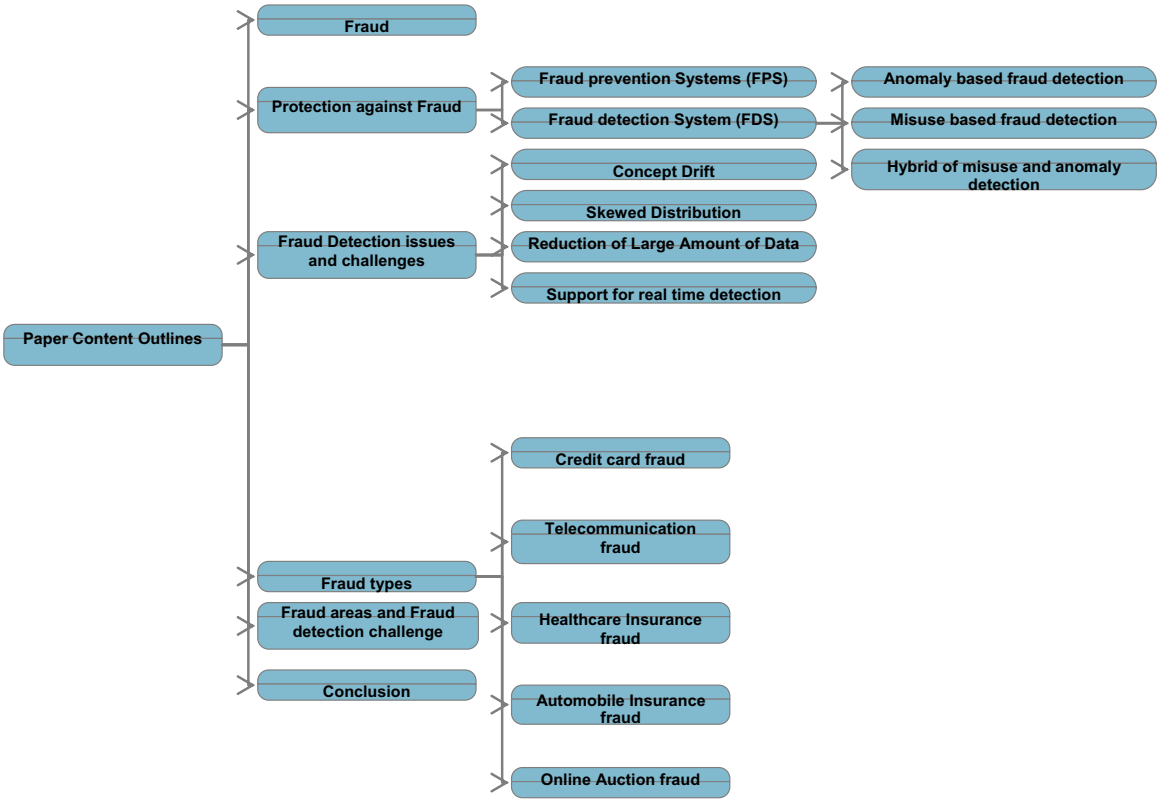

Fig. 1. Outline of this survey.

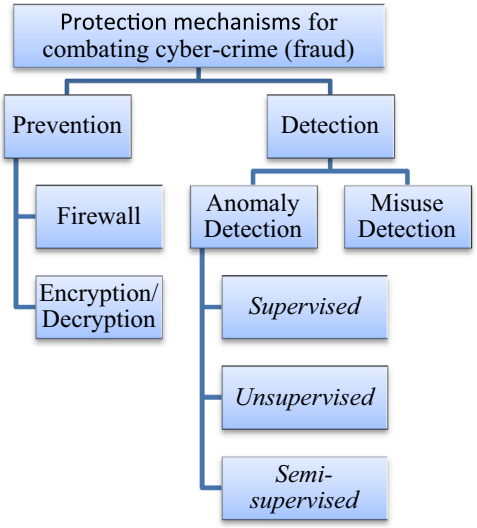

Fig. 2. Protection systems against fraud.

Saravanan et al., 2014). Data mining involves statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and subsequent knowledge from large databases (decision support systems and intelligent systems). These systems have several main advantages: (1) fraud pattern are obtained automatically from data; (2) specification of "fraud likelihood" for each case, consequently that efforts in investigating suspicious cases can be prioritized; and (3) revelation of new fraud types that were not defined before (Li et al., 2008). Data mining methods consist of six main categories which are Classification, Clustering, Regression, Outlier detection, Visualization and Prediction (Edelstein, 1997; Noor et al., 2015). Each of these methods is supported by specific techniques. For example, neural network technique and support vector machine technique are used for data mining classification method. K-means technique is used for data mining clustering method. Furthermore, data mining has incorporated many techniques from other domains such as statistics, machine learning, pattern recognition, database and data warehouse systems, information retrieval, visualization, algorithms, high-performance computing, and many application domains (Han et al., 2012). Recently, fraud detection integrates anomaly based detection approach and misuse based detection approach by using data mining techniques (Allen, 2000; Sasirekha et al., 2012).

A. Anomaly based fraud detection

Anomaly or outlier detection approach is used by FDS and it relies on behavioral profiling methods, where it models each individual's behavioral pattern, monitoring it for any deviation from the norm (Jyothsna and Rama Prasad, 2011). Anomaly based FDS are adopted by numerous authors in different fraud areas (Ghosh and Reilly, 1994; Dorronsoro et al., 1997; Taniguchi et al., 1998; Brause et al., 1999). Anomaly based FDSs have the potential to detect novel fraud. Therefore, it is mostly used by the FDS literature (Sun et al., 2006). This method can be further categorized into three types; unsupervised, Semi-supervised and supervised anomaly detection (Akhilomen, 2013).

- *Supervised*
  Supervised learning techniques require a data set that has been labeled as "fraud" and "nonfraud" and involves training a classifier. This is the most common learning approach. The major advantage of supervised learning is that all classes outputs manipulated by the algorithm of this approach are meaningful to humans, and it can be easily used for discriminative pattern

classification and data regression. However, supervised learning has several limitations. The first one is caused by the difficulty of collecting supervision or labels. When there is a huge volume of input data, it is prohibitively expensive, if not impossible, to label all of them. Second, sometime it is extremely hard to find distinctive label, there are uncertainties and ambiguities in the supervision or labels. These limitations may obstruct the implementations of the supervised learning approaches in some cases. Therefore, unsupervised learning and Semi-supervised learning are used to overcome these disadvantages (Liu and Wu, 2012). Supervised learning encompasses many algorithms include:

- Classification algorithms
  For example artificial neural network, K-nearest neighbors, trees, logistic regression, Naïve–Bayes and support vector machine (SVM) techniques.
- Regression algorithms
  For example linear regression, simple regression and logistic regression.
- *Unsupervised*
  Unsupervised learning techniques detect fraudulent in an unlabelled test data set under the assumption that majority of the instances in the data set is nonfraud. Unlike supervised technique, unsupervised means there is no class label for model construction. The main benefit of using unsupervised approach is that it does not rely on accurate identification for label data which is often in short supply or non-existent (Bolton and Hand, 2001). Two simple classic algorithms employed in unsupervised learning are:
- Clustering algorithms like K-means techniques.
- Dimensionality reduction algorithms such as: Principal Component Analysis (PCA)
- *Semi-supervised*

Semi-supervised learning lies between supervised and unsupervised learning since it involves a small number of labeled samples and a large number of unlabelled samples. The main goal of Semi-supervised approach is to train a classifier from both labeled and unlabeled data (Zhu et al., 2011; Akhilomen, 2013). Semi-supervised learning has more advantage compared to supervised learning because it achieves better performance by utilizing both labeled and unlabeled data, but with fewer labeled instances. Furthermore, Semi-supervised learning also provides a computational model to understand human category learning, where most of the input is self-evidently unlabelled (Xiaojin Zhu and Goldberg, 2009).

B. Misuse based fraud detection

In misuse detection approach, fraudulent behaviors are first defined by using fraudsters signatures, and then other behaviors are defined as normal behaviors. Misuse approach adopted by FDS utilizes rule-based, statistics, or a corresponding heuristic methods to reveal the happening of specific suspicious transaction (Hand and Crowder, 2012). Misuse detection is expert system which is considered as a simple and fast detection mechanism. But it has major limitation because it is not possible to detect all different kinds of frauds because it only looks for known patterns of misuse (Wei et al., 2012).

C. Hybrid of misuse and anomaly detection

Some researchers have been proposing a hybrid approach in which anomaly detection and misuse detection models are combined to get optimum results (Kundu et al., 2006; Sherly and Nedunchezhian, 2010; Sasirekha, 2012). This is due to that misuse detection's incapability to detect novel fraud; meanwhile, anomaly detection suffers from the lack of generalization capability and presence of high false alarm rates (Mule and Kulkarni, 2014). However, according to the literature, anomaly based FDSs is the most commonly used approach (Sun et al., 2006; Akhilomen, 2013).

5. Fraud detection issues and challenges

Fraud detection is a complex domain; we may find that a fraud detection system is prone to fail, has a low accuracy rate, or gives many false alarms. It is extremely difficult for electronic commerce systems to handle fraud problem forcing them to incur heavy losses. This happens because fraud detection systems need to deal with multiple challenges to be taken into account. Several challenging properties that fraud detection must deal with will be presented in this section. Fig. 3 shows distribution of FDS articles based on issues and challenges. The statistics are based on number of papers published between 1994 and 2014. The focus is taken from the most prevalent types of e-frauds: credit card, telecommunication, healthcare insurance, automobile insurance, and online auction frauds.

5.1. Concept drift

There are a several definitions for concept drift issue in the literature. In data mining, concept drift refers to the phenomenon that the underlying model (or concept) is changing over time (Abbass et al., 2004). FDSs work in dynamic environment where behavior of legitimate user or fraudster is continuously changing is called the drift phenomenon concept (Gama et al., 2013). For example, in credit card area, the cardholder behavior may be subject to change due to a variety of external causes; for example, the transaction amount and frequency are closely related to the spending habits of a person which is actually influenced by income, resource availability, and lifestyle of a person, which may change with time (Malekian and Hashemi, 2013). In addition, fraudster's tricks are continuously evolving and detection has to adapt to these new fraud types (Dal Pozzolo et al., 2014).

Furthermore, concept drift primarily refers to an online supervised learning scenario when the relation between the input data and the target variable changes over time. Whereas, in supervised learning, the aim is to predict a target variable y given a set of input features X. In the training instance that are used for model building, both X and y correspond to input data and target variable, respectively. In the new instance on which the predictive model is applied, X is known, but y is not known at the time of prediction, and the relation between the input data and the target variable may change (Gama et al., 2013). Concept drift is a big concern particularly in online learning where detection model is updated immediately but based is based upon outdated data, so

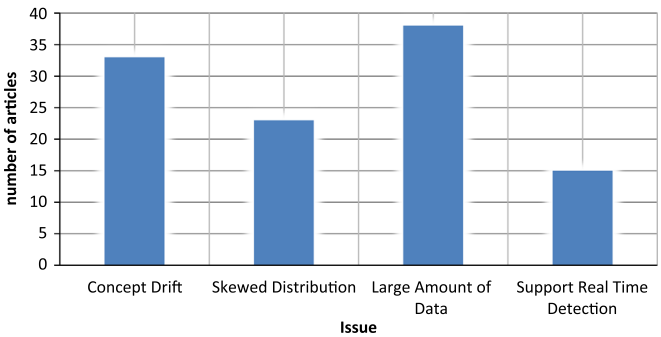

Fig. 3. Distribution of FDS articles based on issues and challenges between 1994 and 2014.

when new data arrives, the model may be misleading and gives many false alarms. Therefore, attention has been dedicated in the literature to deal with non-stationary behavior and dynamically update fraud detection model (Malekian and Hashemi, 2013). So, the use of adaptive learning algorithms to handle concept drift issue is required. Adaptive learning algorithms can be seen as advanced incremental learning algorithms that are able to update detection model for streaming evolving data over time (Bolton and Hand, 2002; Phua et al., 2005; Oza, 2005; Gama et al., 2013). Zliobaite (2010) expressed the definition of incremental learning with drift as follows: incremental learning process at every time t, where a historical data is available, a target instance X_{t+1} arrives, the task is to predict label Y_{t+1}. For that the learner L_t is built in training phase by using all or selection from historical labeled data $X_{historical}=(X_1,\ldots,X_t)$. This is as illustrated in Fig. 4. By incremental learning process, the label y_{t+1} becomes available with X_{t+1} will be part of history to predict X_{t+2}.

The common taxonomy for existing learners responsive to a concept drift in adaptive FDSs is categorised into two groups based on when the adaptive function is activated: this can be either evolving based or regulated based as shown in Fig. 5 (Wang et al., 2003; Zliobaite, 2010; Ross et al., 2012). Evolving based approach is when the learner automatically adapts its behavior in staying up-to-date with the stream dynamics. Meanwhile, regulated based approach is when concept drift and classification is handled as separate problems. The designed concept drift detectors will flag when there are changes take place, and then some reactions should be taken. Generally, the advantage of regulated based is not only due to its adaptation to concept drift, but also in providing needed information about drift has taken place. For example, when FDS detects a credit card fraud, it is then required to take further investigative action of the behavior of the fraudsters. Regulated based method does not using frequently like evolving based method to handle concept drift in the fraud area (Ross et al., 2012; Gomes et al., 2011).

5.2. Skewed class distribution

Skewed distribution (imbalanced class) is considered as one of the most critical issues faced by FDS. Generally, the imbalanced class issue is the situation where there are much fewer samples of fraudulent instance than normal instance (Maes et al., 2002). In a supervised learning approach, the class imbalance problem happens when the minority class is very small, leading to numerous problems such as disability of learners to discover patterns in the minority class data (Stolfo et al., 1997). Furthermore, imbalance class has a serious impact on the performance of classifiers that are tend to be overwhelmed by the majority class and ignore the minority class (Liu et al., 2012).

For clarification, the 2009 University of California San Diego

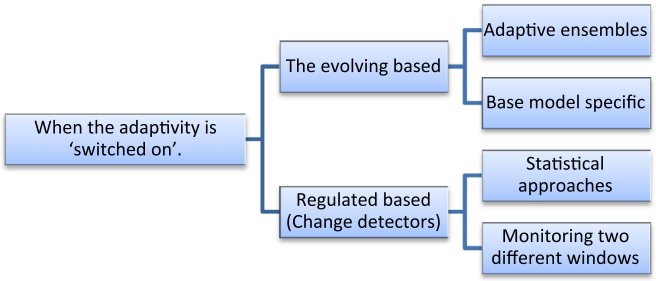

Fig. 5. Adaptive learning approaches.

Data Mining Contest (UCSD) dataset will be used to illustrate the imbalanced class issue. The UCSD dataset is a real dataset of E-commerce transactions and it is used to detect anomalous E-commerce transactions. The training dataset contains 100,000 transactions of 73,729 customers spanning over a period of 98 days; and the test set consists of 50,000 transactions. The training data is highly imbalanced, which consists of 97,346 normal transactions (majority class) and only 2654 fraudulent transactions (minority class), as shown in Fig. 6. Percentage of normal transactions (majority class) it is around 97–3% fraudulent transactions (minority class). Therefore, a balancing mechanism is required to make this data balanced with ratio of 1:1 between normal and fraudulent class to handle class imbalance. This will make fraud extremely easy to detect, because the difference between minority and majority class samples can be recognized effectively. Data balancing approaches can be categorised into two different levels, data level and algorithmic level. Fig. 7 shows the balanced approaches and techniques (López et al., 2012).

A. Data level methods

Balancing techniques at data level are used as a pre-processing step to rebalance the data set or remove the noise before the application of other classification algorithms. In FDS literature, most researchers employ data level balancing techniques, for example undersampling or oversampling approaches.

- Under sampling approach removes part of the data in the majority class (Chen, 2006). A wide number of the proposed fraud detection systems utilized the under sampling approach to balance the training data.
- Over sampling approach replicates the data in the minority class. Over sampling approach is rarely used because it causes overfitting of a model, particularly with the existence of a noisy data. Also, over sampling does not result in more information being included in the training set, which leads to a very complex model (Hofmann, 2012). Alternatively, SMOTE (Synthetic

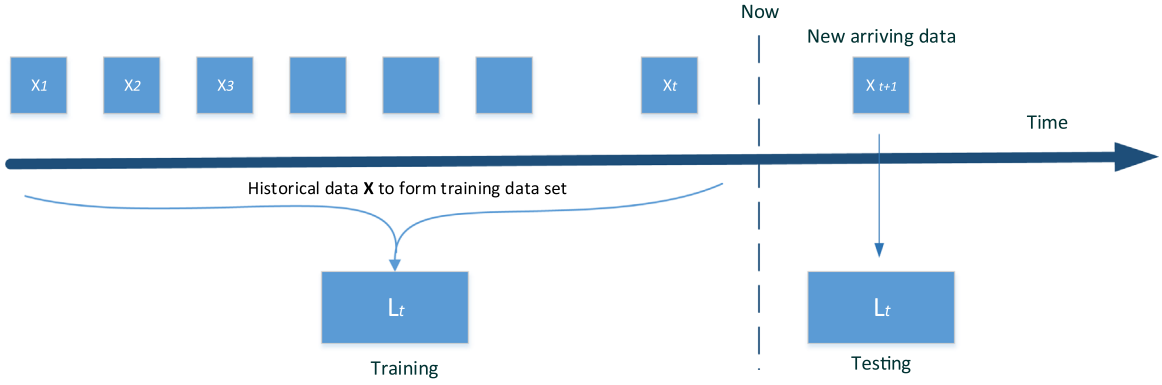

Fig. 4. Incremental learning in time t.

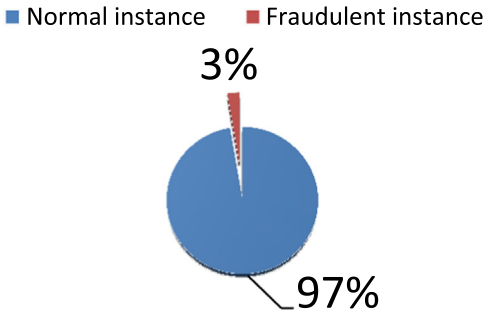

Fig. 6. Imbalance training UCSD dataset.

Minority Over Sampling Technique) (Chawla et al., 2002) is used for fraud detection which is the better alternative to over sampling approach. SMOTE oversamples the minority class by generating synthetic minority examples in the neighborhood of observed ones. (Dal Pozzolo et al., 2014) conducted several experiments over different kinds of balanced data level techniques (Under, SMOTE, EasyEnsemble) to find out the most efficient technique for credit card fraud detection system.

A. Algorithmic level methods

These kinds of classification algorithm deal with fraudulent class. The algorithmic level method employed:

- A cost-sensitive learning to deal with skewed class distribution. Cost-sensitive learning puts a cost-variable to misclassification of the different classes by assuming that there is a cost-matrix available for the different type of errors. Cost-matrix created to bias the model to minimize cost or maximize benefit. In the cost-matrix formulation, costs are associated with those predictions: false negative (if the true label is fraud and it is classified as normal), false positive (if the true label is normal and it is classified as fraud), true negative (if the true label is normal and it is classified as normal), and true positive (if the true label is fraud and the FDS classified as fraud). The classification for each instance may give only two entries (false positive, true negative) or (false negative, true positive) (Zadrozny et al., 2003). In FDS literature, there are main two approaches that have been proposed to utilise cost-sensitive learning for imbalanced class: metacost-thresholds or use of learners which are not sensitive to the class imbalance problem (Hofmann, 2012). These methods are used frequently in fraud detection system to balance training data.
- Using the learner itself to handle skewed distribution, which is another algorithmic method used in the FDS literature. These learners are either resistant to the class imbalance problem through inherent properties of the learner, as in the case of the Repeated Incremental Pruning to Produce Error Reduction (RIPPER) algorithm as presented in Chan et al. (1999). Furthermore, learners are hardened against the problem through internal modification as in the case of K-NN or the SVM learners.

Generally, data methods perform better than algorithm methods. This due to the fact that data methods are easier to implement and do not lead to the increase in training time or resources needed. Therefore, most FDS literature utilizes data level balancing techniques (Hofmann, 2012; Dal Pozzolo et al., 2014).

5.3. Reduction of large amount of data

Large-scale and high dimensions of fraud data set and presence of numbers of features/attributes/inputs/variables make the process of data mining and detection extremely difficult and complicated (Hilas and Sahalos, 2007). Besides, this situation also slows down the detection process. Therefore, the existing FDSs use data reduction approaches to reduce the size of data set (Viaene et al., 2004) producing small model size which may be useful with respect to real-time processing (Onderwater, 2010). In addition, small data will reduce the size of model, consequently reducing the computation time (Lane and Brodley, 1999). Data reduction approaches include dimensionality reduction and numerosity reduction (Kamber and Pei, 2012). Dimensionality reduction includes many strategies, namely data compression, feature selection and feature construction are the most common and frequently used strategies in FDSs. Data compression strategy compresses the representation of original data through the use of data compression techniques such as in Brockett et al. (2002), Ai et al. (2009) and Ju and Lu (2011). Meanwhile, features selection is another dimensionality reduction strategy, the most significant and relevant features are selected to be used in model construction. Feature selection has been adopted by Tsang et al. (2014) and

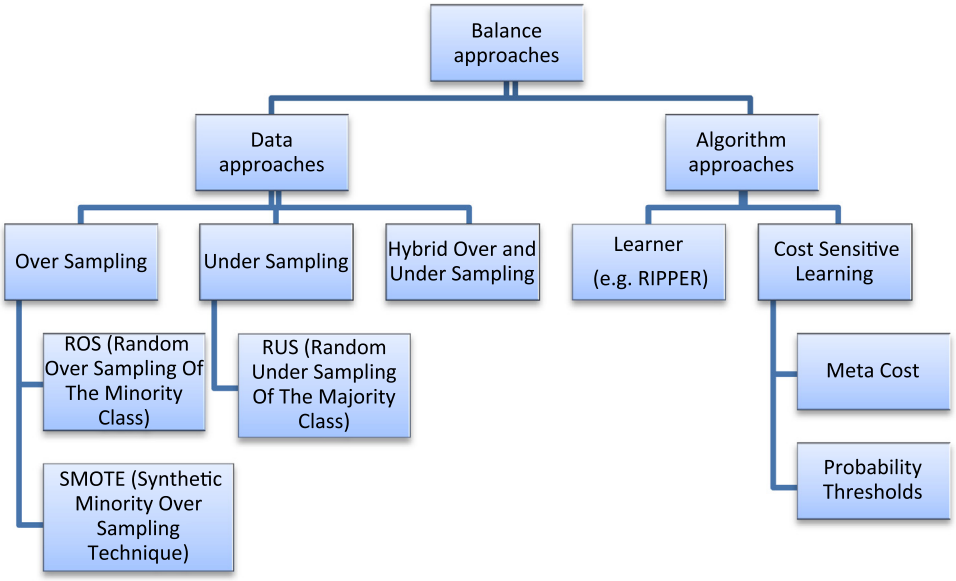

Fig. 7. Imbalance class handling approaches.

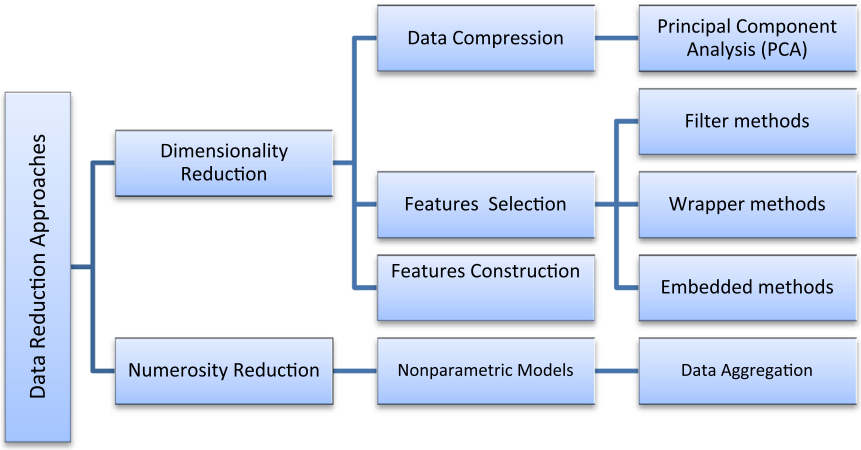

Fig. 8. Data Reduction Strategies.

Almendra and Enachescu (2011). Three feature selection methods are used in FDS: filter methods, wrapper methods and embedded methods. Filter methods act as preprocessing algorithm to rank the features wherein the highly ranked features are selected and applied to a predictor. In wrapper methods, the feature selection criterion is the performance of the predictor i.e. the predictor is wrapped in a search algorithm which will find a subset which gives the highest predictor performance. Embedded methods include variable selection as part of the training process without splitting the data into training and testing sets (Sánchez-Maroño et al., 2007). Next, feature construction is where a small set of more useful features is derived from the original set. Meanwhile, in numerosity reduction, the data are replaced by smaller representations like using data aggregation (Liu et al., 2010; Jha et al., 2012; Dal Pozzolo et al., 2014). Data reduction approaches include dimensionality reduction and numerosity reduction as presented in Fig. 8 (Kamber and Pei, 2012).

5.4. Supports real time detection

Fraud detection systems work in two different modes which are offline detection or online detection that is based on different fraud types. Where there are areas have real time applications which require online fraud detection. For example, a fraud in the online payment application in the credit card area needs immediate detection and response; as well as similar situation happening in an online auction. In contrast, there are applications that require offline detection. Although offline detection does not require instant reaction, yet researchers are still continuously working to optimize the detection process. Online fraud detection should be able to deal with limited resources (time and memory) in ensuring that the detection process works efficiently. Therefore, the efficiency of any proposed online fraud detection solution does not only benefit from the reduced amount of data (as in Section 5.3) but also from the reduced computational complexity of methods used for detection. This paper will focus on online detection, exploring the approaches used to achieve efficient real time detection in an aspect of resources consumption.

6. Fraud areas

Almost any technological system that involves money and services can be compromised by fraudulent acts, for example credit card system, telecommunication system, health care insurance system, etc. (Almeida et al., 2008). Fig. 9 shows the most common areas of frauds. This section is going to address fraud happening to the five most prevalent areas, which are credit card, telecommunication, health care insurance, automobile insurance and online auction areas.

Fig. 10 shows a statistics of published work related to the five fraud areas from 1994 to 2014. From the figure, it is prevalent that bank fraud is the most researched area. Insurance fraud is the third popular area, which has been the main subject of several studies since it may include be infused with other areas such as healthcare insurance fraud, automobile insurance fraud, home insurance fraud and crop insurance fraud. Telecommunication and Internet marketing are the least studied areas during the specified period of time. This survey paper will focus mainly on credit card fraud which is under bank fraud, healthcare and automobile fraud under

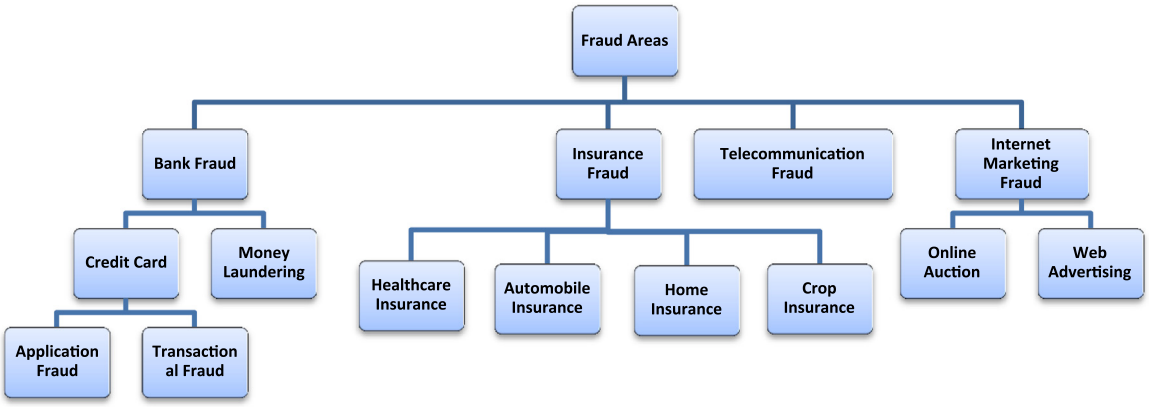

Fig. 9. Taxonomy of The Most Common Areas of Frauds.

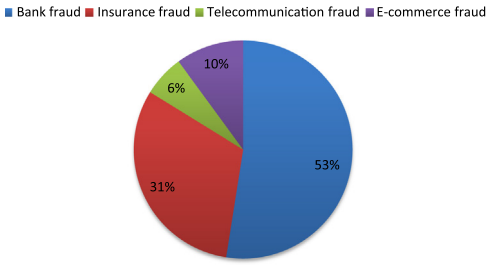

Fig. 10. Overview for the quantity of most researched area of fraud.

insurance fraud, telecommunication fraud and online auction fraud under Internet marketing fraud.

6.1. Credit card fraud

The term credit is used to describe the method of buying and selling goods without having money. Credit card is a small plastic card to provide the credit service to customer (Delamaire et al., 2009; Raj et al., 2011). Credit card is very popular and plays an important role in electronic commerce and online money transaction area which is growing every year. As a result of the growing usage of credit card, fraudsters try to find more opportunities to commit frauds that can cause huge losses to cardholders and banks (Sherly and Nedunchezhian, 2010). Credit card fraud is considered as common type of credit fraud (Potamitis, 2013). Credit card fraud takes many forms, existing literatures classify them into several common categories. Laleh and Azgomi (2009) classify the credit fraud based into:

I. Offline credit card fraud: Happens when the plastic card is stolen by fraudsters, using it in stores as the actual owner. This is an uncommon type of fraud because financial institutions will immediately lock the lost card when cardholders report about the theft.
II. Online credit card fraud: A popular and very dangerous fraud, credit cards' information are stolen by fraudsters to be used later in online transactions by Internet or phone. Another name for this kind of fraud is “cardholder not present” fraud; whereas the card's details are given over one of the following methods: Skimming, Site Cloning, Credit Card Generators or phishing (Patidar and Sharma, 2011).

There is another classification for credit card fraud, types into two types, namely application fraud and behavioral fraud. (Delamaire et al., 2009). This classification is based on fraudster's strategy on committing fraud. Application fraud occurs when fraudsters enter wrong information and details into the application form for opening new credit card. Fraudsters may use other persons' information to obtain credit cards or get their new credit cards by using false personal information with the intention of never repaying the purchases (Bolton and Hand, 2002). On the other hand, behavioral fraud occurs when fraudsters obtain cardholder details to use them later for sales which are made on a cardholder present basis. These sales include telephone sales and E-commerce transactions, where only the card details are required (Bolton and Hand, 2002). Ghosh and Reilly (1994) categorized credit card fraud into:

1. Lost cards and stolen cards.
2. Counterfeit cards.
3. Theft of cards from the mail or non-receipt of issue (NRI).
4. Mail/telephone order fraud. In such cases, the purchaser is not physically present before the merchant at the time of the transaction, and there is no card imprint that can be obtained as a record of the transaction.
5. Merchant fraud can involve the “laundering” of phony merchant receipts, garnering large sums of money for transactions that never occurred.
6. The cardholder makes purchases on the card for which he/she has no intention of paying. This is called bankruptcy fraud.

Patidar and Sharma (2011) proposed another credit card fraud categorization. It was divided into three categories, which are, traditional card related frauds, merchant related frauds and Internet frauds.

6.1.1. Credit card fraud detection

Credit card fraud detection is one of the most explored domains of fraud detection (Brause et al., 1999; Bolton and Hand, 2001; Sherly and Nedunchezhian, 2010). Numerous authorization techniques are used to prevent credit card frauds, such as signatures, credit card number, identification number, cardholder's address, expiry date, etc. However, these techniques are not enough to hinder credit card fraud. Therefore, there is a need to use fraud detection approaches which analyse data that can detect and eliminate credit card fraud (Sherly and Nedunchezhian, 2010). Mostly, the strategy of credit card fraud detection is pattern recognition by analyzing user spending behavior automatically. Customer spending behavior contains information about the transaction amount, time gap since last purchase, day of the week, item category, customer address, etc. Anomaly based fraud detection is mostly used for credit card fraud detection system in which the cardholder's profile is made up by analyzing the cardholder spending behavior pattern. In doing so, any incoming transaction that is inconsistent with the cardholder's profile would be considered as suspicious (Malekian and Hashemi, 2013). Profile approaches can be done based on owner approach or operation approach as classified by Chandola et al. (2009). Owner approach is where each credit card user is profiled based on his/her credit card usage history. Any new coming transaction is compared to the user's profile and suspected as a fraud if it does not match the profile. Meanwhile, operation approach detects fraudulent transactions from transactions taking place at a specific geographic location. On the other hand, misuse detection approach has always been performed to detect credit card frauds, where fraudulent persons' fingerprints patterns will be used to learn on the credit card FDS (Lei, 2012). Generally, misuse detection approach is seldom employed for credit card FDS or other fraud areas. This is due to its inability to detect new fraud. Therefore, in Table 3 we just focus on anomaly based approach as well as other fraud types.

As we can observe from Table 3, the most frequently used data mining methods to create credit card FDS is the classification method, in specific the neural networks techniques under supervised learning manner.

6.1.2. Credit card fraud detection issues and challenges

A. Concept drift

Credit card area is being addressed mostly by researchers compared to other fraud areas with respect to the concept drift issue. This is because credit card holders always change their behavior that may due to specific circumstance (e.g. Christmas holidays), and in this period, the purchasing power of users will be increased. If FDS does not consider this as normal change, it will be considered as fraudulent behavior and the alarms will be triggered, locking the cardholder's transaction, leading to the decrease to the bank reputation in the future. Therefore, the credit card FDS needs to discriminate and classify fraudulent and legitimate transactions effectively. Also, credit card FDS must be capable to capture and adapt the drifting behavior of

Table 3
Research on approaches and techniques in credit card FDS.

Detection type	Detection tool	Learning approach	Data mining categories	Techniques	References
Anomaly based fraud detection methods	**Data mining**	**Supervised**	Classification	Decision trees	Kokkinaki, 1997; Minegishi and Niimi, 2011
				Artificial neural networks	Ghosh and Reilly, 1994; Dorronsoro et al., 1997; Brause et al., 1999; Maes et al., 2002; Kim and Kim, 2002; Syeda et al., 2002; Zaslavsky and Strizhak, 2006; Quah and Sriganesh, 2008; Guo and Li, 2008; Omlin and Bénard, 2009; Patidar and Sharma, 2011; Sahin and Duman, 2011
				Hidden Markov model	Srivastava et al., 2008; Falaki et al., 2012; Kumari et al., 2014; Mule and Kulkarni, 2014
				Artificial immune systems	Gadi et al., 2008; Brabazon et al., 2010; Wong et al., 2012; Halvaiee et al., 2014; Seeja and Zareapoor, 2014
				K-Nearest neighbours	Halvaiee et al., 2014
				Support vector machine	Chen et al., 2004; Chen et al., 2005; Chen, 2006; Ju and Lu, 2011; Bhattacharyya et al., 2011
				Genetic algorithm	Duman and Ozcelik, 2011; Wu et al., 2007
				Fuzzy	Bentley et al., 2000
		Unsupervised	Clustering	Self-organizing map algorithm	Zaslavsky and Strizhak, 2006

cardholders, updating the detection model to that behavior over time. Consequently, credit card FDS should have low false alerts but high detection accuracy. For that reason, researchers used several approaches for handling concept drift in credit card system. These approaches will be categorized based on their taxonomy in Section 5.1, namely the evolving based approach and regulated based approach. Most existing adaptive FDSs that handle concept drift are using *evolving based approach* which includes Adaptive ensembles and Base model specific techniques. For instance, Fan (2004), Wang et al. (2003) and Dal Pozzolo et al. (2014) employed *evolving learning approach* under adaptive ensemble classifier technique for handling concept drift. Wang et al. (2003) designed an effective credit card fraud detection framework that is mining concept drifting data streams using weighted ensemble classifiers. They trained ensemble of classification models C4.5, RIPPER, naive Bayesian, etc., from sequential chunks of credit card data. Dal Pozzolo et al. (2014) proposed an adaptive model for credit card fraud detection system. The model addresses fluctuation or evolution as well as the behavior of the regular card holder or the fraudster. They proposed and compared three different approaches that handle the unbalanced problem in a changing environment static approach, update approach and forgetting approach. Fan (2004) proposed a detection framework based on a simple cross-validation decision tree ensemble method that handles concept drift. He applies this model to a number of areas; one of them is credit card fraud detection. In addition, evolving learning approach can be applied by using specific base model, in which the adaptively is achieved by managing specific model parameters or design. For instance, Wong et al. (2012) proposed adaptive credit card FDS based on Artificial immune systems (AIS). They used this method in anomaly detection, attempting to reduce credit card fraud. Phua et al. (2012) proposed resilient adaptive fraud detection for identity crime in the credit application. This model is based on whitelisting and detecting spikes of similar applications. Whitelisting uses real social relationships on a fixed set of attributes. Furthermore, Brause et al. (1999) and Paasch (2008) achieved the adaptability for data base of credit card transactions by using artificial neural networks (ANN). The evolution of architectures enables ANNs to adapt their topologies to cardholders drift without human intervention and thus provides an approach to automatic ANN design.

On the other hand, *regulated based approach* is adopted by Sherly and Nedunchezhian (2010) and Malekian and Hashemi (2013) that handles concept drift by monitoring two different profiles technique. Sherly and Nedunchezhian (2010) proposed adaptive FDS using both misuse and anomaly detection methods with incremental BOAT algorithm to find out fraud and changing spending behavior of legitimate user. Meanwhile, Malekian and Hashemi (2013) addressed the general problem of concept drift management in a profile based credit card fraud detection system by using two profiles, temporal profile and initial profile.

B. Skewed distribution issue (Imbalance class)

Fraudulent credit card transactions have very small percentage of total number of the transactions, which might cause obstructions to the performance of credit card FDS. Particularly, in credit card system, the misclassification of legitimate transaction is causing customer dissatisfaction which is considered more harmful than fraud itself. As we pointed out in Section 5.2, *algorithmic level* and *data level* are approaches used for dealing with imbalance class issue. In *data level approach*, Ryman-tubb and Krause (2011), Sahin and Duman (2011), Bhattacharyya et al. (2011), Phua et al. (2012) and Duman et al. (2013) utilize undersampling for skewed class for credit card fraud detection system. Meanwhile, Stolfo et al. (1997) employed oversampling approached over fraud transaction in pre-processing phase of

credit card fraud detection system.

On the other hand, *algorithmic level approach* is adopted through the use of cost-sensitive learning method or use learner itself to handle skewed distribution. For example, Sahin et al. (2013) used cost-sensitive classifiers in addressing the class imbalance problem. In addition, Dorronsoro et al. (1997) used nonlinear discriminant analysis (NLDA) neural models to deal with the imbalance class issue. Ju and Lu (2011) adopted improved Imbalance Class weighted support vector machine (ICW-SVM) to treat imbalance credit card transactions data. Bentley et al. (2000) proposed a neural fraud detection system with fraud density map to improve the detection efficiency in context of biased train data due to skewed distribution of data. Pozzolo et al. (2013) proposed to use a racing method to select the most appropriate strategy for a given unbalanced task adaptively. To obtain a high detection rate from imbalance input data, Chen (2006) using a binary support vector system (BSVS) and genetic algorithm (GA). Minegishi and Niimi (2011) proposed the construction of a Very Fast Decision Tree (VFDT) learner that can be applied to imbalanced distribution data streams. Seeja and Zareapoor (2014) proposed FraudMiner model with the ability to handle class imbalance by entering the unbalanced data directly to classifier. Their models present a better result compared to the oversampling technique.

C. Reduce large amount of data

As described earlier, reduced data strongly affect the performance of FDS. It is important in the credit card system because it reduces transaction processing time as well as the complexity in processing a transaction. Credit card attributes are used to decide the cardholders consuming habits which are correlated dramatically with cardholders' characteristics. On average, there are around 25 attributes found in a credit card, for example Customer income, Customer age, Customer profession, Number of cards used, Credit card type, Credit grade, Credit line, Book balance, Times of using card, Times of overdraft, Time bracket, Times of overdraft, Times of bad debt, Times of overdraft but not bad debt, Card using frequency, Overdraft rate, Growth rate of shopping, Average daily spending etc. To solve this, the use of data reduction approach, namely *dimensionality reduction* and *numerosity reduction* as presented in Section 5.3 is important.

Numerosity reduction approach as in Whitrow et al. (2008), Krivko (2010), Jha et al. (2012) and Dal Pozzolo et al. (2014) utilize aggregation, which is a nonparametric method for aggregating credit card transactions to capture consumer buying behavior prior to each transaction and uses these aggregations for model estimation to identify fraudulent transactions.

On the other hand, *dimensionality reduction approach* was adopted in credit card FDS through the application of Principal Component Analysis (PCA) (Ju and Lu, 2011) in order to reduce the dimension of credit card training dataset. Furthermore, Sherly and Nedunchezhian (2010) used embedded method for selecting a relevant features to be used in the model, whereas this method is performed as part of ID3 decision tree. Meanwhile, Paasch (2008) and Lei (2012) implemented wrapper approach where subsets of features are found by employing genetic algorithm for global search which find a subset heuristically.

D. Supports real time detection

This section will review the different algorithms used by researchers to support credit card FDS with efficient online detection. The algorithms are selected based on their capability to speed up the online processing of FDS. For example, Kundu et al. (2009) proposed Hybridization of BLAST and SSAHA methods, which are resulted in increasing the speed of processing, which enables online detection of credit card fraud.

Sherly and Nedunchezhian (2010) developed online credit card fraud detection using BOAT (Bootstrapped Optimistic Algorithm for Tree Construction) algorithm which reduce training time. Quah and Sriganesh (2008) introduced real time credit card fraud detection using self-organization map (SOM) to decipher, filter and analyse customer behavior for detection of fraud. SOM can filter out the number of transactions that need to be sent for review, therefore reducing the overall processing time, cost and complexity. Minegishi and Niimi (2011) proposed credit card fraud detection using online type decision tree construction and verification of generality, known as Very Fast Decision Tree (VFDT).

6.2. Telecommunication fraud

Telecommunication fraud is a problem that has grown dramatically over the past 10 years (Tawashi, 2010). Fraud in mobile telecommunications is a complex and dynamic problem for telecommunication operators. This is because these frauds threaten both the prepaid and post-paid services. In addition, fraud can be committed to fixed and mobile telephone lines (Held et al., 2001). Fixed line fraud is committed against telephone companies; this as fraudster gain access to switchboard and sell other people ability to make calls through the switchboard (Action Fraud, 2015). Mobile fraud is unauthorized use, tampering or manipulation of a cellular phone or service. Generally, the main goal behind commit fraud in both types of telecommunication (fixed line, mobile line) is to gain services and calls by illegal ways (Held et al., 2001). Based on global fraud loss survey announced by Communications Fraud Control Association (CFCA), in 2013, a fraud loss was recorded at $46.3 billion USD, a 15% increase from 2011. As a percent of global telecom revenues , fraud losses are approximately 2.09% —a 0.21% increase from 2011(CFCA, 2013). This is due to the large number of telecommunication fraud recorded from different categories.

Gossett and Hyland (1999) grouped telecommunication fraud into four categories:

Contractual fraud: fraudster uses telecom services with no intention to pay the service charge; for example subscription fraud and Premium Rate fraud.

Hacking fraud: fraudsters in this category breached the systems of business and take advantage of available resources illegally. Examples of such fraud are PABX (Private Automatic Branch Exchange) fraud and network attack.

Technical fraud: fraudsters in this category capitalize on weaknesses exist in mobile system technology. Such fraud needs high technical knowledge. Examples of such fraud are Cloning and Technical Internal fraud.

Procedural fraud: frauds under this group involved attacks against the procedures implemented to reduce the risk of exposure to fraud, and often attack the weaknesses in the business procedures used to grant access to the system. Examples of such fraud are Roaming fraud, Voucher ID duplication, and Faulty vouchers.

On the other hand, Cortesão et al. (2005) classifies telecommunication fraud according to three areas which are:

Motive: the main reason behind commit fraud.

Means: the nature or form of the fraud used to satisfy the motive.

Methods: the facilities and tools that are used to commit fraud.

There are many types of frauds that threaten the telecommunication sectors, which are considered as the most popular fraud area. It is estimated that more than 200 variants of telecom

Table 4
Types of telecommunication fraud.

Fraud type	Description
Subscription fraud	One of the most prevalent types of telecom frauds. The fraudsters tend subscribe telecom services by using false or fake personal information with no intention to pay the fees. In this type, the fraudsters could use the services for personal purpose or profit motivated to nonexistent business. (Kuşaksizoğlu, 2006; Rosas and Cesar, 2009; Farvaresh and Sepehri, 2011).
Superimposed fraud	Fraudsters take over a legitimate account and use services without the necessary authority for using it and it would appear as phantom calls in the bill (Bolton and Hand, 2002; Yusoff et al., 2013). There are several ways to carry out superimposed fraud, including mobile phone cloning and obtaining calling card authorization details (Laleh and Azgomi, 2009).
Premium Rate fraud	Offenders make large numbers of calls to premium rate service number from a subscriber's account without their knowledge (Van Heerden, 2005).
Roaming fraud	Perpetrator performs high value activities over visited public network with intention to escape the payment. In that time, the subscriber's home public network obliged to pay the charges to the victim visiting a public network (Macia-Fernandez et al., 2009).
Prepaid fraud	Use stolen or counterfeit card that offers non-existence services.
SIM Surfing	Fraudster steals services by robbing someone's SIM card or using it without owner knowledge. This kind of fraud could be permanent surfing or incidental surfing (Suksmono and Nugraha, 2006).
SIM Cloning fraud	Fraudsters duplicate someone's SIM card, and then the cloned card can be utilized by another person. The main reason behind this fraud is the bills of cloned card will be paid by original subscriber (Suksmono and Nugraha, 2006; Kuşaksizoğlu, 2006).
SIM BOX/Gateway Bypass fraud	SIM Boxes (GSM Gateways) are used illegally to bypass standard network interconnections in order to make traffic appear as local mobile calls (Active Fraud Eliminator, 1999).
Private Branch Exchange Hacking (PBX) fraud	Offenders illegally utilize subscriber's telephone lines and services, in order to make costly and numerous long-distance calls that will be paid by subscriber. This fraud also called Toll fraud (Fonebox Australia Group, 2011).
Voucher fraud	This fraud could take two different cases, the first case, the fraudsters attempt to guess the voucher numbers by submitting random numbers. The second case, fraudsters submit valid voucher numbers which already have been used before. However, all these unsuccessful attempts do not influence telecommunication company, but the numerous attempts will increase the system load and overall operational costs (Cortesão et al., 2005).

fraud exist in the telecommunications industry (Tsung et al., 2007). There are many literatures that address these frauds (Ghosh, 2010). Table 4 summarizes and lists the most studied telecom frauds by researchers.

Subscription fraud and superimposed fraud are the most prevalent types of telecom frauds. Therefore, they are addressing frequently in the literature compared to others types of telecommunication fraud.

6.2.1. Telecommunication FDS

Anomaly based fraud detection is usually used for telecommunication FDS. Each subscriber's extracted profile based on his/her CDR patterns are used to detect abnormal behaviors. These profiles are based either on CDR (e.g., number of calls, call duration, call type) or subscriber demographic properties (e.g., age, gender, region) or both. CDR is very useful in extracting user behavior (Farvaresh and Sepehri, 2011). Telecommunication FDS based on profile approach relies on a comparison of recent and long term behavior histories derived from the toll ticket data. If there is a significant change in the pattern, the alarms will be triggered (Held et al., 2001). Table 5 shows research done on the approaches and techniques used in telecommunication FDS.

As we can observe in Table 5, a hybridization of supervised techniques and unsupervised techniques have been used in order to obtain the best results. For example, Taniguchi et al. (1998)* first used a feed-forward neural network based on supervised learning to learn a discriminative function and classify subscribers by using summary statistics. Then, Gaussian mixture model is used to model the probability density of subscribers' past behavior so that the probability of current behavior can be calculated to detect any abnormalities from past behavior. Lastly, Bayesian networks are used to describe the statistics of a particular user and the statistics of different fraud scenarios. Aside from that, there are cases where multiple supervised techniques have been combined together, such as research by Held et al. (2001)* that used fuzzy rules and neural networks.

6.2.2. Telecommunication FDS issues and challenges

A. Concept drift

Customer call data are categorized permanently under concept drift due to many reasons, one of them being customers change their behaviors due to the introduction of new services introduced by telecommunication companies. The consequences of ignoring concept drift when mining for classification models can be catastrophic (Black and Hickey, 2002). In telecommunication area, the most current adaptive FDSs are based on evolving learner approach to detect drift; specially using base model specific technique. For example, the study by Fawcett and Provost (1997) is considered one of the earliest study that addresses adaptive FDS. Proposed adaptive fraud detection system has been applied to the problem of detecting cellular cloning fraud based on a database of call records. This paper presents a framework that has been used the rule based method. Which it works through generating the indicators (features) to use automatically by monitors (models) in order to detect a fraud and launch alarms. Sanver and Karahoca (2009) designed adaptive fraud detection model by utilizing an adaptive neuro-fuzzy inference system in mobile telecommunication networks. The proposed model achieved high precision of fuzzy based classification system and adaptability (backpropagation) property of neural networks in classification of data. Akhter and Ahamad (2012) developed adaptive telecommunication fraud detection model by using combination of neural, rule based and case based technology. Saravanan et al. (2014) proposed a model using Naïve–Bayesian classification to calculate the probability and an adapted version of KL-divergence to identify the significant difference between a normal user and a suspected user on the basis of subscription.

On the other hand, Jiang et al. (2007) and Tsung et al. (2007) built regulated based adaptive telecommunication FDS that detects drift through the use of statistical process control SPC technique. The primary advantage of using this technique is it has low resources complexity (time and memory) compared to other regulated learner approach in detecting concept drift.

Table 5
Research on approaches and techniques in telecommunication FDS.

Detection type	Detection tool	Learning approach	Data mining categories	Techniques	References
Anomaly based intrusion detection methods	**Data mining**	**Supervised**	Classification	Decision trees	Black and Hickey, 2002; Hilas and Sahalos, 2007
				Artificial neural networks	Moreau et al., 1997; Burge et al., 1997; Taniguchi et al., 1998*; Shawe-Taylor and Howkev, 1999; Held et al., 2001*; Boukerche et al., 2002; Hilas and Sahalos, 2006; Hilas and Mastorocostas, 2008*; Mohamed and Bandi et al., 2009; Mohamed and Fuad et al., 2009; Krenker et al., 2009; Qayyum et al., 2010; Akhter and Ahamad, 2012
				Hidden Markov model	Hollmen, 1999; Sun et al., 2006
				Artificial immune systems	Boukerche et al., 2004; Graaff, 2011
				Case-based reasoning	Almeida et al., 2008
				Bayesian classification	Saravanan et al., 2014; Taniguchi et al., 1998* Baharim et al., 2008
				Support vectors machine	Kim et al., 2003; Dong et al., 2004*
				Rule based	Murad and Pinkas, 1999; Moreau et al., 1996; Hilas, 2009*; Rosset et al., 1999
				Fuzzy neural network	Held et al., 2001*; Sanver and Karahoca, 2009; Qian et al., 2008
		Unsupervised	Clustering	Fuzzy logic	Bentley et al., 2000
				PCA	Hilas and Mastorocostas, 2008*; Dong et al., 2004*
				Distance based	Ferreira et al., 2007
				Gaussian mixture	Taniguchi et al., 1998*; Yusoff et al., 2013
				Hierarchical agglomerative clustering	Hilas and Mastorocostas, 2008*
	Statistical mining		Statistical method	Discriminant analysis	Olszewski, 2012
	Visual and data mining		Visualization	Data visualization	Cox, 1997

* A hybridization of supervised techniques and unsupervised techniques have been used in order to obtain the best results.

B. Skewed Class Distribution (imbalanced class)

The number of existing studies that handle imbalanced class in customers call data are quite small. For example, data level method has been used in Farvaresh and Mehdi (2011) by employing undersampling approach to balance between fraudulent and legitimate users. Meanwhile, Held et al. (2001) adopted oversampling approach. In order to avoid biasing the neural network towards the legitimate samples, the study made on the fraudulent samples were repeated three times.

C. Reduce large amount of data

There are several of benefits of dimensionality reduction for telecommunication data since its datasets have a large number of features. Usually, dataset in telecommunication area consists of Call Detail Records (CDR) which cover the customers numbers of caller and called, date and time when the call was made and call duration (Augustin et al., 2012). Normally, the total numbers of features in that record are around 56 features for each record. Hence, data reduction strategies have been performed widely for telecommunication FDS. For instance, in *numerosity reduction* approach, Hilas and Sahalos (2006) accumulate characteristics by applying weekly aggregation of the user's behavior. In addition, they implement *dimensionality reduction* through Principal Component Analysis (PCA) which is performed in order to transform the input vectors into uncorrelated ones. Also, Farvaresh and Sepehri (2011) employed PCA technique in pre-processing phase of FDS for detecting subscription fraud in telecommunication in order to reduce dimension.

Through the use of data compression technique under the same *dimensionality reduction* approach, Sun et al. (2006) operated optimal Lempel–Ziv (LZ) data compression technique to construct an end user's mobility profile for anomaly intrusion detection in wireless cellular networks. In this study, each user's itinerary was modeled as an *m*th-order Markov source. Meanwhile, Dong et al. (2004) proposed feature extraction method named GPCA based on IG (information gain) PCA to improve SVM accuracy and training time. Furthermore, Kim et al. (2003) used feature selection and extraction to discover indicators corresponding to changes in behavioral indicatives of fraud. Hilas (2009) performed feature extraction to dimension reduction by using classical time series analysis.

D. Supports real time detection

Supporting real time detection in telecommunication system is a big challenge. This is because normal profiles are usually extremely difficult to build due to the continuous mobility of end users (Sun et al., 2006). Therefore, researchers in this area aim to propose an FDS having the ability to respond quickly to user mobility and in the same time, qualified enough to detect fraudulent activities rapidly. Delay in the detection of telecommunication fraud might lead to intolerable losses and potential exploitation by fraudsters. As a matter of fact, there is no common approach to follow in order to apply the desired characteristics except using appropriate algorithms as presented in Section 5.4. Further, in telecommunication area, there is another method that supports real time detection using mobility management mechanism as accomplished by Sun et al. (2006) and Al-Fayoumi and Shilbayeh (2013). Meanwhile, Krenker et al. (2009) proposed a system for mobile-phone fraud detection based on a bidirectional artificial neural network (bi-ANN). The proposed system has the ability to detect fraud not only using offline processing, but also in real time. Furthermore, Hollmen (1999) proposed real time telecommunication FDS based on a stochastic generative mode and EM algorithm, which is trained in an incomplete data setting and is further refined with gradient-based discriminative training, which considerably improved the results.

6.3. Healthcare insurance fraud

Nowadays, healthcare insurance systems have become the main concern of modern life. These systems are working to support people with low income to pay the high costs of healthcare (Yang and Hwang, 2006). As a consequence, this system is being a target for fraudsters and busters. Global Fraud Study (2012) presents the financial losses for hundred countries due to healthcare fraud. For example, it is estimated that between $600 and $850 billion annually is lost to fraud, waste, and abuse in the US healthcare system, with $125 to $175 billion of this amount is due to fraudulent activity (Kelley, 2009). The healthcare system is so complex and confusing to most people. It consists of many rules and regulations. In addition, the system contains many parties such as physician, insurance company, insurance and health center. The complicated systems and regulations make it harder to discover fraudulent activities. There could be many kinds of frauds and they vary depending on their natures and positions in this system (Chen and Gangopadhyay, 2013). For instance, Sparrow (2000) classified healthcare insurance fraud types based on two different methods: hit-and-run and steal a little, all the time. In hit-and-run method, fraudsters simply submit many fraudulent claims in short time, receive payment, and disappear. While in steal a little, all the time method, fraudsters work to ensure fraud goes unnoticed and bill fraudulently over a long period of time. The most common known healthcare fraud types are listed in Table 6 (Kirlidog and Asuk, 2012; Thornton et al., 2013):

From these categories, based on statistics provided by Liu and Vasarhelyi (2013), Phantom claims fraud, Duplicate claims fraud and Kickbacks fraud were the most popular issue among researchers compared to other healthcare frauds.

6.3.1. Healthcare insurance fraud detection system

Raw data for healthcare fraud detection came mostly from insurance claims (e.g. information about the participation of an insurance subscriber and a service provider), general practitioners data (e.g. age, gender, etc.), or clinical-instance data (e.g. measuring blood pressure, examining respiration, and medicine treatment) (Liu and Vasarhelyi, 2013). Recently, research trend in the healthcare FDS uses zip or geolocation data of providers and their participations (Musal, 2010; Liu and Vasarhelyi, 2013). FDS of healthcare insurance is similar to the telecommunication and credit card, from the context of using supervised technique, unsupervised technique, or hybrid of unsupervised and supervised technique, as in Table 7.

As we can see from Table 7, healthcare fraud detection system is similar to telecommunication system in terms of using supervised approaches fundamentally for designing fraud detection models.

6.3.2. Healthcare fraud detection issue and challenges

A. Concept drift

In health care insurance area, the fraudsters tend to change their behavior of committing the fraud. On the other hand, the behavior of legitimate insured is changing due many health circumstances. This could lead to misclassification, which by classifying the legitimate insured as fraudster and vice versa. There are few studies that addressed concept drift in healthcare insurance FDS that evolve based on adaptive FDS. For example, Yamanishi et al. (2004) proposed SmartSifter as a program for online unsupervised outlier detection using finite mixtures with discounting learning algorithms, and the proposed model is adapted to non-stationary sources of data. Yang and Hwang (2006) designed a process-mining framework that utilizes the idea of clinical pathways to facilitate the automatic and systematic construction of an adaptable and extensible detection model. Lu et al. (2006) introduced an adaptive fraud detection method successfully identified actual fraudsters among real health and auto insurance data. They used Adaptive Benford's Law, which is a digital analysis technique combined with a reinforcement learning technique. The adaptive approach is based on deviations from the expected Benford's Law distributions as an indicator of anomalous behavior.

B. Reduce large amount of data

Healthcare systems are rich with amount of raw data since they come from many sources. That raw data are high dimensional in nature and could include hundreds of attributes. The mostly used raw data in health care fraud detection are: insurance claims data, e.g. participation of an insurance subscriber and a service provider; general practitioners data, e.g. information of service providers; and clinical instance data, e.g. patient treatment measuring blood pressure, examining respiratory, medicine treatment (Liu and Vasarhelyi, 2013). There are healthcare FDSs that employed all these data in their detection procedure, others use only some of the available raw data. Therefore, the number of features used for healthcare fraud detection can range from several to fifty. In the healthcare system, the features are usually identified manually by domain experts (Li et al., 2008; Ortega and Ruz, 2006). This article will focus only in the automatic approaches that are used to reduce data amount. In this case, *dimensionality reduction* approach has been used in several studies, for example, Chen and Gangopadhyay (2013) applied this approach by proposing dimension reduction method which is spectral clustering to reduce the dimension of healthcare data. The proposed spectral clustering method is able to analyze and discover the relationships between physicians and their references. Meanwhile, Ng et al. (2010) applied propositionalization approaches to spatio-temporal health data. This approach uses systematic feature construction/extraction

Table 6
Types of healthcare insurance fraud.

Fraud	Description
Phantom claims	The healthcare provider (healthcare center) presenting a bill to the healthcare insurer for services not provided.
Duplicate claims	The healthcare provider (health care center) presenting invoices to an insurer by using the same claims.
Bill padding	Submitting claims for unneeded ancillary services to Medicaid
Upcoding	Presenting claims whose reimbursement value more than the services provided or the insurance company will pay.
Unbundling	Presenting (reporting) excessive numbers of claims for different services that should be charged as one service.
Excessive or unnecessary Services	Introduced medical unneeded services for patient.
Kickbacks	Is colluding between provider and patient to take commission for illegal service.
Claims in short time	Reporting numbers of claims in for same insured in short time.
Unpaid installments	Reporting claims for insured has not paid any installments.
Incorrect dates	Reporting claims with incorrect dates that could be prior to or after than the beginning of the insurance period.
Medications without examination	Invoices for medications without the medical check-up or examination.
Excessive numbers of small bills	Excessive numbers of manual invoice demands whose amounts are smaller than the usual inspection limit.

Table 7
Research on approaches and techniques in healthcare insurance FDS.

Detection type	Detection tool	Learning approach	Data mining categories	Techniques	References
Anomaly based intrusion detection methods	**Data mining**	**Supervised**	Regression Classification	Regression analysis Decision trees Artificial neural networks Hidden Markov model Support vectors machine Rule based Fuzzy neural network	Musal, 2010* Mailloux et al., 2010 He et al., 1998*; Ortega and Ruz, 2006 Tang et al., 2011 Francis et al., 2011 Yang and Hwang, 2006; Shan et al., 2008; Tsai et al., 2014 Major and Riedinger, 2002
		Unsupervised	Clustering	Gaussian mixture Distance-based methods	Yamanishi et al., 2004 Konijn and Kowalczyk, 2011; Musal, 2010*
	Statistical data mining		Statistical method	Benford's law distributions	Lu and Boritz, 2005; Durtschi et al., 2004

* A hybridization of supervised techniques and unsupervised techniques have been used in order to obtain the best results.

techniques to transform multi-relational, multi-dimensional data into feature vectors and then plug them into well-understood and efficient algorithms (Kramer et al., 2000; Ng et al., 2010). Francis et al. (2011) determined a set of features that have provided optimal performance in their experiments. Yang (2003) applied filter method for feature selection by using Markov blanket filter.

C. Supports real time detection

Currently most fraud detection systems on healthcare insurance systems are static, lack real time detection and depend totally on human interaction to give final decision about suspicious claims (Travaille et al., 2011). Therefore, there are researchers that introduced solutions to speed up the detection time. Francis et al. (2011) employed SVM to provide a real world speed up method for medical fraud detection experts in their work. Tsai et al. (2014) also reduced the large labor cost and processing time of the existing medical insurance fraud detection systems by using a knowledge engineering methodology to analyse problems and construct knowledge model, including the domain schema and rules.

6.4. Automobile Insurance fraud

Automobile insurance is a contract between an automobile insurance company (insurer) and the car owner (insured) to cover all the billing in case of car damage, stolen or accident. Recently, insurance companies have suffered due to dishonest insured claims. Therefore, automobile insurance fraud become the main issue for companies and consumers (Artís et al., 2002). An automobile insurance system is similar to healthcare insurance system from the aspects of the parties involved in the fraud such as, drivers, chiropractors, garage mechanics, lawyers, police officers, insurance workers and others. Furthermore, automobile insurance fraud is being encountered in a variety of domains and it comes in many different types and sizes. For example, simple form when making cheat in the insurance claim; while sophisticated form when groups of individuals are collaborating in order to commit fraud (like make fake accident) (Šubelj et al., 2011). ACFE (2009) defined all schemes used to defraud the automobile insurance company; which is listed and concluded in Table 8 based on the most common and most addressed issues in the literature on automobile insurance fraud detection system.

Staged vehicle fraud and rental car fraud are considered the most prevalent fraudulent behavior in this area (Šubelj et al., 2011; Brockett et al., 1998).

6.4.1. Automobile insurance fraud detection system

The available literature on automobile insurance fraud detection is divided into theoretical literature in which claims are audited to deter fraud, such as in Crocker (2002) Dionne (2002) Lincoln (2003) Dionne et al. (2005) and Furlan et al. (2011), and statistical analysis literature of claims by using data mining techniques under the anomaly detection approach by recognizing the presence of suspicious behavior from insured claims as in Table 9 (Tennyson, 2001; Tennyson and Salsas-forn, 2008). Artís et al.

Table 8
Types of automobile insurance fraud.

Fraud	Description
Ditching	The perpetrators work to dispose of their vehicles to gain funding from insurers.
Past posting	The fraudsters try to get the compensation from insurers based on old accidents occurred before they obtain the insurance.
Vehicle repair	This scheme involves the billing of new parts on a vehicle when used parts were actually use as replacements.
Vehicle smuggling	The cheaters go to buy a vehicle with high price. The vehicle is put under an insurance policy to the maximum, with minimum deductible theft coverage. After that, the cheaters go to transfer the vehicle to another port and report the stolen case to collect money from the insurance company.
Phantom vehicles	Insurance issued for fake, non-existing vehicles.
Staged accidents	The fraudsters organize some kind of car accident to launch fake claims in order to obtain financial compensation of insurers.
Vehicle Identification Number (VIN) switch	The damaged vehicle is sold and a fraudulent claim is made while it is being repaired. Originally, the insurance beneficiary replaces the Vehicle Identification Number with stolen vehicle which has the same specifications. This means there are two different vehicles having the same insurance policy and one vehicle Identification Number.
Rental car fraud	A person does not need to own a vehicle to commit automobile fraud. There are several schemes that can be perpetrated using rental cars. The most prevalent involve property damage, bodily injury, and export fraud.

Table 9
Research on approaches and techniques in automobile insurance FDS.

Detection type	Detection tool	Learning approach	Data mining categories	Techniques	References
Anomaly based intrusion detection methods	**Data mining**	**Supervised**	Regression	Logistic models	Wen and Wang, 2005; Artís et al., 2002; Weisberg et al., 1998; Dionne, 1997; Bermúdez et al., 2008
			Classification	Decision trees	Pérez et al., 2005; Bhowmik, 2011
				Artificial neural networks	Xu et al., 2011; Viaene et al., 2005; Brockett and Golden, 2006
				Naïve Bayes	Bhowmik, 2011; Viaene et al., 2004
				Support vectors machine	Tao et al., 2012
				Fuzzy neural network	Derrig and Ostaszewski, 1995; Pathak et al., 2005
		Unsupervised	Clustering	Self-organizing map algorithm	Brockett et al., 1998
	Visual and data mining		Visualization	Data visualization	Šubelj et al., 2011

(2002) stated that the audit strategy is not perfect, and claims may be misclassified. Instead, the authors tend to exploit statistical analysis to detect fraud. Šubelj et al. (2011) utilized Iterative Assessment Algorithm (IAA), a social network that is used to detect fraud. Their methodologies do not need any labeled data. Brockett et al. (1998) used Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. Wen and Wang (2005) employed multinomial logit (MNL) and nested logit (NL) models to estimate the influence of the insured and claim characteristics on the probability of committing fraud. Artís et al. (2002) and Dionne (1997) proposed a logit model for fraud detection, taking into account misclassification of the type of claim. Weisberg and Derrig (1998) proposed a multiple linear regression model to select indicators of different types of fraud suspicions. Derrig and Ostaszewski (1995) used fuzzy set techniques to classify claims. Viaene et al. (2005) showed the explicative capabilities of neural network classifiers with automatic relevance determination weight regularization, and reported the findings from applying these networks for personal injury protection automobile insurance claim fraud detection. The automatic relevance determination function has played an important role in the success of this proposed model since feature selection is a big concern in automobile fraud detection system, as several authors also addressed this issue. We will go in depth on this discussion in Section 6.2.4 B. Brockett and Golden (2006) explored and compared two statistical methods and two artificial neural network methods for the prediction of financial hazard in life insurers. Tao et al. (2012) proposed a fuzzy support vector machine model with dual membership to handle the overlapping problem of insurance fraud samples. Bermúdez et al. (2008) used an asymmetric or skewed legit model using Bayesian analysis for fraud detection in the Spanish automobile insurance claims. Brockett et al. (2002) proposed statistical and a priori classification and principal components analysis of RIDIT score (PRIDIT) method to detect fraud in the automobile insurance industry. Pathak et al. (2005) introduced a fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims. Pinquet (2007) used a two-equation model (a bivariate probit model) for audit and fraud detection in automobile insurance. Viaene et al. (2004) applied AdaBoost Naïve Bayes scoring to insurance claims fraud.

6.4.2. Automobile insurance fraud detection system issues and challenges

Automobile insurance fraud area has specific characteristics as well as specific challenges. For example, real time support for detection challenge in the credit cards area does not exist in the automobile insurance area.

A. Skewed distribution

Only a small portion of accidents participate in fraudulent (skewed class distribution) cases making them extremely difficult to detect. Data level method has been used under the sampling approach as in Šubelj et al. (2011) and Pérez et al. (2005). Hofmann (2012) conducted a comprehensive survey and comparison of data level methods (SMOTE, undersampling, oversampling) against algorithms level techniques over automobile insurance fraud.

B. Reduce large amount of data

Ordinarily, automobile insurance can cover some or all of the following parties: the insured party, the insured vehicle, third parties involved (e.g. property damage and bodily injury), personal injury protection, bodily injury liability and so on. Automobile insurers claims data set is used for fraud detection which consists of thousands of claim observations with around 90 features that could be categorised into: Policy Information; Claim Information

e.g. Accident date, Report date and Type of injury; Outpatient Medical Provider Information e.g. Provider type and Amount billed; Attorney Information; Claim Handling Information e.g. Type of investigation and Result of Investigation (D'Arcy, 2005).

In automobile insurance FDS for data reduction, *dimensionality reduction* approach is mostly used. For instance, (Ai et al., 2009; Brockett et al., 2002) used Principal Component Analysis (PCA) for dimensionality reduction of a personal injury protection insurance claims data set from the Automobile Insurance Bureau. Viaene et al. (2005', 2004) utilized automatic relevance determination (ARD) weight regularization for personal injury protection automobile insurance claim fraud detection. The ARD objective function hyper parameter scheme provides a means for soft input selection as it allows the determination of which predictor variables are most informative. Xu et al. (2011) employed rough set reduction to generate a set of reductions which can keep the consistency of the automobile insurance dataset.

6.5. Online auction fraud

As E-commerce develops rapidly, online auction is also becoming more popular. Auction websites such as Yahoo and eBay have been growing at a considerable rate (Liaw et al., 2006). An online auction is one of the most popular profit Internet business models, because online auction activities are not constrained by time or physical store locations. Flexible use and vast profit gained from the activity attract offenders to defraud in order to cash on the lucrative online trading market. The Internet Complaint Center reported that online auction is one of the top two most dangerous Internet crimes in recent years, putting online auction participants at big risk (Chang and Chang, 2012). Online auction frauds have been classified into six categories by the Internet Fraud Complaint Centre (IFCC) which is: 1) non-delivery of goods, 2) misrepresentation of the items, 3) triangulation, 4) fee staking, 5) selling of black-market goods, 6) multiple bidding and shill bidding. Fraud categories number 5 and 6 can be termed as cheating (Jenamani et al., 2007). Dong et al. (2009) classified the types of online auction fraud based on time periods into three categories in which the fraudulent behavior can take place: pre-auction, in-auction and post-auction which is probably the most prevalent form of online auction fraud. Moreover, Chang and Chang (2014) categorized online auction fraud according to fraudster attitudes into four types: Aggressive, Classic, Luxury and Low-profiled. Cheating could be committed either by a bidder (buyer) or an auctioneer (seller). Table 10 concludes the common types of online auction fraud from the victim's viewpoint as presented in Chua and Wareham (2004) and Lee et al. (2010).

Based on the literature, we found that bid shielding fraud is most prevalent fraud type in online auction.

6.5.1. Online auction fraud detection system

Online auction fraud is increasing rapidly, therefore several detection schemes are used by researchers. Aleem and Antwi-Boasiako (2011) are grouping the schemes into three; feedback anomaly detection schemes, data mining schemes and agents based-trust management schemes.

Feedback anomaly detection schemes uses a reputation system which is an available countermeasure for buyers to evaluate a seller's credit. The mechanism is scoring the reputation of the trader by accumulating the feedbacks from trading partners. For example, positive feedbacks will increase the accumulated score by 1 while negative feedback lowers the accumulated score by 1 (Chang and Chang, 2011). Feedback scheme has been used extensively in the past for online auction fraud detection, but it is ineffective; it is easy to manipulate and create fake overrated reputations (Chau et al., 2006; Chang and Chang, 2011; Aleem and Antwi-Boasiako, 2011).

Data mining schemes are widely used now to detect online auction fraud. Generally, online auction FDS procedure consists of two basic steps: (1) construct features which extract user profiles and transaction histories of suspended accounts in order to discriminate between legitimate trader and fraudster; and (2) build a detection model based on these constructed features (Chang and Chang, 2012; Chang and Lee, 2012; Chau et al., 2006). Classical classification algorithm is adopted by researchers in building the detection model, especially decision trees (Shao et al., 2002). Table 11 lists data mining techniques used by both trends of research. Agent-based trust management schemes handle trust and identity problem by using multiple interacting intelligent agents (Ba et al., 2003; Jaiswal et al., 2004; Wang et al., 2004).

The table shows that most researchers use decision tree technique under the classification approach for building online auction FDS.

6.5.2. Online auction fraud detection system issues and challenges

A. Concept drift

Adaptive FDS that handles concept drift is frequently considered in many studies of online auction system. Adaptive FDS works in monitoring the behavior of a bidder's account to determine if he/she is a fraudster, and then updates detection model based on returned result. In online auction area, the concept drift issue is tackled by numerous authors. One of the adopted adaptive FDS method is based on regulated learner approach which uses the monitoring distribution of profiles technique. In their detection models, (Chang, 2009; Chang and Chang, 2010, 2011) schemed the profiling phased method of fraudster history where this method can detect fraud as early as possible. In addition, the profiling phased method shows possible patterns of behavior changes in successive phases by using specific clues (features) in the transaction histories of online auctions. Furthermore, their

Table 10
Types of online auction fraud.

Fraud type	Description
Competitive shilling	A corrupt seller performs agreement with bidders to place bids in auctions without having intention to win. The main goal of this fraud is to inflate the price of the item in order to increase profit of the seller (Yokoo et al., 2004; Tsang et al., 2014).
Bid shielding	Illegitimate bidding to preserve a low price.
Non-delivery of goods	Seller never sends the goods.
Multiple bidding	A bidder can place multiple bids on the same item using different aliases (Yokoo et al., 2004).
False bids	A seller cheats in a second-price sealed-bid auction by looking at the bids before the auction clears and submitting an extra bid just below the price of the highest bid. Such extra bids are often called false bids (Jenamani et al., 2007).
Bid shading	The bidder (buyer) bids on auction for an item with a price far below than the item is worth. This way, the bidder will win the auction at a minimum price (Ford et al., 2012; Johnson, 2012).
Credit card phantom	Fake transactions for illegal loan sharking through illegal conspiracy between the seller and buyer.

Table 11
Research on approaches and techniques in online auction FDS.

Detection type	Detection tool	Learning approach	Data mining categories	Techniques	References
Anomaly based intrusion detection methods	**Data mining**	**Supervised**	Classification	Logic regression	Dong et al., 2012*; Maranzato et al., 2010
				Decision trees	Chang and Lee, 2012; Chau et al., 2006; Chau and Faloutsos, 2005; Ku et al., 2007* Shao et al., 2002; Ochaeta, 2008; Tsang et al., 2014*; Almendra, 2013; Lin et al., 2012
				Artificial neural networks	Dong et al., 2012*; Ford et al., 2012; Tsang et al., 2014
				k-Nearest-Neighbour classifier	Chang and Chang, 2012
				Bayesian classification	Goel et al., 2010
				Support vectors machine	Ochaeta 2008*; Almendra and Enachescu, 2011
				Association rule analysis	Shah et al., 2003
		Unsupervised	Clustering	Clustering graph and network data (social networkAnalysis.)	Ku et al., 2007*
				k-means	Bapna et al., 2004; Chang and Chang, 2014
				Hierarchical agglomerative clustering	Hou and Rego, 2007

* A hybridization of supervised techniques and unsupervised techniques have been used in order to obtain the best results.

proposed models have the capability to deal with inconsistent features of fraudster's problem by hybrid phased models developed in which every single account contains all fraudster features in all various predefined phases.

In contrast, Xu et al. (2009), Ford et al. (2012) and Tsang et al. (2014) proposed evolving based adaptive FDSs for online auction that rely on base learner techniques. Xu et al. (2009) proposed a real time model checking method for detecting abnormal bidding behaviors. This model will be updated as the auction state changes with new bids and time. Ford et al. (2012) designed framework of real time self-adaptive classifier for identifying suspicious bidders in online auctions by using an incremental neural network approach. Tsang et al. (2014) mitigated the changes in fraudulent behavior issue by proposing a shill detection framework that uses agent to generate synthetic data set of arbitrary types of fraud. Then, the generated data are passed to supervised methods of neural network and decision trees to detect shilling fraud.

B. Skewed distribution

In auction data, skewed distribution exists when the number of legitimate actions exceeds the number of fraudulent actions, In auction data, skewed distribution exists when the number of legitimate actions exceeds the number of fraudulent actions; that will make the classification accuracy of users to fraudulent or legitimate is extremely weak. Chau and Faloutsos (2005), Chang and Chang (2011), Almendra (2013) and Tsang et al. (2014) applied undersampling balanced approach for online auction fraud detection system.

C. Large amount of data

Reducing data in online auction systems is a significant task, because the auction data come from different source such as data of bidder account, bidder transactions, network level data, or feedback reputation system data (Tsang et al., 2014). The data that come from all resources are extremely valuable and need to be included in the detection process. These data include features that represent bidder account like Auction Count, Reputation, Bid Amount, Excess Bid and so forth (Tsang et al., 2012). Meanwhile, feedback reputation system contains rating scale from 1 to 5, or using several measures (friendliness, prompt response, quality product, etc.), or averaging rather than totalling feedback scores (Resnick et al., 2000).

Therefore, researchers in this area give their utmost effort to reduce data through the use of different *dimensionality reduction or numerosity reduction* approach. For example, Liu et al. (2010) aggregated third parties' feedbacks on the sellers. Meanwhile, Almendra (2013) used aggregation to obtain product categories for several years.

Tsang et al. (2014) improved the detection model performance by selecting the features based on the results from correlation analysis and Principal Component Analysis (PCA) to discover which attributes were redundant, or which values were similar across all users. Almendra and Enachescu (2011) reduced the number of features by implementing forward selection procedure using an SVM classifier with a linear kernel. Chau and Faloutsos (2005), and Chang and Chang (2010, 2012) defined relevant features by using filter approach which provides a measurement of the features usefulness in discriminating the fraudulent and legitimate classes.

D. Supports real time detection

In online auction, instantaneous support and detection is extremely important (Chang and Chang, 2012). In this context, Ford et al. (2012) presented a real time self-adaptive classifier framework which is fast enough to be utilized in a real time environment. In the same area, Xu et al. (2009) designed a technique to detect shilling behaviors in live online auctions.

In online auctions, the earliness of detection is supported by Chang (2009) and Chang and Chang (2011, 2012) through proposed novel two stage phased modeling framework that integrates hybrid phased models with a successive filtering procedure to identify latent fraudsters by examining the phased features of potential fraudsters' lifecycles. The main idea behind this method is phasing the transaction history of trader to detect latent fraud. In this method, 100% phase indicates the full transaction history of trader, and 80% phase indicates 80% transaction history of the trader. Meanwhile, M (100%) is a detection model learned by 100% Phase. They argued that in M (100%), fraud transaction history can be easily identified, but may not identify a potential fraudster during the latency period. Therefore, they tend to use other phased detection models built in earlier phases. If the phased model matched a suspect account's behavior, then a fraud alarm should be triggered. In online auction, Chiu et al. (2011) also presented a feasible method to detect fraudulent accounts using social network analytical metrics and data-mining approaches. This method is effective and works almost instantaneously with an acceptable classification accuracy rate.

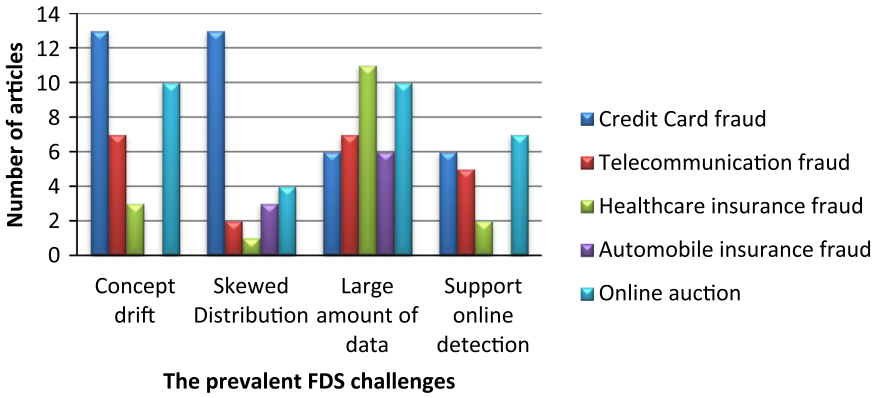

Fig. 11. The prevalent FDS challenges over studied fraud areas.

7. Discussion and analysis

In previous sections, we addressed the 5 areas of frauds (credit card fraud, medical insurance fraud, automobile insurance fraud, online auction fraud and telecommunication fraud), while in each domain, we highlighted the most common fraud types. Further, we defined the detection approaches and techniques. In addition, the existing challenges (concept drift, supports real time detection, skewed distribution, and large amount of data) for each domain and how they have treated were presented.

As FDS faced many challenges in every area, authors are trying to alleviate these challenges in their proposed FDS by using appropriate methods and approaches. From previous sections, it is obvious that FDS challenges have been treated efficiently. This is due to the characteristics of the work in those areas of fraud, for instance, concept drift challenge has made severe impact on real time (online) services such as credit card, telecommunication and online auction areas. In contrast, healthcare insurance area and automobile area do not suffer much from this issue because they have slight dynamism and work offline. Fig. 11 illustrates our observation on the relationship between FDS challenges and fraud areas using number of studies made to handle such challenges. As evident, concept drift issue received great attention in credit card fraud and online auction fraud areas. The same case applies for handling skewed distribution issue, this matter was covered extensively in credit card fraud area compared to other fraud domains. On the other hand, large amount of data issue has gained considerable concern by researchers from all areas of fraud particularly in health care fraud and online auction fraud. Lastly, supports real time detection issue obtained a fair portion of consideration in online auction fraud, credit card fraud and telecommunication fraud. This is due to online credit card and online auction areas require immediate fraud detection, thus numerous studies try to come up with FDS that support real time detection efficiently. Also, from Fig. 11, we see that credit card area is the most researched area for all types of fraud among the five areas selected.

It should be highlighted that, issues and challenges faced by the fraud detection system are not limited to those that have been highlighted in this paper. There are also other challenges that obstruct the performance of FDS, for example, overlapping class challenge, noisy data challenge and misclassification cost-issue. These challenges have negative impacts to FDS, such as creating improper and complex detection model with weak predictive accuracy.

- Overlapping data
  Overlapping data is a phenomenon that occurs when the fraudulent transaction looks as legitimate transaction and vice versa. This is because the fraudsters try every time to make their transaction very similar to normal transaction in order to mislead the FDS and reduce the detection rate (Kim and Kim, 2002; Travaille et al., 2011). Therefore, the treatment for overlapping data should be considered when designing FDS; that might be achieved by choosing suitable classifier and proper option as (Chen et al., 2005) in credit card fraud detection.
- Noisy data
  Fraud data set tend to be noisy, incomplete and inconsistent. Therefore, most researchers attempt to apply data cleansing routines to smooth out noisy data, complete missing values and correct inconsistencies in the data (Kamber and Pei, 2012). Since filling up missing data is considered as a kind of treatment for a noisy data, many researchers handled noisy data issue in the preprocessing stage before the detection process. In fraud detection system, there are several data smoothing techniques available to filter out the noisy data. This type of data will negatively affect the effectiveness and efficiency of a classifier as well as reducing its predictive accuracy (Philip and Sherly, 2012). There is another approach used to treat noisy data which is human inspection to detect suspicious values (Yang and Hwang, 2006). In telecommunication area, Baharim et al. (2008) has presented the leveraging missing values method using the Naïve–Bayes approach posterior to rule-based classifier to analyse the probability of corrupted and missing values in CDR, which then consequently led to the discovery of usable record that lies in rejected CDR. Lu and Boritz (2005) took into account the missing data over the proposed fraud detection model in healthcare area. They utilised an Adaptive Benford algorithm which tunes the distribution of digit frequencies to account for any missing data cut-off and produces a threshold cut-off for various ranges of digits.
- Misclassification cost

Misclassification cost is the error costs of false positive (genuine transaction identified as fraudulent transaction) and false negative (fraudulent transactions not identified). Misclassification costs (false positive and false negative error costs) are unequal, uncertain, can differ from instance to instance, and can change over time. In fraud detection, a false negative error is usually more costly than a false positive error (Phua et al., 2005). In the FDS literature, the misclassification cost-issue is tackled through reducing the costs of losses by maximizing the percentage of correct classification of fraudulent transactions while minimizing the false ones. For example, in credit card area, Sahin et al. (2013) designed a new cost-sensitive decision tree approach which minimizes the sum of misclassification costs. In the same area, Duman and Ozcelik (2011) handled misclassification problem by combined metaheuristic approaches, namely the genetic algorithms and scatter

search. In the automobile insurance fraud detection area, Viaene et al. (2007) used cost-sensitive approach to analyse the effects of taking into account information on damages and audit costs early on in the screening process.

8. Conclusion

Fraud cases have increased in recent years, particularly in important and sensitive technical areas. Hence, there is a dire need to combat fraud. Fraud prevention and detection are the proper protection mechanism against fraud. Fraud prevention alone is not sufficient. Fraud detection is proposed to protect vital services in the technical systems. This survey article has explored the state-of-the-art fraud detection systems in five areas of fraud. Furthermore, the fraud detection approaches and techniques have been categorized and reviewed. Which it is noticed that most fraud detection systems in all areas use supervised approach. In addition, the most commonly used fraud detection technique is artificial neural networks (ANN), support vector machines (SVM), rule-induction techniques, decision trees, logistic regression, and meta-heuristics such as genetic algorithms. These techniques can be used alone or combined with an ensemble or meta-learning techniques to build strong detection classifiers. Next, the challenges that hinder the performance and efficiency of fraud detection systems were discussed. It is observed that each fraud area has its specific characteristics and faces different challenges. For example, supports real time detection issue that is crucial in the credit cards area is not important for the Automobile insurance area. Lastly, another set of challenges that obstruct the performance of FDS are highlighted, namely, overlapping class, noisy data and misclassification cost-issues. The impacts of these challenges to FDS are improper model construction and extremely complex detection model with weak predictive accuracy.

References

Index. In data mining. In: Kamber, Jiawei Han Micheline, Pei, Jian (Eds.), The Morgan Kaufmann Series in Data Management Systems, Third ed. Morgan Kaufmann, Boston, pp. 673–703. http://dx.doi.org/10.1016/B978-0-12-381479-1.00022-8.

Abbass, Hussein A., Bacardit, Jaume, Butz, Martin V., Llor, Xavier, 2004. Online Adaptation in Learning Classifier Systems: Stream Data Mining, 217.

Action Fraud, 2015. UK's National Fraud and Internet Crime Reporting Centre. 1, pp. 1689–1699. doi: http://dx.doi.org/10.1017/CBO9781107415324.004.

Active Fraud Eliminator, 1999. Active Fraud Eliminator.

Ai, Jing, Brockett, Patrick L., Golden, Linda L., 2009. Assessing consumer fraud risk in insurance claims: an unsupervised learning technique using discrete and continuous predictor variables. August 2014, pp. 37–41. doi: http://dx.10.1080/10920277.2009.10597568.

Akhilomen, John, 2013. Data mining application for cyber credit-card fraud detection system. In Lecture Notes in Engineering and Computer Science. pp. 1537–1542.

Akhter, Mohammad Iquebal, Ahamad, Mohammad Gulam, 2012. Detecting telecommunication fraud using neural networks through data mining. Int. J. Sci. Eng. Res. 3 (3), 1–5.

Aleem, Azeem, Antwi-Boasiako, Albert, 2011. Internet auction fraud: the evolving nature of online auctions criminality and the mitigating framework to address the threat. Int. J. Law Crime Justice 39 (3), 140–160. http://dx.doi.org/10.1016/j.ijlcj.2011.05.003.

Alexopoulos, Panos, Kafentzis, Kostas, Benetou, Xanthi, Tagaris, Tassos, 2007. Towards a generic fraud ontology in e-government detection in the e.

Al-Fayoumi, Mustafa A., Shilbayeh, Nidal F., 2013. Cloning sim cards usability reduction in mobile networks. J. Netw. Syst. Manag. 22 (2), 259–279. http://dx.doi.org/10.1007/s10922-013-9299-8.

Allan, Tareq, Zhan, Justin, Dako, South, 2010. Towards Fraud Detection Methodologies e.

Allen, Julia, 2000. State of the Practice of Intrusion Detection Technologies. January.

Almeida, Miguel Pironet San-bento, 2009. Classification for Fraud Detection with Social Network Analysis (Miguel Pironet San-Bento Almeida Dissertation for the obtaining of a Masters Degree in Engenharia Informática e de Computadores Júri).

Almeida, Pedro, Jorge, Marco, Cortesão, Luís, Martins, Filipe, Vieira, Marco, Gomes, Paulo, 2008. Supporting Fraud Analysis in Mobile* Telecommunications Using Case-Based Reasoning. pp. 562–572.

Almendra, V., 2013. Finding the needle: a risk-based ranking of product listings at online auction sites for non-delivery fraud prediction. Expert Syst. Appl. 40 (12), 4805–4811. http://dx.doi.org/10.1016/j.eswa.2013.02.027.

Artís, Manuel, Ayuso, Mercedes, Guillén, Montserrat, 2002. Detection of automobile insurance fraud with discrete choice models and misclassified claims. J. Risk Insur. 69 (3), 325–340. http://dx.doi.org/10.2307/1558681.

Association of Certified Fraud Examiners, 2002. Report to The Nations on Occupational Fraud And Abuse.

Association of Certified Fraud Examiners, 2009. Insurance fraud handbook.

Augustin, Simon, et al., 2012. Telephony fraud detection in next generation networks. In: Proceedings of the AICT 2012 – 8th Advanced International Conference on Telecommunications. pp. 203–207.

Ba, Sulin, Whinston, Andrew B., Zhang, Han, 2003. Building trust in online auction markets through an economic incentive mechanism. Decis. Support Syst. 35 (3), 273–286. http://dx.doi.org/10.1016/S0167-9236(02)00074-X.

Baharim, Khairul Nizam, Kamaruddin, Mohd. Shafri, Jusof, Faeizah, 2008. Leveraging missing values in call detail record using naïve bayes for fraud analysis. In: Proceedings of the 2008 International Conference on Information Networking. January, pp. 1–5. http://dx.doi.org/10.1109/ICOIN.2008.4472791.

Bapna, Ravi, Paulo Goes, Alok Gupta, Yiwei Jin. 2004. User Heterogeneity and Its Impact on Electronic Auction Market Design: An Empirical Exploration. MIS Quarterly 28 (1). Management Information Systems Research Center, University of Minnesota: 21–43. http://www.jstor.org/stable/25148623.

Behdad, Mohammad, Barone, Luigi, Bennamoun, Mohammed, French, Tim, 2012. Nature-inspired techniques in the context of fraud detection. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) 42 (6), 1273–1290. http://dx.doi.org/10.1109/TSMCC.2012.2215851.

Belo, Orlando, Vieira, Carlos, 2011. Applying User Signatures on Fraud Detection in Telecommunications Networks. pp. 286–299.

Bentley, Peter J., Kim, Jungwon, Jung, Gil-ho, Choi, Jong-uk, 2000. Fuzzy darwinian detection of credit card fraud. In: Proceedings of the 14th Annual Fall Symposium of the Korean Information Processing Society. 1–4.

Bermúdez, Ll., Pérez, J.M., Ayuso, M., Gómez, E., Vázquez, F.J., 2008. A Bayesian dichotomous model with asymmetric link for fraud in insurance. Insur. Math. Econ. 42 (2), 779–786. http://dx.doi.org/10.1016/j.insmatheco.2007.08.002.

Bhattacharyya, Siddhartha, Jha, Sanjeev, Tharakunnel, Kurian, Westland, J. Christopher, 2011. Data mining for credit card fraud: a comparative study. Decis. Support Syst. 50 (3), 602–613. http://dx.doi.org/10.1016/j.dss.2010.08.008.

Bhowmik, Rekha, 2011. Detecting auto insurance fraud by data mining techniques. J. Emerg. Trends Comput. Inf. Sci. 2 (4), 156–162.

Black, Michaela, Hickey, Ray, 2002. Classification of customer call data in the presence of concept drift and noise. pp. 74–87.

Bolton, Richard J., Hand, David J., 2002. Statistical fraud detection : a review. Stat. Sci. 17 (3), 235–255.

Bolton, Richard J., Hand, David J., 2001. Unsupervised profiling methods for fraud detection. In: Proceedings of Credit Scoring Credit Control. pp. 235–255.

Boukerche, Azzedine, Sechi, Mirela, Notare, M. Annoni, 2002. Behavior-based intrusion detection in mobile phone systems. J. Parallel Distrib. Comput., 1476–1490. http://dx.doi.org/10.1006/jpdc.2002.1857.

Boukerche, Azzedine, Regina, Kathia, Juc, Lemos, Sobral, Bosco, Sechi, Mirela, Annoni, Moretti, 2004. An artificial immune based intrusion detection model for computer and telecommunication systems. Parallel Comput. 30, 629–646. http://dx.doi.org/10.1016/j.parco.2003.12.008.

Brabazon, Anthony, Cahill, Jane, Keenan, Peter, Walsh, Daniel, 2010. Identifying online credit card fraud using artificial immune systems.

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. http://dx.doi.org/10.1109/TAI.1999.809773.

Brause, R., Langsdorf, T., Hepp, M., 1999. Credit Card Fraud Detection by Adaptive Neural Data Mining.

Brockett, Patrick L., Golden, Linda L., 2006. A comparison of neural network, statistical methods, and variable choice for life insurers' financial distress prediction. J. Risk Insur. 73 (3), 397–419.

Brockett, Patrick L., Xia, Xiaohua, Derrig, Richard a, 1998. Using Kohonen's self-organizing feature map to uncover automobile bodily injury claims fraud. J. Risk Insur. 65 (2), 245. http://dx.doi.org/10.2307/253535.

Brockett, Patrick L., Derrig, Richard A., Golden, Linda L., Alpert, Mark, 2002. Analysis of RIDITs. 69, 3, pp. 341–371.

Burge, P., et al., 1997. Fraud detection and management in mobile telecommunications networks. Burge, P., Shawe-Taylor, J., Cooke, C. 437, pp. 28–30.

CFCA, 2013. Communications Fraud Control Association (CFCA) announces results of worldwide telecom fraud survey. Communications Fraud Control Association. pp. 0–1.

Chan, Philip K., 1999. Loss due, to illegitimate, behavior in, and many cases. Distributed data mining in credit card fraud detection.

Chandola, Varun, Banerjee, Arindam, Kumar, Vipin, 2009. Anomaly detection : a survey. ACM Comput. Surv. 41 (3), 1–58. http://dx.doi.org/10.1145/1541880.1541882.

Chang, Jau-Shien, Chang, Wen-Hsi, 2014. Analysis of fraudulent behavior strategies in online auctions for detecting latent fraudsters. Electron. Commer. Res. Appl. 13 (2), 79–97. http://dx.doi.org/10.1016/j.elerap.2013.10.004.

Chang, Wen Hsi, Chang, Jau Shien, 2011. A novel two-stage phased modeling framework for early fraud detection in online auctions. Expert Syst. Appl. 38 (9),

11244–11260. http://dx.doi.org/10.1016/j.eswa.2011.02.172.
Chang, Wen-Hsi, Chang, Jau-Shien, 2012. An effective early fraud detection method for online auctions. Electron. Commer. Res. Appl. 11 (4), 346–360. http://dx.doi.org/10.1016/j.elerap.2012.02.005.
Chang, Jau-shien, 2009. An Early Fraud Detection Mechanism for Online Auctions Based on Phased Modeling 3. Phased Modeling Framework for Early. pp. 743–748.
Chang, Jau-shien, Lee, Ching-fen, 2012. Cost-effective online auction fraud detection by genetic feature selection.
Chang, Wen-Hsi, Chang, Jau-Shien, 2010. A multiple-phased modeling method to identify potential fraudsters in online auctions. In: Proceedings of the 2010 Second International Conference Computer Research Development. pp. 186–190. http://dx.doi.org/10.1109/ICCRD.2010.50.
Chau, Duen Horng, Faloutsos, Christos, 2005. Fraud Detection in Electronic Auction. In: Proceedings of European web mining forum, EWMF 2005, ECML/PKDD.
Chau, Duen Horng, Pandit, Shashank, Faloutsos, Christos, 2006. Detecting Fraudulent Personalities in Networks of Online Auctioneers. pp. 103–114.
Chaudhary, Khyati, Yadav, Jyoti, 2012. A review of fraud detection techniques: credit card. Int. J. Comput. Appl. 45 (1), 39–44.
Chawla, Nitesh V., Bowyer, Kevin W., Hall, Lawrence O., Philip Kegelmeyer, W., 2002. SMOTE : synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.
Chen, Rong-Chang, 2006. A new binary support vector system for increasing detection rate of credit card fraud. Int. J. Pattern Recognit. Artif. Intell. 20 (2), 227–239.
Chen, Rong-Chang, Chiu, Ming-Li, Huang, Ya-Li, Chen, Lin-Ti, 2004. Detecting credit card fraud by using questionnaire-responded transaction model based on support vector machines. In: Proceedings of IDEAL. pp. 800–806.
Chen, Rong-Chang, Luol, Shu-Ting, Lee, Vincent C.S., 2005. Personalized approach based on SVM and ANN for detecting credit card fraud. In: Proceedings of the IEEE International Conference on Neural Networks and Brain. Beijing, China. pp. 810–815.
Chen, Song, Gangopadhyay, Aryya, 2013. A novel approach to uncover health care frauds through spectral analysis. In: Proceedings of the 2013 IEEE International Conference on Healthc. Informatics. September, pp. 499–504. http://dx.doi.org/10.1109/ICHI.2013.77.
Chiu, Chaochang, Ku, Yungchang, Lie, Ting, Chen, Yuchi, 2011. Internet auction fraud detection using social network analysis and classification tree approaches. Int. J. Electron. Commer. 15 (3), 123–147. http://dx.doi.org/10.2753/JEC1086-4415150306.
Chua, C.E.H., Wareham, J., 2004. Fighting Internet auction fraud: an assessment and proposal. Computer 37 (10), 31–37. http://dx.doi.org/10.1109/MC.2004.165.
Cortesão, Luis, Martins, Filipe, Rosa, António, Carvalho, Pedro, 2005. Fraud Management Systems in Telecommunications : a practical approach.
Cox, Kenneth C., 1997. Brief Application Description Visual Data Mining : Recognizing Telephone Calling Fraud. 231, pp. 225–231.
Crocker, Keith J., 2002. Insurance fraud and optimal claims settlement strategies*. J. Law Econ. XLV.
D'Arcy, Stephen, P., 2005. Predictive Modeling in Automobile Insurance: A Preliminary Analysis.
Dal Pozzolo, Andrea, Caelen, Olivier, Borgne, Yann-A.ël Le, Waterschoot, Serge, Bontempi, Gianluca, 2014. Learned lessons in credit card fraud detection from a practitioner perspective. Expert Syst. Appl. 41 (10), 4915–4928. http://dx.doi.org/10.1016/j.eswa.2014.02.026.
Delamaire, Linda, Abdou, Hussein, Pointon, John, 2009. Credit card fraud and detection techniques : a review. Banks Bank Syst. 4 (2).
Derrig, Richard A., Ostaszewski, Krzysztof M., 1995. Fuzzy techniques of pattern recognition in risk and claim classification. J. Risk Insur. 1995, 447–482.
Desai, Anita B., Deshmukh, Ravindra, 2013. Data mining techniques for fraud detection. Int. J. Comput. Sci. Inf. Technol. 4 (1), 1–4.
Dionne, Georges, 2002. Replacement cost endorsement and opportunistic fraud in automobile insurance*. J. Risk Uncertain.
Dionne, Georges, 1997. Development of an Expert System for the Automatic Detection of Automobile Insurance Fraud.
Dionne, Georges, Picard, Pierre, Paris, Descartes F., Giuliano, Florence, 2005. Optimal Auditing with Scoring Theory and Application to Insurance Fraud LABORATOIRE D ' ECONOMETRIE Insurance Fraud. November.
Dong, Fei, Shatz, Sol M., Xu, Haiping, 2009. Combating online in-auction fraud: clues, techniques and challenges. Comput. Sci. Rev. 3 (4), 245–258. http://dx.doi.org/10.1016/j.cosrev.2009.09.001.
Dong, Fei, Shatz, Sol M., Xu, Haiping, Majumdar, Dibyen, 2012. Price comparison: a reliable approach to identifying shill bidding in online auctions? Electron. Commer. Res. Appl. 11 (2), 171–179. http://dx.doi.org/10.1016/j.elerap.2011.12.003.
Dong, Wang, Quan-yu, Wang, Shou-yi, Zhan, Feng-xia, L.I., Da-zhen, Wang, 2004. A feature extraction method for fraud detection in mobile communication networks. In: Proceedings of Fifth World Congress on Intelligent Control Automation (IEEE Cat. No. 04EX788). 2, pp. 1853–1856. doi: http://dx.doi.org/10.1109/WCICA.2004.1340996.
Dorronsoro, R., Ginel, Francisco, Carmen, S., Santa Cruz, Carlos, 1997. Neural fraud detection in credit card operations. IEEE Trans. Neural Netw. 8 (4), 827–834.
Duman, Ekrem, Ozcelik, M. Hamdi, 2011. Detecting credit card fraud by genetic algorithm and scatter search. Expert Syst. Appl. 38 (10), 13057–13063. http://dx.doi.org/10.1016/j.eswa.2011.04.110.
Duman, Ekrem, Buyukkaya, Ayse, Elikucuk, Ilker, 2013. A novel and successful credit card fraud detection system implemented in a Turkish Bank. In: Proceedings of 2013 IEEE 13th International Conference on Data Mining Workshops. December 2013, pp. 162–171. http://dx.doi.org/10.1109/ICDMW.2013.168.
Durtschi, Cindy, Hillison, William, Pacini, Carl, 2004. The Effective Use of Benford' s Law to Assist in Detecting Fraud in Accounting Data. 99, 99, pp. 17–34.
Edelstein, Herb, 1997. Data Mining: Exploiting the Hidden Trends in Your Data. DB2 Online Mag. 2, 1.
Falaki, S.O., Alese, B.K., Adewale, O.S., Ayeni, J.O., Aderounmu, G.A., 2012. Probabilistic credit card fraud detection system in online transactions. Int. J. Softw. Eng. Appl. 6 (4), 69–78.
Fan, Wei, 2004. Systematic Data Selection to Mine Concept-Drifting Data Streams. pp. 128–137.
Farvaresh, Hamid, Sepehri, Mohammad Mehdi, 2011. A data mining framework for detecting subscription fraud in telecommunication. Eng. Appl. Artif. Intell. 24 (1), 182–194. http://dx.doi.org/10.1016/j.engappai.2010.05.009.
Farvaresh, Hamid, Mehdi, Mohammad, 2011. A data mining framework for detecting subscription fraud in telecommunication. Eng. Appl. Artif. Intell. 24 (1), 182–194. http://dx.doi.org/10.1016/j.engappai.2010.05.009.
Fawcett, Tom, Provost, Foster, 1997. Adaptive fraud detection. Data Min. Knowl. Discov. 1 (3), 291–316. http://dx.doi.org/10.1023/A:1009700419189.
Ferreira, P., Alves, Ronnie, Belo, Orlando, Ribeiro, Joel, 2007. Detecting Telecommunications Fraud based on Signature Clustering Analysis. In: Proceedings of Bus. Intell. Work. 13th Port. Conference Artificial Intelligence.
Fonebox Australia Group, 2011. PABX Fraud Alert – Notice to our customers regarding Toll Fraud Your PABX could expose you to Toll Fraud. PABX Fraud, also known as Toll Fraud, is causing extensive financial loss to organisations each year and is now impacting on Australian businesses. February, pp. 4000–4001.
Ford, B.J., Xu, H., Valova, I., 2012. A real-time self-adaptive classifier for identifying suspicious bidders in online auctions. Comput. J. 56 (5), 646–663. http://dx.doi.org/10.1093/comjnl/bxs025.
Francis, Charles, Pepper, Noah, Strong, Homer, 2011. Using support vector machines to detect medical fraud and abuse. In: Proceedings of Conference of the IEEE Engineering Medicine Biology Society. pp. 8291–8294.
Furlan, Štefan, Vasilecas, Olegas, Bajec, Marko, 2011. Method for selection of motor insurance fraud management system components based on business performance. Technol. Econ. Dev. Econ. 17 (3), 535–561. http://dx.doi.org/10.3846/20294913.2011.602440.
Gadi, Manoel Fernando Alonso, Wang, Xidi, Lago, Alair Pereira do, 2008. Credit card fraud detection with artificial immune system. In: Bentley, Peter J., Lee, Doheon, Jung, Sungwon (Eds.), Artificial Immune Systems SE – 11. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 119–131. http://dx.doi.org/10.1007/978-3-540-85072-4_11.
Gama, Joao, Bifet, Albert, Pechenizkiy, Mykola, Bouchachia, Abdelhamid, 2013. 1 A survey on concept drift adaptation. ACM Comput. Surv. 1 (1).
Ghosh, Mahuya, 2010. Telecoms fraud. Comput. Fraud Secur. 7, 14–17. http://dx.doi.org/10.1016/S1361-3723(10)70082-8.
Ghosh, S., Reilly, D.L., 1994. Credit card fraud detection with a neural-network. In: Proceedings of the Twenty-Seventh Hawaii International Conference on System Science. 3, pp. 621–630.
Global Fraud Study, 2012. Report to the nations 2012 global fraud study. Letter from the President & CEO.
Goel, Ankit, Xu, Haiping, Shatz, Sol M., 2010. A Multi-State Bayesian Network for Shill Verification in Online Auctions*. In Proceedings of the 22nd International Conference on Software Engineering and Knowledge Engineering, SEKE'2010. pp. 279–285.
Gomes, João Bártolo, Menasalvas, Ernestina, Sousa, Pedro A.C., 2011. Learning recurring concepts from data streams with a context-aware ensemble. In: Proceedings of the 2011 ACM Symposium on Applied Computing. SAC'11. p. 994. http://dx.doi.org/10.1145/1982185.1982403.
Gossett, Phil, Hyland, Mark, 1999. Classification, Detection and Prosecution of Fraud on Mobile Networks. 1, pp. 2–4.
Graaff, A.J., 2011. The Artificial Immune System for Fraud Detection in the Telecommunications Environment. pp. 1–4.
Guo, Tao, Li, Gui-yang, 2008. Neural Data Mining for Credit Card Fraud Detection. pp. 12–15.
Halvaiee, Soltani, Akbari, Neda, Kazem, Mohammad, 2014. A novel model for credit card fraud detection using artificial immune systems. Appl. Soft Comput. 24, 40–49. http://dx.doi.org/10.1016/j.asoc.2014.06.042.
Han, Jiawei, Kamber, Micheline, Pei, Jian, 2012. Data mining concepts and techniques. In: Kamber, Jiawei Han Micheline, Pei, Jian (Eds.), Data Mining (The Morgan Kaufmann Series in Data Management Systems), Third ed. Morgan Kaufmann, Boston, pp. 585–631. http://dx.doi.org/10.1016/B978-0-12-381479-1.00013-7.
Hand, David J., Crowder, Martin J., 2012. Overcoming selectivity bias in evaluating new fraud detection systems for revolving credit operations. Int. J. Forecast. 28 (1), 216–223. http://dx.doi.org/10.1016/j.ijforecast.2010.10.005.
He, Hongxing, Wang, Jincheng, Graco, Warwick, Hawkins, Simon, 1998. Application of neural networks to detection of medical fraud. Expert Syst. Appl. 13 (4), 329–336.
Held, Claudio M., Perez, Claudio A., Estevez, Pablo A., 2001. Subscription fraud prevention in telecommunications using fuzzy rules and neural networks. Expert Syst. Appl.
Hilas, Constantinos, Sahalos, John, 2007. An application of decision trees for rule extraction towards telecommunications fraud detection. In: Apolloni, Bruno, Howlett, RobertJ., Jain, Lakhmi (Eds.), Knowledge-Based Intelligent Information and Engineering Systems SE – 139. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg,

pp. 1112–1121. http://dx.doi.org/10.1007/978-3-540-74827-4_139.
Hilas, Constantinos S., 2009. Designing an expert system for fraud detection in private telecommunications networks. Expert Syst. Appl. 36 (9), 11559–11569. http://dx.doi.org/10.1016/j.eswa.2009.03.031.
Hilas, Constantinos S., Mastorocostas, Paris As, 2008. An application of supervised and unsupervised learning approaches to telecommunications fraud detection. Knowl.-Based Syst. 21 (7), 721–726. http://dx.doi.org/10.1016/j.knosys.2008.03.026.
Hilas, Constantinos S., Sahalos, John N., 2006. Testing the Fraud Detection Ability of Different User Profiles by Means of FF-NN Classifiers. pp. 872–883.
Hofmann, Markus, 2012. A comprehensive survey of methods for overcoming the class imbalance problem in fraud detection by Dr Peter Brennan. June.
Hollmen, Jaakko, 1999. Call-based Fraud Detection in Mobile Communication Networks using a Hierarchical Regime-Switching Model.
Hou, Jianwei, Rego, Cesar, 2007. A classification of online bidders in a private value auction: evidence from eBay. Int. J. Electron. Mark. Retail. 1 (4). http://dx.doi.org/10.1504/IJEMR.2007.014847.
Internet Crime Complaint Center, 2014. Internet Crime Report.
Jaiswal, Ashutosh, Kim, Yongdae, Gini, Maria, 2004. Design and implementation of a secure multi-agent marketplace. Electron. Commer. Res. Appl. 3 (4), 355–368. http://dx.doi.org/10.1016/j.elerap.2004.06.005.
Jenamani, Mamata, Zhong, Yuhui, Bhargava, Bharat, 2007. Cheating in online auction – towards explaining the popularity of English auction. Electron. Commer. Res. Appl. 6 (1), 53–62. http://dx.doi.org/10.1016/j.elerap.2005.12.002.
Jha, Sanjeev, Guillen, Montserrat, Westland, J. Christopher, 2012. Employing transaction aggregation strategy to detect credit card fraud. Expert Syst. Appl. 39 (16), 12650–12657. http://dx.doi.org/10.1016/j.eswa.2012.05.018.
Jiang, Wei, Au, Tom, Tsui, Kwok-Leung, 2007. A statistical process control approach to business activity monitoring. IIE Trans. 39 (2007), 235–249. http://dx.doi.org/10.1080/07408170600743912.
Johnson, Andy, 2012. Automated Inference of Shilling Behavior in Online Auction Systems. Defense Committee, Sol Shatz, Chair and Advisor.
Ju, Chunhua, Lu, Qibei, 2011. Research on credit card fraud detection model based on class weighted support vector machine. J. Converg. Inf. Technol. 6 (1), 62–68. http://dx.doi.org/10.4156/jcit.vol6.issue1.8.
Jyothsna, V., Rama Prasad, V.V., 2011. A review of anomaly based intrusion detection systems. Int. J. Comput. Appl. 28 (7), 26–35.
Kelley, Robert, 2009. White Paper Where can $700 Billion in Waste be cut Annually from the U. S. Healthcare System? October.
Kim, Hyun-Chul, Pang, Shaoning, Je, Hong-Mo, Kim, Daijin, Bang, Sung Yang, 2003. Constructing support vector machine ensemble. Pattern Recognit. 36 (12), 2757–2767. http://dx.doi.org/10.1016/S0031-3203(03)00175-4.
Kim, Min-jung, Kim, Taek-soo, 2002. A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection. pp. 378–383.
Kirlidog, Melih, Asuk, Cuneyt, 2012. A fraud detection approach with data mining in health insurance. Procedia – Soc. Behav. Sci. 62, 989–994. http://dx.doi.org/10.1016/j.sbspro.2012.09.168.
Koh, Hian Chye, Tan, Gerald, 2005. Data mining applications in healthcare. J. Healthc. Inf. Manag. 19 (2), 64–72.
Kokkinaki, A.I., 1997. On Atypical Database Transactions: Identification.
Konijn, Rob M., Kowalczyk, Wojtek, 2011. Finding Fraud in Health Insurance Data with Two-Layer Outlier Detection Approach. pp. 394–405.
Kou, Yufeng, Lu, Chang-tien, Sinvongwattana, Sirirat, 2004. Survey of fraud detection techniques. In: Proceedings of 2004 IEEE International Conference on Networking Sensing and Control. 2, 3, pp. 749–754.
Kramer, Stefan, Lavrač, Nada, Flach, Peter, 2000. Relational Data Mining. In Sa/ueso D/uezeroski, (Ed.). Springer-Verlag, New York, Inc., New York, NY, USA, pp. 262–286.
Krenker, Andrej, Volk, Mojca, Sedlar, Urban, Bešter, Janez, Kos, Andrej, 2009. Bidirectional artificial neural networks for mobile-phone fraud detection. ETRI J. 31 (1), 92–94. http://dx.doi.org/10.4218/etrij.09.0208.0245.
Krivko, M., 2010. A hybrid model for plastic card fraud detection systems. Expert Syst. Appl. 37 (8), 6070–6076. http://dx.doi.org/10.1016/j.eswa.2010.02.119.
Ku, Yungchang, Chen, Yuchi, Chiu, Chaochang, 2007. A Proposed Data Mining Approach for Internet Auction Fraud Detection. pp. 238–243.
Kumari, Nitu, Kannan, S., Muthukumaravel, A., 2014. Credit card fraud detection using Hidden Markov Model – a survey. Middle-East J. Sci. Res. 20 (6), 697–699. http://dx.doi.org/10.5829/idosi.mejsr.2014.20.06.11387.
Kundu, A., Panigrahi, S., Sural, S., Majumdar, A.K., 2009. BLAST-SSAHA hybridization for credit card fraud detection. IEEE Trans. Dependable Secur. Comput. 6 (4), 309–315.
Kundu, Amlan, Sural, Shamik, Majumdar, A.K., 2006. Two-Stage Credit Card Fraud Detection Using Sequence Alignment. pp. 260–275.
Kuşaksizoğlu, Bülent, 2006. Fraud detection in mobile communication networks using data mining.
López, Victoria, Fernández, Alberto, Moreno-Torres, Jose G., Herrera, Francisco, 2012. Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. Expert Syst. Appl. 39 (7), 6585–6608. http://dx.doi.org/10.1016/j.eswa.2011.12.043.
Laleh, Naeimeh, Azgomi, Mohammad Abdollahi, 2009. A Taxonomy of Frauds and Fraud Detection Techniques. 1, pp. 256–267.
Lane, T., Brodley, C.E., 1999. Temporal sequence learning and data reduction for anomaly detection. ACM Trans. Inf. Syst. Secur. 2 (3), 295–331.
Lee, Byungtae, Cho, Hyungjun, Chae, Myungsin, Shim, Seonyoung, 2010. Empirical analysis of online auction fraud: credit card phantom transactions☆. Expert Syst. Appl. 37 (4), 2991–2999. http://dx.doi.org/10.1016/j.eswa.2009.09.034.
Lei, Liang, 2012. Card fraud detection by inductive learning and evolutionary algorithm. In: Proceddings of the 2012 Sixth International Conference on Genetetic Evolutionary Computing. 156, August, pp. 384–388. http://dx.doi.org/10.1109/ICGEC.2012.70.
Li, Jing, Huang, Kuei-Ying, Jin, Jionghua, Shi, Jianjun, 2008. A survey on statistical methods for health care fraud detection. Health Care Manag. Sci. 11 (3), 275–287. http://dx.doi.org/10.1007/s10729-007-9045-4.
Liaw, Horng-Twu, Juang, Wen-Shenq, Lin, Chi-Kai, 2006. An electronic online bidding auction protocol with both security and efficiency. Appl. Math. Comput. 174 (2), 1487–1497. http://dx.doi.org/10.1016/j.amc.2005.06.016.
Lin, Shi-Jen, Jheng, Yi-Ying, Yu, Cheng-Hsien, 2012. Combining ranking concept and social network analysis to detect collusive groups in online auctions. Expert Syst. Appl. 39 (10), 9079–9086. http://dx.doi.org/10.1016/j.eswa.2012.02.039.
Lincoln, Robyn, Wells, Helene, Petherick, Wayne, 2003. An Exploration of Automobile Insurance Fraud. Humanities & Social Sciences papers. Paper 64. http://epublications.bond.edu.au/hss_pubs/64.
Liu, Xu-ying, Wu, Jianxin, Zhou, Zhi-hua, 2012. Exploratory undersampling for class-imbalance learning. IEEE Trans. Syst. Man. Cybern., 1–14.
Liu, Q., Wu, Y., 2012. Supervised learning. Encycl. Sci. Learn.
Liu, Qi, Vasarhelyi, Miklos, 2013. Healthcare fraud detection : a survey and a clustering model incorporating Geo-location information.
Liu, Xin, Kaszuba, Tomasz, Nielek, Radoslaw, Datta, Anwitaman, Wierzbicki, Adam, 2010. Using stereotypes to identify risky transactions in internet auctions. In: Proceedings of the 2010 IEEE Second International Conference on Social Computing. August, pp. 513–520. DOI:http://dx.doi.org/10.1109/SocialCom.2010.81.
Lookman Sithic, H., Balasubramanian, T., 2013. Survey of insurance fraud detection using data mining techniques. Int. J. Innov. Technol. Explor. Eng. 3 (2013), 62–65.
Lu, Fletcher, Boritz, J. Efrim, 2005. Detecting Fraud in Health Insurance Data : Learning to Model Incomplete Benford' s Law Distributions.
Lu, Fletcher, Boritz, J. Efrim, Covvey, Dominic, 2006. Adaptive Fraud Detection Using Benford' s Law. pp. 347–358.
Macia-Fernandez, Gabriel, Garcia-Teodoro, Pedro, Diaz-Verdejo, Jesus, 2009. Fraud in roaming scenarios: an overview. IEEE Wirel. Commun.
Maes, Sam, Tuyls, Karl, Vanschoenwinkel, Bram, Manderick, Bernard, 2002. Credit card fraud detection using Bayesian and neural networks. In: Proceedings of the 1st international Naiso Congress on Neuro Fuzzy Technologies.
Magalla, Asherry, 2013. Security, Prevention and Detection of Cyber Crimes. Tumaini University Iringa University College. Cyber Crime. Prepared by Asherry Magalla (LL . M-ICT LAW-10919), Supervised by Dr . Puluru (2013).
Mailloux, Allan T., Cummings, Stephen W., Mugdh, Mrinal, 2010. A decision support tool for identifying abuse of controlled substances by forward health medicaid members. J. Hosp. Mark. Public Relat., 37–41. http://dx.doi.org/10.1080/15390940903450982.
Major, John A., Riedinger, Dan R., 2002. EFD: A hybrid knowledge/statistical-based system for the detection of fraud. J. Risk Insur. 69 (3), 309–324.
Malekian, Donia, Hashemi, Mahmoud Reza, 2013. An Adaptive Profile based Fraud Detection Framework For Handling Concept Drift.
Maranzato, Rafael, Pereira, Adriano, 2010. Fraud detection in reputation systems in e-markets using logistic regression categories and subject descriptors. In: Proceedings of the 2010 ACM Symposium on Applied Computing. São Paulo, pp. 1454–1459.
Minegishi, Tatsuya, Niimi, Ayahiko, 2011. Proposal of credit card fraudulent use detection by online-type decision tree construction and verification of generality. Int. J. Inf. Secur. Res. 1 (4), 229–235.
Mohamed, Azlinah, Bandi, Ahmad Fuad Mohamed, Tamrin, Abdul Razif, Jaafar, Md Daud, Hasan, Suriah, Jusof, Faeizah, 2009. Telecommunication fraud prediction using backpropagation neural network. In: Proceedings of International Conference on Soft Computing and Pattern Recognition. pp. 259–265. http://dx.doi.org/10.1109/SoCPaR.2009.60.
Moreau, Yves, Preneel, Bart, Cooke, Chris, 1996. Novel Techniques for Fraud Detection in Mobile Telecommunication Networks.
Moreau, Yves, Verrelst, Herman, Vandewalle, Joos, 1997. Detection of mobile phone fraud using supervised neural networks : a first prototype. In: Proceedings of the International Conference on Art@cial Neural Networks Proceedings, ICA'97. pp.1065–1070, Elektrotechniek Esat, and Katholieke Universiteit Leuven.
Mule, Komal, Kulkarni, Madhuri, 2014. Credit Card Fraud Detection Using Hidden Markov Model (HMM). 11.
Murad, Uzi Pinkas, Gadi, 1999. Unsupervised Profiling for Identifying Superimposed. pp. 251–261.
Musal, Rasim Muzaffer, 2010. Two models to investigate medicare fraud within unsupervised databases. Expert Syst. Appl. 37 (12), 8628–8633. http://dx.doi.org/10.1016/j.eswa.2010.06.095.
Ng, K.S., et al., 2010. Detecting non-compliant consumers in spatio-temporal health data: a case study from medicare Australia. In: Proceedings of the 2010 IEEE International Conference on Data Mining Workshop. December, pp. 613–622. http://dx.doi.org/110.1109/ICDMW.2010.146.
Ngai, E.W.T., Wong, Yong Hu, Y.H., Chen, Yijun, Sun, Xin, 2011. The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. Decis. Support Syst. 50 (3), 559–569. http://dx.doi.org/10.1016/j.dss.2010.08.006.
Noor, N.M.M., Hamid, S.Ha, Mohemad, R., Jalil, Ma, Hitam, M.S., 2015. A review on a classification framework for supporting decision making in crime prevention. J. Artif. Intell. . http://dx.doi.org/10.3923/jai.2015.17.34
Ochaeta, Karen Elisa, 2008. Fraud Detection for Internet Auctions: A Data Mining Approach.

Ogwueleka, Francisca Nonyelum, 2011. Data mining application in credit card fraud detection system. J. Eng. Sci. Technol. 6 (3), 311–322.
Olszewski, Dominik, 2012. A probabilistic approach to fraud detection in telecommunications. Knowl.-Based Syst 26, 246–258. http://dx.doi.org/10.1016/j.knosys.2011.08.018.
Omlin, Christian, Bénard, Wiese, 2009. Credit card transactions, fraud detection, and machine learning: modelling time with LSTM recurrent neural networks. Innovations in Neural Information Paradigms and Applications. pp. 231–268.
Onderwater, Martijn, 2010. Detecting unusual user profiles with outlier detection techniques. September.
Oppliger, Rolf, 1997. Internet security: firewalls and beyond. Commun. ACM 40 (5), 92–102. http://dx.doi.org/10.1145/253769.253802.
Ortega, Pedro A., Ruz, Gonzalo A., 2006. A Medical Claim Fraud/Abuse Detection System based on Data Mining : A Case Study in Chile.
Oza, Nikunj C., 2005. Online bagging and boosting. In: Proceedings of Conference on Systems, Man and Cybernetics. 3, pp. 2340–2345.
Pérez, Jesús M., Muguerza, Javier, Arbelaitz, Olatz, Gurrutxaga, Ibai, Martín, José I., 2005. Consolidated Tree Classifier Learning in a Car Insurance Fraud Detection Domain with Class Imbalance. pp. 381–389.
Paasch, Carsten A.W., 2008. Credit Card Fraud Detection Using Artificial Neural Networks Tuned by Genetic Algorithms (Ph.D. thesis). Hong Kong University of Science and Technology.
Pathak, Jagdish, Vidyarthi, Navneet, Summers, Scott L., 2005. A fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims. Manag. Audit. J. 20 (6), 632–644. http://dx.doi.org/10.1108/02686900510606119.
Patidar, Raghavendra, Sharma, Lokesh, 2011. Credit card fraud detection using neural network. Int. J. Soft. Comput. Eng. 13–14
Pejic-Bach, Mirjana, 2010. Invited paper: profiling intelligent systems applications in fraud detection and prevention: survey of research articles. In: Proceedings of the 2010 International Conference on Intelligent Systems, Modelling and Simulation. IEEE, pp. 80–85. http://dx.doi.org/10.1109/ISMS.2010.26.
Philip, Nimisha, Sherly, K.K., 2012. Credit card fraud detection based on behavior mining. TIST Int. J. Sci. Technol., 7–12.
Phua, Clifton, Lee, Vincent, Smith, Kate, Gayler, Ross, 2005. A comprehensive survey of data mining-based fraud detection research. Artif. Intell. Rev.
Phua, Clifton, Smith-Miles, Kate, Lee, Vincent, Gayler, Ross, 2012. Resilient identity crime detection. IEEE Trans. Knowl. Data Eng. 24 (3), 533–546. http://dx.doi.org/10.1109/TKDE.2010.262.
Pinquet, Jean, 2007. Selection bias and auditing policies for insurance claims. J. Risk Insur. 74 (2), 425–440.
Potamitis, Giannis, 2013. Design and Implementation of a Fraud Detection Expert System Using Ontology – Based Techniques. University of Manchester (A dissertation submitted to the University of Manchester Giannis Potamitis School of Computer Science Table of Contents).
Pozzolo, Andrea Dal, Caelen, Olivier, Waterschoot, Serge, Bontempi, Gianluca, 2013. Racing for Unbalanced Methods Selection. pp. 24–31.
Qayyum, Sameer, Mansoor, Shaheer, Khalid, Adeel, Halim, Zahid, Baig, A. Rau, 2010. Fraudulent call detection for mobile networks. In: Proceedings of 2010 International Conference on Information and Emerging Technologies. June, pp. 1–5. http://dx.doi.org/10.1109/ICIET.2010.5625718.
Qian, Yu-hua, 2008. Fraud detection in telecommunication: a rough fuzzy set based approach. Information Technology, In: Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008, Hong Kong, pp. 12–15.
Quah, Jon T.S., Sriganesh, M., 2008. Real-time credit card fraud detection using computational intelligence. Expert Syst. Appl. 35 (4), 1721–1732. http://dx.doi.org/10.1016/j.eswa.2007.08.093.
Raj, S. Benson Edwin, Portia, A. Annie, 2011. Analysis on Credit Card Fraud Detection Methods. International Conference on Computer, Communication and Electrical Technology – ICCCET2011, 18th & 19th March, 2011.
Rebahi, Yacine, Nassar, Mohamed, Magedanz, Thomas, Festor, Olivier, 2011. A survey on fraud and service misuse in voice over IP (VoIP) networks. Inf. Secur. Tech. Rep. 16 (1), 12–19. http://dx.doi.org/10.1016/j.istr.2010.10.012.
Resnick, Paul, Zeckhauser, Richard, Friedman, Eric, Kuwabara, Ko, 2000. Reputation systems: facilitating trust in internet interactions. Commun. ACM 43 (12), 45–48. http://dx.doi.org/10.1145/355112.355122.
Richhariya, Pankaj, 2012. A Survey on Financial Fraud Detection Methodologies. 1, 1, pp. 14–24.
Rosas, Eugenio, Cesar, Analide, 2009. Telecommunications Fraud : Problem Analysis – an Agent-based KDD Perspective.
Ross, Gordon J., Adams, Niall M., Tasoulis, Dimitris K., Hand, David J., 2012. Exponentially weighted moving average charts for detecting concept drift. Pattern Recognit. Lett. 33 (2), 191–198. http://dx.doi.org/10.1016/j.patrec.2011.08.019.
Rosset, Saharon, Murad, Uzi, Neumann, Einat, Idan, Yizhak, Pinkas, Gadi, Amdocs Israel Ld, 1999. Discovery of Fraud Rules for Telecommunications Challenges and Solutions. pp. 409–413.
Ryman-Tubb, Nick F., Krause, Paul, 2011. Neural Network Rule Extraction to Detect Credit Card Fraud. pp. 101–110.
Sánchez-Maroño, N., Alonso-Betanzos, A., Tombilla-Snaromán, M., 2007. Filter methods for feature selection – a comparative study … Data Eng. …. pp. 178–187.
Sahin, Yusuf, Bulkan, Serol, Duman, Ekrem, 2013. A cost-sensitive decision tree approach for fraud detection. Expert Syst. Appl. 40 (15), 5916–5923. http://dx.doi.org/10.1016/j.eswa.2013.05.021.
Sahin, Y., Duman, E., 2011. Detecting Credit Card Fraud by Decision Trees and Support Vector Machines. I.
Sanver, Mert, Karahoca, Adem, 2009. Fraud Detection Using an Adaptive Neuro-Fuzzy Inference System in Mobile Telecommunication Networks.
Saravanan, P., Subramaniyaswamy, V., Sivaramakrishnan, N., Arun Prakash, M., Arunkumar, T., 2014. Data Mining Approach For Subscription-Fraud Detection in Telecommunication Sector. 7, 11, pp. 515–522.
Sasirekha, M., 2012. A defense mechanism for credit card fraud detection. Int. J. Cryptogr. Inf. Secur. 2 (3), 89–100. http://dx.doi.org/10.5121/ijcis.2012.2308.
Sasirekha, M., Thaseen, I. Sumaiya, Banu, J. Saira, 2012. An Integrated Intrusion Detection System for Credit Card Fraud Detection. pp. 55–60.
Seeja, K.R., Zareapoor, Masoumeh, 2014. FraudMiner : a novel credit card fraud detection model based on frequent itemset mining. Sci. World J. 2014.
Sethi, Neha, Gera, Anju, 2014. A Revived Survey of Various Credit Card Fraud Detection Techniques. 3, 4, pp. 780–791.
Shah, Harshit S., Joshi, Neeraj R., Sureka, Ashish, Wurman, Peter R., 2003. Mining eBay : Bidding Strategies and Shill Detection. pp. 17–34.
Shan, Yin, Jeacocke, David Murray, D. Wayne, Sutinen, Alison, 2008. Mining Medical Specialist Billing Patterns for Health Service Management.
Shao, Hua, Zhao, Hong, Chang, Gui-Ran, 2002. Applying data mining to detect fraud behavior in customs declaration. pp. 1241–1244.
Shawe-Taylor, John, Howkev, Keith, 1999. Detection of fraud in mobile telecommunications. Inf. Secur. Tech. Rep. 4 (1), 16–28.
Sherly, K.K., Nedunchezhian, R., 2010. Boat adaptive credit card fraud detection system.
Singh, Anshul, Narayan, Devesh, 2012. A survey on hidden markov model for credit card fraud detection. Int. J. Eng. Adv. Technol. 3, 49–52.
Sparrow, M.K., 2000. License To Steal: How Fraud bleeds America's health care system.
Srivastava, Abhinav, Kundu, Amlan, Sural, Shamik, Senior Member, 2008. Credit card fraud detection using hidden markov model. IEEE Trans. Dependable Secur. Comput. 5 (1), 37–48.
Stolfo, Salvatore J., Fan, David W. Lee, Wenke, Prodromidis, Andreas L., Chan, Philip K., 1997. Credit Card Fraud Detection Using Meta-learning : Issues and Initial Results. 1.
Šubelj, Lovro, Furlan, Štefan, Bajec, Marko, 2011. An expert system for detecting automobile insurance fraud using social network analysis. Expert Syst. Appl. 38 (2011), 1039–1052. http://dx.doi.org/10.1016/j.eswa.2010.07.143.
Suksmono, Andriyan B., Nugraha, Tedi, 2006. A Research on Usage Pattern and Analysis Technique for Communication Fraud : SIM Cloning and Surfing. pp. 1–6.
Sun, Bo, Yu, Fei, Wu, Kui, Xiao, Yang, Senior Member, Leung, Victor C.M., 2006. Enhancing security using mobility-based anomaly detection in cellular mobile networks. IEEE Trans. Veh. Technol. 55 (4), 1385–1396.
Syeda, Mubeena, Zbang, Yan-qing, Pan, Yi, 2002. Parallel granular neural networks for fast credit card fraud detection. In: Proceedings of the IEEE International Conference on Fuzzy Systems. pp. 572–577.
Tang, Mingjian, Mendis, B. Sumudu U., Murray, D. Wayne, Hu, Yingsong, Sutinen, Alison, 2011. Unsupervised Fraud Detection in Medicare Australia. pp. 103–110.
Taniguchi, Michiaki, Haft, Michael, Hollmkn, Jaakko, Tresp, Volker, 1998. Fraud Detection in Communication Networks Using Neural and. pp. 1241–1244.
Tao, Han, Zhixin, Liu, Xiaodong, Song, 2012. Insurance fraud identification research based on fuzzy support vector machine with dual membership. In: Proceedings of the 2012 International Conference on Information Management Innovation Management and Industrial Enginerring. October, pp. 457–460. http://dx.doi.org/10.1109/ICIII.2012.6340016.
Tawashi, Hiyam Ali El., 2010. Detecting Fraud in Cellular Telephone Networks.
Tennyson, Sharon, Forn, Pau, 2002. Claims auditing in automobile insurance: fraud detection and deterrence objectives. J. Risk Insur. 69 (3), 289–308.
Tennyson, Sharon, 2001. Claims Auditing in Automobile Insurance.
Thornton, Dallas, Mueller, Roland M., Schoutsen, Paulus, van Hillegersberg, Jos, 2013. Predicting healthcare fraud in medicaid: a multidimensional data model and analysis techniques for fraud detection. Procedia Technol. 9, 1252–1264. http://dx.doi.org/10.1016/j.protcy.2013.12.140.
Travaille, Peter, Thornton, Dallas, Müller., Roland M., 2011. Electronic Fraud Detection in the U.S. Medicaid Healthcare Program: Lessons Learned from other Industries, pp. 1–10.
Tripathi, Krishna Kumar, Pavaskar, Mahesh A., 2012. Survey on credit card fraud detection methods. Int. J. Emerg. Technol. Adv. Eng. 2 (11).
Tsai, Yao-hsu, Ko, Chieh-heng, Lin, Kuo-chung, 2014. Using CommonKADS method to build prototype system in medical insurance fraud detection. J. Netw. 9 (7), 1798–1802. http://dx.doi.org/10.4304/jnw.9.7.1798-1802.
Tsang, Sidney, Koh, Yun Sing, Dobbie, Gillian, Alam, Shafiq, 2014. Detecting online auction shilling frauds using supervised learning. Expert Syst. Appl. 41 (6), 3027–3040. http://dx.doi.org/10.1016/j.eswa.2013.10.033.
Tsang, Sidney, Dobbie, Gillian, Koh, Yun Sing, 2012. Evaluating fraud detection algorithms using an auction data generator. In: Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops. IEEE, pp. 332–339. http://dx.doi.org/10.1109/ICDMW.2012.34.
Tsung, Fugee, Zhou, Zhihong, Jiang, Wei, 2007. Applying manufacturing batch techniques to fraud detection with incomplete customer information. IIE Trans. 39 (6), 671–680. http://dx.doi.org/10.1080/07408170600897510.
Van Heerden, Johan H., 2005. Detecting Fraud in Cellular Telephone Networks. December.
Viaene, S., Derrig, R.A., Dedene, G., 2004. A case study of applying boosting naive bayes to claim fraud diagnosis. IEEE Trans. Knowl. Data Eng. 16 (5), 612–620. http://dx.doi.org/10.1109/TKDE.2004.1277822.

Viaene, S., Dedene, G., Derrig, R., 2005. Auto claim fraud detection using Bayesian learning neural networks. Expert Syst. Appl. 29 (3), 653–666. http://dx.doi.org/10.1016/j.eswa.2005.04.030.

Viaene, Stijn, Ayuso, Mercedes, Guillen, Montserrat, Gheel, Dirk Van, Dedene, Guido, 2007. Strategies for detecting fraudulent claims in the automobile insurance industry. Eur. J. Oper. Res. 176 (1), 565–583. http://dx.doi.org/10.1016/j.ejor.2005.08.005.

Vinicius Almendra and Denis Enachescu, 2011. A supervised learning process to elicit fraud cases in online auction sites. In: Proceedings of the 2011 13th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing. September. pp. 168–174. http://dx.doi.org/10.1109/SYNASC.2011.15.

Wang, Yan, Tan, Kian-Lee, Ren, Jian, 2004. PumaMart: a parallel and autonomous agents based internet marketplace. Electron. Commer. Res. Appl. 3 (3), 294–310. http://dx.doi.org/10.1016/j.elerap.2004.01.003.

Wang, Haixun, Fan, Wei, Yu, Philip S., Han, Jiawei, 2003. Mining Concept-Drifting Data Streams Using Ensemble Classifiers. pp. 226–235.

Wang, Shiguo, 2010. A comprehensive survey of data mining-based accounting-fraud detection research. In: Proceedings of the 2010 International Conference on Intelligent Computation Technology and Automation, ICICTA. 1, pp. 50–53. doi: http://dx.doi.org/10.1109/ICICTA.2010.831.

Wei, Wei, Li, Jinjiu, Cao, Longbing, Ou, Yuming, Chen, Jiahang, 2012. Effective detection of sophisticated online banking fraud on extremely imbalanced data. World Wide Web 16 (4), 449–475. http://dx.doi.org/10.1007/s11280-012-0178-0.

Weisberg, HerbertI., Richard A., Derrig, 1998. Quantitative methods for detecting fraudulent automobile bodily injury claims. Risques 35, 75–101.

Wen, Chieh-Hua, Wang, Ming-Jyh, 2005. Discrete choice modeling for bundled automobile insurance policies. J. East. Asia Soc. Transp. Stud. 6, 1914–1928.

Whitrow, C., Hand, D.J., Juszczak, P., Weston, D., Adams, N.M., 2008. Transaction aggregation as a strategy for credit card fraud detection. Data Min. Knowl. Discov. 18 (1), 30–55. http://dx.doi.org/10.1007/s10618-008-0116-z.

Wong, Nicholas, Ray, Pradeep, Stephens, Greg, Lewis, Lundy, 2012. Artificial immune systems for the detection of credit card fraud: an architecture, prototype and preliminary results. Inf. Syst. J. 22 (1), 53–76. http://dx.doi.org/10.1111/j.1365-2575.2011.00369.x.

Wu, Chih-Hung, Tzeng, Gwo-Hshiung, Goo, Yeong-Jia, Fang, Wen-Chang, 2007. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. Expert Syst. Appl. 32 (2), 397–408. http://dx.doi.org/10.1016/j.eswa.2005.12.008.

Xu, Haiping, Bates, Christopher K., Shatz, Sol M., 2009. Real-Time Model Checking for Shill Detection in Live Online Auctions*. pp. 134–140.

Xu, Wei, Wang, Shengnan, Zhang, Dailing, Yang, Bo, 2011. Random Rough Subspace Based Neural Network Ensemble for Insurance Fraud Detection. In: Proceedings of the 2011 Fourth International Joint Conference on Computational Sciences and Optimization. IEEE, pp. 1276–1280. http://dx.doi.org/10.1109/CSO.2011.213.

Yamanishi, Kenji, Takeuchi, Jun-ichi, Williams, Graham, Milne, Peter, 2004. On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. Data Min. Knowl. Discov. 8 (3), 275–300. http://dx.doi.org/10.1023/B:DAMI.0000023676.72185.7c.

Yang, Wan-Shiou, Hwang, San-Yih, 2006. A process-mining framework for the detection of healthcare fraud and abuse. Expert Syst. Appl. 31 (1), 56–68. http://dx.doi.org/10.1016/j.eswa.2005.09.003.

Yang, Wan-Shiou, 2003. A Process Pattern Mining Framework for the Detection of Health Care Fraud and Abuse (A Thesis Of By Wan-Shiou Yang In Partial Fulfillment of the Requirements for the Degree Of Doctor of Philosophy).

Yokoo, Makoto, Sakurai, Yuko, Matsubara, Shigeo, 2004. The effect of false-name bids in combinatorial auctions: new fraud in internet auctions. Games Econ. Behav. 46 (1), 174–188. http://dx.doi.org/10.1016/S0899-8256(03)00045-9.

Yusoff, Mohd Izhan Mohd, Mohamed, Ibrahim, Bakar, Mohd. Rizam Abu, 2013. Fraud detection in telecommunication industry using Gaussian mixed model. In: Proceedings of the 2013 International Conference on Research and Innovation Information Systems. November, pp. 27–32. http://dx.doi.org/10.1109/ICRIIS.2013.6716681.

Zadrozny, Bianca, Langford, John, Abe, Naoki, 2003. Cost-sensitive learning by cost-proportionate example weighting. In: Proceedings of the 3rd IEEE International Conference on Data Mining, ICDM'03.

Zareapoor, Masoumeh, Alam, M. Afshar, 2012. Analysis of credit card fraud detection techniques : based on certain design criteria. Int. J. Comput. Appl. 52 (3), 35–42.

Zaslavsky, Vladimir, Strizhak, Anna, 2006a. Credit card fraud detection using self-organizing maps. Inf. Secur 18 (2006), 48–63.

Zhu, Shunzhi, Wang, Yan, Wu, Yun, 2011. Health care fraud detection using non-negative matrix factorization. In: Proceedings of the 2011 6th International Conference on Computer Science & Education, ICCSE. IEEE, pp. 499–503. DOI: http://dx.doi.org/10.1109/ICCSE.2011.6028688.

Zhu, Xiaojin, Goldberg, Andrew B., 2009. Introduction to Semi-Supervised Learning.

Zliobaite, Indre, 2010. Learning Under Concept Drift: an Overview. CoRR abs/1010.4.