A Comparison of Machine Learning Algorithms for Fraud Detection

BY

Michael Johnson, University of South Wales
May 2014

Abstract

Today most peple and commecial organistions are economically dependent on Internet-based systems, whether directly or not. In some ways, financial information itself can be considered a critical asset, for without money we cannot function. As with anything of value, tangible or not, it is at risk of theft through fraud. This paper presents a review of proposed and deployed systems that are designed for detecting suspected financial theft through the use of stolen credentials. As most electronic financial transactions invariably involve the use of credentials, the core principles behind the reviewed systems might also be applied to the detection of unauthorisd access to user accounts on a corporate network, which makes this even more relevant to the field of computer security.

1. Introduction

Perhaps the fundamental problem, which financial institutions, banks and payment processors have invested heavily in attempting to resolve, is that identity cannot be attributed to individuals with absolute certainty. A bank or payment processor can only verify credentials and not people. Credentials can be stolen from one person and used by another. This is the problem that the reviewed systems address, and they must somehow distinguish between legitimate and fraudulent transactions through unauthorised access. One important thing that should be noted is that fraud can take several other forms than identity fraud. It could involve insider threats, misuse or money laundering, but this paper exclusively addresses fraud involving stolen credentials.

Of course, a very similar problem exists in the world of corporate network security, where user accounts and user authentication are involved. Like a bank account's Chip-and-PIN, a username and password might be obtained by a malicious party. This means the detection methods reviewed could potentially are not limited to the protection of bank accounts, but might potentially be applied to the broader field of network security.

But the specifics of authentication are irrelevant and beyond the scope of this review, which instead deals with the detection of fraudulent account activity after unauthorised access has been gained. That said, a reliable and accurate fraud detection system could actually be useful in assessing the effectiveness of a given authentication method after it is implemented.

The following fraud detection methods are reviewed and later evaluated: * Signature-based fraud detection: Comparing current activity with that associated with known fraud cases.

- Data mining: The aggregation and processing of data sets.
- Expert systems: Formalised and static method of encoding fraud detection expertise.
- Neural network: Adaptive system for detecting deviations from behaviour patterns.
- Bayesian systems: Evidence-based determination system.
- Improvised Competitive Learning Network: A system designed for efficiency and pre-emptiveness.

- Multi-layered hybrid model: A combination of methods for accuracy, reliability and efficiency.

To illustrate the difficulty of the problem the systems aim to solve, a typical financial institution or payment processor might handle millions of transactions per day, from across large geographical regions, and there would be somewhere between 70 and 80 variables per transaction. (Hand, 2007) As will become apparent, this leads to several design criteria being identified.

Authors have different methods of categorising fraud detection systems in order to make comparisons between them. Generally the two main categories are systems that compare input data against signatures of known attacks, and systems that detect anomalies. These are also broadly referred to as supervised and unsupervised systems elsewhere in literature related artificial intelligence.

Fraud detection methods in particular can also be categorised as proactive and reactive, the difference being whether a fraud detection system is capable of identifying, pre-empting and terminating a suspicious transaction before it is completed. The reactive methods are what Michael Edge and Pedro Sampaio (2009) term the 'store now query later' approach, where detection happens after a transaction is completed. A combination of methods can be referred to as a 'hybrid' or fusion system.

2. Materials and Methods

The review was performed through a critical analysis of multiple academic papers. Most of them were published post-2009, as the aim here is to provide a review of what is considered the 'state-of-the-art'. It is also critical that relatively recent source material is used, as the technologies for electronic commerce and transactions are rapidly evolving, as are the countermeasures used for preventing economic crime. Another aim of this paper is to provide a balanced review, where a variety of systems are briefly described and evaluated, rather than focussing on just two proposed methods. This approach has the advantage of providing a much broader view of automated fraud detection, and from this it might be possible to identify commonalities, design goals and perhaps make predictions regarding how methods might develop in the future.

3. Results

3.1 Signature-Based Fraud Detection

By far the simplest, straightforward and perhaps least intensive method of fraud detection is signature-based.

Signature-based systems are characterised as involving the comparison between current account activity and datasets of behaviours associated with known cases of fraud. Of course, this would involve stream mining in order to extract knowledge from data streams in network traffic if the system is implemented for detection in real-time.

However, such a system does have a fundamental drawback - it is not intelligent. Unless it is implemented in conjunction with a more elaborate method, it can only make comparisons with predefined signatures.

Drawing on a past comparison of signature-based methods for detecting fraud within large data sets. It's unclear whether the authors are referring to methods that could be deployed for dynamic data streams and real-time network activity.

One of the questions posed by the authors is whether 'Proactive Fraud Management' techniques can be deployed that detect potential fraud prior to the completion of transactions. The difficulty here, with the signature-based method, is that current account behaviour must be compared against datasets of activity associated with known fraud cases. This would increase the latency of performing a transaction. The idea of implementing 'thresholds' has been suggested, as the logic is easy to implement for real-time transactions, but the problem is the thresholds will vary between accounts. For example, one user might routinely withdraw large sums, while another user might draw minor sums rarely.

3.2 Data Mining

As is commonly known, data mining deals with large volumes of data aggregated with the objective of being able to profile objects, activities or behaviours. This is especially true of 'social networking' and targeted advertising, where companies might sift through large amounts of user-generated data in order to build profiles of users. It should be immediately obvious that this principle could also be applied to detecting fraud. As with social networks, financial institutions have datasets from which they can build profiles of account holders, and perhaps identify patterns within data aggregated from millions of accounts - patterns that might not be apparent within data from a much smaller group.

But there are differences between financial institutions and companies that own social networks, and these are related to the resources available. Data mining, when applied to social networks, is used for collecing data in near real-time, and the technologies utilised for this have been under heavy research and development for at least a decade. Companies such as FaceBook and Google also have the advantage of not only being able to collect that data from their own servers, but they have the financial resources to store the aggregated data in large data centres. Financial institutions, on the other hand, must somehow aggregate that data across multiple payment processors, ATM opertors and venues that accept Chip-and-PIN payments.

Another technical challenge is the raw data, or traffic generated by electronic payment methods, must be filtered in some way, so the fraud detection system (or at least the initial stages of it) focusses only on datasets that are relevant. Optimising the performance of fraud detection necessitates discarding the 'background noise' at the earliest possible stage.

Of course, if there are between 70 and 80 variables involved per transaction, some form of data mining, mining for relevant data within data streams, might actually be a necessity.

Having aggregated the data and sorted it into datasets, a financial institution could readily perform a range of statistical analyses to identify patterns, trends and perhaps information to add to a rule-based or expert system. Alternatively, the sorted data could become the input for Bayesian and 'associated rule' systems, so the application is not limited to statistical methods. (Li and Yen, et al, 2012)

According to David Hand (2007) and several authors, one of the most common forms of analysis is logistic regression, which estimates the probability that a transaction would be fraudulent. This is done in a very different way to Bayesian systems, discussed later in this paper, which makes a comparison between multiple opposing hypotheses by weighing evidence. Logistic regression deals with statistical likelihood that a transaction is fraudulent.

Commonly deployed as logistic regression might be, Siddhartha Bhattacharyya, et al, (2010) have proposed the use of another method, known as 'random forests', that could prove more suited to analysing mined data. The 'random forests' are so called because they are collections of decision trees, formed by the recursive partitioning of data. Conceivably this formation could happen in real-time as the data is aggregated. This makes random forests ideally suited for data mining applications.

Additionally, the random forests method, as demonstrated by Bhattacharyya, can reveal the number of observations made and the probability that a given object would fall into a given category. Of course, probabilities can change over time for a large system with millions of active user accounts, so the data must be resampled to keep the system current. This is commonly referred to as 'bagging'.

Of three systems evaluated, the Random Forests method was considered to provide the best overall performance. But the tests were only partialy based on real-world data, and the system was fine tuned. (Bhattacharyya, Jha, et al. 2010) Being computationally more efficient, and essentially rule-based, the method could be ideal for real-time fraud detection that might terminate suspect transactions before they are completed, and with fewer false positives than a more primitive system.

3.3 Patterns within Aggregated Data

This section discusses specifically the analysis of large data volumes beyond the technical background of data mining methods, as it is inevitable that fraud detection, both

automated and manual, must eventually be performed on large volumes of collected data. The rules that applied to a world in which ATMs were the only electronic financial system the account holder are obsolete. As mentioned, a typical account holder might interact with numerous payment systems and processors across a more geographically distributed area, and the amounts involved could vary between large cash withdrawals to much smaller online 'micropayments'. Somehow the fraud detection system must be capable of determining patterns within all these transactions, and across millions of account holders. A critical design goal for future systems might therefore be the capacity to identify patterns within aggregated transactional data instead of concentrating on individual transactions. (Jha, Guillen and Westland, 2012)

The main disadvantage with such an approach should be immediately obvious - this, by itself, would be a reactive method that processes the mined aggregated data after the transactions have been completed. Even if the method was optimised for efficiency, it would be for the processing of historical data that is not reflective of any individual tansaction.

The authors did suggest that aggregated transaction data could be partitioned in such a way that a detection system could focus on datasets where the probability of transactions being fraudulent is high, but again those datasets must be sorted from the greater pool of collected data. One suggestion would be to obtain the datasets by filtering the overall collected data through an optimised rule-based system that is discussed in the next section. After that point, more sophisticated forms of analyses might be applied.

3.4 Expert Systems

The expert systems are defined as a system that codifies or formalises expertise into a rigid, rule-based system. Algorithmically, an expert system would consist of a series of IF... THEN... ELSE statements into a type of 'decision tree'. Being non-adaptive, static, and limited to a fixed set of conditions, it has the disadvantage of requiring modification whenever a significant change in operating conditions arise. Expert systems do have the advantage of being generally less computationaly expensive, and this has implications for other current and proposed detection systems.

Although Constantinos Hilas' paper decribed expert systems in the context of detecting telecommunications-related fraud, the basic principles can be applied to any system that involves identities, authentication and user accounts. (Hilas, 2009) Both telecommunications and financial transactions involve real-time data streams, potentially large numbers of users at a given moment, telecommunications usage can be quantified in a vaguely similar wy to ATM usage, and identifiable data types from which user behaviours could be profiled. Noteably, we also find a Chip-and-PIN system is used for authenticating cellphone network subscribers, and there are perhaps as many cellphone subscribers as there are account holders with Chip-and-PIN cards.

In both, efficiency and the capability of processing real-time data are design goals for this method. Moreover, Hilas' discusses the use of expert systems for detecting fraud within large organisations, which could again be applied to the customers of a financial institution. The system in which users of a large network can use their accounts to make long-distance expensive calls can be analogous to a large number of users on a system making cash withdrawals.

Some degree of locality could also be expected, or at least geo-location data. Where users in the network Hilas described might have a tendency to use the same phones, real-world ATM users might make their cash withdrawals from ATMs within a given geographical area as a matter of routine. This is a behaviour trait that could be profiled, so a fraud detection system that focused on individual transactions might notice any deviation from that. This provides one of the rules in Hilas' expert system - was an account accessed from a terminal outside a region we'd typically expect it to be?

Although this principle might contradict that discussed for data mining and pattern recognition, it does demonstrate that rule-based and expert systems could be applied to services with large numbers of users for low-level anomaly detection.

3.5 Neural Networks

A neural network can be summarised as being as a series of mathematical functions that produce a score from a number of inputs, and in the context of fraud detection, would indicate

suspected fraud if that score reaches a given threshold. What determines the intelligence and the learning itself are the 'weights' given to specific nodes within the neural network as it adapts to inputs. (Alford, 2013)

However, the capabilities of artificial neural networks are limited, being effective only where restricted ranges of inputs are concerned. The more variation there is across the inputs, and the wider the range, the harder it becomes to detect anomalies. Neural networks are most useful for processing data within a given range that is incomplete or noisy, such are the predictive abilities. They could, at best, be used for detecting sudden changes in account activity, but obviously this alone is not strong enough evidence of fraud. (Estevez, Held and Perez, 2006) Estevez et al. (2006) and Alford (2013) also seem to be in agreement on the other weakness of the neural network approach: Neural networks can only indicate inconsistency and anomalies, and cannot reveal why the inconsistency or anomaly occurred. This would then make it extremely difficult to determine why transactions were flagged as being suspicious, in the absence of another method that could reveal the cause. Another implication of this is the difficulty in assessing the effectiveness of other security measures, in particular authentication. It is for these reasons that Bayesian systems are more commonly found than neural networks throughout this review. The next section will discuss the intrinsic features of Bayesian systems that make them highly appropriate for fraud detection.

3.6 Bayesian Belief Networks

A variation on the artificial neural network concept is the Bayesian Belief Network (BBN) The next step up from artificial neural networks is the Bayesian Belief Network (BBN), and the consensus across several of the papers reviewed appears to be that BBN is considered a far more accurate and adaptable alternative. On the surface, BBNs appear to be ideal for fraud detection. The question is whether the BBN systems are scalable enough to track millions of user accounts. The other question is whether BBNs are suitable enough, being probabilistic, for terminating suspect transactions without acting on too many false positives.

Rather like the human mind, a Bayesian system is one that makes decisions based on evidence available. Bayesian systems are considered more 'transparent', as the factors (the evidence) on which the output is based are known, as is the weighting of each type of evidence. The designer or developer is then forced to consider how much weight should be given to each type of input. Represented graphically, a Bayesian network can communicate an argument in a clear, structured way. (Fenton, 2012) How would this be applied to detecting fraud? It could be argued that a Bayesian system is almost ideal, because it can determine the probability that a given transaction is fraudulent based on the available evidence, and whatever action could be taken if the probability exceeds a threshold. Compare this to an artificial neural network, which simply indicates when activity somehow deviates from learned behaviour.

These systems are more probabilistic in nature. They work on the assumption there is a known ratio between legitimate and fraudulent transactions, and therefore the probability that the system has correctly flagged a suspected transaction is known.

Based on financial fraud reports supplied by Deloitte dating between 2000 and 2006, Wei Zhou and Gaurav Kapoor (2011) surmise that fraud methods are evolving in relation to changing industry-wide practices. This would mean that a long-term countermeasure would either need to be adaptive or intrinsically suited to handling conditional changes. Neural networks that require intervention and fundamental adjustments to their design are also limited in this respect also. These limitations were pointed out, that artifician neural networks aren't suited to changing conditions or volatile input. (Zhou and Kapoor, 2011)

3.7 Improvised Competitive Learning Network (ICLN)

John Zhong Lei and Ali Ghorbani (2011) sought to address two problems specific to fraud detection applications that are metioned elsewhere in this review: Firstly, the volume of Internet traffic generated by millions of transactions per day, a volume that is increasing. Secondly, Lei and Ghorbhani argue that most real-world fraud detection does not happen until after transactions are completed, and therefore fraud detection systems in general are currently ineffective for preventing financial losses. (Zhong and Ghorbani, 2011)

Therefore a highly efficient system is required.

The ICLN and its supervised learning counterpart, has also been developed to be implemented as a network intrusion detection measure.

According to the authors, the other anticipated problem with the reactive system is that detection happens after the transaction is completed, and therefore it might take a matter of weeks for a fraudulent transaction to appear on a bank statement and for the customer to susequently report the incident to the financial institution. Within that time, of course, the behaviour associated with the account would have changed so the fraudulent activity might be considered normal by whatever learning algorithm was in place.

We have identified another criterion, or area for improvement. The ideal system would preempt the completion of a fraudulent transaction. Such a system must therefore be highly efficient, fast and reliable.

3.8 Hybrid Fraud Detection Systems

Proposed by Suvasini Panigrahi, et al, this consists of 1) rule-based system, 2) Dempster-Shafer adder, 3) History database and 4) Bayesion component. (Panigrahi, Kundu, Sural, et al. 2009)

The authors proposing this method, as with several other authors, have singled out Bayesian belief networks as being one of the optimal systems for detecting fraud with a much higher degree, and they sought to improve on this with a hybrid system that combines Bayesian learning with other methods.

Since a rule-based system might be the fastest, and therefore the most suitable for examining large volumes of real-time data, although with less accuracy, it forms the initial filtering component in the hybrid model. From this, only the transactions considered suspect might be examined further. This could prove to be an intrinsic weakness in the proposal, as it has already been identified that fraud methods can evolve faster than the countermeasures. Conversely, a rule-based system might lead to the detection of most fraud, if only a minority of fraudsters are using methods that defeat such a system.

To address this, Suvasini Panigrahi, et al, have included something they termed 'outlier detection' as part of their rule-based system. The concept underpinning this is a familiar one - outlier detection is another less flexible method of detecting deviations from normal ATM usage. For example, if an account holder tends to withdraw very similar ammounts on a daily basis from a group of ATMs within a given area, each occurrence would be a point within a cluster. If, for some reason, the account holder withdraws a much larger amount from a cash machine hundreds of miles away, this occurrence would be a point noticeably outside that cluster, and therefore an indication that credentials required to make withdrawals might have been stolen. Of course, the system is not as effective if the account holder's behaviour is more random, or the account is a new one.

Here the historical data is stored in a 'Transaction History Database' (THD), and the authors had focussed on the transaction times, using a 'fraud frequency table' and 'good frequency' table as references for the rule-based system.

This rule-based system is supplemented by a 'Dempster-Schafer Adder' as a second or third stage, and this functions in a similar way to the Bayesian system in calculating the probability, from multiple data points, of a transaction being fraudulent. But potentially the Dempster-Schafer adder becomes ineffective when the probabilities of a transaction being fraudulent and legitimate are given equal weight. It seems the Dempster-Schafer adder and the Bayesian system are used in conjunction as a matter of redundancy.

In terms of accuracy and reliability, we would expect a hybrid model such as this to be optimal, but the authors did not discuss the computational efficiency of a real-world implementation or how it would scale to a large distributed system with many users. Their system was benchmarked from generated data that only simulated the activities of a group of account holders.

4. Conclusion and Discussion

Throughout the review, it has been demonstrated there is a very delicate balance between efficiency and accuracy when deciding which automated fraud detection methods to implement. It is a balance between a system that

functions with minimal computational overhead but less accuracy, and an accurate system that is computationally more expensive. No single method can provide accuracy and speed. Efficiency and accuracy have therefore become design goals for fraud detection systems, and a common method of addressing this is to use a rule-based system for real-time detection, and more sophisticated systems for analysing historical data. The only real question is how the two should be integrated.

Several of the authors have come to the realisation that an ideal fraud detection system would be a combination of two or more methods, perhaps in the form of a lightweight rule-based system that passes suspected transactions to a more sophisticated machine learning system. More specifically, we could say that an ideal system would appear to have an Expert System as its initial component, and a more adaptable Bayesian Belief Network as a latter stage.

Why is there a need for rule-based systems that are less accurate, or an expert system that evolves slower than the methods for committing fraud? One of the biggest reasons would be that fraud detection in itself is near useless unless it can prevent financial loss somehow. For that, the system must be capable of preventing fraud. At worst, it should limit the amount that could be stolen over a period of time, and at best it could prevent the first transaction being completed if unauthorised access was gained to a bank account.

So real-world fraud detection has two classes of system. Several authors discuss the methods in terms of being pro-active and reactive. Pro-active methods detect and terminate a fraudulent transaction before it is completed, and reactive methods detect fraud after the fact. The comparison is rather the same as that between Intrusion Detection Systems (IDS) and Intrusion Prevention Systems (IPS), where only the latter is capable of actually preventing network-based attacks.

There are other design goals, namely scalability, adaptability and extensibility. However effective the reviewed systems are for today's payment methods and online transactions, they must be applicable to future methods also. This is critical, as methods that prove effective today could be entirely unsuitable in future. Again, it is predicted that the development of lightweight, pro-active and reliable methods will be of primary importance as the number of Internet-based transactions continue to increase. It is also predicted that future systems will be developed for integration with more commonly-known data mining methods, the latter already being developed to an advanced level by companies such as Google and FaceBook.

5. References

Battacharyya, S. Jha, S. Tharakunnel. Westland, C. 2010. Data mining for credit card fraud: A comparative study. *Decision Support Systems*. Volume 50 (Issue 3). [PDF]. http://dx.doi.org/10.1016/j.dss.2010.08.008. (15th April 2014).

Edge, M. Sampaio, P. 2009. A survey of signature based methods for financial fraud detection. *Computers & Security*. Volume 28 (Issue 6). [PDF]. http://dx.doi.org/10.1016/j.cose.2009.02.001. (28th March 2014).

Estevez, P. Held, C. Perez, C. 2006. Subscription fraud prevention in telecommunications using fuzzy rules and neural networks. *Expert Systems with Applications*. Volume 31 (Issue 2). [PDF]. http://dx.doi.org/10.1016/j.eswa.2005.09.028. (31st March 2014).

Fenton, M. 2012. *Bayesian belief networks: an overview*. [WWW]. http://www.eecs.qmul.ac.uk/~norman/BBNs/

BBNs.htm. (23rd January 2014).

Hand, D. 2007. *NATO ASI: Mining Massive Data sets for Security - Statistical techniques for fraud detection, prevention, and evaluation*. [PDF]. http://langtech.jrc.ec.europa.eu/mmdss

2007/htdocs/Presentations/Docs/

MMDSS_Hand_PUBLIC.pdf

(20th January 2014).

Hilas, C. 2009. Designing an expert system for fraud detection in private telecommunications networks. *Expert Systems with Applications*. Volume 36 (Issue 9). http://dx.doi.org/10.1016/j.eswa.2009.03.031. (10th March 2014).

Jha, S. Guillen, M. Westland, C. 2012. Employing transaction aggregation strategy to detect credit card fraud. *Expert Systems with Applications*. Volume 39 (Issue 16). [PDF]. http://dx.doi.org/10.1016/j.eswa.2012.05.018. (1st May 2014).

Li, S. Yen, D. Lu, W. Wang, C. 2012. Identifying the signs of fraudulent accounts using data mining techniques. *Computers in Human Behaviour*. Volume 28 (Issue 3). [PDF]. http://dx.doi.org/10.1016/j.chb.2012.01.002. (28th March 2014).

Panigrahi, S. Kundu, A. Sural, S. Majumdar, A. 2009. Credit card fraud detection: A fusion approach using Dempster?Shafer theory and Bayesian learning. Information Fusion. Volume 10 (Issue 4). [PDF]. http://dx.doi.org/10.1016/j.inffus.2008.04.001. (20th April 2014).

Zhong Lei, J. Ghobani, A. 2011. Improved competitive learning neural networks for network intrusion and fraud detection. *Neurocomputing*. Volume 75 (Issue 1). [PDF]. http://dx.doi.org/10.1016/j.neucom.2011.02.021. (5th February 2014).

Zhou, W. Kapoor, G. 2011. Detecting evolutionary financial statement fraud. *Decision Support Systems*. Volume 50 (Issue 3). [PDF]. http://dx.doi.org/10.1016/j.dss.2010.08.007. (12th March 2014).