Detecting Cellular Fraud Using Adaptive Prototypes.

Peter Burge and John Shawe-Taylor

Department of Computer Science
Royal Holloway University of London
Egham, Surrey, England. TW20 0EX
{Peteb,jst}@dcs.rhbnc.ac.uk

Abstract

This paper discusses the current status of research on fraud detection undertaken as part of the European Commission-funded ACTS ASPeCT (Advanced Security for Personal Communications Technologies) project, by Royal Holloway University of London. Using a recurrent neural network technique, we uniformly distribute prototypes over Toll Tickets, sampled from the U.K. network operator, Vodafone. The prototypes, which continue to adapt to cater for seasonal or long term trends, are used to classify incoming Toll Tickets to form statistical behaviour profiles covering both the short and long-term past. These behaviour profiles, maintained as probability distributions, comprise the input to a differential analysis utilising a measure known as the Hellinger distance[5] between them as an alarm criteria. Fine tuning the system to minimise the number of false alarms poses a significant task due to the low fraudulent/non fraudulent activity ratio. We benefit from using unsupervised learning in that no fraudulent examples are required for training. This is very relevant considering the currently secure nature of GSM where fraud scenarios, other than Subscription Fraud, have yet to manifest themselves. It is the aim of ASPeCT to be prepared for the would-be fraudster for both GSM and UMTS.

Introduction

When a mobile originated phone call is made or various inter-call criteria are met the cells or switches that a mobile phone is communicating with produce information pertaining to the call attempt. These data records, for billing purposes, are referred to as Toll Tickets. Toll Tickets contain a wealth of information about the call so that charges can be made to the subscriber. By considering well studied fraud indicators these records can also be used to detect fraudulent activity. By this we mean interrogating a series of recent Toll Tickets and comparing a function of the various fields with fixed criteria, known as triggers. A trigger, if activated, raises an alert status which cumulatively would lead to investigation by the network operator. Some example fraud indicators are that of a new subscriber making long back-to-back international calls being indicative of direct call selling or short back-to-back calls to a single land number indicating an attack on a PABX system. Sometimes geographical information deduced from the cell sites visited in a call can indicate cloning. This can be detected through setting a velocity trap.

Fixed trigger criteria can be set to catch such extremes of activity, but these absolute usage criteria cannot trap all types of fraud. An alternative approach to the problem is to perform a differential analysis. Here we develop behaviour profiles relating to the mobile phone's activity and compare its most recent activities with a longer history of its usage. Techniques can then be derived to determine when the mobile phone's behaviour changes significantly. One of the most common indicators of fraud is a significant change in behaviour.

The performance expectations of such a system must be of prime concern when developing any fraud detection strategy. To implement a real time fraud detection tool on the Vodafone network in the U.K, it was estimated that, on average, the system would need to be able to process around 38 Toll Tickets per second. This figure varied with peak and off-peak usage and also had seasonal trends. The distribution of the times that calls are made and the duration of each call is highly skewed. Considering all calls that are made in the U.K., including the use of supplementary services, we found the average call duration to be less than eight seconds, hardly time to order a pizza.

In this paper we present one of the methods developed under ASPeCT that tackles the problem of skewed distributions and seasonal trends using a recurrent neural network technique that is based around unsupervised learning. We envisage this technique would form part of a larger fraud detection suite that also comprises a rule based fraud detection tool and a neural network fraud detection tool that uses supervised learning on a multi-layer perceptron. Each of the systems has its strengths and weaknesses but we anticipate that the hybrid system will combine their strengths.

The following section discusses in more detail the concept of behaviour profiling for the purposes of performing a differential analysis. This is followed, in section 3, by the neural network prototyping technique and the way these prototypes are used to generate behaviour profiles. In section 4 we describe the workings of the fraud engine and follow up with some preliminary results. Lastly we discuss how we cater for changing distributions.

Behaviour profiling

For a differential analysis we need information about the mobile phone's history of behaviour plus a more recent sample of the mobile phone's activities. An initial attempt might be to extract heuristic information from the Toll Tickets and store it in record format. For this simple scenario we would need to consider two windows or time spans over the sequence of transactions for each user. The shorter sequence could be called the Current Behaviour Profile (CBP) and the longer sequence the Behaviour Profile History (BPH). Both profiles could be treated as finite length queues. When a new Toll Ticket arrives, relating to a given user, the oldest entry from the BPH would be discarded and the oldest entry from the CBP would move to the back of the BPH queue. The new record encoded from the incoming Toll Ticket would then join the back of the CBP queue.

Clearly, in practice, it is not optimal to search and retrieve a history of transaction records from a database prior to each calculation on receipt of a new Toll Ticket. Instead we compute a single behaviour profile record which we store in a database using the International Mobile Subscription Identity (IMSI) as the primary key. As a new Toll Ticket arrives, for a particular subscriber, the profile record is simply updated with information reduced from the Toll Ticket. In order to preserve the concept of two different time spans over the Toll Tickets, we will need to decay the influence of previous Toll Tickets, on the profiles, before adding in information from the new one. By applying two different decay factors we can maintain the concept of a CBP and a BPH. Of course we have to be careful not to dilute information by applying a decay factor and thus introducing false information to the behaviour profile. The following section describes a prototyping technique, based on the Second Maximal Entropy Principle by Grabec[1] which enables us to construct statistical behaviour profiles that are simple to decay without introducing false behaviour patterns.

Prototyping

Prototyping is a method of forming an optimal discrete representation of a naturally continuous random variable. The processing of continuous random variables by discrete systems generally reduces empirical information. Neural Networks are capable of forming optimal discrete representations of continuous random variables through their ability to converge, by lateral interaction, to stable uniformly distributed states, von der Malsburg[4].

Grabec[1] introduces a technique to dynamically generate prototypical values to span a continuous random variable as samples are taken from it. He also suggests an extension to generate prototypes for multi dimensional random variables. In its simplest form his method resembles the more well known self organisation technique developed by Kohonen 1989 [3]. However, Grabec's method does not restrict the prototypes to lie on a two dimensional manifold, but allows them to form their own topology. Only one pass through the training set is required giving rise to the potential for online adaption.

Grabec introduced the second maximal entropy principle stating that

*The mapping of a continuous random variable X into a set of K discrete prototypes **Q** reduces the empirical information by the least amount if a uniform distribution ($\mathbf{P}(q_i) = \frac{1}{K}, i = 1 \ldots K$), corresponding to the absolute maximum ($S_Q = \log K$) of information entropy, is assigned to **Q**.*

When considering the set of all possible Toll Tickets, we clearly have a dimension to represent every parameter, from a Toll Ticket, that we wish to include in the analysis. Each parameter of a Toll Ticket can assume a range of values and is thus itself a random variable. Grabec's technique enables us to create a number of prototypes that dynamically and uniformly span the set of samples from a download of Toll Tickets taken from a live network. Owing to the fact that there are so few fraudulent Toll Tickets in comparison to non fraudulent ones, in a live network download, the prototypes will organize themselves as if the data were totally fraud free. The resulting set of prototypes will enable us to classify future incoming Toll Tickets with minimal loss of empirical information.

To distribute the prototypes over the input stream of Toll Tickets, we set up an iterative procedure that computes the change in the current value of the K prototypes **Q**

$$\Delta q_{lm}^{(i+1)} = B_{lm} - \sum_{k \neq l}^{K} \sum_{i \neq m}^{M} C_{lmki} \Delta q_{ki} \qquad ;$$
$$l = 1 \ldots K; \ \mathrm{m}=1 \ldots \mathrm{M}$$

which starts with $\Delta q_l^{(0)} = B_l$. The coefficients are determined by the expressions:

$$C_{lmki} = \left[\delta_{mi} - \frac{(q_{lm} - q_{km})(q_{li} - q_{ki})}{2\sigma^2} \right] \exp\left[\frac{-(q_l - q_k)^2}{4\sigma^2} \right]$$

$$B_{lm} = \frac{K}{N+1} \left\{ \begin{array}{l} (X_{N+1,m} - q_{lm}) \exp\left[\frac{-(X_{N+1} - q_l)^2}{4\sigma^2} \right] \\ -\frac{1}{K} \sum_{k=1}^{K} (q_{km} - q_{lm}) \exp\left[\frac{-(q_k - q_l)^2}{4\sigma^2} \right] \end{array} \right\}$$

where $\sigma = \frac{S}{K^{1/M}}$ approximates the standard deviation and S is the expected range of **X**

Figure 1 below shows the distribution of 50,000 international calls randomly sampled over a two month period from the Greek network operator Panafon. In this figure we only consider the two parameters TT-Charging-Start-Time and TT-Call-Duration.

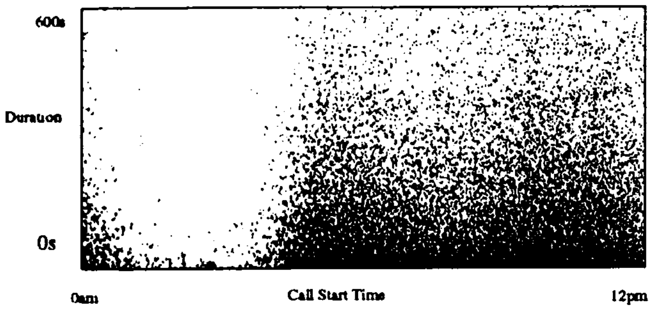

Figure 1 - 50,00 international calls from the Panafon network.

Figure 2 demonstrates the result of applying the neural network prototyper to the above data specifying that it is to develop 50 prototypes.

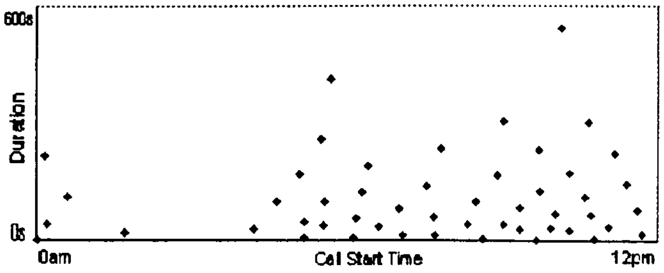

Figure 2 - 50 prototypes of the 50,000 international calls from the Panafon network.

Constructing profiles

Using the K Toll Ticket prototypes Q, we can now encode future Toll Tickets as feature vectors v

where $v_j = \frac{\exp(-\| X_{N+1} - Q_j \|)}{\sum_{j=1}^{K} \exp(-\| X_{N+1} - Q_j \|)}$.

Note that $\sum_{j=1}^{K} v_j = 1$ and so the feature vector can be viewed as a probability distribution.

Feature vectors are generated for National Calls, International Calls and for the use of supplementary services. The CBP and the BPH are now also formed as probability distributions using two different decay factors α and β to maintain the concept of two time spans over the Toll Tickets. After the encoding of each Toll Ticket into the feature vector v, each element of the CBP is updated in the following way

$$C_i = \alpha C_i + (1 - \alpha) v_i .$$

Where C_i is the *ith* entry of the CBP. Note that $\sum C_j = 1$.Following this, both the CBP and BPH are presented to the fraud engine as will be discussed in the following section.

The BPH is then updated using the CBP as follows:

$$H_i = \beta H_i + (1 - \beta) C_i$$

where H_i represents the *ith* entry in the BPH. This also makes $\sum H_j = 1$.

The fraud engine

The task of the fraud engine is to take the user profile record, consisting of the CBP and the BPH, and calculate a measure known as the Hellinger distance

$$d=\sum_{i=0}^{K}\left(\sqrt{C_i}-\sqrt{H_i}\right)^2$$

where C and H are the CBP and BPH respectively and K is now the number of entries in the profile record. This is a natural measure to take when comparing two probability distributions. The Hellinger distance will always be a value between zero and two where zero is for equal distributions and two represents orthogonality. The Hellinger distance in this scenario can be seen as a measure of how erratic the behaviour is. The fraud engine then checks if d is less than some threshold value, which is determined by experiment through performance tuning, and if necessary raises an alarm.

The value of d can be further used to indicate how severe the change in behaviour was. This means that alarms can be prioritised for investigation by the network operator. Owing to this fact, we are not so concerned about setting low alarm thresholds because if sorted into order of severity, only the upper part of the ordered list need be investigated.

Figure 3 shows a BPH and a CBP for a subscriber who was displaying what might be considered as acceptable behaviour, i.e. the fraud detection tool had not raised an alarm. The subscriber did not make international calls however did use supplementary services such as voice mail and call forwarding.

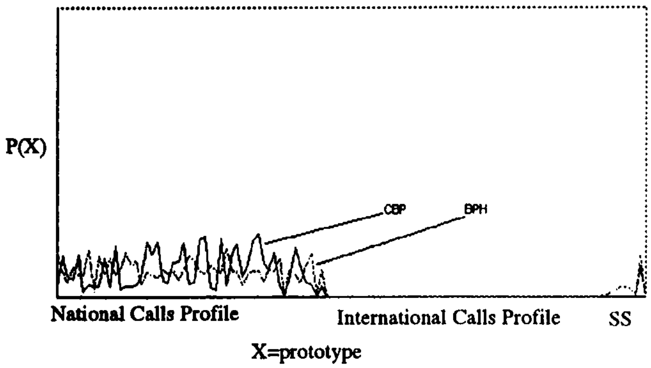

Figure 3 - A CBP and a BPH of a subscriber exhibiting acceptable behaviour.

Figure 4 show the CBP and BPH of a subscriber who raised an alarm. It is interesting to note that the spike is for the BPH which indicates that the alarm was raised through a sudden drop in activity. One fraud scenario that could account for this is handset theft where the thief is unlikely to call the subscribers voice mail service.

On its own however, one would probably not consider a drop in activity as a primary fraud indicator however it could be used to support other indications.

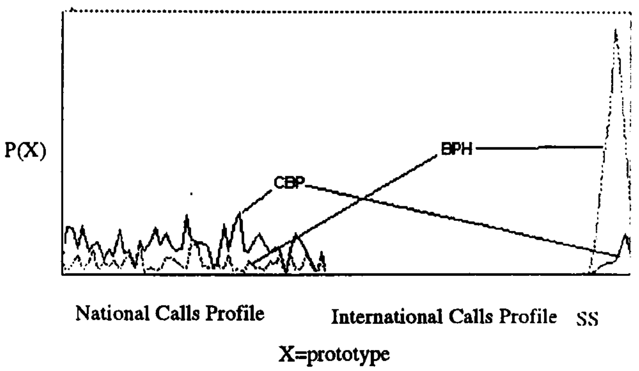

Figure 4 - A CBP and a BPH of a subscriber who raised an alarm.

Experimental results

The fraud detection tool is still in the very early stages of development. Initial trials have been performed using parameters that have yet to be performance tuned, with very pleasing results. For these initial trials, we only developed prototypes of the 2D space of call start time verses call duration.

To develop the prototypes we used a two months download of Toll Tickets relating to all calls made by new subscribers. 50 prototypes were generated relating to national calls, 50 for international calls and 10 for calls relating to supplementary services. The decay factors α and β used to maintain the CBP and the BPH were 0.9 and 0.98 respectively. Behaviour profiles were developed over a two week period before detection started.

Two data sets were used for the trial itself. The first consisted of the Toll Tickets for a subset of non fraudulent subscribers. The second comprised the Toll Tickets for subscribers whose service had been barred due to the detection of overlapping calls, indicative of cloning. These Toll Tickets were converted from the analogue TACS network into GSM format.

Even though these fraudsters were all detected for overlapping calls, our fraud detection tool would base its analysis on behaviour change. For this reason we considered this trial a valid first attempt. The result was that our fraud detection tool detected 75% of the fraudsters whilst only misclasifying 4% of the valid subscribers as fraudsters. However as the list was ordered according to the severity of the fraud the 4% of misclassified subscribers appeared at the bottom of the fraudster list.

Online adaptation of prototypes

By considering a window over the sequence of Toll Tickets, within which prototypes are uniformly distributed, we can ensure that the prototypes adapt to natural trends in behaviour such as with seasonal drift. This ensures that the feature vectors being constructed always comprise maximum information content, enabling clear distinctions between Current and History profiles.

Grabec's technique encodes the input Toll Tickets X as prototypes Q according to the probability distribution

$$p_r(x) = \frac{1}{K(2\pi\sigma)^{M/2}} \sum_{k=1}^{K} \exp\left[\frac{-\|x-q_k\|^2}{2\sigma^2}\right]$$

where $\sigma = S/K^{1/M}$.

We propose that with hindsight, we could focus the generation of prototypes on Toll Tickets that prove to be of most relevance to fraud scenarios. For example, concentrating less on medium duration national calls. We propose that this be implemented by an adaptive critic, Haykin[2], whereby information concerning the success or failure of a scheme is used to direct the future learning experience, determining regions of the Toll Ticket space critical to any analysis. A negative effect of an evolving set of prototypes is the impact on very infrequent users whose profiles will become unstable showing untrue behavioural anomalies. We are currently investigating this issue, however question how useful a differential analysis of this category of user is in the first place.

There are a number of parameters, in addition, relating to the operation of the fraud detection tool that need to be optimised. The first is the number of prototypes that will be used to encode the Toll Ticket space. Next are the two decay factors α and β which determine the relative lengths of the short and long term past. These decay factors could well be critical to the success or failure of the system. Further parameters such as the number of Toll Tickets, or the number of days, that the behaviour profiles should be developed over before detection starts and the alarm threshold need to be determined by experiment.

The above ideas are currently being developed and put into practice for the next generation of our fraud detection tool.

References

[1] Grabec I, Modelling of Chaos by a Self-Organizing Neural Network. Artificial Neural Networks. Proceedings of ICANN, Espoo, Vol 1, pp 151-156, Elsevier publications.

[2] Haykin S, 1984, Neural Networks A Comprehensive Foundation. Macmillan College Publishing Company.

[3] Kohonen T, 1988, Self-organization and Associative Memory, Springer Verlag, Berlin, pp 119.

[4] Malsburg, Chr. Von der, 1973, Self-Organization of orientation sensitive cells in the striate cortex. Kybernetic 14, pp 85-100.

[5] Pitman E, 1979, Some basic theory for statistical inference. Chapman and Hall, London.