Prudent Fraud Detection in Internet Banking

Oarabile Omaru Maruatona
Internet Commerce Security Laboratory
University of Ballarat
o.maruatona@icsl.com.au

Peter Vamplew
School of Science, IT and Engineering
University of Ballarat
p.vamplew@ballarat.edu.au

Richard Dazeley
School of Science, IT and Engineering
University of Ballarat
r.dazeley@ballarat.edu.au

Abstract- **Most commercial Fraud Detection components of Internet banking systems use some kind of hybrid setup usually comprising a Rule-Base and an Artificial Neural Network. Such rule bases have been criticised for a lack of innovation in their approach to Knowledge Acquisition and maintenance. Furthermore, the systems are brittle; they have no way of knowing when a previously unseen set of fraud patterns is beyond their current knowledge. This limitation may have far reaching consequences in an online banking system. This paper presents a viable alternative to brittleness in Knowledge Based Systems; a potential milestone in the rapid detection of unique and novel fraud patterns in Internet banking. The experiments conducted with real online banking transaction log files suggest that Prudent based fraud detection may be a worthy alternative in online banking.**

Keywords- RDR, Prudence, RM, RDM, Online banking Fraud Detection

I. INTRODUCTION

Traditionally, banks do not publicise specific details of their Fraud Detection (FD) systems [1]. Despite this fact, there have recently been a number of software vendors who publicised some information about their Internet banking FD tools and the market share of such tools in commercial banks worldwide. For example, the Proactive Risk Manager (PRM) was reported to be used by the top 20 banks in the world and in more than 40 countries [2]. Similarly, the SAS Fraud Management system is reportedly used in debit and credit card FD solutions at more than 43000 online banking sites [3]. Although the information publicised is relatively simplified and primarily released for marketing purposes, the details exposed give a good indication of the generic architecture used by commercial banks to detect fraud in online banking transactions. This research has found out from different software vendors' white papers that there is a consistent use of a combination of a Rule-Based System (RBS) with an Artificial Neural Network (ANN). The Falcon Fraud Manager, a payment card fraud detection tool comprises an RBS and ANN as its main components [4]. The PRM white paper likewise reports that the PRM debit/credit card and internet banking system uses a hybrid structure featuring an RBS and ANN [2]. In a similar fashion, the SAS Fraud Management system's architecture is reported to mainly include an RBS and an ensemble of Self Organising Neural Networks (SONNA) [3].

The RBS approaches employed by the FD systems profiled earlier have been criticised on a number of aspects. One of the disapprovals is based on the knowledge acquisition bottleneck phenomenon, where the process of transferring knowledge to an ES is indirect (involves an expert and knowledge engineer), labour intensive and usually restricted to a specific context [5]. Maintenance in these systems has also been described as laborious, costly and difficult [6], [5]. This is because a typical adjustment in such an RBS would involve a knowledge engineer and the domain expert. Usually if the original knowledge acquisition involved an expert and a knowledge engineer, then maintaining the system will require the expert to ensure that the latest changes do not render the old knowledge invalid. These factors, and the fact that maintenance in these cases is performed as an additional task to knowledge acquisition [7], make maintaining these systems a time consuming and costly endeavour. Another main technical limitation of these commercial FD systems is that they are brittle. Brittleness occurs when a KBS does not realise when its knowledge is

978-0-7695-4940-8/12 $26.00 © 2012 IEEE
DOI 10.1109/CTC.2012.13

inadequate for a particular case [8]. Brittleness, also likened to a lack of common sense in expert systems [9] results in systems giving illogical conclusions to cases they have not been trained for. For online banking FD systems, this limitation poses a serious threat with a potential loss of significant amounts of money.

Ripple Down Rules (RDR) was introduced in around 1988 to counteract KA and maintenance limitations of traditional KBS [10]. RDR is an incremental technique whereby KA and maintenance are essentially integrated and usually not requiring the additional services of a knowledge engineer. RDR has since been used in commercial applications including in intelligent web browsers, help desk systems, online shopping systems and numerous more applications [11]. RDR was originally developed to produce single classifications but was extended into Multiple Class RDR (MCRDR), a method essentially similar to RDR but with an added capability to produce multiple classifications. This paper presents two RDR techniques previously unused in online banking FD applications. Rated MCRDR (RM) and Ripple Down Models (RDM) are two successful Prudence Analysis (PA) techniques. PA is an RDR technique and was discovered as an alternative to KBS brittleness. A brittle KBS does not realise when its knowledge is inadequate for a particular case [8]. Consequently, a prudent KBS is one with a mechanism to issue warnings or alerts whenever a current case is beyond the system's expertise. A prudent system therefore has the ability to signal alerts whenever it encounters a case beyond its range of expertise. The rest of the paper is arranged as follows: sections 2 and 3 explain the RM and RDM techniques respectively, section 4 presents a number of experimental results showing the capabilities of RM and RDM using real Internet banking transaction logs and section 5 concludes the paper.

II. RATED MCRDR (RM)

RM is a hybrid approach and combines MCRDR with an ANN [12]. The method's functionality is based on [5]'s premise that if patterns of fired MCRDR rules were collected, they can provide an additional context about a domain. The existence of such a pattern is due to a conscious or subconscious relationship between the rules in the expert's mind. Groups of these patterns can then be assigned values representing their contribution to different particular tasks [5]. RM has a MCRDR output simplifying mechanism which decodes MCRDR conclusions into indexes of binary inputs for the ANN. The MCRDR output generally can be a set of terminating rules, or the classifications produced for a given case. For this project, the set of terminating rules was used as the MCRDR output for every case. This particular choice of MCRDR output was based on recommendations from [5]'s previous experiments with a range of different outputs .The terminating rules are indexed and assigned a 0 or 1 value depending on whether the particular rule was fired for the current case or not. Figure 1 illustrates an overview of RM.

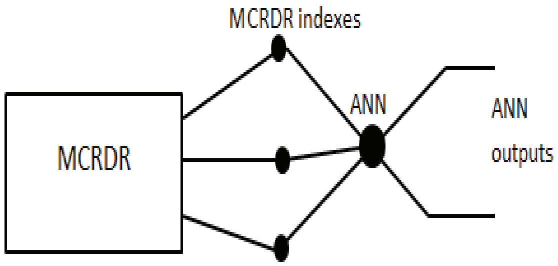

Figure 1. RM schematic

In RM, the indexed binary set is fed into an ANN such that each firing terminating rule produces a 1 input for the ANN, and a 0 input for non firing terminal rules. For example, in the diagram at Figure 1, if only the first and third terminating rules fired, then the resulting indexed binary word would be 101 which is also the input for the ANN. The ANN uses two main learning approaches. If the input does not change (or if there are no new rules added to MCRDR), a standard back-propagation algorithm with a sigmoid thresholding function is engaged. If a new rule is added in MCRDR, an additional input is created for the ANN. Additional inputs may pose threats to the ANN's learning and conservation of already learned information. In addressing this threat, new shortcut connections are introduced from the newly created input to each output node. These shortcut connections are assigned weights that enable the network to retain previously learned content and resolve to the correct class immediately [12]. The shortcut weights are calculated using the single step initialisation formula (see equation 1)

$$w = ƶ\left(\log\left(\frac{fnet + \partial + 0.5}{0.5 - (fnet + \partial)}\right)\right) - (((\sum A) + (\sum B))/m)$$

(1)

A and B are the weighted sums at the hidden and output nodes respectively, ƶ is the step distance modifier in the range of 0 to 1. The step modifier is the rate of adjustment of for the new features and determines how quickly the shortcut weights adjust to the correct output. m is the number of newly added inputs and ∂ is the sum of differences between the network calculated outputs and the target outputs (or error sum value) at an output neuron.

The ANN learns the patterns of the fired rules for each classification as the MCRDR generates different classifications. After learning, the ANN is capable of detecting contradictions or misclassifications from the MCRDR output. The prudence part of the system is engaged each time there are inconsistencies between the two components, and a warning is issued. The system issues a warning each time the MCRDR and the ANN produce different classification.

III. RIPPLE DOWN MODELS (RDM)

RDM has two main components, the MCRDR part and an outlier detection component. RDM first generates MCRDR outputs to the complementary outlier detection component. In RDM, the output passed to the outlier detector is not necessarily classifications but models. A model consists of situated profiles and each situated profile contains a number of profiles corresponding to the number of attributes in a case. In a numerical dataset for example, a profile for a given attribute could be the minimum and maximum values for that attribute. After RDR generates a model, it is passed to the outlier detector. RDM has two outlier detection functions: the Outlier Estimation with Backward Adaptation (OEBA) for continuous attributes and the Outlier Detection for Categorical Attributes (OECA) for discrete attributes [13].

In OEBA, a case's attribute profiles are grouped as a Situated Profile and organised according to the conclusions generated by RDR. For example, an OEBA Situated Profile will contain minimum and maximum values for each attribute for the corresponding RDR classification. For each classification produced by RDR, a Model comprising the Situated Profile(s) is returned to the outlier detection component. A Range Probability for each profile in the Situated Profile is then calculated. A range Probability indicates the likelihood of the case value being a part of the profile. Depending on the case values' similarity or difference to the each profile's upper and lower bound, and whether the Range Probability is less or greater than a set threshold, an outlier is flagged. For this project, if outliers are identified in more than 2/5 of a case's attributes, then an anomaly is flagged. Ideally, an anomaly should be only flagged for incorrect classifications by RDR. If an outlier was flagged incorrectly, then Backward Adaptability adjusts the appropriate profiles' minimum and maximum values.

In OECA, each profile keeps a set of an attribute's observed values and a corresponding *M* value. The *M* value is computed for each attribute value and denotes a probability of the value being observed after *v* other values from *k* observations [9]. A New Value Ratio (NVR) is also computed for each case value and the returned profile. The NVR is the ratio of the current attribute's *M* value and the *M* value for the last updated value in the profile [9]. An anomaly is flagged when the NVR of a case is greater than a set threshold. As in OEBA, an outlier is flagged when a case has more than 2/5 anomalies. Figure 2 shows the general architecture of RDM. Ideally, OEBA/OECA flags outliers when the RDR conclusion is not right. This way, a warning will be issued at the right time. As in RM, the prudence of RDM depends on the effectiveness of the warning system. An overview of RDM is shown in Fig. 2.

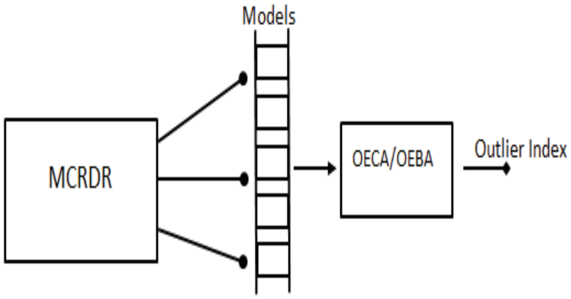

Figure 2. Overview of RDM components

IV. EVALUATION APPROACH

Two common obstacles in developing and testing KBS are the difficulty of the process itself and cost of actual expertise. Usually, it is often hard (to test a KBS) and expensive to get a readily available expert for controlled tests [14], [15]. In RDR, a popular workaround has been the use of simulated experts. Simulated experts have been used extensively in testing RDR methodologies [5], [14]. A simulated expert is essentially a mapping of the cases (input) to the conclusions (correct output) used to confirm (and train) the KBS being built. For each dataset, the simulated expert loads a rule-set generated from a data mining tool such as See5 [16]. Usually, a smart or clever simulated expert will have a perfect mapping of rules to their corresponding classes. In some cases, the simulated expert is not a perfect mapping of rules to their correct classes. Irrespective of the simulated expert's level of expertise, the objective is usually to determine how well the expert system learns and not just how accurate the teacher (simulated expert) is. If the expert system learns well from an imperfect simulated expert, it will usually learn just as well from a perfect simulated expert. In this paper, simulated experts with varying levels of expertise were used. The results give a good indication of the system's performance relative to the quality of the simulated expert. The Relative Accuracy (RA) of a system in this context refers to the proportion of system's accuracy relative to the simulated expert's accuracy. This measure determines how well the system learns. RA is calculated using the simple formula below:

$$RA = \left(\frac{SA}{SEa}\right)\% \qquad (2)$$

Where SA is the simple Accuracy (TP+TN/FP+FN+TP+TN), SEa is the simulated expert's accuracy.

A lack of real data for the development of practical fraud detection systems has often been cited as one the biggest obstacles in the advancement of sophisticated fraud detection tools [17]. The unavailability of real data also limits the commercial applicability and relevance of the methods and algorithms with potential use in FD. Consequently, it becomes a hard endeavour to justify the application of a new technique in a domain in which the technique has not been tested. In the few cases where real data is used to test a given method, the results reveal how much work is yet to be done to have robust FD tools. The results also indicate that some techniques have immense potential in a particular FD domain. Two sets of data were used for these experiments; the Poker Hand dataset and a real dataset of commercial online banking transactions. The Poker-Hand dataset, available from the UCI repository [18] was used to demonstrate the general capability of PA. The other dataset was collected from log files of a commercial online banking system. The de-personalized and obfuscated data effectively comprises more than 15 features of users' online banking transactions. These transaction log files were divided into unique groups of exclusively numerical and exclusively categorical features. The purpose of dividing the dataset into categorical and numerical sets was to compare RDM's outlier detection components, OECA for categorical data and OEBA for numerical data. Table 1 presents a description of the datasets and the simulated experts used with them. The table describes the number of instances (cases) in each dataset, whether the data is numerical or categorical, how many rules the data's simulated expert has and the accuracy of the simulated expert.

TABLE 1. DATASET DESCRIPTION

Dataset	Cases	Type	SE Rules	SE Acc.
Poker	2000	num	83	60%
OT_A	1755	num	72	60%
OT_B	1755	cat	49	73%

The performance of both RM and RDM was evaluated using the Balanced Accuracy (BA) measure. The BA is calculated using each system's True Positives (TP), True Negatives (TN), False Negatives (FN) and False Positives (FP). The TP in this case is when a warning was issued correctly. TN is when a warning was not issued correctly. FN includes instances when a warning was not issued but should have. When a warning was issued incorrectly, then this is a FP [9], [15]. Equation 3 shows the formula for BA.

$$bA = TP(\alpha/T)/\beta + TN(\beta/T)/\alpha \qquad (3)$$

where α = TN + FP, β= TP+ FN and T= TN + TP + FN + FP.

Tables 2, 3 and 4 show RM and RDM's performance metrics including the BA and RA of each of the systems with the Poker-Hand dataset and the two sets of online transaction logs. In Table 2, results from the Poker-Hand are shown. Tables 3 and 4 present RM and RDM's performance metrics from the numerical and categorical online banking transaction features respectively. It must be noted that RM was tested with two different settings where z values were set to 0.01 in the first instance (RM*a*) and 0.95 in the second instance (RM*b*).

TABLE 2. THE POKER-HAND DATASET RESULTS

System	TP	TN	FP	FN	BA %	RA %
RM*a*	750	418	559	273	57.7	97.3
RM*b*	621	580	394	405	60	100
RDM	72	934	43	951	52.3	83.8

TABLE 3. NUMERICAL INTERNET TRANSACTIONS (OT_A) RESULTS

System	TP	TN	FP	FN	BA %	RA %
RM*a*	372	335	869	179	55.1	67.1
RM*b*	371	317	867	200	52.5	65.3
RDM	74	1123	61	497	39.6	100

TABLE 4. CATEGORICAL INTERNET TRANSACTIONS (OT_B) RESULTS.

System	TP	TN	FP	FN	BA %	RA %
RM*a*	1089	59	185	422	30.8	89.6
RM*b*	1066	75	174	440	35.9	89.1
RDM	1481	195	54	25	81.2	100

The accuracy of RM and RDM in general datasets are impressive as shown in Table 2. Other tests on general datasets using these two algorithms is reported in detail in [5] and [9] for RM and RDM respectively. A preliminary direct comparison of these methods also using public datasets is detailed in [19]. Apart from the online banking transaction tests reported in this paper, the two algorithms have also been tested with specialised datasets. RM has been tested in a document management system and RDM was tested with computer network traffic data [20], [9]. The performance of both RM and RDM relative to the simulated expert is noteworthy, with an average of 88% across three datasets, three simulated experts and over 200 rules. RM was particularly good on the Poker Hand dataset with average RA of 99%. This asserts Dazeley [5]'s proposal that patterns of fired MCRDR rules provide an additional context about a domain. However, the validity of such a context depends largely on the domain (or type of data) as shown by RM's average performance with the numerical online banking transactions (OT_A). RDM's RA was perfect in the two online banking transaction datasets. RDM's idea of homogenising cases into models seems to be especially effective for both the numerical and categorical parts of online banking transactions. The system effectively learnt everything from the simulated expert in both datasets. In the categorical transactions, RDM outperformed the simulated expert with a BA of 81% compared to the simulated expert's accuracy of 73%. Overall, the high RA across all datasets indicates that given a good simulated expert, both RM and RDM have a high learning capability. In terms of the detecting frauds in online banking transactions, these systems have proved their relevance and potential in a commercial setting.

V. CONCLUSIONS

In most commercial systems, a hybrid approach to FD seems to be a reliable path as proved by the consistent use of a Rule-Based system and neural network in most commercial FD tools. These RBS use a less innovative, slow and potentially costly way of adding and updating knowledge. Furthermore, the systems are incapable of warning an administrator in case a transaction with previously unseen features is conducted. For an online banking system, the potential loss of funds presented by such transactions cannot be ignored. The RDR based Prudence methods have been shown to overcome the KBS limitations stated earlier. This paper presented an application of two of these methods in commercial online banking FD. Test results indicate that RM and RDM are a practical and viable alternative in detecting internet banking fraud. The tests conducted in this paper are part of a bigger research project whose objective is to develop an Integrated Prudence Analysis (IPA) method for the rapid detection of fraud in internet banking systems. Future work includes testing the IPA with more transaction from an online banking system. The research work and experiments completed so far suggest that RDR Prudence presents a potentially useful fraud detection method in commercial Internet banking systems.

VI. REFERENCES

[1] R Bolton and D Hand, "Statistical fraud detection: A Review ," *Statistical Science*, vol. 17, no. 3, pp. 235-249, August 2002.

[2] ACI Worldwide. (2011, January) ACI Worldwide. [Online]. http://www.aciworldwide.com/igsbase/igstemplate.cfm/SRC=DB/SRCN=/GnavID=15

[3] SAS, "SAS Fraud Management," SAS Institute Inc, Technical Report 2007.

[4] FICO. (2011, January) FICO. [Online]. http://www.fico.com/en/products/dmapps/pages/fico-falcon-fraud-manager.aspx

[5] R Dazeley, "To the Knowledge Frontier and Beyond: A Hybrid System for Incremental Contextual-Learning and Prudence Analysis," University of Tasmania, PhD Thesis 2007.

[6] F Hayes-Roth and N Jacobstein, "The State of Knowledge-Based Systems," *Communications of the ACM*, pp. 27-39, 1994.

[7] D Richards, "Knowledge-Based System Explanation: The Ripple-Down Rules Alternative," *Knowledge and Information Systems*, pp. 2-25, 2003.

[8] P Compton, P Preston, G Edwards, and B Kang, "Knowledge based systems that have some idea of their limits," *CIO*, vol. 15, pp. 57-63, June 1996.

[9] A Prayote, "Knowledge Based Anomaly Detection," University of New Soth Wales, PhD Thesis 2007.

[10] P Compton and K Horn, "Maintaining an Expert System," *Applications of Expert Systems*, pp. 366-385, 1989.

[11] D Richards, "Two decades of Ripple Down Rules research," *The Knowledge Engineering Review*, vol. 24, no. 2, pp. 159-184, 2009.

[12] R Dazeley and B Kang, "Detecting the Knowledge Boundary with Prudence Analysis," in *AI 2008*, Auckland, 2008, pp. 482-488.

[13] A Prayote and P Compton, "Detecting Anomalies and Intruders," in *AI06*, Hobart, 2006.

[14] P Compton, P Preston, and B Kang, "The Use of Simulated Experts in Evaluating Knowledge Acquisition," in *Knowledge Acquisition for Knowldge Based Systems Workshop*, Banff, 1995.

[15] R Dazeley, S Park, and B Kang, "Online knowledge validation with prudence analysis in a document management application," *Expert Systems with Applications*, vol. 38, pp. 10959-10965, 2011.

[16] Rulequest Research. (2012) Data Mining Tools See5 and C5.0. [Online]. http://rulequest.com/see5-info.html

[17] Varun Chandola, Arindam Banerejee, and Vipin Kumar, "Anomaly Detection : A Survey," *ACM Computing Surveys*, pp. 1-72, September 2009.

[18] R Cattral and F Oppacher. (2007, January) Poker Hand Data Set. Dataset. [Online]. http://archive.ics.uci.edu/ml/datasets/Poker+Hand

[19] O.O Maruatona, P Vamplew, and R Dazeley, "RM and RDM, a Preliminary Evaluation of two Prudent RDR Techniques," in *The Pacific Rim Knowledge Acquisition Workshop*, Kuching, 2012.

[20] R Dazeley, S.S Park, and B.H Kang, "Online knowledge validation with prudence analysis in a document management application," *Expert Syst. Appl.*, pp. 10959-10965, 2011.