Detecting credit card fraud by ANN and logistic regression

Article · June 2011

DOI: 10.1109/INISTA.2011.5946108

2 authors:

Yusuf Sahin
Marmara University
4 PUBLICATIONS **256** CITATIONS
SEE PROFILE

Ekrem Duman
Ozyegin University
51 PUBLICATIONS **916** CITATIONS
SEE PROFILE

Detecting Credit Card Fraud by ANN and Logistic Regression

Yusuf Sahin[1] and Ekrem Duman[2],

[1] Department of Elec.& Electronics Eng., Marmara University, Goztepe Campus, 34722, Istanbul, Turkey
ysahin@marmara.edu.tr
[2]Department of Industrial Eng., Dogus University, Acibadem, Istanbul, Turkey
eduman@dogus.edu.tr

Abstract. With the developments in information technology and improvements in communication channels, fraud is spreading all over the world, resulting in huge financial losses. Though fraud prevention mechanisms such as CHIP&PIN are developed, these mechanisms do not prevent the most common fraud types such as fraudulent credit card usages over virtual POS terminals through Internet or mail orders. As a result, fraud detection is the essential tool and probably the best way to stop such fraud types. In this study, classification models based on Artificial Neural Networks (ANN) and Logistic Regression (LR) are developed and applied on credit card fraud detection problem. This study is one of the firsts to compare the performance of ANN and LR methods in credit card fraud detection with a real data set.

Keywords: Credit card fraud detection, ANN, logistic regression, classification

1 Introduction

Fraud can be defined as wrongful or criminal deception intended to result in financial or personal gain [1], or to damage another individual without necessarily leading to direct legal consequences. The two main mechanisms to avoid frauds and losses due to fraudulent activities are fraud prevention and fraud detection systems. Fraud prevention is the proactive mechanism with the goal of disabling the occurrence of fraud. Fraud detection systems come into play when the fraudsters surpass the fraud prevention systems and start a fraudulent transaction. Nobody can understand whether a fraudulent transaction has passed the prevention mechanisms. Accordingly, the goal of the fraud detection systems is to check every transaction for the possibility of being fraudulent regardless of the prevention mechanisms, and to identify fraudulent ones as quickly as possible after the fraudster has begun to perpetrate a fraudulent transaction. A review of the fraud detection systems can be found in [2-5].

With the developments in the information technology and improvements in the communication channels, fraud is spreading all over the world with results of huge financial losses. Though fraud can be perpetrated through many types of media, including mail, wire, phone and the Internet, online media such as Internet are the most popular ones. Because of the international availability of the web and ease with

which users can hide their location and identity over Internet transactions, there is a rapid growth of committing fraudulent actions over this medium. Furthermore, with the improvements in the bandwidth of internetworking channels, fraudsters have the chance to form fraud networks among themselves through information change and collaboration all over the world. As a result, frauds committed over internet such as online credit card frauds become the most popular ones because of their nature.

Credit card frauds can be made in many ways such as simple theft, application fraud, counterfeit cards, never received issue (NRI) and online fraud (where the card holder is not present). In online fraud, the transaction is made remotely and only the card's details are needed. A manual signature, a PIN or a card imprint are not required at the time of purchase. Though prevention mechanisms like CHIP&PIN decrease the fraudulent activities through simple theft, counterfeit cards and NRI; online frauds (internet and mail order frauds) are still increasing in both amount and number of transactions. There has been a growing amount of financial losses due to credit card frauds as the usage of the credit cards become more and more common. Many papers reported huge amounts of losses in different countries [2, 6-7]. According to Visa reports about European countries, about 50% of the whole credit card fraud losses in 2008 are due to online frauds.

Credit card fraud detection is an extremely difficult, but also popular problem to solve. Firstly, there comes only a limited amount of data with the transaction being committed, such as transaction amount, date and time, address, merchant category code (MCC) and acquirer number of the merchant. There are millions of possible places and e-commerce sites to use a credit card which makes it extremely difficult to match a pattern. Also, there can be past transactions made by fraudsters which also fit a pattern of normal (legitimate) behavior [8]. Moreover, the problem has many constraints. First of all, the profile of normal and fraudulent behavior changes constantly. Secondly, the development of new fraud detection methods is made more difficult by the fact that the exchange of ideas in fraud detection, especially in credit card fraud detection is severely limited due to security and privacy concerns. Thirdly, data sets are not made available and the results are often censored, making them difficult to assess. Because of this problem, there is no chance of benchmarking for the models built. Even, some of the studies are done using synthetically generated data [9-10]. None of the previous studies with real data in the literature give details about the data and the variables used in classifier models. Fourthly, credit card fraud data sets are highly skewed sets with a ratio of about 10000 legitimate transactions to a fraudulent one. Lastly, the data sets are also constantly evolving making the profiles of normal and fraudulent behaviors always changing [2-4].

The most commonly used fraud detection methods are ANNs, rule-induction techniques, decision trees, Support Vector Machines (SVM), LR, and meta-heuristics such as genetic algorithms, k-means clustering and nearest neighbor algorithms. These techniques can be used alone or in collaboration using ensemble or meta-learning techniques to build classifiers.

Past data in the credit card data warehouses are used to form a data mart representing the user profiles of the customers. These profiles consist of variables each of which discloses a behavioral characteristic of the customer. These variables may show the spending habits of the customers with respect to geographical locations, days of the month, hour of the day or MCCs. Later on, these variables are used to

build a model to be used in the fraud detection systems to distinguish fraudulent activities which show significant deviations from the profile of the customer stored in the data-mart.

In this study, a credit card fraud detection system based on a number of ANN and LR methods is developed. In this system, each account is monitored separately using suitable descriptors, and the transactions are attempt to be identified and flagged as legitimate or normal. The identification will be based on the suspicion score produced by the classifier models developed. When a new transaction is going, the classifier can predict whether the transaction is normal or fraud.

The rest of this paper is organized as follows: Section 2 gives some insights to the structure of credit card data. Section 3 gives a brief explanation of the classification methods used to develop the classifier models of the credit card fraud detection system given in this paper, and continues with giving the details of our approach. Section 4 gives the results and a short discussion about the results. Section 5 concludes the study and shows directions for future work.

2 Structure of the Credit Card Data

The credit card data used in this study are taken from a national bank's credit card data warehouses with the required permissions. The past data in the credit card data warehouses are used to form a data mart representing the card usage profiles of the customers. Though some of the customers may have more than one credit card, each card is taken as a unique profile because customers with more than one card generally use each card for a different purpose. Every card profile consists of variables each of which discloses a behavioral characteristic of the card usage. These variables may show the spending habits of the customers with respect to geographical locations, days of the month, hours of the day or merchant category codes (MCC) which show the type of the merchant where the transaction takes place. Later on, these variables are used to build a model to be used in the fraud detection systems to distinguish fraudulent activities which show significant deviations from the card usage profile stored in the data-mart.

The number of transactions for each card differs from one to other; however, each transaction record is of the same fixed length and includes the same fields. Hand and Blunt gave a detailed description of the characteristics of credit card data [11]. These fields range from the date and hour of the transaction to the amount, transaction type, MCC code, address of the merchant where the transaction is done and etc. The date and hour of the transaction record shows when the transaction is made. Transaction type shows whether this transaction is a purchase or a cash-advance transaction. MCC code shows the type of the merchant store where the transaction takes place. These are fixed codes given by the members of the VISA International Service Association. However; however, many of these codes form natural groups. So, instead of working with hundreds of codes, we grouped them into 25 groups according to their nature and the risk of availability to commit a fraud. The goods or services bought from merchant stores in some MCC codes can be easily converted to cash. As a result, transactions belonging to these MCC codes are more open to fraud and more risky

from the transactions belonging to others. The grouping of the MCC codes are done according to both the number of the fraudulent transactions made belonging to each MCC code and the interviews done with the personnel of the data supplier bank with domain expertise about the subject.

The distribution of the data with respect to being normal or fraudulent is highly imbalanced with a ratio of about 20000 normal transaction records to one fraudulent transaction record. So, to enable the models to learn both types of profiles, some under sampling or oversampling techniques should be used. Instead of oversampling the fraudulent records by making multiple copies or etc., we use stratified sampling to under sample the legitimate records to a meaningful number.

3 Approach

There are a lot of studies done on credit card fraud detection. The general background of the credit card systems and non-technical knowledge about this type of fraud can be learned from [12] and [13]. Most of the credit card fraud detection systems are using supervised algorithms such as neural networks [7, 9, 14-21]. In this study, the performance of classifier models built by using a number of different ANN methods and the well-known LR methods are compared. To the best of our knowledge, there is only one previous study in the literature applying both ANN and LR to the credit card fraud detection problem [20]. This study is a more comprehensive one using many ANN and LR methods on four different data sets with different fraudulent transaction / legitimate transaction ratios.

ANN refers to a group of nonlinear, statistical modeling techniques inspired and derived from the structure of the human brain. ANNs can be used in modeling any complex transactional pattern, so they are well suited to the credit card fraud detection problem [21]. The basic element of a neural network is a neuron which accepts many inputs, sums them, applies a (usually nonlinear) transfer function, and generates the result, either as a model prediction or as input to other neurons. A neural network is a structure of many such neurons connected in a systematic way. The most well known neural networks used are feed-forward neural networks, also known as multilayer perceptrons. The neurons in such networks (sometimes called units) are arranged in layers. Typically, there is one layer for input neurons (the input layer), one or more layers of internal processing units (the hidden layers), and one layer for output neurons (the output layer). Each layer is fully interconnected to the preceding layer and the following layer if there is any. The connections between neurons have weights associated with them, which determine the strength of influence of one neuron on another. Information flows from the input layer through the processing layer(s) to the output layer to generate predictions. By adjusting the weights of the connections during training to match predictions to target values for specific records using a proper error function, the network generate better and better predictions by each iteration. However, as the number of iterations reaches a certain level, the network begins to memorize the input data and overfitting occurs. So, the number of iterations must be neither too small so that the network can learn the patterns, nor too large so that overfitting does not occur.

Logistic regression is a well-established statistical method for predicting binomial or multinomial outcomes. Multinomial Logistic Regression algorithm can produce models when the target field is a set field with two or more possible values. See below for more information. Binomial Logistic Regression algorithm is limited to models where the target field is a flag, or binary field.

In this study, the ANN and LR classifier models are built using IBM SPSS PASW Modeler with the relevant modeling methods implemented in it. There are six different ANN methods defined in PASW Modeler. These are RBFN, Quick, Dynamic, Multiple, Prune and Exhaustive Prune methods. A radial basis function network (RBFN) is a special kind of neural network. It consists of three layers: an input layer, a hidden layer (also called a receptor layer), and an output layer. The input and output layers are similar to those of a multilayer perceptron. However, the hidden or receptor layer consists of neurons that represent clusters of input patterns, similar to the clusters in a k-means model. These clusters are based on radial basis functions, or functions of the distance between the RBF's center and a vector of input values. The connections between the input neurons and the receptor neurons (receptor weights) are trained in essentially the same manner as a k-means model. When the quick method is selected, a single neural network is trained. By default, the network has one hidden layer containing max $(3*(n_i+n_o)/20)$ neurons, where n_i is the number of input neurons and n_o is the number of output neurons. The network is trained using the back-propagation method. When the dynamic method is selected, the topology of the network changes during training, with neurons added to improve performance until the network achieves the desired accuracy. There are two stages to dynamic training: finding the topology and training the final network. When the multiple method is selected, multiple networks are trained in pseudo-parallel fashion. Each specified network is initialized, and all networks are trained. When the stopping criterion is met for all networks, the network with the highest accuracy is returned as the final model. The prune method is, conceptually, the opposite of the dynamic method. Rather than starting with a small network and building it up, the prune method starts with a large network and gradually prunes it by removing unhelpful neurons from the input and hidden layers. Pruning proceeds in two stages: pruning the hidden neurons and pruning the input neurons. The two-stage process repeats until the overall stopping criteria are met. The exhaustive prune method is a special case of the prune method.

Logistic regression methods built in PASW Modeler are stepwise, enter, forwards and backwards. The stepwise method can be used in multinomial LR; however, the other methods can be used in both binomial LR and multinomial LR.

Before starting the modeling phase, the collected data is pre-processed. As mentioned earlier, the distribution of data with respect to the classes is highly imbalanced. The time period that is used to build our sample included 978 fraudulent records and 22 million normal ones with a ratio of about 1:22500. So, stratified sampling is used to under sample the normal records so that the models have chance to learn the characteristics of both the normal and the fraudulent records' profile. To do this, first of all, the variables that are most successful in discriminating the fraudulent and legitimate transactions are found. Then, these variables are used to form stratified samples of the legitimate records. Later on, these stratified samples of the legitimate records are combined with the fraudulent ones to form three samples

with different fraudulent to normal record ratios. The first sample set has a ratio of one fraudulent record to one normal record; the second one has a ratio of one fraudulent record to four normal ones; and the last one has the ratio of one fraudulent to nine normal ones.

The variables which form the card usage profile and the methods used to build the model make the difference in the fraud detection systems. Our aim in defining the variables used to form the data-mart is to discriminate the profile of the fraudulent card usage by the fraudsters from the profile of legitimate (normal) card usage by the card holders. We will be content with mentioning about the type of variables used but regarding the privacy, confidentiality and security concerns, we are not allowed to talk on the full list of variables (a variable is a certain level of deviation from a personal average statistics). The variables are one of the five main types: all transactions statistics, regional statistics, MCC statistics, daily amount statistics and daily number of transactions statistics.

The chosen methods to build classifier models are RBFN, Quick, Dynamic, Multiple, Prune and Exhaustive Prune methods from the ANN algorithm family implemented in PASW Modeler; and stepwise, enter, forwards and backwards methods both in multinomial logistic regression (MLR) and binomial logistic regression (BLR). Stepwise method can only be used in MLR. All these methods are used to develop models using the three data sets. These methods and the parameters used with these methods are given in Table 1. The dynamic method for ANN takes no parameters from the user and runs with default values.

Table 1. Input parameters for ANN classifier models

Model		Quick	Prune	Multiple		RBFN
Model Parameters	Alpha	0,9	0,9	0,9	Alpha	0,9
	Initial Eta	0,3	0,3	0,3	RBF Clusters	20
	High Eta	0,1	0,1	0,1	Eta	0,4
	Low Eta	0,01	0,01	0,01	RBF Overlap	1,0
	Eta Decay	30	30	30	-	-
	Layer1	20	20	2	-	-
	Layer2	15	15	20|27|22	-	-
	Layer3	10	10	3|5|4	-	-
	Persisten	200	100	200	Persistence	30
	Sample%	70%	70%	70%	Sample%	70%

For the binomial and multinomial regression models, score is used as the entry criterion and likelihood ratio is used as the removal criterion. A value of 0.05 is required for entry; and a value of 0.1 is required for the removal of a variable from the function.

As mentioned above, three different sets are formed to compare the performances of the methods under these different conditions. In the first set, there is one normal transaction for each fraudulent one. In the second set, there are four normal transactions for each fraudulent one and there are nine normal ones for each fraudulent one in the third set. For each sample, 70% of the data, both 70% of the normal transactions and 70% of the fraudulent transactions, are taken as the training

set for the models; and 30% of the data are taken as the test set to evaluate the performance of the models developed. The training set and test set of each sample is designed to have the same fraudulent records as the sets of other samples. The training and test set sizes are given in Table 2. Every model for a sample uses the same training and test data set so that an exact comparison of the performance of the models for each sample becomes possible.

Table 2. Training and test set sizes for each sample

Samples		# of Records	
		Training Set	Test Set
1F-to-1N	Normal	681	292
	Fraud	681	292
1F-to-4N	Normal	2723	1168
	Fraud	681	292
1F-to-9N	Normal	6130	2626
	Fraud	681	292

4 Results and Discussions

In this study, 13 alternative models based on ANN and LR were built using the relevant training data set for each sample. To evaluate these models, we used the remaining transactions in the relevant test sets for each model. The fraudulent transactions in the test sets of the samples are identical. Accuracy rates were used to describe the usefulness of the models. Accuracy is probably the most commonly used metric to measure the performance of targeting models in classification applications. However, the number of fraudulent transactions caught is also an important performance indicator. The performance of the classifier models built using the mentioned ANN and LR methods with respect to accuracy, and the number of fraudulent transactions in the test set caught by the models are given in Table 3. The left column shows the methods used to build the models, the columns named as "Training" show the prediction accuracy of the models on the training data set of the given samples, and the columns named as "Testing" show the prediction accuracy of the models on the testing data set of the given samples. The columns named as "Frauds Caught" show the number of fraudulent transactions in the data set caught by the model built.

From Table 3, it is clear that ANN models outperform the LR models when their performances are compared over the test data sets. Performances of the models are similar over the training data sets, so it just shows that LR models overfit more the training data. As the number of the training data increases, this overfitting behavior becomes more remarkable and due to the biased characteristic of the class distribution of the training data sets, number of frauds caught in the test set decreases though accuracy of the models increase. In credit card fraud detection number of fraudulent transactions caught is more important than the accuracy of the model. Accordingly, models built using the set with one normal to one fraud ratio are more successful than

the models built using the other sample sets. Among the MLR models, Stepwise MLR is the champion in both accuracy performance and catching fraudulent transactions except the last case.

Table 3. Performance of classifiers with respect to accuracy and # of frauds caught

Classifier Model	Set-1F-To-1N			Set-1F-To-4N			Set-1F-To-9N		
	Training	Testing	Frauds Caught	Training	Testing	Frauds Caught	Training	Testing	Frauds Caught
Quick (ANN)	90.46%	90.22%	264	91.86%	92.05%	244	94.51%	94.69%	184
Dynamic (ANN)	92.29%	90.57%	268	92.39%	92.47%	249	94.83%	94.55%	186
Multiple (ANN)	90.82%	89.88%	265	92.77%	92.12%	245	95.13%	94.21%	193
Prune(ANN)	92.66%	88.68%	258	93.65%	92.40%	239	94.45%	94.89%	205
Exh. Prune (ANN)	90.46%	88.85%	259	91.92%	92.19%	250	95.60%	95,00%	201
RBFN (ANN)	86.27%	87.48%	265	90.98%	91.92%	229	93.36%	94.31%	154
Stepwise (MLR)	90.68%	89.02%	255	93.16%	92.26%	233	95.35%	94.28%	186
Enter (MLR)	90.97%	77.19%	210	92.83%	89.11%	189	94.39%	92.77%	157
Enter (BLR)	90.97%	77.19%	210	92.83%	89.11%	189	94.39%	92.77%	157
Forwards (MLR)	91.34%	88.34%	249	93.89%	91.71%	218	95.39%	94.65%	190
Forwards (BLR)	90.01%	88.51%	253	91.72%	89.73%	195	95.40%	94.65%	190
Backwards (MLR)	90.97%	77.19%	210	92.83%	89.11%	189	94.39%	92.77%	157
Backwards (BLR)	90.38%	78.90%	209	93.86%	91.58%	215	95.40%	94.62%	190

5 Conclusions

As usage of credit cards become more and more common in every field of the daily life, credit card fraud has become much more rampant. To improve security of the financial transaction systems in an automatic and effective way, building an accurate and efficient credit card fraud detection system is one of the key tasks for the financial institutions. In this study, 13 classification methods were used to build fraud detecting models. The work demonstrates the advantages of applying the data mining techniques including ANN and LR to the credit card fraud detection problem for the purpose of reducing the bank's risk. The results show that the proposed ANN classifiers outperform LR classifiers in solving the problem under investigation. However, as the distribution of the training data sets become more biased, the performance of all models decrease in catching the fraudulent transactions.

As a future work, instead of making performance comparisons just over the prediction accuracy, these comparisons will be extended to include the comparisons over other performance metrics, especially the cost based ones.

References

1. Hobson, A.: The Oxford Dictionary of Difficult Words. The Oxford University Press, New York (2004)
2. Bolton, R. J., Hand, D. J.: Statistical fraud detection: A review. Statistical Science 28(3), 235--255 (2002)

3. Kou, Y., Lu, C.-T., Sirwongwattana, S., Huang, Y.-P.: Survey of fraud detection techniques. In: Proceedings of the 2004 IEEE International Conference on Networking, Sensing and Control, Taipei, Taiwan (2004)
4. Phua, C., Lee, V., Smith, K., Gayler, R.: A comprehensive survey of data mining-based fraud detection research. Artificial Intelligence Review (2005)
5. Sahin, Y., Duman, E.: An overview of business domains where fraud can take place, and a survey of various fraud detection techniques. In: Proceedings of the 1st International Symposium on Computing in Science and Engineering, Aydin, Turkey (2010)
6. Leonard, K. J.: Detecting credit card fraud using expert systems. Computers and Industrial Engineering, 25, (1993)
7. Ghosh, S., Reilly, D. L.: Credit card fraud detection with a neural network. In: Proceedings of the 27th Hawaii International Conference on system Sciences, (1994)
8. Mena, J.: Investigate Data Mining for Security and Criminal Detection, Butterworth-Heinemann, Amsterdam (2003)
9. Aleskerov, E., Freisleben, B., Rao, B.: CARDWATCH: A neural network based data mining system for credit card fraud detection. In: Computational Intelligence for Financial Engineering, 220-226 (1997)
10. Chen, R., Chiu, M., Huang, Y., Chen, L.: Detecting credit card fraud by using questionnaire-responded transaction model based on SVMs. In: Proceedings of IDEAL2004 (2004)
11. Hand, D. J., Blunt, G.: Prospecting gems in credit card data. IMA Journal of Management Mathematics, 12 (2001)
12. Dahl, J.: Card Fraud. In: Credit Union Magazine (2006)
13. Schindeler, S.: Fighting Card Fraud in the USA. In: Credit Control, House of Words Ltd. (2006)
14. Brause, R., Langsdorf, T., Hepp, M.: Neural data mining for credit card fraud detection. In: Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence (1999)
15. Chen, R.-C., Luo, S.-T., Liang, X., Lee, V. C. S.: 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 (2005)
16. Dorronsoro, J. R., Ginel, F., Sanchez, C., Cruz, C. S.: Neural fraud detection in credit card operations. IEEE Transactions on Neural Networks, 8 (1997)
17. Hanagandi, V., Dhar, A., Buescher, K.: Density-Based Clustering and Radial Basis Function Modeling to Generate Credit Card Fraud Scores. In: Proceedings of the IEEE/IAFE 1996 Conference (1996)
18. Juszczak, P., Adams, N. M., Hand, D. J., Whitrow, C., Weston, D. J.: Off-the-peg and bespoke classifiers for fraud detection. Computational Statistics & Data Analysis. 52(9) (2008)
19. Quah, J. T., Sriganesh, M.: Real-time credit card fraud detection using computational intelligence. Expert Systems with Applications. 35(4) (2008)
20. Shen, A., Tong, R., Deng, Y.: Application of classification models on credit card fraud detection. In: International Conference on Service Systems and Service Management, Chengdu, China (2007)
21. Syeda, M., Zhang, Y., Pan, Y.: Parallel granular neural networks for fast credit card fraud detection. In: Proceedings of the 2002 IEEE International Conference on Fuzzy Systems (2