Credit Card Fraud Detection with a Neural-Network

Sushmito Ghosh and Douglas L. Reilly

Nestor, Inc.

Abstract

Using data from a credit card issuer, a neural network based fraud detection system was trained on a large sample of labelled credit card account transactions and tested on a holdout data set that consisted of all account activity over a subsequent two-month period of time. The neural network was trained on examples of fraud due to lost cards, stolen cards, application fraud, counterfeit fraud, mail-order fraud and NRI (non-received issue) fraud. The network detected significantly more fraud accounts (an order of magnitude more) with significantly fewer false positives (reduced by a factor of 20) over rule-based fraud detection procedures. We discuss the performance of the network on this data set in terms of detection accuracy and earliness of fraud detection. The system has been installed on an IBM 3090 at Mellon Bank and is currently in use for fraud detection on that bank's credit card portfolio.

Credit Card Fraud Problem

Credit card fraud is a growing problem in the credit card industry. In the US alone, losses from all types of credit card fraud are projected to exceed $850 million, representing a 10% increase in fraud losses over 1991 [1]. Though small when compared to credit card losses due to charge-offs of seriously delinquent accounts (charge-offs accounted for $8.5 billion of losses in 1992), fraud represents an increasing percentage of total charge volume, indicating that it is growing faster than the credit card business itself. From 1988 through 1991, the size of the fraud problem grew from 8 basis points to over 20.

Although credit card fraud takes many forms, there are several principal categories. Fraud due to lost cards and stolen cards generally accounts for a certain "base level" of fraud activity. The size of this base level can be affected by general economic conditions (e.g., times of high unemployment are correlated with increases in fraud losses due to lost and stolen cards). Fraud due to counterfeit cards has become a growing problem over the past several years, despite the more sophisticated card manufacturing technologies (holograms on the cards) and the encrypting of information on the magnetic stripe. Obviously, counterfeit tends to be a more organized and systematic problem in certain areas, as opposed to the more "opportunistic" and thus randomly driven nature of most fraud due to lost or stolen cards.

A special category of stolen cards has become a major problem in recent years: the theft of cards from the mail. This so-called NRI (non-receipt of issue) fraud affects issuers at the time of both new card issues as well as re-issues. Certain geographic regions of the country are more at risk than others for NRI. In some areas, the problem has been so severe that issuers have used alternate methods of card delivery (courier as opposed to mail), as well as special card activation programs. With card activation, a card is blocked (listed in the banks' authorization systems as an account for which transaction requests will be denied) until the customer calls the bank to verify card receipt. The bank uses the call to establish that the caller is the rightful cardholder by asking a small number of background information questions taken from the card member's application (if it is a new issue) or cardholder information file. Such programs, though expensive, have led to reductions in NRI losses.

Additional fraud schemes involve the submission of fraudulent applications for a card. In such cases, perpetrators obtain access to actual personal background and financial information and use this data to submit an application, specifying a mail drop to which the card should be sent. If a card is issued in such circumstances, even cardmember activation cannot prevent the card from falling into the wrong hands, since the perpetrator, who authored the phony application information, can provide this information during the phone call to activate the account.

Still another source of fraud losses is mail order/telephone order fraud. In such cases, the purchaser is not physically present before the merchant at the time of the transaction, and there is no card imprint that can be obtained as a record of the

Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences, 1994

transaction. Efforts to combat such MO/TO fraud have included verification of address information over the phone with the cardholder at the time of the purchase.

All of the above tend to be examples of cardholder or account fraud. There is also fraud that originates at the merchant. Such merchant fraud can involve the "laundering" of phony merchant receipts, garnering large sums of money for transactions that never occurred. In some cases, merchant collusion results in the merchant's establishment being used as a location at which account information is copied during the course of legitimate transactions. This information is subsequently used to produce counterfeit duplicate cards, which are then used elsewhere for fraudulent transactions. In this scheme, the collusive merchant is regarded as a "point of compromise"; all such counterfeited cards have transaction histories that can be traced back to use at the given merchant's establishment.

Additionally, there is an entire area of fraud that is often referred to as abuse. In this category, the cardholder makes purchases on the card for which he/she has no intention of paying. In some cases, this is pre-meditated activity that occurs just prior to the cardholder's filing for personal bankruptcy. Losses due to this "bankruptcy fraud" problem are not considered part of the credit card fraud problem itself, and are reported as part of charge-off losses. At least one estimate puts the size of bankruptcy fraud in the neighborhood of $2.65 billion in 1992, which would make it larger than all other fraud losses combined [1].

Current Methods of Fraud Detection

The diversity of fraudulent activity as evidenced by the many forms of fraud makes detection of fraudulent behavior a non-trivial task. At most banks, some part of the review process of new applications for new credit cards involves routine information checks to spot possible fraudulent applications. (In some cases, scrutiny of application forms for telltale methods of handwriting has led investigators to spot fraudulent applications submitted by organized criminal elements. Some fraudulent applications submitted by a Nigerian fraud ring have been caught in this way.)

Once a card has been issued, however, most banks rely upon periodic scrutiny of account behavior to determine if there is suspicion of fraud. In particular, banks have developed a series of rule-based checks against which all portfolio activity is reviewed. Such checks might specify nominal limits on the number of transactions that should reasonably be expected to occur in a single day. This excessive transaction report might also be limited to a count of transactions above some threshold on purchase amount.

These fraud rules are developed as a result of historical analyses of past fraudulent behavior in the portfolio. However, most banks use only the most basic of statistical analyses to develop the fraud rules, leading in most cases to rule sets that consist of a set of simple threshold conditions on account variables. Not surprisingly, the use of more sophisticated technologies for fraud detection can lead to dramatically improved results. In particular, when viewed as a problem in pattern recognition, the problem of fraud detection is an excellent application for an appropriately chosen neural network solution. Increasingly, a number of problems in financial services are being viewed in terms pattern recognition problems for which neural network solutions may be developed [2].

The Mellon Bank Fraud Detection Feasibility Study

A feasibility study was done for Mellon Bank to determine the effectiveness of a neural network for fraud detection on their credit card portfolio. The feasibility study consisted of training a neural network-based system on a sample of good and fraud accounts, followed by the execution of a blind test of the trained model on a separate, unsampled, holdout set of transactions. In the training set, transactions were provided together with good/fraud labels, while in the blind test set, no fraud labels were initially provided.

The study was designed to simulate the effectiveness of a fraud detection system deployed as a postprocessing step to the bank's authorization system. When a transaction arrives for authorization, it is characterized by a stream of authorization data fields that carry information identifying the cardholder (account number) and characteristics of the transaction (e.g., amount, merchant code). There are additional data fields that can be taken in a feed from the authorization system (e.g., time of day). In most cases, banks do not archive logs of their authorization files. Only transactions that are forwarded by the merchant for settlement are archived by the bank's credit card processing system. Thus, a data set of transactions was composed from an extract of data stored in Mellon's settlement file. In this extract, only that authorization information that was archived to the settlement file was available for model development. In particular, day but not time of transaction was available, as were amount of transaction and SIC code of the merchant. Transaction denials (authorization requests that were denied as opposed to authorized) were also not available.

Additionally, payment information on the account was also available as an extract from the settlement file. Non-financial information (e.g., date of issue or date of last re-issue) was also included as a data field that was available at the time of a transaction request for authorization.

One of the objectives of the study was to determine if the use of a wide variety of information characterizing the transaction would be helpful in developing improved fraud detection capability. What was envisioned was a system that could review each transaction in the context of the recent history of account transactions and payments, along with other non-financial data on the account, to determine likelihood of fraud.

In the design of neural network-based pattern recognition systems, there is always a process of feature extraction that is applied to the input "raw" data fields in order to present to the network a set of inputs that is meaningfully organized in terms of the particular pattern recognition problem at hand. In this case, values from a set of 50 data fields were combined to produce a set of 20 features that were used as input to the network. These features can be roughly grouped in the four categories shown in Figure 2.

Current transaction descriptors can include such features as the amount of the transaction, the day and time at which it is occurring, (though time of day was not available for the Mellon study) as well as the standard industry code (SIC) of the merchant, a numeric code representing the nature of the merchant business (e.g., jewelry store, consumer electronics, restaurant, hotel, etc.) History descriptors contain features characterizing the use of the card for transactions and the payments made to the account over some immediately prior time interval. (For the Mellon study, the length of the history period for the account and payment-related variables was on the order of 8-10 weeks.) Other descriptors can include such factors as the date of issue (or most recent re-issue) of the card. This can be important for the detection of NRI fraud.

The transaction data for the training set consisted of a sample of the transactions on Mellon's portfolio during the months of January through June, 1991. The original set of portfolio transactions was sampled in such a way that all fraud transactions were included, while a sample of the good transactions was chosen so as to have a ratio of roughly 30 good accounts for each fraudulent account in the training set. An account was considered fraudulent if, during the course of the time represented by the training set, it had at least one transaction labelled as fraudulent. Altogether, some 450,000 transactions (as opposed to accounts) were used in the training set. The entire portfolio available for training consisted of approximately 650,000 accounts.

As is often the case in dealing with real-world data, some effort was expended to validate the contents of data fields and to ensure that the data set did contain the full complement of fraudulent activity from the time period in question. To the extent that only a partial picture of account activity or fraudulent activity on the account is available during modeling, the model will train less effectively to detect fraud.

The neural network used in this fraud detection feasibility study is the P-RCE neural network [3]. The P-RCE is a member of the family of radial-basis function networks that have been developed for application to pattern recognition. The P-RCE is a three-layer, feed-forward network that is distinguished by its use of only two training passes through the data set. The first training pass involves a process of prototype cell commitment in which exemplars from the training set are stored in the weights between the first and second (middle) layer cells of the network. A final training pass determines local a posteriori probabilities associated with each of these prototype cells. P-RCE training is not subject to problems of convergence that can afflict gradient-descent training algorithms. The P-RCE network and networks like it have been applied to a variety of pattern recognition problems both within and beyond the field of financial services, from character recognition to mortgage underwriting and risk assessment [4,5,6,7]. More detail on the P-RCE network is provided in the Appendix.

In this study, the P-RCE output layer consisted of a single cell that outputs a numeric response that can be considered as a "fraud score". This is analogous to credit scoring systems that produce a score, as opposed to a strict probability. The objective of the neural network training process is to arrive at a trained network that produces a fraud score that gives the best ranking of the credit card transactions. If the ranking were perfect, all of the high scoring transactions down to some threshold would be fraud; below this threshold, only good transactions would be ranked. However, perfect separability of frauds from goods is not possible due to the inherently non-separable nature of the fraud and good distributions in the selected pattern recognition space.

Final evaluation of the trained network was done on the Blind Test data set. The Blind Test data represented an unsampled set of all Mellon Bank's transactions for the two month period October-

November 1991. (Note that the model development data set was taken from transactions prior in time to that of the Blind Test data.) The network was tested on all (roughly) 2,000,000 transactions that were authorized in this two month period. Results are presented in the following figures.

Figure 3 shows one measure of fraud detection accuracy: a rank curve of the the percentage of fraudulent transactions detected by the neural network plotted against the number of cardholder accounts that would be flagged by the system for review on a daily basis. We see from the shape of the curve that extremely high accuracies in fraud detection are obtained for those high fraud score values at which a small number of accounts are flagged for review each day. Further down the fraud score axis, the slope of the accuracy curve begins to flatten out. At lower fraud score cutoffs, where the number of accounts that are reviewed per day are in the neighborhood of 50 accounts, nearly 40% of all fraudulent transactions will appear among the accounts under review.

By comparison, the effectiveness of Mellon's prior fraud detection efforts on this same data were such as to require them to review approximately 750 accounts per day, yielding, on the average, only one detected fraudulent account per week. The improvement in detection performance is undeniably substantial.

However, an increase in accuracy of detection does not necessarily, by itself, bring a comparable improvement in terms of economic benefit. Most fraud is fast moving, beginning and ending over a three day period. It is possible to imagine a system that might, with great accuracy, detect the last fraudulent transaction to occur on the card. Such a system would bring little economic benefit to the bank. It would have detected the fraud, but detected it too late to achieve any savings. Consequently, we also measure the earliness of the fraud detection.

Figure 4 shows a histogram of a breakdown of when the fraud was detected for two different operating points: a fraud score threshold at which, on average, one fraud account is detected per day and a lower detection threshold at which, on average, two fraud accounts are detected per day. Two observations are important to note. The first is that a significant percentage of detections are detections of accounts on either the first or second day of fraud activity. At the 1-fraud-per-day operating point, 50% of detected accounts are detected on either the first or second day of fraud activity.

Secondly, if the fraud detection threshold is reduced so as to catch more frauds, in particular to the point where the detections are averaging two fraud accounts per day, the percentage that are detected on either the first or second day of fraud activity climbs to 60%. Thus, lowering the fraud detection threshold can result not only in more fraud detections but also in earlier detection. This early detection provides the bank the opportunity to take action to prevent future use of the card for fraudulent transactions, thus reducing the average "run" on the card and the average dollars lost to fraud per fraud account.

Another measure of the detection performance concerns the types of fraud that the system is detecting. Figure 5 presents a breakdown, at the two different operating points, one fraud per day and two frauds per day, of the detected frauds as a percentage of the population of the major fraud categories. Importantly, the system is detecting fraud across all the major fraud categories.

Since this is an historical study, it is possible to determine, in retrospect, a direct measure of savings that would have resulted from use of this technology and system for fraud detection. We compute the savings on a particular fraud account by summing the dollars associated with all fraudulent transactions on that account subsequent to (and in some cases including) the first detected transaction, given a particular fraud score detection threshold. The total fraud dollar savings is then summed for all fraud accounts. It is convenient to express this savings as a percentage of the total dollar fraud losses. Figure 6 presents a plot of the percentage fraud loss saved versus total number of account reviews per day. (The explicit dependence on fraud score threshold is eliminated by casting this in terms of the number of accounts reviewed per day, which is directly determined from the fraud score threshold.)

Several graphs are presented in Figure 6. In the topmost curve (open squares), the percentage dollar fraud loss savings is plotted for a system that anticipates that the first detected transaction can be blocked at the point of sale. This requires that the system compute its fraud score for a given transaction request and message the authorization system in timely fashion to transmit a "Denial" response to the merchant. We call this the "real-time response" mode. The third curve from the top (open circles) shows the economic benefit for a system in which the first detected transaction is not blocked, but all subsequent transactions are blocked. This corresponds to a system that is perhaps scoring the transaction a short time interval after it has already been authorized. We might call this the "post-processing" mode. Each of these two curves gives rise to another curve, which removes from consideration all transactions less than $50. These transactions would not be seen by Mellon's authorization system either because the transaction amount was less than the merchant's floor-limit, or because of Visa and MasterCard stand-in processing. Depending upon the mode of implementation and the selected operating point for fraud detection, savings in the range of 20% to in excess of 40% can be achieved.

Importance of Payment Information

In order to determine the contribution of the payment-related features to the accuracy of the model, a comparison was done of the performance of two different models, one trained with the benefit of payment-related features and one without. This comparison was done on the training set, (as opposed to the Blind Test set), where the good accounts had been sampled by a 1:17 factor. The graph below shows the performance of the two different models. At many points along the rank curve, the model trained without the benefit of payment-related features produces nearly twice as many accounts for review as the model trained with the benefit of payment-related features. Naturally, this factor will be influenced by the mix of fraud types in the portfolio. However, Figure 7 does present very firm evidence for the importance of payment-related information as input features to a neural network-based fraud detection system.

Installed Software

The fraud detection feasibility study has led to an installation of the fraud detection software (FDS™) at Mellon Bank. The installation configuration chosen by Mellon Bank involves the execution of the scoring function of FDS periodically during the day (as of this writing, every two hours), scoring all transactions that have been authorized since the last scoring run. Additionally, there is a final scoring that occurs at the end of the day after account posting, to process the settlement file. The settlement file consists of transactions that were authorized by Mellon Bank and have now come in for posting, transactions that were authorized by MasterCard and Visa as "stand-in" processing and transactions of amounts less than the merchant's floor limit which were not authorized. (For transactions below agreed-upon amounts, the credit card associations provide "stand-in" authorization in order to help reduce overall transaction traffic on the communications network. These transactions are transmitted electronically to the bank at day's end.)

The FDS scoring functions execute at roughly 200,000 transactions per CPU hour on Mellon's IBM mainframe, a model 3090-600J. A rapid processing "Pre-scoring" neural network is available in the FDS software to enhance scoring speed of execution. The above quoted speed does not reflect the use of that network. Prior experience indicates that a factor of 10 increase in speed is achieved when this pre-screening network is used.

The Mellon FDS™ fraud detection system has been fully operational since March 3, 1993. In-field performance consistent with or better than that of the feasibility study reported in this paper is being achieved.

Conclusion

The alarming increase in fraudulent credit card usage has stressed the fraud management systems currently in use at banks and other institutions that process credit card transactions. In a rigidly controlled test on real world data from Mellon Bank's credit card portfolio, a neural network-based fraud detection system has been shown to provide substantial improvements in both accuracy and timeliness of fraud detection. The feasibility study demonstrated that due to its ability to detect fraudulent patterns on credit card accounts, it is possible to achieve a reduction of from 20% to 40% in total fraud losses, at significantly reduced caseload for human review. Software implementing this neural network approach to fraud detection has been successfully installed and integrated into the production environment on Mellon Bank's mainframe computer, and Mellon is achieving fraud loss reductions consistent with those predicted in the study.

Acknowledgements

The authors wish to express their appreciation to Mr. Philip Samson for his participation in the feasibility study.

Appendix

Over the years, a great diversity of artificial neural network (ANN) architectures and learning techniques has been developed. Very general ANN architectures require enormous processing element inter-connectivity that can make them impratical for software simulation on complex pattern recognition problems involving large scale data sets.

Fortunately, it is now well known that the three-layer feed-forward architecture permits arbitrary mappings between input and output data distributions [8]. The three-layer, feed-forward, radius-limited perceptron network permits accurate modeling of statistical distributions when the data may be described by a small number of Gaussian-like distributions. This type of network is ideal for compact, non-linearly meshed class regions. Among the several training procedures known for such networks [9,10], Reilly et al. defined a rapid, non-gradient-descent procedure [11, 12] which requires limited connectivity for error correction.

Feed-forward, three-layer networks that combine radius-limited perceptrons with inner-product perceptrons are currently called radial basis function networks (RBFs)[9,10]. The notion of radius-limited perceptrons was discussed by some of the earliest neural network investigators [13], and RBF networks themselves have received much attention in the past [14,15,16,17].

In an RBF network, cells of the input layer transmit the input pattern vector to all of the cells in the second layer. The second layer cells are radius-limited perceptrons which compute activations according to:

$$x^2{}_i = G(d_i) = G\Big(\Big(\sum_{j=1}^{N^1} (x^1{}_j - \omega^1{}_{ij})^2\Big)^{1/2} - \theta_i\Big). \qquad \text{A.1}$$

In the above notation, superscripts identify the layer index. $\underline{\omega}^k$ refers to the weight matrix connecting the kth layer to the (k+1)th layer; N^k is the number of cells in the kth layer; $G(\cdot)$ is a reverse step function such that $G(d_i) = 0$ for $d_i > 0$ and $G(d_i) = 1$ for $d_i < 0$. The classification cells in the output layer use a step function $F(\cdot)$, $F(d_i) = 0$ for $d_i \leq 0$ and $F(d_i) = 1$ for $d_i > 0$, to compute activations according to:

$$x^3{}_i = F(d_i) = F\Big(\sum_{j=1}^{N^2} \omega^2{}_{ij} x^2{}_j\Big). \qquad \text{A.2}$$

This network performs well when class regions are separable, that is when a boundary of some kind exists between them. However, many problems are characterized by class regions which share points in the feature space. These shared regions are said to be non-separable, and are best characterized by probability density functions (pdfs) $p(f \mid C)$, which specify the dependence of the distribution for class C on the random input pattern $\underline{f}$. Bayes rule permits classification of patterns in non-separable regions if the pdfs are known.

In the P-RCE network, the reverse step activation function of the second layer cells is replaced by an exponential:

$$x^2{}_i = \exp(-d^2{}_i), \; d^2{}_i = | \underline{x} - \underline{\omega}^1{}_i |^2 \qquad \text{A.3}$$

and the activation function of the third layer cells is the identity function so that

$$x^3{}_i = \sum_{j=1}^{N^2} \omega^2{}_{ij} x^2{}_j \qquad \text{A.4}$$

Training of the P-RCE network begins with the simple procedural learning algorithm for the RCE network described in detail in [11] and [18]. An illustration is provided in Figure A.1. A pattern from the training set is presented to the network, and if any cells are present in the second layer, they are activated according to Equation A.1; output cells are activated according to Equation A.2. If an output cell of a different class than the pattern is activated, then the thresholds of the middle-layer cells which activated it are adjusted so that they are no longer activated by the pattern (Figure A.1b). In addition, if none of the middle-layer cells connected to the correct output cell are active, then a new cell is added to the network with weights which correspond to the location of the pattern (Figure A.1c). At the same time, the new cell is connected to the corresponding output layer cell with a unit weight. The new cell's threshold is then initialized to the distance between the input pattern and the nearest weight vector of a cell belonging to another class.

As a result of the learning procedure, the cells of the second layer of the network form a "covering" of the class regions that can represent arbitrarily non-linear interclass boundaries. (Figure A.1d). This mapping of the training data distributions typically requires from four to six passes through a randomly ordered training set, depending on the complexity of the distributions and the ordering of the data in the training set .

For overlapping class distributions, pdfs may be estimated by maintaining a count of correctly classified patterns which fall within the radius of each of the middle layer of cells [12]. This local counter of correctly classified patterns is stored in the second layer weight $\omega^2{}_{ij}$ which links the jth cell of the middle layer with the ith output cell. Thus during training, correctly identified patterns of class i, which elicit activity in middle-layer cell j, cause an increment to the weight ω^2_{ij}:

$$\omega^2_{ij}(t+1) = \omega^2_{ij}(t) + 1. \qquad A.5$$

Patterns which are not correctly identified by middle-layer cell j do not alter the value of ω^2_{ij}. Note that since the radial threshold of cell j may "shrink" as a result of RCE learning, the counter ω^2_{ij} may not be an accurate estimate of correctly identified patterns during RCE training. To ensure that the second layer of weights correctly estimates the local pdfs, a final pass on the full training set is required in which no addition of second layer cells and no adjustment of the cell thresholds occurs.

References

[1] *The Nilson Report*, Issue 540, January 1993.

[2] R. T. Trippi, E. Turban (eds), Neural Networks in Finance and Investing, Probus Publishing Company (1993).

[3] C. L. Scofield, D. L. Reilly, "Into silicon: real time learning in a high density RBF neural network," In *Proc. IEEE Int. Conf. on Neural Networks*, Seattle, WA, July 1991. Vol. I, pp. 551-556.

[4] D. L. Reilly and L. N. Cooper, "An overview of neural networks: early models to real world systems," in An Introduction to Neural and Electronic Networks, ed. S. F. Zornetzer, J. L. Davis and C. Lau, 227-248, Academic Press, (1990).

[5] R. Rimey, P. Gouin, C. L. Scofield and D. L. Reilly, "Real-time 3-D Object Classification using a learning system." *Proceedings of SPIE—the International Society of Optical Engineers*, 726, 552-558.

[6] E. Collins, S. Ghosh, C. L. Scofield, "An application of a multiple neural network learning system to emulation of mortgage underwriting judgments," in Neural Networks in Finance and Investing, ed. R. T. Trippi and E. Turban, 289-304, Probus Publishing Company, (1993).

[7] D. L. Reilly, E. Collins, C. L. Scofield and S. Ghosh, "Risk assessment of mortgage applications with a neural network system: an update as the test portfolio ages," in Neural Networks in Finance and Investing, ed. R. T. Trippi and E. Turban, 305-311, Probus Publishing Company, (1993).

[8] B. Irie and S. Miyake, "Capabilities of three-layered perceptrons," in *Proc. IEEE Second Int. Conf. on Neural Networks*, San Diego, CA, July 1988, vol. I, pp. 641-648.

[9] Moody and C. Darken, "Learning with localized receptive fields," in Proc. of the 1988 Connectionist Models Summer School, D.S. Touretzky, G.E. Hinton and T.J. Sejnowski, eds., Morgan Kaufmann Publishers, San Mateo, CA, 1989, pp. 133-143.

[10] S.J. Nowlan, "Max likelihood competition in RBF networks," Technical Report CRG-TR-90-2, Dept. of Computer Science, University of Toronto, Canada, 1990.

[11] D.L. Reilly, L.N. Cooper and C. Elbaum, "A neural model for category learning," *Biol. Cybern.*, vol. 45, 1982, pp.35-41.

[12] C.L. Scofield, D.L. Reilly, C. Elbaum and L.N. Cooper, "Pattern class degeneracy in an unrestricted storage density memory," in *Neural Information Processing Systems*, Denver, CO, 1987, D.Z. Anderson, ed., American Institute of Physics, New York, NY, 1988, pp. 674-682.

[13] M.L. Minsky and S. Papert, Perceptrons. MIT Press, Cambridge, MA, 1969.

[14] D.L. Reilly, C.L. Scofield, P.R. Gouin, R. Rimey, E.A. Collins and S. Ghosh, "An application of a multi-neural network learning system to industrial part inspection," in *Proc. of ISA Conference*, Houston TX, Oct, 1988.

[145] K.A. Marko, J. James, J. Dosdall and J. Murphy, "Automotive control system diagnostics using neural nets for rapid pattern classification of large data sets," in *Proc. IJCNN*, Washington D.C., June 1989, vol II. pp. 13-16.

[16] P. Zemany, W. Hogan, E.C. Real, D.P. Morgan, L. Riek, D.L. Reilly, C.L. Scofield, P. Gouin and F. Hull, "Experiments in discrete utterance recognition using neural networks," IEEE Central New England Second Biennial Conference on Acoustics, Speech, and Signal Processing, 1989.

[17] C.L. Scofield, L. Kenton and J.C. Chang. Invited Paper, COMPCON Spring '91, San Francisco, CA, February, 1991.

[18] D.L. Reilly, C.L. Scofield, C. Elbaum and L.N. Cooper, "Learning system architectures composed of multiple learning modules," *Proc. IEEE First Int. Conf. on Neural Networks*, San Diego, CA, June 1987, vol. II, pp. 495-503.

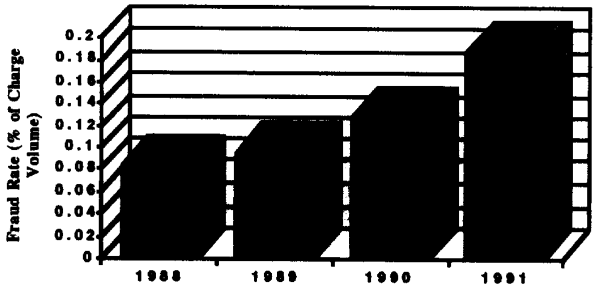

Figure 1: Fraud growth as percentage of charge volume, 1988-1991. (Source: Visa, MasterCard)

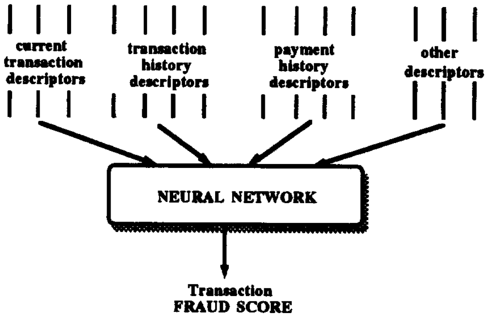
current transaction descriptors
transaction history descriptors
payment history descriptors
other descriptors
NEURAL NETWORK
Transaction FRAUD SCORE

Figure 2: Groups of input features characterizing each transaction to the network

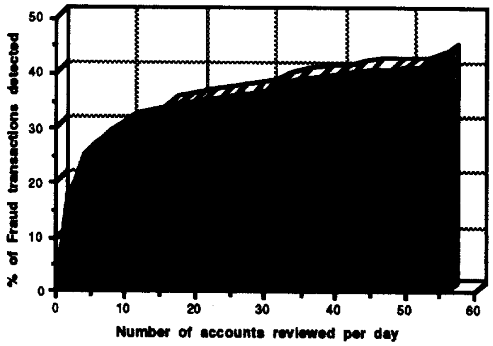

Figure 3: Percentage of fraud transactions detected as a function of number of accounts flagged for review by the system each day.

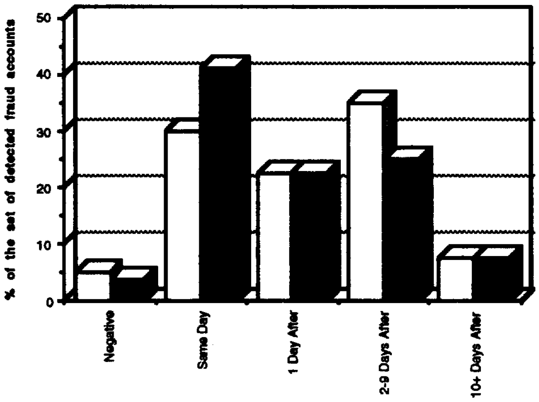

Figure 4: Percentage of detected accounts, broken down by date of detection, referenced to the day of the first observed fraud transaction. Small negative bin is an artifact of uncertain data labelling that may not have accurately established first fraud transaction. Light histograms correspond to the operating point at which 1 fraud is detected per day; dark histograms correspond to the operating point at which 2 frauds are detected per day.

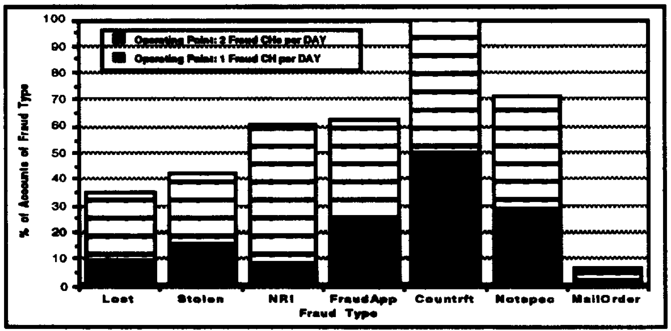

Figure 5: Fraudulent accounts detected at two different operating points, broken down by category. Within each category, number of frauds detected is expressed as a percentage of total fraudulent accounts belonging to that category.

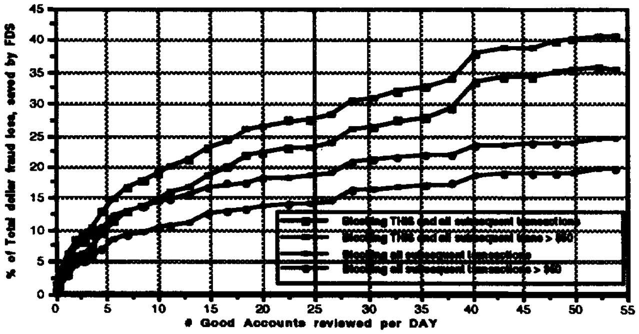

Figure 6: Percentage dollar fraud loss savings as a function of different installation-related assumptions.

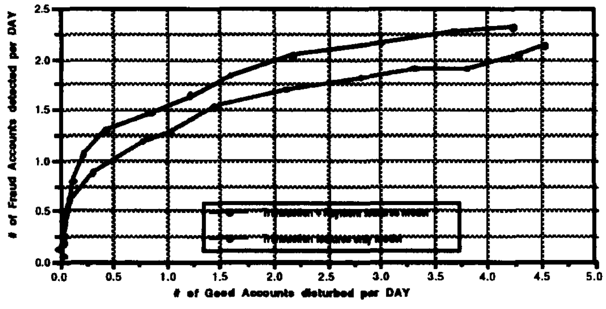

Figure 7: Performance comparison of models trained with and without payment-related information. Detection accuracy is quoted on sampled data set. (Good account sampling factor is 1:17.)

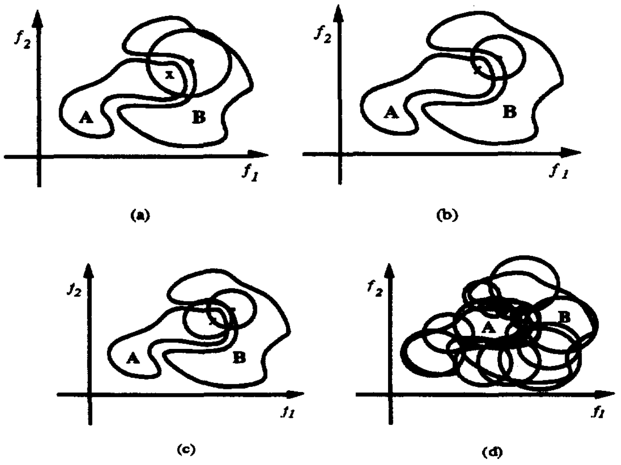

Figure A.1: (a) A training pattern from class region A, represented by an ``x", is presented to the network and activates a second layer cell belonging to class B, represented by the "•", because it falls within the cell's influence region (the circle.) (b) That cell's "region-of-influence" (defined by the cell threshold, θ) is decreased to the point where the pattern no longer activates the cell. (c) New cells are added to the second layer when no cells of the correct class are activated by the pattern. (d) Class regions are defined by the final assemblage of second layer, radius-limited cells.