*Peter J. Bentley, *Jungwon Kim, **Gil-Ho Jung and ***Jong-Uk Choi
*Department of Computer Science, University College London
**Department of Computer Science, SungKyunKwan University
*** Department of 888 , Sangmyung University
e-mail: J.Kim@cs.ucl.ac.uk

Fuzzy Darwinian Detection of Credit Card Fraud

Peter J. Bentley, *Jungwon Kim, **Gil-Ho Jung and ***Jong-Uk Choi
*Department of Computer Science, University College London
**Department of Computer Science, SungKyunKwan University
*** Department of 888 , Sangmyung University
e-mail: J.Kim@cs.ucl.ac.uk

요 약

Credit evaluation is one of the most important and difficult tasks for credit card companies, mortgage companies, banks and other financial institutes. Incorrect credit judgement causes huge financial losses. This work describes the use of an evolutionary-fuzzy system capable of classifying suspicious and non-suspicious credit card transactions. The paper starts with the details of the system used in this work. A series of experiments are described, showing that the complete system is capable of attaining good accuracy and intelligibility levels for real data.

1. INTRODUCTION

Fraud is a big problem today. Looking at credit card transactions alone, with millions of purchases every month, it is simply not humanly possible to check every one. And when many purchases are made with stolen credit cards, this inevitably results in losses of significant sums.

The only viable solution to problems of this scale is automation by computer. Just as computers are used for credit scoring, risk assessment and customer profiling, it is possible to use computers to assess the likelihood of credit card transactions being "suspicious". Such automated detection can be performed by using simple statistical techniques, or by applying 'rules of thumb' to claims. However, the fingerprints of fraudulent activity may be diverse and complex, resulting in the failure of these traditional methods. This motivates the use of newer techniques, called machine learning or *pattern classification*, which are capable of finding complex non-linear 'fingerprints' in data.

This paper investigates one such technique: the use of genetic programming to evolve fuzzy logic rules capable of classifying credit card transactions into "suspicious" and "non-suspicious" classes. The paper follows on from [1] and [2], describing the application of the committee-decision making system to a new problem.

2. SYSTEM OVERVIEW

This section describes the evolutionary fuzzy system used (with different setups) as members of a committee. Full details of this system can be found in [2].

The system developed during this research comprises two main elements: a Genetic Programming (GP) search algorithm and a fuzzy expert system. Figure 1 provides an overview.

2.1 CLUSTERING

Data is provided to the system in the form of two comma-separated-variable (CSV) files: training data and test data. When started, the system first clusters each column of the training data into three groups using a one-dimensional clustering algorithm. A number of clusterers are implemented in the system, including C-Link, S-Link, K-means [5].

After every column of the data has been successfully clustered into three, the minimum and maximum values in each cluster are found. These values are then used to define the domains of the membership functions of the fuzzy expert system [6].

2.2 MEMBERSHIP FUNCTIONS

Three membership functions, corresponding to the three groups generated by the clusterer, are used for each column of data. Each membership function defines the 'degree of membership' of every data value in each of the three fuzzy sets: 'LOW', 'MEDIUM' and 'HIGH' for its corresponding column of data. Since every column is clustered separately, with the clustering determining the domains of the three membership functions, every column of data has its own, unique set of three functions.

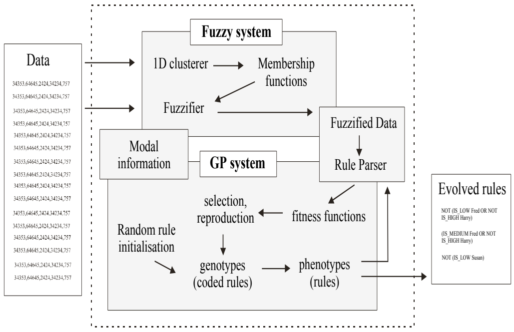

Figure 1 Block diagram of the Evolutionary-fuzzy system.

The system can use one of three types of membership function: 'non-overlapping', 'overlapping', and 'smooth' [2]. The first two are standard trapezoidal functions, the third is a set of functions based on the arctangent of the input in order to provide a smoother, more gradual set of 'degree of memberships'.

Whichever set of membership functions are selected, they are then shaped according to the clusterer and used to fuzzify all input values, resulting in a new database of fuzzy values. The GP engine is then seeded with random genotypes (coded rules) and evolution is initiated

2.3 EVOLVING RULES

The implementation of the GP algorithm employs many of the techniques used in GAs to overcome some of the problems associated with simple GP systems. For example, this evolutionary algorithm uses a crossover operator designed to minimise the disruption caused by standard GP crossover, it uses a multiobjective fitness ranking method to allow solutions which satisfy multiple criteria to be evolved, and it also uses binary genotypes which are mapped to phenotypes.

2.3.1 Genotypes and Phenotypes

Genotypes consist of variable sized trees, where each node consists of a binary number and a flag defining whether the node is binary, unary or a leaf, see figure 2. At the start of evolution, random genotypes are created. Genotypes are mapped onto phenotypes to obtain fuzzy rules, e.g. the genotype shown in fig. 2 maps onto the phenotype:

"(IS_MEDIUM (Height OR IS_LOW Age) AND IS_MEDIUM Age)".

Currently the system uses two binary functions: 'OR' and 'AND', four unary functions: 'NOT', 'IS_LOW', 'IS_MEDIUM', 'IS_HIGH', and up to 256 leaves (column labels such as "Date", "PolicyNumber", "Age", "Cost"). Depending on the type of each node, the corresponding binary value is mapped to one of these identifiers and added to the phenotype. The mapping process is also used to ensure all rules are syntactically correct, see [2].

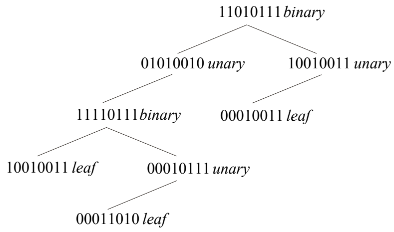

Figure 2: An example genotype used by the system.

2.3.2 Rule Evaluation

Every evolved phenotype (or fuzzy rule) is evaluated by using the fuzzy expert system to apply it to the fuzzified training data, resulting in a defuzzified score between 0 and 1 for every fuzzified data item. This list of scores is then assessed by fitness functions which provide separate fitness values for the phenotype, designed to:

i. minimise the number of misclassified items.
ii. maximise the difference between the average scores for correctly classified "suspicious" items and the average scores for "normal" items.
iii. maximise the sum of scores for "suspicious" items.
iv. penalise the length of any rules that contain more than four identifiers (binary, unary, or leaf nodes).

2.3.3 Rule Generation

Using these four fitness values for each rule, the GP system then employs the SWGR multiobjective optimisation ranking method [4] to determine how many offspring each pair of rules should have.

Child rules are generated using one of two forms of crossover. The first type of crossover emulates the single-point crossover of genetic algorithms by finding two random points in the parent genotypes that resemble each other, and splicing the genotypes at that point. By ensuring that the same type of nodes, in approximately the same places, are crossed over, and that the binary numbers within the nodes are also crossed, an effective exploration of the search space is provided without excessive disruption [3]. The second type of crossover generates child rules by combining two parent rules together using a binary operator (an 'AND' or 'OR'). This more unusual method of generating offspring (applied approximately one time out of every ten instead of the other crossover operator) permits two parents that detect different types of "suspicious" data to be combined into a single, fitter individual. Mutation is also occasionally applied, to modify randomly the binary numbers in each node by a single bit.

The GP system employs population overlapping, where the worst Pn% of the population are replaced by the new offspring generated from the best Pm%. Typically values of $Pn = 80$ and $Pm = 40$ seem to provide good results. The population size was normally 100 individuals.

2.3.4 Modal Evolution

Each evolutionary run of the GP system (usually only 15 generations) results in a short, readable rule which detects some, but not all, of the "suspicious" data items in the training data set. Such a rule can be considered to define one mode of a multimodal problem. All items that are correctly classified by this rule (recorded in the modal database, see figure 1) are removed and the system automatically restarts, evolving a new rule to classify the remaining items. This process of modal evolution continues until every "suspicious" data item has been described by a rule. However, any rules that misclassify more than a predefined percentage of claims are removed from the final rule set by the system.

2.4 ASSESSMENT OF FINAL RULE SET

Once modal evolution has finished generating a rule set, the complete set of rules (joined into one by disjunction, i.e., 'OR'ed together) is automatically applied to the training data and test data, in turn. Information about the system settings, which claims were correctly and incorrectly classified for each data set, total processing time in seconds, how the data was clustered and the rule set are stored to disk.

2.5 APPLYING RULES TO FUZZY DATA

The path of evolution through the multimodal and multicriteria search space is guided by fitness functions. These functions use the results obtained by the Rule Parser - a fuzzy expert system that takes one or more rules and interprets their meaning when they are applied to each of the previously fuzzified data items in turn.

This system is capable of two different types of fuzzy logic rule interpretation: traditional fuzzy logic, and *membership-preserving* fuzzy logic, an approach designed during this research. Depending on which method of interpretation has been selected by the user, the meaning of the operators within rules and the method of defuzzification is different. Complete details of the fuzzy interpretation methods are provided in [2].

2.6 COMMITTEE DECISIONS

As should now be apparent, the evolutionary-fuzzy system has a number of very different parameters that can be used at any one time. What may be a good setup for one data set is not so good for another. In addition, the multiple results generated by multiple different system setups need to be assessed against multiple criteria. To achieve this, the system equips a multi-model decision aggregation system. The user can now set up as many as four different versions of the system and have them run in parallel on the same data set. The committee decision maker employs aggregation of weighted normalised values for accuracy and importance [1]. The default weighting values were 0.3 and 1.0 for accuracy and importance, respectively. Once every rule set has been assigned a score, the set(s) with the highest score for each committee member are reported to the user. The committee decision maker then performs the same analysis globally, finding the globally most accurate and intelligible rule set(s), then assigning every rule set a score based on globally aggregated, weighted, normalised values. The best overall rule set(s) are then reported to the user. For full details, see [1].

3. APPLYING THE SYSTEM TO CREDICT CARD DATA

3.1 DATA

The data used in this work was gathered from a domestic credit card company. Even though the company provided real credit card transaction data for this research, it required that the company name was kept confidential. The data was gathered from January to December of 1995 and a total of 4000 transaction records were provided, each with 96 fields. 62 fields were selected for the experiments. The excluded 34 fields were regarded as clearly irrelevant for distinguishing the credit status. (Examples include the client code number and the transaction index number.) The details of selected field names were not allowed to be reported. In order to allow the fuzzy rule evolution of the system, the collected data was labeled as "suspicious" or "non-suspicious". These labels were made by following the heuristics used in the credit card company. Specifically, when the customer's payment is not overdue or the number of overdue payment is less than three months, the transaction is considered as "non-suspicious", otherwise it is considered "suspicious".

To prepare a training set and a test set, we employed a simple cross-validation method. We held one-third of the data for testing and used the remaining two-thirds for training. The system executed its rule-evolution three times on three different training data sets. For each run, the system replaced the training set with the other third of the data set. This cross-validation was performed in order to ensure the evolved rule sets were not biased by a certain group of training set. By comparing the three different evolved rules based on three different groups of training data set, the final rule set is expected to represent the features of the entire data set. Unfortunately, the distribution of collected credit card transaction data was not even for each class. It had a larger number of examples for the "non-suspicious" class than for the "suspicious" class. The total number of items belonging to the smaller size of "suspicious" class was 985. This number is large enough to be divided into three subsets. Thus, the four committee members with identical experiment setups were run three times on each data subset respectively. The examples included in each set are shown in Table 1.

Exp	"SUSPICIOUS"		"NON-SUSPICIOUS"	
	Training	Test	Training	Test
1	1-656	657-985	1-2000	2000-3015
2	329-985	1-328	1001-3015	1-1000
3	657-985 & 1-328	329-656	2001-3015 & 1-1000	1001-2000

Table 1. Credit card data distribution for three experiments. The number in this table shows the IDs of examples belonging to each set. Exp stands for the experiment.

3.2 EXPERIMENTS

Three sets of experiments were performed with the committee decision system and the four different setups of fuzzy rule evolver were run for each experiment:

	[A] Fuzzy Logic with non-overlappingMFs					[B] Fuzzy Logic with overlapping MFs					[C] MP-Fuzzy Logic with overlapping MFs					[D] MP-Fuzzy Logic with smooth MFs				
	R	Training		Test		R	Training		Test		R	Training		Test		R	Training		Test	
		TP%	FN%	TP%	FN%		TP%	FN%	TP%	FN%		TP%	FN%	TP%	FN%		TP%	FN%	TP%	FN%
1	3	6.09	3.81	10.4	3.35	2	100	0	100	85.1	16	10.9	5.79	100	100	5	48.6	5.79	42.5	10.3
2	2	44.1	5.79	47.8	9.45	3	100	1.67	99.7	6.38	3	1.37	5.64	99.7	100	10	41.6	5.79	47.6	12.5
3	3	46.8	5.18	46.9	6.09	3	100	5.78	100	5.79	4	1.67	5.64	86.9	100	16	42.7	5.94	42.9	6.40

Table 2 Intelligibility (number of rules) and accuracy (number of correct classifications of "suspicious" items) of rule sets for test and training data. **R** shows the number of rules in the generated rule set and **TP** and **FN** is represented in %.

1. standard fuzzy logic with non-overlapping membership functions
2. standard fuzzy logic with overlapping membership functions
3. membership-preserving fuzzy logic with overlapping membership functions
4. membership-preserving fuzzy logic with smooth membership functions

(Previous work had shown that varying these aspects of the system caused the largest variation in behaviour [2].)

All four committee members were trained on one selected training set and test set. This resulted in different rule sets being generated for this problem, each with different levels of intelligibility and accuracy.

3.3 RESULTS AND ANALYSIS

Table 2 presents the results of the experiments. The accuracy of the system is described by a True Positive (TP) prediction rate and a False Negative (FN) error rate. The TP is the rate that the predicted output is "suspicious" class when the desired output is "suspicious" class. The FN is the probability of which the predicted output is "suspicious" when the desired output is "non-suspicious" class. The desired system will have a high TP and a low FN.

As Table 2 explains, committee member [B] provides the most accurate and intelligible classifications for all experiments with this data. The best accuracy overall is achieved by [B], detecting 100% of the "suspicious" claims for both on the training and the test set, whilst showing that 5.79% of false negative error, which is relatively low. In addition, the most accurate and intelligible rule sets that are generated by [B] contain just three rules. Overall, the best rule set as reported by the committee decision maker is for experiment 2:

(IS_LOW field57 OR field50)
IS_MEDIUM field56
(field56 OR field56)

and for the experiment 3:

(Filed49 OR Field56)
(IS_LOW Field26 OR field15)
IS_MEDIUM field56

These best rule sets are clearly dominated by the field56. This implies that this field seems to be the single best indicator of "suspicious" case. In summary, the prediction results of these best rule sets are satisfying in terms of the accuracy and intelligibility.

Another interesting observation is that the results of experiments rapidly change depending on the specific experiment setup. While [B] setup always generated the good rule sets, [C] setup provided almost meaningless rule sets, which showed nearly random prediction results. The setup [D] showed the consistent results, which the differences of TP and FN for both the training and the test sets are within 6%, but the best result is not satisfying. These results show again the large variance of committee member performance and illustrate the validity of the committee-decision maker approach for this problem.

In addition, from [A] and [B]'s results, it could be implied that the data set used in the experiment 1 seems to have somewhat different characters from other two data sets. The quite large difference, about 40% for TP in [A] and 80% for FN in [B] represent that the importance of data sampling during the fuzzy rule evolution stage.

4. CONCLUSION

This paper has described the application of a committee-decision-making evolutionary fuzzy system for credit card evaluation. The results for this real-world problem confirm previous results obtained in [1] for real home insurance data. They illustrate that the use of evolution with fuzzy logic can enable both accurate and intelligible classification of difficult data. The results also show the importance of committee-decision making to help ensure that good results will always be generated.

REFERENCES

[1] Bentley, P. J., "Evolutionary, my dear Watson: Investigating Committee-based Evolution of Fuzzty Rules for the Detection of Suspicious Insurance Claims", In the Proceeding of GECCO' 2000, July 8-12, Las Vegas, Nevada, USA, pp** - **, 2000.

[2] Bentley, P. J., "Evolving Fuzzy Detectives: An Investigation into the Evolution of Fuzzy Rule", A late-breaking paper in GECCO '99, July 14-17, 1999, Orlando, Florida USA, pp. 38-47, 1999.

[3] Bentley, P. J. & Wakefield, J. P., "Hierarchical Crossover in Genetic Algorithms", In Proceedings of the 1st On-line Workshop on Soft Computing (WSC1), (pp. 37-42), Nagoya University, Japan, 1996.

[4] Bentley, P. J. & Wakefield, J. P., "Finding Acceptable Solutions in the Pareto-Optimal Range using Multiobjective Genetic Algorithms", Chawdhry, P.K.,Roy, R., & Pant, R.K. (eds) Soft Computing in Engineering Design and Manufacturing. Springer Verlag London Limited, Part 5, 231-240, 1997.

[5] Hartigan, J. A , *Clustering algorithms*. Wiley, NY, 1975.

[6] Mallinson, H. and Bentley, P.J. "Evolving Fuzzy Rules for Pattern Classification", In Proc. of the Int. Conf. on Computational Intelligence for Modelling, Control and Automation - CIMCA'99, 1999.