A Neuro-fuzzy Approach to Bad Debt Recovery in Healthcare

Donghui Shi
Department of Computer Engineering
School of Electronics and Information Engineering
Anhui Jianzhu University
Hefei, China 230601
sdonghui@gmail.com

Jozef Zurada
Department of Computer Information Systems
College of Business
University of Louisville
Louisville, KY 40292
jozef.zurada@louisville.edu

Jian Guan
Department of Computer Information Systems
College of Business
University of Louisville
Louisville, KY 40292
jeff.guan@louisville.edu

Abstract

In the U.S. the healthcare industry is often plagued by unpaid bills, collection agency fees, and outstanding medical testing costs. All these factors contribute significantly to the rising cost of healthcare. Health care providers often have to treat patients on credit, especially in emergency and trauma cases. Unlike financial institutions health care providers do not collect financial information about their patients. This lack of information makes it difficult to evaluate whether a particular patient-debtor is likely to pay his/her bill. In recent years researchers have started to recognize the potential of data mining methods in improving our understanding of medical bad-debt, but there is relatively little research that examines the effectiveness of data mining methods in classifying bad debt in healthcare. This paper evaluates the effectiveness of an adaptive neuro-fuzzy inference system (ANFIS) in classifying bad debt in the healthcare context. The data analysis and evaluation of the performance of the ANFIS model are based on a fairly large unbalanced data sample provided by a healthcare company, in which cases with recovered bad debts are grossly underrepresented. Computer simulation shows that ANFIS is a viable method which produced under some scenarios good classification accuracy. More in-depth interpretation of the results, including nonlinear interaction between various factors, is provided through the analysis of the control surfaces generated by ANFIS and receiver operating characteristic (ROC) charts. Finally the paper also shows the potential of data mining models to classify unknown cases, which are a potential source of revenue recovery.

1. Introduction

The healthcare costs in the US have been rising at a very fast rate during the last two decades. One of the causes for the rising costs is unpaid medical bills. The healthcare industry, specifically hospitals and clinical organizations, are often plagued by bad-debt, collection agency fees and outstanding medical testing costs [2], [6], [10]. More recently the rising uninsured and underinsured are adding to the budget pressures. The bad debt issue in healthcare is not only affecting the bottom line, but it also has an impact on a healthcare organization's ability to provide care [6]. Recovering bad debt has become a serious matter and may even result in hospitals suing patients. Patients who miss court hearings related to their healthcare debts may be imprisoned [5]. Pesce [7] argues that hospitals should invest in modern information technology to reduce bad-debts. Though scarce, emerging predictive analytics models based on patient profiles show promise in better debt payment and collection fee reduction [3].

Despite an increasingly obvious and urgent need for predictive and classification models of bad debt in healthcare, there appears to be relatively little academic research on this very important topic. An early study by (Zollinger *et al*. [11] examined patient bad debt data using a regression model and identified several institutional variables such as total hospital charge and the total hospital revenue and patient variables such as marital status, gender, diagnoses, insurance status, employment status, and discharge status were significant factors in recovering unpaid hospital bills. Buczko [1] analyzed data on charges assigned to bad debt for 82 short-stay hospitals in Washington. The author confirmed that unpaid care has become a serious problem in hospital finance because of increasing number of uninsured patients and declining hospital revenues. Veletsos [10] described a more comprehensive study on using predictive modeling software such as IBM Intelligent Miner and DB2 for bad-debt recovery. The model is based on a variety of data variables, including credit factors, demographic information, and previous organizational

978-1-4799-2504-9/14 $31.00 © 2014 IEEE
DOI 10.1109/HICSS.2014.361

payment patterns. The model yielded approximately $200,000 in savings.

The work by [10] was substantially extended by Zurada and Lonial [12], [13]. Zurada and Lonial examined several different data mining tools and investigated their comparative performance in recovering bad debt [12], [13]. Computer simulation showed that the logistic regression, neural network, and the ensemble (combined) models produced the best overall classification accuracy, and the decision tree was the best in classifying "good" cases. They also provide more in-depth and meaningful interpretation of the results by analyzing the percent response, lift and ROC charts. The models were also used to score the "unknown" cases, which were not pursued by a company. The neural network model classified more "unknown" cases into "good" cases than any other remaining models. This may potentially provide an additional source of recoverable income. In this study computer simulation was performed on the same data set. More recently a report describes the use of predictive software in IBM SPSS to improve the bad debt collection effort and boost revenue. Though no details (such as models and methods used) are provided, the report describes that "one hospital saw a 30% reduction in bad-debt write-offs, a 12 percent increase in self-pay collection rates, and $25,000 per month reduction in agency fees."[3].

Predicting whether a particular customer is likely to repay a healthcare debt is an inherently complex and unstructured process. What makes this process especially difficult in the healthcare context is the hospital's inability to obtain detailed financial information concerning the patients. Unlike a financial institution which would collect information and carefully evaluate whether to extend a loan, healthcare institutions must often admit a patient and perform the necessary medical procedures on credit knowing very little about the particular patient. Thus, due to moral, legal and practical constraints, healthcare providers in the U.S. often become unwilling creditors to a multitude of borrowers. A healthcare institution is handicapped by having only a small number of independent attributes of the patient-debtor to evaluate [7]. In addition, some of debt defaults may be attributed to unforeseen events (i.e. divorce, death, loss of employment) or be governed by factors that may be difficult or impossible to detect in the attributes of the consumer (i.e. stability of marriage, general health, job stability). Given all of the difficulties described above it is not surprising that the healthcare institution that provided data for this paper recovered bad debts from only about seven percent (7.3%) of the non-paying patients.

This paper examines and compares the effectiveness of a neuro-fuzzy method (ANFIS) under different scenarios in classifying bad debt. The target/dependent variable in a fairly large data set provided by a healthcare company represents the following three classes: 1: "good" customers (those who repaid the debt or made partial payments to repay the debt); 2: "bad" customers (those who defaulted or refused to repay the debt); and 3: "unknown" customers (those who were not pursued). Due to the low recovery rate, the number of "good" customers is vastly underrepresented in the data set. To build and test the models, we only used cases representing "good" and "bad" customers, rejecting all "unknown" cases. The models were then used to score all "unknown" cases into "good" or "bad" which could provide additional revenue to the company.

We ran computer simulation for 5 different scenarios and used different membership functions for ANFIS. The best models in terms of the correct classification accuracy rates and the global performance were obtained when only the bad debt cases with the highest claim amounts were used. The interpretation of the control surfaces generated by ANFIS gave a unique and preliminary insight into the interaction between the input factors and the probability of default/recovery.

The paper is organized as follows. Section 2 discusses the basic characteristics of ANFIS used in the study. Section 3 describes the data sample and section 4 presents the experiments and simulation results. Finally, section 5 concludes the paper and makes recommendations for future work.

2. An adaptive neuro-fuzzy inference system for bad debt recovery

Neural fuzzy inference systems have emerged from the fusion of artificial neural networks and fuzzy inference systems. These systems combine learning/training and optimization abilities of artificial neural networks with human-like reasoning using if-then fuzzy rules offered by fuzzy inference systems. Neuro-fuzzy inference systems have formed a popular framework for modeling real world problems including classification. ANFIS is one of the better known neuro-fuzzy inference systems [4]. One of the advantages of ANFIS is its ability to generate fuzzy sets represented by membership functions and fuzzy rules from preexisting input-output data pairs available in the data set. Figure 1 shows the architecture of the ANFIS bad debt classification model in this paper. The model has 4 (m) inputs representing the 4 patient characteristics described in section 3. Each of the inputs has 2 (n) membership functions. The model uses a typical

ANFIS architecture with an additional node at the output end representing a discrimination function that classifies the output as either "good" customer (debt repaid) or "bad" customer (debt unpaid) with a user specified threshold value. The model uses a Takagi, Sugeno, and Kang (TSK) type fuzzy inference system and has two sets of trainable parameters: the antecedent (premise) membership function parameters and the consequent (polynomial) parameters [8], [9]. A typical TSK rule has the following structure:

If X_1 *is* $A_{1,j}$ *and* X_2 *is* $A_{2,j}$ *and* ... X_m *is* $A_{m,j}$
Then $f = r + p_1 X_1 + p_2 X_2 + \ldots + p_m X_m$

where $A_{i,j}$ is the jth linguistic term (such as high, low) of the ith input variable X_i, m is the number of inputs, f is the estimated output, and finally r and p_i are the consequent parameters to be determined in the training process. The architecture in Figure 1 is described as follows:

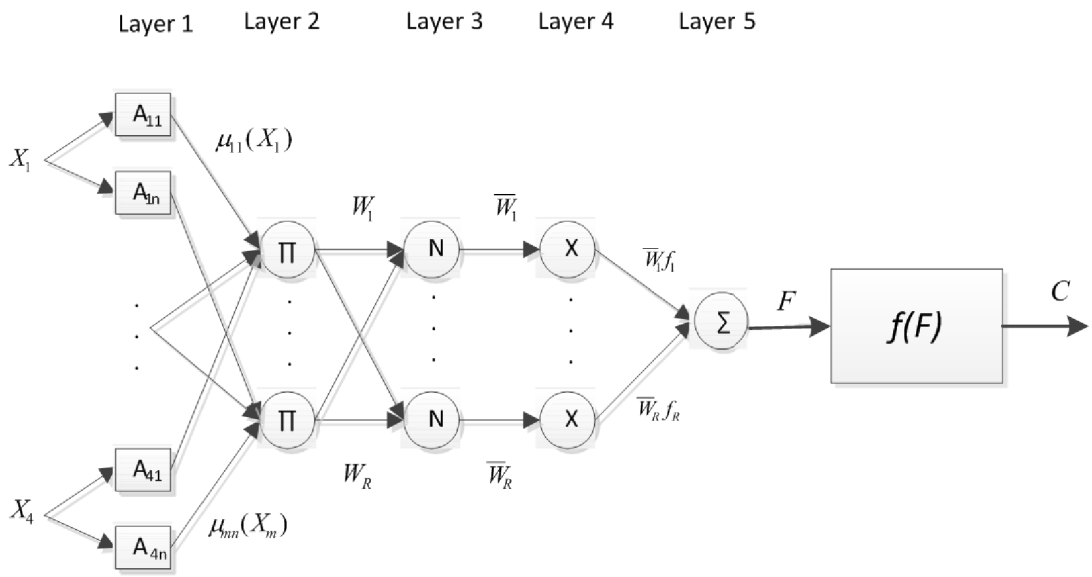

Figure 1. Architecture of the ANFIS bad debt recovery model

Layer 1: This layer contains the membership functions with adaptive parameters or premise parameters. The number of nodes (N=8) in the first layer is the product of the input size (m=4) and the number (n=2) of the membership functions for each input variable, or $N=m\times n$. The output of each node is defined as

$$O_{ij} = \mu_{ij}(X_i), for\ i = 1, m, j = 1, n$$

where μ_{ij} is the jth membership Gaussian function (four other functions have been used in this study) for the input X_i and is given as follows:

$$\mu(X) = exp\left\{-\left[\left(\frac{x-c}{a}\right)^2\right]^b\right\}$$

where a, b, and c are the premise parameters.

Layer 2: This layer calculates the firing strength of each rule and the output in this layer represents these firing strengths. The output is the product of all of its inputs as follows:

$$O_k = W_k = \mu_{1,i}(X_1)\mu_{2,i}(X_2)\cdots\mu_{m,i}(X_m)$$

for $k=1,R$ and R is the number of rules.

Layer 3: This layer normalizes the weighing factor of each of the input nodes k as follows:

$$O_k = \overline{W}_k = \frac{W_k}{W_1 + W_2 + \cdots + W_R}$$

Layer 4: the output of this layer represents a weighted value of the first order fuzzy if-then rule as follows:

$$O_k = \overline{W}_k f_k$$

where f_k is the output of the kth fuzzy rule as follows:

$$If (X_1\ is\ A_{11}) and\ (X_2\ is\ A_{22})\ and\ \cdots\ (X_m\ is\ A_{mn})$$

$$Then\ f_k = \sum_{i=1}^{m} p_{ij}X_i + r_k$$

where p_{ij} and r_k are called the consequent parameters and $j = 1, n$ and $k = 1, R$.

Layer 5: Finally this single node layer computes the overall output (F) of the ANFIS model as the sum of all the weighted outputs of the previous layer as:

$$O = F = \sum_{k=1}^{N} \overline{W}_k f_k$$

where f_k represents the output of the kth TSK-type rules as defined in layer 4.

Finally, the last module is a discriminant function *f(F)* which receives *F* as input and maps it to output *C* which is one of two values, "good" customer or "bad" customer. The parameters, both the premise parameters and consequent parameters, are learned/optimized in the training process. Two parameter optimization methods are used in training. The first method is backpropagation and the second method is a hybrid method that uses a mixture of backpropagation and least squares.

3. The data sample

The healthcare company, which is the subject of this study, relied on only four simple factors to determine whether the bad debt was recoverable: (1) Patient Age (PA), (2) Patient Gender (PG), (3) Injury Diagnosis Code (IDC), and (4) Dollar Amount of the Claim (DAC). In all likelihood, the four factors constituted all of the information about the patient-debtor that was available to the healthcare company. Furthermore, aside from the amount owed, the information appears to be only tangentially related to the probability that a particular bad-debt could be recovered. The dataset contains 6,319 cases with an outstanding balance of $2,388,999. The dependent variable, Status, represented "good", "bad", and "unknown" cases, respectively. After eliminating cases which had at least one missing value, we obtained a data set containing 6180 cases unequally divided into 449 "good" cases (group 1), 2,835 "bad" cases (group 2), and 2896 "unknown" cases (group 3).

Table 1. Summary of the descriptive statistics for the variables used in the models

Status	Patient Gender (PG)	Patient Age (PA)	Injury Diagnosis Code (IDC). (Most frequently represented shown only [>5%].) Code Range %	Dollar Amount of the Claim (DAC) [in US $]
Overall (groups 1, 2, and 3)	Female (N=2884, 52.9%) Male (N=3235, 47.1%)	Mean=34 St. Dev=19 Min=0 Max=88 S_k=0.5	"800"-"829" 16 "840"-"849" 21 "870"-"899" 12 "920"-"924" 15 "958"-"959" 14	Mean=$387 St. Dev=$1,470 Min=$1 Max=$40,508 Sum=$2,388,999 S_k=12.9
1 (Good) N=449	Female (N=248, 55.2%) Male (N=201, 44.8%)	Mean=34 St. Dev=18.9 Min=0 Max=88 S_k=0.5	"800"-"829" 13 "840"-"849" 23 "870"-"899" 10 "920"-"924" 13 "958"-"959" 16	Mean=$1,052 St. Dev=$3,442 Min=$3 Max=$40,508 Sum=$472,461 S_k=7.2
2 (Bad) N=2835	Female (N=1331, 47%) Male (N=1504, 53%)	Mean=31.6 St. Dev=23 Min=0 Max=100 S_k=0.6	"800"-"829" 18 "840"-"849" 17 "870"-"899" 15 "920"-"924" 8 "958"-"959" 19	Mean=$417 St. Dev=$1,248 Min=$1 Max=$19,568 Sum=$1,182,350 S_k=7.8
3 (Unknown) N=2896	Female (N=1335, 46%) Male (N=1561, 54%)	Mean=28 St. Dev=19 Min=0 Max=100 S_k=0.8	"800"-"829" 14 "840"-"849" 19 "870"-"899" 14 "920"-"924" 10 "958"-"959" 20	Mean=$254 St. Dev=$1,081 Min=$1 Max=$30,976 Sum=$734,188 S_k=19.2

To learn more about the distribution of the variables within the data set and to find out whether any transformation of the variables was needed, we performed a simple bivariate exploratory data analysis. The results are summarized in Table 1. The table reveals that for the DAC variable the average dollar amount of the recovered cases (group 1), not recovered (group 2), and not pursued (group 3) are $1,052, $417, and $254, respectively. Furthermore, the table shows that the total amounts for the DAC variable for each of the 3 groups are $472,461, $1,182,350, and $734,188, respectively. Thus it appears that the company used common sense and some procedure that allowed it to target the patients with larger debt amounts and ignore those with smaller debt amounts. Furthermore, the skewness coefficient Sk=19.2 shows that the distribution of the DAC variable is very positively skewed, especially for group 3, which suggests that

small dues were simply not pursued. Because for cases belonging to group 3 ("unknown" cases) debt collection was not pursued by the subject company, we did not use this group to build the models. It is possible that the low debt recovery rate might have been caused by the healthcare company using primarily the amount owed to determine which bad debts to target. The purpose of the data mining techniques is to use the seemingly unrelated factors such as the patient's gender, age and type of injury to determine the likelihood that a particular patient-debtor will pay his/her overdue bill. Therefore, each case from group 3 can be used to test the final models to find out whether the models would classify them as a "good" or "bad" case. As a result, the data set used to build and test the models contained 3,284 cases divided unequally between 449 "good" cases (group 1) and 2,835 "bad" cases (group 2). To improve the distribution of the DAC variable and obtain better prediction results, we computed and used log10(DAC) instead of DAC.

4. The experiments and simulation results

Computer simulation, which consists of five scenarios, was conducted using MatLab Fuzzy Logic Toolbox. In each scenario 50% of the records in the data set were randomly allocated to building the models, whereas 25% of the records were allocated to the models' validation and testing each. Since classification accuracy rates may vary significantly for different partitions/splits of the data set, this process was repeated 50 times and the reported classification rates on the test sets were averaged over the 50 runs to eliminate the classification bias resulting from random splits of the data set and to increase the reliability and generalizability of the results. We used the back-propagation method for training the fuzzy inference system (FIS) membership function parameters and GENFIS1 function to generate the initial FIS. We used two membership functions per input variable and five types of different membership functions. These are two Gaussian membership functions (gauss2mf and gaussmf), generalized bell-shaped membership function (gbellmf), the difference between two sigmoid membership functions (dsigmf), and triangular membership function (trimf). To compare the models' performances across the different scenarios, we used the overall correct classification accuracy rates as well as the rates for good and bad cases. We also utilized the ROC charts, which depict the global performances of the models within the [0, 1] range of cutoffs. Low and high probability cutoffs tend to be in the upper right and lower left areas, respectively, of the ROC curves. To interpret the results we also used 3-dimensional control surfaces generated by ANFIS.

In scenario 1 we used all the 449 good cases and all the 2835 bad cases. In scenario 2 we used the 449 good cases and randomly selected 898 bad cases. In scenario 3 we used all the 449 good cases with randomly selected 449 bad cases. The best overall rates were obtained for cutoff=0.85, cutoff=0.6, and cutoff=0.5 for scenarios 1, 2, and 3, respectively. Tables 2-4 depict the classification rates for the three scenarios. It seems that a choice of the membership function and the number of bad cases, which are different for each of the three scenarios, affect the correct classification accuracy rates. The overall rates are approximately between 60% and 66% across the three scenarios and the best rates appear to be associated with the generalized bell-shaped membership function (gbellmf). The good rates and the bad rates are within the [50%, 64%] range and [60%, 71%] range, respectively. However, for the first three scenarios the ANFIS generally produced worse results than those obtained from the five models (logistic regression, ensemble, memory-based reasoning, neural network and decision tree) described in the Zurada and Lonial papers [12], [13].

The ROC charts generated for scenarios 1 through 3 and gbell membership function confirm these observations (Figure 2). The three curves show that the global performances of the three models are unsatisfactory and they are only somewhat better that of the worthless model. This situation dramatically changes for scenario 4.

In scenario 4 we used all the good cases (449 cases) and 449 bad cases with the highest DAC values. The results of this scenario are given in Table 5. Compared to scenarios 1-3, the overall, good, and bad correct classification accuracy rates improved very significantly. The overall rates vary between 78.3% and 83.1%. The good rates and bad rates are within the ranges [69.2%, 72.3%] and [84.2%, 95.8%], respectively. The improvement in the overall rates was mainly caused by the dramatic increment in the rates for bad cases. The reason may be that the bad cases which have larger DAC values have some apparent features that are easy to identify. These features are DAC, PA, IDC, and to a lesser extent PG. The relationships between any two of these features and the probability of recovery are presented in Figures 3-8. Also, the ROC curve drawn for scenario 4 and the model using a dsigmf membership function attests to the good quality of the model. One can see that the model created in scenario 4 exhibits the best global performance at all probability cutoffs and significantly outperforms the models created in scenarios 1-3. The area under the curve for scenario 4 is the largest. In this

scenario ANFIS generated significantly better results than those obtained from the five models in the Zurada and Lonial studies [12], [13], especially in overall and bad classification accuracy rates.

Table 2. The classification accuracy rates in [%] for Scenario 1 with cutoff=0.85

	gauss2 mf	gbell mf	dsig mf	gauss mf	tri mf
Overall	60.2	64.9	63.6	64.9	64.2
Good	52.3	54.6	55.5	57.8	59.3
Bad	61.5	66.6	65.0	66.1	64.9

Table 3. The classification accuracy rates in [%] for Scenario 2 with cutoff=0.6

	gauss2 mf	gbell mf	dsig mf	gauss mf	tri mf
Overall	63.4	66.1	63.2	64.0	63.5
Good	51.5	50.6	52.8	50.3	51.3
Bad	69.2	70.8	68.7	70.9	69.5

Table 4. The classification accuracy rates in [%] for Scenario 3 with cutoff=0.5

	gauss2 mf	gbell mf	dsig mf	gauss mf	tri mf
Overall	61.6	63.7	62.8	63.1	62.5
Good	61.4	63.5	61.9	63.9	63.5
Bad	61.9	64.2	64.0	62.7	62.0

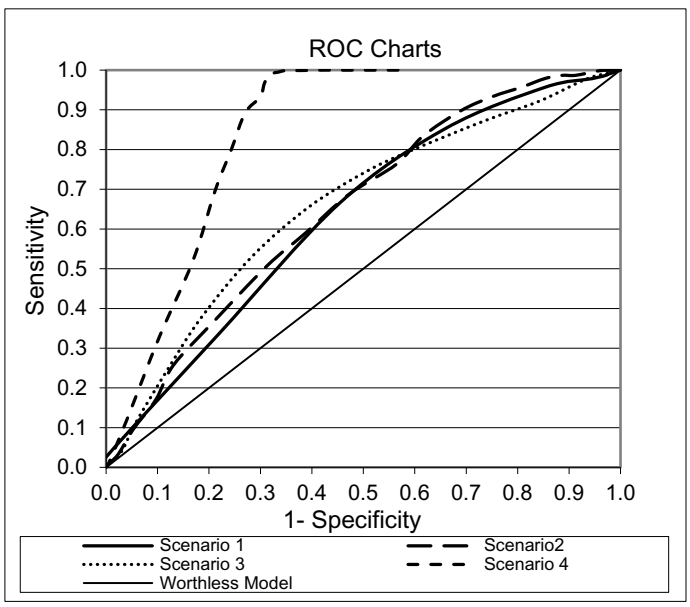

Figure 2. The ROC charts for the best models in the four scenarios

Table 5. The classification accuracy rates in [%] for Scenario 4 with cutoff=0.5

	gauss2 mf	gbell mf	dsig mf	gauss mf	tri mf
Overall	78.3	81.9	83.1	80.4	82.2
Good	72.3	69.7	70.3	71.0	69.2
Bad	84.2	94.3	95.8	89.7	94.9

In the last scenario, scenario 5, the data set was divided into two groups. The first group contains cases whose IDC frequency of occurrence > 5%, i.e., the codes are among the most common. The second group contains cases whose IDC range <= 5%, i.e., the codes are among the least frequently seen in the dataset. The first group contains 291 good cases and 291 bad cases. Both sets of cases were randomly selected from these cases containing IDC values in the top 5% most common codes. The second group contains 158 good cases and 158 bad cases. Both sets of cases were randomly selected from these cases containing IDC values in the least common 5% codes. The models built and tested for the least frequent IDCs tend to perform much better in terms of overall, good, and bad rates than the models for the most common IDCs (Table 6). There does not appear to be an obvious explanation for this particular set of results. However, further investigation into the diseases, disorders, and/or symptoms by a healthcare organization may shed light on this interesting association.

Table 6. The classification accuracy rates in [%] for Scenario 5 with cutoff=0.5

Injury diagnosis code range > 5%					
	gauss2 mf	gbell mf	dsig mf	gauss mf	tri mf
Overall	55.7	58.1	57.9	57.8	59.0
Good	63.7	64.3	60.4	58.6	62.0
Bad	48.8	52.3	56.0	57.5	56.3
Injury diagnosis code range ≤ 5%					
Overall	65.1	67.2	64.9	68.1	65.6
Good	79.0	69.1	74.5	67.2	75.4
Bad	52.4	65.4	55.8	69.6	56.2

Figures 3-8 depict 3-dimesional surfaces generated by ANFIS and they offer a more insight into nonlinear and complex interactions between the variables and the probability of debt recovery/default. They have been plotted for scenario 4 and the membership function dsigmf. For example, Figure 3 represents the probability of default/recovery versus the (log(DAC)) and the IDC. The control surface clearly shows that it is less likely to recover debt for larger values of DAC and for the lower values of the IDC, approximately those IDC codes ≤ 12. As far as the Patient Age (PA) is concerned, Figure 4 shows that older patients are less likely to pay their debt, especially when the amount owed is large. It may be a challenge to fully interpret the local peak in Figure 5. It suggests that younger patients with lower IDCs may be more prone to default. Examination of Figure 6, Figure 7, and Figure 8 also provides interesting insights. For example Figure 6 shows that larger DAC values are likely to lead to default regardless of age. Interpretation of Figure 7 is less straightforward but it seems to suggest that defaults tend to peak for IDC codes around 7. Finally Figure 8 suggests for both genders default rates increase with PA values.

Finally we will discuss the classification of the unknown cases. Exploration of unknown cases to recover potential revenue is important as these unknown cases represent a huge potential recovery source. The objective here is to explore a method to allow healthcare organizations to pursue those unknown cases that are most likely to be recovered. In the dataset used in this study the majority of the cases are labeled "unknown". As mentioned earlier these cases contain customers who were <u>not</u> targeted for debt collection, or in other words, the debt was not pursued. As a result, they were labeled "unknown" because it was impossible to determine whether those customers would have paid the debt off or defaulted upon it. We used the best models in the first four scenarios to predict/classify the 2896 unknown cases. The function evalfis() in MatLab was used to compute a recovery (good case) probability of a unknown case using the best model in each of the different scenarios described earlier in this section. The output of the prediction is either "0" representing "good cases" or "1" representing "bad cases." If the return value was less than the threshold, the corresponding case would be classified as a good case (a recoverable case); otherwise the case was classified as a bad case. Two probability thresholds were selected in each case. For the model in the first scenario for example the first threshold is 0.15. This value was chosen to reflect the fact that the model in scenario 1 was built with cases where only one third of them were good cases. The second threshold value (in each case) was the first threshold value + 0.1. The increase in threshold value was intended to see its effect on the resulting classification. In other words the higher threshold value is the more likely it is for the resulting cases to be recovered. The thresholds for scenarios 2 and 3 in Table 7 were similarly selected.

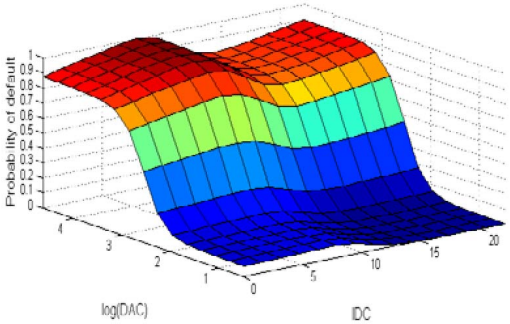

Figure 3. Control surface

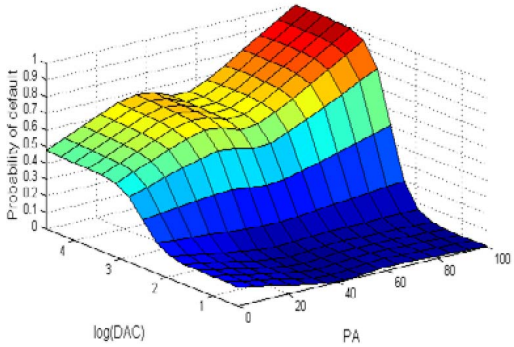

Figure 4. Control surface

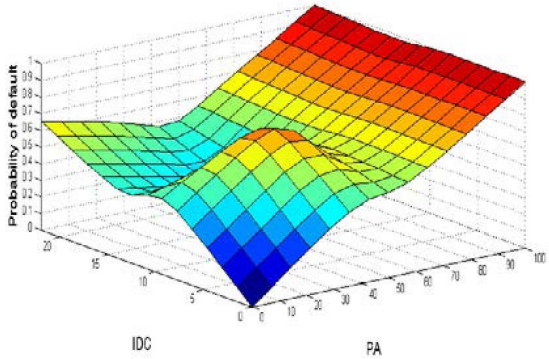

Figure 5. Control surface

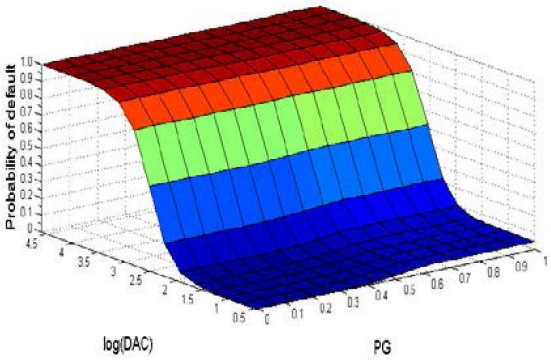

Figure 6. Control surface

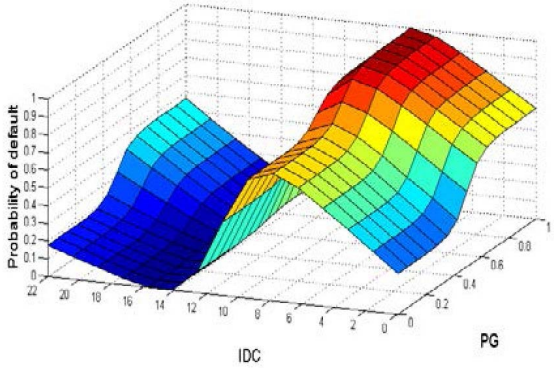

Figure 7. Control surface

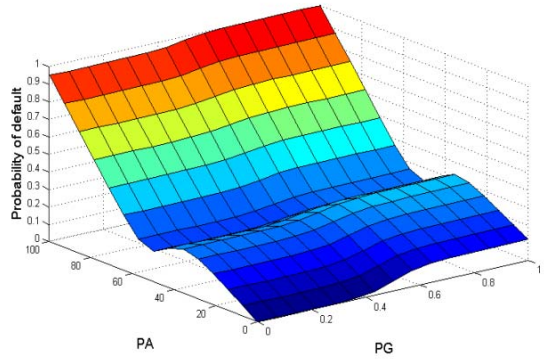

Figure 8. Control surface

Table 7. The results from the classification of the "unknown" cases

Scenario	Recovery probability threshold	Number and percentage of cases classified as "good"	The potential amount recovered [$]
1	0.15	917/2896 31.66%	576,720
	0.25	153/2896 5.28%	352,772
2	0.4	837/2896 28.90%	590,990
	0.5	170/2896 5.87%	304,927
3	0.5	1503/2896 51.90%	676,730
	0.6	697/2896 24.07%	394,014
4	0.5	2171/2896 74.97%	139,195
	0.6	2001/2896 69.10%	110,481

In Table 7 one can see that scenario 3 has the best predictive result as the potential amount of recovery is $676,730 when the recovery probability threshold is 0.5. However, when the threshold value increases to 0.6, the number of cases/recovery amount decreases since the resulting cases are potentially more likely to be recovered. The results for the other models can be similarly interpreted/analyzed. The model from scenario 4 produced the worst results in terms of the recovery amount because the model was trained with cases having the highest DAC values. Since the unknown cases contain a lot of cases with low DAC values, the poor results were expected. This result is interesting as healthcare organizations now could use this method to explore unknown cases, cases that they currently ignore. In addition, our models can provide the patient financial service department a list of "unknown" patients sorted from the most likely to pay the bill to the least likely to pay the bill.

5. Conclusion

The paper explores the effectiveness of ANFIS in recovering bad debt in the healthcare context. The data analysis and evaluation of the performance of the various test scenarios is based on a fairly large unbalanced data sample provided by a healthcare company, in which cases with recovered bad debts are underrepresented. This research was motivated by an urgent and recognized need to better understand the effectiveness of data mining methods in bad debt recovery in the healthcare industry and relatively low level of academic interest in this field. This paper describes a study that explores the effectiveness of ANFIS in classifying bad debts. Five different test scenarios were designed and tested. The best rates were obtained for scenario 4 in which the bad cases with the largest DAC values were used. The results and the approach in this study could potentially help health-care organizations target those customers with a high level of debt, thus improving their return on debt recovery efforts. The control surfaces present revealing relationships between the probability of recovery/default and the two other variables. Finally the paper also shows the ability of data mining models to classify unknown cases, which are a potential source of revenue recovery. This preliminary study shows the potential of data mining models to classify bad debts using data sets whose features contain very tangential information about the patients. An important fact about the dataset used in the study is the high proportion of unknown cases. These unknown cases often represent huge amounts of possible recoverable revenue.

Further research can examine the effectiveness of data mining models in bad debt recovery by focusing on classification of unknown cases as well as in-depth examination of control surfaces and interpretable rules generated by ANFIS. It would also be interesting to examine specific IDCs to find out if patients who suffered from a more serious injury (such that may lead to disability) are less likely to pay the debt off than patients who were treated for minor injuries. Finally, it may be advisable to explore the effect of data clustering on classification accuracy. The bad debt data could be divided into subsets with clustering algorithms and a classification model is created for each subset. Given the good classification rates obtained for the high DAC cases such a data set obviously could be divided into clusters by debt amounts but other cluster boundaries could also be explored.

Acknowledgment: We express gratitude to Professor Subhash Lonial, Ph.D., Department of Marketing, College of Business, University of Louisville, for providing the data set which was used for computer simulation in this paper.

6. References

[1] Buczko, W., "Factors Affecting Charity Care and Bad Debt Charges in Washington Hospitals", *Hospitals and Health Services Administration*, 39(2), 1994, 179-191.

[2] Galloro, V., "Bad News on Bad Debt", *Modern Healthcare*, 33(43), Oct 2003, 8.

[3] IBM, "Helping Hospitals Capture More Revenue," http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=YTC03052USEN&appname=wwwsearch, June 12, 2013.

[4] Jang, J. S. R., "ANFIS: Adaptive-network-based fuzzy inference system. "*IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, 1993, 665-685.

[5] Lagnado, L., "Hospitals Try Extreme Measures to Collect their Overdue Debts, *The Wall Street Journal*, Oct 30, 2003.

[6] Pell, M.B., "Patients in Arrears Face Collectors", *The Atlanta Journal-Constitution*, June 6, 2011.

[7] Pesce, J., "Stanching Hospitals' Financial Hemorrhage with Information Technology", *Health Management Technology*, 24(8), 2003, 12.

[8] Sugeno, M. and G. T. Kang, "Structure identification of fuzzy models," *Fuzzy Sets and Systems*, 28, 1988, 15-33.

[9] Takagi, T., Sugeno M., "Fuzzy identification of systems and its application to modeling and control," *IEEE Transactions on Systems Man and Cybernetics*, 15(1), 1985, 116-132.

[10] Veletsos, A., “Getting to the Bottom of Hospital Finance”, *Health Management Technology*, 24(8), 2003, 30.

[11] Zollinger, Terrell W., Sawyell, Robert M. Jr., Chu, David K. W., Ziegert, Andrea, Woods, John R., LaBov, Dean, "A Determination of Institutional and Patient Factors Affecting Uncompensated Hospital Care", *Hospital & Health Services Administration*, Chicago, 36(2), 1991, 243-256.

[12] Zurada, J., and Lonial, S., “Application of Data Mining Methods for Bad Debt Recovery in the Healthcare Industry”, in *Proceedings of the 6th International Conference on Databases and Information Systems*, J. Barzdins (Ed.), Vol. 672, 2004, 207-217.

[13] Zurada, J., and Lonial S., “Comparison of the Performance of Several Data Mining Methods for Bad Debt Recovery in the Healthcare Industry”, *Journal of Applied Business Research*, 21(2), Spring 2005, 37-54.