Journal of International Technology and Information Management

Volume 24 | Issue 4 Article 1

2015

An Adaptive Neuro-Fuzzy System with Semi-Supervised Learning as an Approach to Improving Data Classification: An Illustration of Bad Debt Recovery in Healthcare

Donghui Shi
Anhui Jianzhu University

Jozef Zurada
University of Lousiville

Jian Guan
University of Louisville

Sandeep Goyal
University of Louisville

An Adaptive Neuro-Fuzzy System with Semi-Supervised Learning as an Approach to Improving Data Classification: An Illustration of Bad Debt Recovery in Healthcare

Donghui Shi
Department of Computer Engineering
School of Electronics and Information Engineering
Anhui Jianzhu University, Hefei
CHINA
Prometeo Researcher
Universidad Técnica Particular de Loja
ECUADOR

Jozef Zurada
Jian Guan
Sandeep Goyal
Department of Computer Information Systems
College of Business
University of Louisville
USA

ABSTRACT

Business analytics has become an increasingly important priority for organizations today as they strive to achieve greater competitiveness. As organizations adopt business practices that rely on complex, large-scale data, new challenges also emerge. A common situation in business analytics is concerned with appropriate and adequate methods for dealing with unlabeled data in classification. This study examines the effectiveness of a semi-supervised learning approach to classify unlabeled data to improve classification accuracy rates. The context for our study is healthcare. The healthcare costs in the U.S. have risen at an alarming rate over the last two decades. One of the causes for the rising costs could be attributed to medical bad debt, i.e., debt that is not recovered by healthcare institutions. A major obstacle to debt classification, hence better debt recovery, is the presence of unlabeled cases, a situation not uncommon in many other business contexts. There is surprisingly very little research that explores the performance of computational intelligence and soft computing methods in improving bad debt recovery in the healthcare industry. Using a real data set from a healthcare organization, we address this important research gap by examining the performance of an adaptive neuro-fuzzy inference system (ANFIS) with semi-supervised learning (SSL) in improving debt recovery rate. In particular, this study explores the role of ANFIS in conjunction with SSL in classifying unknown cases (those that were not pursued for debt collection) as either a good case (recoverable) or a bad case (unrecoverable). Healthcare institutions can then pursue these potentially good cases and improve their debt recovery rates. Test results show that ANFIS with SSL is a viable method. Our models generated better classification accuracy rates than those in prior studies. These results and their analysis show the potential of ANFIS with SSL models in classifying unknown cases, which are a potential source of revenue recovery for health care organizations. The significance of this

  ISSN: 1543-5962-Printed Copy ISSN: 1941-6679-On-line Copy

research extends to all types of organizations that face an increasingly urgent need to adopt reliable practices for business analytics.

Keywords: bad debt recovery, healthcare industry, adaptive neuro-fuzzy inference system (ANFIS), semi-supervised learning, classification

INTRODUCTION

Organizations today realize that they have to embrace evidence-based decision making to achieve greater competitiveness (McAfee & Brynjolfsson, 2012). In an increasingly digitized environment organizations are adopting new and improved methods in analytics or business analytics (BA) (Holsapple et al., 2014). Thus the investigation, adoption, and application of business analytics have become a priority for chief information officers. Yet there is relatively little academic inquiry into this increasingly important area from both a practice and academic perspective (Holsapple et al., 2014). A direct result of the explosive growth in data from digitization is the challenge posed by imbalanced data sets (He & Garcia, 2009), particularly when the imbalanced data contain a disproportionately large number of unlabeled data records (Chapelle et al., 2006). Traditional classifiers only use labeled data (features/target pairs). However, labeled data can often be difficult, expensive to obtain or simply unavailable. Example situations include automatic classification of web pages and classification of medical bad debt (Zurada & Lonial, 2004). With the increasing proliferation of complex and large-scale data in organizations, a better understanding of the applicability and effectiveness of such BA approaches is therefore warranted. This study examines a semi-supervised learning (SSL) approach to improve data classification in the presence of unlabeled data. The context for our study is healthcare, i.e., classifying bad debt in a very important sector of the economy, the healthcare sector.

The healthcare costs in the US have been steadily rising during the last two decades. In 2013 U.S. health care spending increased 3.6 percent to reach $2.9 trillion, or $9,255 per person, the fifth consecutive year of slow growth in the range of 3.6 percent and 4.1 percent. The share of the economy devoted to health spending has remained at 17.4 percent since 2009 as health spending and the Gross Domestic Product increased at similar rates for 2010 - 2013 (https://www.heartland.org/policy-documents/national-health-expenditures-2013-highlights).

One of the factors that contributes to such high costs is medical bad debt (Albright, 2013; Galloro, 2003; Kutscher, 2013; Pell, 2011; Veletsos, 2003). Medical bad debt includes unpaid patient bills for medical treatment, outstanding medical testing costs, and collection agency fees. According to Galloro (2003), the Nashville-based Hospital Corporation of America's provision for bad debt rose to an astounding 10.3% of its net revenue for just one quarter of 2003, compared to an already high 8.3% of revenue in the same quarter the previous year of 2002. Even when unpaid bills are eventually paid, hospitals typically end up paying 30% to 50% of recovered bad-debt revenue to outside collection agencies (Veletsos, 2003). With the recent legislative changes (i.e., the Affordable Care Act), there are indications of decrease in the number of uninsured patients. A number of these newly insured patients are enrolled in high-deductible plans (Albright, 2013), and the bills from patient deductibles may add to the bad debt recovery issue. 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 (Pell, 2011). According to an analysis from Citi projects bad

  ISSN: 1543-5962-Printed Copy ISSN: 1941-6679-On-line Copy

debt could reach $200 billion by 2019 (Kutscher, 2013). Recovering bad debt has become a serious matter and has led hospitals to suing patients in several states in the U.S. According to information provided by collection attorneys, consumer advocates, and court records, some hospitals use extreme measures to collect bad debt and even seek arrest of patients who miss court hearings related to their healthcare debts (Lagnado, 2003).

Thus a pressing issue in medical bad debt recovery is an effective approach to identifying cases of debt that are worth recovering. However, there is curiously little available academic research on bad debt recovery in the health care context. But there is a great deal of research on scoring, managing, and recovering distressed debts from a loan, a credit line, or an accounts receivable. For example, Murgia and Sbrilli (2012) compare the performance of a neural network, integer programming, hidden Markov model, logistic regression, regression tree, and Bayesian classifier in distressed debts recovery. Murgia and Sbrilli report that the neural network scored the recovery rate of each distressed debt better than the other models. However, these advances in the reference literature (e.g., finance) may not be readily applicable to the healthcare context. Predicting whether a particular patient is likely to repay a healthcare debt is an inherently complex and unstructured process. What makes this process difficult in the healthcare context, especially in emergency rooms and acute care facilities, is their inability to obtain detailed financial information concerning the patients. Unlike a financial institution which would collect financial, social, and personal information about customers and carefully evaluate whether to extend them a loan, healthcare institutions must often admit a patient and perform the necessary medical procedures on credit with very little knowledge about that particular patient. This lack of information makes it difficult to predict whether a patient-debtor will pay his/her bill or not. Thus, due to moral, legal and practical constraints, healthcare providers in the U.S. often become unwilling creditors to a multitude of borrowers. According to Pesce (2003) a healthcare institution is handicapped by having only a small number of independent attributes of the patient-debtor for evaluation. This lack of attributes situation can be further exacerbated by the presence of a disproportionately large subset of debt cases with a reduced chance for recovery. In other words the available data can be very imbalanced, containing a large number of unlabeled cases. The imbalanced nature of the data can present significant challenges to traditional methods of data classification. Despite an increasingly obvious and urgent need for predictive and classification models of bad debt in the healthcare industry, academic research on this very important topic appears to be surprisingly scarce.

Therefore improving bad debt recovery rate is important and will remain important to the healthcare industry. And such improvement can be made possible by better models for classifying bad debt in the presence of a large set of unknown cases. In this study we explore an approach to improve the performance of classification models through semi-supervised learning (SSL). In particular this study explores the role of an adaptive neuro-fuzzy inference system ANFIS in conjunction with SSL in classifying unknown cases (those that were not pursued for debt collection) as either a good case (recoverable) or a bad case (unrecoverable). ANFIS is one of the best-known classification models that combine the benefits of fuzzy logic and neural networks (Jang, 1993). However, models like ANFIS depend on the availability of known input-output pairs. In an imbalanced data set where a large subset of cases are unlabeled, a classification model such as ANFIS would not be able to utilize the unlabeled data. SSL, in conjunction with ANFIS, makes possible learning from unlabeled data as well (Chapelle et al., 2006; Drummond, 2003). Thus

  ISSN: 1543-5962-Printed Copy ISSN: 1941-6679-On-line Copy

healthcare institutions can then pursue a hitherto untapped source and improve their debt recovery rates. Test results obtained with data from a healthcare organization show that ANFIS with SSL is a viable approach.

The rest of the paper is organized as follows. Section 2 summarizes the prior literature on debt recovery and presents the basic tenets of ANFIS with SSL used in the study. Section 3 describes the data sample and the experiment. Section 4 presents the simulation results as well as compares them to two previous studies. Finally, section 5 concludes the paper and provides recommendations for future work.

BACKGROUND

Though there is a healthy amount of existing work on debt recovery in the finance area, very little is available in healthcare debt recovery. An early study by Zollinger et al. (1991) examined a sample of 985 patients from 28 Indiana hospitals 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, as significant factors in recovering unpaid hospital bills. Similarly, Buczko (1994) analyzed data on charges assigned to bad debt for 82 short-stay hospitals in Washington. Buczko confirmed that unpaid care has become a serious problem in hospital finance because of the increasing number of uninsured patients and declining hospital revenues. Using data from approximately 2400 patients of the Florida Hospital in Orlando, Veletsos (2003) presented a more comprehensive study on using predictive modeling software such as IBM Intelligent Miner and DB2 for bad-debt recovery. The final model, as described by Veletsos, is based on a variety of data variables, including credit factors, demographic information, and previous organizational payment patterns. The model yields approximately $200,000 in savings. Pesce (2003) argued that hospitals should invest in modern information technology to reduce bad-debts, which along with other factors such as billing errors, insurance underpayments, and inability to collect accurate patient and payer information throughout delivery of care, account for 13% of a hospitals' lost revenue each year.

Zurada and Lonial (2004, 2005) were among the first studies that leveraged the advanced computational intelligence tools in recovering bad debt. Their results show that the logistic regression, neural network, and the ensemble models produce the best overall classification accuracy, and the decision tree is the best in classifying cases for which the debt has been recovered. The models presented by Zurada and Lonial were also used to score the "unknown" cases – those that were not pursued by a company. The neural network model classifies more "unknown" cases into "good" (recoverable) cases than any other models tested in the studies. Bradley and Kaplan (2010) argued that predictive analytics can identify root causes and trends causing missing charges or bad debt to provide executives with strategic intelligence to prevent further revenue leakage or violated contract terms as well as accelerate the resolution of credit accounts and reduce bad debt. More recently, IBM reported that the use of predictive software in IBM SPSS can improve the bad debt collection effort and boost revenue. Though no details (such as models and methods used) were provided, the report states 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" (IBM, 2013). Finally, Shi et al. (2014) examined the effectiveness

of a neuro-fuzzy method (ANFIS) under numerous scenarios in classifying bad debt. ANFIS is a powerful inference system that integrates both neural networks and fuzzy logic principles and has the ability to approximate nonlinear functions (Jang, 1993). Work done by Shi et al. showed that the classification accuracy rates and scoring of "unknown" cases using ANFIS outperformed the methods used in the studies by Zurada and Lonial (2004).

Work done by Shi et al (2014) was a significant first step in introducing neuro-fuzzy methods to addressing complex and unique debt recovery issues. In this research, we aim to build upon and go beyond this existing literature by using ANFIS in conjunction with semi-supervised learning (SSL). SSL is a class of machine learning methods that make use of unlabeled data for training (Chapelle et al., 2006). While ANFIS is now established as one of the best known classification methods, it still depends on the availability of known input-output pairs. Such pairs would require a balanced dataset, which is often not practically feasible. With the proposed SSL unlabeled data (unknown cases) could be labeled through the use of already labeled data for considerable improvement in data classification rates, and does not require a balanced dataset. In the case of bad debt recovery in the healthcare context SSL presents an attractive potential solution to improve debt recovery by allowing more debt cases to be pursued and creating a better model for debt classification. More specifically, this study explores an approach to improve the classification rate of bad debt cases using ANFIS with SSL. The use of SSL provides a viable method for improving a bad debt classification model that incorporates the unlabeled cases, which are otherwise ignored or not pursued. In doing so, we compare and contrast our classification accuracy rates to those in two previous studies—namely Shi et al. (2014) and Zurada & Lonial (2005). The target/dependent variable in a fairly large data set provided by a healthcare institution 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). As both ANFIS and SSL are crucial to the reclassification of data in our experiment, in the next section we describe the basic principles of ANFIS and SSL (Chapelle, et al., 2006).

METHODS

Adaptive Neuro-fuzzy Inference System

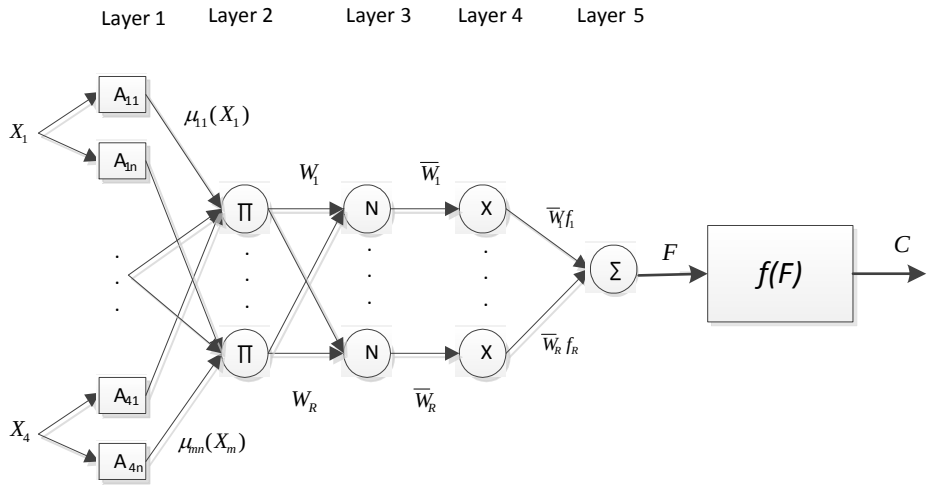
Neural fuzzy inference systems have emerged from the fusion of artificial neural networks and fuzzy inference systems. These systems combine learning/training and optimization abilities of artificial neural networks with human-like reasoning using if-then fuzzy rules offered by fuzzy inference systems. 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 (Jang, 1993). 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 2—namely, (1) Patient Age (PA), (2) Patient Gender (PG), (3) Injury Diagnosis Code (IDC), and (4) Dollar Amount of the Claim (DAC). 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 (Sugeno & Kang, 1988; Takagi & Sugeno, 1985). A typical TSK rule has the following structure:

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

where $A_{i,j}$ is the j^{th} linguistic term (such as high, low) of the i^{th} 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 2 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 j^{th} 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 k^{th} 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 k^{th} TSK-type rules as defined in layer 4.

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

Semi-supervised learning

Semi-supervised learning leverages both labeled and unlabeled data examples. It aims to combine unsupervised learning, which utilizes no data labels, and supervised learning, which utilizes data for which labels are present. Many studies have shown that adding unlabeled data to a small amount of labeled data as examples can produce considerable improvement in learning accuracy (https://en.wikipedia.org/wiki/Semi-supervised_learning). The learning algorithm generally uses a smoothness assumption, which states that if two examples are relatively close in feature space, then their corresponding class outputs should be close in class space. In a semi-supervised learning a classifier is iteratively built on its own predictions. Under-sampling is one of the most common sampling methods used to process the class imbalanced problem by removing some data from majority classes (Drummond, 2003). In the following algorithm, first use under-sampling to sample the balanced data from the labeled data, construct a classifier based on the data, and then used to classify unlabeled data. Typically the most confidently predicted examples are iteratively inserted into the training set and a new classifier is generated (Chapelle et al., 2006). Self-training is a wrapper method for semi-supervised learning. A basic classifier could be implemented using a *k*-nearest neighbor method, support vector machines, neural networks, or logistic regression in the wrapper algorithm.. In this paper we propose an ANFIS-based self-training algorithm. An

ANFIS model was used as a basic classifier. 60% of the data were used for the training set and 40% for the validation set in constructing an ANFIS-based predictive model. Under-sampling was used to solve the imbalanced problem. To our best knowledge an ANFIS-based classifier has not been used in a self-training algorithm.

An ANFIS-based self-training algorithm for the imbalanced data follows.

Input: L_I – the original labeled data set; U – unlabeled data set

Output: L_F – final labeled data set with all cases from U classified as good or bad

Repeat:

- Merge all good cases from L_I and the same number of bad cases randomly selected from L_I to produce a new dataset D with balanced class labels
- Use D to construct an ANFIS-based predictive model F using 60% for the training set and 40% for the validation set
- Classify unlabeled data set U with F (the ANFIS-based model)
- Compute the error
- Select case u with the minimum error and move the case from U to the L_F: $L_F=L_I+u$; $U=U-u$

Until U is null.

THE FIELD EXPERIMENT

The healthcare company, whose data were used in this study, relied on only four simple input factors to determine whether a 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 6117 cases with a total outstanding balance of $2,381,453. The dependent variable, Status, comprises of 449 "good" cases (group 1), 2833 "bad" cases (group 2), and 2835 "unknown" cases (group 3).

The "good" cases are significantly underrepresented in the data set. 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 calculated the descriptive statistics. The results are summarized in Table 1. The table shows 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 $256, respectively. The table also shows that the total amounts for the DAC variable for each of the 3 groups are $472,461, $1,182,142, and $727,850, respectively. The skewness coefficient (S_k=19.1) shows that the distribution of the DAC variable is very positively skewed for group 3, which suggests that small debts were simply not pursued. 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. Because we found the DAC to be significantly skewed for all three groups, we used log(DAC) instead of DAC to improve the distribution of the DAC variable and obtain better prediction results.

Unknown cases are those for which the hospital did not pursue for debt recovery. We conducted some preliminary analysis to determine these factors. As described later, a simple descriptive

analysis showed that the average outstanding dollar amount for unknown cases was $256. Compared to the average dollar amount for good cases ($1,052) and bad cases ($417), this amount was low. It is plausible that the healthcare institution classified some cases as unknown primarily because of the dollar amount owed. It is reasonable to argue that it may not be economically feasible for the healthcare institution to pursue the low dollar amount cases because of the collection agency fees. We, however, argue that if the unknown cases can be classified as good cases or bad cases, the healthcare institution can consider pursuing good cases as the likelihood of recovering debt is high.

For cases belonging to group 3 ("unknown" cases) debt collection was not pursued by the subject company, but these unknown cases may represent a potential source of debt recovery. Given that the unknown cases represent a little over 30% of debt for the healthcare institution that provided us data, such a recovery would have a significant impact on their bottom line. The purpose of our study is to use the seemingly unrelated factors such as the patient's gender, age, the dollar amount of debt, and type of injury to determine the likelihood that a particular patient-debtor will pay his/her overdue bill. To build the ANFIS model we only used all the cases representing "good" customers and an equal number of randomly selected "bad" cases. The model was then used in SSL to classify the unknown cases. The labeling of "unknown" cases into "good" or "bad" could provide additional revenue to the company.

Table 1: Summary of the descriptive statistics for the variables.

Status	Patient Gender (PG)	Patient Age (PA)	Dollar Amount of the Claim (DAC)
Overall (groups 1,2,3)	Female (N=2884, 47.1%) Male (N=3233, 52.9%)	Mean= 30 St. Dev=21 Min=0 Max=100 S_k=0.7	Mean=$389 St. Dev=$1,477 Min=$1 Max=$40,508 Sum=$2,381,453 S_k=12.8
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	Mean=$1,052 St. Dev=$3,442 Min=$3 Max=$40,508 Sum=$472,461 S_k=7.2
2 (Bad) N=2833	Female (N=1331, 47.0%) Male (N=1502, 53.0%)	Mean=31.6 St. Dev=23 Min=0 Max=100 S_k=0.6	Mean=$417 St. Dev=$1,248 Min=$1 Max=$19,568 Sum=$1,181,142 S_k=7.8
3 (Unknown) N=2835	Female (N=1305, 46.0%) Male (N=1530, 54.0%)	Mean=28 St. Dev=19 Min=0 Max=100 S_k=0.8	Mean=$256 St. Dev=$1,091 Min=$1 Max=$30,976 Sum=$727,850 S_k=19.1

Table 2: The Description of the seven most frequently occurring IDCs and their frequencies.

IDC Group	Code Range	Code Description	Frequency of Occurrence [%]			
			Overall	Good	Bad	Unknown
4	"800-829"	Fractures	16	13	18	14
5	"830-839"	Dislocations	4	4	5	4
6	"840-848"	Sprains & Strains	18	22	17	19
9	"870-897"	Open Wounds	14	10	15	14
12	"910-919"	Superficial Injuries	5	2	5	5
13	"920-924"	Contusion w/Intact Skin	9	13	8	10
18	"958-959"	Complication & Unspecified Injuries	19	15	19	20

Out of 24 IDC groups numbered 1-24, Table 2 shows the seven most frequently occurring IDC groups. These seven IDC groups account for about 85% of all IDC groups. The other 17 IDCs are not shown in the table because the frequency of occurrence of each is between 1% and 3%. As a result, for these less frequent IDCs we do not have a sufficient number of cases to build and test the models. These unlisted IDCs fall into the following categories: Miscellaneous I, Miscellaneous II, Lung Disorders, Head Injuries (Ex. Fractures), Internal Injuries, Blood Vessel Injuries, Late Effects of Injuries, Crushing Injuries, Effect of Foreign Body Entering Orifices, Burns, Nerve and Spinal Cord Injuries, Poisoning by Drugs, Toxic Effects of Non-medicinal Substances, Other Unspecified Effects of External Causes, Complication of Surgical & Medical Care, Other Effects of Medical Care, and Accident Cause Codes.

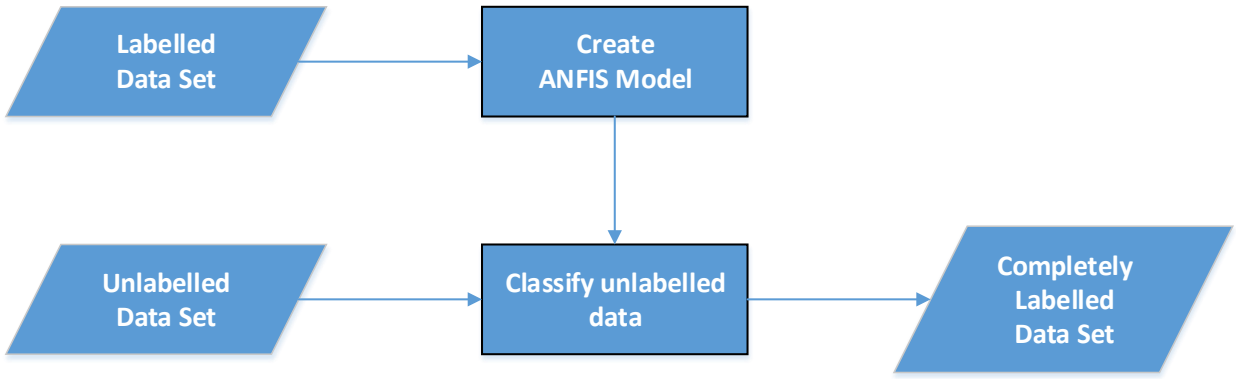
As is common among healthcare institutions, the healthcare institution that provided data for this study recovered bad debts from only 7.3% of the non-paying patients. Due to the low recovery rate, the number of "good" customers is vastly underrepresented in the data set. To build the ANFIS model we used all the cases representing "good" customers and an equal number of randomly selected "bad" customers. The model was then used in SSL to classify the unknown cases. The labeling of "unknown" cases into "good" or "bad" provides a potential source of additional revenue to the company.

We tested our approach with different membership functions for ANFIS using first unclustered data and then the data clustered by seven most frequently occurring diagnostic codes.

Originally the labeled data set contained 3282 cases of which 449 were good cases and 2833 were bad cases. Unlabeled data set contained 2835 cases. After employing ANFIS with SSL, the final data set contained 6117 cases of which 1657 were classified as good cases and 4460 as bad cases. Post classification there were no unlabeled cases as they were classified either as good or bad cases. The block diagram Figure 2 depicts our model of ANFIS with SSL for medical bad debt recovery.

ISSN: 1543-5962-Printed Copy ISSN: 1941-6679-On-line Copy

Figure 2: The diagram of ANFIS with SSL.



RESULTS

In this section, we first present the results of classification of unlabeled data. In order to determine the efficacy of our model, we compare the results of our model to those of prior research—namely, Shi et al. (2014) and Zurada and Lonial (2005). Further validity of our results is provided using the comparison of the descriptive statistics of re-classified data with the descriptive statistics of unclassified data. To develop a better understanding of classifying unlabeled data using ANFIS and SSL, in general, and bad debt recovery in the healthcare industry, in particular, we conduct several post-hoc analysis. First, we develop non-linear response surfaces that demonstrate the complex relationship between bad debt recovery and a variety of variables (e.g., age) that predict bad debt recovery for the healthcare industry. Second, we analyze and present results of these complex non-linear models using clustered data.

Testing was conducted using MatLab Fuzzy Logic Toolbox. The initial ANFIS model was built with all the good cases (499) and an equal number of bad cases randomly selected from the 2833 bad cases. 50% of this data set containing 998 cases were randomly allocated to building the model, 25% of the cases were allocated to the model's validation and the remaining 25% for testing. Normally in such models the classification accuracy rate is determined by the rate of correctly predicted classes (as compared to the actual classes). Because the 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 possible 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 tested five different types of 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 a triangular membership function (trimf).

The number of the premise parameters with input membership functions are 4 for gauss2mf, 3 for gbellmf, 4 for dsigmf, 2 for gaussmf, and 3 for trimf. The number of consequent parameters with output member functions is 5, and the number of output membership functions is 16, which is the number of rules generated in a Sugeno-type fuzzy inference system. The output membership

function type is linear. The total number of parameters can be calculated as (the number of input membership functions × the number of premise parameters + the number of consequent parameters × the number of rules). For example, for gaussmf the total number of parameters is: (4*2+5*16)=88.

Comparison with results from previous studies

To compare the models' performances, 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, to compare the global performance of the models created in this study with those presented in scenario 3 of Shi et al. (2014) and the results from Zurada and Lonial (2005). In both of these previous studies and in this study a comparable data set was used. To interpret the results we also used 3-dimensional control surfaces generated by ANFIS for the unclustered data containing all 24 IDCs and the data clustered by the five most frequently occurring IDCs shown in Table 2.

Table 3 through 5 show the correct classification accuracy rates for unclustered data at the 0.5 cut-off point from the current study, Shi et al. (2014), and Zurada and Lonial (2005). Table 3 shows that the choice of the membership function affects the classification accuracy rates. The overall rates are between 74.8% and 76.7% and the best rates came from the model using the generalized bell-shaped membership function (gauss2mf). The rates for good cases are within the [80.9%, 84.0%] range and the rates for bad cases are within the [66.9%, 70.9%] range. ANFIS with SSL (Table 3) generated better results than those obtained from ANFIS alone (Shi et al., 2014; Table 4) as well as those results from the three best models (neural network, logistic regression, and ensemble model) reported by Zurada & Lonial (Table 5). For example, one can see in Table 4 that the overall correct classification accuracy rates oscillate between 61.6% and 63.7%, whereas the same rates in Table 5 for the three selected models are within the [72.3%,75.0%] range. The notable improvement in the overall correct classification accuracy rates in this study could be mainly attributed to the increment in the rates for good cases and decrement in the rates for bad cases as a result of labeling the unknown cases through SSL. This is important because it leads to a higher recovery of debt.

Table 3: The correct classification accuracy rates in [%] from this study: ANFIS with SSL.

	gauss2 mf	gbell mf	dsig mf	gauss mf	tri mf
Overall	76.7	76.3	74.8	75.9	75.5
Good	84.0	83.1	82.6	80.9	81.5
Bad	69.3	69.6	66.9	70.9	69.3

Table 4: The correct classification accuracy rates in [%] for Scenario 3: ANFIS alone (Shi et al., 2014).

	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

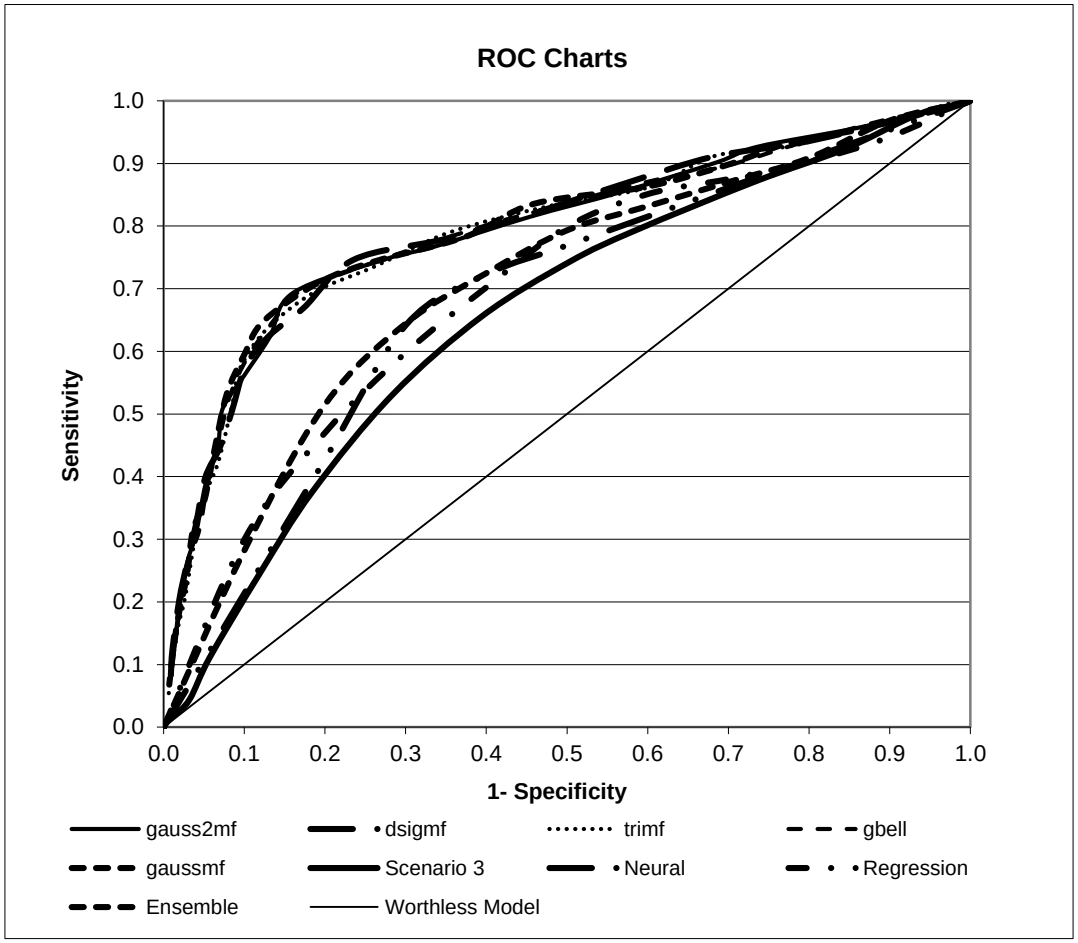
Table 5: The correct classification accuracy rates in [%] for the three best models from the Zurada and Lonial (2005) study.

	Neural Network	Logistic Regression	Ensemble Model
Overall	72.3	75.0	73.7
Good	67.9	71.4	71.4
Bad	76.8	78.6	75.9

One can also compare the global performance of the models at a continuum of cut-offs from within the range [0,1] by examining the areas under the ROC charts presented in Figure 3. The straight line represents the worthless model, identical to completely random guesses. The more the curves push up and to the left upper corner of the coordinates (0,1), the better the models are. Low and high probability cutoffs tend to be in the upper right and lower left areas, respectively, of the ROC curves (Fawcett, 2006). The five partially overlapping curves on the leftmost part of the chart demonstrate that there is little difference among the global performances of the five models, one for each of the five different membership functions—namely, the two Gaussian membership functions (gauss2mf and gaussmf), generalized bell-shaped membership function (gbellmf), the difference between two sigmoid membership functions (dsigmf), and a triangular membership function (trimf). However, all five of these models significantly outperform the model by Shi et al. (2014) and the model by Zurada and Lonial (2005) (represented by four curves just above and to the left of the straight line).

ISSN: 1543-5962-Printed Copy ISSN: 1941-6679-On-line Copy

Figure 3: ROC charts for the current study, Shi et al. (2014), and Zurada & Lonial (2005).



Results of labeling unknown cases

To obtain a deeper understanding of our models and to theoretically validate the validity of our models, we evaluated the descriptive statistics of just the unknown cases as presented in Tables 6 and 7. Table 6 shows the descriptive statistics of 2835 unknown cases that were classified as 1208 good cases and 1627 bad cases. The group of unknown cases has a larger proportion of males (54%) than females (46%) with average DAC of $256 and average age of 28 years. It appears that younger female patients with an average age of 19 years are more likely to pay off their medical debt than males (59.7% vs. 40.3%.). These descriptive statistics also demonstrate that models learn to pursue larger debts with an average of $454 while ignoring lower debts with an average of $110. The results of the descriptive analysis also show that for certain types of injuries significantly more unknown cases are classified as good cases than bad cases. More specifically, these results show that it is easier to recover bad debt for injuries that represent fractures ("800-829"), sprains & strains ("840-848"), and open wounds ("870-897") than for injuries related to contusion w/intact skin ("920-924") and complication & unspecified injuries ("958-959"). One explanation could be that it may be easier to recover bad debt for injuries with more readily defined diagnosis such that patients are sure about the cost of treatment. When the chances of prolonged treatment are higher,

ISSN: 1543-5962-Printed Copy ISSN: 1941-6679-On-line Copy

patients might perceive they may be unable to pay for the treatment and decide not to pay for smaller amounts that they can potentially pay.

Table 7 represents the final descriptive statistics for the resulting cases after labeling the unknown cases with ANFIS using SSL. The format of the table is very similar to that of Table 6, except that there are no unknown cases in Table 7 as they have already been classified as good cases or bad cases. Comparing Table 2 with Tables 6 and 7 confirms the findings described above. For example, females are more likely to repay their medical bad debt than males, as mentioned before. It also appears that younger patients are less likely to default on their bad debt. Also, it is much more likely to recover the bad debt for medical procedures represented by the following IDCs: Fractures ("800-829"), Sprains & Strains ("840-848"), and Open Wounds ("870-897") and less likely for Contusion w/Intact Skin ("920-924") and Complications & Unspecified Injuries ("958-959"). There does not appear to be an obvious explanation for this last set of results based on IDCs. However, further investigation into the diseases, disorders, and/or symptoms by a healthcare organization may shed light on these interesting findings. The 2835 unknown cases, which have now been classified as 1208 good cases, can potentially bring ($1,020,834-$472,461)=$548,373 in additional revenue (Tables 2 and 7). Finally the models obviously learn to pursue larger debts.

Table 6: The descriptive statistics of the unknown cases classified as good and bad cases by semi-supervised learning

Status	Patient Gender (PG)	Patient Age (PA)	Most frequently occurring Injury Diagnosis Code (IDC). Code Range	%	Dollar Amount of the Claim (DAC)
3 (Unknown) N=2835	Female (N=1305, 46.0%) Male (N=1530, 54.0%)	Mean=28 St. Dev=19 Min=0 Max=100 S_k=0.8	"800"-"829" "840"-"848" "870"-"897" "920"-"924" "958"-"959"	14 19 14 9 20	Mean=$256 St. Dev=$1,091 Min=$1 Max=$30,976 Sum=$727,850 S_k=19.1
(Good) N=1208	Female (N=721, 59.7%) Male (N=487, 40.3%)	Mean=19 St. Dev=17.1 Min=0 Max=100 S_k=1.5	"800"-"829" "840"-"848" "870"-"897" "920"-"924" "958"-"959"	23 22 18 7 3	Mean=$454 St. Dev=$1,628 Min=$1 Max=$30,976 Sum=$548,373 S_k=10.4
(Bad) N=1627	Female (N=584, 35.9%) Male (N=1043, 64.1%)	Mean=33 St. Dev=18.2 Min=0 Max=100 S_k=0.6	"800"-"829" "840"-"848" "870"-"897" "920"-"924" "958"-"959"	7 16 11 12 30	Mean=$110 St. Dev=$240 Min=$1 Max=$5,318 Sum=$179,477 S_k=13.0

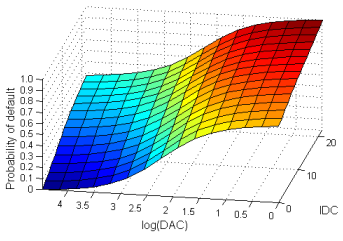
Table 7: The final descriptive statistics after semi-supervised learning for overall cases, good cases, and bad cases.

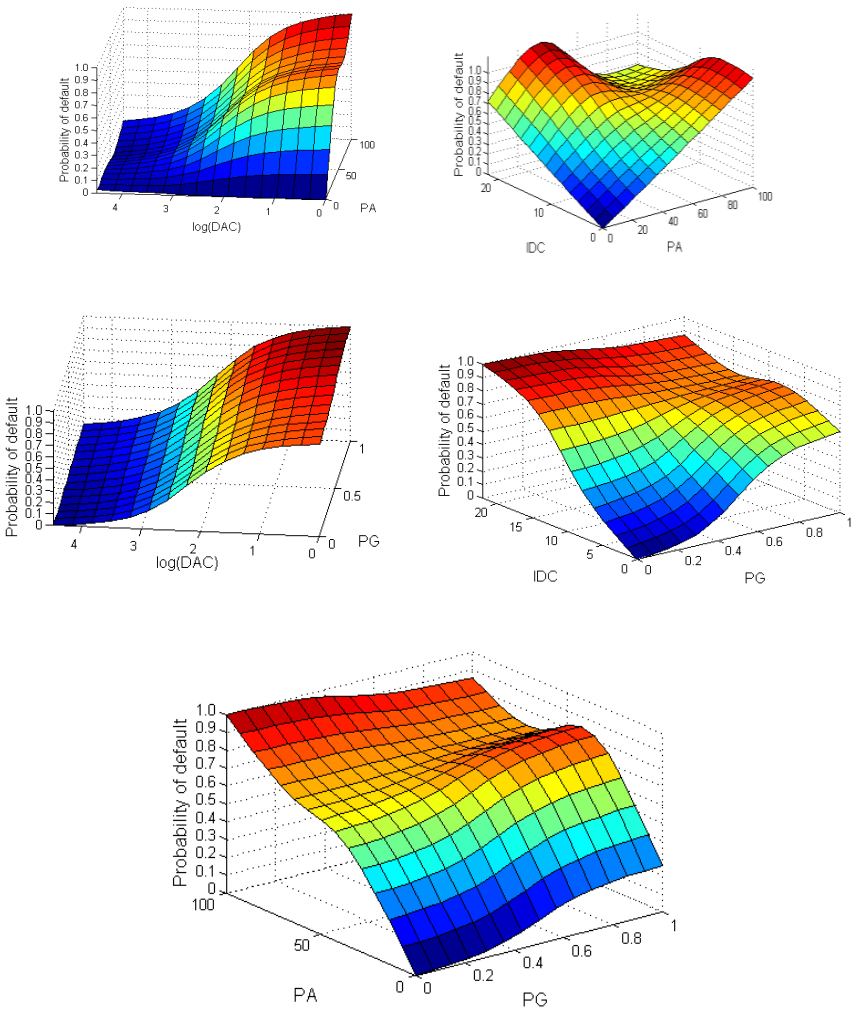
Status	Patient Gender (PG)	Patient Age (PA)	Most frequently occurring Injury Diagnosis Code (IDC). Code Range %	Dollar Amount of the Claim (DAC)
Overall (groups 1,2) N=6117	Female (N=2884, 47.1%) Male (N=3233, 52.9%)	Mean= 30 St. Dev=21 Min=0 Max=100 S_k=0.7	"800"-"829" 16 "840"-"848" 18 "870"-"897" 14 "920"-"924" 9 "958"-"959" 19	Mean=$389 St. Dev=$1,477 Min=$1 Max=$40,508 Sum=$2,381,453 S_k=12.8
1 (Good) N=1657	Female (N=969, 58.5%) Male (N=688, 41.5%)	Mean=23 St. Dev=18.8 Min=0 Max=100 S_k=1.1	"800"-"829" 19 "840"-"848" 22 "870"-"897" 15 "920"-"924" 9 "958"-"959" 6	Mean=$616 St. Dev=$2,281 Min=$1 Max=$40,508 Sum=$1,020,834 S_k=10.4
2 (Bad) N=4460	Female (N=1915, 42.9%) Male (N=2545, 57.1%)	Mean=32.4 St. Dev=21.6 Min=0 Max=100 S_k=0.6	"800"-"829" 14 "840"-"848" 17 "870"-"897" 13 "920"-"924" 9 "958"-"959" 24	Mean=$305 St. Dev=$1,016 Min=$1 Max=$19,568 Sum=$1,360,619 S_k=9.5

Impact of Patient Characteristics on Bad Debt

To generate more insight into nonlinear and complex interactions between the variables and the probability of bad debt default/recovery, we developed six 3-dimensional surfaces generated by ANFIS with SSL (Figure 4). These surfaces have been plotted for the gbell membership function as this function generated stable classification rates. For example, the chart log(DAC) vs. IDC indicates that patients who owe smaller amounts and have injuries represented by IDC codes in the higher-numbered groups such as "920-924" (Contusion w/Intact Skin) and "958-959" (Complications & Unspecified Injuries) are more likely to default. Apparently, the healthcare company has not been interested in pursuing small debts. From the log(DAC) vs. PA plot one can conclude that it is much more difficult to recover bad debt from older patients. The default rate starts to rise sharply around age 30 or so. The remaining charts show that males are more likely to default than females. These six charts generally confirm our observations described earlier.

Figure 4: Six control surfaces.





Analysis through clustered data

We also tested our approach after the data were clustered into seven groups of the most frequently occurring IDCs (Table 2). The control surfaces in Figures 5 through 7 represent the relationships between log(DAC) vs. PA, log(DAC) vs. PG, and PA vs. PG for the data representing the three most frequently occurring IDCs. These are fractures ("800-829"), sprains & strains ("840-848"), and complication & unspecified injuries ("958-959"). As discussed earlier, it appears that it is more likely to recover the bad debt for the first two IDC groups than for the third one. Examination of the control surfaces for these three specific and more finely defined subsets of data by IDCs offers insight unavailable from results obtained from the entire dataset without clustering.

The three control surfaces in Figure 5 represent cases in the IDC group representing Fractures. The control surface of log(DAC) vs. PA indicates that the highest probability of default (the peak) occurs for the middle age patients with relatively low values of DACs. The chart of log(DAC) vs.

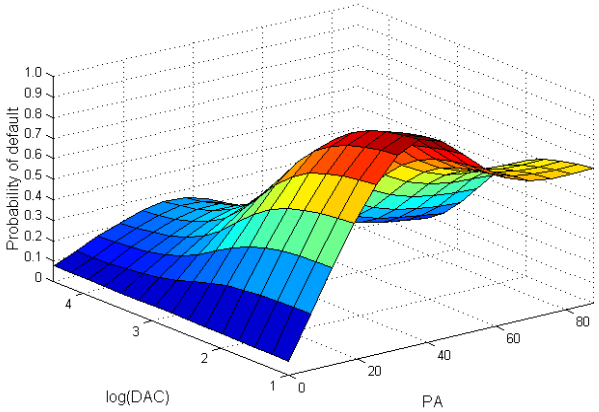
PG is somewhat counterintuitive as it seems that patients with smaller DACs are more likely to default. However, as mentioned earlier, the company chose to pursue larger debts and the models learn to pursue higher debts. Another possible explanation for this counter-intuitive finding is that these bills were for initial diagnosis that reflected major health complications and patients decided to either pursue treatment elsewhere or not to pursue treatment at all. The control surface of PA vs. PG confirms the previous the finding that older (middle-age) male patients are more likely to default.

The three control surfaces in Figure 6 represent Sprains and Strains. For this specific group of injuries, it is evident from the first two charts that patients are more likely to default on smaller debts. Senior patients over 65 years old are less likely to default than younger patients (the first chart), and females are less likely to default than males (the second chart). However, the third chart shows a sharp rise in the probability of default for older females over 65 years old, whereas for males this increase is more moderate.

Figure 7 shows the three control surfaces for Complications & Unspecified Injuries. As described earlier it is harder to recover bad debt for this specific group of patients exhibiting these types of injuries. The first chart shows that regardless of the debt amount the probability of default for younger patients is constant and approximately equals to 0.6. However, for older patients and low amount of debt the probability of default approaches 1.0. Interestingly, older patients with larger debt are less likely to default. The second chart shows similar patterns. Males and females are more likely to default on smaller debts, and for larger debts males are more likely to default.. The third chart shows that regardless the gender the probability of default is very high for young and middle age patients. However, the probability of default is low for old females and very high for older males.

The charts from the clustered data provide additional support for conclusions drawn from the results using the unclustered data as well as offering additional insight at a finer level of analysis.

Figure 5: Three control surfaces for IDC=4 ("800"-"829": Fractures).



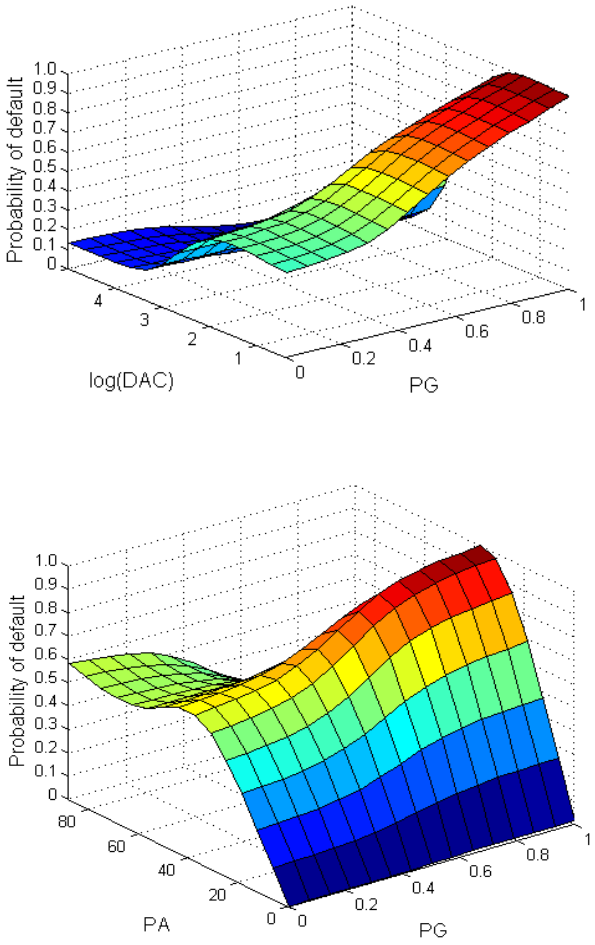
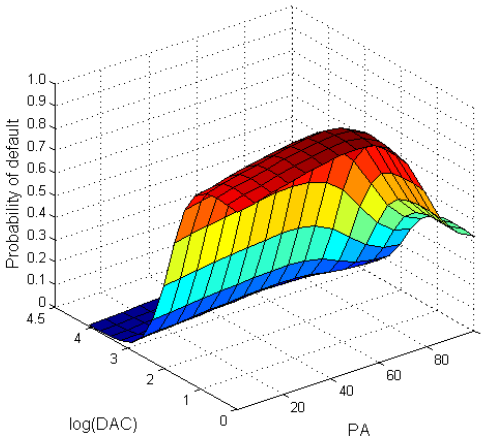


Figure 6: Three control surfaces for IDC=6 ("840-848": Sprains & Strains).



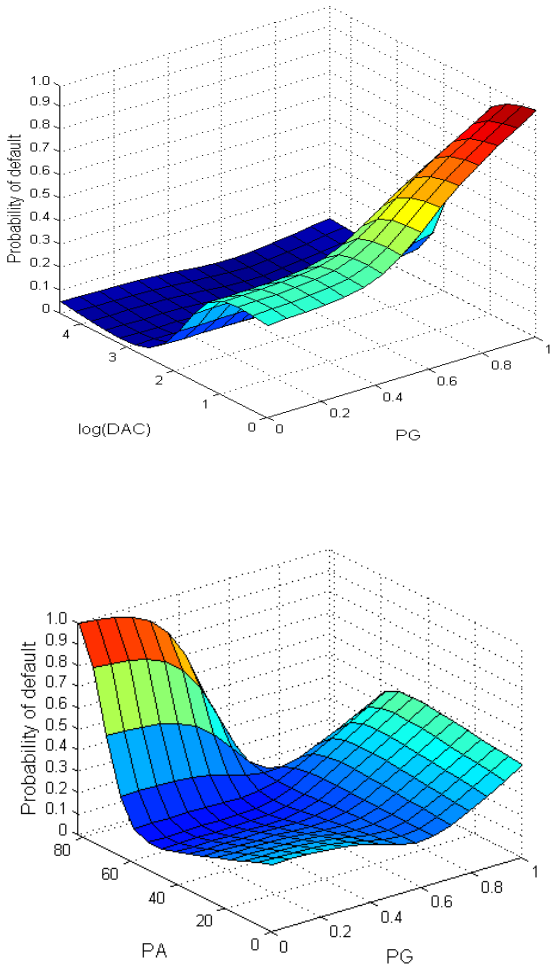
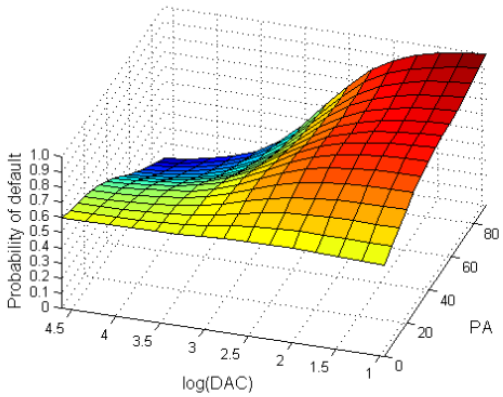
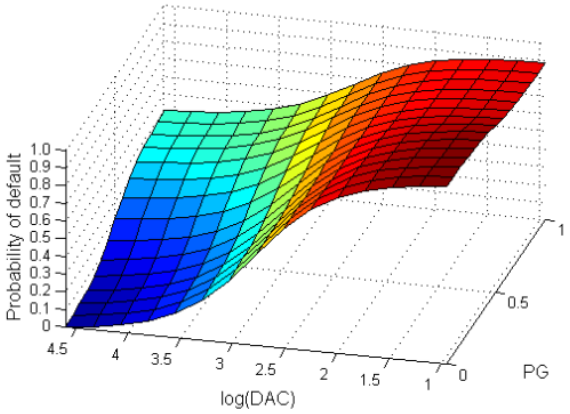
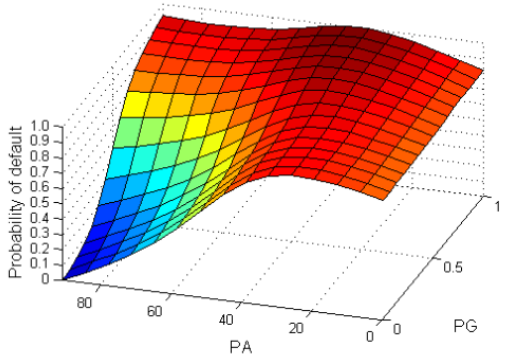


Figure 7: Three control surfaces for IDC=18 ("958"-"959": Complications & Unspecified Injuries).







CONCLUSION

The paper explores the effectiveness of ANFIS with SSL in building classification models using unlabeled data. This study addresses an important class of problems that will pose increasing challenges to today's organizations as they embrace business analytics to stay competitive. Data from a real-life healthcare organization was used in the study. A key characteristic of this data set, a characteristic common to many other contexts other than healthcare, is the presence of a large set of unlabeled data records. Our models generated better classification accuracy rates than ANFIS alone and three other models created and tested in the previous studies. Computer simulation was performed for unclustered data and data clustered by the seven most frequently occurring IDCs. The results indicate that ANFIS with SSL could potentially lead to a higher bad debt recovery rate through classification of unknown cases. Five different models were designed and tested. The results and the approach in this study could potentially help healthcare organizations target specific group of customers to improve their return on debt recovery efforts. The example control surfaces for all data and the data clustered by IDCs reveal interesting relationships between the probability of recovery/default and the three other variables. Though there does not appear to be an obvious interpretation for some interesting results, further

investigation into the diseases, disorders, and/or symptoms by a healthcare organization may shed more light on these interesting findings.

An interesting aspect of the paper shows the ability of the models to classify unknown cases, which are a potential source of revenue recovery. This study shows the potential of computational intelligence and soft computing 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. Though our study was conducted with data from a healthcare organization, the approach can be easily adapted for similar situations where the data exist an unbalance with a large number of unknown cases.

Acknowledgment: This work was partially supported by (1) the Prometeo Project of the Secretariat for Higher Education, Science, Technology and Innovation of the Republic of Ecuador, (2) Anhui Provincial Natural Science Foundation of China (1508085MF114), and (3) Technology Foundation for Selected Overseas Chinese Scholar (2014).

REFERENCES

Albright, E. (2013, 11 July). Obama won't offer relief on bad debt for hospitals. *Pharma & Healthcare*, 1-2.

Bradley, P., & Kaplan, J. (February 2010). Turning hospital data into dollars. *Healthcare Financial Management*, i-v.

Buczko, W. (1994). Factors affecting charity care and bad debt charges in Washington hospitals. *Hospitals and Health Services Administration*, 39(2), 179-191.

Chapelle, O., Scholkopf, B., & Zien, A. (Eds.). (2006). *Semi-Supervised Learning*. MIT Press, Cambridge, MA.

Drummond, C., & Holte, R. C. (2003). C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. In *Workshop on learning from imbalanced datasets II*, 11.

Fawcett, T. (2006). An Introduction to ROC Analysis. *Pattern Recognition Letters*, 27 (8), 861–874.

Galloro, V. (October 2003). Bad news on bad debt. *Modern Healthcare*, 33(43), 8.

Haibo, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. *Decision Support Systems*, 64, 130–141.

IBM. (2013, 12 June). Helping hospitals capture more revenue. http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=YTC03052USEN&appname=wwwsearch.

Jang, J. S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23, 665-685.

Kutscher, B. (2013, 17 August). Targeting bad debt: Hospitals getting proactive on billing. *Modern Healthcare*, 1-3.

Lagnado, L. (2003, 30 October). Hospitals try extreme measures to collect their overdue debts. *The Wall Street Journal*.

McAfee, A., & Brynjolfsson, E. (October, 2012). Big data: the management revolution. *Harvard Business Review*, 3-9.

Murgia, G., & Sbrilli, S. (2012). A decision support system for scoring distressed debts and planning their collection. *In Methods for Decision Making in Uncertain Environment*: *Proceedings of the XVII SIGEF Congress*, J. Gil-Aluja and Antonio Terceño (Eds), World Scientific Publishing Co., Singapore, 6, 69-89.

National Health Expenditures 2013 Highlights. (n.d.). http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/highlights.pdf.

Pell, M. B. (2011, June 6). Patients in arrears face collectors. *The Atlanta Journal-Constitution*.

Pesce, J. (2003). Stanching hospitals' financial hemorrhage with information technology. *Health Management Technology*, 24(8), 12.

Shi, D., Zurada, J., & Guan, J. (2014). A Neuro-fuzzy approach to bad debt recovery in healthcare. *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS'47)*. (R. Sprague, Ed.), IEEE Computer Society Press, 2888-2897.

Sugeno, M., & Kang, G. T. (1988). Structure identification of fuzzy models. *Fuzzy Sets and Systems*, 28, 15-33.

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

Veletsos, A. (2003). Getting to the bottom of hospital finance. *Health Management Technology*, 24(8), 30.

Zollinger, T. W., Sawyell, R. M. Jr., Chu, D. K. W., Ziegert, A., Woods, J. R., & LaBov, D. (1991). A determination of institutional and patient factors affecting uncompensated hospital care. *Hospital & Health Services Administration*, Chicago, 36(2), 243-256.

Zurada, J., & Lonial, S. (2004), Application of Data Mining Methods for Bad Debt Recovery in the Healthcare Industry. *The Proceedings of the 6th International Baltic Conference on Databases and Information Systems, Baltic DB&IS'2004,* J. Barzdins (Ed.), 672, 207-217.

Zurada, J., & Lonial S. (2005, Spring). Comparison of the performance of several data mining methods for bad debt recovery in the healthcare industry. *Journal of Applied Business Research*, 21(2), 37-54.