AN OPTIMUM CREDIT LENDING DECISION

Monireh Sadat Mirtalaei - Peiman Pazhoheshfar - Emad Roghanian - Morteza Saberi - Zahra Saberi

Abstract

Nowadays, the concept of Trust can be found in a number of different fields including sociology, business and computing. Trust is playing an important role in the e-business and virtual environment. Exchanging goods and services across an interactive digital network, forces the companies to build their transactions based on the level of trust of their customers. One of the recent services which are widely provided by internet is credit score computation. Till now, almost all of the studies in this field are trying to improve the accuracy rate of the proposed algorithms and this is the first times that trust concept is used in credit scoring domain. In this paper, a trust- based approach for credit scoring is proposed which lets the financial institutions granting their credit, based on the level of trust of the customers. The mathematical formulas are proposed to link the credit scoring domain and trust concept. At the classification stage, an Artificial Neural Network (ANN) - based model is used. To show the applicability and superiority of the proposed algorithm, it is performed on a credit card dataset obtained from UCI repository.

Key words: Trust; Business Intelligence; Credit Scoring; Artificial Neural Network

JEL Code: D81,G32,H81.

Introduction

With the rapid growth in credit industry and the management of large loan portfolios, credit scoring models have been extensively used for the credit admission evaluation. The credit scoring models are developed to classify loan customers either as a good credit group (accepted) or a bad credit group (rejected) with their related characteristics such as age, income and marital status or based on the data of the previous accepted and rejected applicants (Chen and Huang, 2003; Evans and Krueger,2011).

Many different credit scoring models have been developed by the banks and researchers in order to solve the classification problems, such as Linear Discriminant Analysis (LDA), Logistic Regression (LR), Multivariate Adaptive Regression Splines (MARS), Classification and Regression Tree (CART), Case Based Reasoning (CBR), and Artificial Neural Networks (ANNs) (Chuang and Lin, 2009; Fan et al, 2011).

From the extensive survey of ANN applications in business, it indicates that ANN shows promise in various areas where nonlinear relationships are believed to exist within the datasets, and traditional statistical approaches are deficient. In credit prediction, the nonlinear features of ANNs make them a potential alternative to traditional parametric (e.g. LDA and LRA) and nonparametric (e.g. KNN and decision tree) methods (Chen and Huang, 2003; Goo & Huang, 2008).

Today, requesting for a loan or credit would be as hard as clicking a tab on your keyboard. The traditional long process of granting credit to the customers has changed its place by just filling an applicant form on the related websites. Therefore, due to the intense competition of credit card issued by banks, more and more people can easily apply a credit card without carefully examining of their credit by the banks. This reckless expansion policy has increased the delinquency rate in the banks (Chuang and Lin, 2009; Ferrary, 2003).

Although the existing proposed models for credit assessment can help the financial institutions to predict the future credit status of a customer, but none of them talk about the trust level of the clients. In the world of interactive digital network and virtual environment, trust plays a really vital role in handling the transactions. By categorizing the customers to different groups and based on their level of trustworthiness, the managers would be able to apply a proper policy in order to keep or ignore the new clients.

The paper is organized as follows. Section 1 provides an introduction to the concept of trust. Section 2 is dedicated to the proposed algorithm. Application of the methodology on a real case study is going to be carried out during Section 3. The final section of the paper offers conclusions.

1 Concept of Trust

Today, trustworthiness is a foundation for the success of e-business. Currently trust is garnering the attention of those who employ websites to make available information, services or products to others (Corritore et al, 2003; Charness et al, 2011; Wang & Gordon, 2007).

We use the proposed concept in Hussain et al, (2004) as the following:

‘Trust is defined as the belief that the Trusting Agent has in the Trusted Agent’s willingness and capability to deliver a mutually agreed service in a given context and in a given Timeslot.’

According to the Hussain et al, (2006), 7-level trustworthiness is defined with both numeric and non-numeric measures for the evaluation of the trustworthiness. Each level is described and the semantics (linguistic definitions), postulates and visual representations are discussed. The linguistic definition of each level provides the meaning of the confidence or the trust that the Trusting Agent has in the Trusted Agent (See Table 1).

Tab. 1: Seven levels of trustworthiness and corresponding semantics, linguistic definitions and approximate user defined interval ranges

Trustworthiness Scale (Ordinal Scale)	**Semantics** (Linguistic Definitions)	**Percentage Intervals** (User defined)	**Trustworthiness Value** (User defined)
Level -1	Unknown Agent	N/A	$x = -1$
Level 0	Very Untrustworthy	0-19%	$x = 0$
Level 1	Untrustworthy	20-39%	$0 < x \leq 1$
Level 2	Partially Trustworthy	40-59%	$1 < x \leq 2$
Level 3	Largely Trustworthy	60-79%	$2 < x \leq 3$
Level 4	Trustworthy	80-90%	$3 < x \leq 4$
Level 5	Very Trustworthy	90-100%	$4 < x \leq 5$

1.1 Level -1: Unknown trustworthiness

Level -1 is defined as the trustworthiness level assigned to a Trusted Agent by the Trusting Agent, when the Trusting Agent is unable to ascertain an estimate or carry out a measurement of trust in the Trusted Agent, in a given context and timeslot and for a given initiation of the association.

1.2 Level 0: Very Untrustworthy

Level 0 is defined as the trustworthiness level that the Trusting Agent assigns to the Trusted Agent when the Trusting Agent cannot trust the Trusted Agent at all, in the specific context and in that particular timeslot.

1.3 Level 2: Partially Trustworthy

Level 2 is defined as the trustworthiness level that the Trusting Agent assigns to the Trusted Agent when the Trusting Agent has around 50% confidence in the Trusted Agent in a given context, timeslot and initiation of the association.

1.4 Level 3: Largely Trustworthy

Level 3 is defined as the trustworthiness level that the Trusting Agent assigns to the Trusted Agent when the Trusting Agent has basically positive trust or a confidence of around 70% in the Trusted Agent in a given context and timeslot and for the given initiation of the association.

1.5 Level 4: Trustworthy

Level 4 is defined as the trustworthiness level that the Trusting Agent assigns to the Trusted Agent when the Trusting Agent has a confidence of over 80% in the Trusted Agent in a given context and timeslot and for a given initiation of the association. The Trusted Agent satisfying this criterion is also referred to as trustworthy and will, in an interaction, behave almost exactly as the Trusting Agent expected.

1.6 Level 5: Very Trustworthy

Level 5 is defined as the trustworthiness level that a Trusting Agent assigns to a Trusted Agent that is completely trustworthy and the assignment of level 5 indicates a very positive trust or a confidence of over 90% from the Trusting Agent to the Trusted Agent in a given context, timeslot and for a given initiation of the trust relationship. A Very Trustworthy agent is also referred to as completely trustworthy, and in any given situation, the Trusted Agent is certain to behave exactly as the Trusting Agent expects.

2 Proposed Algorithm

As we mentioned before, the main reason to introduce the proposed algorithm is using the concept of trust in credit scoring and easing the process of policy making for granting the credit to the customers. The main structure of the intelligent approach is explained following.

2.1 Step 1: Defining the Data

The Selection of observation set is done as the first step of proposed methodology.

2.2 Step 2: Normalizing the data

The database may contain both large values and really small ones. For removing the possible dominating effect of large values, the data set is normalized according to the following formula (1). As the result, all the values lie between 0 and 1.

$$X_{normalized} = \frac{X - Min(X)}{Max(X) - Min(X)} \qquad (1)$$

2.3 Step 3 : Dividing the data set

According to the proposed algorithm, after selecting the most effective features, the data set is divided into two subsets: Training set and Test set. The training set is used for computing the gradient and updating the network weights and biases. The second subset is the test set. The error on the test set is monitored during the training process. Here, the available data is split into a training data set, containing 90% of the data, and a validation data set, containing 10% of the data.

2.4 Step 4: Constructing the best applicable ANN

The plausible architecture of ANN models are constructed at first. Then the ANN models are run by training data set and tested by validation data set. Among the different networks, the feed forward neural networks or multi-layer perceptron (MLP) are the most commonly used in engineering. In this step, the models are run. Transfer function, learning algorithm and number of neurons are parameters that are considered in ANN architect constructing (100 is considered as a maximum).Table 2 shows the list of different learning methods, used in this paper. Also, log-sigmoid and hyperbolic tangent sigmoid are used as transfer functions. Totally, 1800 ANNs are run and trained in this step.

In order to assess the performance of ANNs, their accuracy are evaluated respect to mean absolute percentage error (MAPE) explained in (2).

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{x_{ANN} - x_{original}}{x_{original}} \right|}{n} \qquad (2)$$

2.5 Step 5: Labeling the results by different trust levels

The final result of the selected ANN would be scores which are associated to the credit status of the test group customers. In order to determine the trust level of these customers, the output of ANN must convert to suitable trust level. For this purpose, Formula (3) has been proposed:

$$\begin{aligned} &B(i) = Min\,C + i * \left(\frac{S}{T} \right) \quad for\ i = 0,1,\dots,5 \\ &T = Max\,T - Min\,T \qquad (3) \\ &S = Max\,C - Min\,C \end{aligned}$$

That:

Max T=Maximum value of the Trust levels

Min T= Minimum value of the Trust levels

Max C= Maximum value of the computed scores

Min C= Min value of the computed scores

The following rule base is used for converting ANN output to trust level:

IF *ANN output =B(0)* **THEN** *Trust Level*=Very Untrustworthy

IF *B(0)<ANN output≤B(1)* **THEN** *Trust Level*= Untrustworthy;

IF *B(1)<ANN output≤B(2)* **THEN** *Trust Level*=Partially Trustworthy;

IF *B(2)<ANN output≤B(3)* **THEN** *Trust Level*=Largely Trustworthy;

IF *B(3)<ANN output≤B(4)* **THEN** *Trust Level*= Trustworthy;

IF *B(4)<ANN output≤B(5)* **THEN** *Trust Level*= Very Trustworthy;

3 Case Study

The proposed methodology is applied and explained to an actual data set.

3.1 Step 1

In this paper, we are going to use the credit card dataset obtained from UCI repository. These samples are associated to German Bank. The database includes 1000 observations which are described by 8 input features. Here, we are going to classify these customers based on their trust level and, by the use of the cited features.

3.2 Step 2

According to the proposed formula in the previous section, we normalize the available data set.

3.3 Step 3

Splitting the available data into a training data set, containing 90% of the observations, and a validation data set, containing 10% of the observations is done in this step.

3.4 Step 4

All of the ANN models are run by testing the data and then the efficiency of each model is evaluated by MAPE. 1800 ANNs are categorized to 18 groups. Table 2 shows the architect of 18 groups and the minimum MAPE of each group. Examination of Table 2 shows that ANN with Levenberg-Marquardt back propagation as a learning algorithm and Hyperbolic tangent sigmoid as first transfer function learning and 15 neurons in its hidden layer is the preferred ANN. The lowest MAPE value is 15 %.

3.5 Step 5

Step 5: For this step, the Formula (2) is going to be applied for determining the amount of new trust ranges. As you can see in this formula, we need to compute the lowest achievable credit score and also the highest possible one. In order to get these scores, we apply 20 different test sets to the selected net, and then the results are accumulated to find the minimum and maximum values. Another point which should be noted here, is that the trustworthiness scale of (-1) is not applicable in the case of credit scoring. A score is assigned to any new

Tab. 2: MAPE value for ANN models in 18 categories

MLP Model number	Learning method	Number of Neurons in first hidden layer	First transfer function	Second transfer function	MAPE
1	LM	64,76,93	Log-sigmoid	Linear	0.18
2	SCG	35,89	Log-sigmoid	Linear	0.175
3	RP	1	Log-sigmoid	Linear	0.175
4	OSS	4	Log-sigmoid	Linear	0.16
5	GDX	74	Log-sigmoid	Linear	0.19
6	GDA	12	Log-sigmoid	Linear	0.185
7	CGB	67	Log-sigmoid	Linear	0.155
8	BR	37,58	Log-sigmoid	Linear	0.17
9	BFG	5	Log-sigmoid	Linear	0.165
10	LM	15	Tan- sigmoid	Linear	0.15
11	SCG	45	Tan- sigmoid	Linear	0.17
12	RP	2	Tan- sigmoid	Linear	0.16
13	OSS	31	Tan- sigmoid	Linear	0.165
14	GDX	23	Tan- sigmoid	Linear	0.165
15	GDA	15,77	Tan- sigmoid	Linear	0.185
16	CGB	61	Tan- sigmoid	Linear	0.165
17	BR	31,47	Tan- sigmoid	Linear	0.17
18	BFG	2	Tan- sigmoid	Linear	0.16

customer based on the information in their applicant form. So the status of those customers will be predicted and they do not remain unknown for the company. Thus, our model is composed of 6 levels.

$$\begin{aligned} MaxT &= 5 \\ MinT &= 0 \\ MaxC &= 2.12 \\ MinC &= 1.13 \\ T &= 5 - 0 = 5 \\ S &= 2.12 - 1.13 = 0.99 \approx 1 \end{aligned}$$

$$B(i) = 1.13 + i * \left(\frac{1}{5}\right)$$

$$\begin{aligned} B(0) &= 1.13 \\ B(1) &= 1.13 + (1)(0.2) = 1.33 \\ B(2) &= 1.13 + (2)(0.2) = 1.53 \\ B(3) &= 1.13 + (3)(0.2) = 1.73 \\ B(4) &= 1.13 + (4)(0.2) = 1.93 \\ B(5) &= 1.13 + (5)(0.2) = 2.13 \end{aligned}$$

The definition of levels is explained in the Table 3:

Tab. 3: Trust – based credit scores rating

Trust – Based Score Scale	Semantic	Score Value
Level 0	Very Weak (VW)	x = 1.13
Level 1	Weak (W)	1.13 < x ≤ 1.33
Level 2	Medium (M)	1.33 < x ≤ 1.53
Level 3	Medium High (MH)	1.53 < x ≤ 1.73
Level 4	High (H)	1.73 < x ≤ 1.93
Level 5	Very High (VH)	1.93 < x ≤ 2.13

As the final step, we examine the test set and label the results according to the Table 3. Here we illustrate the results of one of these test sets. Table 4, shows the different categorizations which are done based on our new trust levels.

Tab. 4: Results of applying proposed algorithm on real data

	VH	H	MH	M	Other
Good	18	29	9	4	8
Percentage	26.5	42.6	13.2	5.9	11.8
	VW	W	M	MH	Other
Bad	0	8	14	4	6
Percentage	0	25	43.8	12.5	18.8

As it is shown in Table 4, about 26.5% of the real good customers are labeled as **Very High** in case of trust levels. Therefore, the decision managers can apply the encouraging policies for this group. Because of high level of trustworthiness, their information would be reliable and the institution can make profit by dealing with them. The overall policy of a customer- oriented company should be keeping this group of customers satisfied.

42.6 is the percentage of good customers which are calssified as **High** in case of their trustwothiness value. The company can use policies which motivate these customers to increase their scores up to the best level (Very High).

Annually feedback on any improvement in their scores can be helpful in this field.

The next group, **Medium High** , can be named as a frontier class. Customers of this group can be considered either as a good credit or the bad ones. A more explanation will be the quasi-equal percentage of MH group in both mentioned tables. Of course, by applying true encouragement policies managers would be able to push these customers forward. They should remind them different advantages of being categorized as very high and high trust values customers.

The **Medium** group are known as partially untrustworthy customers. Their information can not be reliable all the times. In different timeslots, they either act as a good customers and in more cases as the bad ones. To apply warning policies for this group of customers may help them to improve their scores.

The financial institutions can not rely on the **Weak** group of customers. Their financial situation is not stable and that may cause the institution to be in loss. Using punishment policies (e.g. increase the installment rate or put limitation on the amount of granting credit) will lead these customers to act better on their financial resposibilities.

The **Very Weak** class or the least score, is related to the customers which are not reliable at all. Their documents are full of delay or non-payment regarding their financial obligations. These group of custmors wil be certain loss for the company. Therefore, it would be better for the managers to ignore them.

We have another column in our tables which is called Other. The numbers lie down on these columns, show the percentage of neural network error in categorizing the customers. Of course, the lower is the better.

Conclusion

In this paper, a trust-based approach for credit scoring is proposed. This is the first study that proposes an integrated algorithm capable of predicting the credit status of customers based on their trust levels. This model will be very applicable in case of e-business and virtual environment where, the concept of trust plays an important role in dealings. For showing the capability of the proposed algorithm in categorizing the customers, a free of noise real data set is employed. Due to their proved predicting ability, at the classification stage, an ANN- based model is used. More than 1800 ANNs are trained by 90% of the available data and then tested through remaining 10%. In order to assess the performance of ANNs, their accuracy was evaluated respect to mean absolute percentage error (MAPE). The ANN with the lowest MAPE value would be the best. In the next step, a formula has proposed to calculate the new ranges of trust levels according to the computed credit scores. Finally, the customers are labeled based on their trust values. The 6 levels of categorizations will help the decision managers choose a proper policy for different type of customers.

References

Charness, G., Du, N. & Yang, C. (2011). Trust and trustworthiness reputations in an investment game, Games and Economic Behavior, 72(2), 361-375.

Chen, M. C., & Huang, S. H. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. Expert Systems with Applications, 24(4), 433–441.

Chuang, C. & Lin, R. (2009). Constructing a reassigning credit scoring model. Expert Systems with Applications, 36 (2), 1685-1694.

Corritore, C., Kracher, B. & Wiedenbeck, S. (2003). Online trust: Concepts, evolving themes, a model. Int. J. Human-Computer Studies 58, 737-738.

Evans, A. M. & Krueger J. I. (2011). Elements of trust: Risk and perspective-taking, Journal of Experimental Social Psychology, 47 (1), 171-177.

Fan, Z., Suo, W., Feng, B. & Liu, Y. (2011). Trust estimation in a virtual team: A decision support method, Expert Systems with Applications, Volume, 38(8), 10240-10251.

Ferrary, M. (2003). Trust and social capital in the regulation of lending activities, Journal of Socio-Economics, 31(6), 673-699.

Goo, J. & Huang, C.D. (2008). Facilitating relational governance through service level agreements in IT outsourcing: An application of the commitment–trust theory, Decision Support Systems, 46(1), 216-232.

Hussain, F. K., Chang, E. & Dillon, T. (2004). Trustworthiness and CCCI metrics in P2P communication. International journal of Computer Systems Science & Engineering, 173-175.

Hussain, F. K., Chang, E. & Dillon, T. (2006). Trust and reputation for service-oriented environments: Technologies for building business intelligence and consumer confidence. West Sussex, England: John Wiley and Sons.

Kursunoglu, A. (2009). An investigation of organizational trust level of teachers according to some variables, Procedia - Social and Behavioral Sciences, Volume 1(1), 915-920.

Wang, L. & Gordon, P. (2007). Trust and institutions: A multilevel analysis, Journal of Socio-Economics, 40(5), 583-593.

Contact

Monireh Sadat Mirtalaei

Department of Industrial Engineering, University of Tafresh, Tafresh, Iran

m.mirtalaie@gmail.com

Peiman Pazhoheshfar

Young Researchers Club, Tafresh Branch, Islamic Azad University, Tafresh, Iran

p.pazhohesh@gmail.com

Emad Roghanian

Department of Industrial Engineering, K.N Toosi University of Technology, Tehran, Iran

E_roghanian@kntu.ac.ir

Morteza Saberi

Institute for Digital Ecosystems & Business Intelligence, Curtin University of Technology, Perth, Australia

M.Saberi.ie@gmail.com

Zahra Saberi

Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran

Zahra.saberi.aut@gmail.com