BAYESIAN CREDIT SCORING MODEL WITH INTEGRATION OF EXPERT KNOWLEDGE AND CUSTOMER DATA

Taner Bilgiç
Bogazici University

International Conference
24th Mini EURO Conference
"Continuous Optimization and Information-Based Technologies in the Financial Sector"
(MEC EurOPT 2010)
June 23–26, 2010, Izmir, TURKEY

ISBN 978-9955-28-598-4
R. Kasımbeyli, C. Dinçer, S. Özpeynirci and L. Sakalauskas (Eds.): MEC EurOPT 2010
Selected papers. Vilnius, 2010, pp. 324–329

BAYESIAN CREDIT SCORING MODEL WITH INTEGRATION OF EXPERT KNOWLEDGE AND CUSTOMER DATA

Işık Biçer[1], Deniz Seviş[2], Taner Bilgiç[1]

[1]*Department of Industrial Engineering, Boğaziçi University, Turkey*
[2]*Graduate Program of Systems and Control Engineering, Boğaziçi University, Turkey*

Abstract: Bayesian classification method is one of the effective classification methods in credit scoring applications. Application of this method to credit scoring provides several advantages, which are suggested in the literature. In this research, a Bayesian credit scoring model is applied for classification of personal loan customers. The contribution of this paper is two-fold. First, we apply a Bayesian classification model to credit scoring of personal loan customers with integration of knowledge of credit experts and personal loan data. The dependency relations between the variables are constructed based on the knowledge of credit experts, and the parameters are estimated by Expectation-Maximization (EM) algorithm with real personal loan data of customers obtained from a Turkish bank. Second, we assess the effectiveness of the Bayesian credit scoring model based on additional personal loan data of the bank customers. The results show that the Bayesian credit scoring model is more effective than the logistic regression credit scoring model which has been currently used in the bank.

Keywords: bayesian classification; credit scoring; expectation-maximization algorithm.

1. Introduction

Credit is the loan that can only be granted by authorized financial institutions or banks to the customers who apply for credit. After a credit application is taken by a creditor, an assessment process is performed in order to decide whether to approve or reject granting credit to the applicant depending on the registered customer information (Thomas, 2000). This assessment process is known as *credit scoring* in finance literature, which is a classification method aiming to distinguish the desired customers who will fully repay from defaulters. Most recently, creditors have been more interested in scoring their credit customers and charging lower interest rates for the customers with good scores; or charging higher rates or not approving the credit application for the customers with bad scores (Thomas, 2000). The importance of credit scoring has been increasing in recent years depending on the profitability and regulatory issues. According to Basel Accords, the risk weight of default loans is % 100, which means that banks should reserve more capital for these loans instead of using the capital and making profit (Basel Manual, 2006).

The interest in credit scoring methods increased with the advent of the credit cards. The increasing number of the credit card applicants makes the creditors to automate the lending decision that, in turn, results in the use of mathematical methods for credit scoring purposes (Thomas, 2000). There have been several methods applied to credit scoring of customers in literature such as discriminate analysis, linear regression, logistic regression, non-parametric smoothing methods (i.e. Generalized Additive Models (GAM)), mathematical programming, genetic algorithms, neural networks, graphical models (i.e. Bayesian networks), support vector machines, and others (Hand *et al.*, 1997a; Hand *et al.*, 1997b; Chen *et al.*, 2003; Huang *et al.*, in press). Among these methods, generalized additive models are effective in classification of customers with large number of parameters. This method is more powerful in detecting the nonlinear patterns; however, it may cause over-fitting problems due to complex smoothing terms (Liu *et al.*, 2007). Multivariate adaptive regression spline (MARS) method is another non-parametric method that is effective in fitting nonlinear multivariate functions in high dimensional problems allowing both additive and interactive effects of the predictors to determine the response variable (Taylan *et al.*, 2007). The main focus of these methods is to estimate a dependent (response) variable from a set of independent variables (predictors). On the other hand, graphical models are different from these methods since it allows the construction of dependency relations within the predictors (Hand *et al.*, 1997b). In general, none of these methods is the best alternative and that is the reason why credit scoring has always been based on a pragmatic approach to the credit granting problem (Thomas, 2000). As a result, the performance of any credit scoring model has to be assessed based on the characteristics of the classification (Hand *et al.*, 1997a).

Bayesian networks have been applied to credit scoring models for several reasons. First, Bayesian networks are effective in understanding the relations between the variables of interest. The number of variables in a credit scoring model can be large depending on the information of applicants, and Bayesian networks help to eliminate some variables which do not influence the credit scores of the customers (Hand *et al.*, 1997b). Second, some decision nodes can be added to Bayesian networks in sequence and the optimum set of decisions maximizing the utility of the decision maker can be reached (Chang *et al.*, 2000). Third, Bayesian networks allow integration of knowledge and data in a manner that once the network is constructed manually based on the initial knowledge, the conditional probabilities of the variables can be calculated depending on the data of the variables. To the best of our knowledge, the integration of expert knowledge and customer data has not been yet applied to credit scoring modeling through Bayesian classification methods. In this research, we first apply Bayesian classification method with integration of expert knowledge from credit experts and customer data from a Turkish bank to credit scoring modeling. In our application, the dependency relations within the variables of Bayesian network of credit scoring model are determined by credit experts. Then, the parameters (conditional probabilities) are estimated via Expectation-Maximization algorithm. We also assess the effectiveness of this credit scoring model based on additional data obtained from the bank. The results show that Bayesian credit scoring model is more effective than the logistic regression credit scoring model which is currently being used in the bank.

The rest of this paper is organized as follows. In Section 2, our Bayesian credit scoring model is described in detail. Section 3 covers reject inference and assessment results. In the last section, we draw conclusions and make suggestions for further research.

2. Description of Model

Decision of the model variables which are included in Bayesian credit scoring model is one of the important parts while developing the structure of the network. As a starting point, 27 variables are selected which have been (mostly) used in the current logistic regression credit scoring model of the bank. Separately, one additional variable (Credit rating variable) representing the credit scores of the customers is added to the Bayesian network. Then the variables that are not in the *Markov blanket* of the credit rating node are eliminated, since the credit rating node is conditionally independent of the rest of the network once the nodes in its Markov blanket (which consists of the parent nodes, the child nodes, and the other parents of the child nodes) are given (Chang *et al.*, 2000). As a result, the variables of the Bayesian network are selected as age, hometown, city of residence, monthly income, fixed expense, credit amount, marital status, number of children, house owning status, cellular phone operator, education level and credit rating as performance node based on knowledge and experience of credit experts. The structure of the network is shown in Fig. 1.

As shown in Fig. 1, some relations between the variables are easy to understand. For instance, age must have a direct influence on the monthly income, since income of a customer is expected to increase with experience in a profession. On the other hand, it is not so easy to find relations between some other variables in the network. At first glance, one cannot find a relationship between cellular phone number and credit rating. However, cellular phones have been widely used over 10 years, and there are 3 operators providing service for this purpose. The customers whose cellular phone numbers start with 0505, 0542, 0532, 0535, 0533 should have good payment history for at least three years. Then, the customers are expected to repay their credit debts fully with higher probabilities.

Associated with the credit rating variable, customers are categorized into 2 groups as good or bad. If a personal loan of a customer goes into default then the customer is classified as bad, otherwise the customer is classified as good (Thomas *et al.*, 2002). One of the important parts for classification of customers as good or bad is the determination of the suitable time horizon. This time period covers the duration starting from granting credit for a customer to the time when the customer is classified as good or bad. It is shown in Thomas (2000) that the optimal time period is between 12 to 18 months and this time period should not exceed 24 months. Our sample data consists of personal loan data of 5024 customers whose credits are granted in 2007. Among these customers, the ones whose credits go into default before 2009 are classified as bad, and others are classified as good. Therefore, time horizon for classification of customers is between 12 to 24 months depending on the beginning time in 2007. There may be some cases

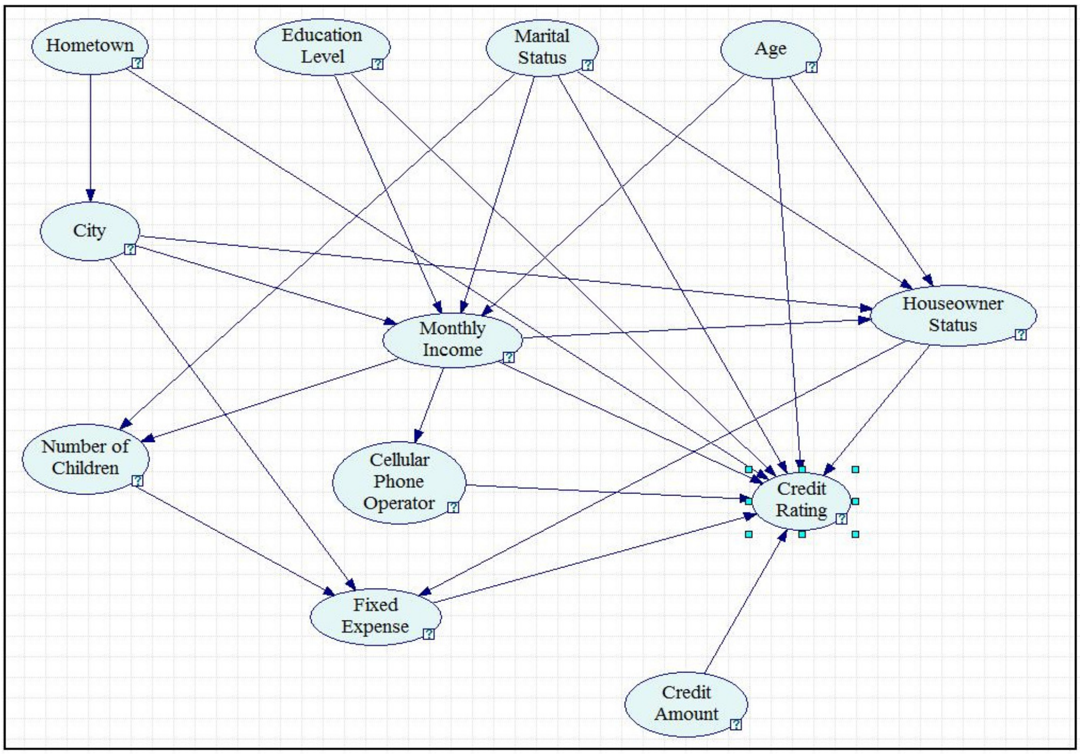

Fig. 1. The Network of Bayesian Credit Scoring Model

that a loan goes default after the first day of 2009. In such cases, the customers are also classified as good, since the profit made during the time horizon may cover the loss of the bank (Thomas, 2000).

After construction of the network, the parameters of the variables are estimated via Expectation-Maximization (EM) algorithm. EM algorithm is a general method of finding the maximum-likelihood estimate of the parameters from a given data set. In EM applications, the expectations of the log-likelihood functions of the parameters are first formulated at each observation. Then, the parameters which maximize the log-likelihood functions are calculated (Bilmes, 1998; Alpaydin, 2004; Mitchell, 1997). In our case, we use *complete data log-likelihood estimates* since credits are not granted unless complete data of customers is not obtained. At the end, the maximum likelihood estimates of the parameters are found by EM algorithm. An application of EM algorithm to estimation of parameters in Bayesian networks is illustrated in Example 1.

Example 1

Suppose that there exist n observations in a data set for the Bayesian network shown below, and for each variable, there are two states as a_1, a_2; b_1, b_2; c_1, c_2.

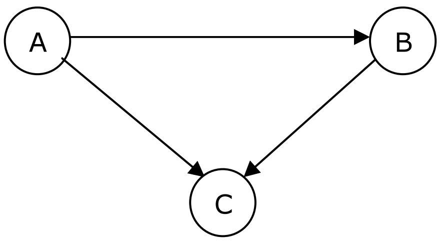

Fig. 1

It is observed that $A = a_1$, $B = b_1$ and $C = c_1$ for the r^{th} observation, where $r < n$. Then, $P(c_1 \mid a_1, b_1)$ is computed as:

$$P(c_1 \mid a_1, b_1) = \frac{P(c_1, a_1, b_1)}{P(a_1, b_1)}. \tag{1}$$

Let the starting and ending probabilities of $P(c_1, a_1, b_1)$ be denoted by θ_0 and θ_1, respectively. Then, the log-likelihood function under binomial distribution is formulated as:

$$L(\theta_1 \mid \theta_0, r) = \ln(Const.) + \left(\theta_0 \cdot (r-1) + 1\right) \cdot \ln(\theta_1) + (1-\theta_0) \cdot (r-1) \cdot \ln(1-\theta_1). \tag{2}$$

The maximum likelihood estimate of θ_1 is then found as shown in Equation 3.

$$\hat{\theta}_1 = \frac{\left(\theta_0 \cdot (r-1) + 1\right)}{r}. \tag{3}$$

Similarly, $P(a_1, b_1)$ can be calculated from Equation 2 and Equation 3, and $P(c_1 \mid a_1, b_1)$ can be obtained from Equation 1.

3. Assessment of Bayesian Credit Scoring

One of the major problems related to the assessment of credit scoring models is unavailability of complete data of all applications. Once a credit application is declined it is then impossible to know the good/bad status of the declined applicant. This causes a reject bias when any assessment is done. The easiest way to cope with the reject bias is to approve all applications. However, it is practically impossible since the risk associated with this process is very high. Ash and Meester (2002) proposes obtaining supplemental bureau data for unobserved customers. If any credit application of rejected customers is approved by other banks, the payment history of the rejected customers can be gathered by credit bureaus. On the other hand, banks usually use the similar scoring techniques and it is hardly possible that rejected applicants can be approved for their other credit applications by other banks. Apart from being hard to find credit history of rejected customers from credit bureaus, it is also costly that members of the credit bureaus should pay to learn the payment history of each customer. There are other techniques such as reclassification, reweighting and extrapolation which are suggested in the literature to cope with the negative effects of reject inference (Ash and Meester, 2002; Crook and Banasik, 2004; Parnitzke, 2005). In this research, we use an extrapolation technique to reduce the negative effects of rejection bias. We group the customers into 10 bands depending on the probability values of being good for the customer. The band of a customer is calculated as:

$$B_i = \lfloor P(G \mid C_i) \cdot 10 \rfloor. \tag{4}$$

In Equation 4, B_i is the band of customer i which is the largest integer smaller than $P(G \mid C_i) \cdot 10$, and $P(G \mid C_i)$ is the probability of being good for customer i in which C_i is the registered customer information set of the customer. To make assessment of Bayesian credit scoring model, we obtain another data set consisting of 1850 personal loan customers in 2008. During that process, it is observed that there are no customers assigned to first three bands. This is resulted from the cut-off score such that the applications of the customers whose probability values are lower than 0.4 are rejected. Since the proportion of the rejected customers are approximately estimated as % 20, 125 customers holding the characteristics of the band in which they are involved are generated for each of the three bands to complete the data set.

3.1. Assessment of Bayesian Credit Scoring

Performance of any credit scoring model is evaluated based on the percentage of the customers whose status is predicted correctly. Let A denote the credit decision variable, and Y be the variable denoting the status of a customer. A is equal to 0 when credit is not granted to the customer; and it is 1 otherwise. Likewise, Y is equal to 0 when credit of the customer goes into default; and it is 1 otherwise. Moreover, let $N_{A_i Y_j}$ denote the number of customers with the states of $A = i$ and $Y = j$. Then, the contingency table in Table 1 shows the number of customers depending on their credit decisions and status. The first diagonal

Table 1. Contingency Table

	$A=0$	$A=1$
$Y=0$	$N_{A_0Y_0}$	$N_{A_1Y_0}$
$Y=1$	$N_{A_0Y_1}$	$N_{A_1Y_1}$

in contingency table ($N_{A_0Y_0}$, $N_{A_1Y_1}$) express the correctly classified customers, and the second diagonal shows the wrongly classified customers ($N_{A_1Y_0}$, $N_{A_0Y_1}$).

The correct classifications of good and bad customers do not have the same importance from the point of view of the creditors. A loss from a bad applicant is expected to be larger than profit obtained from a good customer in today's competitive credit market. Hence, in our evaluation, we use different weights for $N_{A_0Y_0}$ and $N_{A_1Y_1}$ as 2:1, although the weights of the elements in the first diagonal of the contingency table were taken equal by some researchers (Parnitzke, 2005; Crook and Banasik, 2004). Then, the performance indicator (PI) is formulated as below.

$$PI = \frac{2 \cdot N_{A_0Y_0} + N_{A_1Y_1}}{2 \cdot N_{Y_0} + N_{Y_1}}. \tag{5}$$

PI values depending on cut-off scores for Bayesian credit scoring model is provided in Fig. 2. The maximum PI value for logistic regression credit scoring model is % 62.15 that is also equal to the PI value of the Bayesian credit scoring model when the cut-off score is equal to 0.4.

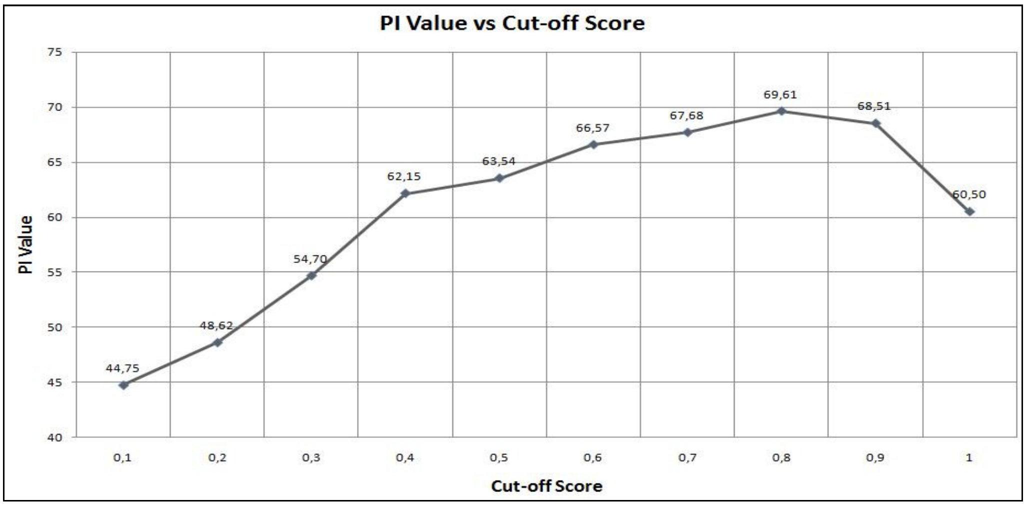

Fig. 2. Weighted Percentage of Correctly Classified Customers

PI values for Bayesian credit scoring model exceed the maximum PI value of the logistic regression credit scoring model when cut-off scores are equal to 0.5, 0.6, 0.7, 0.8 and 0.9. The maximum PI value for Bayesian credit scoring model is achieved at 0.8, and the gap between the maximum PI values of Bayesian and logistic credit scoring models is % 7.46. Therefore, by taking the maximum PI values of both models into consideration, Bayesian credit scoring model is % 7.46 more effective than the logistic regression credit scoring model.

4. Conclusion & Further Research

Bayesian network has become very popular in recent years. Its application areas have been expanding consistently. In our work, credit scoring was facilitated by Bayesian network model in that the structure is developed with credit experts and the parameters are estimated by using customer data obtained from a

Turkish bank. A comparative study is also illustrated in order to compare the effectiveness of the Bayesian credit scoring model with logistic regression credit scoring model which has been currently used by the bank. The results show that the Bayesian credit scoring model performs better than logistic regression credit scoring model in classification of customers.

Further research can be conducted on improvement techniques in the quality of expert knowledge. The quality of expert knowledge is important since the development of the structure depends on it. We suggest that further research would be stimulated by learning the structure with knowledge from multiple experts. Having a pool of several experts, each expert would state its belief about the structure of the model. A posterior structure would then be induced by the individual structures. Apart from that, different databases might probably result in different structures when structure is learnt from data. Having n different databases resulting in n different structures, a posterior structure would be developed. A comparison of these two methods applied in credit scoring would be valuable for further research (Richardson and Domingos, 2003).

References

Alpaydin, E. 2004. *Introduction to machine learning*. MIT Press, Cambridge, MA 02142.

Ash, D.; Meester, S. 2002. Best Practicing in Reject Inferencing, in *Conference Presentation on Credit Risk Modeling and Decisioning.* Available from Internet: <http://fic.wharton.upenn.edu/fic/ash.pdf>.

Bank for International Settlements. 2006. International convergence of capital measurement and capital standards, in *Basel Manual*.

Bilmes, J. A. 1998. *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*: tech. rep. International Computer Science Institute.

Chang, K. C.; Fung, R.; Lucas, A.; Oliver, R.; Shikalo, N. 2000. Bayesian networks applied to credit scoring, *IMA Journal of Mathematics Applied in Business and Industry* 11(1): 1–18.

Chen, M. C.; Huang, S. H. 2003. Credit Scoring and Rejected Instances Reassigning through Evolutionary Computational Techniques, *Expert Systems with Applications* 24(4): 433–441. doi:10.1016/S0957-4174(02)00191-4

Crook, J.; Banasik, J. 2004. Does reject inference really improve the performance of application scoring models? *Journal of Banking and Finance* 28(4): 857–874.

Hand, D. J.; Henley, W. E. 1997a. Statistical Classification Methods in Customer Credit Scoring: A Review, *Journal of the Royal Statistical Society* 160(3): 523–541. doi:10.1111/j.1467-985X.1997.00078.x

Hand, D. J.; McConway, K.; Stanghellini, E. 1997b. Graphical models of applicants for credit, *IMA Journal of Mathematics Applied in Business and Industry* 8(2): 143–155.

Huang, S. H.; Chen, M. C.; Wang, C. J. Credit Scoring with a Data Mining Approach based on Support Vector Machines, Expert Systems with Applications (in press). Available from Internet: <http://www.sciencedirect.com>.

Liu, W.; Cela, J. 2007. *Improving credit scoring by generalized additive model.* SAS Global Forum 2007.

Mitchell, T.1997. *Machine Learning.* McGraw Hill, Ch. 6 p.

Parnitzke, T. 2005. *Credit scoring and the sample selection bias:* Working Papers on Risk Management and Insurance. Available from Internet: <http://symposium.fbv.uni-karlsruhe.de/10th/papers/Parnitzke%20-%20Credit%20scoring%20and%20the%20sample%20selection%20bias.pdf>.

Richardson, M.; Domingos, P. 2003. Learning with Knowledge from Multiple Experts, in *Proceedings of the Twentieth International Conference on Machine Learning*, 621–634.

Taylan, P.; Weber, G. W. 2007. *Multivariate adaptive regression spline and continuous optimization for modern applications in science, economy and technology.* Submitted to TOP.

Thomas, L. C. 2000. A survey of credit and behavioral scoring: forecasting financial risk of lending to consumers, *International Journal of Forecasting* 16(2): 149–172. doi:10.1016/S0169-2070(00)00034-0

Thomas, L. C.; Edelman, D. B.; Crook, J. N. 2002. Credit Scoring and its applications, in *SIAM Society of Industrial and Applied Mathematics*, Philadelphia.