Research Article

A Personal Credit Rating Prediction Model Using Data Mining in Smart Ubiquitous Environments

Jae Kwon Bae and Jinhwa Kim

1Department of Management Information Systems, Keimyung University, 1095 Dalgubeoldaero, Dalseo-gu, Daegu 704-701, Republic of Korea
2School of Business, Sogang University, 1 Sinsu-dong, Mapo-gu, Seoul 121-742, Republic of Korea

Correspondence should be addressed to Jae Kwon Bae; jkbae99@kmu.ac.kr

Received 16 November 2014; Revised 1 February 2015; Accepted 9 February 2015

Academic Editor: Damien Sauveron

Copyright © 2015 J. K. Bae and J. Kim. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study suggests a methodology called a smart ubiquitous data mining (UDM) that consolidates homogeneous models in a smart ubiquitous computing environment. It tests the suggested model with financial datasets. It basically induces rules from the dataset using diverse rule extraction algorithms and combines the rules to build a metamodel. This paper builds several personal credit rating prediction models based on the UDM and benchmarks their performance against other models which employ logistic regression (LR), Bayesian style frequency matrix (BFM), multilayer perceptron (MLP), classification tree methods (C5.0), and neural network rule extraction (NR) algorithms. To verify the feasibility and effectiveness of UDM, personal credit data and personal loan data provided by a Financial Holding Company (FHC) were used in this study. Empirical results indicated that UDM outperforms other models such as LR, BFM, MLP, C5.0, and NR.

1. Introduction

The Global Financial Crisis in 2008 resulted from unreliable credit ratings assigned to mortgage-backed securities by world-renowned credit rating agencies with vested interests. Consequently, credit rating prediction has been a critical issue in the financial markets and global banking systems. The purpose of credit rating prediction is to classify the applicants into two classes: applicants with good credit histories (accepted) and applicants with bad credit histories (rejected). The predictive accuracy of credit ratings is critical to financial institutions’ profitability.

Various data mining algorithms have been adopted for predicting credit ratings, including classical statistical methods, tree-structured classification methods, and artificial neural networks (ANN). Desai et al. [1] investigated the accuracy of credit rating prediction models by using the personal loan information of three United States credit unions. They have been conducted that have compared ANN with other traditional classification algorithms in the field of credit prediction models, since the prediction accuracies of ANN are better than linear discriminant analysis (LDA) and logistic regression (LR). Yap et al. [2] illustrated using data mining algorithms to improve assessment of credit worthiness using credit rating prediction models. They compared a data mining based credit scorecard model, LR, and decision trees (DT) and did not detect a significant difference in classification error rates. Chung and Suh [3] suggested ANN and DT based credit rating prediction models to estimate the utility value of credit card delinquents. Huang et al. [4] suggested credit rating prediction models using support vector machines (SVM) and back-propagation neural networks (BPNN) to evaluate credit ratings and used 21 variables from the United States and Taiwanese financial markets. The results showed that SVM achieved accuracy comparable to that of BPNN. Ince and Akta [5] compared the performance of several credit rating prediction models applied to credit loan data. They used four traditional statistical methods, multiple discriminant analysis (MDA), LR, neural networks (NN), and classification and regression trees (CART), and suggested that CART obtained the best accuracy performance. Cubiles-De-La-Vega et al. [6] used the credit card dataset of Peruvian microfinance institutions to develop credit prediction models.
by using supervised classification techniques. Their proposed model exhibited better performance than LDA, LR, multilayer perceptron (MLP), and CART.

Studies on credit rating prediction have been conducted extensively. However, financial institutions and credit scoring agencies do not actively adopt the various credit rating prediction models presented in these studies for real-world applications because of their poor accuracy. To the best of our knowledge, the structural problems of financial institutions, which can contribute to the poor classification accuracy of credit rating prediction models, have not been considered thus far. Therefore, this study focuses on such problems.

Financial Holding Companies (FHCs) in Korea have various subsidiaries such as banking, credit cards, investment and securities, life insurance, real estate trust, data system, and saving bank. These subsidiaries have been building independent financial databases (DBs) and decision-making models for solutions to financial decision-making problems.

There are problems in these databases such as data duplication, inconsistency, and isolation, which lead to inefficient use of the databases. These problems can cause incorrect and poor decision-making. In order to facilitate decision-making, the financial DBs of all the subsidiaries should be integrated, and the financial data scattered among them should be stored and managed in the integrated DB. Accordingly, this study attempts to present a ubiquitous data mining- (UDM-) based credit rating prediction model that integrates the financial datasets (credit loan datasets) of FHC, and the integrated DB is used to solve financial decision-making problems. UDM-based credit rating prediction models are combining the knowledge in the form of rule sets extracted from data mining algorithms in smart ubiquitous environments.

Based on these considerations, we develop UDM for credit rating prediction that combines rules extracted using Bayesian style frequency matrix (BFM), decision trees (C5.0), and neural network rule extraction (NR). The proposed model is applied to the personal credit data provided by FHC in Seoul, Republic of Korea. After building credit rating prediction models with a dataset of personal credit using LR, NN, BFM, C5.0, and NR, the forecasting capabilities of the benchmarking models and those of UDM were compared.

The remainder of this paper is organized as follows. Section 2 introduces the UDM-based credit rating prediction models. Section 3 presents the research design, including the datasets, preprocessing, feature selection, and adjustment of class distribution. Section 4 presents the main empirical results of the experiments. Lastly, the empirical findings and future research are discussed in Section 5.

2. Concepts of Ubiquitous Data Mining (UDM) Classifier

With information and inspiration from learning processes in cognitive and neural sciences, this study suggests a model, which simulates human learning and decision-making processes. Each of us understands the same situation or event differently. The description from each can, therefore, be different. The basic assumption in this study is that humans are smart enough to combine pieces of knowledge from different sources. A smart person can combine advices from many people and utilize this combined knowledge for further decision-making. A group intelligent system such as a Wiki can be an example of this. The basic idea of group intelligence system is that knowledge from many is better than one. Like human, each model has its own specialty in the processes of inducing knowledge and unique coverage in domain knowledge inherent from its algorithm or heuristic.

This study suggests a methodology which is called UDM. It firstly induces knowledge in the form of rules from various algorithms such as BFM, C5.0, and NR in a smart ubiquitous computing environment. It then systematically consolidates these rules to build an integrated metamodel. Figure 1 shows the conceptual model of UDM. The key idea of UDM is to combine a number of classification models into one such that the performance of the consolidated model is better that that of the original individual classification models in their classification accuracy and efficiency. The objective of the study is to increase the accuracy in personal credit rating prediction by utilizing unique characteristics of diverse algorithms found in the literature. The tests show that UDM produces better performance than single classification model because it consolidates diverse rules from different algorithms.

Figure 2 shows how UDM is working with 3 steps: (1) constructing different models using datasets, (2) inducing rules from these models, and (3) consolidating these rules.

Step 1: Model Construction in Smart Ubiquitous Computing Environments. In the first stage, credit rating prediction models are built using four data mining algorithms (i.e., LR, BFM, NN, and DT) based on an integrated financial DB in smart ubiquitous computing environments. In this study, we used a stepwise logistic regression method to build a logistic-regression-based credit rating prediction model, and we used a case-based learning (CBL) algorithm (i.e., BFM) to build a Bayesian-network-based credit rating prediction model. In the case of NN, we established a standard three-layered back-propagation network. In the case of DT, we used the measure

![Figure 1: The pictorial model of UDM.](image-url)
Step 2: Rule Extraction Using BFM (Bayesian Style Frequency Matrix), C5.0 (Decision Tree), and NR (Neural Network Rule Extraction). To convert a heterogeneous model into a rule set, a uniform knowledge representation format, rules are induced from rule extraction algorithms such as BFM, C5.0, and NR. These rules are again consolidated into one metamodel, which combines these rules into one repository (consolidated knowledge base) of rules for the model. In Figure 2, rules extracted from the BFM are denoted by "R(Rule Set)F," whereas those extracted from C5.0 are denoted by "R(Rule Set)C." Moreover, multiple rules are extracted from neural networks using the NR. In Figure 2, rules extracted using the NR are denoted by "R(Rule Set)N."

Step 3: Rule (Knowledge) Integration of BFM, C5.0, and NR for UDM. UDM-based decision-making models are combining the knowledge in the form of rule sets extracted from BFM, C5.0, and NR. The rule sets R(Rule Set)F, R(Rule Set)C, and R(Rule Set)N in step 3 are merged into a cumulative rule set \[ R(Rule Set)_F + R(Rule Set)_C + R(Rule Set)_N \]. Each rule in the consolidated knowledge base from UDM serves as an agent as it is in the form of IF condition THEN decision.

3. Data Collection and Sampling

The raw data used in our experiment comprises personal credit data, personal loan applications data, and personal loan data acquired from FHC, Korea’s leading financial holding institution, with 30-million-strong customer base and business network exceeding 1,200 branches, the largest in Korea. With total assets exceeding US$ 284 billion and sales exceeding US$ 22 billion in 2013, it wields considerable capital influence and enjoys strong brand loyalty. As of July 2014, FHC has ten domestic subsidiaries (banking, credit cards, securities, life insurance, asset management, real estate trust, venture capital, credit information, data system, and savings
Table 1: Description of predictor variables.

| Attribute | Description                                      |
|-----------|--------------------------------------------------|
| X1        | Average balance for the last six months          |
| X2        | Previously granted credits                       |
| X3        | Loans granted in the last year                   |
| X4        | Previously denied loans                          |
| X5        | Total number of fees paid in credit history      |
| X6        | Number of arrears                                |
| X7        | Cash dispenser amount                            |
| X8        | Other cash dispenser total amounts               |

The data used herein comprises loan transactions of 10,062 customers, from January 2004 to December 2008, and includes various types of information used by credit scoring managers to make decisions regarding loan approval for customers, including personal characteristics of borrowers (age, income, job, academic ability, marital status, housing, property, etc.), loan information (purpose of loan, type of interest, loan amount, loan duration, etc.), and credit information (payment history, credit amount, credit history, etc.). After eliminating missing and abnormal cases, 8234 cases remained. Typically, the number of bad customers (applicants with bad credit) is much smaller than the number of good customers. The data consist of 4117 bad cases (customers with bad credit) and 4117 good cases (customers with good credit). 12 variables are selected via preliminary screening in the form of two-sample $t$-tests, and the remaining 8 variables are finally selected by stepwise LR. Table 1 shows the input variables used in this study. The variables finally selected include the average balance for the last six months, previously granted credits, loans granted in the last year, previously denied loans, total number of fees paid in credit history, number of arrears, cash dispenser amount, and other cash dispenser total amounts.

To perform an appropriate comparison of the prediction (classification) models the final dataset was randomly split into two subsets: a training set of 70% (5764 customers) and a validation set of 30% (2470 customers). A systematic method was required to evaluate and compare the classification models to determine the performance for our specific problem. Therefore, this study implemented fivefold cross validation to minimize the impact of random variation in the training set [7]. In fivefold cross validation, the original sample is randomly partitioned into five equal size subsamples.

### 4. Results of Experiment

Table 2 summarizes the results of the number of extracted rules and average number of rules using BFM, C5.0, NR, and UDM. From the five datasets (datasets 1–5), 35 rules were extracted using BFM, 82 rules were extracted using C5.0, 41 rules were extracted using NR, and 158 rules were extracted using UDM. The average number of rules for the BFM was found to be 7, representing the least number of rules extracted in the algorithms, and the average number of rules for C5.0 was found to be 16.4, representing the highest number of extracted rules. UDM is an accumulation of the rules extracted from the three above-mentioned rule extraction algorithms, and the average number of rules was found to be 31.6. The greater the number of accumulated rules, the more comprehensive the credit rating prediction model is.

Table 3 and Figure 3 compare the classification performance of LR, NN, BFM, C5.0, NR, and UDM using 5-fold cross validation. We can evaluate the prediction performance using the accuracy rate (also referred to as hit ratio), which is calculated by dividing the number of correct predictions by the total number of predictions. The main purpose of this cross validation procedure is to obtain the average accuracy rates for all iterations in the five sets (five iterations per set). Fivefold cross validation is employed to enhance the generalizability of the test results [8, 9]. UDM has the highest average classification accuracy with 78.67%, followed closely by NN (72.75%) and NR (70.04%). The result shows that the performance of UDM is superior to that of the other models such as LR, NN, BFM, C5.0, and NR.

We use the Wilcoxon signed-rank test to examine whether the predictive performance of UDM is significantly better than that of other single models. Like the $t$-test,
Table 3: Performance comparison of 6 models.

| Dataset number | Result | LR    | NN    | BFM   | C5.0  | NR    | UDM   |
|----------------|--------|-------|-------|-------|-------|-------|-------|
| Dataset 1      | Training | 70.69%| 72.22%| 69.00%| 71.79%| 70.06%| 81.02%|
|                | Validation | 66.52%| 71.38%| 69.35%| 69.55%| 68.99%| 80.65%|
| Dataset 2      | Training | 72.41%| 72.62%| 70.85%| 72.95%| 69.78%| 76.04%|
|                | Validation | 70.53%| 72.19%| 68.87%| 68.42%| 66.56%| 74.53%|
| Dataset 3      | Training | 69.83%| 73.11%| 72.22%| 72.05%| 71.34%| 77.20%|
|                | Validation | 70.20%| 72.71%| 69.27%| 69.19%| 72.23%| 77.65%|
| Dataset 4      | Training | 72.41%| 75.62%| 71.95%| 72.41%| 73.47%| 81.59%|
|                | Validation | 69.88%| 75.06%| 69.15%| 67.37%| 70.61%| 81.21%|
| Dataset 5      | Training | 69.83%| 73.06%| 71.48%| 70.35%| 72.31%| 78.45%|
|                | Validation | 68.14%| 72.39%| 69.31%| 68.91%| 71.82%| 79.31%|
| Avg.           | Training | 71.03%| 73.33%| 71.10%| 71.91%| 71.39%| 78.86%|
|                | Validation | 69.05%| 72.75%| 69.19%| 68.69%| 70.04%| 78.67%|

Note: LR (logistic regression), NN (neural networks, MLP), BFM (Bayesian style frequency matrix), C5.0 (decision tree), NR (neural network rule extraction), UDM (ubiquitous data mining): FM + C5.0 + NR.

Table 4: Wilcoxon signed-rank test (validation sets).

| Dataset number | UDM – LR | UDM – NN | UDM – BFM | UDM – C5.0 | UDM – NR |
|----------------|----------|----------|-----------|------------|----------|
| Dataset 1      | −22.115  (0.000) | −19.650  (0.000) | −21.327  (0.000) | −17.520  (0.000) | −15.284  (0.000) |
| Dataset 2      | −16.323  (0.000) | −6.708   (0.000) | −18.638  (0.000) | −5.874   (0.000) | −8.652   (0.000) |
| Dataset 3      | −24.580  (0.000) | −6.894   (0.000) | −20.956  (0.000) | −10.198  (0.000) | −6.403   (0.000) |
| Dataset 4      | −29.279  (0.000) | −15.440  (0.000) | −24.052  (0.000) | −12.606  (0.000) | −11.950  (0.000) |
| Dataset 5      | −27.795  (0.000) | −10.747  (0.000) | −22.077  (0.000) | −5.390   (0.000) | −10.250  (0.000) |

Note: z-statistics, P value.

The Wilcoxon test involves comparisons of differences between measurements, and, hence, it requires the data to be measured at an interval level of measurement [10]. Table 4 shows the results of the Wilcoxon signed-rank test for evaluating the classification performance of UDM. As we can see from Table 4, the performance of UDM is significantly better than that of the other models (1% or 5% significance level for the most of datasets). The results prove that multiple rule sets derived by UDM are more accurate than other traditional single models.

5. Conclusions and Future Directions

A critical issue in financial decision-making is to forecast, accurately and timely, business failure such as credit rating prediction, bankruptcy prediction, and delisting decision prediction. Credit rating prediction influences important managerial decisions affecting financing and firm value. The performance of the UDM is compared with that of LR, NN, BFM, C5.0, and NR to prove the efficiency of the suggested model using 5-fold cross validation process. Multiple tests with 5 different training and validating sets confirm the efficiency of the model from UDM.

To demonstrate the effectiveness of proposed model, UDM is performed on a credit loan dataset obtained from FHC in Korea. The result shows that the performance of UDM is superior to that of other models such as LR, NN, BFM, C5.0, and NR. This enhancement in the predictability of credit ratings can significantly contribute to the correct credit admission evaluation of loan customers, and, hence, FHC and international financial institutions can employ UDM for better credit appraisal, risk analysis, and lending decisions.

Our study has the following limitations, which require further investigation. Firstly, the results from the study should be generalized. Our study uses only a single selected dataset for system validation. However, only one dataset may not be reliable for making a conclusion. Other problem domains (credit risk assessment, corporate dividend payments prediction, insurance fraud detection, and delisting decision prediction) should be considered in further research. Secondly, the investigations could be extended to include other financial products such as mortgages, derivatives, and corporate loans. Lastly, further research may consider nonfinancial and macroeconomic variables for UDM inputs or to develop time-series credit rating prediction models that include the change of credit status in every period.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.
References

[1] V. S. Desai, J. N. Crook, and G. A. Overstreet Jr., "A comparison of neural networks and linear scoring models in the credit union environment," *European Journal of Operational Research*, vol. 95, no. 1, pp. 24–37, 1996.

[2] B. W. Yap, S. H. Ong, and N. H. M. Husain, "Using data mining to improve assessment of credit worthiness via credit scoring models," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13274–13283, 2011.

[3] S.-H. Chung and Y. Suh, "Estimating the utility value of individual credit card delinquents," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3975–3981, 2009.

[4] Z. Huang, H. Chen, C.-J. Hsu, W.-H. Chen, and S. Wu, "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decision Support Systems*, vol. 37, no. 4, pp. 543–558, 2004.

[5] H. Ince and B. Aktan, "A comparison of data mining techniques for credit scoring in banking: a managerial perspective," *Journal of Business Economics and Management*, vol. 10, no. 3, pp. 233–240, 2009.

[6] M.-D. Cubiles-De-La-Vega, A. Blanco-Oliver, R. Pino-Mejías, and J. Lara-Rubio, "Improving the management of microfinance institutions by using credit scoring models based on Statistical Learning techniques," *Expert Systems with Applications*, vol. 40, no. 17, pp. 6910–6917, 2013.

[7] S. M. Weiss and C. A. Kulikowski, *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*, Morgan Kaufmann, San Francisco, Calif, USA, 1st edition, 1991.

[8] G. Zhang, M. Y. Hu, B. E. Patuwo, and D. C. Indro, "Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis," *European Journal of Operational Research*, vol. 116, no. 1, pp. 16–32, 1999.

[9] D. Lee, J. H. Park, and H.-R. Bae, "Comparison between planned and actual data of block assembly process using process mining in shipyards," *The Journal of Society for e-Business Studies*, vol. 18, no. 4, pp. 145–167, 2013.

[10] W.-C. Hur, H.-R. Bae, S. Kim, and K.-S. Jeong, "A method for business process analysis by using decision tree," *The Journal of Society for e-Business Studies*, vol. 13, no. 3, pp. 51–66, 2008.
