Optimization of credit scorecard combinations based on quantum annealing algorithm

Bingze He [*, #], Xingyi Ji [#] and Jiaxin Lv

School of Electronic Information Engineering, South-Central Minzu University, Wuhan, China

* Corresponding Author Email: 1093724684@qq.com

[#]these authors contributed equally to this work.

Abstract. With the development of the times, the usage of loan-related business is increasing. However, most of the credit score card portfolio methods used nowadays are manually designed and supplemented with computer calculations, so the bank yields are low. Therefore, it is of great interest to study the application of quantum computers in credit score card portfolio optimization for better rating of customer credit ratings. Based on the given credit score card data, a linear programming model is developed. Firstly, the data is processed into two sets of one-dimensional data by Python and the working range of the two sets of data is found. Secondly, after deriving the objective function expressions according to the given conditions, the data were visualized by building a 3D grid matrix with the pass rate and bad debt rate as the XOY plane and the final benefit to the bank as the Z axis. Subsequently, the decision variables in the linear programming model are converted into binary variables, and the original model is transformed into a QUBO model by converting the constraints into binary variables. After analyzing the QUBO model, the quantum annealing method is chosen to solve the optimal solution of the objective function. And based on this, the case of three selected credit score card combinations and any three credit score cards are considered.

Keywords: Linear programming model; 3D grid matrix; credit scorecard; quantum annealing method.

1. Introduction

In a bank's credit card or related loan business, customers need to have their credit ratings rated by vetting rules to decide whether to qualify for a loan. These vetting rules are called credit score cards, and they can be a combination of one or multiple rules, while each credit score card has multiple thresholds, each corresponding to a different pass rate and bad debt rate. The banks, on their part, want to optimize the combination of review rules, i.e., credit score cards, in order to maximize bank revenue. However, most of the credit score card combination methods used nowadays are manually designed and supplemented with computer calculations, so the yield is slightly low. Therefore, it is of great interest to study the application of quantum computers in credit score card portfolio optimization to better rate customer credit ratings and further increase banks' profits [1].

Suppose 100 credit score cards are given, and each credit score card can be set with 10 thresholds. In this paper, a bank revenue model is developed to calculate the corresponding credit score card combinations and their thresholds when the bank's final revenue is at its maximum.

2. Modeling of bank earnings

2.1. Data pre-processing

The given data is initially processed using Excel to obtain a 100*10 array of numbers, and the data is pre-processed in Python to decompose a set of two-dimensional data into two sets of mixed one-dimensional data.

2.2. Page Numbers

The mathematical standard type of general linear programming is as follows [2]:

$$max\ z = \sum_{j=1}^{n} c_j x_j \tag{1}$$

$$s.t. \begin{cases} \sum_{j=1}^{n} a_{ij} x_j = b_i \\ x_j \geq 0 (j = 1,2,3 \cdots, n) \end{cases} \tag{2}$$

The general form of the QUBO model is as follows [3]:

$$\min_{x} \sum_{i=1}^{n} q_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} q_{i,j} x_i x_j \tag{3}$$

2.3. Model Building

By using the panda's database in Python, two sets of two-dimensional data are decomposed by circular filtering.

$$t = \{i \mid i > 0.5\} \tag{4}$$

$$h = \{i \mid i < 0.5\} \tag{5}$$

The calculation gives:

$$0.7 < t < 0.99 \tag{6}$$

$$0.003 < h < 0.088 \tag{7}$$

The final income of the bank can be defined as:

$$Z = I - Y \tag{8}$$

Where interest income can be defined as:

$$I = m \cdot r \cdot t \cdot (1 - h) \tag{9}$$

Bad debt losses can be defined as:

$$Y = m \cdot t \cdot h \tag{10}$$

Then the final return to the bank at this point can be changed to

$$Z = m \cdot r \cdot t \cdot (1 - h) - m \cdot t \cdot h \tag{11}$$

After using meshgrid to synthesize two sets of one-dimensional data with t and h as XOY surface respectively, and generating 3D grid matrix with bank final revenue as Z axis, as shown in Figure 1.

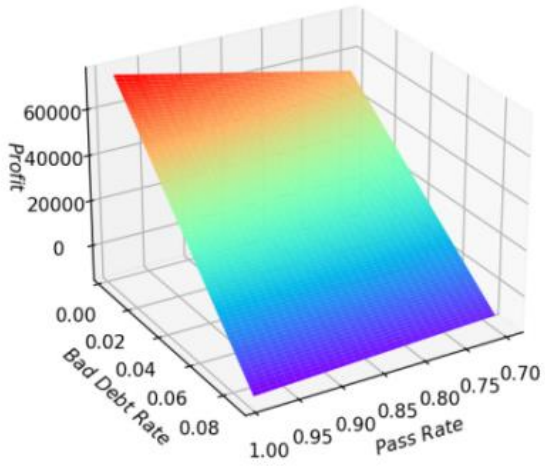


Fig 1. Revenue Grid Matrix.

After converting the decision variables r and h in this model into 0-1 binary variables, the model is transformed into QUBO form.

Define x_i as the binary variable corresponding to the ith decision variable, and q_i and $q_{i,j}$ as the coefficients between x_i and x_j in the objective function. The objective function of the model corresponding to the QUBO form is as follows:

$$\min_x \sum_{i=1}^{n} q_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} q_{i,j} x_i x_j \tag{12}$$

Due to the large amount of data, we consider using the simulated annealing algorithm for the optimization and calculation of the QUBO model.

First we define the energy function of the simulated annealing algorithm for this model condition [4]:

$$E(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} q_{ij} x_i x_j + \sum_{i=1}^{n} q_{ij} x_i \tag{13}$$

Second, a binary number x is chosen as the initial point for annealing with a random number, and a second number y is chosen as the second annealing point within the neighborhood structure of x. Both satisfy the following equation:

$$\begin{cases} When\ E(x) > E(y), x = y \\ When\ E(x) < E(y), x = y\ under\ certain\ conditions \end{cases} \tag{14}$$

When E(x) < E(y), a certain condition is p = exp(-ΔE/T), where ΔE = E(y) - E(x).

When E(x) is smallest, it is the optimal solution for this iteration.

Therefore, the QUBO model is implemented using Python to find the optimal solution for the use of the 49th credit score card, and the final maximum benefit to the bank is $61,172.0.

3. Optimal combination of selected credit card thresholds

It is necessary to find the optimal solution of the threshold combination among Card1, Card2, and Card3 to achieve the maximum gain.

On the model in the previous section, the selection of credit score cards is changed to the first three cards and $i \in \{1,2,3\}$ is defined, and the binary variable x_i is whether the ith card is selected, $x_i \in \{0,1\}$, when $x_i = 0$, which means that the card is not selected, and vice versa, when $x_i = 1$ means that the card is not selected.

Define $j \in \{1,2,3,\cdots,10\}$, y_{ij} as the jth threshold of the ith card, $y_{ij} \in \{0,1\}$

Define t_{ij} as the pass rate for the jth threshold of the ith card.

Define h_{ij} as the bad debt rate for the jth threshold of the ith card.

Define R_i as the revenue of the ith card.

At this point the objective function of the maximum gain from the combination of the thresholds of the three cards changes to

$$\max \sum_{i=1}^{3} x_i \cdot R_i \tag{15}$$

It is necessary to ensure that at least one credit score card exists for measurement, and it is necessary to ensure that there is only one threshold for one credit score card, while defining C as the linear maximum combination of the thresholds of the three cards and y_i as the threshold of the ith card.

Thus the constraints can be obtained as follows:

$$\sum_{i=1}^{3} x_i \geq 1 \tag{16}$$

$$\sum_{j=1}^{10} y_{ij} \leq 1 \tag{17}$$

$$\sum_{i=1}^{3} x_i \cdot y_i \leq C \tag{18}$$

Defining $t_{1,j1}$ as the pass rate for the j_1th threshold of the first card, the formula for the average of the corresponding thresholds for the 3 cards is as follows:

$$(t_{1,j_1} + t_{2,j_2} + t_{3,j_3})/3 \tag{19}$$

Define P_i to denote the sum of bank pass rates for the jth value of the i-th card and C_i to denote the sum of bank bad debt rates for the j-th value of the i-th card.

$$P_i = \sum_{j=1}^{10} t_{ij} \tag{20}$$

$$C_i = \sum_{j=1}^{10} t_{ij} \tag{21}$$

With the above constraints, the model in the above section is modified and the objective function is transformed into a QUBO model.

Denoting x_i as $\frac{1-z_i}{2}$, where z_i is a 0-1 binary variable, the objective function is then transformed into

$$Z = \frac{1}{2}\sum_{i=1}^{100}(P_i - C_i)Z_i - \frac{1}{2}\sum_{i=1}^{100} P_i \tag{22}$$

The model is processed using the quantum annealing method to obtain the energy function as follows [5]:

$$H = \sum_{i=1}^{3} x_i \cdot R_i - \lambda(\sum_{i=1}^{3} x_i \cdot y_i) \tag{23}$$

The QUBO model is implemented using Python to find the optimal solution with Card1 threshold of 8, Card2 threshold of 1, and Card3 threshold of 2. The maximum value of the bank return obtained is $27914.820.

4. Optimal combination of arbitrary credit card thresholds

In this section, the credit score cards are selected randomly out of one hundred, thus defining

$$i \in \{1,2,\cdots,100\} \tag{24}$$

The selected card number:

$$j \in \{1,2,3\} \tag{25}$$

Define the binary variable $x_{i,j}$ to indicate whether the ith card, the jth card, is selected as the jth card.

$$x_{i,j} = \{0,1\} \tag{26}$$

Define P_i to denote the sum of the bank pass rate under the jth value of the i-th card

$$P_i = \sum_{j=1}^{10} t_{ij} \tag{27}$$

C_i Denotes the sum of the bank bad debt rate for the jth value of the i-th card

$$C_i = \sum_{j=1}^{10} t_{ij} \tag{28}$$

To ensure that there is one and only one threshold for the selected credit score card, the following constraint is obtained:

$$\sum_{i=1}^{100} x_{i,1} \le 1, \sum_{i=1}^{100} x_{i,2} \le 1, \sum_{i=1}^{100} x_{i,3} \le 1 \tag{29}$$

$$\sum_{i=1}^{100} x_{i,1} + \sum_{i=1}^{100} x_{i,2} + \sum_{i=1}^{100} x_{i,3} \ge 1 \tag{30}$$

$$\sum_{j=1}^{3} x_{i,\,j} = 1, i \in \{1,2,...,100\} \tag{31}$$

After improving the model in the previous section, the QUBO model under the constraints of this problem can be obtained as

$$Z = \frac{1}{2}\sum_{i=1}^{100}(P_i - C_i)Z_i - \frac{1}{2}\sum_{i=1}^{100} P_i \tag{32}$$

The energy function model for the quantum annealing method is constructed as follows:

$$H = \sum_{i=1}^{3} x_i \cdot R_i - \lambda(\sum_{i=1}^{3} x_i \cdot y_i) - \sum_{i=1}^{3} z_i^{\,2} \tag{33}$$

The QUBO model is implemented using Python to find the optimal solution as Card1 picks the 7th credit score card, Card2 picks the 32nd credit score card, and Card3 picks the 48th credit score card, and the maximum bank return obtained is $43,880.9712.

5. Summary

The model is flexible in selecting credit score cards and corresponding thresholds for calculation, which greatly improves the flexibility of the model. In this model, by using an analog quantum computer to calculate the QUBO problem, the speed and efficiency of the solution are greatly improved. By using binary variables, the constraints are converted into QUBO form, which makes the overall architecture of the model simpler and easier to understand.

When building the model, the model only considers a single credit scorecard and threshold, resulting in a poorly adaptable model that cannot be adapted to the full range of conditions in the database. The model is limited by the lack of data and related constraints, and may not be able to adapt to the full range of conditioned databases. The model was built without considering the impact on the overall stability of the model when different credit scorecards were selected, which may lead to inaccurate results.

References

[1] Wang Rui. Research on the application of Boosting family algorithm in credit score card model [D]. Liaoning:Dalian University of Technology,2021.

[2] Ding Li. Teaching ideas of linear programming model solutions[J]. Gansu Science and Technology Vertical,2006,35(2):211-212.

[3] Wen K, Ma Y, Wang P, et al. A research report on credit score feature screening based on optical quantum computing[J]. Network Security and Data Governance, 2022, 41(9):13-18.

[4] Liu, Shuang-Shuang, Li, Jun. Research on simulated pre-distortion based on simulated annealing method[J]. Data Communication,2023(2):42-45,54.

[5] Zhang Hongtao,Xiong Hongmei,Tu Lingying. An improved quantum annealing algorithm[J]. Journal of Jiangxi Normal University (Natural Science Edition),2016,40(5):473-475.