บทความวิจัย (Research Article)

The Optimal Course Bidding Strategy under Limited Resource Constraint using Genetic Algorithm

Thanatchaporn Sri-art and Prabhas Chongstitvatana*

Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University
* ผู้รับผิดชอบบทความ prabhas.c@chula.ac.th

Received: 15 Jun 2023
Revised: 16 Jul 2023
Accepted: 22 Jul 2023

Abstract

In a course bidding system, there are more students than the number of available seats for a course. To enroll a course, students have to bid using their tokens and the system will fill up the available seats with the top bidders. Since the students have a limited number of tokens, they have to allocate their tokens wisely. In this paper, we apply a genetic algorithm to search for the best way to allocate the tokens such that it maximizes the probability of successful enrollment. To estimate the probability, we train a logistic regression on the course registration data and the model achieves 78.39% accuracy. By using the synthesized dataset, we compare the effectiveness of tokens suggested by the genetic algorithm and other approaches such as heuristics and excel built-in solver. The results from several experiments with different scenarios and settings suggest that the genetic algorithm tends to provide a set of tokens that produce the highest probability of successful enrollment compared to the other approaches.

Keywords: Genetic algorithm, Bidding strategies, Resourse allocation, Constraint optimization, Machine learning

1. Introduction

Several industries have applied the auction theory for the problem of allocating a limited resource in a situation where the demand is higher than the supply. In the education industry, for example, the auction theory and a bidding mechanism are used for the application of course registration. The bidding system will prioritize the higher bids over the lower bids, and fill in the available seats for a course with the top bidders. Therefore, the students have to competitively bid the course in order to get enrolled. However, most students do not have access to the past statistics of the course registration data, so they usually bid a very high price for a high demand course, resulting in wasting the resource (token). To solve this problem, [3] use a machine learning approach and [4] use a quantum computing approach to estimate the token price for bidding a course. However, in a real-world scenario, students usually enroll multiple courses all at once. Thus, the students have to decide how to allocate their tokens based on individual recommended tokens. In addition to that, these tokens are likely not the optimal solution since they are separately determined for a set of courses.

In this paper, we propose a genetic algorithm to search for an optimal set of tokens for course bidding that maximizes the average probability of successful enrollment in a fixed number of courses. One challenge is that how to estimate the probability. We tackle this challenge by training a logistic regression model on the past registration data to estimate a probability of successful enrollment of a course. Note that other machine learning models such as neural network can also be used to achieve this task, but we choose the logistic regression due to its simplicity and its interpretability. Our datasets contain a course identification number (course id), a limit seat available (limit), and a number of tokens (tokens) as features. Our goal is to build a token recommendation system to help students effectively allocate their tokens while having a high probability to win the bidding.

To this end, we utilize the trained logistic regression model as part of the objective function of the genetic algorithm. One of the reasons that we choose the genetic algorithm is that it is a stochastic process and its evolutionary procedures allow for escaping a local minimum.

2. Literature review

2.1 Course Bidding

Graves et al. [1] applied an auction mechanism to the course registration system in University of Chicago Graduate School of Business to optimally allocate course sections to students. It also takes into account student preferences and bidding points during the allocation.

According to Sonmez et al. [2], course bidding was designed to deal with the situation where demand is higher than supply. It promotes competition between students in terms of price bidding. Students who put higher bid will have higher chance of being enrolled for a course compared to other students who bid lower. This bidding mechanism rely on a marketing mechanism which promotes capitalism. Courses with limited seats will be allocated properly to students who pay more price and the price they pay implies their willingness in getting enroll for the courses. Bidding strategies are important as they help students to have competitive advantages to win the bid.

J. Chonbadee, [3] train three machine learning models, including decision tree, random forest, and artificial neural network (ANN), for predicting a token for a course. As a result, the ANN achieves the best performance with RMSE of 3.98. After that, L. Suraphan. [4] train a quantum neural network (QNN) on the same dataset used in [3] to predict a token for a course and achieve RMSE of 6.38, while having a less number of model parameters. Both [3] and [4] can suggest a token for one course at a time, while in our work, we can suggest a set of tokens by considering multiple courses all at once.

2.2 Genetic Algorithm

Genetic algorithm is an optimization procedure inspired by the biological theory to search the optimal solution by simulating the natural evolution. The algorithm uses binary representation and converts the process into selection, crossover and mutation [5-6].

The algorithm works in an iterative manner starting from generating a fixed number of random bitstrings as an initial population. In each iteration, the algorithm evaluates each of the bitstrings in the population set with the objective function to get the corresponding objective scores, tracks the scores, and uses the selection, crossover and mutation operations to evolve the bitstrings within the current population to generate the new one for the next iteration. After a fixed number of iterations is reached, the algorithm returns the solution (i.e., the bitstring) with the highest objective score.

The algorithm selected the parents based on their fitness score of two pairs of individuals (parents). Individuals with high fitness have more chances to be selected for reproduction in the next generation. The parents with high fitness score were selected to crossover involves a random split on the bitstrings after split point then create the child from the first parent to the end of the second parent. Mutation flipped bit of the child it can be candidate solutions.

The genetic algorithm is implemented in several industries such as in the financial section, [7-8] optimized the bidding strategy for the profit maximization.

The education industry implements the genetic algorithm into optimal school resources such as course scheduling problem, reducing energy consumption and electricity costs [9], creating course schedule with limited factors for the flexibility of classrooms and time arrangements [10-12].

3. Methodology

3.1 Estimating the Probability of Successful Enrollment

We utilized a logistic regression model to predict a probability of successful enrollment. Logistic regression model is a linear model that can be used to perform a classification tasks as sigmoid function is apply to cap the output of the linear model to be within the range between 0 and 1 and then we can threshold the value to determine a positive or negative class [13-14]. Mathematically, the hypothesis $h_\theta(\vec{X})$ of the logistic regression model is

$$h_\theta(\vec{X}) = \sigma(\theta^T \vec{X}) \tag{1}$$

where $\vec{X}$ is a set of input features, θ is a set of parameters of the logistic regression model, and $\boldsymbol{\sigma}$ is the logistic sigmoid function define as $\sigma(z) = \frac{1}{1+e^{-z}}$ The output of the logistic regression model can be interpreted as

$$\boldsymbol{p(Y = 1|\vec{X}; \theta)} \tag{2}$$

where $Y \in \{0, 1\}$ is a target label such that **0** represents a negative class and **1** represents a positive class as shown in Figure 1.

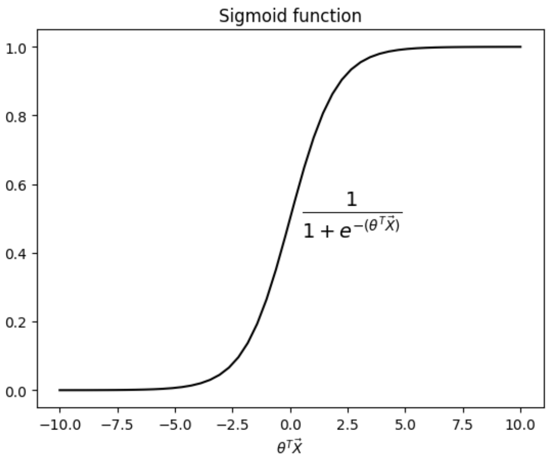

Figure 1 Sigmoid function can classify output within the range between 0 and 1

3.2 Searching for a set of tokens

In our problem, we use the genetic algorithm to search a set of tokens to recommend for course bidding, our objective is to find a set of tokens that maximizes the average probability of successful enrollment of a fixed number of courses. The objective function is defined as follow:

Objective function $Maximize\ P(Y = 1|\vec{X})$ (3)

$$P(Y = 1|\vec{X}) = \begin{cases} 0, & \sum_{i=1}^{N} x_1^i > B \\ \sum_{i=1}^{N} \frac{1}{1 + e^{-(\vec{\theta}^T \vec{X}^i)}}, & x \geq 0 \end{cases}$$

where, $\vec{X}^i$ is a three-dimensional feature vector in $\mathbb{R}^3$ for the i-th course, N is a number of courses, $\vec{X}^i$ = [x_1^i, x_2^i, x_3^i] is a feature vector such that x_1^i is a token for the i-th course, x_2^i is a course number for the i-th course, x_3^i is a limited number of seats for the i-th course, B is a balance token of a student and θ is a three-dimensional vector in $\mathbb{R}^3$ representing fitted parameters of the logistic regression model, and $P(Y = 1|\vec{X})$ is an average probability of successful enrollment given a set of features. Note that for the case where the sum of the tokens does not exceed the balance token, the probability of successful enrollment is given by the equation of the logistic regression model. Figure 2 illustrates the high-level.

4. Experiment

4.1 Data Preparation

The information obtained from the database. The course bidding mechanism starts with 1,000,000 tokens and requires a minimum bid of 50,000 tokens for each course. The statistics are as follows: 5,456 records, with a minimum of 50,000 tokens, a mean of 118,307.95 tokens, a standard deviation of 87,677.84 tokens, and a maximum of 703,905 tokens.

To estimate the probability of successful enrollment, we trained a logistic regression model on the past registration records in the course bidding system. The dataset consists of 5,456 records which we used 4364 (80%) records for training and 1092 (20%) for testing. For each of the records, we have four features include course id, limit, and tokens, and one target value which is a registration status that we preprocess it so that its value can be either 1 and 0, representing successful enrollment and unsuccessful enrollment, respectively. Our logistic regression model achieves an accuracy of 78.39% on the test set.

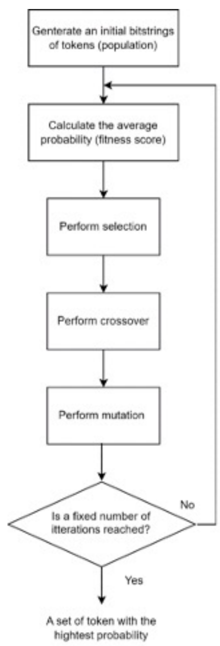

Figure 2 The high-level overview of the algorithm overview of our algorithm

4.2 Course bidding mechanism

Students are initially given the same number of tokens and can bid on more than one course. The maximum number of course per student in fixed (5 in the experiment). All students start with the same number of tokens. In this study, the number of tokens is fixed for each year (two semesters). Students have to strategically decide their study plan each year. The course bidding method prioritizes higher bids above lower offers and fills available seats in a course with the highest bidders. In the event of a tie, seats will be assigned to the bidder who submitted first. If a course's bid is unsuccessful, the system will return the tokens.

4.3 Token spending strategies

In our experiment, we set up four strategies for recommending a set of tokens to enroll courses as baselines to compete with the genetic algorithm approach. We constrained each of the strategies to recommend 50,000 tokens at a minimum and the total tokens suggested for 5 courses cannot exceed the balance token of a student. The four strategies include:

4.3.1 Random token strategy (Random)

The random token strategy randomly suggests a token for each course within a range between the minimum token (50000) and the balance token that a student has available. It also has a constraint that the sum of total tokens for all courses cannot exceed the balance token.

4.3.2 Average balance token strategy (Average)

The average balance token strategy suggests an equal amounts of tokens for every course by dividing the balance token with the number of courses.

4.3.3 Minimum token strategy (Minimum)

The minimum token strategy suggests a minimum token (50000 token align with the exploration) for each course.

4.3.4 Excel built-in solver strategy (Solver)

[15] and [16] applied a linear programming using a Microsoft Excel's solver to effectively allocate limited resources. Similarly, we utilize Generalized Reduced Gradient Nonlinear Solving (GRG Nonlinear) [17] which is one of the basic methods for solving nonlinear problems. The GRG Nonlinear method is used to suggest tokens with the following settings:

- Set the objective such that the sum of the tokens suggested for 5 courses equals to the balance token of each student.
- Set a constraint such that the minimum token suggestion for each course is greater or equals to 50,000 tokens.
- Set a solving method to the GRG Nonlinear method to find a solution.

The result returned from the solver is a set of tokens suggested for enrollment of the corresponding courses for a student.

4.4 Set up genetic algorithm

We represent a token suggested for each course as a bitstring using 20 bits, assuming that the token suggested for each course will not exceed one million. Since we have 5 courses for each student, concatenating the bitstrings for the courses together produces us a chromosome of 100 bits. The number of chromosomes in our population is set to 100. The crossover rate is set to 90%, the mutation rate is set to 10%, and the algorithm is executed for 1000 iterations.

4.5 Experiment setup

To evaluate the effectiveness of the genetic algorithm approach and the four token spending strategies, we synthesize course bidding registration into 2 scenarios as the following:

Scenario 1, The new student group. The data consists of 41 students receiving the 1,000,000 initial tokens in the first semester.

Scenario 2, The existing student group. The data consists of 78 students with a balance of tokens from previous academic year auctions.

Table 1 The results of the average probability of successful enrollment of 5 courses were estimated by logistics regression model for each strategy that sends their set of token suggestions for the student in scenario 1

Student id	Course id	Balance token	Avg Prob. GA	Avg Prob. Random	Avg Prob. AVG	Avg Prob. Min	Avg Prob. Solver
Student1	2110432, 2110433, 2110479, 2110413, 2110511	1,000,000.00	83.33% ± 0.26%	56.93% ± 4.43%	63.00%	33.96%	56.27%

Table 2 The results of the average probability of successful enrollment of 5 courses were estimated by logistics regression model for each strategy that sends their set of token suggestions for the student in scenario 2

Student id	Course id	Balance token	Avg Prob. GA	Avg Prob. Random	Avg Prob. AVG	Avg Prob. Min	Avg Prob. Solver
Student1	2110498, 2110433, 2110593, 2110481, 2110477	370,240.00	61.81% ± 0.10%	59.83% ± 1.22%	61.13%	53.35%	61.64%

Both scenarios have information for each of the students including a student id, a balance token, 5 course numbers that the student would like to enroll in, and the corresponding limit seats available of the courses that we look up from the database.

4.6 Evaluation

To evaluate the performance of different strategies, we suppose that all students would like to enroll in 5 courses. Each of the token spending strategies and the genetic algorithm is then applied to generate a set of suggested tokens for each student. After that, we use the logistic regression model to predict the probability of successful enrollment of each of the 5 courses based on the suggested tokens, course id, and the corresponding limit. We then summarize the returned probabilities into a single number using average as shown in Figure 3.

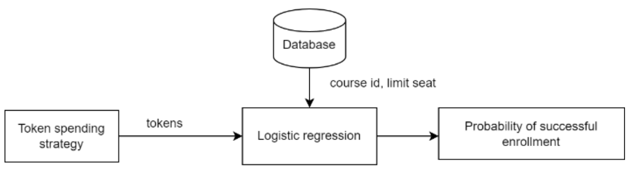

Figure 3 Token Spending Strategy include random, average, minimum, solver and genetic algorithm send input to logistic regression to get the probability of successful enrollment

Since the genetic algorithm and the random strategy are non-deterministic approaches, we ran them for 30 iterations and report the average probability of successful enrollment plus and minus standard deviation for each of the students. For other strategies, we report a single number corresponding to the average probability of successful enrollment for each student. Tables 1 and 2 provide scenarios of student 1 of 41 students in scenario 1 and student 1 of 78 students in scenario 2.

By contrasting the results in Tables 1 and 2. We count the number of students who acquire the highest probability from a set of tokens recommended by the strategies shown in Table 3.

Table 3 The number of students that get the highest probability from a set of tokens suggested by each strategy comparing between scenario 1 and scenario 2

Strategy	**Scenario 1**	**Scenario 2**
	Number of students	**Number of students**
Genetic algorithm	41	49
Random	0	0
Average	0	1
Minimum	0	0
Solver	0	28

As shown in Table 3, in scenario 1, 41 students were given the highest average probability of successful enrollment by a set of tokens recommended by the genetic algorithm, whereas, in scenario 2, 49, 28 and 1 students were given the highest average probability of successful enrollment by a set of tokens recommended by the genetic algorithm, the solver strategy and the average strategy, respectively.

Note that the genetic algorithm completely beats all other strategies in scenario 1, where each student has a large number of balance token, as it tries to find the best way to allocate the balance token into different courses such that the average of the returned probabilities evaluated by the logistic regression is maximized. However, having large number of balance tokens does not affect the way the other baseline strategies such as the random strategy, the average strategy and the minimum strategy allocate the balance token.

5. Conclusion

In this paper, we study the course registration using a token bidding system where each student has a limited number of tokens and each opening course has a limited number of seat available. Due to limited resources mentioned previously, it is beneficial to have a token suggestion system that recommends a set of tokens for bidding multiple courses with a highest average probability of successful enrollment.

To implement the token suggestion system mentioned previously, we employ machine learning and genetic algorithm. We start by framing the problem into two stages; 1) estimating a probability of successful enrollment for each course using a logistic regression model which its outputs always lie within the 0 and 1 range so that we can interpret them as probabilities, 2) searching a set of tokens that gives the highest average probability of successful enrollment (estimated by the logistic regression model trained in the first stage) while ensuring that the total sum of tokens does not exceed the budget (balance tokens) constraint by using a genetic algorithm.

We compare the genetic algorithm against multiple token suggestion strategy baselines that most of them based on heuristics. Our experiment shows that the tokens suggested by the genetic algorithm approach give higher average probabilities of successful enrollment for most of the time. One main advantage of our approach is that using genetic algorithm allows us to recommend a set of tokens for multiple courses with the associated average probability of successful enrollment instead of a single token prediction for a single course without a probability as in [2] and [3].

Recommending a set of tokens is very beneficial to students because they do not have to determine by themselves how to distribute their tokens across multiple courses. In addition to that, having access to the estimated probability making them aware of the associated risk so they can make an informed decision.

Furthermore, the set of tokens suggested by the genetic algorithm are the optimal solution that maximizes the average probability of successful enrollment. Therefore, students will likely to win the bid. Our approach is very generic and it can be applied to other constraint optimization problems not limited to the course bidding problem. We can also improve the accuracy of the logistic regression model by performing more feature engineering to obtain features with high predictive signals.

6. Acknowledgements

The author would like to express my deepest appreciation to Chaipat Suwannapoom for his valuable advice and encouragement throughout this paper. This work would not have been possible without his valuable advice to help makes the process much smoother and easier than expected. Furthermore, I had the pleasure to work with my advisor, Professor Prabhas Chongstitvatana provides opportunities and guidance on the concept on a number of occasions while conducting research during the preparation of this paper. Lastly, I would be remiss in not mentioning my family and my friends are always believed in me.

7. Reference

[1] Graves, Robert L., Linus Schrage, and Jayaram Sankaran, “An auction method for course registration,” *INFORM JOURNAL ON APPLIED ANALYTICSI.*, Vol. 23 (5), pp. 81-92, 1993.

[2] Sonmez, Tayfun Oguz and Ünver, M. Utku. “Course Bidding at Business Schools” *Boston College Working Papers in Economics, Boston College, Department of Economics.* 2005.

[3] J. Chonbadee, P. Krerk and C. Prabhas, “Token allocation for course bidding with machine learning method”. *18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). 19-22 May. Chiang-Mai Thailand* : 1168-1171, 2021.

[4] L. Suraphan and C. Prabhas, “Quantum Neural Network model for Token allocation for Course Bidding”. *25th International Computer Science and Engineering Conference (ICSEC). 18-20 November. Chiang-Rai Thailand* : 273–276, 2021.

[5] Mitchell, Melanie, *An Introduction to Genetic Algorithms. Cambridge*. MA: MIT Press, 1996.

[6] Jason Brownlee. (15 March 2023). *Simple Genetic Algorithm From Scratch in Python* **[Online] Available: https://machinelearningmastery.com/simple-genetic-algorithm-from-scratch-in-python/”**

[7] Azadeh, A., Ghaderi, S. F., Pourvalikhan Nokhandan, B., & Sheikhalishahi, M., "A new genetic algorithm approach for optimizing bidding strategy viewpoint of profit maximization of a generation company," *Expert Systems with Applications*., Vol. 39 (1), pp. 1565–1574, 2012.

[8] Mathur, S.P.S., Arya, A., Dubey, M., "Optimal bidding strategy for price takers and customers in a competitive electricity market," *Cogent Engineering*., Vol. 4 (1) pp. 1-15, 2017

[9] Y. Liu, H. Ming, X. Luo, L. Hu and Y. Sun., "Timetabling optimization of classrooms and self-study rooms in university teaching buildings based on the building controls virtual test bed platform considering energy efficiency," *Building Simulation*., Vol. 16, pp. 263-277, 2023.

[10] Xu, J., "Improved Genetic Algorithm to Solve the Scheduling Problem of College English Courses," *Complexity*., pp. 1-11, 2021.

[11] Wang, Y.-Z. Wang., "Using genetic algorithm methods to solve course scheduling problems," *Expert Systems with Applications*., Vol. 25 (1), pp. 39-50, 2003.

[12] Arias-Osorio, J., & Mora-Esquivel, A., "A solution to the university course timetabling problem using a hybrid method based on genetic algorithms,", *DYNA*., Vol. 87 (215), pp. 47–56, 2020.

[13] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Singapore, 2006, pp. 205—206.

[14] S. Raschka, *Python machine learning*. 1st ed., Birmingham, UK: Packt Publishing Ltd, 2015.

[15] Srinaul P., *Applying mathematical programming for resource allocation problem to reduce production planning time*. Thammasat University, 2016.

[16] Homsap K., *Applications of linear programming for materials planning in small-medium furniture business*. Thammasat University, 2017.

[17] Frontline Systems. (1 February 2023). *Excel solver – GRG nonlinear solving method stopping conditions* **[Online] Available: https://www.solver.com**