An investigation into the school canteen evaluation and reform

Lingpeng Chen, Minen Lyu, Chengle Zheng and Zijie Xu

***Abstract*—Establishing a sound cafeteria evaluation system is an essential part of school construction. A reasonable cafeteria stall ranking mechanism will help us identify popular cafeteria stalls, and the school can find the right direction in cafeteria reform to improve student satisfaction. However, for The Chinese University of Hong Kong, Shenzhen (CUHKSZ), the current cafeteria evaluation and reform system is not effective enough. Therefore, in this project, we establish a new evaluation model for the CUHKSZ campus cafeteria and rank the quality and satisfaction of all cafeteria stalls based on our model. The implementation of the new model is mainly based on the Bayesian ranking, modified page ranking, and Borda count.**

I. Introduction

Meals are essential to a student's campus life, and having a well-evaluated canteen system is a vital aspect of school infrastructure. CUHKSZ has undergone several canteen reform programs, but students' satisfaction with the cafeteria remains low despite these efforts. Complaints about the canteen are frequent, and many students prefer to eat the off-campus meal. As students of CUHKSZ, we recognize the importance of providing an effective evaluation system to guide future canteen reform programs. To address this issue, we propose a new evaluation system that collects and analyzes students' evaluations of each stall in the cafeteria. We then apply the collected data to a new evaluation model that combines Bayesian ranking, modified page ranking, and Borda count to create a more reasonable canteen improvement strategy. We aim to enhance the overall student experience by involving them in evaluating and reforming the canteen. In this project, we will discuss the details of our proposed evaluation system and how it can create a more student-centric canteen. This new evaluation model will provide a comprehensive understanding of the student's needs and expectations for the cafeteria, leading to more effective canteen reform programs.

II. Research Questions

This study aims to develop a ranking model to reflect the students' favor for each stall, which can contribute to the school canteen evaluation and reform strategy. Therefore, the following questions are asked:

1. What aspects are needed to evaluate a cafeteria stall?

2. How to use data from these aspects to rank all stalls reasonably and comprehensively?

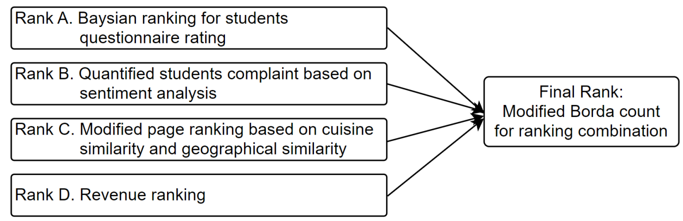

Fig. 1. Pipeline for cafeteria stall ranking algorithm

III. Overview

A. Participants

The questionnaire was completed by undergraduate students from CUHKSZ. In the survey, participants rated each stall (on a scale of 0 to 5), including the closed stalls. Participants also had the option not to rate to indicate they had not eaten in a specific stall.

B. Data Collection

An online survey was conducted from March 13th to March 20th, 2023, using a questionnaire created on the Wenjuanxing app and shared on WeChat. The app automatically collected the results. In addition, data from student complaints and cafeteria revenue were obtained from the Administrative Services Office (ASO). These data sources provide insights into the performance and safety of the cafeteria stalls from different aspects.

C. Model pipeline

Our model comprises five modules (shown in Fig. 1).

1) Bayesian ranking - rank 1: For the data collected by the questionnaire, we use the rating for Bayesian ranking. As the number of people who rated each stall may vary, we use the Bayesian Ranking strategy to determine their ratings. All the rating values were stored in a vector.

2) Complaint Analysis model - rank 2: For the complaints provided by ASO, we propose a Natural Language Processing (NLP) model to analyze complaints about the cafeteria in order to evaluate the quality of food stalls based on students' opinions. We aim to determine the degree of students' dissatisfaction by conducting a sentiment analysis of their complaints. Based on the quantitative complaint analysis, we form a vector to be a part of the stalls' performance evaluation.

3) Modified Page Ranking - rank 3: As some stalls may sell similar cuisine and compete with one another, we build a graph based on both the similarity between the types of stalls sold and the geographic distance between them and then apply the Page Ranking algorithm to determine the importance of each stall. Stalls that are more similar and closer in proximity have a more severe competition relationship and thus have a higher chance of being replaced.

4) Revenue Rating - rank 4: We rank the overall revenue from high to low directly.

5) Modified Borda Count for Ranking Combination - final rank: Based on rank 1, rank 2, rank 3, and rank 4, we set their corresponding weights w1, w2, w3, and w4. After summing up the Borda Counts of the four ranks, we get an overall weighted Borda Count. This Borda Count gives out the final ranking result.

D. Evaluation

As seen above, our ranking model takes into consideration three factors: student opinions (rank 1 and 2), stall information (rank 3), and stall revenue (rank 4). Each rank is assigned a weight, which needs to be determined through evaluation. Our evaluation algorithm is based on survey results. We obtain the participants' favorite and least favorite stalls from each survey sample. For the ranks generated by our model, we assume that our college removed the final 10% of the stalls in the rank. For example, if we have 30 stalls, we will remove three at the end of the rank. We develop a cost function as follows: if a participant's least favorite stall is among these three stalls, the cost function will be reduced by 1, but if a participant's favorite stall is among these three stalls, the cost function will be increased by 2. We aim to minimize the cost function by tuning the weight parameters (w1-w4).

IV. Methodology

Until April 20th, we have partially finished the first four models of our design (i.e., Bayesian Ranking, Complaint Analysis model, Modified Page Ranking, and Revenue Rating). In the following section, we will elaborate on each part of the pipeline (shown in Fig. 1).

A. Bayesian Ranking

For the rating result of our model, we directly used the formula of Bayesian rating in the textbook[1]:

$$r_i = (NR + n_i r_i)/(N + n_i), \tag{1}$$

where i is the index of each stall, N is the total number of reviews with an average rating of R, and n_i is the number of reviews of the ith stall with an average rating of r_i.

The main difficulty of this model in our research was caused by the questionnaire design. In the beginning, we hoped to collect as much data as possible. In this case, for the first version of the questionnaire, all participants were required to rate all 45 stalls. However, the completion rate was low since we designed too many questions in the questionnaire. To solve this problem, we set a minimum number of questions for submission so that participants could end the questionnaire instead of giving up when tired. In the second version of the questionnaire, each user needed to rate all stalls of at least two canteens. In this case, we received 234 questionnaires. Most participants chose to rate the Student Center (SC) 1st-floor canteen (165) followed by SC 2nd-floor canteen (112) and Shaw College canteen (98). Muse College canteen received 84 ratings, while only 46 participants rated for Harmonia College canteen. Then, we calculated the Bayesian rating for each stall.

According to the Bayesian ranking formula, for the stalls with too few raters, Bayesian ranking would drag their scores close to the average. For example, stalls in Harmonia College canteen ranked close to average. Especially the Harmonia Hot Pot, whose initial rating was the lowest among all stalls (2.9), ranked 31st instead of the bottom 10%. To modify the model, we denote the minimum, maximum, and average amount of rating that each stall received (i.e., N_{min}, N_{max}, N_{avg}) as N. Compared with the original Bayesian rating result, the top stall varies, and the least popular is always the SC 1st-floor Korean bibimbap. Concentrating on the ranking of Harmonious College Canteen, we can see the differences between the highest and lowest rankings of the stalls are all larger in these three models than in the original model. To some degree, our modification of the N avoids the stalls of Harmonia College being dragged to the average.

B. Complaint Analysis model

1) Complaint sentiment analysis: To perform sentiment analysis on complaints, we used the Sentiment140 dataset consisting of 1.6 million labeled tweets. We split the dataset into an 80 percent training set and a 20 percent test set. Using unsupervised learning techniques with Word Embedding, we mapped each word in the training dataset to a 300-dimensional vector in a high-dimensional space, where semantically similar words are closer to each other in this space. This allowed the neural network to understand the semantic relationships between words better. So we could convert each sentence into a vector composed of vectors, i.e., a matrix. In order to make the inputs of the neural network consistent, we padded each sentence vector to a length of 300. So we can have a matrix with a (300, 300) shape for each comment sentence. (The first dimension represents the sentence length (each sentence is padded or truncated to 300 words), and the second dimension represents the dimensionality of a word vector, which is 300 in our case).

After the embedding layer, a dropout layer is added to the neural network to prevent overfitting by randomly dropping out some neurons. We then use the Long Short-Term Memory (LSTM) layer. LSTM is a type of recurrent neural network (RNN) architecture that is designed to overcome the issue of vanishing gradients in traditional RNNs. LSTMs are well-suited for processing sequential data, where the relationships between the inputs are meaningful, and they have demonstrated state-of-the-art performance on various NLP tasks. Finally, a fully connected layer is used to map the output of

TABLE I
THE INDEX OF EACH STALL

Index	Canteen	Stall	Index	Canteen	Stall
1	SC 1st floor	Kaifanle (small bowl of dishes)	24	SC 2nd floor	Pork Belly Chicken
2	SC 1st floor	Breakfast (bun shop)	25	SC 2nd floor	Zijin Eight Knife Soup Powder
3	SC 1st floor	Guilin Rice Noodle	26	SC 2nd floor	Hongli Village Rice Roll
4	SC 1st floor	Steak Meal	27	Shaw canteen	Snake
5	SC 1st floor	Korean bibimbap	28	Shaw canteen	Noodles
6	SC 1st floor	Cantonese Siu Mei	29	Shaw canteen	Small Bowl of Dishes
7	SC 1st floor	Chongqing Noodles	30	Shaw canteen	Dumplings
8	SC 1st floor	Spicy Incense Pot	31	Muse canteen	Optional Dish
9	SC 1st floor	Braised Chicken (with) Rice	32	Muse canteen	Cantonese Siu Mei
10	SC 1st floor	Guokui	33	Muse canteen	Noodle
11	SC 1st floor	Lanzhou Beef Lamian	34	Muse canteen	Bibimbap
12	SC 1st floor	Shuyi Tealicious	35	Muse canteen	Teppanyaki
13	SC 1st floor	Fruit	36	Muse canteen	Bar
14	SC 1st floor	Teppanyaki (Cancelled)	37	Harmonia canteen	Bar
15	SC 1st floor	Beef Rice (Cancelled)	38	Harmonia canteen	Bread
16	SC 1st floor	Light Salad(cancelled)	39	Harmonia canteen	Salad
17	SC 1st floor	Rice Noodle (cancelled)	40	Harmonia canteen	Hot Pot
18	SC 2nd floor	Small Bowl of Dishes	41	Harmonia canteen	Fry Chicken
19	SC 2nd floor	Optional Dish	42	Harmonia canteen	Noodle
20	SC 2nd floor	Beef Brisket Noodles	43	Harmonia canteen	Steak Meal
21	SC 2nd floor	Cantonese Siu Mei	44	Harmonia canteen	Optional Dish
22	SC 2nd floor	Spicy Incense Pot(cancelled)	45	Harmonia canteen	Southeast Asian restaurant
23	SC 2nd floor	Water bar (cancelled)			

the LSTM layer to a scalar value representing the degree of sentiment.

Since our dataset only had two labels (positive and negative sentiment), we used binary cross-entropy as the loss function and the Adam optimizer. After training, we found that the model converges at the seventh epoch, achieving an accuracy of 78 percent on the training set and 79 percent on the test set (shown in fig. 2).

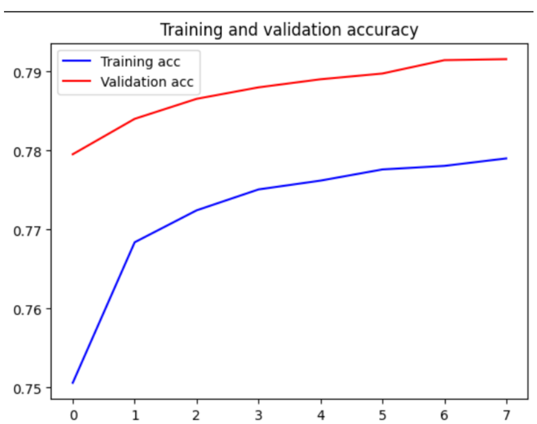

Fig. 2. Training result for sentiment analysis model

The model's output ranges from 0 to 1. We use two thresholds to classify emotions into three categories for experimental results. Models with an output value greater than 0.5 are considered neutral, those between 0.15 and 0.5 are considered negative, and those below 0.15 are considered extremely negative. The predicted results of some comments have been shown in Table II, and the emotional difference can be seen from these comments.

TABLE II
THE COMPLAINT ANALYSIS

Complaint	Categories	Values
Hair found in the dish.	1	0.59
Shouldn't the canteen ensure that there are people serving every minute during operating hours?	2	0.23
f**k!! This is the worst	3	0.017

Category 1 is neutral, 2 is negative and 3 is extremely negative

2) Data analysis for evaluation results: To be continue. The semantic analysis results will be used to form a rank (rank 2) to evaluate the stall performance.

C. Modified Page Ranking

1) Graph for geographical distance: To represent the geographical distance between each stall, we considered whether the stall is on the upper or lower campus and the canteen where the stall locates. On the same campus, stalls have a weak geographical similarity (grey edges, weight = 1), while in the same canteen, stalls have strong geographical similarity (black edges, weight = 2). Due to the close distance between the first and second floors of the Student Center, their stalls have a medium similarity (brown edge, weight = 1.5). To simplify the diagram, we use 1 to 45 to represent 45 different stalls (I). The geographical distance between the stalls is shown in the figure below.

We can calculate each stall's closeness centrality, which represents how close the node is to other nodes. [1]

$$c_i = \frac{n-1}{\sum_j d_{ij}} \tag{2}$$

where i is the index of each stall, n is the total number of stalls and d_ij is the distance between $stall_i$ and $stall_j$. If one stall is closer to others, its closeness centrality c_i is larger, which

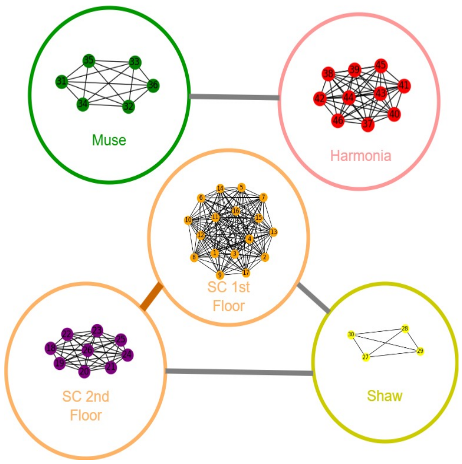

Fig. 3. Graph for geographical similarity

means there is more severe competition and thus more likely to be replaced. Therefore, regarding geographical similarity, we consider the reciprocal of closeness centrality as the score of each stall. The higher the score, the more important this stall is.

$$G_i = \frac{\sum_j d_{ij}}{n-1} \tag{3}$$

where i is the index of each stall, and G_i is the score of $stall_i$'s geographical similarity. Here is the result:
SC 1_{st} floor canteen: G_{1-17} = 1.293
SC 2_{nd} floor canteen: G_{18-26} = 1.431
Shaw canteen: G_{27-30} = 1.897
Muse canteen: G_{31-36} = 1.667
Harmonia canteen: G_{37-45} = 1.400

2) Graph for cuisine similarity: For the similarity graph, we refer to the five layers classification strategy from a patent– "A Dish Recommendation Strategy and Device" [2]. These layers are cuisine category, cooking method, main ingredients, secondary ingredients and seasoning. However, since we are classifying the stalls instead of the dishes, secondary ingredients and seasoning are too detailed to be considered. Therefore, we deleted the last two layers.

In our model, the stalls are divided into clusters layer by layer. The first layer divided stalls into four categories: Chinese food, western food, bibimbap, and water bar. Within each category, we split into different classes. For example, Chinese food contains small bowl of dishes, snack, Cantonese Siu Mei, noodles and water-boiled. Due to the different ingredients, stalls in noodles class are further grouped into sub-classes, including rice noodles, wheaten noodles and dumplings.

According to our classification strategy, the similarity graph is shown in Fig.4 and similarity scales are defined as 2, 1 and 0.5. Connected components are sub-classes and each edge has a similarity scale of 2. Black rectangles with black labels framing connected components represent classes that have a similarity scale of 1. The orange dash lines and orange labels specify the main categories. Stalls within the same main category but different classes have a scale of 0.5.

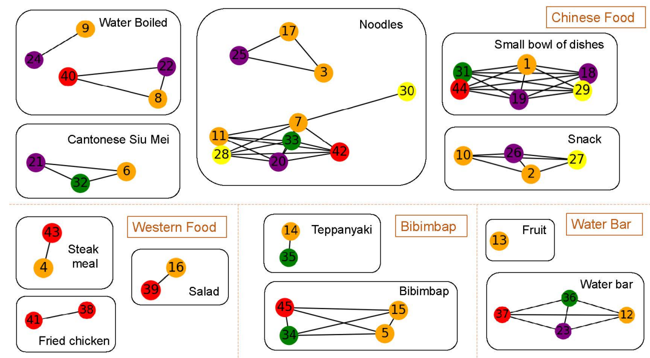

Fig. 4. Graph for cuisine similarity

D. Revenue Rating

Initially, we used the average monthly revenue of each stall to rank. However, different months have different crowd flows. An average value cannot represent the precise popularity of stalls. In addition, the revenue data for some months were missing, leading to differences in the amount of available data for each stall. To address these issues, we plan to use the Netflix algorithm, where we treat each month as a user and each stall as a movie, to better fill in the missing data.

V. Conclusion and Future work

In the past few weeks, we developed a new evaluation model for the school canteen and the first four steps of the model have been finished basically.

In the following weeks, we will finish the rest of the model (data analysis for quantified complaints results, data fusion for page ranking, data filling for revenue rating, modified Borda count for final ranking combination and model evaluation) and figure out the reform strategies for the canteens.

Contribution of Each member

Data collection: Chengle Zheng (40%), Minen Lyu(30%), Lingpeng Chen(20%),and Zijie Xu(10%)

Questionnaire Design and Bayesian Ranking: Lyu Minen (100%)

NLP model and ranking: Lingpeng Chen(100%)

Modified Page Ranking: Chengle Zheng (cuisine similarity) (100%) and Zijie Xu (geographical distance) (100%)

Revenue Rating: Chengle Zheng (100%)

References

[1] M. Chiang, *Networked life: 20 questions and answers*. Cambridge University Press, 2012.

[2] X. Zhou, *A dish recommendation strategy and device*, CN107391531A, 2020. [Online]. Available: https://patents.google.com/patent/CN107391531A/zh.