GROUP ASSIGNMENT

TECHNOLOGY PARK MALAYSIA

CT075-3-2 DTM

DATA MANAGEMENT

APD2F2211CS(DA)/APU2F2211CS(DA)

HAND OUT DATE: 15 DECEMBER 2022
HAND IN DATE: 17 MARCH 2023
WEIGHTAGE: 25%

---

1 **INSTRUCTIONS TO CANDIDATES:**
2 **Submit your assignment at the administrative counter**
3 **Student are advised to underpin their answer with the use of references (cited using the Harvard Name System of Referencing)**
4 **Late submission will be awarded zero (0) unless Extenuating Circumstances (EC) are upheld**
5 **Cases of plagiarism will be penalized**
6 **The assignment should be bound in an appropriate style (comb bound or stapled).**
7 **Where the assignment should be submitted in both hardcopy and softcopy, the softcopy of the written assignment and source code (where appropriate) should be on a CD in an envelope / CD cover and attached to the hardcopy.**
8 **You must obtain 50% overall to pass this module.**

Data mining topic:

Application of Data Mining Techniques in Food Recommendation on platforms/apps

Table of Contents

List of Figures

1 Introduction

Figure 2.1.1 Uber Eats

The Uber Eats app is an online food ordering app that serves as a portal to its restaurant partners all over the world. The food delivery methods in the system include cars, bikes, etc. The system is a three-sided system including restaurants, eaters, and delivery side. This three-sided recommender system is a unique challenge because every side has its own benefits that need to be considered. While some relationships may be straightforward, such as more orders benefiting restaurants and delivery partners by increasing their income, other trade-offs are more complex. For instance, optimizing solely for eater conversion may result in newer restaurants not receiving enough orders to stay on the platform, leading to a decrease in restaurant supply. This can further reduce restaurant diversity, potentially deterring potential eaters from using the platform. Therefore, creating an intricate solution that balances all marketplace aspects is crucial for the success of a three-sided marketplace recommender system.

Uber Eats uses sophisticated machine-learning models and algorithms to generate tailored suggestions for meals and restaurants, improving the customer experience and making it simpler to browse. These models analyze customer data, such as past orders and searches, and use that information to suggest relevant and appealing options.

2 Motivation for using data mining in the area

2.1 Demands for personalized services

The motivation for the food recommendation industry to utilize data mining is that it helps to analyze the data of the customers. Food delivery platform like Uber Eats has massive customer personal data and meal preferences setting within their database, by conducting data mining and analysis techniques to these data, they can have better customer profiling. By comparing their behavior with other similar users, they can make a cross-comparison and recommend possible flavor foods to them, this technique is called collaborative filtering (CR). Furthermore, by using clustering techniques, Uber Eats can help their restaurants to promote their food by using classification through machine learning. As a result, the restaurant in their platform can easily reach their targeted customers and increase their revenue. Hence, every customer using Uber Eats has an interface that shows different kinds of food that comply with their preferences.

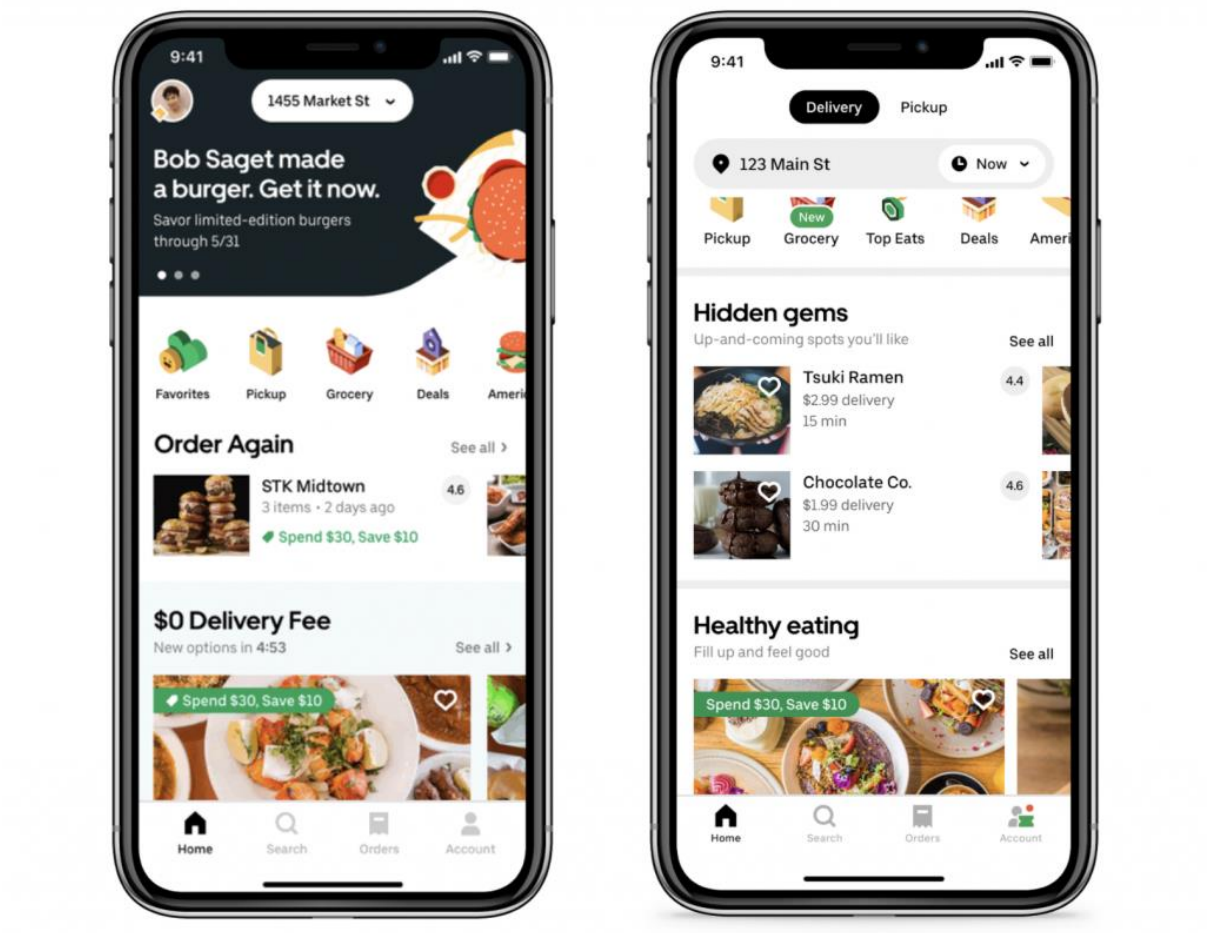

Figure 2.1.1 main carousel of Uber Eats

2.2 Increase platform management efficiency

Besides, data mining applications can increase the efficacy of the Uber Eats platform and their restaurant. Uber Eats can allocate their company resources and determine the policy for their food senders according to the delivery peak hours, the food collection and sending efficacy in terms of restaurants' distance, etc. On the other hand, restaurants, can check on their sales every day to every month, and decide how many ingredients and food they have to prepare every day, re-organize their kitchen to reduce the food preparation duration or involve more staff in the restaurant during peak hours.

In terms of customer retention or preventing their eaters' base from churn, data mining contributes to this topic by using algorithms and data modeling. For illustration, by knowing the behavior of most churn customers, Uber Eats management can easily know the possible reason for customer churn, then provide solutions to get the customer back to their platform like sending them vouchers and making targeted promotions regarding the food they usually order. (Mirus, 2018)

3 Data Mining/Machine Learning Techniques used

3.1 Deployment of Graph Neural Network (GNN)

To make the food recommendation within the Uber Eats platform more accurate and efficient, they establish the graph neural network (GNN). Their machine learning algorithms learn the properties within each of the nodes in the graph. Every node comprises information like customer preferences, order history, favorite food categories, and so on. Then, it calculates the vector for every node to identify the connection and similarity between the nodes. The algorithm will aggregate the similar and high relationship nodes together in the system. The system will recommend the food to each of its customers according to the search depth set, food choice probability, and the properties of neighborhood nodes.

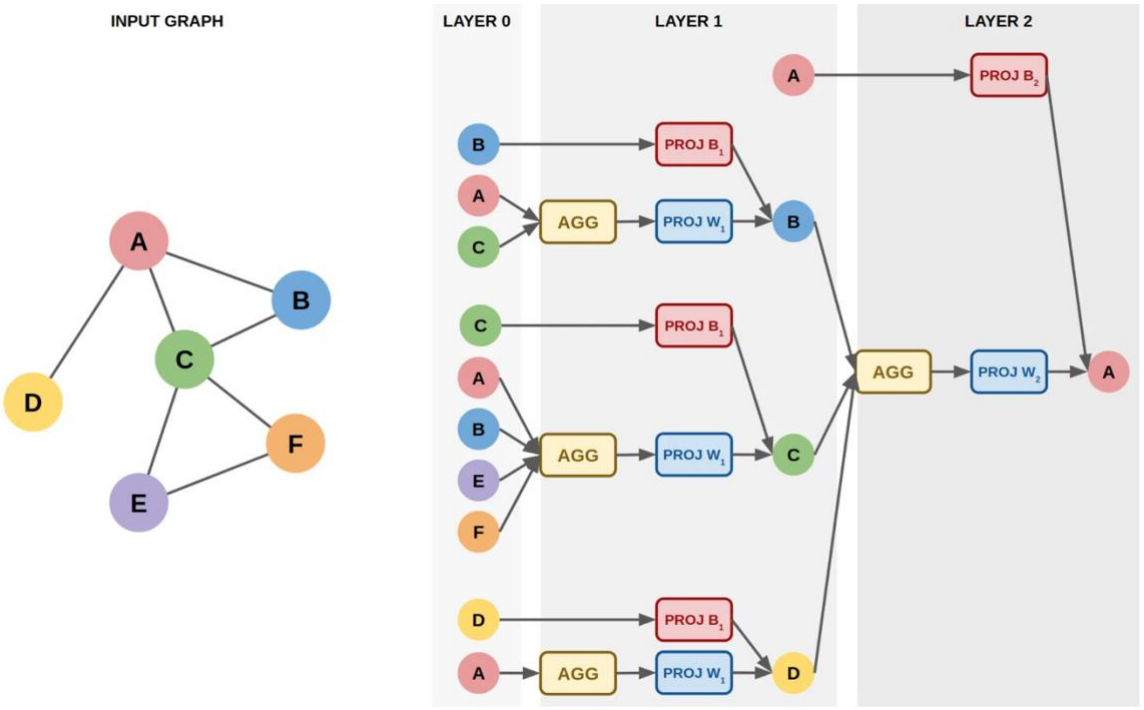

Figure 3.1.1 Graph Neural Network (GNN)

In the main interface, Uber Eats has a recommendation marquee and personalized recommendations. These two dynamic features are determined by the results of the candidate generation and personalized ranking benchmarks.

For the algorithm of benchmark, they use the bipartite graph that has two nodes: the users and dishes information node, and the order history node. Uber Eats uses the GraphSAGE algorithm from GitHub and makes some changes to it. By using an excellent sampling strategy, the system can obtain the information nodes immediately and concatenate two nodes' information together.

Through this algorithm, the candidate generation benchmark will be calculated by filtering the information search scope according to the customer's personal information like location, dish, and restaurant preferences. Meanwhile, the personalized ranker filters the database information by using other relevant information like ordering time and date and customer ordering behavior. These two features have dramatically increased the speed for customers from searching to ordering food.

3.2 Improvement of Algorithms

However, GraphSAGE is only a foundation algorithm on GitHub, Uber Eats keeps improving it by integrating the edge weight calculation of “Low-Rank Positive” into this algorithm so that the edges with lower rank will not be neglected in this algorithm. (Jain et al., 2019)

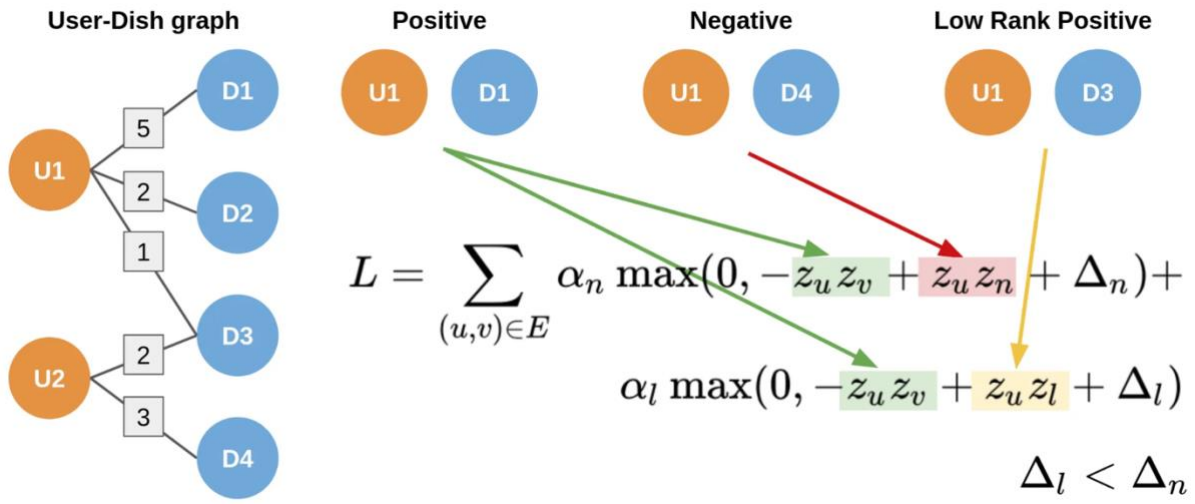

Figure 3.2.1 Improvement on the algorithm of adding "Low-Rank Positive"

For further illustration of the bipartite graph in the Uber Eats platform, it is a variant of the Knowledge Graph (KG). Knowledge Graph (KG) is a cutting-edge technology that enables the system to more conveniently and efficiently handle the labeled data that is provided by real-world factual information (background knowledge) through the form of a graph. These data play a vital role to be processed to become the labeled data sources for machine learning training. Through the semantic enrichment method, it segmented the data into pieces, adding nodes, labels, and edges to itself, helping in finding answers for the user’s query by machine learning and the retrieval path will reuse for the next time. (IBM, 2023)

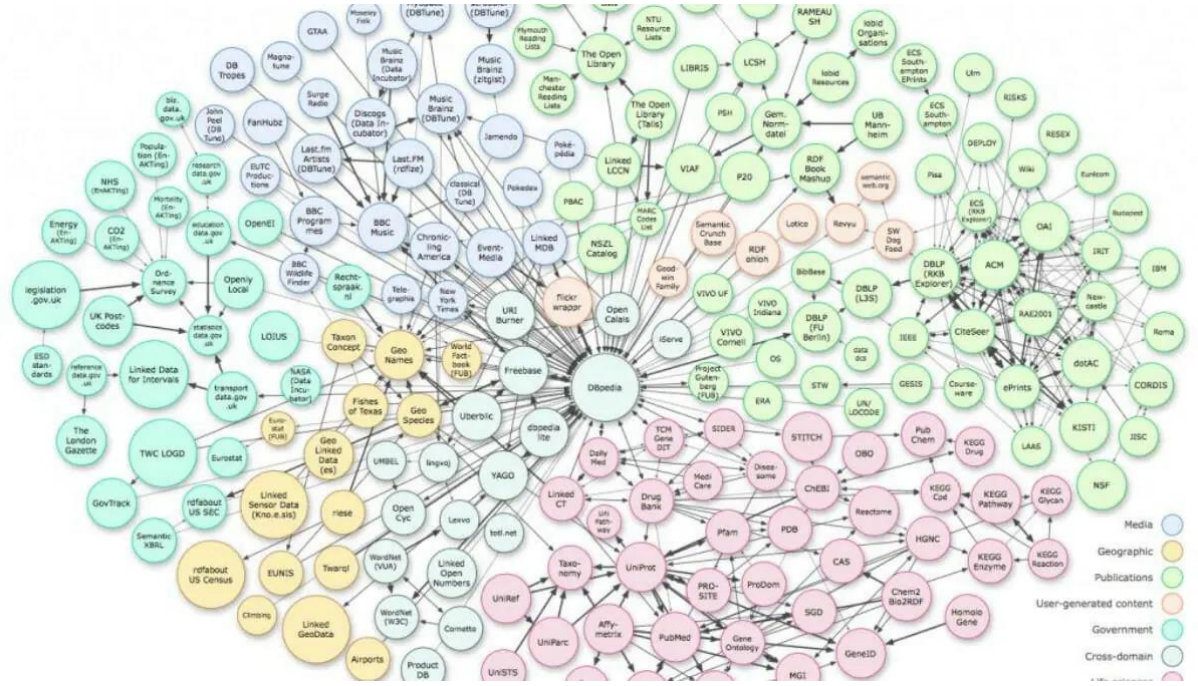

Figure 3.2.2 Knowledge Graph

4 New /prospective opportunities produced by the analysis

4.1 Query Expansion (QE)

Since the system is supposed to have a search function for customers to search for their favorite restaurants and cuisines, query expansion techniques should be applied to best describe the personalized requirements in the original user query.

Query expansion techniques (QE) is one of the information retrieval methods that is widely used in adding more information to the user's query to make the displayed result to be more comprehensive and closer to the user's expected result. Nowadays, QE has been broadly utilized in several areas like emails, medical, social media, etc. Meanwhile, through the research, nature-inspired algorithms including accelerated PSO, and firefly algorithms have been proven to be able to visibly increase the efficiency of QE. (Khennak & Drias, 2018) This can be used in the Uber Eats platform to help the system find the restaurants and foods when the user's input query is ambiguous, and the algorithm could not find the result merely using the original query. However, the result should be aligned with the fact and available in the nearby areas of customers. Applying QE can help customers to find the foods they might never consider before but satisfy their requirements or their expectations. (Hamad et al., 2018)

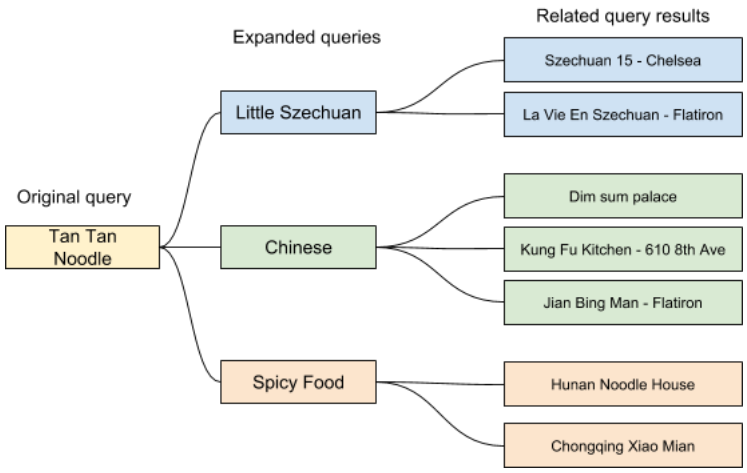

Figure 4.1.1 Query Expansion of "Tan Tan Noodle"

4.2 Nutrition intake monitoring

Nowadays, the fast-paced society has triggered serious imbalanced diets among people as they have no time to deal with their diet and take care of the nutrition intake in their daily life since they are all busy in working. Most of them are relying on convenience food like frozen food, partially prepared food, and fast food. These foods tend to be high in calories, unhealthy fats, and sugar.

Diet unbalancing might lead to severe health conditions and suffering from chronic and cardiovascular diseases. However, converting personal health conditions, eating habits, and routine diet practices to the data in the system is quite challenging. By using food recommendation techniques realized using data mining and analysis, we can guide users to have good dietary habits even if they are ordering food from food delivery platforms like Uber Eats, which food is commonly known as non-healthy food for the public.

In the past, most of the researchers treat the data as a matrix and classification problem, they neglected the special condition that the user might have. For example, the food recommendation has not considered the supplementary information provided by users when they ask for breakfast with bread in their user query. Another point is the recommendation of the system has only followed the users' interested recipe; it has not avoided suggesting food that is not suitable for users like allergic food. Meanwhile, it only suggests the food based on the user preferences extracted from user history regardless of the nutrition of the recipe. Hence, the paper published in 2021 proposes an idea of limited question answering, called Knowledge Base question answering (KBQA), which can eliminate the situation

aforementioned. The system to provide customized consultation using KBQA is called Personalized Food REcommendation via Question answering (pFoodReQ). (Chen et al., 2021)

The methodology proposed in this paper is similar to the search function that is available in Uber Eats. Uber Eats can add the health care domain knowledge to its platform since people nowadays are more emphasized their health and lifespan besides earning money.

The ACM's survey in 2015 reveals that most user nowadays has a high demand for the food recommendation system to enhance their diet quality and make sure they have adequate nutrition routine intake. Even though the suggested diet might not be their preference before, they are also willing to adjust their diet preferences to have a healthier meal. The respondents in the survey are majority requiring an easy-access system or application which can help them in leveraging professional nutrition knowledge. (Ge et al., 2015)

According to research published in 2022, Europe will account for 33% of the growth in the market for health and wellness foods over the projected period. The market in this region is anticipated to grow faster than those in other regions due to consumers' growing preference for organic food over inorganic food products. The UK and France are the two biggest markets in Europe for foods that promote health and well-being. (Technavio, 2022)

On the other hand, a qualified and thorough industry analysis predicts that the Asia-Pacific Health and Wellness Food Market would expand significantly between 2021 and 2028. Functional food, healthy snacks, drinks, fortified and healthy bakery items, BFY foods, chocolate, and other products are the market segments. Due to a rise in the number of health-conscious consumers in the area and a rising demand for nutritious foods, China now has a dominant position in the industry. The study includes a competitive analysis of the key market competitors, such as Danone, PepsiCo, Alter Eco, Nestlé, and Kellogg Co. (Bhat, 2022)

Hence, raising healthy food awareness and changing the stereotype of people towards unhealthy delivery food will be a huge task for Uber Eats and other platforms. If Uber Eats be the first to integrate the nutrition concept into its platform using data mining and analysis, it is expected that Uber Eats will keep growing in the future.

Figure 4.2.1 Forecasting of Global Health and Wellness food market

Figure 4.2.2 Forecasting of the Health and Wellness food market in the Asia Pacific region

4.3 Multi-objective engineering

Eaters in the real world usually have multiple requirements and expectations simultaneously. For example, the algorithm might consider factors such as distance, delivery time, price, and customer reviews when suggesting a restaurant or dish. However, Uber Eats needs to help them to find a tradeoff among all of the competing requirements. Customers need the system to introduce them to the restaurants that they favor, but it will lead to unfairness if the algorithms only introduce those famous restaurants to the customers. Hence, Uber Eats employs machine learning algorithms for multi-objective optimization, which allows the platform to balance multiple criteria when making recommendations to customers. By leveraging data and advanced algorithms, Uber Eats can continuously improve its services and provide personalized recommendations to customers across the globe.

Hence, Uber Eats develops a framework for quadratic programming. The developers weighted the importance of each objective and set the optimization methodology by ranking. These can be deployed using the concept of Lagrangian duality and Karush-Kuhn-Tucker (KKT) conditions from nonlinear programming. Consequently, the objective function and the constraints can be combined into one expression. Then, it can be calculated to a coefficient number and weighted using A/B testing. The different combinations of objectives and constraints will then be deployed and tested online. (Wang et al., 2018)

5 Challenges faced in this area

5.1 Choice diversity

The traditional way in the food recommendation industry is always using algorithms that predict the diner's favorite food users only based on the relevance between nodes. However, it may not always be the ideal method. The relevance model may infer that a user prefers a particular sort of cuisine when they place an order at a specific kind of restaurant and provide them with a list of restaurants that serve that food. The diners might not only or always order particular food, but the relevance-based algorithm only promotes the food they have ordered in the past. But ironically, diners do not like to get so many similar push notifications from the application. Sometimes, they might not even know what kind of food they are interested in. Hence, the conventional algorithms are more probably seem to make them feel boring and annoying. The algorithms should be able to help the eaters to explore their favorite food list, trying new menus and foods.

In summary, the diversity of food and the possibility that a consumer might desire to experience various cuisines should not be ignored. Uber Eats is now keen to create a tailored diversification algorithm that considers both relevance and diversity while making suggestions to solve this issue.

5.2 Restaurant exposure

To maintain the Uber Eats platform market and keep striving to expand its business scope, Uber Eats has to make sure its restaurant partners reach the performance target and find out those underrating restaurants. To make sure all the restaurant partners have a fair exposure chance in the system, they need to apply the new multi-armed bandit (MAB) framework rather than using conventional algorithms.

Due to this, Uber Eats decided to base its algorithm on the upper confidence bound (UCB), one of the techniques in the MAB framework. They may calculate a UCB score for each restaurant based on variables such as past impressions, overall clicks, and boosting factors. A new restaurant will now start out with a high UCB score, helping it to rank well and gain more exposure. The UCB score will eventually drop as it receives more views and clicks, and the algorithm will return its attention to other quantitative factors like relevancy. This approach gives new restaurants a fair chance to gain traction on the platform while ensuring that popular and relevant options remain at the forefront of the user experience.

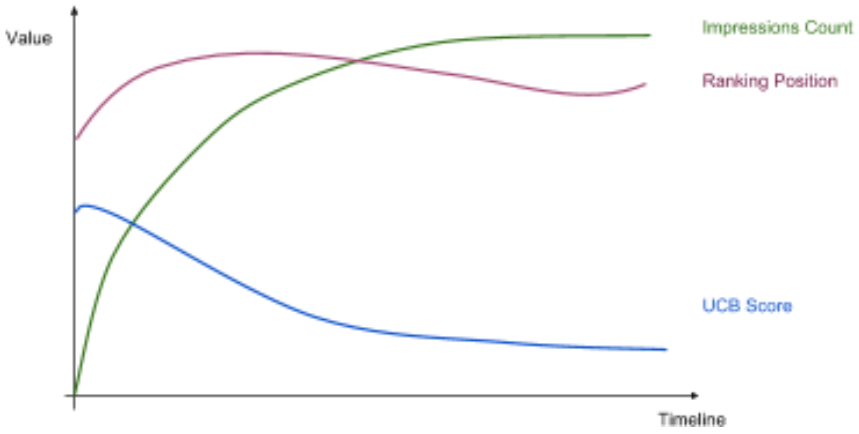

Figure 5.2.1 trend of UCB scoring for the new restaurants

5.3 Exploitation and exploration tradeoff

The exploitation and exploration trade-off are fundamental dilemmas in decision-making, especially in the context of optimization and machine learning.

Exploitation refers to deciding based on the current best-known information, to maximize the expected payoff. This is typically the best approach when the underlying distribution of the data or system is well understood, and the goal is to optimize performance on a specific task.

On the other hand, exploration refers to choosing actions that may not be optimal based on the current knowledge, to gather more information and reduce uncertainty about the system or data. This is often necessary when the distribution of the data or system is not well understood, or when the goal is to discover new insights or improve performance on multiple tasks simultaneously.

The trade-off between exploitation and exploration arises because optimizing for one of them can lead to suboptimal performance with a higher payoff if the underlying data or system changes while over-emphasizing another one can lead to slow progress or wasted effort. Balancing these two approaches is essential in many decision-making scenarios in Uber Eats, including multi-objective optimization and reinforcement learning. (Wang et al., 2018)

Hence, the Uber Eats team applies the Bayesian optimization with a contextual multi-armed bandit (MAB) to balance the trade-off and find the optimal weight combination for tunable coefficients in a multi-objective optimization framework. This involves modeling the payoff of different coefficient values as a probabilistic function and using Bayesian optimization to search for the best values.

They also provide fresh sets of coefficients for training data that are less biased. The application of the epsilon-greedy approach enables a balance between the utilization of the most advantageous coefficients at the moment and the investigation of new coefficients.

6 The future direction of research in this area

6.1 Knowledge Graph Combination

Further enhancement Uber Eats is planning to do is that they try to reinforce the recommendation task in the system together. They need to combine into a bipartite graph together by combining the entities and information into a single knowledge graph. However, this enhancement requires a heavy workload of solving the calculation flaws and aggregation function. Combining the restaurant and dish recommendations can help the system to provide a more complete understanding of a user's preferences and allow for more accurate and relevant recommendations. Furthermore, the system may be able to make better predictions for new items or users that have limited or no historical data if there is only combined KG exists in the system. The overall scalability and efficiency of the recommendation system can reduce the number of separate models that need to be maintained and reduce the complexity of the overall system architecture.

6.2 Meta graph learning

The Uber Eats team always faces a problem in expanding the cities that provides Uber Eats food delivery. When their platform newly comes to a new area, there might be not enough data for accurate recommendations since the prediction of conventional methods is highly dependent on historic data. Hence, the Uber Eats team is now moving forward to the Meta graph learning.

Meta graph learning is a technique that involves using a pre-trained graph embedding model to learn the embeddings of a new graph that has little or no historical data. It is one of the machine learning methods that combine multiple graphs with different structures and types of entities and relationships to generate better recommendations. The pre-trained model captures the common features and relationships of the entities in the graph, which can be used to infer the embeddings of new entities.

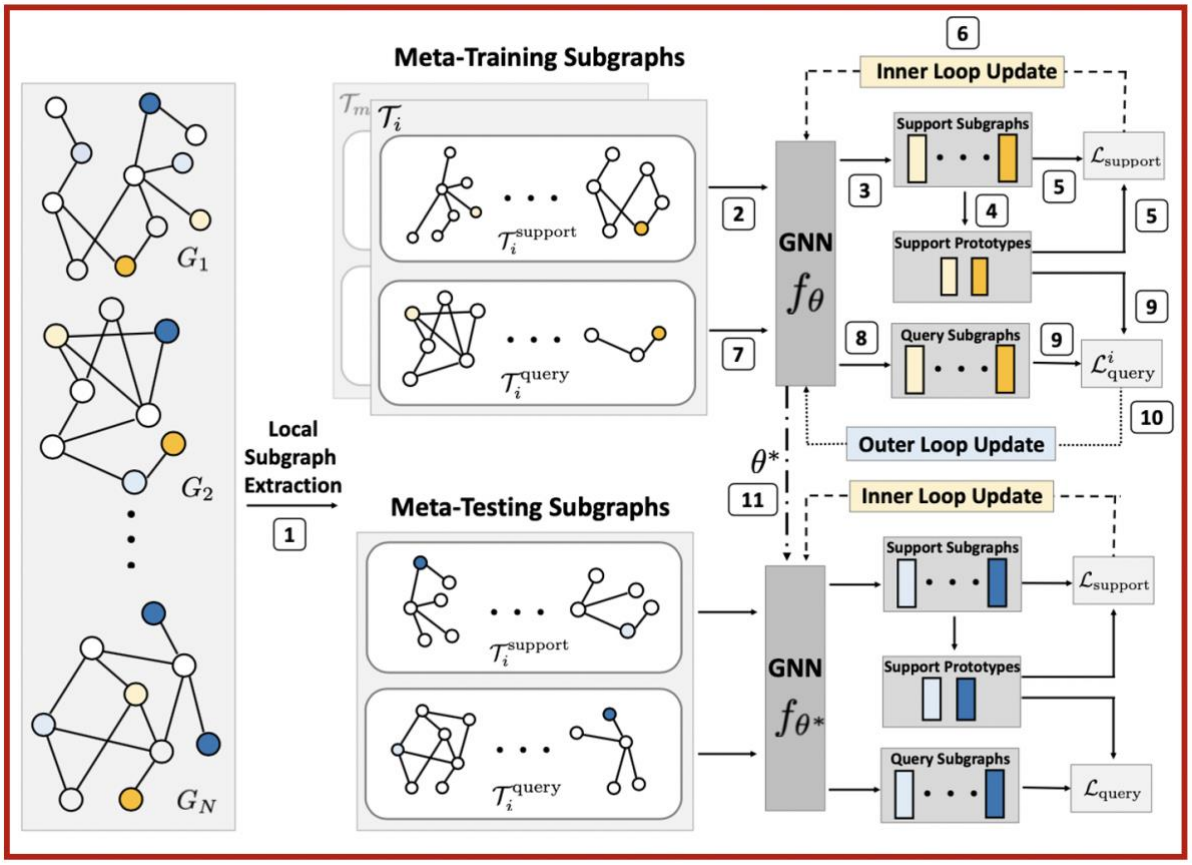

Figure 6.2.1 Meta graph learning

By applying meta graph learning, Uber Eats can provide reasonable recommendations to users even in situations with data scarcity, such as in new cities where there may be limited historical data. The system can leverage data from multiple sources, such as user behavior data from other cities or demographic data, to generate more accurate recommendations for new cities. The meta graph can also be updated over time as more data becomes available, leading to continuous improvements in recommendation quality. This approach has shown encouraging results, suggesting that it has the potential to address the limitations of traditional graph-based recommendation systems.

7 Conclusion

In conclusion, the application of data mining and machine learning techniques has revolutionized the food recommendation industry, as exemplified by the case of Uber Eats. By analyzing customer data and behavior, personalized recommendations can be made to customers, leading to customer satisfaction and increased revenue for restaurants. Algorithms such as Graph Neural Networks and bipartite graphs have been deployed to achieve this, with continuous improvements being made to these algorithms.

Uber Eats has recognized the importance of personalized recommendations and has implemented various machine-learning techniques to improve its food recommendation system and restaurant exposure on its platform. They have developed a tailored diversification algorithm that takes both relevance and diversity into account to provide choice diversity in their recommendations. Furthermore, they have adopted a multi-armed bandit framework with the upper confidence bound methodology to ensure fair exposure for all restaurant partners.

In addition, we have identified three key areas of opportunity for Uber Eats to further improve its platform through data mining and analysis techniques: query expansion, nutrition intake monitoring, and multi-objective engineering. These opportunities can help Uber Eats to provide more personalized and relevant recommendations to its users, promote healthier eating habits, and enhance the overall user experience. By adopting these data-driven approaches, Uber Eats can remain competitive in the fast-growing food delivery market and attract and retain more customers.

Overall, machine learning can be used to improve user experience and optimize business outcomes in food recommendations. The efficiency and customization of the food suggestion business have significantly increased thanks to the use of data mining and machine learning techniques, making it better for customers.

8 References

Bhat, I. (2022, December 27). *Food Market is Surging to Witness Huge Demand at a CAGR of 9.9% during the forecast period 2029*. Data Bridge Market Research. https://www.linkedin.com/pulse/asia-pacific-health-wellness-food-market-surging-witness-indu-bhat/

Chen, Y., Subburathinam, A., Chen, C.-H., & Zaki, M. J. (2021). Personalized Food Recommendation as Constrained Question Answering over a Large-scale Food Knowledge Graph. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. https://doi.org/10.1145/3437963.3441816

Ge, M., Ricci, F., & Massimo, D. (2015). Health-aware Food Recommender System. *Proceedings of the 9th ACM Conference on Recommender Systems*. https://doi.org/10.1145/2792838.2796554

Hamad, F., Liu, I., & Xian Xing, Z. (2018, June 10). *Food Discovery with Uber Eats: Building a Query Understanding Engine | Uber Blog*. Uber Blog. https://www.uber.com/en-MY/blog/uber-eats-query-understanding/?_ga=2.70384022.741903276.1678630163-637696985.1678630163

IBM. (2023). *What is a Knowledge Graph? | IBM*. IBM - United States. https://www.ibm.com/my-en/topics/knowledge-graph

Jain, A., Liu, I., Sarda, A., & Molino, P. (2019, December 4). *Food Discovery with Uber Eats: Using Graph Learning to Power Recommendations | Uber Blog*. Uber Blog. https://www.uber.com/en-MY/blog/uber-eats-graph-learning/

Khennak, I., & Drias, H. (2018). Data mining techniques and nature-inspired algorithms for query expansion. *Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications*. https://doi.org/10.1145/3230905.3234631

Mirus. (2018, May 11). *Why Your Restaurant Business Needs A Data Mining Solution*. Restaurant Data and Reporting Blog; Mirus. https://blog.mirus.com/why-your-restaurant-business-needs-a-data-mining-solution

Technavio. (2022, August 8). *Health and Wellness Food Market size to grow by USD 362.15 Bn, 33% of the market's growth will originate from Europe - Technavio*. PR Newswire: Press Release Distribution, Targeting, Monitoring and Marketing.

https://www.prnewswire.com/news-releases/health-and-wellness-food-market-size-to-grow-by-usd-362-15-bn-33-of-the-markets-growth-will-originate-from-europe---technavio-301600733.html

Wang, Y., Ning, Y., Liu, I., & Zhang, X. (2018, September 10). *Food Discovery with Uber Eats: Recommending for the Marketplace | Uber Blog*. Uber Blog. https://www.uber.com/en-MY/blog/uber-eats-recommending-marketplace/?_ga=2.70384022.741903276.1678630163-637696985.1678630163

WordLift. (2022). *What is a Knowledge Graph? A comprehensive Guide - WordLift Blog*. WordLift Blog. https://wordlift.io/blog/en/entity/knowledge-graph/