The First INFORMS TSL Data-Driven Research Challenge (TSL-Meituan 2024): Background and Data Description

Academia: Lei Zhao[1], Hai Wang[2],
Meituan: Yile Liang[3], Donghui Li[3], Jiuxia Zhao[3], Xuetao Ding[3], Jinghua Hao[3], and Renqing He[3]

[1]Department of Industrial Engineering, Tsinghua University
[2]School of Computing and Information Systems, Singapore Management University
[3]Meituan

This version is updated on February 25, 2024.

Background

In on-demand food delivery (OFD), consumers continuously place orders from various locations. In response, the platform promptly gathers these newly initiated orders, channels them to the respective merchants, and then assigns dedicated couriers for pick-up and delivery within the promised time. The platforms act as intermediaries, linking a multitude of consumers, merchants, and couriers within the ecosystem, and strive to balance gains and losses among them. In this system, consumers expect timely delivery of their food, merchants aim to serve their food fresh to satisfy their customers, couriers seek to deliver enough orders with minimal effort, and the platform aims to operate the service efficiently.

In Meituan's food delivery service, the dispatch system conducts batch order assignments at intervals of either 30 seconds or 1 minute. During each dispatch cycle, the system identifies available couriers for recently placed orders. It then evaluates the Matching Degree (MD) between the orders and their corresponding couriers, which considers factors that include route convenience, overtime risk, and courier willingness to accept the orders. This evaluation process requires extensive computations for Pick-up and Delivery Route Planning (PDRP) to simulate the courier's behavior after accepting orders. Following this, the system resolves a multi-objective Many-orders-to-One-courier Assignment (MOA) problem, matching orders with the most suitable courier to optimize overall MD scores. Furthermore, to create densely packed order structures for efficient deliveries and to find the most suitable courier, at each dispatch moment, the system decides not only who should be assigned the orders but also the best timing for the assignment. This means it may delay assigning some new orders to a later dispatch cycle.

The dispatch system aims for long-term optimality across all dispatch cycles over the day, rather than seeking immediate returns in a single cycle. This approach is in line with the cumulative nature of the key performance measures, such as courier efficiency, on-time delivery rates and couriers' rejection rates.

In this context, couriers often handle multiple delivery tasks simultaneously (including order pick-up and delivery), while also receiving assignments for additional orders. The MD between a courier and new orders is influenced by their existing orders and current routes. Furthermore, couriers have the option to decline or transfer orders, enabling them to devise more comfortable routes and reduce the workload.

Challenges

1. Computational complexity in real time. On the one hand, the MD scores, based on the PDRP outcomes and operational guidelines, are non-additive. Specifically, the MD score for assigning multiple orders together to a courier does not equal the sum of the MD scores if each order were assigned individually to the same courier. Therefore, modeling the MOA problem usually involves intensive MD score calculations, which suffers from combinatorial explosion. On the other hand, the MOA problem itself is categorized as an NP-hard integer programming problem, known for its extremely large search space. Developing online algorithms that perform effectively for the MOA is an exceptionally challenging task. Moreover, the rapid movement of couriers requires that assignment decisions be made within a very short time, e.g., 10 seconds. This tight timeframe is crucial to ensure the synchronization of courier status between the information acquisition phase and the actual assignment moment. Consequently, the platform tends to favor one-order-to-one-courier assignments during each dispatch cycle, a strategy that reduces computational volume and complexity, albeit at the cost of limiting comprehensive order pooling.

2. Full of uncertainty. Uncertainty is an inherent characteristic of the food delivery process. Factors such as supply-demand variability, traffic jams, and extended food preparation times by the merchants significantly affect couriers' behaviors and the service quality of their delivery orders. However, these uncertainties cannot be completely eliminated or accurately predicted by the system. Furthermore, the absence of adequate real-time monitoring of couriers' and merchants' statuses exacerbates the negative impact of these uncertainties. As a result, it is impossible for the system to precisely predict the couriers' delivery processes, such as the exact timing of order pick-up or delivery. Relying on modeling the MD score and making dispatch decisions based on these inaccurate estimations can reduce the effectiveness of these decisions, negatively impacting the experiences of consumers, couriers, and merchants. In addition, the effects of these uncertainties are further intensified by factors such as the couriers' workload and adverse weather conditions.

3. Multi-objective balance. As outlined above, the system is tasked with balancing the diverse objectives of different stakeholders. These objectives need to be appropriately represented in the MD scores. Meanwhile, careful trade-offs among these objectives

are essential to achieve long-term optimality for the entire system and all involved stakeholders.

Sample Questions

1. Improving solution quality vs. computation performance: Investigate algorithms or methods that can optimize the MOA problem with reduced computation overhead. The practical primary objective functions include the incremental distance (i.e., detour) traveled by a courier upon accepting an order, and the delays beyond the promise deliver time of current orders, to assess consumer experience.

2. Objective design improvement in the MOA: Explore ways to refine the MD score model at each dispatch moment. This involves decomposing long-term and global goals into actionable steps at different dispatch moments, aiming to avoid short-sighted decisions at the dispatching moments.

3. Modeling benefits of postponing orders: Explore methods to quantify the advantages of delaying certain orders, particularly in achieving long-term and global optimality, and how this strategy impacts the overall system efficiency.

4. Enhancing courier behavior understanding: Develop models to predict courier delivery processes in uncertain environments, like precise timings for order pick-up and delivery, to improve system responsiveness and decision-making. Derive valuable insights from skilled couriers' behaviors for better order assignments.

5. Balancing diverse objectives in order assignment: Study approaches to balance different objectives, such as courier efficiency and safety, merchant experience, and customer satisfaction, particularly from the perspective of long-term system optimality.

6. New assignment mechanisms to improve consumer experience: Explore innovative order assignment mechanisms that could mitigate the impact of courier rejection and delays on the entire order delivery wave, aiming to enhance customer satisfaction.

7. Estimating spatiotemporal order structure: Explore methods to forecast the spatial and temporal distributions of future orders to guide couriers' real-time locations, thereby improving the balance between supply and demand.

Data Description

The dataset comprises information that describes the process of food order dispatching. These features are categorized into three parts: order, courier, and assignment inputs.

1. Order: geographic coordinates of the pickup and delivery locations, estimated delivery time, and meal preparation time. It also tracks key milestones from the creation to the completion of the order. See Tables 1 and 2 for details (File: all_waybill_info_meituan.csv). For non-existent time data, the corresponding timestamp is represented by 0. Table 1/2 contains 654,343 rows in total.

2. Courier: basic profile and records of a "wave." A wave is a sequence of actions on served orders in which the courier starts with an idle status, accepts orders, and completes all deliveries. See Table 3 for details (File: courier_wave_info_meituan.csv). Table 3 contains 206,748 rows in total.

3. Assignment inputs: orders to be assigned and candidate couriers. At a dispatch time, orders to be assigned can be retrieved from Table 4 (File: dispatch_waybill_meituan.csv). Details of candidate couriers eligible for the orders, e.g., the couriers' current geographical coordinates and the orders they are presently carrying, can be retrieved from Table 5 (File: dispatch_rider_meituan.csv). Table 4 contains 15,921 rows in total; Table 5 contains 62,044 rows in total.

[1] The platform does not use this metric to directly evaluate courier performance. Instead, it sets a distinct and customized performance assessment for couriers, considering the order acceptance time, delivery distance, and associated delivery challenges.

Column	Description	Format	Example	Schema
Order ID	Unique and anonymized identifier of an order	String (Long)	57635	order_id
Waybill ID	Unique and anonymized identifier of a waybill. An order may have several waybill IDs; a waybill is created when an order enters the dispatch system for assignment; a new waybill is created when the order is refused by a courier and re-enters the assignment process	String (Long)	654338	waybill_id
Date	Date of order creation	String	'20231212'	dt
Area ID	Unique and anonymized identifier of the business district	String	12	da_id
Sender latitude	Latitude of the pickup point	Long	39996108	sender_lat
Sender longitude	Longitude of the pickup point	Long	116481447	sender_lng
Recipient latitude	Latitude of the delivery point	Long	39996108	recipient_lat
Recipient longitude	Longitude of the delivery point	Long	116481447	recipient_lng
Sender POI ID	Unique and anonymized identifier of the merchant	String	578	poi_id
Order create time	Unix time when the order is created	Second/Long	1702357786 (2023-12-12 13:09:46)	platform_order_time

Table 1: Order Data.

Column	Description	Format	Example	Schema
Promise deliver time	Unix time when the order is promised to be delivered to the customer[1]	Second/Long	1702357786 (2023-12-12 13:09:46)	estimate_arrived_time
Estimated meal preparation time	Unix time when meal preparation is complete, i.e., courier cannot fetch the order before this time	Second/Long	1702357786 (2023-12-12 13:09:46)	estimate_meal_prepare_time
Order push time	Unix time when the order enters the dispatch system for assignment	Second/Long	1702357786 (2023-12-12 13:09:46)	order_push_time
Waybill dispatch time	Unix time when a waybill is assigned to a courier (courier can refuse the assignment)	Second/Long	1702357786 (2023-12-12 13:09:46)	dispatch_time
Courier ID	Unique and anonymized identifier of the courier to whom a waybill is assigned at the current dispatch time	String/Long	1159	courier_id
Courier location latitude	Latitude of the courier location when the waybill is assigned to him/her	Long	39996108	grab_lat
Courier location longitude	Longitude of the courier location when the waybill is assigned to him/her	Long	116481447	grab_lng
Is waybill grabbed	Whether the waybill is accepted by the courier	Boolean	1	is_courier_grabbed
Waybill grab time	Unix time when the waybill is accepted by the courier	Second/Long	1702357786 (2023-12-12 13:09:46)	grab_time
Waybill fetching time	Unix time when the waybill is fetched by the courier	Second/Long	1702357786 (2023-12-12 13:09:46)	fetch_time
Waybill arrived time	Unix time when the waybill is delivered by the courier	Second/Long	1702357786 (2023-12-12 13:09:46)	arrive_time
Is prebook order	Whether the order is a prebook order	Boolean	1	is_prebook
Is weekend	Whether the order date is weekend	Boolean	1	is_weekend

Table 2: Order Data (Continued).

Column	Description	Format	Example	Schema
Date	Date of a wave	String	'20231212'	dt
Courier ID	Unique and anonymized identifier of a courier	String (Long)	1159	courier_id
Wave ID	Unique and anonymized identifier of the wave	Long	12	wave_id
Wave start time	Unix time of the courier's acceptance of the first order in the wave	Second/Long	1702357786 (2023-12-12 13:09:46)	wave_start _time
Wave end time	Unix time of the courier's delivery of the last order in the wave	Second/Long	1702357786 (2023-12-12 13:09:46)	wave_end _time
Order set	Set of order IDs in the wave	Set	(57635, 57636, 57637)	order_ids

Table 3: Courier Data.

Column	Description	Format	Example	Schema
Date	Date of order creation	String	'20231212'	dt
Dispatch time	Unix time when the order is assigned to a courier	Second/Long	1702357786 (2023-12-12 13:09:46)	dispatch _time
Order ID	Unique and anonymized identifier of an order	String (Long)	57635	order_id

Table 4: Assignment Inputs: Orders to be Assigned.

Column	Description	Format	Example	Schema
Date	Date of order creation	String	'20231212'	dt
Dispatch time	Unix time when the order is assigned to the candidate courier	Second/Long	1702357786 (2023-12-12 13:09:46)	dispatch _time
Candidate courier ID	Unique and anonymized identifier of the order's candidate courier	String (Long)	1159	courier_id
Courier location latitude	Latitude of the courier's location at the dispatch time	Long	39996108	rider_lat
Courier location longitude	Longitude of the courier's location at the dispatch time	Long	116481447	rider_lng
Courier on-hand order set	Set of the courier's on-hand orders at the dispatch time	Set	(57635, 57636, 57637)	courier _waybills

Table 5: Assignment Inputs: Candidate Couriers.