TSL-Meituan Data-Driven Research Challenge: Supplementary Document

Research Challenge Committee

Abstract

This supplementary document provides a few clarifications on the dataset for the research challenge.
Version: October 15, 2024.

1. Introduction

The dataset contains detailed information on the food delivery process. The features are divided into three categories: orders (File: *all_waybill_info_meituan.csv*); couriers (File: *courier_wave_info_meituan.csv*); and assignment inputs (Files: *dispatch_waybill_meituan.csv* and *dispatch_rider_meituan.csv*).

In this supplementary document, we clarify certain data issues regarding the dataset and the "Background and Data Description" document.

2. Data Clarification

2.1. Latitude and Longitude

For privacy protection purposes, all latitude and longitude data in the dataset have been uniformly shifted but not scaled. One may assume that the impact on the spherical distance due to the latitude shift can be neglected.

2.2. Wave Start Time

In the dataset (File: *courier_wave_info_meituan.csv*), each entry represents a wave. A wave is defined as a sequence of actions in which the courier starts in an idle status, accepts orders, completes all deliveries, and returns to the idle status.

Wave start time is the earliest waybill grab time (*grab_time*) in a wave. The *wave_start_time* in the dataset may be problematic due to incorrect indexing. For example, the *wave_start_time* 1665976317 in Table 1 is incorrect. The correct time should be 1665976313, which can be obtained from Table 2.

Table 1: Example Wave of Issue 2.2

courier_id	dt	wave_id	wave_start_time	order_ids
1	20221017	2	1665976317	264834,107524,334962, 125682,493460

Table 2: Grab Time of Each Order in the Example Wave

order_id	waybill_id	grab_time
107524	111084	1665976313
334962	366543	1665976317
493460	559321	1665978002
264834	284850	1665976658
125682	130466	1665976562

2.3. Assignment Inputs

In the dataset (Files: *dispatch_waybill_meituan.csv* and *dispatch_rider_meituan.csv*), the column *dispatch_time* should be interpreted as the checkpoint time for the dispatch decision, instead of the time "when the order is assigned," as in Tables 4 and 5 of the "Background and Data Description" document. At each checkpoint, orders (File: *dispatch_waybill_meituan.csv*) and couriers (File: *dispatch_rider_meituan.csv*) are **candidates** to be assigned, but are not guaranteed to be assigned. In addition, about 1.81% of the entries may not perfectly follow the rule due to the complexity of the actual business.

2.4. Rejected Waybill

In the dataset (File: *all_waybill_info_meituan.csv*), each entry represents a waybill. If a waybill is rejected by the assigned courier, its grab time (*grab_time*), fetch time (*fetch_time*), arrive time (*arrive_time*), and courier location coordinates (*grab_lat* and *grab_lng*) are unavailable and set to 0. See the second row in Table 3 for an example.

Table 3: Example of Issue 2.4

waybill_id	is_courier_grab	grab_time	fetch_time	arrive_time	grab_lat	grab_lng
1	1	1665936737	1665937344	1665937798	45906005	174530062
8	0	0	0	0	0	0

2.5. Estimated Meal Ready Time

In the dataset (File: *all_waybill_info_meituan.csv*), the column *estimate_meal_prepare_time* should be interpreted as the **estimated** meal ready time of the waybill, instead of the time "when meal preparation is complete," as in the "Background and Data Description" document. As a result, there may be waybills in which the fetch time (*fetch_time*) is earlier than the estimated meal ready time (*estimate_meal_prepare_time*).

2.6. Area ID

In the dataset (File: *all_waybill_info_meituan.csv*), the column *da_id* corresponds to "Area ID" in Table 1 in the "Background and Data Description" document. Due to the complexity of the actual business, *da_id* is associated with geographic locations and other factors such as order type. As a result, different *da_id* may have geographically overlapped areas.

2.7. Wave ID

In the dataset (File: *courier_wave_info_meituan.csv*), the combination of date (*dt*), courier ID (*courier_id*), and wave ID (*wave_id*) uniquely identifies a wave conducted by a courier on a day. See Table 4 for an example.

Table 4: Example of Issue 2.7

courier_id	dt	wave_id
0	20221017	1
0	20221017	2
0	20221018	1
0	20221018	2
0	20221018	3
1	20221017	1
1	20221017	2

Acknowledgement

We thank participants in the data-driven research challenge for their feedback. We also thank the research assistants (Kaile Yan, Xiaoyi Wang, Siyuan Yi, Dongyu Zhang, Yinghao Li) from Tsinghua University for their analysis of the dataset.