

物流大数据标准及案例研究

蒋凡

百度外卖研发中心, 北京 100086

摘要

介绍了在城市物流配送领域的数字化和智能化改造过程中,如何面对大数据标准化的挑战,并从调度系统和开放平台两方面入手,提出了提高单次配送效率和节省多次配送成本的标准化解方案。案例为物流大数据标准化技术提供了可供参考的分析思路、实施案例和创新经验。

关键词

城市物流 ;O2O ;大数据 ;标准化

中图分类号 :TP319

文献标识码 :A

doi: 10.11959/j.issn.2096-0271.2017042

Logistics big data standardization and case study

JIANG Fan

Baidu Waimai R&D Center, Beijing 100086, China

Abstract

In the field of logistics distribution, the solution of data and intelligent transformation process was introduced to solve how to face the challenges of big data standardization, and how to improve the single distribution efficiency and save multiple distribution cost standard solution from the scheduling system and open platform. The case provides a reference for the analysis of logistics data standardization technology, the implementation of cases and innovative experience.

Key words

urban logistics, O2O, big data, standardization

1 引言

随着移动互联网技术对国民经济日益深入的影响过程,作为现代经济生活的底层支柱,物流行业也越来越体现出标准化、数据化和智能化的特点。这里标准化又是所有数据化和智能化改造升级的起点,负责对日常海量的物流数据从底层标准上建立统一的规范,进而影响基于大数据的上层模型和算法创新。可以说,标准化的质量和程度,决定了大数据和人工智能所能达到的高度;标准化工作中遇到的问题 and 应对策略,也都融合在现代物流行业在“互联网+”升级浪潮的解决方案之中^[1]。

在国家指导层面上,2015年3月国务院印发了《深化标准化工作改革方案》。方案指出,要“更好发挥标准化在推进国家治理体系和治理能力现代化中的基础性、战略性作用,促进经济持续健康发展和社会全面进步”。2015年11月,国家标准化管理委员会又进一步印发了《物流标准化中长期发展规划(2015—2020年)》,明确了到2020年我国物流标准化工作的发展目标、主要任务、重点领域、重点工程等。

在具体实施层面上,百度外卖自从2014年5月成立以来,专注于城市物流配送领域,在物流行业的数据化和智能化改造方面探索出诸多成果。这其中很大程度上得益于物流标准化建设的工作,为百度外卖物流配送效率和体验提升提供了持久动力^[2]。

2 物流调度系统设计架构

物流调度系统是所有涉及线下运力

调配应用的最核心环节,也是关联诸多因素的综合系统。需要依托海量历史订单数据、配送/驾驶员定位数据、精准的商户特征数据,针对配送/驾驶员实时情景(任务量、配送距离、并单情况、评级),对订单进行智能匹配,实现自动化调度以及资源的全局最优配置,在保证系统效率的前提下,最大限度地提高用户体验。

具体来说,包括以下几部分内容。

(1) 综合交通数据采集平台

综合交通数据采集平台主要提供多源异构数据的采集与清洗功能,解决数据的不一致性问题,有效提高数据质量评价与数据清洗技术,主要功能包括以下几个方面:

- 根据业务范围及应用类型提供数据采集中间件与信息采集工具;
- 实现连续、实时、高效、准确的交通运输领域中结构化与非结构化大数据的获取与采集;
- 结合物联网、互联网和卫星定位技术开发智能终端,实现在动态环境下的交通数据实时采集;
- 实现数据清洗和预处理功能,解决数据冗余和不相容性并消除噪声的数据集成与融合技术。

(2) 物流运输调度决策平台

物流运输调度决策平台主要提供实时数据环境下物流调度规划、下发、反馈、评估和追踪定位技术,主要功能包括以下几个方面:

- 采集并计算所在地区道路、桥梁、河流、障碍物等地理经纬度和距离信息;
- 实时采集点对点物流运输需求数据和物流配送人员的动态信息;
- 进行物流运输路径规划和配送任务指派的决策判断;
- 结合应用场景,线下模拟真实物流运输的多维指标及演进状态;

- 设计并实现分布式、高并发、大容量的流式计算框架；

- 采集并存储所有调度参数信息，提供历史调度场景的追查定位服务。

(3) 物流运力预警及分配平台

物流运力预警及分配平台主要提供跨平台数据环境下物流运力警报的评估、发布和关闭以及相应级别下分配物流运力的技术，主要功能包括以下几个方面：

- 多维度采集公共平台和行业领域综合交通业务数据，比如天气、节假日、重大社会活动、运力资源等；

- 实时采集目标区域内的运力指标数据，评判行业运力紧张状况，预估时间、空间、持续范围及程度；

- 结合智能终端设备，下发并适时调整运力警报分级，通报实时状态；

- 针对不同分级警报状态，有差异化地采取降级应对措施，提供运力灾备和最低限度服务能力。

3 物流配送标准化的必要性

结合百度外卖的业务实践经验发现，标准化对于物流配送的必要性，或者说能实际发挥重要作用的地方主要有两点：

- 精准描述用户的运力需求和平台的配送能力，便于抽象到模型层面，实现算法调度和行程规划，提高物流配送效率；

- 消除信息差异带来的平台扩充成本，适应现实环境中复杂多变的配送场景，便于统一接入调度系统，灵活扩展到更多物流行业，满足更多配送需求。

因此，在物流配送标准化建设中，要注意从提高单次配送效率和节省多次配送成本两方面着重设计标准化的解决方案，做到有的放矢。具体来说，可以从标

准化在调度系统和开放平台的应用中观察到更多细节。

4 标准化在调度系统中的应用

标准化在调度系统中主要是解决低效率的传统外卖配送形式和行程规划问题，将原先需要依赖相关人员经验数值和手工处理的流程改造成可大规模自动处理的标准化数据和可灵活配置复用的标准化流程。

(1) 订单样本数据标准化

调度模型机器学习算法需要千万级的训练样本，告诉学习算法可以从中抽取哪些特征值，才能经过数学模型求出最优参数^[3]。从百度外卖平台的历史订单数据中抽取实例不难，难点在于如何认定这些订单的实际出餐时间、骑士到店时间和用户收餐时间。由于现实网络环境限制和实际操作规范管理难度，并不是采集到的原始数据就可以直接利用，通常需要利用骑士轨迹多次回传记录，并结合地理形态和道路交通属性对缺失数据和异常数据进行预测和校验，才能清洗掉由于商户和骑士操作不规范导致的干扰数据。

(2) 商户和菜品数据标准化

对于餐饮外卖平台来说，商户和菜品信息标准化是提供优质服务体验的基础，也是数据标准化的另外一个重要内容。通常有关商户和菜品的信息都是由餐饮运营方提供，平台负责录入和存储。但这样采集到的数据并不规范，每家餐厅对经营类别、菜品名称的定义标准都各不相同，需要平台方提供一整套的商户和菜品的标准化体系。比如，在经营类别方面，首先制定出17个大类、110个小类的分级体系，并为每个商户明确其对应的经营范围标签；其次对数以百万计的原始菜品名称进行去重、

合并,得到归一化后的标准化菜名,对应到菜品数据库中唯一的标识符;在此基础上,针对每个菜品建立菜式、菜系、口味、食材、做法等各类标签,并进一步得到商户在这些菜品标签上的分布情况。

构建全面的商户和菜品标签体系可以应用在很多方面,比如,通过商户的品牌、营业面积等信息可以估算该商户的产能,通过菜品在材料、价格、烹饪方法上的差异可以估算制作时间。

(3) 虚拟场景构建标准化

虚拟调度系统用来线下模拟物流配送的真实场景及变化过程,以商圈为粒度构建由骑士、用户、配送事件、场景参数等虚拟数据组成的虚拟场景,达到线下演算、优化调度算法系统的效果。标准化的虚拟场景不仅是对静态数据提供可量化、可复用的规范,而且还需要定义一系列的虚拟物流配送动作,才能在线下完成真实场景的变化过程。

不管接入什么样的受测调度算法,设置什么样的初始状态参数,虚拟调度系统都能按照预设的物流运力和调度压力运转完整的多轮配送流程,并统计得出受测调度算法的整体效果指标。比如,在虚拟调度系统里,一次完整的调度配送涉及用户、骑士和调度员3个角色,每次配送行为可以包括下单、取餐、送餐、上岗和下岗等操作,监控和评估环节需要统计的指标包括配送时间、配送距离、空驶距离、准时率、并单率、压单率等。当启动虚拟调度系统以后,同时有多个并发进程在以某个时钟频率实时更新虚拟场景状态,并触发某个时段后各个角色的新行为。

5 标准化在开放平台中的应用

标准化在开放平台中主要是解决复杂

多变的外部数据结构和业务需求带来的成本问题,将原先需要一对一制定方案的封闭式合作模式改造成一对多复用方案的开放式合作模式。

不管是接入不同商户自有的订单管理系统,还是接入不同物流公司的配送管理系统,开放平台负责提供标准化的商户订单生成、制作、取货、配送的全过程字段定义以及标准化的能对接多套物流公司的元调度接口。这一整套物流标准化解方案包含了内外部系统的交互数据标准(比如订单格式)、交互操作标准(因为存在需要协同合作的多个机构或者系统)以及对外部合作方的服务要求标准等。

双方通过超文本传输协议(hypertext transfer protocol, HTTP)方式交互数据,第三方平台通过POST方式发送提交内容,内容参数为JSON串。双方需要保证数据传输的完整性和安全性,每次发送请求都有响应(响应返回格式为纯文本),安全验证目前采用数据签名的方式。

接口采用静态KEY+参数方式进行签名,以防止数据在过程中被篡改,参数值为非数字、字符串的参数不参与签名。签名的具体步骤如下。

步骤1 将请求字段按照字典顺序进行排序,将字段key与value用“=”连接,字段之间用“&”相连,例如out_order_id=12121& immediate_deliver=1&...

步骤2 在步骤1中拼接好的字符串结尾拼“&APPID&APPKEY”;

步骤3 对步骤2的结果进行sha1方式签名,获得签名sign;

步骤4 将签名参数sign=”3da541559918a808c2402bba5012f6c60b27661c”,放入请求字段中。

对于第三方平台新订单的格式要求为:请求URL: {SERVER_HOST}/api/createorder。

请求参数则包括了订单号、用户联系电话号、用户经纬度坐标、用户配送地址、期望送达时间等字段。具体部分字段的示例见表1。

常用的交互操作包括取消订单（请求URL：`{SERVER_HOST}/api/cancelorder`）和批量查询订单（请求URL：`{SERVER_HOST}/api/listordermulti`）。

对于骑士的配送服务要求则分别定义了接单取餐、送餐和完成订单3类标准接口。以接单取餐为例可以看到请求字段为：

```
{
  "order_id": 100001,
  "confirm_time": 1439544470,
  "sign": "jsadieuddye73625wrstdedfcvfgbh096lhogit6",
  "delivery_name": "John",
  "delivery_phone": "13288909981",
  "push_time": 1428374657
}
```

返回字段为：

```
{
  "error_no": 0,
  "message": "success",
}
```

表1 请求参数部分字段示例

参数字段	类型	必填	描述
out_order_id	字符串类型(string)	是	商户订单号, 不允许重复
order_index	整数类型(int)	是	该店今天该平台的第几号订单
user_phone	string	是	用户手机号
user_address	string	是	用户配送地址
user_name	string	是	用户姓名
user_longitude	string	是	用户经度坐标, 请使用百度坐标(BD-09)保留6位小数
user_latitude	string	是	用户纬度坐标, 请使用百度坐标(BD-09)保留6位小数

6 标准化能力通用性及效果评估

以上标准化能力已经被证明在物流配送行业具有较强的通用性。

- 在百度外卖自有物流业务中得到应用, 通过大规模地在百度物流配送业务运用智能化调度系统, 调度效率提升明显, 配送准时率达98.78%, 平均送达时长为32 min。

- 完成对多家品牌商户订单管理系统的对接, 比如: 和合谷、正一味、眉州东坡酒楼, 每天承接这些入口商户带来的物流订单, 无缝纳入百度物流的运力资源。

- 完成对众包物流运力的服务对接, 每天发布高峰时期的峰值订单, 由富余的众包运力负责配送。

- 完成能兼容多家物流配送公司的元调度接口测试, 比如: 许鲜网、多点网, 混合调度多家异质运力资源, 做到兼容并包, 统一调配, 实现了错峰时段的运力最大化利用。

7 对城市物流领域典型问题的贡献

百度外卖在物流标准化建设的工作为物流配送效率和体验提升提供了持久动

力,也对城市物流领域在提升效率、评估效果和降低成本等典型问题上探索出了有益经验,做出了相应贡献。

(1) 提高调度系统机器学习的优化效果

综合利用各种地理定位数据和校验手段,获得精准的标准化订单样本数据。结合餐饮领域特点构建全面的商户和菜品标签体系,构建标准化的分层商户经营范围和多维度菜品数据库。

在此基础上,才能够利用机器学习方法从千万量级的训练样本训练产出预测误差在7 min以内的出餐时间预估,才能够实现多目标优化规划算法将原先40多分钟的平均配送时长缩短到32 min,极大改变了整个城市短途物流配送行业的生态。

(2) 建立模拟系统线下效果的评估能力

通过离线的标准化处理和在线的异步处理模式,实现精准推演。通过模拟真实约束情况将多变的场景变量(比如商圈、天气、整体运力)在高层作为调优参数进行优化,尽可能地模拟多维限制条件下的真实约束情况。模拟系统会分商圈、分时间地统计每个调度场景下的订单分布数据,解析成骑士在岗率、平均压单数、订单单位位置密度等参数的基础物理分布函数,并作为刻画该调度场景的约束条件组。

模拟系统还可以进一步调整这些分布函数的参数,得到人工设定的约束条件组,从而模拟出更复杂丰富的设定场景。以这些与动态场景相关的参数组作为调度算法的输入约束条件,调用模拟系统,反复推演虚拟订单的分配过程,通过梯度下降优化算法,求解出多目标下的最优解。

(3) 实现物流平台调度能力的横向输出

受限于业务场景的复杂性,各个不同配送场景下的运力难以互相调剂,表现为:

时段、空间上的运力不均匀现象。如果能够将这些孤立的运力资源区间打通,将有效地发挥城市物流运力的作用,提高运力使用效率。

对物流调度的相关元素做标准化处理,抽象出更基本的元调度接口,极大降低了各种异质运力的接入成本,实现城市内自由运力的互联互通。通用的元调度(meta dispatch)将所有的调度行为抽象成最核心的操作和数据结构,对外提供简易可行的通用接口^[4]。

8 结束语

通过在外卖配送行业的实践可以发现,遵循国家层面在标准化建设的指导意见,适应并积极推动移动互联网技术为物流配送领域带来的挑战和变革,解决一线业务实际需求面临的大数据标准化问题,就一定能很好地解决大数据技术在应用层面遇到的工程问题,并为其他人工智能创新提供支持。百度外卖物流在这一领域开展了一些有益尝试,并将在标准化方案研制和应用推广方面持续深入进行下去。尤其是在物流运输需求预测及供需调配方面,有必要进一步提供大数据环境下城市物流运输的预测需求、调配供给和平衡供需等技术,未来可能涉及的研究方向包括:基于历史交易和消费信息,采集并存储特定行业领域的物流运输需求大数据;利用数据挖掘和机器学习方法,分析物流运输需求在时间、空间和人群上的分布规律和性质特点;动态调配可供运力资源在人员数量、工具类型、投放能力等指标上的配比,以适应需求变化;主动采用动态定价、划分服务范围、人群画像和推送消息等技术手段,引导运输需求发生变化,使之趋于均匀有效分布。

参考文献:

- [1] 华岗, 顾德道, 刘良华, 等. 城市大数据: 内涵、服务架构与实施路径[J]. 大数据, 2016, 2(3): 2016026.
HUA G, GU D D, LIU L H, et al. Urban big data: connotation, service architecture and implementation path[J]. Big Data Research, 2016, 2(3): 2016026.
- [2] 蒋凡, 徐明泉, 崔代锐. 基于外卖物流配送大数据的调度系统[J]. 大数据, 2017, 3(1): 104-110.
JIANG F, XU M Q, CUI D R. Scheduling system based on takeaway logistics big data[J]. Big Data Research, 2017, 3(1): 104-110.
- [3] 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述[J]. 计算机应用研究, 2012, 29(8): 2806-2810.
SUN Z J, XUE L, XU Y M, et al. Overview of deep learning[J]. Application Research of Computers, 2012, 29(8): 2806-2810.
- [4] 蒋凡. 综合交通大数据应用技术国家工程实验室申请报告[R]. 北京: 百度外卖物流调度平台, 2016.
JIANG F. Application report for national engineering laboratory for integrated transport big data application technology[R]. Beijing: Baidu Waimai Logistics Scheduling Platform, 2016.

作者简介



蒋凡(1979-), 男, 百度外卖研发中心技术委员会主席、主任架构师, 中国计算机学会专业会员, 主要研究方向为物流调度、推荐系统、数据挖掘。

收稿日期: 2017-01-17