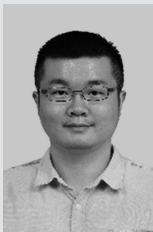


# 基于外卖物流配送 大数据的调度系统

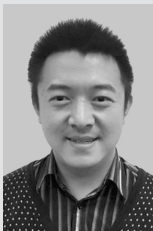
*Scheduling system based on takeaway  
logistics big data*



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



徐明泉 (1988-), 男, 百度外卖研发中心架构师, 主要研究方向为物流调度、数据挖掘。



崔代锐 (1985-), 男, 百度外卖研发中心技术总监, 主要研究方向为商业智能、大数据分析。

中图分类号: TP399

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2017013

## 1 引言

餐饮O2O行业连接线上线下的订餐需求,将传统的到店消费模式改造成更加灵活便捷的到家消费模式,极大降低了用户的消费成本和商户的固定成本。外卖平台除了促进线上需求向线下转化,也包括达成订单的最后一公里任务——基于城市道路交通状况的物流配送<sup>[1]</sup>。高效的物流配送能力是决定外卖平台商业模式成败的关键,也是O2O经济区别于传统经济的根本,即运用城市交通大数据相关的云计算、深度学习和可视化技术提升行业效率,创造并满足新的民生需求<sup>[2]</sup>。

因此,智能调度系统是外卖物流的核心环节,依托海量历史订单数据、骑士(送餐员)定位数据、精准的商户特征数据,针对骑士实时情景(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。

然而在外卖物流调度的真实场景中,用户点了餐就希望能按时送到,骑士上了路就希望每趟路线能多配送几单,商家接了餐就希望骑士快来取餐,平台则关心如何以最小的运力承接最大的配送压力,而且能扛住高峰时段突如其来的订单量。更加困难的是,这些目标有时就是互相矛盾的,满足了一方,势必会影响另一方,调度订单是非常复杂的多目标动态规划的决策过程。

智能调度系统需要将以上所有因素都考虑在内,实时采集整个商圈里各方的动态数据,在1 ms内做出在时间跨度和空间范围内的最优分配序列,让骑士轨迹能无缝衔接起整个配送流程,让每个环节耗费

的时间降到最低,让分摊到有限运力上的配送成本费用降到最低。

实践证明,在智能物流调度系统实施之前,物流订单需要由调度员手工分配,每个骑士每天最多配送10单,每单配送时长超过1 h。实施之后,百度物流调度每天所有订单都由算法自动选择最优化的方案调配,每个骑士的配送效率翻倍、收入翻数倍,每单配送时长节省50%以上,覆盖全国100多个城市。

## 2 物流调度的云端计算模式

影响订单分配的因素很多,从订单生成那一刻开始,调度系统就要考虑到订单的取餐地址、用户的配送地址、商圈内的骑士数量和状态、订单的预期送达时间。每一个订单并不是孤立存在的,要想得到全局最优的配送方案,还必须考虑到这一时段内其他订单的配送情况,尽可能做合并,提高整体的配送效率。如果再考虑到不同城市、商圈、天气、节假日、工作日和商圈骑士运力配置等情况,事情就变得更加复杂。

这是一个极其复杂的多目标动态优化问题<sup>[3]</sup>,可以表示为:

$$\min(f_1(x), f_2(x), \dots, f_n(x)), \text{s.t. } g(x) \leq 0 \quad (1)$$

其中,假设 $f_n(x)$ 代表单个指标的优化目标,举例如下。

$f_1(x)$ : 即时订单的配送时间越少越好。

$f_2(x)$ : 预约订单的送达时间距离预约时间点越短越好。

$f_3(x)$ : 每单的配送距离越小越好。

$f_4(x)$ : 骑士全天的总配送单量越多越好。

$f_5(x)$ : 订单被并联配送的比例越多越好。

$f_6(x)$ : 骑士空驶距离越少越好。

$g(x)$ 代表业务限制的约束条件:

- 不能超出预计配送时间30 min;

- 单个骑士不能同时配送若干单以上；
- 同一商圈同一时段骑士之间未完成单量不能相差若干单以上；
- 新骑士必须每天保证能被分配到若干单以上。

要满足这些优化目标和约束条件的直接计算复杂度太高,需要采取逐层建模的方式来降低复杂度,其基本逻辑具体如下。

#### (1) 降维代价函数

将通用的参数变量,比如时间、距离,在底层作为基本限制条件进行采集和转换,尽可能地对代价函数进行降维,拟合成线性多项式函数,减少计算复杂度。

假设在时刻 $T$ 有一笔订单 $O$ 要分配给候选的骑士 $R_1, R_2, \dots, R_n$ ,需要分别预估若由骑士 $R_x$ 配送这笔订单的实际开销,比如:骑士的到店时间、等餐时间、送餐时间、交付时间、配送里程,计算式可以表示为:

$$f(x) = k_1 \times f_1(x) + k_2 \times f_2(x) + \dots + k_n \times f_n(x) \quad (2)$$

计算出单次分配的综合代价打分,其中每项参数 $k$ 根据经验给出初始设置值,后续在模拟系统中经过迭代优化调整。

#### (2) 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、订单出单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。

以与这些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) 求解最优解

在真实场景中,基于单次分配代价函数,采用二分图最大匹配算法求解多次分配下的最优解,尽可能地找出多个骑士和多笔订单间的最优组合,提高并单率或骑士人效,减少骑士空驶率<sup>[4]</sup>。

假设可分配骑士为 $R_1, R_2, \dots, R_n$ ,待分配订单为 $O_1, O_2, \dots, O_n$ ,分别计算每一组分配 $\langle R_x, O_x \rangle$ 的代价得分,找出全局最优的骑士—订单分配组合,使得总代价最低。问题可以转化成传统的完备匹配下的最大权匹配问题:在一个二分图内,左边的骑士节点为 $X$ ,右边的订单节点为 $Y$ ,对于每组左右连接 $X_i Y_j$ 有权重 $W_{ij}$ ,即配送代价得分,求一种匹配使得所有 $W_{ij}$ 的和最大,即总配送代价得分最小。

#### (4) 调控虚拟调度订单转为真实派单的节奏

将待分配订单存储在云端服务器,维护订单分配的虚拟队列,通过设定压单时间窗口来调控虚拟调度订单转换为真实派单的节奏。

云分配算法同时维护实际分配订单队列(下发到骑士客户端,不可更改)、虚拟分配订单队列(对应虚拟分配的骑士,可以随时更改)和待分配订单队列(未关闭压单时间窗口或未找到合适分配时机)。这样可以避免订单过早被分配给骑士后,因为情况发生变化,比如某个订单的延误导致后续配送代价函数估算失效,原先最优的分配不再成立。

最终,为了衡量调度系统优化效果,需要将系统配送效率和用户配送体验结合起来,统计在存在并单的情况下,系统为了完成订单配送实际耗费的时间成本,这样才能反映调度系统的整体效果。例如并单配送的2单,40 min送到,单均成本则是 $40/2=20$  min,比只送1单提升了1倍效率。

但如果并单效率不高,过于分离的2单被并联在一起,第1单30 min送到,第2单50 min送到,虽然节省了部分配送路程,但迟送到的那单会拖慢整个订单分组的单均成本,单均成本是 $50/2=25$  min。

按照这个计算方法统计一下调度系统上线前后的效果,如图1所示,可以看到以有效时间计算的单均成本从40多分钟下降到20多分钟,单均成本得到明显改善。

以上复杂的调度规划、海量的数据建模和实时的平台响应,都需要有强大的计算能力在背后支撑。为了得到最合适的订单调配,指派时机可能有多个,每个指派时机上各种可能的组合也会随着商圈订单规模的增大而呈指数级增长。为了尽可能地选择最优指派时机,系统还可能会做预测,尝试各种分支副本情况下的优化空间。如此一来,后台系统需要的计算能力就会非常高。总之,物理上计算能力的容量决定了系统的最终效果。

为此,依托百度云计算积累下来的技术优势,并结合百度外卖的实际应用场景,设计了分布式、高并发、大容量的流式计算框架,尽可能将独立的计算任务拆分到不同的机群上运行,得到整体最优的计算效果。

经测算,目前的计算框架完全可以支撑百度外卖未来每天千万级订单、秒级10亿次计算的动态调度、数据建模和实时平台响应的运算规模。

### 3 深度学习智能预测算法

百度外卖每天百万级的订单量为大数据深度分析技术提供了理想的应用场景。具体来说,调度算法的最终效果除了云端的规划计算能力之外,非常依赖每一个基础变量预估值的准确性,尤其是在不同环节耗费的时间。

相比配送环节而言,商户出餐环节的耗时更长,也更不可控。菜品类型、价格、数量,商户品牌、规模、堂客食比,下单日期、时段都会影响商户出餐时间。如果估算不准,骑士到店时间过早则会白白浪费运力,过晚又会延误本可以早点取餐的订单,还会损伤一些菜品的品相和口感。

为了解决这个难题,采用深度学习方法,让系统自动学习利用并组合产生新特征的方法,借助深层次神经网络模型,能够更加智能地利用不同层次的特征,对样本数据中蕴含的规律进行更加准确、有效的表达<sup>[5]</sup>。

具体来说,在以下3个方面结合餐饮O2O领域特点做出了开创性贡献。

- 从百度外卖平台的历史订单数据中抽取千万级训练样本,其中难点在于如何认定这些订单的实际出餐时间,需要借助骑士轨迹、停驻坐标等数据清洗出由于商户和骑士操作不规范导致的干扰数据。

- 构建全面的商户和菜品标签体系,获取商户在竞品和本品的运营信息,作为系统的输入特征。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。

- 系统对不同出餐时长区间的菜品的预估时间要求误差是不同的,处于中段区间的菜品,对预估时间的要求误差最严,为

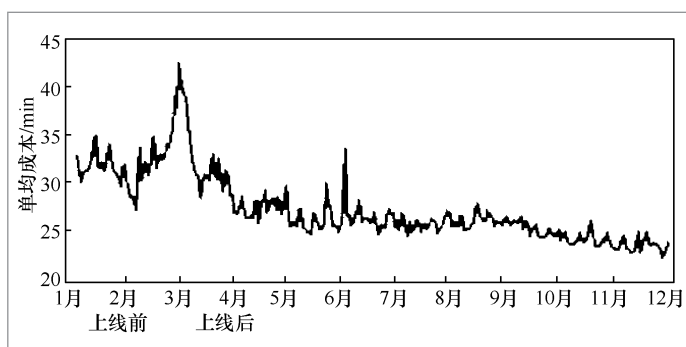


图1 调度系统上线前后的单均成本的变化

此设计了针对不同时长区间样本敏感的深度学习评估框架。

最终在十万级测试订单样本上,深度学习模型对出餐时间的预估准确率在不同时长区间上,相比原有模型分别能有3%~14%幅度的提升,在95%的订单上能将37 min的平均送餐时间再缩短约0.8 min。

为了衡量调度系统的优化效果,计算整体的配送时间平均值变化还不够。例如,10笔订单的平均配送时间缩短1 min,有可能配送最慢的2单各延迟了1 min,配送最快的3单各缩短了4 min,但这并不是预期的理想效果,反而可能会有更多的用户体验离开满意时间区间。因此,还需要统计所有用户等待配送时间的整体占比变化,观察是否有越来越多的用户体验进入满意时间区间,更加全面地衡量调度系统优化效果。

按照这个计算方法统计一下调度系统上线前后的效果,如图2所示,可以看到等待配送时间分布曲线明显地向送餐时间更短的左方偏移,调度效果得到全面改善。

## 4 可视化平台

可视化技术能够结合业务逻辑、地理

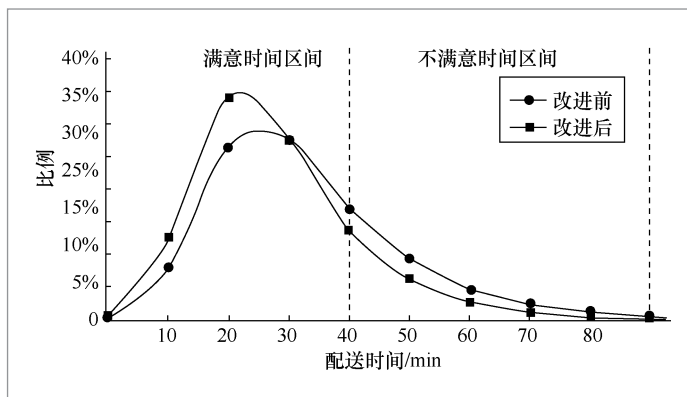


图2 调度系统上线前后等待配送时间分布曲线

信息和人群画像,从多角度展现信息、观察趋势<sup>[6]</sup>。物流调度逻辑非常抽象,但又会切实影响整个环节参与者的感受,因此需要可视化技术将复杂的数据和逻辑转化为可以理解和交互的图形界面,帮助骑士、用户和开发人员更好地理解和使用整个调度系统。好的可视化技术能够建立骑士和用户对调度系统的信任感,降低开发人员发现、定位并解决系统问题的成本,进一步发挥出大数据在提高物流调度平台效率、改进用户消费体验方面的作用。

具体来说,百度外卖物流调度平台在可视化方面的成果具体如下。

- 实时回传服务器并向用户展现骑士的运行轨迹,其中难点在于骑士端采集骑士定位数据需要在精度和性能两者之间进行平衡,在用户端需要有稳定的长连接技术保证轨迹数据的有效性和及时性。

- 采集并存储每一次调度用到的所有参数数据,方便以后可以回溯定位到任意一笔订单的分配理由。系统能够回放出当时将订单A分配给骑士B的所有细节参数和判断理由,方便相关骑士和研发人员定位排查调度算法逻辑错误或底层采集的数据问题。只需要调整可能有问题的特征值或权重,平台就能从那个时间点出发,按照新的模型重新演算一遍调度分配过程,验证效果。

- 调度系统在掌握了商圈内每个商户所有订单的地理分布和配送效果之后,能够绘制出商户配送范围内的配送效率分布图。由于受到地理位置、道路特点和业务密集程度等因素的影响,这份地形图并不是基于商户地址的同心圆,而是类似衡量山丘高度的等高线图,可以清晰明了地告知业务人员影响物流配送效率的瓶颈在哪里,如何更合理地划分商户配送范围,调整商圈运力分布。

• 这些可视化技术降低了相关角色（例如用户、骑士、研发和业务人员）使用和理解调度平台的难度，用更直观易懂的方式将背后复杂海量的大数据规律呈现出来，基于应用客户端和Web页面的人机交互技术则可以帮助使用者进一步调控和分析感兴趣的过程和维度。从某种意义上讲，放大了云端调度计算和深度智能学习的能力和效果。

## 5 结束语

百度外卖物流智能调度系统采用云计算和人工智能深度学习技术，将复杂的调度问题分层处理，并在日益增长的物流配送大数据的基础上，不断精细化调度模型依赖的状态预估数值，不断提高调度模型的多目标规划能力，同时通过大量运用平台可视化技术，以实时、图表化和可交互的方式将系统运行状态呈现出来，最终在平台运力效率和用户体验时间上得到优化效果。

展望未来，更加智能的调度系统需要依赖人工智能技术从物流配送大数据中主动发现可以改进的方向，能够有针对性地解决用户体验和平台效率中存在的普遍问题和长尾问题。例如，利用深度学习模型对大量基础特征的组合训练能力，自动构造打分算法依赖的调度因素的组合或约束结构，形成多层反馈神经网络，找出最优的分配方案；针对极端自然天气现象及复杂社会环境因素等引起的短时期局部范围内城市交通堵塞或行驶困难状况，通过多维度采集公共平台和行业领域综合交通业务数据，围绕如何及时评判行业运力紧张状况的时间空间持续范围及程度，并采取分级预警和调控措施等内容展开研究；基于行业领域的物流运输需求大数据，挖

掘城市区域内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，围绕如何精准预测，并在此基础上通过技术手段干预物流运输需求，达到在时间和空间维度上趋于均匀有效分布状态等内容展开研究。

## 参考文献:

- [1] 熊刚, 董西松, 朱凤华, 等. 城市交通大数据技术及智能应用系统[J]. 大数据, 2015042.  
XIONG G, DONG X S, ZHU F H, et al. Big data technologies and intelligent application system for urban transportation[J]. Big Data Research, 2015042.
- [2] 潘柱廷, 程学旗, 袁晓如, 等. CCF大专委2016年大数据发展趋势预测——解读和行动建议[J]. 大数据, 2016, 2(1): 105-113.  
PAN Z T, CHENG X Q, YUAN X R, et al. Developing trend forecasting of big data in 2016 from CCF TFBD: interpretation and proposals[J]. Big Data Research, 2016, 2(1): 105-113.
- [3] 刘淳安. 动态多目标优化进化算法及其应用[M]. 北京: 科学出版社, 2014.  
LIU C A. Dynamic multi-objective optimization evolutionary algorithm and application[M]. Beijing: Science Press, 2014.
- [4] 唐敏, 关键, 邓国强, 等. 一种求解二部图最大匹配问题新算法及其应用[J]. 计算机系统应用, 2012, 21(3): 72-75.  
TANG M, GUAN J, DENG G Q, et al. A new algorithm and application of solving maximum matching problem of bipartite graph[J]. Computer Systems & Applications, 2012, 21(3): 72-75.
- [5] 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.
- [6] 张宏鑫, 盛风帆, 徐沛原, 等. 基于移动终端日志数据的人群特征可视化[J]. 软件学报,

2016, 27(5): 1174–1187.

- ZHANG H X, SHENG F F, XU P Y, et al. Visualizing user characteristics based on mobile device log data[J]. Journal of Software, 2016, 27(5): 1174–1187. □

行业会讯

# 2016中国大数据技术大会 在北京盛大召开

2016年12月8日,作为大数据领域规模最大、最具影响力的IT盛会,2016中国大数据技术大会(BDTC 2016)在北京新云南皇冠假日酒店盛大开幕。本次大会历时3天,以更加国际化的视野,邀请众多国内外大数据专家齐聚一堂,从政策法规、技术实践和产业应用等角度深入探讨大数据落地后的挑战,与往届相比,本届大会更关注行业应用最佳实践,突出应用和数据的深度融合,并首次将人工智能、高性能计算等热点话题引入大会,作为大数据产业界、科技界与政府部门密切合作的重要平台,本届大会吸引了数千名大数据技术从业者与技术专家到场参会。

BDTC 2016由中国计算机学会(CCF)主办,CCF大数据专家委员会承办,中国科学院计算技术研究所、中科天玑数据科技股份有限公司与CSDN共同协办。

为了更好地探讨大数据技术生态系统的现状和发展趋势,交流大数据技术实践经验,进一步推进大数据技术创新与应用,展示国内外大数据领域的最新成果,BDTC 2016除第一天的全体大会外,主办

方还精心策划了16场专题技术和行业论坛,涵盖了大数据分析生态系统、大数据云服务、HPC大数据、推荐系统、数据安全与隐私保护、人工智能、网络与通讯、政策法规与标准化、工业与制造业、数据库、金融、精准医疗和生物医药大数据、交通旅游与出行等主题、数据开放与政府治理高层沙龙。超过130位技术专家,奉上了100多场精心准备的技术演讲,与参会者聚而论道。

本届大会的主题是“聚焦行业最佳实践,应用与数据深度融合”。大会主题亮点:回顾10年大数据历程,聚焦最佳实践;垂直领域应用百花齐放,数据和应用加速融合;紧跟大数据技术前沿,突出产学研用;关注安全和隐私保护,以数据开放创新政府治理。

大会期间,组委会开设了各种方便互动和社交的交流环节和场所,每一场精彩讲演之后都进行了深入的提问和讨论。作为大数据技术与应用深度结合的新起点,BDTC 2016已经成为产业界、科技界与政府部门密切合作的新平台,为推动我国大数据的产学研用做出了重大贡献。 □



## 2015年《大数据》高被引论文 Top10

排名	论文信息
1	高丰. 开放数据:概念、现状与机遇[J]. 大数据, 2015(2): 2015014. GAO F. Open data: concepts, status and opportunities[J]. Big Data Research, 2015(2): 2015014.
2	李涛, 曾春秋, 周武柏, 等. 大数据时代的数据挖掘——从应用的角度看大数据挖掘[J]. 大数据, 2015(4): 2015041. LI T, ZENG C Q, ZHOU W B, et al. Data mining in the era of big data: from the application perspective[J]. Big Data Research, 2015(4): 2015041.
3	李国杰. 对大数据的再认识[J]. 大数据, 2015(1): 2015001. LI G J. Further understanding of big data[J]. Big Data Research, 2015(1): 2015001.
4	董诚, 林立, 金海, 等. 医疗健康大数据: 应用实例与系统分析[J]. 大数据, 2015(2): 2015021. DONG C, LIN L, JIN H, et al. Big data in healthcare: applications and system analytics[J]. Big Data Research, 2015(2): 2015021.
5	王融. 关于大数据交易核心法律问题——数据所有权的探讨[J]. 大数据, 2015(2): 2015018. WANG R. Discussion on the legal core question of the data ownership in big data trade[J]. Big Data Research, 2015(2): 2015018.
6	周大铭. 我国政府数据开放现状和保障机制[J]. 大数据, 2015(2): 2015015. ZHOU D M. Status and safeguard mechanisms of chinese government data opening[J]. Big Data Research, 2015(2): 2015015.
7	黄宜华. 大数据机器学习系统研究进展[J]. 大数据, 2015(1): 2015004. HUANG Y H. Research progress on big data machine learning system[J]. Big Data Research, 2015(1): 2015004.
8	杨琪, 龚南宁. 我国大数据交易的主要问题及建议[J]. 大数据, 2015(2): 2015017. YANG Q, GONG N Y. Reflections on big data exchange of china[J]. Big Data Research, 2015(2): 2015017.
9	李翠平, 蓝梦微, 邹本友, 等. 大数据与推荐系统[J]. 大数据, 2015(3): 2015026. LI C P, LAN M W, ZOU B Y, et al. Big data and recommendation system[J]. Big Data Research, 2015(3): 2015026.
10	朱扬勇, 熊贲. 大数据是数据、技术, 还是应用[J]. 大数据, 2015(1): 2015007. ZHU Y Y, XIONG Y. Defining big data[J]. Big Data Research, 2015(1): 2015007.

数据来源: 中国知网; 统计日期: 2016-12-30

# 电信科学

Telecommunications Science

www.telecomsci.com.cn

- 中国通信学会和人民邮电出版社携手打造
- 中文核心期刊
- 中国科技核心期刊
- 被国内外核心数据库和科技文摘期刊收录

## 聚焦自主创新 助推信息通信发展

视点聚焦

运营技术广角

综述

专题

研究与开发



邮发代号：2-397 月刊 每月20号出版 2017年12期 48元/期 全年价576元

### 邮局订阅

凭邮发代号，在全国各地邮局（所）订阅

### 发行部订阅

将订阅信息以传真或邮件方式发送至发行部，也可致电发行部提交订阅信息

电话：010-81055597

传真：010-81055464

E-mail: zhangyaning@bjxintong.com.cn

### 银行汇款

户名：北京信通传媒有限责任公司

开户行：中国工商银行北京体育馆路支行

账号：0200008109200044661

### 邮局汇款

北京信通传媒有限责任公司发行部

地址：北京市丰台区成寿寺路11号邮电出版大厦8层

邮编：100078



邮发代号：2-537 国外代号：C9118 定价：35.00元

ISSN 2096-0271



9 772096 027179