

起底滴滴数据科学团队

数据驱动超复杂线下场景 但拒绝“唯数据论”

本文为清华数据科学研究院联合大数据文摘发起的年度白皮书《顶级数据团队建设全景报告》系列专访的第一篇内容。《报告》囊括专家访谈、问卷、网络数据分析，力求为行业内数据团队的组建和高校数据人才的培养提供指导性意见。

作者 刘俊豪

面对疫情这样的重大社会事件，数据科学团队能做什么？

16万、37987名、1500万公里，这是滴滴数据科学团队在医护车队项目中交出的答卷。

2019年腊月二十九武汉“封城”，大量医护人员出行不便，滴滴随即组建医护保障车队，为医护人员免费提供出行服务。除夕当晚，50辆车投入运营。如今，即使防疫进入常态化，但每次回想起春节期间的医护车队项目，滴滴数据科学与智能部高级数据科学总监李伟健还是充满了感慨。

3个多月的时间里，共有300多名司机加入武汉医护车队，累计为武汉16家医院近2万名医护人员提供了近50万单服务。而在全国15座城市，共近16万名司机自愿报名加入滴滴医护车队，总计服务37987名医务工作者，行驶总里程超过1500万公里。

能够在短时间内组织运营医护车队，除了高效的线下能力，滴滴多年来积累的出行数据和团队用数据解决问题的经验也很关键。李伟健介绍道，医护车队上线初期主要依靠工作人员手动匹配，为了提高接单效率，滴滴紧急为医护人员研发了线上产品，第二天，武汉医护人员就可以在APP线上接单。

滴滴数据科学与智能部也在第一时间加入，在订单级别从数据角度针对医护人员的用车规律和出行场景进行实时分析。比如，他们发现，早上七点是医务人员的上下班高峰，很多医生下班后不会回家，而是前往酒店等。

除此之外，对出行高峰、出行热区等的预判，也能有效帮助业务团队提前对司机进行调度，更高效地保障医护人员的出行。整个春节，李伟健都在和同事一起，在相对较小样本的环境下输出了大量分析结果，有效地支撑着医护车队项目决策的快速迭代。

海量数据背后，是滴滴数据科学体系的支持和承接，在大数据文摘采访几位负责人的过程中，隐藏在滴滴的数据基因也渐渐显露出来。

数据体系团队四大模块 助力业务可持续发展

作为一家老牌互联网公司，数据思维一直贯穿着滴滴各项业务的发展。

早在出租车时代的各项服务中，滴滴的数据思维就已经显现，以司乘规模、供需匹配等交易环节为中心，数据分析评估已经覆盖到了体验、司乘生态、城市交通安全等众多场景。

2017年，滴滴正式组建数据科学部，他们的目标是用数据为滴滴的运营和产品提供洞见，帮助业务在快速的迭代中科学决策，实现“数据驱动”，一方面要重视数据积累，另一方面也要辩证地看待和使用数据。

这对滴滴数据科学团队的后续发展也起到了一定的影响，在数据科学团队内部，奉行着这样一条不成文的规律，先找准业务中最需要利用数据的模块，在这些领域中体现出数据价值后，再将数据决策扩展到其他业务模块上去。

也正是秉承着这样的传统，滴滴的数据科学家们也天然地和业务部门很是亲近。

据滴滴技术副总裁、数据科学与智能部的负责人赖春波介绍道，滴滴的数



据体系分为四大模块，大数据架构、数据平台、数据治理、数据科学。在职位划分中，下面三大模块多为工程师、产品经理、数据开发工程师。数据科学分为数据分析师和数据科学家。他们数量最多，以“嵌入式”的方式，分布在不同的业务部门中。其中，数据科学团队，需要在业务形态中实现广泛的运营智能、产品智能和决策智能，助力业务可持续发展。

数据科学：通过系统的数据挖掘和主动深入的业务分析，看清业务发展方向和要素，提出策略建议，帮助业务实现用户价值与商业价值；并通过科学的实验设计和评估，辅助管理层更快更准确地进行业务决策，保证决策质量；数据治理（DG）：通过系统、管理流程、意识提升等手段，体系化治理全公司数据资产，向前赋能，提高数据使用效率，发挥数据生产力；数据平台（DP）：通过工具产品，向前提升生产效率、可靠性和可扩展性；大数据架构（Dinf）：构建稳定可靠、低成本、高性能的大数据基础设施，赋能业务。

2017年首份《顶级数据团队建设全景报告》调研了解，顶级数据团队一般具有相似的特征：所在组织或机构数据驱动战略明确，数据团队运作高效；高层需要设置清晰的数据团队建设目标并将数据纳入决策流程；数据团队的高效运作则需要优秀的团队领导、合理的组织架构和多样化的人才。

高层中心化的数据指导部门对于一家公司的数据科学团队建设的效用显而易见，包括联想集团、瓜子二手车等公司，都设置有中心化的组织，统一领导公司数据化运营。

滴滴也不例外。根据赖春波介绍，由于滴滴有网约车、车主服务、两轮车、代驾、出租车等多个业务群，滴滴的数据科学家也就很自然地分散在不同的业务部门里。为了能更全面准确赋能业务，滴滴组建了数据科学委员会，增强跨业务数据科学家间的交流和协作，同时对复杂问题进行决策和迭代数据体系建设。

与瓜子二手车的“技术委员会”不同的是，滴滴的数据科学委员会成员占比最多的是数据分析师，他们每季度开

会一次，主要针对公司的规划服务和长期定位等进行商讨。

而这些例行会议并不只是技术交流。毕竟除了技术能力和批判性思考的能力外，一个好的数据分析师还需要足够的商业能力、战略视野、影响力、领导力和同理心等素养，每次会议也不可避免地涉及到相关领域的讨论。

“分析师需要把自己脑袋中的东西放到别人脑袋，是靠嘴吃饭的。”赖春波笑称。

不过要想真正提升产品、运营和决策的智能化，只靠一张嘴是远远不够的。赖春波介绍，数据科学团队每周会产出四五十份的专题分析研究和每周几千次的实验和评估，这些都随时可能影响到公司决策。前者会呈金字塔式排列，最顶端的体系化和方向性研究是真正实现辅助战略的决策智能，投入的精力也更多；后者主要针对业务或产品的方案进行评估，相对更加自动化和流程化。

构建智慧交通 数据共建共享很关键

如今，滴滴已经成为国内最大的一站式移动出行服务平台，每天处理的数据量高达4875TB。但滴滴想做的还远不止于此。赖春波说，滴滴希望能帮助构建智慧城市，在交通汽车产业做得更好。

要实现这个目标，仅靠滴滴内部数据是不够的，需要从更大的社会维度进行数据的共建共享。据介绍，滴滴正携手国家预警信息发布中心、各地交通管理部门以及行业合作伙伴，进一步丰富平台天气特征、路网信息，积极鼓励司机和乘客进行交通上报，加强数据的完善。

也正是得益于与外部的众多合作，滴滴在二十多个城市基于平台车辆数据及城市交警卡口、地磁等多元数据，落地了包括智慧信号灯、智慧交通诱导屏、交通信息系统等智慧交通项目。不仅如此，滴滴还向学界免费开放脱敏后的出行场景数据，助力学界更好地进行前沿探索。

从2017年接入ofo，2018年正式托管小蓝单车，到上线自有品牌青桔，伴随着用户骑行数据的不断完善，滴滴还

与公交集团开展定制公交、实时公交等合作，用户出行生态就能够多维度进行描绘。

但得到数据还只是第一步，全球AI咨询公司Pactera Edge副总裁Rajeev Sharma在提到数据系统时曾说过这么一句话，“你面对的，是一个有生命的、会呼吸的系统”。

数据在体系内的应用则更加重要。对供给和需求进行有效预测，并提前进行调度，是提升网约车效率和服务体验的关键。温度、降水、司乘活动，以及是否是工作日等都会对供需情况产生影响，基于海量数据和机器学习等算法，滴滴也能模拟未来供需和历史供需、司乘补贴、城市、天气、节日关系，进而对未来的供需情况进行预测，进而更加精细化地提前进行调度，进一步提升成交率和司乘两端的出行体验。

而在安全层面，数据分析也在发挥巨大的价值。由于疲劳驾驶危害极大，通过大量的分析，自去年6月起，滴滴在《道路交通安全法实施条例》要求的基础上还上线了防疲劳驾驶的长时策略，司机达到一定计费时长后休息6小时才能上线；之后又进一步在车载录像设备中设计了疲劳驾驶预警系统，能在设备中自动进行图像处理与分析，检测司机的疲劳特征，在司机进入疲劳状态前语音提醒司机注意行车通风，及时休息。

相关分析还显示，不仅仅是在夜间，在凌晨或午后一些其他时段，司机重度疲劳的概率也会更高，为此，滴滴扩大了易疲劳时段的覆盖范围，开始在易疲劳时段对全量司机进行实时的加强提醒，以进一步帮助提升驾驶安全。

依赖数据，但不“唯数据论”

这些工作能够顺利完成，与滴滴内部已形成共识的数据文化思维有密切的关系。实际上，广受社会关注的安全事件发生后，滴滴还开始在网约车引入“安全派单”，这对数据模型的要求也变得更高。派单系统需要分析评估接单场景可能存在的风险，充分考虑乘客的性别、出行习惯、订单时间、订单距离等订单特征和司机的驾驶习惯、历史订单信息、投诉记录等服务质量分级，在此前全局

派单的基础上进一步综合分析司机和乘客是否合适出行。

赖春波也坦言，安全派单一定程度上可能也会伤害乘客体验，出现女乘客深夜有时更难打到车的情况，策略还需要不断调整优化。

也正是有了这样的案例，对于如今的滴滴来说，数据面临的问题不在于对“数据驱动”本身的质疑上，而在于当评估业务中一个不可量化的任务时，如何把握好“度”。

交通行业是一个系统工程，很多环节面临的问题都没有也没有任何一个完美的模型可以解决。一边要千方百计从技术上去做更好的算法，但同时也需要企业更多考虑现实和线下场景的复杂性。

滴滴数据治理和数据平台部负责人王勇总结称，“要依赖数据做决策，但不能只依赖数据做决策”。

在他看来，和电商业务不同，滴滴的双边交易市场是实时一对一撮合交易，是处于更加零和博弈的状态。同时，移动出行的行业渗透率有限，滴滴兼具线上和线下特征，线下数据显得十分重要，但行业内专家相对有限，这就需要大家持续探索利用数据进行不断地试错迭代，沉淀方法论和战略视野。

李伟健也表示，正是这些让滴滴场景里的实验和评估等数据科学问题变得更加独特和有价值。比如当滴滴在乘客端做了实验优化后，却发现因为司机端运力被抢夺导致结果不理想。“这里面有很多有挑战性的问题，我们也非常欢迎优秀的统计人才加入”。

据介绍，为了更好地将数据文化贯彻到整个公司内部，滴滴数据体系还面向全员开设了以提升数据能力、培养数据思维为目标的能力提升课程，并结合技术分享、训练营等多样化方式展开交流。

针对数据科学家，则会在专业技能之外会侧重培养商业理解能力、洞察力和影响力，以进一步提升通用素质和专业技能。在学习互动过程中，加强团队信任，通过切实案例理解数据驱动，这对每个岗位来说都是必要的。

在数据应用落地上，为了让数据、业务、工程三方合作更和谐，滴滴成立了data business partner团队，既要强调客户意识导向的文化，也要强调价值牵引技术驱动的背景，很多时候只依靠任何一方都是不行的。

在谈到对数据人才的期待时，滴滴表示，他们会重点关注以下五个方面：

- 用户导向思维，数据发挥价值要跟创造价值联系在一起；
- 认识数据的边界和局限性；
- 要有同理心，数据方虽不直接负责业务，但需要了解彼此目标；
- 布道者角色，让数据文化落地，影响更多人；
- 敬畏数据安全和隐私。E



BIG DATA DIGEST
大数据文摘

（本文由《大数据文摘》杂志授权转载）