

# 数据竞赛助力大数据产业 生态系统健康发展

*Data competition helps healthy development  
of big data industry ecosystem*



陈娟(1991-),女,就职于中国科学院计算技术研究所,中国计算机学会(CCF)大数据专家委员会对外合作部主任,主要研究方向为社会媒体分析及信息传播。参与创立了CCF大数据系列赛事品牌——大数据与计算智能大赛(BDCI),BDCI得到了教育部、中共中央网络安全和信息化委员会办公室、科学技术部等多个部委及地方政府的指导与支持,与百度、腾讯、阿里巴巴、国家电网有限公司、中国移动通信集团公司、国双科技有限公司等知名企业建立了良好的合作关系,中国科学院计算技术研究所、微软亚洲研究院等研究机构,清华大学、北京大学、麻省理工大学等国内外知名高校积极参与,共吸引国内外1 000余所高校、1 200余家企事业单位及80余所科研机构的30 000余人参加,参赛人员来自20多个国家和地区。目前BDCI已是国内大型算法挑战综合赛事之一。

中图分类号:TP183

文献标识码:A

doi: 10.11959/j.issn.2096-0271.2018045

## 1 引言

大数据正在开启第三次信息化浪潮，中国正迎来发展数字经济的历史性机遇。立足新时代，发展数字经济是顺应时代趋势、抢抓发展机遇的重大战略。大数据一词多次出现在政府工作报告中，2018年大数据的主要表现特征可概括为：非结构化数据呈爆发式增长，对数据资产运营与管理的复杂度要求更高；各领域对大数据服务的需求将进一步增强，新技术、新业态、新模式将不断涌现；多个大数据产业集群将崛起，大数据和行业应用深度融合。概括来讲，经过几年的发展，大数据进入高质量、高速发展阶段。

然而，当前我国的高端大数据专业人才非常匮乏。“政、产、学、研、用”各界紧密交流，力图通过多种方式缓解巨大需求和人才匮乏这一矛盾。其中，关于大数据及人工智能领域算法挑战的大型活动和赛事近几年受到了各界人士的欢迎和参与。各个领域对数据竞赛的关注点在哪？为什么有如此多的大型企业、参赛团队、政府机构等愿意参与其中？对大数据高端人才的培养有哪些促进作用？本文将对这些问题进行解答。

## 2 数据竞赛的作用和意义

数据竞赛参与方的构成如图1所示。按照行业发展的正常规律，一个产业如果要形成产业优势，就必须形成技术、资源和人才优势，这是帮助上下游实现业务创新升级的基础，是促进生态良性健康发展的基石。数据竞赛作为一种创新业态，对内汇聚产学研用多方资源，对外影响、助力

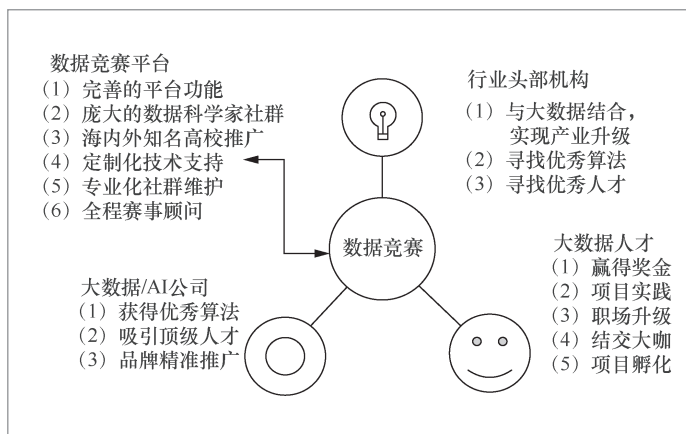


图1 数据竞赛参与方构成及关注点

整个产业生态。在大数据生态系统中，不论是政府、行业头部机构、大数据/人工智能(artificial intelligence, AI)公司，还是数量庞大的大数据人才，都在积极地为产业生态赋能，也都能从数据竞赛中获得价值。

### 2.1 行业头部机构的关注点

在大数据生态系统中，包括政府、行业头部企业在内的组织机构扮演着重要的角色，大数据与行业深度融合，也就有了更丰富的应用场景，这将为行业自身变革提供有效的方式。

近几年，各地政府在数据竞赛方面进行了积极尝试。以2018年中国计算机学会(CCF)大数据与计算智能大赛为例，中电科大数据研究院有限公司利用贵阳市政府的开放大数据，设立了1 500万元的“大数据+政府治理”开放项目基金，用于解决政府决策、政府治理、政务管理等方面的问题。

此外，国家电网有限公司、中国电力投资集团公司、中国航天科工集团公司等大型企业对于数据竞赛也格外重视，这些企业成为行业主动求变的先行军。以“AIIA杯”(航天科工站)小样本下的卫星图像典型

目标识别为例,通过数据和业务需求的发布,积极探索大数据人才外脑、社会众包的问题解决新模式。而从大赛的组织结构来看,大赛指导单位为国家发展和改革委员会、工业和信息化部、中共中央网络安全和信息化委员会办公室,主办单位为中国人工智能产业发展联盟、中国航天科工集团公司、浙江省宁波市政府。这足以看出,各级政府和头部企业正在积极推进大数据和传统行业的深度融合,使之成为一股创新原力。综合分析近几年的赛事可以发现,国内外许多传统行业头部企业作为数据竞赛的出题方已成为一种常态。

## 2.2 大数据/AI公司的关注点

数据科学家是大数据领域的普遍需求。时下的大数据已进入以数据广泛关联、跨域融合和深度应用为特征的智慧化阶段。数据已经成为战略资源及经济资产,通过机器学习方法来挖掘、分析海量数据,鼓励学科交叉、跨界合作,探索以大数据为基础,涉及政府治理、产业升级等的算法及解决方案已经成为时代发展的迫切需求。数据大赛通常都是把握数据挖掘的深度和应用的广度,面向重点行业和应用领域征集需求的,因此大数据/AI公司非常愿意参与其中,以实现人才、算法、品牌的“一石三鸟”的目的。

近几年,大数据/AI公司举办的单项赛事异常火热,如阿里巴巴“图像和美”团队联合香港理工大学纺织及制衣学系共同举办的2018 Fashion AI全球挑战赛、搜狗和今日头条等携手发起的“AI Challenger全球AI挑战赛”、京东JData算法大赛、平安科技的2018产险数据建模大赛等。其中,首届京东JData算法大赛创造了单赛题大数据类赛事参赛规模世界纪录,在一个半月的时间内,仅凭一道赛题就吸引了7 346人、

4 240支队伍报名参赛,在大数据人才圈产生了巨大影响力,获得了远远超过预期的算法方案,并有效辅助了京东技术驱动战略的推广。同样,一些知名的数据竞赛也不同程度地实现了获得出色算法、吸引优秀人才、品牌精准推广的目的。表1列举了一些典型实例,展示了数据竞赛给出题方带来的实际价值。

## 2.3 大数据人才的关注点

在大数据生态系统中,人才是核心。一项赛事有超过万名专业人士参加比赛,驱动大数据人才参加数据竞赛的动力是什么?

### 2.3.1 大数据课题研究的良好机会

根据教育部最新发布的数据,截至2018年,共有283所高校获批数据科学与大数据技术专业。很多高校导师把数据竞赛的题目作为学生的研究课题、毕业设计课题。出题方会提供大量的场景数据,题目和实际的应用场景相结合,这往往超出了高校自身所具有的资源,实现了很好的“学与用”的结合。在一些数据竞赛的决赛现场,由导师带队与选手集体答辩的情况比比皆是。

数据竞赛不仅仅在高校、科研院所得到了重视,很多商业机构也把数据竞赛题目作为培养团队成长的良机。很多知名企业组织内部团队参加与自身业务相关的数据竞赛,并多次获得大奖。

### 2.3.2 “五子棋”

获得奖金是竞赛选手能力最直接的佐证。但奖金并不是唯一的奖励,选手的竞赛诉求呈现多样化趋势,用“五子棋(奖金、实践、职场、圈子、孵化)”来形容选手诉求更加全面。

表1 数据竞赛给出题方带来的实际价值

赛题名称	赛题任务	技术属性	技术难点	赛题效果	赛题方评价
京东商城·高潜用户购买意向预测	使用京东历史销售数据,预测未来5天内,用户是否购买有购买意愿的商品及购买的商品名	机器学习、回归分析	不同用户有不同的购买用户行为,对应的用户画像复杂多样,在特征选择上存在难点	上线一周内突破了企业设置的极限成绩超30%,直接应用于京东“618购物节”,有效助力了京东大数据战略推广	让更多算法精英脱颖而出,为中国互联网电商未来的发展贡献出更多力量
国双科技有限公司·判决书中的金额提取	抽取判决书中费用类型、具体金额及相关细项,经数据分析,判断某地域某业务类型判决金额范围,辅助法官判决	自然语言处理、信息抽取	判决书由人工录入,格式不规范、数据表达方式丰富,需要准确表达及展示;金额信息在句子中出现的位置多变	参赛团队取得了0.807 07的好成绩,超过了主办方预期,方案经过简单调整后,直接应用于实际业务,成为法律判决业务的重要模块之一	从实际产品出发,在解题思路各支队伍有很多创新点,为判决书中涉案金额的精准提取提供了有效的解决方法
驭势科技(北京)有限公司·自动驾驶交通标志检测	在真实场景的图片数据中,检测出所有的交通标志的位置	图像识别	交通标志种类多,大小、角度不一,在真实行车环境中,受到天气、光照等影响,检测更难	超过500支队伍参与本赛题,优秀获奖者赛题成果直接及间接应用于驭势科技无人驾驶技术中	技术上得到了很大突破,获得了实际应用
国家电网有限公司·用电行为异常分析	根据用户用电量、电能表停走、电流失流、窃电行为等数据,训练模型判别是否存在窃电行为	用户异常行为检测、分类预测、机器学习	每个智能电表对应的用电情况复杂;需要从复杂的用电行为找到有效的检测特征	建立窃电检测模型,识别用户窃电行为,为国家电网减耗增效,监督和发现异常用户,提供技术支持和保障	有极高的行业应用价值,可解决企业类实际管理运营难题,可为行业提供更优的解决方案

“直入职场,获得实习与实践机会”已经成为大赛出题方抛出的普遍橄榄枝。在很多大赛中,参赛者可前往出题企业内部,研究超大量级的相关数据,进行训练及模型调优。对参赛者来说,这是非常具有吸引力的实践机会。目前,国内巨头IT企业(如阿里巴巴、腾讯等)已经将大赛参赛经验及参赛成绩作为入职招聘的重要参考项。

开启职场绿色通道也让众多获奖选手实现了职场进阶,赢在了职场起跑线。当在校内感叹工作难找时,一份2018届互联网校招高薪清单在网络流传,清单显示了众多知名互联网企业给技术类岗位应届生的年薪。在清单中,有20多家企业年薪突破了30万元,而在数据竞赛的参赛选手中,以算法见长的人数比例最为突出。获奖团队直接加盟出题的大数据公司的事件并不罕见。

此外,以赛交友、优质项目孵化也是数据竞赛的价值所在。例如CCF大数据计算智能大赛中,每年的参赛团队过万,屡屡有落地当地政府的团队出现,其中有的团队已经实现了C轮过亿融资。

## 2.4 数据竞赛平台的作用

国内外的数据竞赛开展得如火如荼,这也得益于专业数据竞赛平台的出现,DF(DataFountain)数据竞赛平台就是其中的代表。专业的数据竞赛平台往往拥有完善的平台功能、庞大的数据科学家社群、海内外知名高校推广、定制化技术支持、专业化社群维护、全程赛事顾问,为企业提供无忧的数据采集、数据脱敏、赛题抽象及精细化运营、个性化品牌策划方案。数据竞赛平台在大数据及人工智能的产业发展中,正在发挥着不可替代的作用。

### 3 结束语

当前,我国各产业都在深入挖掘大数据的价值,研究大数据的深度应用。大数据在各行业的全面、深度渗透,将有力

地促进产业格局重构,驱动生产方式和管理模式变革,推动制造业向网络化、数字化和智能化方向发展,大数据将成为中国经济新一轮快速增长的新动力和拉动内需的新引擎。希望更多人关注并参与数据竞赛,为我国的大数据产业发展贡献力量。 □