

数据大爆炸: KDD 2016

Microsoft
Research

微软亚洲研究院

作者: 张富峥、孙宇

2016-09-07 17:54:20

举报

阅读量: 14573

二十大KDD新星都有谁?



饕餮盛宴

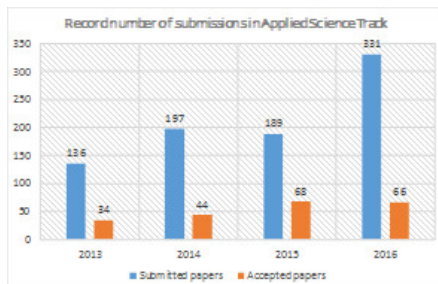
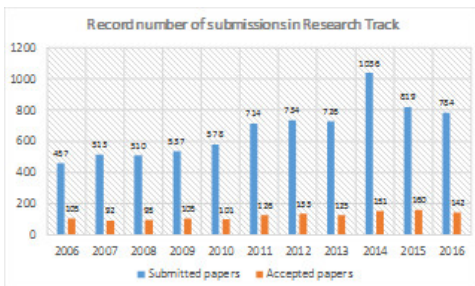
ACM SIGKDD国际会议(简称KDD)是数据挖掘领域的顶级国际会议,由ACM(计算机协会)的数据挖掘及知识发现专委会(SIGKDD)负责组织筹办。在现今如火如荼的大数据时代,从学术研究到工业应用,从科学发现到医疗卫生服务,来自不同领域的研究者们和工业界的弄潮儿们共享和交流数据挖掘的创新理念及先进方案的需求与日俱增。KDD作为数据挖掘领域的顶级年会,为该领域研究成果的展示和实战经验的交流提供了一个理想的场所。由于KDD的交叉学科性以及广泛的应

自然语言处理、数据挖掘、云计算计算、高性能计算以及大数据存储等众多领域的专家学者。

今年的KDD于8月13日至17日在美国旧金山召开。此次KDD大会堪称史上最大规模！有超过2700名来自学术界和工业界人士参与此次盛会，与往届会议参会人数相比可谓呈现爆炸式增长。旧金山市中心耸入云天的希尔顿饭店大楼里处处可见脸上洋溢着笑容与会人员；熙熙攘攘的旧金山市街头也是随处可见三三两两结伴而行的行人胸前摇摆着明亮而鲜艳的KDD参会人员标牌。这些参会人员从世界的各个角落赶来，来感受KDD大会节日般喜庆与热烈。这样的盛会也让这些从事数据分析和数据挖掘相关的从业人员，找到了大家庭般的归属感。

此次KDD大会不仅参会人员众多，而且这些人全都是重量级的业界大牛！在这里已经不知道如何列举这些业界大牛了，因为实在是太多了而且难分伯仲。如果真要列举，可能一页纸都不够。简单描述来说就是这样的情形，在会场上随便走出几步，你就会遇到大概十个人，这十个人里面有四五个人会看上去年长一些，他们要么是世界知名大学里数据挖掘、机器学习的学科领头人，要么是世界知名公司里的技术高管；剩下五六个人可能看上去年轻一些，而他们要么是各个大学里年轻的教授或是实验室里领头的博士研究生，要么是各大公司里的技术骨干。总之，你遇到的人不是已经耀眼闪亮的业界明星，就是正在冉冉升起的业界新星！这次大会也是参加过的众多会议里面，大牛数量最多密度最高的一个。

本届KDD共吸引了1115篇投稿，其中研究专题投稿论文784篇，最终有142篇录用；应用数据科学专题投稿论文331篇，录用66篇。今年KDD的应用数据科学专题由于放开了过去只接收工业界投稿的限制，吸引了大量来自高校的学者投稿，因此今年的331篇投稿较之去年的189篇投稿有了大幅度的提升。应用数据科学专题的录用文章中，微软和领英也领先于其他的企业。



创新与实践齐飞

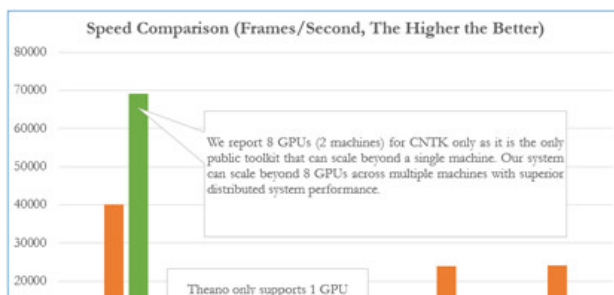
数据科学专题，包括前导辅导 (Tutorial) 进阶研究辅导 (Ordinary tutorial) 和实践辅导 (hands-on tutorial)；这里前者注重理论分析与证明，后者注重动手开发与实践。大会还邀请了众多知名科技公司高管来分享自己团队在解决公司具体问题时践行的方法和积累的经验。这是KDD作为一个学术会议务实主义的一面，因为她很好地有一部分注意力放在了如何系统地解决现实工业界中的实际问题上。可能正是这样的脚踏实地务实精神，也因为开会地点位于著名的硅谷附近，大会吸引了许多公司的数据科学家和研发工程师前来参会交流，营造了一个让学术界与工业界沟通交流的和谐融洽的平台。

此次大会的另一个创新点是各个赞助商公司提供了招聘人才的机会。由于大会在业界无与伦比的声望以及注重开发实践的务实主义精神，她吸引了许许多多业界知名的公司前来资助、招聘人才，包括微软、谷歌、苹果、脸书、腾讯等国际知名企业。在会议正式开始之前，大会为参会者提供机会向自己感兴趣的公司提交简历。在大会开始之后，每个赞助企业都在人才交流峰会场有自己独立的展台，为参会人员提供企业信息咨询、招聘职位介绍、以及现场报名招聘甚至现场面试的机会。除此之外，一些总部位于旧金山市区的公司还在会议期间，特别提供了参观公司的机会，让参会者能够走进公司内部，感受其开放的工程师文化。微软也在自己的展台积极介绍了微软与数据挖掘机器学习紧密相关的产品，吸引了非常多的参会人员前去交流。我们也尝试寻访了很多展台并与一些公司进行了后续聊天，收获颇丰，在这个过程中感受到有很多的公司多样的文化，以及业界对数据挖掘相关的人才的渴求。



微软的长袖善舞

作为世界顶尖的研究中心之一，微软在本次大会上的表现格外亮眼。在大会伊始的主题报告上，微软新英格兰研究院和微软纽约研究院院长Jennifer Chayes对稀疏大规模网络的建模和推理进行了精彩风趣的解析。在应用数据科学的邀请报告上，作为社交网络六度空间理论的著名学者，微软纽约研究院的首席研究员Duncan Watts深刻剖析了大数据时代下计算社会学所关注的诸如金融系统的系统性风险以及流行病传播演化之类的大问题和大挑战。在偏重于实践应用的实践辅导上，微软展示了能够实现多机器多GPU运行的开源大规模深度学习工具CNTK以及如何Spark或者Hadoop等分布式平台上用R语言进行大规模地数据挖掘。



微软在今年的KDD中发表了11篇高质量的论文, 其中有5篇来自位于北京的微软亚洲研究院。这些论文覆盖了数据挖掘相关的众多方向, 比如推荐系统、迁移学习、社交网络以及神经网络。从理论到应用, 这些论文也涉及到了多个新兴的应用领域, 比如在线广告中用户的注意力分析、基于聊天的推荐场景以及基于搜索日志的疾病监控。

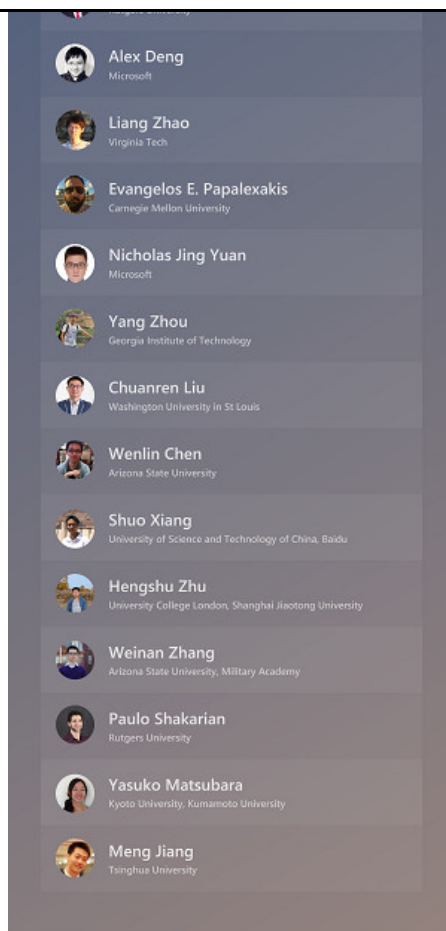
我们团队在这次KDD上发表了2篇论文。一篇文章是《智能个人助手中基于环境的用户意图推断》(Contextual Intent Tracking for Personal Assistants), 很荣幸这篇文章获得了这次KDD应用数据科学专题的最佳学生论文奖, 这篇文章的第一作者是孙宇同学。我们的论文内容是跟用户信息推荐相关的, 根据用户的上下文信息来推测用户当前的意图, 并以此意图来推荐相关的资讯或者主动提供帮助。总体感觉, 我们的论文能够获得最佳学生论文的原因是我们研究的问题本身紧密结合了当前公司个人助理产品的开发, 问题很实际, 也有很广泛的应用价值。解决问题的方法会对微软小娜等个人助理产品上的推荐有助于提高用户使用个人助理的粘性和满意度, 对其他公司个人助理产品, 例如苹果公司的Siri和谷歌的Google Now等, 也都有很实际的参考意义。尽管我们在问题定义、实验和论文筹备的过程中投入了大量时间和精力, 但文章能得奖也离不开微软亚洲研究院紧密结合开发的研究精神和对实习生的悉心培养, 以及研究院导师们远见卓识、高屋建瓴的指引和在研究遇到困难时耐心的指导。同时, 大会能将奖项颁发给这篇论文也再一次印证了大会注重工业界研发实践的务实精神。

我们团队的另一篇论文是《更好的推荐系统: 当协同过滤遇见知识库》(Collaborative Knowledge Base Embedding for Recommender Systems), 这篇KDD论文的灵感起源于我们团队与微软小娜的内容推荐项目的合作。在最初使用传统协同过滤算法时由于数据稀疏性而效果不尽如人意时, 我们想到了知识库中的语义网络、文字描述以及图片等信息都可能被用来提高推荐系统的质量。在如何使用异构信息这个问题上, 我们使用了异构网络嵌入以及深度学习等技术来提取异构数据中的语义信息, 并和协同过滤算法进行了深度融合, 在电影和书籍数据集上分别验证了推荐算法的有效性。

KDD新星

正如前面所说的, 这场KDD大会大咖云集。也有这样一批年轻学者, 他们不仅在近年的KDD大会中持续贡献了高质量的论文, 在后续也极具发展潜力。微软学术搜索在近期基于KDD大会的作者及论文等相关信息, 结合微软学术大数据, 评选出了二十位在近6年KDD里上升最快的学术新星, 名单如下。看看你心仪的大牛有没有名列其中吧





作者简介:

张富峰, 微软亚洲研究院副研究员, 中国科学技术大学计算机博士。研究领域为数据挖掘和人工智能, 方向包括推荐系统、深度学习、社交网络、情感检测、时空数据挖掘和大规模机器学习系统。



孙宇, 目前于墨尔本大学攻读博士学位, 本科毕业于中国人民大学计算机系, 曾实习于微软亚洲研究院, 并参与微软人工智能助手Cortana推荐相关的工作。



推荐阅读



丈母娘的无理要求招来了警察

人都说：“坑爹”，这明明是丈母娘坑女婿呀！



据说自主品牌三缸发动机能全面超越合资 这款车最有话语

说起三缸发动机，大多数人的第一反应是迷之微笑。



古代生活大揭秘：古人没有厕纸，如何解决“后股之忧”？

卫生纸真是个伟大的发明啊



从转子发动机到创驰蓝天 马自达的情怀你懂多少？

当魂动遇上大理、当那一抹经典的红映衬在古城的蓝天白云下、当澎湃动力遇上洱海旁的高速，前不久马自达在云南举行了全系车型试驾会，没错是全系车型，想上土坡豁，CX-5CX-4开着走！想来一把速度与激情，阿特兹昂科塞拉任你挑！不管是哪一款，你都能在这次的...



别再拿着父母的钱装逼了

01

换一换 查看更多

转发 24

评论 2

17

快速开通微博你可以查看更多内容，还可以评论、转发微博。

微博精彩

- 热门微博 热门话题
- 名人堂 微博会员
- 微相册 微游戏

手机玩微博



扫码下载，更多版本戳这里

认证&合作

- 申请认证 开放平台
- 企业微博 链接网站
- 微博标识 广告服务

微博帮助

- 常见问题
- 自助服务

大家正在搜: 好奇双11

[首页](#) [视频](#) [发现](#) [游戏](#) [注册](#) [登录](#)

[关于微博](#) [微博帮助](#) [意见反馈](#) [舞弊举报](#) [开放平台](#) [微博招聘](#) [新浪网导航](#) [社区管理中心](#) [微博社区公约](#) [京ICP证100780号](#) [互联网药品服务许可证](#)

Copyright ? 2009-2015 WEIBO 北京微梦创科网络技术有限公司 [京网文\[2014\]2046-296号](#) [京ICP备12002058号](#) [增值电信业务经营许可证B2-20140447](#)

[中文\(简体\)](#) ▼

服务热线: 4000 960 960 (个人/企业) 服务时间9:00-21:00 4000 980 980 (广告主) 服务时间9:00-18:00 (按当地市话标准计算)