

# 第十一届中国R会议(合肥)

# 会议手册

2017 6.17-6.18 中国科学技术大学



中国科学技术大学管理学院  
The School of Management, University of Science and Technology of China



中国科学技术大学国际金融研究院  
International Institute of Finance



CAPITAL OF STATISTICS  
PROFESSION, HUMANITY & INTEGRITY



大数据分析与应用  
安徽省级重点实验室



科大讯飞 MERKLE IT大咖说

主办: 中国科学技术大学管理学院 / 中国科学技术大学国际金融研究院 / 统计之都

协办: 大数据分析与应用安徽重点实验室 战略合作伙伴: 科大讯飞 / 美库尔 直播支持: IT大咖说



# 第十届中国 R 会议 (合肥)

联合主办



中国科学技术大学

管理学院



中国科学技术大学  
UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA

国际金融研究院  
International Institute of Finance

中国科学技术大学

国际金融研究院



CAPITAL OF STATISTICS  
PROFESSION, HUMANITY & INTEGRITY

统计之都

共同协办



大数据分析与应用

安徽省重点实验室

# 欢迎辞

2017 年，中国 R 会议已走过十载春秋。在这样一个值得纪念的年份，R 会也首次来到四大科教名城之一的合肥。

R 会发展至今，已经不只局限于一个和 R 语言有关的会议，它的内涵已经从 R Language 被扩大到 Research，是数据科学的一场盛会。在这里，可以看到量化金融如何从不同层面进行市场博弈，可以看到生物统计在人类医疗水平进步中的重要贡献，也可以看到目前大热的机器学习如何进一步影响我们的生活。

我们由衷感谢各位学术专家、业界精英以及技术大咖从百忙之中抽空参加本次会议，分享和数据科学的相关故事，为中国数据科学的发展添加动力。

十而复一，迎来新一轮起点的 R 会，和首次举办 R 会的合肥，又将碰撞出怎样的火花。

让我们拭目以待。

第十届中国 R 会议（合肥）筹备委员会  
二零一七年六月十七日

# 目录

欢迎辞	i
<b>会议介绍</b>	<b>1</b>
主办机构	1
协办机构	5
战略合作伙伴	6
直播支持	8
第十届中国 R 会议（合肥）筹备委员会	9
<b>会议日程及演讲信息</b>	<b>10</b>
主会场（17 日上午 8:30-下午 17:30，西区礼堂）	15
谭昶：讯飞大数据的实践与思考	15
张渊浩：R 在客户关系管理中的应用	15
王占锋：函数型数据的过程分析方法	15
谢金贵：医疗数据分析	15
刘淇：社会化行为数据挖掘方法及应用	16
周矿：复杂网络置信社团结构挖掘	16
史东辉：A Data-Mining Approach to Identification of Risk Factors in Aviation Safety	17
熊熹：大规模线上实验与机器学习	17
讯飞大数据专场（17 日下午 14:00-17:30，西活学术报告厅）	19
张弢：R 语言在教育大数据上的应用	19
袁晶：Applied Science in Bing China and XiaoIce	19
朱琛：AI * HR：用数据改变招聘	19
刘丹月：交通大数据分析与可视化	20
量化金融专场（17 日下午 14:00-17:30，西活多功能厅）	21
刘钊：一个指数的悲剧	21
李翛然：如何制造一次成功的投资	21
薛昆：大数据时代的量化投资	21
李茂：上证 50 成分股的“社交网络”	22
美库尔营销大数据专场（18 日上午 8:30-12:00，西活学术报告厅）	23
李闻宇：多渠道广告营销的绩效评估	23
马海平：讯飞用户标签挖掘方法	23
马朝良：天使前的天使：创新项目众筹-融资与营销融合	23
潘宇曦：运营商大数据应用	24
狗熊会专场（18 日上午 8:30-12:00，西活多功能厅）	25
成慧敏：数据融合与信用风险评估	25
陈昱：移动程序化广告	25
周扬：基于车联网数据的商业价值探索	25
周静：证券分析师的价值分析	25
王菲菲：高校创业数据分析	26
生物与医疗专场（18 日上午 8:30-12:00，特种楼学术报告厅）	27
金博：AI+ 慢性病管理”使精准医疗成为可能	27
黄帅：Smart Monitoring for Complex Diseases by Collaborative Learning and Selective Sensing	27

---

赖颖蕾: Detecting concordance and discordance changes among a series of large-scale data sets . . . . .	28
蒋宇康: 眼底图像自动识别与诊断 . . . . .	29
软件工具专场 (18 日下午 14:00-17:30, 西活学术报告厅) . . . . .	30
李舰: R 与深度学习的应用 . . . . .	30
陈天龙: Detection and Tracking . . . . .	30
黎思言: Great Again or Stronger Together? Sentiment Analysis About Book Reviews on Amazon . . . . .	30
李亚光: Consistent Multiple Change-point Detection and R implementation . . . . .	31
徐刚: 高频金融数据的非参数分析方法 . . . . .	31
机器学习与优化专场 (18 日下午 14:00-17:30, 西活多功能厅) . . . . .	32
邓世名: Targeted Sampling and Pricing Strategy with Imperfect Targetability and Customer Learning . . . . .	32
杨滔: DataBrain, 基于 R 语言开发的机器学习引擎 . . . . .	32
邓琪: First Order Methods for Fast Linear Programming in SHUFE . . . . .	32
黄训蓬: 基于低秩近似的一般性增量矩阵分解框架 . . . . .	33
<b>参会指南</b> . . . . .	<b>34</b>
中国科学技术大学地图 . . . . .	34
校园周边设施 . . . . .	35
温馨提示 . . . . .	38

# 会议介绍

对统计学而言，这或许是最好的时代。

信息技术的蓬勃发展让海量数据触手可及。大数据时代已然到来。在过去的许多年里，我们经历了大数据相关产业的高速发展，诸如生物医药、金融、移动互联网、车载信息技术等。毫无疑问，数据科学的巨大需求正在悄然兴起，统计学正大有可为。

对统计学而言，这或许是最坏的时代。

大数据相关科学与产业的发展太过迅速，随之而来的是具有复杂结构的海量的庞大数据集。在许多情况下，现有的统计模型无法拟合非结构化数据。通常，我们也无法算出大规模数据下的最大似然估计。在方兴未艾的大数据相关产业与科学面前，我们的领域知识显得如此苍白无力。放眼世界或立足中国，统计学都正面临着严峻挑战。

## 主办机构

### 中国科学技术大学管理学院

中国科学技术大学管理学院于 1995 年 6 月成立，管理学院的使命是创造管理思想与工具，培养管理英才与领袖。

管理学院遵循交流、开放和国际联合的原则办学，与北美、澳大利亚、日本、新加坡及香港、台湾地区的大学与研究机构建立了紧密的合作关系和培养计划。管理学院聘有数十名国内外知名专家和学者作为顾问和兼职教授，他们的参与使管理学院的教学和研究与社会需要密切结合。广泛的社会联系为高年级和研究生提供了大量的社会实践机会。

基础扎实、融会贯通、视野开阔、实际工作能力强是科大管理学院毕业生的特点。管理学院有一支实力雄厚的研究与教学队伍。十余年来，已培养了数百名中高级工商管理人才；毕业生在社会上做出了优秀成绩，很多已成为企业、政府和金融部门的管理骨干。

管理学院本科生前两年的课程突出数理和基本技能训练，除统计专业由数学系招生并进行前期教学培养外，其他学生入学后前一年不分专业，后三年根据学生自己的兴趣在院内选择专业（不包括统计专业），并根据各专业的要求进行专业理论和专业技能培养。统计分析手段和计算机应用技术的训练将贯穿于整个培养过程之中。

学院下设工商管理系、管理科学系、统计与金融系，以及 MBA、EMBA 专业学位教育中心和 EDP 高层管理者培训与发展项目中心。构建了高质量的管理科学与工程、统计学和工商管理三个一级学科人才培养和科研体系，同时设有科技创新管理与创业研究所、决策科学实验室、统计研究所、电子商务实验室、安徽省金融信息重点实验室、可持续发展与创新研究中心、工商管理创新研究中心、服务科学与信息技术实验室、科普产业研究所等有影响力的研究机构。统计学和管理科学两个国家一级学科排名均进入全国前十，学院培养了国家杰出青年 3 人、国家优秀青年 4 人、中组部青年千人计划 1 人、中国科学院百人计划 5 人，承担了基金委创新研究群体等一系列重要科研项目。

近 5 年来，管理学院的科研能力呈现强劲的上升势头，已稳居全国管理科学与工程学科前十位。2007 年大陆教育部学位与研究生教育发展中心公布的管理科学与工程一级学科排名中，管理学院位居全国第 10 位。管理学院的教师 2008-2012 年一共发表了 669 篇同行评审期刊论文（每年论文数都超过了 120 篇），人均 7 篇。其中国际刊物学术论文为 426 篇，SSCI、SCI 和 EI 检索的文章 299 篇。这五年间还出版了研究专著 4 部、书籍 10 本、咨询和研究报告 45 份。取得了令人瞩目的成果。从系别上看，在发表的 669 篇同行评审期刊论文中管理科学与工程系、工商管理系分别为 477 篇和 192 篇。2008-2012 年一共获得了 57.689 百万元人民币的研究经费（约合 9.1721 百万美元），人均 62 万元。其中国家基金 42.907 百万元，企业基金 7.574 百万元，国际合作

研究基金 7.208 百万元。尤其是 2012 年，获得的经费大幅增长，全年经费总额为 26.734 百万元，占五年经费总额的 46%；其中国家基金 21.998 百万元，占五年获得国家基金经费总额的 51%。

管理学院于 2015 年 2 月获得 AACSB (the Association to Advance Collegiate Schools of Business, 国际精英商学院协会) 认证——系全球首家按照 AACSB 最新标准获得认证的单位，并在 AMBA2016 年全球大会上，从 AMBA 主席 Sir Paul Judge 先生手中，庄重地接过了 AMBA 的认证证书，至此中国科大管理学院获得 AACSB 和 AMBA 两大国际认证。

在今后的发展阶段，管理学院将继续以“顶天（理论水平高）立地（实践能力强）”的教研宗旨，围绕既定的战略目标和长期规划，健康有序、踏实践健地开展各项教学、科研和管理工作，以“红专并进理实交融”的校训为指南，本着“创一流成果服务中国育一流人才建设中国”的愿景目标，努力创造体现国际水平、中国情景、国家需求的一流成果，致力培养具有创新思维、全球视野、社会责任的管理领袖和管理精英，为社会发展和经济建设做出更大贡献。



### 中国科学技术大学国际金融研究院

中国科大国际金融研究院是合肥市政府和中国科大全面战略合作、共建协同创新平台的又一力作，以“数据深度，世界高度”为办学使命，以金融、大数据、国际化与交叉学科为办学抓手，打造世界一流金融学者的荟萃高地、国际化金融创新人才的培养基地、中国金融的开放创新创业平台、金融产业调转促推手。

中国科大国际金融研究院位于合肥徽州大道与烟墩路交口东南方向，系合肥市滨湖国际金融后台服务基地核心位置。项目用地 60 余亩，项目签约暨开工仪式于 2016 年 11 月 19 日举行，目前各项工作正在紧锣密鼓地进行中。合肥市负责研究院基本建设工作，计划在五年内初步建成国内领先的金融学科教育、金融人才培养基地。

中国科大国际金融研究院的研究方向主要涵盖商务数据分析、互联网金融、金融大数据、物流供应链金融、智慧金融及量化投资、创新创业等方面，充分彰显了国际化特色、数理工具特色（数学学院、计算机学院、信息学院等共建）及互联网特色（互联网+、大数据、物联网等融合）。

在运作方面，中国科大国际金融研究院以教学科研为核心层，以产学研结合为紧密层，以多方合作共嬴为合作层，打造基础研究平台、开放平台和教育平台；以高端国际化办学方式，与西雅图华盛顿大学 Foster 商学院和圣路易斯华盛顿大学 Olin 商学院等国际一流商学院合作办学，汇聚一流师资，培育高端金融人才；以服务地方为己任，加强与“一行三会”及金融机构的合作，进一步发挥中国科大在科研创新领域的优势，充分利用中国科大海外校友资源，定期举办金融前沿高峰论坛及高端培训，培养高端金融人才，促进地区产业升级。

中国科大国际金融研究院将充分利用合肥的区位、产业、环境优势和中国科大人才、科教、平台资源，助推

合肥金融跨越式发展和人才高地建设，助力中国科大建设世界一流大学、一流学科，创出安徽乃至全国政产学研合作的新典范，打造国际互联网金融和数理金融的引领者。



## 统计之都

“统计之都”(Capital of Statistics, 简称 COS) 网站成立于 2006 年 5 月 19 日，其主旨为传播统计学知识并将其应用于实际领域。纵观现今国内统计学理论和应用的发展，一方面我们不难发现统计学在应用领域的巨大潜力——现代管理、咨询、商业、经济、金融、医药、生物等等，无不需要数据的力量，而另一方面我们也不得不承认，国内统计学的应用很大程度上受理论的制约——无论是应用界的人们对统计学基础理论知识的欠缺，还是学术界所研究的理论对应用领域问题的轻视。“统计之都”网站便是基于这样的认识而创建的。我们希望，统计理论研究者能充分关注应用问题，而统计应用者也能正确把握统计学基本知识，将统计学这门应用学科真正的潜力开发出来。“统计之都”为非赢利性质网站，但大力欢迎所有商界和研究领域的朋友与我们在实际应用问题上合作。我们的口号是：

中国统计学门户网站，免费统计学服务平台

我们怀着“十年磨一剑”的决心，要将“统计之都”创建成中国的统计学“正直、人本、专业”的社区；我们抱着“己欲立而立人、己欲达而达人”的信条，要将“统计之都”以免费统计学服务平台的形式坚持办下去。我们希望“统计之都”在专业知识体系上有真正的王者风范，在面对用户需求时却又以谦恭的态度为大家服务。统计之都（下文简称 COS）虽以网站和论坛起家繁荣，但是随着越来越多喜爱统计的朋友们加入，大家对于线下活动和书稿撰写翻译等等的需求也越来越旺。目前，COS 的线下活动从一年两次的 R 会议，逐渐发展到沙龙、交流会、竞赛、讲座、培训等等。我们希望更多的新鲜血液可以就近加入 COS 的线下活动中。COS 线下活动总结：

1. 中国 R 会议：目前已开展到第十届，分别在北京、上海、广州、杭州、西安、南昌、武汉等地举办。历届会议纪要和幻灯片共享都可以在 COS 主站上找到：<http://china-r.org/>
2. 线下沙龙：目前我们在北京、上海和广州深圳开展线下沙龙活动。不同于规模庞大的 R 会，沙龙形式更为轻巧，注重讨论交流。目前已经举办过 37 期，目前主要在北京，每月举办，详情参见统计之都微信公众号。
3. 海外在线视频沙龙：我们在 Google Hangouts 举办在线沙龙，主要由海外嘉宾来分享学术、生活中的点点滴滴。目前已经举办 18 期：<http://meetup.cos.name/>.
4. 书籍出版，包括写作和翻译。如《Dynamic Documents with R and knitr》(2nd edition) 谢益辉著，《Implementing Reproducible Research》谢益辉等著，《数据科学中的 R 语言》李舰、肖凯著，《R 语言实战》

高涛、肖楠、陈钢翻译,《ggplot2: 数据分析与图形艺术》统计之都翻译,《R 语言核心技术手册》刘思喆、李舰、陈钢、邓一硕翻译,《R 语言编程艺术》陈堰平、邱怡轩、潘岚峰等翻译,《R 数据可视化手册》肖楠、邓一硕、魏太云翻译,《R 语言统计入门》邓一硕、郝智恒、何通翻译,《数据科学实战》冯凌秉、王群锋翻译,《R 语言实战》(第 2 版) 王小宁、刘撷芯、黄俊文翻译,《Rcpp:R 与 C++ 的无缝结合》寇强、张晔翻译,《R 绘图系统》呼思乐、张晔、蔡俊翻译,《R 语言编程实战(暂定)》(待出版) 冯凌秉翻译,《量化投资与 R》(待出版) 邓一硕、冯凌秉、杨环翻译,《金融风险建模与投资组合优化》(待出版) 邓一硕、郑志勇等翻译等等。



## 协办机构

### 大数据分析与应用安徽省重点实验室

大数据分析与应用安徽省重点实验室于2017年正式获得安徽省科技厅批准认定，实验室依托中国科学技术大学计算机学院，联合科大讯飞股份有限公司、安徽广电传媒产业集团有限责任公司共同建设。针对国家战略规划和安徽省创新发展对大数据技术的迫切需求，实验室瞄准大数据分析的当前挑战及发展前沿，在中国科大、科大讯飞公司及安徽广电集团的已有科研基础上，将在大数据基础理论、大数据分析技术、大数据应用以及大数据处理平台等几个方向有所突破，拟建设成为国内领先、国际上有重要影响的大数据创新学术团队、大数据方向高端人才培养高地，打造国内综合性强、开放性好的大数据产学研基地，引领安徽省大数据产业发展，推动相关产业转型升级。



## 战略合作伙伴

### 科大讯飞有限公司



科大讯飞股份有限公司成立于 1999 年，是一家专业从事智能语音及语言技术、人工智能技术研究，软件及芯片产品开发，语音信息服务及电子政务系统集成的国家级骨干软件企业。2008 年，科大讯飞在深圳证券交易所挂牌上市，股票代码：002230。

科大讯飞作为中国智能语音与人工智能产业领导者，在语音合成、语音识别、口语评测、自然语言处理等多项技术上拥有国际领先的成果。科大讯飞是我国唯一以语音技术为产业化方向的“国家 863 计划成果产业化基地”、“国家规划布局内重点软件企业”、“国家高技术产业化示范工程”，并被原信息产业部确定为中文语音交互技术标准工作组组长单位，牵头制定中文语音技术标准。

自 90 年代中期以来，在历次的国内外语音合成评测中，各项关键指标均名列第一；2003 年、2011 年，科大讯飞两次荣获“国家科技进步奖”；2005 年、2011 年两次获得中国信息产业自主创新最高荣誉“信息产业重大技术发明奖”；2008 年至今，科大讯飞连续在国际说话人、语种识别评测大赛中名列前茅；2014 年，科大讯飞首次参加国际口语机器翻译评测比赛（IWSLT）即在中英和英中互译方向中以显著优势勇获第一；2016 年，国际语音识别大赛（CHiME）科大讯飞取得全部指标第一；在认知智能领域，相继获得国际认知智能测试（WSC）全球第一、国际知识图谱构建大赛核心任务全球第一。

随着移动互联网时代的到来，科大讯飞率先发布了全球首个提供移动互联网智能语音交互能力的讯飞开放平台，并持续升级优化。基于该平台，科大讯飞相继推出了讯飞输入法、灵犀语音助手等示范性应用，并与广大合作伙伴携手推动各类语音应用深入到手机、汽车、家电、玩具等各个领域，引领和推动着移动互联网时代大潮下输入和交互模式的变革。

基于拥有自主知识产权的世界领先智能语音技术，科大讯飞已推出从大型电信级应用到小型嵌入式应用，从电信、金融等行业到企业和消费者用户，从手机到车载，从家电到玩具，能够满足不同应用环境的多种产品。科大讯飞已占有中文语音技术市场 70% 以上市场份额。

2014 年，随着人工智能时代的到来，科大讯飞推出了“讯飞超脑计划”，旨在实现一个中文的认知智能计算引擎，未来将引领在家居、教育、客服、医疗等领域的智能应用。2015 年，科大讯飞重新定义了万物互联时代的人机交互标准，发布了对人工智能产业具有里程碑意义的人机交互界面——AIUI。2016 年，围绕科大讯飞人工智能开放平台的使用人次与创业团队成倍增长，带动超百万人进行双创活动。截至 2017 年 4 月，讯飞开放平台在线日服务量超 35 亿人次，合作伙伴达到 25 万家，用户数超 10 亿，以科大讯飞为中心的人工智能产业生态持续构建。

科大讯飞在语音技术核心研究和产业化方面的突出成绩引起了社会各界的广泛关注，习近平、李克强等多位党和国家领导人视察过科大讯飞技术产品，对智能语音技术的重大价值及科大讯飞做出的创新工作均给予充分肯定。国家各行业部委和权威机构已形成了科大讯飞“中文语音产业国家队”的基本共识。

## 美库尔商务信息咨询有限公司

Merkle Inc.，世界顶级的大数据营销咨询公司。公司创立于 1971 年，总部设在美国首都大华府地区，全公司 3,800 多名员工分布在全美 16 座城市，以及英国伦敦、布里斯托、西班牙巴塞罗那和中国上海、南京。Merkle 为全球超过 200 家的国际大型品牌提供咨询服务，涉及行业包括汽车、多渠道零售、保险、金融服务、抵押贷款、商业银行业务、信用卡业务、非盈利性基金募集、消费品包装行销、生物制药和旅游休闲等领域。公司于 2009 年初起，拓展部分核心业务到中国，为全球和本土企业提供客户关系市场营销服务，主要业务包括：市场营销策划、客户关系管理、统计建模和分析，数据处理分析、数据库开发、管理和服务。

### 【全球办公地点分布】



### 【公司业务及主要客户】



### 【关注 Merkle】



图 1: 新浪微博



图 2: 微信订阅号: MerkleChina



图 3: 微信服务号: merkle

## 直播支持

IT 大咖说



IT 大咖说，IT 垂直领域的大咖知识分享平台，践行“开源是一种态度”，通过线上线下开放模式分享行业 TOP 大咖干货，技术大会在线直播点播，在线直播知识分享平台。75+ 合作社区，每周 15+ 场技术大会精彩分享，2000+ 业内大咖资源。让程序猿、攻城狮不再遗憾，随时随地，想看就看！IT 大咖说，让智慧流动起来！

## 第十届中国 R 会议（合肥）筹备委员会

### 主席

林 枫

### 秘书团

(按姓名拼音排序)

戴玮宸 董梦皓 胡 蝶 姜昊茗 李亚东  
李叶蓁 李芸琪 马千驰 宋思航 童世炜  
王柔琳 宣婷婷 徐丹冶 徐 骞 周豆豆

### 志愿者

(按姓名拼音排序)

陈 丹 崔 笛 董 仕 董玉林 杜鑫杰  
桂 林 韩汶衍 侯晓天 刘许成 刘轩铭  
罗 潘 吕建成 马荣贵 钱 芳 檀佳欣  
王江丽 王 震 王梓苏 徐建军 徐自强  
张 辉 张冒莉 张猛猛 钟 立 邹 纯

# 会议日程及演讲信息

主会场：17 日上午 08:30-12:10，西区学生活动中心二楼大礼堂

时间	嘉宾	题目
08:30-08:40	林枫	回顾和感谢
08:40-08:50	院系领导	关注与鼓励
08:50-09:35	谭昶	讯飞大数据的实践与思考
09:35-10:20	张渊浩	R 在客户关系管理中的应用
10:40-11:25	王占锋	函数型数据的过程分析方法
11:25-12:10	谢金贵	医疗大数据分析
主持人		邓思怡、赵如玉

主会场：17 日下午 14:00-17:30，西区学生活动中心二楼大礼堂

时间	嘉宾	题目
14:00-14:45	刘淇	社会化行为数据挖掘方法及应用
14:45-15:30	周旷	复杂网络置信社团结构挖掘
16:00-16:45	史东辉	A Data-Mining Approach to Identification of Risk Factors in Aviation Safety
16:45-17:30	熊熹	大规模线上实验与机器学习
主持人		童世炜

讯飞大数据专场：17 日下午 14:00-17:30，西区学生活动中心二楼学术报告厅

时间	嘉宾	题目
14:00-14:45	张弢	R 语言在教育大数据上的应用
14:45-15:30	袁晶	Applied Science in Bing China and Xiaoice
16:00-16:45	朱琛	AI * HR：用数据改变招聘
16:45-17:30	刘丹月	交通大数据分析与可视化
主持人		林枫

量化金融专场：17 日下午 14:00-17:30，西区学生活动中心三楼多功能厅

时间	嘉宾	题目
14:00-14:45	刘钊	一个指数的悲剧
14:45-15:30	李翛然	如何制造一次成功的投资
16:00-16:45	薛昆	大数据时代的量化投资
16:45-17:30	李茂	上证 50 成分股的“社交网络”
主持人		陈丹

美库尔营销大数据专场：18 日上午 08:30-12:00，西区学生活动中心二楼学术报告厅

时间	嘉宾	题目
08:30-09:15	李闻宇	多渠道广告营销的绩效评估
09:15-10:00	马海平	讯飞用户标签挖掘方法
10:30-11:15	马朝良	天使前的天使：创新项目众筹-融资与营销融合
11:15-12:00	潘宇曦	运营商大数据应用
主持人		陈敏

狗熊会专场：18 日上午 08:30-12:00，西区学生活动中心三楼多功能厅

时间	嘉宾	题目
08:30-09:10	成慧敏	数据融合与信用风险评估
09:10-09:50	陈昱	移动程序化广告
10:00-10:40	周扬	基于车联网数据的商业价值探索
10:40-11:20	周静	证券分析师的价值分析
11:20-12:00	王菲菲	高校创业数据分析
主持人		张猛猛

**生物与医疗专场：18 日上午 08:30-12:00，特种楼二楼学术报告厅**

时间	嘉宾	题目
08:30-09:15	金博	AI+慢性病管理”使精准医疗成为可能
09:15-10:00	黄帅	Smart Monitoring for Complex Diseases by Collaborative Learning and Selective Sensing
10:30-11:15	赖颖蕾	Detecting concordance and discordance changes among a series of large-scale data sets
11:15-12:00	蒋宇康	眼底图像自动识别与诊断
	主持人	罗潇

**软件工具专场：18 日下午 14:00-17:30，西区学生活动中心二楼学术报告厅**

时间	嘉宾	题目
14:00-14:45	李舰	R 与深度学习的应用
14:45-15:30	陈天龙	Detection and Tracking
15:50-16:35	黎思言	Great Again or Stronger Together? Sentiment Analysis About Book Reviews on Amazon
16:35-17:00	李亚光	Consistent Multiple Change-point Detection and R implementation
17:00-17:30	徐刚	高频金融数据的非参数分析方法
	主持人	桂林

机器学习与优化专场：18 日下午 14:00-17:30，西区学生活动中心三楼多功能厅

时间	嘉宾	题目
14:00-14:45	邓世名	Targeted Sampling and Pricing Strategy with Imperfect Targetability and Customer Learning
14:45-15:30	杨滔	DataBrain, 基于 R 语言开发的机器学习引擎
16:00-16:45	邓琪	First Order Methods for Fast Linear Programming in SHUFE
16:45-17:30	黄训蓬	基于低秩近似的一般性增量矩阵分解框架
	主持人	崔笛

## 讯飞大数据的实践与思考

谭昶 科大讯飞

时间：8:50-9:35

**简介：**2014年7月于中国科学技术大学计算机应用与技术专业获得博士学位，现任科大讯飞大数据研究院副院长，中国计算机学会大数据专家委员会委员。主要从事数据挖掘和推荐系统技术的研究和推广应用，在大数据技术、智慧城市、个性化推荐方面有着多年的研究和实践经验。

**摘要：**以大数据、人工智能等为代表的新一代信息技术，正在从每一个细微的角落改变整个人类社会，更加精准、更加智能的信息服务需求快速增长。作为人工智能技术的领军企业和一家拥有8.9亿用户的互联网企业，如何运用人工智能技术从自有大数据中深度挖掘更多价值，促使大数据和人工智能技术更加紧密的结合，正是科大讯飞所要探索、实践和反复思考的。

## R 在客户关系管理中的应用

张渊浩 美库尔

时间：9:35-10:20

**简介：**美库尔数据分析部高级经理，图卢兹一大经济学院硕士。在美库尔从事客户关系管理(CRM)的分析工作，工作内容主要包括数据整合，数据可视化，数据建模，消费者行为分析，营销策略分析等。对于数据可视化(R Shiny, D3)和数据建模在市场营销上的应用有浓厚兴趣。

**摘要：**对于营销来说，成功的定义是找到合适的人，在合适的渠道发送合适的消息。美库尔结合多年数据分析的经验，开发了众多的模型来解决上述问题。本报告会介绍美库尔数据分析部是怎样借助R，在众多数据和茫茫人海中上下求索，通过建模解决渠道和目标人群的问题。美库尔也借助可视化工具(R Shiny)，优化营销执行的流程，包括营销前期的规划，预测以及后期的营销结果的评估和展现。

## 函数型数据的过程分析方法

王占锋 中国科学技术大学

时间：10:40-11:25

**简介：**中国科学技术大学统计与金融系副教授。分别于2003年和2008年获中国科学技术大学统计学学士和理学博士学位。主要从事统计渐近理论，生物统计，函数性数据分析等领域的研究。曾主持国家自然科学青年基金一项，参与国家重点自然科学基金一项和自然科学基金一项。2014年获得安徽省高等学校省级教坛新秀称号。

**摘要：**本报告主要基于3个不同函数型数据集：executive function research data, motor learning data 和 upper limb function evaluation data，提出三种不同的统计模型和3种不同的分析方法，并利用R语言去实现所提的统计方法。

## 医疗大数据分析

谢金贵 中国科学技术大学

时间：11:25-12:10

**简介：**中国科学技术大学管理学院副教授，管理科学系副主任，NUS-USTC 商业数据分析中心联合主任，中国运筹学会随机服务与运作管理分会秘书长，中科大医疗服务研究中心主任。2010 年获清华大学管理科学与工程博士学位，美国哥伦比亚大学商学院联合培养博士。在国际知名期刊发表文章十多篇。

**摘要：**介绍医疗大数据产业发展的驱动力、产业发展前景，医疗大数据的应用，研究的方法和技术，介绍该领域目前的发展现状及未来的发展潜力。

## 社会化行为数据挖掘方法及应用

刘淇 中国科学技术大学

时间：14:00-14:45

**简介：**博士，中国科学技术大学计算机学院副教授。主要研究数据挖掘与知识发现、机器学习方法及其应用，着重于针对用户行为数据（如消费数据、社交数据、教育数据等）的建模和应用研究。在重要国际学术会议和期刊共发表论文 60 余篇，2011 年获得数据挖掘领域顶级国际会议之一 IEEE ICDM 的最佳研究论文奖，还获中科院院长特别奖、KSEM 2013 最佳论文奖以及 SDM 2015 最佳论文候选奖、中科院优博等重要学术奖励，入选了中科院青年创新促进会。主持了多项国家、省部级以及与知名公司（如微软、腾讯、科大讯飞）的合作研究项目。担任了 CCF 大数据学术会议 (BigData)2015-2017 的宣传主席、是包括 IJCAI、KDD、WWW、AAAI、ICDM、CIKM 等国际会议的程序委员会委员以及国际期刊 TKDE、TKDD、TC、TSMC-C、TIST 等的审稿人、是 FCS 青年 AE。

**摘要：**在当今大数据时代，社会个体在诸多领域的行为轨迹能够被大量的记录和保存下来，如消费行为数据、社交行为数据、教育相关的行为数据等，这些社会化行为数据是学习用户行为模式和了解社会运行规律的重要信息来源，有着迫切的研究价值和应用需求。然而，应用场景的多样性、领域特征的复杂性等，使得传统的数据挖掘结果难以有效满足社会化场景的实际需求，在已有方法基础上，必须发展新的理论框架和处理算法。为此，本报告将以各领域记录的社会化行为数据为研究对象，以数据驱动的用户理解为核心，从数据挖掘方法与领域知识联合建模的角度，介绍我们近期在社会化行为数据挖掘方法及应用方面的一些研究成果。

## 复杂网络置信社团结构挖掘

周旷 西北工业大学

时间：14:45-15:30

**简介：**法国雷恩第一大学信息科学博士，现任西北工业大学理学院应用数学系助理教授。参加国家自然科学基金《基于精细可靠度评估的优化证据推理及应用》及《带有属性信息缺失的不完整数据广义信任分类》课题。研究方向是统计模式识别、应用概率统计、复杂网络社团挖掘、证据理论及其应用。

**摘要：**社团结构是复杂网络的一个重要特征。现实世界中的复杂网络通常具有高度不确定性。一方面，不同社团之间广泛存在重叠；另一方面，网络中通常存在有关社团结构的不确定先验信息。证据理论，又称置信函数理论，在不确定信息的表达和处理方面具有独特优势，已经被广泛应用于模式识别、数据聚类、信息融合、图像处理等领域。本报告主要介绍基于证据理论模型的复杂网络置信社团结构挖掘算法，探讨网络不确定社团结构的度量和发现方法，并讨论如何利用少量不确定监督信息改善无监督社团发现算法的性能。

## A Data-Mining Approach to Identification of Risk Factors in Aviation Safety

史东辉 安徽建筑大学

时间：16:00-16:45

**简介：**男，1966 年 8 月出生，教授，博士，硕士生导师，安徽建筑大学，电子与信息工程学院计算机工程系，美国路易斯维尔大学访问学者，厄瓜多尔普罗米修斯学者。主要研究方向为人工智能、机器学习、数据挖掘、信息管理等，近期主持安徽省自然科学基金、人社部留学科技活动等项目，参加厄瓜多尔普罗米修斯项目、美国路易斯维尔大学合作项目等国际交流项目。在《Expert System with Applications》，《Journal of Organizational Computing and Electronic Commerce》，《Journal of International Technology and Information Management》等国际期刊发表论文，连续 5 年在信息系统领域知名国际会议“Hawaii International Conference on System Sciences”上发表论文。

**摘要：**We present a data-mining approach to incident risk factor identification and analysis using incident data from the Federal Aviation Safety Board. Timely and accurate analysis of these reports is vital to the improvement of aviation safety. But extensive manual identification of risk factors by human experts and incomplete data present daunting challenges. The approach presented in this paper is an attempt to address both of these challenges. First, topical mining techniques are used to convert underutilized textual data in the form of incident narratives to serve as input. Second, data streaming algorithms allow efficient training of risk factor classification models. The results show promise of such an approach not only as a potential approach to process daily influx of incident reports but also a more efficient way of building a classification model. The identification of risk factors is an important step to closing the gap between incident reporting and safety management systems.

## 大规模线上实验与机器学习

熊熹 京东

时间：16:45-17:30

**简介：**京东推荐算法工程师。2015 年加入京东，一直致力于机器学习算法在京东个性化与推荐业务中的应用，目前主要负责个性化业务中大规模线上实验，和用户体验优化等。曾在国内外知名大公司和研究机构从事复杂实验设计的理论和实践工作，并持续跟踪大规模线上实验与机器学习在其中应用的前沿研究。由于在利用人工智能技术提升个性化用户体验以及更全面科学地定义个性化对京东的贡献等工作上的突出贡献，曾获得 2016 年度 CTO 特别奖。

**摘要：**大规模线上实验在京东每一天都在发生，大到一个全新的模块乃至平台上线，小到一个 icon 颜色，样式的更改，主要以 AB 试验的形式进行。大多数试验遵循直觉，数据收集和整理的工作冗长，但是对需要测试的指标以及收集到的数据的验证工作比较简单，容易造成区分度不足乃至和真实结论南辕北辙的情况。本次报告会详细介绍线上实验的基本科学原则，实施细节，容易犯的错误；并结合 google, 微软, LinkedIn, Amazon 等公司的最新研究论文，以及京东个性化推荐中的实践，从案例中学习如何使用机器学习和人工智能技术来验证数据一致性，降低误差，以指导算法优化的方向等。

## R 语言在教育大数据上的应用

张弢 科大讯飞

时间：14:00-14:45

**简介：**研究生学历，现科大讯飞大数据研究院研究员，研究领域为教育大数据挖掘工作、统计分析、机器学习。曾任南京烽火公司大数据研发工程师，从事网络安全大数据挖掘工作。

**摘要：**分享本人与 R 相关工作，重点介绍 R 在教育大数据上的应用。从师生的行为等来挖掘教育大数据的价值，进而使得学生知道自己的擅长与短板，让教师更加了解学生，并更好指导学生。

## Applied Science in Bing China and XiaoIce

袁晶 微软亚洲研究院

时间：14:45-15:30

**简介：**博士，现任微软高级科学家，Bing 中国应用科学团队负责人，此前为微软亚洲研究院研究员。他于 2007 年本科毕业自中国科学技术大学少年班学院，专业为计算数学，后于 2012 年获得该校计算机软件与理论博士学位，师从陈国良院士。他已在数据挖掘、普适计算、地理空间系统等领域的国际顶级会议和期刊上发表了 50 余篇论文，并多次获得包括 SIGSPATIAL, ICDM, KDD 在内的国际会议的最佳论文奖项。他曾作为程序委员会主席或领域主席参与组织了多次国际学术会议，并长期担任多个国际顶级会议（如 KDD, WWW, AAAI）的程序委员会委员。他于 2011 年被评为微软学者，现为 ACM 和 IEEE 会员。

**摘要：**I will introduce our recent applied science work and the related products in Bing China and XiaoIce.

XiaoIce is an advanced natural language chat-bot developed by Microsoft. It is primarily targeted at the Chinese community on the micro blogging service Weibo. The conversation is text based. The system learns about the user and provides natural language conversation. Microsoft gave XiaoIce a compelling personality and sense of “intelligence” by systematically mining the Chinese Internet for human conversations.

## AI \* HR：用数据改变招聘

朱琛 百度

时间：16:00-16:45

**简介：**朱琛高级工程师目前担任百度人才智库实验室（Talent Intelligence Center）核心研究人员，负责百度内部商业智能和企业人才管理大数据领域的科研工作和项目落地。他作为核心研究人员主持或参与了多项百度公司内部重要人工智能项目，曾作为种子项目入围 2016 年“百度最高奖”，并获得百度“最佳团队”等荣誉。曾在 KDD、AAAI、KAIS、ICDM 等人工智能相关领域的顶级期刊和会议上发表多篇学术论文，2 次获得最佳论文奖或提名奖，并申请了多项人力资源智能化相关专利，相关成果被《哈佛商业评论》等权威媒体广泛报道。

**摘要：**近年来，随着人工智能和大数据相关技术的高速发展，许多传统领域开始主动参与到智能化革命之中。规范化重复性的工作开始逐渐被机器所取代；而与此同时，具有创造性思维的人才将成为各大公司最宝贵的财富。基于以上背景，本次报告将重点介绍百度如何充分发挥其在人工智能和大数据领域的天然技术优势，在

招聘智能化领域开展的诸多前瞻性探索，以期帮助公司进行更高效准确的人才招聘，帮助求职者更清晰地规划个人发展蓝图。

## 交通大数据分析与可视化

刘丹月 科大讯飞

时间：16:45-17:30

**简介：**先后在天猫商业智能部门和讯飞大数据研究院任数据分析师，经历过三届双十一的洗礼，现主要研究交通和城市大数据分析。

**摘要：**借力 R 语言，展示交通大数据的分析方法和分析成果，包括城市交通规划，拥堵问题，通勤职住，市民出行等。

## 一个指数的悲剧

刘钊 摩根史丹利华鑫

时间：14:00-14:45

**简介：**中国科学技术大学金融工程专业博士，深圳证券交易所博士后，中国量化投资俱乐部执行理事，曾任摩根士丹利华鑫基金助理总经理、数量化投资总监，管理大摩多因子、大摩量化配置、大摩量化多策略三只公募基金，最高管理规模超过 100 亿元。其中，大摩多因子曾荣获 2014 年《中国证券报》“三年期开放式股票型持续优胜金牛基金”奖、2015 年《中国证券报》“三年期开放式股票型持续优胜金牛基金”奖、东方财富网“2014 年最受欢迎基金”奖；大摩量化配置曾荣获《中国证券报》“2014 年度开放式股票型金牛基金”、《证券时报》“2014 年明星基金”奖。

**摘要：**从一个失败的指数入手，谈量化投资的有效性、逻辑和开发模型过程中常见的错误。

## 如何制造一次成功的投资

李翛然 奇点创世

时间：14:45-15:30

**简介：**北京奇点创世信息技术有限公司创始人。先后从事过寿险精算，投资银行工作。于 2014 年创办北京奇点创世信息技术有限公司，主要业务领域为二级市场金融风险管理系统。现已有 10 余家金融机构、私募基金采用该系统为客户和自营交易提供风险管理及投资顾问服务。其主要工作经历覆盖了一级市场的发行，尽职调查，搭建企业信用分析系统，二级市场的量化分析，风险管理 saas 系统。

**摘要：**量化投资是近年来金融行业最火热的话题之一，从高频，套利交易，Alpha，事件驱动，FOF 投资，新的投资方法层出不穷，那么，到底一个投资者该如何选择策略？这些策略的背后逻辑到底有哪些优点缺点？在传统的金融学术和量化投资实战之间的巨大鸿沟有没有一些通用的方法论可以让一个新人成长？这次简短的分享可以让大家对传统投资到量化投资有一个清晰而理性的认识，同时可以对自己将来的投资生涯提供很多有意义的指导和帮助。

## 大数据时代的量化投资

薛昆 通联数据

时间：16:00-16:45

**简介：**通联数据董事总经理，优矿团队负责人。毕业于清华大学计算机系，拥有 10 年以上量化投资，金融平台研发和大数据研究经验。曾任另类投资对冲基金 Dillon Read Capital Management 量化分析师，瑞士银行投行全球量化团队董事，以及摩根大通量化研究团队副总裁等职位。

**摘要：**随着量化认知度越来越高，量化投资策略同质化也日趋严重，如何利用另类数据来挖掘非相关超额收益已成为各家机构争相研究的方向。优矿除了提供特色多维度的另类数据，也提出了信号众筹这样产学研结合的投研新思路，充分挖掘高校闲置信号为买方服务。同时应对市场快速的风格切换，除了需要多资产多策略

并行发展，有效的风险管理也是必不可少的一环，优矿自主研发适应中国市场环境的风险模型为机构用户提供了专业的适合 A 股市场的度量和控制风险的方案。

## 上证 50 成分股的“社交网络”

李茂 天津理工大学

时间：16:45-17:30

**简介：**李茂，天津理工大学理学院，讲师，硕士。研究领域：统计计算、金融计量。与高蓉老师合作出版译著《量化金融 R 语言初级教程》、《量化金融 R 语言高级教程》。高蓉，杭州电子科技大学经济学院，讲师，博士。研究领域包括资产定价、实证金融、数据科学应用。出版教材《实验经济学》，译著《数据科学入门》等。

**摘要：**网络模型可以用来解释大规模的自然和社会系统。在金融学中，我们使用网络方法来研究金融市场和金融机构的相互关联，由此识别风险在金融系统中的传播，为风险管理的投资决策提供参考。自 2007-08 年金融危机以来，金融网络模型已经成为金融计量领域一个新的重要研究方向。本报告以上证 50 指数成分股票为例，展示了一种基于偏相关的网络建模方法。我们建立了稀疏的网络拓扑结构，给出了板块的系统重要性，并且讨论了网络特征与市场表现的动态联系。

## 多渠道广告营销的绩效评估

李闻宇 美库尔

时间：8:30-9:15

**简介：**美库尔跨渠道媒体分析和数据产品副总监。领导团队为医药、零售、保险、金融、商旅娱乐等行业的跨国公司在北美和亚太地区的业务提供广告媒体组合优化的分析咨询服务。此前曾任职于阳狮集团和奥姆尼康集团。本科毕业于北京大学，并获得美国南加州大学统计学硕士及国际关系博士学位。

**摘要：**信息化时代广告媒体无处不在，它们影响受众行为的方式也日益增多。在商业活动中，我们迫切需要分辨出哪些媒体平台和广告营销活动是最有效的，并据此优化资源配置提高收益。美库尔作为业界的领跑者之一，在跨渠道媒体分析优化领域积累了多年的经验并进行了持续的创新。本次分享将介绍我们的核心解决方案以及行业的一些前沿方法论的应用。

## 讯飞用户标签挖掘方法

马海平 科大讯飞

时间：9:15-10:00

**简介：**马海平，女，1986 年 5 月生，2013 年 6 月毕业于中国科学技术大学计算机科学与技术学院并获得工学博士学位。博士期间，参与了 863、国家自然科学基金、国家科技支撑计划等国家科研项目 10 余项，以及和阿里巴巴、诺基亚研究院、科大讯飞等国内知名 IT 企业技术合作项目。以第一作者在国际顶级会议 WWW 和国际重要期刊 Neurocomputing (SCI 检索), IJITDM (SCI 检索) 发表高质量学术论文 3 篇，共发表论文 7 篇，专利 2 篇，著作 1 本。2013 年 7 月加入科大讯飞股份有限公司，现任大数据研究院研究主管，主要从事大数据挖掘算法的研究、及在计算广告、金融等领域落地应用。

**摘要：**大数据时代，将多来源的用户数据进行有效集成，完成从基本数据到标签构建，最终形成用户 360 度全景画像，是实现用户“数据化”，用户数据“价值化”的有效手段。讯飞基于其海量真实的用户大数据资源和人工智能技术，实现用户标签挖掘，构建了讯飞用户数据平台。该用户数据平台已成功应用于智能营销、智能广告投放、个性化推荐等多个业务。本次简短的分享可以让大家对讯飞用户数据平台背后的技术有个初步的认识。

## 天使前的天使：创新项目众筹-融资与营销融合

马朝良 中国科学技术大学

时间：10:30-11:15

**简介：**马朝良，中国科学技术大学管理学院博士后。2016 年毕业于中国科学技术大学管理学院，获金融工程专业博士学位。学术研究领域主要包括创新与互联网金融、众筹、众包与众创等，在国内外学术期刊上发表相关研究论文数篇。POMS、中国技术经济学会会员，并担任 Technovation、Chinese Management Studies 等国际期刊的匿名审稿人。

**摘要：**本报告主要讲述创新项目众筹的融资与营销融合问题及众筹营销的主要影响因素问题。随着“大众创业、万众创新”的发展，创新项目的融资手段不断增加，众筹融资已经成为创新项目的融资新渠道。创新项目筹

资方不仅可以通过众筹筹集到相关资金，还可以通过分析众筹投资者来预估产品的市场前景。作为一种融资和营销融合的模式，众筹得以迅速发展。本报告将探讨众筹营销的主要模式，并对众筹营销的影响因素进行重点分析，研究影响众筹融资绩效的创新项目在线信息，并将相关结果应用到众筹的实际操作中去。

## 运营商大数据应用

潘宇曦 美库尔

时间：11:15-12:00

**简介：**美库尔高级分析与市场企划经理，上海交通大学生物信息学博士，拥有人工智能算法研发及多年电信、信息技术、汽车等行业数据的量化分析、挖掘与策略咨询经验。

**摘要：**基于众多商家苦于缺乏行业客户洞察与获客渠道的现状，报告着重介绍如何运用运营商数据帮助行业客户更好地刻画潜在目标客户，从而制定针对性的营销和广告策略，进行精准客户触达以及后续客户管理的解决方案。

## 数据融合与信用风险评估

成慧敏 考拉征信

时间：8:30-9:10

**简介：**考拉征信联合研究组组长，中国人民大学统计学院硕士，曾在电信、金融、互联网等领域从事数据分析工作。对数据分析成果在业务和产品上的应用有浓厚兴趣，擅长数据集成与治理、用户行为画像研究、信用风险建模。

**摘要：**本次主要分享考拉征信依托海量数据为银行等提供信用卡申请评分模型构建服务，介绍模型设计和数据处理的一般流程。

## 移动程序化广告

陈昱 蓬景数字

时间：9:10-9:50

**简介：**北京大学光华管理学院博士生在读。研究兴趣为空间统计，深度学习等方向。狗熊会蓬景数字联合研究组组长。

**摘要：**本报告介绍移动端的程序化广告，即在移动设备上为广告主的精准营销需求提供全方位的服务。包括以多种多样的数据为基础的、建立在传统互联网广告业务的基础上的，结合移动互联网的特点及优势的多种面向效果的解决方案。本报告还将介绍蓬景数字以及狗熊会联合研究组在针对不同的场景以及数据的相关研究工作，包括基于数据的方法自动对广告出价、定向、投放、反馈、优化等各个环节。

## 基于车联网数据的商业价值探索

周扬 彩虹无线

时间：10:00-10:40

**简介：**彩虹无线数据团队负责人，浙江大学业界导师，主要研究方向包括：UBI 模型、数据产品及数据可视化。

**摘要：**当前，汽车行业整体处于数据来源一致性低、数据质量一般且可用性不强的基本状态。车联网数据作为采集频度高、数据质量好、来源稳定的数据源，成为了汽车主机厂商及周边行业的关注热点。其中包括的方向有：车险 UBI、无人驾驶、汽车营销、汽车后市场、车辆的生产制造及工况研究等核心方向。此次演讲，将基于彩虹无线多年来在车联网大数据行业的深耕，围绕实际商业应用场景，结合统计模型及算法，分享基于车联网数据商业应用的实践案例。

## 证券分析师的价值分析

周静 对标科技

时间：10:40-11:20

**简介：**中国人民大学统计学院助理教授，北京大学光华管理学院管理学博士，研究上关注复杂网络数据建模、营销模型、消费者行为分析等，研究论文发表于 Journal of business and economic Statistics、Science China Mathematics、营销科学学报等国内外权威杂志上。在产业实践上，对客户流失预警模型、用户欺诈模型等相关模型具有丰富的实战经验。热衷案例创作，是微信公众号狗熊会精品案例的作者之一。

**摘要：**证券分析师就是给市场提供投资建议的人，投资人听了他的建议交易股票，分析师赚取交易的佣金提成。这种人可不容易的；首先严格的准入条件，要去考试拿资格证书，其次，严格的语言用词限制，发出之前必须过内审，必须实事求是，不能引起市场恐慌，再次严格的监管个人的语言和行为，以防各种内部交易。但是国内分析师的评级可信吗？分析师有用吗？肯定有用，为什么？各大券商都花重金建立自己的分析团队，要是没用早就都开除了。那怎么知道谁有用谁没用，尤其是中小机构和个人投资者，自己没有什么判断能力，市场上有 4000 多位分析师，每天发出将近 500 份研报，平均都在千字左右，怎么看啊，看谁的啊！这是一个甄别分析师分析能力的问题，分析师分析的准不准主要是看他推的股票涨没涨，涨了多少，是不是跑赢了行业，他是最初价值的发现者，还是人云亦云的跟风者，亦或是发了报告股票就下跌的悲催者。对标科技通过汇总所有历史上的分析师行为数据进行挖掘，以收益率为核心通过统计分析，为每一位分析师建立数据模型，鉴别分析师的分析能力，并将不同行为特点的分析师归类，为中小投资者提供精准化的分析师群体行为的实时动态跟踪、关键信息的实时提醒并积累投资人行为；最终制定自己的投资策略。

## 高校创业数据分析

王菲菲 百分点

时间：11:20-12:00

**简介：**北京大学光华管理学院商务统计与经济计量系博士研究生，2012 年毕业于中国人民大学统计学院，获经济学学士学位。感兴趣的研究领域有：文本挖掘，贝叶斯分析等。目前的研究课题集中在文本挖掘领域，尤其是主题模型在营销和社交网络方面的应用。

**摘要：**如何科学有效的挖掘增量人才是扩充国家人才战略储备的重要方面。本案例依托某全国性的高校创业大赛，采集进入复赛的参赛团队作为创新创业人才的有效信息，并设计了一套完整的创新创业人才评价体系。在该评价体系中，我们将公民道德素质作为创新创业人才的必备要求，并从性格特质、个人能力和外部资源三个维度进行综合测评，整体的评分机制采用预设分值与模型调整相结合的方式。

## AI+ 慢性病管理”使精准医疗成为可能

金博 大连理工大学

时间：8:30-9:15

**简介：**大连理工大学副教授。致力于数据挖掘、大数据分析、创新管理、商务智能等领域的科学研究。主持了国家自然科学基金青年项目、辽宁省高校科研项目、国家重点实验室开放课题等课题，参与科技部国家重点研发计划“精准医疗研究”项目、国家自然科学基金重大研究计划培育项目和面上项目、863 计划项目等国家级课题。在相关领域重要国际期刊及会议上发表论文 60 余篇，近年来多篇论文在数据挖掘领域顶级会议（KDD、AAAI、ICDM、SDM、PAKDD 等）收录，担任数据挖掘领域三大顶级会议 KDD、ICDM、SDM 的程序委员，是 ACM、IEEE 和 CCF 高级会员。

**摘要：**调查显示，慢性病及其并发症的急性发作已成为威胁我国老年患者健康的最主要因素。以帕金森症、阿兹海默症等神经系统慢性退行性疾病为研究对象，针对临床医学研究中的慢性病并发症评估、药品不良反应预测、联合用药推荐等难题，采用机器学习和医疗大数据分析的方法，在前期积累的海量医疗数据基础上，构建人工智能 + 慢性病管理的模式，以数据为驱动，使精准医疗成为可能，为提高我国医疗信息服务水平、合理利用医疗资源、探索新的慢性病并发症个性化治疗模式提供理论与实践支撑。

## Smart Monitoring for Complex Diseases by Collaborative Learning and Selective Sensing

黄帅 华盛顿大学

时间：9:15-10:00

**简介：**Dr. Shuai Huang is an Assistant Professor at the Department of Industrial and Systems Engineering at the University of Washington. He received a B.S. degree on Statistics from the University of Science and Technology of China in 2007 and a Ph.D. degree on Industrial Engineering from the Arizona State University in 2012. He is also an adjunct faculty member at the Department of Biomedical Informatics and Medical Education (BIME) and the Integrated Brain Imaging Center (IBIC) at the University of Washington. Dr. Huang develops methodologies for modeling, monitoring, diagnosis, and prognosis of complex networked systems such as the brain connectivity networks, manufacturing systems, and disease progression process of complex diseases that have multiple stages and pathways. He also develops statistical and data mining models to integrate massive and heterogeneous datasets such as neuroimaging, genomics, proteomics, laboratory tests, demographics, and clinical variables, for facilitating scientific discoveries in biomedical research and better decision-makings in clinical practices. His research is funded by the National Science Foundation, National Institute of Health, Juvenile Diabetes Research Foundation, Helmsley Foundation, and several biomedical research institutes. Dr. Huang currently serves as Associate Editor for the IIE Transactions in Healthcare Systems Engineering and Quality Technology and Quantitative Management.

**摘要：**The emerging data-rich environments in healthcare hold great promises to accelerate the paradigm transition of U.S. healthcare from reactive care to preventive care. One question is how we could translate the big disease data into better care management of preclinical or diseased patients. While these diseases manifest complex progression process, involving both temporal dynamics and spatial evolution, how could we model, monitor, and modify these processes are challenging problems. The challenges mainly lie on three aspects:

disease modeling, monitoring, and prognosis. For example, diseases such as Alzheimer's disease and Type 1 Diabetes share the commonality that they involve slow and predictable progression processes. Knowing how a disease progresses is helpful, particularly if we'd like to prevent the disease as early as we could for maximum therapeutic efficacy and improved quality of life. The modeling of the progression process is statistically challenging given the high-dimensionality of the data (e.g., tens of thousands variables), the mixed types variables, and the data's longitudinal nature. Another commonality of these diseases is that, since they are chronic conditions, being able to recognize subtle symptoms that indicate significant clinical events or suggest worse outcomes is crucial for preventative care. Further, patients need to be dynamically prioritized by their projected risk for resource allocation optimization. This needs robust models that build on the statistical knowledge provided by disease modeling and monitoring, to guide the selection of high-risk patients for targeted care. Thus, my works collectively work towards the goal of smart monitoring. Such a smart monitoring method will provide data-driven decision-making capabilities for better disease management, leading to efficient targeted screening and affordable care, better treatment planning, and improved quality of life for both patients and caregivers.

## Detecting concordance and discordance changes among a series of large-scale data sets

赖颖蕾 乔治华盛顿大学

时间：10:30-11:15

**简介：**Dr. Yinglei Lai is Professor of Statistics at The George Washington University. His research interest is to develop statistical and computational methods in bioinformatics, computational biology and biostatistics. He received his B.S. in Information & Computation Sciences and Business Administration from the University of Science and Technology of China in 1999. Dr. Lai received his Ph.D. in Applied Mathematics (Computational Biology) from the University of Southern California in 2003. After his postdoctoral training at Yale University School of Medicine, he joined as a faculty member in the Department of Statistics at the George Washington University in 2004.

**摘要：**With the current microarray and RNA sequencing technologies, two-sample genome-wide expression data have been increasingly collected in biological and medical studies. Differential expression analysis and gene set enrichment analysis have been frequently conducted. The related statistical software in R has been widely used. Integrative analysis can be conducted when multiple data sets are available. In practice, concordant and discordant molecular behaviors among a series of data sets can be of biological and clinical interest. There is still a lack of statistical methods and software for these types of integrative analysis.

We have proposed a mixture model based approach to the integrative analysis of multiple large-scale two-sample expression data sets. Since the mixture model is based on the transformed differential expression test P-values (z-scores), it is generally applicable to the expression data generated by either microarray or RNA sequencing platforms. The mixture model is simple with three normal distribution components for each data set to represent down-regulation, up-regulation and no differential expression. However, when the number of data sets increases, the model parameter space increases exponentially due to the component combination from different data sets. To achieve a concordant and discordant integrative analysis for a series of data sets, We have introduced two model reduction strategies. The related statistical computing has been implemented in R.

We demonstrate our methods on the recent TCGA RNA sequencing data. To illustrate a concordant integrative analysis, we apply our method to a series of data sets collected for studying two closely related types

of cancer. To illustrate a discordant integrative analysis, we apply our method to a series of data sets collected for studying different types of cancer. Interesting disease-related pathways can be detected by our integrative analysis approach.

## 眼底图像自动识别与诊断

蒋宇康 中山大学

时间：11:15-12:00

**简介：**中山大学数学学院统计学本科生，华南统计研究中心和 R Square 成员，对数据分析与 R 语言有着浓厚的兴趣。在华南统计研究中心的学习工作中，接触并分析过用户收视数据、基因数据、生存数据、图像数据等。

**摘要：**医学图像处理是当今非常热门的一个话题。其中包括 CT、MRI 等三维图像重构，多种疾病的诊断，病变部位标记等等。现阶段，医学眼底图片需要专业的医生来读图诊断，并且标记各种眼部疾病，需要大量的人力，花费大量的时间。基于这样的现状，我们通过自动化诊断系统诊断眼底疾病，节省医生时间，提高工作效率。本次分享将介绍，从眼科医生的实际需求出发，如何自动对视盘、视杯与黄斑进行识别；对血管静动脉进行标定；对出血、渗出、微血管瘤进行检测。

## R 与深度学习的应用

李舰 微家实业

时间：14:00-14:45

**简介：**现任微家实业 CDO，“统计之都”核心成员之一，曾任 Mango Solutions 中国区数据总监。台湾辅仁大学博士生在读，北京大学软件工程硕士，中国人民大学统计学学士。专注于数据科学在行业里的应用，著有《数据科学中的 R 语言》一书，曾在华东师范大学、浙江大学等高校任兼职导师，讲授数据科学相关的专业课程。

**摘要：**近年来，随着人工智能在图像识别、围棋竞技等领域的突破，带来了一波投资热潮，也使得深度学习变得火热。此外，TensorFlow、MXNet 等优秀工具的成熟也使得深度学习的门槛一再降低，在很多领域都实现了成功的应用。本次报告试图用一种最简单的方式介绍深度学习技术的来龙去脉和实现原理，并分享图像识别的应用案例。此外还会介绍使用主流工具进行实际操作的方式，并对未来的可能发展方向进行展望。

## Detection and Tracking

陈天龙 中国科学技术大学

时间：14:45-15:30 邮箱：[wiwjp619@mail.ustc.edu.cn](mailto:wiwjp619@mail.ustc.edu.cn)

**简介：**Mathematics and Applied Mathematics and Computer Science double degree in USTC.

Research Experiences:

International Genetically Engineered Machine Competition Golden Prize

Software Development and Deeping learning in Bioinformatics, Biomedical Cybernetics Laboratory, Harvard

Elastic Optical Networks, Large Scale Computingand Networking Lab, USTC

Classification of skin diseases based on image data, Graphics & Geometric Computing Lab, USTC

**摘要：**The main thing I am going to show is a Detection and Tracking framework, which realizes capturing object from a video and tracking it. Firstly, I detect object at the start frame by using methods of RCNN, SPP,Fast-RCNN,Faster-RCNN,YOLO,SSD and YOLO-V2. Then, the bounding- box of the previous result is conveyed to the Tracking algorithm. I mainly focus my research on KCF and MDnet , by which the tracking performance of the object is tested. One thing need mentioning is that after a certain period of time, the location of the object is corrected using the former detection algorithm.

## Great Again or Stronger Together? Sentiment Analysis About Book Reviews on Amazon

黎思言 中央财经大学

时间：15:50-16:35

**简介：**中央财经大学应用统计专业硕士，中国人民大学五校联合培养大数据分析硕士。参与《基于出租车数据分析北京市城市道路支路利用率》，聚类分析医学类微信公众号推文，“A Combination of CNN and VADER For Sentiment Classification”等项目。

**摘要：**我们收集了希拉里和特朗普在 2016 年美国大选期间发表的两本畅销书在亚马逊上的所有书评。然后使用情感分析和文本分类方法对两本书的书评进行了分析。我们得到的信息有：在亚马逊的书评中，对特朗普畅销书的评价呈现出积极的情感特征，书评内容和美国未来发展紧密相关。相反，对希拉里畅销书的评价呈现出消极的情感特征。

## Consistent Multiple Change-point Detection and R implementation

李亚光 中国科学技术大学

时间：16:35-17:00

**简介：**中国科学技术大学统计金融系博士研究生，2012 年毕业于华南理工大学，获理学学士学位。感兴趣的研究领域有：高维数据分析，change-point，变量选择等。目前的研究课题集中在高维数据的结构识别，尤其是高维数据的变点检测问题。

**摘要：**在变点分析中，对特定时间序列的多变点检测问题一直是统计学中的热点问题。R 语言是从事统计研究的必备工具。本次报告将介绍常见的多变点检测方法和常用的 R package，特别的本次报告还将介绍对一类非平稳时间序列模型以及线性回归模型的多变点检测方法及其 R 语言实现。最后通过模拟数据和实例说明该检测方法的有效性。

## 高频金融数据的非参数分析方法

徐刚 中国科学技术大学

时间：17:00-17:30

**简介：**中国科学技术大学统计与金融系硕士研究生在读，2014 年毕业于中国科学技术大学少年班学院，获统计学学士学位。研究兴趣为非参数统计，小波分析，金融数据分析等方向。

**摘要：**金融数据时常会有不连续的跳跃点，很多时候这些跳跃点是不可忽略的。因此分析带跳跃的高频金融时间数据相比于传统的连续过程有显著不同。本报告主要介绍利用核方法和小波方法在跳跃点未知时识别这些跳跃点，并对含有跳跃点的高频金融降噪。

## Targeted Sampling and Pricing Strategy with Imperfect Targetability and Customer Learning

邓世名 华中科技大学

时间：14:00-14:45

**简介：**邓世名博士，华中科技大学管理学院，教授，博士生导师。楚天特聘学者，入选教育部新世纪优秀人才计划，武汉市东湖高新区3551人才支持计划，湖北省“百人计划”。北京大学物理学士，计算机辅修学士。2003年获得美国加利福尼亚大学伯克立分校工业工程与运筹学系博士学位。多年从事管理科学和运营管理中的研究和实践。曾在美国国际企业规划管理系统软件界领先的Oracle总部工作。在定价和制造联合决策，投资成本避税理论，供应链风险管理与合同设计理论，半导体可重组生产装备的产能规划研究等多项领域都有创新性的研究成果。

**摘要：**We investigate how sellers should design targeted sampling and pricing strategy when consumers have product fit uncertainty and/or the targeting method cannot perfectly identify customers with specific traits. We consider three cases: (1) perfect targetability and perfect learning; (2) imperfect targetability; and (3) imperfect learning. We reveal that the decisions of targeted pricing and targeted sampling are affected by each other. With the presence of imperfect targetability, an over-targeting problem of choosing too many attributes in grouping consumers may occur. The consumer grouping decisions depends on both the degree of targetability and consumers' ability to learn product with provided samples.

## DataBrain, 基于 R 语言开发的机器学习引擎

杨滔 桃树科技

时间：14:45-15:30

**简介：**杭州桃树科技有限公司创始人及首席执行官，拥有超过十年机器学习技术研究与应用经验。奥克兰大学机器学习博士，悉尼科技大学博士后。曾任阿里巴巴集团数据科学家，建立淘宝网数据科学团队，首创聚划算爆款模型。曾任F团首席科学家，建立F团数据化运营体系。

**摘要：**桃树科技产品DataBrain是一款机器学习引擎。DataBrain为数据分析师、数据工程师、信息工程师以及数据科学家提供端到端的数据建模工具。基于DataBrain所产生的信用风控、精准营销、个性化推荐和定投策略等模型已经嵌入多家银行、券商、互联网公司和IT企业的业务流程，大幅度提高企业运营效率。DataBrain降低机器学习和人工智能门槛，提供端到端的建模工具，让任何人都可以成为数据科学家。

## First Order Methods for Fast Linear Programming in SHUFE

邓琪 上海财经大学

时间：16:00-16:45

**简介：**博士，上海财经大学信息管理与工程学院讲师，博士毕业于美国佛罗里达大学，本科毕业于上海交通大学，运筹优化与机器学习领域青年研究学者。主要研究方向为运筹优化，以及优化算法在深度学习中的应用，有若干篇学术论文在机器学习及可视化领域顶级会议、期刊发表。目前参与信管学院、交叉学院与管科中心机器学习与并行优化软件开发。

**摘要：**线性规划是数学优化的核心问题之一，在管理科学和数据科学中有极其重要的作用，并被广泛而深入地研究。在求解线性规划的问题上，虽然传统的方法（单纯形法或内点法）能获得很高的精度，但它们并不能有效推广到大数据问题。本发言将介绍在大数据时代线性规划算法设计的挑战，以及基于一阶算法的线性规划求解方案。

## 基于低秩近似的一般性增量矩阵分解框架

黄训蓬 中国科学技术大学

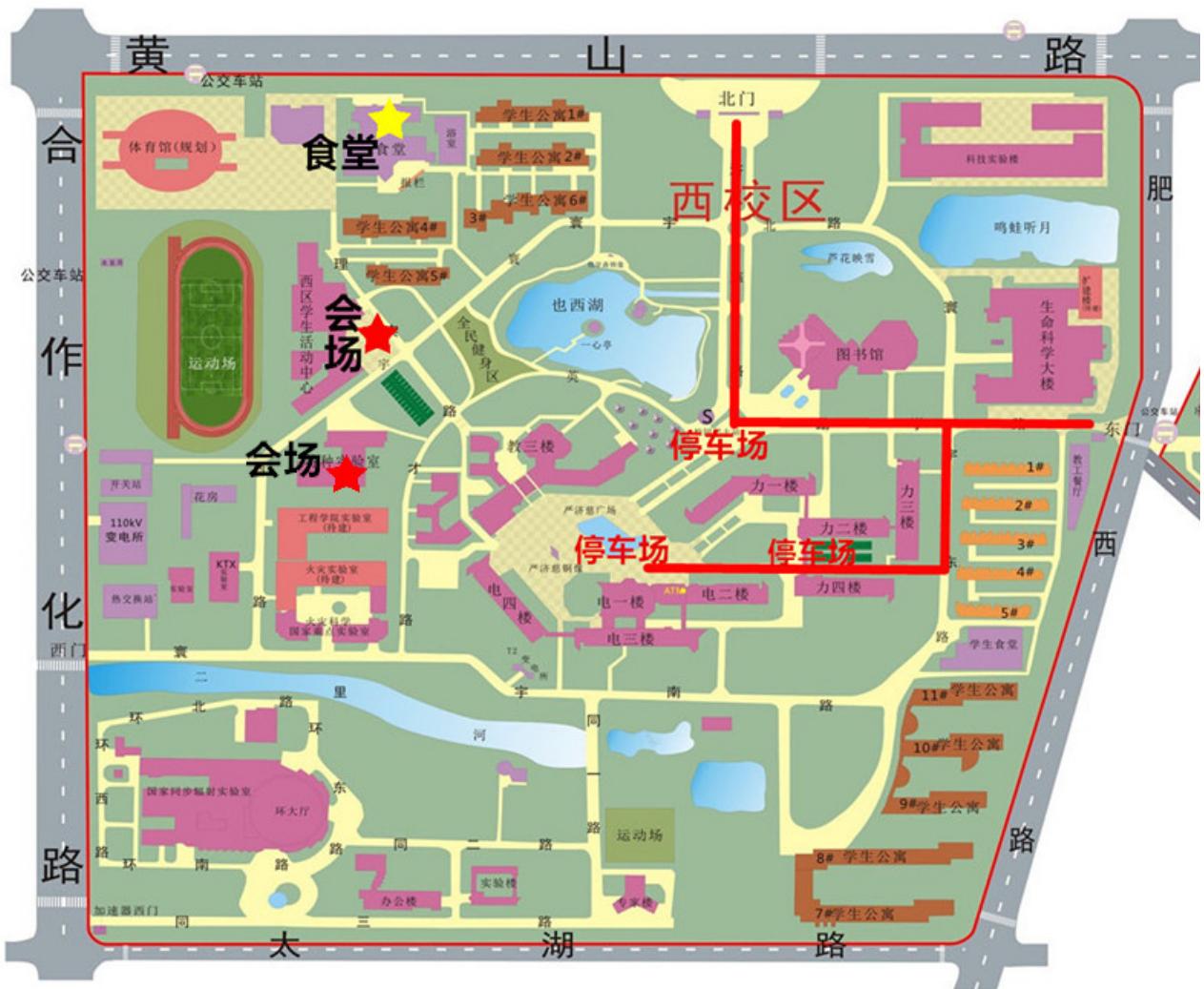
时间：16:45-17:30

**简介：**黄训蓬，中国科学技术大学在读研究生。主要研究方向为增量学习与随机优化，着重于使用低秩近似方法解决高维优化变量运算时的时空复杂性问题。以第一作者主持的工作：“Incremental Matrix Factorization: A Linear Feature Transformation Perspective”已被 IJCAI-17 接收，另有 2 篇论文于 ICDM-17、Information Science 在投。

**摘要：**矩阵分解作为协同过滤中最常用的技术之一，有着高效、易于实现等一系列优点。但是由于矩阵分解本身是批量式的模型，如何在线场景中随着已知元素的增加，增量地更新潜在特征矩阵是一个亟待解决的问题。目前大部分已有增量矩阵学习方法都有着较大的局限性，或是针对某些特定矩阵分解模型进行设计，或是有着较强的使用条件限制。在该报告中我们将介绍了一种一般性的增量矩阵分解框架，从低秩近似的角度对该方法进行了解释，给出了某些特殊情况下增量学习部分训练误差的上界。此外在多个数据集上进行的实验，验证了该方法从数据存储角度实现增量更新的同时保证了分解结果的高效与精确。

# 参会指南

## 中国科学技术大学地图



注：

1. 红星标注了会场位置。
2. 黄星标注了食堂位置。
3. 红色路线表示从北门或东门前往停车场的行车路线图；黄色路为可行路线。

## 校园周边设施

### 餐饮

中国科大东区附近

**老乡鸡（科大西门店）**

金寨路鸿基广场旁（24 小时营业）

**瑞辣重庆潮人火锅**

金寨路与太湖路交口向东 150 米瀚海星座 B 座 105

**麦芒餐厅**

金寨路 155 号黄金广场 2 幢 1 层

**风波庄酒家**

宿松路与太湖路交口南 100 米

**合肥獭祭日本料理**

金寨路高科技广场北一座 23 号

**嘿鱼捞**

金寨路 133 号

**永和豆浆**

宿松路 155 号

中国科大西区附近

**宽板凳老灶火锅**

东至路与黄山路交口向南 200 米路东

**巴蜀人家**

励学路与槽郢路交叉口东北 100 米

**大野屋韩国料理**

肥西路文馨苑小区 B-1 号商铺

**重庆酒楼**

肥西路文馨园 b 栋

## 住宿

### 中国科大东区附近

#### 晨都宾馆（宿松路店）

包河区黄山路与宿松路交叉口北行 100 米，距离中国科学技术大学东校区 526 米

#### 迷你快捷宾馆

包河区宿松路与太湖路交叉口，距离中国科学技术大学东校区 555 米

#### 盛源旅馆

包河区望江东路供电局对面，距离中国科学技术大学东校区 1.1 公里

#### 怡园宾馆（桐城路店）

包河区桐城路 444 号曙光新村对面，距离中国科学技术大学东校区 1.1 公里

#### 星星宾馆

包河区望江路与金寨路交叉口向东 100 米路南，距离中国科学技术大学东校区 1.2 公里

#### 宜临国际大酒店（桐城路店）

包河区桐城路 33 号（芜湖路与桐城路交口向南 100 米），距离中国科学技术大学东校区 1.5 公里

#### 紫云宾馆

包河区徽州大道 1140 号（中科大南区东门斜对面），距离中国科学技术大学东校区 2.0 公里

#### 丰祥快捷宾馆

包河区徽州大道 1150 号（中国科技大学南校区东门对面），距离中国科学技术大学东校区 2.1 公里

#### 包河招待所

包河区宣城路与前进巷交叉口西约 10 米，距离中国科学技术大学东校区 2.6 公里

### 中国科大西区附近

#### 宜临国际大酒店（合作化路店）

蜀山区合作化南路 19 号（合作化南路与休宁路交汇处或合作化南路高架休宁路出口），距离中国科学技术大学西校区 1.8 公里

#### E 居酒店式公寓（百盛苑小区店）

蜀山区黄山路与肥西交口，百盛苑小区内（科技大学西区 20 米），距离中国科学技术大学西校区 300 米

#### 腾龙宾馆

包河区桐城南路与望江路交口向南 200 米，距离中国科学技术大学西校区 2.4 公里

**安昊安商务宾馆**

包河区金寨路与东流路交口万豪（万瑞）花园酒店东 200 米路北，距离中国科学技术大学西校区 2.4 公里

**府台酒店**

蜀山区合作化北路 158 号，距离中国科学技术大学西校区 2.5 公里

## 温馨提示

1. 为保证会场的卫生，会场内严禁有色饮料，严禁进食与吸烟，垃圾请随身带走。
2. 为营造一个良好的会议环境，请您将手机调至静音或振动模式。请勿在会场内大声喧哗或频繁走动；拍照请关闭闪光灯。
3. 各个分会场容纳人数有限，请您根据自己的需要选择感兴趣的专场。部分会场可能会出现爆满的情况，请您及时更换会场，以保证各个会场的秩序与环境，感谢您的理解与支持。
4. 若遇紧急情况，疏散时请听从工作人员指挥，保证安全。





---

战略合作伙伴

---



---

直播支持

---





The image features a large, three-dimensional red question mark centered against a background composed of numerous overlapping, semi-transparent text words in various sizes and orientations. The words include "量化金融" (Quantitative Finance), "数据挖掘" (Data Mining), "机器学习" (Machine Learning), "精准医疗" (Precision Medicine), "生物信息" (Bioinformatics), "理论统计" (Theoretical Statistics), "互联网挖掘" (Internet Data Mining), and "时空数据" (Spatial Temporal Data). These words are repeated in a grid-like pattern across the entire image.



中国科学技术大学



统计之都



科大讯飞



美府尔