

第十届中国 R 会议（兰州） 暨西北地区数据科学会议



自由的统计语言

主办：

兰州财经大学统计学院

统计之都

协办：

甘肃省大数据产业技术创新联盟

战略合作伙伴：

知几基因



目 录

R 语言简介.....	4
第十届中国 R 会议(兰州)筹备委员会.....	5
机构简介.....	6
兰州财经大学简介.....	6
兰州财经大学统计学院简介.....	7
统计之都简介.....	8
知几基因简介.....	10
会务信息.....	11
从车站、机场抵达兰州财经大学段家滩校区路线规划.....	11
会议日程.....	12
第十届中国 R 会议(兰州)暨西北地区数据科学会议日程.....	12
嘉宾职务及所在单位（按演讲顺序）.....	13
演讲嘉宾及演讲内容简介.....	14
人工智能化大数据分析.....	14
大数据认知与关键技术思考.....	14
基于电信网络的面向用户特征的大数据分析平台.....	15
大数据挖掘利器—R 语言.....	15
北京遇上西雅图 - 心理学遇上大数据.....	16
数据融合与信用风险评估.....	16
R 语言与地学建模：从函数到可视化.....	17
R 语言与互联网数据价值发现.....	17
兰州财经大学和平校区实景标识图.....	18

第十届中国 R 会议(兰州)暨西北地区数据科学会议交流与资讯



Rconference 兰州 微信公众号

中国 R 会议（兰州）QQ 咨询群



中国 R 会议（兰州）微信咨询群

R 语言简介

R 是用于统计分析、绘图的语言和操作环境，是属于 GNU 系统的一个自由、开源软件，用于统计计算和统计制图的优秀工具。它是一套完整的数据处理、计算和制图软件系统。其功能包括：数据存储和处理系统；数组运算工具（其向量、矩阵运算方面功能尤其强大）；完整连贯的统计分析工具；优秀的统计制图功能；简便而强大的编程语言；可操纵数据的输入和输出，可实现分支、循环，用户可自定义功能。

与其说 R 是一种统计软件，不如说 R 是一种数学计算的环境，因为 R 并不是仅提供若干统计程序，使用者只需指定数据库和若干参数便可进行统计分析。R 的思想是：它可以提供一些集成的统计工具，但更大量的是它提供各种数学计算、统计计算的函数，从而使使用者能灵活机动的进行数据分析，甚至创造出符合需要的新的统计计算方法。

第十届中国 R 会议(兰州)筹备委员会

主 席：陈鑫鹏

总干事：张 慧

干 事：范博凯

组 长：徐湘钰 罗家鑫 马乾 夏丽丽 崔亚奇

组 员：高思梦 张梦瑶 李辉 杨晨茜 任巧丽

魏益琼 藏 媛 李娜 李雪超 马爱琴

钟添添 杨燕燕 李娇 张晓娟 刘 畅

机构简介

兰州财经大学简介

兰州财经大学肇始于 1952 年成立的甘肃省人民政府行政干部学校；1958 年，升格为甘肃财经学院，隶属甘肃省人民政府领导，实施本科教育；1981 年，成立兰州商学院，先后隶属原国家商业部、国内贸易部领导；1998 年，实行中央与地方共建、以甘肃为主的管理体制；2003 年，增列为硕士学位授权单位，开展研究生教育；2015 年，更名为兰州财经大学，是黄河上游甘青宁三省（区）唯一一所财经类普通高等学校。

学校设有会计、金融、统计、工商管理、经济、国际经济与贸易、信息工程、财税与公共管理、法学、马克思主义、商务传媒、外语、艺术、农林经济管理、国际教育、继续教育等 16 个学院以及体育教学部、MBA 教育中心、经济管理实验教学中心（大学生创新创业教学中心）等 3 个教学部（中心）；学校现有教职员工 1300 余人，其中专任教师 1007 人，具有教授、副教授职称教师 681 人，博士、硕士研究生学历教师 715 人，专任教师中享受国务院政府特殊津贴专家、省部级优秀专家、教育部新世纪优秀人才支持计划入选者、甘肃省领军人才、甘肃省“333”“555”人才工程人选、甘肃省学科带头人、甘肃省高等学校教学名师等高层次人才 90 余人（次）。

兰州财经大学统计学院简介

兰州财经大学统计学科建设始于 1985 年，以计划统计本科专业为起点，依托统计学院展开。统计学科自 1997 年起先后五轮顺利入选甘肃省重点学科，并于今年进入甘肃省一流特色学科建设行列，现拥有硕士一级授权资格、应用统计硕士专业学位培养资格，支持建成了“甘肃省电子商务技术与应用重点实验室”和甘肃省高校人文社会科学重点研究基地“甘肃经济发展数量分析研究中心”。

兰州财经大学统计学是甘肃省飞天学者设岗学科、统计学院是中国高等教育学会高等财经教育分会统计学专业协作组主任委员单位、“甘肃省大数据产业技术创新联盟”发起和理事单位，与国内主流统计学科建立了畅通的联系渠道，并通过合作共享甘肃移动“网上丝绸之路大数据中心”与“西北中小企业云计算中心”、甘肃省科技情报研究所“文献资源平台”等社会资源，与实际统计业务部门建立了密切的合作关系。

统计之都简介

“统计之都” (Capital of Statistics, 简称 COS) 网站成立于 2006 年 5 月，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。

“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：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门户网站；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”的建设需要您的帮助，如果您从统计之都网站中获取了有价值的帮助。

甘肃省大数据产业技术创新联盟简介

甘肃省大数据产业技术创新联盟(<http://www.gs-bigdata.com/>)是受甘肃省工信委和甘肃省科学技术厅的委托，由产、学、研、用各界相关单位发起并筹备，依托中国科学院寒区旱区环境与工程研究所；甘肃省大数据技术产业创新联盟有效的整合产、学、研、用各方资源，充分发挥其自身优势，探索建立长效稳定的大数据产学研用合作机制，突破大数据产业发展的核心技术，形成大数据产业技术标准，搭建有效的大数据技术产业合作交流平台，培养大数据技术产业人才。

经过近 30 年的信息技术发展和应用，已经形成了具有特色的科学数据资源，交通、金融、国土等各行业数据资源，形成大数据有关的基础设施，开展应用的云计算支持平台，形成了大数据应用的初步技术体系，形成了大数据的初期应用示范，围绕数据价值链形成了一批企业，为我省大数据产业发展奠定了基础。

知几基因简介

“GenePlan（知几基因）”专注于提供与基因科学有关的健康类服务。创始团队于 2014 年组建。“GenePlan（知几基因）”的独立检测研究室设于成都，同时与国内数家基因检测机构、分析及医疗机构合作。

“GenePlan（知几基因）”现提供三种类型的基因体检服务。分别为“专业版”、“深度版”和“旗舰版”。与传统体检相比，基因体检可以在任何时间、地点独立完成。基于基因检测技术，可以比传统体检检测出更多遗传病风险，及早通过基因点位变异，发现肿瘤的早期风险。同时结合靶向用药，对癌症临床治疗起到显著的效果。

除此之外，“GenePlan（知几基因）”还运营公众号“知几基因研究院”，以及积极与地方医疗机构合作，开展基因检测与体检健康相关的合作。

会务信息

从车站、机场抵达兰州财经大学段家滩校区路线规划

①机场下车公交线路：先乘坐机场巴士城关区线（1 站）：兰州中川机场上车—兰州东方大酒店下车。换乘 110 路公交（7 站）：兰州大学上车—段家滩。步行 280 米到达兰州财经大学段家滩校区。

②火车站下车公交线路：在兰州车站可坐 110 公交（9 站）：兰州车站上车—段家滩。步行 280 米到达兰州财经大学段家滩校区。

③火车西站下车公交线路：1 路车（7 站）：西客站东出口站上车—南关十字下车。换乘 138（13 站）：南关十字上车—省第三人民医院（段家滩）下车。步行 160 米到达兰州财经大学段家滩校区。

9 月 16 日早上 07:30—08:00 参会者在兰州财经大学段家滩校区小喷泉广场集合乘车前去会场

会议须知

1. 会议期间请随身佩戴代表证，凭代表证出入会场。
2. 报告期间请将您的手机关闭。
3. 请您根据会议各项安排，按时集中会合，按照工作人员的引导进行活动。
4. 会议期间，我们为您提供全天服务。

联系人：范博凯 17693183095

陈鑫鹏 18317666055

会议日程

第十届中国 R 会议(兰州)暨西北地区数据科学会议日程

日期	时间	内容	嘉宾	单位	主持人	
9.15	14:00-18:00	参会者报到、领取会议资料				
9.16	07:30-08:00	参会者段家滩校区小喷泉集合乘车前去会场				
	08:00-09:00	参会者入场				
	09:00-09:10	兰州财经大学校领导致欢迎辞			庞智强	
	09:10-09:20	兰州财经大学统计学院院长庞智强教授致辞				
	09:20-09:30	林祯舜博士代表统计之都致辞				
	09:30-10:00	大会合影				
	10:00-10:30	人工智能化大数据分析	陈江教授	中国人力资源与社会保障部	林祯舜	
	10:30-11:00	大数据认知与关键技术思考	张耀南主任/研究员	中国科学院西北生态环境资源研究院		
	11:00-11:10	休息				
	11:10-11:40	基于电信网络的面向用户特征的大数据分析平台	李吉元高级 IT 技术专家	中国移动甘肃有限公司	田茂再	
	11:40-12:10	大数据挖掘利器——R 语言	谢佳标数据挖掘专家	平安人寿		
	12:10-14:10	午餐及午休 午餐地点：兰州财经大学和平校区教职工餐厅（清真餐厅在学生餐厅一楼）				
	14:10-14:40	北京遇上西雅图——心理学遇上大数据	翁称誉高级经理	微软(中国)	韩海波	
	14:40-15:10	数据融合与信用风险评估	成慧敏博士	中央财经大学		
	15:10-15:20	休息				
15:20-15:50	R 语言与地学建模：从函数到可视化	罗立辉高级工程师	中国科学院西北生态环境资源研究院	韩君		
15:50-16:20	R 语言与互联网数据价值发现	黄恒君博士	兰州财经大学			
16:20-16:30	第十届中国 R 会议（兰州）暨西北地区数据科学会议学生组委会主席陈鑫鹏发言					
16:30-17:00	参会者致远楼停车场集合乘车返回兰州财经大学段家滩校区					

嘉宾职务及所在单位（按演讲顺序）

陈 江	教授	中国人力资源和社会保障部
张耀南	主任/研究员	中国科学院西北生态环境资源研究院
李吉元	高级 IT 技术专家	中国移动甘肃有限公司
谢佳标	数据挖掘专家	平安人寿
翁称誉	高级经理	微软(中国)
成慧敏	博士	中央财经大学
罗立辉	高级工程师	中国科学院西北生态环境资源研究院
黄恒君	博士	兰州财经大学

演讲嘉宾及演讲内容简介

人工智能化大数据分析

陈江¹

中国人力资源和社会保障部 教授

从经济的角度出发，数据分析可以从宏观和微观两个方面着眼；从用户的角度出发，数据应用可聚焦在 ToB（企业机构），ToC（个人客户）和 ToS（社会）。从这个层面来看，当下大数据行业应用最为广泛的精准营销、客户画像以及金融方面的个人征信等，其实是仅仅针对微观 ToB 和 ToC 的层面。本报告以经济金融领域为例探讨人工智能和大数据分析结合的一些问题和推进。

大数据认知与关键技术思考

张耀南²

中国科学院西北生态环境资源研究院 主任/研究员

通过举例说明大数据的应用来认识大数据各环节的价值体现，从而说明在发挥大数据价值中存在的 key 技术问题，并概括出大数据价值链所涉及的关键技术，并对 key 技术的发展做一些展望。

¹陈江，美国雪城大学管理学院财务金融系和 Kiebach International Business Research Center 的荣誉退休教授，目前担任美国 Scientific Computing Associates Corp. 的首席数据科学家、中国人力资源和社会保障部智力引进专家，参与指导政策仿真实验室的建设和运行，另外是山东省科学院战略所经济产业研究室的智库人才专家，是国内在学术及应用方面都经验丰富的跨领域专家。

²张耀南，甘肃省政协常委，九三学社甘肃省委副主委，中国科学西北生态环境资源研究院大数据中心主任、所级公共技术服务中心主任，甘肃省资源环境科学数据工程技术研究中心主任，甘肃省大数据产业技术创新联盟秘书长，中科院超级计算兰州分中心主任，甘肃省高性能网格计算中心主任。主要从事基于“数据—模型—计算”的地学 e-science 研究，基于大数据的地学模型表征研究，数据科学与数据工程研究。

基于电信网络的面向用户特征的大数据分析平台

李吉元³

中国移动甘肃有限公司 高级 IT 技术专家

甘肃移动高级 IT 技术专家，从事经营分析及大数据相关技术工作多年。演讲内容主要是：电信运营商大数据相关工作介绍，包括数据源，大数据处理技术，大数据系统架构，大数据应用等。

大数据挖掘利器--R 语言

谢佳标⁴

平安人寿 数据挖掘专家

R 语言作为一个可进行交互式数据分析和探索的强大平台，起着举足轻重的作用，R 语言的免费开源使得很多公司用来它处理数据、展示数据、分析数据、完成模型。本主题将快速带领大家体验 R 语言在数据挖掘方面的便捷及强大，最后对基于 R 的企业级大数据分析平台--Microsoft R Server 各功能进行介绍。

³李吉元，中国移动通信集团甘肃有限公司高级 IT 技术专家，从事经营分析及大数据相关技术工作多年。

⁴谢佳标，资深 R 语言用户，有十年以上数据挖掘工作实战经验，多次在中国 R 语言大会上作主题演讲。撰写书籍《R 语言与数据挖掘》、《数据实践之美》和《R 语言游戏数据分析与挖掘》等。

北京遇上西雅图 - 心理学遇上大数据

翁称誉⁵

微软（中国） 高级经理

人有一双眼，一眼回顾过去，一眼放眼未来，现今人人都知道大数据是目前企业最宝贵的资产，数据之所以有价值，古老的历史都已经印证。翁经理的演讲分六部分：①美国与中国不同的地方；②最古老的大数据应用；③心理学在大数据的实践；④数据的价值在哪里？数据创造解决方案还是解决方案决定数据来源；⑤最性感的解决方案；⑥未来的发展方向。

数据融合与信用风险评估

成慧敏⁶

中央财经大学 博士

关于面向小微商户以及个人消费的小微信贷是当前互联网金融的重要发展方向，并且正在经历爆发式增长。在这个增长过程中，如何在没有实物抵押的情况下，通过互联网大数据分析，实现快速准确征信是一个非常重要的问题。为此，不同的数据都可以做出一定的贡献。研究过追踪用户手机行为数据，利用 R 软件建立互联网征信的信用评分模型。研究表明，用户手机行为对于用户信用评估具有重大作用。本次分享主要基于考拉征信信用评分模型，介绍利用 R 软件进行模型设计和数据处理的一般流程。

⁵翁称誉，毕业于中欧国际工商学院 MBA，目前任职于 微软(中国) 有限公司,担任大数据与人工智能解决方案部行业解决方案高级经理，曾任职于苹果(中国)与仁宝电脑，是制造业与零售业，专案管理，与物联网与人工智能等领域有丰富的行业经验。

⁶成慧敏，中央财经大学统计与数学学院博士生在读，考拉征信联合研究组组长。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。

R 语言与地学建模：从函数到可视化

罗立辉⁷

中国科学院西北生态环境资源研究院 高级工程师

以其最新开发的 PIC (Permafrost Indices Computing) 模型为例，从简单的方程构建到复杂的多层次的函数计算，在从多点到空间的模拟、统计分析和可视化，系统的介绍基于 R 语言的地学建模。

R 语言与互联网数据价值发现

黄恒君⁸

兰州财经大学统计学院 博士

分享主要结合科研、教学实例，说明 R 语言在互联网数据（包括数字、文本、地理信息和图像）采集与分析中的应用。

⁷罗立辉，中国科学院西北生态环境资源研究院副研究员、中欧联合培养博士、硕士生导师。先后主持了国家自然科学基金青年基金、中科院青年人才成长基金、中国博士后面基金、中国博士后特别资助等项目。在国际和国内杂志发表论文 10 多篇，并担任《Science of the Total Environment》、《科研信息化技术与应用》等多个期刊的审稿人。指导的硕士研究生开发的“机器人艺术家”获得了 2016 年微创客高校挑战赛三等奖，开发的多套科学计算软件在 Github 上共享。

⁸黄恒君，经济学博士，兰州财经大学统计学院副院长，副教授，硕士生导师，入选陇原创新人才扶持计划；中国统计教育学会理事、高等教育分会常务理事；甘肃经济发展数量研究中心特聘研究员。研究兴趣为异源异构数据整合、聚类算法、函数型数据分析，应用领域主要涉及统计名录库构建及其数据质量问题。

兰州财经大学和平校区实景标识图

