



## 北京理工大学图书馆“大数据创新学习中心”活动通知

### 一、宗旨与理念

随着信息技术的发展，在社会、科学研究领域逐步形成了人-机-物和谐发展格局，其中的大数据分析遍及社会、科研、学习等各类活动中。虽然大数据分析离不开统计学与信息技术的紧密结合，但目前与数据挖掘相关的最新统计学方法与计算方法还没有得到有效的普及。为抓住大数据时代的脉搏，我们需要与数据挖掘的前沿团队或师资紧密合作，以全新的形式推进大数据挖掘最新理论及方法的学习与应用实践。

**理念：**为适应大数据研究与应用的需要，北京理工大学图书馆数据港湾中心开辟创新学习空间——“大数据创新学习中心”的，将长期开展大数据挖掘学习与交流活动，该活动通过组建“大数据分析研究小组”，师生共同参与统计分析、各类数据的数据挖掘、计算算法及商业智能应用的学习与研究，以提升广大师生的数据分析能力。

**宗旨：**该活动的开展，旨在解决数据分析的困惑：1) 统计学方法的迅速发展对计算方法、数据储存都提出了更高的要求，对计算机的依赖日趋显著，数据港湾中心将突破计算资源的硬件限制，提供性能更高的计算平台，以满足大数据分析的硬件需求；2) 为让师生尽快掌握现代统计技术、计算机技术与数据挖掘技术，“数据挖掘系列活动”将提供与国际、国内大师学习与交流机会，研讨各类大数据的数据分析理论、方法及其应用实践。

通过创新学习活动，一方面，提高师生自身数据分析能力，达到学为所用的目的；另一方面，为广大师生、科研工作者搭建合作交流平台。最终形成具有竞争力的数据分析与深度挖掘的团队，有能力承担相关的研发任务。

### 二、指导教师

该中心隶属于北京理工大学图书馆，整合校内外优势师资资源，邀请校内知名学者做“学习指导教师”，同时，还聘请国内外知名专家做学术研究与应用实践指导。

目前邀请的校外专家有：陈定庚（北卡罗来纳州大学教堂山分校杰出教授、美国统计协会 Fellow）、黄健华（美国德州农工大学统计系教授）、朱砂（英国牛津大学大数据研究中心主任）、漆桂林（东南大学教授）、刘超（北京航空航天大学副教授）、胡芳槐（上海海翼知信息科技有限公司研发及产品总监）、丁军（上海海翼知信息科技有限公司 CEO）、林荟（杜邦公司商业数据科学家）、陈堰平（雪晴数据网创始人）、王安（北京布本智能科技有限公司首席）、杨伟（中国中医科学院助理研究员）。

图书馆专职组织与管理教师：逢金辉、李枫、赵霞、刘春涛等。

国外合作机构：美国北卡罗来纳州大学教堂山分校、英国牛津大学、美国德州农工大学等。

国内合作学校：中国人民大学、清华大学、北京航空航天大学、中国科学院大学。

国内合作机构：雪晴数据网、上海海翼知信息科技有限公司、北京清博大数据科技有限公司等。

### 三、适用人员

北京理工大学教师及科研工作者、博士、硕士、本科学生；北京地区其他高校教师、博士、硕士学生；北京地区各类研究机构及公司员工。

### 四、活动方式

以师生“学中用，用中思”、教学相长、共同进步为原则，以教师引导、学生参与、自主学习、共同实践为主要方式开展活动。

活动主要包括：

#### （1）统计与数学模型专题讨论

每周一次定期讨论，学生分小组报告统计分析、数据挖掘最新模型、算法及其应用、科学研究的数学方法及相关软件工具的学习与使用等。

#### （2）计算机领域知识学习

由专职老师指定学习内容、参考资料，学生自主学习，定期讨论，共同进步。

#### （3）大数据分析操练

友情提供大数据数据源，方便大家下载数据，用学习到的知识进行分析实践。

#### （4）大数据系列主题讲座

每个月邀请数据挖掘领域的资深专家做大数据系列讲座，欢迎大家前来聆听。

#### （5）知识工程专题学习

知识表达、知识推理方法、知识挖掘技术、知识工程的文本挖掘与情报信息抽取、知识工程的应用与技术开发等；基于对象计算的知识揭示和组织；大规模数据抽取方法,知识图谱化方法,科学数据的知识图谱、知识图谱的产业应用等。

#### (6) 专业领域大数据分析研讨

邀请专业领域数据分析专家,紧密结合专业领域应用实例,以讲座、咨询、培训和专题讨论形式解答学员数据分析的疑难问题,有效提升学员解决复杂问题的能力。

#### (7) 大数据前沿理论与方法培训

根据研讨学习的需要,适当邀请业界知名专家进行短期培训,以实现快速提高师生的数据分析基础,该模块为收费活动,费用由参加培训学员均摊。

#### (8) 高水准英文论文写作活动

为保障学习研讨所形成的成果形成国际高水准的学术论文,特别邀请 Juelich Torsten 博士参与中心活动,传授英文文献阅读方法、高水准论文写作授课等,该活动模块所产生的授课费用由参加听课的学员均摊(注:每课时授课费 1200 元)。

#### (9) 大数据挖掘新技术方法及应用的竞赛活动

根据学员的学习需要,适当组织数据分析竞赛活动。

### 四、学术研究

将与国际、国内各高校、知名公司密切合作,在数理统计理论、政府统计、商务统计、生物统计、大数据技术等领域进行科研探索,促使广大师生、科技工作者掌握灵活机动的数据分析技能,适应大数据时代发展需要。

### 五、活动开始仪式

时间地点: 2016 年 12 月 29 日, 15: 00—17: 00 在北京理工大学图书馆报告厅

领导讲话: 龙腾教授(北京理工大学校长助理)

黄河燕教授(北京理工大学计算机学院院长)

唐胜景教授(北京理工大学研究生院副院长)

徐文国教授(北京理工大学图书馆馆长)

主题报告: 黄河燕教授主持

#### 1) 报告一: 知识工程与大数据分析

报告人: 漆桂林博士, 东南大学教授

主要内容: 知识工程是人工智能的重要分支之一,知识的获取和基于知识的推理能力是人类高级智能的体现。知识工程涉及到机器学习、知识表示与推理、自然语言处理和数据存储等技术。本报告将介绍知识工程的发展及应用,主要包括

以下方面：1) 知识工程发展历程；2) 知识表示与推理；3) 知识挖掘与知识融合；4) 知识工程应用案例等。

## 2) 报告二：市场营销中的非结构数据分析

报告人：林荟博士，杜邦公司（美国）商业数据科学家

主要内容：介绍非结构数据分析在传统企业市场营销扮演的重要角色：1) 如何建立从自动化的数据获取（API，网页爬取），安全构架，数据分析到结果交付整个生态系统？2) 如何使用 R 进行关键数据分析（会提供数据和可重复的代码）？3) 如何进行有效的跨团队合作（数据科学家，数据工程师以及市场营销人员）？

## 3) 报告三：中文知识图谱构建的方法研究及产业化探索

报告人：胡芳槐博士，上海海翼知信息科技有限公司 创始人 研发总监

主要内容：多源异构数据如何构建中文知识图谱的方法研究及知识图谱的产业化探索：1) 知识图谱背景介绍 2) 如何构建中文知识图谱，有哪些方法？3) 知识图谱的产业化应用探索等。

## 六、报名方式及名额

广大师生自愿，报名邮箱：pangjinhui1@126.com

报名方式：通过邮箱 pangjinhui1@126.com 报名，报名信息包括：姓名、年龄、学习（或工作）单位及专业、联系电话。

人员数额限制：为保证大数据创新学习中心活动的质量，人数上限为 80 人。

## 七、免费资源

学习收费为免费（短期培训、英文写作授课除外），图书馆教室、研讨室、数据库资源使用免费。

## 八、活动周期及主题

每期活动主题由专家与学员共同商讨制定，每期活动时间跨度不超过 4 个月。

每周活动一次，地点在北京理工大学图书馆报告厅或多媒体教室（小分组自主讨论学习不限活动次数）。

## 九、校外顾问简介

陈定庚博士，美国统计协会 Fellow、Wallace Kuralt 杰出教授，北卡罗来纳大学教堂山分校统计发展咨询协会主席，曾任罗切斯特大学教授、佐治亚南方大学的 Karl E. Peace 著名讲座教授、美国生物制药和政府机构 Monte Carlo 模拟、临床试验统计学和卫生统计学领域的资深专家顾问。他发表论文 100 余篇，出版

临床试验方法、Meta 分析、因果推论和大数据高级统计方法等相关领域的专著 12 本。

**漆桂林**博士，东南大学教授、博导，担任中国中文信息学会语言与知识计算专业委员会副主任和中国计算机学会中文信息技术专业委员会专委委员。2006 年获英国贝尔法斯特女皇大学计算机科学博士学位，师从人工智能界著名专家 Weiru Liu 教授。2006 年--2009 年，在德国 Karlsruhe 大学 AIFB 研究所做博士后研究，师从语义 Web 界国际知名专家 Rudi Studer 教授。发表高质量学术论文 100 余篇，特别是在国际人工智能联合会议 (IJCAI)、AAAI 人工智能会议 (AAAI)、知识表示与推理会议 (KR)、不确定性推理会议 (UAI)、语义网会议 (ISWC) 发表论文 20 余篇。

**黄建华**博士，美国德州农工大学 (Texas A&M University) 统计系教授，中国人民大学统计与大数据研究院学科建设委员会主任。主要从事非参、半参、函数型数据、纵向数据、大维协方差矩阵估计、统计机器学习及应用统计方面的研究。

**朱砂**博士，英国牛津大学博士后，现为牛津大学大数据研究所研究中心主任，研究领域包括：构建系统发育树方法研究，种群遗传学，全基因组数据分析，生物信息软件开发，独立开发了生物信息软 (hybrid-coal, hybrid-Lambda, pfDwconv)。

**林荟**博士，现任杜邦公司商业数据科学家，曾经在 Iowa State University 先后担任兽医学院统计咨询师及商学院统计咨询师，研究兴趣在预测模型，机器学习，数据可视化，市场营销调查分析，消费者行为分析，自然语义处理和文本挖掘，健康与疾病统计分析等方面。

**陈堰平**，雪晴数据网创始人，主要从事统计咨询、数据分析、开发基于 R 语言的商业解决方案，曾给微软、招商银行、惠普、花旗银行、东方航空、中国电信做过培训和咨询。译作有《R 语言编程艺术》《实用数据分析》。

**丁军**博士，上海海翼知信息科技有限公司 CEO，长期从事知识图谱构建及应用方面研究，在国内最先提出专利垂直知识图谱构建及应用解决方案。在行业应用领域推行了全国企业商业知识图谱、中外创投知识图谱、海洋鱼类知识图谱、全国专利知识图谱，相关研究发表在国际知名会议 ISWC 上。

**王安**，北京布本智能科技有限公司首席分析官，曾任钱方银通科技有限公司首席分析师。中国科学技术大学理论力学学士，中国科学院计算力学硕士，北京大学光华管理学院 MBA。此前曾带领团队为多家大中型银行，保险公司提供风险管理，数据分析咨询服务。

**刘超**博士，人民大学统计学博士，现任北京航空航天大学数学与系统科学学院副教授，硕士生导师，曾主持国家社科基金青年项目《高维数据的稀疏主成分

分析及应用研究》、教育部人文社科青年项目《稀疏主成分分析的理论研究及其应用》等科研项目。

**杨伟**博士，中国人民大学统计学博士，现工作于中国中医科学院中医临床基础医学研究所，从事临床评价统计方法及大规模观察性数据挖掘方法的研究。目前已发表论文 24 篇，已参编学术著作 2 部，SCI 论文 2 篇，参与 1 项科技部重大专项研究，主持一项国家自然科学基金项目。

**胡芳槐**博士，上海翼知创始人，现任上海海翼知信息科技有限公司研发及产品总监，知识图谱和本体学习 5 年以上的研究以及产业化经验，多个相关方向的国家项目和上海市政府项目骨干成员国内最早研究中文知识图谱构建并进行产业化探索，并在中文图谱构建、机器学习方面在国际知名会议和期刊上发表多篇相关论文。