CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY

IT大咖说
知识共享平台

R语言构筑“人工智能+教育”应用服务

华东师范大学教育学部教育信息技术学系

吴 永 和 博士 研究员 博导

第十届中国R 语言会议（上海会场）

2017.12.2

教育信息技术学系
DEPARTMENT OF EDUCATION INFORMATION TECHNOLOGY

ercde² 華東師範大學
上海数字化教育装备工程技术研究中心
Shanghai Engineering Research Center of Digital Education Equipment

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY

IT大咖说
知识共享平台

抛砖引玉

R语言的教育应用系统构建与前景分析

华东师范大学教育学部教信系

吴 永 和 博士 研究员 博导

第九届中国R 语言会议（上海会场）

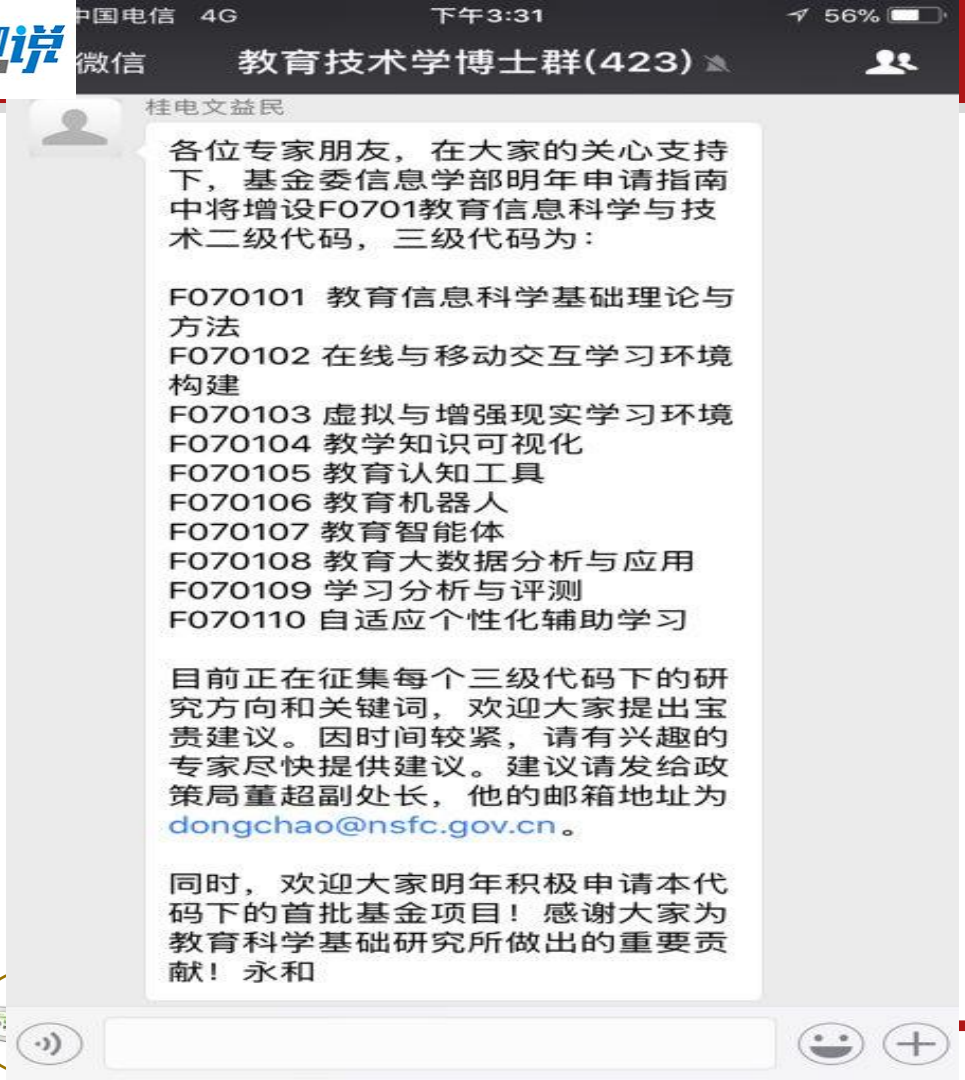
2016.12.10

教育信息技术学系
DEPARTMENT OF EDUCATION INFORMATION TECHNOLOGY

ercde² 華東師範大學
上海数字化教育装备工程技术研究中心
Shanghai Engineering Research Center of Digital Education Equipment

抛砖引玉

抛砖引玉


中国电信 4G
下午3:31
56%
微信
教育技术学博士群(423)
桂电文益民
各位专家朋友，在大家的关心支持下，基金委信息学部明年申请指南中将增设F0701教育信息科学与技术二级代码，三级代码为：
F070101 教育信息科学基础理论与方法
F070102 在线与移动交互学习环境构建
F070103 虚拟与增强现实学习环境
F070104 教学知识可视化
F070105 教育认知工具
F070106 教育机器人
F070107 教育智能体
F070108 教育大数据分析与应用
F070109 学习分析与评测
F070110 自适应个性化辅助学习
目前正在征集每个三级代码下的研究方向和关键词，欢迎大家提出宝贵建议。因时间较紧，请有兴趣的专家尽快提供建议。建议请发给政策局董超副处长，他的邮箱地址为dongchao@nsfc.gov.cn。
同时，欢迎大家明年积极申请本代码下的首批基金项目！感谢大家为教育科学基础研究所做出的重要贡献！永和

Intel 资助的“互联网+”教育课题

抛砖引玉

课题1：基于混合式学习的教和学行为数据分析指标体系及数据结构模型研究

分课题1：微课程与进阶练习学习行为分析

分课题2：信息化教学课堂教学行为的大数据分析

分课题3：作业与练习学生学习行为分析

分课题4：基于数字资源的在线学习行为分析

分课题5：学习社区学生交互行为大数据分析指标体系设计与实现

课题2：国内外教育大数据实践应用的比较研究

课题3：学生综合素质评价的数据采集、分析与评价

分课题6：学生综合素质评价的数据采集、分析与评价（浙江）

分课题7：学生综合素质评价的数据采集、分析与评价（上海）

抛砖引玉

数据驱动的精准学习研究

—— 一个团队行动研究的实例

华东师范大学教育学部教育信息技术学系

吴 永 和 博士 研究员 博导

第二届全国实证教育研究论坛（师生对话专场）

2016.10.29

教育信息技术学系
DEPARTMENT OF EDUCATION INFORMATION TECHNOLOGY

ercde² 華東師範大學
上海数字化教育装备工程技术研究中心
Shanghai Engineering Research Center of Digital Education Equipment

“三”个项目：浦江人才、上海科委、高峰高原

抛砖引玉

基于教育大数据的学习分析教育应用创新研究
（2014年9月）

- 教育大数据
- 学习分析
- 应用创新研究

面向移动互联网的自主学习系统研制及应用示范
（2014年6月）

- 移动化联网
- 自主学习系统
- 研制及应用示范

基于学习分析的学习资源精准服务研究
（2015年12月）

- 学习分析
- 学习资源
- 精准服务研究

提纲

Contents Page

R语言构筑“人工智能+教育”应用服务

一、人工智能+教育的发展趋向

二、R的机器学习包及教育应用

三、推进R语言的教育应用实践

人类自身的学习?

机器学习?

深度学习?

人类通过使用和创造工具，变得更聪明、更智慧

人工智能发展的七个阶段（时间维度）

AI（Artificial Intelligence）
发展历程

萌芽阶段	诞生阶段（1943-1956）	黄金阶段（1957-1973）	第一次低谷（1974-1980）	繁荣阶段（1981-1987）	第二次低（1988-199
1.1206，加扎利的可编程自动人偶。 2.1863，Samuel Butler的《机器中的达尔文》。 3.马略卡哲学家拉蒙·柳利开发了“逻辑机”。	1.1950，图灵测试。 2.1951,闵斯基和Dean Edmonds建造了第一台神经网络机(SNARC)。 3 .1955，Newell和Simon开发了“逻辑理论家”。 4.1956，达特茅斯会议：AI的诞生	1.1959，萨缪尔创造“机器学习”一词。 2 .1961 , James Slagle 开发SAINT。 3.1964，丹尼尔·鲍勃罗开发“STUDENT”。 4.1966，第一个通用型移动机器人Shakey。 5.1968年，Terry Winograd开发SHRDLU。 6.1973，英国发布莱特希尔人工智能调查报告。	1.1976，计算机科学家拉吉·瑞迪发表《Speech Recognition by Machine: A Review》。 2.1978，卡内基梅隆大学开发XCON。 3.1979，斯坦福大学自动驾驶汽车Stanford Cart。 4.1980，早稻田大学的Wabot-2机器人。	1.1981，日本“第五代计算机项目”研究。 2.1982，John Hopfield证明一种新型神经网络。 3.1986，第一辆无人驾驶奔驰汽车建造。 4.1986，Rumelhart等主编论文集“分布式并行处理”问世。 5.1987，约翰·斯卡利提到“Knowledge Navigator”。	1.1988，Judea Pea《Probabilistic Rea in Intelligent Systen 2.1988，罗洛·卡发Jabberwacky。 3.1988，IBM沃森心发表《A sta approach to la translation》。 4.1990，Rodney发表了《lephants Play Chess》。

1.1924，美国教育心理学家普莱西年试制出第一台用于测验的机器。

1.1954，斯金纳发表《学习的可续和教学的艺术》，推动了程序教学运动的发展。

1.1960，世界上第一个计算机辅助教学系统：PLATO系统。
2.1965年，费根鲍姆等开始研究历史上第一个专家系统DENDRAL系统。
3.1970，J.R.Carbonell提出智能型计算机辅助教学（ICAI）的构想。
4 . 1973 ， Hartley & Sleeman提出智能教学系统（ITS）的基本框架。

1.1975，Collins等人研制了教授学生探索降雨原因的根源的WHY系统；
2.1977，Wescourt等设计了辅助Basic语言教学的BIP系统；
3.1977，MIT开发用于逻辑学、概率、判断理论和几何学训练的WUMPUS游戏系统。

1.1982，Sleeman&Brown提出智能导师系统（ITS）概念。
2.1983，AISB组织第一个明确的AIED研讨会。
3.1984，梅瑞尔提倡教学设计自动化（AID）研究。
4.1987，开发用于辅助教学设计决策的专家系统原型ID Expert 系统。

1.1992，Brusilovsky 提出智能授导系统（Intelligent Tutoring Systems）ITEM/IP，这是一个支持学习与授导的整合智能系统。
2.1992，第一届美国人工智能学会移动机器人比赛。

1. 1996，Brusilovsky 等人早开发了第一个自适应教学系统。
2.2011，韩国教育科学技术部颁布《推进智慧教育战略》规划。
3.2011，首届学习分析技术与知识国际会议举行。
4.2013，麻省理工学院Ehsan Hoque等人研发社交技能训练系统MACH。

AIE （Artificial Intelligence in Education）
发展历程

JOURNAL OF DISTANCE EDUCATION / Frontier Discovery

构筑“人工智能 + 教育”的生态系统*

吴永和[1] 刘博文[1] 马晓玲[2]

(1.华东师范大学 教育学部 教育信息技术学系，上海 200062；
2.华东师范大学 经管学部 信息管理系，上海 200241)

[摘 要] 近期国内召开的人工智能教育论坛，进一步促进了“人工智能+教育”研究的深入和应用推广，同时也为人工智能(AI)向“人工智能+教育”发展提供了技术基础。AI 发展过程可分别从时间和技术两个维度进行划分，其在发展过程中不断与教育结合，产生了大量的研究成果，这为“人工智能+教育”提供了理论指导。通过查新关键词云图和共词网络图分析进程了目前的研究现状，认为国内外各企业教育智能产品应用，是推进“人工智能+教育”的动力；国家颁布了一系列人工智能发展战略政策，特别强调教育应用，为“人工智能+教育”提供了政策保障，进而从学育条件、特征、作用三方面阐述了“人工智能+教育”的内涵，从应用形态、技术架构、业态趋向等要素构建了“人工智能+教育”的生态系统，并阐述了“人工智能+教育”的人才培养体系。

[关键词] 人工智能；人工智能+教育；智能教育；教育人工智能；EAI；教育智能代理；教育 Agent；生态系统；人才培养
[中图分类号] G40-057 [文献标识码] A [文章编号] 1672-0008(2017)05—0027—13

一、研究的提出

人工智能(Artificial Intelligence，简称 AI)是伴了人类的体力和脑力。“阿尔法狗”先后战胜二位世界围棋冠军李世石、柯洁，掀起了全球性人工智能发展的新一轮热潮。2016 年 5 月 23 日，国家发改委联合科技部、工信部等共同制定了(“互联网+”人工智能三年行动实施方案)[1]，首次明确提出到 2018 年基本建立人工智能产业、服务和标准化体系，实现核心技术突破，培育若干全球领先的人工智能骨干企业，形成千亿级的人工智能市场应用规模。2017 年 3 月 5 日，国务院总理李克强在(政府工作报告)中提到，要加快培育壮大新兴产业，全面实施战略性新兴产业发展规划，包括人工智能产业。这是“人工智能”首

7 月 8 日，国务院印发了《新一代人工智能发展规划》[4]，为加快人工智能的创新应用，提出了“智能教育”，为“人工智能+教育”指明了发展方向。

近年来，国内一些学者纷纷展开了相关研究。如，闫志明等提出教育人工智能(EAI)，界定了教育人工智能的内涵[5]；马玉慧等研究了智慧教育时代我国人工智能教育应用的发展路径[6]；刘清堂等对智能教学技术的发展与展望进行了研究[7]。可见，教育作为人工智能的主要应用领域之一，正向“人工智能+教育”的新阶段迈进。

那么，在今天这样一个人工智能快速发展的时代背景下，“人工智能+教育”的发展现况与重点如何？其生态系统如何构建？本文拟从人工智能的发展

http://dej.zjnu.edu.cn

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

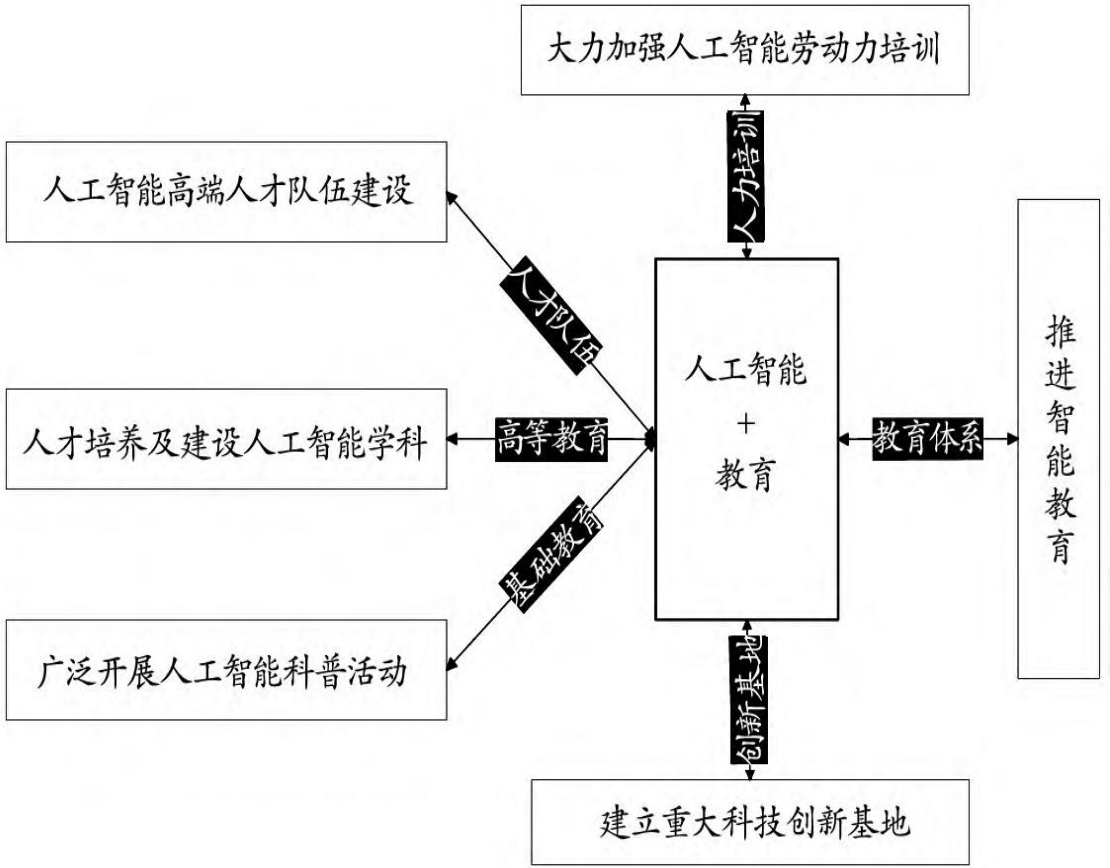
人工智能发展的三个阶段（技术维度）

发展阶段	特征	教育应用案例
计算智能	能存会算	存储与传递海量学习资源、智能学生信息管理系统
感知智能	能听会说、能看会认	语言教学、口语测评、图像搜题
认知智能	能理解、会思考	个性化学习、自主学习

中国颁布的与人工智能相关的一系列政策

时间	颁布主体	法律法规	相关内容	关键词频次
2015.5	国务院	中国制造 2025	提出加快推动新一代信息技术与制造技术融合发展，把智能制造作为两化深度融合的主攻方向。研究制定智能制造发展战略；加快发展智能制造装备和产品；推进制造过程智能化；深化互联网在制造领域的应用。	“智能”一词出现 74 次。
2015.7	国务院	国务院关于积极推进“互联网+”行动的指导意见	“互联网+”人工智能是 11 个重点行动之一，提出“依托互联网平台提供人工智能公共创新服务，加快人工智能核心技术突破，促进人工智能在智能家居、智能终端、智能汽车、机器人等领域的推广应用”。	“智能”一词出现 71 次。“人工智能”一词出现 9 次。
2016.1	国务院	“十三五”国家科技创新规划	在发展引领产业变革的颠覆性技术中，提到重点开发人工智能等技术。	“智能”一词出现 96 次；“人工智能”出现 5 次。
2016.3	国务院	国民经济和社会发展第十三个五年规划纲要	在信息化重大工程的“互联网+”行动里也提及人工智能。	“智能”出现 24 次；“人工智能”出现 3 次。
2016.5	国家发改委、科技部、工信部、中央网信办	“互联网+”人工智能三年行动实施方案	总体思路是“贯彻落实创新、协调、绿色、开放、共享发展理念，以提升国家经济社会智能化水平为主线”。人才培养是保障措施之一，要鼓励开展人工智能基础知识和应用培训，完善高校的人工智能相关专业、课程设置，注重人工智能与其他学科专业的交叉融合，鼓励校、所、企间开展合作，建设一批人工智能实训基地，支持人才赴海外交流。	“智能”一词出现 120 次。“人工智能”一词出现 54 次。
2017.7	国务院	新一代人工智能发展规划	规划将我国新一代人工智能发展的战略目标分三步走。提出“智能教育”，利用智能技术加快推动人才培养模式、教学方法改革，构建包含智能学习、交互式学习的新型教育体系。建立以学习者为中心的教育环境育服务，实现日常教育和终身教育定制化。	“智能”一词共出现 224 次。“人工

新一代人工智能发展规划——智能教育与人才培养



中华人民共和国中央人民政府
www.gov.cn

国务院 | 总理 | 新闻 | 政策 | 互动 | 服务

首页 > 信息公开 > 国务院文件 > 科技、教育 > 科技

索 引 号：000014349/2017-00142　　主题分类：科技、教育\科技
发文机关：国务院　　成文日期：2017年07月08日
标　　题：国务院关于印发新一代人工智能发展规划的通知
发文字号：国发（2017）35号　　发布日期：2017年07月20日
主 题 词：

国务院关于印发
新一代人工智能发展规划的通知

国发〔2017〕35号

各省、自治区、直辖市人民政府，国务院各部委、各直属机构：

现将《新一代人工智能发展规划》印发给你们，请认真贯彻执行。

国务院

2017年7月8日

（此件公开发布）

新一代人工智能发展规划

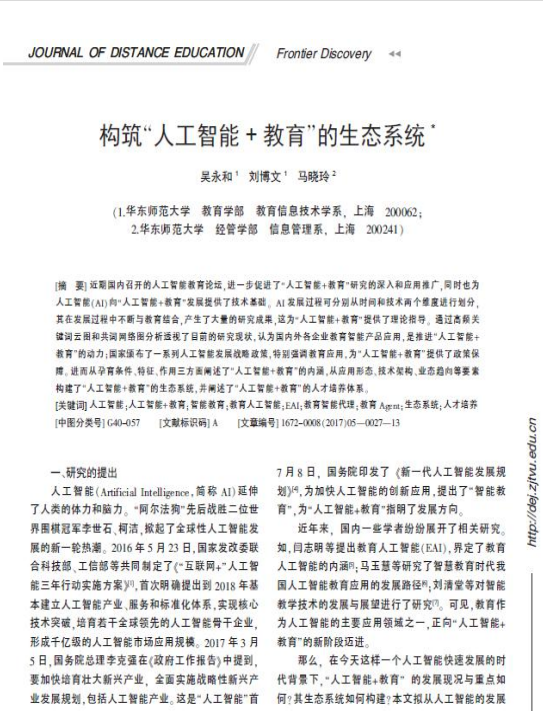
人工智能的迅速发展将深刻改变人类社会生活、改变世界。为抢抓人工智能发展的重大战略机遇，构筑我国人工智能发展的先发优势，加快建设创新型国家和世界科技强国，按照党中央、国务院部署要求，制定本规划。

一、战略态势

人工智能发展进入新阶段。经过60多年的演进，特别是在移动互联网、大数据、超级计

智能教育

利用智能技术加快推动人才培养模式、教学方法改革，构建包含智能学习、交互式学习的新型教育体系。开展智能校园建设，推动人工智能在教学、管理、资源建设等全流程应用。开发立体综合教学场、基于大数据智能的在线学习教育平台。开发智能教育助理，建立智能、快速、全面的教育分析系统。建立以学习者为中心的教育环境，提供精准推送的教育服务，实现日常教育和终身教育定制化。



JOURNAL OF DISTANCE EDUCATION Frontier Discovery

构筑“人工智能 + 教育”的生态系统*

吴永和[1] 刘博文[1] 马晓玲[2]

(1.华东师范大学 教育学部 教育信息技术学系，上海 200062；
2.华东师范大学 经管学部 信息管理系，上海 200241)

[摘 要] 近期国内召开的人工智能教育论坛，进一步促进了“人工智能+教育”研究的深入和应用推广，同时也为人工智能(AI)向“人工智能+教育”发展提供了技术基础。AI发展过程可分别从时间和技术两个维度进行划分，其在发展过程中不断与教育结合，产生了大量的研究成果，这为“人工智能+教育”提供了理论指导。通过高频关键词云图和共词网络图分析适视了目前的研究现状，认为国内外各企业教育智能产品应用，是推进“人工智能+教育”的动力；国家颁布了一系列人工智能发展战略政策，特别强调教育应用，为“人工智能+教育”提供了政策保障。进而从孕育条件、特征、作用三方面阐述了“人工智能+教育”的内涵，从应用形态、技术架构、业态趋向等要素构建了“人工智能+教育”的生态系统，并阐述了“人工智能+教育”的人才培养体系。

[关键词] 人工智能；人工智能+教育；智能教育；教育人工智能；EAI；教育智能代理；教育 Agent；生态系统；人才培养
[中图分类号] G40-057 [文献标识码] A [文章编号] 1672-0008(2017)05—0027—13

一、研究的提出

人工智能(Artificial Intelligence，简称 AI)延伸了人类的体力和脑力。“阿尔法狗”先后战胜二位世界围棋冠军李世石、柯洁，掀起了全球性人工智能发展的新一轮热潮。2016年5月23日，国家发改委联合科技部、工信部等共同制定了《“互联网+”人工智能三年行动实施方案》[1]，首次明确提出到2018年基本建立人工智能产业、服务和标准化体系，实现核心技术突破，培育若干全球领先的人工智能骨干企业，形成千亿级的人工智能市场应用规模。2017年3月5日，国务院总理李克强在《政府工作报告》中提到，要加快培育壮大新兴产业，全面实施战略性新兴产业发展规划，包括人工智能产业。这是“人工智能”首

7月8日，国务院印发了《新一代人工智能发展规划》[4]，为加快人工智能的创新应用，提出了“智能教育”，为“人工智能+教育”指明了发展方向。

近年来，国内一些学者纷纷展开了相关研究。如，闫志明等提出教育人工智能(EAI)，界定了教育人工智能的内涵[6]；马玉慧等研究了智慧教育时代我国人工智能教育应用的发展路径[8]；刘清堂等对智能教学技术的发展与展望进行了研究[7]。可见，教育作为人工智能的主要应用领域之一，正向“人工智能+教育”的新阶段迈进。

那么，在今天这样一个人工智能快速发展的时代背景下，“人工智能+教育”的发展现况与重点如何？其生态系统如何构建？本文拟从人工智能的发展

http://dej.zjtvu.edu.cn

“人工智能教育”高频关键词云图

JOURNAL OF DISTANCE EDUCATION Frontier Discovery

构筑“人工智能 + 教育”的生态系统*

吴永和[1] 刘博文[1] 马晓玲[2]

(1.华东师范大学 教育学部 教育信息技术学系，上海 200062；
2.华东师范大学 经管学部 信息管理系，上海 200241)

[摘 要] 近期国内召开的人工智能教育论坛，进一步促进了“人工智能+教育”研究的深入和应用推广，同时也为人工智能(AI)向“人工智能+教育”发展提供了技术基础。AI 发展过程可分别从时间和技术两个维度进行划分，其在发展过程中不断与教育结合，产生了大量的研究成果，这为“人工智能+教育”提供了理论指导。通过查新关键词主题和共词网络图分析该视域下目前的研究现状，认为国内外各企业教育智能产品应用，是推进“人工智能+教育”的动力；国家层面下一系列人工智能发展战略政策，特别强调教育应用，为“人工智能+教育”提供了政策保障，进而从孕育条件、特征、作用三方面阐述了“人工智能+教育”的内涵，从应用形态、技术架构、业态趋向等要素构建了“人工智能+教育”的生态系统，并阐述了“人工智能+教育”的人才培养体系。

[关键词] 人工智能；人工智能+教育；智能教育；教育人工智能；EAI；教育智能代理；教育 Agent；生态系统；人才培养
[中图分类号] G40-057 [文献标识码] A [文章编号] 1672-0008(2017)05—0027—13

一、研究的提出

人工智能(Artificial Intelligence，简称 AI)easy了人类的体力和脑力。“阿尔法狗”先后战胜二位世界围棋冠军李世石、柯洁，掀起了全球性人工智能发展的新一轮热潮。2016 年 5 月 23 日，国家发改委联合科技部、工信部等共同制定了(“互联网+”人工智能三年行动实施方案)[1]，首次明确提出到 2018 年基本建立人工智能产业、服务和标准化体系，实现核心技术突破，培育若干全球领先的人工智能骨干企业，形成千亿级的人工智能市场应用规模。2017 年 3 月 5 日，国务院总理李克强在(政府工作报告)中提到，要加快培育壮大新兴产业，全面实施战略性新兴产业发展规划，包括人工智能产业。这是“人工智能”首

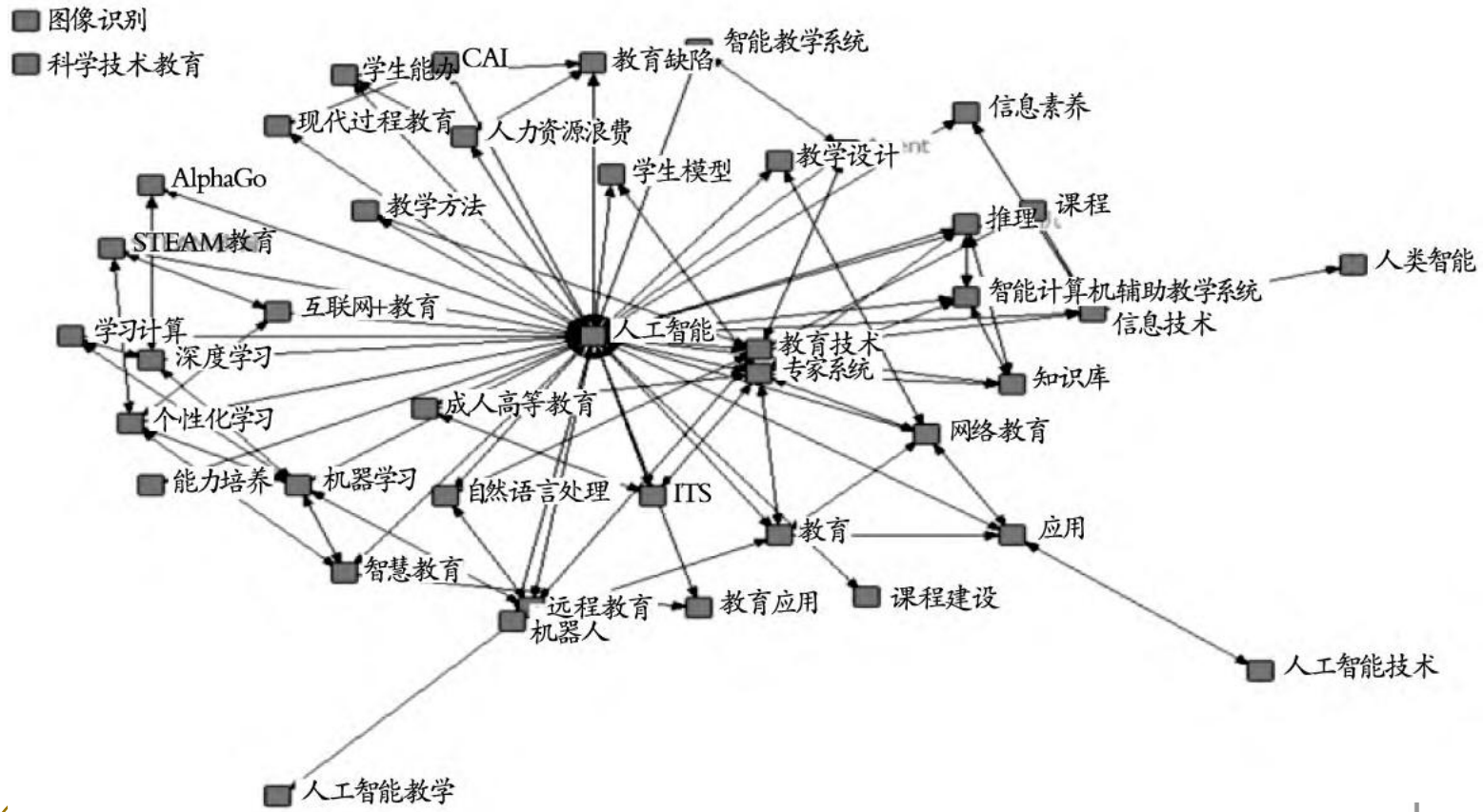
7 月 8 日，国务院印发了《新一代人工智能发展规划》[4]，为加快人工智能的创新应用，提出了“智能教育”，为“人工智能+教育”指明了发展方向。

近年来，国内一些学者纷纷展开了相关研究。如，闫志明等提出教育人工智能(EAI)，界定了教育人工智能的内涵[5]，马玉慧等研究了智慧教育时代我国人工智能教育应用的发展路径[6]；刘清堂等对智能教学技术的发展与展望进行了研究[7]。可见，教育作为人工智能的主要应用领域之一，正向“人工智能+教育”的新阶段迈进。

那么，在今天这样一个人工智能快速发展的时代背景下，“人工智能+教育”的发展现况与重点如何？其生态系统如何构建？本文拟从人工智能的发展

http://daj.zjhu.edu.cn

“人工智能教育”高频关键词共词网络图



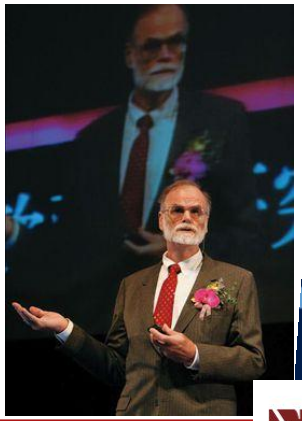
科学研究范式

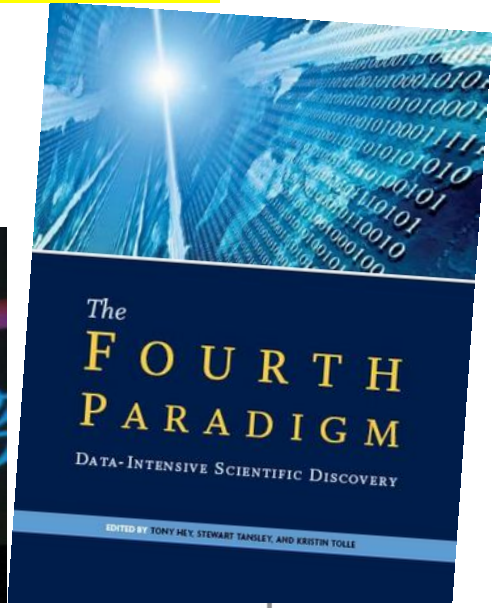
实证式
（科学研究）
第一范式

理论式
（理论推演）
第二范式

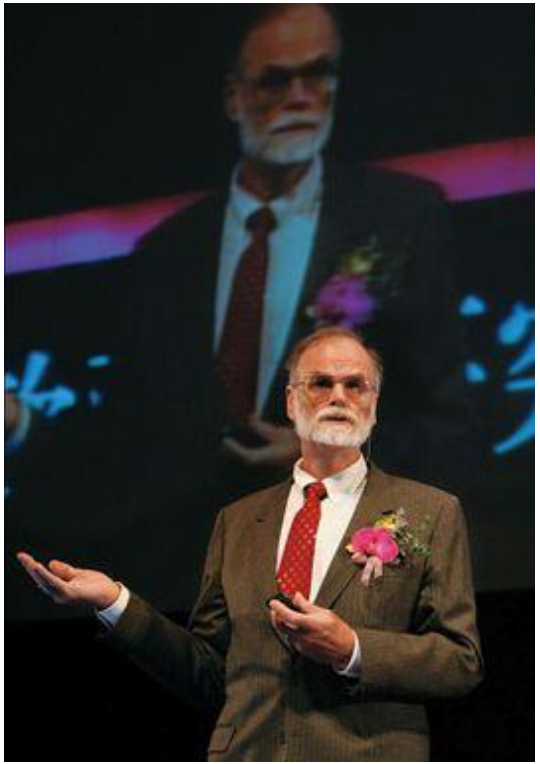
电脑计算
（计算机仿真）
第三范式

数据密集型科学
第四范式





研究范式：第四范式



几千年前的科学，以记录和描述自然现象为主，称为“**实验科学**”，**即第一范式**，其典型案例如钻木取火；

数百年前，科学家们开始利用模型归纳总结过去记录的现象，发展出“**理论科学**”，**即第二范式**，其典型案例如牛顿三定律、麦克斯韦方程组、相对论等；

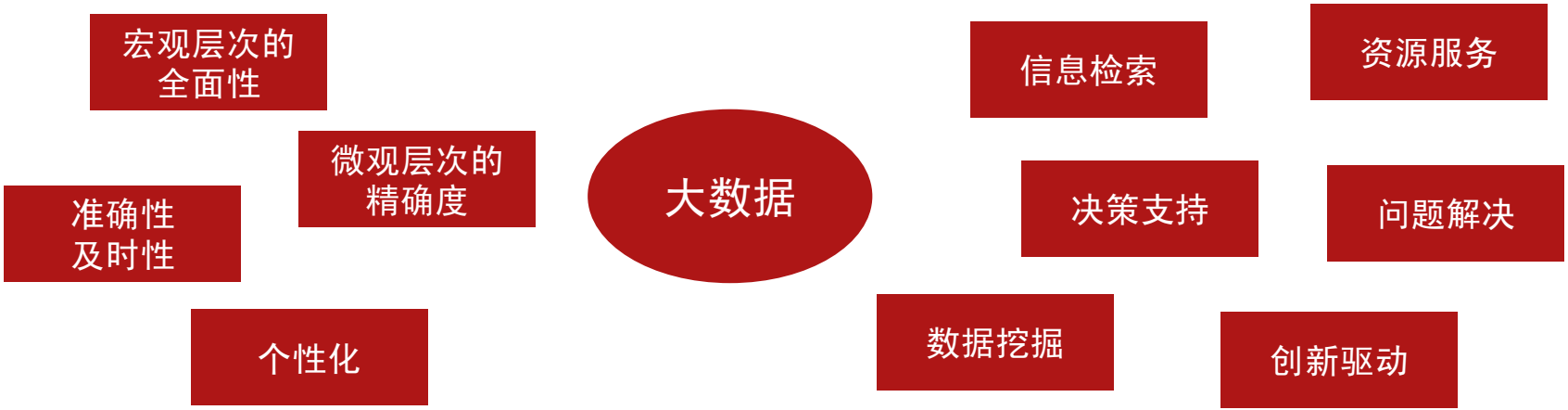
过去数十年，科学计算机的出现，诞生了“**计算科学**”，对复杂现象进行模拟仿真，推演出越来越多复杂的现象，其典型案例如模拟核试验、天气预报等；

Jim Gray认为今天，以及未来科学的发展趋势是，随着数据量的高速增长，计算机将不仅仅能做模拟仿真，还能进行分析总结，得到理论。也就是说，过去由牛顿、爱因斯坦等科学家从事的工作，未来可以由计算机来做。

Jim Gray将这种科学研究的方式，成为**第四范式，即数据密集型科学**。

图灵奖得主、关系数据库的鼻祖Jim Gray The Fourth Paradigm: Data-Intensive Scientific Discovery（《第四范式：数据密集型科学发现》）

大数据的价值



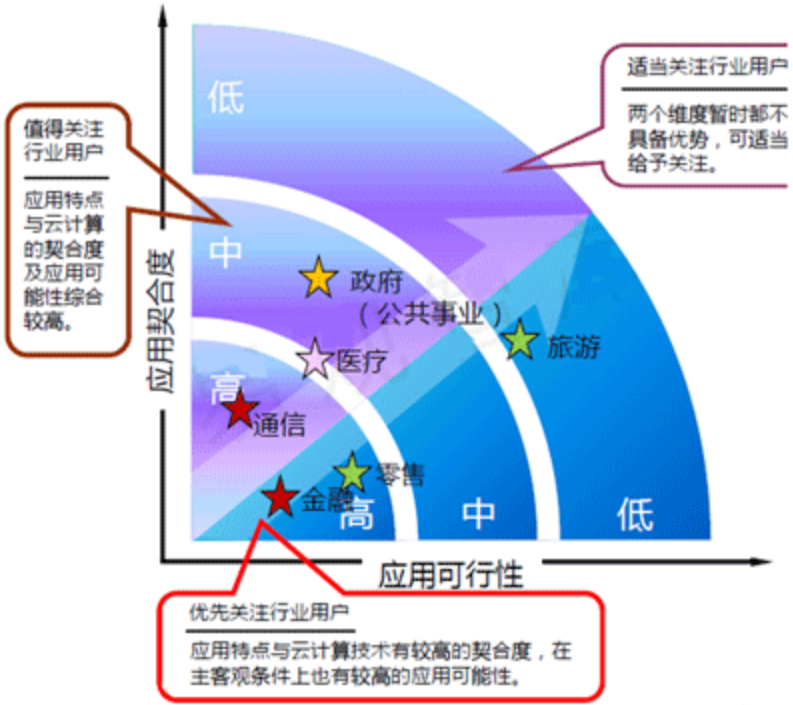
数据是基础性资源，也是重要生产力。 ——李克强

我们正迎来一个从IT（信息技术）到DT（数据技术）变革的新时代，IT时代把人变成了机器，而DT时代把机器变成了智能化的人。 ——马云

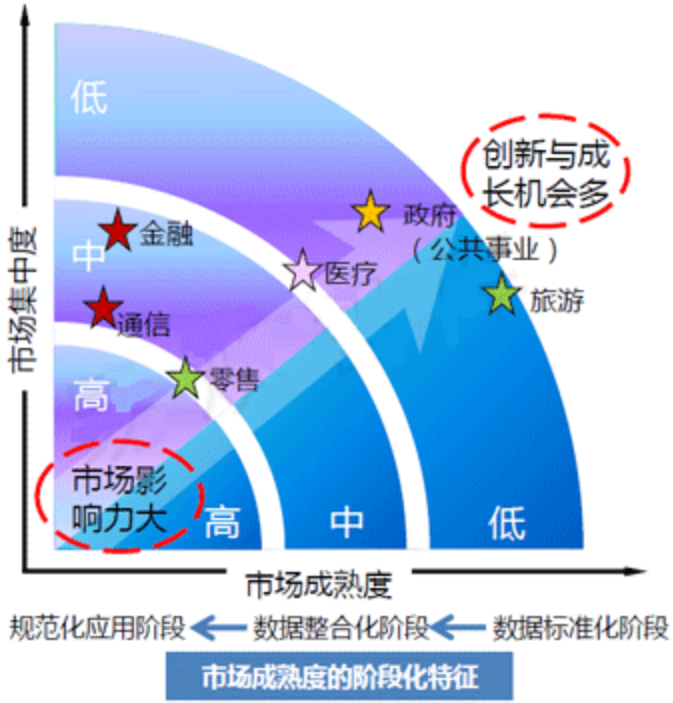
数据是国家基础性战略资源，是21世纪的“钻石矿”。——《大数据产业发展规划》

大数据、教育大数据与学习分析

大数据应用领域盘点



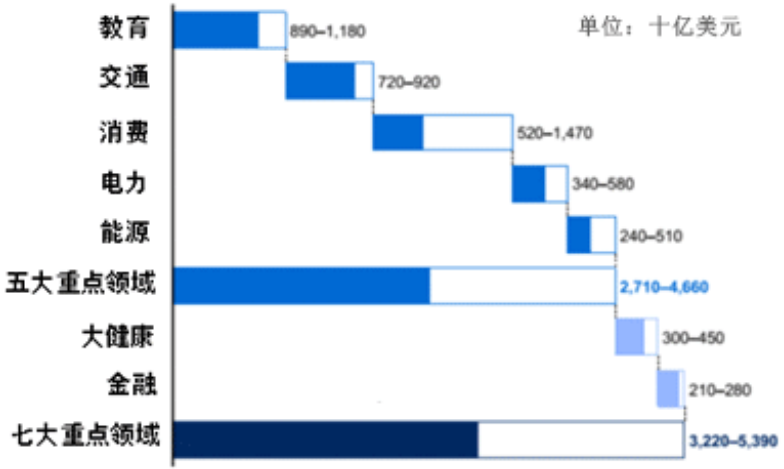
中国大数据市场行业契合度及应用可能分析



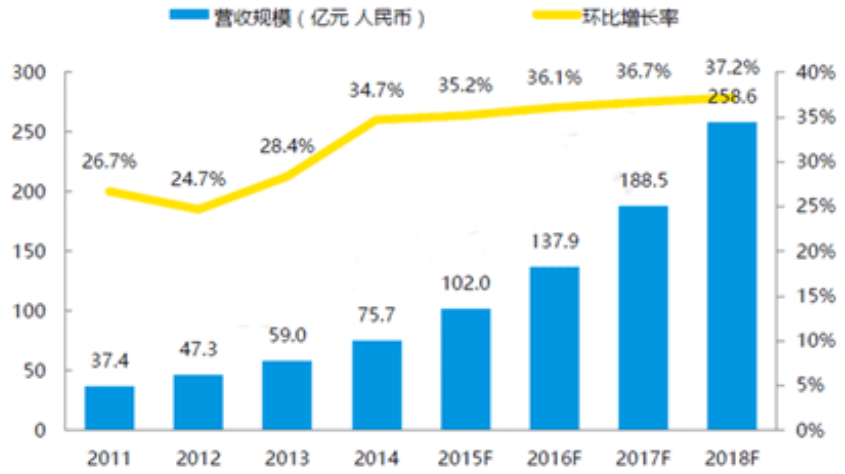
中国大数据市场集中度与成熟度分析

大数据产业规模

在全球七大重点领域内（包括教育、交通、消费、电力、能源、大健康以及金融），大数据的应用价值预计在32200-53900亿美元之间。



全球七大重点领域大数据应用潜在价值



2011-2018年中国大数据产业营收规模及增速

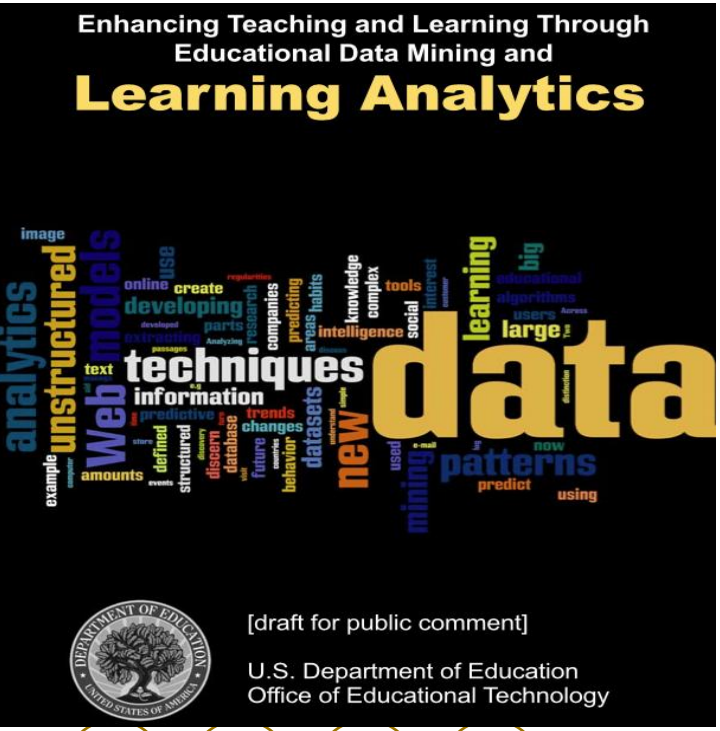
文献分析—中国知网：教育大数据-2017.10.1

华东师范大学 EAST CHINA NORMAL UNIVERSITY

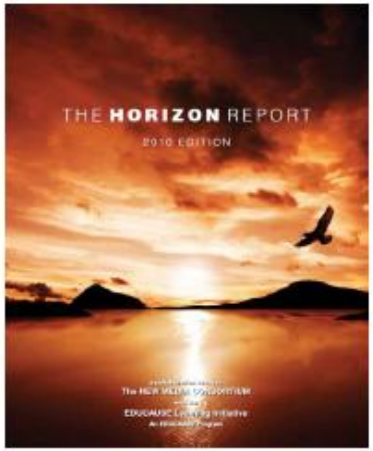
CAPITAL OF STATISTICS PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

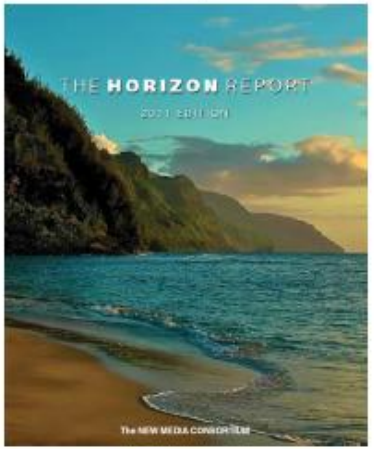
学习分析

Draft U.S. Department of Education Report released April 12,2012 http://1.usa.gov/GDFpnI



NMC Horizon Report





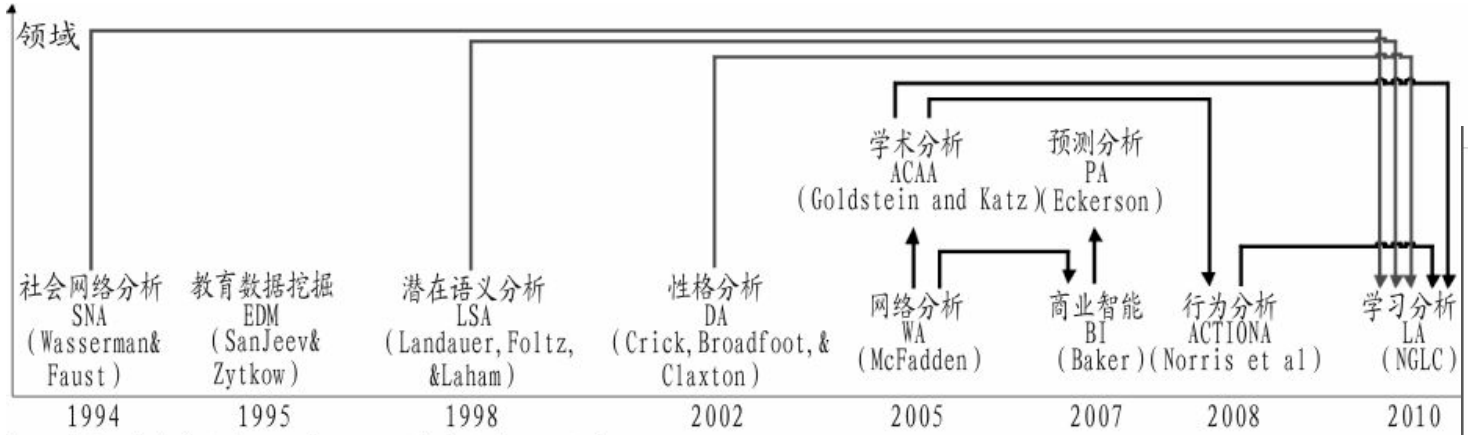
"I keep saying that the sexy job in the next 10 years will be statisticians," said Hal Varian, chief economist at Google. "And I'm not kidding."

Steve Lohr, NY Times, August 5, 2

地平线报告：学习分析from Horizon Report 2010 to 2017

Year	短期（1年以内）Short-Term(within a year)		中期（2到3年）Mid-Term(two or three years)		长期（4到5年）Long-Term(four or fives years)	
2017	创客空间	机器人	分析技术	虚拟现实	人工智能	物联网
2016	自带设备	学习分析(Learning Analytics)与自适应学习	增强与虚拟现实	创客空间	情感计算	机器人学
2014	翻转课堂	学习分析 Learning Analytics	3D打印	游戏与游戏化	自我量化	虚拟助手
2013	大规模在线开放课程	平板电脑	游戏与游戏化	学习分析 Learning Analytics	3D打印	可穿戴技术
2012	移动应用	平板电脑	基于游戏的学习	学习分析 Learning Analytics	基于手势的计算	物联网
2011	电子书	移动技术	增强现实	基于游戏的学习	基于手势的计算	学习分析 Learning Analytics
2010	移动计算	内容开放	电子图书	简易扩真现实	基于手势的计算	可视数据分析 Visual Data Analytics

学习分析溯源时间线



JOURNAL OF DISTANCE EDUCATION Theory Front

学习分析：教育信息化的新浪潮*

综述分析：《学习分析：教育信息化的新浪潮》（《远程教育杂志》2013.1 CSSCI）

CAPITAL OF STATISTICS PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

文献分析—中国知网—关键字：学习分析-2017.10.1

华东师范大学 EAST CHINA NORMAL UNIVERSITY

CAPITAL OF STATISTICS PROFESSION, HUMANITY & INTEGRITY

IT大咖说 知识共享平台

学习分析技术是测量、收集、分析和报告有关学生及其学习环境的数据，用以理解和优化学习及其产生的环境的技术，通过对不同情境下学生学习创造的大量数据的收集和推断来评价学习进步并及时发现潜在的问题。

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

R语言构筑“人工智能+教育”应用服务

提纲

Contents Page

一、人工智能+教育的发展趋向

二、R的机器学习包及教育应用

三、推进R语言的教育应用实践

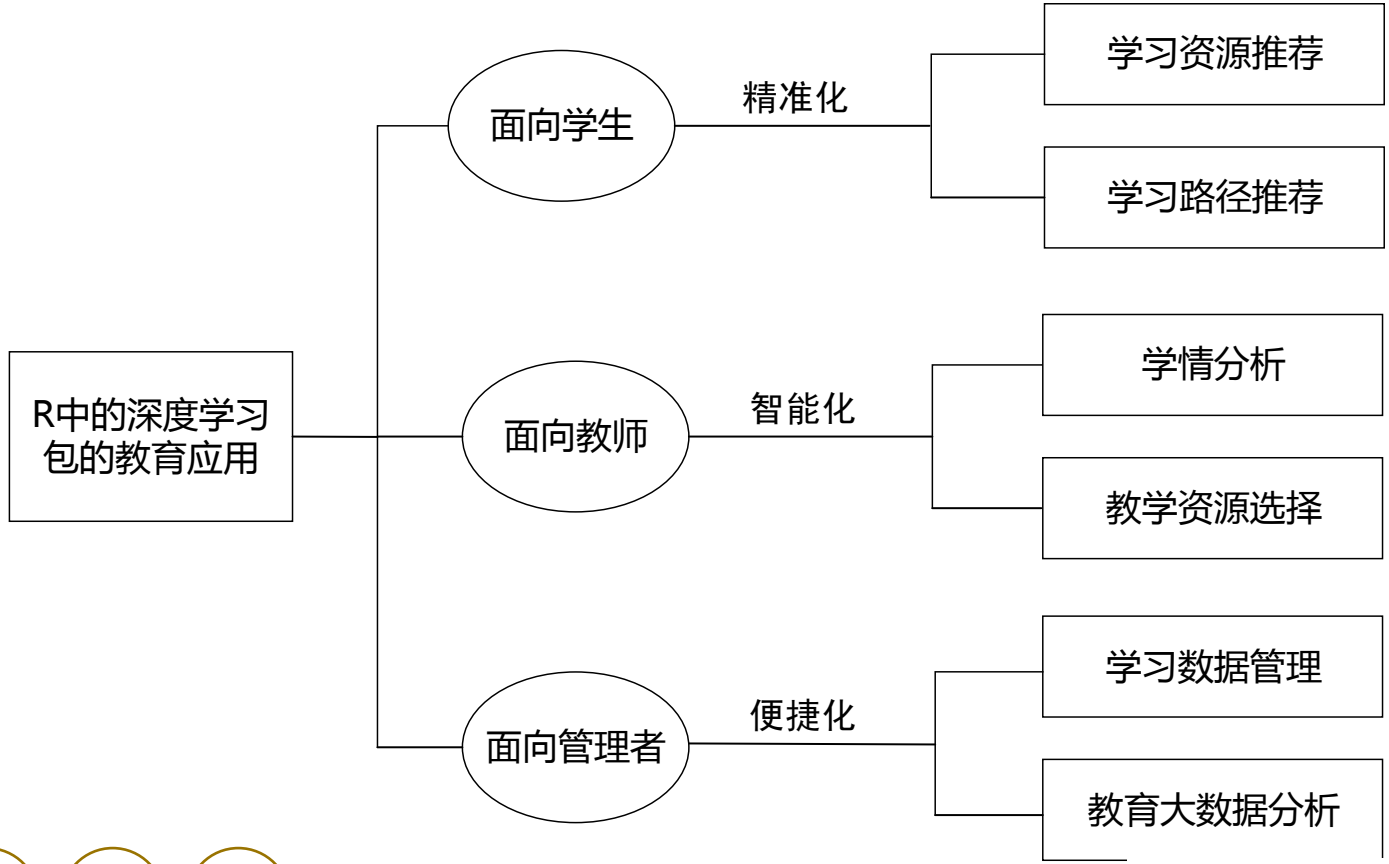
R应用中的机器学习包

中文名称	英文名称	个数
神经网络与深度学习	Neural Networks and Deep Learning	8个
递归划分	Recursive Partitioning	14个
随机森林	Random Forests	12个
正则化和收缩法	Regularized and Shrinkage Methods	25个
推进和梯度下降	Boosting and Gradient Descent	9个
支持向量机与核函数	Support Vector Machines and Kernel Methods	5个
贝叶斯方法	Bayesian Methods	2个
遗传算法优化	Optimization using Genetic Algorithms	3个
关联规则	Association Rules	2个
模糊规则系统	Fuzzy Rule-based Systems	2个
模型选择与验证	Model selection and validation	6个
其他程序	Other procedures	3个
综合的包	Meta packages	4个
统计学习要素	Elements of Statistical Learning	1个
用户图形界面	GUI	1个
其他	Others	2个
合计		99个

R中的深度学习包

R包的类型	R包的名称	R包的基本介绍
神经网络与深度学习 Neural Networks and Deep Learning	nnet	单隐藏层神经网络
	RSNNS	提供接口连接到斯图加特神经网络模拟器(Stuttgart Neural Network Simulator, SNNS)
	FCNN4R	提供接口连接到C++编写的开源库(Fast Compressed Neural Network, FCNN)，该库可用于人工神经网络的计算
	rnn	递归神经网络
	darch	限制玻尔兹曼机、深度信念网络
	deepnet	前馈神经网络、限制玻尔兹曼机、深度信念网络、堆叠自动编码机
	RcppDL	去噪自动编码机、堆叠去噪自动编码机、限制玻尔兹曼机、深度信念网络
	h2o	前馈神经网络、深度自动编码机

R中的深度学习与教育服务



华东师范大学 EAST CHINA NORMAL UNIVERSITY
CAPITAL OF STATISTICS PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

学习资源与学习路径推荐

与传统的浅层机器学习相比，深度学习具有优秀的自动提取抽象特征的能力，并且随着分布式计算平台的快速发展，大数据的处理能力得到极大的提升，使得近年来DL在工程界得到广泛的应用，包括图像识别，语音识别，自然语言处理等领域，并取得比传统机器学习更好的效果提升。另一方面，智能推荐系统，本质上是从一堆看似杂乱无章的原始流水数据中，抽象出用户的兴趣因子，挖掘出用户的偏好，因此把深度学习与推荐系统相结合成为近年来DL发展的一个新热点，事实上，Spotify，Netflix，Facebook，Google等公司早已经对如何把深度学习应用到推荐系统中有了很多深入研究，并在实际应用中取得了很好的效果提升 。

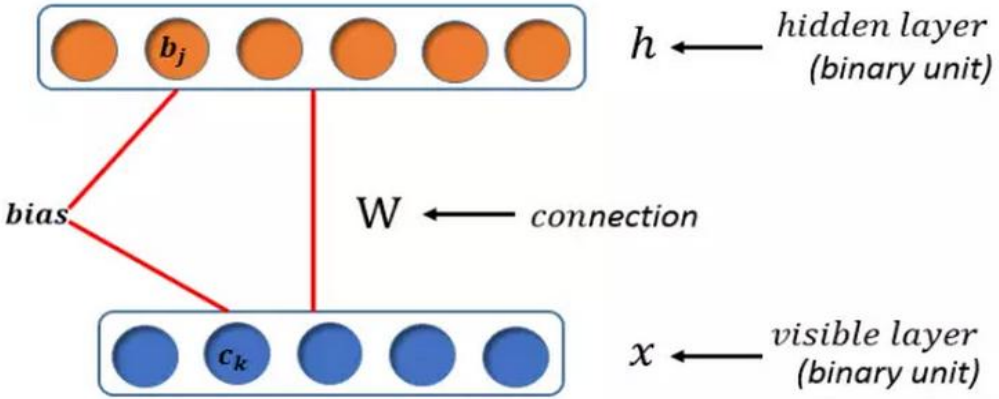
案例1：R中的深度学习与学习资源推荐

基于深度学习的推荐系统分类的二维体系，左侧部分对神经网络模型进行了说明，右侧部分则说明了整合模型。

darch包的应用：受限玻尔兹曼机（RBM）与学习资源推荐系统

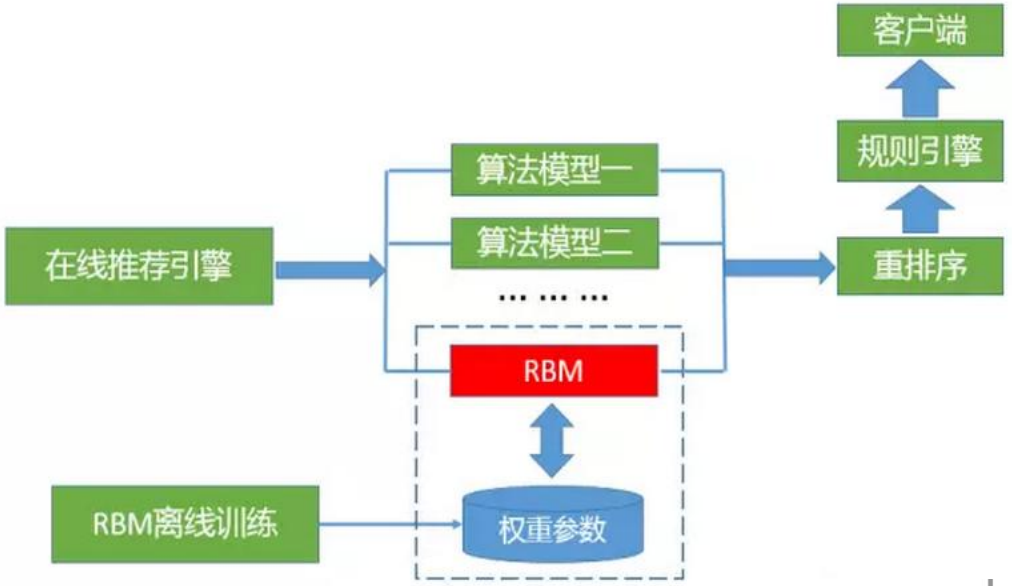
RBM网络结构定义

我们首先给出RBM的网络结构定义，RBM是由可视层和隐藏层相互连接的一个完全二分图网络结构，如下图所示，每一个神经元是一个二值单元，也就是每一个神经元的取值只能等于0或1：



darch包的应用： 受限玻尔兹曼机（RBM）与学习资源推荐系统

将RBM应用到学习资源推荐系统中，我们不但能得到已评分资源的新的评分数据，还能对未评分的资源的得分进行预测，这些未评分资源的分数从高到低的排序形成推荐列表。



案例2： R中的贝叶斯网络推理与学习路径推荐

学习路径推荐算法性能比较

比较维度 / 算法	参数设置	求解效率	空间开销	鲁棒性	运行机制	优势
遗传算法	适应性函数，交叉概率，变异概率	慢	中	强	内在并行	结构化处理组合优化问题
蚁群算法	启发信息，信息素	较快	中	强	分布式正反馈并行	动态组合优化问题
粒子群算法	速度 v，位置 p	快	大	强	信息共享机制	处理连续优化问题
人工神经网络	误差公式，权重	较快	中	强	动态反馈	分类、预测准确性好
AprioriAll 算法	规则，支持度，置信度	较快	大	中	项目序列集	适合稀疏数据集处理
Bayesian 网络推理	概率矩阵、权重公式	快	大	强	动态反馈	分类、聚类、因果关系分析，预测

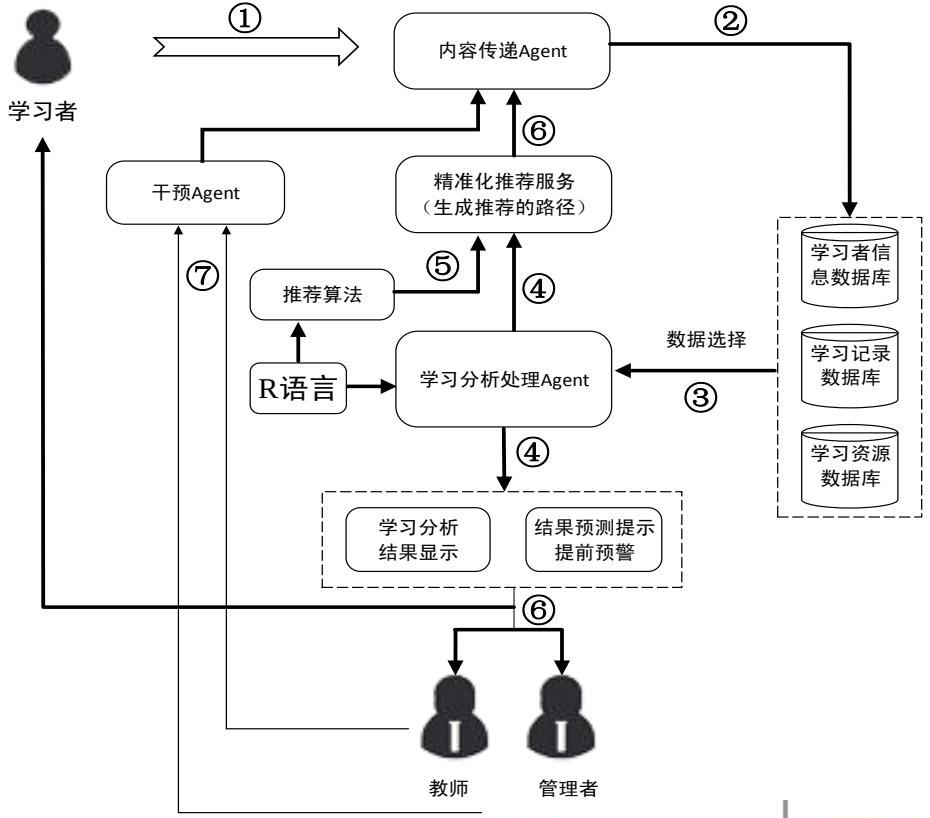
案例2： R中的深度学习与学习路径推荐

R包的类型	R包的名称	R包的基本介绍
贝叶斯方法 Bayesian Methods	BayesTree	贝叶斯可加回归树，最终的模型是由许多较弱学习器组成的，但跟随机森林部分提到的集成思想不一样
	tgp.MXM	贝叶斯不稳定性、半参数非线性回归、树状高斯过程(包括贝叶斯CART和线性模型)
	bnlearn	包含贝叶斯网络的结构学习、参数学习和推理三个方面的功能，其中结构学习包含基于约束的算法、基于得分的算法和混合算法，参数学习包括较大似然估计和贝叶斯估计两种方法。

案例2： R中的深度学习与学习路径推荐

基于R语言的学习路径推荐系统应用模式

R语言bnlearn包的应用：
利用贝叶斯网络推理实现学习路径的推荐



華東師範大學
EAST CHINA NORMAL UNIVERSITY

提纲

Contents Page

R语言构筑“人工智能+教育”应用服务

一、人工智能+教育的发展趋向

二、R的机器学习包及教育应用

三、推进R语言的教育应用实践

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY

IT大咖说
知识共享平台

研究生课程

课程名称：教育大数据与学习分析的研究
英文名称： Big Data in Education and Learning Analytics
任课教师　吴永和、 孙妍妍、 冯翔等
助教：钟薇 （邮件：xzw0706@163.com ）

课程类别	专业选修课	学时	36
学分	2	授课方式	教师讲座（讲课）+学生报告+学生研讨（课堂+微信+BB）+项目研究（实践）

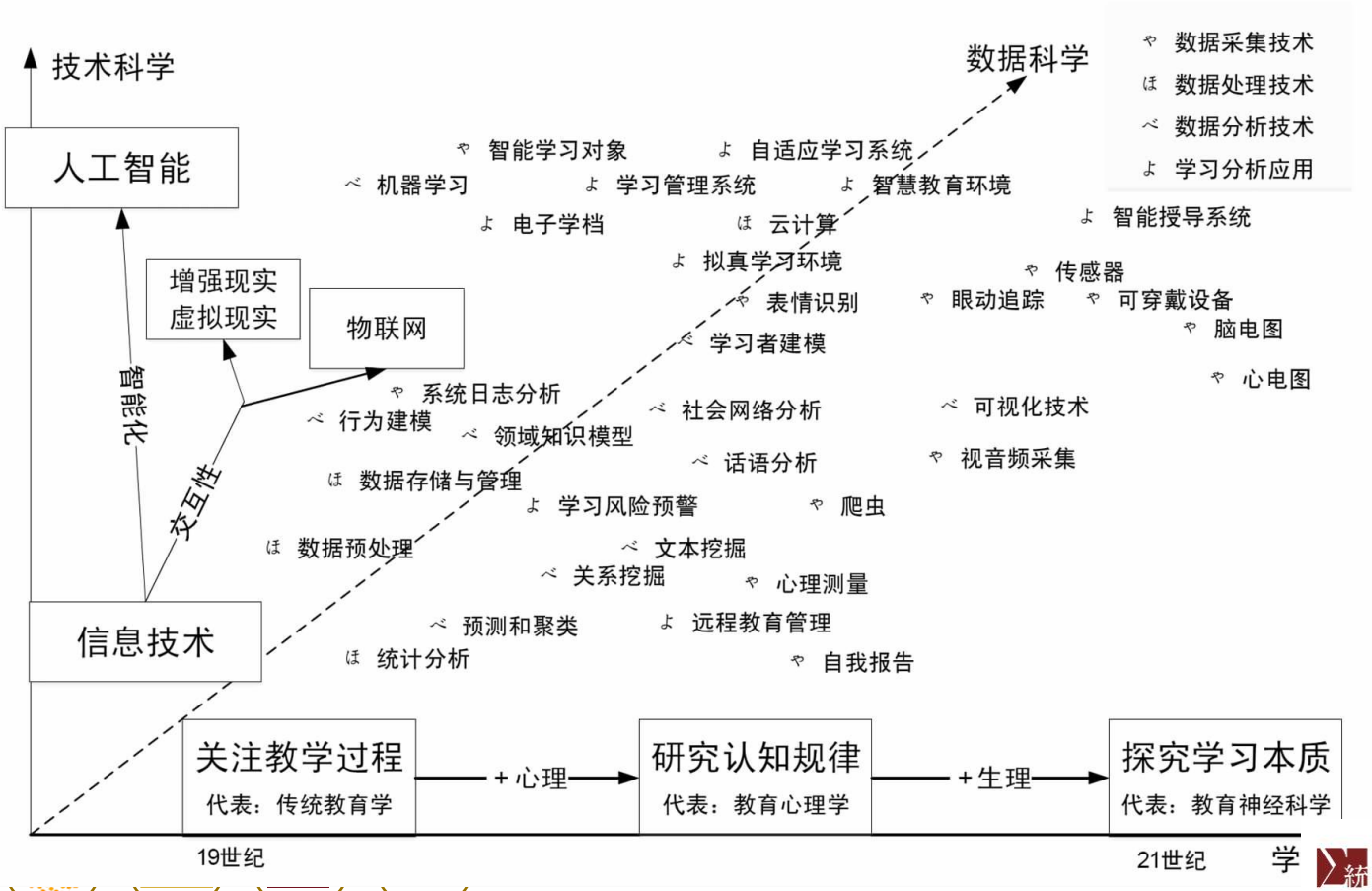
大数据与学习分析研讨班

华东师范大学教育学部教育信息技术学系

吴 永 和 博士 研究员 博导

2017.9—2018.1

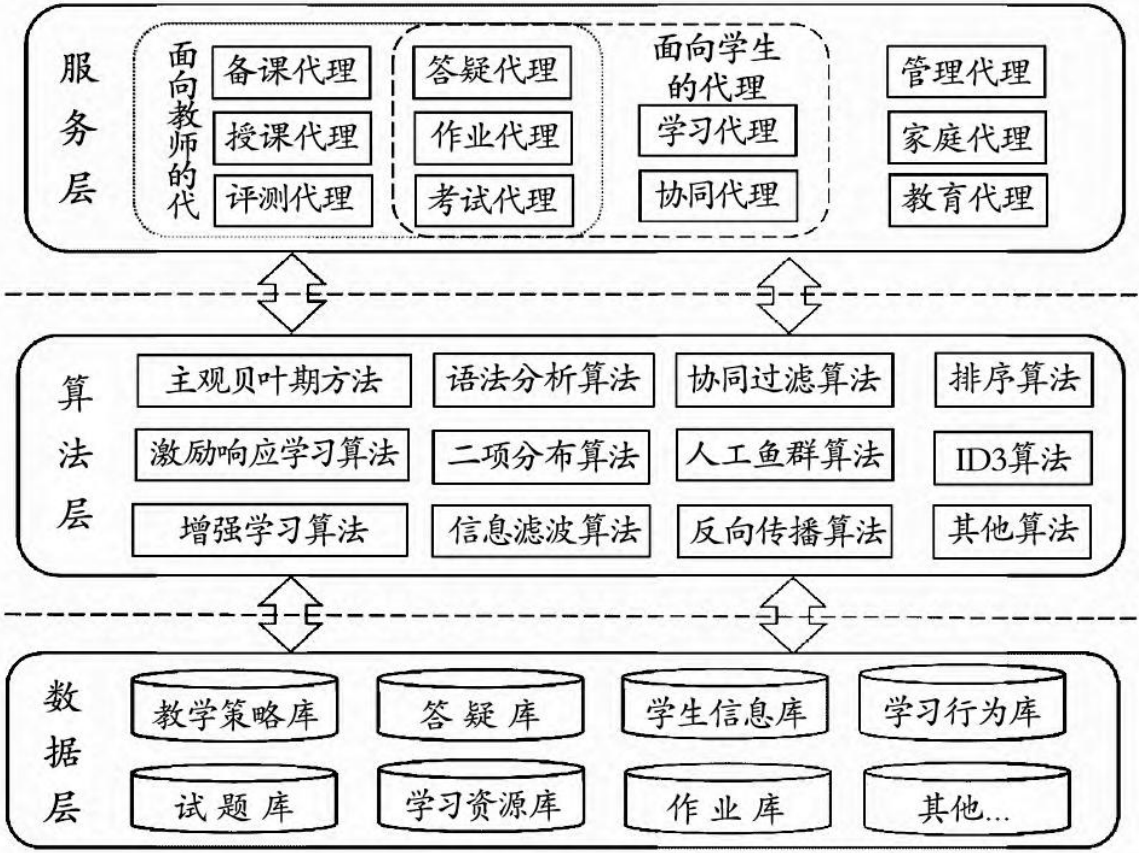
学习分析的跨学科技术支持



学习分析技术的发展和挑战

——第四届学习分析与知识国际会议评析

(教育代理)智能体层次架构



JOURNAL OF DISTANCE EDUCATION // Frontier Discovery

构筑"人工智能 + 教育"的生态系统*

吴永和[1] 刘博文[1] 马晓玲[2]

(1.华东师范大学 教育学部 教育信息技术学系，上海 200062；
2.华东师范大学 经管学部 信息管理系，上海 200241)

[摘 要] 近期国内召开的人工智能教育论坛，进一步促进了"人工智能+教育"研究的深入和应用推广，同时也为人工智能(AI)向"人工智能+教育"发展提供了技术基础。AI发展过程可分别从时间和技术两个维度进行划分，其在发展过程中不断与教育结合，产生了大量的研究成果，设为"人工智能+教育"提供了理论指导。通过查阅关键词云图和共词网络图分析国视了目前的研究现状，认为国内外各企业教育智能产品应用，是推进"人工智能+教育"的动力；国家颁布了一系列人工智能发展战略政策，特别强调教育应用，为"人工智能+教育"提供了政策保障，近而从孕育条件、特征、作用三方面阐述了"人工智能+教育"的内涵，从应用形态、技术架构、业态趋向等要素构建了"人工智能+教育"的生态系统，并阐述了"人工智能+教育"的人才培养体系。

[关键词] 人工智能；人工智能+教育；智能教育；教育人工智能(EAI)；教育智能代理；教育 Agent；生态系统；人才培养
[中图分类号] G40-057 [文献标识码] A [文章编号] 1672-0008(2017)05—0027—13

一、研究的提出

人工智能(Artificial Intelligence，简称 AI)延伸了人类的体力和脑力。"阿尔法狗"先后战胜二位世界围棋冠军李世石、柯洁，掀起了全球性人工智能发展的新一轮热潮。2016 年 5 月 23 日，国家发改委联合科技部、工信部等共同制定了《"互联网+"人工智能三年行动实施方案》[1]，首次明确提出到 2018 年基本建立人工智能产业、服务和标准化体系，实现核心技术突破，培育若干全球领先的人工智能骨干企业，形成千亿级的人工智能市场应用规模。2017 年 3 月 5 日，国务院总理李克强在《政府工作报告》中提到，要加快培育壮大新兴产业，全面实施战略性新兴产

7 月 8 日，国务院印发了《新一代人工智能发展规划》[2]，为加快人工智能的创新应用，提出了"智能教育"，为"人工智能+教育"指明了发展方向。

近年来，国内一些学者纷纷展开了相关研究。如，闫志明等提出教育人工智能(EAI)，界定了教育人工智能的内涵[3]，马玉慧等研究了智慧教育时代我国人工智能教育应用的发展路径[4]，刘清堂等对智能教学技术的发展与展望进行了研究[5]。可见，教育作为人工智能的主要应用领域之一，正向"人工智能+教育"的新阶段迈进。

那么，在今天这样一个人工智能快速发展的时代背景下，"人工智能+教育"的发展现况与重点如

http://obj.zjhu.edu.cn

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

R的教育应用现状

国外	Li, Yuelin (2006)	使用R 语言作为一种教学的工具和数据分析的软件来优化教学。
	Velasquez Henao, Juan David(2012)	在教室课堂上使用R 语言来进行模式分类，主要通过使用复杂的图形来解决数值问题和表征概念。
	Asaoka and RyoRussell(2013)	使用R 语言做地理地形的3D 可视化，以此来辅助地理教学。

R的教育应用现状

国内	关彦辉，闫朝辉，朱述知，李国军，张毅宁，欧祖军等	R软件在统计教学中的应用
	尹楠等	R语言在教育统计中的应用
	吴睿，张俊丽等（2016） 王星等（2017）	R语言在数据挖掘中的使用。利用网络爬虫等相关技术，我们可以捕捉到我们需要的教育数据，帮助研究工作顺利的展开。
	魏利明和陈文冬等（2015）	通过R语言统计建模，用来预测用户满意度。这对教育领域的学习者画像建模和预测学习者满意度有指导和借鉴作用。

R语言的教育应用系统构建与前景分析*

吴永和[1] 王浩楠

（1. 华东师范大学 教育学部教育信息技术学系 上海

[摘 要] 在当下数据密集型时代，教育数据的积累日益增多，但缺少
处理工具，导致了教育数据分析欠佳，其价值无法得到较好地挖掘，而
数据处理问题提供了较好的平台。R语言作为一款开源免费，具有统计
分析工具，国内外在不同领域对R语言的需求和应用都较为广泛，在教
可能为教育数据处理模式的转变创新贡献巨大的力量。本文通过梳理R
计类软件的优势对比，以及在教育中应用的现状等方面，进而分析了教
R语言作为教育领域数据处理工具的需求，示例了R语言在教育中的应用
教育应用发展的路径与架构，并指出了R语言在教育领域发展所存在的
更有效地服务于教育数据处理，提出了常态化应用的发展趋势及前景分
探索搭建了R语言教育应用社区和宣传平台，对解决教育领域数据处理
考。

[关键词] R语言；教育应用；大数据；开源免费；开源社区

一、引言

人类通过使用和创造工具，变得更聪明、更智慧。2016年诺贝尔化学
子机器的设计与合成”，获奖科学家解决了微观世界中的工具问题，而
具的作用同样较为重要。图灵奖得主、关系数据库的鼻祖Jim Gray提出
——数据密集型科学[1]；他认为今天以及未来科学的发展趋势是，随着
计算机将不仅仅能做模拟仿真，还能进行分析总结，得到理论。也就是
因斯坦等科学家从事的工作，未来可以由计算机来做。由此可见，工具
大。

2016年，国家发布的《国家十三五””规划纲要》中明确提出“实施国家大数据战略计划”[2]，将大数据列为国家今后的战略性发展方向。推动大数据的发展和利用，在各行业中发挥其作用，所创造出的价值将不可估量。随着国家“互联网+”计划的实施，以及教育信息化的不断普及与数字化学习的深入应用，教育领域中也逐渐累积了海量的数据，教育大数

基于 STEM 教育的大学生创新能力培养路径探究

——以 R 语言与 3D 打印在高数应用的实证研究为例

摘要：STEM 教育是培养学生创新能力的一种重要方式，结合 STEM 教育理念，将辅助工具应用于高等数学学习，有助于大学生创新能力的培养和学习兴趣的提升。通过分析 STEM 教育应用现状和目前大学生创新能力的培养路径情况，设计和开发了基于 R 语言与 3D 打印结合的高数函数学习活动，并以此为核心，选取合适的实验对象进行了实证研究。设计和实施了 STEM 教育实践活动探究实验，考察并应证了该工具及活动对大学生创新能力培养和学习兴趣方面的促进反馈作用。同时分析了影响大学生创新能力培养的诸多因素和学习者对 STEM 教育活动的学习需求，以及实施活动中存在的问题，对探究大学生创新能力的培养路径提供了一定的参考。

关键词：STEM 教育；大学生创新能力培养；实证研究；R 语言与 3D 打印；高等数学三维函数

一、引言

“为什么我们的学校总是培养不出杰出人才？”著名的“钱学森之问”对学校培养创造性人才的模式提出了严肃拷问。创新是一种发散性的高阶思维能力，是国家和社会发展的核心动力（林金辉,1995）。创新也是 21 世纪学习的一种重要能力（Trilling et al.，2012）。近几年，创新一直获得相当高的关注度。在心理、经济、数学及其他学科，创新已经变成了最重要的研究方向（Abdullaetal, 2017）。高等教育已被广泛认为是推动创新的最重要途径，从高校中走出的大学生理应是创新创业中极具活力的新生代力量。1999 年，联合国教科文组织发布《21 世纪的高等教育宣言》，将“创业技能和创业精神”作为高等教育的基本目

R 语言在课堂教学数据处理中应用的研究*

吴永和[1] 李若晨[1]

（1. 华东师范大学 教育学部教育信息技术学系 上海 200062）

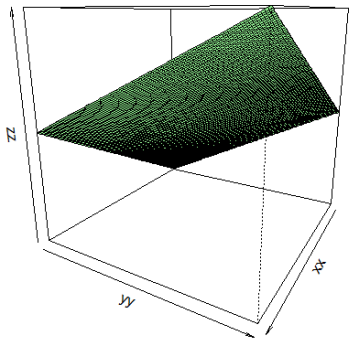
教育的核心环节，也是教育量化研究难点。R 语言这一统计分析和可视化工具正飞速
具有成为课堂教学数据处理普适性工具的潜力。本文探究电子书包互动课堂教学中产
结构及处理模型，从技术的视角探究 R 语言在教学中的应用体系，进而提出基于 R 语
处理系统框架，结合教学需求、R 技术方法和应用案例，从利益相关者角度阐述 R 语言
中的应用前景，分析并提出了其应用研究的方向。
堂教学；教育大数据；学习分析；教学分析系统

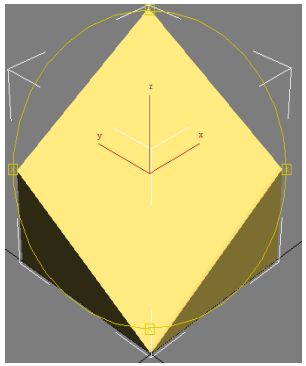
[文献标识码]A

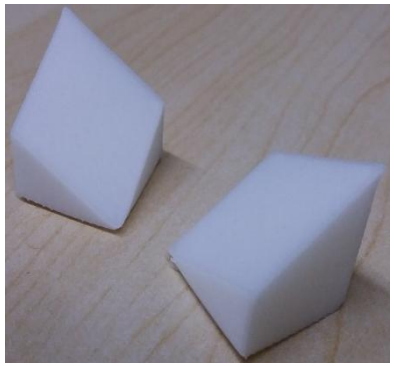
围绕一定的教学任务，组织教材和教学方式，以班级教学为主要形式，指导
支能的过程，是教师“教”和学生“学”的相结合或者相统一的活动[5]。它
遍使用的一种手段，也是教育的核心环节。2015 年和 2016 年地平线报告将
视为高等教育的主要发展趋势之一。混合式学习将传统学习与网络化学习
各优势互补，以获得最佳的学习效果[2]。传统课堂教学与数字化学习资源（电
线学习平台的有机结合，构筑了“数字化教育生态新环境”[6]。数字化教育
的教育大数据为课堂教学中的“数据密集型科学[4]”研究奠定基础。
据处理是学习分析在课堂教学领域中的应用分支。学习分析是测量、搜集、
者及其所在学习情境中产生的数据，以此预测学习表现、改善教学实践、优
术[8]。它通过对不同情境下学习者学习活动创造的大量数据的收集和挖掘来
共预警，优化教学策略和教学设计，构建学习资源服务生态环境等。作为学
中的重要组成部分，课堂教学数据处理主要侧重于使用专业工具，通过一定
去，对课堂教学数据进行采集、处理并赋予一定意义，使之成为能真实反应
水平或课堂教学成果等的有用信息。其中，工具的选择和使用对数据处理结
有较大影响。
牛 Excel、SPSS、SQL 等虽然能完成当前大部分数据处理需求，但作为商业
普遍较低，部分功能与教育数据处理需求契合度不足。选择切合教学实践的
应课堂教学数据处理的迫切需求，R 语言作为一款开源的统计分析和可视化

案例1： R语言与3D打印相结合--r2stl

```
###截取一个平面（初中典型）
x <- seq(-10, 10, length= 100)
y <- x
f <- function(x,y) { r <- x+y;r}
z <- outer(x, y, f)      # z <- outer(x, y, “+”)
r2stl(x, y, z, filename=“jiemian.stl", show.persp=TRUE)
```







基于 STEM 教育的大学生创新能力培养路径探究

——以 R 语言与 3D 打印在高数应用的实证研究为例

摘要：STEM 教育是培养学生创新能力的一种重要方式，结合 STEM 教育理念，将辅助工具应用于高等数学学习，有助于大学生创新能力的培养和学习兴趣的提升。通过分析 STEM 教育应用现状和目前大学生创新能力的培养路径情况，设计和开发了基于 R 语言与 3D 打印结合的高数函数学习活动，并以此为核心，选取合适的实验对象进行了实证研究。设计和实施了 STEM 教育实践活动探究实验，考察并应证了该工具及活动对大学生创新能力培养和学习兴趣方面的促进反馈作用。同时分析了影响大学生创新能力培养的诸多因素和学习者对 STEM 教育活动的学习需求，以及实施活动中存在的问题，对探究大学生创新能力的培养路径提供了一定的参考。

关键词：STEM 教育；大学生创新能力培养；实证研究；R 语言与 3D 打印；高等数学三维函数

一、引言

“为什么我们的学校总是培养不出杰出人才？”著名的“钱学森之问”对学校培养创造性人才的模式提出了严肃拷问。创新是一种发散性的高阶思维能力，是国家和社会发展的核心动力（林金辉,1995）。创新也是 21 世纪学习的一种重要能力（Trilling et al.，2012）。近几年，创新一直获得相当高的关注度。在心理、经济、数学及其他学科，创新已经变成了最重要的研究方向（Abdullaetal. 2017）。高等教育已被广泛认为是推动创新的最重要途径，从高校中走出的大学生理应是创新创业中极具活力的新生代力量。1999 年，联合国教科文组织发布《21 世纪的高等教育宣言》，将“创业技能和创业精神”作为高等教育的基本目

案例2：国培数据分析

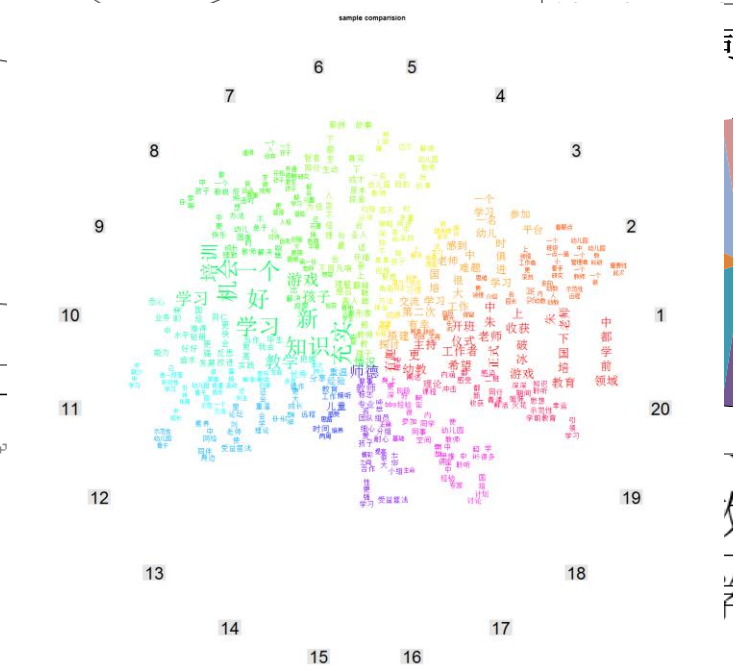
词类分布

学生的情绪和态度 2%

教师的情绪和态度 1%

教学者行为与品质 18%

现有教育信息化状况 5%



描述

评判

情感

对同伴、教学者所表

15% 现有教育信息化状况 5%

教学平台

CAPITAL OF STATISTICS PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

案例3——华东师范大学大夏学堂平台

基于 LMS 的文理科使用情况差异的比较研究——以上海某高校平台为例

小组成员:肖敏 危怡 师睿 付楚昕 李雪娇 吴秉聪

1.文献综述

1.1 学习管理系统概述

教育信息化的迅速发展对技术支持的学习环境提出更高要求，学习管理系统（LMS, learning management system）作为远程开放教育的主要媒介（张汉玉等，2013），具有组织、跟踪、评估、发送、呈现、管理学习内容和学习活动，促进学习者交互等一系列功能（骆莎&黎加厚，2007），成为远程开放教育中成为主流的支持工具。学习管理系统系统的使用教师为课堂建构一个将课内外教学连接起来的学习环境。

IEEE 1484 标准提出了关于学习管理系统的整体架构 LTSA(Learning Technology System Architecture)，作为信息化教学系统的抽象模型，该系统由四个模建（学习者、评估、教学代理和投递）和 2 个数据库（学习资源及学习者记录）组成（李华等，2013），如图 1 所示。此外，学习管理系统的发展收到了学习设计理念（learning design）的影响，强调“活动”和“序列”，“活动”则处于“内容”的上一个层次（曹晓明&何克抗，2006）。基于此将本研究涉及的 15 项主要功能分消息通讯、内容创作、学习资源、学习支持、学习评价五大模块。

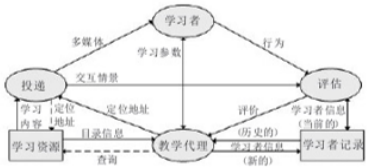


图 1 学习管理系统的整体架构 LTSA

1.2 数据挖掘在学习管理系统中的应用

聚类分析是在没有给定划分类别的情况下，根据数据相似度进行样本分组的一种方法。与分类模型需要使用类标记样本构成的训练数据不同，聚类模型可以建立在无类记得数据上，是一种非监督的学习算法。聚类的输入是一组未标记的样本，聚类根据数据自身的距离或者相似度将其分为若干组，划分的原则是组内距离最小化组件距离最大化。Talavera和Gaudioso（2004）提出使用聚类挖掘学生数据来发现反映用户行为的模式。 他们提出了协作管理模型，以表征非结构化协作空间中的类似行为组。 Mor（2004）通过将教师专业知识与通过系统使用分析获得的学习经验相结合，扩展了SCORM标准的测序能力，以纳入推荐行程的概念。 他们使用聚类算法对学生进行分组。 Castro，Vellido，Nebot和Minguillon（2005）发现虚拟校园用户的分组结构的非典型行为。 他们建议使用生成地形图模型和聚类模型来表征在线学生群体。

关联规则分析也成为购物篮分析，最早是为了发现超市销售数据库中不同的商品之间的关联关系，关联规则分析是数据挖掘中最活跃的研究方法之一，目的是在一个数据集中找出各项之间的关联关系，而这种关系并没有在数据中直接表示出来。目前最常用的经典算法是Apriori算法，对离散化的课程指要表执行支持度和可信度为特定参数值的Apriori算法，会出现一系列因果规则的支持度和可信度。

关联规则挖掘应用于基于网络的教育系统可以：（1）创建可以推荐在线学习活动或学习捷径的委托代理；（2）诊断学生学习问题并向学生提出建议；（3）自动引导学习者的学习行为并推荐学习资料；（4）确定哪些学习材料最适合推荐给用户；（5）确定有品性差异的学生群组间的特征模式；（6）从学生的使用信息中发掘值得关注的关联关系以便向课程创作者提供反馈；（7）从学习者行为模式中找出联系；（8）找寻学生经常犯的错误；（9）引导搜寻最合适学生学习的转换模式；（10）通过确定什么最吸引用户来优化E-learning入门内容（吴青等，2015）。

1.3 文理科课程中学生学习行为差异

对于学科的分类,教育学者的看法各不相同。法国哲学家孔德在19世纪上半叶提出了被称为“实证等级论”的科学分类体系,直至今天依然被视为中小学和大学课程分科的基本依据。杜威（1963）认为，按照惯例，人们把学科划分为三类：获得实践技能的特殊学科，如阅读、书写、计算和音乐；主要为获得知识的“知识性”学科，如地理、历史，工作的技能和知识扩增不占重要地位，而更注重抽象思维的“推理的”、“训练性的”学科，如数学、形式语法等。因此文理科课程具有本质意义上的差别，从而也会影响学生的学习行为，造成学生在线上的行为差异。

此外，文理科课程中学生的学习习惯、学习行为等都会显现出不同，龙迟和杨瑜（2015）对文理科学生的信息选择行为进行了差异分析，发现文理科课程中对学生信息检索能力的要求存在差异。胡选萍（2011）则对学生文理科思维影响因素进行了探析，除此之外大部分文理科对比的研究都以学生作为分类对象，而很少以课程作为对象，本研究则致力于探寻学习管理平台中不同学科的学习行为差异。

2.文理科点击量描述性统计

2.1 文科课程程序模块点击概况

文科共 5 门课程，包括：学术英语阅读、数学教学设计、马克思主义原理概论、汉语言文学、语言学概论，学生总数为 683 人。对点击量不为 0 的所有功能模块的点击量利用 SPSS 软件进行描述统计，得到如下表：

表 1 文科程序模块点击描述性统计

	N	极小值	极大值	均值	标准差
公告	683	0	137	7.31	14.999
词汇表	683	0	3	.97	.323
博客	683	0	128	1.35	7.848
内容项	683	0	633	67.13	74.213
电子邮件	683	0	3	.92	.198
工具区	683	0	14	.40	1.441
讨论板	683	0	1	.96	.974
我的课堂日历	683	0	1	.96	.930
作业	683	0	1435	60.74	166.180
小组	683	0	86	2.13	5.985
成绩中心	683	0	996	4.92	56.742
日历	683	0	4	.92	.197
联系人	683	0	11	.92	.935
我的成绩	683	0	187	9.35	15.527
任务	683	0	1	.92	.390
grading_applicatio n_name	683	0	92	.98	3.998
BBCourse_service_ Level_label测验	683	0	1	.96	.930
retention_applicat ion_name	683	0	7	.92	.360
测	683	0	1773	176.94	323.130
有效的 N（列表状态）	683				

从上表的均值来看，在文科中，LMS 平台点击最多的功能模块是讨论板，为 84 次，其次是内容项，为 68 次。去掉总点击量低于 10 以及一些静态的程序模块，作饼图如下：

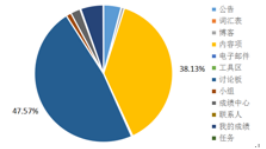


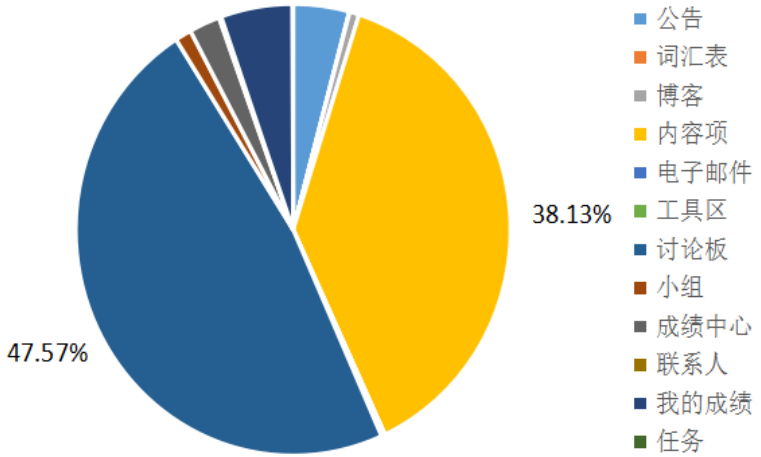
图 1 文科总程序模块点击分布饼图

今日优选
快剪辑
搜索 Web 和 Windows
CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

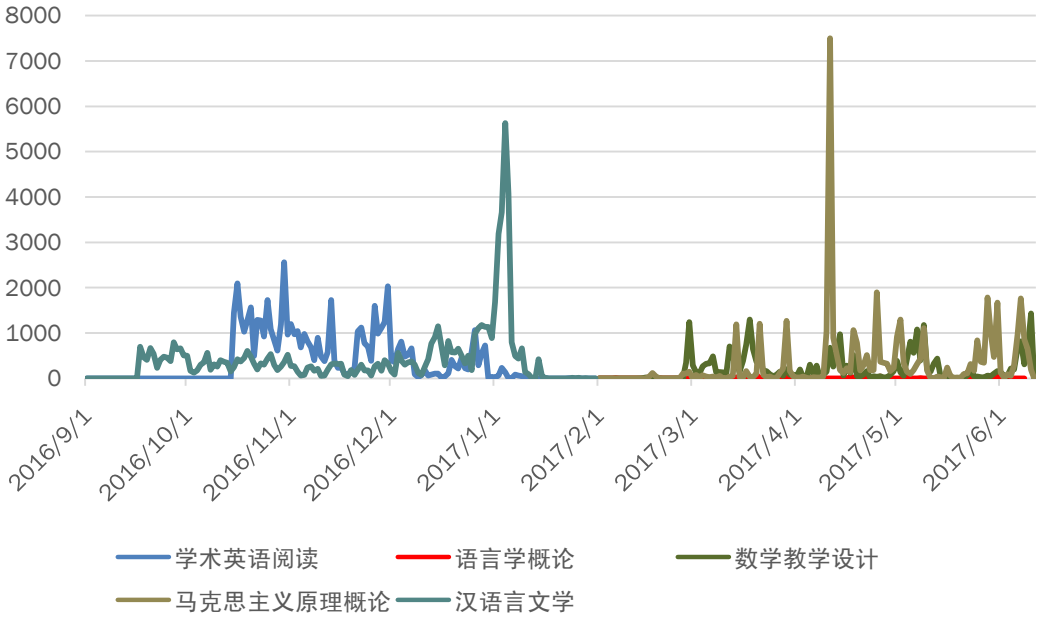
案例3 ——华东师范大学大夏学堂平台

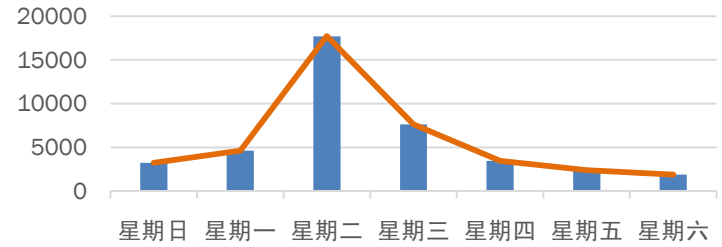
描述统计量

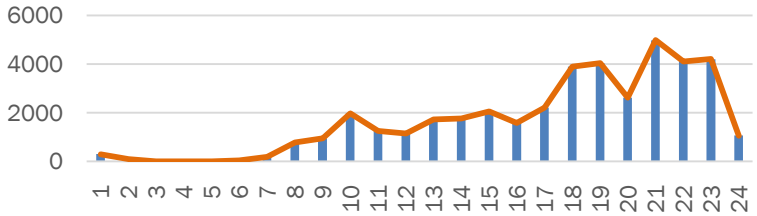
	N	极小值	极大值	均值	标准差
公告	683	0	137	7.31	14.989
词汇表	683	0	3	.07	.323
博客	683	0	128	1.25	7.810
内容项	683	0	633	67.13	74.213
电子邮件	683	0	3	.02	.198
工具区	683	0	14	.46	1.411
花名册	683	0	1	.00	.054
档案夹主页	683	0	1	.00	.038
讨论板	683	0	1410	83.74	186.103
小组	683	0	86	2.13	5.985
成绩中心	683	0	990	4.03	56.742
日志	683	0	4	.01	.167
联系人	683	0	11	.32	.935
我的成绩	683	0	187	9.35	15.527
任务	683	0	6	.09	.396
grading.application.name	683	0	52	.08	1.990
@X@course.service_level_label@X@ 消息	683	0	1	.00	.038
retention.application.name	683	0	7	.02	.363
总	683	0	1773	176.04	223.136
有效的 N （列表状态）	683				



案例3 ——华东师范大学大夏学堂平台



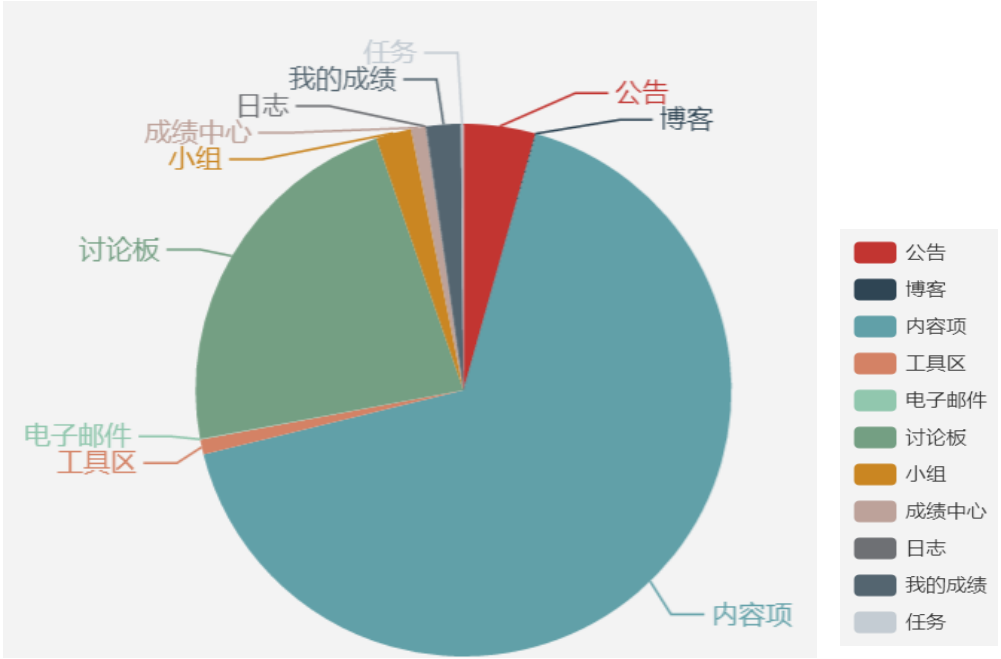




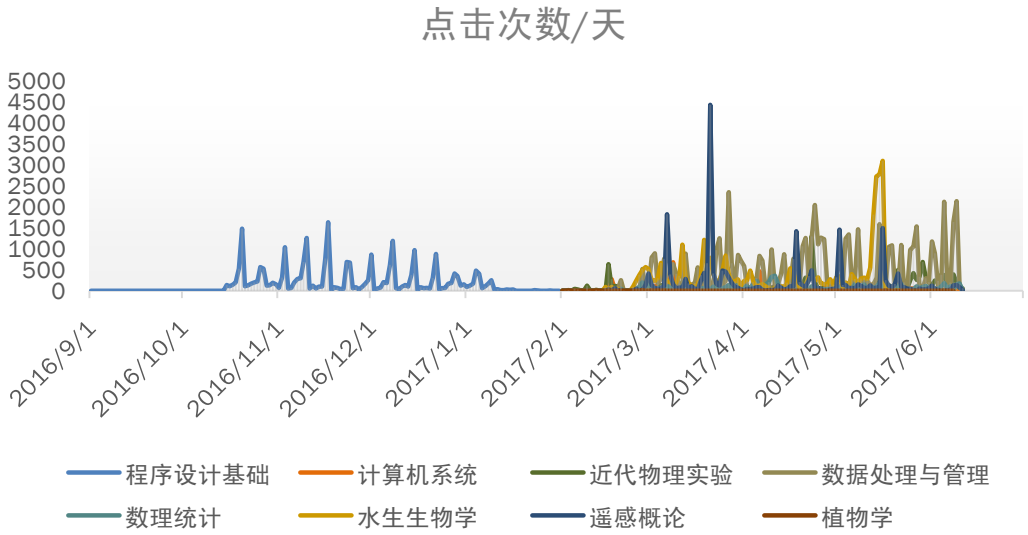
CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

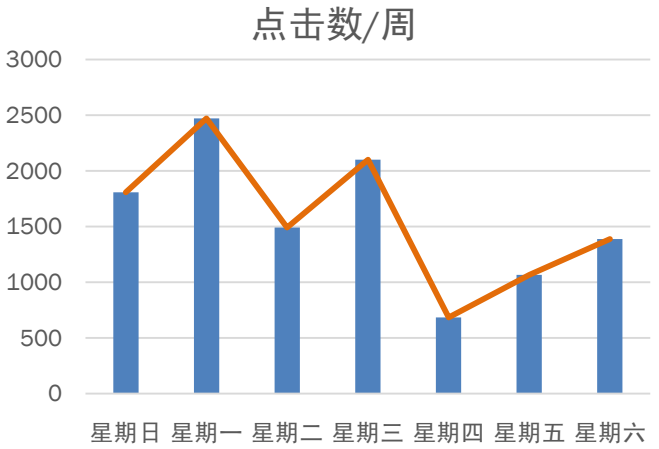
案例3 ——华东师范大学大夏学堂平台

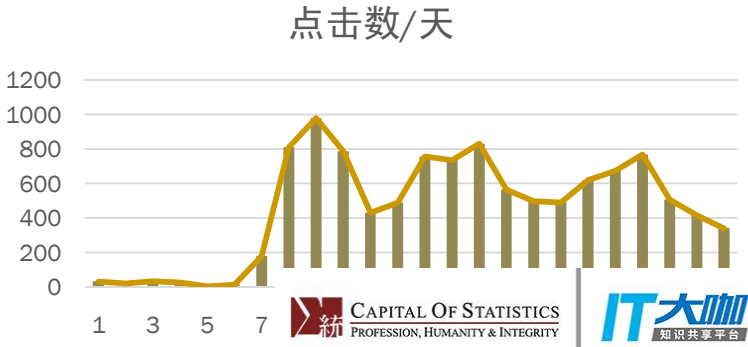
	N	极小值	极大值	均值	标准差
公告	898	0	170	7.10	13.527
博客	898	0	5	.05	.337
内容项	898	0	833	109.59	106.920
花名册	898	0	5	.03	.248
工具区	898	0	42	1.49	3.845
档案夹主页	898	0	1	.01	.088
电子邮件	898	0	9	.07	.532
讨论板	898	0	1298	37.19	107.818
小组	898	0	157	3.49	10.895
成绩中心	898	0	1002	1.42	34.051
日志	898	0	6	.10	.488
我的成绩	898	0	76	3.41	7.090
任务	898	0	9	.24	.880
有效的 N （列表状态）	898				



案例3 ——华东师范大学大夏学堂平台



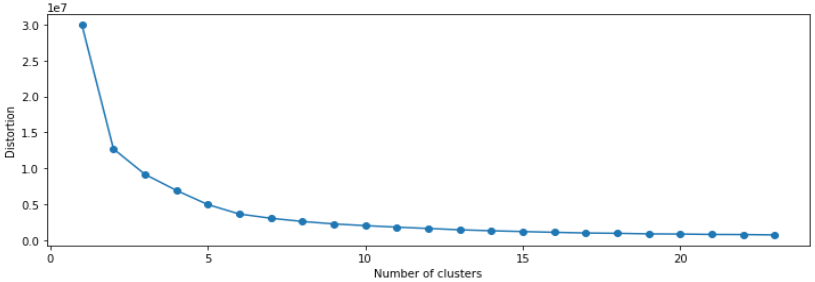


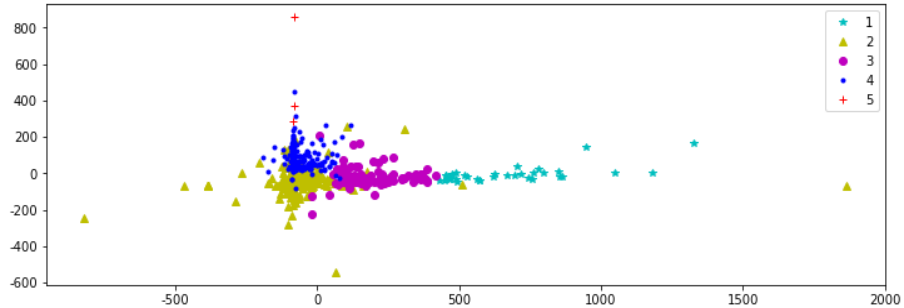


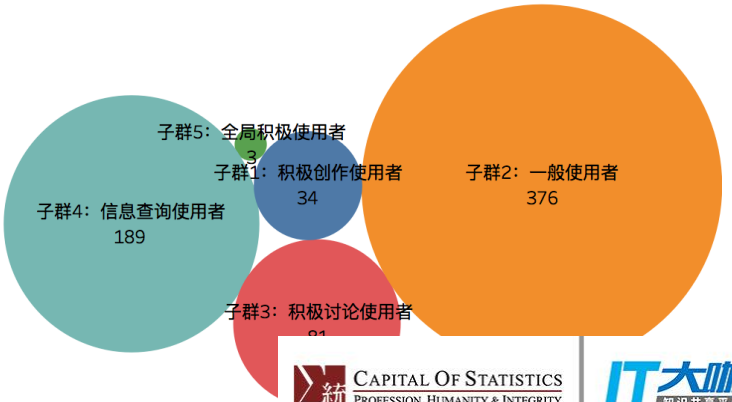
CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

案例3 ——华东师范大学大夏学堂平台

文科聚类分析



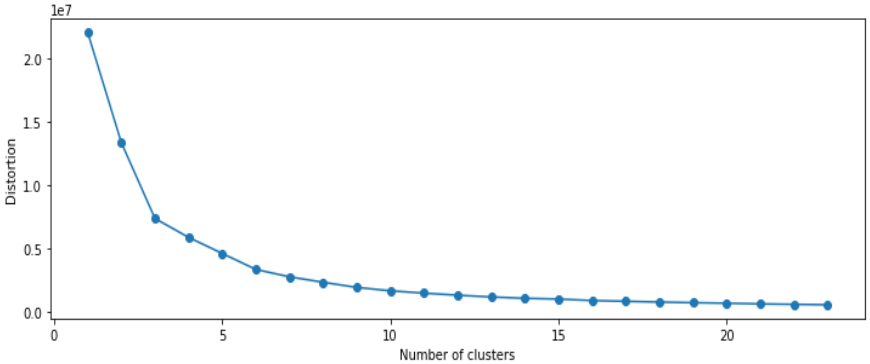


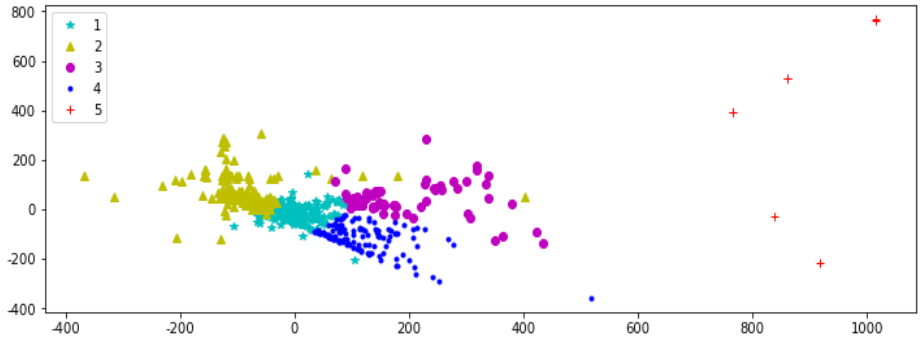


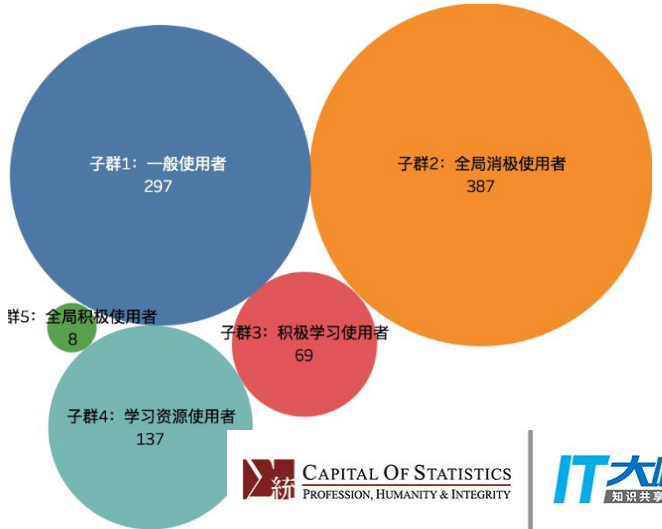
CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

案例3 ——华东师范大学大夏学堂平台

理科聚类分析



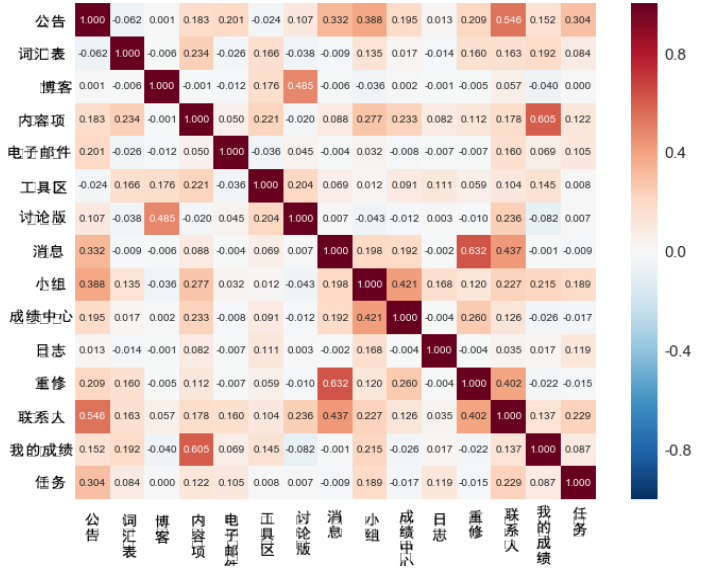




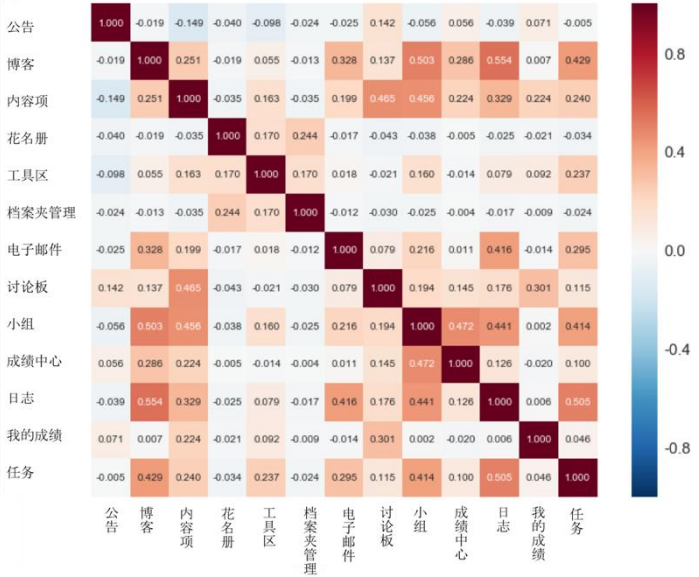
CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

案例3 ——华东师范大学大夏学堂平台

文理科关联规则分析



文科聚类分析



理科聚类分析

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

案例3 ——华东师范大学大夏学堂平台

文理科模块访问差异比较

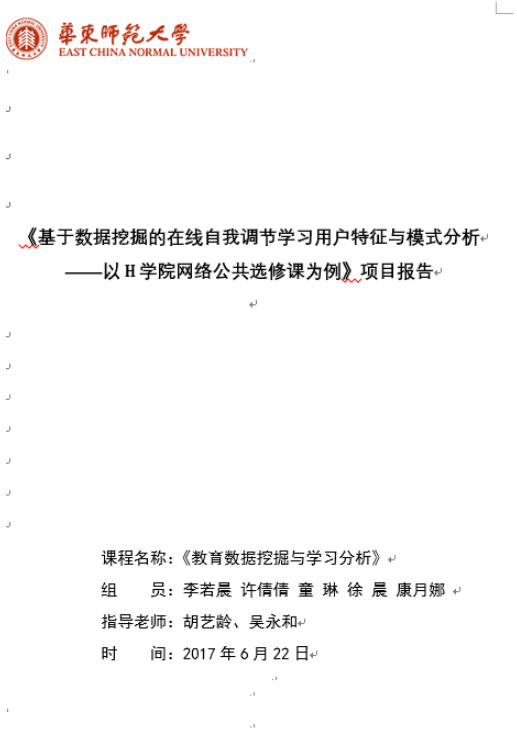
文科学生更加关注学习评价，学习评价与学习资源相关，积极的使用者会在课程学习的过程中通过小组讨论的方式完成创作。

理科学生主要集中在内容创作和学习支持上，注重在线学习中的自我创造能力和交互能力，积极的使用者关注教师的状态，以教师的任务为导向。而学习资源使用者则时刻关注自身学习的发展动向，并常使用日志和博客的形式关注教学或调整自己的学习步调。

可见，理科学生较文科学生能更全面的使用大夏平台，而文科学生则偏向通过对学习资源的使用来提高成绩。

首都 CAPITAL OF STATISTICS PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

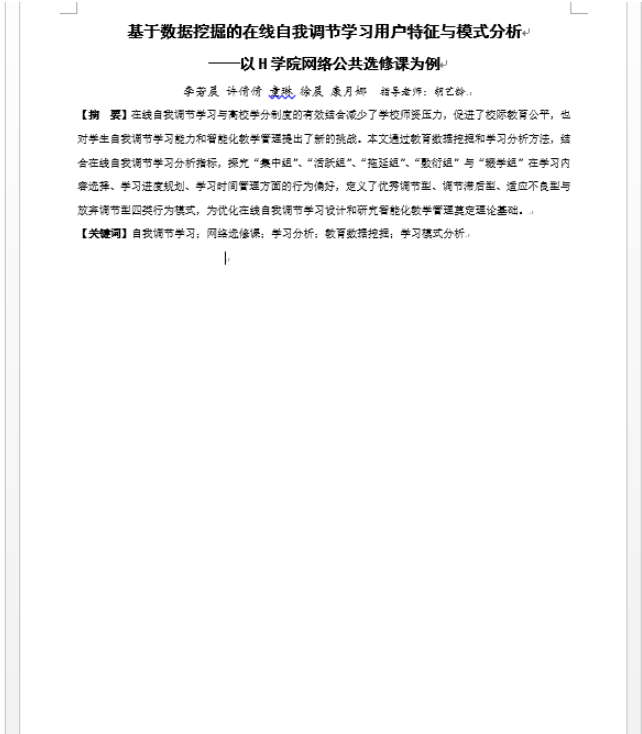
案例4 ——基于数据挖掘的在线自我调节学习用户特征与模式分析——以H学院网络公共选修课为例



華東師範大學
EAST CHINA NORMAL UNIVERSITY

《基于数据挖掘的在线自我调节学习用户特征与模式分析
——以 H 学院网络公共选修课为例》项目报告

课程名称：《教育数据挖掘与学习分析》
组　　员：李若晨 许倩倩 童 琳 徐 晨 康月娜
指导老师：胡艺龄、吴永和
时　　间：2017 年 6 月 22 日

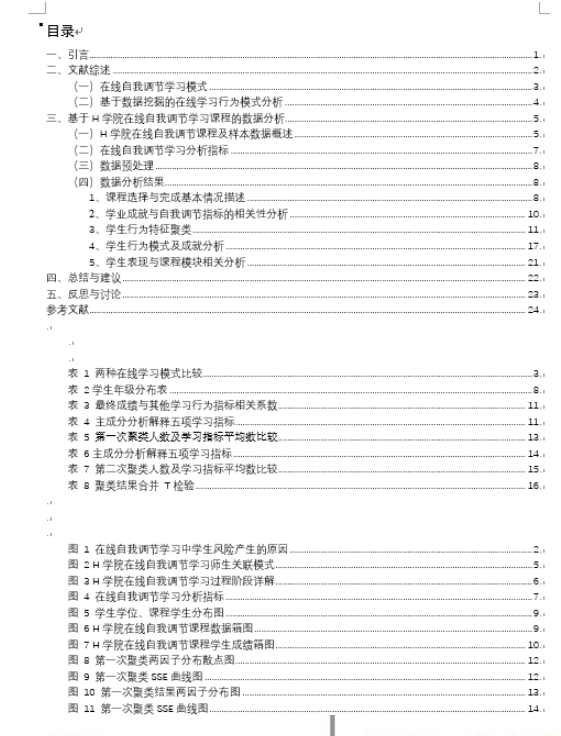


基于数据挖掘的在线自我调节学习用户特征与模式分析
——以 H 学院网络公共选修课为例

李若晨 许倩倩 童琳 徐晨 康月娜　指导老师：胡艺龄，

【摘　要】在线自我调节学习与高校学分制度的有效结合减少了学校师资压力，促进了校际教育公平，也对学生自我调节学习能力和智能化教学管理提出了新的挑战。本文通过教育数据挖掘和学习分析方法，结合在线自我调节学习分析指标，探究“集中组”、“活跃组”、“拖延组”、“勤奋组”与“顺学组”在学习内容选择、学习进度规划、学习时间管理方面的行为偏好，定义了优秀调节型、调节滞后型、适应不良型与放弃调节型四类行为模式，为优化在线自我调节学习设计和研究智能化教学管理奠定理论基础。

【关键词】自我调节学习；网络选修课；学习分析；教育数据挖掘；学习模式分析



目录

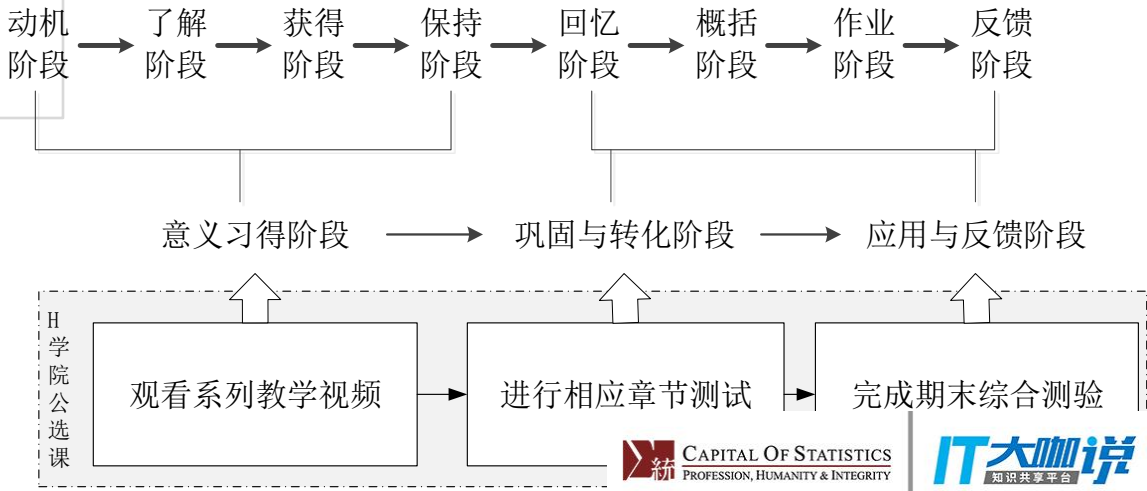
一、引言 1
二、文献综述 2
（一）在线自我调节学习模式 3
（二）基于数据挖掘的在线学习行为模式分析 4
三、基于 H 学院在线自我调节学习课程的数据分析 5
（一）H 学院在线自我调节课程及样本数据概述 5
（二）在线自我调节学习分析指标 7
（三）数据预处理 8
（四）数据分析结果 8
1、课程选择与完成基本情况描述 8
2、学业成就与自我调节指标的相关性分析 10
3、学生行为特征聚类 11
4、学生行为模式及成就分析 17
5、学生表现与课程模块相关分析 21
四、总结与建议 22
五、反思与讨论 23
参考文献 24

表 1 两种在线学习模式比较 3
表 2 学生年级分布表 8
表 3 最终成绩与其他学习行为指标相关系数 11
表 4 主成分分析解释五项学习指标 11
表 5 第一次聚类人数及学习指标平均数比较 13
表 6 主成分分析解释五项学习指标 14
表 7 第二次聚类人数及学习指标平均数比较 15
表 8 聚类结果合并 T 检验 16

图 1 在线自我调节学习中学生风险产生的原因 2
图 2 H 学院在线自我调节学习师生关联模式 5
图 3 H 学院在线自我调节学习过程阶段详解 6
图 4 在线自我调节学习分析指标 7
图 5 学生学位、课程学生分布图 9
图 6 H 学院在线自我调节课程数据箱图 9
图 7 H 学院在线自我调节课程学生成绩箱图 10
图 8 第一次聚类两因子分布散点图 12
图 9 第一次聚类 SSE 曲线图 12
图 10 第一次聚类结果两因子分布图 13
图 11 第一次聚类 SSE 曲线图 14

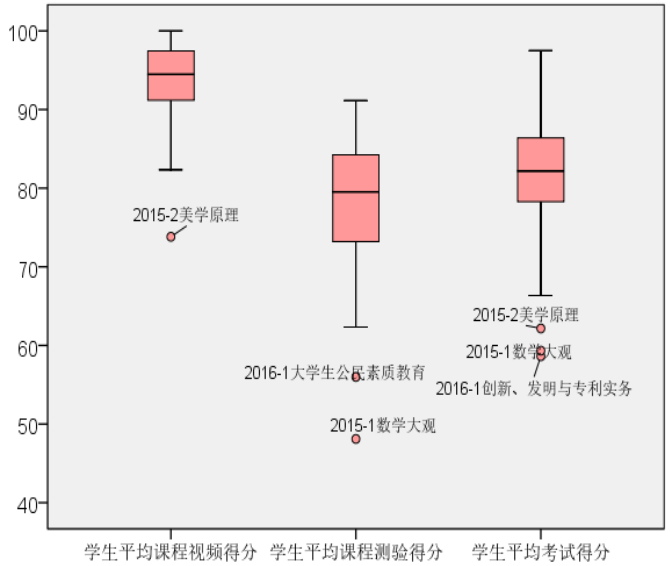
CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

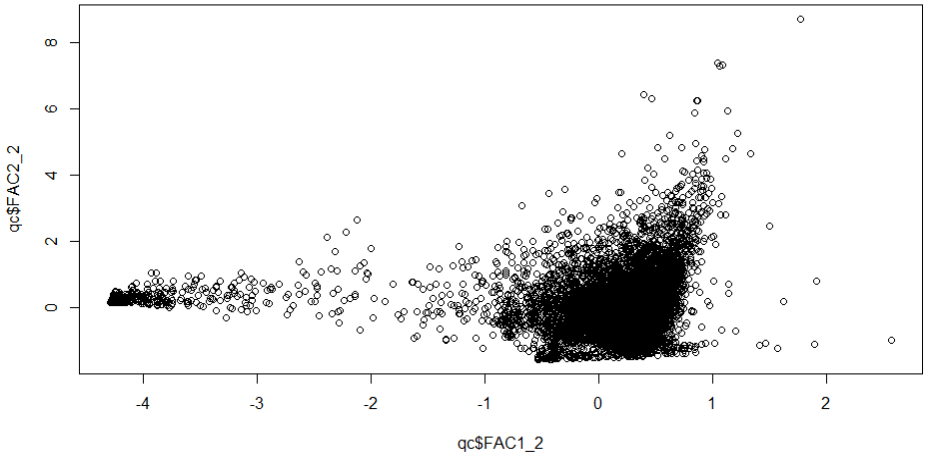
案例4 ——基于数据挖掘的在线自我调节学习用户特征与模式分析——以H学院网络公共选修课为例

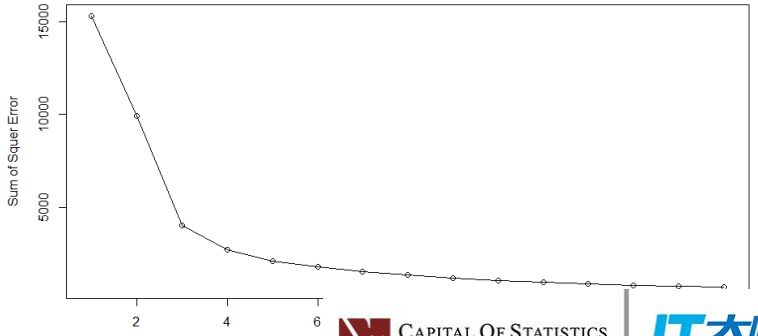


CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

案例4 ——基于数据挖掘的在线自我调节学习用户特征与模式分析——以H学院网络公共选修课为例

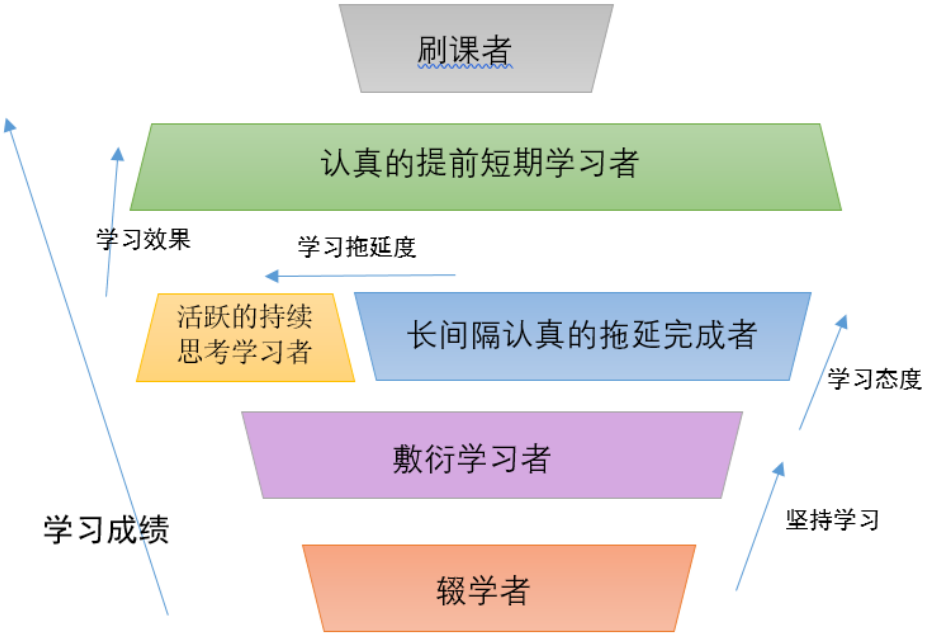




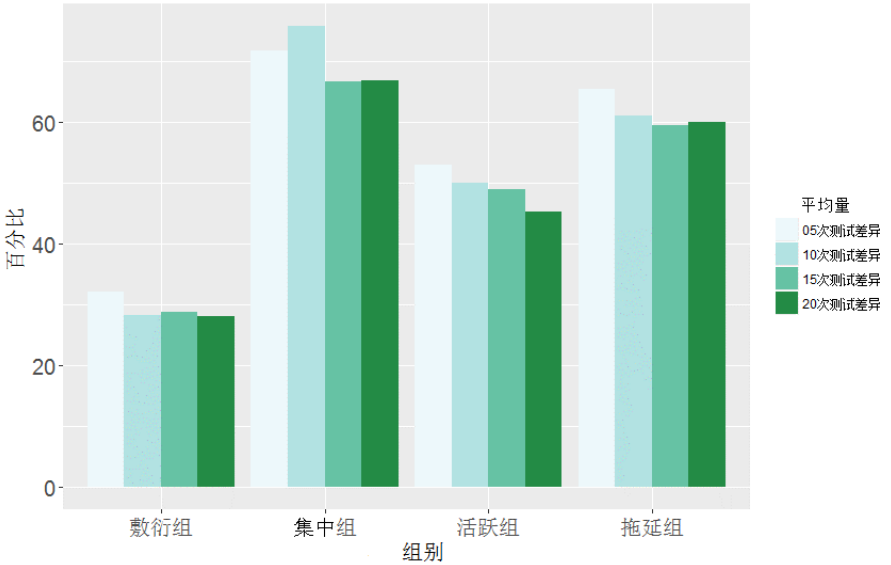


CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

案例4 ——基于数据挖掘的在线自我调节学习用户特征与模式分析——以H学院网络公共选修课为例



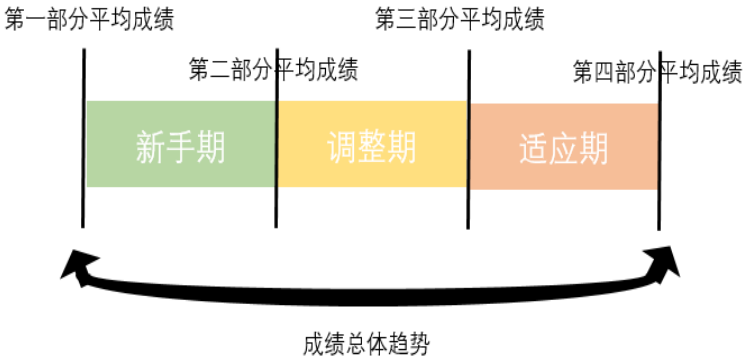
在线自我调节学习用户特征模型图



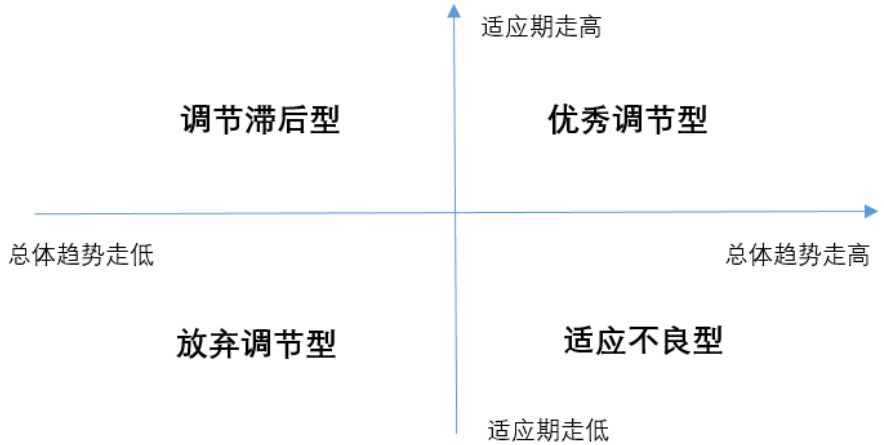
在线自我调节

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

案例4 ——基于数据挖掘的在线自我调节学习用户特征与模式分析——以H学院网络公共选修课为例



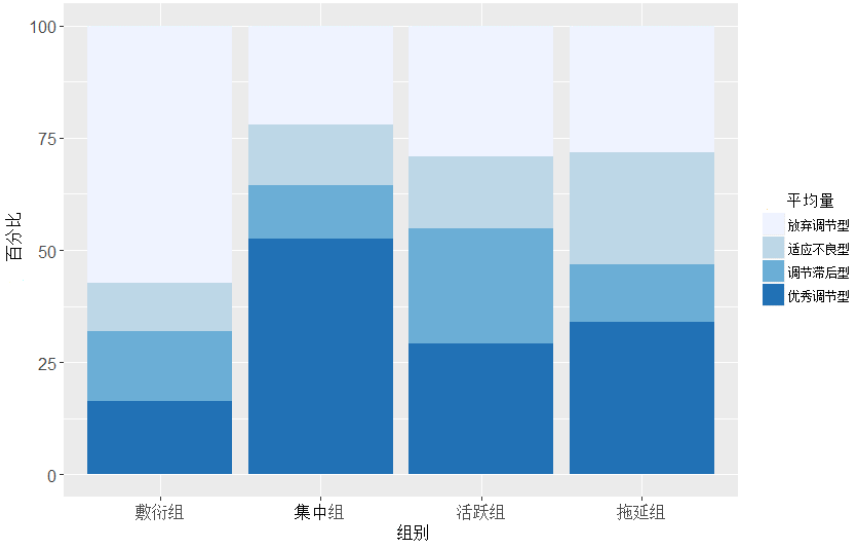
在线自我调节学习行为模式三阶段示意图



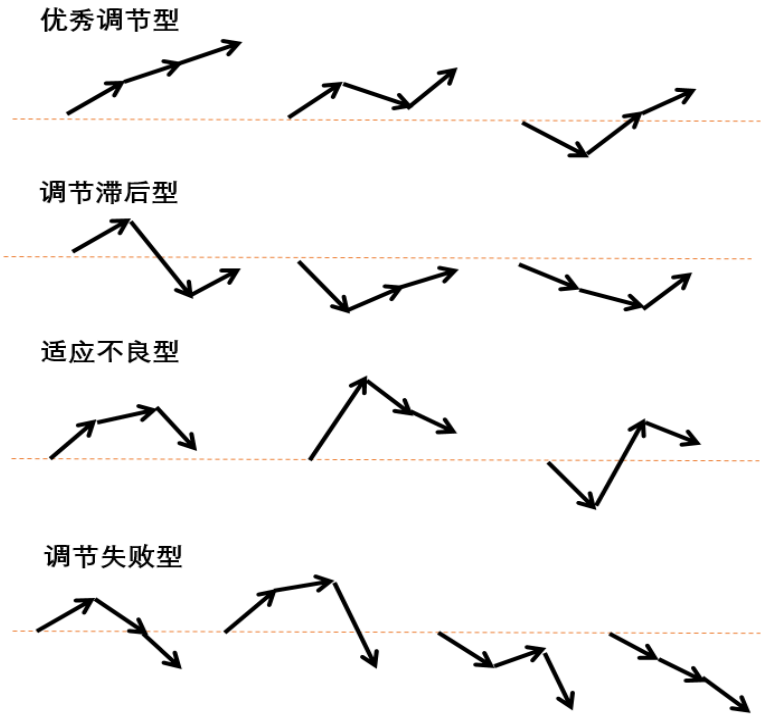
在线自我调节行为模式模型

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

案例4 ——基于数据挖掘的在线自我调节学习用户特征与模式分析——以H学院网络公共选修课为例



不同学习类型学生行为模式比例图



案例4 ——在线创客社区讨论深度分析研究-以D网站为例

目录

一．研究背景 1
二．研究设计 5
2.1 研究样本的选取 5
2.2 研究方法 7
2.3 分析框架 8
2.4 编码和一致性检验 9
三．数据分析 13
3.1 对于帖子的总体描述 13
3.2 对各领域数据的解释 15
3.3 帖子讨论与社交网络之间的关系 17
四．研究结论 21
4.1 创客社区的高活跃度人群过于集中 22
4.2 创客讨论社区缺乏深度学习的引导 22
4.3 创客讨论社区讨论区域功能混乱 22
五．优化策略 23
5.1 成员等级梯度化 23
5.2 培养创客导师引导学习者深度学习 24
5.3 主题板块多样化 25

在线创客社区讨论深度分析研究-以 D 网站为例

小组成员：张明心 单俊豪 郭日发 胡梦华

一．研究背景

随着信息时代的到来，互联网发展迅速，数字技术已经逐渐渗透到了人们生活的方方面面。互联网与教育的结合，是教育领域的一大创新。而近几年创客运动的兴起，更是给教育领域带来了新的生机与活力。

创客，最早起源于美国 MIT 比特与院子研究中心发起的 Fab Lab 创新项目，该项目的核心理念是个人创意，个人设计，个人制造，旨在构建以用户为中心的创新制造环境。创客呈现出的是一种无拘无束、自由发挥、全民参与的社会文化。它可以锻炼参与者承受失败额能力，因为一件成功作品的背后一定会经历了多次的失败。同时创客秉承着“做中学”的核心理念，在解决问题的过程中不断学习知识，并通过实践将其运用和巩固提升。

创客运动，可以理解为“互联网+DIY”，是指任何人都可以利用身边的一切资源，无论是软件还是硬件，通过动手操作将自己的想法变成现实，并且通过互联网平台将其分享给全世界。[1]近年来，各种创客空间、创客社区、创客大赛、创客嘉年华的不断出现，大大促进了创客之间的智慧交流和碰撞，诞生了一大批极具创意的产品。

创客运动与教育的“碰撞”，正在慢慢改变传统的教育理念、组织、模式与方法，创客教育应运而生。创客教育是一种融合信息技术，秉承“开放创新、探究体验”教育理念，以“创造中学”为主要学习方式和以培养各类创新型人才为目的的新型教育模式。[1]最开始的创客教育是一种比较简单的、面向对象较少的、技术要求相对较低的教育，比如创客运动兴起的典型标志 Arduino 采用 2 层板结构和 16MHz 频率的低技术来实现，但是这种初期的创客教育对于培养参与者的技术能力和创造能力有很大的局限性。[2]随着“互联网+”的广泛应用，技术的不断进步以及材料成本的下降，创客教育逐渐开始面向大众。

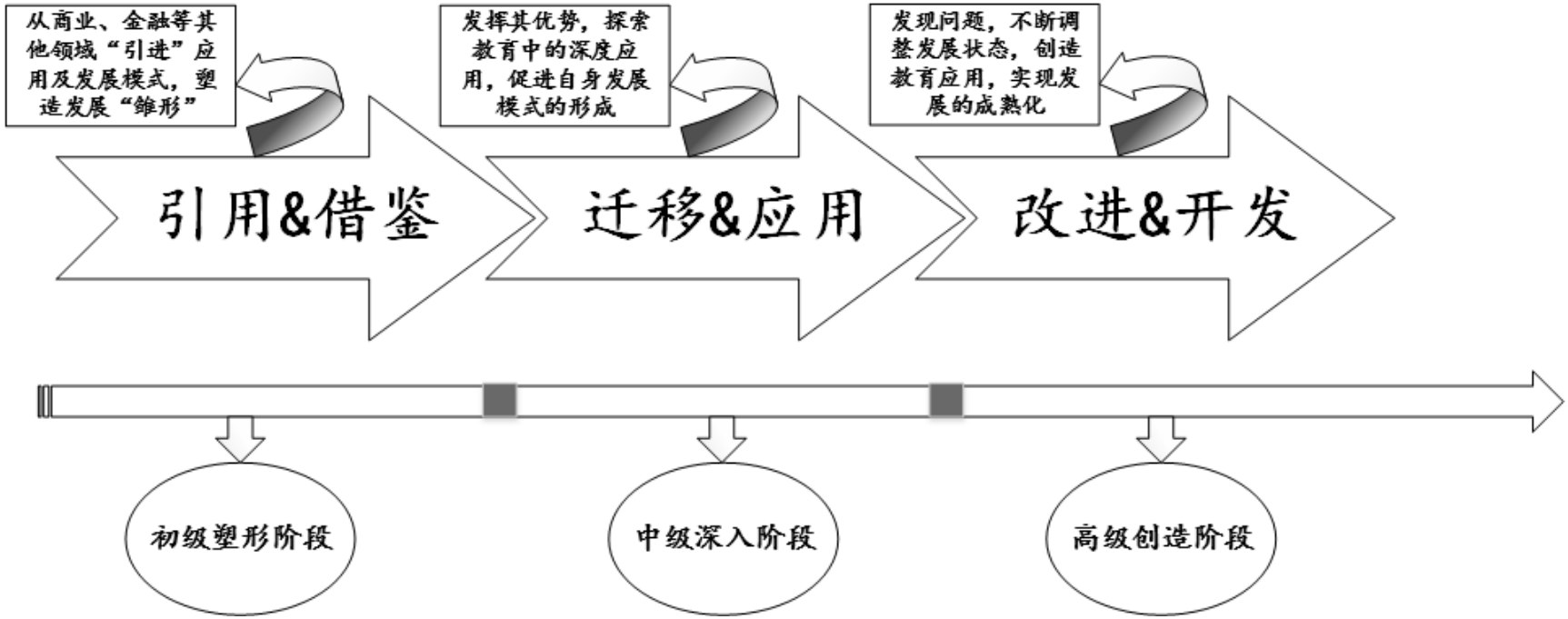
我们可以从两个角度去理解创客教育：一种是“创客的教育”，旨在培养创客人才，可以通过开设专门的创客课程，建立创客空间，配备专业化的指导教师进行实施；另一种是“创客式教育”，旨在应用创客的理念与方式去改造教育，需要将创客运动倡导的“动手操作、实践体验”理念融入各学科教学过程，开展基于创造的学习。然而无论是“创客的教育”还是“创客式教育”，最终的教育目标是一致的，即培养具有创新意识、创新能力和创新思维的创新型人才。[1]

目前，创客教育在国内外线上和线下都有广泛的应用。作为创客运动的发源地，美国的创客教育在全球也是最具规模和影响力的。美国的创客教育得到政府部门的高度重视和社会力量的大力支持。目前，全美已有 MIT、哈佛大学、斯坦福大学等在内的近百所高校开设了创客空间，不限年级、不限学科，面向所有学生开放。[3]较之传统的实验室，高校创客空间具有更高的开放性、自由度和灵活性，正在成为培养拔尖创新人才的新试验田。

以美国的联合国国际学校为例，该学校于 2013 年初建立了合作实验室（Co Laboratory），配备了各种制作工具和材料，如螺丝刀、硬纸板、米尺、3D 打印机、等等，还有用于演示交流的交互式墙体

[mw_shl_code=cpp,true]#include #inclu
教程 (1)
25
CAPITAL OF STATISTICS PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

R语言教育应用生态发展脉络图

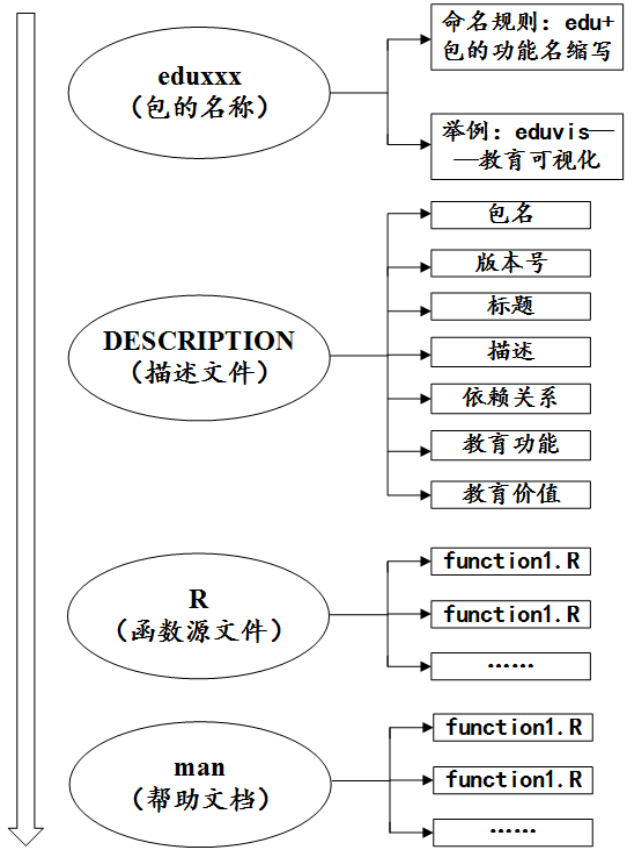


发展趋势及应用前景

国家大数据战略规划的引导实践
数据统计及可视化的主流工具
“人工智能+教育”的指导思想
学习分析技术的数据挖掘手段
为数据驱动的学习和评估提供有效工具和平台
开源特性对教育大众的影响
推动精准教育的持续发展
R

R语言教育包的开发架构设计

- 提出了明确的命名规则，即education（教育）的简写“edu”+功能名的方式。
- DESCRIPTION中即包括基本信息（包名、版本号、标题、描述、依赖关系等），也包括包的教育功能、教育价值以及针对的教育问题等。
- R（实现包的功能的函数源文件），遵循R语言包编写的原则，包括教育功能的程序代码及标注。
- man (帮助文档)，用简明的语言，阐述清楚包的操作指南，并且附有典型的实际案例，帮助用户快速的上手使用。



R语言教育社区建立方式

- 项目驱动的教育社区
- 开放融合的生态系统
- 更新迭代的社区激励方式
- 形成具有教育特色的复杂自适应循环系统

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说
知识共享平台

自然基金—教育技术

国家自然科学基金委信息学部增设F0701教育信息科学与技术二级代码：

- F070101 教育信息科学基础理论与方法
- F070102 在线与移动交互学习环境构建
- F070103 虚拟与增强现实学习环境
- F070104 教学知识可视化
- F070105 教育认知工具
- F070106 教育机器人
- **F070107 教育智能体**
- **F070108 教育大数据分析与应用**
- **F070109 学习分析与评测**
- **F070110 自适应个性化辅助学习**

教育数据集（Harvard DataVerse）

华东师范大学 EAST CHINA NORMAL UNIVERSITY

Dataverse 关于 User Guide 支持 注册 登录

Harvard Dataverse

Metrics 2,871,970 Downloads

Contact Share

Education Find Advanced Search Add Data

- Dataverses (42)
- Datasets (11,131)
- Files (5,148)

Dataverse Category
Research Project (11)
Journal (8)
Organization or Institution (8)
Researcher (7)
Research Group (2)

Metadata Source
Harvested (10,099)
Harvard Dataverse (6,222)

Publication Date
2015 (5,597)
2012 (2,487)
2008 (1,490)
2011 (1,161)
2016 (934)

More...

Subject
Social Sciences (1,643)
Medicine, Health and Life Sciences (359)
Arts and Humanities (123)
Other (30)
Agricultural Sciences (26)

More...

Author Name

1 to 10 of 16,321 Results Sort

Education Dataverse (Murray Research Archive)
May 12, 2015 Murray Research Archive Dataverse

Education Dataverse (UNESCO Institute for statistics)
Oct 1, 2016 UNESCO Institute for Statistics Dataverse

Medical **Education** Research Dataverse (University of Illinois at Chicago)
Feb 2, 2012

Dan Koretz Dataverse (Harvard School of Education)
Jun 6, 2007

African Higher **Education** Data Dataverse (Centre for Higher Education Transformation)
Apr 6, 2014
Open data collected by the Centre for Higher **Education** Transformation (CHET) on various aspects of African higher **education**, including enrolments, graduates, staff data and knowledge outputs.

Project Star Dataverse (Health and Education Research Operative Services)
Oct 8, 2008
Health and **Education** Research Operative Services (HEROS), Incorporated is a 50
is to assess the outcomes of programs and initiatives that impact children, families
and initiatives will achieve greater outcomes as a result

CAPITAL OF STATISTICS PROFESSION, HUMANITY & INTEGRITY

IT大咖说 知识共享平台

R语言教育应用微信群

目标和愿景

1. 建立R语言教育应用社区，开放、交流、分享、共进。
2. 构建R语言教育应用平台，推进数据驱动精准的教育教学。
3. 推动R语言教育应用服务，促进教育教学创新、变革、发展。

CAPITAL OF STATISTICS
PROFESSION, HUMANITY & INTEGRITY
IT大咖说 知识共享平台

谢谢聆听与指正！

吴永和
华东师范大学教育学部教育信息技术学系
上海数字化教育装备工程技术研究中心

EAST CHINA NORMAL UNIVERSITY