

大数据智能与 图书馆服务创新

刘柏嵩@2018.10.30, 无锡
长三角地区高校图书馆联盟馆长论坛

2018年10月29日

lib.nbu.edu.cn

汇报提纲

 一、进展与挑战

 二、场景与实践

 三、趋势与展望

一.进展与挑战

什么是智能

智力或智能

- 是指生物一般性的精神能力。这个能力包括以下几点：理解、计划、解决问题，抽象思维，表达意念以及语言和学习的能力。

智力三因素理论 (Robert Sternberg)

- 成分性智力，指思维和问题解决所依赖的心理过程
- 经验智力，指人们在两种极端情况下处理问题的能力：新异的或常规的问题。
- 情境智力反映，在对日常事物的处理上，它包括对新的和不同环境的适应，选择合适的环境以及有效地改变环境以适应你的需要。



Robert Sternberg
(1949-) 是美国心理学家和心理测量学家。他是康奈尔大学人类发展教授。

智能与智慧的区别

- ▣ 智能是一种**能力**，涵盖语言、逻辑思维、空间认知、肢体运作、音乐审美、人际交流、自我反省等。
- ▣ 智慧是一种**状态**，智慧让人深刻地理解人、事、物、社会、宇宙、现状、过去、将来，拥有思考、分析、探求真理的能力。

什么是人工智能

人工智能（AI）

也称作机器智能，是指由人工制造出来的系统所表现出来的智能。通常人工智能是指通过普通计算机实现的智能。

人工智能研究

研究内容：包括[认知建模](#)、[知识学习](#)、[推理及应用](#)、[机器感知](#)、[机器思维](#)、[机器学习](#)、[机器行为](#)和[智能系统](#)等。

研究动机：包括[推理](#)，[知识](#)，[规划](#)，[学习](#)，[交流](#)，[感知](#)，[移动和操作物体](#)的能力等。

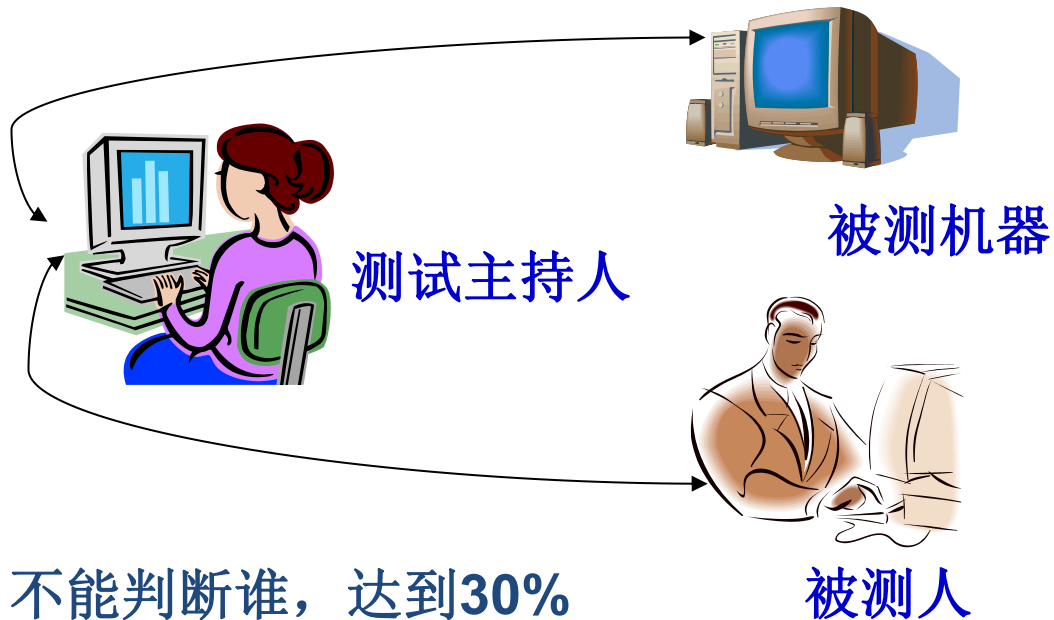
基础知识：包括搜索和数学优化，逻辑，基于概率论和经济学的方法等

应用系统：目前有大量的人工智能应用系统，如AlphaGo，Siri等。

图灵论题

- 在定义智能时，英国科学家图灵做出了贡献：如果一台机器能够通过图灵实验，那它就是智能的；让人在不看外型的情况下不能区别是机器的行为还是人的行为时，这个机器就是智能的；
- “如果机器在某些条件下，能够非常好地模仿人回答问题，以至提问者在相当长时间里误认它不是机器，那么机器就可以被认为是能够思维的。——图灵《计算机器与智能》1950年
- 注：获得图灵奖就等于物理学家获得诺贝尔奖一样，图灵在理论上奠定了计算机产生的基础，没有他的杰出贡献世界上根本不可能有这个东西，更不用说什么网络了。

图灵 (Turing) 测试



图灵测试的误读

- 📖 一个常见的说法是把他提出的“图灵测试”作为人工智能的标准或定义：
 - 👉 (1) 这个说法不符合人工智能的史实；
 - 👉 (2) 把图灵测试作为人工智能的定义是不恰当的
 - 👉 (3) 其实图灵本人并不是这个意思。
- 📖 该测试在人工智能学术界颇有争议。

人工智能普及应用



2017年5月

中国乌镇

AlphaGo 3:0 完爆

世界围棋冠军柯洁九段!

由CMU团队研发的Libratus战胜4位德州扑克顶级选手：无限德州扑克是一种“不完全信息博弈”，相较于围棋难度更大。



AlphaGo Zero

- 老大AlphaGo Lee, 2016年3月 4:1 打败韩国棋手李世石
- 老二AlphaGo Master, 2017年5月 3:0 打败中国棋手柯洁
- 老三AlphaGo Zero, 2017年10月宣布 自学围棋40天, 打败所有人

规则

棋盘/黑子/白子

棋子被围起来就死

刚开始, 随机下子

热衷于吃子, 完全不顾死活

发现了如果先占住棋盘的边和角, 后面占便宜

学会了如何「打劫」、「征子」

懂得看「棋形」

棋士柯洁 10月19日 02:22

一个纯净、纯粹自我学习的AlphaGo是最强的... 对于AlphaGo的自我进步来讲... 人类太多余了

老三战绩

第3天, 下了490万局棋, 打败老大

第21天, 败了老二

第40天, 完整走过了一个人类棋手学棋的全过程

语音识别

从实验室走向市场

■ 2012年，微软发布同声传译产品

采用深度神经网络方法，一举将语音识别错误率降低近30%，推动了语音识别的产品化进程



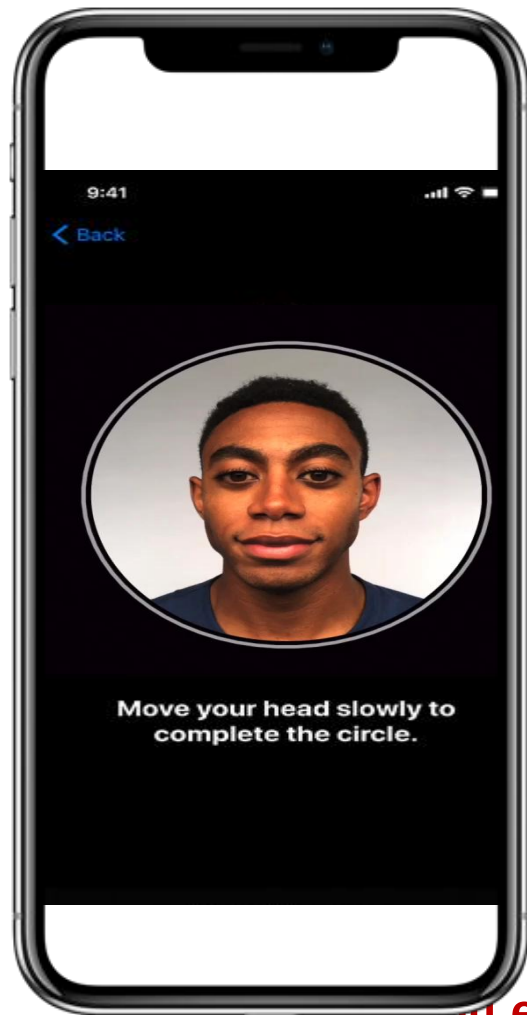
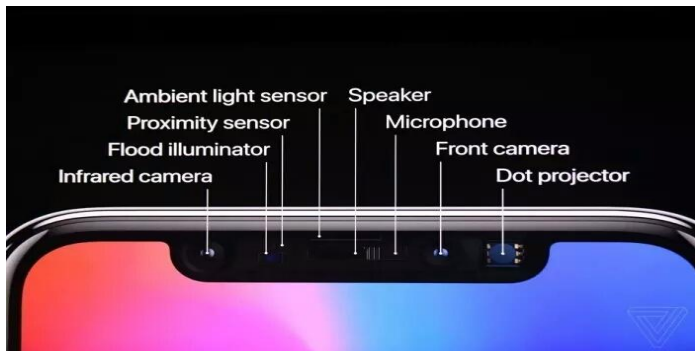
■ 2017年，微软文字转录超过人类转录员水平

微软语音到文字转录词错率为5.1%低于人类单一转录员的5.9%

人脸识别

超越人类水平:达99.53%

- 2017年9月，苹果公司发布产品Face ID
采用深度神经网络方法，在真实生活场景中
将人脸识别错误率降低至百万分之一。



影像识别

医学影像识别

取代人工读片成趋势



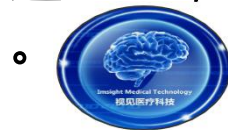
- 2016年，肺癌检出率超过放射技师水平

2016年，美国Enlitic公司发布了基于深度神经网络的癌症检测系统，用于从 X光、CT 扫描、超声波检查、MRI 等图像中发现恶性肿瘤。



- 2017年，自动读片速度超过医师检测速度

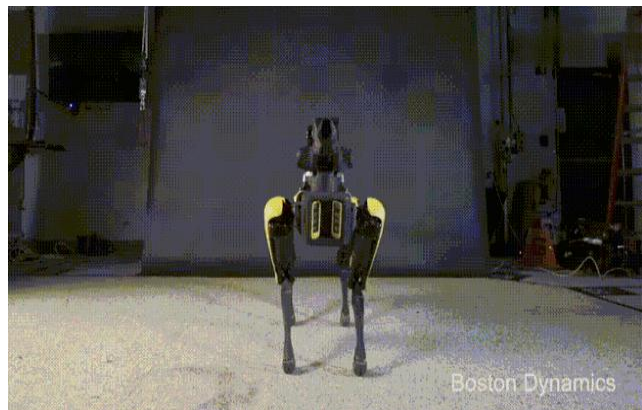
2017年，视见医疗科技公司基于深度神经网络的人工智能影像识别系统，对病理组织切片的识别率高达99%，并减少10分钟检测时间。



视见医疗科技

仍在不断突破

🏠 LYNA: 在有谷歌AI 算法辅助的情况下，6名有美国医生从业执照的专科病理学家的诊断时间平均缩短了一半。-20181014



基于大数据的精准医学

组学

(基因组、转录组、蛋白质组、代谢组.....)

大数据 (人群、队列)

基因型与表型的关联

建立表型组数据库

生物信息学 生物网络

系统生物学

当前国际两大科学前沿的交汇

在此基础上融合临床检验、影像学等指标

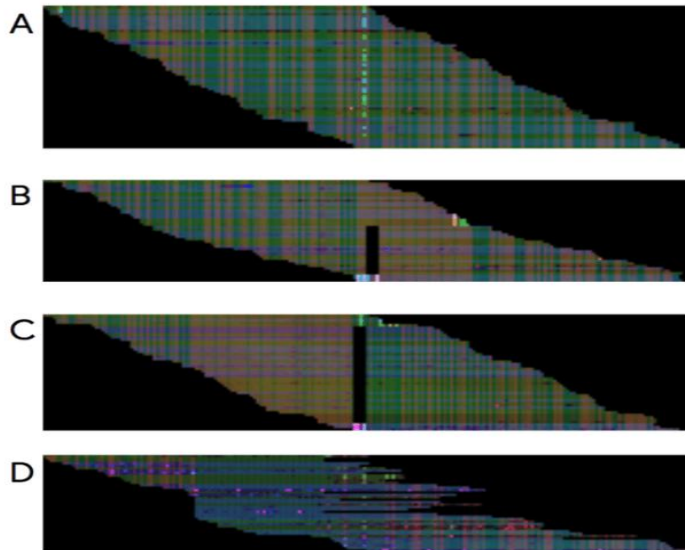
“人工智能+基因” 新的应用进展

IBM Watson Health - 个人健康领域的认知计算系统

IBM Watson Health



Google 发布基于深度学习AI技术的基因组变异精准识别工具
DeepVariant



阿里AI鉴黄师上线

阿里Ai鉴黄师
阿里安全，用技术解决社会问题

耳听八方
无论你说汉语、英语、日语，还是东北话、四川话、广东话……我都能听辨出来。

火眼金睛
超过99.5%色情图片我都能认出来。

一抵十万
我一天能处理10亿张图片
完胜10万人的审核量
谁能有我效率高？

咬文嚼字
发帖，发评论，发弹幕……都要小心
羞羞的词都能被我过滤掉。

模型养成
纵有图片千万张，我一天就能学完！

阿里安全
ALIBABA SECURITY

近期其他进展

- 🏠 MIT人工智能实验室最新研究成果：AI系统不仅可以识别假新闻，还能辨别个人偏见
- 🏠 斯坦福人工智能模型实现地震的精准预测；
- 🏠 清华大学「九歌」—基于深度学习的中国古典诗歌自动生成系统
 - 🎨 基于超过 30 万首的诗歌进行训练学习，能够产生集句诗、绝句、藏头诗、宋词等体裁的诗歌

AI让我们

 Know more:知道更多

 Do more: 做到更多

目前主要突破只在深度学习



人工智能的三要素



A 算法

B 数据

C 算力



AI四个重要时间节点

- 🏠 第一代（1956-1970）模仿人
- 🏠 第二代（1971-2000）数据驱动
- 🏠 第三代（2001-2016）大数据+深度学习
- 🏠 第四代（2016-至今）超级人工智能
 - 🌊 IA: 智能增强，通过数据来增强人的创造和智能
 - 🌊 II: 智能基础设施，象网约车和餐馆推荐系统

吴军，《超级智能时代的两个关键因素》CNCC2018,10月26号

智能引擎助力信息技术：从算法到系统和生态



经过近10年的增长，2018年8月2日苹果公司股票市值首次突破1万亿美元大关

公司	市值 (美元)	AI举措
苹果公司 (1976年)	1.01万亿	重新架构人工智能、SIRI和机器学习团队，由谷歌搜索和人工智能前主管John Giannandrea 领导
微软公司 (1975年)	8412亿	人工智能技术下的智能云与智能边缘计算
谷歌公司 (1998年)	8551亿	人工智能和安全是谷歌公司优先考虑
亚马逊 (1995年)	9017亿	云计算、智能音箱Echo

钢铁、电力、硬件、软件的发展之路

公司	首次超过的市值 (美元)	现在市值 (美元)
美国钢铁公司 United States Steel Corporation(1901 年)	1901年首次超过10亿	54.53亿
通用电气公司(1892年)	1955年首次超过100亿	1134亿
IBM公司(1911年)	1987年首次超过千亿	1340亿
微软公司(1975年)	1999年首次超过5000亿	8447亿
苹果公司(1976年)	2018年首次超过1万亿	1.05万亿
下一个? (计算服务)		

- 2007年中石油上市当天市值7.7万亿人民币，是首家突破万亿美元市值公司，目前市值为1387亿人民币。

人工智能的发展：云、网、端、芯

Worldwide IaaS Public Cloud Services Market Share, 2016-2017 (Millions of U.S. Dollars)

Company	2017 Revenue	2017 Market Share (%)	2016 Revenue	2016 Market Share (%)	2017-2016 Growth (%)
Amazon	12,221	51.8	9,775	53.7	25.0
Microsoft	3,130	13.3	1,579	8.7	98.2
Alibaba	1,091	4.6	670	3.7	62.7
Google	780	3.3	500	2.7	56.0
IBM	457	1.9	297	1.6	53.9
Others	5,902	25.0	5,392	29.6	9.5
Total	23,580	100.0	18,213	100.0	29.5

Source: Gartner (August 2018)

亚马逊、微软、谷歌、IBM在云服务上的份额（2017年度和2016年度）

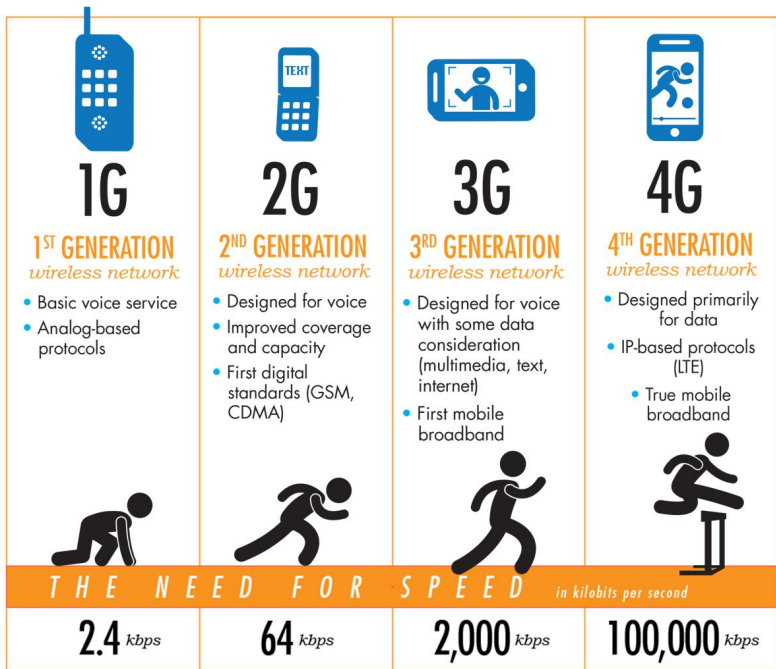
Public Cloud Leadership by Region – Q1 2018

Rank	Worldwide	North America	EMEA Region	APAC Region	Latin America
Leader	AWS	AWS	AWS	AWS	AWS
#2	Microsoft	Microsoft	Microsoft	Alibaba	Microsoft
#3	Google	Google	Google	Microsoft	Google
#4	Alibaba	IBM	IBM	Google	Salesforce
#5	IBM	Salesforce	Salesforce	Tencent	IBM

Source: Synergy Research Group

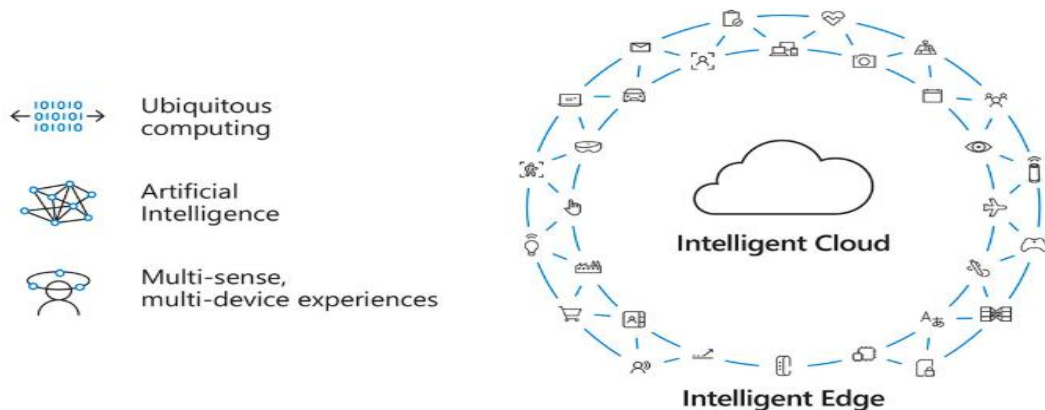
各地区云计算份额排列
(EMEA: Europe, the Middle East and Africa)

人工智能的发展：云、网、端、芯



新一轮人工智能热潮下的公司发展：云、网、端、芯

- 预计到2020年，智能设备总数将超过200亿台。这些设备将非常智能，甚至无需一直连接到云端，就能具备看、听、推理、预测的能力。
- 边缘计算可以理解为是指利用靠近数据源的边缘地带来完成的运算程序
- 由于数据本地化计算与存储，边缘计算能够在一定程度上保护数据隐私



新一轮人工智能热潮下的公司发展：云、网、端、芯

类别	特点	Training--Inference	代表
CPU			Intel
通用芯片	通用性强、功耗高、性能好	1--4	英伟达(其GPU占全球AI芯片市场的50%以上份额)、AMD
半定制FPGA (field-programmable gate arrays)	低延时、低功耗、可编程	2--3	赛灵思(Xilinx)、英特尔的Mobileye和Altera、微软Project Brainwave等
全定制ASIC (application-specific integrated circuit)	针对特定任务优化、性能稳定、功耗可控	3--2	Intel (Nervana)、谷歌(TPU)、寒武纪、地平线、Facebook(NNP)
神经形态	功耗低、处于早期研制	?	IBM (True North), Qualcomm(spiking networks), Intel (spiking circuitry)

AI 芯片-装满了 TPU 的谷歌云服务器



TensorFlow 框架下的神经网络，驱动了谷歌的图像识别、翻译、搜索和游戏；

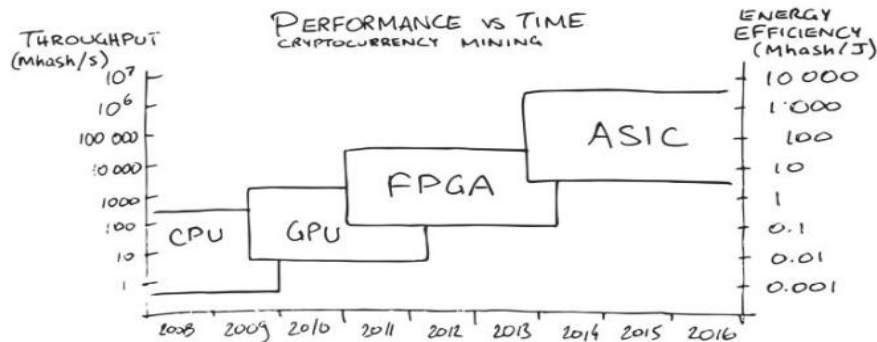
TPU 效率要比通用CPU高 30-80 倍；

目前已为全球 10 亿人提供日常服务

新一轮人工智能热潮下的公司发展：云、网、端、芯

- Benchmarking: Accuracy, Power, Throughput, Cost
- IBM用大规模的模拟存储器阵列训练深度神经网络（DNN），在计算能效（28,065 GOP/sec/W）和通量（3.6 TOP/sec/mm²）上，超过了当前GPU的两个数量级。

Any system that is very power efficient will tend to



equivalent-accuracy accelerated neural-network training using analogue memory, *nature*, 558, 60-67(2018)

形成了AI的三层生态系统

🏠 基础层

🏠 GPU云计算平台+深度学习平台+大数据

🏠 英伟达、英特尔、寒武纪

🏠 技术层（部分通过图灵测试）

🏠 听说（语音识别与语音合成）：科大讯飞

🏠 看（视觉识别）：旷视科技

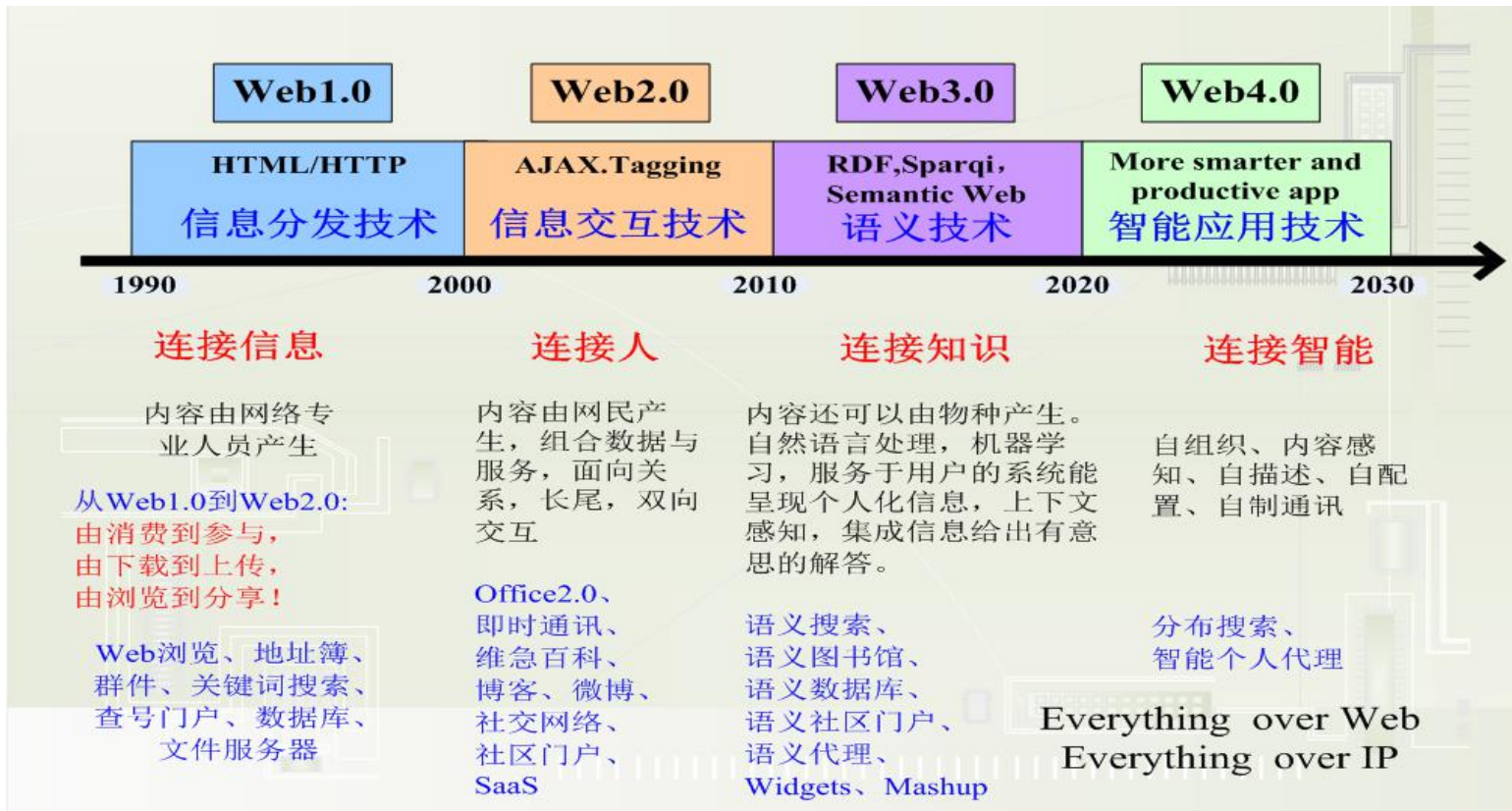
🏠 读写（自然语言处理）：微软、百度

🏠 应用层

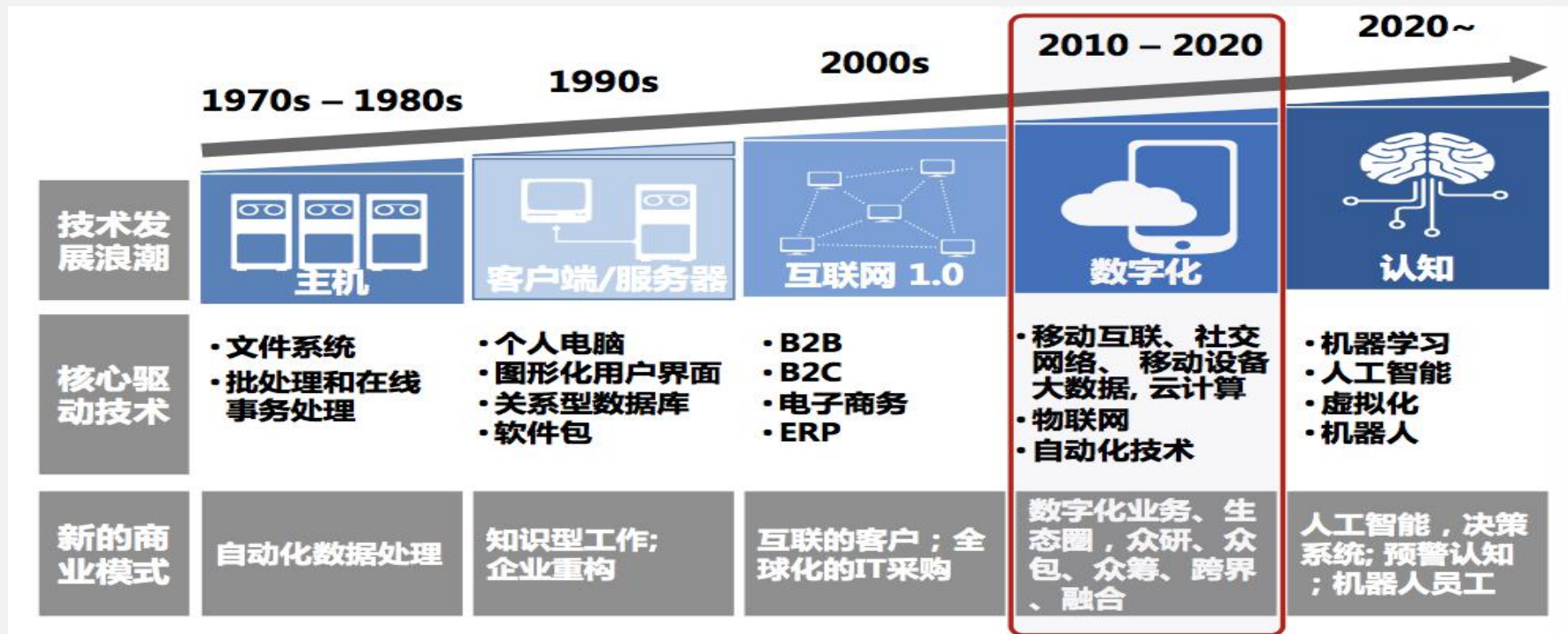
🏠 AI的SaaS应用：交通（无人驾驶）、教育、金融、医疗

互联网的技术演进

Packet Switching Invented 1964
Hypertext Invented 1965
ARPANET 1969
TCP/IP Created 1972
WWW Created 1989
Mosaic Created 1993
Age of eCommerce Begins 1995
Semantic Web 1999



新技术科技革命进程加快



Artificial Intelligence

Big Data

Cyber-Physical-System

智慧图书馆的提出



——刘兹恒，北京大学信息管理系

目前还是概念化阶段

🏠 概念很美好

🏠 现实很骨感

🎡 业务互联程度不高：

🏠 数据商应用商百花齐放，数据标准并不统一

🎡 管理与服务效率不高

🎡 用户体验并不便利

图书馆当前面临的挑战

🏠 业务互联方面：

- 📖 单馆业务的纸电分离；
- 📖 线上线下业务基本割裂。

🏠 管理与服务效率方面：

- 📖 难以实时感知读者需求；
- 📖 无法全面掌握空间与资源的利用情况。

🏠 用户体验方面：

- 📖 社交功能缺乏；
- 📖 主动推送、个性化服务欠缺。

智慧化路径的可能选择

- 📖 大数据涌现和计算能力大幅提升的趋势；
- 📖 图书馆如何顺应人工智能带来的红利：
- 📖 **赋能路径**：从数字化-》数据化-》智慧化

二、场景与实践



真诚图书馆

沈从文



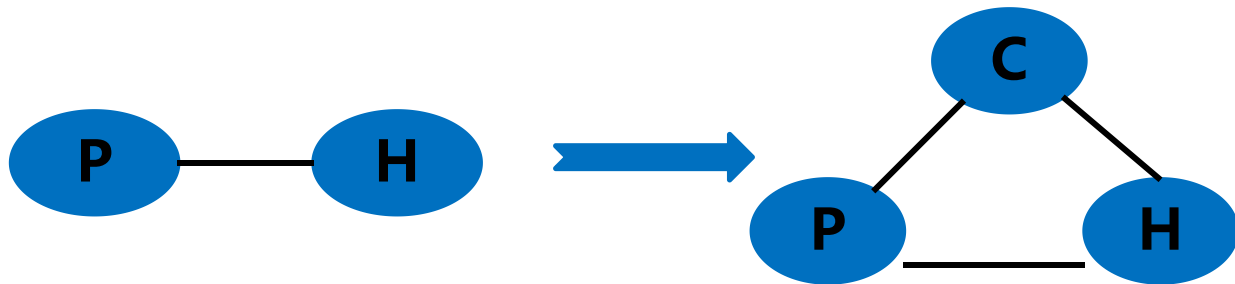


学术报告厅



从二元空间进入新的三元空间

世界原来是二元空间：**人类社会空间**（H）和**物理空间**（P）。但近年来，信息力量的迅速壮大，已长成除P、H两极之外的新一极：**信息空间**（C）。

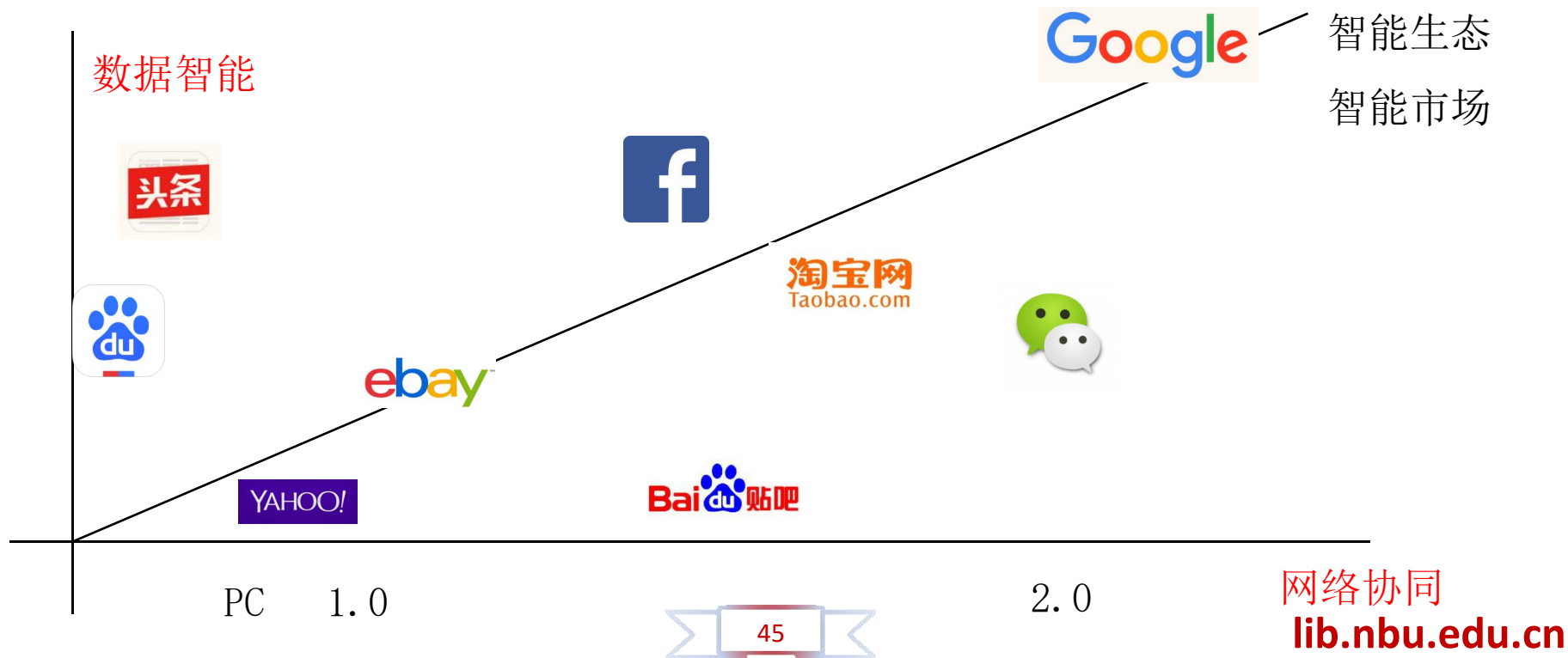


近50年来，信息空间成长壮大的历程

- 信息来自于人类社会：书籍、交流、媒体、计算机
- 信息开始互联：互联网、移动通讯和搜索
- **信息绕过人类**，直接来自于物理世界：传感器和物联网
- 产生大数据：信息快速膨胀，人类已无法单独处理
- 从大数据必然会走向**大知识**，并推动人类认识与控制能力的大变化

新商业文明DNA的两个螺旋

有两个螺旋：一个是网络协同，一个是数据智能



图书馆智能应用场景

🏠 服务:

- 📖 1. 自动问答
- 📖 2. 智能搜索, 智能推荐
- 📖 3. 虚拟助理 (VPA)、学术助理
- 📖 4. 智能技术帮助基层民众、特殊群体阅读

🏠 空间:

- 📖 5. 上架机器人, 无人图书馆
- 📖 6. 智能导引、座位推荐

🏠 资源:

- 📖 7. 自动分类、自动摘要、自动标引
- 📖 8. 精准情报、学者画像、读者画像

1、数据来源

🏛️ 数据是城市最重要的自然资源

🌈 (王坚@云栖大会2017)

🏛️ 数据更是图书馆的重要资源

🌈 是**智能的唯一可靠基础**；与数字图书馆不同

🏛️ 数据主要来源:

🌈 学术资源；图书馆空间

🌈 用户数据

图书馆的“大数据”

NBUL&IC
NBUL&IC



图书馆域三类主要数据

📖 学术资源

📖 书目数据（元数据）：GB级

📖 数字资源（扫描及全文数据）

📖 TB级-》PB级；资源**数据化**

📖 读者数据

📖 GB级-》PB级（校内业务系统、个人空间）

📖 **行为数据**（流通借阅、数字阅读、评价行为）

📖 外部数据

📖 互联网及关联数据

大数据与传统数据的差别

在线

大数据必须是永远是在线的，是随时能调用的。不在线的数据不是大数据。只有在线的数据才能马上被计算、被使用。

实时

瞬间完成匹配呈现，这才叫大数据。

全貌

大数据还有一个最大的特征，它不再是样本思维，它是一个全体思维。以前一提到数据，人们第一个反应是样本、抽样，但是大数据不再抽样，不再调用部分，我们要的是所有可能的数据，它是一个全貌。其实叫全数据比大数据更准确。

传统的数据应用与大数据的区别

传统

基于目标收集

有意识参与

单向

大数据

自动收集

无意识参与

双向

数据从四个维度产生作用

NBUL&IC

NBUL&IC

📖 把数据变成知识和智慧去指导行动：

📖 **决策**：通过数据让我们更好理性决策，不是基于经验，而是通过数据：校级、部门级

📖 **优化**：通过数据分析，进行服务的优化。

📖 **推广**：基于读者的个性化需求，进行精准推荐与阅读推广：读者画像。

📖 **创新**：通过数据的方式，形成新的服务，新的业务增长点：

📖 对学习、科研和管理方面



增加与读者接触点

📖 记录每一位读者的行为数据

📖 到馆的服务

👤 好的借阅环境只是基础

👤 读者感知、互动

📖 不到馆的服务

👤 接触点分布在网站和移动APP中

📱 支付宝：公众号+余额宝+快车

👤 如何更多更方便地收集到读者的兴趣

👤 如何让读者参与服务中

加强对读者的洞察

📖 读者的交互数据（线上线下）都会产生很多的数字足迹

- 📖 他们什么时候开始搜索论文、图书关键词
- 📖 他们的兴趣爱好是什么
- 📖 他们喜欢电子还是纸质图书
- 📖 他们喜欢坐在哪里阅读
- 📖 他们经常预约研讨间否
- 📖 除了他自己的专业资源外，还爱看哪些资源

📖 将各类数据进行关联

2、数据分析

🏠 学术资源

- 📖 全文级-知识点提取

- 📖 精准化-数据抽取，包含**科学数据**

🏠 用户行为数据

- 📖 个性化地对待每一个读者

- 📖 个性化服务：包含线上和线下

🏠 管理数据

- 📖 安全与运维


学术资源的全文分析

全文全媒体的知识点抽取

实体识别

 人物；事件；关联关系

数据抽取

 时间、GDP

图表抽取

引用关系及引用态度

基于全文的知识服务



书是输入



提取语义关系



分解：知识元

Basic information:

full name	Wen Jiabao
date of birth	September 1942 (age: 68 years)
place of birth	Tianjin, Tianjin

Images:



Leadership positions:

official position	Premier
country	China
start date	2003-03-16

Timeline:

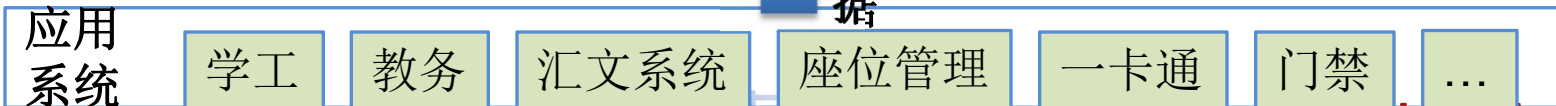
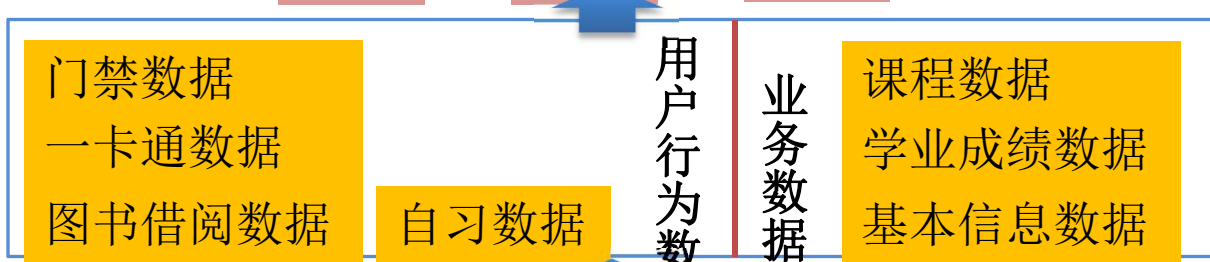
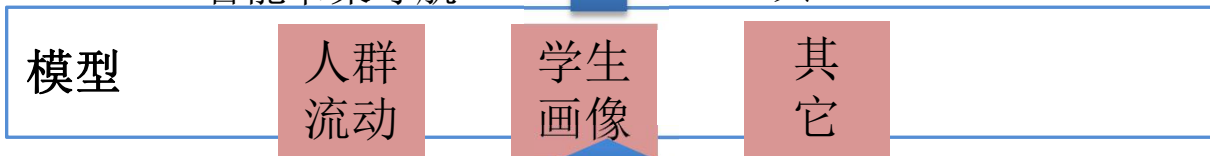


Wen Jiabao

基于需求组合

高校的智能应用场景

- | | | |
|------|--------------|------------|
| 决策服务 | ● WIFI热点部署优化 | ● 挂科预警 |
| | ● 餐厅座位、饭菜安排 | ● 失联预警 |
| | ● 阅览座位分布优化 | ● 大学课程设置指导 |
| | ● 图书采购 | ● 资源个性化推送 |
| | ● 智能书架导航 | ● 其它 |



◆ ◆
感知 用户 精准
无

人脸识别技术

术 室内定位技

我校的AI应用探索

业务
数据

- 学生画像
- 优秀学生特征挖掘
- 消费分析
- 失联预警
- 人脸识别

智能
学习

智慧
生活

业务数据情况

数据来源

学工系统

- 籍贯、民族、出生年月、政治面貌等

教务系统

- 专业、班级、学院、课表、课程、成绩等

汇文系统

- 当前借阅、历史借阅、书目信息等

科研管理系统

- 科研项目、科研成果

门禁系统

- 门禁数据等

座位管理系统

- 自习数据等

一卡通系统

- 消费、餐厅、用水数据等

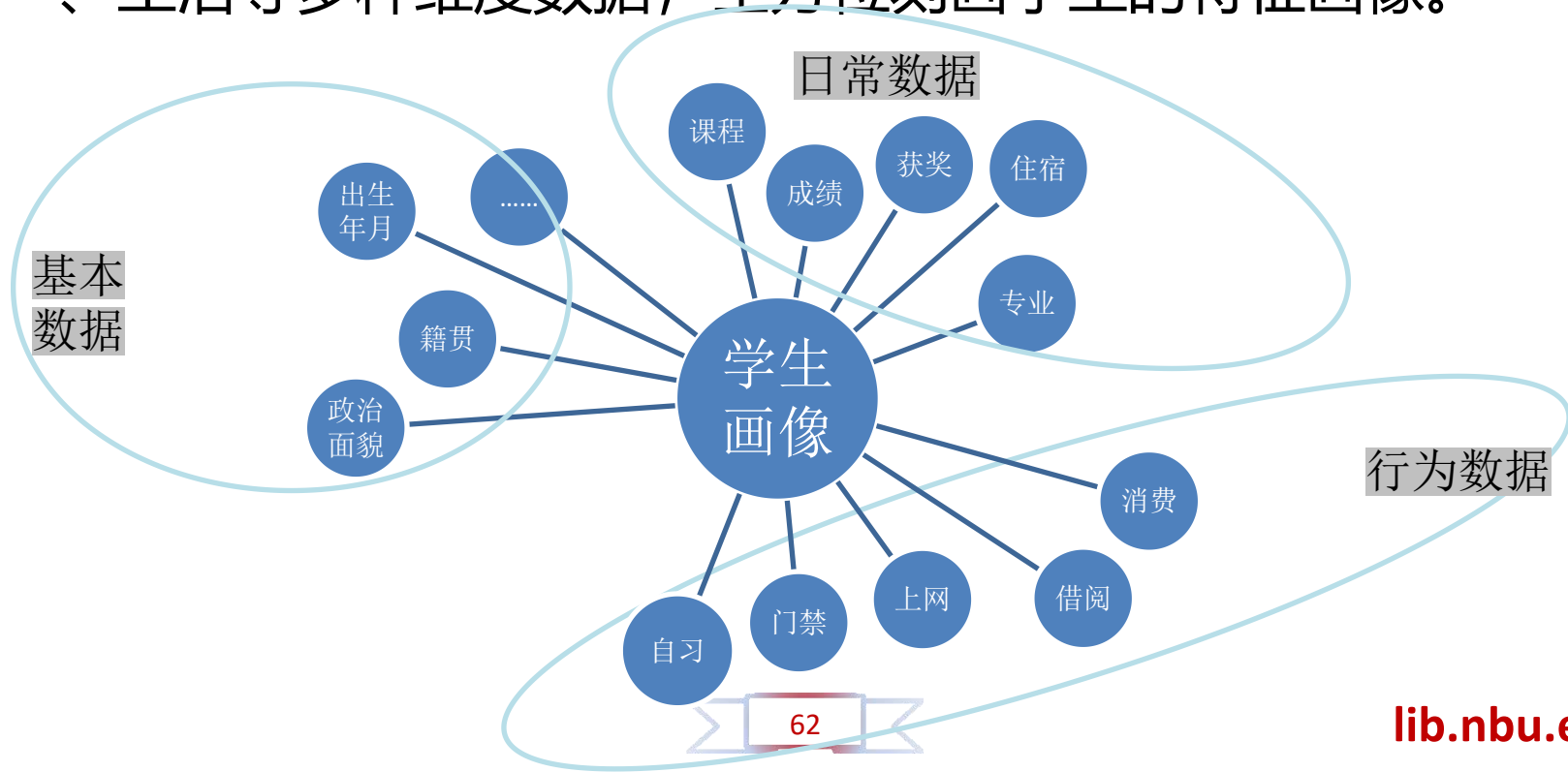
业务数据情况

数据量

数据名称	数据总量 (条)
基本信息数据	10.2873万
学业成绩数据	160.0948万
自习行为数据	350.3231万
图书借阅数据	500.3918万
一卡通数据	9000多万
课表数据	25.6463万
共计	亿级

高校学生画像

- 学生画像描绘学生在校期间的学习、生活等信息，综合学习、生活等多种维度数据，全方位刻画学生的特征画像。



+ 人物画像、班级画像

通过挖掘人物属性、上网日志行为、校园内部应用数据，并结合多方领域数据补充，形成庞大人物标签体系。将细节到人、班级进行细颗粒的用户、班级画像。为数据分析提供支持。

人口属性：

年龄、性别、年龄、电话、政治面貌、关键人

兴趣爱好：

篮球、健身、养生、机车、电影、没事、电子竞技.....

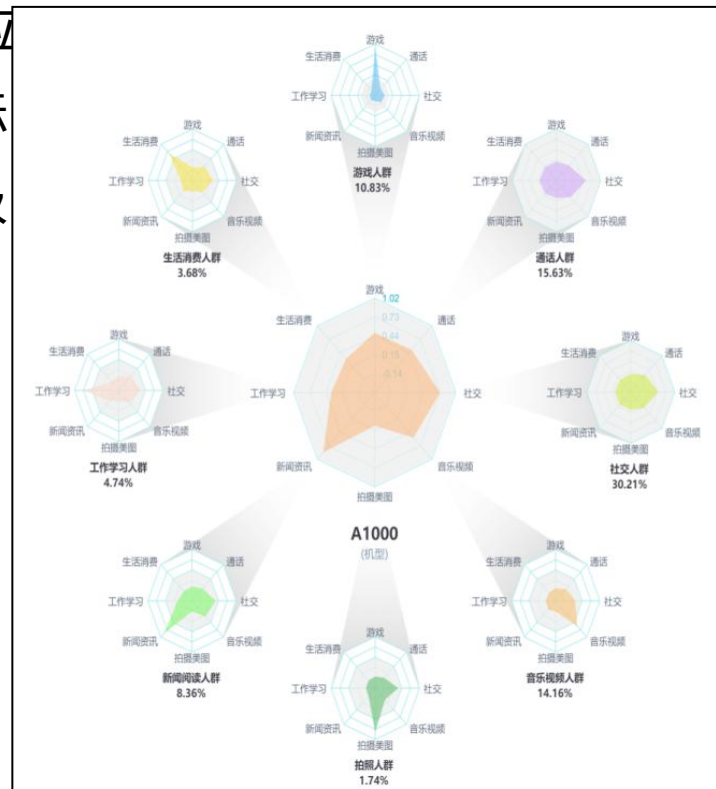
商业兴趣：

旅游、加电数码、体育运动

学习属性：

逃课、图书馆、挂科、请假、成绩、学业、科研、学历

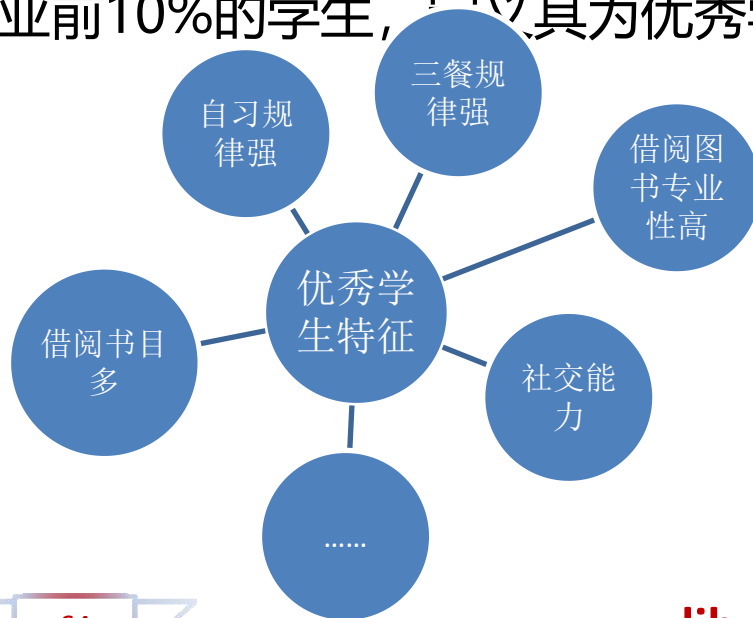
.....



学霸画像

- 针对优秀学生，结合优秀学生的相关数据，对学生自习规律性、学生阅读爱好和学生社交活跃度等情况，在不同学院、年级的优秀学生群体数据进行不同维度的数据分析和对比。
- 在校学生单个学期成绩排名专业前10%的学生，定义其为优秀学生。

- 基本信息数据
- 学业成绩数据
- 课程数据
- 一卡通数据
- 自习行为数据
- 图书借阅数据



读者画像

- 📖 做到 “千人千面+每时每刻”
- 📖 基于在物理空间和数字空间的数据收集
- 📖 引入互联网数据
- 📖 行为分析
 - 👤 用户建模
 - 👤 社交关系分析：同学，兴趣组，师兄弟
 - 👤 关联分析

数据就是决策



可以低成本积累**活数据**，机器通过数据处理，**直接**产生管理**决策**。

互联网时代跟传统时代比较起来，一个是做完整的实时数据记录，一个是要事后去搜集数据用于支持决策，完全不一样。

读者能跟你实时互动，给你反馈

对于管理者来说，最重要的还是你的**洞察力**。在你所在的这个业务环境下，用什么样的方法将数据、读者体验、算法有机地结合在一起，让读者体验每时每刻都得到改进

数据驱动采访与决策

🏠 建立图书馆“数据集市”

- 📊 实时看到整体运行状况和用户需求

🏠 馆藏政策：纸质书借阅率，刊的保障率

- 📊 自动统计分析本校的专业和学科

- 📊 自动匹配、生成纸质与数字资源采访列表

🏠 决策分析

- 📊 空间布局：读者的学习习惯，粉丝的需求

- 📊 与学业成绩的关联；与科研产出的关联

- 📊 对学科（团队）的数据分析

图书馆运营数据成为学校决策依据

- 📖 科研绩效分析——管理决策
- 📖 学校ESI分析——团队优化、集中冲击
- 📖 学院发文引用分析——学科优势
- 📖 社科绩效分析——武大评估体系
- 📖 通过平台化策略-实现快速高效
 - 📖 分为三个层面：校领导；部门、学院；教师
 - 📖 工具化、轻量化
 - 📖 学术头条

学术图谱与学术评价

🏠 [Open Academic Graph](#): 集成了

🌐 微软学术图谱 (MAG): 1.66亿

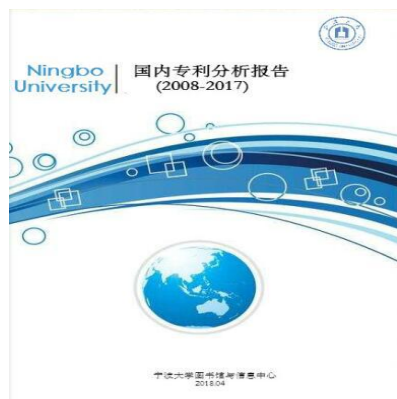
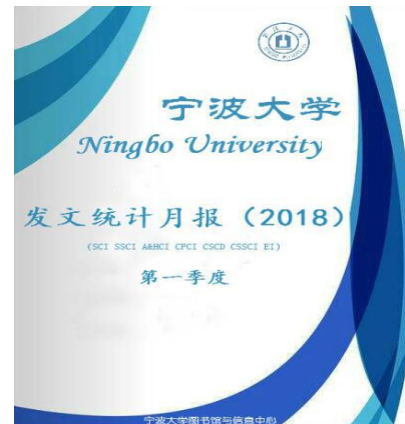
🌐 AMiner学术图谱: 1.55亿

🏠 CNCC2017提出构建全球学术大数据:

🌐 一是集成丰富的学术知识数据;

🌐 二是数据共享; 三是服务共享。

🏠 学术地图画像、人才迁徙图



+按“数”找人（才）



通过数据分析，辅助人事部门获取全球顶尖技术人才信息和建立人才关系网络。

压力与褥疮形成的原因及预防

莫元雪 梁仕英 周云君

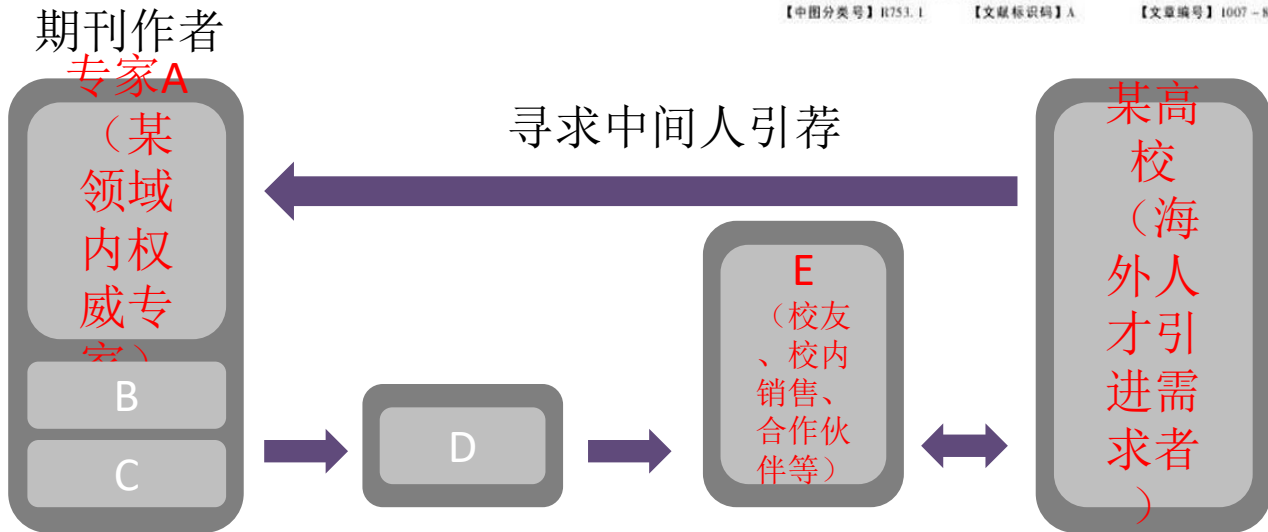
广西省柳城县人民医院·广西 柳城 545200

【关键词】 褥疮；压力；原因分析；预防

【中图分类号】R753.1

【文献标识码】A

【文章编号】1007-8517(2011)09-0014-02



E与学校为强人际关系人物

人才评价

表 1-XXX 学者发文年度情况

	2008	2010	2011	2012	2013	2014	2015	2016	2017	合计
发文量	1	2	4	3	3	3	5	6	1	28
WOS 中被引	259	3	104	133	14	12	13	18	0	556
被引合计	277	4	116	148	15	13	15	19	0	607

表 2-XXX 通讯作者发文年度情况

	2010	2011	2012	2013	2014	2015	2016	2017	合计
发文量	2	2	1	2	2	4	4	1	18
WOS 中被引	3	51	119	3	11	13	9	0	209
被引合计	4	59	133	4	12	15	10	0	237

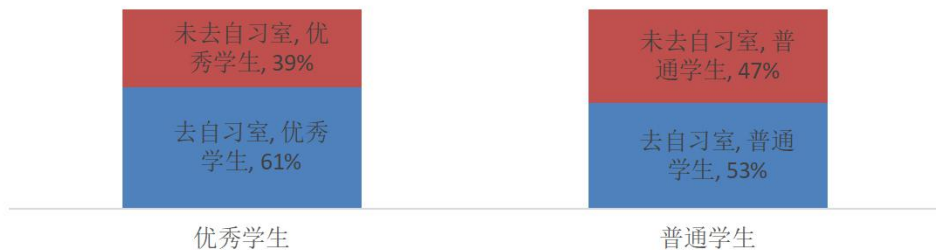
表 12-XXX 学者发文的 ESI 学科表现

序号	ESI 学科	WOS 论文数	学科规范化的引文影响力	被引频次	高被引论文	国际合作论文	引文影响力	期刊规范化的引文影响力
1	Plant & Animal Science	9	2.28	196	1	1.634	21.78	1.3
2	Molecular Biology & Genetics	6	0.31	14	0	0.175	2.33	1.08
3	Biology & Biochemistry	5	0.9	51	0	0.765	10.2	2.07
4	Environment/Ecology	5	0.12	8	0	0.12	1.6	0.7
5	Multidisciplinary	1	8.6	249	0	18.684	249	5.4
6	Clinical Medicine	1	1.18	20	0	1.501	20	3.88
7	Computer Science	1	0	0	0	0	0	0

读者行为-优秀学生特征

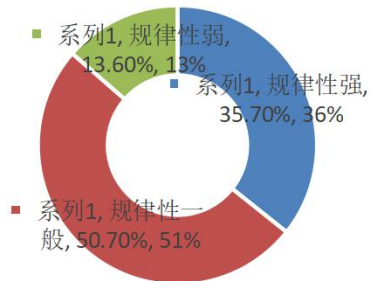
优秀学生和普通学生是否去自习室情况

■ 去自习室 ■ 未去自习室



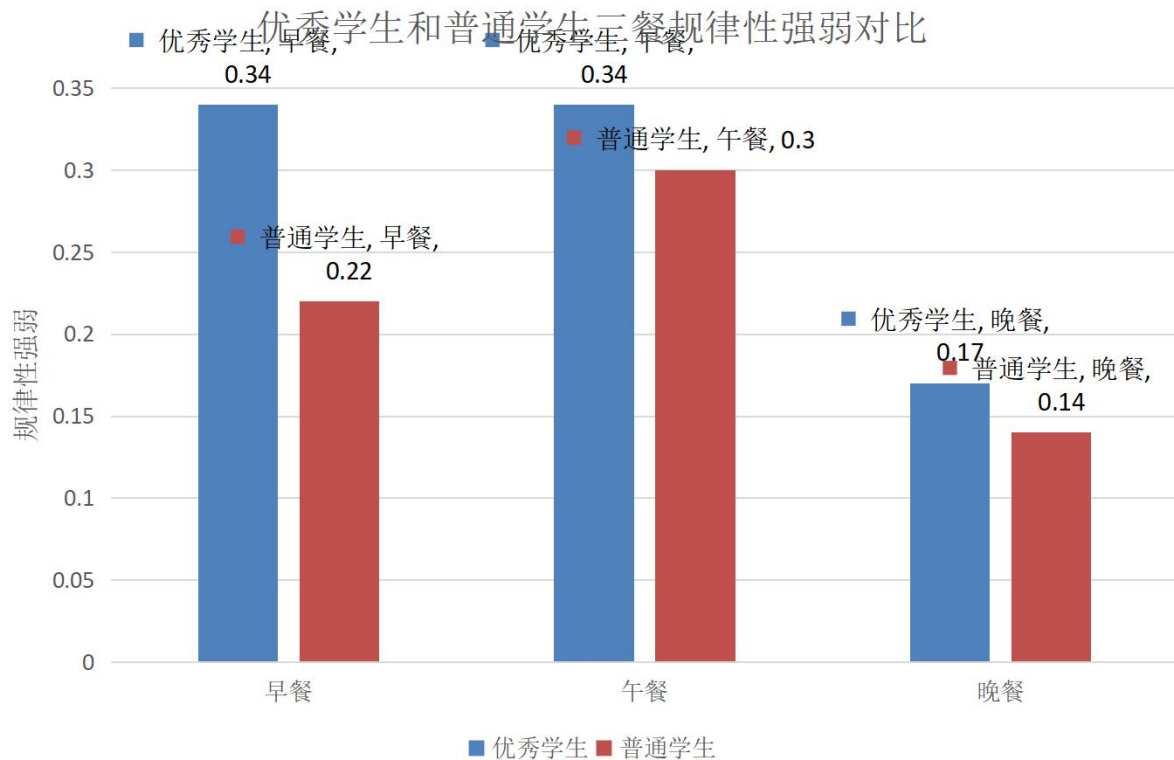
📖 优秀学生去图书馆的比例大于普通学生去图书馆的比例。


优秀学生去自习室规律性强弱分布



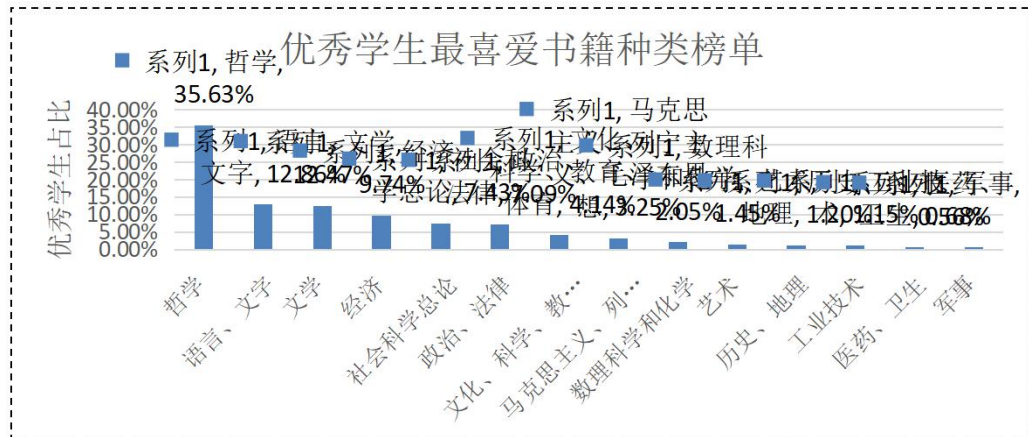
📖 36%的优秀学生去自习室的规律性强，13%的优秀学生去自习室规律弱。

优秀学生行为规律性分析

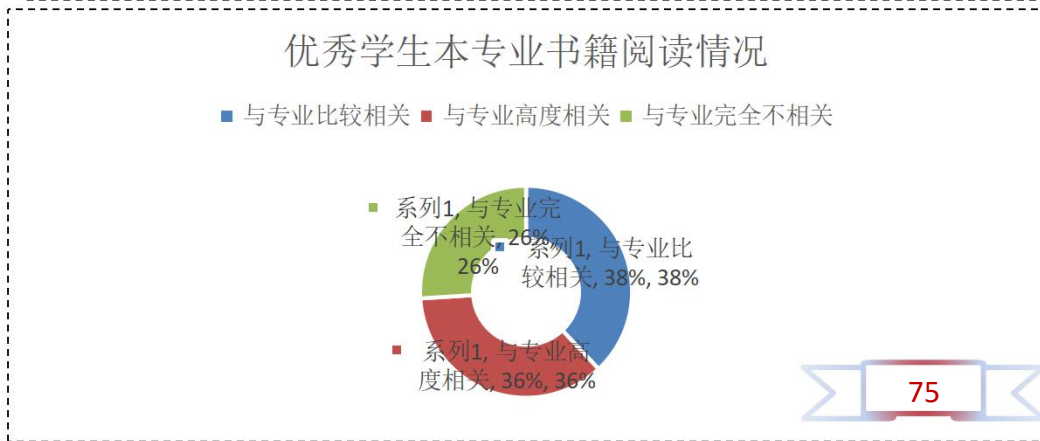


 饮食规律分析是根据学生一日三餐去食堂消费的时间构成时间序列分布，结果表明优秀学生三餐的规律性都强于普通学生。

优秀学生阅读爱好分析



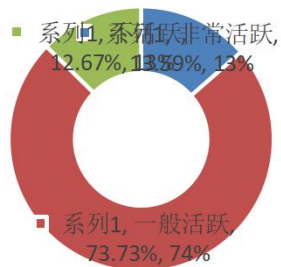
多数优秀学生喜欢阅读哲学类的书籍，而榜单上的其他书籍多为一些专业性书籍。



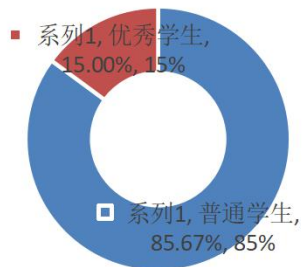
在学生最喜欢书籍种类中，38%的优秀学生其阅读爱好跟其专业高度相关，36%的优秀学生其阅读爱好与专业比较相关。

优秀学生社交行为分析

优秀生活活跃度情况

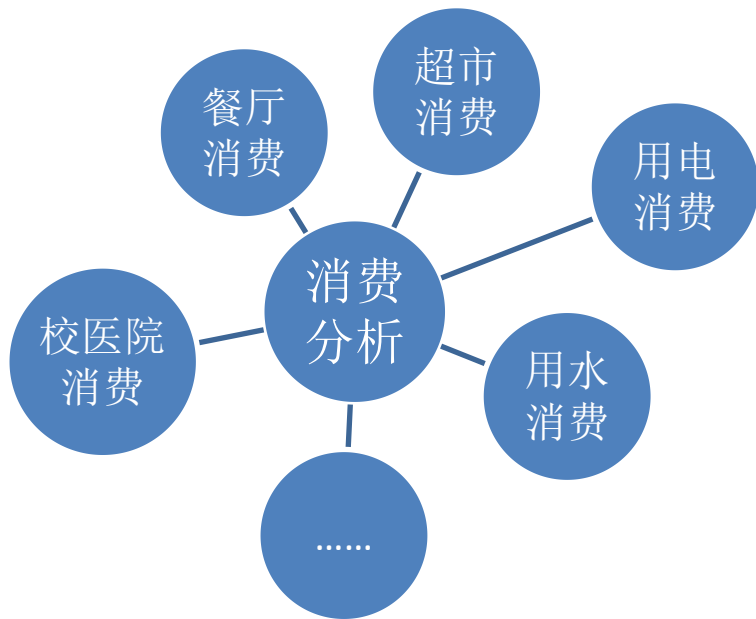


优秀学生最亲密朋友分布图



- 社交活跃度是衡量这个人与他人互动频率、效果等的量化的综合指标。
- 有13%的优秀学生在社交方面表现为活跃，13%的优秀学生在社交方面表现为不活跃。
- 15%的优秀学生其关系最亲密好友也是优秀学生。

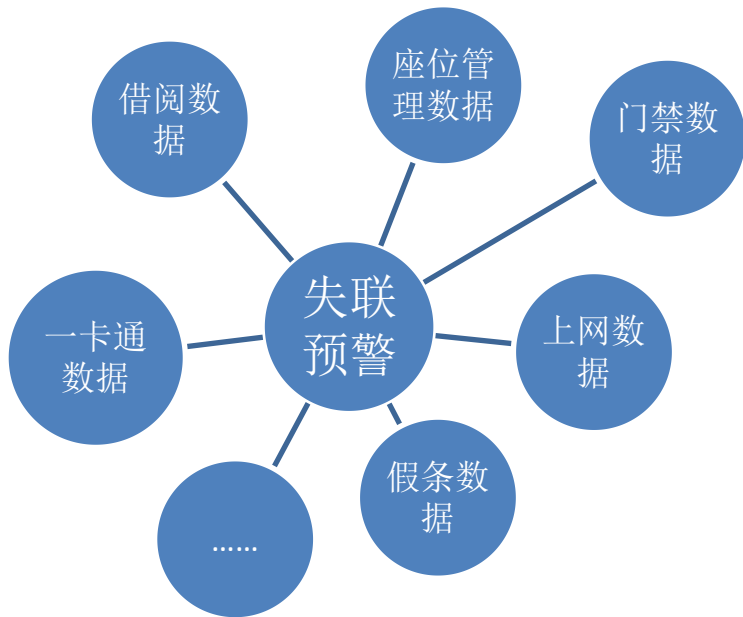
学生消费行为分析



- 通过一卡通数据对师生在食堂、超市、医院、用水、用电等账单及消费习惯进行分析，以全面掌握自己在校期间的消费报告。
- 包括根据月度、学期查看各消费类型金额和占比，各个消费类型的消费排名，喜欢光顾的商店、餐厅

家乡味道

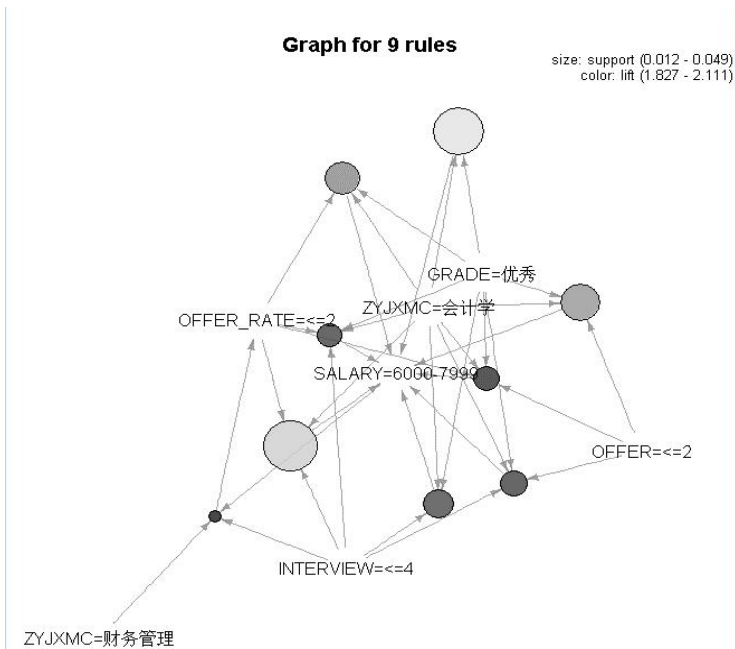
失联预警



- 根据学生校内消费、上网、进出门禁等行为数据，结合请假信息、校历信息及管理教学活动信息
- 预判学生不在校的可能性，将预警信息通过邮件推送给相应老师。

人才培养的关联分析

- 毕业生签约数据挖掘—专业、成绩、面试数、offer数、面试offer比、月薪之间的关联关系



条件	结果	支持度	置信度	提升度	说明
{面试数 <= 4, 专业=财务管理, 月薪=6000-7999}	= {面试offer比 > <= 2}	0.012	0.85	2.111	财务管理专业月薪在6000-7999的学生中面试offer比较低, 说明财务管理专业学生要求税前月薪在6000-7999时比较容易拿到offer
{offer数 <= 4, 专业=信用管理, 成绩=优秀}	= {月薪=6000- > 7999}	0.007	0.909	1.942	信用管理成绩优秀的学生税前月薪一般是6000-7999
{面试数 <= 4, 专业=金融学, 成绩=优秀}	= {月薪=6000- > 7999}	0.008	0.885	1.890	金融学成绩优秀的学生税前月薪一般是6000-7999
{专业=会计学, 成绩=优秀}	= {月薪=6000- > 7999}	0.046	0.855	1.827	会计学成绩优秀的学生税前月薪一般是6000-7999

- 会计学、金融学、信用管理成绩优秀的学生月薪一般是6000-7999
- 成绩普通的学生通过增加面试数也可以找到月薪较高的工作

3、智慧情报

语义搜索

 数据平台：学术数据集市

智能咨询

 在线客服

 个人助理

 扎克伯格定律（每隔一年互联网内容翻番）

知识服务

 无连接-》弱连接-》强连接（信用建立）-》知识社群
（圈子）

自动写作、机器翻译

📖 当面临海量的信息时，深度学习表现了特殊的能力，信息越多维度准确性就越高，很容易就超过人类

🌟 《纽约时报》主编是机器人——Blossom

👤 据调查显示，一半以上的读者看后分不清到底是不是人写的。

🌟 华尔街的高盛公司：智能投资机器人

👤 实时完成上市公司的公告、财报、行业分析，并发现行业上下游关系、供应链关系，形成投资“知识图谱”。

🌟 百度：**为你写诗**

📖 Google神经机器翻译：已接近人类专业水平

语义搜索-评价科学家的影响力

NBUL&IC



FAQ Contact Sign in

Cut through the clutter

Find peer-reviewed research from the world's most trusted sources

Search over 40 million papers from ArXiv, PubMed, and more...

Try: Oliver W. Press Radioimmunotherapy Deep Learning

基于**4000万**篇文献进行语义分析；而且持续增加。

识别论文中的不同部分（引言、方法、结果、图表），识别作者对所引用文献的态度；能在一定程度上**理解文献内容**的人工智能搜索引擎；

并可对科学家或科研组织对后续研究的**影响力评估**。



语音AI

🏠 谷歌Duplex:

🌐 2018 I/O开发者大会上，AI会自己打电话，能够模仿真人语气、语速，以流畅的人机交互方式帮助用户完成美发沙龙和餐馆的预定操作；

🏠 知识引擎研究计划:

🌐 阿里巴巴宣布联合清华和浙大学等5所高校发布「藏经阁」研究计划。

智能咨询

📖 咨询机器人

📖 语音识别

- 📖 实时记录读者的问题，并显示在屏幕上，系统搜索并调取优秀（资深）馆员以往对这些问题的回复；
- 📖 咨询馆员快速学习，并形成专题知识库。

📖 语音合成

- 📖 “读小说”、“朗诵诗歌”给读者听；
- 📖 读者录音=》读者语音合成；
- 📖 咨询问题及回答自动翻译成方言或英语、法语等

唐代人物资料的数据化

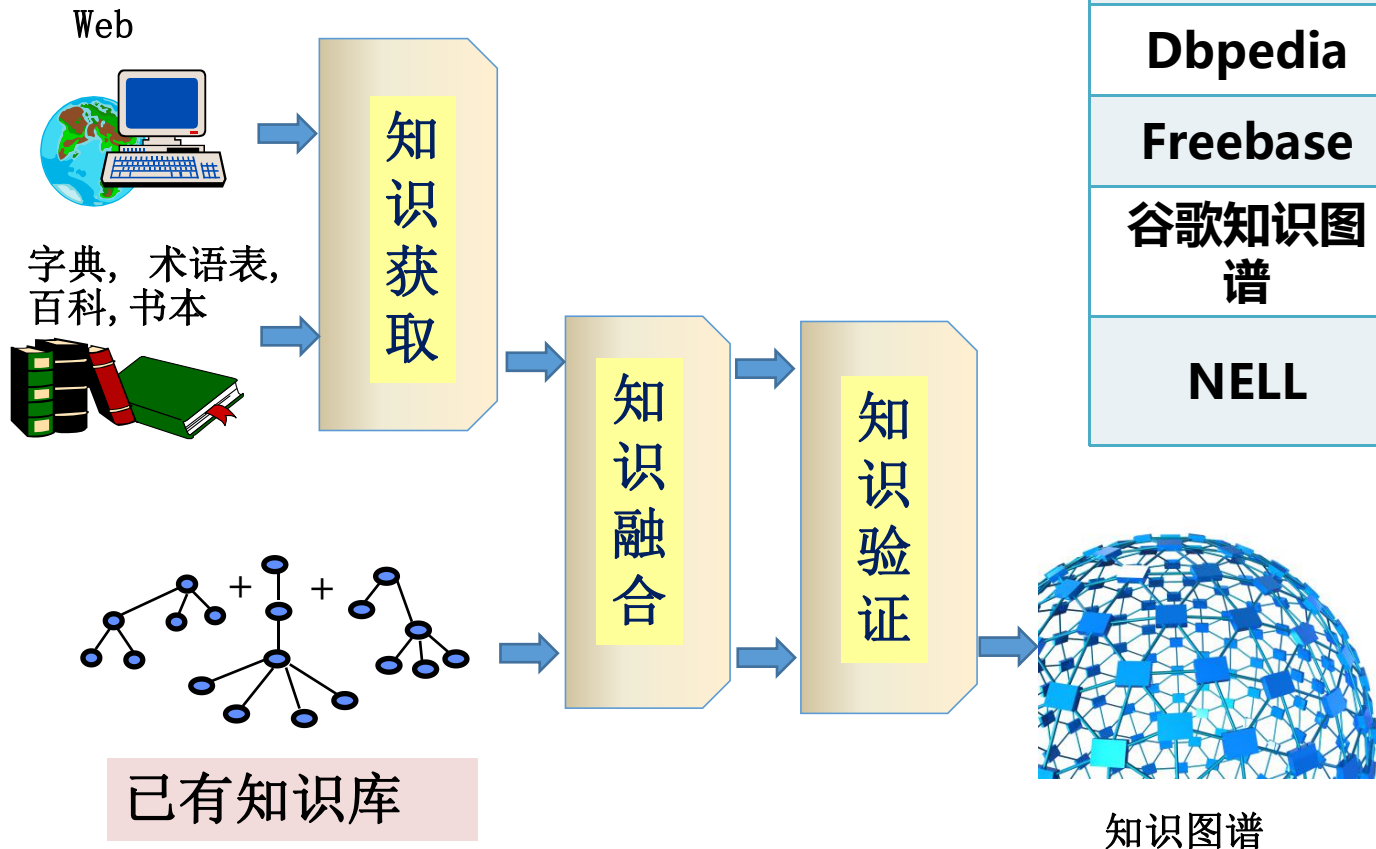
📖 史料的电子化-》“数据化”

📖 45,000多人的姓名、时间、地址、职官、入仕途径、著作、社会区分、亲属关系、社会关系、财产、事件

📖 训练计算机判断古籍文本中人物之间的关系

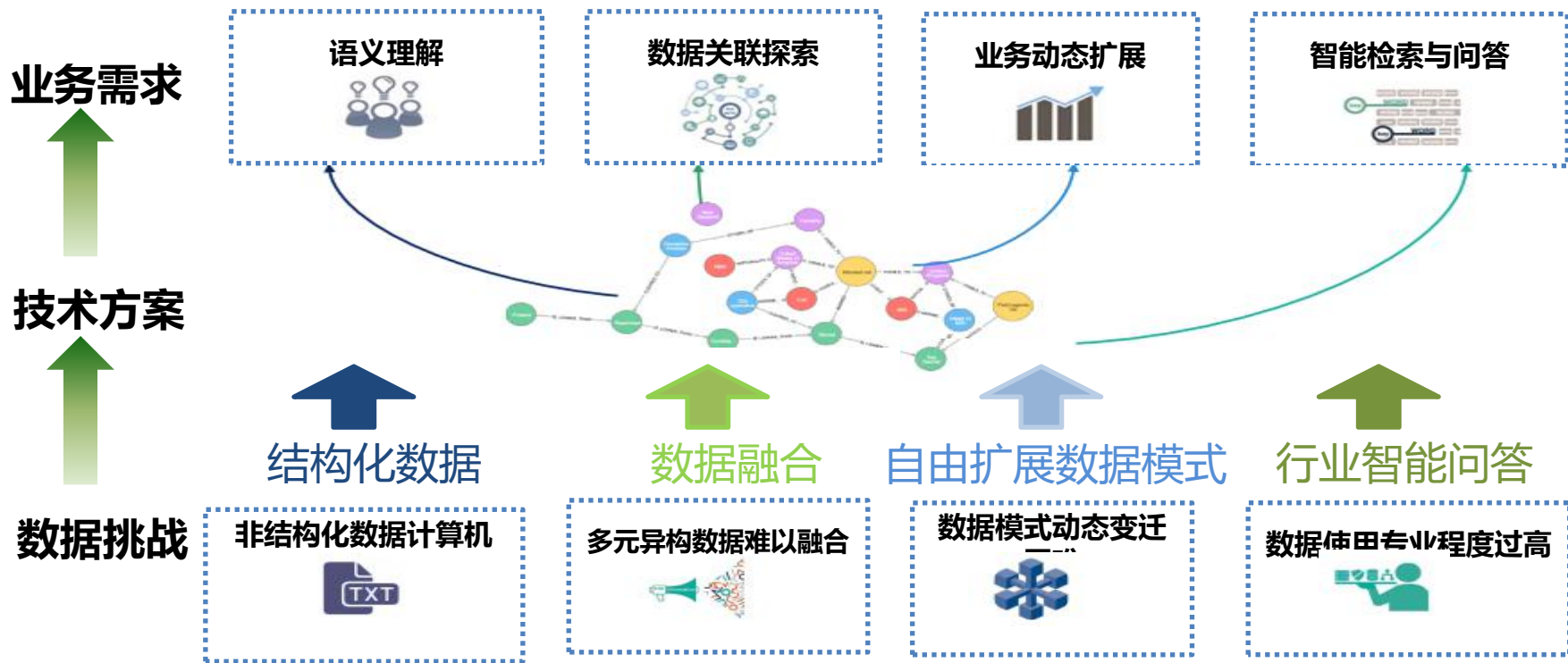
📖 哈佛大学、北京大学、“中央研究院”历史语言研究所共同主持

知识图谱构建



名称	规模
Yago	1千万实体, 35万类别, 1.8亿事实, 100种属性, 100语言
Dbpedia	4千万实体, 250类别, 5亿事实, 6000种属性,
Freebase	2千5百万实体, 2000主题, 1亿事实, 4000种属性
谷歌知识图谱	5亿实体名字, 35亿条事实
NELL	3百万实体名字, 300类别 500属性, 100万事实 1千5百万学习规则

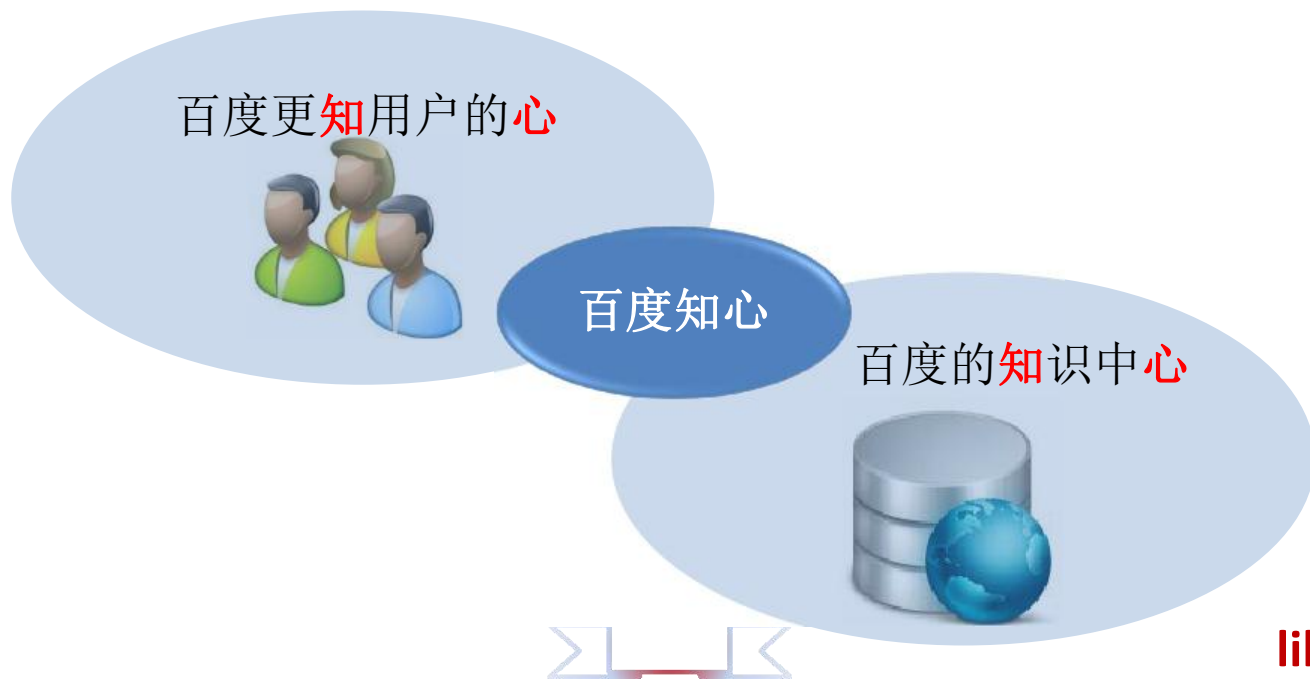
知识图谱助力智能情报



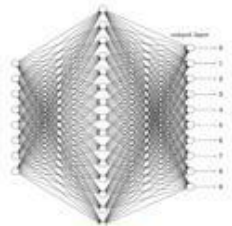
大脑处理各种各样的数据，涉及到智能语音处理和知识图谱，如知识推理、智能问答和可视化等

例， 百度知心： 百度的知识图谱

🏠 知心的两层含义：



人工智能的第二大技术驱动



深度学习（隐性模型）

- 面向某一个具体任务（如下围棋，识别猫，人脸识别，语音识别等）
- 需要海量训练数据
- 需要强大的计算力

知识图谱（显性模型）

- 可广泛用于不同任务
- 从海量数据中进行知识学习和挖掘
- 可理解、可解释，类似人类的思考方式



	深度学习	知识图谱
场景示例	 人脸识别	 语音助手
目前进展	在一些任务上已经接近或超过人类	在知识量上超过人类，在知识推理上不如人类
任务范围	面向具体任务，难以迁移	广泛适用于不同任务
可解释性	较难解释	可解释性强
数据量	海量训练数据	海量知识数据
未来趋势	未来深度融合	

知识图谱赋能

- ❏ 第一，采用智能爬虫、自然语言处理、众包、机器学习和行业词库等方式去处理数据源；
- ❏ 第二，利用 ETL、知识融合和知识存储去进行数据融合和存储；
- ❏ 第三，采用动态知识模型、实体知识图谱构建、知识推理引擎等方式去进行业务建模；
- ❏ 第四，用户体验方面，采用智能问答、语义检索、智能推荐和可视化分析这四种方式去触及用户。



其他的智慧情报应用场景

- 人工智能助教：jill watson
 - 在线课程，400个学生，1万个问题
 - 2016春季上岗，3个月下来，未被学生发现
 - 基于三年来4万多个问题-答案进行学习。
- 个人学术助理
 - 学者投稿智能引导
 - 自动情报服务、学术热点追踪
- 智能参考咨询
 - **自动简报**：《基于情报3.0工作思路的自动简报系统设计与实现》201802
 - 专题特色库的自动化智能化
 - 机器文献综述、自动文摘；
 - 自动编目，自动标引



4、智慧空间

交互化

 让你的图书馆更懂读者

在线化

 让你的图书馆永不打烊

数据化

 让你的图书馆生产数据；识别读者及其与空间的联系

知化

 让你的图书馆更聪明

Amazon Go=》 淘宝无人超市



无人超市：扫码购物、自动付款；2016年12月、2017年4月推出
无人餐厅：刷脸、手势识别、个性化推荐；2017年10月推出
永辉超市的智能购物车；2018年10月

交互化-视觉识别

阅面算法IP提供完整的算法矩阵

人脸检测

属性识别

人脸追踪

人体检测

人脸识别

人体追踪

在人脸上可以追踪106个点位



交互化-人脸识别

🏠 技术平台

👉 百度人脸识别+公安身份证数据

🏠 让各类系统与读者连接

👉 人脸闸机、空间预约签到

👉 图书自主借还、自助打印



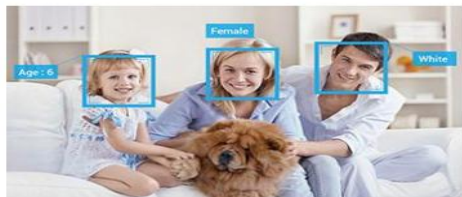
人脸检测与追踪
难度:★★★★



五官关键点检测
难度:★★★



人脸像素解析
难度:★★★★



表情、性别、年龄、种族分析
难度:★★★



活体检测与验证
难度:★★★★★



大规模人脸检索 (1 : N)
难度:★★★★★★★

百度人脸识别技术

应用案例



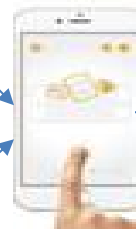
人脸图像

公安数据

身份证



二代身份证



前端人脸、证件采集 | 云端人脸比对

提升读者体验



- **基础服务**

- 进馆
- 自助打印复印
- 超期罚款

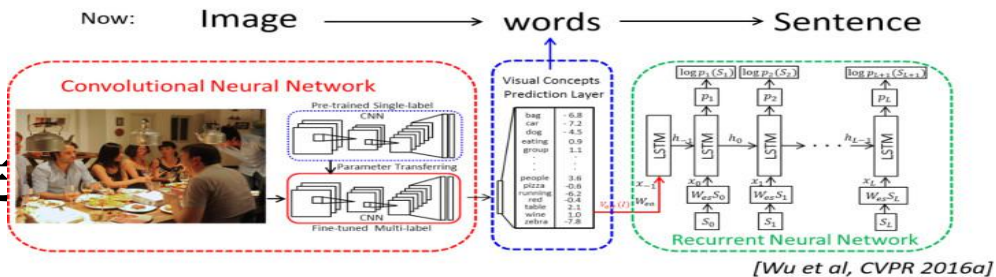
- **精准服务**

- 推荐服务
- 信用服务
- 占座问题
- 阅读报告

- **浙江理工大学图书馆**
lib.zjhu.edu.cn

从 Vision 到 Language 再到 Action

- 跨领域图像识别
- 视觉与语言的结合是一个非常好的方向，引出了 image captioning 和 VQA
- 学习理解多种模态的信息，能与真实环境进行一定程度的交互



Q: How many mammals in this image?
Common-Sense: dog and cat are mammals, bird is not.



Q: Why are they wearing such bright colour?
Common-Sense: bright colour can be used for safety reasons.



Q: Which object in this image is most related to entertainment?
A: TV. **R:** Television → Performing Arts → Entertainment



Q: List close relatives of the animal. **A:** Donkey, horse, mule, asinus, hinny, etc



Q: List common properties of the two images
A: Scene concepts: transport infrastructure;



Q: Which image is the most related to chef?
A: The left one. **R:** Found kitchen, oven, microwave.



Q: Which image is the most related to programmer?
A: The right one. **R:** Found computer and mouse.



Which furniture in this image can I lie on?
Pred. QF: (ClassFor, ObjSect, Image)
Keywords: "lie on"
Pred. SF: a sofa is usually to sit or lie on
Pred. Answer: sofa
GT Answer: sofa



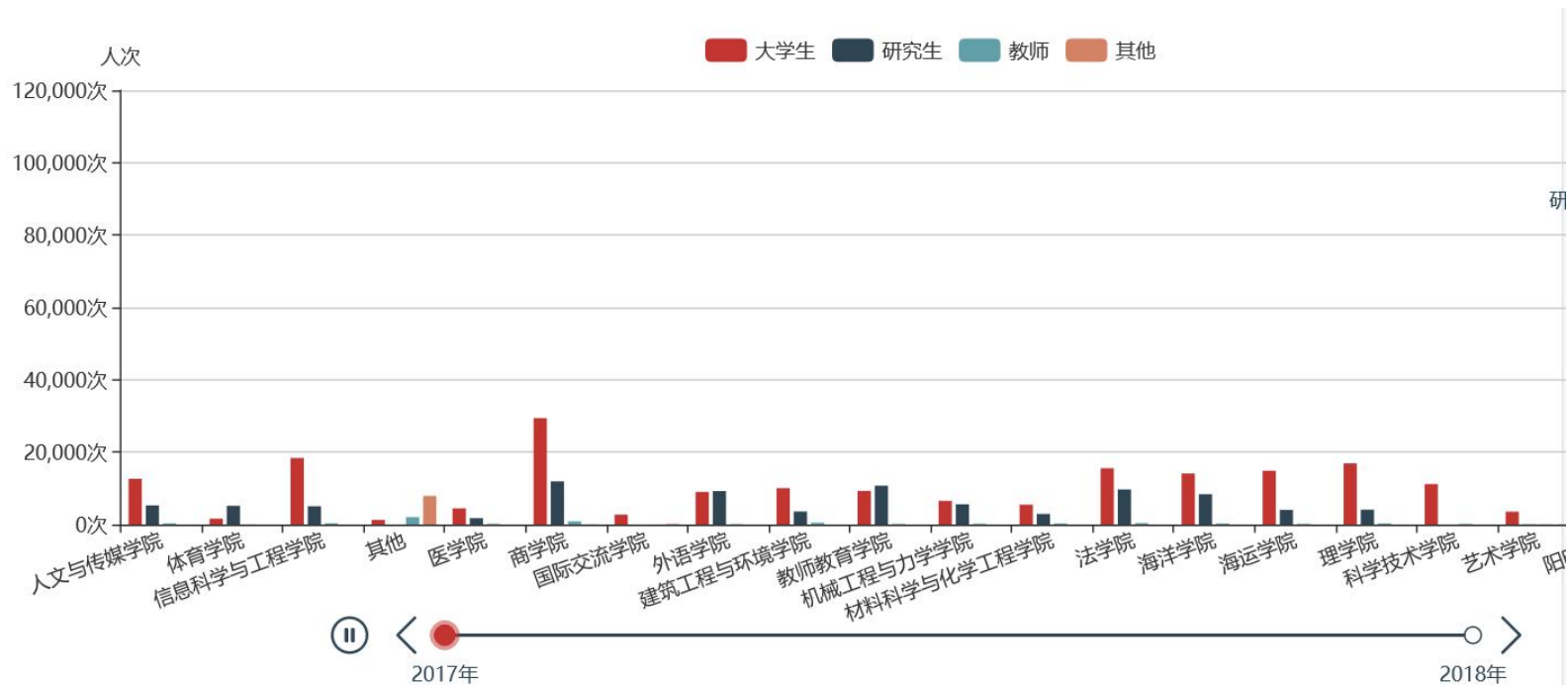
Which object in this image can I ride?
Pred. QF: (ClassFor, ObjSect, Image)
Keywords: "ride"
Pred. SF: motorcycle is used for riding
Pred. Answer: motorcycle
GT Answer: motorcycle



读者入馆数据感知



读者归属机构自动生成



当年进馆人次

3110654

今日进馆量

12429

电子资源点击总量

364882

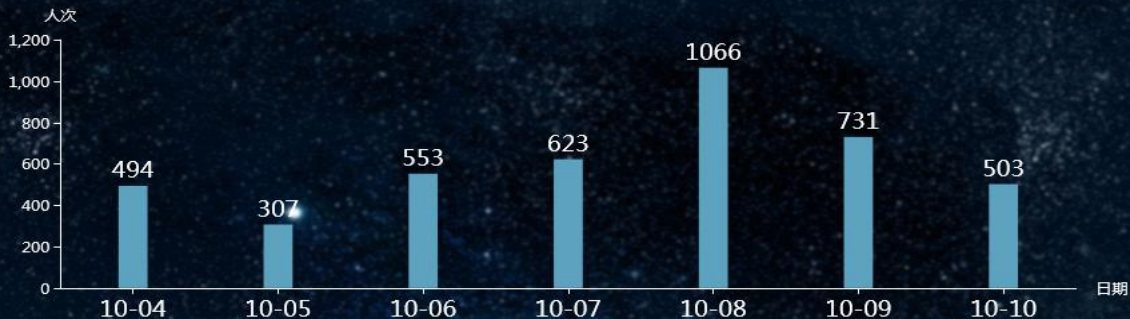
当年借阅书量

444581

图书馆藏书数量

1892946

近7日借阅量



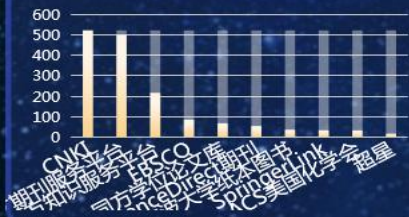
座位



研修间



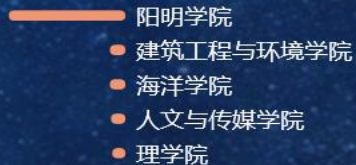
电子资源点击量



新书增量



昨日借阅学院排名



存包柜



交互化-语音识别

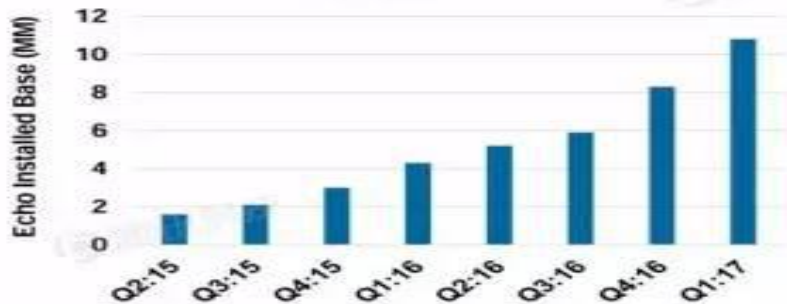
基于语音的家庭平台前端=
语音可以代替打字...

亚马逊ECHO从2014年11月至2017年5月的演变

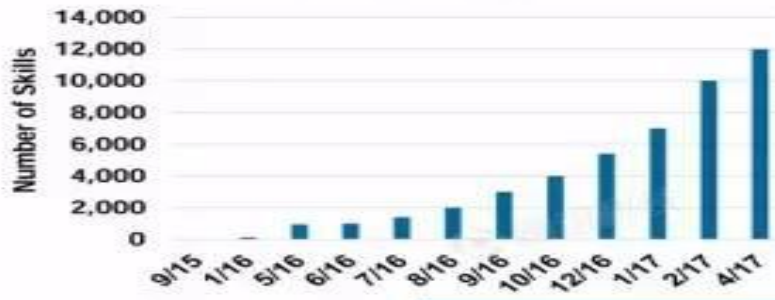


ECHO= 购物 + 媒体
ECHO LOOK = 购物 + 推荐
ECHO SHOW = 视频 + 语音电话

亚马逊ECHO设备在美国市场的
安装量



亚马逊ECHO第三方应用插件的
数量不断增加



智能音箱->智能空间

- 🏠 一进门, “ Alexa, turn on the light” , 灯就亮了, 因为已经把 Echo 和房里飞利浦 Hue 智能灯相连;
- 🏠 读者通过**语音与图书馆服务交互**
- 🏠 可以把房子 (研讨间) 变成**智能系统**: 语音输入两秒钟的事, 然后系统之间协同。



交互化-“屏幕”时代

🏠 目前已有50亿张数字屏幕：

🌐 每年还有38亿个新屏幕制造。

🏠 古代我们是“言语之民”

🏠 “书籍之民”

🏠 “屏幕之民”：正在进行。

🏠 AI与屏幕的结合

🏠 亚马逊今年5月推出的 Look 和 Show

🏠 Look 会启动全尺寸照片或视频拍摄

🏠 用机器学习算法和时尚专家意见给你参考

怎样才能快速进行智能化服务

第一，数据化往往是一个很昂贵的过程，
你是否有创造性找到一种方法能够**完成数据的初始化**？

第二，**机器学习逻辑是否贯穿你的整个业务过程**？针对一个足够大的问题，你能否找到新的算法，挖掘数据背后的洞察？

第三，互联网思维经常讲**上线、迭代、优化**。核心原因是，不上线就没有用户反馈，就不知道往哪个方向优化。

精准服务要建立在和用户的持续性互动关系之上，在这种**持续性互动**中，对产品（服务）进行迭代和优化，从而更加精准。能够这样做的产品是一种**“活”的产品**。而要建立一个活的产品，就必须**以数据智能作为产品的核心**。

数据化-宁大馆重新规划所有业务

🏠 推进线上线下有机融合OMO

📖 资源：纸质与数字融合；方便用户贡献UCG

📖 空间：馆舍、座位、资源等可远程实时查询

🏠 实时感知到读者，个性化推荐资源

🏠 服务：Reading&Research

📖 阅读服务：阅读单-》阅读圈-》学习圈

📖 学科服务：精细化、个性化、场景化

亚马逊：从线上到线下的融合



在线化-业务重构

🏠 大部制重组

🎨 技术总部+业务再造

🏠 服务总部，8个学科团队：70%馆员

🎨 与空间管理、纸质书刊管理结合

🏠 相关学科的资源调研与筹购

🎨 学科服务+阅读推广:每个团队对应2-3个学院

🏠 技术支撑：线下向线上融合

🏠 块的工作：虚拟团队来支撑（查新、培训）

知化-懂读者

场景（借书）：

- 1、进入藏书区域，系统根据历史使用情况，对新书进行个性化推送介绍；
- 2、书架驻留时，根据历史浏览、借书记录，告诉该用户历史书架浏览借书情况，提供本班级或学院同学借阅情况；
- 3、该书架已被借出的图书有哪些，对感兴趣的图书，可直接提交转借申请。

场景（阅览）：

- 1、进入阅览区域，利用室内定位与座位管理系统等，提醒历史座位使用情况及正在使用情况；
- 2、如果该阅览区域无空座的情况下，推荐空闲座位（分安静区与自由区）；
- 3、共同圈子中，其他成员阅览分布情况。

李磊
2016



一层大厅

图书馆信息	
当前时间:	2015年11月23日
在馆人数:	128 + 1
馆内温度:	22°C
馆内湿度:	40%
座位剩余:	428/1712
新入馆读者信息	
持卡姓名:	李磊
在借图书:	6册
在约空间:	二楼研讨室03室

当前使用率: 25%

刷一卡通进入

一门式管理

信息显示屏

the entrance guard screen

门禁显示屏实时更新图书馆信息，通过一卡通刷卡进馆统计进馆人数，刷卡同时显示持卡人预约信息。

5、精准情报

- 📖 精准医学=》精准情报
- 📖 理解读者：个性化、精准化
 - 👤 在掌握使用数据的基础上
 - 👤 精准服务;个性化服务
- 📖 理解内容：数据化、语义化
 - 👤 从元数据级到全文级
 - 👤 自动分类;自动摘要生成
 - 👤 学术标签生成

精准为什么是今天

“精”是精细

互联网不仅能做到个性化到每一个人，还能个性化到这个人在某个时间点的状态、情绪、行为。

“准”是越来越准确，而这只有通过智能化才能实现。这个时代最大的特点是开始接受不确定性，尝试用统计的方法来逼近准确性。

(Google翻译的案例，与传统工业设计的

114



从人找资源=》为资源找人

📖 图书推荐

- 👤 基于借阅历史、学科、内容、社会关系、热门搜索
- 👤 线下阅览习惯、其他书（网）店的热门书

📖 空间资源：座位推荐

- 👤 基于个人习惯
- 👤 馆藏与学科背景相似度、活动与个人兴趣匹配

📖 学术资源：学术头条

- 👤 从搜索引擎——向推荐引擎转变

实现读者业务运营



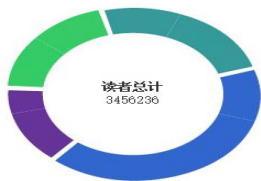
整合读者在实体馆和数字馆的使用日志。



基于读者行为数据的智慧推荐，每个读者看到的服务内容都有差异。



读者分类统计



学生: 25664

教师: 25664

校友: 25664

馆员: 25664



1天活跃用户数
26,313
占总数的百分比: 100.00% (26,313)

7天活跃用户数
158,052
占总数的百分比: 100.00% (158,052)

14天活跃用户数
306,773
占总数的百分比: 100.00% (306,773)

30天活跃用户数
710,831
占总数的百分比: 100.00% (710,831)

阅读推广数据化、互联网化

微书、实体书店、借阅配送



移动应用：智慧图书馆APP



智慧图书馆功能



借还书校内物流配送



借书

搜书

预定
书本

配送
信息

等书
上门

还书

提交
申请

配送
信息

上门
取书

还书
入库

读者操作

配送员操作

学术服务个性化

学术
资源

- 学术交流的工具不断变革和创新
- 被动服务转向主动服务
- 大众服务转向个性化服务
- 传统文献库转向开放型学术资源

学术
头条

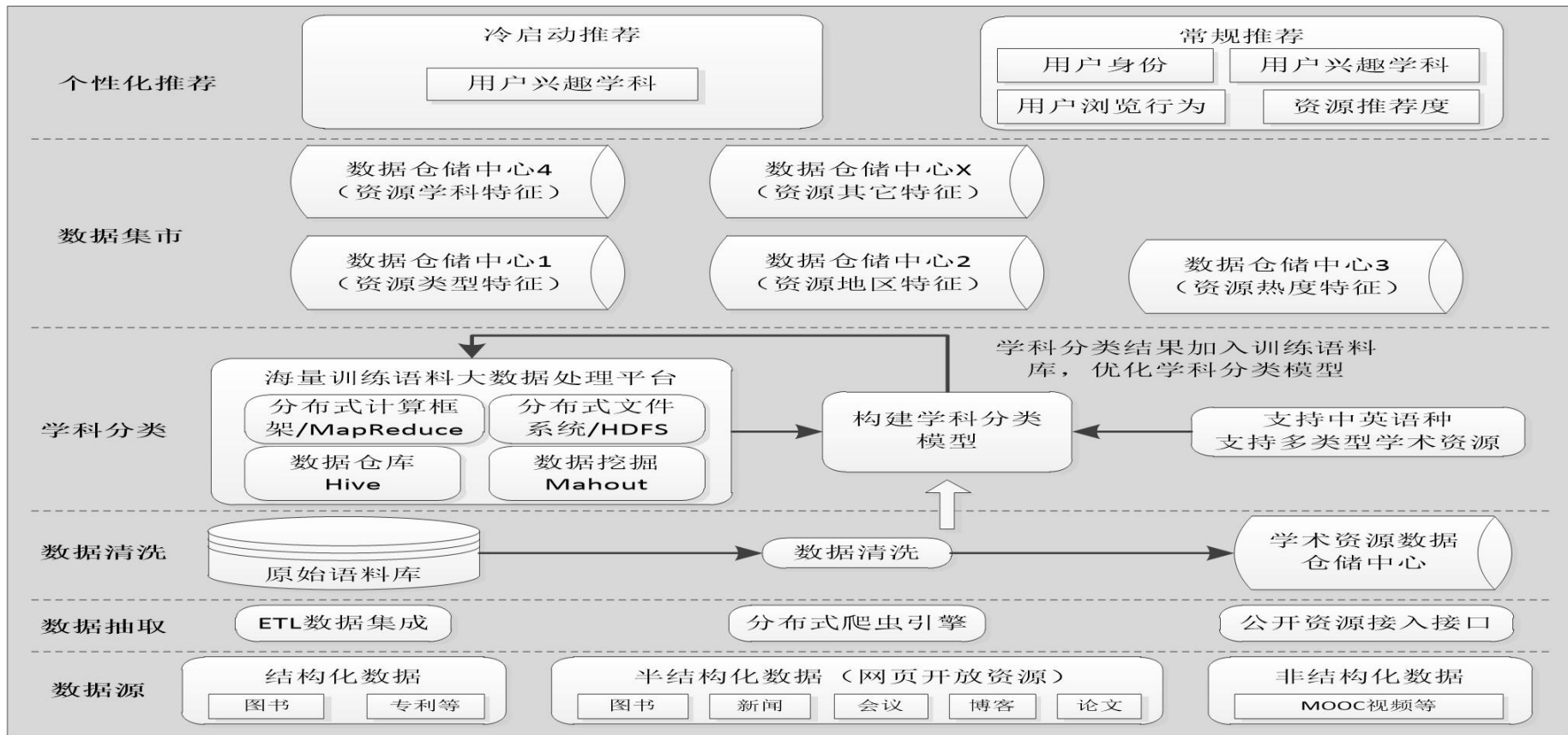
AI产品化：让算法利用足够的数据（学术资源），使得产品（学术头条）运行起来，然后通过产品来获取用户，用户在提供更多的数据……周而复始。

学术头条



- 以用户为中心,让资源去找人
- 海量学术资源的动态推荐
- 基于大数据的深度学习,对用户信息、行为动态等方面进行捕捉,多维度制作精准的用户画像,进而形成个性化、智能化、精准化的推荐
- 资源类型包括:期刊、图书、论文、会议、专利、新闻、博客等
- 人工智能:自动收割、学科分类、标签生成

学术头条框架

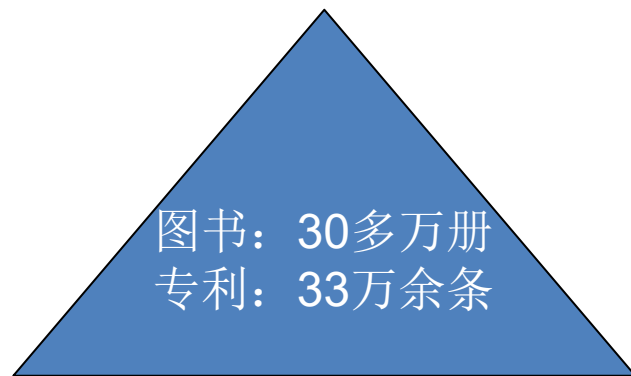


数据更新自动化

学术论文	年份	数量（篇）
	2013	268.682万
	2014	278.8001万
	2015	245.3132万
	2016	235.6011万
	2017	129.4645万

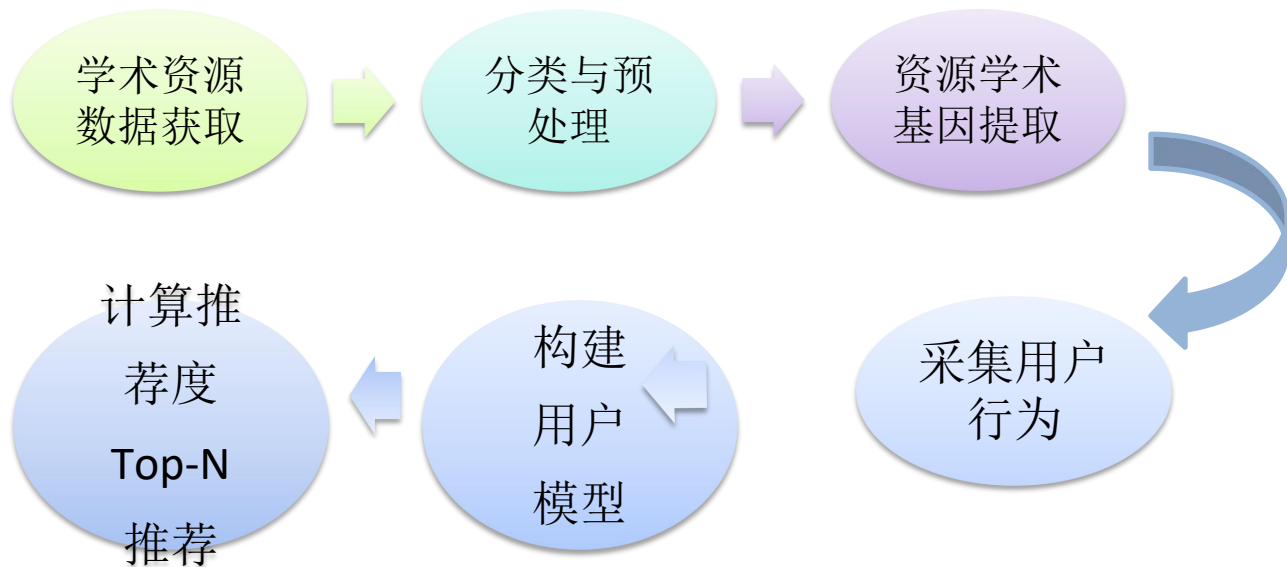
学术会议	网站源	数量
	Ourglocal、WikiFPC、丁香园、中国化学会、中国学术会议在线、中国教育网、医学全在线 医学生物会展网、学术会议网 新浪财经一会展、新浪财经网 科学网、艾会网.....	两万余条 实时抓取，日更新量50 余条

学术新闻	网站源	数量
	科学网 http://www.sciencenet.cn/	10万余条 实时抓取，日更新量 1000余条
	中新网 http://www.chinanews.com/	
	Nature中文 http://www.natureasia.com/zh-cn/	
	生物帮 http://www.bio1000.com/news/	
	中国社会科学网 http://www.cssn.cn/	
	中国科技网 http://www.stdaily.com/	
.....		

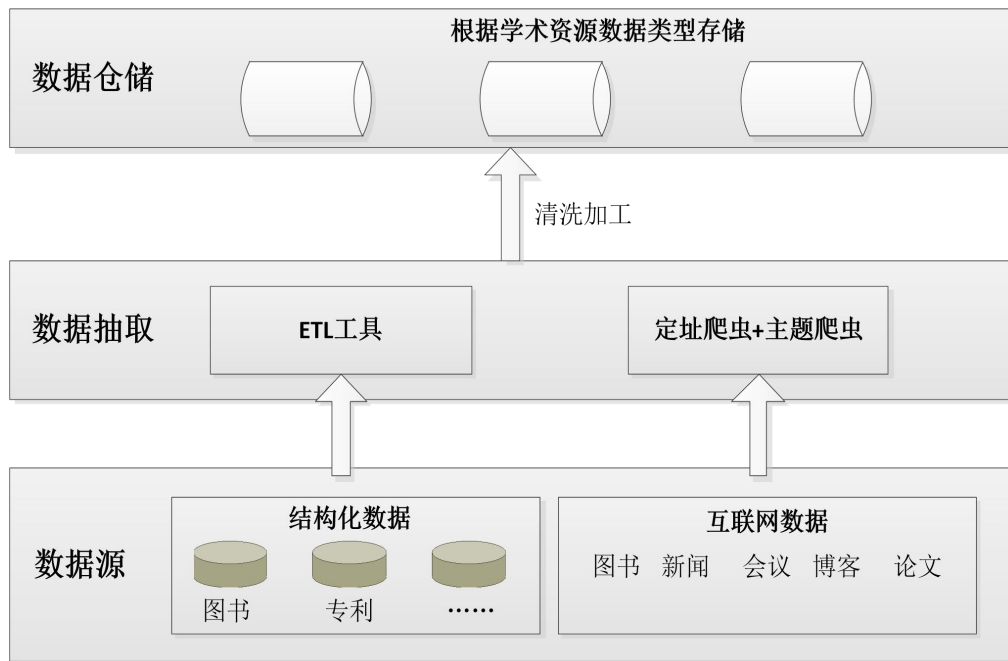


图书：30多万册
专利：33万余条

学术头条流程



学术资源自动收割



- 🏠 资源**自动**收割
- 🏠 整合各种互联网资源
- 🏠 **主动推荐**多类型资源：
：论文、图书、会议、
、专利、新闻、博客
- 🏠 资源类：20亿条
- 🏠 用户数据：每天100亿条



推荐模型动态更新



- 阅读
- 收藏
- 搜索
- 不感兴趣
- 点赞

实时更新，实时影响推荐资源顺序



- 阅读
- 收藏
- 搜索
- 不感兴趣
- 点赞

两分钟更新所有资源的全局热度

学术基因提取

- 🏠 学术资源的多标签
- 🏠 学术特征库的建立数十个特征, n个标签
 - 👤 关键词: 大数据、深度学习、超级计算机
 - 👤 题材分类, 分类虚构故事的体裁
- 方法: 决策树、SVM、神经网络等监督学习模型, 用来预测已标记数据

资源的个性化动态推荐

资源

- 资源类型
- 关键词分布
- LDA主题分布
- 学科分布

用户

- 用户身份
- 浏览、阅读时长
- 评分、分享、收藏

刚开始

- 用户喜欢xxx等学科的资源

有部分
用户数
据时

- 用户喜欢xxx等学科的资源
- 用户喜欢xxx主题方面的资源
- 用户喜欢xxx关键词方面的资源

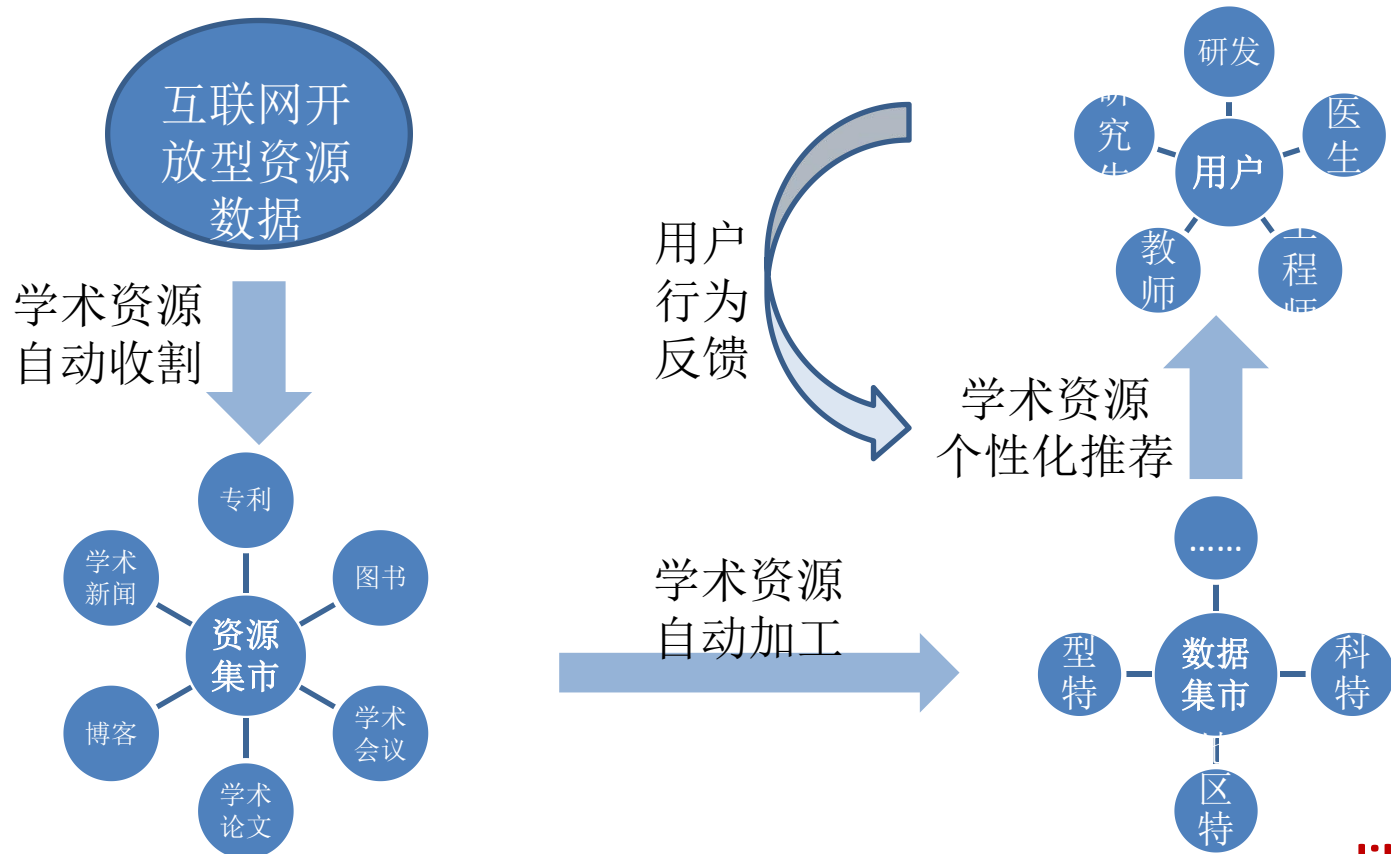
有大量用
户数据时

- 推荐与该用户兴趣相投、拥有共同经验用户喜欢的学术资源

基于学术资源内容的个性化推荐策略

基于协同过滤+矩阵分解模型+上下文信息

构建学术资源的数据闭环



三、趋势与展望

人工智能的三个阶段

📖 弱智能（ANI）：感知智能

🦉 从基本的计算智能进化而来：利用大量算力

🦉 认知智能（智能1.5）

🏠 能够理解、思考、交流：语义解析、对话交互

📖 强智能（AGI）：人类级别的智能

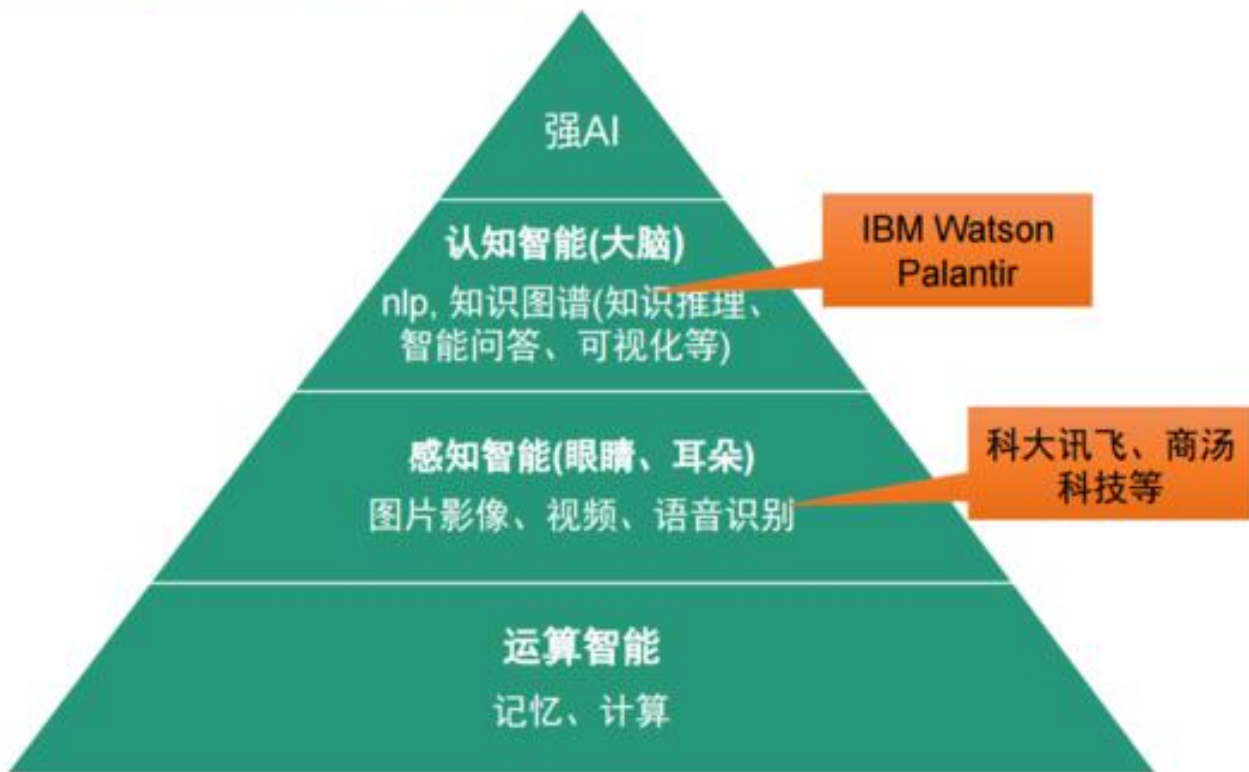
🦉 让机器有自主智能是最终理想：强大、通用

📖 超人工智能（ASI）



人工智能的技术金字塔

AI 研习社 · 知识地图




人工智能在互联网时代的四大浪潮

- 📖 第一阶段在互联网阶段：广泛应用搜索购物
- 📖 第二阶段AI赋能商业化：AI应用在商业领域
- 📖 第三阶段实体世界的智能化：
 - 👤 对用户行为全方位的理解和数据化，OMO线上线下融合，线下的场景充分数据化，通过智能化进行赋能。
- 📖 第四阶段是全面的智能化：如智能驾驶。



AI+ 还是 +AI?

AI+行业：从0到1

 在AI技术成熟之前，这个行业、产品从未存在过。比如自动驾驶，亚马逊的Echo智能音箱、苹果的Siri语音助手。在人工智能技术未突破前，不存在这样的产品。因为AI，创造出了一条全新的产业链。

AI+ 还是 +AI? (续)

🏠 行业+AI: 从1到N

- 👉 在行业本身一直存在，产业链条成熟；以前完全靠人工，效率比较低，现在加入AI元素后，使得行业效率有了明显提高。比如安防、医疗等领域

🏠 涟漪效应

- 👉 当一个AI应用找到第一批用户时，他们使用的行为和记录被后台记录下来；再对这种行为和记录进行迭代的改进，当再把该应用投向第二批用户的时候，其行为已经比第一代提升了

图书馆行业的互联网思维

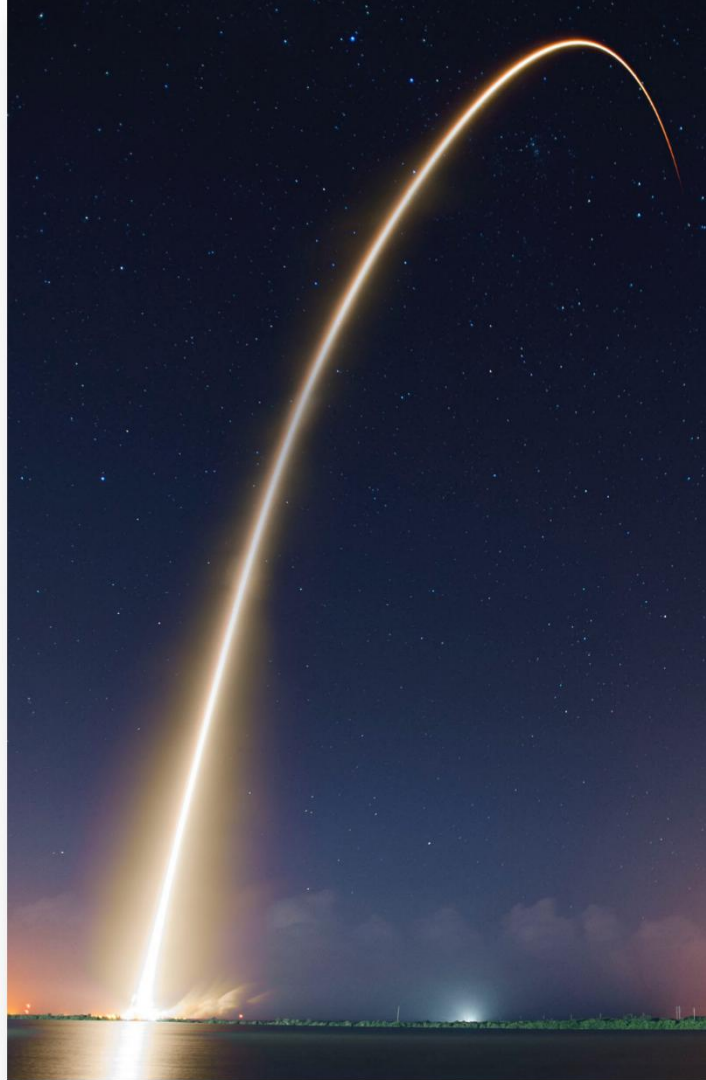
- 📖 通过系统平台建设，整体提高了我们高校图书馆的服务水准；
- 📖 跟国外同行相比，差距越来越小，有的方面甚至超越；
- 📖 通过共同描绘一种AI愿景，与业界形成合力，不断迭代，不断完善，不断推进。



知识经济的时代最核心的资源，一个是数据，一个是知识工人（知识工作者）。资本“Capital”，在未来再也不会占据那么大的重要性。

大数据只是知识经济落地最实在的切入口，当云计算有了大数据、有了互联网这样的基础后，这三者是三位一体的。

一切才刚刚开始，未来还有太多太多的可能性。如果不是100倍空间的话，至少几十倍的空间是能看到的。



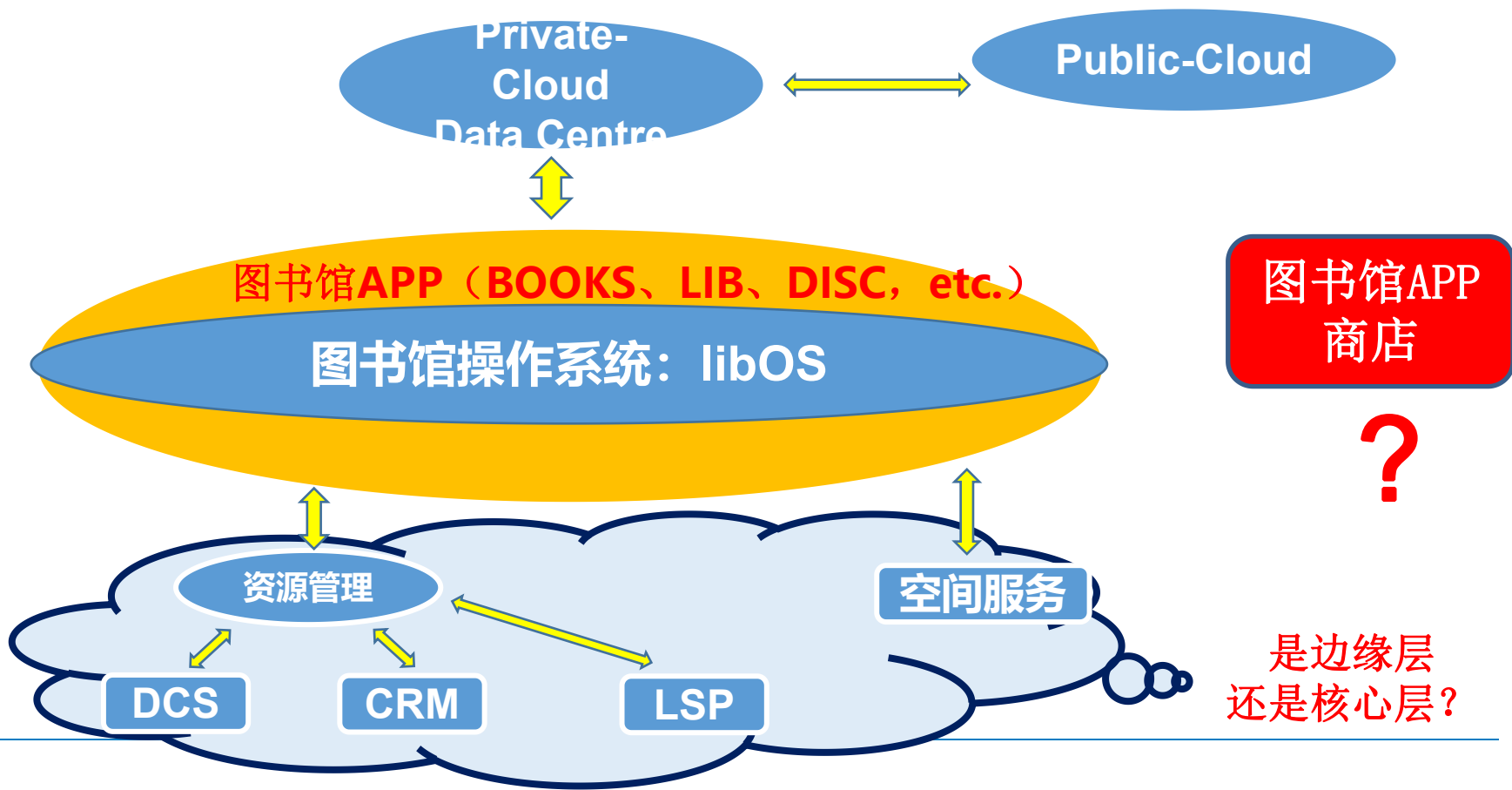


无 智 能 不 商 业

未来的世界将是一个万物互联、人工智能无所不在的世界，一切商业智能化是未来最重要的一个趋势

什么叫智能化？核心是未来商业的决策会越来越多地依赖于机器学习、依赖于人工智能，机器在很多商业决策上将扮演非常重要的角色，它能取得的效果超过今天人工运作带来的效果。

智能图书馆的II：情报互联网



智能图书馆II：打通三条主线



1、**服务**管控一体化集成与优化

SERVICE

价值链服务：空间优化 - 读者画像 - 资源推荐 -
服务交付

感知智能



2、**资源**协同一体化生态

RESOURCE

数字化生态圈：读者需求 - 采访 - 编目 - 本馆 -
区域- 全国-全球

认知智能



3、**技术**及设备全生命周期管理

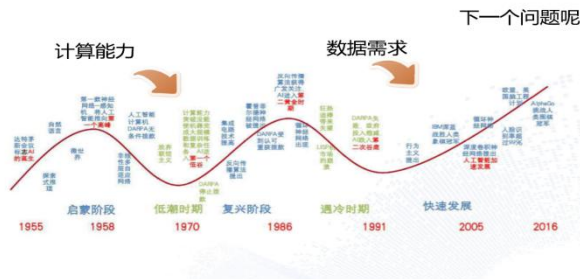
TECHNOLOGY

项目 - 设计 - 建造交付 - 状态监控-运行维护 - 报废退出

运算智能

AI展望

昨天



历史是螺旋进步的，人工智能走到今天并非一帆风顺，经历了三个浪潮。

141

今天

语音识别



图像识别



自然语言处理



伴随大数据时代的到来，计算能力的大幅提升，人工智能技术在越来越多的领域取得了突破或者长足的进步。

141

明天

- 1、逻辑 (logical)
- 2、语言文字 (linguistic)
- 3、空间 (spatial)
- 4、音乐 (musical)
- 5、肢体动作 (kinesthetic)
- 6、内省 (intra-personal)
- 7、人际 (inter-personal)
- 8、自然探索 (naturalist)
- 9、图形图像 (Graphics)

AI离人类智能还有多远？对比哈佛大学心理学家加德纳的多元智能理论，在空间、音乐和肢体运动方面有差距，在内省、人际和自然探索方面尚无可比性。

如何评价目前人工智能取得的成果

🏠 现在的人工智能是**没有理解**的人工智能

👤 须满足五个条件：一个是数据，一个是知识；
确定性信息、完全信息；静态的、单任务；有
限领域

🏠 我们现在离真正的人工智能还有一段很长的路

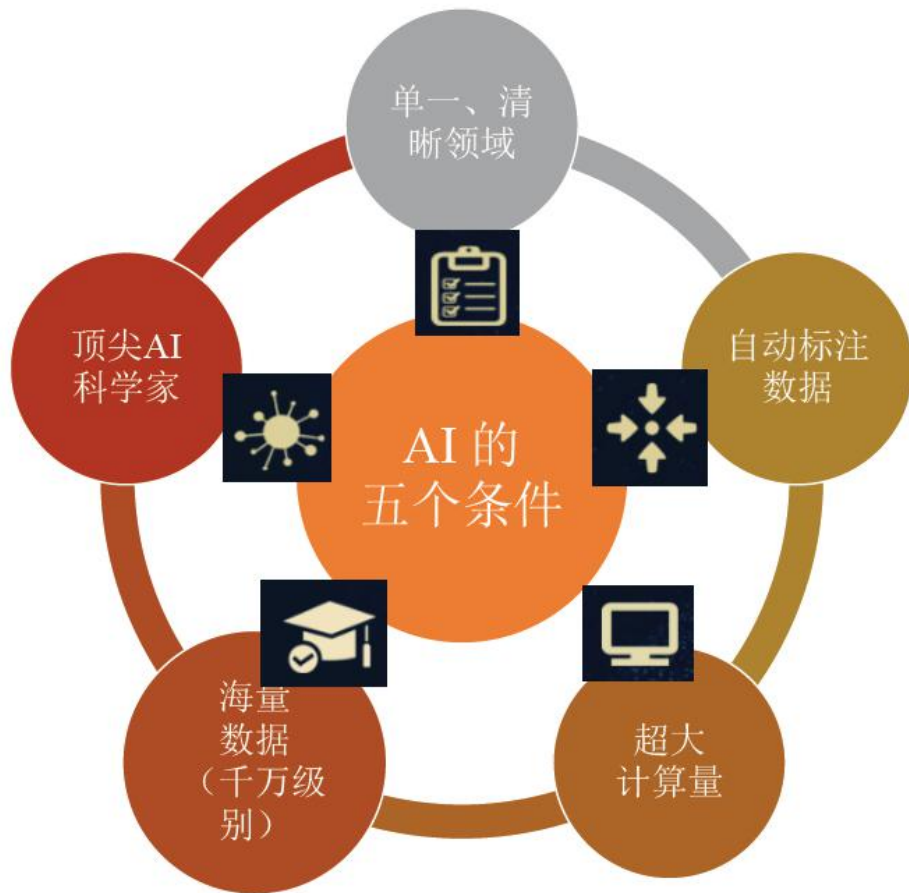


满足四个条件机器才能超过人类

受四个条件限制：

- 📖 头一个条件当然是需要有大量的数据，
- 📖 第二个是完全信息，
- 📖 第三个是确定性，
- 📖 第四个是单领域和单任务。

什么领域AI能够超过人



- 充分的**数据**：需要有超级大的数据量，它不能举一反三，它要举多反三
- 信息具有**确定性**：数据是要有标注的，不是到网上弄一堆数据
- （完全信息博弈）**完全信息**：应用一定是单一领域，这个领域越垂直、越细越好。一个人工智能客服的应用不可以用来做导游，应用单一、一定要非常清晰不跨领域
- 有懂行的专家：去选择有效的神经网络模型，调整各种网络参数
- 需要超大计算量，机器容易甩开人类

最近新提出的图灵测试标准

🏠 威诺格拉德模式挑战

👉 测试「常识」，AI、事物和文化规范是如何互相影响的

🏠 人类的标准化测试

👉 AI将接受人类在小学中学阶段面临的考试，不给任何宽限

🏠 物理图灵测试

👉 要求被测试的机器理解任务内容，找到解决方法

🏠 I-Athlon

👉 AI总结音频中的内容叙述视频中发生的情节，即时翻译执行其他任务

指数时代-from 《爆裂》

- ▣ **技术的变化速度**会超过人类的适应速度。
- ▣ 指数时代有三个特点：一是不对称性；二是复杂性；三是不确定性。
- ▣ 破局的九个原则：涌现优于权威；拉力优于推力；**指南针优于地图**；风险优于安全；违抗优于服从；**实践优于理论**；多样性优于能力；韧性优于力量；系统优于个体。



图书馆智能的进程表

🏠 感知图书馆：弱智能-2020年

📖 让图书馆能更懂读者；

📖 利用大数据+AI听说读看能力。

🏠 认知图书馆：中智能-2025年

📖 图书馆能够**理解**读者、**思考**交流

📖 全新知识服务：深度语义解析、对话式交互

🏠 自主图书馆：强智能-2030年

📖 有一定自主智能、**行动**能力；灵活、强大而通用



关键在于构建智慧图书馆体系

- 🏠 图书馆服务智能化
- 🏠 资源融合与个性化
- 🏠 管理全面感知化
- 🏠 连接无缝化
- 🏠 用户体验智慧化
- 🏠 智慧图书馆的评价体系



数据决定智慧

Be more

体验更多

图书馆要 保存每一条数据

总结：走向智慧图书馆

📖 大数据：大量海量实时更新的数据

- 📖 本地中央库：全方位收集各类数据
- 📖 网上云平台：提升对师生的分析和服务能力
- 📖 本地化APP：有效地将每一服务、每一份资源推送给用户

📖 借力：三个大趋势的汇聚

- 📖 **海量数据**（Big Data）《= 互联网、物联网
- 📖 线上线下融合**OMO**：引进各种技术积木，为本馆所用

未来已来

- 🏰 金融、商业的城堡由数据的砖块堆砌而成，随处可见人工智能的缝隙
- 🏰 图书馆的智慧之光正在闪现：
 - 📖 大势所趋，积木式创新；
 - 📖 数据红利+AI让师生感受到被服务；
 - 📖 图书馆：联盟+赋能，打造爆款服务
- 🏰 即将进入 **“智能图书馆”** **新范式** AI Iniside

谢谢！

请批评指正！

lbs@nbu.edu.cn

