电子商务导论

朱桂祥
南京财经大学信息工程学院
江苏省电子商务重点实验室
电子商务信息处理国家级国际联合研究中心
电子商务交易技术国家地方联合工程实验室

范围:

C1:电子商务大数据分析导论
C2:数据采集与预处理
C3:轨迹大数据挖掘
C4:电子商务欺诈与反欺诈
C5:推荐系统
C6:案例分析

教材推荐:

[1]电子商务数据分析 大数据营销 数据化运营 流量转化，杨伟强 著，人民邮电出版社
[2]大数据管理与应用导论，曹杰、李树青 著，科学出版社
[3]机器学习，周志华 著，清华大学出版社
[4]商务数据分析与应用，胡华江，杨甜甜 著，电子工业出版社

推荐的论文:

[1] Guo Q, Zhuang F, Qin C, et al. A survey on knowledge graph-based recommender systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2020.
[2] Wu S, Zhang W, Sun F, et al. Graph neural networks in recommender systems: A survey[J]. arXiv preprint arXiv:2011.02260, 2020.
[3] Zhang S, Yao L, Sun A, et al. Deep learning based recommender system: A survey and new perspectives[J]. ACM Computing Surveys (CSUR), 2019, 52(1): 1-38.
[4]伍之昂, 王有权, 曹杰. 推荐系统托攻击模型与检测技术[J]. 科学通报, 2014, 59(7): 551-560.
[5]黄立威, 江碧涛, 吕守业, 等. 基于深度学习的推荐系统研究综述[J]. 计算机学报, 2018, 41(7): 1619-1647.

评分规则
点名5%
提问5%
作业和实验(5% 6次 30%)
课程报告60%

相似度检测超过30%计0分。

视频热身话题讨论：

虚拟现实VR(Virtual Reality)
增强现实AR(Argumented Reality)
混合现实MR (Mixed Reality) 和电子商务的关系

虚拟现实购物不再是科幻电影的桥段。

“BUY+ The First Complete #VR Shopping Experience. Watch how everything, from perusal to purchase, takes place inside a VR environment. pic.twitter.com/JvYfOYokqF

— Alibaba Group (@AlibabaGroup) October 20, 2016”

视频短片观看网站
:https://haokan.baidu.com/v?vid=3427168685531592589&pd=bjh&fr=bjhauthor&type=video

Big Data Analytics and E-commerce

虚假评论评论检测 (第4章)

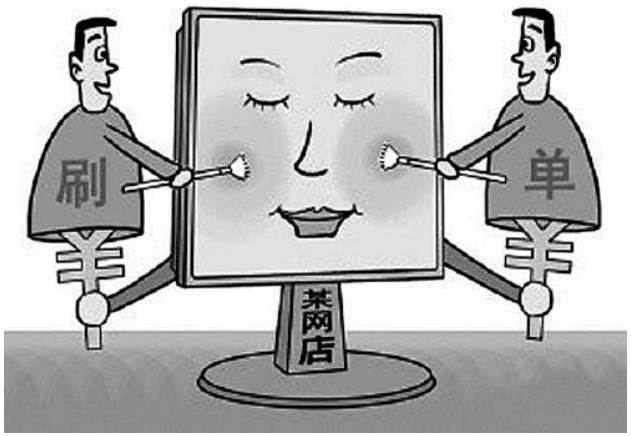
恶意用户：大多具有“操榜”行为
电子商务平台：劫持用户情感、操控榜单等
社交平台：传播广告、舆论等特定消息
恶意用户日渐猖獗，引起学术界和工业界的重视
2001年，Sony Pictures承认雇佣枪手来推广其电影
Yelp和TripAdvisor虚假评论超过6%
第三方水军雇佣平台大量涌现：“刷客网”

https://www.forbes.com/sites/emmawoollacott/2017/09/09/exclusive-amazons-fake-review-problem-is-now-worse-than-ever/#1b54b4f07c0f

电子商务大数据分析

虚假评论检测问题 (第4章)

讨论
(1) 虚假评论的原因是什么？
(2) 怎样使用现有的技术手段解决该问题？
(3) 开放性的其他问题？

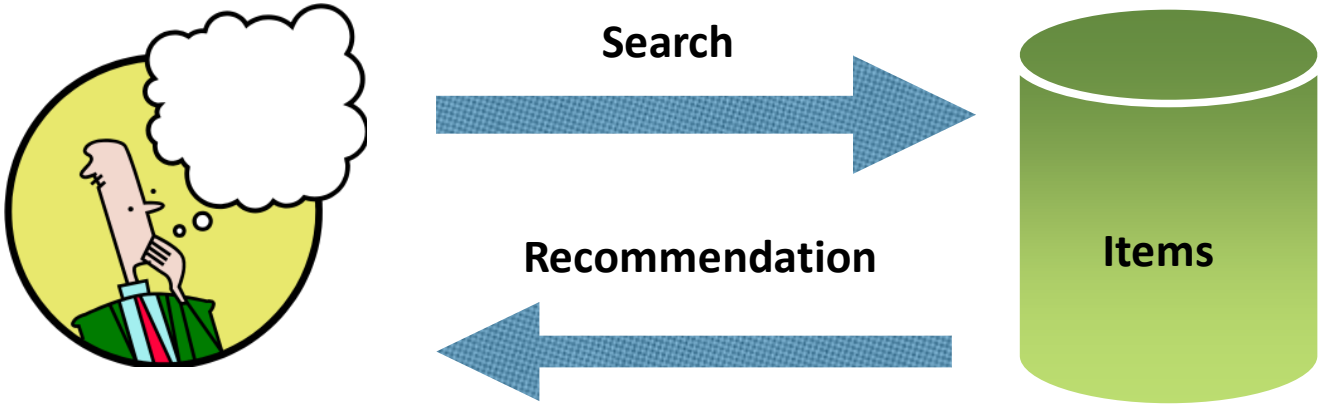


[1]https://www.forbes.com/sites/emmawoollacott/2017/09/09/exclusive-amazons-fake-review-problem-is-now-worse-than-ever/#1b54b4f07c0f
[2] http://tech.sina.com.cn/i/2015-10-19/doc-ifxivsce6919323.shtml

电子商务大数据
(第5章 推荐系统)

问题：Web 2.0呈现出”信息过载“，仅依靠搜索引擎很难满足用户的个性化需求，转被动拉取为主动推送

个性化推荐模型与系统：建立在海量数据挖掘基础上的一种高级商务智能平台，以帮助电子商务网站为其顾客购物提供个性化的决策支持和信息推荐服务



电子商务大数据
(第5章 推荐系统)

协同过滤简单来说是利用某兴趣相投、拥有共同经验之群体的喜好来推荐用户感兴趣的信息，个人通过合作的机制给予信息相当程度的回应（如评分）并记录下来以达到过滤的目的进而帮助别人筛选信息。

电子商务大数据

(第5章 推荐系统)

最早被使用的推荐算法，它的思想非常简单：根据用户过去喜欢的物品（内容），为用户推荐和他过去喜欢的物品相似的物品。而关键就在于这里的物品相似性的度量，这才是算法运用过程中的核心。基于内容的推荐最早主要是应用在信息检索系统当中，所以很多信息检索及信息过滤里的方法都能用于其中。

什么是电子商务?

维基百科的定义

电子商务，简称电商，是指在互联网或电子交易方式进行交易活动和相关服务活动，是传统商业活动各环节的电子化、网络化。电子商务包括电子货币交换、供应链管理、电子交易市场、网络购物、网络营销、在线事务处理、电子数据交换（**EDI**）、存货管理和自动数据收集系统。在此过程中，利用到的信息技术包括互联网、万维网、电子邮件、数据库、电子目录和移动电话。

（***E-commerce is the activity of buying or selling of products on online services or over the Internet. Electronic commerce draws on technologies such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems.***

）

https://en.wikipedia.org/wiki/E-commerce

电子商务发展的简要历史
(1) 1979: Michael Aldrich 建立第一家在线商店
(2) 1995: Jeff Bezos 创建了 *Amazon.com*.
eBay 由程序员Pierre Omidyar 创立。
(3)1999: Alibaba集团在中国成立.。
(4) 2001: Alibaba.com 在2001年12月盈利。
(5) 2002: eBay 收购 PayPal 价值15亿美金.。
(6) 2003: Amazon.com 发布第一个盈利年报。
(7) 2015: Amazon.com 占据了美国电子商务市场增长的半壁江山。

问题?

阿里和Amazon的主要区别？

Thanks Giving Day
(感恩节, 十一月最后一个周四)
Black Friday
(黑色星期五)
Cyber Monday
(剁手星期一)
V.S. 双十一
11, 26, 2018 V.S. 11, 11, 2018

电子商务的不同类型

B2B (business-to-business)
B2C (business-to-consumer)
B2B2C (business-to-business-to-consumer)

C2B (consumer-to-business)
C2C (consumer-to-consumer)
e-government
O2O (online-to-offline)
……

问题：下面的商家属于哪种类型

JD.com （京东）
https://2.taobao.com/ (闲鱼)
https://www.ele.me/home/ (饿了么)

从电子商务到社交商务

案例分析 索尼公司(sony.com)

索尼公司作为消费电子公司，和 samsung(三星)，sharpusa (夏普), LG等公司存在激励的竞争关系。
公司的利润从 2009年到2012年快速下滑。

为了应对业绩压力，索尼公司引入了社会化电商运作。

解决方案
社交网络和社区运行。

YouTube: 分析培训视频。
Linkedin: 吸引新的消费者。

结果：
消费者的信任上升。

消费者的营销为公司带来了快速的增长。
35,000,000 粉丝在 Facebook

从电子商务到社交商务(social commerce[1])

社交商务可以定义为把社交网络的口碑效应应用于电子商务[2]。
应用社交媒体技术重塑商业，把现有的商品和服务市场转化为以社交网络为中心，用户驱动的市场[3]。

Facebook上星巴克社交商务的例子[1].
个性化的页面。
有组织的定期活动。
丰富的社交信息，包括视频和图片。
信息的分享行为。
把信息链接到社团内部的人员。

传统的电子商务仅仅是在线上进行订单和支付。

[1]Huang Z, Benyoucef M. From e-commerce to social commerce: A close look at design features[J]. Electronic Commerce Research and Applications, 2013, 12(4): 246-259.
[2]Dennison, G., Bourdage-Braun, S. and Chetuparambil, M. Social commerce defined. White paper #23747, IBM Corporation, Research Triangle Park, NC, November 2009.
[3] Wigand, R. T., Benjamin, R. I., and Birkland, J. Web 2.0 and beyond: implications for electronic commerce. In Proceedings of the 10th International Conference on Electronic Commerce, Innsbruck, Austria, August 2008, ACM Press, New York, NY, 2008.

从电子商务到虚拟世界的商务

电子商务的新兴技术

(1)电子钱包 (E-wallet)
(2)智能客服(Instant Messaging Intelligence)
(3) 推荐系统(Recommender Systems)
(4) 虚拟现实和增强现实(VR, AR)

智能客服(Instant Messaging Intelligence)

智能客服背后的技术：
语音识别，自然语言处理，自然语言理解，知识图谱，语音合成等人工智能相关技术融合。

[1]https://sv.baidu.com/videoui/page/videoland?pd=bjh&context={%22nid%22:%2236570091915669160 67%22,%22sourceFrom%22:%22bjh%22}&fr=bjhauthor&type=video

电子商务中的数据：

(1)结构化数据：使用关系型数据库如MySQL、Oracle、SQL Server等来表示和存储成二维形式的数据。

学 号 Sno	课程号 Cno	成绩 Grade
200215121	1	92
200215121	2	85
200215121	3	88
200215122	2	90
200215122	3	80

结构化数据的优点：
存储规范，便于查询，易于修改。

结构化数据的缺点：
对于不满足规则的数据，不容易存储。

电子商务中的数据：
(2)非结构化数据：没有固定格式的数据。

比如word文档，pdf文档，语音文件mp3,视频文件,图片文件jpg。

非结构化数据的优点：
一般用二进制进行存储。

非结构化数据的缺点：
一般整体存储，特征并不明显。

电子商务中的数据：

(3)半结构化数据：通过(key,value)调整相关的信息，而且value的类型不固定。可以是文本，列表等。
比如xml，json等。

半结构化数据的优点：
存储比较灵活。

半结构化数据的数据库是节点的集合，每个节点都是一个叶子节点或者一个内部节点。叶子节点与数据相关，数据的类型可以是任意原子类型，如数字和字符串。每个内部节点至少有一条外向的弧。每条弧都有一个标签，该标签指明弧开始处的节点与弧末端的节点之间的关系。一个名为根的内部节点没有进入的弧，它代表整个数据库。

半结构化数据的例子：

比如学生的实验报告。

电子商务中的数据按照产生的对象分为四类：

(1)交易数据，比如电子商务的购物数据。
(2)消费者行为数据：比如评论数据，点击数据，浏览数据。
(3)移动数据:由手机App产生的相关数据。
(4)机器和传感器数据：比如线下体验店的监控视频数据，由无人购物结算的商品扫码信息等。

[1]https://sv.baidu.com/videoui/page/videoland?pd=bjh&context={%22nid%22:%2236570091915669160 67%22,%22sourceFrom%22:%22bjh%22}&fr=bjhauthor&type=video

大数据

大数据的时代背景
信息产业的三个高峰：
第一个是信息高速公路，由Internet的建设驱动，造就了移动，联通和电信等巨头。
第二个是互联网化，由WWW的应用驱动，造就了BAT等巨头。移动互联网化是一个小高峰。
第三个是大数据，云计算和人工智能。

第三次高峰不仅仅影响信息产业，而且会驱动一次新的工业革命，因为很多行业会受到影响。

[1]https://sv.baidu.com/videoui/page/videoland?pd=bjh&context={%22nid%22:%2236570091915669160 67%22,%22sourceFrom%22:%22bjh%22}&fr=bjhauthor&type=video

对比三次工业革命

	第一次工业革命	第二次工业革命	第三次工业革命
时间	18世纪60年代到19世纪40年代	19世纪70年代到20世纪初	21世纪初至今
能源	蒸汽	电力	计算
材料	金属	化学	数据
工艺	机器制造	精密仪器	证析
特征	规模化	自动化	个性化

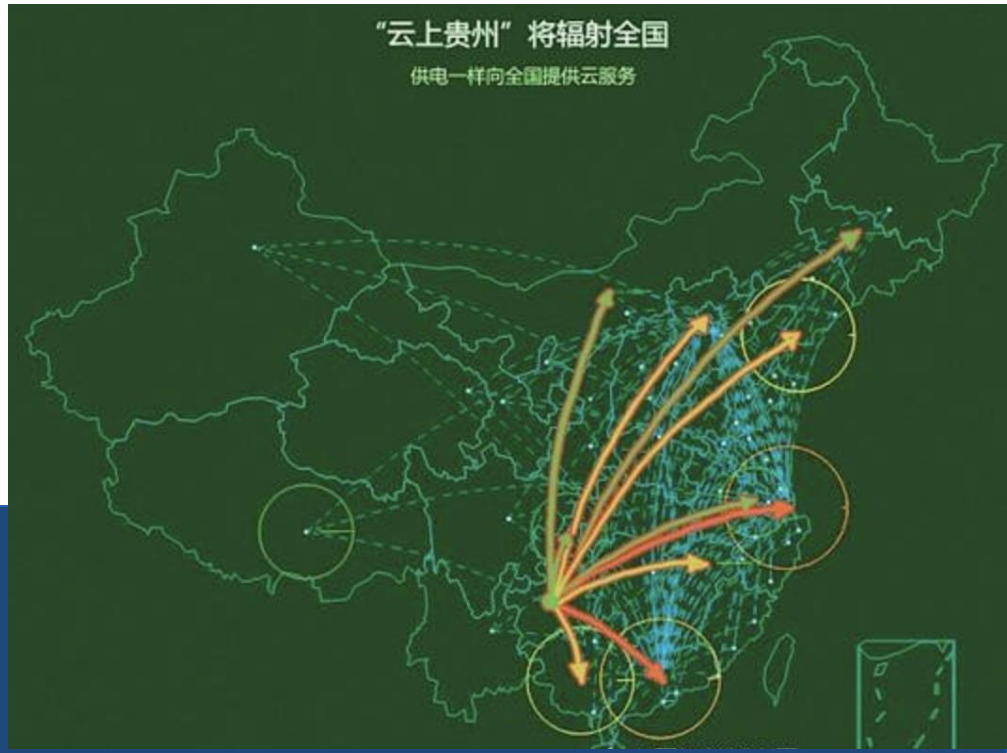
[1]https://sv.baidu.com/videoui/page/videoland?pd=bjh&context={%22nid%22:%2236570091915669160 67%22,%22sourceFrom%22:%22bjh%22}&fr=bjhauthor&type=video

大数据

*计算*是大数据时代的*新能源*:
书桌上的笔记本电脑，口袋里的计算器，云端的计算资源，正在分析天气数据的科学计算大型机，正在优化物流库存和运输方案的云计算中心，正在支持车载GPS导航的内嵌芯片。你不需要知道他们从哪里来，是微软云，阿里云还是京东云等。

计算资源被整合起来，向消费者或是厂家收费。像电费一样，对计算能力收费，在很多生产，创新中计算是一种新的能源。

中国的计算资源布局。
服务器的硬件成本越来越低，运营成本越来越高。
包括土地租赁成本，数据输入输出带宽成本，服务器运行所需的温度，湿度等消耗的电力成本。服务器管理的人力成本。 云上贵州计划



大数据

数据是大数据时代的***新材料***:

我们生产数据产品，提供基于数据的服务，都是建立在原始数据(raw data)的基础上。

认识到数据的基础地位，每一个企事业单位，高校和科研院所，政府机构都有责任把数据保留存储下来。

数到用时方恨少。
建立数据的战略储备，在安全的条件下，尽可能的向社会开放。

大数据开放的隐私保护问题

大数据

证析是大数据时代的先进工艺技术：

商品具有排他性，你有我没有。
数据具有共享性，大家都可以得到。

所谓证析就是大数据分析。基于海量数据进行实证分析，并得到洞见。

新的就业机会：数据科学家，数据工程师。

大数据

个性化是大数据时代的最显著的商业特征：

大数据时代最显著的商业特征是个性化，为每一个终端消费者提供专属的产品和服务。

“商业属于个性化的”，市场细分满足不能消费者的目标。

比如通过个人基因序列的分析，能够分析某个个体早期可能患有某些疾病的风险，进行个性化的治疗。

好莱坞知名女星：安吉丽娜 朱莉由基因测序技术发现其携带某种错误基因，有87%的可能性患有乳腺癌。勇敢的朱莉进行了乳腺切除手术，从而避免了未来患癌症的风险。史蒂夫 乔布斯对自己的DNA进行了测序，一般人患胰腺癌仅能存活几个月，而针对性的治疗使得其生命延续了9年。未来随着技术的成熟，个性化医疗的成本会下降，普通人也可以享用。

大数据的定义

大数据（Big Data）是指大小超出了常用的软件工具在运行时间内可以承受的收集、管理和处理数据能力的数据集。大数据是目前存储模式和计算模式与能力不能满足存储与处理现有数据集规模这一现状而产生的相对概念。

大数据是一种新兴的规模庞大而且复杂的数据集，传统的数据处理程序并不足以应对。这些挑战包括感知，存储，转换，预处理，分析，共享，可视化，查询等各种数据分析任务。

大数据的特征4V

规模性(Volume):不是TB,而是PB,EB或ZB。

多样性(Variety):大数据的类型可以包括网络日志、音频、视频、图片、地理位置信息等，具有异构性和多样性。

高速性(Velocity): 处理速度快，时效性要求高，需要实时分析而非批量式分析，数据的输入、处理和分析连贯性地处理，这是大数据区分于传统数据挖掘最显著的特征。

价值高(Value):大数据分析可以从中得出高价值的数据内容，帮助我们进行更好的研究。因此我们需要对未来趋势与模式做可预测分析，利用机器学习、人工智能等进行深度复杂分析。

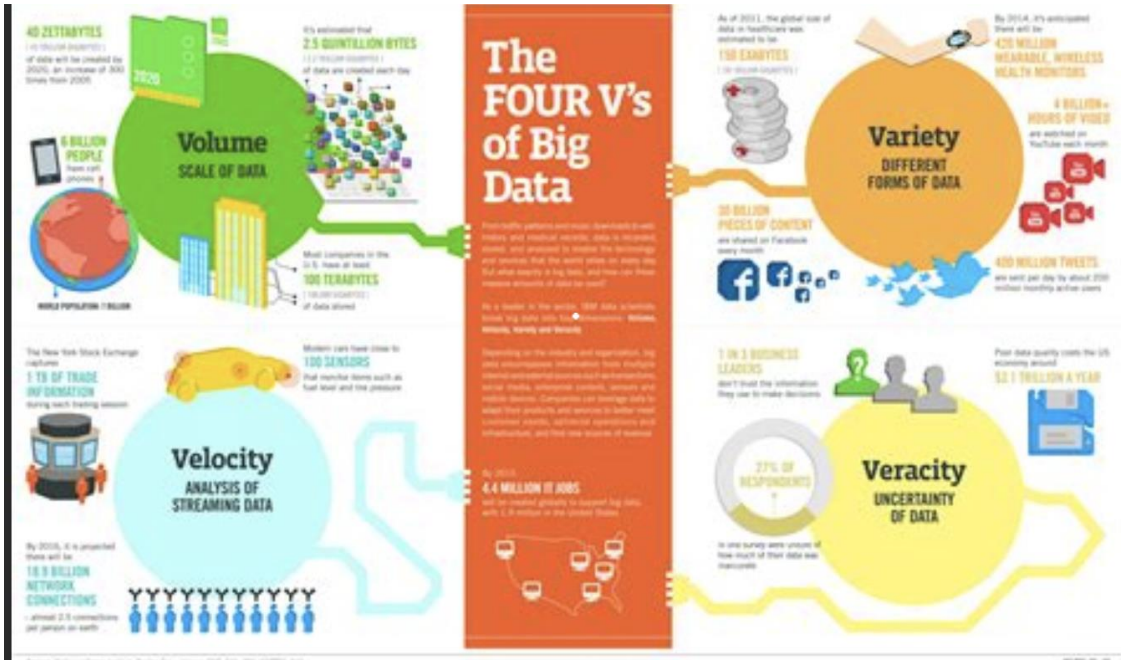
大数据

大数据的特征

4V?

5V?(Veracity)真实性

6V?(Visualization)
可视化



大数据的融合

(1)数据的全空间：涵盖复杂的多维多源的数据集合。
(2)数据的融合：数据综合加工处理。
(3)数据的互补：互补数据的融合才可以使系统发生质的飞跃

大数据的价值
(1)统计呈现洞见：

快递员的通话记录的故事:对北京市所有用户的电话呼叫行为进行分析，通过对呼入和呼出频率的统计，找到了北京所有快递员的号码。快递员打电话特别多，接电话相对少。外卖员则相反。通过快递员顺藤摸瓜，找到了所有网购行为消费者的电话。
通过通话记录分析，可以知道消费者是否买车(4S店通话记录)，是哪家银行的客户(经常接收短信提醒)。

大数据的价值:

(2)关联蕴含价值：

寻找因果关系是人类科学发展的目标。少量关联关系是寻找因果的利器。

沃尔玛通过数据分析发现，和尿布一起购买最多的商品是啤酒。
美国的太太们经常嘱咐丈夫在下班后为小朋友买尿布，他们在买尿布之余也带回来了啤酒。

从数据记录中找到关联的最简单的办法是“关联规则挖掘”，寻找商品中的相关性，更好地指导销售策略的制定。把相关性强的商品放在货柜比较近的位置。

股票市场走势预测和公众情绪在论坛上的表达相关。

大数据的价值
(3)预测指导决策：

预测某个消费者有多大概率点击某个广告或购买某个商品。

阿里巴巴大数据竞赛的课题。

这类问题分四个步骤：

数据清洗，特征提取，把特征输入训练模型，模型的融合。

大数据的价值

(4):大数据的外化

利用与业务无关的数据解决业务中的问题，并把自身的数据贡献出来，解决外面和本业务无关的各种各样的问题。

从行为数据预测学生成绩：

电子科技大学2万名在校学生的85项数据，包括基本信息，历年成绩，选课记录，图书馆记录，食堂超市消费记录，门禁记录，医疗记录，党团活动等。

在寝室呆的越久，统计而言，成绩越差。

进图书馆越多，统计而言，成绩越好。

吃早饭越多，洗澡越规律，成绩越好。

从洗衣服，进出寝室，洗澡等生活规律性发现，生活越规律，成绩越好。

前后连续在食堂刷卡---定义为不期而遇。

从连续打卡人员的发现闺蜜，朋友，情侣或是孤独症患者，对学校心理咨询提供建议名单。

电子商务大数据分析[1].

研究在电子商务环境下的大数据，涉及大量的事务，点击流数据，语音数据，视频数据等电商场景 [2].

大数据分析用于发掘持续变化的模式，事件和机遇，并用于发现新的商业机会和保持敏捷的能力。

案例研究
Alimama (阿里妈妈) [3]

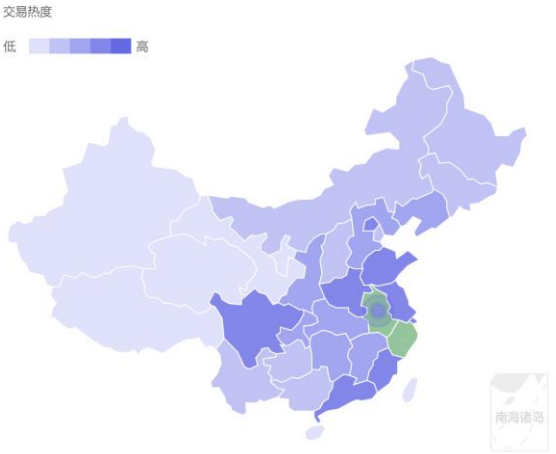
大数据驱动的电商平台
https://www.alimama.com/index.htm

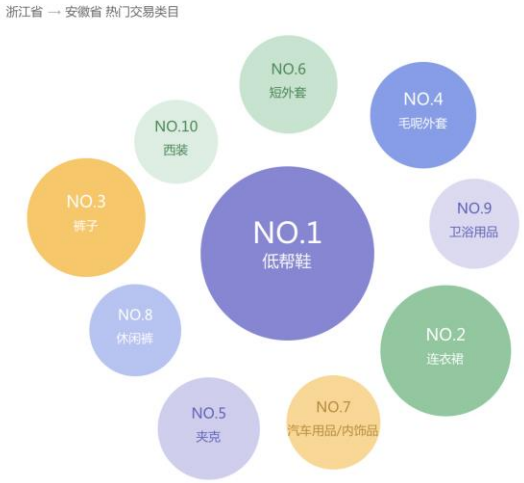
[1] Akter S, Wamba S F. Big data analytics in E-commerce: a systematic review and agenda for future research[J]. Electronic Markets, 2016, 26(2): 173-194.
[2] Davenport, T.H., 2012. The Human Side of Big Data and HighPerformance Analytics. International Institute for Analytics, pp. 1–13.
[3] https://sv.baidu.com/videoui/page/videoland?pd=bjh&context={%22nid%22:%2237020856328291952 99%22,%22sourceFrom%22:%22bjh%22}&fr=bjhauthor&type=video

电子商务大数据分析.

电子商务大数据的四个创新点:
(1)个性化数据导购
用户行为分析，浏览，点击，收藏，评论等数据进行个性化的推荐。
(2)专业化的数据服务模式
淘宝指数
(3)低交易成本的商品流通模式
智能化的仓储和物流。
(4)构建垂直细分的服务模式
从大型综合的京东淘宝，向细分的蘑菇街，唯品会等公司过渡。

案例研究
淘宝指数[1]





[1] https://shu.taobao.com/area/?spm=a2oaa.0.0.0.30716a84BwhyGi

电子商务大数据分析

020电商大数据融合：
线下商业可以到线上吸引客户，消费者在线上完成商品筛选，在线下进行消费体验。

(1)多源异构的数据和信息的融合
局部挖掘获得局部模式，经过全局学习形成全局统一的模式。
(2)知识的融合

模型训练的融合