

$P(A) = \sum p(\omega)$
 $1. A \cap B' \quad \omega \in A$
 $2. A \cap B$
 $3. A' \cap B$
 $4. A' \cap B'$
 $S_n = a^n \frac{a^{n-1}}{a^{-1}} = a^{2n-1}$
 $y = ax^2 + bx + c$
 $x_{1,2} = \frac{-b \pm \sqrt{D}}{2a}$

$V(k,n) = \frac{n!}{(m-k)!}$
 $\vec{u} + \vec{v} = \vec{z}$
 $(a+b)^n = \binom{n}{0} a^n b^0 + \binom{n}{1} a^{n-1} b^1 + \dots + \binom{n}{n} a^0 b^n$
 $e = 2.718281828$
 $\int f(\varphi(x)) \varphi'(x) dx = \int f(u) du$

A	B
1	1
1	0
0	0
0	1
0	0

$\lim_{n \rightarrow \infty} a_n = a$
 $P(A \cap B) = P(A) \cdot P(B)$
 $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{\lim_{n \rightarrow \infty} a_n}{\lim_{n \rightarrow \infty} b_n} = \frac{a}{b}$
 $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \lim_{n \rightarrow \infty} \log_a \sqrt[n]{r} = \frac{1}{5} \log_a r$
 $P(A|B) = \frac{P(A \cap B)}{P(B)}$
 $z_1 \cdot z_2 \cdot \dots \cdot z_n = \sqrt[n]{\prod_{i=1}^n z_i}$
 $y = x^2$

个性化数据挖掘 应用与发展

—— 达观数据

大数据时代实践分享

达观数据 知名的数据技术服务企业

达观数据
DATA GRAND

达观数据是知名数据技术服务企业，拥有中国领先的数据挖掘、自然语言理解、智能推荐、搜索引擎技术，拥有二十余项国家发明专利，入选Xtecher中国100家人工智能创新企业

达观团队获中国著名投资机构真格基金等联合注资，是上海大数据联盟理事单位，由原百度、腾讯、盛大、阿里等公司技术专家组成

达观数据已为数百家企业开发部署了卓越的大数据技术服务，成功为客户创造了巨大业绩





第一部分

数据挖掘应用的发展趋势

大数据挖掘从**整体**统计向**个体**挖掘迈进

整体统计

从整体上统计数据
走势、人群分部、
业务情况

缺点：难以通过
整体数据观察细分
情况



分群统计

按不同的维度对人群
进行细分，针对一类
人群完成统计和分析

缺点：划分方法和
划分精度会带来问题



个体分析

为每个用户进行数据
建模和画像分析，
刻画精确，服务个性化

优点：充分发挥个性化
数据挖掘的潜力

你了解你的用户吗？

你的用户的浏览、购买、搜索、排序行为，你是否了解？



你不知道**用户从哪里来**

你不知道**用户的性别年龄喜好**

你不知道**用户什么时候会流失**



但这些都**非常有价值！**

大数据并不神秘，有很多实际化的应用，涉及生活中方方面面

经典案例：比父亲更早知道女儿怀孕

曾经有一位男性顾客到一家Target超市投诉，商店竟然给他还在读书的女儿寄婴儿用品的优惠券。这家全美第二大零售商，会搞出如此大的乌龙？但经过这位父亲与女儿进一步沟通，才发现自己女儿真的已经怀孕了。

原来，Target有一个庞大的数据库，里面记录了每一位顾客的消费内容、时间等信息，通过对这些信息的挖掘和分析，塔吉特准确地识别出哪些顾客是孕妇，甚至估算出她们的预产期。

企业投入大量预算搞一场营销活动



零星的转化



一堆客户联系方式



客户联络名
单.xlsx

躺在Excel表格里

你不知道你的粉丝是谁，他们从哪里来，你不知道如何再次找到你的粉丝，下一场活动又是同样的投入同样的结果，对企业发展来说是非常不健康的。

用户从哪里来

达观大数据服务



通过大数据，分析不同渠道的引流效果，计算成本和收益
帮你寻找到最优的获取新用户的渠道

为你生成用户的行为标签

属性标签：

性别

年龄

城市

职业

爱好

行为标签：

浏览

点击

收藏

购买

评论



生成多维度数字化的用户画像



帮你自动分类筛选特定类型的用户，查看每个用户的偏好

智能推荐

- 推荐数据
- 推荐配置

搜索引擎

- 基本数据
- 行为分析
- 搜索配置

精准营销

- 用户画像**

设置

- 账号设置
- 宽表设置
- 访问设置
- 名称映射
- 功能布局

用户属性	level_1	req_2014	req_2013	req_2012	req_2015	req_2011	req_2010	req_2009	level_2	req_2016	更多
其他	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
ID	昵称	邮箱	电话	活跃程度	消费能力	操作					
10327318	18369188183			720	90	查看详情					
10545005	冬天不冷了吗			720	29	查看详情					
1668342	曲荷风			720	295	查看详情					
5033557	eugenieli			717	55	查看详情					
4520826	yingmanne			717	21	查看详情					
4567953	15976747892			716	158	查看详情					
6494530	藍love海			716	18	查看详情					

精准营销 用户详情

用户ID: 10327318
 用户昵称: 18369188183
 手机号码:
 活跃程度: 720
 消费能力: 90

用户画像表

标签	重生	宠文	女强	宝宝	婚恋	爽文	腹黑	专情	种田	明星
类别	都市	穿越	豪门	架空	异能	玄幻	青春	仙侠	职场	校园
用户属性	level_3 r...									

用户流失预测和挽回方法

建立用户画像以后，可以实时对用户的行为进行追踪和分析

用户什么时候会流失

什么原因导致的流失

什么方法能挽回



找不到想要的商品

比竞争对手贵

操作繁琐，流程麻烦

客服态度不好，没反应

达观大数据挖掘和预测服务

“用户浏览网站时喜欢看什么，对哪些内容感兴趣，如何让用户在网站停留更长时间？”

“根据以往的销售数据统计，如何预测企业未来一段时间内的销量走势？”

“如何准确地预测用户的诉求，将用户寻找的内容，超前提供给用户？”



达观数据
预测引擎



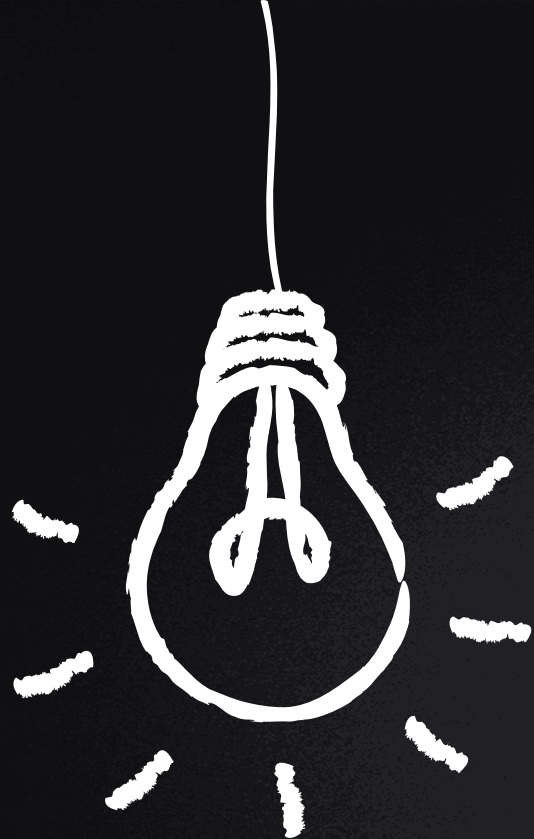
智能推荐

搜索引擎

数据预测

数据抓取

语义分析



第二部分

对传播路径的数据预测

只统计过去的的数据意义有限

商品数量统计

商品种类统计

商品销量统计

只能做到对过去数据的总结和统计，然后变成



产品销售报
告.xlsx

躺在Excel表格里

面向未来的数据预测才有更大价值

除了这些基础功能以外，我们还需要对商品统计结果继续分析挖掘

商品销量预测

热卖商品预测

库存预测

通过对商品销量、热卖的预测，来帮助企业很好地解决库存问题

传统企业离大数据并不遥远

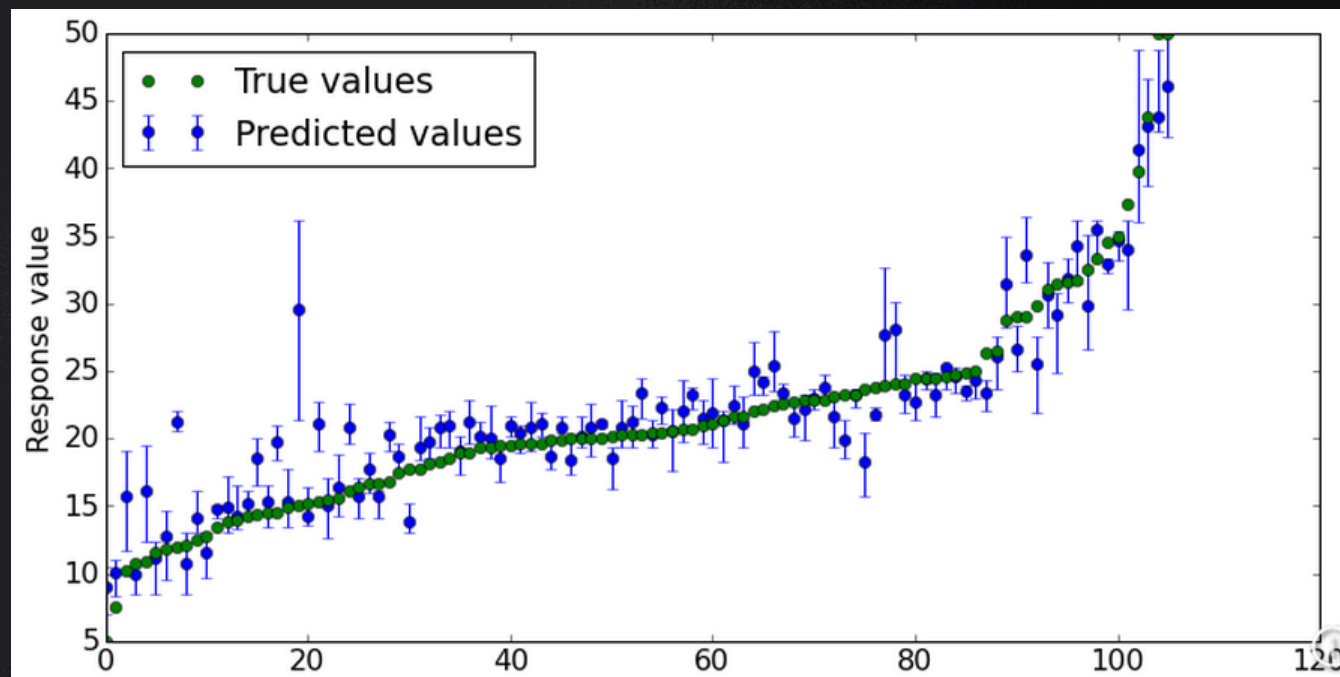
经典案例：沃尔玛的大数据产业链

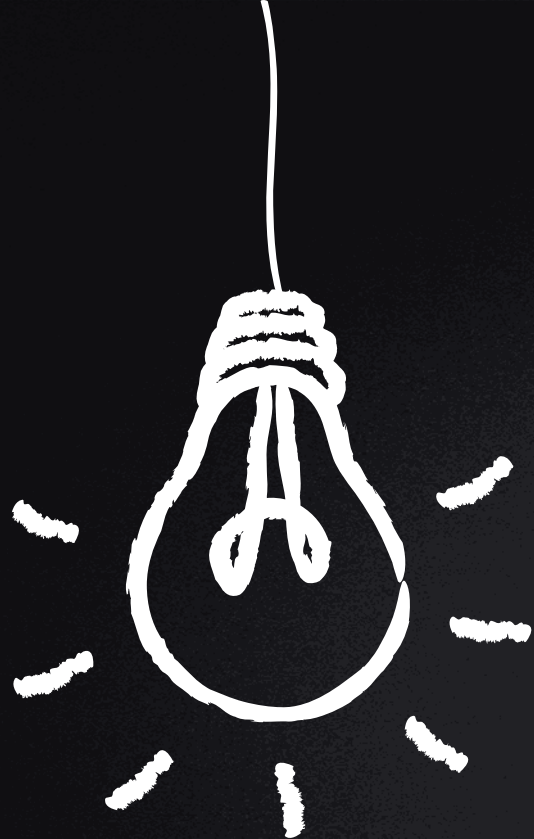
沃尔玛开发了一个叫做Retail Link的大数据工具，通过这个工具供应商可以事先知道每家店的卖货和库存情况，从而可以在沃尔玛发出指令前自行补货，这可以极大地减少断货的情况和供应链整体的库存水平。

通过在整条供应链上分享大数据技术，沃尔玛引爆了零售业的生产效率革命。

结合行业动态和历史数据

已有成熟的方案完成预测

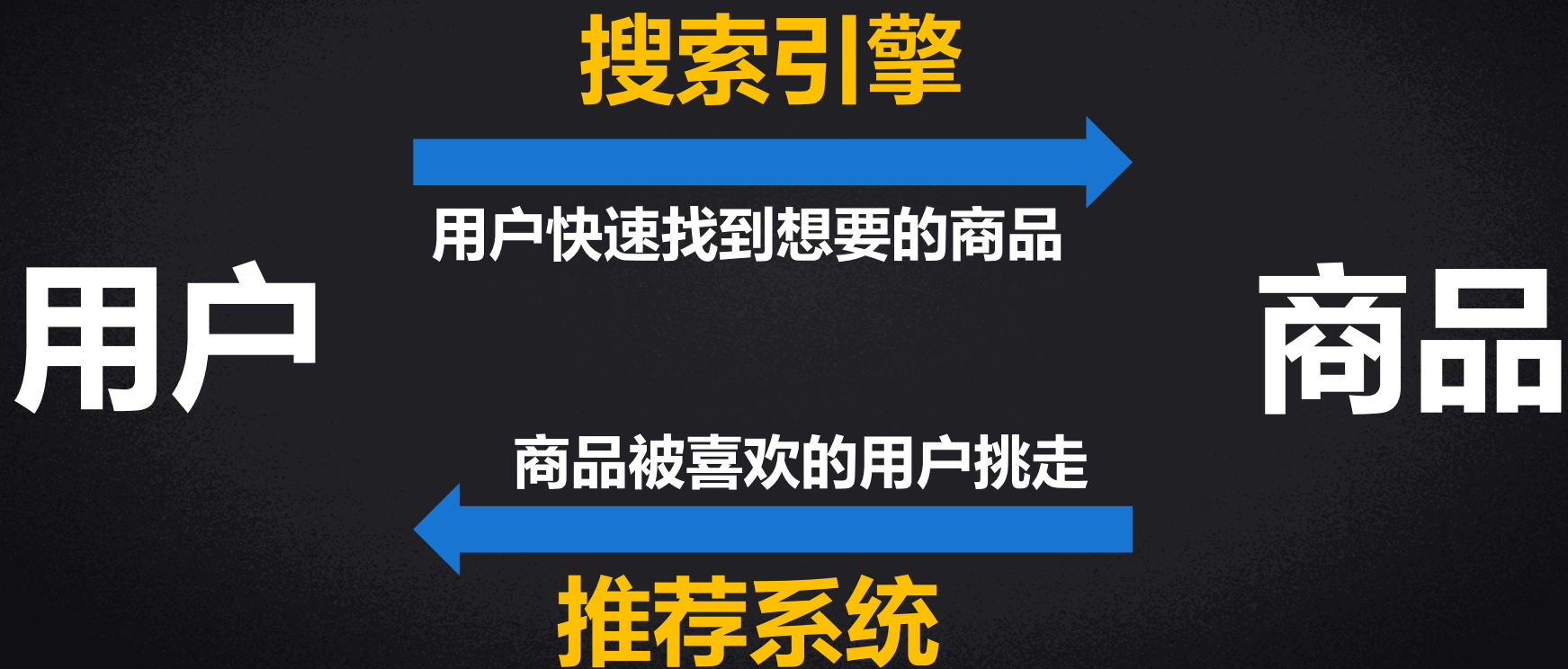




第三部分

通过大数据实现信息高效传递

连接用户与商品的两种方式



拼音识别

自动变换

自动纠错

达观搜索系统

响应快

搜索提示

相关搜索

前百度团队 倾力打造 堪比百度搜索的站内搜索

用户在店铺内的行为数据分析



根据行为数据预测用户需求



润物细无声的个性化推荐

我们接触最原始的“**推荐系统**”，应该就是服装店的导购员了



导购员总会根据你看上去的年纪、身材，不厌其烦地为你推荐商品，他们凭借丰富的经验、阅人无数的经历来为自己的推荐提供可靠的依据。

而现在这一切都只需要交给

算法

个性化技术在推荐系统的常见应用形式

看过此商品的顾客最终购买

9.0% 的用户购买本商品，还有：



3.6% 购买

古奇天伦硬汉长城底休闲鞋5102 咖啡色

市场价 ¥ 538 售价 ¥ 158



2.7% 购买

古奇天伦欧美百搭风反毛牛皮日常休闲男鞋5559 咖啡色

市场价 ¥ 538 售价 ¥ 158

看过本商品的还看了



蒲公国际大奖小说系列 (共6册)

¥60.50



小牛顿科学馆全集 (全60册, 马英)

¥624.00



彩色世界童话全集 (全60册, 12位)

¥381.20



美丽故事绘本 第一辑

¥64.40

推荐菜 | 环境 | 价目表

招牌乒乓鲜虾净云吞(359) 冰火菠萝油(237) 查厘士绿茶(186) 咖喱牛腩(159) 查厘士避风塘虾(79) 烧味拼盘(72) 查厘士炒饭(52) 玫瑰就油鸡(50) 港式奶茶(49) 奶油脆皮猪仔包(40) 招牌乒乓鲜虾云吞面(34) 咖喱鱼蛋(29) 海南鸡饭(29) 更多



招牌乒乓鲜 ¥ 35



冰火菠萝油 ¥ 10



查厘士绿茶 ¥ 24



咖喱牛腩 ¥ 58



查厘士避风 ¥ 78

更多

搜缘分，就在眼前 我要找 女 年龄 22 - 28 岁 地区 上海 搜索 高级搜索 >>



奔三D女子很美

28岁，上海，160cm，本科，5000~10000元

终有那么一个人，会让彼此满足也倍感幸福。...

查看详情 >>



冬眠的泰迪熊

32岁，上海，182cm，大专，2000~5000元

人生可能不会一帆风顺，但是只要坚持寻找自己...

查看详情 >>



愤怒的小鸟

28岁，上海，155cm，硕士，2000~5000元

虚拟的世界靠谱嘛？呵呵。所以希望真诚交友！...

查看详情 >>



-品味人生-

32岁，上海，175cm，本科，5000~10000元

希望在这里能找到那个心爱的你，最好是成都女...

查看详情 >>



性感天后时尚魅力大片拍摄花絮

"花絮 魅力"高清视频



Coco复古贵妇短片
编辑点评：生活！



彩绘舞者惊现太古里三里屯!

"三里屯"最热视频



北京三里屯保时捷体验中心 活动花絮

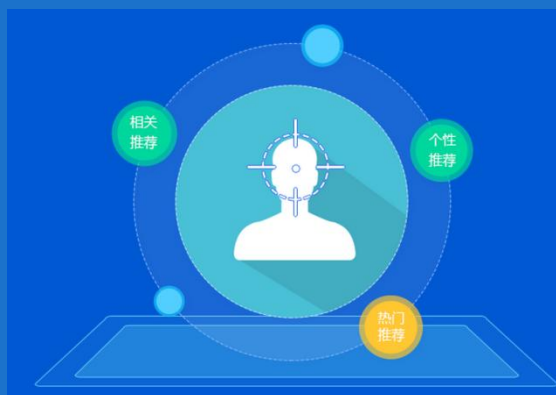
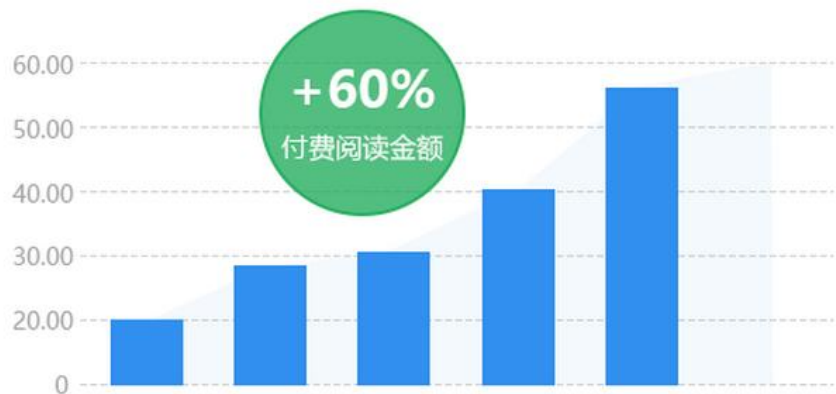
"三里屯 花絮"相关精选

达观个性化推荐：分析每个用户

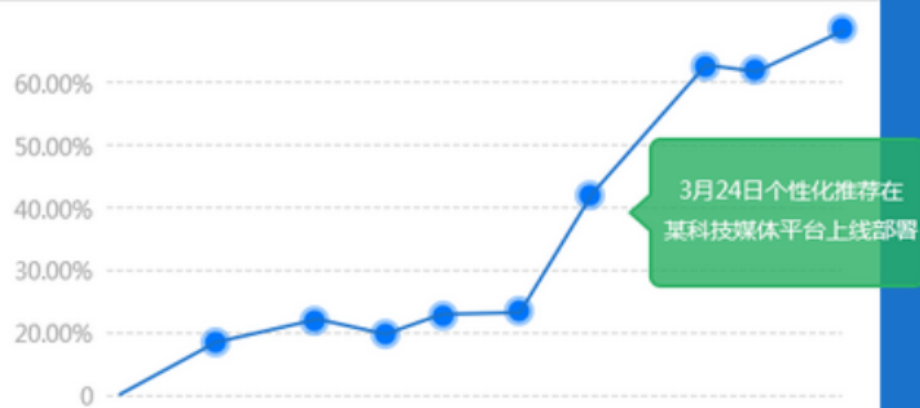
达观数据
DATA GRAND

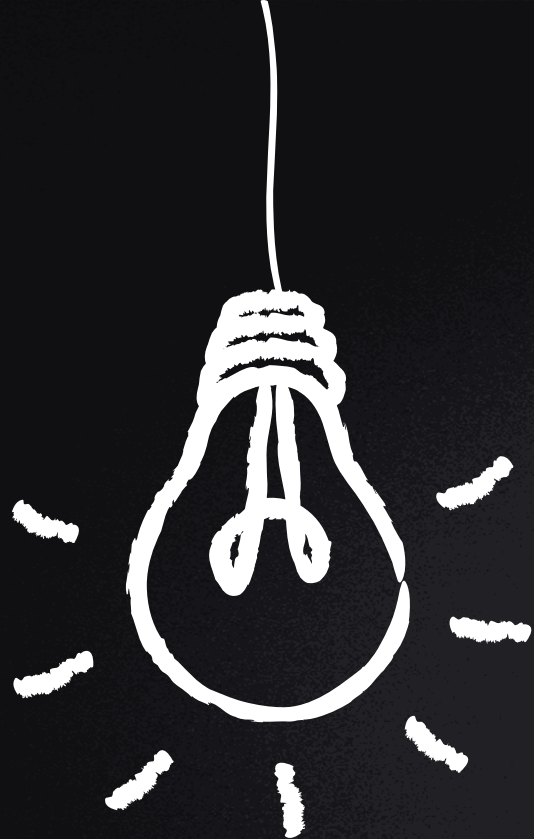
- 不再是将用户分群体来进行整体的刻画，而是为每个用户单独进行行为分析
- 根据每个用户的历史行为来生成一个模型，分析偏好的内容，个性化的推荐感兴趣的内容
- 自动的完成内容和用户的撮合交易，促进企业经营业绩的提升

某小说网站接入达观推荐服务前后付费阅读金额



个性化推荐在某科技平台上线部署前后对比





第四部分

达观提供的大数据挖掘服务

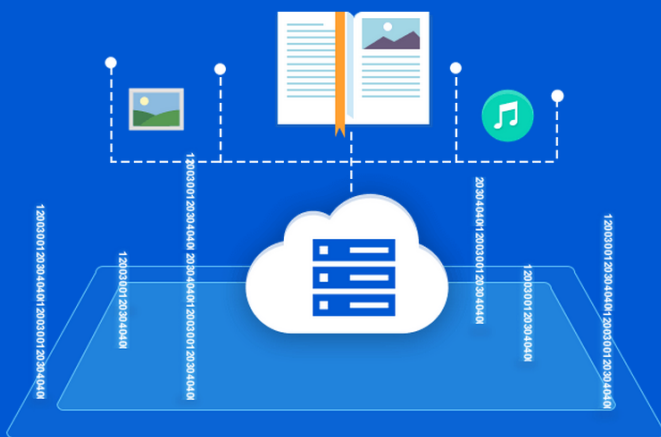
达观数据提供的企业技术服务

文本挖掘

从海量文本中精炼高质量信息

高精度的挖掘结果，毫秒级的处理速度，高度稳定的云服务，简单至极的使用方法。达观自动化的文本挖掘服务为企业节省大量人力物力，提升经济效益。

立即体验

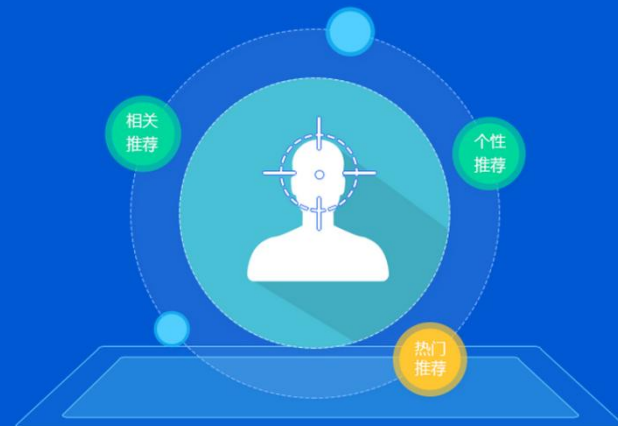


推荐系统

引导用户发现自己的需求

世界级冠军团队为您提供扩展性强、可用性高、响应速度快的精准智能推荐系统。

申请试用



搜索引擎

从海量数据中快速准确找到目标信息

智能的用户搜索意图识别，秒级的实时索引更新，基于用户点击反馈的智能排序，低成本的高速接入。

申请试用



数据抓取

从海量数据中快速采集目标信息

达观利用集群分布式的数据抓取系统，全面、快速、实时地抓取互联网公开信息，为用户持续提供海量目标数据。

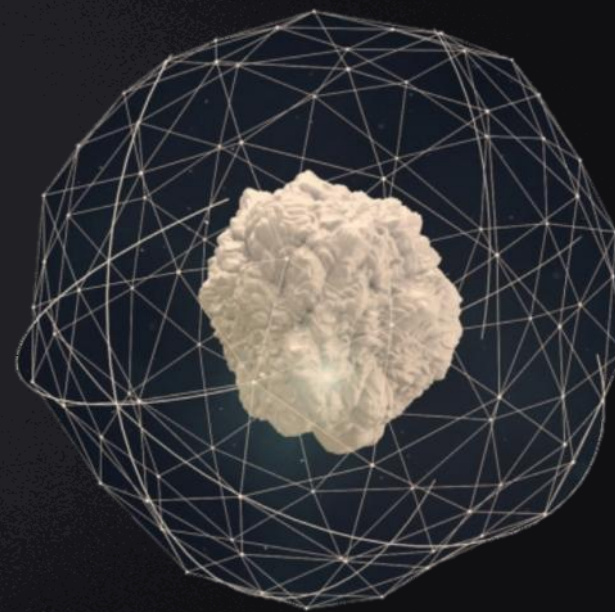
申请试用



机器学习+大数据将代劳越来越多的工作

达观数据
DATA GRAND

- 判断一个商品是否会热卖
- 判断一个用户会喜欢哪些商品
- 判断一个商品最适合的售价
- 判断一个商品适合与那些商品一起售卖
- 自动获取行业情报数据并进行分析
- ○ ○ ○ ○ ○



达观 帮客户更好的运用数据



最卓越的技术



灵活的部署方案



贴心的服务维护



客户业绩迅速提升



欢迎访问达观数据公众号