

大数据，且挖掘且珍惜

叶邦宇

中国科学院信息工程研究所

ICST国家工程实验室

agenda

- 1, 什么是大数据
- 2, 大数据的应用/场景
- 3, 四类大数据相关技术
- 4, 研究现状
- 5, 就业
- 6, 参考书和建议

What is Big Data???

- 大数据？多数据！
- 1000亿个文件，每个文件1KB
- OR

What is Big Data???

- 大数据？多数据！
- 1000亿个文件，每个文件1KB
- OR
- 一个文件，大小为100000GB

What is Big Data???

- 大数据？多数据！
- 1000亿个文件，每个文件1KB
- OR
- 一个文件，大小为100000GB

What is Big Data???

谷歌官方博客发文称其索引的网页数目已超过1万亿


2008-07-26 17:14:28 来源: 网易科技报道 [网友评论 5 条](#) [进入论坛](#)

今天Google在其官方博客发表文章,称Google索引的网页数目已经超过1万亿。

[网易科技讯](#) 7月25日消息,今天Google在其官方博客发表文章,称Google索引的网页数目已经超过1万亿。

Google的索引工作于1998年开始运行,当时收集的网页数目为2600万个。10年后的今天,这一数目达到了惊人的1万亿。

在文中Google也承认互联网非常之大。互联网上到底在多少个独立的惟一页面? Google表示自己也搞不清楚。严格来说,网页数目几乎是无穷尽的。

另外Google还扼述了自己如何索引到所有这些网页的:首先搜索爬虫机器人找到一组互相链接的网页,顺着其中的链接抓取新的网页。然后再通过新网页包含的链接,抓取到更多的新网页和新链接。在这个过程中,Google找到了超过1万亿个链接,但并非所有这些链接都指向惟一性内容,因为许多不同的链接指向同样的网页。在移除了重复的链接后,剩下的就是Google今天索引到的1万亿个惟一性链接。而互联网上的链接都在以每天数十亿的数目逐日增加。(蒋彬) 

What is Big Data???

Google Web Search: 1999 vs 2010

- 文档数: 数千万 to 数万亿
---100000X
- 更新频率: 几个月to 数十秒
---50000X
- 查询所需时间: <1s to <0.2s
---5X

task1

- 泉州师院，男女比例4:6还是3:7？

task2



task3

热门商品推荐



匹克篮球服套装运动服吸汗背心

¥48.00

换一组

淘

task4

- 奥巴马竞选团队利用数据挖掘寻找潜在投票者，为其中特定人群设计竞选广告。竞争激烈的州，通过数据挖掘获得微弱优势可能至关重要。



task5

- 微软研究院的科学家David Rothschild在2012年利用大数据预测美国总统大选，51个选区中50个选区预测均正确，准确率达到98%



2012 PRESIDENTIAL ELECTION
美国大选



task6

- 谷歌的科学家们利用大数据，通过分析哪些地区的用户在谷歌上搜索“哪些是治疗咳嗽和发热的药物”这样的搜索数据，来预测流感的爆发。谷歌的预测模型的准确度高达97%。



task7

#直击总决赛#自1985年总决赛实行2-3-2赛制以来，系列赛打成1-1情况下，拿下第三场的球队夺冠几率为92.3%（12-1）。

[查看大图](#) | [向左转](#) | [向右转](#)



关注NBA官方微信帐号

I. 输入NBA_Big, 添加好友

II. 扫描左侧二维码



weibo.com/nba

6月12日 11:50 来自专业版微博 | 举报

转发(2042) | 收藏 | 评论(665)

task8

- 美国印第安纳大学的约翰·博伦通过跟踪Twitter社交网站上股民的发言情绪，可以对3天后道琼斯工业平均指数进行预测，预测精度高达86.7%



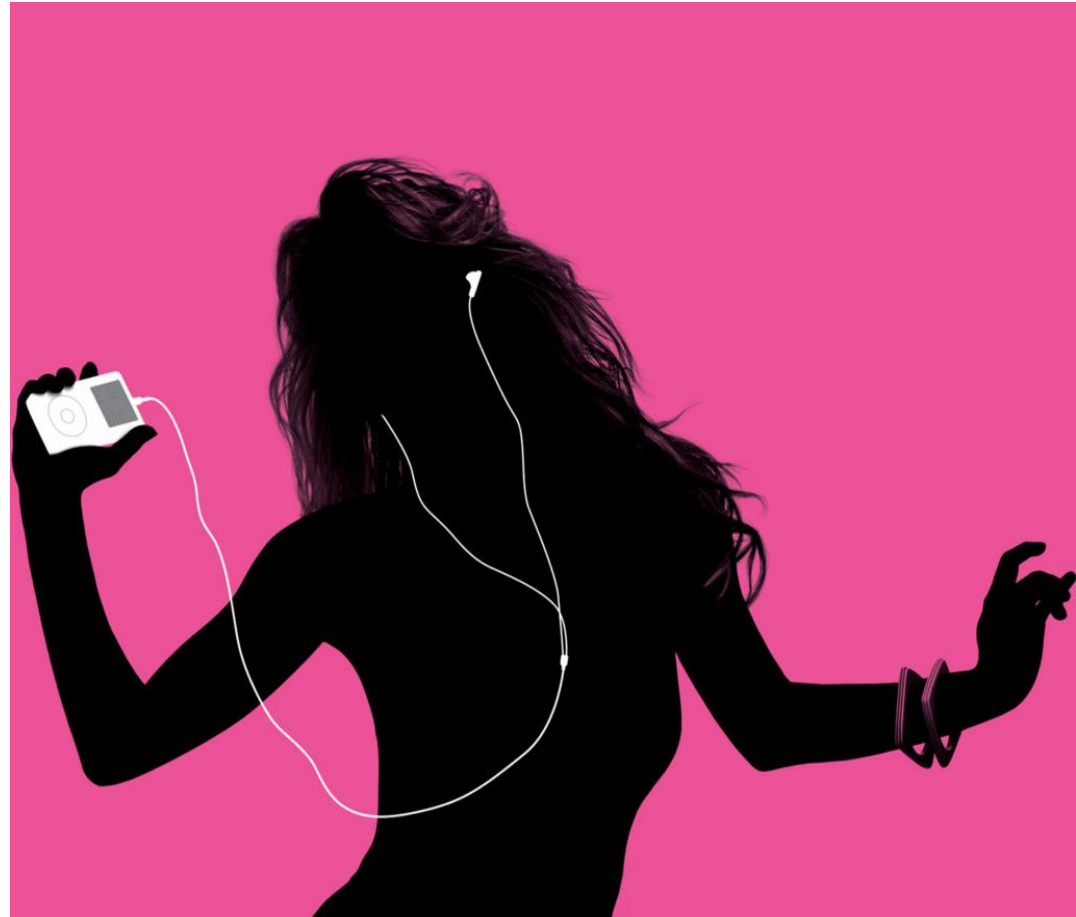
task9

- David Rothschild对奥斯卡奖进行预测，除了最佳导演奖外，其他奖项全部命中

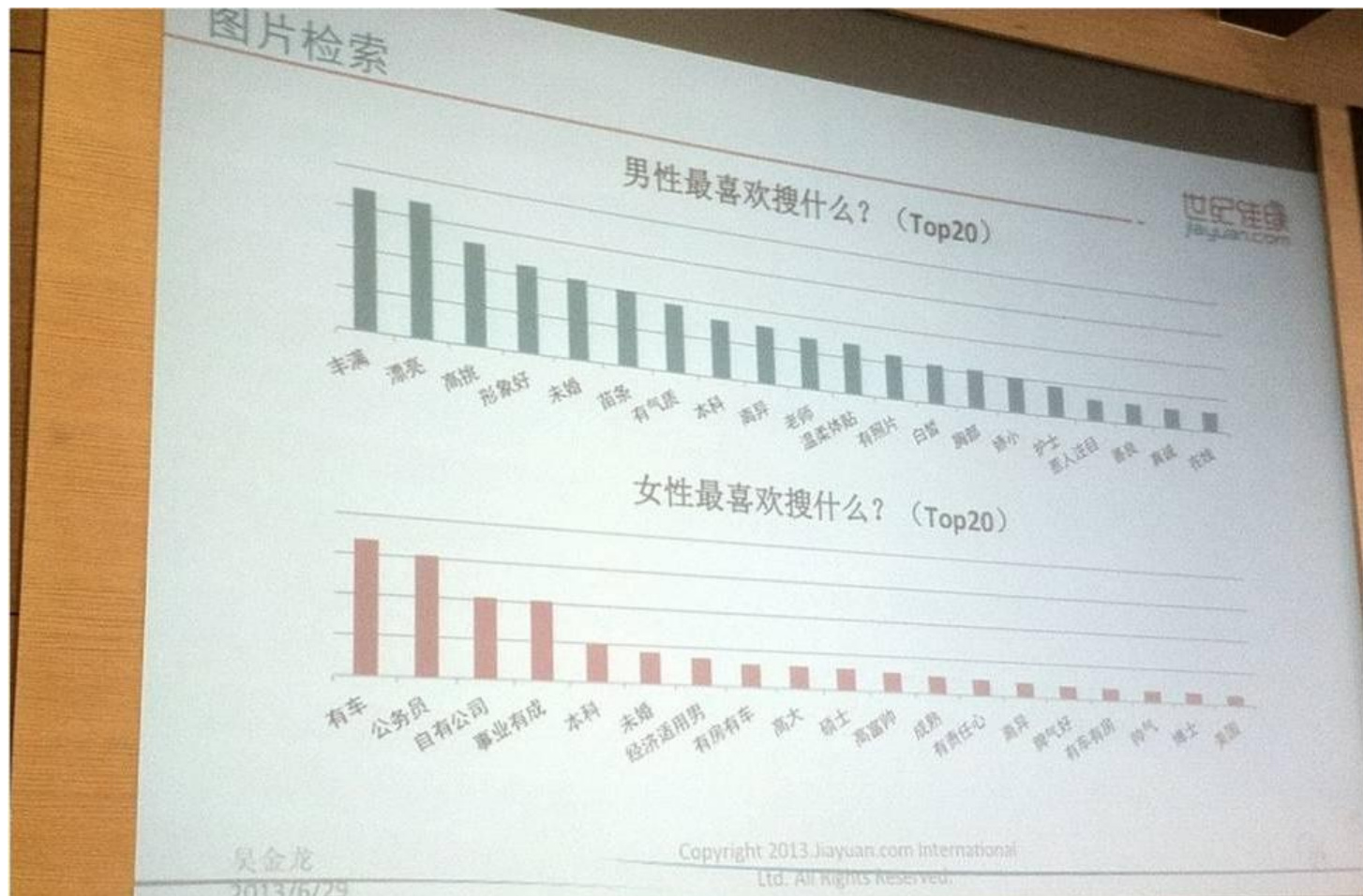


task10

- 美国某疾病治疗中心跟踪观察病人的心跳等，来给病人推荐适当风格的音乐，来辅助治疗疾病



task11



task12

- 专家通过数据挖掘方法得出，对城镇就业人员整体而言，最优退休年龄为64.14岁



task13

- ManWin，色情行业的帝国企业，全球排名第一的成人站就是他们控股，每个月能拿到大约16亿浏览者的数据资料

task13

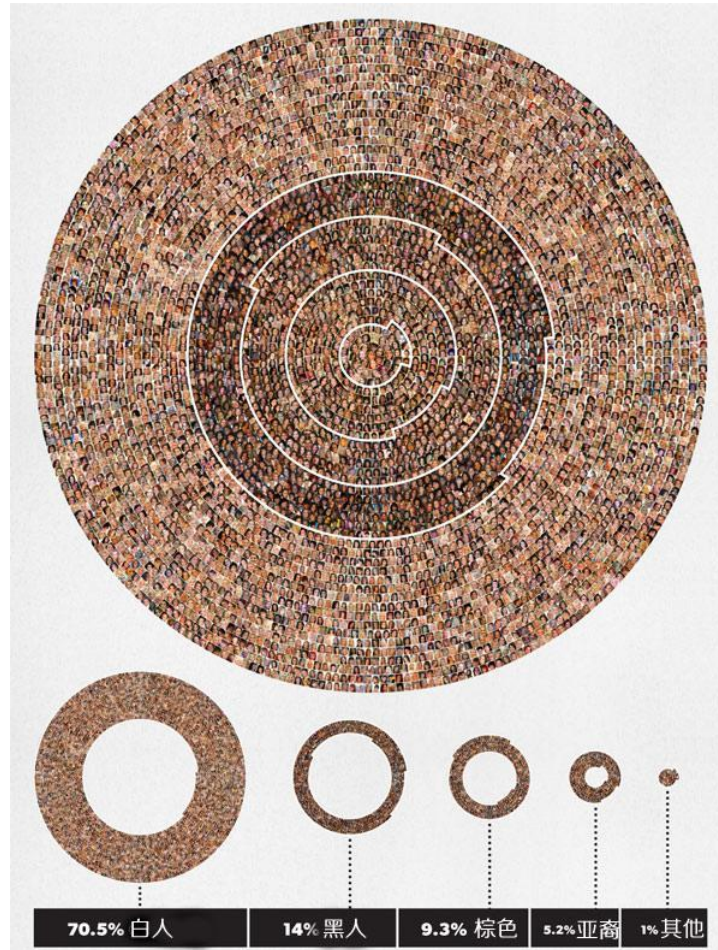
最受欢迎的艳星大众脸



欧美艳星的头发颜色



task13



task13

- 排在第一位的罩杯是B，最常见的尺寸是34B。紧随其后的是C，其次是D，接着是DD(欧美标准，大约相当于F杯)
- 最勤劳的男艳星汤姆拜伦一共拍了2549部色情片，和1127位女星合演过
- 最勤劳的女艳星尼娜-哈特林,她和199个男艳星合演过

task13

- 10%到30%的女性会只拍一部作品就销声匿迹
- 对作品的标题进行分析时，发现排名第一的关键词是**TEEN**，有近**2000**部作品标题中包含这个词。而排名第二的是**MILF**，排名第三的是**Wife**

task13

- 利用大数据：
- 1.帮助色情帝国转型
- 2.带来的性岗位就业
- 3.带来的产品与技术创新

task14



conclusion

- 1, 家庭生活
- 2, 医疗
- 3, 电子商务
- 4, 游戏/娱乐
- 5, 科学研究
- 6, 国民政策制定/总统选举
-

Four categories of Big Data

- Fetching 采集
- Storage 存储
- Mining 挖掘
- Processing 处理

Four categories of Big Data

- Fetching 采集
- Storage 存储
- Mining 挖掘
- Processing 处理

Subjects...

- 1, Data Mining, 数据挖掘
- 2, Machine Learning, 机器学习
- 3, Natural Language Processing 自然语言处理
- 4, Recommender System, 推荐系统
- 5, Social Network, 社交网络
- 6, Search Engine , 搜索引擎

Social Network

- 微博
- Twitter
- Facebook
- 有趣的应用：
 - a, 社团发现
 - b, 僵尸粉
 - c, 情感分析/性别判断/学历.....
 - d, 原词挖掘

判断一个粉丝是不是黑粉

特征

-
- 转发多，原创少
- 关注多，粉丝少
- 内容基本上都是广告
- 头像基本上都是美女图片
-

判断一个粉丝是不是黑粉

- If(粉丝数<=10&&关注数>=1000)
- if(原创数<=100&&转发数>=1000)
- if(广告所占比例>=90%)
-

Machine Learning

- 分类和回归
- 算法:
 - Linear regression,
 - Logistic regression,
 - SVM,
 - kNN,
 - Deep Learning,
 -

Machine Learning

百度识图-上传图片 查询  x

← → ↻

 **Baidu 图片**
识图 测试版

粘贴图片网址  | [从本地上传](#)

提示: 您也可以把图片拖到这里

[玩快乐猜图](#) [赢礼品](#)

 **找到该图片不同尺寸229张**
原图尺寸: 500x370
筛选该图片的其他尺寸:
[全部尺寸](#) [大尺寸](#) [中尺寸](#) [小尺寸](#) [精确尺寸](#) » [取消尺寸筛选](#)

 200x200 jpg	 333x226 jpg	 112x135 jpg	 539x404 jpg
			

Why we need machine learning

- WHY



Why we need machine learning

- Drawbacks of Rule-based

- 1, not elegant 不优美

- 2, hard to maintain 不好维护

- 3, lack of Explanations of probability 木有概率解释

Machine learning

- 1, 特征选择(feature selection)
- [转发数, 粉丝数, 关注数....]
- 2, 标记数据(labelling training data)
- ID 转发数 粉丝数 关注数 y/n
- 1 10 20 21 0
- 2 2300 10 3000 1
-

Machine learning

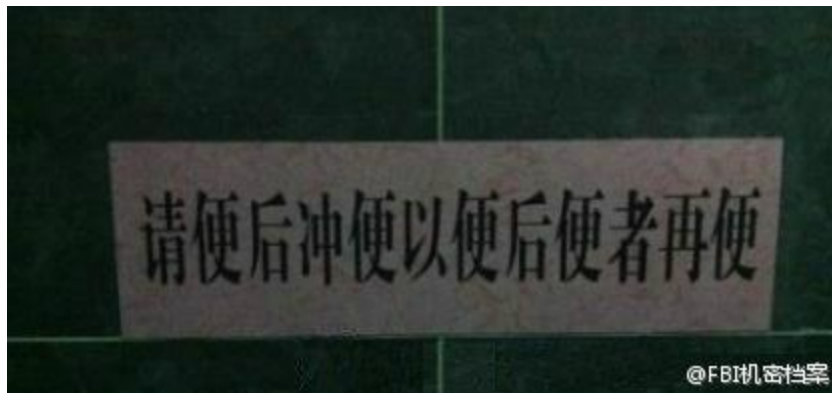
- 3, 模型选择(model selection)
- $Y = a_1 * \text{转发数} + a_2 * \text{粉丝数} + a_3 * \text{关注数}$
- 4, 模型训练(training)
- => 确定参数 a_1 , a_2 , a_3 的值
- 5, 得到模型
- $Y = 0.0052 * \text{转发数} + 0.019 * \text{粉丝数} + 0.0743 * \text{关注数}$

Data Mining

- 算法: K-means,BFR,CURE.....
- 应用 : 疾病诊断

NLP

- 女人如果没有了男人就恐慌了
- 明日逢春好不晦气，来年倒运少有余财
- 此屋安能居住，其人好不悲伤



NLP

@全球震惊创意 🏆

冬天：能穿多少穿多少；夏天：能穿多少穿多少。「转」

📌 收起 | 🖼️ 查看大图 | ↶ 向左转 | ↷ 向右转



6月18日 12:18 来自新浪微博

👍(49) | 转发(590) | 评论(131)

NLP

- **【2014年高考语文题】** 央视报道否认马航否认之前否认交通部长否认过的由大马军方否认外媒的否认消息。——请考生答：这个消息属实吗？

Recommender System

- Amazon:

“查看此商品的顾客也查看了”

“购买此商品的顾客也购买了”

浏览更多商品

您浏览过

查看此商品的顾客也查看了



《古兰经》与伊斯兰文化
王新生
平装
★★★★☆ (6)
¥38.00 ¥ 29.40



古兰经
马坚
精装
★★★★★ (53)
¥ 63.40



石头读古兰经
马石头
平装
★★★★★ (1)
¥28.00 ¥ 21.00



谁代表伊斯兰讲话? 十几亿穆斯林的
真实想法
约翰·L·埃斯波西托(John...
平装
★★★★☆ (9)
¥29.00 ¥ 22.90



伊斯兰教史
金宜久, 任继愈
平装
★★★★☆ (13)
¥42.00 ¥ 33.60

> [查看或编辑您最近浏览过的商品](#)

Recommender System

- 喜欢还是不喜欢？ 喜欢是多喜欢？

Recommender System

- **rating**
 - rating=1代表非常讨厌
 - rating=2代表讨厌
 - rating=3表示一般般
 - rating=4表示喜欢
 - rating=5表示非常喜欢

Recommender System

	大闹天宫	爸爸去哪儿	澳门风云
A:	1	2	5
B:	5	5	4
C:	5	5	1
D:	1	1	5
E:	1	2	4

Recommender System

- 1, Collaborative Filtering （协同过滤）
- Neighborhood-based （基于邻居）

User-based （基于用户）

Item-based （基于商品）

- Model-based （基于模型）

SVD

SVM

.....

Recommender System

	大闹天宫	爸爸去哪儿	澳门风云
A:	1	2	?
B:	5	5	4
C:	5	5	1
D:	1	1	5
E:	1	2	4

User-based

- $\text{Rating}(D, \text{澳门风云}) = 5$
- $\text{Rating}(E, \text{澳门风云}) = 4$

User-based

- $\text{Rating}(D, \text{澳门风云}) = 5$
- $\text{Rating}(E, \text{澳门风云}) = 4$
- $\text{Prediction}(A, \text{澳门风云})$
- $= (5 + 4) / 2 = 4.5$

User-based

- $\text{Rating}(D, \text{澳门风云}) = 5$
- $\text{Rating}(E, \text{澳门风云}) = 4$
- $\text{Prediction}(A, \text{澳门风云})$
- $= (5 + 4) / 2 = 4.5$
- $= 5 * 0.5 + 4 * 0.5 = 4.5$

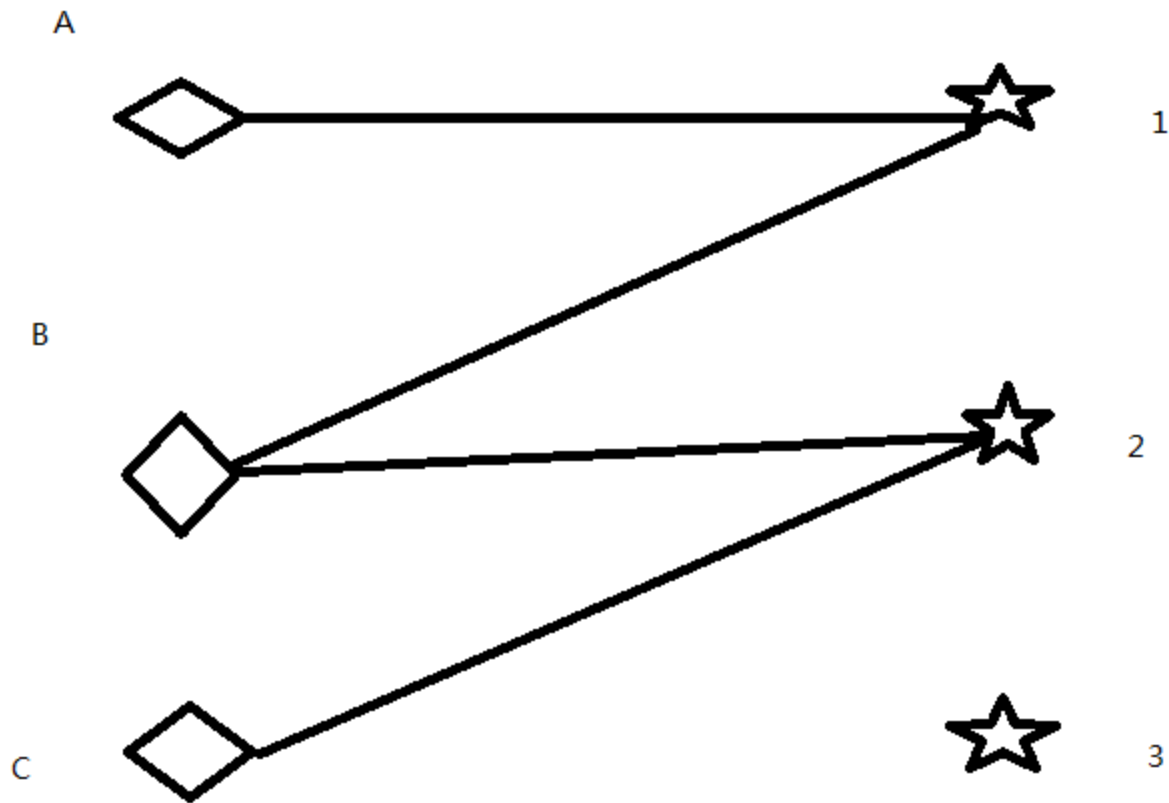
Different weights

- $\text{Rating}(D, \text{澳门风云}) = 5$
- $\text{Rating}(E, \text{澳门风云}) = 4$
- $\text{Prediction}(A, \text{澳门风云})$
- $= 5 * 0.8 + 4 * 0.6 = 6.4$

Normalization

- $\text{Rating}(D, \text{澳门风云}) = 5$
- $\text{Rating}(E, \text{澳门风云}) = 4$
- $\text{Prediction}(A, \text{澳门风云})$
- $= (5 * 0.8 + 4 * 0.6) / (0.8 + 0.6) = 6.4 / 1.4 = 4.57$

Graph-based



Recommender System

- 2, Content-based

User profile

Item description

Recommender System

- 3, Knowledge-based

Constraint-based （基于约束）

Cased-based （基于实例）

Recommender System

- 4, Hybrid approach

Recommender System

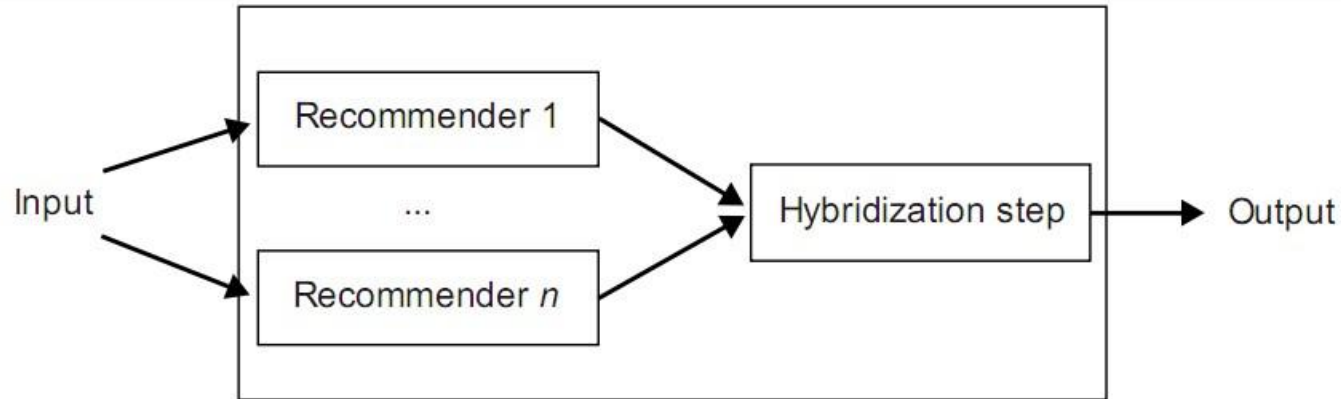


Figure 5.3. Parallelized hybridization design.

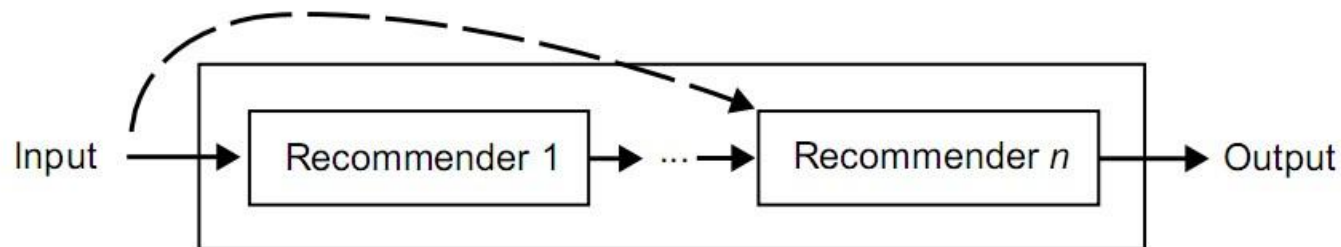


Figure 5.4. Pipelined hybridization design.

Recommender System

- What to do right now and next?
- 1, Cold start /Data sparsity problem
- 2, Scalability of model
- 3, Online-learning
- 4, Explanations in recommender systems
- 5, Attacks and protections

.....

Search Engine

- 1, TF-IDF
- 2, PageRank算法
- 3, SEO

Search Engine

- 机长神秘电话

TF-IDF

- 分词

机长 神秘 电话

- 搜索

搜索和“机长神秘电话”相关的文档，呈现给用户。

TF-IDF

- 哪些网页可能是相关的呢？
- 考虑两个方面：
- 1, “机长”、“神秘”、“电话”这些词出现的次数越多越好

网页1: 机长 神秘 电话 机长 神秘 电话

网页2: 机长 神秘 电话

TF-IDF

- 2, 这些词, 权重显然不尽相同

网页1: 机长 神秘 电话 电话

网页2: 机长 机长 神秘 电话

TF-IDF

- 词项出现的次数:

词项频率, TF, Term Frequency

- 词项在多少个文档里有出现:

逆文档频率, Inverse Document Frequency

IDF

$$\text{IDF} = \log(N/n)$$

N代表文档总数, n代表包含该词的文档数

TF-IDF

- TF，词项频率

一个单词在一个文档上出现的次数

- IDF，逆文档频率

一个单词在所有文档上出现的情况

PageRank

- 作弊...
- 自己说自己牛逼没有用，关键是别人说你牛逼。
- 自己说自己牛逼没有用，关键是牛逼的人说你牛逼。
- ===》 PageRank

SEO

- SEO, Search Engine Optimization
- 搜索引擎优化

Four categories of Big Data

- Fetching
- Storage
- Mining
- Processing 处理

Why

- Tremendous Data 数据量太大
- One Machine even with many CPU cores is not enough at all... 一台机器没有办法了。。。
- So? 怎么办?

Why

- 一台机器搞不定？
- 我有很多台机器！！！！

Distributed Framework

- Hadoop
- MapReduce

---We want to count all the books in the library.
You count up shelf #1, I count up shelf #2.
That's map. The more people we get, the
faster it goes. 你统计，我也来统计

---Now we get together and add our individual
counts. That's reduce 把我们的合并在一起

What is so-called key -value

- Key:学生的学号
- Value:
- 男生女生?
- 身高
- 体重
- 生日
- ○ ○ ○ ○

What is so-called key -value

- 数据结构: $\text{Hash}(\text{key})=\text{value}$
- 数据库: 主键, 自增ID
- 不同的key可以有不同的value, 也可以有相同的value。
- Key不同, 说明这两个东西就不同, 就不是同一个东西。

An example of MapReduce

- 大量的文件，每个文件包含了某天内，大量的登陆的QQ号和登陆时间
- **Task:** 如何统计这一天每个QQ号的登陆次数？

An example of MapReduce

- 596369874 2013-08-13 08:21:00
- 179863256 2013-08-13 02:00:03
- 2596369514 2013-08-13 11:05:25
- 543649874 2013-08-13 17:32:48
- 543649874 2013-08-13 18:25:37

.....

An example of MapReduce

- Map
- Shuffle
- Reduce

Map

- 输入：文本文件



Map

- 输出（key-value）
- 5110284382, 1
- 334719252, 1
- 6371062339, 1
- 3616108415, 1
- 3481058668, 1
- 5110284382 , 1
-

Shuffle

- 1, Partition
- 2, Spill
- Combine
- 5110284382, 3
- 334719252, 3
- 6371062339, 2
- 3, Merge
- 5110284382, [3,2,1]

Shuffle

- 输入：内存里的key-value
- 输出：磁盘文件

Reduce

- 输入：Shuffle生成的磁盘文件
 - 1，从哪里拿？
 - 2，什么时候拿？
- 输出：磁盘文件

Advantages

- automatic parallelization

自动并行化

- load balancing

负载均衡

- network and disk transfer optimizations

网络和磁盘传输的优化

- robustness

鲁棒/容错

Jeff Dean



Astounding 'Facts' About Google's Most Badass Engineer, Jeff Dean

- 1, When Jeff Dean designs software, he first codes the binary and then writes the source as documentation
- 2, Compilers don't warn Jeff Dean. Jeff Dean warns compilers.
- 3, Knuth mailed a copy of The Art of Computer Programming to Google. Jeff Dean autographed it and mailed it back

Astounding 'Facts' About Google's Most Badass Engineer, Jeff Dean

- 4, When Jeff has trouble sleeping, he MapReduces sheep
- 5, Jeff Dean builds his code before committing it, but only to check for compiler and linker bugs

Astounding 'Facts' About Google's Most Badass Engineer, Jeff Dean

- 7, Jeff Dean sorts his phone contacts by their vcard's md5 checksums
- 8, You name three pointers, Einstein, Euler, and Turing, when you de-reference them, all you get is Jeff Dean
- 9, Unsatisfied with constant time, Jeff Dean created the world's first $O(1/n)$ algorithm

Mahout

- Java-based and Hadoop-based package
基于java和Hadoop
- Implementation of many dm/ml/rs algorithms
实现了数据挖掘、机器学习、推荐系统算法
- Easy use and very powerful
非常强大，容易上手

But....

- When your model and algorithm can not work
- When you want to optimize your model
- When you want to use some models that has not been implemented in mahout yet
- When you want to design a new model by yourself....

Four categories of Big Data

- Fetching
- **Storage 存储**
- Mining
- Processing

NoSQL

- NoSQL means Not Only SQL
- What happened here...

Mysql

- Id 姓名 年龄 班级 身高 体重 籍贯
- 1 张三 20 1 173 60 泉州
-
- Id----key
- Others---value

Mysql

- Schema-based

Hard to scale horizontally

- Join and transactions

Unnecessary sometimes

- Traditional applications

Poor performance on paralleling

- Disk-based

Bad performance on time-consuming(1s or so)

MongoDB

- DataBase(数据库) VS DataBase(数据库)
- Collection(集合) VS Table(表)

Schemaless

- Document(文档) VS Record(记录/行)

Key-value

MongoDB is dead?

- Drawbacks

Takes a lot of memory / memory-cost heavily

Poor performance on LOCK

Many bugs exist in driver can result in data lost

Redis

- Key-value 内存存储系统。没有表的概念
- Value: 非常丰富，可以是：
- String
- List
- Set
- HashMap

Redis---hashmap

```
insert("张三", //key
{
  '年龄': '23',
  '电话': '13888888888'
  '身高': '173'
  '年龄': '25',
  '籍贯': '福建泉州'
})
```

Redis---hashmap

```
insert("李四", //key
{
'出生年月' : '1990-09-22',
'电话' : '13999999999'
'身高' : '181'
'年龄' : '25',
'政治面貌' : '共产党员'
})
```

Redis---set

- insert(“泉州师院计算机08一班学号” //key
{
“080308001”,
“080308002”,
“080308003”,
.....
“080308050”,
}

Redis---set

- 新浪微博：求杨幂和刘诗诗的共同关注

Redis---set

- insert(“杨幂” , {11,678,9923012.....364})
- key value
- insert(“刘诗诗” , {889,678,9923.....11})
- key value
- 共同关注？

Redis---set

- insert(“杨幂” , {11,678,9923012.....364})
- key value
- insert(“刘诗诗” , {889,678,9923.....11})
- key value
- 共同关注？
- Sunion

Redis

- 1, 效率高, 毫秒级别
- 2, 数据结构非常丰富
- 3, 支持java、C++、python等语言
- 4, master/slave replication
- 5, disk persistence
- 6, Publish/Subscribe

What should we do?

- 1, 怎么更快处理数据?
- 2, 怎么利用数据?

Note that...

很多时候，重要的不是算法牛逼，而是数据给力：数据多、全、准

研究现状

- 医疗，教育，金融，互联网.....
- 相关牛人：（新浪微博）

@南大周志华

@余凯_西二旗民工

@老师木

@王斌_ICTIR

@邓侃

@xlvector_Hulu

就业形势

- Google, facebook, 阿里巴巴(Alibaba), 百度(Baidu), 腾讯(Tencent), 360, 豆瓣, 微策略, 网易, 新浪微博, [hulu](#), 世纪佳缘, 美团网, 去哪儿网, 大众点评网.....
- 高校/研究所
- 银行/医疗机构
- 需要科学家, 也需要工程师
- When it comes to Fujian.....

就业形势

GZ老杨同志V: 广州鑫亚集团股份有限公司董事长张长德在老杨读书汇分享大数据在实际中的应用

收起 | 查看大图 | 向左转 | 向右转



今天 16:40 来自三星android智能手机

转发(1) | 收藏 | 评论(1)

就业形势

[首页](#)[找职位](#)[找猎头](#)[找企业](#)

推荐系统算法工程师

招聘企业：汇杰 ?

汇报对象：部门经理

下属人数：0人

年薪：30-55万 | 免费应聘

所属行业：互联网/移动互联网/电子商务

企业性质：私营·民营企业

工作地点：杭州

所属部门：交易平台

企业规模：5000-10000人

发布日期：2013-05-31

岗位职责：

1. 参与淘宝重要推荐场景算法的研发和优化
2. 参与基础数据和算法的研发和优化
3. 跟踪互联网领域相关算法进展和发展趋势

就业形势

- 年龄： 30岁以下
- 职级： T9-T10
- 名额： 9人
- 部门： IDL
- 薪酬： 100万RMB+

大数据？创业不？

- 1，耐得住寂寞
- 搞得到数据
- 2，不求大而全
- 只做小而精
- 3，处处可用大数据技术

Suggestions

- 1, 数学

高等数学

线性代数, 概率论

变分法、高等代数、矩阵论

- 2, 算法和数据结构

动态规划, 贪心

K-d Tree, Treap, AVL Tree, Hash.....

An application of Algorithm

- How to implement this ?



您要找的是不是: [hello](#)

《Hello Baby》综艺 (更新至2013-03-15) 高清在线观看_百度视频



KBS

简介: 韩国一档育儿节目, 请当红明星来养育小孩子. 第一季为少女时代; 第二季为SHINee; 第三季为T-ARA; 第四季为利特(SUJU)和Sistar; 第五季为MBLAQ; 第六季为B1A4; 第七季为Boyerien...

[查看全部视频](#)

Hello Baby最新视频:

风行网

百度视频整合优质资源



[Baby远行日本寻...](#)



[爸爸们肤质大公开](#)



[光旻荣旻变身人体...](#)



[小鸟叔来袭, 展现...](#)

video.baidu.com/ 2013-6-27

[hello_百度百科](#)

impl

- 用户输入的词记为 w （比如上图中的hrlllo）

impl

- 用户输入的词记为 w （比如上图中的hrllo）
- 用户想输入的词实际上是 c （对应hello）

impl

- 用户输入的词记为 w （比如上图中的hrllo）
- 用户想输入的词实际上是 c （对应hello）
- 选择 c 使得 $P(c|w)$ 最大

impl

- 用户输入的词记为 w （比如上图中的hrlllo）
- 用户想输入的词实际上是 c （对应hello）
- 选择 c 使得 $P(c|w)$ 最大
- $P(c|w) = P(c) * P(w|c) / P(w)$

impl

- 用户输入的词记为 w （比如上图中的hrllo）
- 用户想输入的词实际上是 c （对应hello）
- 选择 c 使得 $P(c|w)$ 最大
- $P(c|w) = P(c) * P(w|c) / P(w)$
- $P(c) * P(w|c)$

impl

- 用户输入的词记为 w （比如上图中的hrllo）
- 用户想输入的词实际上是 c （对应hello）
- 选择 c 使得 $P(c|w)$ 最大
- $P(c|w) = P(c) * P(w|c) / P(w)$
- $P(c) * P(w|c)$
- 如何计算 $P(c)$ 和 $P(w|c)$ ？

Suggestions

- 3, 计算机编程高级语言
- C/C++/Java
- 4, 脚本语言
- Ruby/Perl/Python
- 5, 英语
- Reading/Writing/Speaking

Step by Step

- 1, 看点数据挖掘的书
- 2, 看点Hadoop的书
- 3, 采集数据
- 4, 我就是搞大数据的~~~不服来辩

References

- 李航
《统计学习方法》，清华大学出版社

References

- 韩家炜
《数据挖掘：概念与技术》，机械工业出版社

References

- 《Data Mining and Analysis:
Fundamental Concepts and Algorithms》

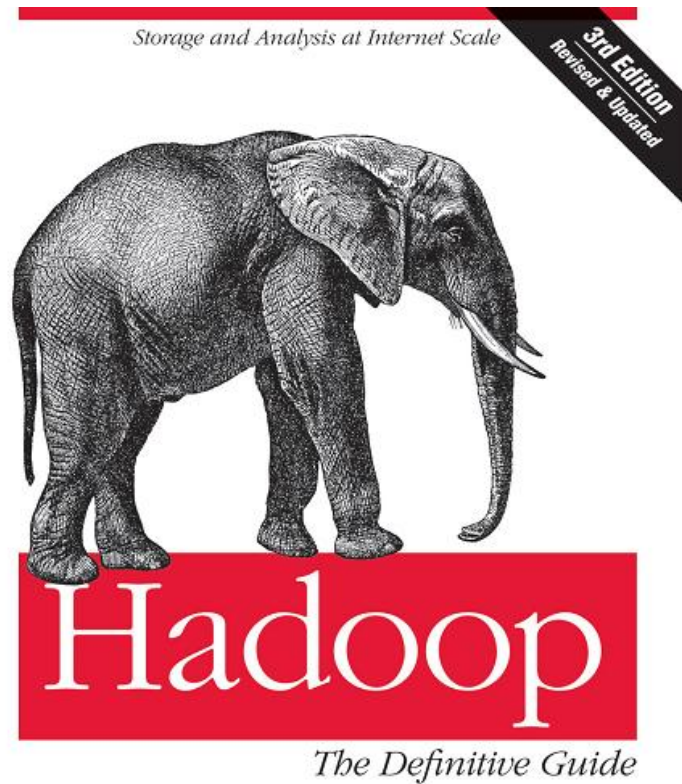
Mohammed J. Zaki

Wagner Meira Jr

References

- Dietmar Jannach
Markus Zanker
Alexander Felfernig
Gerhard Friedrich
《推荐系统》，人民邮电出版社

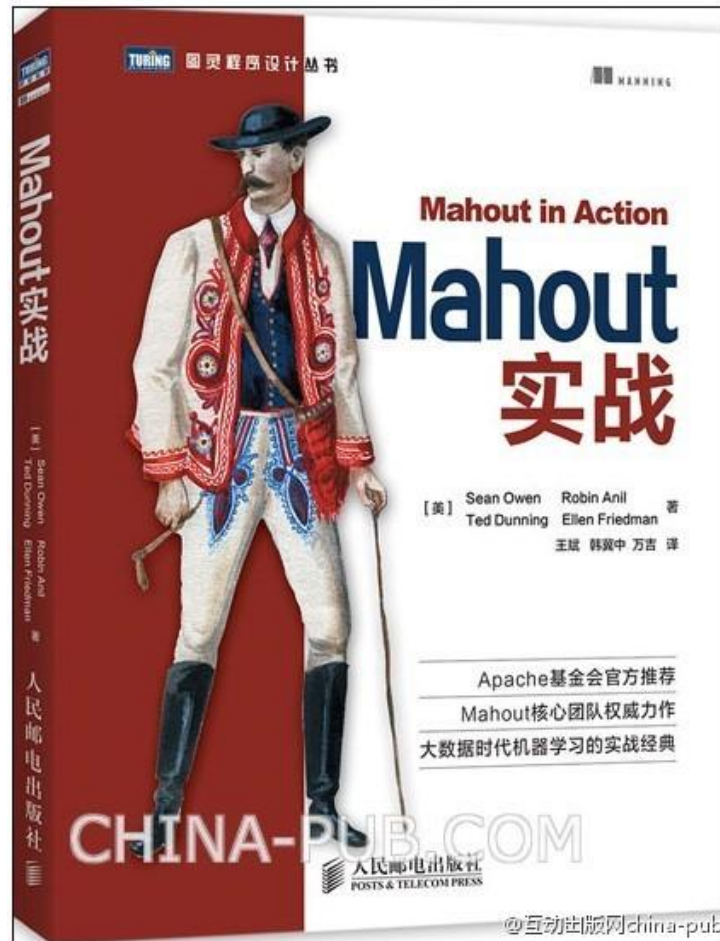
References



O'REILLY®

Tom White

Mahout in action



休闲读物



小结

- 1, 什么是大数据?
- 2, 应用和场景
- 3, ml, dm, rs, sn, nlp, se

Hadoop

NoSQL

- 4, 建议
- 5, 参考书

Acknowledgments

Special thanks go to :

- Zeng Taisheng at qztc for the arrangement
- Wang Jie at ICT,CAS for reviewing the slides
- Wang Chonghua at IIE,CAS for tips for making the title
- Meng Xingquan at qztc for advice

Q&A

- Thanks A Lot
- Any questions?