

www.qconferences.com

www.qconbeijing.com



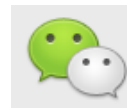
QCon北京2014大会 4月17—19日

伦敦 | 北京 | 东京 | 纽约 | 圣保罗 | 上海 | 旧金山

London · Beijing · Tokyo · New York · Sao Paulo · Shanghai · San Francisco

QCon全球软件开发大会

International Software Development Conference



软件
正在改变世界!

360 推荐技术那些事

杨浩 (yanghao@360.cn)

20131012

个人简介

• 行业背景

- 面向服务计算（博士）
- 模式识别（实习生）
- 搜索（大搜索Query分析）
- 推荐（视频、导航、搜索onebox）

BUPT

MSRA

BAIDU

360

• 技术背景

- C++
- Python/Perl
- SOA
- Machine Learning
- Linux



OUTLINES

- 为什么要做推荐
 - 更高导流，更高收益？
- 推荐现状
 - 无视需求，无视场景，无视质量，无视反馈的“强行推荐”
- 如何做推荐
 - “用户行为 + 领域知识 + 微创新”
- 一些推荐实践
 - 推荐指标、产品设计、技术实现、架构支持
- 推荐后续规划
 - 标签推荐、Social推荐
- Q&A

1、为什么要做推荐

搜索 VS 推荐

搜索 (引擎)

- 赚钱
- 强需求 (明确需求)
- 覆盖率低
- 点击率高
- 导流高

推荐 (引擎)

- 能不赚钱吗?
- 弱需求 (非明确需求)
- 覆盖率高
- 点击率低
- 导流更高

业界实例:



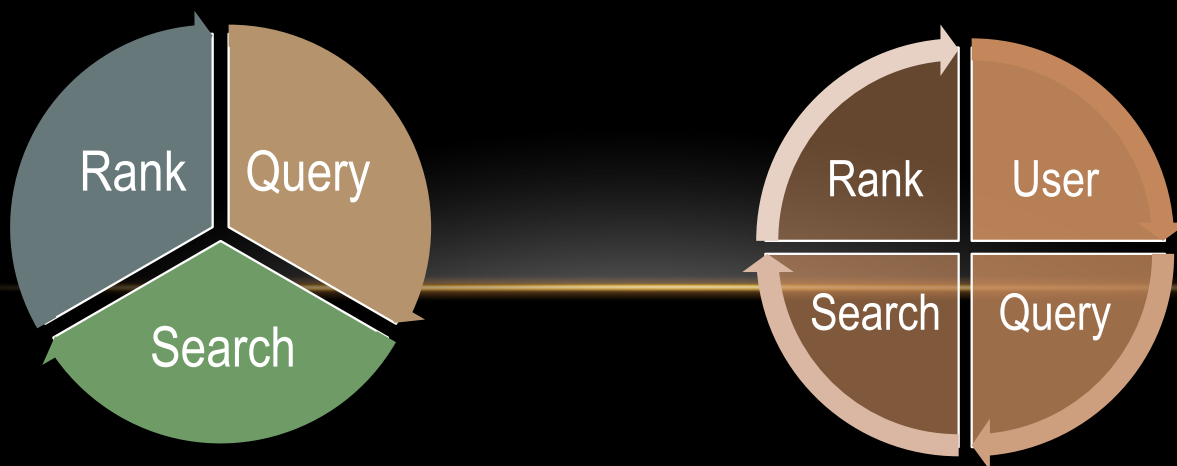
35% of sales result from recommendations



75% of Netflix views result from recommendations

搜索VS推荐

- 搜索 **垫基石**，推荐 **任鸟飞**
 - 搜索不好，用户第一感觉很差，推荐无从谈起
- 搜索通用，推荐专业、垂直
- 搜索流程：Query – Search – Rank
- 推荐流程：User – Query – Search – Rank
 - 可以结合用户，分析不同的深层需求 (Query)，进行多种类型推荐
 - Search 和 Rank 可以借鉴搜索技术



2、推荐现状

貌似相关下的“推荐轰炸”

- 相关：
 - 用户行为和推送内容具有一定相关性
- 推送轰炸：
 - 无视时机
 - 无视感受
 - 无视产品质量
 - 无视用户反馈



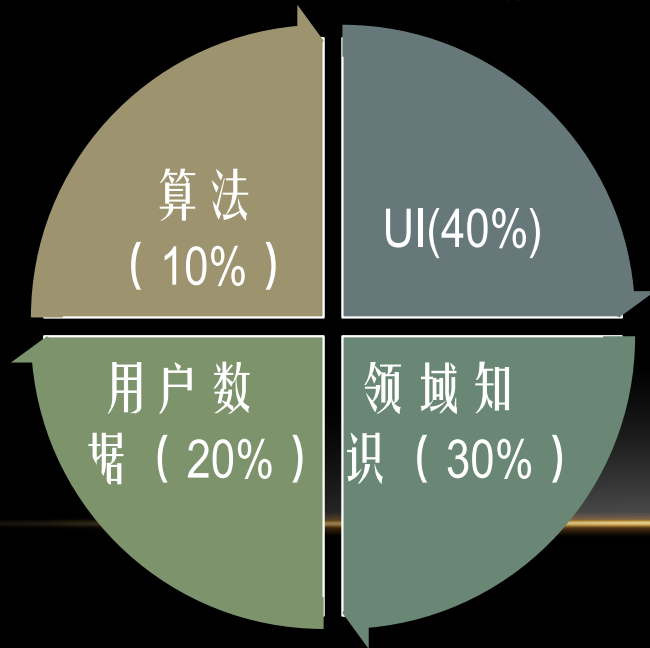
好的推荐是什么？

- 相关的推荐：
 - 用户行为和推荐内容一定要有一定的相关性
- 把握住时机的推荐：
 - 解决时效性，什么时候、什么地点、什么上下文进行推荐
- 经得起诱惑的推荐：
 - 商业利益和用户需求的均衡
- 可理解的推荐：
 - 高质量的推荐原因挖掘
- 及时反馈的推荐：
 - 及时响应用户反馈，更多可能是从产品而非技术的方法解决

3、如何做推荐

如何做推荐

- 一个好的产品（张栋 – RecSys 2010）
 - UI，离用户近，直接影响用户
 - 领域知识，是否真的懂
 - 用户数据，是否真的做过
 - 算法，是否比别人好，黏住用户



推荐用户需求

- 推荐用户需求分析
 - 基于用户Session
 - 实际数据说话
 - 结合领域（专家）知识
 - 半自动
 - 微创新
 - 产品
 - 技术

推荐用户需求

- 基于Session的用户需求分析
 - 频道偏爱需求
 - 基因偏爱需求
 - 追剧需求
 - 推荐原因可理解
 - 频道切换
 - 基因切换
 - ...

推荐用户需求

- UI
 - 频道偏好、切换
 - 基因偏好及切换
- 领域知识
 - (电视)追剧
- 数据、算法
 - 推荐Item生成
 - 推荐原因挖掘



3、一些推荐实践

推荐指标

- 目标
 - 导流
 - 覆盖率
 - CTR
 - ABtest

推荐总导流 (DATE: 2012-11-04)

PV	Click	CTR
305347	32040	10.5%

详情

首页&频道页 统计 (DATE: 2012-11-04)

电影

	PV	Item 点击	CTR	- 换一换点击	历史影片点击	左右切换点击
Cookie	3907042	645898	16.53%	- 336704	35750	481824
Mid	66053	7816	11.83%	-	-	16655
总数	3973095	653714	16.45%	- 336704	0.92%	498479

电影详情页 (Item2Item) (DATE: 2012-11-04)

TotalPV	TotalClick	TotalCTR	A 方案 PV	A 方案点击	A 方案 CTR	B 方案 PV	B 方案点击	B 方案 CTR
10323916	153636	14.88%	928969	137802	14.83%	1034226	158341	15.31%

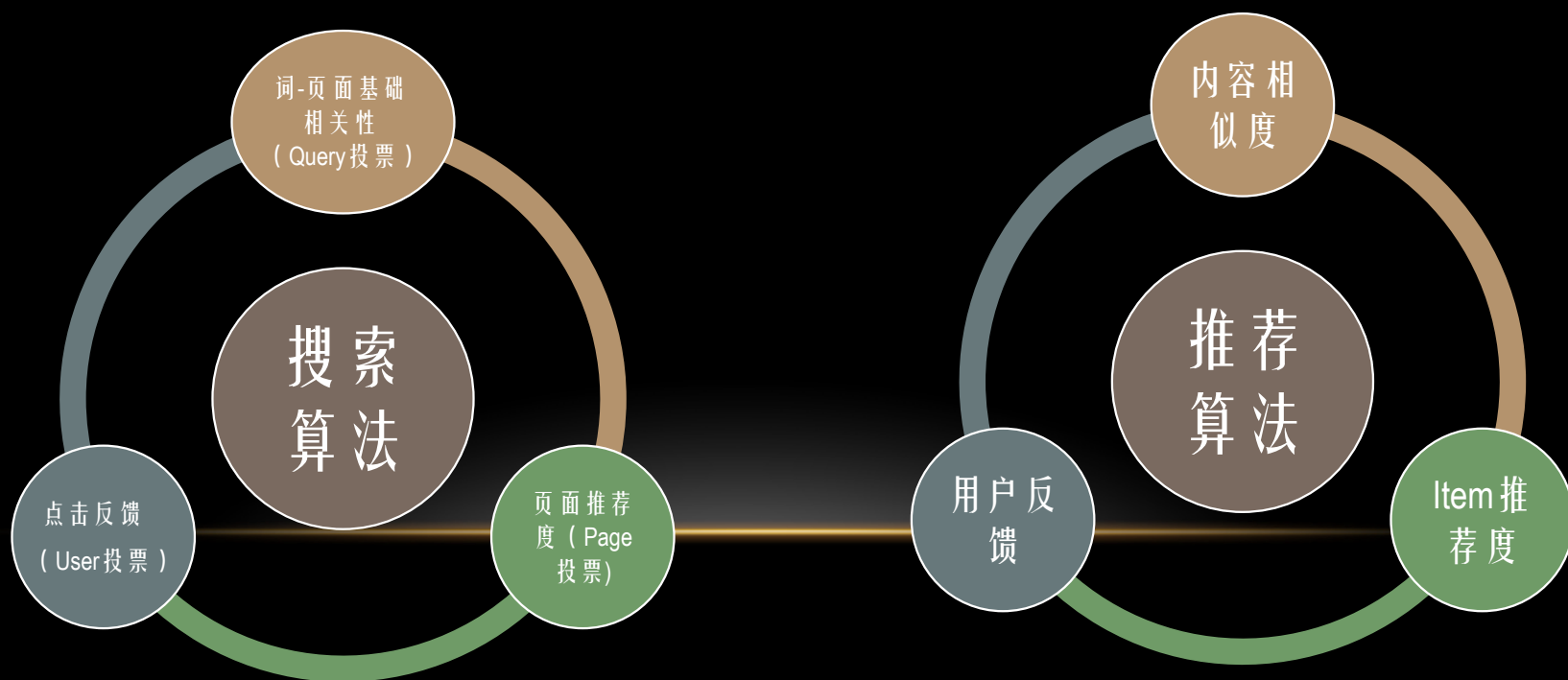
产品设计（UI & 领域知识）

- 需求满足
 - 类型、基因偏好
 - 追剧
 - 切换等



技术实现（数据 & 算法）

- 他山之石可以攻玉
 - 借鉴搜索Rank思路
 - 基于推荐算法
 - 结合自身数据微创新



技术实现（数据&算法）

- 算法对比(OFFLINE)
 - Recommendation handbook, Francesco Ricci, 2011
 - Top-rated: recommend a list of hottest

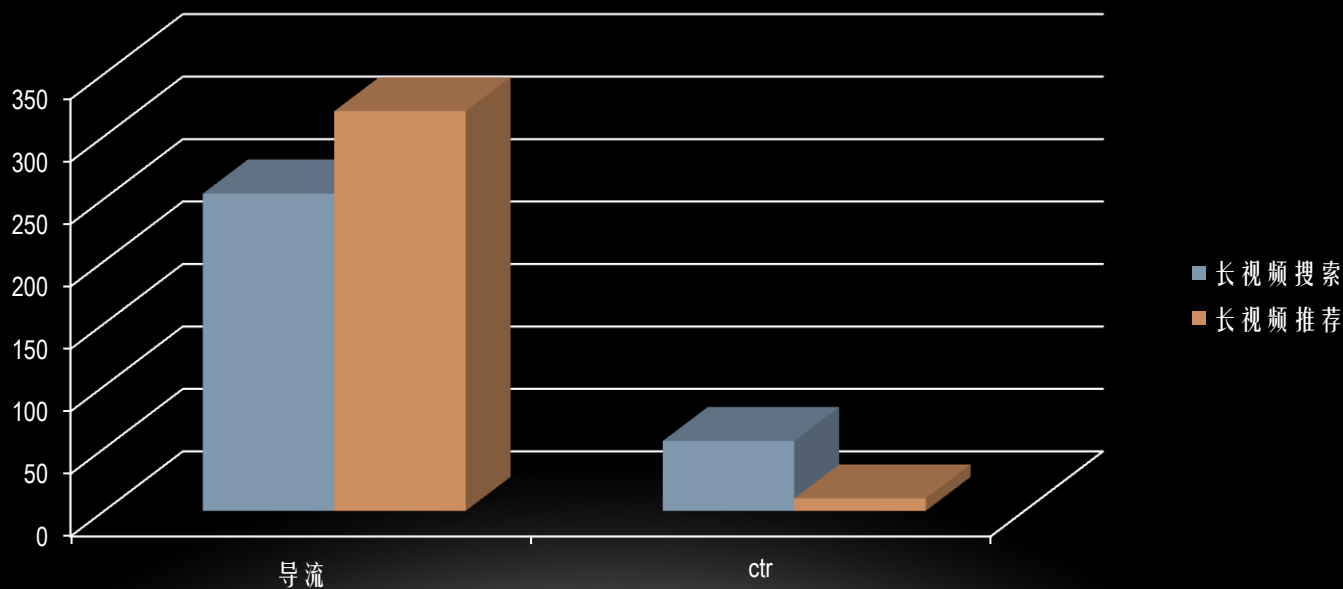
Algorithm	Parameter	Recall	
		3 months	6 months
Item-based-CF	$k = 10$	16.8%	14.9%
	$k = 50$	18.7%	16.4%
	$k = 100$	19.0%	16.6%
	$k = 150$	18.8%	16.5%
SVD-CF	$l = 5$	15.1%	12.7%
	$l = 15$	12.6%	13.3%
	$l = 25$	10.9%	11.5%
	$l = 50$	9.3%	9.9%
	$l = 100$	6.3%	8.0%
LSA-CB	$l = 50$	1.9%	1.7%
	$l = 100$	2.3%	2.3%
	$l = 150$	2.4%	2.4%
	$l = 200$	2.5%	2.5%
Top-rated		12.2%	7.7%

技术实现（算法）

- 微创新
 - 推荐时效性挖掘：
 - 及时反馈、爆发时效性、周期时效性
 - 推荐原因挖掘：基于随机行走模型的视频标签转移
 - 标签会在相似电影间进行传递，最终达到平衡
 - $$\text{NEW_VIDEO_TAG} = \text{VIDEO_TAG} * a + \text{TRANS_VIDEO_TAG} * (1 - a)$$

指标评估

- 推荐算法（效果）
 - CTR：搜索大于推荐
 - 导流：**推荐大于搜索**（某些领域不一定适用，如电商）



4、后续推荐规划

后续规划

- 推荐后续规划
 - 技术：
 - 标签推荐：深挖推荐原因 & 用户标签
 - Social推荐：朋友、名人推荐
 - 产品：搜索、导航、app等等

毛泽东

Q 360推荐: 沁园春 满江红 水调歌头 十六字令 毛泽东选集 换一换

[百科](#) [毛泽东](#) [360百科](#)

简介: 毛泽东 (1893年12月26日-1976年9月9日), 字润之 (原作咏芝, 后改润芝); 笔名: 子任、二十八画生 ("二十八画"即"毛泽东"这三个字的繁体总共的笔画数, 国际...
[逝世37周年](#) - [个人简介](#) - [邓小平评价](#) - [生平经历](#) - [查看全部>>](#)
[baike.so.com/doc/1184930.html](#) 2013-10-27

毛泽东 360图片 - 举报图片

image.so.com 2013-10-27 相关推荐: 毛泽东遗体 毛泽东的妻子 韶山毛泽东故居

毛泽东 (约45900个相关视频) 高清在线观看 360视频搜索

历任中国共产党主要领导人

				
陈独秀	向忠发	王明	博古	华国锋

历任中华人民共和国主席

				
李先念	杨尚昆	江泽民	胡锦涛	习近平

历任全国政协主席

				
周恩来	邓小平	邓颖超	李先念	李瑞环

Q&A

特别感谢 QCon上海合作伙伴

