GOME

机器学习排序实践

邱宝军 杨骥 李冰青

国美在线大数据中心

国美大数据研究院

2016年7月1日

目录

排序

推荐

vs.广告

vs.搜索

排序→机器学习排序

搜索GITC

GITC全球互联网技术大会 **Perfect**

What is GITC？GITC全球互联网技术大会以汇集行业精英、促进技术交流、加深商务合作以及推动行业发展为大会宗旨。通过针对互联网技术领域专门设计的先进开放会议会 ...

www.thegitc.com

全球互联网技术大会_百度百科 **Excellent**

GITC（全球互联网技术大会）项目于2013年5月正式发起，通过针对互联网技术领域专门设计的先进开放会议会展内容及形式，联合iTech Club（中国互联网技术

baike.baidu.com/view/10900771 2016-5-31

GITC 2014新闻发布会暨开票仪式在京召开|GITC_滚动新闻 ... **Good**

2014年8月6日"GITC 2014全球互联网技术大会新闻发布会暨开票仪式"于北京五洲皇冠国际酒店隆重召开。作为GITC 2014全球互联网技术大会筹备召开的正式启动 ...

tech.sina.com.cn › 新浪科技 › 滚动新闻 › 2014GIT... 2014-9-30

GITC（团体儿童智力测验）_百度百科 **Fair**

中文名 团体儿童智力测验 外文名 GITC 优 点 客观、精练、易操作、经济快速等 主要特点 适用于团体快速施行，非语言量表

baike.baidu.com/subview/11415563/14249887.htm 2016-1-17

Greater Illinois Title Company 翻译此页 **Bad**

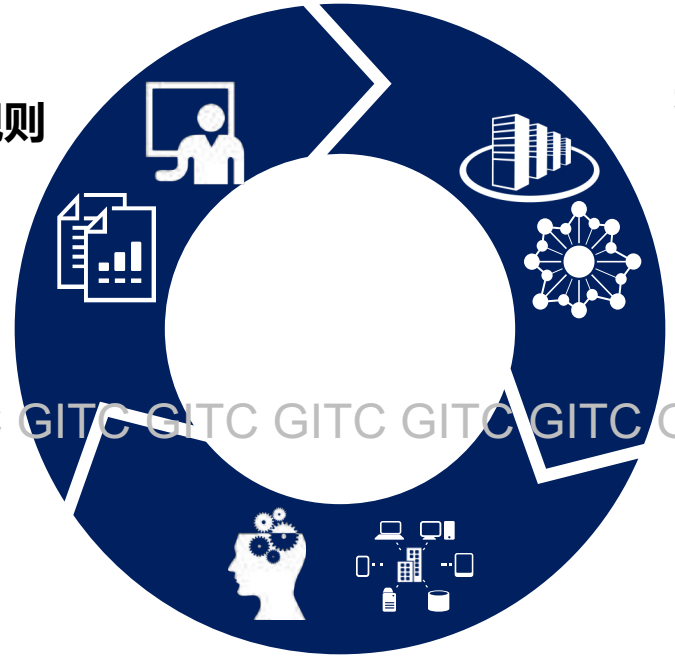
As the largest title agency in Illinois, Greater Illinois Title Company provides a single point of contact for all title and closing

搜索词、目标文档(网站)、搜索人、时间、环境...

Query	URL	Label	Page-Rank	BM25	...	Featu-re_N
GITC	Thegitc.com	Perfect	0.6	0.9		100
GITC	...	Excellent	0.9	0.9		66
GITC	...	Good	0.8	0.7		75
GITC	...	Fair	0.9	0.6		32
GITC	...	Bad	0.2	0.65		120
国美	...	Perfect	0.6	0.8		150
国美	...	Perfect	0.5	0.7		50
...						

人工 vs 机器学习排序

人工规则



人工规则 vs. 机器学习

（广告 vs. 搜索 vs. 推荐）

机器学习排序(Learning to Rank)

- ✓ Pointwise
  排序问题转化为多类分类问题或者回归问题，相关度相同的为一类
- ✓ Pairwise
  排序问题转化为二分类问题，由两两之间的偏序关系得到全局的排序
- ✓ Listwise
  直接对排序结果进行优化

机器学习排序-流程

训练数据

<Query, Doc> → 标注

A B C D E F G

	Label
B	+
E	+
A	-
C	-
D	-

	Label
B-A	+
B-C	+
B-D	+
C-E	-
D-E	-
E-A	+

特征工程

模型训练
（&离线测试）

线上部署
（&A/B测试）

特征工程-从实体到特征

用户属性、
购买力、
品类倾向

查询量、
转化率、
品类分

UV、PV、
订单量、
转化率

店铺评分、
店铺流量、
店铺评价

当地温度、
污染指数、
人口数量

……

用户　查询　商品　商家　环境

关键词匹配、价格匹配等

区域销量、商品天气匹配等

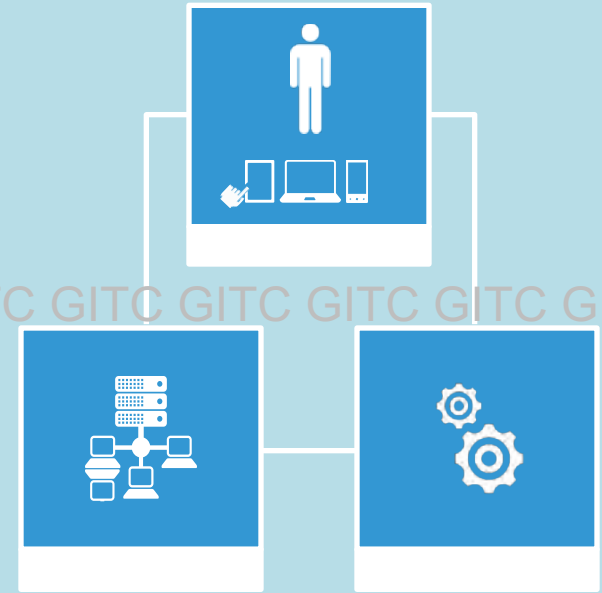
品类匹配、购买力匹配等

常用模型-（推荐系统常用算法）

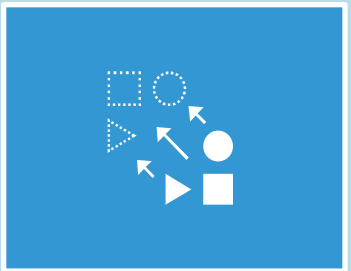
- 协同过滤
- 回归模型（LR,SVR...）
- 矩阵分解（SVD,LFM,SVD++,...）
- 马尔科夫链
- 聚类 (K-means, ...)
- Deep Learning（RBM, CNN, RNN, ...）
- LDA
- Association Rules
- GBDT/RF
- ...

机器学习排序-部署效率

离线预处理



在线处理



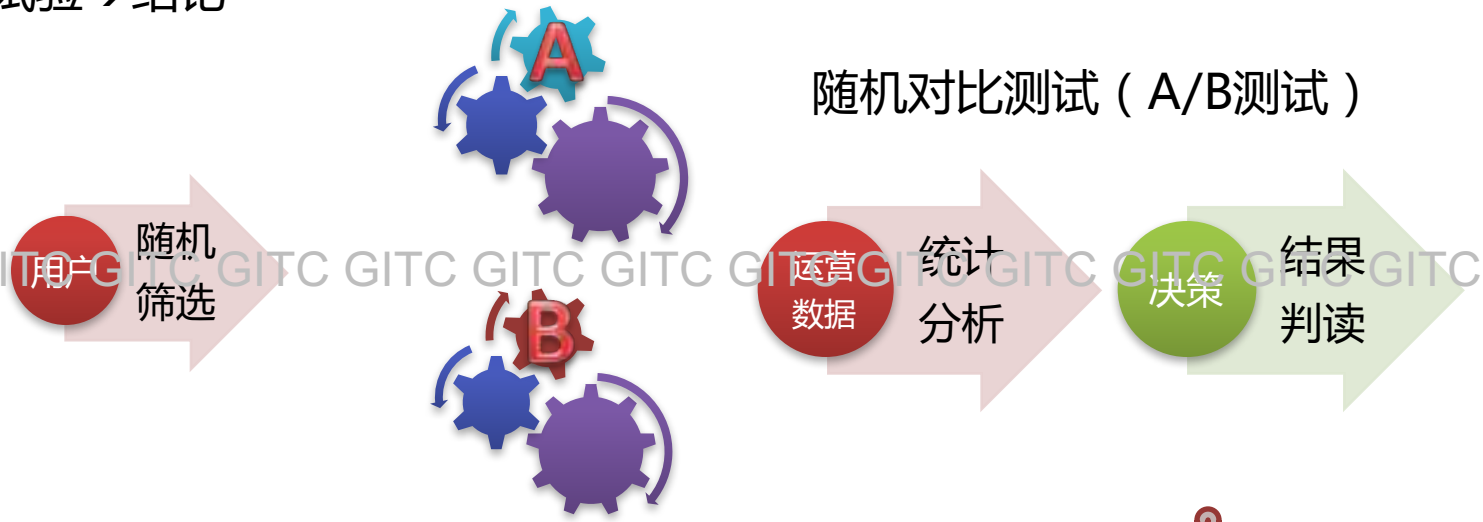
精确度换速度

空间换速度

量化与A/B测试

- 可量化目标：分类准确率、NDCG、点击率、转化率
- 假设→试验→结论

随机对比测试（A/B测试）



德鲁克：如果你不能衡量它，就无法增长它

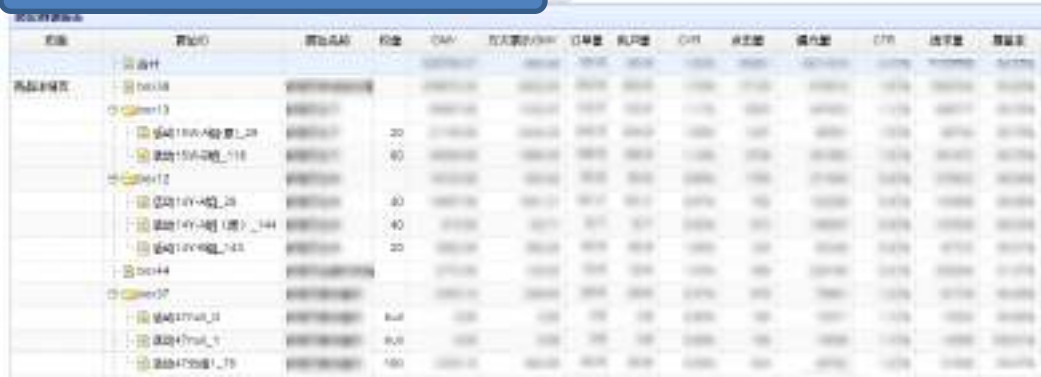
国美推荐A/B测试配置及量化

第一步：更改A组权重

第二步：更改B组配置及流量

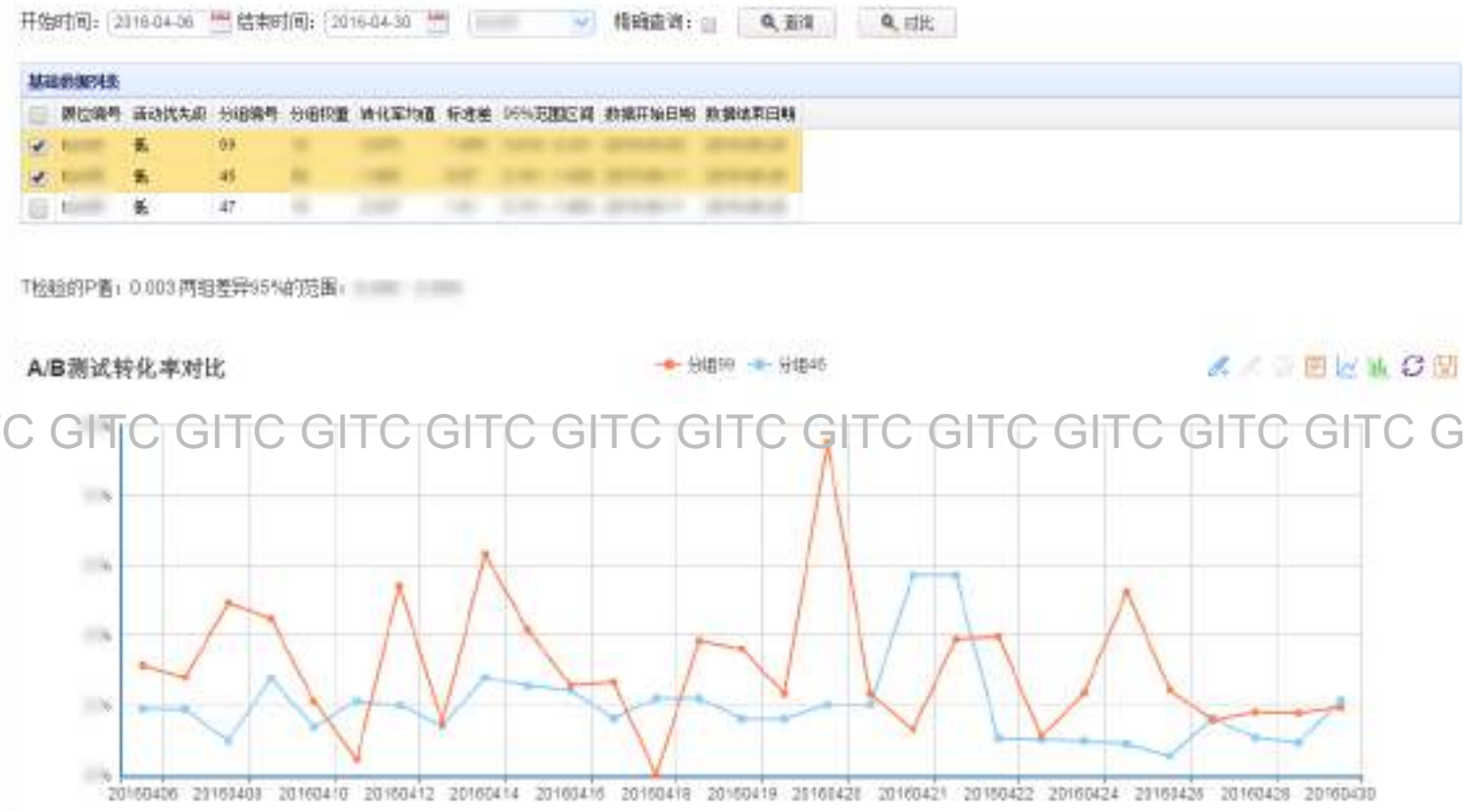
第三步：完成分组流量配置及分组备注

持续观察KPI指标

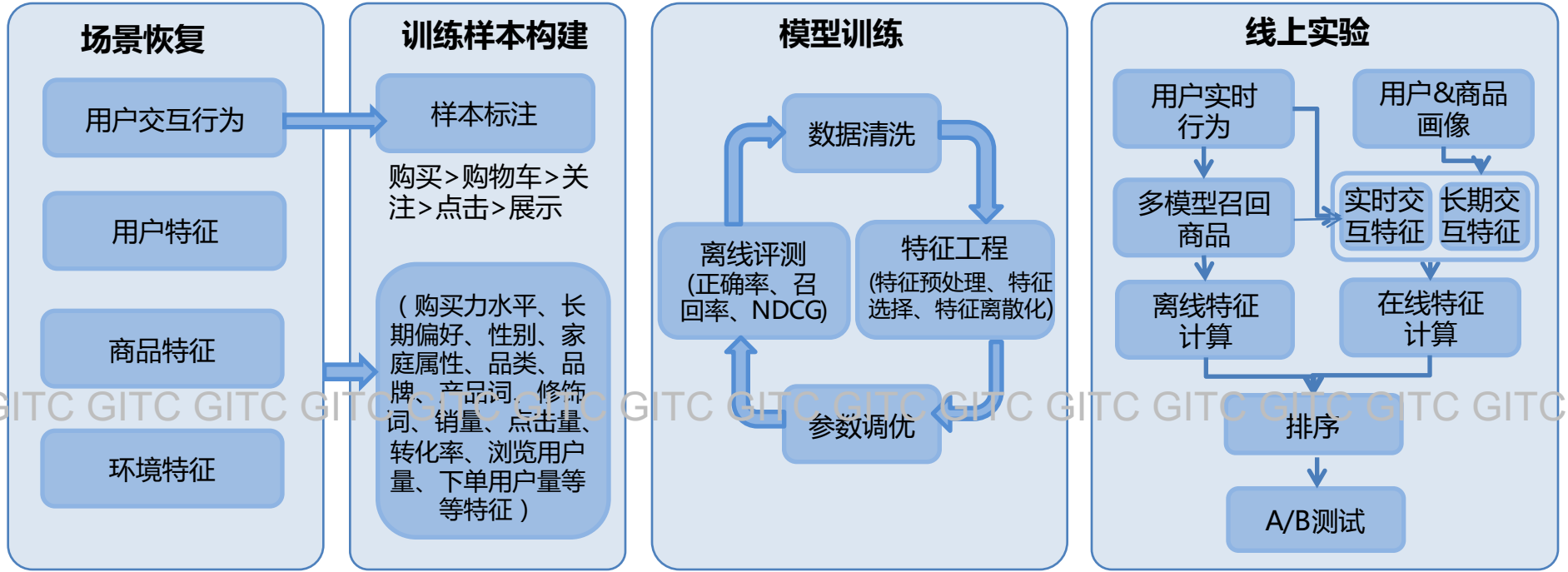


国美推荐A/B测试量化效果

对比A/B测试效果

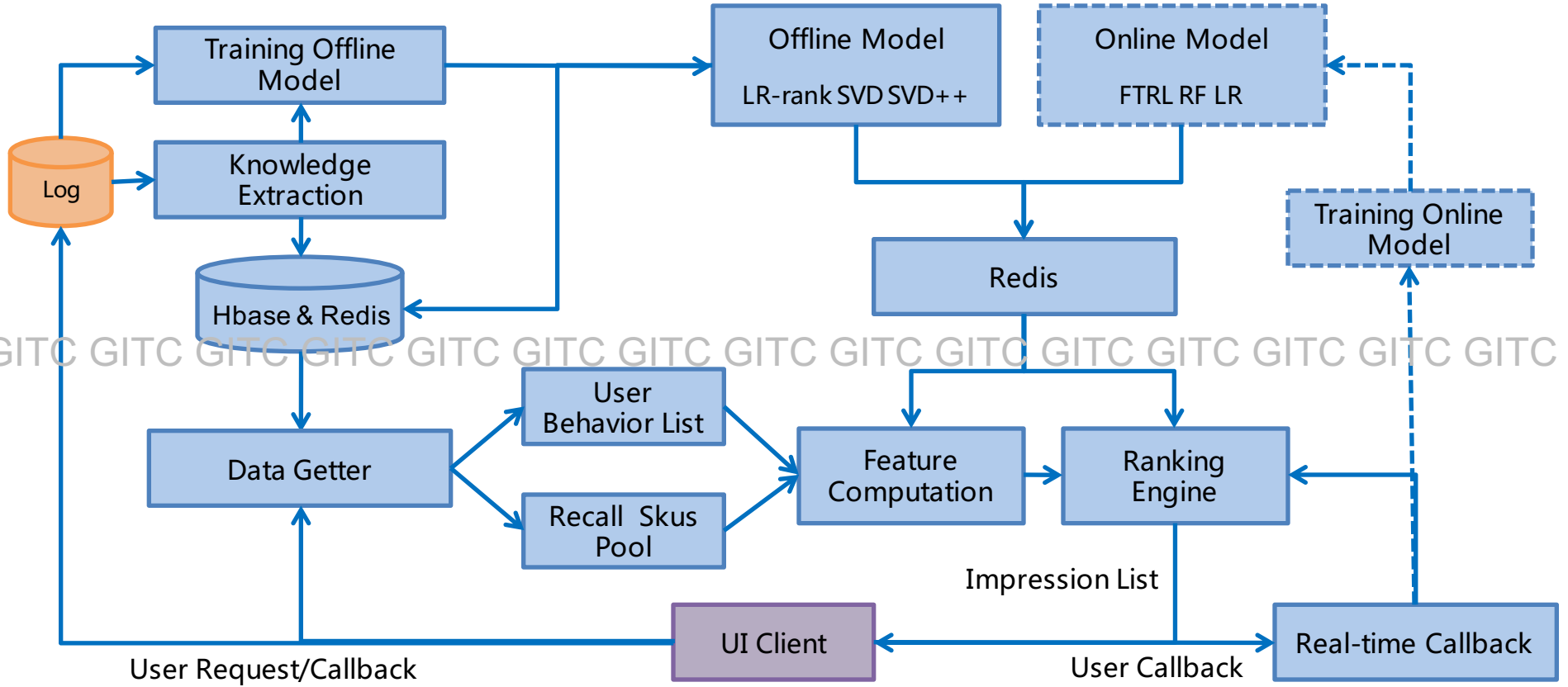


机器学习排序-小结



目录

推荐排序实践：架构



推荐排序实践：在线vs离线

在线实时排序

- 过滤： 业务需求（无库存、成人用品、促销商品）
- 实时特征计算：人vs.商品，准实时行为特征
- 多样性：多模型、多数据源融合
- 新颖度
- CTR/CVR

- 在线A/B测试

离线数据挖掘

- 隐语义模型数据： 矩阵分解
- 协同过滤：UserBasedCF、ItermBasedCF
- 用户画像：品类、品牌、性别、年龄、购买力等
- 商品画像：价格指数、适用人群和地理区域、购买周期等
- 计算机视觉特征

- 离线测试

推荐排序实践：推荐离线模型的构建流程

1. 收集用户行为，包括点击、加购、关注、下单等

2. 对行为进行过滤，比如：join(白名单)、统计截断、position-bias、多次加/删购等处理

3. 制定行为评分规则，生成评分矩阵

4. 训练矩阵分解模型

5. 导入缓存

推荐排序实践：基于SVD的评分预测

- 用户因子向量 $p_u \in \Re^f$ 和商品因子向量 $q_v \in \Re^f$
- 评分偏置(bias) $b_{uv} = \mu + b_u + b_v$ （μ 是整个网站评分的均值，b_u和b_v是用户/商品较网站评分的偏差）
- 用户对某商品v的预测评分为 $r'_{uv} = b_{uv} + p_u^T q_v$

- SVD++ 加入隐式反馈

$$r'_{uv} = b_{uv} + q_v^T \left(|R(u)|^{-\frac{1}{2}} \sum_{j \in R(u)} (r_{uj} - b_{uj}) x_j + |N(u)|^{-\frac{1}{2}} \sum_{j \in N(u)} y_j \right)$$

- 其中
  - $q_v, y_v, x_v \in \Re^f$ 是商品的三个隐向量
  - $R(u)$代表被用户u 打过分的商品集合
  - $N(u)$代表用户隐式选择的商品集合（即打分/未打分）

推荐排序实践：环境特征

某地高温
橙色预警

利用天气API定位某地高温橙色预警，该地区用户进入网站，首页猜你喜欢推荐空调、风扇等商品：

某地雾霾
红色预警

利用天气API定位某地雾霾红色预警，该地区用户进入网站，首页猜你喜欢推荐空气净化器、口罩等商品：

推荐排序实践：深度学习与视觉特征

图像标注

对国美全站的商品图像进行标注：

1. 最主要的工作是数据清洗，即把不能和品类对应的图片删除或者重新进行品类校准
2. 统计国美全站最近一年内各品类下商品总数的分布；然后按照分布进行图片抽样

提取特征

图片的特征分为两个部分，一是通过深度学习得到的特征，二是图像色局部特征：

利用caffe训练CNN，将倒数第二层输出作为Feature Learning的结果提取出来 **(Deep Learning + Transfer Learning)**

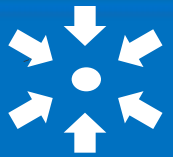
利用局部特征算子(SIFT,kaze等)提取出图像的局部特征

降维

分别对前面两种特征进行降维处理:

用积量化(Product Quantization)的方法对CNN特征进行降维

用Fisher Vector对图像局部特征进行降维



匹配

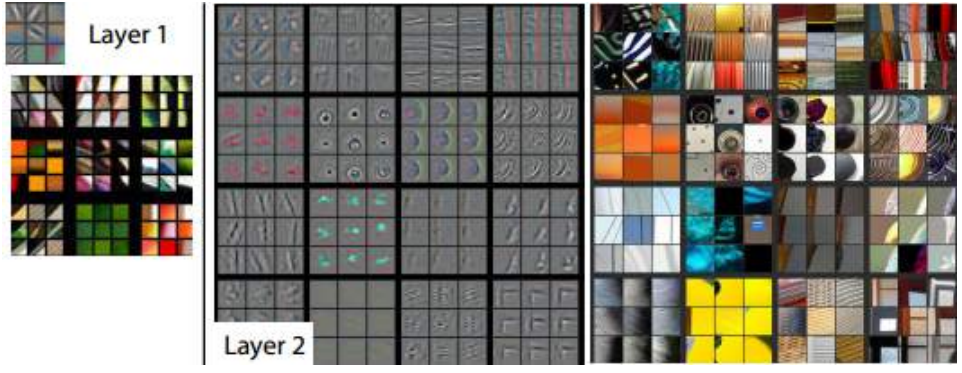
采用最近邻搜索的方法找出每一个商品的相似商品集合

深度学习特征

手工特征：

SIFT [Lowe 99]
Spin Images [Johnson&Herbert 99]
Textons[Malik et al. 99]
RIFT[Lazebnik 04]
GLOH[Mikolajczyk&Schmid 05]
HOG[Dalal&Triggs 05]

深度学习特征：



Zeiler&Furges，ECCV2014

推荐排序实践：提升推荐个性化及转化率

未经机器学习排序，网站首页猜你喜欢结果

使用机器学习提升推荐转化率

利用机器学习排序就是从数据中自动学习模式，在若干限定条件下，找出全局最优或者局部最优的近似值

利用机器学习排序后，网站首页猜你喜欢推荐商品

推荐排序实践：A/B测试监测分析

分组ID	分组名称	权重	CTR均值	CTR标准差	CTR95%范围区间	CVR均值	CVR标准差	CVR95%范围区间
29	W-A组(原)							
115	A组							
116	W-B组							
119	W-C组							

谢谢

Thanks!