

LARGE SCALE E-COMMERCE RECOMMENDER SYSTEM



YUAN QUAN / ALIBABA GROUP

Homepage Recommendation

- Personalized real-time recommendation
- Based on click/purchase behavioral



根据浏览推荐 根据购买推荐



纯一良品韩版童装 儿童纯棉V领短袖T恤 男
46.00元
淘宝网



【限时特价+送电保包】Nokia/诺基亚
138.00元
天猫



(美国GOFIT)时尚新型 锻炼器/阻力伞/健身
99.00元
淘宝网



纯一良品韩版童装 儿童纯棉螺纹五角星短袖
46.00元
淘宝网



【168元包邮 送电保包】Nokia/诺基亚
168.00元
天猫



56速度训练阻力伞 可调运动跑步伞 球健身
33.00元
淘宝网



纯一良品韩版童装 儿童纯棉薄款短袖T恤 男
38.00元
淘宝网

Personalized Best-seller List

我的排行榜



天猫国庆特价 雅丽品牌 宁夏
均价: 100元



中秋大促 新货 宁夏枸杞子 极
均价: 38元



淘金币 头茬 宁夏中宁极品
均价: 36元



雅丽新货宁夏 枸杞500g特价
均价: 50元



新货不加工 特价直销农家自
均价: 37元

Item Detail Page recommendation

天猫 TMAIL.COM

月成交记录 (1771件) 服务质量 给我推荐 售后服务

看了此商品的会员通常还看了

韩都衣舍韩国2012秋冬新品韩版女装纯色连帽长袖连衣裤
¥188.00
★★★★★ (512人评价)

韩都衣舍2012秋冬新品韩版女装纯色套头低腰长袖连衣裤
¥99.0 ¥198.00
★★★★★ (3636人评价)

韩都衣舍2012秋冬新品韩版女装纯色套头低腰长袖连衣裤
¥208.00
★★★★★ (2589人评价)

韩都衣舍韩国2012秋冬新款女装翻领纯色修身长袖连衣裤
¥131.6 ¥188.00
★★★★★ (1210人评价)

买了此商品的会员通常还买了

韩都衣舍韩国2012秋冬新品拼接显瘦修身加
¥128.00
★★★★★

韩都衣舍韩国2012秋冬新品修身显瘦包臀假
¥139.00
★★★★★

韩都衣舍2012秋冬新品韩版女装蝙蝠型长款
¥298.00
★★★★★

韩都衣舍韩国2012秋冬新品女装高腰显瘦
¥86.4 ¥108.00
★★★★★

韩都衣舍韩版女鞋2012秋款新品牛皮超
¥378.00
★★★★★

看过这本书的人也在看

如何教孩子性
格好 父母必
备
3.99元

妈妈不生气孩
子很争气: 18
年教子心经
2.80元

和儿子一起成
长
5.30元

Buy-after-View Recommendation

Relevance Recommendation

一淘 优惠购

索尼 DSC-HX30V 高清数码相机 1800万像素 20倍长焦 正品行货

新品 HX30

¥2008.00

满1000减100 下单立减100

限时抢购 省30.64元 优惠购

其他商家报价

品牌	1号店	当当网 (自营)	苏宁易购
索尼 DSC-HX30V	¥1888.00	¥1998.00	¥2049.00
差价	¥27.60元	¥89.90元	¥161.00元

价格走势

品牌: 索尼数码相机 型号: HX30 颜色: 黑色 尺寸: 112.3x67.1x39.5mm 重量: 190g 分辨率: 1800万像素 镜头: 20倍长焦 显示屏: 3英寸 触摸屏: 支持 电池: 锂电池 包装: 原厂包装

浏览此商品的用户还浏览了

购买此商品的用户还购买了

浏览此商品的用户最终购买了

一淘

商品 婴儿床



搜索"婴儿床"的用户还买了



儿童手套 婴儿防抓手套
纯棉 防止宝宝抓伤

2.00元

儿童防抓手套



四皇冠 拉比正品 绒绒
小熊方被

149.00元

儿童绒被



(包邮)婴爱婴儿床围四
件套

135.00元

婴儿床围



四省包邮笑巴喜 工字
型 豪华款婴儿摇篮

128.00元

婴儿摇篮



婴~姿飒爽 尿布纱布
尿片 竹纤维尿布婴儿

13.00元

纱布尿片



韩国毛绒公仔手机链
手机挂件/挂饰品/汽车

4.00元

毛绒玩具

Many others ...

Mobility & Context-ware



Group-buy Reco

-- user with similar intention



Ebook Reco -- Content Modeling



Alipay Post-purchase Reco -- Predict Next Best Action

2012.10.17 16:09	付款-买童车 交易号: 2012101700	 可爱的小丑鱼尼莫儿童 自行车、童车卡通车 10.00元
2012.10.15 14:57	付款-1015午餐 交易号: 2012101500	
2012.10.12 13:16	付款-转账 交易号: 2012101200	 3d木制立体拼图星空 拼图9片拼图七巧板 智 24.75元
2012.10.09	付款-10.9午饭	

E-Commerce Data Model

- Behavioral data

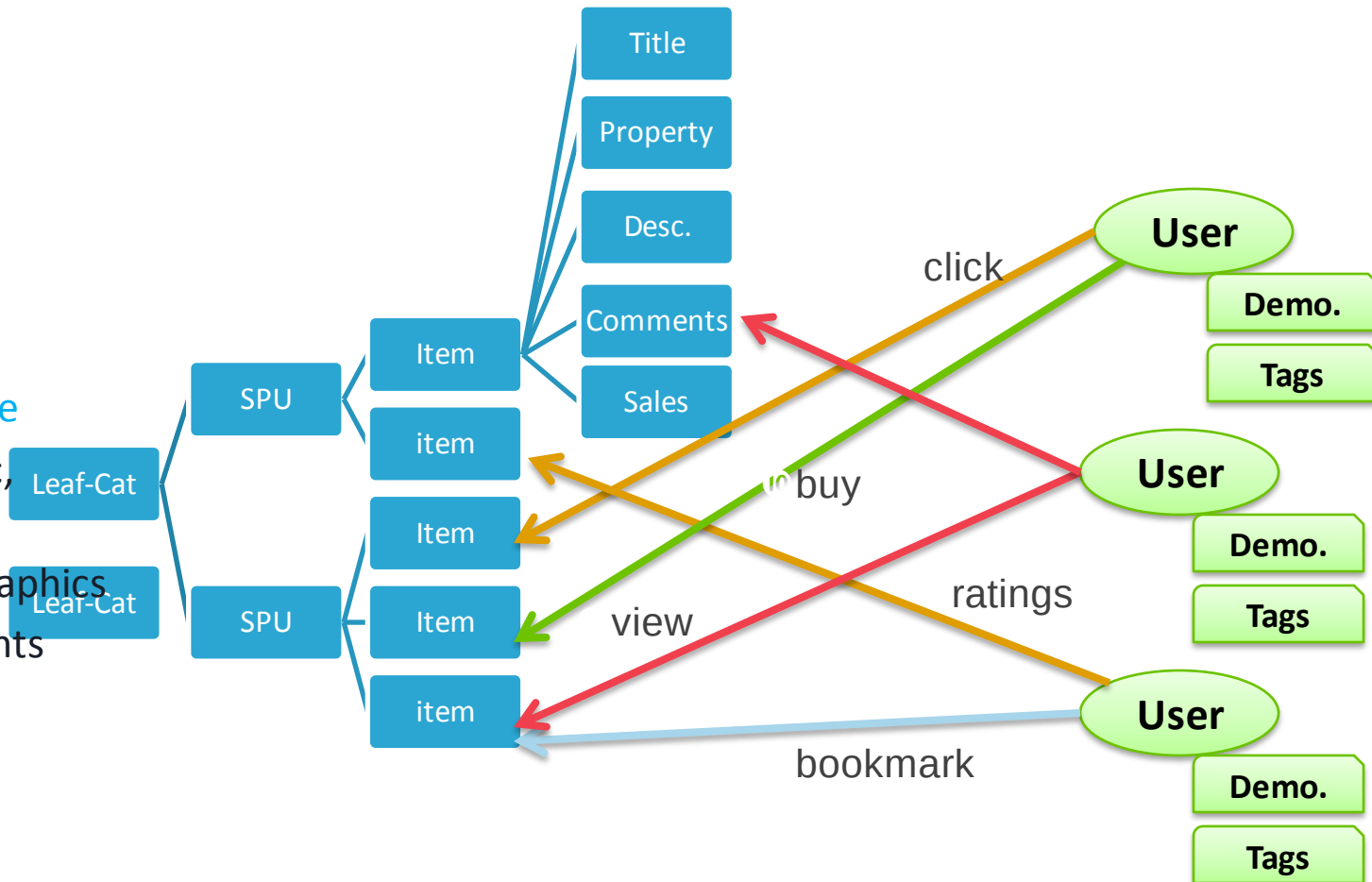
- View
- Click
- Bookmark
- Purchase
- Ratings

- Content data

- Categorical tree
- Item title, desc, properties
- Users' demographics
- Users' comments

- Context data

- loc, time, area
- Mobile

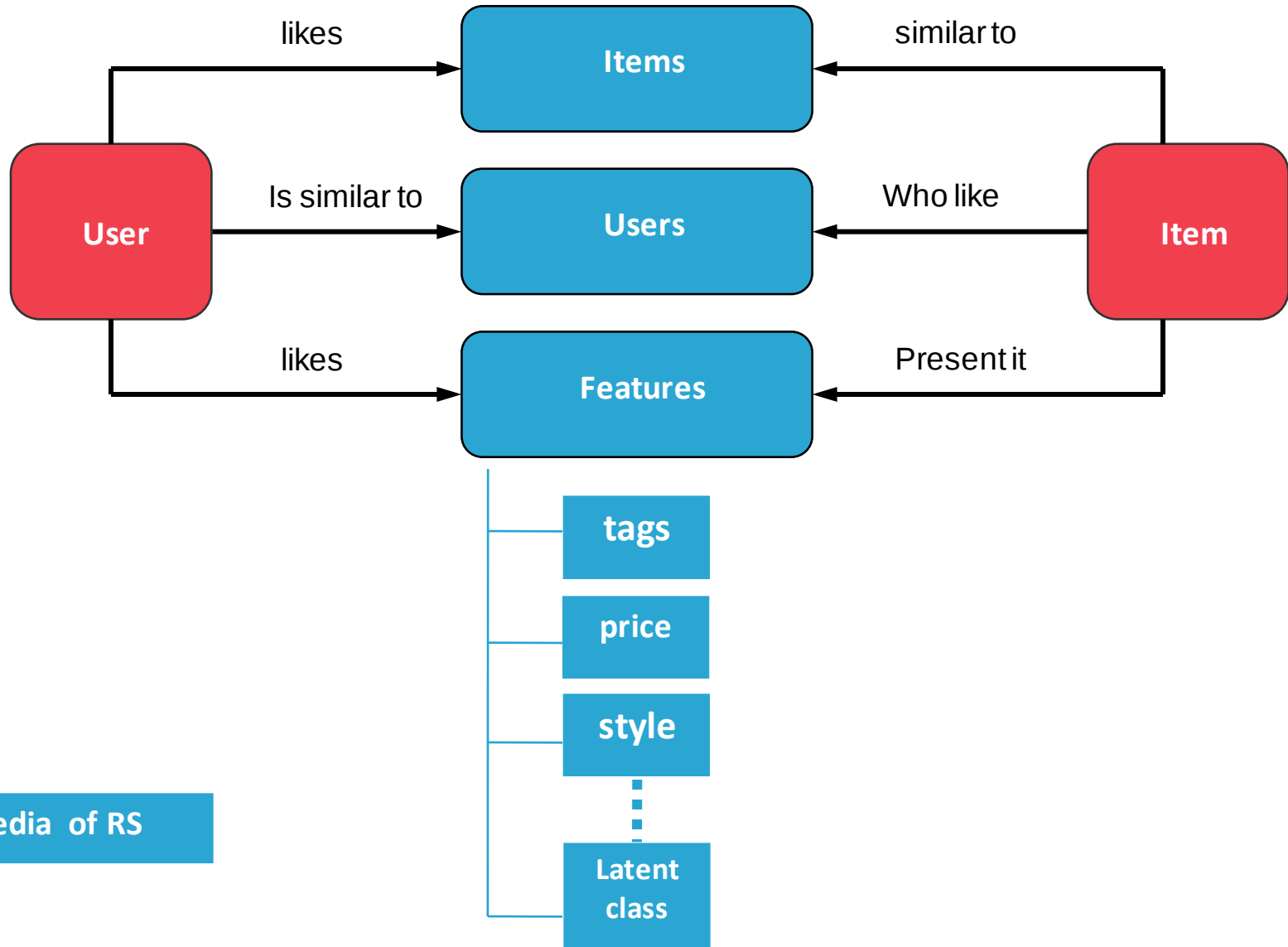


The Power of Recommendation in Wireless

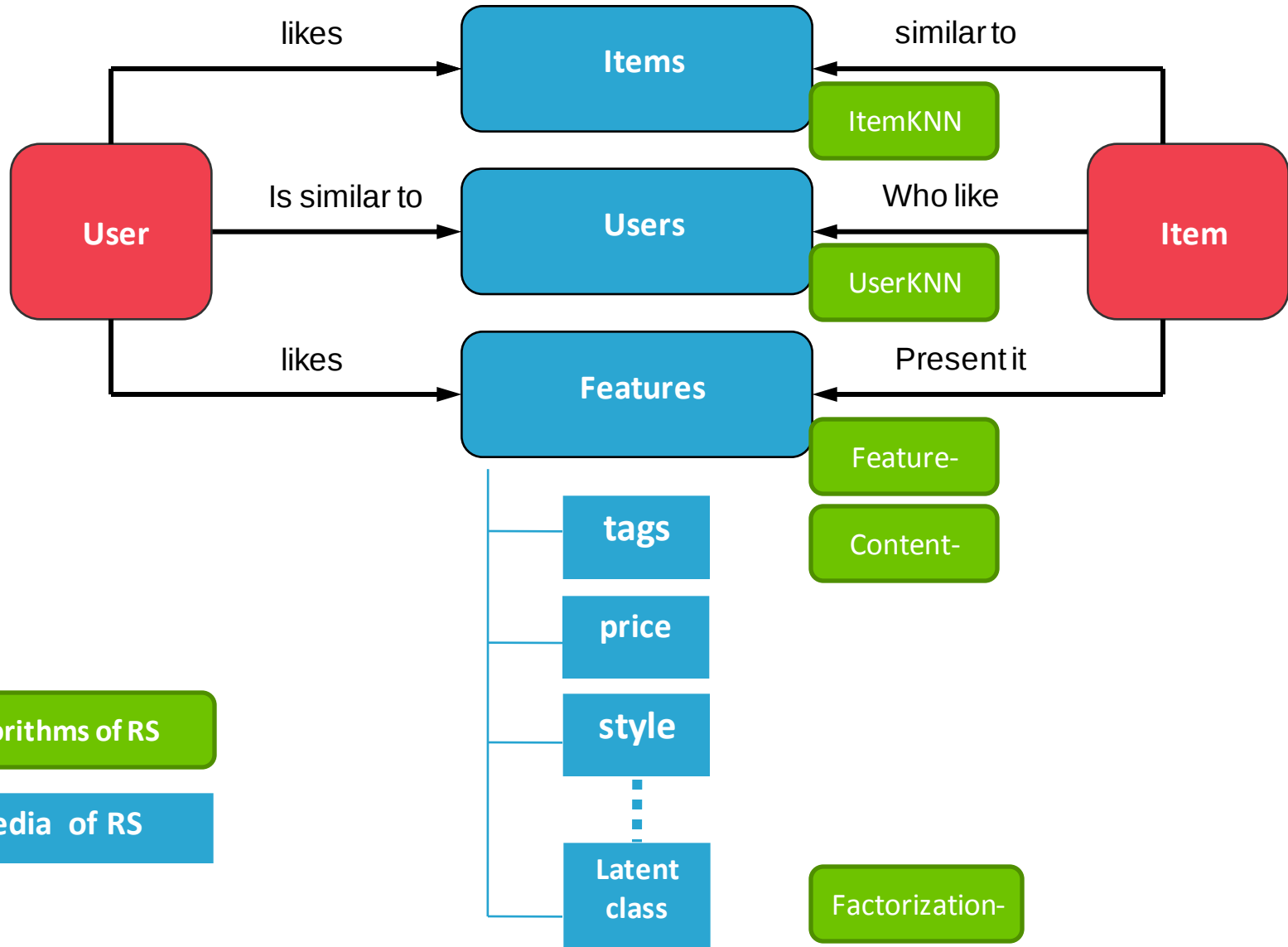
- In more than 2000 shops in m.taobao.com, choose to select auto-RS to replace manually recommendation
- Key metrics:
 - CTR: Click-through rate
 - CVR: Conversion rate
 - pCVR: $\text{CTR} * \text{CVR}$
- The facts
 - Auto-RS outperform manual-RS by 75.7% in CTR
 - Auto-RS outperform manual-RS by 58.4% in pCVR
 - For more than 10K categories, auto-RS only lags behind manual-RS in two (sports shoes, bedclothes)

Recommender Review

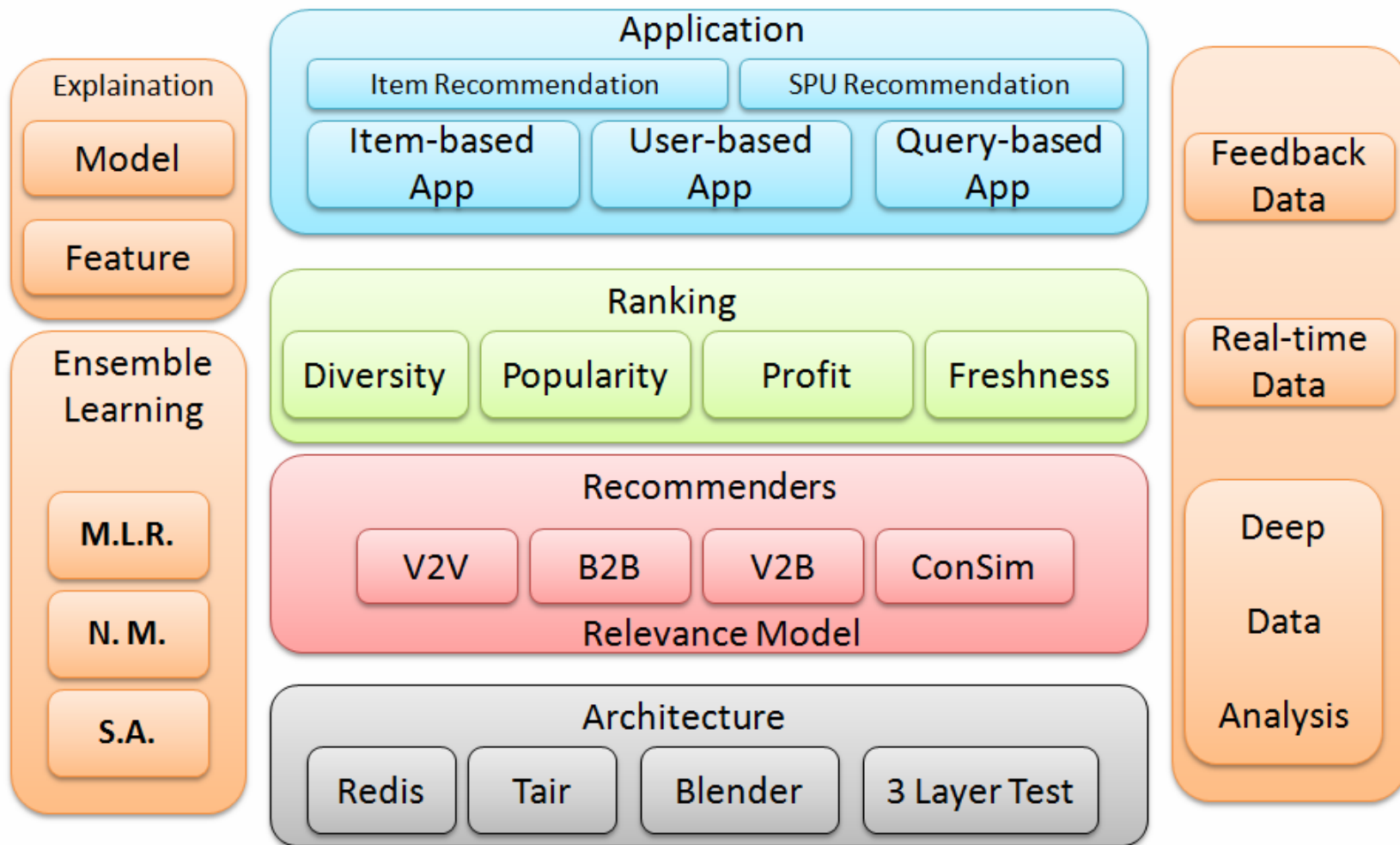
Link user with items in a reasonable way



Recommender Review cont.



eTao RECommender(eTREC) v1



About the Data

- Input: <userID, <item1, item2, ... itemN>>
- Output: <ItemMain: itemReco1, itemReco2, ..., itemRecoN>

	I1	I2	I3	I4
U1	0	0	1	1
U2	1	0	1	0
U3	0	1	0	1
U4	1	0	1	1

User-Item Matrix



	I1	I2	I3	I4
I1	1	0.2	0.4	0.5
I2	0.1	1	0.3	0.7
I3	0.2	0.4	1	0.9
I4	0.1	0.3	0.6	1

Similarity Matrix

Jaccard Review

- Numerator re-thinking
 - Considering each user as a voter
 - Rewrite denominator as a function of user and item pair: $f(u, i, j)$
- when $f(u, i, j)$ equals to 1, it means the classical jaccard.
 - Voter's weight is equally

$$Sim(I_i, I_j) = \frac{\sum_{u=1}^n f(u, I_i, I_j)}{|U(I_i) \cup U(I_j)|}$$

$$f(u, I_i, I_j) = 1$$

Jaccard Review

- In numerator, each voter's weight is dependent on its voting frequency
 - vote more, voter's weight less
 - 20+% (offline recall) gain compare to Jaccard
 - Tuning α , gained 2%

$$f(u, I_i, I_j) = \frac{1}{N(u)}$$

$$f(u, I_i, I_j) = \frac{1}{N(u)^\alpha}$$

Data Sparsity Issue

- Data is not so sparse
 - $\text{Sparsity} = 1 - (\text{U-I Pairs}) / (\text{Nu} * \text{Ni})$
 - Another metric: **active user size / active item size**
 - Example
 - 630M registration user, 150M active user
 - 740M total items, items viewed 80M, items bought 8M
 - 80M items were explored by 150M users, avg. 1.875 per item
 - Content /metadata is rich
 - Item titles, categories, properties, descriptions, etc
- Make it possible to do item-level, SPU-level recommendation directly
 - Currently no clustering, no approximation
 - Future will do for better quality

Understanding users' intent

- For example, in item detail page
 - “Does user want to buy more besides the main item?”
- Relevance VS. Complementary
 - Different data & algorithm
 - How to balance?



Relevance

 凯莱雅温泉泳衣 女分体裙式小胸聚拢遮肚比基尼 ¥ 98.00	 韩国 温泉泳衣 女分体裙式平角 保守带袖 遮手臂 ¥ 98.00	 SASE 2012新款 金格尔 韩国分体平角女游泳衣 泳装 ¥ 50.00	 金格尔温泉泳衣 女小胸比基尼钢托聚拢三件套 ¥ 120.00	 2012韩国 温泉游泳衣 女连体裙式 保守显瘦遮 ¥ 88.00
---------------------------------------	---	---	---------------------------------------	---

 SASE 2012新品 金格尔 韩国版连体三角 女游泳衣 温泉 ¥ 50.00	 硅胶 泡泡游泳帽 大头长发泳帽 护耳防水防着温 ¥ 26.00	 金格尔 高档成人泳裤 男平角 时尚正品温泉 大码 ¥ 30.00	 Tmall 品牌保证 泳镜/游泳镜 男女游泳眼镜防雾防水 潜水镜 正 ¥ 25.00	 正品硅胶/隐形文胸贴加厚聚拢蕾丝泳透气 ¥ 25.00
---	--	---	--	------------------------------------

Complementary

Understand users' intent

- $P(r \mid u, c, f)$
 - c: category
 - f: purchase flow
 - u: user's personal preference
 - r: recommendation type: relevant / complementary
- Purchase flow
 - Homepage -> Search result page -> Comparable page -> Item detail page -> Purchase page -> Post purchase page
 - Users' intent changes in purchase flow
- An example
 - Category's impact

Dynamic Layout optimization

电子/电工	4.79%
床上用品/布艺软饰	17.99%
家居饰品	28.32%
网络设备/网络相关	35.50%
宠物/宠物食品及用品	44.23%
尿片/洗护/喂哺/推车床	48.84%
保健品/膳食营养补充剂	50.47%
汽车/用品/配件/改装	50.85%
书籍/杂志/报纸	52.90%
美容护肤/美体/精油	57.05%
孕妇装/孕产妇用品/营养	59.75%
生活电器	87.03%
笔记本电脑	87.50%
影音电器	88.22%
ZIPPO/瑞士军刀/眼镜	88.64%
饰品/流行首饰/时尚饰品新	92.38%
箱包皮具/热销女包/男包	92.38%
珠宝/钻石/翡翠/黄金	92.38%
音乐/影视/明星/音像	92.38%
古董/邮币/字画/收藏	92.38%
手机	92.38%
特色手工艺	92.38%
品牌手表/流行手表	92.38%

- Dynamically allocation of relevant and complementary
- Primary online test
 - Boosting CTR by 6+%

Category's impact - watch

看了此商品的会员通常还看了



限量版冠琴正品 超薄间金男士手表 休闲商务手表 自动机械表
¥889.0 ¥1270.00
★★★★★ (32人评价)



新款特价 冠琴手表 最酷男表 登上运动系列多功能表 手表
¥399.0 ¥798.00
★★★★★ (7人评价)



限量包邮 冠琴正品 时装表 自动机械表 商务男表 手表
¥560.0 ¥1120.00
★★★★★ (103人评价)



卖疯了 冠琴正品 时尚钨钢手表 玫瑰金女士手表 永不褪色 女表
¥298.0 ¥596.00
★★★★★ (280人评价)

买了此商品的会员通常还买了



爆款 2012特价全防水 女士手表钨钢女表 情
¥298.0 ¥596.00
★★★★★



冠琴正品 秒杀爆款-5 色可选 自动机械表 情
¥680.00
★★★★★



爆款 时装表 罗马钻面 钨钢男士手表 情侣对
¥298.0 ¥596.00
★★★★★



包邮 冠琴正品 真陶瓷 防水女表时装表 女士
¥596.00
★★★★★



冠琴正品 多功能手表 背透全自动机械手表男
¥336.0 ¥480.00
★★★★★

Category's impact - furniture



花木家居 布艺 蕾丝 桌布 餐桌布 台布椅套椅垫

¥36.0



花木家居 布艺软饰 餐桌布 台布 贡缎镂空绣花 粉

¥59.0



花木 布艺软饰 贡缎镂空 飞梭水溶花边 餐桌布台

¥29.0



花木家居 布艺软饰 桌布 台布餐桌布 手工丝带绣

¥24.0



花木 布艺软饰 提花绣花 镂空 多功能盖巾 桌布台

¥9.9



花木 布艺软饰 餐垫杯垫 隔热垫花瓶垫 雕绣 绣花

¥4.5



花木 全棉圈毛棉 乳胶底 进门垫 脚垫 防滑垫 地垫

¥18.0



花木 布艺软饰 餐巾盒抽 纸盒纸巾套餐巾套 田园

¥15.0



花木家居 布艺软饰 田园 小碎花 蕾丝 桌脚套 椅

¥1.5



花木家居 布艺软饰 餐垫 盘垫 隔热垫 杯垫 桌布

¥5.0

Understand users' intent - Bookmark

- Behavioral data

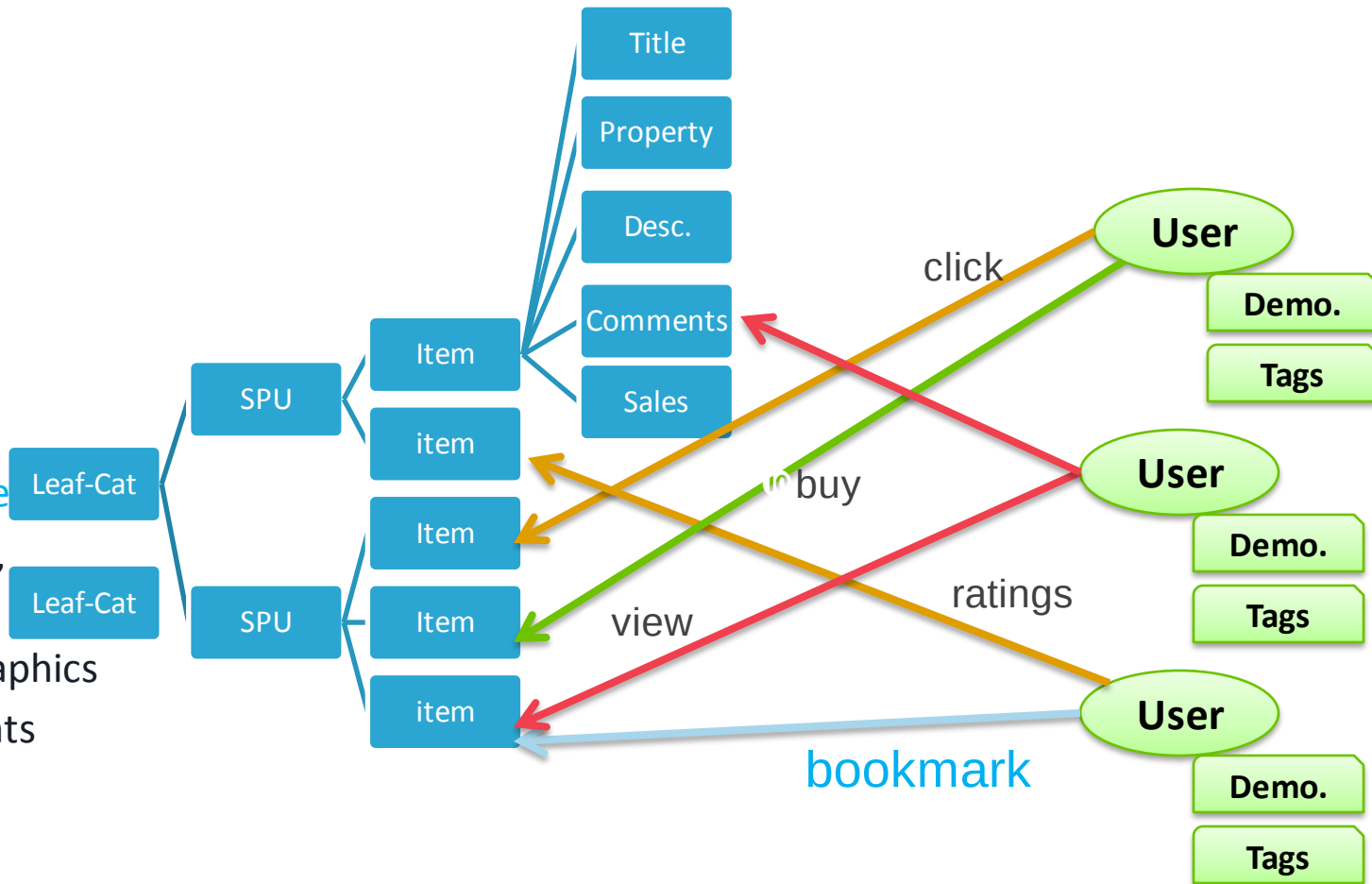
- View
- Click
- Bookmark
- Purchase
- Ratings

- Content data

- Categorical tree
- Item title, desc, properties
- Users' demographics
- Users' comments

- Context data

- loc, time, area
- Mobile



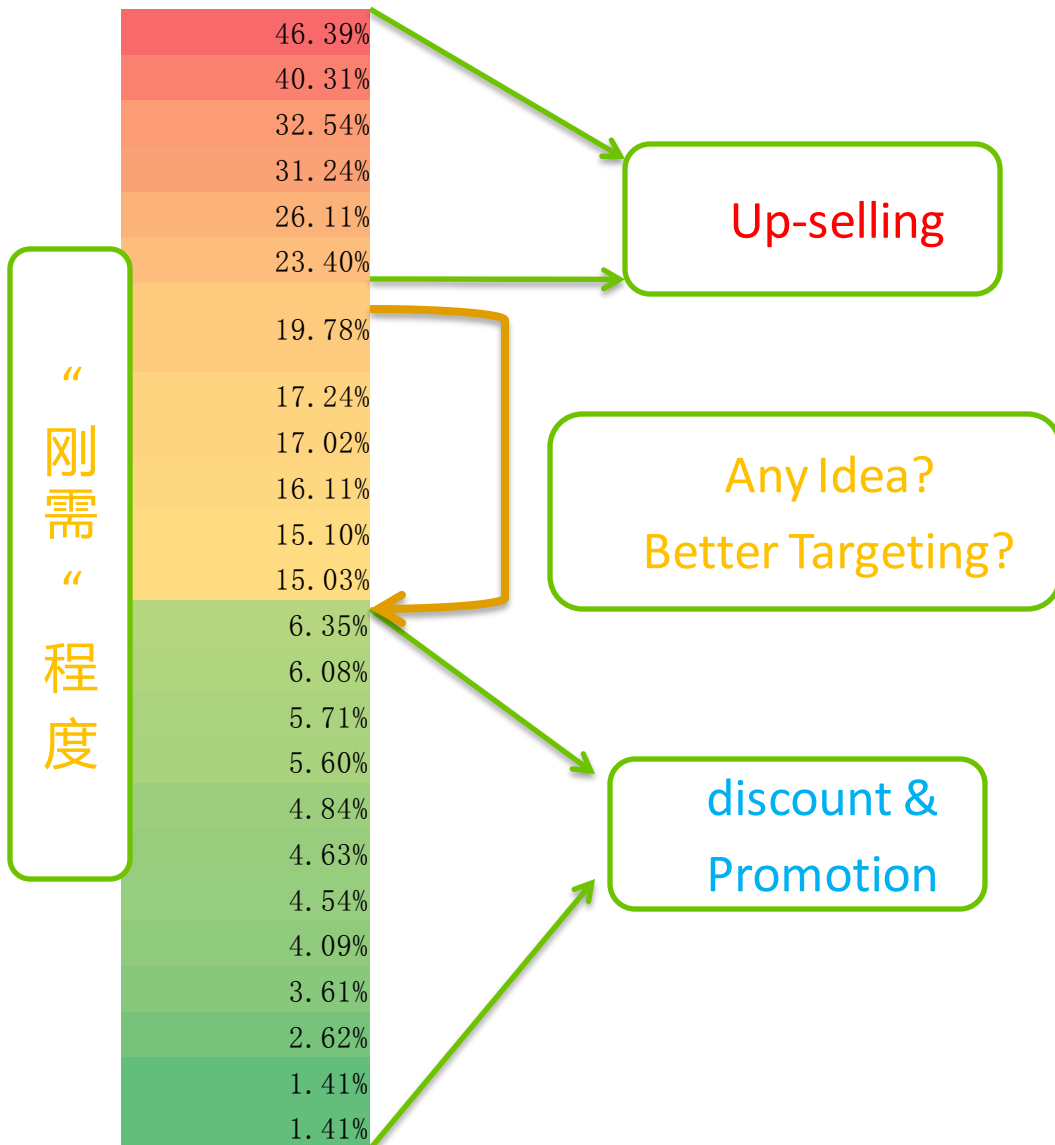
Understanding user behavioral – Bookmark II

- Bookmark != Buy
- Customers' price sensitivity
 - As item's price goes up, people become more cautious



Category's stickiness

网络游戏点卡
腾讯QQ专区
网游装备/游戏币/帐号/代练
移动/联通/电信充值中心
网游垂直市场根类目
网店/网络服务/软件
OTC药品/医疗器械/隐形眼镜/计生用品
书籍/杂志/报纸
电影/演出/体育赛事
本地化生活服务
个人护理/保健/按摩器材
手机号码/套餐/增值业务
大家电
女鞋
台式机/一体机/服务器
女装/女士精品
数码相机/单反相机/摄像机
住宅家具
笔记本电脑
自用闲置转让
珠宝/钻石/翡翠/黄金
外卖/外送/订餐服务(垂直市场)
房产/租房/新房/二手房/委托服务
新车/二手车

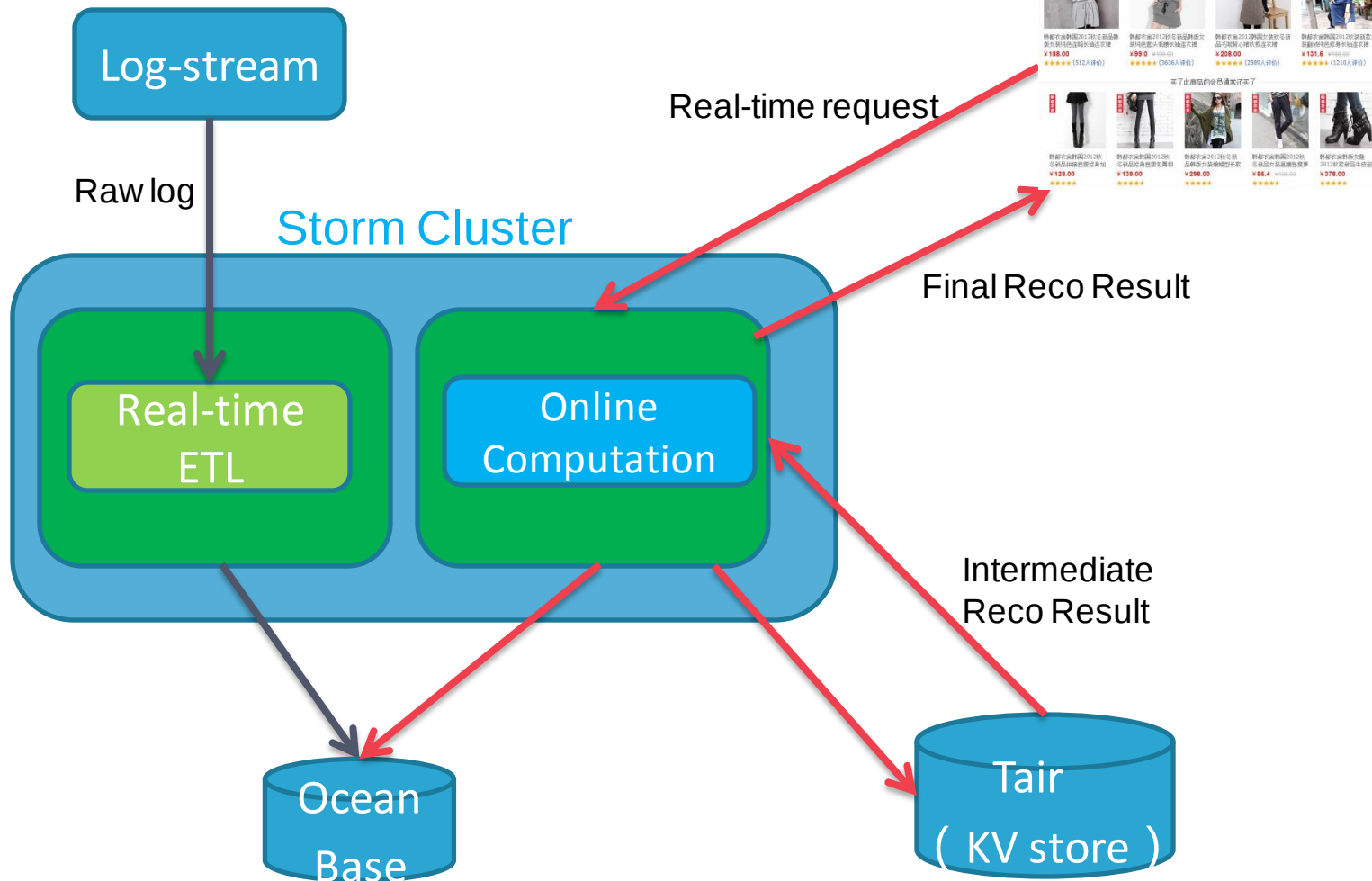


Algo. Summary

- ItemCF
 - Heavily rely on
 - It (and variants) covers 80% of typical scenarios
 - High efficiency
 - MapReduce implementation; 16 min on Tmall data, 2h or so on whole site
 - Will be released internally 7th. Nov, 2012, named **eTREC_v1**
- Content-based
- Offline hybridization & Online Ensemble
- Exploring
 - Implicit matrix factorization (IMF)
 - learning to rank
- Future
 - Markov model / Tag-based RS / RBM / online learning etc.

Real-time Recommendation

A Dilemma: Real-time VS. Performance



Thanks & QA