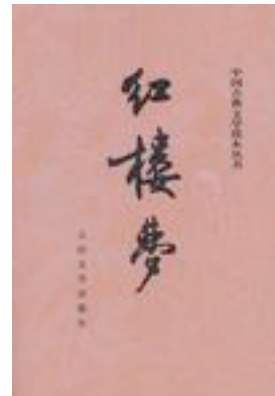


寻路推荐

——豆瓣推荐系统实践

阿稳 豆瓣算法组

今天讲什么



三个案例

怎么讲



三步流程

点在哪？

- CF与Content-based
- 专家推荐
- 从算法到系统

困境I

喜欢看"叶问"的人也喜欢·····



梅兰芳



非诚勿扰



新警察故事



闪电狗



女人不坏



硬汉



通缉令



冬荫功



大搜查之女



霍元甲

头部效应

UGC的尴尬



胖子 说:

“ 这些用户都走进了同一家电影院 ”

3年前 回应 赞 (1) 转播

CF正确地反映了数据空间，但得到了毫无意义的结果

为什么我就是知道它们不搭调



豆瓣成员常用的标签(共3038个)

动作(21536) 香港(11508) 2008(3802) 功夫(1450) 传记(1287) 剧情(633) 武侠(619) 历史(522)



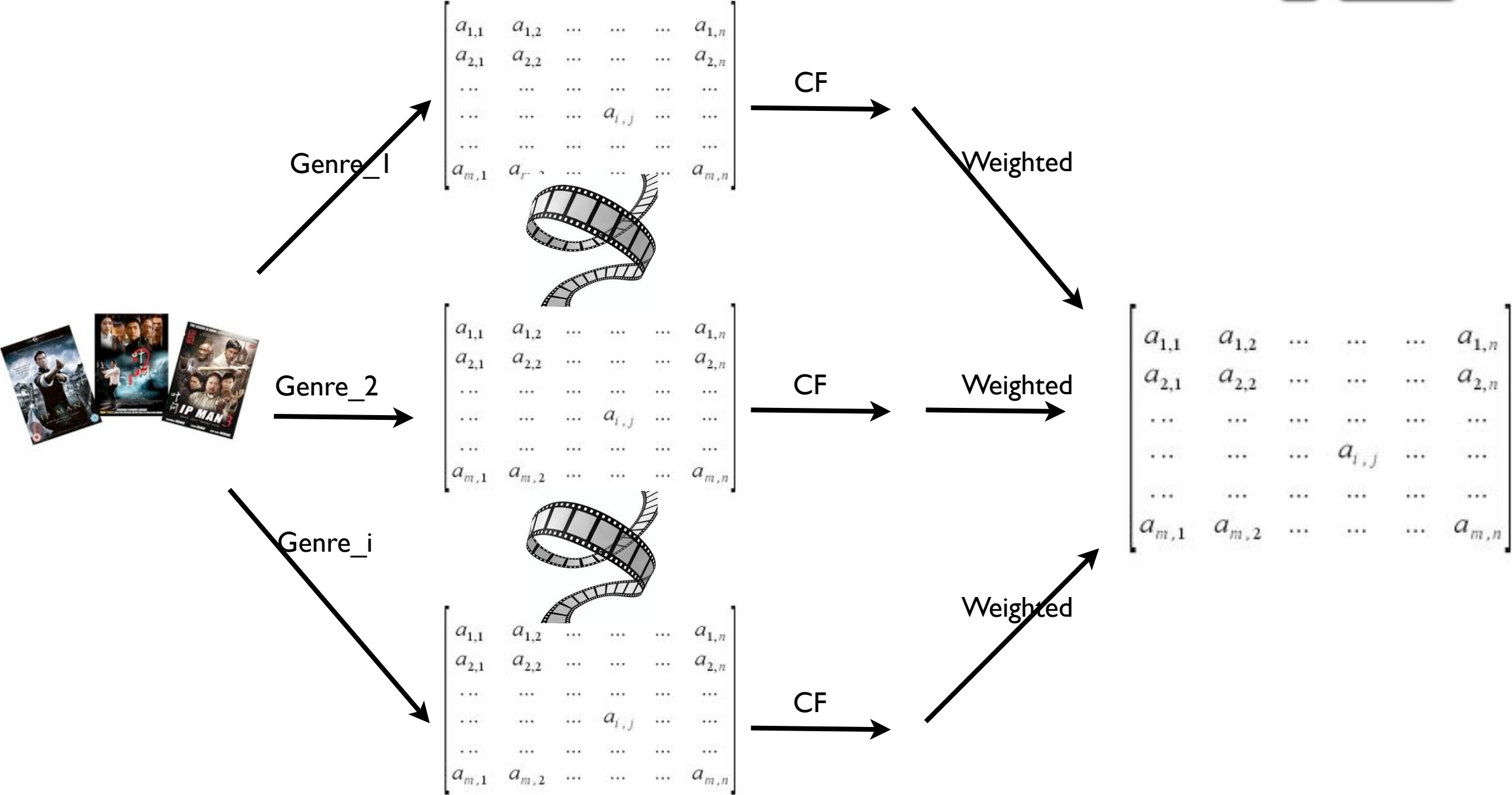
豆瓣成员常用的标签(共2234个)

陈凯歌(12130) 梅兰芳(8836) 中国电影(6085) 传记(5863) 黎明(4236) 章子怡(4216) 中国(3726) 大陆(2909)



Content-based分组+Weighted-CF

42



喜欢看“叶问”的人也喜欢·····



叶问2：宗师传
奇



导火线



新警察故事



十月围城



杀破狼



大兵小将



通缉令



冬荫功



盗佛线



霍元甲

算法改变世界

困境2

我们需要一个新的读书豆瓣猜



它要经常变的，这样我才会不时地来看看

要推荐新的书，这样才有惊喜，红楼梦天龙八部拜托就不要推了

我的阅读兴趣有很多方面，能不能不要只推文学类的

我不相信机器，你总得告诉我为什么要推荐这个吧

冷启动？新颖度？多样性？时效性？推荐解释？CF的问题？



The Wisdom of the Few

A Collaborative Filtering Approach Based on Expert Opinions from the Web

Xavier Amatriain
Telefonica Research
Via Augusta, 177
Barcelona 08021, Spain
xar@tid.es

Neal Lathia
Dept. of Computer Science
University College of London
Gower Street
London WC1E 6BT, UK
n.lathia@cs.ucl.ac.uk

Josep M. Pujol
Telefonica Research
Via Augusta, 177
Barcelona 08021, Spain
jmps@tid.es

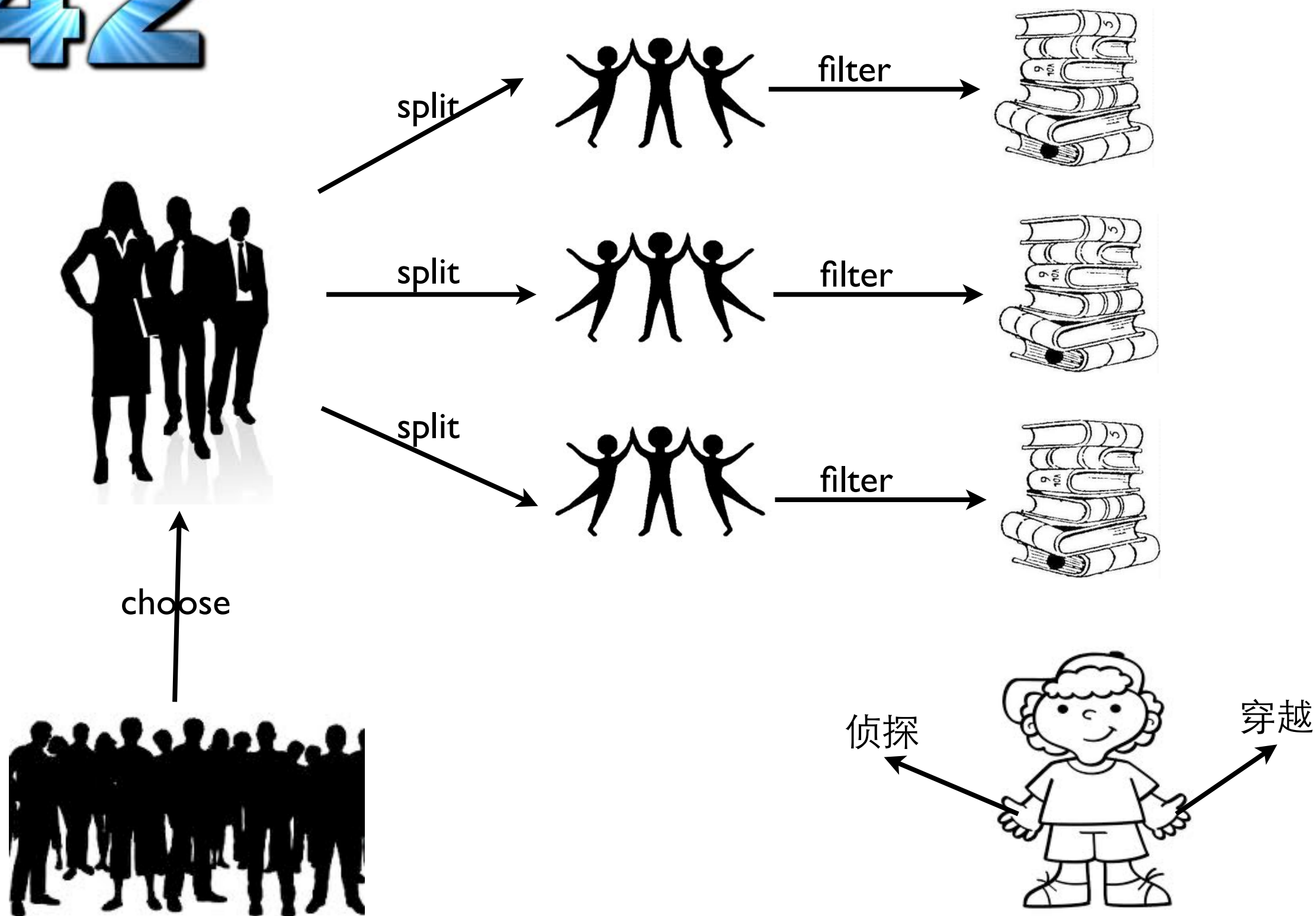
Haewoon Kwak
KAIST
Computer Science Dept.
Kuseong-dong, Yuseong-gu
Daejeon 305-701, Korea
haewoon@an.kaist.ac.kr

Nuria Oliver
Telefonica Research
Via Augusta, 177
Barcelona 08021, Spain
nuriao@tid.es

<http://www.wentruen.net/blog/?p=1034>

在大部分的场合，我们需要的并不是与自己相似的用户
的推荐，而是与自己相似的专家的推荐

42



不明确面目的群体：统计特征

解决新颖性、多样性、分众性

有明确面目的个人：特征描画

解决个性化匹配

豆瓣猜

豆瓣猜你可能感兴趣的新书



豆瓣猜你的阅读兴趣

全部

历史

小说

美国

随笔

旅行

心理学

社会学

编程

政治

数学

社会

商业

经济学

科普

文学

金融

中国

文化

互联网

哲学

法律

管理

散文

传记

创业

设计

思维

台湾

经典



什么是豆瓣猜

巴黎，很烦人

散文&旅行

毕远月 / 2011-8 / 上海人民出版社 / 39.00元 / 平装

世界上很少有哪一座城市会像巴黎那样被夸得天花乱坠。可撇开那些关于巴黎的陈词滥调，真正的巴黎到底是什么样子的？作者以巴黎人的眼光，用异于平常作者的独特视角，向读者展示日常生活中真实有趣的巴黎面貌，有一种透过指缝看风景的绝妙感觉。全书分为四个章节，通过“好的，待会见（巴黎人说完再见并不各奔东西，跟所有人都拥抱亲吻一遍后还是依依不舍，这是一座无...

中信出版社 (小日)

9人推荐

标签: 经济学 / 管理

管理-企业管理 (枫之舞)

4人推荐

标签: 传记&管理

算法改变世界

困境3



怎么解决新用户问题

怎么灵活地调整算法的参数

怎么在多个算法间进行调度

怎么评价不同算法的质量

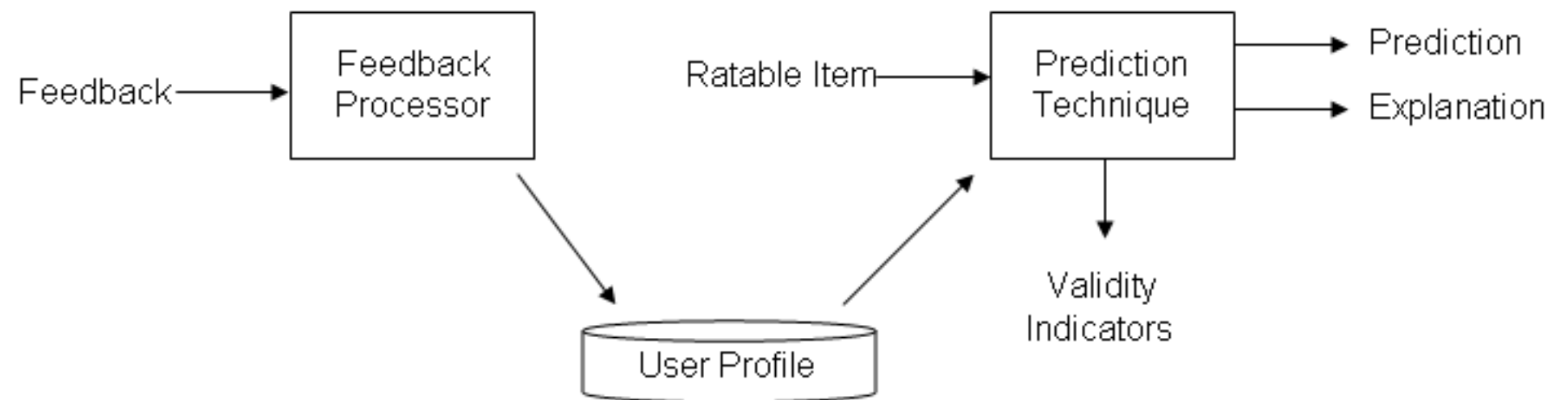
怎么平衡不同用户群在不同状态的需求

怎么根据用户实时反馈做出响应



一个算法不能解决所有问题，多个算法要解决很多问题

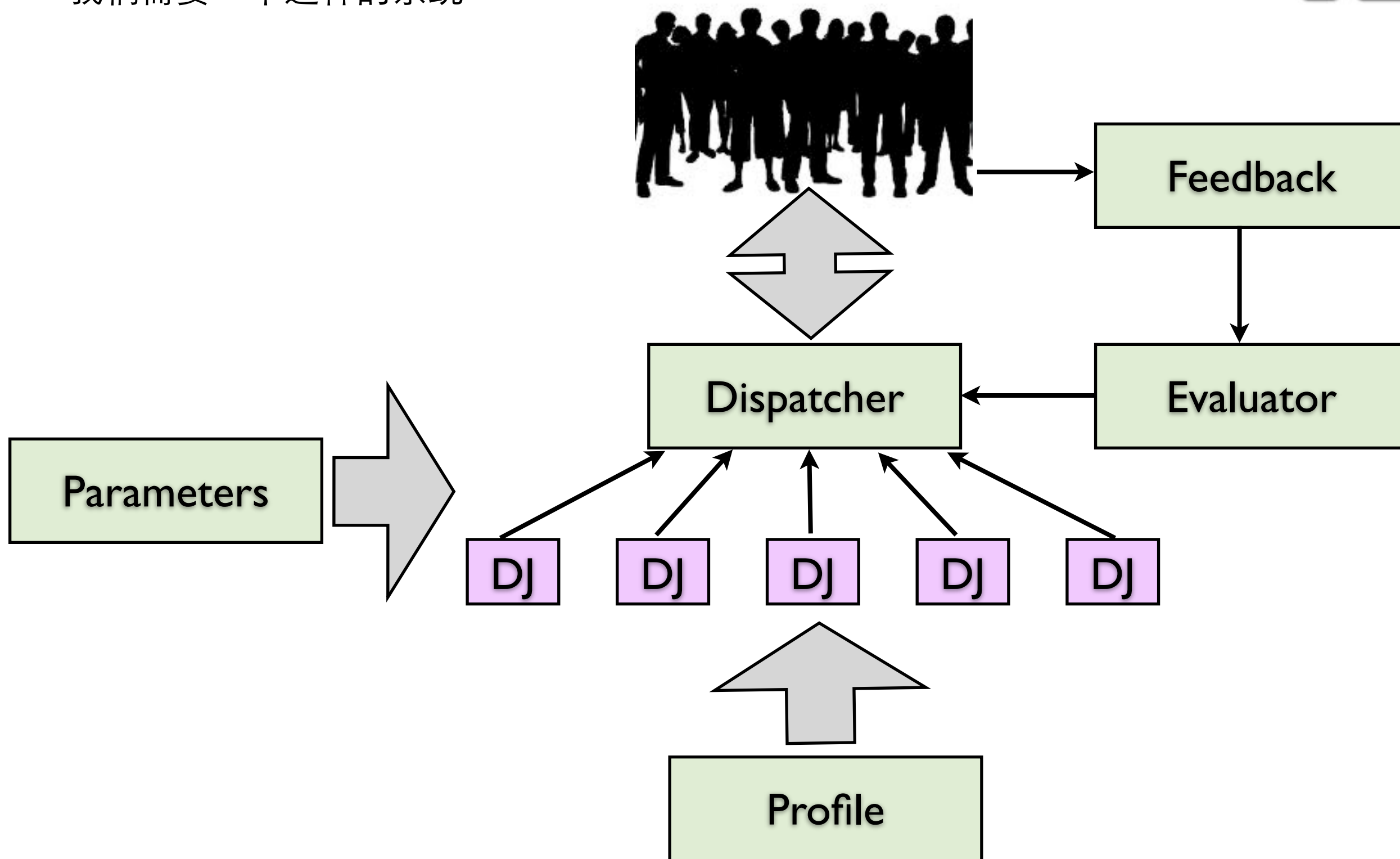
Duine的思路



<http://www.wentruer.net/blog/?p=1034>

42

我们需要一个这样的系统



无数据无真相，该系统在当时的表现：

全站用户满意度：0.75

新用户满意度：0.6

表现最好的DJ算法红心率：18%

对比热门歌曲推荐DJ（baseline），满意度：0.4，红心率=垃圾桶率=5%

算法改变世界

更多的挑战， 更多的算法机会

- 尽快地让用户度过冷启动阶段
- 保持用户听歌会话的一致性
- 新歌曲进入/劣质歌曲淘汰的闭环
- 公共频道的个性化
- 移动端的推荐算法
- 新的产品形式

随着用户成长的推荐系统



Q & A

<http://www.douban.com/people/wentruer/>

<http://www.wentruer.net/blog/>

<http://www.weibo.com/wentruer>



<http://site.douban.com/jobs/>

本讲稿感谢豆瓣，感谢<http://images.google.com/>，图片非本人版权，纯属借用