

相关性反馈在 推荐系统中的应用

李北涛

Tumblr 数据科学工程总监



QCon

全球软件开发大会

成为软件技术专家 的必经之路

[北京站] 2018

2018年4月20-22日 北京·国际会议中心

7折

购票中, 每张立减2040元

团购享受更多优惠



识别二维码了解更多

AiCon

全球人工智能与机器学习技术大会

助力人工智能落地

2018.1.13 - 1.14 北京国际会议中心



扫描关注大会官网



极客时间

重拾极客精神·提升技术认知

下载极客时间App

获取有声IT新闻、技术产品专栏，每日更新



扫一扫下载极客时间App

SPEAKER INTRODUCE



李北涛

Tumblr 数据科学与工程总监

大学就读于中国科学技术大学少年班。后赴美国加州大学圣芭芭拉分校留学。获计算机工程博士学位。主要研究领域在机器学习，数据挖掘，和多媒体技术。毕业后在多家Startup任职。涉足搜索，电子商务，社交网络等领域。现任Tumblr数据科学与工程总监。擅长于创造性开发产品和算法。对用户体验和产品方向有较深理解。

TABLE OF CONTENTS 大纲

- Tumblr 推荐系统介绍
- 设计目标
- 相关性反馈和加强学习
- 几个实际问题
- 效果和总结
- Q&A

TABLE OF CONTENTS 大纲

- Tumblr 推荐系统介绍
- 设计目标
- 相关性反馈和加强学习
- 几个实际问题
- 效果和总结
- Q&A

Tumblr

- 社交分享平台
- 3.75亿 博客
- 1500亿 帖子
- 84%用户通过移动访问



主要推荐产品

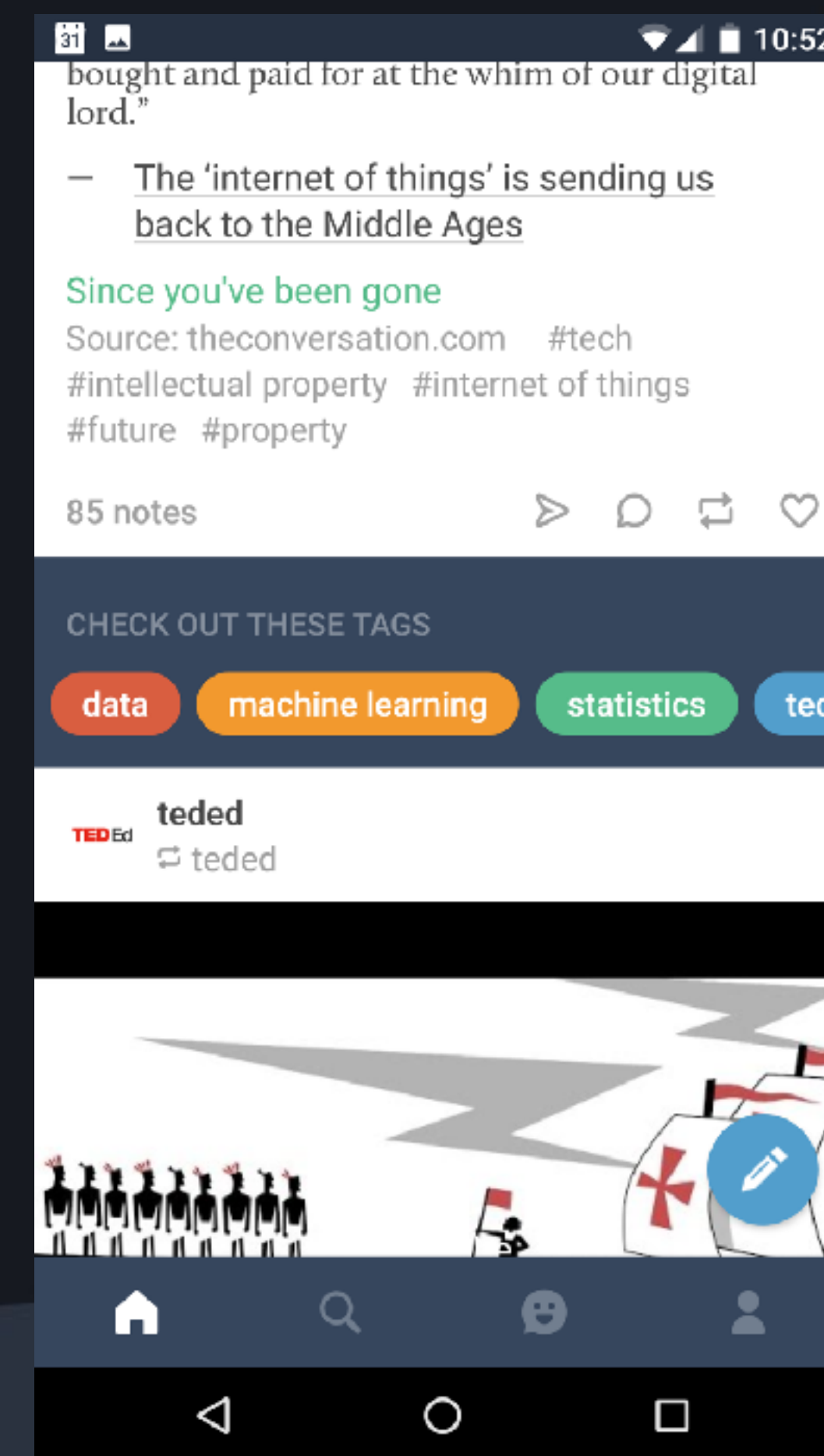
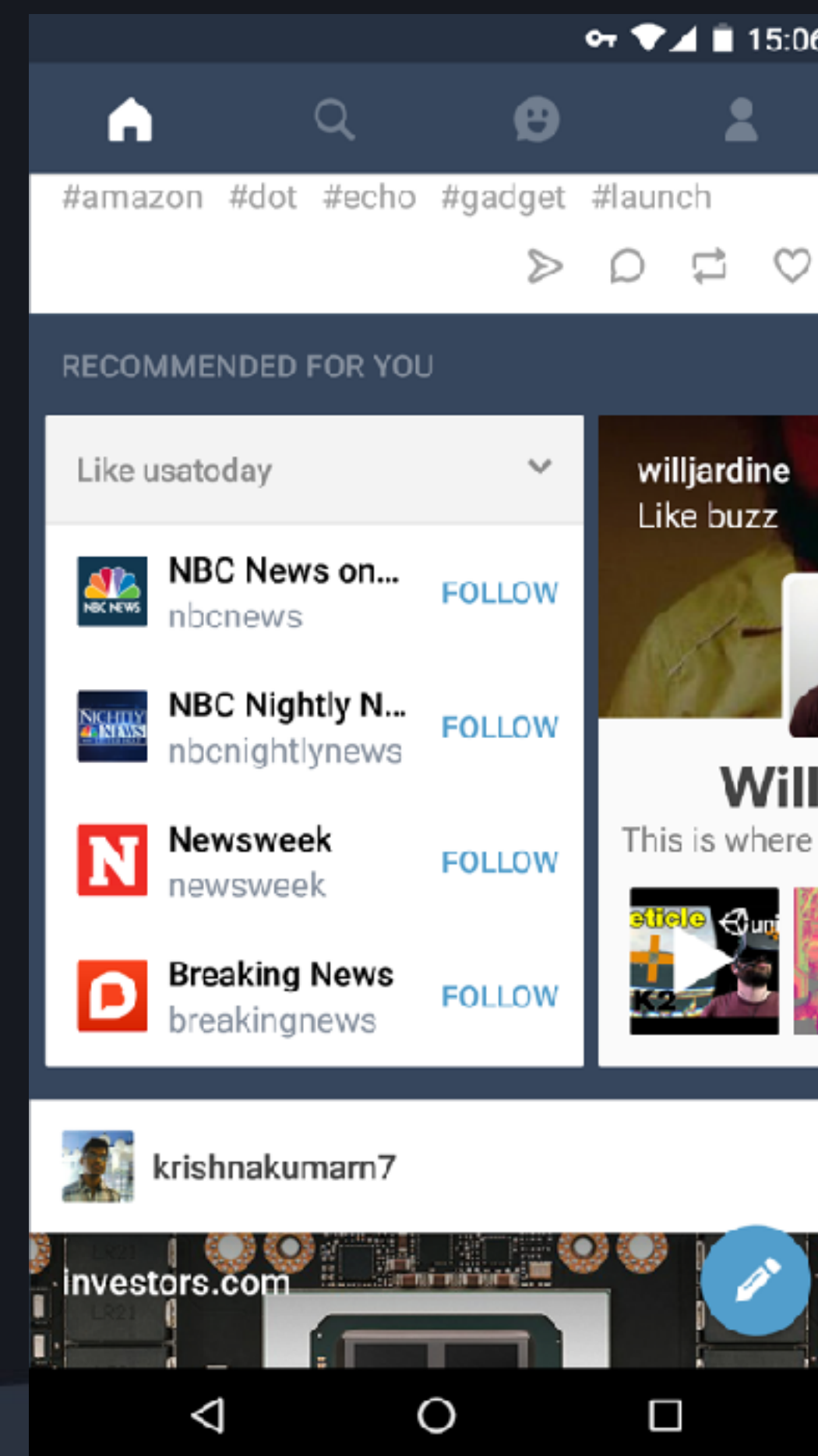
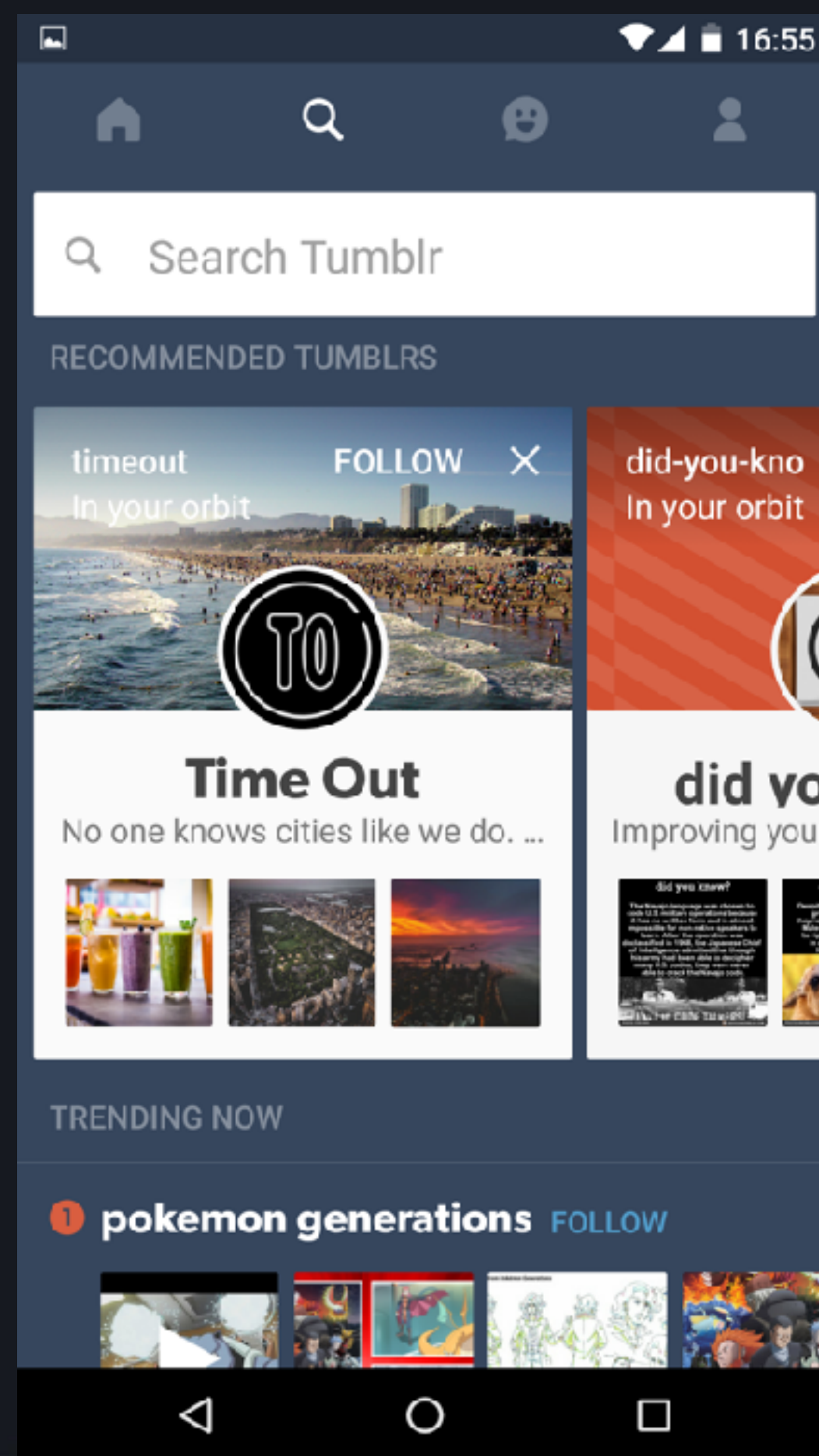
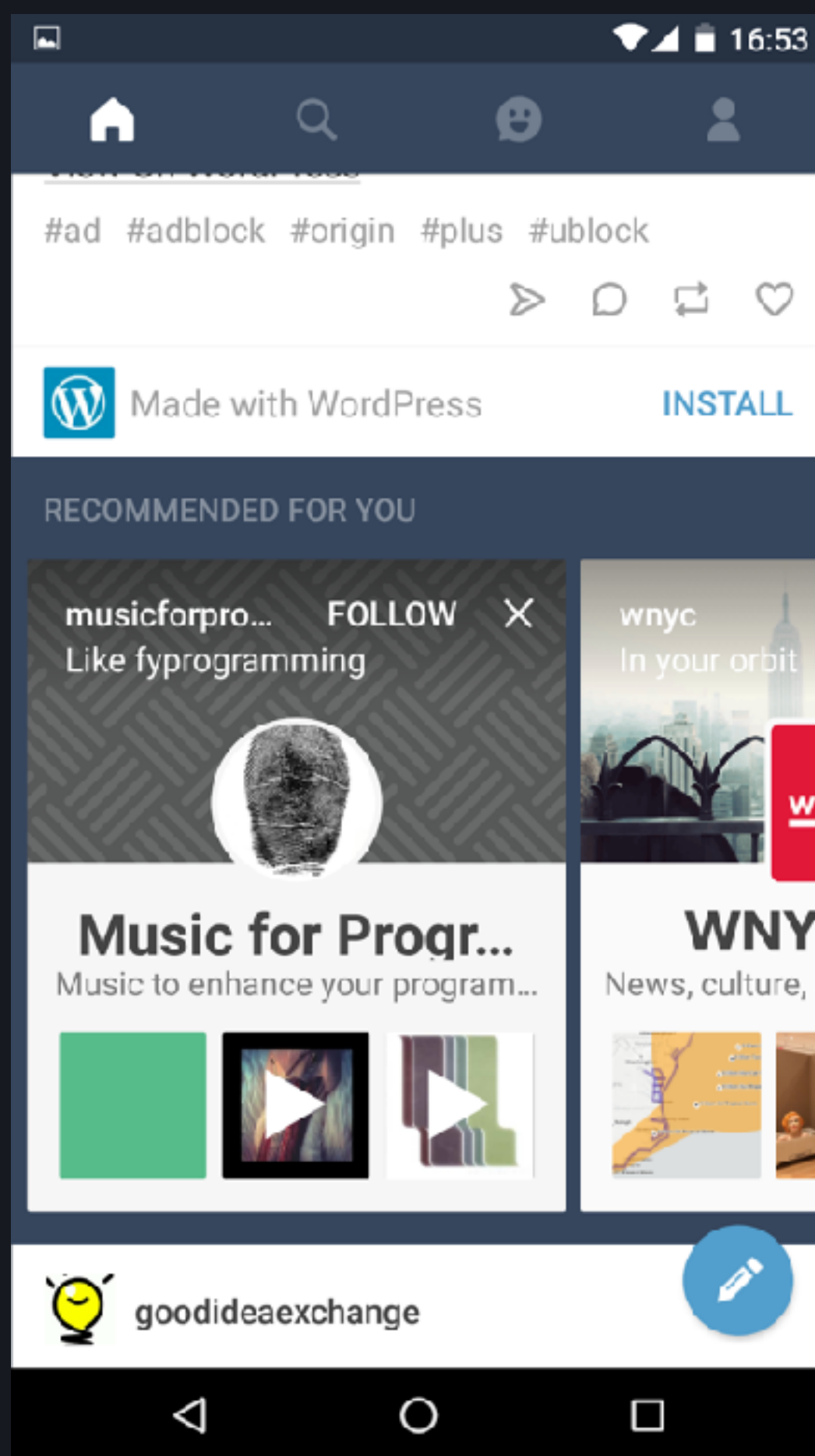


TABLE OF CONTENTS 大纲

- Tumblr 推荐系统介绍
- 设计目标
- 相关性反馈和加强学习
- 几个实际问题
- 效果和总结
- Q&A

设计目标

- 通用性
- 即插即用
- 快速迭代

通用性

- 推荐源的抽象定义
 - 一个包括N个从user到item(s)的映射的集合

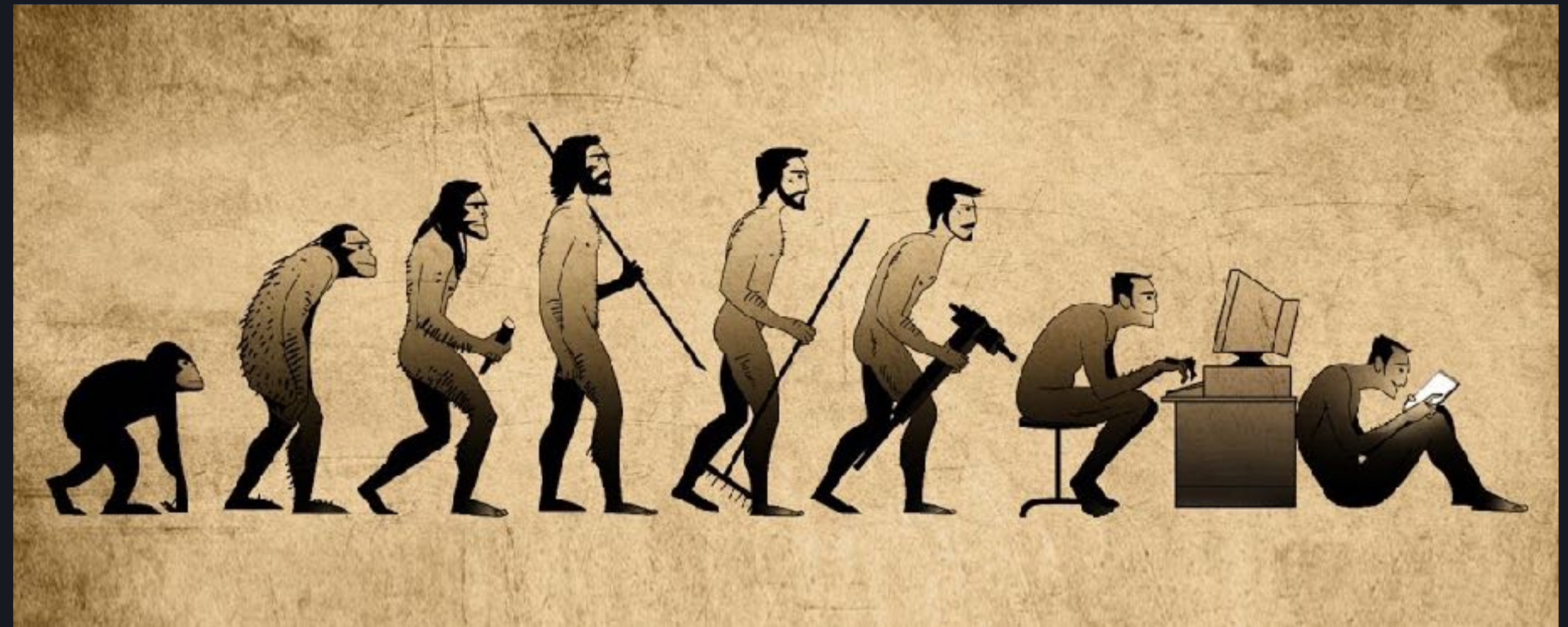
即插即用

- 支持推荐源的无缝添加和删减



快速迭代

- 实时自动调整推荐源的比重，实现用户体验优化。
- 支持不同推荐源的组合。自动分析用户反馈，达到最佳组合。
- 加快推荐系统进化速度。



推荐系统架构

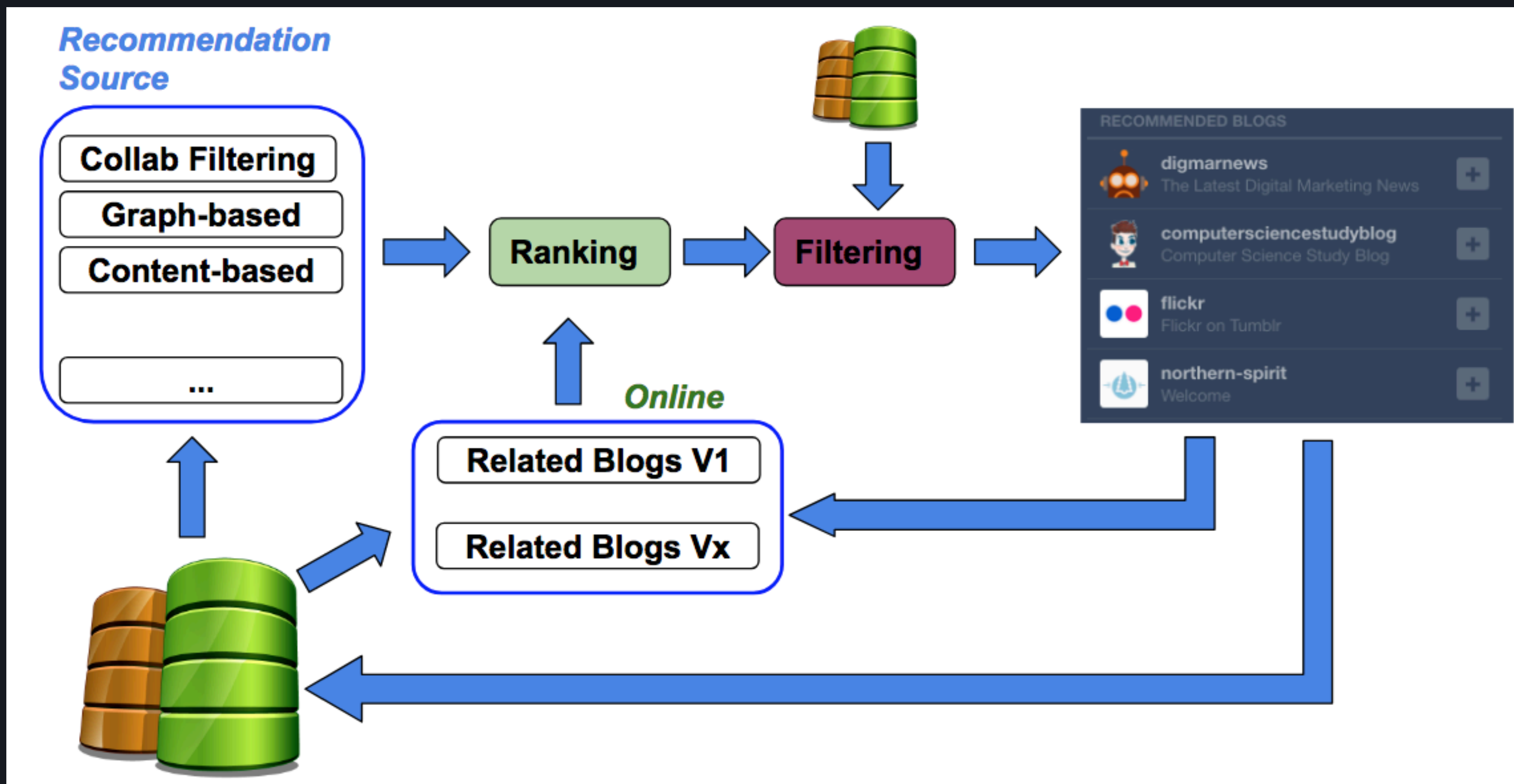


TABLE OF CONTENTS 大纲

- Tumblr 推荐系统介绍
- 设计目标
- 相关性反馈和加强学习
- 几个实际问题
- 效果和总结
- Q&A

相关性反馈

- 用户反馈
 - Follow （强正反馈）
 - Reject （强负反馈）
 - Ignore （弱负反馈）

反馈数据稀疏

- 用户数据多层级聚合
- 博客级别正负反馈聚合
- 推荐源级别正负反馈聚合

推荐特性

- 每个推荐有一系列不同粒度的特性
 - 基于内容（标签：“梅西”，“巴萨”，“足球”，“体育”）
 - 基于相似性（相似特性，共同朋友）
- 一个反馈信息和多个推荐特性相关
 - 充分利用反馈信息

反馈信息

- (用户, 推荐特性) 级别
- 推荐特性级别
- 自动过渡

曝光率(Exposure)权重

- 显示权重
 - 排名位置
 - 产品单元
- 用户权重
 - 新用户 / 老用户
 - 不同地区

曝光率计算

- 曝光环境
 - (UK-新用户-Dashboard-Rank2)
 - (US-老用户-Explore-Rank3)
- 总曝光强度为其在各个环境里的曝光强度之和

量化反馈信息

- 对每个曝光环境计算全局平均值
- 根据推荐特性在不同曝光环境的分布，计算期望值
- 根据实际推荐效率，计算推荐特性的反馈积分

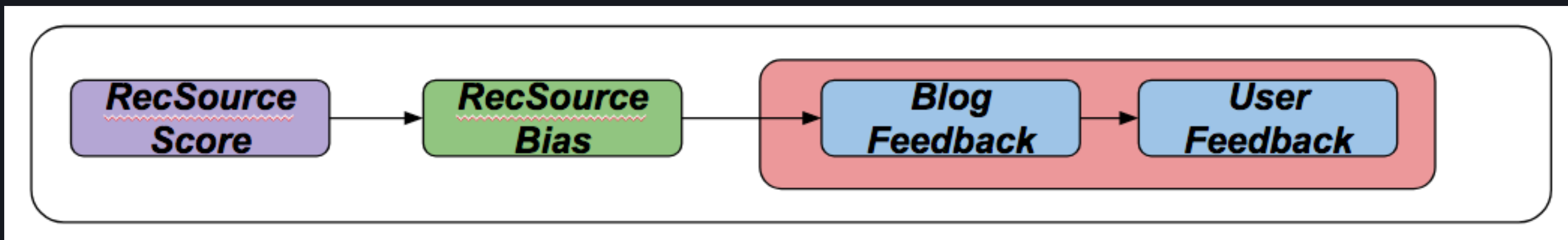
TABLE OF CONTENTS 大纲

- Tumblr 推荐系统介绍
- 设计目标
- 相关性反馈和加强学习
- 几个实际问题
- 效果和总结
- Q&A

曝光环境要分多细

- 不能太粗
 - 曝光强度不精确
- 不能太细
 - 没有足够数据支持计算期望值

整合推荐源

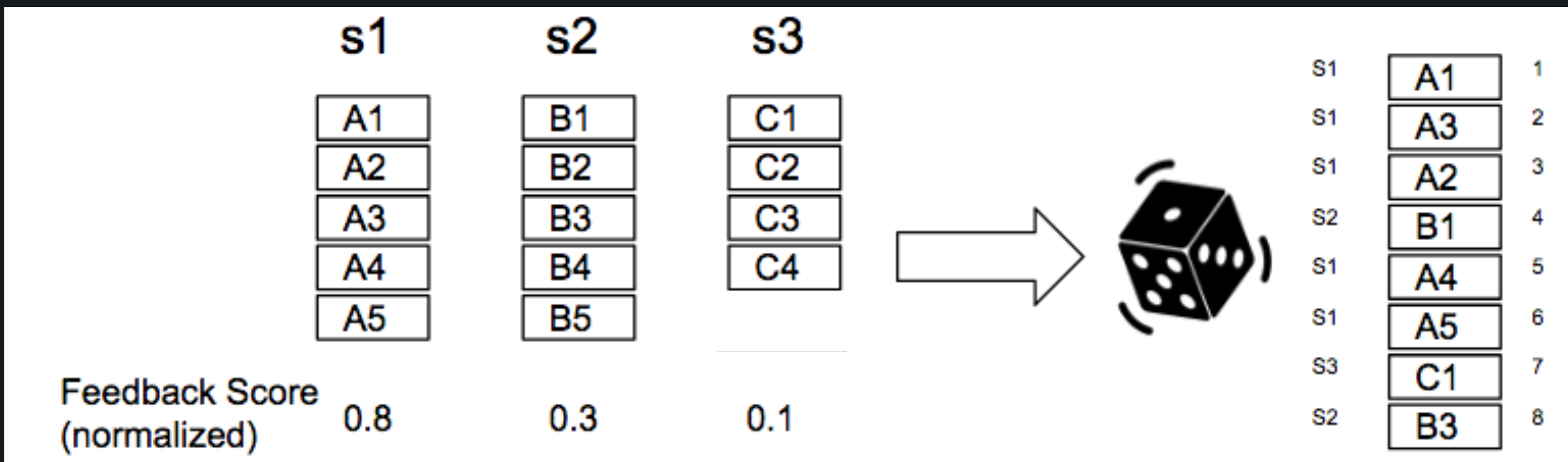


- 怎么比较不同推荐源的分數？
- 强耦合： 要求不同推荐源分數可比较。
- 怎么处理推荐数量非常不同的推荐源？

推荐源取样算法

- 在每个推荐源内部排序
- For each $i \leftarrow (1, K)$:
 - 加权随机选取一个推荐源
 - 从推荐源中加权随机选取一个Item

例子



新推荐源

- 初始反馈值
- 全局平均值
- 经验估计值
- 自动过渡到用户实际反馈值

加强学习

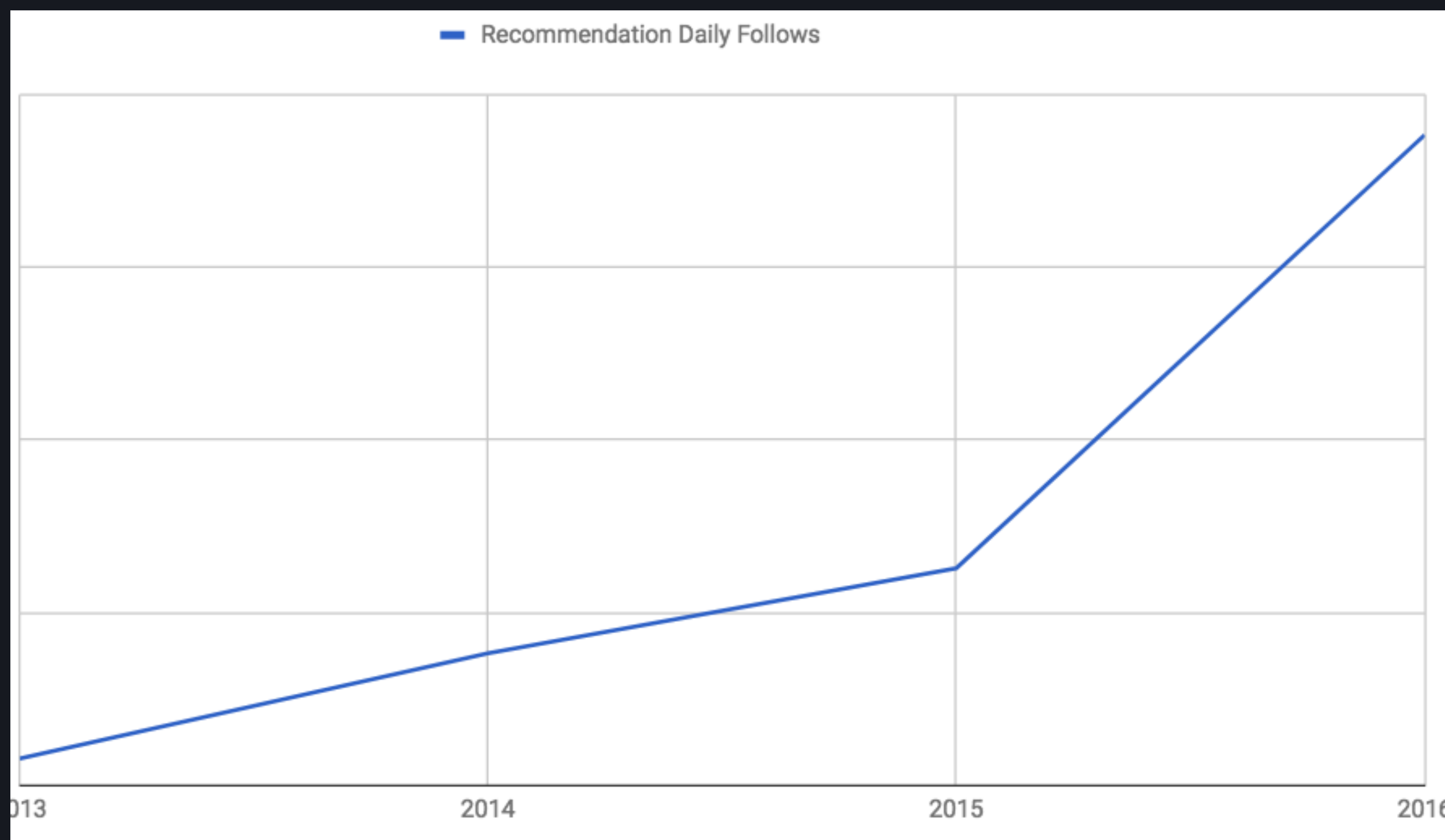
- Multi-armed bandit
 - 根据每个推荐源的效率，自动提高或降低推荐源的使用次数。

TABLE OF CONTENTS 大纲

- Tumblr 推荐系统介绍
- 设计目标
- 相关性反馈和加强学习
- 几个实际问题
- 效果和总结
- Q&A

实际效果

- 快速迭代
- 发布 150 推荐版本
- 推荐效率增加20倍
- 推荐贡献： 1% -> 50% 的移动Follows



其他应用领域

- 搜索
- 电商
- 广告
- 相关内容

总结

- 通用性，即插即用，快速迭代
- 推荐源之间零耦合
- 利用用户反馈加强学习

联系方式

- Email: beitao.li@gmail.com
- 微信: [beitaoli](#)

THANK YOU

如有需求，欢迎至 [讲师交流会议室] 与我们的讲师进一步交流

