

微博机器学习介绍



Weibo Machine Learning

微博研发中心 - 机器学习团队

CONTENTS

微博机器学习团队

01 | 微博简介

02 | 机器学习团队业务

03 | 机器学习关键技术

微博—中国最大的社交媒体平台

- 微博是一个基于社交关系和社交媒体的信息分享、传播以及获取的平台
- 微博建立一个以内容为纽带，连接人、人、人和组织、人和物、人和兴趣的社交媒体网络



用户规模



日活跃用户

2 亿



月活跃用户

4.62 亿



移动端占比

93%

* 以上数据基于2018年财报

CONTENTS

微博机器学习团队

01 | 微博简介

02 | 机器学习团队业务

03 | 机器学习关键技术

微博机器学习团队

微博机器学习研发部是一个高度国际化的专业算法团队，主体设立在北京总部，同时在北美硅谷和杭州研发中心也设立了分支。国内部分注重前沿落地应用到业务，北美研发中心注重引进硅谷先进智慧，合力创建世界一流水平的算法团队。

微博机器学习团队

业务应用



关系Feed



兴趣Feed



Push



访客



同城



榜单



正文页推荐



机器学习 算法



排序



计算机视觉



NLU



语音

机器学习 架构



分布式计算



高性能计算



基础架构



大数据



大规模集群



北京总部



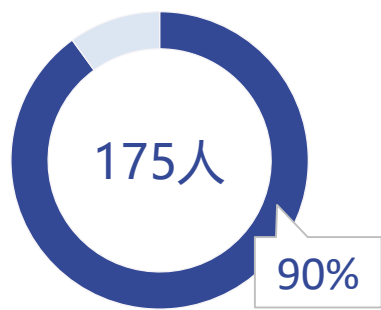
杭州机器学习



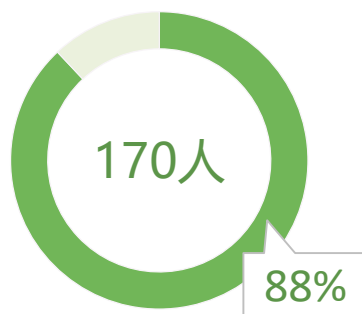
北美研发中心



机器学习团队在职人员教育背景



硕士及以上学历

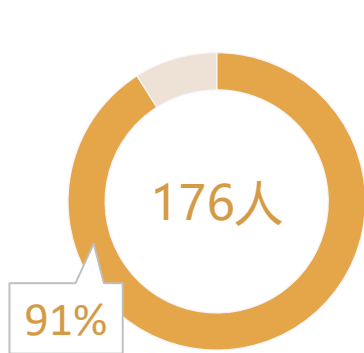


985/211院校

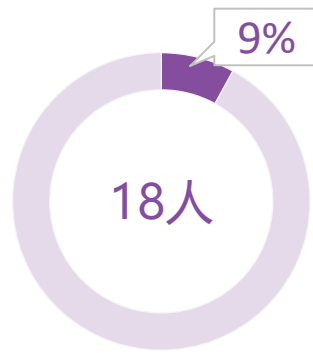


海外教育背景

机器学习团队在职人员工作背景

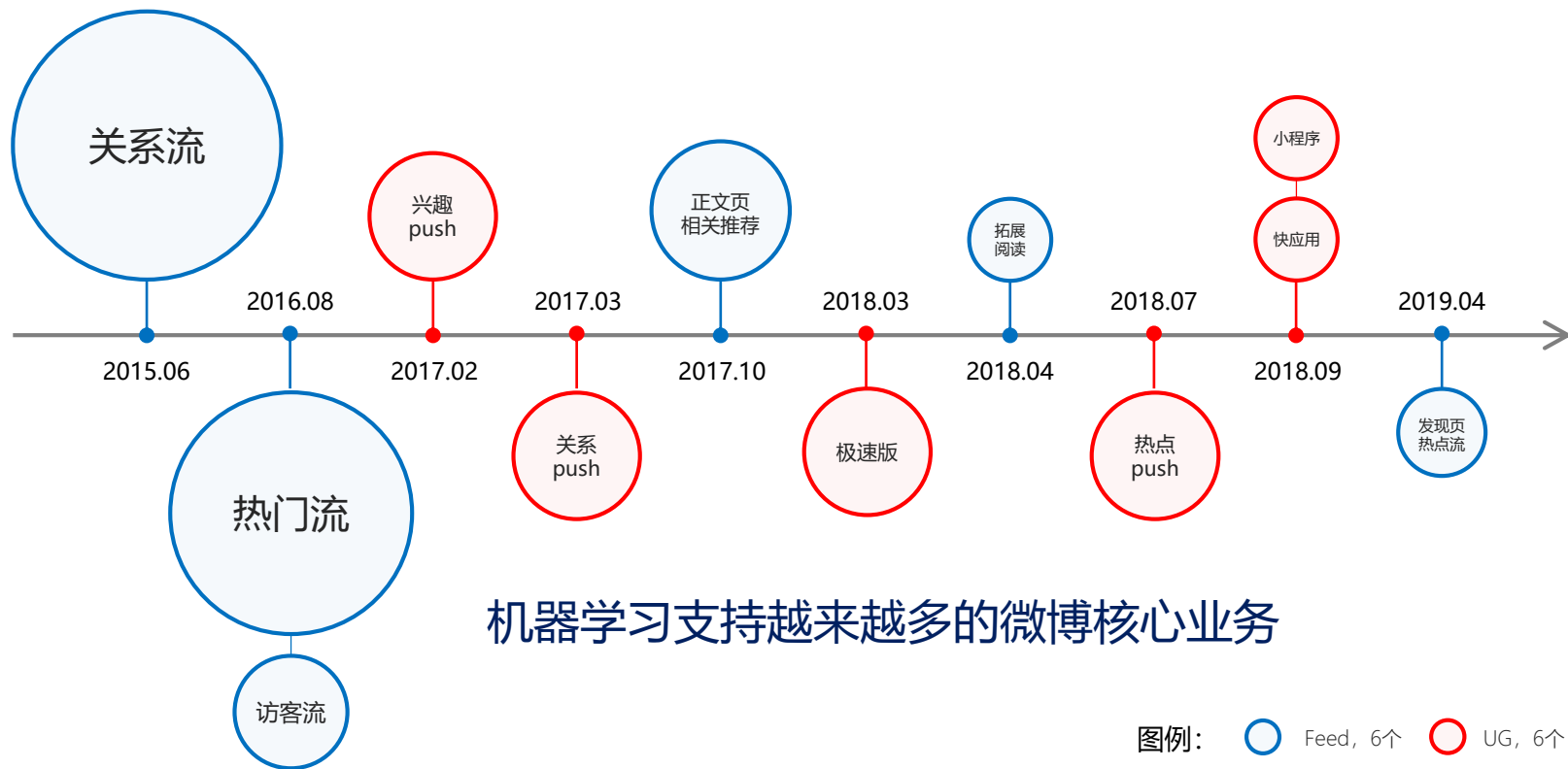


知名企业工作背景



海外工作背景

机器学习在微博发挥越来越重要的作用



微博机器学习——热门微博

- 热门微博是基于用户兴趣进行内容分发的个性化信息流。
- 机器学习技术在热门微博的快速发展中起到了至关重要的作用。
- 基于超大规模样本和特征的新一代CTR预估模型成为了热门微博推荐算法中的核心技术。



微博机器学习—关系流

- 微博关注流是以关注关系为纽带连接用户与内容的订阅式信息流。
- 关注流排序模型以内容、用户、关系、行为等海量数据为基础，采用超大规模机器学习算法连接用户与信息，提升用户每天百亿次曝光的内容消费效率，增强用户黏性。
- 已将LR+GBDT、大规模FM、深度学习等算法成熟地应用于排序业务中，并取得了显著的业务效果。



微博机器学习—用户增长

- 用户增长业务作为公司的战略级项目，承担着获取新用户、激活老用户、提升留存并最终推动微博用户的快速增长的责任。
- 团队不断应用前沿算法，训练微博用户增长模型，有效提升了微博用户从拉新、留存、活跃的全生命周期的阅读体验。
- 在当前日益激烈的竞争环境下，以大数据+深度模型的新型用户增长战略正在持续优化。



微博机器学习—技术里程碑



CONTENTS

微博机器学习团队

01 | 微博简介

02 | 机器学习团队业务

03 | 机器学习关键技术

机器学习团队—技术方向



涵盖多个人工智能技术方向

机器学习支持完整的微博内容生态链

微博内容生态



以内容消费与分发为中心，加强了内容理解，开始投入技术助力内容生产

内容消费与分发

通过人工智能技术，对多样性内容进行深度理解及智能分发，达到「千人千面」的效果

多样性内容



智能分发

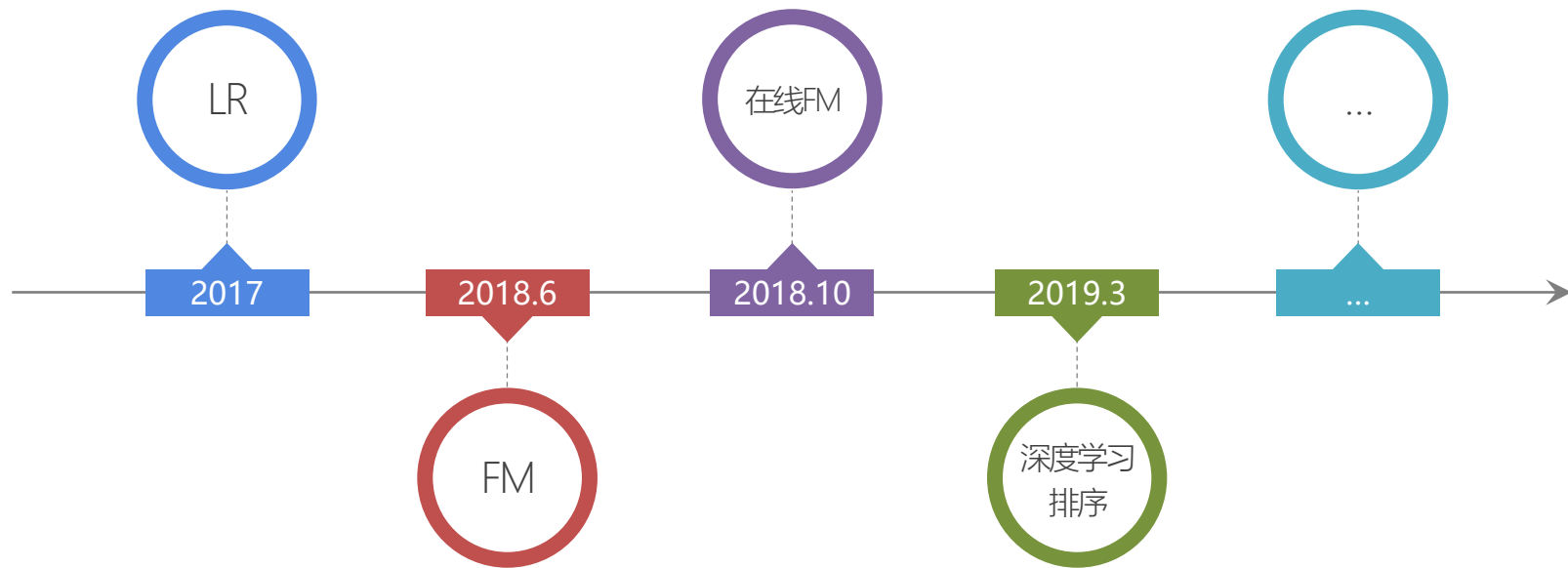
- 微博
- 视频
- 图片
- 文章
- 话题
- 问答



千人千面



推荐技术的进化



= 推荐技术的创新

我们积极探索推荐新技术，除了业务落地外，还在一流国际会议发表了相关成果：

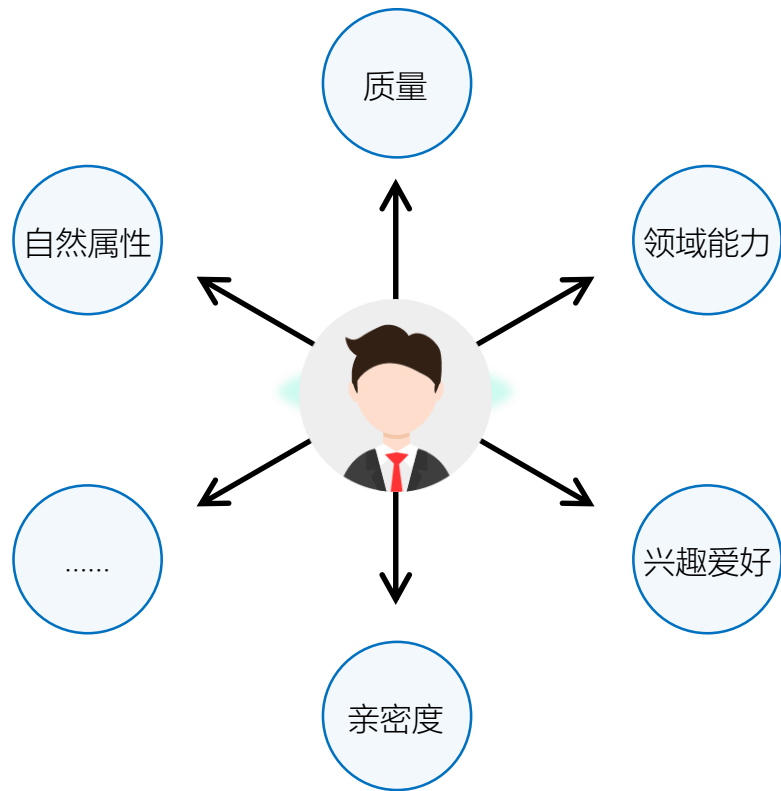
1.RecSys 2019(推荐领域最好的国际会议,The ACM Recommender Systems Conference)

FiBiNET: Combining Feature Importance and Bilinear feature Interaction for Click-Through Rate Prediction, Tongwen Huang,Zhiqi Zhang, Junlin Zhang

2.ICDM 2019

FAT-DeepFFM: Field Attentive Deep Field-aware Factorization Machine, Junlin Zhang,Tongwen Huang,Zhiqi Zhang

用户理解—用户画像



覆盖4亿MAU



自然属性



LBS位置



职业情况



教育背景



增值活跃



汽车偏好



游戏偏好



音乐偏好

内容理解—文本图片理解

文本理解

图片理解

内容分类

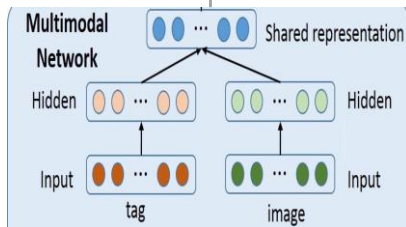
吸引力

时效性

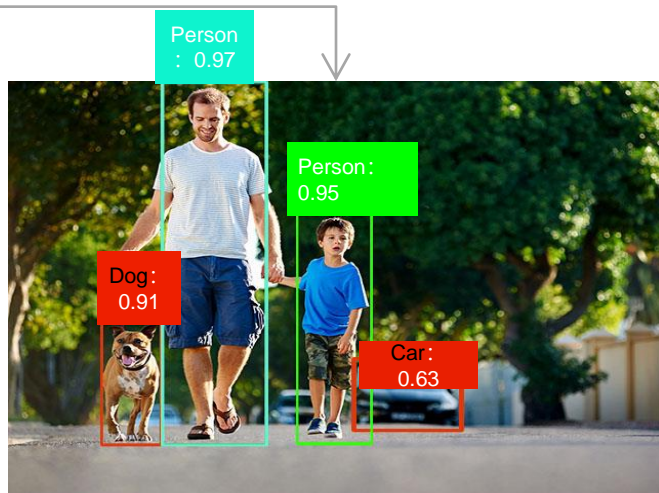
质量分析

垃圾广告识别

情感分析

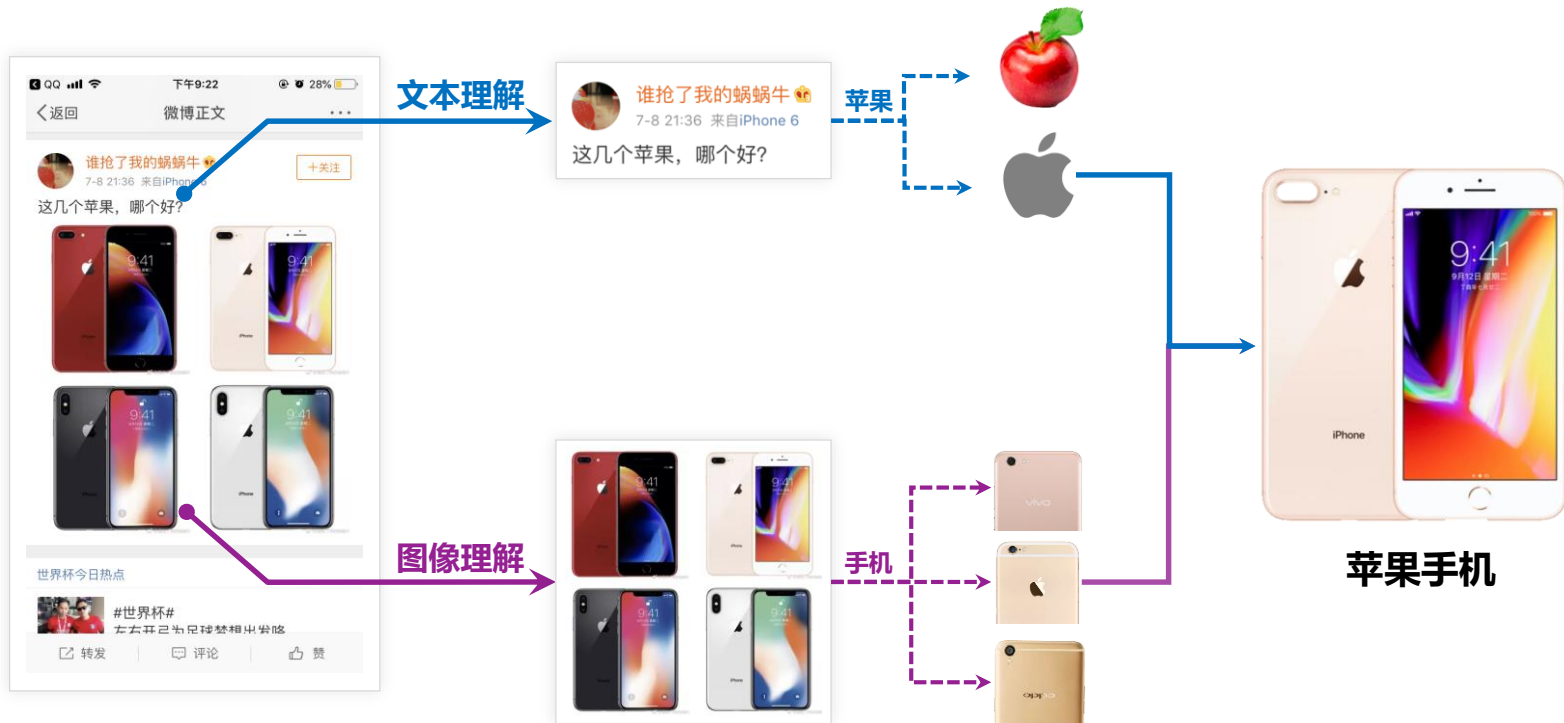


多模态融合更好地理解内容



内容理解—多模态融合

>> **多模态融合** 文本、图像视频等更好地理解内容标签 <<



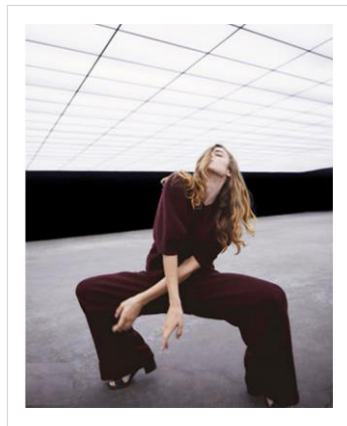
内容理解—视频理解

>> 文本信息量少，加强**视频理解**，提升视频推荐精准度和播放时长 <<

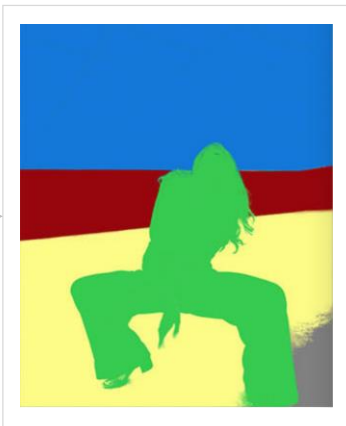


内容生产

>> 尝试探索技术促进内容生产<<



语义分割



背景替换



举例：人物背景替换

技术影响力 (2018年)

01 走出去

- 参与会议48人次
- 海外会议2场
- 国内会议17场

02 对外输出

- 对外演讲38人次
- 技术媒体发表文章10余篇
- 多次荣获“明星讲师/优秀出品人”荣誉
- 涵盖“全球人工智能与机器学习大会 (AICon)”等工业界一流会议

03 请进来

- 邀请业界知名专家交流30多次
- 嘉宾来自于Facebook、谷歌、阿里、北大、清华、中科院等国际国内著名互联网公司 & 著名高校
- 每周1次“七点夜校”内部培训

感谢您的聆听



Weibo Machine Learning

微博研发中心 - 机器学习团队