



全球

World Of Tech 2017

2017年12月1日-2日 · 深圳中洲万豪酒店

软件开发技术峰会

DEVELOPMENT



# 当AI遇到信息流

新时代的信息分发

**刘彦东**

网易传媒 技术副总裁

# AI的时代已经到来

- 用户数据仍在快速增长
- 机器学习已经成熟，深度学习方兴未艾
- 计算能力不断增强: GPU, TPU
  - TPU is 15x ~ 30x faster than GPU
- AI可以渗透的场景无处不在

# 什么是信息流 ( News Feed )

- 以推荐为基础，将信息呈现在用户面前
- 内容是多样的
- 场景极其丰富
- 信息选取是被动的
- 重新定义人和信息的连接



3

**信息流全都满足，得到了极大的发展**



# 前信息流时代

- 主动获取内容
  - 搜索引擎
  - 门户
  - 直接访问内容网站
- 信息单一
- 效率低
- 千人一面



# 信息流的优势

- 理想的AI场景化应用
  - +AI
  - 内容生产已经持续了几十年
  - 大规模分发需要AI
- 算法可以复制，数据却不可多得
- 体量可以变得很大
- 商业变现: 精准定向广告

# 信息流的形式





# 信息流产品的体量

- 已有数家大公司入场
- 今天有几千万的日活跃用户
- 每天新增几十万篇内容
  - 精品原创
  - 海量自媒体内容(图文, 短视频)
  - 社区, 直播, 线下活动, etc
- 时长在几十分钟

海量用户+长时长 = 海量用户数据

# 信息流的价值

- 连接用户和内容，解决信息过载
- 信息获取更加便捷
- 泛娱乐内容 -> 填补碎片时间



# 如何定义成功的信息流产品？

- 推荐算法为核心->预测用户需求
- 过去人找信息（网站访问，搜索引擎）
  - 片面的，单点信息需求
- 现在是全面的,个性化的信息满足



# 信息流产品的构成

- 内容生产，过滤，加工
- 推荐算法为核心，理解用户并把合适的内容在合适的时间传递给用户
- 收集用户反馈改进模型
- 精准定向广告
- 重视生态建设
- 工作很综合
- 难度很大





# 内容形式

- 文本：主题，人物，事件，文章质量，文章热度，etc
- 图片/视频：清晰度，类别，语义，低俗率，是否有广告，etc
- 音频：转化为文本
- 拓展：知识图谱





# 内容处理

- 高效的处理、分析、挖掘、理解、组织内容
- 海量图片、视频、文字、音频
- 充分理解内容
  - 文本分析
  - NLP
  - Computer Vision
  - Data Mining
  - 知识图谱



# 理解用户需求

- 内容形式：图文，视频，短咨询，音频？
- 内容属性：要闻，娱乐，体育赛事...
- 其他要求：高清视频？流量限制？
- 场景需求：天气状况？路况？体育赛事结果？

# 建立用户画像

用户画像用来刻画用户

## 用户行为画像

根据用户历史行为数据，为用户打标签

## 上下文画像

根据用户当前访问行为为用户打标签

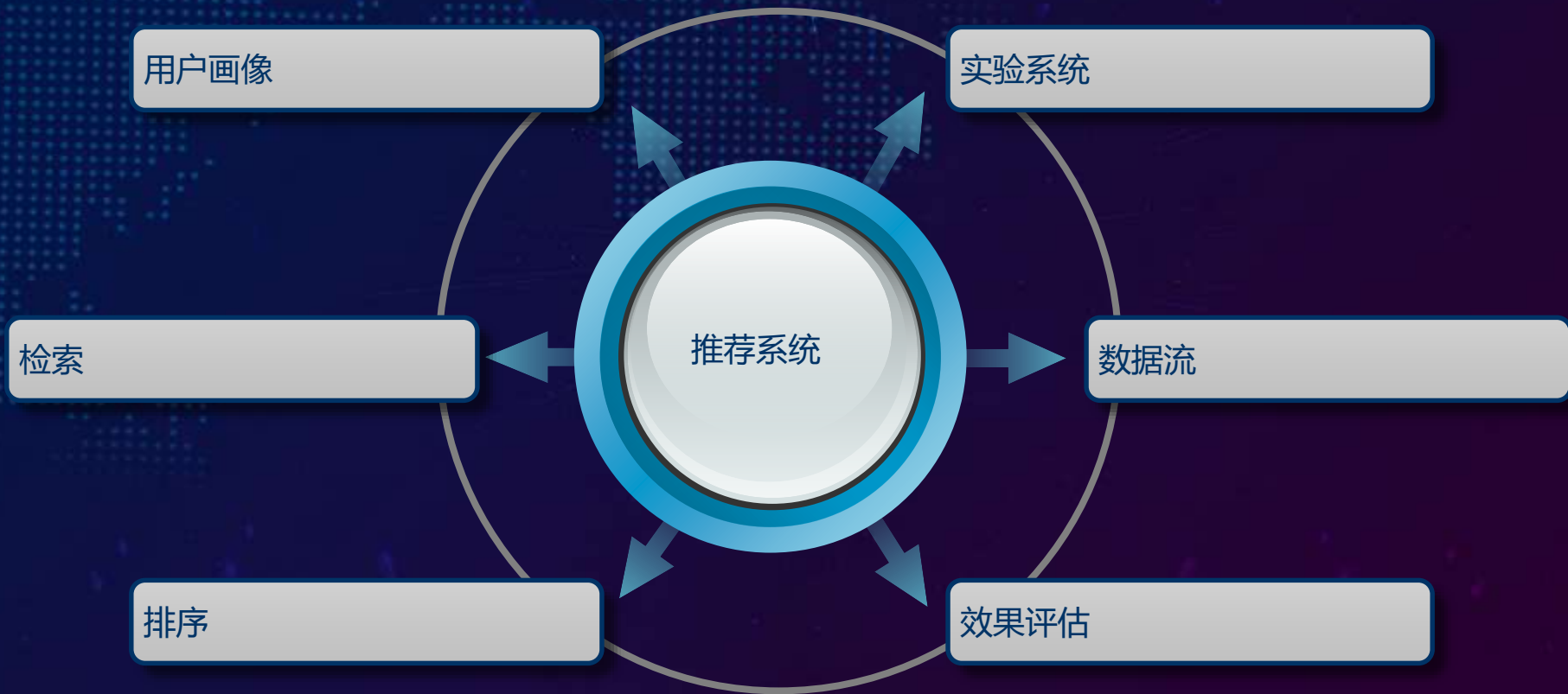
## 人口属性画像

描述年龄，性别，教育程度等标签

# AI算法是核心

- 基于用户兴趣的推荐引擎
- 用户建模
  - 性别，年龄，收入，职业
  - 阅读历史，兴趣点，阅读时间
- 上下文：
  - 时间地点，网络条件，天气...

# 推荐系统架构





# 实验迭代提升效果

## 离线实验

- 1 通过日志获取用户行为
- 2 将数据按规则分成训练集和测试集
- 3 训练集训练模型，测试集预测
- 4 离线指标评测测试集的预测结果

## 用户调查

用户调查需要一些真实用户在待测系统上完成一些任务，观察并记录他们的行为，通过分析这些行为了解推荐效果

## 在线实验

在推荐系统上做AB测试，将用户根据一定规则分成不同组，不同组采用不同算法

# 信息流生态培养

- 内容是起点，用户是终点，形成闭环，互相促进
- 大量内容来自PGC/UGC
- 内容创作、审核、过滤
- 通过人工标注、人工运营来促进AI的应用
- 通过用户反馈正向引导内容生产

# 信息流的未来

- 形式会更加多样
  - 图文
  - 视频
  - 音频
  - 知识问答
- 粘性越来越强
- 会是用户获取信息的最重要的渠道之一

# 场景会更加细分

- 晨读/夜读
- 天气
- 体育赛事结果
- 重大要闻
- 追剧
- 打发时间
- 路况信息
- ...



# AI会发挥更大价值

- 随着深度学习的进一步发展，AI会更“好用”
  - Scalability
  - 推荐精度进一步提升，场景判断更准确
  - 内容理解更全面（图像，视频，音频）
  - 理解更深入：视频里的人物，故事情节
- AI将渗透到更多的方面中去
  - 内容过滤，分类
  - 机器人写稿
  - 机器人互动
  - ...



# 机器学习、深度学习会扮演更大角色

- 图像，视频分类
  - 深度学习(CNN, RNN, etc)
- 明星脸识别
- 语音识别
- 三俗，不适内容识别
- 从图文行为推荐视频
  - 迁移学习
- 智能写稿

Thank you!