

人工智能驱动的内容生产与分发

马艳军 博士



AiCon

全球人工智能与机器学习技术大会

助力人工智能落地

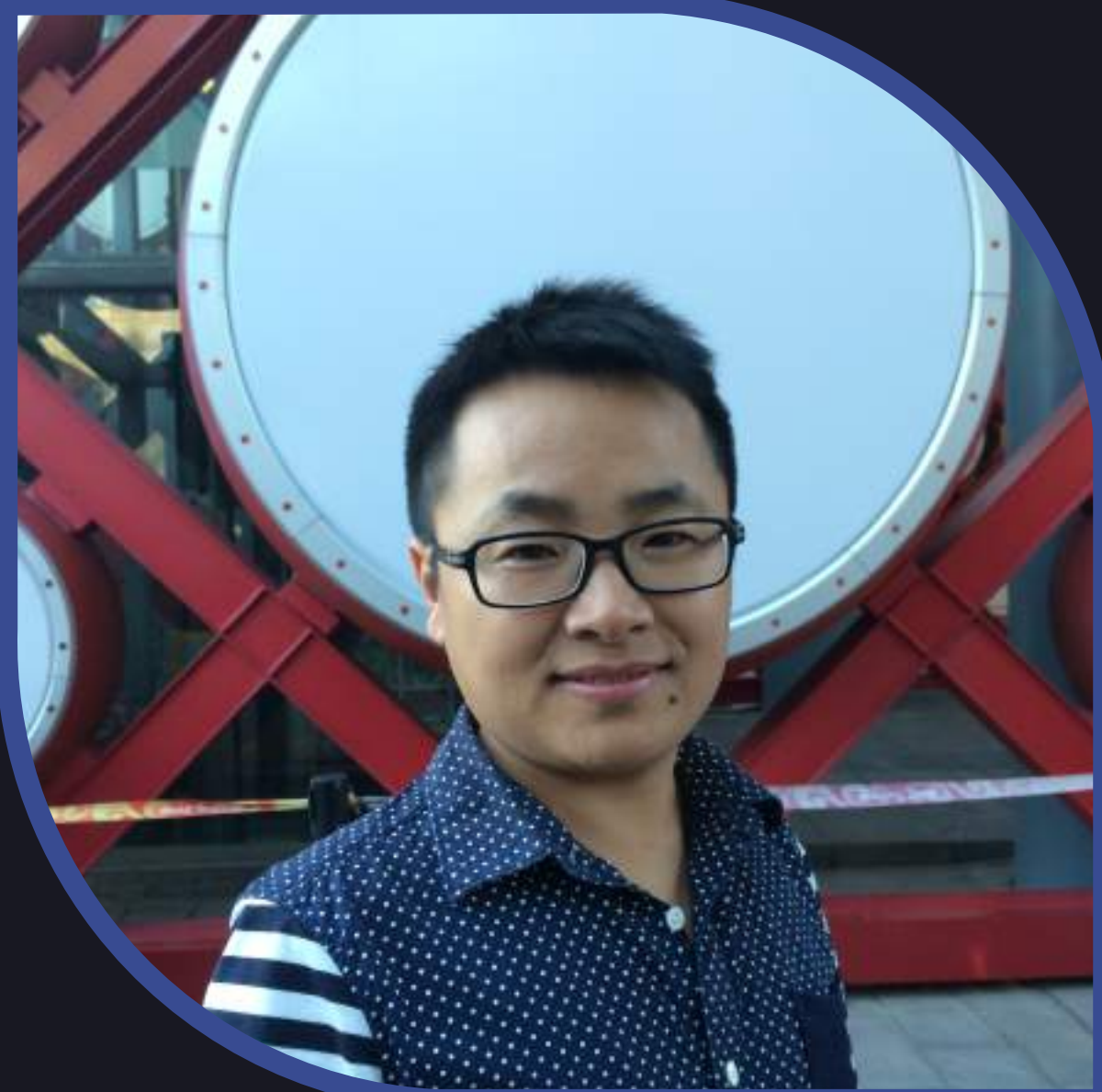
2018.1.13 - 1.14 北京国际会议中心



扫描关注大会官网

SPEAKER

INTRODUCE



马艳军

百度 首页业务部主任研发架构师

马艳军博士，现在在百度主要从事自然语言处理、机器翻译、推荐技术的研发工作，相关研发成果广泛应用于百度搜索、信息流等产品中。

马博士曾参与欧盟第七框架计划（FP7）、863计划等国内外项目，发表论文20余篇，申请国内外技术发明专利20余项，并多次担任ACL，IJCAI等国际权威会议领域主席和审稿人，相关成果曾获得2016年国家科技进步二等奖。

提纲

- 内容消费的行业现状
- 人工智能辅助内容生产
- 内容质量控制模型
- 内容分发/推荐系统

- 内容消费的行业现状

什么是内容消费？

一种直接或间接以**内容产品**和**内容服务**为消费对象的经济活动

媒介独立

内容传播的媒介相对独立，传播媒介的属性限定了内容的类型，如视频类内容几乎只能在电视频道上传播，供需调节难

角色单一

内容生产成本高；图文内容、视频内容的生产者之间有很大的壁垒，内容的供给以生产者为主，内容产量相对低

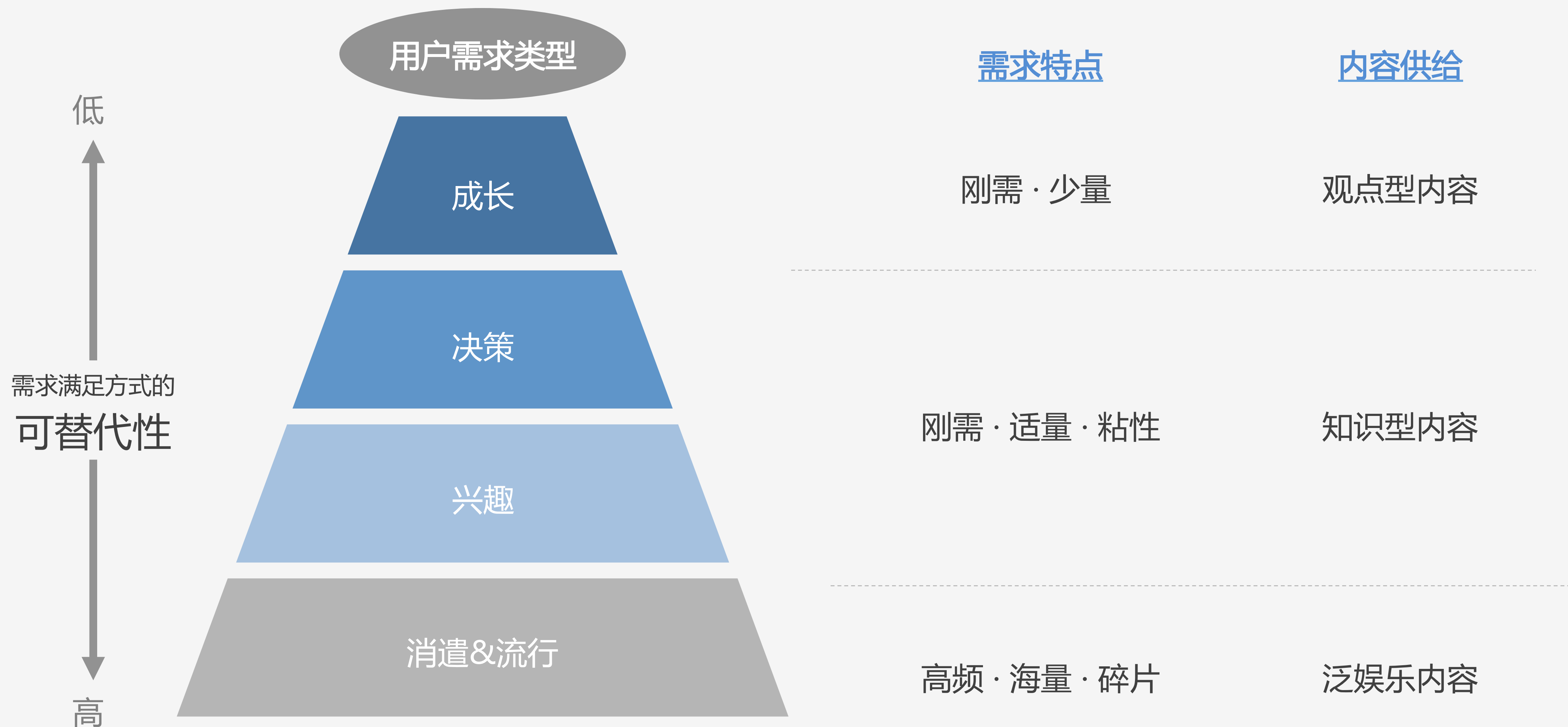
单向消费

内容从生产者、传播渠道、到用户，全程单向传播，一次内容消费的完成就意味着一次传播行为的终结



内容消费满足了用户的什么需求？

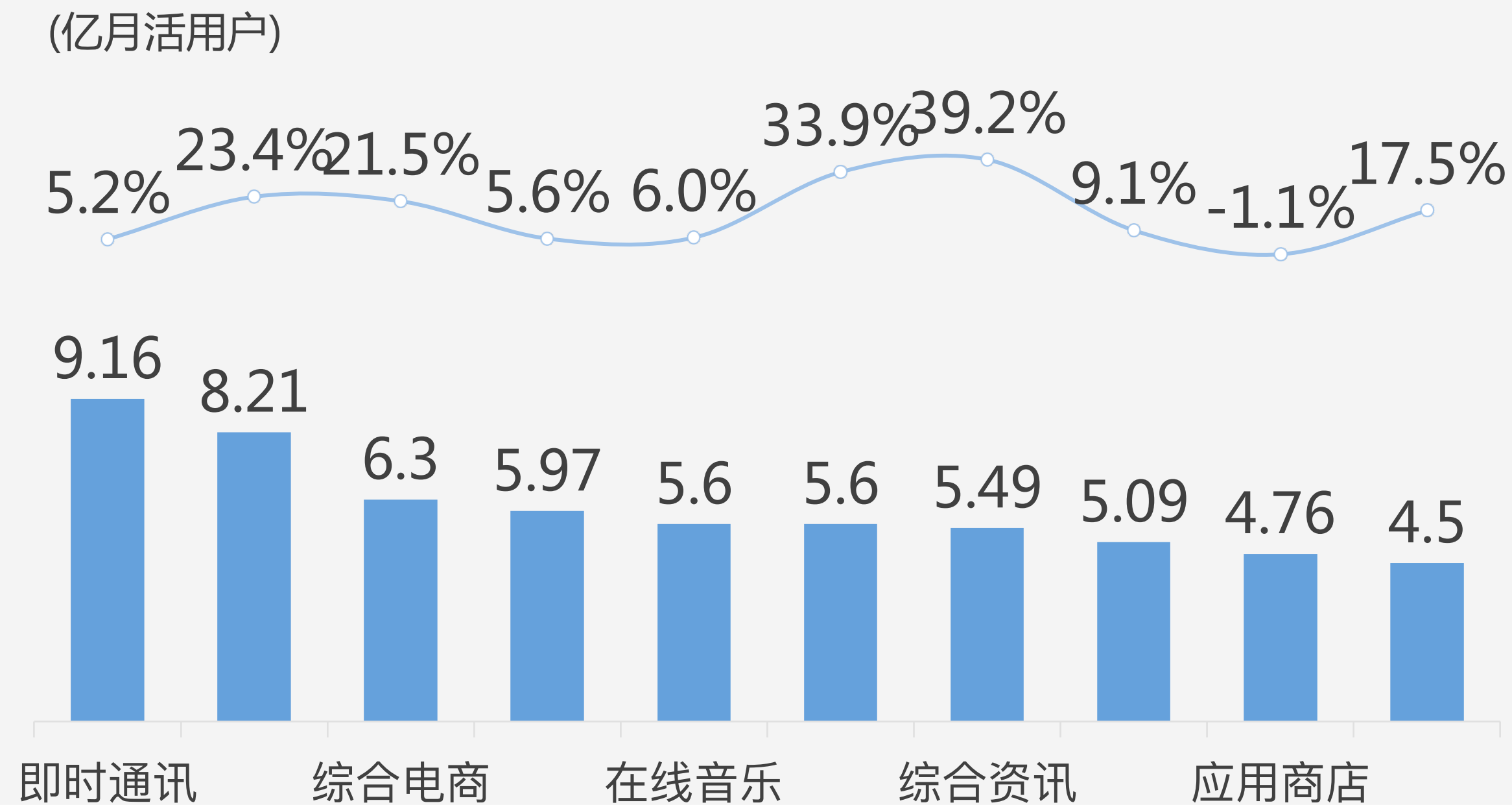
覆盖每一个用户生活的方方面面



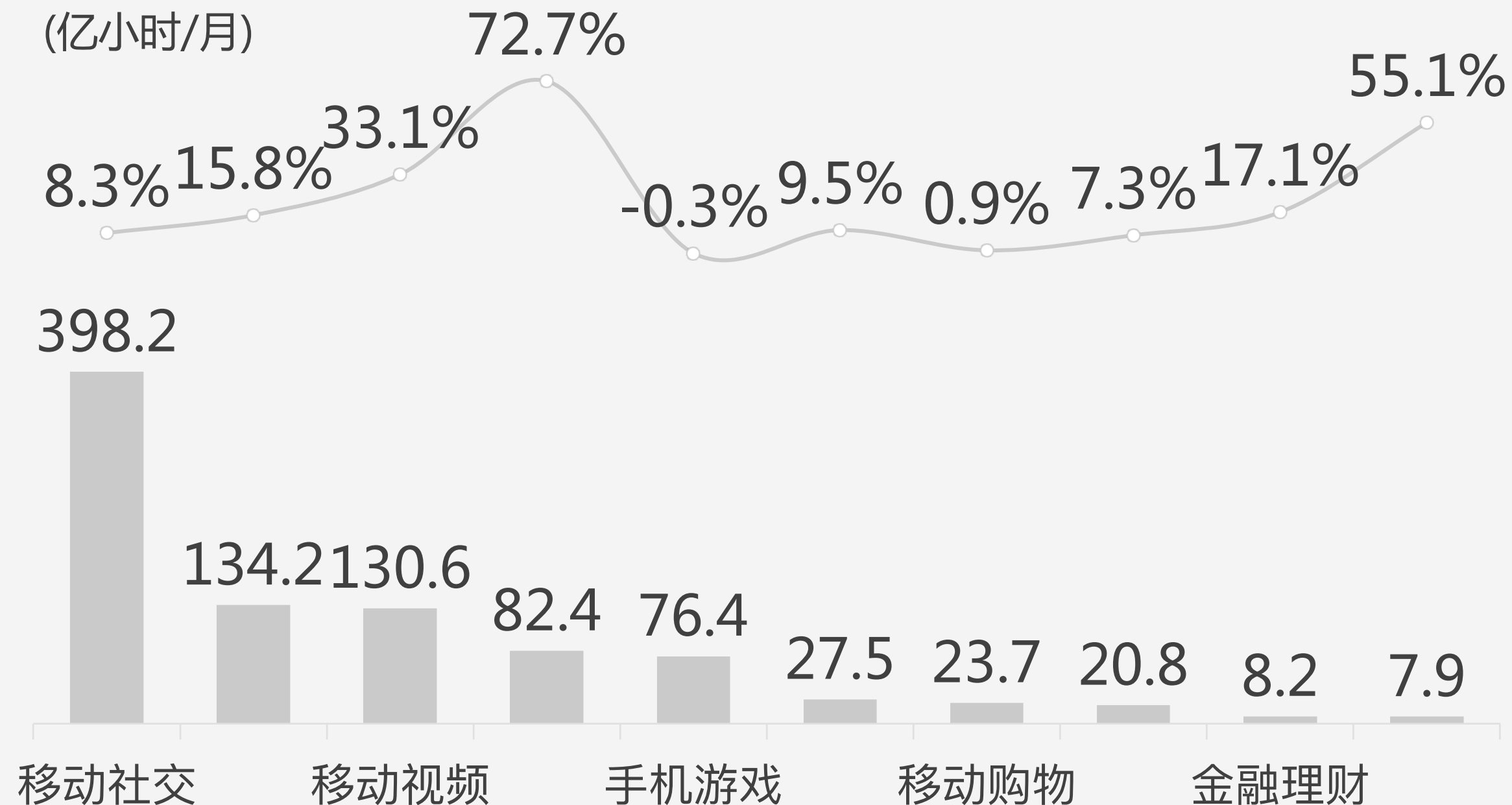
内容消费为什么重要？

1. 内容消费已经发展成为了一个影响所有人日常生活的庞大产业

Top 10行业用户规模及同比增长率 (2017年6月)



Top 10行业用户总时长及同比增长率 (2017年6月)

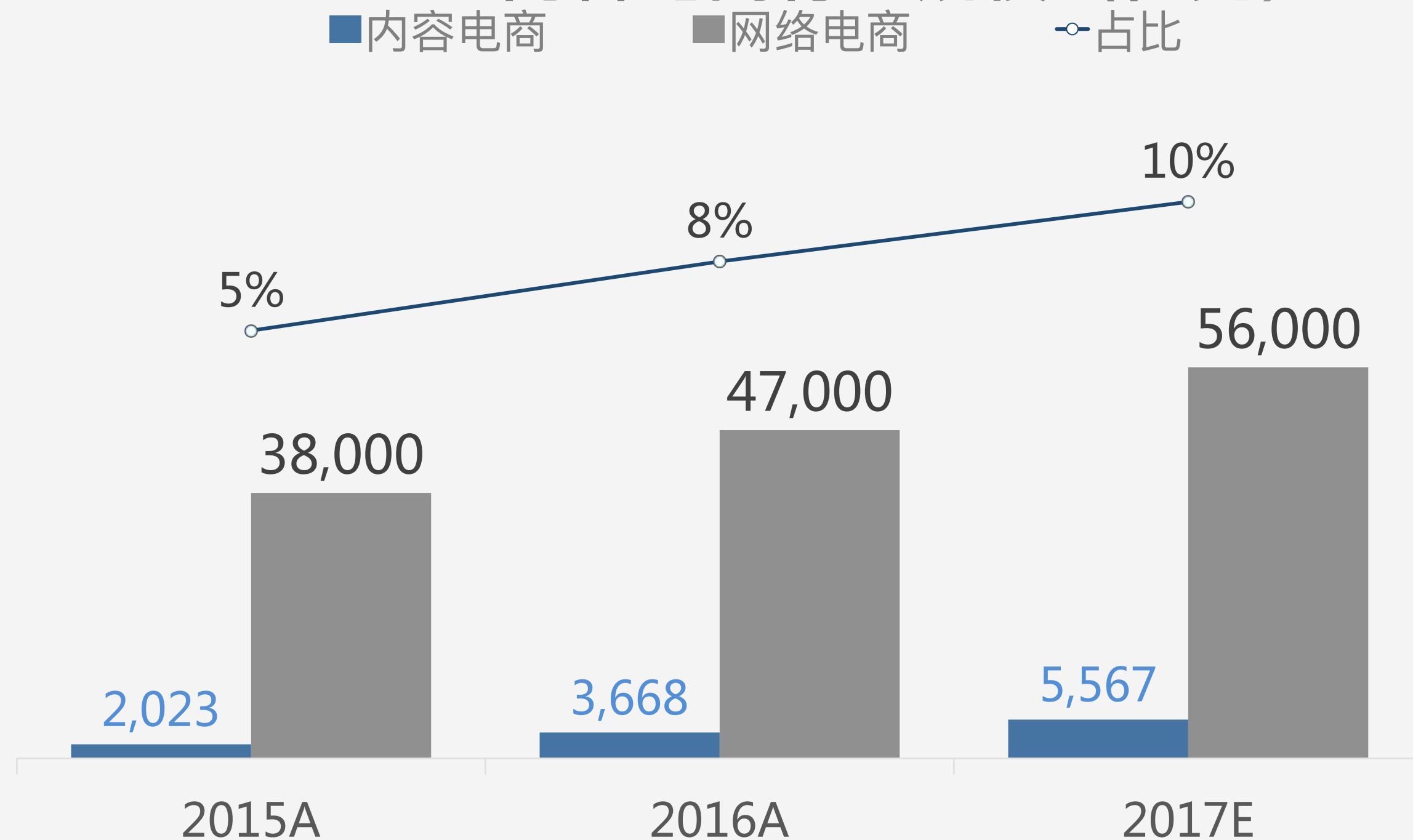


数据来源：Top10行业用户规模、用户总时长和同比增长率来源于Quest Mobile的《移动互联网2017年Q2夏季报告》

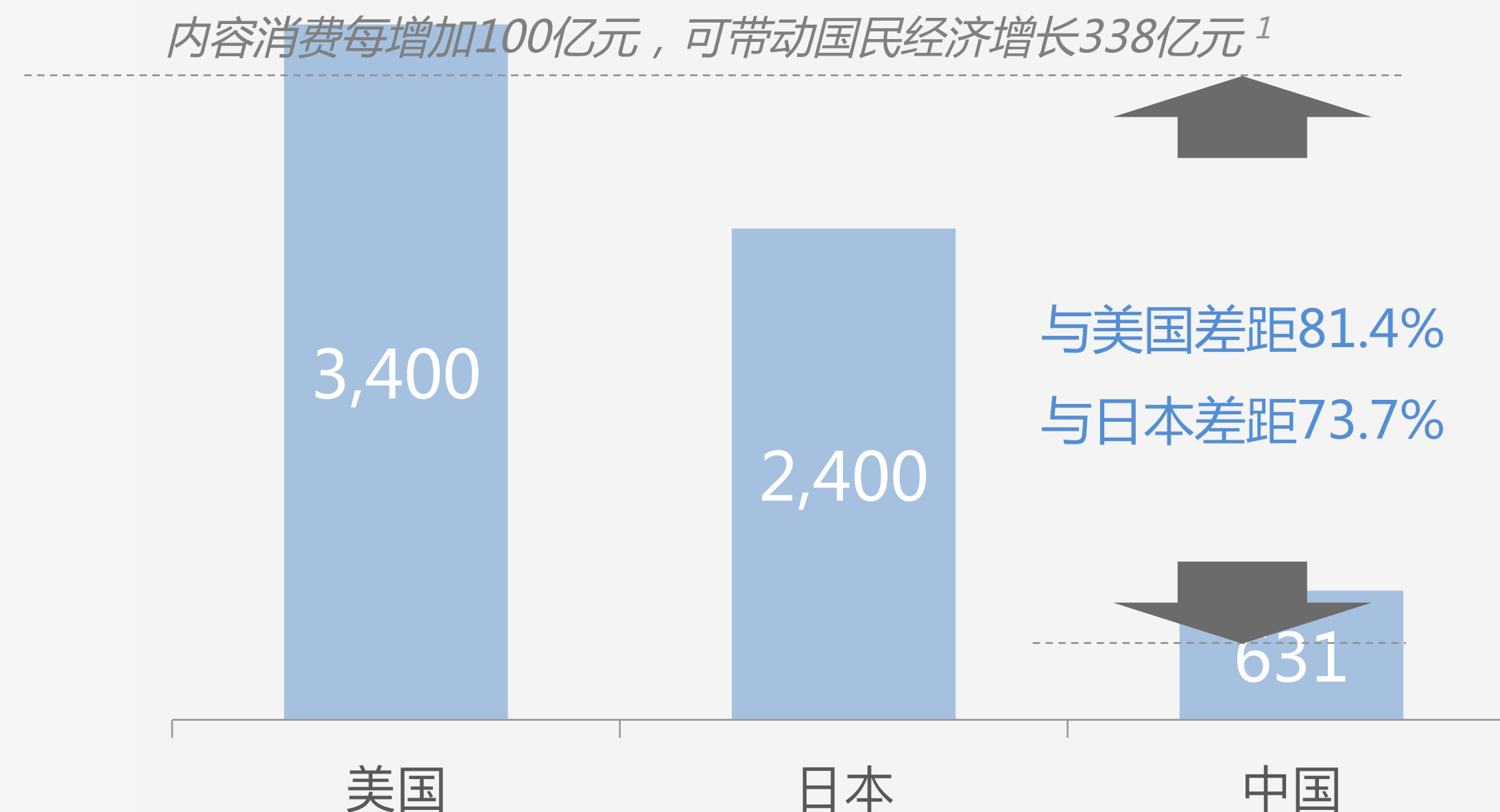
内容消费为什么重要？

2. 更是国民经济杠杆，关乎社会主流价值观的塑造

2015-2017内容电商行业规模（亿元）



人均信息消费（美元/年）



数据来源：1) 2013年工业和信息化部电信研究院数据；2) 人均信息消费：世界银行及恒丰银行研究院商业银行研究中心统计数据

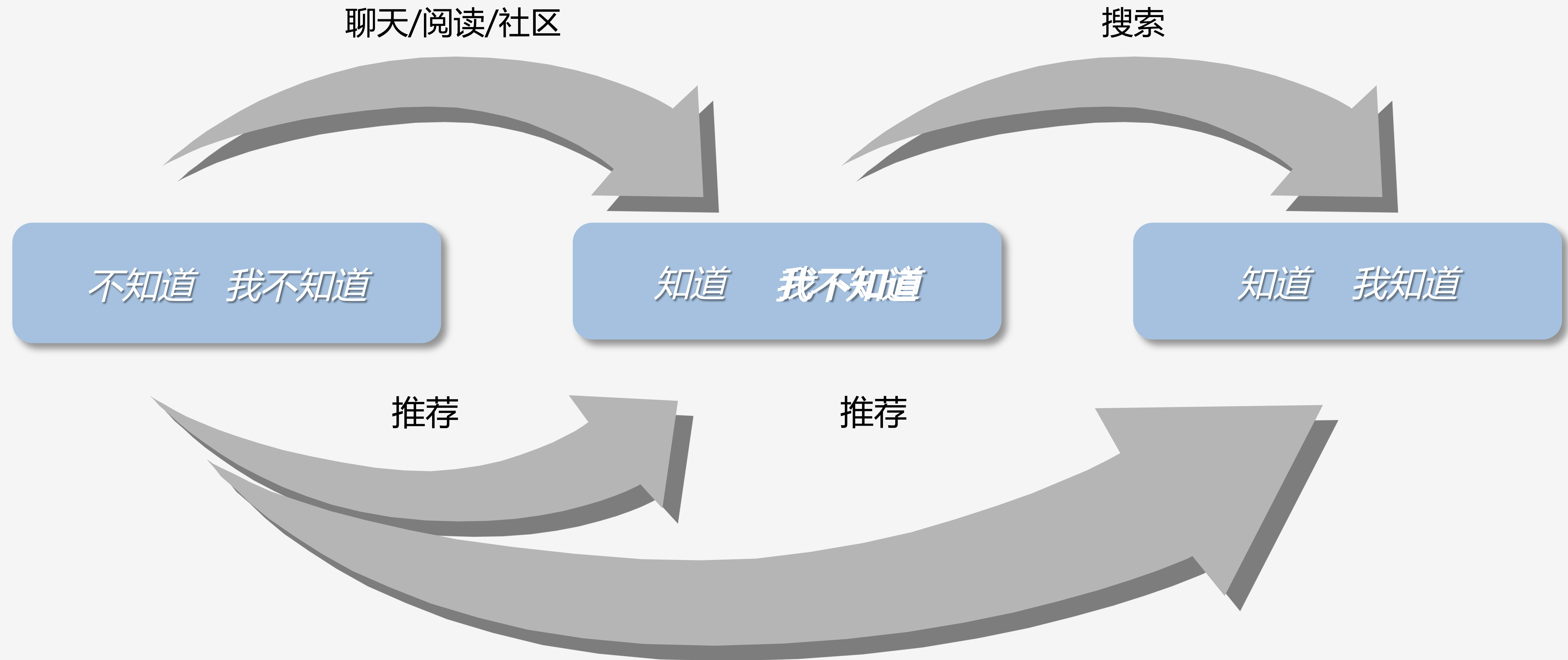
内容消费行业地图

人工+机器，多种消费形态组合



内容消费的主要形态

从搜索到推荐，是内容获取方式的必然演进



内容消费的三大变化趋势

媒介融合 · 角色融合 · 消费升级



媒介融合

- 一个媒介上可以获取多种内容，同一内容也可以在多个媒介上分发
- 获取内容的方式大发展，从传统的人找内容，发展到内容找人



角色融合

- 生产门槛降低，人人都是生产者，人人都是消费者，内容生产者猛增
- 从专业生产（PGC）发展到社会化生产（UGC），内容海量爆发



消费升级

- 从关注到最终的分享可循环，内容散落在消费各个环节，影响力扩大
- 消费结构变化，用户需要更多专业垂直、精细、多元化的内容

需要解决的问题

作者深度赋能

优质内容识别

精准个性推荐

百度智能驱动的内容消费生态

始于2016年6月

精准个性化推荐及推送

- 搜索：从关键字，到语音、图片交互
- 个性化推荐：百度特有的超大规模实时个性化推荐系统，千亿规模参数下的多目标最优化
- 推送：基于用户兴趣点、地理位置、天气等综合因素



定制生产、优质生产、高效生产

- 选题推荐：大数据+知识图谱
- 内容选材：知识图谱+自动配图
- 图片处理：智能裁图
- 文字处理：自动标题推荐+语义纠错
- 视频识别：视频去重+长短关联
- 智能写作：AI智能写作

多维度识别优质内容资源

- 质量先验：基于半监督学习构建对于篇章质量的全方位的理解框架，多维度丰富信息
- 质量后验：基于NLP和CNN，识别用户正负反馈，评价内容质量

搜索+推荐，满足用户内容需求

双引擎

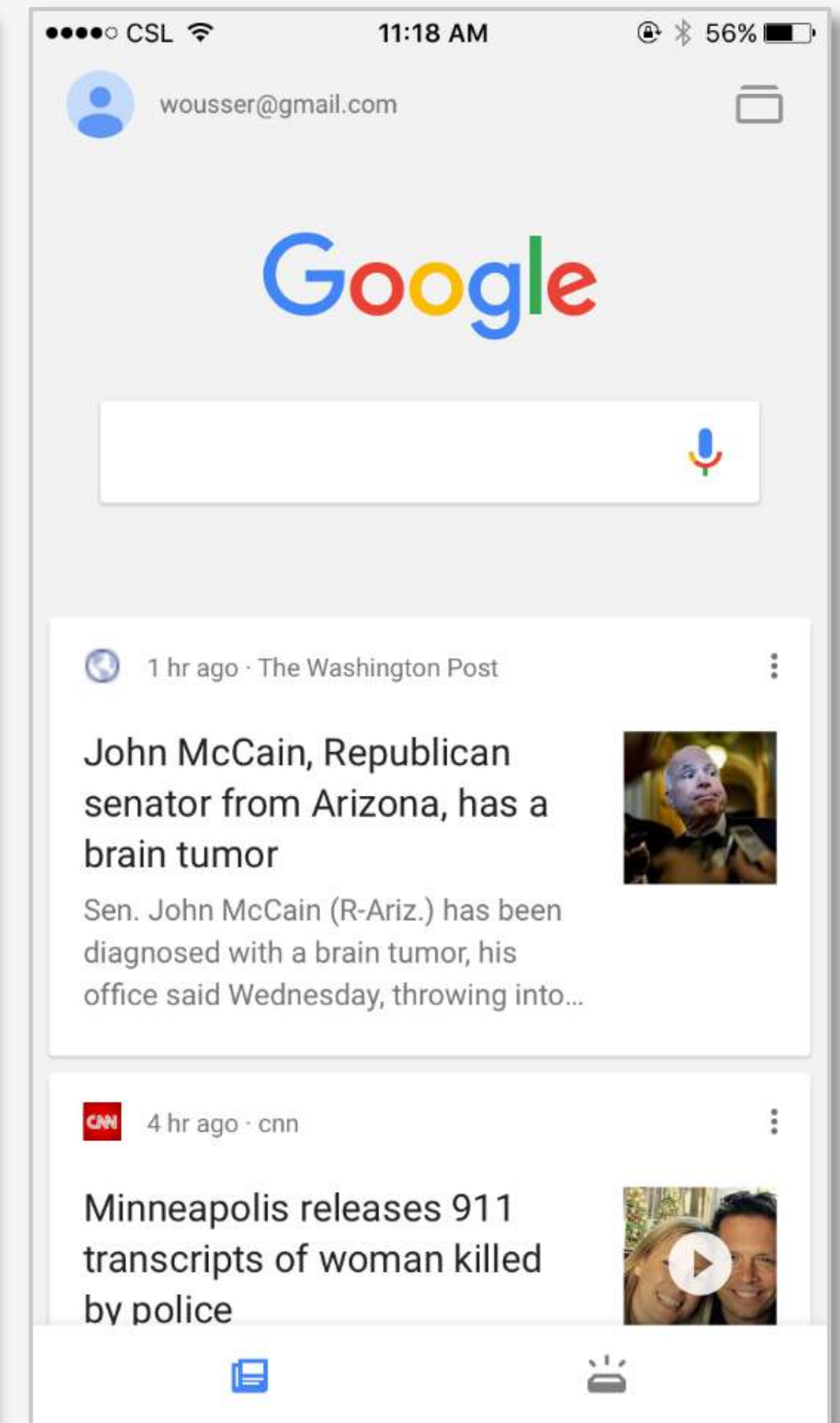
手机百度，“搜索+推荐”双引擎内容分发

极速搜索

一站搜索全网，搜文字、搜图片，快速识别，帮助用户主动获取所需内容

精准推荐

精选资讯、视频、小说、漫画等优质内容，结合智能个性化推荐算法，满足用户潜在内容消费需求



Google Feed, 2017年7月上线

- 人工智能辅助内容生产

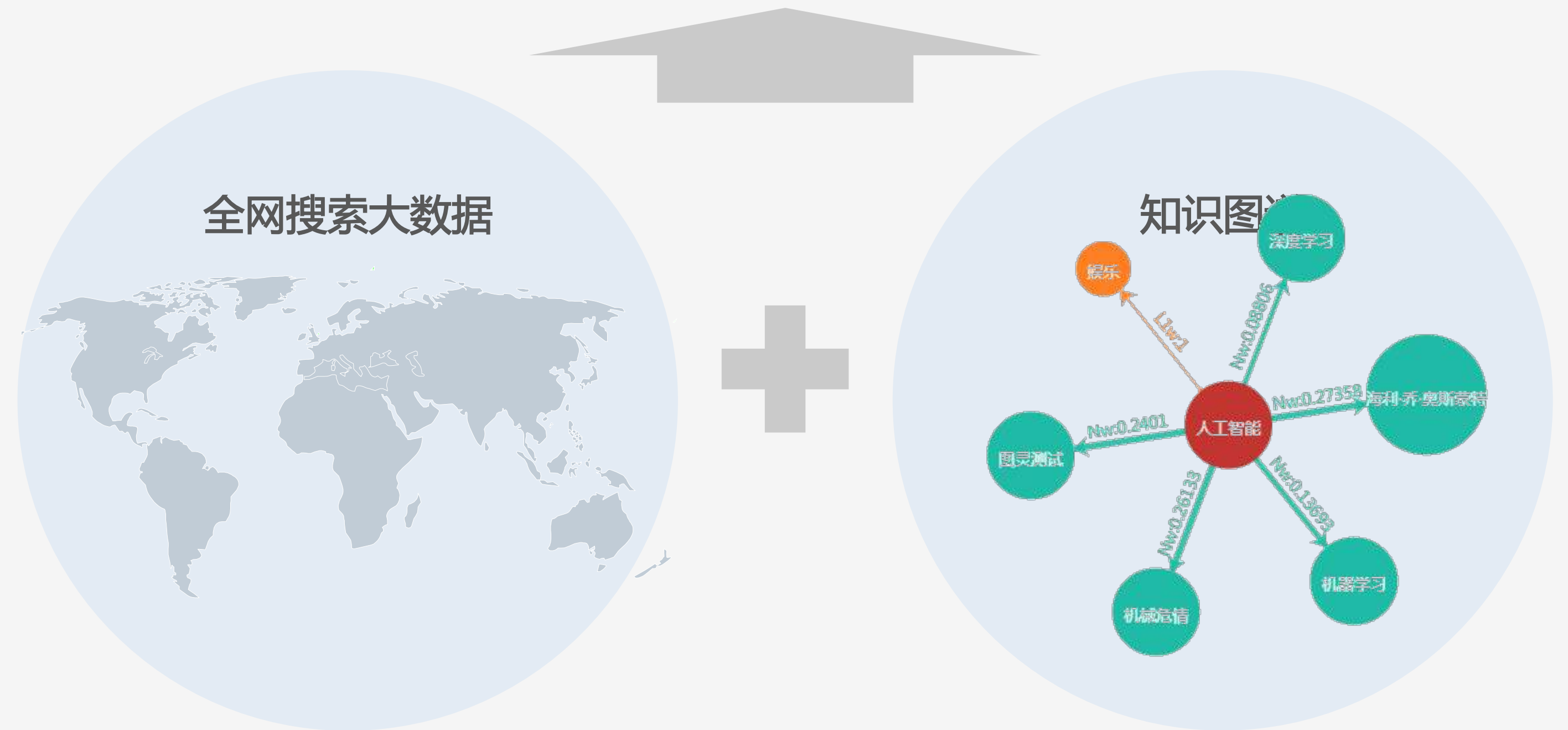
内容生产流程 1

选题推荐 · 内容选材 · 图片处理 · 文字处理 · 视频识别 · 智能写作

实时热点，建议选题，稀缺主题，.....

选题推荐

- 基于7亿互联网民每天60亿次搜索请求的大数据积累，挖掘实时热点事件，结合用户画像和知识图谱技术，智能推荐给相关作者，邀请作者写作，解决作者选题环节的痛点
- 相比普通作者自选题文章平均用户点击率提高2.7倍



内容生产流程 2

选题推荐 · 内容选材 · 图片处理 · 文字处理 · 视频识别 · 智能写作

内容选材

- 基于知识图谱的内容推荐：通过语义分析技术识别作者写作主题，结合百度百科图谱、指数、百科、正版图库等产品，帮助作者发现写作素材

自动配图

- 基于图像识别能力的自动配图功能，根据图像清晰度、美观度、图文相关性等特征选取优质图片
- 模糊图片识别准确率71%，清晰图片识别准确率96%以上；自动配图准确率90%



The screenshot displays a content production interface. On the left, a text editor shows a paragraph about the actor Yang Mi. A large image of Yang Mi is selected in the main content area. On the right, a sidebar provides a knowledge graph for Yang Mi, including a biography, a list of works, a gallery of official images, a Baidu Index score of 60880, and related news and recommendations. The interface also includes a 'Writing Assistant' button and a 'Correction' icon.

正文

作为开年第一部爆款大剧，《三生三世十里桃花》以双台双网收视第一、网... 引得网友感叹：“我去机场大吧上有人看，地铁上有人看，到了机场还是... 杨幂演技的讨论更是层出不穷。一直以来她都人气爆棚却被评价演技一般... 我们完全可以说这次杨幂简直开了挂：她一人挑战了五个角色，并且将这... 象深刻的情节段落。有网友更是评价道：“有如突然开窍了一样，真真把力

杨幂

百科目录

1. 早年经历
2. 演艺经历
3. 个人生活
4. 主要作品
5. 社会活动
6. 获奖记录
7. 人物评价
8. 人物争议

正版图片 [更多>](#)

百度指数 60880 [查看30日趋势>](#)

相关新闻 [更多>](#)

流量女王杨幂:请叫我“人民艺术女演员”
杨幂遇刁钻问题还能微笑回答,难怪能红这么久

相关推荐

- 刘恺威 杨幂老公
- 小糯米 刘恺威杨幂的亲生子女
- 唐嫣 杨幂闺蜜
- 刘诗诗 杨幂闺蜜

内容生产流程 3

选题推荐 · 内容选材 · 图片处理 · 文字处理 · 视频识别 · 智能写作

智能裁图

- 利用基于深度学习的图像处理技术，自动识别图片主要内容区域并进行截取
- 准确率97%以上，百家号文章头图的不合格率降低60%

