

内容生产流程 4

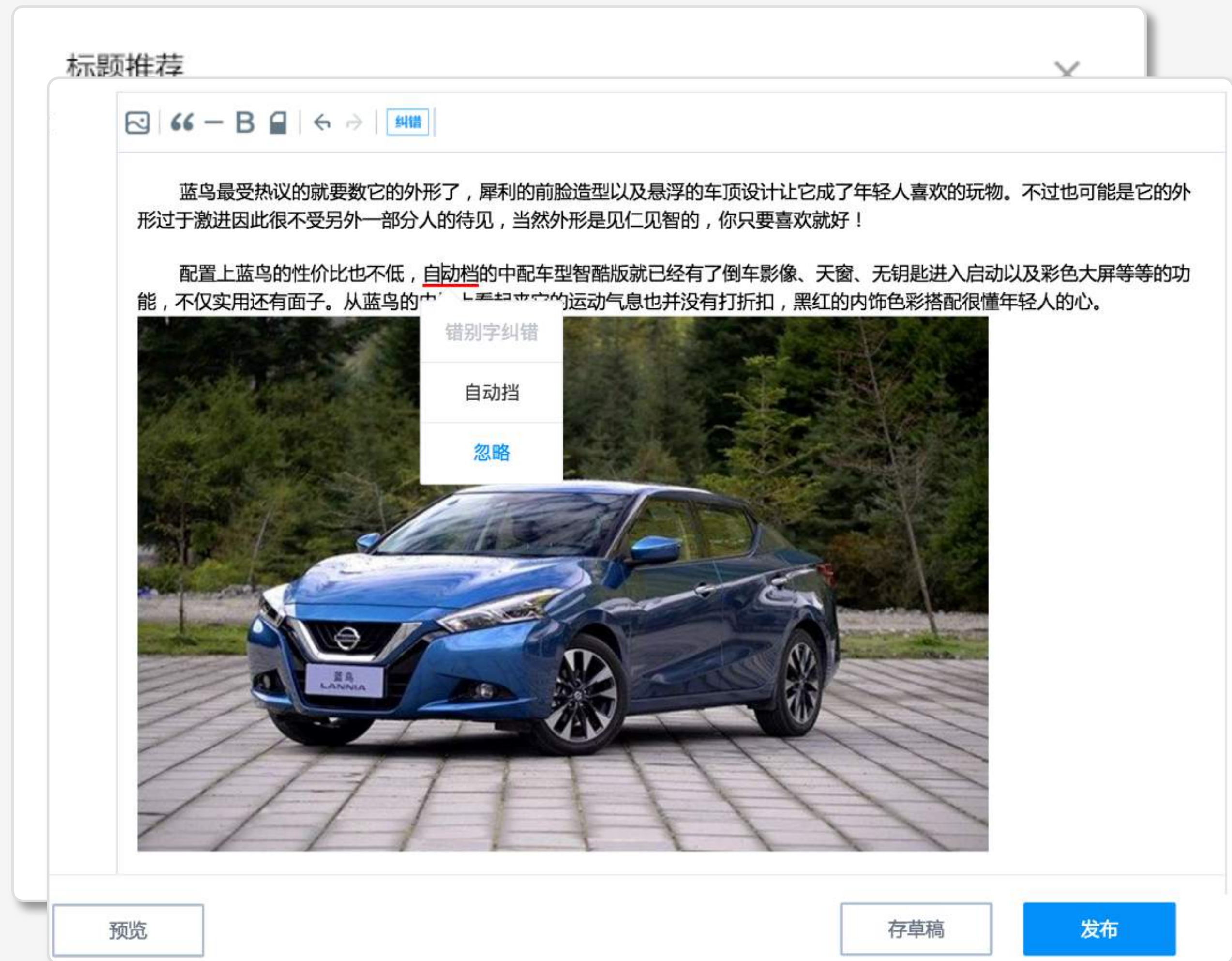
选题推荐 · 内容选材 · 图片处理 · 文字处理 · 视频识别 · 智能写作

标题推荐

- 基于语义理解技术，识别文章内容，推荐优质标题

语义纠错

- 运用机器学习的能力，让机器理解文章中的句子含义，来找出不符合语义的错别字
- 纠错准确率99%



内容生产流程 5

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

视频识别

- 视频查重：基于图像及语音的比对技术，识别内容近似的短视频，应用于审核前原创作者保护，线上降低重复视频推荐；准确率达到97.5%



重复



- 长-短视频关联：将影视截取的短视频匹配到其来源长视频，利用知识图谱获取关联信息、进行相关推荐；准确率90%



--短视频片段获取长视频结构化信息--
影视剧名称：三生三世十里桃花
演员：杨幂，赵又廷，张智尧、迪丽热巴
年份：2017
类型：古装剧

内容生产流程 6

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

百度智能写作示例

智能写作

- 大数据+知识图谱，自动挖掘现实生活中动态热度变化，再通过算法自动生成文章

AP

2014年美联社就启用机器人进行财经报道，每季度能完成对4000家公司的财报报道，此前靠记者仅能完成400家

Y!
YAHOO!

2015年起，机器人用于体育新闻报道，完成梦幻橄榄球赛报道，还在行文中大展幽默（技术支持为自动化洞察力公司，其开发的软件Wordsmith在16年共写了15亿篇文章，涉及金融、房地产、体育等多领域）

腾讯财经

2015年9月，推出自动化新闻写作机器人Dreamwriter，发布了关于消费价格指数的报道

新华社
XINHUA NEWS

2015年11月启用机器人写稿系统“快笔小新”，供职于体育部、经济信息部和中国证券报，撰写体育和财经稿件

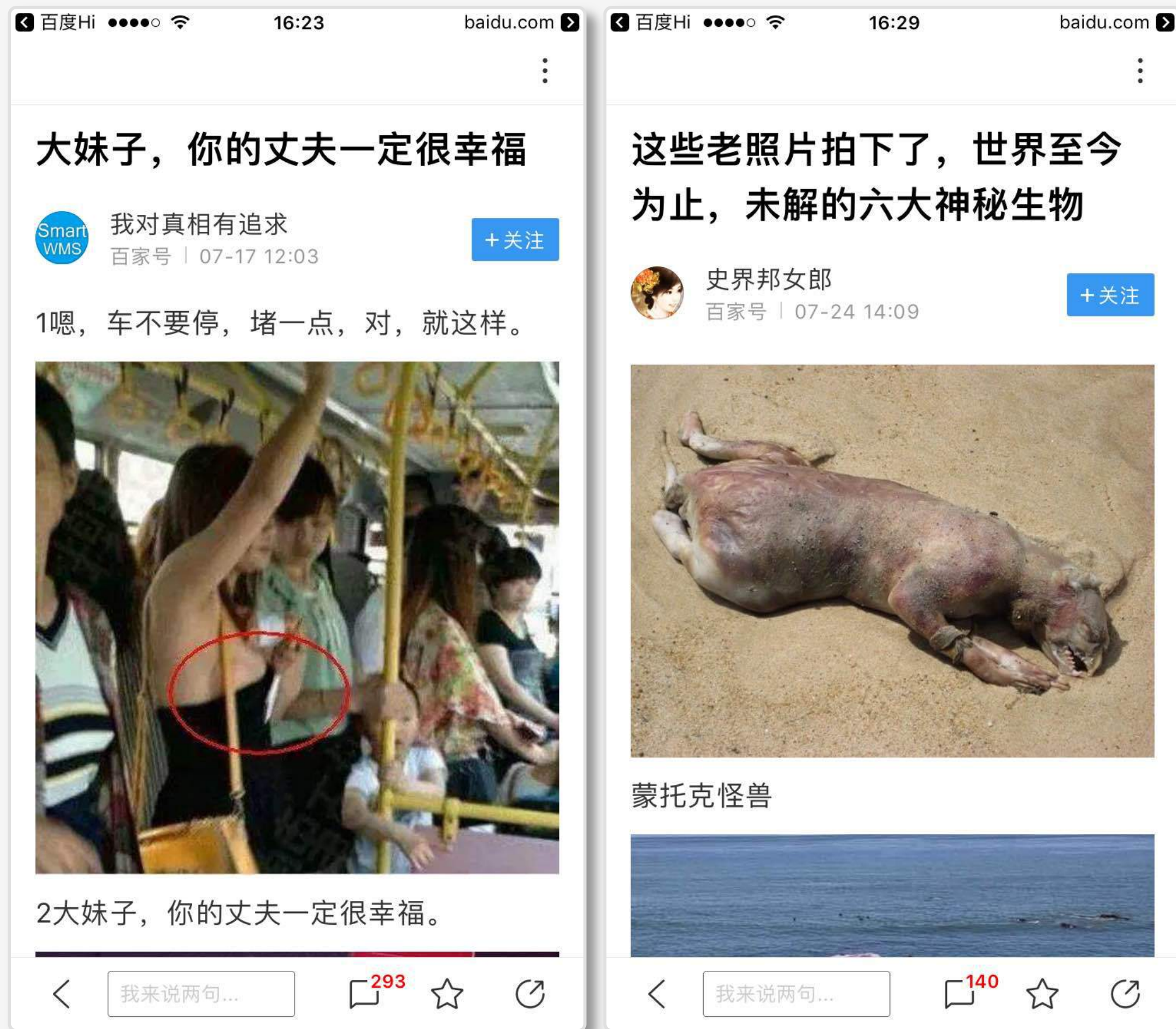


- 内容质量控制模型

内容质量控制模型

构建全方位的篇章质量理解框架，**多维度**理解内容质量

低质内容示例



篇章质量理解框架

低质内容识别

- 内容低俗
- 内容令人不适
- 标题党
- 旧闻新发
- 恶意广告
- 基础文法硬伤

内容优质度计算

- 可读性
- 时效性
- 原创度
- 源权威度
- 内容影响力
- 题材吸引力

基础数据

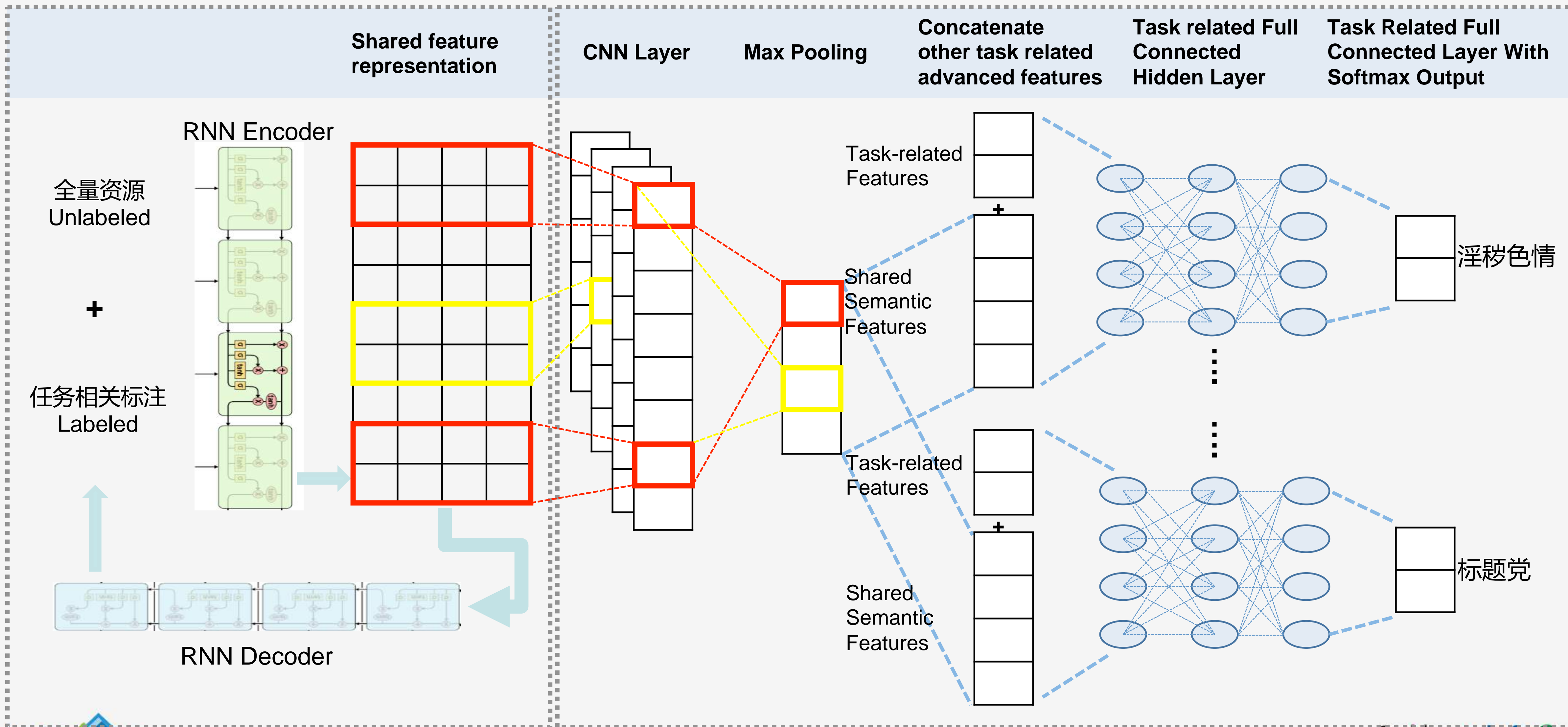
- 作者行为
- 全网用户行为
- 作者信息
- Feed用户行为

基础技术

- 网页搜索
- 机器学习
- 自然语言理解
- 数据挖掘

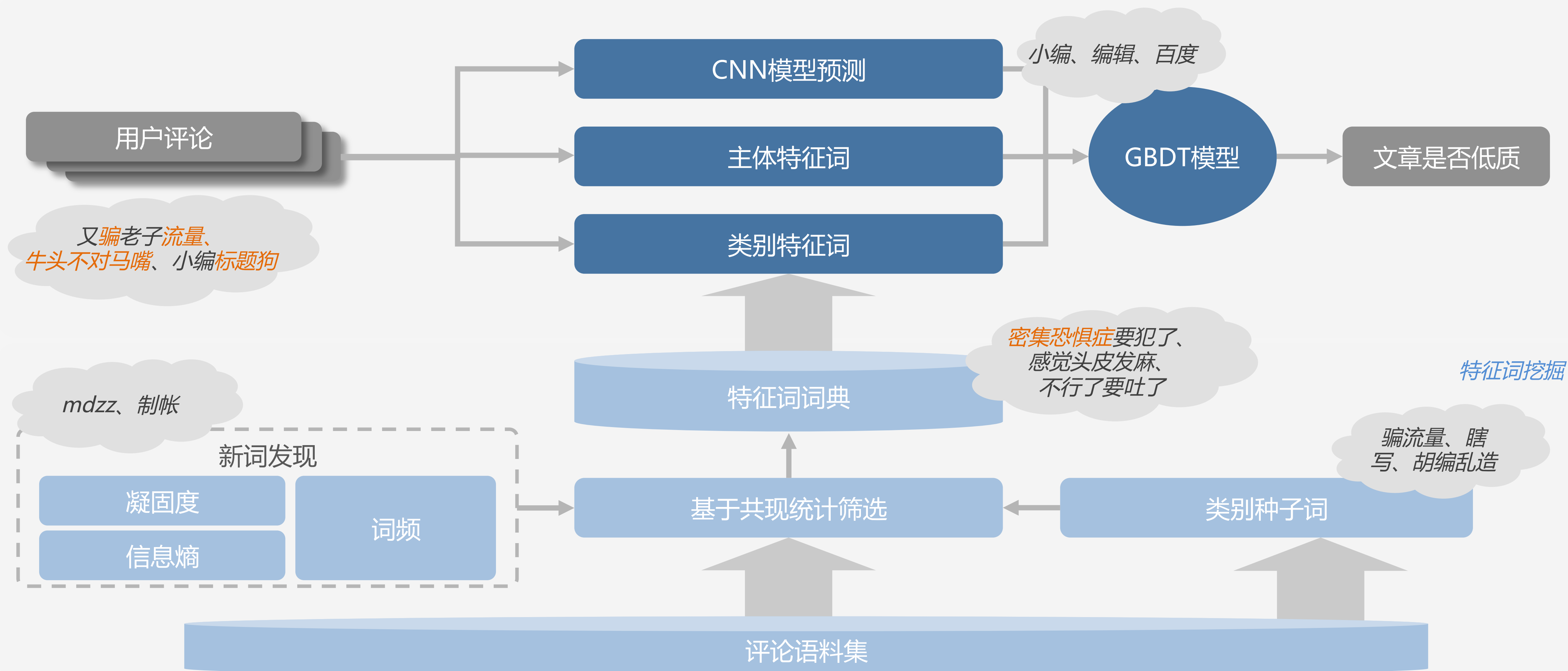
内容质量控制模型

先验质量识别：基于半监督的多目标统一内容分类框架



内容质量控制模型

后验质量识别：基于NLP语义理解的评论正负反馈分析

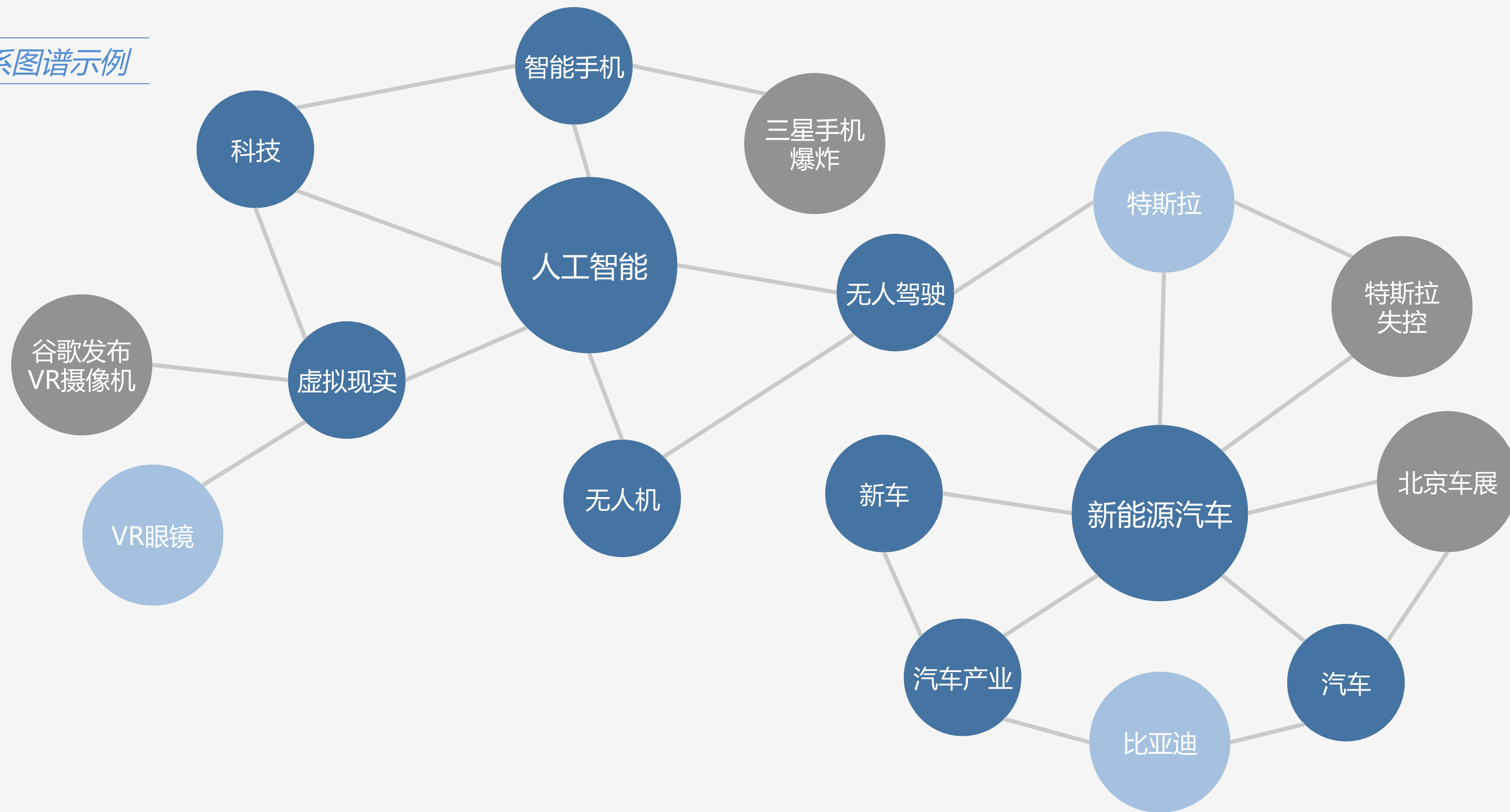


- 内容分发/推荐系统

内容分发/推荐系统

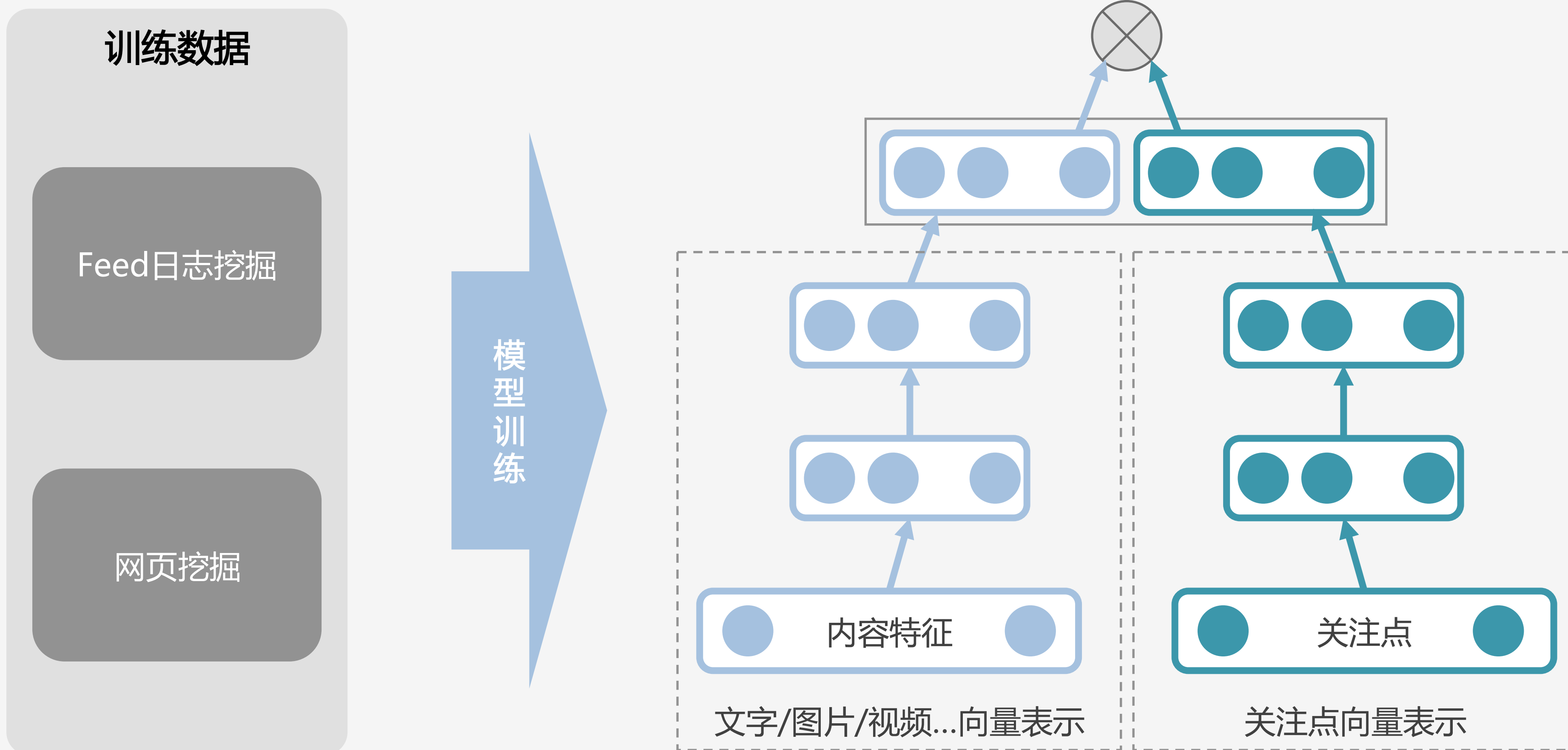
基于知识图谱，构建关注点关系图谱

关注关系图谱示例



内容分发/推荐系统

内容理解：关注点抽取



内容分发/推荐系统

用户理解：用户兴趣挖掘

- 数据来源

- 行为数据：搜索query、Feed阅读、贴吧、全网浏览数
- 画像数据：百度用户画像数据

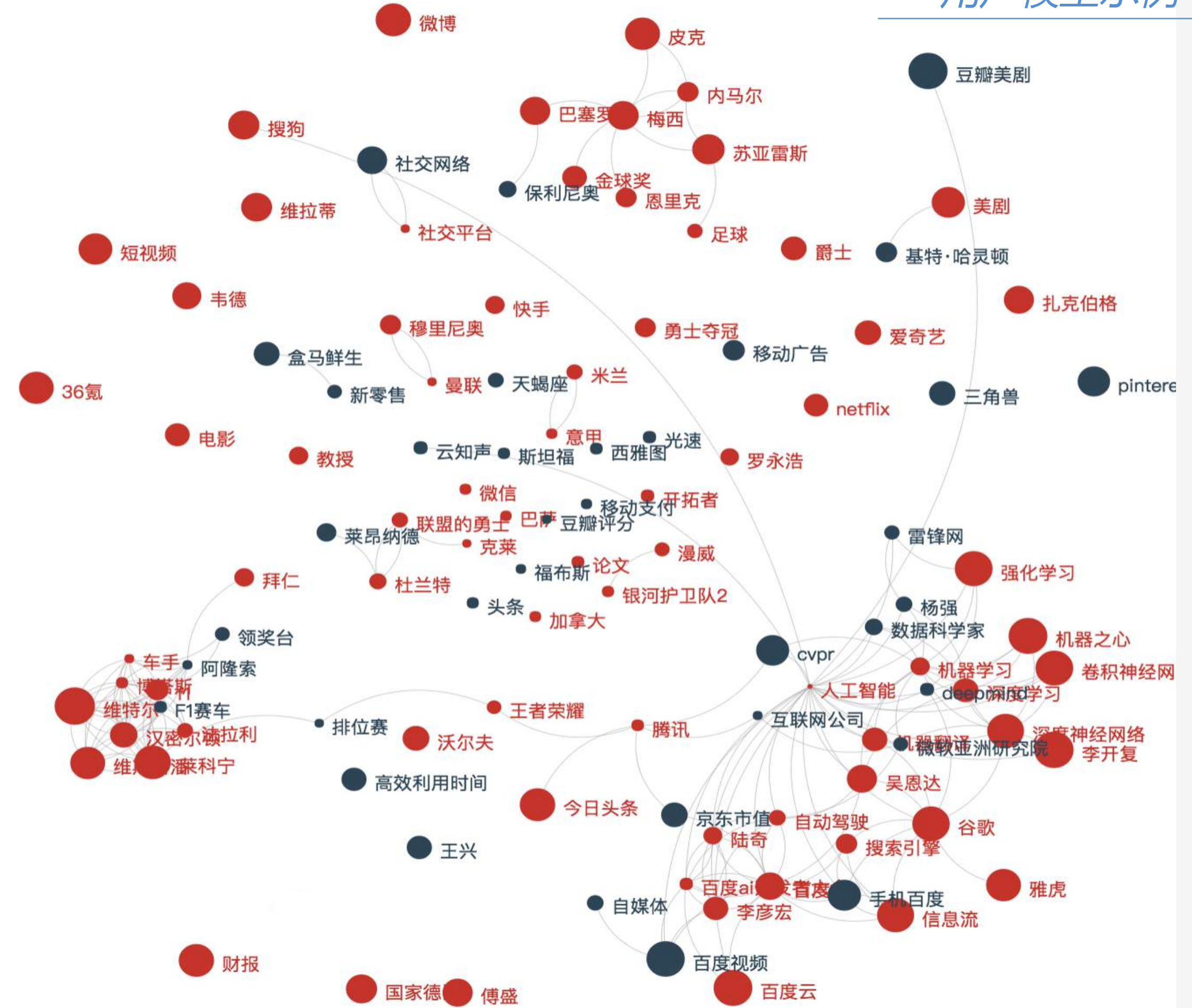
- 基于行为数据的挖掘策略

- 针对搜索Query和Feed阅读内容，基于内容理解技术提取兴趣主题和关注点
- 根据关注点/主题在内容中的权值、展现数、点击数计算相应的强度
- 根据点击数和持续周期，区分长期和短期兴趣点

- 基于画像数据的挖掘策略

- 画像兴趣体系到Feed兴趣体系映射
- 基于画像属性的人群划分和推荐触发

用户模型示例

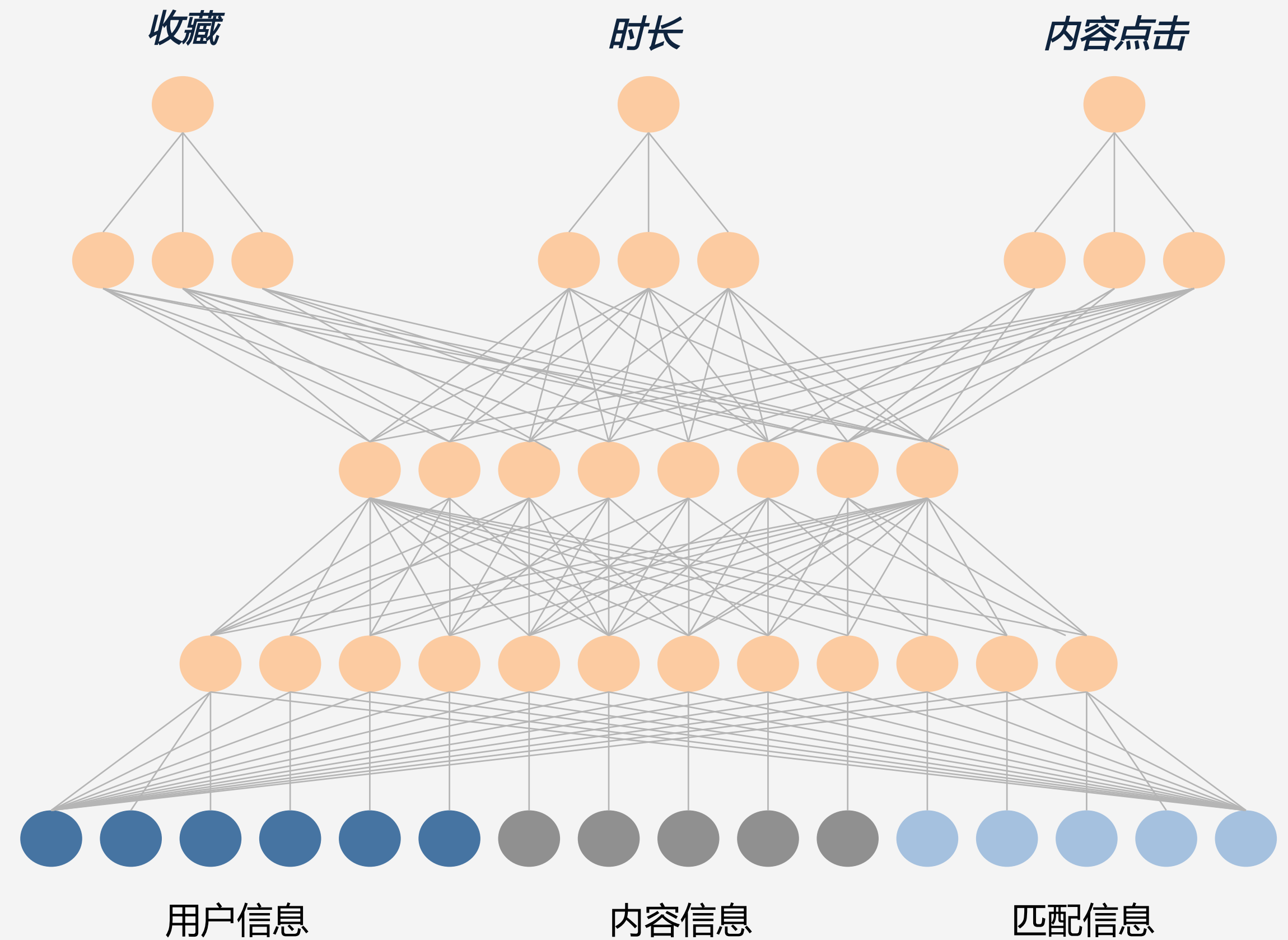


内容分发/推荐系统

推荐策略：多维度的推荐队列召回、排序和融合

推荐价值相关的多目标学习系统 (Multi-Task Learning

- 模型输入：千亿规模参数
 - 用户信息：兴趣、属性、偏好，等
 - 内容信息：吸引力、质量、相关性，等
 - 匹配信息：用户-内容
- 模型输出：多目标
 - 内容点击、时长、收藏、评论、分享，等



内容分发/推荐系统

DNN上线，核心兴趣点强相关内容推荐准确度显著提升

DNN上线前



DNN上线后



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

