

# 什么决定用户满意度？

王树森

# 什么决定用户满意度？

- 用户主观认为某歌比某度好……为什么大家有这种感受？
- 业界共识：相关性、内容质量、时效性是影响用户满意度的三大因素。
- 移动互联网时代，个性化、地域性也会影响用户满意度。
- 通用搜索引擎迭代优化的目标是让用户更满意。
  - 如何让用户更满意？提升相关性、内容质量、时效性、个性化。
  - 如何评价用户体验？留存、有点比等客观指标；人工评估的主观指标。

# 相关性 (Relevance)

# 相关性的定义

- 相关性是查询词  $q$  与文档  $d$  两者的关系。
- 相关性是客观标准，不取决于用户  $u$ 。（如果大多数有背景知识的人认为  $(q, d)$  相关，则判定为相关。）
- 相关性是语义上的，不是字面上的。（相关是指  $d$  能满足  $q$  的需求或回答  $q$  提出的问题。）
- 查询词  $q$  可能有多重意图。只要  $d$  命中  $q$  的一种主要意图，则  $(q, d)$  算相关。

# 相关性模型

- 召回、粗排、精排均需要计算相关性。
- 召回完成之后，候选文档量级为数万；召回海选阶段用文本匹配分数或双塔 BERT 模型粗略地估计相关性。
- 粗排阶段候选文档量级为数千，用双塔 BERT 模型或浅层交叉 BERT 模型计算相关性。
- 精排阶段候选文档量级为数百，用交叉 BERT 模型计算相关性。

# 内容质量

# EAT

- EAT 是谷歌提出的内容质量评价标准 [1]。
- 专业性 (Expertise)：作者有专业资质，比如医生、记者等。
- 权威性 (Authoritativeness)：作者、网站在领域内有影响力，不会被用户质疑。
- 可信赖 (Trustworthiness)：作者、网站的名声好坏。

参考文献：

1. Google's Search Quality Evaluator Guidelines.

# EAT

- 对于 **your money or your life** 方面的查询词，EAT 是排序的重要因子。
- Your money :
  - 金融理财（保险、投资、报税、贷款、转账）。
  - 电商、购物。
- Your life :
  - 医疗健康（诊断建议、用药建议、医院介绍、减肥）。
  - 法律等严肃的主题（诉讼、移民、选举、离婚、收养）。
  - 对人生有重大影响的主题（高考、择校、出国、就业）。



# 文本质量

- 文字的质量：文章写得好不好？
  - 文章的价值：文章是否清晰、全面，事实是否准确，信息是否有用。
  - 作者的态度和水平：写作是否认真、写作的专业程度、写作的技巧。
  - 宠物饲养：解释清楚、事实正确，对读者有价值。
  - 笑话：好笑。
  - 影评：清晰、全面、公正、深刻。

# 文本质量

- 文字的质量：文章写得好不好？
  - 文章的价值：文章是否清晰、全面，事实是否准确，信息是否有用。
  - 作者的态度和水平：写作是否认真、写作的专业程度、写作的技巧。
- 文章的意图：有益、有害？
  - 有益：分享有用的知识、攻略、亲身经历。
  - 有害：虚假信息、软广、散布仇恨、男女对立、发泄情绪。
- 标题党、图文不一致、虚假引流标签、堆砌关键词……

# 文本质量

- 文本质量不是一个分数，而是很多个分数，在搜索排序中起作用。
- 对于每个文本质量分数，都有一个专门训练的模型。
  - 模型：BERT 等 NLP 模型、CLIP 等多模态模型。
  - 数据：制定分档规则，然后人工标注。
- 在文档发布、或被检索时，用模型打分，分数存入文档画像。（搜索排序时直接读取文档画像。）

# 内容质量小结

- EAT：专业性、权威性、可信赖，针对作者和网站。  
(对金钱、健康类的查询词尤为重要。)
- 文本质量：文字质量、文章意图、标题党这样的负面信号，针对文档本身。
- 图片质量（或视频质量）：分辨率、有无水印、是不是截图、图片是否清晰、图片的美学……
- 内容质量不是一个分数，而是很多分数，会在搜索排序中用到。

# 时效性

# 时效性

文档的年龄在排序中起多大作用？

- query = “最新房贷政策”、“美元汇率”。
- query = “泰国旅游”、“新荣记探店”。
- query = “搞笑萌宠”、“宝宝湿疹”。

# 时效性

- 优化搜索时效性的关键是识别查询词的时效性意图（即查询词对“新”的需求）。
- 分类：突发时效性、一般时效性（强/中/弱/无）、周期时效性。
- 识别方法：数据挖掘、语义模型。

# 突发时效性

- 查询词涉及突发的新闻、热点事件。
- 如果查询词带有突发时效性，那么用户想看最近发布的文档。
  - “阿根廷足球”有没有突发时效性？
  - “王力宏”有没有突发时效性？



# 突发时效性

- 查询词涉及突发的新闻、热点事件。
- 如果查询词带有突发时效性，那么用户想看最近发布的文档。
- 识别方法：以数据挖掘为主。
  - 挖掘站内搜索量激增的查询词。
  - 挖掘站内发布量激增的关键词。
  - 爬取其他网站的热词。

# 突发时效性

- 查询词涉及突发的新闻、热点事件。
- 如果查询词带有突发时效性，那么用户想看最近发布的文档。
- 识别方法：以数据挖掘为主。
- 为什么不能用 BERT 等自然语言模型？
  - 人擅长什么，深度学习就擅长什么。
  - 人擅长判断文字质量、相关性，BERT 也擅长。
  - 如果不借助新闻媒体，人无法判断突发时效性。

# 一般时效性

- 只看查询词字面就可以判断时效性意图的强弱。（无需知道近期是否有大新闻。）
- 按需求强度分为 4 档：强、中、弱、无。
- 例：某品牌薅羊毛、黄金价格、上海楼市新闻、在美国更新中国护照、单反相机测评、苏联笑话、82版射雕。
- 识别方法：BERT 等语义模型。

# 周期时效性

- 周期时效性：在每年特定时间表现为突发时效性，在其他时间表现为无时效性。
- 例：双十一、春晚小品、高考作文、奥斯卡。
- 可以不做任何处理。（当查询词表现出突发时效性时，会被算法挖掘到。）
- 可以通过人工标注、数据挖掘识别周期时效性查询词。

# 个性化

# 个性化

- 考虑到不同用户有不同偏好，搜索引擎可以根据用户特征做排序（类似推荐系统）。
- 用预估点击率、交互率来衡量用户对文档的偏好。
- 结合相关性、内容质量、时效性、个性化（预估点击率和交互率）等因子对候选文档排序。

# 为什么需要个性化和点击率模型？

- 查询词越宽泛（例如“头像”），就越需要个性化排序。
  - 宽泛的查询词（例如“头像”）相关的文档数量巨大，其中小部分是用户感兴趣的。
  - 精准的查询词（例如“权力的游戏龙妈头像”）不需要个性化。
- 预估点击率和交互率有利于提升相关性和内容质量。
  - 相关且高质量的文档更容易被点击、点赞、收藏、转发。
  - 与 BERT 等语义模型互补，解决 bad case。
  - 即便是非个性化排序，也会用模型预估点击率和交互率，有助于提升排序效果。

# 总结

- 相关性：查询词和文档在语义上有多相关。
- 内容质量：分为 EAT、文本质量、图片质量等几类，每一类都有多种分数。
- 时效性：意图分为突发时效性、一般时效性、周期时效性。
- 个性化：反映在预估点击率、交互率上。



# 思考题

- PageRank：根据网页之间的链接关系给网页打分。如果很多重要网站的页面指向网页  $x$ ，则认为网页  $x$  也重要。
- 问题：PageRank 属于搜索排序用到的哪一类因子？
  - A. 相关性
  - B. 权威性 (EAT)
  - C. 文本质量
  - D. 时效性
  - E. 个性化

**Thank You!**

<http://wangshusen.github.io/>