

个性化搜索词推荐

赵海臣

搜索词相似度矩阵-训练数据清洗

🌀清洗出最近365天的用户搜索词记录

- ★ 按照userId - search_word - data_date格式将用户的搜索记录进行清洗；
- ★ 按照userId - data_date为session来区别search_word，使用布尔余弦相似度来计算距离，因此session是关键区分因素；

搜索词相似度矩阵训练

🌀使用cf算法的item相似度--布尔余弦计算公式，按userId-data_date的session粒度进行搜索词相似度计算。

- ★ 由于用户的搜索词中有较大的噪声，因此，我们需要对搜索词进行过滤：
 - 过滤掉出现次数低于10个session的噪声搜索词
- ★ 最近搜索词，每个搜索词保留最近的300个搜索词；

购买关联词关联度矩阵训练

🌀与搜索词相似度矩阵流程一样，只是训练数据清洗使用的是搜索订单-搜索关键词

★ session都一样，只考虑发生了订单的搜索关键词。

搜索词推荐

✎ 使用用户最近一个月的搜索词进行推荐：

★ 使用搜索词相似度矩阵进行cf推荐。

THE END

THANK YOU!