

A wealth of data 数据财富

2020年3月5日 22:25

目标

确定可能的设计功能 → 评论筛选关键词

态度?

好: 宣传点

坏: 改进点

时间趋势 → 关键词 freq 趋势 → 加分

问题要求

a) 评价与星级关联:

评价: score

星级: star

加权: 50%? → point

同一商品根据 point 加权得到 mark

权重增加: 高于平均长度, 绿标, 高认可度
(但数值也要高)

权重降低: 太短, 未购买, 认可度低,

权重惩罚: (ban 用户 ID) 星级与评价相反,
关键词众多但与众不同

b) 数据可靠性: 可能要加大低权重惩罚? 可能对过于严重的要 ban

两方面: 恶意/善意比和平均 mark?

拐点分析: 是否应加入低权重(占比)的曲线?

数量

c) 找前者拐点特征

拐点处: 绿标用户?

低权重用户?

滚雪球?

d) 滚雪球

e) 情感分析(可提前写)