**1. 现行通用的算法的问题，尤其是针对我们的需求的弱点**

a.评论的一句话很长，对不同方面的正负面评价都在一句话里，普通的断句模型没有办法断开并且分别给两个维度做正面和负面的判断

b.产品属性不全、或者不灵活

c.情感分析不准，一方面是因为特殊词汇例如“不硌脚”机器不认识，另一方因为a中的断句问题，导致具体某个属性的情感不准

举例：不臭脚、不硌脚、有一双感觉不是很合适等

**2.你用了怎样的方案改进或突破；**

a.开发断句模型，不再仅靠标点符号断句，可以把一句话中对不同维度的评价断开

b.基于关键词的情感分析及属性提取，话题归类

可以直接给一句话打上“舒适-正面”，‘耐磨-正面’的标签

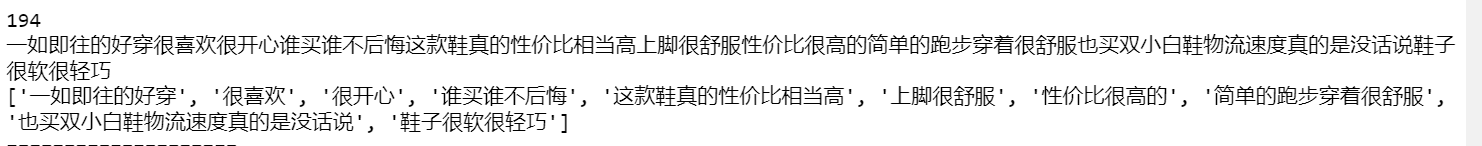
有适用场景，为何人购买，是否首购复购再也不会回购，为了偶像购买，为了品牌购买等精神认同标签，

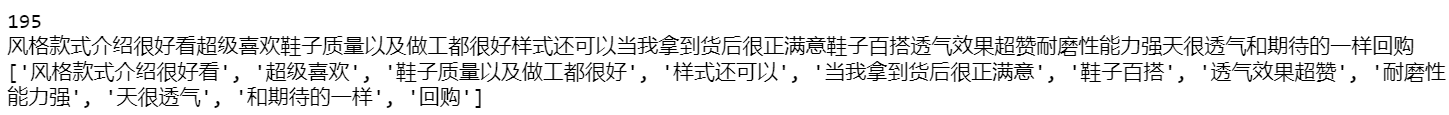
c.建立了一套产品评论的关键词库（辅助b）

d.评论聚类发现新话题，可以灵活的发现新的主题或者热点，比如说反光，比如为某些代言人购买

**3. 效果改进的结果**

a.断句模型能解决一句话不断的问题，把维度分开，不过还是会有断错的地方





b.情感分析能对不同维度做更加正确的情感判断



c.聚类效果也还不错，这个好像更没法对比衡量了