按star rating进行每一星级下review body中形容词的词频统计

在每个商品内部，给定相同benchmark，筛选出符合条件的形容词并排序，再在各星级之间做交叉比对，保留了相邻星级之间最多在三个中同时存在的形容词，汇总如下（具体的词频统计见附录的三张表）

Popular single adjective word people used for different score

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Rating | Most popular adjective | | | | | |  |
| 1 | Bad | Disappointed | Hard |  |  |  |  |
| 2 | Fine | Bad | Hard | Expensive |  |  |  |
| 3 | Fine | Easy | Hard | Cheap | Wish |  |  |
| 4 | Fine | Easy | Overall | Nice | Wish | Happy | Perfect |
| 5 | Easy | Happy | Nice | Perfect |  |  |  |

可以看到

评分为1、2的评论中多出现bad

评分为3、4的评论中多出现wish

评分为4、5的评论中多出现happy、nice、perfect

评分为1、2、3的评论中多出现hard

评分为3、4、5的评论中多出现easy

评分为2、3、4的评论中多出现fine

评分为1的评论中常常出现disappointed

评分为4的评论中常常出现overall

评分为2的评论中常常出现expensive（仅限hairdryer）

评分为3的评论中常常出现cheap（仅限microwave）

使用N-gram模型，得到2-gram and 3-gram together的频率统计

同样的方法，汇总如下（具体的词频统计见附录的三张表）

Popular single frequent phrases people used for different score

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rating | Most popular phrases | | | | |
| 1 | Stay away | Not buy | Waste money | Stop work |  |
| 2 | Not recommend | Waste money | Not think | Stop work |  |
| 3 | Not know | Not want | Stop work | Look good | Look nice |
| 4 | Really like | Work great |  |  |  |
| 5 | Highly recommend | Great price | Work great | Look great | Love |