

亚马逊在线平台评论分析

摘要：在亚马逊的在线平台中，亚马逊为客户提供了对各类产品的购买进行评分和评价的机会，客户可以根据这些数据来协助自己的购买决策。同样公司也可以根据这些数据，来改变自身产品销售策略或是改善自身产品的功能。

本研究综合了star_rating、review_body、review_headline、vine和purchase数据，运用TF-IDF模型、朴素贝叶斯算法、逻辑回归分类、层次分析法、相关性分析和多项式拟合完成了对评论的定量并得到了一个模型，通过该模型本研究能得到一条评论对产品更客观的评分。识别了评分与时间的关系模式，判断在某一时间段产品的声誉是否正在上升或下降。



关键词： 1.词频统计2. TF-IDF模型 3. 层次分析法 4. 相关性分析5. 多项式拟合

对数据的量化与挖掘

review_headline			review_body			review_headline			review_body		
count	11468		11470	count	1615		1615	count	18939		18937
unique	7696		11197	unique	1346		1601	unique	12611		18085
top	Five Stars		good	top	Five Stars		None available.	top	Five Stars		good
freq	1284		20	freq	149		4	freq	1875		72

1.1
Hair dryer

1.2
microwave

1.3
pacifier

-Python对review_headline与review_body进行词频分析，得到3个频率最高的词语

star_rating	sum	Y	N
1	1032	739	293
2	639	509	130
3	999	859	140
4	2096	1804	292
5	6704	5900	804

2.1
Hair dryer

star_rating	sum	Y	N
1	402	116	285
2	112	56	56
3	134	92	42
4	300	246	51
5	667	579	85

2.2
microwave

star_rating	sum	Y	N
1	1192	468	117
2	945	440	65
3	1426	678	98
4	2716	1255	191
5	12660	6950	936

2.3
pacifier

-探究已购买过产品的客户和未购买过产品的客户对产品的评分情况



三种不同类型客户的权重及满意度与参考价值的权重

product_review	vine but no purchase	no vine but purchase	no vine or purchase	result_weight
vine but no purchase	1.0000	0.3333	3.0000	0.2426
no vine but purchase	3.0000	1.0000	7.0000	0.6694
no vine or purchase	0.3333	0.1429	1.0000	0.0880

score	satisfaction	Reference value	result_weight
satisfaction	1.00	3.00	0.75
Reference value	0.33	1.00	0.25

层次分析法建立模型

score	product_review	stars_ratings	result_weight
product_review	1.00		
stars_ratings		1.00	

score	product_review	stars_ratings	result_weight
product_review	1.00	3.00	0.75
stars_ratings	0.33	1.00	0.25

分析：客户可以分为4种类型：

- 1.既是评论员又购买了产品
 - 2.是评论员但没有购买产品
 - 3.不是评论员但购买了产品
 - 4.既不是评论员又没有购买产品
- 我们认为这四种不同客户所给出的评分与评价的参考价值不同，所以我们依然使用层次分析法去确定四种客户类型的权重。

我们建立的客观评分模型为：

$$Y = x * (A * a + B * b) + 5 * z * c$$

(满意度权重：x 参考价值权重：

z 评分：A 评分权重：a 评论：B 评论权重：b

评论者类型权重：c)

实验讨论：从以上相关性分析与挖掘可看出，在总体上来看一星评分出现前后的评分态度并没有什么太大相关性，可是在hairdryer和pacifier这两款产品早期一星评分对后续评分的影响是负相关的。这两款产品早期一星评价对后续评分可能出现跟风性评分，而在后期可能由于客户对产品更加了解或者说客户对于评分更加的理智，所以对后续的评分影响并不大，冲microwave的相关性变化可看出这款产品的客户更加理性并没有被一星客户所影响，但是有一段时间一星评分等于后续客户的态度是有少许负面影响的，总结来说，但产品刚上市客户对产品不了解恶意一星评分可能会对产品后续的评分以及销量产生较大的负面影响，而在产品经过市场的检验以后，人们对产品的评价也更加理性。