

早上好，阳光： 基于评分和评论的分析

总结

点评系统在电子商务的演变过程中扮演着巨大的角色。客户使用它来帮助做出最佳的购买决策，而供应商可以利用它来收集关于如何进行改进的意见。因此，从供应商的角度来看，学会最有效地从堆积如山的客户声音中提取最有价值的信息是至关重要的。

首先，基于我们在提供的数据集中发现有用的属性，我们提出了基于评分和评论的分数(RRBS)模型，它定义了产品的分数——基于从评分和评论中提炼出来的信息的数据度量——精细地量化和描述了客户对产品的反应。在 RRBS 模型的基础上，我们进一步引入了声誉模型，它给出了一个基于时间的产品声誉随时间周期的量化。我们的模型评估结果表明，量化后的声誉与相应产品的销量在趋势上有很强的相关性。这一事实直观地证明了我们模型的正确性。

然后，为了预测一个产品在未来的潜在成功或失败，我们提出了成功预测模型，利用我们之前提出的模型所建立的知识基础。基于一个产品的成功与其销量密切相关的前提，我们能够通过计算出的声誉序列来预测产品未来的声誉，然后将预测值与预先设定的阈值进行比较，从而解决成功预测任务。因此，可以有效地预测产品的成功程度。

接下来，我们通过记录一系列相同评级(高或低)之后的客户评论来评估星级评级对客户具体评论的激励，并计算它们是正面、负面或中性的机会。事实证明，一系列的高评分确实会激发更多的正面评论;另一方面，我们无法自信地确定一系列低评分对差评的煽动。随后，我们通过将评论内容与分级词进行匹配，从评论文本中发现评分水平与特定质量描述符之间的关联。结果显示，积极词和高评分之间存在强烈的联系，同样，消极词和低评分之间也存在强烈的联系。

最后，为了协助阳光公司即将开展的业务，我们提出了我们设计的营销策略和每个产品的一些重要设计特征。关键词:RRBS 模式;信誉模型;相关性分析;



关注数学模型

获取更多资讯

内容

1 介绍	2
1.1 问题重述	2
1.2 文献综述	2
1.3 数据清理	3
1.4 建模框架	.
2 个假设和命名法	3
3 RRBS 模式	4
3.1 模型概述	5
3.2 评价向量	5
3.3 审查向量	6
3.3.1 基于评论的测度方程	8
3.3.2 情绪极性矩阵	9
4 信誉模型	10
4.1 时间序列	10
4.2 量化的声誉	11
4.3 模型的评估	11
4.3.1 量化美誉度与销量的趋势相似性	11
4.3.2 肯德尔 τ 的方法	12
5 成功预测模型	
5.1 高斯过程回归	14
5.2 声誉映射	15
5.3 成功阈值	15
6 结果	16
6.1 评价星评的煽动性	16
6.2 评论措辞与评分等级的关联	18
6.3 我们的策略	19
6.3.1 营销策略	19
再重要的设计特征	19
7 敏感性分析	20
8 优点和缺点	
8.1 优势	20
8.2 缺点	21
9 结论	21
10 我们的信	22



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

1 介绍

“当你倾听消费者的声音时，神奇的事情就会发生。”

——爱彼迎首席营销官乔纳森·米德霍尔

电子商务在过去几十年里在世界各地蓬勃发展，现在已经成为最繁荣、最有前途的现代产业之一。在这一过程中，点评系统一直扮演着巨大的角色。它不仅可以帮助顾客在做购买决定时找出最适合自己的东西，还可以帮助卖家意识到他们的产品存在什么问题，从而可以改进。因此，对于企业来说，知道如何最大限度地利用信息技术，并将这些知识应用于改进产品设计和营销策略是非常重要的。

1.1 问题重述

阳光公司即将在市场上推出三款新产品，包括微波炉、婴儿奶嘴和吹风机。亚马逊网站上出售的微波炉、婴儿奶嘴和吹风机的顾客评论和评级数据被提供，每一个都有时间标签。需要分析顾客对其他公司竞争产品的评价和评论，以找出最佳的在线营销策略和可能的设计特点，他们可以使用，使他们的产品更符合顾客的要求。

为了实现我们的目标，具体来说，我们需要：

- 一旦他们在在线市场推出了他们的三种新产品，我们需要根据评论和评级为阳光公司确定最有信息的数据度量。
- 发现基于时间的衡量标准和模式，可以暗示一个产品的声誉的增长或下降趋势。
- 将基于文本的度量和基于评分的度量结合起来，以发现最能表明产品潜在成功或失败的度量。
- 找出特定的星级评分是否会引发更多的评论。
- 找出基于文本的评论的评级水平和特定质量描述符是否密切相关。

1.2 文献综述

早期的工作提出了几种利用在线产品评论和评分的方法。2009 年，Koren 等人[4]提出隐语义模型(Latent Factor Models, LFM)来解决评分预测任务。2013 年，McAuley 等人[5]利用评论中丰富的信息，利用潜在狄利克雷分配(LDA)实现主题分布。2016 年早些时候，Tan 等人提出了 rating - Latent Topics (RBLT)框架，将文本评论与用户的情感倾向相结合，从而使推荐更加准确。

然而，评论中一些有价值的信息在介绍中并没有得到充分利用作品，比如不同评论的文本长度不同，情感强度不同 的具体关注数学模型



获取更多资讯

字(如。“fine”没有“good”那么积极强烈。而且，现代电商推荐系统中的其他重要属性，如有用性评分、评测者可信度、评测者是否为验证买家等，都很少被考虑在内。对评论和评分组合的量化还没有提出，基于评论和评分的数据度量也没有提供。

1.3 数据清洗

在对提供的数据进行观察后，我们注意到每个数据集中都有一些放错位置的数据，其中一些评论项目属于错误类别的产品。例如，在文件 hair_dryer 中。TSV 具有 product_parent 值 #466871293(唯一标识一个产品)的评论项属于一个发胶产品，而不是吹风机。类似地，在文件微波炉。TSV product_parent 值#311592014 的评论项属于冰箱，而不是微波炉。我们发现这些错误几乎只发生在那些评论记录为个位数的产品上。加上稀疏数据对我们建模贡献不佳的看法，我们省略了所有具有单位数评论记录的产品项目。

1.4 建模框架

我们的建模框架如图 1 所示。

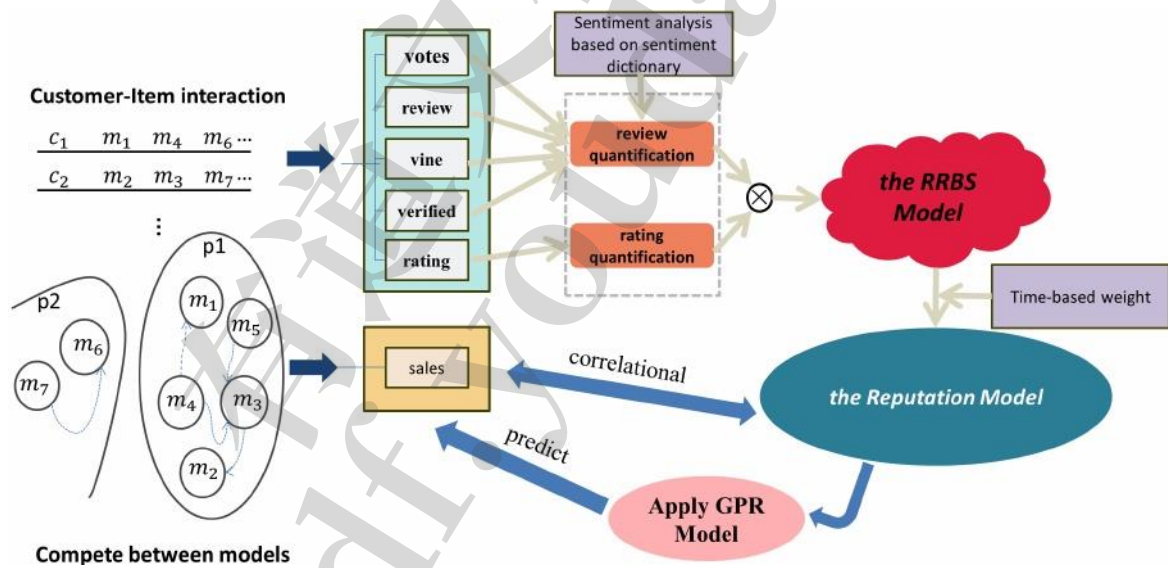


图 1:建模框架

2 个假设和命名法

为了简化我们的模型，我们做以下假设:

关注数学模型
获取更多资讯

- Asm. 1 提供的数据不包括来自竞争对手或恶意客户的垃圾评论。
- Asm. 2 每条评论的内容都是合法且公正的。
- Asm. 3 一个产品的评论记录数等于该产品上线以来累计的销量，即。，在提供的任何数据集中都没有遗漏或排除的记录。
- Asm. 4 客户的评论和评分是在他们购买产品的当天发布的，即。，忽略了包装、发货、用户确认质量等环节的时间消耗。
- Asm. 5 每个客户最多有一种产品类型(微波炉、安抚奶嘴或电吹风)的评论和评级记录。

我们有一个在建模过程中定义的主要符号列表，如表 1 所示。

3 RRBS 模型

在这一节中，我们提出了 RRBS 模型(即基于评分和评论的分数模型)作为数据度量，结合了星级评分和评论的文本信息。在确定量化过程时，我们考虑了所提供数据中几乎所有的影响属性。因此，一旦他们的三款新产品推出，一个信息量非常大的数据度量就会呈现给阳光公司来跟踪。

表 1:我们在未来讨论中使用的符号。

象征	描述
i	产品类型(可以是微波炉、奶嘴或吹风机), $i = 1,2,3$
k	某一类产品的唯一型号标识符
t	时间戳标识符(单位:月)
n	某一型号在特定月内的评级和审核记录数量
$score_i^k(t)$	衡量一个产品模型收到的客户反应
α	评分向量——一个 n 维行向量
β	复习向量——一个 n 维列向量
λ	评级权重参数
一个	情感极性矩阵——一个 $n \times n$ 的对角矩阵
Φ	复习-量化向量——一个 n 维列向量
θ_j	葡萄树的 因素
s_j	情感强度因子
v_j	效度的因素
沪江	Helpfulness-rating 因素
L_j	Review-length 因素
ϵ_j	修正项
分数	模型得分的时间序列
$Repi^k$	量化的声誉



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

3.1 模型概述

作为我们指标的一部分，我们根据公司收到的评级和评论，将分数定义为公司特定产品的衡量标准。它定量地指出了产品在获得客户积极反馈方面做得好还是不好。我们将指标定义为:

$$score_{i,k}^{(t)} = \alpha^{\circ \lambda} \beta$$

(1)

地点:

- $I \in \{1,2,3\}$ 表示 I 的乘积 $_{th}$ 类型;
- k 表示 k_{th} I 的模型(由 “product_parent” 值唯一标识) $_{th}$ -类型产品;
- t 表示 t_{th} 月;
- α 是一个 n 维行向量, 其值由星级评分决定;
- β 是 n 维列向量, 其值由评论和其他影响因素决定;
- λ 是一个影响因子, 从影响评分的角度决定评分的权重价值。我们最初将其值设置为 1。

3.2 α 的定义——评级向量

α 的值与产品的星级评分直接相关, 从 1 到 5 不等。而 1 的值表示 “非常不满意”, 5 的值表示 “非常满意”。我们考虑两个星级评级, 一个值为 1, 另一个值为 5, 具有相同的强度程度(即。 , 一个表示 “非常” 不满意, 另一个表示 “非常” 满意)。同样, 4 星评级被认为与 2 星评级具有相同的强度。而 3 星评级代表的是中立立场, 我们认为这种立场相对不那么激烈。因此, 我们进行了一个映射过程, 如表 2 所示, 以使用统一的度量来量化一个星级评级的正负程度。

表 2:从星级评分值到其对应的映射评分值(即强度程度)的映射。

星级	映射的评级
1 星和 5 星	3.
2 星和 4 星	2
三星 级	1

如上所述, α 是一个 n 维行向量, 只与星级相关。 α 中的 n 个非零元素表示一个产品在一个月内的 n 个映射评级值。我们可以得到:

$$\alpha = \left(\alpha_{i,k,1}^{(t)} \quad \alpha_{i,k,2}^{(t)} \quad \cdots \quad \alpha_{i,k,j}^{(t)} \quad \cdots \quad \alpha_{i,k,n}^{(t)} \right)$$

(2)

其中 $\alpha_{i,k,j} (t) \in \{1,2,3\}$ 。



关注数学模型
获取更多资讯

3.3 β 的定义——评论向量

β 是一个 n 维列向量，其值与评论和其他一些影响因素密切相关。它综合考虑了评论文本的情感强度、不同评论的长短、有用性评分、评论人的权威性(即评论人的影响力)等因素。他/她是否是 Amazon Vine 的评论人¹⁾以及评测者是否是经过验证的买家。因此，可以利用所提供数据中的信息，并有效地将文本评论和其他非数值影响因素整合到我们的数据度量中。

我们将其定义为:

$$\beta = A\Phi = \left(\beta_{i,k,1}^{(t)} \quad \beta_{i,k,2}^{(t)} \quad \cdots \quad \beta_{i,k,j}^{(t)} \quad \cdots \quad \beta_{i,k,n}^{(t)} \right) \tag{3}$$

地点:

- A 是一个 n 乘 n 的对角矩阵，用于描述评论的情感极性;
- Φ 是一个 n 维列向量，用于量化语义信息和评论的其他影响属性。

Φ -the 评论定义——量化向量

为了量化评论的语义信息和其他影响属性，我们将 Φ 定义为 n 维列向量，如式(4)所示:

$$\Phi = \left(\phi_{i,k,1}^{(t)} \quad \phi_{i,k,2}^{(t)} \quad \cdots \quad \phi_{i,k,j}^{(t)} \quad \cdots \quad \phi_{i,k,n}^{(t)} \right)^T \tag{4}$$

在 $\phi_{i,k,j}^{(t)}$ 是一篇文本评论的语义信息，以及其他影响属性的度量，它来自 j_{th} k 的顺序 $_{th}$ I 的模型 $_{th}$ 类型的产品。

我们将这个衡量标准定义为:

$$\phi_{i,k,j}^{(t)} = \theta_j \cdot s_j^{v_j} \cdot h_j \cdot L_j \tag{5}$$

θ 的定义_j-藤因子

据亚马逊[1]网站称，亚马逊选出的最值得信任的评论人都被邀请参加亚马逊的 Vine 项目，在这个项目中，他们可以发布关于新产品和预发布产品的意见，以帮助其他客户做出明智的购买决定。

虽然 Vine 的评论者对所提供的产品有独立的意见，但供应商不能影响、修改或编辑这些评论。因此，我们有理由相信，Vine 评测员发布的评论比常规评论更专业、更值得信赖。相应地，当评论是 Vine 评论时，我们通过定义 θ ，使我们衡量的价值内容更多_j 为:

$$\theta_j = \begin{cases} 2, & \text{如果这篇评论是由 Vine 评论人写的} \\ 1, & \end{cases} \tag{6}$$

1.亚马逊 Vine 评测员是亚马逊 Vine 计划的会员，受邀领取产品
免费换取公正的评论。



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

s的定义_j-情感强度因子

大多数时候，当人们对他们购买的产品发表评论时，往往会包含一些表达明确情感和态度的具体词汇。还值得注意的是，评论人所选择的词语的情感强度有所不同。例如，“好”比“好”表达的正面情绪强度更高。也就是说，我们能够量化不同单词的不同情感强度，因此，在衡量语义信息时，一篇评论试图表达的信息是可分级的。

带着这种心态，我们试图找到一个现有的数据集，其中包含相当数量的常见形容词，(最重要的是)每个形容词都有一个等级，表明其情感强度。幸运的是——感谢伟大的开源社区——我们能够获得一个名为 Afinn[6]的 Python 库，它由 2,477 个分级的单词组成，满足了我们的确切要求。然后我们使用这个词集来匹配提供的数据中所有评论的内容，过滤掉没有出现在任何评论中(我们不需要)的单词。

结果，我们得到了一个分级词表(表 3 中显示了部分内容)，它将正负词汇分类为 3 组不同的等级(3 为最高，1 为最低)。

表 3:通过将 Afinn 词集与所有评论的内容进行匹配而获得的分级词表的一部分。一些出现频率最高的评分词被列在这个表中。注意:我们根据自己对单词情感强度的了解修改了其中一些单词的等级。

年级	积极的	负
3.	杰出的，激动的，了不起的，了不起的，难以置信的，辉煌的，了不起的	灾难性的，可怕的，死亡的，可笑的，可怕的，糟糕的，讨厌的，绝望的
2	善良、可爱、强大、舒适、干净、有用、满意、乐于助人	坏、失望、错、无用、消极、贫穷、不快乐、烦人
1	公平的，精细的，容易的，凉爽的，清晰的，安全的，可接受的，有能力的	陌生、辛苦、困难、嘈杂、有限、愚蠢、不确定、虚假

因此，我们可以定义 s_j 为 j 的文本评论内容的情感强度 $_{th}$ 的次序(的 k_{th} I 的模型 $_{th}$ 类型的产品)。其值定义为:

$$s_j = \begin{cases} \eta & \text{如果匹配的单词有 } \eta \text{ 的最大等级, 否则} \\ 0, & \end{cases}$$

(7)

v的定义_j-效度因子

直觉上，如果评论不是由经过验证的买家发布的，它的有效性就有点可疑。因此，考虑到未经验证的买家评论的潜在不真实性，我们降低了这类评论在我们的度量中的权重。我们定义 v_j 为了表示一个评论的有效性，当它较低时，相应的评论对我们衡量产品的分数的影响较小。

$$v_j = \begin{cases} 1, & \text{如果评论人是经过验证的买家,} \\ 0.1, & \text{则不然} \end{cases}$$

(8)



关注数学模型
获取更多资讯

h 的定义 j -有用性评分因子

毫无疑问，评论的有用性评级作为我们衡量的一个影响因素是不能被忽略的。它不仅说明了一条评论的可信性和整体质量，还暴露了一条评论潜在的垃圾性质(被误导或误解)。这使得决定一篇评论是否应该在我们的衡量中有更多或更少的权重变得更容易。

因此，我们定义 h_j 作为我们测量中的有用性分数，如公式(9)所示。当还没有人投票给它有用性时，它的值应该是 1，使它在我们的测量中具有有用性的原始权重。如果“有帮助”的票数与总票数之比大于 0.5，我们认为这个评论是一个有帮助的评论，因此它的权重应该高于 1；反之，如果该比率低于 0.5，权重应小于 1，则为无益评论。

$$h_j = \begin{cases} e^{\frac{\text{helpful_votes}_j}{\text{total_votes}_j} - 0.5}, & \text{if } \text{total_votes}_j > 0 \\ 1, & \text{否则} \end{cases}$$

(9)

在 helpful_votes_j “有帮助”的票数和 total_votes_j 是评论的总投票数。

L 的定义 j -评论长度因子

Chua et al.[3]提供的见解是，评论的有用性与其深度/长度呈正相关。从经验上讲，我们知道如果一篇评论在某种程度上很长，那么假设它对目标有更多的描述是合理的，准确性也更有可能得到保证。

因此，我们可以定义 L_j 作为仅由评论长度贡献的评论有用性的度量。一篇评论包含的单词越多，我们认为这篇评论就越有价值。然而，值得注意的是， L_j By review length 应该是非线性的，并且增长率应该随着 review length 的增加而降低，否则它会过多影响我们对产品评分的衡量。因此，我们定义 L_j 为:

$$L_j = \frac{1}{2} \log_{10} \text{len}_j$$

(10)

在 len_j 是评论里的 j 字的字面数量吗 θ_k (k 的顺序 θ_k I 的模型 θ_k 类型的产品)。

3.3.1 基于评论的测度方程

将式(6)、(7)、(8)、(9)、(10)代入式(5)得到:

$$\phi_{i,k,j}^{(t)} = \theta_j \cdot s_j^{v_j} \cdot e^{\frac{\text{helpful_votes}_j}{\text{total_votes}_j} - 0.5} \cdot \frac{1}{2} \log_{10} \text{len}_j$$

(11)

这是我们定义的方程，用于衡量文本评论的语义信息和其他影响属性，考虑了评论者的权威、评论的有效性、评论的情感强度、评论的有用性评级和评论的文本长度。

请注意 s_j 的值 j 可以为 0，为:

•评价的措辞可能是相当中立和明确的情绪(eith positive or 否定的)无法推断:



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

•评论中的一些特定单词可能暗示一种情感，具有一定程度的强度，但不包括在 *Afinn* 的分级词集中，使得对此类评论的情感强度进行评估是不可实现的。

根据公式(11)，我们知道一个带有 s 的评论 j 值为 0 也有 ϕ 的值 $\phi_{i,k,j}$ 值为 0，意味着该测度将这样的评论视为“完全垃圾”的评论;这使得该评论与其对应的星级评级相乘的结果为 0，意味着该评级间接失去了其效果。这是不合适的，我们希望在这种情况下保留我们的数据度量的效果。因此，我们通过引入一个修正项来解决这个问题 ϵ_j ;其值定义为:

$$\epsilon_j = \begin{cases} 1, & \text{if } s_j = 0 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

通过将这个修正项添加到我们的测量中(式(11)), 我们有:

$$\phi_{i,k,j}^{(t)} = \theta_j \cdot s_j^{v_j} \cdot e^{\frac{\text{helpful_votes}_j}{\text{total_votes}_j} - 0.5} \cdot \frac{1}{2} \log_{10} \text{len}_j + \epsilon_j \tag{13}$$

现在计算 ϕ 的值 $\phi_{i,k,j}$ 将为 1，如果 s_j 为 0，使得对应的评论有了初始权重;由此，由式(1)可知，该评论及其对应的星评共同决定的分数，在这种情况下仅取决于星评。

通过合理地整合评论中几乎所有有用的信息，我们可以有效地衡量评论的价值，从而可以过滤掉写得很差、没有什么值得关注的评论，并从值得倾听的评论中保留有价值的信息。

3.3.2 情感极性矩阵

A 是一个 $n \times n$ 的对角矩阵，用于描述评论的情感极性。我们给 A 的一个元素赋值+1，当且仅当 (该元素)对应的评论被测量为正。类似地，如果对应的评论被测量为负的，我们给元素赋值 - 1，如式(14)所示:

$$a_{i,k,j}^{(t)} = \begin{cases} +1, & \text{如果对应的评论暗示了积极的情绪} \\ -1, & \text{如果对应的评论暗示了消极的情绪} \end{cases} \tag{14}$$

至于如何判断评论是正面的还是负面的，我们知道，我们在匹配评论时使用的 *Afinn* 词集是由带符号的分级词组成的。例如，在词集中，“完美”一词的评分为+3，“不完美”一词的评分为- 2。因此，在经过一个情感词和评论内容的匹配过程后，我们提取结果的符号并进行检查。如果是积极的信号，我们认为这篇评论是积极的;同样，如果是负号，我们认为这篇评论是负面的。

对于 A ，我们有:

$$A = \text{diag}(a_{i,k,1}^{(t)}, a_{i,k,2}^{(t)}, \dots, a_{i,k,j}^{(t)}, \dots, a_{i,k,n}^{(t)})$$



(15) 关注数
学模型

获取更多资讯

我们还考虑了一些评论中可能包含“not”这样的词，从而颠倒了下一个词的意思。为了评估这类案例的影响，我们进行了统计分析。如果“否定-形容词”组合(例如，“不好”)出现的频率低到可以忽略，那么我们可以相信，如果不解决这个问题，这种可能性对我们的结果几乎没有影响。如表 4 所示，这类词出现的百分比为 $1166/101435 = 1.15\%$;因此，忽略它们是合理的。

4 信誉模型

在 3.1 节中，我们定义分数来衡量客户对产品的总体反应的积极/消极，基于它收到的评级和评论。基于此，在本节中，我们进一步引入了一个基于时间的指标——声誉——它直观地描述了一个产品在一段时间内积累的声誉。相应地，我们基于 RRBS 模型提供的度量(第 3 节)，建立了我们的声誉模型来量化产品的基于时间的声誉。

4.1 时间权重序列

直观地说，某一时刻的声誉多少依赖于它之前的声誉，而不是突然变化，无论它在过去或未来的状态如何。一般来说，一个产品在某一时刻的声誉与其整个生命周期的声誉是紧密相连的。带着这种心态，我们定义了时间权重序列的概念来描述时间 t 时声誉中先前声誉的影响，用 t 维行向量 γ 表示:

$$\gamma = (\gamma_1 \quad \gamma_2 \quad \dots \quad \gamma_m \quad \dots \quad \gamma_t)$$
 (16)

在 γ_m 分数在 m 中的影响权重是多少 $_{th}$ 月(以 1_{st} Month 是产品发布的第一个月)。我们考虑产品在国外的声誉 $_{th}$ 月在 $(m-1)$ 中更容易受到之前声誉的影响 $_{th}$ 月份比 $(m-2)$ 月份更容易受到影响 $_{th}$ 月。我们假设 γ 的增长 $_m$ 按时间是指数增长，并在此基础上进行归一化;因此，我们得到:

$$\gamma_j = \frac{a^j}{a^t}, \quad a > 1$$
 (17)

其中 a 是一个常量，其值设置为 1.1。

表 4:所有评论中常见形容词(如“good”)和“negative -adj”组合(如“not good”)的出现次数。形容词是通过一个开源来区分的

Python 文字处理库[2]和外观使用 Excel 进行统计。

	常见的形容词	数	“Neg-adj”组合。	数
	伟大的	4466	不确定	148
	好的	3055	不重	40
	小的	2963	不估组	34
总计	



关注数学模型
获取更多资讯

4.2 声誉量化

首先，根据公式(1)，我们定义了一个 t 维的列向量，用分数表示，来表示一个产品的 t 个月的分数：

$$score = \begin{pmatrix} score_{i,k}^{(1)} & score_{i,k}^{(2)} & \dots & score_{i,k}^{(m)} & \dots \text{分数} & \begin{pmatrix} t \\ i,k \end{pmatrix} \end{pmatrix}^T$$

(18)

其中 $score_{i,k}^{(m)}$ 为第 i 类产品的第 k 个模型在第 m 个月的得分。

然后将声誉的量化定义为 $Rep_{i,k}^{(t)}$ ，如式(19)所示

$$代表_{i,k}^{(t)} = \gamma \cdot \text{分数}$$

(19)

其中 I 表示产品型号的类型; K 表示 K_{th} I 的模型 $_{th}$ 类型产品; T 表示 T_{th} 月份(产品推出以来)。

从公式(17)我们知道时间权重序列 γ 遵循单调递增的指数模式，这意味着一个产品来自最近一个月的分数比来自遥远月份的分数对新声誉的影响更大。因此，通过对产品声誉的量化，我们应该能够解决观察和预测在线市场中产品声誉的增加或减少的任务。

4.3 模型评估

根据经验知识，一个有良好声誉的产品(即顾客对它有多好)可以比一个声誉较差的产品吸引更多的销量。基于此，我们可以说，产品声誉的增长可以刺激成比例的销售额增长，而另一方面，当产品声誉受损时，销售很可能会出现下降。

话是这么说的，我们可以解决评估我们的模型和措施的正确性的任务——这是棘手的，因为声誉这个术语非常抽象，没有统计数据供我们参考——通过比较我们量化的声誉的趋势和产品的实际销售的趋势(来自评级和评论记录)。因此，在我们用我们的度量方法量化一个产品的声誉之后，我们可以在量化的声誉趋势和实际销售趋势之间进行比较;这样，我们的模型和度量的正确性就得到了评估。

4.3.1 量化的声誉和销量之间的趋势相似性

我们从这三种产品中每一种选择了 3 款最畅销的车型来呈现并比较它们的美誉度和销量的趋势。然后，对于每个产品模型，我们得到一个时间序列²量化的声誉和销售的时间序列，并分别将它们呈现在散点图中，其中的点大致相连，以揭示趋势。每个产品模型的结果如图 2 所示。接下来，为了更全面，我们还从每类产品中选择了 1 个相当不好卖的模型，并进行同样的操作，结果如图 3 所示。

有了这两张图的结果，我们可以看到我们量化的声誉和实际销量的趋势

2A 时间序列是一系列按时间顺序索引(或列出或图形化)的数据点。



关注 数学模型

获取更多资讯

非常相似，可能高度相关。为了进一步证明这个想法，我们接着进行相关性分析。

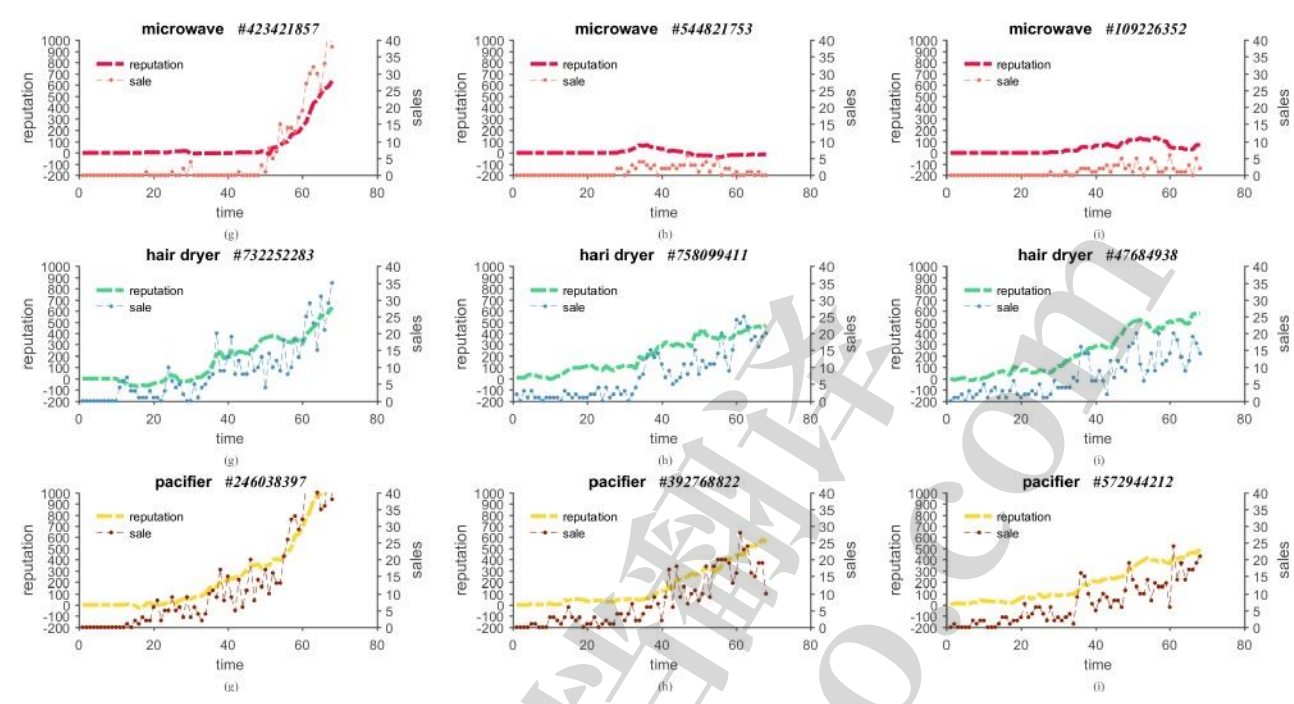


图 2.3 种产品中每种最畅销的 3 种型号的声誉和销量的可视化趋势。

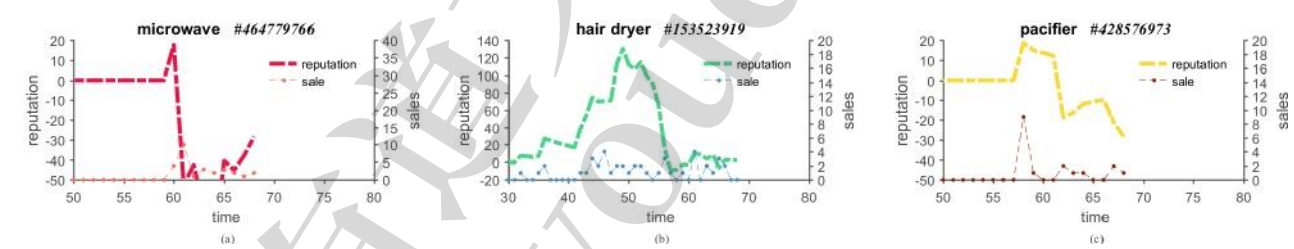


图 3.3 种产品中各有 1 种销售较差的模型的声誉和销量的可视化趋势。

4.3.2 Kendall’ s Tau 方法

为了评价美誉度和销量之间的趋势相似性，我们采用 Kendall’ s Tau 方法[9]来研究美誉度趋势和销量趋势之间的相关性。

Kendall’ s Tau 方法的过程如下:假设 $\{x_1, \dots, x_i, \dots, x_T\}$ 和 $\{y_1, \dots, y_i, \dots, y_T\}$ 是两个随机生成的时间序列，其中 $i = 1, 2, \dots, T$ ，我们分别通过 x 的值对两个序列进行排序操作 i 和 y_i 。然后，我们得到两个排序好的时间序列——比如 $\{x_{(1)}, x_{(2)}, \dots, x_{(T)}\}$ (命名为时间序列 $\{x_{(1)}, \dots, x_i, \dots, x_T\}$)和 $\{y_{(1)}, y_{(2)}, \dots, y_{(T)}\}$ (称为时间序列的顺序统计量 $\{y_1, \dots, y_i, \dots, y_T\}$)—where $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(T)}$ and $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(T)}$.

x_j 从排序序列 $\{x_{(1)}, x_{(2)}, \dots, x_{(T)}\}$ ，假设它在其 k 处 i_{th} 位置， $1 \leq k \leq T$ 的个数称为 x 的秩 j_i ，用 p 表示 j 。同理，从排序序列 $\{y_{(1)}, y_{(2)}, \dots, y_{(T)}\}$ 我们得到了 y_j 用 q_j 表示的秩。令 (x_i, y_i) 和 (x_{2j}, y_j) 与关注数学模型



$i = 1, \dots, T$ 和 $j = i + 1, \dots, T$ 是原始时间序列中的两个数据对。如果 $(p_j - p_i)(q_j - q_i) > 0$ ，我们说两个数据对是一致的;否则，我们说它们是不协调的。设 P 表示协调对的数量， Q 表示不协调对的数量，由此得出 $P + Q = N(N- 1)/2$,Kendall’ s tau 可定义为:

$$\tau_K(X,Y) \triangleq \frac{P-Q}{P+Q} = \frac{P-Q}{N(N-1)/2}$$

(20)

因此，我们知道当两个时间序列 $\{x_1, \dots, x_i, \dots, x_T\}$ 和 $\{y_1, \dots, y_i, \dots, y_T\}$ 具有完全相同的趋势，然后我们有 $\tau_K(x, y) \rightarrow 1$;当两个序列具有完全相反的趋势时，我们有 $\tau_K(x, y) \rightarrow -1$;而当两个序列完全不相关时，我们有 $\tau_K(x, y) \rightarrow 0$ 。也就是说， $|\tau_K(X, Y)| \approx 1$ 表明两个序列的趋势高度相关，而 $|\tau_K(X, Y)| \approx 0$ 表明它们完全不相关。

采用这种方法，与时间序列 $\{x_1, \dots, x_i, \dots, x_T\}$ 是被量化的声誉和 $\{y_1, \dots, y_i, \dots, y_T\}$ 作为销售，我们得到的结果(如图 4 所示)表明声誉和销售趋势之间的高度相关性:大多数计算的相关性都超过 0.6，这被认为是 strong。在安抚奶嘴组中计算的相关系数接近 0.8，这也表明了一个安抚奶嘴的声誉与其销量的相关性最大。

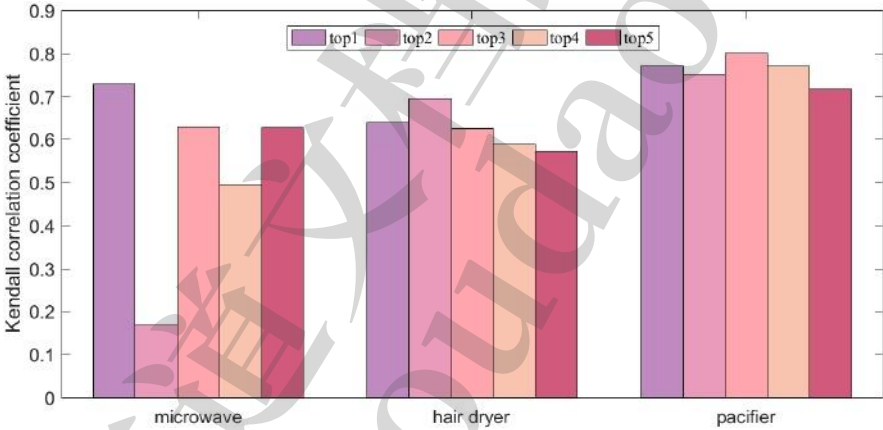


图 4:按产品类型划分为 3 组的不同模型的计算相关性。每一组包含 5 个模型的相关性，top1 表示本组中最畅销的模型，top2 表示第二畅销的模型，ramining 3 表示同样的方式。

5 成功预测模型

基于经济学知识，一个产品的成功可以很好地通过其随时间推移的销售/订单数量来量化。一种产品的销售量越多，它在市场上的优势就越大，因为它可以获得良好的利润，为可能的增强(这需要花钱)腾出空间。

在 4.3.1 节中，我们了解到一个产品的销售趋势与其声誉趋势高度相关。结合我们所掌握的知识，如果一个产品的美誉度增加，那么它的销量就有可能在相似程度内增加，因此我们可以假设这个产品是有可能的

成功;另一方面，一个产品的声誉下降可能意味着它潜在的失败。此外，我们注意到我们的销售额在一定程度上呈上升趋势。因此，如果我们



关注数学模型

获取更多资讯

将产品的成功程度与销量进行回归为了预测产品未来的成功或失败，预测可能很容易出错。

我们在第 3 节中构建的 RRBS 模型有效地结合了基于文本的评论中的星级评级水平和信息，而在第 4 节中引入的声誉模型提供了一种基于时间的衡量方法，很好地量化了产品随时间的声誉。考虑到我们之前提出的模型的强大能力，以及声誉与销售相比变化不那么突然的事实，我们意识到，在预测产品的成功时，通过将我们的声誉度量替换为实际销售，并根据计算的声誉序列预测产品的未来声誉，可能会是一个很好的方法。

5.1 高斯过程回归

高斯过程(GP)是一种通用的监督学习方法，旨在解决回归和概率分类问题，而使用高斯过程先验来推断连续值被称为高斯过程回归(GPR)[7]。我们认识到 GPR 有几个我们觉得方便的属性，例如，它在小型数据集上工作得很好，并且有能力提供预测的不确定性度量。

因此，我们采用 GPR 来拟合产品的声誉系列，从而预测产品在未来的声誉(晚于提供的数据中的最新标记日期)，并最终使预测产品未来的成功成为可能。

我们首先尝试估计最畅销的奶嘴模型的声誉(使用“parent_product”值#246038397)。其评级和审核记录的日期范围在 2010 年 1 月至 2015 年 8 月之间。我们首先用我们的声誉模型计算得到其声誉的时间序列，然后我们采用 GPR 对时间序列进行拟合(去掉 2015 年 7 月和 8 月的值)，并对其在 2015 年 7 月和 8 月的声誉进行预测，置信区间为 95%。结果显示，这样的短期预测具有较高的可靠性，如图 5 所示。

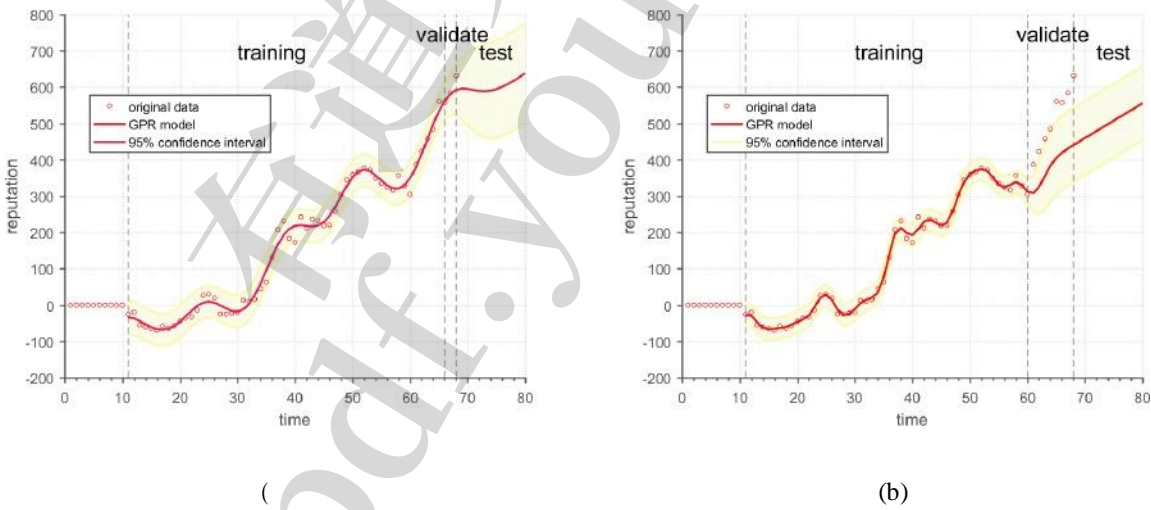


图 5:子图(a)是进行短期预测(预测几个月后的声誉)的结果;子图(b)是长期预测的结果(预测几年后的声誉)。

通过观察，我们意识到过度拟合的可能性，因此我们获得了相同模型声誉的时间序列(修剪日期范围为 2010 年 1 月至 2014 年 12 月)，并进行预测它在 2014 年 1 月之后的声誉。如图 5(b)所示，一些真实的数据点超出了



获取更多资讯

置信区间，表示我们预测的误差。经过深入分析，我们得出结论，误差的发生是因为对声誉的计算是有时间权重序列的(4.1 节)，使得遥远的声誉的影响较小。因此，在预测遥远的未来的声誉时，可能会有些不准确;因此，在进行预测时，应该使用所有现有的记录。

5.2 声誉映射

考虑到 2015 年 8 月之后的预测声誉可能在价值上有很大的差异，我们可能无法有效地找到收敛区间，这使得很难评估一个产品的成功。因此，我们寻求找到一种映射方法来映射我们的非收敛声誉预测，以便映射后的声誉在一个明显的收敛区间内波动，而映射前后的声誉值应该表明相同的趋势。我们采用以下函数作为我们的映射方法：

$$p_{i,k}(t) = \frac{1}{1 + e^{-\hat{Rep}_{i,k}(t)}}$$

(21)

其中- $\hat{Rep}_{i,k}(t)$ 为第 t 个月的预测声誉值。可以看出， $p(t) \rightarrow 1$ 如果 $\hat{Rep}_{i,k}(t) \rightarrow +\infty$;而 $p(t) \rightarrow 0$ 如果 $\hat{Rep}_{i,k}(t) \rightarrow -\infty$ 。因此，我们映射的声誉 $p(t)$ 具有收敛区间，同时也具有与原始的相同的趋势。

5.3 成功阈值

要决定一个产品是成功还是失败，我们需要确定一个门槛。首先将我们的阈值定义为一个常量，并将其值设置为 0.5:

$$threshold_{i,k}^{(t)} = 0.5$$

(22)

当评估一个产品未来的成功程度时，我们首先通过我们的声誉模型来计算它的评级和评论记录来获得它的声誉的时间序列，接下来我们对获得的时间序列进行回归分析并预测产品的未来声誉，然后将我们预测的声誉映射到 $p_{i,k}(t)$ ，最后，我们决定一个产品在未来是否会成功如果 $p_{i,k}(t) \geq 阈值_{i,k}^{(t)}$ 且如果 $p_{i,k}(t) < 阈值_{i,k}^{(t)}$

使用我们上面所述的方法，我们对一些模型在未来的潜在成功进行预测。表 5 列出了一些结果。

表 5:部分产品模型的成功预测。

预测/产品类型	吹风机	微波	奶嘴
会成功的	#732252283、#47684938、 #486774008、 #328811288……	#423421857、 #109226352、 #771401205、 #943347999……	#246038397、#392768822 #667171015、 #343148814……
注定失败	#862140913、#932955324 #945323010、 #768062995……	# 392967251, # 618770050…	#565545704、#671647927 #130912179、 #181638405…

不过，因为一套产品中的不同型号在发布上差别很大

日期(由他们第一次评分和评论记录的日期决定), 如果我们将阈值设置为
的成功预测可能会不准确。假设一个新设计的



有道文档翻译
pdf.youdao.com

微波是在 2015 年 8 月左右推出的，有很大的商业潜力，但由于它是一个年轻的产品，一开始的口碑也比较低，如果我们在评估未来的成功时忽略了产品的年龄，它在未来可能会被认为是一个失败的产品——很有可能不是。相应地，我们修改成功的阈值，使其成为时间的函数，如式(23)所示：

$$\text{阈值}_{i,k}(t)=0.5-0.1e^{-10(t-\tau_{th})}$$

(23)

其中 $\tau_{th} < t$ 表示 τ_{th} 是产品发布的月份。

修改后，我们重新评估了上述同一组模型，得到了略有不同的结果，其中电吹风 #945323010、#767062995 和奶嘴 #130912179、#181638405 被重新预测为未来可能成功。通过在数据集中查阅他们的评论和销量，我们发现这些重新预测是合理的，因为从统计学上看，它们似乎很有希望。因此，我们的模型在预测产品未来的成功时是可靠的。

6 结果

6.1 星级评分的激励性评价

在发现具体的星级评分是否煽动了更多的某类评论时，我们主要考虑两种情况：当看到一系列高评级时，客户是否倾向于写某类评论？而当看到一系列的低评分时，客户是否倾向于写某种类型的评论？

我们把高评级定义为星评级 4 或 5，而把低评级定义为星评级 1 或 2。为了更好地解决本节的任务，我们还将评论分为三种不同的类型，包括正面评论、中性评论和负面评论。正如之前在 3.3 节中定义的那样，类似于评论分数计算的过程——但这次我们不包括修正术语——我们考虑使用 β 进行评论 $i,k,j^{(t)} > 0$ 作为积极的评论；

(t)带 β 的评论 $i,k,j^{(t)} < 0$ 被认为是负面评论；和带有 β 的评论 $i,k,j^{(t)} = 0$ 被认为是中性的评论。

为了观察客户在看到一系列高评级后发布的评论的特定类型，我们可以将一组评论(属于相同的产品模型，由“parent_product”属性标识)按时间顺序排序，然后记录几次高评级发生后立即发布的评论类型。为了使我们的解决方案更加清晰，我们将 k 个评分(“高”或“低”类型中的任何一个)随后发生的事件定义为 k -event。因此，一旦高评级的 k 事件发生，我们可以记录下一个评论的类型，并在重复这个过程相当多的次数后——足够频率表明概率。通过改变 k 的值，我们可以推断客户在看到不同数量的后续特定评级后，是否会写某种类型的评论。

我们首先应用我们的方法，但只考虑在整个生命周期内(从他们的评论和评级记录的数量中)所有三种类型产品的销售额最高的前 5 个模型，以发现一些模式。所选的模型列在表 6 中。

然后我们遍历他们的高评分和低评分的 k -事件($k = 2,3,4,5$)，记录每 k -事件后的评论类型；这样我们就可以得到这三个事件的频率

评论类型(跟随 k 事件)。通过对模型的 num -进行加权求和
订单的 ber 作为权重，我们可以得到不同类型的近似概率



表 6:销量前 5 的车型。

产品类型	销量前 5 的型号
吹风机	#732252283、#758099411、#47684938、#235105995、 #694290590
微波	#423421857、#544821753、#109226352、#771401205、 #827502283
奶嘴	#246038397、#392768822、#572944212、#450475749、 #812583172

顾客可能在 k-event 之后发布的评论。结果列在表 7 和表 8 中。

$k =$ $k =$ $k =$

表 7:在高评分的 k 个事件($k = 2、3、4、5$)之后，三类评论出现的估计概率，分别为每种类型的产品测量。注意: P_+ 为评论类型为“正面”的概率; P_0 是评论类型为“中性”的概率; P_- 是评论是类型“否定”的概率。

$P_+/P_0/P_-$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
电 吹 风	0.5875/0.2040/0.2085	0.5899/0.2022/0.2078	0.5721/0.2141/0.2137	0.5723/0.2142/0.2134
微 波	0.4856/0.2981/0.2163	0.4710/0.3032/0.2258	0.5088/0.2982/0.1930	0.5301/0.2892/0.1807
奶 嘴	0.5839/0.2579/0.1582	0.5753/0.2650/0.1597	0.5892/0.2593/0.1515	0.5891/0.2518/0.1591

表 8:分别为每类产品测量的高评分 k 事件($k = 2,3,4,5$)后，三类评论出现的概率估计。

$P_+/P_0/P_-$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
吹风机	0.3040 / 0.1486/0.5472	0.3472 / 0.1736/0.4791	0.1666 / 0.1666/0.6665	0/0/1
微波	0.3750 / 0.2083 // 0.4167	0.2500 / 0.2500/0.5000	0.5000 / 0/0.5000	0.5000 / 0/0.5000
奶嘴	0.0714 / 0.6429/0.2857	0.4120 / 0.1675/0.4205	0.2675 / 0.3089/0.4236	0.1156 / 0.2320/0.6524

从表 7 中我们可以看到三种类型的产品有一些相似之处:当看到一个高评分的 k 事件发生时，所有类型产品的客户都最有可能发布正面评论，平均概率为 55.23%。因此，我们可以得出一个粗略的结论，顾客在看到一系列的高评分被发布后，会被鼓动去写正面评论。

然而，相比之下，表 8 中的结果看起来并不是很理想。对于吹风机来说，在一系列低评分之后出现负面评论的概率在正面评论和中性评论中是最大的，这是合理的。然而对于微波炉来说，当 $k = 4,5$ ，即负面评论和正面评论的概率相等时，就会出现不可信的结果。

事实上，我们发现畅销模型的低评级数量明显低于高评级的数量——这是有道理的，因为它们是畅销书。因此，后续的低评级很少出现，这使得我们的测量容易出错。为了得出更精细的结论，我们不再只覆盖热卖模型，而是将我们的测量范围扩大到每类产品中 95% 的模型。尽管如此，后续出现 4 个以上低评级的频率与高评级相比还是相当低的。因此，我们无法信心满满地评估低评分对客户差评的煽动。

综上所述，我们确实发现高评分会激发更多的正面评论;另一方面，我们无法确定低评分是否煽动了负面评论。



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

6.2 评论措辞与评分水平之间的关联

根据经验，我们假设当人们根据他们对产品的感受(可以从他们的评论中推断出)对产品进行评级时，评级水平可能与评论的内容有所关联。

为了验证我们的理论，我们计算了我们的评分词在所有评论中的出现，同时记录他们(即，单词出现在评论中)不同的相应的星级评级水平。注意，同一个单词在同一评论中多次出现只计算一次。基于此，我们将一个单词的星级定义为其特定的评级水平占与之相关的所有评级的百分比。

特别地，我们得到了一个 5 星率最高的前 5 个单词的列表(表 9)。从结果中可以看出，所有的单词都是积极形容词，5 个单词中有 4 个的等级为 3(最高强度级别)，每个单词的 5 星率都超过 80%。另一方面，我们也实现了一个 1 星率最大的 top-5 单词列表(表 10)。并且从结果中，我们可以看到所有的单词都是负面的单词，5 个单词中有 3 个的评分为 3，每个单词的 1 星评分都超过 35%(对于一星评分来说，这仍然是一个很高的比例——人们通常不会给那么低的评分)。

由于小样本量无法证明全部，我们接着计算所有频繁词(所有评论中单词出现次数超过 50 次)的高星级(星级评分超过 3)。而我们的结果显示，37.84%的词有超过 80%的高星级;所有这些词都是积极词，其中 60.87%的词具有最高的积极强度。另一端，我们也计算了所有频繁出现的单词的低星率(星级低于 3)，10.29%的单词的低星率超过 50%，这些单词都是负面词汇，100%的单词的负能量强度水平等于或大于 2。这样的结果揭示了情感词(也可以作为质量描述词)和评分等级之间的强烈关联。

表 9:5 星评分最高的 5 个词。完整的前 30 名词汇，请参见附录中的表 13。

词	数	年级	1-Star-Rate	2-Star-Rate	3-Star-Rate	4-Star-Rate	5-Star-Rate
太棒了	92	3.	0.0435	0.0109	0.0435	0.0326	0.8696
美妙的	138	3.	0.0072	0.0145	0.0290	0.0870	0.8623
爱	52	3.	0.0192	0.0192	0.0192	0.0962	0.8462
可爱的	241	3.	0.0166	0.0249	0.0332	0.0954	0.8299
最喜欢的	258	2	0.0078	0.0271	0.0271	0.1085	0.8295

表 10:1 星评分最高的 5 个单词。完整的 top-20 词见附录中的表 14。

词	数	年级	1-Star-Rate	2-Star-Rate	3-Star-Rate	4-Star-Rate	5-Star-Rate
危险的	52	3.	0.6346	0.0962	0.0769	0.0385	0.1538
可怕的	54	3.	0.5000	0.1111	0.0741	0.1481	0.1667
可怜的	94	2	0.4255	0.2447	0.1383	0.0638	0.1277
失望	196	2	0.3622	0.2245	0.1837	0.0867	0.1429
可怕的	61	3.	0.3607	0.1639	0.1475	0.1148	0.2131



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

6.3 我们的策略

对于阳光公司来说，为了提高他们在市场上的竞争力，我们充分利用了我们在模型中获得的知识，并提出了一些他们可以采用的营销策略和可能的重要设计特征。

6.3.1 营销策略

我们注意到，随着进入不同的特定市场而来的潜在竞争强度可能会有所不同，因为客户对特定市场内特定产品的既定偏好不同。例如，如果很大一部分客户是月光公司的奶嘴的忠实客户，那么阳光公司就很难诱使这些客户购买他们的奶嘴。

为了衡量客户的这种“偏好”，我们可以利用我们的声誉测量。我们可以对所有产品模型的声誉进行加权平均，这些声誉是由产品的销量加权得到的。然后，通过结果的规模，我们可以判断一个市场是否建立了某种强烈的忠诚度。使用这种方法，我们得到了表 11。可以看到，安抚奶嘴市场的忠诚度最强，而微波炉的忠诚度最弱。也就是说，在安抚奶嘴市场上竞争，是最费力气的。

据此，我们提出以下营销策略：

- 如果阳光公司有足够的精力，我们建议他们在销售安抚奶嘴上花更多的精力，例如，分配更多的广告预算给他们的安抚奶嘴；
- 如果阳光公司没有足够的精力腾出来(由于短缺或预算有限)，我们建议他们将有限的精力更多地投入到销售微波炉上，因为他们可以在至少一个市场上追求优势；
- 无论如何，阳光公司都应该不断跟踪我们的分数和声誉衡量他们的产品，因为他们可以很好地反映客户对产品的看法。

6.3.2 重要的设计特征

为了确定客户最关心的重要特征，我们计算了每类产品的所有评论中单词的外观。进行一些操作来提取表明设计特征的质量描述符：

- 1 省去不是名词的词，因为名词描述的是客户需要什么具体特征；
- 2 省略不代表产品具体质量的词。

表 11:三个不同市场的竞争强度估计。

产品类型	声誉权重平均
吹风机	239.7633
微波	184.0208



关注数学模型
获取更多资讯

然后我们得到了每类产品提到最多的前 3 个特征的表格，如表 12 所示，它们代表了阳光公司在设计产品时最应该考虑的最重要的特征。从中我们可以假设客户期望所有三类产品的合理价格。而具体地说：

- 客户可能希望吹风机能够产生足够的热量，并有高质量的电线；
- 微波炉需要有足够的空间，客户服务也很有价值，因为机器随时可能坏。
- 安抚奶嘴是给婴儿用的，父母理所当然地希望它们有良好的质量和ap-大小合适，让宝宝咬起来舒服。

表 12:每类产品提到最多的前 3 个特点。

	吹风机	微波	奶嘴
词(计数)	价格(1063)	空间(183)	质量(405)
	热量(960)	价格(147)	规模(395)
	线(805)	服务(137)	价格(333)

7 敏感性分析

我们对两个主要参数进行灵敏度分析： λ (式(1))和 a (式(17))。首先，我们将 a 的值从- 10% 更改为+10%，每次步长为 5%。结果如图 6(a)所示，在 38 之前的信誉度趋势大致只略有变化，但随着时间戳的推移，它变得相当敏感。这是合理的，因为 RRBS 模型中的分数是指数加权和求和以获得声誉。因此， a 是一个需要仔细确定的关键参数。然后，类似地，我们将 λ 的值从- 10%改变为+10%，并评估其灵敏度。如图 6(b)所示，它似乎没有明显的敏感性。

综上所述，我们的两个参数 a 中的一个被测试为敏感参数。但总的来说，我们的模型具有相当程度的鲁棒性。

8 优点和缺点

8.1 优势

- 信息的充分利用: 在设计我们的措施时，我们不仅结合了星级评分和评论内容，还考虑了提供的数据中的其他有用属性；
- 良好的模型表现: 我们量化的声誉趋势与实际销售之间的高度相关性表明我们的模型具有良好的表现，我们的测量是准确的；
- 良好的方法论选择: 我们揭示了积极词汇之间的强烈联系以及高评分，负面和低评分，多亏了我们选择的方法;关注数学模型



获取更多资讯

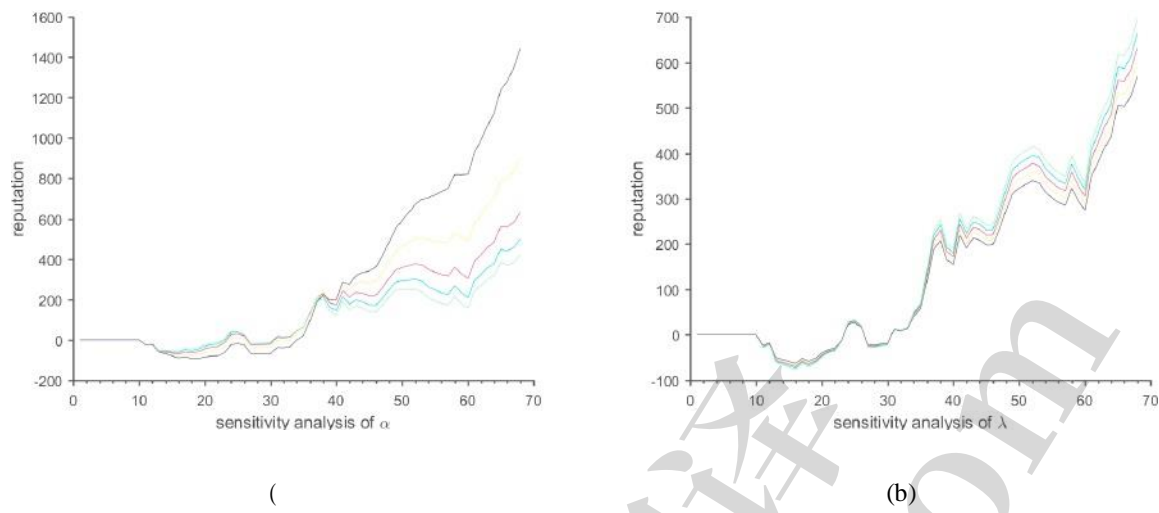


图 6:我们两个主要参数 λ 和 α 的灵敏度分析结果。

- 鲁棒性高:一般来说, 我们的模型对参数-的值变化不敏感发疯。

8.2 弱点

- 模型设计中一定程度的任意性:当设计我们的部分模型;
- 我们获得的词集的局限性:我们在 *RBSS* 模型中用于匹配评论内容的词集在一定程度上缺乏大小, 在测量产品得分时遗漏了一些信息;

9 结论

在我们的工作中, 我们提出了 *RRBS* 模型和声誉模型, 以提供一种基于评分、基于评论和基于时间的度量, 精细地描述了一个产品在一段时间内的声誉。使用这个度量, 我们能够预测一个产品未来潜在的成功或失败。我们还发现, 一系列的高评级可以激发更多的正面评论, 但无法确定一系列的低评级是否具有类似的效果。通过词匹配和外观计算, 也发现了评分水平与评论的特定质量描述符之间的强关联。最后, 我们向阳光公司提供了我们设计的营销策略, 并指出了一些他们可以研究的产品设计特征, 从而提高他们在市场上的整体竞争力。



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

10 我们的信

来自:团队#2004647

收件人:阳光公司市场总监时间:2020 年 3 月 9 日

尊敬的阳光公司市场总监:

感谢您聘请我们为顾问!我们已了解你们的具体要求,并对我们任务的可行性进行了充分评估。在这里,我们向您提供我们的详细解决方案。

我们彻底检查了与竞争对手产品相关的评论和评级数据。我们发现,数据的很多属性,以及实际的评分水平和评论内容,都可以被利用,从而得出一个信息量相当大的衡量标准,供你追踪。为了对它们进行量化,我们提出了基于评分和评论的得分模型,它引入了我们定义的数据度量——得分。在我们的模型中,一个产品的分数描述了客户对该产品的反应。它基于评级和文本评论。在计算一个产品的评分时,我们会考虑:

- 其收到的评级水平;
- 它的评论是否来自藤评家;
- 评论内容的情感强度(如“公平”不如“精彩”强烈,两者都是褒义词);
- 其评论是否来自经过验证的买家;
- 评论的有用性评级,这有助于过滤已经存在的评论
对其他客户没有帮助的;
- 评论的长度。

通过将这些因素结合在一起,我们的评测应该能够从各种声音中挑选出最有价值的声音,一旦你的产品发布,你的公司应该高度重视这些声音,因为它们包含了至关重要的信息,让你知道你的产品有什么问题,以及如何改进。

基于我们的分数衡量,我们进一步提出了声誉模型,它直接定义了一个产品的声誉-这是一个基于时间的衡量方法,用于量化一个产品在一段时间内在客户中的声誉。该度量被证明在描述产品的实际(但抽象)声誉方面是相当准确的,由量化的声誉趋势和产品的销售之间的强相关性支持。

经验知识表明,产品的成功与其销量密切相关。因为我们知道我们的声誉测量是可靠的,并且与产品的销售密切相关,我们可以利用它来表明产品未来的成功。我们通过采用高斯过程回归来解决这个问题,事实证明,我们可以根据计算出的声誉的时间序列,很好地预测一个产品在不久的将来的声誉。因此,我们能够在短期内预测一个产品在未来的潜在成功或失败;虽然在做长期预测时,它缺乏准确性。

至于大家所关心的具体问题,我们想了解一下 se-
特定评级的 Ries 是否会引发某种类型的评论。通过记录客户的评论权关注数学模型



获取更多资讯

在一系列相同的评分之后，通过计算它们是积极的、消极的或中性的几率，我们成功地发现，一系列高分确实会引发更多的积极评论——这表明声誉显著增加。但是，由于缺乏证据，我们无法确定一系列的低评分是否会煽动负面评论。

为了解决另一个问题，即具体的质量描述符是否与评分水平存在强烈关联，我们首先获得了一个包含 2477 个单词的词集，这些词都是根据其情感强度进行评分的(说“精彩”比“公平”更积极强烈)。然后，我们匹配提供的数据中的所有评论内容，并计算单词的外观和相关评分水平。结果显示它们之间明显有很强的联系。例如，在所有包含“awesome”(正面强度最大)一词的评论中，86.96%的评论与 5 星评级相关。另一方面，在所有包含“危险”(负面强度最大)的评论中，有 63.46%的评论与 1 星评分相关。

最后，通过我们的建模和方法论，我们设计了一套你可以采用的营销策略，以及一套客户最看重的每个产品的设计特征。我们首先对每个市场上的所有产品模型的声誉进行加权平均，这些声誉是由产品的销量加权的，因此我们有一个衡量你的公司将遇到的潜在竞争的方法，因为一个市场上的平均声誉可以表明该市场的客户忠诚度。结果显示，在安抚奶嘴市场上竞争所花费的精力最多，在微波市场上竞争所花费的精力最少。因此，我们提出以下营销策略：

- 如果您的公司有足够的精力，我们建议您花更多的精力销售安抚奶嘴，例如，分配更多的广告预算给您的安抚奶嘴；
- 如果您的公司没有足够的精力(由于短缺或预算有限)，我们建议您将有限的精力更多地放在销售微波炉上，因为您可以追求至少一个市场的优势；
- 无论如何，您的公司都应该不断跟踪我们的分数和声誉测量因为它们可以很好地反映客户对产品的看法。

随后，我们将评论与质量指定名词进行匹配，并计算它们的外观。然后我们为每个产品挑选了 3 个最具外观的词，它们表明了客户最重视的设计特征。结果显示，顾客期望所有三种产品都有一个合理的价格。而具体分析：

客户可能希望吹风机能够产生足够的热量，并有高质量的电线；
微波炉需要有足够的空间，客户服务也很有价值，因为机器随时可能坏。
安抚奶嘴是给婴儿用的，父母理所当然地希望它们有良好的质量和ap-
大小合适，让宝宝咬起来舒服。

如果你想了解更多细节，可以参考我们的论文。我们很乐意与您讨论我们的解决方案细节。



关注数学模型
获取更多资讯

参考文献

[1] i.亚马逊。COM, 什么是 amazon vine?, 于 2020 年 3 月 7 日登陆。

[2] S. BIRD 和 E. LOPER, NLTK: The natural language toolkit, 发表在 2004 年 7 月西班牙巴塞罗那的 ACL 互动海报和演示会议论文集上, 计算语言学协会, 第 214-217 页。

[3] a . CHUA 和 S. BANERJEE, 《理解评论有用性作为评论人声誉、评论等级和评论深度的函数》, 《信息科学与技术协会学报》, 66(2014)。

[4] Y. KOREN、R. BELL 和 C. VOLINSKY, 《推荐系统的矩阵分解技术》, 《计算机》, 2009 年第 42 期, 第 30-37 页。

[5] J. MCAULEY 和 J. LESKOVEC, Hidden factors AND Hidden topics: Understanding rating dimensions with review text, Proceedings of the 第七届 ACM Conference on Recommender Systems, RecSys 13, 纽约, 美国, 2013 年, 计算机械协会, 第 165172 页。

[6] f . Å. 尼尔森, 阿芬, 2011 年 3 月。

[7] C. E. RASMUSSEN 和 C. K. I. WILLIAMS, 《机器学习的高斯过程(自适应计算和机器学习)》, 麻省理工学院出版社, 2005 年。

[8] Y. TAN、M. ZHANG、Y. LIU 和 S. MA, rating - latent topics: Understanding users AND items with ratings AND reviews, 《第二十五届人工智能国际联合会议论文集》, IJCAI16, AAAI 出版社, 2016, 第 26402646 页。

[9] W. XU, C. CHANG, Y. S. HUNG, S. K. KWAN, P. C. W. FUNG, Order statistics correlation coefficient as a novel association measurement with applications to biosignal analysis, IEEE Transactions on Signal Processing, 2007, 55 (2007), pp. 5552-5563。



关注数学模型
获取更多资讯

附录

afinn-matcher.py

#这个程序用于将评论与 AFINN word set import csv 提供的分级词进行匹配

进口 nltk

进口再保
险

来自 afinn import afinn

Def split(分隔符，字符串，maxsplit=0):

 regexPattern = '|' .join(map(re.escape, delimiters))
 return re.split(regexPattern, string, maxsplit)

Delims = "， .;?<>()!"

af = AFINN(语言='en')

Adjs = {}

Files=["安抚器.csv"、"微波.csv"、
"hair_dryer.csv")

CNT =
0

对于文件中的文
件:

 open(file, "r", encoding="utf-8", errors='ignore')为 f:
 reader = csv.DictReader(f)

 对于 line in reader:

 CNT +=
 1

 若 CNT % 100 == 0:

 print(“读行” + str(cnt) +
 , dict size= " + str(len(adjs))

 评论=line["review_headline"] + line["review_body"]
 toks = split(delims, 评论)

 toks = set(滤镜(None, toks))

 #打印(托克
 斯)

 Words = nltk.pos_tag(toks)

 For I in range(len(words)):

 Tup = words[i]

 if tup[1] == "JJ":

 词 = tup[0]

 Word = Word.lower()

 Score = af.score(word)

 如果 score ==
 0:

 继续

 Adjs [word] = [sum(x) for x in zip(

 adj. 得到(词,[0]* 7),[1,0,0,0,0,0,0))评级 =
 int((“star_rating”)行)

 如果评级> 5 或评级< 1:

 继续

Adjs [word][rating] += 1
Adjs [word][6] = score

Out = "matched_words.csv"

以 open(out, "w", encoding="utf-8")
作为 f: writer = csv. Writer (f,
delimiter=", ")为键，值在 adjs.items():
行 =[键]+值

writer.writerow(行)



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

RRBS-reputation.m

```
因为我= 2:长度(S (:
1))
    指数=找到(项目== cell2mat
(S(我,5)));
    如果国防部(计时
器,12)~= 0
        如果 review_date(我,2)~=国防部计
时器,12)
            计时器=时间+
1;
        结束
    其他的
        如果
review_date(我,2)= 1
            计时器=时间+
1;
        结束
    结束
    如果~ isempty(指
数)
        如果
K(我,1)= 0
            β = 1;
        其他的
            如果
K(我,1)> 0
                β =葡萄树(我,1)* K(我,1)^ (我,1)验证* exp (helpful_of_total_votes(我,1))* 0.5
                *瞧
                β =葡萄树(我,1)* (1)* (abs (K(我,1)))^ (我,1)* exp (helpful_of_total_votes 验证
(结束
            结束
        记录(定时器 1 指数)=(定时器 1 指数)+ 1;
        记录(定时器 2 指数)=(定时器 2 指数)+ rating_123(我,1)。* β ;
    结束
结束

开始= 0 (1);
α = 1.1;
指数= 1:长度(记录(1 1:))
    exponent_window_origina =
[];
    开始(指数)=找到(记录(:,1,指数)> 0,1,'第一次');
    记录(1:开始(指数)1,3,指数)= 0(开始(指数)1,1);j = 1:长度(记
录(:,1,1))开始(指数)+ 1
        exponent_window_origina = exponent_window_origina α ^ [j];
    结束
因为我=开始(1、索引):长度(记录(:,1,1))
    exponent_window = exponent_window_origina./ α ^ (我开始(指数)+ 1);
    记录(我 3 指数)= exponent_window(1, 1:我开始(指数)+ 1)*记录(开始(指数):我,2
```


有道文档翻译
pdf.youdao.com



关注数学模型
获取更多资讯

表 13:五星评分最高的 30 个单词

词	数	年 级	1-Star-Rate	2-Star-Rate	3-Star-Rate	4-Star-Rate	5-Star-Rate
太棒了	92	3.	0.043478261	0.010869565	0.043478261	0.032608696	0.869565217
美妙的	138	3.	0.007246377	0.014492754	0.028985507	0.086956522	0.862318841
爱	52	3.	0.019230769	0.019230769	0.019230769	0.096153846	0.846153846
可爱的	241	3.	0.01659751	0.024896266	0.033195021	0.095435685	0.829875519
最喜欢的	258	2	0.007751938	0.027131783	0.027131783	0.108527132	0.829457364
耐用	174	2	0.011494253	0.040229885	0.040229885	0.103448276	0.804597701
神奇的	102	3.	0.029411765	0.029411765	0.058823529	0.078431373	0.803921569
完美的	467	3.	0.023554604	0.019271949	0.029978587	0.160599572	0.766595289
优秀的	164	3.	0.042682927	0	0.036585366	0.158536585	0.762195122
美丽的	57	3.	0.087719298	0.01754386	0.070175439	0.087719298	0.736842105
可爱的	338	2	0.00591716	0.026627219	0.082840237	0.147928994	0.736686391
负担得起的	56	2	0	0	0.071428571	0.196428571	0.732142857
很高兴	59	3.	0.016949153	0.033898305	0.016949153	0.203389831	0.728813559
容易	1490	1	0.01409396	0.019463087	0.040268456	0.203355705	0.722818792
很高兴	222	3.	0.018018018	0.022522523	0.031531532	0.207207207	0.720720721
满意	55	2	0.018181818	0.054545455	0.054545455	0.163636364	0.709090909
超级	204	3.	0.034313725	0.029411765	0.078431373	0.156862745	0.700980392
健康的	50	2	0.04	0.08	0.04	0.14	0.7
伟大的	3706	3.	0.031840259	0.036157582	0.062601187	0.177010254	0.692390718
有趣的	78	3.	0.064102564	0.064102564	0.051282051	0.128205128	0.692307692
免费的	208	1	0.057692308	0.033653846	0.067307692	0.158653846	0.682692308
快乐	848	3.	0.053066038	0.048349057	0.036556604	0.183962264	0.678066038
爱	163	3.	0.055214724	0.049079755	0.049079755	0.171779141	0.674846626
清洁	494	2	0.046558704	0.046558704	0.062753036	0.172064777	0.672064777
推荐	114	2	0.087719298	0.070175439	0.052631579	0.122807018	0.666666667

有效的	55	2	0.072727273	0.036363636	0.018181818	0.218181818	0.654545455
自然	243	1	0.057613169	0.037037037	0.065843621	0.189300412	0.650205761
价值	299	2	0.107023411	0.043478261	0.053511706	0.157190635	0.638795987
强大的	792	2	0.025252525	0.041666667	0.112373737	0.198232323	0.622474747
可靠的	52	2	0.096153846	0.038461538	0.096153846	0.153846154	0.615384615



关注数学模型
获取更多资讯

表 14:1 星评分最高的 20 个单词

词	数	年 级	1-Star-Rate	2-Star-Rate	3-Star-Rate	4-Star-Rate	5-Star-Rate
危险的	52	2	0.634615385	0.096153846	0.076923077	0.038461538	0.153846154
可怕的	54	3	0.5	0.111111111	0.074074074	0.148148148	0.166666667
可怜的	94	2	0.425531915	0.244680851	0.138297872	0.063829787	0.127659574
失望	196	2	0.362244898	0.224489796	0.183673469	0.086734694	0.142857143
可怕的	61	3	0.360655738	0.163934426	0.147540984	0.114754098	0.213114754
无用的	56	2	0.321428571	0.25	0.196428571	0.142857143	0.089285714
错误的	129	2	0.271317829	0.108527132	0.131782946	0.100775194	0.387596899
坏	353	3	0.246458924	0.101983003	0.144475921	0.175637394	0.331444759
积极的	56	2	0.232142857	0.107142857	0.089285714	0.107142857	0.464285714
某些	51	1	0.176470588	0.058823529	0.215686275	0.215686275	0.333333333
弱	57	2	0.175438596	0.245614035	0.263157895	0.140350877	0.175438596
安全	111	1	0.171171171	0.045045045	0.063063063	0.135135135	0.585585586
清晰的	149	1	0.167785235	0.046979866	0.134228188	0.22147651	0.429530201
困难	234	1	0.132478632	0.128205128	0.183760684	0.247863248	0.307692308
宽松的	50	3	0.12	0.12	0.12	0.08	0.56
前	123	2	0.113821138	0.146341463	0.089430894	0.170731707	0.479674797
硬	372	1	0.112903226	0.10483871	0.13172043	0.174731183	0.475806452
想要	63	1	0.111111111	0.063492063	0.111111111	0.158730159	0.555555556
价值	299	2	0.107023411	0.043478261	0.053511706	0.157190635	0.638795987
负	123	2	0.105691057	0.032520325	0.056910569	0.284552846	0.520325203

有道文档翻译
pdf.youdao.com



关注数学模型
获取更多资讯