

狮身人面像之谜:破解亚马逊的秘密 评分和评论

总结

我们见证了大众在线市场的崛起。例如，亚马逊，最大的在线平台之一，价值约 9150 亿美元。在客户执念原则的指导下，它为客户提供了一个从 1 到 5 的产品评级的机会。此外，买家可以提交一条文字信息，即评论，来表达他们对产品的感受。这些评分和评论的海量数据提供了大量有待挖掘的信息。基于文本的信息分析或基于评分的价值分析已经受到广泛关注，但还没有一种方法可以将两者结合起来，特别是对于在线市场而言。

为了应对上述挑战，本文提出了一种新颖的 CE-VADER 混合模型用于评论中的情感分析，将消息分为强正面、弱正面、中等、弱负面和强负面五组。实证结果表明，所提出的五组分类模型与五星评级体系具有良好的相关性。在此基础上，提出了一种结合基于文本和基于评分的信息评价模型。我们挑选出每个产品 1% 信息量最大的评论和评级来评估属性并提出销售策略。

我们提出基于文献中的微分方程模型的“声誉”率来评估产品的声誉。然后，我们采用自回归(AR)模型作为时间序列预测方法，预测未来的“声誉”率和每个产品的潜在成功或失败。AR 模型在验证集上显示出较高的准确性，最大均方根误差(RMSE)为 0.131。安抚奶嘴的声誉很好，预测会成功，而微波炉和吹风机的声誉很差，预测会失败。结果显示，与连续五星或一星评级序列的比例相关。最后，我们分析了特定的单词和描述符，以发现它们与评分的相关性。

根据我们的实证结果，我们提出了一些自信的销售策略和建议，如引入市场的时机选择，根据星级评分有针对性的调整等。我们给阳光公司的市场总监写了一封信，总结我们的分析和结果，以及我们的建议。

我们的框架显示出了很强的准确性、鲁棒性。它可以很容易地用我们的源代码实现到其他数据。

关键词:基于文本的度量，信息性文本选择，信誉量化，销售策略形成。



关注数学模型
获取更多资讯

狮身人面像之谜:破解亚马逊评分和评论的秘密

2020 年 3 月 9
日

内 容

1 介绍	3
2 个假设和符号	4
2.1 假设	4
2.2 符号	4
3 信息化评价模型	4
3.1 星评向量编码形式.....	5
3.2 基于文本度量的上下文熵维德混合模型.....	5
3.2.1 人工标注种子词.....	7
3.2.2 上下文熵块(CE).....	7
3.2.3 维达块	9
3.2.4 提出 CE-VADER 用于情感分析.....	9
3.3 基于文本和基于评分的措施相结合.....	10
3.4 模型实现、敏感性分析和结果.....	11
4 测量基于时间的模式的差分方程	11
4.1 基于差分方程模型	11
4.2 模型实现、灵敏度分析及结果.....	12
5 预测潜在的成功或失败	14
5.1 预测未来声誉的时间序列.....	14
5.2 评估成功或失败的潜力.....	14
5.3 模型实现和结果	14
6 具体评级和描述分析	15
6.1 具体星级评级与评级频率相关.....	16
6.2 具体质量描述符与评分等级的相关性.	18



有道文档翻译
pdf.youdao.com

6.2.2 模型实施与结果.....	19
设计特征的吸引力分析	20
8 个销售策略和建议	个
优点和缺点 21	
9.1 优势	21
9.2 缺点	22
10 结论 22	
给阳光公司市场总监的一封信	
附录	24
附录 A 标注了种子词和频次	24
附录 B 不同关键词组中关键词出现的次数附录 C 最具信息量的评分和评论前 1%	25
附录 D VADER 情感分析源代码	26
附录 E 信息量算法源代码	36
附录 F 信誉计算源代码	36
附录 G 贝叶斯模型源代码	37
附录 H 时间序列预测源代码	37
附录 一词云图片源代码	38
	38

有道文档翻译
pdf.youdao.com



关注数学模型
获取更多资讯

1 介绍

我们的社会见证了许多在线市场的崛起，全球总市值为 4.3 万亿美元[1]。与传统平台相比，在线市场的一个显著特征是对文本和评级的大规模审查。其中，最受关注的是亚马逊，作为其最成功的[1]。亚马逊还为顾客提供自由表达感受的机会，并对他们购买的产品进行评分。

之前的工作[2]表明，客户在购买平台上的产品之前，会很大程度上参考评论和评分。平台可以通过查看这些评论来调整销售策略。因此，评分和评论既可以为其他潜在买家提供参考，也可以为分析客户需求提供海量数据，有助于制定适应性策略。通过充分利用这些数据，可以实现买家和平台的双赢。

其中最大的挑战之一是评论文本的复杂性和多样性[3,4]。本文提出了一种新颖的情感分析模型作为基于文本的度量来解决这个问题。在本文中，我们开发了一系列模型，作为基于文本、基于评分和基于时间的度量的组合，以挑选出信息量最大的评分和评论进行跟踪。我们还构建了一个新颖的评价框架，以量化每个产品的声誉，并预测潜在的成功或失败。然后，我们分析了连续同星评分、词描述符和产品声誉之间的相关性。我们将我们的模型在三种不同类型产品生成的真实数据集上实现，这三种产品分别是安抚奶嘴、微波炉和吹风机。

研究人员指出，有必要研究在线平台应在何时以及如何根据消费者评论或评级[5]调整营销传播策略。我们在本文中根据我们的分析和结果提出了几个销售策略和建议。

论文的其余部分组织如下。在第 2 节中，我们列出了模型构建中的主要假设，并介绍了本文中经常使用的符号。在第 3 节中，提出了一种新颖的信息评估模型。它由最先进的 CE[6]和 VADER[7]混合组成，用于评论文本中的情感分析。然后，我们提出“重要性”率作为一种基于文本的度量(即重要性)的组合。，即我们提出的 CE-VADER 模型)和基于评级的度量(即，星级评级和有用的投票)来表明评论和评级的信息量有多大。据我们所知，我们首次提出了一种基于评论文本的情感分析模型。在第 4 节中，我们采用差分方程模型作为骨干来测量每个产品的时间模式。此外，在这一节中提出了“声誉”率来衡量声誉的增长或下降。在第 5 节中，我们采用了一个自回归模型(AR)来预测声誉在未来时域的变化，并提出了一个模糊系统来预测每个产品的潜在成功或失败。关于我们的模型在给定数据上实现的结果的更多细节可以在第 6、7、8 节中找到。第 9 节讨论了所提出的模型和框架的优点和缺点。我们在第 10 节中总结。所有源代码都附在附录 D-I 中，可以很容易地实现到其他数据集上。



关注数学模型
获取更多资讯

2 个假设和符号

2.1 假设

为了简化我们的模型并消除复杂性，我们在这篇文献中做了以下主要假设。所有的假设一旦被用于我们模型的构建中，就会被重新强调。

假设 1。在线市场运行稳定。也没有出现疫情爆发等严重影响网购产业链的情况。

假设 2。评级和评论描述了顾客对他们购买的产品的真实体验和感受。评论文本中的情感反映了一个人对产品的感受。

假设 3。客户的绝大多数个体差异，如经济地位和教育水平，都被忽略了。

假设 4。发货需要一些时间。一些客户更希望在收到购买的产品后的一段时间内进行评论。

假设 5。消费者在购买产品时更关注负面评论，如低星评分或差评。

2.2 符号

在这项工作中，我们在模型构建中使用了表 1 中的命名法。其他不常用的符号一旦被使用，就会被介绍。

表 1:本文献中使用的符号

象征	定义	类型
id	评论 id	字符串
sid	星级，下标是其关联评论 id	标量
hvid	有帮助的投票，下标是其关联评论 id	标量
掉	评论文本，下标是其关联评论 id	字符串
rdid	评论日期，下标是其关联评论 id	日期
VEC	5 类种子词相关强度的向量编码	映射
INT	评论重要性率和关联评级	映射
小鬼	某一时期产品的信誉率	映射
代表		映射

3 信息化评价模型

在这一节中，我们提出了“重要性”来评估评论文本和星级的信息量。我们考虑的信息量最大的因素是评论文本的情感。在这篇文献中，我们提出了 CE-VADER 模型来解决评论文本中的情感分析问题。我们的模型将文本分为五组:强正面、弱正面

正面、中度、弱负面和强负面在五星评级方案的一致性上。然后我们提出的“重要性”将纳入基于文本的衡量标准，星级评级



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

与他们的忠诚，相关性。重要性越高，信息量就越大。这一节的其余部分安排如下。在 3.1 节中，我们将积分星率转换为向量形式。在 3.2 节中，我们提出 CE-VADER，一种用于基于文本的度量的混合模型。在 3.3 节中，我们引入了“重要性”来计算评论和星级评级一起的信息量。在第 3.4 节中，我们在 3 类产品的真实数据集上实现了我们的模型，以指示 1% 信息量最大的评论和星级评分，并分析模型敏感度。

3.1 星级评分的向量编码形式

消费者在购买后可以在亚马逊上自由地对产品进行 1 - 5 星的评价。一星代表满意度最低，五星代表满意度最高。1 - 5 星的评分本身就是一个足够的衡量标准。将基于评分的衡量与基于文本的衡量相结合，我们将在下一节中讨论。我们希望在本节中将星级评级转换为矢量形式。

首先，我们分别从给定的数据中计算出电吹风、婴儿安抚奶嘴和微波各星率的比例，如图 1 所示。我们观察到，婴儿安抚奶嘴获得高星级评级的比例最高，而微波的星级评级较低。科技含量高的产品也面临更多的质量问题，这符合实际预期，说明星级评分确实可以反映消费者的满意度。

不过，我们希望将评分转换为一种等效的 5- 维向量编码形式。表示星率为 $s \in \{1,2,3,4,5\}$ ，s 的向量编码形式可以用 $VEC(s) = (VEC^1_s, \dots, vec^5_s)^T \in R^5$ 其中定义的组件为:

$$vec^i_s = \frac{e^{\frac{|i-s|^2}{2\sigma_0}}}{\sum_{j=1}^5 e^{\frac{|j-s|^2}{2\sigma_0}}}$$

(1)

在 σ_0 是一个可调参数，决定了我们模型的鲁棒性，越大鲁棒性越强。映射 VEC 是一对一的，因此我们声称转换后的形式等同于星级评级。此外，根据我们的定义，我们可以发现: i) $s = \operatorname{argmax}_i \{vec^i_s\}$, (二) $\sum_{i=1}^5 vec^i_s = 1$ 。VEC(s) 编码为每个分量的概率向量表示它们被相关星评级的可能性，例如，4 星率被评级 4 星的可能性最高，被评级 3 星或 5 星的可能性第二高。

3.2 上下文熵 VADER 混合模型用于基于文本的度量

在这篇文献中，我们构建了一个新颖的基于评论文本的情感分析模型。为了简单起见，文本得分的情感被视为衡量产品成功或失败的唯一事实，例如，积极的态度通常表明产品成功的可能性更高，而相反，消极态度表明产品失败的可能性更高。

在本节中，我们提出了一种上下文熵和 VADER[7]混合模型，即 CE-VADER 来解决评论文本中的情感分析挑战。该模型制作

由两个模块组成:上下文熵(CE)模块和 VADER 模块。CE 模型在对股市新闻[6]进行情感分析时表现出了很高的能力，但在简言之存在局限性



关注数学模型
获取更多资讯



图 1:根据给定数据，安抚奶嘴、微波炉、吹风机的星级评分分布。

背景例如，这篇文献中的综述。虽然 VADER 模型在简短的在线纹理方面优于最先进的自然语言模型，但它们的准确性在很大程度上取决于预先列出的词典单词。CE-VADER 的两个块将分离结果一个 5 维概率向量，每个分量代表作为其关联组的概率。在一个投票块之后，我们提出的模型将把评论分类到五个组中的一个，以及一个强度，显示它被分类到组的强度。通过混合这两个模型，我们表明 CE-VADER 可以在星级评级的一致性下将评论上下文分类为五个组。

本节的其余部分安排如下。首先，我们展示了我们为 CE 模型生成种子词的策略，并扩展了 VADER 块的黄金标准列表。然后我们最高级地介绍了 CE 和 VADER 模型。最后提出了 CE- VADER 混合模型。我们的模型将评论上下文根据其强度分为五类:强正面、弱正面、中等、弱负面和强负面。在下一节中，我们将根据分类结果和强度提出一个评论保真度。

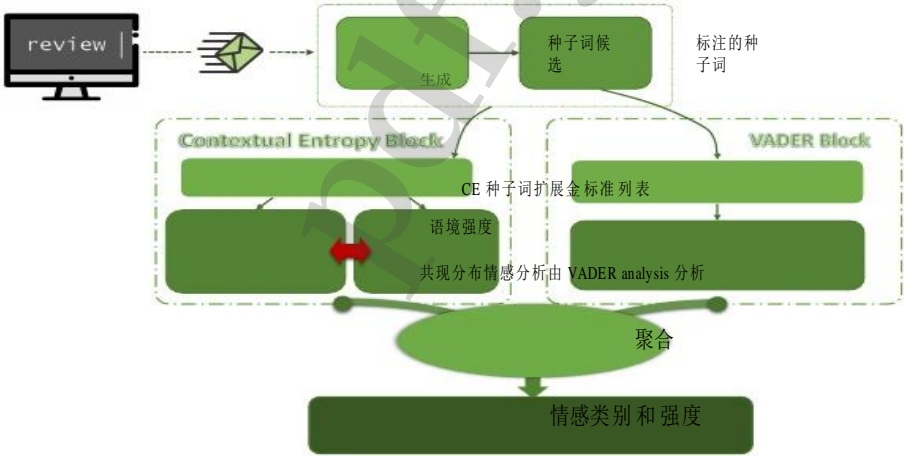


图 2:我们提出的 CE-VADER 模型的整体架构。该模型由两个积木，即 CE 积木和 VADER 积木。



关注数学模型
获取更多资讯

3.2.1 手动标注种子词

我们将 80% 的数据作为训练集，剩余的 20% 作为评估的测试集。来自训练集的评论体中的句子被分解为分离的单词，其中统计了它们的频率。高频情感词被挑选出来作为种子词，由我们人工标注，而低频情感词则被丢弃。标注者(我们小组成员之一)将结合他的专业知识自然语言处理知识，将所有选定的情感词分类为五组，即。，强正面，弱正面，中度，弱负面和强负面。我们没有粗糙地对“正”或“负”[6]进行两组注释，而是将每一组详细地细分为“强”和“弱”两个子组，并多留出一个标记为“中度”的组。五组注释策略旨在与 1 - 5 星评级分数相关联，例如，“弱积极”映射到四星评级。我们将种子词的五组表示为 G_i ，其中 $i = 1, 2, 3, 4, 5$ 。

从带有“正”或“负”标签的训练集生成的代表词分别用冷色调或暖色调描绘在图 3 中。标注的五类种子词附在附录 A 中。



图 3:部分代表性种子词演示。标注为“积极”的词用浅色调着色,而标注为“消极”的词用深色调。字号越大,词频越高。

3.2.2 上下文熵块(CE)

上下文熵块采用部分上下文熵模型[6]作为主干架构。部分上下文熵模型可以同时考虑评论上下文中最具代表性单词的候选词与生成的种子词之间的共现强度和上下文分布。

我们使用一个向量来编码评论中单词与其上下文之间的强度。更具体地说，表示 k 的左右上下文 w_k 在 n 个单词的评论语境中， $R = w_1 w_2 \cdots w_k w_{k+1} \cdots w_n \{w_1 w_2 \cdots w_{k-1}\}$ 和 $\{w_{k+1} w_{k+2} \cdots w_n\}$ 分别。注意，考虑到在线评论的简短上下文风格，我们将 `review` 设置为一个完整的目标，而不是像 `ref.[6]` 中那样将其分解为句子。向量的维度取决于评论的长度，即 n 。表示记录单词左上上下文的向量

wk 为 $vleft(wk) = (vwleftk1, vleftwk2, \dots, vwleftkN)$ 和记录右侧上下文的向量



关注数学模型
获取更多资讯

($v_{w_1}^{right}, v_{w_2}^{right}, \dots, v_{w_N}^{right}$), 其中下标 i 代表的是 I^{th} 向量的分量。 v 的值 v_{wi} 表示 w 之间的共现强度 k 和 w_i 。为了计算我们即将讨论验证的可能性距离, 所有向量都必须以概率表示的形式存在。 $N \leq N$ 是上下文向量的维度, 它统计了评论 R 的上下文中不同单词的数量。

权重 v_{wi} 公式为:

$$v_{wi} = \sum_j e^{-\frac{|j-k|^2}{2\sigma}} \tag{2}$$

它同时考虑了空间因素和距离因素。这里我们采用高斯函数来测量距离, 因为我们持有越近的假设, 它获得的权重越大。然后, 为了使向量变成概率表示, 我们通过以下方式对其进行归一化:

$$v_{wi} \leftarrow \frac{v_{wi}}{\sum_i v_{wi}} \tag{3}$$

然后将候选词或种子词转换为概率向量表示。然后我们将采用 Kullback-Leibler (KL) 距离[8](记为 $KL(\cdot || \cdot)$)来测量向量 $v(c_i) = (v(c_i)^{left}, v(c_i)^{right})$ 我的 i^{th} 候选词和

$v(seed_j) = (v(seed_j)^{left}, v(seed_j)^{right})$ of the j^{th} seed word, that is:

$$D(c_i || seed_j) := KL(v(c_i) || v(seed_j)) = \sum_{k=1}^N P(d_k | c_i) \log \frac{P(d_k | c_i)}{P(d_k | seed_j)} \tag{4}$$

在 $P(d_k | c_i)$ 和 $P(d_k | seed_j)$ 表示 k 的概率权重 i 的左(或右)上下文向量的维度和种子 j 分别, N 是上下文概率向量的维度。由于 KL-距离的非对称性, 即 $D(c_i || seed_j) \neq D(seed_j || c_i)$, 我们采用如下对称距离(记为 $SD(\cdot, \cdot)$):

$$SD(seed_j, c_i) = SD(c_i, seed_j) := D(c_i || seed_j) + D(seed_j || c_i) \tag{5}$$

$D(c_i || seed_j)$ 和 $D(seed_j || c_i)$ 不仅使 $SD(\cdot, \cdot)$ 的对称性, 而且说明了它们的左右上下文分布。然后我们将定义候选词 c 之间的相似度(记为 $SI(\cdot, \cdot)$) _{i} 和种子词 $seed_j$ 基于对称距离由:

$$SI(c_i, seed_j) := \frac{1}{1 + SD(c_i, seed_j)} \tag{6}$$

$SI(c_i, seed_j) \in (0,1)$ 衡量两个词之间的相似度, 两个词的相似度越大。特别是, $SI(c_i, seed_j) = 1$ 当且仅当 $c_i = seed_j$ 。在计算完候选词和种子词的相似度后, 我们将提出候选词 c 之间的相似度 i 以及五组之一的 G_j 由:

$$SI(c_i, G_j) := \frac{1}{|G_j|} \sum_{seed_k \in G_j} SI(c_i, seed_k) \tag{7}$$

$SI(c_i, G_j) \in (0,1)$ 度量候选词 c 的相似度 i 对 j 的相似性 i 群的煽情种子词。评论 R 与 G 的相似性 j 定义为:

$$\mathbb{S}\mathbb{I}(R, \mathbf{G}_j) := \frac{1}{N} \sum_{i=1}^N \mathbb{S}\mathbb{I}(c_i, \mathbf{G}_j)$$



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

$SI(R, G_j) \in (0, 1]$ 度量评论 R 与 j 的相似性th 群中的情感种子词，越大相似度越高。CE 块的输出是一个 5- 维的概率向量 $CE(R) = (CE_R^1, \dots, ce_R^5)^T \in R^5$ 。 j^{th} 组件 CE_R^j 强度是否与 G 族有关 j ，其定义为：

$$CE_R^j = \frac{SI(R, G_j)}{\sum_{i=1}^5 SI(R, G_i)} \tag{9}$$

3.2.3 维德块

VADER[7]是一个简单的基于规则的模型，用于一般的情感分析，特别是针对社交媒体的文本风格。基于人工生成的黄金标准词汇特征列表[7]，VADER 不需要任何训练数据，这显示了其扩展到广泛的情感分析任务的潜力。因此，在这篇文献中，我们基于我们生成的种子词扩展了 good- standard 的词汇特征列表，以解决评论上下文的情感分析问题。此外，我们还将原来的 4 个组预测结果扩展为 5 个组预测结果，即。、强阳性、弱阳性、中度、弱阴性和强阴性在我们 CE 块的一致性上。读者可参考 ref.[7]了解更多关于维德架构的细节。扩展的黄金标准列表附在附录 a 中，维德块的结果也是一个带有 j 的 5- 维概率向量th 分量是 j 对应的强度th 种子词组 G_j 。表示评论 R 的结果向量为 $VADER(R) = (VADER_R^1, \dots, vader_R^5)^T \in R^5$ 。

3.2.4 提出 CE-VADER 用于情感分析

我们提出的 CE-VADER 混合模型由两个块组成:CE 块和 VADER 块。这两个块都将为评论上下文提供一个 5- 维的概率向量 $CE(R)$ 和 $VADER(R)$ ，其中每个分量 R 是其对应组 G 的强度 j 。我们将采用两个向量的平滑凸线性组合作为我们 CE-VADER 模型的最终强度概率向量(表示为 $INT(\cdot)$)，即：

$$INT(R) := softmax(\lambda CE(R) + (1 - \lambda)VADER(R)) \tag{10}$$

其中 λ 是融合系数，在本文献中我们设置 $\lambda = 0.5$ 来同等权衡两个块的结果。我们使用 $softmax(\cdot)$ 作为我们的平滑函数，它由：

$$softmax(x_1, x_2, \dots, x_s) = \frac{1}{\sum_{i=1}^s e^{x_i}} (e^{x_1}, e^{x_2}, \dots, e^{x_s}) \tag{11}$$

正如我们的实证结果显示，经过平滑后，强度向量会与星级表现出很强的一致性。

提出的 CE-VADER 在评论 R 上的情感分类结果为 G 的组名 j 与分量 $INT(R)$ 相关联的 $= (INT_R^1, \dots, int_R^5)$ 值最大，用 j 表示 $j_0 = \operatorname{argmax}_j \{int_j R\}$ ，与其对应的强度 $int_{j_0} R$ 。



3.3 结合基于文本和基于评分的措施

唯一的评论 id 记为 id 。评论标题 Rh_{id} ，评审机构 Rb_{id} ，星评 s_{id} ，回顾日期 rd_{id} ，有益的投票 hv_{id} ，产品标题 pt_{id} ，产品 id $p_{id} \in \{B, M, H\}$ (B, M, H 分别代表婴儿奶嘴、微波炉、电吹风)都与下标评论 id 相关联。由于产品标题、产品家长和产品 id 之间存在很强的相关关系(差异发生租期总和小于 0.01%)，即一旦给出其中一个我们几乎可以唯一地告诉另外两个，我们在本文献中只使用产品标题来描绘产品。我们没有考虑市场因素，因为他们都来自美国。表示配对 $P(id) = (R_{id}, 年代_{id}, 高压_{id}, id, pt_{id}, p_{id})$ ，其中 $R_{id} = (Rh_{id}, Rb_{id})$ 是包括标题和正文在内的整个评论文本。

我们提出了每个评论的重要性率(记为 IMP)，通过取其有益的投票，星级评价指标 $VEC(s_{id})$ 之间的对应关系 id 和基于文本的度量 $INT(R_{id})$ ，审查文本清晰度。由以下公式定义。

$$IMP(id) := (1 + hv_{id}) \cdot exp[-\alpha(1 - \frac{INT(R_{id}) \cdot VEC(s_{id})}{||INT(R_{id})|| ||VEC(s_{id})||})] \cdot exp[\beta(\sum_{i=1}^5 int_R^i log(int_R^i))] \tag{12}$$

其中 $1 - \frac{INT(R_{id}) \cdot VEC(s_{id})}{||INT(R_{id})|| ||VEC(s_{id})||}$ 是文本的度量和评级 id 和 $VEC(s_{id})$ 之间的余弦距离是 k 吗
基于测度，计算两者对应关系的保真度， $\alpha > 0$ 是其关联权重。 $-\sum_{i=1}^5 int_R^i log(int_R^i)$ 是文本上下文的熵，越低清晰度越高， $\beta > 0$
为关联权重。重要性越高，说明评论文本和评分的信息量越大。

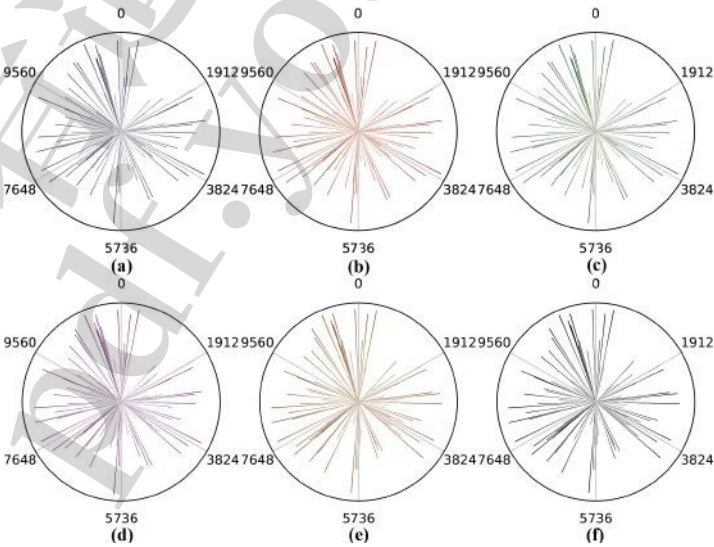


图 4:在 “Hair dryer” 数据集上实现了我们提出的模型。我们展示了由不同参数(α , β)产生的前 1%信息量最大的评论和关联评级的分布。线性越长，我们的模型表明的信息量越大。(α , β) 的设置为:a)(1,1);b)(1、3);c)(1、5);d) (2, 1);e) (4,1);f)(5, 5)。如图所示，我们的模型在两个参数上表现出了鲁棒性。



3.4 模型实现、敏感性分析及结果

我们针对给定的数据实现了我们提出的模型。我们设置参数 $\sigma_0 = 1$ 在等式(1)中。3.3 节提出的模型包含两个参数 α , β 。我们首先分析了这两个参数的敏感性。

如图 4 所示，我们在“电吹风”数据集上实现了我们提出的模型，我们的模型显示的前 1% 信息量最大的评论在两个参数 α 和 β 上显示了很强的鲁棒性。我们还利用 DTW 相似度来量化我们的模型在两个参数上的鲁棒性，如图 5 所示。DTW 相似度越小，说明两个排名越相似。关于 DTW 相似度的更多细节，读者可以参考 ref.[9]。我们设置 $\alpha = 1$, $\beta = 1$ 为基线，计算 DTW 相似度，最大值为 14.3，在评论数量为 11470 条的情况下，这是一个很小的值。再次，我们的模型对 α 和 β 显示出了很强的鲁棒性。

前 1%信息量最大的评论和根据它们对三种产品的排名列出的相关评级附在附录 C。

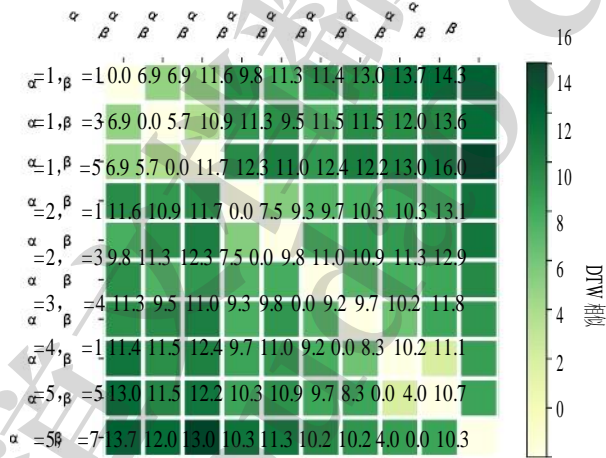


图 5:我们的模型在不同 α 和 β 值的“吹风机”数据集上实现的 DTW 相似度。我们设置 $\alpha = 1$, $\beta = 1$ 作为基线。最大相似度是 14.3，考虑到我们的总评论 11470 条的情况，这个值很小。这显示了我们模型的鲁棒性。

4 个差分方程来测量基于时间的模式

在本节中，我们构建了一个基于差分方程的模型来描述产品声誉的变化。本节的其余部分安排如下。在 4.1 节中，我们详细制定了我们的模型。在 4.2 节中，我们分析了模型敏感度，并在三种产品上实现了模型。

4.1 基于差分方程的模型

我们提出“声誉”来制定产品 P 在时间 t 附近的声誉。将声誉表示为 REP.然而，我们假设产品的声誉会逐渐改变



关注数学模型
获取更多资讯

通过买家评论文本和星级评分。我们将采用差分方程来表示声誉。声誉的差异(即。 , 声誉的增长率)可以表述为:

$$\Delta \text{REP}_{t_0, \theta}(T, P) := \frac{1}{2z} \left(\sum_{T-t_0 \leq rd_{id} \leq T, p_{id}=P} \text{IMP}(id) \cdot (\theta s_{id} + (1-\theta) \text{argmax}_j \{ \text{int}_{R_{id}}^j \} - 3) \right) \quad (13)$$

其中 Z 为归一化常数定义为:

$$Z = \sum_{T-t_0 \leq rd_{id} \leq T, p_{id}=P} \text{IMP}(id) \quad (14)$$

为简单起见, 我们对 ΔREP 的值进行归一化 $\text{REP}_{t_0, \theta}(T, P)$ 到 $[-1, 1]$, 负值与负面感觉或一星和二星评级相关联, 而正值与正面感觉或四星或五星评级相关联。

参数 $\theta_0 \in [0, 1]$ 是我们在 3.2 节中提出的 CE-VADER 模型的星级评级和基于文本的度量的权重系数。我们假设运输产品需要一段时间, 并且一些客户更喜欢在购买后的一段时间后进行评论。因此, 我们认为所有 id 都满足 $T - T_0 \leq rd_{id} \leq T$ 为时间标签, 其中 T_0 为阈值。注意, 买家几乎有 90 天的时间留下反馈, 因此 $t_0 \leq 90$ 。在这篇文献中, 我们设置 $\theta = 0.5$ 为星级和评论上下文的同等权重。并且我们设置 $t_0 = 10$ 是这个文献。那么, 声誉可以用下面的差分方程来表示:

$$\text{REP}_{t_0, \theta}(T, P) - \text{REP}_{t_0, \theta}(T-1, P) = \Delta \text{REP}_{t_0, \theta}(T, P) - \text{PEN}(T, P) \quad (15)$$

其中 $\text{PEN}(T, P)$ 是惩罚因子, 制定了低星评级或负面评论在多大程度上破坏了声誉, 因为我们假设买家更关注那些负面评论。它的表述如下。

$$\text{PEN}(T, P) = k_2 \times \text{sigmoid}(k_1 \times \text{REP}_{t_0, \theta}(T-1, P)) \times \frac{\#\{id | \theta s_{id} + (1-\theta) \text{argmax}_j \{ \text{int}_{R_{id}}^j \} \leq k_3\}}{\#\{id | T-t_0 \leq rd_{id} \leq T, p_{id}=P\}} \quad (16)$$

$k_1 k_2 k_3$ 为惩罚因子的阈值。我们设定 $k_3 = 2$ 作为要量化的阈值

的“差评”。因此, $\frac{\#\{id | \theta s_{id} + (1-\theta) \text{argmax}_j \{ \text{int}_{R_{id}}^j \} \leq k_3\}}{\#\{id | T-t_0 \leq rd_{id} \leq T, p_{id}=P\}}$ 传播的是“负的

评论”。 sigmoid 由 $\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \in [0, 1]$ 定义。我们设置 sigmoid 项来亲密化一种社会行为, 即一旦一件产品赢得了良好的口碑, 人们就会更加重视“差评”。 k_2 值惩罚因子与增长率的比值, 即。 Δ 代表。

4.2 模型实现、敏感性分析及结果

作为 4.1 中我们模型的构建, 有两个可调参数, 即。 $k_1 k_2$ 。我们在三类产品的数据上实现了我们提出的模型。并展示我们的模型在这两个参数上的敏感性。

图 6 描绘了参数设置为 k 时的声誉曲线 $\theta = 0.5$ 和 $k_2 = 20$ 。

如图所示，“微波”为负值信誉率拥有不好的声誉;而“安抚者”则赢得了良好的口碑。没有一款产品有稳定或单调的增长或 a



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

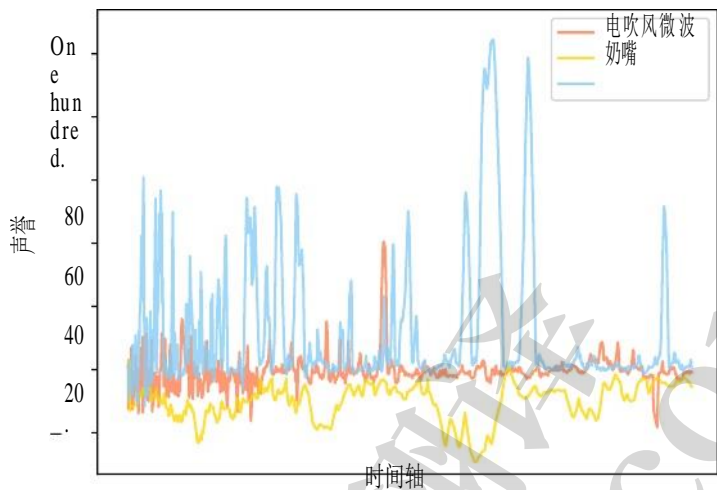


图 6:带 k 的产品声誉曲线 $k_1 = 0.5$ 和 $k_2 = 20$ 。

信誉率下降。然后，我们分析了我们的模型对两个参数的敏感性，即。 k_1 和 k_2 。我们的模型在 k 上显示出极强的鲁棒性 k_1 以及对 k 的温和敏感性 k_2 图 7。如图 7-B 所示， k 的不同设置 k_2 会导致“安抚者”数据上的信誉率增长率不同，但趋势相同。但是， k 的小值 k_2 会导致指数级的大声誉增长率。我们之前讨论过， k 的小设定 k_2 表明对负面评论的关注较少。

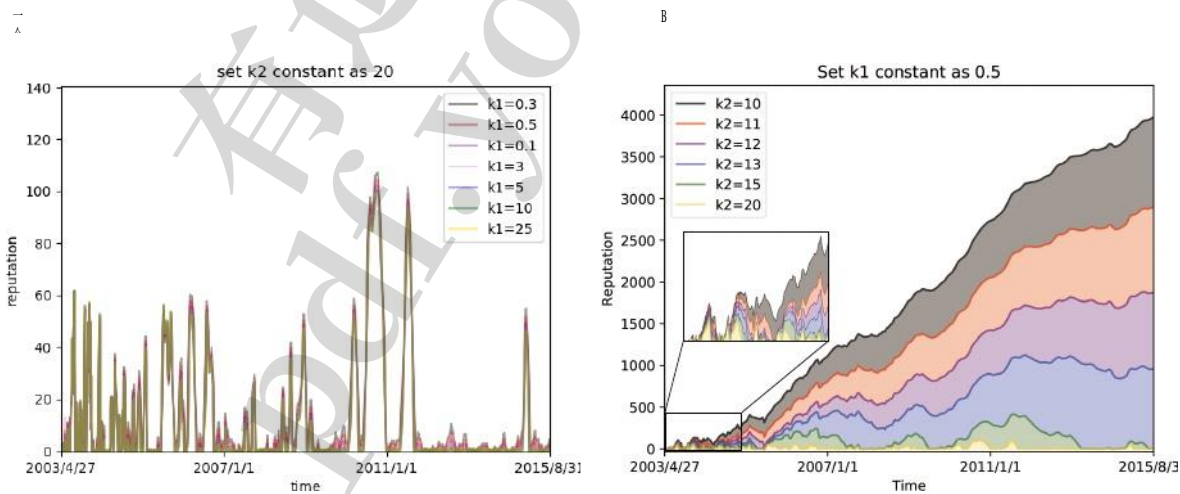


图 7:参数 k 的灵敏度分析 k_1 和 k_2 。A)我们的模型在 k 上显示出极端强的鲁棒性 k_1 。B)我们的模型显示了对 k 的敏感性 k_2 。



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

预测可能的成功或失败

在最后一节中，我们构建了一个映射 $REP_{t_0} : [T_0 T_1] \times \{b, m, h\} \rightarrow r_+$ 在 $[T_0 T_1]$ 为给定审核日期的时间范围。在本节中，我们将采用时间序列预测方法来预测 REP 的值 $_{t_0}$ 在扩展的未来域上，即 $[T_1 T_2 \times \{B, M, H\}]$ 为一些 $T_2 > T_1$ 在 5.1 节中。基于预测的 REP_{t_0} 在未来的时间域，我们将在 5.2 中评估每个产品的成功或失败潜力。我们在 5.3 节展示了真实数据上的结果，并预测了它们的潜在成功或失败可能性。

5.1 预测未来声誉的时间序列

我们采用自回归模型(AR)作为时间序列预测，以预测信誉率未来的增长或下降。AR 模型可以用下面的差分方程来描述。

$$REP_{t_0, \theta}(T, P) = a_0 + \sum_{k=1}^p a_k REP_{t_0, \theta}(T - k, P) + \varepsilon_T \tag{17}$$

在 ε_T 就是白噪音。AR 模型去搜索系数(a_0 , 一个 $_1, \dots, a_p$)在等式(17)中拟合给定的时域数据 $[T_0 T_1]$ 具有最小的均方根误差(RMSE)。

5.2 评估成功或失败的潜力

我们从 T 定义了产品 P 的平均声誉 $_0 T_1$ 为:

$$\overline{REP_{t_0, \theta}^{T_0 T_1}(P)} := \frac{1}{T_1 - T_0} \int_{T_0}^{T_1} REP_{t_0, \theta}(T, P) dT \tag{18}$$

同样，我们可以定义产品 P 在未来时域的预测平均声誉 $[T_1, \quad t]$ 为 $t < t_2$ 基于 5.1 节构建的模型如下。

$$\overline{REP_{t_0, \theta}^{T_1 t}(P)} := \frac{1}{t - T_1} \int_{T_1}^t REP_{t_0, \theta}(T, P) dT \tag{19}$$

由给定数据计算得到的平均声誉与 Eq.(18)和 Eq.(18)预测的平均声誉之比定义为声誉变化率。定义为:

$$\gamma(t, P) := \overline{REP_{t_0, \theta}^{T_1 t}(P)} / \overline{REP_{t_0, \theta}^{T_0 T_1}(P)} \tag{20}$$

我们提出了一个预测产品 p 成功或失败的模糊评价系统，系统的整体架构如图 8 所示。我们将结果分为四类，即“强成功”、“弱成功”、“弱失败”、“强失败”。

5.3 模型实施与结果

我们将我们的模型实现到给定的数据中，在等式(17)中 $p = 50$ 。我们在给定的时间域 $([T_0, T_1]) \subset [T_0, T_1 + 14 \cdot t_0]$ 拟合回归系数 (a_0, a_1, \dots, a_p) 。



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

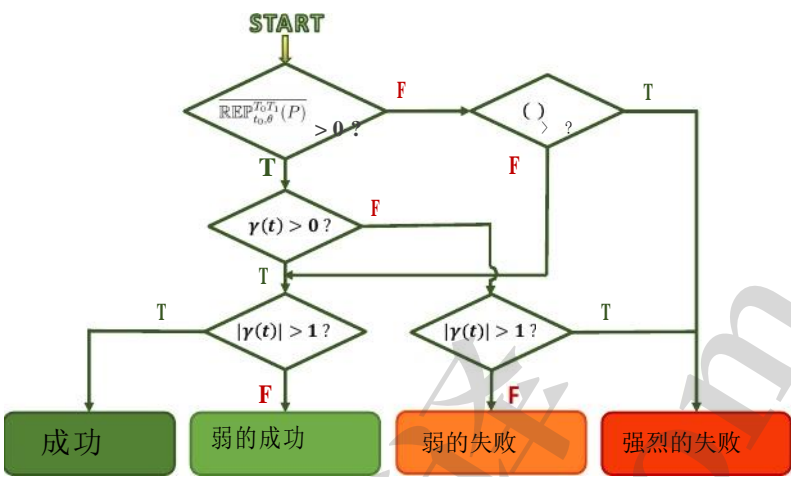


图 8:我们提出的用于预测每个产品潜在成功或失败的模糊评估系统的整体架构。

后 14 种方法用于评估 AR 模型。，在时域上 $[34 T_1 + 14 t_0 T_1]$ 。在评价领域，我们发现该模型能够很好地拟合和预测“声誉”率，最大 RMSE 率为 0.031。我们评估了域 $[T_1 T_2]$ 用 $T_2 - T_1 = 2^l (T_1 - T_0)$ 。安抚奶嘴、微波、电吹风的预测信誉率分别如图 9、10、11 所示。根据 5.2 节提出的框架，它们被预测为“弱成功”、“强失败”和“强失败”。

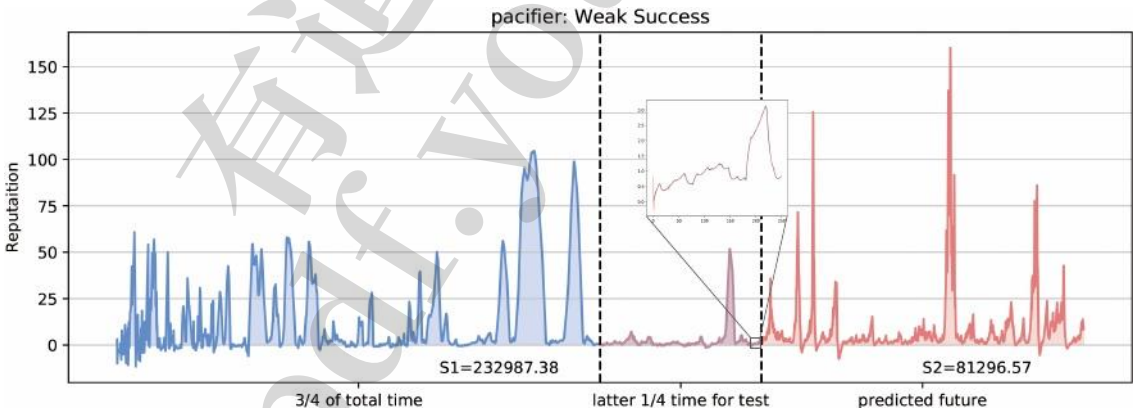


图 9:模型在“安抚者”数据上的实现。预测为弱成功。

6 具体评级和描述分析

在本节中，我们在 6.1 节和 6.2 节中重复分析具体评分和描述符。



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

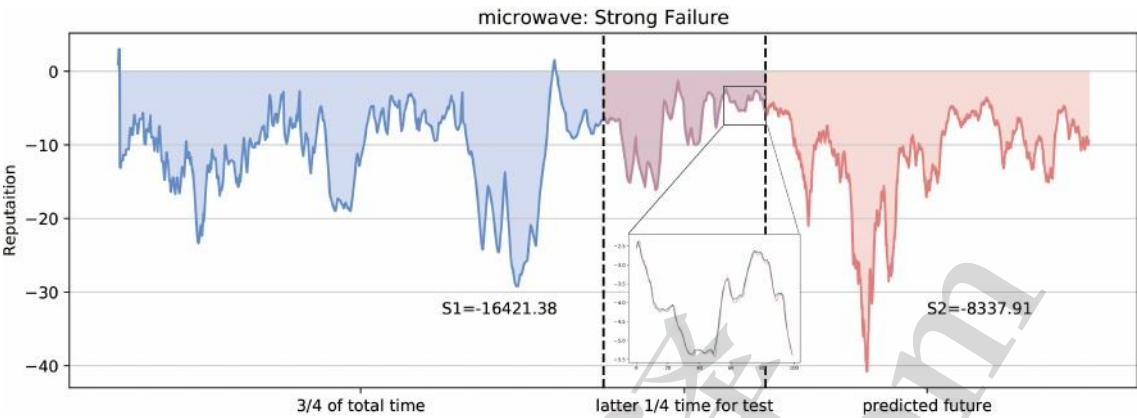


图 10:模型在“微波”数据上的实现。被预测是一个强烈的失败。

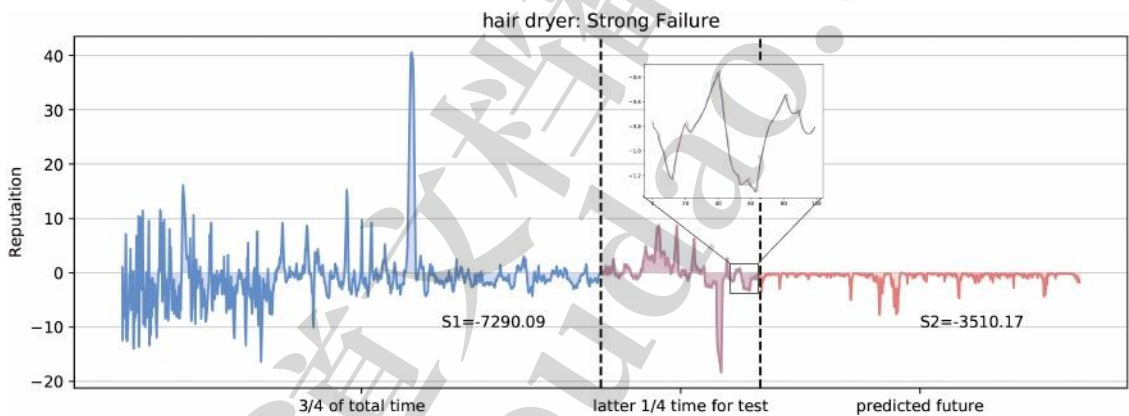


图 11:模型在“吹风机”数据上的实现。被预测为强势失败。

6.1 具体星级评级与评级频率的相关性

在本节中，我们首先分析了特定星级评级与声誉率之间的关系，即。，一星评级的强负面案例和五星评级的强正面案例。我们使用我们提出的声誉率作为基于文本的度量。如图 12 所示，我们可以观察到，五星评级比例的增长导致了“声誉”率的即时提高。而在五星评级比例达到峰值后，“信誉率”立即下降。信誉率的增长，也与一星评级比例的下降有着强烈的关联。

然后，我们评估了星级评分与图 13 中有帮助的投票之间的相关性。再一次，在所有类型的数据中，五星评级比其他四种评级显示了与有用投票的强得多的相关性。婴儿奶嘴，微波炉和吹风机。

而且，在之前的工作[10]中也有研究指出，人们认为所有网络评论(包括正面和负面)中的极端案例是更有用的信息，而不是我们这一期中被分类为“强正面”或“强负面”的评论文本和一星或五星的评价。

因此，在这篇文献中，我们想强调极端评分的影响



关注数学模型
获取更多资讯

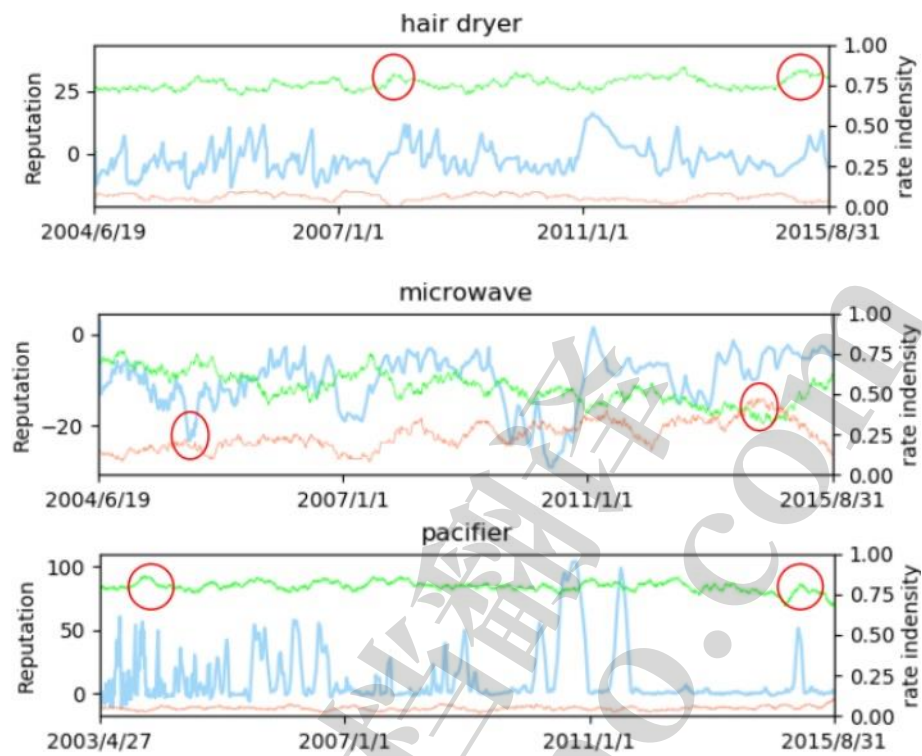


图 12: “口碑”率与一星或五星评分比例的相关性。蓝色线表示声誉时间序列，绿色线表示五星评分比例，橙色线表示一星评分比例。红圈突出的典型例子是，较高的五星评分比例导致信誉率瞬间上升，而一星评分领先下降。

即。，一星和五星的评分，而不是 2 星到 4 星的评分，基于上述三个事实。

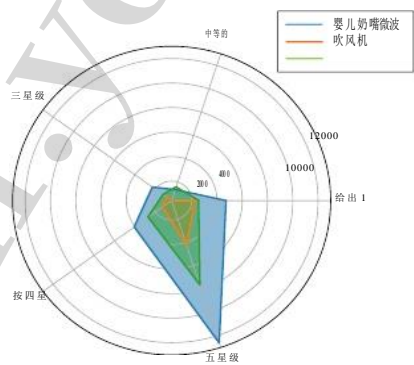


图 13:星级评分与有益投票之间的相关性。与五星评分相关的评论总是被认为是有帮助的。

通过验证评价为 1 星或 5 星产品相关的评论文本词频模式，我们发现消费者不太可能购买与低星评分相关的产品，而更乐于购买评级为高星的产品。我们举了一个具体的案例
与一星或五星评级相关的 x 个评论的连续出现



关注数学模型
获取更多资讯

以人们为基准，电吹风、安抚奶嘴 $x = 10$ ，微波炉 $x = 5$ 。由于五星评分的百分比较低，以及它拥有的坏名声(如我们之前在 4.2 节中讨论过的)微波相比于其他两种情况，我们专门为它设置了 $x = 5$ 的五星评分比例较低。我们的实证结果表明， x 越小，频率越高，而 x 越大，频率越低。为了简单起见，我们如上所述设置了一个适中的 x 。

如表 2 所示，在声誉良好和声誉不好的两种情况下(声誉由我们的模型在第 4 节中表示)，连续出现五星评分后，会引发更多的评论。这表明人们在连续出现高星级评分后，会更愿意对产品进行评分。而在连续发生一星评分后，人们对声誉较差的产品进行评分的意愿较低。、微波炉和电吹风。看到连续出现一星评级，人们强烈被鼓动对口碑好的产品进行评级。

表 2:评分频率与一系列极端评分案例的对比，整体评分频率与婴儿奶嘴、微波炉、电吹风数据的平均口碑。评分频率越小，说明买家对产品的评分频率越高。五星评分的连续出现，总是煽动着人们去评分。声誉差的产品连续出现一星评分，在强烈煽动人们为声誉好的产品投票的同时，导致评分频率大大降低。

产品	额定频率 经过一系列的 五星级评级(天)	额定频率 经过一系列的 一星评级(天)	整体 频率(天)	声誉
婴儿奶嘴	1.9444	1.56000	2.37828	好
微波	8.66667	30.61904	24.39975	坏
吹风机	1.58207	16.37500	4.29293	坏

图 14 表示了从 2012 年 8 月 31 日到 2015 年 8 月 31 日吹风机上评分总次数发生显著变化时的月度星级评分总次数以及一星、五星评分次数。我们可以观察到，五星评分的集中总是与评分总数的增加相关联，而一星评分的连续出现则导致评分数量的下降。

6.2 具体质量描述符与评级水平的相关性

6.2.1 用于评估的朴素贝叶斯模型

一些词通常与情感感觉强烈相关，导致不同的评级水平，例如，“热情”与积极和高兴的感觉可能导致更高的星级评级和“失望”与消极的感觉。在这一节中，我们采用 Naïve 贝叶斯模型来识别具体的单词如何与评级水平相关联。具体单词 w 的星评 s 的 Naïve 贝叶斯分布概率描述如下：

$$P(s|w) = \frac{P(w|s)P(s)}{P(w)}$$

(21)

通过评估概率 $P(s|w)$ ，我们可以衡量特定单词 w 与星率 s 之间的相关性。



关注数学模型
获取更多资讯

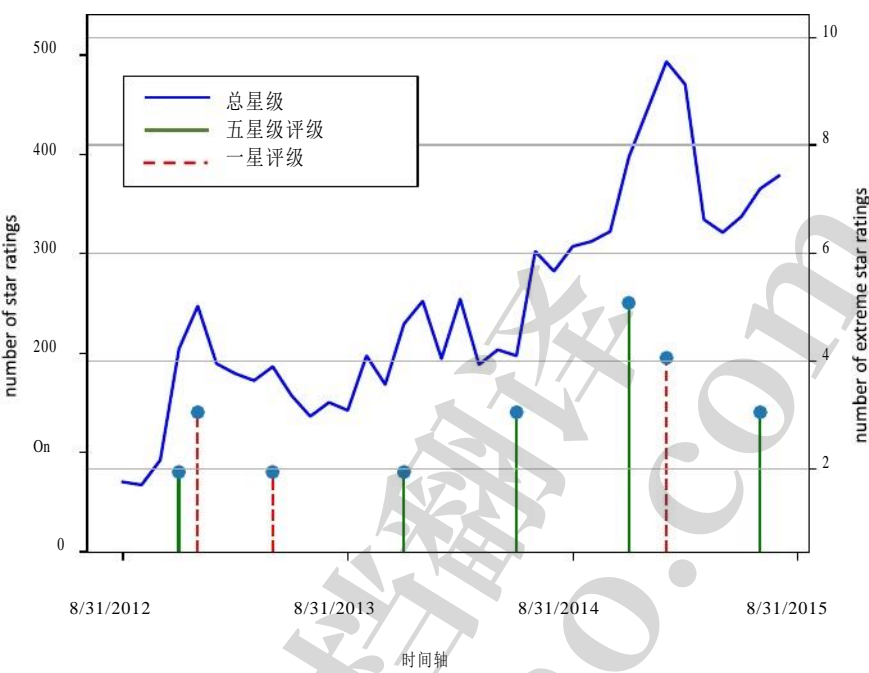


图 14:2012 年 8 月 31 日至 2015 年 8 月 31 日，电吹风数据一星或五星评分与整体星级评分之间的相关性的代表性例子。

6.2.2 模型实施与结果

$P(s|w)$ 越高，单词 s 和星级评分 s 的相关性越强。在提供的海量数据下，我们使用频率来近似 Eq.(21)中的概率，即 $P(s) = \frac{\#(id|sid=s)}{\#(id)}$ 。我们展示了与五星和一星相关的十大词汇

分别在表 3 中。

表 3:分别与五星评分和一星评分相关的十大词汇。

话说强烈 与 五星级评级	$P(s = 5 w)$	话说强烈 与 1 星评价	$P(s = 1 w)$
容易	0.8518	或	0.1416
爆米花	0.8333	出	0.1434
爱	0.7809	不	0.1483
功能	0.7778	没有	0.1513
按钮	0.7692	回来	0.1546
伟大的	0.7685	热	0.1718

多	0.7452	因为	0.1759
高	0.7250	个月	0.1940
可以	0.7200	从	0.2083
也	0.7196	干燥机	0.2125



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

7 设计特征的吸引力分析

在本节中，我们从前 1%信息量最大的评论中提取描述婴儿奶嘴、微波炉和吹风机性能、外观等的关键词。通过计算这些关键词在所有评论中的频率，并分析前 1%信息量最大评论的内容，我们告知了一些对消费者有吸引力的设计特征，这些关键词组和关键词出现次数在附录 B 中。

- 1.婴儿安抚奶嘴:根据表 4，婴儿安抚奶嘴的尺寸、外观、便利性、安全性是消费者最关心的。人们更关心的是它对宝宝是否合适和安全，所以尺寸和安全材料的各种替代是必要的。此外，可爱的图案也很受宝宝们的欢迎。
- 2.微波:从表 5 可以看出，消费者最关心的是微波炉的外观、价格、组成，可见人们比较看重的是微波炉的性价比。很多评论对多功能但价格昂贵的微波炉表示欢迎。当产品价格不菲的时候，人们更喜欢获得尽可能多的功能，比如旋转烧烤架、精确的计时器、多种使用模式等。
- 3.电吹风:根据表 6，消费者最关心的是电吹风的功率和外观，说明消费者更喜欢功率大的便携式电吹风。同时，安全性和工作量影响用户体验，对于阳光公司来说，在增加功率和安全安静的操作环境之间保持平衡至关重要。特别是一些特殊的设计可能会刺激消费者购买，例如折叠手柄可以提高吹风机的便携性。

8 销售策略和建议

在本节中，我们根据前几节的结果，针对婴儿奶嘴、微波炉和电吹风这三种精准产品，为阳光公司提出有信心的网络营销策略和建议。包括(给阳光公司的)具体理由在内，我们最自信的建议如下：

- (一般推荐)帮助度高且与强烈感受相关的评论通常是信息量大的。实时收集这些评论和星级评价。像关注星级评级一样关注文本评论。
- (一般推荐)始终保持高比例的五星评分，因为它与产品的声誉有很强的关联。
- (一般推荐)更强调负面评论或与低星评分相关的评论，如果产品声誉不好，例如低星评分比例高。并立即做出反馈和调整，因为这些评论或评级会继续破坏产品的声誉。大多数给一星评级的人表示，一旦他们听到公司负责人的意见，他们就会更新我的评论。大约 30%的一星评级与保修服务不及时有关或者是公司的反馈。



关注数学模型
获取更多资讯

- (吹风机的具体理由)吹风机的条件似乎很差。在我们评估的前 100 个信息量最大的评论中，有 13 个评论报告了一个吸烟或抛出火花或抛锚的案例，这可能会导致一些危险。由于电吹风的美誉度下降，我们建议大家在其美誉度上升之前，不要将电吹风投放市场，这有利于产品的顺利启动。
- (微波炉的具体理由)微波炉在这三种产品中口碑最差。它有最高比例的一星评级和强烈的负面评论。在信息量最大的前 100 条评论中，大家普遍吐槽的是条件差，有着火的潜在危险。我们的模型暗示了它潜在的失败。因此，我们建议等到微波的声誉恢复并保持稳定。
- (安抚奶嘴的具体理由)安抚奶嘴口碑好，好评比例最高，五星评分最高。我们的模型也预测它会成功，所以把它投放到市场的时机很好。然而，在信息量最大的前 1% 的评论中，他们报道了漏水的案例。企业应注意连接处或接头部分。
- (一般建议)奶嘴、微波炉、电吹风作为低成本产品，同时对消费者的专业要求较低。产品信息越完整，因买卖双方信息不平等而造成的损失就越小。当然，信息的真实性是需要保证的。
- (一般推荐)我们的研究结果表明，一系列的五星评分可以煽动消费者评论。因此，我们建议您在产品五星评分较多的情况下加大促销力度，形成正面反馈。
- (一般建议)创意和吸引人的广告将增加产品的销售。在线市场使消费者更容易浏览商品的多种选择，这导致了公司之间的激烈竞争。设计有趣和醒目的广告可以有效地增强产品的竞争力，所以阳光公司应该在广告上进行适当的投资，以获得更多的利润。信息量最大的前 1% 评论表明，一些消费者对公司网站非常关注。

9 优点和缺点

9.1 的优势

- 1.新鲜事物。据我们所知，我们首次提出了一个 CE-VADER 混合模型，用于在线市场上基于文本的情感评估。
- 2.准确性。50 阶 AR 模型在验证时域上的最大 RMSE 为 0.031。基于文本的度量与基于评分的度量具有良好的相关性。



关注数学模型
获取更多资讯

- 3.泛化。我们提出的框架可以自由地应用于任何数据集，例如，任何在线平台的任何产品的评论和星级评级。
- 4.鲁棒性。我们的模型对大多数参数表现出了极大的鲁棒性。

9.2 缺点

- 1.耗时的人工标注。手工标注 CE-VADER 模型从评论中生成的种子词非常耗时。
- 2.CE-VADER 模型不能用特殊的变异规则识别同一词的不同形式。CE-VADER 模型不能识别动词的过去式、名词的复数形式和具有特殊变化的形容词的比较级形式，如 is(对 was 的表示)、children(对 child 的表示)和 better(对 good 的表示)。
- 3.遗漏了其他潜在的相关因素。在分析具体的评分和描述时，我们没有考虑类似亚马逊的促销营销策略。

10 的结论

为了破解亚马逊评分和评论的秘密，我们提出了一系列新颖的模型来解决从选择信息量最大的评论到识别评论质量描述符的子问题。所提出的模型实现了较高的准确性和鲁棒性。

- 1.信息评价模型可以将基于文本的度量与基于评分的度量相结合，其中我们提出了一种新颖的 CE-VADER 混合模型用于情感分析，作为基于文本的度量。我们可以用提出的模型对每条评论和评分的信息量进行排序。信息量与有用的投票相关联。更具体地说，帮助投票越多，他们拥有的五星评级越多，评论机构越长，他们越有可能被评价为有价值的评论。不过，请记住，信息熵高的中等评级评论，正面评论和负面评论兼有，用“然而”、“但是”等词标注，也有很大的参考价值。
- 2.我们利用差分方程模型构建“信誉度”来量化婴儿奶嘴、微波炉和吹风机这三种产品的信誉度。婴儿安抚奶嘴的美誉度为正，电吹风的美誉度为弱负，微波的美誉度最差。我们通过改进的 AR 算法，预测了这三种产品未来的信誉度趋势。
- 3.在分析星级评分和特定词的分布时，我们通过使用连续的极端评分和一组特殊词来识别特殊评论描述符。连续极端评分对总销量有明显影响，特殊词的出现能够以高概率判断评论的评分。

11 写给阳光公司市场总监的一封信



尊敬的阳光公司营销总监:

根据您的要求,我们分析了贵公司拟引进销售的婴儿奶嘴、微波炉、电吹风在亚马逊上的竞品评分和评论。我们形成了三个模型,称为信息性评估模型来评估每条评论的信息量,基于差分方程的模型来制定产品声誉的变化,基于时间序列预测的评估模型来预测产品的潜在成功或失败。并得到了一些有意义的结果,这些结果有助于吸引消费者和制定适应性的在线销售策略。

首先,构建信息性评价模型,帮助用户方便地跟踪信息性评论和评分;这个模型可以根据星级评分、评论文本和有帮助的投票来评估每条评论是否有信息量,然后根据信息量对评论进行排名。我们认为,在处理大量评论信息时,这种信息性评价对你的工作至关重要,因为信息性评论往往为设计特征提供更有建设性的输入,因此在声誉分析中更具参考价值。

我们通过筛选前 1% 最具信息量的婴儿奶嘴、微波炉和吹风机的评分和评论来获得消费者的偏好和关注。我们分析得出的流行设计特征如下:

- 婴儿安抚奶嘴:需要各种尺寸和安全材料的替代。此外,可爱图案很受宝宝们的欢迎。
- 微波炉:具有旋转烧烤架、精准计时器、多种使用模式等功能的多功能微波炉深受消费者欢迎。
- 电吹风:消费者更喜欢小功率的电吹风(在安全的前提下)。一些额外的设计,如折叠把手,有利于吸引消费者。

然后,我们建立基于时间的措施来预测产品在网上市场声誉的增加或减少。我们的模型可以准确地预测一个产品在未来很长一段时间内的声誉变化。这种精确的预测对于你在声誉下降之前生成策略是有意义的。

基于声誉预测,我们规定未来声誉增加或减少的程度,作为产品潜在成功或失败的判断。我们的分析表明,总体而言,婴儿奶嘴是潜在成功的产品,而微波炉和吹风机的失败风险更大。

根据我们的分析结果,我们为贵公司制定了合理的销售策略:(1)我们建议贵公司在口碑上升时,将微波炉和吹风机投放市场。而婴儿安抚奶嘴的口碑正在上升,最好现在就投放市场。(2)产品信息越完整,因买卖双方信息不平等而造成的损失就越小。(3)建议您在产品五星评分较多的情况下加大促销力度,形成正面反馈。(4)当你的产品声誉下降时,专注于一星评分和评论。

感谢您在百忙之中抽出时间阅读我的来信。希望我们的建议能对您有所帮助。
的帮助。

MCM 战队# 2002116



关注数学模型
获取更多资讯

参考文献

[1] Peter C Evans 和 Annabelle Gawer。平台企业的崛起:全球调查。2016。

[2] Seyed Pouyan Eslami, Maryam Ghasemaghaei 和 Khaled Hassanein。消费者觉得哪些网上评论最有帮助?多方法调查。《决策支持系统》, 113:32 - 42,2018。

[3]理解在线评论有用性的决定因素:一项元分析调查。《决策支持系统》, 102:1 - 11,2017。

[4]胡雅涵, 陈宽沁。预测酒店评论有用性:评论可见性的影响, 以及酒店星级与评论评分的互动。《国际信息管理杂志》, 36(6,A 部分):929 - 944,2016。

[5] Y. 陈和谢 j.j。线上消费者评论:口碑作为营销传播组合的新元素。《管理科学》, 54(3):477-491,2008。

余良志, 吴正龙, 张培灿, 朱宣寿。利用上下文熵模型对情感词及其强度进行扩充, 用于股市新闻的情感分类。《基于知识的系统》, 41:89 - 97,2013。

[7]克莱顿·J·赫托和埃里克·吉尔伯特。Vader:一种基于简约规则的社交媒体文本情感分析模型。发表于2014年第八届国际AAAI网络日志与社交媒体会议。

[8]所罗门·库尔巴克。信息论与统计学。约翰·赖利和儿子们。公司。纽约, 1959 年。

[9] Ira Assent, Marc Wichterich, Ralph Krieger, Hardy Kremer, Thomas Seidl。用于时间序列数据库中高效相似性搜索的预测 dtw。《计算机科学与技术》, 2(1):826-837,2009。

[10] Sangwon Park, Juan L. Nicolau。线上消费者评论的不对称效应。《旅游研究年鉴》, 50:67 - 83,2015。

附录

附录 A 标注了种子词和频次

我们手动标注了五组种子词, 如下图所示。

- 1.强烈正面:很棒(3006)、推荐(1174)、完美(761)、最佳(656)、最爱(368)、完美(183)、棒极了(125)、优秀(114)。



关注数学模型
获取更多资讯

- 2.弱阳性:喜欢(3624)、爱(2766)、爱(2753)、容易(2154)、好(1424)、好(1286)、可以(1054)、可爱(1002)、喜欢(876)、好(855)、更好(821)、能(738)、适合(731)、右(687)、喜欢(647)、愿望(644)、想要(585)、快乐(577)、更容易(521)、适合(395)、喜欢(311)、高兴(229)、干净(193)、可爱(191)、可爱(191)、有趣(181)、爱(163)、好(157)、好(153)、帮助(152)、高兴(141)、容易(131)、美妙(131)、兴奋(118)、安慰(114)、温暖(114)、爱(103)、可爱!(99), enjoy (96), enjoy (86), cool(85)。
- 3.温和的:宝贝(4998), 她(3852), 她(3678), 他(3537), 能(2913), 他(2658), 使用(2446), 儿子(2239), 女儿(2114), 更多(1867), 他(1184), 她(763), 宝贝(626), 容易(619), 礼物(556), 价值(513), 超级(489), 玩具(448), 想要(437), 好(407), 不会(400), 较小(302), 不会(303), 女孩(293), 附加(293), 精确(282), 给予(251), 可能(218), 扔(213), 重(210), 朋友(205), 更大(189), 品牌(171), 问题(169), 更大(167), 任何人(166), 担心(166), 附加(165),(160),快(159),更喜欢(158),失去了(157), 保健(157),取代(157),(151),孙女(150),(146)问题,快速(142),便宜(142),女儿的(140)、安全(139)、小(136),(136),(131),(129),(128),喜欢(128 年),清洁(123),孔(123),礼物(120),很容易(118),简单(118),(113), 便宜(110),(106),除非原(107),略(105),护理(103),礼物(98),薄(94),侄子(94),保存(94),松(93),调整(92), 更好(92), 脏(87)。
- 4.弱否定:不(1283), 没有(1011), 不能(780), 硬(734), 少(354), 问题(346), 不是(297), 不能(285), 然而(252), 困难(248), 还(244), 不能(211), 然而(175), 不是(173), 问题(149), 不是(143), 然而(132), 击中(126), 反对(124), 更困难(118), 浪费(118), 错误(102), 麻烦(101), 痛苦(90), 虽然(180)。
- 5.强烈否定:但是(7128), 失望(170), 不(90), 没有(113)。

附录 B 不同关键字组中关键字出现的次数

baby 安抚奶嘴的关键词出现次数列在表 4 中，共有 18937 条评论。

表 4:婴儿安抚奶嘴不同关键词组的关键词出现次数

集团	关键字出现的次数
大小	3402
外观	3399
方便	3138
安全	2374
组件	706

在总共 1616 条评论中，电吹风的关键词出现次数列在表 5 中。

电吹风的关键词出现数列在表 6，共 11470 条评论中。



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

表 5:微波不同关键词组的关键词出现次数

集团	关键字出现的次数
外观	2350
组件	2289
价格	2234
设置	1049
安全	617
权力	594

表 6:电吹风不同关键词组的关键词出现次数

集团	关键字出现的次数
权力	6774
外观	4656
特殊的设计	3973
头发质量	2666
安全	1544
工作容积	1542

附录 C 前 1%的信息最丰富的评级和评论

根据我们的模型在“吹风机”数据集上预测的“重要性”率，我们列出了前 1%最具信息量的评论及其相关评级。

- 1.<没有问题>(5星评级):好于预期。很棒的产品，出色的交付。
- 2.<伟大-强力小旅行>(5星评级):我个头小，肩部有问题，这款非常紧凑，重量也很轻。它带着力量装了一拳!
- 3.<好产品，爱它!>(5星评分):我的头发比较卷曲厚实，使用这款产品让它更容易打理。增加光泽和柔软度..不是油性/重..爱死它了。
- 4.<买的好质量>(5星评分):我经常用它，我是一个发型师，用它来帮助我的卷发客户，它从不让我失望。价格好，质量好，买得好。
- 5.<浓密的头发,不用担心。>(5星评分):爱死这款烘干机了，干得又好又快。为我的妈妈买了一份礼物,她绝对喜欢它!她的头发很浓密，吹风机对她来说很好用。
- 6.<太好了!适合我的浓密的头发。>(5星评分):我有一头浓密的及肩长发，通常需要 35 - 40 分钟才能吹干。有了这台吹风机，20 分钟就干了，头发看起来顺滑柔软。我本来对整件事持怀疑态度，但它让我的头发发生了很大的变化。可拆卸的
滤镜是决定因素。我把干衣机扔了，就因为我洗不干净。优秀的产品!

7.<我爱这台烘干机!!!!>(5星评分):作为一名专业发型师我已经厌倦了昂贵的吹风机，吹风机干头发的效果并不比便宜的好。然后，我读到

这台烘干机的评论，我再次想，是啊。所以，我买了它，我必须买



关注数学模型
获取更多资讯

有道文档翻译
pdf.youdao.com

这个吹风机真棒。它一半时间就能把头发吹干，还能让头发光滑有光泽!!!终于找到了一款能兑现承诺的

- 8.<很棒的烘干机，很棒的价格>(5 星评级):很棒的烘干机，效果很好。我需要的时候有很多热量和音量，很好的冷却按钮。能快速吹干头发。我的头发死后，亚马逊在一夜之间把它给了我，所以我只有一天的头发不好。还不错，因为我没时间去实体店买东西。
- 9.<卓越的价值!>(5 星评分):这个好便宜但是产品不-质量过硬!!!爱它，经常用!
- 10.<优秀，强大的动作>(5 星评级):我们已经连续使用这款吹风机几个星期了。总的来说，这是一个强大的吹风机，可以让你的头发快速干燥，而且热量平衡极佳。没有廉价吹风机能吹出的热点。是的，它是笨重的，但那是因为质量建设。我毫不怀疑，这款烘干机将继续使用多年。
- 11.<漂亮整洁的小烘干机>(5 星评级):我妻子在我们住的汽车旅馆用过一个。她很想要一个，所以我就买了。你的老太太也会喜欢的，因为它很强大，而且有一个绳卷。之后你会恨自己，因为她抱怨吹起来不像以前那样用力，你不得不拆开它把她的头发从里面挖出来。但女人就是这样.....Blah Blah Blah 抱怨 Blah Blah
如果你能在她需要干她湿漉漉的头之前，把这个东西拆开，清洗干净，再按照工作顺序重新组装好，那就算是幸运了。我觉得核弹做起来会更容易。
- 12.<好产品，好价格。>(5 星评分):这是我的第二个同款吹风机。(Conair Ionic Conditioning Pro Style 1875 瓦电吹风)[[ASIN:B00005O0MZ Conair Ionic Conditioning Pro Style 1875 瓦电吹风]]我想再来一遍。
- 13.<爱它!!史上最好的烘干机>(5 星评级):它让我的头发如此好看，没有冻住!!!!!!我很伤心，我不小心弄掉了我的三年后，我有它与我，但我防卫地再次购买它!!
- 14.<非常适合细头发!>(5 星评分):我的头发很细，不喜欢做造型。使用这款吹风机，我有;

Volume
Body
Curl

，而且，我不需要使用任何产品就能做到这一点。我相信关键不仅在于烘干机，还在于你用的洗发水。我用的是含角蛋白的洗发水，搭配离子超轻量护发素。最重要的是——这款吹风机太棒了!
- 15.>(5 星评级):我的头发很浓密，这确实缩短了我的干燥时间。我用完后用冷键，它似乎让我的头发更容易打理一些。
- 16.<最好的投资!>(5 星评价):我现在会少去发型师那里存钱。这台吹风机吹干我头发的速度比我之前用过的吹风机快了两倍。我真希望我几年前就买了这个吹风机!
- 17.<优秀!!>(5 星评分):我的客房浴室里有一个，客人可以用。你不需要在厕所的下方找一个。



关注数学模型
获取更多资讯

- 18.<T>(5 星评分):这款烘干机太棒了，你拥有的力量是首屈一指的..，有点重，但从它留下的效果来看，值得一试。我很开心。
- 19.<我老婆很喜欢(这个我做得很好)>(5 星评分):它会严重地让你家里的灯变暗，但是男人可以用它干头发！
- 20.<我喜欢这个吹风机！>(5 星评级):它能快速吹干我的头发，体积小大到可以装进行李箱，有可伸缩的线绳，让我的头发出奇的丝滑顺滑。而且价格也是你无法比拟的。只要这款还在生产，我就再也不买别的干衣机品牌/型号了。
- 21.<爱它！>(5 星评级):虽然小，但包了一些严重的热度和功率！
- 22.<爱它！>(5 星评分):我大概买了另外 5 个吹风机还退了，才留了这一个。我和我的孩子都有很多头发，要花很长时间才能吹干。这款吹风机是最好的;它吹干我们头发的速度比以前的快得多。不用担心不小心关掉或者改变设置，冷拍也不用按住按钮，如果常规设置对你来说不够快，甚至还有涡轮设置。这帮我们早上节省了好多时间!我在考虑再买一个放在另一个浴室里，这样我们两个人就可以同时吹干头发了。一开始看起来确实有点大，但没过多久就习惯了。其他吹风机的气流太大，我无法控制自己的发型，或者是高热量设置太热，但这台的设置刚刚好。
- 23.<既方便又很棒的产品。>(5 星评级):正是我们想要的。
- 24.<染过色的头发不再干枯>(5 星评分):因为好评如潮，我买了这个。我染了头发(为了遮灰)，吹干头发总是花太长时间，头发看起来很干，到处都有一些翘起来。我买了这个，天哪!!干起来花了一半的时间，而且没有飞毛!! !我得到了很多赞美。我很喜欢，但从没想过会有人注意到。我确实得到过这样的赞美。记住我已经过了年纪，头发似乎随着年龄的增长而变得干燥，染发并没有什么帮助，除了遮盖住灰色，我为此感到高兴。但是如果你觉得你的头发变得太干或卷曲，头发会飞起来——这个可以解决问题。这是一次性的购买，非常值得。3 个温度和一个我喜欢的很酷的按钮。我能说什么呢，哦，它没有那么重，和其他普通重量一样，除非你想用便宜得多的吹风机吹一下，发出很多噪音。我做过这样的事……这就是炸弹!!别再把时间和金钱浪费在让人抓狂的吹风机上了，对自己好点，买这个牛逼的吹风机吧。多保重，你一秒都不会后悔。
- 25.<作品太棒了！>(5 星评分):非常实惠，评价准确。比我的老款安静，速度也快。有了 Conair 的质量，就不用花很多钱买烘干机了。
- 26.<好干燥！>(5 星评级):在纽约的一家酒店发现了这颗小宝石。吹干头发，不会留下毛躁和飘逸。折叠式把手，方便收纳和出行。



关注数学模型
获取更多资讯

- 27.<大吹风机!>(5 星评分):这款吹风机是我用过的最轻、最安静的吹风机。事实上，我老公现在也想要一个自己的!
- 28.<那么热。爱与低噪音>(五星评分):我爱爱爱这个吹风机....洗头座 30 分钟我就搞定了...
- 29.<爱它!!>(5 星评分):以前每天都害怕吹干头发。头发又长又厚，过了半个小时还是干不了。这款吹风机最多 3-4 分钟就能让我的头发完全吹干。它完全改变了我的生活。我又买了一个就在手边。你永远不知道吹风机能用多久。
- 30.<爱它!>(5 星评分):我买了这台烘干机来换旧的，哇，真不一样。我的头发很柔软，很有光泽。我甚至可以用它来吹干我 1 岁女儿的头发，但我不能用我的旧头发，因为即使在较低的设置下，它也会变得太热。它干得快，不会烧焦。绝对物有所值!
- 31.<超级小发电机!!>(五星评分):我已经有一台了，但因为我经常旅行，所以我家里有一台，旅行拖车也有一台。一个小烘干机里有很多热量和电力。
- 32.<IN LOVE>(5 星评分):I LOVE eeeeeeeee IT
- 33.<爱它>(5 星评分):一年后还在用这个.....爱这款可伸缩的线
- 34.<我爱它>(5 星评级):比我的旧款更快，线卷在里面不占用额外空间，我觉得很方便。
- 35.<爱它>(5 星评分):每天都用
- 36.<我喜欢它>(5 星评级):他们说的都是，它占用很少的空间，而且能起到作用
- 37.<爱它>(5 星评级):这个是我用了很多年的那个的替代品。
- 38.<伟大!>(5 星评价):贵但物有所值。专业的吹风机，专业的效果。减少在浴室里吹头发的时间
- 39.<爱它!>(5 星评分):我头发很厚，这个宝宝让我的头发很快就干了，没有毛躁!爱它!
- 40.<很棒的低档吹风机>(5 星评级):这款吹风机非常适合旅行。可折叠的手柄便于打包。脐带牵引器很棒
- 41.<工作原理就像烘干机一样...>(5 星评分):我用吹风机的目的，除了烘干，就没别的了...哈哈很简单，in-expensive，还吹了很多热风...

它很好地达到了目的...
- 42.<爱它!>(5 星评分):干得好快棒!我从法国带了一个过来，那是我妈妈多年前买的，但不幸的是，由于电压差，我永远无法得到相同的功率。我订购了这个，它是一样的
(颜色除外)。把头发常规剪成两半。
它会冒烟，发热很多，所以要小心!
(我有黑色放松的头发)



关注数学模型
获取更多资讯

- 43.<非常有效!>(5 星评分):我曾考虑花一大笔钱买一台高端吹风机, 但后来我意识到这很愚蠢...而且我很便宜, 所以我买了这个, 说实话, 我太开心了!我的头发很长(长到后背中间), 而且很厚, 这台吹风机不到 10 分钟就能把我的头发吹干。我上一个吹风机用了。ev。呃, 为了避免我的丈夫盯着我看, 看着他眼中充满仇恨, 因为我花了很长时间准备, 所以他不让我吃我们要吃的饭, 我不会费力地吹干和卷我的头发, 而是直接摇我的自然发型, 但有了这台新吹风机, 我可以做我的头发, 而且仍然可以快速走出家门(相对而言)。

我强烈推荐这款吹风机!没必要把辛苦赚来的钱浪费在一台高端烘干机上;这台就够了!
- 44.<看起来很棒>(5 星评级):不再需要回归
- 45.<牛逼产品>(5 星评级):我们有一款 18 岁以上的, 终于出了!新款超越了之前的型号!
- 46.<伟大的烘干机>(5 星评级):买这个是为了替换(意外)被踩到的那个。它具有非常专业的品质, 没有高昂的价格。
- 47.<Great power>(5 星级):快速干燥, 充满光泽-我第一次接触这款吹风机是在我住过的一家 5 星级酒店-它能快速吹干头发, 充满光泽。
- 48.<伟大的烘干机>(5 星评级):想要和我们 Y 店墙上的一样的品牌和功率。< br />完美。正是我的配偶想要的。
- 49.<很棒的吹风机>(5 星评级):我的头发用这款吹风机很柔软, 不会卷曲。而且, 它不像我以前那个更便宜的烘干机那样吵。
- 50.<伟大的电吹风>(5 星评级):这款电吹风功率很大。壁挂式可以让它不碍事, 但在需要的时候很容易拿到。这是为我们更换的吹风机, 当另一个不工作时, 我们购买了相同的一个。
- 51.<伟大的产品!>(5 星评价):我在一本杂志上读到过 John Frieda 的全卷吹风机, 它完全没有辜负它的评价。干得快, 让头发柔软, 没有毛躁, 还很有质感。它让我的卷发第一次有了自觉性。我强烈推荐这款产品。没有它我就活不下去。
- 52.<很棒的>(5 星评级):这款烘干机正是我想要的, 非常轻, 而且效果很好
- 53.<伟大的>(5 星评分):完全符合预期
- 54.<不再有纠结的电线!>(5 星评级):这款吹风机功率大, 发热设置不同。价格和质量都很棒, 非常重负荷。自动后坐力线是加分项!
- 55.<好干燥!>(5 星评级):买这款吹风机, 是为了换掉之前用了多年坏掉的 Hot Tools 吹风机。这款吹风机甚至比我以前用过的还要好用!它的重量轻了很多, 而且仍然有和上一个一样的快干力。我把

到处都是, 因为没有其他吹风机能像它那样快速、轻松地吹干我的多层长发。



关注数学模型
获取更多资讯

- 56.<美妙。潮湿的。负担得起的。几分钟就能搞定。优秀的>(5 星评级):设置非常容易和快速。不需要工具。实际上只需要几分钟。我放了一个吧台凳坐着，用毛巾盖住通风口。就像别人说的，用毛巾盖在新鲜空气孔上是需要的，这样才能保持蒸汽在里面。我不知道为什么人们会抱怨通气孔。花 2 美元你可以给他们买一条毛巾，或者花 2000 美元你可以买一个不同风格的蒸汽桑拿。说真的，朋友们，买条毛巾吧。再加点香薰盐，简直就是幸福！区区 200 块，很难找到这种极乐。我在 30 分钟内达到几乎令人痛苦的高温，然后降低温度，上车，重置计时器。图上显示的是有点偏离，因为门是从一侧打开的，而不是图上的中间打开的。不过不用担心。它很好。我会再买一次的。
- 57.<世界上最好的吹风机!>(五星评分):牛逼!!!!!!!!!!!!!!!!!!!!!!
- 58.<爱死这个吹风机了!>(5 星评分):这款吹风机很棒，它重量轻，干头发的时间是我用过的所有其他吹风机的一半。
- 59.<物美价廉的吹风机>(5 星评级):收到这款吹风机的时候我真的很惊讶，因为产品的质量很棒
这款吹风机非常实惠，外观和手感都很好，能发出很好的热量。我一定会把这款产品推荐给朋友，用的很开心。
- 60.<我们爱它>(5 星评级):我把护发素加到日常生活中。我老公发现，当他同时使用洗发水和护发素，并让护发素开了几分钟后，真的很有帮助
- 61.<爱我的吹风机>(5 星评分):没有它我活不下去!
- 62.< >附近最好的(5 星评级):这是我购买的第三个壁挂。喜欢它的便利性，而且不用放在抽屉里。
- 63.<我有过最好的吹风机!!>(5 星评级):没有。
- 64.<爱它>(5 星评级):几个月前买了这个吹风机，用来换旧的(另一个牌子)。这款新的吹风机和我上一个吹风机一样强大(我很喜欢)，但是安静多了，之后我的头发也不会有卷曲的感觉。我唯一不满的是，吹风机的手柄底部有点脱皮。
- 65.<超级安静>(5 星评分):这个非常非常安静。它有最长的线。我在烘干机上见过的最长的。我不懂离子这个东西。
- 66.<太好了!>(5 星评级):这个让我的晒干时间缩短了一半!虽然比我预想的要重一些，但还是超出了预期。
- 67.<优秀!!>(五星评分):这玩意对我老茧的脚和角质层都有神奇效果。除了抗菌，你不能错过这款产品。很少使用，所以它会持续使用很长时间。
- 68.<伟大的吹风机>(5 星评级):直播这个吹风机
- 69.<很棒的吹风机>(5 星评级):这款吹风机真的能让你的头发更柔软，我

我已经用了一个月了，我只有在睡觉前或者出门前头发湿了才用它，它只需要 5 分钟就能吹干我的头发(我大的那个用了 30 分钟)，我强烈推荐它。



关注数学模型
获取更多资讯

- 70.<一个很棒的烘干机。>(5 星评级):当你不想坐在吹风机下或用吹风机吹干头发时，这就是你需要的。当它吹干你的头发时，我可以带着它在房子里到处走。
- 71.<完美>(5 星评级):我从我的国家拿了一个吹风机。但它不工作，因为。不同的电压。

所以我不得不买一个新的，然后我选择了它，我在两天后得到了它。好快啊!

It is perfect@!
- 72.<完美>(5 星评分):正是我想要的。虽小，但功能强大。
- 73.<伟大的产品!>(5 星评分):这是一款很棒的烘干机!我的第一个在持续使用了 5 年之后终于坏了，我最后尝试了一个类似的 Gold N Hot，第一次使用它就停止工作并且过热。我很快订购了 Andis，没有任何问题!我有一头长长的卷发，它在 20 分钟内就能把我的头发吹干。一旦我穿完刚买的这款，我一定会第三次购买这个产品。
- 74.<好主意>(5 星评级):我们有一个没有储物间的小浴室。这个吹风机太完美了。不用的时候很碍事。不占用抽屉空间。甚至还有夜灯。
- 75.<伟大的产品>(5 星评级):干得快，让我的头发有光泽、直发。它有点吵，但是它会释放很多空气，所以我想这是意料之中的。
- 76.<伟大的产品>(5 星评级):烘干机重量轻，传递大量热量。快速吹干、抚平头发。将头发干燥时间缩短一半。不会回到标准干燥机。
- 77.<了不起的!>(5 星评级):即使是最细的头发也能卷起来，卷起来能卷一整天。瞬间升温，不会烫到头发。发胶和/或定型啫喱可以延长卷发的寿命，但不是必须的。我可以在 15- 20 分钟内完成整个头部，而且我有长发。我发现唯一的缺点是，如果你烫到自己，你肯定知道。孩子们使用它不安全，因为它会变得很热。
- 78.<作品太棒了!>(5 星评分):快速吹干头发。我的头发很厚，这台吹风机让我的时间缩短了一半。我会向任何人推荐它。
- 79.<伟大的产品>(5 星评分):我一模一样的吹风机用了 20 年!我一直买它是有原因的，尽管它似乎用了 4-5 年就会死。夜灯非常适合主浴室使用。现在 we 不会在半夜打开正常灯把对方吵醒了。
- 80.<神奇:)>(五星级评级):我一直在使用这个洗面奶去年 3 年了,我所能说的就是这张脸洗奇迹,它清除所有我的痤疮,粉刺留下的所有我的伤疤,并使我的皮肤柔软,当然这可不是仅仅是因为洗面奶,我不得不穿上面部乳液为粉刺真的消失但是这个洗面奶也没有什么,不重皮肤上相比其他许多粉刺洗面奶我试过了
- 81.<世上没有安静的吹风机。>(5 星评级):这款吹风机虽然不安静，但噪音比一些吹风机有吹口哨的噪音或噪音要忍受得多



关注数学模型
获取更多资讯

超级大声。我喜欢它可折叠和可伸缩的特性，这让它更容易收纳和旅行。

- 82.<超棒的烘干机>(5 星评分):我已经连续两年想要这个圣诞礼物了，终于我家有人买了，现在他们想要一个!
- 83.<The best>(5 星评级):我用过的最好的吹风机。我的第一个每天用了 12 到 15 年，记不清有多少年了。我猜它死于染发年 110 岁。
- 84.<超棒的烘干机>(5 星评分):我在我未婚妻的公寓住了几晚的时候拿了这个烘干机，这样我就可以在我不在家的时候把它藏在他的地方。它折成两半，绳子缩回到手柄里，让它变得非常小。我立刻对它的力量印象深刻，我的头发很厚，但我可以在几分钟内完全吹干它。我们结婚买了房子，它一直陪伴着我，成为我主要的吹风机。我每天都用它，到现在已经 3 年了，它一点也不出毛病，绝对值得!
- 85.<伟大的购买。>(5 星评分):虽然是为了旅行才买的，但比起普通大小的烘干机，我更喜欢它。
它是粉色的，但是图片上看起来是紫色的，这就是我给它 4 分而不是 5 分的唯一原因。
- 86.<很棒的吹风机!>(5 星评分):没想到电吹风居然能让你的头发感觉柔软!这款做到了，2 秒就把我的头发吹干了。好吧-也许不是 2 秒，但它真的很快。
- 87.<Happy Customer>(5 星 评分):符合我的预期。
- 88.<完美产品>(5 星评级):产品及时发货，与广告完全一致。不太多，也不太少.....能帮我和洗完澡后的吉娃娃干头发。
- 89.<最佳>(5 星评级):现在有吹风机几个月了...没有问题不过热....最好的投资
- 90.<The Best>(5 星评级):唯一一款侧边有通风口的吹风机。这对我的工作至关重要，更新的后置通风口对我来说毫无价值。
- 91.<爱它。>(5 星评级):烘干机准时送达。它功能强大，2 分钟就能把我的短发吹干，而且不重。爱死它了。
- 92.<我拥有过的最好的>(5 星评级):这款吹风机我怎么说都不够好。我有一头长长的卷发，花了一个多小时才吹干。现在我可以准时去上班了。我之前的烘干机更贵，甚至没有那么好，也没有那么快。再也没有理由花更多的钱了。我是一个非常满意的顾客。毫无疑问，下次我需要新吹风机的时候，我会成为回头客的。
- 93.<我爱它>(5 星评级):独创的产品。简单方便。我把它粘在镜子上，然后就呆在那里了。不用再用一只手拿着烘干机刷牙了
用另一只手梳理我的头发!物有所值



关注数学模型
获取更多资讯

- 94.<五星>(五星评级):气压大
- 95.<我有过最好的吹风机!>(五星评分):快速吹干头发，不炸!!!
- 96.<非常好用>(5 星评级):它具备一个好的烘干机应该具备的一切功能。遗憾的是，这条可伸缩的线没有它应有的那么好，但可以管理。
- 97.<爱它!!我买的东西再满意不过了。>(5星评级):我买这个吹风机是为了换一个从朋友那里借的，因为电转换器用错了烧坏了。因为价格实在是太合理了，又因为包邮要达到 25 美元，所以我买了两个，自己留了一个。我每天都用这台烘干机，而且非常满意。我的头发有点长(略低于肩膀)，这款吹风机轻盈有力，能快速吹干我的头发。它可以做一个很棒的旅行吹风机。它比大多数酒店房间里的烘干机好用得多。这个价格你不会出错的。它有双电压的事实是一个很好的优点。
- 98.<爱它>(5 星评分):爱这款吹风机。一切如其所述。说真的，人们，按住插头的同时按下缩回按钮，没有人会受伤!
- 99.<我爱它>(5 星评级):拥有我想要的一切，去 000 000 000 000 000 000 000 d
- One hundred.很棒 的旅行吹风机。>(5 星评级):尺寸和功率刚刚好，适合旅行。
- 101.<这东西太棒了!>(5 星评分):其实我是在找别的东西的时候偶然发现了这个，心想 "why not";。我对这次购买非常满意。正如另一位评论者所说的那样，我有很好的肩长头发，而且不用一个小时就能吹干。(当然，当头发湿漉漉的时候，我也没有试图吹干头发。我会用毛巾把头发包起来，做一会儿其他事情，然后再尝试吹干。)总之，一开始我很担心，因为包装和大部分说明都是日文的。但我在低电压的时候用了，还好。我用普通的发刷吹干头发，因为我还不知道怎么用圆发刷。我吹干了大部分头发，然后把它倒过来用，让我的头发有一些提振和弹性，它确实做到了。而且，我的头发看起来也不会像使用普通吹风机时那样卷曲(我相信这是因为有时候普通吹风机既会损伤头发，又容易打结)。我的头发实际上看起来既光滑又有光泽(就像我平烫过一样)。如果你喜欢自己的头发被吹干或熨平的样子，但又看起来损伤小得多，我会尝试一下这个方法。我认为你可以只使用这个吹风机和一点精华液(取决于你的头发有多自然直)。我很兴奋地使用它，并试图弄清楚圆刷。
- 102.<伟大的产品>(5 星评分):我老婆试过很多电吹风，我不敢相信她对这款电吹风有多满意…完全值得!
- 103.<爱它>(5 星评级):三速(噪音)两热设定。我四岁的儿子不再抱怨吹头发了，因为速度更慢了。物美价廉。
- 104.<伟大的价值!>(5 星评分):瞬间加热，非常快干头发，比我用过的其他牌子都安静得多…非常值这个价。



关注数学模型
获取更多资讯

- 105.<完美>(5 星评级):完全做到了它需要做的。坚固耐用。变得如此热，让你想要打你的妈妈。买它。
- 106.<爱这款吹风机>(5 星评级):这款吹风机比我之前的吹风机干得快多了，而且看起来还能提供更顺滑的触感。推荐。
- 107.<最好的吹风机…句号!>(五星评分):真实故事....我花了大约 500- 700 在不同的头发和吹风机上。从便宜的到贵的我都买了，这是迄今为止我拥有的最好的吹风机。我是一名黑人女性，有着天然的 4c 头发，这意味着又厚又粗糙，通常当吹干我的头发时，我的头发感觉很脆，但这款吹风机并没有让我的头发感觉像那样。当然，在吹干我的头发之前，我用了 Alikay Naturals 产品的 LOC 方法，它变成了我需要的直发。我超级开心和满足。我还用了发夹 yo 吹干我的头发，它没有弹掉，我也没有在发夹中折断牙齿。炸药!如果它弹开了，就用耐热双面胶带把它粘住。
- 108.<爱这只吹风机>(5 星评级):这只吹风机我已经用了好几年了，直到插头上的浪涌保护器按钮一直冒出来。我在沃尔玛买了一个 Conair 电吹风换掉，回到家插上电源，简直不敢相信它听起来这么便宜，声音这么大。我立刻退了回去，发现这款雷明顿电吹风在亚马逊上有售，我很高兴。这款吹风机我完全买账了。电机安静，干燥时间短。唯一的问题是刚开始感觉有点重。但用了几次之后，我想我已经习惯了，因为现在这个问题已经不存在了。
- 109.<别让便宜的价格骗了你>(5 星评级):我的电吹风坏了，我需要一个新的，但我发现他们不再生产了。我在当地的化妆品供应链上买了一个吹风机，想着它会和我亲爱的已故的那个一样好用。错了!它让我的头发感觉非常干燥，尽管我在吹干头发之前用了产品。我本来打算投资一个 CHI，但在看到大多数关于那个型号的负面评论后，我决定攒钱买这个 Andis 吹风机。我的头发又粗又长(超过我的肩膀 4 英寸)，我正在从松弛的发型过渡到自然的发型。我有大约 3-4 英寸的新生长，其余的仍然是放松的，所以直发型暂时有效。这款吹风机就是一切!它不仅方便钱包，还能让我的头发超级快地变干，同时又不会把头发里的水分吸光。它只需要不到 20 分钟就能吹干我的头发，速度甚至比我原来的吹风机还快，还能把我的发根吹得相当直(我用的是小齿梳附件)。它握起来很方便，而且很烫，不会烫到我的头皮。再加上我的 Sedu 平烫，我的头发几乎是沙龙级的，超级柔软有光泽!我会把这款推荐给和我发质一样的人。我很欣慰!
- 110.<很棒的吹风机>(5 星评级):我买了这个来取代我买的另一个雷明顿
- 用了几年了。到目前为止，我对我买的东西很满意。我选择这款是因为我以前买过的雷明顿的质量很好，而且这款似乎和其他几款没什么区别。质量。吹风机确实有轻微的橡胶感。我从来没有过这样的吹风机。不能说我在乎;只是不一样。没有好坏之分。

我确实特别选择了这一个，因为附加的扩散器。对于我们这些卷发的人来说必不可少，而且效果很好。



关注数学模型
获取更多资讯

- 111.<大干燥机…>(5 星评级):拥有一段时间。目前没有问题。和我第一天买的时候一样好用。一些物美价廉的功能
- 112.<诚实>(5 星评分):我和老婆每天用这个产品至少 10 年了，终于用坏了，只好买个新的。
- 113.<很棒的吹风机包装成小号!>(5 星评分):完美尺寸。就是我想要的!!不抱怨!
- 114.<伟大的产品>(5 星评级):我用了 Andis 电吹风至少 8 年，直到最近它在我身上死掉。
替换的很棒;产品还是和以前一样可靠!
- 115.<很棒的烘干机>(5 星评级):对我来说效果很好，因为我每天都用。这款产品没有投诉。使用和操作方便。

附录 D VADER 情感分析源代码

```
从 vaderSentiment。 vaderSentiment 导入 SentimentIntensityAnalyzer
#VADER 免费获取 athttps://github.com/cjhutto/vaderSentiment
Data = pd.read_csv(loadurl,sep= '\t')[[' star_rating ', ' review_headline ', ' review_body ']]
维达号模型
def 变换 (df):
    d = df.values.tolist
    ()
    对于 d 中的元素:
        Vs1 = analyzer.polarity_scores(str(元素
        [1]))l1=list(Vs1.values())
        Vs2 = analyzer.polarity_scores(str(元素
        [2]))l2=list(Vs2.values())
        element.append (l1,
        l2)
    返回 pd.DataFrame (d)
data=变换(数据)
数据。列=[ '明星' , '头' , '身体' , 'h_sn' , 'h_wn' , 'h_m' , 'h_wp' , 'h_sp' , 'b_sn' , 'b_wn' , 'b_m' b_wp" 、 "b_sp "]
data.to_csv (outputurl)
```

附录 E 信息性算法的源代码


```
def cosdistance (v1,
v2):
    vec1 v1
    =
    vec2 v2
    =
    返回 1-fenzi / (fenmu1 * fenmu2)
def 熵 (v1):
    总 = 0
    vec v1
    =
    For i in range(5):
        Total +=
        vec[i]*math.log(vec[i])返回 Total
```

#将数据更改为 one-hot-vector 形式



关注数学模型
获取更多资讯

```
#数据导入
data = pd.read_csvloadurl
()

intvec = onehotvec(数据(“星”)).values.tolist
()

vecvec =数据[[p1,p2,p3,p4,'p5']] .values.tolist ()

帮助=pd.read_csv (r想:\Problem_C_Data \ hair_dryer.tsv ',9 = '\
t”)(“ helpful_votes” )。 values.tolist ()

date=pd.read_csv(r'd:\Problem_C_Data\hair_dry .tsv ', sep= '\t ')[ 'review_date ' ] #
获取信息价值

卡尔
= []

For I in range(len(help)):

    x =(1 +帮助[我])*战俘(np.e, (1) * cosdistance (intvec[我],vecvec[我]))*战
    俘(np.e 1 *
                                熵(vecvec[我]))

    cal.append (x)

结果= pd.DataFrame (cal)
```

附录 F 信誉计算的源代码

```
def 乙状结肠
(x):

    返回 1/(1 +战俘(np.e, (1) * x))

def get_reputation (df):

    data = df.values.tolist
    ()

    =数据(:,1)

    曲线= np.zeros .tolist (len(数
    据))()

    曲线[0]= 1

    #参数设置

    p =
    0.75

    n =
    0.25

    k1 =
    25

    k2 =
    20

    # k =
    2

    For I in range(1,len(data)):

        如果数据
        [我 ][0]> = 0:

            如果数据
            [我 ][3]= = 0:

                曲线[我]=曲线(张)+ (pn) *数据
                [我][0]

            其他:

                曲线[我]=曲线(张)+ (pn) *数据[我][0]乙状结肠(k1 *曲线(张))* k2
                *数据

                [我][1]/[我][3]
                的数据

    其
```

```
他：
    如果数据
    [我][3] == 0:
        曲线[我] = 曲线(张) + (p + n) * 数据
            [我][0]

    其他：
        曲线[我] = 曲线(张) + (p + n) * 数据[我][0] 乙状结肠(k1 * 曲线(张)) *
            k2 * 数据

    [我][2] / [我][3]
    的数据

    返回 pd.DataFrame(曲线)

#计算
curve1 = get_reputation
(inputdata)
```

附录 G 贝叶斯模型源代码

```
def word_freq (df, s):
    data = df(身体的).values.tolist ()
    = 0

    For I in range(len(data)):
        if s in data[I]:
            返回 数 / len(数据)

def countfive_prob(数据):
    d = 数据
    ( “星” ).values.tolist() = 总
    len (d)

    数 = 0
    返回 数 / 总

def countfive_word (df, s):
    数 = 0
```



关注数学模型
获取更多资讯

```

    For I in range(len(data)):
        如果数据
        [我][0] = 5:
            += 1
            如果 s 在
            str(data[i][1])中 :
                数 += 1

    返回 数/总

def 火车(str):
    s =
    str

    Data =pd.concat([data1,data2,data3], axis=0)
    Data = Data [[' head ', ' body ', ' 0 ', ' star
    ']] five_prob=countfive_prob(Data)

    word_star = countfive_word(数据、
    年代)

    word_frequency = word_freq(数据、
    年代)

    Result =word_star*five_prob/
    word_frequency 返回结果

textdata = pd.read_csv
loadurl()[[ ' review_body' , ' review_head']] wordlist =
pd.read_csv (loadurl2)

#获取每个高频率单词的 beyes prob 值
结果 = []

For I in word:
    result.append((我,火车
    (i)))

r = pd.DataFrame(结果)
```

附录 H 时间序列预测源代码

```

从 statsmodels.tsa. ar_model 导入 AR

#训练自回归
模型 = AR(train)
Model_fit = model.fit()
Window = model_fit.k_ar
Coef = model_fit.params

#在测试中沿时间步向前走
History = train[len(train)-window:]
History = [History [i] for i in range(len(History))]
Predictions = list()

For t in range(len(test)):
    Length = len(history)

    Lag = [history[i] for i in range(length-window,length)]
    yhat = coef[0]

    For d in range(window):
        Yhat += coef[d+1] * lag[window-d-1]

    predictions.append (yhat)

error = mean_squared_error(测试, 预测)print(
测试 MSE: %。3f '%错误)
```

print('测试 RMSE: %。 3f'% math.sqrt(错误))

附件我词云图片源代码

来自 wordcloud 导入 wordcloud, ImageColorGenerator, STOPWORDS
txt=pd.read_csv(loadurl)

```
my_wordcloud = WordCloud(  
    background_color = '白',  
  
    掩码= imageio.imread                                #参数设置  
  
    max_words =  
    100,  
  
    Max_font_size =100,  
    random_state=50,  
    width=1000,  
  
    Min_font_size =20,  
    relative_scaling=0.4,  
    color_func=genclr,  
  
).generate  
(txt)  
  
# 显示词云图片  
plt.imshow(my_wordcloud)
```



关注数学模型
获取更多资讯