**主题因素对商品评论的影响**

摘要：

本文采用神经网络方法对服装、鞋子和珠宝饰品的商品评论进行有用性和极性预测，其输入层为LDA从语料库中提取的主题-文档矩阵。然后，通过主题-文档矩阵训练得到神经网络参数，进一步计算每个主题与预测结果之间的关系。实验结果表明，在服装、鞋子和珠宝饰品的评论中，主要有19个主题。然而，对评论的有用性和极性的预测并没有取得良好的性能，因为输入层除了主题因素外没有考虑其他影响有用性和极性的因素。通过神经网络参数计算关系强度，分别选取对有用性和极性影响最大的主题。服装的美感和设计、服装的尺寸、颜色、评论者的观点和评论者提供的信息是影响评论有用性的前5个主题。同时，家庭对商品的感受、商品的性价比、情感的表达、珠宝饰品的主题、服装的舒适度是影响评论极性的前5个主题。这项研究将帮助卖家了解客户在购买某一类产品时关心什么，并告诉评论者如何编辑他们的评论，以便为买卖双方提供更多有用的信息。

1.简介

在电子商务飞速发展的今天，人们逐渐倾向于通过互联网购买自己的需求。从家具、电器等大型设备，到服装、食品等小型商品，甚至在线节目、游戏软件等非物质产品都可以通过网络支付方便地接入，大大节省了消费者的时间和精力。虽然我们的物质生活和精神生活得到了极大的改善，但电子商务中也存在着一些隐藏的问题，阻碍了电子商务环境的积极快速发展。例如，付款和运输过程中的产品质量问题和服务问题。这些问题不仅损害了客户的利益，也损害了企业的信誉和信誉。

解决这些问题的有效办法是加强买卖双方以及买卖双方的互动。受到Web 2.0启发的解决方案并没有过多地关注增加更多的直接利润，因为与导致信息技术和Web内容普及的Web 1.0相比，Web 2.0增强了网络作为媒介和社会的功能，创造了更加个性化的信息,但这并不意味着电子商务评论不会给电子商务卖家带来任何好处。事实上，丰富的顾客评论为潜在购买者提供了决策信息参考，对卖家的产品和服务进行了有效的监督和反馈，从而打破了人们选择网购的心理障碍。以上优点提高了购物的可靠性，也间接给整个电子商务行业带来了可观的利润。虽然商品评论的兴起不可避免地会导致一些不公平竞争，比如使用“互联网水军”来影响产品的评级，但强交互性给电子商务带来的积极方面仍然是不可磨灭的。

目前，几乎所有的电商平台都为客户提供了评论功能，其中最完整的学术分析平台是美国的亚马逊。亚马逊的优势在于：（1）客户评论的排序和评价机制。可以根据评审人员的所有评论来评估其内容的有用性，并将其排在所有评审人员的前面。审稿人会通过他们的排名在一定程度上感受到激励和监督。（2）有用的一票。消费者可以选择阅读页面上的评论是否有用，很多平台也在效仿，但消费者在阅读评论时很少按有用按钮。（3）评论有组织性。每个评论都可以追溯到产品和审稿人，甚至是审稿人评论过的所有产品。这些优点使得使用文本处理技术分析影响评论有用性的因素成为可能。

已有许多研究对影响评论有用性的因素进行了分析。然而，这些论文更多地关注非文本或主题因素，如长度和极性。此外，学者们更喜欢回归方程法而不是文本挖掘算法。因此，本文主要探讨评论的主题与评论的有用性和极性之间的关系。首先，采用LDA (Latent Dirichlet Allocation)方法从商品评论语料库中提取主题信息。然后，利用LDA得到的文档主题矩阵作为神经网络的输入，预测评论的有用性和极性。最后，提取神经网络的参数，计算输入和输出之间的关系值，以发现影响评论有用性和极性的主题因素。

在本文中，第一节介绍了本研究的背景。第二节是本文可能涉及到的相关工作。第三节描述了评价主题因素对商品评论影响的实验过程。第4节展示了实验结果，并详细说明了本次工作与以往研究的异同。最后，第五部分是本文的结论。

2.相关工作

**2.1影响评论有用性的因素：**

根据已有的研究，商品评论的特征包括数量、价(极性)和质量三个方面。这些特征会对产品的销售和评论的有用性产生积极或消极的影响。

在数量方面，许多学者发现评论的数量对产品销售有正向影响。其中，有学者不仅发现了评论数量与产品销量之间的关系，而且还提出评论的效价与产品销量之间的相关性并不像评论数量那样强。路德维希等人对评论效价的进一步研究得出结论:对于风险规避者来说，负面评论比正面评论更有价值。这一结论是通过比较正面评论和负面评论对产品销售的转化率得出的。评论质量是一个比较抽象的概念，主要代表评论的可靠性。质量的衡量是不固定的，因为实际情况总是变化的。常见的度量指标是审稿人的可信度、评论的长度和评论的可读性。

商品评论影响产品销售的原因是它们为潜在客户提供了有关购买决策的信息。除了评论数量的宏观影响外，每条评论的内容也会对买家的决策产生影响。投票的有效性的作用每个评论评论页面可以量化的影响买家的评论,这使得它的学者来衡量的各种特征的影响商品评论评论的有用性。

关于不同评论特征对评论有用性的影响，已有大量的研究得出了重要的结论。例如，注释长度越大，注释就越有用;带有极端星级的评论往往更有用;评论的有用性还会受到非内容项的影响，如产品类型、评论人的特征等。上述研究都没有探讨主题与评论有用性之间的关系，尽管评论中包含的主题信息显然对买手的决策有显著影响。

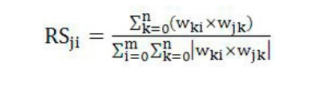
由于文本挖掘技术的复杂性，很少有研究采用文本挖掘来研究主题与评论有用性之间的关系。虽然Cao、Duan和Gan使用了潜在语义分析(Latent Semantic Analysis, LSA)提取了评论中包含的话题因子，并分析了这些话题对评论有用性的影响，但他们没有解释每个因素的含义。在此基础上，Ahmad和Laroche也使用了LSA，但他们进一步解释了影响评论有用性的各种主题。此外，本研究还根据不同的星级来分别考虑正面和负面评价，从而导致不同效价下的结论不同。研究表明，增加正面评论有用性的主题是商品功能和技术参数，而增加负面评论有用性的主题是订单和运输服务。但是，研究选择的商品类型可能会影响分析的结果，从而降低结论的普遍性。因此，本文选择了另一种商品类型来分析选题的影响，以便对已有的研究进行比较和补充。

**2.2．LDA和神经网络：**

LDA是一种文档的主题生成模型，与LSA一样属于主题模型。它们都可以将文档-词矩阵分解为文档-主题矩阵和主题-词矩阵，从而达到降维的目的。LDA与LSA不同，LSA通过矩阵计算忽略不重要信息来降维，LDA根据预先设置的主题号，利用贝叶斯估计和抽样方法对语料库进行训练，直接得到文档-主题矩阵和主题-词矩阵。虽然LSA不需要预先设置主题的数量，但仍然需要使用SVD (singular Value Decomposition)方法根据奇异值人工确定提取因子的数量。如果奇异值太多或者奇异值非常相似，模型最终也会忽略很多的信息。同时，LDA方法不需要丢失任何文本信息，但是很难确定合适的主题数量。

不同于已有的研究采用LSA，本文选择LDA提取评论的主题信息。这是因为在SVD的过程中很难确定要提取的主因子个数，而LDA的主题个数可以通过实验来检验，最后通过比较不同模型的对数似然值来确定。Log似然性和perplexity都是评价LDA模型有效性的指标。当对数似然值越大，复杂度越低时，认为模型越好。

神经网络是一种模拟生物大脑神经元工作模式的监督学习方法。与传统的回归模型相比，神经网络往往能达到更高的精度。神经网络不需要预先设置线性、逻辑回归等回归函数的形式来表示某种线性或非线性关系。相反，该模型简单地设置了大量神经元来表示自变量和因变量之间的所有复杂关系，大大简化了工作。但缺点是复杂的关系像黑盒子一样通过模型的训练过程自动计算出来。经过训练过程，人类最终得到的只是各种参数，难以像回归模型中的参数那样用来解释自变量和因变量之间的关系：

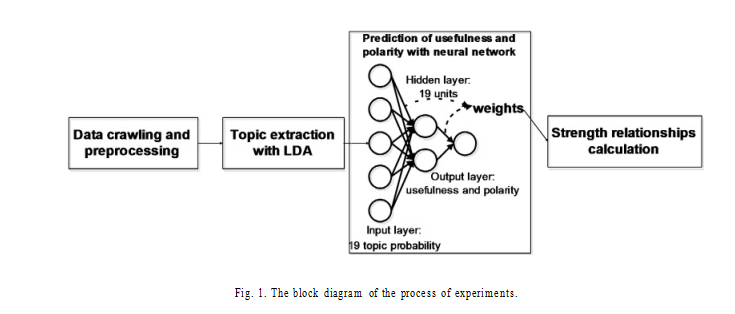


其中，wki表示隐含层第i个输入单元与第k个单元之间的权值;wjk表示隐含层JTH输出单元与KTH单元之间的权值;RSji意味着力量第i输入单元(自变量)与jth输出单元(因变量)之间的关系。

已有的评论有用性预测研究表明，神经网络模型的预测效果优于多元线性回归方法。但是实验选择的输入变量是20个与商品特征、评论者特征和评论深度相关的因素，而不是评论文本中的话题因素。因此，本文也采用了神经网络的方法，主要关注话题因子与评论有用性的关系以及评论的极性。

3.方法

下图是该方法的框图和实验过程。以下部分将详细说明该方法。



**3.1．数据源和预处理**

本文的原始语料库是在美国亚马逊网站上收集的商品评论。爬虫程序是由Python在selenium包的帮助下编写的。爬行策略是收集所有关于\_Clothing, Shoes & Jewelry产品的评论，这些评论是由网站上排名前2000位的评论者发布的。2017年7月，爬取任务最终获得了1249位评论者的23220条评论。

在预处理过程中，需要对注释的内容进行清理。同时，评论的有用分数和极性分数必须进行离散化，才能作为因变量使用。

数据清理由使用nltk和wordnet包的Python程序完成。最后，还有23212条评论进入下面的分析。清洁详情如下:

使用NLTK包的相关语料库删除标点符号和数字。

使用NLTK包的停止字列表来删除注释中的停止工作。

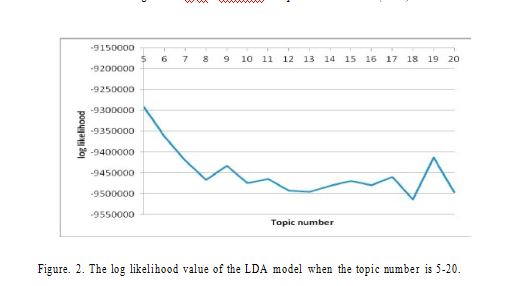
利用wordnet包中的morphy函数提取注释中单词的词干。

有用性得分是有用票数除以总票数的值。以0.6为边界，有用性评分大于等于0.6的评论被认为是高有用的评论。而极性评分大于3星的评论则为正面评论。

**3.2.基于LDA的主题抽取**

在使用LDA生成文档-主题矩阵和主题-词矩阵之前，需要准备注释语料库的主题 – 单词矩阵，由python中sklearn包的文本特征提取模块完成。然后，在LDA包中输入word - document矩阵，生成LDA模型。Python中的LDA包允许模型通过设置主题的数量和最大迭代次数来实现收敛。每次迭代完成时，都会有一个模型的对数似然值。当模型收敛时，对数似然值趋于稳定。最大迭代次数设置为1000次。我们发现所有的实验在1000次迭代内都达到了收敛。但是选题号的确定还需要通过多次实验来验证。为了降低题目讲解的难度，题目的数量不宜过大。同时，为了保证主题的差异化，主题的数量不宜过少。最后，设置主题号范围为5-20，通过比较不同主题号LDA模型的对数似然值来判断最佳主题数。

下图为主题号为5 ~ 20时LDA模型的对数似然值折线图。由图2可知，如果我们要求主题的个数尽可能大，那么log likelihood的指标在主题数为19时就会达到最大值。因此，选取19作为LDA模型的最优主题数，19也是神经网络输入层的单元数。



**3.3.基于神经网络的有效性和极性预测**

在LDA得到的文档-主题矩阵中，每个评论包含19个主题及其对应的主题概率。将19个主题的概率作为神经网络的输入变量，神经网络是一个三层结构，有一个隐含层。与输入层相似，第二层和隐藏层的单元数也设置为19。输出层只包含一个单元，它代表注释的有用性或极性。

利用sklearn包中的MLP Classifier模块进行神经网络计算。在模型评价过程中，以f1分作为评价指标，采用k次交叉验证消除偶发情况，评价分类器的平均性能。

本文的最终目标是找到所有与评论的有用性和极性相关的主题。因此，在对神经网络进行训练，得到相邻层之间的连接权值后，使用(1)分别计算19个主题与评论的有用性和极性之间的强度关系。

4.结果与讨论

**4.1.注释主题的解释**

在LDA模型中，每个主题的意义可以用主题-词矩阵来解释。当LDA的主题号设置为19时，每个主题的前30个单词及其说明如表1所示:

在本节中，我们将根据表1中每个主题前30个单词来解释主题的一般含义。

很明显，话题1、4、7、8、10、11、12都与某一类产品的材料、功能和设计有关。主题1:珠宝饰品的材料、宝石及包装;主题4主要介绍内衣、夹克、雨衣的防水保暖功能;话题7是关于连衣裙的面料以及胸腰位置的设计。题目8是关于衬衫和毛衣领子和袖子的设计;Topic10和11分别与鞋子和包的材质和设计有关;题目12涉及到手表的功能和设计。

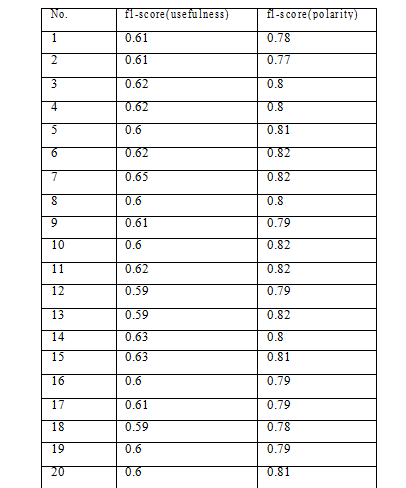
Topic 0、Topic 2、Topic 15、Topic 18与某一种产品无关，而是反映了产品的某些特性。Topic 0主要表达审稿人对产品品牌和风格的感受;题目2是对颜色的描述;Topic 15是审稿人对衣服尺寸的评论和建议;话题18与穿着舒适度有关。

第5题和第6题在意义上相似。主题5中的词汇大多与支付、包装和运输服务有关。虽然主题6也包含了与运输相关的词汇，但更多的是关于价格和质量的词汇。所以题目6代表产品的性价比。主题3、9、13、14所表达的内容与产品无关。Topic 3中形容词较多，是审稿人的主观情感表达;Topic 9包含了很多hope、seem、think、expect可能的同义词，表达了审稿人对产品的期待和意见;话题13包含两种说话方式。一种是从其他评论者那里表达自己对某些产品的不同感受。二是为其他买家提供信息;话题14出现了很多关于家庭的词汇。评论者通常会表达他们对家人所穿产品的感受，或者他们的家人是否喜欢他们收到的产品。很难判断16和17谈论的是什么话题，因为这两个话题中的词很少表达具体的意思。但是话题16包含了一些与时间有关的词汇，而话题17包含了一些与数量有关的词汇。

|  |  |  |  |
| --- | --- | --- | --- |
| 不。 | 前30个单词 |  | 解释 |
| 0 | 喜欢品牌合身穿使短裤看起来很棒好一个不同的买好尺寸得到一双多线真的也觉得去标签说甚至风格爱内衣要一点 | 对品牌和风格的感受。 | |
| 1 | 美丽的礼物项链戒指块石盒子珠宝手镯链银黄金吊坠脸可爱华丽优雅的钻石透明眼镜扣魅力水晶闪耀太阳镜闪亮来金属集完成 | 珠宝配饰。 | |
| 2 | 颜色颜色腰带蓝色胸罩黑色帽子图片领带红色深粉红色的头明亮的白色衣服匹配女孩不同的围巾可爱的棕色婴儿扣绿色可调节的孩子头发灰色的照片 | 颜色。 |  |
| 3. | 真的很好，使外观设计很好，风格有趣，加上不同的，一起惊奇，似乎细节风扇独特的五而奇妙的图表，接近酷昂贵的超级猫，别致的触感给经典 | 情感的表达。 | |
| 4 | 裤子保暖夹克就像御寒冬天轻重量水风格厚外套天气厚男人凉爽热手套落层睡袍住睡衣雨里衣领天 | 保暖功能。 | |
| 5 | 说不能知道买肯定的人告诉妻子让到真正看到钱问必须巨大的卖家花交易喜欢远没有什么课程叫天包意思朋友想法坏 | 运输和支付服务。 | |
| 6 | 伟大的好质量价格看推荐看非常便宜优秀的绝对值得服装期待高准确快速运输绝对惊喜令人敬畏支付体面的价值持久印象交易派对到万圣节 | 产品性价比的感觉。 | |
| 7 | 连衣裙顶部腰部长度拉伸看织物裙子显示身体高中等平坦臀部性感套装区域胸围英寸材料形状切割美丽的风格背部皮肤胸部膝盖缝短 | 服装各部分的面料和设计。 | |
| 8 | 衬衫柔软的材料长洗喜欢顶部织物袖子宽松短干棉中按钮打底裤长度薄毛衣挂休闲洗打印脖子低黑色收缩颜色坦克夏天 | 衬衫各部分的面料和设计。 | |
| 9 | 好像有点漂亮，想得多了，似乎想得少了，期待可能会更好，其实希望希望的方式很薄，当然，也许我认为几乎有点结束，生病的更大的一种可能 | 审稿人的期望和观点。 | |
| 10 | 鞋子脚靴子双鞋宽跟脚趾支持凉鞋走走皮革脚踝窄鞋底拱拖鞋跑步舒适宽度蕾丝好滑半看运动鞋翻盖试带 | 鞋子各部分的材料和设计。 | |
| 11 | 包口袋拉链背带皮革内钱包持有使用侧携带开放钱包前小背包肩保持两个东西空间足够处理卡旅行车厢背包手机项目手提箱 | 钱包包的面料和功能设计。 | |
| 12 | 手表耳环、表带、光集易看要看多背格、手腕保持却给简单的手、秒针黑色可以轻易地来取走方式比较显剪 | 手表的功能与设计。 | |
| 13 | 评论购买产品收到亚马逊看到请给星总是快乐的帮助项目知道提供找到图片免费归还照片诚实的意见有用失望更新然而满基是的 | 强调他们有  买了产品  根据别人的  评论,或提供  帮助  别人。 | |
| 14 | 爱买完美的小可爱的超级大女儿要年丈夫快乐月儿子也孩子圣诞节还高兴已经秋天学校赞美担心玩完美兴奋等年龄 | 参考家人对产品的意见或强调产品是为家人购买的。 | |
| 15 | 尺寸适合订购小号大号紧身run牛仔裤完美大号尺寸真正的小女人通常返回xl额外的臀部舒适我们订购紧身通常宽松，但建议牛仔布一半最少 | 对服装尺寸的意见和建议。 | |
| 16 | 即使第一次回来，得到时间，最后一次，再花几年时间去做最好的事情 | 关于时间的东西。 | |
| 17 | 一得到使用也需要来很多很多两个找到看更好喜欢的易新难带店三件东西一得到持有理由容易使开始站得特别干净 | 一些关于数量 | |
| 18 | 穿舒适感觉穿在工作袜子周围足够的日子对厚对细舒适通常每一个高极底的房子时尚停留需要夏天真正发现经常舒适问题讲究 | 穿着舒适。 | |
|  |  |  | |

**4．2．评论有用性和极性的预测结果**

下表显示了对评论有用性和极性预测的评价。20次交叉验证，采用f1-score指标评价两个模型的平均绩效。



从表中可以看出，利用话题因素很难预测评论的有用性和极性，因为两个模型的f1-score分别小于0.65和0.82。这可能是因为话题并不是决定评论有用性和极性的唯一因素，文章中没有考虑到的其他因素，如评论的深度、产品的特性等，也会影响评论的有用性和极性。为了获得更高的预测性能，在接下来的研究中，最好同时考虑主题和非主题因素。

**4.3.主题和评论之间的强度关系有用性和极性**

虽然神经网络的性能还不够好，但我们仍然可以利用模型的参数计算强度关系，以发现不同主题对因变量的贡献。

如下表所示，如果根据主题与因变量的强度关系对主题进行排序，则与评论有用性最相关的主题为:主题7、15、13、2和9。关于服装的话题与实用性最相关的原因可能是因为顾客特别喜欢买衣服重视美观和设计，因此评论经常被编辑得很好，成为其他女孩的重要参考信息;同时，必须仔细核实衣服的尺寸，因为如果衣服不合适，买家将面临退货，这是一件麻烦的事情;另外，对颜色的评价也会对产品的实用性产生影响，可能是因为网站上的图片与实际产品存在色差;审稿人对产品的意见和积极提供的信息对其他人也很重要。

与评论极性最相关的话题是:主题14、6、3、1和18，表明影响评论者对所购买产品满意度的主要因素包括家庭对产品的感受、产品的性价比、情感的表达、珠宝饰品和服装的舒适度。

与Lee和Choeh7的研究相比，本文的结果略有不同。Lee and Choeh提出，对技术方面、核心功能和美学方面的积极评论主要影响有用性，而本文的结论是服装的设计、衣服的尺寸和颜色与评论有用性相关。此外，本文也认为审稿人提供的信息是重要的，而比较研究没有考虑到这一方面。最后，Lee and Choeh认为负面评论中的服务也是有用性的主要因素，而本文的研究发现服务并不是影响评论有用性的首要因素。



Lee和Choeh的研究和本文选择的产品类别不同，可能导致了两项研究的分歧。他们选择了对厨房用具的评价作为主题。由于产品数量大，价格高，客户需要考虑保修问题，更注重运输和支付服务。而本文所选择的服装、鞋子、首饰配件等产品大多都是小而不贵的。所以客户通常不需要担心运输损坏和保修问题。同时，服装和鞋子的尺寸是这类产品最重要的特征，就像厨房电器的技术方面一样。所以衣服的尺寸对评论的有用性有很大的影响是合理的。

1. 结论

本文采用LDA方法从商品评论语料库中提取主题信息。对19个主题进行解释并输入到神经网络中，以预测评论的有用性和极性。最后利用神经网络的参数计算话题与评论的有用性和极性之间的相关性。在此过程中，产生了五个对评论有用性和极性影响最大的主题。这项研究将帮助卖家了解客户在购买某一类产品时关心什么，并告诉评论者如何编辑他们的评论，以便为买卖双方提供更多有用的信息。

在分析评论有用性的影响因素时，本文选择了一种不同于已有研究的新产品类别。通过对两项研究的比较，发现了一些联系和区别，丰富了已有的结论。然而，神经网络模型对于评论有用性和极性的预测并没有取得很好的效果，这可能是因为话题因素并不是影响评论有用性和极性的唯一因素。目前许多成功的评论有用性研究都将非话题因素作为主要研究对象，这意味着未来的研究可能会关注如何将话题因素与非话题因素结合起来，以提高分类器的性能。此外，应用文本挖掘方法对评论有用性进行分析的研究还不够多。因此，未来需要找到更有效的方法来揭示话题与评论有用性之间的相关性。