在线评论的有用性预测方法

摘要：

分析、预测在线评论的有用性，有利于提高在线评论信息质量、辅助消费者进行购买决策和促进在线平台的优化设计具有十分重要的意义，现有工作的特征选择大多数采用语法特征（计数标点、虚词、词性标记）、语义特征（评价元组、情感极性）、体裁特征（属性或写作风格的标记）、元数据特征（依靠文本以外的数据，如评论的得票数、发表时间、评论星级）。现有工作缺点，对文本内容挖掘深度不够，情感要素分析未具体到方面级，评论效用的有用性影响因素解释性不强，未考虑专业评论者（刷单，返好评的专业写手）对评论有用性的影响，本文针对这些缺点，提出对评论文本进行方面级的提取，分析方面级情感。然后基于异常检测的原理找到“专业评论者”所侧重的方面级，对不同方面依据重要性进行排序。方面级的重要性用作对评论有用性预测的重要特征，可以增强评论文本有用性分析的可解释性。

1.引言：

电子商务的快速发展使网络购物已经成为消费者购物的一种重要方式。在线评论是消费者获取信息、做出购买决策参考的重要来源，而在信息分享阶段，在线评论又是消费者创造的信息。Mudambi 和 Schuff（2010）认为，在线评论是消费者在电子商务平台或第三方评论等网站上发布的为其他消费者提供参考的关于产品的评估信息。国际著名市场研究公司 Jupiter Research 调查显示，超过 75% 的消费者会将在线评论作为网络购买决策的重要参考依据。但随着在线评论数量的急剧增长，在线评论质量变得参差不齐，在海量的在线评论中，消费者很难发现有用的在线评论，企业管理者也很难发现能够提供有关产品建议的有用的在线评论，因此在线评论有用性吸引了众多学者的关注，成为在线评论领域的研究热点。

2.相关工作：

2.1 评论有用性预测

现有的工作以预测评论有用性，辨别高质量评论为目标，主要关注评论内容，内容特征主要包括，语法特征，语义特征，体裁特征，还有另外考虑元数据特征（例如该产品参数，评论点赞数，评论者综合信誉等）。用到的技术有文本数据挖掘和自然语义处理，文本情感分析等。由于特征较多且复杂，建模方法多使用机器学习或深度学习的方法，例如支持向量机（Support Vector Machine，SVM ）,随机森林（Random forest，RanF）,卷积神经网络（Convolutional Neural Networks, CNN）,循环神经网络（Recurrent Neural Network, RNN）,长短期记忆网络（Long Short-Term Memory，LSTM）等。

现有工作缺点，对文本内容挖掘深度不够，情感要素分析未具体到方面级，评论效用的有用性影响因素解释性不强，未考虑专业评论者（刷单，返好评的专业写手）对评论有用性的影响，本文针对这些缺点，提出对评论文本进行方面级的提取，分析方面级情感。然后基于异常检测的原理找到专业评论者所侧重的方面级，对不同方面依据重要性进行排序。方面级的重要性用作对评论有用性预测的重要特征，可以增强评论文本有用性分析的可解释性。

2.2方面级情感分析（实体方面、观点、情感）

文本情感分析可以按分析粒度分为篇章级情感分析、句子级情感分析和方面级情感分析。早期阶段，篇章级和句子级情感分析任务是研究的重点。它们假设一段文本只有一种情感，并对给定文本进行分析并判断其整体的情感极性是正面、负面或者中性等。篇章级和句子级情感分析的方法仅仅是处理对象文本的长度的差异。目前，针对文本整体进行情感分析的研究技术已经相对成熟，达到了较好的性能。然而，对文本进行整体情感分析会掩盖其细节，且整体情感并不能反映人们对意见目标的、细粒度的情感表达。如果只关注整体情感而忽略具体细节，可能会在推荐系统、问答系统等现实应用中计算出有误的结果。因此，为了进行更完整的情感分析，系统需要发现文本评论的各个方面对象，并确定文本针对每个方面所表达的情感信息，这就是方面级情感分析技术。

2.3 VAE模型在异常检测中的应用

这是利用变分自动编码器重构概率的异常检测方法。重建概率是一种考虑变量分布变异性的概率度量。重建概率具有一定的理论背景，使其比重建误差更具有原则性和客观性，而重建误差是自动编码器（AE）和基于主成分（PCA）的异常检测方法所采用的。利用变分自动编码器的生成特性，可以推导出数据重构，分析异常的根本原因。

3.模型构建

本文利用异常检测的原理利用文本评论和用户行为提出了一种半监督的方面级评论有用性评价方法，通过用户在评论中表达对商品或服务不同方面的看法，根据评论文本的可读性、信息量和用户行为特征分类出高质量评论和低质量评论。对高质量评论进行方面级提取，并对方面级的评论进行情感极性分析，使用条件变分自编码器（Conditional Variational Auto-Encoder，CVAE）学习正常用户方面级情感的数据分布，以检测出异常用户的数据分布。异常用户会对产品方面级进行极大的夸赞或者贬低，使用注意力机制（Attention mechanism）将异常用户所关注的方面级提取出来并做方面级的重要性排序。依据学习到的方面级排序，可以对新的评论的有用性做出预测，如果新的评论的在异常用户所关注的方面级给出极高或极低评价，那么可以预测该评论是一条高质量的异常评论，有疑似雇佣写手、专业评论的嫌疑，可以让该评论的有用性得分降低。对于低质量文本评论，同样依据方面级的情感极性分析、用户行为进行异常检测，分类出，低质量的正常评论和低质量的异常评论，低质量的正常评论代表着普通大众的评价，可以作为商品的常态评价显示给消费者，低质量的异常用户，可以隐藏或删除。

3.1高低质量评论分类

对于评论R=[], w代表词，方面序列A也是由w构成，

A=[],其中方面序列可能包含一个或多个词。考虑距离方面词中心为k个词的序列片段L。当方面词为一个单词时，L=[],其中为方面词，当方面词为多个单词时，L=[],其中为方面词的第一个词，为最后一个词。

计算方面词的个数，再考虑用户的历史评论数量，给出综合的高低质量评分。

通过如下公式获取所述目标商品所对应的多个评论中每个评论的质量评分：

；

其中，为所述目标商品所对应的多个评论中任意一个评论i的质量评分，为所述评论i中方面词的个数，为的最大值，为所述多个评论中每个评论对应的用户的历史所发评论数量，为的最大值，为所述Na的比例系数，为所述Nn的比例系数。

3.2方面级情感分析（对文本进行 目标、方面、观点、情感的提取）

采用长短期记忆网络（Long Short-Term Memory，LSTM）模型进行方面级情感分析，该LSTM模型的输入是训练数据集中的评论文本对应的词向量和方面表示向量，首先将词向量和方面表示向量进行串联，得到串联数据，之后将串联数据输入LSTM模型进行上下文编码，LSTM相比循环神经网络（Recurrent Neural Network， RNN）解决了长期依赖的问题，将词向量和方面表示向量进行串联的是在编码上下文信息时可以把方面信息考虑进去，具体公式如下：



其中，为训练数据集的评论文本中第i个词所对应的词向量，为训练数据集的评论文本中的方面词所对应的方面表示向量，是方面情感分析模型的隐藏层的输出数据。

3.3基于异常检测的有用性预测

模块构建条件变分自编码器（Conditional Variational Auto-Encoder，CVAE）模型进行异常检测。异常检测任务是一个半监督框架，仅使用正常实例的数据来训练CVAE，重建概率是一种考虑变量分布变异性的概率度量。利用变分自动编码器的生成特性，可以推导出数据重构，分析异常的根本原因。CVAE是附带条件的变分自编码器，使用的条件是评论者行为。

由基于质量的评论分类模块和方便级情感分析模块，得到评论文本，评论文本方面级情感。将正常评论者的评论文本和方面级情感串联得到数据X=[]。数据Y为对应的评论者的行为，具体包括历史发布评论数量、有用性投票得分、品牌偏移得分、评级偏差、评论者日最多评论量,Y=[]。

假设评论者行为信息y与隐变量z没有直接关系，因此，

条件概率：

变分下界：

首先使用正常评论者数据和对应的评论者行为数据，训练CVAE得到、，、分别是p和q的概率分布参数。

CVAE的构成包括编码和解码阶段，使用包含异常评论者的数据集,和对应的评论者行为数据，概率编码器和解码器分别对潜在变量空间和原始输入变量空间中各向同性正态分布进行参数化。

在测试中，对于一个样本)，从训练的概率编码器中抽取L个样本，对于来自编码器的每一个样本，概率解码器输出均值和方差参数。使用这些参数，计算从分布产生原始数据的概率，平均概率用做重构概率。

计算的公式为：

Reconstruction probability(i) =

这里的 是在概率解码器解码L个样本得到的。若重构概率小于阈值α则数据)则为异常数据为非有用性评论。

4.实验

5.分析

6.总结

附录：方面级情感分析

定义1：观点（Opinion）：观点对象包含五个元素：（表达者，实体，方面，时间，情感）。其中，表达者是观点表达的主体；实体是主体评价的客体；方面是实体的一个方面；情感一般包括正面、负面和中立等；时间是发表观点的时间。

定义2：实体（Entity）：实体是评论的目标，这个目标可以是产品、服务、话题、时间、人物、组织或事件等。方面级情感分析研究中的每个观点都有一个目标实体，可以是被评价的实体本身，也可以是实体的一部分、一个模块或相关实体。

定义3：方面（Aspect）：方面是评论的对象实体属性（如在餐厅评论中，一个方面可以是食品的价格，质量等）。方面是一个较高层次的概念，相对于评论文本包含的词汇，方面概念集合规模很小。

  早期研究将评论中的方面信息划分为两类：显式方面与隐式方面。显式方面是指在评论文本中显式出现的评价对象方面表述，隐式方面则是暗含在评论文本语义中的评论对象。2014年，SemEval将方面级情感分析任务的方面对象分为方面术语和方面类别。方面术语是显式方面表述，方面类别则为方面表述不显示出现在评论文本的情况提供了获取方面信息的解决思路。SemEval提供的基准数据集，由人工标注了评论文本中的方面术语、方面类别和情感极性信息，相关定义如下。

定义4：方面术语（Aspect Terms）：方面术语是出现在评论文本中的词语序列，是评论对象实体方面的表述信息，通常有较大的词汇量。方面术语可以是一个词也可以由多个词连接组成。

定义5：方面类别（Aspect Category）：方面类别由评价对象实体及其属性组合而成。

定义6：情感极性（Sentiment Polarity）：情感极性是指评论文本中表达的正面(positive)、负面(negative)、中性(neutral)或者冲突(conflict)的情感。其中，冲突(conflict)是指评论文本对评价对象既表示了正面情感又表达了负面情感。

评论有用性研究

研究的必要性：在线评论多而且影响用户决策。

用户依赖：

据报道，大量的消费者的购买行为依赖在线评论

Murphy, R., 2018. Local Consumer Review Survey 2018, retrieved June 20, 2020, from https://www.brightlocal.com/research/local-consumer-review-survey/.

评论的量多：

目前，在Yelp上发表的评论超过2.14亿篇，并且这一数字以每年12%的速度增长（Yelp，2020年）。

造成的问题：

信息过载

成倍增长的在线评论已经造成了信息过载的问题（Peter 2019）（Han-fen Hu, Anjala S. Krishen 2019）.

Peter Gordon Roetzel. 2019. Information overload in the information age: a review of the literature from business administration, business psychology, and related disciplines with a bibliometric approach and framework development. Business Research (2019), 479-522. DOI 10.1007/s40685-018-0069-z

Han-fen Hu, Anjala S. Krishen,When is enough, enough? Investigating product reviews and information overload from a consumer empowerment perspective,Journal of Business Research,Volume 100,2019,Pages 27-37,ISSN 0148-2963.

在线评论分析：

在线评论中阅读者所给出的Review helpfulness是信息质量的关键因素

Baseline：

作为一个回归问题

文章：BERT Feature Based Model for Predicting the Helpfulness Scores of Online Customers Reviews2020

考虑：比例

数据集：Amason

方法：SVR, M5P, N.Net

评价指标:MSE

作为一个分类任务：

文章：An attention model of customer expectation to improve review helpfulness prediction. In: The 42nd European Conference on Information Retrieva

考虑：比率

数据集Amazon, Yelp

方法RandF, SVM, CNN, DNN

指标AUC

一个方法的指标和评价的指标 数据集

评论有用性分析

Motivation：依据评论文本，通过（总评、时间差、方面、方面情感，情感强度）计算出评论者的写作能力，细化评论文本的有用性，作为刻化文本有用性的指标

1. 不同产品，优质评论差别大
2. 针对某产品利用得分最多的前K项（利用总评、时间差、方面、方面情感、情感强度、评论者写作能力）
3. 计算评论与前k项评论的相似度作为评论者的写作能力（细化评论的有用性）