

基于用户和产品表示的情感分析和评论质量检测联合模型*

吴 璠, 王中卿, 周夏冰, 周国栋



(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

通讯作者: 王中卿, E-mail: wangzq@suda.edu.cn

摘 要: 情感分析旨在判断文本的情感倾向,而评论质量检测旨在判断评论的质量.情感分析和评论质量检测是情感分析中两个关键的任务,这两个任务受多种因素的影响而密切相关,同一个产品的情感倾向具有相似的情感极性;同时,同一个用户发表的评论质量也具有一定的相似性.因此,为了更好地研究情感分类和评论质量检测任务的相关性以及用户信息和产品信息分别对情感分类和评论质量检测的影响,提出了一个情感分析和评论质量检测联合模型.首先,使用深度学习方法学习评论的文本信息作为联系两个任务的基础;然后,将用户评论及产品评论作为用户的表示和产品的表示;在此基础上,采用用户注意力机制对用户的表示进行编码,采用产品注意力机制对产品的表示进行编码;最后,将用户表示和产品表示结合起来进行情感分析和评论质量检测.通过在 Yelp2013 和 Yelp2015 数据集上的实验结果表明,该模型与现有的神经网络模型相比,能够有效地提高情感分析和在线评论质量检测的性能.

关键词: 情感分析;评论质量;用户表示;产品表示;联合模型;注意力机制

中图法分类号: TP18

中文引用格式: 吴璠,王中卿,周夏冰,周国栋.基于用户和产品表示的情感分析和评论质量检测联合模型.软件学报,2020,31(8):2492-2507. <http://www.jos.org.cn/1000-9825/5895.htm>

英文引用格式: Wu F, Wang ZQ, Zhou XB, Zhou GD. Joint model for sentiment analysis and review quality detection with user and product representations. Ruan Jian Xue Bao/Journal of Software, 2020,31(8):2492-2507 (in Chinese). <http://www.jos.org.cn/1000-9825/5895.htm>

Joint Model for Sentiment Analysis and Review Quality Detection with User and Product Representations

WU Fan, WANG Zhong-Qing, ZHOU Xia-Bing, ZHOU Guo-Dong

(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: Sentiment analysis aims to judge the emotional tendency of the text, while the review quality prediction aims at judging the quality of the review. Sentiment analysis and review quality detection are two key tasks in sentiment analysis, these two tasks are closely related by many factors, the reviews on the same product have similar opinion polarity, and the quality of reviews from the same user tend to be similar. Therefore, this study proposes a joint neural model to learn sentiment analysis and quality prediction in order to better study the correlation between sentiment classification and review quality prediction tasks and the impact of user information and product information on sentiment classification and review quality prediction respectively. First of all, this study employs a deep representation learning approach to learn textual information of reviews, serving as the basis to connect two tasks, and then uses the user reviews and product reviews as the representation of the user and the representation of the product, on the basis, a user attention is adopted to encode user information in user representation, and a product attention is used to encode product information in product representation, and finally both user and product representations are jointly integrated for sentiment analysis and quality prediction with attention mechanism.

* 基金项目: 国家自然科学基金(61806137, 61702518, 61836007, 61702149); 江苏省高等学校自然科学研究基金(61702149)

Foundation item: National Natural Science Foundation of China (61806137, 61702518, 61836007, 61702149); Natural Science Foundation of the Higher Education Institutions of Jiangsu Province, China (61702149)

收稿时间: 2019-01-25; 修改时间: 2019-05-06; 采用时间: 2019-07-29

The experimental results on the Yelp2013 and Yelp2015 datasets show that the proposed model can effectively improve the performance of sentiment analysis and online review quality prediction compared with the existing neural network models.

Key words: sentiment analysis; review quality; user representation; product representation; joint model; attention mechanism

情感分析旨在判断在线评论的情感倾向,挖掘出用户对该产品的情感态度,分析用户的产品喜好;而评论质量检测则是研究在线评论的质量,衡量在线评论对消费者及商家的参考价值.近年来,随着亚马逊、Yelp、IMDB等在线评论网站的迅速发展,情感分析受到了研究人员和行业的广泛关注.然而,情感在评论中可能受到多种因素的影响,例如评论的对象、作者的相关属性、评论的质量等,其中,评论的质量对情感的分析有着较大的影响.一般来说,低质量的评论倾向于直接表达对目标的情感,如例 1 所示.

例 1: The service was great. The food was really good. I can't wait to bring my mom here!(服务很棒,食物真的很好吃,我等不及要带我妈来了!)

例 1 这条评论直接表达该商家的服务很好,食物很好吃.但是不同的消费者有自己独特的口味爱好和个人需求,因此该条评论对消费者的参考价值很有限.而高质量的评论通常在表达观点时更加复杂,一般会仔细分析商家的服务态度、地理位置、食物口味等特征,针对目标商家做出不同方面的分析.如例 2 所示,该条评论对该商店的地理位置、停车位、商店的卫生状况进行了全面的评价,消费者在做出选择时就可以根据自己的需求和喜好进行选择,该条评论对潜在消费者的参考价值就比较大.因此,将情感分析和评论质量检测任务一起研究具有一定的意义.本文提出了一个联合模型,可以同时从评论中获取用户的情感倾向和评论的质量信息,对于每一条评论,既给出了该条评论的评论质量,也给出了该条评论的情感倾向.

例 2: Good location. Parking lot is often full and parking is limited, fresh coffee. Store typically clean and fairly busy too. (位置很好,就是停车位有限,咖啡新鲜,商店很繁忙但还是干净整洁.)

情感分类和评论质量检测任务一直是自然语言处理领域的研究重点,国内外有关的研究成果如下:例如:Pang 等人^[1,2]将情感分类作为文本分类问题的一个特例,他们将用户标记的情感极性或者评分作为标签,并且使用机器学习算法构建带有文本特征的情感分类器.而现有的评论质量检测工作大多集中在选取高效的特征上面,广泛使用的特征包括日期特征^[3]、产品特征^[4]和语义特征^[5].但是以上研究都是将文本的情感倾向和评论质量独立研究的,其实在社交媒体上,针对同一产品的评论具有相似的情感极性,因为针对某一产品的情感倾向比较客观,口碑较好的产品评论大多都是正面的,口碑不好的产品评论大多是负面的;而评论的质量则更加的主观,取决于发表评论的用户,一直发表高质量评论的用户再次发表评论的质量一般来说也较高,而一直发表低质量评论的用户再次发表的评论质量较低的可能性也较大,因此,来自相同用户的评论质量也趋于相似.由以上分析可知,产品信息和用户信息对情感分类和评论质量检测任务分别有着不同的影响.

针对上述问题,本文提出了情感分析和评论质量检测联合模型,并结合注意力机制研究用户信息和产品信息分别对情感分析任务和评论质量检测任务的影响.首先,利用评论的文本信息作为联系两个任务的基础;此外,为了研究用户信息和产品信息分别对情感分析和评论质量检测的影响,本文在文本信息中进一步引入了评论的用户信息和产品信息;同时,为了放大用户信息和产品信息的作用,模型利用注意力机制分别对用户的表示和产品的表示进行编码;最后,将用户表示和产品表示结合起来进行情感分析和评论质量检测.实验表明,本文提出的情感分析和评论质量检测联合模型是有效的;同时,在情感分类和评论质量检测任务中分别加入针对用户信息和产品信息的注意力机制,以进一步提升联合模型的性能.本文的主要贡献有以下几点.

- (1) 提出了一个联合神经网络模型来同时进行情感分析和评论质量检测任务;
- (2) 采用文本的共享表示和注意力机制来共同捕获情感分析和评论质量检测两个任务的相关性;
- (3) 利用用户的表示和产品的表示来研究情感分析和评论质量检测之间的相互作用.

本文第 1 节主要介绍情感分析、评论质量检测 and 联合学习模型的相关工作.第 2 节介绍评论质量检测 and 情感分析联合模型.第 3 节介绍如何利用注意力机制将用户信息和产品信息融入联合模型来进一步提高模型的性能.第 4 节给出实验过程和实验结果的详细分析.最后,在第 5 节中总结全文,同时提出下一步的研究方向.

1 相关工作

1.1 情感分类

传统的情感分析方法主要有两大类:基于知识的方法和基于机器学习的方法.基于知识的方法主要利用句法规则、情感词典等外部资源对文档中情感词进行加权求和,得到文档的总体情感倾向.Turney^[6]在文献中提出了一种简单的无监督算法,该算法通过评论中包含的形容词或副词短语的平均语义将评论分为推荐和不推荐两类,在4个不同领域的评论数据集上实验结果分类平均准确率达到了74%.Zhang等人^[7]提出了一种基于规则的方法:首先根据情感词典得到句子的情感倾向,再根据句子的情感倾向得到整个文档的情感倾向.与以上基于知识的方法不同,使用机器学习的方法主要在于特征的提取^[8-10].Pang等人^[2]按照不同的方法提取特征,如Unigram、Bigram、POS、Position等,然后对比分析SVM(支持向量机)、ME(最大熵模型)、Naïve Bayes(朴素贝叶斯)这3种机器学习方法的实验结果,最终发现,在IMDB影评数据集上,SVM结合Unigram特征提取方法效果最好,正确率达到了82%以上.近几年,随着深度学习的发展,各种神经网络的方法也应用到了情感分类当中.Socher等人^[11,12]使用树状的长短时记忆网络来改善语义的表示,以进一步提高情感分类的准确度.王青等人^[13]提出一种基于注意力机制的双向LSTM模型,对中文商品评论进行情感分类.实验结果表明,该模型在中文商品评论二分类任务和三分类任务中均获得了较好的准确率、召回率和F1值.Yang等人^[14]提出了句子级别和文档级别的层次化注意力模型,以提高情感分类的准确率.Li等人^[15]用CNN层替代注意力机制RNN层去提取最重要的分类特征,将上下文信息和变换后的特征结合起来进行情感分类,该方法在实验中取得了全面优于已有方法的准确率和F1值.

1.2 评论质量检测

评论的质量是以该条评论对潜在消费者购物帮助的大小来衡量的,评论描述产品特征越具体,对潜在消费者购物的帮助越大,该条评论的质量就越高.在线评论质量检测研究可以看作传统的文本分类问题,为了提高评论质量检测的性能,评论质量检测的研究的重点在于选取不同的特征.Larkey^[16]利用评论的文本特征如评论平均长度、单词平均长度和不同单词的数量来构建回归模型,进行评论质量检测.Kim等人^[17]研究表明,评论发表距今的时间也会影响评论的质量.孟园等人^[18]研究发现,评论内容的信息增益、产品特征词的信息增益、评论客观情感倾向度、内容差异性对评论质量检测具有明显作用.胡学钢等人^[19]考虑评论的文本属性、评论者属性和店铺属性对评论质量的影响,构建了一个在线评论质量预测模型,在实验数据集上具有较高的准确率与召回率.除了上述特征外,评论质量检测研究广泛使用的一些特征还有日期特征^[20]、语义特征^[21]和情感词典^[22]等.最近,由于论证文本经常出现在产品评论中,Liu等人^[23]使用论证信息,例如议论的百分比和结论的比率来预测评论的质量.Ghose等人^[24]研究评论者以往在社交网站发表的评论总数、评论获得的有用率、评论者身份等特征对评论质量的影响.为了研究用户信息对评论质量检测的影响,Zhang等人^[25]提出一个根据作者的历史评论进行评论质量检测的模型.

1.3 联合学习模型

联合学习模型能同时解决多个任务,与独立模型相比,联合模型有两个主要优点:首先,它能允许不同任务之间的交互,从而通过信息共享提高准确度;其次,与传统的方法相比,它减少了一个任务到另一个任务的错误传播.因此,联合模型也被广泛运用于自然语言处理任务中,其中就包括分词和词性标注^[26,27].Shi等人^[28]提出了一个利用条件随机场(CRF)进行分词和词性标注的联合模型.Boonkwan等人^[29]提出了一个简单而有效的长短时记忆(LSTM)神经网络联合模型来解决分词和词性标注问题,在5个不同数据集上的实验结果表明,该模型具有很强的竞争力.同时,联合模型在联合解析和命名实体识别上也有应用.Finkelt等人^[30]提出一种新的模型,它利用额外的单任务注释数据来提高联合模型的性能,他们的模型在连接多个单任务模型和联合模型之前共享特征权重时使用了层次结构.实验表明,层次联合模型比只对联合标注数据进行训练的联合模型效果有所提高.在联合解析和语义角色标注上面也使用了联合模型,Yang等人^[31]提出了一个联合模型,该模型利用预测器中的

知识来优化意见抽取子任务.实验结果表明,联合方法在单独处理子任务意见抽取问题上明显要优于传统方法.

近年来,对联合模型的研究热度有增无减.Li 等人^[32]提出了一种神经网络联合模型来同时提取生物医学实体及其关系.Kurita 等人^[33]提出了基于神经网络的中文分词、词性标注和依赖解析的联合模型.Wang 等人^[34]提出了基于神经网络的实体抽取和关系抽取联合模型.许银洁等人^[35]提出了一个主题情感联合模型(JST),该模型能够同时识别文本中的主题和情感.李大宇等人^[36]面向电影评论数据提出了一个标签情感联合模型,该模型可以同时挖掘出电影评论数据领域的标签以及该条评论的情感倾向.

与以往情感分析和评论质量检测的研究不同,以往的研究都是将情感分析和评论质量检测视为两个独立的任务,而本文首次提出了一个联合模型来同时学习这两个任务;此外,还利用注意力机制研究用户表示和产品表示分别对这两个任务的影响.

2 情感分析和评论质量检测联合模型

在本节中,我们初步提出了一个同时进行情感分类和评论质量检测的联合模型,该模型使用 LSTM 模型来学习情感分类和评论质量检测的共享文本表示.

基于联合深度学习模型的情感分析和评论质量检测整体架构如图 1 所示,在联合模型中,我们利用长短时记忆网络(LSTM)学习评论的文本信息来作为两个任务的共享表示,基于文本信息的共享表示,使用 softmax 层进行回归分析,经过 softmax 层得到每一类对应的概率,取概率最大的一类作为最终的情感分析和评论质量检测的结果.

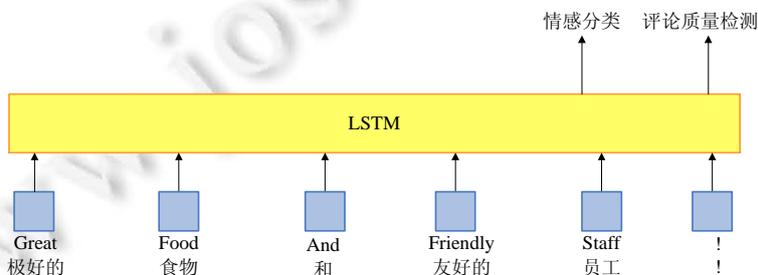


Fig.1 Joint model framework

图 1 联合模型框架

2.1 共享文本信息的表示

原始输入可以将每条评论表示为由 $\{W_1, W_2, W_3, \dots, W_i\}$ 组成的一个词序列,首先,将评论中的单个词 W_i 转换为实值向量 X_i ,得到一个矩阵: $M=(X_1, X_2, X_3, \dots, X_i)$;然后,我们使用 LSTM 模型来学习评论的文本信息.相比传统的 RNN 模型,LSTM 通过引入多种门,使其能够长时间的保留重要的信息,从而在某种程度上缓解了在传统 RNN 中出现的梯度消失问题.LSTM 神经网络将定长的矩阵 M 进行训练,得到隐藏向量序列 $\{h_1, h_2, h_3, \dots, h_i\}$.在 LSTM 网络中,通过当前时刻的向量 X_t 和先前时刻的向量 H_{t-1} 得到隐藏向量 $h_t=LSTM(H_{t-1}, X_t)$.LSTM 神经网络在递归神经网络(RNN)的基础上增加了门控机制,添加了输入门(input gate)、细胞门(cell gate)、遗忘门(forget gate)和输出门(output gate).数据在 LSTM 神经网络中经过这些门,来筛选需要保留的信息和遗弃不需要的信息.

$$i_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}) \tag{1}$$

$$f_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}) \tag{2}$$

$$o_t = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1} + b^{(o)}) \tag{3}$$

$$u_t = \tanh(W^{(u)}x_t + U^{(u)}h_{t-1} + b^{(u)}) \tag{4}$$

$$C_t = i_t \cdot u_t + f_t \cdot c_{t-1} \tag{5}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{6}$$

其中, σ 为激活函数, $\text{sigmoid}, \tanh$ 表示双曲正切激活函数, i, f, o, c 和 C_t 分别表示 t 时刻对应的输入门、遗忘门、输出门和细胞门 4 种门结构, x_t 是在时刻 t 时的输入词向量. 本文使用 $H=h_n$ 作为情感分类和评论质量检测两个任务的共享文本表示 X , 在大多数自然语言处理任务中, LSTM 确实展示了很好的性能.

2.2 情感分析

情感分析的主要目的是挖掘评论中蕴含的感情色彩, 即分析文本的情感倾向于正面还是负面. 本文将情感分析看作是一个回归问题, 情感分析旨在预测评论文本的情感标签, 其目的是为了找出评论者或作者在某些话题上或针对某一产品的观点态度. 通过情感分析, 可以缓解社交网络上的评论信息爆炸问题, 使得用户在较短的时间内获得针对某一产品的大众评论情感倾向, 为消费者购买商品提供决策支持. 在实验中, 我们使用多层感知器训练二进制分类器来预测给定评论的极性, 将评论的文本信息表示为输入向量 H , 使用 LSTM 的隐藏层来学习更高层次的特征. 由隐藏层生成的 H_p 作为 softmax 输出层的输入, 最后经过 softmax 层预测情感极性标签 P_p .

$$H_p = \sigma(w_p^h H + b_p^h) \quad (7)$$

$$P_p = \text{softmax}(w_p, H_p + B_p) \quad (8)$$

其中, w_p^h, b_p^h, w_p, B_p 均为实验模型的参数.

2.3 评论质量检测

在情感分析研究的早期阶段, 人们大多将研究重心放在自然语言处理和数据挖掘技术的观点抽取问题上. 但是一个不容忽视的问题就是, 垃圾评论、虚假评论等质量低下的评论给研究工作带来了极大的阻碍. 近年来, 随着网络平台的快速崛起, 人们交流意愿的增强, 网络评论的数量迅速增长, 但评论良莠不齐, 评论质量得不到保证, 快速筛选有价值的评论信息较难. 因此, 如何在大量的评论当中检测出评论的质量, 是情感分析研究的前提. 评论质量检测可以看作是一项回归任务, 在实验中, 我们使用多层感知器训练二进制分类器来预测给定评论的评论质量, 将评论的文本信息表示为输入向量 H , 使用 LSTM 的隐藏层来学习更高层次的特征. 由隐藏层生成的 H_R 作为 softmax 输出层的输入, 最后经过 softmax 层预测评论的质量标签 P_R .

$$H_R = \sigma(w_r^h H + b_r^h) \quad (9)$$

$$P_r = \text{softmax}(w_r, H_r + B_r) \quad (10)$$

其中, w_r^h, b_r^h, w_r, B_r 均为实验模型的参数.

2.4 模型训练

对于情感分析和评论质量检测联合模型的训练, 我们的训练目标是 minimized 均方误差损失, 联合模型的联合损失函数如下所示.

$$L(\theta) = \sum_{i=1}^N (P_{p_i} - y_{p_i})^2 + \sum_{i=1}^N (P_{r_i} - y_{r_i})^2 + \frac{\tau}{2} \|\theta\|^2 \quad (11)$$

其中, P_p 和 P_R 分别是模型预测的情感极性和评论质量结果, y_p, y_R 是评论原本的情感和质量标签, θ 是模型的参数集, τ 是 L2 正则化的参数. 本文使用 AdaGrad 梯度下降算法进行模型训练过程中的优化, 实验中使用 SkipGram 算法来生成词向量.

3 基于用户和产品表示的联合模型

上节中提出的利用评论文本信息的 JLSTM 联合模型虽然可以同时进行情感分析和评论质量检测任务, 但是每个任务之间是独立的, 没有考虑到两个任务之间的相互作用, 也不能研究用户和产品的表示分别对这两个任务的不同影响. 并且, 在研究过程中我们发现, 这些在线评论中有些词或者句子表现出强烈的用户偏好, 而另一些则更倾向于表达产品特征. 例如, 对于评论“Amazing food! Customer service was excellent even with it being so busy on opening day! Highly recommend Three Cheese Cauliflower soup and Asparagus soup!(令人惊叹的食物, 虽然开业当天很忙, 但顾客服务还是很好! 强烈推荐芝士花椰菜汤和芦笋汤!)”, “食物令人惊叹”和“顾客服务

很好”等描述了产品的特点,而“强烈推荐”则表现出强烈正向的用户情感倾向.显然,这两种信息对情感分析和评论质量检测任务有着不同的影响,这意味着该条评论在用户视图和产品视图有着不同的语义表示.一般来说,来自相同产品的评论趋向于具有相似的情感极性;同时,来自同一个用户的评论质量也具有一定的相似性.为了解决这一问题,我们将上节提出的联合模型进行了改进,将用户信息和产品信息考虑进来,同时进行情感分析和评论质量检测任务.

结合了用户信息和产品信息的情感分析和评论质量检测联合模型如图 2 所示,首先,利用与上节模型相同的 LSTM 神经网络来学习评论的共享文本信息,得到情感分析隐藏向量 H_P 和评论质量检测隐藏向量 H_R ;然后,构建产品的表示来捕获产品信息对情感分析的影响,进一步构建新的考虑产品信息的情感分析向量 V_P ;另一方面,为了捕获用户信息对评论质量检测的影响,我们同样构建了结合用户信息的用于评论质量检测的向量 V_R .在本章节中,我们将详细描述如何使用注意力机制将用户信息和产品信息结合到情感分析和评论质量检测联合模型当中去.

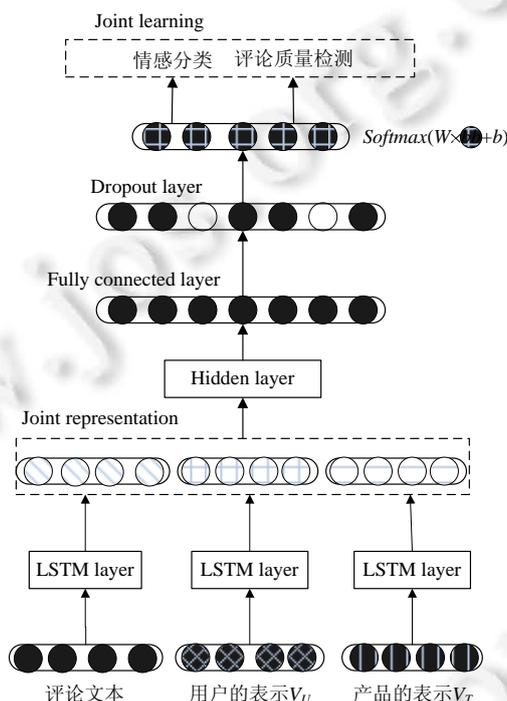


Fig.2 Joint model diagram of sentiment classification and review quality detection

图 2 情感分类和评论质量检测联合模型图

3.1 用户的表示

在实验中,考虑到用户发表评论的相似性,我们利用用户以往在社交网站上发表的评论信息,通过 LSTM 层来构建用户的表示,用户的表示图.如图 3 所示,首先,我们将某个用户抽取出的 $\{e^d(r_{u_1}), e^d(r_{u_2}), e^d(r_{u_3}), \dots, e^d(r_{u_n})\}$ 这些评论信息作为输入,经过 LSTM 神经网络 $h_{u_i} = LSTM(H_{u_{i-1}}, e^d(r_{u_i}))$,我们得到了一系列隐藏状态向量 $\{h_{u_1}, h_{u_2}, h_{u_3}, \dots, h_{u_i}\}$,这些 h_{u_i} 分别为不同用户的表示.在具体实验中,对于每个用户,我们随机抽取其发表的 6 条评论,不足 6 条评论的我们就抽取了最大评论条数,然后利用 LSTM 模型,基于这些历史评论分别构建每个用户的表示.

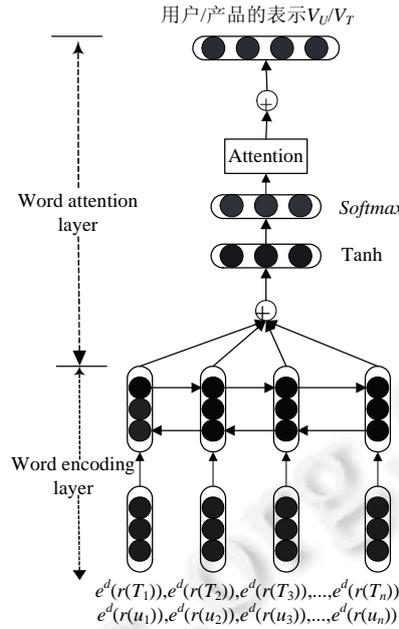


Fig.3 User/Product representation

图3 用户/产品的表示

注意力机制可以从众多信息中选择出对当前任务目标更关键的信息,在模型训练时,让神经网络高度关注某些信息.例如在情感分析任务中,注意力机制可以通过注意力特征图,让模型在训练过程中赋予与情感分类相关的形容词等词的权重,来正确判断评论的情感极性.Mnih 等人^[37]首次将注意力机制用在图像分类任务中,发掘了注意力机制在图像处理领域的优势,同时,也使注意力机制成为自然语言处理领域的研究热点.同样,在情感分类和评论质量检测任务中,每一条评论对于用户的表示重要性也是不同的.因此,我们引入了基于注意力机制的用户信息来提取相对较为重要的评论,将用户具有代表性的评论汇总起来,形成用户的表示.

将用户的表示 $\{h_{u_1}, h_{u_2}, h_{u_3}, \dots, h_{u_n}\}$ 作为输入,经过注意力机制进行编码,构建一个隐藏状态 h_{u_i} 的加权和.最后输出引入用户注意力机制的用户的表示向量 V_U 如公式(12)所示.

$$v_U = \sum_{i=1}^T \partial_i h_{u_i} \tag{12}$$

其中, T 为输入状态的数目, ∂_i 为输入状态 h_{u_i} 的权重, $\partial_i \in [0,1]$ 并且 $\sum_i \partial_i = 1$. ∂_i 的计算公式如公式(13)、公式(14)所示.

$$\partial_i = \frac{\exp(u_i)}{\sum_{j=1}^T \exp(u_j)} \tag{13}$$

$$u_i = \text{Tanh}(W_U h_{u_i} + b_U) \tag{14}$$

其中, W_U 为模型中的权重, b_U 为模型中的偏置.

3.2 产品的表示

我们将针对某一个产品抽取出的 $\{e^d(r_{u_{T_1}}), e^d(r_{u_{T_2}}), e^d(r_{T_3}), \dots, e^d(r_{T_n})\}$ 这些评论信息作为输入,经过 LSTM 神经网络 $h_{T_i} = \text{LSTM}(H_{T_{i-1}}, e^d(r_{T_i}))$, 我们得到了一系列隐藏状态向量 $\{h_{T_1}, h_{T_2}, h_{T_3}, \dots, h_{T_n}\}$, 这些 h_{T_i} 分别为不同产品的表示.在具体实验中,对于每个产品,我们随机抽取不同用户针对该产品发表的 6 条评论,不足 6 条评论的我们就抽取了最大评论条数,然后利用 LSTM 模型,基于这些历史评论分别构建每个产品的表示.产品的表示如图 3 所

示.

和用户的表示一样,每一条评论对于产品的表示重要性也是不同的,因此,在产品的表示中也引入了注意力机制.将产品的表示 $\{h_{t_1}, h_{t_2}, h_{t_3}, \dots, h_{t_n}\}$ 作为输入,经过注意力机制进行编码,构建一个隐藏状态 h_{t_i} 的加权和.最后输出引入产品注意力机制的产品的表示向量 V_T 如公式(15)所示.

$$v_T = \sum_{i=1}^T \partial_i h_{t_i} \tag{15}$$

其中, T 为输入状态的数目, ∂_i 为输入状态 h_{t_i} 的权重, $\partial_i \in [0, 1]$ 且 $\sum_i \partial_i = 1$. ∂_i 的计算公式如公式(16)、公式(17)所示.

$$\partial_i = \frac{\exp(t_i)}{\sum_{j=1}^T \exp(t_j)} \tag{16}$$

$$t_i = \text{Tanh}(W_T h_{t_i} + b_T) \tag{17}$$

其中, W_T 为模型中的权重, b_T 为模型中的偏置.

3.3 产品的表示对情感分析的影响

由上文的分析可知,口碑较好的产品情感倾向一般积极的较多,口碑较差的产品情感倾向一般消极的较多.我们随机选取了 Yelp 数据集上的几个商家.一家名为 Hollywood Café on Fillmore 的餐厅,该餐厅口碑良好,综合评分较高;另一家名为 The Cheesecake Factory 的餐厅,该餐厅服务态度差,卫生情况糟糕,口碑不好.统计的评论星级分布情况见表 1.口碑较好的产品,大众对其是认可的,评分星级相对集中在 4、5 颗星;而口碑较差的产品评分星级主要集中在 1、2 星级.均方差能够反映一个数据集的离散程度,由表 1 可以看出,同一个产品评分的均方差较小,基本上控制在 1.5 以内.这说明针对同一个产品的大众情感倾向是相似的.而同一个用户的情感倾向可能会因为产品不同而不同,因此,产品信息与情感分类之间的相关性远远超过了用户信息.

Table 1 Data set merchant rating star examples

表 1 数据集商家评分星级示例

商家名称	1 星	2 星	3 星	4 星	5 星	均方差
Brew cha	3	4	8	15	57	1.057
Hollywood cafe on Fillmore	0	0	1	7	13	0.597
The cheesecake factory	20	9	14	8	12	1.504
Tads steak restaurant	25	10	6	6	2	1.233

我们随机抽取实验数据集上某产品的两条评论,如例 3 和例 4 所示.可以看到,这些用户都对这一产品表达了消极的情感态度,两条评论的情感倾向是一致的.但是评论的质量却是不同的,例 3 是低质量的评论,而例 4 是高质量的评论.从例 3 和例 4 我们可以发现,同一产品一般具有一致的情感倾向,口碑较好的产品评论情感倾向大多为正面,口碑不好的产品评论情感倾向大多为负面,但评论质量与产品之间没有直接的联系.

例 3: Food is hard to eat, service is not good, will not come again.(食物很难吃,服务也不好,下次不会再来.)

例 4: This bakery was just okay, not so satisfying. My advice is keep on walking to the next one. I will not come back that's for sure. I feel like because we had our backpacks on, we were treated less than the business people that were coming in. so please show a little more attentiveness and care next time!! Also the food was just okay, but our turkey sandwiches lacked flavor or something. It could have been better overall, waste of time and money.(这家面包店一般,不是很满意,我的建议是继续走到下一个.我肯定是不回来的.我觉得因为我们背着双肩包,所以我们的待遇比商务人士要少.所以请在下次表现出更多的关注和关怀!!食物还不错,但是我们的火鸡三明治缺乏了一点味道,它本来可以更好的,浪费时间和金钱.)

为了捕获产品的表示对情感分类的影响,我们将结合产品信息注意力模型的输出向量 $V_p \in R^{d-1}$ 用于情感分类,该向量由 h_{p_i} 的加权和计算得到.

$$v_p = \sum_{i=1}^{n_p} \beta_i h_{p_i} \tag{18}$$

其中, β 为输入状态 h_p 的权重, $\beta_i \in [0, 1]$ 并且 $\sum_i \beta_i = 1$, n_p 是隐藏变量大小. 对于每一个 h_p , 我们使用前馈神经网络来计算其语义相关性, 并且向量化表示为 V_p . β_i^t 和 u_i^t 计算如下.

$$u_i^t = \text{Tanh}(W_p h_p + W_T h_T^{t-1} + W_p v_p + b) \quad (19)$$

$$\beta_i^t = \frac{\exp(u_i^t)}{\sum_{j=1}^T \exp(u_j^t)} \quad (20)$$

3.4 用户的表示对评论质量检测的影响

评论质量与用户信息之间的相关性也比产品信息更大. 我们随机抽取实验数据集上某用户发表的两条评论, 如例 5 和例 6 所示. 这两条评论来自同一个用户, 我们可以发现, 该用户的评论质量普遍偏低. 因为该用户总是喜欢简单直接地表达他对该产品的喜好, 并没有详细描述该产品的细节. 不同的用户对产品的需求不一样, 用户体验也不一样, 所以该用户的评论对消费者的参考价值不大. 因此, 来自同一用户的评论质量往往具有一定的相似性, 而来自同一产品的评论质量没有直接的关系, 但评论质量可能会因为不同的用户而有所不同.

例 5: The TV situation is almost epic, so any sporting event will be easy to watch. A few drink specials each night, but Bud Light bottles are just \$3.50 anyway if you in that mood. (电视情况几乎是史诗般的, 所以任何体育赛事都很容易观看. 如果你有心情的话, 每天晚上都有一些特价饮料, Bud Light 啤酒的售价也不过 3.50 美元.)

例 6: This pasta clearly has a chipotle kick, but is worth the sweat. Lots of giant rigatoni with some tomatoes and cream sauce and a lot of food for only \$9.50. It comes with not enough garlic bread, so get extra or have some ready at home. Enjoyable for one time for me, but my spicy tolerance level is very gringo, but it was tasty. (这份意大利面显然很有刺激感, 是值得大汗淋漓. 巨大的通心粉配一些西红柿、奶油酱和很多食物, 价格仅为 9.5 美元. 它没有足够的大蒜面包, 所以多吃点或者在家准备一些. 对我来说是一段很享受的时光, 因为我的辛辣忍耐力是非常好的, 但它是美味的.)

为了捕获用户的表示对评论质量检测的影响, 我们将结合用户信息注意力模型的输出向量 $V_R \in R^{d+1}$ 用于评论质量检测^[38], 该向量由 h_{R_i} 的加权和计算得到.

$$v_R = \sum_{i=1}^{n_r} \beta_i h_{R_i} \quad (21)$$

其中, β_i 为输入状态 h_{R_i} 的权重, $\beta_i \in [0, 1]$ 并且 $\sum_i \beta_i = 1$, n_r 是隐藏变量大小. 对于每一个 h_{R_i} , 我们使用前馈神经网络来计算其语义相关性, 并且向量化表示为 V_R . β_i^r 和 u_i^r 函数计算如下.

$$u_i^r = \text{Tanh}(W_R h_{R_i} + W_T h_T^{t-1} + W_R v_R + b) \quad (22)$$

$$\beta_i^r = \frac{\exp(u_i^r)}{\sum_{j=1}^R \exp(u_j^r)} \quad (23)$$

4 实验结果与分析

本节首先介绍 Yelp 数据集的概况, 然后给出实验设置, 包括衡量性能的指标、神经网络模型中超参数的设置等; 其次介绍基准系统, 用于和本文结合用户信息和产品信息的联合模型进行比较. 本文共进行了 3 组实验: (1) 第 1 组实验是为了与不同模型实验结果进行比较; (2) 第 2 组实验分别做情感分析和评论质量检测任务, 研究用户信息和产品信息不同因素对两个任务的影响; (3) 第 3 组实验在联合模型下同时做情感分析和评论质量检测任务, 并且将基于注意力机制的用户信息和产品信息融入文本信息, 对比第 2 组实验模型的性能, 进行不同网络结构的比较. 最后给出本文提出的模型在 Yelp 数据集上的性能和分析.

4.1 数据集

本文使用 yelp2013 和 yelp2015 数据集.由于两数据集过于庞大,本文实验中只使用了其中部分数据.本文选取了与 yelp2013 大小相当的数据集,即抽取了 yelp2015 数据集上前 17 389 条评论.实验中,数据集分布见表 2.

Table 2 Statistics of Yelp corpus

表 2 Yelp 数据集的统计数据

	Yelp2013	Yelp2015
#class	5	5
#doc	78 966	1 569 264
#users	1 310	132 747
#products	1 635	102 245
#sens/doc	10.89	8.97
#words/sen	17.38	18.15

本文采用均方误差作为情感分析和评论质量检测的评价指标,因为均方误差能够很好地反映预测评论质量(usable label)和情感得分(sentiment label)与真实评论质量和情感得分的偏离程度.评价标准 MSE 的公式如下.

$$MSE = \frac{\sum_i^N (gr_i - pr_i)^2}{N} \quad (24)$$

其中, N 为整个数据集评论数, gr_i 和 pr_i 分别为真实评论质量或情感得分和预测的评论质量或情感得分.

4.2 实验设置

4.2.1 文本预处理

在 Yelp 数据集中,每个评论至少包括以下信息.

- (1) 评论的质量评分(usable score),该项作为评论的质量指标,Useful Score 越大,证明潜在用户认为该条评论的对其购物的帮助更大,评论的质量越高.在实验中,为了剔除 usable score 极端值,将 usable score 的上限设置为 10;同时,将 usable score 转化为 0~1 之间的数值,在评论质量检测时作为评论质量检测标签.
- (2) 评论的星级评分(star),该项作为评论的情感指标,Stars 分为 5 个等级(1-5 分),Stars 星级指数越高,表明用户该条评论的情感倾向越积极.stars 的上限为 5,在实验中,我们将 stars 转化为 0~0.5 之间的数值,在情感分析时作为情感分析标签.
- (3) 评论的文本信息.
- (4) 评论的用户 ID 信息.
- (5) 评论的对象商家或产品的 ID 信息.

4.2.2 参数设置

本实验将数据集划分为训练集、测试集:训练样本数量为 2 800 条评论,测试数据为 3 000 条评论.实验中,神经网络激活函数使用的激活函数为 sigmoid,Dropout 层接受比例为 0.2,batch_size 为 15,采用 Adam 算法优化模型参数,全连接层隐藏单元为 128 个.

4.3 实验结果及分析

4.3.1 不同模型实验结果比较

为了能够与已有的工作进行比较,体现神经网络深度学习的发展潜力,我们选取了传统机器学习性能较好的支持向量机 SVM 模型与 LSTM 基准模型进行比较(见表 3).

- SVM 模型:作为传统机器学习的代表,SVM(support vector machine,<http://svmlight.joachims.org/>)在很多文本分类任务中都取得了优秀的效果.本文以单个词 Unigram 作为特征,使用传统机器学习中 SVM 模型进行情感分析和评论质量检测.
- LSTM 模型:长短时记忆网络 LSTM 在自然语言处理的很多任务中都表现优异,因此我们采用 LSTM 模

型分别进行情感分类和评论质量检测任务,将实验结果与本文提出的结合用户信息和产品信息的联合模型性能进行比较.

- JLSTM 模型:同时进行情感分析和评论质量检测任务的联合模型,由于很少有工作是基于情感分类和评论质量检测的,因此 JLSTM 模型可以看作是一个较强的基准模型.
- JNET 模型:同时引入用户信息注意力机制和产品信息注意力机制的联合模型.

Table 3 Comparison of experimental results of different models

表 3 不同模型实验结果对比

模型名称	Yelp2013		Yelp2015	
	Sentiment	Quality	Sentiment	Quality
SVM	0.142	0.119	0.145	0.168
LSTM	0.125	0.093	0.126	0.161
JLSTM	0.123	0.093	0.125	0.157
JNET	0.114	0.084	0.117	0.127

通过对 SVM,LSTM,JLSTM 和 JNET 模型比较,我们发现,神经网络 LSTM 模型较传统的机器学习方法,性能的确得到了提高;并且本文提出的结合用户信息和产品信息的情感分析和评论质量检测 JNET 联合模型与基准模型相比,实验性能在基准系统的基础上得到了进一步的提高.通过比较 Yelp2013 和 Yelp2015 年的实验数据结果,我们发现,Yelp2015 上情感分析的 MSE 与 Yelp2013 的相似,但是评论质量检测的 MSE 与 Yelp2013 的有所差异.分析 Yelp 数据集的语料,觉得可能是 2013 年 Yelp 官方网站才刚刚推出官方评论质量评分,大众的关注度和积极性不高,给的评论质量评分比较随意和大众化,基本上都是比较中性的分数,分数差异不大.随着时间的推移,评论数量的不断加大,人们逐渐意识到评论质量检测的重要性,同时也由于评论质量不像情感分类,口碑好的产品一般都是较为积极的评论,口碑差的产品一般都是比较消极的评论,评论质量具有一定的主观性,每个人会根据自己的喜好和需求来给出评论质量,所以在 Yelp2015 数据集上面,评论质量检测的均方误差 MSE 要比情感分析的均方误差 MSE 大.

4.3.2 不同因素对实验结果的影响

为了研究产品信息和用户信息不同因素对情感分析和评论质量检测任务的影响,我们进行了第 2 组实验,实验结果见表 4.

- LSTM+User 模型:该模型在仅利用文本信息的 LSTM 模型的基础上引入了评论的用户的表示,分别进行情感分析和评论质量检测任务.
- LSTM+Product 模型:该模型在仅利用文本信息的 LSTM 模型的基础上引入了评论的产品的表示,分别进行情感分析和评论质量检测任务.
- LSTM+User+Product 模型:该模型在 LSTM 模型的基础上加入评论的用户的表示和产品的表示,实现情感分析和评论质量检测.

Table 4 Influence of different factors on the experimental results

表 4 不同因素对实验结果的影响

模型名称	Yelp2013		Yelp2015	
	Sentiment	Quality	Sentiment	Quality
LSTM	0.125	0.093	0.126	0.161
LSTM+User	0.120	0.088	0.123	0.148
LSTM+Product	0.119	0.089	0.121	0.149
LSTM+User+Product	0.117	0.086	0.119	0.145
LSTM+Att+User+Product	0.121	0.086	0.122	0.135

通过对 LSTM 与 LSTM+User 和 LSTM+Product 的比较发现,用户信息和产品信息对情感分类和评价质量预测都是有效的.此外,LSTM+User 在评估评论质量方面更有效,而 LSTM+Product 在情感分析方面更有效.同时,通过 LSTM+Att+User+Product 的实验结果我们可以看出,注意力机制在单个的情感分类和评论质量检测模型中对实验结果的提升性能并不大.

4.3.3 不同网络结构对实验结果的影响

同时,为了分别研究用户信息和产品信息对评论质量检测和情感分类联合模型的影响,我们使用以下模型进行了第3组实验,实验结果见表5.

- JLSTM 模型:同时进行情感分析和评论质量检测任务的联合模型.
- JLSTM+User 模型:在联合模型的基础上进一步引入了用户信息.
- JLSTM+Product 模型:在联合模型的基础上进一步引入了产品信息.
- JLSTM+Att+User 模型:在联合模型的基础上进一步引入了基于注意力机制的用户的表示.
- JLSTM+Att+Product 模型:在联合模型的基础上进一步引入了基于注意力机制的产品的表示.
- JNET 模型:同时引入用户信息注意力机制和产品信息注意力机制的联合模型,相比较上面的几种模型,实验性能进一步得到了提升.

Table 5 The influence of different network configuration on the experimental results

表 5 不同网络结构对实验结果的影响

模型名称	Yelp2013		Yelp2015	
	Sentiment	Quality	Sentiment	Quality
LSTM	0.125	0.093	0.126	0.161
JLSTM	0.123	0.093	0.125	0.157
JLSTM+User	0.120	0.088	0.124	0.153
JLSTM+Product	0.118	0.087	0.124	0.147
JLSTM+Att+User	0.120	0.087	0.118	0.139
JLSTM+Att+Product	0.117	0.086	0.121	0.144
JNET	0.114	0.084	0.117	0.127

通过对 LSTM 和 JLSTM 的比较发现,具有共享表示的联合模型 JLSTM 对情感分类和评论质量检测更加有效.这说明了情感分析与评论质量检测具有一定的相关性.并且,JLSTM+Att+User 在评论质量检测方面更有效,而 JLSTM+Att+Product 在情感分类方面更有效.这也验证了同一个产品的情感倾向具有一致性,同一个用户发表的评论质量具有相似性.同时,对比 JNET 模型与表 4 中的 LSTM+Att+User+Product 模型的实验结果,我们发现,JNET 模型性能的提高是在于联合模型而不是注意力机制.此外,我们提出的同时结合基于注意力机制的用户表示和产品表示的联合模型 JNET 优于所有其他模型的性能.

4.3.4 不同训练样本数量百分比的影响

为了研究不同训练样本数量对实验结果的影响,我们在实验结果相对更加正常的 Yelp2015 数据集上进行了不同训练样本数量百分比的实验.图 4 显示了基准模型 LSTM、同时输出情感分析和评论质量的多输出模型 Multi-Output 以及 JNET 模型在不同百分比训练样本下的实验结果.横轴为训练样本的数量百分比,纵轴为实验的 MSE 值.

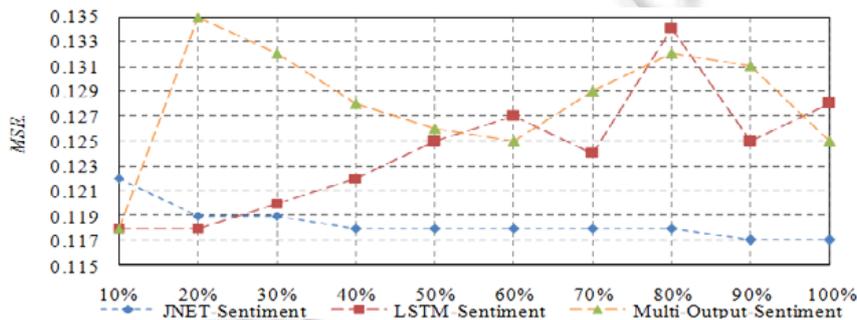


Fig.4 Experimental results of sentiment analysis under different percentages of training samples

图 4 不同数量百分比训练样本下情感分析实验结果

显然,从图 5 中我们可以发现,随着训练样本的增加,JNET 模型的 MSE 值不断的减小并趋于稳定,并且 JNET

模型的 MSE 值基本上低于 LSTM 和 Multi-Output 模型;同时,JENT 模型在不同训练样本下的实验结果更加稳定.这说明我们的模型也可以减少需要训练的样本数量.

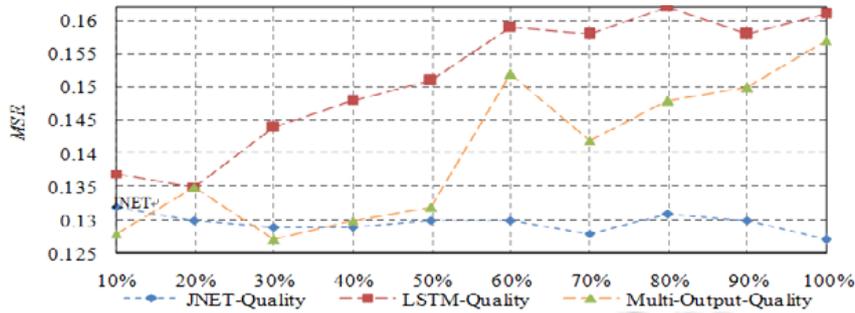


Fig.5 Experimental results of review quality detection under different percentages of training samples

图 5 不同数量百分比训练样本下评论质量检测实验结果

4.3.5 基于用户信息和产品信息的案例分析

为了更好地研究用户信息和产品信息的作用,我们进行了案例分析,表 6 和表 7 给出了一些案例,证明了加入用户信息和产品信息确实能进一步提高模型的准确度.

Table 6 Example output of LSTM and LSTM+User

表 6 LSTM 和 LSTM+User 模型案例分析

Review	LSTM	LSTM+User
	Quality	Quality
I still cannot understand for the life of me why every one of my friends wants to go here. The service is terrible, and the drinks aren't even that low priced.(我仍然无法理解为什么我的每个朋友都要来这里.服务很糟糕,酒水的价格也不低.Useful=1)	1	0
I recognize you are a talented award winning chef. However, I am wholeheartedly disappointed in what you've done to Pischke's ... Pischke's was a fun, friendly, unpretentious, comfort food joint. And now it's gone. Old Pischke's would've gotten five stars. (我知道你是一位才华横溢的获奖厨师.然而,我对你对 Pischke 所做的一切感到非常失望...Pischke 是一家有趣、友好、朴实无华、舒适的餐馆.现在它不见了.老 Pischke 会得到五星的.Useful=3)	1	0

Table 7 Example output of LSTM and LSTM+Product

表 7 LSTM 和 LSTM+Product 模型案例分析

Review	LSTM	LSTM+Product
	Sentiment	Sentiment
I really enjoyed my stay at Legacy ... The check in process was simple and fast and the front desk girl was really sweet. The property is huge, with a nice pool, kids playground, and golf course. The location is out of the way with not a whole lot nearby. However, the airport was 10 minutes away and all the great restaurant in Phoenix were only about 15 minutes away ... (我真的很喜欢我在 Legacy 的住宿...办理入住手续简单快捷,前台小姐也很贴心.房子很大,设有漂亮的游泳池、儿童游乐场和高尔夫球场.这个地方偏僻,人烟稀少.然而,机场离这里只有 10 分钟的路程...Stars=4)	0	0.5
Looking for a place in Phoenix to get some early breakfast Easter Sunday and we came across Lo-Lo's. The place is small and quaint ... First thing I noticed was that the whole staff was really friendly, welcoming, and attentive ... The food came out quickly and piping hot! (我们在凤凰城找个地方吃复活节早餐时遇到了洛洛餐厅,这个地方小而古雅...我首先注意到,所有的员工都非常友好、热情、专注...食物很快就来了,而且是滚烫的!Stars=5)	0	0.5

通过表 6 和表 7 的案例分析可以看出,表 6 的第 2 个例子,发表评论的用户仅仅说明了以前的 Pischke 餐厅他很喜欢,而对现在的 Pischke 餐厅的厨师不满意,但是并没有具体服务质量、饭菜口味等原因,所以对潜在消费者的参考价值不大.因此在加入用户信息后,模型做出的判断 helpful 值为 0,相比 LSTM 基准模型更加接近于网站消费者给出的评分 useful=3.通过分析表 7 中的第 1 个例子,该用户对 Legacy 酒店的服务、设施各方面都

是比较满意的,只是提到了位置偏僻、人烟稀少,但是后面也指出了交通还是很方便的.综合看来,表达的还是一个正向的情感,因此,加入了产品信息的 LSTM+Product 模型做出的判断更加准确.

4.3.6 基于联合模型的案例分析

表 8 给出了一些案例来证明我们提出的联合模型的有效性,案例分析的结果表明,本文提出的联合模型 JNET 相比 LSTM 模型结果更加准确,因为联合模型能将情感分析和评论质量检测任务联系起来.

Table 8 Example output of LSTM and JNET

表 8 LSTM 和 JNET 模型案例分析

Review	LSTM		JNET	
	Sentiment	Quality	Sentiment	Quality
I went back. Had the chicken again. It was much better and now Stacy get's four. count em ... four Stars! (我又回去吃了鸡肉.Stacy 现在好多了,得了四分...嗯...四颗星!Useful=1,Stars=4)	0	0	0.5	0
Went here because of seeing it on food network. I find it highly overrated ... It's basic, super greasy, not healthy at all breakfast but definitely would not go out of my way to get here. It was fine, Denny's is better. I tried to be healthy and ordered something with spinach and eggs. (来这里是因为在美食频道看到的,我发现它被高估了.这里是最基本的、超级油腻、不健康的早餐,但以我的方式绝对不会来这里...很好,Denny's 更好.我试着保持健康,点了一些菠菜和鸡蛋.Useful=2,Stars=2)	1	1	0	0

通过分析表 8 中的第 2 个例子我们可以看出,用户对在美食频道看见的这家餐厅并不是很满意,但是在评论中出现了“很好”“更好”这些词,使得基准模型 LSTM 将其误判为正向的情感,纵观整个评论和结合上下文语境,我们发现用户表达的情感还是更加倾向于负面,从“油腻”“不健康”“高估”等词可以看出来.并且在评论中没有出现该餐厅具体的价格、种类、地理位置等关键信息,因此对消费者的参考价值也不是很大.而本文提出的 JNET 模型能够正确地判断该条评论的情感倾向和评论质量.

5 结束语

近年来,用户生成的评论数据作为用户观点的重要载体越来越受到重视,同时,社交网站上评论增多,也为情感分类和评论质量检测工作带来了新的挑战,挖掘评论中包含的情感倾向,对潜在顾客的购物倾向和商家的营销策略具有重要的应用价值,而社交媒体上的评论数量众多,评论质量良莠不齐,更新速度快.因此,评论质量检测研究工作尤为重要,评论质量检测可以快速筛选出高质量的在线评论,节省时间,提高效率.

情感分类和评论质量检测这两个任务之间存在一定的相关性,此外,在情感分类和评论质量检测联合模型中,应该同时考虑用户信息和产品信息的影响.来自同一产品的评论具有相似的情感倾向,而来自同一用户的评论质量也具有一定的相似性.因此,本文提出了一种将用户信息和产品信息相结合的情感分析和评论质量检测联合模型,具体来说:首先,用深度学习方法学习评论文本的共享表示;然后,利用注意力机制分别对用户信息和产品信息进行编码.此外,我们还设计了一种组合策略,以充分利用这两种信息进行训练和预测.实验结果表明,本文提出的 JNET 模型情感分类和评论质量检测效果与其他神经网络模型相比都得到了明显的改进.

致谢 衷心感谢为本文的研究工作提供帮助的老师和同学.

References:

- [1] Pang B, Lee L. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In: Proc. of the 43rd Annual Meeting of the Association for Computational Linguistics. University of Michigan, 2005. 115–124.
- [2] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. In: Proc. of the 2002 Conf. on Empirical Methods in Natural Language Processing. EMNLP, 2002. 79–86.
- [3] Liu Y, Huang X, An A, *et al.* Modeling and predicting the helpfulness of online reviews. In: Proc. of the 8th IEEE Int'l Conf. on Data Mining. Pisa: ICDM, 2008. 443–452.

- [4] Kim S, Pantel P, Chklovski T, *et al.* Automatically assessing review helpfulness. In: Proc. of the 2006 Conf. on Empirical Methods in Natural Language Processing. Sydney: EMNLP, 2006. 423–430.
- [5] Yang Y, Yan Y, Qiu M, *et al.* Semantic analysis and helpfulness prediction of text for online product reviews. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing of the Asian Federation of Natural Language Processing. Beijing: ACL, 2015. 38–44.
- [6] Turney PD. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proc. of the Meeting on Association for Computational Linguistics. 2002. 417–424.
- [7] Zhang C, Zeng D, Li J, *et al.* Sentiment analysis of Chinese documents: From sentence to document level. Journal of the Association for Information Science and Technology, 2009,60(12):2474–2487.
- [8] Wen X, Shao L, Xue Y, *et al.* A rapid learning algorithm for vehicle classification. Information Sciences: An Int'l Journal, 2015, 295(C):395–406.
- [9] Chen B, Shu H, Coatrieux G, *et al.* Color image analysis by quaternion-type moments. Journal of Mathematical Imaging and Vision, 2015,51(1):124–144.
- [10] Gu B, Sheng VS, Wang Z, *et al.* Incremental learning for ν -support vector regression. Neural Networks, 2015,67(C):140–150.
- [11] Socher R, Perelygin A, Wu JY, *et al.* Recursive deep models for semantic compositionality over a sentiment treebank. In: Proc. of the Conf. on Empirical Methods in Natural Language Processing. 2013. 1631–1642.
- [12] Tai KS, Socher R, Manning CD. Improved semantic representations from tree-structured long short-term memory networks. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics. 2015. 1556–1566.
- [13] Wang Q, Cheng L. Research on the attention mechanism-based bidirectional LSTM model for the sentiment classification of Chinese product reviews. Software Engineering, 2017,20(11):4–6 (in Chinese with English abstract).
- [14] Zichao Y, Diyi Y, Chris D, *et al.* Hierarchical attention networks for document classification. In: Proc. of the Conf. of the North American Chapter of the Association for Computational Linguistics. San Diego, 2016. 1480–1489.
- [15] Li X, Bing L, Lam W, *et al.* Transformation networks for target-oriented sentiment classification. In: Proc. of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne: ACL, 2018. 946–956.
- [16] Larkey LS. Automatic essay grading using text categorization techniques. In: Proc. of the 21st Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 1998. 90–95.
- [17] Kim SM, Pantel P, Chklovski T, *et al.* Automatically assessing review helpfulness. In: Proc. of the 2006 Conf. on Empirical Methods in Natural Language Processing. Sydney, 2006. 423–430.
- [18] Meng Y, Wang HW. Evaluating online reviews based on text content features. New Technology of Library and Information Service, 2016,32(4):40–47 (in Chinese with English abstract).
- [19] Hu XG, Cheng FX, Zhang YH. Research on impact factors of online reviews' helpfulness based on product reviews data. Application Research of Computers, 2016,33(12):3559–3561 (in Chinese with English abstract).
- [20] Liu Y, Huang X, An A, *et al.* Modeling and predicting the helpfulness of online reviews. In: Proc. of the Int'l Conf. on Data Mining. IEEE, 2009. 443–452.
- [21] Yang Y, Yan Y, Qiu M, *et al.* Semantic analysis and helpfulness prediction of text for online product reviews. In: Proc. of the 23rd Int'l Conf. on Parallel & Distributed Systems. IEEE, 2015. 38–44.
- [22] Martin L, Sintsova V, Pu P. Are influential writers more objective? An analysis of emotionality in review comments. In: Proc. of the Int'l Conf. on World Wide Web. ACM, 2014. 799–804.
- [23] Liu H, Gao Y, Lv P, *et al.* Using argument-based features to predict and analyse review helpfulness. In: Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing. EMNLP, 2017. 1358–1363.
- [24] Ghose A, Ipeirotis PG. Estimating the socio-economic impact of product reviews: Mining text and reviewer characteristics. IEEE Trans. on Knowledge & Data Engineering, 2011,23(10):1498–1512.
- [25] Zhang J, Lin Y, Huang T, *et al.* Evaluating reviews quality based on review content and reviewers expertise. In: Proc. of the Int'l Conf. on Database Systems for Advanced Applications. 2018. 36–47.
- [26] Ng HT, Jin KL. Chinese part-of-speech tagging: One-at-a-time or all-at-once? Word-based or character-based? In: Proc. of the Conf. on Empirical Methods in Natural Language Processing. Barcelona, 2004. 277–284.

- [27] Zhang Y, Clark S. Joint word segmentation and POS tagging using a single perceptron. In: Proc. of the Association for Computational Linguistics. Ohio, 2008. 888–896.
- [28] Shi Y, Wang M. A dual-layer CRFs based joint decoding method for cascaded segmentation and labeling tasks. In: Proc. of the 20th Int'l Joint Conf. on Artificial Intelligence. Hyderabad, 2007. 1707–1712.
- [29] Boonkwan P, Supnithi T. Bidirectional deep learning of context representation for joint word segmentation and POS tagging. IEEE Trans. on Multimedia, 2017,18(7):184–196.
- [30] Finkel JR, Manning CD. Hierarchical joint learning: Improving joint parsing and named entity recognition with non-jointly labeled data. In: Proc. of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala: DBLP, 2010. 720–728.
- [31] Yang B, Cardie C. Joint inference for fine-grained opinion extraction. In: Proc. of the Annual Meeting of the Association for Computational Linguistics. 2013. 1640–1649.
- [32] Li F, Zhang MS, Fu GH, *et al.* A neural joint model for entity and relation extraction from biomedical text. BMC Bioinformatics, 2017,18(1):Article No.198.
- [33] Kurita S, Kawahara D, Kurohashi S. Neural joint model for transition-based Chinese syntactic analysis. In: Proc. of the Annual Meeting of the Association for Computational Linguistics. 2017. 1204–1214.
- [34] Wang SL, Zhang Y, Che WX, *et al.* Joint extraction of entities and relations based on a novel graph scheme. In: Proc. of the 27th IJCAI-ECAI. 2018. 4461–4467.
- [35] Xu YJ, Sun CH, Liu YZ. Joint sentiment model integrating user characteristics. Journal of Computer Applications, 2018,38(5): 1261–1266 (in Chinese with English abstract).
- [36] Li DY, Wang J, Wen Z, *et al.* Labeled joint aspect sentiment model for movie reviews. Journal of Frontiers of Computer Science and Technology, 2018,12(2):300–307 (in Chinese with English abstract).
- [37] Mnih V, Heess N, Graves A, *et al.* Recurrent models of visual attention. In: Proc. of the 27th Advances in Neural Information Processing System (NIPS 2014). 2014. 2204–2212.
- [38] Wu F, Wang ZQ, Zhou XB, *et al.* Neural review quality prediction via textual and user information. Journal of Chinese Information Processing, 2019,33(9):107–114,140 (in Chinese with English abstract).

附中文参考文献:

- [13] 王青,成璐.基于注意力机制的双向 LSTM 模型在中文商品评论情感分类中的研究.软件工程,2017,20(11):4–6.
- [18] 孟园,王洪伟.基于文本内容特征选择的评论质量检测.现代图书情报技术,2016,32(4):40–47.
- [19] 胡学钢,陈方鑫,张玉红.在线商品评论有用性影响因素研究.计算机应用研究,2016,33(12):3559–3561.
- [35] 许银洁,孙春华,刘业政.考虑用户特征的主题情感联合模型.计算机应用,2018,38(5):1261–1266.
- [36] 李大宇,王佳,文治,等.面向电影评论的标签方面情感联合模型.计算机科学与探索,2018,12(2):300–307.
- [38] 吴璠,王中卿,周夏冰,等.基于文本和用户信息的在线评论质量检测.中文信息学报,2019,33(9):107–114,140.



吴璠(1994—),女,硕士,CCF 学生会员,主要研究领域为自然语言处理.



周夏冰(1988—),女,博士,讲师,CCF 专业会员,主要研究领域为自然语言处理,机器学习.



王中卿(1987—),男,博士,讲师,CCF 专业会员,主要研究领域为自然语言处理.



周国栋(1967—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为自然语言处理.