

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.

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

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

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,Usable 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.

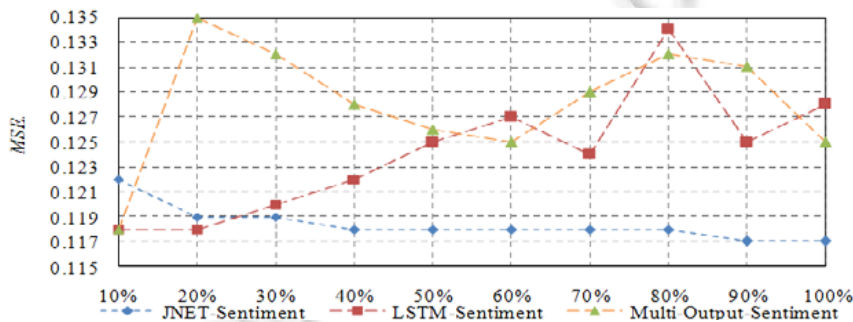


Fig.4 Experimental results of sentiment analysis under different percentages of training samples

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.

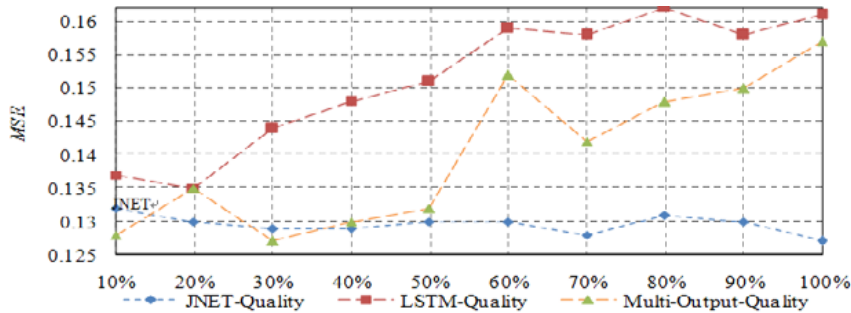


Fig.5 Experimental results of review quality detection under different percentages of training samples

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.