

文章编号: 1003-0077(2019)09-0123-06

基于循环实体网络的细粒度情感分析

贾川, 方睿, 浦东, 康刚

(成都信息工程大学 计算机学院, 四川 成都 610225)

摘要: 目前, 深度神经网络模型已经在文本情感分析领域取得了较好的效果, 但是对于属性相关的细粒度的情感分析任务, 现有研究方法的效果仍有待改进。该文提出了一种基于循环实体网络来进行细粒度情感分析的方法, 在网络中嵌入预定义的评价属性类别信息, 利用扩大的内部记忆链来抽取与每个属性类别相关的情感特征, 并通过动态记忆单元控制与属性相关情感信息的远距离依赖, 然后, 对于给定的单个属性类别, 利用注意力机制从内部记忆链中抽取该属性类别的情感特征进行分类。该文提出的方法在 Sentihood 数据上与目前精度最高的方法相比, 取得了近 1 个百分点的提升, 而且模型的收敛速度更快。

关键词: 细粒度情感分析; 循环网络; 属性嵌入

中图分类号: TP391

文献标识码: A

Recurrent Entity Networks for Fine-grained Sentiment Analysis

JIA Chuan, FANG Rui, PU Dong, KANG Gang

(School of Computer Science, Chengdu University of Information Technology, Chengdu, Sichuan 610225, China)

Abstract: At present, deep learning models have achieved good results in text sentiment analysis. Following this thread, we propose a method based on recurrent entity networks for fine-grained sentiment analysis. In this method, predefined evaluation attribute categories information is embedded in networks, and the sentiment features about each attribute categories are extracted through the expanded internal memory chains. The dynamic memory unit controls the distant sentiment dependence about attribute categories. Finally, for a given attribute category, attention mechanism is applied to extract sentiment features from internal memory chains. In our experiment, the proposed method achieves a nearly one percentage point improvement in Sentihood data compared to the current highest accuracy method.

Keywords: fine-grained sentiment analysis; recurrent networks; attribute embedding

0 引言

随着互联网上由用户生成的评论文本数据不断增多, 此类数据的研究价值和商业价值也越来越明显, 而情感分析或观点挖掘是针对此类文本数据的主要利用方式^[1]。早期对于文本评论的情感分析是对整个评论文本进行情感分类, 假设评论文本中已经包含且只包含感兴趣的评价对象, 分类的关键在于从评论文本中手工构造或自动生成情感特征, 而不考虑情感特征对应的评价对象信息^[1-2]。常见的研究, 如文档级情感分析和句子级情感分析^[3-4]。一

般的评论中往往包含多个评价对象, 或未包含感兴趣的评价对象, 且多个评价对象可能对应不同的情感倾向, 因此将整个评论文本看作整体进行情感分类略显粗糙, 限制了情感分析的作用。细粒度的情感分析任务, 也叫作基于属性的情感分析任务, 是针对评论文本中不同的评价属性, 识别出各个属性对应的情感倾向。通常, 一个评论文本中会出现针对多个对象表达的不同情感倾向, 比如, 对于评论“这家店的菜很好吃, 但是老板态度有点差”, 细粒度的情感分析需要识别出针对“菜”和“老板态度”两个不同方面的情感倾向, 得出<菜, 积极>和<老板态度, 消极>的评价结果。因此, 相比句子级或文档级

收稿日期: 2018-12-10 定稿日期: 2019-04-08

基金项目: 四川省科技厅重点项目(2017GZ0331)

情感分析任务,基于属性的细粒度情感分析可以从评论文本中获得更精准的用户评价信息,从而更有利于商家了解用户需求和改进产品及服务质量。

1 相关研究

目前已有研究中,存在将属性看作属性词和属性类别的两种不同处理方式^[5]。基于属性词的细粒度情感分析将任务看作序列标注,属性词为出现在评论文本中的具体词,标注出属性词之后,再抽取针对每个属性词的情感特征。基于属性类别的细粒度情感分析将属性类别看作预定义类别集合,属性类别词不一定出现在评论文本之中,对于每一个预定义的属性类别需要从文本中抽取对应的情感特征进行情感分类。

针对属性相关的细粒度情感分析,关键在于抽取与属性相关的情感特征。早期研究中使用的方法主要依赖于繁琐复杂的特征工程^[6-7],通过人工定义规则来提取特征,或者将浅层词向量特征与情感词典等人工定义特征结合形成针对属性的情感特征^[8],然后,利用机器学习方法进行情感分类,分类效果取决于人工特征以及领域专家的知识。近些年,随着神经网络理论和应用的不断发展,以及注意力机制^[9]和外部扩展记忆^[10]等网络计算机制的提出,针对属性的细粒度情感分析问题也涌现出不少使用神经网络的方法^[11-13]。得益于循环神经网络的序列处理特性,研究^[13-14]利用长短期记忆网络(Long short-term Memory Network, LSTM)或双向长短期记忆网络(bi-LSTM)抽取和识别文本中的情感特征。由于基于属性的情感分析需要提取对应于该属性的情感特征,Tang 等^[11]尝试在提取情感特征时将属性词与上下文词区别对待,提出 TD-LSTM 和 TC-LSTM 两种方式。其研究表明,在模型的序列处理每一个阶段都考虑属性特征的 TC-LSTM 可以取得当时最高的准确率。随后,Wang 等^[13]同样将属性表示与文本词序列表示拼接作为 LSTM 的输入,并在序列输出中结合属性特征表示来计算注意力权重,提出 ATAE-LSTM 模型。该模型还首次尝试用预定义的属性词嵌入矩阵对属性词进行嵌入表示,该方法在 SemEval 2014 Task4^①任务上取得了较好的效果。

常见的 LSTM 网络或门控循环网络(Gated Recurrent Unit Network, GRU)网络的内部隐状态通常为一个向量,使模型内部能够保存的信息有

限。因此,Weston 等^[10]提出使用可读写的记忆网络来扩展模型学到的知识,用于问答任务中的知识推理。同时,受多个计算层能够识别高层抽象特征^[15]的思想启发,Sukhbaatar 等^[16]对记忆网络进行多层扩展,发现具有多层计算的记忆网络在问答和语言建模等多个任务上均取得很好的效果。随后,相关研究^[17-18]将外部记忆的思想应用于针对属性词的细粒度情感分析,通过多层注意力从外部记忆中抽取针对属性的情感特征。但是以上方法在多个注意力层计算中,对记忆模块中的每一部分通过注意力层独立计算权重,舍弃了循环网络的序列处理步骤,而近期的研究表明,对文本输入的序列处理有利于提高整体性能^[19-20]。因此,Liu 等^[21]提出一种带延迟更新单元的循环实体网络,应用于属性情感分析,其方法主要受 Henaff 等^[20]提出的循环实体网络的启发,在 LSTM 中嵌入属性依赖的实体信息,通过扩大的内部记忆链,在序列处理输入文本的每一步都动态更新所有的内部记忆链。目前,Liu 等^[21]的方法在 Sentihood^[22]数据集上取得了最好的效果。本文提出的方法也受 Liu^[21]和 Henaff^[20]研究的启发,利用扩大内部记忆链的 LSTM 网络对评论文本进行序列处理,在给定属性类别的情况下进行属性情感特征抽取,但是不同于 Liu^[21]的方法,本文在循环实体网络单元中嵌入预定义的属性类别信息,使得在保持模型扩展性的同时更有利于提取属性类别相关的情感特征。实验表明,本文提出的方法在 Sentihood 数据集上取得了更高的准确率。

2 研究方法

上述研究中,基于属性词的细粒度情感分析^[8,11-12,23],要求用于训练和预测的评论文本数据中的属性词已经进行了标注识别,或者将属性词识别作为一个子任务^[24],然后再对识别出的属性词抽取情感特征进行分类。基于属性类别的细粒度情感分析^[5,13,21]省略了对属性词进行识别的需要,只需针对预定义的属性类别抽取与属性类别相关的情感特征。比如,针对上文提到的评论“这家店的菜很好吃,但是老板态度有点差”,基于属性类别的细粒度情感分析可以得到<味道,积极>和<服务,消极>的评价结果。本文认为在诸如电商平台、大众

^① <https://www.aclweb.org/portal/content/semEval-2014-task4-aspect-based-sentiment-analysis>

服务等领域,识别针对属性类别的情感倾向可以方便商家从整体上了解产品和服务质量,更具有实用价值。因此本文将基于属性类别的细粒度情感分析作为研究任务。

基于属性类别的细粒度情感分析,需要模型准确地识别并抽取与属性类别相关的情感特征。本文以 Henaff 等^[20]提出的循环实体网络为基础,以预训练词向量为输入,在循环实体网络中嵌入预定义的属性类别,并通过双向循环实体网络和注意力机制抽取属性情感特征,最后进行 softmax 分类。模型的总体结构如图 1 所示。

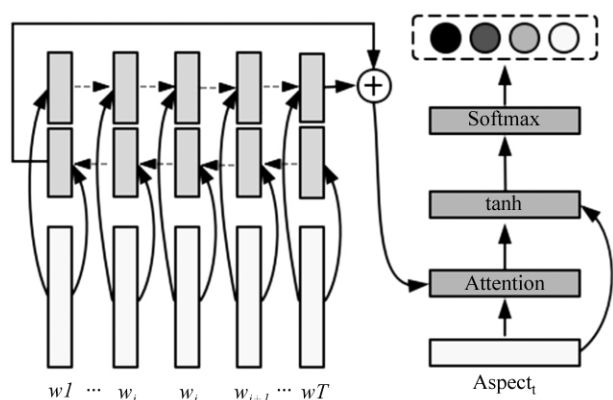


图 1 模型总体结构

在对输入评论文本转换为词嵌入序列后,使用双向循环实体网络抽取情感特征,循环实体网络结构如图 2 所示。

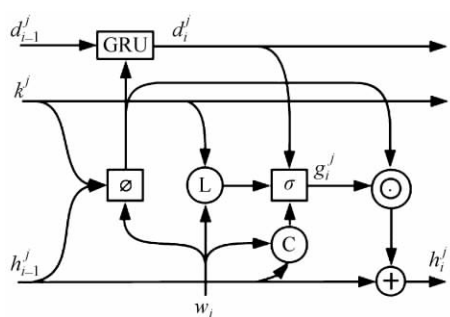


图 2 循环实体网络

该网络结构以 LSTM 网络为原型, w_i 为评论文本中的词,具体为词向量表示, k^j 为针对第 j 个预定义属性类别上的平均词嵌入表示, k^j 的个数为预定义的属性类别的个数,每一个类别 k^j 对应一个状态记忆链 h^j ,训练过程中更新对应属性类别的记忆状态, h_{i-1}^j 为循环实体网络上一步中针对属性类别 j 的隐层状态。对于具有长距离依赖关系的属性类别情感特征,类似于 Liu 等^[21]通过一个 GRU 单

元作为动态记忆模块来进行控制。GRU 单元的内部状态链的个数与预定义属性类别个数相同,每一个内部状态链 d_{i-1}^j 为针对属性类别 j 的上一步隐层状态, d_i^j 和 h_i^j 分别为当前针对属性类别 j 的动态记忆模块隐层状态和循环实体网络隐层状态。 g_i^j 为当前循环实体网络针对属性类别 j 的门控单元。对于循环实体网络的每一步输入 w_i ,计算针对属性类别 j 的临时状态 ${}^\circ h_i^j$,然后通过门控机制结合动态记忆更新输入状态 h_{i-1}^j 得到新的状态 h_i^j 。具体计算公式如式(1~4)所示。

$${}^\circ h_i^j = \phi(U h_{i-1}^j + V k^j + W w_i) \quad (1)$$

$$d_i^j = \text{GRU}({}^\circ h_i^j, d_{i-1}^j) \quad (2)$$

$$g_i^j = \sigma(w_i \cdot h_{i-1}^j + w_i \cdot k^j + v \cdot d_i^j) \quad (3)$$

$$h_i^j = h_{i-1}^j + g_i^j \odot {}^\circ h_i^j \quad (4)$$

式中 U, V, W, v 为训练参数,在单个方向上的每一层之间共享, ϕ 和 σ 代表非线性激活函数, \odot 表示逐元素相乘。临时状态 ${}^\circ h_i^j$ 由前一步记忆状态的线性变换、属性类别信息的线性变换、当前输入词的线性变换三部分信息决定。对应的门控单元 g_i^j 也由三部分控制,其中, $w_i \cdot h_{i-1}^j$ 代表内容信息, $w_i \cdot k^j$ 代表位置信息,当输入信息 w_i 与属性类别表示 k^j 相符,且携带属性类别 j 的情感特征时,二者将通过门控机制对状态 h_{i-1}^j 提供更大的更新量。 $v \cdot d_i^j$ 代表远距离依赖信息,如果当前输入词的信息与属性类别 j 存在远距离以来,则该部分为当前状态 h_i^j 提供较大的更新信息量,因此,动态记忆模块为属性相关的长距离依赖的情感特征提供了控制能力。当前的循环实体网络内部状态由前一步状态 h_{i-1}^j 与更新量 $g_i^j \odot {}^\circ h_i^j$ 相加表示。之后,对于更新后的状态 h_i^j 通过式(5)遗忘过期信息:

$$h_i^j = \frac{h_i^j}{\|h_i^j\|} \quad (5)$$

式(5)中 $\|h_i^j\|$ 为 h_i^j 的二范数,通过该方式让每一层中的隐层状态 h_i^j 取值限定在单位距离之内,从而达到在包含的所有信息中丢弃过期信息的作用。以上公式得出的 h_i^j 为循环网络在评论文本的单个方向上对各个属性类别 k^j 的隐层状态表示,同样在文本的另一个方向上,进行和上面一样的处理,最后将两个方向上的循环网络的最终隐层状态相加,从而得到双向循环实体网络的最终内部状态表示,如式(6)所示。

$$h_i^j = \vec{h}_i^j + \overleftarrow{h}_i^j \quad (6)$$

对于给定的属性类别,通过一个注意力层,计算给定属性类别 a 和所属实体 e 的情感特征 u 。然后,

类似于 Henaff^[20] 在输出模块将 u 转换并进行 softmax 情感分类。如式(7~10)所示。

$$p^j = \text{softmax}(k^j W_{att}[e, a]) \quad (7)$$

$$u = \sum_j p^j h_m^j \quad (8)$$

$$\hat{y} = \text{softmax}(R\phi(H_u + a)) \quad (9)$$

$$\text{Loss} = \text{CrossEntropy}(y, \hat{y}) \quad (10)$$

公式中, W_{att} 、 H 、 R 为训练参数, a 为属性类别的嵌入向量表示, e 为属性 a 的实体嵌入向量, m 为句子的长度, 式(10)中 \hat{y} 为预测值。最后, 通过交叉熵损失训练模型。

3 实验分析

本文实验所采用的数据集为 Sentihood^[22], 该数据集是针对不同地区的不同属性类别的情感评价, 其中, 每一个样本都标注了评价属性类别、所属地区和情感极性。属性类别包括价格、交通、安全、生活等 12 个方面, 所属地区包含 2 个, 情感类别包含积极或消极, 数据集共包含 5 215 条样本, 其中, 3 862 条样本是针对一个地区的不同方面评价。为方便对比, 本文和之前的研究^[20]一样, 采用数据集中最常见的 4 个属性类别(general, price, transit-location, safety)进行方法验证, 在评论文本中未提及的属性类别的情感类别判断为空, 按照 7 : 1 : 2 的比例划分训练集、验证集和测试集, 并将每条样本拆分成对单个属性类别和所属实体的情感类别标注。最终数据量分布如表 1 所示。

表 1 数据样本分布

	训练集	开发集	测试集
样本数	15 013	3 750	7 518

模型输入采用预训练的 Glove 词向量^[25] (300-D, 42B 个词)来初始化文本嵌入矩阵。对于包含多个词的属性类别, 用多个词的词向量平均来表示属性类别。单个隐藏层记忆链的维度设为 300, 参数矩阵 H 、 U 、 V 、 W 维度大小为 $\mathbb{R}^{300 \times 300}$, 延迟更新信息权重 $v \in \mathbb{R}^{300}$, 情感分类层权重 $R \in \mathbb{R}^{300 \times 3}$, 批量大小为 64, 学习率为 0.05, 未使用学习率衰减, 正则化参数为 0.002, 最终, 属性相关情感特征进行 dropout, 保留比例 0.8, 模型迭代训练 500 次。对于未表达情感极性的属性类别预测其情感标签为未提及, 为了避免未提及情感类别的样本数过多而导致

的数据分布不平衡问题, 在训练过程中按照未提及的类别数量对其他类进行降采样。

3.1 横向对比实验

对比实验阶段, 本文与其他在 Sentihood 数据集上进行属性级别细粒度情感分析的方法做了横向对比, 主要包含: (1) Saeidi 等^[22] 在 Sentihood 数据集上提出的基线模型 LR、LSTM-Final 以及 LSTM-Loc, LR 为包含 n 元特征和词性特征的逻辑回归分类器, LSTM-Final 为双向 LSTM 的最终隐状态拼接, LSTM-Loc 为考虑地区实体在文本中所处位置的双向 LSTM 的最终隐状态拼接, 隐状态为一维; (2) Ma 等^[26] 提出的引入常识信息的模型 SenticLSTM; 隐状态为一维; (3) Tang 等^[12] 提出的基于记忆网络的模型 TDLSTM, 使用文本嵌入作为外部记忆矩阵; (4) Liu 等^[21] 提出的嵌入实体信息的循环实体网络 EntNet, 隐状态为扩大的记忆链, 但未明显针对多个属性类别。验证指标采用情感分类的准确率和 AUC 值。对比实验结果如表 2 所示, 其中“—”表示未报告的实验指标, “*”表示平均 5 轮实验结果。

表 2 验证集上不同模型的指标对比

模型	准确率	AUC 值
LR	87.5	90.5
LSTM-Final	82.0	85.4
LSTM-Loc	81.9	83.9
TDLSTM	82.6	—
SenticLSTM	89.3	—
EntNet	91.0	94.8
Our model*	91.9	95.3

对比实验表明, 通过扩大记忆链的方式, 本文的方法和 Liu^[21] 的方法都取得了比较高的指标。而 Tang 等^[12] 将文本嵌入矩阵作为外部记忆的方法也是希望模型能够记住更多信息, 用于识别不同属性对应的情感特征, 但是, 在给定评论文本后, 不同属性类别的情感特征应该是固定的, 各属性类别之间的情感特征不应有太多交集。通过扩大记忆链的方式, 相当于给每一个预定义的属性类别分配一条记忆通道, 从而在模型训练过程中可以针对不同的属性类别更新相对应的内部记忆, 相对常见的循环神经网络而言, 扩大了内部记忆的容量, 而 Tang 等^[12] 的方法在更新记忆的过程中针对不同属性类别的更

新并不明显。由此,本文提出的方法和 Liu 等^[21]提出的方法证明,使用扩大记忆链的方式可以取得更好的效果。

相对 Liu 在 EntNet 中只考虑了属性所属的实体,本文将预定义的属性类别嵌入到模型中,使网络在循环处理文本序列时能够更好地识别针对不同属性类别的情感特征,提高了网络的特征表示能力,从而取得了更高的准确率和 AUC 值。

3.2 纵向对比实验

为进一步验证嵌入属性类别信息对抽取属性相关情感特征的影响是否稳定,本文在实验数据集上根据属性类别对应的样本个数,由多到少选择了多个属性类别,分别将不同数量的属性类别嵌入模型中进行了纵向对比实验。实验结果如表 3 所示。

表 3 嵌入不同数量属性类别的模型指标

类别个数	准确率	AUC 值
4	91.9	95.3
5	91.8	94.7
6	91.7	95.6
7	92.6	95.3
8	91.5	94.6
9	92.1	95.6

从表 3 可看出,随着属性类别嵌入的增多,对更多的属性类别进行细粒度情感分析的准确率指标和 AUC 指标并没有明显降低,这说明通过嵌入属性类别来增强模型对属性相关情感特征的抽取是确实有效的。而且,当针对不同任务存在不同数量的预定义属性类别个数时,该方法可以有很好的扩展能力。

另一方面,与 Liu 的方法相比,在循环实体网络中嵌入预定义的属性类别信息使得模型在训练中收敛速度上更快。如图 3 所示,横坐标表示训练次数,纵坐标表示平均损失。其原因在于,嵌入属性类别信息后,网络可以直接考虑到与属性类别相关的情感特征,因此在特征表示和损失计算时,模型对属性情感特征更敏感,从而有利于加快训练更新。

4 总结

本文提出了一种嵌入属性类别的循环实体网络来进行针对属性类别的细粒度情感分析,使得序列处理文本数据时,可以更好地针对不同属性类别抽

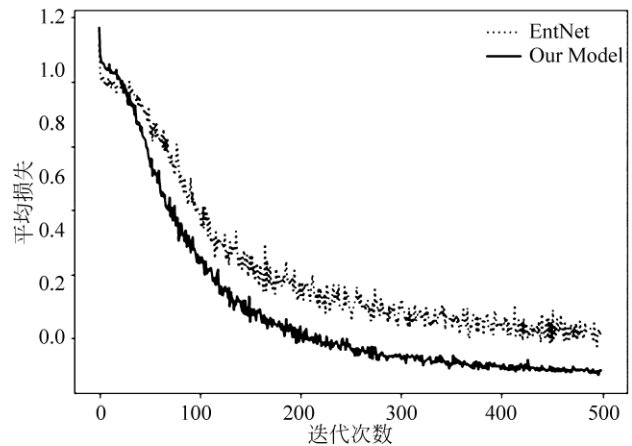


图 3 分别嵌入实体和属性后模型的收敛速度

取特定的情感特征,该方法在 Sentihood 数据集上取得了当前最高的准确率。后期会在此基础上探索如何将外部知识库与循环实体网络结合,以进一步提高模型的知识推理和特征表示能力。

参考文献

[1] B. Liu. Sentiment analysis and opinion mining[J]. Synthesis Lectures on Human Language Technologies, 2012,5(1):1-167.

[2] Pang B, Lee L, et al. Opinion mining and sentiment analysis. Foundations and trends[R]. Information Retrieval, 2008,2(1-2),1-135.

[3] N Kalchbrenner, E Grefenstette, P Blunsom. A convolutional neural network for modelling sentences[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL). Stroudsburg, PA. 2014: 655-665.

[4] Z Yang, D Yang, C Dyer, et al. Hierarchical attention networks for document classification[C]// Proceedings of HLT-NAACL. 2016: 1480-1489.

[5] W Xue, T Li. Aspect based sentiment analysis with gated convolutional networks[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL2018). 2018: 2514-2523.

[6] J Wagner, P Arora, S Cortes, et al. DCU: Aspect-based polarity classification for SemEval Task 4[C]// Proceedings of International Workshop on Semantic Evaluation. 2014.

[7] S Kiritchenko, X Zhu, C Cherry, et al. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews[C]// Proceedings of International Workshop on Semantic Evaluation. 2014: 437-442.

[8] D-T Vo, Y Zhang. Target-dependent Twitter sentiment classification with rich automatic features[C]//

- Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI2015). 2015: 1347-1353.
- [9] D Bahdanau, K Cho, Y Bengio. Neural machine translation by jointly learning to align and translate[J]. arXiv:1409.0473v2, 2014.
- [10] J Weston, S Chopra, A Bordes. Memory Networks [C]//Proceedings of International Conference on Learning Representations(ICLR). 2015.
- [11] D Tang, B Qin, X Feng, et al. Effective LSTMs for target-dependent sentiment classification[J]. arXiv: 1512.01100v2, 2016.
- [12] D Tang, B Qin, T Liu. Aspect level sentiment classification with deep memory network[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP2016). 2016. 214-224.
- [13] Y Wang, M Huang, L Zhao. Attention-based LSTM for aspect-level sentiment classification[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016: 606-615.
- [14] J Liu, Y Zhang. Attention modeling for targeted sentiment[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. 2017: 572-577.
- [15] Y LeCun, Y Bengio, G Hinton. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [16] S Sukhbaatar, J Weston, R Fergus. End-to-end memory networks[C]//Proceedings of Advances in Neural Information Processing Systems. 2015: 2440-2448.
- [17] L Dong, F Wei, C Tan, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014: 49-54.
- [18] P Chen, Z Sun, L Bing, et al. Recurrent attention network on memory for aspect sentiment analysis [C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017: 452-461.
- [19] M Seo, S Min, A Farhadi, et al. Query-reduction networks for question answering[C]//Proceedings of the 5th International Conference on Learning Representations(ICLR 2017), 2017.
- [20] M Henaff, J Weston, A Szlam, et al. Tracking the world state with recurrent entity networks[C]//Proceedings of the 5th International Conference on Learning Representations(ICLR 2017), 2017.
- [21] F Liu, T Cohn, T Baldwin. Recurrent entity networks with delayed memory update for targeted aspect-based sentiment analysis[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies (NAACL-HLT 2018). 2018: 278-283.
- [22] M Saeidi, G Bouchard, M Liakata, et al. SentiHood: targeted aspect based sentiment analysis dataset for urban neighbourhoods[C]//Proceedings of the 26th International Conference on Computational Linguistics (COLING). 2016: 1546-1556.
- [23] M Zhang, Y Zhang, D T Vo. Gated neural networks for targeted sentiment analysis[C]//Proceedings of AAAI. 2016: 3087-3093.
- [24] W Wang, S J Pan, D Dahlmeier, et al. Coupled multi-layer attentions for co-extraction of aspect and opinion terms[C]//Proceedings of AAAI. 2017.
- [25] J Pennington, R Socher, C Manning. Glove: Global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical.
- [26] Y Ma, H Peng, E Cambria. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM[C]//Proceedings of AAAI. 2018.



贾川(1993—), 硕士研究生, 主要研究领域为自然语言处理, 机器学习, 深度学习。
E-mail: amanchuan@163.com



浦东(1994—), 硕士研究生, 主要研究领域为数据挖掘、机器学习、自然语言处理。
E-mail: puudong@foxmail.com



方睿(1974—), 通信作者, 硕士, 教授, 主要研究领域为大数据处理、数据挖掘、数据安全。
E-mail: fangrui@cuit.edu.cn