

文章编号: 1003-0077(2021)10-0090-11

融合情感词交互注意力机制的属性抽取研究

程 梦, 洪 宇, 尉 桢楷, 姚建民

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘 要: 在现有的属性抽取研究中, 注意力建模多采用全局或局部的自注意力机制, 未能较好地利用句子本身所含有的情感词信息。然而, 句子中所需抽取的属性大多存在若干与其具有强相关性的情感词。因此, 利用情感词辅助注意力的建模可增强注意力的准确性。该文提出一种融合情感词的交互注意力机制, 将文本中所有的情感词按序排列, 并通过双向长短时记忆网络编码原始文本, 利用全连接神经网络和高速网络编码排列的情感词; 然后, 利用情感词编码与原始文本编码建模交互注意力, 从而使模型在情感词的辅助下精确地定位文中所包含的属性; 最终, 使用条件随机场进行属性标记。该文利用 2014、2015 语义评估属性级情感分析官方评测数据进行实验, 验证了上述方法的有效性, 该方法在三个基准数据集上 F_1 值分别提高了 5.53%、2.90% 和 5.76 个百分点。

关键词: 情感词; 交互注意力; 属性抽取; 条件随机场

中图分类号: TP391

文献标识码: A

Aspect Extraction via Interactive Attention with Sentiment Words

CHENG Meng, HONG Yu, WEI Zhenkai, YAO Jianmin

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

Abstract: Current aspect extraction methods adopt global or local self-attention mechanism without discriminating the sentiment words in the sentence. This paper proposes an aspect extraction method via interactive attention with sentiment words. We arrange all the sentiment word in the text in order, and apply Bi-directional Long-Short Term Memory network to encode sentence, and adopt fully connected neural network and High Way network to encode sentiment words. Then, we model the interactive attention with the sentence and sentiment words representations, which can precisely locate aspects with the help of sentiment words. Experimental results on three datasets from SemEval 2014-2015 show that our method can effectively improve the performance of aspect extraction by 5.53%, 2.90% and 5.76% respectively.

Keywords: sentiment words; interactive attention; aspect extraction; conditional random fields

0 引言

属性级情感分析 (Aspect Based Sentiment Analysis, ABSA) 任务致力于识别特定实体或属性所表达的情感, 并判断情感的褒贬极性, 是一种细粒度的情感分析任务。

属性级情感分析主要分为三个子任务^[1]: 属性抽取 (Aspect Extraction), 也叫作评价对象的抽取 (Opinion Target Extraction), 主要目标是识别出给定句子中关联于特定情感表述的实体或属性; 属性

归类 (Aspect Category Detection), 主要目的是将被抽取的实体或属性归并至一组预定义的类别集合; 属性情感极性 (Aspect Polarity) 判断, 主要目的是判断关联于属性或实体的情感极性 (一般分为正向、负向和中性)。本文主要聚焦于属性抽取子任务。

属性抽取任务定义为: 对用户所评论的文本, 抽取该用户评价的实体或属性。该实体或属性代表用户的评价对象。评价对象通常为一个单词或短语, 比如, 例 1 下划线标定的名词短语即为属性抽取的目标样本。

收稿日期: 2020-02-25

定稿日期: 2020-04-18

基金项目: 国家自然科学基金 (61672367, 61672368, 61773276, 61836007)

[例 1]

评论句子: Fish burger is the best dish, it tastes fresh.

(译文: 鱼汉堡是最好的菜, 它的口感新鲜。)

传统属性抽取方法主要采用基于监督学习的自动标记方法, 该方法将属性抽取任务转化为序列标记任务。其中, Jakob 等^[2]将条件随机场(Conditional Random Fields, CRF)模型首次应用于评价对象抽取的研究, 该方法在融入多种手工特征条件下, 在属性抽取的任务上取得了较好的效果, 在电影评论数据集^[3]上 F_1 值达到了 70.2%。近期, 大量研究将神经网络模型应用到属性抽取任务, 包括循环神经网络(Recurrent Neural Network, RNN), 双向循环神经网络(Bi-directional Recurrent Neural Network, Bi-RNN), 基于长短时记忆单元(LSTM)的 RNN 网络, 以及双向长短时记忆网络(Bi-LSTM)。这类方法均取得了显著的性能提升^[4]。在此基础上, Toh 等^[5]将 Bi-RNN 与 CRF 结合, 并加入多种手工特征, 进一步优化属性抽取的性能, 并于 2016 年语义评测(Semantic Evaluation, SemEval)属性级情感分析评测任务中^[6]综合成绩排名第一。

目前, 注意力机制^[7](Attention Mechanism)在多种自然语言处理任务上均取得了显著的性能提升, 其突出的优势在属性抽取任务上也得到学者的关注。例如, Wang 等^[8]使用成对多层注意力机制进行属性与情感词的联合抽取。Li 等^[9]利用窗口化的历史注意力机制捕获与属性词相关的局部历史信息。该类方法在加入注意力机制的前提下可显著提升属性抽取的性能。然而, Wang 等^[8]分别建模属性抽取通道和情感词抽取通道的全局注意力, Li 等^[9]建模局部的历史注意力信息, 均未能较好地利用文本中所包含的情感词信息辅助注意力的建模。

[例 2]

评论句子: The service is excellent and the decor is great.

(译文: 服务非常好, 装饰很棒。)

评价属性: service (服务), decor (装饰)

在属性抽取任务中, 待抽取的属性词大多存在若干与其具有强相关性的情感词, 比如, 例 2 中属性“service”存在与其强相关的情感词“excellent”, 属性“decor”存在与其强相关的情感词“great”。因此, 可利用上述强相关性, 将情感词信息融入待抽取文本的每个词项, 从而通过文本中的情感词“excellent”

和“great”辅助属性词的抽取。

本文提出一种融合情感词的交互注意力机制, 并将其应用到基于 Bi-LSTM 的 CRF 模型中。该注意力机制将文本中所有的情感词按序排列, 并与原始文本建模交互注意力。从原始文本的每个目标词项角度出发, 可获取被排列的情感词的重要程度, 从被排列的情感词角度出发, 可得知文本中哪些词项与情感词相关, 进而通过以上两种相关性, 将情感词融入文本表示中, 促进属性词的抽取。在排列情感词的编码过程中, 为了尽量保证情感词与原始文本的语义空间相似性, 本文利用高速网络^[10](High Way Network)将情感词映射到与原始文本相似或相同的语义空间。此外, 本文引入“是否情感词”向量, 以标识文本中每个词项是否为情感词, 从而为原始文本和排列的情感词建立初始弱相关性。

本文对融入了上述交互注意力机制的属性抽取系统进行测试。在国际公开数据集 SemEval 2014^[1]和 2015^[11]上进行测试, 均获得了性能的提升。在三个基准数据集上, F_1 值分别提高了 5.53%、2.90% 和 5.76%。

全文组织结构如下: 第 1 节回顾属性抽取相关工作; 第 2 节介绍融合情感词交互注意力的 LSTM 模型; 第 3 节给出实验数据, 并进行结果分析; 第 4 节总结全文。

1 相关工作

在属性级情感分析任务中, 属性抽取子任务已经积累了大量优秀研究成果。早期研究制定多种规则进行属性抽取研究。Hu 等^[12]首次制定一系列依存的关联关系规则进行评价对象抽取研究。Zhuang 等^[3]针对影评中的评价对象—评价观点二元组, 通过二者之间的依存关系实现抽取工作。Blair 等^[13]将主观句中的名词和名词短语按其出现的频率加权排序, 并按照权重选取所需抽取的属性。Wang 等^[14]将用户的评价对象和观点使用 Bootstrapping 方法进行交替识别。

此外, 对于主题模型, 学者发现在特定的条件下, 比如限定抽取评价对象类别的情况下, 可运用主题模型抽取评价对象。Mei 等^[15]在评价对象抽取的研究中发现, 概率潜在语义模型(Probabilistic Latent Semantic Analysis, PLSA)可促进评价对象的抽取。Titov 等^[16]针对评价对象抽取的粒度问题, 在潜在狄利克雷分配(Latent Dirichlet Alloca-

tion, LDA)模型的基础上,提出一种多粒度的 LDA 模型。Lin 等^[17]针对主题和情感词的联合抽取任务,提出一种联合主题模型,用于同时抽取主题和情感词。Mukherjee 等^[18]收集评价对象的种子集合,利用半监督的方法建立联合主题模型用于评价对象抽取。

在监督学习应用方面,可制定多种标签模式,通过标签所在的位置标识出文中存在的评价对象,因此,序列标记方法逐渐成为属性抽取的主流方法。Jin 等^[19]使用隐马尔科夫模型(Hidden Markov Model, HMM)学习标签的序列表示,通过序列标签标记评价对象。Jakob 等^[2]首次将 CRF 模型单独用于评价对象抽取研究。Li 等^[20]发现评价文本的结构有助于评价对象的抽取,比如链式结构、连接结构和句法结构,因此,可将文本结构特征化,并融入 CRF 模型中。

近期,由于神经网络模型的广泛应用,大量研究开始将其应用到属性抽取任务中。Xu 等^[21]使用递归自动编码器模型,同时抽取观点词、评价对象,以及它们之间的关系。Liu 等^[4]使用基于 LSTM 的循环神经网络(RNN),同时融入了词性的分布语义表示、自动学习词语之间的关联,并且在多个数据集上证明其优于 CRF 方法。Toh 等^[5]将 Bi-RNN 与 CRF 结合,并加入多种手工特征,以此优化属性抽取的性能。Wang 等^[22]通过 RNN 对依存树进行编码,学习单词的特征表示,并将表示学习的结果输入到 CRF 进行序列标注。Wang 等^[8]通过一种成对多层注意力机制(Coupled Multi-layer Attentions),捕获句子中各个单词之间的直接依存关系和间接依存关系,进行属性词和情感词的联合抽取。Li 等^[23]利用双通道 LSTM 与记忆力机制,实现属性和情感词的联合抽取。Xu 等^[24]使用卷积神经网络(Convolutional Neural Networks, CNN),并同时利用领域词向量和通用词向量提升属性抽取性能。Li 等^[9]利用窗口化的历史注意力,借此关注每个目标词项的局部历史信息。Ma 等^[25]探究了端到端模型在属性抽取任务中的应用。Dai 等^[26]使用规则扩充训练数据,并利用基于 BiLSTM 的 CRF 模型进行属性抽取。

其中,上述研究与本文研究工作最为相关的是 Toh 等^[5]以及 Xu 等^[24]的研究,本文以 Toh 等^[5]的系统为基线系统,并将融合情感词的交互注意力机制融入该系统中。由于传统的 RNN 难以处理长期依赖关系,而 LSTM 相比于 RNN 能更加有效地学

习长期依赖关系,因此我们在 Toh 等^[5]基线系统的基础上,以更适宜于学习长期依赖关系的 LSTM 代替系统中的埃尔曼型(Elman-type)RNN。在系统的输入层,本文利用 Xu 等^[24]提出的双词向量思想,并引入“是否情感词”向量,以标识文本中每个词项是否为情感词。在系统的输出部分,本文采用 Luo 等^[27]提出的堆叠 RNN(stacked RNN),在交互注意力计算的下游堆叠一层双向门控循环单元(Bi-directional Gated Recurrent Unit, Bi-GRU),用于建模交互注意力的输出向量。最后,输出层使用 CRF 解码出真实预测标签。与 Wang 等^[8]使用成对多层注意力机制建模文本的全局注意力和 Li 等^[9]使用历史注意力建模局部历史信息不同,本文所提融合情感词的交互注意力机制能利用文本本身存在的情感词辅助注意力建模,通过交互注意力实现文本和情感词排列之间的交互,从而将情感词融入文本的表示中,提高属性抽取的效果。此外,本文利用高速网络将情感词映射到与原始文本相似或相同的语义空间,使情感词与原始文本能在相同语义空间中进行计算。

2 融合情感词交互注意力的 LSTM 模型

本文采用序列标记方法处理属性抽取任务,使用的标签模式为 BIO 标签。其中,B 代表属性短语的开始,I 代表属性短语的中间词或结尾词,O 代表非评价对象(例 3)。

[例 3]

句子: Best spicy tuna roll, great Asian salad.

标签: O B I I O O B I O

(译文:最好的辣金枪鱼卷,美味的亚洲沙拉。)

本文针对属性抽取任务,利用原始文本存在的情感词辅助注意力的建模,提出一种融合情感词的交互注意力机制。本文模型如图 1 所示,分别为:词向量层、编码层、融合情感词的交互注意力层、堆叠 Bi-GRU 层以及 CRF 解码层。下面将对每一层的计算模型给出详细介绍。其中融合情感词的交互注意力层为本文重点。

2.1 词向量层

本文利用 Xu 等^[24]提出的双词向量,将预训练的通用词向量 $\mathbf{G} \in \mathbf{R}^{d_G \times |V|}$ 和领域词向量 $\mathbf{D} \in \mathbf{R}^{d_D \times |V|}$ 拼接作为词的初始化词向量。除此之外,本文使用随机初始化的“是否情感词”向量 $\mathbf{O} \in \mathbf{R}^{d_O \times 2}$

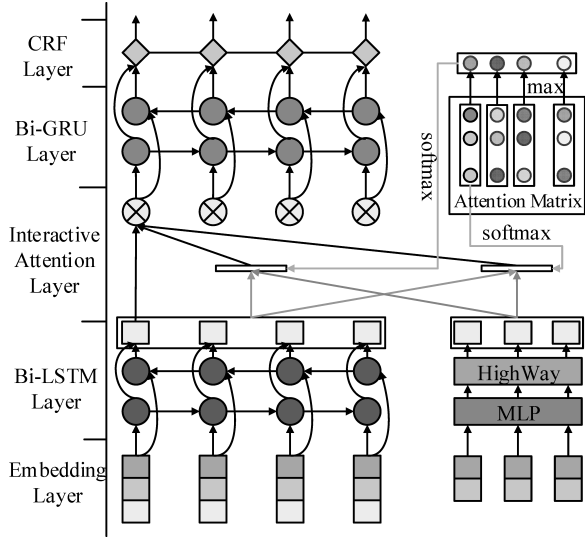


图1 融合情感词的交互注意力机制图

与词向量拼接作为词的最终向量表示。其中, $|V|$ 代表词表大小, d_G 代表通用词向量的维度, d_D 代表领域词向量维度, d_O 代表“是否情感词”向量的维度。因此, 对于待判别句子 $S = \{s_1, s_2, \dots, s_n\}$, 其初始化语义向量表示为 $W = \{w_1, w_2, \dots, w_n\}$, $w_i = [G(s_i); D(s_i); O(s_i)]$; 对于句子所包含的情感词, 本文将所有情感词按其在句子中出现的顺序进行排列, 获取排列的情感词 $S^S = \{s_1^S, s_2^S, \dots, s_m^S\}$, 其初始化向量表示为 $W^S = \{w_1^S, w_2^S, \dots, w_m^S\}$, $w_i^S = [G(s_i^S); D(s_i^S)]$, 其中, $[\cdot]$ 表示向量的拼接。

2.2 编码层

编码层需要分别编码待判别句子和句子中按序排列的情感词。对于待判别句子初始化语义向量表示 W , 本文使用 Bi-LSTM 编码, 获取每个单词的隐层表示 $H = \{h_1, h_2, \dots, h_n\}$ 。将单词 w_i 的正向 LSTM 编码表示 \vec{h}_i 和反向 LSTM 编码表示 \bar{h}_i 拼接, 获取 w_i 由 Bi-LSTM 编码后的联合语义向量表示为 h_i , 计算模型如式(1)~式(3)所示。

$$\vec{h}_i = \text{LSTM}(\vec{h}_{i-1}, w_i, \vec{\theta}) \quad (1)$$

$$\bar{h}_i = \text{LSTM}(\bar{h}_{i+1}, w_i, \bar{\theta}) \quad (2)$$

$$h_i = [\vec{h}_i; \bar{h}_i] \quad (3)$$

其中, $\vec{\theta}$ 表示正向 LSTM 神经元中的参数, $\bar{\theta}$ 表示反向 LSTM 神经元中的参数。

借用 Liu 等^[10]使用高速网络调整多任务学习中各任务的语义空间的思路, 本文利用高速网络调整原始文本和情感词的语义空间, 使二者在同一语义空间中计算。对于排列的情感词初始化向量表示

W^S , 由于 W^S 中各个单词之间没有前后时刻的联系, 因此无法像待判别句子那样使用 BiLSTM 进行编码。此外, 为了方便后续计算, 需将每个情感词 w_i^S 的维度转化为与原始文本编码 h_i 相同的维度, 因此, 本文利用 tanh 函数激活的全连接神经网络对排列的情感词同时进行编码和维度转化。在此之后, 使用高速网络进行语义空间的转化, 获取情感词编码的隐层表示 $H^S = \{h_1^S, h_2^S, \dots, h_m^S\}$, 计算方式如式(4)~式(6)所示。

$$h_i^S = t \odot g(W_h h_i^M + b_h) + (1 - t) \odot h_i^M \quad (4)$$

$$t = \sigma(W_t h_i^M + b_t) \quad (5)$$

$$h_i^M = \tanh(W_M w_i^S + b_M) \quad (6)$$

其中, 式(6)为使用 tanh 函数激活的全连接网络计算, 式(4)和式(5)为高速网络的计算, t 表示转移门, 使用 sigmoid 激活函数, $(1 - t)$ 表示携带门, g 表示激活函数, 本文使用的 RELU 激活函数, \odot 表示对应元素相乘。

2.3 融合情感词的交互注意力层

本节为本文重点, 在编码层分别编码原始文本和排列的情感词之后, 使用交互注意力进行原始文本和情感词之间的交互建模(下文简称情感交互层)。

情感交互层首先使用原始文本编码 H 和情感词编码 H^S 计算注意力相似度矩阵 $A \in \mathbf{R}^{m \times n}$, 其中 A 的每个元素 $A_{i,j}$ 由原始文本和情感词交互计算获得, 计算如式(7)所示。

$$A_{i,j} = W_{hs} [h_i; h_j^S; h_i \odot h_j^S] + b_{hs} \quad (7)$$

在此之后, 使用注意力相似度矩阵分别从两个角度计算原始文本注意力向量和情感词注意力向量, 并将此二者注意力向量进行融合。

从文本看情感词角度: 对于原始文本中的每个目标词项 i , 取注意力相似度矩阵 A 中与目标词项 i 对应的第 i 列 $A_{:,i}$, 作为当前词看情感词的注意力得分, 并使用 softmax 函数归一化该注意力得分, 从而得知情感词对于当前目标词项 i 的重要程度 a_i^S 。利用归一化的注意力得分与情感词加权求和, 计算从当前目标词 i 看情感词角度的注意力 $c_i^S \in C^S$, $C^S = \{c_1^S, c_2^S, \dots, c_n^S\}$ 。计算如式(8)~式(9)所示。

$$c_i^S = \sum_{j=1}^m a_{ij}^S h_j^S \quad (8)$$

$$a_i^S = \text{softmax}(A_{:,i}) \quad (9)$$

从情感词看文本角度: 对于所有的情感词, 取注意力矩阵中每列的最大值 $a'_i \in a'$, 获取原始文本

对于所有的情感词的注意力得分 $a' = \{a'_1, a'_2, \dots, a'_n\}$, 并使用 softmax 函数归一化该注意力得分, 从而获得原始文本对于所有情感词的重要程度 a 。利用归一化的重要性程度和原始文本加权求和, 计算从情感词看文本角度的注意力 $c \in C$, 并将 c 复制 n 次获得 $C = \{c_1, c_2, \dots, c_n\}$ 。计算如式(10)~式(12)所示。

$$c = \sum_{j=1}^n a_j h_j \quad (10)$$

$$a = \text{softmax}(a') \quad (11)$$

$$a'_i = \max(A_{:,i}) \quad (12)$$

注意力向量融合: 本文将从文本看情感词角度的注意力向量和从情感词看文本角度的注意力向量进行交互融合, 获得融合情感词的交互注意力向量 $R = (r_1, r_2, \dots, r_n)$ 。如此, 对于每个词项, 其注意力向量 r_i 均融合了情感词信息, 从而以情感词信息辅助属性的抽取。计算如式(13)所示。

$$r_i = [h_i; c_i^S; h_i \odot c_i^S; h_i \odot c_i] \quad (13)$$

2.4 堆叠 Bi-GRU 层

参考 Luo 等^[27]的堆叠 RNN, 本文在情感交互层的输出后堆叠一层 Bi-GRU, 用于对情感交互层的输出进行进一步的建模。在实验过程中, 本文尝试使用 BiLSTM 网络进行堆叠, 由于其效果与 Bi-GRU 相近, 因此, 本文选取参数相对较少的 Bi-GRU 作为堆叠的 RNN 网络。与编码层使用的 BiLSTM 类似, 将前向 GRU 的输出与反向 GRU 的输出拼接, 获得堆叠 Bi-GRU 的建模表示 $U = \{u_1, u_2, \dots, u_n\}$ 。堆叠 Bi-GRU 层的计算如式(14)~式(16)所示。

$$\vec{u}_i = \text{GRU}(\vec{u}_{i-1}, r_i, \vec{\theta}_u) \quad (14)$$

$$\overleftarrow{u}_i = \text{GRU}(\overleftarrow{u}_{i+1}, r_i, \overleftarrow{\theta}_u) \quad (15)$$

$$u_i = [\vec{u}_i; \overleftarrow{u}_i] \quad (16)$$

其中, $\vec{\theta}_u$ 表示正向 GRU 神经元中的参数, $\overleftarrow{\theta}_u$ 表示反向 GRU 神经元中的参数。

2.5 CRF 解码层

本文将 U 输入到 CRF 进行解码, 获取句子中每个单词的预测标签 $L = \{l_1, l_2, \dots, l_n\}$, 其中, $l_i \in \{B, I, O\}$ 。将 U 由全连接层进行降维, 获取 CRF 输入的发射分数 $E = \{e_1, e_2, \dots, e_n\}$, 通过优化 CRF 的序列对数似然损失函数训练本文的属性抽取模型, 其损失函数如式(17)所示。

$$L = \sum_{i=1}^n \log P(y | e_i; W_E, b_E) \quad (17)$$

其中, W_E 为 CRF 层的转移矩阵, b_E 为偏置。预测时, 使用维特比算法根据发射分数 E 和转移矩阵 W_E 推算出预测标签。

3 实验

3.1 实验数据及设置

为了验证本文方法的有效性, 本文在 2014、2015 SemEval 属性级情感分析的三个不同数据集上进行实验分析。该数据集包含电脑(laptop)领域和餐馆(restaurant)领域。其来源如下: 2014 年语义评测任务四^①电脑领域(SemEval 2014 task 4 laptop, 14 LAPT), 2014 年语义评测任务四餐馆领域(SemEval 2014 task 4 restaurant, 14 REST), 2015 年语义评测任务十二^②餐馆领域(SemEval 2015 task 12 restaurant, 15 REST)。其中, 各个数据集中均标注了每个句子存在的属性词。

实验数据设置如下: 三个数据集中, 训练样本数量分别为 3 045、3 041 和 1 315。测试样本数量分别为 800、800 和 685。实验过程中, 参考 Xu 等^[24]随机从训练数据中选取 150 条样本作为开发集。经过开发集的划分后, 各个数据集的训练集、开发集以及测试集样本数量如表 1 所示。

表 1 语料统计

数据集	来源	训练集	开发集	测试集
S1	14 LAPT	2 895	150	800
S2	14 REST	2 891	150	800
S3	15 REST	1 165	150	685

3.2 评价方法

与 Wang 等^[8]相同, 本文采用 F_1 值作为评价指标。评价过程中, 只有当模型预测结果与标准答案完全一致, 才能计算为一个正确的预测答案。例如, 标准答案为“spicy tuna roll”, 如果模型预测结果为“tuna roll”, 则不是正确答案。只有当模型预测结果与标准答案完全匹配, 才为正确的预测答案。

① <http://alt.qcri.org/semeval2014/task4/>

② <http://alt.qcri.org/semeval2015/task12/>

3.3 参数设置

在双词向量中,通用词向量使用 100 维的 GloVe 词向量^①,领域词向量使用 Xu 等^[24]分别使用电脑领域和餐馆领域数据训练出的 100 维电脑领域词向量和餐馆领域词向量^②,“是否情感词”向量使用 20 维随机初始化的向量。实验过程中,词向量固定,训练过程中不更新,“是否情感词”向量随着训练过程不断更新。对于领域词向量的未登录词,本文使用 fastText^③获取其词向量,对于通用词向量的未登录词,本文随机初始化 100 维向量作为其词向量。本文使用 Wang 等^[8]手工标注的情感词,为每个句子提取出其中存在的情感词。学习率 (Learning Rate) 固定为 0.001,批量 (Batch Size) 输入的大小设为 64。BiLSTM 隐层维度和堆叠的 Bi-GRU 隐层维度均设置为 128。为了避免过拟合,本文在词向量层和交互层之后加入 dropout,在上述两层之后加入部分 dropout (partial dropout)^[28],其中 dropout 比率(dropout rate)设置为 0.5。为了避免神经网络初始化的随机性引起的实验性能波动,本文在实验中以相同的超参数(hyper-parameters)训练出 5 个模型进行测试,并选取 5 个模型测试结果的平均值作为最终的实验性能。

3.4 实验配置

为了与本文所提融合情感词的交互注意力机制 (Sentiment Words Incorporated Interactive Attention, 以下简称为 SIIA) 作对比,本文将对对比实验分为三组。第一组为单任务属性抽取系统:

CRF: 使用基础模板,并利用 crfsuite^④ 工具训练 CRF,用于属性抽取。

IHS-RD^[29] 和 **DLIREC^[30]**: 分别为 14 LAPT 和 14 REST 属性抽取第一名评测系统。二者均为融入多种手工特征的 CRF 方法。

WDEmb: Yin 等^[31]提出将词向量、线性上下文以及依存句法作为特征输入 CRF 模型,来优化属性抽取的性能。

EliXa^[32]: 15 REST 属性抽取第一名评测系统。该系统使用多种不同的手工特征,并利用开源工具^⑤训练属性抽取模型,是一种 HMM 方法。

LSTM: Liu 等^[4]提出将预训练的词向量输入 LSTM,并通过全连接得到各个单词标签的概率分布,最终得到预测标签。

DE-CNN: Xu 等^[24]提出同时使用通用词向量

和领域词向量,并利用多层 CNN 进行属性抽取,可大幅提高属性抽取的性能。

Seq2Seq4ATE: Ma 等^[25]首次提出使用端到端模型进行属性抽取,并在实验过程中使用了融入位置的注意力机制。

第二组为多任务联合学习系统。包括: ①属性与情感词联合抽取的多任务学习系统,其利用情感词抽取任务与属性抽取任务相互促进,从而提升属性抽取性能,且同时提高情感词抽取性能;②属性及其情感极性联合学习系统,通过属性抽取任务与情感极性判断任务相互促进,提升属性抽取的性能。

RNCRF: Wang 等^[22]通过 RNN 对依存树进行编码,学习单词的特征表示,并将表示学习的结果输入到 CRF 进行序列标注。

MIN: Li 等^[23]利用双通道 LSTM 实现属性和情感词的联合抽取,并利用记忆力机制实现双通道 LSTM 的记忆交互。

CMLA: Wang 等^[8]提出通过一种成对多层注意力机制 (Coupled Multi-layer Attentions),捕获句子中各个单词之间的直接依存关系和间接依存关系,进行属性词和情感词的联合抽取。

HAST: Li 等^[9]提出使用局部的历史注意力 (Truncated History-Attention) 和选择转移网络 (Selective Transformation Network) 进行属性与情感词的联合抽取。

RINANTE: Dai 等^[26]使用规则扩充训练数据,并使用共享的 BiLSTM 进行属性与情感词的联合抽取。

DORE: Luo 等^[27]提出交叉共享单元,并将其应用到堆叠的对偶 RNN 网络上,同时使用额外的属性长度预测和词典引导的情感极性预测任务进行属性及其情感极性联合抽取。该模型在输入部分同时参考了 Xu 等^[24]提出的双词向量的思想。

第三组为本文所提融合情感词的交互注意力机制的 SIIA 系统及其各部分拆解系统。

G+BCRF、**D+BCRF** 和 **GD+BCRF**: 基线系统使用基于 BiLSTM 的 CRF 系统 (下文简称为 BCRF)。LSTM 的细胞单元使用 Greff 等^[33]提出

① <https://nlp.stanford.edu/projects/glove/>

② <https://github.com/howardhsu/DE-CNN>

③ <https://github.com/facebookresearch/fastText>

④ <https://sklearn-crfsuite.readthedocs.io/en/latest/tutorial.html>

⑤ <https://github.com/ixa-ehu/ixa-pipe-nerc>

的 LSTM 单元,其中 G+BCRF 代表使用通用词向量 GloVe,D+BCRF 代表使用领域词向量,GD+BCRF 同时使用通用词向量和领域词向量。

GD+BCRF+IA: 在 GD+BCRF 的基础上加入交互注意力,其中情感词编码过程中未使用高速网络进行语义空间的映射,且未使用“是否情感词”向量。

GD+BCRF+IA+H: 在 GD+BCRF 的基础上加入交互注意力,且情感词编码过程中使用高速网络进行语义空间的映射,未使用“是否情感词”向量。

GD+BCRF+IA+H+S(SIIA): 本文完整的属性抽取系统,在 GD+BCRF 的基础上加入交互注意力,其中情感词编码过程中使用高速网络进行语义空间的映射,并同时使用“是否情感词”向量。

SIIA+DICT: 在本文完整系统 SIIA 的基础上,将 Wang 等^[8] 手工标注的情感词替换为 MPQA^① 词典,本文选取词典中强主观词语,构成最终的情感词典,并以此词典来获取数据集中每个句子包含的情感词。

3.5 实验结果与分析

3.5.1 总体实验结果分析

本文总体实验结果如表 2 所示,其展示了第一组和第二组以及本文 SIIA 系统的实验结果。由表 2 可发现,本文系统 SIIA 在三个数据集上均取得了最优的性能。

表 2 不同模型实验结果对比 (单位:%)

Model	14 LAPT	14 REST	15 REST
CRF	72.77	79.72	62.67
IHS-RD	74.55	79.62	N/A
DLIREC	73.78	84.01	N/A
WDEmb	75.16	84.97	69.73
EliXa	N/A	N/A	70.05
LSTM	75.00	82.06	64.30
DE-CNN	81.59	N/A	N/A
Seq2Seq4ATE	80.31	N/A	N/A
RNCRF	78.42	84.93	67.74
MIN	77.58	N/A	N/A
CMLA	77.80	85.29	70.73
HAST	79.52	85.61	71.46
RINANTE	80.16	86.76	69.9
DORE	82.61	N/A	N/A
SIIA	82.49	87.01	74.87

分析第一组单任务学习系统实验结果可知,基础的 CRF 模型在三个数据集上的表现都较差,而相比于其他加入大量手工特征的模型 IHS-RD,DLIREC,加入手工特征的模型都较基础的 CRF 模型有所提升,由此可说明手工特征可在一定程度上促进属性抽取。WDEmb 进一步使用词向量作为 CRF 模型的输入,性能得到进一步的提升,因此词向量对于属性抽取任务来说是一项重要的输入表示。

对于深度学习模型 LSTM,其在各个数据集上的表现均未能取得较好的效果,而加入双词向量的多层 CNN 模型和融入位置注意力的端到端神经网络模型均能取得较好的效果,由此可说明双词向量和注意力对于属性抽取的有效性。

分析第二组属性与情感词联合抽取模型可发现,相比融入依存树编码的 RNCRF,使用双通道注意力的 MIN,使用多层成对注意力的 CMLA 以及使用历史局部注意力的 HAST,在未加入任何手工特征的情况下均能取得与 RNCRF 可比的性能或更优的性能。而 DORE 在设计的交叉共享注意力的基础上,在 14LAPT 上取得了最优的效果。由此可说明,设计能捕获特定信息的注意力模型有利于属性抽取。此外,RINANTE 利用规则扩充训练数据进行属性抽取的研究,并在 14 REST 上取得了最优的性能,由此可说明数据量对于深度学习模型的重要性。

其中,相比于第一组单独使用 LSTM 的单任务模型,第二组使用 RNN 编码依存树的 RNCRF 联合学习模型以及双通道的 LSTM 联合学习模型效果均优于单独使用 LSTM 的单任务模型。由此可发现,属性与情感词的联合抽取有助于属性的抽取,由于二者呈现出一种相互促进的关系,使得属性抽取与情感词抽取都能取得较单任务学习更优的效果。而 DORE 同时利用了双词向量、注意力机制以及多任务联合学习,该模型在结合以上三个研究点优势的基础上,于 14LAPT 上取得了最优的性能。

本文系统 SIIA 相较于 14 LAPT 最优系统 DORE 取得了可比的效果,相较于 14 REST 最优系统 RINANTE 和 15 REST 最优系统 HAST 取得了最优的性能,其属性抽取性能分别提升 0.25% 和 3.41%。由于本文系统 SIIA 在输入层使用了通用词向量和领域词向量,且本文系统融入情感词的交互注意力,使其可在情感词的辅助建模中促进属性

① <http://mpqa.cs.pitt.edu/>

抽取,如本文引言例 2 所示,可利用情感词“excellent”和“great”辅助模型发现属性词“service”和“decor”。此外,本文同时结合双词向量和融入情感词的交互注意力思想,可在少量的数据上取得较好的效果,相较于属性与情感词联合抽取系统,并使用大量数据的 RINANTE 具有一定的优越性。

3.5.2 系统拆解实验结果分析

表 3 展示了本文系统各部分拆解实验结果,即第三组实验结果。由拆解实验结果对比分析发现,相较于使用通用词向量的 G+BCRF 与使用领域词向量的 D+BCRF,同时使用通用词向量和领域词向量的 GD+BCRF 性能得到较大程度的提高,以此可进一步验证上一节所提到的双词向量的有效性。原因在于,使用双词向量可使句子中的通用词汇和领域词汇得到更加精确的表示,比如,对于通用词汇“and”,使用大量数据训练出的通用词向量可更加精确地表达其含义,而对于领域词汇“speed”(电脑领域)和“drink”(餐馆领域),则更加倾向于使用领域数据训练出的领域词向量精确地表达其含义。对于电脑领域的词汇“speed”,其表达的意义为电脑或处理器的速度,而在通用领域,其表达的速度一般为每小时多少公里;对于餐馆领域的词汇“drink”,其表达的意义为饮料,而在通用领域,其表达的意义一般为喝。因此,针对上述分析,本文同时利用通用词向量和领域词向量,以精确地表达句子中不同词汇的意义。

表 3 系统拆解实验结果对比 (单位:%)

Model	14 LAPT	14 REST	15 REST
G+BCRF	76.96	84.11	69.11
D+BCRF	78.36	84.86	69.45
GD+BCRF	80.72	85.22	71.03
GD+BCRF+IA	81.52	85.98	73.12
GD+BCRF+IA+H	81.93	86.25	74.05
SIHA+DICT	81.82	86.35	74.25
SIHA	82.49	87.01	74.87

为了验证本文所提融合情感词的交互注意力机制的有效性,本文在 GD+BCRF 的基础上加入交互注意力,并且不使用高速网络进行语义空间的映射,同时在网络的输入层,不使用“是否情感词”向量(见 GD+BCRF+IA)。实验结果证明,加入交互注意力之后,GD+BCRF+IA 比基础模型 GD+BCRF

在三个数据集上 F_1 值分别提升 0.8%、0.76% 和 2.09%。由此可证明融入情感词的交互注意力机制的有效性。该注意力机制将句子中所有的情感词按顺序排列,并利用文本与情感词的交互注意力,通过注意力权重,捕获情感词对于文本中每个词项的重要程度,将情感词信息融入文本的每个词项表示中,从而达到以情感词辅助属性词抽取的效果。例如,本文引言例 2 中将情感词“excellent”和“great”融入“service”和“decor”的表示中,从而实现以情感词辅助发现属性词。

为了进一步验证本文所提高速网络进行语义空间映射的有效性,在 GD+BCRF+IA+H 实验中,对于排列的情感词编码过程,使用高速网络进行语义空间的映射,以将其转换为与原始文本使用 LSTM 编码后相同或相似的语义空间,从而使注意力相似度矩阵来自相同的语义空间向量。由于在编码的过程中,待判别句子使用 BiLSTM 进行编码,该编码考虑了时序的信息,而对于排列的情感词,由于各个单词之间没有联系,因此无法使用考虑时序信息的 BiLSTM 进行编码,而使用 tanh 函数激活的全链接神经网络进行编码。考虑到不同的编码方式可能会导致语义空间的不同,因此,本文使用高速网络进行排列情感词的语义空间映射,以期望将其转化为与带判别句子相同或相似的语义空间。实验结果证明了高速网络语义空间映射的有效性,相较于未使用高速网络的 GD+BCRF+IA 模型,在三个基础数据集上 F_1 值分别提升 0.41%、0.27% 和 0.93%。

在本文完整的系统 GD+BCRF+IA+H+S (简称为 SIHA) 中,其进一步在输入层使用了“是否情感词”向量,相较于未使用“是否情感词”向量的模型 GD+BCRF+IA+H,SIHA 在三个数据集上 F_1 值分别提升 0.56%、0.76% 和 0.82%。使用“是否情感词”向量,在编码过程中让句子知道哪些位置的词为情感词,从而起到类似于位置信息的作用,例如,例 2 中,通过“是否情感词”向量,模型可感知在第四时刻和第九时刻的单词为情感词,且模型在学习的过程中可间接感知属性词与情感词之间的距离,从而使二者建立初始弱相关关系。同时,经过观察与分析发现,原始数据中大多数的属性词都不含有情感词,只有少部分属性词会包含情感词,例如“fresh mozzarella”包含情感词“fresh”。因此,可利用“是否情感词”在编码过程中捕获上述信息,以达到促进属性抽取的效果。

为了进一步验证本文所提方法的有效性,本文

在 SIIA+DICT 实验中将 Wang 等^[8]手工标注的情感词替换为 MPQA 情感词典,本文选取 MPQA 词典中的强主观词构成情感词典,以此获取数据集中每个句子中所包含的情感词。实验结果表明,使用情感词典获取的情感词方法相较于基础模型 GD+BCRF,在三个数据集上 F_1 值分别提升 1.1%、1.13% 和 3.22%。相较于其他前沿方法,使用情感词典的方法能达到最好的效果(15REST)或可比的效果(14LAPT 和 14REST)。由于在实际使用情况中,多数情况下只能部分标注情感词或不能手工标注情感词。因此,本文增设此组实验证明,在无法完全手工标注情感词的情况下,使用情感词典获取句子中包含的情感词可成为一种更加方便的选择,并能达到与手工标注情感词可比的效果。

综合以上各部分实验,在原始使用通用词向量的 G+BCRF 基础模型上,本文同时在输入层使用通用词向量和领域词向量,在建模过程中,加入本文所提融入情感词的交互注意力,同时使用高速网络进行语义空间的映射,并在输入层加入“是否情感

词”向量。最终,本文所提完整的模型相较于基础模型 G+BCRF 在三个数据集上 F_1 值分别提升 5.53%、2.90% 和 5.76%。

3.5.3 注意力矩阵可视化分析

[例 4]

评论句子: The service is excellent, the decor is great, and the food is delicious and comes in large portions.

(译文:服务非常好,装饰很棒,食物美味并且量多。)

评价属性: service(服务), decor(装饰), food(食物), portions(分量)

本文将注意力相似度矩阵可视化,以此分析融合情感词的交互注意力的作用。图 2 展示了例 4 的注意力相似度矩阵,其中图中横坐标代表待测试句子,纵坐标代表该句子中所有的情感词。图中颜色的深浅代表注意力相似度分数的大小,颜色越深代表相似度分数越高。

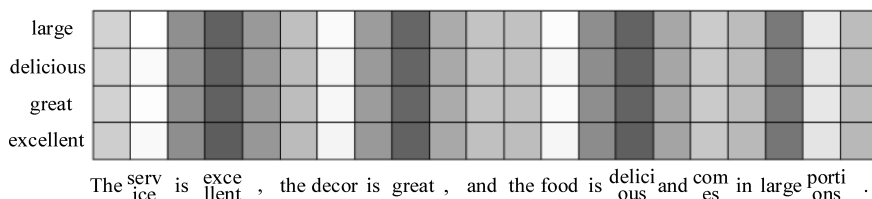


图 2 注意力相似度矩阵图

从图中可发现,在情感词“excellent”“great”“delicious”和“large”的辅助下,文本中属性词与情感词相对于排列的情感词相似度得分呈现两个极端的表现,其中属性词与排列的情感词相似度得分普遍很低(如图 2 中的属性词“service”“decor”“food”和“portions”颜色最浅。),与之相反的是,情感词与排列的情感词之间会得到最高的相似度得分(如图 2 中的情感词“excellent”“great”“delicious”和“large”颜色最深。)由此现象可发现,属性词与情感词呈现出一种负相关性,二者为一种对立的极端。因此本文利用这种对立的相关性,在注意力建模过程中,将与属性词对立相关的情感词信息融入文本的表示中,在此过程中,属性词会融入较少的情感词信息(属性词对应的颜色较浅)而文本中其他词会融入较多的情感词信息(其他词对应的颜色较深),从而达到以情感词辅助发现属性词的目的。

3.5.3 显著性检验分析

为了检验模型性能提升的显著性,本文进行了显著性检验分析实验。本文选取 14 REST 和 15 REST

最优上表现最优的模型 RINANTE 和 HAST, 14 LAPT 上表现次优的 DECNN(由于 DORE 实验以属性及其情感极性对为主要实验结果进行分析,其实验数据的设置与本文不同,且对于属性抽取,该论文未汇报其基础模型的属性抽取实验结果,因此,本文选取 14 LAPT 上表现次优的 DECNN 模型进行显著性检验分析)以及本文模型 SIIA 和其可选的替代模型 SIIS+DICT 进行显著性检验分析。实验中,计算各个模型的基础模型与最终模型在其所使用数据集上的显著性得分 P 值^[34],如表 4 所示。

表 4 显著性得分表

Model	P 值
DECNN	0.081 7
HAST	0.007 0
RINANTE	0.045 3
SIIA+DICT	0.023 8
SIIA	0.017 8

分析表 2 可发现,除 HAST 外,本文模型 SIIA 和 SIIA+DICT 显著性 P 值均优于其他模型(P 值小于 0.05 代表性能提升显著, P 值越小,效果越优),且均小于 0.05,由此可说明本文所提模型对性能的提升是显著的,同时可证明,本文所提模型在实际使用中的效果优于 DECNN 和 RINANTE。除此之外,SIIA 和 SIIA+DICT 的显著性 P 值较为接近(相差 0.006),说明二者的性能提升效果较为相似。由此可证明,在实际使用过程中,可使用更加方便的情感词典代替手工标注的情感词,从而达到模型效果的提升。

4 总结与展望

本文提出一种融合情感词的交互注意力机制,旨在通过交互注意力机制,以文本中所包含的情感词辅助属性词的抽取。此外,为了保证情感词和原始文本语义空间的一致性,本文使用高速网络进行情感词语义空间的映射。实验证明,本文所提方法在三个基准数据集上均能得到一定程度的性能提升。然而本文方法仍然存在一定的局限性,例如,需要使用情感词,在未来的工作中,考虑使用自动发现的情感词来辅助属性词的抽取。

参考文献

- [1] Pontiki M, Galanis D, Pavlopoulos J, et al. SemEval-2014 Task 4: Aspect based sentiment analysis[C]//Proceedings of the 8th International Workshop on Semantic Evaluation, 2014: 27-35.
- [2] Jakob N, Gurevych I. Extracting opinion targets in a single-and cross-domain setting with conditional random fields[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2010: 1035-1045.
- [3] Zhuang L, Jing F, Zhu X Y. Movie review mining and summarization[C]//Proceedings of the 15th ACM International Conference on Information and Knowledge Management. ACM, 2006: 43-50.
- [4] Liu P, Joty S, Meng H. Fine-grained opinion mining with recurrent neural networks and word embeddings [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2015: 1433-1443.
- [5] Toh Z, Su J. NLANGP at SemEval-2016 Task 5: Improving aspect based sentiment analysis using neural network features[C]//Proceedings of the 10th International Workshop on Semantic Evaluation, 2016: 282-288.
- [6] Pontiki M, Galanis D, Papageorgiou H, et al. SemEval-2016 Task 5: Aspect based sentiment analysis [C]//Proceedings of the 10th International Workshop on Semantic Evaluation, 2016: 19-30.
- [7] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv: 1409.0473, 2014.
- [8] Wang W, Pan S J, Dahlmeier D, et al. Coupled multi-layer attentions for coextraction of aspect and opinion terms[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence, 2017: 3316-3322.
- [9] Li X, Bing L, Li P, et al. Aspect term extraction with history attention and selective transformation[J]. arXiv preprint arXiv: 1805.00760, 2018.
- [10] Liu L, Shang J, Ren X, et al. Empower sequence labeling with task-aware neural language model[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018.
- [11] Pontiki M, Galanis D, Papageorgiou H, et al. SemEval-2015 Task 12: Aspect based sentiment analysis [C]//Proceedings of the 9th International Workshop on Semantic Evaluation, 2015: 486-495.
- [12] Hu M, Liu B. Mining and summarizing customer reviews[C]//Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2004: 168-177.
- [13] Blair Goldensohn S, Hannan K, McDonald R, et al. Building a sentiment summarizer for local service reviews[C]//Proceedings of the WWW Workshop on NLP in the Information Explosion Era, 2008, 14: 339-348.
- [14] Wang B, Wang H. Bootstrapping both product features and opinion words from Chinese customer reviews with cross-inducing[C]//Proceedings of the 3rd International Joint Conference on Natural Language Processing, 2008.
- [15] Mei Q, Ling X, Wondra M, et al. Topic sentiment mixture: Modeling facets and opinions in weblogs [C]//Proceedings of the 16th International Conference on World Wide Web. ACM, 2007: 171-180.
- [16] Titov I, McDonald R. Modeling online reviews with multi-grain topic models[C]//Proceedings of the 17th International Conference on World Wide Web. ACM, 2008: 111-120.
- [17] Lin C, He Y. Joint sentiment/topic model for sentiment analysis [C]//Proceedings of the 18th ACM Conference on Information and Knowledge Management. ACM, 2009: 375-384.
- [18] Mukherjee A, Liu B. Aspect extraction through semi-supervised modeling [C]//Proceedings of the 50th Annual Meeting of the Association for Computational

- Linguistics, 2012: 339-348.
- [19] Jin W, Ho H H, Srihari R K. OpinionMiner: a novel machine learning system for web opinion mining and extraction[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2009: 1195-1204.
- [20] Li F, Han C, Huang M, et al. Structure-aware review mining and summarization[C]//Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics, 2010: 653-661.
- [21] Xu L, Liu K, Zhao J. Joint opinion relation detection using one-class deep neural network[C]//Proceedings of the 25th International Conference on Computational Linguistics; Technical Papers, 2014: 677-687.
- [22] Wang W, Pan S J, Dahlmeier D, et al. Recursive neural conditional random fields for aspect-based sentiment analysis[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2016: 616-626.
- [23] Li X, Lam W. Deep multi-task learning for aspect term extraction with memory interaction[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2017: 2886-2892.
- [24] Xu H, Liu B, Shu L, et al. Double embeddings and CNN-based sequence labeling for aspect extraction[J]. arXiv preprint arXiv: 1805.04601, 2018.
- [25] Ma D, Li S, Wu F, et al. Exploring sequence-to-sequence learning in aspect term extraction[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 3538-3547.
- [26] Dai H, Song Y. Neural aspect and opinion term extraction with mined rules as weak supervision[J]. arXiv preprint arXiv: 1907.03750, 2019.
- [27] Luo H, Li T, Liu B, et al. DOER: Dual cross-shared RNN for aspect term-polarity co-extraction[J]. arXiv preprint arXiv: 1906.01794, 2019.
- [28] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization[J]. arXiv preprint arXiv: 1409.2329, 2014.
- [29] Chernyshevich M. IHS R&D belarus: Cross-domain extraction of product features using crf[C]//Proceedings of the 8th International Workshop on Semantic Evaluation, 2014: 309-313.
- [30] Toh Z, Wang W. DLIREC: Aspect term extraction and term polarity classification system[C]//Proceedings of the 8th International Workshop on Semantic Evaluation, 2014: 235-240.
- [31] Yin Y, Wei F, Dong L, et al. Unsupervised word and dependency path embeddings for aspect term extraction[J]. arXiv preprint arXiv: 1605.07843, 2016.
- [32] Vicente I S, Saralegi X, Agerri R. EliXa: A modular and flexible ABSA platform[C]//Proceedings of the 9th International Workshop on Semantic Evaluation, 2015: 748-752.
- [33] Greff K, Srivastava R K, Koutník J, et al. LSTM: A search space odyssey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(10): 2222-2232.
- [34] Dror R, Baumer G, Shlomov S, et al. The Hitchhiker's guide to testing statistical significance in natural language processing [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018: 1383-1392.



程梦(1994—),硕士研究生,主要研究领域为属性级情感分析、属性抽取。
E-mail: dedream@outlook.com



尉桢楷(1995—),硕士研究生,主要研究领域为属性级情感分析、属性抽取。
E-mail: zkyl666@gmail.com



洪宇(1978—),通信作者,博士,教授,主要研究领域为信息检索、信息抽取。
E-mail: tianxianer@gmail.com