

文章编号: 1003-0077(2022)03-0136-10

结合多种注意力机制的方面词提取方法

张名芳, 相艳, 邵党国, 熊馨

(昆明理工大学 信息工程与自动化学院, 云南 昆明 650504)

摘要: 方面词提取是方面级情感分析中最重要的子任务之一, 其旨在从评论文本中找出意见目标。当前对于方面词提取主要使用卷积神经网络(Convolutional Neural Networks, CNN)和双嵌入的方法, 但传统的 CNN 模型受限于卷积核感受野, 不能很好地获取全局信息。为此, 该文提出了一种基于双嵌入和多种注意力的方面词提取模型。联合使用 non-local 网络能够更好地捕获长范围依赖关系, 使用与跳跃连接相结合的空间注意力能够更好地捕获文本的字符特征。该文模型在 Laptop 数据集和 Restaurant 数据集上分别进行了实验, F_1 值分别为 83.39% 和 76.26%。与多个基线模型相比, 该文提出的模型性能更优。

关键词: Non-local 网络; 空间注意力; 方面词提取; 跳跃连接; 双嵌入

中图分类号: TP391

文献标识码: A

Aspect Terms Extraction Based on Multiple Attention Mechanisms

ZHANG Mingfang, XIANG Yan, SHAO Dangguo, XIONG Xin

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650504, China)

Abstract: Aspect terms extraction is the task of extracting the entity attributes commented in the opinions, which is a sub-task of aspect-based sentiment analysis. To improve the current solution base on convolutional neural networks (CNN), this paper proposes an aspect terms extraction model based on double embedding and multiple attentions. It better captures the long-range dependencies by a joint Non-local networks, betters capture the characters in the text by the spatial attention combined with jump connections. Experimented on the laptop dataset and the restaurant dataset, the proposed method achieves F1 values of 83.39% and 76.26%, respectively, better than multiple baseline models.

Keywords: Non-local networks; spatial attention; aspect terms extraction; jump connections; double embedding

0 引言

方面级情感分析的目的是从非结构化评论中提取结构化意见(方面词和观点术语), 并确定其情感极性(正面、负面或中性)。例如, 在评论“Boot time is super fast, around anywhere from 35 seconds to 1 minute.”中, 客户意见可以概括为<Boot time, fast>, 它的情绪极性为正面。方面词提取旨在从评论文本中找出意见目标, 即“Boot time”, 它描述了计算机“operating system”的属性, 而“fast”是它

的观点术语。

目前用于方面词提取任务的方法, 主要有无监督的方法和有监督的方法。对于无监督方法, Hu 和 Liu^[1] 使用频繁模式挖掘来提取方面词, Qiu 和 Liu^[2-3] 提出使用基于规则的方法, 其思想是基于评论语句的句法或依存结构手动设计一些规则, 以通过特定的规则来提取方面词和观点术语。随后基于文档主题生成模型(Latent Dirichlet Allocation, LDA)^[4-7] 开始被广泛用于产品评论方面词的提取。Yin 等^[8] 提出了 WDEmb, 使用线性上下文嵌入和依赖路径嵌入作为输入来增强无监督条件随机场

收稿日期: 2020-05-03 定稿日期: 2020-06-15

基金项目: 国家自然科学基金(61462054, 61732005, 61672271, 61741112); 云南省自然科学基金(2017FB098); 国家博士后面上科学基金(2016M592894XB); 云南省科技厅(2015FB135); 云南省重大科技项目(2018ZF017)

(Conditional Random Field, CRF)的效果。在最近的研究中,变分自编码和神经网络也被应用到了方面词提取任务中,Liao 等^[9]提出了一种新颖的神经网络,能够结合全局和局部表示来发现方面词(或主题),文章使用了长短时记忆网络(Long Short-Term Memory, LSTM)用来构建文本的局部特征信息,使用典型的变分自编码(VAE)的编码层构建文本的全局特征,之后在 VAE 的解码层与 LSTM 相结合,以期能从文本的局部特征中提取到有用的信息。这些无监督模型能够在缺少人工标注的数据时,从产品评论中抽取具有一定意义的方面词或方面属性,但是往往无法充分利用文本的特征信息,且无法找出精确的方面词。

对于有监督的方法,大多数的研究者将其视为序列标注问题。在传统的序列模型中,例如条件随机场^[10-11]因能联合考虑相邻词的标注结果而受到较多关注,但缺点是无法有效地利用词义特征以及单词间的依赖关系特征。LSTM^[12-13]和诸如支持向量机的分类模型也被用于方面词的提取。Li 等^[14]使用了方面词检测历史来帮助预测当前时刻可能的方面词,当评论中出现多个方面词时,也可以利用它来识别后续出现的不常见的方面词。同时文章还用了意见摘要信息来加强方面词和观点术语之间的相关关系。Wang 等^[15]使用多层注意力网络联合提取方面词和观点术语,其中每一层都由张量算子的几个注意力组成。它在不使用句法信息的情况下对方面词-观点术语关系进行建模,使这两个任务可以通过注意力机制共享信息。Xu 等^[16]使用双重嵌入,将一般嵌入和特定领域嵌入以拼接的方式作为 CNN 模型的输入,并让 CNN 模型确定哪些嵌入具有更有用的信息。实验表明,与单独使用通用领域嵌入或特定领域嵌入相比,双嵌入机制在方面词提取上具有更好的性能,但文章没有探讨当不同领域的词向量输入权重不同时,对于模型最终结果的影响。Shu 等^[17]使用一种称为受控 CNN(Ctrl)的改进的 CNN 进行有监督的方面词提取,Ctrl 具有两种类型的控制模块,通过异步参数更新,它可以防止过拟合并显著提高 CNN 的性能。Luo 等^[18]提出了一种以注意力机制为基础的 Dual cross-shared RNN 框架(DOER)以同时生成输入句子的所有方面词提取和评论的极性,文章分别使用一个双向循环神经网络以提取每个任务的各自表示形式,以及一个交叉共享的单元来考虑它们之间的关系。

通过对现有模型的结果分析,发现以上的方面

词提取模型仍然存在不足,CNN 模型在运行效率上和其他模型相比具有明显的优势,但 CNN 模型在运算时,受到卷积核的限制,只考虑了卷积核范围内各元素间的相互影响,从而导致最终提取的特征缺失了文本的部分全局信息;同时卷积计算有时对于不同位置的特征的关注度不够,从而对网络的最后输出结果有影响。Attention 模型在进行特征提取的时候,能够根据具体的任务关注特定的信息,从而可以在一定程度上去除不相关信息对模型训练的影响,但也会由于特征聚合带来一定的干扰。

基于以上原因,本文提出了一种基于双嵌入的使用 Non-local 网络^[19]和结合跳跃连接空间注意力的方面词提取模型。通过 Non-local 网络能够获得文本的长范围依赖关系信息,而空间注意力^[20]能够保留文本中局部最重要的特征和平均特征。该模型将 CNN 模型、LSTM 模型和注意力模型各自的优点进行融合,能够提取到有效且全面的特征信息。

本文的贡献总结如下:

(1) 由于传统 CNN 模型中受卷积核范围的限制,本文采用了 Non-local 网络来获取更大的感受野,以便充分考虑文本全局信息间的相互影响,帮助模型获取全局特征。

(2) 采用了与跳跃连接相结合的空间注意力机制,能够使模型对局部信息给予更多的关注,从而使方面词的界限更加清晰。

(3) 将 Non-local 网络和空间注意力网络进行了结合,使模型通过 Non-local 网络在长范围内依赖关系信息,又通过空间注意力网络获得对某些字符更大的关注,使得最终的模型得到了更好的 F_1 值。

1 方法

本节先给出本文模型的整体框架,如图 1 所示。本模型包括四个部分:①嵌入层,将不同的权重分配给一般词向量和特定领域词向量作为模型的嵌入;②Non-local 网络层,由注意力机制来捕获单词间长距离依赖关系;③空间注意力层,模型通过对不同位置的关注来获取对方面词提取最有用的特征;④Bi-LSTM 层及输出层,与 CNN 相比,它能够获取文本的语序信息,并且在短句中能够较好地捕获方面词之间的前后依赖关系,最后使用全连接层和 softmax 函数输出。本节将详细地介绍模型的各个组成部分。

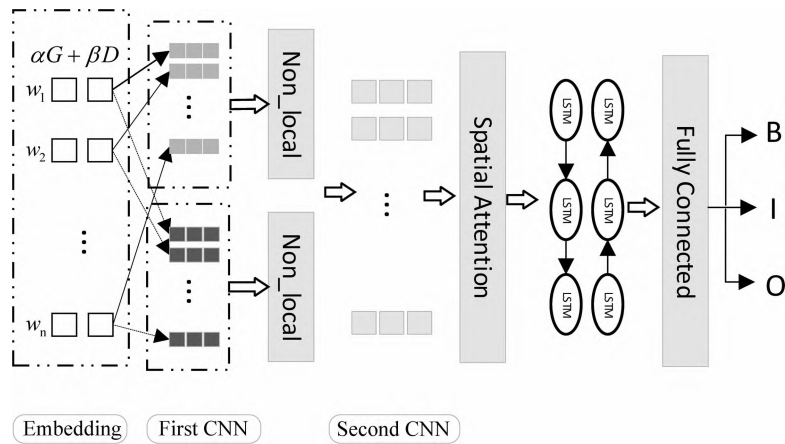


图1 模型整体框架

1.1 任务描述

对于一个特定领域的评论文本,表示为 $S = \{w_1, w_2, \dots, w_i, \dots, w_N\}$, 其中 N 是最大句子长度。对于任意单词 $w_i \in S$, 方面词提取的任务是找到与其对应的标签 $t_i \in T$, 其中 $T = \{B, I, O\}$, “B”“I”“O”分别代表方面词的开头、方面词内部的词和其他词。例如, “Boot/B time/I is/O super/O fast/O, /O around/O anywhere/O from/O 35/O seconds/O to/O 1/O minute/O. /O”是带有标签的句子, 其中方面词为“Boot time”, 这种 BIO 的编码方式普遍应用于自然语言处理(Natural Language Processing, NLP)任务。

1.2 嵌入层

受 Xu 等^[16]的启发, 本文使用双重嵌入作为初始词嵌入。双重嵌入包含两种类型: 通用词嵌入和领域词嵌入, 它们的区别在于是否由特定领域文本构建的语料库训练所得。对于初始词嵌入 $x = \{w_1, w_2, \dots, w_i, \dots, w_n\}$, 每个单词 w_i 由特征向量 $H \in \mathbb{R}^{d \times |v|}$ 初始化, 其中 $d=300$ 为词向量维度, v 是单词表的大小, n 为句子最大长度, 并且对于每个句子, 使用 0 填充使之与长度最大的句子对齐。 $H \in \mathbb{R}^{d \times |v|}$ 由通用词向量嵌入 $G \in \mathbb{R}^{d \times |v|}$ 和特定领域词嵌入 $D \in \mathbb{R}^{d \times |v|}$ 以不同权重相加获得如式(1)所示。

$$\left. \begin{aligned} H &= \alpha G + \beta D \\ \alpha + \beta &= 1 \end{aligned} \right\} \quad (1)$$

其中 G 和 D 的词向量维度为 300。对于单词表外的词, 采用随机生成。

1.3 Non-local 网络层

传统的 CNN 模型受卷积核大小的限制, 只考虑

卷积核范围内各元素之间的相互影响, 通过堆叠卷积核来获取较大的感受野, 忽略了卷积核范围之外的元素对卷积核内元素的影响, 是典型的局部操作。而在文本处理中, 单词间往往存在较长距离的依赖关系。因此, 如果只考虑卷积核范围内元素之间的相互关系, 将会使得最终提取的特征中丢失部分有意义的信息。受 Wang 等^[19]的影响, 原本的非局部注意力是基于图片滤波领域的非局部均值滤波操作思想提出的, 它通过直接计算两个位置之间的相关系数来表示它们之间的依赖关系。本文引入非局部注意力机制, 用于直接计算两个单词特征间的相互关系, 得出原评论与自身的相关系数矩阵, 使得网络模型能够充分考虑到卷积核之外的特征对卷积核内部特征的影响, 从而帮助模型获得更加全面的信息。

卷积神经网络能够很好地描述数据的局部特征, 因此在 Non-local 网络层之前, 本文使用 CNN 层进行局部特征的提取, 按照如下形式进行卷积操作, 如式(2)所示。

$$\left. \begin{aligned} F_j^{(l)} &= f^{i \times i}(x) + b^{(l)} \\ l &= 1 \end{aligned} \right\} \quad (2)$$

第一层 CNN 为并列的两个卷积操作, 本文采用两种不同的卷积核, 以获取不同的视野。其中, $i \in \{3, 5\}$, $f^{3 \times 3}$ 表示使用大小为 3 的卷积核对输入的文本向量进行卷积操作, $b^{(l)}$ 为偏置, l 表示第 l 层卷积, $j = (1, 2)$ 为第一层卷积的序号, 滤波器个数均为 128。同时, 在卷积后, 应用激活层将线性的输入数据转换为非线性数据, 使得数据在拟合过程中具有一定的区分能力。在本文模型中选用 ReLU 作为激活函数, 并在每一个卷积操作后使用 Non-local 网络层提取全局信息。该网络的具体结构如图 2 所示。

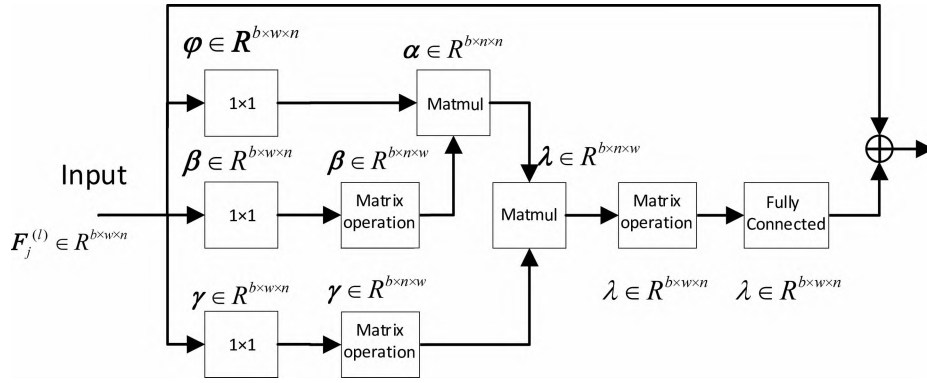


图 2 Non-local 网络的结构

图 2 中各参数的具体含义如下： n 表示输入文本的长度， w 表示特征的维度， φ 、 β 和 γ 表示三个不同的通道，即将同一数据送入不同的路径进行运算。如图 2 所示，将 CNN 模型中卷积之后的结果分别送入 φ 、 β 和 γ 通道，进行卷积核为 1×1 的卷积操作，并将 β 通道的数据转置，接着将 φ 、 β 通道的结果进行矩阵相乘以获得位置对应矩阵 $n \times n$ ，该矩阵中的每一个元素表示不同位置元素之间的相关系数。通过 γ 通道对原始数据进行卷积操作以获得每个元素的对应输入 $g(x_j)$ ，如式(3)所示。

$$y = \frac{1}{D(x)} \sum_{j} f(x_i, x_j) g(x_j) \quad (3)$$

这里， x_i 是输入， y 是输出， i 是当前元素对应的位置索引， j 是文本中其他元素对应的位置索引。 $f()$ 是计算任意两个元素之间相关系数的函数， $g(x_j)$ 表示位于 j 处的元素输入， $D(x)$ 为归一化常量。即 f 用来衡量不同位置元素之间的相关系数，通过求解当前元素和其他所有元素的相关系数 $f(x_i, x_j)$ ，并与相对应元素的输入 $g(x_j)$ 相乘，作为

j 处元素对 i 处元素的贡献，如式(4)、式(5)所示。

$$f(x_i, x_j) = \beta(x_i)^T \varphi(x_j) \quad (4)$$

$$D(x) = \sum_{j} f(x_i, x_j) \quad (5)$$

上式中， x_i 和 x_j 分别表示位于不同位置的两个元素输入， φ 和 β 分别表示对元素的输入进行卷积操作， $D(x)$ 表示对输入的所有元素之间的相关系数进行求和，作为归一化常量。

最后，本文对两组 Non-local 网络层的结果进行拼接，作为下一层的输入，如式(6)所示。

$$f^n = [y_1; y_2] \quad (6)$$

其中， y_1 和 y_2 为两个卷积操作经过 Non-local 网络输出的结果。

1.4 空间注意力层

原本的空间注意力^[20]是指对图像进行特征提取时，找到更加重要的像素特征位置的机制。因为文本设置为二维，本文在这里进行了修改。该网络的具体结构如图 3 所示。

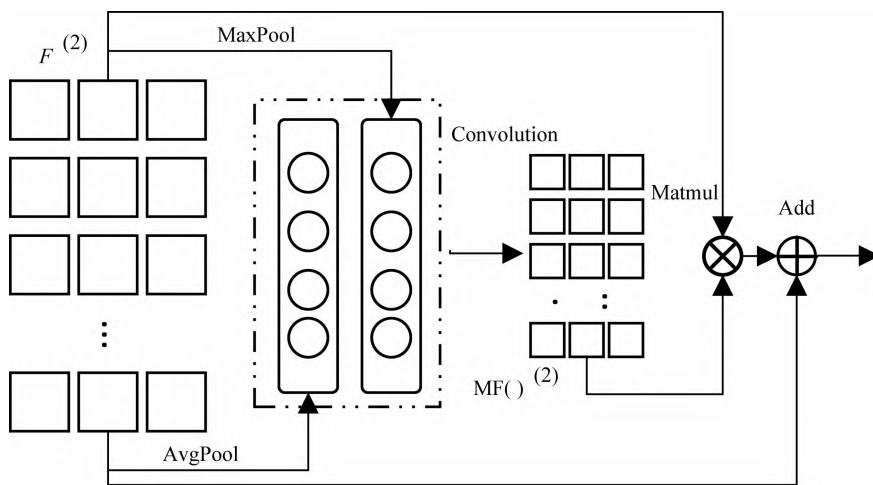


图 3 空间注意力层的结构

图3中各参数的具体含义如下：为了更高效地提取有用信息，本文使用CNN进行进一步的特征提取，如式(7)所示。

$$F^{(l)} = f^{i \times i}(f^n) \quad (7)$$

其中， $i=5$ 表示使用 5×5 的卷积核对输入的文本向量进行卷积操作，滤波器个数为256个。随后分别使用最大池化MaxPool和平均池化AvgPool保留文本中局部最重要的特征和平均特征，然后再对池化的结果进行卷积，模拟这些特征的重要程度，从而得出对每个输入字的注意力大小，如式(8)所示。

$$M(F^{(2)}) = f^{7 \times 2}([\text{AvgPool}(F^{(2)}); \text{MaxPool}(F^{(2)})]) \quad (8)$$

AvgPool、MaxPool分别表示平均池化操作和最大池化操作，卷积核大小为 7×2 。最后，本文对空间注意力提取后的结果进行跳跃连接，即提取空间特征前与提取空间特征后的结果相加，以增加模型的拟合能力，如式(9)所示。

$$sa = M(F^{(2)}) \times F^{(2)} + F^{(2)} \quad (9)$$

1.5 Bi-LSTM及输出层

由于句子中双向的信息均有助于序列建模，因此本文采用既能捕获文本双向信息又能处理梯度消失问题的Bi-LSTM网络来进行句子上下文特征的提取。如式(10)~式(13)所示。

$$\vec{h}_l = \text{LSTM}(a^{<l-1>}, sa) \quad (10)$$

$$\overleftarrow{h}_l = \text{LSTM}(a^{<l-1>}, sa) \quad (11)$$

$$h_l = [\vec{h}_l, \overleftarrow{h}_l] \quad (12)$$

$$h = [h_1, h_2, h_3, \dots, h_n] \quad (13)$$

其中， $a^{(i)}$ 表示当前记忆单元的隐藏层状态。 \vec{h}_l 和 \overleftarrow{h}_l 分别表示Bi-LSTM的双向信息，在第*i*个单词位置在过去和未来记忆网络的隐藏状态。 $h_l \in R^{n \times s}$ 表示两个方向的隐藏状态的组合，其中， n 为句子长度， s 表示隐藏层的维度。

方面词提取是序列标注任务，本文采用了softmax函数输出序列标注的结果，如式(14)~式(15)所示。

$$y = \sigma(hw + b) \quad (14)$$

$$s_i = \frac{e^{y_i}}{\sum_i e^{y_i}} \quad (15)$$

其中， $w \in R^{s \times \text{tags}}$ 、 $b \in R^{n \times \text{tags}}$ 为全连接层的参数。

本文采用交叉熵函数计算损失，通过模型训练使损失函数最小化，如式(16)所示。

$$L = -\frac{1}{n} \sum_{i=1}^n [t_i \ln s_i + (1 - t_i) \ln(1 - s_i)] \quad (16)$$

其中， s_i 为预测类别， t_i 为实际类别。

2 实验与分析

2.1 数据集

本文在SemEval挑战的两个基准数据集上进行了实验，数据集描述如表1所示。第一个数据集来自SemEval-2014任务4子任务1的笔记本电脑评论领域，该数据集总共有3845条数据。第二个数据集来自SemEval-2016任务5子任务1的饭店评论领域，该数据集总共有2676条数据。

表1 数据集描述

数据集	训练集		测试集	
	句子	方面词	句子	方面词
SemEval-14 Laptop	3 045	2 358	800	654
SemEval-16 Restaurant	2 000	1 743	676	622

2.2 实验参数设置

本文进行实验的词向量维度均设为300。如果在使用的词向量中未找到当前训练的单词，则由随机初始化的生成向量来表示。其他超参数如表2所示。

表2 模型的超参数设置

超参数	数值	
Mini-batch大小	128	
CNN层数	2	
Bi-LSTM层数	1	
空间注意力层数	1	
Non-local网络层	1	
Learning rate	0.000 1	
optimization	Adam	
自选参数	Laptop	Restaurant
α	0.6	0.2
β	0.4	0.8

2.3 评估方法

本文所采用的评价指标是方面词作为一个整体的准确率 P 、召回率 R 和 F_1 -score。其中每一个类别的指标的计算如式(17)、式(18)所示。

$$P = \frac{TP}{TP + FP} \quad (17)$$

$$R = \frac{TP}{TP + FN} \quad (18)$$

$$F1 = \frac{2PR}{P + R} \quad (19)$$

其中, TP 是真阳的数目, FP 是假阳的数目, FN 是假阴的数目。

2.4 对比模型

为了验证本文所提模型的有效性, 本文从主流的深度学习模型中, 选择以下与本文相关的具有代表性的模型作为 baseline 对比。

CRF: 是使用词向量嵌入的传统条件随机场。

IHS_RD: 是 Chernyshevich^[10] 在 SemEval-2014 任务中提交的基于 IHS Goldfire 语言处理器的方面词提取模型, 其在 CRF 模型中使用了丰富的词汇、句法和统计功能。

WDEmb: 是 Yin 等^[8] 提出的以词嵌入、线性上下文嵌入和依赖路径嵌入作为输入的增强型 CRF。

RNCRF: 由 Wang 等^[21] 提出, 是具有基于依赖关系树的递归神经网络和 CRF 进行方面词和观点术语共同提取的联合模型。

MIN: 由 Li 和 Lam^[13] 提出, MIN 分别使用两个 LSTM 进行方面词和观点术语的提取, 并且他们额外使用了一个 LSTM 网络来判别评论句子中是否含有情感词。

CMLA: 由 Wang 等^[15] 提出, CMLA 使用多层注意力网络联合提取方面词和观点术语, 其中每一层都由张量算子的几个注意力组成。一种注意力用于提取方面词, 另一种注意力用于提取观点术语。它们可以交互学习, 在方面词和观点术语之间双重传播信息。

THA&STN: 由 Li 等^[14] 提出, THA&STN 利用方面词检测历史来帮助预测当前时刻可能的方面词, 当评论中出现多个方面词时, 也可以利用它来识别后续出现的不常见的方面词。

DE-CNN: 由 Xu 等^[16] 提出, DE-CNN 综合使

用通用领域词向量信息与特定领域词向量信息, 使用纯 CNN 进行方面词提取, 并且取得了当前最佳效果。

Ctrl: 由 Shu 等^[17] 提出, Ctrl 使用一种称为受控 CNN(Ctrl)的改进的 CNN 进行有监督的方面词提取。该模型具有两种类型的控制模块, 通过异步参数更新, 提高了 CNN 的性能。该模型在标准方面词提取数据集上实现了最新的结果。

实验结果如表 3 所示。通过观察表 3 可以得到以下结论: 以 F_1 评分作为评价标准, 在机器学习中, CRF 是最原始的序列标注方法, 无法有效地利用词义特征和语序信息, 所以 F_1 值最低, 与 WDEmb 相比, HIS_RD 在条件随机场增添了更加丰富的信息嵌入, 在 Laptop 数据集上分别取得了 0.44% 和 1.15% 的提升, 但仍没有利用语序信息。

表 3 F_1 值比较结果 (单位: %)

模型	数据集	
	Laptop	Restaurant
CRF*	74.01	69.56
HIS_RD	74.55	—
WDEmb	75.16	—
LSTM*	75.25	71.26
RNCRF	78.42	—
CMLA	77.80	72.77
MIN	77.58	73.44
THA&STN	79.52	73.61
DE-CNN*	81.59	74.37
Ctrl*	82.73	75.64
Our Model*	83.39	76.25

注: *表中给出的结果为 5 次实验平均值。

在深度学习的方法中, LSTM 是最常用做序列标注的网络模型, RNCRF 在基于依赖关系树的递归神经网络基础上使用了条件随机场作为辅助, 除此之外还使用了丰富的人工特征, 相比单纯的 LSTM, 在 Laptop 数据集上获得了 3.17% 的提升。CMLA 首次在方面词提取任务上使用了注意力机制, 它利用多层耦合注意力网络, 可以利用全局观点信息进行方面词提取, 但由于该模型为方面词与观

点术语联合提取模型,使用注意力机制传播方面词与观点术语之间的特征影响了单任务的结果。MIN 中有两个组件,第一个组件使用了两个 LSTM 来分别提取方面词和观点术语,第二个组件用以识别该句子是否存在情感表达,以确保方面词源于表达了情感的句子,相比 CMLA 在 Restaurant 数据集上提升了 0.67%。THA&STN 能够利用观点摘要信息有助于加强与方面词之间的联系,同时它利用方面词检测的历史信息来帮助当前时刻的方面词预测,对于检测同一评论中出现的多个方面词的效果更好。DE-CNN 使用卷积神经网络进行方面词提取,证明了卷积神经网络在序列标注方面的有效性。DE-CNN 引入了外部知识,改善了在不同领域下单纯使用通用领域特征嵌入可能存在的误差,相对于 THA&STN,在两个数据集上分别获取了 2.07%和 0.76%的提升,但是仍存在对于单词间长依赖关系捕获不足和对特定的字符关注过少的缺点。Ctrl 在 DE-CNN 的基础上使用了两种控制模块,取得了方面词单独提取的最佳效果,但是 Ctrl 仅提升了 CNN 的性能,改善了 CNN 网络过深可能带来的梯度消失问题,相比 DE-CNN 在两个数据集上获得了 1.14%和 1.27%的提升。

具体来说,作为深度学习的代表方法,MIN 的识别性能最低。而 Ctrl[17]在 baseline 中取得了 82.73%和 75.64%的最高 F_1 值。这表明特定领域的外部知识特征对方面词识别的有效性。然而,该模型却没有对文本的词序信息进行处理,同时对卷积核之外的信息没有考虑,因此本文模型与之相比在 F_1 值上获得 0.66%和 0.61%的提升。这是因为本文的模型使用了 Non-local 网络,有助于模型提取到全局信息,Non-local 网络可以直接计算两个位置之间的相关系数来表示它们的依赖关系,使模型可以获得更加丰富的单词间依赖关系,并且不考虑欧氏距离,使得 Non-local 网络可以获得长范围的依赖关系,这种单词间的长依赖关系对于长方面词的提取非常重要。同时,本文使用了引入跳跃连接机制的空间注意力,能够有效关注特定的字符信息,对于提取单个单词构成的方面词也有一定的帮助。

2.5 消融分析

为了验证本文提的方法中多种注意力机制以及

跳跃连接的有效性,本文设计以下模型并在 Laptop 数据集上进行了实验与分析。

CNN-Bi-LSTM: 使用基于双嵌入的 CNN-Bi-LSTM 提取字符信息和语序信息。

CNN-Bi-LSTM + Non-local: 在 CNN-Bi-LSTM 的基础上,只使用 Non-local 网络层获取长范围依赖关系。

CNN-Bi-LSTM + SA: 在 CNN-Bi-LSTM 的基础上,只使用空间注意力获取重要的字符信息。

CNN-Bi-LSTM + Ctrl-SA: 在 CNN-Bi-LSTM 的基础上,使用加入跳跃连接机制的空间注意力网络。

CNN-Bi-LSTM + Non-local + SA: 在 CNN-Bi-LSTM 的基础上,使用 Non-local 网络层获取长范围依赖关系,并使用空间注意力获取重要的字符信息。

表 4 给出 CNN-Bi-LSTM 的实验结果,并将其作为基线方法。其中 CNN-Bi-LSTM 的 F_1 值在 laptop 数据集上最低,这显示了通过多种注意力机制获取全局和局部信息的有效性。

表 4 消融模型实验结果 (单位: %)

Laptop*	P	R	F_1
CNN-Bi-LSTM	83.23	81.19	82.20
CNN-Bi-LSTM + Non-local	83.31	82.42	82.86
CNN-Bi-LSTM + SA	82.49	82.11	82.29
CNN-Bi-LSTM + Ctrl-SA	82.47	82.72	82.59
CNN-Bi-LSTM + Non-local + SA	82.18	83.18	82.67
Our model	83.18	83.34	83.26

注: *每组实验选取最接近平均值的结果,仅保留两位小数。

同 CNN-Bi-LSTM 相比,单独使用 Non-local 网络或者空间注意力机制的模型与联合使用 Non-local 网络和空间注意力机制的模型均能大幅度提高模型的 R 值。尤其当单独使用 Non-local 网络时,模型能够获得更好的 P 值。这是因为相比单纯的 CNN-Bi-LSTM,它能够更好地获得文本单词间的长范围依赖关系,从而使方面词的界限更加清晰,案例如表 5 所示。

表 5 案例 1

①	Exceptionally	easy	to	set	up	—	—	—
gold	O	O	O	B	I	O	O	O
non	O	O	O	B	I	O	O	O
CB-1	O	O	O	B	I	I	I	I

当使用跳跃连接机制的空间注意力,比使用单纯的空间注意力在 R 值的提升上效果更加明显。说明对文本中某些局部信息给予更多关注有助于找到更多的方面词,其案例如表 6 所示。相反单独使用空间注意力或者跳跃连接的空间注意力时,虽然能够提取出更多的信息,但精度却不够, P 值反而有所下降。

表 6 案例 2

	...	easy	and	intuitive	to	setup	or	configure
gold	...	O	O	O	O	B	O	B
ctrl-sa	...	O	O	O	O	B	O	B
CBI	...	O	O	O	O	B	O	O

当联合使用 Non-local 网络和跳跃连接空间注意力机制时,模型的 P 值相对于 CNN-Bi-LSTM 基本一致,同时又获得了更好的 R 值。这可能是由于 Non-local 网络提供了输入文本的全局信息,初步对文本的依赖关系进行了筛选,而后跳跃连接空间注意力机制在此基础上选取最重要和最具普适性的特征,从而提升了预测的准确率,其案例如表 7 所示。

表 7 案例 3

	I	love	the	lit	up	keys	and	screen	display
gold	O	O	O	B	I	I	O	B	I
ctrl-sa	O	O	O	B	O	I	O	B	I
non	O	O	O	O	O	I	O	B	I
our	O	O	O	B	I	I	O	B	I

2.6 超参数分析

本文改变 α 与 β 的比例,以模拟通用领域和特定领域词向量的不同组合对本文提出的模型的提取性能的影响,并在 Restaurant 和 Laptop 两个数据集中给出了结果。为了简化计算,本文将 α 与 β 之和设置为 1,并以 0.1 为步长改变 α 和 β 的值以调整通用领域和特定领域词向量的比率。其余参数与表 2 一致。

在宏观层面上,图 4(a)和图 4(b)都显示出先上升后下降的趋势。可以看出,当大量特定领域信息用作联合嵌入层时,由于通用信息稀少,该模型的性能较差。当在一定范围内逐渐增加通用领域嵌入的比例时,本文提出的模型的性能将会提高。但是,超出适当的范围,本文模型的性能会降低,这是因为太多的通用领域特征可能会稀释特定领域的专有术语的特征信息,从而使联合嵌入层对一些专有名词不太敏感。从图 4 的实验结果可以得出以下结论:本文的算法对 α 和 β 的值基本上是敏感的,并且在不同的数据集中,合适的比例是不同的。

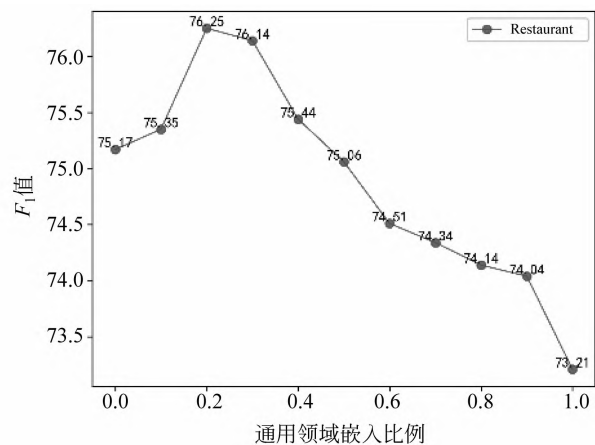
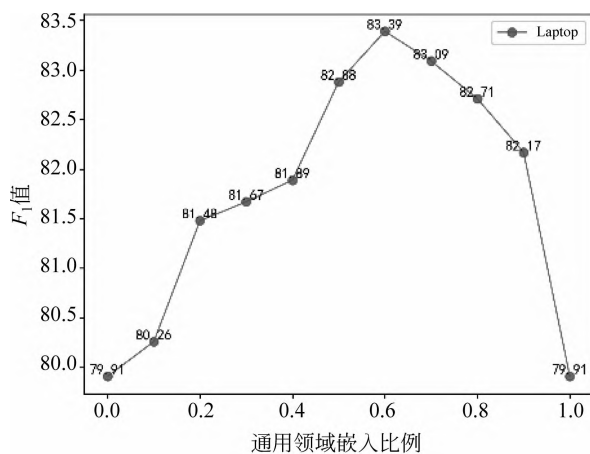


图 4 α 和 β 对模型整体结果的影响
所有结果均为 5 次实验平均值

① gold,ctrl-sa,CBI,non 分别代表真实标签、CNN-Bi-LSTM + Ctrl-SA,CNN-Bi-LSTM,CNN-Bi-LSTM + Non-local,代表句子已结束,不存在单词或标点字符。

3 结论

针对产品评论的方面词提取问题,本文提出了一种结合多种注意力的方面词提取模型。本文模型通过联合通用领域词向量和特定领域词向量来捕获更加有用的文本特征,使用 Non-local 网络层获取长范围依赖关系,使用空间注意力获取重要的字符信息;同时为了缓解网络过深带来的问题和提高模型的拟合能力,本文使用了引入跳跃连接机制的空间注意力机制来提高字符的提取效果。最终,与不同模型的对比结果也验证了本文提出的模型的优越性。

参考文献

- [1] Minqing Hu, Bing Liu. Mining and summarizing customer reviews [C]//Proceedings of the Tenth {ACM} {SIGKDD} International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, August 22-25, 2004.
- [2] Guang Qiu, Bing Liu, Jiajun Bu, et al. Opinion word expansion and target extraction through double propagation[J]. Computational Linguistics, 2011, 37(1): 9-27.
- [3] Qian Liu, Zhiqiang Gao, Bing Liu, et al. Automated rule selection for aspect extraction in opinion mining [C]//Proceedings of the 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015.
- [4] Qiaozhu Mei, Xu Ling, Matthew Wondra, et al. Topic sentiment mixture: modeling facets and opinions in weblogs [C]//Proceedings of the 16th International Conference on World Wide Web, Banff, Alberta, Canada, 2007.
- [5] He Y, Lin C, Alani H. Automatically Extracting Polarity-Bearing Topics for Cross-Domain Sentiment Classification [C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics; Human Language Technologies, 2011.
- [6] Chen Z, Mukherjee A, Liu B. Aspect Extraction with Automated Prior Knowledge Learning [C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014.
- [7] Blei D M, Ng A Y, Jordan M I, et al. Latent Dirichlet Allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022.
- [8] 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.
- [9] Ming Liao, Jing Li, Haisong Zhang, et al. Coupling Global and Local Context for Unsupervised Aspect Extraction [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019.
- [10] Maryna Chernyshevich. Ihs r&d belarus: Crossdomain extraction of product features using crf [C]//Proceedings of the 8th international workshop on semantic evaluation. Dublin, Ireland, 309-313, 2014.
- [11] Toh Zhiqiang, Wang Wenting. DLIREC: Aspect Term Extraction and Term Polarity Classification System [C]//Proceedings of the 8th International Workshop on Semantic Evaluation, 2014.
- [12] Pengfei Liu, Shafiq Joty, Helen Meng. Fine-grained opinion mining with recurrent neural networks and word embeddings [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, Lisbon, 2015.
- [13] Xin Li, Wai Lam. Deep multi-task learning for aspect term extraction with memory interaction [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 2017.
- [14] Xin Li, Lidong Bing, Piji Li, et al. Aspect term extraction with history attention and selective transformation [C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, Sweden, 2018.
- [15] Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, et al. Coupled multi-layer attentions for coextraction of aspect and opinion terms [C]//Proceedings of the 31st Conference on Artificial Intelligence, San Francisco, California, 2017.
- [16] Hu Xu, Bing Liu, Lei Shu, et al. Double embeddings and cnn-based sequence labeling for aspect extraction [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 2018.
- [17] Lei Shu, Hu Xu, Bing Liu. Controlled CNN-based sequence labeling for aspect extraction [J/OL]. arXiv preprint arXiv: 1905.06407, 2019.
- [18] Huaishao Luo, Tianrui Li, Bing Liu, et al. DOER: Dual Cross-Shared RNN for Aspect Term Polarity Co-Extraction [C]//Proceedings of the 57th Conference of the Association for Computational Linguistics, Florence, Italy, 2019.

- [19] Xiaolong Wang, Ross Girshick, Abhinav Gupta, et al. Non-Local Neural Networks [C]//Proceedings of the Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018.
- [20] Sanghyun Woo, Jongchan Park, Joon Young Lee. CBAM: Convolutional Block Attention Module [C]//Proceedings of the Computer Vision, Germany, 2018.
- [21] Wang W, PanS 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, Austin, Texas, USA, 2016.



张名芳(1995—), 硕士研究生, 主要研究领域为数据挖掘、自然语言处理。
E-mail: 1254116691@qq.com



相艳(1979—), 通信作者, 博士, 讲师, 主要研究领域为医学图像处理、自然语言处理等。
E-mail: 50691012@qq.com



邵党国(1979—), 博士, 副教授, 主要研究领域为数据挖掘、文本处理。
E-mail: huntersdg@163.com

第二十八届全国信息检索学术会议 (CCIR2022) 征稿启事

信息检索旨在满足人类在互联网上快速准确地获取信息与知识的需求, 研究成果将支撑国家战略决策, 推动互联网和人工智能领域的发展, 提升整个社会的生产效率, 并对社会生活各个领域产生重大影响。全国信息检索学术会议(CCIR)由中国中文信息学会(CIPS)举办, 一路伴随着中国互联网产业的成长, 是信息检索领域的旗舰会议。

第二十八届全国信息检索学术会议(The 28th China Conference on Information Retrieval, CCIR 2022)将包含一系列学术活动。会议议程除包含传统的海内外知名学者的大会报告、会议论文报告、Poster 交流、评测活动外, 还将组织青年学者论坛、博士生指导论坛以及面向热点研究问题的前沿讲习班等。

第二十八届全国信息检索学术会议(CCIR2022)将于2022年09月16-18日在美丽的山城重庆举行, 此次会议将由中国中文信息学会主办, 重庆理工大学承办。

征文范围具体相关内容可查阅中国中文信息学会官方微信公众号。

征文要求

投稿到 CCIR 2022 的论文必须是未公开发表, 并且不能同时被投稿到其他会议和期刊。

本次会议接收中英文论文。中文论文: 全文不超过 8000 字, 格式要求参照《中文信息学报》论文投稿模版。所有被 CCIR2022 录用的中文论文将以论文摘要集的形式发放给与会者, 会议将评选出优秀学生论文。中文论文模板: <http://jcip.cipsc.org.cn/CN/column/column33.shtml>。英文论文: 格式要求参照 Springer 的论文模板, 长度为 12 页。录用论文将由 Springer LNCS 或 LNAI 出版, EI 索引。

投稿方式

登录 <https://easychair.org/conferences/?conf=ccir2022> 提交

重要日期

- 投稿截止日期: 2022 年 6 月 10 日
- 录用通知日期: 2022 年 7 月 10 日
- 会议举行日期: 2022 年 9 月 16 日-9 月 18 日

(中国中文信息学会信息检索专委会)