

基于交互注意力图卷积网络的方面情感分类

潘志豪 曾碧 廖文雄 魏鹏飞 文松

广东工业大学计算机学院 广东 广州 510006

(pzh@mail2.gdut.edu.cn)

摘要 基于方面的情感分类任务旨在识别句子中给定方面词的情感倾向性。以往的方法大多基于长短时记忆网络和注意力机制,这种做法在很大程度上仅依赖于建模句子中的方面词与其上下文的语义相关性,但忽略了句中的语法信息。针对这种缺陷,提出了一种交互注意力的图卷积网络,同时建模了句中单词的语义相关性和语法相关性。首先使用双向长短时记忆网络来学习句子的词序关系,捕捉句中上下文的语义信息;其次引入位置信息后,通过图卷积网络来学习句中的语法信息;然后通过一种掩码机制提取方面词;最后使用交互注意力机制,交互计算特定方面的上下文表示,并将其作为最后的分类特征。通过这种优势互补的设计,该模型可以很好地获得聚合了目标方面信息的上下文表示,并有助于情感分类。实验结果表明,该模型在多个数据集上都获得了优秀的效果。与未考虑语法信息的 Bi-IAN 模型相比,该模型在所有数据集上的结果均优于 Bi-IAN 模型,尤其在餐厅领域的 REST14, REST15 和 REST16 数据集上,该模型的 $F1$ 值较 Bi-IAN 模型分别提高了 4.17%, 7.98% 和 8.03%;与同样考虑了语义信息和语法信息的 ASGCN 模型相比,该模型的 $F1$ 值在除了 LAP14 数据集外的其他数据集上均优于 ASGCN 模型,尤其在餐厅领域的 REST14, REST15 和 REST16 数据集上,该模型的 $F1$ 值较 ASGCN 模型分别提高了 2.05%, 1.66% 和 2.77%。

关键词: 交互注意力机制;双向长短时记忆网络;图卷积网络;方面情感分类;语义信息;语法信息

中图法分类号 TP391

Interactive Attention Graph Convolutional Networks for Aspect-based Sentiment Classification

PAN Zhi-hao, ZENG Bi, LIAO Wen-xiong, WEI Peng-fei and WEN Song

School of Computer, Guangdong University of Technology, Guangzhou, Guangdong 510006, China

Abstract Aspect-based sentiment classification aims at identifying the sentiment polarity of the given aspect in a sentence. Most of the previous methods are based on long short-term memory network(LSTM)and attention mechanisms, which largely rely on the semantic correlation between aspects and contextual words in the modeled sentence, but ignore the syntactic information in the sentence. To tackle this problem, an interactive attention graph convolutional network(IAGCN) is proposed to model the semantic correlation and syntactic correlation of words in a sentence. Firstly, IAGCN starts with a bi-directional long short-term memory network(BiLSTM) to capture contextual semantic information regarding word orders. Then, the position information is introduced and put it into the graph convolutional network to learn the syntactic information. After that, aspect representation is obtained through mask mechanism. Finally, the interactive attention mechanism is used to interactively calculate and generate the aspect-specific contextual representation as the final classification feature. Through this complementary design, the model can obtain a good contextual representation that aggregates the aspect target information, and is helpful for sentiment classification. Experimental results show that the model achieve a good performance on multiple datasets. Compared with the Bi-IAN model without considering the syntax information, our model are superior to Bi-IAN model on all datasets, especially on the REST14, REST15 and REST16 datasets in the restaurant domain. Our model improves by 4.17%, 7.98% and 8.03% on $F1$ scores respectively compare with the Bi-IAN model. Compared with the ASGCN model, which also takes semantic information and syntax information into account, the $F1$ scores of our model is better than that of the ASGCN model in all datasets except LAP14 dataset, especially on the REST14, REST15 and REST16 datasets in the restaurant domain. Compared with the ASGCN model, the $F1$ scores of our model is increased by 2.05%, 1.66% and 2.77% respectively.

到稿日期:2021-01-23 返修日期:2021-06-08

基金项目:国家自然科学基金(61876043);广东省自然科学基金(2018A030313868);清远市工业高新技术领域技术攻关项目(2020KJJH039)

This work was supported by the National Natural Science Foundation of China(61876043), Natural Science Foundation of Guangdong Province(2018A030313868) and Key Technology Projects in High-Tech Industrial Field of Qingyuan(2020KJJH039).

通信作者:曾碧(zb9215@gdut.edu.cn)

Keywords Interactive attention mechanism, Bi-directional long short-term memory network, Graph convolutional networks, Aspect-based sentiment classification, Semantic information, Syntactical information

1 引言

近年来,各种社交平台和电商平台的用户评论呈爆炸式增长,如何从众多评论中提取出用户对产品或服务在某个方面的具体情感,对商家改进自家产品与服务,以及指导其他用户选购商品都极具现实意义。基于方面的情感分类(Asspect-Based Sentiment Classification, ABSC)指,给定句子和句子中的实体方面,推断该方面的情感倾向性。例如,在“The decor is not special at all but their amazing food makes up for it.”中,对于“decor”的方面,其情感倾向是消极的;而对于“food”方面,其情感倾向是积极的。相比句子级别的情感分类任务,基于方面的情感分类任务是一种更加精细、更具挑战性的情感分类任务,可以捕捉到句子级别的情感分类任务所捕捉不到的句子内部的情感倾向性。

在早期的研究中,基于方面的情感分类任务被当作一般的情感分类任务,大多采用机器学习的方法^[1],通过人工进行大规模的数据预处理,利用情感词典人工设计特征工程并训练分类模型等需要花费大量的时间和精力。随着深度学习技术的发展,各种神经网络被应用在基于方面的情感分类任务中并取得了不错的结果。相比机器学习的方法,基于深度学习的方法可以让神经网络自动学习方面词和上下文之间的语义关系。这些方法^[2-5]大多数基于卷积神经网络(Convolutional Neural Network, CNN)^[6]、长短期记忆网络(Long Short-Term Memory Network, LSTM)^[7]和注意力机制(Attention Mechanism)^[8]来学习句子中的语义表示,并建模方面词和上下文之间的关系。近年来,随着图卷积神经网络(Graph Convolutional Networks, GCN)^[9]的出现,基于方面的情感分类任务也引入了 GCN。研究人员使用语法解析工具来解析句子,利用句子的语法依赖关系树进行信息传递,完成基于方面的情感分类任务^[10-11]。这种基于依赖关系树的图卷积方法可以从语法关系上缩短方面词和观点词之间的距离,并利用句子间的语法关系进行信息传递,但这种方法高度依赖语法解析工具,一旦解析错误就不可避免地会导致错误的积累。

为了充分利用句子的语法依赖,同时缓解由于解析错误而导致的问题,一个简单的解决方法就是将句子的语义关系和语法关系结合。受到 Zhang 等^[11]和 Ma 等^[12]的启发,本文提出了一种基于交互注意力机制的图卷积神经网络的方法(Interactive Attention Graph Convolutional Network, IAGCN),通过使用双向长短期记忆网络(Bi-directional Long Short-Term Memory Network, BiLSTM)、GCN 和交互注意力机制,结合句子与方面的语义信息和语法信息进行方面情感分类。具体来说,首先通过一个 BiLSTM 建模句子之间的词序关系并获得上下文的语义信息,然后在句子的语法依赖关系树上应用 GCN 来捕捉句子的语法信息和单词间的依赖关系,并通过掩码机制分别获取 BiLSTM 和 GCN 的方面词表示,进而通过交互注意力机制,交互建模 BiLSTM 和 GCN 中的方面词表示与上下文表示之间的注意力,最后将它们

拼接整合得到最终的特定方面的上下文表示,用于情感分类。

2 相关工作

基于方面的情感分类任务是情感分类任务中的一个分支,其是情感分类任务中一个细粒度的任务,旨在根据给定的句子和方面词进行情感分类。

基于机器学习的方法^[13-14]一直被用于解决情感分类任务,其需要对文本进行预处理,对数据进行人工标注,对特征进行提取,构建分类模型,但这会耗费大量的人力成本和时间成本。Li 等^[14]通过卡方统计进行特征词的提取,并用 TF-IWF 方法赋予特征词权重,最后用 SVM 方法用于文本分类,有效提高了短文本分类的性能。

随着深度学习的快速发展,许多神经网络也被应用于基于方面的情感分类任务中,其大多数是基于 LSTM 来建模句子中的单词序列。Tang 等^[3]基于 LSTM 能够捕捉单词之间的长距离依赖的特性,将句子分为具有方面词的左侧部分和具有方面词的右侧部分,并使用两个 LSTM 分别对这两个部分进行建模,然后将这两部分的隐状态作为分类的特征。Wang 等^[5]第一次提出结合 LSTM 和注意力机制的模型,通过计算方面词和上下文之间的注意力,来选择句子中与方面相关的重要单词。Ma 等^[12]使用两个 LSTM 分别对上下文和方面词进行建模,并引入交互式注意力网络来交互学习上下文和方面词之间的注意力权重,由此分别生成特定上下文的方面词表示和特定方面的上下文表示,用于分类。Huang 等^[2]使用 LSTM 生成的方面词表示和上下文表示,通过 AOA(Attention-Over-Attention)模型进行交互。

除了基于 LSTM 的方法,还有一些研究人员使用其他的神经网络模型用于解决基于方面的情感分类任务。Tang 等^[4]从问答系统中受到启发,设计了由多个计算层组成的深度记忆网络(deep memory network),以集成方面词的信息,并将其应用于基于方面的情感分类任务中。Li 等^[15]基于 CNN 的方法提出了 T-Net 模型,通过一个组件来生成句子中单词的特定方面目标的表示,同时结合了一种用于保留 LSTM 层中原始上下文信息的机制,弥补了 CNN 不能很好地处理多目标情感的缺陷。Huang 等^[16]结合 CNN 和门控机制,通过参数化的方式将方面信息结合到 CNN 中。Xu 等^[17]基于胶囊网络和交互注意力机制的方法,使用序列卷积来分别提取上下文和方面词的特征,在一定程度解决了特征重叠的问题。Sun 等^[18]提出了一种结合区域卷积神经网络、双向门控循环单元和高速公路网络的模型 RCNN-BGRU-HN,在保留句子时序关系的同时,大大降低了模型训练的时间代价。Yin 等^[19]综合了 CNN 有效提取关键词语和 RNN 捕捉长距离依赖的优点,并结合注意力机制提出了 ATT-DGRU 网络,用于方面级别情感分类任务,有效地提高了检测复杂语句情感的准确性。

上述用于解决基于方面的情感分类任务的方法大多是将句子建模成单词序列,基于 CNN、RNN 和注意力机制的结合,将方面信息嵌入到句子表示中,以建模句子和方面词之间

的关系,很大程度上忽略了句子的语法结构,因此很难找到一些和方面词距离较远的观点单词。为了解决这个问题,一些研究基于图的模型来整合句子的语法结构,如Huang等^[10]提出了一种基于图注意力网络(Graph Attention Network, GAT)来显式地利用单词间的语法依赖关系。Zhang等^[11]通过语法解析工具解析句子来得到依赖关系树,然后在依赖关系树上使用GCN来利用语法信息和单词依赖关系。

3 基于交互注意力的图卷积神经网络

本文模型的目标是给定句子和句子中的方面词,预测方面词的情感倾向性。本文提出的基于交互注意力的图卷积神经网络(IAGCN)的整体框架如图1所示。该框架主要分为以下几层:1)输入层和BiLSTM层,对数据集进行预处理,然后基于GloVe^[20]方法初始化词嵌入,将输入层得到的词向量作为输入,经过BiLSTM网络,建模上下文之间的语义特征,获取句子的序列表示;2)GCN层,将BiLSTM的输出作为GCN的输入,建模上下文之间的语法关系,获取句子的语法特征;3)Mask层,通过掩码机制,分别获得BiLSTM层和GCN层的方面表示;4)方面注意力层,对BiLSTM层输出的上下文

表示和GCN层输出的方面表示进行注意力操作,获取特定方面的BiLSTM层的上下文表示;同理,对GCN层输出的上下文表示和BiLSTM层输出的方面表示进行注意力操作,获取特定方面的GCN层的上下文表示;5)输出层,将特定方面的BiLSTM层的上下文表示和特定方面的GCN层的上下文表示进行拼接,得到最终的特定方面的上下文表示,最后经过一个线性层和softmax层进行情感分类。

3.1 输入层和BiLSTM层

给定一个由 n 个单词组成的句子 $s = \{w_1, w_2, \dots, w_{a+1}, \dots, w_{a+m}, \dots, w_{n-1}, w_n\}$, 其中句中的 w_i 表示句子 s 的第 i 个单词。方面词为 $a = \{w_{a+1}, \dots, w_{a+m}\}$, 方面词的长度为 m 。通过GloVe词向量^[20]初始化句子的词嵌入,得到 $s = \{x_1, x_2, \dots, x_{a+1}, \dots, x_{a+m}, \dots, x_{n-1}, x_n\}$, 其中 $x_i \in \mathbb{R}^{d_e}$, d_e 是单词嵌入的维度。然后把词向量输入到BiLSTM网络,捕捉单词间的词序关系和语义信息,得到隐藏层向量 $H = \{h_1, h_2, \dots, h_{a+1}, \dots, h_{a+m}, \dots, h_{n-1}, h_n\}$, 其中 $h_i \in \mathbb{R}^{2 \times d_h}$ 。为了简化和节省空间,本文用 $LSTM(x_i)$ 和 $LSTM(x_i)$ 分别表示 x_i 执行前向和后向的LSTM操作。具体计算如下:

$$h_i = [LSTM(x_i); LSTM(x_i)], i \in [1, n] \tag{1}$$

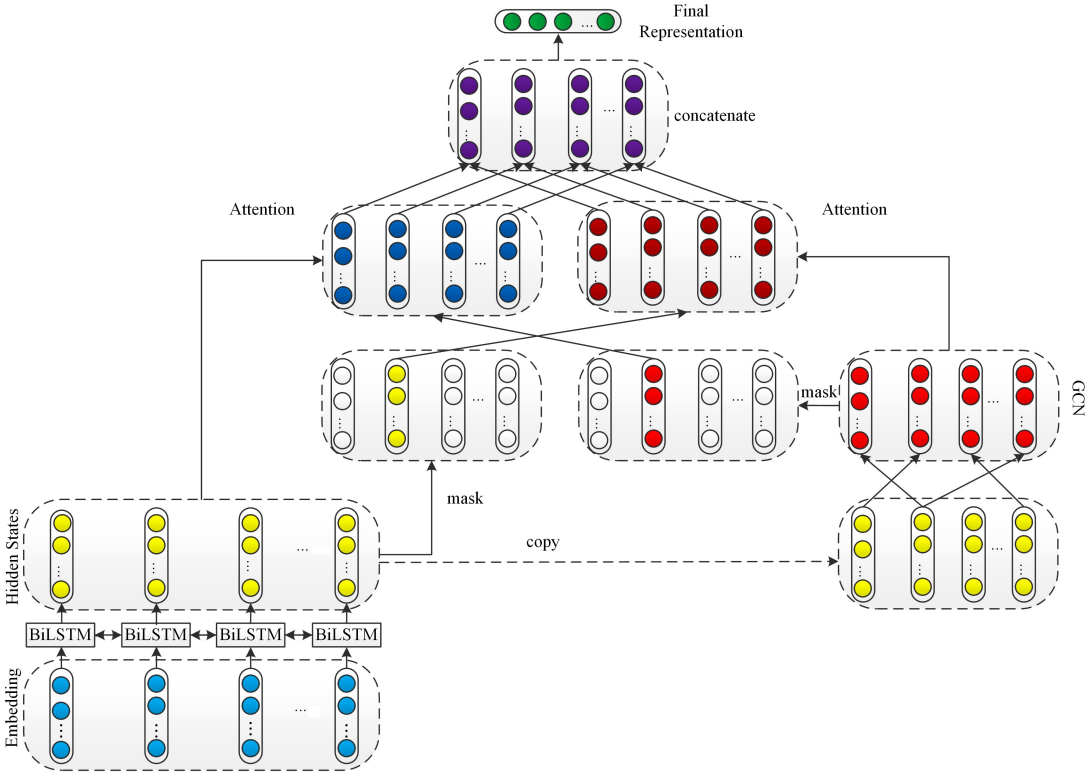


图1 IAGCN模型架构图
Fig. 1 Architecture of IAGCN model

3.2 位置编码

之前多项研究^[21-22]认为,句中每个单词对方面词的重要性不同,越靠近方面词的单词,对方面词的影响就越大。根据这个原理,本文采用以往研究中使用的方法^[4,15],为隐藏层向量引入一个位置嵌入函数,用于增强方面词附近的上下文单词对方面词的重要性。因为BiLSTM是一种时序模型,计算时已经隐含了单词序列的位置信息,所以只需额外为GCN

添加位置信息即可。具体来说,就是在进行图卷积操作之前,为GCN的输入乘以一个位置嵌入函数。位置嵌入函数如下:

$$p_i = \begin{cases} 1 - \frac{a+1-i}{n} & , 1 \leq i < a+1 \\ 1 & , a+1 \leq i \leq a+m \\ 1 - \frac{i-a-m}{n} & , a+m < i \leq n \end{cases} \tag{2}$$

$$g_i = p_i \cdot h_i \tag{3}$$

其中, $p_i \in \mathbb{R}$ 为第 i 个单词的位置权重, $\mathbf{g}_i \in \mathbb{R}^{2 \times d_h}$ 为带有位置信息的隐藏层输出。

3.3 图卷积网络

GCN 可以看作是传统 CNN 的一种改进,可以对非结构化数据的局部信息进行编码,在处理包含丰富关系的图数据时非常有效,可以有效地捕获结点之间的关系。同时 GCN 也可以理解为沿边传递信息的信息传递网络,它可以沿着依赖关系树的边直接在上下文单词和方面词之间传递信息,显著缩短了按单词顺序传递消息的距离,同时减少了消息传递过程中的噪声。为了充分利用句子的语法信息,本文通过在句子的语法依赖树上使用 GCN,来获取句子中单词的语法信息,有效利用依赖关系路径来转换、传播和聚合句中的单词信息。

具体来说,本文通过使用 spaCy 工具¹⁾来解析句子,构造句子的语法依赖树,从而获得代表语法关系的邻接矩阵 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 。若句子中的第 i 个单词和第 j 个单词在语法结构上有联系,则 $\mathbf{A}_{ij} = 1$, 否则 $\mathbf{A}_{ij} = 0$ 。同时,为每个结点添加自连接,即矩阵 $\mathbf{A}$ 的对角线 $\mathbf{A}_{ii} = 1$ 。本文使用文献[9]中改进过的归一化的图卷积公式来更新结点,结点更新操作如下:

$$\tilde{\mathbf{g}}_i = \sum_{j=1}^n \mathbf{A}_{ij} \mathbf{W} \mathbf{g}_j \quad (4)$$

$$\mathbf{g}_i = \text{ReLU} \left(\frac{\tilde{\mathbf{g}}_i}{d_i + 1} + \mathbf{b} \right) \quad (5)$$

其中, $d_i = \sum_{j=1}^n \mathbf{A}_{ij}$ 为第 i 个单词在依赖关系树上的度数; $\mathbf{W}$ 为权重矩阵, $\mathbf{b}$ 为偏置,它们均为可训练的参数。

3.4 方面词的表示

本节将通过掩码机制从 BiLSTM 层和 GCN 层的输出中获取蕴含了上下文语义信息和上下文语法信息的方面词表示。具体而言,即把方面词以外的隐藏层向量设为 0,仅保留方面词的表示:

$$\mathbf{h}_i = \begin{cases} 0, & 1 \leq i < a + 1 \\ \mathbf{h}_i, & a + 1 \leq i \leq a + m \\ 0, & a + m < i \leq n \end{cases} \quad (6)$$

$$\mathbf{g}_i = \begin{cases} 0, & 1 \leq i < a + 1 \\ \mathbf{g}_i, & a + 1 \leq i \leq a + m \\ 0, & a + m < i \leq n \end{cases} \quad (7)$$

其中, $\mathbf{h}_i$ 和 $\mathbf{g}_i$ 分别为 BiLSTM 和 GCN 层的方面词表示。此时的方面词表示由于在训练过程中都是随整个句子一起参与训练的,因此比单纯地仅输入方面词的网络输出具备更多的信息,不仅具有 BiLSTM 的词序信息和语义信息,还具有 GCN 的位置信息和语法信息。

3.5 交互注意力机制

本节将通过一种基于检索的交互注意力机制,来生成 BiLSTM 层和 GCN 层特定方面的上下文表示,具体方法是:从 BiLSTM 层上下文的隐状态向量中检索与 GCN 层方面词语义相关的重要特征,从 GCN 层上下文的隐状态向量中检索与 BiLSTM 层方面词语法相关的重要特征,并为每个上下文词分配不同的注意力权重,由此生成了从不同角度考虑的

与方面词相关的上下文信息。通过这种交互注意力的方法,可以在引入句子的语法信息的同时,减小语法解析器解析错误带来的影响,充分利用句子的语义和语法信息。具体计算如下:

$$\tilde{\alpha}_i = \sum_{i=1}^n \mathbf{g}_i \mathbf{h}_i \quad (8)$$

$$\alpha_i = \frac{\exp(\tilde{\alpha}_i)}{\sum_{i=1}^n \exp(\tilde{\alpha}_i)} \quad (9)$$

$$\tilde{\beta}_i = \sum_{i=1}^n \mathbf{h}_i \mathbf{g}_i \quad (10)$$

$$\beta_i = \frac{\exp(\tilde{\beta}_i)}{\sum_{i=1}^n \exp(\tilde{\beta}_i)} \quad (11)$$

本文采用点积的方法计算方面词和上下文词之间的相关性。式(8)~式(11)中, α_i 为 GCN 层方面词对 BiLSTM 上下文的隐状态向量的注意力权重, β_i 为 BiLSTM 层方面词对 GCN 层上下文的隐状态向量的注意力权重。然后计算特定方面的 BiLSTM 层的上下文表示 $\mathbf{r}_i^{\text{BiLSTM}}$ 和特定方面的 GCN 层的上下文表示 $\mathbf{r}_i^{\text{GCN}}$ 。具体计算过程如下:

$$\mathbf{r}_i^{\text{BiLSTM}} = \sum_{i=1}^n \alpha_i \mathbf{h}_i \quad (12)$$

$$\mathbf{r}_i^{\text{GCN}} = \sum_{i=1}^n \beta_i \mathbf{g}_i \quad (13)$$

经过尝试多种不同的特征融合方法,本文发现简单地进行拼接就可以获得最好的效果。因此,将 $\mathbf{r}_i^{\text{BiLSTM}}$ 和 $\mathbf{r}_i^{\text{GCN}}$ 进行拼接,得到最终的特定方面的上下文表示。最后对整个句子特定方面的上下文表示进行求和,得到蕴含了语义信息和语法信息的特定方面的句子的整体表示 $\mathbf{r}$ 。

$$\mathbf{r}_i = [\mathbf{r}_i^{\text{BiLSTM}}; \mathbf{r}_i^{\text{GCN}}] \quad (14)$$

$$\mathbf{r} = \sum_{i=1}^n \mathbf{r}_i \quad (15)$$

3.6 输出层

得到最终的句子表示后,将其输入到一个线性层,经 softmax 归一化,获得情感倾向性的概率分布:

$$\mathbf{p} = \text{softmax}(\mathbf{W}_f \mathbf{r} + \mathbf{b}_f) \quad (16)$$

其中,权重矩阵 $\mathbf{W}_f \in \mathbb{R}^{4d_h \times d_p}$ 和偏置 $\mathbf{b}_f \in \mathbb{R}^{d_p}$ 均为可训练的参数, d_p 为情感倾向性的类别数。

3.7 训练

模型使用标准的梯度下降算法,通过最小化交叉熵损失和 L2 正则项进行训练:

$$\text{Loss} = - \sum_i^C p_i \log \hat{p}_i + \lambda \Theta_2 \quad (17)$$

其中, C 表示数据集的集合, p 和 $\hat{p}$ 分别表示真实标签和预测标签, Θ 表示所有可训练的参数, λ 表示 L2 正则化的系数。

4 实验

4.1 实验设置

本文代码使用 Python 语言编程实现,版本为 3.8,采用了 Facebook 公司的 Pytorch²⁾深度学习框架,版本为 1.8.1。实验环境配置如下:操作系统为 Ubuntu18.04;CPU 为 Inter 的 i9-9900X@3.5 GHz;显卡为 GTX2080Ti;内存为 16 GB。

¹⁾ <https://spacy.io/>

²⁾ <https://pytorch.org>

本文选用 5 个数据集作为实验语料:由推特帖子组成的 TWITTER^[23]、SemEval 2014 task4 的 REST14 和 LAP14^[24]、SemEval 2015 task 14 的 REST15^[25]、SemEval 2016 task 5 的 REST16^[26]。其中后 4 个数据集包含了餐厅领域和笔记本电脑领域。所有数据集都分为训练集和测试集,与以往的研究一样^[3-5,12,15],本文移除了情感倾向性冲突的样本和没有明显方面词的样本。表 1 列出了各个数据集的统计信息。

表 1 数据集的统计信息
Table 1 Dataset statistics

Dataset		# Pos	# Neu	# Neg
TWITTER	Train	1 561	3 127	1 560
	Test	173	346	173
LAP14	Train	994	464	870
	Test	341	169	128
REST14	Train	2 164	637	807
	Test	728	196	196
REST15	Train	912	36	256
	Test	326	34	182
REST16	Train	1 240	69	439
	Test	469	30	117

本文实验选用 300 维度的 GloVe^[20] 词向量初始化词嵌入,所有模型的权重参数均采用均匀分布进行初始化。隐藏层的维度设为 300,选用 Adam 作为优化器,学习率为 0.001, L2 正则化系数为 10^5 ,批处理大小为 32。

实验随机初始化运行 3 次,并取 3 次运行结果的均值作为最终的实验结果。实验选择准确率和宏观 F1 值作为评估指标,因为当数据集中的样本不平衡时,宏观的 F1 值更适合作为评估指标。

表 2 模型在数据集上的结果

Table 2 Experimental results on datasets

Model	TWITTER		LAP14		REST14		REST15		REST16	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
SVM	63.40	63.30	70.49	N/A	80.16	N/A	N/A	N/A	N/A	N/A
ATAE-LSTM	69.65	67.40	69.14	63.18	77.32	66.57	75.43	56.34	83.25	63.85
MemNet	71.48	69.90	70.64	65.17	79.61	69.64	77.31	58.28	85.11	65.99
IAN	72.50	70.81	72.05	67.38	79.26	70.09	78.54	52.65	84.74	55.21
TD-GAT	72.20	70.45	75.63	70.74	81.32	71.72	80.38	60.50	87.71	67.87
ASGCN	72.15	70.40	75.55	71.05	80.77	72.02	79.89	61.89	88.99	67.48
Bi-IAN	72.01	69.97	73.67	68.68	79.40	69.90	76.94	55.57	86.80	61.72
IAGCN	72.69	70.95	74.40	69.71	82.14	74.07	79.27	63.55	88.96	69.75

(1)本文提出的 IAGCN 模型在 4 个数据集上的 F1 值都达到了最优的效果,且准确率也是最优或是接近最优的。尤其是在 REST14,REST15,REST16 数据集上,比第二名的 F1 值高 2.05,1.66,1.88。值得一提的是,尽管 IAGCN 模型在 TWITTER 数据集上达到了最优,但与其他模型的差距并不明显,这可能是由于 TWITTER 数据集上每个样本句子上仅包含一个方面,整个数据集的复杂度并不大,因此 IAGCN 模型的结果无法与其他模型的结果拉开一个较大的距离。

(2)基于图和语法信息的方法 TD-GAT、ASGCN 和本文的 IAGCN,相比其他完全没有考虑语法信息的方法,其 F1 值至少高了 1.03,说明语法信息确实有利于基于方面的情感分类任务。

(3)本文的 IAGCN 模型在 LAP14 数据集上的表现

4.2 对比实验

为了全面评估 IAGCN 模型,本文选择了以下几个具有代表性的基线与本文提出的 IAGCN 模型进行比较。

(1)SVM^[1]:针对 SemEval 2014 task 4 的基于特征提取的机器学习方法。

(2)ATAE-LSTM^[5]:基于 LSTM 和注意力机制的经典模型,且通过将方面词和上下文进行拼接来进行信息交互,探索方面和上下文之间的联系。

(3)MemNet^[4]:基于记忆网络的方法,将注意力模型与外部记忆相结合,以计算每个上下文词对方面词的重要性。

(4)IAN^[12]:使用两个 LSTM 分别对上下文和方面词进行建模,并引入交互注意力机制网络分别生成特定上下文的方面词表示和特定方面的上下文表示,用于分类。

(5)TD-GAT^[10]:基于图注意力网络来显式地利用单词之间的依赖关系。

(6)ASGCN^[11]:本模型的直接竞争对手之一,在依赖树上使用 GCN 来利用语法信息和单词依赖。

(7)Bi-IAN^[27]:本文模型的直接竞争对手之一,基于 Bi-LSTM 提取上下文和方面词的表示,并使用交互注意力机制来建模它们间的关系,是 IAN 模型的改进版。

上述方法中,前 4 种方法分别是基于机器学习的 SVM 模型、深度学习的 LSTM、记忆网络和注意力机制的经典模型,后两种是最新的基于图的方法和语法集成的方法。为了实验的客观性,所有模型都运行 3 次后取平均值,结果如表 2 所列。由表 2 中的结果可知。

不佳,虽然高于其他没有考虑语法信息的方法,但逊于 TD-GAT 模型。这可能是因为 LAP14 中的方面含有笔记本电脑领域中大量的专有名词,这无疑加大了对识别方面情感任务的难度。

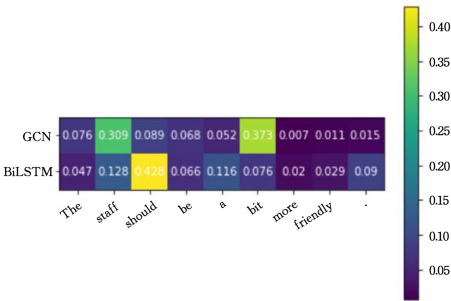
(4)从 IAN 与 Bi-IAN 的结果可看出,Bi-IAN 在各个数据集上的结果大多优于 IAN,说明相比 LSTM,BiLSTM 由于双向建模上下文的原因,可以捕捉到更丰富和完整的语义信息。

(5)与 Bi-IAN 相比,本文的 IAGCN 在所有数据集上的准确率和 F1 值都优于 Bi-IAN,说明 IAGCN 挖掘并利用了 Bi-IAN 所忽略的语法信息,这也验证了 GCN 可以有效地捕获单词间的语法信息,从而在基于方面的情感分类任务中取得了更好的结果。

(6)与 ASGCN 相比,本文的 IAGCN 在 5 个不同的数据集中 $F1$ 值都有不同程度的提升。这说明交互注意力网络可以充分利用语义信息和语法信息,并在一定程度上减轻语法解析错误带来的后果。

4.3 可视化分析

为了更好地了解 IAGCN 的工作原理,本文提供了对两个测试集的例子注意力得分,以进行可视化,如图 2 所示。第一个例子为预测出句子“The staff should be a bit friendly.”中方面词“staff”的情感倾向性,如图 2(a)所示。这是一句蕴含虚拟语气的句子,表达了与陈述语气相反的情感倾向性,需要深度结合上下文的内容才能作出正确的判断。图 2 中第一行是 GCN 的方面词表示对 BiLSTM 的上下文表示的注意力权重,第二行是 BiLSTM 的方面词对 GCN 的上下文表示的注意力权重。可以看出,GCN 通过语法依赖的信息重点关注到了“bit”这个词,而 BiLSTM 则重点关注了“should”这个虚拟语气词,通过交互注意力机制把两种特征融合后,可进一步加强两者共同关注的重点,从而预测出句中方面词的真正情感。



(a)Staff should be a bit friendly



(b)Great food but the service is dreadful

图 2 样例分析
Fig. 2 Case study

第二个例子为预测出句子“Great food but the service is dreadful.”中方面词“food”的情感倾向性,如图 2(b)所示。这是一个基于方面的情感分类任务中经典的例子。该句子同时拥有两个方面词“food”和“service”,这两个方面词表达的情感倾向性截然相反。由图可以得知,GCN 虽然同时关注了“Great”和“dreadful”,但注意力权重最深的是“dreadful”;而 BiLSTM 的注意力重点除了“Great”外,还关注了句中情感转变的转折词“but”,且对“dreadful”的注意力权重极低。通过特征融合后,两者不仅进一步加强了对观点词“Great”和转折词“but”的注意力权重,还削减了对“dreadful”的注意力权重,

从而通过不同网络间的优势互补,进一步减小了作出错误判断的可能性。

结束语 本文针对基于方面的情感分类任务,提出了一种交互注意力的图卷积网络模型。该模型充分利用了句中的语义信息和语法信息,并有效地减轻了由语法解析器解析错误带来的噪声。

在未来的工作中,我们将继续探究语义信息和语法信息更深层次的特征融合方法,学习如何获得更好的特定方面的上下文表示,从而有效地利用方面信息得到更好的分类结果。

参 考 文 献

[1] KIRITCHENKO S,ZHU X,CHERRY C,et al. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews [C]// Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014:437-442.

[2] HUANG B,OU Y,CARLEY K M. Aspect level sentiment classification with attention-over-attention neural networks [C]// International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. 2018:197-206.

[3] TANG D,QIN B,FENG X,et al. Effective LSTMs for target-dependent sentiment classification[J]. arXiv:1512. 01100,2015.

[4] TANG D,QIN B,LIU T. Aspect level sentiment classification with deep memory network[J]. arXiv:1605. 08900,2016.

[5] WANG Y,HUANG M,ZHU X,et al. Attention-based LSTM for aspect-level sentiment classification[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016:606-615.

[6] LECUN Y,BOTTOU L,BENGIO Y,et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE,1998,86(11):2278-2324.

[7] HOCHREITER S,SCHMIDHUBER J. Long short-term memory[J]. Neural Computation,1997,9(8):1735-1780.

[8] MNH V,HEESS N,GRAVES A. Recurrent models of visual attention[J]. Advances in Neural Information Processing Systems,2014,27:2204-2212.

[9] KIPF T N,WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609. 02907,2016.

[10] HUANG B,CARLEY K M. Syntax-aware aspect level sentiment classification with graph attention networks[J]. arXiv: 1909. 02606,2019.

[11] ZHANG C,LI Q,SONG D. Aspect-based sentiment classification with aspect-specific graph convolutional networks[J]. arXiv:1909. 03477,2019.

[12] MA D,LI S,ZHANG X,et al. Interactive attention networks for aspect-level sentiment classification [J]. arXiv: 1709. 00893, 2017.

[13] JIANG L,YU M,ZHOU M,et al. Target-dependent twitter sentiment classification[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics; Human Language Technologies. 2011:151-160.

[14] LI C B, DUAN Q J, JI C H, et al. Method of Short Text Classification Based on CHI and TF-IWF Feature Selection[J]. Journal of Chongqing University of Technology (Natural Science), 2021, 35(5): 135-140, 222.

[15] LI X, BING L, LAM W, et al. Transformation networks for target-oriented sentiment classification[J]. arXiv: 1805. 01086, 2018.

[16] HUANG B, CARLEY K M. Parameterized convolutional neural networks for aspect level sentiment classification[J]. arXiv: 1909. 06276, 2019.

[17] XU Z D, CHEN B Y, WANG X, et al. Research on capsule network-based for aspect-level sentiment classification[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(3): 284-292.

[18] SUN Z F, WANG J. RCNN-BGRU-HN Network Model for Aspect-based Sentiment Analysis[J]. Journal of Computer Science, 2019, 46(9): 223-228.

[19] YIN J, CHI K K, HUAN R H. Aspect-level Sentiment Analysis of Text Based on ATT-DGRU[J]. Computer Science, 2021, 48(5): 217-224.

[20] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1532-1543.

[21] ZHANG C, LI Q, SONG D. Syntax-aware aspect-level sentiment classification with proximity-weighted convolution network [C]// Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2019: 1145-1148.

[22] ZHAO P, HOU L, WU O. Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification[J]. Knowledge-Based Systems, 2020, 193: 105443.

[23] DONG L, WEI F, TAN C, et al. Adaptive recursive neural net-

work for target-dependent twitter sentiment classification[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (volume 2: Short Papers). 2014: 49-54.

[24] 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 (SemEval 2014). 2014: 27-35.

[25] 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 (SemEval 2015). 2015: 486-495.

[26] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2016 task 5: Aspect based sentiment analysis[C]// International Workshop on Semantic Evaluation. 2016: 19-30.

[27] CHEN H, YANG Y, DU S D. Research on Aspect-Level Sentiment Analysis of User Reviews[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(3): 478-485.



PAN Zhi-hao, born in 1997, postgraduate. His main research interests include natural language processing and emotion analysis.



ZENG Bi, born in 1963, professor, is a senior member of China Computer Federation. Her main research interests include machine learning and big data applications.

(责任编辑:柯颖)