

许敏聪, 宛艳萍. 基于语法类型依赖的图注意力网络细粒度情感分析方法[J]. 智能计算机与应用, 2024, 14(5): 27-35. DOI: 10.20169/j.issn.2095-2163.240504

基于语法类型依赖的图注意力网络细粒度情感分析方法

许敏聪, 宛艳萍

(河北工业大学 人工智能与数据科学学院, 天津 300401)

摘要: 现实生活中人们在网络上会产生大量的评价性文本, 对这些文本进行细粒度情感分析, 可以帮助决策者快速了解人们对某类产品或事件的看法和意见, 提供决策支持。细粒度情感分析旨在利用神经网络来对输入句子进行编码, 提取目标与其上下文之间的关系, 从而进行情感倾向的判断。最近, 人们把图神经网络应用于该任务, 取得了很好的结果。然而这些方法没有利用依赖类型标签信息, 也没有考虑到噪声的影响, 为解决这两个问题, 本文研究了一种新型图神经网络拓展来利用依赖类型信息, 并应用本文提出的语法语义交互融合的方法来降低噪声影响, 在4个公开数据集上取得了最优的效果。

关键词: 自然语言处理; 深度学习法; 情感分析; 图注意力网络

中图分类号: TP391.1

文献标志码: A

文章编号: 2095-2163(2024)05-0027-09

Aspect-based sentiment analysis based on dependency syntax type graph attention network

XU Mincong, WAN Yanping

(School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China)

Abstract: In current life, a large number of evaluative texts could be generated on the Internet. Sentiment analysis of these texts can help decision makers quickly understand people's views and opinions on a certain type of product or event, and provide decision support. Aspect-level sentiment analysis aims to use neural network to encode the input sentences, extract the relationship between the aspect and context, and judge the emotional tendency. Recently, graph neural network has been applied to this task and achieved fine results. However, these methods do not use the dependency type label information, and do not take into account the impact of noise. To solve these two problems, the paper studies a new graph neural network extension to utilize the dependency type information, and applies the syntax-semantic interaction proposed in this paper to reduce the influence of noise. The fusion method has achieved the best results on 4 public datasets.

Key words: natural language processing; approach of deep learning; sentiment analysis; graph attention network

0 引言

在互联网上会存在大量的用户评价文本, 对用户评论情感极性进行细粒度分析, 就能为决策者提供决策支持。在现实应用方面, 以华为云提出的模型为例, 针对网上的汽车领域的评论, 该模型可以分析汽车在多个属性、诸如“外观”、“动力”、“消耗”等的正负面评价。如此一来, 就有助于用户在产品选型的时候做一个对比, 也有利于厂家针对评论对产品做相应改进。

细粒度情感分析 (Aspect-based Sentiment

Analysis, ABSA) 的任务是预测文本中特定方面的情感极性^[1], 与文本级和句子级情感分析任务不同, 细粒度情感分析是以实体为中心的, 在许多情况下, 需要识别同一上下文中多个方面的不同情感极性。例如, 在“价格是很不错的, 尽管服务不是很好 (The price is reasonable, although the service is poor.)”这句话中, “服务 (service)” 的情绪极性为负, 而“价格 (price)” 的情绪极性为正; 这种情况下, 模型如果不能捕捉到每个方面的重要上下文信息, 就可能会预测错误。

目前, 细粒度情感分类工作可根据结构依赖语

基金项目: 河北省高等学校科学技术研究重点项目 (ZD2014051)。

作者简介: 许敏聪 (1999-), 男, 硕士研究生, 主要研究方向: 自然语言处理。

通讯作者: 宛艳萍 (1968-), 女, 副教授, 主要研究方向: 机器学习与智能计算, 自然语言处理。Email: wanyp_ok@126.com

收稿日期: 2023-05-08

哈尔滨工业大学主办 ◆ 学术研究与应用

法信息与否可分为两大类,即传统的不依赖语法信息的方法和依赖于图神经网络的使用语法结构的方法。

大多数传统方法的第一步是将句子分为2部分,即给定的方面和上下文,两者结成句子-方面对,然后从每个部分提取特征,结合特征进行预测。传统模型采用 CNN (Convolutional Neural Network) 建模^[2]、RNN (Recurrent Neural Network) 建模^[3]、递归卷积神经网络建模^[4]、胶囊网络建模^[5]、基于注意力机制的 RNN 方法^[6]、基于记忆网络 (MemNet)^[7] 以及在 NLP 领域产生广泛影响的预训练语言模型 BERT^[8] 都用于处理该项分类任务。研究认为 BERT 模型包含隐含的语法语义信息,这对

下游任务非常有益。

近年来,随着图神经网络的兴起,研究者们发现把图神经网络应用于细粒度情感分析任务中可以缓解上下文长距离依赖的问题,促进模型对句子的解析以及特征表示,从而提升模型效果。具体来讲,常规模型对文本的解析是线性解析,图神经网络使用语法依赖树解析句子,语法树的图结构可以指导网络重要节点的信息交换,缩短重点网络节点之间的距离。以图 1 给出的句子为例,例句中方面词“price”和情感词“reasonable”的距离间隔在常规模型解析中为 1,在图神经网络的距离间隔为 0,可看出句子在图神经网络中的解析距离比常规模型更短。

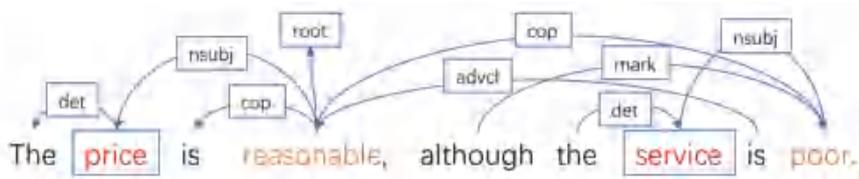


图 1 句子及其语法依赖树的示例

Fig. 1 An example of a sentence and its grammar dependency tree

使用语法结构的方法的主要步骤是将语法依赖树转换为邻接矩阵,输入到图卷积网络^[9] (Graph Convolutional Networks, GCN) 或图注意力网络^[10] (Graph Attention Networks, GAT) 等图神经网络中编码,进行提取特征和处理后得到输出。文献[11]为让模型更好地关注更重要的上下文单词,利用语法依赖树上的距离关系来引导注意力机制。文献[12]发现用 GCN 编码句法特征更有利于实现目标单词和上下文之间的信息交换。Huang 等学者^[13]应用 GAT 对依赖树进行建模,将 GAT 层集成到 LSTM (Long Short-Term Memory) 中,实现了良好的效果。文献[14]对句子中方面词和上下文应用位置信息与语法信息交互建模,在使用图注意力网络交互注意力学习语法信息的同时还学习到位置信息。

研究发现,这些实验中使用的模型忽略了语法树中携带的依赖类型标签信息,从而平等地对待语法树中的所有单词关系,这可能导致模型预测失败。另一方面,由于这些模型对语法依赖树是直接使用,这可能会导致噪声问题,究其原因在于模型对句法分析效果的不完善和输入句子的随机性,所以在引用依赖树时会产生噪声。

针对这 2 个问题,本文提出了一种语法类型依赖图注意力网络 (Dependency Syntax Type Graph Attention Network, DST_GAT),通过结合单词关系

和彼此间的依赖类型来解析文本,并应用新的语法语义融合方法,从而增强模型效果。本文的贡献在于:

(1) 该模型在解析语法依赖树基础上将依赖类型特征融入注意力机制,使用一种新的扩展注意力机制来学习文本信息,由此模型可以更好地捕捉单词之间的关系,从而解决了未使用语法树中依赖类型信息的问题。

(2) 本文提出一种新的仿射融合机制 (Biaffine Fusion Function, BFF),通过在利用语法和语义互相交换信息的同时,采用函数融合的方法降低了引用语法依赖树时产生的噪声对系统的影响,并通过实验验证了其有效性。

(3) 本文的模型在 Laptop、Restaurant、Twitter 和 MAMS 四个数据集上达到了最优的效果。实验结果表明,使用 DST_GAT 模型可以获得更好的结果,验证了本文提出的句法依赖拓展机制及语法语义融合机制的重要性。

1 方法

本次实验的目的是预测出给定目标词序列的情感极性。给定一个长度为 n 的句子 $S = \{w_1, w_2, \dots, w_i, \dots, w_{i+m}, \dots, w_n\}$, 长度为 m 的目标词序列 $T = \{w_i, w_{i+1}, \dots, w_{i+m-1}\}$, 根据语法关系生成语法树 $G =$

(V, A, R) 来进行建模, 其中 V 包含所有节点(单词) $\{w_1, w_2, \dots, w_n\}$, A 是邻接矩阵($A \in \mathbb{R}^{n \times n}$), 如果单词 w_i 和 w_j 之间存在依赖关系, $A_{ij} = 1$, 否则 $A_{ij} = 0$, R 为关系标签矩阵, 其中, 如果 $A_{ij} = 1$, 则 R_{ij} 记录 A_{ij} 的对应关系标签, 否则 $R_{ij} = \text{None}$ 。然后预测句子 S 中给定目标词 T 的情绪的极性 $y \in \{1, -1, 0\}$, 其中 1, -1 和 0 分别表示正、负和中性。

实验模型的总体架构如图 2 所示。由图 2 可知, 架构主要由 3 部分组成: 上下文编码器、依赖关系类型编码器和分类器。上下文编码器可以被视为

一种传统模型, 可将上下文编码器层(例如, BiLSTM、CNN、BERT)应用于表征学习语义信息。DST_GAT 编码器是一种基于语法的编码器, 是将语法信息整合到句子建模过程中, 从而生成包含语法的单词嵌入。实验使用 DST_GAT 编码器和常规上下文编码器, 对 2 个编码器的输出使用池化函数, 得到 2 个编码器语法和上下文的特征表示, 分别表示为 h_{syn} 和 h_{con} 。最后采用仿射融合机制 BFF 来融合 h_{syn} 和 h_{con} , 通过该机制降低语法树中噪声的影响, 得到最后的表示 h_f , 将 h_f 输入分类器进行分类。

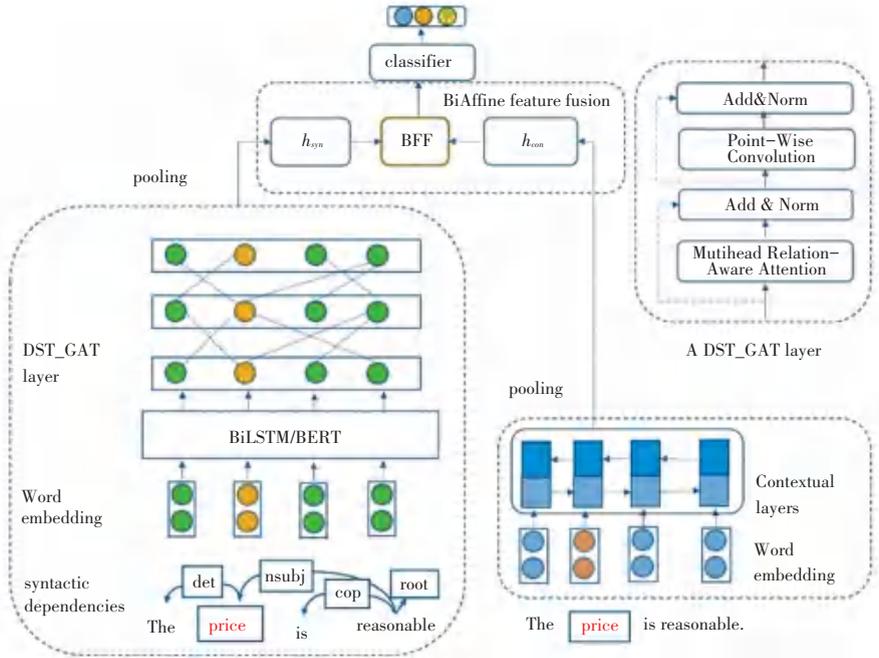


图 2 DST_GAT 模型
Fig. 2 DST_GAT model

1.1 上下文编码器

本文使用 2 种类型的上下文建模模型。一种是 BiLSTM 模型, 被广泛用于分类等许多任务。另一种是 BERT 模型, 已在大规模原始文本上进行了预训练, 并在许多任务上取得了良好的效果。

(1) BiLSTM。首先使用 BiLSTM 来对上下文信息进行建模。在以往工作^[12]的基础之上, 实验使用 GloVe 词嵌入 $v_i \in \mathbb{R}^{d_e}$, 标签嵌入 $t_i \in \mathbb{R}^{d_l}$ 以及位置信息嵌入 $p_i \in \mathbb{R}^{d_p}$ 作为输入特征, 单词 w_i 的表示为三者的并列 $x_i = [v_i; t_i; p_i]$, 其中 d_e, d_l 和 d_p 分别表示单词、标签和位置信息嵌入的维度。

给定一个文本的词嵌入序列 $x = \{x_1, x_2, \dots, x_n\}$ 输入至 BiLSTM 编码器中, 然后经过前向和后向传播生成相应的隐藏状态 $\vec{h} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n\}$ 和 $\overleftarrow{h} = \{\overleftarrow{h}_1, \overleftarrow{h}_2, \dots, \overleftarrow{h}_n\}$, 最后, 再将二者拼接并输出最后的

隐藏状态 $h = \{\vec{h}; \overleftarrow{h}\}$ 。

(2) BERT。BERT 是 Google 于 2018 年提出的一种基于 Transformer^[15] 的预训练语言模型, 由于其包含丰富的语法语义信息, 使用 BERT 模型可以优化系统, 显著提高分类精度。

本文将序列重构为“[CLS]”+句子+“[SEP]”+给定目标词+“[SEP]”作为 BERT 的输入。随后 BERT 模型生成一个与重构长度相同的新序列, 记为:

$$h = \{h_0, h_1, \dots, h_n, h_{n+1}, h_{n+2}, \dots, h_{n+1+m}, h_{n+2+m}\} \quad (1)$$

其中, h_1, h_2, \dots, h_n 是输入单词序列的输出上下文表示, $h_{n+2}, \dots, h_{n+1+m}$ 对应给定目标词的输出嵌入。在这里, 使用了 h_1, h_2, \dots, h_n 进行池化和特征融合。与 BiLSTM 不同, 基于 BERT 的模型不需要

GloVe、标签信息 POS-tag 或位置信息嵌入作为额外的输入特征,因为这些特征已经包含在模型的预训练过程中。

1.2 语法类型依赖图注意力网络

语法类型依赖图注意力网络(DST_GAT)的目标是使用句法依赖路径,让单词之间的信息进行更好的交换。与标准 GAT 网络相比,DST_GAT 方法可以额外利用语法依赖关系,从而生成更多信息表示。本节中,首先提出了在未使用依赖关系的 $G=(V, A)$ 上运行的基线 GAT 模型,然后介绍 DST_GAT,对应的依赖关系图为 $G=(V, A, R)$ 。

1.2.1 vanilla GAT

图注意力网络(GAT)是图神经网络的一种变体,可以将目标节点的语法上下文信息传播到目标节点。其传播方式如下:给定一个具有 N 个节点的依赖图,其中每个节点与一个单词嵌入向量相关联,一个 GAT 层通过聚集邻域的隐藏状态来计算节点的表征,对应的输入为一组节点特征序列 $\vec{h} = \{\vec{h}_1, \vec{h}_2, \dots, \vec{h}_i, \dots, \vec{h}_n\}$, $\vec{h}_i \in \mathbb{R}^d$,其中 N 为节点数, d 为每个节点的特征数,输出为由该层生成的一组新的节点特征 $\vec{h}' = \{\vec{h}'_1, \vec{h}'_2, \dots, \vec{h}'_i, \dots, \vec{h}'_n\}$, $\vec{h}'_i \in \mathbb{R}^d$ 。如果使用 L 层 GAT 网络,可以将 L 跳的特征传播到目标节点。在这个过程中,扩展并使用多头注意力有助于学习深层信息,稳定这个过程中图注意力的学习过程。基于多头注意力的 GAT 的聚合过程可以描述为:

$$\mathbf{h}_i^l = \parallel_{z=1}^z \sigma \left(\sum_{j \in N(i)} a_{ij}^z \mathbf{w}_i^z \mathbf{h}_i^{l-1} \right) \quad (2)$$

其中, $\mathbf{h}_i^l \in \mathbb{R}^d$ 为 l 层第 i 个节点的隐藏状态,其大小是词特征向量的维度,即大小为 d 的向量,这里 d 表示词特征向量的维度;“ \parallel ”表示向量连接; z 表示注意力头部的数量; $j \in N(i)$ 是其邻居单词索引,如果 $j \notin N(i)$,则 $a_{ij}^z = 0$,权重 a_{ij}^z 取决于 \mathbf{h}_i^l 对 \mathbf{h}_j^{l-1} 的依赖程度; $\mathbf{w}_i^z \in \mathbb{R}^{\frac{d}{z} \times d}$ 表示第 l 层第 z 个头部的参数矩阵; $\mathbf{h}_j^{l-1} \in \mathbb{R}^d$ 为 $l-1$ 层第 j 个节点的隐藏状态; σ 表示 sigmoid 激活函数。

在这里,原 GAT 模型中使用一个单层前馈神经网络进行计算 a_{ij}^z 的权重 e^{ij} :

$$e^{ij} = \vec{a}^T (\mathbf{w}_i^z \mathbf{h}_i^{l-1}, \mathbf{w}_j^z \mathbf{h}_j^{l-1}) \quad (3)$$

由 $\mathbf{w}_i^z, \mathbf{w}_j^z \in \mathbb{R}^{\frac{d}{z} \times d}$ 和 $\mathbf{h}_i^{l-1}, \mathbf{h}_j^{l-1} \in \mathbb{R}^d$ 可以计算出 $\mathbf{w}_i^z \mathbf{h}_i^{l-1}$ 和 $\mathbf{w}_j^z \mathbf{h}_j^{l-1}$ 是长度为 $\frac{d}{z}$ 的向量,两者连接后

得到一个大小为 $\frac{2d}{z}$ 的向量, \vec{a} 为一个大小为 $\frac{2d}{z}$ 的权重向量,两者相乘得到一个实数,即 e^{ij} ,再对这个结果应用 LeakyReLU 激活函数,使用 softmax 函数归一化得到最终结果:

$$a_{ij}^z = \text{softmax}_j (\text{LeakyReLU}(e^{ij})) \quad (4)$$

式(4)完全展开后,注意力系数计算公式可以表示为:

$$a_{ij}^z = \frac{\exp(\text{LeakyReLU}(\vec{a}^T [\mathbf{w}_i^z \mathbf{h}_i^{l-1} \parallel \mathbf{w}_j^z \mathbf{h}_j^{l-1}]))}{\sum_{j \in N(i)} \exp(\text{LeakyReLU}(\vec{a}^T [\mathbf{w}_i^z \mathbf{h}_i^{l-1} \parallel \mathbf{w}_j^z \mathbf{h}_j^{l-1}]))} \quad (5)$$

与原论文不同,为提高模型中信息的融合程度,期望在语法中能够充分利用语义信息,达到降低模型噪声的效果,本文的模型中使用缩放点积注意力函数 f ,这是一个注意力函数:

$$a_{ij}^z = \frac{\exp(f(\mathbf{h}_i^{l-1}, \mathbf{h}_j^{l-1}))}{\sum_{h_j^{l-1}} \exp(f(\mathbf{h}_i^{l-1}, \mathbf{h}_j^{l-1}))} \quad (6)$$

$$f(\mathbf{h}_i^{l-1}, \mathbf{h}_j^{l-1}) = \frac{(\mathbf{W}_Q^z \mathbf{h}_i^{l-1})(\mathbf{W}_K^z \mathbf{h}_j^{l-1})}{\sqrt{d/z}} \quad (7)$$

其中, $\mathbf{W}_Q^z, \mathbf{W}_K^z \in \mathbb{R}^{\frac{d}{z} \times d}$ 表示第 l 层第 z 个头的参数矩阵。

1.2.2 语法类型依赖图注意力网络

前文提及的 vanilla GAT 使用邻接矩阵作为结构信息,但在这个过程中忽略了语法类型依赖特征。语法类型依赖图注意力网络(DST_GAT)将关系特征融入注意力计算和聚合过程,继而计算 2 种注意力分布,即:节点注意力和依赖关系注意力,再将其组合作为特征聚合的最终注意力权重。

节点注意力 e^N 的计算方式与式(7)相同:

$$e_{ij}^N = \begin{cases} f(\mathbf{h}_i^{l-1}, \mathbf{h}_j^{l-1}), & j \in N(i) \\ -\text{inf}, & \text{otherwise} \end{cases} \quad (8)$$

关于依赖关系注意力的计算,实验将节点间依赖关系作为输入,令 w_i 和 w_j 之间的关系表示为 R_{ij} ,将 R_{ij} 转换为向量 $r_{ij} \in \mathbb{R}^{d_r}$,其中 d_r 是关系嵌入的维度,然后计算依赖关系注意力。

依赖关系注意力 e^R 权重由式(9)运算求出:

$$e_{ij}^R = \begin{cases} f(\mathbf{h}_i^{l-1}, r_{ij}), & j \in N(i) \\ -\text{inf}, & \text{otherwise} \end{cases} \quad (9)$$

通过式(9),可以计算出 $l-1$ 层节点与其相邻节点关系特征 r_{ij} 之间的关系权重,也就是依赖关系注意力。在这个过程中,因为 r_{ij} 所表示的关系嵌入在多个层次上和注意力头之间是共享的,所以 r_{ij} 还

可以承载更丰富的关联信息,在优化句法依赖关系表示的同时可以优化模型,最终得到的模型训练结果不仅仅是依赖关系注意力最优解,也有益于模型优化。

最后,将上述2种类型的注意力 e_{ij}^N 和 e_{ij}^R 合并并标准化得到融合注意力 \hat{a}_{ij} :

$$\hat{a}_{ij} = \frac{\exp(e_{ij}^N + e_{ij}^R)}{\sum_{j' \in N(i)} \exp(e_{ij'}^N + e_{ij'}^R)} \quad (10)$$

其中, $j' \in N(i)$ 表示 \hat{a}_{ij} 的邻居节点。

在得到包含语法依赖关系和节点关系的融合注意力 \hat{a}_{ij} 后,进行关系特征聚合,实验使用 h_i^l 的相邻单词的隐藏特征向量 h_j^{l-1} 及其对应的语法关系向量 r_{ij} 作为输入,来通过 h_j^{l-1} 更新 h_i^l :

$$h_i^l = \sigma \left(\sum_{j \in N(i)} \hat{a}_{ij}^l (w_v^l h_j^{l-1} + w_r^l r_{ij}) \right) \quad (11)$$

其中, $w_v^l \in \mathbb{R}^{d \times d_l}$ 是一个参数矩阵。为更好地了解深层特征,本文采用多层关系图注意力网络。

1.3 池化和特征融合

研究使用上下文编码器和DST_GAT编码器后,获得上下文和语法特征向量,分别表示为 $\{h_i, h_{i+1}, \dots, h_{i+m-1}\}$ 和 $\{h'_i, h'_{i+1}, \dots, h'_{i+m-1}\}$ 。然后在这些向量上应用池化函数,获得2个全局表示 h_{con} 和 h_{syn} :

$$h_{con} = pool(h_i, h_{i+1}, \dots, h_{i+m-1}) \quad (12)$$

$$h_{syn} = pool(h'_i, h'_{i+1}, \dots, h'_{i+m-1}) \quad (13)$$

其中, $pool$ 是一个普通的池化函数,本实验采用的是平均池化。

为了学习包含上下文和语法特征的复合表示,本文借鉴BiAffine仿射方法,提出一种新的仿射约束融合机制BFF,将语法表示和上下文表示通过融合的方式有效地交换特征,减少不正确的依赖树所导致的错误,由此降低语法树中噪声的影响,融合函数如下:

$$h'_{syn} = softmax(h_{syn} W_1 (h_{con})^T) h_{syn} \quad (14)$$

$$h'_{con} = softmax(h_{con} W_2 (h'_{syn})^T) h_{con} \quad (15)$$

其中,参数矩阵 $W_1, W_2 \in \mathbb{R}^{d \times d}$ 。 h_{con} 和 h_{syn} 与一个可学习参数 W_1 和 W_2 相乘后再分别与对方相乘,然后通过对其施加归一化函数得到一系列参数,由此来指导自身学习。通过这种互相融合学习的方式,来更充分地学习上下文和语法深层信息,达到降低噪声的目的。

然后将语法表示 h'_{syn} 和上下文表示 h'_{con} 二者相

加得到最终输出 h_f :

$$h_f = h'_{syn} + h'_{con} \quad (16)$$

1.4 分类器

分类器是一个全连接网络,将融合表示 h_f 作为输入,并计算每个情绪类别 c 的概率:

$$P(y=c) = \frac{\exp(Wh_f + b)_c}{\sum_{c' \in C} \exp(Wh_f + b)_{c'}} \quad (17)$$

其中, W 和 b 是可调的模型参数, C 是情感类的集合。

给定一组训练实例 $D = \{d_1, d_2, \dots, d_n\}$,训练目标为L2正则化的交叉熵损失:

$$l = - \sum_{l=1}^N \sum_{c \in C} I(y=c) \log(p(y=c)) + \lambda \|\Phi\|^2 \quad (18)$$

其中, I 表示一个指示函数; N 表示训练示例的数量; λ 表示一个正则化超参数; Φ 表示模型中的所有参数集。

2 实验

本文在4个基准数据集上进行了实验,包括Restaurant数据集、Laptop数据集、Twitter数据集和MAMS数据集^[16]。这些数据集有3个情绪极性,即:积极、中性和消极。

在Restaurant, Laptop和Twitter这3个数据集中,大多数句子只包含一个方面词或者包含多个相同情感的方面词,这样可能会造成基于方面的情感分析任务退化成句子级别的情感分析任务。数据集MAMS的特点是一个句子中一定包含至少两个方面词,并且同一个句子中至少有2个方面词情感极性是不同的。表1统计了数据集的数据。

表1 数据集统计

Table 1 Dataset statistics

数据集	方法	积极	中性	消极
Laptop	train	976	445	851
	test	377	167	128
Restaurant	train	2 164	637	807
	test	727	196	196
Twitter	train	1 507	3 016	1 528
	test	172	336	169
MAMS	train	3 380	5 042	2 764
	test	400	607	329

2.1 实验设置

实验使用2种类型的上下文编码器:基于

BiLSTM 的编码器和基于 BERT-BASE 的编码器。对于基于 BiLSTM 的编码器,采用 300 维 GloVe 向量来表示单词,POS 标签、位置和关系嵌入的大小设置为 30。输入字嵌入的 *Dropout* 率为 0.7, *L2* 正则化项 $\lambda = 10^{-5}$ 、采用学习率为 10^{-3} 的 Adamax 优化器进行模型训练。

对于基于 BERT 的编码器,采用预训练的 BERT 进行微调。BERT 嵌入的 *dropout* 率为 0.1, 正则化项 $\lambda = 10^{-5}$ 、采用学习率为 2×10^{-5} Adam 优化器进行训练。在 DST_GAT 编码器中,每一层使用 5 个注意力头,实验使用 Deep Biaffine 解析器来获取依赖树。模型采用准确率(*Acc*)与宏平均 *F1* 值(*MF1*)作为评价指标。

2.2 基线

本实验模型与现有的一些基准模型进行比较。各个模型简介如下。

(1)SVM^[17]:支持向量机(SVM)模型应用于广泛的特征进行分类,利用 *n*-gram 特征、解析特征和词汇特征进行方面级情感分类。

(2)IAN^[18]:使用 2 个 LSTM 网络分别对句子和方面进行建模。可使用句子中的隐藏状态来生成目标的注意向量,反之亦然。基于这 2 个注意向量,输

出一个句子表示和一个目标表示进行分类。

(3)TNet^[19]:将 BiLSTM 嵌入转化为特定目标嵌入,并使用 CNN 进行编码。

(4)MGAN^[20]:利用 BiLSTM 捕捉上下文信息,利用多粒度注意机制捕捉方面和上下文之间的关系。

(5)AEN^[21]:采用注意力编码网络进行特征表示,并对目标和上下文之间的语义交互进行建模。

(6)BERT-SPC^[21]:将目标词和句子输入预训练的 BERT 模型,然后使用池化嵌入进行分类。

(7)DGEDT^[22]:融合了 GCN 和 Transformer 的细粒度情感分类模型,并使二者交互学习表征信息。

(8)CDT^[23]:将语法依赖树与图卷积神经网络集成,将二者用于方面表征学习,进行细粒度情感分类任务。

(9)ASGCN^[24]:首先使用图卷积神经网络学习基于方面的情感分类,同时在文本信息表征中应用注意力机制。

(10)R-GAT^[25]:提出了一种面向方面的依赖树结构,然后用基于关系表征的 GAT 对新的依赖树进行编码。

2.3 实验结果

不同模型的实验结果见表 2。

表 2 实验结果

Table 2 Experimental results

模型	Laptop		Restaurant		Tweets		MAMS	
	准确率 (<i>Acc</i>)	宏平均 <i>F1</i>						
SVM	70.49	-	80.16	-	63.40	-	-	-
IAN	72.05	67.38	79.26	70.09	72.50	70.81	76.60	-
TNet	76.54	71.75	80.69	71.27	74.90	73.60	-	-
MGAN	75.39	72.47	81.25	71.94	72.54	70.81	-	-
AEN	73.51	69.04	80.98	72.14	72.83	69.81	66.72	-
BERT-SPC	78.99	75.03	84.46	76.98	73.55	72.14	82.82	81.90
DGEDT	76.80	72.30	83.90	75.10	74.80	73.40	-	-
CDT	77.19	72.99	82.30	74.02	74.66	73.66	80.70	79.79
ASGCN	75.55	71.05	80.77	72.02	72.15	70.40	-	-
R-GAT	77.42	73.76	83.30	76.08	75.57	73.82	-	-
DST_GAT	78.26	74.64	82.86	77.12	75.12	75.21	82.44	81.44
DST_GAT+BRET	80.86	77.51	85.92	79.88	77.18	76.23	83.56	82.97

表 2 展示了不同模型的实验结果。与不依赖句法特征的模型(SVM、IAN、TNet、MGAN、AEN)相比,DST_GAT 模型的结果有显著提高,这证明了语法依赖树的积极作用。此外,与最新的采用依赖树但没

有关系标签作为输入的模型(CDT、ASGCN、R-GAT)相比,DST_GAT 在所有数据集上也提供了良好的结果。预训练的 BERT 模型可以显著提高模型的性能,在 BERT 的帮助下,该模型比所有基线都取

得了更好的结果,在 Laptop、Restaurant、Twitter 和 MAMS 数据集上的准确度分别为 80.86%、85.92%、77.27% 和 83.56%。实验模型在不借助使用外部资源(如情感词典等)的情况下达到最优实验结果,这证明本文提出的模型在面对复杂句子处理能力上的优异能力。

3 分析

为了进一步验证 DST_GAT 的每个组件对性能的影响,本文对结构信息和关系信息的影响进行了

消融研究。考虑 3 个消融基线进行比较,具体如下。

(1)Transformer:将关系图注意层替换为自我注意层,且未使用 BFF 模块,这个模型被视为没有使用依赖树的对比模型。

(2)GAT:用图注意层代替关系图注意层,取消语法依赖特征模块及关系融合双约束模块。

(3)DST_GAT-BFF:移除了 DST_GAT 的关系特征双约束模块,保留语法依赖特征模块。

各个模型的性能对比见表 3。

表 3 消融实验结果
Table 3 Ablation results

模型	Laptop		Restaurant		Tweets		MAMS	
	准确率 (Acc)	宏平均 F1	准确率 (Acc)	宏平均 F1	准确率 (Acc)	宏平均 F1	准确率 (Acc)	宏平均 F1
Transformer	74.31	71.79	80.78	72.10	72.78	70.23	79.65	78.87
GAT	76.24	73.03	81.31	75.02	73.25	72.38	81.21	80.32
DST_GAT-BFF	76.93	72.82	82.34	75.41	74.52	73.19	81.45	80.86
DST_GAT	78.25	74.63	82.83	77.12	75.19	75.16	82.44	81.44

Transformer 模型在所有数据集上的表现明显逊色于其他所有模型,证明了关系图注意层相对于自我注意层的优越性。GAT 模型在整体上的 Acc 值和 F1 值上基本低于除 Transformer 以外的其他模型,这证明了语法依赖树标签及特征融合在指导情感分类中的作用。

在证明语法依赖树的成功的同时,同样发现一些问题,如在 Laptop 数据集的对比中,没有语法依赖特征模块的模型的 GAT 比拥有语法依赖特征模块的 DST_GAT-BFF 在 F1 值的表现更好。除了在数据集介绍中提到的情感分析任务退化问题外,认为产生这种情况的原因和以下因素有关。

从数据集角度看,Laptop 数据集仅 2 944 条数据,在所有数据集中最少,又有更多隐式表达情感的样本,分析难度更大,可能导致欠拟合状态。在模型方面,尽管拥有语法依赖特征模块的 DST_GAT-BFF 在计算节点注意力系数时应用了点积注意力的方式降低了部分噪声,但未使用语法语义融合模块 BFF 来充分利用语义信息,从而导致对隐性情感表达不够敏感,本次研究认为语义信息的缺失是影响 DST_GAT-BFF 在具有更多隐性表达数据集上性能表现较弱的主要原因。而 GAT 并未使用语法依赖树,受语法结构影响较少,相较而言学习的隐式语义信息更多,所以部分表现优于未使用语法语义融合

模块 DST_GAT-BFF。在后续数据集上的实验中,随着数据集规模的增大,系统学习隐性语义信息越来越丰富,从 Restaurant 到 Tweets、尤其是再到拥有 12 522 条数据的 MAMS 数据集,都鲜明地显示出本文所提出实验模型的优越性。尽管 Tweets 数据集语法相较于其他数据集并不是很规则,但在 6 728 条数据的训练下,可以看到在 Tweets 数据集上 DST_GAT-BFF 模型性能已经高于 GAT 模型,证明了 DST_GAT-BFF 可以通过更大量的学习来弥补欠拟合和受噪声影响的状态。由此可以得出以下结论:

语法依赖特征模块可以帮助模型充分利用语法信息进行学习,学习更多关于语法结构的信息。关于学习语法信息过程中产生的噪声,在这个模块中采用了点积注意力计算注意力系数的方式来初步降低噪声,融合学习语义信息。随后通过关系特征融合模块做进一步优化,通过语法和语义融合过程中进行交互学习,从而达到降低语法依赖树噪声和优化模型的效果,该模块是对语法依赖标签模块的互相补充和优化。接下来又进行了模型深度的测试实验。

图 3 显示了在 Restaurant 数据集上不同层数的 DST_GAT 和 DST_GAT+BERT 模型的精度曲线。实验考虑了从 1~8 的不同层数,两者初始精度较低,在 2 层中分别达到 82.87% 和 85.92% 的最佳结果。本文认为在第二层的时候达到最佳效果的原因是随

着网络层数的增加,模型过拟合问题逐渐突出,同时模型中的噪声的强度也随之增加,因而也影响了模

型的性能,从而导致模型性能随深度增加而整体呈降低趋势。

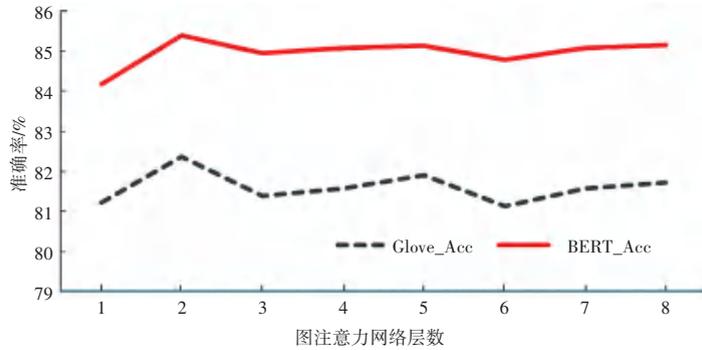


图3 图注意力网络层数对模型精度的影响

Fig. 3 Effect of layers of graph attention network on model accuracy

4 结束语

本文提出了一种基于语法图注意力网络的细粒度情感分析模型,扩展了解析关系特征的图注意力网络编码器,提出了语法语义交互融合机制来降低模型噪声和学习语义信息,研究了语法依赖结构在细粒度情感分类中的应用。后续在4个公开数据集上进行的实验表明,DST_GAT模型可以有效地对这些特征进行编码,并取得最优效果,也证明了语法标签信息对于细粒度情感分类非常有用。实验还进一步研究了不同模块和网络深度对模型性能的影响,验证模型不同模块的功能。

参考文献

[1] SHEN Weizhou, YANG Yunyi, QUAN Xiaojun, et al. Relational graph attention network for aspect-based sentiment analysis [C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020; 3229-3238.

[2] 王义, 沈洋, 戴月明. 基于细粒度多通道卷积神经网络的文本情感分析[J]. 计算机工程, 2020, 46(5): 102-108.

[3] CASTELLUCCI G, SIMONE F, CROCE D, et al. UNITOR: Aspect based sentiment analysis with structured learning [C]// International Workshop on Semantic Evaluation. Dublin, Ireland; ACL, 2014; 761-767.

[4] LAI Siwei, XU Liheng, LIU Kang, et al. Recurrent convolutional neural networks for text classification [C]// Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, USA; AAAI, 2015; 2267-2273.

[5] 滕磊, 严馨, 徐广义, 等. 使用胶囊网络的细粒度情感分析方法[J]. 小型微型计算机系统, 2020, 41(12): 2550-2556.

[6] WANG Yequan, HUANG Minlie, ZHU Xiaoyan, et al. Attention-based LSTM for aspect level sentiment classification [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, USA; ACL, 2016; 606-615.

[7] TANG Duyu, QIN Bing, LIU Ting. Aspect level sentiment classification with deep memory network [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, USA; ACL, 2016; 214-224.

[8] DEVLIN J, CHANG Mingwei, LEE K, et al. BERT: Pretraining of deep bidirectional transformers for language understanding [C]// The North American Chapter of the Association for Computational Linguistics. Minneapolis, USA; ACL, 2019; 4171-4186.

[9] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks [J]. arXiv preprint arXiv: 1609.02907, 2016.

[10] VELICKOVI P, GUCURULL G, CASANOVA A, et al. Graph attention networks [J]. arXiv preprint arXiv: 1710.10903, 2017.

[11] HE Ruidan, WEE S L, HWEE T N, et al. Effective attention modeling for aspect-level sentiment classification [C]// Proceedings of the 27th International Conference on Computational Linguistics. New Mexico, USA; ICCL, 2018; 1121-1131.

[12] SUN Kai, ZHANG Richong, MENSAH S, et al. Aspect-level sentiment analysis via convolution over dependency tree [C]// Conference on Empirical Methods in Natural Language Processing. Hong Kong; ACL, 2019; 5678-5687.

[13] HUANG Binxuan, CARLEY K. Syntax-aware aspect level sentiment classification with graph attention networks [C]// Conference on Empirical Methods in Natural Language Processing. Hong Kong; ACL, 2019; 5468-5476.

[14] 韩虎, 吴渊航, 秦晓雅. 面向方面级情感分析的交互图注意力网络模型[J]. 电子与信息学报, 2021, 43(11): 3282-3290.

[15] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS). Long Beach, USA; NIPS Foundation, 2017; 5998-6008.

[16] JIANG Qingnan, CHEN Leitai, XU Ruifeng, et al. A challenge dataset and effective models for aspect-based sentiment analysis [C]// Conference on Empirical Methods in Natural Language Processing. Hong Kong; ACL, 2019; 6279-6284.

[17] KIRITCHENKO S, ZHUXiaodan, CHERRY C, et al. Detecting aspects and sentiment in customer reviews [C]// The 8th International Workshop on Semantic Evaluation (SemEval 2014).

- Dublin, Ireland; ACL, 2014: 437-442.
- [18] MA Dehong, LI Sujian, ZHANG Xiaodong, et al. Interactive attention networks for aspect-level sentiment classification [C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia; Melbourne, 2017: 4068-4074.
- [19] LI Xin, BING Lidong, LAM W, et al. Transformation networks for target-oriented sentiment classification [C]// Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia; ACL, 2018: 946-956.
- [20] FAN Feifan, FENG Yansong, ZHAO Dongyan. Multi-grained attention network for aspect-level sentiment classification [C]// Proceedings of the 56th Conference on Empirical Methods in Natural Language Processing. Punta Cana, Dominican; dblp, 2018: 3433-3442.
- [21] SONG Youwei, WANG Jiahai, JIANG Tao, et al. Attentional encoder network for targeted sentiment classification [J]. arXiv preprint arXiv:1902.09314, 2019.
- [22] TANG Hao, JI Donghong, LI Chenliang, et al. Dependency graph enhanced dualtransformer structure for aspect-based sentiment classification [C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. ACL, 2020: 6578-6588.
- [23] SUN Kai, ZHANG Richong, SAMEL M, et al. Aspect-level sentiment analysis via convolution over dependency tree [C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong; dblp, 2019: 5679-5688.
- [24] CHEN Zhang, LI Qiuchi, SONG Dawei. Aspect-based sentiment classification with aspect-specific graph convolutional networks [C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong; dblp, 2019: 4567-4577.
- [25] HUANG Binxuan, CARLEY K. Syntax-aware aspect level sentiment classification with graph attention networks [C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong; dblp, 2019: 5469-5477.