

融合句法树多信息学习方面级情感分析

张文豪¹, 廖列法^{1,2}, 王茹霞¹

(1. 江西理工大学 信息工程学院, 江西 赣州 341000; 2. 江西理工大学 软件工程学院, 南昌 330000)

摘要:近年来的方面级情感分析基本都是单一地进行语义信息或语法信息的挖掘, 未建立语义信息和语法信息之间的关联, 且已有模型大多都是单一地将词的相对距离或语法距离嵌入模型中, 忽略了相对距离和语法距离对方面词的联合影响, 同时未充分考虑单词在依存句法树中的位置关系。建立一种融合句法树多信息学习的方面级情感分析模型 MILFST, 有针对性地利用不同神经网络的优点进行模型构建, 以获得更为丰富的信息。通过双向长短时记忆网络捕捉文本序列的信息, 根据依存句法树的树形结构更新序列信息, 将相对距离和语法距离位置信息嵌入文本序列中, 分别通过卷积神经网络和图卷积网络学习语义信息和语法信息。通过注意力机制实现语义信息和语法信息的优化融合, 并将融合后的信息输入 Softmax 分类器中进行情感极性分类。实验结果表明, 在 Twitter、Lap14、Res14、Res15、Res16 数据集上, MILFST 模型的准确率和 F1 值分别为 74.27% 和 73.14%、77.74% 和 74.27%、82.50% 和 74.54%、81.73% 和 66.15%、89.61% 和 71.57%, 模型中的树形结构有助于对信息的捕获, 同时兼顾语法信息与语义信息的学习有利于方面词情感极性判断。

关键词: 语义信息; 语法信息; 卷积神经网络; 图卷积网络; 注意力机制

开放科学(资源服务)标志码(OSID):



中文引用格式: 张文豪, 廖列法, 王茹霞. 融合句法树多信息学习方面级情感分析[J]. 计算机工程, 2023, 49(10): 72-79.

英文引用格式: ZHANG W H, LIAO L F, WANG R X. Aspect-level sentiment analysis for multi-information learning by fusing syntax trees[J]. Computer Engineering, 2023, 49(10): 72-79.

Aspect-Level Sentiment Analysis for Multi-Information Learning by Fusing Syntax Trees

ZHANG Wenhao¹, LIAO Liefu^{1,2}, WANG Ruxia¹

(1. School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, Jiangxi, China;

2. School of Software Engineering, Jiangxi University of Science and Technology, Nanchang 330000, China)

[Abstract] In recent years, aspect-level sentiment analysis has mostly focused on mining semantic or syntactic information, without establishing an association between the two. In most existing models, the relative or syntactic distance of words is embedded into the models, ignoring the joint impact of relative and syntactic distances on aspect words and thus failing to fully consider the positional relationship of words in dependency syntax trees. In this study, an aspect-level sentiment analysis model, MILFST, is established for multi-information learning. This model integrates the syntactic tree and is expressly constructed to utilize the advantages of different neural networks with the purpose of obtaining richer information. The information of text sequences is captured through a bidirectional long and short term memory network, whereby the sequence information is updated based on the structure of the dependency syntax tree. With relative and syntactic distance positional information embedded into the text sequence, the learning of semantic and syntactic information is achieved through Convolutional Neural Networks (CNN) and Graph Convolutional Networks (GCN), respectively. The fusion of semantic and syntactic information is optimized through the attention mechanism, and the fused information is input into the Softmax classifier for sentiment polarity classification. The experimental results show that on the Twitter, Lap14, Res14, Res15, and Res16 datasets, the accuracies of the MILFST model are 74.27%, 77.74%, 82.50%, 81.73%, and 89.61%, respectively, and the F1 values are 73.14%, 74.27%, 74.54%, 66.15%, and 71.57%, respectively. The tree structure in the model helps capture information, and balancing the learning of syntactic and semantic information is beneficial for determining the emotional polarity of aspect words.

基金项目: 国家自然科学基金(71761018)。

作者简介: 张文豪(1997—), 男, 硕士研究生, 主研方向为自然语言处理、情感分析; 廖列法, 教授、博士; 王茹霞, 硕士研究生。

收稿日期: 2022-11-15 **修回日期:** 2022-12-15 **E-mail:** 1694374795@qq.com

[Key words] semantic information; syntactic information; Convolutional Neural Networks (CNN); Graph Convolutional Networks (GCN); attention mechanism

DOI: 10.19678/j.issn.1000-3428.0066266

0 概述

随着大数据时代的到来,互联网上产生的数据呈爆炸式增长,对数据分析的需求也越来越大。方面级情感分析是数据分析中的一种,旨在分析用户评论数据中的情感极性(如消极、中性或积极),方便企业推荐用户感兴趣的内容以及了解用户的喜好。

方面级情感分析早期大都采用机器学习方法^[1]进行分析,需要大量的人力对数据进行预处理,还需要通过人工建立情感词典并训练出分类模型,这对研究人员提出了很高的要求,且过程耗费精力,容易出错。随着深度学习技术的不断发展,各种神经网络模型^[2-3]被用于方面级情感分析任务中,大部分方法是通过神经网络结合注意力机制^[4-5]来完成情感分析。文献[6]利用长短时记忆网络单独对方面词和上下文进行建模,并利用交互式注意力网络分别学习并生成方面词和上下文的交互特征。文献[7]提出一种细粒度注意力机制,用于捕获方面词和上下文之间词级别的信息交互,然后将粗粒度注意力与细粒度注意力相结合用于最终的情感极性预测。上述方法大多通过神经网络模型学习句子中的语义信息,然后通过注意力机制获得文本序列与方面词的交互信息,它们大多都忽略了对语法信息的获取。

近年来,图卷积网络(Graph Convolutional Networks, GCN)^[8]被广泛应用于情感分析任务中。文献[9]利用GCN进行文本分类,然后通过依存句法树完成方面级情感分析。这种结合依存句法树和GCN的方法可以缩短方面词与情感词之间的距离,并利用GCN进行语法信息的学习。文献[10]利用句子的依存关系构建邻接矩阵,通过GCN学习句子中的语法关系,验证了GCN可以正确捕获句法信息和远程单词依赖关系。文献[11]提出一种新的网络架构,利用多层图注意力网络将情感特征词与方面词相连接。文献[12]认为普通的邻接矩阵缺少了词与词之间的依存关系和共现关系,其将层次句法图和层次词汇图(Hierarchical Lexical Graph, HLG)代替普通的邻接矩阵来进行语法学习。例如,“food was okay, nothing special.”句子中的“nothing special”单词对在训练集中出现了5次,并且该单词对表示消极的情感极性,在这种情况下,如果要正确预测“food”方面词的情感极性,则需要有其他信息来抵消句中出现的“okay”的积极性。

受上述方法的启发,本文提出MILFST模型,该模型同时进行语义信息和语法信息的学习,通过使用层次词汇图代替普通的依存句法树来完成语法信息的学习,并借助层次词频图代替普通的依存句法树进行图卷积网络的语法信息学习。在此基础上,

将相对距离和语法距离信息融合后嵌入模型中。本文的主要工作如下:

1)将相对距离和语法距离融合后嵌入模型中,同时考虑单词的相对距离和语法距离对方面词的影响。

2)考虑单词在依存句法树上不同位置对分析结果的影响,分析单词在句法树中的高度以及单词节点所含有的度数在文本序列中的重要程度。

3)同时对语义信息和语法信息进行学习,使文本序列含有更丰富的信息,更充分地提取方面词信息。

4)引入层次词汇图来代替普通的依存句法树,既能捕捉语法信息,又能关注词汇之间的共现关系,从而提高语法学习的效率。

1 相关工作

通过神经网络模型处理方面级情感分析任务,大多都是使用词嵌入层^[13-14]作为模型的起始层,主要是因为词嵌入层能将文本序列信息一一映射到低维的向量空间,然后通过各种深度学习模型来对映射到低维向量空间中的信息进行学习。卷积神经网络(Convolutional Neural Networks, CNN)早期是用在图像处理任务中,主要通过卷积核提取图像的局部信息。自从CNN被用在自然语言处理任务中^[15],研究人员开始利用CNN结合其他模型进一步提取特征信息。文献[16]采用多层CNN来并行地对上下文进行建模,然后利用注意力机制来关联上下文和方面词之间的信息。文献[17]提出一种CNN结合门控机制的模型,该模型过滤与方面词信息无关的情感特征,然后根据方面词有选择性地输出相关的情感特征。

有研究表明^[18-19],在方面级情感分析任务中,句中每个词对方面词都有不同程度的影响,距离方面词越远的词对方面词影响越小,距离方面词越近的词对方面词影响越大。因此,句中单词和方面词之间的距离可以分为相对距离和语法距离。文献[20]利用加权卷积网络来探索相对位置关系和语法距离关系,研究结果表明,语法距离关系比相对距离关系更有利于方面词分类。文献[21]将语法位置信息融合到模型中,实验结果证明了语法位置信息能够让模型更好地理解上下文和方面词之间的关系。

随着情感分析研究的不断深入,语法信息被不断地挖掘。例如,通过使用语法解析工具,可以将句子“The food was excellent as well as service, however, I left the four seasons very disappointed.”解析成一棵具有语法信息的依存句法树,如图1所示。句子中有3个方面词,即 food、service 和 seasons。从

句中可以推断出方面词“food”和“service”是积极的,“seasons”是消极的。以往许多模型在一些情况下会关注错误的单词。在图1中,对于“service”方面词,很多模型会错误地关注“well”,但是从整个句子来

看,判断“service”方面词的情感极性依赖的是“excellent”而不是“well”。从依存句法树上看,“service”通过依存关系“attr”和“acomp”与“excellent”相连接,减少了“well”对方面词的影响。

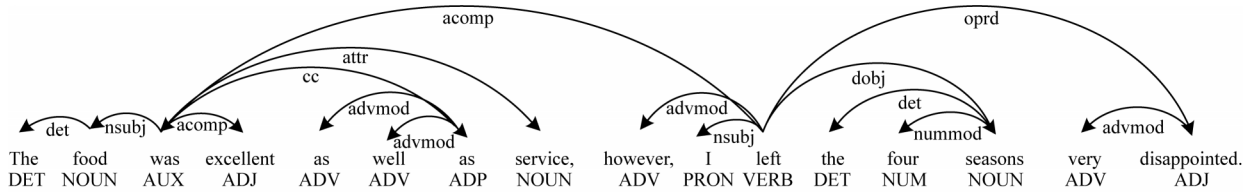


图1 依存句法树

Fig.1 Dependency syntax tree

图卷积网络的优点是能够很好地捕获依存句法树的结构信息,弥补其他模型难以捕获语法信息的缺陷。文献[22]通过键值记忆网络有选择性地利用依赖信息,然后根据记忆机制对不同的依赖信息进行加权,有效地甄别无用的信息。文献[12]使用层次句法图和层次词汇图来进行语法信息学习。层次词汇图将语料库中2个词出现的频率作为连接因素,根据这个因素对每一个句子构建层次词汇图。如图2所示,在语料库中,很多词共同出现的概率很低,只有少数一些词的共现概率很大。因此,层次词汇图根据共现次数的对数的正态分布频率进行分组。其中, d_1 和 d_2 分别表示共现频率为 2^0 和 2^1 的词对, $d_3, \dots, d_5$ 表示共现频率在 $[2^{k+1}, 2^{k+2}]$ ($1 \leq k \leq 5$)区间的词对, d_6 表示共现频率大于 2^6 的词对。最后,基于词汇共现频率构造层次词汇图 $HLG = \{V_d, E_d, R_d\}$,其中, V_d 表示图中的节点集合, E_d 表示图中的边集合, R_d 表示词对共现关系的集合。

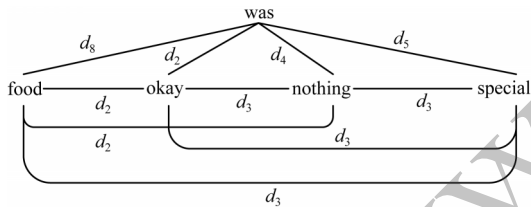


图2 层次词汇图

Fig.2 Hierarchical lexical graph

2 模型结构

给定一个文本序列 $s = \{w_1, w_2, \dots, w_n\}$ 和方面词序列 $a = \{w_i, w_{i+1}, \dots, w_{i+m-1}\}$, 其中, 文本序列中含有 n 个词, 方面词序列中含有 m 个词, 并且方面词序列 a 是文本序列 s 中的一个子序列。本文模型结构如图3所示, 模型总共分为7层, 分别是词嵌入层、隐藏层、位置嵌入层、学习层、注意力层、特征提取层和池化层。其中, 学习层分为2个模块, 分别是语义学习模块和语法学习模块, 注意力层也分为2个部分, 分别是语义优化注意力和语法优化注意力。

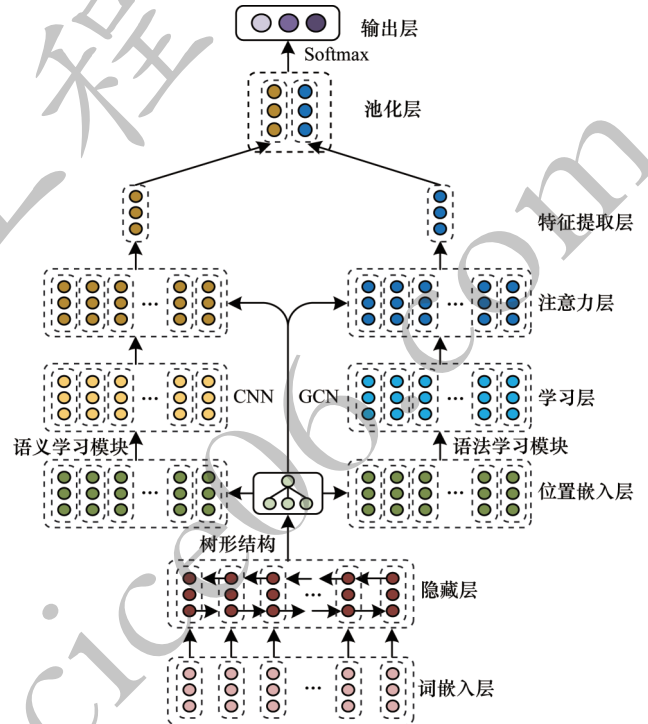


图3 MILFST 模型结构

Fig.3 MILFST model structure

2.1 词嵌入层和隐藏层

本文使用 GloVe 作为词嵌入层, 通过 GloVe 预训练模型, 将文本序列中的每个词一一映射到向量空间中, 然后每个词都由向量表示, 词向量 $w_i \in \mathbb{R}^d$, 其中, d_w 表示词向量的维度。再将词向量输入到双向长短时记忆网络中获得隐藏层向量 $H = \{h_1, h_2, \dots, h_n\}$ 。隐藏层向量计算公式如下:

$$h_i = [\overrightarrow{\text{LSTM}}(w_i), \overleftarrow{\text{LSTM}}(w_i)], i \in [1, n] \quad (1)$$

其中: $\overrightarrow{\text{LSTM}}$ 表示前向执行长短时记忆网络; $\overleftarrow{\text{LSTM}}$ 表示后向执行长短时记忆网络。

2.2 位置嵌入层

2.2.1 树形结构

依存句法树代表的是一个句子的语法信息, 现有研究目前还未利用依存句法树中节点的度和深度进行信息更新, 更多的是利用依存句法树进行图卷积操作。

句法树表示的是一个句子的语法信息,如图4所示,该树由图1中的依存句法树简化而来。“The”“as”“well”和“very”等词都是一些修饰词,这些词在语法中起到加强的作用,但是,模型在学习一些信息时会受到这些词的影响。例如,单独“well”表示好的意思,在句中,它只是作为一个连词,起连接的作用,如果只是对这个句子进行学习,“well”会影响“service”的判断。在句法树中,“well”距离“service”较远,对“service”的影响较小。在此句中的修饰词可能会在其他句中修饰含义相反的词,从而方面词情感极性的判断。从依存句法树中可以看出,往往一些修饰词或者噪声词的深度更深,而且该节点的度都是1。因此,在学习句子信息之前,需要尽量降低一些修饰词和噪声词的影响。

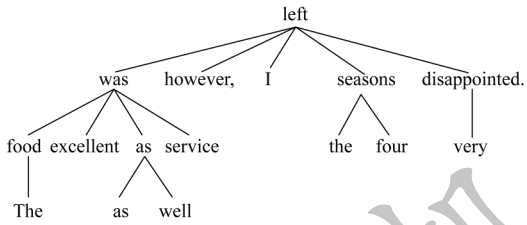


图4 简化的依存句法树

Fig.4 Simplified dependency syntax tree

式(2)、式(3)分别是词节点的度和深度的权重公式:

$$d_e = \frac{d_i}{d_{\max}} \quad (2)$$

$$g_h = 1 - \frac{g_i}{g_{\max} + 1} \quad (3)$$

其中: d_i 和 $d_{\max}$ 分别表示当前节点的度和依存句法树中最大的度; g_i 和 $g_{\max}$ 分别表示当前节点所在的深度和依存句法树中最大的深度。最后,根据句法树的2个权重公式进行信息更新,如下:

$$h_{i,t} = \frac{d_e + g_h}{2} \cdot h_i \quad (4)$$

2.2.2 位置编码

本文采用位置编码来关联上下文和方面词之间的关系。位置编码分为语法位置编码和相对位置编码。语法位置编码在依存句法树中考虑单词与方面词之间的距离。例如,在图4中,方面词“food”与“excellent”通过“was”相连,因此,它们的语法距离为2。语法位置编码计算公式如下:

$$s_i = \begin{cases} 1 - \frac{d_i}{n}, & 0 \leq i < t, t+m \leq i < n \\ 1, & t \leq i < t+m \end{cases} \quad (5)$$

其中: d_i 表示方面词与其他词在依存句法树中的距离; n 表示句中单词的个数。

相对位置编码是根据每个词与方面词的距离不同进行权重分配,距离方面词越远的词信息就越弱,距离方面词越近的词信息就越强。根据此分配方

法,相对位置编码计算公式如下:

$$r_i = \begin{cases} 1 - \frac{t+1-i}{n}, & 1 \leq i < t+1 \\ 1, & t+1 \leq i \leq t+m \\ 1 - \frac{i-t-m}{n}, & t+m < i \leq n \end{cases} \quad (6)$$

通过式(5)和式(6)对上下文进行位置编码后,需要将编码后的信息嵌入模型中,嵌入后隐藏层表示如下:

$$p_i = \frac{e_{s,i} + e_{r,i}}{2} \quad (7)$$

$$h_{i,p} = p_i h_{i,t} \quad (8)$$

其中: $e_{s,i}$ 表示语法位置嵌入表示; $e_{r,i}$ 表示相对位置嵌入表示。

2.3 学习层

一个句子同时包含语义信息和语法信息,从以往的研究中可以得出^[12,18],这2种信息对于方面词情感极性判断都有作用。本文模型设计2个模块分别对语义信息和语法信息进行学习。

2.3.1 语法学习模块

图卷积是在传统卷积神经网络上的一种改进,主要是对图结构进行操作。对于图结构中的每个节点,图卷积都要考虑该节点自身的特征信息以及它所有邻居节点的特征信息,这样可以有效地获取2个相连节点之间的信息。在进行方面级情感分析时,可以根据依存句法树中的边进行信息的传递与获取。例如,对于图4中的方面词“service”,在依存句法树中该节点连接的是“was”节点,“was”节点连接的是“excellent”节点。因此,在进行信息获取时,“excellent”节点会将特征信息先传入“was”节点中,然后再将特征信息传入“service”中。在进行信息获取时,这缩短了相对位置中传递信息的距离,有效减少了传递过程中的噪声。

由于普通的图卷积操作中不含带标记边的图,因此本文利用文献[12]中的层次词汇图进行语法学习,该图不仅包含句法树结构,还包含词之间的共现关系。通过该图卷积,能够将相同共现关系的词合并到虚拟节点,然后对合并的所有虚拟节点进行信息更新。虚拟节点信息更新如下:

$$h_{i,l} = \text{ReLU}\left(w_l \cdot \left(\oplus_r \tilde{h}_{i,l,r}\right)\right) \quad (9)$$

其中: $\oplus_r$ 表示不同共现关系类型的连接; l 表示层数; w_l 表示第 l 层中的权重; $\tilde{h}_{i,l,r}$ 表示每一种共现关系 r 的表示。

2.3.2 语义学习模块

在图卷积出现之前,大多数的模型都是通过对语义的学习来提高对方面词情感极性的判断。本文使用多层卷积神经网络对语义信息进行学习。卷积神经网络通过卷积核将一个文本序列对应的矩阵以

滑动窗口的方式提取相邻单词之间的信息。卷积核的高通常设为3,长为词向量的维度。因此,在进行卷积操作时,往往会被一些噪声词影响。本文模型在语义学习之前,在位置编码时就对噪声词进行削弱。因此,相对于文献[16]中的模型,本文模型在进行语义学习时减少了一些对噪声词的操作。利用式(10)进行语义信息的学习:

$$h_{i,c,l} = \text{CNN}(h_{i,p}) \quad (10)$$

其中:CNN表示卷积操作; l 表示卷积神经网络层数。

2.4 注意力层

在学习层进行语义和语法学习之后,需要对这2个部分信息进行优化融合。考虑到语义信息和语法信息存在一些差异性,直接将两者进行注意力融合会导致融合不充分。因此,为了更好地融合这2种信息,借助树形结构化的信息分别对语义信息和语法信息进行注意力融合优化。注意力层分为语义优化注意力和语法优化注意力2个部分,分别对语义信息和语法信息进行融合优化。2个部分具体如下:

1) 语义优化注意力。

语义优化注意力主要是将语义学习模块学到的信息 $h_{i,c,l}$ 与树形结构化信息进行交互优化,如式(11)~式(13)所示:

$$\tilde{a}_i = h_{i,t} h_{i,c,l} \quad (11)$$

$$a_i = \frac{\exp(\tilde{a}_i)}{\sum_{i=1}^n \exp(\tilde{a}_i)} \quad (12)$$

$$S_{i,a} = \sum_{i=1}^n a_i h_{i,c,l} \quad (13)$$

其中: $\exp$ 表示以 e 为底的指数函数; $S_{i,a}$ 表示语义优化注意力的输出。

2) 语法优化注意力。

语法优化注意力主要是将语法学习模块学习到的语法信息 $h_{i,l}$ 与树形结构化信息进行交互优化,如式(14)~式(16)所示:

$$\tilde{\beta}_i = h_{i,t} h_{i,l} \quad (14)$$

$$\beta_i = \frac{\exp(\tilde{\beta}_i)}{\sum_{i=1}^n \exp(\tilde{\beta}_i)} \quad (15)$$

$$G_{i,\beta} = \sum_{i=1}^n \beta_i h_{i,l} \quad (16)$$

其中: $G_{i,\beta}$ 表示语法优化注意力的输出。

2.5 特征提取层

方面词的信息对方面词情感极性的判断尤为重要,特征提取层则是对方面词信息进行提取,分别从语义信息和语法信息中提取方面词的信息。方面词信息提取公式如下:

$$a = \begin{cases} 0, 1 \leq i < t, t+m \leq i \leq n \\ h_i, t \leq i < t+m \end{cases} \quad (17)$$

其中: a 表示方面词信息。使用式(17)分别提取语义信息和语法信息中的方面词信息,分别记为 a_s 和 a_g 。

2.6 池化层

在提取方面词信息之后,将提取的方面词信息进行整合,对 a_s 和 a_g 进行拼接。方面词信息整合公式如下:

$$o_a = [a_g, a_s] \quad (18)$$

将整合后的方面词信息通过池化层进行最大池化操作,进一步筛选方面词的有效信息,如下:

$$o_p = \max_pooling(o_a) \quad (19)$$

其中: $\max_pooling$ 表示最大池化。

2.7 输出层

将得到的方面词信息输入Softmax层进行分类,分类结果分为积极性质、中性性质和消极性质3类。最终的情感极性分类结果如下:

$$y = \text{Softmax}(W o_p + B) \quad (20)$$

其中:Softmax表示分类器; W 表示权重矩阵; B 表示偏置矩阵。

2.8 模型训练

本文模型使用标准梯度下降算法进行参数的学习训练,训练算法为交叉熵损失函数和L2正则化项:

$$\zeta(\theta) = \sum_{i=1}^c \hat{y}_i \log_a y_i + \lambda \sum_{\theta \in \Theta} \theta^2 \quad (21)$$

其中: c 表示情感标签的个数; $\hat{y}_i$ 表示模型的输出情感值; y_i 表示标签中的真实值; λ 表示L2正则化参数; Θ 表示模型中所用到的参数。

3 实验验证

3.1 实验环境及参数设置

本文模型是在搭载Intel® Xeon® W-2123 CPU @ 3.60 GHz的CPU和NVIDIA GeForce RTX 2080 Ti的GPU服务器上运行的。实验使用Python语言进行编程,Python版本为Python 3.8.8,编程工具为PyCharm,版本为PyCharm 2021.2.1社区版。

模型使用GloVe^[14]作为词嵌入层对文本进行初始化,初始化词向量维度为300,双向长短时记忆网络的隐藏层维度与词向量维度相同,设置为300。选择Adam作为模型的优化器,模型学习率为0.001,L2正则化项系数为0.000 01,批量训练大小为32。

3.2 实验数据集

实验所使用的数据集是5个公开的英文方面词情感分析数据集,分别是Twitter^[23]、SemEval-2014 Task 4中的Lap14和Res14^[24]、SemEval-2015 Task 12中的Res15^[25]和SemEval-2016 Task 5中的Res16^[26]。SemEval-2014 Task 4含有对笔记本评价的数据和对餐厅评论的数据。数据集标签分布如表1所示。

表 1 数据集标签分布

Table 1 Datasets label distribution

数据集	积极		中性		消极	
	Train	Test	Train	Test	Train	Test
Twitter	1 561	173	3 127	346	1 560	173
Lap14	994	341	464	169	870	128
Res14	2 164	728	637	196	807	196
Res15	912	326	36	34	256	182
Res16	1 260	469	69	30	439	117

3.3 评价指标

实验采用 2 个评价指标,分别是准确率(计算中用 A_{Acc} 表示)和 Macro-F1(计算中用 F_1 表示)。 A_{Acc} 表示分类正确的样本数占比,计算如式(22)所示。 F_1 是精确率和召回率的调和平均,计算如式(23)~式(25)所示。

$$A_{\text{Acc}} = \frac{T_{\text{TP}} + T_{\text{TN}}}{T_{\text{TP}} + T_{\text{TN}} + F_{\text{FP}} + F_{\text{FN}}} \quad (22)$$

$$F_1 = \frac{2P \times R}{P + R} \quad (23)$$

$$P = \frac{1}{c} \sum_{i=1}^c \frac{T_{\text{TP},i}}{T_{\text{TP},i} + F_{\text{FN},i}} \quad (24)$$

$$R = \frac{1}{c} \sum_{i=1}^c \frac{T_{\text{TP},i}}{T_{\text{TP},i} + F_{\text{FP},i}} \quad (25)$$

其中: T_{TP} 表示分类正确的积极标签的样本数; T_{TN} 表示分类正确的消极标签的样本数; F_{FP} 表示分类错误的积极标签的样本数; F_{FN} 表示分类错误的消极标签的样本数; c 表示情感类别的标签种类个数。

3.4 对比实验及分析

为了评估 MILFST 模型的有效性,选择如下具有代表性的模型进行对比:

1)SVM^[1],该模型通过支持向量机方法进行方面级情感分类。

2)ATAE-LSTM^[27],该模型是将文本序列与方面词进行拼接以融合信息,通过 LSTM 和注意力机制进行情感极性分类。

3)GCAE^[17],该模型使用卷积神经网络提取特征信息,然后通过门控机制过滤与方面词信息无关的情感特征,进而有选择性地输出相关的情感特征。

4)IAN^[6],该模型利用 LSTM 分别对文本序列和方面词进行建模,利用交互注意力交叉融合文本序列和方面词信息,并生成特定的文本序列信息和方面词信息。

5)AOA^[28],该模型利用注意力机制关注重要信息的特点,将方面词和句子以联合的方式进行建模,进而捕获方面词和句子之间的交互信息。

6)ASGCN^[10],该模型利用依存句法树建立一个依存句法图,然后通过图卷积操作进行语法信息学习。

7)TD-GAT^[11],该模型利用单词之间的依赖关系进行情感特征传递。

8)Bi-GCN^[12],该模型在依存句法树上提出层次语法图和层次词汇图,通过双层交互式图卷积网络学习语法信息。

在上述对比模型中:前 5 个模型基于机器学习、深度学习进行语义信息学习,这些模型没有学习语法信息;后 3 个模型通过依存句法关系对句子的语法信息进行学习。各模型在数据集上的实验结果如表 2 所示,最优结果加粗标注。

表 2 各模型的实验结果

Table 2 Experimental results of each model

%

模型	Twitter		Lap14		Res14		Res15		Res16	
	A_{Acc}	F_1	A_{Acc}	F_1	A_{Acc}	F_1	A_{Acc}	F_1	A_{Acc}	F_1
SVM	63.40	63.30	70.49	—	80.16	—	—	—	—	—
ATAE-LSTM	69.65	67.40	69.14	63.18	77.32	66.57	75.43	56.34	83.25	63.85
GCAE	71.64	69.88	69.90	65.71	78.57	68.06	77.85	59.63	86.29	65.87
IAN	72.50	70.81	72.05	67.38	79.26	70.09	78.54	52.65	84.74	55.21
AOA	72.30	70.20	72.62	67.52	79.97	70.42	78.17	57.02	87.50	66.21
ASGCN	72.15	70.40	75.55	71.05	80.77	72.02	79.89	61.89	88.99	67.48
TD-GAT	72.20	70.45	75.63	70.74	81.32	71.72	80.38	60.50	87.71	67.87
Bi-GCN	74.16	73.35	74.59	71.84	81.97	73.48	81.16	64.79	88.96	70.84
MILFST	74.27	73.14	77.74	74.27	82.50	74.54	81.73	66.15	89.61	71.57

从表 2 可以看出,基于深度学习的模型普遍优于基于机器学习的模型,这是因为深度学习能够自主学习句子中的特征信息,捕捉的信息更为丰富和完整。将前 5 个模型和后 4 个模型进行对比可以看出,前 5 个模型的指标值明显小于后 4 个模型,这主要是因为后 4 个模型考虑了语法信息,而前 5 个模型

未考虑语法信息,这也验证了句子的语法信息能够有效帮助模型识别方面词的情感极性。

ASGCN 模型和 TD-GAT 模型在 5 个数据集中实验结果相差不大:在 Twitter 数据集中,TD-GAT 模型的准确率和 F1 值都大于 ASGCN 模型,但是相差不大;在 Lap14 数据集、Res14 数据集和 Res15 数据集

中,TD-GAT模型的准确率高于ASGCN模型,但是F1值小于后者;在Res16数据集中,虽然TD-GAT的准确率比ASGCN低,但是F1值却比ASGCN高。因此,只是单独地通过依存句法树进行语法信息学习,很难再提高准确率。基于此,Bi-GCN在依赖图中增加了词与词之间的连接关系,从表中数据可以看出,Bi-GCN在增加连接关系之后准确率和F1值明显高于前2个模型。

与ASGCN和TD-GAT相比,本文MILFST模型在5个数据集中准确率和F1值都更优,这说明模型在考虑语法信息时也不能忽略语义信息,语义信息和语法信息对方面词情感极性的判断都很重要。

与Bi-GCN模型相比,MILFST模型在5个数据集上结果几乎都有提升,除了在Twitter数据集上

F1值小于Bi-GCN外,在其他数据集上F1值都高于Bi-GCN,在另外4个数据集上F1值分别提高2.43、1.06、1.36和0.73个百分点,在5个数据集上准确率分别提高0.11、3.15、0.53、0.57和0.65个百分点。这也验证了模型同时对文本序列的语义信息和语法信息进行学习是有意义的。

3.5 消融实验

为了验证MILFST模型中每个组件对模型性能的重要性,设置一系列的消融实验进行测试,实验结果如表3所示。其中:w/o pos表示消融了位置信息;w/o tree表示消融了树形结构化信息;w/o atn表示消融了注意力层;w/o sy表示消融了语法学习模块;w/o se表示消融了语义学习模块;w/o asp表示消融了特征提取层。

表3 消融实验结果

Table 3 Results of ablation experiment

模型	Twitter		Lap14		Res14		Res15		Res16		%
	A_{Acc}	F_1	A_{Acc}	F_1	A_{Acc}	F_1	A_{Acc}	F_1	A_{Acc}	F_1	
w/o pos	73.55	71.99	76.48	72.12	80.98	72.65	80.44	66.03	88.96	70.01	
w/o tree	73.26	71.49	75.54	71.88	82.23	74.18	80.25	65.67	88.14	70.71	
w/o atn	74.13	72.56	78.84	75.07	80.62	70.99	80.07	64.99	88.79	69.95	
w/o sy	73.55	71.57	75.70	72.12	81.42	73.80	80.25	64.67	89.12	73.34	
w/o se	73.12	71.68	74.60	69.35	80.00	70.80	79.70	66.06	88.47	70.59	
w/o asp	73.69	71.92	74.45	70.95	79.64	70.41	79.70	62.41	89.28	69.86	
MILFST	74.27	73.14	77.74	74.27	82.50	74.54	81.73	66.15	89.61	71.57	

从表3的消融实验结果可以看出:

1)在消融位置信息和树形结构化信息后,w/o pos和w/o tree的性能指标下降,说明每个词的位置信息和树形结构化信息会对方面词的判断结果产生影响。在5个数据集中,w/o pos在Res14数据集上下降幅度大于w/o tree,在另外4个数据集上w/o tree下降幅度更大,说明树形结构化信息对方面词的影响比位置信息更大。

2)在消融了注意力层之后,在Lap14数据集上,模型性能指标反而有所上升,在其他数据集上都有所下降,说明在Lap14数据集上学习的语义信息和语法信息已经能够表达方面词的情感极性,但是注意力机制过度聚焦信息,反而对语义信息和语法信息的融合有抑制作用。在另外4个数据集上,注意力机制能有效融合语义信息和语法信息。

3)在实验中分别消融了语法学习模块和语义学习模块后,w/o sy和w/o se性能指标都有下降,但是w/o se的下降幅度比w/o sy更大,说明语义学习模块和语法学习模块在MILFST模型中同样重要,但是MILFST模型对语义信息的学习依赖性更高一些。

4)在消融了特征提取层后,w/o asp性能指标有所下降,说明在最后对方面词进行特征提取是必

要的。

从整个实验结果来看,将模型中的一部分组件进行消融之后,模型的性能指标有明显下降,说明模型中的每一个部分都十分重要,缺一不可。消融每一个组件后模型的性能指标下降幅度也各不相同,说明每个部分对于不同数据集的敏感程度不同,主要取决于每个数据集中句子的完整性和长度。

4 结束语

本文提出一种同时学习语义信息和语法信息的MILFST模型。该模型在学习语义信息和语法信息之前,通过文本序列在依存句法树中的位置信息以及其自身的相对位置信息和语法位置信息对文本序列进行更新,从而减轻文本序列中存在的一些噪声词的影响,使得在进行信息学习之前原本的文本序列中就含有语法相关信息。实验结果表明,MILFST模型能够有效学习语义信息和语法信息,且模型中的各个部分都对情感极性判断有积极作用。

经研究发现,MILFST模型高度依赖语句的完整性,对于不完整语句的分类效果不理想。下一步将尝试使用知识图谱和情感知识来提高模型对方面词信息的理解,增强语义信息和语法信息的学习效果,进一步提高模型对方面词信息的分类性能。

参考文献

- [1] KIRITCHENKO S, ZHU X D, CHERRY C, et al. NRC-Canada-2014: detecting aspects and sentiment in customer reviews[C]//Proceedings of the 8th International Workshop on Semantic Evaluation. [S. l.]: Association for Computational Linguistics, 2014: 437-442.
- [2] PORIA S, CAMBRIA E, HAZARIKA D, et al. A deeper look into sarcastic tweets using deep convolutional neural networks[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1610.08815>.
- [3] TANG D, QIN B, FENG X, et al. Effective LSTMs for target-dependent sentiment classification[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1512.01100>.
- [4] MNIH V, HEES N, GRAVES A, et al. Recurrent models of visual attention[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1406.6247>.
- [5] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1409.0473>.
- [6] MA D, LI S, ZHANG X, et al. Interactive attention networks for aspect-level sentiment classification[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1709.00893>.
- [7] FAN F F, FENG Y S, ZHAO D Y. Multi-grained attention network for aspect-level sentiment classification[C]//Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. [S. l.]: Association for Computational Linguistics, 2018: 3433-3442.
- [8] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1609.02907>.
- [9] YAO L, MAO C S, LUO Y. Graph convolutional networks for text classification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 7370-7377.
- [10] ZHANG C, LI Q, SONG D. Aspect-based sentiment classification with aspect-specific graph convolutional networks[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1909.03477>.
- [11] HUANG B, CARLEY K M. Syntax-aware aspect level sentiment classification with graph attention networks[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1909.02606>.
- [12] ZHANG M, QIAN T Y. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis[C]//Proceedings of 2020 Conference on Empirical Methods in Natural Language Processing. [S. l.]: Association for Computational Linguistics, 2020: 3540-3549.
- [13] 施荣华, 金鑫, 胡超. 基于图注意力网络的方面级别文本情感分析[J]. 计算机工程, 2022, 48(2): 34-39.
- [13] SHI R H, JIN X, HU C. Aspect-level text emotion analysis based on graph attention network[J]. Computer Engineering, 2022, 48(2): 34-39. (in Chinese)
- [14] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation[C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. [S. l.]: Association for Computational Linguistics, 2014: 1532-1543.
- [15] CHEN Y. Convolutional neural network for sentence classification[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1408.5882>.
- [16] ZHANG S Q, XU X Y, PANG Y W, et al. Multi-layer attention based CNN for target-dependent sentiment classification[J]. Neural Processing Letters, 2020, 51(3): 2089-2103.
- [17] XUE W, LI T. Aspect based sentiment analysis with gated convolutional networks[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1805.07043>.
- [18] LI X, BING L, LAM W, et al. Transformation networks for target-oriented sentiment classification[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1805.01086>.
- [19] TANG D, QIN B, LIU T. Aspect level sentiment classification with deep memory network[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1605.08900>.
- [20] ZHANG C, LI Q C, SONG D W. 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. New York, USA: ACM Press, 2019: 1145-1148.
- [21] PHAN M H, OGUNBONA P O. Modelling context and syntactical features for aspect-based sentiment analysis[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [S. l.]: Association for Computational Linguistics, 2020: 3211-3220.
- [22] TIAN Y H, CHEN G M, SONG Y. Enhancing aspect-level sentiment analysis with word dependencies[C]//Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. [S. l.]: Association for Computational Linguistics, 2021: 3726-3739.
- [23] DONG L, WEI F R, TAN C Q, 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. [S. l.]: Association for Computational Linguistics, 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. [S. l.]: Association for Computational Linguistics, 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. [S. l.]: Association for Computational Linguistics, 2015: 486-495.
- [26] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2016 Task 5: aspect based sentiment analysis[C]//Proceedings of the 9th International Workshop on Semantic Evaluation. [S. l.]: Association for Computational Linguistics, 2016: 19-30.
- [27] WANG Y Q, HUANG M L, ZHU X Y, et al. Attention-based LSTM for aspect-level sentiment classification[C]//Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing. [S. l.]: Association for Computational Linguistics, 2016: 606-615.
- [28] HUANG B, OU Y, CARLEY K M. Aspect level sentiment classification with attention-over-attention neural networks[EB/OL]. [2022-10-05]. <https://arxiv.org/abs/1804.06536>.