

doi:

一种结合序列与语法关系的方面级情感分

针对的问题 现在方法的缺陷 我们的解决方案 我们的
优点 析方法

凌键军¹, 李志鹏¹, 陈丹阳¹, 钟诚^{1,2}

(1. 广西大学 计算机与电子信息学院, 广西 南宁 530004;

2. 广西高校并行分布于智能计算重点实验室, 广西 南宁 530004)

特点一: 结合序列结构和语法
结构关系
特点二: 一种新的融合方式

摘要: 目前大多数方面级情感分类的研究主要是使用注意力机制结合依赖树的方法来对方面词及其上下文关系进行建模。然而这些方法过于关注句子的语法结构情感特征信息, 它们忽略了序列结构关系中的情感特征信息, 只凭借语法结构的情感特征信息没有办法处理复杂的句子关系。所以, 这些方法没有能够充分获取到句子的情感特征信息, 限制了模型的分类效果。针对这些问题, 该文提出了一种能够结合两种句子结构关系信息的序列结构和语法结构信息融合的网络(Sequence-sYntax Information Fusion Network, SYFN)模型。这个模型能够同时处理句子的序列结构和语法结构关系, 并且充分利用两种结构关系的情感特征信息进行情感分类。实验结果表明, 该模型与基线模型相比性能有较明显的提升。提升了多少个点

关键词: 神经网络; 自然语言处理; 方面级情感分析; 语法依赖树; 信息融合

中图分类号: TU 473.1

文献标识码: A

文章编号:

An aspect-level sentiment classification method combining sequence and syntactic relationships

JianJun Ling¹, Zhipeng Li¹, Danyang Chen¹, Cheng Zhong^{1,2}

(1. School of Computer, Electronics and Information, Guangxi University, Nanning 530004, Guangxi, China;

2. Key Laboratory of Parallel, Distributed and Intelligent Computing in Guangxi Universities and Colleges, Nanning, Guangxi, 530004, China)

Abstract: Most current research on aspect-level sentiment classification mainly uses attention mechanism combined with dependency tree to model aspect words and their contextual relationships. However, these methods pay too much attention to the sentiment feature information of the syntactic structure in sentences. They ignore the sentiment feature information in the sequence structure relationship, so they have no way to deal with the complex sentence relationship only based on the sentiment feature information of the syntactic structure. Therefore, these methods are not able to fully extract the sentiment feature information of the sentences, which limits the classification effect of the model. To address these issues, this paper proposes a Sequence-sYntax Information Fusion Network (SYFN) model that can combine the sequence structure and syntactic structure information of sentence. This model can simultaneously process the sequence structure and syntactic structure relationships of sentences, and make full use of the sentiment feature information of the two structural relationships for sentiment classification. The experimental results show that the performance of this model is significantly improved compared with the baseline model.

Key words: neural network; natural language processing; aspect-level sentiment classification; dependency tree; information fusion

收稿日期: 2021-12-01 ; **修订日期:** 2022-04-11

基金资助: 国家自然科学基金项目(61962004); 广西大学引进人才博士科研启动基金项目()

通讯作者: 陈丹阳(1988 -), 女, 黑龙江齐齐哈尔人, 广西大学讲师, 硕士生导师, 博士; E-mail: chendanyang@gxu.edu.cn。

0 引言

方面级情感分类是一种细粒度的文本情感分类任务，它可以识别出句子中方面词所表达的情感极性，情感极性一般分为积极、消极和中性3类。如果一个句子包含多个方面词，则分别对每个方面词都进行情感极性判断。例如，在图1的句子中，这个句子包含了两个方面词“performs”和“price”，它们的情感极性分别是“积极”和“消极”，方面级情感分类的目标是分别预测句子中这两个方面词的情感极性，被广泛应用于情感聊天机器人^[1-2]和推荐系统^[3-4]等领域。

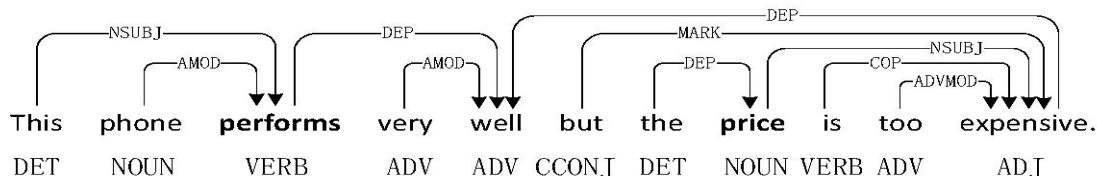


图1 一个带有依赖关系树的示例句子

Fig.1 An example sentence with its dependency tree

方面级情感分类的核心问题是如何从句子中获取与方面词相关的情感信息。起初，一些研究者将神经网络模型应用于方面级情感分类中，取得了一定的成绩。Tang等^[5]使用长短记忆网络(LSTM)从句子的上下文中提取特征信息，获取与方面词相关的情感特征信息进行分类。Ma等^[6]分别对方面词和上下文进行建模，利用注意力机制针对方面词学习上下文的特征向量，从而提取有利于方面级情感分类的特征信息。Song等^[7]使用注意力编码器网络(AEN)从单词嵌入中挖掘出丰富的语义信息。这些模型虽然利用了句子与方面词之间的序列结构信息，但却忽视了语法结构信息在方面级情感分类中的重要性，难以分析复杂的方面词关系。

近年来，一些研究者尝试将语法依赖树应用于方面级情感分类。Zhang等^[8]提出了一种基于特定方面的图卷积网络(ASGCN)模型，模型使用图卷积网络(GCN)结合语法依赖树对方面词及上下文关系进行建模。然而该模型只是简单地利用语法依赖树的邻接矩阵关系，丢失了其他的语法依赖信息。同时，GCN网络忽视了句子中每个单词对方面词的重要程度不同的这一特点，平等地处理各个单词之间的关系，有失偏颇。为了更好地利用语法依赖树的信息，Wang等人^[9]提出了一种面向方面词的语法依赖树的关系图注意网络(R-GAT)模型。该模型修剪语法依赖树，人为构造面向方面词的依赖树关系，使得模型得到更丰富的语法信息，提升了模型的性能。然而，面向方面的依赖树的构建方式是主观的，对语法依赖树的修剪会导致重要语法信息的丢失。同时，单词之间依赖关系的编码过于简单，无法充分表示单词之间的语法关系。

针对以上问题，本文提出了一种序列结构和语法结构信息融合的网络(Sequence-sYntax Imformation Fusion Network, SYFN)模型。SYFN模型结合了语法依赖关系、词性标注信息以及相对位置信息多种句子语法与语义知识，能够同时处理句子的序列结构信息和语法结构信息，学习获得更多的情感特征知识，增强模型的情感表达能力。SYFN模型主要由结构关系网络和多层融合网络组成。结构关系网络包含多层，每一层都包含三个主要模块：序列结构网络、语法结构网络和结构信息融合网络。序列结构网络能够处理句子的序列结构关系，提取句子序列结构的情感特征信息。语法结构网络能够结合语法依赖树信息，提取句子语法结构中所表达的情感特征信息。结构信息融合网络能够融合两种结构关系的情感特征信息，增强模型对句子关系总体情感的建模能力，使模型能够更好地捕获与方面词相关的情感特征信息。多层融合网络能够整合结构关系网络中每一层的结构融合情感特征信息，有效地将下层简单关系的情感特征信息和高层抽象、复杂的情感特征信息相结合，进一步提高了该模型的情感分类效果。

本文的主要贡献如下：

1) 本文设计了两种结构网络来处理句子的序列结构信息和语法结构信息。它们可以处理多种语

- 1 仅仅是单纯描述 没有进行分析 在讲到一种结构时要简要地分析他们的特点 作用
- 2 最好能够将我们的方法与其他人的方法进行一个对比 (在这里比较的是 方法中的结构 而不是整个网络)

本文设计了一种结合了序列和语法信息的模型，能够提取更加充分的信息

采用一种更加有效的信息融合方式

本文采用迭代思想设计实验，有力说明了该方法的有效性(前面也得描述一下)

在这里
可以在结构
后面加上这
样做的优点
好处 比以
前方法的优
点

法与语义知识, 获取句子中两种结构关系的情感特征信息。

2) 本文设计了一种融合两种结构关系情感特征信息的结构信息融合网络。它可以充分利用这两种结构的情感特征信息, 结合这两种情感特征信息的优势, 更好地提取与方面词有关的情感特征信息。

本文设计了一个整合多层的结构融合情感特征信息的网络。它可以结合下层简单情感特征信息和高层抽象情感特征信息, 同时处理句子的简单关系和复杂关系, 提高模型的情感分类效果。

1 理论与方法

在本文模型中, 假定输入句子序列 $s = \{w_1, w_2, \dots, w_n\}$, 方面词输入序列 $a = \{a_1, a_2, \dots, a_m\}$, 其中 a 是 s 的子序列, n 和 m 分别是句子和方面词的长度。方面级情感分类的主要目的是预测句子 s 在给定方面词 a 下的情感极性。图 2 中展示了本文模型的总体框架和数据流。

1.1 预处理模块

该模块包括数据预处理和结构信息预处理两部分。数据预处理用于从句子中提取位置信息、语义信息和语法结构信息。结构信息预处理用于初始化两种结构网络的输入信息。

1.1.1 数据预处理

首先, 使用上下文编码器对句子 s 进行编码, 得到句子的上下文词嵌入向量 $e = \{e_1, e_2, \dots, e_n\}$ 。在输入的句子 s 中, 根据句子中每个单词与方面词之间的相对距离, 可以得到句子中每个单词的相对位置信息 $p_i (i \in n)$ 。位于方面词左边的位置信息设为负数, 位于方面词右边的位置信息设为正数。位置信息可以为模型提供句子的序列结构信息。 p_i 的具体计算方法如公式 (1) 所示。

$$p_i = \begin{cases} i - j_1, i < j_1 \\ 0, i \in [j_1, j_m] \\ i - j_m, i > j_m \end{cases} \quad (1)$$

其中, j_i 表示方面词在句子中的位置, 方面词可以包含多个单词。对相对位置信息 p_i 进行编码, 得到位置信息向量 $p = \{p_1, p_2, \dots, p_n\}$, 维度为 d_p 。

为了获取更多的语法和语义知识, 使用句法解析器对句子 s 进行解析, 获取句子的词性标注信息 POS 和语法依赖树关系图。将 POS 词性标注进行编码向量化, 从而获得句子的词性标注信息向量 $T = \{t_1, t_2, \dots, t_n\}$, 维度为 d_t 。从解析器获得的语法依赖树, 包含着句子的语法结构关系, 可以为模型提供语法结构信息。通过构造依赖关系词典, 对依赖关系进行编码, 得到句子的依赖关系图矩阵 D 如公式 (2) 所示。

$$D = \begin{bmatrix} dep_{11} & \cdots & dep_{1n} \\ \vdots & \ddots & \vdots \\ dep_{n1} & \cdots & dep_{nn} \end{bmatrix}, \quad dep_{ij} \in D \quad (2)$$

其中, dep_{ij} 是句子中单词 i 和单词 j 的语法依赖关系的数字化表达。

1.1.2 结构信息预处理

本文模型中, 序列结构网络和语法结构网络的输入数据是不同的。序列结构网络的输入数据包

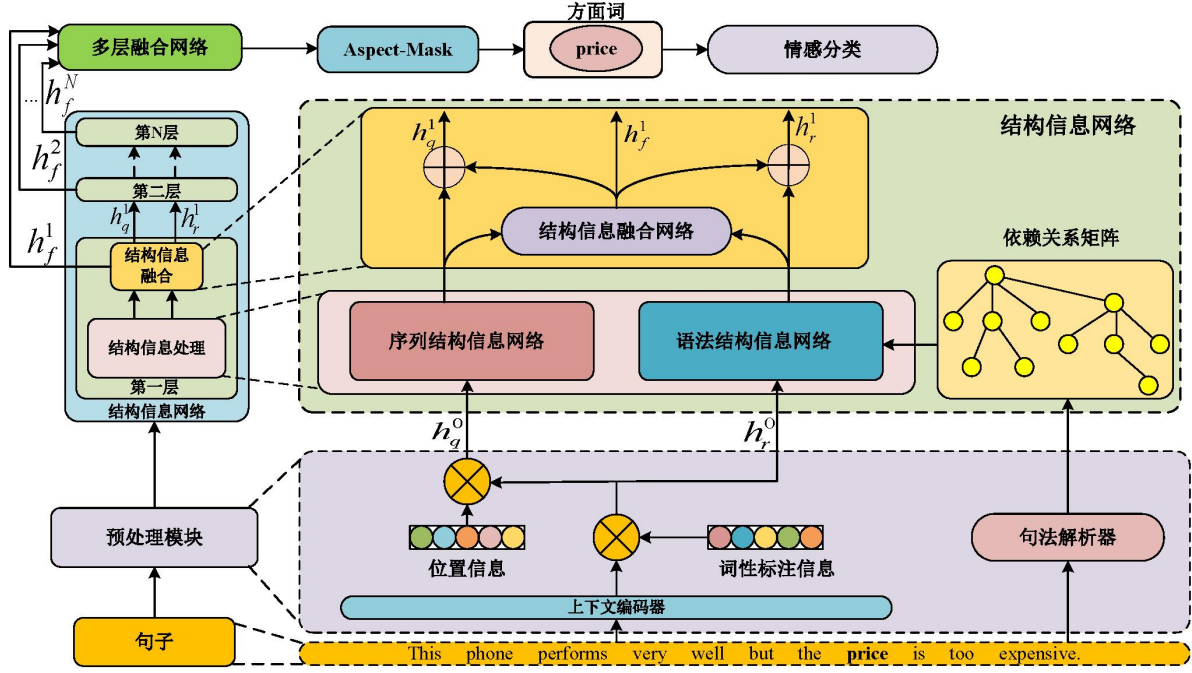


图 2 SYFN 模型的总体框架

Fig.2 Overview structure of SYFN model

括句子单词向量 \mathbf{e}_i 、POS 词性标记信息 \mathbf{t}_i 和单词的相对位置信息 \mathbf{p}_i 。序列结构网络的初始输入数据 \mathbf{h}_q^0 可以通过组合三种信息得到，如公式(3)所示。

$$\mathbf{h}_q^0 = \text{concat}(\mathbf{e}_i; \mathbf{t}_i; \mathbf{p}_i) \quad (3)$$

语法结构网络的输入数据可分为两部分。一部分是输入数据 \mathbf{h}_r^0 ，其中包含句子单词向量 \mathbf{e}_i 和 POS 词性标记信息 \mathbf{t}_i ，具体如公式(4)所示。另一部分是语法结构信息的依赖关系图矩阵 \mathbf{D} 。

$$\mathbf{h}_r^0 = \text{concat}(\mathbf{e}_i; \mathbf{t}_i) \quad (4)$$

1.2 结构信息网络

该模块是一个多层网络结构，每一层网络主要由序列结构信息网络、语法结构信息网络和结构信息融合网络组成。

1.2.1 序列结构信息网络

序列结构信息网络能够处理句子的序列结构关系，通过分析单词之间的序列结构关系，提取句子的序列结构特征信息。本文利用 transformer^[10]中编码器单元的变体结构提取句子的序列结构关系信息。序列结构信息网络主要包括归一化层、多头自注意层和前馈神经网络层，如图 3 所示。

序列结构信息网络的输入数据为 \mathbf{h}_q^L ， L 代表网络层数， q 表示序列结构，当 $L=0$ 时， \mathbf{h}_q^0 为第一层网络中序列结构信息网络的输入数据。输入数据经过归一化层进行标准化处理，得到标准化的数据 $\mathbf{h}_{q,norm}^L = \text{norm}(\mathbf{h}_q^L)$ 。

多头自注意力 (multi-head self-attention, MHSA) 机制使模型能够关注句子中不同维度的特征信息，使模型能够提取到句子更丰富的特征信息。使用多头注意力机制提取句子的内部关系，能够更好地保留原句子的序列结构信息和语义信息，并且不引入外部冗余信息。经过注意力机制可以得到句子的注意力分数 score_q^L ，

$$\text{score}_q^L = \text{MHSA}(\mathbf{h}_{q,norm}^L) \quad (5)$$

多头自注意力机制为了实现并行计算，会丢失数据的位置信息，因此为了保留数据的序列结构

信息, 需要为数据补充位置信息。没有经过多头注意力机制的输入数据中保留着句子的位置信息, 结合注意力分数 $score_q^L$ 和输入数据 h_q^L , 便可得到包含位置信息和注意力分数的数据 $h_{q,self}^L$, 如公式(6)所示。

$$h_{q,self}^L = score_q^L + h_q^L \quad (6)$$

最后将数据输入到前馈神经网络层中, 可以得到序列结构网络输出 $h_{q,o}^L$,

$$h_{q,o}^L = W_2 relu(W_1 norm(h_{q,self}^L) + b_1) + b_2 \quad (7)$$

其中, W_1 , W_2 , b_1 , b_2 为可学习参数, $norm$ 为标准化处理, $relu$ 为激活函数。

1.2.2 语法结构信息网络

该网络用于提取句子的语法结构信息。结合语法依赖树信息有助于模型从语法关系层面提取句子的情感特征信息, 提高方面级情感分类的效果。在该网络中, 我们设计了一个基于图注意机制的语法结构模型来处理句子的语法结构信息, 这个模型能够获得更多的语法关系信息。

该网络的输入数据分为两部分。一部分是句子序列的特征向量 h_r^L , 其中, L 代表网络层数, r 表示语法结构。当 $L=0$ 时, 为第一层网络中语法结构信息网络的输入数据。另一部分是与输入句子相对应的依赖关系图矩阵 D , 它可以将语法结构关系引入到模型中。该网络将依赖关系图与图注意机制相结合, 获得了比邻接矩阵更多的信息, 有利于学习到更多的语法信息知识。首先, 对输入数据进行标准化处理 $h_{syn}^L = norm(h_r^L)$, 然后结合图注意机制获取句子的语法依赖关系信息, 句子中第 i 个单词与句子中所有单词的图注意力 $h_{rel,i}^L$ 的计算公式如下:

$$h_{rel,i}^L = ||_{m=1}^M \sum_{j \in (1,n)} \beta_{ij}^{Lm} (W_m^L h_{syn,j}^L + W_d^L dep_{ij}) \quad (8)$$

$$\beta_{ij}^{Lm} = \frac{\exp(\alpha_{ij}^{Lm})}{\sum_{j=1}^n \exp(\alpha_{ij}^{Lm})} \quad (9)$$

$$\alpha_{ij}^{Lm} = \sigma(W_i h_{syn,i}^L + W_r dep_{ij} + b_i) \quad (10)$$

其中, $||$ 表示向量连接, σ 为激活函数, j 表示句子中的第 j 个词, W_m^L 、 W_d^L 、 W_i 、 W_r 和 b_i 为可学习参数, M 表示图注意力头的个数。语法结构信息网络的最后输出 $h_{rel,o}^L$ 计算如下:

$$h_{rel,o}^L = concat(h_{rel,1}^L; \dots; h_{rel,n}^L) \quad (11)$$

1.2.3 结构信息融合网络

序列结构信息和语法结构信息的融合可以结合这两种结构关系的优势, 能够提取到融合两种结构关系的情感特征信息, 提高模型的情感分类能力。在这个网络中, 我们使用门控机制来融合这两种结构关系的特征信息, 其计算公式如下:

$$\delta = \sigma(W_q^L h_{q,o}^L + W_r^L h_{rel,o}^L + b^L) \quad (12)$$

$$h_{fus}^L = (1 - \delta) \cdot h_{q,o}^L + \delta \cdot h_{rel,o}^L \quad (13)$$

其中 W_q^L 、 W_r^L 和 b^L 为可学习参数, σ 为激活函数, h_{fus}^L 为两种结构信息的融合特征。将融合特征信息 h_{fus}^L 添加到下一层的输入信息中, 有利于提高下一层网络的情感特征信息的提取效果。下一层中两种结构网络的输入信息如下:

$$h_q^{L+1} = h_{fus}^L + h_{q,o}^L \quad (14)$$

$$h_r^{L+1} = h_{fus}^L + h_{rel,o}^L \quad (15)$$

1.3 多层融合网络

在多层融合网络中,低层的网络可以捕获句子单词之间简单的句法关系信息,高层的网络可以捕获得句子单词之间复杂的句法关系信息。多层融合网络结合了来自高低层的结构融合特征信息,让模型能够同时处理简单和复杂的句子关系,提高了模型的情感分类的效果。

因为越高层次的情感特征信息,就越接近人类的情感表达。结合高层和低层的结构融合特征信息 h_{fus}^l ,应当给予高层的融合特征信息更大的权重,融合高低层特征信息 h_{layer}^l 的公式具体如(16)所示。

$$h_{layer} = \sum_{l=1}^N \frac{l}{N} W_l h_{fus}^l + b_l \quad (16)$$

其中, l 为融合特征信息 h_{fus}^l 对应的层数, N 为网络总层数, W_l 和 b_l 为可学习参数。最后使用 $mask$ 从多层融合特征信息 h_{layer}^l 中获取到方面词位置的特征信息 $h_{asp} = mask(h_{layer}^l)$ 。

1.4 情感分类

将方面词的特征信息 h_{asp} 输入到 $softmax$ 函数中,计算其在不同情感类别中的概率 P ,然后通过最小化带 L_2 正则项系数的交叉熵损失函数来调节模型参数,损失函数如公式(17)所示。

$$loss = - \sum_{i=1}^N \sum_{c \in C} I(y=c) \log(P(y=c)) + \lambda \|\theta\|^2 \quad (17)$$

其中, c 为情感类别, I 是一个指示函数, N 是训练实例的数量, λ 是一个正则化超参数, θ 表示模型中的所有参数集。

2 实验与分析

本部分主要包括实验数据、实现细节、基线模型、实验结果与分析、消融研究和案例分析 5 个部分。

2.1 实验数据

本文在 Rest14 和 Laptop^[11]、Twitter^[12]、MAMS^[13] 4 个公开基准数据集上评估了 SYFN 模型的性能。数据集的情感分类包含积极、中性和消极三种类型。其中,Rest14 数据集和 Laptop 数据集中每个条数据包含一个或多个方面词, Twitter 数据集中每条数据只包含一个方面词, MAMS 数据集中的每条数据都包含至少两个具有不同情感极性的方面词。表 1 展示了四个数据集在训练集和测试集中的情感类别的分布情况。

表 1 实验数据集统计信息

Tab.1 Statistics of experiment datasets

数据集 \ 类别	积极		消极		中性	
	Train	Test	Train	Test	Train	Test
Rest14	2164	727	807	196	637	196
Laptop	976	337	851	128	455	167
Twitter	1507	172	1528	169	3016	336
MAMS	3380	400	2764	329	5042	607

2.2 实现细节

本文使用 pytorch 1.7 在 CentOS 7.4 系统和 Tesla T4 GPU 上进行实验, 模型训练使用 Adam^[14] 优化器。使用 The Biaffine Parser^[15] 解析句子得到 POS 词性标注信息和语法依赖树关系, POS 词性标注、位置关系和语法依赖树的向量化维度设置为 30。模型的词嵌入初始化分别使用 300 维的 GloVe 词嵌入向量^[16]结合 BiLSTM 网络(隐藏层大小为 50)和英文版的 bert-base-uncased^[17]预训练模型。

对于使用 GloVe 词嵌入向量的 SYFN-GloVe 模型, 训练批次 $num_epoch = 60$, 批次大小 $batch_size = 16$ 。输入数据的 $input_dropout = 0.3$, 正则化系数 $\lambda = 0.00001$, 优化器的学习率 $learning_rate = 0.001$ 。对于 Rest14、Laptop、Twitter 和 MAMS 数据集, 模型网络层数和注意力头数分别设置为 (3, 8), (2, 8), (3, 5), (4, 10)。

对于使用 bert 预训练模型的 SYFN-BERT 模型, 训练批次 $num_epoch = 30$, 批次大小 $batch_size = 32$ 。bert 模型的输出维度设为 100。输入数据的 $input_dropout = 0.1$, 正则化系数 $\lambda = 0.00001$ 。优化器的学习率 $learning_rate = 0.00001$, bert 模型的学习率 $bert_lr = 0.00002$ 。对于 Rest14、Laptop、Twitter 和 MAMS 数据集, 网络层数和注意力头数分别设置为 (2, 5), (2, 4), (3, 4), (2, 4)。

本文使用准确率 (Accuracy, Acc) 和宏平均 (Macro-F₁, F₁) 两个指标来评估模型的性能, 独立重复每个实验 10 次, 并计算平均值。

2.3 基线模型

本文将 SYFN 模型与具有代表性的基线模型进行比较。基线模型如下:

- 1) ATAE-LSTM^[18](2016): 将方面词嵌入和注意力机制引入方面级情感分类中。
- 2) IAN^[6](2017): 使用两个 LSTM 模型和注意机制交互学习方面词和上下文表示。
- 3) MGAN^[19](2018): 利用一种 BiLSTM 来捕获上下文信息, 并利用一种多粒度的注意力机制来捕获方面词和上下文之间的关系。
- 4) AEN^[7](2019): 使用多头自注意力的编码器分别对方面词和上下文进行建模, 之后再次使用多头注意力机制将两者结合。
- 5) CDT^[20](2019): 使用单层图卷积神经网络结合语法依赖树关系学习方面词的情感表示。
- 6) ASGCN^[8](2019): 使用具有注意力机制的图卷积神经网络结合语法依赖树学习方面词与上下文的关系。
- 7) kumaGCN^[21](2020): 使用一个潜在的图结构来补充句法特征。
- 8) RGAT^[9](2020): 对原始的句法依赖树进行重构和剪枝, 生成面向方面词的依赖树, 构造各个单词与方面词的关系, 并使用图注意力网络 (GAT) 编码新的语法依赖树学习方面词的表示。
- 9) BERT-PT^[22](2019): 在预训练的 BERT 模型上使用后训练的方法, 以提高阅读理解和目标方面情绪分类的表现。
- 10) BERT-SPC^[23](2019): 将方面词和句子以 “[CLS]” + 句子 + “[SEP]” + 方面词 + “[SEP]” 的格式输入一个预先训练过的 BERT 模型进行编码, 然后使用池化层进行分类。
- 11) DualGCN^[24](2021): 使用两个图卷积网络分别考虑句子的语法和语义信息, 并使用正则项对模型进行约束。

2.4 实验结果及分析

为了更好地展示模型的效果, 本文使用了 GloVe 和 bert 两种类型的上下文编码器。表 2 中展示了 SYFN 模型和基线模型的实验结果对比, 名字中包含 “syn” 的模型表示模型结合了语法知识。从表中可以看出, 不管是在 GloVe 模型还是 bert 模型中, 我们模型的表现基本上优于所有的基线模型。

表 2 在 4 个数据集上 SYFN 模型与基线模型的性能比较 (%)

Tab.2 Performance comparison of SYFN model with baseline models on four datasets (%)

编码器	模型	Rest14		LapTop		Twitter		MAMS	
		Acc	F ₁	Acc	F ₁	Acc	F ₁	Acc	F ₁
GloVe	ATAE-LSTM	77.20	-	68.70	-	-	-	-	-
	IAN	78.60	-	72.10	-	-	-	-	-
	MGAN	81.25	71.94	75.39	72.47	72.54	70.81	-	-
	AEN	80.98	72.14	73.51	69.04	72.83	69.81	-	-
	CDT(syn)	82.30	74.02	77.19	72.99	74.66	73.66	80.70	79.79
	ASGCN(syn)	80.77	72.02	75.55	71.05	72.15	70.4	-	-
	RGAT(syn)	83.30	76.08	77.42	73.76	75.57	73.82	-	-
	kumaGCN(syn)	81.43	73.64	76.12	72.42	72.45	70.77	-	-
	DualGCN(syn)	83.02	75.15	77.53	73.78	74.00	71.68	-	-
Bert	AEN-BERT	83.12	73.76	79.93	76.31	74.71	73.13	-	-
	BERT-PT	84.95	76.96	78.07	75.08	-	-	-	-
	BERT-SPC	84.46	76.98	78.99	75.03	73.55	72.14	82.82	81.9
	RGAT-BERT(syn)	86.60	81.35	78.21	74.07	76.15	74.88	-	-
	DualGCN-BERT(syn)	86.06	78.53	80.22	77.15	76.22	75.09	-	-
ours	SYFN-Glove	83.80	76.61	77.85	74.32	75.99	74.06	82.37	81.52
	SYFN-BERT	86.86	81.56	80.34	76.41	76.46	75.97	84.42	83.87

具体来说,在同一种编码器的模型中,具有语法知识的模型在四个数据集上的表现明显优于不具有语法知识的模型。这说明在方面级情感分类中,结合语法知识能够有效提高情感分类的效果。而 SYFN 模型比包含语法知识的模型效果更好,情感分类的性能超过了所有只利用其中一种结构知识的模型,这说明 SYFN 模型能够结合两种句子结构知识的优势,证明结构知识融合能够提高情感分类的效果。在 MAMS 数据集的实验结果上,尽管 MAMS 数据集中每条数据都包含了至少两个方面项和不同的情感极性,但 SYFN 模型仍然取得了最优的效果。这说明 SYFN 模型能够处理复杂的句子关系,能够在复杂的句子关系中避免噪声的干扰并准确地提取方面词的情感特征信息,充分展现 SYFN 模型结合两种结构知识信息的优越性。

最后,SYFN 模型不管是在简单的句子关系还是复杂的句子关系中都取得了卓越的效果,表明 SYFN 模型能够处理各种各样的句子关系,体现了 SYFN 模型结合两种结构知识的优点。

2.5 消融研究

为了进一步研究模型 SYFN 网络的每个组成部分对性能的影响,在数据集 Rest14 和 Laptop 上对使用 GloVe 词嵌入向量的 SYFN 网络模型进行了消融实验研究。本实验设计了四种消融模型来与

表 3 消融模型描述信息

Tab.3 Ablation Models Descriptions

消融模型	详细描述
SYFN w/o Syn	SYFN 模型去掉语法结构模块,只保留序列结构模块的网络模型。
SYFN w/o Seq	SYFN 模型去掉序列结构模块,只保留语法结构模块的网络模型。
SYFN fus in last-layer	SYFN 模型中序列结构网络和语法结构网络只在最后一层网络进行结构信息特征融合并将最后一层的融合特征信息用于情感分类的网络模型。
SYFN clas in last-fus	SYFN 模型中序列结构网络和语法结构网络在每一层都进行结构特征信息融合,但是只将最后一层融合特征信息用于情感分类的网络模型。

SYFN 模型进行比较。具体描述如表 3 所示。

SYFN clas in last-fus 模型的分类效果比 SYFN fus in last-layer 模型中有所提升,表明在每一层的网络中融合两种结构特征信息,能获得到更多的情感特征信息,将融合信息传递到后面的网络中,提升模型的分类效果。此外,本文的 SYFN 模型分类效果优于 SYFN clas in last-fus 模型,这表明结合每一层的结构融合信息能够提升模型的性能,SYFN 模型中低层的网络获取句子简单的结构特征信息,高层的网络获取句子复杂的结构特征信息。结合高低层融合信息使模型能够兼顾简单和复杂的情感特征信息,提高模型的情感表达能力。

总的来说,消融实验的结果表明 SYFN 模型中两种结构信息网络、结构信息融合网络和多层融合网络在方面级情感分类中有着巨大的作用,结合这三种网络的优势,让模型能够更全面地获取到方面词的情感特征信息。

表 4 消融研究的实验结果(%)

Tab.4 Experiment Results of Ablation Study(%)

模型	Rest14		LapTop	
	Acc	F ₁	Acc	F ₁
SYFN	83.80	76.61	77.85	74.32
SYFN w/o Syn	81.74	73.71	75.16	70.42
SYFN w/o Seq	81.76	73.36	75.32	70.76
SYFN fus in last-layer	82.04	74.84	75.95	71.83
SYFN clas in last-fus	82.67	75.35	76.27	72.49

2.6 案例分析

为了更好地研究 SYFN 模型在结合序列和语法结构信息方面的优势,将 SYFN 模型与没有句法结构信息的 AEN 模型和具有句法结构信息的 RGAT 模型在特定例子中进行比较。表 5 中展示了三个模型在案例中的预测结果,表中用粗体突出显示了方面字。

在第一个例子中,AEN 模型因为无法理解连接词“but”而错误地关注了单词“Great”,误判了方面词“service”的情感极性。与之相比,RGAT 模型和 SYFN 模型都能够对方面词进行正确的分类,这说明了结合语法结构信息的模型能够有效地处理复杂的句子关系,能够从句子中准确提取到方面词的情感特征。

在第二个例子中,连词“but”和否定词“not”一起出现,这种情况让句子中的句法关系更加复杂。面临这种情况,不仅仅是 AEN 模型,包含语法结构信息的 RGAT 模型也对方面词“price”进行了错误的分类。这意味着只包含语法结构知识的情感分类模型也不能处理这种复杂的关系。SYFN 模型能够准确地对方面词的情感极性进行分类,这表明本文中结合两种结构信息的网络和多层融合策略可以有效地处理各种复杂的句子关系。

表 5 AEN、RGAT 和 SYFN 三个模型在案例上的预测结果

Tab.5 Preprediction results of the three models AEN, RGAT and SYFN

句子	方面词: 极性	AEN	RGAT	SYFN
Great food but the service was dreadful!	food : pos	pos √	pos √	pos √
	service : neg	pos ×	neg √	neg √
The food was good, but not at all worth the price .	food : pos	neg ×	pos √	pos √
	price : neg	pos ×	pos ×	neg √
Slow service and the environment just so-so but the food is not bad.	service : neg	neg √	neg	neg √
	environment : neu	neg ×	neg ×	neu √
	food : pos	neg ×	neg ×	pos √

第三个例子中存在着三个方面词,而且每个方面词的情感极性都各不相同,这种情况下预测单个方面词的情感极性时,它们会互相干扰。同时句子中还存在着连词“but”和否定词“not”,这大大增加了情感分类的难度。AEN 和 RGAT 模型表现不佳,说明只凭借一种句子结构关系无法处理这种复杂的情况。SYFN 模型仍然表现得很好,它可以准确地预测各个方面词的情感极性。这表明 SYFN 模型中将两种结构关系和多层信息融合相结合的方法可以有效地防止噪声的干扰,并能够从复杂的句子关系提取出方面词对应的情感特征信息。

3 结语

本文中提出的 SYFN 方面级情感分类模型能够有效利用句子的序列结构信息和句法结构信息,从两种结构关系中准确提取出方面词的情感特征信息。SYFN 模型能够同时处理句子的序列结构关系和句法结构关系,然后生成两种结构关系的融合特征信息,使模型能够充分地从句子的结构关系中提取到方面词的情感特征信息。结合多层网络的融合信息,使模型能够同时处理句子简单和复杂的句子关系。这种多层网络融合的机制能够提升模型的信息获取能力,进一步加强了模型的情感表达能力。实验结果表明,本文的 SYFN 模型能够有效应对各种复杂的句子关系和噪声干扰,在方面级情感分类上取得了优异的成果。

参考文献

- [1] Li Q, Chen H, Ren Z, et al. EmpDG: Multi-resolution Interactive Empathetic Dialogue Generation[C]// Proceedings of the 28th International Conference on Computational Linguistics. 2020.
- [2] Lan T, Mao X L, Wei W, et al. PONE: A Novel Automatic Evaluation Metric for Open-Domain Generative Dialogue Systems:, 10.1145/3423168[P]. 2020.
- [3] Jin Y C, Zhao K, Joty S, et al. ANR: Aspect-based Neural Recommender[C]// the 27th ACM International Conference. ACM, 2018.
- [4] Zhang K, Qian H, Liu Q, et al. SIFN: A Sentiment-aware Interactive Fusion Network for Review-based Item Recommendation[J]. 2021.
- [5] Tang D, Qin B, Feng X, et al. Effective LSTMs for Target-Dependent Sentiment Classification[J]. Computer Science, 2015.
- [6] Ma D, Li S, Zhang X, et al. Interactive Attention Networks for Aspect-Level Sentiment Classification[J]. 2017.
- [7] Song Y, Wang J, Tao J, et al. Attentional Encoder Network for Targeted Sentiment Classification[J]. 2019.
- [8] Zhang C, Li Q, Song D. Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks:, 10.18653/v1/D19-1464[P]. 2019.
- [9] Wang K, Shen W, Yang Y, et al. Relational Graph Attention Network for Aspect-based Sentiment Analysis[J]. 2020.

- [10] Vaswani A , Shazeer N , Parmar N , et al. Attention Is All You Need[J]. arXiv, 2017.
- [11] Pontiki M , Galanis D , Pavlopoulos J , et al. SemEval-2014 Task 4: Aspect Based Sentiment Analysis[J]. Proceedings of International Workshop on Semantic Evaluation at, 2014.
- [12] Li D , Wei F , Tan C , et al. Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification[C]// Meeting of the Association for Computational Linguistics. 2014.
- [13] Jiang Q, Chen L, Xu R, et al. A challenge dataset and effective models for aspect-based sentiment analysis[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). 2019: 6280-6285.
- [14] Kingma D , Ba J . Adam: A Method for Stochastic Optimization[J]. Computer Science, 2014.
- [15] Dozat T , Manning C D . Deep Biaffine Attention for Neural Dependency Parsing[J]. 2016.
- [16] Pennington J , Socher R , Manning C . Glove: Global Vectors for Word Representation[C]// Conference on Empirical Methods in Natural Language Processing. 2014.
- [17] Devlin J , Chang M W , Lee K , et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. 2018.
- [18] 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.
- [19] Fan F , Feng Y , Zhao D . Multi-grained Attention Network for Aspect-Level Sentiment Classification[J]. 2018.
- [20] Sun K, Zhang R, Mensah S, 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). 2019: 5679-5688.
- [21] Chen C, Teng Z, Zhang Y. Inducing target-specific latent structures for aspect sentiment classification[C]//Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP). 2020: 5596-5607.
- [22] Xu H , Liu B , Shu L , et al. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis:, 10.48550/arXiv.1904.02232[P]. 2019.
- [23] Zhang M , Qian T . Convolution over Hierarchical Syntactic and Lexical Graphs for Aspect Level Sentiment Analysis[C]// Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2020.
- [24] Li R , Chen H , Feng F , et al. Dual Graph Convolutional Networks for Aspect-based Sentiment Analysis[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021.