

DOI:10.13624/j.cnki.issn.1001-7445.2023.1156

结合句子序列与语法关系的方面级情感分类方法

凌键军^{1,2}, 李志鹏^{1,2}, 陈丹阳^{1,2*}, 王翔宇³, 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. 广西高校并行分布与智能计算重点实验室, 广西 南宁 530004;
3. 腾讯云计算(北京)有限责任公司, 广东 深圳 518000)

摘要:为了研究句子结构关系对方面级情感分类模型性能的影响,针对当前方面级情感分类方法中过于重视句子中语法关系的作用,而忽视了句子的序列结构关系的问题,提出一种结合句子序列与语法关系的信息融合网络(sequence-syntax information fusion network, SYFN)模型。SYFN 模型分别处理句子的序列与语法关系信息,并在融合 2 种结构关系信息的基础上进一步结合网络的高低层融合信息。这些改进措施使模型能够结合高低层网络的关系融合信息去处理各种复杂的句子关系,能够充分利用结构关系信息进行情感分类。实验结果表明,SYFN 模型与基线模型相比性能有较明显的提升。

关键词:深度学习;自然语言处理;方面级情感分类;语法依赖树;信息融合

中图分类号:TP391.1 **文献标识码:**A **文章编号:**1001-7445(2023)05-1156-11

An aspect-level sentiment classification method combining sequence and syntax

LING Jianjun^{1,2}, LI Zhipeng^{1,2}, CHEN Danyang^{1,2*},
WANG Xiangyu³, ZHONG Cheng^{1,2}

- (1. School of Computer, Electronics and Information, Guangxi University, Nanning 530004, China;
2. Key Laboratory of Parallel, Distributed and Intelligent Computing of
Guangxi Universities and Colleges, Nanning 530004, China;
3. Tencent Cloud Computing (Beijing) Co., Ltd., Shenzhen 518000, China)

Abstract: In order to study the influence of sentence structure relationship on the performance of aspect level sentiment classification model, a new information fusion network (sequence-syntax information fusion network, SYFN) model combining sequence and syntax relationships is proposed to solve the problem that the current aspect level sentiment classification methods pay too much attention to the role of grammatical relationship in sentences, while ignoring the sequence structure relationship of sentences. In the SYFN model, the sequence and grammatical relationship information of sentences are processed

收稿日期:2023-02-25; **修订日期:**2023-04-04

基金资助:国家自然科学基金项目(61962004)

通信作者:陈丹阳(1988—),女,黑龙江齐齐哈尔人,讲师,博士;E-mail:chendanyang@gxu.edu.cn。

引文格式:凌键军,李志鹏,陈丹阳,等. 结合句子序列与语法关系的方面级情感分类方法[J]. 广西大学学报(自然科学版),2023,48(5):1156-1166.

separately, and based on the fusion of the two structural relationship information, the high and low layer fusion information of the network is further combined. These improvements enable the model to combine the relationship fusion information of high and low layer networks to deal with various complex sentence relationships, and fully utilize structural relationship information for sentiment classification. The experimental results show that the performance of the SYFN model is significantly improved compared with the baseline models.

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

0 引言

方面级情感分类^[1]是一种细粒度的文本情感分类任务,可以识别出句子中方面词所表达的情感极性,方面词的情感极性一般分为积极、消极和中性3种类型。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,在句子“This phone performs very well but the price is too expensive.”中包含了2个方面词“performs”和“price”,它们在句子中的情感极性分别是“积极”和“消极”,方面级情感分类任务的目标是分别判断句子中2个方面词的情感极性。

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,取得了一定的成果。Tang等^[2]使用长短记忆网络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^[3]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^[4]使用注意力编码器网络(attentional encoder network, AEN)从单词嵌入中挖掘丰富的语义信息。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^[5]提出了一种基于特定方面的图卷积网络模型,模型使用图卷积网络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,图卷积网络^[6]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^[7]提出了一种面向方面词的语法依赖树的关系图注意力网络(relational graph attention networks, RGAT)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,面向方面的依赖关系的构建方式过于主观,对语法依赖树的修剪和重构会导致句子中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,本文中提出了一种结合句子序列与语法关系的融合网络(sequence-syntax information fusion network, SYFN)模型。SYFN模型结合语法依赖关系、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模型主要由结构信息网络和多层融合网络组成。结构信息网络包含多层网络结构,每一层网络中包含3个主要部分:序列结构网络、语法结构网络和结构信息融合网络。其中,基于多头注意力机制的序列结构网络能够快速提取句子序列的前后语义信息。结合关系图注意力方法的语法结构网络能够获取到更多的与方面词相关的语法结构信息。结构信息融合网络能够将2种结构关系信息进行融合,有效提取句子中与方面词最相关的部分情感特征信息。多层融合网络使用权重机制有选择地结合高低层网络的结构关系融合信息,使模型能够同时关注低层网络单词之

间的简单关系和高层网络中短语之间的复杂关系,有效地结合简单和复杂的关系信息,提升模型分类性能。另外,采用一种迭代式的消融实验方法,对融合模型进行不断改进,最终得到最优结果。

1 理论与方法

在本文模型中,假定输入句子序列 $s = \{w_1, w_2, \dots, w_n\}$, 方面词序列 $a = \{a_1, a_2, \dots, a_m\}$, 其中方面词 a 是句子 s 的单词子序列, n 和 m 分别是文本句子和方面词的单词个数。方面级情感分类任务的主要目的是预测句子 s 在给定方面词 a 的情感极性。

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 2 个部分。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 2 种结构信息网络的输入信息。

1.1.1 数据预处理

首先,使用上下文编码器对句子 s 中的单词 w 进行编码,得到句子的上下文词嵌入向量集合 $e = \{e_1, e_2, \dots, e_n\}$, 其中 $e_i (i \in n)$ 是单词 w_i 的向量化编码。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,可以得到句子中第 i 个单词的相对位置信息 $r_i (i \in n)$ 。位置信息可以为模型提供句子的序列结构信息,计算单词的相对位置信息时,将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。句子中第 i 个单词的相对位置信息 r_i 的具体计算方法如式(1)所示。

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

式中 j 表示方面词在句子中的位置,方面词可以包含多个单词。对整个句子中所有单词的相对距离信息集合 $r = \{r_1, r_2, \dots, r_n\}$ 进行向量化编码,得到句子中各个单词的相对位置信息的向量集合 $p = \{p_1, p_2, \dots, p_n\}$, 其中 p_i 表示句子中第 i 个单词的相对位置信息 r_i 的向量化编码,其向量维度为 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和语法依赖树关系图。将词性标注信息进行编码向量化,从而获得句子的词性标注信息向量集合 $t = \{t_1, t_2, \dots, t_n\}$, 维度为 d_t 。从句法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。通过语法依赖树构造单词之间的依赖关系词典,对依赖关系进行编码,得到句子中各个单词之间的依赖关系矩阵 D 如式(2)所示。

$$D = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \cdots & \vdots \\ d_{n1} & \cdots & d_{nn} \end{pmatrix}, \quad (2)$$

式中 d_{ij} 是句子中第 i 个单词和第 j 个单词之间依赖关系的向量编码。

1.1.2 结构信息预处理

在本文模型中,序列结构信息网络和语法结构信息网络的输入数据是不同的。

序列结构信息网络的输入数据包括句子单词向量 e 、词性标注信息 t 和单词的相对位置信息 p 。序列结构信息网络的初始输入数据 h_q^0 可以通过组合 3 种信息得到,如式(3)所示。

$$h_q^0 = \text{concat}(e; t; p), \quad (3)$$

式中: q 表示序列结构网络; 0 表示第一层网络; concat 表示对向量进行拼接操作。

语法结构信息网络的输入数据可分为 2 个部分。一部分输入信息是输入数据 h_r^0 , 其中包含句子单词向量 e 和词性标注信息 t , 具体如式(4)所示。

$$h_r^0 = \text{concat}(e; t), \quad (4)$$

式中 r 表示语法结构网络。

另一部分的输入信息是语法结构信息的依赖关系集合 D 。

1.2 结构信息网络

结构信息网络是一个多层结构的网络,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。该网络主要利用多头注意力机制(multi-head self-attention, MHSA)提取句子的序列结构关系信息。

假设序列结构信息网络的输入数据为 h_q^L , L 代表网络总层数,当 $L=0$ 时, h_q^0 为第一层序列结构信息网络的输入数据。序列结构网络的输入数据经过归一化层进行标准化处理,得到标准化的数据信息 $h_{q,norm}^L = \text{norm}(h_q^L)$, 其中 norm 表示归一化处理。

多头自注意力机制使模型能够从不同特征维度关注句子单词之间的语义信息,使模型能够提取到句子更丰富的特征信息。序列结构信息网络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关系信息和语义信息,并且不引入外部冗余信息。数据 $h_{q,norm}^L$ 经过注意力机制可以得到句子的注意力分数 s_q^L , 如式(5)所示。

$$s_q^L = \text{MHSA}(h_{q,norm}^L) \quad (5)$$

然而,多头自注意力机制为了实现并行计算,将丢失数据单词之间的位置信息,因此序列结构信息网络中为了保留数据的序列结构信息,需要为数据补充单词之间的位置信息。网络中没有经过多头注意力机制的数据 h_q^L 仍然保留着句子单词的位置关系,通过结合注意力分数 s_q^L 和数据 h_q^L , 网络模型能够得到同时包含位置信息和注意力分数的数据 $h_{q,s}^L$, 计算公式为

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

最后,将数据 $h_{q,s}^L$ 输入到前馈神经网络层中,得到式(7)中的序列结构网络输出 $h_{q,o}^L$ 。

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

式中: W_1 、 W_2 、 b_1 、 b_2 分别为可学习参数; relu 为激活函数; o 表示信息输出。

1.2.2 语法结构信息网络

语法结构信息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,从而能够获得更多的语法关系信息。

语法结构信息网络的输入数据分为 2 个部分:一部分是句子序列的特征向量 h_r^L , 当 $L=0$ 时, h_r^0 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矩阵 D , 它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,对语法结构网络的输入数据 h_r^L 进行标准化处理 $h_{r,s}^L = \text{norm}(h_r^L)$, 其中下标 s 表示语法结构信息网络;然后,使用图注意机制结合句子的语法依赖关系矩阵 D 获取句子单词之间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 $h_{r,i}^L$ 的计算如式(8)–(10)所示。

$$h_{r,i}^L = \parallel_{m=1}^M \sum_{j \in (1,n)} \beta_{ij}^{Lm} (W_1 h_{r,s,j}^L + W_2 d_{ij}) \quad (8)$$

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

$$\alpha_{ij}^{Lm} = \sigma(W_3 h_{r,s,i}^L + W_4 d_{ij} + b_1) \quad (10)$$

式中: \parallel 表示向量连接; σ 为激活函数; j 表示句子中的第 j 个单词; W_1 、 W_2 、 W_3 、 W_4 、 b_1 分别为可学习参数; M 表示图注意力头的个数。语法结构信息网络的最后输出 $\mathbf{h}_{r,o}^L$ 如式(11)所示。

$$\mathbf{h}_{r,o}^L = \text{concat}(\mathbf{h}_{r,1}^L; \mathbf{h}_{r,2}^L; \cdots; \mathbf{h}_{r,n}^L) \quad (11)$$

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 2 种结构关系的优势,从句子中提取到融合 2 种结构关系的情感特征信息,提高模型的情感表达能力。在结构信息融合网络中,使用门控机制融合 2 种结构关系的特征信息,其计算公式分别为

$$\delta = \sigma(W_5 \mathbf{h}_{q,o}^L + W_6 \mathbf{h}_{r,o}^L + b_2), \quad (12)$$

$$\mathbf{h}_f^L = (1 - \delta) \mathbf{h}_{q,o}^L + \delta \mathbf{h}_{r,o}^L, \quad (13)$$

式中: W_5 、 W_6 、 b_2 分别为可学习参数; \mathbf{h}_f^L 为 2 种结构信息的融合特征;下标 f 表示信息融合。将融合特征信息 \mathbf{h}_f^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 2 种结构网络的输入信息为

$$\mathbf{h}_q^{L+1} = \mathbf{h}_f^L + \mathbf{h}_{q,o}^L, \quad (14)$$

$$\mathbf{h}_r^{L+1} = \mathbf{h}_f^L + \mathbf{h}_{r,o}^L, \quad (15)$$

式中 \mathbf{h}_q^{L+1} 、 \mathbf{h}_r^{L+1} 分别为下一层序列结构信息网络和语法结构信息网络的输入。

1.3 多层融合网络

在多层融合网络中,低层的网络可以关注句子单词之间简单的句法关系,高层的网络可以关注句子短语之间复杂的句法关系信息。多层融合网络能够结合来自高低层的结构融合特征信息,使模型能够同时处理简单和复杂的句子关系,提高模型的情感分类的效果。

在多层网络中,高层网络的情感特征信息比低层网络的情感特征信息更加接近人类真实的情感表达。结合高层和低层网络中的结构融合特征信息 \mathbf{h}_f^L , 应当给予高层的融合特征信息更大的权重。融合高低层网络的特征信息 \mathbf{h}_f^L 可以得到多层融合特征信息 \mathbf{z}' , 具体计算公式为

$$\mathbf{z}' = \sum_{l \in (1, N)} \frac{l}{N} W_l \mathbf{h}_f^l + b_l, \quad (16)$$

式中: l 为当前网络层数; \mathbf{h}_f^l 为第 l 层的融合特征信息; N 为网络总层数; W_l 和 b_l 为可学习参数。最后使用掩码机制 mask 方法从多层融合特征信息 \mathbf{z}' 中获取到方面词位置的向量特征信息 $\mathbf{h}_a = \text{mask}(\mathbf{z}')$, a 表示方面词。

1.4 损失函数

将方面词的特征信息 \mathbf{h}_a 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,损失函数的计算公式为

$$L = - \sum_{i \in (1, n)} \sum_{c \in C} I(y=c) \log[P(y=c)] + \gamma \|\theta\|^2, \quad (17)$$

式中: c 为情感类别; I 为一个指示函数; γ 为一个正则化超参数; θ 为模型中的所有参数集。

2 实验与分析

2.1 实验数据

本文中在 Rest14、Laptop^[8]、Twitter^[9]、MAMS^[10] 这 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、消极 3 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 2 个具有不同情感极性的方面词。数据集统计信息见表 1。

表1 数据集中不同情感分类的数据数目

Tab. 1 The number of different sentiment classifications in the datasets

| 数据集 | 积极 | | 消极 | | 中性 | |
|---------|-------|-----|-------|-----|-------|-----|
| | 训练集 | 测试集 | 训练集 | 测试集 | 训练集 | 测试集 |
| Rest14 | 2164 | 727 | 807 | 196 | 637 | 196 |
| Laptop | 976 | 337 | 851 | 128 | 455 | 167 |
| Twitter | 1 507 | 172 | 1 528 | 169 | 3 016 | 336 |
| MAMS | 3 380 | 400 | 2 764 | 329 | 5 042 | 607 |

2.2 参数设置

使用 PyTorch 1.8 在 CentOS 7.4 系统、Tesla T4 GPU 上进行实验,模型训练使用 Adam^[11] 优化器。使用句法解析器 The Biaffine Parser^[12] 解析句子得到词性标注信息和语法依赖树关系,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结合双向长短记忆网络的 300 维的 GloVe^[13] 词嵌入向量和英文版的 bert-base-uncased^[14] 预训练模型进行词向量编码。

对于使用 GloVe 词嵌入向量的 SYFN-GloVe 模型,每次训练 60 个批次,每批次包含 16 条数据。输入数据的 dropout 率为 0.3,正则化系数为 0.000 01,优化器的学习率设为 0.001。对于 Rest14、Laptop、Twitter 和 MAMS 这 4 个数据集,训练过程中 SYFN 模型的网络层数和注意力头数分别设置为(4,8)、(4,5)、(2,5)和(2,10)。

对于使用预训练语言模型 BERT 编码的 SYFN-BERT 模型,每次训练 30 个批次,每批次包含 32 条数据。BERT 模型的输出维度设为 100。输入数据的 dropout 率为 0.1,正则化系数为 0.000 01。优化器的学习率设为 0.000 01,BERT 模型的学习率为 0.000 02。对于 Rest14、Laptop、Twitter 和 MAMS 4 个数据集,网络层数和注意力头数分别设置为(3,5)、(2,5)、(2,4)和(2,10)。

使用准确率(accuracy, ACC)和宏平均(macro- F_1 , F_1) 2 个指标评估模型的性能,独立重复每个实验 10 次,并计算平均值。

2.3 基线模型

将 SYFN 模型与方面级情感分类领域的一些具有代表性的基线模型进行比较。这些基线模型的具体描述如下:

- ① IAN^[3]:使用 2 个长短记忆网络模型和注意机制交互学习方面词和上下文表示。
- ② MGAN^[15]:使用双向长短记忆网络获取句子的上下文信息,并结合多粒度的注意力机制来提取方面词和上下文之间的语义关系。
- ③ AEN^[4]:使用基于自注意力的编码网络分别对方面词和上下文的语义信息进行建模。
- ④ CDT^[16]: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。
- ⑤ ASGCN^[5]: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。
- ⑥ RGAT^[6]:对句子的语法依赖树进行重构,生成面向方面词的语法依赖树,并主观构造单词与方面词之间的依赖关系,最后结合图注意力网络学习方面词的句法特征信息。
- ⑦ BERT-PT^[17]:在 BERT 预训练语言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。
- ⑧ BERT-SPC^[18]:将方面词和句子以“[CLS]+句子+[SEP]+方面词+[SEP]”的格式输入 BERT 预训练语言模型进行编码,然后使用池化层进行分类。
- ⑨ DualGCN^[19]:使用 2 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。
- ⑩ BERT4GCN^[20]:利用 BERT 预训练语言模型的中间层知识和位置信息,并使用图卷积网络结

合依赖树语法知识进行情感分类。

⑪ T-GCN-BERT^[21]:使用能够区分依赖树关系和从不同层次学习的感知图卷积网络结合语法知识进行方面级情感分类。

2.4 实验结果及分析

为了更好地展示模型的效果,使用了 GloVe、BERT 这 2 种类型的上下文编码器。在 4 个数据集上 SYFN 模型与基线模型的性能比较见表 2,名字中包含“syn”的模型表示模型结合了语法知识,“—”表示模型没有在改数据集上进行实验;加粗字体表示该数据集上的最优结果。从表 2 可见,不管是在 GloVe 模型还是 BERT 模型中,SYFN 模型的表现基本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

| 编码器 | 模型 | Rest14 | | Laptop | | Twitter | | MAMS | |
|-------|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | ACC | F_1 | ACC | F_1 | ACC | F_1 | ACC | F_1 |
| GloVe | 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 | | |
| | DualGCN(syn) | 83.02 | 75.15 | 77.53 | 73.78 | 74.00 | 71.68 | | |
| | 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 |
| BERT | 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 | | |
| | BERT4GCN(syn) | 84.75 | 77.11 | 77.49 | 73.01 | 74.73 | 73.76 | — | — |
| | T-GCN-BERT(syn) | 85.54 | 78.86 | 80.25 | 76.92 | 76.16 | 74.44 | 81.73 | 81.12 |
| SYFN | 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 4 个数据集上的表现明显优于不具有语法知识的模型,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,说明 SYFN 模型能够结合 2 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 2 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,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 2 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 2 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,本文中将采用迭代的思想逐步验证 SYFN 模型中各个组件的效果和作用,在数据集 Rest14、Laptop 上使用 GloVe 词嵌入向量的 SYFN 网络模型进行了消融实验研究。

本次消融实验中设计了 4 种类型消融模型与 SYFN 模型进行比较,4 种类型消融模型分别为

①SYFN w/o Syn 模型: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;②SYFN w/o Seq 模型: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;③SYFN fus in last-layer 模型: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关系信息融合并将最后一层的融合信息用于情感分类的网络模型;④SYFN clas in last-fus 模型: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。消融研究的实验结果见表3。

表3 消融研究的实验结果

Tab. 3 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 |

从表3可见,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2种结构特征信息,能获取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,本文中的SYFN模型分类效果优于SYFN clas in last-fus 模型,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2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巨大的作用,结合这3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。

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模型与没有语法结构信息的AEN模型和具有语法结构信息的RGAT模型在特定例子中进行比较。AEN、RGAT和SYFN模型在案例上的预测结果见表4,表中用粗体突出显示了方面词。

表4 AEN、RGAT和SYFN模型在案例上的预测结果

Tab. 4 Preprediction results of the AEN, RGAT and SYFN models

| 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✓ |

在第1个句子中,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,说明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 2 个句子中,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,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,表明本文中结合 2 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。

第 3 个句子中存在着 3 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”,大大增加了情感分类的难度。AEN 和 RGAT 模型表现不佳,说明只凭借一种类型的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,可以准确地预测各个方面词的情感极性,表明 SYFN 模型中将 2 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。

2.7 模型参数量

本文中的 SYFN 模型使用了更为灵活的特征融合方法,同时也使用了更多的参数。文中基于 GloVe 编码和基于 BERT 编码的 2 种 SYFN 模型在不同数据集上的参数量见表 5。

表 5 SYFN 模型在不同数据集上的参数量
Tab. 5 Parameters of the SYFN model on different datasets $\times 10^6$

| 模型 | Rest14 | Laptop | Twitter | MAMS |
|------------|--------|--------|---------|--------|
| SYFN-Glove | 5.0 | 3.87 | 9.63 | 7.86 |
| SYFN-BERT | 110.37 | 110.38 | 110.73 | 110.37 |

从表 5 可见,基于 BERT 编码的模型参数量远远大于基于 GloVe 编码的模型参数量,这是因为预训练模型 BERT 中包含着庞大的参数量,导致 SYFN-BERT 模型参数量的大量增加。同时可以看到,在 SYFN-GloVe 模型中,Twitter 和 MAMS 数据集的模型参数量多于 Rest14 和 Laptop 数据集的模型参数量,这是因为在 Twitter 和 MAMS 数据集的文本数据中包含的单词种类个数远多于 Rest14 和 Laptop 数据集,导致在对文本句子数据进行单词向量化时需要更多的参数。

3 结论

本文中提出了一种结合句子序列和语法关系的方面级情感分类(SYFN)模型,它能够有效地利用句子的序列语义信息和语法结构信息,从 2 种信息角度提取到与方面词更加相关的情感特征信息。此外,多层网络融合的机制进一步加强了模型对 2 种结构信息的利用能力,使模型能够同时处理简单和复杂的句子关系,最大限度地获取到与方面词相关的情感信息,提高模型的分类能力。实验结果表明,SYFN 模型能够处理各种复杂的句子关系,在方面级情感分类中拥有优秀的情感分类能力,同时文中还在消融实验中验证了 SYFN 模型各个组件设计思想的合理性。

参考文献:

- [1] 王婷,杨文忠. 文本情感分析方法研究综述[J]. 计算机工程与应用, 2021,57(12):11-24.
- [2] TANG D, QIN B, FENG X, et al. Effective LSTMS for target-dependent sentiment classification[C]//Proceedings of the 26th International Conference on Computational Linguistics, Osaka: ACL, 2016:3298-3307.
- [3] MA D, LI S, ZHANG X, et al. Interactive attention networks for aspect-level sentiment classification[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2017: 4067-4074.

- [4] SONG Y, WANG J, TAO J, et al. Attentional encoder network for targeted sentiment classification[DB/OL]. (2019-04-01)[2023-06-13]. <https://arxiv.org/abs/1902.09314v2>.
- [5] ZHANG C, LI Q, SONG D. 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. Hong Kong: Association for Computational Linguistics, 2019:4567-4577.
- [6] 徐冰冰,岑科廷,黄俊杰,等. 图卷积神经网络综述[J]. 计算机学报, 2020,43(5):755-780.
- [7] WANG K, SHEN W, YANG Y, et al. Relational graph attention network for aspect-based sentiment analysis [C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Hong Kong: Association for Computational Linguistics, 2020:3229-3238.
- [8] 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, COLING 2014. Dublin: The Association for Computer Linguistics, 2014:27-35.
- [9] LI D, WEI F, TAN C, 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. Baltimore: The Association for Computer Linguistics, 2014:49-54.
- [10] 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. Hong Kong: Association for Computational Linguistics, 2019:6280-6285.
- [11] KINGMA D, BA J. Adam: a method for stochastic optimization[DB/OL]. (2014-12-22)[2023-06-13]. <https://arxiv.org/abs/1412.6980>.
- [12] DOZAT T, MANNING C D. Deep biaffine attention for neural dependency parsing[DB/OL]. (2017-03-10)[2023-06-13]. <https://arxiv.org/abs/1611.01734>.
- [13] PENNINGTON J, SOCHER R, MANNING C. Glove: global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha: ACL, 2014:1532-1543.
- [14] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019:4171-4186.
- [15] FAN F, FENG Y, ZHAO D. Multi-grained attention network for aspect-level sentiment classification[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018:3433-3442.
- [16] 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. Hong Kong: Association for Computational Linguistics, 2019:5679-5688.
- [17] XU H, LIU B, SHU L, et al. BERT post-training for review reading comprehension and aspect-based sentiment analysis[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language. Minneapolis: Association for Computational Linguistics, 2019:2324-2335.
- [18] ZHANG M, QIAN T. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Online: Association for Computational Linguistics, 2020:3540-3549.
- [19] 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. Online: Association for Computational Linguistics, 2021:

6319-6329.

- [20] XIAO Z, WU J, CHEN Q, et al. BERT4GCN: Using BERT intermediate layers to augment GCN for aspect-based sentiment classification[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Punta: Association for Computational Linguistics, 2021:9193-9200.
- [21] TIAN Y, CHEN G, SONG Y. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Online: Association for Computational Linguistics, 2021:2910-2922.

(责任编辑 梁碧芬)