

基于自注意力门控图卷积网络的特定目标情感分析

陈佳伟, 韩芳*, 王直杰

(东华大学 信息科学与技术学院, 上海 201620)

(* 通信作者电子邮箱 yadiahah@dhu.edu.cn)

摘要: 基于特定目标的情感分析旨在预测句子中不同方面表达的不同情感倾向。针对之前利用循环神经网络(RNN)结合注意力机制的网络模型所带来的训练参数多且缺少对相关句法约束和长距离词依赖机制解释的问题, 提出自注意力门控图卷积网络MSAGCN。首先, 模型采用多头自注意力机制编码上下文词和目标, 捕获句子内部的语义关联; 然后, 采用在句子的依存树上建立图卷积网络的方法获取句法信息以及词的依存关系; 最后, 通过带有目标嵌入的门控单元(GTRU)获取特定目标的情感。与基线模型相比, 所提模型的准确率和调和平均值F1分别提高了1%~3.3%和1.4%~6.3%; 同时, 预训练的BERT模型也被应用到当前任务中, 使模型效果获得了新的提升。实验结果表明所提出的模型能更好掌握用户评论的情感倾向。

关键词: 特定目标情感分析; 自注意力机制; 图卷积网络; 门控机制; BERT

中图分类号: TP389.1 **文献标志码:** A

Aspect-based sentiment analysis with self-attention gated graph convolutional network

CHEN Jiawei, HAN Fang*, WANG Zhijie

(College of Information Science and Technology, Donghua University, Shanghai 201620, China)

Abstract: Aspect-based sentiment analysis tries to estimate different emotional tendencies expressed in different aspects of a sentence. Aiming at the problem that the existing network model based on Recurrent Neural Network (RNN) combined with attention mechanism has too many training parameters and lacks explanation of related syntax constraints and long distance word dependence mechanism, a self-attention gated graph convolutional network was proposed, namely MSAGCN. First, the multi-headed self-attention mechanism was used to encode context words and targets, thus capturing semantic associations within the sentence. Then, a graph convolutional network was established on the sentence's dependency tree to obtain syntactic information and word dependencies. Finally, the sentiment of the specific target was obtained through the GTRU (Gated Tanh-ReLU Unit). Compared with the baseline model, the proposed model has the accuracy and F1 improved by 1%-3.3% and 1.4%-6.3% respectively. At the same time, the pre-trained Bidirectional Encoder Representations from Transformers (BERT) model was also applied to the current task to further improve the model effect. Experimental results verify that the proposed model can better grasp the emotional tendencies of user reviews.

Key words: aspect-based sentiment analysis; self-attention mechanism; Graph Convolutional Network (GCN); gating mechanism; Bidirectional Encoder Representations from Transformers (BERT)

0 引言

基于特定目标的情感分析 (Aspect-Based Sentiment Classification, ABSC) 是一种细粒度的自然语言处理任务, 是情感分析领域^[1] 一个重要的分支。

传统的机器学习方法^[2-3] 主要基于特征工程, 然而这类方法需要大量的人力和物力来设计特征, 并且需要获得足够多的语料来构造词典。支持向量机 (Support Vector Machine, SVM)^[2] 在该领域取得了可观的性能。然而, 与大多数传统的机器学习方法类似, 这种方法过于繁琐, 不适合手工设计特征。另外, 当数据集发生变化时, 方法的性能也会受到很大影响。因此, 基于传统机器学习的方法通用性差, 难以应用于多种数据集。

由于基于神经网络的方法具有很强的捕获原始特征的能

力, 因此近年来的研究工作越来越多地与神经网络相结合。词嵌入^[4] 是基于深度神经网络方法的基础, 它将自然语言表示为连续的低维向量, 通过对向量的操作来学习自然语言的交互特性。Word2Vec (Word to Vector)^[5]、PV (Paragraph Vectors)^[6] 和 GloVe (Global Vectors)^[7] 等便是预训练好的词向量。为了将目标词融入到模型中, Tang 等^[8] 提出 TD-LSTM (Target-Dependent Long-Short Term Memory) 模型来扩展长短期记忆 (Long-Short Term Memory, LSTM) 模型, 分别使用两个单向 LSTM 对目标词的左上下文和右上下文进行建模。随着注意力机制模型成功应用于机器翻译^[9], 目前已经有一些研究使用注意力机制生成特定目标的句子表示^[10-12], 还有一些研究根据目标词转换句子表示^[13]。除了基于注意力机制的模型, 卷积神经网络 (Convolutional Neural Network, CNN) 已经被用于捕获上下文中描述特定目标的多词短语的语义特征。多

收稿日期: 2019-12-24; 修回日期: 2020-02-27; 录用日期: 2020-03-08。

基金项目: 国家自然科学基金资助项目 (11972115, 11572084)。

作者简介: 陈佳伟 (1994—), 男, 江苏南通人, 硕士研究生, 主要研究方向: 自然语言处理、机器学习; 韩芳 (1981—), 女, 山西朔州人, 教授, 博士, 主要研究方向: 智能系统、神经动力学; 王直杰 (1969—), 男, 浙江台州人, 教授, 博士, 主要研究方向: 神经网络、数字信号处理、工业控制软件。

注意力卷积神经网络(Multi-ATTention Convolutional Neural Network, MATT-CNN)^[14]将多种注意力机制与CNN结合,能获取更深层次的情感特征信息。门控卷积网络(Gated Convolutional network with Aspect Embedding, GCAE)^[15]采用CNN和门控机制有效地选择给定目标词的情感特征,且该模型可以并行计算,提升了训练速度。

大多数之前的工作存在以下几个问题:第一,句子的建模依赖于循环神经网络(Recurrent Neural Network, RNN)(如LSTM、门控循环单元(Gated Recurrent Unit, GRU)和注意力机制。然而,RNN很难并行化,并且需要大量的内存和计算量;另外,RNN缺乏在长时间范围内捕获语义依赖关系的能力。第二,尽管注意力机制能关注到句子的重要部分,但是不善于捕捉句子中上下文单词和目标之间的句法依赖关系,导致模型将语义依赖关系不相关的上下文词作为给定目标的描述。例如“*Its size is ideal and the weight is acceptable*”,基于注意力机制的模型通常会把 *acceptable* 作为目标 *size* 的描述。为了解决这个问题,He等^[16]在注意力权重上施加句法约束,但没有充分发挥句法结构的作用。

基于此,本文提出基于自注意力门控图卷积网络,并在3个公开数据集上对模型的效果进行了验证,实验结果表明,本文模型能有效地解决上文所提到的问题。本文的主要工作如下:

1)采用多头自注意力机制对目标和句子编码,从而获取其隐藏语义特征和两者之间的语义关联,摒弃RNN编码并且减少训练参数。

2)在对句子进行依存句法分析的基础上,采用两个图卷积网络(Graph Convolutional Network, GCN)分别对句子和目标抽取句法信息并且捕获词的依存关系,解决长距离词依赖机制。

3)将预训练的BERT(Bidirectional Encoder Representations from Transformers)应用到此次任务中,相比非BERT模型获得了更好的效果。

1 相关工作

1.1 模型定义

对于长度为 n 的句子序列 $w^s = \{w_1^s, w_2^s, \dots, w_n^s\}$ 和长度为 m 的目标序列 $w^t = \{w_1^t, w_2^t, \dots, w_m^t\}$,其中 w^t 是 w^s 的子序列。本文的目的是预测句子 w^s 中目标 w^t 的情感极性。

本文提出的基于自注意力门控图卷积网络的结构如图1所示,主要包括词嵌入层、自注意力编码层、图卷积层、门控单元和输出层。

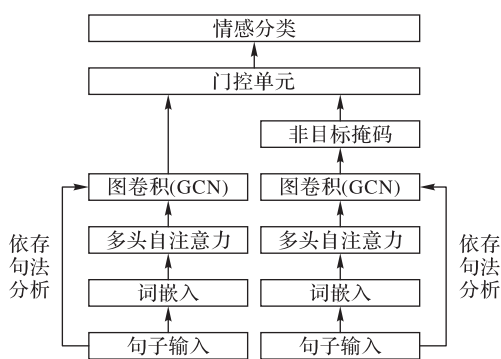


图1 自注意力门控图卷积网络

Fig. 1 Self-attention gated graph convolutional network

词嵌入层把句子转换成模型的输入;自注意力编码层编码句子和目标,进行语义特征的预提取;图卷积层捕捉上下文词和目标的句法依赖关系,非目标掩码获取目标的特征表示;门控单元提取情感特征。词嵌入层有两种:GloVe嵌入层和BERT嵌入层。

1.2 词嵌入层

1.2.1 GloVe嵌入层

定义 $M \in \mathbf{R}^{d_{emb} \times |V|}$ 为预训练的GloVe嵌入矩阵,其中 d_{emb} 是词向量的维度, $|V|$ 是单词个数。然后把每一个单词 $w^i \in \mathbf{R}^{1 \times |V|}$ 映射为其相应的词向量 $e_i \in \mathbf{R}^{d_{emb} \times 1}$,该词向量是词嵌入矩阵 M 中的列向量。

1.2.2 BERT嵌入层

BERT^[17]嵌入层是一个预训练的Seq2Seq语言理解模型,使用预训练的BERT生成序列的词向量。BERT层有足够的力量抓取文本的语义特征。为了获得更好的效果,模型在训练过程中进行微调,采用两个独立的BERT嵌入层分别建模句子特征和目标特征。

1.3 自注意力编码层

为了更好地编码句子和目标的语义特征,采用Transformer结构^[18]中提出的一种新的注意力机制——多头自注意力机制表示句子和目标。多头自注意力机制执行多个注意力功能来计算每个上下文单词的注意力分数。与其他自注意力机制相比,缩放点积注意(Scaled Dot product Attention, SDA)更快,而且在计算时更有效,故本文使用SDA作为注意力函数。

假设 X_{SDA} 是通过嵌入层嵌入的输入表示。SDA定义如下:

$$SDA(X_{SDA}) = \text{Softmax} \left(\frac{Q \cdot K^T}{\sqrt{d_k}} \right) \cdot V \quad (1)$$

$$Q, K, V = f_x(X_{SDA}) \quad (2)$$

$$\begin{cases} Q = X_{SDA} \cdot W_q \\ K = X_{SDA} \cdot W_k \\ V = X_{SDA} \cdot W_v \end{cases} \quad (3)$$

其中: Q, K 和 V 分别是上层隐状态乘以它们各自的权重矩阵 $W_q \in \mathbf{R}^{d_h \times d_q}, W_k \in \mathbf{R}^{d_h \times d_k}, W_v \in \mathbf{R}^{d_h \times d_v}$,这些权重矩阵在学习的过程中是可训练的。维度 d_q, d_k, d_v 等于 d_h/h ,其中 d_h 是隐状态的维度。把每个头学习到的注意力表示拼接再乘以矩阵 W^{MH} 进行转换。在本文模型中,注意力头的个数 h 设置为12。假设 H_i 是每个注意力头学习到的表示,有:

$$MHSA(X) = \tanh(\{H_0; H_1; \dots; H_h\} \cdot W^{MH}) \quad (4)$$

其中:“;”表示向量的拼接, $W^{MH} \in \mathbf{R}^{hd_h \times d_h}$ 。另外, $\tanh$ 激活函数用来增强MHSA编码器语义表示的学习能力。

1.4 基于依存树的图卷积

1.4.1 依存句法分析

句法是指句子的各个组成部分的相互关系,依存句法分析是句法分析中的一种方法。依存句法分析通过分析语言单位内成分之间的依存关系揭示其句法结构,主张句子中核心动词是支配其他成分的中心成分,而它本身却不受其他任何成分的支配,所有受支配成分都以某种依存关系从属于支配者。本文使用spaCy对句子进行依存句法分析,示意结果如图2所示。

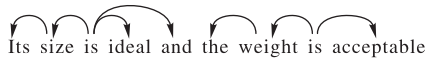


图2 依存句法分析图

Fig. 2 Dependency syntax analysis graph

1.4.2 图卷积

为了解决句法约束和长距离依赖机制,本文在对句子进行依存句法分析的基础上建立图卷积网络(GCN)^[19]。具体地,同时对文本和目标采用图卷积操作以充分考虑两者的句法约束关系和长距离词依赖机制。GCN可以被认为是对传统CNN编码非结构化数据局部信息的一种改进。具体地说,通过对给定句子构建依存树获得依存矩阵 $A \in \mathbf{R}^{n \times n}$,其中 n 是图的节点个数。为了方便,定义第 l 层的第 i 个节点的输出为 h_i^l ,其中 h_i^0 代表节点 i 的初始化状态。对于一个 L 层的图卷积网络, $l \in [1, 2, \dots, L]$, h_i^l 是节点 i 的最终状态。为了让节点更好地了解上下文,本文把预抽取的特征作为节点的特征输入GCN,然后用归一化因子的图卷积运算^[10]更新每个节点表示,具体过程如下所示:

$$\tilde{h}_i^l = \sum_{j=1}^n A_{ij} W^l g_j^{l-1} \quad (5)$$

$$h_i^l = \text{ReLU}(\tilde{h}_i^l / (d_i + 1) + b^l) \quad (6)$$

其中: $g_j^{l-1} \in \mathbf{R}^{1 \times 2d_h}$ 是前一层GCN第 j 个节点的输出表示; $h_i^l \in \mathbf{R}^{1 \times 2d_h}$ 是当前GCN层的输出表示; $d_i = \sum_{j=1}^n A_{ij}$ 是树中第 i 个标记的程度系数;权重矩阵 W^l 和偏置 b^l 是可训练参数。

本文没有将 h_i^l 的特征表示直接输入GCN层,而是首先对它进行位置编码:

$$g_i^l = F(h_i^l) \quad (7)$$

其中, $F(\cdot)$ 是位置编码函数,用于减少依存句法分析过程中产生的噪声和偏差,增强目标周围词的重要性。位置编码函数如下所示:

$$F(h_i^l) = q_i h_i^l \quad (8)$$

$$q_i = \begin{cases} 1 - \frac{r+1-i}{n}, & 1 \leq i < r+1 \\ 0, & r+1 \leq i \leq r+m \\ 1 - \frac{i-r-m}{n}, & r+m < i \leq n \end{cases} \quad (9)$$

其中, $q_i \in \mathbf{R}$ 是第 i 个单词的位置权重。 L 层GCN的最终输出是 $H^L = \{h_1^L, h_2^L, \dots, h_{r+1}^L, \dots, h_{r+m}^L, \dots, h_{n-1}^L, h_n^L\}$, $h_i^L \in \mathbf{R}^{1 \times 2d_h}$ 。

为了获取目标特征,对非目标词的隐状态向量进行掩码操作,并且保持目标词状态不变:

$$h_t^L = 0; 1 \leq t < r+1 \text{ 且 } r+m < t \leq n \quad (10)$$

掩码层的输出是面向目标的特征 $H_{\text{mask}}^L = \{0, \dots, h_{r+1}^L, \dots, h_{r+m}^L, \dots, 0\}$ 。通过图卷积,这些特征 H_{mask}^L 已经感知到了目标周围的上下文,这种感知方式同时考虑了句法依赖和长距离多词关系。

1.5 门控单元

带有目标嵌入的门控单元(Gated Tanh-ReLU Unit, GTRU)在每个位置 t 与两个卷积神经元相连。情感特征 c_i 的计算如下:

$$c_i = s_i \times a_i \quad (11)$$

$$a_i = \text{ReLU}(X_{i:i+k} * W_a + V_a v_a + b_a) \quad (12)$$

$$s_i = \tanh(X_{i:i+k} * W_s + b_s) \quad (13)$$

其中: v_a 是图卷积层后经过非目标掩码后的目标特征向量。

式(12)和(13)包含两个相同的卷积操作:特征 a_i 通过ReLU激活函数接收目标信息 v_a , $a_i \in [0, 1]$,表示目标与该位置文本的语义相似度; $s_i \in [-1, 1]$ 为该位置卷积后的情感。

1.6 模型训练

本文的模型包括使用GloVe作为词嵌入和使用BERT预训练语言模型对网络结构进行微调。对于前者,把整个句子作为模型的输入;对于后者,把“[CLS]+整个句子+[SEP]+目标+[SEP]”作为句子建模的输入,把“[CLS]+整个句子+[SEP]”作为目标建模的输入。

模型采用带交叉熵损失和 L_2 正则化的标准梯度下降法进行训练:

$$L(\theta) = - \sum_{i=1}^c \hat{y}_i^c \ln(y_i^c) + \lambda \sum_{\theta \in \Theta} \theta^2 \quad (14)$$

其中: $\hat{y}_i^c \in \mathbf{R}^{1 \times c}$ 是真实标签, y_i^c 是输出层给出的预测标签, c 表示标签的类别; λ 是 L_2 正则化项的系数; Θ 是参数集。

2 实验与结果分析

2.1 实验环境

实验配置:CPU为Intel i7 8700K, RAM为8 GB, GPU为Nvidia GTX 1080, GPU加速库为CUDA 8.0.61, CUDNN 6.0.21, 实验系统为ubuntu 16.04, 开发环境为Pytorch, 开发工具为PyCharm, 开发语言为Python。

2.2 数据集和超参数

本文在以下3个数据集上进行实验:SemEval 2014数据集^[20],包括餐厅评论和笔记本电脑评论;ACL推特评论数据集^[21]。这些数据集带有三种情感标签:积极、中性和消极。每个数据集的训练集和测试集数量如表1所示。

表1 数据集信息

Tab. 1 Dataset information

数据集	积极用户评论条数		中性用户评论条数		消极用户评论条数	
	训练集	测试集	训练集	测试集	训练集	测试集
笔记本	994	341	870	128	464	169
餐厅	2164	728	807	196	637	196
推特	1561	173	1560	173	3127	346

本文实验中,GloVe词嵌入在训练过程中参数没有更新,但是本文使用预训练的BERT对模型进行微调。GloVe词嵌入维度为300,BERT维度为768,隐藏层的维度设置为300,使用Adam作为优化器,学习率为 1×10^{-3} ,BERT层在微调的过程中需要极小的学习率,设置为 2×10^{-5} 。另外,图卷积的层数设置为2,门控单元中卷积核的大小设置为3、4、5,卷积核的个数为100。本文采用准确率 acc 和调和平均F1来评估模型的性能。

2.3 模型比较

为了全面评价和分析本文模型的性能,在上述3个数据集上进行实验,并与多种基线模型进行对比。实验结果表明,本文模型MSAGCN和MSAGCN-BERT的准确率较高,尤其是基于BERT的MSAGCN-BERT模型。基线模型如下:

1) LSTM^[22]利用最后一层隐状态向量预测情感极性。

2) ATAE-LSTM (Attention-based LSTM with Aspect Embedding)^[10]利用注意力机制帮助模型关注与目标方面更相关的上下文。同时,该模型在每一个词的嵌入中都加入了目标信息,通过学习上下文和目标之间的隐藏关系来增强模型性能。

3) IAN (Interactive Attention Networks)^[11]通过两个LSTM

网络分别生成目标和上下文的表示,并利用交互注意力机制建模目标和上下文之间的关系。

4)GCAE^[15]通过两个卷积层各自建模目标信息和情感信息,两个门控单元根据给定的目标控制情感的流向,最终对情感作出判别。

5)ASGCN(Asspect-Specific Graph Convolutional Network)^[23]利用LSTM建模上下文信息,并以此作为节点特征输入图卷积网络抽取目标特征,并结合注意力机制获得最终上下文表示。

6)注意力编码网络(Attentional Encoder Network, AEN)^[24]通过注意力机制编码上下文和目标,通过特定目标注意力层抽取带有目标信息的句子表示,并将其与上下文和目标表示拼接从而进行交互学习。

7)BERT-SPC^[24]预训练BERT模型的句子匹配分类任务,该模型将输入序列转换为“[CLS]+句子+[SEP]+目标+[SEP]”。

2.4 结果分析

为了保证实验结果的准确性,本次实验的模型(包括基线模型)均运行在相同的实验环境下。各模型的准确率 acc 和 $F1$ 值的结果如表2所示。其中, $F1$ 为准确率和召回率的调和平均。

表 2 各模型的准确率和F1值对比 单位: %

Tab. 2 Comparison of accuracy and F1 value of each model unit: %

模型	笔记本		餐厅		推特	
	acc	$F1$	acc	$F1$	acc	$F1$
基线模型	LSTM	69.59	61.90	77.05	62.64	68.35
	ASGCN	72.88	68.03	79.82	70.12	70.66
	ATAE-LSTM	68.65	62.17	77.77	66.73	67.63
	IAN	68.50	59.84	77.95	65.41	68.35
	GCAE	70.06	62.65	78.84	69.26	70.23
	AEN	70.06	63.37	77.32	61.66	68.79
	BERT-SPC	75.71	69.40	83.48	75.49	71.82
本文模型	MSAGCN	71.94	67.00	80.45	69.67	70.81
	MSAGCN-BERT	79.94	74.77	85.45	78.50	74.42

从表2的实验结果可以看出,相较于基线模型,本文提出的模型在三个数据集上的效果都有一定程度的提升,MSAGCN-BERT模型相较于其他基线模型效果得到了显著的提升,充分说明预训练BERT模型在本次任务中的重要性。相较于BERT-SPC模型,MSAGCN-BERT模型在三个数据集上准确率提高了2~4个百分点,说明了针对BERT下游任务设计特定网络结构是非常重要的。但是,相较于ASGCN模型,MSAGCN模型在笔记本数据集上的效果有所下降,经过分析,可能的原因是笔记本数据集对于句法依赖机制不是特别敏感。

使用循环神经网络建模句子和目标的模型(如LSTM、ATAE-LSTM、IAN)效果总体上不够理想,原因是模型很难记住长距离信息。同时,使用注意力机制的模型(如ATAE-LSTM、IAN)效果也没有获得明显的提升,原因是当句子包含多个目标且目标属性类似时,注意力机制可能会关注到其他目标的相关信息。使用卷积神经网络的模型(如GCAE)效果相较于循环神经网络模型效果有所提升,原因是CNN善于捕捉语义特征,但是由于缺乏对长距离词依赖机制的考虑,模型效果同样不尽如人意。注意力编码网络(AEN)避免了递归计算,但由于没有对嵌入的隐藏语义进行建模,其整体性能并不好,最后一层注意力机制的结果本质上是单词嵌入的线性组合。

2.5 模型分析

为了进一步研究本文模型各部分对性能的影响程度,对模型进行了消融实验,结果如表3所示。

从表3的实验结果可以看出,消融后的模型在准确率 acc 和 $F1$ 评价指标两方面效果都不如原模型。这一结果表明,所有这些被丢弃的部分对于提高模型性能都是非常重要的。首先,在每个图卷积层前取消对句子和目标表示的位置编码(MSAGCN w/o pos),结果显示模型在三个数据集上准确率 acc 和 $F1$ 值都有轻微下降,表明目标周围的上下文词的语义对最终的结果是重要的。另外,用LSTM网络(MSAGCN-LSTM)取代自注意力机制编码句子和目标,结果显示模型效果在三个数据集上普遍下降。最后,对句子建模取消图卷积操作(MSAGCN w/o cgcen)以及直接利用卷积对目标抽取特征(MSAGCN w/o tgcen),结果显示模型效果相较于原模型有明显的下降;同时发现,本次的消融比前两次的效果有了进一步的下降,表明利用图卷积获取句子依存信息对最终的情感分析结果起到了关键性作用。

表 3 消融实验结果对比 单位: %

Tab. 3 Comparison of ablation results unit: %

模型	笔记本		餐厅		推特	
	acc	$F1$	acc	$F1$	acc	$F1$
MSAGCN	71.94	67.00	80.45	69.67	70.81	68.97
MSAGCN w/o pos	70.69	64.93	79.20	68.36	69.08	66.71
MSAGCN-LSTM	71.16	66.29	79.02	67.82	68.21	65.99
MSAGCN w/o cgcen	70.69	64.01	77.95	66.37	68.50	65.49
MSAGCN w/o tgcen	68.03	59.46	77.86	67.51	69.80	67.76

为了验证本文提出的模型是否比结合注意力机制的RNN模型更轻量级,将本文模型的大小和基线模型中结合注意力机制的RNN模型进行对比,结果如表4所示。由于BERT本身具有模型大的特性,本文提出的基于BERT的模型大小也会相应变大。ATAE-LSTM、IAN、ASGCN都是基于注意力的RNN模型,因为RNN编码后的句子的所有隐藏状态必须同时保存在记忆单元中才能执行注意力机制,所以这些模型的记忆优化将更加困难,模型更大。

表 4 模型大小对比

Tab. 4 Model size comparison

模型	模型大小	
	参数/ 10^6	内存/MB
ATAE-LSTM	2.53	13.89
IAN	2.17	12.64
ASGCN	2.17	12.51
MSAGCN	1.72	10.69
MSAGCN-BERT	115.11	439.42

3 结语

针对RNN结合注意力机制的网络模型所产生的训练参数多且缺少对相关句法约束和长距离词依赖机制解释的问题,本文提出了基于自注意力门控图卷积网络模型解决特定目标情感分析任务。本文的模型避免了类似RNN结构的重复计算,利用自注意力捕获目标和其上下文词之间的交互语义,同时执行多个注意力功能捕获更加丰富的不同层面的语义信息。另外,考虑到距离目标词更近的上下文对该目标的情感重要程度更高,模型对句子采用位置编码。本文将编码后的句子作为特征并结合句子依存结构进行两次图卷积操作:一次作为句子的更深层次语义表示;另一次再通过掩码机

制过滤掉非目标词特征,只保留高级别目标特征,从而捕捉到上下文词和目标的句法依赖关系。利用门控机制提取与目标词有关的情感特征。还将预训练的BERT语言模型应用于此任务,进一步增强了模型的性能。实验和分析验证了本文模型的有效性。如何利用神经网络中大规模的情感词汇以及如何将外部常识融入神经网络获得更好的情感分析效果是我们未来的研究工作。

参考文献 (References)

- [1] 周立柱,贺宇凯,王建勇. 情感分析研究综述[J]. 计算机应用, 2008, 28(11): 2725-2728. (ZHOU L Z, HE Y K, WANG J Y. Survey on research of sentiment analysis[J]. Journal of Computer Applications, 2008, 28(11): 2725-2728.)
- [2] 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. Stroudsburg: Association for Computational Linguistics, 2014: 437-442.
- [3] VO D T, ZHANG Y. Target-dependent twitter sentiment classification with rich automatic features[C]// Proceedings of the 24th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2015: 1347-1353.
- [4] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model[J]. The Journal of Machine Learning Research, 2003, 3:1137-1155.
- [5] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]// Proceedings of the 26th International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2013: 3111-3119.
- [6] LE Q, MIKOLOV T. Distributed representations of sentences and documents[C]// Proceedings of the 31st International Conference on Machine Learning. New York: JMLR.org, 2014: 1188-1196.
- [7] 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. Stroudsburg: Association for Computational Linguistics, 2014: 1532-1543.
- [8] 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. Stroudsburg: Association for Computational Linguistics, 2016: 3298-3307.
- [9] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. [2019-11-24]. <https://arxiv.org/pdf/1409.0473.pdf>.
- [10] 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. Stroudsburg: Association for Computational Linguistics, 2016: 606-615.
- [11] MA D, LI S, ZHANG X, et al. Interactive attention networks for aspect-level sentiment classification[C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 4068-4074.
- [12] 李丽双,周安桥,刘阳,等. 基于动态注意力GRU的特定目标情感分类[J]. 中国科学:信息科学, 2019(8): 1019-1030. (LI L S, ZHOU A Q, LIU Y, et al. Aspect-based sentiment analysis based on dynamic attention GRU [J]. SCIENTIA SINICA Informationis, 2019(8): 1019-1030.)
- [13] LI X, BING L, LAM Y, et al. Transformation networks for target-oriented sentiment classification[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2018: 946-956.
- [14] 梁斌,刘全,徐进,等. 基于多注意力卷积神经网络的特定目标情感分析[J]. 计算机研究与发展, 2017, 54(8): 1724-1735. (LIANG B, LIU Q, XU J, et al. Aspect-based sentiment analysis based on multi-attention CNN [J]. Journal of Computer Research and Development, 2017, 54(8): 1724-1735.)
- [15] XUE W, LI T. Aspect based sentiment analysis with gated convolutional networks[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2018: 2514-2523.
- [16] HE R, WEE S L, NG H T, et al. Effective attention modeling for aspect-level sentiment classification[C]// Proceedings of the 27th International Conference on Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2018: 1121-1131.
- [17] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2019-10-24]. <https://arxiv.org/pdf/1810.04805.pdf>.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 6000-6010.
- [19] KIPF T N, WELLMING M. Semi-supervised classification with graph convolutional networks [EB/OL]. [2019-10-25]. <https://arxiv.org/pdf/1609.02907.pdf>.
- [20] PONTIKI M, GALANIS D, PAVLOPOULOS J, et al. SemEval-2014 task4: aspect based sentiment analysis[C]// Proceedings of the 8th International Workshop on Semantic Evaluation. Stroudsburg: Association for Computational Linguistics, 2014: 27-35.
- [21] DONG L, 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. Stroudsburg: Association for Computational Linguistics, 2014: 49-54.
- [22] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [23] ZHANG C, LI Q, SONG D. Aspect-based sentiment classification with aspect-specific graph convolutional networks [EB/OL]. [2019-11-13]. <https://arxiv.org/pdf/1909.03477.pdf>.
- [24] SONG Y, WANG J, JIANG T, et al. Attentional encoder network for targeted sentiment classification [EB/OL]. [2019-11-15]. <https://arxiv.org/pdf/1902.09314.pdf>.

This work is partially supported by the National Natural Science Foundation of China (11972115, 11572084).

CHEN Jiawei, born in 1994, M. S. candidate. His research interests include natural language processing, machine learning.

HAN Fang, born in 1981, Ph. D., professor. Her research interests include intelligent systems, neuro-dynamics.

WANG Zhijie, born in 1969, Ph. D., professor. His research interests include neural network, digital signal processing, industrial control software.