



基于注意力与双通道网络的方面级情感分析

摘要

针对方面级情感分析任务不能充分兼顾句法全面性与语义关联性,且大多数研究中使用的图卷积仅考虑信息自上而下的传播,忽略了信息自下而上的聚合等问题,本文提出了基于注意力与双通道网络的情感分析模型.该模型在扩展依存表示的同时使用自注意力获取具有语义关联的信息矩阵,使用双通道网络结合全局句法与语义关联信息,双通道网络分别侧重于自上而下传播的语义特征与自下而上聚合的结构特征.通道内的图卷积输出会与信息矩阵进行交互注意力起到残差互补的作用,然后通过平均池化完成通道内的任务.最后将基于语义与基于结构的决策融合得到最终的情感分类特征.实验结果表明该模型在三个公开数据集上的准确率与 F1 值均有提升.

关键词

注意力机制;双通道网络;决策融合;图卷积

中图分类号 TP391

文献标志码 A

收稿日期 2021-11-05

资助项目 国家自然科学基金(61273229,51705260);2016年度江苏高校“青蓝工程”科技创新团队培养对象

作者简介

杨春霞,女,博士,教授,研究方向为大数据分析、自然语言处理.y.cx@163.com

¹ 南京信息工程大学 自动化学院/江苏省大数据分析技术重点实验室/大气环境与装备技术协同创新中心,南京,210044

0 引言

自然语言处理(Natural Language Processing, NLP)旨在对文本数据进行分析处理.情感分析作为 NLP 中的一项任务,可以通过对文本分析得到用户想要表达的情感^[1].方面级情感分析(Asspect-Based Sentiment Analysis, ABSA)作为情感分析中的细粒度任务,可以针对语料中的特定方面分析出每句话的情感极性.随着数据量的激增,对文本进行情感分析可以反映出大众对互联网服务、产品分析、企业改进等方面的一些重要意见.

当前大多数 ABSA 任务通过句法的改进或语义信息的获取,研究其对于情感分析任务的影响.其中句法信息大多通过依存树获取文本中的结构.一般地,研究者会基于依存树进行修剪,通过改进句法结构用以提高情感分类的结果.此外,文本之间存在语义关联,通过挖掘深层次的语义信息也能取得不错效果.以往的研究表明无论从句法结构的角度或是语义信息的角度都能够很好地完成情感分类任务.但是大多数研究没有考虑到语义信息之间的关联性,并将其同时与句法结构结合,因此本文在扩展句法结构的同时获取到语义间的关联信息,可以完成两者的结合.

使用图卷积网络(Graph Convolution Network, GCN)可以将句法结构与语义关联信息相结合,用以完成情感分类任务.图卷积对于特征提取具有一定的优势,但是大多数研究仅自上而下传播考虑信息的语义特征,缺少自下而上传播聚集信息的结构特征.文献[2]在社交媒体谣言检测中结合两种传播方式,同时考虑因果与结构特征得到很好的结果,所以如何将因果特征与结构特征同时提取的方式融入 ABSA 任务之中是亟待解决的一个问题.有研究者使用多个网络通道处理不同的信息,通过融合多个通道的特征用以完成各自任务,如文献[3]在情感分析任务中使用两种不同输入信息的卷积网络以获取不同的特征信息,然后通过信息间的交互得出最终表示.于是本文考虑将自上而下传播的语义信息与自下而上聚合的结构信息分别作为两种网络,提出基于注意力与双通道网络的方面级情感分析模型,即双通道卷积模型(Two Channel Convolution Model, 简称 TCCM 模型).实验结果表明, TCCM 模型在 Laptop、Restaurant 与 Twitter 三个公开数据集上的准确率与 F1 值均有一定提升.

本文主要贡献如下:

1) 考虑到依存树生成的节点表示需要包含更全面的句法, 本文扩展了依存表示, 使依存信息中可以包含更全面的句法信息. 同时将编码后的信息通过自注意力完成语义信息的关联, 经过自注意力机制处理的信息可以看作有边的加权有向图, 用以表示语义信息相关性, 以达到兼顾句法全面性与语义关联性的目的.

2) 双通道网络分别为自上而下传播的语义特征网络与自下而上聚合的结构特征网络, 每个通道内都含有图卷积、交互注意力与平均池化. 双通道网络会得到基于语义特征与结构特征的两类决策, 将两类决策融合即为最终的情感特征.

1 相关工作

句法依存树可以提取句子结构中的语法信息, 许多研究对于依存树的修剪或句法结构的改善提出了各自的见解, 例如: 文献[4]通过多种解析, 生成不同句法结构的依存树, 并将其组合成一张有向图网络进行训练; 文献[5]构建一种以方面词为根节点的依存树, 与关系图注意力网络结合实现了情感预测; 文献[6]赋予依存树位置信息, 以此构建带有位置与权重信息的矩阵, 结合图卷积得到带有位置信息的表示; 文献[7]使用软剪策略修剪依存树, 通过实验验证了软剪依存树有助于效果提升. 这些方法虽然通过改善句法结构或修剪依存树取得了不错的效果, 但是难免会丢失部分有用的句法信息, 从而造成句法信息的缺失. 本文同时考虑句法信息的全面性与语义信息的关联性, 在句法结构方面进行了简单扩展, 使之成为包含更全面句法信息的依存表示.

注意力机制(简称注意力)可以赋予句子中每一个单词不同的权重, 用以代表单词的重要性. 它可以专注于特定的输入, 改善情感分析任务中的性能, 例如: 文献[8]提取上下文中表示长距离情感特征的实验, 采用的是多重注意力机制; 文献[9]采用多层注意力机制将上下文单词信息分配不同的权重, 最后一层注意力的结果即为输出的分类特征; 文献[10]使用交互注意力机制判断单词的不同影响力, 说明注意力对于关键信息的筛选提取存在着正向增益. 本文则使用自注意力进行语义信息关联, 通过权重分配出关联性较大的语义信息, 并与全面句法相结合, 最终达到同时兼顾句法结构改进与语义信息关联的目的.

图卷积网络可以进行特征的提取, 将句子中的

句法与语义融入情感分析中. 文献[11]将双向图卷积应用在关系抽取任务上, 证明了正反两个方向的卷积可以更好地捕获邻居节点间的信息; 文献[2]将双向图卷积应用在社交媒体检测中, 使用从父节点向子节点散布的自上而下的传播方式表示因果特征的提取, 从子节点向父节点聚合的自下而上的聚合方式表示结构信息的提取, 最后通过合并得到出色的效果. 本文考虑将这两种传播方式运用在 ABSA 任务中, 用以分别获取丰富的语义或结构特征.

近年来有研究者尝试通过多个通道网络的特征融合取得进一步的效果提升. 文献[2]在社交媒体检测中使用两种网络分别代表两种不一样的传播, 最终以融合的方式得出分类表示; 文献[3]在情感分析任务中使用两种不同输入图卷积网络, 以差分正则化等方式的融合也取得了不错的效果. 因此, 多网络融合也是一个值得考虑的点. 将多网络通道与两种传播方式相结合得出基于语义特征的决策与基于结构特征的决策, 经过融合即为本文基于注意力的双通道模型.

2 模型概述

本文的 TCCM 模型框架主要包含词嵌入、双向长短期记忆(Bi-Directional Long Short-Term Memory, BiLSTM)层、语义关联层、双通道网络模块、融合层、分类层, 其中每个通道网络中又包含图卷积层、交互注意力层、平均池化层. 具体结构如图1所示.

2.1 词嵌入与扩展的依存表示

本文使用文献[12]的方法对初始语料进行词嵌入. 指定一句话由 k 个单词组成, 可以表示为 $\{w_1, w_2, \dots, w_x^n, \dots, w_x^{n+L}, \dots, w_{k-1}, w_k\}$. 其中 $\{w_x^n, \dots, w_x^{n+L}\}$ 表示含有 L 个方面词. 句中的 $\{w_1, w_2, \dots, w_{k-1}, w_k\}$ 表示基于方面词的上下文单词. 对初始语料进行映射后得到词嵌入矩阵, 其形式为 $A \in \mathbf{R}^{H \times d_h}$, 其中 H 代表词汇表大小, d_h 代表词嵌入的维度.

利用自然语言处理工具包将初始语料处理生成句法依存树, 以此构建句法依存树的邻接矩阵 \mathbf{W} 作为句法结构. 在依存表示中的节点添加自循环如 $\mathbf{W}_{aa} = 1$. 若节点之间存在依存关系则 $\mathbf{W}_{ab} = 1$, 若节点之间不存在依存关系则 $\mathbf{W}_{ab} = 0$. 为了获得全面句法信息从而构建扩展的依存表示 \mathbf{Q}_{ab} . 其中与原始句法结构不同的点在于使用的初始矩阵为全 1 矩阵, 而非全 0 矩阵. 本文认为句子中所有单词之间本存在相应联系, 单从置 0 与置 1 的角度出发会割舍一些

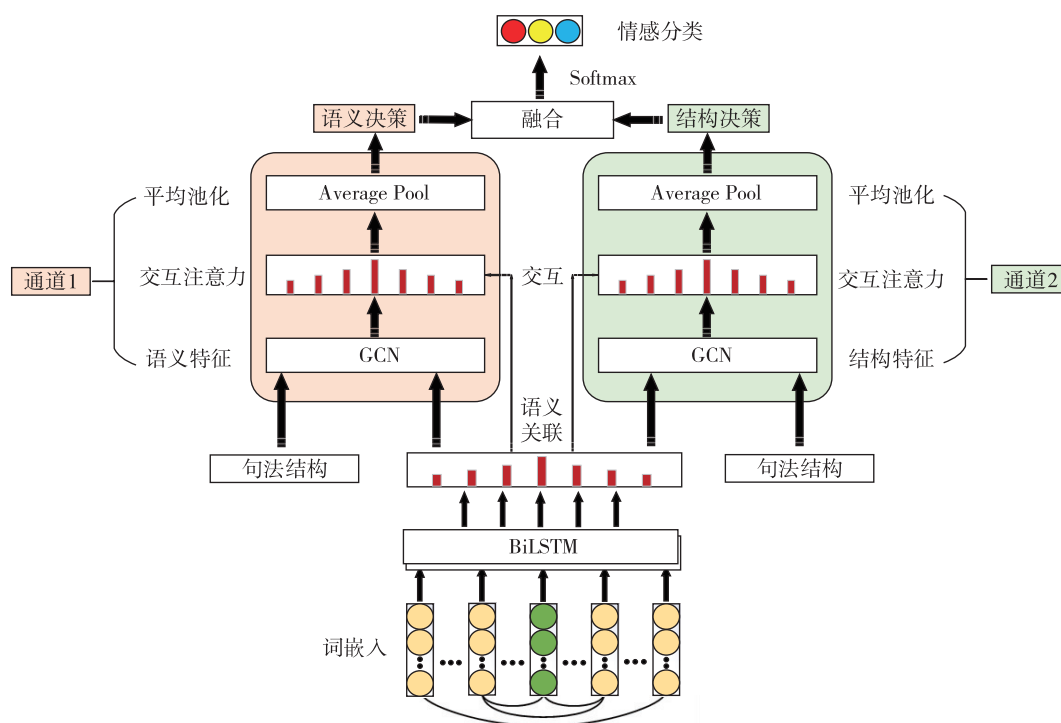


图1 TCCM 模型结构

Fig. 1 TCCM model structure

句法信息.在全 1 的初始矩阵上进一步更新节点关系,可以使得每个词之间存在关联的同时也能很好获取相应的节点关系,从而结构方面包含更全面的句法信息.将依存信息进行扩展是为与语义关联信息结合,兼顾句法信息全面性与语义信息关联性才是本文的目的.

2.2 BiLSTM

本文使用 BiLSTM 进行编码.传统的 LSTM 仅能单向处理信息,而 BiLSTM 可从正反两个方向对词向量进行处理,将正向 LSTM 与反向 LSTM 的输出合并得到含有上下文信息的隐层表示.LSTM 的计算公式如下:

$$d_t = \sigma(W_d \cdot [h_{t-1}, v_t] + b_d), \quad (1)$$

$$e_t = \sigma(W_e \cdot [h_{t-1}, v_t] + b_e), \quad (2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, v_t] + b_o), \quad (3)$$

$$\tilde{c} = \tanh(W_c \cdot [h_{t-1}, v_t] + b_c), \quad (4)$$

$$c_t = d_t * c_{t-1} + e_t * \tilde{c}, \quad (5)$$

$$h_t = o_t * \tanh(c_t), \quad (6)$$

式中: d_t, e_t, o_t 分别代表遗忘门、输入门、输出门; W_d, W_e, W_o, W_c 表示权重矩阵; b_d, b_e, b_o, b_c 为偏置值; σ, \tanh 分别表示 sigmoid 激活函数与双曲正切函数; c_{t-1}, c_t 分别表示 $t-1, t$ 时刻记忆细胞状态.本文使用两层 BiLSTM,计算方法如下:

$$\vec{h}_t = \overrightarrow{\text{LSTM}}(\vec{h}_{t-1}, s_t), \quad (7)$$

$$\overleftarrow{h}_t = \overleftarrow{\text{LSTM}}(\overleftarrow{h}_{t-1}, s_t), \quad (8)$$

$$h_t = [\vec{h}_t, \overleftarrow{h}_t], \quad (9)$$

其中: h_t 表示拼接两层正反向 LSTM 后的输出; s_t 表示 t 时刻输入.记 BiLSTM 对上下文编码后得到的隐层表示为 H^s , 如式 (10) 所示:

$$H^s = \{H_1, \dots, H_x^n, \dots, H_x^{n+L}, \dots, H_k\}, \quad (10)$$

式中 H^s 为经过 BiLSTM 编码后的隐层表示, 其中 $\{H_x^n, \dots, H_x^{n+L}\}$ 为方面词, $\{H_1, \dots, H_k\}$ 为基于方面词的上下文单词.

2.3 语义关联

对于编码后的信息一般没有分配信息间的权重,无法根据权重判断语义信息间的关联程度,所以本文使用自注意力机制完成语义信息的关联.自注意力机制处理后的信息可以看作有边的加权有向图,可用以表示语义信息的关联性.具体计算方式如下:

$$Q_1 = H^s \cdot (H^s)^T, \quad (11)$$

$$v_t^1 = \frac{\exp(Q_1)}{\sum_{i=1}^n \exp(Q_1)}, \quad (12)$$

$$\mu_1 = \sum_{i=1}^n v_i^1 \mathbf{H}^s, \quad (13)$$

其中 \mathbf{Q}_1 表示上下文中单词对方面词的相关性, v_i^1 表示方面词对上下文的注意力权重, μ_1 表示经过自注意力后的权重信息表示。

2.4 双通道网络

2.4.1 图卷积

图卷积网络可以从图的层面完成邻域信息提取。大多数情感分析任务的研究中自上而下传播考虑了信息的语义特征,缺少自下而上聚合信息结构特征,于是本文使用双通道图卷积网络融合两种信息传递的方式,分别得到基于语义特征与结构特征的信息表示。通道1与通道2中同时包含了图卷积层、交互注意力层、平均池化层。不同点在于通道1中使用如图2左图所示卷积方式,从父节点向子节点散布的自上而下的传播方式表示着语义特征的提取,通道2中使用如图2右图所示卷积方式,从子节点向父节点聚合的自下而上的聚合方式表示着结构信息的提取,这样双通道可以分别得到基于语义特征的情感决策与基于结构特征的情感决策,使之可以在求和平均层中融合两种决策的情感选择。

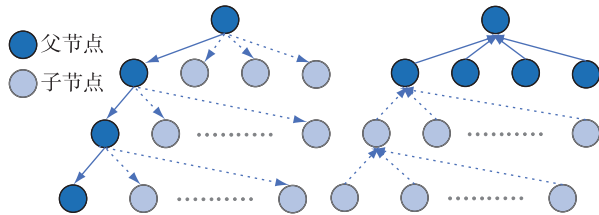


图2 两种图卷积方式

Fig. 2 Two graph convolution methods

通道1和通道2的图卷积具体计算方式分别如式(14)和式(15)所示:

$$\vec{\mu}_i = \rho \left(\sum_{j=1}^n \mathbf{Q}_{ab} \mathbf{W}^i \mu_j + \mathbf{b}^i \right), \quad (14)$$

$$\overleftarrow{\mu}_i = \rho \left(\sum_{j=1}^n \mathbf{Q}_{ab}^T \mathbf{W}^i \mu_j + \mathbf{b}^i \right), \quad (15)$$

式中 μ_1 为权重信息, \mathbf{W}^i 代表权重矩阵, \mathbf{b}^i 代表偏置值, ρ 代表 ReLu 激活函数, \mathbf{Q}_{ab} 为依存表示, \mathbf{Q}_{ab}^T 为转置的依存表示, $\vec{\mu}_i$ 为基于语义特征的图卷积输出, $\overleftarrow{\mu}_i$ 为基于结构特征的图卷积输出。

2.4.2 交互注意力

双通道图卷积网络中为了避免图卷积后产生的信息缺失,本文使用交互注意力机制进行互补。使用图卷积的输出与语义信息进行交互注意力,可以缓

解经过图卷积神经网络后产生的梯度消失问题,起到类似于残差网络的作用。具体计算方法如下:

$$\mathbf{Q}_2 = \mathbf{H}^s \cdot (\vec{\mu}_i)^T, \quad (16)$$

$$\mathbf{Q}_3 = \mathbf{H}^s \cdot (\overleftarrow{\mu}_i)^T, \quad (17)$$

$$v_i^2 = \frac{\exp(\mathbf{Q}_2)}{\sum_{i=1}^n \exp(\mathbf{Q}_2)}, \quad (18)$$

$$v_i^3 = \frac{\exp(\mathbf{Q}_3)}{\sum_{i=1}^n \exp(\mathbf{Q}_3)}, \quad (19)$$

$$\mu_2 = \sum_{i=1}^n v_i^2 \mu_1, \quad (20)$$

$$\mu_3 = \sum_{i=1}^n v_i^3 \mu_1, \quad (21)$$

其中: $\mathbf{Q}_2, \mathbf{Q}_3$ 分别表示通道1与通道2中图卷积与经过自注意力后信息表示的相关性; v_i^2, v_i^3 分别表示结合了隐层输出的双通道图卷积的注意力权重; μ_2, μ_3 分别表示通道1与通道2中图卷积经过交互注意力后的信息表示。

2.4.3 平均池化

双通道分别在经过图卷积与注意力的处理后方面词中包含了上下文信息的句法特征与语义关联信息,于是本文采取置0的方法,分别选取 μ_2, μ_3 中的方面词作为输入,将非方面词置为0,方面词即为原有表示。具体方法如下:

$$\mu_2 = 0, \mu_3 = 0, 1 \leq i < n, n+l < t \leq n, \quad (22)$$

其中: $\mu_2 = 0, \mu_3 = 0$ 为被置为0的非方面词; n 与 $n+1$ 为方面词的索引。接着将 μ_2, μ_3 中的方面词分别平均池化,用以提取方面词表示中的特征,具体方法如下:

$$\mu_2^s = f(\{\mu_{H^n}, \mu_{H^{n+1}}, \dots, \mu_{H^{n+l}}\}), \quad (23)$$

$$\mu_3^s = f(\{\mu_{H^n}, \mu_{H^{n+1}}, \dots, \mu_{H^{n+l}}\}), \quad (24)$$

其中: $f(\cdot)$ 表示将方面词经过平均池化函数; $\{\mu_{H^n}, \mu_{H^{n+1}}, \dots, \mu_{H^{n+l}}\}$ 为只含有方面词的表示; μ_2^s, μ_3^s 分别表示带有通道1中语义特征信息的方面词表示与带有通道2中结构特征信息的方面词表示。

2.5 融合层

将双通道中带有各自特征信息的方面词表示进行求和平均,得到同时包含语义特征信息与结构特征信息的决策:

$$\mu^s = (\mu_2^s + \mu_3^s) / 2, \quad (25)$$

式中 μ^s 为融合两种决策的最终表示。

2.6 分类层

将融合两种决策的方面词表示 μ^s 送入 Softmax 函数中,得到用以情感分类的概率,通过损失函数 (L) 不断优化模型参数:

$$L = - \sum_i^C \sum_j^Z y_i^j \log y_i^j, \quad (26)$$

式中 i 为数据样本的索引, j 为情感类别索引, C 为样本总数, Z 为情感类别数, y_i^j 为情感极性, y^j 为预测的情感极性.

3 实验

3.1 实验环境

本文的实验平台如表 1 所示.

表 1 实验平台

Table 1 Experimental platform

系统	Windows 10
CPU	Intel(R) Core(TM) i7-9750H
显卡	NVIDIA GeForce GTX 1660 Ti with Max-Q Design
存储	8.0 GB
语言	Python 3.8
平台	Pytorch 1.6.0
工具	Pycharm 2020

3.2 数据集

本文选取了 Laptop、Restaurant 与 Twitter 评论三个公开的数据集. Laptop、Restaurant 选自国际语义评测大会 SemEval-2014 Task (<https://aclanthology.org/S14-2004/>) 中的任务 4. Twitter 选自 ACL-2014 (<https://aclanthology.org/P14-2009/>). 三个数据集中包含消极、中性和积极三种不同情感极性, 具体数据分布个数如表 2 所示.

表 2 数据集数据分布情况

Table 2 Data distribution of three data sets

情感极性	Laptop	Restaurant	Twitter
积极	976	2 164	1 561
中性	455	637	3 127
消极	851	807	1 560
积极	337	727	173
中性	167	196	346
消极	128	196	173

3.3 实验参数与评价指标

3.3.1 参数设置

本文使用 300 维的词向量进行初始化, 使用两

层 BiLSTM、一层 GCN, 在 Laptop 与 Restaurant 数据集上使用 Adam 优化器进行优化, 在 Twitter 数据集上采用 Adamax 优化器进行优化. 为了防止过拟合采用了 dropout. 具体参数设置如表 3 所示.

表 3 参数设置

Table 3 Parameter setting

参数设置	值
Embedding_dim	300
Batch size	32
Learning rate	0.01
Num_epoch	30
input_dropout	0.7
gcn_dropout	0.7
gcn_hidden_dim	200
rnn_hidden	100

3.3.2 评价指标

本文采用准确率 (Accuracy, A) 和 MacroF1 值为评价指标, 具体公式如下:

$$P = \frac{TP}{TP+FP}, \quad (27)$$

$$R = \frac{TP}{TP+FN}, \quad (28)$$

$$MacroP = \frac{1}{n} \sum_{i=1}^n P_i, \quad (29)$$

$$MacroR = \frac{1}{n} \sum_{i=1}^n R_i, \quad (30)$$

$$MacroF1 = \frac{2 \times MacroP \times MacroR}{MacroP + MacroR}, \quad (31)$$

$$A = \frac{T}{N}, \quad (32)$$

其中 P 表示精确率, R 表示召回率, n 表示类别数, TP 表示正样本预测为正, FP 表示正样本预测为负, FN 表示负样本预测为负, TN 表示负样本预测为正, $MacroF1$ (简称 $F1$) 表示各个类别 $F1$ 的平均值, T 为正样本数, N 为总样本数.

3.4 对比实验

本文选用以下模型作为对比, 且在相同实验环境下运行, 结果具体如表 4 所示.

1) IAN^[13]: 同时建模方面词与上下文信息, 可以使得方面词与上下文信息用注意力交互融合.

2) AOA^[14]: 使用长短期记忆神经网络同时对方面词和文本建模, 以此关注句中重要内容.

3) AEN-Glove^[15]:使用含有注意力机制的编码器对目标和上下文信息建模,可以挖掘出词嵌入中的交互式语义信息.

4) R-GAT(Stanford)^[5]:通过修剪依存树,重塑了一个以方面为基础的依存树,使用关系图注意力编码树结构.

5) ASGCN^[16]:使用图卷积网络处理依存关系,利用句间句法依存结构来解决长期依存问题.

6) Repwalk^[17]:提出一种新型神经网络,使用多路径语法图,并在图上进行随机游走策略.

7) CDT^[18]:提出卷积依赖模型,识别句中特定方面的词语情感,将依存树与图卷积融合进行表征学习.

本文选取的对比模型可以分为含有注意力机制与图神经网络两类,其中 IAN、AOA、AEN-Glove 主要利用注意力机制完成分类任务,R-GAT、ASGCN、Repwalk、CDT 均使用依存树获取语法信息并利用图神经网络完成分类任务,但是这些模型中仅包含单一通道.所有模型均不使用预训练语言模型.

3.5 对比实验分析

由表 4 可知,使用注意力机制或图神经网络的模型都有着不错的效果,但本文同时结合注意力与双通道网络的 TCCM 模型效果更好.相比于使用单通道的模型,双通道网络通过两种不同的信息传递方式更好地获取到侧重于语义特征或结构特征两种信息,对情感分析任务的提升有着一定的帮助.相比于使用注意力的模型,使用自注意力可以获取语义信息的关联性并与扩展的句法结构搭配作为图卷积的输入,兼顾了两种信息的 TCCM 模型自然可以取得更好的分类效果.

表 4 不同模型的实验结果

Table 4 Experimental results of different models %

模型	Laptop		Restaurant		Twitter	
	A	F1	A	F1	A	F1
IAN	72.05	67.38	79.26	70.09	72.50	70.81
AOA	72.88	67.48	80.53	69.84	72.25	69.96
AEN-Glove	73.51	69.04	80.98	72.14	72.83	69.81
R-GAT	73.55	69.32	79.55	70.00	71.65	69.80
ASGCN	76.01	71.44	81.16	72.14	72.39	70.80
Repwalk	76.80	73.59	82.32	73.87	73.84	72.32
CDT	76.56	72.34	81.46	73.59	74.37	72.79
TCCM	77.65	73.81	82.64	74.92	75.87	73.49

3.6 模型分析实验

3.6.1 扩展依存表示的有效性

为了说明扩展依存表示的有效性,本文建立了 WEDR(Without Extended Dependency Representation)模型代表消融扩展依存表示的模型,并在三个数据集上进行了对比实验,实验结果如表 5 所示,实验中保留了注意力与双通道图卷积网络.

表 5 扩展依存表示的有效性

Table 5 Validity of extended dependency representation %

模型	Laptop		Restaurant		Twitter	
	A	F1	A	F1	A	F1
WEDR	73.22	68.11	80.49	68.99	73.80	72.03
TCCM	77.65	73.81	82.64	74.92	75.87	73.49

从表 5 可以看出 WEDR 和 TCCM 准确率和 F1 值相差较多,主要原因是本文模型将扩展依存表示与注意力相结合,以兼顾全面句法信息与语义信息间的关联性为目的,而消融了依存表示后的 WEDR 模型在无法包含更全面句法信息的前提下使用自注意力会导致信息的丢失以及梯度的弥散.实验结果表明本文模型可以包含更全面的句法信息,并与注意力机制的结合能够更好地完成任务.

3.6.2 双通道网络的有效性

为了证明使用双通道网络的有效性,本文构建了 SFCN(Semantic Feature Channel Network)模型代表语义特征通道网络,即只考虑语义特征的信息决策,SCCN(Structural Feature Channel Network)模型代表结构特征通道网络,即只考虑结构特征的信息决策,两种模型都保留了扩展的依存表示与注意力机制.

由表 6 可知,无论是消融了结构特征的 SFCN 模型还是消融了语义特征的 SCCN 模型效果都略低于双通道的 TCCM 模型,说明双通道图卷积网络可以通过语义特征以及结构特征两个方面更好地进行信息获取.因为父节点向子节点散布的传播偏向于语

表 6 单通道消融

Table 6 Single channel ablation %

模型	Laptop		Restaurant		Twitter	
	A	F1	A	F1	A	F1
SFCN	75.83	71.34	81.65	72.27	74.51	73.08
SCCN	76.19	72.55	81.74	72.53	74.09	71.83
TCCM	77.65	73.81	82.64	74.92	75.87	73.49

义特征的提取,子节点向父节点聚合的方式偏向于结构信息的提取,通过两个角度的决策融合可以更好地完成情感分类任务,证明了本文融合双通道网络决策的可行性,双通道的融合可以使得分类结果有着一定提升.

3.6.3 注意力的有效性

为了验证注意力机制的有效性,建立了 NAM (No Attention Mechanism)、NSAM (No Self Attention Mechanism)、NIAM (No Interactive Attention Mechanism) 模型分别代表消融注意力、消融自注意力、消融交互注意力的模型.注意力消融如表 7 所示.实验结果表明注意力的存在是有一定效果的,两种注意力机制的结合可以很好完成语义信息的关联,以及弥补信息的丢失.其次在实验中分别对自注意力与交互注意力进行消融,以准确率为指标研究单个注意力在三个数据集上对于模型的影响,结果如图 3 所示.实验结果表明结合两种注意力的 TCCM 模型的准确率高于一保留交互注意力的 NSAM 模型和只保留了自注意力的 NIAM 模型.所以自注意力和交互注意力均有存在的必要性.自注意力可以更好选择关注对情感分类任务重要的信息,获取语义信息间的关联性;经过图卷积处理的数据存在一定程度的信息丢失,而交互注意力在双通道网络中起到残差互补的作用,将图卷积的输出与权重信息进行二次交互,用以解决经过图卷积层后造成的信息缺失.

综上所述,采用单一注意力机制的模型效果低于同时结合两种注意力的模型,从而证明本文中使用了两种注意力机制的 TCCM 模型比单一注意力模型更优越.

表 7 注意力的有效性

Table 7 Effectiveness of attention %

模型	Laptop		Restaurant		Twitter	
	A	F1	A	F1	A	F1
NAM	76.09	72.58	81.65	72.08	74.09	72.27
TCCM	77.65	73.81	82.64	74.92	75.87	73.49

4 结束语

为了解决兼顾句法信息全面性与语义信息关联性的问题,同时为了将两种信息传递方式融合至情感分析任务中,本文提出了 TCCM 模型,使之可以在考虑全面句法结构信息时得到语义信息的关联程度.其中扩展的依存表示包含更全面的句法信息,自

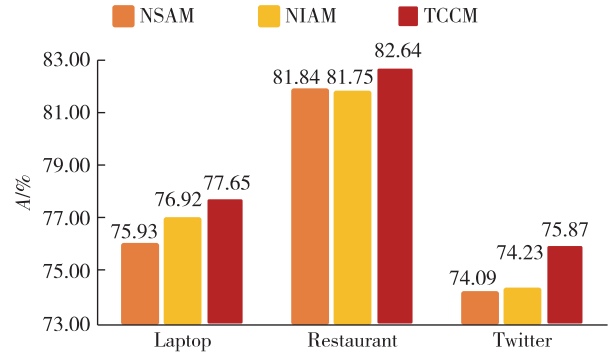


图 3 注意力机制消融

Fig. 3 Attention mechanism ablation

注意力可以获取语义关联性,双通道图卷积网络通过两个通道实现两种不同的信息决策,通道中的交互注意力避免经过图卷积层后产生信息缺失问题.最后融合双通道的决策并用 Softmax 完成情感分类.通过三个公开数据集上的实验结果,证明该模型的有效性和可行性.此外,对于每部分的消融实验也能证明各个模块安排的合理性.下一步的工作可以考虑如何将语义分析与句法结构分析分开进行,从而实现更精准的语义与句法交互,也可以考虑在任务中使用预训练语言模型以提高分类效果.

数据申明

本文使用 Glove 词向量模型对初试语料进行词嵌入,可在 DOI: 10.3115/v1/D14-1162, URL: <https://aclanthology.org/D14-1162> 进行查阅. Laptop、Restaurant 数据集选自国际语义评测大会 SemEval-2014 Task 中的任务 4,可在 DOI: 10.3115/v1/S14-2004, URL: <https://aclanthology.org/S14-2004> 进行查阅. Twitter 数据集选自 ACL-2014,可在 DOI: 10.3115/v1/P14-2009, URL: <https://aclanthology.org/P14-2009> 进行查阅.其余实验数据均在本文实验环境下运行得出.

参考文献

References

- [1] 罗嘉,王乐豪,涂姗姗,等.基于 LSTM-BLS 的突发气象灾害事件中公众情感倾向分析[J].南京信息工程大学学报(自然科学版),2021,13(4):477-483
LUO JIA, WANG Lehao, TU Shanshan, et al. Analysis of public sentiment tendency in sudden meteorological disasters based on LSTM-BLS[J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition), 2021, 13(4): 477-483
- [2] Bian T, Xiao X, Xu T Y, et al. Rumor detection on social

- media with bi-directional graph convolutional networks [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1):549-556
- [3] Li R F, Chen H, Feng F X, 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. DOI: 10.18653/v1/2021.acl-long.494
- [4] Hou X C, Qi P, Wang G T, et al. Graph ensemble learning over multiple dependency trees for aspect-level sentiment classification [J]. arXiv e-print, 2021, arXiv:2103.11794
- [5] Wang K, Shen W Z, Yang Y 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, 2020. DOI: 10.18653/v1/2020.acl-main.295
- [6] Chen G M, Tian Y H, Song Y. Joint aspect extraction and sentiment analysis with directional graph convolutional networks [C] // Proceedings of the 28th International Conference on Computational Linguistics, 2020. DOI: 10.18653/v1/2020.coling-main.24
- [7] 王晓霞, 钱雪忠, 宋威. 基于注意力与图卷积网络的关系抽取模型 [J]. 计算机应用, 2021, 41(2):350-356
WANG Xiaoxia, QIAN Xuezhong, SONG Wei. Relation extraction model via attention-based graph convolutional network [J]. Journal of Computer Applications, 2021, 41(2):350-356
- [8] Chen P, Sun Z Q, Bing L D, et al. Recurrent attention network on memory for aspect sentiment analysis [C] // Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017. DOI: 10.18653/v1/D17-1047
- [9] Tang D Y, Qin B, Liu T. Aspect level sentiment classification with deep memory network [C] // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016. DOI: 10.18653/v1/D16-1021
- [10] 马远, 程春玲. 融合左右双边注意力机制的方面级别文本情感分析 [J]. 计算机应用研究, 2021, 38(6):1753-1758
MA Yuan, CHENG Chunling. Joint left and right attention mechanism for aspect-level text sentiment analysis [J]. Application Research of Computers, 2021, 38(6):1753-1758
- [11] Fu T J, Li P H, Ma W Y. GraphRel: modeling text as relational graphs for joint entity and relation extraction [C] // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019:1409-1418
- [12] 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 (EMNLP), 2014. DOI: 10.3115/v1/D14-1162
- [13] Ma D H, Li S J, Zhang X D, et al. Interactive attention networks for aspect-level sentiment classification [J]. arXiv e-print, 2017, arXiv:1709.00893
- [14] Huang B X, Ou Y L, Carley K M. Aspect level sentiment classification with attention-over-attention neural networks [C] // International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. Springer, Cham, 2018:197-206
- [15] Song Y W, Wang J H, Jiang T, et al. Attentional encoder network for targeted sentiment classification [J]. arXiv e-print, 2019, arXiv:1902.09314
- [16] Zhang C, Li Q C, Song D W. 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), 2019:4560-4570
- [17] Zheng Y W, Zhang R C, Mensah S, et al. Replicate, walk, and stop on syntax: an effective neural network model for aspect-level sentiment classification [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(5):9685-9692
- [18] Sun K, Zhang R C, 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), 2019:5683-5692

Aspect level sentiment analysis based on attention and dual channel network

YANG Chunxia¹ XU Ben¹ GUI Qiang¹ HAN Yu¹

¹ School of Automation/Jiangsu Key Laboratory of Big Data Analysis Technology (B-DAT)/ Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET), Nanjing University of Information Science & Technology, Nanjing 210044

Abstract In view of the problems that aspect level sentiment analysis tasks cannot give full consideration to syntactic comprehensiveness and semantic relevance, and the graph volume used in most studies only considers the top-

down dissemination of information and ignores the bottom-up aggregation of information, this paper proposes a sentiment analysis model based on attention and dual channel network. While expanding the dependency representation, the model uses self attention to obtain the information matrix with semantic relevance, and uses a dual channel network to combine comprehensive syntactic and semantic relevance information. The dual channel network focuses on the semantic features of top-down propagation and the structural features of bottom-up aggregation respectively. The graph convolution output in the channel will interact with the information matrix, pay attention to complement the residual, and then complete the tasks in the channel through average pooling. Finally, the final sentiment classification features are obtained by the fusion of semantic based and structure based decision-makings. The experimental results show that the accuracy and F1 value of the model are improved on three public data sets.

Key words attention mechanism; dual channel network; decision fusion; graph convolution