

面向细粒度情感分类任务的双通道分类模型

汪伟^{1,2}, 马致远^{1,3}, 李清都¹, 李想^{1,2}, 韩士洋^{1,4}

¹上海理工大学机器智能研究院, 上海

²上海理工大学光电信息与计算机工程学院, 上海

³南京大学计算机软件新技术国家重点实验室, 江苏 南京

⁴上海理工大学机械工程学院, 上海

收稿日期: 2023年1月21日; 录用日期: 2023年2月17日; 发布日期: 2023年2月24日

摘要

句法信息对情感分类任务十分重要, 使用GCN来建模这种信息有助于模型关注情感关键词。然而这类模型仅是利用基本语义信息辅助学习句法信息, 且单一地从句法依存角度捕获情感关键词, 忽略了从语义角度发掘情感关键词。另外, 此类模型过于依赖句法信息, 没考虑到使用句法提取工具对分类效果造成的负面影响。针对以上问题, 提出一种双通道分类模型。该模型利用双通道分类结构减弱对于句法信息的依赖性, 同时采用语义情感通道从语义上捕获情感关键词, 进而提升模型获取情感信息的能力。在两个常用中文情感分类数据集上的实验表明, 该模型的 $Micro_F$ 值和 $Macro_F$ 值相较于现有模型均有提升, 模型对比和消融实验验证了双通道分类结构在提升模型分类任务性能上的有效性。

关键词

双通道, 情感分类, 句法信息, 情感关键词

Dual-Channel Classification Model for Fine-Grained Sentiment Classification Task

Wei Wang^{1,2}, Zhiyuan Ma^{1,3}, Qingdu Li¹, Xiang Li^{1,2}, Shiyang Han^{1,4}

¹Institute of Machine Intelligence, University of Shanghai for Science & Technology, Shanghai

²School of Optical-Electrical & Computer Engineering, University of Shanghai for Science & Technology, Shanghai

³State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing Jiangsu

⁴School of Mechanical Engineering, University of Shanghai for Science & Technology, Shanghai

Received: Jan. 21st, 2023; accepted: Feb. 17th, 2023; published: Feb. 24th, 2023

Abstract

Syntactic information plays an important role in sentiment classification, using GCN to model the information can help the model learn sentiment keywords. However, such models only use semantic information to assist learning syntactic information, and capture sentiment keywords from the perspective of syntactic dependency, ignoring the semantic perspective. In addition, such models rely on syntactic information and do not consider the negative impact of using syntax extraction tools on classification results. Giving the aforementioned issues, a dual-channel classification model is proposed. The model uses a dual-channel classification to reduce the dependence on syntactic information, and adopts an attention mechanism to capture semantic sentiment words, thereby improving the ability of the model to obtain sentiment information. Experiments on two commonly used Chinese sentiment classification datasets show that both *Micro_F* and *Macro_F* are improved compared with existing model. Comparative and ablation experiments illustrate the effectiveness of dual-channel classification structure to improve the model's classification performance.

Keywords

Dual Channel, Sentiment Classification, Syntactic Information, Sentiment Keywords

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

文本情感分类是自然语言处理的重要研究方向。随着大数据时代的到来, 互联网上的时事和网购评论等文本数据呈爆炸式增长[1], 其中蕴含着大量带有情感色彩的文本数据。通过对这些数据挖掘和分析, 可以更好地了解民众对于公共事件的看法及购物者给予商品的评价, 有助于政府舆情管控和商户商业决策[2], 因此对文本数据进行情感分类研究具有重要意义。

2. 相关工作

文本情感分类的早期研究主要基于词典、规则以及传统机器学习方法。Bhutekar 等人[3]将文本句子拆分为单词并通过搜索情绪词典找到相应的标签类别, 然后对词标签进行分类得到情感类别。郑等人[4]聚焦于微博的文本顺序和话语结构, 将每条微博文本看作一个数据序列, 利用从数据集中挖掘的类序列规则(Class Sequence Rule, CSR)提取有效特征来支撑模型训练。然而, 在面对复杂的文本数据时, 设计词典、规则以及提取人工特征的方法耗时费力且泛化性差。

近年来, 深度学习方法成为文本情感分类的主流[5], 该方法通过自身误差在神经网络模型中反向传播来训练模型, 训练目标集中在如何获取表义丰富的文本特征。例如, 柴源[6]采用 Word2Vec 模型学习图书评论的文本特征表示。而刘等人[7]以 CNN-BiLSTM 模型为基础融合词/句级特征来增强语义信息和丰富情感特征。此外, Pei 等人[8]还将词性标注和情感极性信息融入预训练模型, 以增强文本语义中的情感信息。然而, 上述模型都是把文本当作单词序列处理, 忽略了文本的句法信息。

句法信息能够直接揭示句子各组成成分之间的依存关系, 有助于帮助情感分类模型关注情感关键词

[9]。图 1 展示了样例文本“我当时真害怕极了”对应的句法依存树，从该句法依存树可以发现“害怕”一词位于句法依存树的中心位置，其与句中所有单词都存在着依存关系，并且中心词“害怕”能够体现整个句子的情感，因此句法信息对于情感分类任务十分重要。

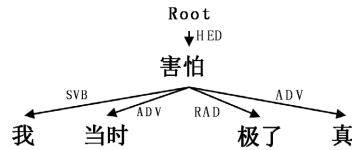


Figure 1. An example of the dependency parsing tree
图 1. 句法依存树样例

像 LSTM (Long Short-Term Memory) 这样的模型虽然能够处理文本序列信息，却无法直接处理句法信息这类树型结构数据。针对树或者图这一类数据的分析处理，Kipf 等人[10]提出了图卷积神经网络模型 (Graph Convolution Networks, GCN)。随后在情感分类任务中，以句法依存树为主要输入的 GCN 模型受到广泛关注。其中，Lai 等人[11]提出一种中文细粒度情感分类模型 Syntax-based GCN，该模型使用 GCN 处理句法信息，利用句子成分依存信息发掘能体现句子情感的情感关键词，在中文微博情感分类任务中取得了 SOTA。随后，Shou 等人[12]在 Lai 等人研究基础上，提出一种面向对话细粒度情感识别任务的分类模型 DSAGCN，该模型是在文本编码器后直接添加注意力层来增强模型学习语义信息，通过增强语义信息来帮助模型学习句法信息，进而使得模型更有效地利用句法依存信息发掘情感关键词。由此可见，使用 GCN 处理树型结构的句法信息可以从句法依存关系角度捕获情感关键词，进而有助于提高情感分类任务的性能。

使用 GCN 建模句法信息可以从句子成分依存方面获取情感关键词，然而在情感分类任务中仅使用句法信息来发现情感关键词是过于单一的，并且在该类模型中文本编码器(例如 BiLSTM)提取的上下文语义向量的主要作用是辅助 GCN 学习句法信息，忽略了进一步利用语义信息来从上下文语义角度发掘对情感分类任务起到重要作用的语义情感关键词。句子中细粒度的语义情感词能体现句子情感极性，其对情感分类任务的作用是不可忽视的。现有研究主要利用注意力机制来发掘文本中的语义情感词[13] [14]。Zhang 等人[13]使用注意力机制识别句子中多个语义情感段，并将其与目标词语义特征进行融合来增强目标词的情感信息，最后对目标词进行情感分类。Later 等人[14]提出了自注意力网络 SANet，在情感分类实验中验证了 SANet 可以通过相邻单词交互发现语义情感词。由此启发，本文在引入 GCN 来建模句法信息的同时增加一条使用注意力机制的通道，使得模型具有关注语义情感关键词的能力。文本使用自注意力机制从上下文语义角度发掘对情感分类任务贡献大的词，并赋予较高权重，最后相加加权后的所有词向量作为句子情感向量表示，进而增强模型对语义信息的挖掘。

另外由于标注数据的成本高，现有中文情感数据集缺少句法标注信息，所以在现有研究[15] [16] [17] 中，依存句法树的建立主要通过使用开源工具(例如：LTP [18], Spacy 等)得到，这些工具的使用无法确保任何时候建立的句法分析结果的准确性，而错误的句法信息将最终影响模型的性能。基于以上研究现状，本文提出一种面向细粒度情感分类任务的双通道分类模型(DCGCSA, Dual Channels of Graph Convolution and Self-Attention)，在从句法依存和上下文语义两个角度发掘情感关键词的过程中，尽可能降低错误句法信息对情感分类任务的负面影响。该模型是双通道分类结构，一条通道侧重于建模情感分类相关的句法信息，另一条通道侧重于从上下文语义信息中筛选出关键的语义情感词，最终分类结果通过一个权重系数来权衡。结合双通道分类结果能使模型分类性能不取决于单一的某个通道性能，从而在从两个方面获取情感关键词信息同时减弱开源工具带来的负面影响。本文贡献有以下三个方面：

1) 提出一种双通道分类模型 DCGCSA, 该模型通过双通道机制不仅可以依据句法依存和上下文语义捕获情感关键词来提高分类性能, 还可以使用双通道机制减弱错误的句法信息带来的影响。

2) 增加一条使用注意力机制的通道。模型不仅可以利用编码器获得的基本语义向量来帮助学习句法信息, 还能够更加充分利用基本语义向量关注对情感分类任务贡献大的语义情感关键词, 增强模型发掘情感信息的能力。

3) 引入通道权衡系数, 优化了双通道机制, 权衡通道对情感分类结果的贡献, 使得模型获取更多有效的情感信息。

3. 双通道分类模型

由于 GCN 对计算机算力要求较高, 考虑设备性能和模型训练时间, DCGCSA 采用算力要求不高的 BiLSTM 作为编码器, 通过共享编码器句法情感关键词通道网络(Syntactic Dependency Sentiment Keywords Channel Networks, SDSK)和语义情感关键词通道网络(Syntactic Sentiment Keywords Channel Networks, SSK)的双通道分类, 总体框架如图 2 所示。

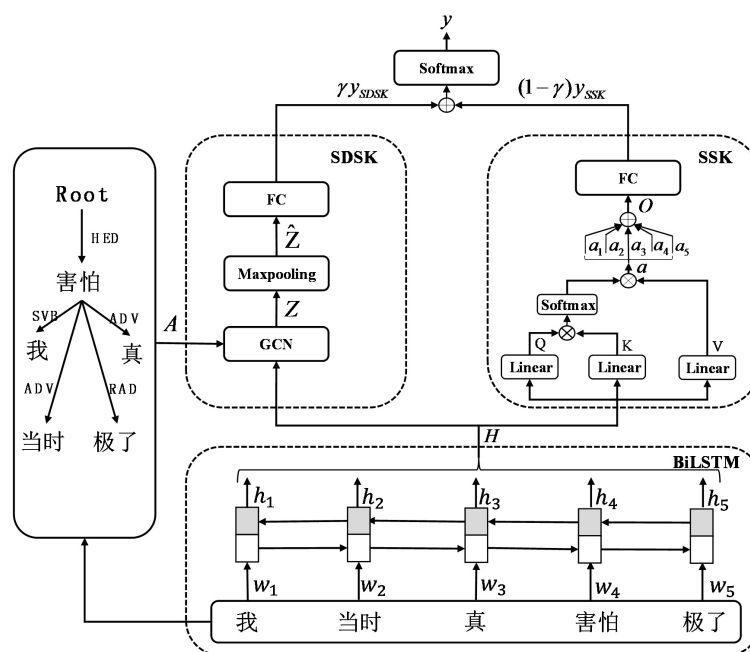


Figure 2. The framework of DCGCSA

图 2. DCGCSA 模型

3.1. 文本编码器

BiLSTM 是 RNN (Recurrent Neural Networks) 的一种, 它由双向 LSTM 构成, 从正向和反向两个角度更好地学习文本句子的上下文语义信息。本文使用 BiLSTM 作为文本编码器, 提取文本语义信息。对于给定文本句子序列 $X = [w_1, \dots, w_i, \dots, w_{|X|}]$, BiLSTM 输出的结果是包含语义信息的特征向量公式(1):

$$H = [h_1, \dots, h_i, \dots, h_{|X|}] = \text{BiLSTM}([w_1, \dots, w_i, \dots, w_{|X|}]) \quad (1)$$

其中 w_i 表示文本句子序列中的每个词, $|X|$ 表示文本句子序列的长度, $h_i \in R$ 表示编码器生成的词特征向量, d 为词嵌入维度, $H \in R^{|X| \times d}$ 表示编码器生成的句子的特征向量。

3.2. 句法情感关键词通道网络

对于情感分类任务, 句法依存树能在句法依存关系上关注到对任务重要的词语。句法依存树的节点为词语, 节点与节点的连接表示词与词之间的依存关系。模型针对句法依存树构建一张图 $Graph = (V, E)$ 。其中, V 是句法依存树中所有词组成的节点集, E 是包含两个词之间所有依存关系的边集。节点与文本特征向量 $[h_1, \dots, h_i, \dots, h_{|x|}]$ 存在一一对应关系。根据 Kipf 等人[10]和 Marcheggiani 等人[19]的研究, 在依存关系中增加词自身依存和反向依存可以增强 GCN 的泛化能力。因此使用“0”、“1”、“1”、“1”分别标记无依存关系、自身依存关系、头到依存和依存到头的依存关系类型。依存关系通常以邻接矩阵形式输入模型, 根据上述规则, 生成句法邻接矩阵 A (图 3)。

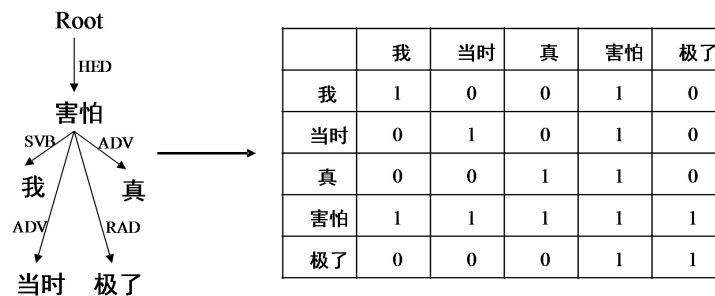


Figure 3. An example of the dependency parsing tree and adjacent matrix
图 3. 句法依存树及其邻接矩阵示例

之后 GCN 利用句法邻接矩阵 A 获取具有文本深层情感信息的句子特征 Z , 见公式(2):

$$Z = \text{ReLU} \left(\tilde{D}^{-\frac{1}{2}} A \tilde{D}^{-\frac{1}{2}} H \theta \right) \quad (2)$$

其中, ReLU 代表激活函数, \tilde{D} 为 A 的度矩阵, θ 为 GCN 神经网络的训练参数。

为了减少计算量和防止过拟合现象, 将句子特征表示 Z 经过最大池化层得到 \hat{Z} 。再将 \hat{Z} 输入全连接层得到情感标签 y_{SDSK} (公式(3))。其中, $\hat{Z} \in R^{l \times d}$, $y_{\text{SDSK}} \in R^{M \times 1}$:

$$\hat{Z} = \text{Maxpooling}(Z) \quad (3)$$

$$y_{\text{SDSK}} = W_{\text{SDSK}} \hat{Z} + b_{\text{SDSK}}$$

其中, M 为预测情感类别的类别数量, $W_{\text{SDSK}} \in R^{M \times d}$ 为全连接层的权重参数, $b_{\text{SDSK}} \in R^{M \times 1}$ 为全连接层的偏置。

3.3. 语义情感关键词通道网络

自注意力机制利用上下文语义信息使得文本序列中的每个单词关注序列中的其他单词, 通过单词与单词之间的交互赋予含有情感信息的单词以其较高的注意力权重。对文本序列的每个文本符号, 自注意力机制首先获取对应位置的查询矩阵 Q 、键矩阵 K 、值矩阵 V , 见公式(4), 其中, W^Q 、 W^K 、 $W^V \in R^{d \times d}$:

$$Q = HW^Q, K = HW^K, V = HW^V \quad (4)$$

利用查询矩阵 Q 和键矩阵 K 计算单词与单词相互之间的注意力权重矩阵, 并将其加权到值矩阵 V (公式(5)):

$$a = [a_1, a_2, \dots, a_{|x|}] = \text{Attention}(Q, K, V) = \text{Softmax} \left(\frac{QK^T}{\sqrt{d}} \right) V \quad (5)$$

加权词嵌入相加得到句子语义情感嵌入(公式(6)), 其中 $a_i \in R^{d \times 1}$, $O \in R^{d \times 1}$:

$$O = a_1 + a_2 + \dots + a_{|X|} \quad (6)$$

最后将 O 输入至全连接层得到情感标签 y_{SSK} (公式(7)), 其中 $y_{SSK} \in R^{M \times 1}$:

$$y_{SSK} = W_{SSK} O + b_{SSK} \quad (7)$$

其中, $W_{SSK} \in R^{M \times d}$, $b_{SSK} \in R^{M \times 1}$ 分别为全连接层的权重和偏置。

3.4. 损失函数

为了使 SDSK 和 SSK 联合且有效发挥作用, 将两个通道的输出相加并过 Softmax 得到模型预测的最终标签 y (公式(8))。其中, γ 是通道的权衡系数。

$$y = \text{Softmax}(\gamma y_{SDSK} + (1-\gamma) y_{SSK}) \quad (8)$$

DCGCSA 使用交叉熵损失来优化 SDSK 和 SSK 的参数。为了控制梯度消失和梯度爆炸问题, DCGCSA 使用随机正交矩阵初始化 BiLSTM 的权重矩阵, 并在损失函数中加入正交化约束(公式(9)):

$$Loss = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic}) + \lambda \sum_i \|W_i^T W_i - I\|^2 \quad (9)$$

其中, N 是样本数量, i 代表样本索引, c 代表情感类别标签, $c \in M$ 。 y_{ic} 是符号函数, 如果索引为 i 样本的真实类别为 c 取 1, 否则取 0。 p_{ic} 表示预测索引为 i 样本属于类别 c 的概率, λ 是正则化系数, I 是单位矩阵, W 为 BiLSTM 的权重矩阵, 通过对随机矩阵进行奇异值分解, 获得的正交或近正交矩阵初始化 W 。

3.5. 伪代码

为了更好地说明 DCGCSA 模型的实现细节, 本小节展示模型训练的伪代码如图 4 所示。

```

算法 1 面向细粒度情感分类任务的双通道分类模型
输入:  $L$ : 标注数据;  $M$ : 输出情感标签候选集合;  $E$ : 训练轮数;  $B$ : 批处理大小;
输出:  $model_{min}$ : 在测试集上的最小损失  $Loss_{test-min}$  模型;
初始化: 一个编码器:  $BiLSTM()$ ; 一个图卷积神经网络:  $GCN()$ ;
          一个句法信息提取工具  $LTP()$ ; 全连接层网络  $FC()$ ;
for  $epoch = 1 \rightarrow E$  do
    按照批处理大小  $B$  加载标注数据  $L$  作为  $Batch$ ;
    for 文本序列  $X$  in  $Batch$  do
      句法邻接矩阵  $A \leftarrow LTP(X)$ ;
      for 词  $w_i \in X$  do
         $h_i \leftarrow BiLSTM(w_i)$ ;
      end for
       $H \leftarrow [h_1, \dots, h_i, \dots, h_{|X|}]$ ;
       $Z \leftarrow GCN(H, A)$ ;
       $y_{SDSK} \leftarrow FC(\text{maxpool}(Z))$ ;
       $Q, K, V \leftarrow HW^Q, HW^K, HW^V$ ;
       $[a_1, \dots, a_i, \dots, a_{|X|}] \leftarrow \text{Softmax}(\frac{QK^T}{\sqrt{d}}) \cdot V$ ;
       $O \leftarrow a_1 + \dots + a_i + \dots + a_{|X|}$ ;
       $y_{SSK} \leftarrow FC(O)$ ;
       $y \leftarrow \gamma \cdot y_{SDSK} + (1-\gamma) \cdot y_{SSK}$ ;
    end for
    if  $Loss_{test} < Loss_{test-min}$  then
       $Loss_{test-min} \leftarrow Loss_{test}$ ;
       $model_{min} \leftarrow model_{test}$ ;
    end if
end for
return  $model_{min}$ 

```

Figure 4. Pseudo code diagram of DCGCSA
图 4. DCGCSA 伪代码示意图

4. 实验设计及结果分析

4.1. 数据集及性能指标

本文实验采用 Lai 等人[11]文献提供的数据集和 SMP2020 数据集, Lai [11]数据集由公开数据集中文计算会议测评的微博情感数据 NLPCC2013 和微博爬取数据构成, SMP2020 数据集来自公开数据集第九届全国社交媒体处理大会会议测评的微博情绪分类数据集。实验中除去 SMP2020 数据集中的中性标签(无情感标签), 并按照官方给定的训练集和测试集作为实验的训练集和测试集。表 1 是数据集 Lai [11]和 SMP2020 的统计信息, 数据集 Lai [11]包含 7 类情感标签, 数据集 SMP2020 包含 5 类情感标签, 其中带有 N/A 标号的表示数据集缺失该类别情感标签:

Table 1. The statistics of the datasets

表 1. 数据集统计信息

情感类别	Lai [12]		SMP2020	
	训练	测试	训练	测试
恐惧	395	49	1220	210
愤怒	2256	235	8345	1508
失望	3943	425	N/A	N/A
悲伤	3100	385	4990	900
快乐	3737	370	N/A	N/A
喜爱	5541	595	N/A	N/A
惊讶	1030	113	2086	374
积极	N/A	N/A	5378	1018
总计	20002	2172	22019	4010

在细粒度情感分类性能评估中, 采用常用的 $Micro_F$ 值和 $Macro_F$ 值作为评价指标。 $Micro_F$ 值反映模型对各个情感类别的分类性能, $Macro_F$ 值反映模型对数据集整体的分类性能。

4.2. 实验配置及超参数设置

本文实验配置: 操作系统 Ubuntu16.04, CUDA 10.2, 编辑环境 Python3.7.9, 深度学习框架 Pytorch1.4.0, CPU 为 Intel(R) Core(TM) i9-9900k CPU @ 3.60 GHz, GPU 为 GeForce RTX 2080Ti。 DCGCSA 的超参数设置如表 2 所示:

Table 2. The hyper-parameters of DCGCSA

表 2. DCGCSA 模型超参数

参数名称	参数值	参数名称	参数值
训练批大小	128	自注意力键嵌入维度	360
词嵌入维度	300	自注意力值嵌入维度	360
BiLSTM 嵌入维度	360	初始化学习率	0.001
GCN 层数	1	权重衰减	$1e^{-8}$

除表 2 中参数设置外, 本次实验还需要调优通道权衡系数 γ 和正交化约束系数 λ 。调优过程: 先令 $\lambda = 0$, 调整 γ (γ 的取值从 0 到 1, 步长为 0.1), 然后修改 λ (λ 的取值从 0 到 5×10^{-9} , 步长为 10 的负幂次方), 再次调整 γ 。经过参数调优, 确定在不同数据集中最优通道权衡系数 γ 和最优正则化系数 λ : 在 $\lambda = 5 \times 10^{-8}$ 、 $\gamma = 0.4$ 时, 使得 DCGCSA 在数据集 Lai [11] 上的性能最优; 在 $\lambda = 5 \times 10^{-8}$ 、 $\gamma = 0.5$ 时, 使得 DCGCSA 在数据集 SMP2020 上的性能最优。

4.3. 实验流程

数据预处理: 数据集数据来自微博, 在实验前要对数据集进行繁体转简体、去除无用 URL 和邮箱等简单地数据清洗和处理。由于微博数据中存在大量的表情包, 而表情包可以明显表达句子的情感, 因此为了融入表情包的信息, 实验前将数据集中的表情包转换成描述相应表情包的文字。

句法依存树的生成: 为了提升模型训练效率, 句法依存树独立于模型, 在模型训练前使用哈工大工具 LTP [18] 生成。生成的句法依存树以邻接矩阵形式保存成序列文件, 当进行模型训练时, 直接加载序列文件即可。

模型训练: 每次训练设置共 200 轮次训练, 取 best 模型; 每个数据集训练 3 次, 取 3 次的 *Micro_F* 和 *Macro_F* 的平均值作为最终实验结果。

4.4. 模型对比实验

本节为了评估 DCGCSA 性能, 对比了该领域相关研究中的模型, 其中带有*标的对比算法实验结果直接引用原文献。为了更准确地评估 DCGCSA 性能, 增加了在数据集 SMP2020 的对比试验, 其中 N/A 代表未获得实验结果。

LSTM: Ji 等人[20]使用 LSTM 最后一层输出向量作整个句子的语义表示, 没有考虑到上下文语义信息。

LSTM-CNN: Lai [11]等人使用 BiLSTM 作为编码器, 提取文本序列的语义表示, 再利用 CNN 整合文本序列的语义表示得到整个句子的语义表示, 该方法考虑到了上下文语义信息, 但是保留了冗余的语义信息。

BiLSTM-Self-Attention: 使用 BiLSTM 作为编码器, 从正反两个方向获取文本序列的语义表示, 再使用自注意力机制赋予上下文中具有重要语义信息的词更高的权重。

Syntax-based GCN: Lai 等人[11]融合句法信息, 利用句法信息挖掘重要的情感关键词, 提出一种结合 GCN 和 BiLSTM 的模型。该模型过于依赖句法依存信息, 且仅使用句法信息发掘情感关键词。另外该方法忽略语义信息的重要性, 只是用语义信息来帮助模型学习句法信息。

DCGCSA 分类性能是在 γ 和 λ 取最优值时测试的($\lambda = 5 \times 10^{-8}$, 在 Lai [11]中 $\gamma = 0.4$, 在 SMP2020 中 $\gamma = 0.5$)。对比试验结果如表 3 所示, DCGCSA 在两个数据集上都取到了最优的效果。

Table 3. Results of model comparative experiment

表 3. 模型对比实验结果

模型	Lai [12]		SMP2020	
	<i>Micro_F</i>	<i>Macro_F</i>	<i>Micro_F</i>	<i>Macro_F</i>
LSTM*	73.42%	68.32%	N/A	N/A
LSTM-CNN*	76.42%	69.89%	N/A	N/A
BiLSTM-Self-Attention	81.63%	78.25%	68.10%	64.20%
Syntax-based GCN*	82.32%	79.93%	68.05%	64.40%
DCGCSA	84.48%	83.00%	69.57%	65.77%

注: 表中黑体字表示最优的分类性能分数。

对比 BiLSTM-Self-Attention 和 LSTM-CNN 可以发现：加入自注意力，可以有效去除冗余的语义信息，使得模型从语义角度捕获情感关键词。对比 DCGCSA、Syntax-based GCN 和 BiLSTM-Self-Attention 的实验结果可以看到，从句法和语义方面进行情感分类的性能差别不大，两者对于情感分类任务都很重要，且可以相互补充，从句法依存和上下文语义双角度捕获情感信息的分类性能(DCGCSA)要优于单角度的分类性能。

通过对比模型在两个数据集的分类性能表现，可以发现 DCGCSA 在数据集 Lai [11] 的表现要优于数据集 SMP2020，为了探究产生这种差异的原因，观察两个数据集的训练样本和测试样本的数量比例，发现数据集 Lai [11] 的比例要更大，这是主要原因之一；另外处理数据集 SMP2020 不够充分也是导致这种差异性的主要原因之一。

4.5. 通道权衡系数 γ

为了探究通道权衡系数 γ 对模型性能的影响，本节保持正则化系数 $\lambda = 5 \times 10^{-8}$ 不变，通过改变 γ 取值，记录模型 DCGCSA 的 *Micro_F* 值和 *Macro_F* 值。实验结果绘制成折线图，见图 5。

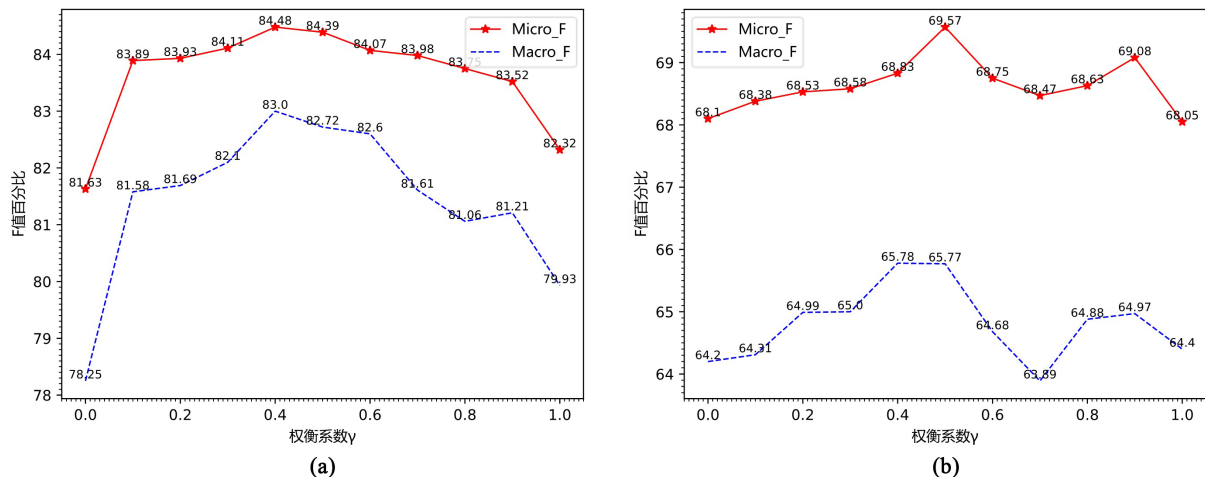


Figure 5. Channel trade factor action chart

图 5. 通道权衡系数作用图

折线图的横轴为权衡系数 γ ，纵轴为 *Micro_F* 和 *Macro_F* 百分比值 (*F* 百分比值)，图 5(a) 是在数据集 Lai [11] 的实验结果，图 5(b) 是在数据集 SMP2020 的实验结果。由图 5(a)、图 5(b) 可知，随着权衡系数 γ 取值增大，DCGCSA 在数据集 Lai [12] 和 SMP2020 的 *F* 百分比值呈现先增加后减小的趋势。在数据集 Lai [11] 中，模型 DCGCSA 在 $\gamma = 0.4$ 时取得最优的分类性能；在数据集 SMP2020 中，模型 DCGCSA 在 $\gamma = 0.5$ 时为最优分类性能，而当 $\gamma = 0.7$ 时，模型的 *Macro_F* 值要低于只有 GCN 的 *Macro_F* 值，出现这种情况的原因可能在于此时 DCGCSA 在某一类别的分类效果比较差，导致整个数据集的性能指标 *Macro_F* 值较低。从图中的变化趋势可以得到模型的性能随着权衡系数 γ 取值改变而变化，说明改变权衡系数 γ 相当于改变 SDSK 和 SSK 对最终分类结果的贡献度，调整权衡系数 γ 可以改变 SDSK 和 SSK 对最终分类结果的影响力度。另外，权衡系数 γ 是 SDSK 的通道贡献系数，而 SSK 的通道贡献系数为 $1 - \gamma$ ，所以在模型 DCGCSA 中 SDSK 和 SSK 对模型的影响属于竞争关系，在减小权衡系数 γ 值，减弱 SDSK 的影响力度时会增强 SSK 的影响力度。当减弱 SDSK 对模型的影响力度时，相当于减弱句法信息对分类的影响，进而减弱开源工具对模型造成的负面影响；而由于 SDSK 和 SSK 存在竞争关系，模型 DCGCSA

会加强 SSK 对最终分类结果的影响, 进而模型 DCGCSA 更加注重从语义角度捕获情感关键词信息。

图 5 中的变化趋势表明双通道分类结构可以让模型不过于依赖句法信息, 减弱使用开源工具造成的负面影响; 并且可以通过调整的权衡系数 γ 取值, 使得模型使用句法提取开源工具带来的负面影响最小化的同时, 最有效综合两个角度获取的情感信息, 进而提高模型的分分类效果。实验结果表明, 使用通道权衡系数可以充分发挥双通道分类的作用, 由于数据集的差异会使得最优权衡系数 γ 的取值不同。

4.6. 正交化约束系数 λ

本节探究正交化约束系数 λ 对模型性能的影响, 在数据集 Lai [11] 中取权衡系数 γ 为 0.4, 在数据集 SMP2020 中取权衡系数 γ 为 0.5, 使得模型的权衡系数处于最优。改变正交化约束系数 λ 的取值(从 0 开始, 每次步长为 10 的负幂次方), 记录模型 DCGCSA 每次 λ 取值的 *Micro_F* 值和 *Macro_F* 值, 结果见表 4。

Table 4. Comparison results of penalty coefficient λ

表 4. 正则化系数 λ 对比结果

惩罚系数 λ	Lai [11]		SMP2020	
	<i>Micro_F</i>	<i>Macro_F</i>	<i>Micro_F</i>	<i>Macro_F</i>
5×10^{-6}	83.83%	82.64%	68.60%	64.82%
5×10^{-7}	83.98%	82.54%	69.37%	65.66%
5×10^{-8}	84.48%	83.00%	69.57%	65.77%
5×10^{-9}	82.92%	81.84%	68.83%	65.00%
0	82.15%	80.05%	68.08%	63.89%

注: 表中黑体字表示最优的分类性能分数, 正则化系数 λ 取 0 表示无正则化约束。

由表 4 可知, 在正则化系数 λ 为 5×10^{-8} 时, 模型 DCGCSA 在数据集 Lai [11] 和数据集 SMP2020 的分类性能取得最优。相较于无正则化约束的 DCGCSA 的表现, 最优正则化系数 ($\lambda = 5 \times 10^{-8}$) 的 DCGCSA 在数据集 Lai [11] 中的 *Micro_F* 值和 *Macro_F* 值分别提高了 2.33% 和 2.95%, 在数据集 SMP2020 中的 *Micro_F* 值和 *Macro_F* 值分别提高了 1.49% 和 1.88%。从最优正则化系数 DCGCSA 分类性能的提升, 可以得出对 BiLSTM 的权重矩阵 W 加正则化可有助于模型 DCGCSA 分类性能的提升。

实验结果表明, 加入正交化约束的 DCGCSA 的性能要优于无正交化约束的 DCGCSA。

4.7. 注意力可视化

为了验证引入自注意力机制的 DCGCSA 具有捕获重要语义情感词的能力。本节使用最优 DCGCSA 对数据样本实例进行注意力可视化(最优 DCGCSA 在数据集 Lai [11] 和数据集 SMP2020 中取最优权衡系数 γ 和最优正则化系数 λ)。按词注意力进行汇总, 并进行规范化处理, 如图 6 所示。

图 6 中的热力图结果是以句进行展示, 每个词都对应着其注意力分数, 分数表示模型 DCGCSA 对词的关注程度, 分数值越大, 关注度越高, 分数的颜色越深。图 6(a) 样例的真实情感标签为“积极”, 图 6(b) 样例的真实情感标签为“恐惧”。

模型 DCGCSA 预测图 6(a) 样例的情感标签为“积极”, 从可视化结果可看出, DCGCSA 在图 6(a) 样例中对“快乐”、“开开心心”和“生日”的关注分数为 0.142、0.107 和 0.048, 要远高于样例中的其

他词,说明模型 DCGCSA 可以捕获“开开心心”、和“快乐”等含有积极语义的情感关键词; DCGCSA 预测图 6(b)样例的情感标签为“恐惧”,对“恐惧”、“吓醒”和“噩梦”的关注分数为 0.208、0.089 和 0.071,说明模型 DCGCSA 可以捕获“恐惧”、“吓醒”和“噩梦”等含有恐惧语义的情感关键词。从样例可视化结果可以得出,加入语义情感关键词通道的 DCGCSA 可以加深利用本文编码器获得的基本语义信息,从上下文语义角度捕获语义情感关键词,从而提高模型的最终分类性能。

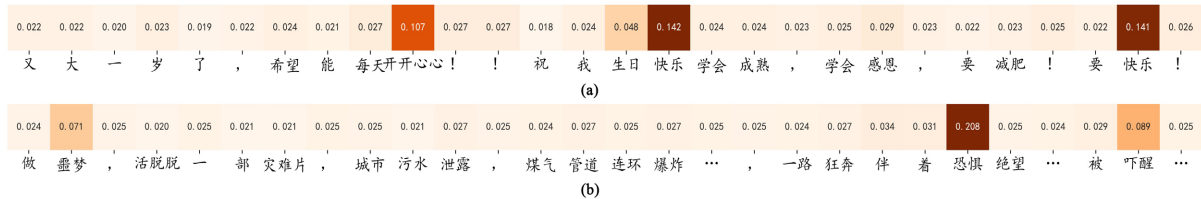


Figure 6. Weight distribution thermodynamic diagram of word attention summary in the sample
 图 6. 样例中词注意力汇总的权重热力分布图

实验结果表明,引入自注意力机制的 DCGCSA 可以筛选出体现句子情感句子语义情感词,赋予这些词更高的权重,从而减少冗余无用的语义信息。

4.8. 消融实验

本节通过设置消融实验进一步验证双通道分类有效性,实验设置模型 DCGCSA 的正则化系数 λ 取 5×10^{-8} 。本节消融实验分别单独去除 GCCN 和 SCN,探究只保留一条通道时的 DCGCSA 的分类性能,实验结果见图 5。

当 $\gamma = 0$ 时,模型 DCGCSA 只存在 SSK。此时模型 DCGCSA 在数据集 Lai [11] 的 *Micro_F* 值和 *Macro_F* 值为 81.62% 和 78.25%,在数据集 SMP2020 的 *Micro_F* 值和 *Macro_F* 值为 68.10% 和 64.20%。而与单 SSK 的 DCGCSA 相比,最优双通道 DCGCSA 在数据集 Lai [11] 中的 *Micro_F* 值和 *Macro_F* 值提升了 2.85% 和 4.75%,在数据集 SMP2020 中的 *Micro_F* 值和 *Macro_F* 值提升了 1.47% 和 1.57%。

当 $\gamma = 1$ 时,DCGCSA 只存在 SDSK。与此时单 SDSK 的 DCGCSA 相比,最优双通道分类的 DCGCSA 在数据集 Lai [11] 中的 *Micro_F* 值和 *Macro_F* 值为 84.48% 和 83.00%,分别提升了 2.16% 和 3.07%;在数据集 SMP2020 中的 *Micro_F* 值和 *Macro_F* 值为 69.57% 和 65.77%,分别提升了 1.52% 和 1.37%。

通过消融实验结果可知,双通道分类机制对模型 DCGCSA 性能提升起到了正向作用,使得模型有效结合了 SDSK 和 SSK 的分类结果,并在最优权衡系数 γ 的调整下,使得 SDSK 和 SSK 联合发挥的作用尽量最大化。

4.9. 混淆矩阵可视化

为了探究模型 DCGCSA 在具体情感类别中的分类性能表现,本节对混淆矩阵结果进行可视化,如图 7 所示。混淆矩阵可视化可以清晰地表现出模型对具体情感类别分类的正确性以及错分后的类别。图中纵轴为样本真实标签类别,横轴为最优 DCGCSA 对该情感类别样本的预测标签,图中的颜色越深则表示权重越大。图 7(a)和图 7(b)分别展示了模型 DCGCSA 在数据集 Lai [11] 和数据集 SMP2020 的混淆矩阵可视化结果。

从图 7(a)可得,深色都出现在对角线上,说明模型 DCGCSA 在数据集 Lai [11] 中能够较好地识别具体情感类别,能够将各个具体情感类别的样例较大概率的预测为正确情感标签。从图 7(b)可得,模型 DCGCSA 在数据集 SMP2020 中能够较好地识别“愤怒”、“悲伤”和“积极”情感类别,能够将 74%

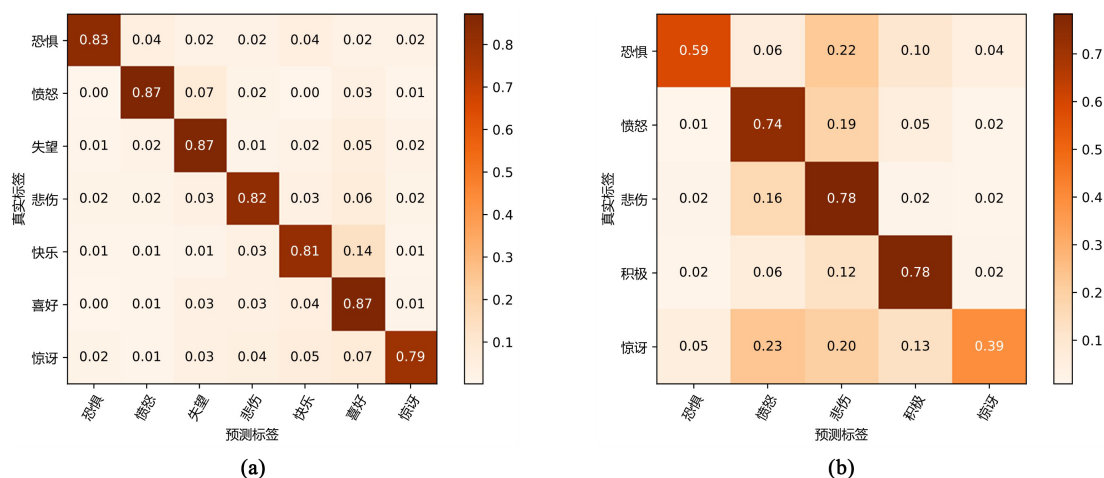


Figure 7. Results of confusion matrix visualization

图 7. 混淆矩阵可视化结果

的“愤怒”样例、78%的“悲伤”样例以及78%的“积极”样例预测为真实标签，但是在“恐惧”和“惊讶”情感类别中的表现较差。22%的“恐惧”情感样例被错误分类为“悲伤”标签，10%的“恐惧”情感样例被错误分类为“积极”标签；23%的“惊讶”情感样例被错误分类为“愤怒”标签，20%的“惊讶”情感样例被错误分类为“悲伤”标签，13%的“惊讶”情感样例被错误分为“积极”标签，造成这现象的原因可能在于数据集 SMP2020 中“恐惧”和“惊讶”情感样本数较少。

5. 结束语

本文提出了一种双通道细粒度情感分类模型 DCGCSA，该模型从句法依存和上下文语义两个角度捕获情感关键词的同时减弱了错误句法信息给模型带来的影响，并通过调优通道权衡系数来增强这种效果。模型对比实验结果表明，捕获语义情感关键词在情感分类任务中具有重要作用，不过分依赖句法信息并能捕获语义情感词的 DCGCSA 的分类性能要优于对比模型。消融实验结果验证了双通道分类机制的有效性。未来研究可以尝试改进模型自动学习通道权衡系数。另外，DCGCSA 的 SDK 只采用了简单的位置映射关系来引入依存句法信息，无法自动筛选错误的句法信息，在未来的工作中可以设计相应机制筛选错误句法信息。

基金项目

国家自然科学基金重点项目(92048205)；南京大学计算机软件新技术国家重点实验室开放课题项目(KFKT2021B39)。

参考文献

- [1] 齐嵩喆, 黄贤英, 朱小飞. 基于权重增强的方面级情感分析模型[J]. 小型微型计算机系统, 2022, 43(4): 747-753.
- [2] 李然, 林政, 林海伦, 等. 文本情绪分析综述[J]. 计算机研究与发展, 2018, 55(1): 30-52.
- [3] Bhutekar, S.D., Chandak, M.B. and Agrawal, A.J. (2012) Emotion Extraction: Machine Learning for Text-Based Emotion. *Proceedings on National Conference on Recent Trends in Computing*, Vol. 1, 20-23.
- [4] 郑诚, 沈磊, 代宁. 基于类序列规则的中文微博情感分类[J]. 计算机工程, 2016, 42(2): 184-189, 194.
- [5] Zhang, L., Wang, S. and Liu, B. (2018) Deep Learning for Sentiment Analysis: A Survey. *WIREs Data Mining and Knowledge Discovery*, 8, e1253. <https://doi.org/10.1002/widm.1253>
- [6] 柴源. 基于 Word2vec 和 SVM 的在线图书评论情感识别系统实现[J]. 电子设计工程, 2022, 30(6): 179-183.

- [7] 刘鑫, 祁瑞华, 徐琳宏, 等. 融合多级特征的俄语推特文本情感分析[J]. 小型微型计算机系统, 2021, 42(6): 1176-1183.
- [8] Ke, P., Ji, H., Liu, S., *et al.* (2020) SentiLARE: Sentiment-Aware Language Representation Learning with Linguistic Knowledge. *Proceedings on Empirical Methods in Natural Language Processing*, November 2020, 6975-6988. <https://doi.org/10.18653/v1/2020.emnlp-main.567>
- [9] Huang, B.-X. and Kathleen, C. (2019) Syntaxaware Aspect Level Sentiment Classification with Graph Attention Networks. *Proceedings on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing*, Hong Kong, November 2019, 5469-5477. <https://doi.org/10.18653/v1/D19-1549>
- [10] Kipf, T.N. and Welling, M. (2017) Semi-Supervised Classification with Graph Convolutional Networks. *Proceedings on Learning Representations*, Toulon.
- [11] Lai, Y., Zhang, L., Han, D., *et al.* (2020) Fine-Grained Emotion Classification of Chinese Microblogs Based on Graph Convolution Networks. *World Wide Web*, **23**, 2771-2787. <https://doi.org/10.1007/s11280-020-00803-0>
- [12] Shou, Y.-T., Meng, T., Ai, W., *et al.* (2022) Conversational Emotion Recognition Studies Based on Graph Convolutional Neural Networks and a Dependent Syntactic Analysis. *Neurocomputing*, **501**, 629-639. <https://doi.org/10.1016/j.neucom.2022.06.072>
- [13] Zhang, J., Chen, C., Liu, P., *et al.* (2020) Target-Guided Structured Attention Network for Target-Dependent Sentiment Analysis. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Cambridge, 5-10 July 2020, 172-182. https://doi.org/10.1162/tacl_a_00308
- [14] Letarte, G., Paradis, F., Giguère, P., *et al.* (2018) Importance of Self-Attention for Sentiment Analysis. *Proceedings on Empirical Methods in Natural Language Processing*, Brussels, 31 October-4 November 2018, 267-275. <https://doi.org/10.18653/v1/W18-5429>
- [15] 张文轩, 殷雁君. 基于依存树增强注意力模型的方面级情感分析[J]. 计算机应用研究, 2022, 39(6): 1656-1662.
- [16] Dai, J., Yan, H., Sun, T., *et al.* (2021) Does Syntax Matter? A Strong Baseline for Aspect-Based Sentiment Analysis with RoBERTa. *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online, June 2021, 1816-1829. <https://doi.org/10.18653/v1/2021.naacl-main.146>
- [17] Dai, A., Hu, X., Nie, J., *et al.* (2022) Learning from Word Semantics to Sentence Syntax by Graph Convolutional Networks for Aspect-Based Sentiment Analysis. *International Journal of Data Science and Analytics*, **14**, 17-26. <https://doi.org/10.1007/s41060-022-00315-2>
- [18] Che, W., Li, Z. and Liu, T. (2010) LTP: A Chinese Language Technology Platform. *Proceedings on Computational Linguistics*, Beijing, August 2010, 13-16.
- [19] Marcheggiani, D. and Titov, I. (2017) Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling. *Proceedings on Empirical Methods in Natural Language Processing*, Copenhagen, September 2017, 1506-1515. <https://doi.org/10.18653/v1/D17-1159>
- [20] Ji, Y.L. and Deroncourt, F. (2016) Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks. *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, June 2016, 515-520.