

标签指导的双注意力深度神经网络模型

彭展望¹ 朱小飞¹ 郭嘉丰²

摘要 在数据集不包含标签文本信息时,现有的显式交互分类模型无法显式计算文本单词和标签之间的语义关系.针对此问题,文中提出标签指导的双注意力深度神经网络模型.首先,提出基于逆标签频次的自动类别标签描述生成方法,为每个标签生成特定的标签描述,用于显式计算文本单词与标签之间的语义关系.在此基础上,使用文本编码器学习具有上下文语境信息的评论文本表示,并提出标签指导的双注意力网络,分别学习基于自注意力的文本表示和基于标签注意力的文本表示.然后,使用自适应门控机制融合这两个文本表示,得到文本最终表示.最后,使用两层前馈神经网络作为分类器,进行情感分类.在 3 个公开的真实数据集上的实验表明,文中模型分类效果较优,可减少计算代价和训练时长.

关键词 情感分类, 标签描述生成, 双注意力, 自注意力, 标签注意力

引用格式 彭展望,朱小飞,郭嘉丰. 标签指导的双注意力深度神经网络模型. 模式识别与人工智能, 2022, 35(2): 175-184.

DOI 10.16451/j.cnki.issn1003-6059.202202008

中图法分类号 TP 391

Label-Guided Dual-Attention Deep Neural Network Model

PENG Zhanwang¹, ZHU Xiaofei¹, GUO Jiafeng²

ABSTRACT Since the text information of labels is not included in some datasets, the semantic relationship between text words and labels cannot be explicitly calculated in the existing explicit interactive classification models. To solve this problem, a label-guided dual-attention deep neural network model is proposed in this paper. Firstly, an automatic category label description generation method based on inverse label frequency is proposed. According to the label description generation method, a specific label description for each label is generated. The generated specific label description is applied to explicitly calculate the semantic relationship between text words and labels. On the basis of the above, review text representation with contextual information is learned by a text encoder. A label-guided dual-attention network is proposed to learn the text representation based on self-attention and the text representation based on label attention, respectively. Then, an adaptive gating mechanism is employed to fuse two mentioned text representations and the final text representation is thus obtained. Finally, a two-layer feedforward neural network is utilized as a classifier for sentiment classification. Experiments on three publicly available real-world datasets show that the proposed model produces better classification performance.

收稿日期:2021-10-25;录用日期:2021-12-27

Manuscript received October 25, 2021;

accepted December 27, 2021

国家自然科学基金项目(No. 62141201)、重庆市技术创新与应用发展专项项目(No. cstc2020jscx-dxwtBX0014)、重庆市教委语言文字科研重点项目(No. yyk20103)资助

Supported by National Natural Science Foundation of China(No. 62141201), Special Project of Chongqing Technological Innovation and Application Development(No. cstc2020jscx-dxwtBX0014), Key Research Project on Language and Text of Chongqing

Municipal Education Commission(No. yyk20103)

本文责任编辑 陈恩红

Recommended by Associate Editor CHEN Enhong

1. 重庆理工大学 计算机科学与工程学院 重庆 400054

2. 中国科学院计算技术研究所 网络数据科学与技术重点实验室 北京 100190

1. College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054

2. Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190

Key Words Sentiment Classification, Label Description Generation, Dual-Attention, Self-Attention, Label Attention

Citation PENG Z W, ZHU X F, GUO J F. Label-Guided Dual-Attention Deep Neural Network Model. Pattern Recognition and Artificial Intelligence, 2022, 35(2): 175-184.

随着互联网的飞速发展,网上购物变得越来越普遍.用户在电商平台购物的过程中会留下相关评论,这对于商家提高服务水平和产品质量具有重要价值.目前,情感分类已成为研究者们重点关注的研究方向.情感分类的目的是给评论文本分配一个情感标签,情感标签反映评论文本的情感倾向.评论情感分类任务可分为文档级别的情感分类、句子级别的情感分类和方面级别的情感分类^[1-2].本文关注文档级别的情感分类,即为输入文档分配一个整体的情感倾向.通常将文档级别的情感分类视为一种文档分类任务^[3-4].

传统的情感分类方法基于人工提取有效特征,这些特征可用于有监督的学习方法^[5-6]或半监督的方法^[7].Gao等^[5]建模用户宽容度和产品流行度,提高情感分类性能.Goldberg等^[7]提出基于图的半监督算法,在标记数据和未标记数据上创建一个图,为这项任务编码,提高未标记样本的预测准确率.但是,上述方法依赖于人工提取有效特征,在面对海量数据时,将会耗费大量的时间与精力.

基于深度学习的方法主要采用词嵌入表示技术表示文本词语,利用深层神经网络从低层的文本表示中学习抽象文本语义特征,如将卷积神经网络(Convolutional Neural Network, CNN)^[8]和循环神经网络(Recurrent Neural Network, RNN)^[9]应用于学习文档表示.相比传统方法,基于深度学习的方法不需要人工提取特征即可取得显著的分类效果.Tang等^[10]提出基于神经网络的层次编码模型,学习有效的评论表示.层次注意力机制^[11-13]也广泛应用于构建评论的有效特征.

近些年,研究者们引入摘要生成模块,提升模型分类性能.Ma等^[14]提出HSSC(Hierarchical Summarization and Sentiment Classification),包含一个摘要层和一个情感分类层,同时优化摘要生成任务和情感分类任务性能,实现更好的情感分类效果.Chan等^[15]进一步提出Dual-view,首先使用编码器学习评论的上下文表示,摘要解码器应用复制机制(Copy Mechanism)^[16]逐字生成评论摘要.然后,摘要视图情感分类器预测生成摘要的情感标签,评论视图情感分类器预测评论的情感标签.在训练过程中,引入不一致性损失函数,惩罚这两个分类器之间

的不一致性.不同于文献[14],Chan等^[15]将评论内容和生成的摘要作为两个不同的视图,分别在其上输出分类结果.相比以往方法,文献[15]方法虽然取得更好的情感分类结果,但同时也面临模型复杂程度较高、在训练时需要消耗更多计算资源的问题.

上述研究方法只是将类别表示作为标签词汇表中的索引,模型缺乏关于分类的细粒度标签指导信息.近年来,研究者们提出利用标签描述带有的语义信息显式指导文本分类,相继提出一些方法.Du等^[17]认为基于深度学习的模型在分类时主要依赖文本级表示,忽略细粒度(词与类之间的匹配信息)分类线索.为了解决此问题,他们提出EXAM(Explicit Interaction Model),利用标签文本信息编码每类,得到可训练的矩阵(每行表示一类),再采用点积作为交互函数,计算目标词和标签之间的匹配分数,提取更丰富的信息,提高分类效果.但是,此方法未考虑不同的标签可能会有相似的文本,影响分类性能.Xiao等^[18]提出LSAN(Label-Specific Attention Network),利用标签语义信息确定标签和文档之间的语义连接,构造特定于标签的文档表示.同时,采用自注意力机制从文档内容中学习标签特定的文档表示.为了实现上述两部分的无缝融合,设计自适应融合策略,有效输出综合的文档表示,构建多标签文本分类器.上述两种方法都依赖于数据集带有标签文本信息描述标签,例如使用“积极”描述简单的类别标签“1”.虽然标签“1”不具有情感语义信息,但单词“积极”具有丰富的情感信息,可用于显式计算文本单词和标签之间的语义关系.然而,在一些应用场景中,无法获得相应的类别标签表述信息,只含有简单的类别标签,标签本身不具有情感语义信息,而且数据集没有关于标签的描述信息,从而无法应用相关方法.

为了解决上述问题,本文提出标签指导的双注意力深度神经网络模型(Label-Guided Dual-Attention Deep Neural Network Model, LGDA).首先,提出基于逆标签频次(Inverse Label Frequency, ILF)的自动类别标签描述生成方法,为每个标签生成特定的标签描述,用于显式计算文本单词与标签之间的语义关系.在此基础上,使用文本编码器学习具有上下文语境信息的评论文本表示,并提出标签指导的

双注意力网络,分别学习基于自注意力的文本表示和基于标签注意力的文本表示.然后,使用自适应门控机制融合这两个文本表示,得到文本最终表示.最后,使用两层前馈神经网络作为分类器,进行情感分类.在3个公开的真实数据集上的实验表明,LGDA的分类效果较优,减少训练时长和计算代价.

1 标签指导的双注意力深度神经网络模型

标签指导的双注意力深度神经网络模型(LGDA)框架包含如下5个部分:标签描述生成、文本编码器、标签指导的双注意力网络、自适应门控机制、情感分类器.模型框架如图1所示.

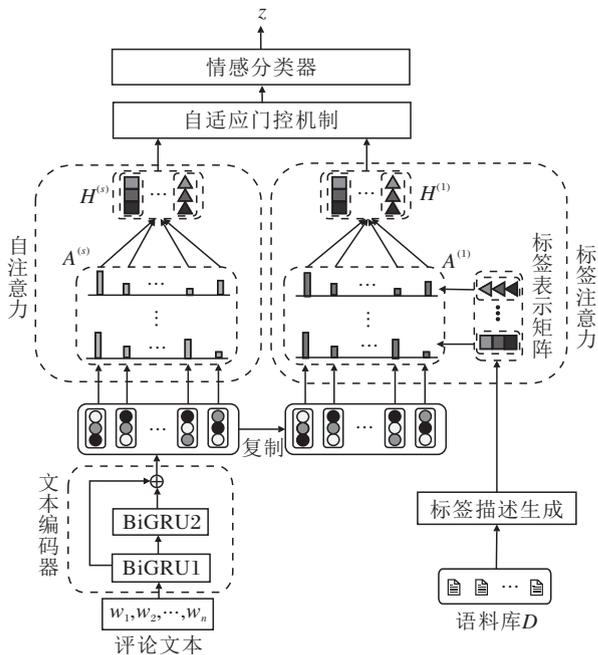


图1 标签指导的双注意力深度神经网络模型框图

Fig.1 Framework of label-guided dual-attention deep neural network model

1.1 标签描述生成

本节探讨如何为每个情感标签生成有效描述,以学习标签指导的文本表示.直观上,一个有效的标签描述应满足2个关键要求.1)相关性.标签描述中的单词应在语义上表示该标签,即每个单词应与标签有很强的相关性.2)区分力.标签描述应具有对不同标签的高区分能力,即标签描述中的单词与该标签的相关性较强,而与其它标签的相关性较弱.

为了解决相关性,测量一个单词对给定标签的

相关性得分.给定一个语料库 D , 一个情感标签 $c \in \{1, 2, \dots, C\}$, 其中 C 表示标签集合大小, 一个单词 w , 一个评论文本 $d \in D$, 计算单词 w 与标签 c 的相关性得分:

$$r_{w,c} = \sum_{d \in D^c} s_{w,d},$$

其中, D^c 表示在 D 中所有标签为 c 的文本, $s_{w,d}$ 表示单词 w 在 d 中的重要性得分.

值得注意的是, 现有多种方法可计算重要性分数 $s_{w,d}$, 鉴于 TF-IDF (Term Frequency-Inverse Document Frequency)^[19] 的良好性能, 本文使用 TF-IDF 计算 $s_{w,d}$. 单词 w 在评论文本 d 中的 TF-IDF 值为

$$s_{w,d} = TFIDF(w, d) = f_{w,d} \cdot \ln \left(\frac{|D|}{f_{w,D}} \right),$$

其中, $f_{w,d}$ 表示单词 w 在 d 中出现的次数, $|D|$ 表示语料库 D 的大小, $f_{w,D}$ 表示语料库中包含单词 w 的评论文本的数量. 计算单词 w 与标签 c 相关性得分:

$$r_{w,c} = \sum_{d \in D^c} f_{w,d} \cdot \ln \left(\frac{|D|}{f_{w,D}} \right).$$

为了处理区分力, 需要评估一个单词对所有情感标签的区分能力, 由此提出逆标签频次 (ILF), ILF 定义如下. 受 TF-IDF 的启发, 基于与逆文档频次 (Inverse Document Frequency, IDF) 相似的策略衡量一个单词的识别能力, 即本文提出的 ILF 中的标签频次 (Label Frequency, LF) 等于单词 w 出现在相应文本中的不同情感标签的数量:

$$LF(w) = \left| \bigcup_{d \in D} \{y_d \mid w \in d\} \right|,$$

其中 y_d 表示文本 d 对应的情感标签. 得到单词 w 相对于标签 c 的基于 ILF 的相关性分数:

$$r_{w,c}^{ILF} = \frac{r_{w,c}}{LF(w)} = \frac{\sum_{d \in D^c} f_{w,d} \cdot \ln \left(\frac{|D|}{f_{w,D}} \right)}{\left| \bigcup_{d \in D} \{y_d \mid w \in d\} \right|}.$$

为标签 c 选取 $r_{w,c}^{ILF}$ 得分最高的 Top- K 单词 $[w_{1,c}, \dots, w_{i,c}, \dots, w_{K,c}]$ 作为标签 c 的描述. 使用词向量矩阵 $\mathbf{E} \in \mathbf{R}^{V \times d_e}$ (V 表示词汇表大小, d_e 表示词嵌入的维度), 把每个描述标签 c 的单词 $w_{i,c}$ 嵌入为词向量 $\mathbf{q}_{i,c} \in \mathbf{R}^{d_e}$, 再通过

$$\mathbf{q}_c = \frac{1}{K} \sum_{i=1}^K \mathbf{q}_{i,c} \in \mathbf{R}^{d_e},$$

得到标签 c 的表示.

同理, 可计算其它标签的表示. 最终得到标签表示矩阵:

$$\mathbf{Q}_L = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_C) \in \mathbf{R}^{C \times d_e}.$$

1.2 文本编码器

在文本编码器模块,可选择 BiGRU(Bi-directional Gated-Recurrent Unit)^[20]、BiLSTM(Bi-directional Long Short-Term Memory)^[21]、Transformer^[22]等作为文本编码器,学习评论文本的隐藏状态.本文与 Dual-view 一致,选择 BiGRU 作为文本编码器.

首先,将评论文本序列 $d = (w_1, w_2, \dots, w_n)$ 输入编码器中,使用词向量矩阵 $E \in \mathbf{R}^{1V1 \times d_e}$ ($|V|$ 表示词汇表大小, d_e 表示词嵌入的维度) 把每个输入单词 w_i 嵌入为词向量 $x_i \in \mathbf{R}^{d_e}$. 为了将评论文本的上下文信息整合到每个单词的表示中,将每个词向量 x_i 输入 BiGRU 中,学习浅层隐藏表示 $u_i \in \mathbf{R}^{d_u}$. 一个 BiGRU 由一个正向 GRU(Gated-Recurrent Unit)(从 x_1 到 x_n 读取词向量序列) 和一个反向 GRU(从 x_n 到 x_1 读取词向量序列) 组成,前向 GRU 和后向 GRU 的隐藏状态为:

$$\begin{aligned}\vec{u}_i &= GRU_{11}(x_i, \vec{u}_{i-1}) \in \mathbf{R}^{d_u/2}, \\ \overleftarrow{u}_i &= GRU_{12}(x_i, \overleftarrow{u}_{i+1}) \in \mathbf{R}^{d_u/2},\end{aligned}$$

其中, GRU_{11} 表示第一个前向 GRU, GRU_{12} 表示第一个后向 GRU. 拼接 \vec{u}_i 和 \overleftarrow{u}_i , 得到 x_i 的浅层隐藏表示

$$u_i = \left[\vec{u}_i; \overleftarrow{u}_i \right],$$

由此得到评论文本的浅层隐藏表示

$$U = (u_1, u_2, \dots, u_n).$$

然后,把浅层隐藏表示输入到第 2 个 BiGRU 中,学习评论文本单词之间更复杂的交互作用:

$$\begin{aligned}\vec{h}_i &= GRU_{21}(u_i, \vec{h}_{i-1}) \in \mathbf{R}^{d_u/2}, \\ \overleftarrow{h}_i &= GRU_{22}(u_i, \overleftarrow{h}_{i+1}) \in \mathbf{R}^{d_u/2}.\end{aligned}$$

拼接 \vec{h}_i 和 \overleftarrow{h}_i , 得到 $h_i \in \mathbf{R}^{d_u}$. 使用残差连接融合 u_i 和 h_i . 残差连接在深度学习中是避免梯度消失问题的标准技术^[23], 计算公式如下:

$$\tilde{h}_i = \gamma h_i + (1 - \gamma) u_i,$$

其中, $\gamma \in [0, 1]$, 表示超参数.

最后得到编码器输出的文本表示

$$\tilde{H} = (\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_n) \in \mathbf{R}^{n \times d_u}.$$

1.3 标签指导的双注意力网络

本文提出标签指导的双注意力网络,目的是寻找评论文本中与不同标签相关的单词,进一步确定整个评论文本的情感倾向.事实上,这个策略对于评论文本情感分类是很直观的.例如,对于评论“...This baby doll is perfect! She loved it as soon

as we handed it to her, and she carries it all around with her...”,情感标签为 5,表示非常积极的态度.显然,单词 perfect 与第 5 类更相关,we, is 等词偏中性,与第 3 类更相关,而整个文本与第 5 类最相关.所以,注意力网络应更关注单词 perfect.下面将使用注意力网络捕获这个特征.

1.3.1 自注意力机制

一个评论文本中不同单词对每个标签有不同的贡献.为了确定文本中与每个标签最相关的单词,采用自注意力机制(Self-Attention Mechanism)^[24].使用

$$A(s) = \text{softmax}\left(W_2 \tanh\left(W_1 \tilde{H}^T\right)\right) \in \mathbf{R}^{C \times n}$$

计算 label-word 注意力分数,其中, $W_1 \in \mathbf{R}^{d_u \times d_u}$, $W_2 \in \mathbf{R}^{C \times d_u}$, 表示可训练学习的自注意力参数. 每行 $A_j^{(s)}$ (n 为评论文本长度) 表示所有单词对第 j 个标签的贡献,即第 j 个标签与所有单词的 label-word 注意力分数.然后,利用 label-word 注意力分数 $A_j^{(s)}$, 得到评论文本沿第 j 个标签的新表示:

$$H_j^{(s)} = A_j^{(s)} \tilde{H}.$$

同理,可计算评论文本沿其它标签的新表示.最后整个矩阵 $H^{(s)} \in \mathbf{R}^{C \times d_u}$ 是由自注意力机制得到的特定于标签的文本表示.

1.3.2 标签注意力机制

自注意力机制只考虑文本内容,可视为基于内容的注意力.众所周知,标签在文本分类中具有特定语义,一般隐藏在标签文本或描述中,对文本分类具有指导作用.然而,一些应用场景中没有标签文本或描述.对于没有标签文本或描述的数据集,使用 1.1 节中方法,得到标签表示矩阵 Q_L . 然后,将标签表示矩阵映射到与 \tilde{H} 相同的向量空间:

$$Q_L^* = \text{ReLU}(Q_L W_3) \in \mathbf{R}^{C \times d_u},$$

其中 $W_3 \in \mathbf{R}^{d_e \times d_u}$ 为可训练的参数.接下来,计算单词和标签表示 Q_L^* 之间的语义关系,情感特征越强的单词与标签表示 Q_L^* 之间的语义相关性越大.为了节省计算代价,本文利用点积计算单词与标签之间的语义关系:

$$A^{(l)} = Q_L^* \tilde{H}^T \in \mathbf{R}^{C \times n}.$$

类似于自我注意力机制,标签指导的特定于标签的文本表示如下:

$$H^{(l)} = A^{(l)} \tilde{H}.$$

在标签的指导作用下,评论文本被重新表示为 $H^{(l)} \in \mathbf{R}^{C \times d_u}$. 因为这个表示是基于标签的,所以称为标签

注意力机制(Label Attention Mechanism).

1.4 自适应门控机制

虽然 $\mathbf{H}^{(s)}$ 和 $\mathbf{H}^{(l)}$ 都是特定于标签的文本表示,但侧重点不同. $\mathbf{H}^{(s)}$ 更侧重于文本内容,而 $\mathbf{H}^{(l)}$ 更注重文本内容与标签表示之间的语义关系. 为了更好地利用这两部分的优势,使用自适应门控机制,从这两部分中自适应地提取有效信息,用于构建全面的标签指导的文本表示.

本文利用2个权重向量 $\boldsymbol{\alpha} \in \mathbf{R}^C$, $\boldsymbol{\mu} \in \mathbf{R}^C$, 决定 $\mathbf{H}^{(s)}$ 和 $\mathbf{H}^{(l)}$ 的重要性,即

$$\boldsymbol{\alpha} = \text{sigmoid}(\mathbf{H}^{(s)} \mathbf{W}_4), \boldsymbol{\mu} = \text{sigmoid}(\mathbf{H}^{(l)} \mathbf{W}_5),$$

其中, $\mathbf{W}_4 \in \mathbf{R}^{d_u}$, $\mathbf{W}_5 \in \mathbf{R}^{d_u}$, 表示可训练的参数, α_j 、 μ_j 分别表示沿第 j 个标签构造最终文本表示时自注意力和标签注意力的重要性. 因此,使用

$$\alpha_j = \frac{\alpha_j}{\alpha_j + \mu_j}, \mu_j = 1 - \alpha_j$$

进行正则化. 然后,利用 α_j 和 μ_j 计算得到沿第 j 个标签的最终文本表示:

$$\mathbf{H}'_j = \alpha_j \mathbf{H}_j^{(s)} + \mu_j \mathbf{H}_j^{(l)}.$$

所有标签的特定于标签的文本表示可描述为矩阵 $\mathbf{H}' \in \mathbf{R}^{C \times d_u}$. 最后,使用均值池化计算得到最终评论文本的表示:

$$\bar{\mathbf{h}}' = \text{AveragePooling}(\mathbf{H}') \in \mathbf{R}^{d_u}.$$

1.5 情感分类器

在得到文本的最终表示后,使用两层前馈神经网络作为分类器. 在网络输出端应用 softmax 函数生成情感标签上的概率分布:

$$P(z | \bar{\mathbf{h}}') =$$

$$\text{softmax} \left(\mathbf{W}_7 \left(\text{ReLU} \left(\mathbf{W}_6 \bar{\mathbf{h}}' + \mathbf{b}_1 \right) \right) + \mathbf{b}_2 \right),$$

其中, $\mathbf{W}_6 \in \mathbf{R}^{d_u \times d_u}$, $\mathbf{W}_7 \in \mathbf{R}^{C \times d_u}$, $\mathbf{b}_1 \in \mathbf{R}^{d_u}$, $\mathbf{b}_2 \in \mathbf{R}^C$, 都为可训练的模型参数.

最后,将概率最高的情感标签作为评论文本的预测情感标签.

本文使用负对数似然作为损失函数:

$$L = -\ln(P(z^* | d)),$$

其中 z^* 表示真实的评论文本情感标签.

2 实验及结果分析

2.1 实验设置

本文使用 Chan 等^[15] 从 Amazon 5-core review repository^[25] 收集的3个公开数据集. 采用如下3个领域的商品评论作为数据集: Sports & Outdoors(表示为 Sports); Toys & Games(表示为 Toys); Home & Kitchen(表示为 Home). 实验中每个数据样本由一个评论文本和一个评分组成. 把评分作为情感标签,取整数,取值范围是[1,5].

为了公平对比,在数据预处理时,遵循 Chan 等^[15] 做法:将所有字母转换为小写并使用 Stanford CoreNLP^[26] 对文本进行分词. 为了减少数据集的噪音,过滤评论文本长度小于16或大于800的数据样本. 把每个数据集随机分割为训练集、验证集和测试集. 数据集的统计信息如表1所示. 在表中, Ave. RL 表示训练集中评论文本的平均长度, L + c 表示情感标签为 c 的数据样本在训练集中的比例.

表1 实验数据集统计信息

Table 1 Statistics of experimental datasets

数据集	训练集	验证集	测试集	Ave. RL	L+1/%	L+2/%	L+3/%	L+4/%	L+5/%
Sports	183714	9000	9000	108.3	3.3	3.9	9.1	22.9	60.8
Toys	104296	8000	8000	125.9	2.9	4.1	11.0	24.0	58.0
Home	367395	10000	10000	120.9	5.3	4.9	9.1	20.1	60.6

2.2 评价指标

从表1中可发现,情感标签的类别分布不均衡,所以本文使用 Macro-averaged F1(M. F1) 和 Balanced Accuracy(B. Acc)^[27] 作为评价指标.

M. F1 用于评估所有不同类别标签的平均 F1. M. F1 赋予每个标签相同的权重,定义如下:

$$M. F1 = \frac{1}{|C|} \sum_{i \in C} \left(\frac{2P_i R_i}{P_i + R_i} \right),$$

其中,

$$P_i = \frac{TP_i}{TP_i + FP_i}, R_i = \frac{TP_i}{TP_i + FN_i},$$

TP_i 表示预测标签为 i 且真实标签为 i 的样本数, FP_i 表示预测标签为 i 且真实标签为非 i 的样本数, FN_i 表示预测标签非 i 且真实标签为 i 的样本数, $|C|$ 表示标签集合的大小.

B. Acc 为准确率针对不平衡数据集的一种变体,定义如下:

$$B. Acc = \frac{1}{|C|} \sum_{i \in C} R_i,$$

$$R_i = \frac{TP_i}{TP_i + FN_i}.$$

在每个数据集的训练集上训练一个 128 维的 word2vec^[28], 用于初始化包括基线模型在内的所有模型的词嵌入. 定义词汇表为训练集中出现次数最频繁的 50 000 个单词. 在实验中 d_e 设置为 128, d_u 设置为 512, γ 设置为 0.5, 批尺寸大小设置为 32. 优化器使用自适应矩估计 (Adaptive Moment Estimation, Adam)^[29], 初始学习率设置为 0.001, 如果验证集损失停止下降, 学习率减少一半.

基线模型可分为如下两组.

1) 只做情感分类的模型, 包括:

(1) BiGRU + Attention. 首先, 使用 BiGRU 单元层^[20] 将输入的评论文本编码成隐藏状态. 然后利用 attention mechanism^[30] with glimpse operation^[31], 从编码器生成的隐藏状态中聚合信息, 生成一个向量. 这个向量再经过一个两层前馈神经网络预测情感标签.

(2) BiLSTM + Attention. 首先, 使用 BiLSTM^[21] 将输入的评论文本编码成隐藏状态. 然后利用 attention mechanism^[30] with glimpse operation^[31], 从编码器生成的隐藏状态中聚合信息, 生成一个向量. 这个向量再经过一个分类器预测情感标签.

(3) Transformer + Attention. 首先使用一个标准的 Transformer^[22], 将输入的评论文本编码成隐藏状态. 然后利用注意力机制把编码器生成的隐藏状态聚合成一个向量. 这个向量再经过一个情感分类器预测情感标签.

(4) DARLM (Differentiated Attentive Representation Learning Model)^[32]. 试图缓解句子分类中的注意力偏向问题. 模型具有两个注意力子网分支和一个示例鉴别器. 这两个分支联合训练, 其中一个分支尽量对所有句子进行分类, 另一个分支对前者不能较好处理的句子进行分类. 最后示例鉴别器选择合适的注意力子网, 进行句子分类.

2) 联合评论摘要与情感分类的模型, 包括:

(1) HSSC^[14]. 联合改进评论摘要和情感分类的模型. 由一个摘要层和一个情感分类层组成. 摘要层

将原文压缩成短句, 情感分类层进一步将文本“概括”成情感类. 层次结构在文本摘要和情感分类之间建立紧密联系, 使这两个任务可以相互改进. 使用摘要压缩文本后, 情感分类器更容易预测较短文本的情感标签.

2) Max^[14]. 使用 BiGRU 单元层^[20] 将输入的评论编码为隐藏状态. 摘要解码器和情感分类器共享这些隐藏状态. 情感分类器利用最大池化将编码器生成的隐藏状态聚合成一个向量, 再通过一个两层前馈神经网络预测情感标签.

3) HSSC + copy. 把 copy mechanism^[16] 应用到 HSSC 中, 作为一个强基线.

4) Max + copy. 把 copy mechanism^[16] 应用到 Max 中, 作为另外一个强基线.

5) Dual-view^[15]. 联合改进评论摘要和情感分类的模型. 首先使用编码器学习评论的上下文表示, 摘要解码器应用 copy mechanism^[16] 逐字生成评论摘要. 然后, 评论视图情感分类器使用编码器输出的上下文表示预测评论的情感标签, 而摘要视图情感分类器使用解码器的隐藏状态预测生成摘要的情感标签. 在训练过程中引入不一致性损失, 惩罚这两个分类器之间的一致.

2.3 情感分类实验结果

本文在 3 个公开的数据集上进行实验, 各模型的情感分类结果如表 2 所示. 表中最优结果使用黑体数字表示, 次优结果使用斜体数字表示.

在 Toys 数据集上, LGDA 的 M. F1 值达到 57.55%, B. Acc 值达到 57.59%, 比次优的 Dual-view 分别提升 1.85% 和 3.53%. 在 Sports 数据集上, LGDA 的 M. F1 值达到 56.89%, B. Acc 值达到 55.70%, 比次优的 Dual-view 分别提升 0.58% 和 1.42%. 在 Home 数据集上, LGDA 的 M. F1 值达到 60.95%, B. Acc 值达到 59.81%, 比次优的 Dual-view 分别提升 0.22% 和 0.18%. 由此表明 LGDA 比基线模型更准确地预测评论情感标签. 另外, Toys、Sports、Home 数据集上训练样本数依次增加, LGDA 在 3 个数据集上的 M. F1 值和 B. Acc 值比 Dual-view 的提升依次减少. 这表明训练数据越充分, 相比 Dual-view, LGDA 提升越少.

表2 各模型在3个数据集上的情感分类结果对比

Table 2 Result comparison of sentiment classification of different models on 3 datasets

模型	%					
	Sports		Toys		Home	
	M. F1	B. Acc	M. F1	B. Acc	M. F1	B. Acc
DARLM	49.60	47.95	50.58	48.67	54.49	53.43
Transformer+Attention	45.67	43.41	47.28	47.08	48.10	47.67
BiLSTM+Attention	55.37	52.93	54.80	53.81	59.65	59.10
BiGRU+Attention	54.21	53.03	53.54	52.82	59.32	58.03
HSSC	53.49	51.99	54.24	53.66	58.51	57.42
Max	53.27	52.64	55.02	53.64	58.31	57.36
HSSC+copy	53.14	52.63	54.38	53.32	58.78	58.02
Max+copy	53.95	52.53	53.52	52.01	58.85	58.05
Dual-view	56.31	54.28	55.70	54.06	60.73	59.63
LGDA	56.89	55.70	57.55	57.59	60.95	59.81

2.4 消融实验结果

为了验证 LGDA 中的自注意力、标签注意力、自适应门控机制及逆标签频次的有效性,在 3 个数据集上进行消融实验,结果如表 3 所示.在表中: - L 表示移除标签注意力; - S 表示移除自注意力; - G 表示移除自适应门控机制,使用拼接; - ILF 表示在生成标签描述时不使用 ILF,只用 TF-IDF; - BiGRU

(+ BiLSTM) 和 - BiGRU(+ Transformer) 表示在文本编码器部分分别使用 BiLSTM 和 Transformer 替代 BiGRU.

考虑到 BiGCN(Bi-level Interactive Graph Convolution Network) 更适用于短文本分类^[33],因此在实验中只新增与 BiLSTM 和 Transformer 的对比.表中黑体数字表示最优结果.

表3 各模型在3个数据集上的消融实验结果

Table 3 Result comparison of ablation experiment of different models on 3 datasets

模型	%					
	Sports		Toys		Home	
	M. F1	B. Acc	M. F1	B. Acc	M. F1	B. Acc
- L	55.65	53.22	56.34	55.71	59.39	58.19
- S	53.49	53.04	56.40	55.16	59.22	58.65
- G	55.89	54.34	55.88	55.12	60.11	59.25
- ILF	56.40	55.36	57.42	57.15	60.81	59.75
- BiGRU(+ Transformer)	46.38	45.07	50.49	50.18	48.71	48.01
- BiGRU(+ BiLSTM)	55.54	53.76	55.53	54.35	60.04	58.88
LGDA	56.89	55.70	57.55	57.59	60.95	59.81

由表 3 可看到,分别移除标签注意力、自注意力、逆标签频次及使用拼接代替自适应门控机制之后,2 个指标值都有不同程度的下降,表明这 4 个部分都有积极作用.另外,在所有数据集上的分类结果显示,使用 BiLSTM 和 Transformer 代替 BiGRU 之后,分类性能均有不同程度的下降,因此使用 BiGRU 比使用 BiLSTM 和 Transformer 更适合 LGDA.

2.5 参数敏感性

为了探究生成标签描述时 K 值的影响,在 3 个数据集上进行参数敏感性实验.分别取 $K = 10, 30,$

$50, 70, 90$, 结果如表 4 所示.

表4 K 不同对指标值的影响Table 4 Influence of different K values on index values

K	Sports/%		Toys/%		Home/%	
	M. F1	B. Acc	M. F1	B. Acc	M. F1	B. Acc
10	55.62	54.24	56.80	56.33	59.79	58.64
30	56.07	54.54	57.20	57.22	59.98	58.82
50	56.89	55.70	57.35	57.59	60.14	58.94
70	55.94	53.95	57.04	56.60	60.95	59.81
90	55.31	53.71	56.99	55.73	59.37	58.43

由表 4 可看到:在 Sports、Toys 数据集上, $K = 50$ 时,分类效果最优;在 Home 数据集上, $K = 70$ 时,分类效果最优。

另外,随着 K 值增大,所有指标值都是先升后降。这是因为当 K 值过小时,选取的关键词较少,信息不充分,无法较好地描述标签,分类性能较差。随着 K 值的增加,选取的关键词增加,信息变得丰富,可更好地描述标签,提升分类效果。当 K 到达某个值时,分类效果最优。若 K 值继续增大,分类效果开始下降,这是因为选取的关键词过多,引入噪音,降低分类效果。

表 5 LGDA 和 Dual-view 的计算代价对比

Table 5 Calculation cost comparison of LGDA and Dual-view

模型	Sports		Toys		Home	
	GPU 内存 /MB	训练时长 /min	GPU 内存 /MB	训练时长 /min	GPU 内存 /MB	训练时长 /min
Dual-view	7654	75	8134	60	7954	160
LGDA	1789	35	1856	30	1824	120

2.7 案例分析

为了更直观地对比 LGDA 与 Dual-view 之间捕获情感特征词的能力,本文将 2 个模型对于同一个评论文本单词的注意力权重可视化,具体如表 6 所示,颜色越深表示权重越大,颜色越浅表示权重越小。由表 5 可看到,Dual-view 虽然关注 love、good 等

2.6 计算代价对比

LGDA 与 Dual-view 的计算代价对比如表 4 所示。表中 GPU 内存表示训练模型时,GPU 内存使用量,训练时长表示模型训练所需时长。在训练过程中,2 个模型的批尺寸大小均设置为 32,采用相同的早停策略。由表可见,LGDA 在 GPU 内存消耗及训练时长上都显著低于 Dual-view。这是因为 LGDA 没有联合摘要生成进行训练,过程更简单,消耗 GPU 内存更小,训练时长更短。但是 LGDA 通过生成标签描述指导模型分类,在简化模型、降低计算代价的同时,可提高分类效果。

积极的情感词,但是分配的权重较小,同时对于与情感无关的词 buy、product 等也分配较大的权重。而 LGDA 对于情感词 love、good、favors、well 分配较大的权重,对于与情感无关的词,分配很少的权重。所以,LGDA 正确预测情感标签,而 Dual-view 预测错误。这表明 LGDA 拥有更好的捕捉情感特征的能力。

表 5 LGDA 和 Dual-view 的注意力权重可视化对比

Table 5 Visualization comparison of attention weights of LGDA and Dual-view

模型	评论文本	预测标签	真实标签
Dual-view	kids love these as party favors, and fit well into a pinata, it was what we expected for the product, just remember the plastic piece is a weight that goes on the nose of the plane and it will fly really good. remind your friends, kids to do this otherwise they won't go far. would buy again.	4	5
LGDA	kids love these as party favors, and fit well into a pinata, it was what we expected for the product, just remember the plastic piece is a weight that goes on the nose of the plane and it will fly really good. remind your friends, kids to do this otherwise they won't go far. would buy again.	5	5

3 结束语

本文提出标签指导的双注意力深度神经网络模型(LGDA)。首先,提出基于逆标签频次的自动类别标签描述生成方法,为每个标签生成特定的标签描述,用于显式计算文本单词与标签之间的语义关系。在此基础上,使用文本编码器学习有效的评论上下文表示,并提出标签指导的双注意力网络,分别学习基于自注意力的文本表示和基于标签注意力的文本

表示。然后,使用自适应门控机制融合这两个文本表示,得到文本最终表示。最后,使用两层前馈神经网络作为分类器,进行情感分类。在 3 个公开的真实数据集上的实验表明,LGDA 分类效果较优,减少计算代价与训练时长。今后将探索更好的生成标签描述的方法,提高分类性能。

参 考 文 献

[1] 何丽,房婉琳,张红艳. 基于上下文保持能力的方面级情感分类

- 模型. 模式识别与人工智能, 2021, 34(2): 157-166.
- (HE L, FANG W L, ZHANG H Y. Aspect-Level Sentiment Classification Model Based on Context-Preserving Capability. *Pattern Recognition and Artificial Intelligence*, 2021, 34(2): 157-166.)
- [2] ZHANG L, WANG S, LIU B. Deep Learning for Sentiment Analysis: A Survey. *WIREs: Data Mining and Knowledge Discovery*, 2018, 8(4). DOI: 10.1002/widm.125.
- [3] QIN Q, HU W P, LIU B. Feature Projection for Improved Text Classification // Proc of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: ACL, 2020: 8161-8171.
- [4] WANG H, WANG S, MAZUMDER S, *et al.* Bayes-Enhanced Long Attention Networks for Sentiment Classification // Proc of the 28th International Conference on Computational Linguistics. Stroudsburg, USA: ACL, 2020: 580-591.
- [5] GAO W L, YOSHINAGA N, KAJI N, *et al.* Modeling User Leniency and Product Popularity for Sentiment Classification // Proc of the 6th International Joint Conference on Natural Language Processing. Stroudsburg, USA: ACL, 2013: 1107-1111.
- [6] QU L Z, IFRIM G, WEIKUM G. The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns // Proc of the 23rd International Conference on Computational Linguistics. Stroudsburg, USA: ACL, 2010: 913-921.
- [7] GOLDBERG A B, ZHU X J. Seeing Stars When There Aren't Many Stars: Graph-Based Semi-Supervised Learning for Sentiment Categorization // Proc of the 1st Workshop on Graph Based Methods for Natural Language Processing. Stroudsburg, USA: ACL, 2006: 45-52.
- [8] KURATA G, XIANG B, ZHOU B W. Improved Neural Network-Based Multi-label Classification with Better Initialization Leveraging Label Co-occurrence // Proc of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, USA: ACL, 2016: 521-526.
- [9] LAI S W, XU L H, LIU K, *et al.* Recurrent Convolutional Neural Networks for Text Classification // Proc of the 29th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI, 2015: 2267-2273.
- [10] TANG D Y, QIN B, LIU T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification // Proc of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: ACL, 2015: 1422-1432.
- [11] YANG Z C, YANG D Y, DYER C, *et al.* Hierarchical Attention Networks for Document Classification // Proc of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, USA: ACL, 2016: 1480-1489.
- [12] YIN Y C, SONG Y Q, ZHANG M. Document-Level Multi-aspect Sentiment Classification as Machine Comprehension // Proc of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: ACL, 2017: 2044-2054.
- [13] 陈金广, 赵银歌, 马丽丽. 面向方面级情感分类的特征融合学习网络. 模式识别与人工智能, 2021, 34(11): 1049-1057.
- (CHEN J G, ZHAO Y G, MA L L. Feature Fusion Learning Network for Aspect-Level Sentiment Classification. *Pattern Recognition and Artificial Intelligence*, 2021, 34(11): 1049-1057.)
- [14] MA S M, SUN X, LIN J Y, *et al.* A Hierarchical End-to-End Model for Jointly Improving Text Summarization and Sentiment Classification[C/OL]. [2021-10-13]. <https://www.ijcai.org/Proceedings/2018/0591.pdf>.
- [15] CHAN H P, CHEN W, KING I. A Unified Dual-View Model for Review Summarization and Sentiment Classification with Inconsistency Loss // Proc of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM, 2020: 1191-1200.
- [16] SEE A, LIU P J, MANNING C D. Get to the Point: Summarization with Pointer-Generator Networks // Proc of the 55th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: ACL, 2017: 1073-1083.
- [17] DU C X, CHEN Z Z, FENG F L, *et al.* Explicit Interaction Model towards Text Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 6359-6366.
- [18] XIAO L, HUANG X, CHEN B L, *et al.* Label-Specific Document Representation for Multi-label Text Classification // Proc of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Stroudsburg, USA: ACL, 2019: 466-475.
- [19] SARI S, ADRIANI M. Applications of TF-IDF Concept to Improve Monolingual and Cross-Language Information Retrieval Based on Word Embeddings // Proc of the International Conference on Advanced Information Science and System. New York, USA: ACM, 2019. DOI: 10.1145/3373477.3373493.
- [20] CHO K, VAN MERRIËNBOER B, GULCEHRE C, *et al.* Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation // Proc of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: ACL, 2014: 1724-1734.
- [21] YOU R H, ZHANG Z H, WANG Z Y, *et al.* AttentionXML: Label Tree-Based Attention-Aware Deep Model for High-Performance Extreme Multi-label Text Classification[C/OL]. [2021-10-13]. <https://arxiv.org/pdf/1811.01727.pdf>.
- [22] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention Is All You Need // Proc of the 31st International Conference on Neural Information Processing Systems. New York, USA: ACM, 2017: 6000-6010.
- [23] HE K M, ZHANG X Y, REN S Q, *et al.* Deep Residual Learning for Image Recognition // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 770-778.
- [24] TAN Z X, WANG M X, XIE J, *et al.* Deep Semantic Role Labeling with Self-Attention // Proc of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI, 2018: 4929-4936.
- [25] MCAULEY J, TARGETT C, SHI Q F, *et al.* Image-Based Reco-

mmentations on Styles and Substitutes // Proc of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA: ACM, 2015; 43–52.

- [26] MANNING C, SURDEANU M, BAUER J, *et al.* The Stanford CoreNLP Natural Language Processing Toolkit // Proc of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Stroudsburg, USA: ACL, 2014; 55–60.
- [27] BRODERSEN K H, ONG C S, STEPHAN K E, *et al.* The Balanced Accuracy and Its Posterior Distribution // Proc of the 20th International Conference on Pattern Recognition. Washington, USA: IEEE, 2010; 3121–3124.
- [28] MIKOLOV T, SUTSKEVER I, CHEN K, *et al.* Distributed Representations of Words and Phrases and Their Compositionality // Proc of the 26th International Conference on Neural Information Processing Systems. New York, USA: ACM, 2013; 3111–3119.
- [29] KINGMA D P, BA J L. Adam: A Method for Stochastic Optimization [C/OL]. [2021–10–13]. <https://arxiv.org/pdf/1412.6980.pdf>.
- [30] BAHDANAU D, CHO K H, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate [C/OL]. [2021–12–13]. <https://arxiv.org/pdf/1409.0473.pdf>.
- [31] VINYALS O, BENGIO S, KUDLUR M. Order Matters: Sequence to Sequence for Sets [C/OL]. [2021–10–13]. <https://arxiv.org/pdf/1511.06391.pdf>.
- [32] ZHOU Q R, WANG X J, DONG X. Differentiated Attentive Representation Learning for Sentence Classification [C/OL]. [2021–10–13]. <https://www.ijcai.org/proceedings/2018/0644.pdf>.
- [33] ZHANG M, QIAN T Y. Convolution over Hierarchical Syntactic

and Lexical Graphs for Aspect Level Sentiment Analysis // Proc of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg, USA: ACL, 2020; 3540–3549.

作者简介



彭展望, 硕士研究生, 主要研究方向为自然语言处理、数据挖掘. E-mail: 51190324101@cqut.edu.cn.

(**PENG Zhanwang**, master student. His research interests include natural language processing and data mining.)



朱小飞(通信作者), 博士, 教授, 主要研究方向为自然语言处理、数据挖掘、信息检索. E-mail: zxf@cqut.edu.cn.

(**ZHU Xiaofei** (Corresponding author), Ph. D., professor. His research interests include natural language processing, data mining and information retrieval.)



郭嘉丰, 博士, 研究员. 主要研究方向为数据挖掘、信息检索. E-mail: guojiafeng@ict.ac.cn.

(**GUO Jiafeng**, Ph. D., professor. His research interests include data mining and information retrieval.)