

# TS-Aug 架构的半监督自训练情感分类算法

郭 卡, 王 芳

(安徽外国语学院信息与数学学院, 安徽 合肥 231200)

[摘要] 网络教学资源的普及使得资源评价的文本数据规模逐步增大. 传统的有监督学习文本分类对标注数据的依赖度较高, 需要足够的数据量和高质量数据才能得到良好的结果. 在网络教学资源的评价文本工作中, 由于标注数据难以获取且质量参差不齐, 使得这一任务的难度越来越高. 针对这一困难, 提出一种 TS-Aug 半监督自训练方案, 通过添加无标签数据并进行伪标签训练, 能在强力数据增广的作用下大幅扩充样本集, 解决数据增广中的过拟合风险. 首先利用标注数据和弱增广策略进行初始化监督训练, 然后利用无标注数据和强增广策略进行半监督训练, 最后使用标注数据进行微调监督训练. 在自建的在线课程评论数据中, 能将分类  $F1$ -Score 从 0.88 提升至 0.95, 表明 TS-Aug 半监督自训练方案在文本分类任务中具有较好的应用前景.

[关键词] 少样本学习, 半监督训练, 数据增广, 情感分类

[中图分类号] TP18; TP391 [文献标志码] A [文章编号] 1672-1292(2024)01-0045-08

## Semi-Supervised Self-Training Sentiment Classification Algorithm Based on TS-Aug Architecture

Guo Ka, Wang Fang

(School of Information and Mathematics, Anhui International Studies University, Hefei 231200, China)

**Abstract:** With the popularity of online teaching resources, the text data size for resource evaluation has gradually increased. Traditional supervised text classification heavily relies on labeled data and requires sufficient and high-quality data to achieve good results. The difficulty of this task has become increasingly high due to the difficulty in obtaining and uneven quality of labeled data. To address this difficulty, this paper proposes a semi-supervised self-training scheme named TS-Aug. By adding unlabeled data and pseudo-labels for training, we can significantly expand the sample set through the aggressive data augmentation, and also solved the overfitting risk in data augmentation. Specifically, the process involves initializing supervised training using labeled data and weak augmentation strategies, followed by semi-supervised training using unlabeled data and strong augmentation strategies, and finally fine-tuning the model with supervised training using labeled data. In our self-built online course comment data, we can improve the classification  $F1$ -score from 0.88 to 0.95. This indicates that the TS-Aug semi-supervised self-training scheme has good applied prospects in text classification tasks.

**Key words:** few-shot learning, semi-supervised training, data augmentation, sentiment classification

“互联网+”已渗透到社会的各个角落, 改变着人们的生活和工作. 在网络环境下, 不断涌现出各种全新的教学模式, 学习者在网络中也更容易张扬个性, 他们可以毫无顾忌地表达自己全部的感情而无须在意别人的眼光, 对网络教学资源的文本评价无疑是最好的体现. 因此, 通过挖掘学习者评论文本中隐藏的情感, 并精确进行情感分类, 从而获得学习者内隐的情感状态, 有重要的研究意义.

在情感分类任务中, 监督学习具有准确性高和可解释性强的特点. 这是因为监督学习模型通过学习输入数据和输出标签之间的映射关系来进行预测, 能够较好地捕捉到输入数据和情感类别之间的关系. 这使得监督学习方法在情感分类任务中更容易理解和解释模型的行为. 然而, 监督学习也存在一些不

收稿日期: 2023-07-01.

基金项目: 安徽省高校自然科学研究项目(KJ2021A1197)、安徽省省级质量工程课程思政教学团队项目(2020kcszjxt034)和安徽外国语学院校级质量工程教学创新团队项目(aw2023jxcxtd06).

通讯作者: 郭卡, 讲师, 研究方向: 深度学习与人工智能. E-mail: 409337713@qq.com

足:(1)监督学习需要训练大量的标注数据,这在某些情况下可能难以获取。例如,对于某些特定领域的文本数据,可能很难找到足够的标注样本来训练一个准确的模型。(2)监督学习模型对输入数据的噪声和异常值比较敏感。若训练数据中存在噪声或异常值,模型可能会受到影响并产生错误的预测结果。因此,在情感分类任务中,需要对数据进行预处理和清洗,以减少噪声和异常值的影响。(3)易出现过拟合现象,尤其是在训练数据较少的情况下,从而导致在新的测试数据上表现不佳。为了避免过拟合,可以使用正则化技术、交叉验证等方法来约束模型的复杂度。

半监督学习是一种介于监督学习和无监督学习之间的机器学习方法。与传统的监督学习方法相比,半监督学习不需要依赖大量标注数据来训练模型,在很大程度上降低了数据采集和标注的成本。同时,半监督学习还能在不牺牲模型性能和准确性的前提下,充分利用未标注的数据来提高模型的训练效果<sup>[1]</sup>。由于情感分类任务通常涉及到对大量文本数据进行情感倾向的判断,而标注这些数据需要大量的人工参与,因此成本较高。半监督学习通过利用未标注的数据来扩充训练集,使得模型能够在更广泛的数据集上进行训练,从而提高模型的泛化能力。此外,未标记的数据可以帮助模型更好地捕捉到数据中的复杂结构和潜在规律,从而减少噪声对模型的影响。总之,半监督学习在情感分类问题中的应用为提高模型性能和准确性提供了一种有效的途径。

本文采用不改变语义的弱增广方式和可能会改变语义的强增广方式,提出在半监督训练过程中为强增广数据生成伪标签,并用平滑参数模型约束主模型训练,最大限度地扩充了训练数据,解决了情感分类中标注数据缺乏的问题。

## 1 相关工作

在深度学习任务中常采用半监督学习的手段来利用无标签样本。近年来,半监督训练方向涌现出不少优秀作品,旨在充分利用无标签数据的自身特征,用以提高模型性能,其中应用最广泛的半监督手段就是基于伪标签的半监督学习<sup>[2]</sup>。无标注数据的伪标签通常来自一个已在监督数据上训练收敛的模型,使用聚类<sup>[3]</sup>或分类的方式生成伪标签,并将伪标签作为数据标签进行监督学习。由于伪标签生成过程中缺乏人工干预,常会出现带偏见的伪标签<sup>[4]</sup>,部分研究者会将这个模型通过特定的处理形成一个独立于主模型之外的模型,称为 Teacher 模型。而半监督训练得到的主模型为 Student 模型,训练过程中会用 Teacher 模型生成的伪标签来指导 Student 模型的训练。

最初用于训练 Teacher 模型的监督数据存在较多的噪声或异常值,导致 Teacher 模型过拟合,需要设法对伪标签进行纠正<sup>[5]</sup>。半监督学习中会基于数据自身的特点,构造无监督的训练任务<sup>[6-7]</sup>,一致性正则就是其中最普适的一种<sup>[8]</sup>,基于平滑假设和聚类假设,具有不同标签的数据点在低密度区域分离,且相似的数据点具有相似的输出。若对一个未标记的数据应用实际的扰动,其预测结果不应发生显著变化,也即输出具有一致性。由于这种方法一般基于模型输出预测向量,不需要具体的标签,所以可应用于半监督学习。通过在未标记数据上构造添加扰动后的预测结果与正常预测结果之间的无监督正则化损失项,提高模型的泛化能力。一致性正则可以与伪标签协同使用<sup>[9-10]</sup>,需要注意的是,并非所有的数据增广方法都适用于半监督学习,必须保证扰动后样本未丢失主要信息以及仍然是一个有意义的样本<sup>[11]</sup>。

## 2 实验方案

### 2.1 数据增广

本文为了避免数据增广对语义的曲解影响到监督学习训练,将数据增广方式分为弱增广和强增广。

#### 2.1.1 弱增广

弱数据增广采用 EDA<sup>[12]</sup>和中英回译<sup>[13]</sup>两种方式,主要目的为扩充同类别监督学习样本。其中,EDA 包含如下操作:同义词替换;同义词插入;随机删除;随机交换。EDA 增广同义词替换实例如下:

老师上课的方式很有趣→老师上课的方式很有意思。

中英回译是指将中文句子翻译成英文,再翻译回中文。由于中文和英文是目前世界上使用人数最多的两种语言,其互译的信息损失量很少,在绝大部分情况下能保证句子语义不变。中英回译增广实例如下:

老师上课的方式很有趣→The teacher's way of teaching is very interesting.→老师的教学方式很有趣。

### 2.1.2 强增广

强数据增广采用不限制同义词的 EDA 增广和中英日三语回译两种方式,主要目的为生成与监督学习样本分布相近的各种不同类别的样本,用于优化分类模型的决策边界. 在 EDA 的替换和插入操作中,不限制同义词,只限制相同词性的替换,会导致语义及情感的变化,例如:老师上课的方式很有趣→老师上课的方式很奇怪,正面评论经过同词性替换之后变成了负面评论.

强增广中选择中英日三语回译是为了增加语义扭曲的概率,具体方式是利用机器翻译模型实现中文→英文→日文→英文→中文的回译过程,例如:老师上课的方式很有趣→The way the teacher teaches is very interesting→先生の教え方はおもしろい→How to teach teachers is interesting→如何教老师很有趣. 经过一系列翻译后,句子的语义发生了显著的变化,这使得强增广的生成样本不仅仅局限于正例.

## 2.2 训练方案

本文的训练方案分为 3 个步骤,分别是监督学习初始化、半监督学习训练和监督学习微调,数据流向及训练准则如图 1 所示.

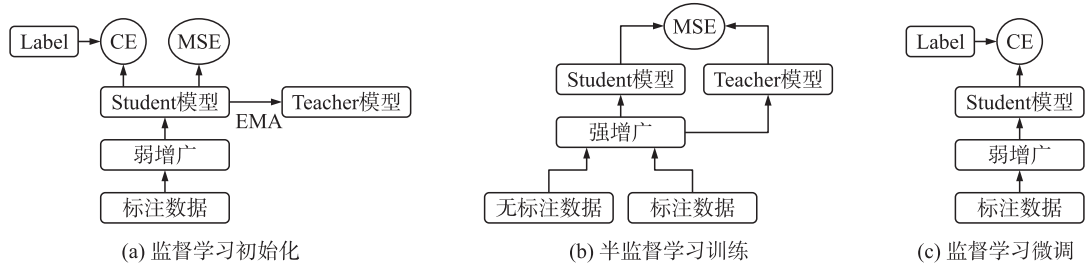


图 1 算法流程图

Fig. 1 Algorithm flowchart

### 2.2.1 监督学习初始化

只使用监督数据,做弱数据增广、损失函数为交叉熵以及均方差一致性正则,计算公式为:

$$\text{Loss} = \text{Loss}_{\text{classification}} + \text{Loss}_{\text{consistency}}; \quad (1)$$

$$\text{Loss}_{\text{classification}} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k y_{ij} * (\hat{y}_{ij}); \quad (2)$$

$$\text{Loss}_{\text{consistency}} = \begin{cases} \frac{1}{n} \sum_{i=1}^n (\tilde{y}_{\text{raw}} - \tilde{y}_{\text{aug}})^2, & \text{if } \tilde{y}_{\text{aug}}[\text{label}] \geq \text{threshold}; \\ \text{ignore}, & \text{if } \tilde{y}_{\text{aug}}[\text{label}] < \text{threshold}. \end{cases} \quad (3)$$

式中,  $\text{Loss}_{\text{classification}}$  为分类损失函数,此处采用交叉熵计算;  $\text{Loss}_{\text{consistency}}$  为一致性正则损失函数,为避免出现过度增广,添加了阈值筛选机制,只选择对应类别的分类得分较高的样本来计算一致性正则,阈值 threshold 取值为 0.5;  $\tilde{y}_{\text{aug}}$  是数据增广后的数据经模型计算得到的输出;  $\tilde{y}_{\text{raw}}$  是未经数据增广的数据经模型计算后得到的输出; label 为监督数据标签. 在训练的同时,使用指数滑动平均(exponential moving average, EMA)的方式生成 Teacher 模型备用<sup>[14]</sup>:

$$W_t = W_t * \alpha + (1 - \alpha) * W_s. \quad (4)$$

式中,  $W_t$  是 Teacher 模型的权重;  $W_s$  是 Student 模型的权重;  $\alpha$  为超参,取值为 0.95,用于控制 Teacher 模型的更新幅度. Teacher 模型在训练过程中会随着 Student 更新,本文 Teacher 模型的更新频率为 100 个 iterations 更新一次.

### 2.2.2 半监督学习训练

使用监督数据 and 无监督数据做强数据增广,由 Teacher 模型生成 one-hot 形式的伪标签,伪标签计算公式为:

$$\text{Label}_{\text{pseudo}} = \begin{cases} -1, & p_{\text{argmax}}(\text{output}) < \text{threshold}; \\ \text{argmax}(\text{output}), & p_{\text{argmax}}(\text{output}) \geq \text{threshold}. \end{cases} \quad (5)$$

式中, output 为 Teacher 模型的输出; threshold 为伪标签的筛选阈值,取值为 0.8. 训练过程中,选择得分较

高的伪标签,未达阈值的样本标签设为-1,分类训练中会被自动忽略. 监督学习和非监督学习通过权重来平衡:

$$\text{Loss} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k y_{s_{ij}} * (\widetilde{y_{s_{ij}}}) + \lambda * \left( -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k y_{u_{ij}} * (\widetilde{y_{u_{ij}}}) \right). \quad (6)$$

式中,  $y_{s_{ij}}$  表示监督学习 (Supervised) 的 label;  $\widetilde{y_{s_{ij}}}$  表示监督学习的模型输出;  $y_{u_{ij}}$  表示非监督学习 (Unsupervised) 的 label;  $\widetilde{y_{u_{ij}}}$  表示监督学习的模型输出;  $\lambda$  为超参,取值为 1.0. 需要注意的是,半监督数据虽然没有标签,却不能视为均衡分布的数据<sup>[15]</sup>,非监督数据的非均衡分布同样会导致模型的过拟合,为了保证训练的稳定性,根据 Teacher 模型生成的伪标签对非监督数据进行重采样.

### 2.2.3 监督学习微调

基于第二步模型,只使用监督数据,做弱数据增广,损失函数为交叉熵分类以及均方差一致性正则. 训练公式与第一步一致. 由于监督学习样本较少,存在两个问题:(1) 第一步中初始化得到的 Teacher 模型效果不佳,自训练时引入的伪标签噪声较多;(2) 第二步中的监督数据被稀释,对模型训练的约束较弱. 需要在监督数据上做“矫正”,才能得到更好的结果<sup>[16]</sup>,流程的伪代码如下:

```
for data_sup, label, data_unsup in balanced_data_loader:
    data_sup, data_unsup = augment( data_sup ), augment( data_unsup )
    if iterations <= sup_iterations:
        student.fit( data, label )
        teacher = ema( student )
    elif iterations > sup_iterations and iterations <= finetune_iterations:
        pseudo_label = teacher.predict( data_unsup )
        student.fit( [ data_unsup, data_sup ], [ pseudo_label, label ] )
    else:
        student.fit( data, label )
```

balanced\_data\_loader 为重采样之后的训练数据; data\_sup 为监督数据; data\_unsup 为非监督数据; augment 代表回译数据增广; pseudo 为 teacher 模型生成的伪标签; ema 代表对 student 模型进行滑动平均从而生成 Teacher 模型. 基线模型选择的是开源项目 Transformers 中的 Bert-Chinese-Base 模型<sup>[17]</sup>. BERT 模型是一种预训练语言模型,主要由 Transformer 编码器和全连接层组成. BERT 的编码器部分是一个标准的 Transformer 结构,由多个自注意力机制、前馈神经网络和残差连接而成. BERT 的预训练过程是基于大规模无监督语言数据进行的,通过最大化上下文中单词之间的语义相似性来学习语言表示.

### 2.3 模型结构

为了适配文本分类任务,本文在 BERT 模型后面添加了一个三层线性层,每个线性层都配有一个 ReLU 激活函数,以保证对非线性特征的分类能力,其具体结构如图 2 所示.

BERT 模型的基本结构包括以下组件:

输入嵌入层 (Input Embedding): 作为模型的入口点,负责将原始样本中的评论文本通过一个矩阵映射为 768 维固定长度的向量表示,把单词从原始的形式转换成一种模型可以处理的数字形式.

位置编码 (Positional Encoding): 由于 BERT 需要理解序列中单词的顺序,位置编码用于对文本中每个字的位置进行编码,使用正弦 (sin) 和余弦 (cos) 函数生成一系列固定值,这些值与词嵌入向量相结合,以使模型能够捕捉序列中的位置信息.

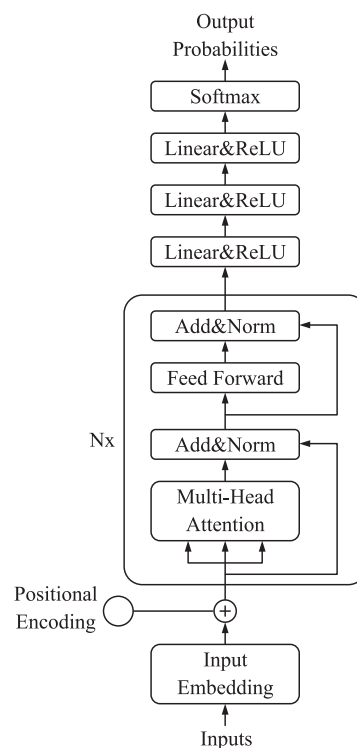


图 2 分类网络结构图

Fig. 2 Network structure



多头自注意力层(Multi-Head Self-Attention):是 BERT 模型的核心部分之一,由两个线性变换和一个非线性变换组成.在多头注意力机制中,模型首先将输入序列通过线性变换得到 Query(查询)、Key(键)和 Value(值).通过计算 Query 和 Key 之间的点积,再除以缩放因子,应用 Softmax 函数来得到注意力权重分布,将注意力权重分布与 Value 相乘,得到加权的 Value 表示.该过程允许模型在不同的"头"中关注输入序列的不同部分,最后将所有"头"的输出拼接起来形成最终的输出,使得模型能够同时考虑多个位置的信息,从而改善了模型的整体性能.

前馈神经网络(Feed-Forward Neural Network):被用来进一步处理位置编码和词嵌入向量相加得到的最终的特征向量,通常由多个全连接层和一个非线性激活函数(如 ReLU 或 GeLU)组成,能够对输入数据进行更深层次的抽象和特征提取.

残差连接(Residual Connection):为了解决深度神经网络训练过程中可能出现的梯度消失或爆炸问题,BERT 模型采用了残差连接,其将一个或多个层之前的输出直接添加到当前层的输出上,允许原始的信号直接绕过一个或多个层,从而保持梯度的流动和传播.这意味着即使深层的梯度由于某些原因变得极小或极大,通过残差连接,梯度仍然可以通过其他路径有效传播,避免了极端情况下的梯度消失或爆炸问题.

层归一化(LayerNorm):与 BatchNorm 不同,LayerNorm 不依赖于 batch size,可以为每一层的每一个样本独立地计算均值和方差,使得 LayerNorm 适用于动态变化的 batch size,并允许模型在处理不同大小的输入时保持一致的性能.这对于 NLP 任务至关重要,因为文本数据的长度通常是动态且多样的,且很多情况下需要处理的是单个样本而不是成批的样本.此外,LayerNorm 还有助于模型的泛化能力.由于归一化是在每个样本的特征上独立进行的,减少了模型对于 batch 中特定样本分布的依赖,使得模型在不同的数据集和任务上更加鲁棒.

BERT 模型的结构设计使其能够有效地理解和处理自然语言数据,通过结合位置编码、多头自注意力机制、前馈神经网络、残差连接以及层归一化等技术,BERT 在各种 NLP 任务中均取得了卓越的表现.

### 3 实验方案

#### 3.1 数据集

自建数据集为来自中国大学 MOOC 的课程评论文本,共 10 万条,其中人工标注数据 8 000 条训练集和 2 000 条测试集,用于监督学习和测试;剩余 9 万条为无标注数据,用于半监督训练.测试集中包含:正面评论、负面评论、中性评论和无关评论 4 个类别各 500 条,训练数据中同样包含正面评论、负面评论、中性评论和无关评论各 2 000 条.标注信息来自课程评论中的星标等级,并经过人工核验.

在监督学习初始化训练过程中会对标注数据进行弱增广,半监督学习训练过程中会分别对标注数据和未标注数据进行强增广,监督学习微调训练过程中会对标注数据进行弱增广.表 1 所示为部分数据增广前后的对比示例.

表 1 数据增广前后的文本样例

Table 1 Examples of text before and after data augmentation

原始文本	标注类别	弱增广		强增广	
		文本	是否扭曲语义	文本	是否扭曲语义
蛮不错的,挺好	正面	很好. 很好	否	好吧,好	否
念 ppt 的无语	负面	阅读 ppt 无语	否	看 ppt	负面变中性
还行还行	中性	不算太坏	否	还好	否
马马虎虎,不太推荐	负面	一般般,不太推荐	否	我不希望你粗心大意	负面变中性
不可多得的课程	正面	珍贵的课程	否	罕见的教训	正面变负面

从表 1 所示样例可以看出,弱增广之后的数据与原文含义相近;而强增广前后的数据差异较大,存在语义扭曲的情况.在强增广条件下,可以产生更丰富的样本,但增广后的数据标签会变得不可用,所以需要采用半监督训练来适配强增广手段.

#### 3.2 评价指标

本文选择的情感分类评价指标为  $F1$ -score,其为精确率和召回率的调和平均数,最大为 1,最小为 0.

$F1$ -score 认为召回率和精确率同等重要,因此在计算时会将两者进行调和平均,计算公式为:

$$F1\text{-score} = \frac{2 * p * r}{p + r}. \quad (7)$$

式中, $p$  表示预测为正例的样本中真正正例的比例; $r$  表示正确识别为正例的样本占总的正例样本数量的比例:

$$p = \frac{TP}{TP + FP}, \quad (8)$$

$$r = \frac{TP}{TP + FN}. \quad (9)$$

式中,TP、FP、FN、TN 为机器学习中常用的指标,用于衡量分类模型的性能. 其中,TP (True Positive) 表示真正例数,即模型预测为正例且实际为正例的数量;FP (False Positive) 表示假正例数,即模型预测为正例但实际为负例的数量;FN (False Negative) 表示假负例数,即模型预测为负例但实际为正例的数量;TN (True Negative) 表示真负例数,即模型预测为负例且实际为负例的数量.

多分类中计算  $F1$ -score 有两种方式,分别是 Micro- $F1$  和 Macro- $F1$ ,由于本文测试集的各类别数量一致,所以两种方式的计算结果一致,可任选其一,本文选择了 Macro- $F1$ :

$$\text{Macro-}F1 = \frac{1}{n} \sum_{i=1}^k F1\text{-score}_i. \quad (10)$$

### 3.3 结果对比与分析

为了验证过度数据增广对监督学习的负面作用,进行了弱增广和强增广两组对比实验.

从图 3 可以看出,弱数据增广能有效提升监督学习的效果,而强数据增广会对监督学习产生负面作用,尤其是数据量较少的情况下影响尤为明显,但当数据增大到一定规模时,数据增广扭曲语义的问题可以得到缓解. 为了消除这种影响,本文首先在 TS-Aug 方案下利用监督数据对第一轮的非增广模型进行微调训练,并分别实施无增广、强增广、弱增广 3 种策略,解决了过度数据增广扭曲语义的问题,得到的效果如图 4 所示.

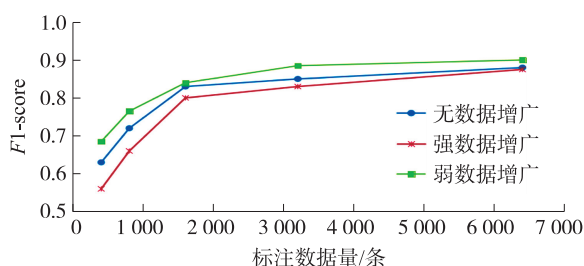


图 3 数据增广强度对监督学习的影响

Fig. 3 Impact of data augmentation intensity on supervised learning

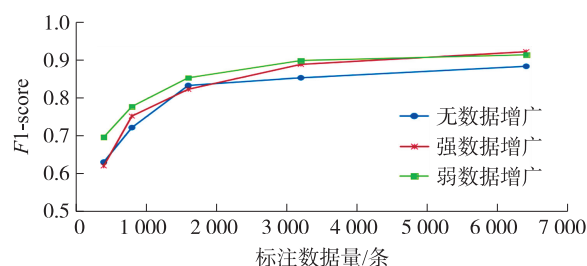


图 4 TS-Aug 方案下数据增广对监督学习的影响

Fig. 4 Impact of data augmentation under TS-Aug on supervised learning

由图 4 可知,在使用 Teacher 指导 Student 模型学习之后,强数据增广的负面作用被消除,且在数据量增大到 6 000 条以上后,强增广的效果与弱增广接近. 对于无监督学习来说,无标注数据获取成本很低,为了探索继续增加数据量之后强增广能否超越弱增广的效果,充分利用所有的标注数据,增量添加无标注数据以期实验更符合实际应用场景. 本文在图 3 使用了 6 400 条标注数据的无数据增广模型的基础上,加入大量无标注数据,进行带无监督数据的 TS-Aug 半监督自训练,得到训练结果如图 5 所示. 可以看出,当无标注数据与标注数据比例接近 1:1 时,3 种数据增广方式的效果差异不大;当无标注数据的比例逐渐提升,达到足够规模后,强数据增广不再对模型训练起负面作用,反而能提升模型效果.

本文对比了弱增广监督学习实验组与 TS-Aug

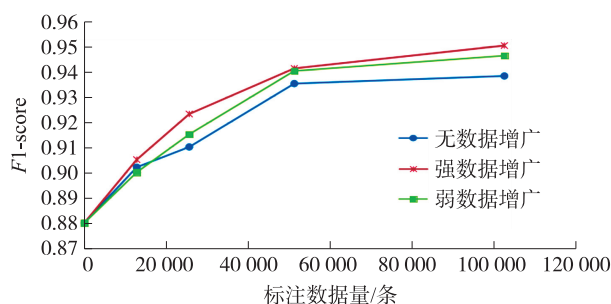


图 5 TS-Aug 与无标注数据训练的效果

Fig. 5 Impact of training with unlabeled data and TS-Aug

半监督训练组的特征分布,如图 6 所示.从图 6 可以看出,加入无标注数据,并使用 TS-Aug 训练策略之后,数据类内分布变得更紧凑,类间距变大,这些都得益于强数据增广带来的样本丰富度的提升.监督学习和 TS-Aug 非监督学习的混淆矩阵对比图如图 7 所示.在监督学习中,预测结果更偏向于正面和负面评论,这两种评论情感鲜明且形式单一,易于学习.在 TS-Aug 半监督学习中,预测结果更加均衡,且无关评论和负面评论的提升最为显著,这两种评论比正面评论和负面评论的内容更丰富、更开放,说明强增广加半监督的手段能够提升模型的集外鲁棒性.

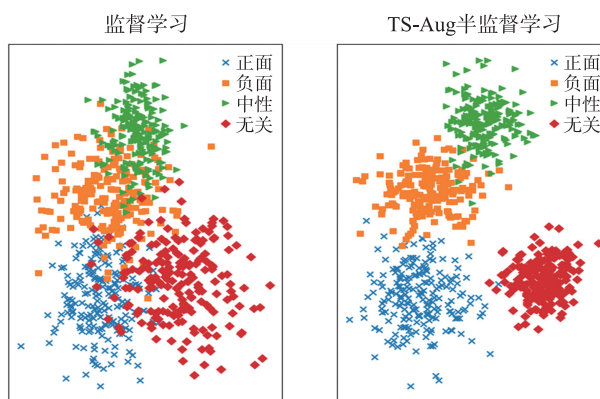


图 6 测试样本的特征分布图

Fig. 6 Distribution of test sample features

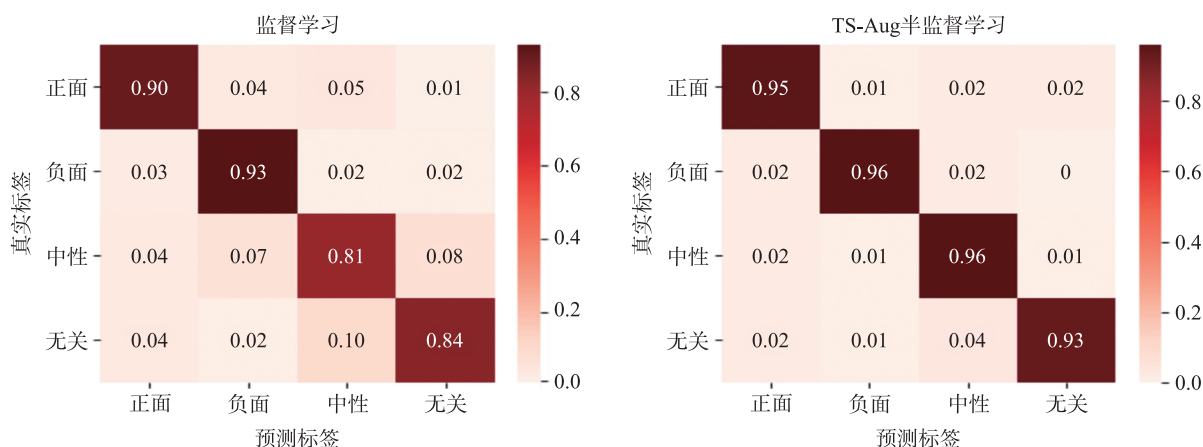


图 7 监督学习与 TS-Aug 半监督学习的分类混淆矩阵图

Fig. 7 Classification confusion matrix of supervised learning and TS-Aug semi-supervised learning

## 4 结论

本文分析了过度数据增广对小样本情感分类任务的负面作用,设计了 TS-Aug 半监督自训练框架,引入大量无监督数据,通过自训练的方式消除过度数据增广的负面影响,使得样本在特征空间中的分布更加紧凑,可分性更强,得到更稳定的少样本半监督情感分类训练方案, $F1$ -score 值从 0.88 提升至 0.95. 本文的半监督自训练技术作为可以同时有效利用有监督数据和无监督数据的学习范式,能降低数据利用门槛,减小数据标注成本,对深度学习在情感分类中的应用研究有重要意义.

## [参考文献] (References)

- [1] 余游,冯林,王格格,等. 一种基于伪标签的半监督少样本学习模型[J]. 电子学报,2019,47(11):2284-2291.
- [2] LEE D H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks[C]//Proceedings of the CML 2013 Workshop on Challenges in Representation Learning. Atlanta, USA: ICML, 2013.
- [3] FINI E, ASTOLFI P, ALAHARI K, et al. Semi-supervised learning made simple with self-supervised clustering[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE, 2023.
- [4] CHEN B X, JIANG J G, WANG X M, et al. Debaised self-training for semi-supervised learning[J/OL]. arXiv Preprint arXiv: 2202.07136, 2022.
- [5] 鲍兆强,王立宏. 基于伪标签纠正的半监督深度子空间聚类[J]. 烟台大学学报(自然科学与工程版), 2023, 36(4): 442-450.
- [6] YANG H F. Contrastive self-supervised learning as a strong baseline for unsupervised hashing[C]//Proceedings of the 2022

- IEEE 24th International Workshop on Multimedia Signal Processing(MMSP). Shanghai, China;IEEE,2022.
- [7] DUAN Y, QI L, WANG L, et al. RDA: Reciprocal distribution alignment for robust semi-supervised learning [C]// Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel;ECCV,2022.
- [8] 廖凌湘,冯林,刘鑫磊,等. 基于信息对齐的半监督少样本学习方法[J]. 计算机工程与设计,2023,44(2):582-589.
- [9] SOHN K,BERTHELOT D,LI C L,et al. Fixmatch:Simplifying semi-supervised learning with consistency and confidence[J]. Advances in Neural Information Processing Systems,2020,33:596-608.
- [10] 宋雨,肖玉柱,宋学力. 基于伪标签回归和流形正则化的无监督特征选择算法[J]. 南京大学学报(自然科学版),2023,59(2):263-272.
- [11] XIE Q Z, DAI Z H, HOVY E, et al. Unsupervised data augmentation for consistency training[J]. Advances in Neural Information Processing Systems,2020,33:6256-6268.
- [12] WEI J,ZOU K. Eda:Easy data augmentation techniques for boosting performance on text classification tasks[J/OL]. arXiv Preprint arXiv:1901.11196,2019.
- [13] SUGIYAMA A,YOSHINAGA N. Data augmentation using back-translation for context-aware neural machine translation[C]// Proceedings of the 4th Workshop on Discourse in Machine Translation(DiscoMT 2019). Hong Kong, China;DiscoMT,2019.
- [14] TARVAINEN A,VALPOLA H. Mean teachers are better role models:Weight-averaged consistency targets improve semi-supervised deep learning results[C]//Proceedings of the 31st International Conference on Neural Information Processing system. Long Beach, USA;NIPS,2017.
- [15] REN Z Z,YEH R A,SCHWING A G. Not all unlabeled data are equal:Learning to weight data in semi-supervised learning [J/OL]. arXiv Preprint arXiv:2007.01293,2020
- [16] SUN Z J,FAN C,SUN X F,et al. Neural semi-supervised learning for text classification under large-scale pretraining[J/OL]. arXiv Preprint arXiv:2011.08626,2020.
- [17] DEVLIN J,CHANG M W,LEE K,et al. Bert:Pre-training of deep bidirectional transformers for language understanding[J/OL]. arXiv Preprint arXiv:1810.04805,2018.

[责任编辑:严海琳]