

基于迁移学习和集成学习的医疗文本分类

郑承宇,王 新,王 婷,徐权峰

(云南民族大学 数学与计算机科学学院,云南 昆明 650500)

摘 要:针对医疗文本语义稀疏、维度过高的问题,提出一种基于迁移学习和集成学习的多标签医疗文本分类算法(Trans-LSTM-CNN-Multi, TLCM)。该算法采用 ALBERT(A Lite BERT)模型内部的多层双向 Transformer 结构对大型语料库展开训练,获取通用领域的文本动态字向量表示。然后,利用医学领域目标数据集通过迁移学习和模型微调技术实现 ALBERT 预训练语言模型在医学领域的文本语义增强。在此基础上,将上述通过迁移学习得到的文本语义增强模型输入到 Bi-LSTM-CNN 集成学习模块,进一步提取医学文本内容的重要信息特征。最后,基于二元交叉熵损失函数构造文本多标签分类器实现医疗文本分类。实验结果表明,通过迁移学习和集成学习的 TLCM 文本分类算法能有效提升医疗文本的分类性能,在中文健康问句数据集上整体 F1 值达到了 91.8%。

关键词:迁移学习;集成学习;ALBERT;Bi-LSTM-CNN;医疗文本;健康问句

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2022)04-0028-06

doi:10.3969/j.issn.1673-629X.2022.04.005

Medical Text Classification Based on Transfer Learning and Ensemble Learning

ZHENG Cheng-yu, WANG Xin, WANG Ting, XU Quan-feng

(School of Mathematics and Computer Science, Yunnan Minzu University, Kunming 650500, China)

Abstract: Aiming at the problems of sparse semantic and high dimension of medical text, a multi-label medical text classification algorithm based on transfer learning and ensemble learning named TLCM (Trans-LSTM-CNN-Multi) is proposed. Firstly, the large-scale corpus is trained through the multi-layer Transformer structure inside the ALBERT (A Lite BERT) model to obtain the dynamic word vector representation of the text. Then, the target data set in the medical field is used to realize the text semantic enhancement in the medical field through transfer learning and model fine-tuning technology based on ALBERT (A Lite BERT) pre-training language model. On this basis, the above-mentioned semantic enhancement model obtained through transfer learning is input to the Bi-LSTM-CNN ensemble learning module to further extract important information characteristics of medical text content. Finally, a text multi-label classifier based on binary cross entropy loss function is constructed to achieve medical text classification. The experimental results show that the text classification algorithm through transfer learning and ensemble learning can effectively improve the overall performance of the model, and finally the overall F1 value on the Chinese health question data set reaches 91.8%.

Key words: transfer learning; ensemble learning; ALBERT; Bi-LSTM-CNN; medical text; health question

0 引言

挖掘隐藏在医疗文本中的丰富语义信息是文本挖掘领域的重要任务之一。医疗文本具有复杂的专业术语等现象,常见的文本分类模型在处理医疗文本时常常面临文本语义稀疏、知识利用不足等问题。传统的主流机器学习算法如 ML-KNN^[1]、ML-DT^[2]、Rank-SVM^[3]在处理数据稀疏、维度过高的大规模数据集时,效果并不理想。近年来,基于大数据驱动的深度学习

习算法如 CNN、RNN 以及其他混合神经网络模型取得了快速发展,在情感分析^[4]、信息检索^[5]、问答系统^[6]等领域得到了广泛应用。基于深度学习算法的文本分类模型为了取得优异的性能,需要大量的时间和精力去采集目标领域的大型数据集和进行模型的参数训练过程,且易造成巨大的计算资源耗费。为了解决这一问题,该文引入文本迁移学习方法,将在通用领域表现较好的预训练语言模型,通过模型微调技术,迁移

收稿日期:2021-04-22

修回日期:2021-08-25

基金项目:国家自然科学基金资助项目(61363022);云南省教育厅科学研究基金项目(2021Y670)

作者简介:郑承宇(1996-),男,硕士研究生,研究方向为数据挖掘、自然语言处理;王 新,教授,硕导,CCF 会员(36839M),研究方向为数据挖掘、推荐系统、机器学习。

到特定目标医学领域数据集上进行文本迁移学习,并微调网络结构,针对医疗文本语义稀疏、维度过高的特点,采用混合神经网络进行集成训练,解决传统文本分类算法对医疗文本语义理解不足的问题。

1 相关研究

2013年,Mikolov等^[7]提出Word2Vec词向量表征模型,推动了深度学习算法在自然语言处理领域中的发展。随后,基于卷积神经网络(convolutional neural network, CNN)、循环神经网络(recurrent neural network, RNN)和长短期记忆网络(long-short term memory network, LSTM)的文本分类模型被广泛应用到句子分类、多模态分析和语义关联等自然语言处理任务中^[8-9]。在医疗文本多标签分类领域,Chen等^[10]提出一种卷积神经网络与循环神经网络在多标签文本分类中的集成应用模型CNN-RNN,同时捕获医疗文本词汇信息和上下文语义信息。Ibrahim^[11]等提出一种用于多标签生物医疗文本分类的混合神经网络模型,分别利用CNN和LSTM模块提取医疗文本最具区分度的特征和上下文信息特征,然后进行特征融合。上述算法结合CNN、RNN和LSTM提取文本特征的优点,有效提升了深度学习算法对医疗文本的分类效果。

Word2Vec词向量表征打破了文本相似词汇不同表达的隔阂,训练完成的词向量表达包含不同文本的词意信息,但面临文本词向量表示无法根据具体语境进行动态调整的问题,从而造成文本语义信息损失。基于此,越来越多的研究者开始采用递进式网络结构进行堆叠式搭建,提取文本特征时充分利用单一网络优势。2018年,Du等^[12]提出一种基于嵌入式语言学习模型(embeddings from language models, ELMO)的深层神经网络学习框架,结合具体语境对文本向量表达进行动态调整,并利用双向长短期记忆网络(bidirectional long-short term memory network, Bi-LSTM)和Attention注意力机制捕获文本上下文的语

义关联,用于生物医学文本的多标签分类,取得了不错的效果。ELMO词向量表征模型可以有效学习文本上下文语义信息进行动态词向量表征^[13],但仍面临着梯度消失以及无法并行化等问题。

针对上述问题,2018年,Devlin等^[14]提出一种基于双向Transformer结构表示的文本预训练语言模型(bidirectional encoder representations from transformers, BERT),利用Transformer结构中的encoder部分代替传统的LSTM模块,在文本向量表示方面展示出了强大的算法性能。BERT模型作为一种高效的文本向量表示方法,其有效性依赖于大量的模型参数训练,需要耗费大量的时间和资源成本。2020年,Lan等^[15]提出了一种轻量化网络结构ALBERT(A Lite BERT),通过采用几种优化策略获得一种轻量化网络模型,在多项自然语言处理任务数据集上超越了BERT模型。基于上述研究,该文提出一种基于迁移学习和集成学习的医疗文本分类算法,从通用领域表现较好的ALBERT预训练语言模型中,通过模型微调技术实现医疗文本的语义增强,然后设计Bi-LSTM-CNN混合神经网络模块进行集成训练,提取医疗文本蕴涵的丰富语义特征,实现医疗文本的深层语义信息抽取。

2 迁移和集成框架

2.1 ALBERT迁移学习框架

ALBERT预训练语言模型在通用领域的中文文本向量表示方面展现出了优越的算法性能,通过采用多层双向的Transformer结构^[16]设计,对大量中文语料库进行训练,得到融合词汇本身和上下文语义信息的动态向量表示,在一定程度上解决了同义词、多义词等现象对文本语义的影响,其模型结构如图1所示。分析ALBERT预训练语言模型的网络结构,通过迁移学习和模型微调技术,将其应用到特定的目标领域,结合医疗领域文本数据集进行训练,可以实现医疗文本的语义增强。

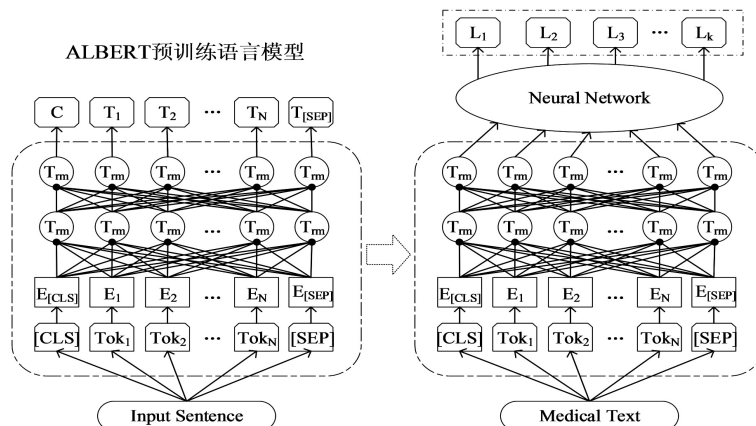


图1 ALBERT模型迁移学习框架

文本领域的迁移学习主要是调整预训练语言模型的第一层网络结构,将在通用领域性能表现较好的预训练网络模型迁移到特定医学领域数据集上进行训练。ALBERT 模型在句子层面按照字符级别进行文本向量输入,基于迁移学习和模型微调技术,使得该模型在特定领域也具有较好的算法性能,在实际应用中具有重要的应用价值。

ALBERT 模型采用多层双向 Transformer 结构设计,多头自注意力机制^[17]是最重要的组成部分,可以计算文中每个词与该句子中所有词的相互关系,并通过这种相互关系调整每个词在句子中的权重从而获得新的表达,其计算原理如下:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h) \mathbf{W}^o \quad (1)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V) \quad (2)$$

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}(\mathbf{Q}\mathbf{K}^T / \sqrt{d_k}) \mathbf{V} \quad (3)$$

$$\text{Softmax}(z)_j = e^{z_j} / \sum_{k=1}^N e^{z_k} \quad (4)$$

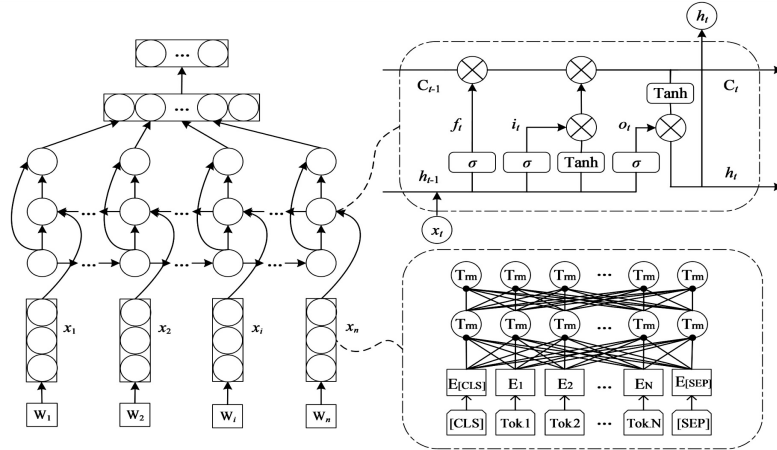


图 2 Bi-LSTM-CNN 集成学习框架

Bi-LSTM 模型考虑了文本的词序信息和语义信息,对带有上下文依赖关系的文本具有很好的建模能力,可适用于各种特定任务的模型构建,是目前常用的文本语义编码模型。Bi-LSTM-CNN 集成学习框架基于 RCNN^[18]混合神经网络模型框架进行搭建,在双向长短期记忆网络后面连接一个卷积层和最大池化层,构成 Bi-LSTM-CNN 集成网络模块,解决了 RNN 循环神经网络梯度消失或梯度爆炸的问题,且充分考虑当前字符所在上下文的语义信息,从而学习文本的全局结构信息。

Bi-LSTM 包含与 LSTM 相同的遗忘门、输入门和输出门三个细胞状态,这三个门协同工作,进行信息的筛选、更新与输出,各细胞状态的更新如下:

$$f = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}) \quad (5)$$

$$i = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}) \quad (6)$$

其中, $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 向量分别由网络输入层的文本序列的每个词的 query、key 和 value 组成,利用多头注意力机制可以将多个网络层的输出矩阵拼接成一个大的词向量矩阵; \mathbf{W}^o 是一个附加权重矩阵,使得拼接后的矩阵维度被统一压缩成序列长度大小; $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V$ 代表每个 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 向量的权重矩阵; d_k 代表每个 $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ 向量的维度大小。

2.2 Bi-LSTM-CNN 集成学习框架

基于 ALBERT 预训练语言模型的文本迁移学习方法,根据不同任务需要可将最后一层 Transformer 结构的输出向量输入到下游任务的深层神经网络中进行训练。医疗文本作为一种复杂的序列信息,具有语义稀疏、维度过高的特点,Bi-LSTM 模型采用双向的序列结构设计,可以更深入地抽取文本序列的全局结构信息,实现序列信息的复杂关系建模。随后,采用 CNN 卷积神经网络模块对 Bi-LSTM 模块获取的文本语义信息特征进行进一步地抽取,可以获得维度更低的深层语义表达,模型结构如图 2 所示。

$$u = \tanh(W^{(u)}x_t + U^{(u)}h_{t-1} + b^{(u)}) \quad (7)$$

$$c_t = i \odot u + f \odot c_{t-1} \quad (8)$$

$$o = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1} + b^{(o)}) \quad (9)$$

$$h_t = o \odot \tanh(c_t) \quad (10)$$

其中, x_t 表示当前输入单元状态, i, o, f, u 分别表示当前输入门、输出门、遗忘门和候选细胞生成, c_t, h_t 表示细胞状态更新和隐藏状态输出, $\sigma(\cdot)$ 表示激活函数, \mathbf{W} 表示权重矩阵, b 表示偏置项。

3 TLCM 文本分类模型

3.1 TLCM 模型

针对医疗文本语义稀疏、维度过高的特点,该文引入 ALBERT 预训练语言模型进行迁移学习,采用多层双向 Transformer 编码器进行医疗文本的动态的词向量表征。在此基础上,利用 Bi-LSTM 双向长短期记忆

网络和 CNN 卷积神经网络搭建 Bi-LSTM-CNN 集成学习模块,实现文本特征的深层语义信息提取。输出层采用二交叉熵损失和 sigmoid 激活函数进行文本的

多标签分类训练,提出一种基于迁移学习和集成学习的医疗文本多标签分类算法 (Trans-LSTM-CNN-Multi, TLCM),算法流程如图 3 所示。

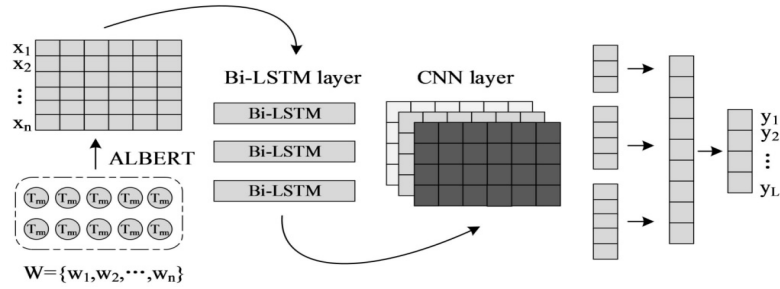


图3 TLCM 模型框架

TLCM 医疗文本分类模型采用 Bi-LSTM 双向长短期记忆网络学习文本序列的上下文依赖关系,即全局结构信息特征,然后利用 CNN 卷积神经网络对 Bi-LSTM 模块获取的文本全局结构信息特征进一步地进行高维特征向量表达,实现更深层次的文本语义抽取,且在一定程度上可以解决医疗文本语义信息表达维度过高、知识利用不足的问题。

3.2 算法流程

基于迁移学习和集成学习的多标签医疗文本分类算法流程如下:

算法 1: 基于 TLCM 模型的医疗文本多标签分类算法。

输入: 训练数据集 $T = \{(w_k, y_k^j)_{j=1}^L\}_{k=1}^N$, 模型参数 θ 。

输出: 各训练参数确定的 TLCA 文本多标签分类网络。

(1) $W^0 = \{w_k = 1/N | k = 1, 2, \dots, N\}$ // 初始化 N 个样本的权值

(2) repeat // 迭代轮次, 重复训练

(3) for all $w = \{(W_k, y_k^j)_{j=1}^L\}_{k=1}^N \in D$ do

(4) for k in range(N)

(5) $w_k = \{w_1, w_2, \dots, w_n\}$ // 读取文本数据 w_k

(6) $T_k = \text{embedding}(w_k)$ // 文本序列化

(7) for o in range(O) // 遍历第 o 层 Trans

(8) $X_k^o = \text{transformer}(T_k)$

(9) end for

(10) for p in range(P) // 遍历第 p 层 Bi-LSTM

(11) $\vec{h}_i^p = \overrightarrow{\text{LSTM}}(X_k^o)$ // 前向传递模块 \vec{h}_i

(12) $\overleftarrow{h}_i^p = \overleftarrow{\text{LSTM}}(X_k^o)$ // 后向传递模块 \overleftarrow{h}_i

(13) $h_i^p = [\vec{h}_i^p, \overleftarrow{h}_i^p]$

(14) end for

(15) for q in range(Q) // 遍历第 q 层卷积模块

(16) $v^q = \text{Convolution}(h_i^p)$ // 卷积

(17) end for

(18) $\mu = \text{MaxPooling}(v^q)$ // 最大值池化

(19) $\bar{w} = \text{FullyConnected}(\mu)$ // 全连接层

(20) $y_k^j = \sigma(\bar{w})$ // 计算当前预测输出 y_k^j

(21) $E = \frac{1}{2} \sum_{k=1}^N \sum_{j=1}^L (y_k^j - y_k^j)^2$ // 计算输出误差 E

(22) end for

(23) 根据 E 计算各参数的梯度

(24) 更新 k 个样本的权值 W^o

(25) end for

(26) until 训练误差 E 收敛到一个较小的值

(27) end

4 实验结果及分析

4.1 实验数据

实验使用的语料库来源于中华医学会提供的 5 000 条公共医疗中文健康问句数据, 每一条中文健康提问数据已被标注为多个医学类别, 且具有 0 个至 6 个不确定的标签数量, 属于典型的多标签分类任务数据集。根据实验需求, 将数据集按照 4 : 1 的比例划分为训练集与验证集。该数据集的样本标签分布统计如表 1 和表 2 所示。

表 1 样本标签个数分布

标签个数	样本数量
$n = 0$	39
$n = 1$	3 550
$n = 2$	1 232
$n = 3$	167
$n = 4$	12

表 2 样本类别分布

标签	所属类别	数量
Category_A	诊断	1 858
Category_B	治疗	3 212
Category_C	解剖学/生理学	1
Category_D	流行病学	746

4.2 评价指标

实验采用 Micro-F1 衡量算法的整体性能, 对于文本分类多标签分类任务, 每条文本文档具有 0 个或多个标签, 每个标签具有“0”或“1”两个类别, Micro-F1 是测试不平衡数据集的重要指标。

$$\text{Micro-F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (12)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (13)$$

Micro-F1 同时考虑了所有标签整体精确率 (Precision) 和召回率 (Recall), 是两者的调和平均值。其中, Precision 表示预测结果为正例的样本里面预测正确的样本所占的比例, Recall 表示实际结果为正例

的样本里面预测正确的样本所占的比例。

4.3 参数设置

该文提出一种基于迁移学习和集成学习的医疗文本分类算法, 模型参数主要包括 ALBERT、Bi-LSTM、CNN 三个模块的参数, 依次调整各个网络模块参数的数值, 可以得到模型的最优超参数。在实验过程中, 迁移学习采用 Google 发布的中文版预训练语言模型, Bi-LSTM-CNN 集成框架采用 ALBERT 模型进行词嵌入表示。模型训练相关参数详细展示如表 3 所示。

表 3 模型参数设置

模型参数	参数(值)	模型参数	参数(值)	模型参数	参数(值)
Embedding size	128	filter 大小	(3,4,5)	学习率	5e-5
Hidden size	4 096	每个 size 下的 filter 个数	128	Batch_size	128
Num attention heads	64	激活函数 (BiLSTM、TextCNN)	ReLU	迭代轮数	32
Num hidden layers	12	池化策略	1_max pooling	最大序列长度	202
词表大小	31 128	Dropout rate	0.1	损失函数	交叉熵
激活函数 (ALBERT)	gelu	L2 正则化	3	优化器	Adam

4.4 实验结果

为了验证所提 TLCM 模型的分类效果, 该文做了 4 组对比实验, 其中 2 组基于 Word2Vec 词向量表示, 另外 2 组基于 ALBERT 预训练语言模型, 分别集成 Bi-LSTM 和 CNN 模型进行对比实验, 每组实验采用相同的测试集和验证集。为了比较模型性能的差异, 对照组实验的 Bi-LSTM 和 CNN 模块涉及参数与该文保持一致。实验采用 Precision、Recall 和 Micro-F1 值作为算法性能评价指标, 实验结果如图 4 和图 5 所示。

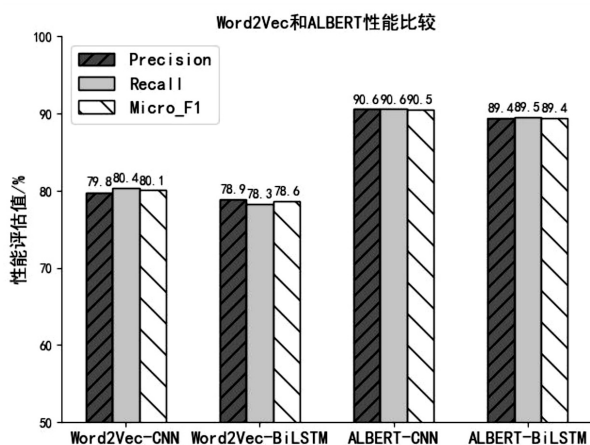


图 4 基于 Word2Vec 和 ALBERT 的算法性能对比

图 5 展示了各模型的分类结果。实验结果表明, 单独引入 Bi-LSTM 和 CNN 模块进行集成训练的 ALBERT 模型, Micro-F1 值分别为 89.4% 和 90.5%, 该文所提基于迁移学习和集成学习的文本分类模型在上述数据集上, Precision、Recall 表现出了优良的算法性能, 三项评测指标分别达到了 91.7%、91.8% 和 91.8%, 说明 Bi-LSTM-CNN 集成模块充分结合两者

的优点, 表现出了优良的算法性能。此外, 如图 4 所示, 基于传统 Word2Vec 词向量表示的文本分类模型在上述数据集中, 模型预测 Micro-F1 值分别为 80.1%、78.6%, 显著低于 ALBERT 预训练语言模型, 这是因为 ALBERT 预训练语言模型可以结合具体语境对文本向量表达进行动态调整, 进一步验证了迁移学习在医疗文本分类领域应用的可行性。此外, 实验采用中文健康问句语料库, 数据规模较小, 标签分布严重不均衡, 一定程度上影响了模型的分类效果, 增加伪数据、多模态数据等, 调整模型相关超参数可进一步提升该算法的多标签分类性能。

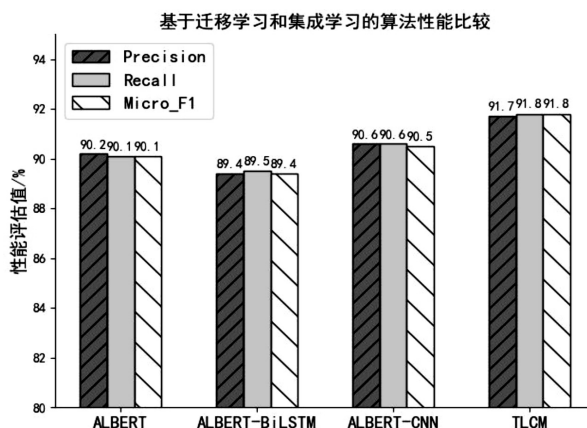


图 5 基于迁移学习和集成学习的算法性能

5 结束语

针对医疗文本语义稀疏、维度过高的问题, 提出了一种基于迁移学习和集成学习的医疗文本分类算法, 迁移学习 ALBERT 预训练语言学习模型, 通过模型微

调技术,实现对目标领域医疗文本的动态词向量表征。在此基础上,构建 Bi-LSTM-CNN 集成学习模块,采用 Bi-LSTM 双向长短期记忆网络学习文本序列的上下文依赖关系,然后利用 CNN 卷积神经网络对 Bi-LSTM 模块获取的文本全局结构信息特征进一步地进行高维特征向量表达,实现更深层次的文本语义抽取。在一定程度上解决了医疗文本语义信息表达维度过高、知识利用不足的问题。实验结果表明,该模型在各项评测指标中均具有良好的表现,基于迁移学习和集成学习的文本分类方法,在小规模医疗文本数据集上表现出了优越的算法性能。在下一步的工作中,扩展模型到多层级和包含隶属关系的医疗文本多标签分类任务中,探讨基于标签序列的文本分类方法,进一步提高模型分类准确率。

参考文献:

- [1] 严玲,周作建,宋懿花,等. 基于 ML-kNN 多标记学习的中医体质辨识模型研究[J]. 世界科学技术-中医药现代化,2020,22(10):3558-3562.
- [2] MORAL-GARCÍAS, MANTAS C J, CASTELLANO J G, et al. Non-parametric predictive inference for solving multi-label classification[J]. Applied Soft Computing, 2020, 88: 106011.
- [3] WU G, ZHENG R, TIAN Y, et al. Joint ranking SVM and binary relevance with robust low-rank learning for multi-label classification[J]. Neural Networks, 2020, 122: 24-39.
- [4] 李磊,吴旭辉,刘继. 融合关键对象识别与深层自注意力的 Bi-LSTM 情感分析模型[J]. 小型微型计算机系统, 2021, 42(3): 504-509.
- [5] 黄名选,蒋曹清,卢守东. 基于词嵌入与扩展词交集的查询扩展[J]. 数据分析与知识发现, 2021, 5(6): 115-125.
- [6] 卢琪,谢艺菲,谢钧,等. 知识图谱在智能问答中的应用研究[J]. 计算机技术与发展, 2021, 31(7): 13-20.
- [7] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv: 1301.3781, 2013.
- [8] 冯进展,蔡淑琴. 融合语义依存关联对的句子相似度的负面在线评论案例推理[J]. 计算机应用研究, 2021, 38(7): 2096-2102.
- [9] LI Y, SONG Y, LUO J. Improving pairwise ranking for multi-label image classification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 1837-1845.
- [10] CHEN G, YE D, XING Z, et al. Ensemble application of convolutional and recurrent neural networks for multi-label text categorization[C]//2017 international joint conference on neural networks (IJCNN). [s. l.]: [s. n.], 2017: 2377-2383.
- [11] IBRAHIM M A, KHAN M U G, MEHMOOD F, et al. GHS-NET a generic hybridized shallow neural network for multi-label biomedical text classification[J]. Journal of Biomedical Informatics, 2021, 116: 103699.
- [12] DU J, CHEN Q, PENG Y, et al. ML-Net: multi-label classification of biomedical texts with deep neural networks[J]. Journal of the American Medical Informatics Association, 2019, 26(11): 1279-1285.
- [13] 赵亚欧,张家重,李贻斌,等. 基于 ELMo 和 Transformer 混合模型的情感分析[J]. 中文信息学报, 2021, 35(3): 115-124.
- [14] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[J]. arXiv: 1810.04805v1, 2019.
- [15] LANZ Z, CHEN M D, GOODMAN S, et al. Albert: a lite bert for self-supervised learning of language representations[J]. arXiv: 1909.11942, 2020.
- [16] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]//Advances in neural information processing systems (NIPS2014). [s. l.]: [s. n.], 2014: 3104-3112.
- [17] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv: 1409.0473, 2016.
- [18] WANG R, LI Z, CAO J, et al. Convolutional recurrent neural networks for text classification[C]//2019 international joint conference on neural networks (IJCNN). [s. l.]: [s. n.], 2019: 1-6.