

文章编号: 1004-4353 (2024) 02-0078-09

基于知识增强的文本分类方法

张博伦¹, 赵亚慧^{1, 2}, 姜克鑫², 卢星华³

(1. 延边大学 融合学院, 吉林 延吉 133002; 2. 延边大学 工学院, 吉林 延吉 133002;
3. 延边大学 外国语学院, 吉林 延吉 133002)

摘要: 为了解决文本分类任务中因部分数据质量差、数据不平衡和数据集过小等原因而导致的分类不准确问题, 提出了一种基于知识增强的文本分类算法。首先, 该算法通过加入外部知识对数据集进行数据增强; 其次, 使用 GloVe 词向量对原始文本和外部知识进行词嵌入, 并使用 CNN、LSTM 和 BERT 模型提取文本特征; 再次, 将提取到的原始文本特征和外部知识文本特征进行融合, 以此得到最终的文本特征; 最后, 将融合后的文本特征送入多层感知机进行分类, 以此得到文本分类的最终结果。在不同数据集上进行实验显示: 在 SST-5 数据集上, 模型 CNN(KB)、LSTM(KB) 和 BERT(KB) 的文本分类准确率比基线模型分别提高了 5.01%、7.92% 和 1.5%; 在 SST-2 数据集上, 模型 LSTM(KB) 和 BERT(KB) 的文本分类准确率比基线模型分别提高了 1.76% 和 1.29%; 在 IMDB 数据集上, 模型 CNN(KB)、LSTM(KB) 和 BERT(KB) 的文本分类准确率比基线模型分别提高了 0.97%、2.87% 和 0.76%。上述结果表明, 该文本分类算法可有效提高文本分类的准确性, 并可为不同领域的文本分类应用提供参考。

关键词: 深度学习; 神经网络; 文本分类; 知识增强; 特征提取

中图分类号: TP391.1 文献标志码: A

Text classification method based on knowledge enhancement

ZHANG Bolun¹, ZHAO Yahui^{1, 2}, JIANG Kexin², LU Xinghua³

(1. College of Integration Science, Yanbian University, Yanji 133002, China;
2. College of Engineering, Yanbian University, Yanji 133002, China;
3. College of Foreign Languages, Yanbian University, Yanji 133002, China)

Abstract: In order to solve the problem of inaccurate classification in text categorization task due to poor quality of some data, data imbalance and too small dataset, a text categorization algorithm based on knowledge enhancement is proposed. Firstly, the algorithm enhances the data set by adding external knowledge. Secondly, the original text and external knowledge are word-embedded using GloVe word vectors and the text features are extracted using CNN, LSTM and BERT models. Thirdly, the extracted original text features and external knowledge text features are fused in order to obtain the final text features. Finally, the fused text features are fed into the multilayer sensing model to obtain the final text features. The experiments on different datasets show that on the SST-5 dataset, the text classification accuracy of CNN(KB), LSTM(KB) and BERT(KB) is improved by 5.01%, 7.92% and 1.5%, respectively, compared with the baseline model, and on the SST-2 dataset, the text classification accuracy of LSTM(KB) and BERT(KB) is improved by 1.76% and 1.5%, respectively, compared with the baseline model. 1.76% and 1.29%, respectively; on the IMDB dataset, the text categorization accuracies of models CNN(KB), LSTM(KB) and BERT(KB) are improved by 0.97%, 2.87% and 0.76%, respectively, over the baseline model. The above results show that the text classification algorithm can effectively improve the accuracy of text classification and can provide good reference for text classification applications in different fields.

Keywords: deep learning; neural networks; text classification; knowledge enhancement; feature extraction

投稿日期: 2023-12-4

基金项目: 国家语委“十三五”科研项目(YB135-76); 延边大学外语语言文学一流学科建设项目(18YLPY13)

第一作者: 张博伦(2001—), 女, 硕士研究生, 研究方向为自然语言处理。

通信作者: 赵亚慧(1974—), 女, 硕士, 教授, 研究方向为智能计算、自然语言处理。

0 引言

目前, 针对文本分类任务的分类方法主要有基于机器学习和基于深度学习两种, 其中深度学习因其能够捕捉到更为丰富的文本特征, 且能够在大量的数据中通过学习来进一步提升其性能而受到学者们的青睐. 为了提高该方法的性能, 近年来一些学者对其进行了改进. 例如: Miyato^[1]采用对抗和虚拟对抗技术进一步提升了长短期记忆网络 (long short-term memory, LSTM) 的分类性能; Chen 等^[2]提出了一种基于表观语义和潜在语义的中文文本分类方法, 研究显示该方法能够很好地挖掘文本的潜在语义; 王海涛等^[3]提出了一种将卷积神经网络 (convolutional neural networks, CNN) 和 LSTM 相结合的提取文本局部特征与输入向量融合的 MLCNN 模型 (Merge-LSTM-CNN), 研究显示该模型能够有效提高文本分类的准确性; 陈可嘉等^[4]提出了 SAttBiGRU-MCNN 文本分类模型, 研究显示该模型可以有效弥补双向门控循环单元 (Bidirectional Gated Recurrent Unit, BiGRU) 模型容易忽略局部特征的缺陷; 孙刘成等^[5]提出了一种基于 LSTM 和注意力机制的文本分类模型, 研究显示该模型捕捉上下文之间关系的能力显著优于 CNN 和 LSTM 模型. 另外, 针对 BERT (bidirectional encoder representations from transformers) 模型, 一些学者对其也进行了多种改进, 如 K-BERT^[6]、ALBERT^[7]、ERNIE^[8]等. 然而, 这些方法和上述相关研究都是通过序列表示或注意力机制来捕捉文本特征与标签之间的关系, 未能充分利用文本背后的外部知识来丰富语义理解和提高分类的准确性.

知识增强方法是一种将外部知识 (如事实、概念、实体关系等) 融入机器学习或深度学习模型中的技术. 研究显示, 使用知识增强方法不仅可以使模型获取所需的外部知识, 还能有效提升模型对文本的理解能力和预测的准确性和鲁棒性^[9-11]. 基于上述研究, 本文为提高文本分类任务的准确率, 提出了一种基于知识增强的文本分类算法, 并在 SST-5、SST-2 和 IMDB 数据集上分别使用 CNN、LSTM 和 BERT 对本文所提算法的有效性进行了验证.

1 基于知识增强的文本分类模型

1.1 模型框架结构

本文构建的基于知识增强的文本分类模型由数据增强层、文本嵌入层、特征提取层、知识融合层和输出层组成, 其具体工作步骤如下:

- 1) 将数据集放入数据增强层中, 以此得到加入外部知识的新数据集. 新数据集中包含原始文本, 以及原始文本对应的外部知识和标签.
- 2) 将原始文本和外部知识送入文本嵌入层进行词嵌入.
- 3) 先利用神经网络模型提取文本特征, 然后将其送入知识融合层, 并采用神经网络中的门控机制控制原始文本特征和外部知识特征的融合比例, 以此得到最终的文本特征.
- 4) 将融合后的数据送入输出层, 使用多层感知机得到文本分类的结果.

1.2 数据增强层

引入数据增强层的目的是通过引入外部知识来提高数据集的质量, 进而增强语言模型的性能. 引入数据增强层的具体操作方法为: 首先, 利用自然语言处理工具对数据集中的单词进行检索, 以此在 Wiktionary 中找到单词对应的单词释义; 然后, 将释义加入数据集中, 将其作为外部知识. Wiktionary 中共包含 999 614 个实体描述. 为保证外部知识的一致性和准确性, 本文将每个实体在 Wiktionary 中的首个释义加入数据集中. 此外, 考虑到有些词汇可能在 Wiktionary 中没有合适的对应形式, 因此本文采用单词的 lemma^[12]形式对其进行处理, 即检索与其词义最接近的词来弥补知识的缺失 (见图 1). 表 1 是 SST-5 数据集集中的两个数

据增强样例, 其中 X_{A1} 、 X_{A2} 表示原文, X_{KA1} 和 X_{KA2} 表示外部知识.

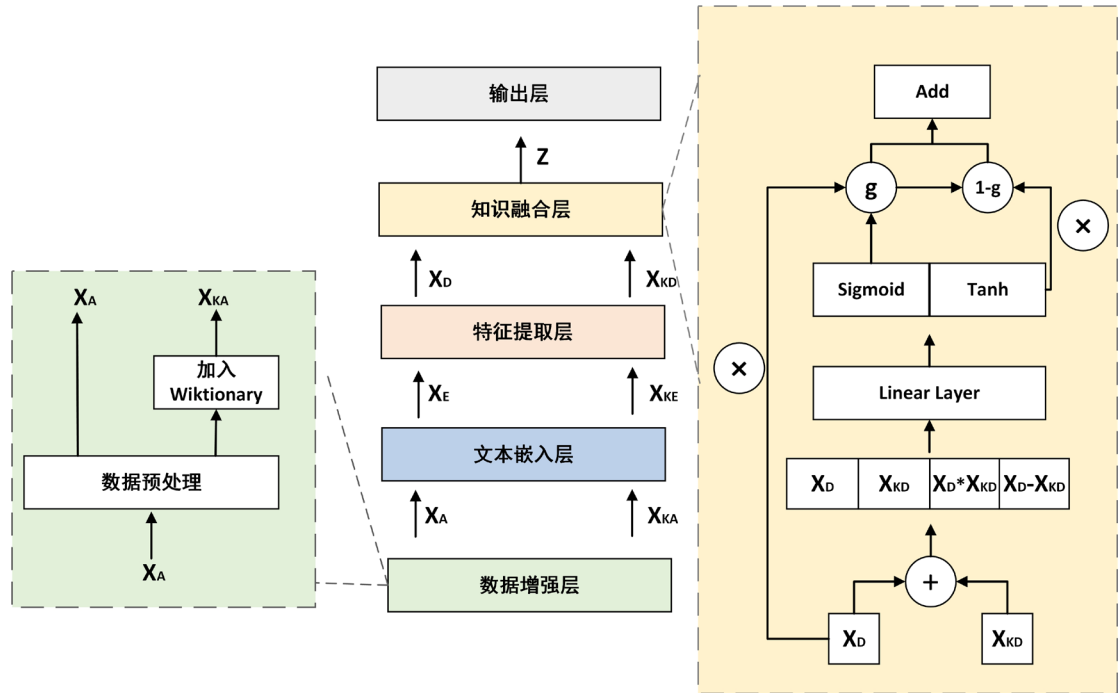


图 1 基于知识增强的文本分类模型图

表 1 数据增强样例

X_{A1}	yet the act is still charming here
X_{KA1}	yet: (nonstandard, West Country) To get.act:(intransitive) To do something. still:(obsolete) To trickle, drip. charming: The casting of a magical carm.
X_{A2}	dramas like this make it human
X_{KA2}	like: (chiefly dialectal, intransitive) To be likely. make: (East Anglia, Essex, obsolete) An agricultural tool resembling a scythe, used to cut (harvest) certain plants such as peas, reeds, or tares. human: (rare) To behave as or become, or to cause to behave as or become, a human.

1.3 文本嵌入层

在处理上述文本 X_A 和外部知识文本 X_{KA} 时, 本文使用 GloVe (global vectors for word representation) 进行词嵌入. GloVe 是一个基于全局词频统计的词表征工具. 由于 GloVe 可通过对大量文本数据进行无监督学习后获得可用于捕捉单词之间语义和语法关系的词向量, 因此它适用于各种自然语言处理任务. 文本嵌入层的表达式为 $X_E = \text{GloVe}(X_A)$, $X_{KE} = \text{GloVe}(X_{KA})$. 其中 X_E 表示对文本 X_A 进行 GloVe 词嵌入后得到的文本词向量; X_{KE} 表示对文本 X_{KA} 进行 GloVe 词嵌入后得到的外部知识词向量. 另外, 在本文提出的基于知识增强的文本分类算法中, 当使用 BERT 提取文本特征时, 由于 BERT 内置了自己的词嵌入机制, 因此该过程不需要单独地对词进行嵌入处理.

1.4 特征提取层

对文本和外部知识进行词嵌入处理后, 本文使用 CNN、LSTM 和 BERT 神经网络进行文本特征提取. CNN 是一种深度学习算法, 其核心思想是利用输入层、卷积层、池化层和全连接层等不同类型的层级组合, 将原始数据转换成更加高级的特征表示, 并通过最后的分类器将特征表示映射到特定类别^[13]. CNN 模型提取文本特征的过程步骤如下:

- 1) 利用卷积层提取出数据特征. 该过程的表达式为 $c_i = f(W_c * [e_{i+k-1} + b_c])$, 其中 e_i 表示词嵌入后第 i 个单词的向量表示, W_c 表示卷积核矩阵, b_c 表示卷积层的偏置项, $f(\cdot)$ 表示卷积操作, k 表示卷积核的大小,

c_i 表示第 i 个卷积核得到的特征图。

2) 利用池化层对输入特征进行映射。该过程本文使用的是最大池化方法, 其表达式为 $p_i = \max(c_{i:i+h-1})$, 其中: h 表示池化操作的大小, p_i 表示第 i 个特征图的池化结果。

3) 利用全连接层将卷积层和池化层提取的特征映射转化为分类结果。该过程的表达式为 $h_j = g(\sum_{i=1}^{n-h+1} W_{fc,j} p_i + b_{fc,j})$, 其中: W_{fc} 表示全连接层的权重矩阵, b_{fc} 表示全连接层的偏置项, g 表示全连接层的激活函数, h_j 表示全连接层输出的第 j 个神经元的结果。

RNN (recurrent neural network, RNN) 是一种能够处理序列数据的神经网络模型。相比于传统的前馈神经网络, 由于 RNN 在处理序列数据时同时考虑了当前的输入数据和历史的输入数据, 因此其具有更好的建模能力, 进而能够捕捉到序列数据中的时间依赖关系和上下文信息。LSTM 模型是 RNN 模型的一种。由于 LSTM 模型在网络结构中引入了多个记忆单元 (LSTM 单元, 每个 LSTM 单元包含 3 个门控单元, 即遗忘门 (forget gate)、输入门 (input gate) 和输出门 (output gate)), 因此其能够有效解决传统 RNN 中存在的梯度消失和梯度爆炸等问题^[14]。LSTM 单元提取文本特征的过程步骤如下:

1) 使用遗忘门来控制上一时刻的隐藏层节点存储的历史信息, 其输入由上一时刻的输出 h_{t-1} 和当前时刻的输入 e_t 构成, 其表达式为 $f_t = \sigma(W_f \cdot [h_{t-1}, e_t] + b_f)$ 。其中 f_t 表示输出门的输出, W_f 和 b_f 分别表示遗忘门的权重和偏置, $\sigma(\cdot)$ 表示 sigmoid 激活函数。

2) 使用输入门来控制当前时刻的输入信息需要被保留到哪些位置, 其包含输入门的输出 i_t 和当前时刻的候选记忆状态 \tilde{C}_t , 其表达式为 $i_t = \sigma(W_i \cdot [h_{t-1}, e_t] + b_i)$, $\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, e_t] + b_c)$, $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$ 。其中 W_i 和 b_i 分别表示输入门的权重和偏置, W_c 和 b_c 分别表示候选记忆状态的权重和偏置, C_t 表示更新的记忆状态, $*$ 表示逐元素乘法。

3) 使用输出门来控制当前隐藏层节点的输入, 确定是否输出给下一隐藏层或者输出层, 其表达式为 $o_t = \sigma(W_o \cdot [h_{t-1}, e_t] + b_o)$, $h_t = o_t * \tanh(C_t)$ 。其中: o_t 表示输出门的输出, h_t 表示 LSTM 单元的输出, W_o 和 b_o 分别表示输出门的权重和偏置。

BERT 是一种基于 Transformer 架构的预训练语言模型, 由于它使用了 masked language model 方法进行预训练, 因此它可以同时考虑文本左侧和右侧的上下文信息, 进而比传统的单向模型能够生成更深层、更丰富的双向语言表示^[15]。BERT 模型提取文本特征的过程步骤如下:

1) 输入文本被分成若干 token, 并添加 [CLS] 用于表示句子的开始和结束, 添加 [SEP] 用于在输入中分隔两个句子。

2) 将这些 token 转换成向量表示, 并在这些向量中嵌入一些位置信息。

3) 对于每个 token, 计算它与其他 token 的交互作用, 并通过多层 Transformer 编码来捕捉文本中的上下文信息。

4) 对于一个句子或一段文本, 取出 [CLS] 对应的向量作为该文本的特征表示, 这个向量可以用于下游任务的输入。

因此, BERT 提取文本特征的表示式为 $f_{\text{BERT}}(X_A) = \text{BERT}(X_A)_{[\text{CLS}]}$ 。其中 X_A 表示输入文本, 因为使用 BERT 提取文本特征时, 无需对文本进行词嵌入, 所以 BERT 模型的输入为原始文本: X_A ; $\text{BERT}(\cdot)$ 表示 BERT 模型的输出向量序列, $\text{BERT}(\cdot)_{[\text{CLS}]}$ 表示取出 [CLS] 对应的向量作为文本的特征表示。

使用 CNN、LSTM 和 BERT 提取文本特征时, 其所组成的神经网络层可表示为 $X_D = M(X_E)$,

$X_{KD} = M(X_{KE})$. 其中 $M(\cdot)$ 表示利用神经网络提取特征; X_D 表示原始文本提取的文本特征, X_{KD} 表示外部知识提取的文本特征.

1.5 知识融合层

知识融合是指将来自不同源头的知识进行整合和转化,生成新的知识. 目前常见的知识融合方法主要有加权平均法和拼接法^[16]. 其中: 使用加权平均法生成的融合数据由于会导致模型不能充分适应数据分布的独特性, 因此该方法会影响模型的应用和准确性; 拼接法生成的融合数据会增加模型的参数, 因此该方法会大幅增加模型计算复杂度. 门控机制是一种能够动态控制信息源之间融合比例的机制, 由于具有能够有效地融合多源知识的优点, 所以近年来被广泛应用于自然语言处理领域^[17]. 因此, 本文采用神经网络中的门控机制控制 X_D 和 X_{KD} 的融合比例, 以此得到本文最终的文本特征表示^[18]. 本文使用的融合函数为:

$$\tilde{x} = \tanh(W_1[x; y; x \odot y; x - y]), \tag{1}$$

$$g = \sigma(W_2[x; y; x \odot y; x - y]), \tag{2}$$

$$z = g \odot \tilde{x} + (1 - g) \odot x. \tag{3}$$

其中: z 表示的是融合后的文本特征表示; W_1 和 W_2 是权重矩阵; g 是一个门控机制, 用于控制 X_D 和 X_{KD} 在输出向量中所占的比重; $\sigma(\cdot)$ 表示 sigmoid 激活函数.

1.6 输出层

本文使用多层感知机获取文本分类的结果, 输出层的表达式为 $y = \text{soft max}(\tanh(ZW + b))$. 其中 W 和 b 都为训练的参数, y 为输入的分类结果.

2 实验及结果与分析

2.1 实验数据

本文使用 SST-5、SST-2 和 IMDB 数据集进行文本分类实验. 各数据集数据分布情况如表 2 所示. 其中: SST 数据集 (Stanford Sentiment Treebank) 是斯坦福大学发布的情感分析数据集, 该数据集由句子组成 (每个句子都被标记为其中一个情感类别), 且句子中涵盖了多种主题, 例如电影、音乐、产品评论, 等等. SST-5 数据集 (Sentiment Analysis on Stanford Treebank) 中包含 5 个情感类别: 非常消极、稍微消极、中立、稍微积极和非常积极. SST-2 数据集是对 SST-5 数据集进行处理后得到的数据集, 即将 SST-5 数据集中的非常消极和稍微消极的样本归类为消极, 将稍微积极和非常积极的样本归类为积极, 同时去除了中立标签的样本. IMDB 电影评论数据集是一个二元情感分析数据集, 由来自互联网电影数据库的 50 000 篇评论组成, 主要包括观众对电影的观感和情感表达. 每部电影包含的评论不超过 30 条, 评论被分为正面情感和负面情感两大类.

表 2 各数据集数据分布情况

数据集	类别数	训练集 / 条	测试集 / 条	验证集 / 条
SST-5	5	8544	2210	1101
SST-2	2	6920	1821	872
IMDB	2	25 000	25 000	—

2.2 对比实验设置

本文使用 Emo2Vec、DAN 和 SWEM-concat 进行对比实验. 其中: Emo2Vec^[19] 是一种基于词向量的情感表示模型, 该模型使用情感词典中的情感标签来扩展每个单词的词向量表示. DAN^[20] 是一种用于文本处理的神经网络模型, 该模型通过对文本中所有词向量相加后取平均值来生成整个文本的表示, 并将得到的文本表示传递给前馈神经网络进行特征提取和分类. SWEM-concat^[21] 是一种基于词向量的文本分类模型, 该模型使用预训练词向量来表示文本中的单词, 并通过池化操作将这些向量合并为文本的表示.

2.3 模型参数设置

本文使用 Python 3.8 进行实验编程和实现, 使用 Pytorch 1.12 作为深度学习框架, 所取的超参数的值如表 3 所示.

表 3 超参数的值

超参数	CNN	LSTM	BERT
词嵌入维度	300	300	—
Epochs	30	30	30
Max_length	128	128	512
Batch_size	48	48	32
Learning_rate	2e-4	2e-5	3e-5

2.4 评价标准

本文使用准确率 (Accuracy) 来评价模型的分类效果, 其表达式为 $A_{\text{Accuracy}} = \frac{r_{\text{TP}} + r_{\text{TN}}}{r_{\text{TP}} + r_{\text{TN}} + r_{\text{FP}} + r_{\text{FN}}}$, 其中 r_{TP} 表示被分类器正确地判定为正例的样本数; r_{TN} 表示被分类器正确地判定为负例的样本数; r_{FP} 表示被分类器错误地判定为正例的样本数; r_{FN} 表示被分类器错误地判定为负例的样本数. 模型文本分类的准确率越高, 表明其文本分类效果越好.

2.5 实验结果与分析

本文提出的基于知识增强的文本分类算法的实验结果如表 4 所示, 其中 Model (KB) 表示应用知识增强算法后的模型. 由表 4 可以看出: 在 SST-5 数据集上, CNN (KB) 模型比 CNN 模型文本分类的准确率提高了 5.01 个百分点; LSTM (KB) 模型比 LSTM 模型文本分类的准确率提高了 7.92 个百分点; BERT (KB) 模型比 BERT 模型文本分类的准确率提高了 1.5 个百分点. 在 SST-2 数据集上, CNN (KB) 模型比 CNN 模型文本分类的准确率降低了 0.87 个百分点; LSTM (KB) 模型比 LSTM 模型文本分类的准确率提高了 1.76 个百分点; BERT (KB) 模型比 BERT 模型文本分类的准确率提高了 1.29 个百分点. 在 IMDB 数据集上, CNN (KB) 模型比 CNN 模型文本分类的准确率提高了 0.97 个百分点; LSTM (KB) 模型比 LSTM 模型文本分类的准确率提高了 2.87 个百分点; BERT (KB) 模型比 BERT 模型文本分类的准确率提高了 0.76 个百分点. 由上述结果可知, CNN、LSTM 和 BERT 模型结合外部知识后, 其文本分类准确率比原模型在大多数情况下都有所提高. 这表明, 应用知识增强算法可提升模型的分类准确率.

另外, 由表 4 还可以看出: 在 SST-5 和 SST-2 数据集上, CNN (KB)、LSTM (KB) 和 BERT (KB) 模型文本分类的准确率均高于 Emo2Vec、SWEM-concat 和 DAN 模型; 在 IMDB 数据集上, CNN (KB) 和 LSTM (KB) 模型文本分类的准确率均高于 Emo2Vec 模型, BERT (KB) 模型以 94.96% 的准确率表现最佳. 整体看来, CNN (KB)、LSTM (KB) 和 BERT (KB) 模型文本分类的效果优于 Emo2Vec、

SWEM-concat 和 DAN 模型 .

表 4 本文模型与其他对比模型的分类结果 %

模型	SST-5	SST-2	IMDB
Emo2Vec	41.60	81.20	83.50
SWEM-concat	46.10	84.30	—
DAN	47.70	—	89.90
CNN	43.18	87.21	86.67
LSTM	41.59	84.80	82.35
BERT	53.20	91.20	94.20
CNN(KB)	48.19	86.34	88.64
LSTM(KB)	49.51	86.56	86.22
BERT(KB)	54.70	92.49	94.96

2.6 知识增强算法在不同模型上的分类实验结果与分析

本文选用 RNN 模型对本文提出的知识增强算法的通用性进行验证,其文本分类实验结果如图 2 所示.由图 2 可以看出:在 SST-5 数据集上,RNN (KB) 模型的文本分类准确率比 RNN 模型提高 8.02 个百分点;在 SST-2 数据集上,RNN (KB) 模型的文本分类准确率比 RNN 模型提高了 2.45 个百分点;在 IMDB 数据集上,RNN (KB) 模型的文本分类准确率比 RNN 模型提高了 2.98 个百分点.上述结果表明,在加入外部知识后,文本分类模型的准确度得到了显著的提升,具有通用性.

2.7 知识增强算法在不同规模数据集上的分类实验结果与分析

本文选用 BERT 模型验证本文提出的知识增强算法在不同规模数据集 (SST-5 数据集) 中的分类效果,实验结果如图 3 所示.由图 3 可以看出:加入外部知识后,当分别使用 10%、30%、50%、80%、100% 的训练数据进行训练时,BERT (KB) 模型的文本分类准确率比 BERT 模型分别提高了 2.15、1.70、1.69、1.50、1.50 个百分点.该结果表明,在不同规模的训练数据中加入外部知识后,其均可起到提升文本分类性能的作用,且训练数据规模越小,提升效果越好.

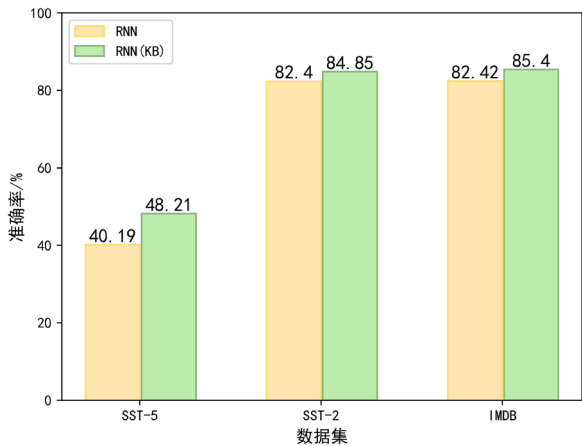


图 2 知识增强算法在不同模型上的分类结果

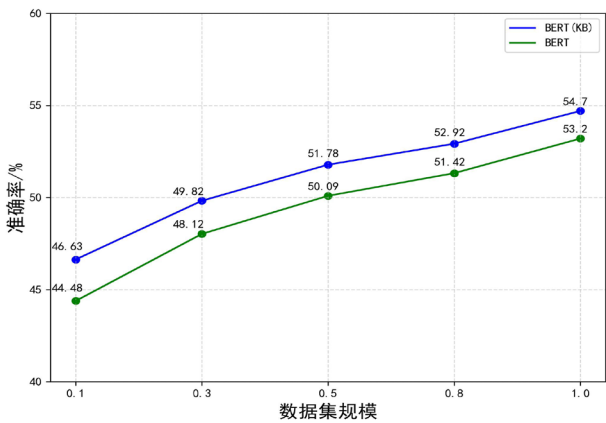


图 3 知识增强算法在不同规模 SST-5 数据集上的分类结果

3 结论

利用 SST-5、SST-2 和 IMDB 数据集, 对文本提出的基于知识增强的文本分类算法进行分析表明: 在 SST-5 数据集上, 本文提出的模型 CNN (KB)、LSTM (KB) 和 BERT (KB) 的文本分类准确率比基线模型分别提高了 5.01%、7.92% 和 1.5%; 在 SST-2 数据集上, 本文提出的模型 CNN (KB) 的文本分类准确率比基线模型虽然下降了 0.87%, 但模型 LSTM (KB) 和 BERT (KB) 的文本分类准确率比基线模型分别提高了 1.76% 和 1.29%; 在 IMDB 数据集上, 本文提出的模型 CNN (KB)、LSTM (KB) 和 BERT (KB) 的文本分类准确率比基线模型分别提高了 0.97%、2.87% 和 0.76%。上述表明, 本文算法能够提高文本分类任务的准确率。在今后的研究中, 我们将探究如何进一步优化知识增强的方法, 以进一步提高该算法的分类效果。

参考文献:

- [1] MIYATO T, DAI A M, GOODFELLOW I J. Adversarial Training Methods for Semi-Supervised Text Classification[C]//5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017: 1-11.
- [2] CHEN Y W, WANG J L, CAI Y Q, et al. A method for chinese text classification based on apparent semantics and latent aspects[J]. Journal of Ambient Intelligence and Humanized Computing, 2015, 6(4): 473-480.
- [3] 王海涛, 宋文, 王辉. 一种基于 LSTM 和 CNN 混合模型的文本分类方法 [J]. 小型微型计算机系统, 2020, 41(6): 1163-1168.
- [4] 陈可嘉, 刘惠. 基于改进 BiGRU-CNN 的中文文本分类方法 [J]. 计算机工程, 2022, 48(5): 59-66, 73.
- [5] 孙刘成, 黄润才. 融合 LSTM 和注意力机制的新闻文本分类模型 [J]. 传感器与微系统, 2022, 41(9): 38-41.
- [6] LIU W, ZHOU P, ZHAO Z, et al. K-BERT: Enabling language representation with knowledge graph[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(3): 2901-2908.
- [7] LAN Z, CHEN M, GOODMAN S, et al. Albert: A lite bert for self-supervised learning of language representations[C]//8th International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2020: 1-17.
- [8] SUN Y, WANG S, LI Y, et al. ERNIE 2.0: A continual pre-training framework for language understanding [C]// Proceedings of the AAAI Conference on Artificial Intelligence. New York: Association for the Advancement of Artificial Intelligence, 2020: 8968-8975.
- [9] GHOSAL D, HAZARIKA D, ROY A, et al. KinGDOM: Knowledge-guided domain adaptation for sentiment analysis[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, 2020: 3198-3210.
- [10] SPEER R, CHIN J, HAVASI C. Conceptnet 5.5: An open multilingual graph of general knowledge[C] // Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. San Francisco: AAAI Press, 2017: 4444-4451.
- [11] 卢嘉荣, 肖红, 姜文超, 等. 基于语料关联生成的知识增强型 BERT[J]. 湖北大学学报 (自然科学版), 2022, 44(6): 732-741.
- [12] HONNIBAL M, MONTANI I. SpaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing[J]. To appear, 2017, 7(1): 411-420.
- [13] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [14] ELMAN J L. Distributed representations, simple recurrent networks, and gram-matical structure[J]. Machine learning,

- 1991, 7(2/3): 195-225.
- [15] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019: 4171-4186.
- [16] ZHANG X, FANG A. MPCNN with Knowledge Augmentation: A model for Chinese text classification[C]// International Conference on Intelligent Computing. Cham: Springer International Publishing, 2022: 141-149.
- [17] 杨璐, 何明祥. 基于门控机制和卷积神经网络的中文文本情感分析模型 [J]. 计算机应用, 2021, 41(10): 2842-2848.
- [18] 姜克鑫, 赵亚慧, 崔荣一. 融合高低层语义信息的自然语言句子匹配方法 [J]. 计算机应用研究, 2022, 39(4): 1060-1063.
- [19] XU P, MADOTTO A, WU C S, et al. Emo2Vec: Learning generalized emotion representation by multi-task training[C]//Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Brussels: Association for Computational Linguistics, 2018: 292-298.
- [20] IYYER M, MANJUNATHA V, BOYD-GRABER J, et al. Deep unordered composition rivals syntactic methods for text classification[C]//Proceedings of the 53rd Annual meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing: Association for Computational Linguistics, 2015: 1681-1691.
- [21] SHEN D, WANG G, WANG W, et al. Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne: Association for Computational Linguistics, 2018: 440-450.