

基于改进卷积神经网络的 短文本分类模型

高云龙^{1,2}, 吴 川¹, 朱 明¹

(1. 中国科学院 长春光学精密机械与物理研究所, 长春 130033;
2. 中国科学院 航空光学成像与测量重点实验室, 长春 130033)

摘要: 基于卷积神经网络, 提出一种基于改进卷积神经网络的短文本分类模型. 首先, 采用不同编码方式将短文本映射到不同空间下的分布式表示, 提取不同粒度的数字特征作为短文本分类模型的多通道输入, 并根据标准知识库提取概念特征作为先验知识, 提高短文本的语义表征能力; 其次, 在全连接层增加自编码学习策略, 在近似恒等的基础上进一步组合数字特征, 模拟数据内部的关联性; 最后, 利用相对熵原理为模型增加稀疏性限制, 降低模型复杂度的同时提高模型的泛化能力. 通过对开源数据集进行短文本分类实验, 验证了模型的有效性.

关键词: 卷积神经网络; 短文本; 概念分布式表示; 稀疏; 自编码

中图分类号: TP181 **文献标志码:** A **文章编号:** 1671-5489(2020)04-0923-08

Short Text Classification Model Based on Improved Convolutional Neural Network

GAO Yunlong^{1,2}, WU Chuan¹, ZHU Ming¹

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Science, Changchun 130033, China;
2. Key Laboratory of Airborne Optical Imaging and Measurement, Chinese Academy of Sciences, Changchun 130033, China)

Abstract: We proposed a short text classification model based on improved convolutional neural network. Firstly, different coding methods were used to map short text to distributed representation in different spaces, and digital features of different granularities were extracted as multi-channel inputs of short text classification model. Extracting concept features from standard knowledge base as prior knowledge to improve the semantic representation ability of short text. Secondly, the self-coding learning strategy was added to the full connection layer, on the basis of approximate identity, the digital features were further combined to simulate the relevance within the data. Finally, the principle of relative entropy were used to increase the sparsity limit of the model, reduce the complexity and improve the generalization ability of the model. The effectiveness of the proposed model was verified by short text classification experiments on the open source dataset.

Keywords: convolutional neural network; short text; concept distributed representation; sparsity; self-coding

收稿日期: 2019-11-13.

第一作者简介: 高云龙(1993—), 男, 汉族, 硕士, 从事机器学习、数据挖掘和图像处理的研究, E-mail: 1458299660@qq.com.

通信作者简介: 吴 川(1974—), 男, 汉族, 博士, 研究员, 从事图像处理和机器学习的研究, E-mail: wuchuan0458@163.com.

基金项目: 国家自然科学基金(批准号: 61401425)和吉林省科技发展规划项目(批准号: 20200571505JH).

文本挖掘是自然语言处理领域中一项较重要的任务,涉及机器翻译、智能问答等许多领域^[1].短文本数据反映了个体对生活方式的总结及社会现象的评论等,常蕴含丰富的普遍规律.不同于传统长文本,短文本有显著的特点^[2]:1) 文本长度短;2) 特征较稀疏;3) 文本格式不统一.短文本的特点使其通常不能提供足够的上下文信息,从而使许多基于词频、字同现的机器学习模型不能有效地提取短文本特征,进而限制了传统的机器学习算法在短文本分类问题上的应用^[3].本文将现有的短文本分类算法分为两类:基于传统语义分析的短文本分类算法和基于深度神经网络的短文本分类算法.前者常通过外部语义信息库丰富短文本的特征表示.在 WordNet^[4]同义词数据库中,每个同义词集合都表示一个基本的语义概念.Chen 等^[5]提出了多粒度主题模型用于精确描述短文本,首先在通用数据集中采用 LDA(latent dirichlet allocation)算法^[6]提取 N 组主题集 T ,再从主题集 T 中选择一个子集构成多粒度的主题空间,将短文本的词特征与主题特征融合成新的短文本特征表示,最后利用支持向量机(SVM)分类器获取较好的分类效果;Sun 等^[7]提出了一种基于 TNG(topic N-gram)的特征扩展方法,首先该方法可同时推断出词的分布式表示及每个主题下短语的分布式表示,并由此构建特征扩展库,然后基于所提取的原始特征,对短文本的主题趋势进行推理,并根据主题趋势在特征扩展库中选择相应的候选词汇与短语,最后将选择结果作为短文本数据特征的扩充,并通过 LDA 与 SVM 算法得到了较好的分类效果;Heap 等^[8]通过使用通用及专用的词向量模型丰富由 Bag-of-words 模型^[9]对低频词汇所生成的词向量,从而提高了多类别短文本的分类精度.而基于深度神经网络的分类算法更关注于解决短文本的分布式表示及特征的无监督获取能力,在短文本分类问题中取得了较好的效果.随着深度学习模型在图像挖掘^[10]、语音识别^[11]等领域的广泛应用,如何利用深度模型处理文本挖掘问题受到广泛关注.Kim^[12]提出了一个基于卷积神经网络(CNN)处理句子分类问题的模型,该模型将句子视为词的序列,并利用预先训练好的词的分布式表示作为模型输入,此外,还按输入词向量的不同将模型分为 4 种变型,其中双道词向量输入模型普遍取得了较好的分类效果;Kalchbrenner 等^[13]提出了一种基于动态的卷积神经网络模型 DCNN,并应用到句子分类问题中,在卷积层提出了一种广义卷积,可充分保留句子边缘词汇的特征信息,在池化层根据卷积层节点的数量定义一种全局的动态 K -Max 策略,从而动态地确定池化层保留的特征数量,DCNN 模型可处理变长的语料数据,具有较强的泛化能力;Zhang 等^[14]构建了短文本的字符分布式表示模型,并通过实验结果验证了模型的有效性.基于神经网络的短文本分类模型通常不需要语言的句法、语法等先验信息,因此在不同类型的数据集或不同的语言中都表现出了良好的扩展性和有效性.

基于此,本文提出一种基于改进卷积神经网络的短文本分类模型 ICNNSTCM.该模型构造了短文本语料多种不同的分布式表示:1) 字符级特征表示和词汇级表示,并作为分类模型 ICNNSTCM 的双道输入,词汇级分布式表示同时作为模型的参数进行调优;2) 提取短文本中包含的实体,在标准知识库中检索实体对应的概念集,并提出 CToST 学习策略计算概念与短文本之间的关联强度,从而构成短文本的概念特征表示.在全连接层增加稀疏自编码策略,即通过非监督学习一个近似恒等的转换函数,利用稀疏性限制保证网络层多数节点处于被抑制的状态,在降低模型复杂度的同时提高模型的泛化能力.本文模型和相关实验均基于 Python 语言及 TensorFlow 框架实现,通过与其他几种短文本分类模型进行对比,实验结果表明,ICNNSTCM 可更好地提取短文本数据的特征,具有较高的分类精度.

1 模型构建

图 1 为基于改进卷积神经网络的短文本分类模型 ICNNSTCM 的框架.采用不同的学习策略分别构建短文本的分布式表示,并作为模型的多通道输入.构造字符级分布式表示即将短文本视为字符的序列,从而捕获短文本的粒子字信息;词汇级分布式表示,即词向量,其几何关系常作为词汇间语义关系的高级抽象;概念级分布式表示,利用标准知识库提取短文本的概念特征,并以此作为模型先验知识,丰富语料表征能力的同时提高了模型参数调优的效率;通过不同窗口的卷积核与 3 种分布式表示分别进行卷积运算,提取卷积层节点的数字特征;池化层采用 K -Max Pooling 保留卷积层节点最大的 K 个特征,并合并为一个一维的长特征序列;在两个全连接层之间增加稀疏自编码学习策略,高层

特征通过与分类器相连,从而预测输入短文本的类别。

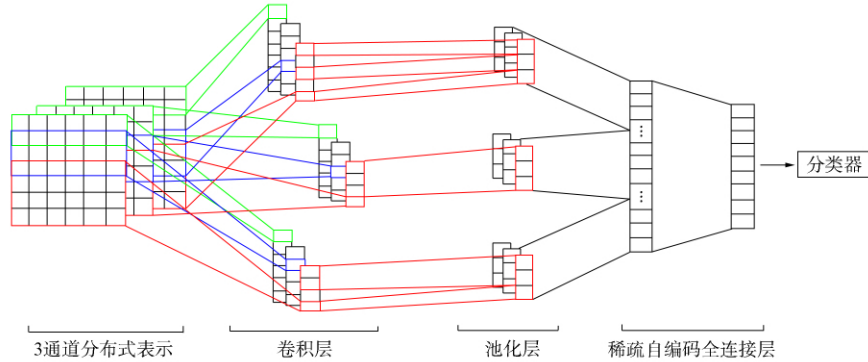


图1 ICNNSTCM 模型结构

Fig.1 Structure of ICNNSTCM model

1.1 字符级分布式表示

通过将输入短文本视为字符的序列,并利用该语言下长度为 m 的字母表 Ω 对字符进行编码量化,从而实现短文本数据字符级分布式表示的构建.考虑到字符数量远小于该语言下词汇的总量,因此采用字符级分布式表示在一定程度上可降低模型的计算复杂度.基于 Bag-of-words 编码方式的思想,将每个字符视为一个词汇,从而得到每个字符的编码;对不包含在字母表 Ω 中的字符,简单地将其编码为零向量.对于长度为 l 个字符的短文本,通过以上处理可得到其维度为 $l \times m$ 的字符分布式表示.本文使用的字母表 Ω 包含 68 个字符:26 个英文小写字母,10 个数字字符及 32 个其他字符。

1.2 概念级分布式表示

概念特征分布式表示将短文本视为互相关联的实体构成的集合,对于集合中的每个实体,通过检索外部标准知识库,可获取其对应的概念,从而更好地实现分类任务.例如短文本 $ST1$:“Rihanna is a super star.”,其中名词实体“Rihanna”对应的概念为“Singer”,因此,分类器更倾向于将 $ST1$ 归类为“娱乐”类别中.此外,实体通常对应多个不同的概念,从而具有较大的歧义性.例如短文本 $ST2$:“She is eating an apple”,对于名词实体“apple”,可得到概念“Phone”和“Apple”,显然对于 $ST2$,“Phone”不是一个正确的概念。

为更好地描述概念与短文本之间的关联强度,本文提出了 CToST 学习策略.对元素个数为 m 的概念集 $C = \{c_1, c_2, \dots, c_m\}$,其对应的词向量可表示为 $CV \in \mathbb{R}^{m \times d}$,短文本 ST 对应的词向量表示为 $ST \in \mathbb{R}^{st \times d}$,其中: d 为词向量维度; st 为短文本 ST 中词的个数.则概念 c_i 与短文本 ST 的关联强度值可表示为

$$\text{corr}_i = g(W(CV_i \cdot ST)^T + b), \quad (1)$$

其中: g 为非线性函数;参数 $W \in \mathbb{R}^d$, $b \in \mathbb{R}^d$.

进一步,为规范化关联强度值,本文采用 Softmax 算子作用于式(1),最终的关键强度值可表示为

$$f_corr_i = \text{Softmax}(\text{corr}_i) = \frac{\exp\{\text{corr}_i\}}{\sum_{k=1}^m \exp\{\text{corr}_k\}}. \quad (2)$$

因此,短文本 ST 的概念级分布式表示可描述为

$$\begin{bmatrix} f_corr_1 \times CV_1 \\ f_corr_2 \times CV_2 \\ \vdots \\ f_corr_m \times CV_m \end{bmatrix}.$$

1.3 稀疏自编码策略

ICNNSTCM 模型中两个全连接层之间的自编码特性是通过增加一个隐含的第三层作为输出层实现的,如图 2 所示.在图 2 所示的三层网络中,尝试学习一个 $h_{w,b}(X) \approx X$ 的恒等函数,从而使得 L_3 层的输出 $\hat{X}$ 近似等于输入 X .

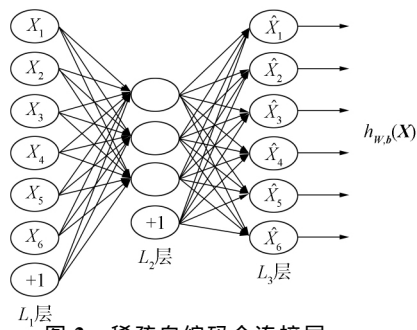


图 2 稀疏自编码全连接层

Fig.2 Sparse and self-coding full connection layer

令 $a^{(1)} = X$ 作为输入层的激活值, 则在给定第 n 层的激活值后, 第 $(n+1)$ 层的激活值可按下式获得:

$$z^{(n+1)} = W^{(n)} a^{(n)} + b^{(n)}, \quad (3)$$

$$a^{(n+1)} = f(z^{(n+1)}), \quad (4)$$

其中 f 为非线性激活函数.

为了使网络学习到更有趣的数据特征, 本文在自编码网络上增加了稀疏性限制. 令 $a_j^{(2)}(X)$ 表示在给定输入 X 的情况下, L_2 层网络节点 j 的激活度, 则节点 j 的平均活跃度可表示为

$$\hat{\rho}_j = \frac{1}{m} \sum_{i=1}^m [a_j^{(2)}(X)], \quad (5)$$

其中 m 表示数据的个数. 进一步, 令 $\hat{\rho}_j = \rho$ (ρ 通常是一个接近于 0 的稀疏性参数值), 以保证网络中大部分节点的活跃度接近于 0. 为了实现这种限制, 本文在优化目标函数中增加了一个基于相对熵的正则化项, 两个变量 ρ 和 $\hat{\rho}_j$ 之间的相对熵可表示为

$$\text{KL}(\rho \parallel \hat{\rho}_j) = \rho \lg \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \lg \frac{1 - \rho}{1 - \hat{\rho}_j}. \quad (6)$$

令 $J(W, b)$ 表示全连接网络的目标函数, 则增加稀疏性约束后网络的优化函数为

$$J_{\text{sparse}}(W, b) = J(W, b) + \beta \sum_{j=1}^{s_2} \text{KL}(\rho \parallel \hat{\rho}_j), \quad (7)$$

其中 s_2 表示 L_2 层中节点的个数.

1.4 算法描述

算法 1 稀疏自编码算法.

输入: 池化层输出特征向量 X ;

输出: L_2 自编码层参数, 包括权重矩阵 W_1 、偏置 b_1 、 L_2 层输出特征向量 $a^{(2)}$;

初始化: 随机设置 L_2, L_3 层权重矩阵 W_1, W_2 和偏置 b_1, b_2 ;

步骤 1) 设迭代次数 $\text{iter} = 0$;

步骤 2) 根据式(3)和式(4)计算 L_2, L_3 层激活值 $a^{(2)}, a^{(3)}$;

步骤 3) $h_{w,b}(X) = a^{(3)}$;

步骤 4) $J(W, b) = \|h_{w,b}(X) - X\|^2$;

步骤 5) 根据式(5)计算 $\hat{\rho}$;

步骤 6) 根据式(6)计算相对熵;

步骤 7) 根据式(7)计算 $J_{\text{sparse}}(W, b)$;

步骤 8) 误差反向传播优化参数 fine-tuning;

步骤 9) $\text{iter}++$;

步骤 10) 如果 $\text{iter} < 1\,000$, 则返回步骤 2).

1.5 模型实现

本文的模型、相关实验均基于 Python 语言及 TensorFlow 框架实现, 并可以同时支持 CPU 和 GPU. 模型参数的选取: 短文本包含字符的长度 $l = 512$, 对于长度超出 512 个字符的短文本将被丢弃, 字符级分布式表示采用 One-Hot 编码; 词向量采用 Word2vec 工具, 并基于 Python 库 Gensim API 实现, 词向量维度值 $d = 200$; 非线性函数 g 选取 Softmax; 卷积核宽度 width 为 3, 4, 5, 并分别对应 100 个卷积节点; 卷积层及全连接层非线性激活函数 f 选取 Sigmoid; 稀疏性参数 $\rho = 0.1$; 第二个全连接层中节点的个数 $s_2 = 1\,024$; batch_size = 50, 迭代次数 epochs = 25; 分类器选择 Softmax 模型, 模型训练采用 Adam 优化算法.

2 实验分析

2.1 数据集

实验采用的标准短文本语料库均可通过下列开源网站获得:

1) MRD(<http://www.cs.cornell.edu/people/pabo/movie-review-data>) 影片评论数据集, 其中每个句子是用户对影片的一条评论, 有正、反面两种属性标签;

2) SST(<http://nlp.stanford.edu/sentiment/>) 情感分析数据集, 是 MRD 数据集的扩展, 共有5种属性标签, 分别为 very positive, positive, neutral, negative, very negative;

3) TREC(<http://cogcomp.cs.illinois.edu/Data/QA/QC/>) 问题属性分析数据集, 问题类别共有6种 (about person, location, numeric information, etc), 将问题与类别进行匹配;

4) CR(<http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>) 顾客意见数据集, 是关于商品正、反面的二元分类数据集。

短文本数据集的相关信息及参数列于表1, 其中: C 为目标类别数; L 为平均句子长度; N 为数据集规模; $|V|$ 为词汇量规模; Test 为测试机规模; CV 为十折交叉验证拆分数据集。

表1 数据集的参数

Table 1 Parameters of dataset

数据集	C	L	N	$ V $	Test
MRD	2	20	10 662	18 765	CV
SST	5	18	11 855	17 836	2 210
TREC	6	10	5 952	9 592	500
CR	2	19	3 775	5 340	CV

2.2 卷积核宽度对模型的影响

卷积层是卷积神经网络中最重要的结构, 通过卷积核与输入数据的局部感受野进行卷积运算, 提取数字特征. 卷积核的维度对所提取特征的粒度有影响, 进而影响模型的分类精度。

组合不同窗口大小的卷积核, 并在短文本数据集上进行分类实验, 结果如图3所示, 其中: [3]表示卷积核的宽度为3, 即一次卷积过程可覆盖3个词汇或者字符; [3,4]表示卷积层含有宽度分别为3,4的卷积核组合; [3,4,5]表示卷积层含有宽度分别为3,4,5的卷积核组合. 由图3可见, 随着窗口组合数的增加, 卷积层可提取不同粒度的特征, 从而学习到更丰富的抽象特征表示, 进而提高了分类精度. 当固定窗口组合数时, 通过改变卷积核的宽度进行实验, 结果列于表2. 由表2可见, 在卷积核宽度组合为[3,4,5]时, 模型取得了分类精度的极值。

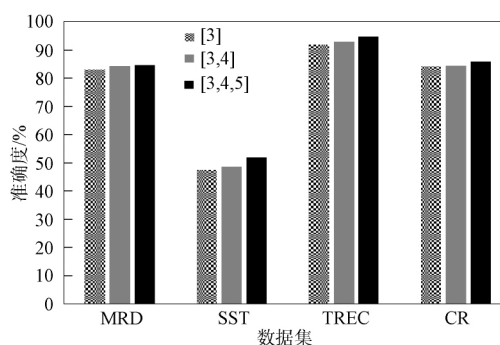


图3 不同宽度组合对分类精度的影响

Fig.3 Influence of different width combinations on classification accuracy

表2 卷积核宽度对分类精度的影响

Table 2 Influence of widths of convolution kernel on classification accuracy

数据集	[1,2,3]	[2,3,4]	[3,4,5]	[4,5,6]	[5,6,7]
MRD	83.5	84.4	84.6	84.2	83.0
SST	47.5	49.3	51.7	48.5	47.2
TREC	92.1	93.1	94.7	93.4	92.8
CR	84.2	84.6	85.8	85.2	84.8

2.3 池化层 K 值选取对模型的影响

ICNNSTCM 模型池化层采用 K -Max 策略选择卷积层较大的 K 个特征值, 并保留其相对顺序. K 值选取对模型泛化误差的影响如图 4 所示. 由图 4 可见, K 值选取通常与短文本数据的平均长度有关, 对平均词汇长度为 18, 19, 20 的数据集 SST, CR, MRD, $K=15$ 时模型取得的泛化误差值最低, 而对平均词汇长度为 10 的数据集 TREC, $K=8$ 时模型分类精度最高. 当 K 值较低时, 模型通常不能保留足够多的特征信息, 而当 K 值较大时, 模型通常表现为过拟合状态, 从而均导致模型分类精度较低.

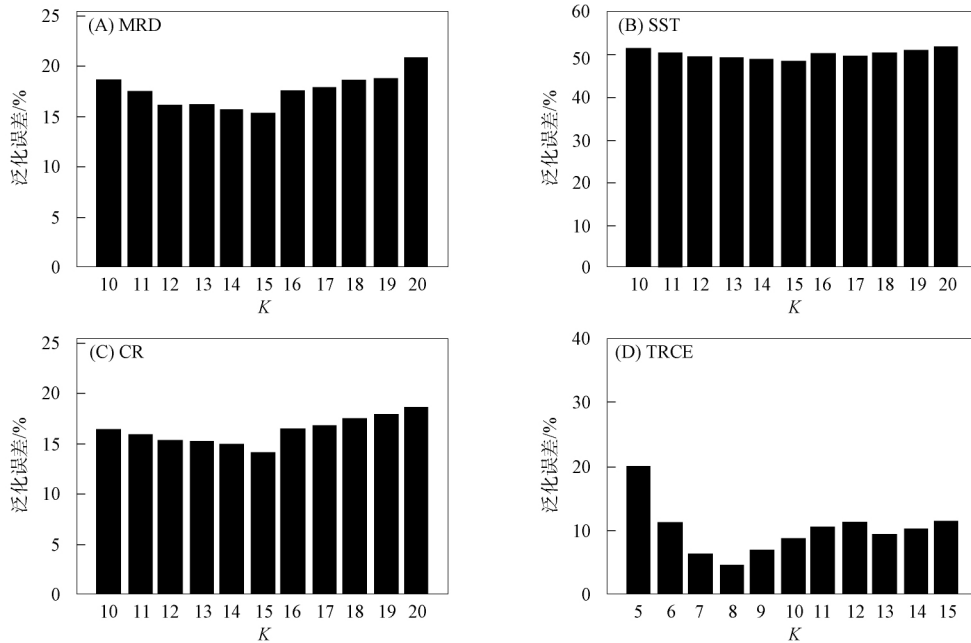


图 4 K 值对模型泛化误差的影响

Fig.4 Influence of K values on generalization error of model

2.4 概念特征表示对模型的影响

本文在短文本字符级和词汇级分布式表示的基础上提出了概念级分布式表示, 利用外部标准知识库为模型增加先验知识, 进一步丰富了短文本的特征描述. ICNNSTCM 模型在不同通道输入情况下的分类精度结果列于表 3. ICNNSTCM-2 模型采用 2 通道输入, 即短文本的字符级和词汇级分布式表示, 而 ICNNSTCM-3 模型则增加了短文本的概念特征分布式表示. 由表 3 可见, 利用先验知识为模型增加概念特征表示, 在一定程度上可提高分类的精确度; 此外, 采用不同的外部标准知识库对模型分类精度也存在一定的影响, 知识库 YAGO^[15] 整体上优于 FreeBase^[16].

表 3 概念分布式表示对模型分类精度的影响

Table 3 Influence of concept distributed representation on classification accuracy of model

数据集	YAGO		FreeBase	
	ICNNSTCM-2	ICNNSTCM-3	ICNNSTCM-2	ICNNSTCM-3
MRD	83.9	84.6	83.1	84.2
SST	48.5	49.8	50.7	51.7
TREC	93.0	94.7	92.8	93.2
CR	84.6	85.8	84.0	84.6

2.5 稀疏性对模型的影响

好的特征具有可区分性^[17]. 为进一步减少特征表示的冗余信息, 通过在 ICNNSTCM 模型自编码层增加稀疏性约束, 以获取输入数据的低维空间表示. 本文对稀疏性概念解释为: 对于激活函数 Sigmoid, 神经元输出接近 1 时表示该节点被激活, 输出接近 0 时表示该节点被抑制, 且同一时刻自编码层大部分节点处于被抑制状态. 稀疏性对模型的影响如图 5 所示. 由图 5 可见, 稀疏特征表示的有效

性通常与短文本数据的平均长度有关,对平均词汇长度为 18,19,20 的数据集 SST,CR,MRD,增加稀疏自编码层可有效降低分类模型的泛化误差,而对平均词汇长度为 10 的数据集 TREC,由于语料的冗余特征较少,增加稀疏约束导致模型无法提取足够多的特征信息,从而降低了模型的分类精度。

2.6 ICNNSTCM 模型的两种变型

ICNNSTCM 模型采用字符级、词汇级和概念级分布式表示作为多通道输入,针对词汇级分布式表示,即词向量是否作为模型优化的参数,本文将模型分为如下两种变型。

- 1) ICNNSTCM-static: 词向量在模型训练过程中保持不变,不作为模型调参的变量;
- 2) ICNNSTCM-dynamic: 词向量在模型训练过程中进行调优,作为模型调参的变量。

在 MRD 数据集上,本文统计了两种变型在迭代调优中泛化误差下降及分类精度上升的收敛过程,分别如图 6 和图 7 所示。由图 6 和图 7 可见:由于 ICNNSTCM-static 在训练过程中优化的参数较少,故其收敛速度高于 ICNNSTCM-dynamic 模型。此外,通过使用与分类问题相关的数据集对词向量进行调优,可提高词向量对数据集的表征能力,从而提高了模型的分类精度。

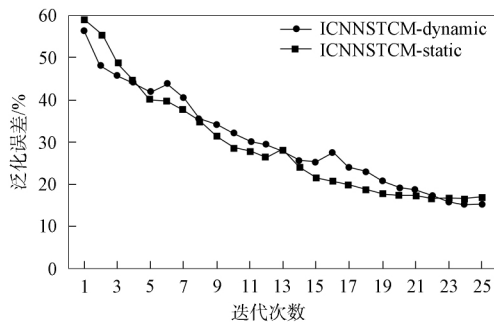


图 6 两种模型泛化误差下降速度对比

Fig.6 Comparison of descent speed of generalization error between two models

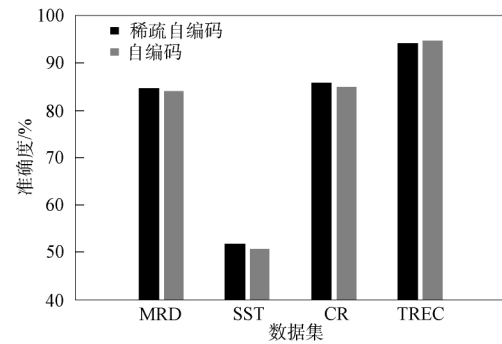


图 5 稀疏性对模型分类精度的影响

Fig.5 Influence of sparsity on classification accuracy of model

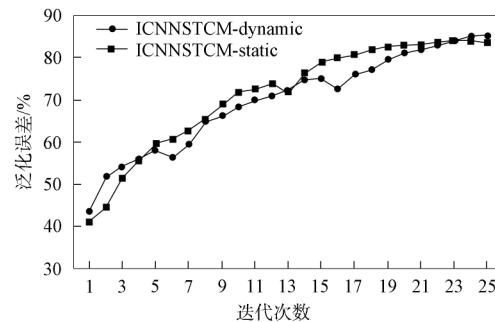


图 7 两种模型分类精度上升速度对比

Fig.7 Comparison of rising speed of classification accuracy between two models

2.7 对比实验

下面通过与几种成熟的短文本分类模型进行对比,验证 ICNNSTCM 模型的有效性,实验结果列于表 4。ICNNSTCM 采用多通道分布式输入,充分提取不同粒度的数字特征,并通过使用与分类问题相关的数据集对词向量进行调优,提高了对于输入短文本的表示能力。此外,在全连接层增加的稀疏自编码策略,在利用近似恒等关系保持特征有效性不变的情况下,通过组合所提取的特征,进一步学习输入数据内部的相关性。实验结果表明,ICNNSTCM 模型在数据集上均取得了较理想的分类效果,实验结果优于目前已有的大部分分类模型。

表 4 ICNNSTCM 模型与其他模型性能对比

Table 4 Performance comparison between ICNNSTCM model and other models

模型	MRD	SST	TREC	CR
ICNNSTCM-static	83.5	48.1	91.6	84.2
ICNNSTCM-dynamic	84.6	51.7	94.7	85.8
CNN-non-static ^[12]	81.5	48.0	93.6	84.3
CNN-multichannel ^[12]	81.1	47.4	92.2	85.0
DCNN ^[13]	—	48.5	93.0	—
SVM ^[18]	—	—	95.0	—
Paragraph-Vec ^[19]	—	48.7	—	—

综上所述,本文提出了一种 ICNNSTCM 模型,并通过相关实验证明了 ICNNSTCM 模型在处理短文本分类问题上的有效性。

参 考 文 献

- [1] FRANCK D. Sequential Short-Text Classification with Neural Networks [D]. Cambridge: Massachusetts Institute of Technology, 2017.
- [2] SONG G, YE Y M, DU X L, et al. Short Text Classification: A Survey [J]. Journal of Multimedia, 2014, 9(5): 635-643.
- [3] RAFEEQUE P C, SENDHILKUMAR S. A Survey on Short Text Nalysis in Web [C]//2011 Third International Conference on Advanced Computing. Piscataway, NJ: IEEE, 2011: 365-371.
- [4] MILLER G A, FELLBAUM C. WordNet Then and Now [J]. Language Resources and Evaluation, 2007, 41(2): 209-214.
- [5] CHEN M G, JIN X M, SHEN D. Short Text Classification Improved by Learning Multi-granularity Topics [C]//Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. New York: ACM, 2011: 1776-1781.
- [6] WANG Z B, MA L, ZHANG Y Q. A Hybrid Document Feature Extraction Method Using Latent Dirichlet Allocation and Word2Vec [C]//2016 IEEE First International Conference on Data Science in Cyberspace. Piscataway, NJ: IEEE, 2016: 98-103.
- [7] SUN B S, ZHAO P. Feature Extension for Chinese Short Text Classification Based on Topical N-Grams [C]//2017 IEEE/ACIS 16th International Conference on Computer and Information Science. Piscataway, NJ: IEEE, 2017: 477-482.
- [8] HEAP B, BAIN M, WOBCKE W, et al. Word Vector Enrichment of Low Frequency Words in the Bag-of-Words Model for Short Text Multi-class Classification Problems [J/OL]. CoRR, 2017-09-18[2019-10-11]. <https://arxiv.org/pdf/1709.05778v1.pdf>.
- [9] NIKOLAOS P, ANASTASIOS T. Bag of Embedded Words Learning for Text Retrieval [C]//2016 23rd International Conference on Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2416-2421.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Image Net Classification with Deep Convolutional Neural Networks [C]//Proceedings of the 25th International Conference on Advances in Neural Information Processing Systems. New York: ACM, 2012: 1097-1105.
- [11] GRAVES A, MOHAMED A R, HINTON G. Speech Recognition with Deep Recurrent Neural Networks [C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE, 2013: 6645-6649.
- [12] KIM Y. Convolutional Neural Networks for Sentence Classification [C]//Proceedings of Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2014: 1746-1751.
- [13] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P. A Convolutional Neural Network for Modelling Sentences [C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2014: 655-665.
- [14] ZHANG X, ZHAO J B J, LECUN Y. Character-Level Convolutional Networks for Text Classification [C/OL]. 2015-09-10[2019-10-11]. <https://arxiv.org/pdf/1509.01626v2.pdf>.
- [15] SUCHANEK F M, KASNECI G, WEIKUM G. Yago a Large Ontology From wikipedia and Wordnet [J]. Web Semantics Science Services and Agents on the World Wide Web, 2008, 6(3): 203-217.
- [16] PASSALIS N, TEFAS A. Bag of Embedded Words Learning for Text Retrieval [C]//2016 23rd International Conference on Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2416-2421.
- [17] LECUN Y, BENGIO Y, HINTON G. Deep Learning [J]. Nature, 2015, 512: 436-444.
- [18] SILVA J, COHEUR L, MENDES A C, et al. From Symbolic to Sub-symbolic Information in Question Classification [J]. Artificial Intelligence Review, 2011, 35(2): 137-154.
- [19] LE Q, MIKOLOV T. Distributed Representations of Sentences and Documents [C]//Proceedings of the 31st International Conference on Machine Learning. New York: ACM, 2014: 1188-1196.