

基于改进激活函数和学习率的双向 LSTM 研究^{*}

李 军¹, 李 明¹, 曾 燕², 李 莉³

(1. 重庆师范大学 计算机与信息科学学院; 2. 重庆师范大学 新闻与传媒学院, 重庆 401331;

3. 西南大学 计算机与信息科学学院, 重庆 400715)

摘要:【目的】目前双向长短期记忆网络(Long short term memory, LSTM)在语音识别、图像识别和情感分类等方面的应用越来越广泛,基于此研究如何提高双向 LSTM 的准确率。【方法】提出一种改进的双向 LSTM,通过对 LSTM 中输入门与输出门激活函数的改进,并结合改进的学习率,能够大大地提高神经网络的收敛速度与准确率。【结果】在不同领域的数据集 MNIST 和 IMDB 上分别进行实验,改进的双向 LSTM 在 MNIST 数据集上的准确率达到 99.61%,比基准双向 LSTM 的准确率提高了 1.172%,在 IMDB 数据集上的准确率则达到了 88.08%,比基准双向 LSTM 的准确率提高了 1.6%,而且在两个数据集上在迭代次数较小的情况下,相比基准双向 LSTM,改进的算法都达到了较高的准确率和较低的损失率。【结论】由此证明改进的双向 LSTM 优于基准双向 LSTM。

关键词:双向 LSTM;激活函数;学习率;输入门;输出门

中图分类号:TP391.4

文献标志码:A

文章编号:1672-6693(2021)02-0070-07

随着神经网络算法的飞速发展,尤其是深度学习算法的快速发展,语音识别、机器翻译和图像识别等应用技术正逐步走进人们的日常生活。双向长短期记忆网络(Long shortterm memory, LSTM)作为一种时间循环神经网络在这些技术领域中的应用十分广泛。随着应用场景的不断扩大,这些技术领域对算法准确度的要求也会越来越高,因此提高双向 LSTM 的准确率就显得尤为迫切。

双向 LSTM 是一种特殊的循环网络架构,可以很好地解决梯度消失或者梯度爆炸的问题,也可以使输出层能够同时获得过去和未来的信息,从而进一步增加信息的依赖与维持。典型的多层神经网络会有许多不同的参数,然而在双向 LSTM 中使用的参数都是一样的,所以会大大减少网络在训练过程中所需要的参数量,也因此缩短了训练时间。传统的双向 LSTM 中,输入门与输出门的激活函数都是 tanh,但是 tanh 激活函数的一大缺点是收敛速度慢,影响训练速率,而且目前对神经网络中学习率的设置大多都是根据先验经验设置或者通过穷举来挑选出最优学习率,这样不仅会消耗大量的时间,也可能达不到合适的学习率,从而影响训练时间和准确率。

本文提出了改进的双向 LSTM,不仅对输入门与输出门的激活函数作了改进,并且对学习率也进行了改进,从而解决了传统双向 LSTM 收敛速度慢,准确率低的问题。实验表明,改进的双向 LSTM 在 MNIST 数据集和 IMDB 数据集上的准确率都要优于基准双向 LSTM。

文献[1-3]中提出将双向 LSTM 循环神经网络用于图像识别与数据预测等领域;文献[4]中提出利用 K-近邻算法对手写数字识别进行研究,准确率达到 89.81%;文献[5]中利用径向基网络对手写数字识别进行研究,准确率达到 98.70%;文献[6]中通过图像增强技术与卷积神经网络的结合,在 MNIST 数据集上的识别准确率达到 99.56%;文献[7]中使用“DropConnect”方法,在 MNIST 数据集上的准确率达到 99.79%,这也是目前在 MNIST 数据集上测试的最高准确率。文献[8-10]中通过改进学习率、融合特征等方式减少了参数数量和消耗时间;文献[11-12]中提出可以通过最大化利用数据和对图像的二值化处理来优化网络。王建仁等人^[13]提出改进的 LeNet-DL 网络,通过优化学习率,在 MNIST 数据集上的识别准确率达到 99.34%;曾文献等人^[14]提

^{*} 收稿日期:2020-09-08 修回日期:2020-12-29 网络出版时间:2021-04-07 15:39

资助项目:国家自然科学基金(No. 61877051);重庆市研究生教育改革重点项目(No. yjg182022);重庆师范大学研究生项目(No. xyjg16009);重庆师范大学教改项目(No. 02020310-0420)

第一作者简介:李军,男,研究方向为深度学习、计算机视觉, E-mail:1429973173@qq.com;通信作者:李明,男,教授, E-mail:55613163@qq.com

网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.N.20210407.1324.002.html

出了深度卷积自编码器神经网络,通过卷积神经网络对特征的提取,利用自编码器对特征信息进行保留,在 MNIST 数据集上的识别准确率达到 99.37%。闵锋等人^[15]提出基于改进主成分分析网络的手写数字识别方法,通过主成分分析网络和压缩感知的结合,得到低维的子空间,保留了高维图像的特征,从而降低了内存消耗,但是准确率不高;侯艳路等人^[16]提出了基于卷积神经网络的 C-RF 模型应用于手写数字识别,通过卷积神经网络提取特征,再输入到 RF 进行分类,提高了泛化能力,减少了训练时间,但是准确率的提高不大。

1 理论准备

1.1 双向 LSTM 和激活函数

图 1 展示了传统 LSTM 元胞的信息处理过程。其中: X_t 表示输入信息; A_{t-1} 表示上文信息; C_{t-1} 表示上一时刻的单元状态; C_t 表示更新后的元胞状态信息; O_t 表示输出的信息; S 代表 sigmoid 激活函数,它输出的值介于 $[0,1]$ 区间,当值为 0 时表示全部抑制,值为 1 则表示全部激活;图 1 中“+”代表求和,“ $\times$ ”代表乘积; $\tanh$ 函数用于生成候选记忆,值域为 $[-1,1]$ 。于是图 1 展示的整个信息处理过程可以描述为:输入信息 X_t 在和上文信息 A_{t-1} 堆叠后被复制成 4 份:第 1 份与遗忘权重 W_1 相乘再加上偏置项 b_1 后,经过激活函数 S 得到遗忘门 f_t ,从而决定从元胞状态中丢弃什么信息;第 2 份与输入权重 W_2 相乘再加上偏置项 b_2 后,经过激活函数 S 得到输入门 I_t ,这样就确定了元胞外部的输入信息;第 3 份与候选记忆权重 W_3 相乘再加上偏置项 b_3 后,经过激活函数 $\tanh$ 得到当前输入的单元状态 U_t ,也即决定了记忆哪些信息,这时将输入信息 I_t 与当前输入的单元状态 U_t 相乘,再加上上一单元经过遗忘后的状态“ $C_{t-1} \times f_t$ ”,就得到了更新后的单元状态 C_t ;第 4 份与输出权重 W_4 相乘再加上偏置项 b_4 后,经过激活函数 S 得到输出门 Y_t ,以此决定下一个状态要输出哪些信息。当前的单元状态 C_t 经过 $\tanh$ 函数后与输出门 Y_t 相乘便可得到元胞的输出 O_t 。这便是 LSTM 元胞处理信息输入到输出的全过程。整个信息处理过程的具体计算为:

$$O_t = \tanh(f_t \times C_{t-1} + I_t \times U_t) \times Y_t.$$

其中: $f_t = S(W_1 \times [A_{t-1}, X_t] + b_1)$, $I_t = S(W_2 \times [A_{t-1}, X_t] + b_2)$, $U_t = \tanh(W_3 \times [A_{t-1}, X_t] + b_3)$, $Y_t = S(W_4 \times [A_{t-1}, X_t] + b_4)$ 。 $[\cdot, \cdot]$ 表示把两个向量连接成一个更长的向量。

在循环神经网络中,由于神经网络层数较多,每层的梯度大小会呈现指数下降或增长,从而出现梯度消失和梯度爆炸问题,而 LSTM 网络能很好地解决这个问题。LSTM 网络由许多互相连接的 LSTM 元胞组成,它的基础架构为双向 LSTM 元胞(图 2)。双向 LSTM 存在两种类型的元胞:前向元胞和后向元胞。双向 LSTM 的输出层,不仅依赖之前的输入,还会依赖后面的元素,这样就增加了信息的交流与维持,充分利用了数据。向前隐藏层从 0 时刻到 n 时刻正向计算一遍,并保存每个时刻向前隐藏层的输出,向后隐藏层从时刻 n 到时刻 0 反向计算一遍,并保存每个时刻向后隐藏层的输出,将每个时刻向前隐藏层和向后隐藏层的输出

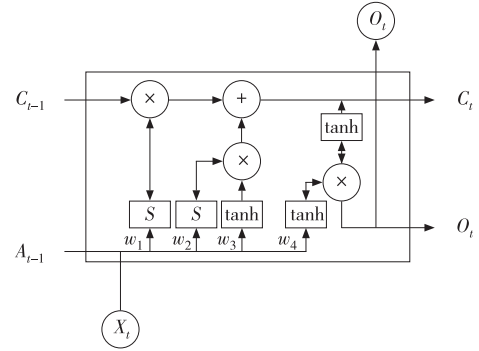


图 1 传统 LSTM 元胞信息处理过程

Fig. 1 Traditional LSTM cellular information processing process

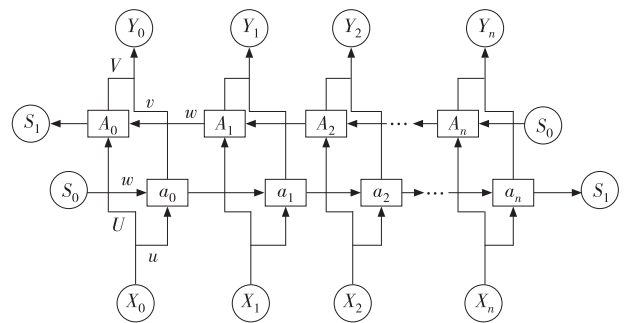


图 2 双向 LSTM 元胞

Fig. 2 Bidirectional LSTM cellular

相结合,就得到了最后的输出。具体某一时刻的输出过程如下: X_1 和 Y_1 分别表示 $t=1$ 时刻的输入和输出, X_1 乘以输入层与向前隐藏层之间的权重 u ,前一时刻向前隐藏层的输出 a_0 乘以向前隐藏层之间的权重 W ,再将两者相加经过激活函数 f 便得到当前时刻向前隐藏层的输出 a_1 ; X_1 乘以输入层与向后隐藏层之间的权重 U ,前一时刻向后隐藏层的输出 A_2 乘以向后隐藏层之间的权重 w ,再将两者相加经过激活函数 f 得到当前时刻向后隐藏层的

输出 A_1, a_1 乘以向前隐藏层与输出层之间的权重 v , A_1 乘以向后隐藏层与输出层之间的权重 V , 再将两者相加经过分类函数 S 得到当前时刻的输出 Y_1 。由此可以得知双向 LSTM 输出的信息不仅是由当前输入的信息决定, 而且需要向前的输出和向后的输出共同决定的。具体计算方法为:

$$Y_1 = S(V \times A_1 + v \times a_1)。$$

其中: $A_1 = f(X_1 \times U + w \times A_2)$, $a_1 = f(X_1 \times u + a_0 \times W)$ 。这里 f 是激活函数, 通常有 Relu 和 tanh 等激活函数; S 是分类函数, 通常用于二分类的是 sigmoid 函数, 用于多分类的是 softmax 函数。

从 LSTM 元胞以及双向 LSTM 元胞的信息处理过程不难看出激活函数在神经网络中起着非常重要的作用, 选择合适的激活函数可以很大程度的提高神经网络的准确率。通常在双向 LSTM 中用到的激活函数有两种, 分别是 sigmoid 激活函数和 tanh 激活函数, 这两个激活函数的具体形式分别为: $S(x) = \frac{1}{1+e^{-x}}$, $\tanh(x) =$

$$\frac{e^x - e^{-x}}{e^x + e^{-x}}。$$

1.2 学习率

学习率是在神经网络进行训练时优化算法中的一个参数。在梯度项前乘以学习率, 从而可以完成权重的更新, 最后收敛到最小值。学习率的设定会对网络的收敛速度和训练结果产生重要的影响。若学习率设定过小, 随着模型的训练, 损失函数值会收敛过慢, 使得模型需要消耗更多时间才会达到较理想的状态, 如图 3 所示。若学习率设定过大, 随着模型的训练, 会使损失函数发生振荡, 不仅消耗更多的时间, 也可能达不到最小值, 如图 4 所示。所以对于神经网络来说, 学习率的设定非常重要。

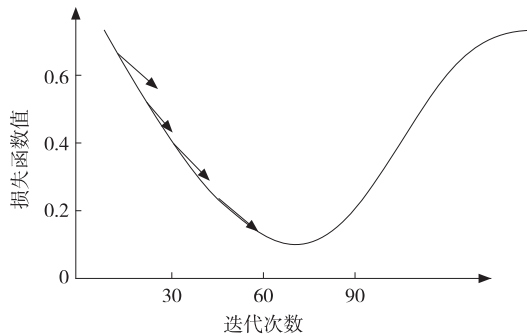


图 3 学习率设置过小的收敛过程

Fig. 3 The convergence process when the learning rate is set too small

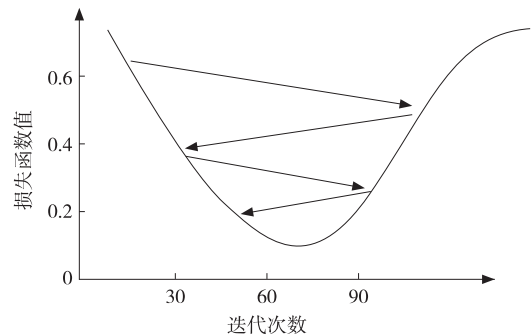


图 4 学习率设置过大的收敛过程

Fig. 4 Convergence process with excessively large learning rate

2 核心算法建立

2.1 激活函数改进

门控采用的是 sigmoid 函数。sigmoid 函数能够使计算结果大多数为 0 或 1, 这样就满足了门控的要求, 即完全抑制和完全激活。在生成候选记忆时采用的是 tanh 函数, 因为 tanh 函数的输出值在 $[-1, 1]$ 区域内, 与集中分布在 0 附近邻域相吻合, 而且 tanh 函数的输入也大多在 0 的附近邻域。相比其他激活函数, tanh 函数在 0 附近邻域拥有更大的梯度, 所以相比其他函数要收敛得更快。

本文提出一种基于 tanh 函数的改进函数, 增大了中心区域的梯度, 同时也不会使梯度过快的饱和, 具体表达式为:

$$f_{\tanh}(x) = \frac{e^{2x} - e^{-\frac{x}{2}}}{e^{2x} + e^{-\frac{x}{2}}}。$$

由此得到的中心梯度曲线见图 5。图 5 中实线代表原始的 tanh 函数, 点划线代表本文提出的函数, 虚线代表梯度更大的函数。

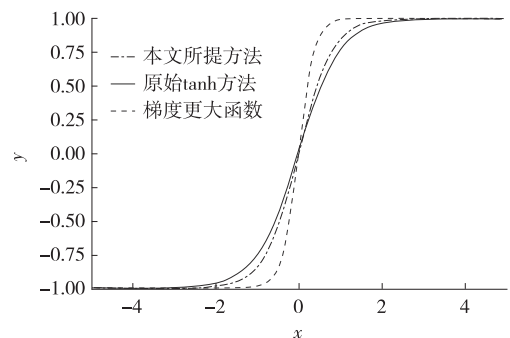


图 5 改进的 tanh 函数

Fig. 5 Improved tanh function

从图 5 中可以看出,本文提出的函数不仅增大了 0 中心区域的梯度,而且使函数的梯度不会过快地饱和,非常接近于原始的 tanh 函数。从原始的 tanh 函数曲线可以看出,一味地增大梯度会导致函数过早的趋于梯度饱和,从而影响收敛速度。

2.2 学习率改进

传统的学习速率设定主要根据先验经验设定和通过穷举来挑选出最优的学习率。但是这些方法,对网络的训练会造成很大的困难,不仅模型训练成本高,会消耗很大的资源 and 时间,最后也不一定会得到较好的学习率。学习率设置过高,收敛速度虽然变快,但是很难达到较高的准确率;学习率设置过小,可以达到较高的准确率,但是收敛速度较慢。

本文提出一种自适应的算法,可以根据迭代的次数来优化学习率,当迭代次数较小时,学习率的降低会比较快,随着迭代次数的增加,学习率变化的越来越慢,这样就能够同时满足收敛快,精度高的问题。具体计算方法如下:

$$l = l_0 \times e^{-\frac{i}{L}}$$

其中: l_0 是初始的学习率,它的设置可以较大, i 表示第 i 次迭代, L 表示网络训练的总迭代次数。从公式中可以看出,当 i 从 0 开始时,代表第 1 次迭代,公式的值为 l_0 ,当 i 很小时, $e^{-\frac{i}{L}}$ 的值大于 0,接近于 1,当 i 增大时, $e^{-\frac{i}{L}}$ 的值会不断减小并趋于 e^{-1} ,这样就能保证学习率不会过低。由 e 为底数的指数函数的特点可知,随着迭代次数的增加,值会变小并且变化率也会变小,也就是迭代次数越高,学习率的变化越慢,使得精确率越高。本文设置 l_0 初始值为 0.005, L 为 600。

2.3 双向 LSTM 改进

将改进的激活函数应用于 LSTM 元胞中,再定义两种类型的 LSTM 元胞,即前向元胞和后向元胞。将前向元胞和后向元胞的输出作为输入,构建双向循环神经网络。

损失函数采用的交叉熵损失函数为:

$$f_{\text{Loss}} = - \sum_{i=1}^n y_i \times \ln(\hat{y}_i)$$

其中: y_i 表示与 i 对应的真实值, $\hat{y}_i$ 表示与 i 对应的预测值。从上式可以看出该函数具有结构简单,计算量小等优点,所以应用较为广泛。

将改进的学习率应用到优化函数当中。优化函数采用的 Adam 优化函数。

原始输出经过 softmax 函数之后,输出符合概率分布的参数。训练过程中为了防止过拟合^[17],本文采用 Dropout 方法^[18],参数设为 0.8。具体来说,使得某个神经元以 0.8 的概率停止工作,在这次训练过程中不会更新权重,也不参加计算,但是它的权值要保留下来,因为下一次运算它有可能会工作。

3 数据集与实验结果

3.1 数据集

本文用来测试的数据集有两种,一种是 MNIST 数据集。MNIST 数据集中含有 60 000 个训练样本和 10 000 个测试样本。其中训练样本具体形式如图 6 所示。

另一种是 IMDB 数据集,该数据集用于分类电影评论的好与坏,其中有 25 000 条作为训练集,25 000 条作为测试集。IMDB 数据集的测试样本形式如表 1 所示。

3.2 对比模型

在 MNIST 数据集上的实验分别采用了 LeNet-DL 网络、K-近邻、卷积自编码器神经网络和基准双向 LSTM 等 4 个模型,然后将这些模型的实验结果与本文改进的双向 LSTM 实验结果进行对比。

在 IMDB 数据集上的实验则采用了 lda1-svd、lda3-svm;LDA、word2vec-svm 和基准双向 LSTM 等 4 个模

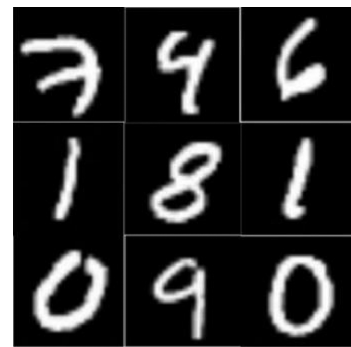


图 6 MNIST 数据集的一个训练样本

Fig. 6 A training sample of MNIST dataset

型,然后将这些模型的实验结果与本文改进的双向 LSTM 实验结果进行对比。

表 1 IMDB 数据集测试样本

Tab. 1 IMDB data set training samples

数据类型	样本的数据内容	标签
测试样本	But these children are amazing and shouldbe praised for what they have done……	正面
测试样本	This is an example of why the majority of action films are the same……	负面
测试样本	That is true to scotland this one is probably unique if you maybe……	正面
测试样本	Not even the Beatles could write songseveryone liked,and although Walter……	负面
测试样本	I liked the look of the film and how shots were set up and i thought it didn't rely……	正面
测试样本	This German horror film has to be one of the weirdest I have seen……	负面

3.3 实验参数

改进的双向 LSTM 中前向 LSTM 元胞和后向元胞都采用 256 个隐藏神经元,遗忘参数设为 0.8,初始学习率设为 0.005。在 MNIST 数据集上采用的是 softmax 分类函数,批处理样本大小为 128,在 IMDB 数据集上采用的是 sigmoid 分类函数,批处理样本大小为 32。同时在 IMDB 数据集上需要统一输入数组长度,本文采用最大句子长度为 500,即大于 500 词的评论截断至 500 词,小于 500 词的用 0 填充。

3.4 实验结果和分析

实验的硬件环境为:CPU 主频为 2 GHz,Intel i5 处理器;显卡型号为 AMD Radeon R5,显存为 2 GB;内存容量为 8 GB。实验的软件环境为:Windows 10 操作系统,Python 编程语言,采用 TensorFlow 深度学习框架。实验首先对模型的收敛速度进行对比。

表 2 为基准的双向 LSTM 和改进的双向 LSTM 在 MNIST 数据集上的收敛情况对比。从表 2 可以看出基准的双向 LSTM 在迭代训练了 140 次达到了 82% 的准确率,损失函数值为 0.519 259,而改进的双向 LSTM 在迭代训练了 140 次的时候,准确率已经达到了 93.359%,并且损失函数值为 0.183 621。通过对比可知:改进的双向 LSTM 在迭代次数相同的情况达到了较高的训练准确率,并达到较低损失函数值。实验证明改进的双向 LSTM 收敛速度更快。

表 2 MNIST 数据集上对应的收敛情况

Tab. 2 Convergence of the corresponding MNIST dataset

迭代次数	基准的双向 LSTM 在 MNIST 收敛情况		改进的双向 LSTM 在 MNIST 收敛情况	
	损失值	准确率	损失值	准确率
70	0.871 526	0.690 00	0.683 741	0.750 00
140	0.519 259	0.820 00	0.183 621	0.933 59

表 3 为基准的双向 LSTM 和改进的双向 LSTM 在 IMDB 数据集上收敛情况的对比。结果表明:基准的双向 LSTM 在迭代了 104 次后,准确率达到 53.88%,损失函数值为 0.691 6;改进的双向 LSTM 在迭代了 104 次后,准确率达到 88.08%,损失函数值为 0.281 5。改进的双向 LSTM 在相同的迭代次数时达到了较高的准确率,而且损失函数值也低于基准双向 LSTM 的损失函数值,不难看出改进的双向 LSTM 收敛速度更快。

表 3 IMDB 数据集上对应的收敛情况

Tab. 3 Convergence of the corresponding IMDB dataset

迭代次数	基准的双向 LSTM 在 IMDB 收敛情况		改进的双向 LSTM 在 IMDB 收敛情况	
	损失值	准确率	损失值	准确率
104	0.691 6	0.538 8	0.281 5	0.880 8

关于准确率的对比结果显示:在 MNIST 数据集上经过 1 630 次迭代训练之后,损失趋于稳定,对 MNIST 数据集进行测试,准确率已经稳定的达到了 99.61%,如表 4 所示。由本文所提方法与目前主流方法进行的实验结

果对比可知,本文提出的方法的准确率要高于其他主流方法。

同样的在 IMDB 数据集上进行实验,改进的双向 LSTM 的准确率达到 88.08%,相比基准双向 LSTM 的准确率有所提高,同样也优于其他对比模型(表 5)。

表 4 MNIST 数据集上采用的方法以及对应的准确率

Tab. 4 The methods used on the MNIST data set and the corresponding accuracy rates

模型	准确率
LeNet-DL 网络	99.34%
K-近邻	89.81%
卷积自编码器神经网络	99.37%
基准双向 LSTM	98.43%
本文改进的 LSTM	99.61%

表 5 在 IMDB 数据集上采用的方法以及对应的准确率

Tab. 5 The methods used on the IMDB data set and the corresponding accuracy rates

模型	准确率
lda1-svd	81.68%
lda3-svm:LDA	80.84%
word2vec-svm	82.20%
基准双向 LSTM	86.48%
本文改进的 LSTM	88.08%

4 结束语

本文提出了改进的双向 LSTM,通过改进 LSTM 中输入门与输出门的激活函数,并结合改进的学习率,在 MNIST 手写数字识别数据集和 IMDB 电影评论数据集上进行实验,实验表明,改进的双向 LSTM 相比基准双向 LSTM 收敛速度更快,准确率更高,由此证明改进的双向 LSTM 优于基准双向 LSTM,未来将会把改进的双向 LSTM 应用到更多的场景中。

参考文献:

- [1] JEONG J H, YU B W, LEE D H, et al. Classification of drowsiness levels based on a deep spatio-temporal convolutional bidirectional LSTM network using electroencephalography signals[J]. Brain Sciences, 2019, 9(12): 348.
- [2] CHENG H J, XIE Z, WU L H, et al. Data prediction model in wireless sensor networks based on bidirectional LSTM[EB/OL]. (2019-08-13)[2021-03-25]. <https://link.springer.com/article/10.1186/s13638-019-1511-4>.
- [3] HUANG Z H, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging[EB/OL]. (2015-08-09)[2021-03-25]. <https://arxiv.org/abs/1508.01991>.
- [4] IIMI N, BUDI W, NUR R K. Handwriting digit recognition using local binary pattern variance and K-nearest neighbor classification[EB/OL]. (2016-09-22)[2021-03-25]. <https://ieeexplore.ieee.org/document/7571937>.
- [5] MAJHI B, PUJARI P. On development and performance evaluation of novel Odia handwritten digit recognition methods[J]. Arabian Journal for Science and Engineering, 2018, 43(8): 3887-3901.
- [6] SHOPON M, MOHAMMED N, ABEDIN M A. Image augmentation by blocky artifact in deep convolutional neural network for handwritten digit recognition[EB/OL]. (2017-04-03)[2021-03-25]. <https://ieeexplore.ieee.org/document/7890867>.
- [7] LI W, ZEILER M D, ZHANG S X, et al. Regularization of neural networks using DropConnect[EB/OL]. (2017-04-03)[2021-03-25]. <https://dl.acm.org/doi/abs/10.5555/3042817.3043055>.
- [8] KESERWANI P, ALI T, ROY P P. Handwritten Bangla character and numeral recognition using convolutional neural network for low-memory GPU[J]. International Journal of Machine Learning and Cybernetics, 2019, 10(12): 3485-3497.
- [9] QIAO J, WANG G, LI W, et al. An adaptive deep Q-learning strategy for handwritten digit recognition[J]. Neural Networks, 2018, 107: 61-71.
- [10] MONTAZER G A, SOLTANSHAH M A, GIVEKI D, et al. Farsi/Arabic handwritten digit recognition using quantum neural networks and bag of visual words method[J]. Optical Memory and Neural Networks, 2017, 26(2): 117-128.
- [11] HOSSEINZADEH H, RAZZAZI F, KABIR E. A weakly supervised large margin domain adaptation method for isolated handwritten digit recognition[J]. Journal of Visual Communication and Image Representation, 2016, 38: 307-315.
- [12] NEMMOUR H, CHIBANI Y. Handwritten digit recognition based on a Neural-SVM combination[J]. International Journal of Computers Applications, 2010, 32(1): 104-109.
- [13] 王建仁, 马鑫, 段刚龙, 等. 边缘智能背景下的手写数字识别[J]. 计算机应用, 2019, 39(12): 3548-3555.

WANG J R, MA X, DUAN G L, et al. Handwritten digit recognition under edge intelligent background[J]. Journal of Computer

Applications, 2019, 39(12): 3548-3555.

- [14] 曾文献, 孟庆林, 郭兆坤. 基于深度卷积自编码神经网络的手写数字识别研究[J]. 计算机应用研究, 2020, 37(4): 1239-1243.
ZENG W X, MENG Q L, GUO Z K. Research on handwritten digit recognition based on deep convolution self-coding neural network[J]. Application Research of Computers, 2020, 37(4): 1239-1243.
- [15] 闵锋, 叶显一, 张彦铎. 基于改进主成分分析网络的手写数字识别方法[J]. 华中科技大学学报(自然科学版), 2018, 46(12): 101-105.
MIN F, YE X Y, ZHANG Y D. Handwritten digit recognition method based on improved principal component analysis network [J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2018, 46(12): 101-105.
- [16] 侯艳路, 丁世飞, 孙统风. 混合深度学习模型 C-RF 及其在手写数字识别中的应用[J]. 数据采集与处理, 2018, 33(2): 343-350.
HOU Y L, DING S F, SUN T F. Hybrid deep learning model C-RF and its application in handwritten digit recognition[J]. Journal of Data Acquisition and Processing, 2018, 33(2): 343-350.
- [17] HAWKINS D M. The problem of overfitting[J]. Journal of Chemical Information and Computer Sciences, 2004, 35(19): 1-12.
- [18] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223.

Research on Bidirectional LSTM Based on Improved Activation Function and Learning Rate

LI Jun¹, LI Ming¹, ZENG Zheng², LI Li³

(1. College of Computer and Information Science, Chongqing Normal University, Chongqing 401331;

2. College of Journalism and Media, Chongqing Normal University, Chongqing 401331;

3. College of Computer and Information Science, Southwest University, Chongqing 400715, China)

Abstract: [Purposes] With the wide application of bidirectional LSTM recurrent neural networks in speech recognition, image recognition and sentiment classification, it has become especially important to improve the accuracy of bidirectional LSTM recurrent neural networks. [Methods] An improved bidirectional LSTM recurrent neural network is proposed, which greatly improves the convergence speed and accuracy of neural network by improving the input gate and output gate activation functions in LSTM and combining with the improved learning rate. [Findings] Experiments were conducted on different domain data sets MNIST and IMDB, and the accuracy reached 99.61% on the MNIST data set, which is 1.172% higher than that of the benchmark bidirectional LSTM, and 88.08% on the IMDB data set, which is 1.6% higher than that of the benchmark bidirectional LSTM, and on both data sets, higher accuracy and lower loss rate are achieved with smaller number of iterations compared to the benchmark bidirectional LSTM. [Conclusions] This proves that the improved bidirectional LSTM recurrent neural network outperforms the benchmark bidirectional LSTM recurrent neural network.

Keywords: bidirectional LSTM; activation function; learning rate; input gate; output gate

(责任编辑 许 甲)