深层神经网络LSTM

问题分析

本文选取长短时记忆网络（LSTM）来解决。。。。。。。，这是一种特殊的循环神经网络，可以有效地规避其他神经网络存在的在长时间的时间序列中有用信息被遗忘的问题，即长期依赖问题。LSTM神经网络近些年来被广泛应用在时间序列的相关问题方面。本问题。。。。。。。。。。。。。。。，适用于该模型。传统递归神经网络（RNN）无差别的记忆所有信息，导致重要信息随着时间的推移丢失。在实际应用中，RNN还容易出现梯度消失和梯度爆炸等问题。因此1997年Hochreiter和Schmidhuber提出了长短期记忆网络模型（LSTM），在RNN模型的基础上进行了改良，有效解决了RNN容易出现梯度消失的问题[1]。1999年，FelixA.Gers等人发现Hochreiter和Schmidhuber提出的LSTM在处理连续输入数据时，需要重置网络内部的状态，否则会导致网络崩溃。因此，他们在原来LSTM的基础上加入了门控机制，在神经元中设置了遗忘门、输入门和输出门这3个“门”结构，使状态信息可以被添加和筛除[2]。使得LSTM能够重置状态。在深层次网络结构中，也能留下重要信息并传递下去，可以解决长期时序信息处理时梯度消失的问题。基于以上特征，LSTM神经网络能够对。。。。。。。做出比较精准的预测。

[1]杨凯,唐卫东,刘诚,贺景龙,姚川.基于LSTM循环神经网络的大地电磁方波噪声抑制[J].物探与化探,2022,46(04):925-933.

[2]孟秋静,杨钢.基于LSTM神经网络模型的液压管路故障诊断方法[J/OL].机电工程:1-9[2022-08-27].http://kns.cnki.net/kcms/detail/33.1088.TH.20220712.1032.009.html

评价

文中采用的LSTM神经网络具有结构简洁、泛化能力强等特点，对。。。。。具有适应性和实用性，并且该方法具有较强的鲁棒性能。

推广

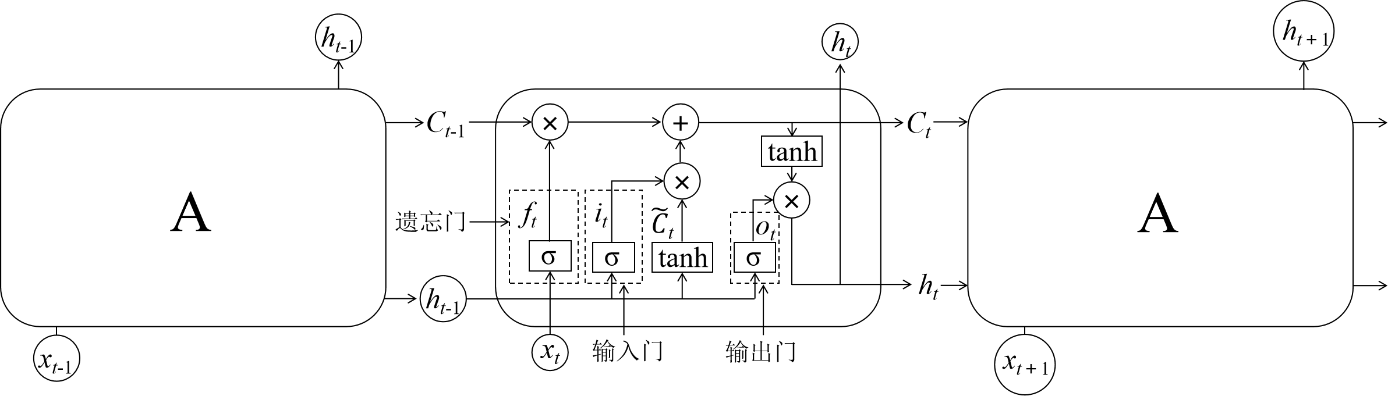
本文运用LSTM神经网络模型解决了。。。。。。。。。。的问题。LSTM神经网络具有强大的提取时序信息的能力，通常用于处理存在长期依赖问题的模型，广泛应用在语音识别、自然语言处理、负荷预测以及时间序列相关的领域，近年来也被逐渐引入到多媒体相关的音频和视频数据分析、图片标题建模，道路运输相关的交通流速预测，医学相关的蛋白质二级结构序列预测等领域之中[1]。

[1]熊波,李肖霖,王宇晴,张瀚铭,刘子君,丁锋,赵必强.基于长短时记忆神经网络的中国地区电离层TEC预测[J].地球物理学报,2022,65(07):2365-2377.

模型建立

LSTM是一种特殊的递归神经网络，它利用时间序列对输入进行分析，可以处理时间维度上的信息，解决RNN神经网络中梯度爆炸和梯度消失的问题。LSTM模型在近些年被广泛使用，适合处理和预测时间序列中时间跨度较大的事件，能够对事物的发展做出比较精准的预测。

一个LSTM细胞拥有三个门，分别是遗忘门（门），输入门（门）和输出门（门）。遗忘门将部分无用信息剔除，输入门确定给记忆细胞添加哪些信息，输出门筛选需要输出的信息。神经元具体结构展开，如图。。所示。

* + - * 1. 图。。 神经元结构(LSTM)

（1）遗忘门，通过时刻的输出和现在输入的数值，可以计算得出遗忘门时刻的输出。通过sigmoid来控制，使输出一个介于0∼1的值，以此来决定是否让上一时刻学到的信息保留或者部分保留，如同式。。。所表示的一样：

|  |  |
| --- | --- |
|  | () |
|  | () |

式中：为遗忘门，取值范围是0∼1，取值越小则遗忘越多，反之则越少；为激活函数sigmoid；为时刻的输出；为时间为时的输入值；为权重参数；为偏置项。

（2）输入门，输入最后时间的数值和目前时间的数值，而后计算得出输入门的输出数值以及和候选单元状态情况，式。。。可表示：

|  |  |
| --- | --- |
|  | () |
|  | () |
|  | () |

式中：为输入门介于0∼1之间；是在时刻从目前输入中提取的的细胞状态值；tanh是双曲正切激活函数；，为权重参数，，为偏置项。

然后需要输入细胞状态，目前的单元状态属于遗忘门的数值和上一时间状态相乘所得然后再加上输入门两部分的相乘结果，如式。。。所示：

|  |  |
| --- | --- |
|  | () |

式中：是同样在时刻但输入之后的细胞状态值；为上一次遗忘关于信息的程度；为输入门由sigmoid层和tanh层两个部分来共同作用，以此来控制候选细胞保留下来的信息。

（3）输出门，输出的数值和输入的数值表示在t时刻的输入门的输入值，而后得出输出门如下：

|  |  |
| --- | --- |
|  | () |
|  | () |

式中：为当前的输出门；是时刻在的控制下对输入进行计算产生的输出；为权重参数；为偏置项。

从以上介绍可以看出，当遗忘门发挥作用时，细胞状态的梯度可以传递给时刻的，这就是LSTM模型解决梯度消失和梯度爆炸问题的关键所在。

算法

（1）将xx的数据作为训练数据集，xxx的数据作为测试数据集，为了数据更快收敛将数据进行归一化处理。

（2）通过遗忘门将我们不需要的信息从细胞状态中移除，以上一步输出的和这一步输入的作为输入，然后为里的每个特征信息输出一个0～1间的值，记为，表示保留多少信息，使和相乘，当为0时表示上个单元的该特征不保留在中。

（3）决定此时细胞状态里存什么，除了需要对上一时刻留下的信息筛选，还要把新的信息选择性的记录在细胞状态中。同样sigmoid决定我们要输入的特征，这个概率表示为，tanh创建一个候选细胞状态值，将会被增加到细胞状态中，然后输入细胞状态时把这两个结合起来。

（4）通过sigmoid层从而决定输出的哪些部分，得到一个初始的输出结果。而后使得细胞状态通过tanh后将值缩放到-1∼1，再与sigmoid的数值相乘得到最终的输出结果。

（5）停止条件（迭代次数？损失函数？）

损失函数选用最小均方误差（MSE）。30个样本计算一次梯度下降，经过1000次训练后, 损失函数值为 0.0107。

设置模型常量：采用函数式构造流程，设置2层LSTM，1层全连接，1层dropout，最后连接1层全连接。预测输入神经元个数为5个，LSTM单元中的神经元个数为10，LSTM单元个数为7，输出神经元个数为1，学习率为0.0007。神经网络变量：随机产生输入和输出层的权重，以及dropout参数。LSTM的输入样本为（samples, timesteps, features）,根据训练集生成时间序列样本数据集为（249，7，1）共249个样本，测试集生成时间序列样本数据集为（125，7，1）共125个样本，lookback设置为7，用前7天的数据来预测下一天的数据。32个样本计算一次梯度下降，经过100次训练后，损失函数值为，效果较理想。