

一种融合 EMD 分解和 LSTM 网络的频谱占用度预测模型

赵晓东¹ 苏公瑾² 李克利² 成杰² 徐江峰¹

1 郑州大学信息工程学院 郑州 450001

2 河南无线电管理信息系统备份中心 郑州 450000

(1024729633@qq.com)

摘要 频谱占用度是衡量频谱利用率、反应频谱分配是否合理的重要依据,但是非稳态的频谱占用度序列为有效的预测带来了巨大的挑战。文中提出了融合 EMD 与 LSTM 的计算模型(EMD-LSTM),该模型首先对原始占用度序列进行经验模态分解(EMD),令其生成含有不同时间尺度的本征模函数(IMF),然后用 Pearson 相关系数选择出相关度高的 IMF,并将其与频谱占用度序列进行融合,最后利用长短时记忆网络(LSTM)对融合序列进行占用度预测。仿真实验结果及分析表明,相比于普通的 LSTM 网络,新的模型在预测频谱占用度变化上有了较大的性能改善。

关键词 频谱占用度;长短时记忆;网络经验模态;分解 EMD-LSTM

中图分类号 TP183

Spectrum Occupancy Prediction Model Based on EMD Decomposition and LSTM Networks

ZHAO Xiao-dong¹, SU Gong-jin², LI Ke-li², CHENG Jie² and XU Jiang-feng¹

1 School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

2 Henan Radio Management Information System Backup Center, Zhengzhou 450000, China

Abstract Spectrum occupancy is an important basis to measure the spectrum utilization rate and reflect whether the spectrum allocation is reasonable. However, the unsteady spectrum occupancy sequence presents great challenges for effective prediction. In this paper, a new computing model (EMD-LSTM) combining EMD and LSTM is proposed. Firstly, the empirical mode decomposition (EMD) of the original occupancy sequence is used to generate the Intrinsic Mode Function (IMF) with different time scales, and then the highly correlated IMF is selected by Pearson correlation coefficient. Then, IMF and spectrum occupancy sequence are fused, and the occupancy sequence is predicted by using the long and short time memory network (LSTM). Simulation experiments and analysis show that, compared with the ordinary LSTM network, the new model has a great improvement in predicting the change of spectrum occupancy.

Keywords Spectrum occupancy, Long-term and short-term memory network, EMD, EMD-LSTM

1 引言

近年来,通信领域快速发展,通信需求不断增长,无线通信依赖的资源有限,通信的实现依赖电磁信道,而信道来自于无线频谱的划分,因此通过对无线频谱的合理划分来实现高效管理,从而实现对频谱资源的有效利用。

频谱管理的 3 个重要因素是:频谱有效利用、频谱共享和经济因素,其中最重要的就是频谱的有效利用。频谱占用度是频谱利用率有效的表达手段,它是某段时间在某个区域内发射信号场强超过门限的时间比例,可以为频谱管理人员提供实时的频谱利用信息,为实现频谱时间、空间复用提供数据支撑;同时,它还能向上级主管部门提供频谱利用率的时空变化轨迹,为频谱管理的宏观政策的制定提供重要依据。

对于频谱占用度的研究从其自身的随机性质开始,文献[1]采用马尔可夫模型,利用信道状态转移概率反映信道状态的相关性来进行频谱使用情况的预测,这种方法在信道采集

数据随时间呈指数增长时遇到很大困难。文献[2]采用自回归移动平均(ARMA)模型来研究频谱占用度;文献[3]对频谱占用状态的时域相关特性进行了研究,构建频谱占用状态变化等效模型;这两种常用的时序研究方法都要求原始序列是平稳序列,对于非平稳的频谱占用度序列通常需要先采用差分平稳化的方法处理数据,但其平稳化效果较差。文献[4]基于数学统计讨论的方法,给出了频谱占用度的某些特定条件下随时间变化的特性,通用性较差;文献[5]结合神经网络算法给出了一种非平稳非线性频谱占用度时间序列分析方法,提出了 EMD 和 ANN 相结合的建模预测模型,该模型对 EMD 分解后的所有特征序列进行预测,将序列结果重构为预测序列,计算量大,也未能规避多个 ANN 网络分别预测带来的噪音问题,在处理数据时未依据 ITU 的一般规则利用 5/15 分钟时长的监测数据计算频谱占用度;此外,ANN 模型在拟合模型时通常假设输入与输出之间相互对应,并不考虑输入维度之间的时间顺序关系,这在序列预测的实践中实际上损

本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家重点研发计划

This work was supported by the National Key Research and Development Program of China.

通信作者:徐江峰(jfxu@zzu.edu.cn)

失了一个重要维度的信息。

另一方面,随着神经网络的发展,研究者意识到长短时记忆网络(LSTM)对于时序数据有着更好的刻画能力,并进行了广泛的研究应用,文献[6-7]对 LSTM 网络的结构原理进行了深入研究,文献[8-10]将 LSTM 网络用于语音识别、流媒体检测、文本情感分析等领域,验证了 LSTM 网络在处理时间序列预测问题时具有良好性能。

近年来,EMD 常被用于信号处理领域,这种以本征模态函数为基函数的信号处理方法克服了传统方法中的非自适应性。文献[11]对其进行了分析和研究,文献[12-15]对 EMD 在处理信号序列变化问题时的有效性进行了研究,证明了在面对非线性序列问题时,EMD 对比传统方法在数据处理中性能更好。

为了探索频谱占用度变化的规律,进一步提高预测结果的准确度,本文提出了一个基于 EMD 和 LSTM 的频谱占用度预测模型,模型中首先将原始序列与 IMF 融合,然后利用 LSTM 网络形成的新模型(EMD-LSTM)进行频谱占用度预测。

2 理论基础

2.1 EMD 分解

对于信号序列的处理,经常采用傅里叶变换进行时域频域的变化,以探究其不同频率上的特征。但是,在进行傅里叶变换时,任何一个局部的处理都是全局性的,而在实际生产环境中,很多序列的真实信息很可能只是一个局部的信号,在短暂的时间内出现然后消失。

EMD,即经验模态分解^[11],是 1998 年 Huang 等提出的,它是在 Hilbert-Huang 变换基础上发展出来的一种非常重要的自适应时频信号处理方法,适用于对非平稳的、非线性的信号进行处理。在处理时,它将非线性非平稳信号序列分解成有限个具有不同特征尺度的平稳序列,每一个平稳序列即为一个本征模态函数(IMF)。

根据以上理论,IMF 要求满足以下条件:

(1)分解后的 IMF 序列中,极值点的个数和过零点的个数必须相等或最多相差一个。

(2)在序列的任意位置,局部极大值点和局部极小值点构成的上、下包络线的平均值为 0,上、下包络线在时间轴上是局部对称的。

但是,在实际分解过程中,严格满足上述两个条件的信号序列几乎不存在,因此不可能严格按照这个标准进行判断。为了保证算法的有效性,Huang 给出了结束筛选的标准,由标准差 SD 来决定:

$$SD = \frac{1}{T} \int_0^T \frac{|h_k(t) - h_{k-1}(t)|^2}{|h_{k-1}(t)|^2} dt \quad (1)$$

设待处理的原始信号为 $x(t)$, $C_i(t)$ 是第 i 个 IMF 函数, $r_i(t)$ 是 $x(t)$ 分离出 $C_1(t), C_2(t) \dots C_i(t)$ 后得到的残差序列,且令 $r_0(t) = x(t), s(t) = x(t)$,则寻找 IMF 的过程如下:

(1)找出 $s(t)$ 中的全部局部极大值和局部极小值,用三次样条插值将所有局部极大值点连接后形成上包络线;同样地,用三次样条插值将所有局部极小值点连接形成下包络线,上包络线与下包络线应包含所有数据。

(2)设 $u(t), l(t)$ 分别为生成的上、下包络线,计算上包络线与下包络线的均值:

$$m(t) = (u(t) + l(t)) / 2 \quad (2)$$

令 $h(t)$ 为 $s(t)$ 与 $m(t)$ 的差:

$$h(t) = s(t) - m(t) \quad (3)$$

(3)如果 $h(t)$ 满足 IMF 的条件,则令 $C_i(t) = h(t)$ 并转(4);否则令 $s(t) = h(t)$ 并转(1)。

(4)从 $r_{i-1}(t)$ 中分离 $C_i(t)$,得到:

$$r_i(t) = r_{i-1}(t) - C_i(t) \quad (4)$$

检查是否满足分解结束条件,若满足则停止分解进入步骤(5),否则令 $s(t) = r_i(t)$,返回步骤(1)继续执行,同时令 $i = i + 1$ 。

(5)分解结束。设最终得到的各个模式分量和残差序列分别为 $C_1(t), C_2(t), \dots, C_n(t)$ 和 $r(t)$,此时有:

$$x(t) = \sum_{i=1}^n C_i(t) + r(t) \quad (5)$$

残差序列 $r(t)$ 是信号序列 $x(t)$ 的集中趋势,分解获得的若干个 IMF 则反映了原始信号序列的不同时间尺度特征。

2.2 LSTM 网络

为了求解存在前后依赖关系的序列问题,Elman 提出了一个预测模型 RNN^[16],它是一个随时间展开的模型,每一个时刻都有输入输出,每个输出都要考虑之前输入序列计算的状态信息。但是普通的 RNN 网络存在着长期遗忘的问题^[14],在训练模型时无法达到很好的效果,因此文献[7]提出了长短时记忆网络(LSTM),其结构如图 1 所示,主要改进在于使误差在网络模型中以常量传播,同时为了让不同时序的输入对当前的影响变得可控,引进了遗忘门、输入门、输出门结构。

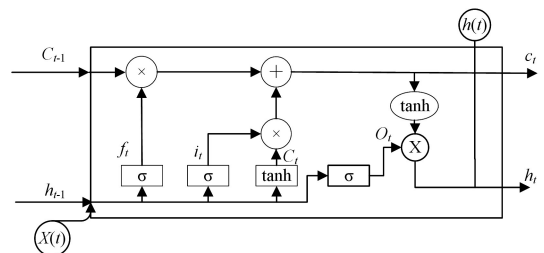


图 1 LSTM 结构

Fig. 1 LSTM structure

图 1 中, C 和 H 分别代表 LSTM 网络中长期状态和当前状态的信息。遗忘门中的 σ 函数在获得了当前的输入 $x(t)$ 和前一个细胞的输出 h_{t-1} 后,通过 σ 函数为上个状态的 C_{t-1} 中的每个值输出一个权重决定其是否保留。

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

输入门中的 σ 函数决定了当前输入和前一个输出结合之后的矩阵, \tanh 函数给出候选向量 \tilde{c}_t , 两者共同决定在下一步更新哪些值。

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

$$\tilde{c}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \quad (8)$$

更新信息分为两部分,第一部分为要丢掉的遗忘信息,是旧状态与 f_t 相乘的结果,第二部分是将是 \tilde{C}_t 与 i_t 相乘获得的更新信息,两者结合就是新的状态信息。

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (9)$$

最后,输出门中先将本层的输入 $X_{(t)}$ 与前一层的输出 h_{t-1} 结合,通过一个 σ 函数来获得一个过滤输出的矩阵 O_t , 细胞状态经过 \tanh 规范化后与 sigmoid 相乘的结果就是本层神

经网络的输出。公式中的 w^* , b^* 分别表示权重矩阵、偏置值, 神经网络通过反向传播更新这两个参数, 使模型能更好地拟合实际问题。

3 EMD-LSTM 模型

采样到的无线频谱信号是一个瞬时脉冲值, 根据国际电联的技术规范, 以 5/15 分钟为区间对无线频谱信号进行处理, 可以生成一个随时间变化的占用度序列。该序列数据的变化较为复杂, 混杂多种噪声, 传统模型在进行训练时经常出现寻找潜在规律困难、拟合准确率低的问题。为了解决这些问题, 本文提出一个融合 EMD 分解和 LSTM 网络的频谱占用度预测模型 (EMD-LSTM), 该模型首先通过 EMD 将序列中的各个特征分解出来, 然后根据各个 IMF 分量的 Pearson 系数大小来确定关键序列, 最后与原始序列相融合, 输入 LSTM 网络模型进行预测。

EMD-LSTM 网络模型如图 2 所示, 基本结构由 EMD 分解层、输入层、隐藏层、输出层、模型优化构成。

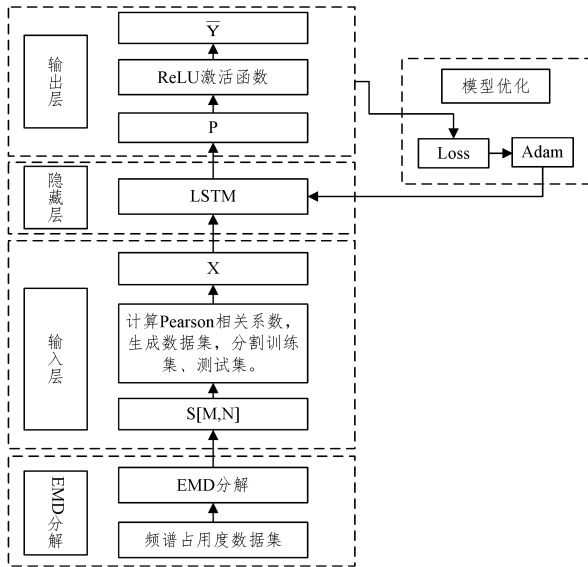


图 2 EMD-LSTM 网络模型
Fig. 2 EMD-LSTM network model

(1) EMD 分解

将原始电平信息进行处理后, 形成一个随时间变化的频谱占用度序列, 对占用度序列进行 EMD 分解获得一个与原始序列等长的多维矩阵, 根据 EMD 分解的原理, 设分解前频谱占用度序列为 $S = \{S_1, S_2, S_3, S_4, \dots, S_t\}$, 分解后生成 M 列 IMF 序列和一系列残差序列 R , 此时对 S 序列中的任意值有:

$$S_i = C_1(i) + C_2(i) + C_3(i) + \dots + C_M(i) + r(i) \quad (10)$$

(2) 输入层

对输入的 IMF 序列 $C_1(t), C_2(t), C_3(t), \dots, C_M(t)$ 以及残差序列 $r(t)$ 计算 Pearson 相关系数, 挑选合适的 IMF 序列融合原始序列构造输入序列, 假设挑选到的分解序列为 $C_l(t), C_m(t), C_n(t), \dots, C_q(t)$, 其中 $0 < l \leq q \leq M$, 假设预测模型以连续的 p 个序列值来预测第 $p+1$ 个值, 则输入 X 可以表示为:

$$X_i = \{C_l(i) + C_m(i) + C_n(i) + \dots + C_q(i)\} \quad (11)$$

$$X = \{X_i, X_{i+1}, X_{i+2}, X_{i+3}, \dots, X_{i+p}\} \quad (12)$$

式(11)中, X_i 代表模型预测时输入的序列中第 i 个值及其关

联的各个选中的 IMF 序列中的信息, 式(12)为某一次预测时的全部输入, 包含了 p 步中的所有的输入信息。

(3) 隐藏层

X 输入隐藏层后, 在由若干层 CELL 构成的网络中传播, 这些 CELL 相互之间既在空间上连接, 同时也在时间上相互连接, 通过 CELL 各个时期的计算, 输出 LSTM 各个时期的输出和状态。

$$H = \{H_i, H_{i+1}, H_{i+2}, H_{i+3}, \dots, H_{i+p}\} \quad (13)$$

$$C_n, H_n = f_{LSTM}(X_n, H_{n-1}, C_{n-1}) \quad (14)$$

其中, H 是隐藏层计算时的各个时期的细胞输出, C 是计算过程中的长期记忆, H_{i+p} 是第 p 步计算完成的输出, 也是隐层对外输出的内容, 根据神经元个数的不同, H_{i+p} 是具有对应维度的向量, 可以把其视为隐藏层在高维空间对序列的抽象预测。

(4) 输出层

输出层由一个 ReLU 激活函数构成, 接受 LSTM 输出 p 时刻的结果, 并将其映射为目标值 \bar{S} , 此时有:

$$\bar{S} = f_{ReLU}(H_{i+p}) \quad (15)$$

(5) 模型优化

1) 目标函数。在模型训练阶段, 前向传播结束后, 计算 loss 值, 其为整个模型的输出误差。本文中的 loss 选择均方根误差 (RMSE), 接着通过 BPTT 进行网络模型的训练。

2) 梯度下降算法。梯度下降算法是神经网络常用的训练手段, 研究者们通过多年的努力, 提出了各种梯度下降算法, 本文实验采用 Adam 算法, 该算法是带动量梯度下降算法结合加速梯度下降算法改进而来的一种算法。其特点在于 Adam 在带动量的梯度下降法的基础上, 引入了平方梯度, 并对速率进行了偏差纠正。

在模型训练时, Adam 算法将自适应地改变梯度下降的速率, 优化模型, 缩小 loss 的大小。

(6) 预测评估

模型训练结束后, 输入带标签的测试集来进行预测, 输出 RMSE, 以评估模型的有效性和稳定性。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S(i) - \bar{S}(i))^2} \quad (16)$$

其中, n 为计算误差时样本集的大小, $S(i)$ 为序列中第 i 个真实值, $\bar{S}(i)$ 为预测值。

EMD-LSTM 模型的设计理念是通过引入额外维度的相关信息来对抗模型训练中的不确定性。对数据进行特征分解, 可以得到数据在不同时间尺度的变化规律; 其次通过相关性计算, 获得那些真正对数据变化贡献较高的特征序列; 最后利用这些高相关性的序列参与模型的训练, 以达到提高模型精度的目的。

EMD-LSTM 模型的详细描述如算法 1 所示。

算法 1 EMD-LSTM

Begin: Step1: 将电平值数据转化为频谱占用度序列

Input value

If time > 5min:

Calc(Spectrum occupancy)

Sequence(Spectrum occupancy)

Return SOsequence

Step2: 将数据通过 EMD 算法分解为 IMF 序列

InputSOsequence

EMD(SOsequence)

```

Return IMFs
Step3:通过 Pearson 相关系数确定权重较大的序列
InputIMFS
pearson(IMFs)
return(IMF.pearson)0.2)
step4:通过神经网络训练模型,并评价模型的优劣
input([sequence,imfs],target)
train(EMD-LSTMtrain_set)
test(EMD-LSTMtest_set)
return train_set.RMSE
return test_set.RMSE
End
    
```

4 实验验证及分析

4.1 无线频谱数据的预处理

从原始数据中选取某地区固定站连续 3 天的 88~108 MHz 频谱信号作为实验样本,用来训练验证模型。监测站以固定 5 s 的频率对目标频段进行扫描,步长频率为 25 kHz,获得瞬时电平值数据 13824000 个。根据国际电信联盟频谱占用度测试标准,每 5 min 计算一次频谱占用度^[17],以选定的连续三天的频谱信号计算频谱占用度,获得如图 3 所示的占用度序列。

作为实验模型的 EMD-LSTM 网络模型,需要对数据进行 EMD 分解以获取数据更多维度的信息。

对数据进行 EMD 分解,获得的 IMF 成分序列和残差序列如图 4 所示,一共分解出 9 个特征序列。

一般地,皮尔逊相关系数从-1 到 +1 之间变化,正值代表正相关,负值代表负相关,其绝对值大小描述相关性有如下关系:

- 0.8~1.0 极强相关
- 0.6~0.8 强相关
- 0.4~0.6 中等程度相关
- 0.2~0.4 弱相关
- 0.0~0.2 极弱相关或无相关

采用 Pearson 相关系数绝对值大于 0.2 的 IMF 序列进行实验,如图 5 所示,采用 $C2(t), C3(t), C5(t), C6(t), r(t)$ 作为选中的 IMF 分量进行实验。将数据处理中获得的数据整理成(输入序列,标签)的形式,根据实际生产环境中的物理意义,时序数据中每个点是 5 分钟的占用度统计结果,综合考虑频谱数据的变化趋势,采用 10 个步长的序列预测下一个序列,即用 50 分钟的数据来预测下一个 5 分钟内的结果。

采用 10 次 10 折交叉验证来测试评估实验模型,将分割出的数据集平均分为 10 份,依次取出 9 份作为训练集,剩下 1 份作为测试集,循环 10 次,将 RMSE 作为主要评价指标输出。

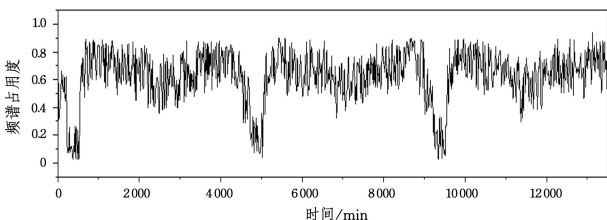


图 3 数据集
Fig.3 Data set

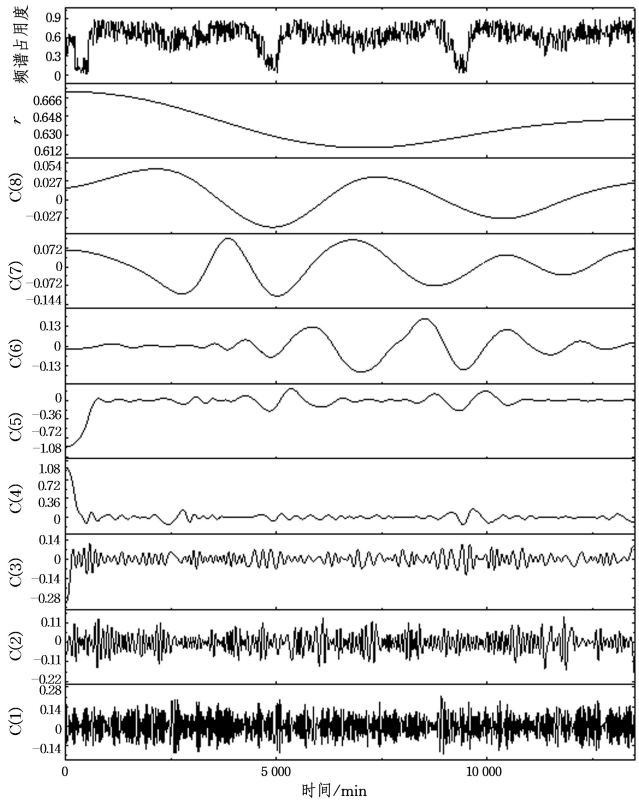


图 4 IMF 序列
Fig.4 IMF sequence

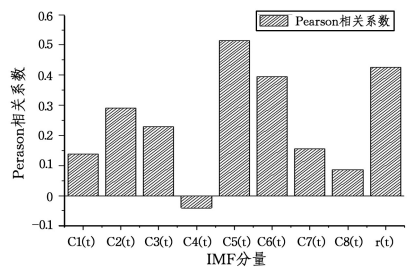


图 5 IMF 序列 Pearson 相关系数

Fig.5 Pearson correlation coefficients of IMF sequence

4.2 实验验证

4.2.1 模型收敛测试

为了评估模型的性能,本文采用了常用的 BPNN 网络、LSTM 网络进行实验对比,通过多次实验给出不同模型的最优实验结果,如图 6 所示。

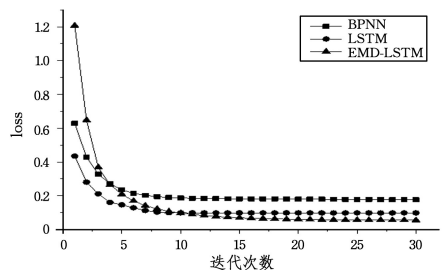


图 6 BPNN,LSTM 和 EMD-LSTM 训练误差下降曲线
Fig.6 Training error decline curve of BPNN,LSTM and EMD-LSTM

图 6 中,BPNN 为两层隐藏层,每层 50 个神经元,将连续 10 个时刻的占用度序列作为输入,输出层为一个 ReLU 函

数,结果映射为一个预测的占用度信息;LSTM网络和EMD-LSTM网络隐藏层都设计为50个神经元,LSTM网络的输入层设计为每步一个占用度信息,将连续的10个占用度信息以时间顺序输入,输出层为一个ReLU激活函数,将隐藏层的信息映射为一个预测占用度信息。

3个预测模型在30次迭代内都能收敛,BPNN在第15次迭代时即达到稳定状态,误差在0.2左右,LSTM网络收敛快速,在第10次迭代时收敛,误差在0.12左右,EMD-LSTM预测模型在20次时收敛,误差大约为0.05左右,其中EMD-LSTM网络模型的误差最小。

4.2.2 预测结果

为了验证预测模型的有效性,选取郑州市某基站一天的频谱占用度信息进行测试。针对目标问题,3个网络预测模型都进行了20次训练预测,选取最好的预测结果。图7—图10分别显示了ARIMA模型及3个预测模型的预测结果与真实结果的对比情况,图中灰色折线圆形图标为预测值,黑色折线方形图标为真实值,在4个网络模型中,ARIMA误差最大,测试误差在0.371左右,BPNN测试误差在0.327左右,LSTM网络优于BPNN,测试误差为0.175,EMD-LSTM最好,测试误差为0.087。

3个模型中,EMD-LSTM网络模型的效果最好,误差最小。

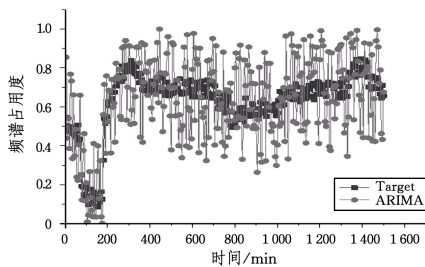


图7 ARIAM预测值和真实值的比较

Fig. 7 Comparison of predicted value and true value of ARIMA

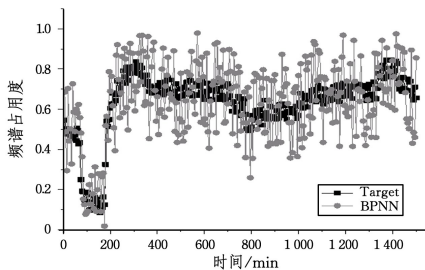


图8 BPNN预测值和真实值的比较

Fig. 8 Comparison of predicted value and true value of BPNN

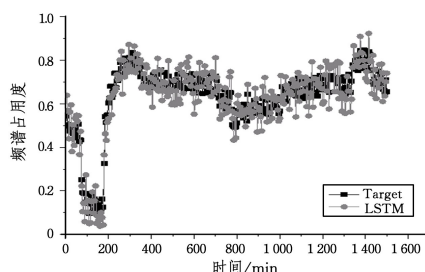


图9 LSTM预测值和真实值的比较

Fig. 9 Comparison of predicted value and true value of LSTM

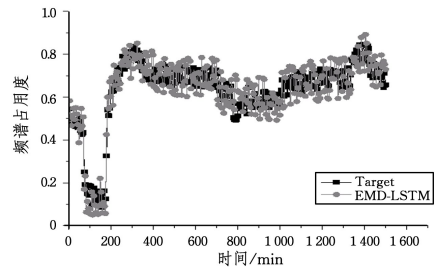


图10 EMD-LSTM预测值和真实值的比较

Fig. 10 Comparison of predicted value and true value of EMD-LSTM

4.3 模型性能分析

本文采用ARIMA模型、BPNN模型和LSTM网络模型作为实验对照组进行对比,采用十折交叉验证后,输出误差均值,如表1所列。

表1 各模型的RMSE值的比较

Table 1 Comparison of RMSE values of each model

模型	测试集	训练集
ARIMA	—	0.3102
BPNN	0.26376	0.24515
LSTM	0.17305	0.12199
EMD-LSTM	0.09858	0.05760

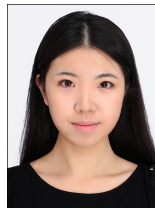
各个实验的结果符合预期,其中ARIMA在处理非线性数据时效果最差,BPNN模型在处理序列数据时,其输入到输出的映射是一次性完成的,没有时间序列的信息参与,因此,其结果较差;LSTM网络模型将输入的序列中的时间信息保留了下来,同时通过隐藏层的循环计算输入先后之间的关系,这种特殊的设计为时间序列的预测提供了理论基础,但是对于单一的序列,其中可能包含的噪音以及多层次的关系无法很好地刻画,LSTM网络模型虽然比BPNN网络有了很大的提升但仍有改进的空间。EMD-LSTM网络在处理频谱占用度序列这种只有一维的随时间变化的序列时,通过设计EMD的处理过程将序列中包含的信息从一维的序列中提取出来,直接提高了输入时的序列信息,为模型更好地刻画问题提供了助益,同时,以Pearson相关系数过滤了潜在的噪音,进一步提高了模型的性能。

本文模型主要是对使用较为频繁的88-108MHz的数据进行的预测研究,在面对其他频段,特别是广播领域较少利用的高频波段时,由于其应用情况复杂,寻找其存在的规律更难,因此需进一步的探索。

结束语 EMD分解近年来广泛地被用于非平稳数据的处理,区别于传统的滤波算法,这种算法能更好地寻找到时序序列的局部时间特征,有利于非平稳序列数据特征的提取。LSTM网络模型通过输入门、遗忘门、输出门的设计,解决了传统RNN结构网络存在的缺陷,提高了处理序列问题的能力。本文在此基础上提出了一种基于EMD的LSTM网络的融合模型,利用LSTM网络的时序预测能力及EMD分解处理方法,剥离出了原始序列中隐含的时间序列信息,又结合皮尔逊相关性检验的方法提取出相关度较高的序列,抑制了无效噪音的产生,得到了一种处理非平稳序列数据的新模型。实验结果证明,针对非平稳的序列数据,新模型有更好的准确性。

tive Guide: Real-time Data and Stream Processing at Scale[M]. O'Reilly Media, Inc. 2017.

- [13] KLEPPMANN M, KREPS J. Kafka, Samza and the Unix philosophy of distributed data[J]. Bulletin of the IEEE CS Technical Committee on Data Engineering, 2015.
- [14] WANG G, et al. Building a Replicated Logging System with Apache Kafka[J]. Proceedings of the VLDB Endowment, 2015, 8(12): 1654-1655.
- [15] BYZEK Y. Optimizing Your Apache Kafka Deployment: Levers for The throughput, Latency, Durability, and Availability[J]. Technical report, Confluent Inc, 2017.
- [16] JUNQUEIRA F, REED B. ZooKeeper: Distributed Process Coordination[M]. Sebastopol; O'Reilly Media, Inc. 2013.



GAO Zi-yan, bachelor. Her main research interests include distributed computing and big data.



WANG Yong, born in 1974, Ph.D, associate professor. His main research interests include parallel and distributed computing.

(上接第 298 页)

参 考 文 献

- [1] LPEZ-BENITEZ M, CASADEVALL F. Discrete-Time Spectrum Occupancy Model based on Markov Chain and Duty Cycle Models[C]// 2011 IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN). IEEE, 2011.
- [2] DALTA D, WYGLINSKI A M, M INDEN G J. A spectrum surveying framework for dynamic spectrum access network[J]. IEEE Transactions on Vehicular Technology, 2009, 58(8): 4158-4168.
- [3] HAM ID E, SITHAMPARANATHAN K, BILL M, et al. Spectrum occupancy prediction using a hidden Markov model[C]// Signal Processing and Communication Systems. Cairns, QLD, IEEE, 2015: 1-8.
- [4] WANG L, XIE S G, SU D L, et al. An Autonomous Detection and Robust Estimation Method of Spectrum Anomaly Based on Time Series Analysis[J]. Acta Electronica Sinica, 2014, 42(6): 1055-1060.
- [5] WEI H H, JIA Y F. A Method for Analysis of Non-linear and Non-stationary Spectrum Occupancy Time Series[J]. Acta Electronica Sinica, 2017, 45(8): 2026-2030.
- [6] GERS F A, SCHMIDHUBER, JÜRGEN, et al. Learning to Forget; Continual Prediction with LSTM[J]. NeuralComputation, 2000, 12(10): 2451-2471.
- [7] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [8] NG Y H, HAUSKNECHT M, VIJAYANARASIMHAN S, et al. Beyond short snippets; Deep networks for video classification [C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.
- [9] KOK I, SIMSEK M U, OZDEMIR S. A deep learning model for air quality prediction in smart cities[C]// 2017 IEEE International Conference on Big Data (Big Data). IEEE, 2017.
- [10] ZENG Z, LI L, CHEN J. Deeply Hierarchical Bi-directional LSTM for Sentiment Classification[J]. Computer Science, 2018, 45(8): 213-217, 252.
- [11] HUANG, NORDEN E. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. Proceedings; Mathematical[J]. Physical and Engineering Sciences, 1998(454): 903-995.
- [12] QI Y Y, YU M. Anti-jamming Method for Frequency Hopping Communication Based on Single Channel BSS and EMD[J]. Computer Science, 2016, 43(1): 149-153.
- [13] MOTIN M A, KARMAKAR C, PALANISWAMI M. Selection of Empirical Mode Decomposition Techniques for Extracting Breathing Rate from PPG[J]. IEEE Signal Processing Letters, 2019: 1-1.
- [14] ZHENG D, CUI G, CAO J, et al. Analysis of Brain-Death EEG Data Using 2T-EMD Algorithm[C]// International Conference on Signal-image Technology & Internet-based Systems. IEEE, 2016.
- [15] GONG B M, WANG W B, ZHAO P. EMD-FSVM Prediction for Nonstationary Time Series [J]. Computer Science, 2014, 41(S2): 57-60.
- [16] ELMAN J L. Finding Structure in Time[J]. Cognitive Science, 1990, 14(2): 179-211.
- [17] International Telecommunication Union. Spectrum occupancy measurement and evaluation[OL]. <https://www.itu.int/rec/R-REC-SM.1880/en>.



ZHAO Xiao-dong, born in 1991, master. His main research interest include big data, artificial intelligence, deep learning.



XU Jiang-feng, born in 1965, Ph.D, professor. His main research interest include big data, cryptography, database.