

# 基于噪声统计的 EMD-LSTM 网络 流量预测方法

谷妙春

(南京理工大学 计算机科学与工程学院, 南京 210094)

**摘要:** 为了提高 LSTM 对网络流量的预测精度, 针对训练样本中存在噪声的问题提出了一种基于噪声统计特性的 EMD 降噪方法, 将每一训练样本经 EMD 分解后得到若干 IMF, 通过分析这些 IMF 确定其中的噪声 IMF, 将各样本中同一位置的噪声 IMF 进行统计平均后再与每一样本中的非噪声 IMF 叠加, 以此实现每一样本的降噪, 并将降噪后的样本作为 LSTM 训练样本; 针对 LSTM 中使用滑动窗口输入训练样本时存在的误差叠加问题使用间隔采样的输入方法构造训练样本; 将两种方法结合提出一种 EMD-LSTM 预测模型; 仿真表明, 相较于传统 LSTM 预测模型, 所提模型具备更优的降噪效果与更准确的预测结果; 此外, 所提预测模型应用于一种基于无人机卸载流量的蜂窝网络, 基于该模型预测结果提出一种无人机活动规划方法以优化无人机长时间工作中返航充电的时间点, 规划后的无人机在同等情况下可以使用更小的缓存队列应对突发流量。

**关键词:** 经验模态分解; 长短期记忆; 流量预测; 无人机; 流量卸载

## Method of an EMD-LSTM Network Traffic Prediction Based on Noise Statistics

GU Miaochun

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** In order to improve the prediction accuracy of long short term memory (LSTM) for network traffic, aiming at the problem of noise in training sample, an empirical mode decomposition (EMD) noise reduction method based on the noise statistical characteristics is proposed. Firstly, each training sample is decomposed by the EMD to obtain several intrinsic mode functions (IMFs), and the noise IMFs are determined by analyzing these IMFs. The noise IMFs at the same position in each sample are statistically averaged and then superimposed with the non-noise IMFs in each sample, the noise reduction of each sample is realized by this way, and the denoised sample is used as the LSTM training sample; Secondly, aiming at the error superposition problem, when the training sample is inputted by using sliding window in LSTM, the training sample is constructed by using the input method of interval sampling. Combining with two methods, an EMD-LSTM prediction model is proposed. The simulation shows that compared with the traditional LSTM prediction model, the proposed model has better noise reduction effect and more accurate prediction result. In addition, the prediction model is applied to the cellular network by using UAV to the offload traffic, based on the prediction results of the model, a UAV activity planning method is proposed to optimize the time of the UAV returning to home for charging during long-term work. This planning method can use the smaller buffer queue to deal with burst traffic under the same conditions.

**Keywords:** EMD; LSTM; traffic forecast; UAV; traffic offloading

## 0 引言

5G 等技术的不断发展促进了各类网络服务的大量应用, 这对网络的性能及服务质量有了更高的要求。流量预测<sup>[1]</sup>作为网络管理、调控的重要方式之一, 利用数据流当前和历史的流量状态数据来预测下一时段的流量状态<sup>[2]</sup>。准确的预测结果有助于优化资源配置, 提高生产效率, 实现供需匹配。

近年来, 随着机器学习的发展, 许多学者使用递归神经网络 (RNN, recursive neural networks) 模型预测流量<sup>[3-5]</sup>。相比于传统的自整合移动平均自回归模型<sup>[6-8]</sup>、广义

自回归条件异方差模型<sup>[9-11]</sup>等预测方法, RNN 通过捕获流量序列的高维特征学习流量序列时序关联性, 避免了繁杂的函数拟合且具备更准确的预测效果。但传统 RNN 存在梯度消失或爆炸问题, 不具备长期依赖性<sup>[12]</sup>。长短期记忆 (LSTM, long short term memory) 作为 RNN 的变种克服了上述梯度问题, 可应用于短期与长期的流量预测。目前基于 LSTM 的流量预测方法广泛应用于各类生产活动中<sup>[13-15]</sup>。

大量的动力学研究表明, 现实中的人类活动普遍遵循着一定的规律性<sup>[16]</sup>, 例如对某一固定区域而言, 其流量的

收稿日期: 2022-06-17; 修回日期: 2022-07-20。

作者简介: 谷妙春 (1996-), 男, 山东临沂人, 硕士研究生, 主要从事无人机通信和流量卸载方向的研究。

引用格式: 谷妙春. 基于噪声统计的 EMD-LSTM 网络流量预测方法[J]. 计算机测量与控制, 2023, 31(2): 21-27.

总体趋势往往表现出一定的周期性,基于这种周期性可以知道下一周期内流量的大致趋势,这对流量预测有着积极的意义。但同时具体的流量数值会受个体随机活动影响表现出很大的不确定性,并且这种数值波动具有非平稳非线性的特点,这又为流量预测带来了一定难度。从信号处理的角度分析,可以将实际的流量看成实际信号,总体的流量趋势看成有用信号,流量的波动看成噪声信号,实际信号即受噪声信号影响围绕有用信号波动。在 LSTM 预测中,通过对实际信号降噪,改善有用信号与噪声信号的信噪比,从而有助于 LSTM 正确学习有用信号,使得预测结果最大程度贴合实际信号的变化趋势,从而提高预测精度。因此在 LSTM 预测前需要对若干周期的原始流量进行降噪处理。

常见的信号降噪方法包括傅里叶变换<sup>[17]</sup>、小波变换<sup>[18-20]</sup>、经验模态分解<sup>[21-23]</sup>(EMD, empirical mode decomposition)等。相较于傅里叶变换与小波变换,EMD 具备自适应性,无需选取基函数,可处理平稳与非平稳序列,克服了傅里叶变换与小波变换的局限性。但现有的 EMD 降噪方法多是基于单个信号的降噪:文献 [21-22] 对 EMD 分解后的不同本征模态函数(IMF, intrinsic mode function)使用不同阈值降噪,但这种做法由于抑制了高于阈值的数值,会在一定程度上破坏实际信号的变化趋势,不能使 LSTM 正确学习有用信号;文献 [23] 基于高斯白噪声统计平均为零的特性将一个含有高斯白噪声的流量序列反复叠加不同随机过程的高斯白噪声以实现降噪。但这种做法由于不确定原始序列中高斯白噪声的模值需要反复叠加,导致降噪效率低下,并且将任意噪声视为高斯白噪声只是一种“近似替换”,实际中的噪声并不一定完全具有高斯白噪声的统计特性,从而不一定能有效、正确抑制噪声。在 LSTM 中使用 EMD 降噪训练样本也多沿用上述 EMD 降噪思路:除了使用上述降噪方法外,文献 [24-25] 基于互信息选择部分由 EMD 分解后的 IMF 将实际信号重构进行预测,但这种做法仍存在趋势破坏问题;文献 [26] 使用某一样本与其临近的序列的平均值作为该样本的降噪结果,但如果选取的临近序列少会影响降噪效果,如果选取的临近序列多同样会破坏实际信号的变化趋势。此外,上述 LSTM 预测方法多着眼于通过降噪训练样本提升预测准确性,但对训练样本的构造缺乏讨论。考虑到序列的时间关联性,上述的 LSTM 预测方法均采取滑动窗口连续的批量输入样本,这会导致长期预测中的误差累积问题,因而只对中短期的预测结果较好。

事实上,每个 LSTM 训练样本中已经包含了噪声的一次随机过程,可将这些噪声信号从训练样本中提取出来,通过分析这些噪声信号确定噪声的统计特性。因此,本文在降噪方法上提出一种基于噪声统计的 EMD 降噪方法:在给定若干流量序列的情况下,将每一流量序列进行 EMD 分解,基于分解后各 IMF 的能量筛选出含噪较高的噪声 IMF,将这些 IMF 进行统计平均得到对应的降噪 IMF,使用降噪

IMF 替换噪声 IMF 再与非噪声 IMF 重构流量序列实现流量序列的降噪,并将降噪后的流量序列作为 LSTM 训练样本。这样既保留了原始流量序列的变化趋势,又充分利用各个样本中包含的噪声,既无需事前知道噪声的统计特性又具有较高的降噪效率。同时在训练样本构造上使用间隔采样选取临近数据与历史同期数据,减少连续滑动窗口带来的累积误差。由此构成的 LSTM 预测模型称为 EMD-LSTM 预测模型。此外,目前对无人机(UAV, unmanned aerial vehicle)的应用研究多是优化 UAV 的活动状态或轨迹实现收益最大化<sup>[25]</sup>,但对 UAV 何时返航充电的问题鲜有讨论,文献 [27-29] 均默认 UAV 电量耗尽时返航充电。本文将该预测模型应用于一种基于 UAV 卸载流量的蜂窝网络,通过该预测模型分析 UAV 在长时间工作下何时返航充电可最小化服务中断时间,提出一种基于 EMD-LSTM 的 UAV 活动规划方法以完善当前对 UAV 完整活动流程缺乏讨论的问题,同时具体说明流量预测如何有助于优化资源配置。

综上,本文工作及创新点如下:

1) 提出一种基于噪声统计的 EMD 降噪方法,该方法相比于传统 EMD 降噪方法无需获取噪声的先验知识,避免因超参设置不当影响降噪效果,同时充分利用样本中噪声的随机过程,具备较高的降噪效率。

2) 使用基于间隔采样的 LSTM 训练样本构造方法,减少预测中的误差累积问题,提升中长期预测的准确性。

3) 将该模型应用于一种基于 UAV 卸载流量的蜂窝网络,提出一种基于 EMD-LSTM 的 UAV 活动规划方法,以最小化服务中断时间为目标优化 UAV 在长时间工作下返航的时间点。

## 1 EMD 分解

EMD 通过将一个序列分解为若干不同尺度的平稳波动项和一个残差项的叠加,其本质是将非平稳序列平稳化过程。EMD 分解中每一个平稳波动项称之为一个 IMF。对于每一个 IMF,需满足以下两个条件:

1) IMF 的极值点和过零点的数目相等或只相差一个。

2) IMF 的上包络线和下包络线关于时间轴对称,即两条包络线在每一个时间点处的均值为 0。

假设蜂窝网络在周期  $T$  内产生的流量序列为  $D(t)$ ,  $D(t)$  表示如下:

$$D(t) = \{D(1), D(2), \dots, D(T)\} \quad (1)$$

将  $D(t)$  按如下方式进行 EMD 分解:

1) 确定  $T$  内  $D(t)$  的全部局部极大值和极小值,通过三次样条插值法求出  $D(t)$  的上包络线  $U(t)$  和下包络线  $L(t)$ ;

2) 求出上下包络线的中位线  $M(t)$ ,  $M(t)$  的表达式如下:

$$M(t) = \frac{U(t) + L(t)}{2} \quad (2)$$

将  $D(t)$  减去  $M(t)$  得到准 IMF 分量  $\phi_1(t)$ ,  $\phi_1(t)$  的表达式如下:

$$\phi_1(t) = D(t) - M(t) \quad (3)$$

3) 判断  $\varphi_1(t)$  是否满足 IMF 条件, 如果是, 则  $\varphi(t)$  为第一个 IMF 分量, 记为  $imf_1$ , 如果不是, 则对  $\varphi_1(t)$  依照  $D(t)$  做相同计算得出  $\varphi_2(t)$ , 判断  $\varphi_2(t)$  是否满足 IMF 条件或停止条件  $SD$ , 如果是则  $\varphi_2(t)$  为  $imf_1$ , 否则重复上述步骤计算  $\varphi_k(t)$ , 直到  $\varphi_k(t)$  满足 IMF 条件或停止条件  $SD$ , 其中停止条件  $SD$  如下:

$$SD = \sum_{t=0}^T \left[ \frac{(\varphi_{k-1}(t) - \varphi_k(t))^2}{\varphi_{k-1}^2(t)} \right] \quad (4)$$

当  $SD$  小于预设的典型值时即可认为  $\varphi_k(t)$  满足 IMF 要求, 一般典型值在区间  $[0.2, 0.3]$  设置。

4) 计算  $R_1 = D(t) - imf_1$ , 将  $R_1$  作为新的待分解序列, 重复上述步骤以此计算  $imf_2, imf_3, \dots, imf_n$ , 直到  $R_n$  为单调序列或常值序列。此时可以将序列  $D(t)$  表示为:

$$D(t) = \sum_{i=1}^n imf_i + R_n \quad (5)$$

至此  $D(t)$  的 EMD 分解完成, 需要指出的是先分解出的 IMF 分量为原始序列高频分量, 也是高噪部分, 后分解出的 IMF 分量为原始序列低频分量, 伴随着较低的噪声,  $R_n$  为残余项, 反映出  $D(t)$  的总体变化趋势。

## 2 基于噪声统计的 EMD 降噪

EMD 降噪的原理是通过计算各 IMF 分量的能量来划分低频分量与高频分量, 对含噪较高的高频分量进行一定的处理以抑制其能量, 从而实现降噪。对任意 IMF, 其能量的计算方式如下:

$$E(imf_k) = \sqrt{\sum_{t=1}^T imf_k^2(t)} \quad (6)$$

式中,  $imf_k(t)$  表示  $t$  时刻  $imf_k$  的值。显然,  $k$  越大, 即 IMF 分量频率越低,  $E(imf_k)$  越大。IMF 能量从高频到低频存在突变, 找到突变位置即可区分高频 IMF 与低频 IMF。突变点  $K_0$  的计算方法如下:

$$\Delta E(K) = \left| \frac{K(n-K)}{n^2} \times \left( \frac{\sum_{k=1}^K E(imf_k)}{K} - \frac{\sum_{k=K+1}^n E(imf_k)}{n-K} \right) \right| \quad (7)$$

$$K_0 = \max_K (\Delta E(K)) \quad (8)$$

找出突变点  $K_0$  后, 对 IMF 高频部分, 即  $imf_1, imf_2, \dots, imf_{K_0}$  进行降噪, 记降噪后的结果为  $\varphi(imf_1), \varphi(imf_2), \dots, \varphi(imf_{K_0})$ 。将降噪后的高频部分与原始序列的低频部分及残余项进行重构, 即可实现对原始序列  $D(t)$  的降噪。降噪结果  $\varphi(D(t))$  表示如下:

$$\varphi(D(t)) = \sum_{k=1}^{K_0} \varphi(imf_k) + \sum_{k=K_0+1}^n imf_k + R_n \quad (9)$$

$\varphi(imf_k)$  的计算依赖于其它原始序列经 EMD 分解后的 IMF。假设从该蜂窝网络采集  $M$  个  $D(t)$  序列, 记第  $m$  个  $D(t)$  序列为  $D_m(t)$ , 分解后的结果为  $imf_{k,m}$ , 则  $\varphi(imf_k)$  与  $\varphi(D_m(t))$  的计算过程如下:

1) 对  $M$  个  $D(t)$  序列依次进行 EMD 分解, 得到  $D_m(t)$  分解结果为  $imf_{1,m}, imf_{2,m}, \dots, imf_{n,m}$ 。

2) 依次找出  $D_m(t)$  的突变点  $K_0^m$ , 进而可得样本总体的突变点  $K_0^*$  为:

$$K_0^* = \text{round} \left( \frac{1}{M} \sum_{m=1}^M K_0^m \right) \quad (10)$$

式中,  $\text{round}(K)$  表示对  $K$  就近取整。此时任意样本  $D_m(t)$  可以表示为:

$$D_m(t) = \sum_{k=1}^{K_0^*} imf_{k,m} + \sum_{k=K_0^*+1}^n imf_{k,m} + R_{n,m} \quad (11)$$

3) 对每一个  $D_m(t)$  中的高频分量  $imf_{1,m}, imf_{2,m}, \dots, imf_{K_0^*,m}$ , 依次计算与其余  $M-1$  个  $D_m(t)$  序列的对应分量的平均, 可得降噪后的高频分量:

$$\varphi(imf_{k,m}) = \frac{1}{M} \sum_{j=1}^M imf_{k,j} \quad (12)$$

4) 将降噪后的高频分量与  $D_m(t)$  中低频分量和残余项重构, 可得  $D_m(t)$  的降噪结果为:

$$\varphi(D_m(t)) = \sum_{k=1}^{K_0^*} \varphi(imf_{k,m}) + \sum_{k=K_0^*+1}^n imf_{k,m} + R_{n,m} \quad (13)$$

由此实现了对  $m$  个  $D(t)$  序列的降噪。

## 3 基于间隔采样的训练样本构造

由于时间序列的连续性, 常见的训练样本构造方法是采用滑动窗口输入序列作为 LSTM 的输入, 但这种方法在预测长期数据时存在误差累计问题。假设对于序列  $\{D_1(t), D_2(t), \dots, D_M(t)\}$ , 窗口大小为  $\theta$ , 在训练 LSTM 预测  $D_m(t)$  时输入的训练样本为  $\{D_{m-\theta}(t), D_{m-\theta+1}(t), \dots, D_{m-1}(t)\}$ 。这种输入方法会导致 LSTM 在预测长期结果时会在短期预测结果的基础上预测。例如预测  $D_{M+2}(t)$ , LSTM 需要的输入为  $\{D_{M-\theta+2}(t), \dots, D_M(t), D_{M+1}(t)\}$ , 而  $D_{M+1}(t)$  是 LSTM 基于序列  $\{D_{M-\theta+1}(t), \dots, D_{M-1}(t), D_M(t)\}$  的预测结果, 这会降低 LSTM 对  $D_{M+2}(t)$  预测的准确性。

鉴于流量序列具有周期性的特性, 可以采用间隔采样的方式基于若干历史同期数据构造训练样本。假设间隔尺度为  $\epsilon$ , 在训练 LSTM 预测  $D_m(t)$  时可构造训练样本  $\{D_{m-\epsilon}(t), D_{m-\epsilon+\epsilon}(t), \dots, D_{m-\epsilon}(t)\}$ 。这样在预测未来多个周期的流量序列  $\{D_{m_1}(t), D_{m_2}(t), \dots, D_{m_n}(t)\}$  时, 使用间隔采样法在预测  $D_m(t)$  ( $1 \leq i \leq n$ ) 时对应的输入序列为:

$$\{D_{m_1-\epsilon}(t), D_{m_1-(\theta-1)\epsilon}(t), \dots, D_{m_1-\epsilon}(t)\} \quad (14)$$

而使用滑动窗口法预测  $D_m(t)$  时对应的输入序列为:

$$\{D_{m_1-\theta}(t), D_{m_1-(\theta-1)(t)}, \dots, D_{m_1-1}(t)\} \quad (15)$$

与滑动窗口相比, 这样的间隔采样方法在输入序列长度不变的情况下尽可能减少使用 LSTM 预测结果再预测, 从而减小累计误差, 提升预测准确性。

## 4 基于 EMD-LSTM 的 UAV 活动规划

基于上述讨论, 本文提出一种基于噪声统计特性的 EMD 降噪方法与 LSTM 结合的流量预测模型: EMD-LSTM 模型。首先采集连续  $M$  个周期的流量序列  $D(t)$ ; 其次对每一序列  $D_m(t)$  进行基于噪声统计特性的 EMD 降噪, 得到降噪后的训练样本  $\varphi(D_m(t))$ ; 接着构造大小  $\theta$ , 间隔尺度

为  $\varepsilon$  的序列构造 LSTM 中预测  $\varphi(D_m(t))$  的训练样本；最后将训练好的 LSTM 用于预测未来周期的流量数据，其输入仍为以同样方式构造的间隔采样序列。

假设某蜂窝网络由 1 个宏基站、 $X$  个潜在热区和  $X$  架 UAV 组成，每架 UAV 为一个潜在热区卸载流量。考虑到所有潜在热区在卸载流程上是等效的，这里以其中一个潜在热区为例。简化后的网络结构如图 1 所示。

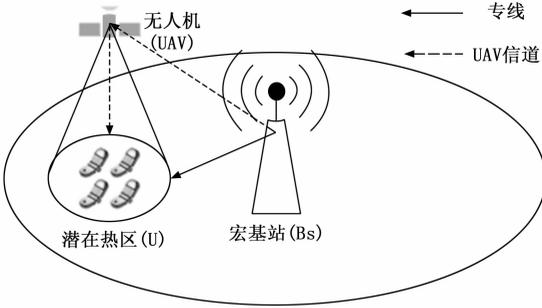


图 1 蜂窝网络结构

宏基站 BS 中存在一个缓存队列  $Q$  以缓存潜在热区  $U$  请求的数据，假设  $t$  时刻  $U$  请求的数据量为  $D(t)$ ，记  $Q(t)$  表示  $t$  时刻  $Q$  中的数据量。BS 可使用一专线以固定速率  $Q_0$  向  $U$  发送数据，同时当  $Q(t) + D(t) - Q_0 > 0$  时可使用 UAV 以固定速率  $Q_1$  发送数据，记  $\Delta Q(t) = Q(t) + D(t) - Q_0$ ，则下一时刻  $Q(t+1)$  的表达式如下：

$$Q(t+1) = \begin{cases} 0 & \Delta Q(t) \leq Q_1 \\ \Delta Q(t) - Q_1 & \Delta Q(t) > Q_1 \end{cases} \quad (16)$$

由于 UAV 续航有限，为了尽可能延长 UAV 的使用时间，当 UAV 在  $t$  时刻发送数据时，UAV 应处于激活状态；而当 UAV 在  $t$  时刻未发送数据时，UAV 应处于休眠状态。记  $a(t)$  表示  $t$  时刻 UAV 的活动状态， $a(t) = 1$  表示 UAV 处于激活状态； $a(t) = 0$  表示 UAV 处于休眠状态。当  $\Delta Q(t) \leq 0$  时 UAV 处于休眠状态；当  $\Delta Q(t) > 0$  时 UAV 处于激活状态。基于 BS 对  $U$  若干历史周期流量序列的采样，使用本文所提 EMD-LSTM 模型预测  $U$  未来一段时间的流量数据，根据  $\Delta Q(t)$  的数值可以知道对应  $a(t)$  的取值，即当前时刻 UAV 是否应活动，进而结合式 (16) 可知所预测时间段内 UAV 的活动序列与队列长度的序列。

记 UAV 的电池容量为  $C$ ，UAV 处于激活状态伴随着  $c_1$  的电量消耗，而当 UAV 处于非激活状态时由于其本身的悬停、移动也伴随着  $c_2$  的电量消耗。当 UAV 电量即将耗尽时，UAV 需返回充电处充电。假设  $U$  距 UAV 充电处的距离为  $l$ ，UAV 的飞行速度为  $v$ ，充电时间为  $t_0$ 。易知 UAV 往返充电处与  $U$  需要的最小时间  $t_c$  为：

$$t_c = \frac{2l}{v} + t_0 \quad (17)$$

于是 UAV 的活动规划方法如下：

- 步骤 1：UAV 更换好电池从充电处飞往  $U$  处待命。
- 步骤 2：当 UAV 飞至  $U$  处时，记该时刻为零时刻。基

于 LSTM 对  $D(t)$  的预测结果可知在未来  $\tau$  时刻内 UAV 的活动状态序列为  $\{a(0), a(1), \dots, a(\tau)\}$ 。考虑到 UAV 需预留返回充电处的电量，则 UAV 可为  $U$  提供服务的最大时间  $\tau_1$  满足：

$$\sum_{t=0}^{\tau_1} 1_{a(t)=1} c_1 + 1_{a(t)=0} c_2 = C - t_c c_2 \quad (18)$$

式中， $1_x$  表示当条件  $x$  为真时， $1_x = 1$ ，否则  $1_x = 0$ 。

记该状态序列中 UAV 首次激活的时间点为  $\tau_0$ 。如果在  $[\tau_0, \tau_1 + t_c]$  时间段内存在某个时间点  $\tau_2$  满足：

$$\sum_{t=\tau_2}^{\tau_2+t_c} a(t) = 0 \quad (19)$$

式中， $\tau_2 \in [\tau_0, \tau_1]$ 。

式 (19) 表示 UAV 在  $\tau_2$  时刻后的接下来  $t_c$  时间段内均处于休眠状态，UAV 可在此时间段内返回充电而不造成服务中断，由此构成的满足条件 (19) 的  $\tau_2$  的集合为  $R_2$ ；如果  $R_2 = \phi$ ，为了尽可能减少服务中断，UAV 应在  $[\tau_0, \tau_1 + t_c]$  时间段中选择时间点  $\tau_3$ ，这样的  $\tau_3$  满足其后的  $t_c$  时间段内处于休眠的时间大于等于  $[\tau_0, \tau_1 + t_c]$  时间段中任意  $t_c$  时间段内处于休眠的时间，即解方程：

$$\min_{\tau_3} \sum_{t=\tau_3}^{\tau_3+t_c} a(t) \quad (20)$$

式中， $\tau_3 \in [\tau_0, \tau_1]$ ，UAV 可在  $\tau_3$  时刻返回充电。记方程 (20) 的解  $\tau_3$  构成的集合为  $R_3$ 。

步骤 3：为了避免 UAV 频繁往返充电，如果  $R_2 \neq \phi$ ，取  $R_2$  中的最大值作为 UAV 返回充电的时间点；如果  $R_2 = \phi$ ，取  $R_3$  中的最大值作为 UAV 返回充电的时间点。记返回充电的时间点为  $\tau_4$ 。

步骤 4：对于 UAV 返回充电前的任意时刻  $t \in [0, \tau_4]$ ，当  $a(t) = 0$  时，UAV 处于休眠状态；当  $a(t) = 1$  时，UAV 处于激活状态并提供传输服务。

步骤 5：当  $t = \tau_4$  时，UAV 返回充电。

## 5 仿真实验

为了能真实反映现实生活中蜂窝流量的特性，仿真选取的数据集为 WIDE 骨干网中采样点 F 于 2021 年 5 月 1 日至 2021 年 6 月 30 日采集的总流量数据<sup>[30]</sup>。该数据由 MAWI Working Group 与 WIDE Project 提供。该采样点记录了其上游网络提供商每日的流量数据，采样间隔为 10 分钟。

基于数据集的实际意义，仿真中  $T = 144$  代表每日流量数据； $\varepsilon = 1\ 008$  代表历史每周同期流量数据； $\theta = 1\ 008$  代表一周流量数据。其余参数为： $Q_0 = 300; Q_1 = 150; l = 3\ 000; v = 3\ 000; C = 120; t_0 = 24; c_1 = 2; c_2 = 1$ 。LSTM 由一个输入层、一个 LSTM 层、一个全连接层、一个回归层构成。LSTM 层有 200 个隐藏单元，解释器为 Adam，初始学习率为 0.005，每迭代 125 次乘以因子 0.2 降低学习率。61 个样本中前 56 个样本用于训练 LSTM 网络，后 5 个样本用于测试，迭代次数为 500。

图 2 和图 3 分别给出了不同工作日和休息日下每日采集

的流量数据。

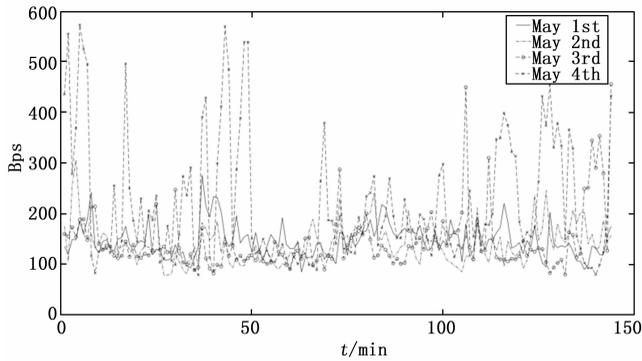


图 2 不同休息日的流量数据

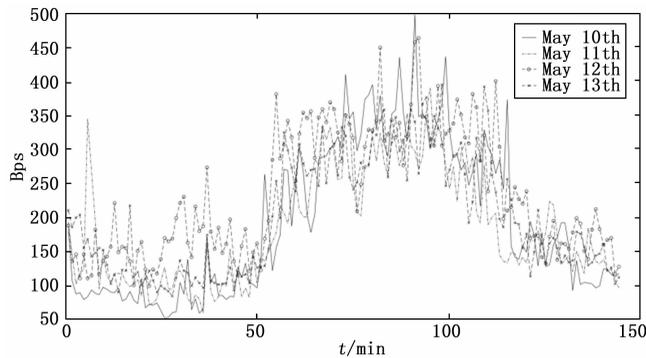


图 3 不同工作日的流量数据

该数据集中每日的流量变化趋势受人类休息日与工作日影响大致为两种。根据图 1 可知，在休息日中，流量不存在明显的变化趋势；在工作日中，流量存在一个大致在每日早 9 点至晚 9 点的流量高峰。同时，不同工作日与休息日下的流量大小与变化趋势虽大致相同但存在明显的波动。此外，除节假日因素外，休息日下的流量趋势只可能出现在每周的周六或周日。实例表明流量具有周期性和高噪声的特点。

图 4 以 5 月 1 日的流量数据为例给出了文献 [19]、文献 [21] 与本文所提 EMD 降噪方法的降噪效果：

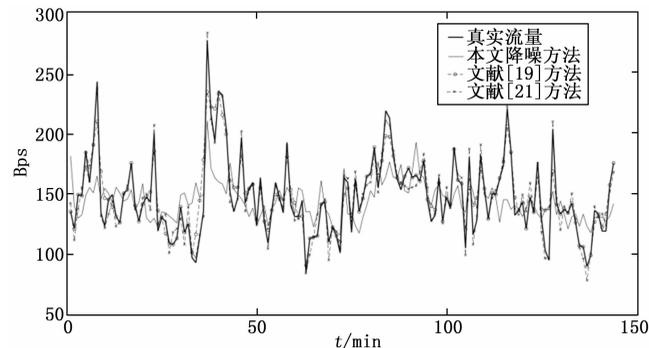


图 4 不同 EMD 降噪方法的降噪效果

不难发现，只有本文所提 EMD 降噪方法获得了比较理想的降噪效果，这是因为文献 [19] 采用的阈值法只会对

超过阈值的部分降噪，因而降噪效果有限；文献 [21] 采用的基于高斯白噪声统计特性的方法在面对非高斯白噪声或未知统计特性的噪声时，并未基于实际噪声的统计特性降噪，因而降噪效果不理想。只有本文所提降噪方法无需获取噪声先验知识，通过提取训练样本中的噪声，分析其统计特性后降噪，获得了相对理想的降噪效果。

图 5 展示了以不同降噪方法的结果作为 LSTM 输入并使用滑动窗口法构造训练样本时 LSTM 对 6 月 25 日至 6 月 30 日流量的预测结果。

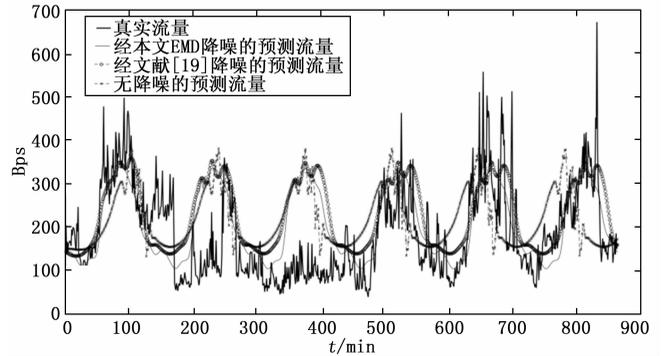


图 5 滑动窗口下不同降噪方法预测结果

根据图 5 可知，使用本文 EMD 降噪方法后的 LSTM 预测结果相较于使用传统 EMD 降噪与未降噪的预测结果更接近真实流量。但三者均在预测 6 月 27 日的流量时均认为流量趋势为工作日而非休息日，这是因为休息日相较于工作日属于偶发的短期事件，基于滑动窗口输入训练样本时休息日的流量数据将作为短期事件被 LSTM 网络“忘记”，因而这种样本构造方式使得 LSTM 对未来每一天流量的趋势预测均为工作日的趋势。

而使用间隔采样方法构造训练样本后 LSTM 的预测结果如图 6 所示。

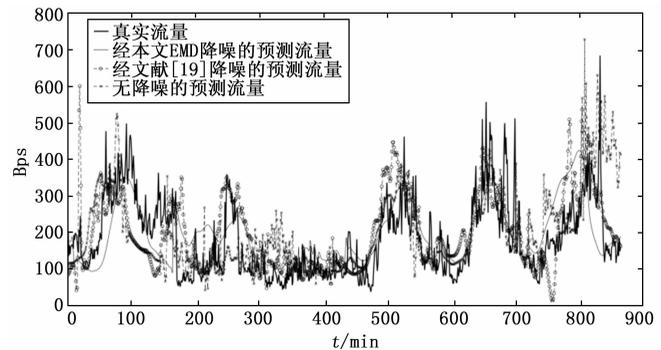


图 6 间隔采样下不同降噪方法预测结果

根据图 6 可知，使用本文 EMD 降噪方法后的 LSTM 预测结果依然优于传统 EMD 降噪以及无降噪的预测结果，并且 LSTM 能正确预测 6 月 27 日的流量趋势为休息日，这是因为对于每周日而言休息日属于高频的长期事件，即使其中混入部分工作日 LSTM 网络也会将这些休息日下的工作日视为短期事件而“忘记”，因而能正确预测未来流量趋

势。因此本文提出的 EMD-LSTM 预测模型相比于传统 EMD-LSTM 预测模型具备更好的预测效果。

为了量化不同方法的预测结果,本文选用平均绝对百分比误差 (MAPE, mean absolute percentage error)、均方根误差 (RMSE, root mean square error) 评估不同方法下的预测结果与真实值之间的差异程度,结果如表 1 所示。

表 1 不同方法的 LSTM 预测效果

构造方法	降噪方法	MAPE	RMSE
滑动窗口	无	0.742 6	113.40
滑动窗口	阈值法	0.697 0	103.63
滑动窗口	本文	0.559 8	94.07
间隔采样	无	0.486 1	103.65
间隔采样	阈值法	0.478 0	93.60
间隔采样	本文	0.462 8	85.84

综上,本文所提 EMD-LSTM 预测模型相比传统 EMD-LSTM 模型具备更优预测结果。

最后,图 7 对比了 6 月 25 日至 6 月 30 日使用本文提出的 UAV 活动规划方法与使用目前 UAV 电量耗尽即返航充电的传统活动方法时  $Q$  的变化以及在使用本文提出 UAV 活动规划方法而使用不同 EMD-LSTM 预测模型时  $Q$  的变化。

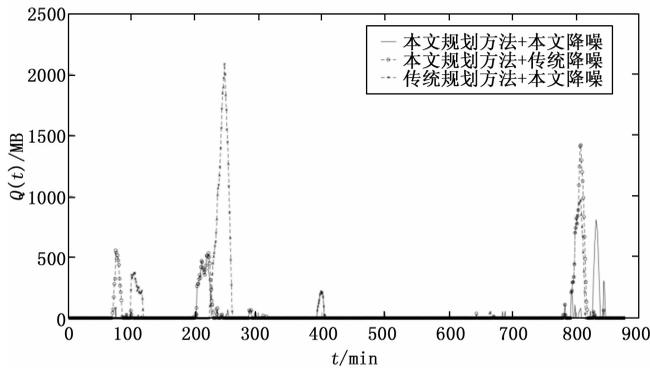


图 7 不同 UAV 活动方法对  $Q$  的影响

根据图 7 可知,传统的 UAV 活动规划方法并未考虑 UAV 的后续使用问题,因而可能会在某些热区存在时段因电量不足而返航充电,从而导致突发数据不能卸载,使得缓存队列中的数据量激增。即主要对应图 7 中 [235, 265] 时间段;而本文提出的 UAV 活动规划方法则考虑了 UAV 后续的使用问题,选择最大续航时间下可工作时段中中断时间最短的时间点返航充电以尽可能减少突发数据因 UAV 返航而无法卸载的问题,因而相较于传统 UAV 活动规划方法缓存队列不存在激增的情况。虽然因为预测精度问题导致缓存队列的使用频率小幅上升,但总体而言使用本文所提 UAV 活动规划方法不仅大幅减小了缓存队列的平均长度,还可以在同等条件下使用更小的缓存队列应对突发流量。而在同样使用本文所提 UAV 活动规划方法的前提下,使用传统的 EMD-LSTM 预测模型因为不能准确预测未

来的流量趋势,因而不能准确控制 UAV 激活的时间点,从而会在某些需要 UAV 卸载流量的时间段因未激活 UAV 导致缓存队列中存在较多的数据。分析数据可知,使用本文所提 UAV 活动规划方法并结合本文所提预测模型所需缓冲区最小大小为 806,平均队列长度为 22.11;使用本文所提 UAV 活动规划方法但使用传统预测模型所需缓冲区最小大小为 1 416,平均队列长度为 38.92;而使用传统 UAV 活动规划方法所需缓冲区最小大小为 2 079,平均队列长度为 65.70。因此本文提出的 UAV 活动规划方法既降低了平均缓存队列长度,又节省了缓冲区空间,结合本文所提 EMD-LSTM 预测模型,进一步强化了效果。

## 6 结束语

本文基于实际蜂窝流量具有周期性的特点,结合若干历史周期流量数据使用 LSTM 预测未来周期流量。为了提高预测精度,提出一种基于噪声统计的 EMD 降噪方法对训练样本降噪;提出使用间隔采样的方法构造训练样本。此外,本文结合实例说明流量预测如何指导生产生活,通过本文提出的 EMD-LSTM 模型规划 UAV 未来一段时间的活动状态,从而选择合适的时间段返航充电,减少 UAV 的服务中断,提升了卸载性能。

### 参考文献:

- [1] WANG Y, LIU Y, GAN Y. Research on combination network traffic forecasting model [C] //2018 IEEE International Conference on Automation, Electronics and Electrical Engineering (AUTEEE). Shenyang, IEEE, 2018: 311 - 314.
- [2] 杨俊. 面向智能工厂的异构网络跨网融合与流量调度研究 [D]. 广州: 华南理工大学, 2020.
- [3] TOKGÖZ A, ÜNAL G. A RNN based time series approach for forecasting Turkish electricity load [C] //2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, IEEE, 2018: 1 - 4.
- [4] BUI V, NGUYEN V H, PHAM T L, et al. RNN-based deep learning for one-hour ahead load forecasting [C] //2020 International Conference on Artificial Intelligence in Information and Communication (ICAIC), Fukuoka, IEEE, 2020: 587 - 589.
- [5] YAHYA M A, HADI S P, PUTRANTO L M. Short-term electric load forecasting using recurrent neural network (study case of load forecasting in central java and special region of Yogyakarta) [C] //2018 4th International Conference on Science and Technology (ICST), Yogyakarta, IEEE, 2018: 1 - 6.
- [6] ABU AMRA I A S, MAGHARI A Y A. Forecasting groundwater production and rain amounts using ARIMA-Hybrid ARIMA; case study of Deir El-Balah city in GAZA [C] //2018 International Conference on Promising Electronic Technologies (ICPET), Deir El-Balah, IEEE, 2018: 135 - 140.
- [7] PUTRA J A, BASBETH F, BUKHORI S. Sugar production forecasting system in PTPN XI Semboro Jember using autore-

- gressive integrated moving average (ARIMA) method [C] // 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), Bandung, IEEE, 2019: 448 - 453.
- [8] XING S, LOU Y. Hydrological time series forecast by ARIMA + PSO-RBF combined model based on wavelet transform [C] // 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, IEEE, 2019: 1711 - 1715.
- [9] TI A, DU Z, ZHANG W. Analysis on the volatility of sustainable stock index and traditional stock index based on GARCH mode [C] // 2019 International Conference on Economic Management and Model Engineering (ICEMME), Malacca, IEEE, 2019: 47 - 50.
- [10] XU Q, et al. Ultra-short-term wind speed forecast based on WD-ARIMAX-GARCH model [C] // 2019 IEEE 2nd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), Shenyang, IEEE, 2019: 219 - 222.
- [11] WANG L, ZHANG Z. Research on shanghai copper futures price forecast based on X12-ARIMA-GARCH family models [C] // 2020 International Conference on Computer Information and Big Data Applications (CIBDA), Guiyang, IEEE, 2020: 304 - 308.
- [12] KUMAR S D, SUBHA D. Prediction of depression from eeg signal using long short term memory (LSTM) [C] // 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, IEEE, 2019: 1248 - 1253.
- [13] UD DIN A Z, AYAZ Y, HASAN M, et al. Bivariate short-term electric power forecasting using LSTM network [C] // 2019 International Conference on Robotics and Automation in Industry (ICRAI), Rawalpindi, IEEE, 2019: 1 - 8.
- [14] PALKAR A, DESHPANDE M, KALEKAR S, et al. Demand forecasting in retail industry for liquor consumption using LSTM [C] // 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, IEEE, 2020: 521 - 525.
- [15] 康守强, 周 月, 王玉静, 等, V. I. MIKULOVICH. 基于改进 SAE 和双向 LSTM 的滚动轴承 RUL 预测方法 [J/OL]. 自动化学报: 1 - 11. [2022-09-14]. DOI: 10.16383/j. aas. c190796.
- [16] XU Z P, WANG Y, WU N Q, et al. Propagation dynamics of a periodic epidemic model on weighted interconnected networks [J]. IEEE Transactions on Network Science and Engineering, 2020, 7 (3): 1545 - 1556.
- [17] BUROVA A. Reducing the error of digital algorithms for deductive signal processing based on their multi-stage discrete fourier transform by the difference digital filters [C] // 2020 22th International Conference on Digital Signal Processing and its Applications (DSPA), Moscow, IEEE, 2020: 1 - 3.
- [18] ALI H H S M, SHARIF S M. Comparison between discrete wavelet transform and maximal overlap discrete wavelet transform as an analysis tool for H. 264/AVC video [C] // 2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCEEE), Khartoum, IEEE, 2018: 1 - 5.
- [19] 石志炜, 张丽萍. 基于改进小波包阈值降噪的滚动轴承故障分析 [J]. 计算机测量与控制, 2019, 27 (5): 58 - 63.
- [20] 高建丰, 周韶彤, 何笑冬. 提升小波改进阈值算法在输气管道泄漏信号降噪处理中的应用 [J]. 计算机测量与控制, 2019, 27 (7): 223 - 226.
- [21] REN G F, LIU Z L. An improved EMD adaptive denoising and feature extraction algorithm [C] // 2019 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Dalian, IEEE, 2019: 1 - 4.
- [22] 周 璇, 鲍长春, 夏丙寅. 融合统计模型与 EMD 的宽带语音增强方法 [J]. 通信学报, 2013, 34 (8): 95 - 101.
- [23] 罗 磊, 黄博妍, 孙金玮, 等. 基于总体平均经验模态分解的主动噪声控制系统研究 [J]. 自动化学报, 2016, 42 (9): 1432 - 1439.
- [24] 李孟威, 史元浩, 杨彦茹, 等. 融合 EMD 和 LSTM 的受热面积灰预测研究 [J]. 电子测量与仪器学报, 2020, 34 (11): 166 - 172.
- [25] 金秀章, 刘 岳, 于 静, 等. 基于变量选择和 EMD-LSTM 网络的出口 SO<sub>2</sub> 浓度预测 [J]. 中国电机工程学报, 2021, 41 (24): 8475 - 8484.
- [26] LIAN H, et al. A short-term LOAD forecasting method based on EEMD-LN-GRU [C] // 2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES). Chengdu, IEEE, 2021: 666 - 673.
- [27] ASHERALIEVA A, NIYATO D. game theory and lyapunov optimization for cloud-based content delivery networks with device-to-device and UAV-enabled caching [J]. IEEE Transactions on Vehicular Technology, 2019, 68 (10): 10094 - 10110.
- [28] SANJAB A, SAAD W, BAŞAR T. A game of drones: cyber-physical security of time-critical UAV applications with cumulative prospect theory perceptions and valuations [J]. IEEE Transactions on Communications, 2020, 68 (11): 6990 - 7006.
- [29] THAKOOR O, GARG J, NAGI R. Multiagent UAV routing: a game theory analysis with tight price of anarchy bounds [J]. IEEE Transactions on Automation Science and Engineering, 2020, 17 (1): 100 - 116.
- [30] MAWI Working Group. MAWI working group traffic archive [DB/OL]. MAWI Working Group, <http://mawi.wide.ad.jp/mawi>.