

# 微调长短期记忆网络群：一种小样本多维时序预测算法

鲍观明

(南京信息工程大学 计算机学院、网络空间安全学院, 江苏 南京 210044)

**摘要：** 多维时序预测问题是单变量时序预测问题的推广，其在现实生活中具有广泛的应用价值，例如交通运输商的运力分配，连锁商店的库存管理，生产线故障预测，以及疾病预测等领域中。本文对多维时序预测问题展开了细致透彻的研究，从薛教授所授的《机器学习》课程中“聚类”章节得到启发，提出了一种能有效优化 LSTM 在小样本多维短期时序预测的算法 FTG-LSTM，并通过实验验证了算法的有效性。首先，本文给出了一般多维时序预测问题的数学表达，给出了小样本多维时序预测问题的定义，并分析讨论经典机器学习算法在目标问题上的欠缺之处，通过实验证实经典机器学习算法的不足。接着，介绍了 LSTM 网络的基本概念，在小数据集上频繁训练 LSTM，并与经典机器学习算法进行比较，发现 LSTM 在小样本多维时序数据上仍能优于经典机器学习算法，随后分析了 LSTM 模型在小样本短期数据上难以表现优异的关键。最后，根据所分析的 LSTM 模型的缺点，提出了一个微调长短期记忆网络群算法 FTG-LSTM，并通过大量的对比实验发现在小样本短期时序预测任务上，本文提出的 FTG-LSTM 优于 LSTM 和其他经典机器学习算法。

**关键词：** 机器学习；小样本多维短期时序预测；长短期记忆网络

## Fine-tuning LSTM Group for Small Sample Multivariate Time Series Prediction

Bao Guanming

(School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China)

**Abstract:** Multivariate time series prediction is an extension of the univariate time series prediction problem and has extensive practical applications in various fields such as transportation capacity allocation, inventory management for chain stores, production line failure prediction, and disease prediction. This paper presents a detailed study on multivariate time series prediction problem and proposes an algorithm called FTG-LSTM that can effectively optimize LSTM for short-term prediction with small sample size. Inspired by the "clustering" chapter of Professor Xue's "Machine Learning" course, this paper first gives the mathematical expression of the general multivariate time series prediction problem, defines the small sample multivariate time series prediction problem, and analyzes the shortcomings of classical machine learning algorithms on the target problem through experiments. Then, the basic concept of LSTM network is introduced, and LSTM is compared with classical machine learning algorithms on a small dataset. It is found that LSTM still outperforms classical machine learning algorithms on small sample multivariate time series data. The key reason why LSTM model performs poorly on short-term data with small sample size is also analyzed. Finally, based on the analysis of the shortcomings of the LSTM model, a fine-tuned long-short-term memory network group algorithm FTG-LSTM is proposed. Through a large number of comparative experiments, it is found that FTG-LSTM is better than LSTM and other classical machine learning algorithms in small sample short-term time series prediction tasks.

**Key words:** Machine Learning; Small Sample Multivariate Short-term Time Series Prediction; Long Short-Term Memory Network

## 1 引言

### 1.1 时间序列预测问题概述

时间序列预测是许多领域中的一个重要问题,包括工业和商业等应用<sup>[1]</sup>。例如,公共交通运营商可以根据人流密度的预测<sup>[2]</sup>,在特定地区合理部署运力以减小人群排队时间,从而提高运营收益。另一个例子是大型零售超市,如果能够准确预测不同商品在不同分区的销售额,就可以避免因供应不合理而导致的收益损失。然而,现实世界中的时间序列通常表现出复杂的动态特性和不稳定性,这给时间序列预测任务带来了挑战。

现实世界的时间序列往往受到多种因素的影响,例如交通流量受到位置、天气和时间点等因素的影响,销售量受到节假日、商品售价、地区特性和气候变化等因素的影响。这些难以捕捉的动态特性使得时间序列预测任务变得复杂<sup>[3]</sup>。在本文中,我们研究小样本多维时序预测问题,即存在多个时序特征的时间序列预测任务,但序列样本数量有限。

传统的时间序列预测算法,如自回归积分移动平均模型 (ARIMA) 及其算法族即使在小样本短期任务中也难以表现出优秀的性能<sup>[4]</sup>,特别是当时序特征稀疏且灵活时<sup>[5]</sup>。近年来,随着深度学习的快速发展,相较于小样本的短期时序预测任务,大样本长序列时序预测任务得到了令人满意的解决。例如,在自然语言处理领域,Transformer 模型通过使用多头注意力机制捕捉多个时序变量之间的相互关系<sup>[6]</sup>,图卷积神经网络 (GCN) 将多维时序数据表示为图结构,并利用图卷积操作来捕捉变量之间的相互关系<sup>[7]</sup>。这些新兴的深度学习模型在海量样本的长时序预测任务中表现出色。然而,少量的数据难以让深度学习模型得到良好的训练,因此这些深度学习模型在小样本短期时序预测任务中的表现平平<sup>[8]</sup>。

### 1.2 小样本多维时序预测解决方法

#### 1.2.1 传统方法

在小样本多维时序预测问题中,传统的时间序列预测方法被广泛研究和应用。其中最常见的方法是自回归积分移动平均模型 (ARIMA) 及其算法族。ARIMA 模型是一种基于线性统计模型的方法,通过对时序数据的自回归 (AR) 部分、差分 (I) 部分和移动平均 (MA) 部分建模,捕捉数据中的趋势和周期性。ARIMA 模型可以根据数据的性质进行参数调整,以适应不同的时序模式。此外,还有一些基于统计分析和时间序列分解的方法,如季节性分解、指数平滑法等。季节性分解方法将时序数据拆分为趋势、季节和残差部分,以便更好地分析和预测其中的模式。指数平滑法则基于移动平均的思想,通过对数据进行平滑处理,提取其趋势和季节成分。然而,传统方法在处理复杂、非线性和高维数据时可能受到限制。这些方法通常依赖于对数据进行线性或统计性假设,并且对于具有长期依赖或非线性关系的数据,其预测能力可能不足。此外,传统方法在选择模型结构和参数时,需要依赖领域专家的经验 and 领域知识,因此在应对新的时序预测问题时可能缺乏灵活性和适应性<sup>[9]</sup>。

#### 1.2.2 深度学习方法

随着深度学习的发展,针对小样本多维时序预测问题,研究人员提出了一些基于深度学习的方法。深度学习方法通过利用神经网络的强大表示学习能力和非线性建模能力,能够更好地处理复杂和高维时序数据。其中,循环神经网络 (RNN) 及其变种,如长短期记忆网络 (LSTM) 和门控循环单元 (GRU),被广泛应用于建模时序依赖关系<sup>[10]</sup>。RNN 通过在网络中引入循环连接,可以捕捉到数据中的时间依赖关系。LSTM 和 GRU 是 RNN 的改进版本,通过门控机制来控制信息的传递和遗忘,从而有效地解决了梯度消失和梯度爆炸等问题,更好地捕捉长期依赖。除了 RNN,卷积神经网络 (CNN) 也被用于多维时序数据的建模和特征提取。CNN 通过卷积操作可有效地提取局部特征,并通过池化操作进行空间降维,使得网络能够更好地处理时序数据。在多维时序预测中,可以将不同变量视为输入的不同通道,利用 CNN 对每个通道进行特征提取,并通过全连接层进行融合和预测<sup>[11]</sup>。此外,注意力机制在多维时序预测中也得到广泛应用。注意力机制可以用于捕捉不同变量之间的重要关系和注意力权重,从而在建模过程中对不同变量赋予不同的重要性。通过引入注意

力机制，网络可以自动学习变量之间的相关性和重要性，提升预测性能<sup>[12]</sup>。

### 1.3 本文研究内容及贡献

作为《机器学习》的课程论文，本文主要是利用机器学习课程中所涉及的深度学习方法对于多维时间序列预测任务进行求解的研究与实践。本文的主要研究内容及贡献可以总结如下：

1. **首先**，对多维时序预测任务进行建模描述，并使用经典机器学习算法对其进行求解；
2. **接着**，使用主流 LSTM 网络对多维时序预测任务进行求解，并对比经典机器学习算法；
3. **最后**，提出了一种微调的长短期记忆网络群算法，以更优的方式解决了短期多维时序预测任务，并通过大量实验验证了算法的有效性。

## 2 多维时序问题定义及经典机器学习算法进行多维时序预测

### 2.1 一般多维时序预测问题定义：

在这一小节，我们讨论一般多维时序预测问题的定义，我们给定一般的多维时序预测问题定义如下：

#### 一般多维时序预测问题定义

(i) 输入：

$$S = \{S_{i,j,r} \mid 1 \leq i \leq n, \quad 1 \leq j \leq k_i, \quad 1 \leq r \leq t\}$$

(ii) 输入描述：

给定序列集合  $S = \{S_1, S_2, S_3, \dots, S_n\}$ ，其中  $n$  是序列样本数量； $S_i = \{S_{i,1}, S_{i,2}, S_{i,3}, \dots, S_{i,k_i}\}$ ，其中  $k_i$  是样本  $i$  的特征特维数，假设  $k_p = k_q = k$ ， $p \neq q$  且  $1 \leq p, q \leq n$ ，即每个样本特征个数相同； $S_{i,j} = \{S_{i,j,1}, S_{i,j,2}, S_{i,j,3}, \dots, S_{i,j,t}\}$ ，其中  $t$  是每个样本的每个特征序列的长度。

(iii) 输出：

有特征  $m$ ， $1 \leq m \leq k$ ，给定  $S = \{S_{i,j,t+h}\}$ ， $1 \leq i \leq m$ ， $1 \leq j \leq k$  且  $j \neq m$ ， $h > 0$ ，输出为： $S_{i,m,t'}, 1 \leq i \leq n, t+1 \leq t' \leq t+h$ 。

特别地，当  $t$  取值较小，且  $k \geq 1$  时，问题演化为小样本多维时序预测问题。

## 2.2 数据集

在本文中, 我们使用 kaggle 竞赛提供的 M5 预测数据集, 该数据样本个数为 30940, 序列长度为 1941, 仅取部分用于实践和算法设计。具体地, 为了测试小样本多维时序预测算法, 我们随机取 20、50、100 个序列制作数据集。该数据集具有多个特征序列, 包括时间类特征 10 个, 销售类特征 2 个, 数据集没有缺失值; 但是数据集各列特征均属于稀疏特征, 含有大量 0 值, 图 1 展示了四个样本的销售量特征, 也是目标特征, 4 类均为所挑选的典型序列, 左上和右下属于较为稀疏类别, 左下和右上为较稠密类别。图 2 展示了 6 件商品 (不同地区) 的历史售价, 可以发现其波动较小, 猜想有可能从分类的角度, 即放弃较大的数据集但是从分类任务的角度做会得到更好的效果。

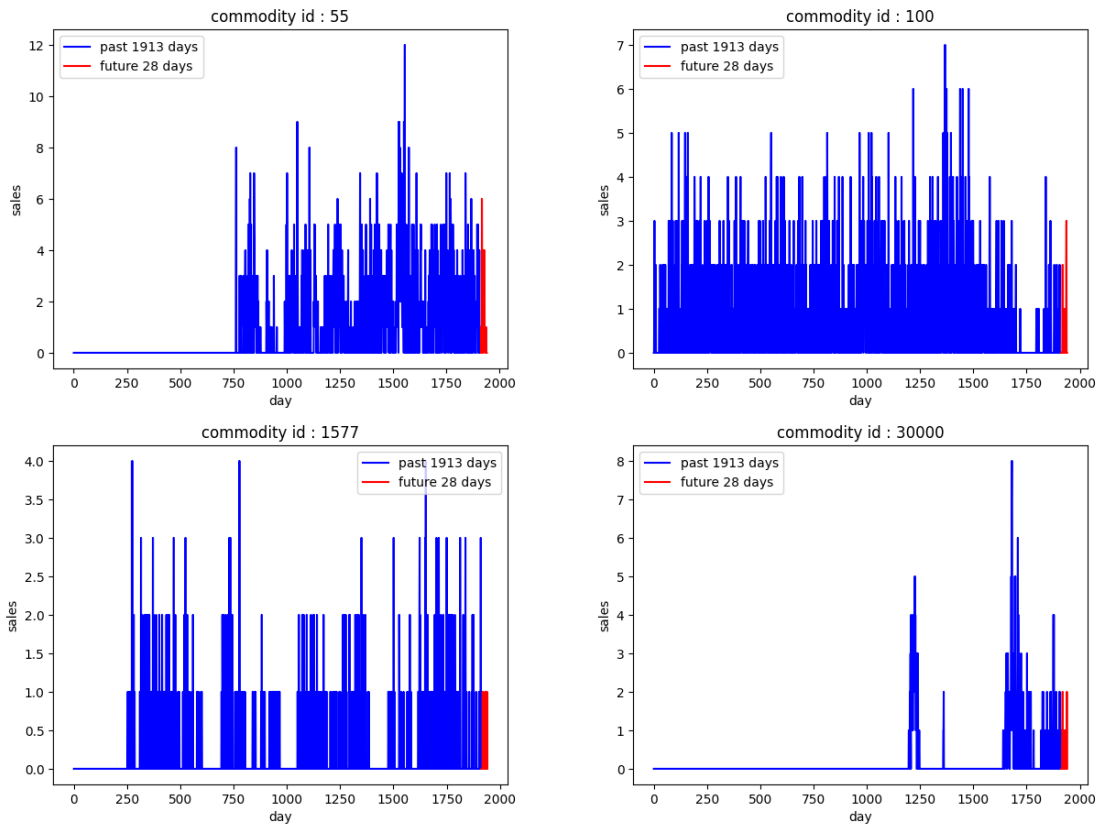
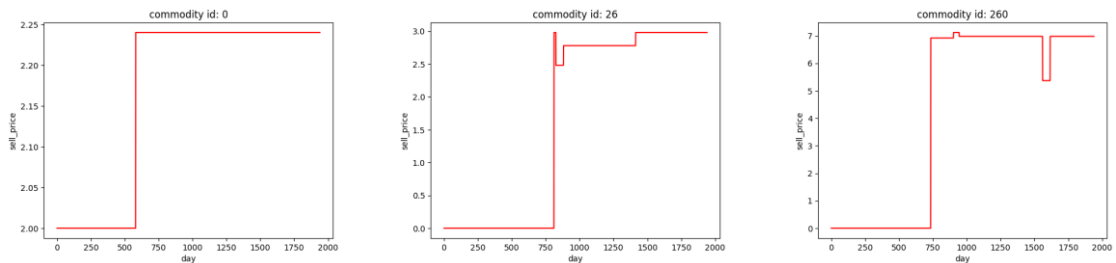


图 1. 销售量特征分布



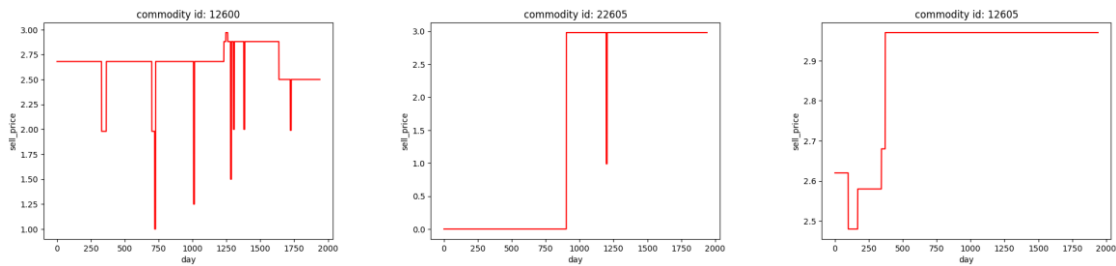


图 2. 售价特征分布

数据预处理：对于时间类别的特征，非数值型特征“snap\_CA”/“snap\_TX”等首先进行整数编码，对于缺失的时间特征使用 0 值填补，对于 sell\_price 数据，使用 pandas 多表查询语句构建与 sales 同样的形状，对于缺失的数据（缺货期间），sell\_price 默认填充为 0。然后按照是否使用时序依赖关系为依据，分别处理成两个小数据集，一个用于经典机器学习算法从回归的角度进行时序预测，一个用于测试和对比 LSTM 和所提出的 FTG-LSTM 算法。

2.3 经典机器学习算法求解

经典机器学习算法在解决多维时序预测任务时面临着一些问题，首先算法的输入输出确定，如果仅仅只看一条序列，那么可以用于训练的样本数将极少，难以精确拟合，尤其时对于本文中所研究的短期序列样本而言，因此必须从另一个角度出发，将序列拆分，舍弃时间点之间的连续性和依赖关系，从而获得一个较大的训练数据集，以便进行充分的拟合<sup>[13]</sup>，即使用经典机器学习算法  $f$  学习映射：

$$S_{i,j,t'} \Rightarrow S_{i,m,t}, j \neq m \text{ 且 } 1 \leq j \leq k$$

为了充分全面的进行实验，我们使用了课上讲过的所有算法，并添加了部分经典机器学习回归算法，对 100 条序列进行预测，序列长度为 1913。实验结果如表 1 所示。

表 1. 经典机器学习算法进行多维时序预测结果（转化为回归问题）

Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
LGBMRegressor	0.34	0.34	3.04	0.72
HistGradientBoostingRegressor	0.34	0.34	3.04	2.8
GradientBoostingRegressor	0.33	0.33	3.05	5.81
XGBRegressor	0.29	0.29	3.15	3.72
MLPRegressor	0.25	0.25	3.24	77.96
RandomForestRegressor	0.2	0.2	3.34	12.13
BaggingRegressor	0.13	0.13	3.48	1.37
SGDRegressor	0.01	0.01	3.71	0.2
BayesianRidge	0.01	0.01	3.71	0.1
RidgeCV	0.01	0.01	3.71	0.11
Ridge	0.01	0.01	3.71	0.08
TransformedTargetRegressor	0.01	0.01	3.71	0.08
LinearRegression	0.01	0.01	3.71	0.11
Lars	0.01	0.01	3.71	0.26
OrthogonalMatchingPursuitCV	0.01	0.01	3.71	0.17
LassoLarsIC	0.01	0.01	3.71	0.11
ElasticNetCV	0.01	0.01	3.71	0.46
LassoCV	0.01	0.01	3.71	0.48

LassoLarsCV	0.01	0.01	3.71	0.29
LarsCV	0.01	0.01	3.71	0.34
KNeighborsRegressor	0.01	0.01	3.71	6.73
PoissonRegressor	0.01	0.01	3.72	0.09
GeneralizedLinearRegressor	0.01	0.01	3.72	0.33
TweedieRegressor	0.01	0.01	3.72	0.1
ExtraTreesRegressor	0.01	0.01	3.72	10.6
OrthogonalMatchingPursuit	0.01	0.01	3.72	0.06
DummyRegressor	0	0	3.74	0.07
ElasticNet	0	0	3.74	0.08
LassoLars	0	0	3.74	0.09
Lasso	0	0	3.74	0.12
NuSVR	-0.04	-0.04	3.81	224.26
SVR	-0.04	-0.04	3.81	105.82
HuberRegressor	-0.07	-0.07	3.86	0.39
LinearSVR	-0.14	-0.13	3.98	6.06
AdaBoostRegressor	-0.21	-0.2	4.1	0.81
DecisionTreeRegressor	-0.38	-0.38	4.39	0.26
PassiveAggressiveRegressor	-0.39	-0.39	4.4	0.17
ExtraTreeRegressor	-0.43	-0.43	4.47	0.2

从表 1 可以看出尽管有一部分机器学习算法的 Adjusted R-Squared 值相对较高,但是其实仍处于较低水准,并不适合该任务。从另一个角度分析,尽管我们舍弃了时间特征的前后依赖性来获得一个较大数据集来预测销售量(回归),但是另一个问题也随之产生了,即特征的严重偏斜<sup>[13]</sup>,参考图 2 中的售价特征。

### 3 主流 LSTM 网络求解

#### 3.1 主流LSTM算法概述

LSTM (Long Short-Term Memory) 是一种经典的循环神经网络 (RNN) 变体,用于处理和建模序列数据。相比于传统的 RNN, LSTM 网络在处理长期依赖性问题表现更出色。LSTM 网络的核心思想是引入了称为“门”的结构,这些门可以选择性地控制信息的流动和记忆的更新。如图 3 所示, LSTM 网络的关键组成部分包括记忆单元,图中标志为 A 的整体,和三个门:输入门、遗忘门和输出门。

输入门:控制新输入信息对记忆单元的影响。它通过一个 sigmoid 激活函数来决定是否接受新的输入,并使用另一个 tanh 激活函数来生成一个新的候选值,用于更新记忆单元的内容,更新公式如下:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)$$

遗忘门:控制之前的记忆是否被遗忘。它通过一个 sigmoid 激活函数来决定是否丢弃之前的记忆,以及保留多少之前的记忆,更新公式如下:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)$$

输出门:控制输出的记忆单元的内容。它通过一个 sigmoid 激活函数来确定输出的哪些部分应该被输出,然后将其传递给一个 tanh 激活函数进行处理,得到最终的输出。

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)$$

细胞状态更新公式如下:

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$

细胞状态（Cell State）更新和输出公式如下：

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$h_t = o_t \odot \tanh(C_t)$$

其中， $x_t$ 为输入数据， $h_{t-1}$ 为上一个时刻的隐藏状态， $W_{xi}, W_{hi}, W_{xf}, W_{hf}, W_{xo}, W_{ho}, W_{xc}, W_{hc}$ 为权重参数， $b_i, b_f, b_o, b_c$ 为偏置参数， $\sigma$ 为sigmoid函数， $\odot$ 为Hadamard积（逐元素相乘）， $\tilde{C}_t$ 为临时的细胞状态， $C_t$ 为当前时刻的细胞状态， $h_t$ 为当前时刻的隐藏状态。

LSTM 网络通过这些门的组合来实现对序列数据的建模和预测。通过输入门、遗忘门和输出门的控制，LSTM 网络可以自动选择性地记住和忘记过去的信息，并从中提取有用的特征和模式。这使得 LSTM 网络能够更好地捕捉长期依赖关系，避免梯度消失或爆炸的问题，从而在处理和预测序列数据时表现更出色<sup>[14]</sup>。

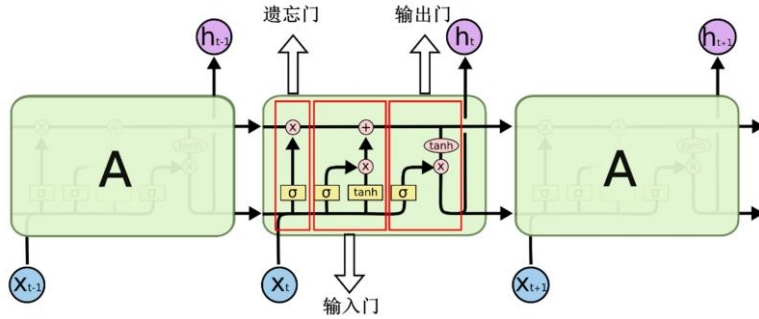


图 3. 经典 LSTM 结构

### 3.2 应用经典LSTM解决小样本短期多维时序预测问题

应用经典 LSTM 解决多维时序预测问题的主要步骤在于：设计 LSTM 网络结构、损失函数选择、LSTM 网络训练。

1.设计 LSTM 网络结构：根据多维时序预测问题的特点，设计包含多个 LSTM 层的神经网络结构。每个 LSTM 层包含若干个 LSTM 单元，每个 LSTM 单元都有输入门、遗忘门和输出门，用于控制信息的输入、遗忘和输出。同时，还需要设置一些全连接层和激活函数，以及输入、输出的维度。

2.损失函数选择：选择合适的损失函数可以有效地衡量模型的预测误差，常用的损失函数包括均方误差（MSE）、平均绝对误差（MAE）和交叉熵等。对于多维时序预测问题，通常使用 MSE 或 MAE 作为损失函数<sup>[15]</sup>。

3.LSTM 网络训练：使用已有的多维时序数据集对设计好的 LSTM 网络进行训练。训练过程中需要设置合适的学习率、批次大小、迭代次数等超参数。训练完成后，可以使用测试数据集对模型进行验证和评估，计算预测误差和准确率等指标。

在实验中，设定 epochs 为 200，lstm\_units 分别为 64、32，学习率 lr 分别为 0.0001、0.001、0.005，分别进行三次训练，训练过程如图 4 所示，从图 4 中可以发现当训练到 30epochs 左右，模型的性能几乎就已经达到最佳了，这再一次展现了经典 LSTM 算法在小样本短期时序数据处理中的不足<sup>[16]</sup>。

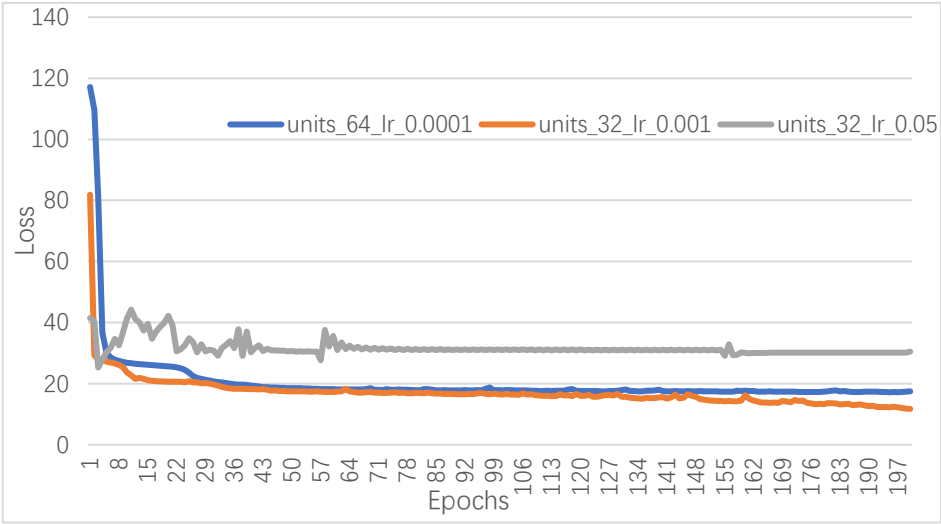


图 4. LSTM 训练过程

在进行四次训练，与上一节中使用相同的数据，对比经典机器学习算法和 LSTM 性能，对比结果如表 2 所示。

表 2. 经典机器学习算法与 LSTM 对比

算法	平均准确率_S1(%)	平均准确率_S2(%)	平均准确率_S3(%)	平均准确率_S4(%)
LGBM	50.23	49.97	53.21	45.61
HGBR	49.79	52.3	55.07	47.71
GBR	48.91	49.79	49.99	46.19
XGBoost	50.02	49.12	52.13	49.12
LSTM	53.44	47.89	55.57	51.21

从上面的对比中可以看出，多数情况下，LSTM 在小样本数据集上短期时序预测表现优于根据回归思想的经典的机器学习算法

4 微调长短期记忆网络群算法（FTG-LSTM）

在本节中，我们首先分析 LSTM 算法在处理短期时序数据时的不足及关键原因，然后介绍我们所提出的算法 FTG-LSTM，一种微调长短期记忆网络群算法，用于提升 LSTM 在小样本短期时序预测任务中的表现，并通过大量的实验证实了其有效性。

4.1 LSTM在小样本多维时序预测任务中存在的挑战

在小样本多维时序预测任务中，LSTM 很有可能存在以下问题：

- (i) 模型泛化能力差：LSTM 在小样本情况下容易出现过拟合，即在训练集上表现很好，但在测试集上表现不佳。这可能导致模型的泛化能力变差，即模型无法很好地适应新的数据集<sup>[17]</sup>；
- (ii) 模型训练困难：在小样本情况下，由于数据量较少，模型很难学习到足够的信息，因此可能需要更长时间的训练或更复杂的模型结构<sup>[18]</sup>；
- (iii) 数据噪声：在小样本情况下，由于数据量较少，数据中可能存在噪声或异常值，这可能会影响模型的预测能力<sup>[19]</sup>；
- (iv) 序列关联性差：在小样本多维时序预测任务中，不同序列之间可能存在较弱的相关性或无关性，这可能会导致模型无法很好地预测这些序列<sup>[20]</sup>。



导致以上问题的关键是小样本造成的欠拟合性，这点可以从上一章最后一小节的分析中得出。应该容易进行下面的分析：在小样本短期多维序列上提升 LSTM 可以从几个可能的方面进行，其一，不严格遵循时序的依赖性，沿用第二部分介绍的经典机器学习算法求解该问题同样的策略，即牺牲时间序列之间的严格时序关系，将多个同类型时间序列粗暴拼接起来，得到一个长序列，使 LSTM 可以得到充分训练，或者在制作 LSTM 输入时，共享不同序列时间段数据，这或许会更合理。其二，数据增强和扩充，既然短期的序列数据 LSTM 难以处理，那么为何不采取手段制作较长期且合理的序列，从这个角度出发可以衍生出更加细致化的特征工程任务，即通过观察序列的历史特征，采取某些方法以不太高的精度在时间维度上扩充时间序列的长度（同时从多个变量）。

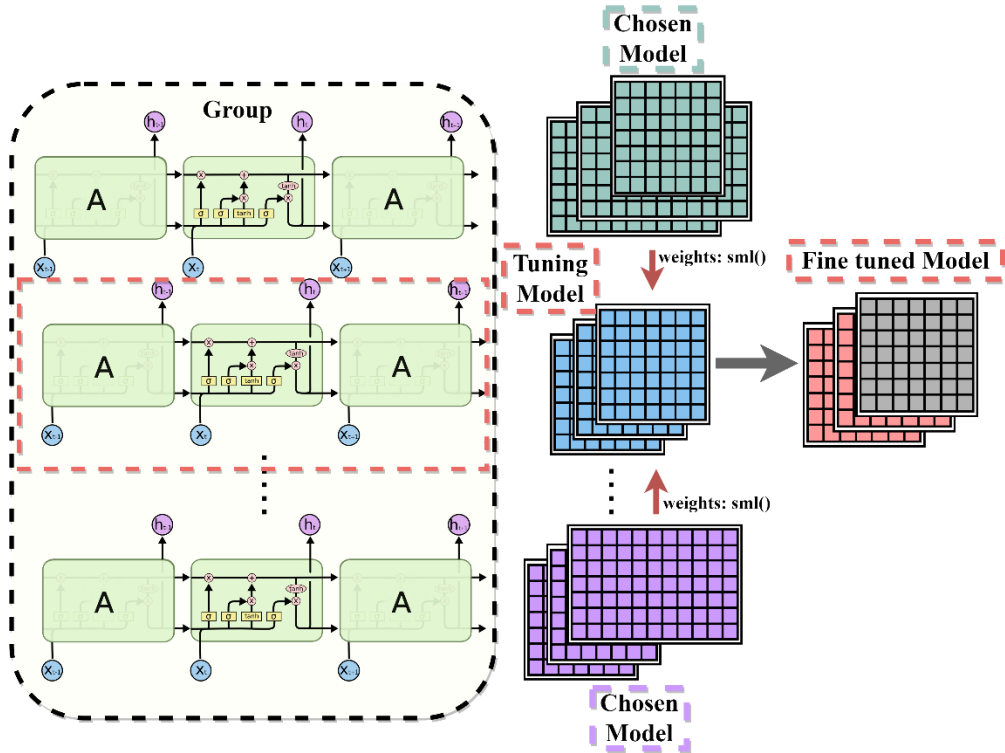


图 5 本文所提出的 FTG-LSTM 框架

4.2 提出的FTG-LSTM算法

在 4.1 节，我们分析了 LSTM 的不足之处，以及 LSTM 在小样本短期时序数据上表现平平的关键原因，并讨论了 LSTM 进一步优化的思路。在这一小节我们详细介绍提出的 FTG-LSTM 算法，其框架如图 5 所示。

FTG-LSTM 算法分为 3 个步骤，首先在一个小样本时序数据集上，即  $n$  为一个较小值（数百或几十），且时序特征的长度较短（数百），对每个  $S_i$  训练一个 LSTM 模型，以一个较小的 epoch 进行（几十），并构成一个 LSTM Group 图 5 中左侧虚线框内的 LSTM 群；然后计算  $S_i$  与  $S_j$  的相似性，本文中使用的是 DTW 指标，并计算参数共享权重；最后使用截断更新方法更新每一个训练集不理想的 LSTM，具体的 FTG-LSTM 算法见算法 1。

## 算法 1. FTG-LSTM

算法 1. 微调长短期记忆网络群算法 (FTG-LSTM) 伪代码

输入: 小样本多维时序数据 $S$ ; 微调权重 $\alpha$ ; 用于微调的比例 $\beta$ ; 欠拟合度量 $acc$ ; 序列相似性指标 $similar$

输出:  $S_{i,m,t'}, 1 \leq i \leq n, t+1 \leq t \leq t+h$

```

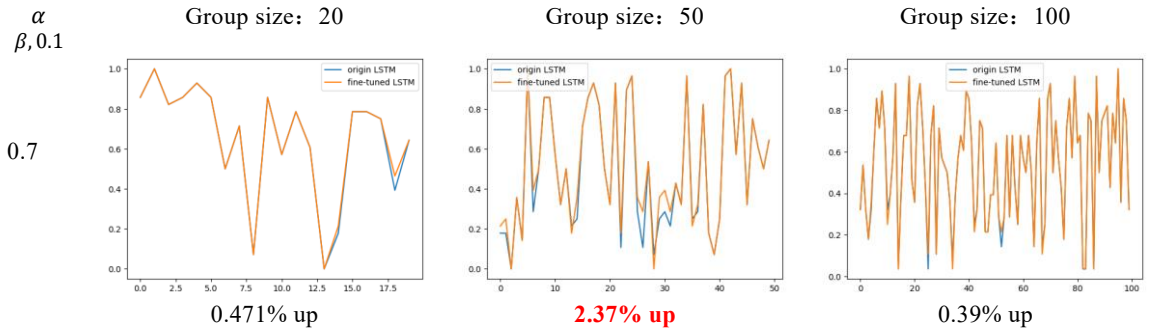
1. TrainSet, TestSet = Split(S)
2. Group =  $\emptyset$ ;
3. For  $i = 1$  to  $n$  do
4.   |  $model = LSTM(S_{i,j,t}, S_{i,m,t})$ 
5.   | Train model (若干 epochs: 前向传播, 计算输出, 计算损失, 反向传播, 参数更新)
6.   |  $Group \Leftarrow Group \cup \{model\}$ 
7. End
8.  $sml \Leftarrow$  zero matrix of size  $n \times n$ 
9. For  $i = 1$  to  $n$  do
10.  | For  $j = i$  to  $n-1$  do
11.  |   |  $sml[i][j] \Leftarrow similar(S_{i,m}, S_{j,m})$ 
12.  |   |  $sml[j][i] \Leftarrow similar(S_{i,m}, S_{j,m})$ 
13.  | End
14. End
15. For model in Group do
16.  | If Accuracy(model, TrainSet) <  $acc$  do
17.  |   | TopNets  $\Leftarrow$  Group 内与 model 训练序列相似性最大的  $(\beta \times n)$  个 model
18.  |   |  $NetWeights \Leftarrow \frac{sml(TopNets)}{\sum sml(TopNets)}$ 
19.  |   |  $model \Leftarrow \alpha \times model + (1 - \alpha) \times NetWeights \times TopNets$ 
20.  | End
21. End
22. return  $S_{i,m,t'}, 1 \leq i \leq n, t+1 \leq t \leq t+h$ ; predict in TestSet with fine-tuned Group

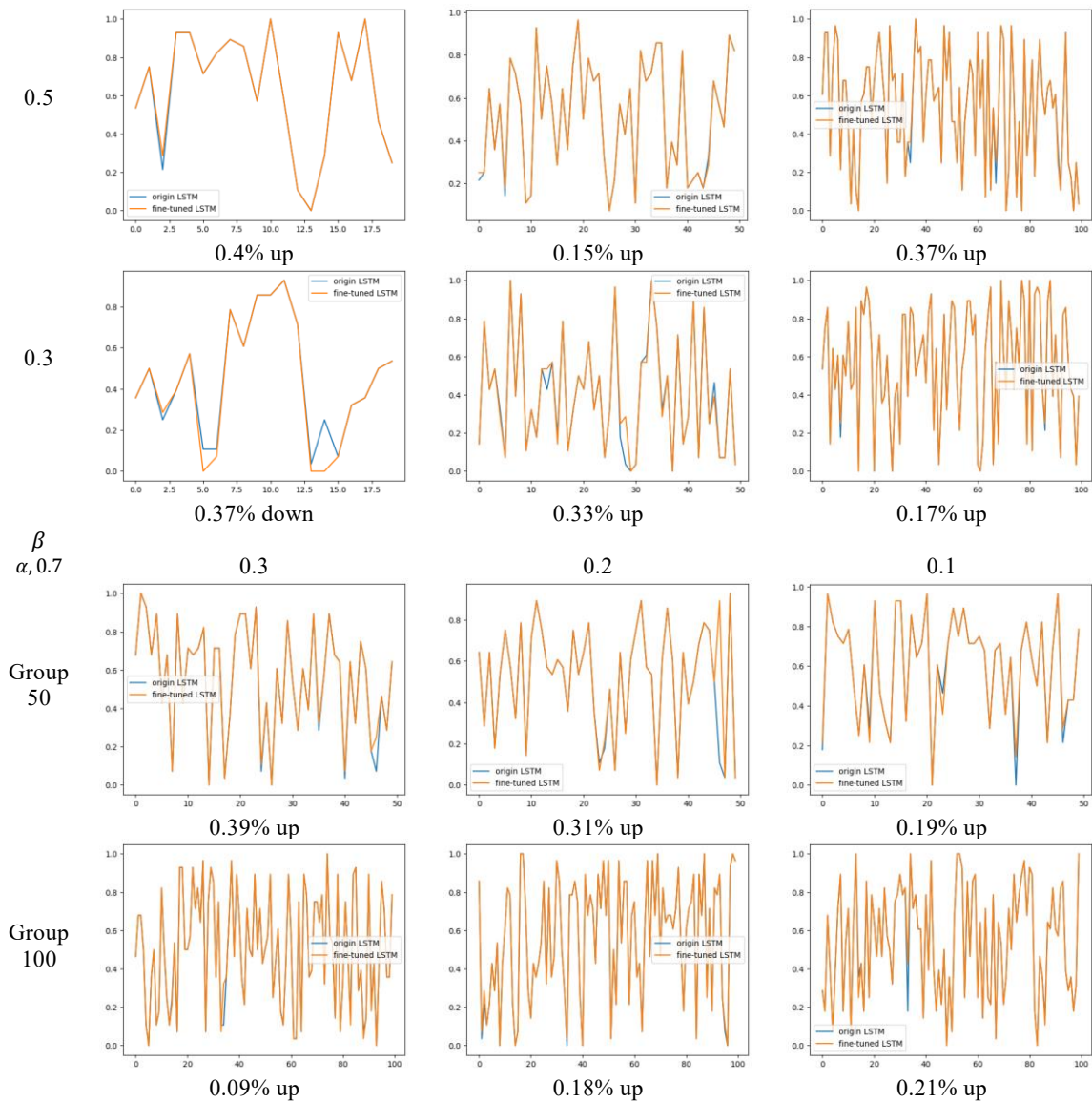
```

## 4.3 实验验证

这一部分实验所用数据为 M5 数据集的一部分, 提取的时间为 150 天, 200 个序列 (每个序列含 10 维稀疏特征), 通过不断调整算法的超参数, 即 $\alpha$ 、 $\beta$ 、Group size, 表 2 展示了 FTG-LSTM 在该小样本短期数据集上的性能, 发现 LSTM 总能提升 Group 内的平均测试集准确率, 且可以观察到在 Group size 越小, 即严格不能数越小时, FTG-LSTM 性能提升更大, 这可以初步认定 FTG-LSTM 对于提升 LSTM 在小样本短期时序数据预测性能上是有效的。

表 3. FTG-LSTM 测试集 Average Accuracy





4.4 算法分析

FTG-LSTM 算法最大的特点是针对小样本多维时序数据，通过微调长短期记忆网络群算法，以解决 LSTM 模型的欠拟合问题。该算法利用序列相似性指标对训练数据进行分组，进而更好地利用数据中的特征，提高预测性能。在模型微调过程中，FTG-LSTM 算法使用与微调模型训练序列相似性最大的一些模型进行权重计算，进一步提高模型的预测精度。FTG-LSTM 算法在多维时序预测问题上具有较好的应用价值，可以帮助解决实际场景中的预测问题，其优缺点总结在表 4 中。

表 4. FTG-LSTM 算法优缺点分析

优点	缺点
能够更好地解决 LSTM 模型的欠拟合问题	算法的计算复杂度较高
可以更好地利用训练数据的特征，提高预测性能，在实验中优于 LSTM 和经典机器学习算法	对于复杂的小样本多维时序数据仍可能存在欠拟合问题

## 5 总结

本文对多维时序预测问题进行了深入研究。多维时序预测是单变量时序预测问题的推广,应用广泛,包括交通、库存管理、生产线故障预测和疾病预测等领域。本文提出了一种名为 FTG-LSTM 的算法,可以有效地优化 LSTM 进行小样本多维时序预测,并通过实验证明了其有效性。首先,本文给出了多维时序预测问题的数学表达,并定义了小样本多维时序预测问题。然后,本文分析了在目标问题上经典机器学习算法的缺陷,并通过实验证实了这些缺陷。接着,本文介绍了 LSTM 网络的基本概念,并在小数据集上频繁训练 LSTM,并将其与经典机器学习算法进行比较。结果表明, LSTM 在小样本多维时序数据上仍然优于经典机器学习算法,并分析了 LSTM 模型在小样本短期数据上难以表现优异的关键因素。最后,本文针对 LSTM 模型的缺点提出了一种名为 FTG-LSTM 的微调长短期记忆网络群算法,并通过大量的对比实验发现, FTG-LSTM 优于 LSTM 和其他经典机器学习算法。总之,本文深入探讨了多维时序预测问题,并提出了一种有效的算法,可以优化 LSTM 进行小样本多维时序预测。实验结果表明,所提出的算法比经典机器学习算法和 LSTM 更优秀,可广泛应用于现实生活中的各种场景。

## References:

- [1] Morid M A, Sheng O R L, Dunbar J. Time series prediction using deep learning methods in healthcare[J]. *ACM Transactions on Management Information Systems*, 2023, 14(1): 1-29.
- [2] Chen W, Van Der Werf J J, Hulscher S. A review of practical models of sand transport in the swash zone[J]. *Earth-Science Reviews*, 2023: 104355.
- [3] Tao P, Hao X, Cheng J, et al. Predicting time series by data-driven spatiotemporal information transformation[J]. *Information Sciences*, 2023, 622: 859-872.
- [4] Keskin S, Külahcı F. ARIMA model simulation for total electron content, earthquake and radon relationship identification[J]. *Natural Hazards*, 2023, 115(3): 1955-1976.
- [5] Babu K, Chakraborty P, Pal M. Planning of Fast Charging Infrastructure for Electric Vehicles in a Distribution System and Prediction of Dynamic Price[J]. *arXiv preprint arXiv:2301.06807*, 2023.
- [6] Kim J, Kang H, Kang P. Time-series anomaly detection with stacked Transformer representations and 1D convolutional network[J]. *Engineering Applications of Artificial Intelligence*, 2023, 120: 105964.
- [7] Lazcano A, Herrera P J, Monge M. A Combined Model Based on Recurrent Neural Networks and Graph Convolutional Networks for Financial Time Series Forecasting[J]. *Mathematics*, 2023, 11(1): 224.
- [8] Qian G, Liu J. Fault diagnosis based on gated recurrent unit network with attention mechanism and transfer learning under few samples in nuclear power plants[J]. *Progress in Nuclear Energy*, 2023, 155: 104502.
- [9] Wu H, Chen S, Ding Y. Comparison of ARIMA and LSTM for Stock Price Prediction[J]. *Financial Engineering and Risk Management*, 2023, 6(1): 1-7.
- [10] Khaldi R, El Afia A, Chiheb R, et al. What is the best RNN-cell structure to forecast each time series behavior?[J]. *Expert Systems with Applications*, 2023, 215: 119140.
- [11] Wang J, Cheng S, Tian J, et al. A 2D CNN-LSTM hybrid algorithm using time series segments of EEG data for motor imagery classification[J]. *Biomedical Signal Processing and Control*, 2023, 83: 104627.
- [12] Du W, Côté D, Liu Y. Saits: Self-attention-based imputation for time series[J]. *Expert Systems with Applications*, 2023, 219: 119619.
- [13] Junaid M, Ali S, Eid F, et al. Explainable machine learning models based on multimodal time-series data for the early detection of Parkinson's disease[J]. *Computer Methods and Programs in Biomedicine*, 2023, 234: 107495.
- [14] Zhao L, Li Z, Qu L, et al. A hybrid VMD-LSTM/GRU model to predict non-stationary and irregular waves on the east coast of China[J]. *Ocean Engineering*, 2023, 276: 114136.
- [15] Li Z, Rao Z, Pan L, et al. Mts-mixers: Multivariate time series forecasting via factorized temporal and channel mixing[J]. *arXiv preprint arXiv:2302.04501*, 2023.
- [16] Shiimoto K. Network Intrusion Detection System Based on an Adversarial Auto-Encoder with Few Labeled Training Samples[J]. *Journal of Network and Systems Management*, 2023, 31(1): 5.
- [17] Qian G, Liu J. Fault diagnosis based on gated recurrent unit network with attention mechanism and transfer learning under few samples in nuclear power plants[J]. *Progress in Nuclear Energy*, 2023, 155: 104502.
- [18] Zabin M, Choi H J, Uddin J. Hybrid deep transfer learning architecture for industrial fault diagnosis using Hilbert transform and DCNN-LSTM[J]. *The Journal of Supercomputing*, 2023, 79(5): 5181-5200.
- [19] Ahmed M R, Islam S, Islam A K M M, et al. An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition[J]. *Expert Systems with Applications*, 2023, 218: 119633.
- [20] Zhang M, Yang D, Du J, et al. A Review of SOH Prediction of Li-Ion Batteries Based on Data-Driven Algorithms[J]. *Energies*, 2023, 16(7): 3167.