



## 基于模型无关优化策略的风电短时回归预测框架设计

丁琦 邱才明 杨浩森 童厚杰

### A Regression Framework Design for Short Term Forecasting of Wind Power Based on Model-Agnostic Meta-Learning Strategy

DING Qi, QIU Caiming, YANG Haosen, TONG Houjie

引用本文:

丁琦, 邱才明, 杨浩森, 等. 基于模型无关优化策略的风电短时回归预测框架设计[J]. 现代电力, 2022, 39(3): 253–261. DOI: 10.19725/j.cnki.1007–2322.2021.0097

DING Qi, QIU Caiming, YANG Haosen, et al. A Regression Framework Design for Short Term Forecasting of Wind Power Based on Model-Agnostic Meta-Learning Strategy[J]. *Modern Electric Power*, 2022, 39(3): 253–261. DOI: 10.19725/j.cnki.1007–2322.2021.0097

在线阅读 View online: <https://doi.org/10.19725/j.cnki.1007–2322.2021.0097>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于深度学习的电网短期负荷预测方法研究

Research on Short-term Load Forecasting Method of Power Grid Based on Deep Learning  
现代电力. 2018, 35(2): 43–48 <http://xddl.ncepujournal.com/article/Y2018/I2/43>

#### 基于相空间重构和长短期记忆算法的电力系统无功负荷预测模型

Reactive Load Forecasting Model Based on PSR-LSTM in Power System  
现代电力. 2020, 37(5): 470–477 <https://doi.org/10.19725/j.cnki.1007–2322.2020.0235>

#### 基于SAE-ELM的电动汽车充电站负荷预测模型

Model of Load Forecasting of Electric Vehicle Charging Station Based on SAE-ELM  
现代电力. 2019, 36(6): 9–15 <http://xddl.ncepujournal.com/article/Y2019/I6/9>

#### 基于Prophet与XGBoost混合模型的短期负荷预测

Short-Term Power Load Forecasting Based on Prophet and XGBoost Mixed Model  
现代电力. 2021, 38(3): 325–331 <https://doi.org/10.19725/j.cnki.1007–2322.2020.0321>

#### 融合分时电价的居民可控负荷优先级控制策略

Priority Control Strategy of Residential Controllable Load Based on Time-of-Use Price  
现代电力. 2021, 38(4): 422–433 <https://doi.org/10.19725/j.cnki.1007–2322.2020.0361>

#### 基于自组织特征神经网络和最小二乘支持向量机的短期电力负荷预测方法

Short-Term Load Forecasting Method Based on Self-Organizing Feature Mapping Neural Network and GA-Least Square SVC Model  
现代电力. 2021, 38(1): 17–23 <https://doi.org/10.19725/j.cnki.1007–2322.2020.0201>

# 基于模型无关优化策略的风电短时回归预测框架设计

丁琦<sup>1</sup>, 邱才明<sup>1,2</sup>, 杨浩森<sup>1</sup>, 童厚杰<sup>1</sup>

(1. 上海交通大学电子信息与电气工程学院, 上海市 闵行区 200240;

2. 华中科技大学电子信息与通信学院, 湖北省武汉市 430074)

## A Regression Framework Design for Short Term Forecasting of Wind Power Based on Model-Agnostic Meta-Learning Strategy

DING Qi<sup>1</sup>, QIU Caiming<sup>1,2</sup>, YANG Haosen<sup>1</sup>, TONG Houjie<sup>1</sup>

(1. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Minhang District, Shanghai 200240, China; 2. School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, Hubei Province, China)

**摘要:** 目前, 风电出力预测面临跨环境、跨传感器设备的多任务挑战, 往往需要对不同的预测目标各自独立地开展针对性训练。鉴于此, 首先提出了一种基于模型无关元学习策略 (model-agnostic meta-learning, MAML) 的短期预测方法, 并基于该方法能够实现对新任务样本快速适应的能力设计了新型回归训练框架。然后结合卷积神经网络-长短期记忆网络、有注意力机制的 Seq2Seq、有自注意力机制的 Transformer、Synthesizer 等时序深度网络模型, 将该框架应用于风力发电预测场景。实验结果表明相比常规的预训练-微调的深度网络训练方法, 所提出的方法在 GEFCom2012 数据集上对各算例实现了均方根误差和均方误差等指标的提高, 同时各模型在短时风电出力为案例的预测任务上的泛化性能获得了一定提升。该训练框架可便捷地将主流深度学习模型和数据集转换为适应 MAML 策略的匹配模式。

**关键词:** 智能电网; 深度学习; 短时电力数据预测; 模型无关优化 (MAML) 策略

**Abstract:** At present, the wind power output prediction has to face with the multi-task challenges including cross environment and cross transducer equipment, so it often needs to conduct targeted training independently for different prediction targets. For this reason, firstly, a short-term prediction method based on model-agnostic meta-learning (abbr. MAML) was proposed. Secondly, based on the ability of the proposed method, by which the new task samples could be rapidly adapted, a new regression training framework was designed. Thirdly,

combining with such sequential depth network models as the convolutional neural network-long and short term memory networks (abbr. CNN-LSTM), the Seq2Seq enhanced with the attention mechanism, the Transformer and Synthesizer enhanced with self-attention mechanism, this framework was applied to the wind power forecasting scene. Experiment results show that comparing with conventional pre-training-fine-tuning deep network training method, the proposed method improves such indicators as root-mean-square error (RMSE) and mean square error (abbr. MSE) on the dataset GEFCom2012 for each computing example, meanwhile, the generalization performance of each model on the prediction task, which takes short-term wind power output as the case, obtains a certain improvement. Besides, this training framework can easily convert the mainstream deep learning regression model and its dataset to the matched pattern adapted to model-agnostic meta-learning (abbr. MAML) strategy.

**Keywords:** smart grid; deep learning; short-term electrical forecasting; model-agnostic meta-Learning (MAML)

**DOI:** 10.19725/j.cnki.1007-2322.2021.0097

## 0 引言

新能源发电设施的出力预测任务在当代高度数字化的电力系统中承担了关键的基础数据支持作用, 如在调度任务中对由复杂传感器网络采样的规模级电力时空数据提供分析服务、或提供量化依据来辅助电力系统的数字化决策等, 从而达到降低能源开销、提供差异化服务等目标<sup>[1]</sup>。特别是近年来融合太阳能、风能等多种新能源的综合能源系统加速入网, 占有比例持续上升并且在

**基金项目:** 国家重点研发计划项目 (2018YFF0214704)  
Project Supported by the National Key R & D Program of China (2018YFF0214704)

部分地区电力系统中逐步占据关键地位。分布式多地部署使得新能源发电端设备极易受到具备时空大数据特性的温度、光照、风力等环境参数影响,导致系统发电端出力变化具有明显的实时性、间歇性,呈现显著的非线性、波动性、不确定性的统计特征,对电力系统稳定安全运行提出了挑战。特别是当前经济恢复的特殊时期,电力数据预测能够有效地为疫情防控和复工复产提供数据支持。在实际负荷预测任务中,往往需要考察基于上述情况在结合气象、气候、地理环境、节假日等因素后生成的多维负荷数据结构体,这对高度依赖相似历史数据特别是气象数据和基于物理模型的传统预测方法提出了严峻的挑战<sup>[2]</sup>,显著提高了多任务预测问题的难度。

目前由于神经网络具有对规模数据进行自动地特征建模等能力,相比经典方法其在多任务预测上的精度指标有显著地提升,因此大量文献基于统计和数据的智能策略提出了许多有效的预测方法并广泛应用于电力负荷预测领域。但从提高神经网络对预测任务的泛化能力出发,重点分析优化和学习训练过程可注意到,多数文献的解决方案呈现典型的二阶预训练-微调分离式训练流程。该流程特点是对已知的规模级标注数据生成预训练模型,在此基础上,重点考察改造模型架构的方法,实现将模型参数迁移至新任务数据集再训练。

如文献[3]从特征融合角度提出,基于聚类算法,首先按时间属性对工作日和周末的短期负荷数据精细分类并分别训练支持向量机预测模型,最后用细菌觅食算法融合多个模型和新类别数据,实现对短期负荷预测精度的提高;文献[4]从数据特征角度,设计二阶迁移方法融合多源域历史负荷数据将 K-means 聚类和门控循环单元 (gate recurrent unit, GRU) 融合作为一次特征提取,并考察基于时间遗忘因子进行二次特征筛选,从而利用二阶迁移方法在极限学习机 (extreme learning machine, ELM) 模型中融合各预测任务的源数据实现精度提升;文献[5]呈现典型的二阶迁移训练特点,首先构建工作日数据的负荷预测网络作为模板网络进行充分训练,然后根据迁移权重选择性冻结模板网络关键权值作为迁移结果后,重新初始化其他结构作为适配节假日数据的预测网络,在新数据上进行微调;文献[6]从多模型综

合架构的角度出发,设计底层深度置信-顶层多任务回归预测框架的多元负荷预测系统,提出以并行的协同训练方式实现模型层面上短期电、热、气等多种负荷预测子任务的直接融合,当预测误差超越允许阈值时,接纳新类型数据再训练,实现对多任务的整体预测框架设计;文献[7]从多源数据融合角度出发,设计多分支输入模型,通过动态分支结构调节用于挖掘、提取历史负荷特征的全连接网络与提取图像数据的标准卷积网络输入,然后嵌入多分支结构的各局部输出,作为底层预测模型的联合输入,从而有效提升了多特征信息的挖掘效能,实现了对多母线多传感器任务在模型层面的融合;文献[8]从重构数据维度的角度出发,首先基于离散小波技术对非线性关系较为复杂的时序数据进行基于离散小波的高低频分解构建扩增数据,然后充分利用 Spark 框架的高效并行训练实现在时钟频率驱动的循环深度网络模型上高低频信号的组合训练,将各组合模型结果取加权评价作为融合的指标;文献[9]从共享权值角度,对智能能量系统 (intelligent energy system, IES) 电、热、冷子任务分别训练独立的长短时记忆网络 (long short-term memory, LSTM) 同时,对各网络剥离若干神经元,拼接成各子任务共享的局部共享 LSTM 网络,从而当在线训练新类型数据时,采取上述典型的硬共享方法迁移共享网络参数进行微调,实现多任务预测精度的提高。

从上述文献比较、分析、总结可见,当前文献对多任务预测的技术路线,或高度依赖于精密设计人工特征筛选,或在流程上选择多种经典模型融合技术的策略组合,或依托对已知高效网络模型施加结构性局部调整。从训练模式和优化方法切入的文献较为稀少。如文献[10]尝试性地将基本二阶优化方法引入负荷预测,通过分割数据构建多个训练模型分批次盲目训练以扩大参数空间探索较优的参数分布,但仅验证了单维度负荷数据情况下多层感知器模型的有效性;文献[11]则侧重整合损失函数,在回归预测的输出模块通过多层线性结构联合多种评价函数作为深度网络训练的综合损失函数,在训练过程中调节各指标的贡献权重。该方法可视为多分支网络融合的一种特例,对优化方法缺少进一步探究。

基于上述分析,为进一步提高预测模型对多源多任务数据集的多维隐含特征融合和参数迁移

能力, 本文从改进深度网络的优化方法的角度, 引入模型无元学习<sup>[12]</sup>(model-agnostic meta-learning, MAML)策略, 并针对性地进行修改, 尝试提升模型预测的泛化性能。MAML 策略及其体系属于元学习中基于优化策略的分支。元学习旨在解决如何学习的问题, 相比典型深度学习学习策略, 主要具备 3 类特征: 训练过程中包含学习子系统、具备可迁移的已训练模型参数、动态调节学习偏差。

MAML 策略基于二阶优化思想, 起初针对呈现数据类别多、单类样本数据规模少的小样本数据学习任务, 其优势在于: 第一, 二阶优化思想强调在模型的训练优化流程中构造双循环, 外循环用于不断对随机建构的任务更新神经网络的初始参数状态, 内循环将外循环训练得到的参数初始状态对新数据集施加若干快速迭代, 实现利用内循环的二阶导信息迫使模型习得更优的初始化参数的思想<sup>[13]</sup>, 从而能够具备比一般的预训练-微调的分离策略对多任务具备更好的快速适应能力; 第二, MAML 策略自身是一种简洁的优化框架, 以二阶优化的角度为多任务回归问题提供了一种能快速兼容多数基于随机梯度下降优化方法的深度学习模型的高效学习策略, 在实践中已验证可取得和设计迁移特征或模型融合方法相媲美的性能提升<sup>[14-15]</sup>, 可直接利用该框架改造已有的深度神经网络模型。

基于上述分析, 本文以多任务风机出力预测为例, 进一步研究 MAML 策略在电力负荷预测场景的应用, 提出一种基于 MAML 策略、面向多任务的回归预测框架。

## 1 基于 MAML 策略的短时预测框架设计

### 1.1 多任务回归预测问题

短时风电出力预测问题呈典型的回归特征。具体考察多任务数据集 $X^{T, N_t, k_t}$ , 其中:  $T$ 为任务数;  $N_t$ 、 $k_t$ 分别为第 $t$ 类任务的采样样本数和传感器输入维度; 匹配的功率数据为 $Y^{T, N_t}$ , 则优化目标可形式化描述为

$$\min_{\theta} \sum_{t=1}^T \sum_{n=1}^{N_t} L(f_{\theta}(X_{t,n}), Y_{t,n}) \quad (1)$$

式中: 回归模型为 $f$ ; 其模型参数为 $\theta$ ;  $L$ 为损失函数。本文中 $L$ 采取均方误差 (mean squared error, MSE)、

均方根误差 (root mean squared error, RMSE)、平均绝对误差 (mean absolute error, MAE)、平均绝对百分比误差 (mean absolute percentage error, MAPE) 作为回归模型训练质量的指标, 在批训练下各指标定义可描述为

$$L_{\text{MSE}} = \frac{1}{N} \sum_i \|f_{\theta}(X_i) - Y_i\|^2 \quad (2)$$

$$L_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_i \|f_{\theta}(X_i) - Y_i\|^2} \quad (3)$$

$$L_{\text{MAE}} = \frac{1}{N} \sum_i |f_{\theta}(X_i) - Y_i| \quad (4)$$

$$L_{\text{MAPE}} = \frac{1}{N} \sum_i \left| \frac{f_{\theta}(X_i) - Y_i}{Y_i} \right| \quad (5)$$

式中:  $N$ 为数据的批大小。

在典型的预训练-微调的二阶方法中, 多任务回归预测问题按任务类别根据实际情况拆分成预训练数据集 $T_0$ 和微调数据集 $T_1$ , 然后分别就 2 个数据集独立执行回归训练, 即:

$$\min_{\theta_0, \theta_1} \sum_{t=1}^{T_0} \sum_{n=1}^{N_t} L(f_{\theta_0}(X_{t,n}), Y_{t,n}) + \sum_{t=1}^{T_1} \sum_{n=1}^{N_t} L(f_{\theta_1}(X_{t,n}), Y_{t,n}) \quad (6)$$

并对各任务 $s$ 重新构建新的随机数据批次, 交替执行周期 $t$ 下的梯度计算:

$$\theta_t^s = \theta_{t-1}^s - \eta \nabla_{\theta_{t-1}^s} L(f_{\theta_{t-1}^s}) \quad (7)$$

显然在该训练策略中, 虽然考察了联合调控损失计算、设计模型结构等融合策略, 对数据集 $T_0$ 和 $T_1$ 仍采取了训练独立、分离优化的方法, 其融合效果存在进一步提升的空间。

### 1.2 基于 MAML 策略的回归预测任务框架

#### 1.2.1 框架介绍与 MAML 策略描述

框架参考 MAML 策略对多分类小样本数据的元学习设计, 将该设计的特点迁移至多任务预测任务, 尝试实现利用单回归模型融合多任务数据的泛化效果。优化目标可描述为

$$\min_{\theta} \sum_{s_1 \in T_1} \sum_{s_0 \in T_0} L(f_{\theta}(X_{s_1}), Y_{s_1}; \theta_t^{s_0}) \quad (8)$$

式中:  $\theta_t^{s_0}$ 为以 $s_0$ 任务体训练 $t$ 周期后的参数。具体如图 1 所示, MAML 策略具体可划分为数据集生成、MAML 训练和微调 3 个阶段。数据集生成是基于综合任务体概念, 为在训练步骤中融合多任务数据, 将各独立不相关的预测任务随机重组为正交的二分数据集, 即 MAML 数据集 $D$ 和微调数据集 $D'$ 。注意到相比标准训练集, 该策略依照时

序关系随机抽取各原始预测任务的部分片段，从而在二分数据集训练过程中确保按照原任务采样的顺序逐次提取各细分数据批次。进而在训练阶段构建了基于双层循环对任务样本的差别损失计算，即首先利用 MAML 数据集  $D$  的数据批次充分训练模型，再迁移模型参数对微调数据集  $D'$  快速适应的策略。

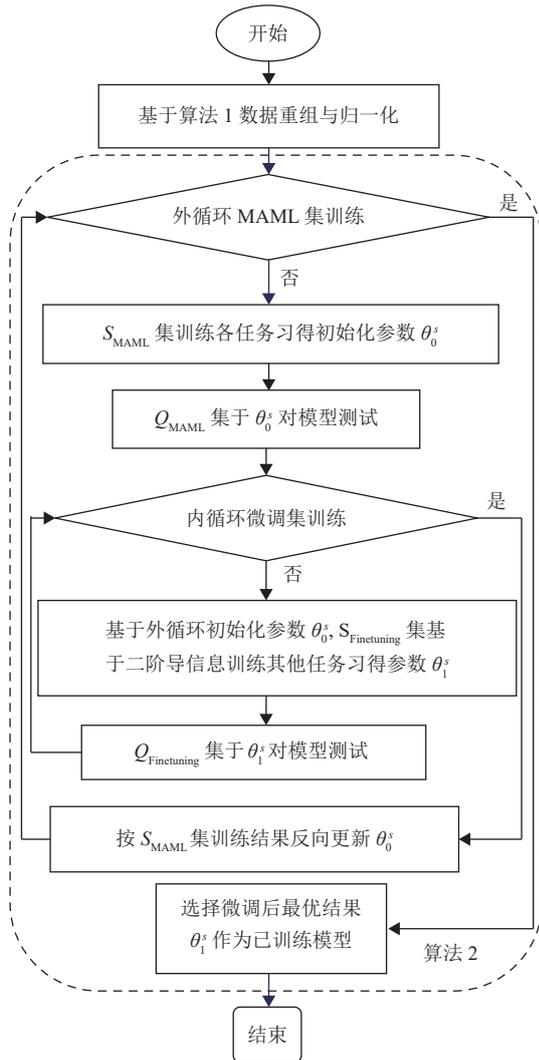


图1 基于 MAML 策略的短时预测框架流程图  
Fig. 1 Framework of MAML strategy-based short-term forecasting

具体地，考察数据集  $D^{T_0}$  和  $D^{T_1}$ ，MAML 框架下的学习过程可概括为基于双循环结构的外循环 MAML 训练过程和内循环微调训练的交替进行，其中前者描述为：

$$\theta_0^* = \operatorname{argmin}_{\theta_0} \sum_{t=1}^{T_0} \sum_{n=1}^{N_t} L_D(f_{\theta_0}(\theta_{\text{init}})) \quad (9)$$

后者描述为：

$$\theta_1^* = \operatorname{argmin}_{\theta_1} \sum_{t=1}^{T_1} \sum_{n=1}^{N_t} L_{D'}(f_{\theta_1}(\theta_0^*)) \quad (10)$$

其中  $\theta_{\text{init}}$  为初始化参数。由公式 (9)(10) 可见，本文采取的 MAML-微调策略与预训练-微调策略的区别是，MAML 策略随机构建了对预训练数据集  $D$  的参数空间，即：

$$\theta_{\text{MAML},t}^* = (\theta_1, \dots, \theta_{t-1} - \eta \nabla_{\theta_{t-1}} L_D(f_{\theta_{t-1}})) \quad (11)$$

如此构建和搜索实现了对各任务体更优的参数初始化设定，从而辅助内循环微调在其他任务数据  $D'$  上训练和搜索的最优解，即

$$\theta_{t,t_1+1}^* = (\theta_{\text{MAML},t}^*, \dots, \theta_{t,t_1} - \eta \nabla_{\theta_{t,t_1}} L_{D'}(f_{\theta_{t,t_1}})) \quad (12)$$

式中： $t, t_1$  分别为外、内循环周期。

### 1.2.2 数据集生成

为生成适于外循环搜索参数初始化分布的分割数据集，使内循环按公式 (12) 习得的  $\theta^*$  对各任务在独立训练下的理想参数  $\theta_s^*$  的距离期望最小，即

$$\min_{\theta^*} \sum_{s \in S} R(\theta^*, \theta_s^*) \quad (13)$$

式中： $s$  为综合任务体； $R$  为参数间的距离度量。首先根据任意预设的容积为  $T_0$  的抽样任务字典，将给定的  $T$  类任务数据  $X^{T, N_t, k_t}$  和功率数据  $Y^{T, N_t}$  按任务维度随机拆分、重组为 MAML 数据集  $D^{T_0, N_t, k_t+1}$  和微调数据  $D^{T-T_0, N_t, k_t+1}$ 。

具体地，对拆分数据，从  $T_0$  中随机抽取  $n_{\text{way}}$  个单任务，每个单任务再随机抽取  $k_{\text{shot}}$  个样本，称该抽样操作生成的数据结构为综合任务体。特别地，为有效应用随机梯度下降 (stochastic gradient descent, SGD) 优化器且不破坏原初数据的时序依赖性，需保持连续抽取  $k_{\text{shot}}$  长的顺序样本。如此多次随机抽取单任务的样本片段并重构为综合任务体后，可进一步将多个综合任务构成反向传播中损失函数计算的单个数据批次。其中，称对二分集  $D'$  和  $D$  再次二分后生成的片段长  $k_{\text{shot}}$  的数据集为支持集  $S^{n_{\text{way}}, k_{\text{shot}}, k_t+1}$ ，剩余样本同操作构成查询集  $Q^{n_{\text{way}}, k_{\text{query}}, k_t+1}$ 。

由此，通过随机抽取的组合方式，实现了将多任务数据按上述“部分任务-部分样本”结构的综合任务体重构为 4 个数据集，即  $S_{\text{MAML}}$ 、 $Q_{\text{MAML}}$ 、 $S_{\text{Finetuning}}$ 、 $Q_{\text{Finetuning}}$ 。具体算法可见附录算法 1。

### 1.2.3 训练策略

区别于对训练-验证集执行反向传播、在测试集执行预测的常规训练，该框架呈现了双层循环、局部误差计算、分组更新的特点。

首先对  $S_{\text{MAML}}$  数据集综合任务体执行推理计算，执行依 SGD 方式的梯度计算，获得该任务体下的模型参数  $\theta_s$ ，但不应用  $\theta_s$  更新该阶段模型参数。

然后进入微调阶段，应用参数  $\theta_s$  生成回归模型副本，对  $S_{\text{Finetuning}}$  数据集执行推理和梯度计算获得  $\theta'_s$ ，再应用参数  $\theta'_s$  对  $Q_{\text{Finetuning}}$  数据集执行验证操作。完成微调阶段后，利用和微调数据集无关的模型参数  $\theta_s$  对  $Q_{\text{MAML}}$  数据集执行推理计算、反向传播和梯度更新并作为本轮的最终更新结果。如此再从  $S_{\text{MAML}}$  取下一个综合任务体重复操作，直至训练完成，以微调阶段最优模型参数  $\theta'$  为最终结果。具体的训练算法可见附录算法 2。

具体地，设第  $s$  个综合任务体初始参数为  $\theta_0^s$ ，则单循环学习下，第  $t$  个训练周期参数  $\theta_t^s$  的梯度更新根据泰勒公式可展开为

$$g_{t,s} = \frac{\partial L}{\partial \theta_0^s} + \frac{\partial^2 L}{\partial \theta_0^s} (\theta_t^s - \theta_0^s) + O(\|\theta_t^s - \theta_0^s\|^2) \quad (14)$$

再考察双循环结构下权值梯度，若第  $t$  次外循环时，内循环完成  $t_1$ ；损失函数  $L$ ；初始参数  $\theta_0$ ；内循环权值更新函数为  $U_t$ ，满足：

$$U_t(\theta_{t,t_1}^s) = \theta_{t,t_1-1}^s - \eta L'_{s,t}(\theta_{t,t_1-1}^s) \quad (15)$$

用公式 (15) 及依梯度传导的链式法则有：

$$g_{\text{MAML}} = L'_t(U_{t_1} \dots (U_{t_1-i}(U_0(\theta_0)))) = \prod_{k=0}^{t_1-1} (I - \eta L'_t(\theta_{t,k}^s)) g_t \quad (16)$$

代入公式 (14) 至公式 (16) 并忽略极小项后，

$$g_{\text{MAML}} \rightarrow L'_t - \eta \sum_{k=0}^{t_1-1} L''_k L'_t - \eta L''_t \sum_{k=0}^{t_1-1} L'_k \quad (17)$$

由上述可见，该框架的核心策略是构造多个随机的综合任务体后，在双层循环中，由外循环对分割生成的时序片段批次按公式 (17) 前项随机搜索更优的初始化情况，由内循环按后项二阶导信息将局部优化后的参数  $\theta_s$  快速适应至对其他任务的回归预测任务，最终实现提升回归模型泛化能力的目标。

## 1.3 预测模型介绍

### 1.3.1 Seq2Seq 预测模型和注意力机制

该模型具备天然适应时序数据张量的编码-

解码结构。具体地，编码器、解码器结构的核心组成为多层长短时记忆网络<sup>[16-18]</sup>，解码器以单层全连接网络输出预测数值。

注意力机制用于解决在长序列输入情况下，Seq2Seq 中编码器编码的语义向量无法充分表征长序列信息、长序列信息在编码时前置信息覆盖等问题。该机制重点聚焦区分 Seq2Seq 模型中编码器和解码器的空间状态的相似性，并对解码器各输入状态执行软寻址操作以提高关键输入的概率占比。

其机制分为二步，首先，计算编码器隐含状态  $h_j$  与编码器隐藏状态  $s_{i-1}$  的关联权重  $e_{ij}$ ，即

$$e_{ij} = V^T \tanh(Ws_{i-1} + Uh_j) \quad (18)$$

式中： $V$ ， $U$  为注意力机制的权值空间。

根据权重系数对编码器输入执行 Softmax 归一化处理获得关联系数  $a_{ij}$

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})} \quad (19)$$

第二步，以  $a_{ij}$  作为权重对编码器首次输入的隐含状态  $h_j$  或多次时序计算后的隐含状态  $s_{i-1}$  取加权作为解码器的部分输入。

### 1.3.2 时序数据的窗口机制

图 2 以 Seq2Seq 模型为案例，描述了经滑窗操作的数据流在计算图中的传播情况。具体地，对序列长度为  $L$  的  $k$  个传感器阵列数据  $x^{L,k}$ ，生成总窗口长度为  $m$ 、其中包含  $u$  个后置预测数据的传感器数据窗口  $x^{m,k}$ ，并匹配生成总窗口长度为  $m-u$  的历史功率数据窗口  $y^{m-u}$ 。传感器窗口  $x^{m,k}$  经由编码器编码生成隐含状态  $h$ ，并协同解码器上一隐含状态  $s$  经注意力模块处理后嵌入到历史功率窗口数据中作为解码器的联合输入，最终经解码计算生成预测长度为  $l$  的回归值  $y^l$  参与反向传播。

特别地，为进一步强化不同窗口间沿采样维度的时序数据的时间耦合关系，减少训练过程的逐批次平均化计算误差。若滑窗的滑动步长为  $t'$ ，则在扫描顺序数据片段时，取本轮窗口对应的上一窗口输出预测值的后  $m+l-t'$  个功率作为下轮采样步骤中历史功率窗口的早期输入。从而基于对历史预测输出的考察，降低了功率预测输出部分沿时序采样的累计误差，一定程度上缓解了对长时序输入功率后端数据的预测影响问题。

### 1.3.3 Transformer 预测模型和自注意力机制

Transformer 网络呈编码-解码结构, 特点是以自注意力机制替代了循环网络结构<sup>[19]</sup>。该机制同样针对长序表达和网络记忆问题, 动态建立了任意长度下长序列数据的长距离依赖关系, 从而以点积形式实现自对齐的学习即

$$Z = \text{softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right)V \quad (20)$$

式中:  $Q$ 、 $K$ 、 $V$  为对同一输入向量经线性变化生成的二维矩阵, 即  $Q = XW$ , 区别于注意力机制中对计算时序的软编码;  $d_K$  为用于缩放点积的方阵  $K$  维度;  $QK^T$  积用于评估输入向量和模型记忆特征间标记对标记的相似度, 从而评估窗口内单个负荷数据对整体采样序列的相对重要性。此外, 由于该结构取缔了循环结构, 可有效提高训练的并行度。

Transformer 网络的编码器由多输入注意力模块和线性层构成, 其中单个多输入注意力模块等效于多个自注意力模块的拼接集成。解码器额外增加了注意力机制, 考察解码器输出  $Q$  和自身自注意力模块变量  $K$ 、 $V$  的相似度。

### 1.3.4 Synthesizer 预测模型

该模型进一步简化了 Transformer 模型的自注意力机制, 可视为对后者的压缩<sup>[20]</sup>。该策略构造了参数化函数  $F$ , 直接将输入变量的序列长度投影至输出变量长度。具体地, 将自注意力模块公式 (20) 的状态变量  $K$  常数化或初始随机化, 将动态自注意力计算的点积操作, 应用双线性层简化至静态矩阵计算, 即

$$Z = \text{softmax}(\text{Relu}(XW_1)W_2)V \quad (21)$$

即 Synthesizer 弱化了自对齐学习中的标记间的交互, 是一种取缔了标记对标记的点积计算的简化变体, 可取得和 Transformer 可竞争的预测性能。Transformer 和 Synthesizer 采取与 Seq2Seq 相同的时序窗口生成机制, 其计算图相当于图 2 以相应模块取缔 LSTM 和注意力模块。

## 2 算例分析

### 2.1 实验设定

#### 2.1.1 数据集介绍与设定

本文采用数据集 GEFCom2012。该数据集记录了 7 个风电场从 2009 年 7 月 1 日—2012 年 6 月 28 日的历史功率出力测量和风力信息, 其中

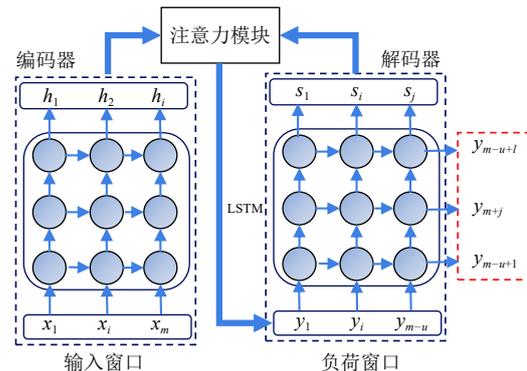


图 2 以 Seq2Seq 模型为例的窗口数据机制示意图

Fig. 2 Schematic diagram of windowed data mechanism taking Seq2Seq model for example

风力信息包括纬向和经向风分量及其对应的风速和风向等 4 类信息, 采样间隔为 1h, 记录形式为在每月奇数日 0:00 和 12:00 开始各独立测量 48 次。为便于对比, 选择 2010 年 12 月 31 日前的数据用于训练, 数据集其余部分用于预测评估。

#### 2.1.2 实验设计与模型参数介绍

为有效对比本文提出的基于 MAML 的回归预测框架和预训练-微调的一般模式, 将本文方法作为实验组, 并设计对照组如下: 将  $S_{MAML}$ 、 $Q_{MAML}$ 、 $S_{Finetuning}$  合并作为训练集、将  $Q_{Finetuning}$  作为验证集。同时为进一步验证框架有效性, 预测模型包括 4 类: 算例部分介绍的有注意力机制 Seq2Seq 和有自注意力机制的 Transformer、Synthesizer, 以及常见的卷积神经网络 (convolutional neural network, CNN)-LSTM 模型, 其中 CNN-LSTM 的架构设计为将经过单层尺寸不变的卷积层和平均池化层处理后的风速信息嵌入历史功率窗口作为 LSTM 网络的输入, 其输出展开后经全连接层变换输出预测功率。如无其他说明, 实验组和对照组采用相同的参数设置, 如注意力机制等, 在训练过程中采取 Adam 优化器。特别地, 为降低连续片段的预测误差积累, 在训练过程中, 将本次窗口的先期输入替代为上次窗口的输出后段。

实验实现基于 Pytorch 1.8 框架; 计算平台为 NVIDIA 1080 GPU, 具体实验参数设置如表 1 所示。

#### 2.2 数据预处理及生成

数据预处理过程分为补偿和归一化操作 2 步。首先对 7 个风电场的风力输入数据和出力数据的缺失部分按插值补偿处理, 其次逐维采取离差标准化预处理, 将数据归一化放缩至  $\pm 1$  之间。

表 1 实验参数和模型参数设定  
Table 1 Setting of experiment parameters and model parameters

参数名称	数值设定
数据批尺寸/个	2
初始学习率 (Transformer/其他)	$5 \times 10^{-3}/1 \times 10^{-4}$
学习率调节周期/次	5
标准训练周期/次	42
MAML训练周期/次	11
微调训练周期/次	8
综合任务体数/个	4
窗口总长/位	24
预测输入长度/位	2
预测输出长度/位	2
窗口滑动步长/位	1
编码器/解码器LSTM深度/层	2/2
编码器LSTM中间变量/个	64
解码器LSTM中间变量/个	64
Transformer/Synthesizer编码器深度/层	2/2
Transformer/Synthesize解码器深度/层	1/1
Dropout设置率	0.5
梯度值约束值	5

为进一步利用功率数据的时序依赖性和提高模型训练效率，本文利用滑动窗口沿时序顺序对原始数据扫描。其中对部分数据缺失的早期窗口或后期窗口采取延展重复末端数值的补偿方式。在利用前滑窗部分模型输出更新历史功率窗口时，直接采纳最近一次的迭代结果以简化重叠部分的处理。

### 2.3 结果可视化与实验分析

本文以 Seq2Seq、CNN-LSTM、Transformer、Synthesizer 模型为算例，重点考察各模型在 MAML-微调模式和预训练-微调模式下因训练模式差异导致的精度变化。为便于可视化，在图 3-6 中截取部分拟合结果片段 (其中蓝色曲线为归一化后输出的真实值，红色曲线为同一模型在 MAML 策略下的结果，绿色曲线为同一模型在预训练-微调策略的结果)，其中横轴仅表征截取片段的相对时间间隔。相应的指标计算结果在表 2-5 中展示，各数值保留 4 位有效数字。

注意到本文提出的基于 MAML 策略的预测框架在对传感器数据和历史数据融合更充分的 Seq2Seq 模型上的表现更优。由于 Synthesizer 模型是 Transformer 模型注意力机制即公式 (20) 中点积计算的轻量化压缩版本，故存在一定的精度损失。同时注意到，虽然各模型在 MAML 训练策略下均取得了对各指标的提升 (部分模型获得

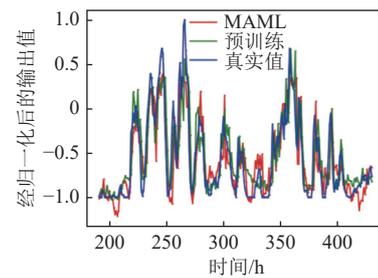


图 3 Seq2Seq 模型在不同训练策略下对部分数据预测效果的局部可视化

Fig. 3 Local visualization of partial data forecasting effect by Seq2Seq model under different training strategy

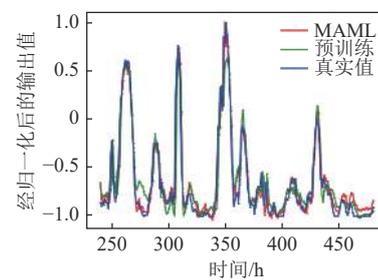


图 4 CNN-LSTM 模型在不同训练策略下对部分数据预测效果的局部可视化

Fig. 4 Local visualization of partial data forecasting effect by CNN-LSTM model under different training strategies

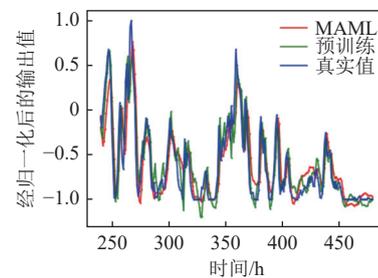


图 5 Transformer 模型在不同训练策略下对部分数据预测效果的局部可视化

Fig. 5 Local visualization of partial data forecasting effect by Transformer model under different training strategies

了较为显著的提升效果)，特别是 MAML 策略对有较大梯度变化的输出峰片段有更好的拟合效果，但对于梯度变化较小、波动性较为明显的输出峰谷片段的拟合效果提升非常有限。由于分割数据集的随机重组实时性对时序数据造成部分破坏，以及窗口生成机制对早期窗口先期输入、后期窗口末端输入缺失部分自动补全的失真影响，使得 MAML 策略的预测结果相比“预训练-微调”策略，预测片段存在更明显的毛刺和噪音。

由上述结果可见，本文提出的基于 MAML

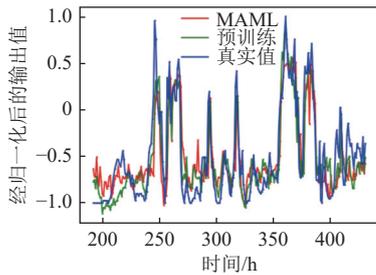


图6 Synthesizer模型在不同训练策略下对部分数据预测效果的局部可视化

Fig. 6 Local visualization of partial data forecasting effect by Synthesizer model under different training strategies

表2 Seq2Seq模型在不同训练策略下的指标对比  
Table 2 Indicator comparison of Seq2Seq model under different training strategies

指标	MAML-微调	预训练-微调
MSE	0.04555	0.08161
RMSE	5.3606	7.1768
MAE	0.1739	0.2504
MAPE	0.4836	0.6024

表3 CNN-LSTM模型在不同训练策略下的指标对比  
Table 3 Indicator comparison of CNN-LSTM model under different training strategies

指标	MAML-微调	预训练-微调
MSE	0.06168	0.08524
RMSE	6.2394	7.3378
MAE	0.1955	0.2588
MAPE	0.5928	0.6512

表4 Transformer模型在不同训练策略下的指标对比  
Table 4 Indicator comparison of transformer model under different training strategies

指标	MAML-微调	预训练-微调
MSE	0.09668	0.1042
RMSE	7.8108	8.1126
MAE	0.2204	0.2356
MAPE	0.6891	0.7338

策略作用在各算例模型时均一定程度上提高了其回归预测指标。结果表明，相比预训练-微调策略会极大受限于切换任务数据集时部分破坏了数据的时序依赖导致拟合难度增加等问题，MAML策略因具备基于构建综合任务体随机抽取任务集合快速训练并对新类型数据有效适应的学习能力，从而部分缓解了上述弊端，提升了模型的泛化能力。

表5 Synthesizer模型在不同训练策略下的指标对比  
Table 5 Indicator comparison of synthesizer model under different training strategies

指标	MAML-微调	预训练-微调
MSE	0.9331	0.1178
RMSE	7.6731	8.2971
MAE	0.2517	0.2863
MAPE	0.6048	0.6512

### 3 结论

本文针对电力时序数据的多任务预测问题，构建了基于MAML策略的新型回归预测训练框架。该框架应用在带有注意力机制的Seq2Seq和CNN-LSTM预测模型后，实验结果表明相比预训练-微调方法，具有良好的迁移能力和一定的工程价值，能够快速改造已有模型，进一步取得泛化性更好、预测精度更高的实际效果。

未来工作将进一步在小样本数据缺失、非平衡时序数据、复杂任务等工况下研究动态窗口设计、扩展部署模型等工作。

(本刊附录请见网络版，印刷版略)

### 参考文献

- [1] 肖泽青, 华昊辰, 曹军威. 人工智能在能源互联网中的应用综述[J]. 电力建设, 2019, 40(5): 63-70.  
XIAO Zeqing, HUA Haochen, CAO Junwei. Overview of the application of artificial intelligence in energy internet[J]. Electric Power Construction, 2019, 40(5): 63-70(in Chinese).
- [2] 邓永生, 焦丰顺, 张瑞锋, 等. 配电网规划中电力负荷预测方法研究综述[J]. 电器与能效管理技术, 2019(14): 1-7.  
DENG Yongsheng, JIAO Fengshun, ZHANG Ruifeng, et al. Research review on power load forecasting methods in distribution network planning[J]. Electrical & Energy Management Technology, 2019(14): 1-7(in Chinese).
- [3] 苏适, 周立栋, 万筱钟, 等. 计及气象因素的用电负荷短期分时分类预测模型与方法[J]. 电力建设, 2017, 38(10): 76-83.  
SU Shi, ZHOU Lidong, WAN Xiaozhong, et al. Time-sharing and classified prediction model for short-term load considering meteorological factors[J]. Electric Power Construction, 2017, 38(10): 76-83(in Chinese).
- [4] 孙晓燕, 李家钊, 曾博, 等. 基于特征迁移学习的综合能源系统小样本日前电力负荷预测[J]. 控制理论与应用, 2021, 38(1): 63-72.  
SUN Xiaoyan, LI Jiazhao, ZENG Bo, et al. Small-sample

- day-ahead power load forecasting of integrated energy system based on feature transfer learning[J]. *Control Theory and Applications*, 2021, 38(1): 63–72(in Chinese).
- [5] 蔡秋娜, 苏炳洪, 闫斌杰, 等. 基于参数迁移的节假日短期负荷预测方法[J]. *电气自动化*, 2020, 42(4): 59–60, 98. CAI Qiuna, SU Binghong, YAN Binjie, *et al.* Forecasting method for holiday short-term load based on parameter transfer[J]. *Electrical Automation*, 2020, 42(4): 59–60, 98(in Chinese).
- [6] 史佳琪, 谭涛, 郭经, 等. 基于深度结构多任务学习的园区型综合能源系统多元负荷预测[J]. *电网技术*, 2018, 42(3): 698–707. SHI Jiaqi, TAN Tao, GUO Jing, *et al.* Multi-task learning based on deep architecture for various types of load forecasting in regional energy system integration[J]. *Power System Technology*, 2018, 42(3): 698–707(in Chinese).
- [7] 范士雄, 刘幸蔚, 於益军, 等. 基于多源数据和模型融合的超短期母线负荷预测方法[J]. *电网技术*, 2021, 45(1): 243–250. FAN Shixiong, LIU Xingwei, YU Yijun, *et al.* Multi-source data and hybrid neural network based ultra-short-term bus load forecasting[J]. *Power System Technology*, 2021, 45(1): 243–250(in Chinese).
- [8] 马天男, 王超, 彭丽霖, 等. 计及需求响应和深度结构多任务学习的电力系统短期负荷预测[J]. *电测与仪表*, 2019, 56(16): 50–60. MA Tiannan, WANG Chao, PENG Lilin, *et al.* Short-term load forecasting of power system considering demand response and multi-task learning based on deep structure[J]. *Electrical Measurement and Instrumentation*, 2019, 56(16): 50–60(in Chinese).
- [9] 孙庆凯, 王小君, 张义志, 等. 基于LSTM和多任务学习的综合能源系统多元负荷预测[J]. *电力系统自动化*, 2021, 45(5): 63–70. SUN Qingkai, WANG Xiaojun, ZHANG Yizhi, *et al.* Multivariate load prediction of integrated energy system based on long short-term memory and multi-task learning[J]. *Automation of Electric Power Systems*, 2021, 45(5): 63–70(in Chinese).
- [10] LEE E, RHEE W. Individualized short-term electric load forecasting with deep neural network based on transfer learning and meta learning[J]. *IEEE Access*, 2021(9): 15413–15425.
- [11] ZANG H, CHENG L, DING T, *et al.* Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning[J/OL]. *International Journal of Electrical Power & Energy Systems*, 2020, 118. <https://doi.org/10.1016/j.ijepes.2019.105790>.
- [12] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]//*International Conference on Machine Learning*. PMLR, 2017: 1126–1135.
- [13] NICHOL A, ACHIAM J, SCHULMAN J. On first-order meta-learning algorithms[EB/OL]. arXiv preprint arXiv, [2018-10-22]. <https://arxiv.org/abs/1803.02999>.
- [14] RAGHU A, RAGHU M, BENGIO S, *et al.* Rapid learning or feature reuse? towards understanding the effectiveness of maml[EB/OL]. arXiv preprint arXiv, [2020-2-12]. <https://arxiv.org/abs/1909.09157>.
- [15] ANTONIOU A, EDWARDS H, STORKEY A. How to train your maml[EB/OL]. arXiv preprint arXiv, [2019-3-5]. <https://arxiv.org/abs/1810.09502>.
- [16] 王轲, 钟海旺, 余南鹏, 等. 基于seq2seq和Attention机制的居民用户非侵入式负荷分解[J]. *中国电机工程学报*, 2019, 39(1): 75–83, 322. WANG Ke, ZHONG Haiwang, YU Nanpeng, *et al.* Nonintrusive load monitoring based on sequence-to-sequence model with attention mechanism[J]. *Proceedings of the CSEE*, 2019, 39(1): 75–83, 322(in Chinese).
- [17] 路宽, 孟祥荣, 孙雯雪, 等. 多层Bi-GRU的Seq2seq网络短期电力负荷预测模型[C]//2017智能电网新技术发展与应用研讨会. 中国电力科学研究院: 北京市海淀区太极计算机培训中心, 2017: 7.
- [18] 杨楠, 贾俊杰, 邢超, 等. 基于E-Seq2Seq技术的数据驱动型机组组合智能决策方法[J]. *中国电机工程学报*, 2020, 40(23): 7587–7600. YANG Nan, JIA Junjie, XING Chao, *et al.* Data-driven intelligent decision-making method for unit commitment based on e-seq2seq technology[J]. *Proceedings of the CSEE*, 2020, 40(23): 7587–7600(in Chinese).
- [19] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[C]//*Advances in Neural Information Processing Systems*. 2017: 5998–6008.
- [20] TAY Y, BAHRI D, METZLER D, *et al.* Synthesizer: rethinking self-attention in transformer models[C]//*International Conference on Machine Learning*. PMLR, 2021: 10183–10192.

收稿日期: 2021-04-21

作者简介:

丁琦(1995), 男, 硕士研究生, 主要研究方向为深度学习在电力系统中的应用, E-mail: 978248121@sjtu.edu.cn;  
邱才明(1966), 男, 教授, 博士生导师, 从事下一代无线通信基础理论与系统、高维统计理论与应用、大数据与人工智能、深度学习理论与应用等研究, E-mail: caiming@hust.edu.cn;  
杨浩森(1996), 男, 博士研究生, 主要研究方向为大数据在电力系统中的应用, E-mail: 31910019@sjtu.edu.cn;  
童厚杰(1997), 男, 硕士研究生, 主要研究方向为图深度学习在电力系统中的应用, E-mail: thj\_926@sjtu.edu.cn.

## 附录 A

式中:  $N$  为数据的批大小

附表 A1 算法 1: 基于 MAML 策略的数据集生成流程

Table A1 Algorithm 1: The generation procedure for datasets based on MAML strategy

算法	基于MAML策略的数据生成
输入	多任务数据集 $X^{T, N, k_t+1}$ , 综合任务体数 $S$ MAML任务集 $T_0$ , 分割比 $r$ , 任务抽取数 $n_{way}$
输出	MAML数据集 $D$ , 微调数据集 $D'$
1	$D \leftarrow DataLoader(X, n_{way}, T_0)$
2	$D' \leftarrow DataLoader(X, N - n_{way}, T \setminus T_0)$
3	<b>function</b> $Set\_Gen(Data, n_{way}, k_{shot}, k, T)$ :
4	$Support\_Set \leftarrow \emptyset$
5	<b>for</b> $s \in S$ <b>do</b>
6	$a \leftarrow random\_select(T, n_{way})$
7	<b>for</b> $p = 1 \rightarrow n_{way}$ <b>do</b>
8	$b \leftarrow random\_select(k, k_{shot})$
9	$Support\_Set \leftarrow Support\_Set \cup Data[a, b]$
10	<b>end for</b>
11	<b>end for</b>
12	<b>return</b> $Support\_Set$
13	<b>end function</b>
14	
15	<b>function</b> $DataLoader(Data, n_{way}, T)$ :
16	$k_{set} \leftarrow \{k_t \text{ if } t \in T\}$
17	$k_{shot} \leftarrow \min\{k_{set}\} * r$
18	$k_{query} \leftarrow \min\{k_{set}\} * (1 - r)$
19	$Support\_Set \leftarrow Set\_Gen(Data, n_{way}, k_{shot}, k_{set}, T)$
20	$Query\_Set \leftarrow Set\_Gene(Data, n_{way}, k_{query}, k_{set}, T)$
21	<b>return</b> $Support\_Set$
22	<b>end function</b>

附表 A2 算法 2: 基于 MAML 策略的短时回归预测训练流程

Table A2 Algorithm 2: The regression training procedure for MAML based short term prediction

算法	基于MAML策略的短时回归训练
输入	多任务数据集 $X$ , 综合任务体数 $S$ , 训练参数 $\alpha, \beta$
输出	已训练模型参数 $\theta'$
1	初始化网络参数 $\theta$
2	构建MAML和微调集 $D, D' \leftarrow DataLoader(X, S)$
3	<b>for</b> $Support, Query$ <b>in</b> $D$ <b>do</b>
4	<b>for</b> $s \in S$ <b>do</b>
5	使用参数 $\theta$ 对 $Support_s$ 推理
6	计算用于微调阶段的参数
7	$\theta^s = \theta - \alpha \nabla L\{f_{\theta}(Support_s)\}$
8	使用 $\theta^s$ 在 $D'$ 上应用公式(10)执行微调
9	<b>end for</b>
10	使用 $\theta_s$ 在 $Query_i$ 上应用公式(9)执行参数更新
11	$\theta \leftarrow \theta - \alpha \nabla L\{f_{\theta^s}(Query_s)\}$
12	<b>end for</b>
13	采取最优的微调结果作为已训练参数 $\theta'$
14	<b>return</b> $\theta'$