

参赛队号:

2021 年（第七届）全国大学生统计建模大赛

参赛学校: 陕西理工大学

论文题目: 大数据驱动的窃电检测研究
——基于模态分解的 GRU-MLR 和 AR 的优化负
荷预测模型

参赛队员: 乔石、余帆、孟磊

指导老师: 王磊

目录

摘要.....	I
Abstract.....	II
一、 引言.....	1
(一) 研究背景	1
(二) 研究目标	2
(三) 研究的理论基础	2
(四) 研究方法与技术路线	4
(五) 研究的创新点	5
二、 基于模态分解的 GRU 和 MLR 优化负荷预测的窃电模型.....	6
(一) 算法原理	6
(二) PSO-CGM-A 模型.....	12
(三) 窃电检测模型	13
(四) 模型指标	14
三、 数据介绍及数据预处理.....	16
(一) 数据介绍.....	16
(二) 数据预处理.....	16
四、 负荷预测及窃电分析.....	20
(一) 原始数据展示及分解.....	20
(二) 预测结果与分析.....	22
(三) 窃电检测模拟结果.....	24
五、 研究的结论和启示.....	27
(一) 主要结论.....	27
(二) 研究的不足与改进方向.....	27
参考文献.....	29
致谢.....	33

表格清单

表 1	混淆矩阵	15
表 2	特征的 Pearson 相关系数.....	17
表 3	ADF 平稳性分析.....	18
表 4	KPSS 平稳性分析.....	19
表 5	原负荷数据分解后的 IMF 分量	22
表 6	IMF9 和 IMF10 分别在 GRU 和 MLR 中预测结果.....	22
表 7	PSO-CGM-A 模型与单一算法模型对比.....	23
表 8	PSO-CGM-A 模型与组合算法模型对比.....	24
表 9	异常时间点	26

插图清单

图 1	GRU 内部结构.....	9
图 2	粒子群算法流程.....	11
图 3	基于模态分解的优化 GRU-MLR 和 AR 的负荷预测模型.....	12
图 4	异常值检测.....	18
图 5	原始电力负荷数据图.....	20
图 6	原负荷数据分解后的 IMF 分量.....	21
图 7	PSO-CGM-A 模型与单一算法模型对比图.....	23
图 8	PSO-CGM-A 模型与组合算法模型对比.....	24
图 9	生成数据（前 150 个数据）.....	25
图 10	生成数据（后 150 个数据）.....	25
图 11	预测数据（前 150 个数据）.....	26
图 12	预测数据（后 150 个数据）.....	26

摘要

“碳中和、碳达峰”成为我国“十四五规划”污染防治攻坚战的主攻目标。在这一背景下，加快调整优化产业结构、能源结构成为能源行业的发展趋势，其中关键技术就包括提高能源利用率，减少不必要的损失，尤其是非技术损失。

窃电，是指用电用户为了少交甚至不交电费，针对电表做出的违法行为，它是造成电能输、配环节中非技术损失的主要原因。用户的窃电行为不仅会干扰电力调度，给电网带来安全隐患，还会直接损害电力公司的经济利益。

本文立足于这一经济背景，针对用户的异常用电行为，提出了一种以数据驱动的基于自适应白噪声的完整经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)的优化 GRU-MLR 和 AR 负荷预测算法的窃电检测模型。该模型首先采用 EEMD 将数据中的负荷值分解为不同的本征模态函数(Intrinsic Mode Functions, IMF), 每个 IMF 都代表了负荷数据不同的特征, 具有平稳性好, 规律性较强的特点。其次分别使用 GRU 神经网络模型和 MLR 模型对模糊熵(Fuzzy Entropy, FE)值高的部分和模糊熵值低的部分分别进行预测, 然后将所预测的结果重构到的完整的预测结果, 最后用自回归(Autoregression, AR)提取负荷数据中的线性成分得到另一个预测结果, 组合两个预测结果, 用粒子群算法(Particle Swarm Optimization, PSO)进行权重优化得到最终的预测结果。此步骤得到负荷预测结果, 最后使用滑动时间窗口对预测负荷和真实负荷进行阈值比较, 判断用电异常点, 得出用户的窃电行为。

本文使用麻省理工公寓负荷公开数据集对该模型进行仿真验证, 结果表明, 该模型的窃电检测召回率达 89.72%, 准确率达到 96%, 证明该模型具有优良的检测性能, 对窃电检测领域的研究具有一定的价值。

关键词: 负荷预测; 窃电检测; EEMD; 模糊熵; GRU 神经网络; 多元线性回归

Abstract

"Carbon neutralization, carbon peaking" has become the main target of pollution control in "the 14th five years plan". In this context, accelerating the adjustment and optimization of industrial structure and energy structure has become the development trend of the energy industry, among which the key technologies include improving energy utilization rate and reducing unnecessary losses, especially non-technical losses.

Stealing electricity refers to the illegal behavior of electricity users aiming at electricity meters in order to pay less or even no electricity charges. It is the main cause of non-technical losses in power transmission and distribution. The electricity stealing behavior of users will not only disturb the power dispatching, bring hidden danger to the power grid, but also directly damage the economic interests of power companies.

Based on this economic background, this paper proposes a data-driven electricity theft detection model based on the optimized GRU-MLR and AR load forecasting algorithm of ensemble empirical mode decomposition (EEMD). Firstly, the load value in the data is decomposed into different intrinsic mode functions (IMF) by EEMD. Each IMF represents different characteristics of the load data, which has the characteristics of good stability and strong regularity. Secondly, GRU neural network model and MLR model are used to predict the part with high fuzzy entropy (FE) value and the part with low fuzzy entropy value respectively, and then the predicted results are reconstructed into the complete prediction results. Finally, autoregression (AR) is used to extract the linear components in the load data to obtain another prediction result and combine the two prediction results, Particle swarm optimization (PSO) is

used to optimize the weight to get the final forecasting results. In this paper.

In this paper, the sliding time window is used to compare the threshold value between the predicted load and the real load to judge the abnormal point of power consumption and get the electricity stealing behavior of users. Finally, the MIT open data set is used to simulate and verify the model. The results show that the recall rate of the model is 89.72%, and the accuracy rate is 96%, which proves that the model has excellent detection performance, and has a certain value for the research in the field of power theft detection.

Keywords: Load forecasting; Electricity stealing detection; EEMD; Fuzzy entropy; GRU neural network; Multiple linear regression

一、引言

（一）研究背景

据统计数据显示，因为用户的窃电行为，我国电力公司每年存在着巨大经济损失^[1]，用户的不正当窃电行为不仅会侵犯电力公司的经济利益，还会给电力公司提供虚假的用户用电数据，这将会影响到电网公司的电力调度，私自接线还会带来安全问题，甚至还会影响电网的安全^[2-4]。如何对窃电问题提出有效的解决措施，已经成为全球电网公司不得不面对的问题。因此，对用户电力负荷进行合理预测，用真实数据生成预测数据，有效甄别终端用户的窃电行为，成为全世界智能电网发展的一个重点^[5-6]。

传统方式进行异常用电检测主要是工作人员利用电表箱数据^[7]，现场记录电压、电流和电量。然后，检测人员根据一段时期采集的用户用电数据，结合以往经验综合判断该客户是否处于异常用电状态或者存在窃电行为^[8]。最后，实地考察异常用户用电情况，及时发现、制止该用电用户的不正当窃电行为。传统方式下的人工窃电检测手段存在诸多弊端，需要投入巨大的人力物力财力^[9]，检测用时过长，效率低下，准确度不高，无法及时止损，这都是窃电行为屡禁不止的重要因素，并且严重制约我国电力行业的快速发展。随着我国智能电网的全面建成，电网朝着数字化，信息化，自动化方向发展^[10]，智能电表的投入应用、在线监测系统的建成，以及电力信息管理不断系统化，都为快速采集用户用电数据、及时发现异常用电提供了技术保障^[11]。海量的数据为大数据分析电力负荷的各种特征提供了一个前提，近年来发展出了很多相关的研究。同时，随着人工智能技术的迅猛发展，利用机器学习可以对采集到的庞大电力数据进行处理^[12]，智能预测用户用电负荷，检测异常用户用电行为，具有效率高、准确度高得优势，为判别异常用电行为提供了新的手段，检测人员可以根据检测结果，实地考察，及时判断并制止用户窃电行为^[12]。

（二）研究目标

电能是我们国家的重要资源之一。在“十四五规划”中，我国全新提出了“碳中和”目标，我国电力企业和供电部门正在进行能源转型，其中包括了提高能源的利用率，同时减少能源的损耗，这其中就包括窃电损失。

传统的窃电检测方法会消耗大量的人力物力，同时反窃电效果有限，随着用电用户数量的激增，传统窃电检测方法的弊端越来越明显^[13]。随着电力行业大数据时代的到来，我们可以从海量的用户用电数据中捕获我们需要的信息，例如用户用电习惯、用户信用指数、台区线损指标等。但是同时，要处理如此庞大的数据成为一个新的挑战。近几年，计算机算力的不断提高使得我们拥有了可以大量处理数据的能力，国内外学者提出了各类窃电检测模型，目的是把检测投入降至最低，同时提高检测效率。

因此，本文的窃电检测研究也主要从这两个方面着手。一方面，我们要不断提高窃电检测效率和精确度，使得更多的窃电用户被检测出来，同时尽量降低误检率；一方面，我们还要控制检测成本，确保窃电检测可以为电力公司带来正收益。

（三）研究的理论基础

由于电力负荷具有一定的季节性，同时影响负荷的因素也很多，如：天气、温度、气压、节假日等^[14]，电力负荷数据常包含不平稳性和非周期性成分存在。目前负荷预测从方法来讲可以分为两类：数学统计预测模型和机器学习预测模型。数学统计模型有自回归、多元线性回归等^[15]，其优点是模型简单，预测速度较快，其缺点是模型的鲁棒性差，只能预测波动性小、周期性强的时间序列。如文献[16]针对数据波动性小，影响因素少的特点使用了多元线性回归进行预测。但是，面对复杂度高、平稳性差、随机性强的负荷数据，机器学习预测模型的优势就比较明显。在机器学习模型中，支持向量回归(Support Vector Regression, SVR)^[17]，

具有良好预测精度，同时还可以快速获得全局解^[18]。然而，它对于大型数据集的可伸缩性是有限的。为了解决这一问题，神经网络因其强大的多变量映射能力而被应用于负荷预测。但传统的反向传播神经网络容易陷入局部最优解，没有充分考虑时间序列的内部影响。递归神经网络(Recurrent Neural Network, RNN)解决了局部最优解问题^[19]，它可以通过隐藏层之间的自连接来把时间的特征紧密联系起来。但 RNN 的梯度爆炸和梯度消失问题限制了其适用性^[20]。在文献[21]通过在 RNN 中添加一个特殊的单元结构，实现了一种缓解这一问题的长短时记忆(Long Short Time Memory, LSTM)神经网络^[22]。但 LSTM 存在收敛较慢的问题，GRU 是 LSTM 最流行的一种变体，它比 LSTM 有更快的收敛速度，使得 GRU 在时间序列预测中更加适用^[23-24]。

虽然机器学习预测模型在某些方面优于数学统计预测模型，但其预测的准确性容易受到数据的影响，低质量的电力数据往往达不到想要的预测结果。两种预测模型都有其各自的应用场景，可以发挥它们特有的优势进行有效预测，但是单一的模型在电力负荷预测中无法弱化的非线性负荷的时间序列，受到这方面的局限，就不能得到准确的预测值。为解决单一模型预测准确度不高的问题，有人提出了混合预测模型^[25]，一般可以有效解决电力负荷非平稳时间序列的预测。例如，如文献[26]提出了经验模态分解(Empirical Mode Decomposition EMD)负荷数据降低信号复杂度使更加平稳，随后使用 AR 模型进行预测，得到了较好的效果。此外，为消除 EMD 存在的模态混叠情况，文献[27]提出了基于集成经验模态分解和支持向量机(Ensemble Empirical Mode Decomposition-SVM, EEMD-SVM)模型^[28]的短期负荷预测模型，该模型让分解后的信号频率均匀分布^[28]，解决了单一模型预测性能差的问题，体现了组合模型的优越性。但是 EEMD 也存在明显的缺点，模态分解耗时长，由于加入了白噪声，使得重构误差变大。同时，通过模态分解后只用平稳性强的分量进行负荷预测，忽视了其他分量对预测结果的

影响，得到的预测结果就不够准确。

文献[29]通过把长短期记忆网络（Long Short-Term Memory, LSTM）应用于序列特征的提取，通过全连接网络对序列特征进行分类，判断了异常用户。文献[30]中使用基于注意力机制的卷积神经网络-长短期记忆模型（LSTM-CNN）进行异常检测，该模型符合工业异常检测的时效性，能够快速检测边缘设备的故障或关闭。文献[31]中 GRU-Bagging 模型对异常用电进行检测，并通过实验证明该模型在不平衡数据集上表现良好，预测准确度达到 86.17%，召回率更是达到 95.34%。

尽管循环神经网络、长短期记忆网络、卷积神经网络（Convolutional Neural Networks, CNN）^[32]等网络模型在用户窃电检测领域中取得了不错的效果，这一类神经网络固有的顺序模式导致了训练样本时不能并行化的问题。同时，对于长序列，内存的限制更是阻碍了对训练样本的批量处理。当序列长度过长时，长短期记忆网络还会遇到训练时间长、参数更新慢，甚至梯度消失等问题，这都阻碍了对窃电检测领域的研究。Google 提出的 Transformer 模型可以很好地解决上述问题^[32]。在时间序列问题上，文献[33]提出了基于 Transformer 的 Informer 模型，它能够高效地捕捉输出和输入之间精确的长程相关性耦合，对窃电检测研究有着重要意义。

（四）研究方法与技术路线

窃电检测方法总体上分为三类：基于系统状态、基于博弈论和基于数据驱动的窃电检测^[7]，三类检测方法各自有各自的优缺点。随着计算机储存、计算能力、算法模型等方面的快速发展，让大数据研究窃电成为可能，基于数据驱动的方法成为大势所趋。智能电表的普及和 AMI 和应用，海量的用电数据时时刻刻从用户终端、电网节点反馈回电力公司，为数据驱动的窃电检测方法提供了可能性。不仅如此，社会信息、天气情况、经济走向等同样会对用电造成影响的非电力数据，也可以用于对用电用户的用电行为中去。从近几年的文献中，我们可以看出，

基于数据驱动的窃电检测模型一般可以分为三类：基于分类、基于聚类以及基于回归。

本文基于前文方法，提出了一种基于模态分解的优化 GRU-MLR 和 AR 短期负荷预测的窃电检测方法。该方法首先通过 EEMD 将负荷数进行分解，得到复杂度低含有不同信息量的 IMF 数据。随后计算每个 IMF 的模糊熵，对模糊熵值高的分量进行 GRU 深度学习模型预测，对模糊熵值低的分量进行 MLR 模型预测，然后把所得的结果进行重构得到完整的预测值，完成第一部分的预测。另一部分充分考虑负荷数据线性特征，采用 AR 模型进行预测，将模态分解 GRU-MLR 模型预测的结果与 AR 模型预测的结果使用粒子群算法进行权重优化，得到最终的预测模型。最后利用阈值判断，对用电用户进行评估。

（五）研究的创新点

本文提出了一种适合大数据环境下的基于模态分解的优化 GRU-MLR 和 AR 负荷预测的窃电检测模型，该模型使用基于回归的窃电检测方法。在研究对象的方法上进行了两点创新：

1. 使用组合预测算法，首先使用 EEMD 算法，对原始的负荷数据进行模态分解，得到若干个包含不同信息的 IMF 和残差；随后计算各 IMF 的模糊熵，根据 GRU 神经网络模型和 MLR 模型各自的性质，将高熵值的 IMF 输入 GRU 神经网络进行预测，其余分量输入 MLR 模型进行预测。将预测的结果重构得到结果，同时使用 AR 模型进行预测，得到预测结果。

2. 我们使用滑动检测窗口对预测负荷曲线和真实负荷曲线对比进行窃电检测，将低于设定阈值的区间和点标记为异常用电区间和点，从而判断该用户是否为窃电用户。

二、基于模态分解的 GRU 和 MLR 优化负荷预测的窃电模型

(一) 算法原理

1. 集合经验模态分解^[34]

经验模态分解(EMD)算法是一种通过本身的数据进行的非线性、非平稳型号分解方法^[26]，具有自适应能力，与小波信号分解相比，不需要选择小波基，解决了小波信号分解在选择小波基的时候不好确定的问题。通过 EMD 分解数据，可以获得含有不同数据特征的分量本征模态函数(IMF)。EMD 的自适应能力强，但是分解后存在的模态混叠现象将影响预测的准确度。

集合经验模态分解(EEMD)算法是 EMD 算法的改进算法^[34]，对原始数据加入白噪声，使得数据在极值点间的分布变均匀，从而使分解之后的模态混叠现象降低。

EEMD 算法具体如下：

①先通过 EMD 算法对信号 $X(t) + \varepsilon_0 n^i(t)$ 进行 N 次分解，然后计算其平均值，可以计算出第一个模态 IMF_1 ：

$$\overline{IMF_1(t)} = \frac{1}{N} \sum_{i=1}^N IMF_1^i(t) \quad (1)$$

②计算 IMF_1 的残余信号 $r_1(t)$ ：

$$r_1(t) = X(t) - \overline{IMF_1(t)} \quad (2)$$

③ 通过 EMD 算法对 $r_1(t) + \varepsilon_1 E_1(n^i(t))$ 进行 N 次分解，计算出第二个模态 IMF_2 ：

$$\overline{IMF_2(t)} = \frac{1}{N} \sum_{i=1}^N E_1(\varepsilon_1 E_1(n^i(t)) + r_1(t)) \quad (3)$$

④ 对于 $k = 2, \dots, K$ ，计算第 k 个残余信号：

$$r_k(t) = r_{k-1}(t) - \overline{IMF_k(t)} \quad (4)$$

⑤ 重复步骤③的计算，得到 $k+1$ 个模态 IMF_{k+1}

$$\overline{IMF_{k+1}(t)} = \frac{1}{N} \sum_{i=1}^N E_1(\varepsilon_k E_k(n^i(t)) + r_k(t)) \quad (5)$$

⑥ 重复步骤④和步骤⑤，直到分解完成，最终得到 K 个模态分量：

残差为：

$$R(t) = X(t) - \sum_{k=1}^K \overline{IMF_k(t)} \quad (6)$$

原始信号被分解为：

$$X(t) = R(t) + \sum_{i=1}^K \overline{IMF_k(t)} \quad (7)$$

2. 模糊熵

模糊熵的是用来表示数据复杂程度的一种方法，模糊熵值的大小，直观的体现了时间序列数据的复杂程度^[35]。

模糊熵方法描述：

① 对于时间序列 $X = [x(1), x(2), \dots, x(N)]$ 设维度为 m ，根据原始时间序列构造 m 维向量为：

$$X_m(i) = [x(i), x(i+1), \dots, x(i+m+1)] - u(i) \quad (8)$$

式中： $i = 1, 2, \dots, N - m + 1$ ， $u(i)$ 为

$$u(i) = \frac{1}{m} \sum_{j=0}^{m-1} x(i+j) \quad (9)$$

② 定义向量 $X_m(i)$ 和 $X_m(j)$ 之间的距离 d_m^{ij} 其对应值差的绝对值最大值，即

$$d_m^{ij} = \max_{k=1, \dots, m-1} (|x(i+k) - u(i) - (x(j+k) - u(j))|) \quad (10)$$

式中： $i = 1, 2, \dots, N - m + 1$ ， 且 $i \neq j$ 。

③ 模糊度函数：

$$D_{ij}^m = \mu(d_{ij}^m, n, r) = e^{-\left(\frac{d_{ij}^m}{r}\right)^n} \quad (11)$$

$$A(x) = \begin{cases} 1, (x=0) \\ \exp[-\ln(2)\left(\frac{x}{r}\right)^2], (x>0) \end{cases} \quad (12)$$

式中： n 为边界梯度， r 为边界的宽度。

④ 定义函数

$$C^m(r) = \frac{1}{N-m} \cdot \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \cdot \sum_{j=1, j \neq i}^{N-m} D_{ij}^m \right) \quad (13)$$

$$C^{m+1}(r) = \frac{1}{N-m} \cdot \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \cdot \sum_{j=1, j \neq i}^{N-m} D_{ij}^{m+1} \right) \quad (14)$$

模糊熵公式表示为：

$$FuzzyEn(x, m, n, r) = -\ln \frac{C^{m+1}(r)}{C^m(r)} \quad (15)$$

3. GRU 神经网络

传统的神经网络不能做到对时间长久记忆，RNN 中引入了循环单元，使得数据间的时间连续性关系被神经网络学习。但是 RNN 存在明显的梯度消失的缺点，时序过长会使 RNN 丢失对初始的部分信息。为了解决这个问题，LSTM 神经网络被提出，它可以连接之前的信息并作用到当前的处理中，被广泛应用于时间序列预测中。LSTM 经过长时间发展，人们提出了很多变体，其中最流行的变体就是 GRU 神经网络。

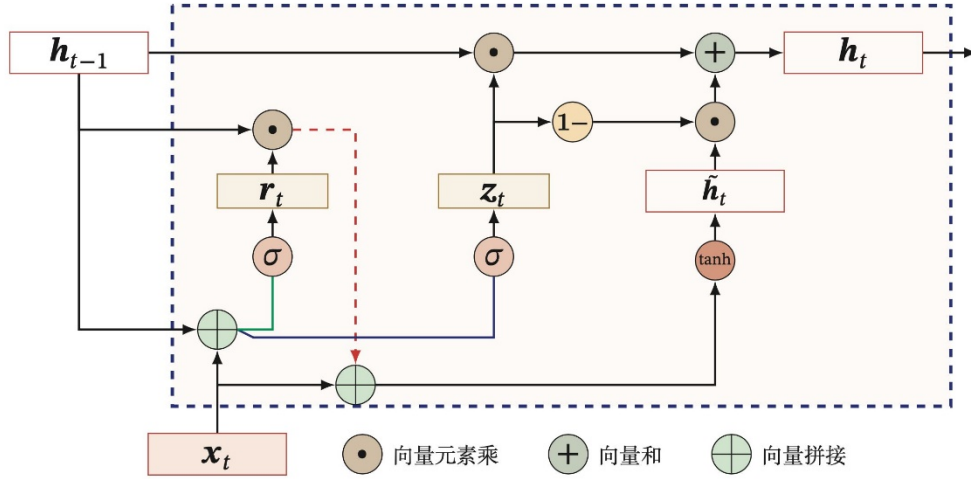


图 1 GRU 内部结构

GRU 采用门控循环神经网络结构，它是由 RNN 和 LSTM 发展而来的，保留了 RNN 的输入和输出，应用了 LSTM 的思想。把原来 LSTM 中的输入门和遗忘门合并为一个更新门，所以 GRU 结构中只有更新门和重置门。更新门控制前一时刻信息能够传递到当前状态的程度。重置门用于选择是否利用之前保留的信息，重置门的值表示对信息的保留程度，值越小遗弃的信息越多^[24]。

公式如下：

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \quad (16)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \quad (17)$$

$$\tilde{h}_t = \tanh(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h) \quad (18)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \quad (19)$$

式中 z_t 和 r_t 分别为更新门和重置门， σ 为 Sigmoid 函数； \tanh 为双曲正切函数； U_z ， U_r ， W_r ， W_z 为训练参数矩阵； \tilde{h}_t 为输入 x_t 和上一隐藏层输出结果 h_{t-1} 的组合^[24]。

4. 多元线性回归

多元线性回归(MLR)是对于简单数据快速预测的模型，它可以输入多变量的数据进行预测，通过找出一个因变量和其他所有自变量之间的关系。在预测复杂

度低的数据时，其训练速度快，训练准确度高的特点表现的很明显。因此本文采用对模糊熵值低的 IMF 分量进行 MLR 预测比其它方法更加准确快捷。

MLR 模型为：

$$Y = X \times \beta + \mu \quad (20)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1n} \\ 1 & x_{21} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nn} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \quad (21)$$

其中： y_i 表示数据中的负荷值； x_{ij} 表示影响负荷的因素值，我们在下文统称为特征值； β_i 表示回归方程的系数； μ_i 表示随机误差。

损失函数定义为：

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (22)$$

5. 自回归模型

自回归(Autoregressive, AR)模型是使用自身作为变量做回归，例如 x_1 到 x_{t-1} 来预测 x_t ，此方法是有线性回归发展而来。自回归模型能充分考虑伏负荷数据的线性特征，其公式为^[14]：

$$X_t = \sum_{i=1}^p \alpha_i x_{t-i} + \beta_i \quad (23)$$

式中 α_i 为常系数， β_i 为随机误差。

6. 粒子群算法

粒子群算法 (PSO) 属于进化算法的一种^[37]。粒子群算法是把求解最优解的问题比喻为鸟类寻食的过程，是一种仿生类算法，通常用来解决单目标最优求解和多目标最优求解，算法结构如图 2：

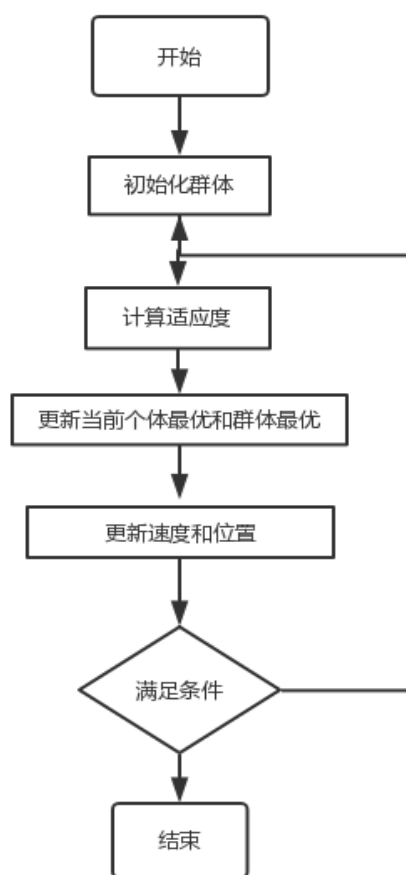


图 2 粒子群算法流程

粒子群算法中，每一个粒子都是一个可行解，每一个粒子在可行解的空间中进行各自的运动，约束条件的设置会用来评价粒子的优劣性。每一次更新中，粒子都会更新自己的速度和位置，更新之后会产生粒子的两个极值^[36]。粒子会依照当前得到的最优解，进行设定的重复检索次数，最终得到最优解。

粒子的位置:

$$z_{it} = [z_{i1t}, z_{i2t}, \dots, z_{idt}]^T \quad (24)$$

粒子的速度:

$$v_{it} = [v_{i1t}, v_{i2t}, \dots, v_{idt}]^T \quad (25)$$

个体最优:

$$p_{it} = [p_{i1t}, p_{i2t}, \dots, p_{idt}]^T \quad (26)$$

群体最优:

$$p_{gt} = [p_{g1t}, p_{g2t}, \dots, p_{gdt}]^T \quad (27)$$

粒子拥有惯性这一特点。他会按照自己的最优状态调整，也会按照群体的最优调整。

所以 $t+1$ 时刻，粒子更新当前速度和位置，用公式表示如下表示

$$v_{id}(t+1) = \alpha v_{idt} + c_1 r_1 (p_{idt} - z_{idt}) + c_2 r_2 (p_{gdt} - z_{idt}) \quad (28)$$

$$z_{id}(t+1) = z_{idt} + v_{id}(t+1) \quad (29)$$

式中， v_{id} 为第 i 个粒子的速度； c 为加速因子； $d=1,2, \dots, D$ ； k 为迭代次数。

(二) PSO-CGM-A 模型

本文将数据中的负荷值进行模态分解预处理之后分别不同模糊熵值的 IMF 进行预测，最后总体考虑数据的线性特征进行结果优化，提出了基于模态分解的优化 GRU-MLR 和 AR 短期负荷预测模型，本文将该负荷预测模型简称 PSO-CGM-A 模型，模型结构图如图 3 所示。

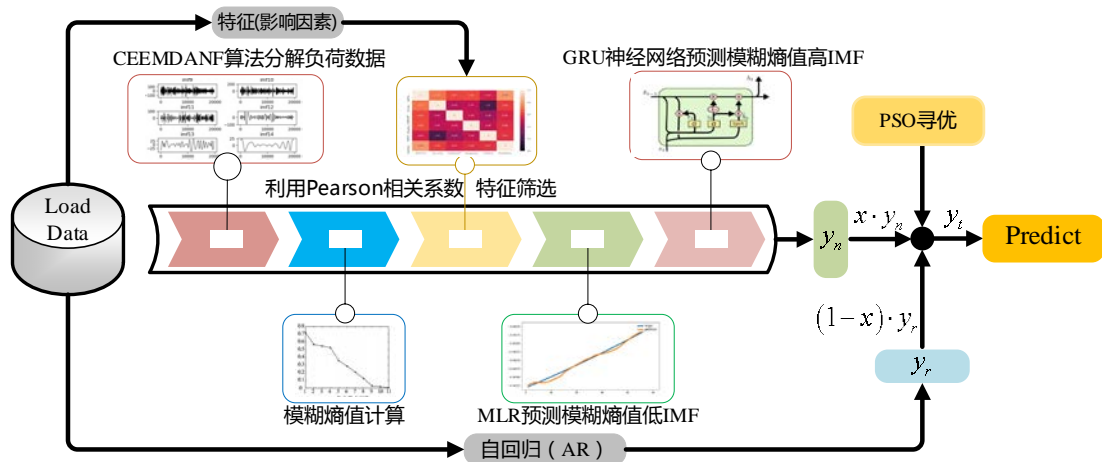


图 3 基于模态分解的优化 GRU-MLR 和 AR 的负荷预测模型

预测部分的数据分为电力负荷数据和特征数据，电力负荷数据通过 EEMD 进行分解，得到若干个 IMF 和残差；随后计算各 IMF 的模糊熵；特征数据通过

Pearson 相关系数进行特征筛选，将相关性较弱的特征排除，保留相关性较强的数据这样可以提高特征的权重，提升预测的精度。

随后计算各 IMF 的模糊熵，根据 GRU 神经网络模型和 MLR 模型各自的性质，GRU 神经网络对复杂度高的数据有很好的预测精度，所以将高熵值的 IMF 输入 GRU 神经网络进行预测，MLR 对线性分量高的部分预测迅速且准确，我们将熵值低的部分输入 MLR 模型进行预测。将预测的结果重构得到结果 y_n ，同时使用 AR 模型进行预测，得到预测结果 y_r 。利用粒子群算对公式(30)寻优，得到最优 x 是的预测结果更加准确。

$$y_t = x \cdot y_n + (1-x) y_r \quad (30)$$

(三) 窃电检测模型

基于大数据的用电异常点通常采用的方法是，以用户的历史用电数据作为翻负荷数据集进行负荷预测，得出的结果与真实负荷进行比对，最后得到异常用电点。这种方法使得在每一个采样点，可以的一个得到比值，当这个值大于我们设定的阈值，我们认为这个采样点是异常点，这在以往模型中的确取得了不错的成绩，公式如下：

$$V_t = \frac{x_t^{true}}{x_t^{pred}} \quad (31)$$

其中， V_t 代表负荷阈值， x_t^{true} 是真实用电数据， x_t^{pred} 是预测用电数据。

但是，用户个体的突发性用电行为，会导致这种单一化的检测模型发生误判。例如，历史用电数据中户主六点下班，下班之后的用户用电行为会增加用电量，但是由于突发性加班，用户用电行为推迟了一个小时，这种情况下，就很容易引发误判。针对时间序列不齐的情况，常见的办法有序列对齐、序列滞后等办法，但是这无疑会大大增加检测成本。因此，本文使用了一种滑动窗口阈值判断，先用滑动时间窗口对一个区间内的用电量进行阈值判断，再从中找出异常起点，判断公式如下：

$$Z_t^w = \frac{\sum_{t=i}^{i+w} x_t^{pred}}{\sum_{t=i}^{i+w} x_t^{true}} \quad (32)$$

其中, Z_t^w 代表在步长为 w 的滑动窗口下真实的负荷阈值, x_t^{true} 是真实用电数据, x_t^{pred} 是预测用电数据。

本文使用的数据集是正常用户的用电数据, 需要根据窃电模式对数据进行转换, 模拟生成异常用户的用电数据集。用户的异常用电行为不是随机的, 其最终目的是减少需要支付的电费, 因此其异常行为有一定的规律。一个偷电用户的典型用电曲线是长时间连续用少量电, 曲线稳定, 即具有连续、少量、稳定的特点。基于窃电用户的用电曲线特征, 以及存在 6 种典型窃电模式的事实^[37,38], 建立了典型窃电模式下的异常用户数据集。

典型的 6 种窃电模式如下所示:

$$\text{模式 1: } P'_{i,j} = \alpha \cdot P_{i,j}, \{0 < \alpha < 1\} \quad (33)$$

$$\text{模式 2: } P'_{i,j} = \begin{cases} P_{i,j} & P_{i,j} \leq P \\ P & P_{i,j} > P \end{cases} \quad (34)$$

$$\text{模式 3: } P'_{i,j} = \max(P_{i,j} - P, 0) \quad (35)$$

$$\text{模式 4: } P'_{i,j} = f(t) \cdot P_{i,j}, \quad f(t) = \begin{cases} 0 & t \in [m, n] \\ 1 & t \in (i, m) \cup (n, j) \end{cases} \quad (36)$$

$$\text{模式 5: } P'_{i,j} = P_{i,j} \cdot \alpha(t), \{0 < \alpha(t) < 1, i \leq t \leq j\} \quad (37)$$

$$\text{模式 6: } P'_{i,j} = \bar{P} \cdot \alpha(t), \{0 < \alpha(t) < 1, i \leq t \leq j\} \quad (38)$$

其中, $P_{i,j}$ 和 $P'_{i,j}$ 是在 i 到 j 时刻中, 居民正常用电负荷和模拟生成的异常用电负荷, \bar{P} 是正常用电前一个月的均值, 在模式 2 和模式 3 中, $\min(P_{i,j}) < P < \max(P_{i,j})$, 在模式 4 中, $i < m < n < j$ 。在模式 5 和模式 6 中, $0 < \alpha(t) < 1$, $i \leq t \leq j$ 。

(四) 模型指标

当我们使用窃电检测模型对用户进行异常用电判断之后, 还需要对模型的准确度进行评估。一方面, 我们希望窃电用户尽可能地被检测出来; 另一方面, 我们还希望正常用户尽可能地被检测为窃电用户。一般来说, 我们使用混淆矩

阵对窃电检测方法优劣进行判断，如表 1 所示。

表 1 混淆矩阵

用户	检测为窃电用户	检测为正常用户
实际为窃电用户	TP(true positive)	FN(false negative)
实际为正常用户	FP(false positive)	TN(true negative)

混淆矩阵将所有用户分为四类，即 TP、FN、FP、TN，其中 TP 和 TN 为分类正确的部分^[12]，这部分占比越高，则表明模型效果越好。同时，我们还定义召回率和准确率：

$$R = \frac{TP}{TP+FN} \quad (39)$$

$$P = \frac{TP}{TP+FP} \quad (40)$$

根据公式我们可以看出，召回率和准确率越高，模型检测效果越好。除了召回率和准确率^[8]，我们通常还会使用 ACC、F1 等指标对模型优劣进行判定。

三、数据介绍及数据预处理

（一）数据介绍

本文采用数据来源于 Umass Smart Dataset 麻省理工数据集，其中有一组公寓数据集包含 2014-2016 年期间 114 个单户公寓的数据，我们选取了 2016 年第一个单户公寓的数据。数据包含两个文件，其一为：公寓用电量数据，采集时间为每分钟采集一次，一共有 503775 条负荷数据，其二为：2016 年全年天气数据，包括温度、湿度、体感温度、风速等 14 个天气因素，采样时间为 1 小时采集一次。

（二）数据预处理

1. 数据初步处理

对原始数据使用 python 分析，发现存在四个方面问题：

① 电力负荷数据采样频率与天气数据记录频次不相同，负荷数据的记录频次为每分钟记录一次，天气数据为每小时记录一次，电力负荷数据中每 60 个数据对应一个天气数据；

② 天气数据中的 windSpeed(风速)和 cloudCover(云层覆盖)分别有 1 个和 979 个缺失值；

③ 天气数据中 icon 和 summary 数据为文字，与分析数据是无法识别文字之间的相关性；

④ 天气数据中包含 14 个因素，但是有些因素对电力负荷的影响很小，在预测时相当于噪声数据，可能会降低预测精度。

针对以上四个方面问题，文本对原始数据进行了以下处理：

① 合并操作：将电力负荷数据中每 60 分钟合并求和，得到采样时间为 1 小时的电力负荷数据，再与天气数据以时间为标度合并为一个整理数据集；

② 缺失值处理：windSpeed 数据分析可知其近似服从正态分布，所以其缺

失值问题通过线性插值法进行填补, cloudCover 数据分析可知缺失值为云层厚度为 0 时没有记录, 所以我们将所有的缺失值填补为 0;

③ 数字替换文字处理: 天气数据中 icon 和 summary 中所有的天气因素类别找出, 分别为其命名为不同的数字;

④ 特征筛选: 通过 Pearson 相关系数分析如表 2, 明显看出有些特征相关性很差, 通过相关性的绝对值大小进行排序, 我们选取前四个作为本文的数据特征, 分别为 windSpeed、dewPoint、temperature、apparentTemperature。

表 2 特征的 Pearson 相关系数

相关特征	Pearson 相关系数
windSpeed	0.141975
windBearing	0.062864
pressure	0.028376
summary	0.027725
cloudCover	-0.027369
visibility	-0.052629
icon	-0.058020
humidity	-0.059597
dewPoint	-0.621811
temperature	-0.639391
apparentTemperature	-0.649639

2. 异常值分析

通过箱盒图对 windSpeed、dewPoint、temperature、apparentTemperature 四个特征进行异常值分析, 如图 4 所示, 我们选取的四个特征都分部都比较合理, 大部分集中在中间, 小部分集中在两边, 没有点存在特殊的分布.

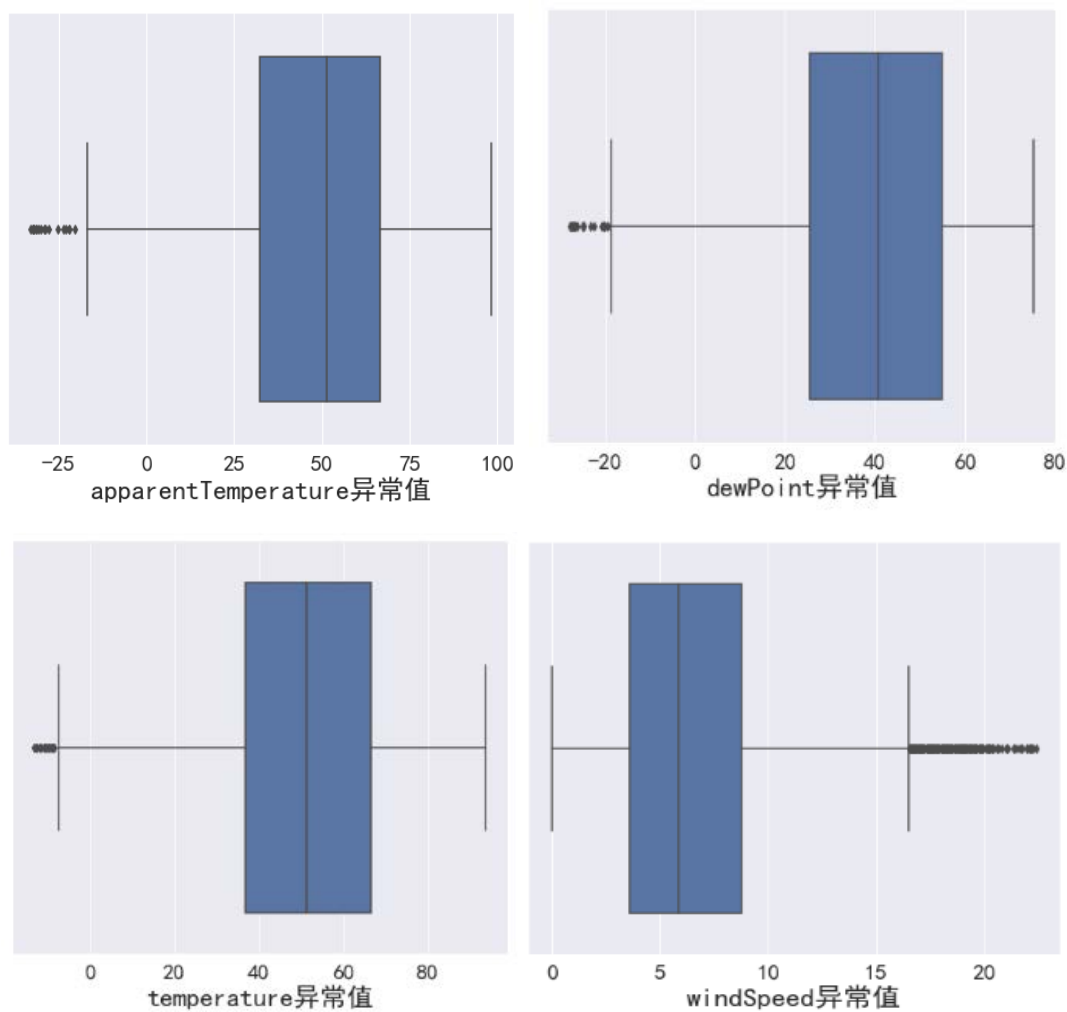


图 4 异常值检测

3. 平稳性分析:

本文采用 Augmented Dickey-Fuller (ADF) 检验和 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) 检验进行平稳性分析, 分析结果如下表 3 和表 4;

表 3 ADF 平稳性分析

ADF Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)
-4.6056	0.00012	-3.4311	-2.8619	-2.5669

ADF 统计量 (-4.6056) 小于临界值 1% (-3.4311), 因此, 我们可以说, 我

们拒绝了显著性水平为 1%的零假设 H_0 , 这意味着时间序列中没有根单位, 因此, 可以认定他是平稳的。

表 4 KPSS 平稳性分析

KPSS Statistic	p-value	Critical	Critical	Critical	Critical
		Value(1%)	Value(2.5%)	Value(5%)	Value(10%)
8.7971	0.0100	0.3470	0.4630	0.5740	0.7390

KPSS 统计量 (8.7971) 高于临界值 1% (0.3470), 因此, 我们不能拒绝显著性水平为 1%的零假设 H_0 , 这意味着时间序列是平稳的或在常数周围是平稳的。

两个测试的结论都是电力负荷时间序列是平稳的, 平稳是时间序列进行时序预测才存在意义, 如不平稳预测结果为伪预测, 所以本文接下来的工作在负荷数据平稳前提下才有实际的意义。

四、负荷预测及窃电分析

(一) 原始数据展示及分解

图 5 为电力负荷的原始数据，可以看出该数据含有大量噪声，而且周期性不明显。

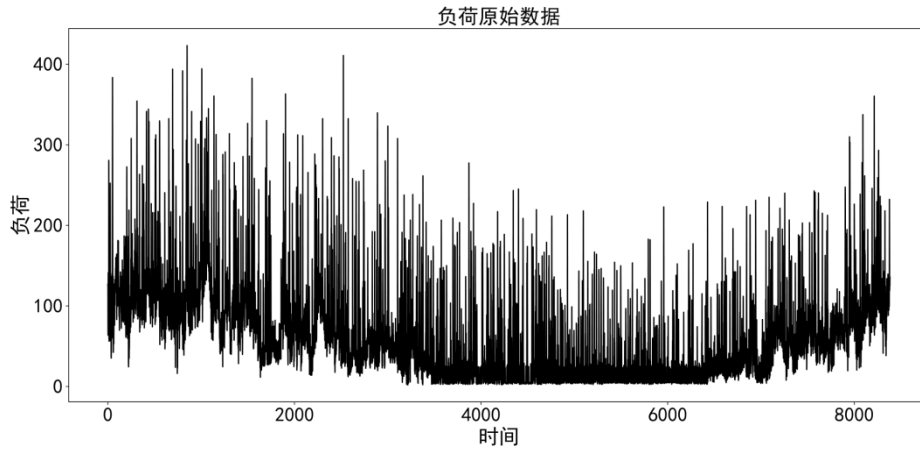


图 5 原始电力负荷数据图

本文对模型的评价采用平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 和均方根误差 (Root Mean Square Error, RMSE) 作为评价指标，公式如下

平均绝对百分比误差 $MAPE$ 公式

$$MAPE = \frac{1}{N} \sum_{T=0}^N \frac{|P_{real} - P_{forecast}|}{P_{real}} \times 100 \quad (41)$$

均方根误差 $RMSE$ 公式:

$$RMSE = \sqrt{\frac{\sum_{T=0}^N (P_{real} - P_{forecast})^2}{N}} \quad (42)$$

数据归一化及 EEMD 分解:

为了使模型能够忽略量纲对数据的影响，本文利用归一化的方法将数据处理，把整体数据压缩到[0, 1]之间，归一化公式如下:

$$y^* = \frac{y - y_{\min}}{y_{\max} - y_{\min}} \quad (43)$$

式中 y^* 为归一化之后的数据, y 为需要归一化的数据, y_{\min} 为最小负荷数据, y_{\max} 为最大负荷数据。

将归一化之后的负荷数据经过 EEMD 分解得到 11 个分量和一个残差, 为了统计方便, 分别记为 IMF1-IMF12。由图 6 直观分析没有出现模态混叠现象, 且具有一定的周期性。

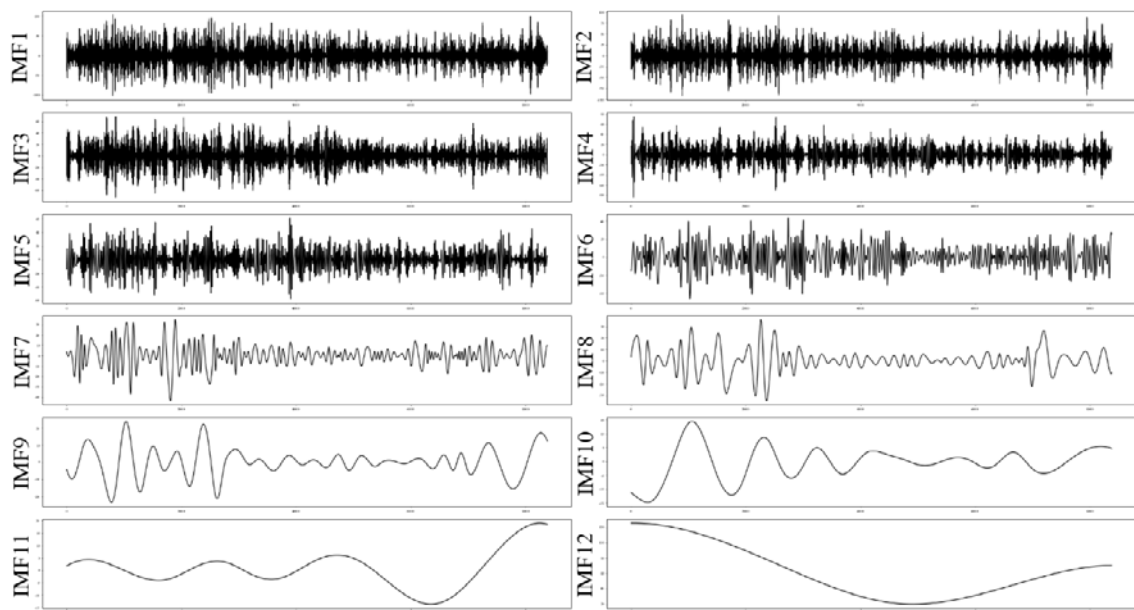


图 6 原负荷数据分解后的 IMF 分量

随后计算各 IMF 分量的模糊熵值, 模糊熵值的大小表示了 IMF 分量包含信息量的多少, IMF1—IMF6 的熵值过大, 计算结果为 \inf (无穷大), 我们将 IMF7—IMF12 的模糊熵值在表 5 中展示, 可以明显看出 IMF7—IMF9 分量中模糊值较大, IMF10—IMF12 分量熵值较小, 且很相近, 通过测试 IMF9 和 IMF10 分量各自在 GRU 神经网络和 MLR 中预测的准确度, 见表 6, 我们可以看出 GRU 神经网络和 MLR 都可以很好的预测 IMF9, 但是 GRU 神经网络在预测 IMF10 时出现了过拟合现象, MLR 可以更好的预测 IMF10, 所以我们用 GRU 神经网络预测 IMF1—IMF9 分量, 用 MLR 预测 IMF10—IMF12 分量。

表 5 原负荷数据分解后的 IMF 分量

IMF	IMF1-6	IMF7	IMF8	IMF9	IMF10	IMF11	IMF12
模糊熵值	inf	6.29	4.87	3.25	2.24	2.64	2.43

表 6 IMF9 和 IMF10 分别在 GRU 和 MLR 中预测结果

IMF 分量	算法	GRU	MLR
IMF9	MAPE	6.71	6.37
	RMSE	0.23	0.06
IMF10	MAPE	12.13	2.19
	RMSE	0.09	0.01

(二) 预测结果与分析

本文将 2016 年 1 月 1 日到 2016 年 10 月 19 日, 共 7019 条数据去预测 2016 年 10 月 20 日到 201 年 11 月 29 日共 984 个数据。本数据有四个特征, 本文采用滑动窗口处理方法, 以 24 小时为时间步长(24 小时中包含 24 个数据)重构神经网络的输入数据, 将维度升至大小为[7019, 24, 4]的子数据集, 最后将数据集输入 GRU 神经网络进行预测。对于 MLR 预测就比较方便, 不需要对数据进行多余处理, 将预处理后的分解数据直接输入 MLR 输入端即可进行预测。

将预测的结果与 AR 模型预测的结果利用 PSO 算法进行寻找最优权重, 本文设置 PSO 参数为: 粒子数量为 10, 迭代次数为 1000 次, 最大值为 1。经过计算可得最优权重为 0.0705。

为了进一步体现该方法的优势, 将本文提出的算法与 2 组单一预测算法和 4 组组合预测模型的预测结果进行对比。

单一算法模型为:

1. 单纯使用 GRU 神经网络预测;
2. 单独使用 MLR 进行预测;

组合预测方法:

1. 经过 EEMD 分解后使用 GRU 神经网络预测, 简称 C-GRU;
2. 经过 EEMD 分解后 MLR 进行预测, 简称 C-MLR;
3. 经过 EEMD 分解, 对模糊熵值低 IMF 分量使用 MLR 模型预测, 模糊熵值高 IMF 分量的部分使用 GRU 神经网络模型预测, 简称 C-GM。

每个训练方案都各自进行参数优化, 达到各预测模型的最佳预测参数。

从表 7 和图 7 可以明显看出 PSO-CGM-A-模型结果远好于其余两种单一预测模型, 他基本还原了原始数据。MLR 模型在预测复杂数据上表现很差, 因为此数据包含了大量了非线性特征; GRU 神经网络模型有不错的预测效果, 但是相遇 PSO-CGM-A-模型比较差, 其中 MAPE 相差 4.39%。

表 7 PSO-CGM-A 模型与单一算法模型对比

预测方法	PSO-CGM-A-	GRU	MLR
MAPE	4.57	8.96	57.04
RMSE	3.94	0.18	37.22

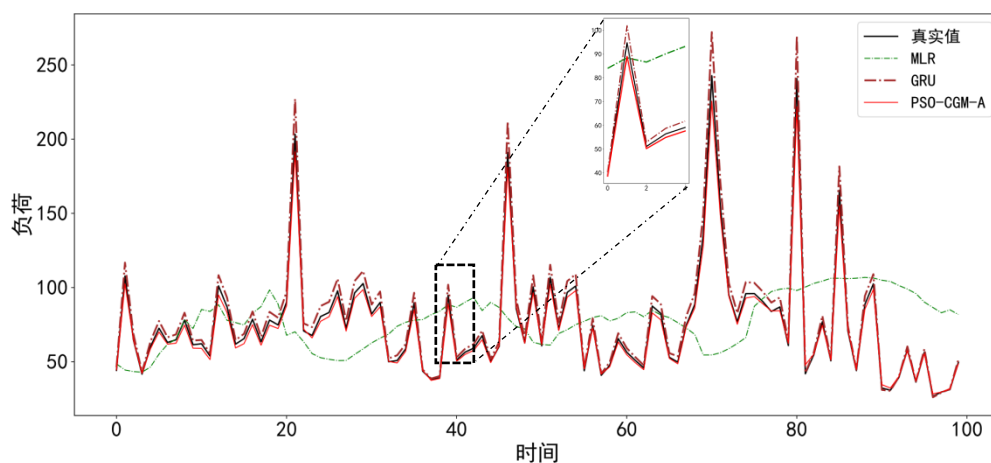


图 7 PSO-CGM-A 模型与单一算法模型对比图

表 8 和图 8 给出了 PSO-CGM-A 模型与组合算法模型对比, PSO-CGM-A 模型明显优于其他组合模型, 值得一提的是 C-MLR 模型表现特别差, 负荷数据经

过 EEMD 后, 模糊熵值高的分量通过 MLR 预测误差很大, 重构叠加时又增大了误差, 使得预测结果没有任何意义。相比 C-GM 模型和 C-GRU 模型, PSO-CGM-A 模型的 MAPE 分别提升 1.34 % 和 2.23 %, 分别提现了合理利用 GRU 神经网络、MLR 的特性和充分考虑复合数据的线性与非线性特征。

表 8 PSO-CGM-A 模型与组合算法模型对比

预测方法	PSO-CGM-A	C-GRU	C-MLR	C-GM
MAPE	4.57	6.80	36.10	5.91
RMSE	3.94	4.06	27.81	2.85

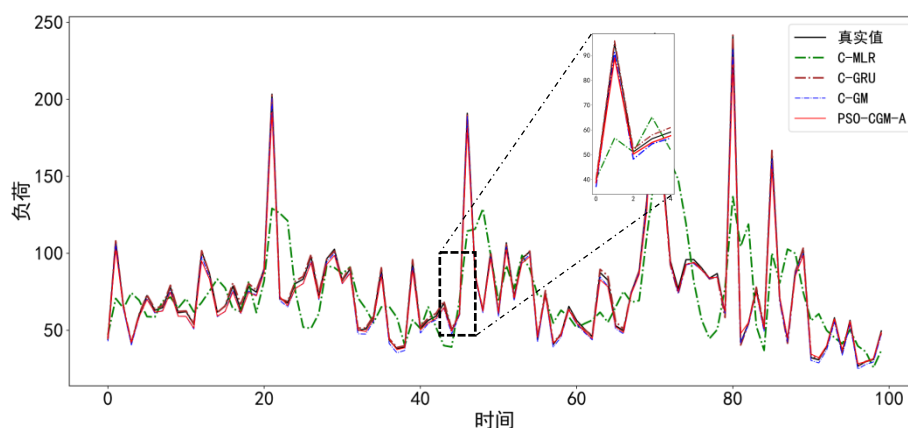


图 8 PSO-CGM-A 模型与组合算法模型对比

(三) 窃电检测模拟结果

根据前面的方法, 我们利用窃电检测六种常见的窃电模式, 可以生成异常用电数据集, 异常用电数据集如图 9、图 10 所示。其中, 我们选取了 300 个采样点 (间隔一小时), 分别带入六种窃电模式, 从中选取 100 个采样点进行异常模拟。我们选取 PSO-CGM-A 模型作为负荷预测模型, 滑动时间窗口与异常点阈值判断作为窃电检测模型, 对该生成的异常数据集进行窃电检测。

如图 11、图 12 所示, 预测数据与生成数据存在明显不同, 我们使用窗口长度为 6 个时间单位、步长为 1 的滑动时间窗口, 从左到右依次滑过整个数据集,

利用两次阈值判断，对用户的用电行为进行异常判断。

首先，设定阈值为 0.8，即每一个滑动时间窗口内的生成数据总和不能低于同一个窗口内预测数据综合的 80%，否则视为存在窃电嫌疑。这样的设定方法充分考虑了用电用户存在与历史用电行为不同的突发行为，例如请假、加班等。但是使用该方法用于窃电检测是不够的。

在使用滑动时间窗口筛选出具有窃电嫌疑的时间区间后，对其中的单个时间采样点再做一次异常检测。我们利用传统的阈值判定方法，本文将阈值设定为 0.95，即异常区间中的每一个采样点，生成数据不能低于预测数据的 95%，否则视为存在窃电嫌疑。

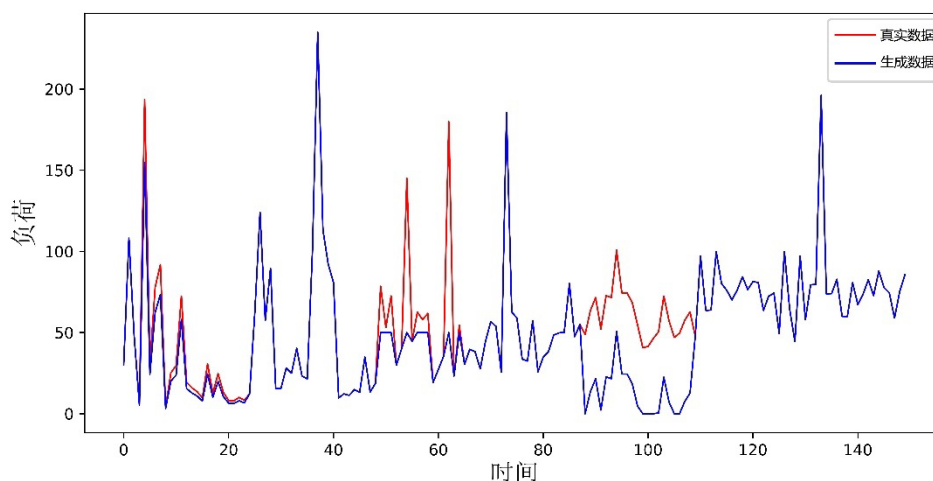


图 9 生成数据（前 150 个数据）

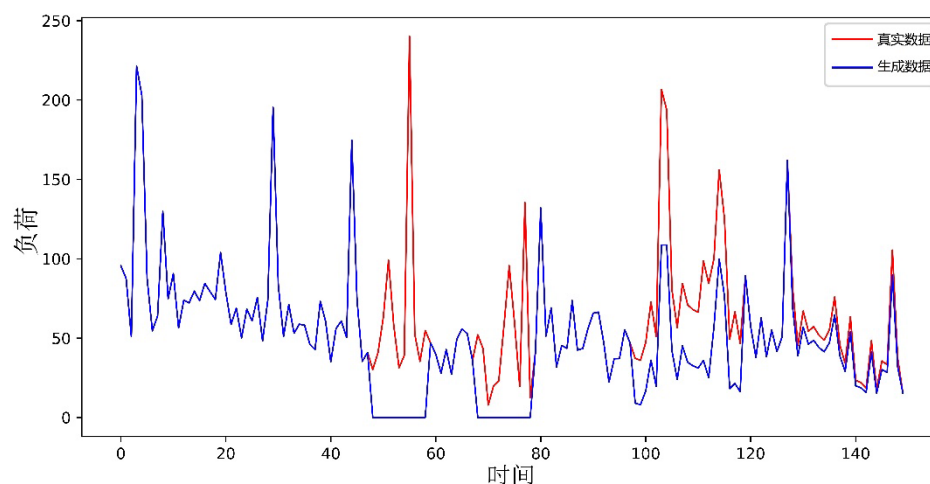


图 10 生成数据（后 150 个数据）

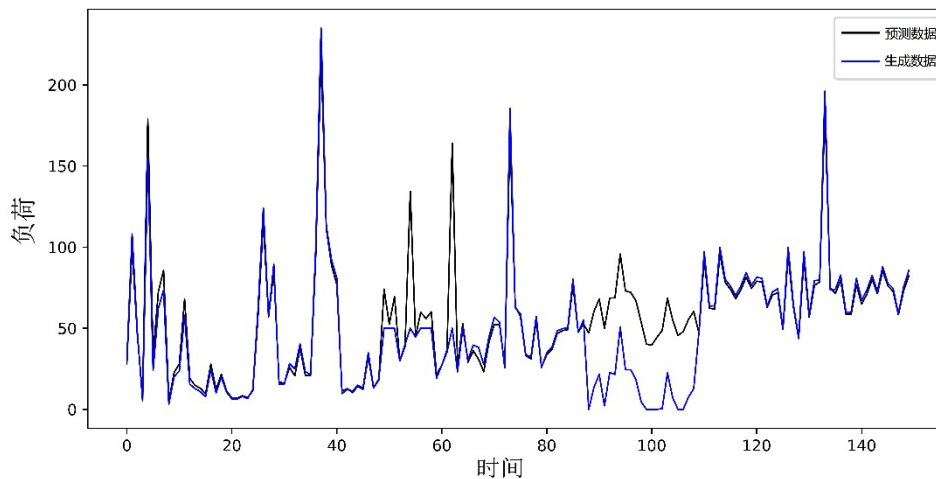


图 11 预测数据（前 150 个数据）

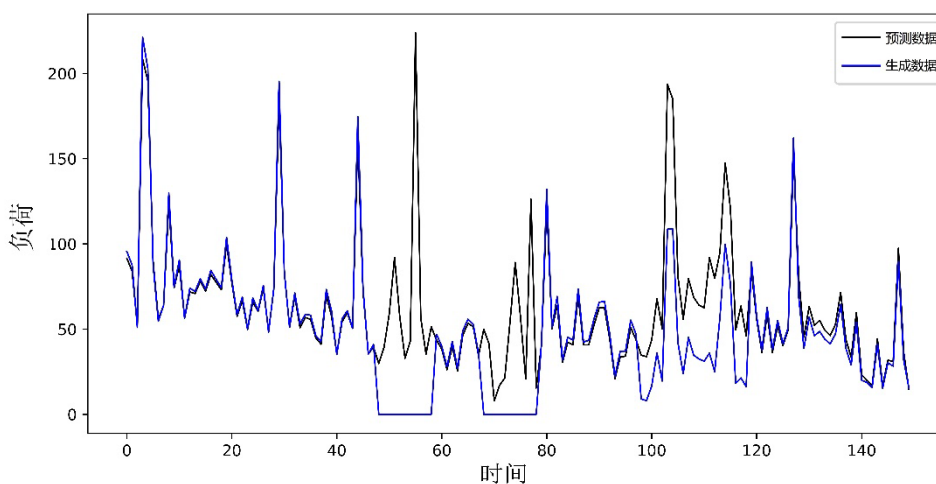


图 12 预测数据（后 150 个数据）

我们通过两次阈值判断，筛选出异常时间点，如表 9 所示。

表 9 异常时间点

用户	检测为窃电用户	检测为正常用户
实际为窃电用户	96	4
实际为正常用户	11	189

由实验数据可得，该模型的窃电检测召回率达到 89.72%，准确率达到 96%，证明该模型具有优良的检测性能，对窃电检测领域的研究具有一定的价值。

五、研究的结论和启示

（一）主要结论

本文使用 EEMD 分解的优化 GRU-MLR 和 AR 负荷预测模型，对用户历史用电负荷进行有效预测，同时提出了一种滑动时间窗口窃电检测方法，从而对客户是否发生异常用电行为做出判断，得出如下几点结论：

1. 该算法对负荷的处理使用 EEMD 分解，将随机性很强的负荷数据分解，得到周期性，平稳性好的 IMF 分量，有助于预测模型的训练；使用模糊熵对不同的 IMF 分量进行计算，将得出的模糊熵值从大到小排序，熵值高的部分进行 GRU 神经网络模型预测，熵值低部分进行 MLR 模型预测，使各个部分的特征能更好的适应模型，提升了预测的精度；使用 AR 模型提取出了负荷数据的线性特征，与 GRU-MLR 预测模型进行权重组合，最后利用粒子群优化权重，得到了更加精确的预测；

2. 提出基于滑动时间窗口的窃电检测模型，减少由于用户用电行为推前或者滞后导致的误检情况；

3. 本文使用了麻省理工公开的公寓电力负荷数据集，同时模型的检测效果使用召回率和准确率进行评估。结果表明，文本所提模型的窃电检测召回率达 89.72%，准确率达到 96%，证明该模型具有优良的检测性能，对窃电检测领域的研究具有一定的价值；

4. 基于数据驱动的负荷预测的窃电检测可以帮助我们精确识别用户的窃电行为。防窃电作为当下保证电力系统安全稳定经济运行的重要部分，窃电检测的工作意义十分重要。

（二）研究的不足与改进方向

本文使用 EEMD 分解的优化 GRU-MLR 和 AR 负荷预测模型及窃电检测模型对生成的异常用电数据进行检测，结果表明该模型具有较高的精准度。但是该

检测模型依然存在某些不足，总结得出以下三点不足和改进：

1. 数据驱动的基于回归的窃电检测模型以负荷预测为理论依据，通过比较真实负荷数据与预测负荷数据的差异判断用户异常用电行为。该方法在充分考虑个人用户用电行为的差异性与特殊性，但同时由于该方法需要对每一位用户进行独立建模，导致算力成本昂贵。因此，该模型适合同其他类型的窃电检测方法一起使用，例如可以先根据台区线损情况，缩小窃电用户范围，再根据用户历史用电情况划分用户信用等级等。

2. 在该窃电检测模型中使用了传统的 GRU 模型，该类模型存在无法并行计算的问题，即数据输入必须满足顺序输入且每次一个数据的要求，这明显不满足在大数据背景下高效检测的要求。在最近的时间序列研究中，Google 提出了可以并行处理数据的 Transformer 模型，今年周提出了基于该模型的 Informer 模型，大大降低了计算成本和计算时间，这对数据驱动的窃电检测模型研究有着重大的意义。在后续的研究过程中，将会对这一部分进行探索。

3. 由于用电负荷数据大部分处于保密状态，未进行公开，因此本文使用的负荷数据是根据已经公开的负荷数据，人为地添加噪声，生成的异常负荷数据。但在实际的电能调度中，有更多的异常情况是我们未曾考虑的，为了进一步对窃电检测进行研究，我们还会与当地供电部门合作，使用真实的负荷数据。

参考文献

- [1] 葛洲, 余英, 郑莎莎, 等. 窃电现象对供电系统的影响及反窃电系统[J]. 科技创业月刊, 2012(12): 198-199.
- [2] 周瑾. 窃电与防窃电[J]. 福建电力与电工, 2004, 03: 70-72.
- [3] 孔令俏. 4563 户窃电用户去年被查处, 挽回经济损失 4284 万元 [N]. 大众日报.
- [4] 龙金光. 2014 年广东查处窃电案 3964 起 [N]. 南方日报 .
- [5] Smith T B. Electricity theft: a comparative analysis[J]. Energy Policy, 2004, 32(18):2067-2076.
- [6] Depuru SSSR, Wang L, Devabhaktuni V. Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft[J]. Energy Policy, 2011, 39(2):1007-1015.
- [7] 窦健, 刘宣, 卢继哲, 等. 基于用电信息采集大数据的防窃电方法研究 [J]. 电测与仪表, 2018, 55 (21): 43-49.
- [8] 陈启鑫, 郑可迪, 康重庆, 等. 异常用电的检测方法: 评述与展望 [J]. 电力系统自动化, 2018, 42 (17): 189-198.
- [9] 韩谷静, 殷小贡, 秦亮, 等. 电能计量设备防电流法窃电新技术 [J]. 电测与仪表, 2004, 44 (502): 29-32.
- [10] 王国峰. 农网智能化的四项内容[J]. 中国电力企业管理, 2011(14):33-34.
- [11] 周文婷, 顾楠, 王涛, 等. 基于数据挖掘算法的用户窃电嫌疑分析 [J]. 河南科学, 2015, 33(10):1767-1772.
- [12] 韩谷静, 殷小贡, 秦亮, 等. 电能计量设备防电流法窃电新技术 [J]. 电测与仪表, 2004, 44 (502): 29-32.
- [13] 陈伟樑. 电力系统窃电特点及反窃电技术分析 [J]. 机电信息,

2011(21):227-229.

[14]Amral N, Ozveren C S, King D. Short term load forecasting using Multiple Linear Regression[C]. Universities Power Engineering Conference, UPEC 2007 International. IEEE, 2007: 1192-1198.

[15]Hermias J P, Teknomo K, Monje J C N. Short-term stochastic load forecasting using autoregressive integrated moving average models and hidden Markov model[C] . International Conference on Information and Communication Technologies. 2017: 131-137.

[16]Saber A Y, AlamA K M R. Short term load forecasting using multiple linear regression for big data[C] . IEEE Symposium Series on Computational Intelligence. Honolulu: IEEE, 2017: 1-6

[17]Ko C N, Lee C M. Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter[J]. Energy, 2013, 49(49): 413-422.

[18]蒋敏, 顾东健, 孔军, 等. 基于在线序列极限支持向量回归的短期负荷预测模型[J]. 电网技术, 2018, 42(7): 2240-2247.

[19]李鹏, 何帅, 韩鹏飞, 等. 基于长短期记忆的实时电价条件下智能电网短期负荷预测[J]. 电网技术, 2018, 12(12): 4045-4052.

[20]Kong W, Dong Z Y, Luo F, et al. Effect of automatic hyper parameter tuning for residential load forecasting via deep learning[C]. 2017 Australasian Universities Power Engineering Conference (AUPEC). IEEE, 2017: 1-6.

[21]K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber, “LSTM: A search space odyssey” IEEE Trans. Neural Networks and Learning Systems, vol. 28, no. 10, pp. 2222-2232, Oct. 2017.

- [22]T. Ergen, and S. S. Kozat, “Efficient online learning algorithms based on lstm neural networks” IEEE Trans. Neural Networks and Learning Systems, vol. 29, no. 8, pp. 3772-3783, 2018.
- [23]周莽, 高僮, 李晨光. 等. GRU 神经网络短期电力负荷预测研究[J]. 科技创新与应用, 2018(33): 52-53, 57.
- [24]陈海文, 王守相, 王绍敏, 等. 基于门控循环单元网络与模型融合的负荷聚合体预测方法[J]. 电力系统自动化, 2019, 43(1): 65-74.
- [25]Qiu X, Ren Y, Suganthan P N, et al. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting [J]. Applied Soft Computing, 2017, 54: 246-255.
- [26]G. F. Fan, S. Qing, H. Wang, W. C. Hong, H. J. Li, “Support vector regression model based on empirical mode decomposition and auto regression for electric load forecasting” Energies, vol. 6, no. 4, pp. 1887–1901, 2013.
- [27]T. Liu, Y. Jin, Y. Gao, “A new hybrid approach for short-term electric load forecasting applying support vector machine with ensemble empirical mode decomposition and whale optimization” Energies, vol. 12, no. 8, p. 1520, 2019.
- [28]Z. Liu, W. Sun, J. Zeng, “A new short-term load forecasting method of power system based on eemd and ss-pso” Neural Computing and Applications, vol. 24, no. 3-4, pp. 973–983, 2014.
- [29]W. Q. Zhao, Z. J. Shen and G. Ji, “Anomaly Detection for Power Consumption Pattern Based on Deep Learning” Electr. Power Autom. Eq, vol. 38, No. 9, pp. 34–38, Sept, 2018, DOI: 10. 16081/j. issn. 1006-6047. 2018. 09. 006.
- [30]Y. Liu, S. Garg, J. Nie, Y. Zhang, Z. Xiong, J. Kang, M. S. Hossain, “ Deep anomaly detection for time-series data in industrial IoT : a

communication-efficient on-device federated learning approach”IEEE Internet Things J, to be publish, DOI: 10. 1109/IIOT. 2020. 3011726.

[31]马云波, 李英娜, 李川. 基于 GRU-Bagging 模型的异常用电行为检测[J]. 数据通信, 2020(04): 30-34+38.

[32] Vaswani, Shazeer N, Parmar N, et al. Attention is all you need[J]. arXiv, preprint arXiv:1706.03762, 2017.

[33]H. Y . Zhou, S. H. Zhang, J. Q. Peng, S. Zhang, J. X. Li, H. Xiong, W. C. Zhang, “Informer: BeyondEfficient Transformer for Long Sequence Time-Series Forecasting” AAAI2020. <https://arxiv.org/pdf/2012.07436.pdf>

[34]李军, 李青. 基于 CEEMDAN-排列熵和泄漏积分 ESN 的中期电力负荷预测研究[J]. 电机与控制学报, 2015, 19(08): 70-80.

[35]杨望灿, 张培林, 王怀光, 陈彦龙, 孙也尊. 基于 EEMD 的多尺度模糊熵的齿轮故障诊断[J]. 振动与冲击, 2015, 34(14): 163-167+187.

[36]李振坤, 陈星莺, 余昆, 刘皓明, 赵波. 配电网重构的混合粒子群算法[J]. 中国电机工程学报, 2008(31): 35-41.

[37]J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed and M. Mohamad, “Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines” IEEE Trans. Power Deliv, vol. 25, No. 2, pp. 1162–1171, Apr, 2010, DOI: 10. 1109/TPWRD. 2009. 2030890.

[38]Z. Marcelo, J. Edgard, P. Marcelo, P. Manoel, V. Zambenedetti and I. Chueiri, “A Tunable Fraud Detection Systemfor Advanced Metering Infrastructure Using Short-lived Patterns” IEEE Trans. Smart Grid, vol. 10, No. 1, pp. 830–840, Jan, 2019, DOI: 10. 1109/TSG. 2017. 2753738.

致谢

值此论文撰写完成之际,回望这将近两个月里的日日夜夜,收获了太多太多,从刚开始的彷徨与困惑,到现在脑海中清晰的设计思路和步骤,我们成长了很多。

我们要感谢我们的指导教师,感谢王教授对我们的这次比赛的关心和精心指导。在论文选题环节与我们多次讨论给出了宝贵的意见,在后期的模型搭建中也帮助我们建立了预测模型与窃电模型,王教授为我们付出的宝贵时间和极大心血,让我们受益匪浅。在此衷心感谢指导教师对我们的无私帮助!

同时,我们还要感谢在一起愉快研究、鼎力支持的成员,大家一起相互支持、帮助,共同克服一个一个的困难和疑惑。还特别感谢师兄在研究领域给予的方向指导和鼓励,谢谢你们!