

# 基于 Zoneout 的跨尺度循环神经网络及其在短期电力负荷预测中的应用



庄世杰 於志勇 郭文忠 黄昉菀

福州大学数学与计算机科学学院 福州 350116

福州大学福建省网络计算与智能信息处理重点实验室 福州 350116

(2942265521@qq.com)

**摘要** 通过精确的电力负荷预测,智能电网可以提供比传统电网更高效、可靠和环保的电力服务。现实生活中,电力负荷数据往往存在着与历史数据较高的时间相关性,而传统的神经网络却很少关注它。近年来,循环神经网络(Recurrent Neural Network,RNN)由于可以很好地捕获在时间上距离很远的数据之间的相关性,因此在电力负荷预测中受到越来越多的关注。但是,由于RNN特有的自循环结构,当采用随时间的反向传播算法(Back-Propagation Through Time,BPTT)进行网络训练时,随着网络层数的增加,很容易发生梯度消失等问题,从而导致预测精度下降。目前已有多种解决梯度消失问题的RNN架构,如长短期记忆网络(Long Short-Term Memory,LSTM)和门控制单元(Gated Recurrent Unit,GRU),但其复杂的内部结构会增加训练时长。为了解决上述问题,文中首先对目前流行的各种RNN架构进行了研究和分析,其次结合最新提出的Zoneout技术,设计了一种跨时间尺度的分模块循环神经网络架构,重点研究了隐藏层模块的随机更新策略,不仅有效解决了梯度消失问题,而且大幅度减少了待训练的网络参数。基于基准数据集和实际负载数据集的实验结果表明,该结构可以获得比目前流行的RNN架构更好的性能。

**关键词:**短期电力负荷预测;循环神经网络;跨时间尺度;Zoneout

**中图法分类号** TP183

## Short Term Load Forecasting via Zoneout-based Multi-time Scale Recurrent Neural Network

ZHUANG Shi-jie, YU Zhi-yong, GUO Wen-zhong and HUANG Fang-wan

College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116, China

Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou University, Fuzhou 350116, China

**Abstract** Because accurate power load forecasting, smart grids can provide more efficient, reliable and environmentally friendly power services than traditional grids. In real life, power load data often has a high temporal correlation with historical data, while traditional neural networks pay little attention to it. In recent years, the recurrent neural network (RNN) has received more and more attention in power load forecasting, because it can well capture the correlation between data with large cross-time scale. However, due to the unique self-connections structure of RNN, when the back-propagation through time (BPTT) is adopted for network training, the problems such as vanishing gradient are prone to occur with the number of network layers increases, resulting in a decrease in prediction accuracy. There are varieties of RNN architectures that can solve the vanishing gradient problem, such as long short-term memory (LSTM) and gated recurrent unit (GRU), but their complex internal structure will increase the training time. In order to solve the above problems, this paper first analyzes and studies RNN and its variants, and then combines the Zoneout function to design a multi-time scale modularized RNN architecture, focuses on the update strategy of hidden layer modules. It not only effectively solves the vanishing gradient problem, but also greatly reduces the number of network parameters that need to be trained. Experimental results based on the benchmark dataset and the real-world load dataset show that this architecture can achieve better performance than the current popular RNN architecture.

**Keywords** Short term load forecast, Recurrent neural networks, Multi-time scale, Zoneout

### 1 引言

负荷预测可以为电力生产部门和运营部门提供决策依据,例如降低生产成本、保障电力供应、制定分时电价、安排设备检修等<sup>[1]</sup>。电力负荷预测根据预测时长通常可以分为4类:长

电力负荷预测是智能电网的重要功能之一。准确的电力

到稿日期:2019-08-07 返修日期:2019-10-18 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61772136,61672159);福建省中青年骨干教师教育科研项目(JT180045)

This work was supported by the National Natural Science Foundation of China(61772136,61672159) and Research Project for Young and Middle-aged Teachers of Fujian Province (JT180045).

通信作者:黄昉菀(hfw@fzu.edu.cn)

期预测、中期预测、短期预测和超短期预测<sup>[2]</sup>。其中,短期负荷预测指预测未来一天到几天的电力负荷值,是本文研究的重点<sup>[3]</sup>。

短期负荷预测研究历史悠久,可以追溯到 20 世纪 50 年代。经典方法包括回归分析<sup>[4]</sup>、时间序列法<sup>[5-6]</sup>和卡尔曼滤波法<sup>[7]</sup>等传统的数理统计方法。20 世纪 90 年代初,随着人工智能的发展,机器学习逐渐被引入短期负荷预测中,如专家系统<sup>[8]</sup>、模糊预测<sup>[9]</sup>、小波分析<sup>[10]</sup>、混沌理论<sup>[11]</sup>、支持向量机<sup>[12]</sup>、聚类分析模型<sup>[13]</sup>和人工神经网络<sup>[14]</sup>等。近年来,随着深度学习的兴起,深度神经网络在各类预测领域越来越受到关注,并在许多基准数据集上取得了优异的性能。深度学习的目标是将多个模块堆叠在一起形成深层网络,以便创建更具表现力的模型,从而可以学习更抽象的数据表示,并实现更好的学习性能<sup>[15]</sup>。作为深度学习神经网络的一种,循环神经网络(RNN)依靠其自身的隐藏层循环结构能够很好地捕捉数据间的时间相关性,已被广泛用于各类时间序列预测问题。但是,随着人们对 RNN 研究的加深,发现了 RNN 中存在的缺陷,采用 BPTT 方法进行网络参数训练时容易出现梯度消失问题,因此无法处理数据间的长期依赖<sup>[16]</sup>。目前,已衍生出多种 RNN 架构来解决梯度消失的问题,主要包含以下几种类型:门架构、跨时间尺度连接、初始化约束和正则化方法等。其中最具有影响力的是以 LSTM 为代表的门架构<sup>[17]</sup>。需要说明的是,门架构由于其复杂的内部结构会产生大量待学习的网络参数,给网络训练带来了很大的困难,并提升了复杂度。针对该问题,本文拟设计一种基于跨时间尺度连接的分模块循环神经网络架构,并利用 Zoneout 技术实现对各个模块的随机更新,在解决梯度消失问题的同时,对短期负荷进行快速准确的预测。

## 2 相关工作

近年来,已有大量的研究工作围绕 RNN 的梯度消失问题展开,致力于解决 RNN 的长期依赖。其中最流行的方法是采用门架构。门架构的基本思想是利用门来控制信息的记忆和遗忘,主要模型包括 LSTM 及其变体 GRU<sup>[18]</sup>。LSTM 的特点是新增一个状态变量来记录历史信息,该状态是在隐藏层中引入细胞单元(cell)、遗忘门、输入门和输出门。LSTM 已在许多时间序列预测领域得到广泛的应用,但是,其增加了大量待学习的权重参数,为网络训练带来了困难。作为 LSTM 的简化变体,GRU 取消了细胞单元,直接对隐藏层状态  $h(t)$  采用线性叠加法则生成下一时刻的隐藏层状态,并把门的个数从 3 减少到 2。虽然 GRU 可以有效降低 LSTM 的复杂度,但是仍然需要耗费一定的时间开销学习更新向量  $u(t)$  和重置向量  $r(t)$ ,对于实时性要求较高的任务来说并不适合。

为了解决门架构训练耗时的不足,研究人员提出了跨时间尺度连接的 RNN 架构。跨时间尺度连接的基本思想是将原本的长期依赖转换为短期依赖,从而解决训练时产生的梯度消失问题。具体的做法是:为隐藏层的神经元设置不同的时间尺度,让模型能够直接学习到不同时刻的信息。相关的模型有 Clock-work RNN(CW-RNN)<sup>[19]</sup>和 Dilated RNN<sup>[20]</sup>。其中,CW-RNN 将隐藏层划分为多个模块,并利用预先定义

的规则为每个模块设定固定的更新周期,例如将模块的周期设为  $2^{i-1}$  ( $i$  表示模块号),从而使得长距离的依赖可以通过周期较长模块的少数几次访问得到。需要说明的是,CW-RNN 的原理与 GRU 非常类似,不同的地方在于 GRU 的更新向量是通过更新门学习得到,其值在  $[0,1]$  之间;而 CW-RNN 的更新向量则是根据当前时刻  $t$  是否能够整除各个模块的周期得到的,其块内元素值保持一致,其值非 0 即 1,无须训练,没有额外开销。与 CW-RNN 不同,Dilated RNN 不是对同一个隐藏层进行分模块处理,而是设置了多个隐藏层,最下面的隐层每个时间步都会循环,较高的隐层循环周期更长。

除了上述两种方法外,还有其他的策略用于解决 RNN 的梯度消失问题,如 Independently RNN (IndRNN)<sup>[21]</sup> 和 Zoneout 技术<sup>[22]</sup>。

## 3 基于 Zoneout 的跨时间尺度 RNN (ZM-RNN)

短期负荷预测相对于中长期负荷预测来说要求更快的预测速度以及更准确的预测精度,以便能在更短的时间内找出负荷变化的规律。而对于 LSTM 和 GRU 这样的门结构模型来说,复杂的训练过程以及大量的训练耗时不适用于短期电力负荷预测,因此需要设计一种简单快速的模型来提高预测的精度,同时缩短模型训练的时间。

借鉴上述相关工作,本文重点对跨时间尺度循环神经网络展开研究,因为跨时间尺度相比门架构而言,在解决长期依赖的同时并没有带来模型复杂度的大幅度提升。通过对其代表模型 CW-RNN 的深入研究不难发现,CW-RNN 的不足在于网络性能依赖预先设定的周期规则,且不够灵活。因此,本文重点研究了隐藏层模块的更新策略,将 Zoneout 技术应用到跨时间尺度循环神经网络,旨在利用随机概率策略代替 CW-RNN 的固定周期策略,构建一个快速简便且更为通用的 RNN 架构用于短期负荷预测。设计该网络的动机在于:

(1)通过对 LSTM 等门架构循环神经网络的研究发现,虽然理论上各个门得到的控制向量的元素值应该在  $[0,1]$  之间,但事实上,表现良好的网络的门向量元素值通常非常接近 0 或者 1。因此,从直觉上,我们认为可以将原本需要耗时训练的门简化为一种简单的二元门(元素值非 0 即 1)来控制信息的记忆与更新,从而节省大量的训练时间。

(2)传统的 Zoneout 技术的特点是每个时刻随机地令某些隐藏层神经元保持其上一时刻的值。这种以神经元为单位的做法存在较大的随机性,容易造成网络性能不稳定。并且传统的 Zoneout 技术应用到 LSTM 网络中,存在着参数过多、训练复杂的问题,不适用于短期负荷预测。为了解决上述问题,本文借鉴 CW-RNN 的做法,首先对隐藏层进行分模块处理,然后对各个模块利用 Zoneout 技术,即每个时刻随机地令某些隐藏层模块内所有的神经元均保持其上一时刻的值。分模块处理的好处有 3 点:1)减少部分模块的更新次数可以显著地提升计算效率;2)利用不同模块更新周期的不同,可以同时兼顾短期依赖和长期依赖;3)达到类似模型集成的效果。短期负荷预测相对于中长期负荷预测来说,具有更大的不确定性,通过分块模式和模型集成能够使 ZM-RNN 充分学习到短期依赖和长期依赖,更适合解决短期负荷预测的问题。

(3)与传统的 CW-RNN 不同,传统的 CW-RNN 方法为

每一个模块都设置了固定的周期,只有当前时刻能够整除某一模块的周期时,该模块内所有神经元的更新向量值才全部置为1,否则全部置为0。这种做法使得网络性能依赖于周期的设定。为了避免人工规则对网络性能的影响,本文建议采用块 Zoneout 技术,在每个时间步为每一模块随机生成一个更新概率值,通过硬阈值策略规定当某一时刻某一模块的更新概率大于阈值时,其模块内所有神经元的更新向量值全部置为1,否则全部置为0。

ZM-RNN 的具体工作流程如下。

(1)将 RNN 的隐藏层划分为  $n$  个模块。假设每个模块的神经元个数为  $m$ ,则隐藏层神经元的总数为  $N_h = n * m$ 。

(2)为每个时刻每个模块随机设置一个更新概率  $P_i(t)$  ( $i=1,2,\dots,n$ )。

(3)设定一个更新阈值  $\epsilon > 0$ ,若  $p_i(t) > \epsilon$ ,则模块  $i$  内所有神经元在时刻  $t$  的更新向量值  $u_i(t) = 1$ ;否则  $u_i(t) = 0$ 。

(4)将所有模块的更新向量值按顺序串联得到时刻  $t$  的更新向量  $\mathbf{u}(t) = [u_1(t), u_2(t), \dots, u_n(t)]^T$ 。

(5)时刻  $t$  的隐藏层状态  $h(t)$  的计算公式为:

$$h(t) = (1 - \mathbf{u}(t)) \odot h(t-1) + \mathbf{u}(t) \odot h(t)' \quad (1)$$

其中,符号  $\odot$  表示向量元素间的相乘,  $h(t)'$  表示候选隐藏层状态,其计算公式如下:

$$h(t)' = \tanh[W_{ih}x(t) + (W_{hh} * \mathbf{M})h(t-1) + b_h] \quad (2)$$

其中,  $W_{ih}$  表示输入层到隐藏层的连接权重,  $W_{hh}$  表示隐藏层的自连接权重,  $b_h$  表示隐藏层偏置。  $\mathbf{M}$  是一个二元掩码矩阵,用于定义隐藏层神经元在相邻时刻间的连接权边是否存在,当  $M_{ij} = 1$  时,表示神经元  $i$  和  $j$  之间存在连接,否则表示不存在连接。矩阵  $\mathbf{M}$  的作用在于防止过拟合,本文采用 CW-RNN 的设置方式,将  $\mathbf{M}$  设置成上三角为 1 的掩码矩阵,以便进行模型的比较。

(6)时刻  $t$  的输出值  $y(t)$  的计算公式为:

$$y(t) = \tanh[W_{ho}h(t) + b_o] \quad (3)$$

其中,  $W_{ho}$  表示隐藏层到输出层的连接权重,  $b_o$  表示输出层偏置。

ZM-RNN 的结构框图如图 1 所示。

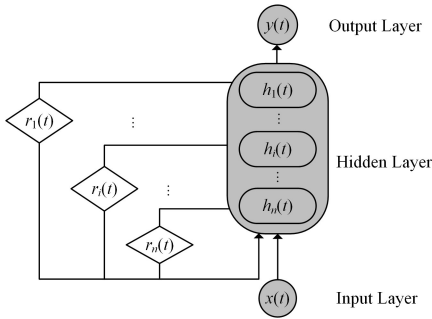


图 1 ZM-RNN 结构图

Fig. 1 Architecture diagram of ZM-RNN

图 1 中,  $x(t)$  和  $y(t)$  分别表示时刻  $t$  的输入值和输出值;  $h_i(t)$  ( $i=1,2,\dots,n$ ) 表示隐藏层模块  $i$  在时刻  $t$  的状态;  $r_i(t)$  表示模块  $i$  在时刻  $t$  的更新步长,其值取决于该模块前后两次更新之间的间隔时间步。

需要说明的是,ZM-RNN 和 CW-RNN 待学习的网络参数相当,均低于具有相同隐藏层神经元个数的其他 RNN 架

构。参数计算的公式如下,其中式(4)和式(6)分别用于计算 RNN 和 GRU 待学习参数的数量;式(5)适用于 LSTM 和 Zoneout-LSTM,式(7)则适用于 CW-RNN 和 ZM-RNN。

$$N_p = N_h^2 + N_h(N_i + N_o + 1) + N_o \quad (4)$$

$$N_p = 4N_h^2 + N_h(4N_i + N_o + 4) + N_o \quad (5)$$

$$N_p = 3N_h^2 + N_h(3N_i + N_o + 3) + N_o \quad (6)$$

$$N_p = m^2 \frac{n(n+1)}{2} + mN_i(N_i+1) + N_o(N_h+1) \quad (7)$$

其中,  $N_p$  表示各个模型所需要计算的参数总数,  $N_h$  为隐藏层神经元个数,  $N_i$  为输入层神经元个数,  $N_o$  为输出层神经元个数,  $m$  为 CW-RNN 和 ZM-RNN 中一个隐层模块的神经元个数,  $n$  为模块总数,  $N_t$  为时刻  $t$  中需要更新的块数数量。显然,由于  $N_h = n * m$ ,且采用部分模块更新策略,ZM-RNN 和 CW-RNN 待学习的网络参数最少。

## 4 实验

为了验证 ZM-RNN 的性能,本文将其与目前流行的 RNN 架构进行比较。实验在两个数据集上进行,一个是时间序列预测任务的基准数据集,一个是真实的电力负荷数据集。

### 4.1 数据集介绍

Mackey-Glass(MG)延时差分系统通常被用作混沌时间序列预测的基准数据集。时刻  $t$  的输入值  $x(t)$  可由下式计算得到:

$$x(t) = x(t-1) + \delta \left( \frac{\alpha x(t-\tau)}{1 + x(t-\tau)^\gamma} - \beta x(t-1) \right) \quad (8)$$

其中,参数设置为  $x(0) = 0.5, \delta = 1, \gamma = 10, \alpha = 0.2, \beta = 0.1$ 。一般情况下,认为当  $\tau > 16.8$  时,该系统具有混沌效果。本文选择  $\tau = 17$  来生成一条长度为 1700 个点的时间序列,如图 2 所示。

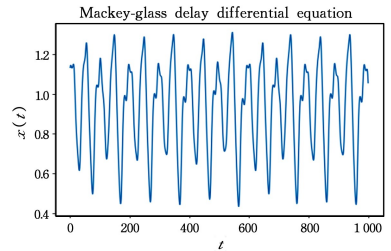


图 2 MG 时间序列

Fig. 2 Macke-Glass time series

真实数据集是由澳洲能源市场运营商(Australian Energy Market Operator, AEMO)提供的 2014—2017 年共 4 年澳大利亚昆士兰地区(Queensland, QLD)的电力负荷值。在原始的数据集中存在部分的数据缺失,这些缺失的电力负荷值会破坏负荷时间序列的排列结构和发展规律,而且会严重影响其预测的精度。因此,需要对负荷数据进行缺失处理,以提高负荷数据的整体质量,降低预测工作的难度。考虑到电力数据的连续性,本文采用取前一天和后一天的电力平均值来取代缺失值的数据补充方法。最后,本文选取每一天的电力负荷最大值,进行 min-max 归一化处理,将数据缩放到  $[0, 1]$  之间,这样做的好处是对于方差小的数据可以增强其稳定性。图 3 为 2017 年 6 月—2017 年 12 月的每天电力负荷最大值的时间序列图。可以看出,真实电力负荷数据集更加不规则。



图3 昆士兰电力负荷序列

Fig. 3 QLD power load series

## 4.2 实验设置

本节给出本文实验涉及的各种 RNN 模型的相关设置。其中 RNN, LSTM, GRU 使用 Tensorflow 库中自带的模型, Zoneout-LSTM 和 IndRNN 采用各自论文提供的代码, CW-RNN 和本文设计的 ZM-RNN 在 Tensorflow 中编写实现。在所有的实验中,我们把数据集划分为训练集、验证集和测试集 3 个部分,通过验证集调整模型的超参数,并在测试集上评估最终模型的性能,详细设置如表 1 所列。

表 1 数据集的划分

Table 1 Partitions of datasets

	MG	QLD
训练集	1500 个点	1095 天
验证集	100 个点	182 天
测试集	100 个点	183 天

针对不同的数据集,本文设置了不同的 batch-size;对于 MG 数据集,设置 batch-size 为 100;对于 QLD 数据集,设置 batch-size 为 64。学习率统一设置为  $1 \times 10^{-4}$ , epoch 设置为 20000 次。考虑到隐藏层层数和初始权重对网络性能的影响,为了公平起见,本文将隐藏层层数统一设置为 1 层,并对所有网络采用统一的初始权重,即采用均值为 0 的截断正态分布来初始化网络权重。最后,将 CW-RNN 和 ZM-RNN 的隐藏层模块数设为 6,目的是保证实验的一致性,以更好地进行对比。

对于网络参数的优化,我们进行了多次调参,利用验证集选取最优的参数并在测试集上测试。例如,对于隐藏层神经元的数量  $N_h$ ,设置其范围为  $\{12, 24, 36, 48, 60\}$ ;ZM-RNN 的更新阈值的范围为  $\{0.1, 0.15, 0.4, 0.5\}$ 。Zoneout-LSTM 的每个神经元的细胞状态更新阈值  $Z_c$  的范围与  $\epsilon$  一致,隐层状态更新阈值  $Z_h$  的范围设置为  $\{0.05, 0.15, 0.2, 0.4\}$ 。

为了评估模型的性能,采用两种不同的评判指标,分别是平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 和标准化均方根误差 (Normalized Root Mean Square Error, NRMSE),它们是目前比较主流的用于时间序列预测的评判标准,其定义为:

$$MAPE = \left( \frac{|y_t - y_t^*|}{y_t^*} \right) \times 100\% \quad (9)$$

$$NRMSE = \frac{\sqrt{\langle (y_t - y_t^*)^2 \rangle}}{\sqrt{\langle (y_t - \langle y_t^* \rangle)^2 \rangle}} \quad (10)$$

其中,  $\langle \cdot \rangle$  表示求平均值,  $y_t$  表示神经网络的输出值,  $y_t^*$  表示真实值。

## 4.3 实验结果

表 2 列出了 MG 数据集的预测结果,可以看出:首先,传

统的 RNN 相对于 LSTM 和 GRU 来说预测准确率较低,说明它无法解决梯度消失带来的长期依赖问题。其次,近几年提出的 Zoneout-LSTM 和 IndRNN 表现效果并不理想。原因在于根据 IndRNN 的原文,其主要在有多个隐藏层和较大时间步长的情况下具有较好的性能,而在单个隐藏层和较短的时间步长的情况下并不具有优势。而 Zoneout-LSTM 既对细胞状态  $c(t)$  采取随机部分抛弃策略,同时又对隐层状态  $h(t)$  的所有神经元采用随机更新策略,存在极大的随机性,因此一些数据集上容易出现表现不稳定的现象。最后, CW-RNN 的性能明显优于其他类型的 RNN 体系结构,这充分说明了跨时间尺度连接可有效解决梯度消失问题,能够很好地捕获长期时间依赖。而本文提出的结合了 Zoneout 技术的 ZM-RNN 取得了最好的预测精度,说明同样是采用跨时间尺度连接的 RNN 架构,随机的块更新策略比固定的块周期策略具有更好的泛化性能。而且,该模型所需训练的网络参数个数远少于采用门架构的 GRU 和 LSTM,计算效率得到了很大的提升。

表 2 MG 数据集的实验结果

Table 2 Experimental results of MG dataset

Models	MAPE/%	NRMSE
RNN	$5.47 \times 10^{-2}$	$2.28 \times 10^{-3}$
LSTM	$3.18 \times 10^{-2}$	$1.59 \times 10^{-3}$
GRU	$3.47 \times 10^{-2}$	$1.83 \times 10^{-3}$
CW-RNN	$2.63 \times 10^{-2}$	$1.14 \times 10^{-3}$
Zoneout-LSTM	$1.06 \times 10^{-1}$	$4.98 \times 10^{-3}$
IndRNN	$1.26 \times 10^{-1}$	$6.69 \times 10^{-3}$
ZM-RNN	$1.17 \times 10^{-2}$	$5.90 \times 10^{-4}$

为了验证在不同的时间步长上所提模型的有效性,本文在真实数据集上分别进行了时间步长为 31 和 62 的实验,即模型需要学习的数据长度分别为 1 个月和 2 个月,这样既能保证网络学习到近期的数据趋势又能够捕捉到长期的季节性变化。实验结果如表 3 和表 4 所列。

表 3 QLD 数据集的实验结果(时间步长为 31)

Table 3 Experimental results of QLD dataset(timestep=31)

Models	MAPE/%	NRMSE
RNN	2.70	0.728
LSTM	2.12	0.685
GRU	1.82	0.615
CW-RNN	1.83	0.652
Zoneout-LSTM	1.71	0.609
IndRNN	2.13	0.831
ZM-RNN	1.66	0.590

表 4 QLD 数据集的实验结果(时间步长为 62)

Table 4 Experimental results of QLD dataset(timestep=62)

Models	MAPE/%	NRMSE
RNN	2.84	0.789
LSTM	1.98	0.681
GRU	1.85	0.605
CW-RNN	1.85	0.584
Zoneout-LSTM	1.65	0.599
IndRNN	2.26	0.776
ZM-RNN	1.59	0.551

从 QLD 数据集的实验结果可知,首先,随着时间步长从 31 增加到 62,传统 RNN 的预测误差有所增加,这充分说明了 RNN 在处理长期依赖关系时存在不足。其次,GRU, CW-

RNN 和 Zoneout-LSTM 的预测性能均好于 LSTM,说明通过降低网络参数数量有助于提高模型的泛化能力。最后,需要说明的是,在该数据集上 Zoneout-LSTM 的表现效果虽然得到了很大的提升,但本文提出的 ZM-RNN 仍然取得了最好的预测精度,这说明针对神经元的随机更新策略会因为数据集规律性不强而产生较大的性能波动,而基于模块的随机更新策略则更为稳定,泛化性能更好,且计算效率更高。

**结束语** 本文设计实现了一种基于 Zoneout 的跨尺度循环神经网络(ZM-RNN),并在时间序列预测的基准数据集和真实的电力负荷数据集中验证了该模型的可靠性。相比现有的循环神经网络模型,ZM-RNN 存在以下两点优势:1)通过在跨尺度循环网络中引入 Zoneout 技术,利用随机概率策略代替了传统的固定周期策略,提高了网络模型的泛化能力;2)对于传统的 Zoneout 技术以神经元为单位进行随机更新而言,ZM-RNN 是以隐藏层模块(模块内包含多个神经元)为对象进行工作的,避免了过度随机带来的网络性能不稳定的问题。需要说明的是,ZM-RNN 的性能仍然会受部分随机性和更新概率阈值等超参数的影响,在未来的工作中,我们将进一步研究跨尺度循环网络中模块更新的自适应策略,以期能进一步提高网络的性能。

## 参 考 文 献

- [1] DU D, CHEN R, LI X, et al. Malicious data deception attacks against power systems: A new case and its detection method[J]. Transactions of the Institute of Measurement and Control, 2019, 41(6): 1590-1599.
- [2] NATARAJA C, GORAWAR M B, SHILPA G N, et al. Short term load forecasting using time series analysis: a case study for Karnataka, India [J]. Int. J. Eng. Sci. Innov. Technol, 2012, 1(2): 45-53.
- [3] HERNÁNDEZ L, BALADRÓN C, AGUIAR J, et al. Artificial neural network for short-term load forecasting in distribution systems[J]. Energies, 2014, 7(3): 1576-1598.
- [4] SONG K B, BAEK Y S, HONG D H, et al. Short-term load forecasting for the holidays using fuzzy linear regression method [J]. IEEE Transactions on Power Systems, 2005, 20(1): 96-101.
- [5] WANG Z, YANG F, HO D W C, et al. Stochastic dynamic modeling of short gene expression time-series data[J]. IEEE Transactions on Manobioscience, 2008, 7(1): 44-55.
- [6] WEI G, WANG Z, SHU H, et al. Robust filtering for gene expression time series data with variance constraints[J]. International Journal of Computer Mathematics, 2007, 84(5): 619-633.
- [7] AL-HAMADI H M, SOLIMAN S A. Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model[J]. Electric Power Systems Research, 2004, 68(1): 47-59.
- [8] RAHMAN S, BHATNAGAR R. An expert system based algorithm for short term load forecast[J]. IEEE Transactions on Power Systems, 1988, 3(2): 392-399.
- [9] WEI G, FENG G, WANG Z. Robust Control for Discrete-Time Fuzzy Systems With Infinite-Distributed Delays [J]. IEEE

Transactions on Fuzzy Systems, 2008, 17(1): 224-232.

- [10] LEE C M, KO C N. Short-term load forecasting using lifting scheme and ARIMA models[J]. Expert Systems with Applications, 2011, 38(5): 5902-5911.
- [11] LIU Y, WANG Z, LIU X. Asymptotic stability for neural networks with mixed time-delays: The discrete-time case[J]. Neural Networks, 2009, 22(1): 67-74.
- [12] LIU L, SHEN B, WANG X. Research on kernel function of support vector machine [M] // Advanced Technologies, Embedded and Multimedia for Human-centric Computing. Springer, Dordrecht, 2014: 827-834.
- [13] BAI Y. Design of Cluster Analysis Model Based on Load Characteristic Curve of Power Consumers [J]. Journal of Chongqing University of Technology (Natural Science), 2018, 32(12): 181-185.
- [14] HIPPERT H S, PEDREIRA C E, SOUZA R C. Neural networks for short-term load forecasting: A review and evaluation[J]. IEEE Transactions on Power Systems, 2001, 16(1): 44-55.
- [15] LÄNGKVIST M, KARLSSON L, LOUTFI A. A review of unsupervised feature learning and deep learning for time-series modeling[J]. Pattern Recognition Letters, 2014, 42: 11-24.
- [16] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2): 157-166.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [18] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. arXiv:1412.3555, 2014.
- [19] KOUTNIK J, GREFF K, GOMEZ F, et al. A clockwork rnn[J]. arXiv:1402.3511, 2014.
- [20] CHANG S, ZHANG Y, HAN W, et al. Dilated recurrent neural networks[C] // Advances in Neural Information Processing Systems. 2017: 77-87.
- [21] LI S, LI W, COOK C, et al. Independently recurrent neural network (indrnn): Building a longer and deeper rnn[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5457-5466.
- [22] KRUEGER D, MAHARAJ T, KRAMÁR J, et al. Zoneout: Regularizing rnns by randomly preserving hidden activations[J]. arXiv:1606.01305, 2016.



**ZHUANG Shi-jie**, born in 1995, undergraduate. His main research interests include computational intelligence, machine learning and deep learning.



**HUANG Fang-wan**, born in 1980, senior lecturer, is a member of China Computer Federation. Her main research interests include computational intelligence, machine learning and big data analysis.