

基于周期性建模的时间序列预测方法及电价预测研究

徐任超¹ 阎威武^{1,2} 王国良³ 杨健程¹ 张曦⁴

摘要 时间序列数据广泛存在于人类的生产生活中,通常具有复杂的非线性动态和一定的周期性.与传统的时间序列分析方法相比,基于深度学习的方法更能捕捉数据的深层特性,对具有复杂非线性的时间序列有较好的建模效果.为了在神经网络中显式地建模时间序列数据的周期性和趋势性,本文在循环神经网络的基础上引入了周期损失和趋势损失,建立了基于周期性建模和多任务学习的时间序列预测模型.将模型应用到欧洲能源交易所法国市场的能源市场价格预测中,结果表明周期损失和趋势损失能够提高神经网络的泛化能力,并提高预测时间序列趋势的精度.

关键词 时间序列预测,深度学习,循环神经网络,周期趋势建模

引用格式 徐任超,阎威武,王国良,杨健程,张曦.基于周期性建模的时间序列预测方法及电价预测研究.自动化学报,2020,46(6): 1136–1144

DOI 10.16383/j.aas.c180712

Time Series Forecasting Based on Seasonality Modeling and Its Application to Electricity Price Forecasting

XU Ren-Chao¹ YAN Wei-Wu^{1,2} WANG Guo-Liang³ YANG Jian-Cheng¹ ZHANG Xi⁴

Abstract Time series data exist widely in human production and life. The real time series data often contain complex nonlinear dynamics and seasonality. Compared with traditional time series analysis methods, deep learning based methods have good modeling effect for the time series with complex nonlinearities but fail to model the seasonality and trend of time series. In order to model the seasonality and trending explicitly in neural networks, this paper introduces seasonal loss and trend loss into recurrent neural networks (RNNs), establishing the time series prediction model based on seasonality modeling and multi-task learning. The suggested method is then applied to the electricity price forecasting on EPEX (European Power Exchange) France market. The experiment results show that seasonal loss and trend loss can improve the generation ability of neural networks and the performance of sequence trend forecasting.

Key words Time series forecasting, deep learning, recurrent neural networks (RNNs), seasonality and trend modeling

Citation Xu Ren-Chao, Yan Wei-Wu, Wang Guo-Liang, Yang Jian-Cheng, Zhang Xi. Time series forecasting based on seasonality modeling and its application to electricity price forecasting. *Acta Automatica Sinica*, 2020, 46(6): 1136–1144

时间序列数据在现实生活中广泛存在,例如金融领域中的交易数据和经济统计数据、消费电商领

域中的用户浏览和购买数据、医疗领域中的医疗器械的信号记录、天气监测站记录的天气指标数据等^[1–4].这些时间序列数据是相应领域非常宝贵的数据资源,对这些数据的准确、有效分析和利用有助于减小人力成本,提高生产效率,提高经济收益^[5].

现实中的时间序列通常具有复杂的非线性动态,这为时间序列预测带来了困难.另外,由于人类活动或自然运动的影响,时间序列数据常常体现出一定的周期性.周期性的提取对时间序列预测有着积极意义.时间序列的趋势同样具有重要的意义,有时甚至作为预测的目的.

传统的时间序列分析方法源自于自回归模型 (Autoregressive model, AR) 和移动平滑模型 (Moving average, MA).自回归移动平滑模型 (Autoregressive model and moving average, ARMA) 和在

收稿日期 2018-10-31 录用日期 2019-06-02

Manuscript received October 31, 2018; accepted June 2, 2019

国家重点研究发展计划基金 (2019YFB1705702), 国家自然科学基金 (60974119, 61533013) 资助

Supported by National Key Research and Development Program of China (2019YFB1705702) and National Natural Science Foundation of China (60974119, 61533013)

本文责任编辑 刘艳军

Recommended by Associate Editor LIU Yan-Jun

1. 上海交通大学电子信息与电气工程学院 上海 200240 2. 系统控制与信息处理教育部重点实验室 上海 200240 3. 上海工程技术大学 电子电气工程学院 上海 201620 4. 南方电网国际有限责任公司 广州 510630

1. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240 2. Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240 3. Shanghai University of Engineering Science, Shanghai 201620 4. China Southern Power Grid International Co., Ltd., Guangzhou 510630

其基础上发展起来的自回归差分移动平滑模型 (Autoregressive integrated moving average, ARIMA) 是时间序列分析的经典方法^[6]。另外, 自回归条件异方差模型 (Autoregressive conditional heteroscedasticity, ARCH)^[7] 和广义自回归条件异方差模型 (Generalized autoregressive conditional heteroscedasticity, GARCH)^[8] 引入了异方差, 对时间序列的波动性进行建模。基于机器学习的支持向量机回归 (Support vector regression, SVR) 和核岭回归 (Kernel ridge regression, KRR) 等方法在时间序列分析上已经有大量的研究^[9-11]。随着深度学习理论的发展, 循环神经网络 (Recurrent neural networks, RNNs) 已经成为处理时间序列数据的主流, 在大量应用场景中取得了不俗的效果^[12-13]。回声状态网络 (Echo state network, ESN) 也是时间序列预测的常用模型^[14]。

传统的时间序列分析方法基于平稳性假设^[15], 对短期平稳的时间序列有较好的预测效果, 但是难以对复杂非线性时间序列数据进行有效建模。SVR 和 KRR 等机器学习方法使用滑动窗口的方式处理预测序列, 忽略了时间序列数据的序列属性, 不能很好地捕捉时间序列中的长时依赖。同时 SVR 和 KRR 受限于模型容量, 难以达到预测非线性时间序列的最佳效果。深度神经网络由于其超强的拟合能力, 在复杂非线性时间序列数据的处理方面有着天然的优势。同样对于分割后的序列窗口, 深度神经网络仍能从中捕捉时间序列的长期依赖。

能源领域是产生大量时间序列的领域, 如电力用量数据、风力序列数据、太阳能数据和电力价格数据等。其中, 电力价格是能源市场上的关键因素, 影响着能源市场的流通和运行。能源市场的自由属性使电力价格具备了金融商品性质, 但其还受到电力传输和需求量的影响。电力传输受电网传输容量的限制, 而电力需求量受人类活动和天气因素的影响。在这些影响因素下, 电力价格数据呈现出长期复杂非线性动态, 体现出高波动性。另外电力价格还存在着明显的周期性。

传统的时间序列预测方法对电力价格的预测已有大量研究和应用^[16], 而循环神经网络在电力价格预测中同样取得了不错的效果^[17]。针对能源领域时间序列的周期性属性, 传统的方法主要采用信号处理方法, 而常规的循环神经网络中并没有对时间序列的周期性进行建模的模块。目前, 许多研究针对数据的周期性展开, 主要研究工作有: Clements 等^[18] 提出了多方程时间序列方法, 在方程中加入了周期设定, 对澳洲能源市场的电力负载进行预测; Anbazhagan 等^[19] 使用离散余弦变换处理原始序列, 并用神经网络进行建模; Rafiei 等^[20] 使用小波

变换处理原始序列, 并用极限学习机对分解结果进行建模等。

针对循环神经网络难以对时间序列数据的周期性直接进行建模的问题, 本文结合时间序列周期分解的思想, 设计了循环神经网络的周期损失和趋势损失, 分别对时间序列中的周期和趋势进行辅助建模; 在多任务学习框架下, 将周期损失、趋势损失和模型自身的损失函数相结合, 联合优化循环神经网络。提出的模型用于能源市场的电力价格预测, 实验结果取得了较好的预测精度, 验证了周期和趋势对时间序列预测的作用, 说明本文提出的周期损失和趋势损失能够有效地辅助模型捕捉数据特征。

1 循环神经网络

循环神经网络是常用的建模时间序列的神经网络模型。循环神经网络使用基于时间的反向传播算法 (Back-propagation through time, BPTT) 来训练网络。

1.1 基础循环单元

基础的循环神经网络及其展开形式如图 1 所示。其数学形式为

$$h_t = f(Ux_t + Wh_{t-1}) \quad (1)$$

$$o_t = g(Vh_t) \quad (2)$$

其中, U , V , W 都是网络的参数, f 和 g 表示激活函数, 可以是 ReLU、sigmoid 和 tanh 等非线性函数。 x_t 表示 t 时刻的输入, h_t 表示 t 时刻网络的隐藏状态, o_t 表示 t 时刻网络的输出。

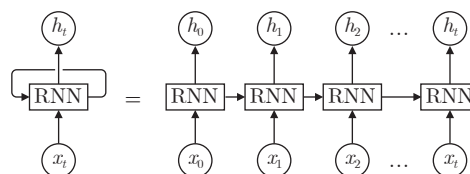


图 1 一个展开的基础循环神经网络

Fig.1 The unfold structure of RNN

基础的循环神经网络在训练时会面对梯度爆炸或梯度消失的问题^[21]。

1.2 长短时记忆网络

为了克服循环神经网络的梯度爆炸和梯度消失问题, Hochreiter 等^[22] 提出了长短时记忆网络 (Long short term memory, LSTM)。LSTM 的内部结构如图 2 所示。

LSTM 的主要思想是引入门控单元和长时记忆单元。门控单元负责控制长时记忆单元中状态的

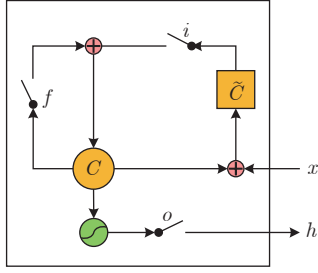


图 2 LSTM 的内部结构示意图

Fig.2 The schematic diagram of LSTM

记忆、修改和遗忘. 同时, LSTM 还具有和基础循环神经网络相同的短期记忆单元.

LSTM 的数学表示为

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (4)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (5)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (6)$$

$$C_t = i_t \tilde{C}_t + f_t C_{t-1} \quad (7)$$

$$h_t = o_t \tanh(C_t) \quad (8)$$

其中, C_t 和 $\tilde{C}_t$ 分别表示 t 时刻的长期记忆和短期记忆, f_t , i_t 和 o_t 分别表示 LSTM 的遗忘门、输入门和输出门在 t 时刻的值. (W_f, b_f) , (W_i, b_i) , (W_o, b_o) 和 (W_C, b_C) 分别表示 LSTM 的遗忘门、输入门、输出门和长时记忆单元的参数. σ 代表 sigmoid 函数. x_t 表示 t 时刻的输入, h_t 表示 t 时刻隐藏状态的输出.

由于引入了门控单元和长时记忆单元, 状态的更新方式从乘性增量更新变为了加性增量更新^[23], 从而一定程度上避免了梯度消失问题.

1.3 门控循环单元

门控循环单元 (Gate recurrent unit, GRU)^[24] 是 LSTM 的一个重要变体. 相比于 LSTM, GRU 进行了一定的简化: 将遗忘门和输入门合并成为一个门控单元, 称为更新门; 去除了长期记忆单元, 与隐藏状态混合, 保留短期记忆单元; 去除了输出门, 改为重置门. GRU 的内部结构如图 3 所示.

GRU 的数学表示为

$$z_t = \sigma(W_z[h_{t-1}, x_t]) \quad (9)$$

$$r_t = \sigma(W_r[h_{t-1}, x_t]) \quad (10)$$

$$\tilde{h}_t = \tanh(W[r_t \times h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - z_t) h_{t-1} + z_t \tilde{h}_t \quad (12)$$

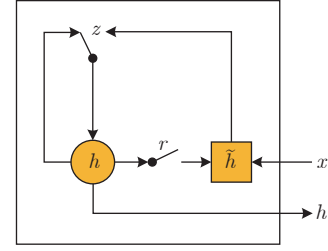


图 3 GRU 的内部结构示意图

Fig.3 The schematic diagram of GRU

其中, $\tilde{h}_t$ 表示 t 时刻的短期记忆, h_t 表示 t 时刻的隐藏状态. z_t 和 r_t 分别表示 GRU 的更新门和重置门在 t 时刻的值. W_z , W_r 和 W 分别表示 GRU 的更新门、重置门和短期记忆单元的参数.

GRU 大大精简了参数, 相同隐藏状态的 GRU 的参数量约为 LSTM 的 2/3, 使得模型更不容易过拟合. 本文使用 GRU 作为默认的循环神经网络模型.

2 基于周期损失和趋势损失的时间序列预测

结合时间序列周期分解, 设计了循环神经网络的周期损失和趋势损失, 将周期损失、趋势损失和模型自身的损失函数相结合, 在多任务学习框架下, 联合优化循环神经网络.

2.1 周期分解

周期分解是一种对时间序列的周期和趋势建模的经典方式.

周期建模是指, 针对周期性的时间序列, 可以显式的考虑序列跨周期的关系. 周期性的一种形式化表述为

$$ACF(x_t, x_{t-T}) \gg ACF(x_t, x_{t-\tau}) \quad \forall \tau : \text{mod}(\tau, T) \neq 0 \quad (13)$$

其中, ACF 代表自相关函数. 周期序列按照周期的延迟序列的自相关性应该远远大于其他的延迟序列.

趋势建模是指除了建模每一个具体时间点的具体数值, 还应该捕捉到一段时间的“整体状态”, 包括总体走势、变化范围等因素.

可以认为, 时间序列是趋势、周期和余差的组合. 时间序列的分解可以是加性的, 也可以是乘性的, 即

$$X_t = T_t + S_t + R_t \quad (14)$$

$$X_t = T_t \times S_t \times R_t \quad (15)$$

其中, X_t 表示原始时间序列, T_t 表示趋势分量, S_t 表示周期分量, R_t 表示余差. 乘性的组合可以通过

对数函数转换为加性组合, 所以这两种组合实质上是一致的。

2.2 周期损失函数和趋势损失函数

时间序列的周期分解启发了周期损失函数和趋势损失函数的提出。

周期损失函数鼓励循环神经网络预测的隐藏状态出现周期性。假设一个循环神经网络的隐藏状态具有周期性, 它的隐藏状态需要满足一定的条件:

$$\cos nt = h_t + h_{t+1} + \cdots + h_{t+T-1} \cong h_{t+1} + h_{t+2} + \cdots + h_{t+T} \quad (16)$$

即有

$$h_t \cong h_{t+T} \quad (17)$$

为了在循环神经网络中加入这样的隐藏状态关系, 在 $[t_1, t_2, \cdots, t_n]$ 的时间窗上定义如下的周期损失函数

$$\ell_{\text{seasonal}}(kT) = \frac{1}{n} \sum_{t=t_1}^{t_n} D(h_t, h_{t+\tau}) \quad (18)$$

其中, $D(*, *)$ 表示距离度量函数, τ 表示隐藏状态之间的时间跨度。

针对周期为 T 的时间序列, 一般取 $\tau = kT$ 。 τ 也可以取 1, 这样式 (18) 可以视为循环神经网络的一种正规化方法。 $D(\cdot, \cdot)$ 一般取平方距离, 此时的周期损失为

$$\ell_{\text{seasonal}}(kT) = \frac{1}{nm} \sum_{t=t_1}^{t_n} \sum_{i=1}^m \|h_t^{(i)} - h_{t+kT}^{(i)}\|^2 \quad (19)$$

其中, m 表示循环神经网络的隐藏状态的维度, $h_t^{(i)}$ 表示 t 时刻隐藏状态的第 i 维的值。更长周期跨度有助于鼓励长距离的周期平稳性。

在很多实际应用中, 每个时间点的数值并不是最受关注的问题, 而关注的重点在于捕获某种趋势。时间序列的趋势可以由一段时间内的均值、最大最小值和波动率来反映。

在 $[t_1, t_2, \cdots, t_n]$ 的时间窗上对时间序列的趋势进行建模, 从均值、最大最小值和波动率的角度可以得到如下的几个趋势损失函数

$$\ell_{\text{trend}}^{\text{MEAN}}(w) = \frac{1}{n} \sum_{t=t_1}^{t_n} \|\text{mean}(\hat{y}_{t-w+1}, \cdots, \hat{y}_t) - \text{mean}(y_{t-w+1}, \cdots, y_t)\|^2 \quad (20)$$

$$\ell_{\text{trend}}^{\text{MAX}}(w) = \frac{1}{n} \sum_{t=t_1}^{t_n} \|\max(\hat{y}_{t-w+1}, \cdots, \hat{y}_t) - \max(y_{t-w+1}, \cdots, y_t)\|^2 \quad (21)$$

$$\ell_{\text{trend}}^{\text{MIN}}(w) = \frac{1}{n} \sum_{t=t_1}^{t_n} \|\min(\hat{y}_{t-w+1}, \cdots, \hat{y}_t) - \min(y_{t-w+1}, \cdots, y_t)\|^2 \quad (22)$$

$$\ell_{\text{trend}}^{\text{VAR}}(w) = \frac{1}{n} \sum_{t=t_1}^{t_n} \|\text{var}(\hat{y}_{t-w+1}, \cdots, \hat{y}_t) - \text{var}(y_{t-w+1}, \cdots, y_t)\|^2 \quad (23)$$

其中, $\ell_{\text{trend}}^{\text{MEAN}}$ 、 $\ell_{\text{trend}}^{\text{MAX}}$ 、 $\ell_{\text{trend}}^{\text{MIN}}$ 和 $\ell_{\text{trend}}^{\text{VAR}}$ 分别代表均值趋势损失、最大趋势损失、最小趋势损失和波动趋势损失。 w 表示趋势窗宽, 一般与序列的时间窗宽度 n 不同。 $\hat{y}$ 是循环神经网络的预测值, y 是真实值。

2.3 基于多任务学习的混合预测模型

在机器学习中, 通常只关注一个优化目标。多任务学习 (Multi-task learning)^[25] 是同时学习多个相关的任务并在这些任务间共享某些底层的特征表示的机器学习方式。在深度学习中, 多任务学习非常普遍。

一般地, 针对多个优化目标, 多任务学习框架中存在共同优化和分开交替优化两种优化方式。共同优化适用于紧密相关的任务, 分开交替优化则适用于不相关任务和对抗性任务。周期性和趋势约束作为时间序列预测的辅助任务, 与预测任务关系密切。因此, 将周期损失和趋势损失引入优化目标进行共同优化是合理的。

本文根据多任务学习的框架, 将周期损失和趋势损失将作为预测任务的共同优化目标, 在原来的损失函数的基础上, 加入了辅助损失函数, 得到时间序列多任务学习混合模型的优化目标函数

$$J = L_{\text{main}} + \lambda_S \ell_{\text{seasonal}} + \ell_{\text{trend}} \quad (24)$$

$$\ell_{\text{trend}} = \lambda_T^{\text{MEAN}} \ell_{\text{trend}}^{\text{MEAN}} + \lambda_T^{\text{MAX}} \ell_{\text{trend}}^{\text{MAX}} + \lambda_T^{\text{MIN}} \ell_{\text{trend}}^{\text{MIN}} + \lambda_T^{\text{VAR}} \ell_{\text{trend}}^{\text{VAR}} \quad (25)$$

其中, λ_S 是周期损失的权重, λ_T^{MEAN} , λ_T^{MAX} , λ_T^{MIN} 和 λ_T^{VAR} 分别代表对应的趋势损失的权重。

引入周期损失和趋势损失的循环神经网络依然可以通过 BPTT 算法来训练。

3 仿真实验

3.1 能源市场简介

欧洲能源交易所 (European Power Exchange, EPEX) 是欧洲最重要的能源交易市场, 覆盖了德国、法国、英国等欧洲最主要的国家。电力在各个国家的电网之间进行交换, 其中一部分是通过事先规划的方式, 而另一部分则通过能源交易所进行交易。

每个国家都会有自己的子市场,例如 EPEX 法国市场. EPEX 法国市场的现货市场主要分为三个子市场: 日前交易市场 (Day-ahead market)、日内交易市场 (Intraday market) 和再平衡市场 (Balancing market). 能源市场的实际交易主要发生在日前交易市场. 通常 EPEX 法国市场的前日交易市场每天形成 24 个报价, 即每小时一个报价.

图 4 展示了 EPEX 法国市场 2012 年~2015 年的一天不同小时的价格的均值和标准差. 可以看出, 电力价格的标准差非常高, 在均值的 1/4 以上, 在个别时间段 (9:00~11:00) 的标准差甚至超出了均值. 这说明 EPEX 法国市场的电力价格具有超高的波动性. 图 5 和图 6 分别展示了 EPEX 法国市场 2012 年~2015 年每年的日级别平均价格和周级别平均价格. 可以看出, 无论是按日的级别还是周的级别对电价进行统计, 每年的电力价格均呈现非常相似的走势, 体现出明显的周期性.

3.2 实验及结果

选用 EPEX 法国市场的电力价格数据作为仿真来源. 数据来自日前电力市场, 时间段取自 2012 年 1 月 1 日~2016 年 6 月 30 日, 其中

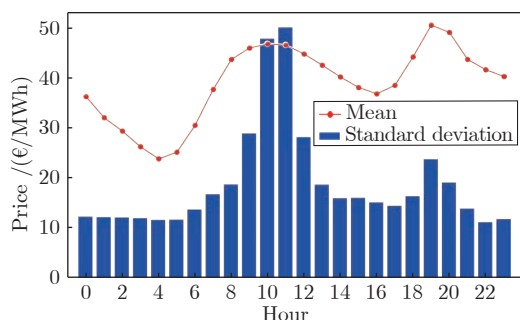


图 4 能源市场价格的高波动性
Fig.4 High volatility of electricity price

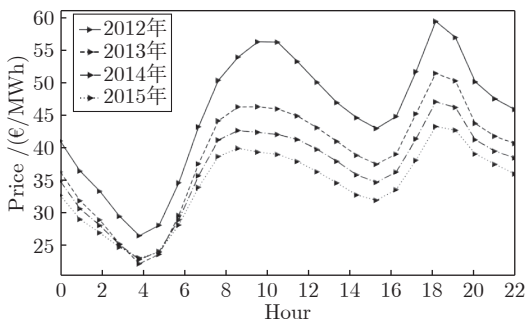


图 5 能源市场价格的日级别周期性
Fig.5 The daily seasonality of electricity price

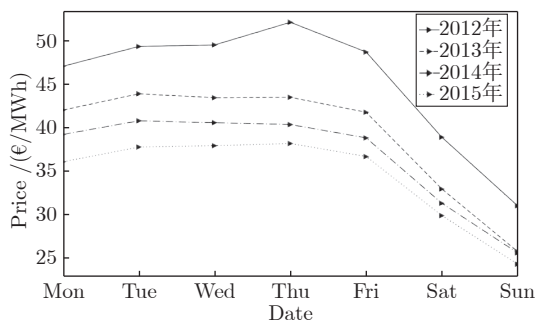


图 6 能源市场价格的周级别周期性
Fig.6 The weekly seasonality of electricity price

2012 年 1 月 1 日~2015 年 12 月 31 日 (共 1 461 天) 的序列数据用来做训练和验证, 2016 年 1 月 1 日~2016 年 6 月 30 日 (共 182 天) 的序列数据用来做测试. 通常每天有 24 个时间点 (夏令时转换日有 23 个时间点, 冬令时转换日有 25 个时间点).

为了保证数据的稳定性, 需要对数据进行预处理, 去除数据的均值和波动率的影响

$$\tilde{x}_t^h = \frac{x_t^h - \bar{x}_t}{std(x)} \quad (26)$$

其中, x 代表所有的序列数据, h 表示序列中的年份信息 (2012 年~2016 年), x_t^h 表示不同年份的在同一时间点 t 的序列值, $\bar{x}_t$ 表示 x_t^h 按年份展开的子序列的均值, $\tilde{x}_t^h$ 为处理后的序列值. 经过这样的变换, 大部分时间点的数据处于 $[-1, 1]$ 区间, 可以避免出现一些数值问题.

实验对比了几种基础的循环神经网络: RNN、LSTM 和 GRU. 为了分析周期损失和趋势损失的作用, 在 GRU 的基础上分别添加周期损失、趋势损失或者两者都添加.

循环神经网络需要调整的超参数有: 隐层大小、优化器、初始学习率、批大小 (batch size)、训练轮数 (epochs) 以及延迟窗宽. 针对引入周期损失和趋势损失的 GRU, 还需要调整的有: 周期损失系数 λ_S 、周期损失中的周期跨度 kT 、趋势损失系数 (λ_{trend}^{VAR} 、 λ_{trend}^{MEAN} 、 λ_{trend}^{MAX} 和 λ_{trend}^{MIN}) 和趋势窗宽 w . 这里考虑能源市场的短期周期, 故 k 取 1, T 取 24. 由于能源市场的价格预测任务往往不关注每日的平均价格和价格的波动预测的误差, λ_{trend}^{MEAN} 和 λ_{trend}^{VAR} 的值都预设为 0, 不再纳入超参数选择中. 能源市场的价格预测比较关注一天内的最大最小值, 因此趋势窗宽 w 设为 24. 训练时间段的最后一个月 (2015 年 12 月) 作为验证时间段. 通过在验证时间段内进行多次实验的方式来选取超参数. 选出来的超参数将

被用于在完整的训练时间段上的训练. 循环神经网络的超参数设置如表 1 所示.

表 1 循环神经网络的超参数设置
Table 1 The hyperparameters of RNN

超参数	具体取值
隐层大小	64
优化器	RMSProp, 配合梯度裁剪
初始学习率	0.001
批大小	64
训练轮数	12
延迟窗宽	14

表 2 列出了周期损失和趋势损失的合理的取值范围.

表 2 周期损失和趋势损失的权重范围
Table 2 Weights range of seasonal loss and trend loss

权重	取值范围
λ_S	0.05~0.15
λ_T^{MEAN}	0
λ_T^{MAX}	0.05~0.1
λ_T^{MIN}	0.05~0.1
λ_T^{VAR}	0

除了与基础的循环神经网络进行对比, 另外还选取了 ARIMA、SVR 和 KRR 作为对比模型. SVR 和 KRR 的超参数均通过在训练时间段内的 5 折交叉验证直接选取, 均选用 3 阶多项式核. 为了满足 ARIMA 模型的静态性要求, 电力价格数据还需要经过特殊处理, 即将序列减去最小值后取对数, 然后进行天级和周级的延迟差分

$$p' = (1 - \beta^{24})(1 - \beta^{168}) \ln(p - \min p + 1) \quad (27)$$

实验采用均方根误差 (Root mean square error, RMSE) 和平均绝对误差 (Mean absolute error, MAE) 作为评价指标

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2} \quad (28)$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |\hat{y}_t - y_t| \quad (29)$$

其中, N 表示测试时间段的时间点数. RMSE 也是各模型训练时的损失指标.

为了更合理地比较趋势损失的影响, 本文还引入了两个额外指标, 最大趋势绝对误差 MAE^{MAX} 和最小趋势绝对误差 MAE^{MIN} , 分别考察模型对每天电力价格最大值和最小值的预测能力

$$MAE^{\text{MAX}} = \frac{1}{M} \sum_{t=1}^M |\max(\hat{y}_{t-w+1}, \dots, \hat{y}_t) - \max(y_{t-w+1}, \dots, y_t)| \quad (30)$$

$$MAE^{\text{MIN}} = \frac{1}{M} \sum_{t=1}^M |\min(\hat{y}_{t-w+1}, \dots, \hat{y}_t) - \min(y_{t-w+1}, \dots, y_t)| \quad (31)$$

其中, M 表示测试时间段划分的趋势窗的数目.

由于神经网络在训练过程中具有一定的随机性, 每次训练的结果都有所不同. 为了保证实验结果的可信性, 对所有循环神经网络采用了重复 10 次训练并取测试结果的平均值和标准差. 各方法在能源价格预测的结果见表 3.

表 3 各种方法的能源价格预测效果对比
Table 3 The result comparisons of different methods for electricity price forecasting

模型	RMSE	MAE	MAE^{MAX}	MAE^{MIN}
ARIMA	6.41	4.77	5.15	4.82
SVR	4.91	3.71	4.27	3.34
KRR	5.14	3.75	3.81	3.78
RNN	5.09±0.24	3.75±0.19	3.72±0.28	3.78±0.19
LSTM	4.90±0.18	3.65±0.17	3.65±0.42	3.61±0.26
GRU	4.83±0.19	3.54±0.06	3.64±0.31	3.56±0.26
GRU, $\lambda_S = 0.1$, $\lambda_T^{\text{MAX}} = 0$, $\lambda_T^{\text{MIN}} = 0$	4.71±0.16	3.49±0.13	3.53±0.28	3.53±0.15
GRU, $\lambda_S = 0.05$, $\lambda_T^{\text{MAX}} = 0$, $\lambda_T^{\text{MIN}} = 0$	4.74±0.11	3.45±0.18	3.53±0.23	3.48±0.26
GRU, $\lambda_S = 0$, $\lambda_T^{\text{MAX}} = 0.1$, $\lambda_T^{\text{MIN}} = 0.1$	4.85±0.16	3.57±0.20	3.41±0.26	3.41±0.18
GRU, $\lambda_S = 0$, $\lambda_T^{\text{MAX}} = 0.05$, $\lambda_T^{\text{MIN}} = 0.05$	4.83±0.11	3.54±0.08	3.39±0.18	3.42±0.15
GRU, $\lambda_S = 0.1$, $\lambda_T^{\text{MAX}} = 0.1$, $\lambda_T^{\text{MIN}} = 0.1$	4.68±0.08	3.45±0.03	3.35±0.13	3.33±0.12
GRU, $\lambda_S = 0.05$, $\lambda_T^{\text{MAX}} = 0.05$, $\lambda_T^{\text{MIN}} = 0.05$	4.60±0.15	3.34±0.12	3.38±0.13	3.27±0.11

从实验结果可以看出, 对于电力价格预测这类复杂的时间序列预测问题, ARIMA 模型的预测效果不佳, SVR、KRR 和 RNN 的效果接近且远远优于 ARIMA 模型 LSTM 和 GRU 的预测效果比 SVR 和 KRR 更好, 而基础的 GRU 在对比模型中效果最佳. 添加周期损失能对 GRU 起到比较好的辅助效果, 有助于降低模型的预测误差, 提升预测效果. 趋势损失最直接的影响是显著降低最大趋势误差和最小趋势误差, 说明模型通过趋势损失捕捉到了数据在时间窗内的趋势信息 (包括电力价格的极值、均值和变化率等), 这对模型捕捉数据特征的能力起到了有力的补充. 另外可以注意到的是, 加入周期损失也能够降低最大趋势误差和最小趋势误差. 这说明周期损失有助于网络捕捉一些数据的本

质特征. 实验结果表明: 同时引入周期损失和趋势损失可以使 GRU 的性能达到最佳.

引入周期损失和趋势损失的 GRU 在最佳的参数下 ($\lambda_S = \lambda_T^{\text{MAX}} = \lambda_T^{\text{MIN}} = 0.05$) 的预测结果如图 7 所示.

可以看出, 引入周期损失和趋势损失的 GRU 对电力价格数据的预测效果不错, 预测曲线对真实数值的走势和变化跟踪准确. 预测曲线具有非常好的周期性, 这体现了周期损失中周期性约束的作用. 同时, 预测曲线更偏向于平稳的预测, 这在一定程度上避免了噪声带来的预测性能损失. 另外, 预测曲线在正常情况下的最大最小值的预测效果很好, 说明趋势损失能够起到有效的监督效果. 但可以观察到的是, 模型在一些极端值 (电力价格的异常攀

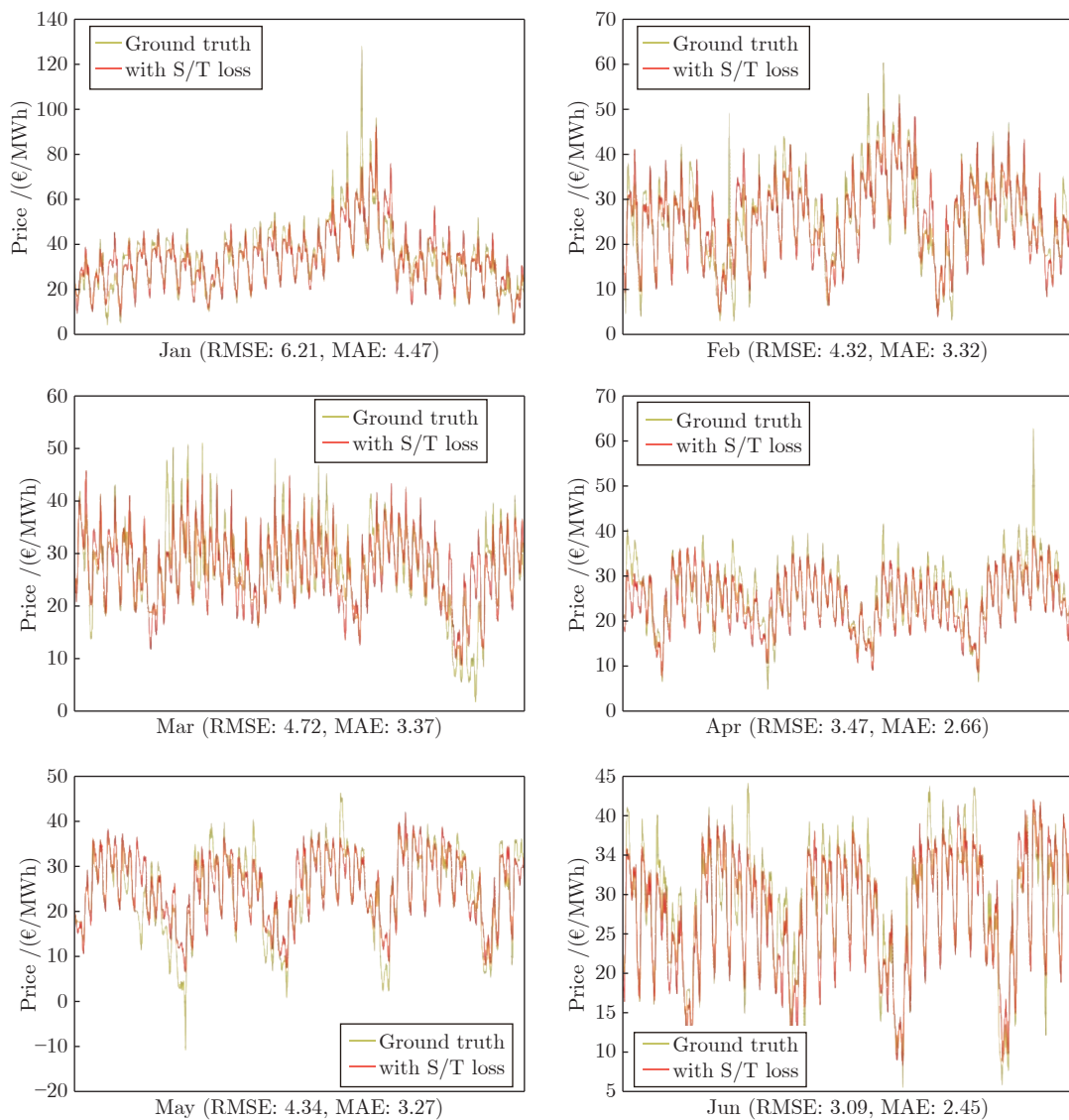


图 7 加入了周期损失和趋势损失的 GRU 模型的预测结果

Fig. 7 The results of electricity price forecasting with seasonal loss and trend loss

升和下降) 上的预测效果不佳. 这涉及到突变的预测问题, 在现有模型框架下难有较好的预测效果.

为了进一步分析周期损失的效果, 将 GRU 和加入周期损失的 GRU 在训练一段时间之后的隐藏状态 h_t 分别抽出, 并将向量 h_t 的均值作对比, 部分结果如图 8 所示.

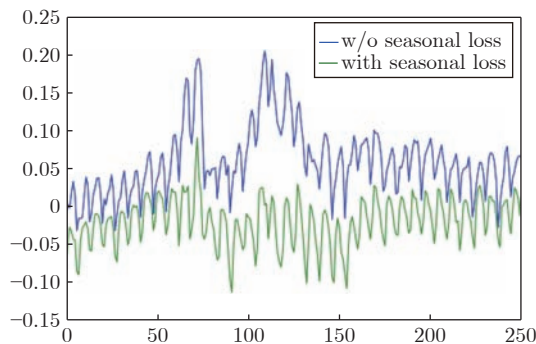


图 8 GRU 在有无周期损失时的隐藏状态均值的比较

Fig.8 Hidden states of GRU with and without the seasonality

由于时间序列本身具有一定的周期性, 使得隐藏状态序列都表现出了一定的周期性. 但是, 使用了周期损失的 GRU 的隐藏状态明显更加平稳, 并且每个周期内的数值更加平滑; 而未使用周期损失的 GRU 的隐藏状态的波动性更大, 并且单个周期内的波动也更大. 这些更大的波动可以认为是捕捉到了更多的噪声. 所以, 周期损失有助于捕捉数据更本质的特征而非噪声.

4 结论

循环神经网络能够提取时间序列中的深层信息, 是非常有效的时序预测方法. 本文通过定义周期损失和趋势损失的合理形式, 试图在循环神经网络内对时间序列的周期和趋势进行建模, 改善模型的性能, 建立了基于周期性建模和多任务学习的时间序列预测模型. 在能源市场价格预测的任务的仿真实验表明: 本文提出的周期损失能够有效引导循环神经网络学习数据的周期性, 提升预测效果, 降低数据噪声的影响. 周期损失还有助于模型捕捉数据的本质特征. 趋势损失则能对提升趋势相关的预测任务的预测效果.

References

- Budish E, Cramton P, Shim J. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 2015, **130**(4): 1547-1621
- He J, Li X, Liao L J, Song D D, Cheung W K. Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence. AAAI Press, 2016. 137-143
- Ren Y F, Wu Y. Convolutional deep belief networks for feature extraction of EEG signal. In: Proceedings of the 2014 International Joint Conference on Neural Networks. Beijing, China: IEEE, 2014. 2850-2853
- Zhang Xi-Lai, Zhao Jian-Hui, Cai Bo. Prediction model with dynamic adjustment for single time series of PM2.5. *Acta Automatica Sinica*, 2018, **44**(10): 1790-1798
(张熙来, 赵俭辉, 蔡波. 针对 PM2.5 单时间序列数据的动态调整预测模型. *自动化学报*, 2018, **44**(10): 1790-1798)
- Keogh E, Chu S, Hart D, Pazzani M. *Segmenting time series: A survey and novel approach*. Data mining in time series databases. 2004.
- Box G E P, Jenkins G M. Time series analysis: Forecasting and control. *Journal of the Operational Research Society*, 1971, **37**(2): 238-242
- Engle R F. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 1982, **50**(4): 987-1007
- Bollerslevb T. Generalized autoregressive conditional heteroscedasticity. *Eeri Research Paper*, 1986, **31**(3): 307-327
- Lu C J, Lee T S, Chiu C C. Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*, 2009, **47**(2): 115-125
- Engel Y, Mannor S, Meir R. The kernel recursive least-squares algorithm. *IEEE Transactions on Signal Processing*, 2004, **52**(8): 2275-2285
- Yang J C, Yan W W, Xu R C, Zhang X. A novel online kernel ridge to forecast next-day electricity price. *International Journal of System Control and Information Processing*, 2018, **2**(4): 317-331
- Zhang J, Man K F. Time series prediction using RNN in multi-dimension embedding phase space. In: Proceedings of the 2002 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2002. 1868-1873.
- Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks. In: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). IEEE, 2013. 6645-6649
- Lun Shu-Xian, Lin Jian, Yao Xian-Shuang. Time series prediction with an improved echo state network using small world network. *Acta Automatica Sinica*, 2015, **41**(9): 1669-1679
(伦淑娴, 林健, 姚显双. 基于小世界回声状态网的时间序列预测. *自动化学报*, 2015, **41**(9): 1669-1679)
- Besse P C, Cardot H, Stephenson D B. Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 2000, **27**(4): 673-687
- Aggarwal S K, Saini L M, Kumar A. Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power and Energy Systems*, 2009, **31**(1): 13-22
- Anbazhagan S, Kumarappan N. Day-ahead deregulated electricity market price forecasting using recurrent neural network. *IEEE Systems Journal*, 2013, **7**(4): 866-87
- Clements A E, Hurn A S, Li Z. Forecasting day-ahead electricity load using a multiple equation time series approach. *European Journal of Operational Research*, 2016, **251**(2): 522-530
- Anbazhagan S, Kumarappan N. Day-ahead deregulated electricity market price forecasting using neural network input featured by DCT. *Energy Conversion and Management*, 2014, **78**: 711-719
- Rafiei M, Niknam T, Khooban M H. Probabilistic forecasting of hourly electricity price by generalization of ELM for usage in improved wavelet neural network. *IEEE Transactions on Industrial Informatics*, 2017, **13**(1): 71-79
- Pascanu R, Mikolov T, Bengio Y. On the difficulty of training

recurrent neural networks. In: Proceedings of the 2013 International Conference on Machine Learning, 2013.

- 22 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation*, 1997, **9**(8): 1735–1780
- 23 Greff K, Srivastava R K, Koutnik J, Steunebrink B R, Schmidhuber J. LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, **28**(10): 2222–2232
- 24 Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv: 1406.1078v1, 2014.
- 25 Caruana R. Multitask learning. *Machine Learning*, 1997, **28**(1): 41–75



徐任超 上海交通大学工学硕士. 主要研究方向为深度学习, 时间序列预测和图像处理.

E-mail: xurechao@yeah.net

(XU Ren-Chao Master student at Shanghai Jiao Tong University. His research interest covers deep learning,

time series forecasting, and image processing.)



阎威武 工学博士, 上海交通大学副教授. 主要研究方向为机器学习, 大数据, 图像处理和智能制造. 本文通信作者.

E-mail: aas_yanw@sjtu.edu.cn

(YAN Wei-Wu Ph.D., professor at Shanghai Jiao Tong University. His

research interest covers deep learning, time series forecasting, and image processing. Corresponding author of this paper.)



王国良 工学博士, 上海工程技术大学讲师. 主要研究方向为过程建模与控制, 火电机组先进控制.

E-mail: glwang@sues.edu.cn

(WANG Guo-Liang Ph.D, lecturer at Shanghai University of Engineering Science. His research interest

covers modeling and advanced control strategy of power process.)



杨健程 上海交通大学自动化系工学硕士. 主要研究方向为深度学习, 医学图像处理和 3D 视觉.

E-mail: jekyll4168@sjtu.edu.cn

(YANG Jian-Cheng Master student at Shanghai Jiao Tong University. His research interest covers

deep learning, medical image processing, and 3D vision.)



张曦 工学博士, 教授级高级工程师. 主要研究方向为统计过程控制, 火电机组先进控制.

E-mail: zhangxi2@csg.cn

(ZHANG Xi Ph.D., professor of engineering at China Southern Power Grid International Co., Ltd.. His research

interest covers statistical process control and advanced control strategy of power process.)