

基于 T2VNN 模型的阀冷系统进阀温度预测^①



陈霖, 周宇

(南京航空航天大学 计算机科学与技术学院, 南京 210016)

通信作者: 周宇, E-mail: zhouyu@nuaa.edu.cn

摘要: 预测进阀温度的变化趋势对阀冷系统中设备运行的安全可靠具有重要的参考价值. 针对传统方法特征提取困难、预测精度低等问题, 提出了一种用于预测进阀温度的 T2VNN (Time2Vec neural network) 模型, 该模型首先通过时间序列表示学习方法 Time2Vec 对进阀温度进行特征提取, 然后结合 TCN 和双向 LSTM 的优势, 并且使用分位数回归来实现概率预测. 最后设计了不同时间步和分位数在多个典型模型上的对比实验, 实验结果验证了 T2VNN 模型具有更高的预测性能, 并且通过消融实验证明了模型中各个组成部分的有效性.

关键词: 进阀温度; 阀冷系统; 时间序列; 表示学习; 分位数回归

引用格式: 陈霖, 周宇. 基于 T2VNN 模型的阀冷系统进阀温度预测. 计算机系统应用, 2022, 31(6): 132-140. <http://www.c-s-a.org.cn/1003-3254/8529.html>

Inlet Valve Temperature Prediction of Valve Cooling System Based on T2VNN Model

CHEN Lin, ZHOU Yu

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Predicting the change trend of inlet valve temperature can provide important reference for the safety and reliability of equipment operation in valve cooling systems. For the problems of difficult feature extraction and low prediction accuracy of traditional methods, a Time2Vec neural network (T2VNN) model is put forward for predicting inlet valve temperature. This model firstly extracts the features of inlet valve temperature by Time2Vec, a time series representation learning method, and then capitalizes on the advantages of the TCN and bidirectional LSTM to achieve probabilistic prediction by quantile regression. Finally, comparative experiments with different time steps and quantiles on several typical models are designed. The experimental results verify that the T2VNN model has higher prediction performance, and the effectiveness of each component in the model is demonstrated by ablation experiments.

Key words: inlet valve temperature; valve cooling system; time series; representation learning; quantile regression

阀冷系统是换流阀的一个重要组成部分, 它将阀体上各元件的功耗发热量排放到阀厅外, 保证晶闸管运行结温在正常范围内. 阀冷系统运行数据主要用于监测采用水冷的换流阀的冷却效果是否有效, 分为内冷水系统和外冷水系统. 其中冷却水进阀温度是反映阀冷系统换热性能的重要指标, 若换热性能出现异常, 将引起冷却水进阀温度上升, 严重时将达到换流阀运行允许温度的最大值, 导致换流阀强迫停运事件的发

生. 进阀温度都是以时间序列形式存储, 可以使用时序分析方法对其进行预测, 实现对设备异常运行状态的提前预警, 将有效提高阀冷系统内设备运行的安全性和可靠性.

时间序列预测在帮助人们进行资源管理和科学决策的过程中扮演着非常重要的角色, 其目的在于研究如何根据历史观测数据对未来进行准确预测. 经典的时间序列预测模型通过给定最近一段时间的历史数据

^① 收稿时间: 2021-08-13; 修改时间: 2021-09-29, 2021-10-08; 采用时间: 2021-10-19; csa 在线出版时间: 2022-05-26

来预测未来. 常用的方法包括 Box-Jenkins 方法^[1] 和 ARIMA 模型^[2]. 在实践中, 预测问题通常更为复杂, 例如, 输入通常包含动态历史特征和长期依赖, 输入和输出之间存在线性和非线性关系等. 然而, 传统的自回归方法^[3-5] 难以捕获序列数据中存在的这些关系. 解决现有方法在时间序列预测中的这种局限性是本文的重点.

最近, 神经网络已经在时间序列预测中得到了广泛应用, 常用于预测多个时间步以后的数据, 预测的时间步越远, 预测难度越大. 循环神经网络 (RNN) 和它的变体长短期记忆网络 (LSTM) 被提出用于对复杂序列数据建模, 例如自然语言^[6], 声音信号^[7] 和视频信息^[8]. LSTM 已经被广泛运用在了预测任务中^[9-11], 在捕获动态时间序列中的重复模式方面表现出强大的能力. 对于大多数深度学习实践者来说, 序列建模等同于循环神经网络, 然而文献^[12] 通过大量实验表明, 基于卷积架构的时间卷积网络 (TCN) 在多个序列建模任务上表现得比循环神经网络更好. TCN 由于其简单性和适用于长时间序列的特点在时间序列建模任务上受到了广泛关注^[13,14]. 本文试图将 LSTM 和 TCN 结合起来使用, 以综合它们的优势.

大多数神经网络模型没能有效地将时间本身作为特征, 通常假设数据是同步的, 即按相等时间间隔采集. 为了帮助神经网络更好地利用时间特征, 一些研究者针对特定问题提出了手动提取时间特征的方法^[15,16], 但这些方法代价高且需要问题相关的专业领域知识. 最近许多研究旨在提出通用的时间序列表示学习方法^[17-19], 通过模型自动提取时间特征, 从而避免手工提取带来的问题.

此外, 在各种确定性的点预测方法的基础上, 考虑到多种不确定性因素, 开展了大量概率性区间预测的研究工作, 即预测目标的完整分布来量化预测的不确定性, 从而估计决策的风险. 对于换流站内的数据来说, 如果预测值高于阈值就需要提前报警, 从预测分布中选取不同程度的估计值依赖于阈值设定的严格程度. 分位数回归已被广泛用于对不同的分位数进行预测, 以便在不做分布假设的情况下对目标分布进行估计. 均值回归又称最小二乘法, 是分位数回归的一种特殊情况, 它关注的是条件平均数, 而分位数回归可以扩展到任意分位数. 最近的许多预测模型^[20,21] 都通过分位数损失函数来实现分位数估计.

本文基于表示学习和组合网络的思想, 结合时间

序列的预测分析方法, 提出了一个深度学习框架 T2VNN (Time2Vec neural network) 并将其应用于阀冷系统进阀温度的预测研究, 如图 1 所示. 首先使用经典的时间表示学习方法 Time2Vec^[22] 对进阀温度的周期信息进行特征提取, 其次将 TCN 和双向 LSTM 构成组合网络对提取后的信息进行编码. 在结果的输出部分添加了传统的自回归线性模型, 使得非线性神经网络对于进阀温度数据规模的变化具有更强的鲁棒性. 最后使用分位数损失函数对模型进行评估, 从而产生不同分位数的预测结果. 实验表明本文方法在真实阀冷系统进阀温度数据集上的预测效果优于传统的时间序列预测方法.

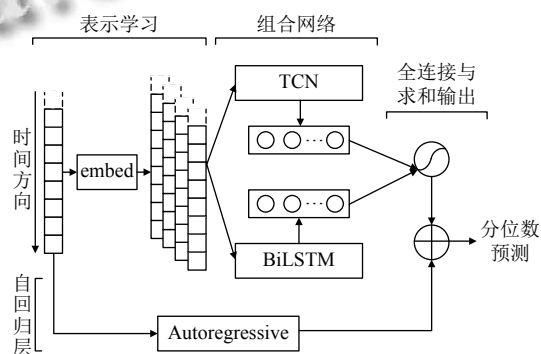


图 1 T2VNN 架构图

1 相关工作

在自然语言处理领域已经存在了大量关于表示学习的研究成果^[23,24], 当处理对象变成时间序列的时候, 许多研究者同样提出了一些通用的表示学习方法, 进一步提升了模型对于时间序列的处理能力. Fortuin 等^[25] 对时间序列的演化进行表示, 同时关注了表示的可解释性. Hyvärinen 等^[26] 通过表征对时间序列中大小均匀的分项来区分它们. Wu 等^[27] 提出一种基于核的估计方法来计算特征嵌入.

神经网络使用过去观察到的时间序列来预测未来未知的时序. 其中, 循环神经网络在多个时间步中维持隐含状态信息, 早期的成果主要包括使用原生的 RNN 模型^[28] 和组合了 ARIMA 和多层感知机的混合模型^[29] 等. Dasgupta 等^[30] 将 RNN 和动态玻尔兹曼机组合在一起. 随着基于卷积架构的 TCN 及其相关模型被提出, 一些研究开始将 TCN 用于时间序列预测. Hewage 等^[13] 将 TCN 用于天气预测, 实验表明 TCN 对比 LSTM 和其他经典的机器学习方法表现更好. Li 等^[31]

结合胶囊网络和 TCN 用于城市交通预测。

Salinas 等^[32]提出了 DeepAR 概率预测模型, 通过事先假设分布来预测未来。DeepAR 以 LSTM 作为基础模型, 通过最大化似然函数来训练。在预测的过程中, 需要从估计的参数分布中多次抽取样本, 作为预测的经验分布。本文方法与 DeepAR 的不同之处在于, 使用了更贴近实际的分位数回归^[33], 这是一种更高效的训练策略, 通过将预测问题表示为分位数回归来直接生成目标变量的分位数估计。Zheng^[34]使用梯度提升方法^[35]来最小化分位数回归的损失函数。

2 框架

在这一部分中, 首先对阀冷系统进阀温度预测问题进行形式化的建模, 然后在接下来的部分中讨论本文提出的 T2VNN 模型的架构细节 (图 1)。最后, 对分位数回归的目标函数和优化策略进行介绍。

2.1 问题建模

中国南方电网阀冷监测设备有很多, 每个设备都有各自的监测数据, 需要对所有的监测数据按设备进行划分, 并按照时间日期进行排列, 筛选出拥有完整时序性的进阀温度数据列表。首先给定一系列通过采集得到的进阀温度序列数据 $Y = \{y_1, y_2, \dots, y_T\}$, $y_1, y_2, \dots, y_T \in \mathbb{R}$, 实际应用中以滚动预测的方式来预测未来一段时间的数据。即在给定 $\{y_1, y_2, \dots, y_T\}$ 的前提下预测当前时间点 h 个时间步以后的数据, 这里记为 y_{T+h} , h 的大小由实际应用背景的需求决定。同样地, 为了预测下一个时间点 y_{T+h+1} , 需要假定 $\{y_1, y_2, \dots, y_T, y_{T+1}\}$ 已经可以获取。所以, 本文将当前时间点 T 的输入矩阵定义为 $X_T = \{y_1, y_2, \dots, y_T\} \in \mathbb{R}^T$ 。在接下来的部分中将详细介绍 T2VNN 模型的各个构建模块。

2.2 Embed 层

表示学习方法 Time2Vec^[22] 作为 T2VNN 模型的 embed 层, 目的是对进阀温度中的周期性和非周期性特征进行特征提取, 使神经网络更好地捕获其中的规律性特征。对于给定时间序列 τ 和输出维度 $k+1$, Time2Vec 的计算公式定义为:

$$t2v(\tau)[i] = \begin{cases} \omega_i \tau + \varphi_i, & \text{if } i = 0 \\ F(\omega_i \tau + \varphi_i), & \text{if } 1 \leq i \leq k \end{cases} \quad (1)$$

其中, $t2v(\tau)[i]$ 代表 $t2v(\tau)$ 的第 i 个输出向量, F 是周期性激活函数, ω_i 和 φ_i 是可学习的参数。在实验中, 为了使算法能够捕获时间序列中存在的周期性模式, F 选定为

正弦函数, $i = 0$ 对应的公式用于捕获时间序列中的非周期性模式。Time2Vec 将进阀温度时间序列分解出多个代表不同周期性以及非周期性的时间序列特征。将这些细化的特征传入到由 TCN 和双向 LSTM 构成的组合网络中, 进一步提取其中的规律性特征。

2.3 TCN-BiLSTM 组合网络

将组合网络 TCN 的输出向量记为 h_t^T , 双向 LSTM 的输出向量记为 h_t^B , 最后使用全连接操作将两者进行组合输出, 结果计算为:

$$h_t^D = W^T h_t^T + W^B h_t^B + b \quad (2)$$

其中, h_t^D 代表组合网络 (图 2 中的上半部分) 的预测结果。接下来将分别介绍其中的两个组成部分。

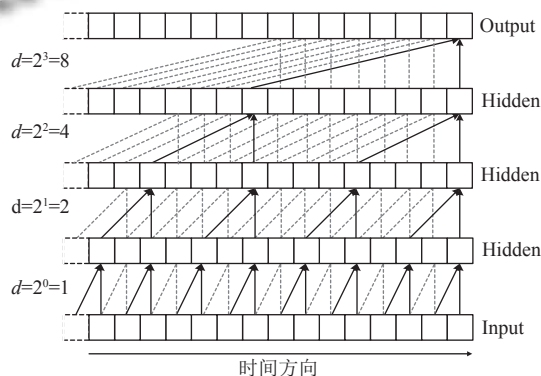


图 2 TCN 模型

(1) 时间卷积网络 TCN

TCN 由扩张因果卷积构成。其中, 卷积操作使用一维卷积, 通过在空缺位置补零将任意长度的多维特征映射为相同长度的输出向量, 并且使用因果卷积在映射的过程中确保当前时间点的输入数据只能通过历史数据计算得到。为了在不增加网络复杂度的情况下扩大对进阀温度历史数据的接受范围, 进一步融合了扩张卷积的思想形成扩张因果卷积。更形式化地说, 对于多维序列特征输入 $x \in \mathbb{R}^{T \times (k+1)}$ (即 Time2Vec 的输出结果) 和卷积核函数 $f: \{0, \dots, k-1\} \rightarrow \mathbb{R}$, 在序列元素 s 上的扩张卷积操作 F 定义为:

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (3)$$

其中, d 是扩张因子, k 是卷积核的大小, $s-d \cdot i$ 代表沿着过去的方向。如图 2 所示, 扩张因子 d 随层数的增加呈 2 的指数级增加, 在确保顶层输出能够表示所有输入的同时扩大了网络的接受范围。最终的输出取最后一层

的最后一个时间点的输出,同时为了提取多组进阀温度序列存在的依赖关系,构造多通道卷积,拼接形成最终的输出向量 h_t^T .

(2) 双向长短期记忆网络 BiLSTM

LSTM 能够缓解 RNN 训练时产生的梯度消失问题,更好地捕获时间序列的长期依赖模式. 因为进阀温度时间序列不仅存在正向的依赖关系,还有可能存在相应的逆向关系,所以在这里使用了双向 LSTM(BiLSTM),可以让输入特征在正序和逆序两个方向上进行传播,从而为预测建立更多的潜在特征. 形式上,将每个时间点的输入记为 x_t ,隐含状态记为 h_t ,内部门控输出记为 i_t, f_t, o_t 和 c_t , LSTM 循环单元的计算公式如下:

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1}) \tag{4}$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1}) \tag{5}$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1}) \tag{6}$$

$$c_t = f \cdot c_{t-1} + i_t \cdot \tanh(W_{cx}x_t + W_{ch}h_{t-1}) \tag{7}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{8}$$

对于双向 LSTM,如图 3 所示,时间点 t 处的隐含状态在正向 LSTM 中记为 \vec{h}_t ,在逆向 LSTM 中记为 \overleftarrow{h}_t ,拼接后的结果记为 h_t . 公式定义如下:

$$\vec{h}_t = \overrightarrow{LSTM}(x_t; \vec{h}_{t-1}) \tag{9}$$

$$\overleftarrow{h}_t = \overleftarrow{LSTM}(x_t; \overleftarrow{h}_{t-1}) \tag{10}$$

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \tag{11}$$

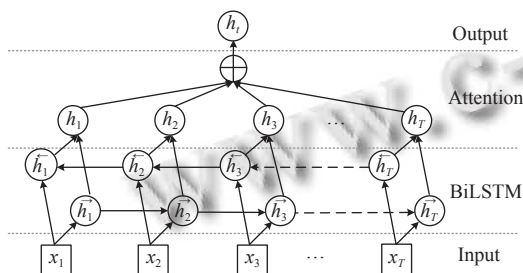


图 3 BiLSTM 模型

在模型的输出部分,为了有选择性地关注到每个时间点的输出信息,引入了 Attention 机制^[36]. 在当前时间点 t 的注意力权重 α_t 可以计算为:

$$\alpha_t = \text{AttnScore}(H_t, h_{t-1}) \tag{12}$$

其中, $H_t = [h_{t-T}, \dots, h_{t-1}]$ 是每个时间点隐含状态的堆叠矩阵, AttnScore 这里代表点乘操作. 最后将加权后的

上下文向量 $c_t = H_t \alpha_t$ 和最后一个时间点的隐含状态 h_{t-1} 拼接后通过线性映射得到双向 LSTM 的最终输出向量 h_t^L , 对应公式如下:

$$h_t^B = W[c_t; h_{t-1}] + b \tag{13}$$

2.4 自回归层

由于神经网络的非线性性质,输出尺度对于输入尺度的变化不敏感,并且由于进阀温度时间序列存在线性和非线性的多种上下文关系,数据规模随时间也会发生较明显的变化. 本文借鉴 LSTNet^[37] 将序列分解成线性和非线性部分,非线性部分通过 Time2Vec 进一步提取其中的周期性和非周期性模式,线性部分通过自回归 (AR) 模型进行特征提取. 将 AR 部分的预测结果记为 $h_t^L \in \mathbb{R}$, 系数记为 $W^{ar} \in \mathbb{R}^w$ 和 $b^{ar} \in \mathbb{R}$, 其中 w 是在输入矩阵上输入窗口的大小, 公式如下:

$$h_t^L = \sum_{k=0}^{w-1} W_k^{ar} y_{t-k} + b^{ar} \tag{14}$$

最后将组合网络部分的输出和 AR 部分的输出相加得到 T2VNN 模型在时间点 t 处的最终预测结果, 记为 \widehat{Y}_t , 对应公式如下:

$$\widehat{Y}_t = h_t^D + h_t^L \tag{15}$$

2.5 目标函数

进阀温度本身存在众多非线性特性以及非均匀分布的残差, 不适合使用平均绝对误差 (MAE) 作为损失函数, 所以本文使用平均绝对偏差 (MAD) 损失函数作为模型训练的目标函数, 以实现分位数预测, 定义为每个未来时间点 t 的标准分位数损失的和, 形式化如下:

$$L^{MAD} = \sum_t [q(y_t - \widehat{y}_t)^+ + (1-q)(\widehat{y}_t - y_t)^+] \tag{16}$$

其中, y_t 和 \widehat{y}_t 分别表示 t 时刻的真实值和模型的预测值. 从式 (16) 中可以看出, 对预测值一定程度上的高估或低估是由加法前后的非对称性来实现的. 前后两部分的比例由分位数 $q \in (0, 1)$ 控制, 当 q 取 0.5 时原损失函数等价于平均绝对误差 (MAE) 损失函数, q 的具体取值由实际环境的要求决定. 最后, 使用 Adam^[38] 来优化目标函数可以获得更快的收敛速度.

3 实验分析

本文使用 5 种方法 (包括本文方法) 在电网领域数据集上针对时间序列预测任务展开实验. 所有的数据

和实验代码可以从网站 <https://github.com/Caleb411/T2VNN> 中获取。

3.1 对比方法

本文使用如下方法展开对比实验:

(1) **Benchmark** 直接使用历史输入最后一个时间点的值作为预测值。

(2) **Autoregressive** 自回归模型是传统时间序列预测方法的典型代表。

(3) **Gradient-Boosting** 是一个经典的机器学习算法,可用于回归和分类问题。为了对比这里使用分位数损失函数。

(4) **DeepAR** 是基于 LSTM 的概率预测模型,为了对比使用其基于分位数回归的形式。

3.2 评估指标

本文使用两个传统的评估指标,定义如下:

(1) 相对平方根误差 (*RSE*):

$$RSE = \frac{\sqrt{\sum_t (Y_t - \hat{Y}_t)^2}}{\sqrt{\sum_t (Y_t - \text{mean}(Y))^2}} \quad (17)$$

(2) 经验相关系数 (*CORR*):

$$CORR = \frac{\sum_t (Y_t - \text{mean}(Y))(\hat{Y}_t - \text{mean}(\hat{Y}))}{\sqrt{\sum_t (Y_t - \text{mean}(Y))^2 (\hat{Y}_t - \text{mean}(\hat{Y}))^2}} \quad (18)$$

其中, $Y, \hat{Y}_t \in \mathbb{R}^n$ 分别表示测试集中预测目标的真实值和预测值的集合。与常用的均方根误差 (RMSE) 不同, *RSE* 消除了数据尺度的影响,可以得到可读性更好的评估结果。*RSE* 越小, *CORR* 越大,模型的效果越好。

3.3 数据描述

实验数据集 valve (<https://github.com/Caleb411/T2VNN/tree/main/data>) 是某电网换流站阀冷系统中内冷水系统各项指标的真实数据,数据的采集频率是半小时一次,时间从 2017 年 12 月 3 日到 2019 年 7 月 3 日,一共包含 29 045 个时间点。该数据集包含 4 个维度的时间序列,分别表示进阀温度、出阀温度、进阀压力和冷却水电导率随时间的变化情况。其中进阀温度是需要重点关注的指标,它的数值变化对于换流站内设备的安全性存在的影响较大,也是本文需要预测的对象。

为了更直观地展示进阀温度随时间的变化情况,绘制如图 4 所示的折线图。从图 4 中可以看出,除了一些明显的极端值外,进阀温度随时间的变化呈现出一

定的周期性。为了使用更清晰的方式对重复模式进行检验,进一步绘制自相关函数 (ACF) 图如图 5 所示。自相关函数描述了进阀温度当前值与其过去值之间的相关程度,包括直接和间接的相关性信息。时间序列包含长期趋势变动、季节变动、循环变动和不规则变动,ACF 在寻找相关性时会综合考虑所有这些变动,它的函数定义如下:

$$R(\tau) = \frac{E[(y_t - \mu)(y_{t+\tau} - \mu)]}{\sigma^2} \quad (19)$$

其中, y_t 表示一系列进阀温度数据, μ 是平均值, σ^2 是方差,这里采用经验无偏估计量来计算自相关性。从图 5 中可以看出,进阀温度时间序列具有高度自相关性的重复模式。这些观察结果对于后续不同方法实验结果的分析非常重要,是判断该数据是否适合建模的前提。

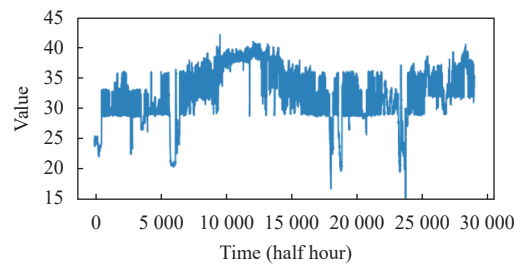


图 4 进阀温度折线图

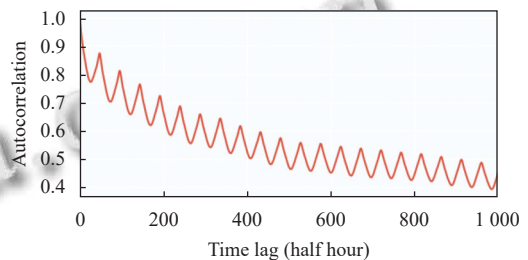


图 5 进阀温度自相关函数图

在将数据输入模型之前,还需要一些预处理操作。从图中可以看出进阀温度随时间变化幅度较大,需要将它们整合到统一的范畴下,然后才能进行后续的建模操作。这里使用标准化的方式,目的是将原始数据分布转化为均值为 0 方差为 1 的标准正态分布。之后将原数据集依据 6:2:2 的比例按时间顺序划分为训练集、验证集和测试集。接下来将对实验过程中涉及的一些细节进行说明。

3.4 实验细节

所有方法首先都需要通过滑动窗口的方式批量采

集数据,形成<输入,输出>的数据对,滑动窗口的长度 T 的取值范围设置为 $\{2^2, 2^3, \dots, 2^6\}$, 实验中根据模型在验证集上的预测效果进行调整. 对比模型 Benchmark 和 Autoregressive 不包含超参数, Gradient-Boosting 使用默认参数, 其中关键超参数梯度提升阶数为 100, 树的最大深度为 3. 本文模型超参数配置如表 1 所示, 其中 k 是表示学习 Time2Vec 部分的维度, filters 和 kernel_size 分别表示组合网络中 TCN 的卷积通道的数目和卷积核的长度. 组合网络中双向 LSTM 的 hidden_size 表示隐含状态的长度, 和 TCN 中的 filters 是相似的概念, 所以这里将它们的值设置为相同. 同时, 对比模型 DeepAR 的隐藏单元长度也设置为 hidden_size, 使模型复杂度与本文模型保持一致. 由于 BiLSTM 是双向的, 所以它经过 Attention 机制后最终的输出长度 output_size 是 hidden_size 的两倍. 本文方法自回归层的输入窗口大小 w 在不超过滑动窗口 T 的前提下从 $\{3, 6, 9, 12\}$ 中选取. 最后, 在模型训练的过程中, batch_size 和 epochs 分别表示数据一次输入的批量大小和迭代次数, 所有对比模型保持一致.

表 1 超参数配置

超参数	值
k	64
filters	256
kernel_size	2
hidden_size	256
output_size	512
batch_size	128
epochs	5

3.5 实验结果

本文使实验从两个角度出发, 一个是不同的预测时间步 $h \in \{3, 6, 9, 12\}$, 另一个是不同的分位数 $q \in \{0.01, 0.25, 0.5, 0.75, 0.99\}$, 并在测试集上评估方法的预测性能. 如表 2 和表 3 所示, 粗体表示每列中最好的评估结果. 表 2 总结了所有方法在不同预测时间步上的评估结果, 预测的时间步越长, 预测难度越大, 此时的分位数统一设置为 0.5. 从表中可以看出, 本文方法 T2VNN 的预测性能明显优于其他方法, 并且在所有的预测时间步中都占据优势. 此外, 当预测时间步为 3 时, 本文方法的 RSE 相比于神经网络基准 DeepAR 降低了 12.8%, 因为 DeepAR 虽然考虑到了序列中存在的长期依赖信息, 但是没能捕获到可能存在的短期局部依赖模式, 而本文方法结合了 TCN 的思想解决了

该问题. 同时也表明本文方法中的 Time2Vec 成功提取到了进阀温度中存在的复杂周期性重复模式, 并且该框架设计对于预测时间步较远的情况依然有效.

表 2 对比实验不同预测时间步评估结果

方法	指标	3	6	9	12
Benchmark	RSE	0.2713	0.4046	0.4782	0.5382
	CORR	0.9632	0.9181	0.8857	0.8555
Autoregressive	RSE	0.3066	0.3860	0.4197	0.4434
	CORR	0.9521	0.9228	0.9081	0.8980
Gradient-Boosting	RSE	0.2288	0.3000	0.3395	0.3652
	CORR	0.9739	0.9540	0.9407	0.9314
DeepAR	RSE	0.2264	0.3000	0.3322	0.3653
	CORR	0.9742	0.9542	0.9447	0.9320
本文方法	RSE	0.1972	0.2790	0.3145	0.3513
	CORR	0.9815	0.9608	0.9507	0.9389

表 3 对比实验不同分位数评估结果

方法	0.01	0.25	0.5	0.75	0.99
Benchmark	0.3412	0.3416	0.3421	0.3426	0.3430
Autoregressive	0.0692	0.3786	0.4377	0.4082	0.0828
Gradient-Boosting	0.0530	0.2701	0.3262	0.2752	0.0360
DeepAR	0.0340	0.2883	0.3176	0.2743	0.0376
本文方法	0.0307	0.2078	0.2738	0.2138	0.0303

表 3 总结了所有方法在不同分位数上损失函数的值, 此时的预测时间步统一设置为 3. 从表中可以看出, Benchmark 和 Autoregressive 由于缺乏分位数预测能力在所有分位数上表现都是最差的. Gradient-Boosting 和 DeepAR 由于拥有较强的分位数感知能力性能有所提升. 本文方法 T2VNN 在所有分位数上损失函数的值都是最低的, 说明 T2VNN 在多种不同的分位数下适应能力更强, 表明使用平均绝对偏差 (MAE) 作为损失函数的正确性. 为了将预测结果可视化, 便于形象地观察预测效果, 在测试集中选取 500 个连续时间点的数据, 将预测结果绘制如图 6 和图 7 所示. 图 6 中展示了分位数为 0.5 (即自回归预测), 预测时间步为 3 的预测结果, 其中蓝线代表预测值, 黑线代表真实值. 图 7 展示了不同分位数的预测结果. 红色区域的下界和上界分别表示分位数 0.25 和 0.75 的预测结果; 黄色区域的边界表示分位数 0.01 和 0.99 的预测结果. 从图中可以看出, 本文方法 T2VNN 能够捕获到未来序列数据中的时间模式和趋势变化.

3.6 消融实验

为了衡量本文 T2VNN 框架中一些部分对于模型整体性能贡献, 设计了相应的消融实验, 即每次从原

方法中移除某一个组件进行实验. 首先, 将移除了不同组件后的 T2VNN 命名如下.

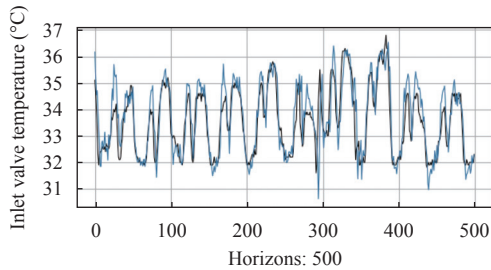


图6 进阀温度自回归预测结果

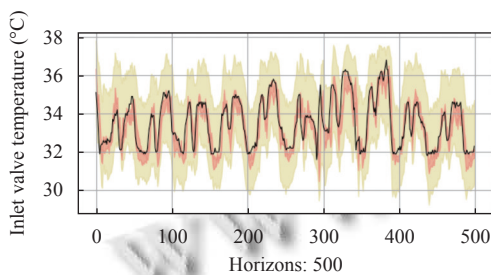


图7 进阀温度分位数预测结果

(1) T2VNNw/oT2V: T2VNN 模型移除时间序列表示学习方法 Time2Vec.

(2) T2VNNw/oAttention: T2VNN 模型组合网络部分移除双向 LSTM 中的 Attention 层.

(3) T2VNNw/oAR: T2VNN 模型移除与表示学习和组合网络并行的自回归层.

移除部分组件会减少模型的参数量, 需要调整部分超参数使得所有模型的参数量相似. 表 4 总结了所有方法在不同预测时间步上的评估结果, 分位数设置为 0.5, 粗体为两个评估指标在每列中最好的结果. 从表中可以看出, 去除 Attention 或 AR 部分后, 模型的性能在所有预测时间步上均出现一定程度的下滑. 去除 T2V 部分后, 除了预测时间步为 12 时 RSE 的值外, 其他情况下也都出现了下滑. 最后, 为了进一步验证本文模型中添加自回归层的必要性, 选取 200 个连续时间点的数据将 T2VNNw/oAR 和 T2VNN 的预测结果绘制如图 8 和图 9 所示, 其中蓝线代表预测值, 黑线代表真实值. 从图中可以看出两者预测的差异主要体现在峰值处, 除去自回归层后模型对于数据趋势的变化不敏感, 自回归层能够在一定程度上改善模型对于数据趋势的捕获能力. 综上所述, 本文方法中的几个关键组成部分对模型整体性能均存在贡献.

表 4 消融实验不同预测时间步评估结果

方法	指标	3	6	9	12
T2VNNw/oT2V	RSE	0.2484	0.3144	0.3146	0.3474
	CORR	0.9705	0.9584	0.9494	0.9388
T2VNNw/oAttention	RSE	0.1993	0.2840	0.3254	0.3610
	CORR	0.9801	0.9602	0.9490	0.9362
T2VNNw/oAR	RSE	0.2025	0.2840	0.3335	0.3733
	CORR	0.9798	0.9599	0.9475	0.9360
本文方法	RSE	0.1972	0.2790	0.3145	0.3513
	CORR	0.9815	0.9608	0.9507	0.9389

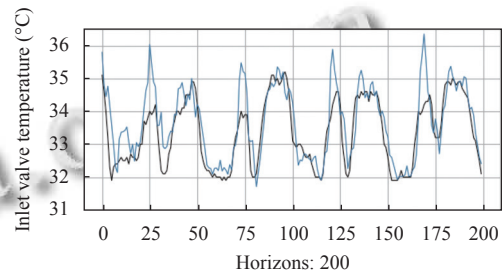


图8 T2VNNw/oAR 预测结果

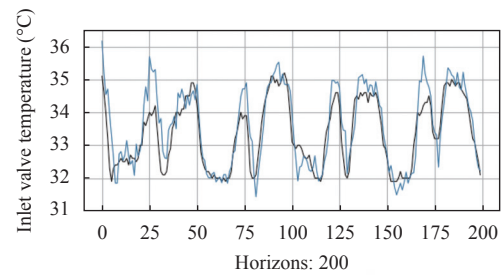


图9 T2VNN 预测结果

4 结论与展望

本文针对阀冷系统进阀温度预测问题提出了一个深度学习框架 T2VNN 模型. 通过结合表示学习方法 Time2Vec 和组合网络的思想, 同时在输出部分添加了自回归模型, 将模型预测过程分解为线性和非线性部分, 用于电网领域阀冷系统中进阀温度的预测研究. 其中组合网络部分结合了 TCN 和双向 LSTM 的优势, 共同捕获序列中存在的依赖关系. 本文方法在多个预测时间步和分位数预测中性能均超过多个典型模型, 成功捕获到了进阀温度历史数据中的周期性与非周期性重复模式, 以及相应的趋势变化. 最后, 通过消融实验证明了本文方法中一些关键组件对整体性能均存在贡献.

参考文献

1 Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of

- Forecasting, 2004, 20(1): 5–10. [doi: [10.1016/j.ijforecast.2003.09.015](https://doi.org/10.1016/j.ijforecast.2003.09.015)]
- 2 Box GEP, Jenkins GM, MacGregor JF. Some recent advances in forecasting and control. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1974, 23(2): 158–179.
 - 3 Janacek G. Time series analysis forecasting and control. *Journal of Time Series Analysis*, 2010, 31(4): 303.
 - 4 Li JH, Chen WY. Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, 2014, 30(4): 996–1015. [doi: [10.1016/j.ijforecast.2014.03.016](https://doi.org/10.1016/j.ijforecast.2014.03.016)]
 - 5 Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 2003, 50: 159–175. [doi: [10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)]
 - 6 Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2014. 3104–3112.
 - 7 van den Oord A, Dieleman S, Zen H, *et al.* WaveNet: A generative model for raw audio. arXiv: 1609.03499, 2016.
 - 8 Donahue J, Hendricks LA, Guadarrama S, *et al.* Long-term recurrent convolutional networks for visual recognition and description. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston: IEEE, 2015. 2625–2634.
 - 9 Trung CV, Dinh MN, Dimla E. Predicting phase-field behavior of brittle fracture model based on LSTM time series forecasting model. *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*. Ho Chi Minh City: IEEE, 2020. 1–6.
 - 10 Yadav A, Jha CK, Sharan A. Optimizing LSTM for time series prediction in Indian stock market. *Procedia Computer Science*, 2020, 167: 2091–2100. [doi: [10.1016/j.procs.2020.03.257](https://doi.org/10.1016/j.procs.2020.03.257)]
 - 11 Karevan Z, Suykens JAK. Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Networks*, 2020, 125: 1–9. [doi: [10.1016/j.neunet.2019.12.030](https://doi.org/10.1016/j.neunet.2019.12.030)]
 - 12 Bai SJ, Kolter JZ, Koltun V. Trellis networks for sequence modeling. arXiv: 1810.06682, 2019.
 - 13 Hewage P, Behera A, Trovati M, *et al.* Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. *Soft Computing*, 2020, 24(21): 16453–16482. [doi: [10.1007/s00500-020-04954-0](https://doi.org/10.1007/s00500-020-04954-0)]
 - 14 Lin L, Xu BL, Wu WC, *et al.* Medical time series classification with hierarchical attention-based temporal convolutional networks: A case study of myotonic dystrophy diagnosis. arXiv: 1903.11748, 2019.
 - 15 Baytas IM, Xiao C, Zhang X, *et al.* Patient subtyping via time-aware LSTM networks. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Halifax: ACM, 2017. 65–74.
 - 16 Kwon BC, Choi MJ, Kim JT, *et al.* RetainVis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Transactions on Visualization and Computer Graphics*, 2019, 25(1): 299–309. [doi: [10.1109/TVCG.2018.2865027](https://doi.org/10.1109/TVCG.2018.2865027)]
 - 17 Neil D, Pfeiffer M, Liu SC. Phased LSTM: Accelerating recurrent network training for long or event-based sequences. arXiv: 1610.09513, 2016.
 - 18 Zhu Y, Li H, Liao YK, *et al.* What to do next: Modeling user behaviors by time-LSTM. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Melbourne: AAAI Press, 2017. 3602–3608.
 - 19 Mei HY, Eisner J. The neural Hawkes process: A neurally self-modulating multivariate point process. arXiv: 1612.09328, 2017.
 - 20 Wen RF, Torkkola K, Narayanaswamy B, *et al.* A multi-horizon quantile recurrent forecaster. arXiv: 1711.11053, 2018.
 - 21 Xu QF, Liu X, Jiang CX, *et al.* Quantile autoregression neural network model with applications to evaluating value at risk. *Applied Soft Computing*, 2016, 49: 1–12. [doi: [10.1016/j.asoc.2016.08.003](https://doi.org/10.1016/j.asoc.2016.08.003)]
 - 22 Kazemi SM, Goel R, Eghbali S, *et al.* Time2Vec: Learning a vector representation of time. arXiv: 1907.05321, 2019.
 - 23 Church WK. Word2Vec. *Natural Language Engineering*, 2017, 23(1): 155–162. [doi: [10.1017/S1351324916000334](https://doi.org/10.1017/S1351324916000334)]
 - 24 Young T, Hazarika D, Poria S, *et al.* Recent trends in deep learning based natural language processing. arXiv: 1708.02709, 2018.
 - 25 Fortuin V, Hüser M, Locatello F, *et al.* Deep self-organization: Interpretable discrete representation learning on time series. arXiv: 1806.02199, 2019.
 - 26 Hyvärinen A, Morioka H. Unsupervised feature extraction by time-contrastive learning and nonlinear ica. *Proceedings of the 30th International Conference on Neural Information Processing Systems*. Barcelona: Curran Associates Inc., 2016. 3772–3780.
 - 27 Wu LF, Yen IEH, Yi JF, *et al.* Random warping series: A

- random features method for time-series embedding. Proceedings of the 21st International Conference on Artificial Intelligence and Statistics. Lanzarote: PMLR, 2018. 793–802.
- 28 Connor J, Atlas LE, Martin DR. Recurrent networks and NARMA modeling. Proceedings of the 4th International Conference on Neural Information Processing Systems. Denver: Morgan Kaufmann Publishers Inc., 1991. 301–308.
- 29 Jain A, Kumar AM. Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing, 2007, 7(2): 585–592. [doi: 10.1016/j.asoc.2006.03.002]
- 30 Dasgupta S, Osogami T. Nonlinear dynamic Boltzmann machines for time-series prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1): 1833–1839.
- 31 Li DZ, Lin C, Gao W, *et al.* Capsules TCN network for urban computing and intelligence in urban traffic prediction. Wireless Communications and Mobile Computing, 2020, 2020: 6896579.
- 32 Salinas D, Flunkert V, Gasthaus J, *et al.* DeepAR: Probabilistic forecasting with autoregressive recurrent networks. International Journal of Forecasting, 2020, 36(3): 1181–1191. [doi: 10.1016/j.ijforecast.2019.07.001]
- 33 Koenker R, Bassett GW. Regression quantiles. Econometrica, 1978, 46(1): 211–244.
- 34 Zheng SF. Boosting based conditional quantile estimation for regression and binary classification. 9th Mexican International Conference on Artificial Intelligence, MICAI 2010. Pachuca: Springer, 2010. 67–79.
- 35 Friedman JH. Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 2001, 29(5): 1189–1232. [doi: 10.1214/aos/1013203450]
- 36 Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv: 1409.0473, 2016.
- 37 Lai GK, Chang WC, Yang YM, *et al.* Modeling long- and short-term temporal patterns with deep neural networks. The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. Ann Arbor: ACM, 2018. 95–104.
- 38 Kingma D, Ba J. Adam: A method for stochastic optimization. arXiv: 1412.6980, 2017.