

DOI: 10.20098/j.cnki.1001-5132.2024.1212

## 基于隶属度的模糊时间序列取水量预测研究

毛小燕<sup>1</sup>, 戴俊杰<sup>1</sup>, 曹磊<sup>2</sup>

(1. 宁波大学科学技术学院 基础学院, 浙江 宁波 315300; 2. 江苏金恒信息科技股份有限公司, 江苏 南京 210044)

**摘要:** 科学准确地预测水厂取水量是实现水资源高效管理与科学决策的关键环节, 更是推进智慧水利建设的重要前提。由于取水量序列具有非线性、模糊性等复杂特征, 且受多种不确定因素影响, 使传统预测模型的精度受到一定限制。本文结合模糊理论和门控循环单元 (GRU) 神经网络的特性, 提出一种基于隶属度的模糊时间序列取水量预测新方法。首先, 利用模糊  $C$  均值聚类 (FCM) 算法从原始序列中构建子隶属度序列; 其次, 采用两种不同策略, 利用 GRU 模型对各子隶属度序列进行预测; 最后, 通过去模糊化处理得出取水量预测值。将新模型应用于浙江省某水厂的取水量序列预测, 并与传统 GRU 模型和 AM-GRU 模型进行对比分析, 结果显示新模型在均方误差 (MSE)、均方根误差 (RMSE) 和平均绝对百分比误差 (MAPE) 上的表现均优于其他模型, 显著提高了取水量的预测精度。

**关键词:** 取水量; 隶属度; 模糊时间序列; 模糊  $C$  均值聚类; 循环神经网络; 门控循环单元

**中图分类号:** TP301

**文献标志码:** A

**文章编号:** 1001-5132 (2025) 06-0082-09

城市供水是城市规划建设中的重要环节, 水厂作为水资源供应的核心主体, 在水资源生产和供给环节中发挥着决定性作用。合理的供需平衡关系不仅能够减少供水设施的建设投资成本, 还能及时应对未来可能出现的用水危机<sup>[1]</sup>。因此, 如何准确预测水厂每日需水量、取水量、供水量尤为重要。

本质上, 水厂历史每日需水量、取水量、供水量能够构成一段简单的时间序列, 而时间序列预测 (Time Series Forecasting, TSF) 就是利用现有已知的观测值来推测未来数据的过程。学者们也提出了各类不同的预测方法, 如差分自回归移动平均<sup>[2]</sup> (ARIMA)、灰色模型<sup>[3]</sup>、基于深度学习的时间序列预测<sup>[4-5]</sup>等。其中, ARIMA 和灰色模型作为传统机器学习算法, 具有所需数据量小、计算简便、准确率较高的特点, 适用于水量的短期预测; 而基于深度学习的时间序列预测虽然数据要求量较大, 参数收敛计算时间较长, 但准确率高, 更适用于拟合较长时间跨度的水量变化情况。

由于各模型的自身局限性, 且现实世界中的不

确定性和非线性同样难以把握, 使得未来值的准确预测极具挑战<sup>[6]</sup>。以日需水量为例, 社会发展、季节气候、假期节庆以及突发事件等都会造成日需水量的波动干扰<sup>[7]</sup>。在此背景下, 模糊集理论被引入时间序列预测领域成为必然趋势。早在 1993 年, Song 等<sup>[8-9]</sup>就提出模糊时间序列预测 (Fuzzy Time Series Forecasting, FTSF), 并成功应用在大学入学人数预测中。近年来, 模糊时间预测方法因其独特的应用价值, 已逐渐成为不同领域的研究热点, 被广泛应用于原油价格预测<sup>[10]</sup>、股价走势<sup>[11]</sup>、清洁能源系统<sup>[12-13]</sup>、传染病防控<sup>[14]</sup>等方面。与传统时间序列预测相比, 模糊时间序列预测通过模糊规则将真实值映射为模糊集合, 再经去模糊化后重新转换为真实值, 有效地提升了模型在复杂场景下的可操作性与适用性。

针对传统模糊时间序列预测模型中有效论域划分和序列数据模糊化处理存在的不足, 众多学者引入直觉模糊集、概率论等理论对其进行优化扩展, 也相继提出了一些改进模型, 例如利用改进狼

收稿日期: 2024-12-23

宁波大学学报(理工版) 网址: <http://journal.nbu.edu.cn/>

基金项目: 浙江省自然科学基金 (LY20A010012)

第一作者: 毛小燕, 副教授, 主要研究方向为模糊数学与 Rough 集理论。E-mail: maoxiaoyan@nbu.edu.cn

群算法来划分模糊区间建立的预测模型<sup>[15]</sup>、基于概率加权策略的概率预测方法<sup>[16]</sup>、基于直觉模糊化的广义预测模型<sup>[17]</sup>等。同时,随着深度学习的快速发展,早期时间序列预测模型已难以满足当下时间序列预测任务的要求。时间序列神经网络作为处理非线性、复杂时间序列数据的专用机器学习工具,催生出许多基于神经网络的时间序列预测的不同模型、方法和程序,例如基于融合 Transformer 的门尼系数预测模型<sup>[18]</sup>、基于 EMD-LSTM 的时间预测方法<sup>[19]</sup>、双向长短期记忆的时间序列预测方法<sup>[20]</sup>等。将模糊集和深度学习相结合应用于时间序列预测,不仅能够有效处理数据中的不确定性和噪声,提升模型对噪声的鲁棒性,还能同时捕捉序列的长期依赖关系和更复杂的非线性模式。其中,模糊预处理可对不确定数据进行软划分,进一步提升模型的抗噪能力。因此针对取水量数据的不确定性、模糊性特点,本文提出一种基于隶属度的模糊时间序列预测方法。新方法首先利用模糊 C 均值(Fuzzy C-Means, FCM)聚类对原始取水量序列进行划分和模糊化,然后建立门控循环单元(Gated Recurrent Unit, GRU)神经网络模型,采用两种策略对隶属度进行预测,最后通过去模糊化得到最终预测值。为验证新方法的有效性,以浙江省某水厂近5年取水量数据作为研究对象,将其与传统预测模型进行对比分析。

## 1 基于隶属度的模糊时间序列模型

### 1.1 模糊时间序列的基本概念

**定义 1**<sup>[8-9]</sup> 设  $U$  为论域,且  $U$  由有限个离散的元素构成,即  $U = \{u_1, u_2, \dots, u_n\}$ , 则定义在该论域上的模糊集合  $A$  可表示为

$$A = \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \dots + \frac{f_A(u_n)}{u_n} \quad (1)$$

其中,  $f_A(\cdot)$  为模糊集  $A$  上的隶属度函数,且  $f_A(\cdot): U \rightarrow [0, 1]$ ;  $f_A(u_i)$  表示  $u_i$  属于模糊集  $A$  的隶属度值,且  $f_A(u_i) \in [0, 1]$ ,  $1 \leq i \leq n$ 。

**定义 2**<sup>[8-9]</sup> 在实数域的一个子集  $Y(t) (t = 0, 1, 2, \dots)$  上定义一组模糊集  $f_i(t)$ , 设  $F(t) = \{f_1(t), f_2(t), f_3(t), \dots\}$ , 则称  $F(t)$  为  $Y(t)$  上的模糊时间序列。

**定义 3**<sup>[8-9]</sup> 设一个模糊时间序列  $F(t) (t = 1, 2, 3, \dots)$ , 令  $F(t-1) = A_i$ ,  $F(t) = A_j$ ,  $R(t, t-1)$  是从

$F(t-1)$  到  $F(t)$  的模糊关系,且满足  $F(t) = F(t-1) \circ R(t, t-1)$ , 即  $A_i \rightarrow A_j$ , 其中“ $\circ$ ”表示合成运算,即称  $R(t, t-1)$  为定义在  $F(\cdot)$  上的一阶模糊关系,其中,  $A_i$  和  $A_j$  分别称为“前件”和“后件”。

**定义 4**<sup>[8-9]</sup> 设模糊时间序列  $F(t) (t = 1, 2, 3, \dots)$  由  $F(t-n), F(t-n+1), \dots, F(t-1)$  通过模糊关系确定,则具有  $n$  个前件的  $n$  阶模糊关系可表示为  $F(t-n), F(t-n+1), \dots, F(t-1) \rightarrow F(t)$ 。

传统模糊时间序列预测模型的实施过程主要包括以下步骤:(1)确定论域并进行划分;(2)定义模糊集和数据模糊化;(3)建立模糊逻辑关系并确定模糊关系矩阵;(4)去模糊化后得到预测值。

### 1.2 模糊 C 均值聚类算法

FCM 聚类算法是一种对目标函数进行优化的聚类方法<sup>[21]</sup>,其核心思想是将数据集划分为若干个簇,同簇的数据相似性大,而不同簇的数据相似性小。对于给定样本集  $X = \{x_1, x_2, \dots, x_N\}$ , 在隶属度  $\sum_{j=1}^k u_{ij} = 1$  的约束条件下,使目标函数  $I_m$  取得最小值,  $I_m$  如下式所示:

$$I_m = \sum_{i=1}^N \sum_{j=1}^k u_{ij}^m \|x_i - c_j\|^2 \quad (2)$$

式中:  $N$  为样本总数;  $k$  为聚类中心数目;  $m$  为控制模糊重叠程度的模糊划分矩阵指数,且  $m > 1$ , 一般  $m$  取 2;  $u_{ij}$  表示样本  $x_i$  对第  $j$  个聚类中心的隶属度;  $\|x_i - c_j\|$  表示样本  $x_i$  与聚类中心  $c_j$  的欧氏距离。目标函数在约束下的拉格朗日式如下:

$$F = \sum_{i=1}^N \sum_{j=1}^k u_{ij}^m \|x_i - c_j\|^2 + \lambda \left( \sum_{j=1}^k u_{ij} - 1 \right) \quad (3)$$

令  $\partial F / \partial u_{ij} = 0$ ,  $\partial F / \partial \lambda = 0$ , 可以得到聚类中心和隶属度矩阵的迭代收敛公式:

$$c_j = \sum_{i=1}^N u_{ij}^m x_i / \sum_{i=1}^N u_{ij}^m \quad (4)$$

$$u_{ij} = 1 / \sum_{r=1}^k \left( \frac{\|x_i - c_j\|}{\|x_i - c_r\|} \right)^{\frac{2}{m-1}} \quad (5)$$

### 1.3 门控循环单元神经网络

门控循环单元<sup>[22]</sup>(GRU)神经网络在传统循环神经网络<sup>[23]</sup>(Recurrent Neural Network, RNN)基础上增加了“门”结构来缓解梯度爆炸和梯度消失现象,避免了长期依赖问题。GRU 在长短期记忆(Long Short-Term Memory, LSTM)神经网络<sup>[24]</sup>

的结构基础上进行改进与优化,摒弃了细胞状态的传递机制,将输入门和遗忘门合并为一个更新门,从而同时实现对上一时刻隐藏状态的选择性遗忘及当前时刻候选状态的选择性记忆。此外,GRU还引入重置门,用于控制上一时刻隐藏状态对当前候选状态的影响程度。

GRU 模型网络结构如图 1 所示,其计算过程如式(6)~(9)所示。图 1 中,  $r_t$  为  $t$  时刻的重置门状态;  $z_t$  为  $t$  时刻的更新门状态;  $x_t$  为  $t$  时刻输入的样本;  $h_{t-1}$  和  $h_t$  分别为  $t-1$  时刻隐藏层的输入和  $t$  时刻隐藏层的输出;  $\tilde{h}_t$  为当前时刻候选状态;  $W_r$ 、 $W_z$  和  $W_h$  分别为重置门、更新门和候选状态的权重矩阵。

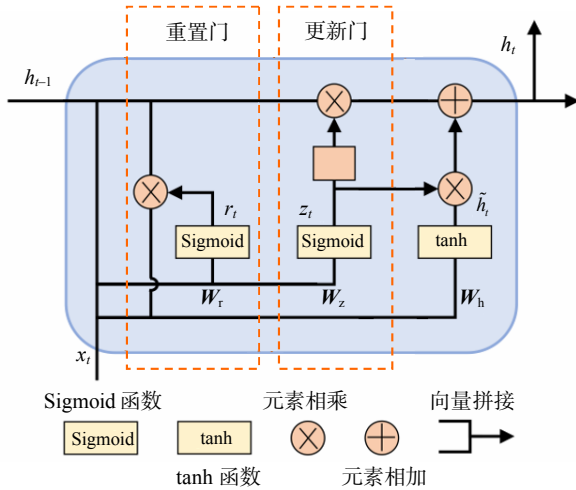


图 1 GRU 模型神经元网络结构

重置门计算公式:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (6)$$

更新门计算公式:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (7)$$

候选状态计算公式:

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t]) \quad (8)$$

隐藏层输出计算公式:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (9)$$

其中,  $\sigma$  和  $\tanh$  分别为 Sigmoid 激活函数和双曲正切激活函数,  $\odot$  表示矩阵的哈达玛积。

#### 1.4 预测模型的建立

GRU 神经网络虽然在时间序列预测任务中取得了不错的效果,但其模型准确率过分依赖数据的完整性和准确性,而且在预测任务中主要考虑了时间维度上的联系,却忽略了时间序列本身存在的模糊性和造成其本身变化的一些不确定性因素。传统模糊时间序列预测模型作为处理不确定性信息的常用方法,已在众多领域得到应用,但其预测模型比较适用于数据量小、预测精度要求不高的场景。因此,本文在传统模糊时间序列预测模型基础上进行改进优化,将深度学习与模糊时间序列相结合,提出一种基于隶属度的模糊时间序列预测方法。新方法主要步骤:(1)确定论域并进行划分;(2)数据模糊化;(3)子隶属度序列的预测;(4)去模糊化后得到最终预测值。基于隶属度的模糊时间序列预测模型的流程如图 2 所示。

##### 1.4.1 确定论域并进行划分

确定原始序列的论域  $U = [x_{\min} - \sigma_1, x_{\max} + \sigma_2]$ , 其中,  $x_{\min}$  和  $x_{\max}$  分别为原始序列的最小值和最大值,  $\sigma_1$  和  $\sigma_2$  是设定的放缩值。利用 FCM 聚类,从原始序列中得到  $k$  个聚类中心,并按升序进行排列,即聚类中心为  $c_1, c_2, \dots, c_k$ 。设相邻聚类中心  $c_r$

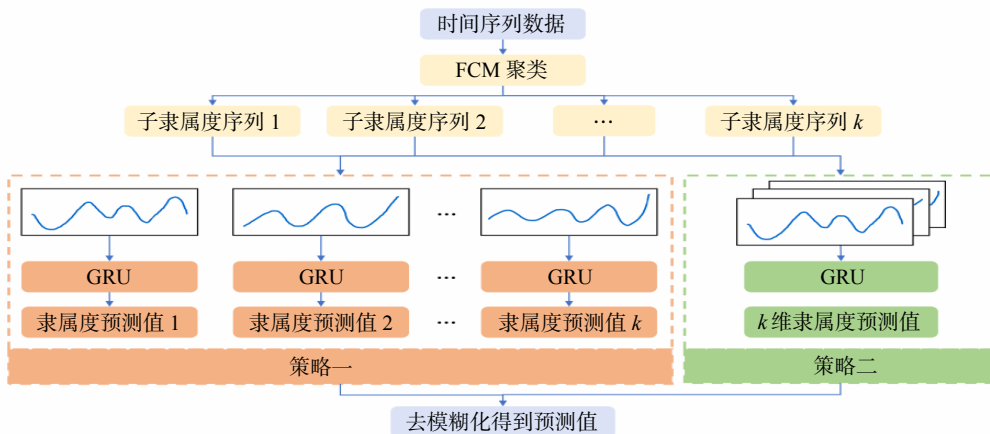


图 2 基于隶属度的模糊时间序列预测模型的流程

和  $c_{r+1}$  的中点为  $l_r$ , 则  $l_r$  是一个子论域的边界值,  $[l_r, l_{r+1}]$  是一个子论域。同理可得论域  $U$  的所有子论域  $[x_{\min} - \sigma_1, l_1], [l_1, l_2], \dots, [l_{k-1}, x_{\max} + \sigma_2]$ 。

#### 1.4.2 数据模糊化

原始序列利用 FCM 聚类得到  $k$  个聚类中心的同时, 还会得到一个隶属度矩阵, 可将此隶属度矩阵拆分为  $k$  个子隶属度序列, 即

$$Y_1 = \{u_{11}, u_{21}, \dots, u_{N1}\}, Y_2 = \{u_{12}, u_{22}, \dots, u_{N2}\}, \dots, Y_k = \{u_{1k}, u_{2k}, \dots, u_{Nk}\} \quad (10)$$

其中,  $Y_1, Y_2, \dots, Y_k$  分别是  $k$  个子论域  $[x_{\min} - \sigma_1, l_1], [l_1, l_2], \dots, [l_{k-1}, x_{\max} + \sigma_2]$  所对应的子隶属度序列;  $u_{ij} (i=1, 2, \dots, N; j=1, 2, \dots, k)$  表示第  $i$  个样本属于第  $j$  个子论域的隶属度, 且  $u_{ij} \in [0, 1]$ , 隶属度越大, 代表第  $i$  个样本隶属于第  $j$  个子论域的程度越大。

#### 1.4.3 子隶属度序列的预测

对原始序列模糊化处理后, 得到子隶属度序列  $Y_1, Y_2, \dots, Y_k$ , 可采用两种不同的策略构建基于 GRU 神经网络的子隶属度序列预测模型。策略一是将 FCM 聚类得到的  $k$  个子隶属度序列分别用 GRU 神经网络模型进行预测; 策略二是将  $k$  个子隶属度序列视为  $k$  维序列, 通过一个 GRU 神经网络模型进行预测。两种策略都具备一定适用性, 但在不同应用场景下的预测效果存在一定差异。

#### 1.4.4 去模糊化得到预测值

将聚类中心和子隶属度序列预测值利用加权平均法得到最终预测值, 算式如下:

$$G(t) = \frac{\sum_{j=1}^k c_j u_{ij}}{\sum_{j=1}^k u_{ij}} \quad (11)$$

式中:  $G(t)$  表示  $t$  时刻的预测值;  $c_j$  表示各聚类中心;  $u_{ij}$  表示  $t$  时刻对应数值隶属于第  $j$  个模糊集的程度。

## 2 实例分析

以浙江省某水厂近 5 年取水量数据作为研究对象。采样频率为每天 1 次, 每次采样记录当天的累积流量 ( $\text{m}^3$ ), 时间跨度为 2018 年 10 月 14 日至 2023 年 5 月 15 日, 共计 1 675 d。由于数据存在部分缺失, 实际可用数据总量为 1 594 条。

### 2.1 数据分析

采样数据为水厂的每日累积流量, 因此需要对现有累积流量数据进行初步处理, 转化为每日取水

量数据。用  $S_d$  表示第  $d$  日累积流量 (即累积取水量), 则第  $d$  日的取水量  $I_d$  可表示成  $I_d = S_d - S_{d-1}$ 。由于采样数据存在某些日期的缺失, 日期并非严格连续, 存在  $S_k$  紧跟着  $S_{k+m} (1 < m < 5)$  的情况, 也就意味着有  $m-1$  日的采样数据缺失。

为了保证数据连续性且便于后续模型训练, 须对缺失值进行填补。常用数据补齐方式有一维线性插值、二次插值、均值法<sup>[25-27]</sup>等。以一维线性插值和均值法两种方法为例, 考虑以下情况: 已知  $I_d$  和  $I_{d+3} (I_{d+3} > I_d)$ , 缺失  $I_{d+1}$  和  $I_{d+2}$ 。根据整体数据的大致变化, 可能存在如图 3 中“情况 1”与“情况 2”所示的两种可能的取水量变化 (假定 4 条曲线各自 4 个点累计和相同, 即累积流量相同)。在这样假设条件下, 分别采用一维线性插值和均值法补齐方式对  $I_{d+1}$  和  $I_{d+2}$  进行计算, 结果显示得到的流量值与真实值损失差距不大 (均值法在“情况 1”中的表现优于一维线性插值, 而在“情况 2”中则相反)。因此在数据量足够多的情况下, 采用任何一种补齐方式都是可行的。

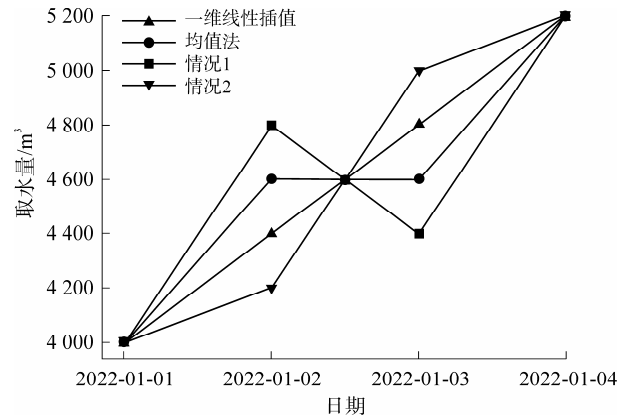


图3 数据分析示意图

与传统时间序列模型 (如 ARIMA) 不同, 深度学习下的神经网络模型 (如 CNN、RNN、LSTM 等) 凭借其与整体数据的长期依赖关系, 展现出强大的非线性能力<sup>[28]</sup>。王志良等<sup>[29]</sup>在黄河流域水文数据研究中发现, 均值法对深度学习模型的非线性补偿能力优于线性插值。均值法通过常数填充缺失值, 虽忽略局部波动, 但允许深度学习模型通过非线性激活函数自主捕捉数据模式, 有效避免了线性插值可能引入的局部趋势偏差, 从而提高模型鲁棒性。因此, 本文最终采用均值填充方式对缺失值进行填补, 相关缺失日期对应取水量算式如下:

$$S_{k+j} = S_k + j \frac{S_{k+m} - S_k}{m}, I_{k+j} = S_{k+j} - S_{k+j-1},$$

$$I_{k+1} = I_{k+2} = \dots = I_{k+m},$$

$$(2 \leq m < 5, j = 1, 2, \dots, m) \quad (12)$$

鉴于研究的每日取水量数值过大,为提升数据表现且加快模型参数收敛速度,在保证不影响原始数据数值排序的前提下,需要对原始数据进行归一化处理。因此,采用最大最小值归一化将数据映射到 0~1 之间,计算公式如下:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (13)$$

式中:  $x_i$  为  $i$  时刻数据;  $x'_i$  为  $i$  时刻的数据归一化值;  $x_{\min}$  和  $x_{\max}$  分别是数据的最小值和最大值。

将归一化后的数据通过模型进行计算,所得结果再经反归一化处理,最后得到取水量的预测值。反归一化计算公式如下:

$$\hat{y}_i = y'_i(x_{\max} - x_{\min}) + x_{\min} \quad (14)$$

式中:  $y'_i$  为模型预测值;  $\hat{y}_i$  为预测的反归一化值。

## 2.2 衡量指标

常用衡量模型有效性指标有均方误差<sup>[30]</sup>、均方根误差<sup>[31]</sup>、平均绝对百分比误差<sup>[32-33]</sup>等。

(1) 均方误差 (MSE):

$$E_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2 \quad (15)$$

(2) 均方根误差 (RMSE):

$$E_{\text{RMSE}} = \sqrt{E_{\text{MSE}}} \quad (16)$$

(3) 平均绝对百分比误差 (MAPE):

$$E_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y'_i - y_i}{y_i} \right| \times 100\% \quad (17)$$

式中:  $y_i$  为取水量数据的真实值;  $y'_i$  为模型的预测值;  $n$  为样本数量。文中选用均方根误差和平均绝对百分比误差作为主要衡量指标。

## 2.3 传统方法的实验结果

将经过预处理的 1 675 条数据进行训练集和测试集的划分,前 80%数据作为训练集用于模型的学习,后 20%数据作为测试集用于验证模型的预测精度。使用经典 GRU 网络模型对原始数据进行拟合学习。设置隐藏层个数为 2 层,时间窗口大小为 3,批处理大小为 64,迭代训练次数为 100。采用归一化后的数据值与模型预测值的均方误差 (MSE) 作为训练的损失函数,使用 Adam 优化算法进行参数

的更新。时间窗口设定为 3,表示需要用前 3 个数据去预测后 1 个数据的方式来训练模型,具体过程如图 4 所示。

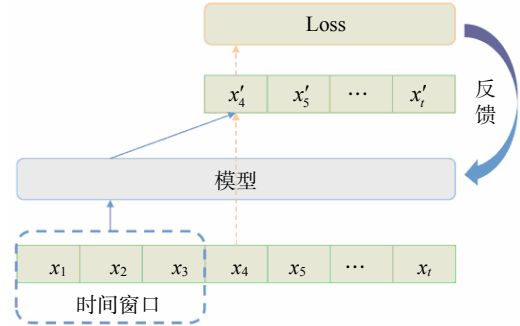


图4 模型训练过程

对训练集进行 100 次迭代训练 (Epoch), 并记录每次训练的误差值, 如图 5 所示, 可以看出整个训练阶段损失平稳下降, 并且在第 80 次迭代后趋于平稳。挑选其中误差最小的模型参数进行取水量预测, 再与测试集进行比较。如图 6 所示, 可很直观地看到模型预测能力符合预期, 其在测试集上的 MSE 和 MAPE 分别为 146 921.27 m<sup>6</sup> 和 8.42%, 这

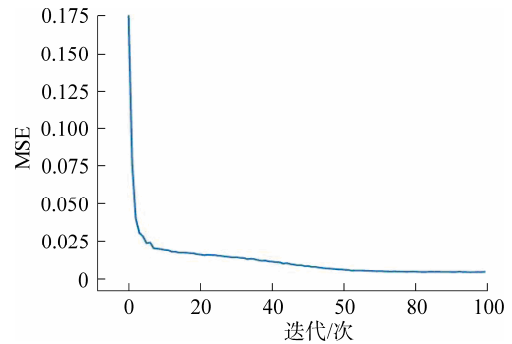


图5 损失函数的迭代训练过程

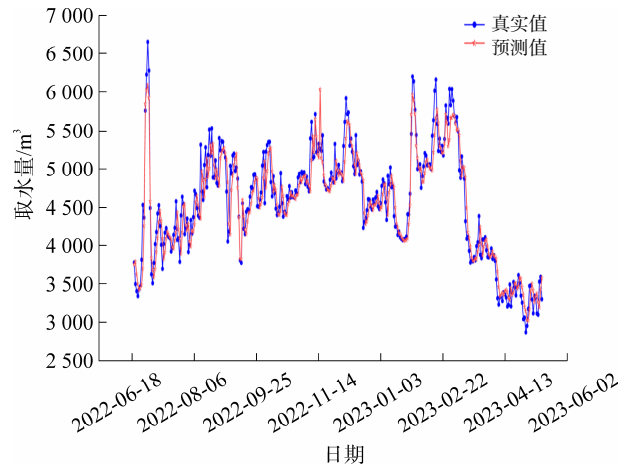


图6 GRU在测试集的预测结果

意味着模型对实际取水量的估计偏差平均不会超过  $\sqrt{146\,921.27(\text{m}^3)^2} \approx 383\text{ m}^3$ 。

建立基于注意力机制 (Attention Mechanism, AM) 优化的 GRU 神经网络模型对训练集进行拟合学习,使用的模型参数、损失函数、优化算法均与 GRU 模型相同。模型训练及测试结果如图 7 和图 8 所示,结果是在测试集上的 MSE 和 MAPE 分别为  $196\,834.28\text{ m}^6$  和  $9.71\%$ , 数据显示 AM-GRU 模型性能不如 GRU 模型。从模型机制上分析,原因可能是引入注意力机制使模型结构复杂度过高,造成过拟合问题,进而对预测结果产生负面影响;其次本例数据量不大,注意力机制难以发挥优势,甚至会受到噪声影响,降低了模型拟合效果。

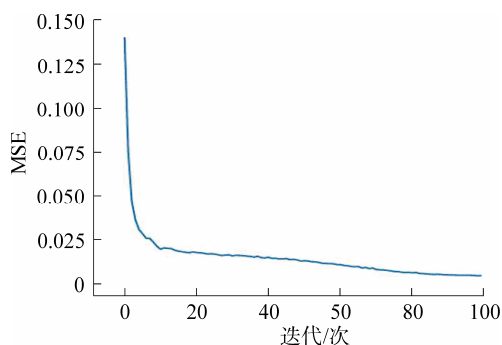


图7 损失函数的迭代训练过程

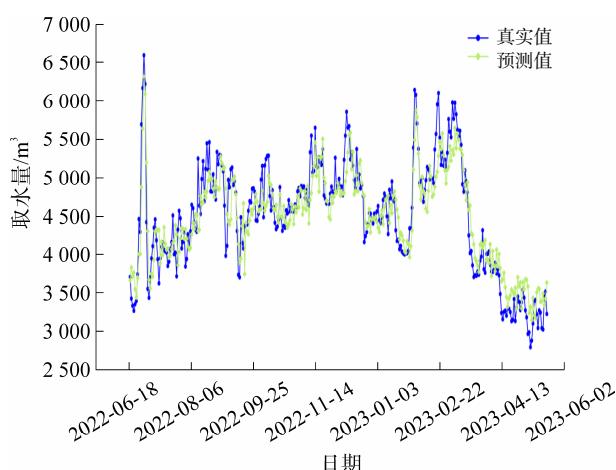


图8 AM-GRU在测试集的预测结果

进一步分析图 6 和图 8 可以发现,不论 GRU 模型还是 AM-GRU 模型,它们都没有很好地拟合极大值和极小值数据,而这些极端值多数是受到实际生活中如政策、紧急情况、突发事件等诸多不确定因素影响而产生的。事实上,多数时间序列模型并没有考虑到数据中存在的不确定因素,因而预测

精度不高,无法达到较高的预期效果<sup>[34]</sup>。

## 2.4 本文方法的实验结果

将原始数据按 8:2 比例划分为训练集和测试集,利用 FCM 聚类对取水量序列进行论域划分和数据模糊化处理,而确定最佳聚类数目对水厂取水状态的划分及后续预测结果至关重要。因此,分别设置聚类数目  $k$  为 2、3、4、5 进行聚类,加权指数  $m$  为 2,最大迭代次数为 100,终止容差  $\varepsilon$  为  $1.0 \times 10^{-5}$ 。为筛选出最优聚类数目  $k$ ,采用聚类有效性指标——划分系数 (Partition Coefficient, PC) 和划分熵<sup>[35]</sup> (Classification Entropy, CE) 进行分析对比。其中,划分系数用于衡量模糊聚类中隶属度矩阵的明确性,其值越接近于 1,表明聚类结构越清晰;划分熵用于衡量隶属度矩阵的模糊性,其值越接近于 0,表明数据点对簇的归属越集中。

对每个候选  $k$  分别进行 FCM 聚类,得到隶属度矩阵  $U = [u_{ij}]_{N \times k}$ , PC 和 CE 算式如下:

$$PC(k) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^k u_{ij}^2 \quad (18)$$

$$CE(k) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^k u_{ij} \log(u_{ij}) \quad (19)$$

对 1 594 条取水量数据进行聚类分析,通过迭代使目标函数  $I_m$  取得最小值,最终得到聚类中心,不同聚类数目的聚类效果如图 9 所示。

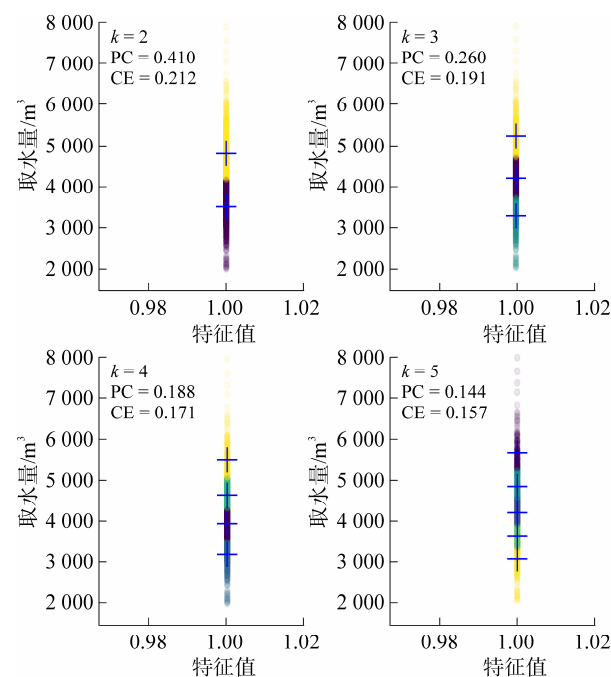


图9 不同聚类中心数目的聚类效果

聚类效果指标 PC 和 CE 的变化趋势如图 10 所示。由图可知, PC 下降速率减缓, 且 CE 优化放缓的  $k$  值为 3。

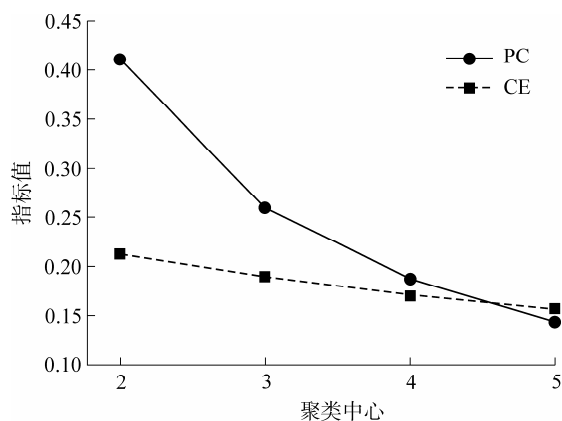


图 10 不同聚类中心数目的 PC 和 CE 变化趋势图

根据两个有效性指标的变化趋势图, 本文选定聚类数目  $k=3$ , 并将取水量划分为“低取水量”“正常取水量”和“高取水量”3 种状态, 分别对应子隶属度序列为  $Y_1$ 、 $Y_2$  和  $Y_3$ 。然后分别采用图 2 中的策略一和策略二方式对子隶属度序列  $Y_1$ 、 $Y_2$  和  $Y_3$  进行建模, 再用 GRU 神经网络模型对子隶属度序列进行预测。最后去模糊化后得到最终取水量预测值。

策略一方式在测试集上的预测结果如图 11 所示, 其均方误差 (MSE) 和平均绝对百分比误差 (MAPE) 分别为  $49\ 131.14\text{ m}^6$  和  $3.52\%$ ; 策略二方式在测试集上的预测结果如图 12 所示, 其均方误差 (MSE) 和平均绝对百分比误差 (MAPE) 分别为  $91\ 675.73\text{ m}^6$  和  $4.70\%$ 。

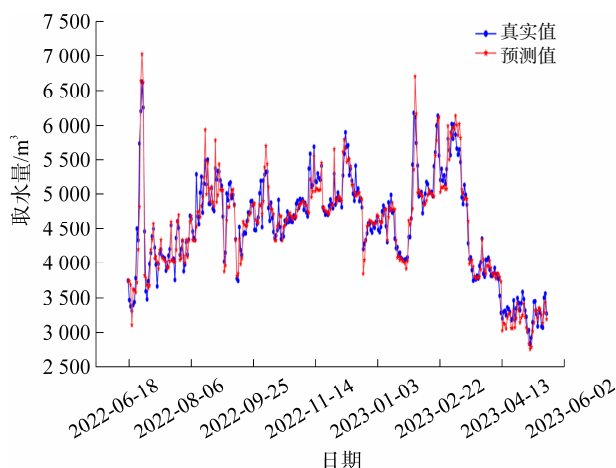


图 11 策略一方式在测试集的预测结果

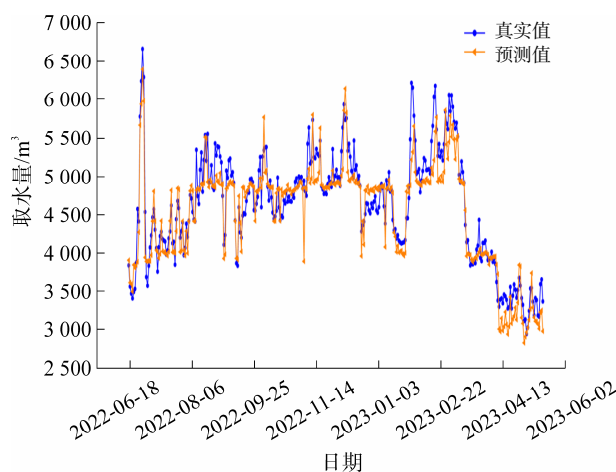


图 12 策略二方式在测试集的预测结果

## 2.5 对比分析

为验证本文方法的有效性, 将其与传统预测方法进行对比分析。表 1 为不同模型的取水量预测结果评价指标值。在对比中可发现, AM-GRU 模型预测效果稍差, 策略一方式的预测性能最好, 对比 AM-GRU 模型, RMSE 和 MAPE 分别减少了  $222.00\text{ m}^3$  和  $6.19$  个百分点。分析结果表明, 本文提出的两种方法预测性能均优于传统模型。综上所述, 本文基于隶属度的模糊时间序列预测方法在取水量预测方面展现出较高的准确性和可靠性, 为取水量预测提供了一种精度较高的解决方案。

表 1 各模型预测结果评价指标值

| 模型      | RMSE/ $\text{m}^3$ | MAPE/% |
|---------|--------------------|--------|
| GRU     | 383.30             | 8.42   |
| AM-GRU  | 443.66             | 9.71   |
| 本文方法策略一 | 221.66             | 3.52   |
| 本文方法策略二 | 302.78             | 4.70   |

## 3 结论

为实现水厂每日取水量的精准预测, 提出一种基于隶属度的模糊时间序列预测方法, 并应用于浙江省某水厂的取水量预测。结果表明, 相较于传统 GRU、AM-GRU 等常用时间序列预测模型, 新方法预测精度更高, 平均相对误差仅为  $3.52\%$ 。

新方法将模糊时间序列与 GRU 模型相结合, 既考虑了取水量数据的模糊性和不确定性, 又保证了时间序列在时间维度上的内在联系, 尤其对极端值的预测方面表现良好, 提高了取水量预测精度。

并且采用两种策略方式对子隶属度序列进行建模,最终预测结果都取得了不错的效果。但新方法在处理大规模数据集时,隶属度矩阵的迭代计算会产生 $O(n^2)$ 时间复杂度,为此未来拟采用 GPU 并行计算与稀疏矩阵优化技术以提升运算效率;后续还将探索基于强化学习的动态划分策略,进一步优化模型性能。

#### 参考文献:

- [1] 尹学康. 城市需水量预测系统的开发研究[D]. 长沙: 湖南大学, 2003.
- [2] 练庭宏, 刘秋娟, 王景成. 基于 ARIMA 时序辨识的需水量预测[J]. 控制工程, 2008, 15(增刊 1):162-164.
- [3] 景亚平, 张鑫, 罗艳. 基于灰色神经网络与马尔科夫链的城市需水量组合预测[J]. 西北农林科技大学学报(自然科学版), 2011, 39(7):229-234.
- [4] 苟非洲, 程玉婷. 基于长短期记忆网络的日供水量预测方法研究[J]. 中国给水排水, 2019, 35(17):79-83.
- [5] 陆维佳, 朱建文, 叶圣炯, 等. 基于多因素长短时神经网络的日用水量预测方法研究[J]. 给水排水, 2020, 56(1):125-129.
- [6] PANIGRAHI S, BEHERA H S. A study on leading machine learning techniques for high order fuzzy time series forecasting[J]. Engineering applications of artificial intelligence, 2020, 87:103245.
- [7] 孙平, 王丽萍, 陈凯, 等. 基于时间序列模型 ARMA 的水厂逐日需水量过程预测方法[J]. 中国农村水利水电, 2013(11):139-142.
- [8] SONG Q, CHISSOM B S. Forecasting enrollments with fuzzy time series-Part I[J]. Fuzzy sets and systems, 1993, 54(1):1-9.
- [9] SONG Q, CHISSOM B S. Forecasting enrollments with fuzzy time series-Part II[J]. Fuzzy sets and systems, 1994, 62(1):1-8.
- [10] ÖZNUR Ö K, KAYMAK Y. Prediction of crude oil prices in COVID-19 outbreak using real data[J]. Chaos, solitons & fractals, 2022, 158:111990.
- [11] 蔺玉佩, 杨一文. 基于模糊时间序列模型的股票市场预测[J]. 统计与决策, 2010, 26(8):34-37.
- [12] SEVERIANO C A, CANDIDO D L E S P, COHEN M W, et al. Evolving fuzzy time series for spatio-temporal forecasting in renewable energy systems[J]. Renewable energy, 2021, 171:764-783.
- [13] SADAIE H J, CANDIDO D L E S P, GUIMARAES F G, et al. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series[J]. Energy, 2019, 175(5):365-377.
- [14] KUMAR N, SUSAN S. Particle swarm optimization of partitions and fuzzy order for fuzzy time series forecasting of COVID-19[J]. Applied soft computing, 2021, 110:107611.
- [15] 鲜思东, 李堂金. 基于改进狼群算法的模糊时间序列预测模型[J]. 控制理论与应用, 2020, 37(7):1637-1643.
- [16] 董文超, 郭强, 张彩明. 一种模糊时间序列概率预测方法[J]. 计算机工程与科学, 2024, 46(8):1493-1502.
- [17] 王鹏, 田宗浩. 基于直觉模糊化的广义模糊时间序列预测模型[J]. 运筹与管理, 2020, 29(3):128-134.
- [18] 杨烨, 卢建刚. 基于融合 Transformer 的门尼系数预测建模研究[J]. 化工学报, 2025, 76(1):266-282.
- [19] 戴邵武, 陈强强, 刘志豪, 等. 基于 EMD-LSTM 的时间序列预测方法[J]. 深圳大学学报(理工版), 2020, 37(3):265-270.
- [20] 管业鹏, 苏光耀, 盛怡. 双向长短期记忆网络的时间序列预测方法[J]. 西安电子科技大学学报, 2024, 51(3):103-112.
- [21] 万红, 钱锐. 模糊 C-均值聚类引导的 Kinect 深度图像修复算法[J]. 计算机应用研究, 2019, 36(5):1564-1568.
- [22] HOUIMLI R, ZMAMI M, BEN-SALHA O. Short-term electric load forecasting in Tunisia using artificial neural networks[J]. Energy systems, 2020, 11(2):357-375.
- [23] KONG W C, DONG Z Y, JIA Y W, et al. Short-term residential load forecasting based on LSTM recurrent neural network[J]. IEEE transactions on smart grid, 2017, 10(1):841-851.
- [24] 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, 42(12):4045-4052.
- [25] 赵家庆, 赵裕啸, 丁宏恩, 等. 电网调度自动化主备系统间模型正确性校验技术方案[J]. 电力系统保护与控制, 2014, 42(19):139-144.
- [26] KABIR G, TESFAMARIAM S, HEMSING J, et al. Handling incomplete and missing data in water network database using imputation methods[J]. Sustainable and resilient infrastructure, 2020, 5(6):365-377.
- [27] VELASCO-GALLEGO C, LAZAKIS I. Real-time data-driven missing data imputation for short-term sensor data of marine systems: a comparative study[J]. Ocean engineering, 2020, 218:108261.
- [28] LIM B, ZOHREN S. Time-series forecasting with deep learning: a survey[J]. Philosophical transactions of the royal society A, 2021, 379(2194):20200209.
- [29] 王志良, 黄珊, 陈海涛. 黄河流域水文数据插补方法比

- 较及应用[J]. 人民黄河, 2020, 47(7):14-18.
- [30] 沐年国, 姚洪刚. 基于注意力机制的循环神经网络对金融时间序列的应用[J]. 现代电子技术, 2021, 44(14): 1-5.
- [31] 张扬. 基于改进深度神经网络的短期电力负荷预测[J]. 科技创新与应用, 2022, 12(25):12-15.
- [32] 鲜思东, 李堂金. 基于改进狼群算法的模糊时间序列预测模型[J]. 控制理论与应用, 2020, 37(7):1637-1643.
- [33] 高凯悦, 牟莉. 基于二次分解和 GRU-attention 的时间序列预测研究[J]. 国外电子测量技术, 2023, 42(2):80-87.
- [34] 丁欣, 谢祥俊, 赵春兰, 等. 基于动态隶属度的模糊时间序列在我国居民消费水平预测上的应用[J]. 模糊系统与数学, 2019, 33(1):164-174.
- [35] 林俊亭, 王帅, 刘恩东, 等. 基于模糊聚类和 CNN-BIGRU 的轨道电路故障预测方法[J]. 振动、测试与诊断, 2023, 43(3):500-507.

## Research on fuzzy time series water consumption prediction method based on membership degree

MAO Xiaoyan<sup>1</sup>, DAI Junjie<sup>1</sup>, CAO Lei<sup>2</sup>

( 1.School of Basic Education, College of Science and Technology Ningbo University, Ningbo 315300, China;

2.Jiangsu Jinheng Information Technology Co., Ltd., Nanjing 210044, China )

**Abstract:** The scientific and accurate prediction of water consumption from water plants is crucial for the efficient management and scientific decision-making of water resources, and is a prerequisite for promoting smart water conservancy construction. Due to the nonlinearity, fuzziness, and other complex characteristics in the water consumption sequence, and those affected by a variety of uncertainties, the accuracy of traditional prediction models is limited to some extent. This paper combines the characteristics of fuzzy theory and gated recurrent unit neural network to propose a new method for predicting water consumption in fuzzy time series based on membership degree. First, the fuzzy *C*-means clustering (FCM) is used to construct sub membership sequences from the original sequence; second, the GRU model is used with two strategies to predict each sub membership degree sequence; finally, the predicted value of water consumption is obtained by defuzzification. The new model proposed in this paper is employed to predict the water consumption sequence of a water plant in Zhejiang Province, with its results being compared with those from the traditional GRU model and AM-GRU model for analysis. The experimental results show that the new model outperforms other models in terms of MSE, RMSE, and MAPE, and it improves prediction accuracy.

**Keywords:** water consumption; membership; fuzzy time series; fuzzy *C*-means clustering; recurrent neural network; gated recurrent unit

( 责任编辑 章践立 )