

DOI:10.3876/j.issn.1000-1980.2022.04.010

机器学习模型在地下水埋深模拟中的适应性分析

牛欣怡,鲁程鹏,卢佳赞,吴成城,刘 波,束龙仓

(河海大学水文水资源学院,江苏 南京 210098)

摘要: 基于京津冀气象、社会资料及地下水埋深数据,构建支持向量机(SVM)、循环神经网络(RNN)和长短期记忆神经网络(LSTM)模型对京津冀地区 13 个城市地下水埋深进行了模拟,并以确定系数、均方根误差、平均绝对百分比误差、纳什系数对 3 个模型的适应性进行了评价。结果表明:LSTM 模型模拟效果最好,其次是 RNN,SVM 最差;不同城市基于 LSTM 模型进行地下水埋深模拟时参数调整最少,适应性最好,SVM 模型参数调整最多。将 3 个模型应用于随机选择的 6 个测站进行验证,在华北地区浅层地下水埋深模拟方面,LSTM 模型模拟精度和可信度最好,适应性最强,是该地区地下水埋深模拟的首选机器学习模型。

关键词: 机器学习;地下水埋深;支持向量机;长短期记忆神经网络;循环神经网络;京津冀

中图分类号:TV211.1⁺² **文献标志码:**A **文章编号:**1000-1980(2022)04-0074-09

Adaptability analysis of machine learning model in groundwater depth simulation

NIU Xinyi, LU Chengpeng, LU Jiayun, WU Chengcheng, LIU Bo, SHU Longcang

(College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)

Abstract: Based on the meteorological data, social data and groundwater depth data of Beijing-Tianjin-Hebei, support vector machine (SVM), recurrent neural network (RNN) and long and short-term memory neural network (LSTM) were constructed to simulate the groundwater depth of 13 cities. The adaptability of three models were evaluated in indicators, such as the determination coefficient, root mean square error, mean absolute percentage error, and Nash coefficient. The results showed that LSTM model performed the best, followed by RNN and SVM. Meanwhile, simulation results of different cities demonstrated the least parameter adjustments and the best adaptability of LSTM model for the groundwater depth simulation, and SVM model parameters had the most parameters adjustments. Three models were applied to 6 randomly selected stations for verification and it was shown that in the shallow groundwater depth simulation in North China, LSTM model had good simulation accuracy and reliability, with strong adaptability. Therefore, it is the first choice in the groundwater depth simulation of the studying area.

Key words: machine learning; groundwater depth; support vector machine; long and short-term memory neural network; recurrent neural network; Beijing-Tianjin-Hebei

1972 年,我国地下水开采量为 196 亿 m³ 左右,至 2014 年末,地下水开采量已达到 1 117 亿 m³,约为 1972 年的 6 倍,在北方一些缺水的地区,如河北、北京、河南等地,其地下水供水量已达到总供水量的 70% 左右^[1]。由于地下水的超量开采,导致很多地区出现海水入侵、地面塌陷、地下水干旱等严重的地质和生态问题^[2-5]。准确模拟地下水埋深对于地下水合理开发利用和保护具有重要的指导意义^[6]。地下水模拟模型大体分为物理模型和数据驱动模型^[7]。基于详细的水文地质资料构建的物理模型主要有具有物理意义的 MODFLOW、GMS 等数值模型,但是物理模型在求解过程中具有工作量大、时间长、模型校准复杂等缺点^[8-9]。人工神经网络等数据驱动模型可以轻松逼近物理系统的复杂行为和响应,并且能够在短时间内优化不同约束集的大量案例场景。相比于物理模型中较多的假设条件、复杂的输入变量和参数精确校准等问题,数据驱

基金项目: 国家重点研发计划(2021YFC3200502);江苏省水利科技项目(2018005);国家自然科学基金(41971027)

作者简介: 牛欣怡(1998—),女,硕士研究生,主要从事水文学及水资源研究。E-mail:3196860667@qq.com

通信作者: 鲁程鹏(1984—),男,教授,博士,主要从事水文学及水资源研究。E-mail:luchengpeng@hhu.edu.cn

引用本文: 牛欣怡,鲁程鹏,卢佳赞,等. 机器学习模型在地下水埋深模拟中的适应性分析[J]. 河海大学学报(自然科学版),2022,50(4):74-82.

NIU Xinyi, LU Chengpeng, LU Jiayun, et al. Adaptability analysis of machine learning model in groundwater depth simulation[J]. Journal of Hohai University(Natural Sciences),2022,50(4):74-82.

动模型的输入变量更易测量和量化,数据限制更低,模型参数适用性更强,而且模拟时间也比较短^[10]。因此,机器学习这类数据驱动模型在埋深模拟中前景广泛,尤其是在缺少水文地质勘测资料的地区,是不可替代的。

支持向量机(support vector machines,SVM)、循环神经网络(recurrent neural network,RNN)和长短期记忆神经网络(long short-term memory,LSTM)是目前在地下水埋深模拟方面应用最广泛的机器学习模型,但由于 SVM 模型和 RNN 模型是基于不同的理论构建,导致其在模拟效果方面有一定的差异^[11-14]。目前国内学者在研究基于机器学习模拟地下水方面主要集中在单一模型的改进和应用,如闫佰忠等^[15-16]基于多变量 LSTM 模型模拟监测井的地下水水位,并与单变量 LSTM、反向传播神经网络模拟结果对比;张钧泳等^[17]建立不同参数的 SVM 模型,探讨单因子、多因子建模反演地下水埋深精度问题;张文鸽等^[18]结合混沌优化方法,构建基于混沌优化的峰值识别最小二乘支持向量机模型,模拟义长灌区地下水位,并将结果与最小二乘支持向量机模拟结果进行比较。然而,对于同一地区,对基于不同理论方法的模拟效果进行比较,尤其是适应性方面的研究比较少。

本文采用 SVM 模型、RNN 模型和 LSTM 模型进行地下水埋深模拟,通过模拟京津冀平原地区 13 个城市地下水埋深情况,比较 3 种模型的模拟精度和适应性,并将 3 种模型应用到随机选择的 6 个测站,以验证更适用于地下水埋深模拟的机器学习模型,以期为今后地下水埋深模拟模型选择以及参数调整提供参考。

1 模型构建

1.1 地下水埋深模拟模型框架

本文气象(降水、平均气温、平均风速)和地下水埋深数据分别来自中国气象数据网记录的 24 个气象观测站和各地区 711 个地下水监测站实测的 2005—2018 年 168 个月的日观测数据,基于月均值利用普通克里金法插值得到京津冀 13 个城市的相应月数据。社会经济类要素(人口、经济生产总值)和地下水开采量来源于各个城市相应年份经济年鉴的年度值,为了计算方便,均匀分配到各月。随后构建 RNN 模型、LSTM 模型和 SVM 模型。在输入变量方面,虽然以多变量因素作为输入比单变量好^[19],但是当因素与地下水相关性较弱时,反而会成为一种噪声影响模型的模拟精度^[20-22]。为了提高地下水埋深模拟模型的准确性,基于 Pearson 相关系数去除与地下水位相关性不显著的变量^[23-24]。随后将显著因素和上月地下埋深作为输入变量进入模型训练。为了保证能够得到较好的训练,防止模型对训练数据的过度拟合,将样本按照时间顺序以训练数据与测试数据为 11 : 3 划分。最后在京津冀地区随机选择 6 个测站的上月埋深和显著因素作为输入变量进入训练好的模型中验证。以上过程通过 ArcGIS 10.3.1 和 Python3.6 实现,模型框架图见图 1。

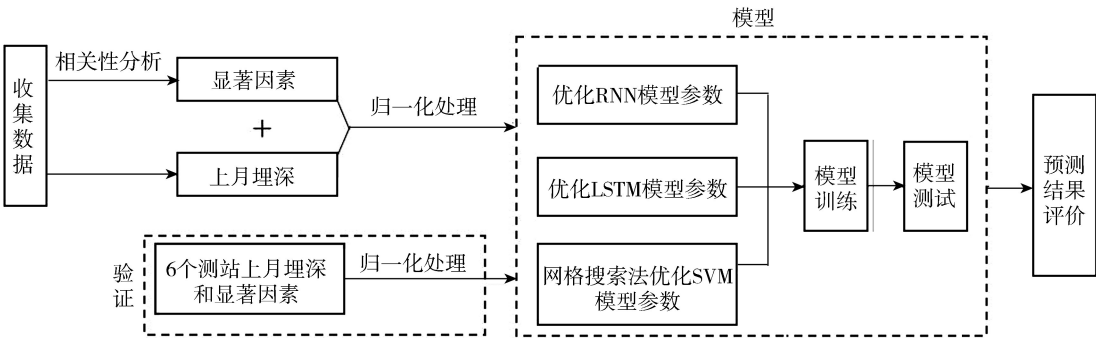


图 1 地下水模拟模型框架

Fig.1 Framework diagram of groundwater simulation model

1.2 SVM

SVM 模型是以统计理论为基础,主要的应用方向是分类、回归。SVM 法的特点是在小样本问题及低维空间数据集不可分的前提下,将其映射到高维空间,从而达到可分的效果^[25],该数据处理方式相对于神经网络可以从很大程度上避免学习或局部极值的出现^[26]。SVM 法的回归计算如下。

首先给定一组训练样本 $M = \{ (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \}$, $(x_i, y_i) \in R (i = 1, 2, \dots, n)$ 。以简单线性式表示:

$$f(\boldsymbol{x}) = \boldsymbol{w}\boldsymbol{x} + \boldsymbol{b} \tag{1}$$

其中 $\boldsymbol{x} = (x_1, \dots, x_n)$ $\boldsymbol{w} = (w_1, \dots, w_n)$ $\boldsymbol{b} = (b_1, \dots, b_n)$

式中: $\boldsymbol{x}$ 为特征向量; $\boldsymbol{w}$ 为超平面的法向量; $\boldsymbol{b}$ 为位移向量。这两个因素决定超平面的方向。

寻找最小的 $\boldsymbol{w}$ 来保证式(1)的平坦。假设在一定精度的 δ 下, 训练样本所有数据都可以用式(1)拟合, 则可将问题转化为凸优化解决:

$$\min J(\boldsymbol{w}) = \frac{1}{2} \|\boldsymbol{w}\|^2 \tag{2}$$

$$\text{s. t. } \begin{cases} y_i - w_i x_i - b_i < \delta \\ w_i x_i + b_i - y_i \leq \delta \end{cases} \quad (i = 1, 2, \dots, n) \tag{3}$$

然后通过引入松弛因子 $\varepsilon_i \geq 0$ 和 $\varepsilon_i^* \geq 0$ 来拟合误差, 将回归估计问题转化成下列问题:

$$\min \frac{1}{2} \|\boldsymbol{w}\|^2 + A \sum_{i=1}^n (\varepsilon_i + \varepsilon_i^*) \tag{4}$$

约束条件:

$$\begin{cases} y_i - w_i x_i - b_i \leq \delta + \varepsilon_i \\ w_i x_i + b_i - y_i \leq \delta + \varepsilon_i^* \\ \varepsilon_i \geq 0 \\ \varepsilon_i^* \geq 0 \end{cases} \quad (i = 1, \dots, n) \tag{5}$$

式中 A 为正则化常数, 用来衡量回归函数的平坦度。再通过拉格朗日乘法求解, 转化成对偶问题。

$$f(\boldsymbol{x}) = \boldsymbol{w}\boldsymbol{x} + \boldsymbol{b} = \sum_{i=1}^n (a_i - \bar{a}_i)(x_i y_i) + b_i \tag{6}$$

对于非线性问题来说, 最后公式表示为

$$f(\boldsymbol{x}) = \boldsymbol{w}\boldsymbol{x} + \boldsymbol{b} = \sum_{i=1}^n a_i y_i K(x_i, x_j) + b_i \tag{7}$$

式中 $K(x_i, x_j)$ 表示各种核函数如线性核、高斯核等。

1.3 RNN

RNN 模型被专门设计用来捕捉时间序列数据中固有的结构^[14], 在模型中是通过传递隐藏层神经元的输出或者状态来实现这一点, 这些神经元代表了在之前时间步骤中学习到的东西, 并且作为下一个时间步骤的额外输入^[27]。RNN 模型中的计算式可以写为

$$h_t = \tanh(Wx_t + Uh_{t-1} + c) \tag{8}$$

$$y_t = Vh_t + c \tag{9}$$

式中: h_t 为隐藏状态; y_t 为输出向量; x_t 为输入向量; W 、 U 、 V 分别为输入、隐藏和输出权值; c 为偏置。RNN 模型的训练采用时间反向传递, 基于神经网络的权值和之前隐藏状态的误差梯度来调整网络权值。由于梯度在这一过程可以呈指数变化, 因此它们往往会消失或者爆炸^[28]。学者们针对这一问题, 又对 RNN 模型进行优化, 提出了长短期记忆神经网络模型。

LSTM 模型是 RNN 模型改进的一种类型, LSTM 模型架构在不激活函数的情况下, 在隐藏单元状态之间强制执行恒定误差流, 从而将梯度问题最小化。LSTM 单元包含 3 个作为门的乘法单元: 遗忘门、输入门和输出门^[29]。因为每个门都是神经元, 它可以通过传递输入、反向传播误差和调整权重的过程来学习哪些输入和细胞状态对模拟输出是重要的。

1.4 数据归一化处理

在模型输入和输出中, 由于不同类型的数据存在着极端值和较多的噪音, 需要进行归一化处理来消除异常值和极端值的影响, 以此来使模型中寻求最优解的过程变得平缓, 加快梯度下降的速度, 并且提高模型精度^[30-31], 本文采用线性归一化对数据进行处理:

$$z^* = \frac{z_{\min}}{z_{\max} - z_{\min}} \tag{10}$$

式中: z 为原始数据; z^* 为无量纲化后的数据; $z_{\max}$ 、 $z_{\min}$ 分别为原始数据中的最大最小值。

1.5 模型评价指标

本文采用确定系数 R^2 、均方根误差 RMSE、平均绝对百分比误差 MAPE^[32]、纳什效率系数 NSE 对 3 种模型模拟效果进行评价。

2 模型实例对比分析

2.1 相关性分析

地下水埋深与影响因素之间的相关性如表 1 所示。京津冀地区中大部分地区主要影响地下水埋深的因素是人口、经济、开采量和降雨,其中人口、经济、开采量的相关性系数都比较高,大体上都达到 0.5 以上,而降水的相关系数大多保持在 0.3 左右,由此可以看出人类活动对于地下水埋深的影响占主导地位,自然因素的影响占比较小。从表 1 可以看出,除了天津、廊坊、沧州,北京和河北大部分城市地下水埋深和开采量呈现负相关,出现这种情况的主要原因是在研究期间地下水开采量虽然逐年减少,但是总量相对较大。如北京市、河北省年均开采量基本保持在 20 亿 m³、140 亿 m³ 左右,且地下水埋深的恢复具有一定的时间滞后性,因此从表面上看开采量逐年减少,地下水埋深逐年增大,但实际上地下水埋深增大的速率在逐渐减小。而天津的年开采量在 5 亿 m³ 左右,当开采量减少时,地下水埋深恢复效果更显著。

表 1 京津冀 13 个城市地下水埋深与多变量之间的相关性

Table 1 Correlation between groundwater depth and multivariate variables in 13 cities of Beijing-Tianjin-Hebei						
地区	相关性					
	开采量	降水	风速	温度	人口	经济
北京	-0.73 **	0.22 **	-0.21 **	0.14	0.87 **	0.67 **
天津	0.70 **	0.14	0.05	0.11	-0.80 **	-0.78 **
石家庄	-0.83 **	0.15 *	-0.02	0.13	0.96 **	0.95 **
唐山	-0.09	0.34 **	-0.02	0.36 **	-0.14	-0.30 **
保定	-0.45 **	0.46 **	-0.28 **	0.34 **	-0.09	0.47 **
廊坊	0.51 **	0.36 **	-0.09	0.32 **	-0.47 **	-0.48 **
张家口	-0.59 **	0.17 *	-0.04	0.08	0.96 **	0.95 **
承德	-0.39 **	0.21 **	0.12	0.16 *	0.68 **	0.70 **
沧州	0.17 *	0.45 **	0.01	0.37 **	-0.25 **	-0.35 **
秦皇岛	-0.61 **	0.15 *	0.12	0.18 *	0.33 **	0.35 **
衡水	-0.48 **	0.40 **	-0.14	0.39 **	0.28 **	0.34 **
邢台	-0.53 **	0.24 **	0.08	0.25 **	0.83 **	0.80 **
邯郸	-0.11	0.40 **	0.07	0.43 **	0.12	0.13

注: * 表示在 $\alpha=0.05$ 水平上显著相关, ** 表示在 $\alpha=0.01$ 水平上显著相关。

2.2 模拟结果分析

基于 RNN 模型、LSTM 模型、SVM 模型得到的结果用 R^2 、RMSE、MAPE、NSE 指标进行评价,结果见表 2,相应的 4 个误差指标箱型图见图 2,模型运行时间见图 3。

表 2 3 种模型性能指标结果

Table 2 Indicator results of three model performance												
城市	R^2			RMSE			MAPE			NSE		
	RNN 模型	LSTM 模型	SVM 模型	RNN 模型	LSTM 模型	SVM 模型	RNN 模型	LSTM 模型	SVM 模型	RNN 模型	LSTM 模型	SVM 模型
石家庄	0.60	0.72	0.54	0.61	0.56	1.11	1.47	1.26	2.81	0.21	0.32	-1.62
保定	0.56	0.63	0.54	0.83	0.71	0.97	2.58	1.97	3.34	0.05	0.30	-0.30
承德	0.86	0.88	0.75	1.17	0.90	1.64	3.17	2.06	5.03	0.61	0.77	0.24
沧州	0.60	0.63	0.52	0.80	0.71	1.02	9.23	7.93	12.35	0.10	0.16	-0.08
秦皇岛	0.56	0.68	0.49	0.77	0.38	0.86	10.74	4.39	11.89	-1.28	0.45	-1.83
北京	0.77	0.89	0.36	0.93	0.57	0.96	2.51	1.71	2.66	0.34	0.67	0.11
廊坊	0.66	0.73	0.48	0.76	0.53	0.82	5.36	4.38	6.42	-1.38	0.14	-1.78
张家口	0.76	0.84	0.34	0.51	0.29	0.60	1.51	0.79	1.61	0.04	0.68	-0.32
邯郸	0.43	0.69	0.30	1.04	0.63	1.16	4.48	2.47	4.86	-0.42	0.47	-0.78
天津	0.59	0.63	0.50	0.36	0.32	0.62	5.95	4.73	10.52	0.21	0.37	-1.34
唐山	0.72	0.75	0.65	0.71	0.59	1.83	5.36	3.95	14.49	0.11	0.38	-1.96
邢台	0.57	0.70	0.49	0.95	0.84	1.03	2.65	2.16	2.98	0.19	0.36	0.03
衡水	0.59	0.68	0.23	0.79	0.71	1.12	5.22	4.32	7.08	0.01	0.20	-0.99

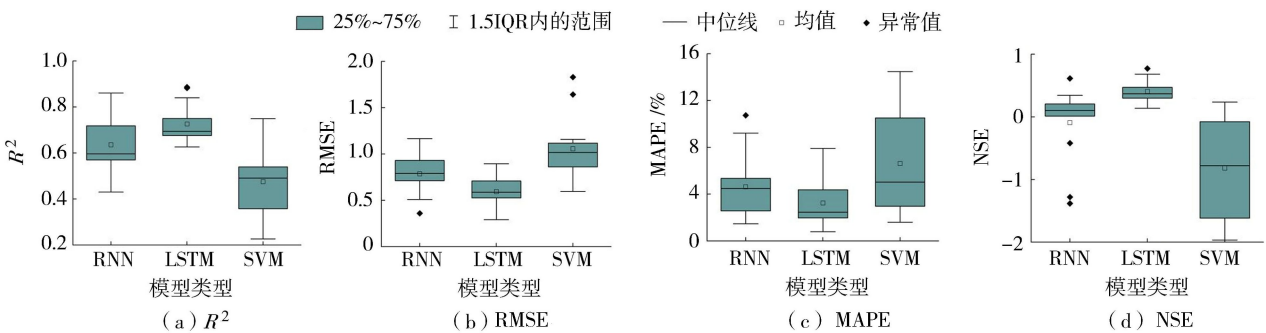


图2 不同模型模拟误差指标分布箱型图

Fig.2 Boxplots of simulation error indexes of groundwater depth simulation with three models

由表2可知,LSTM模型的拟合精度最高,其次是RNN模型,SVM模型表现最差。LSTM模型的拟合精度始终保持在0.6~0.9之间,最大为0.89,最小为0.63,均具有高度相关性,其中有54%地下水埋深数据的拟合度在0.6~0.7,23%的数据拟合度在0.8~0.9之间。利用RNN模型进行模拟的城市一半以上地下水埋深拟合度在0.5~0.6之间,拟合度在0.8~0.9的数据仅占8%。而基于SVM模型进行模拟的城市中,有62%的城市地下水埋深拟合精度在0.2~0.5之间,相关性较低。从图2(a)可以明显看出LSTM模型箱型图分布宽度最窄,SVM箱型图分布宽度最宽,这体现了LSTM模型应用在不同城市的适应性在3个模型中是最好的。

从RMSE、MAPE和NSE3项指标来看,不同城市模型的误差排序从大到小均为SVM模型、RNN模型、LSTM模型。结合3项误差的箱型图可以明显看出,在不同的指标下,LSTM模型的箱型图分布最窄,SVM箱型图分布最宽。在RMSE和NSE两个指标下RNN模型虽然与LSTM模型分布宽度相差不大,但是均出现异常值,而且异常点的个数比LSTM模型多,这是因为RNN模型存在梯度消失或爆炸问题,而LSTM模型可以克服这些问题,使模型运行得更加稳定,从而误差波动范围也小。在3个指标下,SVM模型箱型图分布宽度均大于RNN模型,这是信息流在模型中传递的方向导致,RNN模型具有记忆功能,信息流在模型中是双向传递,而SVM模型中信息流仅仅只是单向传递。综合上述模型指标的结果显示,LSTM模型的泛化能力明显比RNN模型、SVM模型优秀。

从图3可以看出LSTM模型运行时间和SVM模型运行时间基本一致,保持在5.5s左右,LSTM模型和SVM模型运行时间长但波动较小,RNN模型运行时间最短但波动范围最大。出现这种情况的原因除了上文提到的LSTM模型可以消除RNN模型中的梯度消失和爆炸问题外,SVM模型运行时间和LSTM模型运行时间基本一致的原因是SVM模型运行时间包括参数调整和模拟两部分,其中由于SVM模型过多的核函数和相应的参数导致参数调整的时间过长,而SVM模型在模拟部分的时间比较短,是其他模型运行时间的1/10。

3 模型参数适应性和稳定性分析

3.1 参数适应性

在模型应用中,针对不同地区的情况需要适当地进行参数调整,对于SVM模型,由于人工搜索参数存在较大的误差,因此本文采用网格搜索法^[33]对SVM参数进行搜索,得到模型应用在各城市的参数,见表3。

从表3可以看出,虽然13个城市所处的地理位置和气候条件不同,但是在利用模型模拟时有相似之处。RNN模型模拟13个城市的地下水埋深,除张家口、邯郸、衡水的学习率为0.001外,其他城市学习率均为0.01,隐藏层的神经元数量全部保持在30以内,并且数量为30和10的城市占大多数,共达9个,迭代次数大部分都是1000和1500;在LSTM模型下学习率均为0.001,隐藏层神经元除了邯郸、天津、衡水、邢台,其他城市均为30,迭代次数除了保定、天津、衡水,其他均为1500次。除此之外,也可以明显看出张家口、邯郸的RNN模型和LSTM模型的参数完全一样,但是由于LSTM模型比RNN模型更能消除梯度爆炸等因素,导

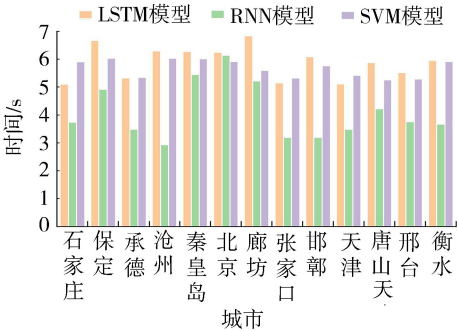


图3 京津冀13个城市地下水模型运行时间
Fig.3 Running time of three models applied in 13 cities of Beijing-Tianjin-Hebei

表 3 不同城市地下水埋深模拟的模型参数

Table 3 Model parameters of groundwater depth simulation in different cities

城市	RNN 模型			LSTM 模型			SVM 模型		
	L	N	E	L	N	E	核函数	C	g
石家庄	0.01	10	1000	0.001	30	1500	rbf	10	0.1
保定	0.01	30	1500	0.001	30	2000	rbf	10	0.1
承德	0.01	10	1000	0.001	30	1500	linear	1	
沧州	0.01	30	800	0.001	30	1500	rbf	10	1
秦皇岛	0.01	15	1500	0.001	30	1500	rbf	30	0.1
北京	0.01	20	2000	0.001	30	1500	sigmoid	8	0.1
廊坊	0.01	30	1500	0.001	30	1500	sigmoid	1	0.1
张家口	0.001	30	1500	0.001	30	1500	poly	0.1	
邯郸	0.001	10	1500	0.001	10	1500	sigmoid	10	0.1
天津	0.01	10	1500	0.001	10	1000	sigmoid	10	0.1
唐山	0.01	10	1000	0.001	30	1500	linear	2	
邢台	0.01	15	1000	0.001	15	1500	sigmoid	10	0.1
衡水	0.001	15	600	0.001	10	1000	rbf	10	1.0

注: L 为模型学习率; N 为隐藏层神经元数量; E 为模型迭代次数; C 为惩罚参数; g 为选择 rbf 函数后自带的参数 gamma。

致 RMSE 和 MAPE 都比 RNN 模型小,使得模拟结果更加精确;在 SVM 模型下,石家庄、保定、沧州、秦皇岛和衡水 5 个城市选择的是 rbf 函数,北京、廊坊、邯郸、天津、邢台选择的是 sigmoid 函数,承德、唐山选择的是 linear 函数,仅有张家口选择的是多项式函数。

综合上述 3 个方面,在 13 个城市中,LSTM 模型变动参数最少,运用参数为 5 组;SVM 模型运用参数次之,运用的参数共达 9 组;RNN 模型运用的参数最多,达到 10 组。但是从参数变量上来看,SVM 模型参数变量最多,应用的核函数有 4 个,再包括 C 和 g 一共有 6 个参数,是其他两个模型的 2 倍,因此虽然 SVM 模型应用的参数组数少,但是根据其模型运行时间来看,其适应性程度低于 RNN 模型。所以在考虑模型参数的适应性以及精确程度方面,首选 LSTM 模型,其次是 RNN 模型,最后是 SVM 模型。

3.2 模拟稳定性

为了验证 RNN 模型、LSTM 模型、SVM 模型模拟的稳定性,随机选择了京津冀地区 6 个测站进行地下水埋深模拟,讨论 3 个模型在新测站的适应性情况,模拟结果如图 4 所示,随后采用 R^2 、RMSE、NSE 指标评价模拟结果,相关参数和指标计算结果见表 4。

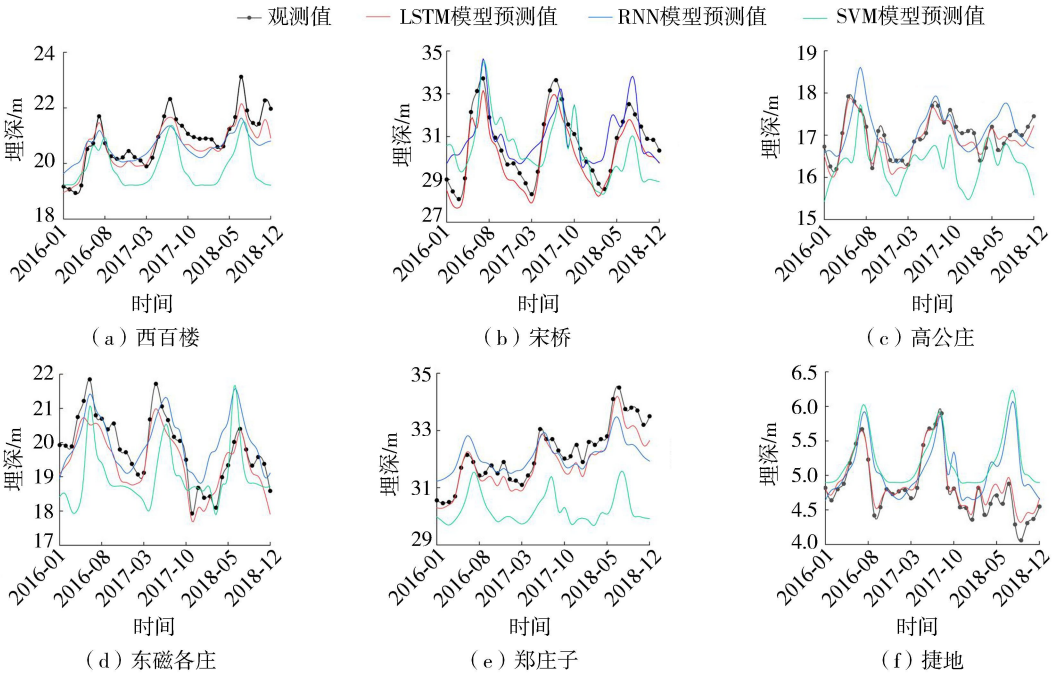


图 4 不同模型 6 个测站地下水埋深模拟结果对比

Fig.4 Simulation result comparison of groundwater depth at 6 measuring stations based on different models

表4 6个测站埋深模拟参数和误差指标

Table 4 Simulation parameters and error indicators for groundwater depth of 6 measuring stations

测站	RNN 模型						LSTM 模型						SVM 模型					
	<i>L</i>	<i>N</i>	<i>E</i>	<i>R</i> ²	RMSE	NSE	<i>L</i>	<i>N</i>	<i>E</i>	<i>R</i> ²	RMSE	NSE	<i>K</i>	<i>C</i>	<i>g</i>	<i>R</i> ²	RMSE	NSE
西百楼	0.01	15	1000	0.858	0.644	0.530	0.001	30	1500	0.950	0.395	0.824	poly	5.0		0.582	1.268	-0.823
宋桥	0.01	15	1000	0.755	1.110	0.512	0.001	30	1500	0.982	0.654	0.831	rbf	1.0	0.4	0.602	1.359	0.269
高公庄	0.01	15	1000	0.574	0.466	-0.010	0.001	30	1500	0.941	0.193	0.827	sigmoid	0.5	1.0	0.374	0.858	-2.429
东磁各庄	0.01	15	800	0.669	0.744	0.392	0.001	30	1500	0.921	0.545	0.674	poly	8.0		0.479	1.240	-0.688
郑庄子	0.01	15	1000	0.796	0.698	0.536	0.001	30	1500	0.979	0.409	0.840	rbf	1.0	0.1	0.406	2.179	-3.525
捷地	0.01	15	1000	0.501	0.484	-0.221	0.001	30	1500	0.991	0.103	0.944	sigmoid	30.0	0.2	0.513	0.569	-0.683

由图4可以直观看出,基于 LSTM 模型模拟的地下水埋深和实测值拟合效果最好,其次是 RNN 模型、SVM 模型。对于 LSTM 模型来说,其基本能够将埋深的每个转折趋势很好地模拟出来。表4结果显示,LSTM 模型在6个站的拟合精度都达到0.9以上, RMSE 最小, NSE 基本都达到0.8以上,表明该模型是可靠的。RNN 模型模拟效果相比于 LSTM 模型来说,转折和峰值表现效果比较差,如宋桥站、高公庄站、郑庄子和捷地站,甚至高公庄和捷地的 NSE 指标为负值,表明模型的结果不可信。对于 SVM 模型来说,只能将埋深的变化趋势大致描述出来,在细节方面忽略比较多,这是由于 SVM 模型中没有记忆功能导致的。而且 SVM 模型的 NSE 系数除了宋桥站,其他均为负数,且比 RNN 模型大得多,表明 SVM 模型在模拟华北平原大部分城市或测站的浅层地下水埋深时稳定性差。这是因为传统的机器学习方法在模拟地下水埋深与输入变量之间的关系方面有限,当输入变量较多时循环神经网络结构可以更好地整合多特征输入和输出变量之间的非线性关系^[34-35]。

从参数的适应性角度来看,6个测站的地下水埋深模拟中,LSTM 模型仅利用1组参数,RNN 模型利用2组参数,而 SVM 模型利用了6组参数,这也表明 LSTM 模型的适应性最优。

4 结 论

a. 在京津冀地区13个城市的地下水埋深模拟中,LSTM 模型模拟效果最好,其次是 RNN 模型,SVM 模型最差,主要原因在于 RNN 模型具有保留记忆的功能,可以从隐藏神经元中提取信息,而 LSTM 模型在此基础上进一步解决了梯度消失的问题,但 SVM 模型只能从与因变量密切相关的数据中提取特征,当处理稍长的数据集时,无法解决数据间的长期依赖问题。在运行时间方面,RNN 模型最短,LSTM 模型其次,SVM 模型由于参数过多,运行时间最长。

b. 在参数适应性方面,不同城市基于 LSTM 模型进行模拟时调整的参数最少,虽然 RNN 模型和 SVM 模型应用的参数组数相似,但是由于 SVM 模型有过多的核函数,因此 RNN 模型参数适应性较好。

c. 将3个模型分别应用到随机选择的6个测站中进行验证,结果表明 LSTM 模型适应性最好,不仅 *R*² 高达0.9以上,NSE 也达到0.8以上,6个测站仅应用了1组参数;RNN 模型效果其次,SVM 模型效果最差,6个测站中有5个测站的 NSE 为负值,没有成功应用。

参考文献:

[1] 卫磊,杨桂莲,鲁程鹏,等. 华北平原超采区浅层地下水埋深变化及控制因素分析[J]. 水资源与水工程学报,2019,30(6):39-44. (WEI Lei,YANG Guilian,LU Chengpeng,et al. Analysis of buried depth variation and controlling factors of shallow groundwater overexploited areas of in North China Plain[J]. Journal of Water Resources and Water Engineering,2019,30(6):39-44. (in Chinese))

[2] 贺国平,周东,杨忠山,等. 北京市平原区地下水资源开采现状及评价[J]. 水文地质工程地质,2005(2):45-48. (HE Guoping,ZHOU Dong,YANG Zhongshan,et al. Current status and evaluation of groundwater resources exploitation in the plain area of Beijing[J]. Hydrogeology & Engineering Geology,2005(2):45-48. (in Chinese))

[3] 谢纪强,游进军,姜纪沂. 基于级差最大化组合赋权法的地下水功能评价[J]. 水资源保护,2021,37(6):54-59. (XIE Jiqiang,YOU Jinjun,JIANG Jiyi. Using level difference maximization method to calculate combined weights for groundwater function evaluation[J]. Water Resources Protection,2021,37(6):54-59. (in Chinese))

[4] 粟晓玲,姜田亮,牛纪革. 生态干旱的概念及研究进展[J]. 水资源保护,2021,37(4):15-21. (SU Xiaoling,JIANG Tianliang,NIU Jiping. Concept and research progress of ecological drought[J]. Water Resources Protection,2021,37(4):15-21.

(in Chinese))

- [5] 王丽娟,宋翘楚,吴为,等. 江苏沿海滩涂地下水埋深及矿化度变化规律[J]. 水利水电科技进展,2021,41(6):52-57. (WANG Lijuan, SONG Qiaochu, WU Wei, et al. Variation law of groundwater depth and mineralization of coastal tidal flats in Jiangsu Province[J]. Advances in Science and Technology of Water Resources,2021,41(6):52-57. (in Chinese))
- [6] 凌敏华,陈万贺. 地下水资源对新疆经济社会支撑作用的定量评价[J]. 水资源保护,2021,37(2):49-54. (LING Minhua, CHEN Wanhe. Quantitative evaluation of groundwater resources supporting economy and society in Xinjiang[J]. Water Resources Protection,2021,37(2):49-54. (in Chinese))
- [7] 郭田丽. 区域地下水埋深的数据驱动预报模型研究[D]. 杨凌:西北农林科技大学,2020.
- [8] 刘鑫,韩宇平,刘中培,等. 基于 BiLSTM-NFC 的地下水埋深模拟方法研究[J]. 人民黄河,2021,43(6):80-85. (LIU Xin, HAN Yuping, LIU Zhongpei, et al. Research on groundwater depth prediction method based on BiLSTM-NFC[J]. Yellow River, 2021,43(6):80-85. (in Chinese))
- [9] 赵振,陈惠娟,秦光雄,等. 青海乌兰盆地地下水资源评价及可开采潜力分析[J]. 水资源保护,2021,37(6):60-66. (ZHAO Zhen, CHEN Huijuan, QIN Guangxiong, et al. Groundwater resources evaluation and exploitable potential analysis in Wulan Basin, Qinghai Province[J]. Water Resources Protection,2021,37(6):60-66. (in Chinese))
- [10] NIKOLOS I K, STERGIADI M, PAPADOPOULOU M P, et al. Artificial neural networks as an alternative approach to groundwater numerical modelling and environmental design[J]. Hydrological Processes,2008,22(17):3337-3348.
- [11] YOON H, JUN S, HYUN Y, et al. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer[J]. Journal of Hydrology,2011,396(1):128-138.
- [12] BILALI A E, TALEB A, BROUZYINE Y. Comparing four machine learning model performances in forecasting the alluvial aquifer level in a semi-arid region[J]. Journal of African Earth Sciences,2021,181:104244.
- [13] 马新双,卢胜勇,李明良,等. 北方浅层地下水超采区地下水模拟模型[J]. 南水北调与水利科技,2011,9(4):134-136. (MA Xinshuang, LU Shengyong, LI Mingliang, et al. Research of groundwater prediction model in shallow groundwater over-exploitation zone in the Northern Area[J]. South-to-North Water Transfers and Water Science & Technology,2011,9(4):134-136. (in Chinese))
- [14] BOWES B D, SADLER J M, MORSY M M, et al. Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks[J]. Water,2019,11(5):1098.
- [15] 闫佰忠,孙剑,王昕洲,等. 基于多变量 LSTM 神经网络的地下水水位模拟[J]. 吉林大学学报(地球科学版),2020,50(1):208-216. (YAN Baizhong, SUN Jian, WANG Xinzhou, et al. Multivariable LSTM neural network model for groundwater levels prediction[J]. Journal of Jilin University(Earth Science Edition),2020,50(1):208-216. (in Chinese))
- [16] 汪云,杨海博,徐建,等. 基于长短期记忆神经网络模型的地下水水位模拟研究[J]. 节水灌溉,2019(10):73-77. (WANG Yun, YANG Haibo, XU Jian, et al. Groundwater level prediction based on long-short-term memory neural network model[J]. Water Saving Irrigation,2019(10):73-77. (in Chinese))
- [17] 张钧泳,丁建丽,谭娇. 基于 SVM 的绿洲荒漠交错带土壤水分与地下水埋深反演[J]. 农业机械学报,2019,50(3):221-230. (ZHANG Junyong, DING Jianli, TAN Jiao. Inversion of soil moisture and shallow groundwater depth based on SVM in arid oasis-desert ecotone[J]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(3):221-230. (in Chinese))
- [18] 张文鸽,黄强,佟春生. 基于混沌优化的支持向量机地下水位动态模拟[J]. 资源科学,2007(5):105-109. (ZHANG Wenge, HUANG Qiang, TONG Chunsheng. Dynamic prediction of groundwater level based on chaos optimization and support vector machine[J]. Resources Science,2007(5):105-109. (in Chinese))
- [19] 黄发明,殷坤龙,张桂荣,等. 多变量 PSO-SVM 模型模拟滑坡地下水位[J]. 浙江大学学报(工学版),2015,49(6):1193-1200. (HUANG Faming, YIN Kunlong, ZHANG Guirong, et al. Prediction of groundwater level in landslide using multivariable PSO-SVM model[J]. Journal of Zhejiang University (Engineering Science),2015,49(6):1193-1200. (in Chinese))
- [20] WU C, ZHANG X, WANG W, et al. Groundwater level modeling framework by combining the wavelet transform with a long short-term memory data-driven model[J]. Science of the Total Environment,2021,783:146948.
- [21] 王丽萍,李宁宁,马皓宇,等. MIC-PCA 耦合算法在径流预报因子筛选中的应用[J]. 中国农村水利水电,2018(9):36-41. (WANG Liping, LI Ningning, MA Haoyu, et al. Research on the application of MIC-PCA algorithm in screening runoff forecast factors[J]. China Rural Water and Hydropower,2018(9):36-41. (in Chinese))
- [22] 梁浩,黄生志,孟二浩,等. 基于多种混合模型的径流模拟研究[J]. 水利学报,2020,51(1):112-125. (LIANG Hao, HUANG Shengzhi, MENG Erhao, et al. Runoff prediction based on multiple hybrid models[J]. Journal of Hydraulic Engineering,2020,51(1):112-125. (in Chinese))
- [23] 吴娟,朱跃龙,金松,等. 三种机器学习模型在太湖藻华面积预测中的应用[J]. 河海大学学报(自然科学版),2020,48(6):542-551. (WU Juan, ZHU Yuelong, JIN Song, et al. Area prediction of cyanobacterial blooms based on three machine learning methods in Taihu Lake[J]. Journal of Hohai University (Natural Sciences),2020,48(6):542-551. (in Chinese))

[24] 刘堃,何祺胜,荆琛琳,等. 基于机器学习的蒸散量插补方法[J]. 河海大学学报(自然科学版),2020,48(2):109-115. (LIU Kun,HE Qisheng,JING Chenlin,et al. Gap filling method for evapotranspiration based on machine learning[J]. Journal of Hohai University (Natural Sciences),2020,48(2):109-115. (in Chinese))

[25] 张楠,夏自强,江红. 基于多因子量化指标的支持向量机径流模拟[J]. 水利学报,2010,41(11):1318-1324. (ZHANG Nan,XIA Ziqiang,JIANG Hong. Prediction of runoff based on the multiple quantity index of SVM[J]. Journal of Hydraulic Engineering,2010,41(11):1318-1324. (in Chinese))

[26] 刘宝成. 基于最小二乘支持向量机的时空序列混合模拟模型及实证研究[D]. 兰州:兰州交通大学,2018.

[27] HU C,WU Q,LI H,et al. Deep learning with a long short-term memory networks approach for rainfall-runoff simulation[J]. Water,2018,10(11):1543.

[28] MA X,TAO Z,WANG Y,et al. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data[J]. Transportation Research Part C:Emerging Technologies,2015,54:187-197.

[29] LIANG C,LI H,LEI M,et al. Dongting Lake water level forecast and its relationship with the Three Gorges Dam based on a long short-term memory network[J]. Water,2018,10(10):1389.

[30] 于国强,张茂省,王根龙,等. 支持向量机和BP神经网络在泥石流平均流速模拟模型中的比较与应用[J]. 水利学报,2012,43(增刊2):105-110. (YU Guoqiang,ZHANG Maosheng,WANG Genlong,et al. Application and comparison of prediction models of support vector machines and back-propagation artificial neural network for debris flow average velocity[J]. Journal of Hydraulic Engineering,2012,43(Sup2):105-110. (in Chinese))

[31] 潘志刚,林祥峰,张继生. 基于支持向量机的桥墩局部冲刷深度预测模型[J]. 水利水电科技进展,2021,41(6):78-81. (PAN Zhigang,LIN Xiangfeng,ZHANG Jisheng. Prediction model for bridge pier local scour depth based on support vector machine[J]. Advances in Science and Technology of Water Resources,2021,41(6):78-81. (in Chinese))

[32] 贺玉琪,王栋,王远坤. BRR-SVR 月降水量模拟优化模型[J]. 水利学报,2019,50(12):1529-1537. (HE Yuqi,WANG Dong,WANG Yuankun. BRR-SVR optimization model for monthly precipitation prediction[J]. Journal of Hydraulic Engineering,2019,50(12):1529-1537. (in Chinese))

[33] 杜聪,邵建华,杨薇,等. 网格搜索法优化的支持向量机室内可见光定位[J]. 激光杂志,2021,42(3):104-109. (DU Cong,SHAO Jianhua,YANG Wei,et al. Support vector machine indoor visible light positioning optimized by grid search method[J]. Laser Journal,2021,42(3):104-109. (in Chinese))

[34] TIAN H,WANG P,TANSEY K,et al. An LSTM neural network for improving wheat yield estimates by integrating remote sensing data and meteorological data in the Guanzhong Plain,P R China[J]. Agricultural and Forest Meteorology,2021,310:108629.

[35] MAIMAITIJANG M,SAGAN V,SIDIKE P,et al. Soybean yield prediction from UAV using multimodal data fusion and deep learning[J]. Remote Sensing of Environment,2020,237:111599. (收稿日期:2021-09-28 编辑:胡新宇)

(上接第49页)

[14] 刘永志,唐雯雯,张文婷,等. 基于灾害链的洪涝灾害风险分析综述[J]. 水资源保护,2021,37(1):20-27. (LIU Yongzhi,TANG Wenwen,ZHANG Wenting,et al. Review of flood disaster risk analysis based on disaster chain[J]. Water Resources Protection,2021,37(1):20-27. (in Chinese))

[15] 黄国如,罗海婉,卢鑫祥,等. 城市洪涝灾害风险分析与区划方法综述[J]. 水资源保护,2020,36(6):1-6. (HUANG Guoru,LUO Haiwan,LU Xinxang,et al. Study on risk analysis and zoning method of urban flood disaster[J]. Water Resources Protection,2020,36(6):1-6. (in Chinese))

[16] 徐宗学,任梅芳,陈浩. 我国沿海城市洪潮组合风险分析[J]. 水资源保护,2021,37(2):10-14. (XU Zongxue,REN Meifang,CHEN Hao. Analysis on urban flooding risk caused by flood tide combination in coastal cities[J]. Water Resources Protection,2021,37(2):10-14. (in Chinese))

[17] 杜钧,GRUMM R H,邓国. 预报异常极端高影响天气的“集合异常预报法”:以北京2012年7月21日特大暴雨为例[J]. 大气科学,2014,38(4):685-699. (DU Jun,GRUMM R H,DENG Guo. Ensemble anomaly forecasting approach to predicting extreme weather demonstrated by extremely heavy rain event in Beijing[J]. Scientia Atmospherica Sinica,2014,38(4):685-699. (in Chinese))

[18] 曹勇,包红军,张恒德,等. 基于快速滚动更新的无缝隙定量降水预报模型[J]. 河海大学学报(自然科学版),2021,49(4):303-308. (CAO Yong,BAO Hongjun,ZHANG Hengde,et al. Seamless quantitative precipitation forecasting model based on rapid rolling update technique[J]. Journal of Hohai University (Natural Sciences),2021,49(4):303-308(in Chinese))

[19] 陶诗言. 中国之暴雨[M]. 北京:科学出版社,1980:14-117.

[20] 杨忠林,赵坤,徐坤,等. 江淮梅雨期极端对流微物理特征的双偏振雷达观测研究[J]. 气象学报,2019,77(1):58-72. (YANG Zhonglin,ZHAO Kun,XU Kun,et al. Microphysical characteristics of extreme convective precipitation over the Yangtze-Huaihe River Basin during the Meiyu season based on polarimetric radar data[J]. Acta Meteorologica Sinica,2019,77(1):58-72. (in Chinese)) (收稿日期:2021-09-28 编辑:胡新宇)