

基于长短期记忆网络的日供水量预测方法研究

苟非洲, 程玉婷

(陕西省水务集团有限公司, 陕西 西安 710004)

摘要: 供水量预测方法分为传统预测法和机器学习法两类。传统方法对于日供水量预测精度有限,机器学习法具有更好的模型刻画能力,可发现传统算法难以捕捉的数据细节。提出了一种基于长短期记忆网络(LSTM)的日供水量预测方法,利用渭南市蒲城县城2016年1月—2018年3月的日供水量数据对该方法进行了验证,并与基于BP神经网络的方法进行了对比。基于LSTM网络方法的预测结果:平均绝对百分比误差为5.7%, R^2 为0.804;基于BP神经网络的预测结果:平均绝对百分比误差为7.3%, R^2 为0.610。可见,基于长短期记忆网络的预测方法具有较高的预测精度和稳定性,是一种有效的预测方法。

关键词: 日供水量预测; 长短期记忆网络; BP神经网络

中图分类号: TU991 **文献标识码:** A **文章编号:** 1000-4602(2019)17-0079-05

Daily Water Supply Forecasting Method Based on Long Short-term Memory Network

GOU Fei-zhou, CHENG Yu-ting

(Shaanxi Water Affair Group Co. Ltd., Xi'an 710004, China)

Abstract: Water supply forecasting method has two categories: traditional forecasting method and machine learning method. The traditional method has limited prediction accuracy for daily water supply, and the machine learning method has better model characterization ability, in addition to the ability to find data details that are difficult to capture by traditional method. A prediction method based on the long short-term memory (LSTM) network model was proposed, and verified using the daily water supply data from January 2016 to March 2018 in Pucheng County, Weinan City. The results were compared with the method of BP neural network. LSTM showed MAPE of 5.7% and R^2 of 0.804, while BP showed MAPE of 7.3% and R^2 of 0.610. The results demonstrated that the prediction method based on the LSTM network model had higher prediction accuracy and stability, and it was an effective prediction method.

Key words: daily water supply forecasting; long short-term memory network; BP neural network

供水量预测既可为水利市政设施的投资决策提供依据,又可为供水信息化管理提供指导,提高水资源利用效率。供水量预测分为短期预测、中期预测、长期预测3种,日供水量预测属于短期预测。日供

水量预测问题是一种时间序列预测问题,即根据时间轴上过去的历史数据预测之后可能产生的结果。供水量预测方法可以分为两类:传统预测方法和基于新技术的预测方法。传统预测方法具有结构简单、计

算负荷低,便于在实际生产过程中应用的优点,但也存在预测精度低、模型泛化能力差等问题^[1],主要包括自回归积分滑动平均模型(ARIMA)^[2]、灰色模型^[3]等。另一大类是机器学习算法,包括支持向量机模型(SVR)^[4]、BP神经网络^[5]、随机森林模型(RF)^[6]、深度信念网络(DBN)^[7]等。机器学习算法具有更好的模型刻画能力,可发现传统算法难以捕捉的数据细节。

利用递归神经网络(RNN)可处理时间序列预测问题,长短期记忆(LSTM)网络^[8]是一种特殊的RNN网络。目前利用长短期记忆网络对日供水量进行预测的研究还较少。为了更好地对日供水量进行预测,笔者尝试采用LSTM网络模型开展区域日供水量预测的相关研究,并利用渭南市蒲城县城区2016年1月—2018年3月的日供水量数据对模型预测结果进行验证。

1 研究方法

1.1 供水量相关因素

区域供水量影响因素多种多样,长期影响因素有城市经济发展水平、人口规模、工业生产能力、气候等,短期影响因素有预测前一段时间的供水量、天气状况、是否节假日等^[1]。在实际生产中,大部分供水企业并未采集与需水量相关的气象、降水等资料,历史供水量数据是供水企业能够收集到的最可靠的数据。有研究认为,把需水量作为单一的模型输入可以增加模型的稳定性和可靠性^[9]。

1.2 基于LSTM的预测方法

RNN单元示意图1。

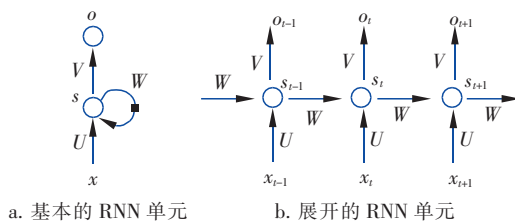


图1 RNN单元示意

Fig. 1 Schematic diagram of RNN unit

基本的RNN单元如图1(a)所示,由输入层、1个隐层和1个输出层组成。“循环”的特性体现在隐层。图1(b)为“循环”展开的RNN单元,可以看出隐层的神经元之间也是带有权值的。随着序列的不断推进,前面的隐层将会影响后面的隐层。RNN进行训练时,由于其网络数学特性,会产生梯度消

失^[10]问题,导致参数在训练过程中无法更新。LSTM网络是一种特殊的RNN网络,改善了基本RNN的很多特性,可以很好地处理具有长期依赖特性的问题^[11]。

典型的LSTM网络结构如图2所示。

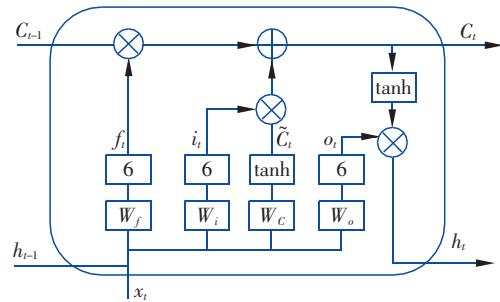


图2 典型的LSTM网络示意

Fig. 2 Schematic diagram of typical LSTM network

标准RNN隐层的单元只有一个tanh函数,LSTM网络由输入门、遗忘门和输出门组成隐层单元。具体数学表达式如下:

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

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

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

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

本研究拟采用双层LSTM网络解决日供水量预测问题,如图3所示。

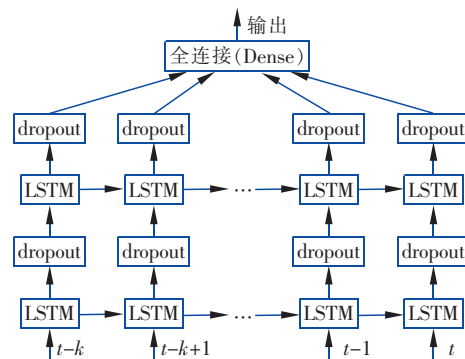


图3 双层LSTM网络示意

Fig. 3 Schematic diagram of double-deck LSTM network

LSTM网络带有循环结构,可通过前向传播算法实现误差计算,利用误差反向传播算法实现对参数的更新。LSTM网络可以处理多输入多输出问题,多层LSTM网络在解决更复杂的问题时可提供更高的预测准确性。

利用 dropout 避免过拟合, dropout 可使一个神经单元和随机挑选出来的其他神经单元共同工作,阻止特征的协同作用,从而增强模型的泛化能力,提高神经网络性能^[12]。

1.3 预测数据评价指标

对于模型预测的结果通常采用一些数据指标进行评价,本研究采用4个预测结果评价指标,分别是均方根误差(RMSE)、平均绝对误差(MAE)、 R^2 、平均绝对百分比误差(MAPE)。 R^2 值越接近1说明预测值越接近真实值,模型表现越好。评价指标的计算见式(7)~(10)。

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (7)$$

$$MAE = \frac{1}{m} \sum_{i=1}^m |(y_i - \hat{y}_i)| \quad (8)$$

$$R = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (9)$$

$$MAPE = \frac{100}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (10)$$

$$SSE = \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (11)$$

$$SST = \sum_{i=1}^m (y_i - \bar{y})^2 \quad (12)$$

$$\bar{y} = \frac{1}{m} \sum_{i=1}^m y_i \quad (13)$$

其中, y_i 为真实测量值; \hat{y}_i 为预测值。

2 案例分析

2.1 数据预处理

对于供水量预测这样的时间序列预测问题,采用机器学习框架进行求解时,首先需要将该问题转化为一个包含输入输出序列的监督学习问题,即将采集的供水量数据形成训练所需的数据集。

一个监督学习问题包含输入参数 X 和输出参数 Y 。模型经过训练就可以完成由输入参数 X 预测输出参数 Y 的工作。对于供水量预测的问题,考虑到可以建立一个单输入单输出的系统,将供水量数据 $d(t)$ 作为输入, $X = d(t)$;供水量输出向后平移一个时间周期, $d(t+1)$ 作为输出, $Y = d(t+1)$ 。即根据第 t 天及之前的供水量数据预测第 $t+1$ 天的供水量数据。

2.2 机器学习训练系统搭建

本研究将 Python 作为开发编程语言,利用 TensorFlow 和 Karas 建立机器学习运行环境。采用陕西省渭南市蒲城城区 2016 年 1 月—2018 年 3

月的日供水量统计数据作为数据集。该数据集共包括 803 个点,其中 60% 的数据(482 个点)作为训练数据集、20% 的数据(160 个点)作为验证集、20% 的数据(161 个点)作为测试集。日供水量数据如图 4 所示。

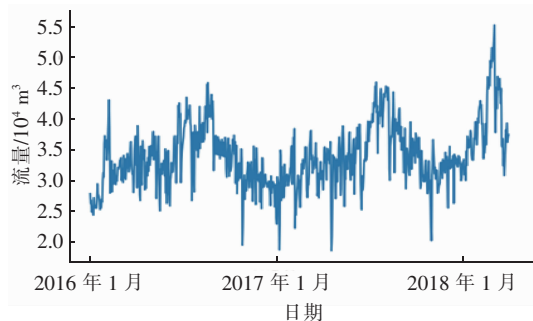


图4 日供水量数据

Fig. 4 Daily water supply data

利用 Karas 建立双层 LSTM 网络,网络单层神经元数为 100。LSTM 网络进行训练的参数 epochs = 60、batch_size = 20。训练过程误差趋势如图 5 所示。可以看出,经过训练,损失函数已收敛,不存在过拟合的现象。

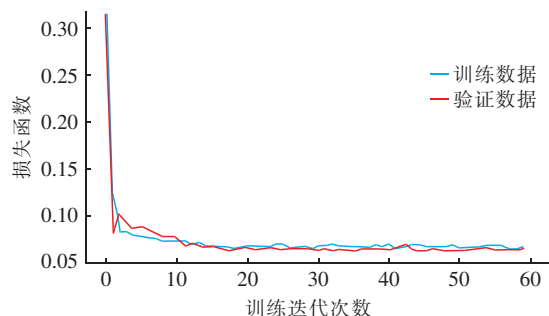


图5 训练过程误差趋势

Fig. 5 Error trend in training process

2.3 预测结果

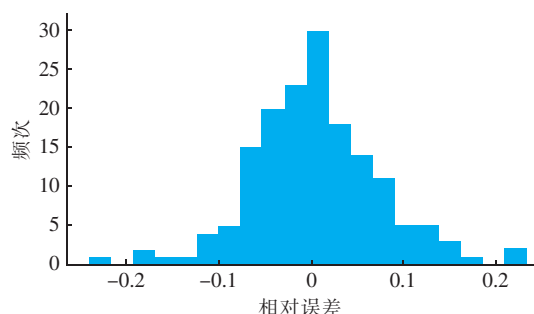
本研究利用训练得到的 LSTM 网络模型对测试集数据进行预测,并与基于 BP 神经网络的预测方法进行对比(基于 BP 神经网络训练数据、测试数据的划分与 LSTM 网络的一致),对比结果见图 6。可以看出,基于 LSTM 网络模型的预测值与实测值拟合度较好,而基于 BP 神经网络的预测曲线和实测曲线有一定差距,尤其在用水量波动较大的时刻。LSTM 模型测试数据集的 RMSE 为 2 709.2、MAPE 为 5.7%、 R^2 为 0.804;BP 神经网络预测结果的 RMSE 为 3 818.4、MAPE 为 7.3%、 R^2 为 0.610。LSTM 网络模型预测结果优于 BP 模型预测结果。



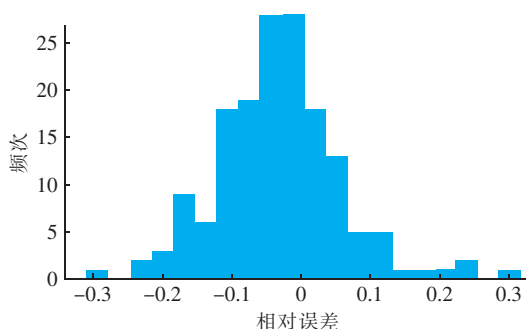
图6 预测数据和实测数据对比

Fig. 6 Comparison between predicted data and measured data

为了更加全面地评估模型预测结果,分别绘制了基于 LSTM 模型和 BP 神经网络模型预测结果与真实数据的相对误差统计直方图,见图 7。可以看出,相对误差均服从正态分布,其中,LSTM 模型相对误差 $< 10\%$ 的数据占总数的 85.1% 、相对误差 $< 5\%$ 的数据占 57.1% ;BP 模型相对误差 $< 10\%$ 的数据占总数的 75.2% 、相对误差 $< 5\%$ 的数据占 47.2% 。这表明,LSTM 模型预测结果误差分布更加集中,基于 LSTM 的预测方法具有一定的优势。



a. LSTM 模型相对误差统计直方图



b. BP 模型相对误差统计直方图

图7 LSTM 模型和 BP 模型的相对误差统计直方图

Fig. 7 Relative error statistics histogram of LSTM model and BP model

在蒲城县城日供水量预测案例中,LSTM 模型表现出了良好的预测精度和稳定性,预测效果优

于传统预测方法。一方面是因为 LSTM 这种反馈性神经网络具有记忆功能,适合处理复杂时间相关性的时间序列学习问题;另一方面是因为 LSTM 模型具有多层结构,可对数据进行复杂变换和映射,有较强的处理非线性数据的能力,能发现数据之间的隐含关系^[1]。

在模型预测过程中发现,数据量的大小对预测结果有明显影响。LSTM 神经网络内部包含海量参数,长时间序列数据是模型预测取得良好结果的重要因素。本研究在 LSTM 网络模型训练中曾加入工作日/周末、天气因素,但未取得明显的模型优化效果,日供水量与其他因素的关联性还需要进一步研究。随着物联网技术和供水网络智能化水平的提高,智能化供水设备将为供水量变化影响因素分析提供更多的手段。

3 结论

搭建了基于 LSTM 的神经网络模型,并用于陕西省渭南市蒲城县城日供水量的预测。预测结果的平均绝对百分比误差为 5.7% 、 R^2 为 0.804 ,误差分布较为集中,预测结果优于 BP 模型。基于长短期记忆网络的神经网络模型预测精度高,预测结果稳定,是预测城市日供水量的有效方法。

参考文献:

- [1] 郭冠呈,刘书明,李俊禹,等. 基于双向长短时神经网络的水量预测方法研究[J]. 给水排水,2018,44(3): 123-126.
Guo Guancheng, Liu Shuming, Li Junyu, et al. Study on water quantity prediction method based on bidirectional long and short time neural network [J]. Water & Wastewater Engineering, 2018, 44 (3): 123 - 126 (in Chinese).
- [2] 田涛,薛惠锋,张峰. 基于 ARIMA 与 GM(1,1) 的区域用水总量预测模型及应用——以广州市为例[J]. 节水灌溉,2018(2): 61-65.
Tian Tao, Xue Huifeng, Zhang Feng. Forecasting models of regional water consumption and their application based on ARIMA and GM(1,1): A case study of Guangzhou City [J]. Water Saving Irrigation, 2018 (2): 61 - 65 (in Chinese).
- [3] 孙强,王秋萍. 融合粗糙集和灰色 GM(1,N) 的西安市供水量预测[J]. 计算机工程与应用,2013,49(11): 237-240.
Sun Qiang, Wang Qiuping. Water supply quantity

- forecast of Xi'an via combination rough sets with GM(1,N) model [J]. Computer Engineering and Applications, 2013, 49(11): 237-240 (in Chinese).
- [4] 牟天蔚, 沈丹玉, 王玲萍. GA-SVR 模型用于日供水量预测的研究[J]. 给水排水工程, 2018, 36(1): 134-136.
- Mu Tianwei, Shen Danyu, Wang Lingping. Research of daily water supply forecast by GA-SVR model [J]. Water Supply & Drainage Engineering, 2018, 36(1): 134-136 (in Chinese).
- [5] 王圃, 唐鹏飞, 白云, 等. 基于多分辨 BP 神经网络的城市日供水量预测模型[J]. 中国给水排水, 2018, 34(11): 51-55.
- Wang Pu, Tang Pengfei, Bai Yun, et al. Forecasting model of daily urban water supply based on multi-resolution BP neural network [J]. China Water & Wastewater, 2018, 34(11): 51-55 (in Chinese).
- [6] 王盼, 陆宝宏, 张瀚文, 等. 基于随机森林模型的需水预测模型及其应用[J]. 水资源保护, 2014, 30(1): 34-37.
- Wang Pan, Lu Baohong, Zhang Hanwen, et al. Water demand prediction model based on random forests model and its application [J]. Water Resources Protection, 2014, 30(1): 34-37 (in Chinese).
- [7] 牟天蔚, 蒋白懿, 沈丹玉, 等. 深度学习框架对城市日供水量预测的研究[J]. 人民黄河, 2018, 40(9): 58-62.
- Mu Tianwei, Jiang Baiyi, Shen Danyu, et al. Study on the prediction of deep learning framework for city daily water supply [J]. Yellow River, 2018, 40(9): 58-62 (in Chinese).
- [8] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.
- [9] Bakker M, Vreeburg J H G, van Schagen K M, et al. A fully adaptive forecasting model for short-term drinking water demand [J]. Environmental Modelling & Software, 2013, 48: 141-151.
- [10] Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions [J]. International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 1998, 6(2): 107-116.
- [11] Graves A. Supervised Sequence Labelling with Recurrent Neural Networks [M]. Berlin: Springer, 2012.
- [12] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: A simple way to prevent neural networks from overfitting [J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.



作者简介: 苟非洲(1975-), 男, 陕西礼泉人, 工学学士, 高级工程师, 主要从事水利市政项目投资运营管理工作。

E-mail: 1464405734@qq.com

收稿日期: 2019-01-20

珍惜资源, 保护环境, 建设美丽中国