

水力发电
Water Power
ISSN 0559-9342, CN 11-1845/TV

《水力发电》网络首发论文

题目: 大坝监测数据的时序预测与补全
作者: 杜曼玲, 高嘉欣, 张礼兵, 罗明清, 陈云天, 胡文波, 田天
收稿日期: 2020-05-12
网络首发日期: 2020-09-28
引用格式: 杜曼玲, 高嘉欣, 张礼兵, 罗明清, 陈云天, 胡文波, 田天. 大坝监测数据的时序预测与补全. 水力发电.
<https://kns.cnki.net/kcms/detail/11.1845.TV.20200928.1554.002.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

大坝监测数据的时序预测与补全

杜曼玲¹, 高嘉欣², 张礼兵³, 罗明清⁴, 陈云天^{2,5}, 胡文波^{2,6}, 田 天^{2,6}

- (1. 中国电建集团国际工程有限公司, 北京 100142;
2. 北京瑞莱智慧科技有限公司, 北京 100084;
3. 中国电建集团昆明勘测设计研究院有限公司, 云南 昆明 650051;
4. 中国电建集团海外投资有限公司, 北京 100142;
5. 鹏城实验室智慧能源工作室, 广东 深圳 518055;
6. 清华大学人工智能研究院, 北京 100084)

摘 要: 大坝安全数据的时间序列预测和补全是大坝安全监测中两个常见问题。在实际应用中, 应对这两类问题已有相应的经验模型, 但由于经验模型过于简单, 上述两类问题上的效果并不理想。采用机器学习中的决策树模型和神经网络模型, 为上述两类问题提供全新的解决方案。基于某水电大坝实测数据, 对不同坝段的大坝安全数据进行预测和补全, 结果表明本文模型相比于传统经验模型预测与补全精度均显著提高, 同时, 该模型还可以对模型输出结果进行不确定性分析, 增强了结果的可靠性。

关键词: 大坝; 监测数据; 时间序列预测和补全; 机器学习

Time Series Forecasting and Imputation of Dam Physical Quantities

DU Manling¹, GAO Jiaxin², ZHANG Libing³, LUO Mingqing⁴, CHEN Yuntian^{2,5}, HU Wenbo^{2,6}, TIAN Tian^{2,6}

- (1. PowerChina International Group Limited, Beijing 100142, China;
2. RealAI, Beijing 100084, China;
3. PowerChina Kunming Engineering Corporation Limited, Kunming 650051, Yunnan, China;
4. PowerChina Resources Limited, Beijing 100142, China;
5. Intelligent Energy Lab, Peng Cheng Laboratory, Shenzhen 518055, Guangdong, China;
6. Institute for AI, Tsinghua University, Beijing 100084, China)

Abstract: The time series forecasting and imputation of dam physical quantities are two central tasks in dam safety monitoring engineering. In practical scenarios, there are empirical models for the two tasks, but the empirical models perform weak in that they are too simple. The solutions to these two problems are provides via decision trees and neural network models, which have achieved good results. Based on the actual dam monitoring data of a hydropower station, the machine learning and neural network models are used to forecast and impute the physical quantity series of different dam sections. The experiments show that the proposed model is significantly better than the empirical models and can provide uncertainties quantification for the predictions which improves the reliability of the model.

Key Words: dam; monitoring data; time series forecasting and imputation; machine learning

中图分类号: TV736 文献标志码: A

0 引 言

目前我国拥有水库大坝 9.8 万余座, 是世界上水库大坝最多的国家, 水坝在统筹防洪、发电、供水、灌溉等方面发挥了重要作用, 是组成国民经济的战略性基础设施之一^[1]。目前水坝面临着数量多、时间长、设备老旧、气候变化、建设条件复杂等多重因素带来的风险, 其安全问题日益突出。如何准确、全面地进行大坝安全监测和预报, 对大坝安全运行和辅助决策具有重要的意义^[2]。

大坝工程正在步入数字大坝和智慧大坝阶段, 尽管现代网络技术使得工程师能够在大坝全生命周期内展开实时、在线、全天候的管理与分析^[3-4], 但由于早期大坝建设中存在的信息化建设不系统、不全面、不统一以及老旧水坝的设施落后甚至缺乏现代监测设备等问题, 使得大坝安全监测普遍面临数据缺失和数据碎片化的挑战^[5]。基于传统数学模型的数值拟合方法难以对残缺的大坝安全数据进行有效的补全与预测。

随着人工智能领域的快速发展, 很多研究者开始采用这一方法解决大坝安全监测和预测问题。赵斌等人应用人工神经网络进行大坝安全数据的预报^[6], 樊琨基于人工神经网络方法建立非线性力学反分析模型解决岩土工程中的复杂非线性问题^[7], 蒋利娟基于线性回归模型利用降水量预测水位数据^[8], 姜成科提出的 GA-LMBP 算法提高大坝安全监测人工神经网络模型的拟合效果和预测精度^[9]。本文采用广泛应用于深度学习的高斯过程回归 (Gaussian Process Regression, GPR)^[10]、LightGBM (Light Gradient Boosting

收稿日期: 2020-05-12

基金项目: 中国电力建设股份有限公司科技项目经费资助(DJ-ZDXM-2018-18)

作者简介: 杜曼玲 (1969—), 女, 湖北浠水人, 高级工程师, 博士后, 研究方向为控制科学与工程; 陈云天 (通讯作者)。

Machine)^[11]、LSTM^[12]等模型进行大坝监测参数的预测与补全,并与传统方法做对比^[13]。

机器学习方法所表现出的自组织性、自适应性、模糊推理能力和自学习能力等优势非常适合解决大坝安全数据补全与预测这一复杂的非线性问题^[14]。本文提出的大坝安全数据的时序预测与补全模型是针对采集的海量数据进行深度、有效分析的前提。基于本文提出的方法所获得的高质量大坝安全数据有利于构建智慧大坝安全评估体系,是智能监控、智能诊断、智能决策的基础,有助于切实提升大坝安全智能管理能力。同时,将专家知识与海量的大坝安全数据通过人工智能技术相结合,是对大坝不同维度物理量之间的深度融合,是实现可感知、可分析、可控制的智能化大坝建设有效途径^[15-18]。

1 研究方法

1.1 基于高斯过程回归的数据插补预处理

高斯过程回归(Gaussian Process Regression, GPR)模型是机器学习领域的一个经典模型。GPR 以其良好的泛用性和可解释性,它在时间序列分析、自动化控制、图像处理等诸多领域都有广泛应用^[19-20]。

由于传感器故障或人工操作的失误,水坝数据中往往存在个别缺失值,此外,由于工况和环境的复杂性,原始数据中还会存在一些随机噪声或粗差。这些不利因素都会影响大坝安全数据补全以及预测工作的正常开展。因此,大坝安全数据补全以及预测的第一步是对原始数据做插值和平滑处理等预处理。

1.2 采用的机器学习算法

1.2.1 长短期记忆神经网络

神经网络能够通过学习来近似拟合输入和输出变量之间的非线性函数关系。它的基本运算单位是神经元,神经元对输出的影响由神经元之前的权重来表现,这个权重会随着网络的训练不断调整。在网络中,每个神经元的输出都通过非线性函数计算得到,而非线性函数的输入是其他神经元输出的代数和。

常规神经网络由于只能构建单一的映射关系,对于时间序列问题的预测效果并不理想^[21]。长短期记忆神经网络(Long Short-term Memory, LSTM)不仅能够利用当前的特征信息,还能够利用先前计算产生的中间结果,实现了无效信息的遗忘和有效信息的加强。因此, LSTM 是解决如位移、渗流等大坝安全数据序列问题的最自然且最合适的理想工具。

1.2.2 LightGBM 模型

决策树是一种常见的机器学习模型,它代表的是对象属性与对象值之间的一种映射关系^[22]。决策树,本质上是针对样本的重要特征不断做出判断,根据每步的判断结果寻找合适的路径,最终得到合适的预测结果。决策树模型具有多种实现方法。LightGBM(Light Gradient Boosting Machine)^[11]模型是众多方法中较为高效的之一,与其他方法相比,LightGBM 具有更高的训练效率、更低的内存使用、支持并行化学习等,这些优秀的特征使得它在实际生产中有比较广泛的应用。

1.3 时间序列预测与补全任务

1.3.1 经验模型

在经验模型中,通常都假设某一时刻的安全监测结果(如位移、渗流等)主要受水压、温度等环境量因素以及时效等因素影响,因此,安全监测结果由水压分量、温度分量和时效分量组成,即 $\delta = \delta_H + \delta_T + \delta_q$ 。

其中 δ 表示位移; δ_H 表示水压分量; δ_T 表示温度分量; δ_q 表示时效分量。从本质上来讲,经验模型就是一个以环境量和时效量为特征的多项式回归模型。

1.3.2 时间序列预测与补全

在大坝安全数据时间序列预测这个问题上,本文尝试了不同的时间序列预测方法,包括前文提到的传统方法、经典的时间序列模型自回归滑动平均模型(Autoregressive moving average model, ARMA)^[23],还有 FCNN(Fully Connected Neuron Network)、LSTM 和 LightGBM 这三个机器学习模型。在三个机器学习模型中,本文用到的特征包括历史的环境量数据(温度,上下游水位,降水)和待预测安全数据的历史数据,应用历史数据的长度和具体问题有关,一般来说应用两个月以内的历史数据即可。

当大坝安全数据由于某些原因(比如传感器长时间故障等)出现大范围缺失的情况下,数据本身的规律变得难以挖掘,数据插补和时间序列预测的模型变得不再适用。为了针对大坝安全展开进一步的研究,必须对这些缺失数据有效补全。对于同一坝段,不同位置的相同类型数据可能有着相似的变化规律,我们可以根据坝段其他位置的信息,去补全坝段的当前位置数据。具体地,本文针对某个缺失数据较多的坝段,找出多个与目标坝段相近且数据较全的坝段,对它们的传感器数据进行插补、平滑等预处理操作。然后,用这些临近坝段的相对完整的数据训练模型,以便模型可以建立起不同点位数据之间的映射关系。最后,将这一模型应用于目标坝段,基于目标坝段中完整的点位数据,补全目标坝段中残缺点位的数据。其中,这个模型可以是简单的线性模型,也可以是本文前面提到的神经网络模型和 LightGBM 模型。

2 结果分析

2.1 高斯过程回归数据预处理结果

在数据预处理阶段,主要应用高斯过程回归(GPR)模型对数据进行平滑和插补,最终数据处理的效果如图 1a(处理前的原始数据)和图 1b(处理后的数据)所示。平滑操作可以有效保留数据的趋势特征,并消除随机噪声与粗差的不利影响,处理后曲线更反映大坝安全数据的变化趋势。

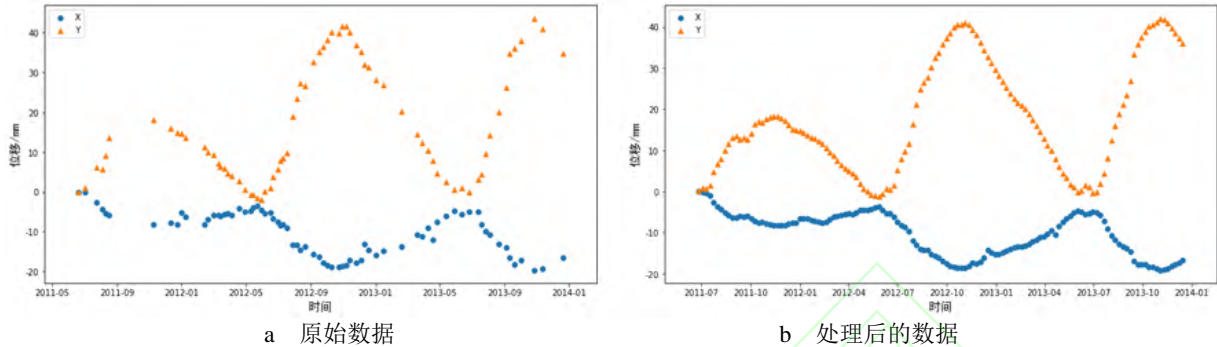


图 1 高斯过程回归数据预处理结果

在数据平滑的过程中,基于高斯过程回归处理的结果的置信度分布如图 2 所示。

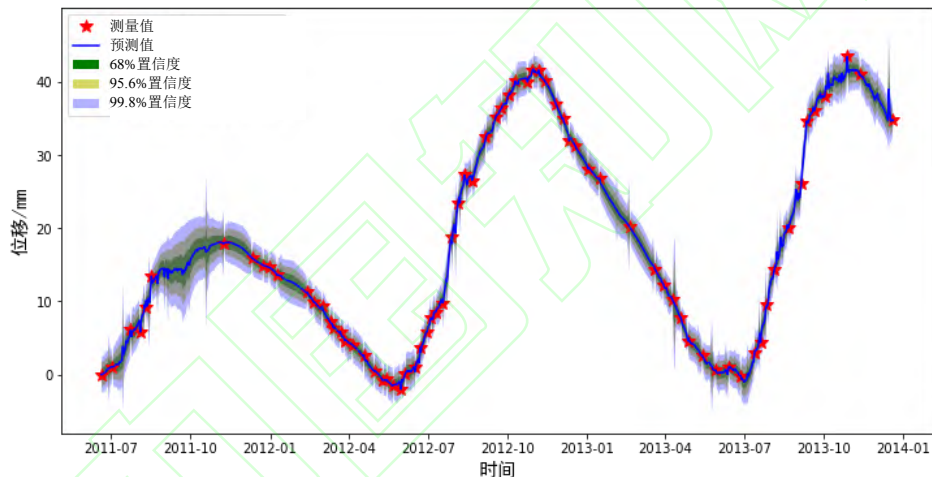


图 2 平滑数据的置信区间

当数据变化趋势较为稳定,预测结果的不确定性就比较小(阴影宽度降低),而当数据的趋势发生变化时,数据的不确定性会有所增加。本文模型给出的结果和专家的主观经验吻合,符合物理机理。

2.2 时间序列预测结果

对于时间序列预测问题,本文对比分析了多种模型,包括经验模型和神经网络模型(FCNN 与 LSTM)。本文以某水电大坝位移数据为例测试。使用 MSE (Mean Square Error, 均方误差,指的是参数估计值与参数真值之差平方的期望值)值描述预测值和真实值的差距,MSE 越小意味着预测精度越高;同时以 R2 score (R 方值,决定系数,反映的是因变量的全部变异能通过回归关系被自变量解释的比例)值描述预测数据的变化趋势和真实数据变化趋势的相似性,它越接近 1 则模型预测精度越高。根据实验结果,经验模型难以在所有坝段均取得良好的效果。而基于深度学习的 FCNN 模型和 LSTM 模型在每个坝段上都取得了明显优于经验模型的良好效果,具有较好的预测稳定性。

本文对不同模型所取得的时长一年预测结果绘图展示,如图 3 所示。纵坐标代表位移数据。蓝色曲线表示真实值,红色带三角标识的曲线是预测值。对比图 3 可知,经验模型能够在一定程度上拟合数据的变化趋势,但是在预测时间点与已知时间点距离较远时精度较差。即随着时间推移,经验模型存在误差累积的问题。然而对于深度学习模型而言,预测结果不但很好反映了真实值的变化趋势,而且在具体数值上也预测准确,具有较好的预测效果。

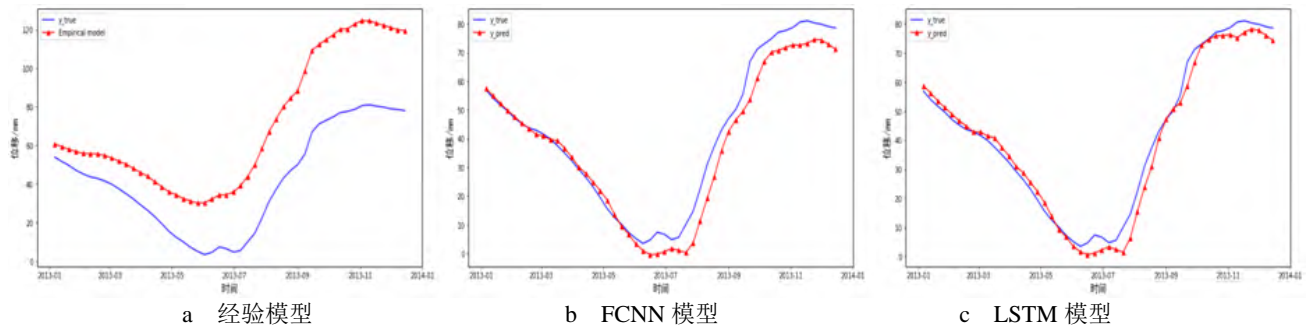


图3 不同模型时序预测结果

2.3 时间序列补全结果

对于时间序列补全问题,本文也分析了传统经验模型、ARMAX(基于 ARMA 的拓展模型,加入了其他通道的数据)、神经网络(FCNN 和 LSTM)以及 LightGBM 模型的效果,实验结果如表 1 所示。实验使用了过去 6 年半的历史数据,其中,训练数据为 2011 年 6 月到 2013 年 12 月间近 2 年半的数据,需要补全的数据为 2014 年 1 月到 2018 年 9 月近 4 年的数据。其中所有数据都已经经过归一化处理。与之前的评价标准相同,在表 1 中,“/”左侧的数值代表的是 MSE,“/”右侧的数值代表的是 R2 score。

表1 某水电站多坝段位移时间序列补全结果

模型	A19.X	A19.Y	A25.X	A25.Y	A29.X	A29.Y	平均
经验模型	0.519/-10.69	0.023/0.63	2.235/-43.92	0.280/-3.40	0.087/-0.42	0.028/0.59	0.529/-9.54
ARMAX	0.083/-0.40	0.186/-1.85	0.167/-2.46	0.191/-2.02	0.185/-1.84	0.212/-2.16	0.171/-1.79
FCNN	0.076/-0.08	0.035/0.59	0.126/-1.52	0.018/0.71	0.025/0.62	0.012/0.85	0.049/0.20
LSTM	0.065/0.13	0.064/0.06	0.103/0.02	0.051/0.02	0.068/0.02	0.062/0.53	0.069/0.13
LightGBM	0.045/0.29	0.041/0.34	0.050/-0.12	0.032/0.50	0.021/0.68	0.018/0.73	0.035/0.40

由表 1 可知,对于补全大范围缺失数据,经验模型并不可靠,在某些坝段上的预测出现了较大的偏差,如 A19.X, A25.X。所以经验模型对于这种大范围的时间序列补全问题几乎是不可用的。ARMAX 作为经典的时间序列模型,可以综合分析其他通道的数据来对待补全通道数据进行估计,在很多问题上都已经证明了它的性能。但是由于缺失数据的时间段过长,ARMAX 模型给出的结果也较差。针对同一问题,以 FCNN 和 LSTM 为代表的神经网络模型取得了可以接受的效果,但是结果不如 LightGBM 稳定。因此,针对大范围数据补全的问题,目前效果最好且精度最高的模型是基于决策树构建的 LightGBM 模型,其补全效果如图 4 所示。图中前半部分是已知的训练数据,后半部分橘色线代表的是真实值,蓝色带三角标识的曲线代表的是预测值。实验表明,LightGBM 模型可以比较准确地预测后 4 年数据的变化趋势。

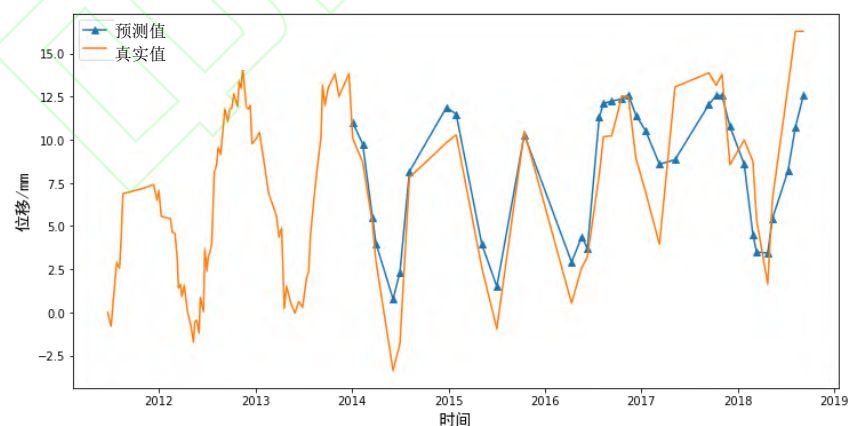


图4 LightGBM 模型实验结果

3 结论

本文采用了以神经网络和决策树为主的机器学习模型来解决水电工程中的大坝安全数据时间序列预测和时间序列补全问题,有利于提升大坝检测数据的质量与完整性,进而促进构建智慧大坝安全评估体系。为了验证模型的有效性,本文根据某水电站的真实数据进行了实验,对比分析了传统经验模型、ARMA 模型、与基于机器学习的 FCNN 模型、LSTM 模型和 LightGBM 模型。根据实验结果可以得出以下结论:

(1)通过 FCNN 和 LSTM 模型对大坝安全数据(位移、渗流等)进行短期的预测是可行的。FCNN 和 LSTM 属于神经网络模型,具有较强的表达能力,可以高效学习并拟合不同类型物理量之间的映射关系。因此该模型可以基于易于获得的环境量特征(温度,上下游水位,降水)和待测安全数据本身的历史数据,对未来短期的物理量进行预测,生成可靠的结果。这一方法有助于降低运维成本并提升监测质量。

(2)通过 LightGBM 模型对大坝安全数据进行长期的时间序列补全是可行的。模型基于学习到的物理量之间的映射关系,结合其他坝段的相同数据作为基础,对长期的缺失数据进行补全。根据本文的实验结果,通过结合高斯过程和 LightGBM 模型可以取得相对较好的补全结果。

(3)通过本文的实验可知,对于大坝安全数据补全与预测这一传统的问题,采用机器学习模型可以对一些由实际工程出发得出来的经验公式进行补充和完善,获得更好的预测效果。机器学习和专家经验的结合有利于提升模型的效果,对于构建智慧大坝安全评估体系极为重要。

参考文献:

- [1] 周建平, 杜效鹄, 周兴波, 等. 世界高坝研究及其未来发展趋势[J]. 水力发电学报, 2019, 38(2): 1-14.
- [2] 杜效鹄. 中国水坝安全状况分析与研究[J]. 水力发电, 2019, 45(2): 1-17.
- [3] 钟登华, 王飞, 吴斌平, 等. 从数字大坝到智慧大坝[J]. 水力发电学报, 2015, 34(10): 1-13.
- [4] 李庆, 林斌. 论智能大坝[J]. 水力发电学报, 2014, 33(1): 139-146.
- [5] 向衍, 盛金保, 刘成栋. 水库大坝安全智慧管理的内涵与应用前景[J]. 中国水利, 2018(20): 34-38.
- [6] 赵斌, 吴中如. BP 模型在大坝安全监测预报中的应用[J]. 大坝观测与土工测试, 1999, 23(6): 1-4.
- [7] 樊琨. 基于人工神经网络的岩土工程力学参数反分析[J]. 河海大学学报(自然科学版), 1998, 26(4): 98-102.
- [8] 蒋利娟. 基于线性回归的水位预测模型研究[D]. 郑州: 河南师范大学, 2017.
- [9] 姜成科. 基于遗传算法的神经网络在大坝变形预报中的应用[D]. 大连: 大连理工大学, 2008.
- [10] QUIONERO-CANDELA J, RASMUSSEN C E. A unifying view of sparse approximate Gaussian process regression[J]. Journal of Machine Learning Research, 2005, 6: 1939-1959.
- [11] KE Guolin, MENG Qi, Finley T, et al. Lightgbm: A highly efficient gradient boosting decision tree[C]//Proceedings of the 31th International Conference on Neural Information Processing Systems, 2017: 3149-3157.
- [12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [13] PAPPAS S Sp, EKONOMOU L, KARAMOVSANTAS D Ch, et al. Electricity demand loads modeling using Autoregressive Moving Average (ARMA) models[J]. Energy, 2008, 33(9):1353-1360.
- [14] 陈云天. 基于机器学习的测井曲线补全与生成研究[D]. 北京: 北京大学, 2019.
- [15] 赵卿. 大坝变形分析多测点统计模型的应用研究[D]. 武汉: 武汉大学, 2010.
- [16] 江新兰. 云计算环境下水坝安全监测模型组合方法研究[D]. 北京: 中国农业大学, 2016.
- [17] 李宁. 基于多源信息融合的复杂系统安全监测及诊断方法研究[D]. 昆明: 云南大学, 2010.
- [18] 姜彤, 杜国倩. 模糊评判法在大坝安全评价中的应用[J]. 人民黄河, 2012(3):117-119.
- [19] WILSON A, ADAMS R. Gaussian process kernels for pattern discovery and extrapolation[C]//Proceedings of the 30th International Conference on Machine Learning-Volume 28, 2013: III-1067-III 1075.
- [20] DEMPSTER A. A generalization of Bayesian inference[J]. Journal of the Royal Statistical Society: Series B, 1968, 30(2): 205-232.
- [21] 张东晓, 陈云天, 孟晋. 基于循环神经网络的测井曲线生成方法[J]. 石油勘探与开发, 2018, 45(4): 598-607.
- [22] FRIEDMAN J H. Greedy function approximation: A gradient boosting machine[J]. Annals of Statistics, 2001, 29(5): 1189-1232.
- [23] CHEN J F, WANG, Weiming, HUANG Chaoming. Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting[J]. Electric Power Systems Research, 1995, 34(3):187-196.