Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

A Time Series Data Cleaning Framework Based on LSTM Prediction Model for Pumped Storage

Changtian Ying[1,2], Weiqing Wang[2,*], Jiong Yu[3], Qi Li[1], Jianhua Liu[1]
[1]School of computer, Shaoxing University, Shaoxing 312000, China
[2]School of Electrical Engineering, Xinjiang University, Urumqi, 830000, China
[3]School of Software, Xinjiang University, Urumqi, 830000, China
*Corresponding Author.

Abstract:

Data security in pumped storage power plants is a critical part of grid development because it is one of the most important power generation technologies in renewable energy. During operation and production, pumped storage unit equipment generates a huge quantity of data, and the fast and accurate storing and processing of millisecond-generated equipment operation data is a critical part of avoiding and detecting equipment operation stability. However, a substantial amount of anomalous and missing data is generated during the collection and transmission of data linked to pumped storage power plants due to the complicated operating environment and operating circumstances of pumped storage units, as well as communication issues and equipment failures. Traditional statistical and machine learning approaches are predicated on complete data sets, and the inclusion of missing data makes it difficult to use and analyze the data sets, reducing their usefulness. Because pumped storage data collected is commonly time series data, this paper employs the HDFS and Spark frameworks to clean and process the anomalous data, which includes data input, data management and cleaning, using long short term memory to supervise, forecast, and fill in the cleaning of the irregular and incomplete data, then using spark to clean and process the anomalous data. To boost efficiency, the data is processed online via spark streaming. The tests were tested using a batch of data obtained from a pumped storage farm to forecast the rack amplitude of a pumped storage unit in order to verify the reliability of the presented prediction model with abnormally absent data. Eclipse was used to setup options and configuration files for Eclipse, Spark, Scala, and HDFS, resulting in an integration platform for Spark. Several evaluations were taken to acquire the average results, taking into consideration practical factors such as platform implementation and reliability, and measuring the real-time performance of these two algorithms, ARIMA and SARIMAX, with the deep learning model LSTM algorithm in a practical setting. The investigations have demonstrated that the prediction findings are generally congruent with real data, with just an error rate of less than 4%, while consuming less effort. LSTM is used to analyze pumped storage data, feature extraction, and prediction in this study, which is implemented on Tensorflow and Spark Streaming platform. The prototype is first trained using historical indicator data, and then the needed information is forecasted using available data, enabling missing and anomalous data to be filled in for data cleansing while considering practical factors such as platform implementation and stability. The study also assesses the stability of pumped storage equipment oscillation by comparing the real-time performance of two statistical models, ARIMA and SARIMAX, with the deep learning model LSTM.

Keywords: *Renewable energy, RNN, Data cleaning, Spark*

I. INTRODUCTION

Hydroelectric power generation is the most technically mature and stable power generation technology among renewable energy generation. In order to ensure the safety and stability of the power system and the quality of power supply, pumped storage power stations are generally used to complement wind power and photovoltaic power generation on the grid [1-5]. This complementary grid connection of renewable and clean energy sources can mitigate the impact on the grid caused by the volatility and intermittency of wind and photovoltaic power generation and improve grid security [6-10]. Pumped storage power stations are the most special type of hydroelectric facility. Pumped storage power stations set power generation and energy storage in one, is the world scale pumped storage power station is a revolutionary change in the form of conventional power generation, is a major innovation in the construction of electricity, but also an important example of ecological civilization. Pumped storage power plants start and stop quickly, flexible working condition conversion, with grid-connected power generation, frequency and phase regulation, peak shaving and valley filling, energy storage and accident backup, etc., to enhance the grid on intermittent energy consumption capacity and power system transmission efficiency, to ensure the safety of the grid are playing a vital role, in response to sudden power supply problems to provide the largest effective protection, the most mature technology of mechanical energy storage methods.

Due to the complex operating environment and operating conditions of pumped storage units, coupled with communication reasons and equipment failures, a large amount of abnormal data and missing data is generated when collecting and transmitting data related to pumped storage plants. Classical statistical learning methods and machine learning methods are based on complete data sets, and the presence of missing data makes the use and analysis of data sets very difficult, making the effectiveness of these methods compromised. In the case of pump turbine operation data, for example, the time series data collected contains a large number of errors such as missing values, outliers and unaligned time tags. The reasons for these errors are varied. Constrained by physical conditions and technical means, certain physical quantities such as guide vane opening and torque are difficult to measure continuously on the unit, and each time the measurement error is relatively large. There are variables such as flow rate and speed where the accuracy of the parameter measurement can often be a bottleneck, as the valid information contained in the data and the error in the measurement are often on the order of magnitude. Up to now, data scientists have had a very detailed analysis of the errors that occur in traditional relational databases. However, the analysis of error types in time series data leaves much to be desired, and deep learning methods can effectively address the problem of data cleaning [11-15].

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

The contributions of this paper are:

Anomaly data prediction and cleaning. In order to make better use of data containing anomalous values for cleaning, this paper proposes a data cleaning method based on LSTM (Long short term memory) prediction with multiple fill and repair, which uses a model to describe the data distribution of data sets containing different anomaly rates and anomaly patterns, and generates multiple new complete data sets using multiple fill and repair methods according to the data distribution. The proposed method is applied to data cleaning in the case of anomalous data, as the pumped storage field data is missing due to human or equipment reasons during the collection process.

Real-time online processing of abnormal data. Pumped storage big data has the characteristics of relatively independent data between devices, little correlation, the same form of data for each device and a large amount of calculation, so pumped storage big data is very suitable for processing using parallel computing techniques. Therefore, in order to achieve real-time, efficient streaming processing of hydroelectric big data, real-time online computing is designed. In order to increase the throughput of the system, Tensorflow and Spark Streaming are used to slice and dice the method from the accuracy and efficiency of dynamic balance cleaning.

II. RELATED WORK

For this problem of time series data cleaning, the following two difficulties were found through the research.

The data volume is large and the error rate is high. The main source of time series data is sensor acquisition. The frequency of these sensor acquisitions is often in the order of seconds. The data collected by the sensors is often not that accurate, some because it is difficult to measure accurately, and some may be distorted due to the power of the sensors themselves, etc. Traditional data cleaning methods are less efficient and less accurate.

The biggest difference between time series data and relational data is that time series are continuous, that is, time series will be continuously generated and stored. Therefore, for time series, it is important that the cleaning algorithm supports online operations (real-time operations). On-line anomaly detection or cleaning algorithms can monitor physical quantities in real time, alerting or cleaning within a reasonable range to ensure the continuous operation of the production line in an efficient and reliable manner.

Existing algorithms often do not support online calculations. (1) There are many different types of data models and different types of data, so it is not appropriate to directly construct rules for identifying abnormal data. When reconstructing abnormal data, it needs to rely on external source data, which is easy to cause errors. (2) Abnormal data processing is often off-line processing, which cannot monitor the data in real time and has low processing efficiency.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

There are two major categories of algorithms for time series prediction as a whole: traditional machine learning algorithms based on statistics (note: machine learning algorithms for short, same below), and deep learning algorithms based on neural networks [16-20]. Commonly used machine learning algorithms, such as exponential smoothing (holt-winters), moving average autoregressive models (ARIMA) and seasonal moving average autoregressive models with a cyclical component (SARIMAX), and commonly used deep learning algorithms, such as deep neural networks (DNNs), recurrent neural networks (RNNs) and long and short term memory networks, which solve the problem of long-term dependence of RNNs. of long and short term memory networks (LSTM) [21-26].

III. THE DATA CLEANING FRAMEWORK

3.1 Problem Analysis

Time series data of pumped storage units is a series of different value changes corresponding to a parameter of a pumped storage unit recorded in relation to the back and forth of time, combined to form a series, a high dimensional data. Define the hydroelectric series s = s[1], s[2],... , where s[i] is the i-th data point value of the hydropower sequence. Each hydropower data point has a timestamp and a range of values, e.g. [0, max] to indicate the maximum error range of all points in the sequence. However each data point should have a similar range to the neighbouring data points, in terms of timing, under the [0, max] range. When a data point exceeds this range of values, or is missing, it needs to be cleaned. Cleaning of a time series involves changing the position of the corresponding data point without changing the time stamp of the time series column. The time series cleaning problem can therefore be converted to analyse and make predictions of outliers based on the time series data and use the predicted values to find a cleaned sequence that meets the range of predicted values.

3.2 LSTM-Based Data Cleaning Framework

Pumped storage units are generally equipped with condition monitoring systems to monitor the operating stability (vibration, oscillation, pressure pulsation, etc.) parameters of the unit in order to determine whether the unit is abnormal and to take timely measures. However, in actual operation, due to various reasons such as sensors, signal conversion devices and data acquisition systems, the data detected by the condition monitoring system is often abnormal and does not reflect the actual operating status of the unit. For example, the performance of the sensors in monitoring the vertical vibration of the upper frame of the pumped storage unit deteriorates over a long period of time, resulting in the peak-to-peak data calculation deviating from the normal trend and not reflecting the actual state of the unit, causing the operation and maintenance personnel to misjudge the operating state of the unit. Therefore, this paper proposes a data cleaning framework for distributed pumped storage plants as shown in Figure 1, using HDFS and Spark framework to clean and process abnormal data, including data input, data monitoring and cleaning, and spark streaming online processing and other aspects.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

The key steps of the cleaning framework include:

3.2.1 Numerical conversion of the data

The original data files were sourced from the database and stored in the form of .CSV files, where the time series fields were long integer values, which were not convenient for data analysis. Distributed time data updates were carried out through the hadoop platform distributed environment and the hive data warehouse tool.

3.2.2 Data splitting by measurement point information

The original data file is the measurement point collection data of all units in all power stations stored in a. CSV file, the actual analysis needs to be carried out in accordance with the hierarchical relationship between the company, power station and unit, and the operation data collected by each unit is exported separately according to the measurement point classification information.

3.2.3 Data abnormal value processing

Pre-data cleaning and outlier processing of equipment production and operation data requires the load value of the unit to be referred to the operation standard, the design of characteristic quantity index gaps, and the use of LSTM algorithm-based time series prediction, and the extreme values, abnormal values and error values in the data are processed.

Establishing a data wide table. Corresponding data tables were established for the basic information of the pumped storage unit (including information on the region, manufacturer, rated capacity, etc.), information on the monitoring part of the unit (including information on the classification of measurement points, characteristic quantities, measurement units, etc.), and information on the unit's operating data (including information on measurement point ID, measurement time, measurement point value), to facilitate further data analysis in the follow-up.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

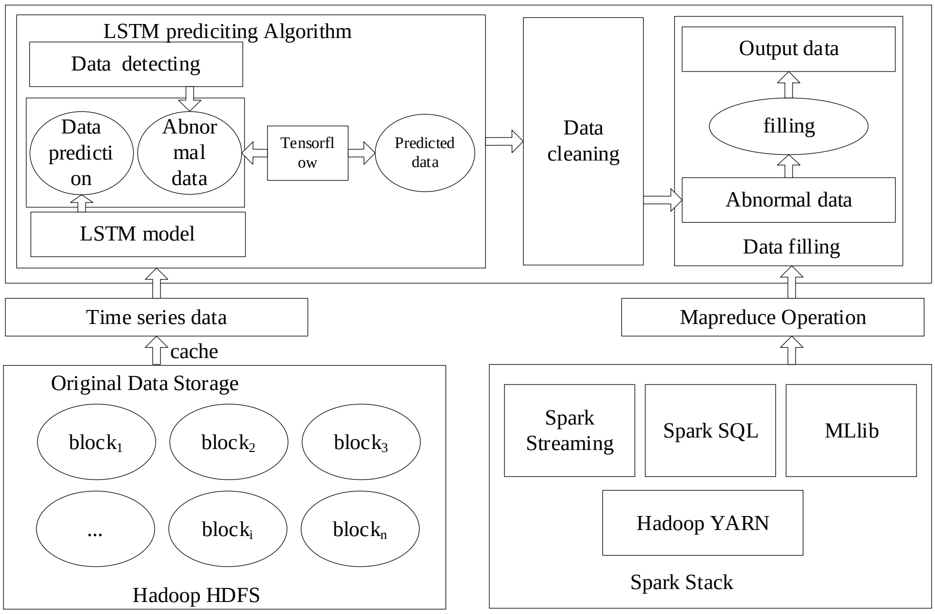


Fig 1: LSTM-based framework for time-series data sequence cleaning

3.3 Time Series Data for Pumped Storage Units

Typical time series data for pumped storage units are the parameters of the operating condition characteristics of the production equipment of a turbine unit. The turbine units are large mechanical rotating equipment and their operating condition characteristics are influenced by and directly related to the hydraulic, electrical and mechanical factors. For the pumped storage unit, the measurement points of each part of the characteristic quantity are counted, as shown in Table I, and its main monitoring measurement points are characterized as follows: combined with the measurement point distribution table of the unit, the characteristic quantity parameters affecting the operating state of the hydraulic turbine unit can be summarized, mainly the process quantity parameters with relatively slow changes, including: stator core temperature, bearing tile temperature, cooling water temperature, oil temperature, etc.; and the working condition parameters with rapid changes, including: rotational speed Vibration (upper and lower frame X, Y, Z axis, top cover, nail core vibration), oscillation (upper and lower water guide bearing X, Y oscillation), pressure pulsation in various parts, active/reactive power, water head (head).

TABLE I. Pumped storage unit data measurement points

Serial number	category		unit	Serial number	category		unit
1	vibrations	upper frame Horizontal X vibration	μm	14	vibrations	Upper X Throw	μm
2		Upper frame horizontal Y vibration	μm	15		Upper Y Throw	μm
3		Upper frame vertical Z vibration	μm	16		Lower X Throw	μm

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

4		Stator core horizontal 1 vibration	μm	17		Lower Y Throw	μm
7		Lower frame horizontal X vibration	μm	20		Steel tube pressure pulsation	kPa
8		Lower frame horizontal Y vibration	μm	21		Inner pressure pulsation between tops	kPa
9		Lower frame vertical Z vibration	μm	22	Pressure pulsation	Pressure pulsation on the outside of the roof	kPa
10		Top cover horizontal X vibration	μm	23		Pressure pulsation behind the guide vane	kPa
11		Top cover horizontal Y vibration	μm	24		Pressure pulsation at the inlet of the worm gear	kPa
12		Vertical Z vibration of the top cover	μm			Pressure pulsation at tailpipe inlet	kPa
13	Rotation speed	Rotation speed	r/min				

3.4 Anomalous Data Detection and Cleaning Algorithms

As shown in Figure 2, the LSTM algorithm for temporal prediction is based on training a model based on the historical data of the indicator and then predicting the required data based on the current data. The specific operation uses the Tensorflow framework, which is a Tensor (tensor: analogous to a multi-dimensional array) for numerical computation in the form of a data flow graph.

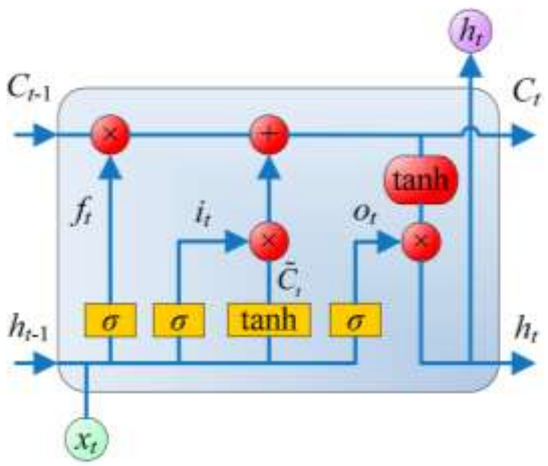


Fig 2: The long short term memory network structure

Step 1. Outlier detection. According to the threshold range, [0, max], the data is analysed to detect outliers and missing points in the time series.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

Step 2. The pre-processed data is converted into a tensor form acceptable to Tensorflow. The data is normalized (to facilitate fast convergence of the model training), the training data is transformed into the format (X: x(1), ..., x(n); Y: x(n+1)), and Reshape into a tensor of the specified shape.

Step3: Define the network structure. Select the LSTM network, set the parameters: input layer (number of layers, number of neurons in the hidden layer), output layer, timestep, initialize the connections (weights) and bias between each neuron.

Step4: Network training. Set the optimizer, loss function, learning rate, activation function, set the training method (epoch, batch_size).

Step5: Production prediction. Save the model, and use the reuse model to predict the outliers, note that the inverse normalization is required.

3.5 Spark Streaming

In classical data cleaning algorithms, the entire data is often modified to achieve a global optimum. The general step is to first collect all the data and then treat all the data as a whole and clean it together. The desired online cleaning of streaming data, especially time series data, is often not supported. In order to ensure online computation, the global optimum for all data points needs to be split into a series of local optima for some of the data points and the received data points are cleaned one by one based on the local optimum as in shown in Figure 3.

Step 1. First, the files are stored on HDFS, which can be read directly by the Spark Streaming program, and then all the read files are reduced to the previously trained neural network model using the Tensorflow framework for deep learning prediction.

Step 2. Load the real-time input data stream into memory in time slice Δt.

Step 3. Construct a DAG task scheduling graph based on the in-memory RDD to execute and schedule the cleaning tasks of the hydroelectric big data in phases.

Step 4. Where the task scheduling module maintains all TaskSet and Task status, and allocates Task according to the remaining resources. Intermediate results and task topology are kept in memory for fast iterative computation and interaction. The abstraction of task topology and streaming data is key in this process, including the definition, loading, starting and online updating of task topology, as well as the flow of data Step 5.

Step 5. The corresponding parallel methods are used for the cleaning of different types of anomalous data. For general non-aggregated processing, such as common data cleaning and data format normalisation, a parallel pipeline processing algorithm based on partitioning is used to enable data processing to be completed quickly in partitioned units and improve the efficiency of data conversion,

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

while for corresponding aggregated processing, partitioned pre-aggregation is used to minimise the frequency of data transfer.

Step 6. For the abnormal data, the prediction results obtained by LSTM algorithm are used, and filled and cleaned.

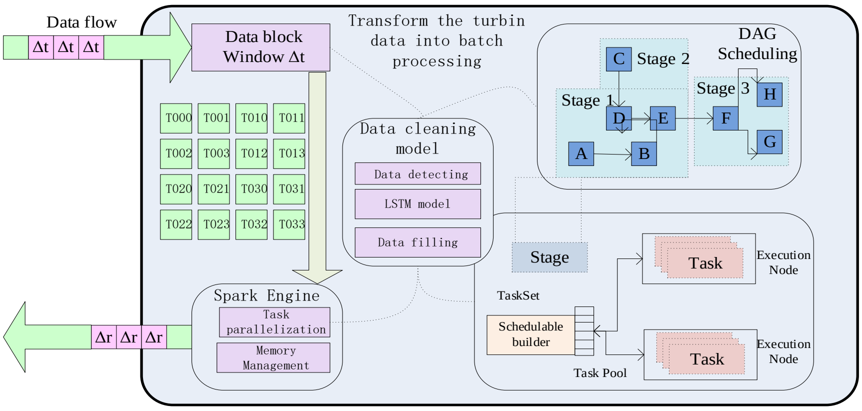


Fig 3: Online data cleaning process based on Spark framework

IV. EXPERIMENTS

4.1 Experimental Configuration

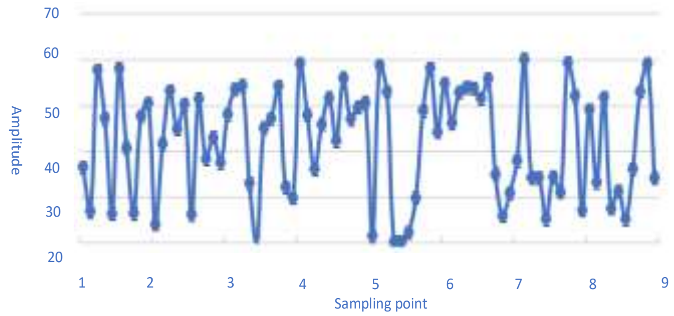
In order to assess the validity of the proposed prediction model with anomalously missing data, the experiments were validated using a batch of data collected from a pumped storage farm to predict the rack amplitude of a pumped storage unit. An integrated development environment for Spark was built using Eclipse to configure parameters and environment variables for Eclipse, Spark, Scala and HDFS. The data is stored in the Hadoop Distributed File System (HDFS) and the trained Tensorflow model is called in Spark, i.e. Tensorflow offline training, Spark real-time prediction, AI production deployment model.

4.2 Model Prediction

Taking the X-axis vibration of the upper stage of the unit as an example, the vibration amplitude data was measured at the average value of the vibration every hour, and 9000 sets of sample data were intercepted continuously, with the first 9000 sets of data as the original time series and the subsequent 900 sets of data as the trend prediction verification object. The experiments show that the average predictive results are generally consistent with real data, with a prediction error of less than 4% and a lower time required.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

(a)



(b)

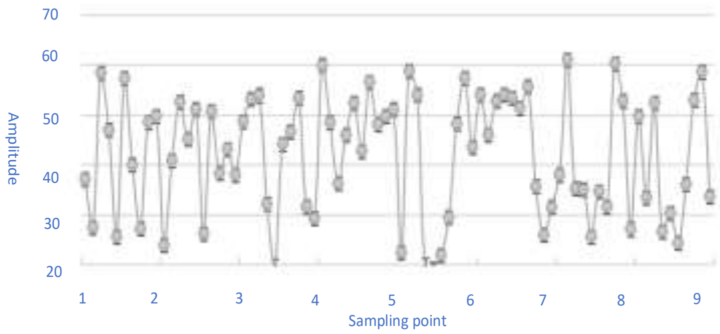


Fig 4: Unit vibration time series (a). Actual measured values (b). Predicted values

4.3 Comparison of Different Algorithms

Taking into account the practical factors such as platform deployment and stability, and comparing the real-time performance of two algorithms, ARIMA and SARIMAX, with the deep learning model LSTM algorithm, in a real environment, multiple comparisons were made to take the average values, and the following conclusions were drawn in table II:

TABLE II. The comparison of different algorithms

Algorithm	Training and prediction methods	Training time	Predicted elapsed time	One day resource consumption	Is it loss value sensitive
ARIMA	One training plus prediction	5min 27s		High	Yes
SARIMA	Generate model once One prediction	7min 43s	51s	Low	Yes
LSTM	Generate model once One prediction call every 15 minutes	4min 12s	35s	Low	No

As can be seen, the LSTM model not only takes much less time to train and predict than the other two models, which cannot meet the demand in real time, but also outperforms the other two models in terms of resource consumption and sensitivity to missing values.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

4.4 Pumped Storage Unit Oscillation Safety Assessment

The mean, variance and anomaly intervals of the oscillation data at each measurement point were calculated by dividing the oscillation range between the extreme and extreme small values into one thousand intervals with equal boundaries, and examining the number of oscillations falling in each interval.

TABLE III. Safety assessment of pumped storage unit

Name of measurement point	Mean value/μm	Variance/nW	Alarm interval/μm
Upward guidance +X	102.9	4.910	128.9~154.6
Lower guidance +Y	137.7	5.66	73.2~120.1
Water-guided oscillation +X	103.0	7.500	7.67~11.6
Upper frame horizontal vibration +X	88.6	0.94	4.28~10.08
Horizontal vibration of the lower frame X	115.4	1.15	7.74~14.7

The amplitude of the oscillation of the pumped storage unit in normal operation is roughly normally distributed, so the normal range of the amplitude of the oscillation can reasonably meet the need for alarm thresholds for abnormalities in actual engineering, and when the oscillation of the unit exceeds the upper threshold of the alarm, early warning treatment can be carried out.

V. CONCLUSION

With the continuous development of smart grids, the amount of data and frequency of data collection of grid equipment monitoring data is getting higher and higher. Inevitably, abnormal data, such as illegal formats, missing values and abnormal values, will appear in the massive data, so there is a strong need for power big data cleaning. In this paper, based on Tensorflow and Spark Streaming architecture, LSTM is used to analyse pumped storage data, feature extraction and prediction. The model is first trained based on the historical data of indicators, and then the required data is predicted based on the current data, so that the missing data and abnormal data can be filled in to achieve data cleansing, taking into account the practical factors such as platform deployment and stability The project also compares the real-time performance of two statistical models, ARIMA and SARIMAX, with the deep learning model LSTM, and provides a brief safety assessment of pumped storage equipment oscillation.

ACKNOWLEDGEMENTS

This research was supported by a project grant from National Natural Science Foundation of China (52067020, 61862060), the joint found of Zhejiang Provincial Natural Science Foundation of China and Huadong Engineering Corporation Limited (Grand No. LHQ20F020001), Scientific Research Project of Zhejiang Provincial Department of Education (Grand No. Y201940952), Higher education research project of Zhejiang Higher Education Society (KT2021126), Teaching Reform Project of Shaoxing University.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
May-June 2022 Page No. 166-178
Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022

REFERENCES

[1] Abdel-Nasser, M., & Mahmoud, K. (2019). Accurate photovoltaic power forecasting models using deep LSTM-RNN. NEURAL COMPUTING & APPLICATIONS, 31(7), 2727–2740. https://doi.org/10.1007/s00521-017-3225-z

[2] Zhang, Y., Qin, C., Srivastava, A. K., Jin, C., & Sharma, R. K. (2020). Data-Driven Day-Ahead PV Estimation Using Autoencoder-LSTM and Persistence Model. IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS, 56(6), 7185–7192. https://doi.org/10.1109/TIA.2020.3025742

[3] Jaseena, K. U., & Kovoor, B. C. (2021). Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks. ENERGY CONVERSION AND MANAGEMENT, 234. https://doi.org/10.1016/j.enconman.2021.113944

[4] Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. IEEE TRANSACTIONS ON SMART GRID, 10(1), 841–851. https://doi.org/10.1109/TSG.2017.2753802

[5] Wang, K., Qi, X., & Liu, H. (2019). Photovoltaic power forecasting based LSTM-Convolutional Network. ENERGY, 189. https://doi.org/10.1016/j.energy.2019.116225

[6] Wu, Y.-X., Wu, Q.-B., & Zhu, J.-Q. (2019). Data-driven wind speed forecasting using deep feature extraction and LSTM. IET RENEWABLE POWER GENERATION, 13(12), 2062–2069. https://doi.org/10.1049/iet-rpg.2018.5917

[7] Srivastava, S., & Lessmann, S. (2018). A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. SOLAR ENERGY, 162, 232–247. https://doi.org/10.1016/j.solener.2018.01.005

[8] Gao, B., Huang, X., Shi, J., Tai, Y., & Zhang, J. (2020). Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. RENEWABLE ENERGY, 162, 1665–1683. https://doi.org/10.1016/j.renene.2020.09.141

[9] Aslam, M., Lee, J.-M., Altaha, M. R., Lee, S.-J., & Hong, S. (2020). AE-LSTM Based Deep Learning Model for Degradation Rate Influenced Energy Estimation of a PV System. ENERGIES, 13(17). https://doi.org/10.3390/en13174373

[10] Kim, D., Kwon, D., Park, L., Kim, J., & Cho, S. (2021). Multiscale LSTM-Based Deep Learning for Very-Short-Term Photovoltaic Power Generation Forecasting in Smart City Energy Management. IEEE SYSTEMS JOURNAL, 15(1), 346–354. https://doi.org/10.1109/JSYST.2020.3007184

[11] Jahangir, H., Tayarani, H., Gougheri, S. S., Golkar, M. A., Ahmadian, A., & Elkamel, A. (2021). Deep Learning-Based Forecasting Approach in Smart Grids with Microclustering and Bidirectional LSTM Network. IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, 68(9), 8298–8309. https://doi.org/10.1109/TIE.2020.3009604

[12] Agga, A., Abbou, A., Labbadi, M., & El Houm, Y. (2021). Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models. RENEWABLE ENERGY, 177, 101–112. https://doi.org/10.1016/j.renene.2021.05.095

[13] Memarzadeh, G., & Keynia, F. (2020). A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets. ENERGY CONVERSION AND MANAGEMENT, 213. https://doi.org/10.1016/j.enconman.2020.112824

[14] Ko, M.-S., Lee, K., Kim, J.-K., Hong Chang Woo and Dong, Z. Y., & Hur, K. (2021). Deep Concatenated Residual Network with Bidirectional LSTM for One-Hour-Ahead Wind Power Forecasting. IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, 12(2), 1321–1335. https://doi.org/10.1109/TSTE.2020.3043884

[15] Jaseena, K. U., & Kovoor, B. C. (2020). A hybrid wind speed forecasting model using stacked autoencoder and LSTM. JOURNAL OF RENEWABLE AND SUSTAINABLE ENERGY, 12(2). https://doi.org/10.1063/1.5139689

[16] Heidari, A., & Khovalyg, D. (2020). Short-term energy use prediction of solar-assisted water heating system: Application case of combined attention-based LSTM and time-series decomposition. SOLAR ENERGY, 207, 626–639. https://doi.org/10.1016/j.solener.2020.07.008

[17] Woo, S., Park, J., Park, J., & Manuel, L. (2020). Wind Field-Based Short-Term Turbine Response Forecasting by Stacked Dilated Convolutional LSTMs. IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, 11(4), 2294–2304. https://doi.org/10.1109/TSTE.2019.2954107

[18] Alavi, S. A., Mehran, K., Vahidinasab, V., & Catalao, J. P. S. (2021). Forecast-Based Consensus Control for DC Microgrids Using Distributed Long Short-Term Memory Deep Learning Models. IEEE TRANSACTIONS ON SMART GRID, 12(5), 3718–3730. https://doi.org/10.1109/TSG.2021.3070959

[19] Shahid, F., Zameer, A., Mehmood, A., & Raja, M. A. Z. (2020). A novel wavenets long short term memory paradigm for wind power prediction. APPLIED ENERGY, 269. https://doi.org/10.1016/j.apenergy.2020.115098

[20] Shuai, H., & He, H. (2021). Online Scheduling of a Residential Microgrid via Monte-Carlo Tree Search and a Learned Model. IEEE TRANSACTIONS ON SMART GRID, 12(2), 1073–1087. https://doi.org/10.1109/TSG.2020.3035127

[21] Zhou, H., Liu, Q., Yan, K., & Du, Y. (2021). Deep Learning Enhanced Solar Energy Forecasting with AI-Driven IoT. WIRELESS COMMUNICATIONS & MOBILE COMPUTING, 2021. https://doi.org/10.1155/2021/9249387

[22] Tyralis, H., & Papacharalampous, G. (n.d.). Boosting algorithms in energy research: A systematic review. https://orcid.org/0000-0002-8932-4997

[23] Devaraj, J., Madurai Elavarasan, R., Shafiullah, G. M., Jamal, T., & Khan, I. (2021). A holistic review on energy forecasting using big data and deep learning models. International Journal of Energy Research, 45(9), 13489–13530. https://doi.org/10.1002/er.6679

[24] Wang, H., Liu, Y., Zhou, B., Li, C., Cao, G., Voropai, N., & Barakhtenko, E. (2020). Taxonomy research of artificial intelligence for deterministic solar power forecasting. In Energy Conversion and Management (Vol. 214). Elsevier Ltd. https://doi.org/10.1016/j.enconman.2020.112909

[25] Manero, J., Béjar, J., & Cortés, U. (2018). Wind energy forecasting with neural networks: A literature review. Computaciony Sistemas, 22(4), 1085-1098. https://doi.org/10.13053/CyS-22-4-3081