BRAIN. Broad Research in Artificial Intelligence and Neuroscience
ISSN: 2068-0473 | e-ISSN: 2067-3957
Covered in: Web of Science (WOS); PubMed.gov; IndexCopernicus; The Linguist List; Google Academic; Ulrichs; getCITED; Genamics JournalSeek; J-Gate; SHERPA/RoMEO; Dayang Journal System; Public Knowledge Project; BIUM; NewJour; ArticleReach Direct; Link+; CSB; CiteSeerX; Socolar; KVK; WorldCat; CrossRef; Ideas RePeC; Econpapers; Socionet.

2024, Volume 15, Issue 2, pages: 437-453 | https://doi.org/10.18662/brain/15.2/561
Submitted: January 18th, 2024 | *Accepted for publication:* May 9th, 2024

Comparative Performance Analysis of Filling Missing Values Algorithms in PdM Systems of UAV

Dragoş Alexandru ANDRIOAIA[1]
Vasile Gheorghiță GĂITAN[2]
Bogdan PĂTRUȚ[3]
Iulian FURDU [4]

[1] Ştefan cel Mare University of Suceava, Suceava, Romania; "Vasile Alecsandri" University of Bacău, Bacău, Romania; dragos.andrioaia@ub.ro
[2] Ştefan cel Mare University of Suceava, Suceava, Romania, vgaitann@usm.ro
[3] Alexandru Ioan Cuza University, Iaşi, Romania, bogdan@info.uaic.ro
[4] "Vasile Alecsandri" University of Bacău, Bacău, Romania; ifurdu@ub.ro

Abstract: *With the development of the IoT domain, the volume of data produced by various applications has also increased. Due to multiple reasons, such as sensor failure, communication system failure, and human errors, the data acquired from the sensors have missing values. The presence of missing values in the dataset affects the informational content of the dataset and thus affects the process of extracting knowledge from the data. In this paper, the authors present a comparative analysis of the performances of the methods of filling in the missing values, such as method, Interpolation, Mean, the K-Nearest Neighbors (KNN), and Random Forests (RF), on the data coming from a Predictive Maintenance (PdM) system that can be used at Unmanned Aerial Vehicle (UAV). The data on which the performance of these methods has been studied comes from a PdM system from the UAVs, used to identify the defects of the Brushless DC (BLDC) motors and estimate the Remaining Useful Life (RUL) of Li-ion batteries.*

Keywords: *missing values data UAV; imputation method; IoT, predictive maintenance; PdM.*

How to cite: Andrioaia, D. A., Găitan, V. G., Pătruţ, B., & Furdu, I. (2024). Comparative performance analysis of filling missing values algorithms in PdM systems of UAV. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 15*(2), 437-453. https://doi.org/10.18662/brain/15.2/561

Introduction

The data collected from the sensors of the Predictive Maintenance (PdM) systems of the Unmanned Aerial Vehicles (UAV) may contain missing values. Incomplete datasets are a common problem in the IoT domain.

A dataset containing missing values can cause distortions due to differences between observed and unnoticed data. Missing values in datasets may also be due to outlier values, most algorithms replace outlier values with missing values. Incomplete datasets can lead to results that are different from those that would have been obtained from a complete dataset (Mohamed Noor, Abdullah, Yahaya, & Ramli, 2014), (Chen, Huang, Lo, Chen, & Lai, 2022), (Imamura, Abedin, Sixian, Tabassum, & Ahmed, 2021), (Friend et al., 2018).

The Remaining Useful Life (RUL) estimate of a component of UAVs may be affected by datasets that have missing values. For this reason, it is necessary to find the best method to estimate the missing values so that the analyzed data reflect the real environment (Samad, Abrar, & Diawara, 2022), (Gungor, Rosing, & Aksanli, 2022), (Khan et al., 2022).

There are two possibilities for filling in the missing values: one refers to univariate time series, and the second is to multivariate time series. The methods of filling in multivariate time series missing values use the entire available dataset to estimate the missing values. A value missing from the *j* dimension of the *i*- features can be derived from the values of the related sensors. Methods to fill univariate time series missing values can generate values in a single dimension, using only available data from that dimension. Methods for filling in univariate time series missing values are present in a much larger number than those for filling in multivariate time series missing values. Among the forms of replacing missing values in multivariate time series, we can mention: the mean, the median, the most frequent value, linear or logistic regression, interpolation, filter-based Kalman methods, etc. (Liu, Dillon, Yu, Rahayu, & Mostafa, 2020).

When the missing data appears randomly, the missing values are often filled with a fixed value, often chosen as a sample mean value (Liu et al., 2020). In other situations, some of those responsible for data analysis delete cases with missing values before further researching them. Both methods of dealing with missing values require low computing resources but are not effective because they do not take into account the distribution of data and lead to an inaccuracy in the subsequent analysis of the data (Okafor & Delaney, 2021).

The methods that allow filling in the missing values can be classified into two types. Within the first type, the missing values are completed based on a mathematical calculation. The second type uses machine learning, which uses the knowledge gained from the dataset to determine the missing values (Dubey & Rasool, 2020).

Missing data can have three types of missing models: missing completely at random, the missing data has no dependence on other data; missing at random, the missing data depends on the other data in the dataset; missing not at random, missing data is due to other missing data which makes missing data unpredictable (Dubey & Rasool, 2020), (Ali, Abu-Elkheir, Atwan, & Elmogy, 2022).

This paper presents a comparative study to identify the best method for completing the missing values from the data from a PdM system within the UAV. The work is organized as follows: Section 1.0 presents the problem and importance of finding the most appropriate method to fill in the missing values of the data collected from the sensors of the PdM systems of the UAV. Section 2.0 presents the main concerns of researchers in the specialized literature on finding the most optimal method for completing the missing values. Section 3.0 offers the ways that have been selected to find the best form to fill in the missing values, as well as the leading indices used to evaluate the performance of regression algorithms. Finally, section 4.0 presents the stands from which the experimental data originated and the methodology by which the data were obtained. In section 5.0, the experimental results obtained are described, and in section 6.0, the conclusions of this study are presented.

State of the art

Regarding the researcher's concerns on the determination of the most effective methods of replacing the missing values, we can specify:

K. Phimmarin performs a comparative analysis on the performance of the methods of replacing missing values as K-Nearest Neighbors (KNN), Cluster-K-Nearest Neighbor (CKNN), Local Least Square (LLS), Cluster-base Local Least Square (CLLS), Iterated Local Least Square (ILLS), and Bayesian Principal Component Analysis (BPCA). The comparison was carried out on five datasets of the same size, coming from different domains. The comparison revealed that BPCA and ILLS were the most effective methods of replacing missing values (Phimmarin KEERIN, 2021).

Y. Fu, H. Liao and L. Lv present a comparative analysis of methods such as Random Forests (RF), Support Vector Regression (SVR), Artificial Neural Network (ANN), Mean, and Multiple Imputation (MI), which can be

used to replace missing values. The analysis was carried out on a free database called UNSODA, in which the properties of the soil are analyzed. The study showed that the RF and MI methods had best performance (Fu, Liao, & Lv, 2021).

T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona analyzes the performance of KNN and missForest methods, which can be used to replace missing values, on a dataset coming from a fan. Data from sensors measuring vibrations and bearing temperatures were used. Experimental results show that the two methods have similar performance and can be successfully used to replace missing values (Emmanuel et al., 2021).

T. Rhaudatunnisa and N. Wilantika compare the performance of three methods that can be used to fill in missing values, from the point of view of filling accuracy but also from the point of view of filling time. The methods used to fill in missing values are: Hot-Deck (HD), K-Nearest Neighbor Imputation (KNNI) and Predictive Mean Matching (PMM). The experimental results suggest that HD Imputation is the most accurate method for replacing missing values (Raudhatunnisa, 2022).

Y. Zhang, C. Kambhampati, D. N. Davis, K. Goode, and J. G. F. Cleland perform a comparative analysis of methods, Naïve Bayes (NB), KNN, Decision Tree (DT) and Multilayer Perceptron (MLP). The performance of the methods was generated on a medical dataset on heart failure. The study suggests that some combinations of methods of generating missing values are suitable for processing clinical data on heart failure (Zhang, Kambhampati, Davis, Goode, & Cleland, 2012).

N. Michikazu, C. Ding-Geng, N. Kunihiro, and M. Yoshihiro investigate the effectiveness of four methods, Mean, Last Observation Carried Forward (LOCF), and MI, that can be used to generate missing values. The performance of the methods was analyzed on a medical dataset. There were studies of the cases in which the missing data were present at the rate of 5%, 30%, and 50% of the total data. Experimental results show that MI is the most efficient method of generating missing values (Nakai, Chen, Nishimura, & Miyamoto, 2014).

C. Velasco-Gallego and I. Lazakis conduct a comparative study by which they try to determine the methods that have the best performance of completing the missing values on the data coming from the ship's sensors in the maritime industry. A comparative study was developed to examine the performance of 20 algorithms. From the analysis performed it was found that the Autoregressive Integrated Moving Average (ARIMA) algorithm has the best performance (Velasco-Gallego & Lazakis, 2020).

Methods used for Detection of Missing Values

In this study, four methods were used to estimate missing values in the datasets of the PdM systems from UAV, and the performance of these methods was compared.

The four methods are the interpolation method, mean value, KNN method, and RF method. Datasets contain data from sensors used to monitor Brushless DC (BLDC) motors (temperature, acceleration, current and voltage) and sensors used to monitor the process of discharging Li-ion batteries (current, voltage).

Use Interpolation to fill in Missing Values

Interpolation is a commonly used method for filling in missing values because it generates new data points in the range of the data series with missing values. Among the most used techniques for estimating missing values, we can mention the interpolation techniques: linear, square, and cubic (Noor, Yahaya, Ramli, & Al Bakri, 2013), (Picornell et al., 2021), (Fan et al., 2020), (Gang, Feng, Xiuyou, Hao, & Jing, 2011).

Linear interpolation is the most common method of interpolation and involves connecting two data points by a line. The linear interpolation equation is (1) (Noor et al., 2013):

$$f_1(x) = f(x_0) + {(((f(x_1) - f(x_0))}\big/{(x_1 - x_0)})(x - x_0) \tag{1}$$

Where: $f1(x)$ is the value of the dependent variable for a value of the independent variable x, x_0 and x_1 are known values of the independent variable.

Polynomials can be also used to approximate functions on a bounded interval $x \in [a, b]$, to find approximations of the intermediate values of the function. Polynomial interpolation finds the degree n polynomial passing through $n+1$ points in the xy plane.

Polynomial interpolation algorithms are expensive in terms of computational resources. Also from the category of polynomial interpolation methods we can mention: Lagrange polynomial interpolation, spline interpolation, polynomial interpolation with finite differences, etc. (Knott, 2018).

Due to the fact that the polynomial interpolation over the entire interval $[x_1, x_n]$ does not always converge, the polynomial interpolation on portions appeared, Spline interpolation, where a polynomial is defined on each portion (Knott, 2018).

Fill in Missing Values Using the Mean

This method replaces the missing values with an mean of all data in the data series, (2) (Sundararajan & Sarwat, 2020), (Pandey, Singh, Sayed-Ahmed, & Abu-Zinadah, 2021):

$$X_i^{'} = \bar{X} = 1/n \sum_{i=1}^{i=n} X_i \tag{2}$$

Where: n is the number of available data, X_i' is the imputed value and X_i is the available data points.

Fill in Missing Values Using KNN

The nearest k neighbors' algorithm is a supervised machine learning algorithm used in classification problems but also regression (P. Keerin, Kurutach, & Boongoen, 2012), (Huang, Wei, Yi, & Liu, 2018), (Bajpai & He, 2020).

In terms of finding the missing values, its unsupervised version is used. Within the unsupervised version of the KNN method there is no learning from the data that has been labeled (Salvador-Meneses, Ruiz–Chavez, & Rodríguez, 2019), (Shokrzade, Ramezani, Akhlaghian Tab, & Abdulla Mohammad, 2021), (Song & Wu, 2020).

Within the unsupervised variant of the KNN method for finding the missing values, the available data from the dataset are used, the one that makes the method preserve the initial structure and properties of the given dataset (Sundararajan & Sarwat, 2020). Each missing value is generated using values from the closest neighbors who have values for features. Features of neighbors are evenly averaged or weighted depending on the distance from each neighbor. Several methods can be used to measure the distance: Euclidean distance, Mahalanobis distance, Minkowski distance, etc. (Okafor & Delaney, 2021).

In the KNN variant with weighted feature mediation, the method looks for the k the closest neighbors to the missing data points, and then the weighted average of the closest values is taken and used to fill in the missing point (Okafor & Delaney, 2021). Minkowski distance of order p from k nearest neighbor is equal to (3) (Sundararajan & Sarwat, 2020):

$$d = \sum_{i=1}^{m} \left| x^k - x_{missing} \right|^{1/p} \tag{3}$$

The missing X_i' value of the i feature is (4) (Sundararajan & Sarwat, 2020):

$$X_i^{'} = \sum_{k=1}^{k} (wx^k) \Big/ k \tag{4}$$

Where: k is the number of close points; w is the value of the weight, and depends on the distance (Sundararajan & Sarwat, 2020). The number of neighbor's k is chosen according to the feature of the data. A measure by which one can choose an optimal value for k would be depending on the variation of the RMSE index (Okafor & Delaney, 2021), (Sundararajan & Sarwat, 2020).

Fill in the Missing Values Using RF Algorithm

RF are supervised machine learning techniques that generalize Decision Tree (DT) assemblies through bagging (Bootstrap Aggregation) that can be used in both classification and regression problems (Reddy & Parvathy, 2022), (Guo, Zhou, Hu, & Cheng, 2019), (Teng et al., 2020).

They are based on assemblies that work on the principle that a group of weak classifiers will form a strong one. Allied forests have weak classifiers formed with the help of DT's (Reddy & Parvathy, 2022), (Feng, Grana, & Balling, 2021), (Verbeke, Baesens, & Bravo, 2018), (Zhou, Lan, Zhou, & Mo, 2020).

Random Forests are successfully used to generate missing values. Generating missing values using Random Forests is considered a regression problem (Schnitzler, Ross, & Gloaguen, 2019).

The number of trees in the alloying forest is selected based on the *nRMSE* index. Within the iterative algorithm, the number of trees will be varied until the minimum value of the *nRMSE* index is discovered, (5) *Var(X)* is the function of calculating the variance of *X*. Dataset *X* is divided in terms of attributes into attributes with present Y_{obs} values and attributes with missing values $Y_{missing}$ (Sundararajan & Sarwat, 2020).

$$nRMSE = \sqrt{(X - X') \Big/ Var(X)}^{\,2} \tag{5}$$

$$s = \left(x_{iter}^{'} - x_{iter-1}^{'}\right)^2 \Big/ \left(x_{iter}^{'}\right)^2 \tag{6}$$

At first, the missing values in the *X* dataset are filled in with an average value. In the next step, the training dataset that has the X_{obs} input data and the Y_{obs} target is introduced into the RF model. The missing values are

fine-tuned in the iterative process until the difference calculated using (6) of the values generated on two future iterations, only decreases (Sundararajan & Sarwat, 2020).

Performance Evaluation of Missing Value Filling Algorithms

Algorithms for filling in missing values can be evaluated using the indicators used to assess regression algorithms within machine learning models. Thus, the most commonly used indicators are: Mean Squared Error (*MSE*), Root Mean Square Error (*RMSE*), and Mean Absolute Error (*MAE*), (7) (de-Prado-Gil, Palencia, Silva-Monteiro, & Martínez-García, 2022), (Garbaya, Kallel, Fakhfakh, & Siarry, 2022), (Mantri, Sharma, & Jayaraman, 2022).

$$\begin{aligned} MAE &= \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \\ MSE &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \\ RMSE &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \end{aligned} \tag{7}$$

Where: y_i is the actual output value; $\hat{y}_i$ is the output value generated by the model.

The algorithms for filling in the missing values are performant when they have as low values as possible for these indicators.

Experimental Stands

In this work were used the data coming from two experimental stands.

The first stand Fig. 1 is used to monitor the parameters of BLDC motors used in UAVs, with the aim of identifying their defects. The stand consists of the BLDC motor, A2208 (Fig. 1, - 5) whose speed is controlled by means of an ESC module by the NodeMCU-32S development board (Fig. 1, - 3). In this stand, the temperature is monitored through the BMP180 sensor (Fig. 1, - 6), the vibration through the MPU9250 sensor (Fig. 1, - 4), but also the ESC (Fig. 1, - 7) supply voltage and the closing current by motor, by means of the INA3221 module (Fig. 1, - 2). Experimental data from the sensors are transmitted by the NodeMCU-32S development board (Fig. 1, - 3) to a Raspberry Pi 4 (Fig. 1, - 1), where they are stored in a SQL database.

The second stand Fig. 2, monitors the voltage and current on each discharge cycle of the Li-ion batteries, with the aim of estimating the RUL of the Li-ion batteries. The Li-ion battery (Fig. 2, - 3) is charged and discharged successively via the NodeMCU-32S module (Fig. 2, - 5) which commands the relay (Fig. 2, - 1) to switch from the charging position to the discharge position. The discharged capacity is calculated on each cycle in real time.

The data is generated using a 3.7 [V] Li-ion battery with a capacity of 500 [mA/h]. To discharge the Li-ion battery was used a BLDC motor, A2208 (Fig. 2, - 7), which is controlled by the NodeMCU-32S module via an ESC (Fig. 2, - 6).

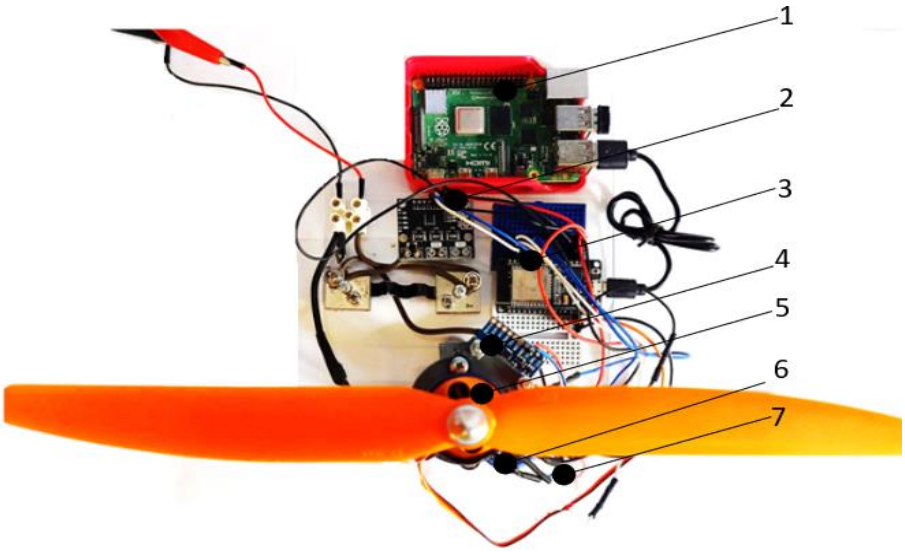

Fig. 1. Stand used for monitoring the parameters of BLDC motors: 1 – Raspberry Pi 4; 2 – INA3221; 3 – NodeMCU-32S; 4 – MPU9250; 5 – A2208; 6 – BMP180; 7 – ESC.

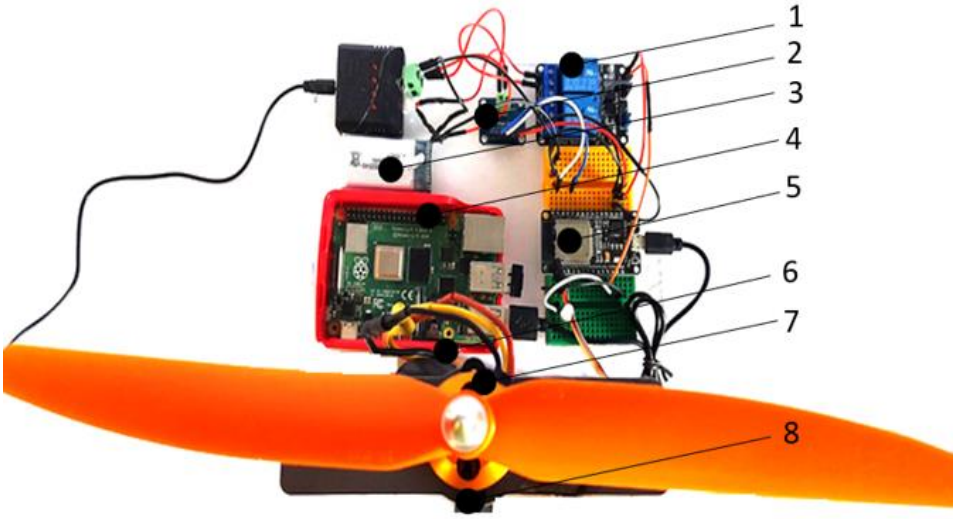

Fig. 2. Stand used to monitor the discharge cycles of the Li-ion batteries: 1 – Electrical relay; 2 – INA219; 3 – Li-ion battery, 3.7 V, 500 [mA/h]; 4 – Raspberry Pi 4; 5 – NodeMCU-32S; 6 – ESC; 7 – A2208; 8 – XL6009.

The supply voltage of the motor was adjusted to 7.5 [V]. To obtain a voltage of 7.5 [V] from the voltage of 3.7 [V] of the Li-ion battery, an XL6009 voltage lift module was used (Fig. 2, - 8).

The current that closes through the motor and the discharge voltage are monitored by the NodeMCU-32S module (Fig. 2, - 5) via module INA219 (Fig. 2, - 2). NodeMCU-32S retrieves the data from the INA219 module and transmits it to a Raspberry Pi 4 (Fig. 2, - 4), where it is stored in an SQL database.

Results and Discussions

The performance of the algorithms used to generate the missing values was analyzed in the *Python 3.6* language.

To evaluate the performance of the algorithms to replace the missing values on the data from the monitoring of the BLDC motor, 28376 data points, generated for three minutes, were used. The monitored parameters were temperature, acceleration on the three axes, voltage, and intensity of the electrical current. Fig. 3 shows the graphs of variation of the parameters monitored within the data set used.

The performance of the algorithms used to replace the missing values on the data generated by monitoring the Li-Ion batteries has been analyzed, using the data on the voltage and intensity of the discharge current during

200 operating cycles, thus several 25341 points were used. The variation graphs of the parameters *I* and *U* on the dataset used are shown in Fig. 4.

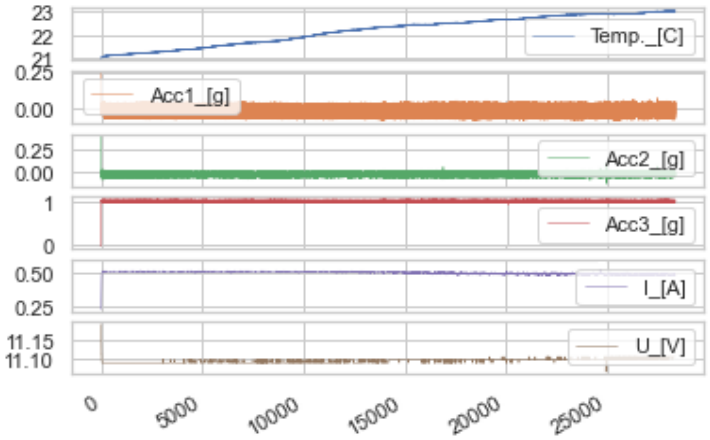

Fig. 3: Variation of *Temp.*, *Acc1, Acc2, Acc3, I* and *U* over the course of 28376 data points.

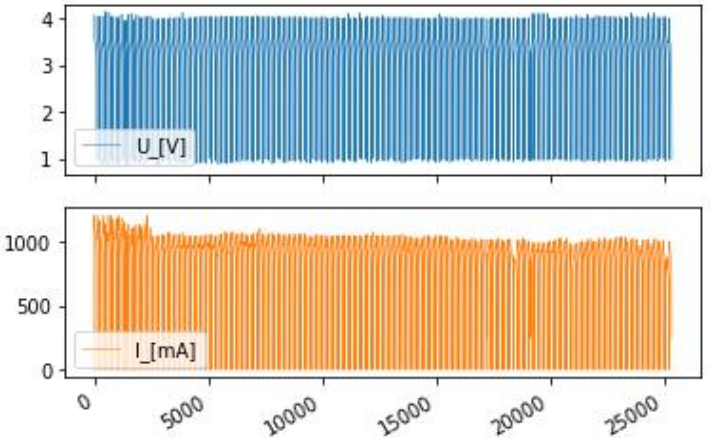

Fig. 4: Variation of *I* and *U* parameters over the course 25341 data points.

Because it would have been difficult to evaluate the performance of the methods of replacing the missing values when the real values are unknown, missing values were entered on randomly chosen positions, to evaluate the ability of the missing values.

Table 1 shows the performances of the missing value replacement algorithms on the data generated by BLDC motor monitoring.

The performance of the algorithms for replacing the missing values on the data generated by monitoring Li-ion batteries over 200 operating cycles is shown in Table 2.

Table 1. Performance of missing value replacement algorithms on data generated by monitoring BLDC motors

Performance index		Linear int.	Mean	KNN	RF
MAE	*Temp.* [°C]	0.0040	0.0988	0.0834	0.0780
	Acc1 [g]	0.0206	0.0246	0.0128	0.0090
	Acc2 [g]	0.0152	0.0134	0.0119	0.0106
	Acc3 [g]	0.0145	0.0195	0.0079	0.0082
	I [A]	0.0006	0.0015	0.0006	0.0004
	U [V]	0.0006	0.0061	0.0054	0.0011
	Mean MAE	**0.0092**	**0.0273**	**0.0203**	**0.01792**
MSE	*Temp.* [°C]	0.0000	0.0135	0.0122	0.0094
	Acc1 [g]	0.0007	0.0009	0.0002	0.0001
	Acc2 [g]	0.0003	0.0002	0.0002	0.0001
	Acc3 [g]	0.0002	0.0004	0.0001	0.0001
	I [A]	0.0000	5.2582	0.0000	0.0000
	U [V]	0.0000	0.0002	0.0003	0.0000
	Mean MSE	**0.0002**	**0.0025**	**0.0022**	**0.0016**
RMSE	*Temp.* [°C]	0.0055	0.1164	0.1100	0.0972
	Acc1 [g]	0.0271	0.0304	0.0172	0.0128
	Acc2 [g]	0.0196	0.0165	0.0155	0.0136
	Acc3 [g]	0.0156	0.0209	0.0123	0.0119
	I [A]	0.0015	0.0022	0.0000	0.0013
	U [V]	0.0015	0.0171	0.0174	0.0023
	Mean RMSE	**0.0118**	**0.0339**	**0.0291**	**0.0232**

Table 2. Algorithms performance to replace missing values on data used to generate by monitoring Li-ion batteries

Performance index		Linear int.	Mean	KNN	RF
MAE	*I* [A]	1.6145	74.9881	25.3267	24.9732
	U [V]	0.0059	0.1751	0.1209	0.0879
	Mean MAE	**0.8102**	**37.5816**	**12.7156**	**12.5276**
MSE	*I* [A]	13.4588	11486.1273	1906.2743	1816.2647
	U [V]	0.0006	0.0447	0.0245	0.0151
	Mean MSE	**6.7297**	**5743.0860**	**953.14721**	**908.1385**
RMSE	*I* [A]	3.6686	107.1733	43.6609	42.6176
	U [V]	0.0253	0.2115	0.1568	0.1230
	Mean RMSE	**1.8469**	**53.6924**	**21.9013**	**21.3645**

To select the best method for filling in the missing values, the performance of the four methods was analyzed through three performance indicators, *MAE, MSE* and *RMSE*.

By analyzing the results obtained for *MAE, MSE* and *RMSE*, presented under Table 1, it is found that when the missing values are filled in using the interpolation method, the obtained values are the closest to the real values and filling in the missing values with the mean value gives the results furthest from the true values.

By analyzing the results obtained for the *MAE, MSE* and *RMSE*, presented in Table 2, it is found that when the missing values are filled in using the interpolation method, the obtained values are the closest to the real values and the method of filling in the missing values with their mean provides the results furthest from the real values.

Conclusion

In this work, the performances of four methods were compared: the Interpolation method, the Mean method, the KNN method, and the RF method, which can be used to fill in the missing values. For example, to compare the performance of the four methods, data were collected from a PdM system to identify BLDC motor defects and estimate the RUL of Li-ion batteries.

By analyzing the values of the performance indicators, the best method proved to be the linear interpolation method. Completing the missing values with the means of the data provides the results furthest from the desired values.

The method with the best results on these data may not produce the same impact on another dataset with a different informational content.

The performance of the candidate methods should be studied on a dataset with the same informational content to choose the best way of completing the missing values

Acknowledgement

The work of Andrioaia Dragos-Alexandru was supported by the project "PROINVENT", Contract no. 62487/03.06.2022 - POCU/993/6/13 - Code 153299, financed by The Human Capital Operational Programme 2014–2020 (POCU), Romania.

References

Ali, A., Abu-Elkheir, M., Atwan, A., & Elmogy, M. (2022). Missing values imputation using Fuzzy K-Top Matching Value. Journal of King Saud University - Computer and Information Sciences, 35(1), 426-437. doi:https://doi.org/10.1016/j.jksuci.2022.12.011

Bajpai, D., & He, L. (2020, 25-26 Sept. 2020). Evaluating KNN Performance on WESAD Dataset. Paper presented at the 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN).

Chen, Y.-P., Huang, C.-H., Lo, Y.-H., Chen, Y.-Y., & Lai, F. (2022). Combining attention with spectrum to handle missing values on time series data without imputation. Information Sciences, 609, 1271-1287. doi:https://doi.org/10.1016/j.ins.2022.07.124

de-Prado-Gil, J., Palencia, C., Silva-Monteiro, N., & Martínez-García, R. (2022). To predict the compressive strength of self compacting concrete with recycled aggregates utilizing ensemble machine learning models. Case Studies in Construction Materials, 16, e01046-e01063. doi:https://doi.org/10.1016/j.cscm.2022.e01046

Dubey, A., & Rasool, A. (2020). Clustering-Based Hybrid Approach for Multivariate Missing Data Imputation. International Journal of Advanced Computer Science and Applications, 11(11), 710-714. doi:https://doi.org/10.14569/IJACSA.2020.0111186

Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., & Tabona, O. (2021). A survey on missing data in machine learning. Journal of Big Data, 8(1), 140-177. doi:https://doi.org/10.1186/s40537-021-00516-9

Fan, J., Zhang, P., Chen, J., Li, B., Han, L., & Zhou, Y. (2020). Quantitative Estimation of Missing Value Interpolation Methods for Suomi-NPP VIIRS/DNB Nighttime Light Monthly Composite Images. IEEE Access, 8, 199266-199288. doi:https://doi.org/10.1109/ACCESS.2020.3035408

Feng, R., Grana, D., & Balling, N. (2021). Imputation of missing well log data by random forest and its uncertainty analysis. Computers & Geosciences, 152, 104763-104772. doi:https://doi.org/10.1016/j.cageo.2021.104763

Friend, J., Hauck, A., Kurada, S., Hartvigsen, T., Sen, C., & Rundensteiner, E. A. (2018, 5-7 Oct. 2018). Handling Missing Values in Multivariate Time Series Classification. Paper presented at the 2018 IEEE MIT Undergraduate Research Technology Conference (URTC).

Fu, Y., Liao, H., & Lv, L. (2021). A Comparative Study of Various Methods for Handling Missing Data in UNSODA. Agriculture, 11(8), 727-755. doi:https://doi.org/10.3390/agriculture11080727

Gang, S., Feng, W., Xiuyou, W., Hao, W., & Jing, C. (2011, 26-27 Nov. 2011). Application Research on Cubic Spline Interpolation Based on Particle

Swarm Optimization in Mine Pressure Missing Data. Paper presented at the 2011 International Conference on Information Management, Innovation Management and Industrial Engineering.

Garbaya, A., Kallel, I., Fakhfakh, M., & Siarry, P. (2022, 3-4 March 2022). Machine Learning Techniques for Solving Constrained Engineering Problems. Paper presented at the 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET).

Gungor, O., Rosing, T. S., & Aksanli, B. (2022). DOWELL: Diversity-Induced Optimally Weighted Ensemble Learner for Predictive Maintenance of Industrial Internet of Things Devices. IEEE Internet of Things Journal, 9(4), 3125-3134. doi:https://doi.org/10.1109/JIOT.2021.3097269

Guo, Y., Zhou, Y., Hu, X., & Cheng, W. (2019, 8-10 Nov. 2019). Research on Recommendation of Insurance Products Based on Random Forest. Paper presented at the 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI).

Huang, J., Wei, Y., Yi, J., & Liu, M. (2018, 10-11 Feb. 2018). An Improved kNN Based on Class Contribution and Feature Weighting. Paper presented at the 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA).

Imamura, Y., Abedin, N., Sixian, L., Tabassum, S., & Ahmed, A. (2021, 13-14 Dec. 2021). Missing Value Imputation for Remote Healthcare Data: A Case study of Portable Health Clinic System. Paper presented at the 2021 9th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC).

KEERIN, P. (2021). A Comparative Study of Missing Value Imputation Methods for Education Data. Paper presented at the Proceedings of the 29th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education.

Keerin, P., Kurutach, W., & Boongoen, T. (2012, 14-17 Oct. 2012). Cluster-based KNN missing value imputation for DNA microarray data. Paper presented at the 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC).

Khan, M., Ahmad, A., Sobieczky, F., Pichler, M., Moser, B. A., & Bukovský, I. (2022). A Systematic Mapping Study of Predictive Maintenance in SMEs. IEEE Access, 10, 88738-88749. doi:https://doi.org/10.1109/ACCESS.2022.3200694

Knott, G. D. (2018). Introduction to Cubic Spline Interpolation with Examples in Python: CreateSpace Independent Publishing Platform.

Liu, Y., Dillon, T., Yu, W., Rahayu, W., & Mostafa, F. (2020). Missing Value Imputation for Industrial IoT Sensor Data With Large Gaps. IEEE Internet of Things Journal, 7(8), 6855-6867. doi:https://doi.org/10.1109/JIOT.2020.2970467

Mantri, V., Sharma, N., & Jayaraman, R. (2022, 22-24 June 2022). Solar Power Generation Prediction for Better Energy Efficiency using Machine Learning. Paper presented at the 2022 7th International Conference on Communication and Electronics Systems (ICCES).

Mohamed Noor, N., Abdullah, M. M. A. B., Yahaya, A. S., & Ramli, N. (2014). Comparison of Linear Interpolation Method and Mean Method to Replace the Missing Values in Environmental Data Set. Materials Science Forum, 803, 278-281. doi:https://doi.org/10.4028/www.scientific.net/MSF.803.278

Nakai, M., Chen, D.-G., Nishimura, K., & Miyamoto, Y. (2014). Comparative Study of Four Methods in Missing Value Imputations under Missing Completely at Random Mechanism. Open Journal of Statistics, 4(1), 27-37. doi:https://doi.org/10.4236/ojs.2014.41004

Noor, M. N., Yahaya, A. S., Ramli, N. A., & Al Bakri, A. M. M. (2013). Filling Missing Data Using Interpolation Methods: Study on the Effect of Fitting Distribution. Key Engineering Materials, 594-595, 889 - 895. doi:https://doi.org/10.4028/www.scientific.net/KEM.594-595.889

Okafor, N. U., & Delaney, D. T. (2021). Missing Data Imputation on IoT Sensor Networks: Implications for on-Site Sensor Calibration. IEEE Sensors Journal, 21(20), 22833-22845. doi:https://doi.org/10.1109/JSEN.2021.3105442

Pandey, A. K., Singh, G. N., Sayed-Ahmed, N., & Abu-Zinadah, H. (2021). Improved estimators for mean estimation in presence of missing information. Alexandria Engineering Journal, 60(6), 5977-5990. doi:https://doi.org/10.1016/j.aej.2021.04.053

Picornell, A., Oteros, J., Ruiz-Mata, R., Recio, M., Trigo, M. M., Martínez-Bracero, M., . . . Rojo, J. (2021). Methods for interpolating missing data in aerobiological databases. Environmental Research, 200, 111391-111401. doi:https://doi.org/10.1016/j.envres.2021.111391

Raudhatunnisa, T. (2022). Performance Comparison of Hot-Deck Imputation, K-Nearest Neighbor Imputation, and Predictive Mean Matching in Missing Value Handling, Case Study: March 2019 SUSENAS Kor Dataset. Proceedings of 2021 International Conference on Data Science and Official Statistics (ICDSOS), 2021(1), 753-770. doi:https://doi.org/icdsos.v2021i1.93

Reddy, P. D., & Parvathy, L. R. (2022, 20-22 Oct. 2022). Prediction Analysis using Random Forest Algorithms to Forecast the Air Pollution Level in a Particular Location. Paper presented at the 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC).

Salvador-Meneses, J., Ruiz–Chavez, Z., & Rodríguez, J. (2019). Compressed kNN: K-Nearest Neighbors with Data Compression. Entropy, 21(3), 234. doi:https://doi.org/10.3390/e21030234

Samad, M. D., Abrar, S., & Diawara, N. (2022). Missing value estimation using clustering and deep learning within multiple imputation framework. Knowledge-Based Systems, 249, 108968-108980. doi:https://doi.org/10.1016/j.knosys.2022.108968

Schnitzler, N., Ross, P. S., & Gloaguen, E. (2019). Using machine learning to estimate a key missing geochemical variable in mining exploration: Application of the Random Forest algorithm to multi-sensor core logging data. Journal of Geochemical Exploration, 205, 106344. doi:https://doi.org/10.1016/j.gexplo.2019.106344

Shokrzade, A., Ramezani, M., Akhlaghian Tab, F., & Abdulla Mohammad, M. (2021). A novel extreme learning machine based kNN classification method for dealing with big data. Expert Systems with Applications, 183, 115293-115311. doi:https://doi.org/10.1016/j.eswa.2021.115293

Song, C., & Wu, B. (2020, 18-20 Dec. 2020). Evaluation Method of the Attack Effect of Network Based on Rough Set and KNN. Paper presented at the 2020 2nd International Conference on Information Technology and Computer Application (ITCA).

Sundararajan, A., & Sarwat, A. (2020). Evaluation of Missing Data Imputation Methods for an Enhanced Distributed PV Generation Prediction. Paper presented at the Proceedings of the Future Technologies Conference (FTC) 2019. FTC 2019. Advances in Intelligent Systems and Computing.

Teng, Z., Chu, L., Chen, K., He, G., Fu, Y., & Li, L. (2020, 6-8 Nov. 2020). Hardware Implementation of Random Forest Algorithm Based on Classification and Regression Tree. Paper presented at the 2020 IEEE International Conference on Information Technology,Big Data and Artificial Intelligence (ICIBA).

Velasco-Gallego, C., & Lazakis, I. (2020). Real-time data-driven missing data imputation for short-term sensor data of marine systems. A comparative study. Ocean Engineering, 218, 108261. doi:https://doi.org/10.1016/j.oceaneng.2020.108261

Verbeke, W., Baesens, B., & Bravo, C. (2018). Profit driven business analytics: a practitioner's guide to transforming big data into added value: John Wiley & Sons.

Zhang, Y., Kambhampati, C., Davis, D. N., Goode, K., & Cleland, J. G. F. (2012, 29-31 May 2012). A comparative study of missing value imputation with multiclass classification for clinical heart failure data. Paper presented at the 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery.

Zhou, Q., Lan, W., Zhou, Y., & Mo, G. (2020, 13-15 Nov. 2020). Effectiveness Evaluation of Anti-bird Devices based on Random Forest Algorithm. Paper presented at the 2020 7th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS).