IET Intelligent Transport Systems

Special Issue: Intelligent Transportation Systems in Smart Cities for Sustainable Environments

Predictive model for battery life in IoT networks

ISSN 1751-956X
Received on 3rd January 2020
Revised 2nd May 2020
Accepted on 27th May 2020
E-First on 30th July 2020
doi: 10.1049/iet-its.2020.0009
www.ietdl.org

Praveen Kumar Reddy Maddikunta[1], Gautam Srivastava[2,3] ✉, Thippa Reddy Gadekallu[1], Natarajan Deepa[1], Prabadevi Boopathy[1]

[1]*School of Information Technology and Engineering, VIT - Vellore, Tamilnadu, India*
[2]*Department of Mathematics and Computer Science, Brandon University, Brandon, MB R7A 6A9, Canada*
[3]*Research Center for Interneural Computing, China Medical University, Taichung 40402, Taiwan*
✉ *E-mail: Srivastavag@brandonu.ca*

Abstract: The internet of things (IoT) is prominently used in the present world. Although it has vast potential in several applications, it has several challenges in the real-world. One of the most important challenges is conservation of battery life in devices used throughout IoT networks. Since many IoT devices are not rechargeable, several steps to conserve the battery life of an IoT network can be taken using the early prediction of battery life. In this study, a machine learning based model implementing a random forest regression algorithm is used to predict the battery life of IoT devices. The proposed model is experimented on ‘Beach Water Quality – Automated Sensors’ data set generated from sensors in an IoT network from the city of Chicago, USA. Several pre-processing techniques like normalisation, transformation and dimensionality reduction are used in this model. The proposed model achieved a 97% predictive accuracy. The results obtained proved that the proposed model performs better than other state-of-art regression algorithms in preserving the battery life of IoT devices.

1 Introduction

The internet of things (IoT) is a technology which allows communication between objects through predominantly wireless telecommunications [1–3]. IoT adoption has bought dramatic changes in information and communication technology. IoT can be understood in two terms, namely ‘internet’ and ‘things’. IoT is an application of the World Wide Web with interconnected devices. IoT allows any real-world, living as well as non-living entities to communicate through the internet, which can be identified uniquely by their integrated chips through the Internet which runs on the communication protocol Transmission Control Protocol (TCP)/internet protocol (IP) [4]. Therefore objects interconnected by IoT are heterogeneous in nature. IoT enables the most effective transmission of information among real-world entities without any human to computer or human to human interactions. This interaction among physically existing objects and their information systems is achieved through the integration of electronics, electrical, information systems and networks. IoT allows the entities under interaction/integration to sense the information from the application domain and controls them remotely. In IoT, the terms of integration and interaction are used interchangeably, as it allows the real-world entities to interact and share data. IoT applies a variety of technologies and protocols to provide more effective means of communication that can be rendered through any computer to computer interaction [5]. This integration of entities in the real-world helps in the automation of several regular tasks. Thus these entities are called as ‘smart objects’. The general architecture of IoT is depicted in Fig. 1. IoT is the convergence of three different visions, namely things oriented, semantic oriented and internet oriented. These visions are consequent from various perspectives of IoT, which includes the entities such as objects Radio-Frequency Identification (RFID), Near-Field Communication (NFC), actuators, smartphones, Unmanned Aerial Vehicle (UAV) and sensors, complex communication mechanism (middleware) and standards for communication (IP), respectively. The ubiquitous computing nature and varied visions of IoT re-defines it as a model that facilitates communication among ‘any object, from any place and at any time’.

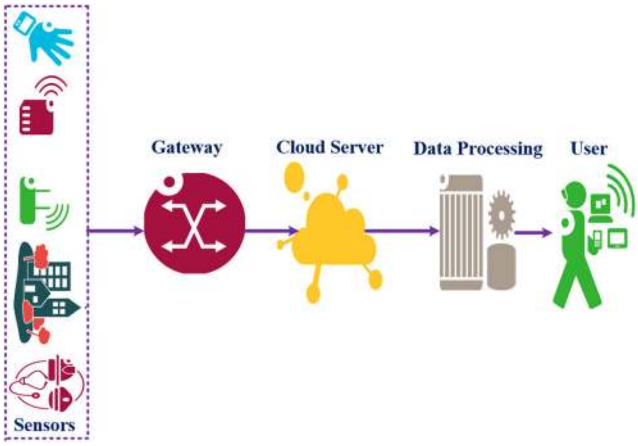

Fig. 1 *IoT architecture*

IoT has its application in almost all fields of digital computing. Some of the major applications include several applications in smart cities such as monitoring structure of health of the building, air quality management, smart garbage management, noise management, smart energy consumption, smart parking etc. [6]. Other applications include smart doorbells, smart transportation, smart restaurants, smart health devices, smart farms in agriculture, smart learning tools, pills enabled digestive sensors, smart medical bottles and smart clothes [7–18]. US National Intelligence Council has recommended IoT as one of the six ‘disruptive civil technologies’. It forecasts that by 2025 all the real-world entities, i.e. all movable and immovable objects will be connected to a node on the internet [19].

Most of these IoT applications rely on batteries. Therefore, the lifetime of the battery used in IoT devices must be monitored continuously for better maintenance, thus avoiding the issues in productivity of the device deployed. Although IoT paves the way for various opportunities by facilitating a sustainable solution to provide access to clean energy all over the world, it has some of the challenges like power or energy management, sensing, security, design complexity and wireless – cloud communication [20]. For a

© The Institution of Engineering and Technology 2020

seamless integration of devices using the internet in IoT, the power consumption and energy utilisation must be carefully monitored and predicted continuously. Several researchers worked on low-power consumption models by considering several factors like protocol for ubiquitous sensor networks which ensures optimisation in acquisition time, power consumption and lesser overhead [21]. Also, the internet engineering task force has taken an initiative of standardising constrained application protocol (CoAP) for allowing seamless integration of low-power consuming devices on the internet [22]. IoT practitioners can utilise this CoAP protocol for their implementation perspectives. A survey made by Gartner suggests that around 26 billion devices will be linked to the internet with low-energy consumption which leads to the promotion of Green IoT thus lowering the power consumption [23]. Machine learning (ML) algorithms [24–27] can play a major role in predicting the battery life of the IoT networks, which in turn help in effectively managing the sensors.

In this work, a novel hybrid Random forest–principal component analysis (PCA) regression algorithm is used to predict an IoT network's battery life. The steps involved in this study are as follows: in the first step, the preprocessing of the IoT data set is performed. In preprocessing, filling up of missing values, normalisation and transformation are performed. In the next step PCA is applied to the preprocessed data to select optimal features. These optimal features are then fed to Random forest regressor for predicting the IoT network's battery life.

The main contributions of this work are summarised below:

(i) Predicting the battery life of the IoT network using random forest regression algorithm.
(ii) Using one-hot encoding scheme to transform the data set.
(iii) Using PCA to select optimal features, remove features which negatively impact the performance of the regression algorithm.
(iv) Evaluate the proposed model's performance with other state of the art regression algorithms.

The organisation of rest of the paper is as follows: related work is discussed in Section 2, in Section 3 background and proposed methodology of the current work is discussed, results are demonstrated in Section 4 and the paper is concluded in Section 5.

2 Related works

Nowadays, IoT sensors have been widely used for monitoring and collecting data for various real-world applications such as health care, agriculture, smart energy management, urban security, transportation, environmental monitoring etc. Many sensor devices are operated using batteries. Therefore, the main challenge in using battery-powered IoT sensors is to predict the batter's lifetime, either recharge or replace the exhausted batteries. Several research works have been carried out on developing models for predicting the lifetime of the battery in IoT devices.

A study was conducted on the lifetime of the batteries and the various patterns of how they discharge the power in the SPHERE environment. The experiment was done on environment sensors and wearable sensor devices incorporated in the SPHERE sensing platform in a residential area. The limitation of the work including abnormal patterns of battery discharge was observed and there was a maximum amount of variation among the devices in the firmware [16].

A narrowband-based IoT has been introduced by third-generation partnership project for solving communication problems with low-energy utilisation with a goal that they should exist for more than ten years without the replacement of a battery. A mechanism has been proposed for the prediction of battery life based on saving energy and the results proved that the energy utilisation using the mechanism is decreased by nearly 34% when compared with the other conventional methods. The proposed work can save battery life between 10 and 34% when implemented in various scenarios. Further investigations were not done on the outcome of contention resolution in various multi-user scenarios [28].

A tool is proposed as a power utilisation model to measure the battery's lifetime in IoT devices by merely assessing the details with respect to the expected network traffic and the feasible level of interferences. The work explains various techniques that can reduce the effect of various components which are used for the consumption of power [29].

Some senor nodes have been powered using solar energy, which is a leading technology in IoT for harvesting the energy. It has been noted that more energy is consumed by IoT sensor nodes when there is an excessive transmission of data. An algorithm is proposed to measure the quantity of energy collected by a solar panel by applying the weighted average of the intensity of light. The error rate of this prediction algorithm is around 0.5% [30].

Location tracking applications are widely used nowadays in smartphones which consists of sensors such as Global Positioning System (GPS), accelerometer, WiFi etc. GPS is the main component for location tracking applications with high-energy consumption and the mobile battery gets drained within a short time. A new location tracking facility, namely SensTrack, is proposed for efficient energy management in smartphones. The limitations of the SensTrack are data processing for accelerometer, optimising energy as well as accuracy and tracking the patterns of multiple mobility [31].

Several methods have been developed to extend the battery life of the industrial IoT wireless sensor network. A platform namely I3Mote is developed with some connected sensor nodes and more I3Motes are integrated to form a maximum coverage of data collectors thus providing a maximum ten years of a lifetime for the connected battery nodes [32]. Sensor nodes are used for data collection and transmitting it to the gateways. These sensor nodes have only a limited supply of energy. To efficiently use the energy, it is advisable to turn off the sensor node after the data transmission is over. In this way, sensor nodes can save their battery life by switching between active and sleep modes automatically. A model is proposed to manage environmental monitoring using a self-monitored algorithm to safeguard the lifetime of the energy stored in the battery. The battery power consumed by the sensor nodes is 2.84% when compared with 12.52% of data read in 2 h cycle. It showed a success rate of 100% transmission for <500 m distance [33].

Wireless sensor nodes powered with battery can be fixed in distant places and they work for a maximum period without the necessity of maintenance. Harvesting energy can be done for recharging the batteries using the available energy sources in the environment. A model is presented with a two-level predictor, which utilises the weather forecasting information from the cloud taken on an hourly basis for the energy prediction, which can be collected simultaneously. The proposed model showed 8% superior results than exponentially weighted moving average predictor when predictions are taken for 24 h [34].

ML algorithms are predominantly used for predicting the lifetime of battery power in sensor nodes, energy harvesting etc. The adaptive power management tool is proposed for harvesting solar energy in sensor nodes using reinforcement learning. For training purposes, historical data is used and the developed tool, which is a power manager, adapt to changes in climate, other parameters and degradation of the battery. The proposed method achieved almost similar energy-neutral operation with 6% less root mean square deviation obtained from energy-neutral operation. The other approaches produced a deviation of >23% compared with energy neutral operation [35].

Rechargeable wireless sensor nodes are becoming popular to increase the life span and reduce the cost of the battery. The information about solar radiation is necessary for planning, deployment and management of self-rechargeable nodes. A forecast model is proposed using ML algorithms for predicting the solar irradiance and the models are developed using leap forward, FoBa, Cubist, bagEarthGCV and spikes lab. The results showed that solar irradiance prediction is accurately evaluated from a few hours to two days by models using ML algorithms in which variations in the season were not considered in weather conditions [36].

© The Institution of Engineering and Technology 2020

A study was conducted on how ML algorithms can be used to predict solar power in a restricted sensor environment with the help of weather data. The results proved that solar energy can be predicted even when there is restricted access to the data. It was observed that the data may contain uncertainty when the predictions are made on weather forecast data and ML algorithm should handle it to improve the accuracy [26].

A model was developed using combined Bayesian techniques, relevance vector machine to predict and forecast the health of the battery. The Bayesian approach is proved to handle uncertainties as it interprets probability distributions on the parameters [37].

To manage the power of energy harvesting, a method is proposed in a wireless sensor network with quality of service (QoS) using a reinforcement learning algorithm. The proposed method produced a duty cycle rate of >50% and increased the remaining energy available in the storage area. The experimental results showed that the proposed method used to obtain QoS demand which was constructed on nearly equivalent residue storage energy outperformed the Kansal's adaptive duty cycling control method [38].

A study was conducted on dynamic power management of energy harvesting in wireless sensor nodes using reinforcement learning and fuzzy approach. The experiment results proved to produce superior energy utilisation with respect to unused battery energy when compared with other existing power management methods. The results showed that the average residual energy of the developed method is 7.5% superior to the other conventional methods with an average exercised duty cycle rate of 45% [39].

As the energy harvesting mechanism has the ability to provide an uninterrupted supply of energy to wireless sensor networks which are battery-powered, it has gained the attention of many researchers. Solar energy is considered to be productive environmental energy for energy harvesting in sensor networks. Therefore, the prediction of the availability of energy in the near future has become a major issue. A novel algorithm was proposed for the prediction of solar energy using Q-learning and the developed algorithm is applied to real-world solar panel data set. The experimental results proved that the proposed algorithm provided much better results for the evaluations taken in long term and can be implemented in the development of Media Access Control (MAC) protocols to predict the quantity of energy harvested in a specific time slot which helps to enhance the performance of wireless sensor networks [40].

A novel algorithm namely RLMan was proposed for energy management in wireless sensor networks for energy harvesting using reinforcement learning. The developed algorithm can provide details about harvested energy as well as the energy consumed in the node. RLMan has proved to provide maximal QoS when compared with other state of art methods such as P-FREEN, LQ-tracker and Fuzzyman. The proposed method does not need extra energy consumption tracking methods in order to get maximum efficiency and QoS. It adjusts to changes in energy consumption [41].

A study was conducted on site-specific prediction models for the generation of solar powers with the help of weather forecasts obtained from the National Weather Service using ML algorithms such as support vector machine (SVM) with multiple kernel functions and linear least square. The experimental results showed that the prediction models developed for seven unique weather forecast measurements using SVM are >27% accurate compared with other forecasting methods. In future, further research can be developed to use onsite solar arrays for power generation for smart homes, data centers etc. [42].

An architecture was proposed for the green cloud environment to provide secure and efficient energy harvesting. The architecture includes a genetic algorithm for the allocation of the virtual machine and multivariate correlation analysis for the detection of denial of service attacks that can happen at data centers [43]. Genetic algorithm and multivariate correlation analysis were used in the proposed framework in order to obtain systematic consumption and utilisation of power. The framework is proved to be secure, efficient and wasted energy is reused.

A deep Q-network based offloading scheme was proposed using reinforcement learning for IoT device, which has been enabled with energy harvesting in order to identify the edge device and also to predict the offloading rate based on the present battery level. The results obtained from simulating the offloading scheme have proved to reduce the computational latency and energy consumption compared with other benchmarking offloading techniques. Furthermore, the research can be extended for evaluating the performance of running real-life applications in order to validate the proposed framework [44].

Data-driven predictive ML models can be used for predicting battery life accurately by making the system to learn from experimental data [45]. Other methods used for the remaining life of the battery are the recurrent neural network, gradient boosted trees, least absolute shrinkage and selection operator optimisation technique and least squares SVM for regression [46].

3 Background and proposed methodology

The power required for an IoT device depends on the application for which it is deployed. Sensors require less energy in sleep mode as it sends only little data whereas, during active mode, it may require more energy as it sends bulk data. However, there are sensors that require more energy all the time. Every component used for IoT application has an impact on the power consumption of the device. The battery life has to be estimated in the IoT devices. Both the rechargeable and non-rechargeable or disposable batteries coexist. These batteries have their own benefits and drawbacks. The three main factors to be considered before choosing battery type for IoT are electrical discharge performance or waste, self-discharge, physical battery features and its safety contributing to its lifetime. Price-wise, disposable batteries are better than rechargeable batteries because of its lower lifetime, whereas lifetime wise, non-rechargeable batteries perform well at the beginning but rechargeable batteries give a better lifetime. Also, depending on the application type, these batteries have to be preferred. Disposable batteries can be preferred for the devices consuming low power or consuming power occasionally for a long period of time (low-drain devices). These primary batteries have higher initial voltage and capacity comparatively but have only a short lifespan. Also, they emit more wastes. Rechargeable batteries can be preferred for high-power consuming devices or devices that drain energy/power faster.

Depending on the type of the battery used, the power consumption and the lifetime should be estimated accordingly. The energy sources have different behaviour, some are stochastic and others are non-deterministic. An in-depth perception of every facet of the different energy system should be given higher importance for effective prediction [23]. Each IoT device embodies an energy management module.

3.1 Measuring power consumption

As IoT is an embedded domain where both software and hardware are integrated for providing the results, predicting the battery life of an IoT device is not only dealt with hardware used, also on the software. Thus, we can make the measurements with the help of the software being deployed on the hardware. This software-based estimation (SBE) is used for determining the different operating states of the hardware deployed. As each operating state consumes different power for being in that state and transit to another state. As SBE provides accurate information for the long term with required customisation in the monitoring of various units deployed, few of the operating system (OS) manufactures too deploy a module for monitoring the energy consumption using SBEs.

A shunt resistor S can also be used for short-term current measurements, which provides an approximation of long-term power consumption. This resistor S is placed in series with the battery and the voltage drop at instant time t is V_t. The current I_t is calculated using ohm's law: $I_t = (V_t/R)$ [47]. A major challenge with low-power devices is its high dynamic range of current, therefore, different resistors can be attached to measure sleep current and active current. As the resistor needs to be attached to

© The Institution of Engineering and Technology 2020

the test device for monitoring the measurements, this method is practically unfit for continuously monitoring the embodied devices.

3.2 Estimating battery life

Once the power consumed by the devices has been measured using either SBE or short-term power measurements, the battery life can be estimated. Consider a linear battery model that has been used. Let Ph be the current measured by the approaches mentioned and N_c be the nominal capability of the battery using the device lifetime LT_h in hours is approximate $LT_h = (N_c/P_h)$. However, these linear battery models do not take into account non-linear effects the resistance developed inside the battery assumes all stored energy in a battery is used and current discharge rate dependence. Researchers noted significant variations in battery lifetime estimated using the model as mentioned above. So while predicting battery life these parameters should be accounted for linear battery model and parameters based on different battery models used must be devised before predicting its lifetime [48].

3.3 ML in predicting battery lifetime

Artificial intelligence (AI), which is the greatest boon of digital computing, is explored in almost all kinds of applications. One of the applications of AI is ML, where machines are trained to behave like a human (human intelligence) and tested for its performance. ML will assist IoT in better prediction. Gartner's report states that almost 80% of IoT projects will incorporate an AI module [49]. In [45], the authors have used AI on their real-time experimental data and devised a new ML model for predicting different battery lifetime. They focused on emerging battery types and observed better results. Deep learning techniques such as convolution neural network (CNN) and artificial neural network (ANN) can be used for predicting battery lifetime.

Although ML models are serving better in most of the applications, there are some limitations. The major drawback of any ML model is data, i.e. it requires enormous data and good featured information. ML models provide excellent results only with useful (meaningful) data and more accurate results with massive data. Therefore, data input determines the performance of the model. Consequently, applying ML models for deterministic problems will yield sparse results. Neural network models are stochastic; they will not provide deterministic values. The neural network models were preferred commonly for battery life prediction of IoT devices. Some of the recommended neural network models for battery life prediction are CNN, ANN and multi-layer perceptron neural network. CNN with meager data set will perform poorly and lead to overfitting problem. So the hypermeters are tuned before proceeding with neural network models to attain global optima by avoiding local minimum. ANN will not perform well for analysing the stochastically independent events [50].

3.4 Linear regression

Regression or prediction is a supervised learning technique through which one can predict the value of the class label which is continuous. Simple linear regression is based on traditional slope–intercept form as shown in (1), where *m* and *b* are the variables, *x* represents input data and *y* represents prediction

$$y = mx + b \tag{1}$$

Another type of regression is multi-variable regression. The equation for multi-variable regression is as

$$f(x, y, z) = w_1x + w_2y + w_3z \tag{2}$$

where *x*, *y* and *z* represent various attributes, w_1, w_2 and w_3 represent constants.

3.5 Random forest regression

Random forest is the Ensemble ML process, which works by creating a multitude of decision-making trees during training and producing a class that is averaged or voted by each tree [51]. Random forest is used in variable selection, classification and regression. The task of random forests is to create decision trees randomly from the training data set. During the training phase, the decision trees will be evaluated. Feature selection plays a vital role as a random forest tries to pick the most important features while making decisions. The random forest has several parameters, one of the parameters focus on the number of decision for the training data set. Another parameter focuses on the number of features. Internally, random forest implements cross-validation and it has appropriate predictive consequences for complicated and non-linear data

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{x}_i - x_i)^2}{\sum_{i=1}^{n} (x_i - \hat{x})^2} \tag{3}$$

where *n* is referred to be the samples $\hat{x}_i$ and x_i are the assumed true values of *i* samples.

3.6 XGBoost regression

The ensemble method, the XGBoost algorithm, is commonly used for high performance and speed applications such as high-energy physics classification, ad-click rate etc. XGBoost is one of the most popular methods used on Kaggle. Many competitions have been won using XGBoost algorithm. XGBoost is ten times faster than most other ML algorithms. XGBoost algorithm uses a boosting technique that produces sequential decision trees where subsequent trees seek to reduce past trees errors. Some of the key features of XGBoost algorithms are regularisation, sparse data handling, effective weighted data handling, parallel learning block structure and out-of-core computing, which helps maximise usable disk space by managing large data sets that do not fit into memory.

The boosting process can be monitored as per (4) given below. Initially, the process will be carried out by taking *F*

$$F_{\mathrm{m}}(x) < -F_{\mathrm{m}} - 1(x) + h_{\mathrm{m}}(x) \tag{4}$$

3.7 Principal component analysis

PCA is a generic method of reducing data dimensionality by determining the most key features. The activities PCA can execute are extraction of functionality, elimination of duplication, data compression, prediction etc. PCA's arithmetic configuration details are given below.

(i) *Mean:* Assume $A_1, A_2, A_3, \ldots A_x$ which denotes the arbitrary variables for a *x* size of the sample. The data set average is an arbitrary function, as can be seen below

$$\bar{A} = \frac{1}{x} \sum_{i}^{x} A_{\mathrm{a}} \tag{5}$$

(ii) *Standard deviation (SD):* In calculating SD it is important to decide the standard range from the data set to a certain value, i.e. measure the distance square from all data points to the data

$$\mathrm{SD} = \sqrt{\frac{1}{x} \sum_{a=1}^{x} (A_i - \bar{A})^2} \tag{6}$$

(iii) *Covariance:* Equation (7) indicates a specification for covariance that is almost equivalent to the specification for variance

$$\mathrm{Cov}(A, B) = \frac{\sum_{a=1}^{x} (A_a - \bar{A})(B_a - \bar{B})}{x} \tag{7}$$

(iv) *Eigenvalues and Eigenvectors of a matrix:* Further, the determination of the Eigenvectors and Eigenvalues is important. If

© The Institution of Engineering and Technology 2020

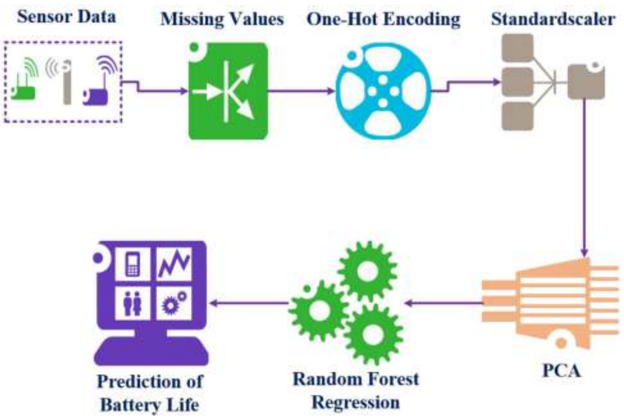

Fig. 2 *Architecture of the proposed model*

Table 1 Data set description

S. No	Name of the attribute	Description
1	beach name	name of the beach in which the sensors are deployed
2	measurement timestamp	date and time in which the data is generated from sensors
3	water temperature	the temperature of the water when sensors capture the data
4	turbidity	it is the haziness of water due to several dissolved particles
5	transducer depth	a transducer is a device that is used to convert changes in the pressure of water into an electrical signal. This attribute says about the depth at which the transducer device is deployed
6	water height	height or depth of water at the location in which sensors are deployed
7	wave period	duration of the waves
8	battery life	this is the class label in the proposed work. The overall life of the battery in the IoT network is captured by this attribute

$\boldsymbol{U}$ is an $x \times x$ matrix, then $\boldsymbol{A} \neq \bar{0}$ is an eigenvector of $\boldsymbol{U}$, where λ is a scalar and $\boldsymbol{A} \neq \bar{0}$

$$[\boldsymbol{U} \parallel \boldsymbol{A}] = \lambda X \tag{8}$$

$$\det\left(\lfloor \boldsymbol{U} - \lambda I = 0 \rfloor\right) \tag{9}$$

For a $\boldsymbol{U}_{x \times x}$ matrix, the distinctive degree of polynomial $\boldsymbol{U}$ is as given in (9). The features obtained using PCA is denoted as fe_b, where $b = 1, 2, 3, \ldots N_f$ and N is the number of features.

3.8 Standard scalar normalisation

In standard scalar normalisation, a feature is standardised by subtracting the mean value and scaled variance to unit. The normalised value depends on mean and variance, this increases the efficiency of the algorithm in terms of throughput and scalability. The standard scalar normalisation is expressed as

$$\hat{M}_i = \frac{M_i - \bar{M}}{\sigma} \tag{10}$$

The proposed architecture is discussed in Fig. 2. The various steps in the proposed model are given below:

(1) Make data collection from several sensors like the temperature of the water, depth of transducer, turbidity, the height of the waves, period of waves and battery life.
(2) Fill the missing values in the data collected using attribute mean.
(3) Using standard scalar method normalise the data generated by sensors between 0 and 1.
(4) Convert the categorical data to numerical data using the one-hot encoding scheme.
(5) In order to reduce the number of attributes in the data generated from step 4 apply PCA. This step also eliminates irrelevant features.
(6) Apply random forest regressor to predict the battery life of the IoT sensors.
(7) Evaluate the performance of the proposed model using measures like mean absolute error, root mean squared error, coefficient of determination and variance.

4 Results and discussion

The proposed model was experimented with the publicly available 'Beach Water Quality – Automated Sensors' data set generated from sensors in the IoT network in the city of Chicago, USA [52]. The sensors in this IoT network capture the data every hour during the summer. A personal laptop with 8 GB RAM, 500 GB hard disk, Windows 10 OS is used for the experimentation. The programming language used for this work is Python 3.7. This data set has 34,923 instances and ten attributes. Two attributes, a label of the measurement timestamp and measurement id, which does not play any role in the battery life, are eliminated. The remaining attributes of the data set are discussed in Table 1.

Before applying the random forest regression algorithm on the IoT network data set, several pre-processing techniques are used in this work, like treating missing values, normalising the data, data transformation and dimensionality reduction. There are numerous missing values in the data set. To treat these values, 'Filling the missing values by attribute means' method, wherein the missing values will be filled up with the mean of the attributes is used. The next step in a pre-processing technique used in the proposed model is a normalisation of the data. To reduce the complexity of the data, the numerical values of the attributes are converted to a range between 0 and 1. To carry out this, 'standard scalar method' is used. To convert the categorical data to numerical data, a popular transformation technique, 'one-hot encoding' is used in this model. After one-hot encoding is applied, the number of attributes is increased to 207. The resultant data set is experimented using three popular prediction/regression supervised algorithms, namely, linear regression, XGBoost and random forest.

The data set may contain some irrelevant and unimportant attributes. Also, the data set may contain the attributes which may negatively impact the performance of the regressor. To eliminate these attributes, a popular dimensionality reduction algorithm, PCA is used in this work with the aim of retaining 90% of the components. After PCA is applied, the number of attributes is reduced to 183 as shown in Table 2. The reduced attributes are then experimented using linear regression, XGBoost and random forest regression algorithms.

© The Institution of Engineering and Technology 2020

Table 2 Data set dimensions

Dimensions	Instances	Attributes
dimensions of the data set	34,923	8
dimensions of the data set after one-hot encoding	34,923	207
dimensions of the data set after one-hot encoding and PCA	34,923	183

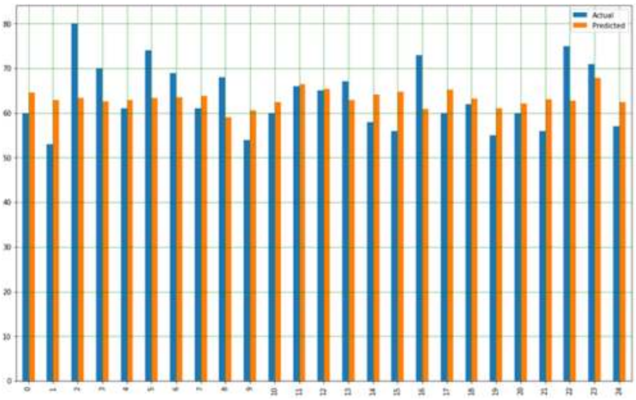

Fig. 3 *Actual values versus predicted value (linear regression)*

Fig. 4 *Actual values versus predicted value (linear regression with PCA)*

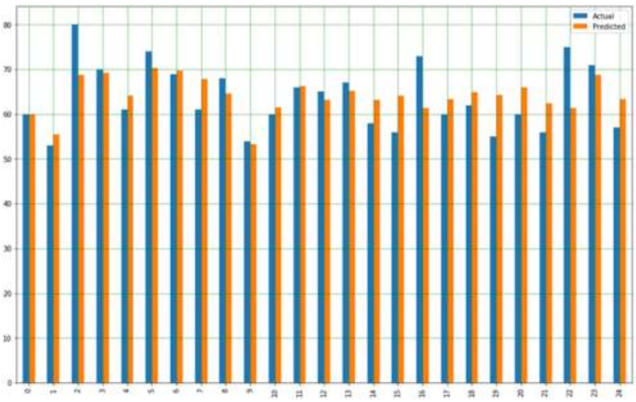

Fig. 5 *Actual values versus predicted value (random forest)*

Figs. 3–8 show the results of these experiments. These figures show the predictions of the algorithms on 25 random instances of the data set. The regression algorithms are evaluated using the metrics, mean absolute error, root mean squared error, coefficient of determination and variance. These results are tabulated in Table 3.

The observations made from the experiments carried out are summarised below:

(i) The proposed model reduces the burden of the regression/prediction algorithms by eliminating unimportant attributes and the

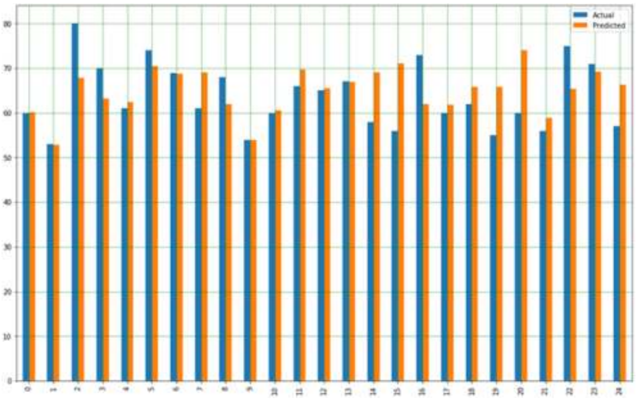

Fig. 6 *Actual values versus predicted value (random forest with PCA)*

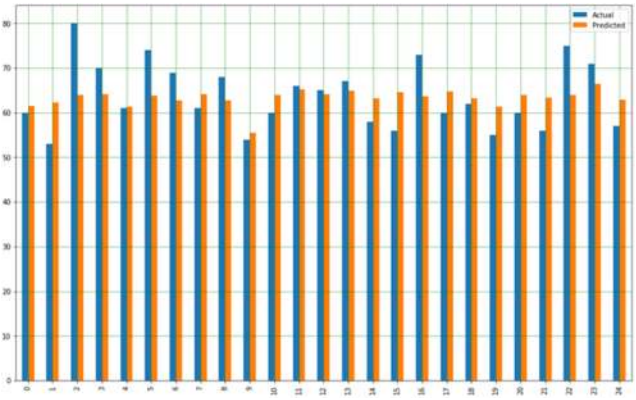

Fig. 7 *Actual values versus predicted value (XGBoost)*

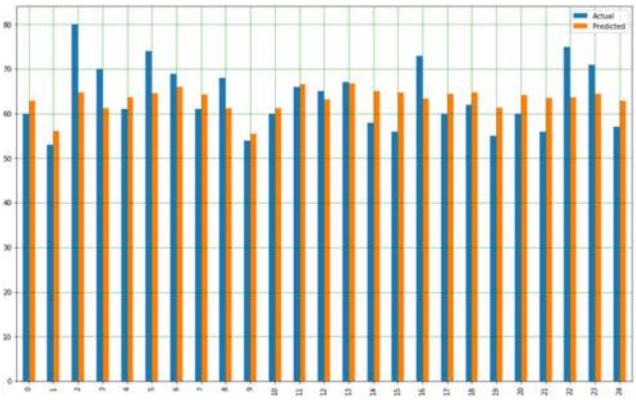

Fig. 8 *Actual values versus predicted value (XGBoost regressor with PCA)*

attributes affecting the performance of the prediction algorithms in a negative manner.

(ii) The performance of the prediction algorithms is enhanced by selecting the optimal features.

(iii) Random forest regression/prediction algorithm outperforms linear regression and XGBoost regression algorithms.

5 Conclusion and future work

IoT has tremendous potential to enhance quality of life of people by incorporating it in many applications like smart cities, smart grids, intelligent home surveillance etc. The actual potential of IoT networks can be realised by overcoming many challenges. In this study, an attempt is made to predict an IoT network's battery life using the PCA-based random forest regression algorithm. During the pre-processing stage, missing values are treated by filling up missing values using attribute mean, standard scalar method is used for normalisation and one-hot encoding method is used for transformation. Then, for feature selection and dimensionality reduction, the PCA algorithm is employed. The reduced dimensions from the data set are then fed to the random forest

© The Institution of Engineering and Technology 2020

Table 3 Performance analysis

Method	Mean absolute error	Root mean squared error	Coefficient of determination R^2 of the prediction	Test variance score
linear regression	988.7187091587965	41,008.537768339535	0.06900683452633649	−25,876,324.98
linear regression with PCA	5.987605663534114	8.852814797391645	0.04594947005928107	−0.21
random forest regressor	4.397746386815284	6.0230991826817935	0.921774027763598	0.44
random forest regressor with PCA	4.559602412198475	6.5947706237259	0.905268993634512	0.33
XGBoost regressor	5.376145536727195	6.725891467126293	0.34547499328589754	0.30
XGBoost regressor with PCA	5.234020186050841	6.8337022007686	0.35282702546408073	0.28

regression algorithm. The performance of the proposed model is then compared with state of the art regression algorithms. The results obtained prove the superiority of the proposed model. In future, the proposed model can be tested on a huge-dimensional data set. It can also be enhanced by using deep neural networks for prediction.

6 Acknowledgments

The authors would like to thank IET for their continued support for our research objectives.

7 References

[1] Iwendi, C., Alqarni, M.A., Anajemba, J.H., *et al.*: 'Robust navigational control of a two-wheeled self-balancing robot in a sensed environment', *IEEE Access*, 2019, **7**, pp. 82337–82348

[2] Jia, X., He, D., Kumar, N., *et al.*: 'Authenticated key agreement scheme for fog-driven IOT healthcare system', *Wirel. Netw.*, 2019, **25**, (8), pp. 4737–4750

[3] Wu, L., Zhang, Y., Ma, M., *et al.*: 'Certificateless searchable public key authenticated encryption with designated tester for cloud-assisted medical internet of things', *Ann. Telecommun.*, 2019, **74**, (7–8), pp. 423–434

[4] Atzori, L., Iera, A., Morabito, G.: 'The internet of things: A survey', *Comput. Netw.*, 2010, **54**, (15), pp. 2787–2805

[5] Iwendi, C., Maddikunta, P.K.R., Gadekallu, T.R., *et al.*: 'A metaheuristic optimization approach for energy efficiency in the iot networks', Software: Practice and Experience, 2020

[6] Gharami, S., Prabadevi, B., Bhimnath, A.: 'Semantic analysis-internet of things, study of past, present and future of iot', *Electron. Gov., Int. J.*, 2019, **15**, (2), pp. 144–165

[7] Dinculeană, D., Cheng, X.: 'Vulnerabilities and limitations of MQTT protocol used between iot devices', *Appl. Sci.*, 2019, **9**, (5), p. 848

[8] Gao, C., Lv, S., Wei, Y., *et al.*: 'M-sse: an effective searchable symmetric encryption with enhanced security for mobile devices', *IEEE Access*, 2018, **6**, pp. 38860–38869

[9] Wang, C., Zhu, L., Gong, L., *et al.*: 'Accurate sybil attack detection based on fine-grained physical channel information', *Sensors*, 2018, **18**, (3), p. 878

[10] Wang, C., Zhao, Z., Gong, L., *et al.*: 'A distributed anomaly detection system for in-vehicle network using htm', *IEEE Access*, 2018, **6**, pp. 9091–9098

[11] Jiao, L., Wu, Y., Dong, J., *et al.*: 'Toward optimal resource scheduling for internet of things under imperfect csi', *IEEE Internet Things J.*, 2020, **7**, (3), pp. 1572–1581

[12] Yin, B., Wu, Y., Hu, T., *et al.*: 'An efficient collaboration and incentive mechanism for internet of vehicles (iov) with secured information exchange based on blockchains', *IEEE Internet Things J.*, 2020, **7**, (3), pp. 1582–1593

[13] Zhou, R., Zhang, X., Wang, X., *et al.*: 'Privacy-preserving data search with fine-grained dynamic search right management in fog-assisted internet of things', *Inf. Sci.*, 2019, **491**, pp. 251–264. Available at: http://www.sciencedirect.com/science/article/pii/S0020025519302968

[14] Zhao, Z., Min, G., Gao, W., *et al.*: 'Deploying edge computing nodes for large-scale iot: A diversity aware approach', *IEEE Internet Things J.*, 2018, **5**, (5), pp. 3606–3614

[15] Liu, S., Guo, C., Al.Turjman, F., *et al.*: 'Reliability of response region: A novel mechanism in visual tracking by edge computing for iiot environments', *Mech. Syst. Signal Process.* ., 2020, **138**, p. 106537

[16] Fafoutis, X., Elsts, A., Vafeas, A., *et al.*: 'On predicting the battery lifetime of iot devices: experiences from the sphere deployments'. Proc. Seventh Int.Workshop on Real-World Embedded Wireless Systems and Networks, Shenzhen, China, 2018, pp. 7–12

[17] Wu, J., Dong, M., Ota, K., *et al.*: 'Application-aware consensus management for software-defined intelligent blockchain in iot', *IEEE Netw.*, 2020, **34**, (1), pp. 69–75

[18] Lin, X., Li, J., Wu, J., *et al.*: 'Making knowledge tradable in edge-ai enabled iot: a consortium blockchain-based efficient and incentive approach', *IEEE Trans. Ind. Inf.*, 2019, **15**, (12), pp. 6367–6378

[19] RM, S.P., Bhattacharya, S., Maddikunta, P.K.R., *et al.*: 'Load balancing of energy cloud using wind driven and firefly algorithms in internet of everything', *J. Parallel Distrib. Comput.*, 2020, **142**, pp. 16–26

[20] Chen, Y.K.: 'Challenges and opportunities of internet of things'. 17th Asia and South Pacific design automation Conf., Sydney, Australia, 2012, pp. 383–388

[21] Djamaa, B., Witty, R.: 'An efficient service discovery protocol for 6lowpans'. 2013 Science and Information Conf., London, U.K, 2013, pp. 645–652

[22] Shelby, Z., Hartke, K., Bormann, C., *et al.*: 'Constrained application protocol (CoAP), draft-IETF-core-coap-13', *Orlando: Internet Eng. Task Force–IETF*, 2012, **1**, pp. 1–109

[23] Abedin, S.F., Alam, M.G.R., Haw, R., *et al.*: 'A system model for energy efficient green-iot network'. 2015 Int. Conf. on Information Networking (ICOIN), Siem Reap, Cambodia, 2015, pp. 177–182

[24] Reddy, G.T., Reddy, M.P.K., Lakshmanna, K., *et al.*: 'Analysis of dimensionality reduction techniques on big data', *IEEE Access*, 2020, **8**, pp. 54776–54788

[25] Reddy, G.T., Reddy, M.P.K., Lakshmanna, K., *et al.*: 'Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis', *Evol. Intell.*, 2019, **13**, pp. 1–12

[26] Kraemer, F.A., Ammar, D., Braten, A.E., *et al.*: 'Solar energy prediction for constrained iot nodes based on public weather forecasts'. Proc. of the Seventh Int. Conf. on the Internet of Things, Linz, Austria, 2017, p. 2

[27] Uddin, Z., Altaf, M., Bilal, M., *et al.*: 'Amateur drones detection: a machine learning approach utilizing the acoustic signals in the presence of strong interference', *Comput. Commun.*, 2020, **154**, pp. 236–245

[28] Lee, J., Lee, J.: 'Prediction-based energy saving mechanism in 3gpp nb-iot networks', *Sensors*, 2017, **17**, (9), p. 2008

[29] Bag, G., Pang, Z., Johansson, M.E., *et al.*: 'Engineering friendly tool to estimate battery life of a wireless sensor node', *J. Ind. Inf. Integr.*, 2016, **4**, pp. 8–14

[30] Shin, M., Joe, I.: 'Energy management algorithm for solar-powered energy harvesting wireless sensor node for internet of things', *IET Commun.*, 2016, **10**, (12), pp. 1508–1521

[31] Zhang, L., Liu, J., Jiang, H.: 'Energy-efficient location tracking with smartphones for iot'. SENSORS, 2012 IEEE, Taipei, Taiwan, 2012, pp. 1–4

[32] Lu, X., Kim, I.H., Xhafa, A., *et al.*: 'Reaching 10-years of battery life for industrial iot wireless sensor networks'. 2017 Symp. on VLSI Circuits, Kyoto, Japan, 2017, pp. C66–C67

[33] Vo, H.: 'Implementing energy saving techniques for sensor nodes in iot applications', *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, 2018, **5**, (17), pp. 1–7

[34] Janković, S., Saranovac, L.: 'Improving energy usage in energy harvesting wireless sensor nodes using weather forecast'. 2017 25th Telecommunication Forum (TELFOR), Belgrade, Serbia, 2017, pp. 1–4

[35] Shresthamali, S., Kondo, M., Nakamura, H.: 'Adaptive power management in solar energy harvesting sensor node using reinforcement learning', *ACM Trans. Embed. Comput. Syst. (TECS)*, 2017, **16**, (5s), p. 181

[36] Sharma, A., Kakkar, A.: 'Forecasting daily global solar irradiance generation using machine learning', *Renew. Sust. Energy Rev.*, 2018, **82**, pp. 2254–2269

[37] Saha, B., Goebel, K.: 'Uncertainty management for diagnostics and prognostics of batteries using Bayesian techniques'. 2008 IEEE Aerospace Conf., Montana, USA, 2008, pp. 1–8

[38] Hsu, R.C., Liu, C.T., Wang, K.C., *et al.*: 'QoS-aware power management for energy harvesting wireless sensor network utilizing reinforcement learning'. 2009 Int. Conf. on Computational Science and Engineering, Vancouver, Canada, vol. **2**, 2009, pp. 537–542

[39] Liu, C.T., Hsu, R.C.: 'mtDynamic power management utilizing reinforcement learning with fuzzy reward for energy harvesting wireless sensor nodes'. IECON 2011-37th Annual Conf. of the IEEE Industrial Electronics Society, Victoria, Australia, 2011, pp. 2365–2369

[40] Kosunalp, S.: 'A new energy prediction algorithm for energy-harvesting wireless sensor networks with q-learning', *IEEE Access*, 2016, **4**, pp. 5755–5763

[41] Aoudia, F.A., Gautier, M., Berder, O.: 'RLMAN: an energy manager based on reinforcement learning for energy harvesting wireless sensor networks', *IEEE Trans. Green Commun. Netw.*, 2018, **2**, (2), pp. 408–417

[42] Sharma, N., Sharma, P., Irwin, D., *et al.*: 'Predicting solar generation from weather forecasts using machine learning'. 2011 IEEE int. Conf. on smart grid communications (SmartGridComm), Brussels, Belgium, 2011, pp. 528–533

[43] Singh, S., Sharma, P., Moon, S., *et al.*: 'Eh-gc: an efficient and secure architecture of energy harvesting green cloud infrastructure', *Sustainability*, 2017, **9**, (4), p. 673

[44] Min, M., Xiao, L., Chen, Y., *et al.*: 'Learning-based computation offloading for IOT devices with energy harvesting', *IEEE Trans. Veh. Technol.*, 2019, **68**, (2), pp. 1930–1941

[45] Severson, K.A., Attia, P.M., Jin, N., *et al.*: 'Data-driven prediction of battery cycle life before capacity degradation', *Nature Energy*, 2019, **4**, (5), p. 383

© The Institution of Engineering and Technology 2020

[46] Zhang, Y., Xiong, R., He, H., *et al.*: 'Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries', *IEEE Trans. Veh. Technol.*, 2018, **67**, (7), pp. 5695–5705

[47] Fafoutis, X., Sørensen, T., Madsen, J.: 'Energy harvesting-wireless sensor networks for indoors applications using IEEE 802.11', *Procedia Comput. Sci.*, 2014, **32**, pp. 991–996

[48] Nguyen, H.A., Förster, A., Puccinelli, D., *et al.*: 'Sensor node lifetime: an experimental study'. 2011 IEEE Int. Conf. on Pervasive Computing and Communications Workshops (PERCOM Workshops), Seattle, USA, 2011, pp. 202–207

[49] CeArley, D., Burke, B., Searle, S., *et al.*: 'Top 10 strategic technology trends for 2018', *The Top*, 2016, **10**, pp. 1–246

[50] Navarro, H., Bennun, L.: 'Descriptive examples of the limitations of artificial neural networks applied to the analysis of independent stochastic data', arXiv preprint:14045598, 2014

[51] Belk, A., Xu, Z.Z., Carter, D.O., *et al.*: 'Microbiome data accurately predicts the postmortem interval using random forest regression models', *Genes*, 2018, **9**, (2), p. 104

[52] District, C.P.: 'Chicago beach sensor life prediction', 2017. Available at https://www.chicagoparkdistrict.com/parks-facilities/beaches Accessed 12 July 2019

© The Institution of Engineering and Technology 2020