Taheri *et al. J Big Data* (2020) 7:8
https://doi.org/10.1186/s40537-020-0284-2

Journal of Big Data

RESEARCH

Open Access

Learning-based power prediction for geo-distributed Data Centers: weather parameter analysis

Somayyeh Taheri[1,2*], Maziar Goudarzi[1] and Osamu Yoshie[2]

*Correspondence: taheris@ce.sharif.edu
[1] Dept. of Computer Engineering, Sharif University of Technology, Tehran, Iran
Full list of author information is available at the end of the article

Abstract

Nowadays, the fast rate of technological advances, such as cloud computing, has led to the rapid growth of the Data Center (DC) market as well as their power consumption. Therefore, DC power management has become increasingly important. While power forecasting can greatly help DC power management and reduce energy consumption and cost. Power forecasting predicts the potential energy failures or sudden fluctuations in power intake from utility grid. However, it is hard especially when variable renewable energies (RE) as well as free cooling such as air economizers are also used. Geo-distributed DCs face an even harder issue: since in addition to local conditions, the overall status of the entire system of collaborating DCs should also be considered. The conventional approach to forecast power consumption in such complicated cases is to construct a closed form formula for power. This is a tedious task that not only needs expert knowledge of how each single cooling or RE system works, but also needs to be done individually for each DC and repeated all over again for each new DC or change of equipment. One alternative is to use machine learning so as to learn over time how the system consumes power in varying conditions of weather, workload, and internal structure in multiple geo-distributed locations. However, due to the wide range of effective features as well as trade-off between the accuracy and processing overhead, one important issue is to obtain an optimal set of more influential features. In this study, we analyze the correlation among geo-distributed DC power patterns with their weather parameters (based on different DC situations and infrastructure) and extract a set of influential features. Afterward, we apply the obtained features to provide a power consumption forecasting model that predict the power pattern of each collaborating DC in a cloud. Our experimental results show that the proposed prediction model for geo-distributed DCs reaches the accuracy of 87.2%.

Keywords: Cloud computing, Geo-distributed Data Center, Machine learning, Weather-based forecasting

Introduction

Nowadays, in the age of big data and more data generation, there is a growing need to store and process large-scale data in real-time which has led to the deployment of cloud computing. The significant growth of the DC market has led to its rapid growth of power consumption as well as cost. By 2025, the DC market is predicted to account for the largest ICT share of global electricity production by 33% and consumes 20% of the world's energy and its carbon footprint is at 5.5% of the global value [1].

The pervasive performance of the cloud, which provides working anywhere, has led to widespread broadcasting of cloud users, around the world. In addition, considerations of globalization, security, and disaster recovery encourage organizations to distribute their DCs over a long geographically distance and across different regions, clearly near to cloud users. These geo-distributed DCs, which replaced the centralized one, offer solutions to deal with the high velocity and high volume of big data generated from geographically dispersed sources. Hence, deploying multiple geo-distributed DCs is the subject of most current efforts by cloud providers such as IBM, Google, and Facebook.

As the underlying infrastructure of the cloud, DCs have to be active all the time to respond to users' requests. Therefore, they consume a great deal of energy and widely affect the local power grid. In addition, electricity cost has turned to one of the most important portions of their operational expenses. For example, DC power consumption in 2018 has been valued at $21.57 billion and is expected to reach $35.09 billion by 2024 [2].

The two key components in DC design are efficiency and clean energy. Due to the environmental impacts, increasing public awareness, and high energy cost, there is a growing desire today among DC providers to use renewable energy and reduce carbon footprints. In this regard, RE is applied as a supplementary energy source in DC. The green DC market investment in 2018 was valued at $43.24 billion and is expected to reach $147.88 billion, by 2024 [2].

Nevertheless, due to the high level of DC power consumption, power management, which reduces the DC power consumption and cost, has become increasingly important. Power management and tuning essentially require a reliable power modeling or forecasting method. While power forecasting can provide new opportunities for participation in the emerging power market which leads to monetary benefits both for the DC as well as the grid.

Practically, the DC power consumption pattern depends on multiple factors such as hardware specifications, workload, cooling requirements, types of applications, etc., and cannot be measured easily. Thus, precise modeling of a DC's power consumption pattern, either at the whole system or its individual component, is not straightforward. Furthermore, it is impractical to perform detailed measurements of power consumption of all existing components, since the measurement infrastructure introduces overhead to the system. Due to these reasons, power forecasting methods have been developed which can predict the power consumption of a DC for a given workload. However, power forecasting is also hard especially when variable renewable energies (RE), as well as free cooling such as air economizers, are used.

Geo-distributed DCs face an even harder issue: Because the incoming load is balanced among them and hence the power pattern of each of geo-distributed DCs depends on others, and in addition to the local conditions, the overall state of the system also affects the power pattern of a DC. Hence, there is a wide range of influential factors that increasingly complicates the power forecasting for a geo-distributed DCs, so that conventional approaches have become inefficient in such complicated cases.

A typical DC has thousands of sensors which produce millions of data points that record the status of the system at any time. Thus, one alternative is to construct a prediction model to learn how the system consumes power form the variation of these data points over time. However, given the clear trade-off between the accuracy and processing overhead, we need to obtain an optimal set of most influential features. In this regard, one approach is to consider the sectors which consume more energy and examine the influential factors in these sectors to extract a set of more significant features.

In this study, we considered the correlation between weather parameters and power pattern of geo-distributed DCs. We use linear regression (LR) to extract an optimal set of most influential features. Afterward, taking advantages of Artificial Neural Network (ANN), we propose a reliable power prediction model for a geo-distributed DC.

The rest of the paper is organized as follows: We start with a background in power distribution of a DC and overview of related works. "Methodology" section discusses the methodology and gives insight into prediction steps and the proposed prediction model. "Simulation setup" section presents the experimental setup and an overview of the proposed approach. Then "Result and discussion" section provides and combines the obtained results and discussion. Finally "Conclusion and future works" section concludes the paper.

Background and related work

Power distribution among geo-distributed DCs

In this section, we first explore the power distribution of a DC as well as the load distribution considerations among geo-distributed ones.

Among different proposed standards, power usage effectiveness (PUE) is widely used by load balancing approaches to distribute the incoming workload among geo-distributed DCs. PUE is the ratio of the DC total power vs. the amount delivered to the computing equipment. Figure 1 depicts the power distribution of a DC and PUE calculation in more details.

As Fig. 1 shows, among different secondary supports, the cooling unit is responsible for a large portion of power consumption and hence, has a significant impact on PUE; as far as up to 40% of a DC power consumption is consumed by cooling system

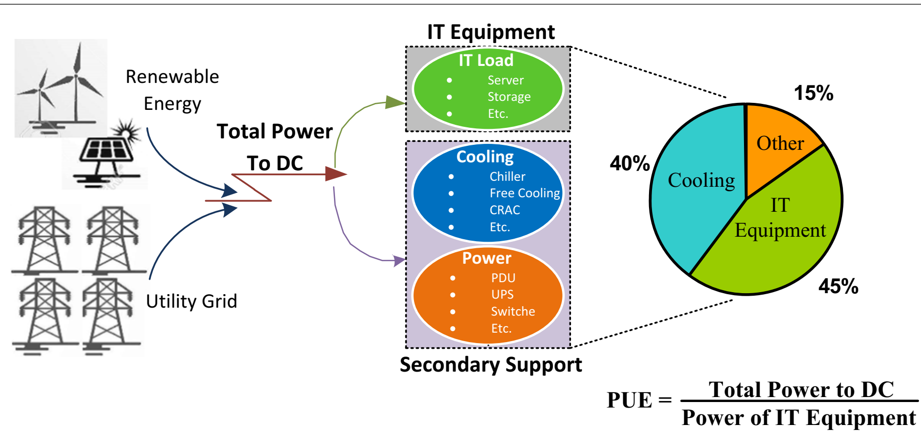

Fig. 1 Overview of DC power distribution

[3]. Recently, utilizing free cooling methods, which takes advantage of the outside cold weather have increasingly considered in DC design. While, according to [4], such this condition, weather parameters such as ambient temperature significantly affect the PUE.

In addition, to reduce the energy expense, most DC owners have persuaded to install RE next to their DC. While weather condition could significantly affects the efficiency as well as the production capacity of the in-site RE and consequently, the DC power pattern. Thus, using variable energy and free cooling, the DC power pattern is associated with weather parameters. Therefore, given the geographical load balancing (GLB) policy among collaborating geo-distributed DCs, the variation of weather condition of each DC will change the GLB decisions and consequently, all DCs power pattern.

DC power prediction

The large amount of data points generated by thousands of sensors inside a typical DC, and the complex interactions between them, makes it difficult to use traditional approaches for an accurate power forecasting. While, machine learning creates new opportunities for power prediction in a DC. However, regarding the recent advances in machine learning, a few studies focused on utilizing them to predict the DC power requirement. Authors in [5] use historical power data to predict the future power consumption. They consider two different time scale for data acquisition and accordingly propose two different prediction models for DC power consumption that is based on the linear recursive Auto Encoder. Li et al. [6] also reinforce their proposed model with an additional layer which corrects the prediction results with respect to the prediction error analysis. Investigates the relation between DC power consumption and weather condition. Authors analyze the correlation between different weather parameters and the amount of a single DC power consumption. They extract more effective features and utilize them to propose a linear regression (LR) based prediction model for power consumption of a typical single DC. Hsu et al. [7] consider the power consumption modeling based on a wide range of input variables and propose a prediction framework using the self-aware computing. The proposed framework dynamically updates the effective training variables according to the input variables interaction analysis as well as the prediction error monitoring. Authors evaluate their framework in a real DC and use it along with the energy management system. Liu et al. [8] estimates the DC power consumption based on the server power consumption modeling. Authors utilize historical data for CPU as well as memory usage in a DC and apply the neural network model to predict the near future values. Afterward, according to the DC power consumption modeling in their work, they estimate the required power and accordingly explore the opportunity of ancillary energy market participation in the DC.

To the best of our knowledge, the existing studies neither pay attention to the power prediction in the geo-distributed DC nor take into account the intricate and interdependent impact of local conditions in global power distribution pattern of a geo-distributed cloud. This work is the first research which considers these issues and analyzes the impact of local weather parameters on geo-distributed DCs power pattern.

To sum up, the contributions of this work are as follows:

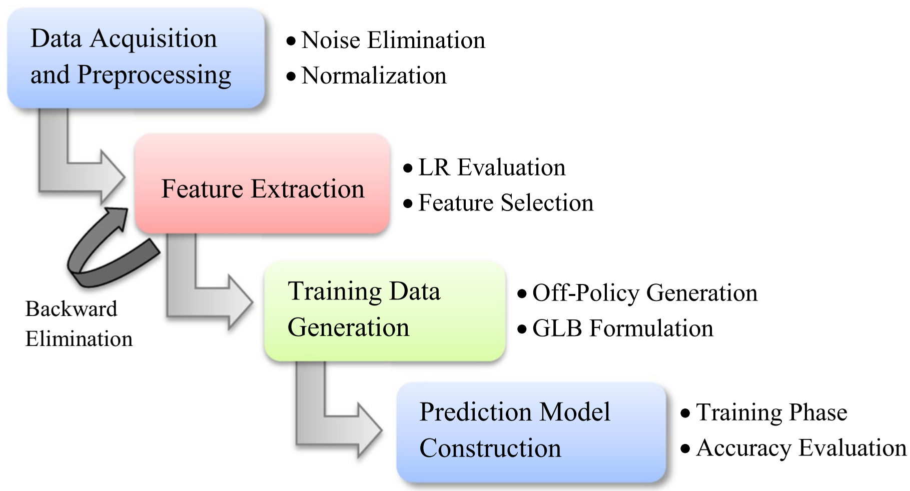

Fig. 2 Prediction steps

- Given the wide range of effective factors in DC power consumption in a geo-distributed cloud, we applied an analyzing algorithm for significant feature extraction to obtain a set of more conducive features, wide enough to reach the acceptable accuracy, yet narrow enough to prevent the unnecessary complexity.
- We formulated a well-known GLB problem and applied its optimization results as training data.
- Instead of using historical power data, we calculated the DC power consumption based on the accurate model of real-time PUE as well as server power consumption model.

Methodology

The prediction procedure includes the historical data acquisition, original data pretreatment for noise elimination, feature extraction, training data generation, and finally prediction model construction. The prediction steps is shown in Fig. 2.

Data acquisition and noise elimination

The historical weather data is acquired from the reliable and high-quality local databases [9–11] through the automatic python script and stored in comma-separated value (.csv) local file. Afterward, we did preprocess to eliminate noise and retrieve the missing data through the linear interpolation. Finally, given the wide range of raw data, we normalized them (feature scaling) to the interval of [1, 0] through *mapminmax()* to provide an easy evaluation as well as comparison.

Most significant feature extraction

Considering the clear trade-off between the accuracy and processing overhead, we eliminate the redundant features which do not effectively contribute to the model training and may lead to over fitting. We utilize LR to evaluate the impact of each feature on the prediction result, because LR has fast computing speed as well as prediction reliability that reaches the result after the finite number of adjustment. Algorithm 1 illustrates the procedure of most significant feature extraction as we considered.

Algorithm 1 Backward Elimination

```
Input:  F^0                            //Initial set of features
        F^1 = {v_0, v_1, ...v_{N-1}}   //A training set with N input features
        R = 0                          //A subset of less significant features
        i = 1                          //Index
Output: F^Best = 0                     //An optimal subset
01: begin
02: Initial  F^Best = F^1
03: γ^best = eval(F^1) //evaluate F^1 with LR
04:  While F^i != F^(i-1)
05:   for each v ∈ F^i do
06:     F' = F^i/v //Eliminate features one by one
07:     γ = eval(F')
08:     if (γ is the same or better than γ^best )
09:       γ^best = γ
10:       R = R U v //Remove Redundant Features
11:     end
12:   end
13:   F^(i+1) = F^i/R
14:   R = 0
15:   i++
16:  end
17: return F^best
18: end
```

Given the number of loop execution in Algorithm 1 (for loop, Line 5), our problem with five input features (including temperature, humidity, pressure, cloudiness and wind speed) has 31 different LR models for each DC. We consider a cloud of four geo-distributed DCs and thus a total of 124 different LR models must be considered. In this regard, to quickly extract the most important features, we first used the *ranking approach* to filter out the irrelevant features and then applied the Algorithm 1 to the remaining three high ranked features. The basis of ranking techniques is on the statistical scores that determine the correlation of features with the outcome variable. Then, the greedy search based on Algorithm 1 creates subsequent models with the left features iteratively and evaluates the overall fit of each model in each iteration for evaluation function (line 7). We applied *Adjusted R-squared* as the γ parameter and evaluate each model considering the model complexity penalty, to control its over fitting. Please note that adjusted R-squared is between 0 and 1 and higher values mean more variance is explained by the model. Table 1 provides the result of ranking approach in each DC.

According to Table 1, three high ranked features in DC1, DC3 and DC4 are the same. Hence, the subsequent models are also the same. Tables 2 and 3 show the result of subsequent models evaluation, based on Algorithm 1, considering the evaluation parameter of adjusted R-squared.

Table 1 Feature ranking result

	Feature ranking
DC1	{Temperature, cloudiness, windspeed, humidity, pressure}
DC2	{Temperature, windspeed, pressure, cloudiness, humidity}
DC3	{Temperature, windspeed, cloudiness, pressure, humidity}
DC4	{Temperature, cloudiness, windspeed, pressure, humidity}

Table 2 R-squared result for subsequent models in DC1, DC3 and DC4

Feature subset of the model	DC1	DC3	DC4
{Temperature, cloudiness, windspeed}	0.799	0.779	*0.838*
{Cloudiness, windspeed}	0.766	0.743	0.535
{Temperature, windspeed}	0.791	*0.779*	0.570
{Temperature, cloudiness}	*0.799*	0.753	0.708
{Temperature}	0.747	0.742	0.658
{Cloudiness}	0.673	0.248	0.535
{Wind speed}	0.416	0.723	0.417

Table 3 R-squared result for subsequent models in DC2

Feature subset of the model	DC2
{Temperature, windspeed, pressure}	0.773
{Wind speed, pressure}	0.45
{Temperature, windspeed}	0. 739
{Temperature, pressure}	0.724
{Temperature}	*0.773*
{Pressure}	0.33
{Wind speed}	0.449

Table 4 Result of identifying most significant feature in DCs

DC	Selected features
DC1	{Temperature, cloudiness}
DC2	{Temperature}
DC3	{Temperature, windspeed}
DC4	{Temperature, cloudiness, windspeed}

The best outcome of R-squared evaluation parameter in the subsequent models are shown in italic, in Tables 2 and 3. Hence, the optimal set of most significant features in each DC would be as Table 4.

Training data generation

In this study, we generated the training data through off-line data generation. This learning technique is useful when the number of training samples is small enough that all training samples can be stored in memory. The policy of training data generation in off-line algorithms is generally independent of the policy applied for predicted data. While on-line method directly uses the predicted values as the control values concurrently.

For this, we formulated power-aware GLB approach as an off-policy control algorithm to generate the historical data of DCs power consumption, as follows:

Power-aware GLB formulation

We modeled the IT power consumption of each DC in time slot t($P_{IT_j,t}$) based on its server type and according to the accurate server power consumption modeling given the CPU as well as memory usage of assigned workload as follows:

$$P_{IT_j,t} = \sum_{i=1}^{|W_t|} X_{i,j}(u_i(p_{activej} - p_{idlej}) + p_{idlej}). \tag{1}$$

where $|W_t|$ is the number of incoming workload during time slot t, u_i is CPU and memory utilization requirement for $workload_i$, $p_{activej}, p_{idlej}$ are respectively the amount of active and idle power consumption of server in DCj, and $X_{i,j}$ is a boolean parameter which represents the assignment of $workload_i$ to DCj:

$$X_{i,j} = \begin{cases} 1 & \text{if } workload_i \text{ is assigned to DCj.} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Clearly $X_{i,j}$ is set through the power-aware GLB problem optimization. The optimization goal is to minimize the total power consumption of incoming workloads by balancing them among DCs, while the power efficiency of each DC is a function of its weather condition. The problem statement and formulation are as follows:

Problem statement: In a distributed cloud with $|L|$ DCs, each with a load-dependent real-time PUE ($PUE_{j,t}$) and available RE ($RE_{j,t}$), distribute the incoming load in a way that minimizes the total power consumption ($Power_t$):

Minimize:

$$Power_t = \sum_{j=1}^{|L|} P_{IT_j,t} PUE_{j,t} - RE_{j,t}. \tag{3}$$

As already mentioned, due to using free cooling in DCs, ($PUE_{j,t}$) is widely affected by weather conditions. Moreover, the amount of available RE, which covers part of the DC power requirement, is also affected by the weather condition.

We formulated the above optimization problem and used the BARON solver through the General Algebraic Modeling System (GAMS) 24.6 to solve it. The optimization problem results are then used as the training data for the prediction model.

Prediction model construction

We developed a neural network framework that learns from actual operations data to predict the power consumption of geo-distributed DCs. We utilized the Artificial Neural Network (ANN) prediction model as Fig. 3.

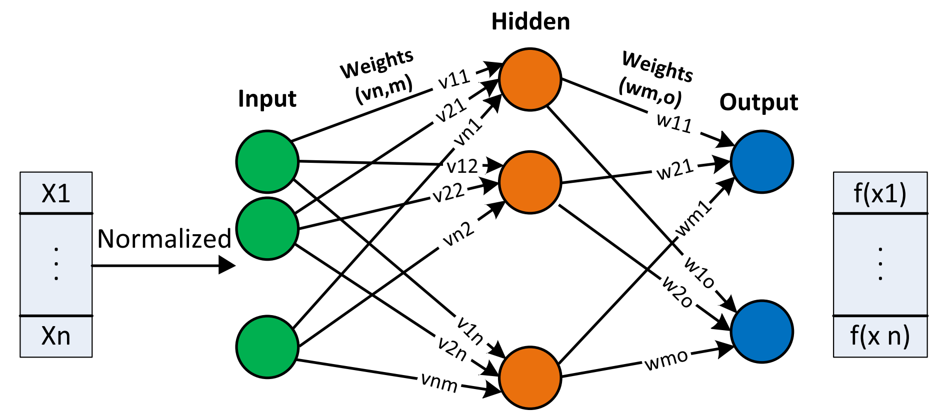

Fig. 3 Overview of a typical Artificial Neural Network

We adopted the feed forward back propagation ANN and used Smooth Rectified Linear Unit (ReLU) as the non-linear activation function to regularize the inner product of the input and weight vectors of the neuron. The Smooth ReLU is defined as (4):

$$f(x) = \begin{cases} x & \text{if input layer.} \\ \frac{e^x}{1+e^x} & \text{otherwise.} \end{cases} \tag{4}$$

We also considered Root Mean Squared Error (RMSE) as well as Mean Absolute Percentage Error (MAPE) to evaluate the model accuracy over unseen patterns as (5), (7):

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(f(x_i) - y_i)^2}. \tag{5}$$

$$MAPE = \frac{100}{M}\sum_{i=1}^{M}\frac{|f(x_i) - y_i)|}{y_i}. \tag{6}$$

where $f(x_i)$ and y_i are the predicted and actual values respectively and M is number of training data. We considered the accuracy percentage as follows [12]:

$$Accuracy(\%) = 1 - MAPE(\%). \tag{7}$$

Finally, to maximize the accuracy, during the training process we update the weights of the interconnected neurons through back propagation. We used an adaptive learning rate optimization algorithm namely the Adam optimization, in response to the back propagated errors [13].

Obviously, the computation complexity of our model is low since (i) the number of neurons in the hidden layer is small and (ii) the frequency of the prediction execution is low.

Table 5 Cloud specification

Location	RE equipment	Server type	Server number	Server spec.	
				P_{active}(kW)	P_{idle}(kW)
Manitoba	Solar	Intel E5506	1500	0.419	0.146
Quebec	None	Intel X5570	1000	0.352	0.153
Minnesota	Wind	AMD EPYC 7601	2000	0.483	0.138
Ontario	Solar + wind	Intel E5-2699	1500	0.529	0.102

Simulation setup

We considered a cloud with 4 geo-distributed DCs located at Manitoba, Quebec, Ontario and Minnesota namely DC1, DC2, DC3 and DC4, respectively. To keep the comprehensive observation, we also equipped our simulated DCs with different kinds of RE. The characteristics of our DCs including the server type and specification as well as the number of servers is shown in Table 5.

To calculate the available RE during GLB formulation, we used the PVWatts calculator [14] for available solar power and the proposed model in [15] for wind power. For solar power, we assumed a fixed photovoltaic array with standard module type facing south at 20-degree tilt angle, 1.1 dc (direct current) to ac (alternative current) size ratio and 96% inverter efficiency. We also considered the diameter of the wind turbine blade 12 meters and normalized the available RE in each DC (except DC2 with no RE) with respect to its maximum IT power requirement. We also captured data from [16] for real-time PUE and scaled it based on our DCs characteristics.

For workload modeling, we employed Google cluster workload through a period of two weeks from May 1–14 available in [11]. We set each time slot to 15 min and extracted the arrival rate as well as CPU and memory utilization requirement over every time slot. Finally, we normalized it with respect to the total amount of available servers in our cloud.

For the prediction model, we considered a model with 9 neurons in the input layer (given the total number of selected features as Table 4 plus the workload information), 40 neurons in the hidden layer (which set through trial and error) and 4 neurons in the output layer (each for power prediction of a DC). Figure 4 represents the overall structure of our model.

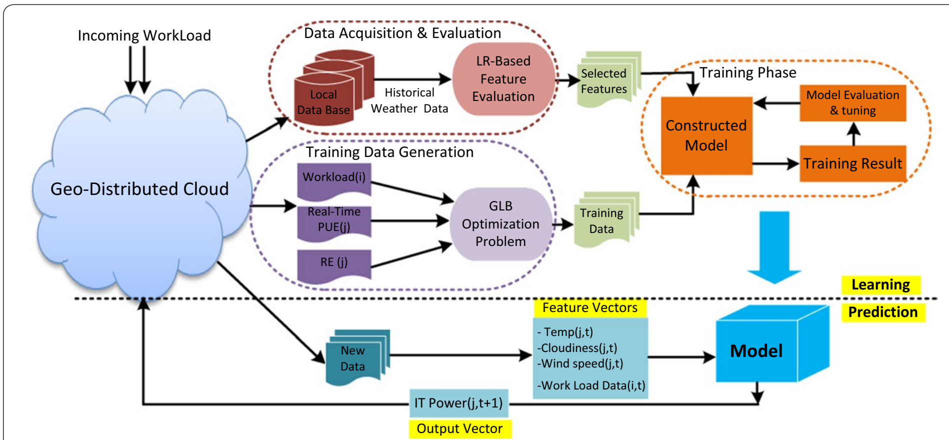

Fig. 4 Overall structure of the proposed model

Table 6 Input variables of prediction model

Parameter	Input variable	Repository	Selected data
U_i	Workload i utilization	Google Cluster [11]	CPU and RAM util. request
$Temp_j$ °C	DCj ambient temp.	Local Data Base [9]	Temperature
$Cloudiness_j$ (Okta)	DCj cloudiness	Local Data Base [10]	Cloud-cover-8
$Windspeed_j$ (km/h)	DCj wind speed	Local Data Base [9]	Wind-speed

Table 6 summarizes the input vectors as well as their repositories.

In our ANN model, we initialized the neural network weights with the uniform distribution between $[1, -1]$, to limit the error and also the formation of the unstable equilibrium. While, due to the error backward propagation through the hidden layers, the identical parameters is necessary.

Finally, our training data set includes 12096 input data through 1344-time samples at 15-min resolution (14 days of operational data). We used 80% of the total data for training and the rest was used as test data.

Result and discussion

System evaluation

We begin our evaluation by considering the DCs' power variation over the consecutive time slots in different approaches of GLB policy. We considered three different approaches as:

- *Power-aware GLB* where the incoming workloads are distributed among DCs considering the amount of RE availability as well as real-time PUE variation (as (3)).
- *Real-time PUE-aware GLB*, where the RE availability is not considered (e.g. due to prediction problems) but real-time PUE is calculated and considered in DCs (as (8)).

$$Minimize: \quad Power_t = \sum_{j=1}^{|L|} P_{IT_j,t} PUE_{j,t}. \tag{8}$$

- *Baseline* where DCs have a given PUE and the RE availability is not considered during GLB procedure (as (9)).

$$Minimize: \quad Power_t = \sum_{j=1}^{|L|} P_{IT_j,t} PUE_j. \tag{9}$$

Figure 5 compares the power consumption of the DCs under these different approaches.

According to Fig. 5 and as we expected, the DCs power consumption is essentially a function of the considerations of the GLB algorithm. However, given the real-time state of the DCs, the more details are taken into account, the fluctuation in the GLB results (and its sensitivity) will be greater. In such this condition, DC's power consumption does not indicate a clear pattern of variations over the consecutive time slots, which is clearly due to a wide range of influencing factors. As a result, the DC power prediction would be more challenging and requires an in-depth analysis of input variables. While,

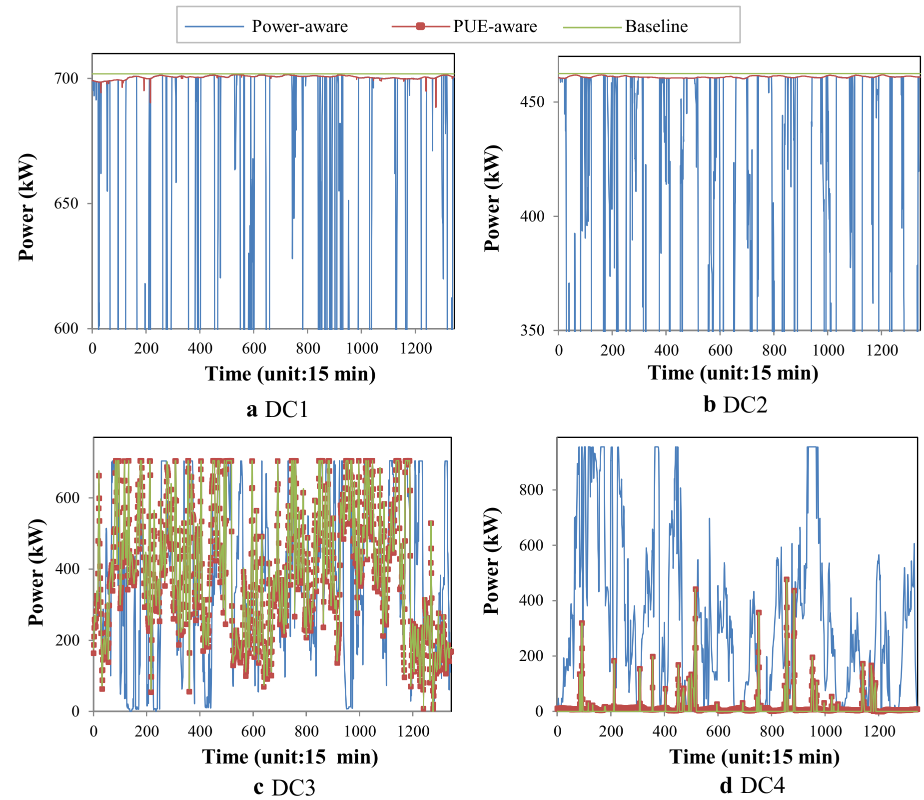

Fig. 5 DCs power consumption pattern in different GLB approaches

in simple GLB approach like baseline, the influencing factors are limited to given data e.g. server type, hardware efficiency, fixed PUE and etc.

Figure 6 depicts the real-time PUE variation among the DCs and over time.

Figure 6 shows that due to the continuous changes in the DC status (e.g ambient temperature, humidity and so on) there is a considerable fluctuation in real-time PUE. At the same time, the wider variations among the real-time DCs status causes the greater fluctuation between their PUE. As a result, this affects GLB decisions and extends the fluctuations of the DCs power pattern over time.

Best-fitted LR model with optimal set of features

In this section, we evaluate the reliability of the feature extraction approach, as we considered in this paper. As already mentioned, we applied LR for significant features extraction and used adjustment R-square as the evaluation parameter among different LR subsequent models. Figure 7 shows the result of DC's power prediction based on the best-fitted real LR equation through the optimal set of most significant features, compared with the DC's actual power data.

We used LR because of its simplicity as well as fast convergence. However, Fig. 7 shows that the output function for each DC is reliable enough, yet we are able to control the processing overhead.

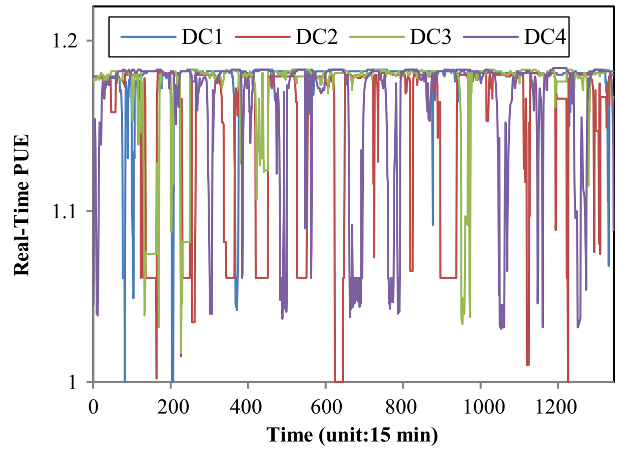

Fig. 6 Variation of Real-Time PUE over the time

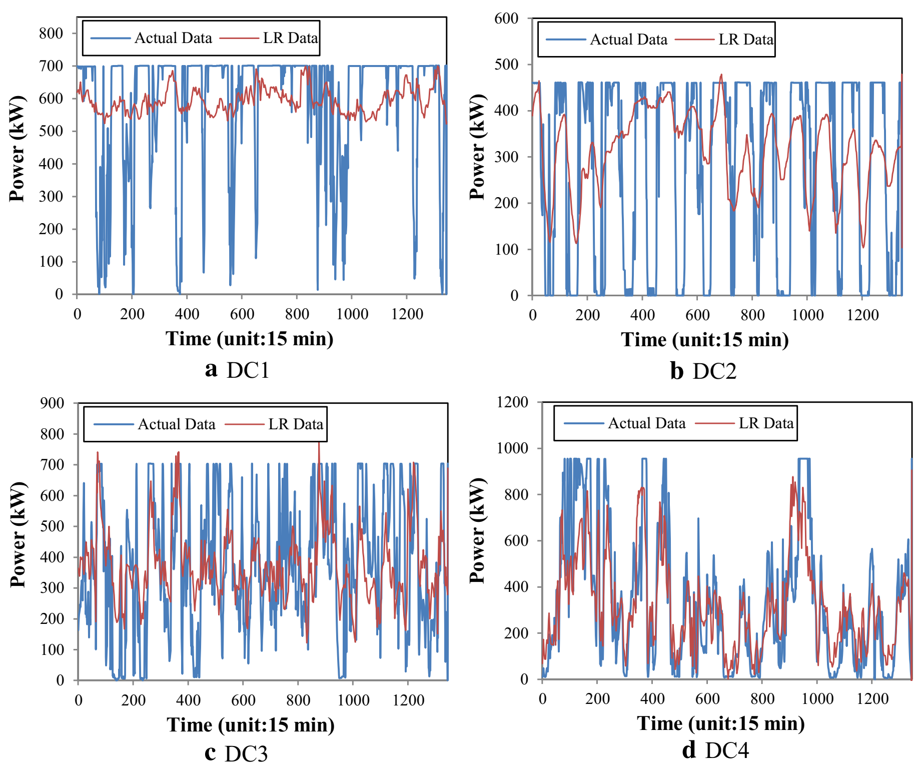

Fig. 7 Comparison of the best-fitted LR model with the actual power pattern

Prediction model evaluation

In this section, we evaluate the accuracy of the prediction model. We applied two evaluation metrics including RMSE as well as MAPE, as mentioned before. During the prediction procedure, we fed the prediction model with the training data and applied some test data to verify the prediction accuracy. Table 7 shows the RMSE as well as MAPE as a function of the training data size vs. the total data.

Table 7 Prediction model evaluation

Metric	Training/total data			
	90%	**80%**	**70%**	**60%**
RMSE	62.2	65.2	80.1	95.7
MAPE	22.8	25.3	28.2	29.5

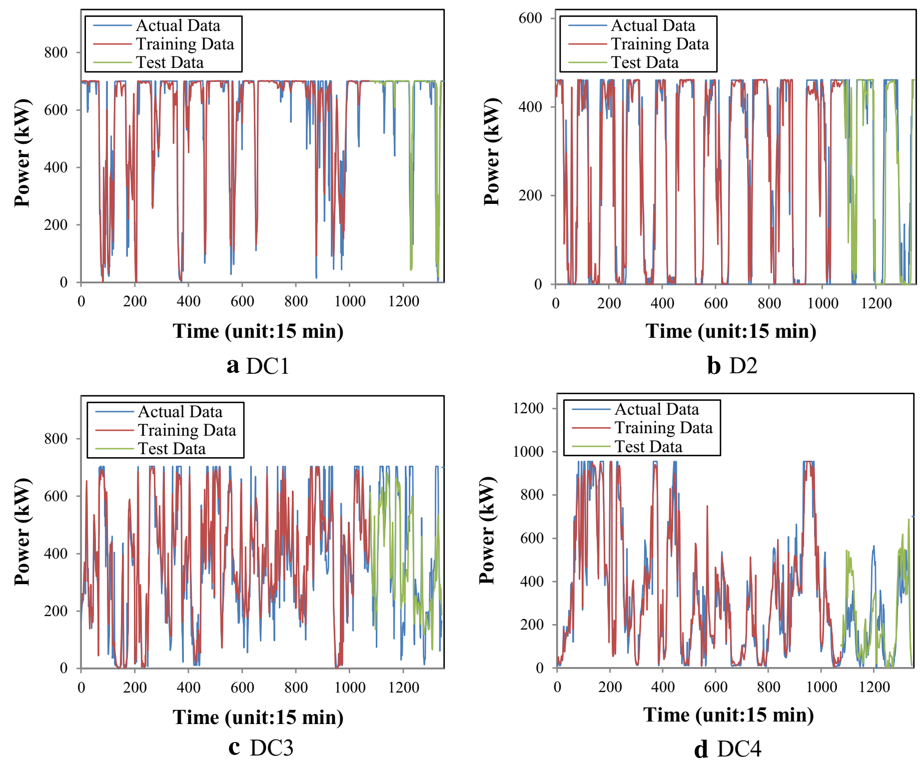

Fig. 8 Prediction model accuracy comparison

As the results show, although for all training data size the prediction result is reliable, however, the larger training data size increases the prediction accuracy. Figure 8 compares the accuracy of our prediction model during the training and test procedure for 80% training data size. We run the 800 epochs with 498s for model construction on Intel(R) Pentium G2120 processor with 4GB RAM.

As Fig. 8 shows, our proposed prediction model is able to predict the challenging problem of geo-distributed DCs power consumption. Moreover, our policy for effective feature extraction (Algorithm 1) is robust enough to achieve an optimal set of most influential features; While examining them reaches a prediction model with the acceptable accuracy as well as the computation overhead.

Conclusion and future works

In this paper, we explored the importance as well the complexity of power pattern prediction in geo-distributed DCs.

Regarding the current energy crisis as well as the share of DCs, power prediction can create new effective ways for energy management and cost reduction, which is the subject of most recent efforts of both the DC owners and the utility grid managers. However, due to the intricate mesh of interdependent input variables, the power prediction for geo-distributed DCs is challenging.

In this study, we used an LR-based feature extraction approach to evaluate the significance of input features. Our algorithm obtains an optimal set of most significant features, while it is fast and has low processing overhead. According to the obtained result in "Best-fitted LR model with optimal set of features" section, the influential feature from DC to DC is different and is a function of DC design considerations.

Afterward, we used ANN to propose an accurate prediction model based on the selected features. Generally, feed-forward ANN provides an important capability which makes it useful for time series forecasting, as follows [17]:

- Robustness to noise, whether in input data or in the mapping function which leads to learning and prediction support even in the presence of missing values.
- Nonlinearity which is a result of the absence of any strong assumptions about the mapping functions and readily learns linear and nonlinear relationships.
- Multivariate inputs and multi-step that provides an arbitrary number of input features and output values to be specified, as well as direct support for multivariate and multi-step forecasting.

Regarding the above capabilities, our proposed ANN-based prediction model reached the accuracy of 87.2%. Nevertheless, investigating other powerful types of neural network methods designed for time series prediction, including Recurrent Neural Networks (RNN) as well as Long Short-Term Memory (LSTM) network which can successfully learn large architectures, is a top priority in our future work.

We are also going to investigate a wider range of historical data. Moreover, leveraging the result of this study for proposing new approaches of power management in geo-distributed DC as well as creating new opportunities for participating in ancillary energy markets are among the goals that we left for future work.

Abbreviations

ANN: Artificial Neural Network; DC: Data Center; GLB: geographical load balancing; LR: linear regression; LSTM: Long Short-Term Memory; PUE: power usage effectiveness; RE: renewable energy; RNN: Recurrent Neural Networks.

Acknowledgements

The authors thanks reviewers for valuable feedback on an early draft of this manuscript.

Authors' contributions

ST performed the primary work of this manuscript. Her contributions include the original idea, literature review, implementation, designing the experiments and initial drafting of the article. MG reviewed the early draft and discussed the final results as well as improved the final manuscript. MG and OY supervised the research. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials
The data used in this paper are publicly online available at [9, 10, 16, 14, 11]. The link for the same is mentioned in "References" section. Moreover, the weather-based raw data files for each DC, as well as their available solar power, as the PVWatts Calculator report, has been uploaded.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Dept. of Computer Engineering, Sharif University of Technology, Tehran, Iran. [2] Graduate School of Information, Production and System, Waseda University, Kitakyushu, Japan.

Received: 18 July 2019 Accepted: 12 January 2020
Published online: 29 January 2020

References
1. Data Center Dynamic. The future of data centers is green. https://www.datacenterdynamics.com/opinions/future-data-centers-green/. 2019.
2. Globe Newswire: Research and Market: Green Data Center Market—growth, trends, and forecast (2019–2024). https://www.globenewswire.com/news-release/2019/06/28/1875927/0/en/Green-Data-Centers-Worldwide-Market-Insight-Report-2019-2024/Power-Segment-to-Hold-a-Significant-Market-Share.html. 2019.
3. Shi Y, Xu B, Wang D, Zhang B. Leveraging energy storage to optimize data center electricity cost in emerging power markets. In: 7th international conference on future energy systems; 2016.
4. Guler H, Cambazoglu B, Ozkasap O. Cutting down the energy cost of geographically distributed cloud DCs. Energy efficiency in large scale distributed systems. Berlin: Springer; 2013.
5. Li Y, Wen Y, Zhang J. Learning-based power prediction for data centre operations via deep neural networks. In: 5th international workshop on energy efficient data centres; 2016.
6. Smpokos G, Elshatshat MA, Iliopoulos I. On the energy consumption forecasting of data centers based on weather conditions: remote sensing and machine learning approach. In: 11th international symposium on communication systems, networks and digital signal processing (CSNDSP); 2018.
7. Hsu YF, Matsuda K, Matsuoka M. Self-aware workload forecasting in data center power prediction. In: 18th IEEE/ACM international symposium on cluster, cloud and grid computing (CCGRID); 2018.
8. Liu N, Lin X, Wang Y. Data center power management for regulation service using neural network-based power prediction. In: 18th international symposium on quality electronic design (ISQED); 2017.
9. Canada historical climate data. http://climate.weather.gc.ca/. 2019.
10. Canada weather stats. http://www.weatherstats.ca/. 2019.
11. Google cluster data. https://github.com/google/cluster-data. 2018.
12. Tofallis C. A better measure of relative prediction accuracy for model selection and model estimation. J Oper Res Soc. 2015;66:1352–62.
13. Kingma DP, Adam JB. A method for stochastic optimization. In: 3rd international conference for learning representations; 2015.
14. PVWatts calculator. http://pvwatts.nrel.gov/pvwatts. 2019.
15. Li C, Qouneh A, Li T. ISwitch: coordinating and optimizing renewable energy powered server clusters. In: 39th annual international symposium on computer architecture (ISCA); 2012. p. 512–23.
16. Facebook public dashboard for real-time PUE. https://www.fbpuewue.com/prineville. 2019.
17. Peter Zhang G. Neural networks for time-series forecasting. Handbook of natural computing. Berlin: Springer; 2012. p. 461–77.

Publisher's Note