Available online at www.sciencedirect.com
ScienceDirect
Energy Reports 8 (2022) 423–429
www.elsevier.com/locate/egyr

The 8th International Conference on Energy and Environment Research ICEER 2021, 13–17 September

Selection of features in reinforcement learning applied to energy consumption forecast in buildings according to different contexts

D. Ramos[a], P. Faria[a,*], L. Gomes[a], P. Campos[b], Z. Vale[a]

[a] *Polytechnic of Porto (P. Porto), R. Dr. Antonio Bernardino de Almeida 431, 4249-015 Porto, Portugal*
[b] *School of Economics and Management, University of Porto (FEP), R. Dr. Roberto Frias S/N, 4200-464 Porto, Portugal*

Received 31 December 2021; accepted 10 January 2022
Available online 3 February 2022

Abstract

The management of buildings responsible for the energy storage and control can be optimized with the support of forecasting techniques. These are essential on the finding of load consumption patterns being these last involved in decisions that analyze which forecasting technique results in more accurate predictions in each context. This paper considers two forecasting methods known as artificial neural network and k-nearest neighbor involved in the prediction of consumption of a building composed by devices recording consumption and sensors data. The forecasts are performed in five minutes periods with the forecasting technique taken into account as a potential to improve the accuracy of predictions. The decision making considers the Multi-armed Bandit in reinforcement learning context to find the best suitable algorithm in each five minutes period thus improving the predictions accuracy in forecasting. The reinforcement learning has been tested in upper confidence bound and greedy algorithms with several exploration alternatives. In the case-study, three contexts have been analyzed.

Peer-review under responsibility of the scientific committee of the 8th International Conference on Energy and Environment Research, ICEER, 2021.

Keywords: Energy management; Learning; Load forecast; Multi-armed Bandit

1. Introduction

The difficulties to balance the consumer needs and the price involved in the electrical energy supply [1] can be overcome with demand response [2]. This optimization can be guaranteed with other factors including forecasting methods [3] and user behavior modeling and learning approaches [4]. Decreases in the costs of power systems operations can be enhanced with the reinforcement deterministic applications in forecasting tasks [5]. A reinforcement application considers an electric vehicle charging station for forecasting tasks as seen in [6] using the Q-learning technique to learn different charging scenarios thus enhancing the accuracy of the forecasting algorithms. A model based RL agent performs schedule tasks optimizations as seen in [7]. Reinforcement learning

* Corresponding author.
E-mail address: pnf@isep.ipp.pt (P. Faria).

https://doi.org/10.1016/j.egyr.2022.01.047

Peer-review under responsibility of the scientific committee of the 8th International Conference on Energy and Environment Research, ICEER, 2021.

is highlighted in the power request application of urban dense areas as seen in [8] to enhance the efficiency of response to consumption fluctuations for the power requests. This is a demand response application integrated in smart grids area that keeps the balance between the satisfaction level and the fluctuation issues. The Q learning algorithm is integrated in emotional applications according to [9] that improves the deep neural network control performance with smaller voltage deviations according to automatic tuning. As established in [10] high performances for power systems require anticipated price predictions, thus a multi-agent reinforcement learning supports the decision making of an artificial neural network to optimize decisions concerning different home appliances in decentralized manner. Reinforcement learning overcomes thermal profile uncertainties in [11] to anticipate thermal energy variations. Forecasting applications concerning office locations provide high relevance to reinforcement learning techniques as mentioned in [12] to deal with nonlinear and complex issues with the benefit of decreasing the forecast errors. Reinforcement learning is adequate for incentive demand response programs seen in [13] adding the reward notion as end users incentives thus decreasing the electricity demand on peak periods. Another role where reinforcement learning plays a relevant role consists in the optimization of energy distribution where it is necessary to balance the energy supply and storage [14]. A similar microgrid application mentioned in [15] of the storage of a photovoltaic system deals with non-linear storage charging and discharging and non-stationary environment. All these applications have the expertise of working on forecasting and learning approaches concerned with the electricity consumption in buildings. However, there is evidence lack to use learning approaches to select the most reliable forecasting algorithm in an office building that monitors and records electricity consumption which is the main goal of this paper. This paper uses the multi armed bandit learning algorithm in order to identify the most reliable forecasting algorithm in different contexts. Evidencing all aspects of the introduction, Section 2 proceeds to present the methodology, Section 3 evidence the case study and results, Section 4 finally presents all the conclusions.

2. Methodology

This section uncovers the several steps of the proposed methodology involved in the process of selecting the most suitable forecasting algorithm in each period according to sensor data. Building offices equipped with electronic devices provide the monitored sensor data involving different measures. Forecasting tasks take into account the historic data to perform prediction and the targets to measure predictions' accuracy. Fig. 1 illustrates the complete process.

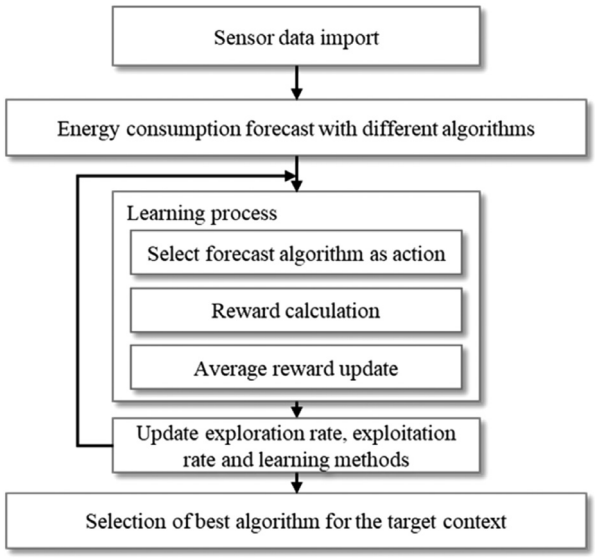

Fig. 1. Proposed methodology.

These tasks take into account the several forecasting algorithms in a reinforcement learning process that evaluates which forecasting algorithm is more accurate in specific periods. The full process is divided in five main steps including sensors data import, energy consumption forecast with different algorithms, learning process,

parameterizations updates involved in reinforcement learning and selection of best algorithm for the target context as illustrated in Fig. 1.

The first step involved in the methodology consists in the import of sensor data (including consumption metering) from electronic devices placed in office buildings. The importation places the data in a structure adaptable by later reinforcement learning tasks. Data is organized according to a time series used to support forecasting tasks. The following step consists in the triggering of forecast of energy consumption according to two different algorithms: artificial neural networks and k-nearest neighbors. The following step consists of a reinforcement learning process used for the decision making that considers either to use artificial neural networks or k-nearest neighbors as the forecasting technique. This learning process considers which forecasting technique to choose according to the forecasting period in question and reward calculations which consider how pragmatic the forecasting technique selection was performed for the period in question. The reward calculation assigns the value 1 to the forecasting algorithm with lower error and the value 0 to the other. This calculation is added to an average reward which look for the reward progress in an accumulated amount of periods. A preliminary application shown in [4] selects rooms of a building where it is intended to improve the accumulated rewards by selecting rooms with better user preferences. Following the reinforcement learning for the period in question, several parameterizations involved in this learning are updated including the exploration and exploitation rates and the learning methods. The exploration rate researches previously unexplored territory for the forecasting algorithms selection while the exploitation rate explores the knowledge involved in a particular forecasting algorithm selection. After the reinforcement learning the best suitable forecasting algorithm for all periods is provided in the last step.

3. Case study and results

This section is structured in two sub-sections, namely 3.1 Case study, and 3.2 Results.

3.1. Case study

The building responsible for the energy control and storage has access to samples of data described by consumption observations while considering weekly patterns separated in five minutes periods. These observations are compared to a forecasting week in five minutes contexts from 18 to 24 November 2019 as seen in Fig. 2. The case study researches the consumption profile in three different scenarios in order to check more accurately the consumption progress according to the different aspects featuring sequences of five minutes. Therefore these three scenarios are classified in “morning”, “afternoon” and “night”. Morning takes sequences of five minutes in periods between 9AM and 12PM, afternoon keeps a sequence of periods from 1PM to 6PM and night provides a sequence of periods from 8PM to 9AM. Morning present a total of 312 observations while afternoon present a total of 360 and night presents a total of 984 observations.

The case study explores the consumption profile in three different scenarios in order to check more accurately the consumption progress according to the different aspects featuring sequences of five minutes. Therefore these three scenarios are classified in “morning”, “afternoon” and “night”. All these periods show very different consumption profiles. The morning period has a lot of activity in 500 W and divergences that describe variations from 500 W to consumption intervals placed in [700, 900], [900, 1100], [1100, 1300] W. In the afternoon period there has a nonlinear behavior that starts in 900 W followed by divergences to 1300 or 500 W later until 2300 W. Night period on the other hand shows a very similar behavior in all five minutes periods described by convergences to the 500 W while few particular periods show few variations between 500 W and 1000 W.

3.2. Results

As stated in the case study, three scenarios were considered to handle the analyzes. These are conducted with a reinforcement learning methodology with decisions concerning the forecasting algorithm selection in each five minutes period. These five-to-five minutes decisions correspond each one to two possible alternatives: KNN and ANN classified as 0 or 1 respectively. The reward is calculated for each five minutes period based on the decision criterion and the forecasting error observations, assigning the value 1 if the selected forecasting algorithm corresponds to the one with lower error otherwise it assigns 0. The average of rewards keeps an historic of rewards defining an accumulated reward divided by the number of periods of five minutes it has applied the decision criterion.

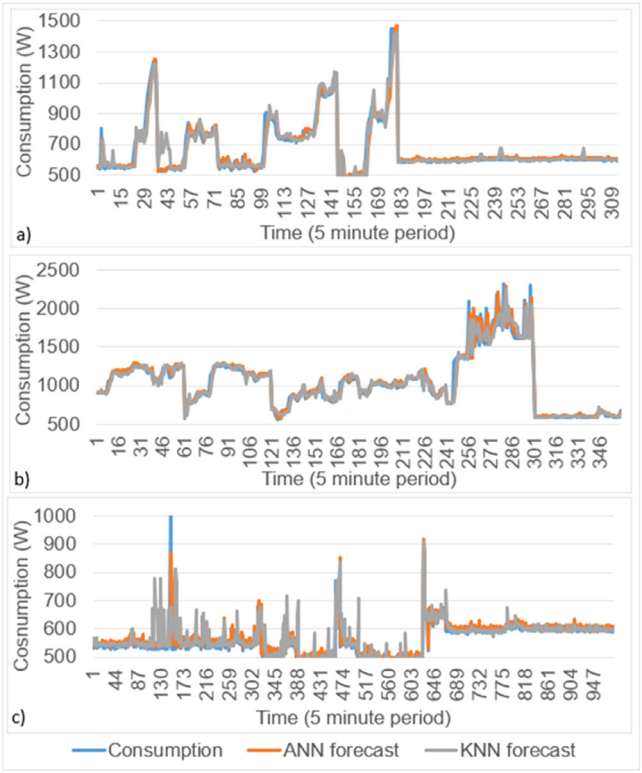

Fig. 2. Consumption profiles in scenarios (a) morning; (b) afternoon; (c) night.

The average reward progress in each five minutes period is verified for the reinforcement learning methods upper confidence bound as seen in Fig. 3.

Morning and afternoon scenarios in Fig. 3 show the average reward in the first five minutes starting in 1 for any exploration/exploitation interpreting a right forecasting selection. This is followed by few wrong forecasting algorithm selections on five minutes context. Following this logic morning converges to an average reward of [0.09, 0.2] in a short sequence of five minutes increasing nearly afterwards to a sequence of nonlinear behaviors describing average rewards of [0.4, 0.6]. Afternoon differs from this converging instead from the initial reward 1 to [0.5, 0.6] in a short sequence of five minutes. Both morning and afternoon tend to converge during the next five minutes to a single average reward which is [0.5, 0.6] for morning depending on the exploration and exploitation rate but almost exactly 0.6 for afternoon. Night shows a wrong forecasting selection for the first five minutes, followed by average reward increase from 0 to [0.5, 0.8]. The average reward tends to decrease to [0.5, 0.6].

Morning and afternoon scenarios show the average reward in the first five minutes starting in 1 for any exploration/exploitation interpreting a right forecasting selection. This is followed by few wrong forecasting algorithm selections on five minutes context. Following this logic morning converges to an average reward usually of [0.3, 0.5] in a short sequence of five minutes increasing nearly afterwards to a sequence of non linear behaviors describing average rewards of [0.5, 0.7]. Afternoon has an initial similar behavior with a reward initially assigned to 1 converging nearly afterwards to [0.4, 0.6] in a short sequence of five minutes. However afternoon differs from morning converging instead during the next five minutes to a single average reward about 0.5. Night shows a wrong forecasting selection for the first five minutes, followed by average reward increase from 0 to [0.5, 0.8].

The average reward tends to decrease to [0.5, 0.6]. Higher exploration rates will result on more convergence final average rewards for the different exploitation rates. While in upper confidence bound methods the different exploitation rates show a similar average behavior, the greedy method tends to have different average reward behaviors during the five minutes periods. The confidence bound is explored all possible scenarios concerning all possible combinations concerning the period, the exploration and exploitation rates. This confidence is a trust in using a forecasting algorithm in general cases. Fig. 4 shows the confidence of each scenario for KNN and ANN.

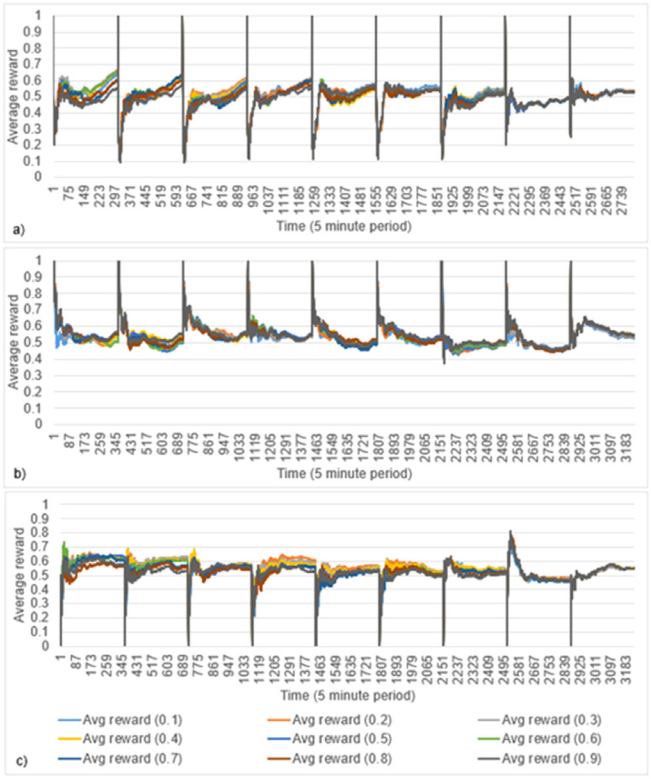

Fig. 3. Historic actions in morning scenario: (a) exploration rate = 0.2; (b) exploration rate = 0.5; (c) exploration rate = 0.8.

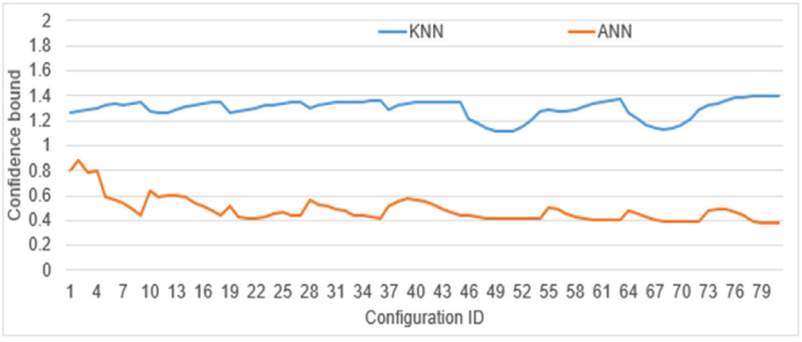

Fig. 4. Confidence bound concerning KNN and ANN decisions for morning scenario.

The confidence bound comparison represented in Fig. 4 shows that KNN is in all possible configurations the forecasting algorithm with lower forecasting error to be applied on general considering contexts of five minutes. KNN shows that the confidence stays in a range between 1 and 1.2 depending on the period, exploration and exploitation used in the configuration. The confidence bound of ANN shows smaller values between 0.4 and 0.9. Despite this, the values provided to the confidence bound of ANN show that in some particular situations ANN is still better than KNN.

4. Conclusion

This paper discusses the forecasting method that looks more adequate in each five minutes context concerning the decisions: k-nearest neighbors and artificial neural networks. Trial and test studies considering alternative

configurations including the exploration, exploitation, period and learning method are considered for a more detailed analysis. The confidence bound graphs show that KNN is clearly the most trustable forecasting algorithm despite the confidence bound for ANN showing that this is not true on specific five minutes periods. The average rewards shows good decision making for the morning, afternoon and night scenarios where confidence bound looks to consist on a more powerful learning method than greedy. However greedy shows to be a less exhausting method that sometimes results in not making few mistakes that result in high average rewards decreases. In the future, alternative contexts will be analyzed in order to improve accuracy of the model.

CRediT authorship contribution statement

D. Ramos: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **P. Faria:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **L. Gomes:** Data curation, Formal analysis, Investigation, Software, Validation, Visualization. **P. Campos:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing. **Z. Vale:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The present work has been developed under the EUREKA - ITEA3 Project (ITEA-18008), Project TIoCPS (ANI|P2020 POCI-01-0247-FEDER-046182), and has received funding from European Regional Development Fund through COMPETE 2020. The work has been done also in the scope of projects UIDB/00760/2020, CEECIND/02887/2017, financed by FEDER Funds through COMPETE program and National Funds through (FCT), Portugal.

References

[1] Faia R, Faria P, Vale Z, Spinola J. Demand response optimization using particle swarm algorithm considering optimum battery energy storage schedule in a residential house. Energies, Energies 2019;12(9):1645.
[2] Faria P, Vale Z. Demand response in electrical energy supply: An optimal real time pricing approach. Energy 2011;36(8):5374–84.
[3] Ramos D, Khorram M, Faria P, Vale Z. Load forecasting in an office building with different data structure and learning parameters. Forecasting 2021;3(1):242–55.
[4] Gomes L, Almeida C, Vale Z. Recommendation of workplaces in a coworking building: A cyber-physical approach supported by a context-aware multi-agent system. Sensors 2020;20:3597.
[5] Ramos D, Faria P, Vale Z, Mourinho J, Correia R. Industrial facility electricity consumption forecast using artificial neural networks and incremental learning. Energies 2020;13(18):4774.
[6] Dabbaghjamanesh M, Moeini A, Kavousi-Fard A. Reinforcement learning-based load forecasting of electric vehicle charging station using Q-learning technique. IEEE Trans Ind Inf 2021;17(6):4229–37.
[7] Huang C, Chen P. Joint demand forecasting and DQN-based control for energy-aware mobile traffic offloading. IEEE Access 2020;8:66588–97.
[8] Shuai H, He H. Online scheduling of a residential microgrid via Monte-Carlo tree search and a learned model. IEEE Trans Smart Grid 2021;12(2):1073–87.
[9] Aladdin S, El-Tantawy S, Fouda MM, Tag Eldien AS. Marla-SG: Multi-agent reinforcement learning algorithm for efficient demand response in smart grid. IEEE Access 2020;8:210626–39.
[10] Pei J, Hong P, Pan M, Liu J, Zhou J. Optimal VNF placement via deep reinforcement learning in SDN/NFV-Enabled networks. IEEE J Sel Areas Commun 2020;38(2):263–78.
[11] Yin L, Zhang C, Wang Y, Gao F, Yu J, Cheng L. Emotional deep learning programming controller for automatic voltage control of power systems. IEEE Access 2021;9:31880–91.
[12] Lu R, Hong SH, Yu M. Demand response for home energy management using reinforcement learning and artificial neural network. IEEE Trans Smart Grid 2019;10(6):6629–39.
[13] Solinas F, Bottaccioli L, Guelpa E, Verda V, Patti E. Peak shaving in district heating exploiting reinforcement learning and agent-based modelling. Eng Appl Artif Intell 2021;102:104235.

[14] Liu T, Tan Z, Xu C, Chen H, Li Z. Study on deep reinforcement learning techniques for building energy consumption forecasting. Energy Build 2020;208:109675.
[15] Wen L, Zhou K, Li J, Wang S. Modified deep learning and reinforcement learning for an incentive-based demand response model. Energy 2020;205:118019.