IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 10, 2022, accepted May 27, 2022, date of publication June 8, 2022, date of current version June 14, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3180754

A Contextual Reinforcement Learning Approach for Electricity Consumption Forecasting in Buildings

DANIEL RAMOS, PEDRO FARIA, (Member, IEEE), LUIS GOMES, (Member, IEEE), AND ZITA VALE, (Senior Member, IEEE)

Intelligent Systems Associate Laboratory (LASI), Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD), Polytechnic Institute of Porto (ISEP/IPP), 4200-072 Porto, Portugal

Corresponding author: Zita Vale (zav@isep.ipp.pt)

This article is a result of the project REal-Time support Infrastructure and Energy management for Intelligent carbon-Neutral smArt cities (RETINA) (NORTE-01-0145-FEDER-000062), supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and grant CEECIND/02887/2017. The authors acknowledge the work facilities and equipment provided by the Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development (GECAD) research center (UIDB/00760/2020) to the project team.

ABSTRACT The energy management of buildings plays a vital role in the energy sector. With that in mind, and targeting an accurate forecast of electricity consumption, in the present paper is aimed to provide decision on the best prediction algorithm for each context. It may also increase energy usage related with renewables. In this way, the identification of different contexts is an advantage that may improve prediction accuracy. This paper proposes an innovative approach where a decision tree is used to identify different contexts in energy patterns. One week of five-minutes data sampling is used to test the proposed methodology. Each context is evaluated with a decision criterion based on reinforcement learning to find the best suitable forecasting algorithm. Two forecasting models are approached in this paper, based on K-Nearest Neighbor and Artificial Neural Networks, to illustrate the application of the proposed methodology. The reinforcement learning criterion consists of using the Multiarmed Bandit algorithm. The obtained results validate the adequacy of the proposed methodology in two case-studies: building; and industry.

INDEX TERMS Consumption forecast, contextual operation, decision tree, reinforcement learning.

I. INTRODUCTION

An important aspect to improve the energy management, namely in the presence of demand response programs, is the forecasting of electricity consuming activities [1]. In fact, the present paper's authors have previously published several works in the literature concerning electricity consumption forecast [2]. K-nearest Neighbors (KNN) and Artificial Neural Networks (ANN) have been proved to be adequate technics for an office building application. However, in some specific periods, here stated as contexts, one of the algorithms is better than the other. Moreover, reinforcement learning has been largely applied to power and energy systems problems [3], providing learning of decisions in complex modeling environments. The authors of the present paper have also used reinforcement learning in buildings environments, despite not for consumption forecasting, in [4].

The electricity consumption forecasting is important to guarantee improved energy management in smart buildings [5]. Therefore, there are in the literature several buildings with data accessibility that research different machine learning techniques on how to achieve more accurate predictions, as in [6].

Buildings equipped with smart grids technology take advantage of data generated from several sources, including smart meters, phasor measurement units, and various sensors [7]. Using such data, forecasting algorithms are essential for prediction activities. Artificial Neural Networks have the advantage of extract and model unseen relationships and features. This ability gifts the neural networks with more robust choices if used the right way [8]. The K-Nearest Neighbour algorithm is an alternative recommended for time series classification. However, the algorithm's performance

The associate editor coordinating the review of this manuscript and approving it for publication was Zhouyang Ren.

requires a minimum quantity of labeled data [9]. The decrease of energy costs may be more effective with the assistance of modeling strategies that combine different forecasting algorithms including Artificial Neural Networks and Random Forest [10]. In fact, the uncertainties of load demand in the energy management present obstacles to achieve accurate forecasts. Reinforcement learning is recommended to overcome complex nonlinear issues with a decision-making ability that optimizes the current solution to be more effective [11, 12]. Reinforcement learning has a strong learning ability and high adaptability gifted with control and decision-making abilities. These are essential to ensure optimal outcomes in different scenarios including in robotics and distributed control [13]. Reinforcement learning is used for different applications according to the problem diversity, including performance improvement. It is also stated that a few applications use reinforcement learning to improve the prediction accuracy with different deep learning techniques, which is the case of this paper. Additionally, the learning method is also discussed being the Q-learning a researched option [14].

Given the results of the above-mentioned literature, the methodology proposed in the present paper aims to, in the first step, identify different contexts using decision trees. Then, reinforcement learning is applied in each context to identify the most accurate forecasting model. It innovates in overcoming the approach of selecting a single forecasting model for all the operational situations in a single consumer or building. For illustration purposes, models based on ANN and KNN forecasting algorithms have been used. The motivation consists in improving the forecasts obtained in recent research published by the authors of this paper [2]. Therefore, the authors reuse several forecasting aspects from [2] including the forecast horizon and forecast strategies. Innovative topics featuring the formation of new contexts with decision tree training and the reinforcement learning evaluation considering the most effective algorithm in different contexts are expected to improve these forecasts. Moreover, the decision tree and reinforcement learning innovative aspects are inspired from recent research published by the authors of this paper, respectively in [15] and [16].

After this introduction, Section 2 explains the proposed contextual approach, Section 3 evidence the details of the case study, and Section 4 presents the obtained results. Finally, Section 5 presents all the conclusions.

II. PROPOSED CONTEXTUAL APPROACH

In this section, it is explained the different phases of the proposed contextual approach. These include obtaining energy consumption forecasts, decision rule-based learning, definition of contexts, learning process, and the selection of the best forecasting algorithm for the target context.

The main goal is to evaluate the best forecasting model for each of different contexts. After obtaining energy consumption forecasts with different algorithms, a decision tree gifted with rule-based learning defines different contexts. Later, a learning process evaluates the best algorithm for different contexts. The first step consists of obtaining energy consumption forecasts for five minutes and according to two algorithms: Artificial Neural Networks and K-Nearest Neighbors.

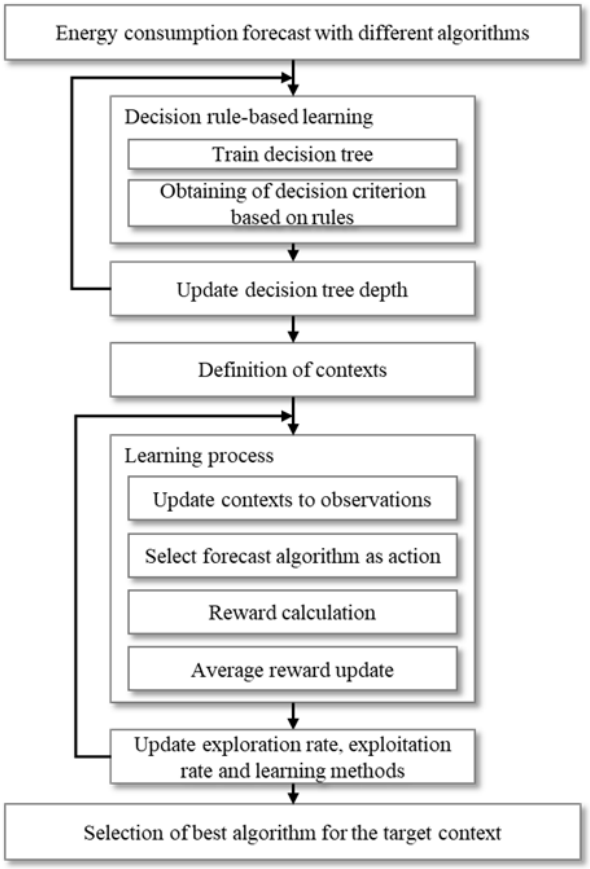


FIGURE 1. Proposed contextual approach.

Afterwards, a rule-based decision learning trains a decision tree with the forecasting data of both algorithms and additional factors from the actual and previous periods.

These factors consider time features including the weekday and the actual period and furthermore consider quantitative data obtained from the previous period including the consumption and two sensor devices data. These last two factors monitored on sensors devices consist of CO2 and a light variable with the value one or zero corresponding respectively to light in the building or no activity at all. These two parameters have been selected in sequence of the validation made in [2].

The learning process arises to evaluate the more suitable forecasting algorithm in different contexts. A set of agents perform this evaluation in an interactive environment through trial and error using feedback from their actions, observations, and rewards. The observations correspond to the contexts defined previously in rule-based decision learning. The agent's action is triggered every five minutes, and it

corresponds to the selection of a forecasting algorithm, either K-Nearest Neighbors or Artificial Neural Networks. The reward is calculated after every five minutes after the agent algorithm selection, representing how good the forecasting algorithm selection was for each actual context. In one hand, Rewards assigned to 0 correspond to scenarios where the selected algorithm is the one with higher forecasting error. On the other hand, rewards assigned to 1 correspond to scenarios where the selected forecasting algorithm has lower forecasting error. Each obtained reward is updated to an average of rewards, measuring the reward performance for all five-minute periods. In other words, the average of rewards measures the algorithm selection performance with lower forecasting error expectations. In each context evaluation, the learning methods and the exploration and exploitation rates are updated. The learning methods may correspond to greedy or upper confidence bound — the exploration rate focus on the angle of unexplored territory for each forecasting algorithm selection.

The exploitation rate focus on the knowledge exploration of a particular forecasting algorithm selection. After evaluating the best forecasting algorithm for all five minutes periods, the multi-agent system is prepared to select the best forecasting algorithm for the target context. Then, according to upper confidence and greedy learning methods, the action is calculated every five minutes according (1) and (2).

$$At = argmax(Qt(a) + c * \sqrt{\frac{lnt}{Nt(a)}} \quad (1)$$

$$At = argmax(Qt(a) \quad (2)$$

where:

- Nt(a) – number of times the action has been selected before time t
- Q(t) – current estimation
- c – degree of exploration
- a – maximizing action

III. CASE STUDIES

In order to illustrate the use of the proposed methodology, the implemented decision tree methodology studies a sample of data obtained from electric devices measuring different units and magnitudes. It has been implemented, in this paper, for two case studies: a building case study, and a industrial case study.

In the building case study, it is contextualized for a whole week from 18 to 24 November 2019 in five minutes periods.

Only a week with five minutes contexts from 18 to 24 November 2019 is considered to compare the same data size studied in recent publications by the authors of this paper [15]. Table 1 presents the decision tree inputs structure with the weekday, the allocated period, the consumption, the light, and the CO2. This table also adds the decision tree output structure with the forecasting algorithm application. Moreover, the input variables with nonlinear behaviors are studied according to their profile during 18 to 24 November 2019 in Fig. 2. Therefore, temporal variables are excluded from the analysis in Fig. 2 keeping however the consumption, light and CO2 profile. The light and CO2 sensors were added to the decision tree structure due to previous research published by the authors of this paper concluding that these two factors have more influence on the consumption [17].

TABLE 1. Decision tree inputs and outputs structure.

Week day	Allocated period	Consumption	Light	CO2	Forecasting algorithm
0	540	549.00	0	13.00	ANN
0	545	541.67	0	13.37	ANN
0	550	553.33	0	14.00	KNN
0	555	561.67	1	13.83	ANN
0	560	553.00	0	13.33	ANN
0	565	561.67	0	13.63	KNN
0	570	550.00	0	13.40	KNN
0	575	553.33	0	13.23	ANN
0	580	549.33	0	13.10	ANN
0	585	543.33	0	13.37	KNN
0	590	806.33	0	13.43	ANN
0	595	577.67	0	13.30	KNN
0	600	543.00	0	13.00	KNN
0	605	552.00	0	13.13	ANN
0	610	556.67	0	13.60	ANN

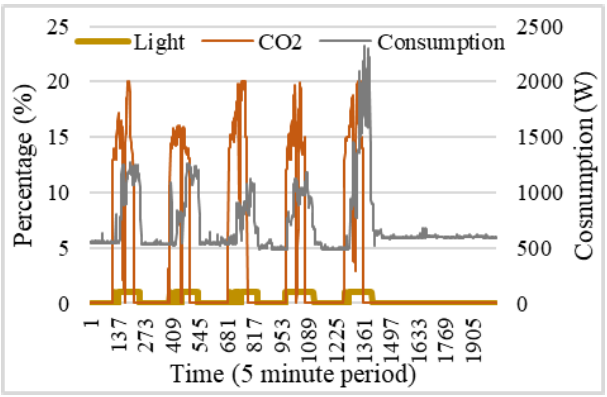


FIGURE 2. Input parameters of train data (decision tree).

The case study researches the different factors according to a weekly profile and five minutes contexts. Five similar patterns are identified, representing the activity data from each day of the week more concretely from Monday to Friday. This is followed by two similar patterns representing the low activity of the weekend. The consumption shows usual variations from 500 to 1500 W, as seen on the patterns from Monday to Thursday. The consumption variation from Friday is shown to be more productive, reaching consumption ranges higher than 2000 W. During the weekend, the consumption behavior is described by variations nearly to 600W. The light intensity describes variations between 0 and 1, representing respectively the absence or presence of light intensity measuring devices. CO2 devices present variations between 0 and 20%. The two sensors present null values during the whole weekend.

The reinforcement learning methodology studies the evaluation of the most suitable forecasting algorithm in five minutes from 18 to 24 November 2019. These five minutes

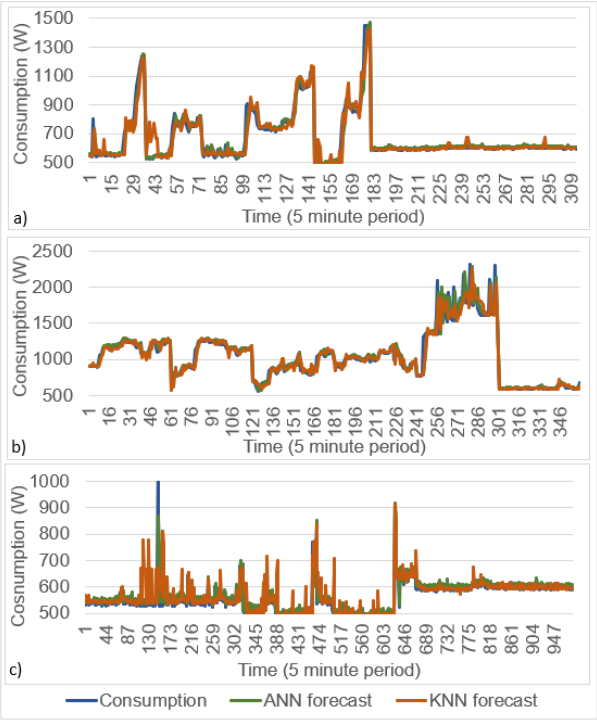


FIGURE 3. Consumption profiles: a) morning, b) afternoon, c) night.

decisions correspond to the forecasting algorithm selection, K-Nearest Neighbors, or Artificial Neural Networks. One week with five minutes contexts is considered to compare with other publications by the authors of this paper [16].

Regarding the industrial case study, which has been included for validation purposes, detailed information is not provided due to space limitations. Further details can be obtained in [18].

IV. RESULTS

In this section are presented the results regarding the use of the proposed methodology. These are obtained with the greedy learning method and according to four selected contexts (SC1, SC2, SC3, SC4).

A. BUILDING

The decision tree approach has been applied to the data in section III, testing different tree depths. Three data samples evidence different day features classified as the morning, afternoon, and night labeled respectively in a), b), and c), as seen in Fig. 3. These three samples correspond to previous known research published by the authors of this paper [16] and are detailed in this case study to support known forecasts in unique and different parts of the day. These forecasts are later used as research during the reinforcement learning evaluation of the most effective algorithm in different contexts.

The k-nearest neighbors and artificial neural networks present very accurate predictions much nearer to the real consumption for almost all five-minute periods. The morning scenario presents consumption variations between 500 and 1500 W. The afternoon scenario presents variations between 500 and 1500 W and between 500 and 2500 W. Finally, the night scenario presents many variations between 500 and 600 W and sequences of 5 minutes reaching 1000W.

TABLE 2. Accuracy of each depth scenario.

Depth	2	3	4	5	6
Accuracy	66.96%	66.96%	66.96%	67.86%	71.43%

The accuracy of the decision tree resulted from the depth parameterization is presented in Table 2.

Table 2 evidence very accurate results for the different depth parameterization values. It is noted that depth parameterizations assigned within ranges between 2 and 4 are not large enough to result in accuracies greater than 66.96%. However, it is possible to obtain higher accuracies by increasing the decision tree depth to values higher than 4. As seen in Table 2, increasing the depth parameterization value to 5 and 6 results in more accurate results, respectively 67.86%, and 71.43%. Therefore, while no real improvements are seen for depth ranges between 2 and 4, parameterization depth value changes to 5 and 6 show accuracy improvements respectively of 0.90% and 4.47%. The reason for these improvements is a higher complexity in the elaboration of decision rules. Therefore, the higher the decision tree depth, the higher the complexity of rules, possibly resulting in more accurate results. The accuracy results obtained in the decision tree feature similar research provided by the authors of this paper [15].

A simple rules elaboration illustrates the decision tree for a depth assigned to the value two as presented in Fig. 4. This scenario is a simple example to summarize the simpler logic presented in the decision tree rules. As identified previously in Table 2, the scenario with decision tree depth assigned to 6 leads to more accurate results. Therefore, the rules split of this scenario is analyzed in List 1. The decision tree presented in Fig. 4 shows very simple rules for depth assigned to 2. Two contexts are identified on the decision tree in Fig. 4 with a) weekday from Monday to Friday and consumption ranges below or equal to 568.833 W or b) weekday from Monday to Friday and consumption ranges higher than 568.833 W. List 1 presents very complex rules for a decision tree depth assigned to 6 corresponding to a total of 46 contexts. These contexts presented many differences, including the day corresponding to a weekday from Monday to Friday or a weekend and specified ranges for consumption (cons), CO2 (CO2), and the period allocated (min). From these 46 contexts, several can be identified within the restrictions defined in a) and b).

Moreover, the selected contexts are identified within the restrictions defined in a) and b) and separating small from large occurrences labeling respectively in SC1, SC2, SC3, and SC4.

The learning phase studies the average rewards and the history of actions for five minutes periods and all exploration

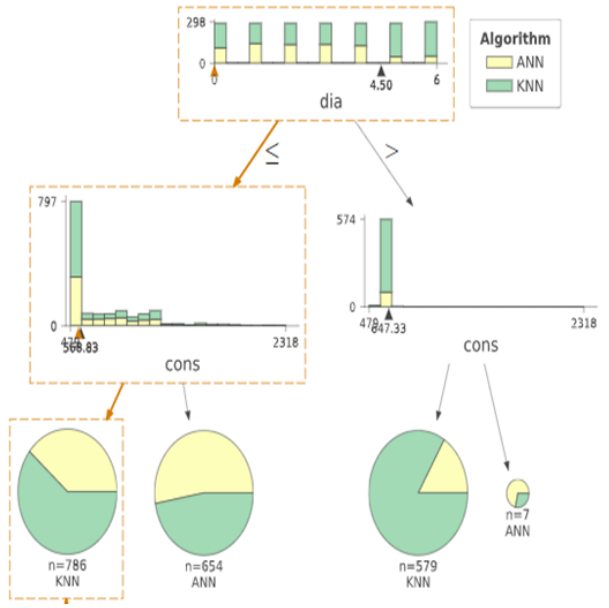


FIGURE 4. Decision tree for depth 2.

and exploitation rates from 0.1 to 0.9 with the greedy learning method. Moreover, this is presented respectively in Fig. 5, and Fig. 6 for four contexts SC1, SC2, SC3, and SC4 labeled respectively in a), b), c), and d).

The average reward alternates every five minutes between 0 and 1, representing algorithm selections with higher and lower forecasting errors. All presented scenarios start with an average reward assigned to 1 in the first five minutes, followed by at least an alternate decision that causes the average reward to converge to an interval between 0.2 and 0.8. Scenario a) has average rewards convergences between 0.7 and 0.8 for low exploration rates. However, it tends to decrease to patterns between 0.4 and 0.7 as the exploration rate increases. Scenario b) has average rewards to converge to 0.6 for lower exploration rates and 0.5 for higher. Scenario c) has average rewards to converge to 0.8 for low exploration rates. However, it tends to decrease to patterns between 0.3 and 0.8 as the exploration rate increases. Scenario d) has average rewards to converge to 0.5. As noted in scenarios b) and d), the increase of the exploration rate makes the different exploitation rates converge towards a more similar pattern.

Thus, the exploitation rates assigned to values 0.1, 0.4, and 0.9 tend to converge to higher average rewards on some scenarios and for the different exploration rates. The historic actions associated with context SC1 and for exploitation rates of 0.9 are illustrated in Fig. 6.

The history of actions is illustrated for context SC2 for the three exploitation rates identified previously as frequent cases to result in higher average rewards. These rates are within 0.9, 0.1, and 0.4, labeled respectively in a), b), and c) in Fig. 7. The historical actions for context SC1 illustrated in Fig. 6 show long sequences of five minutes deciding to use KNN repeatedly. After nearly 75 sequences of five minutes, the history of action finds it essential to alternate between

```
|--- day <= 4.500
|   |--- cons <= 568.833
|   |   |--- CO2 <= 13.350
|   |   |   |--- cons <= 540.833
|   |   |   |   |--- cons <= 485.833
|   |   |   |   |   |--- min <= 417.500 -> class: 1.0
|   |   |   |   |   |--- min >  417.500 -> class: 1.0
|   |   |   |   |--- cons >  485.833
|   |   |   |   |   |--- min <= 1242.500 -> class: 1.0
|   |   |   |   |   |--- min >  1242.500 -> class: 1.0
|   |   |   |--- cons >  540.833
|   |   |   |   |--- min <= 22.500 -> class: 1.0
|   |   |   |   |--- min >  22.500
|   |   |   |   |   |--- min <= 47.500 -> class: 0.0
|   |   |   |   |   |--- min >  47.500 -> class: 1.0
|   |   |--- CO2 >  13.350
|   |   |   |--- min <= 647.500
|   |   |   |   |--- cons <= 556.500
|   |   |   |   |   |--- min <= 627.500 -> class: 0.0
|   |   |   |   |   |--- min >  627.500 -> class: 1.0
|   |   |   |   |--- cons >  556.500
|   |   |   |   |   |--- CO2 <= 13.733 -> class: 0.0
|   |   |   |   |   |--- CO2 >  13.733 -> class: 1.0
|   |   |   |--- min >  647.500
|   |   |   |   |--- cons <= 518.667
|   |   |   |   |   |--- cons <= 502.167 -> class: 0.0
|   |   |   |   |   |--- cons >  502.167 -> class: 1.0
|   |   |   |   |--- cons >  518.667
|   |   |   |   |   |--- cons <= 555.667 -> class: 0.0
|   |   |   |   |   |--- cons >  555.667 -> class: 0.0
|   |--- cons >  568.833
|   |   |...
```

LIST 1. Decision tree rules for depth 6.

KNN and ANN, being this more frequent between 190 and 230 and between 260 and 297 long sequences of five minutes.

The historical actions for context SC2 show two possible behaviors for long sequences of five minutes: either to use repeatedly KNN as seen between 408 and 445 long sequences of five minutes or alternating very frequent between KNN ANN as seen between 445 and 482 long sequences of five minutes.

The history of actions of context SC1 presented in Fig.6, and SC2 presented in Fig. 7 labeled in a), b) and c) suggest a long-term learning approach more capable of alternating more between KNN and ANN according to the five minutes context, rather than repeatedly evaluating for KNN.

Lower exploitation rates tend to repeatedly evaluate more sequences of five minutes as KNN as evidenced in Fig.7 when comparing scenario b) with scenarios a) and b), respectively low and higher exploitation rates. This is understandable as low exploitation rates take more sequences of five minutes to acquire knowledge about KNN. Therefore, scenario a) has the advantage of acquiring more knowledge about a particular forecasting algorithm in fewer periods of five minutes. The historic actions associated with context SC3 and for exploitation rates of 0.9 are illustrated in Fig. 8.

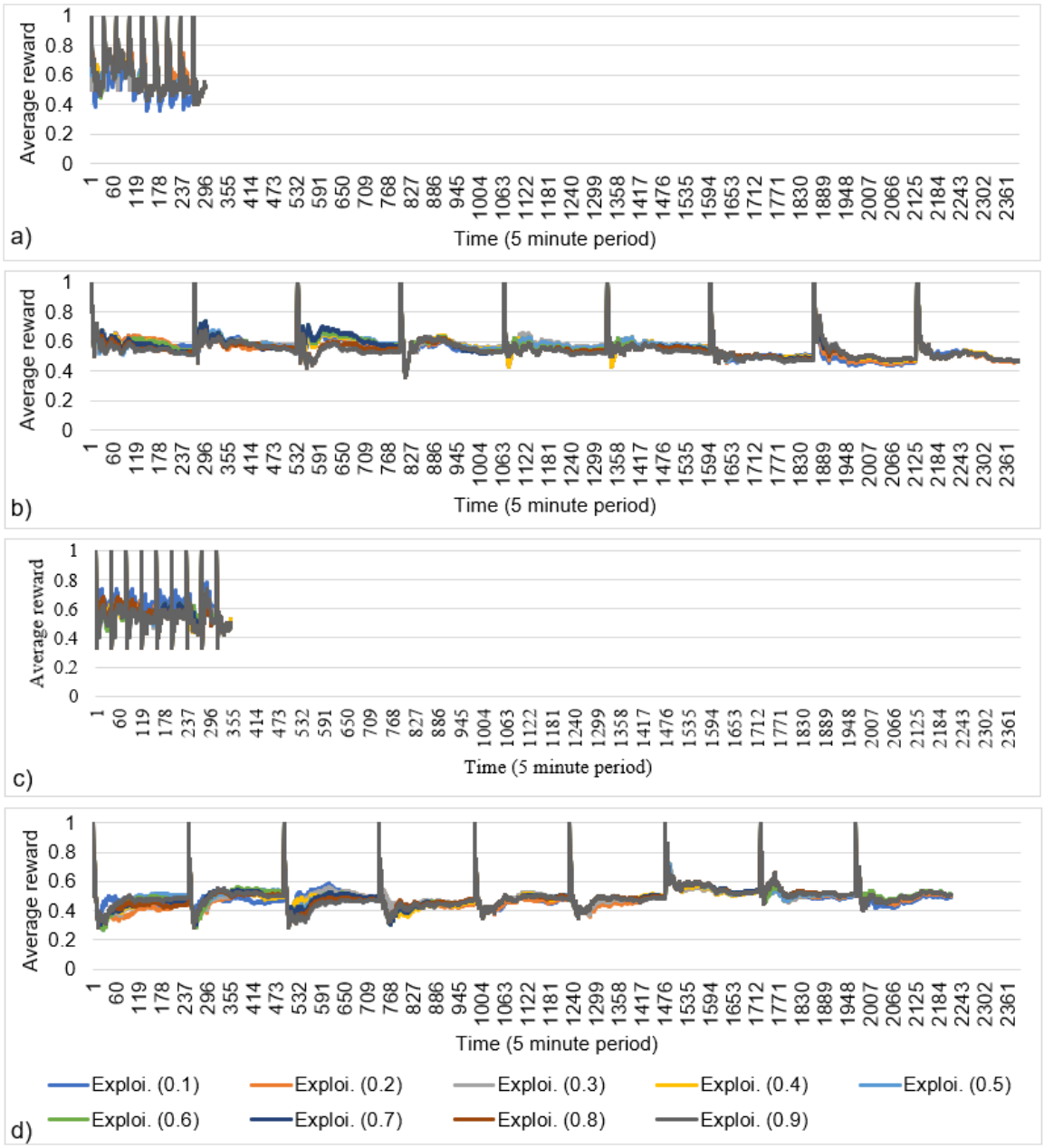


FIGURE 5. The average reward for contexts SC1, SC2, SC3, and SC4.

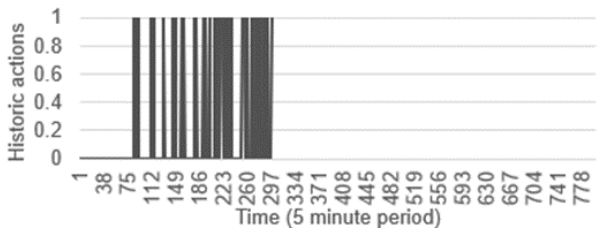


FIGURE 6. Historic of actions for context SC1 and exploitation rate 0.9.

The historical actions are illustrated for context SC4 for the three exploitation rates identified as frequent cases to result in higher average rewards. These rates are within 0.9, 0.1, and 0.4, labeled respectively in a), b), and c) in Fig. 9. The historical actions for context SC3 illustrated in Fig. 8 show long sequences of five minutes deciding to use KNN repeatedly. After nearly 75 sequences of five minutes, the history of action finds it essential to alternate between KNN and ANN. This behavior is presented between intervals of sequences of five minutes, including between 90 and 110, 120 and 150, 152 and 190, 192 and 294, and finally 197 and 334.

The history of actions for context SC4 show two usually and possible behaviors for long sequences of five minutes: either to use repeatedly KNN as seen between 260 and 297 long sequences of five minutes or alternating very

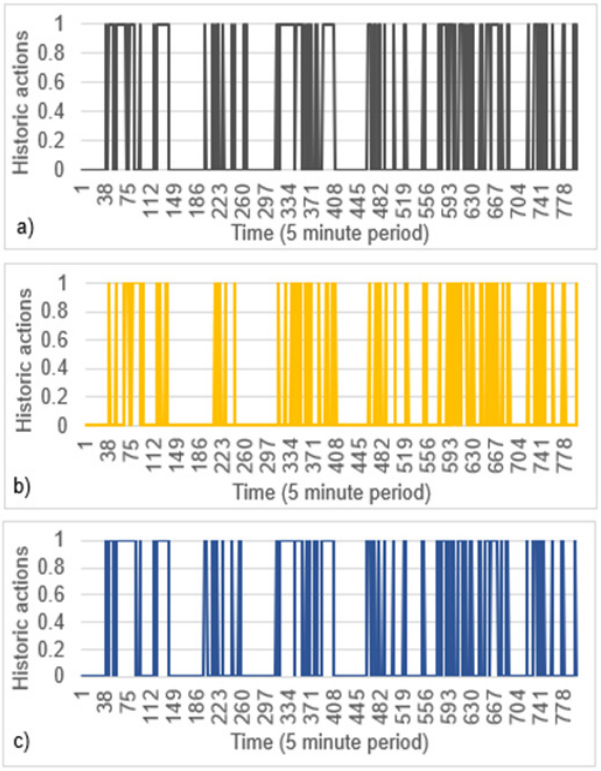


FIGURE 7. Historic actions for context SC2 and scenarios a), b) and c) respectively with exploitation rates 0.9, 0.1, and 0.4.

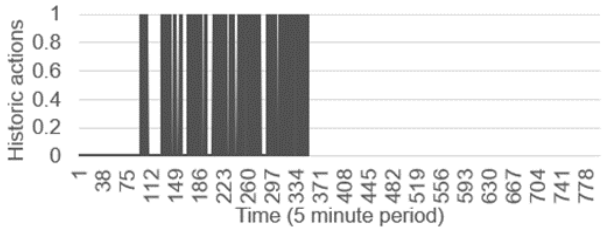


FIGURE 8. Historic of actions for context SC3 and exploitation rate 0.9.

frequent between KNN and ANN as seen between 297 and 334 long sequences of five minutes. Although these two behaviors are usual, the scenario represented in b) with a low exploitation rate of 0.1 shows that the historic of actions is also capable of evaluating small sequences of five minutes periods repeatedly as ANN as seen between 112 and 149 long sequences five minutes.

This is understandable as low exploitation rates need more time to acquire knowledge of ANN on five minutes contexts before having knowledge of both forecasting algorithm and reaching more pragmatic decisions.

The history of actions of context SC3 presented in Fig. 8, and SC4 presented in Fig.9 labeled in a), b) and c) suggest a long-term learning approach more capable of alternating more between KNN and ANN according to the five minutes context, rather than repeatedly evaluating for KNN or ANN. Lower exploitation rates tend to repeatedly evaluate more sequences of five minutes as KNN or ANN, as evidenced in Fig. 9 when comparing scenario b) with scenarios a) and

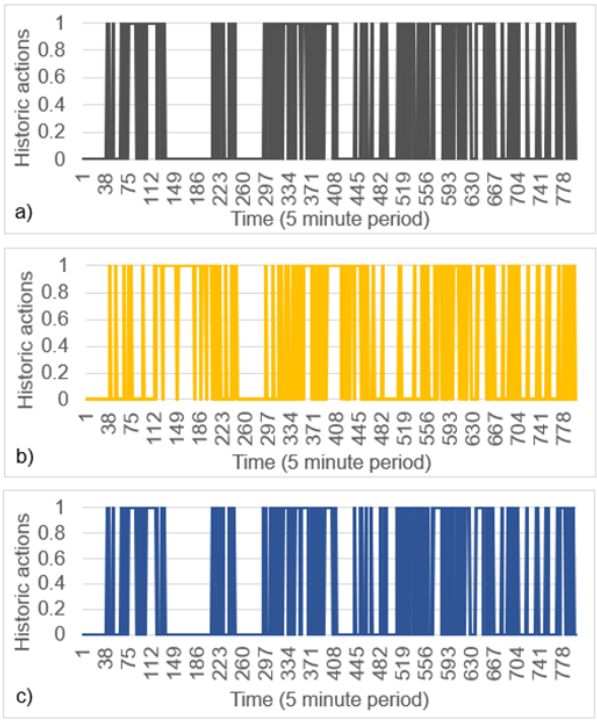


FIGURE 9. Historic actions for context SC4 and scenarios a), b) and c) respectively with exploitation rates 0.9, 0.1, and 0.4.

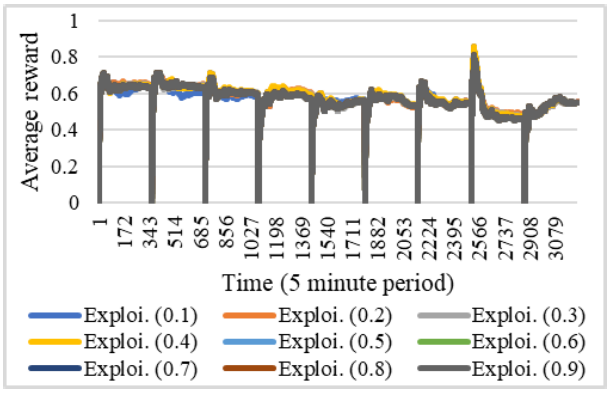


FIGURE 10. Average reward for the whole period (18-24 November 2019).

b), respectively low and higher exploitation rates. This is understandable as low exploitation rates take more sequences of five minutes to acquire knowledge about KNN or ANN. Therefore, scenario a) has the advantage of acquiring more knowledge about a particular forecasting algorithm in less periods of five minutes. It is possible to research the learning phase results for the whole week from 18 to 24 November 2019 with no contexts distinction. This research presents the average rewards for five minutes and all exploration and exploitation rates from 0.1 to 0.9, as illustrated in Fig. 10.

The results obtained in Fig. 10 presents overall average rewards nearly to 0.6, highlighting average rewards above reasonable. It is possible to obtain higher average rewards with context distinction for context SC3 nearly to 0.8 as illustrated in Fig. 5 scenario c).

TABLE 3. Accuracy of each depth scenario in industrial context.

Depth	2	3	4	5	6
Accuracy	61.11%	61.11%	60.42%	61.11%	56.25%

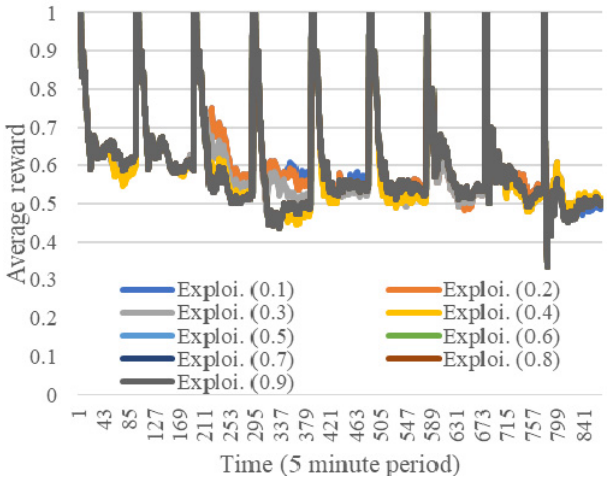


FIGURE 11. Average reward in the industrial context for the whole period (8 to 13 April 2019).

B. INDUSTRY

An identical simulation contextualized in industrial energy consumptions compares the decision tree accuracies and the average rewards with the electrical building simulation previously studied.

The accuracy of the decision tree is obtained for different tree depths according to an industrial use case as visualized in Table 3.

The decision tree accuracies visualized in Table 3 evidence very accurate predictions between 60.42 and 61.11% using decision tree depths assigned to values between two and five. The decision tree loses accuracy while improving the decision tree depth from value five to value six decreasing the accuracy from 61.11 to 56.25%. This is logical as the use of time features and industrial energy consumption has its limitations while elaborating decision rules. Table 3 also evidences the decision tree accuracy decrease from 61.11 to 60.42% while changing the depth from value three to value four. However, a decision tree depth increase from value four to value five, improves the accuracy from 60.42 to 61.11 %.

The average rewards evaluation of the most effective forecasting algorithm application in different five minutes contexts is also studied for the industrial context. This analysis considers all exploration and exploitations rates from 0.1 to 0.9 in the learning phase parameterization with the greedy method application as illustrated in Fig. 11.

The average rewards contextualized in the industrial context show an initial average reward of one for all exploration rates due to the selection of the most effective forecasting algorithm in the first five minutes. This is followed by at least a forecasting algorithm selection with lower accuracy leading to the average reward decreasing from 1 to a lower value between 0.4 and 0.6. The average reward converges to 0.6 for exploration rates between 0.1 and 0.2 and to 0.5 for exploration rates between 0.3 and 0.9 until the last five minutes period evaluation.

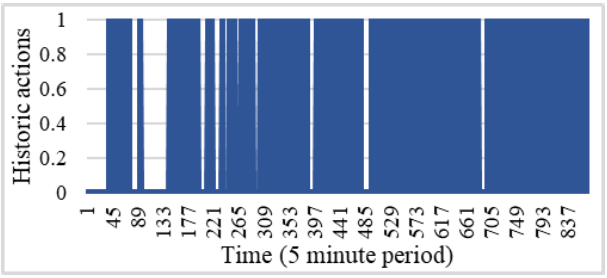


FIGURE 12. Historic of actions for industry, exploitation rate 0.4.

The historic of actions studies the forecasting algorithm application in different five minutes periods. The k-nearest neighbors and artificial neural networks applications are alternated in different five minutes contexts for an industrial application with an exploitation rate assigned to 0.4 as illustrated in Fig. 12.

Alternations between the k-nearest neighbors and artificial neural networks applications in different five minutes contexts can be visualized in many long five minutes sequences. Some examples are observed between 163 and 186 periods of five minutes, between 325 and 379 periods of five minutes. The historic of actions presents another behavior where the k-nearest neighbors algorithm is applied repeatedly in long sequences of periods with five minutes. Some examples are observed including between 1 and 37 sequences of five minutes and between 91 and 145 sequences of five minutes.

V. CONCLUSION

This paper identifies suitable contexts through decision tree rules and analyzes the best forecasting model in different periods. The results obtained for the different decision tree depth values suggest the decision tree is suitable to identify contexts. It is also noted that increasing the depth value higher enough makes the decision rules complex enough to result in more accurate results. The obtained results on the learning phase for the greedy method show average rewards converging to values above reasonable. It is noted that increasing the exploration rate may decrease the final average reward in some contexts. The historic actions present two frequent patterns on long sequences of five minutes: to select KNN or ANN repeatedly or to alternate between KNN and ANN. It also noted that it is advantageous to use large exploitation rates to acquire more knowledge of a particular forecasting algorithm selection in fewer periods of five minutes. Moreover, this motivates to alternate between KNN and ANN on different five minutes contexts faster than for low exploitation rates. An accurate analysis of the learning phase results for the whole period reveals that context use is advantageous for obtaining higher average rewards. The

industrial use case also reaches very accurate decision tree accuracies, however this is limited to a maximum of 61.11% while the electrical building application contextualized in this paper reaches accuracies with maximums of 71.43 %. It is inferred that the less precise decision tree accuracy in the industrial context is because of the lack of sensors data in the decision rules. Moreover, this problem may also explain why the increase of the decision tree depth at some point decreases the accuracy. It is inferred that the rules built in the decision tree training are able to reach stronger logics when including sensors data. The average of rewards analysis on the industrial use case has also obtained above reasonable forecasting algorithm applications in different contexts. The historic of actions contextualized in the industrial use case have shown two similar behaviors leading to either alternating between k-nearest neighbors and artificial neural networks applications or evaluating repeatedly with k-nearest neighbors.

REFERENCES

[1] P. Faria, Z. Vale, and J. Baptista, "Constrained consumption shifting management in the distributed energy resources scheduling considering demand response," *Energy Convers. Manage.*, vol. 93, pp. 309–320, Mar. 2015, doi: 10.1016/j.enconman.2015.01.028.

[2] D. Ramos, M. Khorram, P. Faria, and Z. Vale, "Load forecasting in an office building with different data structure and learning parameters," *Forecasting*, vol. 3, no. 1, pp. 242–255, 2021, doi: 10.3390/forecast3010015.

[3] M. Dabbaghjamanesh, A. Moeini, and A. Kavousi-Fard, "Reinforcement learning-based load forecasting of electric vehicle charging station using Q-learning technique," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4229–4237, Jun. 2021, doi: 10.1109/tii.2020.2990397.

[4] L. Gomes, C. Almeida, and Z. Vale, "Recommendation of workplaces in a coworking building: A cyber-physical approach supported by a context-aware multi-agent system," *Sensors*, vol. 20, no. 12, p. 3597, Jun. 2020, doi: 10.3390/s20123597.

[5] S. Hadri, Y. Naitmalek, M. Najib, M. Bakhouya, Y. Fakhri, and M. Elaroussi, "A comparative study of predictive approaches for load forecasting in smart buildings," *Proc. Comput. Sci.*, vol. 160, pp. 173–180, Jan. 2019, doi: 10.1016/j.procs.2019.09.458.

[6] L. Zhang, J. Wen, Y. Li, J. Chen, Y. Ye, Y. Fu, and W. Livingood, "A review of machine learning in building load prediction," *Appl. Energy*, vol. 285, Mar. 2021, Art. no. 116452, doi: 10.1016/j.apenergy.2021.116452.

[7] M. S. Ibrahim, W. Dong, and Q. Yang, "Machine learning driven smart electric power systems: Current trends and new perspectives," *Appl. Energy*, vol. 272, Aug. 2020, Art. no. 115237, doi: 10.1016/j.apenergy.2020.115237.

[8] T. Ahmad, H. Zhang, and B. Yan, "A review on renewable energy and electricity requirement forecasting models for smart grid and buildings," *Sustain. Cities Soc.*, vol. 55, Apr. 2020, Art. no. 102052, doi: 10.1016/j.scs.2020.102052.

[9] H. Gweon and H. Yu, "A nearest neighbor-based active learning method and its application to time series classification," *Pattern Recognit. Lett.*, vol. 146, pp. 230–236, Jun. 2021, doi: 10.1016/j.patrec.2021.03.016.

[10] M. Zekić-Sušac, A. Has, and M. Knežević, "Predicting energy cost of public buildings by artificial neural networks, CART, and random forest," *Neurocomputing*, vol. 439, pp. 223–233, Jun. 2021, doi: 10.1016/j.neucom.2020.01.124.

[11] M. Jin and J. Lavaei, "Stability-certified reinforcement learning: A control-theoretic perspective," *IEEE Access*, vol. 8, pp. 229086–229100, 2020, doi: 10.1109/ACCESS.2020.3045114.

[12] H. Liu, Z. Zhang, and D. Wang, "WRFMR: A multi-agent reinforcement learning method for cooperative tasks," *IEEE Access*, vol. 8, pp. 216320–216331, 2020, doi: 10.1109/ACCESS.2020.3040985.

[13] M.-L. Li, S. F. Chen, and J. Chen, "Adaptive learning: A new decentralized reinforcement learning approach for cooperative multi-agent systems," *IEEE Access*, vol. 8, pp. 99404–99421, 2020, doi: 10.1109/ACCESS.2020.2997899.

[14] N. V. Varghese and Q. H. Mahmoud, "A hybrid multi-task learning approach for optimizing deep reinforcement learning agents," *IEEE Access*, vol. 9, pp. 44681–44703, 2021, doi: 10.1109/ACCESS.2021.3065710.

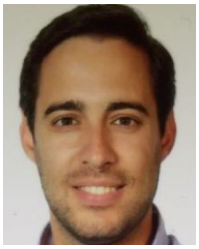
[15] D. Ramos, P. Faria, A. Morais, and Z. Vale, "Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building," *Energy Rep.*, vol. 8, pp. 417–422, Jun. 2022, doi: 10.1016/j.egyr.2022.01.046.

[16] D. Ramos, P. Faria, L. Gomes, P. Campos, and Z. Vale, "Selection of features in reinforcement learning applied to energy consumption forecast in buildings according to different contexts," *Energy Rep.*, vol. 8, pp. 423–429, Jun. 2022, doi: 10.1016/j.egyr.2022.01.047.

[17] D. Ramos, B. Teixeira, P. Faria, L. Gomes, O. Abrishambaf, and Z. Vale, "Use of sensors and analyzers data for load forecasting: A two stage approach," *Sensors*, vol. 20, no. 12, p. 3524, Jun. 2020, doi: 10.3390/s20123524.

[18] D. Ramos, P. Faria, Z. Vale, and R. Correia, "Short time electricity consumption forecast in an industry facility," *IEEE Trans. Ind. Appl.*, vol. 58, no. 1, pp. 123–130, Jan. 2022, doi: 10.1109/TIA.2021.3123103.

DANIEL RAMOS received the B.Sc. degree in informatics engineering from the Polytechnic of Porto, Porto, Portugal, in 2019, and the M.Sc. degree in modeling, data analysis and decision support systems from the University of Porto, Porto, in 2021. He is currently a Researcher at GECAD, Polytechnic Institute of Porto. His research interests include data science, machine learning, reinforcement learning, demand response, and smart grids.



PEDRO FARIA (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Polytechnic of Porto, Portugal, in 2008 and 2011, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Trás-os-montes e Alto Douro, Vila Real, Portugal, in 2016. He is currently a Researcher with GECAD, Polytechnic Institute of Porto. His research interests include demand response, smart grids, and electricity markets.

LUIS GOMES (Member, IEEE) received the Ph.D. degree in computer engineering from the University of Salamanca, Spain, in 2020. He is currently a Junior Researcher at the Polytechnic of Porto-School of Engineering (ISEP). His research interests include smart buildings, microgrids, citizen energy communities, multi-agent systems, artificial intelligence applied to energy management, and the Internet of Things.

ZITA VALE (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Porto, Porto, Portugal, in 1993. She is currently a Professor with the Polytechnic Institute of Porto, Porto. Her research interests include artificial intelligence applications, smart grids, electricity markets, demand response, electric vehicles, and renewable energy sources.