



HAL
open science

Characterization of Residential Electricity Customers via Deep Ensemble Learning

Weixuan Lin, Di Wu

► **To cite this version:**

Weixuan Lin, Di Wu. Characterization of Residential Electricity Customers via Deep Ensemble Learning. IFIP International Workshop on Artificial Intelligence for Knowledge Management (AI4KMES), Aug 2021, Montreal, QC, Canada. pp.75-86, 10.1007/978-3-030-96592-1_6. hal-04120815

HAL Id: hal-04120815

<https://inria.hal.science/hal-04120815v1>

Submitted on 7 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Characterization of residential electricity customers via Deep Ensemble Learning

Weixuan Lin¹[0000-0002-4474-6950] and Di Wu¹[0000-0001-7419-9903]

Department of Electrical and Computer Engineering, McGill University, 3480 Rue
University, Montreal, Quebec, Canada
`di.wu5@mail.mcgill.ca`

Abstract. The household characteristics in an electric grid include the socio-economic status of households, the dwelling properties, the information on the appliance stock, and so forth. These characteristics are significantly beneficial to electric retailers, because they can be utilized to provide personalized services, improve the demand response, and make better energy efficiency programs. However, these privacy-sensitive characteristics (e.g., employment, income, age of residents) require time-consuming surveys, and it is difficult to be gathered in a large scale. In recent years, the increasing availability of electricity consumption data makes it possible to infer household characteristics from residential electricity consumption data. A number of supervised learning methods have been proposed. Among these solutions, features are extracted from the electricity consumption patterns, and the selected features are used to train a classifier or regressor. However, the existed methods depend on a single contributing model, which can be possibly undertrained. To achieve the optimal performance of classifiers for characteristics identification, we propose an ensemble framework based on bagging algorithms. With the proposed ensemble framework, the performance of characteristic identification has been improved.

Keywords: Ensemble learning · supervised classification

1 Introduction

The modern power grids are now facing increasing uncertainties from both the energy consumption side and the energy generation side [19, 6, 8]. Customer identification is of benefit for utility companies to optimize their energy programs. With detailed knowledge socio-economic characteristics of residential households, tailor saving advice can be customized for specific addressees (such as family with children or retirees). Thus, a comprehensive knowledge on customers helps electric retailers to make reliable decisions on the targeting of demand response and energy efficiency programs [17]. However, the socio-economic information of customers are privacy-sensitive and is protected by governments, especially in European [4]. Hence, in most of the cases, the knowledge of utilities about the customers does not reach beyond the address and billing information.

In the development of modern smart grids, smart meters have been becoming widely applied in residential households [22, 23, 7, 20, 12, 13]. With a large amount of fine-grained data of individual costumers collected by smart meters, a deeper insight about the consumption behaviours of customers becomes available. In recent years, machine learning has been recently utilized for power grids [21]. Using data collected from smart meters, researchers have paid attention to models that identify the socio-economic characteristics of residential customers from power consumption patterns. Beckel et al. [4, 3] propose a two-stage model for customer identification: first, temporal features are extracted from the original consumption patterns; second, the extracted features are used as the input for a classifier such as support-vector machines (SVM), k-nearest neighbors algorithm (KNN). With this model, various socio-economic labels have been classified by using residential consumption data. The temporal features extracted in [4] are based on priori-knowledge regarding consumption figures, ratios, temporal properties and statistical properties. Similar works have also been reported. Wang et al. decompose load profiles into partial usage patterns by using sparse coding, and the customers are classified with SVM [18]. Zhong and Tam use discrete Fourier transform and a classification-and-regression-tree (CART) to classify customers into groups based on consumption patterns [26]. Wang et al. replace the manually feature selection in [4] by a deep convolutional neural network (CNN) [17]. And the CNN in [17] is believed to be able to automatically select features from power consumption patterns. A better performance has been reported in [17] compared to [4], because CNN can extract the highly nonlinear relationships between electricity consumption in different time steps.

However, most of existing models are only dependent on a single contributing model, which might lead to an undertrained model. To achieve a better accuracy and more stable performance of a classification model, we propose a framework based on deep ensemble learning, which combines multiple classifiers to boost the accuracy and has showed its effectiveness on different types of applications [9]. This paper is organized as follow. The technical background will be introduced in Section 2. The experiments are conducted and the results are demonstrated in section 3. Last, conclusion will be given in Section 4.

2 Technical Background

2.1 Identification of Residential Customers

To acquire socio-demographic information of residential customers, time-consuming surveys are typically required. Besides, the acquisition of some private information is strictly regulated by governments, especially in Europe. To avoid these time-consuming surveys, researchers have proposed to infer socio-demographic information based on the existing surveys and power consumption of residential households [3, 4, 17]. The presumption of these models is that the power consumption pattern differs between households, and is related to the socio-demographic characteristics of a household. Therefore, classifiers based on super-

vised learning can be built to infer socio-demographic labels from consumption patterns of a residential household.

Households with different characteristics have evident difference on their consumption patterns. Based on the Commission for Energy Regulation (CER) dataset, Figure 1 demonstrates typical power consumption profiles of three customers. The consumption profiles of the three customers are distinct from each other. For customer#1017, most of the consumption of a day is located at the time span of [18, 27] and [31, 40]. The consumption profile of customer#1018 has smaller peak values and less regular profiles compared to customer#1017. The consumption profile of customer#1020 demonstrates less regular pattern than customer#1017 and #1018. Table 1 shows some of the socio-demographic information of the three customers, revealing that these three customers have different living conditions as well. Hence, by comparing the daily consumption profiles in Figure 1 and the socio-demographic information in Table 1, it is believed that the daily consumption patterns are related to the living conditions.

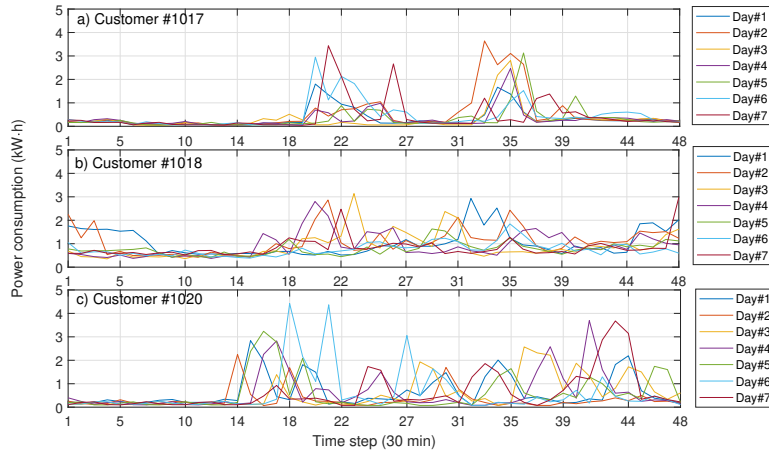


Fig. 1. Power consumption profile of a week for customers #1017, #1018, and #1020.

2.2 Problem Definition

The goal of this paper is to derive socio-demographic information from the power consumption patterns of a customer. This problem can be formulated as a supervised multi-class classification problem [17]: given the power consumption pattern $x_{i,j}$ and label $y_{i,j}$ of the i -th customer and the j -th label, a classification model $\hat{y}_{i,j} = F_j(x_{i,j}, \Theta)$ will be trained such that the loss function $L_j(\hat{y}_{i,j}, y_{i,j})$ is minimized.

Questions	#1017	#1018	#1020
Age	35~65	35~65	<35
Job	Retired	Retired	Not retired
Have children?	no	no	yes
House type	Semi-detached or terraced	Detached or bungalow	Semi-detached or terraced
Year of house	1983	1993	2000
Bedroom number	3	>5	4
Cooker type	Electric	Electric	Electric
Floor area	<100	>100 & <200	>100 & <200

Table 1. Socio-demographi information of customers #1017, #1018, and#1020.

2.3 Ensemble Learning

Ensemble learning has been developed since 1990s [25]. A ensemble model is designed to combine multiple base-learners, so that the error of single base-learners can be compensated by each other. It is expected that ensemble models can outperform single base-learner [15]. Among methods in ensemble learning, Bagging and Boosting are distinct ones and commonly used.

Bagging, i.e. Bootstrap Aggregating, was first proposed in [5] and has been developed into a number of variants [11, 2, 10]. The essence of bagging consist of three parts: making subsets by bootstrapping on dataset, training on subsets, and aggregating single-models into the final model [15]. In this paper, we proposed a ensemble framework based on bagging. Figure 2 illustrates the ensemble diagram. M subsets are bootstrapped from the training set D . For each subset D_i , a $Model_i$ is trained by a base-model. The choice of the base-models includes SVM, FNN, and LSTM. Finally, in the aggregation section, the final model is obtained by averaging the outputs from all base models.

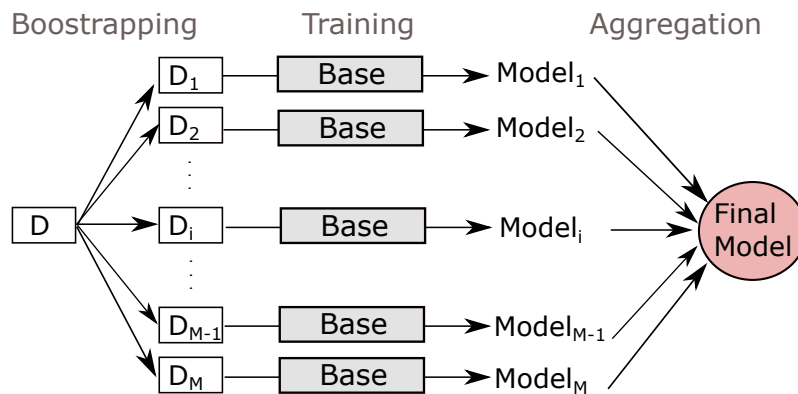


Fig. 2. Schematic of the proposed ensemble algorithm. D : training set. D_i subsets bootstrapped from the training set.

3 Experiments and Results

3.1 The CER Dataset Description

The dataset used in the experiments are provided by Commission for Energy Regulation (CER) [1], which was also used in the previous researches [4, 17, 16, 24]. The CER dataset consists of the electricity consumption of 4232 households and a pre-trial residential survey including various questions. The power consumption spans from the 195th day to the 212th day of 2009, at a interval of 30 minutes. And ten of the questions as selected in [17] are the labels to be classified in the following experiments. Table 2 demonstrates the questions and the total numbers of each answer.

Quest. No.	Socio-demographic question	Answers	Num.
300	Age of chief income earner	<65	3255
		≥65	953
310	Chief income earner is retired?	Retired	1285
		Not retired	2947
401	Social class of chief income earner	A or B or C	2482
		D or E	1593
410	Have children?	Yes	3003
		No	1129
450	House type	Detached or bungalow	1964
		semi-detached or terraced	2189
453	Age of house	>30	2161
		<30	2071
460	Number of bedrooms	<3	2288
		≥4	1944
4704	Cooking facility type	Electrical	2960
		Not electrical	1272
4905	Energy-efficient light bulb percentage	<1/2	2041
		≥1/2	2191
6103	Floor area	<200	1438
		≥200	343

Table 2. Information of pre-trial socio-demographic survey in CER dataset.

In the following experiments, the input consumption profiles have a time span of 7 days. To capture more training samples, sliding window is used as a data augmentation method. Each sampling window scans a consumption profile of 7 days (i.e. $7 \times 48 = 336$ time steps) of a customer. And Adjacent windows are shifted by a time span of 7 day (i.e. 48×7 time steps), in order to pertain the weekly pattern of each customer. Each customer has ten classification labels to be classified. 4232 customers are divided into training/test set with a ratio of 0.7 : 0.3. It is noted that some labels are missed for certain customers. The customers with missing labels are omitted in the experiments.

3.2 Evaluation Metrics

To evaluate the performance of a classification model with M classes, the first step is to statistically obtain a $M \times M$ confusion matrix C , which is used to count the number of correct classifications and the number of misclassifications for each class [14, 4, 17]. Each $C_{m,n}$ denotes the number of samples of class m classified into the class n . Based on the confusion matrix, the following metrics can be derived.

Accuracy The accuracy is defined as the rate of the samples that are correctly classified:

$$Acc = \frac{\sum_{m=1}^M C_{m,m}}{\sum_{m=1}^M \sum_{n=1}^M C_{m,n}} \quad (1)$$

F_1 score In a binary classification problem, the F_1 score is defined based on the number of TP (true positives, i.e. $C_{1,1}$), TN (true negatives), FP (false positives), and FN (false negatives), Pr (precision) and Re recall [14], where

$$Pr = TP / (TP + FN) \quad (2)$$

$$Re = TP / (TP + FP) \quad (3)$$

$$F_1 = 2 \frac{Pr \times Re}{Pr + Re} \quad (4)$$

In the case of multi-class problem, the F_1 score is generalized to macro- F_1 score [14], which is defined as the average value of the F_1 scores of all labels, i.e.

$$F_1^{macro} = \frac{1}{M} \sum_{m=1}^M F_1(m) \quad (5)$$

where M is the total number of classes and $F_1(m)$ is the F_1 score of the class m .

3.3 Experimental Setups

Experiments are first conducted by various individual classifiers including biased guess (BG), Feedforward Neural Network (FNN), Long short-term memory (LSTM) based recurrent neural network, and support-vector machines (SVM). The individual model with the best performance will be chosen as the base model in the ensemble learning framework. Experiments will be conducted by the ensemble learning framework as well. To be specific, the individual models are listed as followed.

Biased Guess (BG): Knowing the proportions of different classes in the training set, the BG model assumes that all cases in the test set have the same class with the largest proportion in the training set.

FNN: A 3-layer FNN with hidden units of (8, 8) are used as a classifier. The input layer has a unit number of 336, and the output layer has the unit number the same as the class number in the label.

LSTM: The LSTM baseline consists of two hidden LSTM layers with units of (8, 8). The input layer has an input shape of 7×48 . The unit number of the output layer is the same as the class number in the target label. And Softmax function is used for the output activation function.

SVM: The SVM baseline uses the kernel of radial basis function and the input is a time-series with 336 steps.

4 Experimental Results

4.1 Baseline Results

Table 3 demonstrates the prediction accuracy and F1 scores of the performance of baselines in test set, and Figure 3 visualizes the accuracy and F1 results. From Table 3 and Figure 3, it is noticed that all labels, SVM performs either better or close to the other baselines. LSTM performs close to SVM in most of the labels. And FNN is consistently the worst among the baselines. Averaging the scores in all labels, as the average accuracy shown in Table 3, SVM and LSTM have 0.669, and 0.651, respectively. While SVM and LSTM have average accuracy better than the averaged accuracy of BG (0.647), FNN has the worst average accuracy of 0.617. In terms of the average F1 scores, SVM scores 0.579, which is similar with the average F1 LSTM of 0.577. Also, FNN underperforms all baselines in average F1 scores, with 0.564.

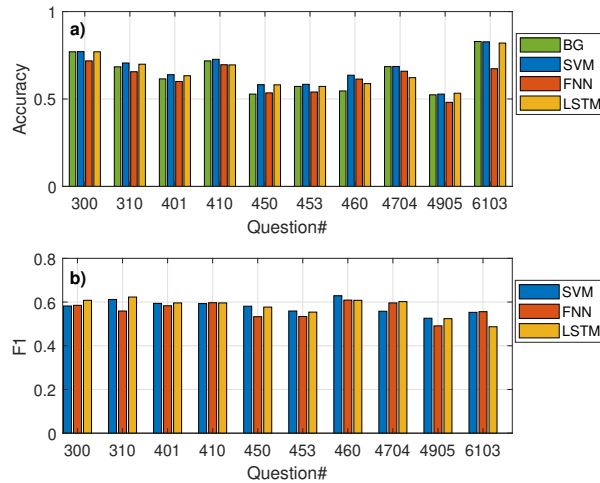


Fig. 3. Visualization of a) accuracy and b) F1 scores of baselines.

Questions	Accuracy				F1		
	BG	SVM	FNN	LSTM	SVM	FNN	LSTM
300	0.770	0.771	0.718	0.770	0.582	0.585	0.608
310	0.684	0.706	0.656	0.699	0.612	0.559	0.623
401	0.615	0.639	0.600	0.633	0.594	0.583	0.596
410	0.718	0.727	0.696	0.695	0.593	0.597	0.596
450	0.528	0.582	0.535	0.581	0.581	0.533	0.577
453	0.572	0.584	0.540	0.572	0.559	0.534	0.554
460	0.546	0.636	0.614	0.588	0.629	0.609	0.608
4704	0.685	0.686	0.659	0.622	0.558	0.596	0.602
4905	0.524	0.528	0.481	0.533	0.526	0.491	0.524
6103	0.829	0.827	0.673	0.820	0.553	0.556	0.487
Average.	0.647	0.669	0.617	0.651	0.579	0.564	0.577

Table 3. Accuracy scores of the baseline models of SVM, FNN, and LSTM.

4.2 Ensemble Model Results

Quest.	Accuracy			F1		
	SVM	FNN	LSTM	SVM	FNN	LSTM
300	0.775	0.766	0.694	0.593	0.618	0.589
310	0.708	0.693	0.652	0.616	0.615	0.583
401	0.643	0.625	0.576	0.595	0.590	0.555
410	0.725	0.712	0.658	0.592	0.599	0.569
450	0.574	0.562	0.571	0.569	0.559	0.566
453	0.583	0.590	0.545	0.557	0.573	0.532
460	0.635	0.624	0.554	0.630	0.618	0.562
4704	0.690	0.693	0.630	0.575	0.604	0.572
4905	0.528	0.515	0.511	0.518	0.514	0.507
6103	0.824	0.788	0.738	0.543	0.557	0.559
Avg.	0.669	0.657	0.613	0.579	0.585	0.560

Table 4. Accuracy scores of the ensemble models with base models of SVM, FNN, and LSTM.

Table 4 shows the accuracy/F1 scores of the ensemble models with different base-models, and Figure 4 visualizes the performance. From the accuracy performance shown in Figure 4, the ensemble-SVM and ensemble-FNN model perform similarly, but ensemble-LSTM has the worst performance in most of the labels. Overall, as shown in the and average F1 in Figure 4, ensemble-SVM and ensemble-FNN have similar average accuracy of 0.669 and 0.659, respectively. But ensemble-LSTM only has an average accuracy of 0.613, significantly smaller than the others. Regarding the average F1, ensemble-SVM, ensemble-FNN, and ensemble-LSTM have 0.579, 0.585 and 0.56, respectively. Therefore, compared to single baseline models shown in Table 3, ensemble-SVM performs similar with single-SVM, ensemble-FNN significantly outperforms single-FNN, and ensemble-LSTM underperforms single-LSTM.

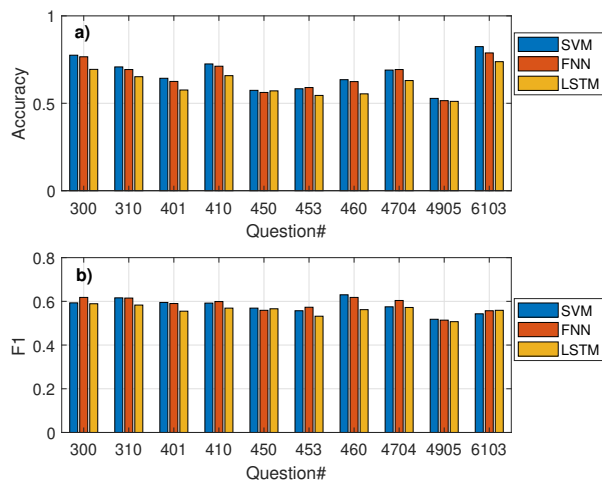


Fig. 4. Visualization of a) accuracy and b) F1 scores of ensemble models with different base-models.

Table 5 quantitatively demonstrates the improvement percentage from a single baseline model to an ensemble model, and Figure 5 visualizes the results. In average, ensemble-SVM improves single-SVM by -0.02% in accuracy and 0.03% in F1. Ensemble-FNN improves single-FNN by 6.45% in accuracy and 3.73% in F1. And ensemble-LSTM improves single-LSTM by -5.59% in accuracy and -2.76% in F1. As shown in Table 5, we can see that with ensemble learning, the classification accuracy/F1 for FNN has been significantly improved. However, ensemble learning shows little improvement on SVM, and the performance of LSTM is deteriorated after the ensemble algorithm. In the future, we will further investigate the results and try to improve the classification accuracy.

Quest.	Improv. in Acc. (%)			Improv. in F1. (%)		
	SVM	FNN	LSTM	SVM	FNN	LSTM
300	0.42	6.71	-9.88	1.94	5.66	-3.13
310	0.32	5.70	-6.67	0.75	10.08	-6.52
401	0.62	4.04	-8.97	0.17	1.25	-6.82
410	-0.26	2.30	-5.33	-0.04	0.18	-4.41
450	-1.34	5.17	-1.74	-2.03	4.89	-1.89
453	-0.06	9.35	-4.78	-0.34	7.37	-3.88
460	-0.15	1.67	-5.76	0.03	1.45	-7.58
4704	0.65	5.29	1.23	3.13	1.42	-4.91
4905	0.00	7.24	-4.02	-1.53	4.71	-3.13
6103	-0.37	17.02	-9.99	-1.77	0.33	14.73
Avrg.	-0.02	6.45	-5.59	0.03	3.73	-2.76

Table 5. Improvement in Accuracy/F1 scores of the ensemble models with base models of SVM, FNN, and LSTM.

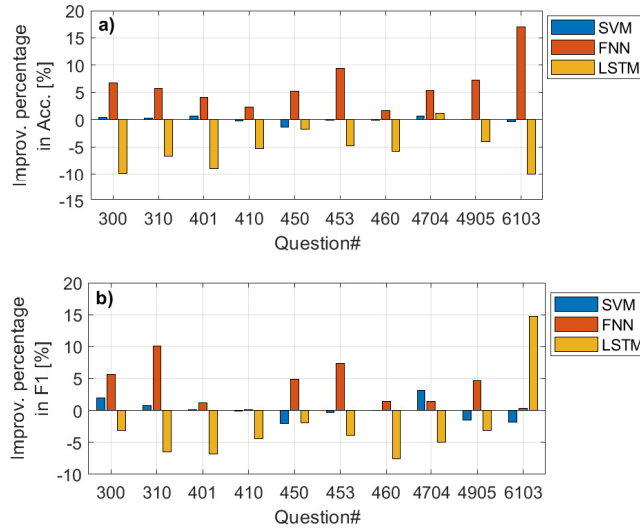


Fig. 5. Visualization of the improvement percentage in a) accuracy and b) F1 scores of ensemble models with base models of SVM, FNN, and LSTM.

5 Conclusion

Customer identification can be of significant importance utility companies to design efficient demand response program. Ensemble learning has shown to be effective for different types of applications. In this paper, we evaluated the performance of ensemble learning with different types of classification models on residential electric customer identification on a real-world dataset. Bagging method has been used to ensemble homogeneous base-models. Experiments have been conducted with base-models of SVM, FNN, and LSTM. Comparison has made between the ensemble models and single models. Results show that Bagging improves the performance of FNN the most, SVM remains the same performance after ensemble, and the performance of LSTM is deteriorated.

References

1. Commission for Energy Regulation (CER): CER smart metering project - electricity customer behaviour trial, 2009-2010 [dataset] (2012), 1st Edition. Irish Social Science Data Archive. SN: 0012-00. <https://www.ucd.ie/issda/data/commissionforenergyregulationcer/>
2. Bauer, E., Kohavi, R.: An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine learning* **36**(1), 105–139 (1999)
3. Beckel, C., Sadamori, L., Santini, S.: Automatic socio-economic classification of households using electricity consumption data. In: *Proceedings of the fourth international conference on Future energy systems*. pp. 75–86 (2013)

4. Beckel, C., Sadamori, L., Staake, T., Santini, S.: Revealing household characteristics from smart meter data. *Energy* **78**, 397–410 (2014)
5. Breiman, L.: Bagging predictors. *Machine learning* **24**(2), 123–140 (1996)
6. Dang, Q., Wu, D., Boulet, B.: An advanced framework for electric vehicles interaction with distribution grids based on q-learning. In: 2019 IEEE Energy Conversion Congress and Exposition (ECCE). pp. 3491–3495. IEEE
7. Dang, Q., Wu, D., Boulet, B.: A q-learning based charging scheduling scheme for electric vehicles. In: 2019 IEEE Transportation Electrification Conference and Expo (ITEC). pp. 1–5. IEEE (2019)
8. Dang, Q., Wu, D., Boulet, B.: Ev charging management with ann-based electricity price forecasting. In: 2020 IEEE Transportation Electrification Conference & Expo (ITEC). pp. 626–630. IEEE (2020)
9. Huang, X., Wu, D., Boulet, B.: Ensemble learning for charging load forecasting of electric vehicle charging stations. In: 2020 IEEE Electric Power and Energy Conference (EPEC). pp. 1–5. IEEE (2020)
10. Jiang, T., Li, J., Zheng, Y., Sun, C.: Improved bagging algorithm for pattern recognition in uhf signals of partial discharges. *Energies* **4**(7), 1087–1101 (2011)
11. Kuncheva, L.I.: Combining pattern classifiers: methods and algorithms. John Wiley & Sons (2014)
12. Lin, W., Wu, D.: Residential electric load forecasting via attentive transfer of graph neural networks. In: IJCAI. pp. 2716–2722. ijcai.org (2021)
13. Lin, W., Wu, D., Boulet, B.: Spatial-temporal residential short-term load forecasting via graph neural networks. *IEEE Trans. Smart Grid* **12**(6), 5373–5384 (2021)
14. Opitz, J., Burst, S.: Macro fl and macro fl. arXiv preprint arXiv:1911.03347 (2019)
15. Sagi, O., Rokach, L.: Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **8**(4), e1249 (2018)
16. Wang, Y., Bennani, I.L., Liu, X., Sun, M., Zhou, Y.: Electricity consumer characteristics identification: A federated learning approach. *IEEE Transactions on Smart Grid* (2021)
17. Wang, Y., Chen, Q., Gan, D., Yang, J., Kirschen, D.S., Kang, C.: Deep learning-based socio-demographic information identification from smart meter data. *IEEE Transactions on Smart Grid* **10**(3), 2593–2602 (2018)
18. Wang, Y., Chen, Q., Kang, C., Xia, Q., Luo, M.: Sparse and redundant representation-based smart meter data compression and pattern extraction. *IEEE Transactions on Power Systems* **32**(3), 2142–2151 (2016)
19. Wu, D.: *Machine Learning Algorithms and Applications for Sustainable Smart Grid*. McGill University (Canada) (2018)
20. Wu, D., Wang, B., Precup, D., Boulet, B.: Boosting based multiple kernel learning and transfer regression for electricity load forecasting. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 39–51. Springer (2017)
21. Wu, D., Wang, B., Precup, D., Boulet, B.: Multiple kernel learning-based transfer regression for electric load forecasting. *IEEE Transactions on Smart Grid* **11**(2), 1183–1192 (2019)
22. Wu, D., Zeng, H., Boulet, B.: Neighborhood level network aware electric vehicle charging management with mixed control strategy. In: 2014 IEEE International Electric Vehicle Conference (IEVC). pp. 1–7. IEEE (2014)
23. Wu, D., Zeng, H., Lu, C., Boulet, B.: Two-stage energy management for office buildings with workplace ev charging and renewable energy. *IEEE Transactions on Transportation Electrification* **3**(1), 225–237 (2017)

24. Yan, S., Li, K., Wang, F., Ge, X., Lu, X., Mi, Z., Chen, H., Chang, S.: Time-frequency feature combination based household characteristic identification approach using smart meter data. *IEEE Transactions on Industry Applications* **56**(3), 2251–2262 (2020)
25. Zhang, C., Ma, Y.: *Ensemble machine learning: methods and applications*. Springer (2012)
26. Zhong, S., Tam, K.S.: Hierarchical classification of load profiles based on their characteristic attributes in frequency domain. *IEEE Transactions on Power Systems* **30**(5), 2434–2441 (2014)