An Ensemble Forecasting for Aggregated Load

Journal:	*IEEE Power Engineering Letters*
Manuscript ID	PESL-00216-2017
Manuscript Type:	Letters
Date Submitted by the Author:	31-Oct-2017
Complete List of Authors:	Wang, Yi; Tsinghua University, Department of Electrical Engineering Chen, Qixin; State Key Lab of Power Systems, Electrical Engineering Department, Tsinghua University Sun, Mingyang; Imperial College London, Electrical & Electronic Engineering Kang, Chongqing; power system institute, Electrical Engineering Department, Tsinghua University Xia, Qing; Tsinghua University, Electrical Engineering
Key Words:	aggregated load forecasting, ensemble forecasting, hierarchical clustering, smart meter data, sub-load profile

An Ensemble Forecasting for Aggregated Load

Yi Wang, *Student Member, IEEE,* Qixin Chen, *Senior Member, IEEE,* Mingyang Sun, *Member, IEEE,* Chongqing Kang, *Fellow, IEEE,* and Qing Xia, *Senior Member, IEEE*

***Abstract*—Traditional load forecasting is typically performed based on its historical load data and relevant factors. With the prevalence of smart meters, fine-grained sub-load profiles provide a valuable opportunity to have a better understanding of aggregated load patterns and further improve the forecasting accuracy. In this paper, a novel ensemble method is proposed to forecast aggregated load based on hierarchical clustering. First, the hierarchical structure of the sub-load profiles is established using hierarchical clustering method, thus constructing different groups of sub-load profiles when the number of clusters is determined. Then, forecasting is conducted on the grouped load profiles individually and these forecasts are summed to form the aggregated load forecast. In this way, different aggregated load forecasts can be obtained by varying the number of clusters. Finally, an optimal weighted ensemble approach is employed to combine these forecasts and provide the final forecasting result. Based on two open datasets, residential load from Irish Customer Behavior Trials (CBTs) and distribution zone substation load from Ausgrid, case studies are conducted to verify the effectiveness and superiority of the proposed method.**

***Index Terms*—aggregated load forecasting, ensemble forecasting, hierarchical clustering, smart meter data, sub-load profile.**

I. Introduction

TRADITIONAL load forecasting is performed based on its historical load data and relevant factors. Recent advances in load forecasting include probabilistic forecasting, hierarchical forecasting, ensemble forecasting, and etc [1]. With the widespread popularity of smart meters, more and more fine-grained sub-load profiles can be measured and collected. Consequently, research on individual load forecasting has also been investigated in [2]. For aggregated load forecasting, a bottom-up approach, implemented based on the smart meter data, is proposed in [3]: forecast them individually and then aggregated the results. To improve the efficient of the forecasting procedure, a clustering based aggregated load forecasting is proposed in [4]: different groups of consumers are first constructed based on their load patterns; afterwards, forecast the load of each group separately; finally, sum the forecasts of different groups to obtain the aggregated load forecast. The optimal number of clusters is determined by cross validation. The results demonstrate that the clustering-based method outperforms the direct forecasting method.

Beyond the aforementioned single-output forecasting methods (i.e. only provide one final forecast value), a series of works have been done on ensemble forecasting methods, which can output multiple forecasts from different models [5]. In general, ensemble forecasting can be classified as homogeneous and heterogeneous methods such as bootstrap aggregating methods and the combination of SVM and ANN [6]. This paper tries to answer the following question: *Is it possible to utilize both ensemble techniques and fine-grained sub-load profiles to further improve the forecasting accuracy?* Unlike the work in [4], we vary the number of clusters to obtain multiple aggregated load forecasts instead of a single forecast. Subsequently, an optimally weighted ensemble approached is used to combine these forecasts and provide the final result. As stated above, the key contributions of this paper are threefold:

1) A novel ensemble forecasting framework is proposed for aggregated load profile. It produces multiple forecasts by varying the number of clusters which is quite different from traditional ensemble methods.
2) Instead of finding the optimal number of clusters, the proposed method searches optimal combination of multiple forecasts and can be flexibly applied to different datasets.
3) Case studies are conducted on two sets of real data (residential and substation load) to verify the effectiveness and superiority of the proposed method.

The remainder of this paper is organized as follows: Section II introduces the proposed clustering based ensemble method for aggregated load forecasting; Section III conducts case studies on Irish residential load data and Ausgrid substation load data, respectively; Section IV draws the conclusion and envisions future works.

II. Proposed Methodology

The target of this paper is to forecast the aggregated load, which contains M sub-load profiles. Let L_t and $L_{m,t}$ denote the total load and the m-th sub-load at time t, the matrix form of the sub-load profiles can be represented as $\mathbf{L}_{M\times T}$. To highlight the idea of the proposed method, only historical load data is employed as input features for constructing the forecasting model. Note that other relevant factors (e.g. temperature) can also be considered in the proposed framework. In general, the proposed method can be divided into three main stages: clustering stage, forecasting stage, and ensemble stage.

1) Clustering Stage: This stage is to establish the hierarchical structure of consumers according to the similarities of their consumption behaviors. First, the representative load profile $L_{m,r,t}$ for each consumer is obtained by normalizing the calculated average weekly load profile to $[0,1]$ domain. The subscript r means representative load here. Thus, the distance matrix $\mathbf{D}_{M\times M}$ among these consumers can be calculated based on Euclidean distance:

$$D_{m,n} = \Big(\sum_{t=1}^{T_W}(L_{m,r,t} - L_{n,r,t})^2\Big)^{\frac{1}{2}} \quad (1)$$

where T_W denotes the number of the time period over one week. It is important to notice that, in this stage, a large

This work was supported by National Key R&D Program of China (No. 2016YFB0900100) (*Corresponding author: Chongqing Kang*)
Y. Wang, Q. Chen, C. Kang and Q. Xia are with the State Key Lab of Power Systems, Dept. of Electrical Engineering, Tsinghua University, Beijing 100084, China. (E-mail: cqkang@tsinghua.edu.cn).
M. Sun is with the Department of Electrical & Electronic Engineering, Imperial College London, London, SW7 2AZ, UK.

number of clustering procedures are required to be performed on different numbers of groups. Therefore, in this research, the agglomerative hierarchical clustering method with single linkage is selected to cluster the customers because of its capability to establish the hierarchical structure and the fact that it does not need to be performed repeatedly.

2) Forecasting Stage: The purpose of this stage is to produce multiple forecasts by varying the number of clusters. When the number of clusters is M, the forecasting is essentially the bottom-up approach; when the number of clusters is 1, the forecasting is performed directly based on historical aggregated load data. In order to diversify the forecasting results, we vary the number of clusters exponentially. Thus, a total of N forecasts will obtained:

$$N = \left\lfloor log_2 M \right\rfloor + 1 \tag{2}$$

where $\lfloor \cdot \rfloor$ denotes the round-down function. For example, $N = 7$ when $M = 100$. The n-th forecast is obtained by summing the forecasts of k_n grouped load profiles, where k_n is expressed as follows:

$$k_n = \min \left\{ 2^{n-1}, M \right\} \tag{3}$$

For example, the set of cluster number is $\mathbf{K} = [1, 2, 4, 8, 16, 32, 64, 100]$ when $M = 100$.

Without loss of generality, one of the most widely used forecasting models, Artificial Neural Network (ANN), is applied to forecast different groups of load profiles. The input of ANN is the lagged values (h denotes the number of time period each day) and calender variables:

$$\mathbf{X}_t = [Week, Hour, L_{t-h}, L_{t-h-1}, L_{t-2h+1}, L_{t-2h}, L_{t-3h}] \tag{4}$$

3) Ensemble Stage: As one of the main contributions in this work, ensemble stage is proposed to calculate the weights $\boldsymbol{\omega}$ for the N forecasts and combine them into final forecast. The ensemble of N forecasts is formulates as an optimization problem where the objective function is to minimize the mean absolute percent error (MAPE) and the constraints include the equations of the combined forecasts, the summation of all the weights, and non-negativity of the weights.

$$\begin{aligned} \hat{\boldsymbol{\omega}} &= \arg\min_{\boldsymbol{\omega}} \sum_{t=1}^{T} \frac{1}{T} \frac{|L_t - \hat{L}_t|}{L_t} \\ s.t.\ \hat{L}_t &= \sum_{n=1}^{N} \omega_n \hat{L}_{n,t}, \quad \sum_{n=1}^{N} \omega_n = 1, \quad \omega_n \geq 0 \ \ \forall n. \end{aligned} \tag{5}$$

The absolute percent error in the objective function can be easily transformed into linear programming (LP) problem by introducing auxiliary decision variables v_t, as follows:

$$\begin{aligned} \hat{\boldsymbol{\omega}} &= \arg\min_{\boldsymbol{\omega}} \sum_{t=1}^{T} \frac{1}{T} \frac{v_t}{L_t} \\ s.t.\ \hat{L}_t &= \sum_{n=1}^{N} \omega_n \hat{L}_{n,t}, \quad \sum_{n=1}^{N} \omega_n = 1, \quad \omega_n \geq 0 \ \ \forall n. \\ v_t &\geq L_t - \hat{L}_t, \quad v_t \geq \hat{L}_t - L_t \end{aligned} \tag{6}$$

4) Whole Algorithm: The whole procedures of the proposed method are presented in **Algorithm 1**. To evaluate the performance of the proposed method, the aggregated and sub-load profiles are divided into three parts: the first part $\mathbf{L}_{tr}$ is used to train the forecasting model for each group load profile; the second part $\mathbf{L}_{en}$ is used to calculate the weights $\boldsymbol{\omega}$ for ensemble; the third part $\mathbf{L}_{te}$ is used to test the performance of the aggregated load ensemble forecasting model.

Algorithm 1 Aggregated Load Ensemble Forecasting

Require: Segmented sub-load profiles $\mathbf{L}_{tr}$, $\mathbf{L}_{en}$, and $\mathbf{L}_{te}$ for training, ensemble, and test forecasting models; set of cluster numbers $\mathbf{K} = [k_1, k_2, ...k_n, ...k_N]$.
Clustering Stage (based on $\mathbf{L}_{tr}$):
Obtain normalized representative weekly load profile for each consumer $L_{m,r,t}$;
Calculate the distance matrix $\mathbf{D}$ among the consumers;
Implement agglomerative hierarchical clustering.
Forecasting Stage (based on $\mathbf{L}_{tr}$ and $\mathbf{L}_{en}$):
for $n = 1 : N$ **do**
  Cluster the sub-load profiles into k_n groups;
  for $j = 1 : k_n$ **do**
    Train the forecasting model f_j for the j-th group based on $\mathbf{L}_{tr}$;
    Forecast the j-th grouped load profiles $\hat{L}_j$ for $\mathbf{L}_{en}$;
  end for
  Calculate the sum of the forecasts of grouped load $\hat{L}_n = \sum_{j=1}^{k_n} \hat{L}_j$.
end for
Ensemble Stage (based on $\mathbf{L}_{en}$):
Solve the optimization problem shown in Eq. 5.
Test Stage (based on $\mathbf{L}_{te}$):
Forecast the load profile in $\mathbf{L}_{te}$ and calculate the MAPE and RMSE;

III. CASE STUDY

In this section, case studies are conducted on two open datasets. In particular, 50%, 25% and 25% of the whole dataset are partitioned into training dataset, test dataset, and ensemble dataset, respectively.

A. Irish Residential Load Data

Residential load data obtained via the Irish Social Science Data Archive (ISSDA) contains over 5000 Irish homes and businesses [7]. After excluding the consumers with large number of zero values, the data of a total of 5237 consumers from July 20, 2009 to December 26, 2010 (75 weeks) are used for forecasting. Fig. 1 shows the weekly predicted and real load profiles. As shown in the figure, the dotted lines are individual forecasts; the blue and red line are the ensemble forecast and actual value, respectively. Table I provides the weights, MAPE, and RMSE of individual forecasts. Regarding the individual forecasts, it can be seen that, instead of using the clustering strategy (i.e. $N > 1$), direct load forecasting based on the aggregated data (i.e. $N = 1$) exhibits the best performance. Nevertheless, the superior performance of the proposed ensemble method can be indicated by the 4.71% and 3.83% lower MAPE and RMSE values, respectively, than

TABLE I
PERFORMANCE OF INDIVIDUAL AND ENSEMBLE FORECASTS FOR IRISH DATASET

N	1	2	4	8	16	32	64	128	256	...	5237	Ensemble
ω	0.634	0	0	0.271	0	0	0.095	0	0	...	0	/
MAPE	***4.25%***	5.05%	5.29%	4.74%	5.55%	4.66%	4.79%	5.09%	5.59%	...	10.31%	**4.05%**
RMSE	***210.95***	229.73	228.01	217.68	244.9	217.64	227.36	232.61	250.27	...	441.33	**202.88**

TABLE II
PERFORMANCE OF INDIVIDUAL AND ENSEMBLE FORECASTS FOR AUSGRID DATASET

N	1	2	4	8	16	32	64	128	155	Ensemble
ω	0	0	0	0	0.113	0	0	0	0.887	/
MAPE	5.68%	5.59%	5.47%	5.27%	5.15%	5.19%	5.13%	5.12%	***5.09%***	**5.08%**
RMSE	223.23	217.4	215.47	208.21	203.91	206.3	204.66	202.73	***202.65***	**202.55**

those of the best individual forecast method. Results also show that the performance of bottom-up approach is much worse than clustering-based method due to the large variation of individual load profiles.

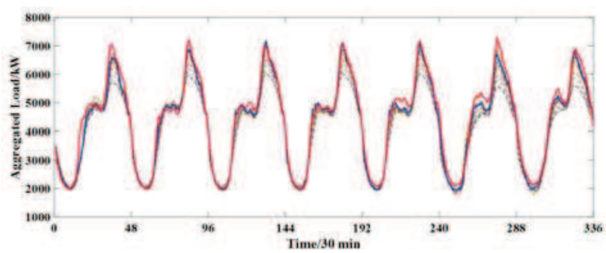

Fig. 1. Predicted and Real Aggregated Individual Load Profiles

B. Ausgrid Substation Load Data

We use the Ausgrid substation load data from May 5, 2014, to April 24, 2016 (103 weeks). After deleting the substations with a large number of non-value, a total of 155 substations data are retained [8]. Thus, nine individual forecasts are obtained by varying the number of clusters. The predicted load profiles and performances are shown in Fig. 2 and Table II, respectively. After the optimization procedure, the weights for forecasts #5 and #9 are 0.113 and 0.887 respectively, whereas the weights for other forecasts are zeros. When comparing the calculated MAPE and RMSE values, it is very interesting to find that, in contrast to Irish dataset, the bottom-up approach (i.e. N=155) have the lowest forecasting errors. The reason of this phenomenon might be that the substation load profiles are more regular than residential load profiles.

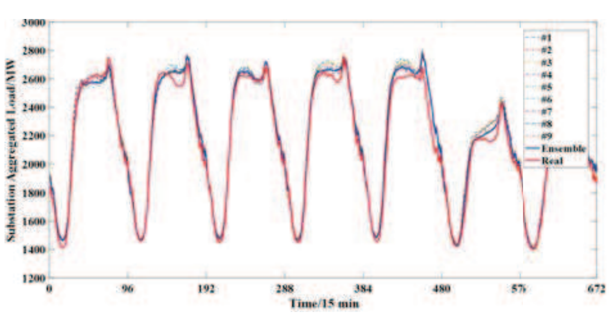

Fig. 2. Predicted and Real Aggregated Substation Load Profiles

IV. CONCLUSION AND FUTURE WORKS

This paper proposes an ensemble forecasting method for aggregated load profile using hierarchical clustering and based on fine-grained sub-load profiles. It is a new way to make full advantages of fine-grained data to further improve the forecasting accuracy of the aggregated load. Case studies on both residential load data and substation load data demonstrate the superior performance of the proposed ensemble method when comparing with the traditional direct or bottom-up forecasting strategies. Further research could focus on extending the ensemble method from point load forecasting to probabilistic load forecasting based on hierarchical sub-load profiles.

REFERENCES

[1] T. Hong and S. Fan, “Probabilistic electric load forecasting: A tutorial review,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.

[2] C.-N. Yu, P. Mirowski, and T. K. Ho, “A sparse coding approach to household electricity demand forecasting in smart grids,” *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 738–748, 2017.

[3] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett, “Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting,” *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1591–1598, 2017.

[4] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, “Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities,” *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, 2015.

[5] S. Li, P. Wang, and L. Goel, “A novel wavelet-based ensemble method for short-term load forecasting with hybrid neural networks and feature selection,” *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1788–1798, 2016.

[6] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, “Ensemble approaches for regression: A survey,” *ACM Computing Surveys (CSUR)*, vol. 45, no. 1, p. 10, 2012.

[7] Irish Social Science Data Archive, “Commission for energy regulation (cer) smart metering project.” http://www.ucd.ie/issda/data/commissionforenergyregulationcer/, 2012.

[8] Ausgrid, “Distribution zone substation information data to share,” http://www.ausgrid.com.au/Common/About-us/Corporate-information/Data-to-share/DistZone-subs.aspx#.WYD6KenauUl, accessed July 31, 2017.