Residential Electricity Load Forecasting: Evaluation of Individual and Aggregate Forecasts (Supplementary Material)

Tri Kurniawan Wijaya, Matteo Vasirani, Samuel Humeau, and Karl Aberer

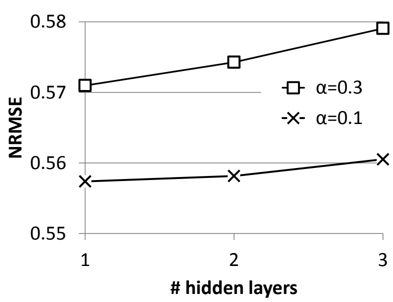

Fig. 1: MLP model evaluation (using NRMSE) using different number of hidden layers and learning rate α on randomly chosen 25 households. The lower the better. For experiments in the main paper, we use one hidden layer and $\alpha = 0.1$.

I. MLP Model Evaluation

Figure 1 shows the MLP model evaluation using:

- learning rate $\alpha = 0.1$ and $\alpha = 0.3$,
- one, two, and three hidden layers, where each hidden layer consists of $(\#features + 1)/2$ neurons.

For experiments in the main paper, we choose the best configuration, i.e., one hidden layer with $\alpha = 0.1$.

II. SVR Model Evaluation

Figure 2 shows SVR model evaluation for individual forecasts using different parameters settings (as described in Section IV-A in the main paper). For experiments in the main paper, we use $C = 100$ and $\gamma = 0.01$. While there are some other combination of C and γ (or settings) which yield better NRMSE, these settings typically require considerably longer running time.

Figure 3 shows the performance for aggregated forecasts. In the main paper, for aggregate forecasts, we use $C = 1000$ and $\gamma = 1$.

C \ γ	0	0.01	0.1	1
1.E+00	0.69	0.62	0.59	0.57
1.E+01	0.62	0.59	0.58	0.57
1.E+02	0.59	**0.58**	0.57	0.58
1.E+03	0.58	0.58	0.57	0.63
1.E+04	0.58	0.57	0.58	0.82
1.E+05	0.57	0.57		

(a)

C \ γ	0	0.01	0.1	1
1.E+00	0.02	0.03	0.02	0.02
1.E+01	0.03	0.02	0.02	0.20
1.E+02	0.02	**0.02**	0.11	0.02
1.E+03	0.02	0.02	0.02	0.02
1.E+04	0.02	0.02	0.02	0.04
1.E+05	0.02	0.11		

(b)

C \ γ	0	0.01	0.1	1
1.E+00	2.1	2.7	3.2	4.6
1.E+01	2.3	2.6	3.1	3.2
1.E+02	2.3	**2.4**	4.0	8.1
1.E+03	2.7	3.0	8.9	45
1.E+04	4.0	8.2	50	222
1.E+05	7.2	52		

(c)

Fig. 2: SVR model evaluation for individual forecasts (as described in Section IV-A in the main paper) on the randomly chosen 25 households: (a) average NRMSE on the validation set given different C and γ, (b) standard deviation on the average, (c) running time. In all of them, the lower the better. While there are some settings which yield better NRMSE, we choose $C = 100$ and $\gamma = 0.01$ because it is much faster.

C \ γ	0.001	0.01	0.1	1	10
1.E-01	0.418	0.415	0.395	0.356	0.318
1.E+00	0.415	0.392	0.226	0.160	0.201
1.E+01	0.391	0.200	0.074	0.070	0.082
1.E+02	0.197	0.073	0.064	0.052	0.082
1.E+03	0.073	0.066	0.059	**0.045**	0.065
1.E+04	0.066	0.065	0.054	**0.045**	0.072
1.E+05	0.066	0.061	0.050	0.050	
1.E+06	0.065	0.057	0.047	0.063	
1.E+07	0.061	0.054	0.046	0.087	
1.E+08	0.063	0.110	0.188	0.149	

Fig. 3: SVM model evaluation (measured by average NRMSE) for aggregate forecasts (as described in Section IV-A in the main paper). The lower the better.

III. Leader-Follower Features

A. Cross-correlation

The cross-correlation between two time series $S^{(1)} = (s_1^{(1)}, ..., s_N^{(1)})$ and $S^{(2)} = (s_1^{(2)}, ..., s_N^{(2)})$ for a certain lag τ is given by:

$$CC(S^{(1)}, S^{(2)}, \tau) = \frac{\mathbb{E}[(s_t^{(1)} - \mu(S^{(1)})) \cdot (s_{t-\tau}^{(2)} - \mu(S^{(2)}))]}{\sigma(S^{(1)}) \cdot \sigma(S^{(2)})} \quad (1)$$

where $\mathbb{E}$ is the expectation operator, μ is the mean and σ is the standard deviation of the time series. Since the goal is a short-term (1h and 24h ahead) prediction, we can set the delay τ to a small value different than 0. In this work, for each pair of households (j,k), we compute the cross-correlation between

T. K. Wijaya, M. Vasirani, S. Humeau, and K. Aberer are with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: {tri-kurniawan.wijaya, matteo.vasirani, samuel.humeau, karl.aberer}@epfl.ch).

MCC

id house	1002	1014	1018	1022	1027	...	6817	...
1002	0.59	0.20	0.11	0.21	0.14	...	0.22	
1014		0.58	0.16	0.20	0.23	...	0.27	
1018			0.67	0.18	0.16	...	0.22	
1022				0.33	0.22	...	0.18	
1027					0.46	...	0.36	
...						...	...	
6817							0.45	

LCC

house	1002	1014	1018	1022	1027	...	6817	...
1002	1	1	1	1	1	...	1	
1014		1	1	1	2	...	1	
1018			1	1	1	...	1	
1022				1	1	...	-1	
1027					1	...	-1	
...						...	...	
6817							1	

Fig. 4: MCC and LCC matrices.

their consumption time series for $\tau \in \{-4, \ldots, -1, 1, \ldots, 4\}$. We then determine the lag $\tau_{j,k}$ with the highest correlation between the two series. A positive value of $\tau_{j,k}$ means that household k leads the consumption of household j. Therefore the consumption of household k at time t contains some information about the future consumption of household j at time $t + \tau$ that can be exploited by the forecasting algorithm.

We store the results in two symmetric matrices, namely the maximum cross-correlation (MCC) matrix and the lag of cross correlation (LCC) matrix. MCC(j, k) contains the maximum cross-correlation between household j and k, while LCC(j, k) contains the corresponding lag $\tau_{j,k}$, for which the cross-correlation is maximal.

Figure 4 shows a portion of the 782 $\times$ 782 MCC and LCC matrices (for 782 households in our experiments). On the diagonal, the MCC matrix contains the correlation between the consumption of a household and the consumption of the same household with lag τ (i.e., auto-correlation). In most cases, a household's consumption is more closely correlated with its own past consumption than with that of any other household. However, in some cases there exists a leader-follower relation between households. For example, household 1027 seems to lead the household 6817, since the correlation between the consumption of household 1027 at time $t-1$ and the consumption of household 6817 is 0.36, which is quite close to 0.45, the auto-correlation with lag 1. See also Figure 5 in the main paper for different illustration.

B. Mutual Information

Determining leaders using cross-correlations means that we are looking for leader-follower relations among households with very similar consumption, but with a delay. However, with the objective of making predictions using sophisticated machine learning algorithms, we are interested in any type of relations between two households consumption time series. Therefore, mutual information could be a more appropriate measure to quantify such relations.

Mutual information expresses the quantity of information that links two random variables. Let X and Y be two discrete random variables. The mutual information between X and Y is defined as:

$$I(X, Y) = H(X) - H(X \mid Y) = H(X) + H(Y) - H(X, Y) \tag{2}$$

where $H(X)$ is the entropy of variable X, $H(X \mid Y)$ is the conditional entropy and $H(X, Y)$ is the joint entropy between X and Y. To generate a random variable X, given a consumption time series S, we discretize S using 8 bits uniform discretization and consider it as a random variable X evaluated each hour. For other alternatives for discretizing energy consumption data, we refer the interested reader to the work by Wijaya et al. [1].

Let $\mathcal{V}^{(1)}$ (resp. $\mathcal{V}^{(2)}$) the set of distinct values that appear in the discretized time series $\widehat{S}^{(1)} = (\widehat{s}_1^{(1)}, ..., \widehat{s}_N^{(1)})$ (resp. $\widehat{S}^{(2)}$). The mutual information between $\widehat{S}^{(1)}$ and $\widehat{S}^{(2)}$ for a certain lag τ is given by:

$$\begin{aligned} I(\widehat{S}^{(1)}, \widehat{S}^{(2)}, \tau) = \\ \sum_{v^{(1)} \in \mathcal{V}^{(1)}} & -P(\widehat{S}_t^{(1)} = v^{(1)}) \cdot \log_2(P(\widehat{S}_t^{(1)} = v^{(1)})) + \\ \sum_{v^{(2)} \in \mathcal{V}^{(2)}} & -P(\widehat{S}_{t-\tau}^{(2)} = v^{(2)}) \cdot \log_2(P(\widehat{S}_{t-\tau}^{(2)} = v^{(2)})) - \\ \sum_{\substack{v^{(1)} \in \mathcal{V}^{(1)} \\ v^{(2)} \in \mathcal{V}^{(2)}}} & -P(\widehat{S}_t^{(1)} = v^{(1)}, \widehat{S}_{t-\tau}^{(2)} = v^{(2)}) \\ & \cdot \log_2(P(\widehat{S}_t^{(1)} = v^{(1)}, \widehat{S}_{t-\tau}^{(2)} = v^{(2)})). \end{aligned} \tag{3}$$

Similar as before, we store the results in two symmetric matrices, namely the maximum mutual information (MMI) matrix and the lag of mutual information (LMI) matrix. MMI(j, k) contains the maximum mutual-information between household j and k, while LMI(j, k) contains the corresponding lag $\tau_{j,k}$, for which the mutual information is maximal.

References

[1] T. K. Wijaya, J. Eberle, and K. Aberer, "Symbolic representation of smart meter data," in *Proceedings of the Joint EDBT/ICDT 2013 Workshops*, ser. EDBT '13. ACM, 2013, pp. 242–248.