Technical University of Denmark

Demand Forecasting at Low Aggregation Levels using Factored Conditional Restricted Boltzmann Machine.

Mocanu, Elena; Nguyen, Phuong H.; Gibescu, Madeleine; Larsen, Emil Mahler; Pinson, Pierre

Published in:

Link to article, DOI:
10.1109/PSCC.2016.7540994

Publication date:
2016

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):

DTU Library
Technical Information Center of Denmark

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Demand Forecasting at Low Aggregation Levels using Factored Conditional Restricted Boltzmann Machine

Elena Mocanu
Phuong H. Nguyen
Madeleine Gibescu
Department of Electrical Engineering
Eindhoven University of Technology, Netherlands
{e.mocanu, p.nguyen.hong, m.gibescu}@tue.nl

Emil Mahler Larsen
Pierre Pinson
Department of Electrical Engineering
Technical University of Denmark,
Lyngby, Denmark
{emlar, ppin}@elektro.dtu.dk

Abstract—The electrical demand forecasting problem can be regarded as a nonlinear time series prediction problem depending on many complex factors since it is required at various aggregation levels and at high temporal resolution. To solve this challenging problem, various time series and machine learning approaches have been proposed in the literature. As an evolution of neural network-based prediction methods, deep learning techniques are expected to increase the prediction accuracy by allowing stochastic formulations and bi-directional connections between neurons. In this paper, we investigate a newly developed deep learning model for time series prediction, namely Factored Conditional Restricted Boltzmann Machine (FCRBM), and extend it for electrical demand forecasting. The assessment is made on the EcoGrid dataset, originating from the Bornholm island experiment in Denmark, consisting of aggregated electric power consumption, local price and meteorological data collected from 1900 customers. The households are equipped with local generation and smart appliances capable of responding to real-time pricing signals. The results show that for the short-term (5 minute to 1 day ahead) prediction problems solved here, FCRBM outperforms the benchmark machine learning approach, i.e. Support Vector Machine.


I. INTRODUCTION

The electrical demand forecasting problem, at various aggregation levels, can be regarded as a highly nonlinear time series prediction problem. The complexity of the consumers' energy producing and consuming technologies and the uncertainty in the influencing factors, yield frequent fluctuations. Traditionally, the short-term forecasting problem is referring to 1 hour and 15 minutes resolutions, but higher resolutions make the problem even more complicated. Moreover, urbanization and electrification trends show that the total energy demand will increase in the future, and the penetration of energy from renewable sources is increasing as well. Future smart grids need a system that can monitor, predict, schedule, learn and make decisions regarding local energy consumption and production in real time. Modeling and predicting energy consumption in smart buildings can provide valuable information to facilitate Demand Response (DR) or Demand Side Management (DSM) programs.

This paper focuses on Deep Learning methods for electrical energy demand prediction, with an application to the aggregated profiles collected from the Danish Island Bornholm within the EcoGrid project [14]. Due to the fact that energy consumption can be seen as a time series problem, it investigates the application of Factored Conditional Restricted Boltzmann Machines (FCRBM) [15], recently introduced stochastic machine learning methods which were used successfully until now to model highly non-linear time series (e.g. human motion style, structured output prediction) [15]–[17]. Consequently,

This research has been partly funded by AgentschapNL - TKI Switch2SmartGrids of Dutch Top Sector Energy. The authors would like to thank to our EcoGrid EU partners and DMI for providing the meteorological dataset.
we adapt the FCRBM architecture for demand forecasting problems by merging the style and feature labels into one, and by rewriting the equations and the derivatives of the learning rules according to the new configuration of the model. As a secondary contribution, we analyze how external factors (e.g., weather conditions, electricity prices) can be used to improve the forecasting accuracy and we propose the use of a Gaussian Restricted Boltzmann Machine to perform feature extraction in a fully automated manner and to reduce their dimensionality.

The remainder of this paper is organized as follows. Section II provides some background knowledge on unsupervised learning with Restricted Boltzmann Machines and Section III presents our proposed method using the FCRBM, including the adaptions necessary for demand prediction. Section IV describes the methodology and data description, followed by Section V where the experiment and results are detailed. Finally, Section VI concludes the paper and presents directions for future research.

II. BACKGROUND

Literature provides a wide range of techniques that can solve the demand forecasting problem. The electrical demand has a non-linear and non-stationary profile, which favours a probabilistic approach. In general, we attempt to model the probability of a data point, \( x \) using a function of the form \( f(x; \theta) \), where \( \theta \) is a vector of model parameters. Learning the model parameters, \( \theta \), can be done by maximizing the probability of a training set of data, or equivalently and often more convenient, by minimizing the negative log \( \log p(x; \theta) \). This is not always a trivial task. In the context of our proposed method, we used another common method to learn the parameters of the model by minimizing the Kullback-Leibler (KL) divergence between the empirical and the approximated distributions of the model, as follows:

\[
\min_{\Theta} \mathbb{KL}(p_{\text{model}}(V|\Gamma; \Theta) || p_{\text{empirical}}(V|\Gamma))
\]  

(1)

where \( \Gamma \) represents the total input set and \( V \) is the total output set. The rest of this section presents the background knowledge useful to the reader to understand the remaining of the paper.

A. Restricted Boltzmann machine

Restricted Boltzmann Machines (RBMs) [18] have been applied in different machine learning fields including, multi-class classification [19], collaborative filtering [20], among others. They are energy-based models for unsupervised learning. These models have stochastic nodes and layers, making them less vulnerable to local minima [15]. Further, due to their multiple layers and neural configurations, RBMs possess excellent generalisation capabilities [13]. Formally, an RBM consists of visible and hidden binary layers. The visible layer represents the data, while the hidden one increases the learning capacity by enlarging the class of distributions that can be represented to an arbitrary complexity [15]. This paper follows a standard notation where \( i \) represents the indices of the visible layer, \( j \) those of the hidden layer, and \( w_{i,j} \) denotes the weight connection between the \( i^{th} \) visible and \( j^{th} \) hidden unit. Further, \( a_i \) and \( b_j \) denote the state of the \( i^{th} \) visible and \( j^{th} \) hidden unit, respectively. According to the above definitions, the energy function\(^1\) of an RBM is given by:

\[
E(v, h) = - \sum_{i,j} v_i h_j w_{ij} - \sum_i v_i a_i - \sum_j h_j b_j
\]  

(2)

where, \( a_i \) and \( b_j \) represent the biases of the visible and hidden layers, respectively. The joint probability of a state of the hidden and visible layers is defined as: \( P(v, h) = \exp(-E(v, h)) / Z \) with \( Z = \sum_{x,y} \exp(-E(x, y)) \). To determine the probability of a data point represented by a state \( v \), the marginal probability is used. This is determined by summing out the state of the hidden layer as: \( p(v) = \sum_h P(v, h) = \frac{\sum_h \exp(-\sum_{i,j} v_i h_j w_{ij} - \sum_i v_i a_i - \sum_j h_j b_j)}{Z} \). Parameters are fitted by maximising the likelihood function. In order to maximise the likelihood of the model, the gradients of the energy function with respect to the weights have to be calculated. Usually, in RBMs maximum likelihood cannot be simply applied due to intractability problems. To deal with these problems, Contrastive Divergence, explained next, was introduced.

B. Contrastive Divergence

In Contrastive Divergence (CD) [21], learning follows the gradient of:

\[
CD_n = D_{KL}(p_0(x)||p_{\infty}(x)) - D_{KL}(p_n(x)||p_{\infty}(x))
\]  

(3)

where, \( p_n(.) \) is the distribution of a Markov chain running for \( n \) steps. Since the visible units are conditionally independent given the hidden units and vice versa, learning can be performed using one step Gibbs sampling, which is carried in two half-steps: (1) update all the hidden units, and (2) update all the visible units. Thus, in \( CD_n \) the weight updates are done as follows: \( w_{i,j}^{n+1} = w_{i,j}^n + \alpha (\langle h_j v_i \rangle_{p(h|v,W)} - \langle h_j v_i \rangle_n) \) where \( \tau \) is the iteration, \( \alpha \) is the learning rate, \( \langle h_j v_i \rangle_{p(h|v,W)} = \frac{1}{N} \sum_{k=1}^N \langle h_j v_i \rangle_{p(h|v,W)}^k \) and \( \langle h_j v_i \rangle_n = \frac{1}{N} \sum_{k=1}^N \langle h_j v_i \rangle_{p(h|v,W)}^k \) with \( N \) being the total number of input instances, and the superscript \( (n) \) indicates that the states are obtained after \( n \) iterations of Gibbs sampling from the Markov chain starting at \( p_0(\cdot) \).

III. FACTORED CONDITIONAL RESTRICTED BOLTZMANN MACHINE

This section presents the adapted mathematical details of the proposed method, namely Factored Conditional Restricted Boltzmann Machine (FCRBM) [15], to achieve an accurate and robust prediction at the low aggregation level of electrical energy demand profiles.

\(^1\)Please note that the energy function of RBM should not be confused with the aggregated electrical energy demand.
where

\[ v_{t(k)} \]

\[ v_{t(i)} \]

\[ h_{t(j)} \]

\[ y_{t(p)} \]

as follows:

\[ p(h_{t,j} = 1 | v_{t,i}, v_{t,<i}, y_{t,p}) = \text{sigmoid}(\hat{b}_{j,t} + h_{j,t}^*) \]

\[ p(v_{t,i} = 1 | h_{t,i}, v_{t,<i}, y_{t,p}) = \mathcal{N}(\hat{a}_{i,t} + v_{i,j}^*, \sigma_{i}^2) \]

where \( \mathcal{N}(\mu, \sigma^2) \) denotes the Gaussian probability density function with mean \( \mu \) and variance \( \sigma^2 \).

C. Learning & Update Rules in FCRBM

The general update rule for all the hyper-parameters \( \theta \) is given by:

\[ \theta_{r+1} = \theta_\tau + \rho \Delta \theta_\tau + \alpha (\Delta \theta_{r+1} - \gamma \theta_r) \]

where \( \tau, \rho, \alpha \) and \( \gamma \) represent the update number, momentum, learning rate, and weights decay, respectively. More details regarding the choice of these parameters are described in [22]. The update rules for each of the weights matrices and biases can be computed by deriving the energy function from (2) with respect to each of these variables (i.e., the factored visible weights, factored label weights, factored hidden weights, and the biases of each of the layers), yielding:

1) Weights update: Three update rules corresponding to each of \( W^v, W^h, W^y \) need to be derived. Firstly, the factored visible weights \( W_{ij}^v \) is computed by deriving the total energy function, provide in (4), with respect to \( W_{ij}^v \) is:

\[ \frac{\partial E(v_{t,i}, h_{t,j} | v_{t,<i}, y_{t,p})}{\partial W_{ij}^v} = -v_{i,t} \sum_{j=1}^{n_2} W_{j,i}^h h_{j,t} \sum_{p=1}^{n_3} W_{pj}^y y_{p,t} \]

Secondly, the factored hidden weights \( W_{ij}^h \) are update. Following the same reasoning we obtain:

\[ \frac{\partial E(v_{t,i}, h_{t,j} | v_{t,<i}, y_{t,p})}{\partial W_{ij}^h} = -h_{j,t} \sum_{i=1}^{n_1} W_{i,j}^v v_{i,t} \sum_{p=1}^{n_3} W_{pj}^y y_{p,t} \]

Thirdly, by deriving the total energy function with respect to \( W_{ij}^y \), we obtain the update rule for the factored label weights:

\[ \frac{\partial E(v_{t,i}, h_{t,j} | v_{t,<i}, y_{t,p})}{\partial W_{ij}^y} = -y_{p,t} \sum_{i=1}^{n_1} W_{i,j}^v v_{i,t} \sum_{j=1}^{n_2} W_{j,f}^h h_{j,t} \]

2) Biases update: The derivatives to find the update rules for the parameters which compose the dynamic biases of the present layer (i.e. \( A_{v,m}, A_{h,m}, A_{l,m} \)) are:

\[ \frac{\partial E}{\partial A_{v,m}} = \sum_{k} A_{k,m}^v v_{k,<t} \sum_{p} A_{p,m}^y y_{p,t} \]

\[ \frac{\partial E}{\partial A_{h,m}} = \sum_{t} A_{i,m}^h v_{i,t} \sum_{p} A_{p,m}^y y_{p,t} \]

\[ \frac{\partial E}{\partial A_{l,m}} = \sum_{t} A_{i,m}^l v_{i,t} \sum_{k} A_{k,m}^v v_{k,<t} \]

Figure 1. The general architecture of FCRBM, where \( v_{t,<i} \) is the conditional history layer (input), \( h \) is the hidden layer, \( y \) is the style layer and \( v \) is the visible layer (output). Where \( \circ \) denotes binary neurons, \( \bullet \) represent the real values and the others are Gaussian value.
Further, the derivatives to find the update rules for the parameters which compose the dynamic biases of the hidden layer (i.e. $B_{j,n}^p$, $B_{k,n}^{v,c}$, $B_{l,n}^i$) are presented as:

\[
\frac{\partial E}{\partial B_{j,n}^p} = -h_{j,t} \sum_k B_{k,n}^{v,c} y_{k,<t} \sum_p B_{p,n}^p y_{p,t}
\] (14a)

\[
\frac{\partial E}{\partial B_{k,n}^{v,c}} = -h_{j,t} y_{k,<t} \sum_j B_{j,n}^h \sum_p B_{p,n}^p y_{p,t}
\] (14b)

\[
\frac{\partial E}{\partial B_{l,n}^i} = -h_{j,t} y_{p,t} \sum_k B_{j,n}^h \sum_k B_{k,n}^{v,c} y_{k,<t}
\] (14c)

Using the energy derivative of the hyper parameters and the Contrastive Divergence expression shown in (3), we can calculate the delta rule leading to:

\[
\Delta W \propto \langle \frac{\partial E}{\partial W} \rangle - \langle \frac{\partial E}{\partial W} \rangle_k, \text{ using eq. (10), (11), (12) (15a)}
\]

\[
\Delta A \propto \langle \frac{\partial E}{\partial A} \rangle - \langle \frac{\partial E}{\partial A} \rangle_k, \text{ using eq. (13) (15b)}
\]

\[
\Delta B \propto \langle \frac{\partial E}{\partial B} \rangle - \langle \frac{\partial E}{\partial B} \rangle_k, \text{ using eq. (14) (15c)}
\]

with $k$ being a Markov chain step running for a total number of $K$ steps and starting at the original data distribution.

### IV. METHODOLOGY AND DATA DESCRIPTION

In this section, we describe the metrics chosen for prediction accuracy assessment, followed by a brief description of the dataset. Finally, we propose and describe an automatic method for feature extraction enforced by our dataset characteristics.

#### A. Metrics for prediction assessment

To quantify the performance of the prediction methods, we used a variety of standard metrics. Firstly, the prediction accuracy is evaluated using three popular metrics capable to put a different penalty on the same error, namely the root mean square error, $RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (v_i - \hat{v}_i)^2}$, the normalized root-mean-square error, $NRMSE[\%] = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (v_i - \hat{v}_i)^2/(v_{\max} - v_{\min})} \cdot 100$, and the mean absolute percentage error, $MAPE = \frac{1}{n} \sum_{i=1}^{n} |v_i - \hat{v}_i|/\max(v_i) \cdot 100$, where $n$ represents the total number of predicted steps, $v_i$ represents the true values for the time-step $i$ and $\hat{v}_i$ represents the value predicted by the model at the same time-step. Secondly, the Pearson Correlation Coefficient ($PCC$) is used to indicate the degree of the linear dependence between the real and the predicted values, as follows:

\[
PCC(v, \hat{v}) = \frac{E[(v - \mu_v)(\hat{v} - \mu_{\hat{v}})]}{\sigma_v \sigma_{\hat{v}}}
\]

where $E[\cdot]$ is the expected value operator with means $\mu_v$ and $\mu_{\hat{v}}$, and standard deviations $\sigma_v$ and $\sigma_{\hat{v}}$, for the true and estimated values, respectively. The PCC value is within the range [-1,1]. The sign of the correlation coefficient defines the direction of the relationship, either positive or negative. Besides using PCC in the demand forecast evaluation process, in the second part of the experiments, the PCC values were used to highlight the most influential factors for the electrical energy demand profiles.

### B. Dataset description

In this work we have used the EcoGrid dataset collected from the Danish Island Bornholm in the first seven months of 2014. The dataset includes the aggregated energy consumption, the real-time price (RTP), forecast prices (DA, HA) and meteorological data [23]. Altogether, this dataset has 50677 records at 5 minutes resolution, each record containing 16 different features, leading to more than 800000 data points. Table I summarizes some basic statistical information about the entire dataset used in the experiments, such as mean and standard deviation for each feature. Furthermore, the last column of Table I shows the correlation coefficient between the electrical energy demand values and the additional information available in the EcoGrid database. Figure 2 shows the aggregated electrical demand composition with a 5 minutes sampling rate which is the basis of our analysis. This aggregated electrical demand involves 1900 customers equipped with local generation. The decreasing trend of the demand with time, observed in Figure 2 is mainly due to the negative correlation with the temperature, i.e. PCC = -0.83, but also suggests that a load shifting may have occurred in the presence of a large renewable energy penetration.

![Figure 2. Electrical demand profile at low aggregation level between 1 January 2014 until 25 June 2014](image_url)
C. Feature extraction

The growth of distributed energy resources (DERs), together with smart appliances, capable of responding to real-time pricing signals yields poor correlations with energy consumption at low time resolutions, as it can be observed in Table I. Thus, constructing a proper combination of the additional information needed in order to improve the forecast accuracy is not a trivial task. Moreover, the extracted information aims to be a non-redundant generalization of the price and weather data. Besides that, from a computational perspective we want to have a lower dimensional data set. One traditional way to perform feature extraction is based on statistical hypothesis testing in order to determine if the distributions of values of a feature for two different classes are distinct. Still this solution creates results which are hard to interpret. A simpler solution is to use Principal Component Analysis (PCA), which is part of a wide area of clustering methods. However, PCA loses its information-theoretic optimality as soon as the data becomes dependent [24].

![Figure 3. General architecture of a Gaussian restricted Boltzmann machine (GRBM) as input to the FCRBM.](image)

More recently, it was shown that restricted Boltzmann machines are capable to learn low-dimensional codes that work much better than PCA as a tool to reduce the dimensionality of the data [25]. Consequently, we propose a combination between a Gaussian Restricted Boltzmann Machine (GRBM) and a FCRBM to perform dimensionality reduction and time series prediction, as depicted in Figure 3. The RBM mathematical details were previously described in Section II-A. This is further enhanced with Gaussian neurons in the visible layer in order to transform the RBM into an GRBM [25].

![Figure 4. The prediction error of the aggregated demand with mean (straight line) and standard deviation (shaded area) for 6 hours, using FCRBM, SVM and persistence methods.](image)

B. Electrical energy demand forecast

To quantify the performance of the proposed method, we used the four metrics described in Section IV-A. The results obtained with FCRBM have been further compared with other forecasting methods, such as SVM and persistence. Traditionally, the persistence method is recommended especially for very short-term forecasting [27], as it simply assumes that a constant value occurs over the forecast horizon.

![Figure 5. MAPE (%) for various forecasting methods.](image)

### Table II. Summary of the Experiments

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Time horizon</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5 minutes</td>
<td>5 minutes</td>
</tr>
<tr>
<td>2</td>
<td>15 minutes</td>
<td>5 minutes</td>
</tr>
<tr>
<td>3</td>
<td>1 hour</td>
<td>5 minutes</td>
</tr>
<tr>
<td>4</td>
<td>6 hours</td>
<td>5 minutes</td>
</tr>
<tr>
<td>5</td>
<td>1 day</td>
<td>5 minutes</td>
</tr>
</tbody>
</table>

The growth of distributed energy resources (DERs), together with smart appliances, capable of responding to real-time pricing signals yields poor correlations with energy consumption at low time resolutions, as it can be observed in Table I. Thus, constructing a proper combination of the additional information needed in order to improve the forecast accuracy is not a trivial task. Moreover, the extracted information aims to be a non-redundant generalization of the price and weather data. Besides that, from a computational perspective we want to have a lower dimensional data set. One traditional way to perform feature extraction is based on statistical hypothesis testing in order to determine if the distributions of values of a feature for two different classes are distinct. Still this solution creates results which are hard to interpret. A simpler solution is to use Principal Component Analysis (PCA), which is part of a wide area of clustering methods. However, PCA loses its information-theoretic optimality as soon as the data becomes dependent [24].
Figure 5 shows an example of aggregated energy demand prediction error for a max lead time of 72 time steps (6 hours ahead) averaged over the five weeks of the testing period, from 21 May 2014 to 25 June 2014. Therein, a slightly better accuracy with a lower variation is visible for FCRBM versus SVM in terms of the MAPE metric. Furthermore, Table III shows the performance of the proposed models for all five scenarios. Overall, FCRBM outperforms the other methods in all metrics, while SVM performs better than persistence for longer time horizons, and persistence performs better than SVM for the short-term scenarios 1 and 2. For a pictorial view of the short-term forecast accuracy of the three methods we depict in Figure 5 an example of the true and forecast aggregated electrical energy demand over 6 hours horizon, with 5 minute resolution.

C. Demand forecast in a price-responsive context

This second set of experiments investigates the possibility to increase the forecast accuracy by using fusion data from the environment in a price-responsive context, over a short period of time (approximately three days).

The price-responsiveness of the electricity demand is observed during a visual analysis of the entire dataset, and highlighted by the differences in the correlations observed in Table I and Table IV in terms of price values. So, we enhanced the FCRBM model with additional information and we analyzed the accuracy of the predictor. More exactly, following the method proposed in Section IV-C, we performed a fully automatic feature extraction computation using a GRBM model from the day-ahead price and cloud cover data. Then this encoded information is placed into the class layer of the FCRBM model.

TABLE V. IMPROVED ACCURACY OF THE AGGREGATED ELECTRICAL DEMAND FORECASTING USING PRICE AND METEOROLOGICAL DATA

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Methods</th>
<th>NRMSE [%]</th>
<th>RMSE</th>
<th>MAPE</th>
<th>PCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Persistence</td>
<td>0.73</td>
<td>27.31</td>
<td>0.85</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>1.69</td>
<td>62.90</td>
<td>1.95</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>FCRBM</td>
<td>0.71</td>
<td>25.21</td>
<td>0.84</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>Persistence</td>
<td>1.27</td>
<td>47.24</td>
<td>1.34</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>2.07</td>
<td>77.30</td>
<td>2.28</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>FCRBM</td>
<td>1.23</td>
<td>45.95</td>
<td>1.31</td>
<td>0.31</td>
</tr>
<tr>
<td>3</td>
<td>Persistence</td>
<td>3.02</td>
<td>112.38</td>
<td>3.03</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>2.91</td>
<td>108.44</td>
<td>3.07</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>FCRBM</td>
<td>2.50</td>
<td>93.06</td>
<td>2.59</td>
<td>0.46</td>
</tr>
<tr>
<td>4</td>
<td>Persistence</td>
<td>12.26</td>
<td>456.10</td>
<td>11.66</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>4.48</td>
<td>166.55</td>
<td>4.43</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>FCRBM</td>
<td>4.30</td>
<td>160.10</td>
<td>4.18</td>
<td>0.88</td>
</tr>
<tr>
<td>5</td>
<td>Persistence</td>
<td>11.76</td>
<td>437.35</td>
<td>10.39</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>SVM</td>
<td>5.64</td>
<td>209.90</td>
<td>4.70</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>FCRBM</td>
<td>5.19</td>
<td>193.14</td>
<td>4.49</td>
<td>0.91</td>
</tr>
</tbody>
</table>
presented the performance of the forecasting methods analyzed for various combinations of input data which include, next to historical values for aggregated electrical demand, also prices and weather conditions. Although our data is multidimensional we have avoided the scalability problems by performing the feature extraction procedure. This leads to a light approach able to generalize over any other time series. It is worth highlighting, that we observed that by adding more external information we can slightly improve the overall accuracy for this real dataset.

VI. CONCLUSION

This paper proposes a powerful stochastic machine learning method to forecast the electricity demand at low aggregation levels, namely the Factored Conditional Restricted Boltzmann Machine. FCRBM has good generalization capabilities and it can be used to accommodate large sets of data, while its exploitation time in real-world settings is on the order of few milliseconds. Secondary, we propose the use of GRBM to extract features from external information and to reduce the dimensionality for the FCRBM. We validate our approach on a real dataset, consisting of 1900 households originating from the Danish island of Bornholm, collected within the EcoGrid project. In order to compare alternative approaches we used four different metrics and two benchmark forecasting methods, namely Support Vector Machine and persistence. On the one hand, the results show that FCRBM outperforms the other two methods, and on the other hand, they suggest that by adding more weather and price information to the FCRBM, its performance may be improved further. This promising method can in the future be applied to a fully automatic real-time prediction and optimal control of electrical energy consumption via demand response in a smart grid context.

REFERENCES


