Combining Pattern Sequence Similarity with Neural Networks for Forecasting Electricity Demand Time Series

Irena Koprinska, Mashud Rana, Alicia Troncoso, and Francisco Martínez-Álvarez

Abstract—We present PSF-NN, a new approach for time series forecasting. It combines prediction based on sequence similarity with neural networks. PSF-NN first generates predictions using the PSF algorithm that are then refined by the neural network component, which also utilizes additional features. We evaluate the performance of PSF-NN using a time series of hourly electricity demands for the state of New South Wales in Australia for three years. The task is to predict an interval of future values simultaneously, i.e. the 24 demands for the next day, instead of predicting just a single future demand. The results showed that the combined PSF-NN approach provides accurate predictions, outperforming the original PSF algorithm and a number of baselines.

I. INTRODUCTION

Time series analysis and time series forecasting are important tasks in many applications in science, economics and industry. In this paper we consider the task of forecasting the future electricity demand for a given region from a time-series of previous electricity demands. More specifically, given a time series of electricity demands measured every hour up to day \( d \), our goal is to forecast simultaneously the 24 hourly demands for day \( d+1 \).

This is an example of short-term demand forecasting. It is used for planning the optimal commitment of generators, setting the minimum reserve and assisting the electricity market participants to make bids in competitive electricity markets. The demand prediction needs to be highly accurate. Under-prediction leads to insufficient supply since large amounts of electricity cannot be stored, while over-prediction increases the electricity price and operational costs.

Predicting short-term electricity demand with high accuracy is challenging. The electricity demand time series is non-linear and non-stationary, with daily and weekly cycles. It is also noisy due to large loads with unknown hours of operation, special events and holidays, extreme weather conditions and sudden weather changes.

Short-term load forecasting has been an active area of research. There are two main groups of approaches: statistical, such as ARIMA, linear regression and exponential smoothing [1-3] and computational intelligence, with Neural Networks (NNs) being the most popular representative of this group [4-7]. NNs offer the following advantages over the model-based statistical methods: they can identify and model non-linear input/output relationships and can learn these complex relationships from a sample of input/output training examples. In contrast, the standard statistical methods fit a pre-specified relationship, e.g. a line using the least square method, which limits the range of relationships they can represent.

Most of the existing approaches mainly focus on predicting the value for a single future point, i.e. at time \( t \) the task is to predict the demand value for time \( t+h \), where \( h \) is the forecasting horizon, typically \( h=1 \). To make predictions for more than one step ahead, they typically use the forecasted values. In this paper we focus on predicting the values for an interval of future points simultaneously, i.e. at time \( t \) the task is to predict \( n \) consecutive future values: \( t+h, t+h+1, ..., t+h+(n-1) \). For example, in our electricity demand task, the goal is to predict simultaneously 24 future values, corresponding to the hourly load for the next day.

Martínez-Álvarez et al. [8] have considered the same problem, predicting \( n \) future values simultaneously, and have proposed the Pattern Sequence-Based Forecasting (PSF) algorithm as a solution. PSF was applied to predict the 24 hourly electricity demands and the 24 hourly electricity prices for the next day. The results showed that it is a very competitive approach outperforming state-of-the-art methods such as ARIMA, support vector regression and Backpropagation Neural Networks (BPNN). In this paper we investigate if it is possible to improve the PSF predictions by combining PSF with a NN in a hybrid approach.

Our contribution can be summarized as follows:

1) We present a new approach for time series prediction, called PSF-NN. It combines forecasts based on similar sequences with neural networks. PSF-NN first generates predictions using the PSF algorithm that are then used in conjunction with additional information as inputs to a BPNN component.

Although we present a case study in electricity demand prediction, PSF-NN is not limited to this particular task. It is a generic approach that can be used to make predictions for different types of time series, where the goal is to predict the values for an interval of future points simultaneously.

2) For the case study discussed in this paper, electricity demand prediction, we have developed three variations of the PSF-NN approach that use different additional information, appropriate for the task, namely the electricity demands from the previous day (PSF-NN1), previous week (PSF-NN2) and both the previous day and week (PSF-NN3).

3) We use a novel method to generate an ensemble of NNs for the NN component of PSF-NN. This is needed in
order to reduce the sensitivity of the NN component to network architecture and initialization of weights and generate more stable forecasts. The results confirm that the ensemble generates more accurate results than a single NN at an acceptable computational cost.

4) We conduct a comprehensive evaluation of the proposed PSF-NN algorithm using electricity demand data for the state of New South Wales (NSW) in Australia for three years. The results show that two versions of the PSF-NN algorithm, PSF-NN1 and PSF-NN3, outperform the PSF algorithm and a number of baselines in terms of prediction accuracy and that these improvements are statistically significant.

5) We investigate the monthly performance of PSF-NN1 and PSF-NN3 and propose a hybrid switch approach that is able to improve further the prediction accuracy.

The rest of this paper is structured as follows. In the next section we review related work on short-term demand forecasting. Section III presents a problem statement and section IV describes the proposed PSF-NN algorithm. Section V summarizes the experimental setup. Section VI presents and discusses the results and Section VII concludes the paper.

II. PREVIOUS WORK

A variety of approaches for short-term demand forecasting have been proposed. They can be broadly grouped into two classes: 1) statistical, e.g. exponential smoothing and ARIMA and 2) computational intelligence, e.g. neural networks and support vector regression.

Examples of the first group are [1-3, 9]. In [1] Taylor, de Menezes and McSharry compared the performance of four methods for predicting the hourly demand for Rio de Janeiro from 1 to 24 hours ahead: ARIMA, double seasonal Holt-Winters exponential smoothing, BPNN and a regression method with principal component analysis. The simplest method, exponential smoothing was shown to be the most accurate. In their subsequent work [2] Taylor and McSharry empirically compared the two best methods from their previous work and also two new methods: an alternative exponential smoothing and periodic autoregression, using hourly data for four European countries. The double seasonal Holt-Winters exponential smoothing was again the best performing method.

A recent study [9] evaluated five novel formulations of exponential smoothing on British and French half-hourly demand data for forecasts up to a day ahead. The results showed that the novel singular value decomposition exponential smoothing was a promising approach, performing similarly to a Holt-Winters exponential smoothing but the best approach was an intraday exponential smoothing. Soares and Medeiros [3] proposed a forecasting model consisting of two elements: deterministic (describing trends and seasonality) and stochastic using linear autocorrelation. A different model was built for each hour of the day. An evaluation using Brazilian hourly data showed that the proposed method obtained good results and outperformed ARIMA and other benchmark methods.

Approaches based on NNs, in particular BPNN, have also been very popular for short-term electricity forecasting, for a comprehensive review see [7]. Reis and Alves da Silva [4] considered 1-24 hours ahead prediction of hourly North American data. They used multilevel wavelet to decompose the demand into several low and high frequency components that were inputs to two BPNN-based approaches. More than one step ahead predictions were made recursively by using the forecasted values. The best approach achieved a mean absolute percentage error of 1.12% for 1-hour ahead prediction. It used both wavelet and non-wavelet features. The wavelet features were extracted from wavelet components A3 and D3 and the non-wavelet features included previous load, temperature and the hour of the day.

Chen et al. [5] also considered predicting the demand 1-day ahead from previous hourly data using wavelet and BPNN. They selected a day that is similar to the day to be forecasted, decomposed the demand for it into two wavelet components and then trained a separate BPNN for each component. Non-wavelet features such as temperature, humidity, cloud cover and precipitation were also used as inputs to the BPNNs. An evaluation using 4 years data for New England was conducted, showing a mean absolute percentage error of 1.24-2.22%.

Fan et al. [6] proposed a hybrid method combining Bayesian clustering and support vector regression. The data was first partitioned into clusters and then 24 support vector regressors were fit to each cluster, one for each hour, to predict the demand for the next day. An evaluation using New York City electricity demand data was conducted showing the effectiveness of the proposed approach.

III. PROBLEM STATEMENT

Given a time series of hourly electricity demands up to day \( d \), the goal is to forecast simultaneously the 24 hourly demands for day \( d+1 \).

Hence, the prediction is based on previous demands only and weather variables are not considered. This is the typical case for short-term demand forecasting. Weather conditions such as temperature and humidity are rarely used for prediction horizons up to a day ahead for two reasons: 1) it is considered that the changes in the weather variables are smooth enough to be captured in the demand time series itself and 2) there are difficulties to obtain reliable weather forecast data [10]. However, recent studies using British electricity data [9, 10] found that although the previous demands were sufficient for making forecasts up to 5 hours ahead, the use of both previous demands and weather variables was beneficial for longer time horizons. Hence, the use of weather variables in conjunction with previous electricity demands requires further investigation for data from different countries, including Australia, and we plan to study this in the future. In this paper we use only previous electricity demands.
IV. THE PSF-NN ALGORITHM

A. The PSF Algorithm

PSF is a very successful algorithm, recently proposed by Martinez-Álvarez et al. [8]. It is a nearest neighbor-based algorithm, a generalization of [11] and an improvement of [12], and uses similarity between sequences of cluster labels.

An instance in the PSF algorithm is a 24-dimensional vector $X_i = (X_{i1},..,X_{i24})$ consisting of the hourly demands for day $i$. The PSF algorithm includes two stages: clustering and prediction. During the clustering step, it groups all vectors $X_i$ from the training data into $K$ clusters using the K-means clustering algorithm and labels them with the cluster number. The prediction step is illustrated in Fig. 1. To predict the demand for day $d+1$ from the test data, PSF extracts a sequence of $W$ consecutive days, from day $d$ backwards, $D_W^d = (d_{i-W+1},d_{i-W+2},...,d_{i-1},d_i)$ with their corresponding sequence of cluster labels $S_W^d$. It matches the sequence of labels against the training data to find sequences that are exactly the same as $S_W^d$; these sequences form the set of equal sequences $ES_d$. For each sequence $ES_d$, it finds the following day and averages the hourly loads for these days; these average values become the 24 hourly predictions for day $d+1$.

![Fig. 1. PSF algorithm – prediction step](image)

If there are no equal sequences of size $W$, the window size is decreased by 1 and the search is repeated until equal sequences are found.

PSF also includes a method for determining the optimal number of clusters $K$ and the optimal window size $W$ using the training data. To select the optimal $K$, the clustering algorithm is run for different $K$s and the obtained clustering results are evaluated by computing the Silhouette, Dunn and Davies-Bouldin indices and taking their majority vote. To select the optimal window length, different lengths $W$ were evaluated using 12-fold cross validation (one fold for each month). The best $W$ is the one that minimizes the averaged error for the 12 folds.

The PSF algorithm was evaluated on energy time series and was shown to outperform state-of-the-art statistical and machine learning methods such as ARIMA, support vector regression and BPNN [8].

B. The proposed PSF-NN Algorithm

In this paper we propose a new hybrid algorithm, called PSF-NN, that combines PSF with NN. There are two key ideas that underpin PSF-NN.

The first idea is to try to refine the PSF prediction by minimizing the difference between the predicted and actual demand using a NN. This correction is learned from previous historical data.

The second idea is to use an extended feature set for the NN component. In addition to the demand predictions computed by PSF, we propose to use the demands from the previous day, previous week or both. This extension is supported by recent research on feature selection for electricity demand prediction [13, 14] showing that the loads from the previous day and previous week at the forecast time are very important for making accurate predictions.

We consider three different feature set extensions: using the demand values from the previous day, the demand values from the previous week and both of them. Based on them, we have developed three different versions of the PSF-NN algorithm: PSF-NN1, PSF-NN2 and PSF-NN3; a summary of the features they use is given in Table 1.

<table>
<thead>
<tr>
<th>To predict the 24 demand values for day $d+1$, use:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$24$ predictions for $d+1$ generated by PSF</td>
</tr>
<tr>
<td>$24$ actual demand values for the previous day, i.e. day $d$</td>
</tr>
<tr>
<td>$24$ actual demand values for the previous week, i.e. day $d$-7</td>
</tr>
<tr>
<td>Total number of features</td>
</tr>
<tr>
<td>PSF-NN1</td>
</tr>
<tr>
<td>$\sqrt{\ }$</td>
</tr>
<tr>
<td>PSF-NN2</td>
</tr>
<tr>
<td>$\sqrt{\ }$</td>
</tr>
<tr>
<td>$\sqrt{\ }$</td>
</tr>
<tr>
<td>PSF-NN3</td>
</tr>
<tr>
<td>$\sqrt{\ }$</td>
</tr>
<tr>
<td>$\sqrt{\ }$</td>
</tr>
<tr>
<td>$48$</td>
</tr>
<tr>
<td>$48$</td>
</tr>
<tr>
<td>$72$</td>
</tr>
</tbody>
</table>

Fig. 2 illustrates the prediction phase of the PSF-NN1 algorithm (i.e. the training of the PSF and NN components have been completed already). To make a prediction for day $d+1$, the PSF component takes as an input the sequence of the $W$ consecutive days, preceding and including day $d$ and outputs the 24 hourly predictions for day $d+1$. The NN component takes as an input the PSF predictions, together with the actual demand values for day $d$ and outputs the final prediction for day $d+1$, i.e. the prediction of the combined algorithm PSF-NN1.
The ensembles are evaluated on the validation set and the best one is selected, the one that is most accurate. This best ensemble is then evaluated on the test data.

Our ensemble method is similar to the one used by Adeodato et al. [15] in their award winning solution for the NN3 Forecasting Competition [16]. However, they first select a good NN architecture and then build an ensemble for it, while we build ensembles with different architectures and select the best one.

V. EXPERIMENTAL SETUP

A. Data

We use electricity demand data for the state of NSW in Australia for three years: 2009, 2010 and 2011. The data is provided by the Australian Energy Market Operator (AEMO) and is publicly available at [17]. It is sampled every hour, so there are 8,760 samples for each year and 8,760*3 = 26,280 samples altogether.

B. Training, Validation and Test Sets

The data from the first two years, 2009 and 2010, is used to train the prediction models and tune their parameters, while the data for 2011 is used to measure the performance of these models.

Although all prediction models use the same test set (2011), the PSF and the combined PSF-NN models are trained differently as they have different requirements.

The training data for PSF is the data for 2010. This data is used to select the parameters $K$ (number of clusters) to group the data into $K$ clusters and $W$ (window size) to match sequences of labels of length $W$ on the historical data.

The combined PSF-NN models require the predictions of PSF for the training of their NN component. The training process for them is two-step. Firstly, the PSF component is trained using the 2009 data and then used to make predictions for 2010. Secondly, these predictions, together with the additional features (actual 2010 demands for the previous day, week or both depending on the PSF-NN version) and the actual targets for 2010, are used to create training and validation data for the NN component: 70% of this data is used for training and 30% is used as a validation set to select the best ensemble of NNs. Recall that the NN component is not a single NN but an ensemble of NNs. This ensemble is then used to predict the test data.

C. Performance Metrics

To assess the performance of the prediction methods, we used the following measures:

1) Mean Absolute Error (MAE)

$$MAE = \frac{1}{N} \sum_{h=1}^{H} \sum_{i=1}^{K} |\hat{X}_{ih} - \bar{X}_{ih}|$$

where $\hat{X}_{ih}$ and $X_{ih}$ are the predicted and actual demand for day $i$ at hour $h$, respectively, $N$ is the number of samples (days) in the test data and $H$ is the number of predicted hours ($h=24$ for our task).
2) Mean Absolute Percentage Error (MAPE), an extension of MAE:

\[
\text{MAPE} = \frac{1}{N} \sum_{h=1}^{H} \sum_{t=1}^{T} \left| \frac{X_{ih} - \hat{X}_{ih}}{X_{ih}} \right| 
\times 100\%.
\]

D. Baselines and Prediction Models Used for Comparison

We also compared the performance of the PSF-NN algorithm with the original PSF and a number of baselines (naïve forecasting methods):

1) B\text{pday}: demand from the previous day at the same time. The prediction for \(X_{ih}\) is given by \(X_{i-1h}\).

2) B\text{pweek}: demand from the previous week at the same time. The prediction for \(X_{ih}\) is given by \(X_{i-7h}\).

3) B\text{mean}: mean demand value in the training data for hour \(h\). The prediction for \(X_{ih}\) will be \(\text{mean}(X_{h})\) over all days in the training data. This is also called ZeroR prediction algorithm.

VI. RESULTS AND DISCUSSION

A. Parameter Selection for PSF

We used the method described in [8] to select the number of clusters \(K\) and window length \(W\) for the PSF algorithm.

1) Selecting the Number of Clusters

The PSF component of the PSF-NN algorithm uses the 2009 data as training data. The \(K\)-means algorithm was used to cluster the training data by varying \(K\) from 2 to 20. The three validity indices (Silhouette, Dunn and Davies-Bouldin) were computed for each of these 19 clusterings; their values are shown in Fig. 4. By keeping in mind that the best value for Silhouette and Dunn is the maximal and the best for Davies-Bouldin is the minimal, we can see that all three indices achieved their best value for \(K=12\). Hence, by taking the majority voting we selected \(K=12\) as an optimal number of clusters for the PSF component of the PSF-NN algorithm.

The same procedure was applied for the standalone PSF algorithm that is used for comparison. PSF uses the 2010 data as training data. The best results were obtained for: 6 clusters (Silhouette) and 14 clusters (Dunn and Davies-Bouldin). By taking the majority vote, \(K=14\) was selected as the optimal number of clusters for the PSF algorithm.

2) Selecting the Length of the Window

Once the optimal number of clusters \(K\) is determined, we can select the optimal window size \(W\). This is done by evaluating the accuracy of the PSF algorithm with the selected \(K\) on the training set using 12-fold cross validation where each fold contains the data for one month. We evaluated the accuracy for different sizes of window in order to find the length providing the minimum error. The results for the PSF component of PSF-NN and the standalone PSF are shown in Table 2. The symbol “-” means that similar sequences of length \(W\) were not found. It can be appreciated that it does not exist similar sequences for windows greater than 6. We can see that in both cases the best accuracy (best MAPE, shown in bold) was achieved for \(W=3\). Hence, we selected \(W=3\) as the optimal window length.

<table>
<thead>
<tr>
<th>(W)</th>
<th>PSF component of PSF-NN</th>
<th>PSF standalone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.83</td>
<td>3.06</td>
</tr>
<tr>
<td>2</td>
<td>3.99</td>
<td>4.64</td>
</tr>
<tr>
<td>3</td>
<td>\textbf{3.28}</td>
<td>\textbf{2.89}</td>
</tr>
<tr>
<td>4</td>
<td>4.04</td>
<td>3.80</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>4.39</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

B. Overall performance

Table 3 shows the accuracy results (MAE and MAPE) of the four prediction methods and the three baselines. Fig. 5 presents graphically the MAE results from Table 3 for visual comparison. Table 4 shows the pair-wise comparison of all prediction methods and baselines for statistical significance of differences in accuracy.
The main results can be summarized as follows:

- The most accurate prediction method is PSF-NN3, followed by PSF-NN1, PSF, PSF-NN2 and the baselines $B_{pday}$, $B_{pweek}$ and $B_{mean}$. All differences in accuracy are statistically significant at $p \leq 0.001$, except the difference between PSF-NN1 and PSF-NN3 and between PSF-NN2 and $B_{pday}$.

- Thus, two of the hybrid methods, PSF-NN3 and PSF-NN1, improved the accuracy of the original PSF algorithm. These improvements were with 14.6% and 11.4% in MAE and 14.9% and 13.1% in MAPE, respectively, and were statistically significant.

- The third hybrid prediction method, PSF-NN2, was not able to achieve an improvement over PSF. This shows that the additional features used by PSF-NN2, the electricity demands from the previous day, were less informative than the demands from the previous day (PSF-NN1) and the demands from the previous day and previous week together (PSF-NN3).

- The two best performing methods, PSF-NN3 and PSF-NN1, achieved similar accuracy (MAE=300.77 and 311.94, respectively) and the differences between these accuracies were not statistically significant. PSF-NN1 is simpler as it uses only 48 features, while PSF-NN3 uses 72. Both are fast to train – their training times were between 12 and 15 minutes.

- All four prediction methods outperformed the baselines in terms of accuracy. The best baseline $B_{pday}$ was only slightly worse than PSF-NN2 and this difference was not statistically significant.

- The ranking of $B_{pday}$ and $B_{pweek}$ is consistent with the ranking of PSF-NN1 and PSF-NN2 and shows again that using the demands from the previous day results in better accuracy than using the demands from the previous week.

Hence, we conclude that the performance of PSF can be improved by using this algorithm as building block in the combined PSF-NN approach. PSF-NN1 and PSF-NN3, outperformed PSF in terms of prediction accuracy and these improvements were statistically significant.

### C. Monthly performance

We investigate whether some months are more difficult to predict than others and if there are differences in the monthly performance of PSF, PSF-NN1, PSF-NN2 and PSF-NN3. Table 5 and Table 6 report the monthly MAE and MAPE, respectively, for the four prediction models. Fig. 6 shows graphically the monthly MAE for visual comparison.

We can see that the most difficult months to predict for all prediction methods were December, February and January. These are the summer months in NSW in Australia. A possible explanation is that the electricity demand during the summer months is more variable as a result of the patchy use of air conditioning in the very hot days and the irregular days due to public holidays (e.g. Christmas and New Year) and school holidays (about from 20 December till the end of January).

### TABLE 5. MONTHLY MAE [MW] OF THE PREDICTION METHODS

<table>
<thead>
<tr>
<th>Month</th>
<th>PSF</th>
<th>PSF-NN1</th>
<th>PSF-NN2</th>
<th>PSF-NN3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>449.93</td>
<td>475.53</td>
<td>495.70</td>
<td>367.91</td>
</tr>
<tr>
<td>Feb</td>
<td>554.48</td>
<td>602.90</td>
<td>773.17</td>
<td>518.59</td>
</tr>
<tr>
<td>Mar</td>
<td>362.30</td>
<td>240.15</td>
<td>323.51</td>
<td>256.97</td>
</tr>
<tr>
<td>Apr</td>
<td>256.06</td>
<td>239.95</td>
<td>299.42</td>
<td>215.87</td>
</tr>
<tr>
<td>May</td>
<td>281.77</td>
<td>197.34</td>
<td>264.66</td>
<td>212.17</td>
</tr>
<tr>
<td>Jun</td>
<td>281.77</td>
<td>197.34</td>
<td>264.66</td>
<td>212.17</td>
</tr>
<tr>
<td>Jul</td>
<td>326.77</td>
<td>262.68</td>
<td>267.37</td>
<td>210.76</td>
</tr>
<tr>
<td>Aug</td>
<td>256.06</td>
<td>239.95</td>
<td>299.42</td>
<td>213.87</td>
</tr>
<tr>
<td>Sep</td>
<td>265.30</td>
<td>243.92</td>
<td>274.01</td>
<td>208.71</td>
</tr>
<tr>
<td>Oct</td>
<td>309.34</td>
<td>302.63</td>
<td>412.54</td>
<td>335.52</td>
</tr>
<tr>
<td>Nov</td>
<td>281.77</td>
<td>197.34</td>
<td>264.66</td>
<td>212.17</td>
</tr>
<tr>
<td>Dec</td>
<td>431.83</td>
<td>376.88</td>
<td>780.09</td>
<td>560.14</td>
</tr>
</tbody>
</table>
TABLE 6. MONTHLY MAE [%] OF THE PREDICTION METHODS

<table>
<thead>
<tr>
<th>Month</th>
<th>PSF</th>
<th>PSF-NN1</th>
<th>PSF-NN2</th>
<th>PSF-NN3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>4.85</td>
<td>5.16</td>
<td>5.28</td>
<td>3.92</td>
</tr>
<tr>
<td>Feb</td>
<td>5.81</td>
<td>5.68</td>
<td>7.35</td>
<td>5.05</td>
</tr>
<tr>
<td>Mar</td>
<td>4.11</td>
<td>2.74</td>
<td>3.72</td>
<td>2.91</td>
</tr>
<tr>
<td>Apr</td>
<td>3.18</td>
<td>2.95</td>
<td>3.69</td>
<td>2.67</td>
</tr>
<tr>
<td>May</td>
<td>3.07</td>
<td>2.10</td>
<td>2.88</td>
<td>2.27</td>
</tr>
<tr>
<td>Jun</td>
<td>3.33</td>
<td>2.61</td>
<td>3.03</td>
<td>2.19</td>
</tr>
<tr>
<td>Jul</td>
<td>3.36</td>
<td>2.68</td>
<td>2.75</td>
<td>2.56</td>
</tr>
<tr>
<td>Aug</td>
<td>3.42</td>
<td>2.56</td>
<td>2.98</td>
<td>2.33</td>
</tr>
<tr>
<td>Sep</td>
<td>3.10</td>
<td>2.83</td>
<td>3.21</td>
<td>2.44</td>
</tr>
<tr>
<td>Oct</td>
<td>3.74</td>
<td>3.61</td>
<td>5.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Nov</td>
<td>4.10</td>
<td>3.80</td>
<td>4.51</td>
<td>3.08</td>
</tr>
<tr>
<td>Dec</td>
<td>5.61</td>
<td>4.73</td>
<td>9.93</td>
<td>7.07</td>
</tr>
</tbody>
</table>

The easiest months to predict are from April to September, i.e. autumn, winter and early spring. These months have less irregular days and also more stable and similar demand profiles across years.

We can also see that PSF1-NN1 is the best prediction method for 6 of the months: March, April, May, August, October and December, while PSF-NN3 is the best method for the remaining 6 months. This motivates the use of a hybrid switch prediction approach that uses PSF-NN1 for the first group of 6 months and PSF-NN3 for the second group of 6 months. The accuracy of the hybrid switch approach will be MAE=284.98 and MAPE=3.16, an improvement with 7-8% over PSF-NN1 and with 5-6% over PSF-NN3.

D. Ensemble of NNs vs single NN

To evaluate the benefit of the ensemble in the NN component of PSF-NN1, PSF-NN2 and PSF-NN3, we compare the performance of these models when using a single NN and when using an ensemble of NNs. All other conditions are the same; in both cases we follow the procedure described in Section IV C, the only difference is in the use of a single NN or an ensemble of NNs.

Table 7 shows the accuracy results of the PSF-NN prediction models when using a single NN. Fig. 7 presents a graphical comparison of the MAE results with and without ensemble.

Fig. 6. Monthly MAE of the four prediction methods

Fig. 7. Ensemble of NNs vs a single NN in PSF-NN prediction methods

We can see that the accuracy was better when using an ensemble for all three prediction methods. The improvement was bigger for the best methods PSF-NN1 and PSF-NN3 and lower for PSF-NN2. The computational cost of using an ensemble was only slightly higher: 12-15 minutes to build and train an ensemble versus 20-30 seconds for a single NN.

Since the training is done offline, the training time requirements of both methods are acceptable. Once trained, both methods were fast at predicting new instances.

Hence, our results show that the use of ensemble of NNs instead of a single NN in the NN component of the PSF-NN prediction methods is beneficial; it produces better accuracy at acceptable computational cost.

VII. CONCLUSION

We presented PSF-NN, a new approach for time series forecasting. It combines pattern sequence similarity with neural networks. PSF-NN first generates predictions using the PSF algorithm that are then used in conjunction with additional features as inputs to a BPNN component. PSF-NN is a generic approach that can be used for different types of time series where the goal is to predict not only the value for a single future point but the values for an interval of future points simultaneously.

To evaluate PSF-NN we considered the task of forecasting the 24 hourly electricity demands for the next day using hourly demands for previous days. We proposed three variations of PSF-NN for this task based on different additional information used (demands from previous day, previous week or both). We conducted a comprehensive evaluation using three years of electricity load data. The results showed that two of these PSF-NN versions, PSF-NN1 and PSF-NN3, outperformed the original PSF algorithm and a number of baselines in terms of prediction accuracy and that these improvements were statistically significant. We investigated the monthly performance of PSF-NN1 and PSF-NN3 and found that they complement each other, which motivated the development of a hybrid
switch approach that further improved the prediction accuracy.

In future work we will investigate if the addition of weather variables is beneficial for short-term demand prediction of Australian data. We also plan to apply PSF-NN to other time series forecasting tasks such as forecasting electricity prices and the amount of solar power produced by photovoltaic plants.

ACKNOWLEDGMENT

Mashud Rana is supported by an Endeavour Postgraduate Award and NICTA Research Project Award (NRPA). We would like also to acknowledge the financial support from the Spanish Ministry of Science and Technology, project TIN2011-28956-C02.

REFERENCES


