Maximum Length Weighted Nearest Neighbor Approach for Electricity Load Forecasting

Tommaso Colombo, Irena Koprinska, and Massimo Panella

Abstract—In this paper we present a new approach for time series forecasting, called Maximum Length Weighted Nearest Neighbor (MLWNN), which combines prediction based on sequence similarity with optimization techniques. MLWNN predicts the 24 hourly electricity loads for the next day, from a time sequence of previously electricity loads up to the current day. We evaluate MLWNN using electricity load data for two years, for three countries (Australia, Portugal and Spain), and compare its performance with three state-of-the-art methods (weighted nearest neighbor, pattern sequence-based forecasting and iterative neural network) and with two baselines. The results show that MLWNN is a promising approach for one day ahead electricity load forecasting.

I. INTRODUCTION

This paper proposes a new machine learning approach for forecasting the hourly electricity loads for the next day. Given a time series of electricity loads measured every hour up to a given day, we aim at forecasting the 24 hourly electricity loads for the next day. This task is categorized as ‘short-term load forecasting’. It is needed for the planning and operation of power systems, to ensure reliable and cost-effective electricity supply. It is also fundamental in deregulated electricity markets, where electricity prices are determined through a bidding process, to support the market participants in their transactions.

Predicting the load for the next day accurately is a challenging task as the electricity time series has a number of nested cycles (daily, weekly, seasonal and yearly), and also shows random fluctuations depending on the household electricity usage, weather changes, large industrial units with irregular hours of operation and other variations.

More generally, the dynamics of the electricity loads influences the behavior of the energy prices. This behavior is complex and has shown large unexpected volatility in the last decade. In this context, a tool providing accurate forecasts of electricity loads and prices, is very useful. Non-linear regression models, such as Neural Networks (NNs), have been successfully used for forecasting financial time series, electricity prices and other commodity prices, and were shown to be able to capture important characteristics such as fat tails, volatility, persistence and leverage effects [1]–[16].

Most of the existing approaches for short-term load forecasting consider one step ahead prediction, i.e. at time $h$ the task is to predict the load for time $h+1$. In this paper we consider predicting all 24 hourly values for the next day simultaneously. This task was previously considered by the Weighted Nearest Neighbor (WNN) approach [17], which is a state-of-the-art approach. Suppose that $X_d$ is a 24-dimensional vector consisting of the hourly loads for a day $d$. To predict the loads $X_{d+1}$ for the next day $d+1$, WNN first finds the $K$ nearest neighbors of the previous day $X_d$; then, the prediction for the new day is a weighted linear combination of the load for the days following the nearest neighbors, where the weights are determined by the distance of the neighbors to $X_d$.

In this paper we propose a new approach for predicting the hourly electricity load for the next day, called Maximum Length Weighted Nearest Neighbor (MLWNN), which extends the state-of-the-art WNN approach in several ways. Our contribution can be stated as follows:

1) Our proposed MLWNN approach can be seen as a generalization of WNN. While WNN finds the most similar sequence of previous days (when determining the nearest neighbors), MLWNN finds the longest sequence of hourly loads with a similarity higher than a given threshold. The best value of this threshold is determined by an optimization procedure. As a consequence, MLWNN is applicable to any forecasting horizon, as it directly operates on a sequence of hourly loads instead of a sequence of days, and it does not require the time series to be organized into 24 dimensional vectors corresponding to each day.

2) We conduct an evaluation of MLWNN using electricity data for two years, for Australia, Portugal and Spain. We compare the forecasting accuracy of MLWNN on the three datasets with WNN and two other advanced forecasting approaches: Iterative Neural Network (INN) [18] and Pattern Sequence-Based Forecasting (PSF) [19], and also with two baselines.

The paper is organized as follows. In Section II we provide an overview of the most important and recent research works in the field of short-term load forecasting. The proposed MLWNN approach is presented in Section III. Section IV presents the experimental setup and Section V shows and discusses the results. Finally, our conclusions are drawn in Sections VI.

II. RELATED WORKS

There are two main groups of approaches for short-term load forecasting: statistical and computational intelligence. Prominent examples of the first group are exponential smoothing, Autoregressive Integrated Moving Average (ARIMA) and
Linear Regression (LR), and notable examples of the second group are NNs and support vector regression.

A. Statistical Approaches

Taylor et al. [20] considered the prediction of the hourly electricity load for Rio de Janeiro, from 1 to 24 hours ahead. They compared four methods: ARIMA, double seasonal Holt-Winters exponential smoothing, backpropagation NN and a PCA-based LR. They found that the most accurate method was exponential smoothing which was also the fastest method. In [21] Taylor and McSharry compared exponential smoothing and PCA-based LR with two new methods (a different formulation of exponential smoothing and a periodic autoregression) using hourly data for Italy, Norway and Sweden. The results again showed that the double seasonal Holt-Winters exponential smoothing was the most accurate method.

Soares and Medeiros [22] proposed a novel forecasting model with two components: deterministic for trends, seasonality and special days, and stochastic that uses linear autocorrelation. A different model was built for each hour of the day. An evaluation using Brazilian hourly data showed that the proposed approach obtained promising results, outperforming ARIMA and other methods. A semi-parametric additive regression method was proposed in [22] and used to forecast half-hourly electricity loads one day ahead for Australian data. A separate model was built for each half hour, using previous electricity loads, calendar and temperature variables. The forecasting method was evaluated offline on historical data and also in real time on site, showing very good results.

Fan and Hyndman [23] proposed a semi-parametric additive regression methodology that was used to forecast the half-hourly electricity loads one day ahead for the states of Victoria and South Australia in Australia. A separate model was built for each half hour, using the previous lagged electricity loads, calendar and temperature variables. The forecasting model showed excellent performance on both historical data and when applied in real time on site.

B. Computational Intelligence Approaches

NN-based approaches are probably the most popular approaches for load forecasting due to their ability to learn the time series from examples and to capture non-linear relationships between the predictor variables and the target variable [24]–[26].

Most of the proposed computational intelligence approaches considered the task of one step ahead prediction (e.g. 1 hour ahead); below we review approaches that predict all 24 hourly values for the next day. Apart from WNN [17] which was already mentioned, there are two other notable approaches that predict the 24 hourly values for the next day: PSF and INN.

PSF [19] is a generalization of WNN, which combines clustering with sequence matching. It first groups all vectors \( X_d \) from the training data into \( K \) clusters and labels them with the cluster number. Then it extracts a sequence of consecutive days, from day \( d \) backwards, and matches the cluster labels of this sequence against the training data to find a set \( ES_d \) of sequences that are the same. It then follows a nearest neighbor approach similarly to WNN, which finds the following day for each element of \( ES_d \) and averages the 24 hourly loads of these following days, in order to produce the final 24 hourly predictions for day \( d + 1 \). The results showed that both WNN and PSF are very competitive approaches outperforming ARIMA, NNs and other methods.

INN [18] is an iterative prediction method. At time \( h \) it makes a prediction for time \( h + 1 \); this prediction is added to the available data and used to make a prediction for time \( h + 2 \) and so on for all 24 points from the forecasting horizon. It uses a mutual information feature selector and a neural network forecasting algorithm. The results showed that it was able to provide accurate predictions, outperforming the non-iterative methods WNN and PSF.

Wavelet-based approaches predicting the load for the next day have also been proposed. Reis and Alves da Silva [27] considered the task of 1-24 hours ahead prediction of hourly North American data. They used multilevel wavelet to decompose the electricity load into several components that were predicted separately by NNs trained with the backpropagation algorithm. Chen et al. [28] also considered the task of predicting the electricity load 1-day ahead from previous hourly loads using wavelet transformation and backpropagation NNs. They selected a day that is similar to the day to be forecasted in terms of weekly index and weather, decomposed the load for this day into two wavelet components and then trained a separate NN for each component. Non-wavelet features such as temperature, humidity, cloud cover and precipitation were also used as inputs to the NNs. An evaluation using four years data for the state of New England was conducted, showing a mean absolute percentage error of 1.24-2.22%.

III. THE PROPOSED APPROACH TO SHORT-TERM ELECTRICITY LOAD FORECASTING

We consider a 'one day ahead' prediction problem: given the hourly loads recorded in the past up to day \( d \), the goal is to forecast the 24 hourly loads corresponding to day \( d + 1 \). More precisely, given a time series \( S(n), n \geq 0 \), of hourly loads that are known up to hour \( h \), we want to determine the hourly loads \( S(h+1), \ldots, S(h+24) \), which means considering a forecasting horizon of 24 time steps. This formulation is a generalization of the standard 'one day ahead' prediction problem, as it does not require a reorganization of the hourly time series data into 24-dimensional vectors. Thus, our proposed approach is general and can be applied to any forecasting horizon \( f \geq 1 \), to predict the samples \( S(h+1), \ldots, S(h+f) \).

We firstly introduce the main parameters of the method, that have to be specified in advance:

- forecasting horizon, an integer value \( f > 0 \);
- number of nearest neighbor patterns (subsequences), an integer value \( K \geq 1 \);
- upper bound of the dimension of the subsequences associated with the nearest neighbors, an integer value \( W \geq 1 \);
- similarity threshold under which two subsequences are not considered as sufficiently similar with respect to a
suitable defined similarity measure, a real-valued parameter $\theta > 0$.

The proposed MLWNN algorithm is based on the following steps, applied at any time instant $h$ using the available data:

1) Let $w$ be a counter variable initialized to 1.
2) Store a dataset of observations in a matrix $Y_w \in R^{(h-f-w+1) \times w}$, where the $i$th row $y_{w,i}$ of $Y_w$, $i = 1 \ldots (h-f-w+1)$, is a vector of $w$ hourly loads:

$$y_{w,i} = [S(i) \ S(i+1) \ldots S(i+w-1)]. \tag{1}$$

We note that the most recent data sample stored in $Y_w$ is $S(h-f)$, and that it is located in the last row and the rightmost column.

3) Let $z_w$ be the vector containing the most recent known $w$ samples of hourly loads:

$$z_w = [S(h-w+1) \ S(h-w+2) \ldots S(h)]. \tag{2}$$

4) Using the Chebyshev distance, find the $K$ nearest neighbors for the vector $z_w$ among the rows of $Y_w$. The Chebyshev distance $D(z_w,y_{w,i})$ between $z_w$ and $y_{w,i}$ is defined as:

$$\|z_w-y_{w,i}\|_\infty = \max_{j=1 \ldots w} |z_w(j)-y_{w,i}(j)|. \tag{3}$$

5) Let $D_{\text{max}}$ be the maximum Chebyshev distance among the found $K$ nearest neighbors and let $\hat{z}_{\text{max}}$ be the maximum absolute value of the elements in $z_w$. If $\frac{D_{\text{max}}}{\hat{z}_{\text{max}}} > \theta$ then go to step (8).

6) Store $w_{\text{best}} \leftarrow w$ as a new value for $w$ and also the reference vector $q_0 \leftarrow z_w$. Store the Nearest Set (NS) of vectors $\{q_1, q_2, \ldots, q_K\}$ using the $K$ nearest neighbors vectors found, where $q_1$ is the nearest vector of $q_0$ and $q_K$ is the furthest nearest vector.

7) If $w < \min\{h-f-K+1, W\}$ then increase $w+1 \leftarrow w$ and go back to step (2).

8) Extract the $f$ following samples for each vector in NS and compute a weighted average to produce the prediction of $S(h+1), \ldots, S(h+f)$, as explained below.

At the end of this procedure, we will obtain the set of nearest neighbors NS for $q_0$; each of them can be associated with the vector containing the $f$ following load values. For example, if $q_n = y_{w_{\text{next}},i}$, $1 \leq n \leq K$, as in (1), then:

$$q_n^{(f)} = [S(i+w) \ S(i+w+1) \ldots S(i+w+f-1)]. \tag{4}$$

The loads to be predicted are $q_0^{(f)}$:

$$q_0^{(f)} = [S(h+1) \ S(h+2) \ldots S(h+f)]. \tag{5}$$

The weighted average, which determines $q_0^{(f)}$, is therefore based on the following formula:

$$q_0^{(f)} = \frac{1}{\sum_{n=1}^{K} \alpha_n} \sum_{n=1}^{K} \alpha_n q_n^{(f)}, \tag{6}$$

where the weights $\alpha_n$ are obtained as follows:

$$\alpha_n = \frac{\|q_K \cdot q_0\|_\infty - \|q_n \cdot q_0\|_\infty}{\|q_K \cdot q_0\|_\infty - \|q_1 \cdot q_0\|_\infty}. \tag{7}$$

Regarding the novelty of the proposed MLWNN approach, there are three main differences with respect to the original WNN method. Firstly, the length of similar vectors we are looking for in the previous data, in order to generate NS, is iteratively and automatically increased. This is based on the assumption that the longer the similar sequence is, the better it represents the local behavior of the time series.

Also, the subsequence is increased based on a similarity threshold whose optimal value is determined using an optimization technique as described in the next section. MLWNN uses the Chebyshev distance while WNN uses the Euclidean distance. The Chebyshev distance is the largest of the element-by-element distances between the two vectors, and it has been chosen to find similarity between sequences by taking under control any divergence on a single element-by-element difference. MLWNN stops increasing the subsequence when the Chebyshev distance between the compared vectors is above the similarity threshold (expressed as a percentage similarity).

We note that if the Chebyshev distance is below the threshold, this means that at least one element-by-element distance is below this threshold.

Finally, WNN uses a window of previous days, while MLWNN does not require this, which means that a larger part of the available data (potentially all available data) can be used to make the prediction. The underlying assumption is that if we use more data, we may be able to find better candidates.

The parameter $w$ in (1) can be considered as the embedding dimension of the time series, that is the number of previous loads that will feed the regression model for estimating the next value to be predicted [19], [29]–[31]. This value can be found by other means, e.g. fractal dimensions and statistical methods, often assuming a chaotic behavior of the observed time series [32]. MLWNN algorithm aims at a more general heuristic for prediction, which tries to overcome the dependence of classical embedding approaches from the performance of estimation of the optimal dimension.

It is also important to note that this approach provides the basis for further extensions, that can utilize more complex regression models and also neural and fuzzy NNs. In addition, after NS has been determined, the predicted values in (5) can be obtained through a nonlinear inference system that replaces equations (6) and (7).

We note again the high flexibility of our proposed MLWNN algorithm, especially with respect to the forecasting horizon $f$. For example, MLWNN can be applied for one hour ahead prediction using the value $f = 1$. In the following sections, we report the results for $f = 24$, for the sake of comparison with WNN, PSN and INN approaches.

### IV. EXPERIMENTAL SETUP

We use electricity load data collected from Australia, Portugal and Spain for two years: from 1 January 2010 to 31 December 2011. The data are sampled every hour, thus the total
number of samples in each dataset is $2 \times 365 \times 24 = 17520$. All datasets are publicly available: the Australian data is for the State of New South Wales (NSW), it is provided by the Australian Energy Market Operator (AEMO) and available from http://www.aemo.com.au. The Portuguese and Spanish data are provided by the Spanish Electricity Price Market Operator (OMEL) and available from http://www. omelholding.es.

The electricity load has three main cycles: daily, weekly and yearly. Figs. 1-3 show the hourly electricity load for the three countries for one month, June 2011. From these plots we can clearly see the daily and weekly cycles, which are correlated with the human, industrial and commercial activities. During the day, the load has a minimum at 4am, a first peak at 9-10 am, stays relatively stable until the end of the working day and then it reaches a second peak at 6-7 pm. As expected, the load during the weekend is lower than the load during the weekdays. We can also see that the load for Portugal and Spain is more variable than the load for Australia, which might be due to greater weather and temperature fluctuations, and also to higher activity variations, especially given that the Australian data is for one region only (the state of NSW).

There is also a yearly pattern of the electricity load, e.g. the load for 2010 is very similar to the load for 2011. This motivates using the data for 2010 as training set to build prediction models and then using these models to predict the load for 2011 (test set data). Although the three datasets have similar cycles, the range of values is different. It is highest for the Spanish data (from 10000 MW to 40000 MW), followed by the Australian data (from 5000 MW to 14000 MW) and lowest for the Portuguese data (from 1000 MW to 8000 MW).

The objective function to be minimized was the Mean Absolute Percentage Error (MAPE), which is defined as:

$$\text{MAPE} = \frac{100}{N} \sum_{n=1}^{N} \left| \frac{S_n - \hat{S}_n}{S_n} \right|,$$

where $\hat{S}_n$ is the estimated value of the time series at time $n$ and $N$ is the total number of samples in the test set (when evaluating performance) or validation set (when tuning parameters).

As previously explained, given that $f$ was set to 24, ML-WNN requires three other parameters to be pre-specified or estimated from data: $\theta$, $W$ and $K$, for which a suitable upper bound was found to be 50. The training phase on the 2010 data was three-fold, due to computational time constraints that we set (i.e., a prediction must be completed in a hour, before a new sample is available and the next one is to be predicted).

Firstly, two suitable initial values for $W$ and $K$ were determined: $W_0$ and $K_0$. Then, a near-optimal value for $\theta$ was found through a global optimization technique. This was a ‘nested’ modification of a full-search global optimization algorithm: every iteration, $m$ values of $\theta$ were drawn at random in a given interval; then, the MAPE was computed $m$ times in order to select the value $\theta^*$ minimizing it; finally, the interval for the next iteration is restricted and centered on the $\theta^*$ value. The output of such an algorithm, although not guaranteeing to find the global optimum, is a ‘funnel’ leading quickly to a local minimum. If the objective function is convex, then it leads to the global minimum.

Successively, the determination of local optima for $W$ and $K$ was addressed: the values of these parameters were iteratively increased until a local minimum was found. This procedure was applied to both parameters, once at a time. For each tuning, the performance was evaluated by measuring the MAPE on a validation set, which consisted of the most recent 25% training data samples. The optimal parameters determined for the three datasets are summarized in Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Australian</th>
<th>Portuguese</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>0.09</td>
<td>0.22</td>
<td>0.19</td>
</tr>
<tr>
<td>$W$</td>
<td>25</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>$K$</td>
<td>8</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

We can see that the value of the parameter $\theta$ is considerably different for each dataset. It is lowest for the least volatile time series (Australia) and highest for the most volatile time series (Portugal). This result reflects the variability of the three time
series for which \( \theta \) is the similarity threshold expressed as a percentage: a lower value (more strict constraint) is suitable for a less volatile time series, while a higher value (more relaxed constraint) is needed when dealing with more volatile time series. The values of \( W \) and especially the values \( K \) are similar for the three datasets (between 7 and 9).

### V. Performance Evaluation

Once the training was completed and the parameters were tuned, MLWNN was evaluated on the test data (the 2011 data) in terms of MAPE. The evaluation procedure consisted of 365 one day ahead predictions, starting from the first hour of each day. The MAPE for each predicted hour, averaged over the whole test set, and the related standard deviations are reported in Table II. The overall average error and standard deviation were also computed and are shown in the last row.

The hourly MAPE results for the three datasets are shown in Fig. 4. Although the graphs are similar, we can see that the peaks in the MAPE error are not aligned: the higher the average error, the more shifted the peak is towards the hours in the middle of the day.

For the sake of comparison with other approaches, we also report the Mean Average Error (MAE), which is defined as:

\[
\text{MAE} = \frac{1}{N} \sum_{n=1}^{N} |S_n - \hat{S}_n|.
\]  

The MAE obtained for each hour, averaged over the 2011 test set, is reported in Table III.

Below we discuss the performance of MLWNN on each dataset in more details. For comparison, we also present the results of the three state-of-the-art approaches mentioned in Sect. II: INN, WNN and PSF. As MLWNN is a generalization of WNN, it is important that we compare the two methods. The other two methods, PSF and INN, were chosen for comparison since PSF is an extension of WNN, and INN was shown to outperform WNN and PSF in [18]. In addition, we also compare the performance of WNN with two baselines: \( B_{\text{pday}} \) and \( B_{\text{pweek}} \). The daily baseline \( B_{\text{pday}} \) simply predicts the loads from the previous day. The weekly baseline \( B_{\text{pweek}} \) predicts the loads from the same day of the previous week.

### A. Australian Data

Table IV presents the comparison results for the Australian dataset. We can see that MLWNN is the most accurate approach, followed by INN, WNN, PSF and the two baselines. We also used the \( t \)-test to evaluate if the differences in accuracy are statistically significant. The results showed that the accuracy of MLWNN is significantly higher (+) than all of the other methods used for comparison at a confidence level of 99%.

Figs. 5, 6 and 7 show the predicted and the actual loads for a month, week and day, respectively, in December 2011. The predicted values are relatively close to the actual values, especially taking into consideration that December is one of
the months with most volatile electricity load in Australia, and hence, difficult to predict.

![NSW: predicted (blue) vs. actual load (red), December 2011.](image)

**Fig. 5.** NSW: predicted (blue) vs. actual load (red), December 2011.

![NSW: predicted (blue) vs. actual load (red), a week in December 2011.](image)

**Fig. 6.** NSW: predicted (blue) vs. actual load (red), a week in December 2011.

![NSW: predicted (blue) vs. actual load (red), a day in December 2011.](image)

**Fig. 7.** NSW: predicted (blue) vs. actual load (red), a day in December 2011.

**B. Portuguese Data**

Table V shows the accuracy on the Portuguese dataset. We can see that MLWNN is more accurate than WNN. Overall, MLWNN is the second best performing approach after INN; WNN is third, followed by the daily baseline and PSF, and finally by the weekly baseline. The t-test for statistical significance showed that the accuracy of MLWNN is significantly lower (-) than INN and significantly higher (+) than all other methods at a confidence level of 99%.

Figs. 8, 9 and 10 show the predicted and actual load for the same month, week and day as for the Australian data. Note that the y-axes are different for the two datasets. The Portuguese dataset is the most difficult to predict. As Fig. 10 shows, the daily pattern for the Portuguese data is more complex and this may be reason for the lower accuracy.

**C. Spanish Data**

Table VI presents the comparison results for the Spanish dataset. MLWNN is the third most accurate approach, after INN and WNN. However, the accuracy of WNN is only slightly higher than the accuracy of MLWNN and this difference is not statistically significant. More specifically, the t-test for statistical significance of the accuracy differences at a confidence level of 99%, showing that the accuracy of MLWNN is statistically higher (+) than PSF and the two baselines, statistically lower (-) than INN and not statistically different (=) than the accuracy of WNN.

![Portugal: predicted (blue) vs. actual load (red), December 2011.](image)

**Fig. 8.** Portugal: predicted (blue) vs. actual load (red), December 2011.

![Portugal: predicted (blue) vs. actual load (red), a week in December 2011.](image)

**Fig. 9.** Portugal: predicted (blue) vs. actual load (red), a week in December 2011.

**TABLE IV**

<table>
<thead>
<tr>
<th>Error/test</th>
<th>MLWNN</th>
<th>INN</th>
<th>WNN</th>
<th>PSF</th>
<th>B_{relay}</th>
<th>B_{pweek}</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>3.10</td>
<td>3.36</td>
<td>3.40</td>
<td>3.96</td>
<td>4.82</td>
<td>5.20</td>
</tr>
<tr>
<td>MAE</td>
<td>277.4</td>
<td>304.9</td>
<td>307.5</td>
<td>352.0</td>
<td>420.5</td>
<td>471.2</td>
</tr>
<tr>
<td>t-test</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**TABLE V**

<table>
<thead>
<tr>
<th>Error/test</th>
<th>MLWNN</th>
<th>INN</th>
<th>WNN</th>
<th>PSF</th>
<th>B_{relay}</th>
<th>B_{pweek}</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>13.42</td>
<td>11.70</td>
<td>14.95</td>
<td>16.18</td>
<td>16.06</td>
<td>19.12</td>
</tr>
<tr>
<td>MAE</td>
<td>476.6</td>
<td>426.4</td>
<td>538.9</td>
<td>589.8</td>
<td>579.6</td>
<td>695.4</td>
</tr>
<tr>
<td>t-test</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**TABLE VI**

<table>
<thead>
<tr>
<th>Error/test</th>
<th>MLWNN</th>
<th>INN</th>
<th>WNN</th>
<th>PSF</th>
<th>B_{relay}</th>
<th>B_{pweek}</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAPE</td>
<td>6.78</td>
<td>5.79</td>
<td>6.03</td>
<td>8.87</td>
<td>9.47</td>
<td>7.45</td>
</tr>
<tr>
<td>MAE</td>
<td>1321.6</td>
<td>1134.6</td>
<td>1179.9</td>
<td>1711.4</td>
<td>1888.0</td>
<td>1460.1</td>
</tr>
<tr>
<td>t-test</td>
<td>-</td>
<td>=</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
ones, for a month, a week and a day, in December 2011, respectively.

In summary, we can conclude that MLWNN is able to obtain very promising results. It outperformed WNN (the algorithm it extends) on the Australian and Portuguese data, and was only slightly less accurate than WNN on the Spanish data. It also outperformed all the baselines and the state-of-the-art approach PSF on all datasets, and produced comparable results to INN. On the Australian data, MLWNN was the best performing algorithm. In terms of computational cost, some optimization efforts might be carried out in future research works, even considering the framework of big data analysis [33] or exhaustive search by quantum computation [34]. Nevertheless, the most ‘strict’ constraint, represented by the computational time that must be less than one hour in case of one hour ahead prediction, is widely respected by the algorithm in almost all the operative situations.

VI. CONCLUSION

In this paper, we presented MLWNN, a new approach for time series forecasting that is applicable for any forecasting horizon. It combines nearest neighbor sequence similarity prediction with optimization techniques. To make a prediction for a future value or a set of values, MLWNN finds similar previous sequences of values and combines them to produce a final prediction. In particular, it finds the longest sequences of values with a similarity higher than a given threshold, where the best value of the threshold is determined by an optimization procedure. In contrast to other advanced approaches, MLWNN directly operates on a single sequence and does not require the time series to be organized into vectors with a particular length (e.g. 24 dimensional vectors for hourly daily electricity loads).

MLWNN was applied and evaluated for predicting the 24 hourly electricity loads for the next day, using data from three different countries. A comparison with several advanced forecasting methods and baselines showed that MLWNN is a promising approach for one day ahead electricity load forecasting. It outperformed WNN, the approach it extends, on two of the datasets and obtained a similar performance on the third one.

In future works, we plan to investigate about the sensitivity of the proposed approach on some parameter settings and to develop further extensions of MLWNN by using advanced regression models based on NNs and fuzzy NNs.

REFERENCES


