

Using the hierarchical temporal memory spatial pooler for short-term forecasting of electrical load time series

E.N. Osegi

Department of Information Technology, National Open University of Nigeria, Lagos, Nigeria

Abstract

In this paper, an emerging state-of-the-art machine intelligence technique called the Hierarchical Temporal Memory (HTM) is applied to the task of short-term load forecasting (STLF). A HTM Spatial Pooler (HTM-SP) stage is used to continually form sparse distributed representations (SDRs) from a univariate load time series data, a temporal aggregator is used to transform the SDRs into a sequential bivariate representation space and an overlap classifier makes temporal classifications from the bivariate SDRs through time. The comparative performance of HTM on several daily electrical load time series data including the Eunite competition dataset and the Polish power system dataset from 2002 to 2004 are presented. The robustness performance of HTM is also further validated using hourly load data from three more recent electricity markets. The results obtained from experimenting with the Eunite and Polish dataset indicated that HTM will perform better than the existing techniques reported in the literature. In general, the robustness test also shows that the error distribution performance of the proposed HTM technique is positively skewed for most of the years considered and with kurtosis values mostly lower than a base value of 3 indicating a reasonable level of outlier rejections.

Keywords Machine intelligence, Short-term load forecasting, Sparse distributed representations, Time series, Overlapping temporal classification

Paper type Original Article

1. Introduction

Short term load forecasting (STLF) has been studied widely by many researchers. Artificial neural network techniques including variants of feed-forward back propagation algorithms, extreme learning machines and deep neural networks have been applied to STLF problems; genetic algorithms including hybrid optimizations have been used for day-ahead forecasts.

© E.N. Osegi. Published in *Applied Computing and Informatics*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) license. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this license may be seen at <http://creativecommons.org/licences/by/4.0/legalcode>

This study received no funding from any source or agency. Special thanks go to the editor and team of reviewers for their valuable and insightful comments and recommendations. Special thanks also go to the MATHWORKS Inc. for providing trial software and their very useful Webinar tutorials.

Declaration of interest: None.

Publishers note: The publisher wishes to inform readers that the article "Using the hierarchical temporal memory spatial pooler for short-term forecasting of electrical load time series" was originally published by the previous publisher of *Applied Computing and Informatics* and the pagination of this article has been subsequently changed. There has been no change to the content of the article. This change was necessary for the journal to transition from the previous publisher to the new one. The publisher sincerely apologises for any inconvenience caused. To access and cite this article, please use Osegi, E. N. (2021), "Using the hierarchical temporal memory spatial pooler for short-term forecasting of electrical load time series", *Applied Computing and Informatics*. Vol. 17 No. 2, pp. 264-278. The original publication date for this paper was 19/09/2018.



Auto-regressive Moving Average (ARMA) models and its variants have also been applied to various STLF problems. However, the problem of predicting in the short-term say in hours or days, and the extensive hyper-parameter tuning in existing techniques has made the task of forecasting the electricity load in an existing power system a continual challenge.

This section presents the related works in this area of forecasting that have used the tools of Artificial Intelligence (AI) or statistical modeling; the emphasis here is to enable the reader gain an understanding on some of the various AI or statistical techniques actively in use in the times past and till this present moment. It also proposes the Hierarchical Temporal Memory – an emerging neural network for streaming time series forecasting as yet another candidate STLF problem solver and states the objective of this research study.

1.1 Related works

While there is no universal approach to solving the STLF problem, research on STLF techniques have indeed been very active with a specific direction in the area of Artificial Intelligence (AI). Some of the various AI techniques may be categorized into hybrid evolutionary optimization approaches, pure or hybrid statistical techniques such as regression modeling and pure or hybrid (ensemble) neural networks.

Evolutionary computing uses the principles of natural competition such as Darwinian criterion for survival of the fittest. In [1], a Hybrid Evolutionary Optimization (HEO) for STLF problems have been proposed; it included the use of Hybrid Features Selection Method (HFS) based on Genetic Algorithms and Rough Sets for optimal selection of features and for reliable predictions of a popular electrical load time series competition dataset (the Eunite dataset). A Hybrid short-term load forecasting model based on principal component analysis (PCA) and a Mind Evolutionary Algorithm (MEA) for optimizing an Elman neural network have been proposed in [2]. The system model has been applied to power systems data in Eastern Europe with improved accuracies reported over the existing Elman without MEA optimization; the Elman network suffers from the local optima problem and the requirement of much iteration before convergence.

On the other hand statistical modeling techniques generally use some statistical measure such as the average (mean) or moving averages, variance or co-variance of the underlying data in conjunction with data regression (data fitting). Often time's auto-regressive or non-linear auto-regressive models in a hybrid fashion have been used for load forecasting [3]. One interesting hybrid statistical technique can be found in [4] where a combination of image processing and statistical techniques was used to perform day-ahead forecast in the California (USA) and Spanish electricity markets; day-ahead forecasts is one popular aspect of the STLF technique employed by power system economic managers and operators for decision making in the power markets. For the image processing step, the authors used a discrete wavelet transform (DWT) which is based on a two-pass signal (data) decomposition stage for image processing. The statistical technique included the Holt-Winters method (an exponential smoothing method) based on triple-smoothing stage and the weighted nearest neighbors technique for modeling a deterministic component (using trend and seasonality factors) and a fluctuation component (using fast changing data dynamics) respectively. The deterministic and fluctuation components were decomposed from the original historical load using a Haar DWT.

In the case of neural network approaches, the attention has been particularly impressive. For instance, one neuron models have been proposed for forecasting electrical load time series with promising results obtained over the Exponential Smoothing (ES) and Auto-regressive Integrated Moving Average (ARIMA) models [5]. In [6], day-ahead electricity price forecasting for Pennsylvania-New Jersey (PJM) interconnection was conducted using a hybrid approach including the back-propagation artificial neural network (ANN) and a weighted least square

(WLS) technique. In particular, they utilized the WLS state estimation (WLS-SE) technique to form a better prediction of the price data fluctuations.

Another hybrid neural technique can be found in [7] where an ensemble approach based on the Extreme Learning Machine (ELM) and a partial least squares regressor for aggregating the ensemble predictor outputs with wavelength pre-processing was proposed. Ensemble approaches avoid the overtraining issue faced by the conventional single ELM neural networks and facilitates the wavelength parameter determination. This method was also used for hourly and day-ahead forecasts of electricity loads (public datasets). Their proposed model was compared with some popular machine learning (ML) approaches and was shown to outperform the existing models with superior results reported.

In [8], point short-term load forecasting was carried out based on Chaos theory and a radial basis function (RBF) neural network. The Chaotic-RBF neural network involved the computation of a Lyapunov index to identify chaos – this however, is an expensive process and may hamper prediction accuracies.

More recently, in [9] was proposed the use of random weight initialization which can help the standard feed-forward Artificial Neural Network (ANN) converge faster. However, this does not overcome the problem of vanishing gradients particularly for networks that run on deeper architectures; more on this problem can be found in [10,11].

Some other important data mining related researches in this field can also be found in [12] which surveyed different data mining techniques for electricity price and demand forecasting, in [13] which uses the pattern similarity sequence (PSS) technique based on labels for electricity price and demand forecasting and the load forecasting technique based on deep learning [14].

While most of these researches have had profound effect on various application domains, the issue still persists as to how accurately these approaches can solve the STLF problem in addition to the need to avoid excessive hyper-parameter tuning and make online (continual) predictions. Thus, there is still room for improvement on existing AI methodologies and schemes.

More recently, cortical-like algorithms such as the Hierarchical Temporal Memory (HTM) have been a promising technique for several real world prediction tasks. However, very little work has been done in the area of applying cortical-like algorithms to electrical load time series data for forecasting in the short-term. These are algorithms that are modelled closely to the way the human brain operate. They learn a sparse distributed representation, form threshold coincidence maps, inculcate the notion of time and hierarchy and are capable of online (continual) learning [15–18]. Also as earlier mentioned, the problem of extensive hyper parameter tuning in the conventional artificial neural networks and the inability of most existing neural models to perform multiple/continual predictions makes cortical-like algorithms particularly very attractive [19].

1.2 Research objective

The primary objective of this research is to determine the effectiveness of the HTM Spatial Pooler's (HTM-SP) continual predictive ability on some open source electrical load time series datasets. In attempt to validate the effectiveness of this important machine intelligent technique, experimental simulations of daily and hourly electrical load forecasts using the HTM Spatial Pooler with an overlapping temporal classification (OTC) scheme are presented.

1.3 Structure of paper

This paper is structured as follows: in [Section 2](#) we describe the HTM model used for making short-term electric load time series forecasts. [Section 3](#) presents the experimental forecasting results on several electrical load time series data. Finally, we give our discussions in [Section 4](#) and conclusion/future work directions in [Section 5](#).

2. Hierarchical Temporal Memory

The Hierarchical Temporal Memory (HTM) is a constrained machine intelligence algorithm and neural network for continual learning tasks [20]. The principle of operation of HTM is based on the formation of sparse distributed representations and then learning and making predictions from these representations using neurobiological principles. HTM is implemented as a suite of algorithms called the Cortical Learning Algorithms (CLA); these algorithms are constrained algorithmic implementations of the operation of the neocortex, the seat of intelligence in the brain [17,20]. Basically, these algorithms consist of two core parts: A HTM Spatial Pooler (SP) for forming sparse distributed representations of real world sensory input or synthetic sensory-like data and a HTM Temporal Pooler (TP) part for making predictions on the SDRs formed by the HTM-SP.

SDRs are basic data structures of the HTM that capture the learning units used in the brain; the SDRs used in the HTM architecture follow the notion of sparse coding earlier suggested for learning sensory inputs in Olshausen and Field [21,22].

A typical neuron model used in a HTM system implementation is given in Figure 1. This model is inspired by neuroscience studies of activity-dependent synaptogenesis which borders on the growth and origin of the biological synapses stimulated by external sensory signals [23]. In this diagram, the connections comprising the green blobs represent proximal synapses which are typically linearly summed to produce a feed-forward activation; distal synapses are represented by segments of blue blobs that are or-ed (logically summed) to give a dendrite spiking neuron activation when they exceed a certain recognition threshold (indicated by the Sigma sign). Feedback and context experiences are formed using these distal connections.

Interestingly, as a machine intelligence cortical learning technique, there is an important difference on how HTM makes time series predictions when compared to other similar neural or machine learning models. In HTM, predictions are performed online in both spatial and

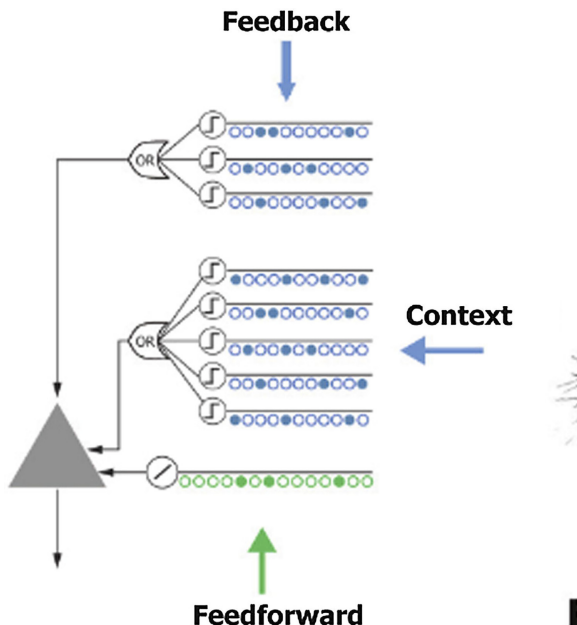


Figure 1.
An HTM Neuron
Model: adapted from
Ref. [24].

temporal domains using sparse distributed representations (SDRs). Rather than simply learning on a training set and later making predictions from the training test – as is the case with most other neural models, HTM continually predicts the underlying causes of the data at the current time step in the context of past data sequences. This is an obvious advantage and is very important particularly for real time processing and data analytic tasks. The Spatial Pooling (SP) and Temporal Pooling (TP) technique for temporal classification are presented in the following sub-sections (Section 2.1 and 2.2).

2.1 HTM Spatial pooling

In HTM, spatial pooling is performed using the notion of SDRs followed by competitive Hebbian learning rules, a Homeostatic excitability control, and an overlapping mechanism for deriving candidate or winner SDR patterns via inhibition [24]. SDRs are formed by activating or deactivating a set of potential synapses or connecting neural links. These synapses are grouped into a set of mini-columns and are spread out in a hypercube based on a set of predefined rules.

Consider a group of mini-columns with a set of potential connecting logical synapses or neurons; these potential connections may be initialized accordingly as:

$$\Pi_i = \{j | I(x_j; x_i^c, \gamma) \ \& \ Z_{ij} < \rho\} \quad (1)$$

where, j = HTM neuron location index in the mini-column, i = mini-column index, x_j = location of the j th input neuron (synapses) in the input space, x_i^c = location centre of potential neurons (synapses) of i th mini-column in a hypercube of input space, γ = edge length of x_j , ρ = fraction of inputs within the hypercube of input space that are potential connections, Z_{ij} = represents a uniformly distributed random number between 0 and 1, I = an indicator function

The indicator function is typically described by Eq. (2):

$$I(x_j; x_i^c, \gamma) = \begin{cases} 1, & \text{if } x_j \subset x_i^c \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

A set of connected synapses are described by a binary matrix, W , which is formulated by conditioning the synapses to a permanence activation rule as:

$$W_{ij} = \begin{cases} 1, & \text{if } D_{ij} \geq \theta_c \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where, D_{ij} = independent and identically distributed (i.i.d) dendrite synaptic permanence values from the j th input to the i th mini-column, θ_c = synaptic permanence threshold

The i.i.d synapse permanence values are described by Eq. (4) as:

$$D_{ij} = \begin{cases} U(0, 1), & \text{if } j \in \Pi_i \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Where a natural topology exists, neighbourhood mini-columns may be inhibited in accordance to the relation given in Eq. (5) otherwise a global inhibition parameter is simply used.

$$N_i = \{j | \|y_i - y_j\| < \varphi, j \neq i\} \quad (5)$$

where, y_i = is the i th HTM-SP mini-column, y_j = is the j th HTM-SP mini-column, i, j = mini-column indexes, φ = inhibition radius control parameter

For creating associations with input patterns, feed-forward inputs to each of the generative spatial mini-columns are computed using a matching technique called the overlap; this concept is diagrammatically illustrated in Figure 2. The overlap is computed as:

$$o_i = b_i \sum_j W_{ij} z_j \tag{6}$$

where, b_i = is a positive boost factor for exciting each HTM-SP mini-column, z_j = input pattern vector seen by the generative HTM neuron

Using Eq. (6), we can calculate the activation of each SP mini-column as:

$$a_i = \begin{cases} 1, & \text{if } o_i \geq Z(V_i, 100 - s) \ \& \ o_i \geq \theta_{stim} \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

$$V_i = \{o_i | j \in N_i\} \tag{8}$$

where, s = target activation density (sparsity), Z = a percentile function, θ_{stim} = a stimulus threshold

The HTM-SP uses a learning rule inspired by competitive Hebbian learning for reinforcing dendrite permanence values [19,24]. The learning rule can be calculated from the formula given in Eq. (9):

$$\Delta D_{ij} = p^+ D_{ij} \circ A^{t-1} - p^- D_{ij} \circ (1 - A^{t-1}) \tag{9}$$

where, p^+ = positive permanence value increment, p^- = negative permanence value increment, A^{t-1} = activation state at time step, t

Finally, boost updating in HTM-SP follows the homeostatic excitability control mechanism comparable to that observed in cortical neurons [25]. Boosting is accomplished in HTM-SP using the following model equations (Eq. 10–Eq. 12):

$$\bar{a}_i(t) = \frac{(T - 1) * \bar{a}_i(t - 1) + a_i(t)}{T} \tag{10}$$

$$\langle \bar{a}_i(t) \rangle = \frac{1}{|N_i|} \sum_{j \in N_i} \bar{a}_i(t) \tag{11}$$

$$b_i = e^{-\beta(a^-_i(t) - \langle \bar{a}_i(t) \rangle)} \tag{12}$$

where, \bar{a}_i = time averaged activation over the last T SDR inputs, T = an integer number denoting the number of Monte Carlo trials to obtain a reasonable activation estimate. $\bar{a}_i(t)$ = the current activity of the i th mini-column at time step t . β = a positive parameter that controls the strength of the adaptation effect.

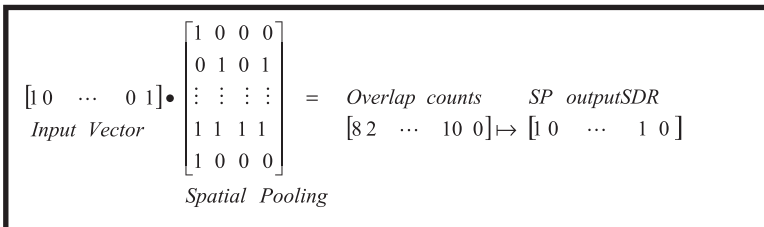


Figure 2.
An Illustrative concept
of Overlap in an HTM-
SP; Source: Ahmad and
Hawkins [15].

As mentioned in Ref. [24], such calculations have been used in previous models of homeostatic synaptic plasticity as in [26,27].

2.2 Temporal classifier

In the proposed HTM system, feed-back associations are built from the HTM Spatial Pooler (SP) SDRs using a temporal overlap classifier. The Temporal classifier uses the overlap technique which is similar to Eq. 6; however predictions are made by performing a match between a set of past SDR observations (used as context) and the current SDR observation. The temporal overlaps through time are obtained using Eq. (13):

$$o_{j_t} = \sum_{j_t} W_{j_t}^{sp} W_{(k-N_c)j_t}^{sp}, \quad N_c < k < j_t, \quad (13)$$

where, N_c = Number of past sample SDRs used as context, k = size of the temporal aggregated (bivariate) sequence through time, j_t = temporal aggregation index number, $W_{j_t}^{sp}$ = bivariate SDRs after temporal aggregation.

2.2.1 Temporal aggregation of HTM-SP SDRs

Temporal aggregation is used in the HTM-SP to build a cause-and-effect data sequence from the SDRs formed initially and then used for an overlapping temporal classification (OTC); such sequences have been assumed to possess a bivariate representational requirement – indeed HTM using an OTC scheme has been proven to be very effective in certain very advanced tasks such as drug discovery [28]. In HTM-SP, adding more variables increases the degree-of-freedom for making effective overlap matches. The temporal aggregation procedure used in the forecast analysis is as follows:

Step 1: Form a single-column vector matrix of length $1:N$ having a width of 1, where N represents the number of sampled sequences SDRs obtained from the HTM-SP stage. The elements in this matrix contain the indexes for temporal aggregation.

Step 2: For each element in the matrix formed in Step 1 greater than 1, perform a modulus operation such that if a remainder exists for the considered element we skip that element, otherwise we select the element; this operation results in single-column vector matrix of length approximately equal to $1:N/2$. The elements in this matrix contain the set of even indexes in the matrix obtained from Step1 at time instance, t . We call this set $A_{t(1)}$.

Step 3: For all elements in the set $A_{t(1)}$, form a concatenation of $A_{t(1)}$ with $A_{t(1)}$ 1-step behind as $\{A_{t(1)} A_{t-1(1)}\}$; this concatenation represent the temporal aggregator index set. We call this set of indexes $A_{t(agg)}$.

Step 4: Using $A_{t(agg)}$ as index sequence, extract SDR patterns obtained from the HTM-SP stage in a temporal aggregated fashion and then perform overlap temporal classification through time.

3. Experiments and results using electrical time series load data

The proposed systems architecture for day-ahead load predictions is shown in Figure 3. In this architecture, the encoder transforms real world input data signals into a binary representation suitable for spatial pooling; then the Hierarchical Temporal Memory Spatial Pooler (HTM-SP) forms sparse distributed representations (SDRs) of the binary representations using the generative procedure described in Section 2.1. The SDRs are

temporally aggregated and predicted using the overlapping temporal classification (OTC) scheme described in Section 2.2.

The Mean Absolute Percentage Error (MAPE) was chosen as the metric for evaluating the performance of the HTM-SP system. The use of MAPE is due to its insensitivity to outliers so it presents an unbiased metric when compared to other metrics such as Mean Squared Error (MSE) or absolute errors which exhibit large fluctuations in values for increasing/decreasing forecast values.

Other metrics for monitoring the performance of the HTM-SP include the skewness and the kurtosis. The skewness measures the asymmetry that exist in the error distribution around its mean; if the skewness value is negative then the error distribution is negatively skewed otherwise it is positively skewed. A skewness of zero implies that the error response is perfectly symmetrical and follows a normal distribution.

The kurtosis on the other hand is used to quantify the outlier prone behavior of the error distribution with respect to a normal distribution which is defined as having a kurtosis equal to 3. A kurtosis greater than 3 indicates that the error distribution is more outlier prone; if it is less than 3, it is said to be less outlier prone. The skewness and kurtosis functions in MATLAB were used for further evaluation of the MAPE error distribution.

The experimental tests were conducted using six different Electrical Time Series Load datasets:

The first two datasets comes from the Eunite Competition datasets organized by the Centre for Intelligent Technology Slovakia; it includes power readings for the years 1997 and 1998 containing a daily MW power reading for 24 h and recorded at 30 min intervals; special days such as Holidays and environmental parameters such as Temperature are also provided. The datasets can be obtained from (<http://eunite.org>). This dataset is open source and has also been used earlier in [1].

The third dataset is based on electric load time series dataset of Polish Power System from 2002 to 2004 [29]; this time series is similar to that of the Eunite competition dataset but with special labels for workdays or weekends.

The next three datasets were obtained from three different electricity markets – the German, French and British electricity markets; these datasets were used to investigate the robustness of the proposed approach. These datasets can be obtained from <https://open-power-system-data.org> and comprises hourly (60 min interval) total load in MW of electricity consumption for the years 2010–2013.

The core parameters used for the HTM-SP simulations are provided in the Appendix. A brief description of the parameters that motivated their use has also being provided in a subsequent sub-section (sub-Section 3.2). Source codes for the simulations performed in this section can be obtained from: <https://www.mathworks.com/matlabcentral/fileexchange/68442>.

3.1 Data forecasting task

For the experiments with HTM, we have only considered the maximum daily (or hourly) power reading i.e. we take the maximum power for the 24-hour duration of each day; the task required here is to continually predict n-days ahead, the power demand of the power system

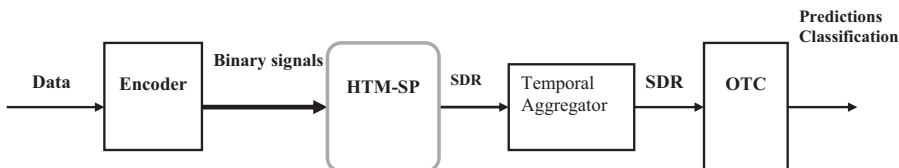


Figure 3.
System for Spatial-
Temporal Predictions
using the HTM-SP.

network based on the data provided. The use of only the maximum power demands makes it difficult for the HTM-SP to make predictions but also has the advantage of dimensionality reduction as the data is then transformed to a univariate time series; we reduce this difficulty by using the temporal aggregation procedure earlier introduced in [Section 2](#) (Sub-section 2.2.1) to form a bivariate distribution for the HTM-SP to learn from.

3.2 Motivation and description of the core HTM-SP system parameters

The core HTM-SP system parameters motivate the analysis of the experimental data presented in this section. The parameters are described succinctly in the following paragraphs.

3.2.1 Number of columns. The number of columns defines the learning extent of the cortical cells used during sensory signal processing. The higher this number, the more likely it is for the HTM-SP system to obtain a match and the more accurate will be the prediction but at the price of higher computational runtimes. Typical values can be as high as 1024 columns or more but a value of 250 was found sufficient to reduce the computational expense.

3.2.2 Initial Synaptic Permanence. This parameter defines how much permanence should be assigned to a cortical neuron prior to HTM-SP processing. By setting this value, we give the HTM-SP system a starting point to begin the wonderful process of activation or deactivation of cortical neurons or cells.

3.2.3 Reduct factor. This is used to adjust the dendritic segment activation threshold to a suitable value. This activation threshold is a function of the overlap metric earlier described in [Section 2](#). Once the activation threshold is met, the permanence's corresponding to the respective dendrite segment is activated and a connection is made to the receiving memory space which is stored in Random Access Memory (RAM).

3.2.4 Boost factor. The boost factor simply defines how much initial boost is used for adaptation (refer to [Eq. 12 Section 2](#), sub-section 2.1). Boosting ensures that sufficient support is given to weaker cortical cells in order to excite them to participate in the inhibition phase of the learning process.

3.2.5 Synapse permanence increment/decrement. This is a reward/punish scheme used to support (reinforce) cells or cortical neurons that do well during the update phase of the learning phase and weaken cells that perform poorly; if a cell contributes to learning then its permanence value is incremented otherwise it is decremented.

3.2.6 Number of past sequences used as context. This parameter plays the very important role of feedback matching or prediction. As mentioned earlier, the HTM-SP system is designed to perform feed-forward and feedback associations. It is this context parameter that makes the feedback function possible. Without providing context, it will be difficult if it not impossible for the HTM-SP to make machine intelligent predictions.

3.3 Experimental results using the Eunite datasets

The mean absolute percentage error (MAPE) values for single day-ahead forecasts using the HTM-SP on the Eunite datasets are given in [Figures 4 and 5](#). This represent a fluctuation of about 0.05% to 0.10% ([Figure 4](#)) for the Eunite 1997 competition data and 0.04% to 0.09% for the Eunite 1998 competition data.

3.4 Experimental results using the Polish dataset

In this section the MAPE values are reported for different n-step forecasts. In the first instance, MAPE values for 1 day-ahead HTM-SP predictions are reported in [Figure 6](#). This represents a fluctuation of about 0.05%–0.35%.

In the second instance, MAPE values for 7 days-ahead forecast are as shown in [Figure 7](#); the MAPE values fluctuate from about 0.05 to 0.35.

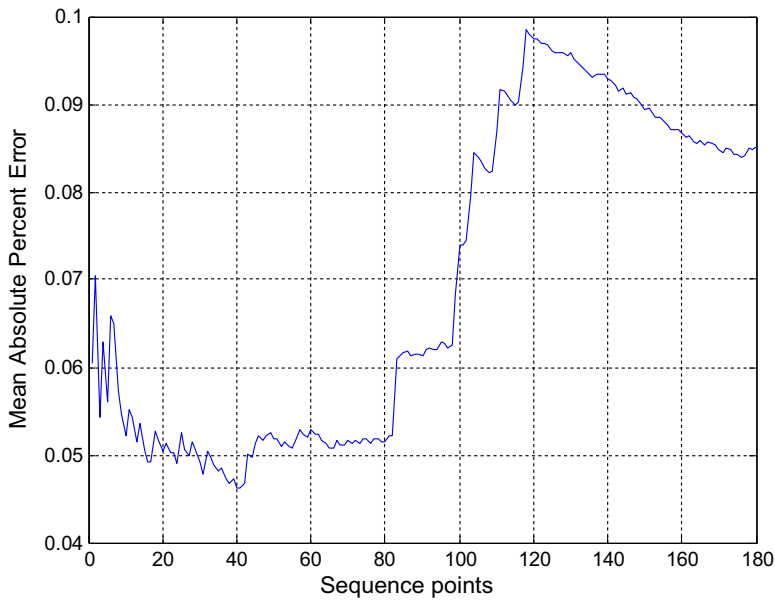


Figure 4.
Error Performance
using the 1997 Eunit
competition dataset for
1 day-ahead forecast.

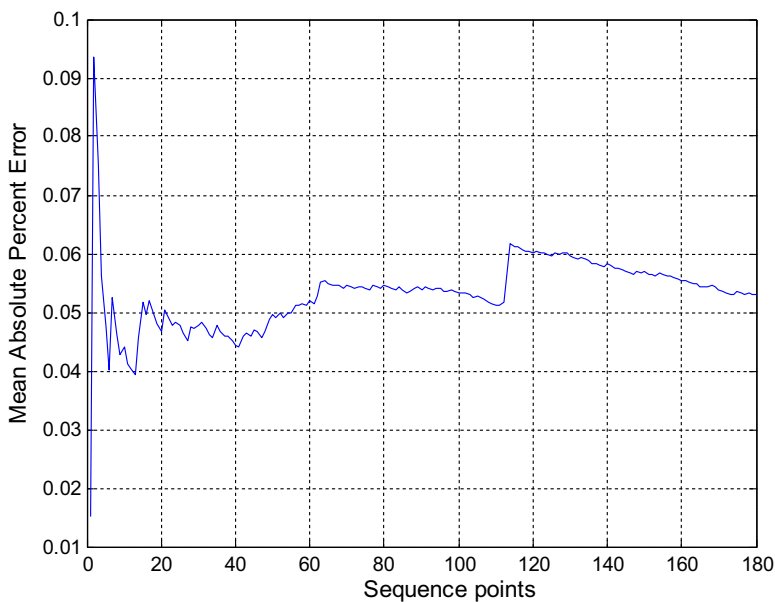


Figure 5.
Error Performance
using the 1998 Eunit
competition dataset for
1 day-ahead forecast.

In [Tables 1 and 2](#), the results showing the maximum value of HTM-SP MAPE for the case of 7 day-ahead and 1 day-ahead forecast of Polish electrical load times series data is compared to that reported in Ref. [\[5\]](#) and Ref. [\[29\]](#) respectively using other techniques; this comparative report clearly shows that the HTM-SP will outperform all these other algorithms.

Figure 6.
Error Performance
using the 2002–2004
Polish power dataset
for 1 day-ahead
forecast.

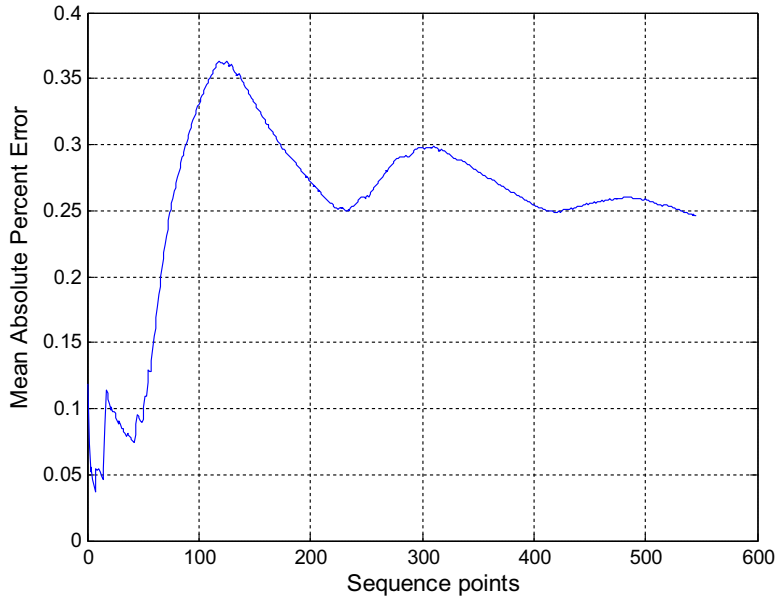
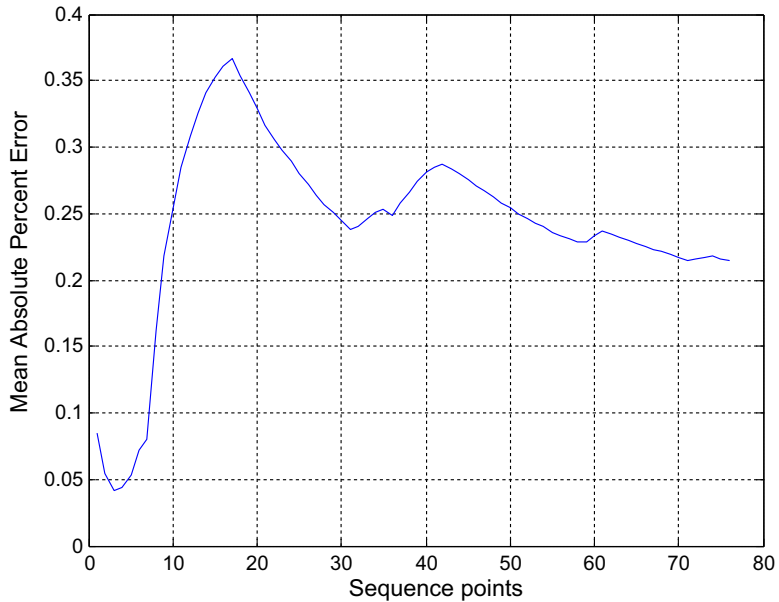


Figure 7.
Error Performance
using the 2002–2004
Polish power dataset
for 7 day-ahead
forecast.



3.5 Experimental results using datasets from the German, French and British markets

In this section, the mean MAPE, skew, and kurtosis values are reported for each market and for each year in consideration in [Tables 3–5](#) respectively. These values are obtained using the HTM-SP default parameters provided in the Appendix used earlier for single day-ahead forecasts.

4. Discussions

In this research, the HTM Spatial Pooler (SP) with an overlapping temporal classification (OTC) technique has been employed to the problem of short-term load forecasting. Experiments have been performed using electrical load time series datasets from the Eunite Competition, the Polish Power System and datasets from three well known electricity markets – the German, English and British power markets. In the whole, the results of these

Technique	MAPE value (%)
ANN	1.44
ARIMA	1.82
ES	1.66
Naïve	3.43
HTM-SP (proposed)	≈0.37

Table 1.
Reported MAPE values for 7 day-ahead forecasts using Polish dataset and different techniques in Ref [5] compared with the proposed HTM-SP.

Technique	MAPE value (%)
RF	1.16
CART	1.42
Fuzzy CART	1.37
ARIMA	1.91
ES	1.76
ANN	1.14
HTM-SP (proposed)	≈0.36

Table 2.
Reported MAPE values for 1 day-ahead forecasts using Polish dataset and different techniques in Ref [29] compared with the proposed HTM-SP; the values in Ref. [29] are due to mean value predictions for the months of January and July.

Electricity Market	MAPE ₂₀₁₀	MAPE ₂₀₁₁	MAPE ₂₀₁₂	MAPE ₂₀₁₃
Germany	0.1772	0.2260	0.1848	0.2366
France	0.3895	0.3604	0.3118	0.4975
Great Britain	0.2208	0.1477	0.1594	0.2077

Table 3.
MAPE Performance for 60 min (1-hour ahead) electricity load time series forecast of the German, French and Great British Electricity markets from 2010 to 2013.

Electricity Market	Skew ₂₀₁₀	Skew ₂₀₁₁	Skew ₂₀₁₂	Skew ₂₀₁₃
Germany	0.7056	0.4465	1.0335	0.2023
France	0.8928	-0.8803	1.0666	-0.4322
Great Britain	1.3735	-0.6409	-0.9399	0.1166

Table 4.
Skew for 1 hour-ahead electricity load time series forecast of the German, French and Great British Electricity markets from 2010 to 2013.

Electricity Market	Kurtosis ₂₀₁₀	Kurtosis ₂₀₁₁	Kurtosis ₂₀₁₂	Kurtosis ₂₀₁₃
Germany	02.4995	03.5009	04.1467	03.5330
France	02.2724	02.3412	02.9740	02.5534
Great Britain	23.1214	02.7369	02.1707	03.7979

Table 5.
Kurtosis for 1 hour-ahead electricity load time series forecast of the German, French and Great British Electricity markets from 2010 to 2013.

experiments indicate that the HTM-SP can continually predict the maximum load demand giving reasonable error accuracies. However, large fluctuations in error values may result due to the effect of seasonality – these effects leads to the peaks and troughs noticed in the MAPE error response plots in [Section 3](#) but they are not so critical as they fall under a much lower level than other techniques reported in the literature (see [Table 1](#)).

The MAPE performance of the German, French and British electricity markets, are as shown in [Table 3](#). These values seem to be much lower for the German and British electricity markets than the French. The measured error performances are also highly positively skewed with kurtosis values much lower than 3 as indicated in [Tables 4 and 5](#) respectively. In particular, the skewness is negative for only the French and British power markets for the years 2011, 2012 and 2013 while the kurtosis values are all below 3 for the French market indicating good outlier rejection capability of the HTM-SP for this class of datasets. The kurtosis of the Great British power market is very high - a value of about 23 for the year 2010 indicating that the error distribution is largely variable; however, for the year 2013, the kurtosis is not so far away from the base value of 3. Also, with the exception of the year 2010, the kurtosis is greater than base value for the German market. Thus, the HTM outlier rejection capability may be compromised due to error variations in some of the datasets considered but the deviations are generally not so critical.

5. Conclusion and future directions

Machine intelligence algorithms such as the Hierarchical Temporal Memory (HTM) based on the Cortical Learning Algorithms, presents an opportunity for industry and academic researchers in power systems to explore the possibility of using more responsive neural models for power demand forecasting. HTM can effectively learn patterns from the data using a continual learning spatial-temporal structured algorithm and help predict the load time series of a power system. This paper has focused on the forecasting in the short term of electrical time series loads using the HTM spatial pooler (HTM-SP) as an online (continual) learning prediction system and classifier. Simulation results using the HTM-SP system showed that it can outperform most existing artificial intelligence (AI)/neural techniques for the task of forecasting daily and hourly load demand using some reported datasets in the literature. This improvement may be attributed to the continual learning processes that occur within the HTM-SP system.

It is therefore recommended that short-term load forecasting algorithms use techniques that encourage continual learning.

However, the HTM-SP predictions may be compromised due to large variations in data which in turn is responsible for the high number of outliers. Thus, it is also recommended to further study the outlier rejection capability of the HTM-SP on novel datasets and devise means for handling them.

Future work should also focus on adapting HTM to conventional algorithms such as Genetic Programming to make the results obtained by its prediction mechanism model expressive. Other neuro-biological continual learning prediction techniques with simpler architecture should also be investigated in the context of forecasting time series.

Also, further real time experimentation may be necessary including real-time implementations to further validate this novel machine intelligence technique and possible variants for power systems load forecasting.

References

- [1] A.I. Saleh, A.H. Rabie, K.M. Abo-Al-Ez, A data mining based load forecasting strategy for smart electrical grids, *Adv. Eng. Informat.* 30 (3) (2016) 422–448.

-
- [2] G. Bao, Q. Lin, D. Gong, H. Shao, Hybrid Short-term Load Forecasting Using Principal Component Analysis and MEA-Elman Network, International Conference on Intelligent Computing, Springer, Cham, 2016, pp. 671–683.
- [3] B. Hayes, J. Gruber, M. Prodanovic, Short-term load forecasting at the local level using smart meter data. In PowerTech, IEEE Eindhoven (2015) (pp. 1-6). IEEE.
- [4] G. Sudheer, A. Suseelatha, Short term load forecasting using wavelet transform combined with Holt-Winters and weighted nearest neighbor models, Int. J. Electric. Power Energy Syst. 64 (2015) 340–346.
- [5] G. Dudek, Forecasting time series with multiple seasonal cycles using neural networks with local learning (pp. 52–63), International Conference on Artificial Intelligence and Soft Computing, Springer, Berlin, Heidelberg, 2013.
- [6] S.S. Reddy, C.M. Jung, K.J. Seog, Day-ahead electricity price forecasting using back propagation neural networks and weighted least square technique, Front. Energy 10 (1) (2016) 105–113.
- [7] S. Li, L. Goel, P. Wang, An ensemble approach for short-term load forecasting by extreme learning machine, Appl. Energy 170 (2016) 22–29.
- [8] D. Niu, Y. Lu, X. Xu, B. Li, Short-term power load point prediction based on the sharp degree and chaotic RBF neural network, Mathemat. Problem. Eng (2015).
- [9] K. Lang, M. Zhang, Y. Yuan, Improved neural networks with random weights for short-term load forecasting, PloS one 10 (12) (2015), e0143175.
- [10] Hochreiter, S., 1991. Untersuchungen zu dynamischen neuronalen Netzen. Diploma, Technische Universität München, 91.
- [11] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neu. Comput. 9 (8) (1997) 1735–1780.
- [12] F.M. Alvarez, A. Troncoso, G. Asencio-Cortés, J.C. Riquelme, A survey on data mining techniques applied to electricity-related time series forecasting, Energies 8 (11) (2015) 13162–13193.
- [13] F.M. Alvarez, A. Troncoso, J.C. Riquelme, J.S.A. Ruiz, Energy time series forecasting based on pattern sequence similarity, IEEE Trans. Knowled. Data Eng. 23 (8) (2011) 1230–1243.
- [14] J.F. Torres, A.M. Fernández, A. Troncoso, F. Martínez-Álvarez, Deep learningbased approach for time series forecasting with application to electricity load. International Work-Conference on the Interplay Between Natural and Artificial Computation, Springer, Cham, 2017.
- [15] S. Ahmad, J. Hawkins, Properties of Sparse Distributed Representations and their Application to Hierarchical Temporal Memory, arXiv preprint. arXiv, 1503 (2015).
- [16] George, D., Jaros, B., 2007. The HTM learning algorithms. Mar, 1, 44.
- [17] Hawkins, J., Ahmad, S., Dubinsky, D. (2010). Hierarchical temporal memory including HTM cortical learning algorithms. Technical report, Numenta, Inc, Palto Alto. https://web.archive.org/web/20110714213347/http://www.numenta.com/htm-overview/education/HTM_CorticalLearningAlgorithms.pdf.
- [18] M. Awad, R. Khanna, Deep Learning. In Efficient Learning Machines, Apress, Berkeley, CA, 2015, pp. 167–184.
- [19] Y. Cui, S. Ahmad, J. Hawkins, Continuous online sequence learning with an unsupervised neural network model, Neu. Comput. 28 (11) (2016) 2474–2504.
- [20] J. Hawkins, S. Ahmad, S. Purdy, A. Lavin, Biological and machine intelligence (BAMI), Init. Online Release (2016) 4.
- [21] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (6583) (1996) 607.
- [22] B.A. Olshausen, D.J. Field, Sparse coding of sensory inputs, Curr. Opin. Neurobiol. 14 (4) (2004) 481–487.
- [23] K. Zito, K. Svoboda, Activity-dependent synaptogenesis in the adult mammalian cortex, Neuron 35 (6) (2002) 1015–1017.

- [24] Y. Cui, S. Ahmad, J. Hawkins, The HTM Spatial Pooler—A Neocortical Algorithm for Online Sparse Distributed Coding, *Front. Comput. Neurosci.* 11 (2017).
- [25] G.W. Davis, Homeostatic control of neural activity: from phenomenology to molecular design, *Annu. Rev. Neurosci.* 29 (2006) 307–323.
- [26] C. Clopath, L. Büsing, E. Vasilaki, W. Gerstner, Connectivity reflects coding: a model of voltage-based STDP with homeostasis, *Nat. Neurosci.* 13 (3) (2010) 344.
- [27] S. Habenschuss, H. Pühr, W. Maass, Emergence of optimal decoding of population codes through STDP, *Neu. Comput.* 25 (6) (2013) 1371–1407.
- [28] V.I.E. Anireh, E.N. Osegi, An online cortical machine learning artificial intelligence technique for drug discovery, *Toxicol. Digest, West Afr. Soc. Toxicol.* 2 (1) (2018) 1–12.
- [29] G. Dudek, Short-term load forecasting using random forests. In *Intelligent Systems 2014*, Springer International Publishing, 2015.

A. Appendix

See (Table 6).

Parameter	Value
Number of Columns	250
Initial Synaptic Permanence	0.21
Reduct factor	2
Boost factor	100
Synapse permanence increment	0.1
Synapse permanence decrement	0.1
Number of past sequences used as Context	2

Table 6.
HTM-SP Parameters.

Corresponding author

E.N. Osegi can be contacted at: nd.osegi@sure-gp.com

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com