A New Entropy-based Feature Selection Method for Load Forecasting in Smart Homes

Omaji Samuel[1,a], Nadeem Javaid[1,*], Asma Rafique[2,b]
[1]COMSATS University Islamabad, Islamabad 44000, Pakistan
[2]Department of Computer Engineering, School of Sciences, Karabuk University, Turkey
[a]omajiman1@gmail.com, [b]asmamscs@gmail.com
*Corresponding author: nadeemjavaidqau@gmail.com

***Abstract*—This paper addresses the challenges of load forecasting that occur due to the complex nature of load in different predicting horizons and as well as the total consumption within these horizons. It is not often easy to accurately fit the several complex factors that are faced with demand for electricity into the predicting models. More so, due to the dynamic nature of these complex factors (i.e., temperature, humidity and other factors that influence consumption), it is difficult to derive an accurate demand forecast based on these parameters. As a consequence, a model that uses hourly electricity loads and temperature data to forecast the next hourly loads is proposed. The model is based on modified entropy mutual information based feature selection to remove irrelevancy and redundancy from the dataset. Conditional restricted Boltzmann machine (CRBM) is investigated to perform load forecasting; accuracy and convergence are improved to reduce the CRBM's forecast error via a Jaya based meta-heuristic optimization algorithm. The proposed model is implemented on the publicly available dataset of GEFCom2012 of the US utility. Comparative analysis is carried out on an existing accurate, fast converging short-term load forecasting (AFC-STLF) model since it has a similar architecture to the proposed model. Simulation results confirm that the proposed model improves the accuracy up to 56.32% as compared to 43.67% of AFC-STLF. Besides, the proposed model reduces the average execution time up to 53.87% as compared to 46.12% of AFC-STLF.**

***Index Terms*—Conditional Restricted Boltzmann Machine, Load Forecasting, Entropy-based Feature Selection, Smart Grid, Jaya Algorithm.**

I. Introduction

FORECASTING of electricity load is one of the vital parameters in the management of power grid system. Today, numerous decision about commitment of generators, setting reserves, maintenance and security, and scheduling of load demand are achieved via forecasting. Furthermore, it ensures a reduce operating cost and reliable supply of electricity.

Categories of electricity load forecasting are made in order to make meaning of the relationships between several forecasting trends; thus, electricity forecast is classified into four predicting horizons: very short-term load forecasting (VSTLF), i.e., a minute to hours ahead; short-term load forecasting (STLF), i.e., a day to weeks ahead; medium-term load forecasting (MTLF), i.e., months to a year ahead and long-term load forecasting (LTLF), i.e., years ahead.

Electricity load forecast with a better accuracy is challenging due to the nature and complexity of the load time series on a daily, weekly and annual basis. In addition, the fluctuation due to random components in individual electricity usage, irregular operating hours, special seasons, intermittent weather changes and industrial growth.

Fewer approaches to MTLF have been proposed; dynamic artificial neural network (ANN) model [1], modified Kalman filter and Walsh transform [2], ANN and fuzzy technique [3], singular value decomposition [4], neural network (NN) and evolution algorithm [5], Gaussian process regression and relevance vector regression [6]. Although, feature selection (FS) has not gained enough improvement in MTLF and it is the focus of this paper.

FS is the method of selecting the set representation of feature variables which is important and adequate to establish a forecasting model. Some applications of FS in research area for electricity load forecast: STLF [7]–[15], VSTLF [16], intelligent load forecasting [17]. Effective FS ameliorates the forecasting accuracy, making it faster to train with minimal complexity; thus, it improves load forecasting.

The major objective of this paper is to demonstrate how modified entropy mutual information (MI) based FS can be implemented to forecast electricity load, more importantly for MTLF. Specifically, the contribution of this paper is presented as follows.

1) An entropy MI-based FS is adopted because of its suitability to identify both nonlinear and linear relationships of the electricity load data; however, this proposed model enhances the work of [19] to remove more redundancy and irrelevancy features, see Subsection III-B.
2) In addition, an auxiliary variable for the FS is proposed. In this way, the four joint discrete random variables can be binary coded. For brevity, 24 hours sliding window are applied to reduce the dimensionality of candidate data while retaining the exact quality of the actual data. Furthermore, the quality of candidate features is evaluated and the final feature subsets are selected based upon ranking.
3) Finally, a state-of-the-art deep learning technique such as conditional restricted Boltzmann machine (CRBM), which is denoted as AR-MTLF is chosen to forecast electricity load. The forecast error accuracy of CRBM is

improved by using Jaya based meta-heuristic optimization algorithm to optimize the learning rate, number of neurons and the root mean square error (RMSE).

The main paper is organized as follows. Section II presents the related work. Section III describes the system model which involves the proposed MI-based FS and forecasting technique. Section IV shows simulations and discussions of the results. Finally, Section V presents the conclusions.

II. Related work

Nowadays, several approaches to load forecasting span from the conventional time series such as exponential smoothing, autoregressive integrated moving average, etc., to the computational intelligence, such as machine learning, i.e., NN, etc. The former approach is linear and model-based while the latter approach can model nonlinear input/output to the corresponding linear relations and has the tendency to learn from the given set of input data. On the other hand, the conventional approach can only fit a model and performs parameter estimation. This section discusses the MI based FS, afterwards, the electricity forecasting techniques.

Developing models for load forecasting involve selection of input features, previous work focuses pre-dominantly on filter and wrapper methods. To get a potential value feature of a hybrid selection scheme that incorporate both filter and wrapper methods, Zhongyi et al. [7] construct hybrid filter wrapper approach for STLF FS on a real-world hourly dataset of 4.5 years. A partial MI-based filter method is used to remove irrelevant and redundant features. The wrapper method combines the firefly algorithm and the support vector regression (SVR) to improve forecasting accuracy. However, the approach may be inefficient due to the high algorithmic complexity and extensive memory requirements. More so, wrong choice in selecting the kernel function parameters may affect performance.

Consideration of weather variations, economic factors, limited number of historical data (i.e., holiday data) as complex nonlinear load forecasting problem, several approaches to address this problem have been suggested. Young-Min et al. [8] present a fuzzy polynomial regression method with an hourly data selection based on Mahalanobis distance. A dominant weather feature to forecast holiday load is incorporated; thus, it provides forecasting accuracy by using the fuzzy polynomial regression method. However, the proposed method is not adaptable for large historical data, oscillation between exact fit value may occur if it falls outside the range of datasets. More so, polynomial regression assumed that the cause and effect relationship between the variables remains unchanged which does not suits the dynamic behavior of electricity load time series.

Jiang et al. [9] propose the date-framework strategy (DFS) to build a pool of features and model the FS technique. A genetic algorithm (GA) binary improved cuckoo search (GABICS) is used to locate a solution within the smallest reduction rate. To achieve robustness and high prediction accuracy, extreme learning machine (ELM) is applied to form the GABICS-DFS-ELM with a minimum subset of features, effectively. However, the framework does not consider spatial information to improve the effectiveness of load forecasting, i.e., the relationship between grids of different states. In addition, the framework only considers date and time series of the electricity load and does not consider other related factors associated to electricity load forecasting.

Grzegorz Dudek [10] presents an artificial immune system (AIS) for STLF; the AIS learns and recognizes antigens (AGs) from the fragment of the previous forecast (input vector) and fragments of the next forecast (output vector). A regression method is the proposed forecast model which uses a clonal selection mechanism to produce specific antibodies, i.e., recognizing AGs through selected features of input vectors and learn output vectors of the fragmented load time series. This proposed model is useful in classification, clustering and in solving optimization problems. However, the system does not includes other parameters such as weather data, thermal property of appliances and customers' behavior.

For efficient power system operation, electricity load demands should be satisfied by the electricity generation. In this regards, Yang et al. [13] propose an improved version of empirical mode decomposition (EMD) known as sliding window EMD (SWEMD) with a FS and hybrid forecast engine. The proposed FS maximizes the relevancy and minimizes the redundancy on a Pearson's correlation (MRMRPC) coefficient. Afterwards, prediction of the load signal performed by an improved Elman neural network (IENN) based forecast engine with an intelligent algorithm to achieve accurate prediction.

Implementing and choosing the best time series model is challenging since different models react differently to the same training data. A considerable amount of feature engineering is needed to find optimal time lag and informative features. Salah et al. [15] use long short-term memory (LSTM) based NN to model aggregated STLF. A GA is used to obtain the optimal time lags and the number of layers for the LSTM model and as well as the predictive performance optimization.

In this paper, limitations of the above existing work in the literature are considered by deriving model parameters for accurate and efficient precision. CRBM is preferred in this paper because of its ability to perform unsupervised (i.e., no labels required) learning and does not compare its output with labels. Moreover, CRBM stacked a layer on top of one another, i.e., using the conditioned hidden layer as the next input layer and iterate, thus, building a multi-layer neural network where each layer represents distinct data abstractions. The forecast error is minimized by using Jaya based meta-heuristic optimization algorithm which optimizes the learning rate and the number of neurons of each layer. The choice of selecting Jaya algorithm over other algorithms in the literature, is that, Jaya does not require algorithm specific control parameters.

Altogether, the system model consists of four modules as shown in Fig. 1: feature selector, a forecaster, optimizer and consumption dynamic which will be analyzed as future work. At first, the feature selector obtained historical time series of load and temperature data as an input, performed normalization and then gets candidates inputs of greater relevant information based on the modified entropy MI-based FS technique. The forecaster module consists of CRBM, which then received the selected candidates from the FS module and is activated by sigmoid and Gaussian activation functions. The next module is the optimizer which consists of an optimization algorithm that performs parameter settings of CRBM and minimizes the forecast error.

III. System model

The proposed system model consists of four modules: a feature selector, a forecaster, an optimizer and consumer dynamics. Initially, the load time series and temperature data are merged and normalized. The feature selector module uses the proposed modified entropy MI-based FS technique to remove redundancy and irrelevancy from the dataset and sort the selected candidates based on ranking. The sorted candidates are constructed based on average, previous and lagged observed data given to the forecaster module. In this module, CRBM is implemented to forecast the load based on the training, validation, and testing dataset. The forecast error is reduced at the optimizer module via Jaya based optimization algorithm using the iterative search process.

A. Data preparation and preprocessing

At first, the processing step begins by merging the electricity load with the temperature data of preceding hours and the moving average of the nth day which is computed using Eq (1) [18].

$$T_{t,n} = \frac{1}{24} \sum_{h=24n-23}^{24n} T_{t-h},\ n = 1, 2, \ldots, k. \tag{1}$$

Where t is the total time period. The merging of electricity load with temperature data is known to have an impact on the electricity consumption behavior. The preprocessing is used to remove zeros and outliers values, and scale the input data to a normalized pattern as given in Eq (2). The normalized data is then divided into three sets: training, validation and testing while still retaining the temporal order; here, the dataset is ready for further analysis.

$$Norm = \frac{X - mean(X)}{std(X)}, \tag{2}$$

where std denotes the standard deviation and X represents the input data.

B. Modified MI based FS

Chandrashekar and Ferat Sahin [23] present a survey of the various FS techniques: Filter methods which uses variable ranking approach as the main criteria for variable selection by ordering (e.g., correlation criteria and MI); wrapper methods which uses the predictor as a black box and the predictor performance as the objective function to evaluate the variable subset (e.g., sequential selection algorithm, heuristic search algorithm); other FSs are unsupervised learning techniques, semi-supervised learning and ensemble feature selection.

In this paper enhances the work of [19] by adding another parameters such as the temperature of preceding hours and the moving average of the observed data. First hour of the first day is used to predict the first hour of the next day and so on. From the Fig. 1, the value of n represents the total number of hours of a day. The value of n is connected to the fine adjustment of CRBM training.

The modified entropy based MI FS is implemented to remove the redundancy and irrelevancy by selecting the best subset that contains the minimal number of features and provide accurate load forecast; In this way, the dimensionality curse is avoided.

$$MI(p, p^t, p^m) = \sum_i \sum_j \sum_k \sum_l pr(p, p^t, p^m) \\ log_2(pr((p, p^t, p^m))), \tag{3}$$

The Eq (3) consists of four joint discrete random variables as defined in Eq (4).

$$MI(p, p^t, p^m, p^q) = \sum_i \sum_j \sum_k \sum_l pr(p, p^t, p^m, p^q) \\ log_2(pr((p, p^t, p^m, p^q))), \tag{4}$$

where $pr(p, p^t, p^m, p^q)$ is the joint probability of the four discrete random variables; p_i is the input discrete random variable. Let p_j^t be the target value, p_k^m is the mean value and p_l^q is the temperature of proceeding hours and the moving average. Hence, the four discrete random variables are important in the FS which is rewritten as:

$$MI(p, p^t, p^m, p^q) = \sum_i \sum_j \sum_k \sum_l pr(p, p^t, p^m, p^q) \\ log_2(\frac{pr(p, p^t, p^m, p^q)}{pr(p)pr(p^t)pr(p^m)pr(p^q)}). \tag{5}$$

If $MI(p, p^t, p^m, p^q) = 0$ this implies that the four discrete random variables are independent. In a likewise manner, if $MI(p, p^t, p^m, p^q)$ is large, then the four discrete random variables are closely related. Finally, if $MI(p, p^t, p^m, p^q)$ is small, then the four discrete random variables are loosely related. Ahmad et al. [19] select the target values as the last hours of the day from training dataset which represent the values of the previous day. In addition, They improve the forecast by including an average behavior. However, including average behavior and target values are not sufficient enough. As consequence, temperatures of the preceding hours as well as the moving average of the target data are also added. From Eq (5), the information is coded in a binary using the Eqs (6, 7 and 8).

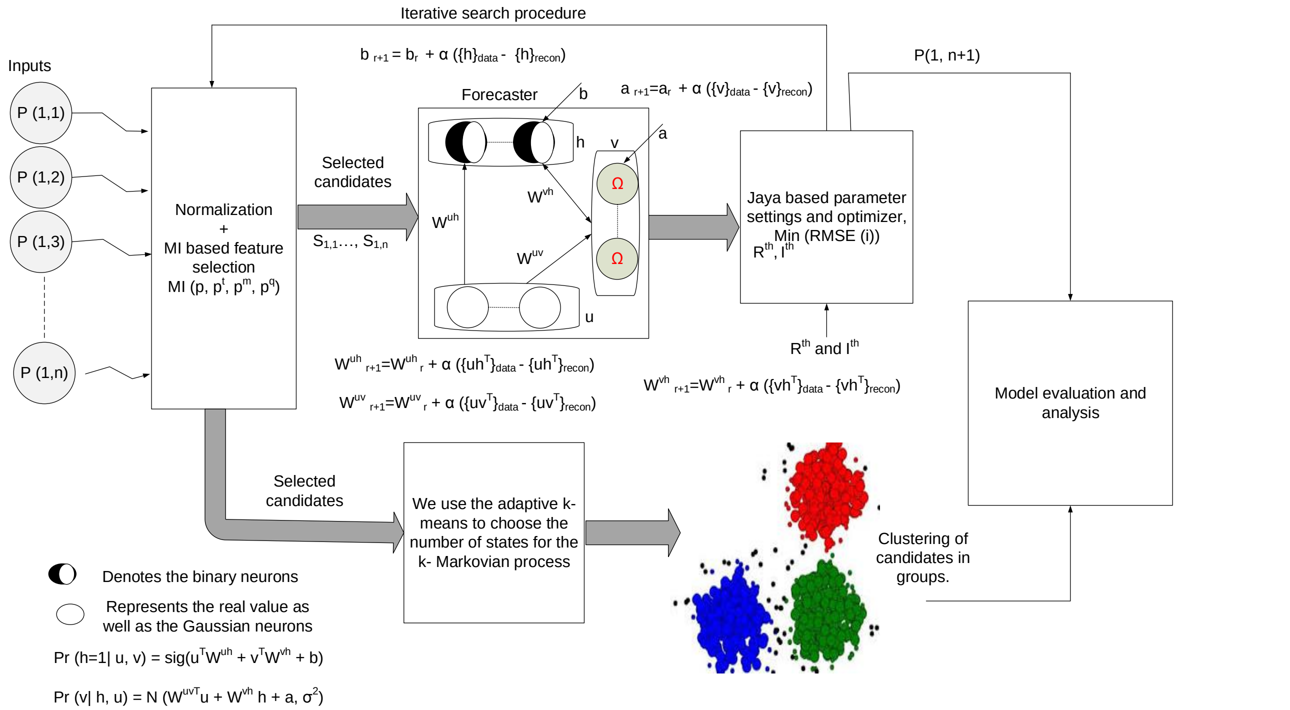

Fig. 1: Proposed system model. P represents the matrix probability of the input load time series and temperature data; W represents the weight of hidden, visible and history layer respectively; a and b are the biases of the visible and hidden layers; MI is the mutual information based feature selection; r is the iteration; α is the learning rate; $._{data}$ denotes the configuration of CRBM after its initialization with training data; $._{recon}$ denotes the configuration of CRBM after the Markov chain is performed; sig and N denote the sigmoid and Gaussian functions respectively.

$$
\begin{aligned}
MI(p, p^t, p^m, p^q) &= pr(p_i = 0, p_j^t = 0, p_k^m = 0, p_l^q = 0) \\
\times log_2(&\frac{pr(p_i = 0), pr(p_j^t = 0), pr(p_k^m = 0), pr(p_l^q = 0)}{pr(p_i = 0)pr(p_j^t = 0)pr(p_k^m = 0)pr(p_l^q = 0)}) \\
&+pr(p_i = 0, p_j^t = 0, p_k^m = 0, p_l^q = 1) \\
\times log_2(&\frac{pr(p_i = 0), pr(p_j^t = 0), pr(p_k^m = 0), pr(p_l^q = 1)}{pr(p_i = 0)pr(p_j^t = 0)pr(p_k^m = 0)pr(p_l^q = 1)}) \\
&+pr(p_i = 0, p_j^t = 0, p_k^m = 1, p_l^q = 0) \\
\times log_2(&\frac{pr(p_i = 0), pr(p_j^t = 0), pr(p_k^m = 1), pr(p_l^q = 0)}{pr(p_i = 0)pr(p_j^t = 0)pr(p_k^m = 1)pr(p_l^q = 0)}) \\
&+pr(p_i = 0, p_j^t = 0, p_k^m = 1, p_l^q = 1) \\
\times log_2(&\frac{pr(p_i = 0), pr(p_j^t = 0), pr(p_k^m = 1), pr(p_l^q = 1)}{pr(p_i = 0)pr(p_j^t = 0)pr(p_k^m = 1)pr(p_l^q = 1)}) \\
&+pr(p_i = 0, p_j^t = 1, p_k^m = 0, p_l^q = 0) \\
\times log_2(&\frac{pr(p_i = 0), pr(p_j^t = 1), pr(p_k^m = 0), pr(p_l^q = 0)}{pr(p_i = 0)pr(p_j^t = 1)pr(p_k^m = 0)pr(p_l^q = 0)}),
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
&+pr(p_i = 0, p_j^t = 1, p_k^m = 0, p_l^q = 1) \\
\times log_2(&\frac{pr(p_i = 0), pr(p_j^t = 1), pr(p_k^m = 0), pr(p_l^q = 1)}{pr(p_i = 0)pr(p_j^t = 1)pr(p_k^m = 0)pr(p_l^q = 1)}) \\
&+pr(p_i = 0, p_j^t = 1, p_k^m = 1, p_l^q = 0) \\
\times log_2(&\frac{pr(p_i = 0), pr(p_j^t = 1), pr(p_k^m = 1), pr(p_l^q = 0)}{pr(p_i = 0)pr(p_j^t = 1)pr(p_k^m = 1)pr(p_l^q = 0)}) \\
&+pr(p_i = 0, p_j^t = 1, p_k^m = 1, p_l^q = 1) \\
\times log_2(&\frac{pr(p_i = 0), pr(p_j^t = 1), pr(p_k^m = 1), pr(p_l^q = 1)}{pr(p_i = 0)pr(p_j^t = 1)pr(p_k^m = 1)pr(p_l^q = 1)}) \\
&+pr(p_i = 1, p_j^t = 0, p_k^m = 0, p_l^q = 0) \\
\times log_2(&\frac{pr(p_i = 1), pr(p_j^t = 0), pr(p_k^m = 0), pr(p_l^q = 0)}{pr(p_i = 1)pr(p_j^t = 0)pr(p_k^m = 0)pr(p_l^q = 0)}) \\
&+pr(p_i = 1, p_j^t = 0, p_k^m = 1, p_l^q = 0) \\
\times log_2(&\frac{pr(p_i = 1), pr(p_j^t = 0), pr(p_k^m = 1), pr(p_l^q = 0)}{pr(p_i = 1)pr(p_j^t = 0)pr(p_k^m = 1)pr(p_l^q = 0)}),
\end{aligned}
\tag{7}
$$

$$+pr(p_i=1,p_j^t=0,p_k^m=1,p_l^q=1)$$
$$\times log_2(\frac{pr(p_i=1),pr(p_j^t=0),pr(p_k^m=1),pr(p_l^q=1)}{pr(p_i=1)pr(p_j^t=0)pr(p_k^m=1)pr(p_l^q=1)})$$
$$+pr(p_i=1,p_j^t=1,p_k^m=0,p_l^q=0)$$
$$\times log_2(\frac{pr(p_i=1),pr(p_j^t=1),pr(p_k^m=0),pr(p_l^q=0)}{pr(p_i=1)pr(p_j^t=1)pr(p_k^m=0)pr(p_l^q=0)})$$
$$+pr(p_i=1,p_j^t=1,p_k^m=0,p_l^q=1)$$
$$\times log_2(\frac{pr(p_i=1),pr(p_j^t=1),pr(p_k^m=0),pr(p_l^q=1)}{pr(p_i=1)pr(p_j^t=1)pr(p_k^m=0)pr(p_l^q=1)})$$
$$+pr(p_i=1,p_j^t=1,p_k^m=1,p_l^q=0)$$
$$\times log_2(\frac{pr(p_i=1),pr(p_j^t=1),pr(p_k^m=1),pr(p_l^q=0)}{pr(p_i=1)pr(p_j^t=1)pr(p_k^m=1)pr(p_l^q=0)})$$
$$+pr(p_i=1,p_j^t=1,p_k^m=1,p_l^q=1)$$
$$\times log_2(\frac{pr(p_i=1),pr(p_j^t=1),pr(p_k^m=1),pr(p_l^q=1)}{pr(p_i=1)pr(p_j^t=1)pr(p_k^m=1)pr(p_l^q=1)}). \tag{8}$$

Auxiliary variable τ_m is introduced for the individual elements and the joint probability is given in Eq (6).

$$\tau_m = 8p^q + 4p^m + 2p^t + p, \tag{9}$$

where $\tau_m \in [0,1,\dots,15]$. τ_{0m} provides the number of elements for which $\tau_m = 0$ in the present column L, which denotes the length of input data (i.e., τ_{1m} gives the number of ones, τ_{2m} gives the number of twos and so on, until τ_{15m} gives the number of fifteens). Fig. 2 presents the number of elements of fifteen auxiliary variables. The individual joint probabilities of the each value of τ_m is given in Eq (10).

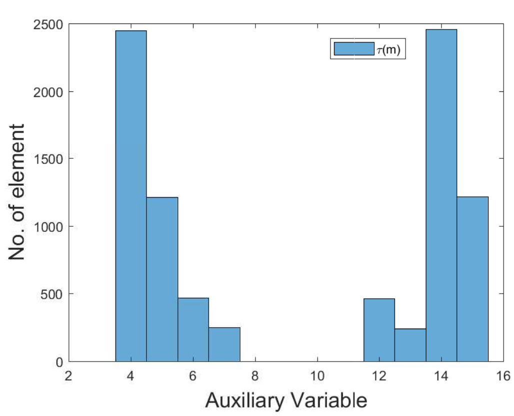

Fig. 2: The auxiliary variables.

$$pr(p=0) = \frac{\tau_{0m}+\tau_{2m}+\tau_{4m}+\tau_{6m}+\tau_{8m}+\tau_{10m}+\tau_{12m}+\tau_{14m}}{L}$$
$$pr(p=1) = \frac{\tau_{1m}+\tau_{3m}+\tau_{5m}+\tau_{7m}+\tau_{9m}+\tau_{11m}+\tau_{13m}+\tau_{15m}}{L}$$
$$pr(p^t=0) = \frac{\tau_{0m}+\tau_{1m}+\tau_{2m}+\tau_{3m}+\tau_{8m}+\tau_{9m}+\tau_{10m}+\tau_{11m}}{L}$$
$$pr(p^t=1) = \frac{\tau_{4m}+\tau_{5m}+\tau_{6m}+\tau_{7m}+\tau_{12m}+\tau_{13m}+\tau_{14m}+\tau_{15m}}{L}$$
$$pr(p^m=0) = \frac{\tau_{0m}+\tau_{1m}+\tau_{4m}+\tau_{5m}+\tau_{8m}+\tau_{9m}+\tau_{12m}+\tau_{13m}}{L}$$
$$pr(p^m=1) = \frac{\tau_{2m}+\tau_{3m}+\tau_{6m}+\tau_{7m}+\tau_{10m}+\tau_{11m}+\tau_{14m}+\tau_{15m}}{L}$$
$$pr(p^q=0) = \frac{\tau_{0m}+\tau_{1m}+\tau_{2m}+\tau_{3m}+\tau_{4m}+\tau_{5m}+\tau_{6m}+\tau_{7m}}{L}$$
$$pr(p^q=1) = \frac{\tau_{8m}+\tau_{9m}+\tau_{10m}+\tau_{11m}+\tau_{12m}+\tau_{13m}+\tau_{14m}+\tau_{15m}}{L}. \tag{10}$$

In the proposed FS based on Eq (10), the $MI(p, p^t, p^m, p^q)$ is derived by means of Eqs (6, 7 and 8). The selected candidates are rank according to the value of MI. In this way, irrelevancy and redundancy are removed. With respect to the forecaster and selected candidates, the selected candidate, $S_{1,1}, \dots, S_{1,n}$ is binary coded. The load pattern is different in all aspects, i.e., days, weeks, seasons, weekends and working days respectively. Moreover, the limitations of [19], which uses a less number of training dataset have been addressed.

C. CRBM based MTLF

The CRBM as shown in Fig.1 shows the configuration of the network. CRBM is the extension of the restricted Boltzmann machine, which is used to model time series and human activities [20]. This paper addresses the parameter setting of CRBM to train the network via an optimization algorithm and the classification of the electricity load from measured data. Since load forest is a time series problem, the CRBM is used to handle the large volume and high dimensional nonlinear data, it is capable of extracting multiple levels of distinct data abstraction. The concept of CRBM works where the higher levels are derived from the lower level ones. The detailed about CRBM is discussed in [20] and the parameters of CRBM used in this paper are presented in Table I.

In this paper, the root mean square error (RMSE) for the validation sample which is termed as the validation error.

$$RMSE(i) = \sqrt{\frac{\sum_{t=1}^{n}(y_t^- - y_t)^2}{n}}, \tag{11}$$

where y_t^- denotes the t^{th} actual load, and y_t represents the t^{th} forecast load. The total time T can represent the hourly,

Parameter	Value
Population size	24
Number of decision variable	2
Maximum iteration	100
Max	0.9
Min	0.1
Number of output layer	1
Number hidden layer	10
Learning rate	0.001
Weight decay	0.0002
Momentum	0.5

TABLE I: Simulation parameters; Max and Min are upper and lower population bound.

daily, weekly, seasonal or yearly time trends. Therefore, the final minimal value of $RMSE$ after a series of iterations is used as the validation error.

D. Forecast error minimization

The forecaster module returns the MTLF value of the next day with a minimum forecast error that represents the capability of the CRBM activation function and algorithm. The RMSE is further minimized using the Jaya optimization algorithm and objective function, mathematically it is written as:

$$\underset{R^{th}, I^{th}}{minimize} \quad RMSE(i) \; \forall \; i \in [1, 2, \ldots, n], \tag{12}$$

where R^{th} and I^{th} are redundancy and irrelevancy thresholds, respectively; which are chosen to be 0.05. Preferring Jaya algorithm over other heuristic algorithms is its tendencies to achieve a globally optimal solution within small execution time. Moreover, it only requires commonly known parameters (i.e., population, elite size, etc.). On the other hand, algorithmic specific control parameters (i.e., crossover probability, mutation probability, etc.) are not required for the optimization.

Jaya algorithm was developed by Rao in 2016 [21], to solve the constrained and non-constrained optimization problem. It serves as a tool for providing optimal solutions in different domains like the microgrid [22], smart grid. The parameters of Jaya based optimization algorithm are shown in Table I. The accuracy equation is given below.

$$accuracy = 100 - RMSE, \tag{13}$$

where the $accuracy$ is measure in (%).

IV. Simulations and Discussions

In regard to the performance evaluation of our proposed AR-MTLF for data analysis. AR-MTLF is compared with AFC-STLF model [19] and is chosen due to its similarity in architecture to our proposed AR-MTLF model. Based on a fair comparison of AFC-STLF with that of ours, the same dataset collected from the hierarchical load forecasting record of GEFCom2012 which contains 4.5 years of hourly temperature and load across 21 zones of the US utility is implemented. The dataset is divided into three parts; for the training set, the first and second are taken from year (2004-2005). Whereas, 2006 for validation and lastly, 2007 for testing as shown in Fig. 3.

This section discusses the simulations in the following format: (A) by hourly load forecasting, (B) by seasonal load forecasting, (C) by performance evaluation in terms of error performance and convergence evaluation.

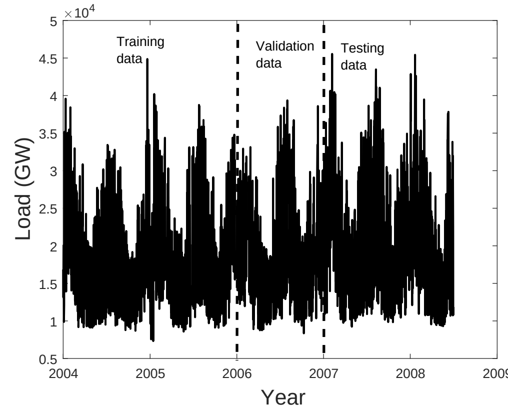

Fig. 3: Data from GEFCom2012 load forecast record.

A. By hourly load forecasting

The hourly demand for electricity normally takes the shape of multiple seasonal trends and can be hours of a day, a month of the year, days of the week, etc., which instantly forms a temporal hierarchy of the calendar variables.

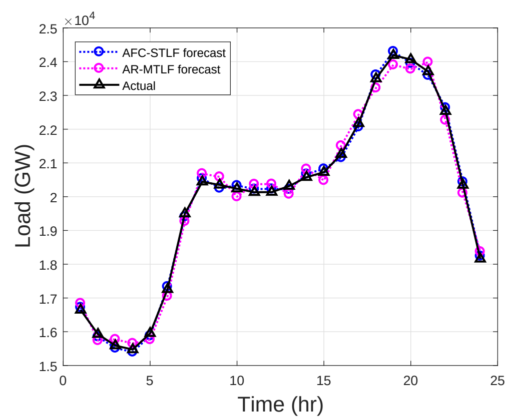

Fig. 4: Hourly load forecast.

The Fig. 4 shows the 24 hour load forecasting of Z_1. It is seen that both model are able to forecast efficiently.

B. By seasonal load forecasting

The Fig. 5 shows the actual load and our proposed AR-MTLF overlaid with the temperature for a summer week. Although, there is no over or under forecast in our proposed model. However, AFC-STLF did not actually learn and train the network efficiently.

The AFC-STLF over and under forecast the winter troughs for these days (11/13-14), and (11/16-22) as shown in Fig. 6.

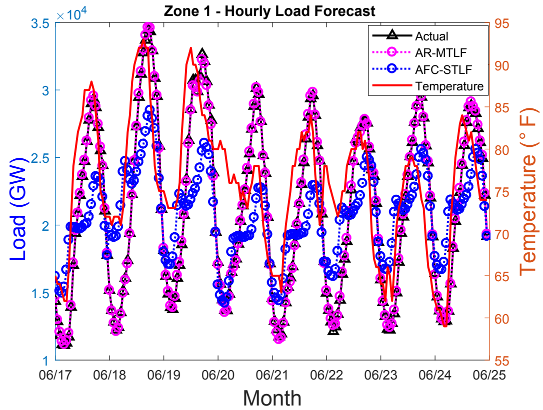

Fig. 5: Summer load prediction.

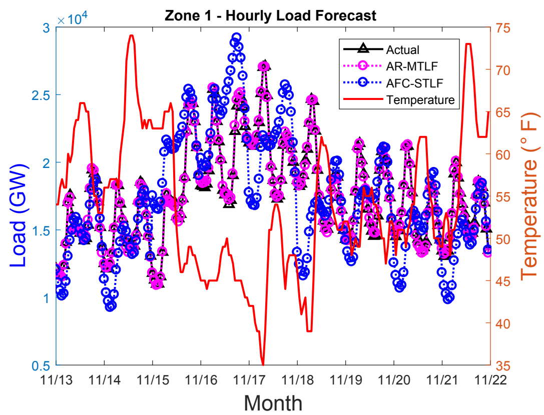

Fig. 6: Winter load prediction.

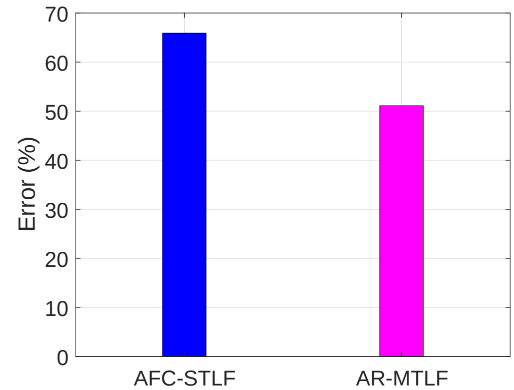

Fig. 7: Error performance.

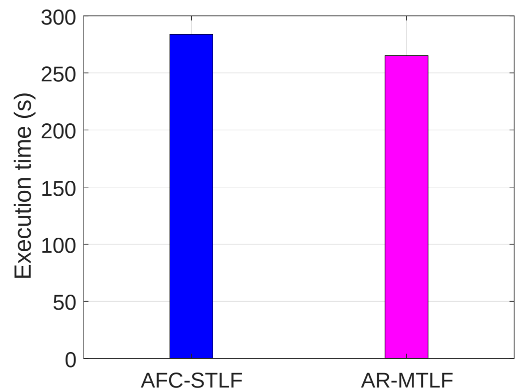

Fig. 8: Convergence rate analysis.

C. By performance evaluation in terms of error performance and convergence evaluation

The performances of the proposed models are measured on a numerical value as shown in Fig. 7. RMSE is used to measure the deviation between the forecast value and the actual value. The smaller the RMSE value is, the higher the accuracy the model achieves. The figure shows that the RMSE value for AR-MTLF is 51.09 as compared to 65.89 of the AFC-STLF. The performance of AR-MTLF is due to the incorporation of the Jaya based optimization algorithm. However, forecast error is minimized at the expense of execution time.

Fig. 8 presents the execution time of AR-MTLF and AFC-STLF to be 213.90 and 249.79 seconds, respectively. Our proposed AR-MTLF minimizes execution time because of following reasons: AR-MTLF uses Jaya based optimization which finds the global optimal solution within the smallest execution time and uses the CRBM which performs better than ANN.

In addition, the AR-MTLF FS process reflects on three parameters: the lagged temperature data of the preceding hours, the moving average of the observed data and average behavior which denotes as the data of the previous hours. In contrast, AFC-STLF considers the last sample and average behavior. As a consequence, AR-MTLF reduces the average forecast error up to 56.32% as compared to 43.67% of the AFC-STLF model.

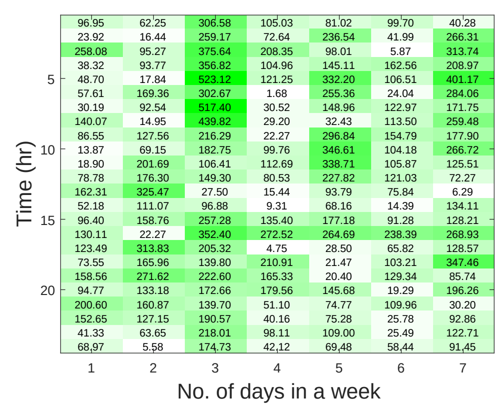

Fig. 9: Heat map for RMSE based on proposed forecast strategy on test dataset: y-axis represents the 24 hours while x-axis represents the days in a week.

Considering the different hours of a day and the number of days of a week, the heat map is derived by relating to the RMSE values of the testing dataset (2007), as depicted in Fig. 9. This illustrates the first 24/7 to avoid verbosity. The minimum RMSE value gives an accurate forecast in that particular hour of the day. Using the model (h =17, d=4) gives

RMSE value of 4.75.

V. Conclusion

Several load forecasts of hourly, daily, monthly and yearly electricity consumption have been proposed by previous researchers. This paper focuses on the hourly electricity load forecast for the month ahead, because of its importance for the outage and operational planning of electric power systems. However, due to the manner of smart meter data collection, scaling of this data is necessary. Here, a modified entropy MI FS is implemented to remove irrelevancy and redundancy from the dataset. Besides, the fundamental relationship between temperature and electricity load for preceding hours is investigated. The newly proposed AR-MTLF achieves approximately 56.32% accuracy, which proves better than the existing AFC-STLF of 43.67% based forecast strategy. The AR-MTLF model has reduced the average execution time up to 53.87% as compared to 46.12% of AFC-STLF. Hence, the modified entropy MI FS and CRBM prove justification for the correctness of the proposed model.

References

[1] Ghiassi, M. D. K. Z., David K. Zimbra, and H. Saidane. "Medium term system load forecasting with a dynamic artificial neural network model." Electric power systems research 76, no. 5 (2006): 302-316.

[2] Bhattacharya, T. K., and T. K. Basu. "Medium range forecasting of power system load using modified Kalman filter and Walsh transform." International Journal of Electrical Power & Energy Systems 15, no. 2 (1993): 109-115.

[3] Gavrilas, M., I. Ciutea, and C. Tanasa. "Medium-term load forecasting with artificial neural network models." In Electricity Distribution, 2001. Part 1: Contributions. CIRED. 16th International Conference and Exhibition on (IEE Conf. Publ No. 482), vol. 6, pp. 5-pp. IET, 2001.

[4] Abu-Shikhah, Nazih, and Fawwaz Elkarmi. "Medium-term electric load forecasting using singular value decomposition." Energy 36, no. 7 (2011): 4259-4271.

[5] Amjady, Nima, and Farshid Keynia. "Mid-term load forecasting of power systems by a new prediction method." Energy Conversion and Management 49, no. 10 (2008): 2678-2687.

[6] Alamaniotis, Miltiadis, Dimitrios Bargiotas, and Lefteri H. Tsoukalas. "Towards smart energy systems: application of kernel machine regression for medium term electricity load forecasting." SpringerPlus 5, no. 1 (2016): 58.

[7] Hu, Zhongyi, Yukun Bao, Tao Xiong, and Raymond Chiong. "Hybrid filterwrapper feature selection for short-term load forecasting." Engineering Applications of Artificial Intelligence 40 (2015): 17-27.

[8] Wi, Young-Min, Sung-Kwan Joo, and Kyung-Bin Song. "Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment." IEEE Transactions on Power Systems 27, no. 2 (2012): 596.

[9] Jiang, Ping, Feng Liu, and Yiliao Song. "A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting." Energy 119 (2017): 694-709.

[10] Dudek, Grzegorz. "Artificial immune system with local feature selection for short-term load forecasting." IEEE Transactions on Evolutionary Computation 21, no. 1 (2017): 116-130.

[11] Liu, Jin-peng, and Chang-ling Li. "The short-term power load forecasting based on sperm whale algorithm and wavelet least square support vector machine with DWT-IR for feature selection." Sustainability 9, no. 7 (2017): 1188.

[12] Dong, Yunxuan, Jianzhou Wang, Chen Wang, and Zhenhai Guo. "Research and Application of Hybrid Forecasting Model Based on an Optimal Feature Selection SystemA Case Study on Electrical Load Forecasting." Energies 10, no. 4 (2017): 490.

[13] Liu, Yang, Wei Wang, and Noradin Ghadimi. "Electricity load forecasting by an improved forecast engine for building level consumers." Energy 139 (2017): 18-30.

[14] Yang, Lintao, Honggeng Yang, Hongyan Yang, and Haitao Liu. "GMDH-Based Semi-Supervised Feature Selection for Electricity Load Classification Forecasting." Sustainability 10, no. 1 (2018): 217.

[15] Bouktif, Salah, Ali Fiaz, Ali Ouni, and Mohamed Adel Serhani. "Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches." Energies 11, no. 7 (2018): 1-20.

[16] Koprinska, Irena, Mashud Rana, and Vassilios G. Agelidis. "Correlation and instance based feature selection for electricity load forecasting." Knowledge-Based Systems 82 (2015): 29-40.

[17] Fallah, Seyedeh Narjes, Ravinesh Chand Deo, Mohammad Shojafar, Mauro Conti, and Shahaboddin Shamshirband. "Computational Intelligence Approaches for Energy Load Forecasting in Smart Energy Management Grids: State of the Art, Future Challenges, and Research Directions." Energies 11, no. 3 (2018): 596.

[18] Wang, Pu, Bidong Liu, and Tao Hong. "Electric load forecasting with recency effect: A big data approach." International Journal of Forecasting 32, no. 3 (2016): 585-597.

[19] Ahmad, Ashfaq, Nadeem Javaid, Mohsen Guizani, Nabil Alrajeh, and Zahoor Ali Khan. "An accurate and fast converging short-term load forecasting model for industrial applications in a smart grid." IEEE Transactions on Industrial Informatics 13, no. 5 (2017): 2587-2596.

[20] Mocanu, Elena, Phuong H. Nguyen, Madeleine Gibescu, and Wil L. Kling. "Deep learning for estimating building energy consumption." Sustainable Energy, Grids and Networks 6 (2016): 91-99.

[21] Rao, R. "Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems." International Journal of Industrial Engineering Computations 7, no. 1 (2016): 19-34.

[22] Samuel, Omaji, Nadeem Javaid, Mahmood Ashraf, Farruh Ishmanov, Muhammad Khalil Afzal, and Zahoor Ali Khan. "Jaya based Optimization Method with High Dispatchable Distributed Generation for Residential Microgrid." Energies 11, no. 6 (2018): 1-29.

[23] Chandrashekar, Girish, and Ferat Sahin. "A survey on feature selection methods." Computers & Electrical Engineering 40, no. 1 (2014): 16-28.