Pakistan Journal of Science (Vol. 66 No. 2 June, 2014)

MACHINE LEARNINGBASED GENERIC LOAD FORECASTINGMODEL FOR NOISY DATA: LESCO CASE STUDY WITH WEATHER INFLUENCE

S. M. Awan[*1,2], M. Aslam[1], Z. A. Khan[3], H. Saeed[2]

[1]Department of Computer Science and Engineering,
[2]Al-Khawarizmi Institute of Computer Science,
[3]Department of Electrical Engineering, UET, Lahore.
Corresponding Author E-Mail:[*]shahidawan@kics.edu.pk

ABSTRACT: Electric load forecasting (LF) involves the projection of peak demand levels and overall energy consumption patterns to support an electric utility's future system and business operations. Short and mid-range predictions of electricity load allow electricity companies to retain high energy efficiency and reliable operation. Absence of such prior planning results in a current crisis like situation in Pakistan, where power generation is not up-to the mark, its fallout is forced load shedding and voltage instability. To solve the problem of accurate LF, a variety of models is reported in literature. However, the accuracy of modeling techniques is extremely dependent on data quality. Since, the data recording in power systems of Pakistan is manual and it contains abnormalities like missing values, outliers, and duplication of records. Observing all the aforementioned problems, authors got motivation to devise such a LF model that can perform well on noisy data of Pakistan power systems and can handle load affecting parameters of this region. In this paper, a customized LF model formulation is presented, which incorporates machine learning techniques for data pre-processing, analysis, and model development.

Key words: Load Forecasting, Artificial Neural Networks, Optimization Techniques, Data Pre-processing.

INTRODUCTION

Load forecasting (LF) (Suganthi and Samuel, 2012)is an essential tool for any electric power supply utility in the world. Accurate estimates of future demand help in planning generation and distribution of electricity(Hahn *et al.,* 2009). It helps in keeping a balance between generation and distribution; increasing efficiency and reliability of the system in an economic manner (Alfares and Nazeeruddin, 2002).

Due to limitation of energy resources, poor planning, environmental factors, and limitation of transmission network; power sector in Pakistan is unable to fulfill the need of its consumers. Still, determining the need and behavior of consumers can help to provide available energy efficiently. Absence of this, results in a current handicapped situation of Pakistan power sector, in which load is either underestimated or overestimated, both situations are not acceptable for a reliable operation of electric utility.

In recent past, a large literature has evolved using efficient models to solve the accurate LF problem. With the advent of modern and powerful techniques, researchers also incorporated these to develop better LF models (Suganthi and Samuel, 2012). Such techniques come from both statistical and artificial intelligence (AI) domains.The statistical category includes time series (Amjady, 2001), regression based method (Charytoniuk *et al.,* 1998), radial basis functions (Xia *et al.,* 2010), and support vector regression (SVR) (Elattar *et al.,* 2010). Whereas, AI methods include expert systems (Liao, 2005), artificial neural networks (ANN) (Amjady and Keynia, 2011) and fuzzy inference (Che *et al.,* 2012). Recently much emphasis is being laid on the use of hybrid models. The most recent survey of such models is reported in (Suganthi and Samuel, 2012), covering all the state of the art methods of LF.A detailed comparison of LF models is presented in (Soares and Medeiros, 2008).

ANN's and support vector machines (SVM) (Vapnik, 1995), emerge as two competitive and successful techniques from AI and statistical domains respectively. AI modelsare discussed and evaluated in (Taylor and McSharry, 2007). In contrast to neural network (NN) theory, SVM (Vapnik, 1995), is a statistical tool for classification and regression. It has greater ability of generalization and to avoid over-fit to data (Jakkula, 2006). Still, ANN is known to suffer the slow convergence and trap into local minimum problems due to its gradient descent (GD) based learning process (Haykin, 1999). Similarly, finding the best parameters for SVM is another issue. There has been a tremendous development to overcome these issues, and a number of optimization techniques(OT's) (Sra *et al.*, 2011) are used to train ANN's and to find the optimum parameters of SVM.

A bunch of load forecasting models developed using such optimization techniques are described in (Suganthi and Samuel, 2012). In the work of (Saini and Soni, 2002), estimation of daily peak demand has been reported based on eleven weather parameters. To overcome the slow convergence problem of neural networks, Levenberg-Marquardt algorithm (LMA) (Wilamowski and Yu, 2010) is used as learning scheme. Simulated annealing (SA) (Aguiar e Oliveira Junior *et al.,* 2012) is used with ANN to solve the electric load forecasting problem (Pai and Hong, 2005).Adaptive PSO(Kennedy, 2010)is used to find optimal parameters for SVM in (Huang Yue *et al.,* 2009). In (Saini, 2008), resilient back propagation (RPROP) is used for training ANN to get 7 days ahead electrical peak load forecasting.

Machine learning assist not only in model development for forecasting but it also helps in data pre-processing; for example, to fill missing values, detect outliers, and remove duplicated values. As data collection and recording is manual in Pakistan power systems (PPS). There are many human errors while recording data. Similarly data contains missing records, and duplication of data is often the case. As accuracy of forecasting models heavily depend on the quality of data under study(Hahn *et al.,* 2009). This also includes dependent and independent factors. Therefore noise in independent factors can also lead to reduced accuracy of forecasts.

Observing the aforementioned problems with PSS and scrutinizing the issues with recorded load data, authors got motivation to solve the problem of accurate LF for noisy data of indigenous power sector.Part of this work is reported in (Awan *et al.,* 2012).The hypothesis behind this study is that modeling techniques can perform well on our datasets, when the data is smooth and free of abnormalities. In order to achieve better accuracy on noisy data sets; this study involves investigation through different statistical and heuristic techniques to suggest an accurate load forecasting model that can fulfill the above mentioned requirements of PSS. For this, authors of this paper have incorporated machine learning techniques for model development, optimization, and data preprocessing. An analytical study is performed on varying behavior of different techniques on predicating futuristic load utilizing different combination of techniques. In this way authors have modeled the complex behavior between different input factors and consumption behavior to predict future loads.

MATERIALS AND METHODS

This experimentation is performed on 4 years (2008-2011) hourly and 6 years (2005-2011) daily peak electricity demand data sets obtained from Lahore Electric Supply Company (LESCO) (http://lesco.gov.pk). Hourly data contain 24 data entries for a single day. Whereas, daily peak data contain only a single entry for each day. This dataset is manually recorded and contains typographical errors, resulting in missing load values and outliers. As energy demand on short term is highly dependent on weather conditions. Sudden change in temperature causes fluctuation on demand graph. Based upon high correlation, different input factors like calendar events, demand of last hour, and previous day, and weather conditions are considered as input parameters in this study. In order to achieve the desired objective of accurate LF model, it is necessary to compare all models on same criteria, for that reason data and features sets are identical for all models. For training the LF models, 70% of data is utilized and remaining 30% is divided into two equal halves for testing and validation. Further, the data is normalized before passing to modeling techniques. Here, min-max normalization (Jain *et al.,* 2005) method is used to scale attribute data set into range [0, 1]. This is defined by the formula given in Eq.1.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Where x' is the normalized value of x, $x_{min}$and x_{max} are the minimum and maximum values of x respectively.

Load data characteristics: There is a complex and nonlinear relationship between the electricity load and its influencing factors. Electric demand on short-term horizon is affected by calendar inputs, weather conditions, and energy price. The calendar events are time of day, day of week, and holidays. It is important to analyze data to find out correlation and dependencies between dependent (load in this case) and independent variables (load affecting variables). As in graph given in Figure1-a relationship between demand and temperature is shown. Here, 5 day profile is presented. It is clear from the graph that spikes in temperature cause increase in demand. However, time of the day is also a major factor in load fluctuation. Whereas, in Figure1-b association of demand with previous values is presented.

STLF process: The goal of accurate demand forecasting can only be achieved, when data is processed for abnormalities before passing it to prediction models. The whole process is divided into four steps, which are input pre-processing, feature selection, normalization, and result interpretation. In order to achieve the desired objective of accurate results of forecasting on noisy data, machine learning techniques are applied on every step of proposed system. The diagram in Figure 2 shows the essential components of this process. These steps are further elaborated in subsequent sections.

Data pre-processing and analysis: Data pre-processing is first and foremost step while building LF models. Pre-processing is applied on data under study to filter outliers, missing, and duplicated values. This effort is carried out to remove irregularities and smooth the load curve, ultimately resulting in more accuracy of forecasts. The pre-processing techniques are applied in following manner.

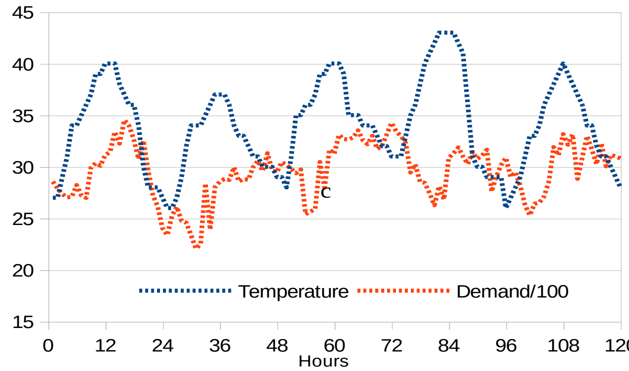

Fig1-a: Effect of temperature on demand

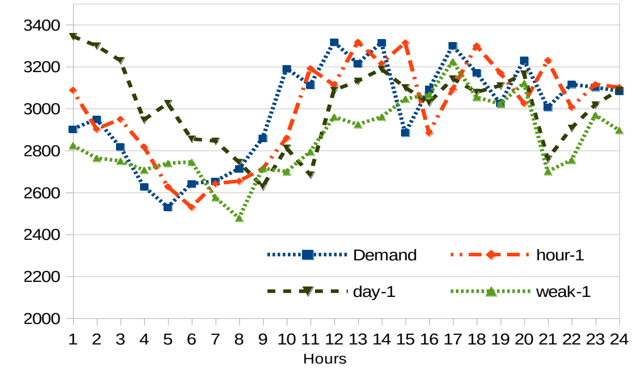

Fig1-b:Relationship of current & previous demands

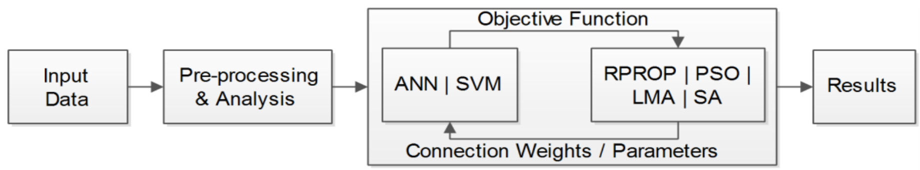

Fig.2: Components of STLF model

Outlier detection: An outlier is defined as an observation that "appears" to be inconsistent with other observations in the data set (Hodge and Austin, 2004). In the case of LESCO, an addition or removal of one digit can change the demand graph drastically. In this study, different techniques are analyzed to process outliers, including Box Plot, Z-Score, modified Z-Score, 2-Sigma (Hodge and Austin, 2004). Though Box Plot method produced most accurate results.

Treating missing values: Missing data is one of the major issues in load forecasting. Missing data arise in almost all serious statistical analyses. Filling missing values is called interpolation, and is a type of regression to treat time series data. After careful observation of LESCO load data, many records are found null or either filled with value '0'.There are various methods for interpolation, including some relatively simple approaches that can often yield reasonable results. In this study, weighted moving average (WMA), regression models, kalman filter, and exponential smoothing methods (Durbin and Koopman, 2012) are utilized. However, WMA and regression models produced more promising results.

Duplicate entries: Duplication in data is removed by taking average of all values recorded for the same instance of time. This process is performed after removing outliers and missing values.

The aforementioned pre-processing methods are applied on load and weather data of LESCO. The graph in Figure 3, shows the five day profile of actual load data before processing and resultant data curve after processing. Actual load data contains missing values represented by '0' and negative values detected as outliers. In the processed curve, both missing and outliers are smoothed by applying filters. Weather data is processed in similar fashion to remove anomalies.

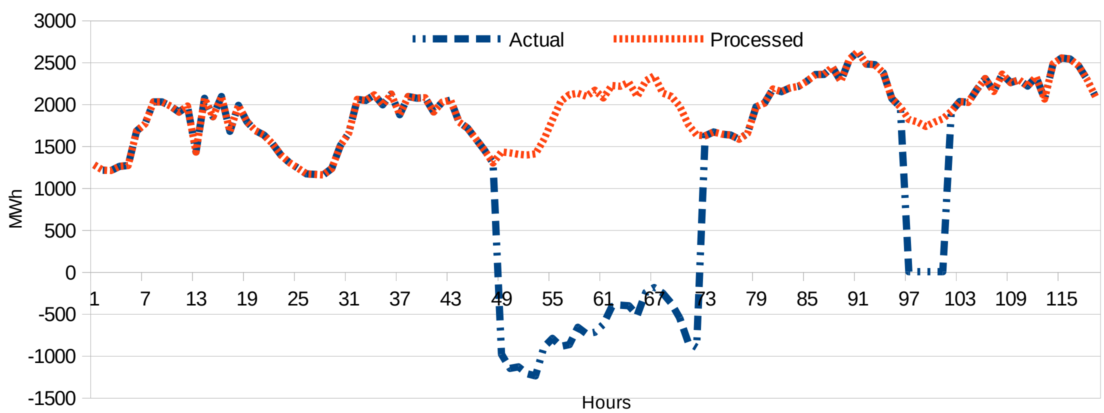

Fig 3: Five day profile of load data with noise and after processing.

Correlation matrix: Correlation is a measure of the extent to which corresponding elements from two sets of ordered data are linked together. It is measured by correlation coefficient, which ranges between -1 and +1. It is positive when the values increase together, and it is negative when one value decreases as the other increases. In the case of LF, it helps detecting the most influencing factors on electric load. Such factors are further selected for model development. The formula given in Eq. 2, is used to compute correlation coefficient *r* of two data series *X*, *Y*. In Table 1, correlation of different factors with demand curve is provided. It is evident that temperature, time of day, and previous load values have high correlation with current demand.

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}} \quad (2)$$

Where *n* is the number of pairs of data.

Table 1. Correlation matrix of input factors with load

Factors	Hour of day	Temperature	Dew	Humidity	Weekday	Previous Hour	Previous Day	Previous Week
Correlation with load	0.582	0.632	0.558	0.313	0.411	0.967	0.830	0.693

Model parameters: In this study, ANN and SVM based models both from statistical and AI domains are evaluated based on their accuracy of results. Four optimization techniques are used to get the optimized set of ANN connection weights and best parameters for SVM based regression model. These learning schemes include RPROP, LMA, SA, and PSO. The standard learning parameters for each technique are used in this experimentation, this include settings of each algorithm, maximum iteration count, and evaluation criteria.

ANN model structure: ANN model used in this experimentation consists of 3 layers of neurons connected to next layer, namely input layer, hidden layer and output layer. Input layer consists of 10 neurons, 7 neurons in hidden layer, and one output layer. Hence, total weight connections are 78.

SVM model: In machine learning, SVM's are supervised learning models that analyze and recognize patterns in data. In order to obtain better generalization performance through SVM, its learning parameters should be optimized. In this study these parameters are tuned by SA method.

RPROP: The parameters used for RPROP training are initial update value, maximum step size, and total iteration count. The initial update value for RPROP algorithm is 0.1; it is used initially for weight delta. Maximum step size is set to 50; it is the maximum value that can be achieved during the complete learning cycle of algorithm. Maximum iteration count for RPROP based model is 3000.

LMA: The LMA has two parameters named as lambda scale and lambda maximum value. The LMA method interpolates between Gauss-Newton Method (GNA) (Wang, 2012) and gradient method. The value of lambda determines which method it is. A lower value results in more usage of GNA and a higher value results in heavier usage of gradient descent. The value of lambda is scaled a teach iteration of algorithm. The initial value of lambda scale is 10.0, and maximum value is $1e^{25}$. The maximum iterations count for this algorithm is also 3000.

SA: The name and inspiration of SA comes from annealing in metallurgy, a method involving heating and controlled cooling of a material. The algorithm starts with a higher value of temperature T, and then it is decreased gradually at each step. In this experimentation start value of T is 10.0, and stop temperature is set to 0.1; which should always be T>=0. The maximum iteration count for this algorithm is 1000.

PSO: PSO is an idea based on social behavior of flock of birds. It models problem as a set of n particles each representing a dimension of solution space. These particles move in solution space in search of optimal solution. The particles follow three principles i.e. evaluating, comparing and imitating. The total population *P* corresponds to number of particles in the swarm, which in this case are 25. Inertia weight *W* controls the local v/s global exploration of search space and should be in range [0, 1]. Here *W* is 0.7. Two parameters, which control the learning rate, are *C1*, *C2*. They have the same value of 1.49. Position limits *P*, determines the limit of search space and velocity limit *V* determines the limit of change of position in search space. Here range of *P* and *V* is [-1.0, 1.0]. The maximum iteration count for PSO is 3000.

Performance evaluation: All of the ANN and SVM based models are ranked on the same criterion, i.e. Mean Absolute Percentage Error (MAPE). It is mostly widely method used to evaluate the forecasting models(Hyndman and Koehler, 2006). It represents the accuracy as percentage of error. MAPE is expressed by the formula given in Eq.3.

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad (3)$$

Where *n* is total number of instances, A_t is the actual load value at time *t*, and F_t is the forecasted value for the same time instance.

Experimentation on LESCO data: This section covers the simulations of aforementioned optimization techniques based ANN and SVM models on hourly and daily peak data. The purpose of this experimentation is to determine the forecasting capabilities of these techniques to predict hourly and daily consumption patterns for next few days.

Hourly demand estimation: In this section aforementioned modeling techniques are evaluated and compared against accuracy of results on hourly data sets. Sample data entries of LESCO are shown in Table 2. It contains hourly demand along other affecting parameters like weather data and calendar data.

Table 2. Data sample of input factors with hourly electric demand

Year	Month	Day	Hour	Temperature	Dew Point	Humidity	Day Type	Holiday	Weekday	Demand
2011	5	29	1	31	68	52	1	0	1	3343
2011	5	29	2	31	68	52	1	0	1	3298
2011	5	29	3	33	66.2	43	1	0	1	3228
2011	5	29	4	35	69.8	44	1	0	1	2945
2011	5	29	5	36	68	39	1	0	1	3024
2011	5	29	6	38	68	35	1	0	1	2854
2011	5	29	7	40	68	31	1	0	1	2844
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2011	5	30	24	28	68	61	0	0	2	3082

Five day demand profile is shown in Fig.4. Where, actual demand curve is plotted against the curves produced by each technique. Actual demand data is plotted for two days and next 3 day profile is estimated. In Fig.4-a, forecasted results obtained by RPROP-ANN model are shown. In Fig.4-b, results of ANN trained by PSO-Jordan are shown. Similarly results achieved by ANN-LMA are shown in Fig.4.c. Forecasted values of SVM trained by SA method are shown in Fig.4-d. To compare the results, estimated demand curves of all four techniques are plotted in a single graph given in Fig.4.e.

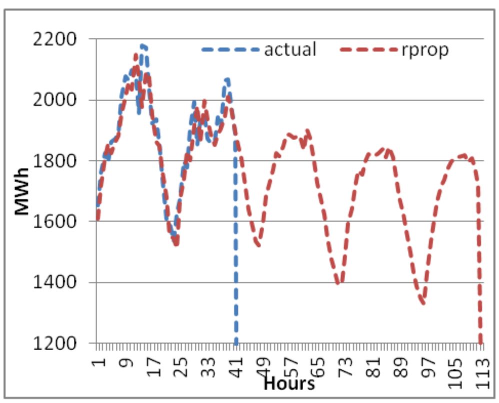

Fig.4-a: Forecasted demand by ANN-RPROP

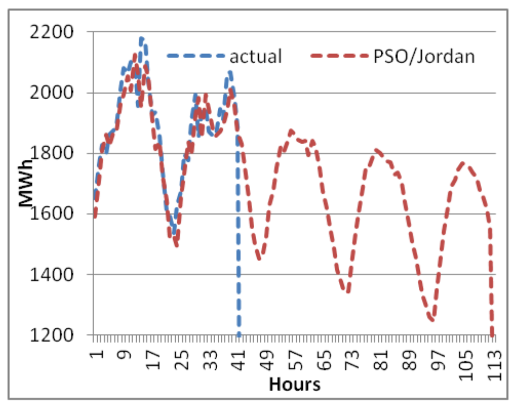

Fig.4-b: Forecasted demand by ANN-PSO

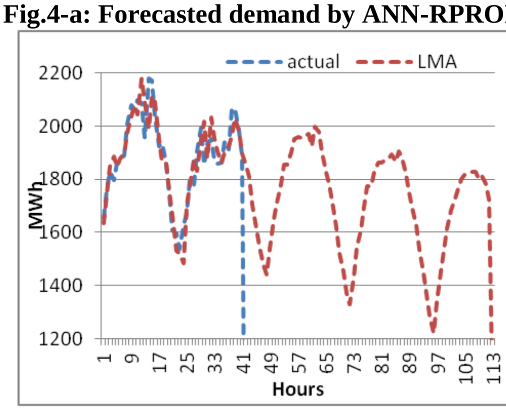

Fig.4-c: Forecasted demand by ANN-LMA

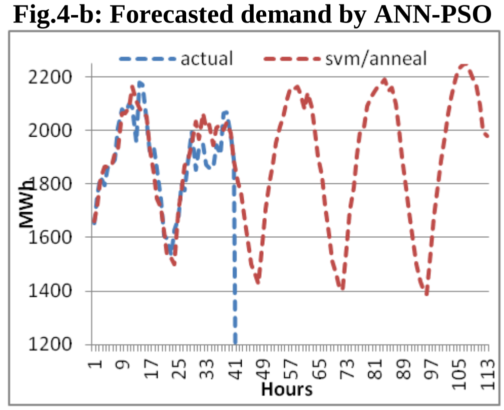

Fig.4-d: Forecasted demand by SVM-anneal

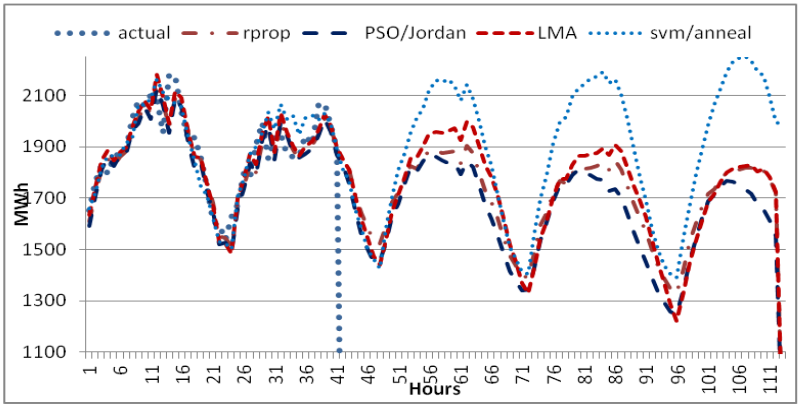

Fig.4-e: Comparison of actual load with forecasted values of four techniques

Daily peak demand estimation: The purpose of this experimentation is to determine the forecasting capabilities of aforementioned techniques to predict daily peak load consumption patterns for next few days. Most influencing factors have been considered that affect the peak load profile fluctuation. These include daily peak weather conditions like max temperature, and calendar inputs include day of week, week of the year, and day of month. Sample daily peak data of LESCO is shown in Table 3. It contains daily peak demand along other affecting parameters like weather and calendar inputs.

Table 3. Sample daily peak data with input factors

Year	Month	Day	Weekday	Week of year	Feels Like	Max Load
2011	4	1	6	14	35.24	2908
2011	4	2	7	14	32.73	2543
2011	4	3	1	15	34.34	2306
2011	4	4	2	15	35.22	2812
2011	4	5	3	15	32.70	2825
2011	4	6	4	15	30.06	2808
2011	4	7	5	15	33.90	2643
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2011	4	30	7	18	47.19	3453

Two months daily peak profile is presented in Fig.5. Where, actual demand curve is plotted against the curves produced by different techniques. Actual demand data is plotted with estimate so forty days and next twenty days profile is estimated by each technique. In Fig.5-a, results obtained by RPROP-ANN are shown. Whereas, in Fig.5-b, results of ANN trained by PSO-Jordan are shown. Similarly results achieved by ANN-LMA are shown in Fig.5.c. Forecasted values of SVM trained by anneal method are shown in Fig.5-d. Estimated demand curves of all the four techniques are compared in Fig.5.e.

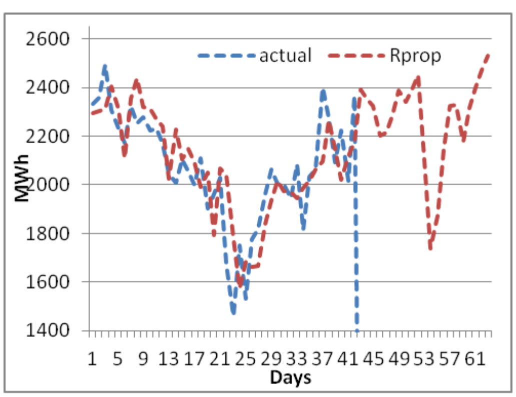

Fig.5-a: Forecasted demand by ANN-RPROP

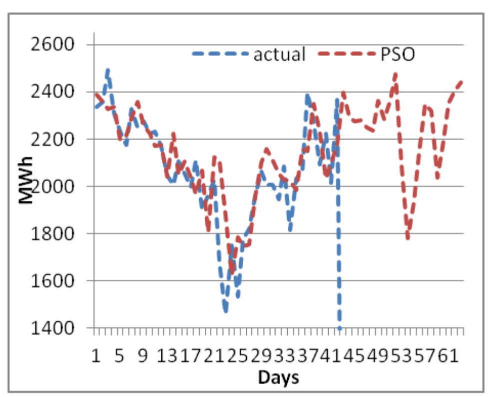

Fig.5-b: Forecasted demand by ANN-PSO

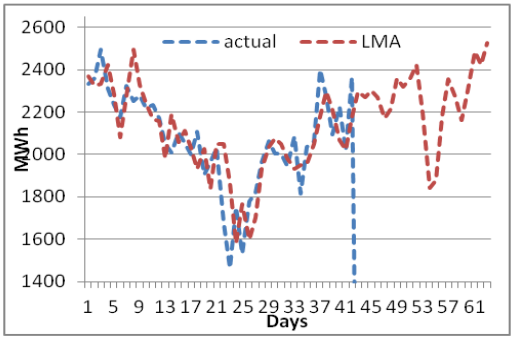

Fig.5-c: Forecasted demand by ANN-LMA

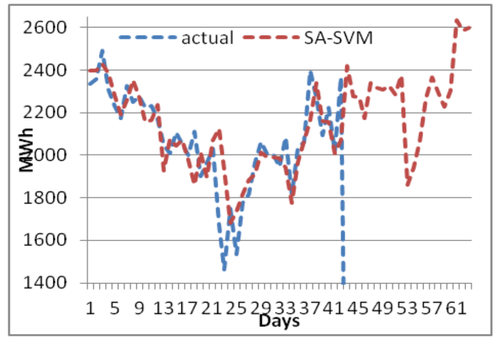

Fig.5-d: Forecasted demand by SVM-anneal

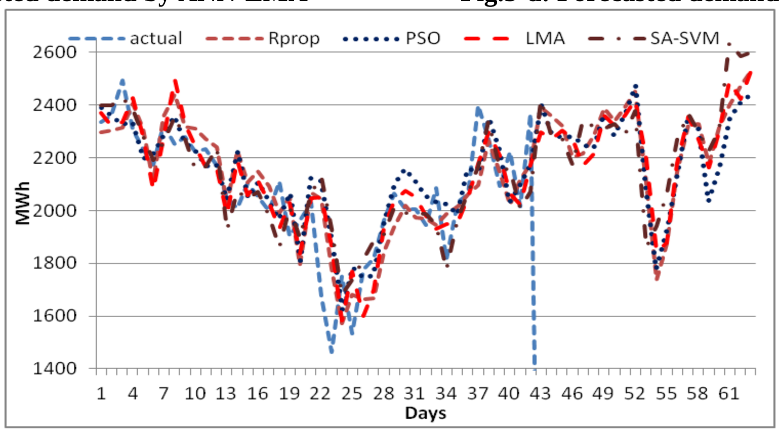

Fig.5-e: Comparison of actual load with forecasted values of four techniques

RESULTS AND DISCUSSION

This experimentation is performed on real electricity demand data collected from LESCO. Load affecting factors, such as weather conditions and calendar events are also considered for the analysis. This proposed methodology incorporates data pre-processing, analysis and LF modeling techniques of machine learning.

Accuracy comparison is presented in Table 4 and5, which is compared on the same criteria by taking MAPE. Training, testing, and validation accuracy is presented in Table 4 and 5. Forecasting accuracy on hourly data is presented in Table.4. It is evident that SA based SVM has outperformed other techniques in terms of accuracy of results. It produced lowest MAPE of training, testing and validation; reaching 97.50% accuracy of forecasts. This comparison is also presented in Fig.6.

Table 4. Accuracy of results obtained by different techniques on hourly load data

Technique	Train Accuracy	Test Accuracy	Validation Accuracy
RPROP	96.95 %	97.22 %	97.06 %
PSO	96.8 %	96.98 %	96.83 %
LMA	96.85 %	97.16 %	97.04 %
SA-SVM	97.36 %	97.65 %	97.57 %

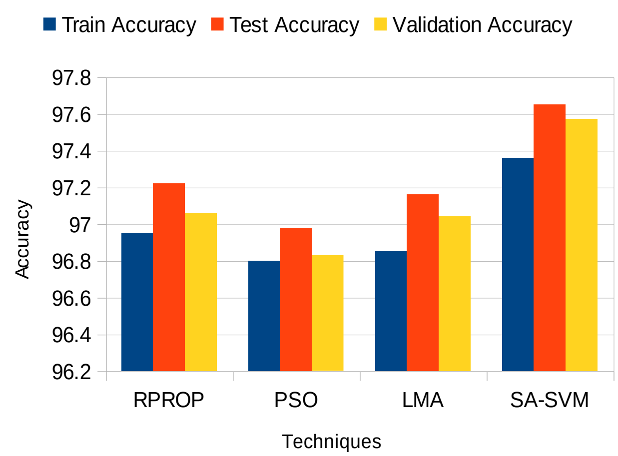

Fig. 6. Accuracy comparison between different techniques on hourly data

While the same technique has produced above 97% accurate results on daily peak demand data. It is evident from the statistics presented in Table.5, that SA based SVM has shown more generalization capability. It avoided over-fitting and under-fitting scenarios as there is comparatively little difference between train, test and validation accuracy. This comparison is also presented in Fig.7.

Table 5. Accuracy of results obtained by different techniques on daily peak data

Technique	Train Accuracy	Test Accuracy	Validation Accuracy
RPROP	96.61 %	97.78 %	97.48 %
PSO	96.37 %	97.85 %	97.55 %
LMA	96.71 %	97.63 %	97.37 %
SA-SVM	97.17 %	97.13 %	97.19 %

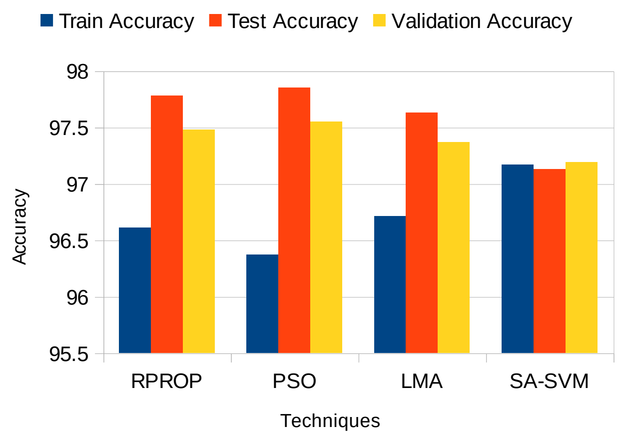

Fig.7. Accuracy comparison between different techniques on daily data

Concluding Remarks and Future Work: In this research work, we contribute by formulating a STLF model for the PSS by integrating machine learning techniques with improved accuracy. We propose an accurate and generic LF model formulation that is fit to indigenous attributes and can fulfill the requirements of our power industry. This formulation has proved our hypothesis to produce the accurate results of forecasts on noisy data sets with appropriate use of machine learning techniques for data pre-processing, analysis, and model development. This experimentation is performed on data collected from one utility. In future, we plan to assimilate other power supply companies. We aim to take into account medium term and long term forecasts on regional and national level.

Acknowledgements: This research is funded by National ICT R&D Fund, Pakistan. Authors are also grateful to National Transmission and Dispatch Company (NTDC) and LESCO, for the co-operation and providing data sets.

REFERENCES

Aguiar e Oliveira Junior, H., Ingber, L., Petraglia, A., Rembold Petraglia, M., Augusta Soares Machado, M. Adaptive Simulated Annealing, in: Stochastic Global Optimization and Its Applications with Fuzzy Adaptive Simulated Annealing, Intelligent Systems Reference Library. pp. 33–62 (2012).

Alfares, H.K., Nazeeruddin, M. Electric load forecasting: Literature survey and classification of methods. Intl. J. Systems Science 33:23–34 (2002).

Amjady, N. Short-term hourly load forecasting using time-series modeling with peak load estimation capability. Power Systems, IEEE Transactions on 16:498–505 (2001).

Amjady, N., Keynia, F. A New Neural Network Approach to Short Term Load Forecasting of Electrical Power Systems. Energies 4:488–503 (2011).

Awan, S.M., Khan, Z.A., Aslam, M., Mahmood, W., Ahsan, A. Application of NARX based FFNN, SVR and ANN Fitting models for long term industrial load forecasting and their comparison, in: Industrial Electronics (ISIE), 2012 IEEE International Symposium On. pp. 803–807 (2012).

Charytoniuk, W., Chen, M.S., Van Olinda, P. Nonparametric regression based short-term load forecasting. IEEE Transactions on Power Systems 13:725–730 (1998).

Che, J., Wang, J., Wang, G. An adaptive fuzzy combination model based on self-organizing map and support vector regression for electric load forecasting. Energy 37:657–664 (2012).

Durbin, J., Koopman, S.J. Time series analysis by state space methods. Oxford University Press (2012).

Elattar, E.E., Goulermas, J., Wu, Q.H. Electric Load Forecasting Based on Locally Weighted Support Vector Regression. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 40:438–447 (2010).

Hahn, H., Meyer-Nieberg, S., Pickl, S. Electric load forecasting methods: Tools for decision making. European J. Operational Research 199: 902 – 907 (2009).

Haykin, S. Neural Networks: A Comprehensive Foundation, Prentice Hall International Editions Series. Prentice Hall (1999).

Hodge, V., Austin, J. A Survey of Outlier Detection Methodologies. Artificial Intelligence Review 22:85–126 (2004).

Huang Yue, Li Dan, Gao Liqun, Wang Hongyuan. A short-term load forecasting approach based on support vector machine with adaptive particle swarm optimization algorithm. IEEE, pp. 1448–1453 (2009).

Hyndman, R.J., Koehler, A.B. Another look at measures of forecast accuracy. Intl. J. Forecasting 22:679 – 688 (2006).

Jain, A., Nandakumar, K., Ross, A. Score normalization in multimodal biometric systems. Pattern Recognition 38:2270 – 2285 (2005).

Jakkula, V. Tutorial on Support Vector Machine

(SVM)(2006).

Kennedy, J. Particle Swarm Optimization, in: Sammut, C., Webb, G. (Eds.), Encyclopedia of Machine Learning. Springer US, pp. 760–766 (2010).

Liao, S.-H. Expert system methodologies and applications—a decade review from 1995 to 2004. Expert Systems with Applications 28:93 – 103 (2005).

Pai, P.-F., Hong, W.-C. Support vector machines with simulated annealing algorithms in electricity load forecasting. Energy Conversion and Management 46: 2669 – 2688 (2005).

Saini, L.M. Peak load forecasting using Bayesian regularization, Resilient and adaptive back propagation learning based artificial neural networks. Electric Power Systems Research 78: 1302 – 1310 (2008).

Saini, L.M., Soni, M.K. Artificial neural network based peak load forecasting using Levenberg-Marquardt and quasi-Newton methods. Generation, Transmission and Distribution, IEE Proceedings- 149: 578–584 (2002).

Soares, L.J., Medeiros, M.C. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. Intl. J. Forecasting 24: 630 – 644 (2008).

Sra, S., Nowozin, S., Wright, S.J. Optimization for Machine Learning. Mit Press (2011).

Suganthi, L., Samuel, A.A. Energy models for demand forecasting—A review. Renewable and Sustainable Energy Reviews 16: 1223 – 1240 (2012).

Taylor, J.W., McSharry, P.E. Short-Term Load Forecasting Methods: An Evaluation Based on European Data. IEEE Transactions on Power Systems 22: 2213–2219 (2007).

Vapnik, V.N. The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA (1995).

Wang, Y. Gauss–Newton method. Wiley Interdisciplinary Reviews: Computational Statistics 4: 415–420 (2012).

Wilamowski, B.M., Yu, H. Improved computation for Levenberg–Marquardt training. Neural Networks, IEEE Transactions on 21:930–937 (2010).

Xia, C., Wang, J., McMenemy, K. Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. International Journal of Electrical Power & Energy Systems 32:743 – 750 (2010).