Learning Energy Demand Domain Knowledge via Feature Transformation

Sanzad Siddique
Department of EECE
Marquette University
Milwaukee, Wisconsin USA
sanzad.siddique@marquette.edu

Richard J. Povinelli
Department of EECE
Marquette University
Milwaukee, Wisconsin USA
richard.povinelli@marquette.edu

***Abstract* — Domain knowledge is an essential factor for forecasting energy demand. This paper introduces a method that incorporates machine learning techniques to learn domain knowledge by transforming the input features. Our approach divides the inputs into subsets and then searches for the best machine learning technique for transforming each subset of inputs. Preprocessing of the inputs is not required in our approach because the machine learning techniques appropriately transform the inputs. Hence, this technique is capable of learning where nonlinear transformations of the inputs are needed. We show that the learned data transformations correspond to energy forecasting domain knowledge. Transformed subsets of the inputs are combined using ensemble regression, and the final forecasted value is obtained. Our approach is tested with natural gas and electricity demand signals. Experimental results show how this method can learn domain knowledge, which yields improved forecasts.**

***Index Terms*--Machine learning, Demand forecasting, Time series analysis, Feature transformation, Domain knowledge.**

I. Introduction

Building forecasting models for energy demand is an active research field in engineering, statistics and econometrics [1-3]. Prior domain knowledge substantially improves energy demand forecasting accuracy. However, domain knowledge may differ between the types of energy to be forecasted. When forecasting in new energy domains, it is likely that there is insufficient domain knowledge to build an accurate forecasting model. This paper proposes an algorithm that can extract domain knowledge from the energy demand signals.

Nonlinear modeling techniques are typically needed to representation domain knowledge [4]. We use machine learning approaches that are capable of modeling nonlinearities. Such machine learning methods have been used for feature extraction [1, 5], input preprocessing [1, 6] and knowledge extraction [7, 8]. The work of Valenzuela [4] is an example of integrating an automatic model discovery algorithm with domain knowledge learning. In this technique, the domain knowledge is learned using a rule extraction technique. In contrast, we propose a two stage forecasting algorithm. The first stage automatically captures energy demand forecasting domain knowledge through nonlinear transformation of the input features. In the second stage, the transformed features are combined via hybridization [2-4] and ensembling [2]. We compare our new algorithm against models with and without domain knowledge.

The research work presented in this paper is motivated by the scenario where sufficient domain knowledge is not present. For example, an experienced gas demand forecaster is planning to start forecasting electricity demand. The forecaster has sufficient historical electricity data available but domain knowledge is limited in the relevant area. The feature extraction method can be used to extract the domain knowledge from the available historical electricity data and a reasonable forecasting can be made based on the extracted feature. Further research in this technique can result in a further improvement in forecasting accuracy when used in conjunction with the available domain knowledge.

It is common to use multiple nonlinear modeling techniques and ensemble the results to forecast the energy demand [2-4]. The paper also uses multiple nonlinear techniques. The contribution of this paper is, it demonstrate feature extraction and the domain knowledge representation ability of nonlinear technique and obtain a better accuracy using the ensemble. Each of the nonlinear techniques are applied with a single input that enables feature extraction and domain knowledge representation.

II. Technical Discussion

The proposed feature transformation algorithm incorporates autoregressive and moving average (ARMA), regression trees (RT), and artificial neural networks (ANN). A brief overview of these methods is presented in this section.

A. Autoregressive and Moving Average (ARMA)

An autoregressive and moving average (ARMA) model combines both autoregressive and moving average terms. It is one of the most commonly used techniques for forecasting time series and was popularized by Box and Jenkins in their time series analysis book [9]. An ARMA model is represented by

We acknowledge Marquette University EECE department for financial support and GasDay for the data and the domain knowledge.

978-1-4799-6415-4/14/$31.00 ©2014 IEEE

$$Y_t = c + \sum_{i=1}^{p} \varphi_i Y_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}, \quad (1)$$

where Y_t indicates the time series value at time *t*. Similarly, Y_{t-i} indicates the value recorded at time *t-i*. The φ's represent the AR coefficient, where φ_i is the coefficient for Y_{t-i}. The θ values are the coefficients of the moving average terms. Additional terms include the constant value *c* and a time-dependent random variable ε_t. The variable *p* is the AR order and *q* is the MA order.

B. Regression Tree(RT)

A regression tree is a special form of a binary decision tree used for building nonlinear regression models. A binary decision tree is a machine learning technique used for the classification, and a regression tree is used for regression. Like a binary decision tree, the decision nodes in a regression tree represent a decision based on the value of a given attribute. The leaves of the tree are learned using the forecasted values. There are fast and reliable algorithms available to learn the nodes and leaves [10]. Regression trees have been used for forecasting [10, 11]. An advantage of using a regression tree is that it can be learned quickly.

C. Artificial Neural Network(ANN)

An ANN consists of fully/partially connected neurons and can often implement effective nonlinear model. The weights of the connection between the neurons can be learned using suitable training algorithm. ANNs are widely used for energy demand forecasting [12-14].

D. Ensemble regression

Ensemble regression [15, 16] uses the outputs from all of the component models in determining the final output. Ensemble regression nonlinearly transforms the component model outputs and learns weights for each of the transformed outputs. If component model outputs were not transformed, ensemble regression would be equivalent to linear regression, where the component model outputs are independent variables, and the weights are regression parameters. Ensemble regression combines the outputs from the different modeling techniques.

III. Proposed Method

Our approach includes two training stages.

A. Stage I training

In the first stage, a set of statistical and machine learning techniques are used for modeling the inputs. A set of candidate models are examined against different subsets of the inputs. Using a search approach, the most appropriate model among the candidate models is chosen for each individual input subset. In our case, the candidate modes were ARMA, RT and ANN. The inputs were previous energy demand, weather inputs, and seasonality. Using our search based approach ARMA, RT, and ANN, were chosen, respectively, to transform previous energy demand, weather inputs, and seasonal information. The feature transformation method is described below.

$$m_j = \underset{m \in \{\text{ARMA, RT, ANN}\}}{\arg\min} \text{MAPE}\left(m\left(\mathbf{X}_j\right)\right), \quad (2)$$

where $\mathbf{X}_j$ is input subset *j*, *m* is a model, MAPE is the mean absolute percentage error of the forecasts of *m* of the training energy demand, and m_j is the model chosen to transform $\mathbf{X}_j$.

A set of candidate models such as ARMA, RT, ANN are used. The input subset $\mathbf{X}_j$ is tested with the candidate models. For example the previous energy demand data is tested with ARMA, RT, and ANN. The candidate model that gives the minimum MAPE is chosen as the model for that input. In this example, the ARMA, ANN, and the RT are selected for the previous energy demand data, the seasonal information, and the temperature input, respectively.

B. Stage II training.

The energy demand forecast outputs generated from each of those selected models are used as the inputs for the second stage of the training: ensemble regression. Our chosen ensemble regression technique uses a generalized linear model with quadratic fitting, and with identity as the link function with a normal random distribution. The ensemble regression is represented as follows.

$$\hat{Y}_t = c + \sum_{j=1}^{M} \gamma_j f\left(m_j\left(X_{j,t}\right)\right), \quad (3)$$

where $\hat{Y}_t$ is the estimated energy demand, *c* is the regression intercept, *M* is the total number of models, γ_i is the regression parameter for model m_j, *f* is a quadratic function and $X_{j,t}$ is the input subset at time *t*.

Once the individual models are trained with the training data, the forecasted output from these models are obtained by using the training data. These outputs are used as inputs for the ensemble regression model. The ensemble regression learns a quadratic transformation and weights for each of the input subsets.

C. Testing

Forecasting using the feature transformation technique is straightforward. Testing data is input into the first stage feature transformation models. The output of the feature transformation models is fed into the ensemble regression, which generates the final forecast.

IV. Data

Our proposed approach is tested with the natural gas and electricity datasets. The gas dataset consists of daily natural gas usage and temperature for 4800 days from a specific location in the United States. The electricity dataset consists of daily load and temperature for 2800 days from another specific location in the United States. The data is normalized due to confidentiality. The normalized natural gas and electricity datasets are shown in Figure 1 and Figure 2, respectively.

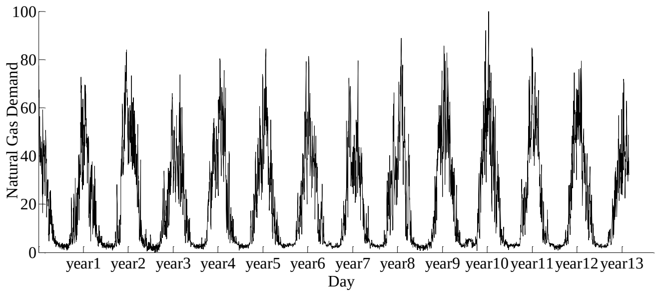

Figure 1: Natural gas demand

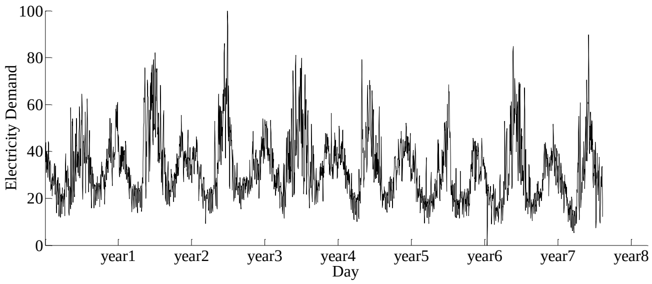

Figure 2: Electricity demand

The first 80% of the data is used for training; the remaining 20% is used for testing. The inputs are 16 autoregressive terms with temperature and day of the year as exogenous inputs.

V. Analysis of Transformed Features

The nonlinear transformation of the input subsets reveals domain knowledge applicable to the input. As an example, *a prior* domain knowledge suggests a nonlinear transformation to the temperature known as heating degree day (HDD) [17] as

$$HDD_t = \max\left(0, T_{ref} - T_t\right), \tag{4}$$

where T_t is the temperature at day t and T_{ref} is the reference temperature. This transformation is seen in Figure 3.

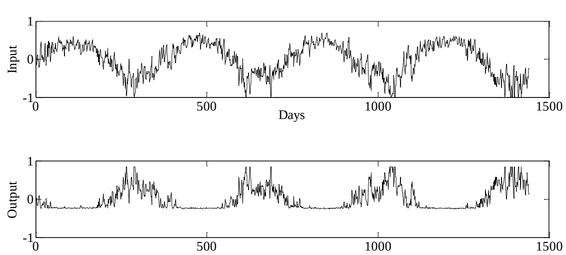

Figure 3: Transformation of temperature for natural gas

Significant improvement in forecasting accuracy can be achieved by applying this domain knowledge [17]. In Figure 3, the RT generates a similar nonlinear transformation of the temperature input. Figure 3 shows the temperature input and the output of the RT model normalized between -1 to 1. This RT model is trained for the gas dataset. A nonlinear transformation is made to the temperature by the feature transformation method, which is similar to having domain knowledge that suggests a similar nonlinear transformation in the temperature represented by (4).

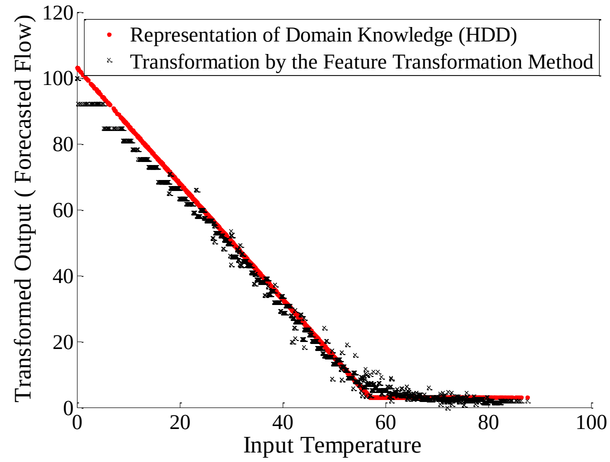

Figure 4: Transformation of temperature for natural gas

If we plot the actual temperature against the RT model output (i.e., the preprocessed and nonlinearly transformed input temperature) in Figure 4, we observe the similar representation of domain knowledge as presented by Equation (4). The Figure-4 presents the transformed output with an expression *1.9 ** max(*0*, *55*-T_t). This indicates a strong correlation between the transformed feature and actual domain knowledge presented by Equation (4).Thus, the feature transformation method is capable of learning domain knowledge.

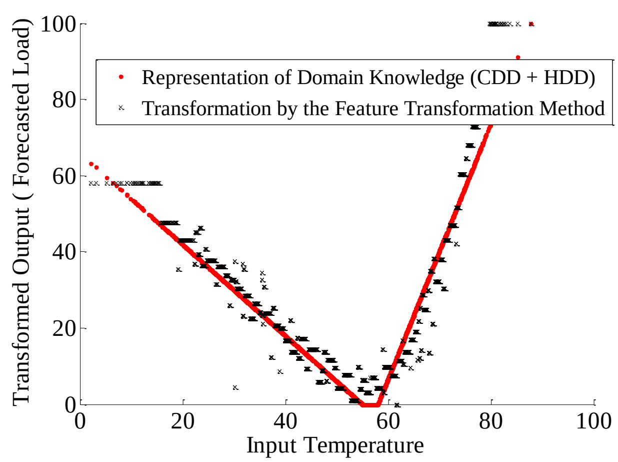

Figure 5: Transformation of temperature for electricity

When tested with the electricity dataset, the transformation of the temperature shows different behavior, presented in Figure 5. The behavior is also consistent with domain knowledge of electricity. Unlike natural gas, where the demand becomes nearly constant after a certain threshold of temperature, the electricity demand increases after the threshold temperature, as suggested by Figure 5. The phenomenon is usually modeled by another nonlinear transformation in the temperature known as cooling degree day CDD) [17] as shown in Equation (5)

$$CDD_t = \max\left(0, T_t - T_{ref}\right), \tag{5}$$

where T_t is the temperature at day t and T_{ref} is the reference temperature.

The transformed output in Figure-5 can be represented by the expression $1.1*\max(0, 55-T_t) + 3.6*\max(0, T_t-58)$, which is a linear combination of Equation (4) and Equation (5), where the constant are chosen to illustrate the correlation with the extracted feature. The feature transformation method automatically performs the *HDD* transformation for the natural gas dataset and both the *HDD* and *CDD* transformation for the electricity dataset without having prior knowledge of the type of energy demand forecasting.

VI. RESULTS AND ERROR ANALYSIS

A set of candidate models (ARMA, RT and ANN) are used. For each set of inputs (AR terms, day of the year, and temperature) a model is selected according to (2). The selected models are presented in TABLE I. Each of the input subsets is modeled using the automatically selected modeling method. The outputs from each of these models are combined using the ensemble regression method. Our approach is tested with the real gas and electricity datasets. Errors are calculated using mean average percentage error (MAPE) using the below formula

$$MAPE = \frac{100\%}{n} \sum_{t=1}^{n} \left(\left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \right), \tag{6}$$

where Y_t is the actual output, $\hat{Y}_t$ is the forecasted output and n is the total number of data points.

To compare the performance, two sets of benchmarks are created using the linear regression model. One benchmark technique uses domain knowledge, and the other does not. The results from the feature transformation method proposed by this paper are compared with the results from both of the benchmarks.

TABLE I. MODEL SELECTION FOR INPUT SUBSETS

Variable	Technique	Attributes
AR terms	ARMA	16 AR order
Temperature	RT	Min parent 53, Min leaf 2
Day of year	RT	Min parent 53, Min leaf 2
Day of week	ANN	1 Hidden layer

Table II shows the test result for the gas dataset by using above input-model sets. TABLE III represents the test result for the electricity dataset also by using the same input model sets. For the electricity datasets, the feature transformation method exceeds the forecasting accuracy of both the benchmarks. For the natural gas dataset, the feature transformation method is far more accurate than the benchmark of no domain knowledge. The accuracy of the feature transformation method is close to the accuracy of benchmark of having domain knowledge. For both the electricity and the gas datasets the feature extraction technique provides better accuracy than the individual machine learning technique such as ANN and RT. This indicates that the selection of different nonlinear feature extraction techniques for different input variables and ensemble the individual model's output yields more accurate result than using a single machine learning technique with all inputs.

This paper demonstrates the feature transformation of temperature input using Figure 4 and 5. The other two inputs, day of year and day of week, are also transformed nonlinearly by the feature transformation technique in a similar fashion and contribute towards the accuracy.

TABLE II. TESTING MAPE FOR NATURAL GAS DATASET

Model	MAPE
Benchmark without domain knowledge	27.13
Benchmark with domain knowledge	7.58
Using only ANN for all inputs	8.13
Using only RT for all inputs	8.11
With feature transformation	7.83

TABLE III. TESTING MAPE FOR ELECTRICITY DATASET

Model	MAPE
Benchmark without domain knowledge	6.51
Benchmark with domain knowledge	4.12
Using only ANN for all inputs	4.37
Using only RT for all inputs	4.25
With feature transformation	3.91

VII. CONCLUSION

The test results show that our feature transformation method is capable of effective automatic input preprocessing. Our technique also is able to learn and represent domain knowledge learned from the data. The feature transformation method even shows better accuracy than the linear model using prior domain knowledge.

One problem of using the feature transformation method is that the output from the regression tree is discrete instead of continuous. Even though other continuous methods are tested with the temperature, a regression tree provided the best result. Further research with other continuous models is needed. Another potential problem with the regression tree is that it may not perform well for the unusual cold or hot days. Test results show that, using the regression tree, the output does not change significantly for exceptionally cold or hot temperatures. This is also an opportunity for further research.

The machine learning techniques introduced by this paper use the historical energy demand data, and the current temperature, and day of the year as inputs. It is also important to include historical temperatures and also other relevant inputs, such as other weather variables and economic variables. Each of these new set of inputs can be modeled using different feature transformation techniques. Also, it can be useful to build mechanisms for learning more complex domain knowledge using the additional sets of inputs.

REFERENCES

[1] M. M. Othman, M. H. H. Harun and I. Musirin, "Forecasting short term electric load based on stationary output of artificial neural network

considering sequential process of feature extraction methods," in *IEEE International Power Engineering and Optimization Conference (PEDCO) Melaka, Malaysia, 2012,* 2012, pp. 485-489.

[2] W. Shuai, T. Ling and Y. Lean, "SD-LSSVR-based decomposition-and-ensemble methodology with application to hydropower consumption forecasting," in *Fourth International Joint Conference on Computational Sciences and Optimization (CSO), 2011,* 2011, pp. 603-607.

[3] S. Asadi, A. Tavakoli and S. Reza Hejazi, "A new hybrid for improvement of auto-regressive integrated moving average models applying particle swarm optimization," *Expert Syst,* vol. 39, pp. 5332-5337, April 2012, 2012.

[4] O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L. J. Herrera, A. Guillen, L. Marquez and M. Pasadas, "Hybridization of intelligent techniques and ARIMA models for time series prediction," *Fuzzy Sets Syst.,* vol. 159, pp. 821-845, 4/1, 2008.

[5] W. Wettayaprasit and P. Nanakorn, "Feature extraction and interval filtering technique for time-series forecasting using neural networks," in *IEEE Conference on Cybernetics and Intelligent Systems, 2006,* 2006, pp. 1-6.

[6] M. Simunek and E. Pelikán, "Temperatures data preprocessing for short-term gas consumption forecast," in *IEEE International Symposium on Industrial Electronics, 2008. ISIE 2008.* 2008, pp. 1192-1196.

[7] K. J. Siddiqui and S. M. Nugen, "Knowledge based system for weather information processing and forecasting," in *Geoscience and Remote Sensing Symposium, 1996. IGARSS '96. 'Remote Sensing for a Sustainable Future.', International,* 1996, pp. 1099-1101 vol.2.

[8] Y. Alsultanny, "Successful forecasting for knowledge discovery by statistical methods," in *Conference on Information Technology: New Generations (ITNG), 2012 Ninth International,* 2012, pp. 584-588.

[9] G. E. P. Box, G. M. Jenkins and G. C. Reinsel, *Time Series Analysis: Forecasting and Control.* Hoboken, N.J.: John Wiley, c2008; 4th ed, 2008.

[10] Ying-Chun Guo, "Knowledge-enabled short-term load forecasting based on pattern-base using classification & regression tree and support vector regression," in *Fifth International Conference on Natural Computation, 2009. ICNC '09,* 2009, pp. 425-429.

[11] H. Mori, N. Kosemura, K. Ishiguro and T. Kondo, "Short-term load forecasting with fuzzy regression tree in power systems," in *IEEE International Conference on Systems, Man, and Cybernetics, 2001,* 2001, pp. 1948-1953 vol.3.

[12] A. Azadeh, S. F. Ghaderi, S. Tarverdian and M. Saberi, "Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption," *Applied Mathematics and Computation,* vol. 186, pp. 1731-1741, 3/15, 2007.

[13] G. Oğcu, O. F. Demirel and S. Zaim, "Forecasting Electricity Consumption with Neural Networks and Support Vector Regression," *Procedia - Social and Behavioral Sciences,* vol. 58, pp. 1576-1585, 10/12, 2012.

[14] E. Erdogdu, "Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey," *Energy Policy,* vol. 35, pp. 1129-1146, 2, 2007.

[15] E. Frank, L. Trigg, G. Holmes and I. H. Witten, "Naive Bayes for Regression," *Machine Learning,* pp. 1-20, 1998.

[16] A. C. de Pina and G. Zaverucha, "Combining attributes to improve the performance of naive bayes for regression," in *IEEE International Joint Conference on Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence),* 2008, pp. 3210-3215.

[17] C. Yan and J. Yao, "Application of ANN for the prediction of building energy consumption at different climate zones with HDD and CDD," in *2nd International Conference on Future Computer and Communication (ICFCC), 2010,* 2010, pp. V3-286-V3-289.