LONG TERM ENERGY CONSUMPTION FORECASTING USING GENETIC PROGRAMMING

Korhan Karabulut[a], Ahmet Alkan[b], Ahmet S. Yilmaz[b]

[a] Department of Computer Engineering, Yasar University, 35500 Izmir, Turkey
korhan.karabulut@yasar.edu.tr,

[b] Department of Electrical and Electronics Engineering,
Kahramanmaras Sutcu Imam University, Kahramanmaras, Turkey
{aalkan@ksu.edu.tr,asyilmaz}@ksu.edu.tr

Abstract- Managing electrical energy supply is a complex task. The most important part of electric utility resource planning is forecasting of the future load demand in the regional or national service area. This is usually achieved by constructing models on relative information, such as climate and previous load demand data. In this paper, a genetic programming approach is proposed to forecast long term electrical power consumption in the area covered by a utility situated in the southeast of Turkey. The empirical results demonstrate successful load forecast with a low error rate.

Key Words- genetic programming, load forecasting, symbolic regression

1. INTRODUCTION

Load management is a capability required by load distribution centers and electric utilities. Providing the balance between supply and demand in energy market and supplying the customers with more efficiency are available via satisfactory and reliable load management. Load forecasting is a requirement for successful load management. An accomplished load forecasting makes possible the trustiest planning for future. Especially, long term load forecasting is guidance for maintenance of electricity installations and construction planning. Therefore, power system engineers and electricity generation/distribution utilities attach importance to load forecasting. Hence, new forecasting methods have been more attractive by power utilities for the reliable and actual estimations.

Load forecasting analyses can be classified as short term, mid term and long term in general. Hourly and daily (24 hours) forecasts can be classified as short term load forecasting. Mid and long term forecasting cover the weekly, monthly, seasonal and annual forecasts.

Short term load variations are quite non-linear due to dependence on knowledge measured in short times from several minutes to an hour. Mid term or intermediate term forecasts cover the duration from a few days to several months. In long term forecasting, forecast time varies between 1 and 10 years. Load profile can be obtained using short term estimations. However, power system planners require the mid and long term load forecasts to make decisions about planning the mid and long term maintenance, preparing the future investment schedules and developing the generation, transmission and distribution systems. Planning of future investment for the constructions depends on the accuracy of the long term load forecasting considerably [1]. Therefore, several estimation methods have been applied for short, mid and long

term load forecasting. Conventional load forecasting techniques are based on statistical methods. Stochastic time series [2, 3], autoregressive models [4], non-parametric regression models [5, 6] were used in load forecasting. Also several soft computing techniques were used as load estimator, such as fuzzy regression [7, 8], self organizing map (SOM) [9], non-fixed neural network [10], dynamic neural networks [11], a combination of regression analysis and a fuzzy inference system [12] and fuzzy regression tree and multi-layer perceptron of ANN model [13].

In this paper, we present a genetic programming approach on the forecasting of long term electrical power consumption of a moderate city in Turkey. We use the genetic programming method to forecast future usage through symbolic regression using annual data of the previous years.

The rest of the paper is structured as follows: section 2 summarizes methods of forecasting an unknown function, section 3 gives the details of the application of genetic programming to load forecasting and section 4 summarizes the results and a conclusion is given.

2. METHODS OF FORECASTING AN UNKNOWN FUNCTION

Approximating an unknown function with sample data is an important practical problem. In order to forecast an unknown function using a finite set of sample data, a function is constructed to fit sample data points. This process is called curve fitting. There are several methods of curve fitting. Interpolation is a special case of curve fitting where an exact fit of the existing data points is expected. In other forms of curve fitting, an approximate fit can be permitted. The term "regression" is used to include many different methods of curve fitting.

Once a model is generated, acceptability of the model must be tested. There are several measures to test the goodness of a model. Sum of absolute difference, mean absolute error, mean absolute percentage error, sum of squares due to error (SSE), mean squared error and root mean squared errors can be used to evaluate models. Minimizing the squares of vertical distance of the points in a curve (SSE) is one of the most widely used methods.

2.1. Genetic Programming and Symbolic Regression

In conventional regression, one has to decide on the approximation function (can be an n-degree polynomial, non-polynomial, or a combination of both) and try to find the coefficients of this selected function. Constructing an approximation function can be a difficult task. There is another form of regression called "symbolic regression". In the symbolic regression problem, the aim is to search a symbolic representation of a model, instead of only searching for coefficients of a predefined model. Genetic programming (GP) method introduced by Koza [14] can be used for the symbolic regression problem. GP searches for the model and coefficients of the model at the same time.

GP simultaneously works on a group of possible solutions instead of a single solution. This group of candidate solutions is called a "population". GP as a whole tries to find the best solution. Since the original function is unknown, GP searches for an approximate hence, an "acceptable" solution. GP is an extension to genetic algorithms where individuals are programs or functions.

In GP, individuals are represented as trees. Elements of the trees are functions and terminals. Terminals are the variables and the functions are operations applied to these variables forming the model together. For example, Fig.1 shows the representation of the simple expression, x*y+x/2.
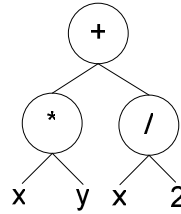

Fig. 1. An example tree representation in genetic programming

In applying genetic programming to a problem, there are five major preparatory steps. These five steps involve determining;
(1) Predefined set of terminals,
(2) Predefined set of primitive functions,
(3) A fitness measure (called a "fitness function" to specify what needs to be done)
(4) The parameters for controlling the run, such as population size, reproduction operators, probabilities of the operator and so on, and
(5) The method for designating a result and the criterion for terminating a run [15].
Genetic programming is an iterative method. The algorithm used by genetic programming in finding solutions is summarized in Fig. 2.
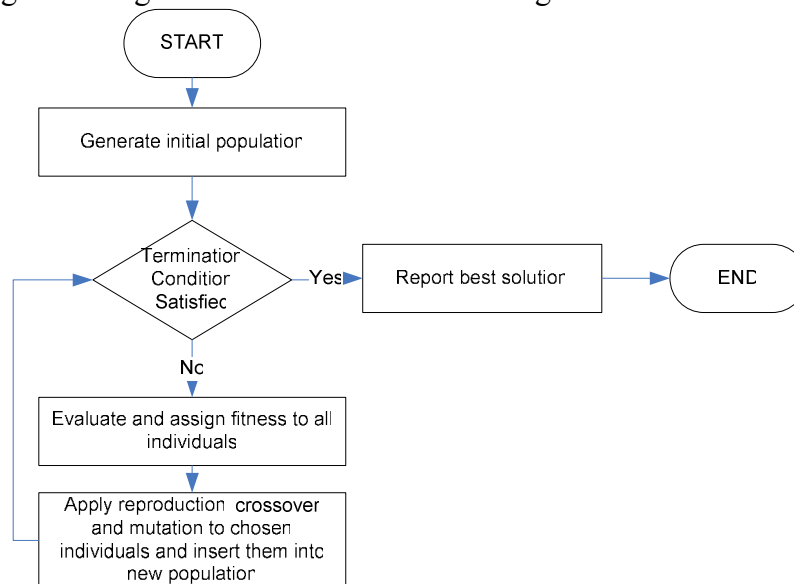

Fig. 2. Simplified form of the algorithm used by genetic programming

The first step in the genetic programming algorithm is the generation of an initial population either by using random compositions of the functions and terminals or by using a predefined strategy.

In the next step, the termination condition is checked. If the termination condition is reached, the process is ended and best so far result is reported. If the termination condition is not reached, the following steps are repeated:

(1) Each individual in the population is evaluated and assigned a fitness value using the fitness function.

(2) A new population is created by applying the following operations. The operations are applied to individuals chosen from the population with a probability based on fitness:

(i) Darwinian Reproduction: Reproduce an existing individual by simply copying it into the new population.

(ii) Crossover: Create two new individuals by genetically recombining randomly chosen parts of two existing individuals using the crossover operation.

(iii) Mutation: Create one new individual from one existing individual by mutating a randomly chosen part of the individual.

(3) The individual that is identified by the method of result designation is reported as the result for the run (e.g., the best-so-far individual). This result may be a solution (or an approximate solution) to the problem [16].

Crossover operation is one of the two primary reproduction operators. In the crossover operation, two solutions are combined to form two new solutions or offspring. A random point is chosen in both of the parents, and the nodes below the crossover points are exchanged between parents. Fig 3 shows an example of crossover operation.
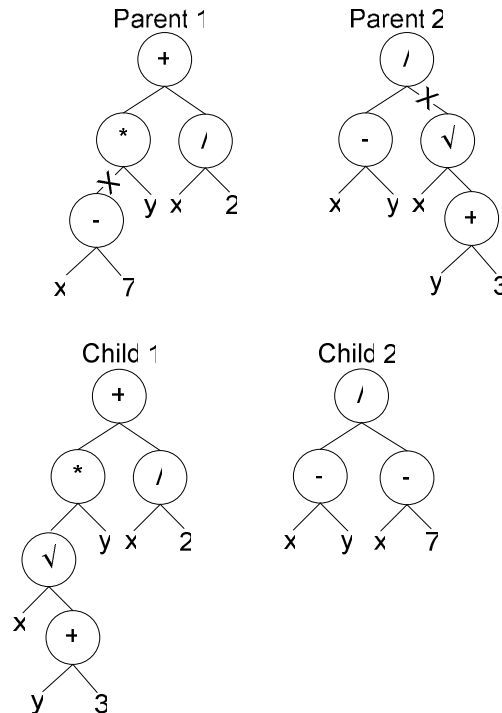

Fig. 3. An example of crossover operation (X denotes the crossover point)

Mutation operation is another important feature of genetic programming. Two kinds of mutation operation are possible. In the first kind, either a randomly chosen

function or a terminal can be replaced with a new function or a terminal. In the second kind, a randomly chosen subtree can be replaced by another subtree. Fig 4 shows an example of each kind of mutation.
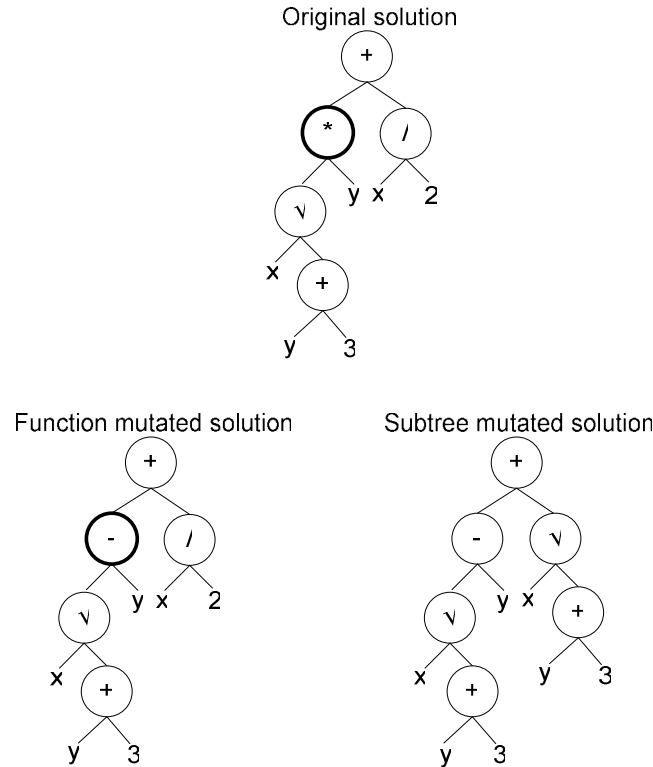

Fig. 4. An example of mutation operation

3. APPLICATION OF GENETIC PROGRAMMING TO LOCAL DISTRIBUTION LOAD FORECASTING

In this study, power consumption data is processed with both conventional regression analysis and genetic programming techniques. We performed all of the computations described in this study using Matlab[TM] (The MathWorks Inc., Natick, MA) on an IBM PC with 3 GHz Pentium IV processor, 512 MB of memory, and Windows-XP[TM] Professional operating system. Curve fitting tool of Matlab (cftool) is used for conventional regression and GPLAB Toolbox for Matlab (available from http://gplab.sourceforge.net) is used for applying genetic programming.

Curve fitting tool of Matlab can be used to fit data using polynomial, exponential, rational, Gaussian and other equations. It also provides statistics to evaluate the goodness of a fit produced.

GPLAB is a free, highly configurable and extendable genetic programming toolbox supporting up-to-date features of the recent genetic programming research.

Curve fitting tool is used for comparison with the genetic programming application. Among the different types of the fit, a 4th degree polynomial and a power equation the following form has produced the best results. Coefficients are calculated with 95% confidence bounds.

The equation found for the 4th degree polynomial is $f(x) = p_1*x^4 + p_2*x^3 + p_3*x^2 + p_4*x + p_5$ with calculated coefficients p_1=2.049, p_2=-5.175, p_3=3.498, p_4=0.7127, p_5=0.4699.

The equation found for the power equation is $f(x) = ax^b + c$ with calculated coefficients a=1.914, b=0.3409, c=-0.3431.

Same data has been used in GPLAB to find a symbolic model. The parameters used in the GP application are listed in Table 1. Standard crossover and mutation operators of the GPLAB are used. GPLAB supports adaptive probabilities for genetic operators. Probabilities of operators generating better individuals in each run are incremented and decremented if they produce worse individuals.

"rand" operator generates a random number between 0 and 1. "mydivide" operator returns the dividend if divisor is zero or the division otherwise. "mypower" operator returns 0 if X1^X2 is NaN (not a number), infinite, or has imaginary part, otherwise returns X1^X2. Fitness function is the standard function of GPLAB used for symbol regression problems. To be able to compare the results of GP with other regression methods, SSE for the best solution found by GP is also calculated. When selecting individuals for reproduction, tournament method is used with a tournament size of 25. The GP has been run for 200 generations with a population size of 200.

Table 1. Parameters used in the genetic programming application

Parameter	Value
Genetic operators	Crossover and mutation
Operator probabilities	Variable
Population diversity calculator	Uniquegen
Selection method	Tournament
Tournament size	25
Tree initialization method	Ramped half-and-half
Operators	rand, plus, minus, times, mydivide, mypower, sin, cos, log
Fitness function	Sum of absolute difference between the expected output value and the value returned by the individual
Number of generations	200
Population size	200

Since values of the data are large, each annual consumption value (yn) is represented as $y_n' = y_n / 10^9$ in all three models to make calculations efficiently.

4. RESULTS

The input to all three methods and the output calculated for existing data are given in table 2. All three models are constructed using input data. After the models have been constructed, power consumption values for the previous years are calculated using these models. Data in the table has also been shown graphically in Fig. 5.

The model generated by GP method has 51 nodes and the depth is 17 levels. GP simulation has been run for 60 minutes.

Table 2. Input to and output from three methods

| Years	Input Data (MWh)	Output of the polynomial model	Error %	Output of the power equation model	Error %	Output of the GP model	Error %
1994	583,835,150	571,222,169	0,0216	529,963,176	0,0923	582,237,085	0,0027
1995	694,312,077	714,291,378	0,0288	762,699,648	0,0985	730,416,316	0,0494
1996	851,470,138	875,475,063	0,0282	926,630,360	0,0883	849,058,233	0,0028
1997	1,071,627,809	1,036,021,424	0,0332	1,057,477,084	0,0132	1,074,374,233	0,0026
1998	1,212,078,708	1,182,096,795	0,0247	1,168,185,943	0,0362	1,209,824,362	0,0019
1999	1,275,954,531	1,304,785,652	0,0226	1,265,107,087	0,0085	1,312,746,495	0,0280
2000	1,395,583,698	1,400,090,607	0,0032	1,351,886,289	0,0313	1,395,695,795	0,0001
2001	1,440,110,142	1,468,932,413	0,0200	1,430,833,849	0,0064	1,465,610,216	0,0174
2002	1,543,594,161	1,517,149,960	0,0171	1,503,517,209	0,0260	1,540,809,210	0,0018
2003	1,569,570,225	1,555,500,277	0,0090	1,571,054,560	0,0009	1,540,302,306	0,0190
2004	1,585,366,488	1,599,658,531	0,0090	1,634,274,951	0,0308	1,604,301,096	0,0118
2005	1,671,939,175	1,672,180,280	0,0010	1,693,812,141	0,0131	1,675,694,099	0,0022
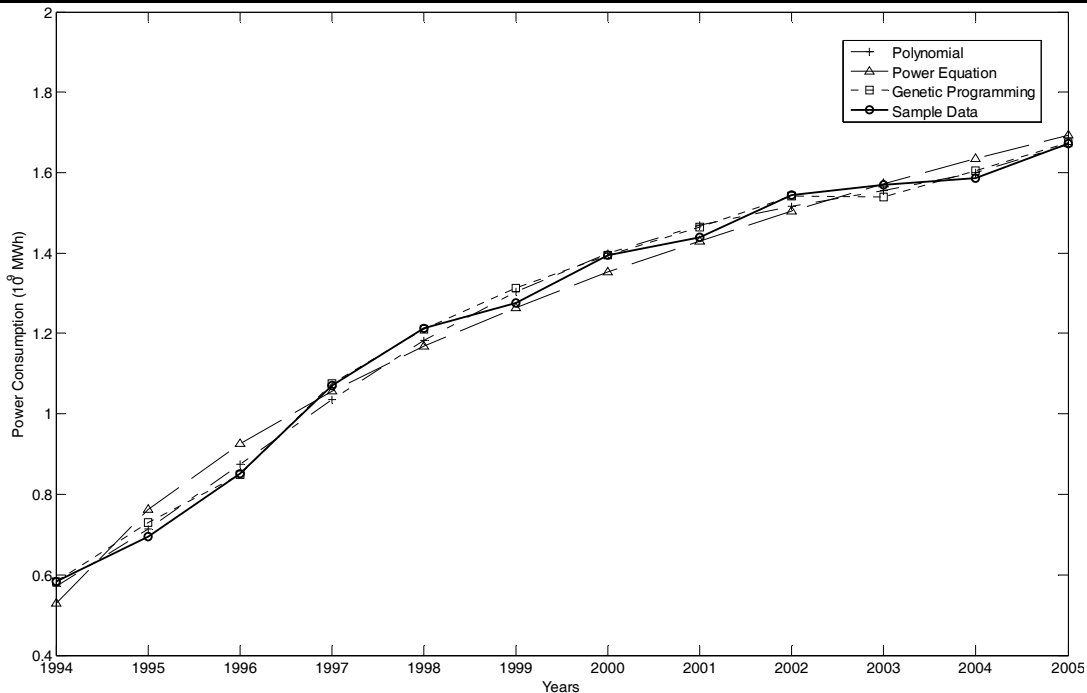
Fig. 5. Comparison of input data and values produced by three methods

Calculated SSE values for the three models are given in table 3. As seen from the table, GP has outperformed the other two methods. Thus, it can be concluded that GP has produced a better model.

Table 3. SSE values of the three methods

Method	SSE value
Polynomial	0.00608797
Power equations	0.02194693
GP	0.00038000

Forecasts for the next years from the three models are given in table 4. Forecasts using polynomial and power equations model have 95% prediction bounds. Forecasts are calculated by providing new years to constructed models and computing the corresponding values. These results are also shown in Fig. 6.

Table 4. Forecasts from three models

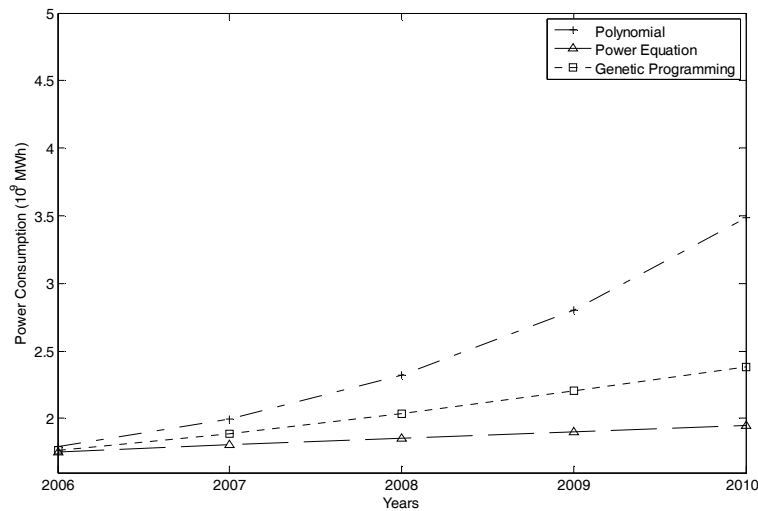| Years	Output of the polynomial model	Output of the power equations model	Output of the GP
2006	1,792,690,212	1,750,162,789	1,767,946,313
2007	1,997,504,667	1,803,724,223	1,888,205,855
2008	2,320,009,115	1,854,819,876	2,036,490,833
2009	2,800,469,415	1,903,716,971	2,205,872,493
2010	3,484,069,565	1,950,639,142	2,384,744,357

Fig. 6. Comparison of forecasts produced by three methods

5. CONCLUSION

Long term power consumption forecasting can provide important information for power distribution centers. As seen from the data, power consumption is in this city is rapidly growing, therefore accurate forecasts can help authorities to make reliable plans.

In this work, a genetic programming based forecasting method is presented. Two other curve fitting methods are also presented for comparison with this technique. Data used in all three models are not preprocessed. Genetic programming technique is used

to form a model and evaluate the parameters for the model. The goodness of the fit produced by the genetic programming method is evaluated using SSE method, which is better than the other two methods of regression. We can say that the genetic programming can be used for electric utility resource planning and forecasting of the future load demand in the regional or national service area effectively.

6. REFERENCES

1. Al-Hamadi, H.M., Soliman S.A.: Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. Electric Power Systems Research 74 (2005) 353–361.

2. Galiana, F., Handschin, E., Fiechter, A.: Identification of stochastic electric load models from physical data. IEEE Trans. Automatic Control 19 (December (6)) (1974) 887–893.

3. Sisworahardjo, N.S., El-Keib, A.A., Choi, J., Valenzuela, J., Brooks, R., El-Agtal, I.: A stochastic load model for an electricity market. Electric Power Systems Research 76 (2006) 500–508.

4. Park, J.H., Park, Y.M., Lee, K.Y.: Composite modeling for adaptive short term load forecasting. IEEE Trans. Power Syst. 6 (1991) 450–457.

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

6. Zivanovic, R.: Nonparametric trend model for short term electricity demand forecasting. In: Fifth International Conference on Power System Management and Control 2002, April 17–19, (2002) 347–352.

7. Papadakis, S.E., Theocharis, J.B., Bakirtzis, A.G.: A load curve based fuzzy modeling technique for short-term load forecasting. Fuzzy Sets and Systems 135 (2003) 279–303.

8. Al-Kandari, A.M., Soliman, S.A., ElHawary, M.E.: Fuzzy short term electric load forecasting. Int j of Electrical power and energy systems, 26 (2004) 111-122.

9. Otavio, A.S., Carpinteiro, A., Agnaldo, J.R., Reis, A., Alexandre P.A., Da Silva, B.: A hierarchical neural model in short-term load forecasting. Applied Soft Computing 4 (2004) 405–412.

10. Huang, H.C., Hwang, R.C., Hsieh, J.G.: A new artificial intelligent peak power load forecaster based on non fixed neural networks. Int j of Electrical power and energy systems, 24 (2002) 245-250.

11. Ghiassi, M., Zimbra, D.K., Saidane, H.: Medium term system load forecasting with a dynamic artificial neural network model. Electric Power System Research. 76 (2006) 302-316.

12. Liang, R.H., Cheng, C.C.: Combined regression-fuzzy approach for short-term load forecasting. IEEE Proceedings-Generation Transmission and Distribution. 147 (2000) 261–266.

13. Mori, H., Kosemura, N., Ishiguro K., Kondo T.: Short-term load forecasting with fuzzy regression tree in power systems. In: IEEE International Conference on Systems, Man, and Cybernetics 2001, vol. 3, October 7–10, (2001) 1948–1953.
14. Koza, J.R.: Genetic Programming: On the Programming of Computers by Means of Natural Selection, MIT Press (1992).
15. Koza, J.R.: Survey of genetic algorithms and genetic programming. In: Proceedings of the Wescon 95 - Conference Record: Microelectronics, Communications Technology, Producing Quality Products, Mobile and Portable Power, Emerging Technologies, San Francisco, CA, 7.-9. November, IEEE, New York, (1995).
16. Koza, J.R.: Genetic Programming, Encyclopedia of Computer Science and Technology. vol. 39 (1998) 29-43.