Daily Energy Demand Forecasting Using A Neural Data Mining Approach with Emphasis on Holidays Treatment

Iván Aquino[1], Samuel A. Oporto[1] Jacqueline K. Chávez[1], César O. Pérez[1], Giovana M. Urbina[1], and Herbert Medina[1]

National University of Engineering, Lima, PERU,
ivaqmo@gmail.com, soporto@wiphala.net, karinajcc@gmail.com
cesaroperez@gmail.com, giovanna.urbina@gmail.com, hmedina@aloe.ulima.edu.pe

Abstract. Monthly energy demand forecasting, peak load demand forecasting and daily energy demand forecasting for the next month are important issues in energy markets. The first and second issue has received much attention from researchers and practitioners in the last decade, the third one not much. The purpose of this paper is to present a neural data mining approach based on quick propagation neural network to the daily energy demand forecasting of the next month for a Regulated Client of Peru. Because energy demand forecasting on holidays are more complex than in nonholidays a new special holidays codifying is presented in the data preparation phase that produce better forecasting results both in holidays and in the whole month than other possible holidays codifying. The results obtained by using the quick propagation neural network and the proposed holidays codifying are compared with those produced by using an ARIMA model. The comparative results show that quick propagation neural network with the proposed codifying form outperform the ARIMA model on holidays of July and August of 2005. Also the neural network outperforms in average the ARIMA model in July 2005, a month that has two holidays.

Key Words: Daily Energy Demand Forecasting, Data Mining, Holidays codifying, ARIMA.

1 Introduction

Load forecasting is a central process in the planning and the operation of electric utilities for nonregulated markets whereas in countries like Peru, where the energy market is regulated, the energy demanded from regulated contracts play an important role for generation companies. Energy demand forecasting for several days helps generators to planning how to fit the energy demand from contracts with the generation scheduled in advanced. The attempts to forecasting energy are usually concentrated in Load Forecasting in short, mid and large terms. Those kinds of models are not suitable for regulated markets where power and capacity are commercialized. Therefore, despite of the fact that mid-term load forecasting has received much attention from researchers and practitioners in the last decade [1],[2],[3],[4] , models for short-term load have been analyzed [5], [6], [7], while attempts to forecasting the demand are few [8],[9], [13], [14]. In this paper a neural data mining approach to daily energy demand forecasting is presented.

An artificial neural network with Quick Propagation training is used as a data mining technique to forecasting the demand for the next thirty days. This technique is used because it is faster than the conventional Back Propagation Neural Network [10] . The data to be used in the experiments was obtained from Egenor, a electric generation company from Peru. This data is preprocessed, structured and mined using the neural network. Also, a new way to code holidays that take into account days before and after holidays is used. The forecasting results obtained by the neural network are compared with those obtained using ARIMA model, which is a technique frequently used in Load Forecasting problems. The comparative results shows that the forecasting error performed by the quick propagation neural network *around 1.6% in July 2005* is better than those from ARIMA *around 2.8% in July 2005.* Also in holidays of July and August of 2005 the quick propagation neural network that uses our proposed holidays codifying clearly gives an average improvement (less error) of around 60% in July (a month with two local holidays) and 3.5% in August (a month with one local holiday) over the forecasting error of ARIMA.

To develop the study the next sections has been organized as follows: Section 2 presents data and task description, Section 3 presents the techniques used in the experiments, Section 4 presents experiments and Section 5 presents conclusions and further Works.

2 Data and Problem Description

2.1 Problem Description

The problem for regulated energy markets such as Peru is to forecast the energy demand for the next month, specially for regulated clients which are distributors, while capacity demand from contracts is established in advanced, the demand of energy is open, and at the end of the month generation companies match the demand to fit contracts demand versus energy generated, the difference of those values must be sold or bought from market, called Spot Market, where the price of energy is variable. Hence, it is necessary to use a suitable forecasting model to planing the best time to scheduling maintenances or other operations that decrease the generation available specially when there is not enough level of water to cover the consumption just with Hydro Plants and to use Thermal Plants, consequently the price of the energy at Spot Market goes up, when these event occurs forecasting is the only way to reduce the risk of expected revenues. Egenor, an electric generation company, has such problem, to forecast the electric energy demand for its regulated clients. The evaluation of the forecasting accuracy is based on the Median Average Percentage Error:

$$\text{MAPE} = 100\frac{\sum_{i=1}^{n}\frac{|Y_i - \hat{Y}_i|}{Y_i}}{n}, \qquad n = 29, 30, 31 \tag{1}$$

where Y_i and $\hat{Y}_i$ are the real and the predicted value of accumulated energy demand on the ith day of the month under forecasting. The goal is to forecast the accumulated energy demand of the next month with minimum MAPE, which is minimum error or its equivalent better accuracy.

2.2 Data Description

The data consist of daily energy demand for 2003, 2004, 2005 and additional characteristics such holiday mark is included. The data scheme provided is: $Data = \{Year, Month, DayOnMonth, DayOnWeek, Holiday, Energy\}$.

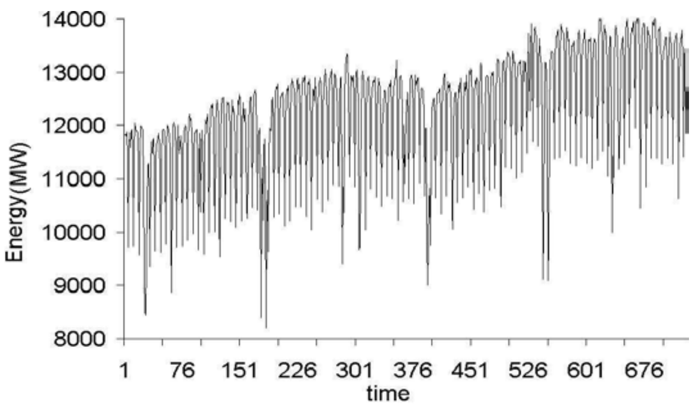

Fig. 1. Energy demand from July 2003 to June 2005

2.3 Properties of Load Demand

Energy demand data given is accumulated per day. Fig 1 gives a simple description of the energy demand data from 2004 to 2005. By simple analysis, it is noticeable that there is a seasonal pattern in each year: a little high demand in summer and low in winter. A linear trend from one year to the next year is also observed. Fig 2 shows the repetitive patterns in the month January 2003 and 2004, this pattern represent the energy demand in a week that is similar for 2004 and 2005. Also the energy demand in weekends is lower than in weekdays.

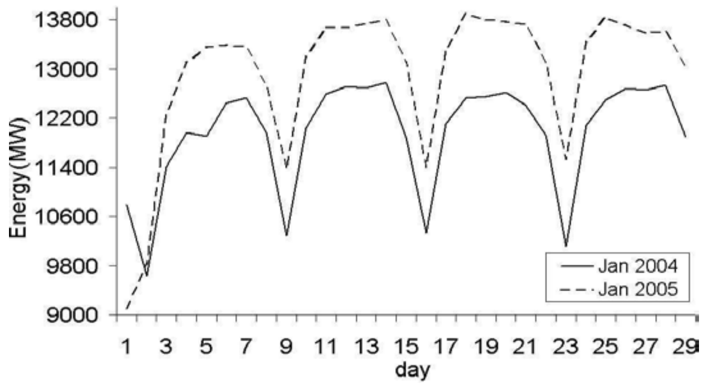

Fig. 2. Energy demand in January 2004 and 2005. January 1st is holiday

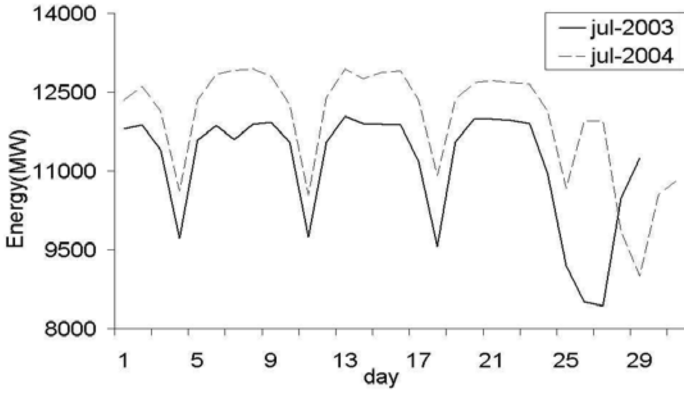

Fig. 3. Energy demand in July, a month that has two local holidays

2.4 Holiday Effects

Holidays are special days that have a high influence in energy demand. Like any other energy data studied in academy, for the data given it is observed that energy demand is lower in holidays than in weekdays as shown in Fig 3. In addition the days before and after the holidays present lower energy demand than other non holidays. Therefore holidays require special treatment as described in [11],[6]. In this work the global model proposed in this paper consider a new way to capture the holiday effect.

3 Data Mining Techniques

3.1 Neural Networks

Quick propagation is a heuristic modification of the back propagation algorithm, introduced by Fahlman in 1988 [12]. Every steps proceeds as in standard back-propagation, but for each weight a copy is kept of the $\partial E/\partial w(t-1)$ the error derivative computed during the previous training epoch, along with the difference between the current and previous values of this weight. The $\partial E/\partial w(t)$ value for the current training epoch is also available at weight-update time.

Then two risky assumptions are made: first, that the error versus weight curve for each weight can be approximated by a parabola whose arms open upward; second, that the change in the slope of the error curve, as seen by each weight, is not affected by all the other weights that are changing at the same time. For each weight, independently, it is used the previous and current error slopes and the weight-change between the points at which these slopes were measured to determine a parabola, then it is jumped directly to the minimum point of this parabola. The computation is very simpler and faster than the Back Propagation training [10], and it uses only the information local to the weight being updated:

$$\mathrm{w_t} = \frac{S_t}{S_{t-1} - S_t} \triangle w_{t-1} \tag{2}$$

where S_t and S_{t-1} are the current and previous values of $\partial E/\partial w(t)$. Of course, this new value is only a crude approximation to the optimum value for the weight [12].

3.2 ARIMA

ARIMA is an autoregressive integrated moving average model used to modeling the serial correlation in the disturbance. An ARIMA usually expressed by ARIMA(p,d,q) is composed by three tools:
1. Autoregressive term with order p where each AR term correspond to the use of a lagged value of the residual in the forecasting equation for the unconditional residual.
2. d indicates the integration order term. Each integration order corresponds to differencing the series being forecast.
3. Moving Average term with order q where this term uses lagged values of the forecast error to improve the current forecast.
All of these characteristics mentioned above can be described by:

$$\phi_p(B)\nabla^d {y_t}^{\lambda,m} = C + \theta_q(B)a_t \tag{3}$$

where:

$$\phi_p(B) = 1 + \phi_1(B) + \phi_2(B)^2 + ... + \phi_p(B)^p \tag{4}$$

$$\theta_q(B) = 1 + \theta_1(B) + \theta_2(B)^2 + ... + \theta_q(B)^q \tag{5}$$

$$(B)/B^{\pm\kappa} y_t = y_{t\pm\kappa} \tag{6}$$

$$\nabla \equiv 1 - B \tag{7}$$

$${y_t}^{\lambda,m} = \begin{Bmatrix} \frac{(y_t+m)^\lambda - 1}{\lambda}, si\lambda \neq 0 \\ ln(y_t + m), si\lambda = 0 \end{Bmatrix} \tag{8}$$

The ARIMA model used in this paper corresponds to the Seasonal ARIMA expressed by $ARIMA(p, d, q) \times (P, D, Q)_S$, where s represents the seasonal period, finally the SARIMA model used can be described by:

$$\phi_p(B)\Phi_P(B^s)\nabla_S^D\nabla^d {y_t}^{\lambda,m} = C + \theta_q(B)\Theta_Q(B^S)a_t \tag{9}$$

4 Experiments and Results

4.1 Data Preparation:

Being aware that data preparation is very important in any data mining procedure, special and careful attention will be taken to analyze in detail the energy demand data.

Feature Selection : It consists in selecting the features that will predict the energy demand data. As we saw in Section 2.3, the energy demand is higher in weekdays than in weekends. In addition, there is a linear trend from one year to the next and as we saw in Section 2.4 the energy demand is lower in holidays than in nonholidays. Therefore including this information: Year,Month,Day in a month, Day in a week and Holiday in the model may be useful for having forecasting accuracy.

Data Transformation : Once the features are selected, a careful transformation process is needed in order to made input variables suitable for the neural network. As year and day of a month are periodic values, and neural networks work better with numeric features than with categorical ones, these features will be encoded with sin and cos functions. It allows us to represent these two input features and eliminate the implicit ordinality in them that troubles the learning process of the neural network. Feature month will be encoded as numeric feature ranging from 1,2 to 12. Holidays demand forecasting is a difficult task due to there is no much data to learn from as shown in [8]. Hence error forecasting in holidays is very high [11]. In [9] holidays are encoded as binary variables, in [8] holidays are encoded as any other weekend. In this work, knowing that holidays affect the day before and after holidays in some degree, they will be encoded with number 3 (the holidays), 1 (the day before the holiday), 2 (the day after the holiday) and 0 (nonholiday).

Training Data : The data for training will be the energy demand of the twelve months before the target month and months equals to the target month but from past years (i.e months from July 2003 to June 2004 will be used as training data to forecast energy demand data in July 2004).

Table 1. Data used for parameters tuning.

Training	Testing
July_{2003}...June_{2004}	July_{2004}
August_{2003}...July_{2004}	August_{2004}
September_{2003}...August_{2004}	September_{2004}

Table 2. Data used for proving the forecasting capability in unseen data: July, August and September of 2005.

Training	Target Month
July_{2003}, July_{2004}...June_{2005}	July_{2005}
August_{2003}, August_{2004}...July_{2005}	August_{2005}
September_{2003}, September_{2004}...August_{2005}	September_{2005}

4.2 Experiments and Results

In the experiments, the target months under forecasting will be July, August and September. July is a month that has two local holidays and August has one local holiday. Table 1 presents the data used for training and testing during the parameters tuning phase. The parameters of neural network with quick propagation algorithm are: hidden layers h, iterations $iter$, learning rate lr and quick propagation coefficient q.

According to this parameters, an experiment that takes $h = 1, 2, 3, ..., 20$, $iter = 50, 55, 60, ..., 500$, $lr = 0.1, 0.2, 0.3, ..., 0.7$ and $q = 0.5, 0.75, 1, ..., 2.5$ were applied to data showed in Table 1. Each combination of the parameters were run ten times in order to get confidence. After this experiments, the parameters that model the data better are: $h = 10$, $iter$=300, lr=0.3 and $q = 1.75$ producing the highest forecasting accuracy in the parameters tuning phase. These parameters are used to forecast the energy demand of July, August and September of 2005 using the before and after holidays codifying described in Section 4.1. In addition this proposed holidays codifying is compared with other 3 ways of holidays codifying making in total 4: a) Holiday=1 used in [9] b) Holiday=3, day before holiday=1, day after holiday=2 the proposed in this paper c) Holiday=Sunday=1 used in [8] d) Holiday=Sunday=3, day before holiday=1 and day after holiday=2. For all of these, nonholidays=0. Also each form of codifying holidays can be treated as numeric or binary resulting in total 8 forms of encoding holidays.

Table 2 presents the data structure to be used in the simulation of the neural network. This last data structure includes one month of data from 2003 in Training data in order to give the neural networks more patterns for learning from the target month as described in Section 4.1 (Data Transformation).

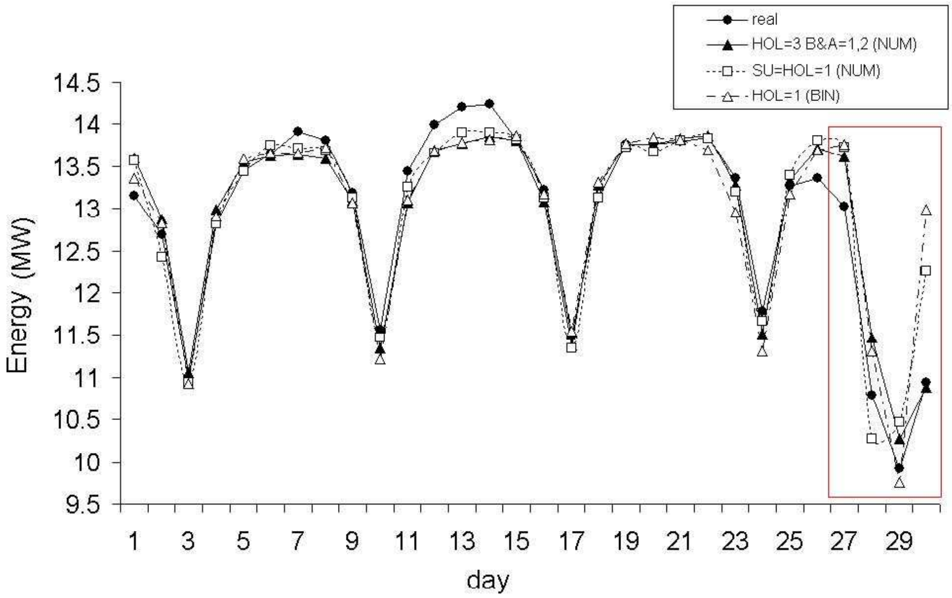

Fig. 4. Forecasting results in July 2005: days 28 and 29 are holidays

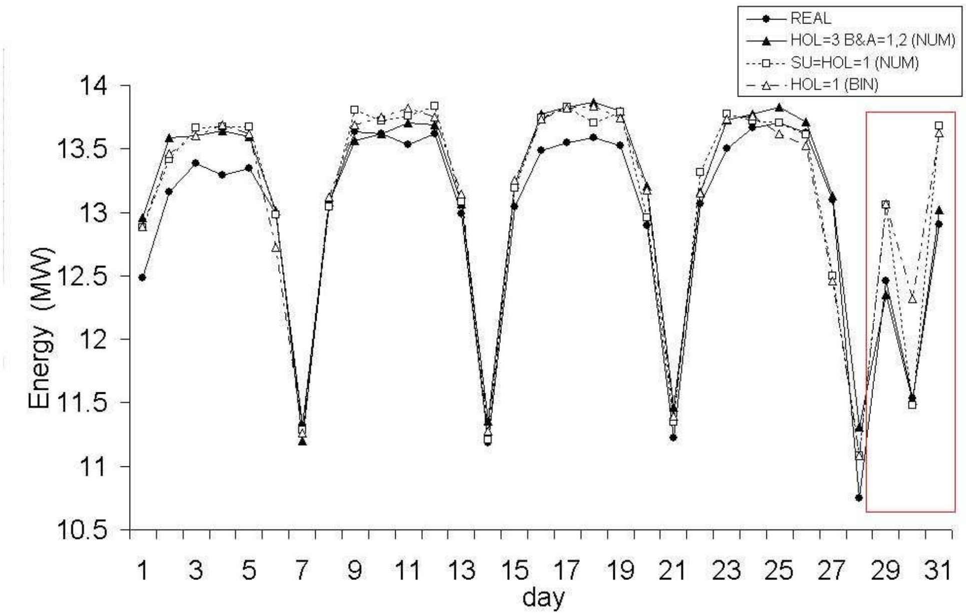

Fig. 5. Forecasting results in August 2005: day 30 is a holiday

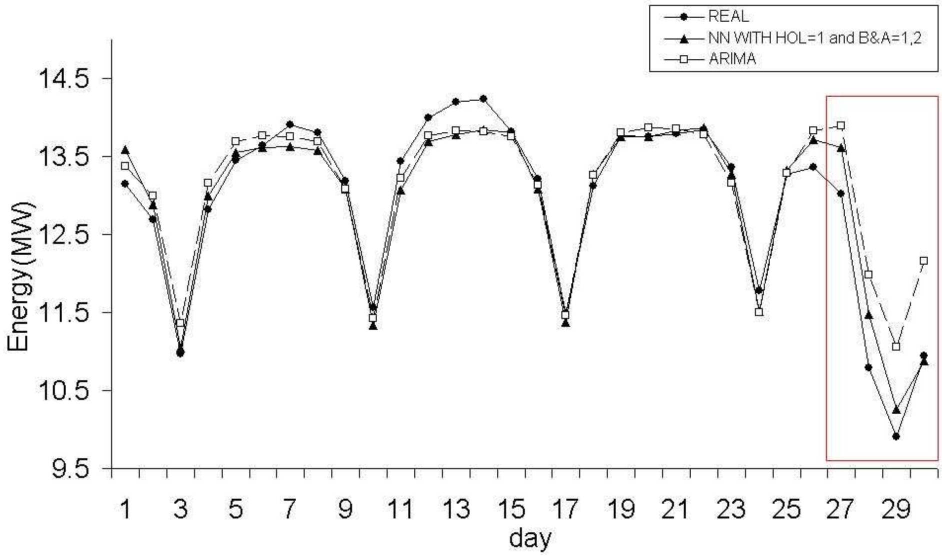

Fig. 6. Comparative results between Neural Network with the proposed holidays codifying and ARIMA model in July 2005: days 28 and 29 are holiday

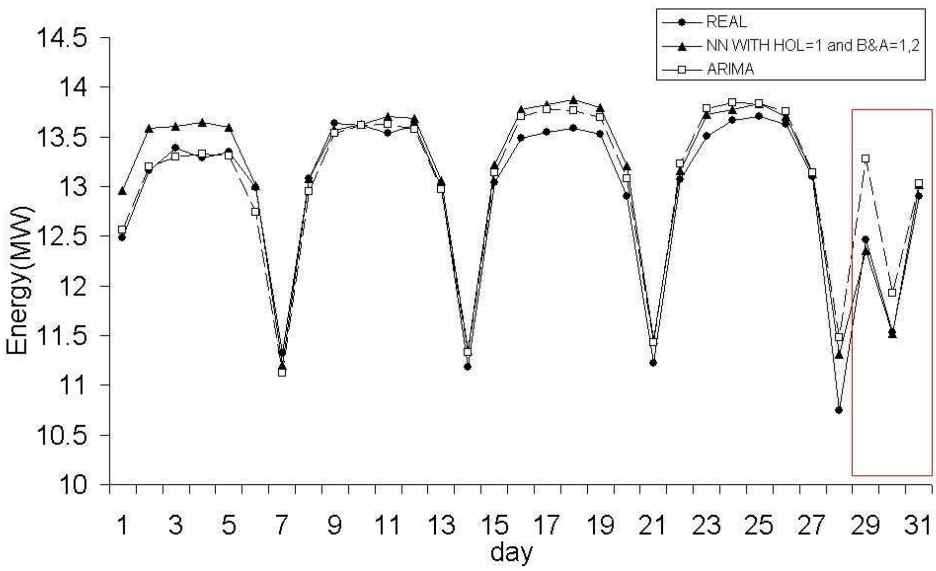

Fig. 7. Comparative results between Neural Network with the proposed holidays codifying and ARIMA model in August 2005: day 30 is a holiday

Table 3. MAPE produced by applying quick propagation neural network using the eight Holidays codifying forms where HOL=holiday, Su=Sunday, B&A=before and after holidays, NUM=numeric, BIN=binary

Target Month	HOL=1		HOL=3 B&A=1,2		Su=HOL=1		HOL=SU=3, B&A 1,2	
	NUM	BIN	NUM	BIN	NUM	BIN	NUM	BIN
July_{2005}	2.08%	2.24%	***1.64%***	1.86%	1.85%	2.30%	2.28%	2.13%
August_{2005}	1.75%	1.77%	***1.45%***	1.60%	1.69%	1.55%	1.65%	1.89%
September_{2005}	***0.82%***	0.95%	0.91%	0.92%	0.85%	1.03%	0.92%	1.17%

Figures 4 and 5 presents the forecasting results in July and August 2005(target months). By using the neural network parameters founded above and the 8 codifying forms, the simulation was run for each combination of month and codifying form. The results of this are presented in Table 3. The Before and After holidays codifying with numeric transformation is the one that produce the most forecasting accuracy (less error) in months with holidays with MAPE=1.64 % in July and MAPE=1.45% in August. In september, the 1 and 0 codifying with numeric transformation is the best with MAPE=0.82%. This could be because september is a month that does not have any holidays and there is no reason to give a special treatment to holidays.

Now, the resuls obtained by the neural network with HOL=1, B&A=1,2 holidays codifying are compared with results obtained by using Autoregressive integrate model average in the same months. The results of these are shown in Figure 6 and 7 for July and August respectively.

In addition the forecasting results produced by the neural network with the proposed holidays codifying gives an improvement of around 60% in July (27,28,29 and 30) and 3.5% in August (29,30 and 31) over the forecasting accuracy of ARIMA (See Table 4).

Moreover, according to the MAPE, the neural network is better than ARIMA model in July but almost equal in August (See Table 4).

5 Conclusions and Further Works

A neural data mining approach that emphasize the coding of days before a holiday, holidays and days after holidays has been developed. Because of the rapidity, quick propagation neural network has been used as data mining technique. Also, a suitable data structure based in taking the last twelve months and months equal to the target month of past years has been used. Then the forecasting results of the proposed approach have been compared with those obtained by using ARIMA model showing that our proposed approach produce high acceptable forecasting errors of 1.64% and 1.45% in July and August 2005 respectively compared to 2.81% and 1.44% produced by the ARIMA model for the same months. Also, in holidays of July and August of 2005 the quick propagation neural network that uses our proposed holidays codifying clearly gives

Table 4. Daily forecasting errors (%) of Neural Network with the proposed holidays codifying and the ARIMA model in July and August 2005. It is observed that Neural Network NN gives an improvement over ARIMA in holidays and days before and after holidays: 27, 28 (holiday), 29 (holiday) and 30 of July and 29,30 (holiday) and 31 of August

FORECASTING ERRORS PER DAY %						
	July			August		
Day	NN	ARIMA	Improvement (%)	NN	ARIMA	Improvement (%)
1	3.42	1.77	-93.20	3.81	0.66	-481.63
2	1.43	2.39	39.96	3.26	0.34	-862.72
3	0.74	3.62	79.60	1.63	0.61	-166.24
4	1.33	2.59	48.63	2.64	0.30	-773.26
5	0.74	1.83	59.42	1.82	0.32	-472.43
6	0.18	0.88	79.59	0.17	1.91	91.14
7	1.97	1.07	-83.97	1.10	1.75	37.49
8	1.58	0.77	-106.56	0.02	0.96	97.59
9	0.74	0.75	1.59	0.48	0.73	34.67
10	1.93	1.19	-61.76	0.04	0.04	-4.74
11	2.72	1.63	-66.88	1.28	0.70	-83.20
12	2.13	1.65	-29.37	0.51	0.34	-52.01
13	2.99	2.60	-15.17	0.55	0.13	-321.51
14	2.75	2.97	7.50	1.55	1.38	-12.43
15	0.04	0.47	90.70	1.36	0.77	-75.82
16	0.99	0.60	-63.71	2.10	1.58	-33.31
17	0.91	0.12	-644.79	2.05	1.63	-25.62
18	1.21	1.08	-11.98	2.07	1.28	-61.03
19	0.07	0.26	71.82	2.01	1.25	-60.69
20	0.02	0.80	97.49	2.39	1.43	-66.73
21	0.22	0.39	44.45	2.13	1.87	-13.73
22	0.18	0.46	61.66	0.70	1.26	44.41
23	0.62	1.50	58.22	1.66	2.05	19.36
24	2.23	2.40	7.28	0.79	1.30	38.68
25	0.46	0.14	-225.72	0.93	0.91	-2.28
26	2.71	3.51	22.61	0.60	0.93	35.71
27	***4.58***	***6.71***	***31.82***	0.23	0.33	28.60
28	***6.37***	***11.10***	***42.56***	5.24	6.84	23.42
29	***3.61***	***11.60***	***68.92***	***0.85***	***6.53***	***86.95***
30	***0.54***	***11.12***	***95.17***	***0.13***	***3.45***	***96.16***
31	1.62	9.19	82.37	***0.93***	***1.00***	***7.02***
MAPE	***1.64***	***2.81***		***1.45***	***1.44***	

an improvement of around 60% in July (a month with two local holidays) and 3.5% in August (a month with one local holiday) over the forecasting accuracy of ARIMA.
Further works include to extend this study to the remaining months. Also it would be interesting to structure further the data prior to be mined by the neural network.

References

1. Teixeira, M.A., Zaverucha, G., Navarro, V., Ferreira, G.: Fuzzy Bayes Predictor in Electric Load Forecasting. International Joint Conference on Neural Networks, 2001. Proceedings of the IJCNN '01. **4** (2001) pp. 2339 - 2342.
2. Teixeira, M.A., Zaverucha, G.: Fuzzy Markov predictor in electric load forecasting. International Joint Conference on Neural Networks, 2002. Proceedings of the IJCNN '02. **3** (2002) pp. 2416 - 2421.
3. Borlea, l., Buta, A., Lustrea, B: Some Aspects Concerning Mid Term Monthly Load Forecasting Using ANN. The International Conference on Computer as a Tool, 2005. EUROCON 2005. **1** (2005) pp. 253 - 256.
4. Tsekouras, G.J., Elias, C.N., Kavatza, S.; Contaxis, G.C.: A hybrid non-linear regression midterm energy forecasting method using data mining. IEEE Bologna Power Tech Conference Proceedings, 2003. **1** (2005) pp. 6.
5. Cho, M.Y., Hwang, J.C., Chen, C.S.: Customer short term load forecasting by using ARIMA transfer function model. IEEE Trans. Power Syst **3** (1988) pp. 908-914.
6. Srinivasan, D., Chang, C.S., Liew, A.C.: Demand forecasting using neural computation, with special emphasis on weekend and public holiday forecasting. IEEE Trans. Power Syst **10** (1995) pp. 1897-1903.
7. Hippert, H.S., Pedreira, C.E., Souza, R.C.: Neural networks for short-term load forecasting: A review and evaluation. IEEE Trans. Power Syst **16** (2001) pp. 44-55.
8. Chen, B., Chang, M., Lin, Ch., : Load Forecasting Using Support Vector Machines:A Study on EUNITE Competition 2001. IEEE Transactions on Power Systems, VOL. 19, NO. 4, November 2004. **19** (2004) pp. 1821–1830.
9. Ortiz-Arroyo, D., Skov, M.K., Quang Huynh: Accurate Electricity Load Forecasting with Artificial Neural Networks. International Conference on Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce. **1** (2005) pp. 94 - 99.
10. Rumelhart, D.E., Hinton,G.E., Williams,R.J.: Learning internal representations by error propagation. D. Rumelhart and J. McClelland editors.
11. Song, K., Baek, Y., Hun Hong, D., Jang, G.: Short term load forecasting for the holidays using fuzzy linear regression method. IEEE Trans. Power Syst **20** (2005) pp. 96-101.
12. Fahlman, S. E.:Faster-Learning Variations on Back-Propagation: An Empirical Study. Proceedings of 1988 Connectionist Models Summer School (1988).
13. Charitoniuk, W., Chen, M.S., Kotas, P., Van Olinda, P.: Demand forecasting in power distribution systems using nonparametric probability density distribution. IEEE Trans. Power Syst **14** (1999) pp. 1200-1206.
14. Tian, L., Noore, A.: Short term load forecasting using optimized neural network with genetic algorithm. 8^{th} International Connference on Probabilistic Methods Applied to Power Systems (2004) pp. 135-140.