

Variable Selection for Five-Minute Ahead Electricity Load Forecasting

Irena Koprinska

School of Information Technologies
University of Sydney
Sydney, Australia
e-mail: irena@it.usyd.edu.au

Rohen Sood

School of Electrical and Information Engineering
University of Sydney
Sydney, Australia
rohens13@gmail.com

Vassilios Agelidis

Centre for Energy Research and Policy Analysis
University of New South Wales
Sydney, Australia
e-mail: vassilios.agelidis@unsw.edu.au

Abstract—We use autocorrelation analysis to extract 6 nested feature sets of previous electricity loads for 5-minute ahead electricity load forecasting. We evaluate their predictive power using Australian electricity data. Our results show that the most important variables for accurate prediction are previous loads from the forecast day, 1, 2 and 7 days ago. By using also load variables from 3 and 6 days ago, we achieved small further improvements. The 3 bigger feature sets (37-51 features) when used with linear regression and support vector regression algorithms, were more accurate than the benchmarks. The overall best prediction model in terms of accuracy and training time was linear regression using the set of 51 features.

Keywords- very short-term electricity load forecasting; prediction; variable selection; autocorrelation analysis

I. INTRODUCTION

Five-minute ahead load forecasting is important in deregulated and competitive energy markets, such as the Australian national electricity market. It is used for a number of decisions, including optimal allocation of generators and setting minimum generation reserves. Accurate forecasting is very important. Reducing the Mean Absolute Percentage Error (MAPE) with 1% for a 10 GW generator saves US\$1.6 million per year [1].

There are only a few published papers on very short-term load forecasting (hours and minutes ahead). Backpropagation Neural Networks (BPNN) are the most popular prediction algorithms used by the research community and industry forecasters. In [2] BPNN was used to predict the load 10 minutes ahead from the load in the previous 20-90 minutes, achieving MAPE of 0.4-1%. In [3] a BPNN for 5-minute ahead load prediction was proposed. The data was preprocessed by applying a logarithmic difference of the consecutive loads. An excellent MAPE of 0.12% was achieved. Taylor [4] used minute-by-minute British load data to predict the load 10-30 minutes ahead. He applied autoregressive and exponential smoothing methods. The

Holt-Winter's smoothing using the intraweek and intraday cycle performed best, with MAPE of 0.4% for 30-minute ahead prediction.

This paper is an extension of our previous work [5], where we used autocorrelation analysis with 4-week sliding window to select a set of 51 variables extracted from the top 6 autocorrelation peaks. In this paper we take a closer look at the variables extracted from each peak and investigate their predictive power and contribution. We construct 6 nested feature sets with increasing size, starting with the variables extracted from the highest peak and gradually adding variables from the other peaks. Our goal is to systematically identify the effect of the 6 feature sets and better understand the underlying process.

The performance of the 6 feature sets is evaluated using 4 algorithms: BPNN, Support Vector Regression (SVR), Linear Regression (LR) and Least Median Squares Regression (LMS). We use electricity load demand data from 2006-2008 for the state of New South Wales (NSW) in Australia, provided by the Australian electricity market operator [6]. We compare with 2 benchmarks: a previous best feature set and the typical prediction model used by industry.

II. FEATURE SETS

The autocorrelation function of a time series with values X consists of the autocorrelations r_1, r_2, \dots , etc. at lags 1, 2, ..., etc. For example, r_1 indicates how the X values 1 lag apart (i.e. the successive values) relate to each other, r_2 indicates how the X values 2 lags apart relate to each other, etc. This function is used to investigate the cyclic nature of the time series. Values close to 0 show lack of autocorrelation while values close to 1 or -1 (i.e. spikes) indicate high positive or negative autocorrelation, respectively.

Fig. 1 shows the autocorrelation function for August 2007 for NSW; the graph for 2006 is very similar and not shown. There are several clusters of high spikes. As in [5] we select the top 6 spikes (shown in Table I in decreasing order) and then extract 6 sets of variables from the areas

around them, as also shown in Table I. The first set of variables is extracted from the cluster around the highest spike and thus contains the most important variables based on the autocorrelation graph, the second set is extracted from the cluster around the second highest spike, etc.

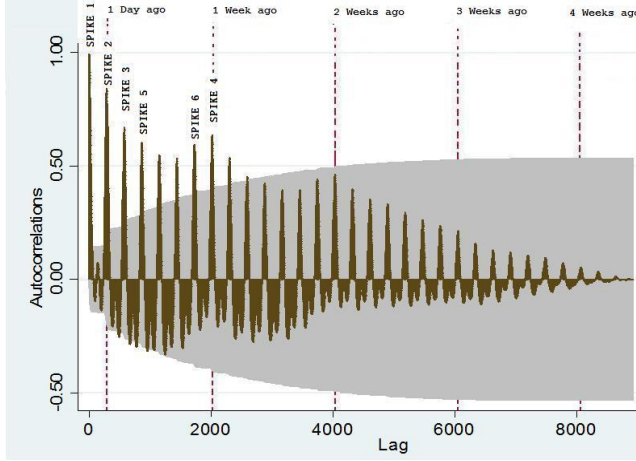


Figure 1. Autocorrelation function of the electricity load for August 2007 (1 lag = 5 minutes)

TABLE I. CLUSTERS OF HIGH SPIKES AND SELECTED VARIABLES

	Lag	Variables	Description
1	0	X_t, \dots, X_{t-8}	Load on the forecast day at $t, \dots, t-8$
2	288	$XD_{t+2}, \dots, XD_{t-4}$	Load 1 day ago at $t+2, \dots, t-4$
3	576	$XD2_{t+2}, \dots, XD2_{t-4}$	Load 2 days ago at $t+2, \dots, t-4$
4	2016	$XX_{t+6}, \dots, XX_{t-7}$	Load 7 days ago at $t+6, \dots, t-7$
5	864	$XD3_{t+2}, \dots, XD3_{t-4}$	Load 3 days ago at $t+2, \dots, t-4$
6	1728	$XD6_{t+2}, \dots, XD6_{t-4}$	Load 6 days ago at $t+2, \dots, t-4$

We then form 6 nested feature sets, with increasing size, as shown in Table II: FS-A, FS-B, FS-C, FS-D, FS-E and FS-all. Each feature set includes the previous set and also new variables from the corresponding spike cluster, e.g. FS-B includes FS-A and the variables from spike cluster 2; FS-all includes FS-E and the variables from spike 6.

TABLE II. SELECTED FEATURE SETS

Feature set	To predict X_{t+j} , use:
FS-A (9 features)	X_t, \dots, X_{t-8}
FS-B (16 features)	FS-A + $XD_{t+2}, \dots, XD_{t-4}$
FS-C (23 features)	FS-A + FS-B + $XD2_{t+2}, \dots, XD2_{t-4}$
FS-D (37 features)	FS-A + FS-B + FS-C + $XX_{t+6}, \dots, XX_{t-7}$
FS-E (44 features)	FS-A + FS-B + FS-C + FS-D + $XD3_{t+2}, \dots, XD3_{t-4}$
FS-all (51 features)	FS-A + FS-B + FS-C + FS-D + FS-E + $XD6_{t+2}, \dots, XD6_{t-4}$

III. EXPERIMENTAL SETUP

We created the prediction models using data from August 2006 and 2007 and tested them on data from August 2008. There were 17,856 instances in the training data and 8,928 instances in the test data.

A. Prediction Algorithms

As in our previous work [5, 7], we applied SVR, LR, LMS and BPNN. BPNN is the most popular algorithm for load forecasting used by researchers and industry forecasters. LR and LMS are standard linear regression algorithms. SVR is a state-of-the-art algorithm. We used the WEKA library [8]. Table III lists the parameters we used.

TABLE III. PREDICTION ALGORITHMS

LR. It uses the least squares method to find the best fit line. We applied the MS's method for variable selection - starting with all variables, at each step the weakest variable is removed until there is no improvement in the error estimate.
LMS. An extension of LR that uses least squared regression functions generated from random subsamples of the data. The final model is the one with the lowest median squared error. We used sample size=4 for the size of the sub-samples.
BPNN. A classical neural network trained with the back-propagation algorithm. We used 1 hidden layer (the number of units in it was set to the average of the number of input and output units), sigmoid transfer functions, learning rate=0.3, momentum=0.2, maximum training epochs =500.
SVR. An extension of the support vector machine algorithm for numeric prediction. SVR with kernel functions can create complex nonlinear decision boundaries while reducing the computational cost. We used polynomial kernel, $\epsilon=0.001$, $C=1$.

B. Benchmark Feature Sets

Table IV lists the 2 feature sets used for comparison. FS1 is the feature set that we found to work best in [7]. To predict the load at a given time, it uses load from the same day (from the previous 5 lags) and load from 7 days ago (at the forecast time and previous 5 lags).

FS-industry is a typical feature set used by industry. It is based on FS1 but applies a natural logarithmic difference as in [3] to improve the data stationarity. This feature set is typically employed in conjunction with BPNN; we will call this combination the ‘‘industry model’’.

TABLE IV. BENCHMARK FEATURE SETS

FS1 (11 features) - to predict X_{t+j}, use:	
X_t, \dots, X_{t-4}	Load on the forecast day at times $t, \dots, t-4$ (5 features).
$XX_{t+1}, \dots, XX_{t-4}$	Load 7 days ago at times $t+1, \dots, t-4$ (6 features).
FS-industry (9 features) - to predict $\ln(X_{t+j}/X_t)$ (that is then transformed back to X_{t+j}), use:	
$\ln(X_t/X_{t-1}), \dots, \ln(X_{t-3}/X_{t-4})$	Logarithmic load differences from the forecast day (4 features)
$\ln(XX_{t+1}/XX_t), \dots, \ln(XX_{t-3}/XX_{t-4})$	Logarithmic load differences 7 days ago (5 features)

C. Performance Metrics

We used the following accuracy metrics:

1) Mean absolute error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |L_actual_i - L_forecast_i|$$

where L_{actual_i} is the actual electricity load, $L_{forecast_i}$ is the forecasted load, i is the 5-minute interval and n is the total number of predicted loads.

2) Relative absolute error (RAE)

$$RAE = \frac{MAE}{MAE_{ZeroR}} \cdot 100 [\%]$$

It measures the MAE of the prediction model used relative to the MAE of a very simple predictor, ZeroR, which predicts the mean value in the training data.

3) Mean absolute percentage error (MAPE), an extension of MAE defined as:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{L_{actual_i} - L_{forecast_i}}{L_{actual_i}} \right| \cdot 100 [\%]$$

IV. RESULTS AND DISCUSSION

Table V shows the performance of the 2 benchmark feature sets and Table VI shows the performance of the 6 feature sets based on the autocorrelation analysis. The number of features in each set is shown in brackets. The MAPE results are also shown in Fig. 2 for visual comparison across all sets and algorithms.

TABLE V. PERFORMANCE USING THE BENCHMARK FEATURE SETS

	LR	LMS	BPNN	SVR
FS1 (11)				
MAE	29.35	29.58	36.21	29.43
RAE [%]	2.34	2.36	2.89	2.35
MAPE [%]	0.296	0.298	0.365	0.297
FS-industry (9)				
MAE	29.41	29.57	28.88	29.42
RAE [%]	2.35	2.37	2.31	2.35
MAPE [%]	0.297	0.299	0.291	0.297

The most accurate prediction models were LR with FS-all and SVR with FS-all. They outperformed the industry model and also the previous best model LR with FS1 [7]. However, SVR was significantly slower than LR – it took 3 seconds to build the LR model and 16 hours to build the SVR model. Thus, overall, considering both accuracy and time, LR with FS-all was the best prediction model.

A comparison across the algorithms on all feature sets indicates that LR was the most accurate algorithm, very closely followed by LMS and SVR, and that BPNN was the least accurate algorithm. The small RAE values (from 2.13% to 2.97%) show that all prediction models significantly improved over the baseline of simply predicting the mean value.

A comparison across the 6 autocorrelation feature sets shows that the 3 smaller feature sets FS-A, FS-B and FS-C were not able to improve over the benchmarks used for comparison. However, the 3 bigger sets, FS-D, FS-E and FS-all, outperformed both the industry model and the benchmark set FS1, with all algorithms except BPNN. For BPNN all sets except the smallest set FS-A outperformed FS1 but neither of them outperformed the industry model (also based

on BPNN but using logarithmic load differences). The less consistent performance of BPNN on the 6 autocorrelation feature sets, in comparison to the other prediction algorithms, may be due to its greater sensitivity to parameter setting.

TABLE VI. PERFORMANCE USING THE FEATURE SETS BASED ON AUTOCORRELATION ANALYSIS

	LR	LMS	BPNN	SVR
FS-A (9)				
MAE	33.63	34.08	37.14	33.65
RAE [%]	2.68	2.72	2.97	2.69
MAPE [%]	0.339	0.343	0.372	0.339
FS-B (16)				
MAE	30.16	30.85	31.23	30.27
RAE [%]	2.41	2.46	2.49	2.42
MAPE [%]	0.305	0.312	0.315	0.306
FS-C (23)				
MAE	29.36	30.16	30.90	29.48
RAE [%]	2.34	2.41	2.47	2.35
MAPE [%]	0.297	0.306	0.312	0.299
FS-D (37)				
MAE	27.16	27.36	34.12	27.19
RAE [%]	2.17	2.18	2.72	2.17
MAPE [%]	0.274	0.276	0.342	0.274
FS-E (44)				
MAE	26.91	27.05	33.26	26.93
RAE [%]	2.15	2.16	2.66	2.15
MAPE [%]	0.272	0.273	0.333	0.272
FS-all (51)				
MAE	26.64	27.42	33.28	26.65
RAE [%]	2.13	2.19	2.66	2.31
MAPE [%]	0.269	0.277	0.332	0.269

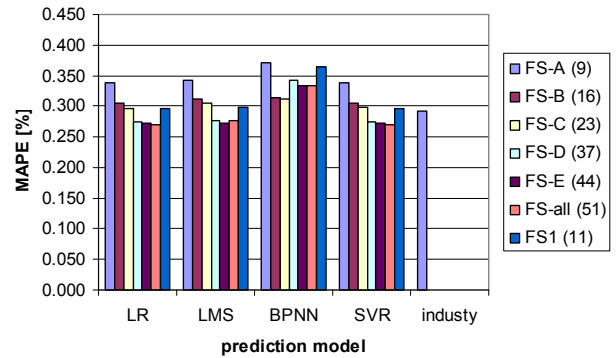


Figure 2. Comparison between all feature sets – MAPE [%]

As the number of features in the 6 sets increases from 9 to 51, the accuracy improves, for all algorithms except BPNN. This improvement is high between the first and second set (from 9 to 16 features) and the third and fourth set (from 23 to 37 features) and small between the last 3 sets (37, 44 and 51 features). It is not possible to directly correlate these improvements with the differences between the autocorrelation peaks from Fig. 1 as the feature sets are nested and contain features extracted from more than one peak.

Hence, the results show that the most important variables are previous loads from the forecast day, 1, 2 and 7 days ago (included in FS-D); the use of additional load variables from

3 and 6 days ago (as in FS-E and FS-all) improves accuracy but not greatly.

As the number of features increases from 9 to 51, the time to build the prediction models also increases. In contrast to SVR, the increased training time was not an issue for LR; its training time was 0.11s using 9 features and 3s using 51 features. Thus, with LR we can use the biggest and most accurate set FS-all.

V. CONCLUSIONS

We evaluated the predictive power of 6 nested feature sets, constructed based on autocorrelation analysis. Three of them, when used with LR, LMS and SVR, were more accurate than the industry model and the benchmark set FS1. In general, as the number of features increases, the accuracy improves. The results showed that the most important variables were extracted from the 4 highest autocorrelation peaks (loads from the same day, 1, 2 and 7 days before); adding variables extracted from the next 2 peaks (loads from 3 and 6 days before) resulted in small further improvements. Overall, the best prediction model was LR with the set of 51 features extracted from all 6 autocorrelation peaks; it was more accurate than the industry model and was also fast to train.

REFERENCES

- [1] B. F. Hobbs, S. Jitrapaikularn, S. Konda et al., "Analysis of the value for unit commitment of improved load forecasts," *IEEE Trans. Power Systems*, vol. 14, no. 4, pp. 1342-1348, 1999.
- [2] W. Charytoniuk, and M.-S. Chen, "Very short-term load forecasting using artificial neural networks," *IEEE Trans. Power Systems*, vol. 15, no. 1, pp. 263-268, 2000.
- [3] P. Shamsollahi, K. W. Cheung, Q. Chen, E. H. Germain, "A neural network based very short term load forecaster for the interim ISO New England electricity market system," in *22nd IEEE PES Int. Conf. Power Industry and Computer Applications (PICA)*, 2001, pp. 217-222.
- [4] J. W. Taylor, "An evaluation of methods for very short-term load forecasting using minute-by-minute British data," *Int. J. Forecasting*, vol. 24, pp. 645-658, 2008.
- [5] R. Sood, I. Koprinska, and V. G. Agelidis, "Electricity load forecasting based on autocorrelation analysis," in *Int. Joint Conference on Neural Networks (IJCNN)*, Barcelona, 2010, in press.
- [6] AEMO, 2009; www.aemo.com.au.
- [7] A. Setiawan, I. Koprinska, and V. G. Agelidis, "Very short-term electricity load demand forecasting using support vector regression," in *Int. Joint Conference on Neural Networks (IJCNN)*, Atlanta, 2009, pp.2888-2894.
- [8] I. Witten, and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2d ed.: Morgan Kaufmann, 2005.