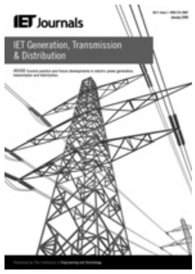
Published in IET Generation, Transmission & Distribution
Received on 21st July 2008
Revised on 7th September 2009
doi: 10.1049/iet-gtd.2009.0297



ISSN 1751-8687

Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network

N. Amjady[1] A. Daraeepour[2] F. Keynia[1]
[1]Department of Electrical Engineering, Semnan University, Semnan, Iran
[2]The Iranian Grid Management Company (IGMC), Tehran, Iran
E-mail: amjady@tavanir.org.ir

Abstract: In a power market, the price of electricity is the most important signal to all market participants. However, electricity price forecast is a complex task due to non-linearity, non-stationarity and volatility of the price signal. In spite of all performed research on this area in the recent years, there is still an essential need for more accurate price forecast methods. Besides, there is a lack of robust feature selection technique for designing the input vector of electricity price forecast, which can consider the non-linearities of the price signal. In this study, a new price forecast method is proposed, which is composed of a modified version of Relief algorithm for feature selection and a hybrid neural network for prediction. The proposed approach is examined on Ontario, New England and Italian electricity markets and compared with some of the most recently published price forecast methods.

1 Introduction

Price forecast is a key information in today's electricity markets. Companies that trade in electricity markets make extensive use of price forecast techniques either to bid or hedge against volatility. However, electricity has distinct characteristics from other commodities. The electrical energy cannot be considerably stored and the power system stability requires constant balance between generation and load. On short time scales, most users of electricity are unaware of or indifferent to its price. Transmission bottlenecks usually limit electricity transportation from one region to another. These facts enforce the extreme price volatility or even price spikes of the electricity market [1]. Besides, volatility in fuel price, load uncertainty, fluctuations in the hydroelectricity production, generation uncertainty (outages) and behaviour of market participants also contribute in the electricity price uncertainty [2].

In recent years, several methods have been applied to predict prices in the electricity markets. Different time series techniques such as dynamic regression and transfer function [3], auto-regressive integrated moving average (ARIMA) [4], mixed ARIMA models [5] and generalised auto-regressive conditional heteroskedastic (GARCH) [6] have been proposed for this purpose. Some other research works suggest artificial neural networks (NNs) [7–10], fuzzy NNs [11, 12] and support vector machines (SVMs) [13] for price forecast. The combination of similar day and NN techniques [14], extended Kalman filter (EKF)-based NN [15], weighted nearest neighbours (WNN) method [16], combination of fuzzy inference system and least-squares estimation [17] and sensitivity-based dynamic model [18] have been also presented for electricity price forecast. Some other researchers proposed specific methods for price spike forecasting [19, 20].

In spite of all performed researches in the area of electricity price forecasting, there is still an essential need for more accurate and stable price forecast methods. Especially, there is a lack of robust feature selection technique for designing the input vector of electricity price forecast, which can consider the non-linearities of the price signal. It is noted that the design of the input vector for electricity price forecast is usually carried out in a discretionary way, mainly based on trial-and-error procedures and on engineering judgement criteria. Recently, some researchers propose correlation analysis [3, 4, 17] and numerical sensitivity analysis [2, 15] for this purpose, which are linear feature selection techniques. However, the electricity price is generally a non-linear mapping function of its input variables [1].

In this paper a new prediction strategy is proposed for day-ahead price forecasting of electricity markets. Contribution of the paper can be summarised as follows:

1. Presentation of a non-linear feature selection technique based on a modified version of Relief algorithm to select relevant input variables for electricity price forecast.

2. Presentation of a new hybrid neural network (HNN) for prediction of electricity prices based on the input variables selected by the modified Relief algorithm. The HNN is composed of NNs and evolutionary algorithms (EAs) in a cascaded structure.

The remaining parts of the paper are organised as follows. In Section 2, the proposed feature selection technique is described. The HNN as the forecast engine is explained in Section 3. In Section 4, the obtained numerical results from the modified Relief algorithm and HNN for price forecast of Ontario, New England and Italian electricity markets are presented and discussed. Section 5 concludes the paper.

2 Proposed feature selection algorithm

Selecting the best set of input features is a crucial preprocessing for the successful application of NNs. The main idea of feature selection is to choose a subset of input variables by eliminating irrelevant features. The consecutive set of features not only has a smaller dimension but also has as much information as the original set. Feature selection can simplify the learning process of the forecasting tool and enhance its generalisation capability for unseen data. In this paper, a modified version of the Relief algorithm is proposed for the relevance analysis of the candidate inputs. The Relief algorithm is considered one of the most successful methods among the existing feature weighting algorithms, due to its simplicity and effectiveness [21, 22]. Relief has been shown to detect relevance well, even when features interact [23], like the case of electricity price forecasting.

Relief is a non-linear instance-based feature selection technique, which assigns relevance values to features by treating training samples as points in the feature space. Each feature's weight reflects its ability to distinguish among the class labels. In the following, we first present the original Relief algorithm, which is limited to classification problems with two classes. Then, our proposed modifications for the original Relief algorithm are explained and so the modified Relief algorithm is introduced, it not only deals with both forecasting and multi-class problems but also is more robust and detects relevance better than the original one.

The original Relief algorithm randomly chooses a number of training samples from the training set (this parameter is user-defined) [21, 24]. Given a randomly selected sample x, the original Relief searches for its two nearest neighbours including one from the same class, called *nearest hit* $NH(x)$, and the other from the opposite class, called *nearest miss* $NM(x)$, based on the Euclidean distance measure defined as follows

$$\begin{aligned} d_{rs} &= \|\boldsymbol{x}_r - \boldsymbol{x}_s\| = \sqrt{(\boldsymbol{x}_r - \boldsymbol{x}_s)(\boldsymbol{x}_r - \boldsymbol{x}_s)} \\ &= \sqrt{\|\boldsymbol{x}_r\|^2 + \|\boldsymbol{x}_s\|^2 - 2\boldsymbol{x}_r\,\boldsymbol{x}_s} \end{aligned} \tag{1}$$

where $\boldsymbol{x}_r$ and $\boldsymbol{x}_s$ are two vectors representing samples r and s, respectively and d_{rs} is Euclidean distance between them (dot sign indicates inner product); $\|.\|$ represents the Euclidean norm: $\|\boldsymbol{x}_r\| = \sqrt{\boldsymbol{x}_r.\boldsymbol{x}_r}$. $NH(x)$ and $NM(x)$ are two samples owning minimum Euclidean distance with x of the same and opposite classes, respectively.

In the original Relief algorithm, the weight of ith feature W_i (initialised to zero at the beginning) is updated according to the following equation

$$\begin{aligned} W_i = W_i + |x^{(i)} - NM^{(i)}(x)| - |x^{(i)} - NH^{(i)}(x)|, \\ i = 1, 2, \ \ldots, \ I \end{aligned} \tag{2}$$

where $x^{(i)}$, $NM^{(i)}(x)$ and $NH^{(i)}(x)$ represent the ith feature of the selected sample x, $NM(x)$ and $NH(x)$, respectively; I is the number of candidate input features. In this way, the weight of all candidate inputs is updated based on the selected sample x. This cycle (finding the nearest hit and nearest miss and updating the weight of features based on (2)) is repeated for all randomly selected samples and then the candidate features are ranked according to the finally obtained weight values. In other words, the original Relief algorithm updates the weight of each feature i (W_i) based on the values of $x^{(i)}$, $NM^{(i)}(x)$ and $NH^{(i)}(x)$. If $x^{(i)}$ and $NH^{(i)}(x)$ have different values, then the ith candidate feature separates two samples with the same class, which is not desirable and so its weight (W_i) decreases based on (2). On the other hand, if $x^{(i)}$ and $NM^{(i)}(x)$ have different values, then the ith candidate input separates two samples with different class labels, which is desirable and so its weight increases. In other words, a relevant feature should largely change between dissimilar samples and slightly change (or does not change at the ideal case) between similar samples.

The original Relief algorithm encounters two inconsistencies for electricity price forecasting. At first, it can be only used for two class problems, but electricity price is a continuous variable and so the nearest hit and nearest miss cannot be found. Secondly, some irrelevant features may also be selected by this algorithm, since a feature may have good behaviour based on the first-order neighbourhood of some randomly selected samples; however, it is not really a relevant feature considering the information content of the whole data set. To remedy these inconsistencies, two modifications are proposed for the Relief algorithm in this paper.

In the first modification, we add a preprocessing stage to the original Relief algorithm to make it applicable for any continuous variable problem. For this purpose, the target variable (here, the electricity price) is standardised to a standard variable with zero mean and unit standard deviation [25]. Standardisation is a well known linear mathematical transformation. Consider the data set $\{y_i\}_{i=1}^N$ with mean value y_{mean} and standard deviation y_{std}. Standardisation maps $\{y_i\}_{i=1}^N$ to another data set $\{z_i\}_{i=1}^N$ with mean z_{mean} and standard deviation z_{std} as follows

$$z_i = (y_i - y_{\text{mean}})\frac{z_{\text{std}}}{y_{\text{std}}} + z_{\text{mean}}, \quad i = 1, 2, \ldots, N \tag{3}$$

Common values for mean and standard deviation of the transformed data set $\{z_i\}_{i=1}^N$ are $z_{\text{mean}} = 0$ and $z_{\text{std}} = 1$, which are also used for the standardisation of the target variable in this paper. With $z_{\text{mean}} = 0$ and $z_{\text{std}} = 1$ we will have

$$z_i = \frac{(y_i - y_{\text{mean}})}{y_{\text{std}}}, \quad i = 1, 2, \ldots, N \tag{4}$$

which is also considered as a kind of normalisation [26]. Considering $z_{\text{mean}} = 0$, we can classify the training samples of the price forecast into two classes based on the positive and negative values of the standardised target variable. So, after the preprocessing stage, the Relief algorithm can be applied to the problem.

In the second modification, we consider k nearest neighbourhoods instead of the first one of the original Relief. Besides, for each neighbourhood all training samples are considered. The k nearest neighbour version seeks for every sample, its k nearest samples from the same class (k nearest hits) and its k nearest samples from the opposite class (k nearest misses) in the feature space. The weight is then the ratio between the sum of average distances over all samples to their k nearest misses and the sum of average distances over all samples to their k nearest hits in projection on that feature. We use the ratio because it can better discriminate the candidate features than the difference operator used in the original Relief algorithm.

Suppose $D = \{p_n = (x_n, y_n)\}_{n=1}^N$ indicates the training set where x_n and y_n represent vector of candidate inputs and target variable for training pattern p_n, respectively. We construct matrix $X = [x_n]_{n=1}^N$ such that ith row of the matrix X indicates the vector of candidate inputs for sample i and jth column of the matrix is a vector containing the values of candidate input j in the training samples of D. For all samples x_n, we find k nearest hits and k nearest misses based on the Euclidean distance and sort them in the matrices $\boldsymbol{NH}_1 = [\boldsymbol{NH}_1(x_n)]_{n=1}^N, \ldots, \boldsymbol{NH}_k = [\boldsymbol{NH}_k(x_n)]_{n=1}^N$ (the set of k nearest hits) and $\boldsymbol{NM}_1 = [\boldsymbol{NM}_1(x_n)]_{n=1}^N, \ldots, \boldsymbol{NM}_k = [\boldsymbol{NM}_k(x_n)]_{n=1}^N$ (the set of k nearest misses) where $\boldsymbol{NH}_k(x_n)$ and $\boldsymbol{NM}_k(x_n)$ denote nearest hit and miss of order k for sample x_n, respectively. Now the difference matrices are defined as follows

$$\mathbf{DNH}_1 = X - \mathbf{NH}_1, \quad \ldots, \quad \mathbf{DNH}_k = X - \mathbf{NH}_k \tag{5}$$

$$\mathbf{DNM}_1 = X - \mathbf{NM}_1, \quad \ldots, \quad \mathbf{DNM}_k = X - \boldsymbol{NM}_k \tag{6}$$

Each row of these difference matrices is related to a sample and each column is related to a candidate feature. Now the overall effect of jth feature in distinguishing the whole samples from their nearest hits and misses up to order k can be calculated as follows

$$\begin{aligned} \mathrm{WNH}_{1j} &= \frac{1}{N}\sum_{l=1}^{N} |\mathbf{DNH}_{1(l,j)}| \\ &\vdots \\ \mathrm{WNH}_{kj} &= \frac{1}{N}\sum_{l=1}^{N} |\mathbf{DNH}_{k(l,j)}| \end{aligned} \tag{7}$$

$$\begin{aligned} \mathrm{WNM}_{1j} &= \frac{1}{N}\sum_{l=1}^{N} |\mathbf{DNM}_{1(l,j)}| \\ &\vdots \\ \mathrm{WNM}_{kj} &= \frac{1}{N}\sum_{l=1}^{N} |\mathbf{DNM}_{k(l,j)}| \end{aligned} \tag{8}$$

where the subscript (l, j) indicates element located in the row l and column j of the respective matrix. The weight of jth feature, based on the k nearest neighbourhoods, can be evaluated as follows

$$W_j = \frac{\sum_{t=1}^{k} \mathrm{WNM}_{tj}}{\sum_{t=1}^{k} \mathrm{WNH}_{tj}} \tag{9}$$

The weight values obtained from (9) are normalised with respect to their maximum such that all the weights be in the range of [0, 1]. Then the candidate inputs are ranked based on their normalised weight values. The candidate inputs with the normalised weight larger than a threshold are selected by the modified Relief algorithm and the other candidates are filtered out. This threshold is determined by the cross-validation technique, which is explained in the next section.

Finally, the parameter of k, that is the number of neighbourhoods, should be determined. Small values for k may result in the selection of irrelevant features (like the problem of $k = 1$). On the other hand, the large values of k increase computation burden and at the same time farther

doi: 10.1049/iet-gtd.2009.0297

neighbourhoods are included, which may be ineffective and even misleading. The Relief algorithm models the original problem as a two class problem such that in each neighbourhood, the nearest hit and miss of the samples (from the same and opposite classes, respectively) are considered. So, this fact motivates the idea of imagining the number of neighbourhoods of the Relief algorithm like the number of levels of a binary decision tree. So, an appropriate choice for the k parameter can be obtained from the following relation

$$k = \mathrm{Round}(\log_2(N)) \tag{10}$$

where N is the number of samples or the number of training patterns; *Round*(.) is a function that rounds the real number to the closest integer value. We examined several other choices for the k parameter, but the value obtained from (10) leads to the best results for feature selection, especially for electricity price forecasting.

It is noted that training samples or instances can be considered as points in the feature space and so we can define distance between these points. In the Relief algorithm, Euclidean distance is used for this purpose, although the other distance measures, such as Bhattacharyya distance, may be considered instead. However, each candidate input is a dimension of the feature space and distance between dimensions cannot be defined. Besides, Euclidean distance is only used to measure distances between samples in the feature space and determines nearest hits and nearest misses for each sample in the modified Relief algorithm. After that the modified Relief measures the relevancy of each candidate input based on its discrimination capability to distinguish between samples and their nearest hits and nearest misses. The relevance weight is assigned to each candidate input based on its discrimination capability (and not the Euclidean distance) as shown in (2) and (9) for the original and modified Relief algorithms, respectively.

3 Proposed forecast method

As the forecast engine of electricity price prediction, a new HNN with the architecture shown in Fig. 1 is proposed in this paper. The HNN is composed of three NNs owning multi-layer perceptron (MLP) structure with one hidden layer [27].

As seen, each NN transfers two kinds of results to the next NN. The first set of results include obtained values for the

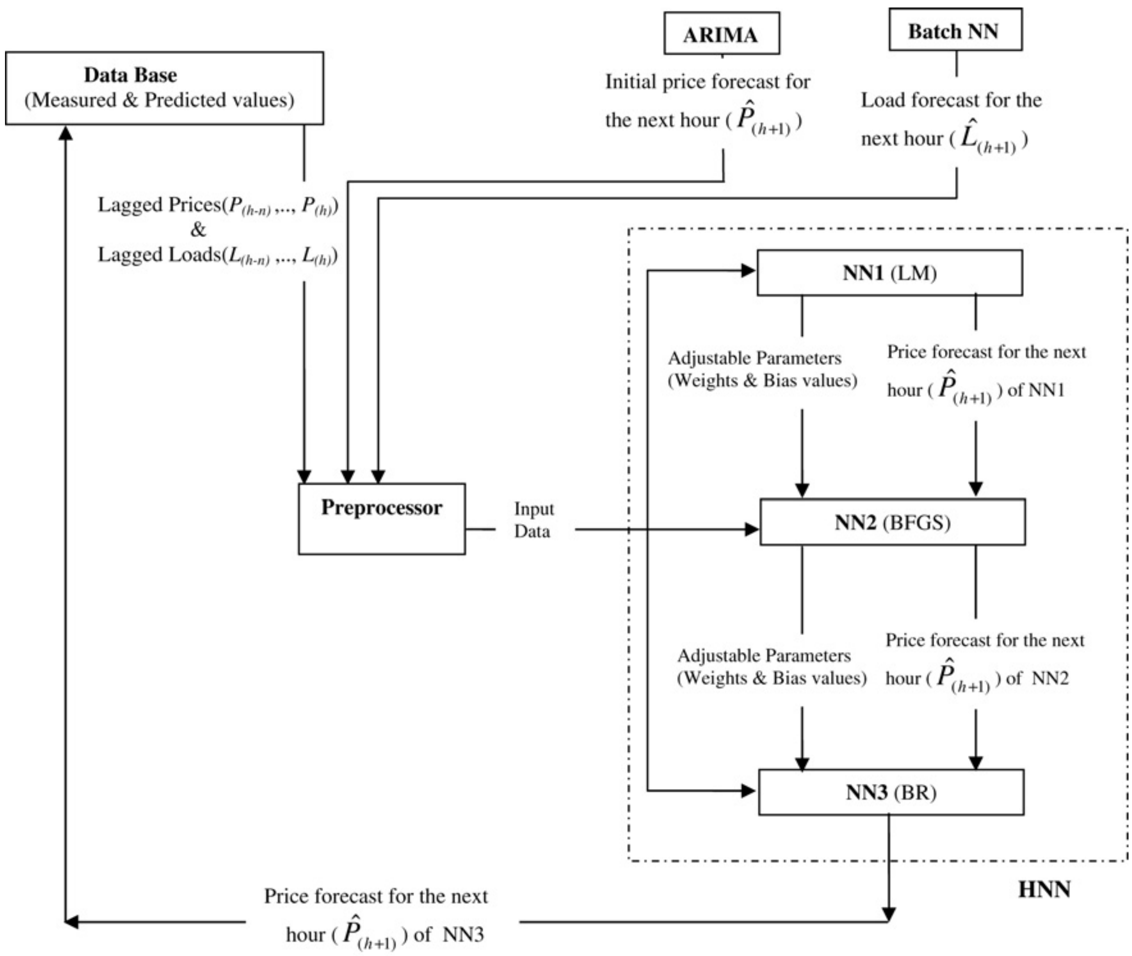


Figure 1 *Structure of the proposed forecast method including the preprocessor, HNN and its initial predictors*

doi: 10.1049/iet-gtd.2009.0297

adjustable parameters, that is, weights and bias values. In other words, each NN transfers its obtained knowledge to the next one and so it can begin its learning process from the point that the previous NN terminated (instead of beginning from a random point). Only the first NN should begin with an initial set of random values for the adjustable parameters. All NNs of the HNN have the same number of input, hidden and output neurons so that the weights and bias values of a NN can be directly used by the next one and then it can increase the obtained knowledge of its previous NN. By suitable selection of training algorithms of the NNs, the HNN can learn much more than a single NN. For the price prediction, we considered three NNs in the HNN, owning LM (Levenberg–Marquardt), BFGS (Broyden, Fletcher, Goldfarb, Shannon) and BR (Bayesian regularisation) learning algorithms, respectively. These are some of the most efficient NN training mechanisms for the prediction tasks and their mathematical details can be found in [12, 28, 29]. LM is a fast learning algorithm. At the beginning of the training phase, usually MLP can rapidly learn about the problem and its training error quickly decreases. So the LM algorithm is selected for the first NN. The BFGS is the most successful quasi-Newton method for the NN training [28]. Especially, this learning algorithm can perform a better search of the solution space provided that it starts from a suitable initial point. So, after the LM algorithm is saturated, the second NN is trained by the BFGS training mechanism to find a fitter solution for the adjustable parameters of the NN. BR learning algorithm minimises a combination of squared errors and weights and then determines the correct combination so as to produce a network that generalises well. So this learning algorithm is considered for the last NN for final tuning of the adjustable parameters and getting the utmost training efficiency. We examined several other configurations for the HNN including the mentioned learning algorithms or other training mechanisms. However, performance of the other configurations is generally less than the proposed one. Besides, we increased number of NNs of the HNN. However, no considerable improvement for price forecast is obtained and in some cases the performance of the HNN degrades. In spite of these experiments, it cannot be claimed that the proposed configuration is the best structure of the HNN. This is a matter that demands further research.

Empirically, we have seen that in many cases, including NN application for price prediction, near optimum solutions are close to each other. At the end of the training phase the learning algorithm may find one of these solutions whereas the better ones might be in its vicinity and unseen for the NN, since the learning algorithms usually search the solution space in a special direction (like the steepest descent). So, this makes the motivation to search around the final solution of the learning algorithm in various directions as much as possible to find a better solution. EA can be a suitable candidate for what is required. The evolution of the proposed EA is based on the following relations

$$\Delta W_{(n+1)} = m\,\Delta W_{(n)} + (1-m)\,gW_{(n)} \quad (11)$$

$$W_{(n+1)} = W_{(n)} + \Delta W_{(n+1)} \quad (12)$$

where W is an adjustable parameter (weight or bias value) of the NN and ΔW indicates change of it. The subscripts n and $n+1$ represent two successive generations (parent and child, respectively) of EA. g is a small random number separately generated for each adjustable parameter. m is the momentum constant. Use of the momentum can smooth the search path decreasing sudden changes. In our examinations, $m = 0.5$ and g is selected in the range of (0, 0.1) for all generations of the EA. At the beginning of the EA, $W_{(0)}$ is the obtained value from the learning algorithm for the adjustable parameter W and $\Delta W_{(0)} = 0$. In each cycle, the EA repeats (11) and (12) until the next generation of all adjustable parameters is obtained. Then, the validation error of the NN [27] is evaluated for the new generation. If the child has less error than its parent, the parent is replaced by the child, otherwise the parent is restored and the next cycle of the EA is executed. So, at the end of the EA, the best examined solution among all generations will be selected. The EA is executed after each learning algorithm of the HNN to enhance the training efficiency at most. So, the whole training phase of the HNN can be completed as shown in Fig. 2. Indeed, the BFGS learning algorithm of NN2 begins from the adjustable parameters obtained by the EA part of NN1 and similarly for the BR of NN3.

To give a better insight about the performance of the HNN in the training phase, a typical curve of its validation error [27] in terms of mean square error (MSE) for the New England Electricity market is shown in Fig. 3. The validation set includes hourly prices of 5 May 2004 and the corresponding training set is its 49 days ago. The vertical axis of Fig. 3 has logarithmic scale. As seen from this figure, LM, BFGS and BR learning algorithms have 50, 60 (110–50) and 65 (175–110) training iterations, respectively. Each of the three EA executions of Fig. 3 has 100 generations. It can be observed that the validation error decreases by about 6231 times (1.62/0.00026). Bearing in mind the validation error is prediction error for the unseen data (the validation set or 5 May 2004 is separate from the training set or its 49 days ago), this shows great generalisation capability of the HNN for the problem of price prediction. Contribution of different learning algorithms and EA executions in the generalisation capability can be seen from Fig. 3. While the NNs of the HNN perform global searches in the solution space, the EAs add the local search capability to the HNN.

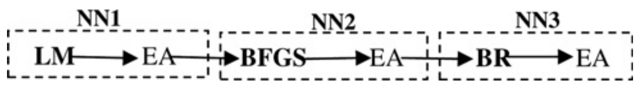


Figure 2 *Construction of the whole training phase of the HNN*

doi: 10.1049/iet-gtd.2009.0297

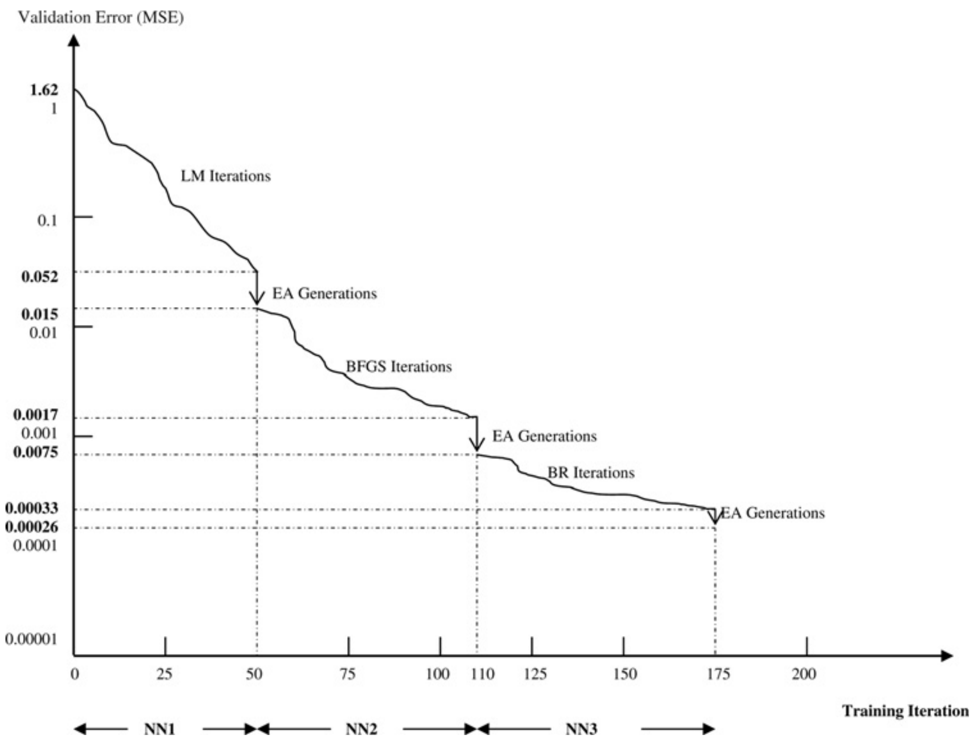


Figure 3 *Typical curve for the validation error of the HNN*

The preprocessor in Fig. 1 performs normalisation and feature selection (modified Relief). Since the candidate inputs like price and load features have different ranges, at first they are linearly normalised in the range of [0, 1] and then given to the modified Relief. In our experiments, history of price and load data are used for price forecast since the price signal is dependent on its previous values (like the effect of short-run trend and daily periodicities [1, 11]) and load is the most important price driver [14, 18]. However, our proposed feature selection technique can consider other exogenous variables like available generation, fuel price and congestion information provided that their data be available. Besides, the predicted values, obtained from the other forecast techniques, can also be useful information for price forecast. As seen from Fig. 1, price forecast of the next hour, that is $\hat{P}_{(h+1)}$, is provided for the HNN by the ARIMA time series [4, 5]. The superscript ^ indicates the predicted value. Besides, load forecast of the next hour, that is $\hat{L}_{(h+1)}$, is also provided for the HNN by a batch NN, owning MLP structure and LM learning algorithm. So the candidate inputs of the modified Relief algorithm include lagged prices and loads (e.g. up to 200 h ago) and their predictions. The input features of the HNN consist of selected inputs by the feature selection technique among the candidate inputs. These input features are given to the first NN (NN1) of the HNN (Fig. 1). It is noted that in our experiments, $\hat{P}_{(h+1)}$ (ARIMA) and $\hat{L}_{(h+1)}$ (batch NN) are always selected by the modified Relief algorithm due to their high relevance to the target variable. The NN2 of the HNN uses the same input features of NN1 except that $\hat{P}_{(h+1)}$ of the ARIMA is replaced by $\hat{P}_{(h+1)}$ of the NN1. Similarly, the NN3 uses $\hat{P}_{(h+1)}$ of the NN2. In other words, the obtained forecast of each NN of the HNN is used by the next one, since price forecast of each block usually has better accuracy than the previous one and so can be a better choice for the input feature of the next block. Indeed, $\hat{P}_{(h+1)}$ is the second kind of results, transferred between the NNs of the HNN as shown in Fig. 1. In other words, each NN of the HNN uses both the obtained knowledge and price forecast of its previous NN. Then it enhances the knowledge and produces a more accurate prediction for the price (Fig. 3).

The ARIMA provides an initial price forecast for the NNs of the HNN. This initial forecast should be as linearly independent as possible from the three NNs' predictions. So, we select a technique different from NNs to provide the initial forecast. ARIMA is a well known and efficient time series technique and so is used to generate the initial price forecast for the HNN. Load demand is a signal with different characteristics from the electricity price signal. Our experience shows that it is usually hard for a single forecaster to simultaneously learn the input/output mapping function of these two different signals and accurately predict them. So, we consider a separate forecaster to generate the required load forecast of the HNN. Since, there is no NN forecaster for load demand in the HNN, we use a NN (named batch NN in Fig. 1) to generate the required load forecast (NN can better model non-linear input/output mapping functions such as the mapping function of load demand).

It is noted that in [8] load forecast has been considered as an input feature in a cascaded structure of NNs. However, we extended this idea for feature selection by consideration of

doi: 10.1049/iet-gtd.2009.0297

both load and price forecasts. Besides, in [8] measured value of $L_{(h+1)}$ has been used for the training phase. However, we consider $\hat{L}_{(h+1)}$ for the training of the HNN (all three NNs). In this way, the HNN can see the load forecast error in the learning phase and so its NNs can be better adapted to the prediction phase when $L_{(h+1)}$ is not available. Finally, our important idea in transferring the obtained knowledge of a NN (its adjustable parameters) to the next one is not seen in [8] (our modified Relief algorithm is not seen in [8] as well).

Now, the whole setup process of the proposed method including the feature selection and training phase of the HNN can be summarised as the following step-by-step algorithm:

1. The modified Relief algorithm selects the most relevant input features from the set of normalised candidate inputs.

2. Three NNs of the HNN are trained by LM + EA, BFGS + EA and BR + EA training mechanisms, respectively. At the end of training phase of each NN, its adjustable parameters and price forecast are transferred to the next one.

3. The degrees of freedom of the whole method including the threshold of the modified Relief algorithm and number of hidden nodes of the NNs of the HNN are fine-tuned by the cross-validation technique. To reduce the number of combinations that should be examined, the two step procedure of [27] has been also used.

Note: All NNs of the HNN have one output node dedicated to price forecast of the next hour. To generate price forecasts of 24 h of the next day, iterative forecasting technique is used in the HNN in which the day-ahead price forecast is reached via recursion, that is, by feeding input variables with the forecaster outputs. In this case, when price of a hour is forecasted, it is used as $P_{(h-1)}$ for the price prediction of the next hour and this cycle is repeated until the price of the next 24 h are predicted. However, this technique may result in the propagation of error in the forecast horizon and so farther hours usually contain more forecast errors [1, 2]. If we can use from more accurate forecasts for initial hours, error propagation decreases and better predictions for all later hours can be obtained. Thus, each NN of the HNN (and also ARIMA), instead of recursive forecasting of the whole 24 h ahead, only predicts the next hour and transfers it to the following block until the final prediction of the HNN for that hour is obtained by the NN3 and inserted to the data base as shown in Fig. 1. This final forecast, owning the least prediction error in our system, is used in the ARIMA and NNs of the HNN to predict its next hour. This cycle is repeated until price of the whole 24 h ahead are forecasted.

4 Numerical results

The proposed method is examined for day-ahead price forecast of Ontario, New England and Italian electricity markets from Canada, US and Europe, respectively. Besides, we compared the method with some of the most recently published price forecast techniques. Data of Ontario, New England and Italian electricity markets have been obtained from their websites [30–32], respectively. For each forecast day, we considered its 50 days ago as the historical data including the day before the forecast day as the validation set and its 49 days ago as the training set in all experiments of this paper (like the validation and training sets of Fig. 3). After forecasting hourly prices of each day, the 50 days data window proceeds by one step and hourly prices of the next day are predicted.

In the first experiment of this paper, reported in Table 1, the performance of the proposed feature selection technique is

Table 1 Weekly MAPE values of the HNN with three different feature selection techniques for the prediction of HOEP in the Ontario electricity market (six test weeks of 2004)

Test week	Numerical sensitivity analysis + HNN (%)	Correlation analysis + HNN (%)	Original relief + HNN (%)	Modified relief + HNN (%)
26 April–2 May	13.45	12.23	10.54	9.46
3–9 May	12.23	11.43	10.52	9.72
26 July–1 August	13.12	12.54	9.79	9.00
2–8 August	12.34	11.37	9.66	9.04
13–19 December	10.45	9.98	9.81	9.36
20–26 December	10.21	9.39	9.21	8.80
average	11.97	11.16	9.92	9.23

IET Gener. Transm. Distrib., 2010, Vol. 4, Iss. 3, pp. 432–444
doi: 10.1049/iet-gtd.2009.0297

evaluated. In this table, the obtained results from the HNN with four different feature selection techniques including numerical sensitivity analysis [2, 15], correlation analysis [3, 4, 17], original Relief algorithm and proposed modified Relief for prediction of hourly Ontario energy price (HOEP) in the Ontario electricity market [30] are presented. For the original Relief algorithm, the first modification introduced in Section 2, that is the preprocessing stage (the standardisation of the target variable) is also used to make is applicable for electricity price forecast. The mean absolute percentage error (MAPE) values in Table 1 are computed as follows

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^{N} \frac{|P_{i\text{ACT}} - P_{i\text{FOR}}|}{P_{i\text{ACT}}} \tag{13}$$

where $P_{i\text{ACT}}$ and $P_{i\text{FOR}}$ are actual and forecasted price of hour i, respectively; N is the number of forecasted hours (for weekly MAPE, N equals to 168). Six test weeks from year 2004 are considered in Table 1 according to [3]. For the sake of a fair comparison, all examined feature selection techniques have the same set of candidate inputs including 200 lagged values of price and load plus predicted price and load, totally $200 + 200 + 1 + 1 = 402$ candidates. As seen from Table 1, the correlation analysis results in slightly better forecast accuracy than the numerical sensitivity analysis. However, both of these methods are linear feature selection techniques whereas the original and modified Relief are non-linear techniques and can better evaluate the non-linear dependencies of the price signal on its input variables. Also, the modified Relief, considering more neighbourhoods, can better evaluate the relevance of each candidate input to target variable than the original Relief. Table 1 shows that the modified Relief outperforms the other methods in all test periods.

For a better illustration of the performance of the numerical sensitivity analysis, correlation analysis and modified Relief algorithm, the obtained results of each method for 1 May 2004 are presented in Table 2. Similar results have been obtained for the other test days. For each feature selection technique in Table 2, selected features (among the 402 candidate inputs) as well as their ranks and weight values are reported, respectively. The weight values of the selected features of each technique are normalised with respect to their maximum. As seen from Table 2, the selected features of the modified Relief algorithm better reflects the dependencies of the price signal. For instance the modified Relief algorithm selects both forecast variables $\hat{P}_{(h+1)}$ and $\hat{L}_{(h+1)}$ (which are highly relevant features) with ranks 2 and 3, respectively. However, the numerical sensitivity analysis cannot detect these features and correlation analysis only selects $\hat{P}_{(h+1)}$ with rank 7.

The effect of short-run trend, that is strong dependencies on the previous neighbouring values, can be seen from the reported results in Table 2. For instance, $P_{(h-1)}$, $P_{(h-2)}$ and $L_{(h-1)}$ are among the selected features of the modified Relief algorithm. As the effect of daily periodicity, $P_{(h-24)}$, $L_{(h-24)}$ and $P_{(h-48)}$ are selected. $P_{(h-168)}$ and $L_{(h-168)}$ are also among the selected features of the modified Relief algorithm, which shows the effect of weekly periodicity. By combination of the effects of short-run trend and daily periodicity, some hours around the same hour in the previous day (such as $P_{(h-23)}$, $P_{(h-25)}$, $L_{(h-23)}$ and $L_{(h-25)}$) are also selected. Moreover, by combination of the effects of short-run trend and weekly periodicity, some hours around the same hour in the previous week (such as $P_{(h-169)}$, $L_{(h-169)}$ and $L_{(h-167)}$) are chosen. Furthermore, by combination of the effects of daily and weekly periodicities $P_{(h-192)}$ and $L_{(h-192)}$ ($168 + 24 = 192$ h ago) as well as $P_{(h-144)}$ and $L_{(h-144)}$ ($168 - 24 = 144$ h ago) are also selected. So, it is seen that most of the selected features by the proposed feature selection technique are sensible.

All mentioned feature selection techniques in Table 2 can give weight to features and rank them. However, a threshold is required for each feature selection technique such that its weighted features with the weight value higher than the threshold are selected as the inputs of the forecast engine. With higher values of the threshold, the pass band of feature selection technique (acting as an irrelevancy filter) becomes narrower such that less features with higher relevance weights can pass it and vice versa. In other words, selection of the threshold is a trade-off between number of selected features by the feature selection technique and their effectiveness for the forecast process. For the test case of Table 2, the threshold of the numerical sensitivity analysis, correlation analysis and modified Relief algorithm has been determined as 0.59, 0.57 and 0.43, respectively, by the cross-validation technique. In other words, for each of these methods, the best value of the threshold obtained by the cross-validation technique is used. It is noted that each of these methods has its own optimum threshold. As seen from this table, the selected features by the numerical sensitivity analysis, correlation analysis and modified Relief algorithm have normalised weights greater than 0.59, 0.57 and 0.43, respectively. Each feature selection technique filters out the remaining candidate inputs with normalised weights lower than its respective threshold.

Although there are some common features among the subsets selected by the feature selection techniques of Table 2, some different features are also observed, which result in different forecast errors using these techniques. However, the forecast accuracy is not linearly proportional to the selected features, since the electricity price is a non-linear mapping function of its inputs and the selected features have different relevance weights.

In the second experiment of the paper reported in Table 3, the proposed forecast method (modified Relief + HNN) is compared with three time series based price forecast techniques proposed in [3] (their feature selection

Table 2 Selected input features among the 402 candidate inputs and their ranks and normalised weights for 1 May 2004 of the Ontario electricity market

Numerical sensitivity analysis			Correlation analysis			Modified relief		
Selected feature	Rank	Normalised weight	Selected feature	Rank	Normalised weight	Selected feature	Rank	Normalised weight
$P_{(h-1)}$	1	1	$P_{(h-1)}$	1	1	$P_{(h-1)}$	1	1
$P_{(h-2)}$	2	0.8212	$L_{(h-1)}$	2	0.8331	$\hat{P}_{(h+1)}$	2	0.7034
$P_{(h-24)}$	3	0.7743	$P_{(h-2)}$	3	0.8069	$\hat{L}_{(h+1)}$	3	0.6923
$P_{(h-25)}$	4	0.7654	$L_{(h-168)}$	4	0.7201	$L_{(h-1)}$	4	0.6859
$L_{(h-1)}$	5	0.76	$L_{(h-24)}$	5	0.7059	$P_{(h-2)}$	5	0.6619
$L_{(h-24)}$	6	0.7543	$L_{(h-2)}$	6	0.7049	$L_{(h-168)}$	6	0.5781
$P_{(h-168)}$	7	0.7538	$\hat{P}_{(h+1)}$	7	0.7029	$P_{(h-168)}$	7	0.5708
$L_{(h-167)}$	8	0.7365	$L_{(h-192)}$	8	0.6905	$L_{(h-24)}$	8	0.5353
$L_{(h-2)}$	9	0.7241	$P_{(h-24)}$	9	0.6809	$P_{(h-24)}$	9	0.5284
$L_{(h-143)}$	10	0.7127	$L_{(h-167)}$	10	0.6797	$P_{(h-23)}$	10	0.5011
$P_{(h-163)}$	11	0.6761	$L_{(h-23)}$	11	0.6758	$P_{(h-25)}$	11	0.5011
$L_{(h-25)}$	12	0.67	$L_{(h-169)}$	12	0.6708	$L_{(h-167)}$	12	0.5011
$L_{(h-166)}$	13	0.6623	$L_{(h-191)}$	13	0.655	$P_{(h-192)}$	13	0.4945
$P_{(h-22)}$	14	0.6467	$P_{(h-3)}$	14	0.6538	$L_{(h-192)}$	14	0.4945
$P_{(h-23)}$	15	0.6452	$L_{(h-125)}$	15	0.6491	$P_{(h-144)}$	15	0.4747
$L_{(h-22)}$	16	0.642	$L_{(h-193)}$	16	0.641	$P_{(h-169)}$	16	0.4619
$L_{(h-23)}$	17	0.6393	$P_{(h-23)}$	17	0.6345	$L_{(h-144)}$	17	0.4555
$P_{(h-192)}$	18	0.6373	$L_{(h-144)}$	18	0.6239	$L_{(h-169)}$	18	0.4492
$P_{(h-48)}$	19	0.6318	$P_{(h-25)}$	19	0.6181	$P_{(h-48)}$	19	0.4429
$L_{(h-148)}$	20	0.629	$L_{(h-48)}$	20	0.5982	$P_{(h-120)}$	20	0.4306
$L_{(h-72)}$	21	0.6228	$P_{(h-168)}$	21	0.5894	$L_{(h-23)}$	21	0.4306
$P_{(h-3)}$	22	0.596	$L_{(h-22)}$	22	0.5882	$L_{(h-25)}$	22	0.4306
			$L_{(h-143)}$	23	0.5861			
			$L_{(h-166)}$	24	0.5825			
			$L_{(h-72)}$	25	0.5747			
			$L_{(h-145)}$	26	0.5738			
			$L_{(h-120)}$	27	0.5731			

technique is correlation analysis) for 24 h ahead HOEP prediction in the Ontario electricity market. MAPE has been defined in (13) and mean absolute error (MAE) is as follows

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |P_{i\mathrm{ACT}} - P_{i\mathrm{FOR}}| \qquad (14)$$

For weekly MAE, $N = 168$, like weekly MAPE. The MAPE and MAE values of the three other methods have been quoted from [3]. Table 3 shows that both the MAPE and MAE values of the proposed method are considerably lower than those of the ARIMA, Transfer Function and Dynamic Regression in the all six test weeks, which indicates price forecast capability of the proposed method. The obtained results of the proposed method for the second test week (3–9 May) are graphically shown in

doi: 10.1049/iet-gtd.2009.0297

Table 3 Weekly MAPE (%) and weekly MAE ($/MWh) for HOEP forecast in the Ontario electricity market

Test week	ARIMA [3]		Transfer function [3]		Dynamic regression [3]		Proposed	
	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE
26 April–2 May	15.9	7.2	15.6	7.1	15.9	7.3	9.46	6.6
3–9 May	18.6	8.2	18	8.2	18.1	8.2	9.72	7.3
26 July–1 August	13.6	6.9	12.3	6.4	13	7.2	9.00	6.7
2–8 August	21.5	8.7	18.3	7.3	19	7.6	9.04	7.1
13–19 December	15.4	9.6	14.8	9.2	14.7	9.3	9.36	8.7
20–26 December	20.8	12	17.5	10.1	18.5	10.7	8.80	9.3
average	17.6	8.8	16.1	8.1	16.5	8.4	9.23	7.6

Fig. 4. As seen, the price forecast curve reasonably follows the actual price curve and error curve has low values.

In the third experiment of this paper, reported in Table 4, the proposed method is examined on the New England electricity market for next day electricity price forecast. For comparison, monthly MAPE and MAE of the forecast prices published by ISO New England [31] as well as those of single SVM and hybrid network [13] are also presented in this table. The hybrid network is composed of the self-organised map (SOM) in the first stage and a group of SVMs in the second stage. For the sake of a fair comparison, the same three test months of [13] have been adopted in this paper. Besides, the results of the other methods have been directly quoted from [13]. Table 4 shows that the proposed method outperforms all other methods in all test months.

In the last experiment of this paper, the proposed method is compared with dynamic price forecast technique [18] on the Italian and New England electricity markets (Table 5). The same test months of [18] have been adopted in this experiment, which include May 2004–March 2005 for the Italian market and February 2004–December 2004 for the New England market. However, in the MAPE definition of [18], the average price of the period considered is used

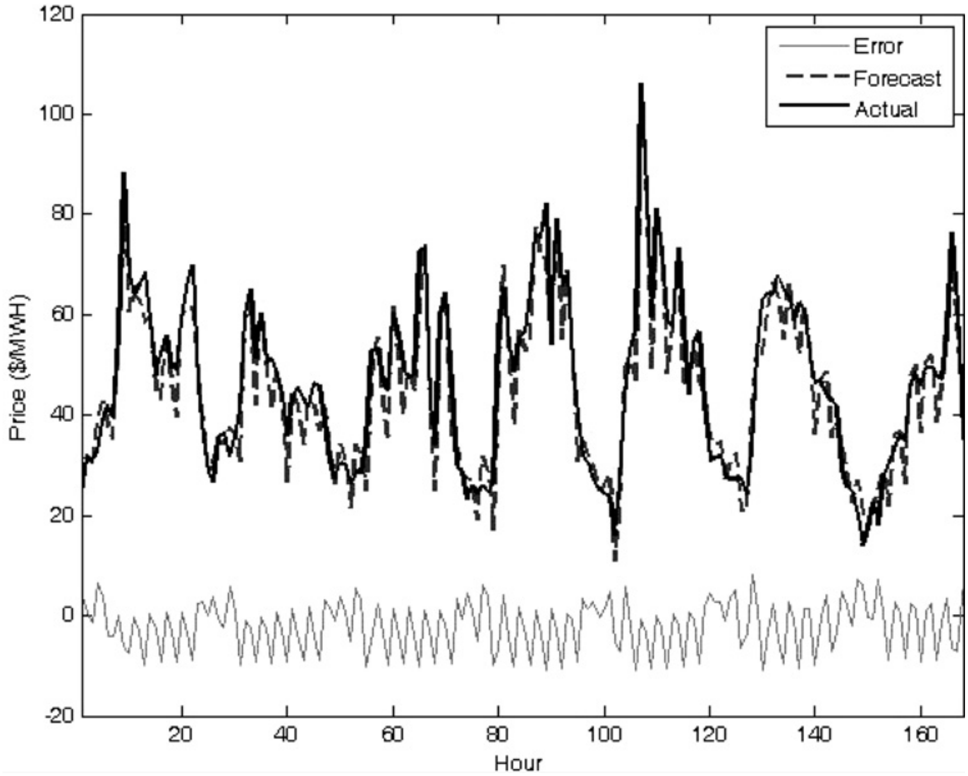


Figure 4 *Obtained results for the Ontario electricity market in 3–9 May 2004*

Table 4 Monthly MAPE (%) and MAE ($/MWh) for price forecast in the New England electricity market

Test month	ISO [13]		Single SVM [13]		Hybrid network [13]		Proposed	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
December 2002	5.53	12.66	4.87	11.31	4.26	10.24	4.22	8.32
March 2005	–	–	4.72	7.17	4.09	6.35	4.08	6.12
August 2005	–	–	8.74	8.85	7.70	7.91	6.71	6.83

Table 5 Monthly MAPE (%) for price forecast in the Italian and New England electricity markets

Test month	Dynamic price forecast [18]		Proposed	
	Italian market	New England	Italian market	New England
1	36.12	6.84	12.12	5.24
2	18.46	6.56	11.23	6.11
3	12.76	9.77	10.14	7.87
4	12.42	6.96	10.64	5.43
5	13.27	8.12	9.75	7.23
6	9.36	8.06	10.56	7.54
7	11.63	6.19	9.54	6.07
8	11.96	6.24	9.73	6.20
9	12.81	7.38	9.25	6.43
10	4.97	7.94	4.81	6.13
11	5.49	7.37	5.35	5.91

in the denominator instead of $P_{i\mathrm{ACT}}$ in (13). For the sake of a fair comparison, the same MAPE definition is also used in this experiment for the proposed method. Table 5 shows that the proposed method on both Italian and New England electricity markets has considerably lower MAPE values than the dynamic price forecast technique.

For the test cases of this paper, the whole running time of the proposed method including execution of the modified Relief, training of the HNN, parameter tuning by the cross-validation and price forecast of the next day is less than 30 min on a simple Pentium P4 personal computer with 512 MB RAM, which is reasonable within a day-ahead decision making framework.

To validate the obtained values for the k parameter from (10), sample price forecast results with the other values of k for the first test case are given in Table 6. In this table, 60%, 80%, 100%, 120% and 140% of the k value obtained from (10) are shown. As seen, the k value obtained from (10), that is 100% k, leads to the best results with the minimum MAPE and MAE values in Table 6.

The proposed HNN is also tested on the classroom problem given in [15] (Example 1 of this reference). The obtained results from the proposed HNN are shown in Table 7 and compared with the results of decoupled extended Kalman filter with U-D factorisation (DEKF-UD) and BP-Bayesian (a combined back propagation learning and Bayesian-based confidence interval method), proposed in [15] for electricity price forecast. The same test conditions of [15] are also considered for the HNN. As

Table 6 Weekly MAPE (%) and weekly MAE ($/MWh) for HOEP forecast in the Ontario electricity market with different values of k parameter

Test week	60% k		80% k		100% k		120% k		140% k	
	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE
26 April–2 May	10.18	7.4	9.73	6.7	9.46	6.6	9.76	6.8	10.52	7.6
3–9 May	10.27	7.5	9.81	7.8	9.72	7.3	10.11	8.1	10.29	7.9
26 July—1 August	9.67	7.1	9.13	6.9	9.00	6.7	9.37	7.0	9.71	7.5
2–8 August	9.64	7.8	9.42	7.4	9.04	7.1	9.55	7.1	9.61	7.7
13–19 December	9.74	9.4	9.57	9.2	9.36	8.7	9.51	9.4	9.79	9.2
20–26 December	9.19	9.5	8.87	9.5	8.80	9.3	8.91	9.5	9.20	9.6
average	9.78	8.1	9.42	7.9	9.23	7.6	9.53	8.0	9.85	8.3

doi: 10.1049/iet-gtd.2009.0297

Table 7 Obtained results for the classroom problem of [15]

	DEKF-UD [15]	BP-Bayesian [15]	Proposed HNN
training time (s)	0.7	2.3	1
prediction error (MAE)	0.029	0.037	0.028

seen, the training time of the HNN is about the DEKF-UD and BP-Bayesian, whereas the prediction error of the HNN is better than these methods.

5 Conclusion

In this paper, a new prediction method composed of modified Relief algorithm for feature selection and HNN as forecast engine is proposed for day-ahead price forecast of electricity markets. The proposed feature selection technique can better evaluate the non-linear dependencies of the price signal on its input variables than the previous linear feature selection techniques. The HNN is composed of NNs, EAs and auxiliary predictors in a cascaded structure with a new data flow among them. The HNN can learn much more than a single NN. The proposed method is examined on Ontario, New England and Italian electricity markets and compared with several recently published price forecast techniques. These comparisons reveal the price forecast capability of the proposed method.

6 References

[1] AMJADY N., HEMMATI M.: 'Energy price forecasting – problems and proposals for such predictions', *IEEE Power Energy Mag.*, 2006, **4**, (2), pp. 20–29

[2] SHAHIDEHPOUR M., YAMIN H., LI Z.: 'Market operations in electric power systems' (Wiley, 2002)

[3] ZAREIPOUR H., CANIZARES C.A., BHATTACHARYA K., THOMSON J.: 'Application of public-domain market information to forecast ontario's wholesale electricity prices', *IEEE Trans. Power Power Syst.*, 2006, **21**, (4), pp. 1707–1717

[4] ZHOU M., YAN Z., NI Y.X., LI G., NIE Y.: 'Electricity price forecasting with confidence-interval estimation through an extended ARIMA approach', *IEE Proc.-Gener. Transm. Distrib.*, 2006, **153**, (2), pp. 187–195

[5] GARCIA-MARTOS C., RODRIGUEZ J., SANCHEZ M.J.: 'Mixed models for short-run forecasting of electricity prices: application for the Spanish market', *IEEE Trans. Power Syst.*, 2007, **22**, (2), pp. 544–551

[6] GARCIA R.C., CONTRERAS J., AKKEREN M.V., GARCIA J.B.C.: 'A GARCH forecasting model to predict day-ahead electricity prices', *IEEE Trans. Power Syst.*, 2005, **20**, (2), pp. 867–874

[7] HONG Y.Y., HSIAO C.Y.: 'Locational marginal price forecasting in deregulated electricity markets using artificial intelligence', *IEE Proc.-Gener. Transm. Distrib.*, 2002, **149**, (5), pp. 621–626

[8] ZHANG L., LUH P.B., KASIVISWANATHAN K.: 'Energy clearing price prediction and confidence interval estimation with cascaded neural network', *IEEE Trans. Power Syst.*, 2003, **18**, (1), pp. 99–105

[9] GUO J.J., LUH P.B.: 'Improving market clearing price prediction by using a committee machine of neural networks', *IEEE Trans. Power Syst.*, 2004, **19**, (4), pp. 1867–1876

[10] GONZALEZ A.M., SAN ROQUE A.M., GONZALEZ J.G.: 'Modeling and forecasting electricity prices with input/output hidden Markov models', *IEEE Trans. Power Syst.*, 2005, **20**, (2), pp. 13–24

[11] AMJADY N.: 'Day-ahead price forecasting of electricity markets by a new fuzzy neural network', *IEEE Trans. Power Syst.*, 2006, **21**, (2), pp. 887–896

[12] RODRIGUEZ C.P., ANDERS G.J.: 'Energy price forecasting in the Ontario competitive power system market', *IEEE Trans. Power Syst.*, 2004, **19**, (1), pp. 366–374

[13] FAN S., MAO C., CHEN L.: 'Next-day electricity-price forecasting using a hybrid network', *IET Gener. Transm. Distrib.*, 2007, **1**, (1), pp. 176–182

[14] MANDAL P., SENJYU T., URASAKI N., FUNABASHI T., SRIVASTAVA A.K.: 'A novel approach to forecast electricity price for PJM using neural network and similar days method', *IEEE Trans. Power Syst.*, 2007, **22**, (4), pp. 2058–2065

[15] ZHANG L., LUH P.B.: 'Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method', *IEEE Trans. Power Syst.*, 2005, **20**, (1), pp. 59–66

[16] LORA A.T., SANTOS J.M.R., EXPOSITO A.G., RAMOS J.L.M., SANTOS J.C.R.: 'Electricity market price forecasting based on weighted nearest neighbors techniques', *IEEE Trans. Power Syst.*, 2007, **22**, (3), pp. 1294–1301

[17] LI G., LIU C.C., MATTSON C., LAWARREE J.: 'Day-ahead electricity price forecasting in a grid environment', *IEEE Trans. Power Syst.*, 2007, **22**, (1), pp. 266–274

[18] BOMPARD E., CIWEI G., NAPOLI R., TORELLI F.: 'Dynamic price forecast in a competitive electricity market', *IET Gener. Transm. Distrib.*, 2007, **1**, (5), pp. 776–783

[19] ZAO J.H., DONG Z.Y., LI X., WONG K.P.: 'A framework for electricity price spike analysis with advanced data mining

doi: 10.1049/iet-gtd.2009.0297

methods', *IEEE Trans. Power Syst.*, 2007, **22**, (1), pp. 376–385

[20] ZAO J.H., DONG Z.Y., LI X.: 'Electricity market price spike forecasting and decision making', *IET Gener. Transm. Distrib.*, 2007, **1**, (4), pp. 647–654

[21] SUN Y.J.: 'Iterative RELIEF for feature weighting: algorithms, theories, and applications', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007, **29**, (6), pp. 1035–1051

[22] DIETTERICH T.G.: 'Machine learning research: four current directions', *Artif. Intell. Mag.*, 1997, **18**, (4), pp. 97–136

[23] CARUANA R., FREITAG D.: 'Greedy attribute selection'. Proc. 1994 Int. Conf. Machine Learning, pp. 28–36

[24] SIKONJA M.R., KONONENKO I.: 'Theoretical and empirical analysis of RELIEFF and RRELIEFF', *Mach. Learning J.*, 2003, **53**, (1–2), pp. 23–69

[25] The MathWorks, MATLAB: [Online] Available at http://www.mathworks.com

[26] LARSEN R.J., MARX M.L.: 'An introduction to mathematical statistics and its applications' (Prentice-Hall, 2000, 3rd edn.)

[27] TSEKOURAS G.J., HATZIARGYRIOU N.D., DIALYNAS E.N.: 'An optimized adaptive neural network for annual midterm energy forecasting', *IEEE Trans. Power Syst.*, 2006, **21**, (1), pp. 385–391

[28] AMJADY N.: 'Introduction to intelligent systems' (Semnan University Press, Semnan, Iran, 2002)

[29] FAUSETT L.: 'Fundamentals of neural networks: architectures, algorithms, and applications' (Prentice-Hall, 1994)

[30] http://www.theimo.com/imoweb/marketdata/marketData.asp

[31] http://www.iso-ne.com/markets/hrly_data/index.html

[32] http://www.mercatoelettrico.org/En/Statistiche/ME/DatiStorici.aspx

doi: 10.1049/iet-gtd.2009.0297