Electric Power Systems Research 80 (2010) 318–327

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Electricity market price spike analysis by a hybrid data model and feature selection technique

Nima Amjady, Farshid Keynia*

Department of Electrical Engineering, Semnan University, Semnan, Iran

ARTICLE INFO

Article history:
Received 15 May 2009
Received in revised form 8 August 2009
Accepted 20 September 2009
Available online 23 October 2009

Keywords:
Price spike occurrence prediction
Hybrid data model
Feature selection technique

ABSTRACT

In a competitive electricity market, energy price forecasting is an important activity for both suppliers and consumers. For this reason, many techniques have been proposed to predict electricity market prices in the recent years. However, electricity price is a complex volatile signal owning many spikes. Most of electricity price forecast techniques focus on the normal price prediction, while price spike forecast is a different and more complex prediction process. Price spike forecasting has two main aspects: prediction of price spike occurrence and value. In this paper, a novel technique for price spike occurrence prediction is presented composed of a new hybrid data model, a novel feature selection technique and an efficient forecast engine. The hybrid data model includes both wavelet and time domain variables as well as calendar indicators, comprising a large candidate input set. The set is refined by the proposed feature selection technique evaluating both relevancy and redundancy of the candidate inputs. The forecast engine is a probabilistic neural network, which are fed by the selected candidate inputs of the feature selection technique and predict price spike occurrence. The efficiency of the whole proposed method for price spike occurrence forecasting is evaluated by means of real data from the Queensland and PJM electricity markets.

1. Introduction

With the introduction of restructuring into the electric power industry, the price of electricity has become the focus of all activities in the power market [1]. Optimal decisions now will be highly dependent on market electricity prices. On the other hand, electricity market has its own complexities due to differences from most 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, e.g., the price spikes of the PJM (Pennsylvania–New Jersey–Maryland) and California markets in 1999 and 2000, respectively [2,3].

The importance of electricity price forecasting on the one hand, and its complexity on the other hand, motivates many research works in the recent years. Stationary time series models such as Auto-Regressive Integrated Moving Average (ARIMA) [4] and non-stationary time series models like Generalized Auto-Regressive Conditional Heteroskedastic (GARCH) [5] have been proposed for this purpose. In [6], four price forecast models are presented and compared. Among these four models, transfer function time series model achieves the best price forecast accuracy. Some other research works proposed neural networks (NN) [7,8] and Fuzzy Neural Networks (FNN) [9,10] for electricity price forecast. Recently, different combinatorial price forecast techniques such as combination of fuzzy inference system (FIS) and least-squares estimation (LSE) [11], combination of similar days and NN techniques [12], wavelet transform + ARIMA [13] and wavelet transform + NN [14] with wavelet transform as preprocessor, Wavelet Neural Network (WNN) with wavelet transform as the activation function of the neural network [15], and combination of NN with evolutionary algorithm [3] have been proposed for the prediction of electricity price. Most of these methods can be effective for normal price forecast, but may encounter problems for price spike prediction [16]. In [1], the probability of occurrence of electricity price spikes is analyzed and it is concluded that the accurate forecast of price spikes is difficult. On the other hand, price spike forecast is important for both suppliers and consumers of an electricity market to construct their bidding strategies and risk management plans [17].

In general, price spike forecast has two main aspects: prediction of price spike occurrence and value. In price spike value prediction, the target feature is a continuous real-valued variable. On the other hand, the problem of price spike occurrence prediction is a binary classification task wherein the target feature is a binary-valued variable indicating whether price spike occurs or not. In

* Corresponding author. Tel.: +98 0341 2121976; fax: +98 0341 2121976.
E-mail address: Keynia@yahoo.com (F. Keynia).

0378-7796/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.epsr.2009.09.015

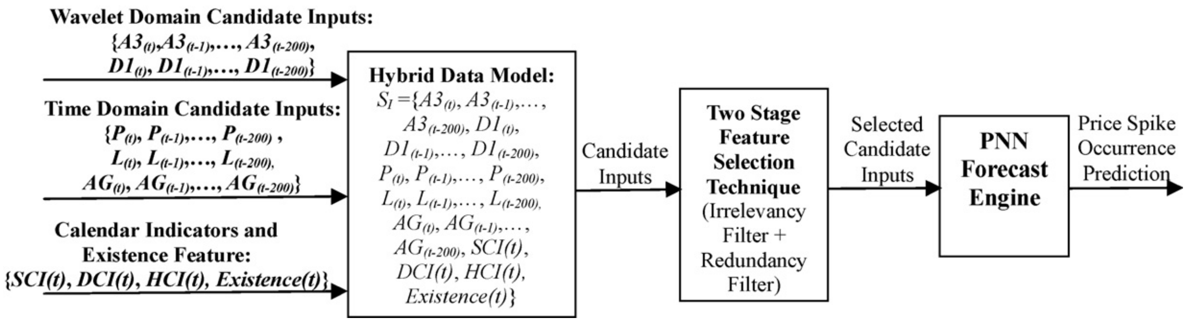

Fig. 1. Structure of the proposed price spike occurrence prediction strategy.

[1,3], a price spike value forecast method is presented where the price spikes are processed by a mathematical transformation. The mathematical transformation and its inverse are added as preprocessor and post-processor to the forecast method to soften the variations of price spikes, respectively. In [18], prediction of price spike occurrence based on forecasted normal price and prediction of price spike value based on the Naïve Bayesian classifier (NBC) and nearest neighboring method is proposed. However, this method cannot accurately forecast price spike occurrence. To remedy this problem, price spike occurrence forecasting by means of classification algorithms including Support Vector Machine (SVM) and probability classifier is proposed in [19]. This research work is continued in [16] where the effect of different kernels of SVM and price spike thresholds is also evaluated. In [17], the construction of bidding strategies considering price spike occurrence prediction is proposed. Application of different classification algorithms for price spike occurrence forecasting has been also evaluated in this reference. In [20,21], price spike occurrence forecasting by ELM (Extreme Learning Machine) and SVM is presented.

It is noted that electricity price time series feature extreme jumps of magnitudes rarely seen in financial markets, and occurring at greater frequency. For instance, the Australian power markets exhibit occasional price spikes several orders of magnitude greater than their mean price [22]. Characteristics of electricity price spikes, such as abrupt occurrence and fast reversion are discussed in [23]. Use of predictions of relevant exogenous drivers, such as the capacity surplus, is proposed in this reference for price spike forecast. In [24], price spikes are considered as the outliers of the price time series and different processing models to handle price spikes are introduced and discussed. In spite of the performed research works in the area, accurate price spike forecast cannot be achieved yet and further research work on price spike prediction is demanded.

Focus of this paper is on price spike occurrence prediction like [16,17,19–21], as the first step of price spike forecast, and a new strategy for this purpose is proposed. The remaining parts of the paper are organized as follows. In Section 2, at first the structure of the proposed strategy for price spike occurrence prediction and data flow among its building blocks are presented. Then, the major components of the proposed strategy are introduced in three separate subsections, respectively. In Section 3, obtained numerical results for price spike occurrence prediction are presented. Section 4 concludes the paper.

2. The proposed strategy for price spike occurrence prediction

The structure of proposed price spike occurrence prediction strategy is shown in Fig. 1. As seen, the proposed strategy is composed of three major components including a new hybrid data model, a novel two-stage feature selection technique and an efficient PNN forecast engine. The hybrid data model constructs the set of candidate inputs for the forecast process. This set is refined by the proposed two-stage feature selection technique to select a minimum subset of the most informative features. The PNN forecast engine is fed by the selected candidate inputs as shown in Fig. 1. The three major components of the proposed strategy are introduced in the next subsections, respectively.

2.1. The proposed hybrid data model

A price that is much higher than the normal price is defined as price spike [17]. So, at the first step, we need a threshold to discriminate price spikes from other prices such that prices higher than this threshold are considered as price spikes. The threshold of $\mu \pm 2\sigma$ in [16–18,20] is proposed for this purpose where μ and σ indicate mean and standard deviation of historical market prices. So, different electricity markets may have different thresholds [16]. In [18], it has been discussed that in the California market, the high price threshold value is US$ 50/MWh and a threshold value of $75/MWh is proposed for the Queensland electricity market. In [3], US$ 150 and 200/MWh are considered as price spike thresholds for the PJM electricity market. In [16], the selection of price spike thresholds is analyzed in more details for the Queensland electricity market. After determining the threshold, historical data related to price spikes can be gathered and spike forecast model can be constructed.

Previous research works on price spike prediction only use time domain information [1,3,16–21]. However, frequency domain can also contain useful information for the forecast process, especially considering that a price spike contains sudden changes in the time domain or equivalently high frequency components in the frequency domain. Fourier Transform (FT) gives the spectral contents of the signal, but it gives no information regarding wherein time those spectral components appear. The Short Time Fourier Transform (STFT) provides the time information by computing different FTs for consecutive time intervals, and putting them together. Consecutive time intervals of the signal are obtained by truncating the signal using a sliding windowing function [25]. STFT gives a fixed resolution at all times. Once the window is chosen, the resolution is set for both time and frequency. Wide analysis window gives poor time resolution and good frequency resolution and vice versa. On the other hand, electricity price series contain several nonstationary features such as trends, changes in level and slope, and seasonalities, to name a few. These features are often the important and challenging parts of the price signal and must be taken into account when dealing with nonstationarity. Hence, the price characteristics challenge the traditional Fourier analysis. Wavelet analysis overcomes the limitations of the Fourier methods by using functions that retain a useful compromise between time location and frequency information. Implicitly, wavelets have a window that automatically adapts itself to give the appropriate resolution [25]. An efficient algorithm to implement discrete