Renewable and Sustainable Energy Reviews 70 (2017) 330–341

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

A new electricity price prediction strategy using mutual information-based SVM-RFE classification

Zhen Shao[a,b,*], ShanLin Yang[a,b], Fei Gao[a], KaiLe Zhou[a], Peng Lin[a]

[a] School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China
[b] Key laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei, Anhui 230009, PR China

ARTICLE INFO

Keywords:
Pattern classification
Electricity market price prediction
SVM-RFE
Time series segmentation
Minimum redundancy maximum relevancy

ABSTRACT

Owing to the central role in electricity market operation, researchers have long sought to investigate the price responsiveness of both electricity supply and consumption sides. From the perspective of demand-side management (DSM), electricity prices prediction can be regarded as a pattern recognition problem of classifying future electricity prices with respect to a predefined threshold. From a fresh perspective this paper develops an efficient framework, called TSS-RFE-MRMR based SVM (Time series segmentation, recursive feature elimination, and minimum redundancy maximum relevance based support vector machine), for providing estimates of price fluctuation over certain valuation domains and modeling high-dimensional electricity market price without adopting additional impact factors. It starts from adopting a novel feature space determination scheme, called principal component analysis-dynamic programming (PCA-DP) based time series segmentation. Then, the RFE-MRMR filter for significant features selection is implemented, where both redundant and less relevant features are progressively eliminated among the potential feature sets. To test the performance of the proposed approach, it is evaluated on Ontario and New York electricity markets and compared with other method. Our experimental results indicate that the proposed approach outperforms other traditional method and present a relatively higher prediction accuracy on the electricity price.

1. Introduction

Over the past decades, with the advances of technology and relaxation of electricity industry regulation, an increasing number of participants join the market, thus promoting further diversification of consumer services. In the deregulated electricity market, the intense competition forces firms make quick and accurate decisions to ensure first mover advantage. Basically, electricity retailing firm stands as a bridge between wholesale side and the customers. Accordingly, in order to improve the competitiveness in electricity retailing, enterprises should concentrate more on confirming the energy procurement strategy and optimal selling price, respectively [1–3].

Nowadays, the competition among retailing firms in electricity market is increasingly dependent on precise evaluation of electricity price fluctuation and proper optimization of the pricing mechanism. Since it is beneficial for supply chain coordination in electric power industry and have become fundamental to electricity market operations, future electricity price estimation have attracted significant attentions among both enterprises and academia. Owing to the central role in electricity market operation, researchers have long sought to investigate the price responsiveness from both electricity supply and consumption side using data at different time scales. Consequently, numerous efforts, dealing with the electricity market price modeling and forecasting, have appeared in the literature [4–7]. Typically electricity market price has been found to be influenced by a large amount of internal and external factors [8–10], for instance, bidding strategies, marginal costs, scheduled and unscheduled generation outages, extreme and unusual weather patterns, and so on. Compared with the electricity demand prediction (especially short term electricity demand), the errors reported in the existing literatures regarding electricity price prediction are higher and more unstable [11,12], which indicates that the fluctuation features of electricity price are hard to detect and capture.

In spite of the above mentioned difficulties, there are a wide variety of approaches that have been successfully utilized to predict the electricity price. Among them, machine learning method [13–15] and statistical time series models [16–18] are the most commonly used approaches. Recently, the application of machine learning approach has gained more attention when relating to other approaches (e.g. traditional econometric approaches), because machine learning ap-

* Corresponding author at: School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China.
E-mail address: shaozhen@hfut.edu.cn (Z. Shao).

http://dx.doi.org/10.1016/j.rser.2016.11.155
Received 18 November 2015; Received in revised form 25 September 2016; Accepted 12 November 2016

proach has been shown to be robust with respect to the unstable variation of price fluctuation. Specifically, in view of recent theoretical developments concerning the volatility pattern recognition and demand prediction, many machine learning techniques such as neural network [19,20], SVM [21,22], extreme learning machine (ELM) [23], adaptive network-based fuzzy inference system (ANFIS) [24] have been presented and discussed under the framework of conventional point prediction. However, simple point prediction is insufficient to measure the inherent uncertainties existing in demand fluctuation and make reasonable judgment of its future trends [25–27]. Thus, to overcome the shortcomings mentioned above, several novel probability density forecast methods are adopted for the price prediction. HY Yamin et al. [28] discussed the probability distribution prediction of electricity price using artificial neural network (ANN). Based on the Monte Carlo simulation, they conducted probability analysis for prices at each hour, respectively. Zhao et al. [29] employed SVM to forecast the point value of electricity price, then a heteroscedastic variance equation (GARCH) based SVM method was proposed to generate the prediction intervals. Similar experiments were conducted in literature [30,31], they implemented a machine learning technique (ELM and ANN, respectively) incorporating the bootstrapping method for the uncertainty quantification. To reduce the arithmetic complexity of functional principal component analysis, Wu et al. [32] applied a recursive dynamic factor analysis algorithm where principal component scores and forecasting interval were predicted recursively using the Kalman filter. Wan et al. [33] introduced a two-stage formulation which combined extreme learning machine and maximum likelihood method for the probabilistic electricity price prediction. Nowotarski and Weron [34] investigated the application of quantile regression averaging (QRA) in the area of electricity spot price forecasts. They confirm that the QRA based prediction is found to be more accurate than the typical individual model. Similarly, Maciejowska et al. [35] utilized the principal component analysis to extend the quantile regression averaging (QRA) approach. The core topic in their research is to automate the process of selecting from a large set of individual forecasting models. To overcome the deficiency of neural networks based approaches in generating prediction intervals, Shrivastava et al. [36] proposed a novel method to directly generate the lower and upper bounds of the future electricity prices using SVM model. Bello et al. [37] developed a nested combination approach for the probabilistic forecast of medium-term hourly electricity price, where this approach consists of two modeling stages, including generating multiple scenarios of uncertain variables and executing a market equilibrium model. Based on the active learning technique and the variational heteroscedastic Gaussian process, Kou et al. [38] established a variational heteroscedastic Gaussian process (VHGP) model for the probabilistic electricity price forecasting. To summarize, traditional way of electricity price prediction for the probability interval generation is to consider the analysis as a statistical regression process. From this perspective, over longer time scales, it is still necessary to evaluate the feasibility of the distributional and parameter assumptions for the presented method.

It is worth noting that, in some real applications (such as DSM), certain price thresholds are sufficient to provide technical support and strategic reference for the decision-making. In this sense, future demand classification, a very different approach which offers measurable price thresholds to the electricity price analysis has been advocated in the literature [11,38–41]. This perspective follows traditional way of pattern recognition to identify the category of the price attributes. The major problem then becomes how historical data set can be handled under the classical classification framework. In recent years, with the maturing both in theory and research, machine learning based approach have been widely used for pattern recognition analysis. Among them, SVM has been recommended not only for the continuous time series modeling but also for pattern classification in the area of electricity market analysis, as well as in many other areas of complex pattern recognition. Despite its superiority, as mentioned before, when the input samples exhibit high dimensional and highly correlated characteristics, it is necessary to identify the maximally relevant and minimally redundant feature sets for the modeling.

Due to the above reasons, practical operation of feature selection has become a top issue in the field of classification application. Currently, many studies attempt to fill these gaps through two different perspectives: (1) Filter method [42] is to investigate the intrinsic characteristics of data sets with respect to the relevant class labels, which is fast and simple but the feature selection process is isolated from the classifier. (2) Wrapper method [43] is to combine the classifier to explore the feature subset, which is computationally expensive but gains higher classification accuracy than the filter method. Motivated by the above, the main purpose of the paper is to provide a complete and practical framework of electricity price prediction from a fresh perspective. By developing a novel framework for combining SVM recursive feature elimination (SVM-RFE), minimum redundancy maximum relevance algorithm, and time series segmentation method, this paper makes two key contributions: (1) In this paper, we concern primarily with the pattern classification method and attempt to propose a novel classification based approach to construct prediction interval of electricity price. In our implementation, the feature size (corresponding to the approximate optimal lag order of the price series) for the classification is defined by a fluctuation trend segmentation approach. After obtaining well-defined feature size, it is more likely to gain a deeper insight into the structure changes throughout a whole time period and construct a proper classifier for recognizing specific pattern of price fluctuation. (2) To achieve better accuracy and computational efficiency, we propose a RFE-MRMR based feature selection strategy for the multiclass classification. This strategy is designed to eliminate the redundancy and irrelevance features by integrating RFE and MRMR method into the SVM classifier. Combining the proposed framework with identified feature sets, it is convenient to generate the recognition results to obtain prediction intervals of the electricity market price.

The rest of the paper is organized as follows. Section 2 proposes a detailed analysis of the TSS-MRMR based SVM-RFE and describes the architecture of the proposed prediction framework. Section 3 presents the experimental results and discussions, and Section 4 concludes the paper.

2. Proposed methodology

The primary task of price regulation is to quantify the future volatility of price series, which is also a fundamental part of electricity market analysis. In this study, we have chosen the perspective of pattern classification to provide fresh insights for electricity price prediction. One of the underlying premises in price classification research is that efficient preprocessing steps of feature elimination and selection will assist classifier in enhancing the modeling accuracy and robustness.

In this section, we present a generic framework for the purpose of modeling and analyzing electricity market price. The theoretical foundation for generating future prediction interval of electricity price can be divided into two parts: Firstly, the basic concept of support vector machine recursive feature elimination (SVM-RFE) is introduced. Secondly, we develop an improved classifier that is based on the integration of SVM-RFE and minimum-redundancy maximum-relevancy (MRMR) algorithm. On this basis, we then sought to provide a novel framework (SVM-RFE with TSS-MRMR filter) for the electricity price prediction, which combines SVM-RFE-MRMR with a sample space identification method, namely time series segmentation (TSS).

2.1. Support vector machine recursive feature elimination

A great deal of previous study on pattern recognition has focused on the supervised classification method (e.g., SVM, Decision tree, and