IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 6, Issue 1, Feb - Mar, 2018
ISSN: 2320 – 8791 (Impact Factor: 2.317)
www.ijreat.org

Application Of Relevance Vector Machine For Classification Of Electricity Prices

Gomathi Rajendran[1], Ganapathy Somaskandan[2]

[1,2]Department of Electrical Engineering, Annamalai University, Chidambaram, Tamil Nadu 608002, India

Abstract
Recently, machine learning techniques have become popular and broadly used in the area of deregulated electricity market. In this paper, a more efficient Relevance Vector Machine (RVM) based classification for prices of electricity market has been presented. The proposed RVM classifier is used to classify the prices as low class and high class prices based on a threshold value. A feature selection algorithm combined with RVM classifier is proposed to reduce the large dimension of input variables into small size thus validating the potential of the whole data representation. The proposed approach has been tested on an Indian Energy Exchange. From the analysis, it has been observed that the proposed classifier provides very high class accuracy with minimum number of variables hence reducing complexity of data analysis and obtaining quicker results.

Keywords: *Deregulated Market, Electricity price Classification, Relevance Vector Machine, Sequence Feature Selection, Support Vector Machine*

1. Introduction

Electricity price forecasting has become increasingly important with the evolvement of the electric power industry into an era of market economy due to deregulation and free competition [1]. The main objective of deregulation is to reduce the cost of electricity through competition. Besides playing a key role in maximizing the benefits, it is also essential to classify the prices of electricity market. In the new competitive environment, most of the decisions made by producers or consumers are dependent on the electricity market [2]. So, electricity price forecasting has become an essential tool in competitive electricity market both for producers and consumers [3]. Electricity price forecasting is used for

various purposes, such as speculation, derivative pricing, risk management and real option valuation. With the accurate price forecasting, the power suppliers can build their bidding strategies to maximise their payoff and achieve the maximum benefit and on the other hand, consumers can minimise their utilisation cost [4]. Accurate forecast of the market prices is an important input to the decision making activities of a generating company for producing energy. Based on the literature survey, electricity price forecasting errors generally range from approximately 5% to 36% and vary based on the technique used and the market analyzed [5]. Electricity price classification used in place of price forecasting, has become a significant study in power system planning in recent years. Price classification is useful as the exact value of future prices is not critically important.

In the last few years, different techniques have been proposed to classify electricity prices. Feed forward Neural Network (FFNN) [6], Cascade-Forward Neural Network (CFNN), Generalized Regression Neural Network (GRNN) [7], Discrete Cosine Transform with Neural Network (DCT-NN) [5], Wavelet Transform, and Particle Swarm Optimization (PSO-FFNN) [8] have been used for classification problems. As the above techniques take long computational time and are infeasible for real time applications, Machine learning based classification has been suggested for classification problems. A very successful approach over supervised learning is the Support Vector Machine (SVM) [9]. SVM, which is based on statistical learning, has also been successfully applied in highly linear and nonlinear feature space. Results have reported that, the number of support vectors grows linearly along with the increase of training samples size, which entails for extra calculation of setting of parameters. On the basis of SVM based price market and evaluation, Core Vector Machine (CVM) [10], Informative Vector Machine (IVM) [11] and Extreme Learning Machine (ELM) [12] have been suggested. Nevertheless, these Machine learning algorithms have some drawbacks due to absence of Probabilistic Interpretation. To overcome the above drawbacks, a Probabilistic Bayesian framework termed as Relevance Vector Machine (RVM) has been proposed for price classification.

This paper presents the suitability of RVM based classifier, by improving its performance using feature selection and feature subset selection process. RVM classifier can classify the prices, within a very short computational time with much better adoptability. The proposed method is trained and tested for an Indian market. The main contribution of the paper is price classification that could be realized in an Indian energy exchange using RVM.

2. RVM Classifier

RVM has been introduced by Tipping as a Bayesian treatment alternative to the SVM. The RVM introduces a prior over the model weights governed by a set of hyper parameters, in a probabilistic framework [13]. One hyper parameter is associated with each weight, and the most probable values are iteratively estimated from the training data. The most compelling feature of the RVM is that it typically utilizes significantly fewer kernel functions compared to the SVM, while providing a similar performance.

Given a training set of input-target pairs $D = \{(x_i, y_i)\}_{i=1}^{N}$, RVM follows the standard probabilistic formulation and assumes that the targets are samples from the model with additive noise:

$$t_i = y(x_i, w) + \varepsilon_i \quad (1)$$

where ε_i are error terms which are generally assumed to be independent identically distributed Gaussian variables with mean zeros and variance σ^2.
The likelihood function can be written as:

$$P(t|\sigma^2) = (2\Pi^2)^{-1/2} \exp\{-\frac{1}{2\sigma^2} ||t - \phi w||^2\} \quad (2)$$

Where ϕ is the design matrix with $\phi = [\phi(x_1),....., \phi(x_l)]^T$, in which $\phi(x_i) = [k(x_i, x_1),.....,(x_i, x_1),l]^T$ and k is a kernel. Maximum likelihood estimation of w and σ^2 will lead to severe over fitting, so RVM encodes a preference for smoother functions by defining an automatic relevance determination Gaussian prior over the weights:

$$P(w|\alpha) = \prod_{i=1}^{l+1} N(w_i | 0, \alpha_i^{-1}) \quad (3)$$

The posterior over the weights is then obtained from Bayesian rule:

$$P(w|t,\alpha,\sigma^2)=(2\Pi)^{-(l+1)/2}|\Sigma|^{-(1/2)}\exp\{-\frac{1}{2}(w-\mu)^T\Sigma^{-1}(w-\mu)\} \quad (4)$$

Where $\Sigma=(\phi^T B\phi+A)^{-1}$, $\mu=\Sigma\phi^T Bt$, and

$A = \mathrm{diag}(\alpha_1,....,\alpha_{l+1}), B = \sigma^{-2} I_l$

By integrating the weights of the product: $p(t|w,\sigma^2)p(w|\alpha)$, RVM obtains the marginal likelihood for the hyper-parameters:

$$P(t|\alpha,\sigma^2)=(2\mu)^{-1/2}|B^{-1}+\phi A^{-1}\phi^T|^{-1/2}\exp-\frac{1}{2}t^T(B^{-1}+\phi A^{-1}\phi^T)^{-1}t\} \quad (5)$$

Because the values of α and σ^2 that maximize the function defined in equation (5) cannot be obtained in closed form [14], RVM considers an alternative formula for iterative re-estimation of α and σ^2:

$$\alpha_i^{new} = \frac{\gamma_i}{\mu_i^2}, (\sigma^2)^{new} = ||t - \phi\mu||^2/(l - \textstyle\sum_i \gamma_i) \quad (6)$$

During re-estimation, many of the α_l approach infinity, and the corresponding weights approach zeros, implying that the corresponding kernel functions can be 'pruned'. The other α_i finally stabilizes at some finite numbers. The corresponding x_i's are called Relevance Vectors (RV). The advantage of RVM is that it needs no extra parameter setting and is more convenient. RVM is a machine learning technique that uses Bayesian inference to obtain parsimonious solutions for probabilistic classification. It is actually equivalent to a Gaussian process model with covariance function:

$$K(x,x') = \sum_{j=1}^{N} \frac{1}{\alpha_j} \varphi(x', x_j) \quad (7)$$

Where φ is the kernel function (usually Gaussian), and $x_1,...x_N$ are the input vectors of the training set. The term training vectors associated with the remaining non-zero weights 'relevance' vectors, in difference to the principle of automatic relevance determination. The optimization process typically continues until the maximum change in α_i values is below a certain threshold or the maximum number of iteration is reached [15]. Finally, the model only depends on the subset of RV of non-zero elements.

Thus, the weights are taken and the non-zero weights are considered as relevance vectors which are strongly bonded and they are taken as low class prices. Then, the remaining irrelevant vectors are weekly bonded and they are taken as high class prices. In this paper, the Bayesian principle based RVMs have been employed in the prices of electricity.

The most important evaluation of the RVM performance is the accuracy which is defined as the ratio of number of prices correctly classified to the total no of prices in the dataset,

$$\text{Accuracy (\%)} = \frac{\text{Number of Prices correctly classified}}{\text{Total number of prices in the data set}} \times 100 \quad (8)$$

$$\text{Misclassification Rate (\%)} = \frac{\text{Number of Prices incorrectly classified}}{\text{Total number of prices in the data set}} \times 100 \quad (9)$$

2.1 Application of RVM to Electricity Price Classification

RVM is a relatively pioneering method for learning separating functions in classification problems. The success of RVM classification relies on a good feature set thus, gathering information, about the behavior of the system obtained from off-line studies. These simulations need two data sets, the training and testing data sets. The training data set is used to originate the price evaluation structure and the testing data is used for validating the developed arrangement. The data handling process is explained below.

2.2 Generation of Data set

A training set consists of input vector and target vector and is combined together with learning methods to train a knowledge data base. This set must adequately represent the entire data. The whole data set consists of 366 days, with prices for 24 hours a day in an Indian market of 2016 which is a leap year. The prices have been classified into two patterns, namely, low and high class prices i.e., low class is considered as correctly classified and high class is considered as misclassified. The data set is generated by considering 80% as training data and 20% as testing data. For a fast evaluation, not all the data are considered as inputs, instead the most effective variables are considered as inputs by using a suitable feature selection process.

In this paper, the Indian energy price is taken from an Indian market [16]. The historical prices P_i have been used as inputs. Here, P_i indicates the price at the i^{th} hour on daily basis. The prices are classified based on a threshold value. The value of the threshold is based on the annual average of the prices in India for the year 2016. The annual average of the price threshold value is taken based on the average of the lowest and highest price of the market. The lowest price for the year 2016 is 712.29 Rs/MWh and highest price is 5249.82 Rs/MWh. By evaluating each price, it is classified as either low or high price. The class distribution based on price threshold is

Class 1: Prices between T_1 and T_2 = low prices
Class 2: Prices between T_2 and T_3 = high prices

Three classification threshold are considered for the market: $T_1 = 0$, $T_2 = 2400.3$, $T_3 = 3000$ with all in Rs/MWh, where T_1, T_2, T_3 represent price floors, annual average price and price cap of the prices respectively.

2.3 Feature Selection

The objective of feature selection is to select a subset from the original feature space which is more informative to target classes in executing machine learning tasks but to ignore the irrelevant and redundant features [17]. Feature selection methods reduce the dimensionality of feature space by removing redundant, irrelevant or noisy data. It brings the immediate effects for application: speeding up a data mining algorithm and increasing the comprehensibility of the mining results [18]. Features may be selected by engineering judgement. But, such selection will be subjective with the possibility of important variables getting rejected. In this paper, Sequential Feature Selection (SFS) approach, a wrapper model, which seeks an optimal linear separation for two classes of data, has been used.

SFS algorithm searches for an efficient subset of features by aggregating the best features. This approach has progressed from one directional search of sequential selection. In one directional search method, once a feature is selected in SFS, there is no way to discard this feature [19]. SFS algorithm is a bottom up search procedure, which starts from an empty feature set S and gradually adds features, selected by some evaluation function that minimizes the Mean Square Error (MSE). At each iteration, features are added. So the new extended feature set should produce a minimum classification error with the addition of any other feature.

2.4 Algorithm [19]

1. Input Y = { $y_1,....y_d$}

The SFS algorithm takes the whole d- dimensional feature set as input

2. 2. Output = { x_j | j=1,2,....k; x_j ε Y} , where k = (0,1,....d)

SFS returns a subset of features ; the number of selected features k ,where k <d has to be specified a priori

3. Initialization :

Initialize the algorithm with an empty set θ ("null set"). So that k=0 (where k is the size of the subset)
Step 1 (inclusion):
$x^+ = \arg\max J(x_k + x)$, where x ε $Y-x_k$
$x_{k+1} = x_k + x^+$
k = k+ 1
Goto step 1

In this step, the feature is added, i.e., x^+ to the feature subset x_k , x^+ is the feature that maximizes our criterion function , that is, the feature that is associated with the best classifier performance if it is added to x_k.
The procedure is repeated until the termination criterion is satisfied

4. Termination: k=p

ηthe feature subset of size k contains the number of desired features p that specified a priori.

2.5 Use of Kernel Function

Having realized the concept of feature selection and classification, in the previous sections, the preceding step is to improve the performance of RVM classifier with the choice of kernel functions. In many classifications, computational problems occur due to the large vectors and the danger of over fitting may happen due to the high dimensionality. The latter problem is solved with application of the maximal margin classifier and so called kernel gives a solution. The kernel function used in RVM is polynomial which is presented in equation (10)

$$\text{Polynomial } k = (x^t(\eta^* x_2)^T + 1)^p \quad (10)$$

Where x denotes input patterns, x_i denotes support vectors, p is degree of the polynomial, η is input scale parameter. Despite its simplicity, RVM greatly reduces the time and space complexities compared with SVM algorithm.

3. Results and Discussion

The proposed RVM based classification approach to electricity prices is performed with an Indian utility system .The required data set for training and testing phases is processed using Tanagra software [20]. Market Clearing Price (MCP) is forecasted depending on the supply and demand of electricity in the power market. The data set is obtained by an hourly basis of whole year data set of 2016. The total system variables are 24 variables initially, indicating the prices of each hour in a day.

It can be seen that for 366 days prices, 217 are found to be low class prices and 149 are found to be high class prices. The training and testing samples are split in ratio of 80% (293 days) for training phase and 20% (73 days) for testing phase as shown in Table 1. Table 2 shows the optimal set of features selected using SFS process. Among all these variables, effective input features have been selected by using sequential based feature ranking

Performance evaluation	Without feature selection		With feature selection	
	Training	Testing	Training	Testing
Accuracy (%)	96.21 (285/293)	96 (68/73)	96.90 (287/293)	100 (73/73)
Misclassi-fication rate (%)	3.78 (3/293)	4 (5/73)	3.09 (5/293)	0 (0/73)

analysis. By having a threshold value T=2400.3, 16 featured inputs have been obtained.

Table1: Data Set for Training and Testing phases

Days	Over all	Training	Testing
Total no of days	366	293	73
Low prices	217	174	43
High prices	149	119	30

Table 2: Feature Selection Process

Table 3 gives the kernel parameters of RVM. It is clear from table 3, that the training accuracy of IVM classifier is improved by choosing optimal kernel parameters. In order to justify this training result, it is to be noted that any good classifier desires to ensure higher value of accuracy and less error rate forever.

Table 3: Choice of Kernel Parameters

Model (kernel)	Alpha max	epsilon	Training accuracy of IVM (%)
Polynomial	e^{-10}	e^{4}	96 (285/293)

Table 4 gives the performance evaluation using RVM classifier with and without feature selection. From Table 4, it is evident that the performance of RVM classifier is improved with selection of good feature set and elimination of redundant data. During testing phase, an overall efficiency of 96 % has been achieved with all possible variables for an Indian market, whereas maximum efficiency of 100% has been achieved with Feature Selection. With the implementation of sequential feature selection, the size of input features has reduced from 24 features to 16 features. Hence, it can be concluded that the electricity prices with feature selection is sufficient for obtaining an efficient classification with maximum accuracy and minimum misclassification rate.

Table 4: Performance Evaluation using RVM classifier

Table 5 shows the classification of electricity prices using RVM classifier with the selected features. The proposed RVM classifier is found to be very superior in classification in terms of higher classification accuracy and less misclassifications as compared to SVM classifier.

Table 5: Comparative results of Electricity Price classification with Feature

Case study	Indian market
No. of pattern variables	24
No. of features selected	16
Dimensionality reduction	33%
Selected features	P_1, P_6, P_5, P_9, P_{14}, P_7, P_{12}, P_{15}, P_{16}, P_{13}, P_{20}, P_{21}, P_{24}, P_{23}, P_{17}, P_{19}

Selection (Testing Phase)

Performance evaluation	SVM	RVM
Accuracy (%)	94.44	100
Misclassification rate (%)	5.55	0
Time in Seconds	1.2	0.3

4. Conclusion

The application of RVM algorithm in classification of electricity prices has contributed the predictions of the large data set. The basic implementation and modification of results in electricity prices are provided with feature selection algorithm. These algorithms have been tested with an Indian market and the results have been analyzed. Results of the Indian market prove that the classification accuracy of the proposed model is 100. The selected Gaussian kernel has identified the resemblance of distinctive input and output of the data set quickly. This has helped to enhance the representativeness of low and high class data in a certain data space so as to improve performance. RVM with feature selection proves to be an optimal approach for classification of prices in electricity market as it provides acceptable classification accuracy with a minimum number of uncorrelated features. This method has proved to be more suitable for large power market so as to significantly improve classification accuracy as compared to other classifiers.

References

1. F. Shu, J.R. Liao, K. Kauhiro, C. Luonan, "An Integrated Machine Learning Model for Day-Ahead Electricity Price Forecasting", In the Proceedings of Power Systems Conference and Explosition, PSCE '06, Atlanta, USA, 2006, pp.1643-1649.
2. N. Amjady, D. Ali, "Day-Ahead Electricity Price Forecasting using the Relief Algorithm and Neural Networks", In the Proceedings of Electricity Market, Fifth International Conference on European, Lisboa, Portugal, 2008, pp. 1-7.
3. W. Sun, J. C. Lu, M. Meng, "Application of Time series based SVM model on Next-Day electricity Price Forecasting under deregulated Power Market", Proceedings of the Fifth IEEE International Conference on Machine Learning and Cybernetics, Dalian, China, 2006, pp. 2373-2378.
4. E.E. Elatter, "Day-Ahead Price Forecasting of Electricity Markets based on Local Informative Vector Machine", IET Generation, Transmission and Distribution, vol. 7, no. 10, 2013, pp. 1063-1071.
5. S. Anbazhagan, N. Kumarappan, "Day-Ahead Deregulated Electricity Market Prices Classification using Neural Network Input Featured by DCT", Electric Power and Energy Systems, vol. 37, no. 1, 2012, pp. 103-109.
6. D. Huang, H. Zareipour, W.D. Rosehart, N.Amjady, "Data Mining for Electricity Price Classification and the Application to Demand-Side Management", IEEE Transactions on Smart Grid, vol. 3, no. 2, 2012, pp. 808-817.
7. S.Anbazhagan, N. Kumarappan, "A Neural Network approach to Day-Ahead Deregulated Electricity Market Prices Classification", Electric Power Systems Research, vol. 86, 2012, pp. 140-150.
8. A. Yadav, P. Rajagopal, and K. Niranjan, "Electricity Price Forecasting and Classification through Wavelet-Dynamic Weighted PSO-FFNN Approach", IEEE Systems Journal, vol. 47, no. 99, 2017, pp. 1-10.
9. C. Gao, E. Bomparb, R. Napoli, H. Zhenz, "Price Forecast in the Competitive Electricity Market by Support Vector Maachine", Physica A: Statistical Mechanics and its Applications, vol. 382, no. 1, 2007, pp. 98-113.
10. H. Adaminejad, I. Shayegani, M. Ohammadi, E. Farjah, "An Algorithm for Power Quality Events Core Vector Machine based Classification", Modares Journal of Electrical Engineering, vol. 12, no. 4, 2013, pp. 50-59.
11. E.E. Elatter, "Day-Ahead Price Forecasting of Electricity Markets based on Local Informative Vector Machine" , IET Generation, Transmission and Distribution, vol. 7,no. 10,2012,pp. 1063-1071.
12. N.A. Shrivastava, B.K. Panigrahi, M.H. Lim, "Electricity Price Classification using Extreme Learning Machine", Neural Computing and Applications, vol. 27, no. 1, 2016, pp. 9-18.
13. M.E.Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine", Journal of Machine Learning Research, vol. 1, 2001, pp. 211-244.
14. M.E.Tipping, "The Relevance Vector Machine", Advances in Neural Information Processing Systems, vol. 12, 2000, pp. 652-658.
15. S. Ghosh, P.P. Mujumdar, "Statistical Downscaling of GCM Simulations to Stream Flow using Relevance Vector Machine", Advances in Water Resources, vol. 31, 2008, pp. 132-146.
16. EP. Smitha, I. Jacob Raglend, "Sequential Wavelet-ANN with Embedded ANN-PSO Hybrid Electricity Price Forecasting for Indian Energy Exchange", Neural Computing and Applications,vol. 28, no. 8, 2017, pp. 2277-2292.
17. C. Te-Shun, K.K. Yen, J. Luo, "Correlation-based Feature Selection for Intrusion Detection Design", In the Proceedings of the 2007 Military Communication Conference (MILCOM), Orlando, USA, 2007, vol. 5, no. 1, pp. 1-7.
18. Y. Pinar, "Filter based Feature Selection Methods for Prediction of Risks in Hepatitis Disease", International Journal of Machine Learning and Computing, vol. 5, no. 4, 2015, pp. 258-263.
19. F. Shirbani, H. Soltanian Zadeh, "Fast SFFS-based Algorithm for Feature Selection in Biomedical Data Sets", Amirkabir International

Journal of Scientific Research, vol. 45, no. 2, 2013, pp. 43-56.

20. Ricco Rakotomalala., Tanagra: a free for research and academic purposes, In the Proceedings of the 2005 European Grid Conference, Amsterdam, Netherlands, 2005, vol. 2, pp. 697-702.