Big Data Analytics for Price Forecasting in Smart Grids

Kun Wang*, Chenhan Xu*, and Song Guo†
*School of Internet of Things, Nanjing University of Posts and Telecommunications,
Nanjing, 210003, China.
†School of Computer Science and Engineering, The University of Aizu,
Aizu-Wakamatsu City, Fukushima, 965-8580, Japan.
Email: *kwang@njupt.edu.cn, *xchank@outlook.com, †sguo@u-aizu.ac.jp

***Abstract*—Demand side management (DSM) is a key mechanism to make smart grids cost efficient using electricity price forecasting issue. Price forecasting method takes the big price data into account, and gives estimates of the future electricity price. However, most of existing price forecasting methods cannot avoid redundancy at feature selection and lack of an integrated framework that coordinates the steps in forecasting. To address this issue, we first propose a new electricity price forecasting framework. It is significant to design a system tool chain based on big data analytics for ensuring that the users can make appropriate decisions. To this end, three algorithms are proposed integratedly. First, feature redundancy elimination is implemented by the fusion of Grey Correlation Analysis (GCA) and ReliefF algorithm. Second, a combination of Kernel function and Principle Component Analysis (KPCA) is designed to achieve dimensionality reduction. Finally, Support Vector Machine (SVM) optimization algorithm based on differential evolution (DE) is proposed to forecast price classification. These three modules jointly power the price forecasting system. Simulation results show the superiority of our proposed framework.**

***Index Terms*—Demand side management; Price forecasting; Big data; Classification; Feature selection; Dimensionality reduction; Smart girds.**

I. Introduction

Demand side management (DSM) is to encourage the consumer to use less energy during peak hours, or to move the time of energy use to off-peak periods, which may make smart grids cost-effective [1]. It helps to alleviate price fluctuating, decrease peak demand, and defer the need for network upgrade [1]. In 2007, DSM helps to save more than 4000 MW peak load in the RFC region, and more than 3000 MW in the MRO region [2]. PJM reported that DSM reduced wholesale energy prices by nearly $400 per MWh during the peak load hours, i.e., $230 million are saved in this time period [2]. DSM enables customers to participate in the operations of modern electricity energy system, where they can have the choice to decrease their energy costs through energy conservation and load shifting. In this way, customers may switch loads on or off relying on the dynamic electricity pricing [3].

Although exact point price forecasting is desired [3], customers are actually satisfied by knowing whether the electricity price surpasses the thresholds, depending on which they can decide to turn on or off the load. In this case, electricity price classification based on the big price data analytic needs to be implemented. One way to implement price classification is to categorize the prices as per the specified thresholds based on the values estimated by point price forecasting methods. Point price forecasting models are built using function approximation techniques, wherein the underlying process of price formation is mimicked by a price model [3]. What's more, forecasting classes requires lower accuracy. In comparison, predicting accurate point price is more challenging due to various influencing factors. Therefore, electricity price classification, especially how to deal with the huge amount of price data, becomes a significant focus in the field of DSM.

To address this issue, recent studies have tried some traditional classifiers, in which Naive Bayes, Neural Network, Decision Tree and Support Vector Machine (SVM) are the most popular ones [4]. According to the survey given in [5], SVM is a classifier with better accuracy but higher complexity. The high dimensionality of DSM features which consist irrelevant and redundant features is a severe challenge to these classifiers, it brings extremely high computational complexity to the training and forecasting processes, and decrease the stability of forecasting. Given a specific forecasting model, training the forecasting model with irrelevant features can greatly degrade the forecasting accuracy, and also decrease the training speed to an intractable level; moreover, when there are too many available features, manually determining the relevant features of the response variable with statistical plots will become impossible [6]. Hence, feature selection is a necessary operation for classifier, which selects the relevant factors of a response variable. ReliefF is a classic algorithm for feature selection, which is improved from simple Relief [6], [7].

However, a common problem in most feature selectors is that redundancy among selected features is not eliminated during selecting process, which means that the selected features may not be representative. Principle Components Analysis (PCA) is a common solution for reducing dimensionality, while PCA is only suitable in dealing with a set of observations that vary linearly [8], but not nonlinear DSM feature space.

In this paper, we present a low-cost model named Select Feature, Principal component analysis and Classifier (SFPC)

978-1-5090-1328-9/16/$31.00 ©2016 IEEE

as shown in the Fig. 1, which consists of feature selector, principle component analysis, and classifier. Data mining approaches are performed from two aspects: selecting vectors corresponding to time series and reducing dimensionality of DSM indexes. The price classification engine consists of a feature selection block, a dimensionality reduction block and a classification block. The main contributions of this paper can be summarized as follows:

- Proposing a new electricity price forecasting system framework SFPC.
- Combining Grey correlation analysis and ReliefF algorithm successfully eliminates redundancy among features and calculates the integrated feature weight.
- To reduce dimensionality for nonlinear DSM features, kernel function and principle component analysis are combined to transform DSM indexes to fewer independent principle components.
- Application of differential evolution solves the problem of finding appropriate training parameters for SVM.

The rest of this paper is organized as follows. In next section, feature selector algorithm is illustrated. Sections III and IV present the dimensionality reduction module and the improved classifier based on DE-SVM, respectively. In section V, we illustrate the system performance by conducting a series of various comparisons. Finally the conclusion is drawn in section VI.

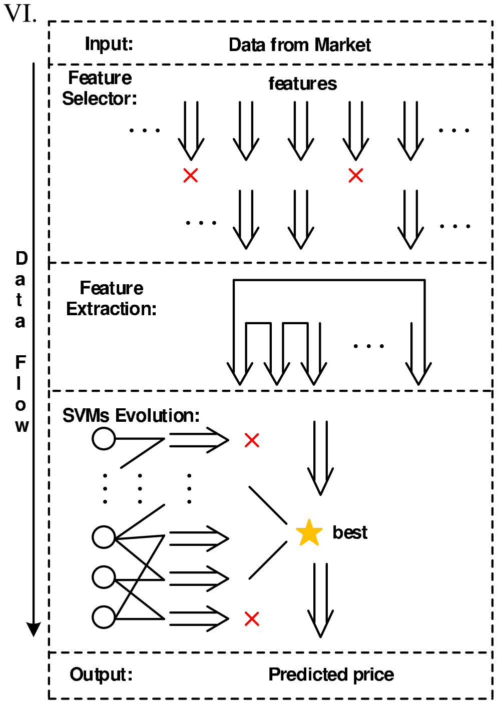


Fig. 1: System Framework

II. Feature Selector based on GCA-ReliefF

In this section, the crucial problem is how to select the critical features from the abstract input data. Some classic algorithms to implement feature selectors are Relief, Correlation Feature Selection (CFS) and Mutual Information (MI). This paper chooses ReliefF algorithm as the feature selector for its wide adaptability and easy implementation. Also, recent study in [9] reveals striking results obtained by the Relief algorithm for the DSM forecasting problem.

We use matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \tag{1}$$

to describe the final electricity price. Rows of the matrix denote the time intervals in hours ahead of the forecasted price. Columns denote the various features used to describe the electricity price. For example, a_{ij} means the jth feature i hour ahead of the aimed price. We can also denote it in time sequence

$$A^T = [\vec{t_1}\ \vec{t_2} \cdots \vec{t_m}], \tag{2}$$

where A^T is the transpose of a matrix A, and

$$\vec{t_k} = \begin{bmatrix} \lambda_1(k) & \lambda_2(k) & \dots & \lambda_n(k) \end{bmatrix} \qquad k \in [1,m]. \tag{3}$$

Because the contribution of each feature varies, it is required to find a solution to determine each feature's weight to the final electricity price. To solve it, we propose the concept of grey correlation analysis. It directly correlates each feature with the final price, and finds the correlation coefficient for each feature.

Grey data processing must be performed before grey correlation coefficients can be calculated. A series of various units must be transformed to be dimensionless. Usually, each series is normalized by dividing the data in the original series by their average. Let the original reference sequence and sequence for comparison be represented as $\lambda_0(k)$ and $\lambda_i(k)$, $i = 1,2,...,n$; $k = 1,2,...,m$, respectively, where n is the size of features to be considered, and m is the total number of time sequence. Data preprocessing converts the original sequence to a comparable sequence. Several methodologies of preprocessing data can be used in grey relation analysis, depending on the characteristics of the original sequence [10]. The original sequence can be normalized as

$$\lambda_i^*(k) = \frac{\lambda_i(k) - min\lambda_i(k)}{max\lambda_i(k) - min\lambda_i(k)}. \tag{4}$$

After the normalization, we can calculate the gray coefficient as

$$\begin{aligned} &\gamma(\lambda_o^*(k),\lambda_i^*(k)) = \frac{\Delta_{min} + \zeta\Delta_{max}}{\Delta_{oi}(k) + \zeta\Delta_{max}}, \\ &\Delta_{oi}(k) = |\lambda_o^*(k) - \lambda_i^*(k)|, \\ &\Delta_{max} = max|\lambda_o^*(k) - \lambda_j^*(k))|,\ j \in [1,n] \ \text{ and } k \in [1,m], \end{aligned} \tag{5}$$

where ζ is the distinguishing coefficient, $\zeta \in (0,1)$, $\Delta_{min} = min_{i,k}|\lambda_o^*(k) - \lambda_i^*(k)|$. Generally, ζ is taken as 0.5 [11].

According to [11], the grey grade is an average of the grey correlation coefficients and is defined as

$$\Gamma_i(\lambda_o^*,\lambda_i^*) = \frac{\sum_{k=1}^{m}\gamma(\lambda_o^*(k),\lambda_i^*(k))}{m}. \tag{6}$$

By comparing the threshold value with every feature's grey correlation grade $\Gamma_i(\lambda_o^*,\lambda_i^*)$, some low correlated features would be removed. We use ϵ to denote the number of removed features.

The realization of grey correlation-based ReliefF algorithm is shown as **Algorithm 1**, where the nearest hit instance is the closest instance within the same class C while the nearest miss instance is the closest instance in the opposite class $\bar{C}$ in the sense of Euclidean distance measure. The function $diff(A,instance,instance)$ computes the difference between the values of attribute for two instances, and its form are as follows:

$$diff = \begin{cases} \mathbf{0} & \text{(If the attribute is discrete,} \\ & \text{ and the two values are same)} \\ \mathbf{1} & \text{(If the attribute is discrete,} \\ & \text{ and the two values are different)} \\ \multicolumn{2}{l}{\textbf{Actual difference normalized to interval [0,1]}.} \\ & \text{ (If the attribute is continuous)} \end{cases}$$

Algorithm 1: Grey correlation-based ReliefF

```
Input: $W[T_k] = 0.0$,$T_k$,$\delta$,$\delta^*$
Output: $A^* = \{T(1),T(2),...,T(m-\alpha)\}$
begin
    initialization: set all weight $W[T_k] = 0.0$, read data from $A[]$
    for $\lambda_i^*(k) \in A$ do
        normalization data: $\lambda_i^*(k) = \frac{\lambda_i(k) - min\lambda_i(k)}{max\lambda_i(k) - min\lambda_i(k)}$;
        find the grey coefficient $\gamma(\lambda_o^*(k),\lambda_i^*(k))$;
        if $\gamma(\lambda_o^*(k),\lambda_i^*(k)) < \delta^*$ then
            remove $\lambda_i^*(k)$ from $A$
        else
            reserve $\lambda_i^*(k)$
        end
    end
    for $k$ from $1$ to $m$ do
        randomly select an instance in $class(C_i)$
        find $k$ nearest hits instance $H_j(C_i)$
        for each $class(C_j) \neq class(C_i)$
        find $k$ nearest miss instance $M_j(C_j)$
    end
    for $T^*_{k=1} to T^*_{k=m}$ do
        $W[T_k] = W[T_k] - \frac{\sum_{j=1}^{k} diff(\vec{A},\lambda_i,H_j)}{m*k} + \frac{\sum_{C \neq class(\lambda)} diff(\vec{A},\lambda_i,M_j(C))}{m*k}$
        reserve all $T_k$ which satisfy $W[T_k] > \delta$
        output the $A^* = \{T(1),T(2),...,T(m-\alpha)\}$,$W[T(k)] > \delta$
    end
end
```

III. DIMENSIONALITY REDUCTION-KPCA

In this section, we study the dimensionality reduction problem for this system. Although PCA is a common solution for reducing dimensionality, there exists a limitation that PCA can only be effectively performed on a set of observations that vary linearly [8], but not suitable for dealing with nonlinear DSM feature space. Thus, we propose to adopt KPCA, which combines kernel function with PCA, to handle this problem.

Remark 1: The features selected by GCA-ReliefF will be the sample for principal component analysis. We assume the matrix of feature selector's output is

$$X = (x_1,\, x_2,\, x_3,\, ...,\, x_p)^T, \tag{7}$$

where $x_1,\ x_2,\ x_3,\ \ldots,\ x_p$ are the variables of p indexes related to regulation clearing price. i.e. x_1 represents real-time congestion component.

To obtain Principle Components (PC) in feature space, we first have to solve the eigenvalue problem

$$\lambda v = C^F v, \tag{8}$$

where C^F is the covariance matrix C of selected features X in feature space F, eigenvalues $\lambda \geq 0$ and $v \in F$. Here, $C^F v$ can be formulated as

$$C^F v = \frac{1}{N}\sum_{i=1}^{N}\langle\phi(x_i),v\rangle\phi(x_i), \tag{9}$$

where $\sum_{k=1}^{N}\phi(x_k) = 0$, $\phi()$ is a nonlinear mapping function for projecting the input vectors to F, and $\langle x,y\rangle$ denotes the dot product between x and y. Hence, problem (8) can be transformed into the problem

$$\lambda\langle\phi(x_k),v\rangle = \langle\phi(x_k),C^F v\rangle, \qquad k = 1,\, ...,\, N, \tag{10}$$

where solutions v with $\lambda \neq 0$ can be expressed as

$$v = \sum_{i=1}^{N}\beta_i\phi(x_i), \tag{11}$$

where $\beta_i(i = 1,\, ...,\, N)$ are the corresponding coefficients.

Hence, it is noted that the eigenvalue problem only involves dot products of mapped shape vectors in the feature space. Since the mapping $\phi()$ may not always be computationally tractable, only dot products of two vectors in the feature space are needed [8]. Then we define an $N \times N$ matrix K by $K_{ij} = \langle\phi(x_i),\phi(x_j)\rangle$.

Proposition 1: *The solution of eigenvalue problem* (10) *can be represented as the solutions of the following problem related to **K** matrix*

$$N\lambda\beta = K\beta. \tag{12}$$

Proof:

Combining Eqs. (10) and Eq. (11), we can get:

$$\lambda\sum_{i=1}^{N}\beta_i K_{ki} = \frac{1}{N}\sum_{i=1}^{N}\beta_i\sum_{j=1}^{N}K_{kj}Kij. \tag{13}$$

Since $k = 1,\, ...,\, N$, Eq. (13) is equivalent to

$$\lambda N K\beta = K^2\beta, \tag{14}$$

where $\beta = [\beta_1, ..., \beta_N]^T$. To find the solution of Eq. (14), the eigenvalue problem should be solved as $N\lambda\beta = K\beta$.

■

In order to reduce the dimensionality, we can pick the first n eigenvectors β_1, β_2, ..., β_n with $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$. Then, β_1, β_2, ..., β_n need to be normalized.

Proposition 2: *Normalizing* β_1, β_2, ..., β_n *leads to the result* $\lambda_k \langle \beta_k, \beta_k \rangle = 1$.

Proof: Normalizing β requires vectors in feature space F to be normalized:

$$\langle v_k, v_k \rangle = 1 \qquad k = 1, ..., n. \tag{15}$$

Combining Eq. (11) with Eq. (15) leads to:

$$\begin{aligned} 1 &= \langle \sum_{i=1}^{N} \beta_i^k \phi(x_i), \sum_{j=1}^{N} \beta_j^k \phi(x_j) \rangle \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} \beta_i^k \beta_j^k \langle \phi(x_i), \phi(x_j) \rangle \\ &= \lambda_k \langle \beta_k, \beta_k \rangle. \end{aligned} \tag{16}$$

■

Then the principal component p of a test vector x can be extracted by projecting function $\phi(x)$ onto eigenvectors $v_k \in F$, where $k = 1$, ..., n,

$$p_k = \langle v_k, \phi(x) \rangle = \sum_{i=1}^{N} \beta_i^k \langle \phi(x_i), \phi(x) \rangle. \tag{17}$$

There are a number of practicable kernel functions K of form $K(x,y) = \langle \phi(x), \phi(y) \rangle$, e. g. linear kernel, sigmoid kernel, and radial basis kernel. After simulation, the results show that radial basis kernel is the most suitable function for DSM features (more details in simulation part).

IV. CLASSIFIER BASED ON DE-SVM

After the processes with feature selector and dimension reduction, low weighted and redundant features have been filtered. The original complex multi-dimensional $(m \times n)$ data has have been simplified in $(m - \alpha \times n - \epsilon)$ variables. This section aims to solve the problem on how to utilize these data to accomplish the final electricity forecasting. The problem is to find a map from the input $\phi(t,x)$ to the output $\Omega(y)$. SVM is a viable and optimal solution.

A. Problem Formulation

The problem can be formulated as

$$f(x,c) = \sum_{i=1}^{D} c_i \lambda_i(x) + b, \tag{18}$$

where b and $c_{i}{}_{i=1}^{\infty}$ are coefficients that have to be estimated from the data. This form of approximation can be considered as a hyperplane in the D-dimensional feature space defined by the functions $\lambda_i(x)$. The dimensionality of the feature space is not necessarily finite, and we will present examples in which it is infinite. The unknown coefficients are estimated by minimizing the following functional coefficients and are estimated by minimizing the regularized risk function

$$\omega(c) = \frac{\sum_{i=1}^{N} |y_i - f(x_i,c)|_\epsilon + \mu c^2}{N}, \tag{19}$$

where μ is a constant and the following robust error function is formulated:

$$\begin{cases} x = 0 & if |y_i - f(x_i,c)| < \epsilon \\ x = |y_i - f(x_i,c)|. & otherwise \end{cases} \tag{20}$$

The function that minimizes the functional in (19) depends on a finite number of parameters, and has the following form:

$$f(x,\alpha,\alpha^*) = \sum_{i=1}^{N} (\alpha^* - \alpha) K(x,x_i) + b, \tag{21}$$

where $\alpha\alpha^* = 0$, $\alpha^* \geq 0$ for $i = 1$, ..., N. In addition, $K(x,y)$ is called kernel function, and describes the inner product in the D-dimensional feature space

$$K(x,y) = \sum_{i=1}^{D} \lambda_i(x) \lambda_i(y). \tag{22}$$

The interesting fact is that for many choices of the set $\{\lambda_i(x)\}_{i=1}^{D}$, including infinite dimensional sets, the form of K is analytically known and very simple, and the features λ_i never need to be computed in practice because the algorithm relies only on computation of scalar products in the feature space. Several choices for the kernel K are available, including gaussians, tensor product B-splines and trigonometric polynomials. The coefficients α and α^* are obtained by maximizing the following quadratic form:

$$\begin{aligned} R(\alpha^*,\alpha) = &-\epsilon \sum_{i=1}^{N} (\alpha_i^* + \alpha_i) + \sum_{i=1}^{N} y_i (\alpha_i^* - \alpha_i) \\ &- \frac{1}{2} \sum_{i,j=1}^{N} (\alpha_i^* + \alpha_i)(\alpha_i^* - \alpha_i) K(x_i,y_i). \end{aligned} \tag{23}$$

B. Optimal Strategies

As discussed before, the different value of SVM's parameter has a gigantic effect on the exactness of regression analysis. Whereas, how to determine the reasonable and efficient parameters is still a challenging problem now. The common approach to select SVM parameters is by using cross validation or gradient descent strategy. However, both solutions are inefficient and may cause unsatisfying results. In this paper, a more reliable method has been proposed, named differential evolution (DE). DE algorithm is similar to genetic algorithms which have analogous operations of initialization, crossover, mutation and selection. Swift convergence and fewer control parameters are the two promising advantages of DE algorithm.

Remark 2: Five Procedures of Proposed DE

$\{1\}$ *Population structure*: The current population, symbolized by principal components, is composed of those D-dimensional vectors $X_i^g = X_1^g$, X_2^g, ..., X_D^g with index g indicating the generation which a vector belongs to. In addition, each vector is assigned a population index, i, which varies from 1 to the population size N_p. Once initialized, DE mutates randomly selected vectors to generate an intermediate population.

{2} *Initialization*: This stage consists in forming the initial population. For example, if our objective is the optimization of the membership functions, the initialization step consists choosing the interval of this function randomly.

{3} *Mutation*: The $(g+1)$-generation equation of the ith mutant vector:

$$V_i^{g+1} = X_*^g + F(X_{r1}^g - X_{r2}^g), \tag{24}$$

$$F_i = F_{max} - \frac{g(F_{max} - F_{min})}{g_{max}}, \tag{25}$$

where X_{r1}^g,X_{r2}^g are randomly selected individuals from the g generation, and X_*^g is the best individual in g generation, F_i is the ith iteration factor, F_{max} and F_{min} are the maximum and the minimum mutant scale factor, and g_{max} is the maximum generation.

{4} *Crossing*: This operation is to increase the variety of generation. The relative vector V_i^{g+1} is mixed with the transferred variable X_i^{g+1} to produce a test vector U_i^{g+1}. The crossing operation is as follows:

$$\begin{cases} u_{i,j}^g = |v_{i,j}^g| & r_{i,j} \leq CR \\ u_{i,j}^g = x_{i,j}^g, & otherwise \end{cases} \tag{26}$$

where $j = 1$, 2, ..., D, D represents the space dimension and CR is the crossover probability. $r_{i,j}$ is a random number in [0,1].

{5} *Selection*: All the individuals in the population have the same opportunity to be selected, regardless of their fitness function value. Selection operation can be described as:

$$\begin{cases} X_i^{g+1} = |U_i^{g+1}| & f(U_i^{g+1}) < f(X_i^g) \\ X_i^{g+1} = X_i^g, & otherwise \end{cases} \tag{27}$$

where $f(U_i^{g+1})$ and $f(X_i^g)$ represent the fitness function of U_i^{g+1} and X_i^g, respectively.

The algorithm of DE-SVM is shown as **Algorithm 2**, where c is a cost penalty, ε is an insensitive loss function parameter and σ is a kernel parameter.

V. Numerical Results

In this section, numerical results are provided to validate the proposed algorithms of this forecasting system. We take the hourly DSM data of the ISO New England Control Area (ISO NE-CA) as the input for the whole forecasting process [12]. Simulation results are divided into the following three parts:

(1) GCA-ReliefF feature selection results;
(2) KPCA results obtained by different kernel functions compared with PCA;
(3) DE-SVM results.

First, Grey correlation analysis and ReliefF method are combined to select features from hourly DSM data in ISO NE-CA during 2014-1-1 to 2014-2-1. In GCA-ReliefF algorithm, one feature seems like a vector containing the indexes corresponding to its time. Since this system is to forecast regulation clearing price (RegCP), the indexes which have low correlation with the price should be eliminated. Here, grey correlation analysis is applied to calculate the correlation between the indexes and RegCP. The results shown in Fig. 2 point out that most indexes' correlation with RegCP are above 0.9. Thus, the correlation threshold can be set as 0.9. In this case, index DA-CC (0.8543) and index RT-CC (NaN) are eliminated. Through the results obtained by grey correlation analysis, the proportion of each retained indexes in features can be calculated. Then, we can run ReliefF to calculate each feature's weight by using **Algorithm 1**. As shown in Fig. 3(a) and Fig. 3(b), weight threshold can be set as 20 to gain the features with higher weight.

Second, in order to reduce dimensionality, and eliminate the correlation of indexes in feature vectors, KPCA is applied to extract components to accomplish this task. The results of

Algorithm 2: Differential Evolution-based SVM

```
Input: $X_i = (0,0,0)$
Output: $F(y) = \{C_1,C_2,...,C_k\}$
begin
    initialization: set $g_{max}$, $F$, $CR$, $Np$,
    random initialize: $X_i = \{c,\sigma,\varepsilon\},i = 1,2,...,Np$
    for $i \in (1,2,...,Np)$ do
        calculate $f_i(X_i) = \frac{1}{TN}\sum_{i=1}^{T} N(\widetilde{y_i} - y_i)^2$
        if $f_i(X_i) < f_{i+1}(X_{i+1})$ then
            reserve $f_i(X_i)$
            compare $f_i(X_i)$ $and$ $f_{i+2}(X_{i+2})$
        else
            reserve $f_{i+1}(X_{i+1})$
            compare $f_{i+1}(X_{i+1})$ $and$ $f_{i+2}(X_{i+2})$
        end
        get the minimum $f_{min}(X_i)$ $denote$ $the$ $X_i$ $as$ $X_*$
    end
    calculate $V_i^{g+1} = X_*^g + F(X_{r1}^g - X_{r2}^g)$,$u_{i,j}$,$X_i^{g+1}$
    get the optimal parameter $X_* = (c_*,\sigma_*,\varepsilon_*)$
    get $X_*$ to SVM regression function
    solve regression function, output $F(y)$
end
```

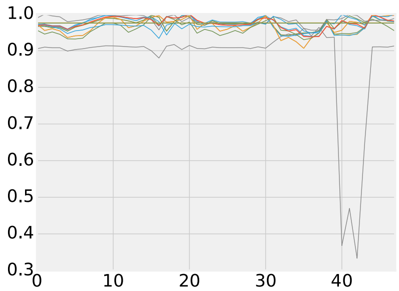
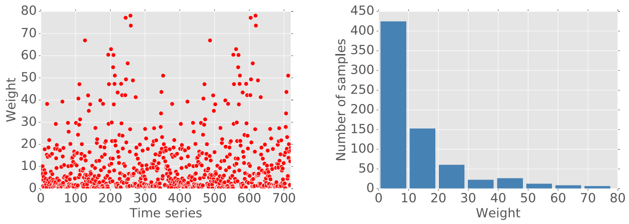


Fig. 2: Grey correlations between indexes and RegCP (index DA-CC (c=0.8543) and index RT-CC (c=NaN) can be eliminated)



(a) Weight distribution of feature vectors during 720 hours

(b) Feature weight during 720 hours (threshold is set as 20, total number of selected features is 128)

Fig. 3: Weight distribution and feature weight during 720 hours

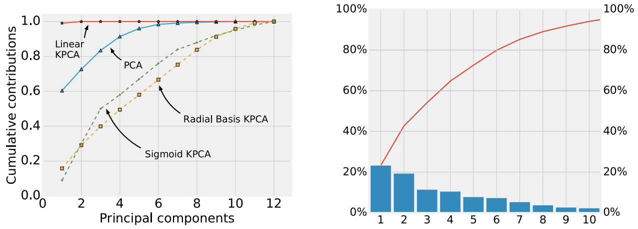
TABLE I: ATTRIBUTES OF DATA

Hour	✓	DA_DEMD	✓	SYSLoad	✓
DA_DEMD	✓	DEMAND	✓	DewPnt	▼
DA_LMP	✓	DA_CC	▲	DryBulb	✓
DA_EC	▼	DA_MLC	✓	RT_MLC	▼
RT_LMP	▼	RT_EC	✓	RT_CC	▲

NOTE: The ▲ means that the feature is irrelevant, the ▼ marks the redundant features, and the ✓ indicates the selected features.

PCA and KPCA are compared by cumulative contribution, which are shown in Fig. 4(a). It is obvious that radial basis KPCA extracts the most principle components when cumulative contributions achieve 95%, which brings enough principal components to ensure the accuracy of forecasting. Thus, radial basis KPCA is chosen for extracting principal components, as shown in Fig. 4(b). The result of feature selection and load reduction is also given in Table. I.

Finally, comparison among DE-SVM, Naive Bayes and Decision Tree on electricity price forecasting is shown in Fig. 5. It can be seen that DE-SVM has better performance than Naive Bayes and Decision Tree, since price curve of DE-SVM fits well with that of real value. In other words, the predicted price classification is pretty close to real value.



(a) Cumulative contribution comparison of PCA, linear KPCA and radial basis KPCA

(b) Cumulative contribution of radial KPCA based on feature output

Fig. 4: Cumulative contribution of different PCA

Acknowledgment

This work is supported by NSFC (61572262); NSF of Jiangsu Province (BK20141427); NUPT (NY214097); Open research fund of Key Lab of Broadband Wireless Communication and Sensor Network Technology (NUPT), Ministry of Education (NYKL201507); Qinlan Project of Jiangsu Province.

VI. Conclusions

In this paper, we proposed a multi-dimensional features based price forecasting model which consists of feature selector, dimensionality reduction, and classifier. Specifically, multi-dimensional features time sequences are used as variables in this model. Additionally, an improved algorithm called GCA-ReliefF is proposed to deal with the time sequence with n-dimensional features. Furthermore, applying KPCA immensely reduces the forecasting running time. Moreover, the improved DE-SVM can capture the best SVM parameters for classification automatically by utilizing DE algorithm. Simulation results reveals that the predicted price classification only has less than 5% errors compared with the real value. In the future, electricity load feedback will be considered in this system.

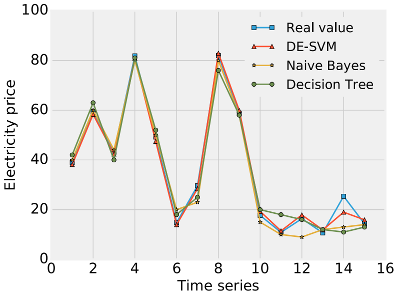


Fig. 5: Comparison between training and test data by DE-SVM

References

[1] R. Zhou, Z. Li, C. Wu, and M. Chen, "Demand Response in Smart Grids: A Randomized Auction Approach," *IEEE Journal on Selected Areas in Communications*, Vol. 33, No. 12, 2015, pp. 2540-2553.

[2] Assessment of demand response and advanced metering: staff report, Federal Energy Regulatory Commission, Sep. 2007. [Online]. Available: http://www.ferc.gov/legal/staff-reports/09-07-demand-response.pdf.

[3] M. Erol-Kantarci, and H. T. Mouftah, "Energy-Efficient Information and Communication Infrastructures in the Smart Grid: A Survey on Interactions and Open Issues," *IEEE Communications Surveys and Tutorials*, Vol. 17, No. 1, 2015, pp. 179-197.

[4] D. Huang, H. Zareipour, W. D. Rosehart, and 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.

[5] A. Fleury, M. Vacher, and N. Noury, "SVM-based multimodal classification of activities of daily living in health smart homes: sensors, algorithms, and first experimental results," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 14, No.2, 2010, pp. 274-283.

[6] Huan Liu, and Hiroshi Motoda, *Computational Methods of Feature Selection*, CRC Press, Boca Raton, FL, 2007, pp. 440.

[7] M. Wu, and Y. Wang, "A feature selection algorithm of music genre classification based on ReliefF and SFS," *IEEE International Conference on Computer and Information Science (ICIS)*, 2015, pp. 539-544.

[8] Z. Du, X. Wang, L. Zheng, and Z. Zheng, "Nonlinear system modeling based on KPCA and MKSVM," *in Proceedings of International Colloquium on Computing, Communication, Control, and Management*, 2009, pp. 61-64.

[9] S. Venkatesan, T. S. Manipriya, and J. S. Kumar, "Demand side management based congestion relief model," *in Proceedings of International Conference on Advances in Electrical Engineering (ICAEE)*, 2014, pp. 1-6.

[10] Y. Daheng, "Optimizing design scheme of energy saving in warehouse building based on grey relational analysis," *in Proceedings of International Conference on Logistics Systems and Intelligent Management*, 2010, pp. 1090-1092.

[11] M. Saman, I. Aydin, and E. Akin, "Grey clustering based diagnosis of induction motor faults," *in Proceedings of IEEE Conference on Signal Processing and Communications Applications*, 2009, pp. 61-64

[12] ISO New England Energy Offer Data [Online]. Available: www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/day-ahead-energy-offer-data