Predicting Short Time Weights using Dimension Reduction Method

Zhao Liyan and Wei Yanna

School of Computer and Remote Sensing Information Technology,
North China Institute of Aerospace Engineering, Box 15, P.O. Box 130,
Langfang, 065000, China

***Abstract* — Taking the overall and correlation of each time point in short time weight value into account, we present a novel method for predicting short time weight values using dimension reduction method. Firstly, we take the thirty-six weight values as a thirty-six dimensional dataset into account, and construct a higher dimensional weight predicting function by conducting the complicated weight variation from the multi-dimensional angle. It utilities dimension reduction method to conduct an efficient manifold learning operation for the higher dimensional function, thus exposing the spatial data attributes and integrity geometric rules of the higher dimensional samples, and revealing the effective information within. We utilize locally linear embedding (LLE) algorithm to conduct the non-linear dimensionality reduction for the thirty-six dimensional weight samples, and predict the weight values in lower dimensional embedding. Then, we use the locally linear embedding algorithm to establish the thirty-six weights forecast. Experimental results demonstrate that the proposed method performs better than those component forecasts.**

Keywords - short time weights; dimensionality reduction; manifold learning; locally linear embedding (LLE)

I. INTRODUCTION

Short time weight predicting is an important issue to be faced in terms of power system scheduling, operation and control. Wide-spread attention has been paid and many predicting methods have been put up [1,2]. However, due to weight diversity, complexity and variability in time and space, the existing methods are still difficult to meet the requirements. One important reason is that there are some limitations on the perspective of research methods. Existing weight predicting researching methods can be divided into two categories: one is the traditional method [2,3] which is represented by time series and regression analysis; another class is artificial intelligence methods [4,5,6] which is represented by artificial neural networks, least squares support vector machine and genetic algorithm. The traditional method is simple and fast. However, based on one-dimensional thinking space and simple model, it is difficult to depict the complex laws of weight changes accurately. By simulating human thinking, artificial intelligence methods, to some degree, reflect nonlinear variation rule. It has high prediction accuracy. However, considering the weighting system from the perspective of the one-dimensional, it may have lost some other effective information. In fact, there is strong relation between the weighting values at every point in one day, for example, temperature, humidity, type of day and other factors measured by days have influences on weight [7]. Therefore, from the multidimensional perspective on the daily weight thirty-six overall dimensional data analysis and establish model to overcome the shortcoming of one-dimensional processing methods and better depict the daily weight variation rule, improving the accuracy of the prediction result and adaptability of predictive methods.

From the perspective of system theory, seeing the weight of thirty-six hours a day as a system, we can propose a new method of depicting and predicting short time weight as a whole and establish a model of high dimensional prediction [8,9]. By the manifold learning, we can have an effective dimensionality reduction of the established high dimensional prediction model and make forecast on the short time weight in space of dimensionality reduction. By locally linear embedding method LLE, a nonlinear dimensionality reduction of thirty-six-dimensional space model can be made, thereby effectively extracting the inherent property of high dimensional spatial data and the overall law. After making a low-dimensional space weight predicting by using Least Square Support vector Machines (LS-SVM), the predicted results of thirty-six hours can be got by reconstituting LLE. The above method will be applied to some local regional power grid to establish model and make prediction. By making a comparison of the results with that in one-dimension, the feasibility and correctness of the thesis can be proved. We present the main contributions as follows in this paper:

1) We present a novel method for predicting short time weight values using dimension reduction method by taking the overall and correlation of each time point in short time weight value into account.

2) We take the thirty-six weight values as a thirty-six dimensional dataset into account, and construct a higher dimensional weight predicting function by conducting the complicated weight variation from the multi-dimensional angle.

3) We utilize dimension reduction method to conduct an efficient manifold learning operation for the higher dimensional function; thus, exposing the spatial data attributes and integrity geometric rules of the higher dimensional samples, and opening out the effective

information it includes.

4) We use the locally linear embedding algorithm to establish the thirty-six weights forecast. Experimental results demonstrate that the proposed method conducts better than those component forecasts.

II. WEIGHT SYSTEM HIGH DIMENSIONAL PREDICTION MODEL

Power weight system is a multi-dimensional, nonlinear system, short time weight forecast, such as the daily weight predicting. Based on existing historical data, we can predict the weight values at many time points in the future or even more than a day (e.g. 24, 42, 94, 258). Let's imagine that there are weight data of thirty-six-point days in N days. $X'(n)=[X_1'(n), X_2'(n), \dots, X_{24}'(n)]^T$ (n=1, 2, … , N). There are 24*N points. The theory of multidimensional time series prediction is not yet mature and generally there are lots of calculations. Therefore, the existing methods are analyzed from the perspective of one dimension. That's to say, considering the $X'(n)$ by the hour type as seen thirty-six $X'(n)$, we can form thirty-six types of training and prediction samples and perform forecast modeling thirty-six times. Reference ten has made relevant calculation on thirty-six hours sequence of weight values in one day in Tianjin-Tangshan and Hebei grid network. They have found that the maximum of the correlation coefficient between adjacent points of time series is 0.8625, the maximum of that is 0.9868 and the maximum of that between of the non-adjacent point is 0.6745. It has showed that there is a strong correlation between the weight values at each time of the day. Dividing that into thirty-six one-dimensional sequence, we can establish models separately. Ignoring the integrity and relevance between the weight value and the time relevance, we have lost a lot of valuable information. We can made an overall analysis of the thirty-six-dimensional data sets, dig up some hidden but available information from the historical from historical perspective, so as to portray more accurately the variation rule of daily weight.

Record the historical weight value in twenty-four hours in N days in this way, $X'=[X_1, X_2, \dots, X_{24}]$ Every X serves as thirty-six-dimensional data, namely, $X'(n)=[X_1'(n), X_2'(n), \dots, X_{24}'(n)]^T$. Because the power weight system itself is complex nonlinear and subjected to many factors and complex fluctuations regularity, it is difficult to directly percept and recognizes the inherent effective information and rules. Only by relying on relevant data analysis and reduction methods can we dig up the data of high dimensional space. One of the important means of solving this high-dimensional spatial data digging up problem is to reduce to reduce the dimension of high-dimensional data [10]. By reducing the dimension to get rid of the noisy and redundant information in raw material, we can convert it to a low-dimensional data losing as less information as possible, discovering the intrinsic properties and overall geometric regularity of high-dimensional data in low-dimensional space and revealing the hidden but effective information. Based on manifold learning theories, an analysis of thirty-six dimensional weight data has been made. By employing locally linear embedding algorithm to reduce the dimensionality, we have obtain the D (d<thirty-six) dimensional data, namely Y $[Y_1, Y_2, \dots, Yn]$ (every dimensionality of Yi is d). D dimension sequence in reduction dimensionality is irrelevant. Thus, we can make prediction the d-dimensional sequence respectively and establish LLE to forecast the d-dimensional predictive value so as to get the predictive value of thirty-six points.

TABLE I. THREE LOW-DIMENSIONAL COORDINATES AND RAW HIGH-DIMENSIONAL WEIGHT VALUES

	Average	**Standard error**	**Variable coefficient**	**Range**
Coordinate	**First dimension**	**Second dimension**	**Second dimension**	**Third dimension**
Correlation coefficient	0.6995	0.6543	0.8569	0.7969

III. MODEL DIMENSIONALITY REDUCTION AND RECONSTRUCTION BASED ON MANIFOLD LEARNING

A. *Manifold Learning Principles*

With the continuous development of information technology, people ability to access and store data is greatly enhanced. Moreover, the data obtained are mostly high-dimensional ones. For high-dimensional data sets, it is difficult to employ existing methods to directly perceive its inherent laws. Therefore, we must resort to a variety of data analysis and reduction methods to understand the data and reveal the hidden but effective information. This process can be called information selection. From data to information is a leap from quantitative to qualitative change, which requires effective learning method. Manifold learning is a non-parametric method for dimensionality reduction of high dimensional data.

In 2000, the "Science" magazine published three papers which has discussed the manifold learning from the cognitive and first used the term manifold learning. It marks the born of manifold learning which is mainly characterized by nonlinearity. Manifold learning is defined as follows: Imagine $Y \in R^d$ is a low- dimensional manifold, and f: $Y \to R^d$ is a smooth embedding (D>d). The data set $\{Y_1\}$ are randomly generated, and become the observation space data $\{X_1 = f(Y_1)\}$ after the mapping of f. Manifold learning is to reconstruct f and (Y_1) on condition of $\{X_1\}$, a observed given sample set. The basic idea of manifold learning method is that manifold within each high-dimensional space has a corresponding manifold within low-dimensional space, tries to find a smooth map, and maps the source of high-dimensional data into the corresponding one of low-dimensional data. The main propose is that without any a priori assumptions, the whole geometric regularity can be found, which means finding the essence of things and inherent laws from the observations. Existing manifold learning methods include: local linear embedding method, isometric mapping algorithms, multidimensional scaling method, Laplacian feature mapping algorithm, etc. The study has found that local linear embedding method can perform better in model analysis of electricity, power weight data.

B. Locally Linear Embedding Method

LLE is a nonlinear dimensionality reduction method which reveals the global nonlinear structure through the union of local linear relationship. With the neighborhood relations of recording data, LLE calculates embedded manifold of high-dimensional input data in a low-dimensional space. Compared with other dimensionality reduction methods, LLE dimensionality reduction method has many advantages: first, it is equipped with adaptively in terms of nonlinear manifold structure data; second, it only involves fewer parameters selection; third, because it maintains the high internal dimensional topology of the data, it has lost smaller information.

Assuming that R^D data located in a high dimensional Euclidean space has a data set $X=\{X_1,X_2,\cdots,X_n\}$, which is situated on nonlinear manifold or its nearby whose intrinsic dimension is d (d<D). We assume that $Y=\{Y_1,Y_2,\cdots,Y_n\}$ is a low-dimensional embedding which embedded in a embedding space R^D. The detailed algorithm is divided into three steps. First, the neighbours can be selected. Calculate neighborhood points of each sample point X_i (i=l,2,...,N) (taking the K points nearest to neighbors or fixed radius spherical neighbourhood). Second, calculates the reconstruction right W_{ij}. In the neighbourhood of X_i, best value of matrix W_{ij} of each X_i is calculated to minimize the following objective function E(W).

$$E(W)=\sum_{i=1}^{n}\left\|X_i-\sum_{j=1}^{k}W_{ij}X_j\right\|^2,\ W_{ij}=0 \quad (1)$$

where, N means the points of every sample; K means the points of neighborhood; X_j (j=1,2,…, K) means k points corresponding to X_i; W_{ij} is a set of weighted values linear reconstructed by X_j (j=1,2,…,K). W_{ij} needs meet the next two constraint conditions, for example, equation (2), where W_{ij} =0, X_j is not the neighborhood point of X_i.

$$\sum_{j=1}W_{ij}=1 \quad (2)$$

X_i means the ith points of samples, X_j means the jth neighborhood point of the ith points of samples.

Third, calculate the embedded value of the d-dimensional Yi to minimize the error of the following reconstruction $\phi(Y)$.

$$\phi(Y)=\sum_{i=1}^{N}\left\|Y_i-\sum_{j=1}^{k}W_{ij}Y_j\right\|^2 \quad (3)$$

Yi is an embedded vector of low-dimensional of each sample point X_i, X_j means the K neighboring points of Yi corresponding. In order to ensure that the formula (2) has only one value, low-dimensional embedding Yi must meet the following two constraint conditions, such as the formula (4).

$$\sum_i Y_i=0,\ \frac{1}{N}\sum_i Y_iY_i^T=I \quad (4)$$

where, I is the unit matrix. Equation (2) is equivalent to finding a feature vector of M, a sparse symmetric positive definite matrix, that is:

$$M=(I-W)^T(I-W) \quad (5)$$

where, W is a weight matrix for the reconstruction. Take M smallest eigenvalues d+1 corresponding feature vector, and then sequence them in ascending order. Leave out the first one, the rest d feature vector consisting of a matrix is low-dimensional embedding vector Y.

C. LLE Reconstruction

There is a set of feature vector in embedded low-dimensional space. Based on this set of criteria, we can find the corresponding data set in a high-dimensional data space. Suppose we have obtained the feature vectors $Y=\{Y_1,Y_2,\cdots,Y_n\}$ by LLE in low-dimensional space and marked the new low-dimensional feature vector Y_0. LLE reconstruction can be used to find the original data points X_0 and approximate value $\overline{X_0}$ in a high-dimensional space based on $X=\{X_1,X_2,\cdots,X_n\}$ and $Y=\{Y_1,Y_2,\cdots,Y_n\}$. The detailed steps are: first: Find K neighboring vectors nearest to the new feature vector Y_0 from $Y=\{Y_1,Y_2,\cdots,Y_n\}$. Second, calculate the weight coefficient W_j to minimize the following objective function E(W).

$$E(W)=\left\|x_0-\sum_{j=1}^{k}W_jY_j\right\| \quad (6)$$

where, Y_0 is the low dimensional vectors to be reconstructed, Y_j (j=1,2,…,k) is k neighboring points of Y_0. The constraint condition is:

$$\sum_{j=1}^{k}W_j=1 \quad (7)$$

Third, $\overline{X_0}$ in high dimensional space can be obtained from reconstruction of formula (8).

$$\overline{x_0}=\sum_{j=1}^{k}W_jY_j \quad (8)$$

Suppose $X_0^{\{j\}}$ is the jth element of vector X_0, the reconstruction error of X_0 can be defined as:

$$RE(X_0)=\frac{1}{D}\sum_{j=1}^{D}\frac{\left|X_0^{\{j\}}-\overline{X}_0^{\{j\}}\right|}{X_0^{\{j\}}} \quad (9)$$

where, $1 \le j \le D$, D is the dimension in high-dimensional space. The reconstruction error of whole data set X can be defined as:

$$TRE = \frac{1}{N \cdot D} \sum_{i=1}^{N} \sum_{j=1}^{D} \frac{\left| X_0^{\{j\}} - \overline{X}_0^{\{j\}} \right|}{X_0^{\{j\}}} \qquad (10)$$

IV. LLE SHORT TIME WEIGHT PREDICTING MODEL AND METHODS BASED ON MANIFOLD LEARNING

Power system weight modelling requires a lot of historical data which are collected mostly by electricity collector or remote-action system. Except for inaccurate or missing data caused by device itself or the process of data transmission, there are various random factors leading to abnormal weight. Therefore, historical weight data often contain non-real data, namely "abnormal data", which need to be removed to ensure the integrity and accuracy of the data. The method employed in this article is to take the average before and after two points of the normal weight value to take place of abnormal points. The weight value fluctuates greatly in one day. The logarithmic transformation enables manifold learning to distributes in low-dimensional space more evenly, improving the effectiveness of manifold learning and reduce the reconstruction error of data set.

Based on statistical theory, support vector machine (SVM) is a new machine learning methods proposed by Vapnik and others in the mid-1990s. Based on the VC dimension and structural risk minimization principle, by answering a quadratic programming problem, this method has solved some practical application problems, such as small samples, nonlinearity, high dimension and local minimum points, etc. Compared to the traditional neural network method based on empirical risk minimization, this method has some advantages, such as fast learning, best within the whole, strong ability of generalization. Therefore, it is considered as an alternative to neural networks. Least squares support vector machine (LS-SVM) is an extension of the standard SVM. Using the squared term indicators to optimize index, and using equality constraints to take place of the standard SVM inequality constraints, it has converted the quadratic programming problem into solving linear equations, which has reduced the computational complexity and accelerated the solving speed. It has found wide-spread application in power system short time weight predicting.

After making overall analysis of pre-processed data, weight value within twenty-four hours a day can be regarded as an observation. The study found that when d is three, the intrinsic dimension d most suits the description for the relationship between the weights [11]. Table 1 shows three low-dimensional coordinates and raw high-dimensional weight value which has a strong correlation coefficient with the mean, standard deviation, variation coefficient, and range of weight in a day and the corresponding correlation coefficient after dimension reduction of daily weight data somewhere in China. As can be seen from Table 1, feature vector sequence of low-dimensional space after high-dimensional weight data dimensionality reduction, it can reflect the dynamic changes of the corresponding original high-dimensional data. The analyzing and modeling can be convenient and accurate for weight predicting. Taking into relevant factors account, we can analyze and establish models of the feature vector sequence of low-dimensional space respectively using LS and SVM and reconstruct LLE of the sequence prediction in low dimensional space to obtain scored twenty-four points weight predicting value prediction date.

TABLE II. PREDICTING RESULTS FROM JUNE 15 TO JUNE 21, 2014

Predicting time	Average predicting error		Improved percentage of this method proposed compared with the one-dimensional component prediction method
	Results using the one dimensional component prediction method	Results using the method proposed in this paper	
June 15	2.26%	2.12%	6.61%
June 16	2.96%	2.13%	28.47%
June 17	2.48%	1.78%	28.81%
June 18	1.79%	1.67%	14.34%
June 19	1.98%	1.54%	27.29%
June 20	1.97%	1.69%	11.35%
June 21	2.67%	1.88%	19.26%

TABLE III. PREDICTING RESULTS OF TWENTY-FOUR IN JUNE 19, 2014

Time/h	Actual weight/MW	Results using the one-dimensional component prediction method		Results using the method proposed in this paper	
		Predicting results/MW	Comparative error	Predicting results/MW	Comparative error
00:00	612.0	615.7	0.45%	606.3	0.28%
01:00	581.2	575.8	0.74%	573.5	0.47%
02:00	556.3	569.5	0.92%	544.5	0.88%
03:00	514.0	523.8	0.37%	529.0	1.93%
04:00	519.0	532.1	0.04%	527.5	0.76%
05:00	525.7	528.1	0.56%	522.5	0.75%
06:00	589.4	586.6	2.98%	573.3	2.74%
07:00	678.5	657.0	2.13%	654.6	2.09%
08:00	823.2	793.6	1.84%	805.2	1.24%
09:00	878.3	869.7	1.15%	862.4	0.05%
10:00	923.5	891.6	2.36%	895.4	2.73%
11:00	889.5	843.1	5.12%	889.6	0.72%
12:00	778.4	756.6	2.14%	761.0	0.85%
13:00	834.6	874.2	0.78%	844.0	1.42%
14:00	891.3	854.9	4.32%	857.4	3.55%
15:00	851.2	812.3	4.35%	852.2	0.76%
16:00	878.7	878.4	1.57%	869.0	1.23%
17:00	854.3	857.6	0.07%	847.0	0.25%
18:00	808.4	834.7	2.05%	826.5	2.87%
19:00	872.9	857.6	2.27%	872.4	0.69%
20:00	846.7	859.2	1.48%	856.0	1.84%
21:00	812.5	856.4	4.47%	824.7	2.45%
22:00	746.2	764.6	0.46%	753.5	0.76%
23:00	659.5	693.4	3.62%	682.3	2.54%

Average predicting error	1.89%	1.36%

As can be seen from the data in Table 2 and Table 3, using the method proposed herein to predict twenty-four hours weight value, the prediction accuracy is improved effectively. According to the prediction results of the regional power grid on June 19, 2014 listed in Table 3, using a one-dimensional component of the forecast, the average prediction error is 1.89%, while the maximum error is 5.04%. As for predicting results used by the proposed methods, the average prediction error is l.36% while the maximum error is 3.27%.

Take the average error 1.36% using the method proposed in this paper on June 19, 2014 for example. As for error control of forecast points within the accuracy range, one-dimensional component prediction is 40.35% while this method 63.7%. Percent of pass using one-dimensional component prediction is 79.89 percent while that of this method 96.03%. Meanwhile, in predicting the weight value of twenty-four hours a day, the time takes of one-dimensional component prediction is 3.68s while that of the method in this paper only 1.36s. Thus, the proposed method of prediction is more accurate and rapid.

V. CONCLUSIONS

Considering the integrity and relevance of weight value of twenty-four hours a day, having a whole analysis and modeling from the perspective of multi-dimensional can depict more accurately the complex daily weight fluctuations rule. By LLE, it can reduce dimensionality nonlinearly of high-dimensional data weight value, get rid of noise and redundant information Using the LS-SVM method, it can make weight predicting in dimension reduction space. Thus, using just a few unrelated sequence of low-dimensional space, we can establish LS-SVM model, reducing the modeling complexity. The numerical examples shows that compared to the traditional one-dimensional component prediction method, the proposed method effectively improves the speed and accuracy.

ACKNOWLEDGMENT

This work is supported by Langfang science and technology support project (2011011028; 2012011024; 2014011051) and Hebei Province science and technology support project (13210707).

REFERENCES

[1] Niu D, Cao SH, Lu J, et al. “Load forecasting technology and its applying. Beijing”: China Electric Power Press, vol. 2, No. 06, pp. 54-65, 2013.

[2] Kang C, Xia Q, Zhang B. “Review of power system load forecast and its development”. Automation of Electric Power Systems, vol. 36, No. 17, pp. 1-11, 2014.

[3] Lei SL, Sun C, Zhou Q, et al. “The research of local linear model of short-term electrical load on multivariate time series”. Proceedings of the CSEE, vol. 28, No. 02, pp. 24-28,2006

[4] Lu N, Zhou JZ, He Y. “Particle swarm optimization-based neural network model for short-term load forecasting”. Power System Protection and Control, vol. 2, No. 38, pp. 65-68, 2010..

[5] Zhao D, Wang M, Zhang J, et al. “A support vector machine approach for short term load forecasting”. Proceedings of the CSEE, vol.28. , No. 4 , pp. 27-30, 2014.

[6] Li G, Liu W. “Bus load short-term forecast based on LSSVM and Markov chain”. Power System Protection and Control, vol. 38 , No. 11 , pp55-60., 2012.

[7] Xie H, Chen Z, Niu DX. “Principle component analysis of data processing for short-term load forecasting”. Power System Technology, vol. 24 , No. 1 , pp. 43-46, 2000.

[8] Seung HS, Lee DD. “The manifold way of perception”. Science, vol. 290 , No. 5500 , pp. 2268-2269.

[9] Chen J, Deng SJ, Huo XM. “Electricity price curve modeling and forecasting by manifold learning”. IEEE Transactions on Power Systems, vol. 23, No.3 ,pp. 877-888, 2011.

[10] Yang FZ. “Data mining technology for high-dimensional data”. Nanjing: Southeast University Press, vol.25 , No.4 , pp.31-33, 2011.

[11] Peter JV, Robert PW. “An evaluation of intrinsic dimensionality estimators”. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17 , No. 1 , pp. 81-86, 1995