2015 IEEE 17th International Conference on High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), and 2015 IEEE 12th International Conf on Embedded Software and Systems (ICESS)

Improving Power Grid Monitoring Data Quality: An Efficient Machine Learning Framework for Missing Data Prediction

Weiwei Shi[1], Yongxin Zhu[1], Jinkui Zhang[1], Xiang Tao[1], Gehao Sheng[1], Yong Lian[1], Guoxing Wang[1], Yufeng Chen[2]

[1]School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

[2]Electric Power Research Institute of Shandong Power Supply Company of State Grid, Shandong Province, China

iamweiweishi@163.com, zhuyongxin@sjtu.edu.cn, {zjksjtu, xiangtaosjtu}@163.com, {shenghe, eleliany, guoxing}@sjtu.edu.cn, chenyufeng169@sina.com

Abstract*—Big data techniques has been applied to power grid for the evaluation and prediction of grid conditions. However, the raw data quality rarely can meet the requirement of precise data analytics since raw data set usually contains samples with missing data to which the common data mining models are sensitive. Though classic interpolation or neural network methods can been used to fill the gaps of missing data, their predicted data often fail to fit the rules of power grid conditions. This paper presents a machine learning framework (OR_MLF) to improve the prediction accuracy for datasets with missing data points, which mainly combines preprocessing, optimizing support vector machine (OSVM) and refining SVM (RSVM). On top of the OSVM engine, the scheme introduces dedicated data training strategies. First, the original data originating from data generation facilities is preprocessed through standardization. Traditional SVM is then trained to obtain a preliminary prediction model. Next, the optimized SVM predictors are achieved with new training data set, which is extracted based on the preliminary prediction model. Finally, the missing data prediction result depending on OSVM is selectively inputted into the traditional SVM and the refined SVM is lastly accomplished. We test the OR_MLF framework on missing data prediction of power transformers in power grid system. The experimental results show that the predictors based on the proposed framework achieve lower mean square error than traditional ones. Therefore, the framework OR_MLF would be a good candidate to predict the missing data in power grid system.

Keywords-missing data prediction; machine learning; support vector machine (SVM); power transformer

I. INTRODUCTION

The practices of data-driven management and decision making have been pervasive and widely used in today's industrial, business and governmental applications after initial successes of big data techniques in internet business. The data quality is regarded as a significant issue of industrial process, market success and decision-making activities [1].

However, more than 41% of the relevant projects would fail if only the original data were used due to the poor or insufficient quality of raw data according to a study by the Meta Group [2]. Missing data which means that electronic data during some period is lost or hidden by uncontrollable factors is one of the major potential flaws in raw data and could result in severe failure. Therefore, the engineers have to sacrifice much time to retrieve this kind of data for further analysis. As a consequence, (semi-)automatic missing data prediction methods have been proposed [3].

A large collection of data mining and statistical methods have been proposed to improve data quality due to missing data. For example, Ma's team proposed a good method for missing data prediction [4]. The algorithm focused on recommender systems using improved collaborative filtering method which outperformed the traditional collaborative filtering method. Nogueira et al [5] solved a practical problem based on the Fast Fuzzy Clustering Algorithm in real world: the prediction of bankruptcy, in which the used data set has missing values. Lei and Wang [6] presents a method for pre-processing the missing observed data by adopting the multiple imputation technique for Macau air pollution index (API) prediction using the Adaptive Neuro-Fuzzy Inference System (ANFIS). The API forecasting performance after missing data pre-processing is better than the conventional case without pre-processing.

In power grid systems, data missing happens so frequently due to the harsh working condition of sensors that classic methods often fail to handle. Expensive critical equipment such as main power transformers are monitored by multiple sensors. Unfortunately, these sensors are not as reliable as the equipment in the harsh open air working condition under the workload of 7*24 hours [7]. Moreover, sensors in remote rural areas such as mountains are usually maintained at an even worse level by workers who received less training than workers in city. Thus, it is normal and inevitable for the sensor system to produce flaw data sets, which lost or hidden some necessary information [8, 9]. These losses affect the data quality so badly that classic data mining and statistical methods alone cannot process these data properly.

In this paper, taking preceding parts into account, we propose our framework incorporating optimized support vector machine (OSVM) and refined SVM (RSVM) models as well as dedicated training strategies. To evaluate our scheme, raw data from power transformer sensors were collected for training and testing.

Our contributions can be summarized as follows:

1) We design an improved prediction models based on OSVM and RSVM;

978-1-4799-8937-9/15 $31.00 © 2015 IEEE
DOI 10.1109/HPCC-CSS-ICESS.2015.16

2) We propose a missing data prediction framework incorporating improved prediction models and data training strategies;

3) We implement and verify the framework and strategies with raw data from power transforms of State Grid of China.

The experimental results show that our novel approaches achieve better prediction accuracy of missing data than classic methods to predict missing data.

The rest of this paper is organized as follows. Section II discusses related work. Then, in Section III, missing data prediction methods of power transformer based on our framework OR_MLF as well as classic neural network (NN), least square SVM (LSSVM) and SVM methods are analyzed. Experimental examples are included in Section IV. Finally, in Section V conclusions are drawn.

II. RELATED WORK

In this section, we give a simplified example of missing data prediction and review several typical methods for prediction of missing data. Then, a short review of SVM for regression is also included.

A. An illustration example of missing data prediction

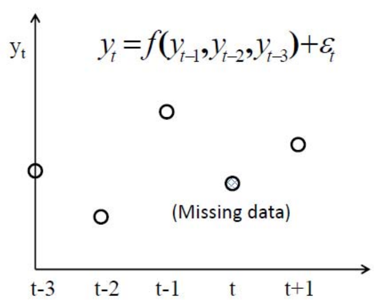

Figure 1. Illustration of missing data prediction

Taking the situation shown in Fig. 1 into consideration, we assume that y_t is missing. The time series model for missing data prediction is as follows:

$$y_t = f(y_{t-1}, y_{t-2}, y_{t-3}) + \varepsilon_t, \tag{1}$$

where *f()* is a nonlinear function and ε_t stands for additive noise. The prediction goal is to predict y_t based on past content, that is y_{t-1} y_{t-2}, and y_{t-3}.

B. Methods for missing data prediction

Simple ways such as eyeballing, calculating appropriate mean value and interpolation could be used in missing data prediction [10]. Although these methods are useful in some particular situations, they might fail when it comes to complex situations, such as consecutive missing data prediction [11-13]. As a result, considering the practical complexity, deeper understanding of original data and more complex methods turn to be necessary. Numerical analysis modeling, state space modeling, and stochastic modeling are alternative solutions to address this problem [14]. Asif et al [15] proposed methods which can construct low-dimensional representation of large and diverse road networks, in presence of missing historical and neighboring data. They used low-dimensional models to reconstruct data profiles for road segments, and impute missing values. To this end these researchers used Fixed Point Continuation with Approximate SVD and Canonical Polyadic decomposition for incomplete tensors to solve the problem of missing data. Experimental evaluation showed that the low-dimensional models could perform data imputation with improved accuracy even in the presence of high percentage of missing data in large road networks. Although it is not suitable for time series prediction, but how to find the similar information and how to extract relevant information are worthy of references. Gao et al [16] proposed a new BP Neural Network algorithm for telecom missing data prediction. The result showed that the algorithm could be used to predict the missing value with good performance. Zhang and Liu [17] used least squares support vector machines to predict missing traffic flow based on spatio-temporal analysis in urban arterial streets. The proposed method showed its capacity of highly accurate computation, remarkable robustness to accurately predict missing values both at low and at high missing rates evidenced the applicability of least squares support vector machines to the traffic field.

In the power grid system research field, most researchers focused on support vector machine [18], artificial neural network [19], and grey model [20]. For the artificial neural networks, they usually need a lager training data set but might be not sufficient in the prediction of missing data as a matter of fact. Grey model is very popular in this field in recent years and works well. But grey model can't deal with the missing data prediction problem in a proper way. However, optimization in SVM algorithms is essential because the traditional support vector machine is not accurate enough in prediction.

C. A short review of SVM for regression

Vapnik and his coworkers originally developed SVM in the 1990s, which is proved to be successful in many significant fields [21]. Here we briefly describe the basic ideas behind SVM for regression.

Given a training data set $T = \{(x_i, y_i)\}_{i=1}^{N}$, collected sampling, some unknown function *g(x)*, with noise. The major objective of regression is to determine a function *f* that approximates *g(x)*, based on the knowledge of *T*. The core idea of SVM is to non-linearly map the input data to some high dimensional space, where the data can be linearly separated in this feature space. The SVM considers approximating functions as follows:

$$f(x, \omega) = \sum_{i=1}^{D} \omega_i \phi_i(x) + b, \tag{2}$$

where the function $\{\Phi_i(x)\}_{i=1}^{N}$ denotes features, and $\{\omega_i\}_{i=1}^{\infty}$ and *b* are coefficients that have to be estimated from the data. So, a linear regression in a feature space is transferred from a nonlinear regression in the low dimensional input space. The unknown coefficients $\{\omega_i\}_{i=1}^{D}$ could be determined from data by minimizing the following function:

$$R(\omega) = \frac{1}{N}\sum_{i=1}^{N} \left| y_i - f(x_i, \omega) \right|_{\varepsilon} + \lambda \|\omega\|^2, \tag{3}$$

where λ is a regulation constant and robust error function defined as

$$\left|y_i - f(x_i,\omega)\right|_\varepsilon = \begin{cases} 0 & if\,|\,y_i - f(x_i,\omega)\,| < \varepsilon \\ |\,y - f(x,\omega)\,| & otherwise \end{cases} \quad (4)$$

is called Vapnik's ε-insensitive loss function, which could minimize the functional in (3). The corresponding minimizing function form is:

$$f(x,\alpha,\alpha^*) = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i)K(x,x_i) + b\,, \quad (5)$$

where $\alpha_i\alpha_i^* = 0$, α_i, $\alpha_i^* \geq 0$ $i = 1, 2, 3, \ldots, N$ and the so called kernel function K describes the inner product in the feature space:

$$K(x,y) = \sum_{j=1}^{D}\phi_j(x)\phi_j(y) \quad (6)$$

Note the fact that the feature Φ_j does not need to be calculated, and the K is analytically known and has a very simple form. Polynomial, Gaussian, and sigmoidal kernels are mostly used in practice. The coefficients α_i, α_i^* are obtained by minimizing the following form:

$$\Omega(\alpha^*,\alpha) = \frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i^* + \alpha_i)(\alpha_i^* - \alpha_i)K(x_i,x_j) - \sum_{i=1}^{N}y_i(\alpha_i^* - \alpha_i) + \varepsilon\sum_{i=1}^{N}(\alpha_i^* + \alpha_i)\,, \quad (7)$$

subject to $\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0$, $0 \leq \alpha_i, \alpha_i^* \leq C$.

Only some of coefficients α_i, α_i^* will be different from zero, and the data points associated to them are called support vectors. Parameters like C and ε are free parameters and left to be determined by user. The main idea is to pick those values α_m, α_m^* for a point x_m on the margin. One x_m would be sufficient, but for stability purpose it is recommended that one takes the average over all points on the margin. More detailed description of SVM for regression can be found in References [22].

III. THE PROPOSED FRAMEWORK

In this paper, we focus on a machine learning framework OR_MLF based on optimized SVM and refined SVM, which takes the advantages of supervised learning and provides an improved fundamental prediction process.

Fig. 2 illustrates the overall structure of OR_MLF: 1) extracting features which include missing data from electronic data gathering equipment; 2) training and optimizing SVM; 3) refining the trained OSVM; 4) testing and comparison with the data in real world. Here, two different types of missing data prediction are to be identified. The first is to predict a single missing data and the second one is to predict consecutive missing data. Both of them could be found in real world.

A. *Extracting features (procedure1)*

For each corresponding sensor facility, its measured samples are represented as a discrete time series - observations are made at fixed time intervals and form a discrete set of values. In order to improve the effect of missing data prediction, contents of these time series are preprocessed through normalization, which replaces nonstandard data values with corresponding values that comply with the standard to adjust values measured on different scales to a notionally common scale. In each case, features $x_1, x_2, x_3, \ldots, x_M$ are extracted as follows:

$$x_k = a + \frac{(D_k - \min D_k)(b-a)}{\max D_k - \min D_k}\,, \quad (8)$$

where k=1, 2, 3, …, M, D_k is the authentic value of the above time series in one same feature. Besides, a and b are generalized to restrict the range of values in the dataset. Here, we set a=1 and b=2. Then, feature vector X= [x_1, x_2, x_3, …, x_M] for SVM is composed.

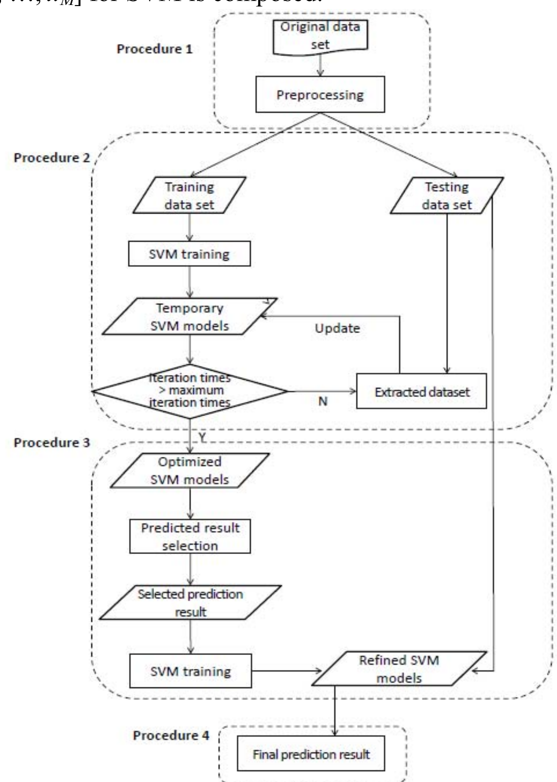

Figure 2. Flow chart of our missing data prediction framework.

B. *Training and Optimizing SVM (procedure2)*

The main process of training and optimizing SVM is shown in Fig. 3. Based on the training data set X after preprocessing, we first separately train each feature of the existing data set by traditional SVM and obtain preliminary SVM models. Second, the temporary prediction results could be achieved by the acquired SVM models. Next, the most relevant samples could be extracted by comparing the mean square errors (MSE) between real original data and the predicted data with respect to ∑ which is set artificially. Then, the training data set is updated by simply merging the extracted data set with the original data set. Finally, the optimized SVM is obtained by updating the preliminary SVM model based on the newly formed data set.

Procedure2. Train Optimized SVM

Input: testing data set TEST, preliminary SVM models SVMs, sampling proportion Σ, maximum iteration number maxIter

Output: optimized SVM OSVM

1: for each testing data set t_0 in TEST do:
2: for *k*=1: *maxIter* do:
3: p_0 = missing data prediction result of t_0 based on corresponding SVM model in SVMs
4: p_0' = sort the prediction result p_0 in ascending order
5: *Index* = find the index of p_0' in p_0
6: *ext_T* = extract data set $x'(k)$ corresponding to *Index* and Σ
7: *new_T* = merge the extracted data set *ext_T* with input testing dataset t_0
8: *OSVM* = update the SVMs(*k*) with *new_T*
9: end for
10:end for

Figure 3. Progress of training optimized SVM.

Taking x_1 for example, some data is lost in x_1 during some period. First we obtain an initial *SVM_*x_1 model by training traditional SVM with x_1. Next the corresponding prediction data p_0 of testing data t1 is calculated by *SVM_x1*, and then the testing data samples are reorganized in ascending order with respect to each sample's prediction error. The extracted samples *ext_T* are later obtained according to the sampling proportion Σ which is set before. The merged data set *new_T* is composed of original testing data set and *ext_T*. Finally, the new OSVM is achieved by updating the *SVM_x1* based on *new_T*. When the iteration times meet the maximum iteration number *maxIter*, the loop is stopped and the final optimized SVM is obtained.

C. Refining SVM model (procedure3)

Procedure 3. Refine SVM

Input: one subset of TEST: t_s, missing data prediction result *P* based on SVMs

Output: Refined SVM model RSVM

1: *E* = calculate the MSE of *P*
2: $[r_1, r_2]$ = find two most minimum result in *E*
3: $[p_1, p_2]$ = find two missing data prediction result corresponding with $[r_1, r_2]$
4: *R* = merge data set t_s with $[p_1, p_2]$
5: *RSVM* = train SVM with data set *R*

Figure 4. Progress of training refined SVM.

Given the fact that the time series of different features would result in different prediction result and taking the strong connection between some features into account, we attempt to refine the obtained SVM model to further narrow the missing data prediction error.

Again, we suppose that some data in x_1 is lost and a series of SVMs have been trained to predict the missing data. As Fig. 4 shows, the MSE E_x_1 of prediction result of respective OSVMs can be calculated first. Then, depending on E_x_1 we could choose the two best predictors $[p_1, p_2]$ among the SVMs and store the corresponding prediction result. Next, the new data set R_x_1 is merged by x_1 and $[p_1, p_2]$, and then the RSVM to predict missing data of x_1 is finally trained.

IV. Experimental Results

In this section, proposed framework OR_MLF is evaluated and the results are compared with existing methods. All the experiments are conducted in the same environment.

A. Case description

In power grid system, the states of transformers are reflected by the content information of some diagnostic gases, including hydrogen (H_2), methane (CH_4), acetylene (C_2H_2), ethylene (C_2H_4), ethane (C_2H_6), and carbon monoxide (CO). These five kinds of gases are designated as x_1, x_2, x_3, x_4, x_5, respectively. First, two hundred samples collected from a 220kV main transformer in State Grid of China are selected as the original training data set. The training data set which contains above five kinds of diagnostic gases is inputted into NN, LSSVM, SVM and OR_MLF respectively. All of the SVM models adopt Gaussian radial basis function as their kernel. Then, four missing data prediction models are produced for testing.

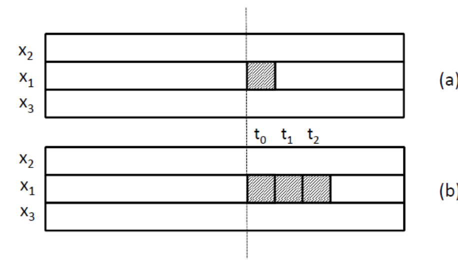

Figure 5. Two practical missing data situations in transformer monitor systems. The single shaded box represents one missing data.

Besides, testing samples t_1, t_2, t_3, t_4, and t_5 are similarly constructed by the above preprocessing procedure from the original data set. Next, to compare the performance of the framework OR_MLF with traditional NN, LSSVM and SVM, the testing samples are inputted to these four different models. Meanwhile, we also take two practical missing data situations into account. As illustrated in Fig. 5, the first situation shows that only one single data is lost during some period while the second indicates that consecutive data is lost.

B. Comparison between OR_MLF with OSVM and traditional methods

1) One single missing data prediction:

TABLE I. Mean Square Error (MSE) of NN, LSSVM, SVM and OSVM for Single missing data predicting in $X_1(H_2)$

	H_2	CH_4	C_2H_4	C_2H_6	CO
NN	0.0054	0.0049	0.0373	0.0312	0.0065
LSSVM	0.0512	0.0514	0.0622	0.0705	0.0484
SVM	0.0046	0.0047	0.0372	0.0379	0.0049
OR_MLF with OSVM	0.0039	0.0026	0.0368	0.0352	0.0042

As Fig. 5 (a) shows, suppose that only one single data in x_1 (H_2) at time t_0 is lost, we attempt to predict the missing data using NN, LSSVM, SVM, and OR_MLF with OSVM. In addition, considering the fact that strong association exists among the five different diagnostic gases, we also try to predict the missing data utilizing different gases, the data set of which might be complete at corresponding time, such as x_2 (CH_4) and x_3 (C_2H_4) in Fig. 5 (a).

To evaluate the performance of kinds of predictors, mean squared error (MSE) is employed to measure the average of the squares of the "errors", that is, the difference between the estimator and what is estimated. Taking testing data set t_1 (H_2) for example, the MSE_t_1 is calculated as follows:

$$MSE\_t_1 = \sqrt{\frac{\sum_{i=1}^{N_{t1}} (p_1^i - t_1^i)^2}{N_{t_1} - 1}} , \tag{9}$$

where t_l^i is one testing data sample and p_l^i is the missing data predicted result and N_{t1} is the total number of t_1.

First, the result in Table I shows that, compared with traditional NN, LSSVM and SVM, the accuracy of missing data prediction in x_1 based on OR_MLF with OSVM is improved from 0.0046 (SVM) to 0.0039. Besides, the result also demonstrates that it is feasible to predict the missing data taking advantage of different gases. For example, if data set x_2 (CH_4) is chosen to predict the missing data in x_1, the performance of NN and SVM is sufficient. Specifically, based on OR_MLF with OSVM, the MSE obtained using x_2 as prediction data (0.0026) is even lower than that is obtained using x_1 itself (0.0039), which means that the OR_MLF with OSVM is more adaptable to predict one missing data by different diagnostic gases than the neural network and other two kinds of SVMs.

2) Consecutive missing data prediction

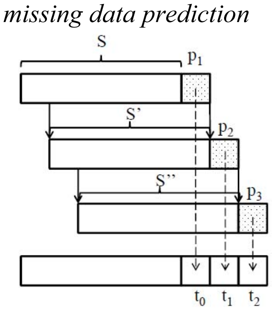

Figure 6. Consecutive missing data prediction method.

As Fig. 5 (b) shows, we also suppose that consecutive data in x_1 (H_2) at time t_0 is lost and endeavor to predict the missing data using NN, LSSVM, SVM, and OR_MLF with OSVM. The consecutive missing data prediction method is depicted in Fig. 6. Suppose that the prediction sample is S and three missing data is to be predicted, the first prediction value p_1 is obtained after S is inputted into one predictor, such SVM. Once new sample S' is produced, then it is inputted into the same predictor and p_2 is now predicted. Similarly, p_3 can also be predicted.

As shown in Fig. 7, for the problem of MSE calculation in terms of consecutive missing data prediction, the predicted data [P_1 P_2 P_3] at t_0, t_1, t_2 is first predicted and then the corresponding MSE is calculated respectively. Since the real original data remain unchanged, it is reasonable to add up the respective MSE at each time spot as the final MSE_s.

MSE Calculation Procedure for consecutive missing data prediction

Input: original data set X, SVM models SVM_m
Output: Summary of MSE MSE_s

1: Initial data $P_1 = []$, $P_2 = []$, $P_3 = []$
2: for each sample S in original data set X do:
3: [p_1 p_2 p_3] = predict three missing data of S using SVM_m
4: [P_1 P_2 P_3] = [P_1 P_2 P_3; p_1 p_2 p_3]
5: end for
6: [MSE_P_1 MSE_P_2 MSE_P_3] = calculate mean square error of [P_1 P_2 P_3]
7: MSE_s = add up MSE_P_1, MSE_P_2 and MSE_P_3
8: return MSE_s

Figure 7. Procedure of calculating consecutive missing data prediction.

In this part, we also try to predict the consecutive missing data utilizing different gases. First, compared with Table I, the result in Table II is totally and reasonably increased. However, most of the predictors based on SVM still present modest MSE. Furthermore, the lowest MSE 0.0126 is achieved utilizing the predictor trained by the dataset x_2, which again demonstrates that the OR_MLF with OSVM is more adaptable to predict the consecutive missing data by different diagnostic gases.

TABLE II. MEAN SQUARE ERROR (MSE) OF NN, LSSVM, SVM AND OSVM FOR CONSECUTIVE MISSING DATA PREDICTING IN X_1(H_2)

	H_2	CH_4	C_2H_4	C_2H_6	CO
NN	0.0185	0.0175	0.1	0.1185	0.0186
LSSVM	0.1302	0.1254	0.117	0.1131	0.1216
SVM	0.0156	0.0157	0.1972	0.1429	0.0207
OR_MLF with OSVM	0.0135	0.0126	0.2084	0.1479	0.0207

C. Comparison between OR_MLF with OSVM and traditional methods

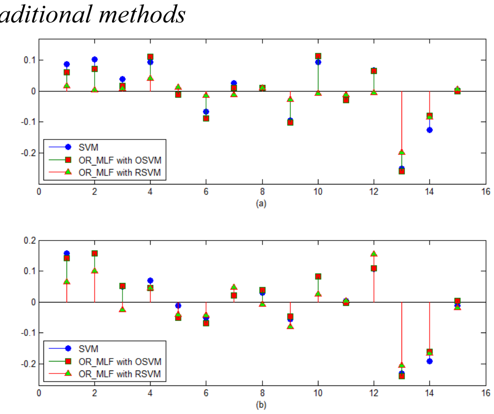

Figure 8. Missing data prediction errors of 15 randomly selected testing samples.

Since the OSVMs show better performance than NN, LSSVMs and SVMs, it is proper to select them as the fundamental SVMs to refine SVM models in OR_MLF.

SVM and OR_MLF with OSVM are chosen as the comparative objects due to their better performance compared with NN and LSSVM. Similarly, the missing data of x1 is predicted using SVM, OR_MLF with OSVM and OR_MLF with RSVM. To better show the performance of different predictors, fifteen testing samples are randomly selected and the corresponding prediction errors are presented in Fig. 8.

Fig. 8 (a) shows the prediction result in the situation when single missing data is to be predicted. The triangle shape stem is more close to the zero line, which means that the predictor under OR_MLF with RSVM outperforms SVM and OR_MLF with OSVM. As for the situation when consecutive missing data during some period is to be predicted, we attempt to predict three consecutive missing data and the prediction result of the third missing data in Fig. 8 (b) reveals that the OR_MLF performs better as well. On the whole, the MSE of the predictor under OR_MLF with RSVM is reduced to 0.0021, which is the best missing data prediction result in our experiment.

V. Conclusion

In this paper, we propose a novel approach for missing data prediction incorporating both optimized SVM and refined SVM to improve the quality of raw data generated by State Grid of China. As expected from our improved SVM approaches, our framework OR_MLF exhibits higher precision in terms of lower MSE than classic SVM and NN methods, e.g., the predictor under OR_MLF with RSVM obtains the lowest MSE 0.0021. Due to the nature of OSVM and RSVM, our framework keeps good robustness while requiring little training time, leading to a strong candidate for online missing data prediction in power grid equipment monitoring and evaluation. Though the OR_MLF framework works well with missing data prediction of power transformers, there are rooms for improvement, such as the optimization of training parameters, which need to be studied in the future.

Acknowledgment

This paper is sponsored in part by the National High Technology and Research Development Program of China (863 Program, 2015AA050204), State Grid Science and Technology Project (520626140020, 14H100000552), State Grid Corporation of China, and the National Research Foundation Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) program.

References

[1] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement". ACM Comput. Surv., 2009. **41**(3): p. 1-52.

[2] D. Luebbers, U. Grimmer, and M. Jarke, "Systematic Development of Data Mining-Based Data Quality Tools", in *Proceedings 2003 VLDB Conference*, J.-C. Freytag, P. Lockemann, S. Abiteboul, M. Carey, P. Selinger, and A. Heuer, Editors. 2003, Morgan Kaufmann: San Francisco. p. 548-559.

[3] V. Tresp and R. Hofmann. "Missing and noisy data in nonlinear time-series prediction". in Neural Networks for Signal Processing [1995] V. Proceedings of the 1995 IEEE Workshop. 1995.

[4] H. Ma, I. King, and M.R. Lyu, "Effective missing data prediction for collaborative filtering", in Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. 2007, ACM: Amsterdam, The Netherlands. p. 39-46.

[5] R. Nogueira, S.M. Vieira, and J.M.C. Sousa. "The prediction of bankruptcy using fuzzy classifiers". in Computational Intelligence Methods and Applications, 2005 ICSC Congress on. 2005.

[6] L. Kin Seng and W. Feng. "Pre-processing for missing data: A hybrid approach to air pollution prediction in Macau". in Automation and Logistics (ICAL), 2010 IEEE International Conference on. 2010.

[7] P. Sheng Kai, J. Joe Air, and C. Chia-Pang. "Conductor Temperature Estimation Using the Hadoop MapReduce Framework for Smart Grid Applications". in High Performance Computing and Communications, 2014 IEEE Intl Conf on. 2014.

[8] L.V. Ganyun, C. Haozhong, Z. Haibao, and D. Lixin, "Fault diagnosis of power transformer based on multi-layer SVM classifier". Electric Power Systems Research, 2005. **74**(1).

[9] N. Solomakhina, T. Hubauer, S. Lamparter, M. Roshchin, and S. Grimm. "Extending statistical data quality improvement with explicit domain models". in Industrial Informatics (INDIN), 2014 12th IEEE International Conference on. 2014.

[10] Z. Lu and Y.V. Hui, "L 1 linear interpolator for missing values in time series". Annals of the Institute of Statistical Mathematics, 2003. **55**(1): p. 197-216.

[11] J. Twisk and W. de Vente, "Attrition in longitudinal studies: How to deal with missing data". Journal of Clinical Epidemiology, 2002. **55**(4): p. 329-337.

[12] H. Junninen, H. Niska, K. Tuppurainen, J. Ruuskanen, and M. Kolehmainen, "Methods for imputation of missing values in air quality data sets". Atmospheric Environment, 2004. **38**(18): p. 2895-2907.

[13] H. Seung Tae and C. Jae Woo. "A New Data Filtering Scheme Based on Statistical Data Analysis for Monitoring Systems in Wireless Sensor Networks". in High Performance Computing and Communications (HPCC), 2011 IEEE 13th International Conference on. 2011.

[14] D.S.C. Fung, "Methods for the estimation of missing values in time series". 2006, Edith Cowan University Perth.

[15] M.T. Asif, N. Mitrovic, L. Garg, J. Dauwels, and P. Jaillet. "Low-dimensional models for missing data imputation in road networks". in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. 2013.

[16] G. Wei, N. Kun, C. Man, and G. Qingyang. "A data prediction algorithm based on BP neural network in telecom industry". in Computer Science and Service System (CSSS), 2011 International Conference on. 2011.

[17] Z. Yang and L. Yuncai. "Missing traffic flow data prediction using least squares support vector machines in urban arterial streets". in Computational Intelligence and Data Mining, 2009. CIDM '09. IEEE Symposium on. 2009.

[18] W. Tang, S. Almas, and Q.H. Wu. "Transformer Dissolved Gas Analysis Using Least Square Support Vector Machine and Bootstrap". in Control Conference, 2007. CCC 2007. Chinese. 2007.

[21] V. Tresp and R. Hofmann, "Nonlinear time-series prediction with missing and noisy data". Neural Comput., 1998. **10**(3): p. 731-747.

[23] M.H. Wang and C.P. Hung, "Novel grey model for the prediction of trend of dissolved gases in oil-filled power apparatus". Electric Power Systems Research, 2003. **67**(1): p. 53-58.

[24] C. Cortes and V. Vapnik, "Support-vector networks". Machine Learning, 1995. **20**(3): p. 273-297.

[25] V.N. Vapnik and V. Vapnik, "Statistical learning theory". Vol. 1. 1998: Wiley New York.