Multivariate Extreme Learning Machine Based AutoEncoder for Electricity Consumption Series Clustering

KAIHONG ZHENG1, JINGFENG YANG2, QIHANG GONG1, SHANGLI ZHOU1, LUKUN ZENG1, AND SHENG LI1

1Digital Grid Research Institute, China Southern Power Grid, Guangzhou 510663, China
2China Southern Power Grid Company Ltd., Guangzhou 510663, China

Corresponding author: Kaihong Zheng (zheng_kh@163.com)

This work was supported in part by the China Southern Power Grid Company Ltd., and in part by the China Southern Power Grid Digital Grid Research Institute Company Ltd.

ABSTRACT Multivariate electricity consumption series clustering can reflect the trend of power consumption changes in the past time period, which can provide reliable guidance for electricity production. The dimensionality reduction-based method is an effective technology to address this problem, which obtains the low-dimensional features of each variate or all variates for multivariate time series clustering. However, most existing dimensionality reduction-based methods ignore the joint learning of the common representations and the variable-based representations. In this paper, we build a multivariate extreme learning machine based autoencoder model for electricity consumption clustering (MELM-EC), which performs common representation learning and variable-based representation learning simultaneously. MELM-EC maps the common representation and multiple variable-based representations to the original multivariate time series and computes the common output weights within a few iterations. Experimental results on realistic multivariate time series datasets and multivariate electricity consumption series datasets demonstrate the effectiveness of the proposed MELM-EC model.

INDEX TERMS Extreme learning machine, representation learning, multivariate time series clustering, electricity consumption clustering.
first used VPCA to construct an informative representation, and then adopted the spatial weighted matrix distance to measure the similarity among all instances. Li [17] first used CPCA to obtain projection coordinate representations, and then adopted the reconstruction error to divide instances. Zheng et al. [18] performed robust matrix factorization on the similarity matrix among all instances and learned the low-dimensional representations. However, most existing dimensionality reduction-based methods ignore the joint learning of the common representations and the variable-based representations. Extreme learning machine based autoencoder (ELM-AE) is a simple and efficient single layer feed-forward neural networks (SLFNs) learning algorithm [19]–[21], where the outputs are equal to the inputs. In addition, ELM-AE obtains only one optimal solution without iterative learning. In this paper, we develop a novel multivariate extreme learning machine based autoencoder framework (MELM-EC) for the MECSC task, which is a joint learning framework for the common representations and the variable-based representations. Extreme learning machine based autoencoder (ELM-AE) is a simple and efficient single layer feed-forward neural networks (SLFNs) learning algorithm [19]–[21], where the outputs are equal to the inputs. In addition, ELM-AE obtains only one optimal solution without iterative learning. In this paper, we develop a novel multivariate extreme learning machine based autoencoder framework (MELM-EC) for the MECSC task, which is a joint learning framework for the common representations and the variable-based representations. As shown in Fig. 1, MELM-EC utilizes the common representations and the variable-based representations to reconstruct the multivariate electricity consumption series, which minimizes the reconstruction error and obtains the common output weights connecting the common representations and the outputs (i.e., the input multivariate electricity consumption series). We think that the common output weights of MELM-EC can represent the features of the input multivariate electricity consumption series. Moreover, unlike iterative learning methods (e.g., back propagation neural networks), MELM-EC utilizes the regularized least mean squares and the coordinate descent method to solve the objective function. In this way, the common output weights can be obtained after a few iterations. Furthermore, the main contributions are summarized as follows:

- This paper presents a multivariate extreme learning machine based autoencoder framework (MELM-EC) to simultaneously perform the common representation learning and the variable-based representation learning, where variable-based variable-based representation learning can help MELM-EC learn more informative common representations.
- MELM-EC utilizes the regularized least mean squares to calculate the output weights and performs clustering on the common representations. Experimental results on realistic datasets demonstrate that MELM-EC is an effective MECSC framework.

The remainder of the paper is organized as follows. Section II introduces the related work about extreme learning machines. Section III details the proposed MELM-EC framework, including the objective function and the inference process. In Section IV, experimental results show the feasibility of the proposed MELM-EC framework. Finally, some conclusions are given in the last section.

II. RELATED WORKS

The extreme learning machine (ELM) model is an efficient learning algorithm of SLFNs [22], [23]. ELM output weights can be determined analytically, while the input weights and biases of the hidden layer are randomly generated. Suppose $X \in \mathbb{R}^{N \times M}$ denotes the input instances, $a \in \mathbb{R}^{M \times J}$ denotes the input weights, and $b \in \mathbb{R}^{J}$ denotes the biases of the hidden layer. Then, the outputs of the hidden layer can be defined as:

$$H = f(Xa + b),$$

where $f(\cdot)$ denotes the activation function (e.g., $f(z) = \frac{1}{1+\exp(-z)}$ is the sigmoid function) and $H \in \mathbb{R}^{N \times J}$ denotes the outputs of the hidden layer (i.e., the hidden features of the inputs). ELM utilizes the least mean square (LMS) to calculate the output weights, which is used for classification and regression. Afterwards, He et al. [24] extends ELM for clustering, where the $k$-means clustering method is performed on the hidden features. Besides the good performance of ELM based algorithms, the ELM clustering method is also very convenient for implementation and computation. Recently, there are some works about ELM clustering.
As ELM has been demonstrated to have good performance in clustering, some researchers developed some novel models on the basis of ELM. For example, Zeng [25] introduced the adaptive graph regularization term into the unsupervised ELM framework and proposed an effective graph learning method. Zhang et al. [26] proposed a residual compensation ELM for regression problem, where residual compensation was performed layer by layer iteratively for remodeling the un-modeling prediction error in the previous layer to improve the accuracy further. Xu et al. [27] proposed a fuzzy granularity neighborhood extreme clustering algorithm on the basis of ELM, which used fuzzy neighborhood rough set to eliminate redundant attributes and introduced the adaptive adjustment mechanism to solve the parameters of unsupervised ELM.

Due to that the input weights and biases of the hidden layer are randomly generated, the hidden features of ELM sometimes cannot represent the input instances well. Extreme learning machine based autoencoder (ELM-AE) was a novel method of neural network which can reproduce the inputs as well as autoencoder [19], [28], [29]. In ELM-AE, the outputs are equal to the inputs, and the output weights can be also calculated by the LMS method. Suppose \( X \) denotes the input instances and \( H \) denotes the outputs of the hidden layer. Then, the objective function of ELM-AE is defined as:

\[
\min L_{ELM-AE} = \frac{1}{2} \| \beta \|^2 + \frac{\lambda}{2} \| X - H \beta \|^2,
\]

where \( \lambda \) is the regularization parameter. Taking \( \frac{\partial L_{ELM-AE}}{\partial \beta} = 0 \), we have

\[
\beta + \lambda H^T (X - H \beta) = 0.
\]

Therefore, the solution of the problem (2) is

\[
\beta = \begin{cases} 
(\frac{1}{\lambda} + H^T H)^{-1} H^T X, & N \geq J \\
H^T (\frac{1}{\lambda} + HH^T)^{-1} X, & N < J
\end{cases}
\]

In this way, the representations of ELM-AE can be defined as

\[
H_{new} = f(X \beta^T).
\]

To address this problem, we propose a novel multivariate extreme learning machine based autoencoder model for electricity consumption clustering (MELM-EC). As shown in Fig. 1, MELM-EC maps the common representation and multiple variable-based representations to the original multivariate time series and minimizes the reconstruction error. In other words, MELM-EC is a joint learning framework for the common representation and the variable-based representations, and it utilizes the common representation and the variable-based representations to reconstruct the corresponding variate series. In this way, the autoencoder mechanism and variable-based representation learning can help MELM-EC learn more informative common representations. Given a multivariate dataset \( X^{ME} = \{X^d\}_{d=1}^D \), where \( X^d \in \mathbb{R}^{N \times M^d} \), \( N \) denotes the number of instances \( X^{ME} \), \( D \) is the number of variables, and \( M^d \) is the length of the \( d \)-th variable \( X^d \). MELM-EC must first initialize the outputs of the variable-based hidden representations \( \{H^d\}_{d=1}^D \) and the outputs of the common hidden representations \( H^c \). Then the outputs of \( H^d \) can be expressed as

\[
H^d = f(X^d a^d + b^d),
\]

where the input weights \( a^d \in \mathbb{R}^{M^d \times J} \) and the biases \( b^d \in \mathbb{R}^J \) need to be orthogonalized, and \( f(\cdot) = \frac{1}{1 + \exp(-\cdot)} \) denotes the activation function. In addition, \( H^c \) is initialized with the mean of \( \{H^d\}_{d=1}^D \).

MELM-EC utilizes the common representation and the variable-based representations to reconstruct the multivariate electricity consumption series, where the objective function to be minimized is

\[
\min \quad L = \frac{1}{2} \sum_d \sigma_d \left( \|H^d \beta^d + H^c B^d - X^d \|_F^2 + \frac{\lambda_1}{2} \sum_d \sigma_d \| H^d - H^c \|_F^2 + \frac{\lambda_2}{2} \sum_d \sigma_d \| \beta^d \|_2^2 + \| B^d \|_F^2 \right).
\]

where \( \beta^d \) denotes the output weights connecting the \( d \)-th variable-based hidden layer and the output layer, \( B^d \) denotes the output weights connecting the common hidden layer and the output layer, \( \sigma_d \) is the parameter to balance different variables (e.g., \( \sigma_d = 1/D \) is used in this paper), \( \sum_d \sigma_d = 1 \), and \( \lambda_1, \lambda_2 \) are hyperparameters. We can see that the objective of MELM-EC consists of these terms: the first term is the reconstruction loss term that minimizes the projection error from the common representation and multiple variable-based representations to the original multivariate time series, the second term is the consistency constraint term that makes the common representations close to the variable-based representations, and the third term is the penalty term that penalizes larger values in the output weights. MELM-EC adopts the coordinate descent method to minimize the objective function. In MELM-EC, when a variable is updated, the values of other variables need to be fixed. Besides, MELM-EC utilizes the regularized least mean squares to obtain the variable, where the value of the variable can be directly calculated by the generalized inverse matrix. In other words, MELM-EC does not need to iteratively update...
the weights of the network like the BP network. The final common representations of MELM-EC can be defined as

$$H_{\text{new}}^c = f \left( \sum_d \alpha_d \left( X_d (B_d)^T \right)^T \right). \quad (8)$$

Therefore, we can perform clustering on the common representations $H_{\text{new}}^c$ (e.g., the clustering model with adaptive neighbors (CAN) [31] is used in this paper).

### A. LEARNING PROCEDURE OF MELM-EC

The objective of MELM-EC can be solved by the coordinate descent method, where four variables $H^c$, $\{H^d_d\}_{d=1}^D$, $\{\beta_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ are updated in turn. In addition, the detailed learning procedure of MELM-EC is provided in **Algorithm 1**.

**Algorithm 1** Learning procedure of MELM-EC

**Input:** A set of a multivariate dataset $X_{\text{ME}} = \{X_d\}_{d=1}^D$, the parameters $\lambda_1, \lambda_2, \{\alpha_d\}_{d=1}^D$, and the number of iterations $K$ (where $K \leq 5$ in this paper).

**Output:** The final common representations $H_{\text{new}}^c$.

1. Initialize the input weights $\{a^d_d\}_{d=1}^D$ and the biases $\{b^d_d\}_{d=1}^D$ that need to be orthogonalized.
2. Calculate the outputs of the variable-based hidden representations $\{H^d_d\}_{d=1}^D$ according to (5).
3. Initialize the common representations $H^c$ with the mean of the variable-based representations $\{H^d_d\}_{d=1}^D$, and $k=1$.
4. Initialize the output weights $\{\beta^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$:
   $$\begin{bmatrix} \beta^d_d \\ B_d^d \end{bmatrix} = \begin{cases} \left( \left[H^d_d H^c \right]^T \left[H^d_d H^c \right] + \lambda_2 I \right)^{-1} \left[H^d_d H^c \right]^T X_d, & N \geq 2J \\ \left[H^d_d H^c \right]^T \left[H^d_d H^c \right]^T + \lambda_2 I \right)^{-1} \left[H^d_d H^c \right]^T X_d, & N < 2J \end{cases} \quad (9)$$
5. while $k \leq K$ do
   6. Fix $\{H^d_d\}_{d=1}^D$, $\{\beta^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ and calculate $H^c$ according to (11).
   7. Fix $H^c$, $\{\beta^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ and calculate $\{H^d_d\}_{d=1}^D$ according to (14).
   8. Fix $H^c$, $\{H^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ and calculate $\{\beta^d_d\}_{d=1}^D$ according to (17).
   9. Fix $H^c$, $\{H^d_d\}_{d=1}^D$, $\{\beta^d_d\}_{d=1}^D$ and calculate $\{B_d^d\}_{d=1}^D$ according to (20).
10. end while
11. Calculate the final common representations $H_{\text{new}}^c$ based on (8).
12. return $H_{\text{ME}}$.

1) FIX $\{H^d_d\}_{d=1}^D$, $\{\beta^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ AND UPDATE $H^c$
If $\{H^d_d\}_{d=1}^D$, $\{\beta^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ is fixed, the problem (7) will become

$$\begin{align*}
\min_{H^c} & \frac{1}{2} \sum_d \alpha_d \left\| H^d_d \beta^d_d + H^c B_d^d - X_d \right\|_F^2 \\
& + \lambda_1 \sum_d \alpha_d \left\| H^c - H^d_d \right\|_F^2 \quad (9)
\end{align*}$$

Setting the derivative of (9) with respect to $H^c$ is equal to zeros, we have

$$\begin{align*}
\sum_d \alpha_d \left( H^d_d \beta^d_d + H^c B_d^d - X_d \right)^T \\
+ \lambda_1 \left( H^c - \sum_d \alpha_d H^d_d \right) = 0. \quad (10)
\end{align*}$$

Therefore, the solution of the common representations $H^c$ is

$$H^c = \left( \sum_d \alpha_d \left( X_d (B_d^d)^T + \lambda_1 H^d - H^d_d \beta^d_d (B_d^d)^T \right) \right) \cdot \left( \sum_d \alpha_d B_d^d (B_d^d)^T + \lambda_1 I \right)^{-1}. \quad (11)$$

2) FIX $H^c$, $\{\beta^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ AND CALCULATE $\{H^d_d\}_{d=1}^D$
If $H^c$, $\{\beta^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ is fixed, the problem (7) will become

$$\begin{align*}
\min_{\{H^d_d\}_{d=1}^D} & \frac{1}{2} \sum_d \alpha_d \left\| H^d_d \beta^d_d + H^c B_d^d - X_d \right\|_F^2 \\
+ \lambda_1 \sum_d \alpha_d \left\| H^c - H^d_d \right\|_F^2 \quad (12)
\end{align*}$$

Setting the derivative of (12) with respect to $H^d_d$ is equal to zeros, we have

$$\alpha_d \left( H^d_d \beta^d_d + H^c B_d^d - X_d \right)^T (\beta^d_d) + \alpha_d \lambda_1 (H^c - H^d_d) = 0. \quad (13)$$

Therefore, the solution of the variable-based hidden representations $H^d_d$ is

$$H^d_d = \left( X_d (\beta^d_d)^T + \lambda_1 H^c - H^c B_d^d (\beta^d_d)^T \right) \left( \beta^d_d (\beta^d_d)^T + \lambda_1 I \right)^{-1}. \quad (14)$$

3) FIX $H^c$, $\{H^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ AND CALCULATE $\{\beta^d_d\}_{d=1}^D$
If $H^c$, $\{H^d_d\}_{d=1}^D$, $\{B_d^d\}_{d=1}^D$ is fixed, the problem (7) will become

$$\begin{align*}
\min_{\{\beta^d_d\}_{d=1}^D} & \frac{1}{2} \sum_d \alpha_d \left\| H^d_d \beta^d_d + H^c B_d^d - X_d \right\|_F^2 \\
+ \lambda_2 \sum_d \alpha_d \left\| \beta^d_d \right\|_F^2 \quad (15)
\end{align*}$$

Setting the derivative of (15) with respect to $\beta^d_d$ is equal to zeros, we have

$$\left( H^d_d \right)^T \left( H^d_d \beta^d_d + H^c B_d^d - X_d \right) + \lambda_2 \beta^d_d = 0. \quad (16)$$

Therefore, the solution of the variable-based output weights $\beta^d_d$ is

$$\beta^d_d = \begin{cases}
\left( \left(H^d_d \right)^T H^c + \lambda_2 \right)^{-1} \left(H^d_d \right)^T \left(X_d - H^c B_d^d \right), & N \geq J \\
\left(H^c \right)^T \left(H^d_d \right)^T + \lambda_2 \right)^{-1} \left(X_d - H^c B_d^d \right), & N < J.
\end{cases} \quad (17)$$
TABLE 1. Performances of MELM-EC and baselines in terms of RI (rand index).

| Datasets                              | MC2PCA | SWMDFC | m-kShape | m-kSC | MLAN  | NESE  | MELM-EC |
|---------------------------------------|--------|--------|----------|-------|-------|-------|---------|
| ArticulatoryWordRecognition           | 0.9891 | 0.8939 | 0.7582   | 0.9510| 0.9899| 0.9756| 0.9909  |
| BasicMotions                          | 0.7910 | 0.7013 | 0.5244   | 0.7718| 0.7628| 0.7449| 0.8308  |
| Epilepsy                              | 0.6126 | 0.6666 | 0.5136   | 0.6044| 0.7330| 0.8897| 0.7145  |
| ERing                                 | 0.7563 | 0.7724 | 0.7701   | 0.7494| 0.7471| 0.7540| 0.7885  |
| Libras                                | 0.8920 | 0.8611 | 0.6605   | 0.9227| 0.9161| 0.9087| 0.9259  |
| NATOPS                                | 0.8818 | 0.7610 | 0.6534   | 0.8348| 0.7992| 0.7637| 0.9001  |
| StandWalkJump                         | 0.5905 | 0.7238 | 0.3485   | 0.6571| 0.7048| 0.6476| 0.7810  |
| UWaveGestureLibrary                   | 0.8828 | 0.8246 | 0.8015   | 0.9259| 0.9434| 0.8553| 0.9601  |

Mean±Std: 0.7995±0.1317 | 0.7756±0.0742 | 0.6288±0.1462 | 0.8021±0.1209 | 0.8245±0.1019 | 0.8174±0.1006 | 0.8615±0.0911

4) **FIX** \( H^c \), \( \{ H^d | d = 1 \} \), \( \{ \beta^d | d = 1 \} \) AND **CALCULATE** \( \{ B^d | d = 1 \} \)

If \( H^c \), \( \{ H^d | d = 1 \} \), \( \{ B^d | d = 1 \} \) is fixed, the problem (7) will become

\[
\min_{\{ B^d | d = 1 \}} \frac{1}{2} \sum_d \alpha_d \left( \| H^d \beta^d + H^c B^d - X^d \|_F^2 + \lambda_2 \| B^d \|_F^2 \right).
\]

(18)

Setting the derivative of (18) with respect to \( B^d \) is equal to zero, we have

\[
( H^c )^T ( H^d \beta^d + H^c B^d - X^d ) + \lambda_2 B^d = 0.
\]

(19)

Therefore, the solution of the common output weights \( B^d \) is

\[
B^d = \begin{cases} 
( ( H^c )^T ( H^c + \lambda_2 I )^{-1} ( H^d )^T ( X^d - H^d \beta^d ), N \geq J \\
( H^c )^T ( H^c ( H^c )^T + \lambda_2 I )^{-1} ( X^d - H^d \beta^d ), N < J 
\end{cases}
\]

(20)

**B. CONVERGENCE ANALYSIS**

In MELM-EC, four variates \( H^c \), \( \{ H^d | d = 1 \} \), \( \{ \beta^d | d = 1 \} \), \( \{ B^d | d = 1 \} \) are calculated iteratively. When computing \( H^c \), \( \{ H^d | d = 1 \} \), \( \{ \beta^d | d = 1 \} \), \( \{ B^d | d = 1 \} \), we have (11) (i.e., the common representations \( H^c \)) is an optimal solution of (9), (14) (i.e., the variable-based hidden representations \( \{ H^d | d = 1 \} \) is an optimal solution of (12), (17) (i.e., the variable-based output weights \( \{ \beta^d | d = 1 \} \) is an optimal solution of (15), and (20) (i.e., the common output weights \( \{ B^d | d = 1 \} \) is an optimal solution of (18). Specifically, like other alternating minimization (ADMM) methods [32], MELM-EC converges to a saddle point or a local minimum.

**IV. EXPERIMENTS**

**A. EXPERIMENTS ON MULTIPLE MULTIVARIATE TIME SERIES DATASETS**

1) **DATASETS AND BASELINES**

We perform evaluation using eight real-world multivariate time series datasets1 (i.e., ArticularyWordRecognition, BasicMotions, Epilepsy, ERing, Libras, NATOPS, StandWalkJump, and UWaveGestureLibrary), where the number of variates varies from 2 to 24 and the length of variates varies from 45 to 2500. Besides, We compare MELM-EC with eight representative methods, including MC2PCA [17], SWMDFC [15], m-kShape [33], m-KSC [33], MLAN [34], NESE [35], multimodal restricted Boltzmann machine (MRBM) [36], and multivariate ELM (MELM). In MLAN and NESE, each variate is regarded as a view. Similar to the work of [24], MELM maps multivariate time series to hidden features and performs k-means clustering on hidden features.

In MELM-EC, the parameter \( \lambda_1, \lambda_2 \) are selected from \( \{ 10^{-3}, 10^{-2}, \ldots, 10^3 \} \), \( \alpha_d = 1/D \) indicates that the importance of all variables is the same, the number of the hidden layer nodes \( J \) is selected from \( \{ 50, 100, 200, 500, 1000, 2000 \} \) and the number of iterations \( r \) is selected from \( \{ 1, 2, 3, 4, 5 \} \). In MC2PCA, the number of reduced dimensions is selected from \( \{ 1, 2, 3, 4, 5, 10 \} \). In m-kShape and m-KSC, the number of clusters is selected from \( \{ 2, \ldots, K \} \). In MRBM, the number of the hidden layer nodes \( J \) is selected from \( \{ 50, 100, 200, 500 \} \) and the number of iterations is set to 200. Two popular measurements [18], rand index (RI) and normalized mutual information (NMI), are used in this paper.

2) **PERFORMANCE COMPARISON**

We report the detailed experimental results of all comparing algorithms on eight datasets in Table 1 and Table 2, where the best results are highlighted in bold. The following observations can be made according to the results: 1) MELM-EC outperforms these baselines, where MELM-EC achieves 7
TABLE 2. Performances of MELM-EC and baselines in terms of NMI (normalized mutual information).

| Datasets                          | MC2PCA | SWMDFC | m-kShape | m-kSC | MLAN  | NESE  | MELM-EC |
|-----------------------------------|--------|--------|----------|-------|-------|-------|---------|
| ArticulatoryWordRecognition       | 0.9338 | 0.5233 | 0.3435   | 0.7922| 0.9435| 0.8486| 0.9558  |
| BasicMotions                     | 0.6736 | 0.5103 | 0.3405   | 0.5539| 0.5648| 0.5250| 0.7327  |
| Epilepsy                         | 0.1726 | 0.1900 | 0.1632   | 0.3811| 0.5318| 0.7600| 0.4044  |
| ERing                            | 0.3357 | 0.4223 | 0.2683   | 0.3478| 0.3185| 0.3778| 0.4391  |
| Libras                           | 0.5773 | 0.5001 | 0.4474   | 0.7244| 0.6141| 0.5418| 0.7554  |
| NATOPS                           | 0.6979 | 0.4716 | 0.3392   | 0.3998| 0.8247| 0.3139| 0.8441  |
| StandWalkJump                    | 0.3495 | 0.4829 | 0.1163   | 0.4606| 0.5545| 0.3985| 0.7090  |
| UWaveGestureLibrary              | 0.5701 | 0.4818 | 0.4194   | 0.7562| 0.8385| 0.5588| 0.8708  |

Mean±Std                           | 0.5388±0.2273| 0.4478±0.1015| 0.3047±0.1084| 0.5773±0.1606| 0.6488±0.1916| 0.5405±0.1738| 0.7139±0.1849|

TABLE 3. Comparative results between MELM-EC and baselines on RI (rand index).

| Datasets                          | Algorithm — A1: MC2PCA; A2: SWMDFC; A3: m-kShape; A4: m-kSC; A5: MLAN; A6: NESE; A7: MELM-EC |
|-----------------------------------|--------------------------------------------------------------------------------------------------|
| ArticulatoryWordRecognition       | A1 > A2, A1 > A3, A2 > A3, A4 > A2, A4 > A3, A5 > A2, A5 > A3, A6 > A2, A6 > A3, A7 > A2, A7 > A3 |
| BasicMotions                     | A1 > A2, A1 > A3, A2 > A3, A4 > A2, A4 > A3, A5 > A2, A5 > A3, A6 > A2, A6 > A3, A7 > A2, A7 > A3 |
| Epilepsy                         | A1 > A3, A2 > A1, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| ERing                            | A1 > A3, A2 > A1, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| Libras                           | A1 > A3, A2 > A1, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| NATOPS                           | A1 > A2, A1 > A3, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| StandWalkJump                    | A1 > A3, A2 > A1, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| UWaveGestureLibrary              | A1 > A2, A1 > A3, A2 > A4, A2 > A3, A3 > A4, A4 > A3, A5 > A4, A5 > A3, A6 > A4, A6 > A3, A7 > A4, A7 > A3 |
| Total order                      | A7(30) > A5(14) > A6(5) > A4(4) > A1(3) > A2(15) > A3(41) |

TABLE 4. Comparative results between MELM-EC and baselines on NMI (normalized mutual information).

| Datasets                          | Algorithm — A1: MC2PCA; A2: SWMDFC; A3: m-kShape; A4: m-kSC; A5: MLAN; A6: NESE; A7: MELM-EC |
|-----------------------------------|--------------------------------------------------------------------------------------------------|
| ArticulatoryWordRecognition       | A1 > A2, A1 > A3, A2 > A3, A4 > A2, A4 > A3, A5 > A2, A5 > A3, A6 > A2, A6 > A3, A7 > A2, A7 > A3 |
| BasicMotions                     | A1 > A2, A1 > A3, A2 > A3, A4 > A2, A4 > A3, A5 > A2, A5 > A3, A6 > A2, A6 > A3, A7 > A2, A7 > A3 |
| Epilepsy                         | A1 > A3, A2 > A1, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| ERing                            | A1 > A3, A2 > A1, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| Libras                           | A1 > A3, A2 > A1, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| NATOPS                           | A1 > A2, A1 > A3, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| StandWalkJump                    | A1 > A3, A2 > A1, A2 > A3, A3 > A2, A4 > A3, A4 > A2, A5 > A3, A5 > A2, A6 > A3, A6 > A2, A7 > A3, A7 > A2 |
| UWaveGestureLibrary              | A1 > A2, A1 > A3, A2 > A4, A2 > A3, A3 > A4, A4 > A3, A5 > A4, A5 > A3, A6 > A4, A6 > A3, A7 > A4, A7 > A3 |
| Total order                      | A7(38) > A5(21) > A4(6) > A1(0) > A6(−6) > A2(−13) > A3(−46) |

best performances on 8 datasets. Moreover, MELM-EC outperforms these baselines in terms of “Mean±Std”, and MELM-EC achieves the least improvements of 4.48% RI and 10.03% NMI; 2) MELM-EC obviously outperforms MLAN and NESE, which indicates that MELM-EC can obtain effective common representations for clustering. To summarize, comparing the proposed MELM-EC against these baselines, our method significantly improved previous results. To make a clearer view of the relative performance between each algorithm, the partial order \(\succ\) measures the relative performance between two algorithms \(A_1\) and \(A_2\) based on pair wise t-test at 5% significance level. If \(A_1\) \(\succ\) \(A_2\) holds, then \(A_1\) is rewarded by a positive score +1 and \(A_2\) is penalized by a negative score −1 [37]. Taking all the partial orders into consideration, the total order \(\succ\) of these algorithms is defined in terms of the accumulated score of each algorithm. Table 3 and Table 4 shows that our proposed MELM-EC presents the best performance in terms of RI and NMI. We also used t-SNE [38] to visualize the common representation \(H^c\) on ArticulatoryWordRecognition. As shown in Fig. 2, it can be seen that there are few outliers in MELM-EC, which also demonstrates its feasibility.
Then, we report performance comparisons between MELM-EC and artificial neural network-based methods (MELM and MRBM) on eight real-world datasets. Fig. 3 visually shows experimental results of MELM-EC and
artificial neural network-based methods. We can see that:
1) MELM-EC and MRBM are better than MELM on eight real-world datasets. Compared with MELM, MELM-EC and MRBM use the autoencoder mechanism to guide the training of the model. This shows that MELM-EC and MRBM can learn more effective features by using the autoencoder mechanism; 2) MELM-EC significantly outperforms MRBM on eight real-world datasets. Unlike MRBM, MELM-EC is a joint learning framework for the common representation and the variable-based representations. This illustrates that variable-based representation learning helps MELM-EC learn more informative common representations.

We analyze the impact of the number of iterations $K$ and the number of the hidden layer nodes $J$ on the algorithm performance. Fig. 4 and Fig. 5 show the RI and NMI performances of MELM-EC on eight real-world datasets as the number of iterations $K$ increases. It can be seen that MELM-EC achieves the better performance on a few iterations, which can greatly reduce the computational complexity of MELM-EC. Fig. 6 and Fig. 7 show the RI and NMI performances of MELM-EC on eight real-world datasets as the number of the hidden layer nodes $J$ increases. It can be seen that MELM-EC achieves the better performance when $J$ is less than or equal to 200, which also can reduce the computational complexity. Moreover, it seems that the performance of MELM-EC is not very sensitive to the number of hidden layer nodes.

B. EXPERIMENTS ON MULTIVARIATE ELECTRICITY CONSUMPTION SERIES DATASET

1) MULTIVARIATE ELECTRICITY CONSUMPTION SERIES DATASET

The realistic data of a region in China Southern Power Grid (CSPG) [6] is used to test our proposed MELM-EC.
The attribute information of the CSPG dataset is given as follows:
- global-active-power: household global minute-averaged active power (in kilowatt);
- global-reactive-power: household global minute-averaged reactive power (in kilowatt);
- voltage: minute-averaged voltage (in volt);
- global-intensity: household global minute-averaged current intensity (in ampere).

Specifically, CSPG contains 2780 multivariate instances of two class (i.e., two different night time periods, including 2 am to 4 am and 8 pm to 10 pm), where the number of variates is 4 and the length of variates is 18.

2) PERFORMANCE COMPARISON

We first test the performance of our proposed MELM-EC on different series combinations of CSPG. Table 5 shows the RI and NMI performances of MELM-EC on CSPG and the combinations containing arbitrary two variates. We can see that MELM-EC achieves the best results on CSPG which consists of four variates, which indicates each variate that can improve the performance of the model. Furthurmore,

| Combination | MELM-EC RI | MELM-EC NMI |
|-------------|-----------|-------------|
| ✓ ✓ ✓ ✓ | 0.7624    | 0.4535      |
| ✓ ✓ ✓ x | 0.7695    | 0.4729      |
| ✓ ✓ x ✓ | 0.7508    | 0.4226      |
| ✓ x ✓ ✓ | 0.5538    | 0.1669      |
| x ✓ x ✓ | 0.7624    | 0.4556      |
| x x ✓ ✓ | 0.7742    | 0.4802      |
| ✓ ✓ ✓ ✓ | 0.7839    | 0.5050      |

Fig. 8 visually shows experimental results of MELM-EC and contrast algorithms on the CSPG dataset. We can see that MELM-EC significantly outperforms contrast algorithms in terms of RI and NMI. Next, we analysis the impact of the number of the hidden layer nodes \( J \) on the CSPG dataset. Fig. 9 shows the RI and NMI performances of MELM-EC on CSPG as the number of the hidden layer nodes \( J \) increases. It can be seen that MELM-EC achieves the best performance when the number of the hidden layer nodes is equal to 50.

V. CONCLUSION

This paper builds an effective multivariate extreme learning machine-based autoencoder framework (MELM-EC) for multivariate electricity consumption clustering, which is a joint learning framework for the common representations and the variable-based representations. The proposed MELM-EC framework utilizes the regularized least mean squares to calculate the output weights and performs clustering on the common representations, where the common representations
and the variable-based representations are used to reconstruct the multivariate electricity consumption series. In addition, the common output weights can be obtained after a few iterations. Experimental results on realistic datasets demonstrate that MELM-EC is an effective MECSC framework.

REFERENCES

[1] A. Ahmadi, A. Mozafarinia, and A. Mohebi, “Clustering of multivariate time series data using particle swarm optimization,” in Proc. Int. Symp. Artif. Intell. Signal Process. (AISP), Mar. 2015, pp. 176–181.

[2] M. Forghani and R. Khani, “A multivariate clustering of AIXindex database for protein numerical representation,” in Proc. 3rd Iranian Conf. Intell. Syst. Signal Process. (ICSPIS), Dec. 2017, pp. 1–4.

[3] S. Hu, Z. Shi, and Y. Ye, "DMDB: Dual-correlated multivariate information bottleneck for multiview clustering," IEEE Trans. Cybern., early access, Oct. 21, 2020, doi: 10.1109/TCYB.2020.3025636.

[4] Y. Tang, Y. Xie, X. Yang, J. Niu, and W. Zhang, "Tensor multi-elastic kernel self-paced learning for time series clustering," IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 1223–1237, Mar. 2021.

[5] L. Zhang, T. Pei, B. Meng, Y. Lian, and Z. Jin, "Two-phase multivariate time series clustering to classify urban rail transit stations," IEEE Access, vol. 8, pp. 167998–168007, 2020.

[6] K. Zheng, P. Li, S. Zhou, W. Zhang, S. Li, L. Zeng, and Y. Zhang, "A multi-scale electricity consumption prediction algorithm based on time-frequency variational autoencoder," IEEE Access, vol. 9, pp. 90937–90946, 2021.

[7] F. Liu, Y. Lu, and M. Cai, "A hybrid method with adaptive sub-series clustering and attention-based stacked residual LSTMs for multivariate time series forecasting," IEEE Access, vol. 8, pp. 62423–62438, 2020.

[8] L. Zhou, G. Du, D. Tao, H. Chen, J. Cheng, and L. Gong, "Clustering multivariate time series data via multi-nomogeneous matrix factorization in multi-relational networks," IEEE Access, vol. 6, pp. 74747–74761, 2018.

[9] J. Liang and Y. Zhou, “Clustering multivariate time series from large sensor networks,” in Proc. Int. Conf. Comput. Syst., Electron. Control (ICCCSEC), Dec. 2017, pp. 563–567.

[10] D. Ienco and R. Interdonato, “Deep multivariate time series embedding clustering via attentive-gated autoencoder,” in Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining, 2020, pp. 318–329.

[11] M. J. Hasan, M. M. M. Islam, and J.-M. Kim, “Multi-sensor fusion-based time-frequency imaging and transfer learning for spherical tank crack diagnosis under variable pressure conditions,” Measurement, vol. 168, Jan. 2021, Art. no. 108478.

[12] A. A. Safi, C. Beyer, V. Unnikrishnan, and M. Spiliopoulou, “Multivariate time series as images: Imputation using convolutional denoising autoencoder,” in Advances in Intelligent Data Analysis XVIII, Konstanz, Germany: Springer, 2020, pp. 1–13.

[13] J. Y. Franceschi, A. Dieuleveut, and M. Jaggi, “Unsupervised scalable representation learning for multivariate time series,” in Proc. 33rd Conf. Neural Inf. Process. Syst., 2019, pp. 1–12.

[14] H. Li, “Accurate and efficient classification based on common principal components analysis for multivariate time series data,” Neurocomputing, vol. 171, pp. 744–753, Jan. 2016.

[15] H. He and Y. Tan, “Unsupervised classification of multivariate time series using VPCA and fuzzy clustering with spatial weighted matrix distance,” IEEE Trans. Cybern., vol. 50, no. 3, pp. 1096–1105, Mar. 2020.

[16] Z.-X. Li, J.-S. Guo, X.-B. Hui, and F.-F. Song, “Dimension reduction method for multivariate time series based on common principal component,” Control Decis., vol. 28, no. 4, pp. 531–536, Apr. 2013.

[17] H. Li, “Multivariate time series clustering based on common principal component analysis,” Neurocomputing, vol. 349, pp. 239–249, Jul. 2019.

[18] K. Zheng, H. Liang, L. Zeng, X. Chen, S. Li, H. Jiang, Q. Gong, S. Li, J. Yang, and S. Zhou, “Robust graph factorization for multivariate electricity consumption series clustering,” Math. Problems Eng., vol. 2021, pp. 1–12, Aug. 2021.

[19] A. H. Z. L. C. Kasun, G. B. Huang, and C. Vong, “Representational learning with extreme learning machine for big data,” IEEE Infell. Syst., vol. 28, no. 6, pp. 31–34, Dec. 2013.

[20] S. K. Padhy, B. K. Panigrahi, P. K. Ray, A. K. Satpathy, R. P. Nanda, and A. Nayak, “Classification of faults in a transmission line using artificial neural network,” in Proc. Int. Conf. Inf. Technol. (ICIT), Dec. 2018, pp. 239–243.

[21] B. K. Panigrahi, P. K. Ray, P. K. Rout, A. Mohanty, and K. Pal, “Detection and classification of faults in a microgrid using wavelet neural network,” J. Inf. Optim. Sci., vol. 39, no. 1, pp. 327–335, Jan. 2018.

[22] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” Neurocomputing, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[23] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, “Robust extreme learning machine for modeling with unknown noise,” J. Franklin Inst., vol. 357, no. 14, pp. 9885–9908, Sep. 2020.

[24] Q. He, X. Jin, C. Du, F. Zhang, and Z. Shi, “Clustering in extreme learning machine feature space,” Neurocomputing, vol. 128, no. 5, pp. 88–95, 2014.

[25] Y. Zeng, J. Chen, Y. Li, Y. Qing, and G.-B. Huang, “Clustering via adaptive and locality-constrained graph learning and unsupervised ELM,” Neurocomputing, vol. 80, pp. 224–235, Aug. 2020.

[26] J. Zhang, W. Xiao, Y. Li, and S. Zhang, “Residual compensation extreme learning machine for regression,” Neurocomputing, vol. 311, pp. 126–136, Oct. 2018.

[27] S. Xu, S. Liu, and L. Feng, “Fuzzy granularity neighborhood extreme clustering,” Neurocomputing, vol. 379, pp. 236–249, Feb. 2020.

[28] N. Zhang, S. Ding, and Z. Shi, “Denoising Laplacian multi-layer extreme learning machine,” Neurocomputing, vol. 171, pp. 1066–1074, Jan. 2016.

[29] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, “Non-iterative and fast deep learning: Multilayer extreme learning machines,” J. Franklin Inst., vol. 357, no. 13, pp. 8925–8955, 2020.

[30] L. Wang and S. Ding, “Multi-view spectral clustering via ELM-AE ensemble features representations learning,” IEEE Access, vol. 8, pp. 198679–198690, 2020.

[31] F. Nie, X. Wang, and H. Huang, “Clustering and projected clustering with adaptive neighbors,” in Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2014, pp. 977–986.

[32] A. Gunawardana and W. Byrne, “Convergence theorems for generalized alternating minimization procedures,” J. Mach. Learn. Res., vol. 6, pp. 2049–2073, Dec. 2005.

[33] M. Özer, A. Sapiezna, A. Abeiliuk, G. Muric, and E. Ferrara, “Discovering patterns of online popularity from time series,” Expert Syst. Appl., vol. 151, Aug. 2020, Art. no. 113337.

[34] F. Nie, G. Cai, and X. Li, “Multi-view clustering and semi-supervised classification with adaptive neighbours,” in Proc. AAAI Conf. Artif. Intell., 2017, pp. 2408–2414.

[35] Z. Hu, F. Nie, R. Wang, and X. Li, “Multi-view spectral clustering via integrating nonembbeding embedding and spectral embedding,” Inf. Fusion, vol. 55, pp. 251–259, Mar. 2020.

[36] N. Srivastava and R. Salakhutdinov, “Multimodal learning with deep Boltzmann machines,” in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 2222–2230.

[37] N. Zhang, S. Ding, and J. Zhang, “Multi layer ELM-RBF for multi-label learning,” Appl. Soft Comput., vol. 43, pp. 535–545, Jun. 2016.

[38] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” J. Mach. Learn. Res., vol. 9, pp. 2579–2605, Nov. 2008.
JINGFENG YANG received the master’s degree from the South China University of Technology, China, in 2009. He is currently a Senior Engineer with China Southern Power Grid. His research interests include electrical energy measurement automation systems, electricity information acquisition systems, electric energy measurement, and big data analysis technology.

QIHANG GONG received the master’s degree from Zhejiang University, China, in 2020. He is currently an Engineer with the Digital Grid Research Institute, China Southern Power Grid Company Ltd. His research interests include electric energy measurement, electrical energy measurement automation systems, electricity information acquisition systems, and electricity technology.

SHANGLI ZHOU received the bachelor’s degree from Tsinghua University, China, in 1991. He is currently a Professor Level Senior Engineer with the Digital Grid Research Institute, China Southern Power Grid. His research interests include electric energy measurement, electrical energy measurement automation systems, electricity information acquisition systems, and electricity technology.

LUKUN ZENG was born in 1992. He received the B.S. degree in communication engineering and the Ph.D. degree in electrical engineering from North China Electric Power University, Beijing, China, in 2013 and 2019, respectively. He is currently working with the Digital Grid Research Institute, China Southern Power Grid. His research interests include electrical energy measurement automation systems and data mining technology.

SHENG LI received the master’s degree from Hunan University, China, in 2019. He is currently an Engineer with the Digital Grid Research Institute, China Southern Power Grid. His research interests include electric energy measurement, electrical energy measurement automation systems, electricity information acquisition systems, and electricity technology.

***