Improved Deep Belief Network and Model Interpretation Method for Power System Transient Stability Assessment

Shuang Wu, Le Zheng, Wei Hu, Rui Yu, and Baisi Liu

Abstract—The real-time transient stability assessment (TSA) and emergency control are effective measures to suppress accident expansion, prevent system instability, and avoid large-scale power outages in the event of power system failure. However, real-time assessment is extremely demanding on computing speed, and the traditional method is not competent. In this paper, an improved deep belief network (DBN) is proposed for the fast assessment of transient stability, which considers the structural characteristics of power system in the construction of loss function. Deep learning has been effective in many fields, but usually is considered as a black-box model. From the perspective of machine learning interpretation, this paper proposes a local linear interpreter (LLI) model, and tries to give a reasonable interpretation of the relationship between the system features and the assessment result, and illustrates the conversion process from the input feature space to the high-dimension representation space. The proposed method is tested on an IEEE new England test system and demonstrated on a regional power system in China. The result demonstrates that the proposed method has rapidity, high accuracy and good interpretability in transient stability assessment.

Index Terms—Transient stability assessment (TSA), representation learning, deep belief network (DBN), local linear interpretation (LLI), visualization, emergency control.

I. Introduction

THE stability of power system is of great significance to the industry and social life. Therefore, maintaining the stability of power system has always been the focus of researchers and engineers. According to the discussions in [1], transient fault is one of the most common and fatal threats to stability of the power system. This concern has been raised and studied since the establishment of the first power system more than a century ago. The energy-based direct method and the time-domain simulation are the two main methods to analyze transient stability [2]. However, many studies have shown that these two methods can not meet the requirements of real-time transient stability assessment (TSA) for applications of large-scale power system.

With the rapid development of artificial intelligence (AI), the AI technologies are being increasingly applied to TSA. The large number of data required in AI model can be obtained through offline simulation of power systems and historical operation data. After the model training is completed, the online measurement data is used to perform power system real-time TSA [3]-[18].

In the scope of machine learning in AI, the TSA problem is treated as a binary classification task, i.e., approximating the transient stability boundary and determining whether the fault clearing state is inside the boundary or not. However, the stability of complex power system is impossible to express in analytical form and the transient stability boundary approximation is highly sophisticated and computationally intensive. Neural networks (NNs) [3]-[6], clustering [7], [8], decision trees (DTs) [9]-[14], and support vector machines (SVMs) [15], [16] as well as the least absolute shrinkage and selection operator (LASSO) [17] are the most frequently used machine learning algorithms. DT and LASSO have simple structures and are easy to interpret. However, in the situation where the classification boundary is highly nonlinear, these methods often fail to achieve good performance. SVM and NN are nonlinear methods, but the predictive ability of SVM is sensitive to the kernel function selection, and advance assumptions about the data distribution need to be made. For conventional shallow NN, its predictive capacity and ability are limited subject to the layer and parameter numbers.

One of the limitations of the conventional machine learning algorithms lies in the feature engineering process. In the existing applications in TSA, the power system variables such as the power angles of generators [5], [9], the active and reactive power flows [10] or the voltage trajectories [7] are directly fed into the machine learning models. Since the stability boundary comprises these features in a highly nonlinear and complex way, it is very difficult to estimate the boundary in the original feature space. Several studies have noticed the problem and tried to use kernel methods such as

S. Wu and W. Hu (corresponding author) are with State Key Laboratory of Power System, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China (e-mail: 523652867@qq.com; huwei@mail.tsinghua.edu.cn).

L. Zheng is with Stanford University, Stanford, California 94305, USA (e-mail: zhengl07@stanford.edu).

R. Yu and B. Liu are with the Southwest Branch, State Grid Corporation of China, Chengdu 610041, China (e-mail: 1458405947@qq.com; 16907860@qq.com).

SVM [15], or the unsupervised embedding methods such as principal component analysis (PCA) [19], to learn higher level representations from the raw data before using a classifier to assess the stability. However, the kernel methods are biased by the design and selection of kernels, while PCA is not appropriate for classifying different categories because the component contributing small principal may often contain important information about sample differences. Hence, a better representative learning method is needed to map the original feature space to a high-dimension representation space to make a more accurate and robust estimation of the boundary.

A deep belief network (DBN) [20] is a non-parametric, multi-layer NN formed by a stack of restricted Boltzmann machines (RBMs) for hierarchical representation learning. DBN is capable of learning complex representations from data due to two advantages. One is its structure of multiple layers which abstract the features layer by layer to get the high-dimension representation. This means the feature abstract process requires no guidance of human feature engineering. In the traditional energy-based TSA method, the energy function comprises potential and kinetic energies which are formulated by the state variables. Furthermore, the state vari-

ables can be represented by the measured algebraic variables. Hence, the above process of energy-based method is hierarchical and similar to the working logic of DBN. The second advantage is that the feature abstract process can automatically learn the data distribution by unsupervised learning methods, which is robust and can be easily generalized. The comparison among DBN, SVM, and PCA is shown in Table I.

The deep learning method has been tremendously successful in many fields such as electroencephalography [21] and drug discovery [22]. So far, some researchers have tried to apply the deep learning algorithm to the study of TSA. In [23], the stacked sparse autoencoder is utilized to predict the post-fault transient stability status of power system. This autoencoder is fed by specific points extracted from the fault on voltage magnitude measurements. However, the special points need to be obtained after getting the voltage trajectory, and the prediction method is still used as a black-box model. In [24], the authors claim to use DBN in TSA for the first time, but the method is not applied to any actual large system. Also, the features used for training are mostly related to the angular velocity and kinetic energy of the generator rotor, which are not easily observable or measurable.

TABLE I
COMPARISON OF REPRESENTATION LEARNING METHODS

Method	Supervised or not	Parametric or not	Linear or not	Data distribution
DBN	Unsupervised/supervised	Non-parametric	Non-linear	Unsupervised pre-training
SVM	Supervised	Parametric	Non-linear	Pre-defined
PCA	Unsupervised	Non-parametric	Linear	None

Representation learning has been popular in recent years, and there are many relative researches [25]. This paper introduces a DBN-based method to handle the transient stability problem based on our previous work in [26]. Apart from giving a more detailed DBN model building and training process considering power system structure characteristics, this paper proposes a local linear interpreter model and tries to illustrate the deep learning model reasonably from two aspects of input-output analysis and visual analysis separately. Since DBN is a highly complex model and the model decision is difficult to understand and interpret for system operators, the proposed interpreter can give an interpretation of the relationship between the system features and the assessment result, and it can also help guide the emergency control strategies if the system is unstable.

The rest of paper is organized as follows. The basic framework and training process are discussed in detail in Section II. In Section III, the interpretation method is given. Section IV proposes the whole procedure for processing the data and the assessing the system stability. An illustrative example of IEEE new England test system and the application in a regional power system of China are presented in Section V. Finally, Section VI discusses conclusions and future work.

II. DBN AND LEARNING ALGORITHM

A. Structure of DBN

A DBN is a multi-layer NN, where any two adjacent layers

can be seen as RBM [20], [27]-[30]. RBM is a generative stochastic NN with an input layer consisting of visible nodes and an output layer consisting of hidden nodes. RBM is designed to learn a probability distribution with hidden nodes $h_k (k = 1, 2, \cdots, K)$ over its visible nodes $v_i (i = 1, 2, \cdots, N)$. The connections are restricted between the visible and hidden nodes, as shown in Fig. 1. By merging one RBM's output layer to another RBM's input layer, more abstract representations can be learned by successive RBMs. Connecting a stack of RBMs to form a DBN, with the representations learned by each RBM hierarchically.
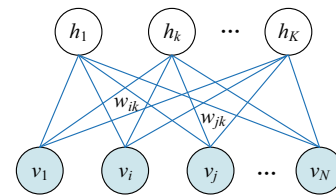

Fig. 1.   Basic structure of RBM.

B. Training Process of DBN

Training a DBN consists of three steps. The first step is the unsupervised pre-training of each RBM, i.e., with no stability information to guide the training process. Then, the trained RBMs are unfolded to form DBN. At last, expected classification accuracy index is used to fine-tune the parameters of DBN [31], [32]. The training process of a DBN is

shown in Fig. 2 and the training direction is shown by the arrow in the figure.

1) Unsupervised Pre-training

As stated above, an RBM is a two-layer NN utilized to represent the visible units by the hidden units. As a typical probability generation model, compared with discriminant model, it is built to establish a joint distribution between the observation data and the label. The connection from input layer to hidden layer represents discriminant mode, and the connection from hidden layer to input layer represents gener-

ation mode. The relationship between the RBM input layer and hidden layer defines the energy function of the system as (1), which is given by Hinton [20], [33].

$$E\left(v,h\,|\,\theta\right) = -\left(a^{\mathrm{T}}v + b^{\mathrm{T}}h + v^{\mathrm{T}}Wh\right) \qquad (1)$$

where v and h are the visible units and hidden units, respectively; W is the weight matrix of the connection between the visible and hidden units; a and b are the bias of the visible and hidden units, respectively; and $\theta = [\,a_i, b_j, w_{ij}\,]$ is the parameter of RBM.
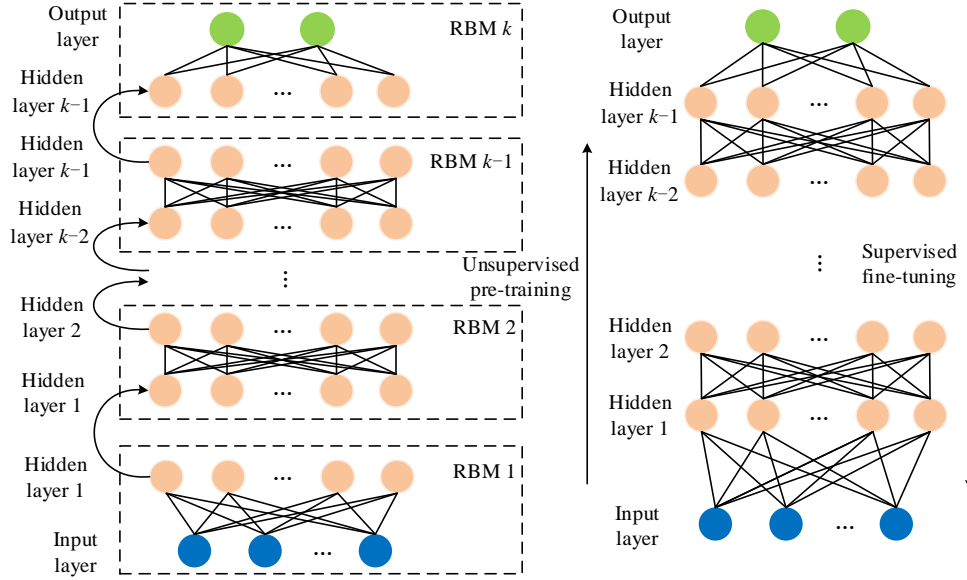

Fig. 2.   Structure of DBN and learning process.

The joint probability distribution over v and h is defined as:

$$P\left(v,h\,|\,\theta\right) = \frac{1}{Z(\theta)}\mathrm{e}^{-E\left(v,h\,|\,\theta\right)} \qquad (2)$$

where $Z(\theta)$ is a normalization factor defined as $Z(\theta) = \sum_v\sum_h \mathrm{e}^{-E\left(v,h\,|\,\theta\right)}$ to ensure that the sum of all possible probability distributions equals to 1. In this paper, v is a scalar vector of power system measurements at the time of fault clearance, including the active power P, reactive power Q, bus voltage U and angle θ. The reason for choosing these variables is that they are easy to be measured and acquired in system dispatching. The probability of the input data is:

$$P_r(v) = \sum_h \frac{1}{Z(\theta)}\mathrm{e}^{-E\left(v,h\,|\,\theta\right)} \qquad (3)$$

Given the learning sample set D, the task of learning RBM is to find the parameters so that the probability likelihood function of D is the largest. Usually, the loss function in the form of a minimum is used as (4).

$$\min_{\theta} L\left(\theta,D\right) = -\sum_{v\in D}\lg P_r(v) = -\sum_{v\in D}\lg\sum_h P_r\left(v,h\right) =$$

$$-\sum_{v\in D}\lg\sum_h\frac{\mathrm{e}^{-E\left(v,h\,|\,\theta\right)}}{Z(\theta)} \qquad (4)$$

An RBM is usually trained by a gradient decent (GD) algo-

rithm, and the practical training tutorial is given by Hinton in [33]. RBM 1 is first trained, then the activation probability of the hidden layer neurons of RBM 1 is used as the input of RBM 2 to train RBM 2, and so on to RBM k. This layer-by-layer learning can be repeated as many times as desired. By pre-training each RBM, the connection relation matrix of each RBM can be obtained without any label information.

2) Unfolding and Supervised Fine-tuning

After pre-training multiple RBMs, all RBMs are unfolded and sequentially connected to form a complete DBN that initially uses the same weights of each RBM. Since the TSA is a binary classification problem, logistic classifier [27] is used in this paper. In the framework of NN, it is quite simple to make the logistic classifier, i.e., adding a neuron with the sigmoid activation function as the output layer on the top of the DBN. Then, the label (the stability status of each simulation scenario) is used to fine-tune the weights by backpropagation algorithms [20]. The loss function is designed according to the specific task, and the back propagation algorithm is used to precisely adjust the parameters of the network, and finally the parameter matrixes $W^{(1)} = W^{(1)} + \Delta^{(1)}, \cdots, W^{(k)} = W^{(k)} + \Delta^{(k)}$ of DBN are obtained, where $\Delta^{(k)}$ represents the amount of learned fine-tuning based on the label information. The fine-tuning procedure is supervised and requires label information to guide the adjustment process.

C. Power Grid Structure Reserved RBM

The weights W of a conventional RBM do not have constraints because each input is assumed to have equal significance and contribution. However, in power networks, the fault propagates along the grid; thus, the connected buses and lines share similar characteristics. In this paper, a new regularization method is presented to consider the power grid connection structure to promote the grouping ability of the input features [29]. The revised RBM is called a power grid structure reserved RBM (SRRBM).

The intuitive idea behind SRRBM is that the input features (visible units) measured from adjacent buses or lines share similar weights. Assume that feature v_i and v_j are measured at bus i and bus j, respectively. Then, the weight w_{ik} connecting v_i and h_k and the weight w_{jk} connecting v_j and h_k should be similar, as shown in Fig. 1. Mathematically, these constraints can be obtained by adding a penalty objective, restricting the weight difference $(w_{ik} - w_{jk})^2$. In power system applications, a natural criterion to measure the connection status between bus i and j is the electrical distance ρ_{ij}. The closer bus i is to bus j, the larger ρ_{ij} is, and $\rho_{ij} = 0$ if bus i and j are not connected. The admittance between different nodes also follows the same rule. In this paper, in order to simplify the calculation, the absolute value of the admittance between the power system nodes is selected as a measure of the electrical distance, i. e., $\rho_{ij} = |Y_{ij}|$, which reflects connection degree between nodes. Considering the constraints of the power grid structure can be regarded as a network smoothing constraint. This constraint refers to that the weight difference of the input matrix of the adjacent nodes is close to zero, so that the features learned by the adjacent nodes are similar, as shown in (5).

$$\Omega(\theta) = \sum_{ij} \rho_{ij} \sum_k \left(w_{ik} - w_{jk}\right)^2 \tag{5}$$

where Ω is the complexity index.

Convert (5) into matrix form:

$$\Omega(\theta) = \sum_k W_k^{\mathrm{T}} TW_k \tag{6}$$

where $T = (T_{ij})$, when $i = j$, $T_{ii} = \sum_{i \neq j} \rho_{ij}$, when $i \neq j$, $T_{ij} = -\rho_{ij}$.

Combing (4) and (6), the loss function of SRRBM is:

$$\min_\theta \tilde{L}(\theta, D) = L(\theta) + \frac{1}{2}\alpha\Omega(\theta) \tag{7}$$

where α is the coefficient of penalty.

Training SRRBM only needs to change the gradient and apply the gradient descent algorithm. Using the SRRBM as the first-level RBM of our DBN model, the SRDBN is regularized to consider the power grid connection. From the perspective of machine learning regularization, the network smoothing constraint is similar to L_2 constraint, which can reduce the variance of the model, thus the new SRDBN structure has a more robust performance on unseen scenarios and yields better generalization ability.

III. DBN DECISION INTERPRETER

The deep learning method can provide high-precision stability assessment results and has good model generalization capability. However, similar to other machine learning algorithms, there is no intuitive physical interpretation for the output of deep learning. Therefore, DBN remains a "black box" model to the system operators and cannot be deduced from the original input features step by step like analytical method to get the conclusion, thus it is difficult to directly interpret the obtained results using the knowledge of power system. Also, it is significant for operators to understand the decisions made by the DBN to trust assessment and carry out successive emergency controls if necessary.

How to reasonably relate the input and output of deep learning model has always been a concern in the field of machine learning. Therefore, building an interpreter of the DBN is as important as building the assessment model itself, which is ignored by most of the previous studies. Essentially, there are two criteria for judging the interpreter: interpretability and local fidelity [34]. To be interpretable means to provide simple and clear qualitative relationships between the input features and the assessment decisions. Reference [27] compares the interpretability and predictive power of several popular machine learning models, as shown in Table II. It is generally believed that a model with good predictive power offers poor interpretability and vice versa. For example, the complexity of a DT can be represented by the number of leaves and that of a linear model is the number of non-zero coefficients.

TABLE II
INTERPRETABILITY VS PREDICTIVE POWER

Parameter	NN	SVM	DT	Linear model	Kernel
Interpretability	Poor	Poor	Fair	Good	Poor
Predictive power	Good	Good	Poor	Poor	Good

Another essential criterion is local fidelity. Although it is impossible for the interpreter to be absolutely faithful with DBN model globally, it is important to be locally faithful. For the actual TSA problem, local fidelity means that the interpreter only needs good performance for a specific type of fault scenario.

The interpretation of DBN model is to find a set of interpreters to link the post-fault power system measurements with the stability assessment decisions made by the DBN. To guarantee a better interpretability, it is necessary to find a cluster of simple functions to fit the relationship between the original input and the DBN model output. Denoting the DBN model as f, the simple interpreter as g and the original input as x, then the interpretation of the DBM model is equivalent to satisfy (8).

$$g(x) \approx f(x) \tag{8}$$

As discussed above, there are two requirements for interpretation function, fidelity and interpretability, which means the model learned by explanatory function should be faithful to the DBN as simple as possible. Thus, the requirements for the interpretation function can be transformed into the optimization problem as (9).

$$\min \xi(x) = \Gamma(f,g) + \Omega(g) \tag{9}$$

where $\Gamma(f,g)$ is the difference between f and g, which reflects the accuracy of g. In this paper, a linear model is used to interpret the DBN and the mean square error to minimize the difference. To decrease the complexity of the interpreter, one direct way is to limit the number of non-zero coefficients in the linear interpreter model so that the L1 regularization is a natural choice. The interpretation method is called local linear interpretation (LLI) as (10).

$$f(x_i) \approx g(x_i) = \sum_j \beta_j x_{i,j} \tag{10}$$

This linear model can be regarded as a first-order Taylor expansion that ignores high-order terms. The coefficient β_j of the linear function can be interpreted as the partial derivative of f relative to the original input feature and as the sensitivity of input feature to the stability domain boundary from the perspective of power system as shown in (11).

$$\beta_j = \frac{\partial f}{\partial x_{i,j}} x_{i,j} \tag{11}$$

The detailed steps are as follows.

1) Vicinity sampling: for a case X that needs to be interpreted, randomly sample N data points $(X^{(1)}, X^{(2)}, \ldots, X^{(N)})$ in the vicinity of the case. Use the DBN model to embed these data points into the representation space $(Z^{(1)}, Z^{(2)}, \ldots, Z^{(N)})$, and then assess the stability status $(f^{(1)}, f^{(2)}, \ldots, f^{(N)})$.

2) Weight assigning: in the representation space, compute the distance between each sampled data point as $d(Z, Z^{(j)})$. The normalized weight of each sampled data point is shown as:

$$\Psi_k = \exp\left(-\frac{d\left(z_i, z_i^{(k)}\right)^2}{\sigma^2}\right) \tag{12}$$

where σ is the standard deviation of all the distance.

Equation (12) indicates that the closer the sampled data is to the case being interpreted, the larger its normalized weight. The weight parameter can help limit the sampled data in the vicinity of the case being interpreted to make the interpreter more robust to noises.

3) Parameter computing: as stated above, loss function $\Gamma(f,g)$ represents the difference between the interpreter and the DBN and can be written as (13) using a mean square error.

$$\Gamma(f,g) = \sum_k \Psi_k \left(f(x_i^{(k)}) - g(x_i^{(k)})\right)^2 = \sum_k \Psi_k \left(f(x_i^{(k)}) - \sum_j \beta_j x_{i,j}^{(k)}\right)^2 \tag{13}$$

And the complexity index Ω is represented by the number of non-zero coefficients of the linear interpreter in (14).

$$\Omega(g) = \sum_j |\beta_j| \tag{14}$$

Therefore, (9) can be written as:

$$\min_{\beta_j} \xi(x_i) = \sum_k \Psi_k \left(f(x_i^{(k)}) - \sum_j \beta_j x_{i,j}^{(k)}\right)^2 + \lambda \sum_j |\beta_j| \tag{15}$$

The left term of right side in (15) is the square error summation term, and the right term is the L1 regularization term, so the local linear interpretation can be transformed into a linear regression problem using the L1 penalty factor, namely the Lasso problem, and λ is the penalty parameter. Like deep learning model, the Lasso problem can be solved optimally using the stochastic gradient descent algorithm. Since (15) is quadratic, optimizing (15) is very fast compared to time-domain simulation.

For TSA problem, from the perspective of topology, the stability domain boundary is a highly nonlinear surface, and the vicinity of any point can be approached by an approximately linear hyperplane. From the point of transient stability, the characteristics that are used to describe the fault are different when the system is affected by different faults. Therefore, it is feasible and necessary to use a linear interpreter in the vicinity of each sample point while the point and its vicinity could relate to a specific fault type and its similar scenarios.

For every case being interpreted, optimizing (15) gives a linear model with limited number of non-zero coefficients. The coefficient β_j indicates the contribution of a feature to the DBN decision. It is easy to know how the features are relevant to the stability status by comparing the value of the coefficients, i.e., for a specific type of fault, the value of β_j shows the importance of active power on different lines for the stability. More importantly, in the vicinity of the sample being interpreted, the stability boundary can be locally linearized, and the coefficient β_j can be regarded as the sensitivity of the power system measurements to the stability boundary.

IV. METHODOLOGY

The flow chart of the DBN-based TSA method is proposed in Fig. 3. The scheme includes two stages: offline training and online assessment. In the offline training stage, DBN is trained with simulation data. In the online assessment stage, the DBN assessment model is triggered whenever a fault occurs. If the system is judged as unstable, the LLI interpreter will be triggered to provide supporting information for the emergency control strategies.

1) Offline training: to evaluate real-time transient stability, first, the system dynamic response database should be established. The data is acquired from extensive transient simulation and prepared in the format of a structured data matrix X:

$$X = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1N} & Q_{11} & Q_{12} & \cdots & Q_{1N} & U_{11} & U_{12} & \cdots & U_{1M} & \theta_{11} & \theta_{12} & \cdots & \theta_{1M} \\ P_{21} & P_{22} & \cdots & P_{2N} & Q_{21} & Q_{22} & \cdots & Q_{2N} & U_{21} & U_{22} & \cdots & U_{2M} & \theta_{21} & \theta_{22} & \cdots & \theta_{2M} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ P_{K1} & P_{K2} & \cdots & P_{KN} & Q_{K1} & Q_{K2} & \cdots & Q_{KN} & U_{K1} & U_{K2} & \cdots & U_{KM} & \theta_{K1} & \theta_{K2} & \cdots & \theta_{KM} \end{bmatrix} \tag{16}$$

where N, M, K are the numbers of transmission lines, buses and different cases, respectively. The data will be cleaned and processed, including a consistency check, outlier check, interpolation, and normalization after being acquired.

In China, each province branch of State Grid Corporation of China needs to perform $N-1$ contingency simulations for every typical operation condition every year, which has accumulated large amounts of transient data. These data have been used to train a practical stability assessment model for a regional power system.
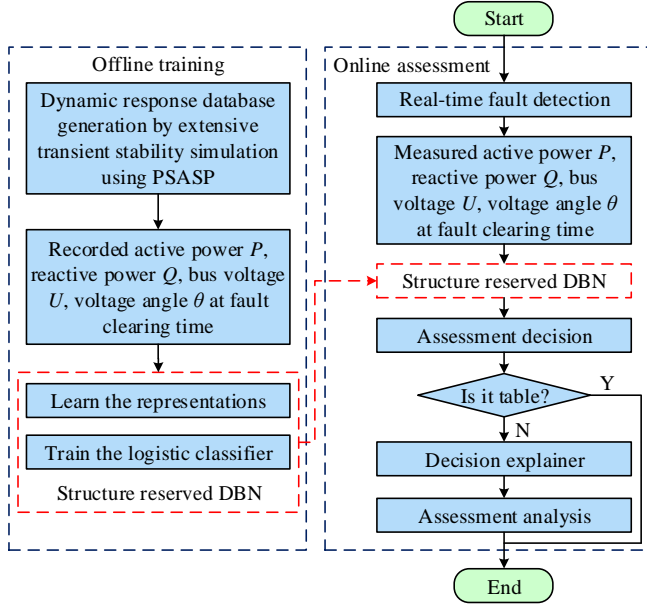

Fig. 3. Flowchart for offline training and online assessment steps of proposed method.

2) Online assessment: after the real-time measurements of the corresponding variables are obtained from the wide-area measurement system (WAMS), the stability status is immediately predicted after the fault is cleared. If the system is assessed as unstable, the LLI interpreter is applied to compute the importance of relevant features. As stated in Section III, the coefficient of the LLI interpreter model can be regarded as a qualitative criterion about how the measurements are related to the stability status. In other words, changing the corresponding features (e. g., generation or load shedding) will help enhance the stability.

V. CASE STUDY AND DISCUSSIONS

A. Generation of Dynamic Response Database

The dynamic response database is generated by time-domain simulations. The uncertainty of faults considered in the paper involves the type, the location of disturbance, and the fault clearing time. The duration of fault is chosen from a normal distribution from 0.1 s to 0.4 s. The long fault lasting time guarantees to generate sufficient unstable cases and patterns for NN to learn from. In the final training dataset, the stable/unstable ratio is 1 : 1.

The transient stability simulation for the IEEE test system is carried out by the PSAT toolbox in MATLAB [35], and those for the real power system are simulated by PSASP [36]. The length of simulation time after fault clearing is 20 s. The system stability is judged at the end of the simulation period. If the difference between rotor angles of any two gen-

erators exceeds 360° at the end of simulation, the case is labeled unstable, otherwise stable.

B. Case Study on Test System

The proposed method is tested on the IEEE new England system with 10 machines and 39 buses, as shown in Fig. 4. The fault is set to three-phase short-circuit grounding of the line, and the fault line is removed after a period of time. In order to avoid the occurrence of islands in the system when the fault line is removed, 35 of 46 lines are selected to participate in the transient scans. The fault location is set at either end of the line, and the fault clearing time is randomly taken from 0.1 s to 0.4 s. P, Q of the 46 lines and V, θ of the 39 buses (angle at generator 30 is a reference) are recorded at the fault clearing time, which form the original input feature of 1 dimension. A total of 3500 samples are generated, and the number of stable samples is equal to the number of unstable samples.
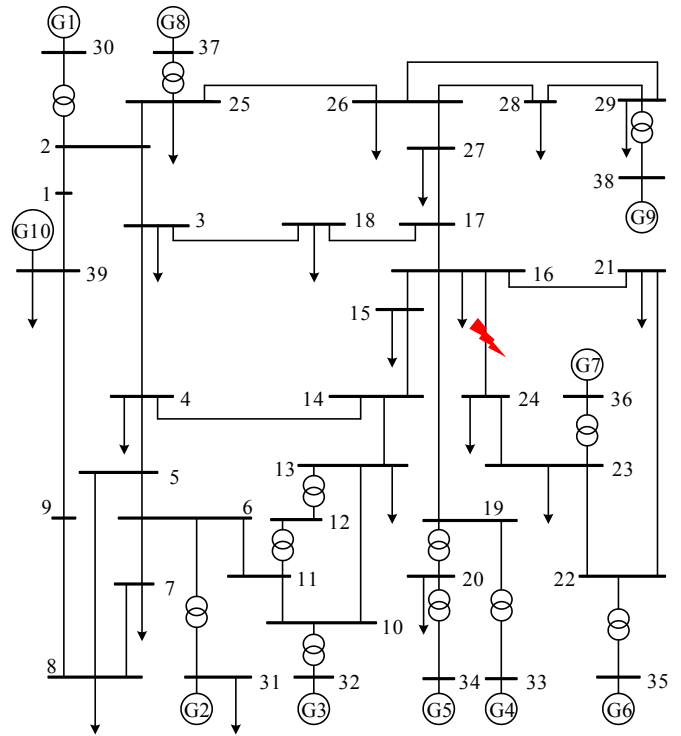

Fig. 4. IEEE new England system with 10 machines and 39 buses.

In the machine learning community, the term of generalization refers to the ability to perform well on new data. In the specific application of TSA, the generalization means the assessment model should be robust in unseen scenarios. This process is very important to make the model practical since time domain simulations can not include all possible operation states or faults. Therefore, the generated samples are divided into three sets as the ratio of 4 : 1 : 2. 2000 of samples are used to learn nonlinear expression as set A, 500 are used to learn classifier as set B, and 1000 are used for testing as set C. In order to reflect the generalization ability of the model, the fault type of set C does not appear in sets A and B. The database generation takes approximately 2 hours of

computation time on a PC with an Intel i5 CPU 3.19 GHz clock and 2 GB RAM.

The assessment ability is evaluated by the misclassification rate that includes two kinds of classification errors, namely, false dismissal (FD) and false alarm (FA). An FD occurs when an unstable case is classified as stable, whereas FAs denote the stable cases deemed as unstable. To illustrate the advantage of the proposed SRDBN model, three models are trained and compared.

1) Polynomial kernel SVM: the SVM model is chosen as a benchmark approach since SVM has achieved state-of-art performance in many TSA studies [9], [15], [16].

2) Conventional DBN: a conventional DBN is used to learn the representations of the original data. The first layer is the conventional RBM, which contains 170 neurons. The network has four hidden layers, and the numbers of neurons in each hidden layer are 1000, 2000, 500 and 30, respectively. In other words, the DBN embeds the data from the original 170-dimension feature space to a 30-dimension representation space.

3) Structure reserved DBN: a SRRBM is used as the first layer of the DBN, all network architecture and configurations are identical to the conventional DBN in model 2.

The training is finished within 1 hour. The misclassification error is presented in Table III. The result shows that DBN-based model achieves better classification performance than the benchmark SVM method. The test results also suggest that the SRDBN performs better than the conventional DBN with "unseen" faults, indicating that the SRDBN model is more practical in the specific application of TSA.

TABLE III
MISCLASSIFICATION ERROR RATE OF TEST SYSTEM

Method	FA (%)	FD (%)	Overall (%)
Polynomial kernel SVM	5.97	5.02	5.50
DBN with RBM	3.78	2.81	3.30
DBN with SRRBM	2.19	1.40	1.80

In order to better show how SRDBN works, the t-distributed stochastic neighbor embedding (t-SNE) algorithm [37] is utilized to map the high-dimension space into a 2-dimension space. The t-SNE is a dimensionality reduction tool that preserves distances in the high-dimension space. Thus, it is suitable for the visualization of high-dimensional data. The 2-dimension projection of the original feature space and the hidden layer space are shown in Fig. 5 and Fig. 6. Note that for the t-SNE algorithms, the axis and the actual point coordinates have no logical meaning, while the relative distance between the points reflects their closeness in the high-dimensional space. From Fig. 5, the stable and unstable samples mix with each other in the original feature space. In contrast, Fig. 6 clearly shows that the samples are gradually divided into two clusters after passing through each hidden layer. Therefore, it is much easier to identify the unstable cases in the representation space than in the original feature space.

In the original input space, the Euler distance between sample points does not reflect the similarity of the samples, while in the representation space, since the SRDBN can learn the global and local features of transient data, the distance of sample points can reflect the difference of instability mode of unstable samples, as shown in Fig. 7. Three instability modes are marked in the figure, i.e., the generator G5 swings relative to other units of the system (mode 1), generators G2, G3, G5 swing relative to other units (mode 2) and generators G5, G9 swing relative to other units (mode 3). The sample points with the same instability mode naturally gather together, indicating that the relative distance in the representation space reflects the similarity of the transient features.
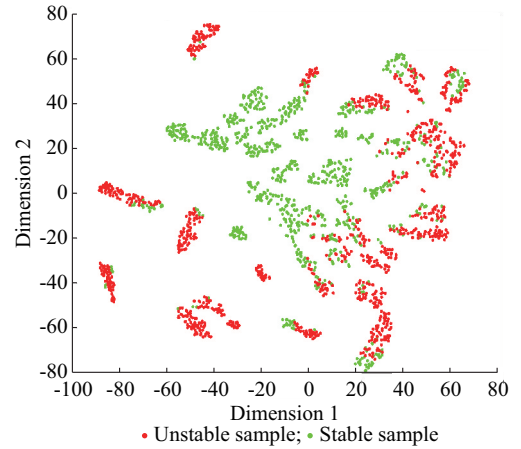

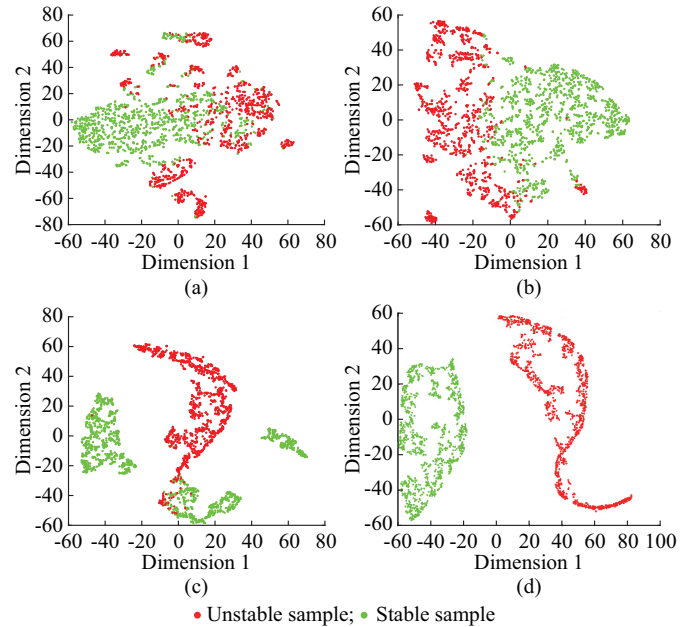Fig. 5.   Two-dimension visualization of original feature space.

Fig. 6.   Two-dimension visualization of hidden layer and representation space. (a) Hidden layer 1. (b) Hidden layer 2. (c) Hidden layer 3. (d) Hidden layer 4.

As for the decision interpretation of DBN, a three-phase short-circuit fault occurring at the end of line 16-24 (Fig. 4)

is considered. The fault lasts 0.18 s, and the system becomes unstable after clearing the fault. Figure 8 shows result of linearization.
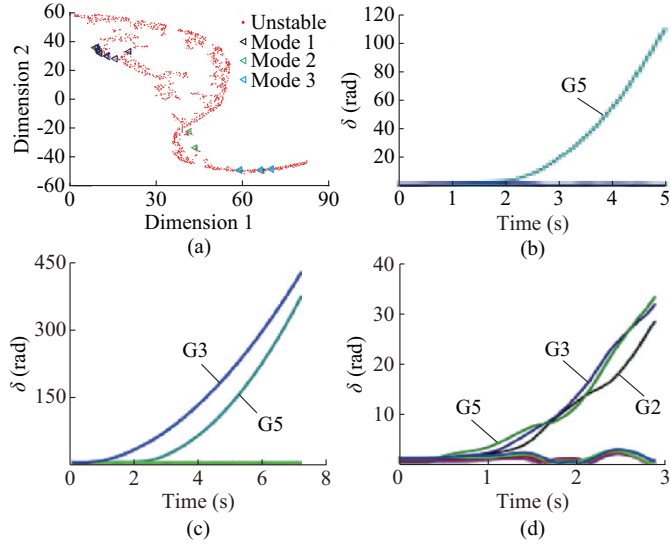

Fig. 7.   2-dimension visualization of the representation space and three in-stability modes. (a) 2-dimension visualization. (b) Mode 1. (c) Mode 2. (d) Mode 3.

The left side of Fig. 8, i.e., the prediction probability indicates the stability assessment of SRDBN, while the middle and right sides list the features most relevant to the stability calculated by LLI. The four features listed are active power

of line 26-29, line 28-29, line 26-28 and line 29-38. The coefficient of 0.09 in the middle part of Fig. 8 is β_j in LLI. $P_{26\text{-}29} \leq -2.54$ indicates that the active power of line 26-29 less than -2.54 is an important reason why the sample point is assessed as unstable compared with other sample points in the vicinity. Combined with Fig. 4, all the features are directed to the generator G38, which is consistent with this case that G38 swings relative to other units of the system. When the fault occurs, by reducing the output of G38 and removing some of the nearby bus load, G38 will be controlled back to the steady state. Therefore, the LLI proposed in this paper can not only find the relationship between the original input and the assessment result, but also help guide the subsequent emergency control.

Fig. 8.   Interpretation of local linearization.

C. Case Study on Real Power System

The method has been demonstrated on Hubei power system in central China, which contains 182 machines, 1300 buses and 3215 transmission lines. The one-line diagram of the real system is shown in Fig. 9.
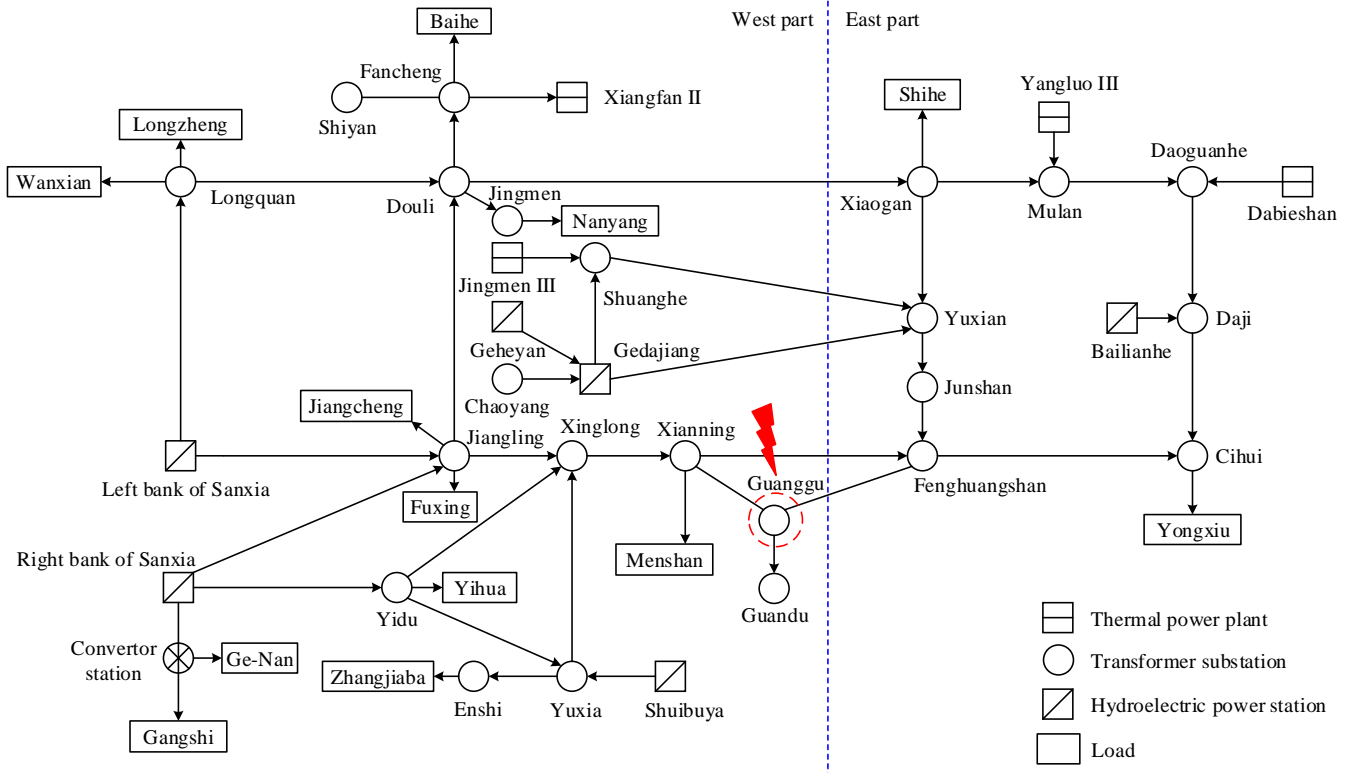

Fig. 9. Diagram of HBPS 500 kV grid structure.

The diagram only contains buses whose rated voltage is 500 kV or higher. A typical operation condition on a summer peak day is used as the initial state of the time domain simulations. The active and reactive power of AC and DC lines as well as the voltage magnitude and angle of buses rated 220 kV and above are recorded. Therefore, the input dimension is 2025. A total of 45000 contingencies are simulated, and all the faults occur on important transmission lines or transformers. The samples are randomly divided into three parts at a ratio of 6 : 1 : 1. 33800 samples are used to learn linear nonlinear expression recorded as set D, 5600 are used to learn the classifier as set E, and the other 5600 are for testing as set F.

Similar to the test system above, the model performance is illustrated with the three methods, except that the architecture of DBN is more complex. The model has 8 layers, and the numbers of neurons of each layer are 1762, 3000, 4000, 6000, 4000, 2000, 1000 and 300, respectively, while the dimension of the representation space is 300. Compared with the IEEE test system, the real system has higher original input space dimensions and more complex stability boundary, so a deeper network is structured for representation learning. Training the SRDBN takes 3.6 GPU hours. The average assessment time is 0.1 s, and the average interpretation time is 0.8 s.

The misclassification rate is shown in Table IV and the visualization of the original feature space and the representation space is shown in Fig. 10. Similar conclusions can be drawn based on the simulations in the regional power system.

TABLE IV
MISCLASSIFICATION ERROR RATE OF REAL SYSTEM

Method	FA (%)	FD (%)	Overall (%)
Polynomial kernel SVM	6.73	6.32	6.55
DBN with RBM	2.44	2.79	2.63
DBN with SRRBM	1.07	1.64	1.37

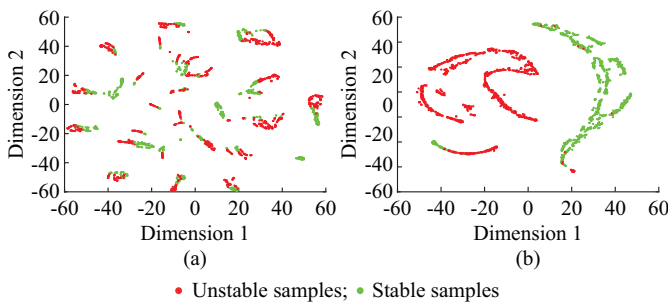• Unstable samples; • Stable samples

Fig. 10.   Two-dimension visualization of original feature space and representation space. (a) Original feature space. (b) Representation space.

In Hubei power system, the west part has the most generation, and the east part has the most loads. A three-phase short-circuit fault located on one of the main 500 kV transmission lines is considered, as shown in Fig. 9. The fault will cause the oscillations to be out of step between the generators of the west part and the east part, as shown in Fig. 11. The interpretation results show that the active power

around one 500 kV power plant (highlighted by the red circle in Fig. 8) and two 220 kV power plants (not shown in the figures) exceed the limit. After decreasing the corresponding power output and tripping nearby loads, the system becomes stable, as shown in Fig. 12.
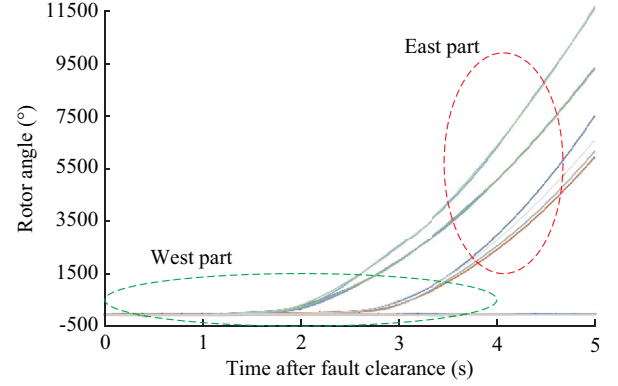

Fig. 11.   Curves of power angles without emergency control.
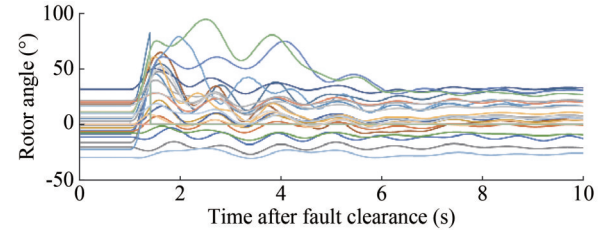

Fig. 12.   Curves of power angles with emergency control.

In addition, the effect of the uncertainty of measurement system and possible errors is investigated by randomly setting a fraction of the input measurements to 0, which simulates the failure of a part of the measuring devices. The permutation rate is set from 1% to 30%, and 1000 experiments are carried out for each permutation rate to compute an average accuracy of the permuted data. The TSA accuracy after the random permutation is shown in Fig. 13. The proposed model can achieve an accuracy of 85% with a permutation rate of 30%, which indicates the robustness of the proposed method when the measurement or the communication system encounters problems.
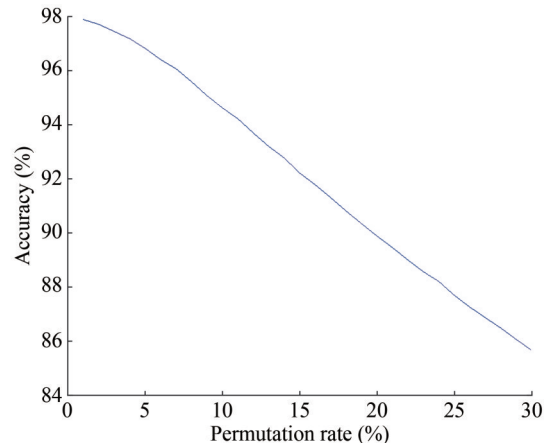

Fig. 13.   Average accuracy at different permutation rate.

VI. Conclusions and Future Work

In this paper, an improved DBN-based method is proposed to assess power system transient stability and provide guidance for emergency control when the system is unstable. The main contributions of this paper are follows.

1) This paper proposes an improved DBN model considering the structure of power system. By modifying the loss function of DBN, the performance, generalization stability and robustness of the model are improved.

2) This paper proposes an interpretation method for DBN decision. The local linearization interpretation transforms the nonlinear stability domain boundary fitting problem into a L_1 regularized linearization fitting problem, and is equivalent to solve the sensitivity relationship between the model output and the original input. It indicates the most important factors relative to system instability and can be used to guide emergency control. The proposed method overcomes the drawback of human feature engineering considering the power grid structure and has higher accuracy, robustness and interpretability when used in real-world applications, while the conventional machine learning method can only give the state of a post-fault system.

3) This paper shows the internal state of DBN network in detail through visualization technology which help operators understand, and analyzes the main pattern of system instability according to the actual cases. The feature separation process is intuitively given by the output of each hidden layer inside the DBN, while the adjacent regions in the high-dimension space reflect similar system instability pattern.

There are still a lot of research work to be carried out in the future such as studying the meaning of each layer in deep structural model learning, how to determine the structure and parameters properly of the deep learning model, etc.

References

[1] IEEE/CIGRE Joint Task Force on Stability Terms and Definitions, "Definition and classification of power system stability," *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 1387-1401, May 2004.

[2] P. Kundur, *Power System Stability and Control*, 2nd ed. New York, USA: McGraw-Hill, 1994.

[3] Q. Zhou, J. Davidson, and A. A. Fouad, "Application of artificial neural networks in power system security and vulnerability assessment," *IEEE Transactions on Power Systems*, vol. 9, no. 1, pp. 525-532, Feb. 1994.

[4] A. Al-Masri, M. Kadir, H. Hizam *et al.*, "A novel implementation for generator rotor angle stability prediction using an adaptive artificial neural network application for dynamic security assessment," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2516-2525, Aug. 2013.

[5] N. Amjady and S. Majedi, "Transient stability prediction by a hybrid intelligent system," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1275-1283, Aug. 2007.

[6] B. Shakerighadi, F. Aminifar, and S. Afsharnia, "Power systems wide-area voltage stability assessment considering dissimilar load variations and credible contingencies," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 1, pp. 78-87, Jan. 2019.

[7] A. D. Rajapakse, F. Gomez, K. Nanayakkara *et al.*, "Rotor angle instability prediction using post-disturbance voltage trajectories," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 947-956, May 2010.

[8] S. Dutta and T. J. Overbye, "Feature extraction and visualization of power system transient stability results," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 966-973, Mar. 2014.

[9] T. Guo and J. V. Milanovic, "Online identification of power dynamic signature using PMU measurements and data mining," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1760-1768, May 2016.

[10] L. Wehenkel, M. Pavella, E. Euxibie *et al.*, "Decision tree based transient stability method: a case study," *IEEE Transactions on Power Systems*, vol. 9, no. 1, pp. 459-469, Feb. 1994.

[11] M. He, V. Vittal, and J. Zhang, "Online dynamic security assessment with missing PMU measurements: a data mining approach," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1969-1977, May 2013.

[12] K. Sun, S. Likhate, V. Vittal *et al.*, "An online dynamic security assessment scheme using phasor measurements and decision trees," *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1935-1943, Nov. 2007.

[13] M. He, J. Zhang, and V. Vittal, "Robust online dynamic security assessment using adaptive ensemble decision-tree leaning," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4089-4098, Nov. 2013.

[14] H. Zhao, Y. Gao, H. Liu *et al.*, "Fault diagnosis of wind turbine bearing based on stochastic subspace identification and multi-kernel support vector machine," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 2, pp. 350-356, Mar. 2019.

[15] W. Zhang, W. Hu, Y. Min *et al.*, "Conservative online transient stability assessment in power system based on concept of stability region," *Power System Technology*, vol. 40, no. 4, pp. 992-998, Apr. 2016.

[16] W. Hu, Z. Lu, S. Wu *et al.*, "Real-time transient stability assessment in power system based on improved SVM," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 1, pp. 26-37, Jan. 2019.

[17] J. Lv, M. Pawlak, and U. D. Annakkage, "Prediction of the transient stability boundary using the LASSO," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 281-288, Feb. 2013.

[18] W. Hu, Z. Lu, S. Wu *et al.*, "Real-time transient stability assessment in power system based on improved SVM," *Journal of Modern Power System and Clean Energy*, vol. 7, no. 1, pp. 26-37, Jan. 2019.

[19] H. Fan, S. Huang, Y. Chen *et al.*, "Power system transient stability assessment based on dimension reduction and cost-sensitive ensemble learning," in *Proceedings of 2017 IEEE Conference on Energy Internet and Energy System Integration*, Beijing, China, Nov. 2017, pp. 1-67.

[20] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, pp. 505-507, Jul. 2006.

[21] F. Movahedi, J. Coyle, and E. Sejdic, "Deep belief networks for electroencephalography: a review of recent contributions and future outlooks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 642-652, May 2018.

[22] F. Ghasemi, A. Fassihi, H. Perez-Sanchez *et al.*, "The role of different sampling methods in improving biological activity prediction using deep belief network," *Journal of Computational Chemistry*, vol. 38, no. 4, pp. 195-203, Feb. 2017.

[23] M. Mahdi and V. M. I. Genc, "Post-fault prediction of transient instabilities using stacked sparse autoencoder," *Electric Power Systems Research*, vol. 164, pp. 243-252, Nov. 2018.

[24] Q. Zhu, J. Dang, J. Chen *et al.*, "A method for power system transient stability assessment based on deep belief networks," *Proceedings of the CSEE*, vol. 38, no. 3, pp. 735-743, Feb. 2017

[25] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798-1828, Jun. 2013.

[26] L. Zheng, W. Hu, Y. Zhou *et al.*, "Deep belief network based nonlinear representation learning for transient stability assessment," in *Proceedings of 2017 IEEE PES General Meeting*, Chicago, USA, Jul. 2017, pp. 16-20.

[27] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, USA: Springer, 2009.

[28] L. Wehenkel, "Machine learning approaches to power system security assessment," Ph.D. dissertation, Department of Electrical Engineering and Computer Science, University of Liège, Liège, Belgique, 1995.

[29] T. Tran, T. Nguyen, D. Phung *et al.*, "Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM)," *Journal of Biomedical Informatics*, vol. 54, pp. 96-105, Feb. 2015.

[30] C. Zhang, C. Chen, M. Gan *et al.*, "Predictive deep Boltzmann machine for multiperiod wind speed forecasting," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1416-1425, Oct. 2015.

[31] J. Goldberger, S. Roweis, G. E. Hinton *et al.*, "Neighbourhood components analysis," in *Proceedings of the 17th International Conference on Neural Information Processing Systems*, British Columbia, Canada, Dec. 2005, pp. 513-520.

[32] R. Salakhutdinov and G. E. Hinton, "Learning a nonlinear embedding

by preserving class neighbourhood structure," *Journal of Machine Learning Research*, vol. 2, pp. 412-419, Jan. 2007.

[33] G. Hinton, "A practical guide to training restricted Boltzmann machines," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 599-619.

[34] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proceedings of International Conference on Knowledge Discovery and Data Mining*, San Francisco, USA, Aug. 2016, pp. 1-10.

[35] F. Milano. (2018, Feb.). Documentation for PSAT. [Online]. Available: http://faraday1.ucd.ie/psat.html

[36] H. Chen and Z. Xu, "Comparison of mathematical models for transient stability calculation in PSASP and PSS/E and corresponding calculation results," *Power System Technology*, vol. 28, no. 5, pp. 1-5, Mar. 2004.

[37] L. van der Maaten and G. E. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579-2605, Jan. 2008.

Shuang Wu received his B.S. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2016, and he is now an Ph.D. candidate in Tsinghua University, Beijing, China. His research interests include power system analysis and control, big data technology in power system.

Le Zheng received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 2011 and 2017, respectively. He is currently a postdoctoral research fellow in Stanford University, Stanford, USA. His research interests include power system stability and control, large-scale wind energy integration, data mining and machine learning applications in power systems.

Wei Hu received the B.S. and Ph.D. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1998 and 2002, respectively, and he is now an Associate Professor at there. His research interests include power system analysis and control, big data technology in power system, multi-type power generator-grid coordination and control, optimization control between renewable energy and energy storage system.

Rui Yu received the B.S. and the M.A. degrees in electrical engineering from Chongqing University, Chongqing, China, in 1999 and 2002, respectively, and he is now a senior engineer in Southwest Branch of State Grid Corporation of China, Chengdu, China. His research field is dispatching and operation of power system.

Baisi Liu received the B.S. and the M.A. degrees in electrical engineering from Chongqing University, Chongqing, China, in 2002 and 2004, respectively, and he is now a senior engineer in Southwest Branch of State Grid Corporation of China, Chengdu, China. His research field is analysis and control of power system.