Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

An artificial intelligence based method for evaluating power grid node importance using network embedding and support vector regression

Hui-fang WANG[†‡1], Chen-yu ZHANG[1], Dong-yang LIN[2], Ben-teng HE[1]

[1]*Department of Electrical Engineering, Zhejiang University, Hangzhou 310027, China*

[2]*State Grid Jiangsu Electric Power Engineering Consulting Co., Ltd., Nanjing 210008, China*

[†]E-mail: huifangwang@zju.edu.cn

Received Mar. 8, 2018; Revision accepted June 17, 2018; Crosschecked June 11, 2019

Abstract: The identification of important nodes in a power grid has considerable benefits for safety. Power networks vary in many aspects, such as scale and structure. An index system can hardly cover all the information in various situations. Therefore, the efficiency of traditional methods using an index system is case-dependent and not universal. To solve this problem, an artificial intelligence based method is proposed for evaluating power grid node importance. First, using a network embedding approach, a feature extraction method is designed for power grid nodes, considering their structural and electrical information. Then, for a specific power network, steady-state and node fault transient simulations under various operation modes are performed to establish the sample set. The sample set can reflect the relationship between the node features and the corresponding importance. Finally, a support vector regression model is trained based on the optimized sample set for the later online use of importance evaluation. A case study demonstrates that the proposed method can effectively evaluate node importance for a power grid based on the information learned from the samples. Compared with traditional methods using an index system, the proposed method can avoid some possible bias. In addition, a particular sample set for each specific power network can be established under this artificial intelligence based framework, meeting the demand of universality.

Key words: Power grid; Artificial intelligence; Node importance; Text-associated DeepWalk; Network embedding; Support vector regression

https://doi.org/10.1631/FITEE.1800146 **CLC number:** TP39; TM74

1 Introduction

Some large blackouts caused by the collapse of certain nodes aroused wide awareness of the need for protection of the key nodes in a power grid (Arianos et al., 2009). According to existing research, most power systems are robust against random attack but fragile against intentional attack. The faults occurring on just a very few key nodes may lead to severe consequences (Albert et al., 2004). Therefore, correctly identifying these key nodes is a necessary step for the differentiated management of power network looking at both safety and economic efficiency (Li and Liang, 2009).

Most existing research can be summarized as using an index system to evaluate node importance. By defining some indices that reflect consequences of the “absence” of the node and then calculating the index value for each node, node importance is ranked accordingly.

Several primary indicators in complex network theory, such as betweenness, shortest path, degree, and closeness, are used to evaluate the importance of power grid nodes (Xu et al., 2010; Nasiruzzaman and Pota, 2011; Ju and Li, 2012; Pan et al., 2014; Fan et al.,

‡ Corresponding author
ORCID: Hui-fang WANG, http://orcid.org/0000-0002-1483-364X

2016; Lin et al., 2017). Among these, several electrical features such as power flow and line impedance are integrated into the index system taking account of the characteristics of the power network (Xu et al., 2010; Nasiruzzaman and Pota, 2011; Pan et al., 2014; Fan et al., 2016; Lin et al., 2017). There is also much literature that establishes an index system using expert knowledge (Xu et al., 2014; da Silva et al., 2016). For example, by categorizing the nodes according to their functions in the network, the internal characteristics and the external influences on the voltage level and active power transmission can be used to establish an index system (Xu et al., 2014); the static performance index (SPI) and dynamic performance index (DPI) are defined to evaluate node importance considering both static and dynamic vulnerabilities of the system (da Silva et al., 2016). Some other documents focus on using some novel indicators. The PageRank (PR) value and hypertext induced topics (HITS) value are calculated for each power grid node to rank their importance based on Internet web page ranking algorithms in Li et al. (2014) and Wang et al. (2017), respectively.

Evaluation methods using an index system can reflect the importance of a power grid node from many critical perspectives. However, power networks vary in scale, structure, and even the tolerance to different consequences caused by the node fault. The method based on an index system cannot reasonably meet the demand of evaluating node importance for all power networks (Bai et al., 2008; Cai et al., 2012). For example, methods based on web page ranking algorithms (Li et al., 2014; Wang et al., 2017) emphasize power flows and the connections between different nodes in the definition of the indices. They are more suitable for finding nodes that are linked with some important loads and generators. Methods based on complex network theory pay more attention to the structural and electrical characteristics of the network (Xu et al., 2010; Nasiruzzaman and Pota, 2011; Pan et al., 2014; Fan et al., 2016; Lin et al., 2017). They are more suitable for finding nodes that play a more important role in power transmission for the whole system. Different methods always give different ranking results for the same power network. This is because different index systems have totally different emphases in their definition of the indices (Li et al., 2014; Tan et al., 2014; Wang et al., 2017). This makes it hard for operators to choose a suitable method for a specific power network. Therefore, it is important to establish a more universal framework for node importance evaluation. This framework should consider not only the physical characteristics of specific power networks, but also the different focuses of the operators when evaluating node importance.

With the coming of the "artificial intelligence era," modeling methods driven directly by data, such as machine learning, provide new insights for node importance ranking. Machine learning methods establish a framework for handling a sample set with data labels generated directly by the object under study (Angra and Ahuja, 2017). Under this framework, a particular sample set can be established for each specific power network. The sample set can reflect the relationship between features of nodes and their corresponding importance. In the data labeling process, a node fault consequence can be labeled according to the specific preference of different power networks. These characteristics make machine learning a more universal and flexible choice given the diversity of power networks. At the same time, machine learning allows us to perform rapid online calculation and use rich information gathered in time-consuming ways offline (Angra and Ahuja, 2017), such as the steady-state and transient simulation data for the power grid. There has been a lot of research on applying machine learning technologies to deal with power system fault consequences, such as transient stability assessment (Xu et al., 2011; Wang et al., 2016). This work is significant, but little attention has been paid to network structure and electrical features at the same time.

The key to apply machine learning methods is to find an effective way to vectorize the features of the object under study. Network embedding (NE) is a prevailing method which aims to represent the features of network entities, such as nodes and edges, in a vector space (Perozzi et al., 2014; Chen et al., 2015; Yang et al., 2015; Grover and Leskovec, 2016). Since the power grid is an important system in the format of the network along with the Internet, social networks, and so on, NE methods have the potential to be applied to deal with power grid problems. In this paper, an NE method is used to vectorize features of power grid nodes with regard to their structural and electrical information. Based on this, a machine learning framework is established to evaluate node importance. Then, a sample set can be established based on a set of steady-state and node fault transient simulations on different operational modes of each specific power network. It can reflect the relationship between features of nodes and their importance. Then, a machine learning method called support vector regression (SVR) is used to automatically mine the valuable information contained in these samples. The case study demonstrates that the proposed method can effectively evaluate node importance for a power grid based on the information learned from the samples. Compared with the traditional methods using an index system, the proposed method can avoid some possible bias. At the same time, this artificial intelligence based method has high computational speed. In addition, a particular sample set for each specific power network can be established under this artificial intelligence based framework, so the demand on universality is also met.

2 Feature extraction for power network nodes

The first step to establish a machine learning framework is to extract vectorized feature expression for the object under study. For a power grid, electrical features of a node will change in different operational modes. These changes may have impacts on the whole network. Therefore, the importance of the nodes will vary with different operational modes of the system. There are two kinds of features of a power network node that reflect its importance. First, electrical features, such as voltage and injected power flow, can reflect the state of the node in a certain operational mode. For example, a node with higher voltage or larger injected flow is more likely to be a critical node. Second, the positional features, which measure the relative location between different nodes, are also essential, because the nodes in a power network are not independent but are connected as a whole. The impact of a single node on the system is exerted through the links between the nodes. Therefore, features extracted for power grid nodes should be able to reflect these two kinds of information and their mutual effect. The text associated DeepWalk (TADW) algorithm, a popular NE method proposed in Yang et al. (2015), is applied in this study to extract the features of power grid nodes. Its main idea is to use low-rank matrix factorization (Yang et al., 2015) to aggregate the transition matrix $\boldsymbol{A}$ and the internal feature matrix $\boldsymbol{Q}$ for the nodes in a network, to extract a vectorized feature expression for the nodes considering the two kinds of features mentioned above. The transition matrix $\boldsymbol{A}$ measures the structure of the network, and the internal feature matrix $\boldsymbol{Q}$ measures the electrical features independent of the network topology.

Given a power network $G=(V, E)$, where V is the set of all nodes, and E is the set of all lines. $|V|$ is the number of nodes in the network. The transition matrix $\boldsymbol{A}\in\mathbb{R}^{|V|\times|V|}$ of the network is a concept in stochastic process theory, reflecting the probability of state transition (Ross, 1983). This transition matrix $\boldsymbol{A}$ can be defined as

$$A_{ij}=\begin{cases}1/d_i, & (i,j)\in E,\\ 0, & (i,j)\notin E,\end{cases} \tag{1}$$

where d_i represents the degree of node i, which equals the number of lines connected to node i.

Based on matrix $\boldsymbol{A}$, a random walk matrix $\boldsymbol{M}\in\mathbb{R}^{|V|\times|V|}$ can then be defined as

$$M_{ij}=\log\left(\frac{[e_i(A+A^2+\cdots+A^t)]_j}{t}\right), \tag{2}$$

where e_i is a $|V|$-dimensional row vector with value 1 at position i and 0 at all other positions, parameter t represents the steps of random walk, and $[\cdot]_j$ represents the j^{th} element of the vector. According to Yang et al. (2015), the feature of the nodes can be extracted and expressed more accurately and effectively when t is set to 2. Therefore, in this study, t is also set as 2. The physical significance of M_{ij} is the logarithmic mean probability for any random walk process to reach node j from node i within t steps. So, matrix $\boldsymbol{M}$ is able to reflect the relative position between any two different nodes in the network from a probabilistic perspective.

The internal feature matrix $\boldsymbol{Q}\in\mathbb{R}^{S\times|V|}$ contains $|V|$ columns. Each column vector in $\boldsymbol{Q}$ represents the feature information of a corresponding node that is independent of network structure in a certain operational mode. Define s as the number of dimensions of these features for each node. With regard to the power network, matrix $\boldsymbol{Q}$ represents the electrical features of the nodes, which may vary a lot under different operational modes. In this study, there are 9 different electrical features summarized, as shown in Table 1, and s is accordingly set as 9.

Given the fact that values of features mentioned in Table 1 vary in ranges and scales, it is necessary to perform normalization processing to ensure the accuracy of modeling. A linear normalization is applied to each feature to guarantee that all the features can be transformed into the range of [0, 1]:

$$f_{\text{norm}} = \frac{f - f_{\min}}{f_{\max} - f_{\min}}, \tag{3}$$

where f is the original value, f_{norm} is the normalized value, and $f_{\min}$ and $f_{\max}$ are the minimum and maximum values among all original values, respectively.

Then, the low-rank matrix factorization algorithm is applied to extract features from matrices. Notice that a matrix can be expressed as the product of several low-rank matrices which carry the feature information of the original matrix. The main idea of low-rank matrix factorization is to find an optimal product expression $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{HQ}$ for matrix $\boldsymbol{M}$ so that the difference between $\boldsymbol{M}$ and $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{HQ}$ can reach a minimum value:

$$\min_{\boldsymbol{W},\boldsymbol{H}} \left\|\boldsymbol{M} - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{HQ}\right\|_{\mathrm{F}}^{2} + \frac{\lambda}{2}\left(\left\|\boldsymbol{W}\right\|_{\mathrm{F}}^{2} + \left\|\boldsymbol{H}\right\|_{\mathrm{F}}^{2}\right), \tag{4}$$

where $\|\cdot\|_{\mathrm{F}}^{2}$ represents the Frobenius norm of a matrix, and $k\times|V|$ elements in $\boldsymbol{W}\in\mathbb{R}^{k\times|V|}$ and $k\times s$ elements in $\boldsymbol{H}\in\mathbb{R}^{k\times s}$ are unknown numbers to be optimized. $\boldsymbol{W}$ and $\boldsymbol{H}$ are low-rank matrices, and k is always smaller than the number of nodes $|V|$. Also, k represents the dimensionality of the final feature vector of a node. λ is the regularization coefficient that is responsible for the sensitivity of parameter adjustment during the optimization process. The larger the value of λ is, the smaller the parameter adjustment in every iteration step is, and the more conservative the optimization process is. After the calculation and optimization steps mentioned above, matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ can be obtained. The feature matrix $\boldsymbol{X}$ of the nodes in the network can then be expressed as

$$\boldsymbol{X} = [\boldsymbol{W}^{\mathrm{T}}, (\boldsymbol{H}\times\boldsymbol{Q})^{\mathrm{T}}] \in \mathbb{R}^{|V|\times 2k}. \tag{5}$$

In matrix $\boldsymbol{X}$, the feature of node i in a certain operational mode taking into account its structural and electrical information can be expressed as a $2k$-dimensional vector $\boldsymbol{X}_i$.

3 Node fault consequence evaluation and fault sample set establishment

A complete supervised machine learning sample should contain both features and labels. The features of each node in a certain operational mode can be extracted by the TADW algorithm mentioned in Section 2. The label of each node can be established based on the evaluation of fault consequences of the node.

Some common indicators for node fault consequence evaluation include power angle difference of generators, node voltage change, and post-fault load losses (Wang et al., 2008). Since different power

Table 1 Electrical features of power grid nodes

Feature	Description
Input active power, f_1 (MW)	Total active power injected to a node
Input reactive power, f_2 (Mvar)	Total reactive power injected to a node
Amplitude of voltage, f_3 (kV)	Amplitude of node voltage
Phase angle of voltage, f_4 (rad)	Phase angle of node voltage
Output active power, f_5 (MW)	Total output active power for generator nodes (equal to 0 if a node is not a generator)
Output reactive power, f_6 (Mvar)	Total output reactive power for generator nodes (equal to 0 if a node is not a generator)
Degree of node, f_7	Number of lines connected to a node
Active power load, f_8 (MW)	Total active power load of a node (equal to 0 if there is no load at a node)
Reactive power load, f_9 (Mvar)	Total reactive power load of a node (equal to 0 if there is no load at a node)

networks may have different tolerance to fault consequences of the nodes, the quantification standard should be set in line with actual conditions. For example, guaranteeing the power supply for the important loads is the priority for some power networks when evaluating node fault consequences, while for some other power networks, maintaining transient stability for the whole system may have the highest priority. Obviously, different evaluation standards may lead to different sample labels, resulting in different machine learning models trained by the samples. Therefore, under the machine learning framework, the flexibility in labeling the data according to the actual characteristics and preference is allowed. Then it can be valued as a more universal framework suitable for different power networks.

The ability to maintain post-fault transient stability is crucial to the operation of a power system. Therefore, in this study, the post-fault power angle difference of generators δ, as a common indicator to evaluate transient stability (Wang et al., 2008), is selected to label the fault consequence of a node. What needs to be stressed is that it is assumed that the power network pays more attention to δ when evaluating node importance, but δ is not the only indicator that can be used to evaluate a node fault. As shown previously, each specific power network may have different focuses and preferences when evaluating node importance. The specific choice can be flexible in this framework.

Fault consequences can be categorized into classes of “transient stable” and “transient unstable” according to whether δ is diverging or not during a certain time period after the fault (Wang et al., 2008). While setting label criteria in detail, the two classes can be further divided into several subclasses according to the degree of divergence. The different transient simulation results of the post-fault power angle difference of generators are shown in Fig. 1.

In Fig. 1, for transient stable cases, if the power angle differences of generators are relatively small, shown as case 1, the system is able to keep a high stable margin. In other words, the system can remain stable after the fault even if the fault lasts longer. However, if the power angle differences of generators are relatively large, shown as case 2, the system cannot remain in high stable margin and may turn into an unstable condition if the fault lasts a little longer. For transient unstable cases, given that the power angle differences of generators can diverge into thousands of degrees in a short period, some faults may be extremely severe, which is shown as case 3. Some other faults, which are less severe than those mentioned above, may only cause the power angle differences of generators to diverge into hundreds of degrees or even at the margin of stable and unstable, which is shown as case 4. Therefore, as shown in Table 2, the severity of node fault consequences can be classified into four categories according to the value of max(δ), which is the maximum power angle difference of any two generators (Wang et al., 2008) within a period after node fault occurrence.

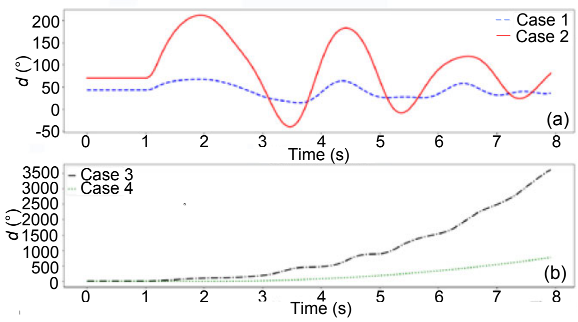

Fig. 1 Different transient simulation results of post-fault power angle difference of generators: (a) stable cases; (b) unstable cases

Table 2 Criteria for labelling node fault consequences

Node fault consequence	Label y
0°<max(δ)≤90°	1
90°<max(δ)≤270°	2
270°<max(δ)≤900°	3
900°<max(δ)	4

In Table 2, label y reflects the severity of the fault, ranking from lowest severity to highest severity in the order of 1 to 4.

Based on the TADW algorithm in Section 2 and the data labeling method mentioned above, the fault sample set for a power network can be established. First, e different kinds of operational modes can be obtained. Under different operational modes, node fault consequences will vary so that enough simulation samples can be accumulated. These samples can reflect the relationship between features of nodes and corresponding importance. Here, random generator outputs and load demands are generated to simulate various operational modes of a system. For a power network with numofLoad loads and numofGen generators, consider an operation mode in which the generator outputs are PGbase_i, QGbase_i (i=1, 2, …, numofGen) and the load demands are PLbase_j, QLbase$j$ (j=1, 2, …, numofLoad) as the basic operational mode. β_j (j=1, 2, …, numofLoad) and γ_i (i=1, 2, …, numofGen) are randomly and independently generated in the ranges of [−0.3, 0.3] and [−0.25, 0.25], respectively. Thus, different operational modes of the system can be obtained by randomly changing the generator outputs and load demands using Eqs. (6) and (7). The slack bus will make up for the imbalance between generator outputs and load demands.

$$\begin{cases} \mathrm{PL}_j = \mathrm{PLbase}_j(1+\beta_j),\ j=1,2,\ldots,\mathrm{numofLoad}, \\ \mathrm{QL}_j = \mathrm{QLbase}_j(1+\beta_j),\ j=1,2,\ldots,\mathrm{numofLoad}, \end{cases} \tag{6}$$

$$\begin{cases} \mathrm{PG}_i = \mathrm{PGbase}_i(1+\gamma_i),\ \ i=1,2,\ldots,\mathrm{numofGen}, \\ \mathrm{QG}_i = \mathrm{QGbase}_i(1+\gamma_i),\ \ i=1,2,\ldots,\mathrm{numofGen}. \end{cases} \tag{7}$$

Then, e different operation modes are simulated under the steady-state condition. The feature matrices $\boldsymbol{X}^1$, $\boldsymbol{X}^2$, …, $\boldsymbol{X}^e$ can be extracted for different nodes by the TADW algorithm. X_i^j represents the feature of node i under operational mode j. Then, the total feature matrix $\boldsymbol{X}'\in\mathbb{R}^{e|V|\times 2k}$ with $e\times|V|$ feature data can be obtained by concatenating all feature matrices $\boldsymbol{X}^1$, $\boldsymbol{X}^2$, …, $\boldsymbol{X}^e$. Finally, a label vector $\boldsymbol{y}\in\mathbb{R}^{e|V|\times 1}$ can be achieved by performing fault transient simulation for each node under each operational mode. The vector $\boldsymbol{y}$ contains $e\times|V|$ fault consequence labels. The feature matrix $\boldsymbol{X}'\in\mathbb{R}^{e|V|\times 2k}$ and the label vector $\boldsymbol{y}\in\mathbb{R}^{e|V|\times 1}$, both with $e\times|V|$ rows, are defined as the fault sample set for this power network.

4 Feature selection

Feature selection is a necessary step in the actual application of machine learning (Jović et al., 2015). The more features there are, the slower the training speed of the machine learning model is. Moreover, unselected features may include some low quality redundant information that will reduce the accuracy of the model.

If the physical significances of features are clear and definite, expert knowledge can be quite useful for evaluating all features. Then, features with strong relevance can be selected manually. However, the features extracted by the TADW algorithm in this study are more like a kind of "fusion feature" considering network structure, electrical features, and their mutual effect. They do not have clear and definite physical significance. Therefore, it is necessary to employ the feature selection algorithm to identify and screen these features.

There are many feature selection algorithms available (Jović et al., 2015), including filter, wrapper, and embedded methods. In this study, a sub-method of an embedded method based on a tree model is applied because of its performance and interpretability. First, a tree-based gradient boosting regression tree (GBRT) model (Friedman, 2002) in the Scikit-learn framework (Pedregosa et al., 2011) is trained by samples in the fault sample set. Then, the importance of each feature is ranked according to how many times the feature is applied in branches of the GBRT model. Features with higher priority according to the ranking are selected and reserved. Assuming that n features are reserved, it is obvious that $n<2k$, and feature matrix $\boldsymbol{X}'$ is then dimensionally reduced to an optimized total feature matrix $\boldsymbol{X}^*\in\mathbb{R}^{e|V|\times n}$, where row vector $\boldsymbol{X}_i^*$ $(i=1,2,\ldots,e\times|V|)$ in $\boldsymbol{X}^*$ represents the optimized feature vector for sample i. The final fault sample set is composed of the optimized feature matrix $\boldsymbol{X}^*$ and label vector $\boldsymbol{y}$.

5 Node importance evaluation based on the SVR model

Among all supervised learning algorithms in machine learning, SVR can better solve the problems of local minima, over-fitting, and under-fitting. Its performance has proved to be impressive and stable in many cases (Basak et al., 2007). Therefore, the SVR model is used in this study to establish the machine learning model.

For $e\times|V|$ samples $\{(X_i^*, y_i)\}$ $(i=1,2,\ldots,e\times|V|)$, all n-dimensional feature vectors can be mapped into a new l-dimensional space using the nonlinear mapping function $p(x)$. It is expected to search for an optimal hyperplane $g(x)=\boldsymbol{D}^{\mathrm{T}}\boldsymbol{p}(x)+b$ in this new space, where $\boldsymbol{D}$ is an l-dimensional weight vector and b is bias term. Therefore, the distances from this optimal hyperplane to all samples $|y_i-g(X_i^*)|$ are smaller than a given precision value τ. The penalty coefficient C (C>0) and non-negative slack variable ε and ε^* are introduced to turn the objective of searching for an optimal hyperplane into a convex quadratic optimization problem:

$$\begin{aligned}&\min L(\boldsymbol{D},b,\varepsilon,\varepsilon^*)=\frac{1}{2}\|\boldsymbol{D}\|^2+C\sum_{i=1}^{e|V|}(\varepsilon_i+\varepsilon_i^*)\\ &\text{s.t.}\\ &\quad y_i-g(X_i^*)\le\tau+\varepsilon_i,\\ &\quad g(X_i^*)-y_i\le\tau+\varepsilon_i^*,\\ &\quad \varepsilon_i\ge0,\ \varepsilon_i^*\ge0.\end{aligned}\tag{8}$$

The hyper-parameter C should be determined in advance to balance the training accuracy and generalization ability of the SVR model. A common predetermination method based on cross validation inspects the performance of different parameters on the validation set (Browne, 2000). In this study, the coefficient of determination (CD) is used as the measurement criterion of performance. After the hyper-parameter C is determined, the optimal hyperplane, which is expressed as a regression function (9), can then be calculated by putting training data into Eq. (8):

$$g(x)=\sum_{i=1}^{e|V|}(a_i-a_i^*)\boldsymbol{p}^{\mathrm{T}}(X_i^*)\boldsymbol{p}(x)+b.\tag{9}$$

where a_i, a_i^*, and b are all parameters to be trained in the SVR model. While calculating these parameters, some mathematical techniques, such as turning calculations into corresponding dual problems or introducing a Gaussian kernel function $V(X_i^*,x)$ to replace $\boldsymbol{p}^{\mathrm{T}}(X_i^*)\boldsymbol{p}(x)$ (Basak et al., 2007), can be used to make the training process easier. Because of space limitations, these techniques will not be discussed in detail in this paper.

After the SVR model is trained using the sample set, the importance of nodes in a particular new operational mode can be calculated using the operational information. The detailed calculation process of the proposed algorithm can be summarized as Fig. 2.

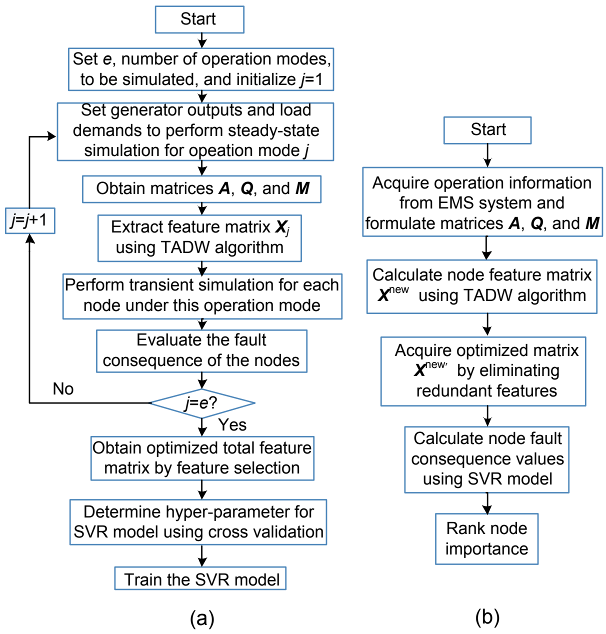

Fig. 2 Flowchart of the proposed algorithm: (a) offline training for the SVR model; (b) online evaluation for node importance

Fig. 2a shows the process of offline training for the SVR model. First, the number of operational modes to be simulated is set to e. Second, the steady-state simulations for all operational modes are performed. For each operational mode j, matrices $\boldsymbol{A}$, $\boldsymbol{Q}$, and $\boldsymbol{M}$ can be calculated accordingly. Then, feature matrix $\boldsymbol{X}^j$ for all nodes in this operational mode is extracted using the TADW algorithm. Third, the fault transient simulation and fault consequence evaluation for each node in each operational mode are performed to determine labels for all samples. After all sample data under e operation modes is obtained, features with lower importance are eliminated to acquire the optimized feature matrix using the feature selection algorithm. Then optimized sample features and sample labels are integrated to determine the optimal hyper-parameter C by a cross-validation process. Finally, the SVR evaluation model is acquired by training on the whole sample set.

Fig. 2b shows the process of online evaluation of node importance. For a particular real-time operational mode to be evaluated, operational information can be acquired from the energy management system, and matrices $\boldsymbol{A}$, $\boldsymbol{Q}$, and $\boldsymbol{M}$ are formulated accordingly.

Then feature matrix $\boldsymbol{X}^{\text{new}}$ is extracted using the TADW algorithm. After that, features with lower importance acquired in process (a) are eliminated to obtain the optimized matrix $\boldsymbol{X}^{\text{new}'}$. Finally, evaluation values of fault consequences for each node can be acquired using the SVR model, and their importance can be ranked accordingly.

6 Case study

In this section, the IEEE-39 power system with 46 branches, 19 loads, and 10 generators shown in Fig. 3 is used to verify the efficiency of the proposed method.

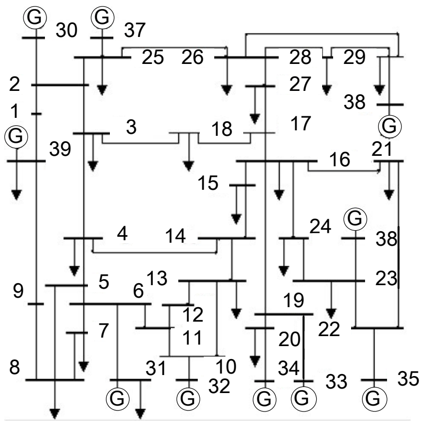

Fig. 3 IEEE-39 power system

MATLAB toolbox Matpower (Zimmerman et al., 2011) and PSAT (Vanfretti and Milano, 2007) are used for the steady-state and transient simulations, respectively. The fourth-order generator model is used in the simulation. Applying three-phase short circuit faults at the nodes at 1.0 s and clearing the faults after 0.2 s to evaluate their fault consequences, the transient process of the system during a period of 7 s after the fault is simulated. After a total of 100 operational modes are obtained by the means described in Section 3, steady-state simulations are performed. Then, for each operational mode, matrices $\boldsymbol{A}$, $\boldsymbol{Q}$, and $\boldsymbol{M}$ are accordingly calculated, and the steady-state feature vectors are extracted for all the 39 nodes using the TADW algorithm. During the TADW feature extraction process, by comparing the network scale of the IEEE-39 system with those of networks studied in Yang et al. (2015), the value of parameter k is set at 30 so that each node can be expressed as a 60-dimensional vector first. This 60-dimensional vector can later be dimensionally reduced using the feature selection algorithm. For each operational mode, there are 39 samples corresponding to the 39 nodes, so that a sample set with 3900 samples can be formed based on the 100-mode simulation process. Finally, transient simulations of certain faults at each node on each operational mode are performed and the label for each sample is obtained by the data labeling method described in Table 2. So far, after the feature selection process, the final sample set consisting of feature matrix $\boldsymbol{X}^*$ and label vector $\boldsymbol{y}$ can be established.

Several parameters, such as parameter λ, the feature selection method, and the number of features eliminated, can influence the result of the final sample set. For this IEEE-39 system, to acquire a more accurate feature expression by performing low-rank matrix factorization, the parameters adjustment of $\boldsymbol{W}$ and $\boldsymbol{H}$ in the optimization iteration process could be bolder, so λ can be given a small value. The results when λ is given the values of 0.05, 0.1, and 0.2 are obtained. For the feature selection process, the GBDT model and the mutual information method are studied. For each feature selection method, different numbers of features eliminated are studied.

The best combination of the parameters is obtained by the cross-validation process. For each combination of the parameters, a different final sample set can be obtained correspondingly. Then, this final sample set can be divided into the training set and the validation set in the proportion of 4:1. Then, the variation range of C is set as [0.5, 15] with a step size of 0.5. For each different value of C, the model is trained using the training set, and the corresponding coefficient of determination (CD) is acquired using the validation set. The best value of hyper-parameter C under each combination of the parameters is then determined as that with the largest corresponding CD value. Finally, the best combination of parameters is determined as that with the largest CD value under the best value of C. When λ=0.1 and the mutual information method is used as the feature selection method, the CD-C curves under different numbers of features eliminated are shown in Fig. 4.

When λ=0.1 and the GBDT model is used as the feature selection method, the feature importance calculated by the model is shown in Fig. 5. The CD-C curves under different numbers of eliminated features are shown in Fig. 6.

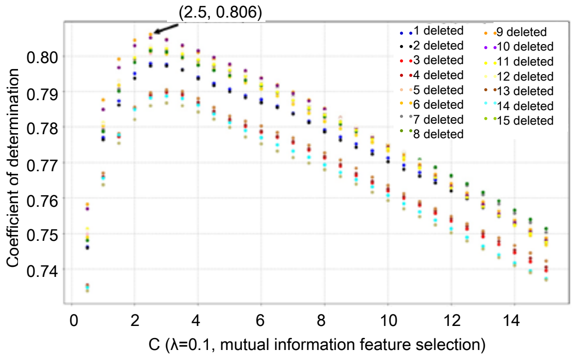

Fig. 4 CD-*C* curves when λ=0.1 and the mutual information method is used as the feature selection method

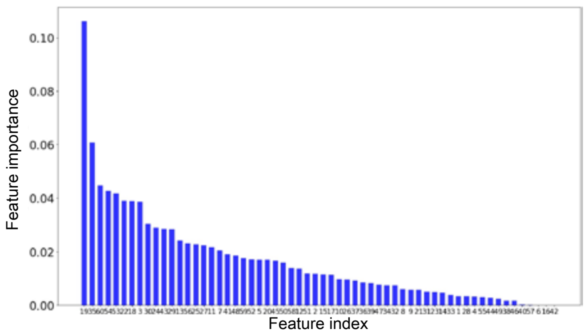

Fig. 5 Feature importance calculated by the GBDT model when λ=0.1

Fig. 6 CD-*C* curves when λ=0.1 and the GBDT model is used as the feature selection method

When the mutual information method is used, the largest CD value is 0.806 when 9 features are eliminated; when the GBDT model is used, the largest CD value is 0.809 when 10 features are eliminated. The results illustrate that the GBDT model is better for feature selection.

When λ is given the values of 0.05 and 0.2, the CD-C curves are shown in Figs. 7–10. The best CD values are 0.774 and 0.781, respectively. The results show that, when λ=0.1, the GBDT model is used for feature selection and the number of features eliminated is 10, the SVR model can obtain its best performance. So, this combination of parameters is chosen as the final choice. Then, the final SVR model can be trained under this best combination of parameters using the whole sample set.

Finally, node importance under a new operational mode can be evaluated using this SVR model:

First, for this operational mode to be evaluated, the corresponding matrices $\boldsymbol{A}$, $\boldsymbol{M}$, and $\boldsymbol{Q}$ can be calculated based on the operational information. Then, the feature vectors for each node are extracted using the TADW method and the 10 features with lower importance mentioned above are eliminated. After that, the evaluation result of the fault consequence of each node and corresponding node importance ranking can be acquired using the SVR model. The results of the most important 15 nodes and the least

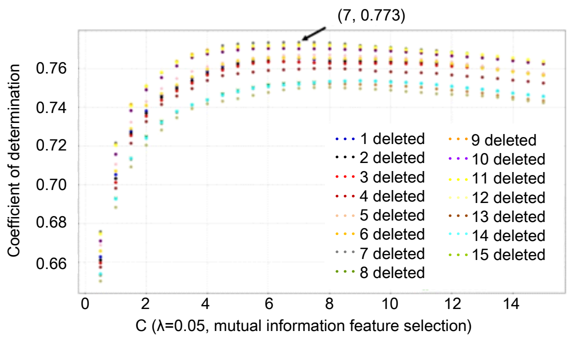

Fig. 7 CD-*C* curves when λ=0.05 and the mutual information method is used as the feature selection method

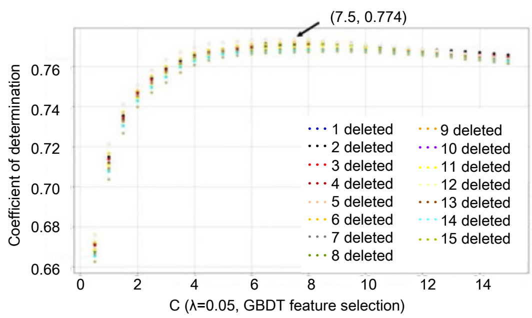

Fig. 8 CD-*C* curves when λ=0.05 and the GBDT model is used as the feature selection method

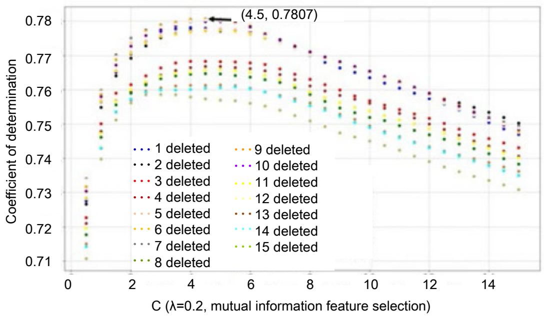

Fig. 9 CD-*C* curves when λ=0.2 and the mutual information method is used as the feature selection method

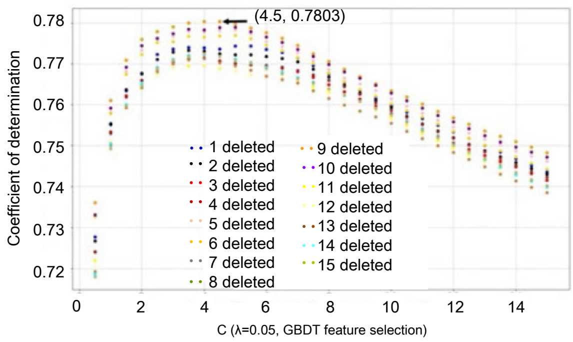

Fig. 10 CD-*C* curves when λ=0.2 and the GBDT model is used as the feature selection method

important 10 nodes are shown in column “Method A” in Table 3. The results are compared with the evaluation results calculated using the electrical betweeness method proposed by Xu et al. (2010) and the PageRank method proposed by Li et al. (2014). These two methods use an index system based on complex network theory and a web page ranking algorithm, respectively. The evaluation results are shown in columns “Method B” and “Method C” in Table 3. The nodes with severe fault consequences under this operation mode according to PSAT simulation are shown in Table 4.

As shown in Table 4, nodes that may cause the maximum power angle difference of generators to reach a value of more than 900° are {16, 17, 19, 20, 22, 24, 25, 28, 29, 33, 34, 38}. The fault of these nodes may cause severe consequences, so much more attention should be paid to these nodes in this operational mode. According to the ranking results evaluated by the proposed method shown in Table 3, these nodes have importance ranks of {2, 5, 1, 6, 3, 7, 12, 14, 9, 10, 4, 8}. It means that nodes {16, 17, 19, 20, 22, 24, 25, 29, 33, 34, 38} are all among the 12 most

Table 3 Evaluation results of node importance ranking by three methods

Ranking	Node index		
	Method A	Method B	Method C
1	**19**	**16**	4
2	**16**	**17**	8
3	**22**	3	**16**
4	**34**	**34**	27
5	**17**	26	15
6	**20**	**24**	3
7	**24**	**25**	6
8	**38**	15	26
9	**29**	39	**24**
10	**33**	14	**17**
11	23	**28**	11
12	**25**	6	9
13	21	**19**	5
14	**28**	18	**20**
15	26	8	23
30	8	28	22
31	36	38	25
32	7	31	29
33	31	32	31
34	39	35	14
35	30	37	21
36	12	36	28
37	1	7	13
38	9	30	1
39	32	12	12

Method A: proposed method; Method B: electrical betweeness method (Xu et al., 2010); Method C: PageRank method (Li et al., 2014)

Table 4 The nodes with severe fault consequences according to PSAT simulation

Index of node that may cause a large power angle difference of generators	Maximum power angle difference (°)
19	6328
16	5945
34	5710
17	5213
22	4820
24	4602
29	4563
20	3584
38	2960
33	2840
25	1562
28	960

important nodes. Node 28 is ranked as the 14^{th} important node, which is not among the top 12 but still has a relatively high ranking. In the 12 most important nodes discovered by SVR, only the fault of node 23, which has a ranking of 11, does not cause a severe transient unstable consequence of the system according to PSAT simulation. As for the 10 least important nodes obtained by SVR, their fault will not cause a severe consequence to the system. The maximum power angle differences of generators are all within 120°. It can be seen that the proposed method can effectively select important nodes with severe fault consequences according to the information learned from the sample set.

As shown in Table 3, only 7 of the 15 most important nodes acquired by method B and 4 of the 15 most important nodes acquired by method C match the nodes with severe fault consequences shown by PSAT simulation. That is to say, the index systems proposed by Xu et al. (2010) and Li et al. (2014) can only partly reflect the transient stability consequences of a power grid node. However, the fault of nodes with a high ranking score according to method B, {3, 26, 15, 39, 14, 6, 18, 8}, and nodes with a high ranking score according to method C, {4, 8, 27, 15, 3, 6, 26, 11, 9, 5, 23}, do not cause any severe unstable consequences. {28, 38} have a very low electrical betweenness value, and nodes {22, 25, 29, 28} have a very low PageRank value. However, they are among the nodes with severe fault consequences. The results illustrate that, as we have asserted before, power systems vary in many perspectives such as scale, structure, and even tolerance on different consequences caused by the node fault. So, a priori index may not be able to cover all the useful information when evaluating node importance, and can hardly meet the demand of universality.

Both methods B and C use an index system. As described in the introduction, different index systems can give very different ranking results. This is clearly shown in Table 3. Method B emphasizes the nodes' contribution of linking the generators and the loads. Therefore, nodes 16 and 17, which have larger electrical betweenness values, are more important. However, method C uses the PageRank algorithm which emphasizes the inflow power of a node, so it gives a totally different ranking result for this same operational mode. In conclusion, due to the bias and subjectivity of the methods using an index system, it is hard for the operators to choose a suitable method for a specific power network. However, the data-driven framework proposed in this study builds the model based on the data generated directly by the specific power network. It can capture different characteristics of different systems, and is more universal than methods based on any index system.

As for calculation speed, it takes about 2.9–3.8 s for PSAT to perform a node fault transient simulation for the IEEE-39 system. To rank the importance of all 39 nodes according to the pure simulation method, transient fault simulation for all 39 nodes needs to be performed and the total time reaches 113.1–148.2 s. The computation of the electrical betweenness index needs multiple calculations of power flows under different output and demand conditions, and the time required is 53 s. The computation of the PageRank index costs 1.6 s. However, the evaluation method proposed in this paper spends only 0.15 s on feature extraction by the TADW algorithm and 0.043 s on importance calculation and ranking by the SVR model. The total time is only 0.193 s. The characteristics of the machine learning framework allow us to perform rapid online calculation and use rich information gathered offline by simulation.

7 Conclusions

In this paper, an artificial intelligence based method using network embedding technology is proposed for importance evaluation of power grid nodes. The SVR model is used to automatically learn the relationships between features of nodes and the severity of their fault consequences. As part of the evaluation process, the TADW algorithm, as an important link, is employed to extract features of the power grid nodes. A case study shows that the proposed method is able to perform reasonable importance evaluation and ranking of the nodes based on the information learned from the sample set. Compared to the traditional methods using an index system, the proposed method has the following important advantages:

First, in terms of universality, under the framework in this paper, a particular sample set can be established for each specific power network. This sample set can reflect specific physical characteristics of the power network. The flexibility in the data labeling process taking account of the specific preference to different node fault consequences also guarantees universality. Thus, the sample set can avoid the possible bias of methods based on an index system given the diversity of power networks.

Second, on calculation speed, the proposed method can perform rapid and efficient online calculation according to the operational information of the power grid. The time consuming steady-state and transient simulation process, which aims at accumulating enough samples for the SVR model, is performed offline.

In further study, some other NE methods can be studied to realize more specific feature extraction for different power networks. Some other machine learning methods, such as deep learning, can also be studied to achieve more accurate results.

Compliance with ethics guidelines

Hui-fang WANG, Chen-yu ZHANG, Dong-yang LIN, and Ben-teng HE declare that they have no conflict of interest.

References

Albert R, Albert I, Nakarado GL, 2004. Structural vulnerability of the North American power grid. *Phys Rev E*, 69(2): 025103(R). https://doi.org/10.1103/PhysRevE.69.025103

Angra S, Ahuja S, 2017. Machine learning and its applications: a review. Int Conf on Big Data Analytics and Computational Intelligence, p.57-60. https://doi.org/10.1109/ICBDACI.2017.8070809

Arianos S, Bompard E, Carbone, et al., 2009. Power grid vulnerability: a complex network approach. *Chaos*, 19(1): 013119. https://doi.org/10.1063/1.3077229

Bai JL, Liu TQ, Cao GY, et al., 2008. A survey on vulnerability assessment method for power system. *Power Syst Technol,* 32(S2):26-30 (in Chinese). https://doi.org/10.13335/j.1000-3673.pst.2008.s2.039

Basak D, Pal S, Patranabis DC, 2007. Support vector regression. *Neur Inform Process Lett Rev*, 11(10):203-224.

Browne MW, 2000. Cross-validation methods. *J Math Psychol*, 44(1):108-132. https://doi.org/10.1006/jmps.1999.1279

Cai ZX, Wang XH, Ren XN, 2012. A review of complex network theory and its application in power systems. *Power Syst Technol*, 36(11):114-121 (in Chinese). https://doi.org/10.13335/j.1000-3673.pst.2012.11.002

Chen WZ, Zhang Y, Li XM, 2015. Network representation learning. *Big Data Res*, 1(3):8-22 (in Chinese).

da Silva AML, Jardim JL, de Lima LR, et al., 2016. A method for ranking critical nodes in power networks including load uncertainties. *IEEE Trans Power Syst*, 31(2):1341-1349. https://doi.org/10.1109/TPWRS.2015.2413847

Fan WL, Ping H, Liu ZG, 2016. Multi-attribute node importance evaluation method based on Gini-coefficient in complex power grids. *IET Gener Transm Distrib*, 10(9): 2027-2034. https://doi.org/10.1049/iet-gtd.2015.0803

Friedman JH, 2002. Stochastic gradient boosting. *Comput Stat Data Anal*, 38(4):367-378. https://doi.org/10.1016/S0167-9473(01)00065-2

Grover A, Leskovec J, 2016. Node2vec: scalable feature learning for networks. Proc 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining, p.855-864. https://doi.org/10.1145/2939672.2939754

Jović A, Brkić K, Bogunović N, 2015. A review of feature selection methods with applications. Proc 38th Int Convention on Information and Communication Technology, Electronics and Microelectronics, p.1200-1205. https://doi.org/10.1109/MIPRO.2015.7160458

Ju WY, Li YH, 2012. Identification of critical lines and nodes in power grid based on maximum flow transmission contribution degree. *Autom Electr Power Syst*, 36(2): 6-12 (in Chinese).

Li CB, Liang JZ, 2009. A novel method of power grid differential planning. *Autom Electr Power Syst*, 33(24):11-15 (in Chinese). https://doi.org/10.3321/j.issn:1000-1026.2009.24.003

Li CB, Liu WC, Cao YJ, et al., 2014. Method for evaluating the importance of power grid nodes based on Pagerank algorithm. *IET Gener Transm Distrib*, 8(11):1843-1847. https://doi.org/10.1049/iet-gtd.2014.0051

Lin ZZ, Wen FS, Wang HF, et al., 2017. CRITIC-based node importance evaluation in skeleton-network reconfiguration of power grids. *IEEE Trans Circ Syst II*, 65(2):206-210. https://doi.org/10.1109/TCSII.2017.2703989

Nasiruzzaman ABM, Pota HR, 2011. Critical node identification of smart power system using complex network framework based centrality approach. North American Power Symp, p.1-6. https://doi.org/10.1109/NAPS.2011.6025194

Pan XD, Wu J, Liu DC, et al., 2014. A method for constructing core backbone grid in differential planning based on importance degrees of components. *Autom Electr Power Syst*, 38(19):40-46 (in Chinese). https://doi.org/10.7500/AEPS20130522006

Pedregosa F, Varoquaux G, Gramfort A, et al., 2011. Scikit-Learn: machine learning in Python. *J Mach Learn Res*, 12:2825-2830.

Perozzi B, Al-Rfou R, Skiena S, 2014. DeepWalk: online learning of social representations. Proc 20th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining, p.701-710. https://doi.org/10.1145/2623330.2623732

Ross S M, 1983. Stochastic Processes. Wiley, New York, SUA.

Tan YD, Li XR, Cai Y, et al., 2014. Critical node identification for complex power grid based on electrical distance. *Proc CSEE*, 34(1):146-152 (in Chinese). https://doi.org/10.13334/j.0258-8013.pcsee.2014.01.017

Vanfretti L, Milano F, 2007. Application of the PSAT, an open source software, for educational and research purposes. IEEE Power Engineering Society General Meeting, p.24-28. https://doi.org/10.1109/PES.2007.385952

Wang B, Fang BW, Wang YJ, et al., 2016. Power system transient stability assessment based on big data and the core vector machine. *IEEE Trans Smart Grid*, 7(5):2561-2570. https://doi.org/10.1109/TSG.2016.2549063

Wang HF, Shan ZB, Ying GL, et al., 2017. Evaluation method of node importance for power grid considering inflow and outflow power. *J Mod Power Syst Clean Energy*, 5(5): 696-703. https://doi.org/10.1007/s40565-016-0234-3

Wang XF, Song YH, Irving M, 2008. Modern Power Systems Analysis. Springer, Boston, USA.

Xu L, Wang XL, Wang XF, 2010. Cascading failure mechanism in power grid based on electric betweenness and active defence. *Proc CSEE*, 30(13):61-68 (in Chinese). https://doi.org/10.13334/j.0258-8013.pcsee.2010.13.010

Xu LX, Liu JY, Liu Y, et al., 2014. Node importance classified comprehensive assessment. *Proc CSEE*, 34(10):1609-1617 (in Chinese). https://doi.org/10.7500/AEPS20130522006

Xu Y, Dong ZY, Meng K, et al., 2011. Real-time transient stability assessment model using extreme learning machine. *IET Gener Transm Distrib*, 5(3):314-322. https://doi.org/10.1049/iet-gtd.2010.0355

Yang C, Liu ZY, Zhao DL, et al., 2015. Network representation learning with rich text information. Proc 24th Int Conf on Artificial Intelligence, p.2111-2117.

Zimmerman RD, Murillo-Sanchez CE, Thomas RJ, 2011. MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Trans Power Syst*, 26(1):12-19. https://doi.org/10.1109/TPWRS.2010.2051168