Topological Convolutional Neural Networks for Transient Stability Assessment on Massive Historical Online Power Grid Data

Xianggen Liu, HaoTian Cui, Yanhao Huang,[1]

Tsinghua University, Beijing 100084, China
Department of Biomedical Engineering, School of Medicine, IDG/McGovern Institute for Brain Research

Abstract

The power system is a highly nonlinear and non-autonomous complex dynamic system with chaos. Based on historical online data generated by dynamic stability analysis(DSA) system, transient stability assessment can be realized and utilized to make power system operate continuously and stably. On the base of huge electric power system operation characteristics, conventional methods such as support vector machine(SVM) and Phasor measurement units(PMUs) rely on the high quality of manually selected features to achieve better results and cannot model the topological structures of the power grids. Here, inspired from the convolutional nerual networks(CNN) for images, we propose a novel deep learning framework, called Topological Convolutional Networks(TCNN), to specifically model the topological structures of power grids and efficiently extract features. Based on adjacent matrix, topological convolution updates the node's representation using the representation of its neighbors, and topological pooling operation summarizes the topological local features. The results show that the proposed model outperforms the other strong baselines by a large margin on the measurements of both compressing the candidate fault set and reducing the misclassification of unstable samples.

Keywords: deep Learning, topological convolution, power grid, transient

☆Fully documented templates are available in the elsarticle package on CTAN.
[1]These authors contributed equally.

1. Introduction

In order to maintain the stable operation of power system, where various disturbances are often encountered [1], it is important to assess transient stability quickly and accurately. Ensuring secure and stable operation of large-scale power systems, and addressing different contingencies are the most formidable challenges that power engineers face today. In particular, short-circuit faults are the dominant inducing factors causing catastrophic power grid accidents and dynamic stability analysis(DSA) is an important approach to deal with these contingencies. The on-line DSA system is generating data continuously, accumulating a great mounts of historical data. At present, the DSA focuses on the fault set of main protection while other fault sets are also important since a small fault may be magnified by networks. If we take more fault sets into consideration, the corresponding computation cost and calculation time are required, such as fault set of back-up protection, doubling the size of total fault set. Under the existing hardware and software conditions, if all of the data are directly used for DSA, it will be difficult to meet the online needs. Therefore, Seeking an effective transient stability assessment(STA) method to compressing the candidant set for DSA with high recall is the right way to address the above problem.

Over the past two decades, there have been several solutions to address this problem [2, 3, 4]. however, they are roughly the same technical route to some extent [5], summarized as: 1) generating data by the power simulation calculation system, 2) selecting the representative characteristic quantity of the system operation according to the existing power system knowledge; 3) designing an appropriate machine learning algorithm such as support vector machine(SVM) [6] and hidden markov model(HMM) [7] for task of transient stability assessment.

However, the above technical route has two significant drawbacks: 1) the strategy of feature selection is manually designated, and there is no guarantee of the completeness of features for STA. 2) Researchers have conducted extensive explorations of power system transient stability assessment methods for machine learning, but the algorithms involved are limited to shallow learning methods such as support vector machine (SVM) [6], artificial neural networks (ANNs) [8, 9] and decision trees (DT) [4]. Shallow learning methods also work not so well due to its limited generalization ability and fitting ability when solving complex classification problems. In recent years, the rapid development of deep learning has provided a brand-new idea for the transient stability assessment of power systems by learning more useful features through a large amount of training data, and further improving the performance of STA [10].

Deep learning [11] is inspired by the cognitive model of the human brain, which is to extract information through multi-level abstraction, make it possible to work with raw input as the data stand. But the deep learning algorithms are not well studied on transient stability assessment. In addition, Commonly used deep learning models such as deep belief networks (DBNs) [12], convolutional neural networks (CNNs) [13], and stacked auto-encoders (SAE) [14] are designed for processing images, texts and audios. the data generated by power systems possess various features of topological structures, which cannot be encoded naturally by the existing models naturally. In on-line DSA systems, a group of differential equations, representing the real physical components and their connections, are solved step by step using numerical techniques in order to get the actual values of the state variables. These state variable values then yield important information regarding transient stability and each variables are highly entangled with all of the others, where physical topology plays a critial role.

To specifically extract topological and high dimensional features of power grids, we present a novel deep learning framework, deep topological convolutional neural networks(TCNN), to represent the state of complex power system and assess its transient stability. The topological convolution and topological pooling operations are the key components of TCNN and effectively extract features for transient stability assessment. Our paper is the first applying topological convolution to extract structured electrical characteristics. Based on adjacent matrix, topological convolution updates the node's representation using the representation of its neighbors and it own. Inspired by maxpooling operation in Convolutional Neural Networks(CNN) [13], we also propose a novel topological pooling operation to summarize the topological local features. Topological pooling transverts the node's representation by selecting the max/mean values among the neighbors. The design of both of the proposed operations fits the topological structure of power grids and utilize the topological features that the previous models have not explored.

We evaluate TCNN on actual industrial dataset generated from a large-scale power grid with four fair metrics. The results show that the proposed model outperforms the other strong baselines by a large margin on the measurements of both compressing the candidate fault set and reducing the misclassification of unstable samples. The further study shows that TCNN can not only extract senseful features and these features can further help SVM to achieve better performance on the above measurements.

2. Related works

2.1. Methods of transient stability assessment

With the expansion of the power grids and the access of large-scale renewable energy sources, the security and stability analysis of the power system are facing more severe challenges. Transient energy functions [15, 16, 17] have been employed to assess the system stability since 1980s. However, this approach has its drawbacks when employed on practical large power grids due to the model simplifications required. An alternative of this direct method is fast time-domain (TD) simulation [18], where a given set of credible contingencies are simulated in an off-line manner to guide the design and tuning of the control system. Meanwhile, accurate TD simulations require complete information of the grid and the disturbance, imposing a heavy computational burden. With the rapid deployment of phasor measurement units(PMUs) [2], a collection of methodologies have been proposed for TSA based on the real-time system operating state. Utilizing the pre- and postcontingency system dynamics, techniques such as piecewise constant-current load equivalent method [19], emergency single machine equivalent [20], and post-disturbance trajectory analysis [3] were developed for on-line TSA. While accurate assessments can be achieved by these methods, their high computational complexities prevent them from being employed in practical post-contingency TSA.

To realize fast real-time TSA status prediction, machine learning and fuzzy logic techniques have been widely adopted as alternative approaches for TSA in recent years, e.g. decision tree methods [4], support vector machine(SVM) [6], artificial neural networks (ANN) [8, 21], and fuzzy knowledge based systems [22]. Different from the conventional analytical methods, these machine learning methods extract the relationship between the system parameters and the corresponding stability conditions utilizing predefined transient stability datasets. The accuracies of these shallow computational models largely depends on the quality of selected features and are limited by their generalization.

2.2. Deep learning methods and graphical neural networks

Recently, deep neural networks have become the prevailing method, as they are more powerful classifiers that can work with raw data *per se*, and achieve promising results [23, 10] in power system. Deep learning has demonstrated distinguished abilities of feature extraction, classification and vast parameter learning, and achieved a wide range of applications in the frontier areas of image recognition, video classification, document understanding. Analogically introducing deep learning to the power system is thus a promising research field.

Neural networks that operate on graphs have previously been introduced in [24, 25] as a form of recurrent neural network. Duvenaud [26] introduced a convolution-like propagation rule on graphs and methods for graph-level classification. Their approach requires to learn node degree-specific weight matrices which does not scale to large graphs with wide node degree distributions. Our model instead uses a single weight matrix per layer and deals with varying node degrees through an appropriate normalization of the adjacency matrix. Our method is based on spectral graph convolutional neural networks, introduced in [27], extended by [28] with fast localized convolutions, and later similified by [29]. Further to these tasks, we deepen the networks by stacking the convolution layer with topological pooling, and thus our model can extract high-level representation and improve scalability and classification performance in large-scale networks.

In a nutshell, the key contributions of our method can be summarized as follows:

- This paper is the first to apply topological/graphical convolution to naturally process topological electrical features.
- We are the fist to propose topological pooling operation to summarize local features with topological structures.
- Experiments on actual industrial dataset demonstrate that the TCNN outperforms several strong baselines by a large margin.
- Verify that the TCNN can extract senseful features. The extracted features can be applied directly to clustering and classification, which can futher improve the classification performance of svm.
- Verify that in the deep network, the empirically statistics features of the power system are still necessary.

The rest of the paper is organized as follows. We briefly describe the transient stability assessment task firstly. In section 4, we present the TCNN model and the corresponding training process. We have performed several experiments and show the results in section 5. Finally we make our conclusion in section 6.

3. Task

Time domain simulation is the most accurate means of determining whether a system will remain stable following a particular contingency. The differential equations of the system are solved step by step using numerical techniques in order to get the actual values of the state variables. When a fault(short circuit or open circuit) occurs, we need to judge whether the system will turn to an acceptable steady state condition without transient instability or voltage instability problems [18] as soon as possible. The above task is called transient stability assessment(TSA).

Thus, we define the transient stability assessment problem as a binary-classification task, stable or not, given the state of current power system, the topological structure and the index of fault alternating current (AC) line. The state of current power system consists of variables of the system before, during the failure, including external injection parameters such as active and reactive power of the system and dynamic parameters such as rotor kinetic energy and acceleration of the generator, and the more details are described at 4.1. The representation of topological structure will described detailly in the following.

3.1. Topological structures of power grids

The power network is a typical complex network with dynamics. An electrical system usually includes transformers, generators, busbars, alternating current (AC) lines, etc. In this paper, the AC line is regarded as the node and the other variables or equipments are regarded as the electrical characteristics aligned the AC lines, and the edges between the nodes are defined here as mutual influences between the nodes, such as generators and busbar.For example, as shown in Figure 1, the parameters of a busbar on the i-th side[2] of AC line belongs to the electrical characteristics of the AC line. The nodes and edges in the abstract topological structure have a practical meaning and simiplifies the problem.

The current state of power system, the topological structure and the index of fault alternating current (AC) line compose the complete information of transient stability assessment. Formally, we build a dataset denoted by $\boldsymbol{X}$ which

[2]There are two sides of a AC line, named i- and j- side.

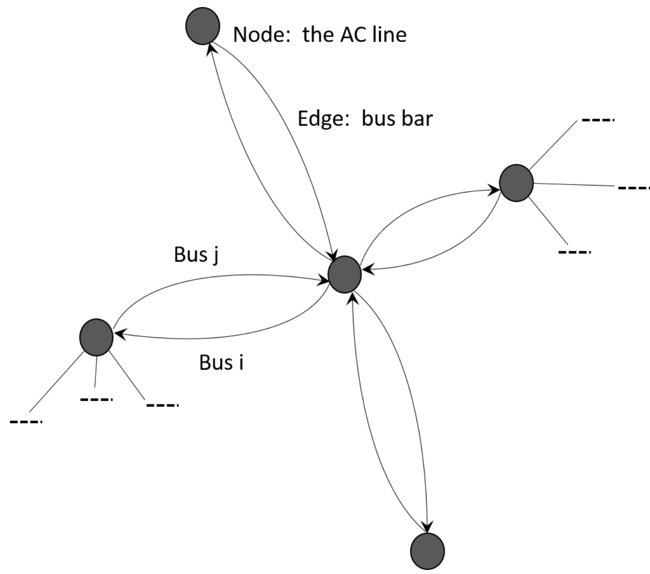


Figure 1: Diagram of the topological structures of of power grids

stores the above information and $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N-1}, \boldsymbol{x}_N]$, where N is the number of samples. Note that the global topological structure is the same and only the fault node is changing instance by instance. Representing the global structure make no sense while the local topological structure play an important role in propagating, amplifying or reducing the fault signal. In addition, It is neccessary to clip the topological graph into a predefined size(K), because the operations (such as spatial convolution) to extract features from topological graph are expensive in terms of model size and computation, both of which grow quadratically with respect to graph size. Here, for each sample of data, we use $\boldsymbol{A}_i$ to represent the topological connection of the i-th sample, that is, the adjacent matrix.

$$\boldsymbol{A}_i = \begin{bmatrix} 1 & a^i_{12} & a^i_{13} & \ldots & a^i_{1K} \\ a^i_{21} & 1 & a^i_{23} & \ldots & a^i_{2K} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a^i_{K1} & a^i_{K2} & a^i_{K3} & \ldots & 1 \end{bmatrix}$$

where a^i_{jk} is a binary varible to denote whether the j-th node and k-th node have a direct connection in the topological of i-th sample.

Besides the topological structure, all the electrical features $\boldsymbol{E}_i$ in this local network, state features(combining local and statistical features, denoted as $\boldsymbol{s}_i$) are also important for the on-line stability transient prediction, and thus all the features of a particular instance can be represented by

$$\boldsymbol{x}_i = [f_i, \boldsymbol{E}_i, \boldsymbol{s}_i, \boldsymbol{A}_i] \tag{1}$$

, where f_i is the index of the fault point in the adjacent matrix, $\boldsymbol{E}_i \in R^{K \times D_e}$, $\boldsymbol{s}_i \in R^{D_s}$, and $A_i \in \mathbb{N}^{K \times K}$. D_e and D_s are the dimension size of electrical features and state features of a node. In power system, the impact of the fault usually propagates along the physical connection. This definition maintains the topological structure of the power network, and meanwhile facilitates the neural network to mine the inner features. The graph structure can retain both the tidal current state parameters and topology information, making some methods catch the relevance of the grid propagation mode with the topological structure.

4. Methodology

4.1. Feature Selection

There are massive information to describe the state of power grid that most of algorithms cannot afford. The selection of candidate attributes should be done to adequately characterize the properties of the system with modest size of features. Based on past experience [30], several parameters such as generator power (real and reactive) outputs, power flows on critical lines, phase angle differences, and information about network topology can be used. This paper considers the stability of the power system, especially the stability of the system when the N-1 fault occurs, and the information related to the two major aspects, the global status of the system and the local characteristics of the fault. For example, when the overall system is stable, a dangerous failure may not cause the network to lose stability; the same failure causes the entire system to become unstable when the system is in a non-stability critical state. Therefore, we divide the characteristics required for stability judgment into two parts: the global characteristics of the power network and the local features of the fault.

4.1.1. Global Features

The description of global feature quantities mainly includes two forms: the first is the physical variables that directly describe original power flow. The second is the populational statistics about the first form, which is refered as to global features since the statistics reflects global distribution of the corresponding physical variable. Table 1 presents the selected original physical variables.

Phy. Variable	Description	Phy. Variable	Description
V	Bus voltage	$cos\theta_L$	Power load factor
θ	Bus voltage angle	P_{AC}	Active power of AC line
P_G	Active power of dynamo	Q_{AC}	Reactive power of AC line
Q_G	Reactive power of dynamo	P_{DC}	Active power of DC line
$cos\theta_G$	Active power factor	Q_{DC}	Reactive power of DC line
P_L	Active load	Q_L	Reactive load

Table 1: The selected physical variables that reflects electrical characteristics of power system.

At the same time, we can design the statistics shown in the Table 2 based on the experience of power experts [30]. According to the above strategy, the

Phy. Variable	Description	Phy. Variable	Description
max	Maximum	$median$	Median
min	Minimum	msd	Standard deviation of median
$mean$	Average value	$1st$	1/4 quantiles
std	Standard deviation	$3st$	3/4 quantiles
$skew$	Measure of skewness	mad	Median absolute deviation
$kurt$	Measure of kurtosis	q_{inter}	Inter-quartile range
$mj10$	Trimmed mean in last 10%	std_{mj10}	Standard deviation of trimmed mean in last 10%

Table 2: The statistics regarding to the above physical variables.

dimension of selected global features(D_s) is 196, resulted from the number of physical varibles multipled by the number of statistics, that is $12 \times 12 = 144$. In addition, to fit the different power level of the electric appliance, the statistics process are performed on 4 different levels of bus voltage, which are: lower than $110V$, higher than $110V$ but lower than $220V$, higher than $220V$ but lower than $1000V$, and higher than $1000V$. Then the dimension of statistic features is 576.

4.1.2. Local Features

Both the structure and control of the actual power system have the hierarchical and regional characteristics. From the perspective of the graphical computational model, the power grids has the characteristics of clique structure [31] but not the same, we refer to this property as cluster, and the clusters are relatively independent and the internal combination is relatively close. Actually the trend data are encoded with some regional information that the AC lines in the same region have a closer distance of electrical characteristics. Therefore, we model the topological features of the power system as an important feature. We take breadth-first search(BFS) from fault AC line as the root and then select K nodes to compose local topological map.

Besides the topological map, the electrical features are divided into two parts: the explicit features and the implicit features. The explicit features refers to the characteristics that reflects the power flow. This part of the features show the electrical status such as busbars and exchanges. Implicit features, such as power load and nearby environments, may implicitly affect the power system. We argues that since each electrical component has different hidden factors, it can be regarded as a time-invariant prior factor. This article innovatively uses the word embedding method [32] which is widely applied in the field natural language processing to represent the implicit features of the nodes(AC lines). Each component of the embeddings is constructed with a unique learnable vector representation (randomized initialization), and optimized through training phase of neural networks, and finally forming an appropriate repretation for implicit prior of the node.

In summary, in order to fully express the explicit and recessive features, the characteristics of the local topology are divided into two parts. We have selected 36-dimensional explicit features, that is, $D_e = 36$ and $D_s = 612$. [3]. The first part is the direct use of static physical quantities to characterize the explicit features, and the second part is the learning vectors for each node as implicit features.

[3] D_s are the dimension size of state features which are the combination of local and global features, thus,$D_s = 576 + 36 = 612$

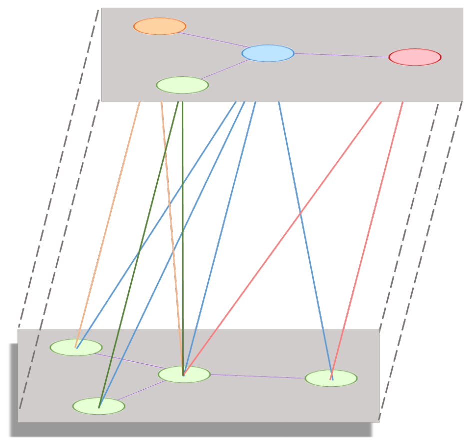


Figure 2: Diagram of the topological structures of of power grids

4.2. Topological Convolution(TConv)

The graph convolutional network was first proposed by [24] in 2005 to deal with non-European structure data. It has stronger expression ability for topological structure data than convolutional networks. After several years of development, several branches have been developed, such as the use of spectral analysis theory in graphical learning. The graph convolutional networks that express and characterize the power grid can accurately capture the characteristics of fault propagation flow in the power grid, while the methods without using graphical information can only catch the shallow features instead of the inner propagation process.

Formally, given a checkpoint of power system, $\boldsymbol{x}$ summarizes the local structures and the electrical stationary and time-dependent characteristics. Here, we omit the subscript i for the sake of brevity, and thus $\boldsymbol{x} = [f, \boldsymbol{E}, \boldsymbol{s}, \boldsymbol{A}]$. Following [10], the topological convolutional network is a convolutional network that acts on the data of the graph structure. The shared convolution kernel is applied to the entire graph to fit a specific objective function based on the data on the graph structure. To be specific, topological convolution fuse all the node information $\boldsymbol{E}^l$ in l-th layer with the weights indicated by adjacent matrix($\boldsymbol{A}$), a single layer of topological convolution operation is as follows:

$$\boldsymbol{H}^{l+1} = TConv(\boldsymbol{A}, \boldsymbol{E}^l) = \delta(\hat{\boldsymbol{A}}\boldsymbol{E}^l\boldsymbol{W}^{l+1}) \tag{2}$$

, where $TConv(\cdot)$ is the topological convolutional function,and $\boldsymbol{W}^l$ is a layer-specific trainable weight matrix; l is the layer index and $\boldsymbol{E}^0$ is the raw node features; $\delta(\cdot)$ denotes an activation function, such as the ReLU$(\cdot) = max(0, \cdot)$. Since multi-layer convolutions will multiply the matrix with A, it will affect the stability of the network by changing the size of the output eigenvectors. The adjacent matrix A is normalized such that the sum of each row is 1 to avoid the above problem. That is

$$\hat{\boldsymbol{A}} = \tilde{\boldsymbol{A}}^{-\frac{1}{2}}\boldsymbol{A}\tilde{\boldsymbol{A}}^{-\frac{1}{2}} \tag{3}$$

$$\tilde{\boldsymbol{A}}_{ii} = \sum_j \boldsymbol{A}_{ij} \tag{4}$$

For the gradient backpropagation in each convolutional layer, the gradients of the inputs is easily computed through matrix calculation. Let $\mathcal{J}$ be the objective of our model, then

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{E}^l} = \frac{\partial \mathcal{J}}{\partial \boldsymbol{H}^{l+1}}\frac{\partial \boldsymbol{H}^{l+1}}{\partial \boldsymbol{E}^l} \tag{5}$$

$$= \frac{\partial \mathcal{J}}{\partial \boldsymbol{H}^{l+1}}\frac{\partial \delta(\hat{\boldsymbol{A}}\boldsymbol{E}^l\boldsymbol{W}^{l+1})}{\partial \hat{\boldsymbol{A}}\boldsymbol{E}^l\boldsymbol{W}^{l+1}}\frac{\partial \hat{\boldsymbol{A}}\boldsymbol{E}^l\boldsymbol{W}^{l+1}}{\partial \boldsymbol{E}^l} \tag{6}$$

. Let $\boldsymbol{B}^{l+1} = \hat{\boldsymbol{A}}\boldsymbol{E}^l\boldsymbol{W}^{l+1}$ and the element-wise calculation is

$$\boldsymbol{B}^{l+1}_{km} = \sum_j \sum_i \boldsymbol{A}_{ki}\boldsymbol{E}^l_{ij}W^{l+1}_{jm} \tag{7}$$

, then the gradient for each elements of $\frac{\partial \boldsymbol{B}}{\partial \boldsymbol{E}}$ is

$$\frac{\partial \boldsymbol{B}^{l+1}_{km}}{\partial \boldsymbol{E}^l_{ij}} = \hat{\boldsymbol{A}}_{ki}\boldsymbol{W}^{l+1}_{jm} \tag{8}$$

4.3. Topological Pooling(TPooling)

From the intuitive sense, the degree of a node in graph suggests the importance of nodes. Inspired by Convolutional Neural Networks(CNN), convolution operation specialize in aggregating spatial information and pooling operation

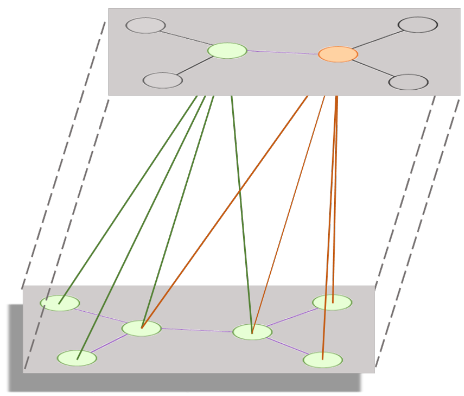


Figure 3: Diagram of the topological structures of of power grids

endows CNN with translation invariant and strong abstraction ability. Here, we present a novel topological pooling(TPooling) operation to summarize the local information and give a robust representation. To get a summarization from a predifined region and ingore the irrelevant details, TPooling operation calculates the max/mean value among the neighborhoods to represent the state of the region and delete the nodes whose degree(d) is smaller than T. Taking $T = 1$ for example, as shown in Figure 3, each green node in the bottom layer represents the features of an AC line and each node is connected to others to form a topological map. When applied topological pooling, the representation of the two important($d > T$) nodes is summarized by fetching the max/mean value at each dimension of features among their own neighbors, and the four unimportant nodes($d > T$, pallid nodes in upper layer in Figure 3) are omitted at the next layer.

Let $TPooling(\cdot)$ be the proposed topological pooling function, the TPooling operation is computed by

$$\boldsymbol{E}_i^l = TPooling(\boldsymbol{H}_i^l, \boldsymbol{A}) \tag{9}$$

$$= max(\mathcal{N}_i^{\,l}), i \in \{1, 2, \ldots, K\}; \quad \forall \boldsymbol{H}_j^l \in \mathcal{N}_i^{\,l}, \text{ if } \boldsymbol{A}_{ij} = 1 \tag{10}$$

, where i and j are the node indexes, $\mathcal{N}_i$ indicates the neignbor node set of node i, and function $max(\cdot)$ denotes elements maximizing operation on each dimension among a neignbor node set. We omit the show the details of gradient backpropagation since the calculation is enssetially the similar to the maxpooling operation in CNN.

4.4. Deep Topological Convolutional Neural Network

In this section, we present our proposed architecture. We stack the two operations several time and combine MLPs to build the deep topological convolutional network, refered to as TCNN, offering the ability to abstract the high level features in power grids. A schematic view of the resulting network is depicted in Figure 4.

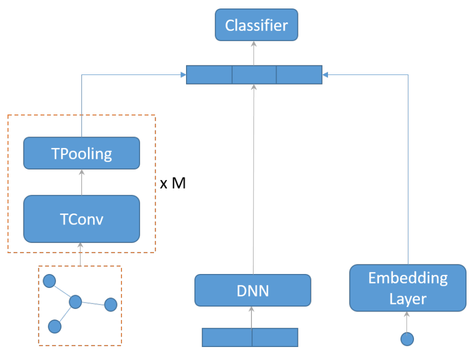


Figure 4: Diagram of the topological structures of of power grids

To be specific, several layers of $TConv$ and $TPooling$ are combined together to process topological data and finally provide a vector($\boldsymbol{E}_f^L$) to represent topological feature of the concerned AC line.

$$\boldsymbol{H}^{l+1} = TConv(\boldsymbol{A}, \boldsymbol{E}^l) \tag{11}$$

$$\boldsymbol{E}^{l+1} = TPooling(\boldsymbol{H}^l, \boldsymbol{A}) \tag{12}$$

. At the same time, the global and local features are processed by a two-layer multiple layer perceptron component and summarized into a vector($\boldsymbol{F}_{gl}$):

$$\boldsymbol{F}_{gl} = MLP_s(\boldsymbol{s}) \tag{13}$$

. In addition, implicit features($\boldsymbol{F}_e$) are obtained from embedding layer

$$\boldsymbol{F}_e = Embedding(f) \tag{14}$$

, where f is the index of the concerned AC line and the function $Embedding(\cdot)$ is a map from AC line index to a vector, which is learnable on each back propogation. Finally, we concatenate the three vectors representation as the final representation of the power grid, and use a softmax predictor to perform the binary classification.

$$\boldsymbol{p} = softmax(Fully([E_f^L, F_{gl}, F_e])) \tag{15}$$

, where function $Fully(\cdot)$ is the fully connection operation and the function $softmax(\cdot)$ is computed as

$$softmax(\boldsymbol{z})_j = \frac{e^{z_j}}{\sum_i e^{z_i}} \tag{16}$$

It should be mentioned that, there are several ways to combine the three vectors while we just take the simplest one and it works well.

Following the literature of supervised learning, the objective function is to minimize the cross-entropy of the samples, which can be formulized as

$$\mathcal{J} = \sum_i log(\boldsymbol{p}_{y^{(i)}}^{(i)}) \tag{17}$$

, where i is the index of sample in the dataset and $y^{(i)}$ denotes the final stability of i-th sample.

5. Experiments

We evaluated TCNN on dataset generated from industrial large-scale power grid. We show the performance of both ultimate Accuracy and other fair metrics; we also have deep analysis into our model.

5.1. Dataset

In this paper, the real-time online trend data from January 1st 2017 to January 3rd 2017 is adopted to construct the dataset, simulated from an industrial large-scale power grid. A sample per 15 minutes and the number of grid busses reached 10,0000. Considering the actual situation of power grids, only one fault occurs at one incident per 15 minutes. In total, 1440 tidal currents in one week were involved, and the total number of failures was 350,000. The fault set was determined by simulation. There were about 35,000 common system instabilities, accounting for 10% of the total. To test our model fairly, we use the data of a certain day as training set, and the half of the data of latter day as validation set, and the other half as test set.

Date	# Sample	# Unstable	# AC lines	# Bus
20170101	39812	98.05	88.43	34.10
20170102	42771	98.05	88.43	34.10
20170103	60256	98.09	88.81	13.17
20170104	xxx	98.05	87.36	11.27
20170105	61601	98.05	87.36	11.27
20170106	58642	98.05	87.36	11.27
20170107	xxx	98.05	87.36	11.27

Table 3: Stasitics of dataset.

5.2. Data preprocessing

Different input parameters have numerical differences in magnitude due to their different measure, which may increase the difficulty of learning. Therefore, it is also necessary to normalize the training set and the test set together to make the data with the statistical uniform distribution. The detailed calculation is as follows:

$$\hat{\boldsymbol{x}} = \frac{\boldsymbol{x} - min(\boldsymbol{X})}{max(\boldsymbol{X}) - min(\boldsymbol{X})} \tag{18}$$

where the function $min(\cdot)$ and $max(\cdot)$ are the minimum and maximum function along the samples.

5.3. Baselines

We compare TCNN with the following baselines:

- Rules. The state of power system is continuous and has local similarity in time. We use the results of last 15 minutes to predict the current state of the power system. We use rules to realize the above process and thus we name it as 'Rules'.
- SVM. We reimplemented the SVM approach proposed by Huang [6].
- MLP. We use three-layer multiple layer perceptron algorithm to process the local and global feature(including embeddings but no topological map).
- CNN. Three-layer traditional convolutional neural network, each layer consists of a convolution and a maxpooling operation. The inputs are also the local and global features, without topological map, the same as MLP.

5.4. Implementation Details

In our experiments, we applied coarse grid search to select hyperparameters and early stopping on the development sets. In designing TCNN and MLP, the hidden size is set to 128, $T = 1$. For CNN, the kernel size is set to 3,and pooling size is set to 2 at each layer, 200 feature maps each filter, dropout rate of 0.5;. The other parameters were initialized by randomly sampling from the uniform distribution in $[-0.01, 0.01]$. For all the models, we used AdaDelta with a learning rate of 0.1 and a batch size of 100. In our model and baselines, the rectified linear units (ReLU) is set as the activation function.

5.5. Metrics

To evaluate TCNN and baselines fairly, we consider the following measurements:

- Recall

$$Recall = \frac{N_{tp}}{N_{tp} + N_{fn}} \tag{19}$$

- Redundancy ratio

$$Rratio = \frac{N_{fp}}{N_{tp} + N_{fp}} \tag{20}$$

- Compression ratio

$$Cratio = \frac{N_{tn} + N_{fn}}{N_{tp} + N_{tn} + N_{fp} + N_{fn}} \tag{21}$$

- Accuracy

$$Acc = \frac{N_{tp} + N_{tn}}{N_{tp} + N_{tn} + N_{fp} + N_{fn}} \tag{22}$$

, where N_{tp},N_{tn},N_{fp} and N_{fn} are the true positive rate, true negative rate, false positive rate and false negative rate respectively. The recall is used to measure the coverage of unstable faults, which is the main indicator, refered to as reliability, and its value should be 1 as much as possible. Redundancy rate is used to measure the misjudgment of stable faults. Its value should be 0 as much as possible; Compression ratio is used to measure the compression ratio of the unstable samples. The larger the $Cratio$ is, the more calculations are reduced. for the original stable calculation, that is, the failure calculation ratio that the dynamic safety assessment can be performed less, is a minor indicator, and the larger the value, the better. Accuracy means the accurate rate between the predictions and the labels. The formmer three indicators are more important than accuracy since there is a more accurate simulating calculation to further comfirm the stablility in the scenario of power grids, but the four indicators have strong consistency to evaluate the methods. Thus, we report them all in the following experiments.

5.6. Performance

Table 4 shows the performance of TCNN and the three benchmark models for the four indicators on the test set. In order to make the models comparable, we fix the recall at 98% through adjusting the threshold of model, and then observe the pros and cons of the other three indicators. It can be seen that the

Date	**Model**	Recall	Redundancy rate	Compression ratio	Accuracy
20170102	Rules	98.09	87.00	39.01	44.91
	MLP	98.02	86.11	45.19	46.21
	SVM	98.05	88.43	34.10	41.57
	CNN	98.10	88.03	36.10	43.04
	TCNN	98.08	**83.98**	**52.05**	**59.91**
20170103	Rules	98.02	87.98	15.99	25.85
	MLP	98.00	88.54	15.34	24.85
	SVM	98.09	88.81	13.17	22.70
	CNN	98.05	88.71	15.21	24.77
	TCNN	98.02	**86.88**	**21.34**	**31.06**
20170104	Rules	98.13	84.32	13.54	25.98
	MLP	98.03	80.91	**45.32**	55.90
	SVM	98.05	87.36	11.27	22.27
	CNN	98.10	88.02	11.01	21.12
	TCNN	98.10	**79.73**	45.21	**56.08**
20170105	Rules	98.07	88.28	17.62	25.92
	MLP	98.00	87.97	18.59	27.42
	SVM	98.14	90.91	10.92	22.35
	CNN	98.09	89.33	10.92	22.57
	TCNN	98.01	**83.00**	**21.63**	**33.04**
20170106	Rules	98.19	65.17	72.13	81.65
	MLP	98.15	59.92	74.17	84.33
	SVM	98.03	83.42	37.65	47.78
	CNN	98.03	73.49	61.00	71.13
	TCNN	98.01	**54.24**	**78.11**	**89.12**
20170107	Rules	98.09	74.29	54.78	63.01
	MLP	98.06	45.91	80.33	88.39
	SVM	97.98	81.42	37.65	47.78
	CNN	98.12	78.23	44.32	52.45
	TCNN	98.06	**38.92**	**86.03**	**93.90**

Table 4: Performance.

performance is consistent in the three indicators and TCNN ourperforms other baselines by a large margin. Compared with Rules and SVM, neural networks performs better since they are deep models. Compared with both of which are also neural model, TCNN has the best performance in the three indicators, proving the proposed two operations have advantages to the extraction of the characteristics of power grids.

5.7. The effectiveness of input features

In order to examine the necessity of the several parts of features (global feature, local feature, topological feature) on the performance of TCNN, we retrain and test the model by reducing the corresponding part of the input features and Finally, the recall indicator is also fixed at 98%, and the other indicators on the test set is compared. As shown in Table 5, the second row indicates TCNN uses all the data and performs best, suggesting all the parts

Date	**Model**	Features	Recall	Redundancy rate	Compression ratio	Accuracy
20170106	TCNN	global+local+topo	98.01	**54.24**	**78.11**	**89.12**
	TCNN	local+global	98.11	65.01	71.52	81.53
	TCNN	local+topo	98.09	73.07	64.01	74.93
	TCNN	global+topo	98.03	84.33	28.39	39.31

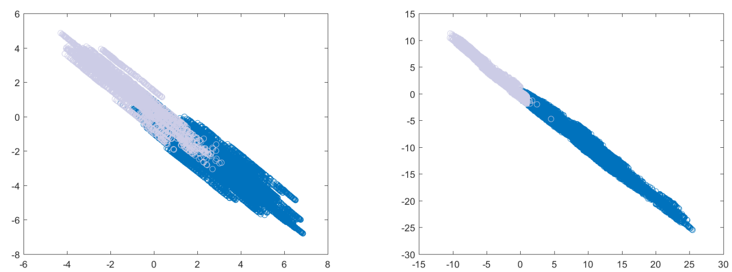
Table 5: Performance.

of features are indispensable. When the global features are missing(third row), the compression rate and accuracy rate both significantly decrease, and the redundancy rate rises significantly. This shows that although topological features are available, global features are crucial. When there is a lack of local features, there is also a large decrease in performance. When there is a lack of topological features, performance is also significantly reduced, but not as big as the formmer two. To sum up, in order to achieve the best performance, adding topological features can further improve the performance of the network, and at the same time, the topology features do not need to do additional processing. From above two experiments, we can conclude that topological information is useful and TCNN can capture it.

5.8. Feature Extraction

5.8.1. Feature visualization

To get an intuitive understanding of what extracted features of TCNN and MLP are like. We apply supervised learning for all samples of January 5th, visualize the samples of January 6th using the 2 dimensional features. The stable samples are dyed blue, and the unstable samples are dyed gray, as shown in Figure 5. The Figure 5(a) is the scatter diagram generated by MLP,and (b) is generated by TCNN. The unstable samples indicated by the features of TCNN are regularly arranged in the upper left corner, with just 2 or 3 outlines, while the two categories of samples are entangled in a large scale. Comparing TCNN and MLP, the features generated by TCNN are more effective, although the boundary is not a straight line, it is clear and easy enough to make classification.



(a) (b)

Figure 5: Feature visualization

5.8.2. Feature Utilization

In the formmer experiments, we have demonstrated TCNN can extract topological features and achieve high performance. It is intriguing to explore whether the extracted features are useful for other models. Therefore, we use the extracted features generated from TCNN as the input of above SVM algorithm. Table 6 shows the performance of the enhanced SVM, denoted as TCNN-SVM, and we also copy the performance of the SVM baseline in the table. We can find that, based on the extracted features from TCNN, there is a significant improve in the three measurements, and expecially, the performance at January 2nd(second row) is even better than TCNN. This experiments validates again, that extracted features are more senseful and general than shallow models.

Date	**Model**	Recall	Redundancy rate	Compression ratio	Accuracy
20170102	SVM	98.05	88.43	34.10	41.57
	TCNN-SVM	98.01	**83.01**	**55.52**	**63.13**
20170103	SVM	98.09	88.81	13.17	22.70
	TCNN-SVM	98.01	**87.95**	**28.49**	**37.79**
20170104	SVM	98.05	87.36	11.27	22.27
	TCNN-SVM	98.12	**82.53**	**33.88**	**39.11**
20170105	SVM	98.14	86.91	10.92	22.35
	TCNN-SVM	98.06	**85.43**	**14.95**	**25.93**
20170106	SVM	98.03	83.42	37.65	47.78
	TCNN-SVM	98.03	**82.44**	**40.27**	**50.33**
20170107	SVM	97.98	81.42	37.65	47.78
	TCNN-SVM	98.18	**60.43**	**65.62**	**80.47**

Table 6: The comparison between the SVM and TCNN-SVM, the latter using the features provided from TCNN .

5.8.3. Embedding Clustering

As explained in Section 4.1.2, embeddings are learned to encode implicit characteristics of AC lines. However,It is not like the embeddings of words, whose meaning is accessible. The AC lines of what counts may be ambiguous in power grids and therefore be challenging to evaluate how well the embeddings are learned. We just focus on the frequency of an AC line occurring fault since fault frequency is important information of each AC line and the ambiguity is minimal, to verify whether the embeddings of AC lines have encoded their fault frequency or not.

Firstly, TCNN is trained with the data of January 5-th and validated with the data of January 6-th to select the best model. Secondly, the 50 dimensional embedding features of the best model, indicating the implicit features of total 10425 AC lines respectively, are reduced to 2 dimensions by principal component analysis(PCA, the largest two prinsipal components are used). Thirdly, each of the AC lines are labelled with its fault frequency, which is calculated by dividing the number of failures by the total number of times. Finally we plot all the embeddings in Figure 6 and dyed by different colors to denote its fault frequency on January 6-th. The darker the dot is, the higher its fault frequency is, and the dots with small size mean they have not appeared in the data of January 6-th.

Fault frequency is easy to learned and can perform as a prior for STA. From the results, we find that the fault frequency of AC lines are in two extremes; some are very high and others are quite small. Intriguingly, the fault frequency increases from right to left with the horizontal axis, as indicated by the arrow. In other words, the largest component of the embeddings encodes the fault frequency of AC lines and other information encoded in embeddings is needed to be discovered in future work.

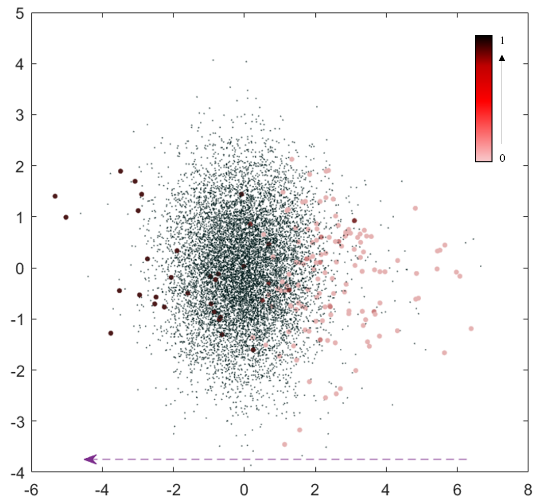


Figure 6: The scatter diagram of AC lines dyed by dyed by different colors to denote its fault frequency. The two axises are the two largest components reduced by PCA.

6. Conclusion and future work

In this paper, we have proposed a novel model, TCNN, that addresses STA problem using topological convolution, topological pooling operations and other deep learning techniques to effective extract features. Experiments show that TCNN achieves the higher performance than baselines by a large extent; that it can extract senseful and meanful features.

In future work, we would like to apply TCNN to handle various tasks in power grid, such as critical clearing time(CCT). Besides, we are also planning to explore the interpretability of the model to guide power engineers to know more about the power grids.

References

[1] J. J. Q. Yu, D. J. Hill, A. Y. S. Lam, J. Gu, V. O. K. Li, Intelligent time-adaptive transient stability assessment system, IEEE Transactions on Power Systems PP (99) (2017) 1–1.

[2] R. O. Burnett, M. M. Butts, P. S. Sterlina, Power system applications for phasor measurement units, IEEE Computer Applications in Power 7 (1) (1994) 8–13.

[3] D. R. Gurusinghe, A. D. Rajapakse, Post-disturbance transient stability status prediction using synchrophasor measurements, IEEE Transactions on Power Systems 31 (5) (2016) 3656–3664.

[4] T. Guo, J. V. Milanovic, Probabilistic framework for assessing the accuracy of data mining tool for online prediction of transient stability, IEEE Transactions on Power Systems 29 (1) (2014) 377–385.

[5] T. W. Wang, L. Guan, Y. Zhang, A survey on application of artificial intelligence technology in power system stability assessment, Power System Technology 33 (12) (2009) 60–65.

[6] Y. Huang, Z. Yu, D. Shi, X. Zhou, Strategy of huge electric power system stability quick judgment based on massive historical online data, Proceedings of the Csee.

[7] J. Yang, J. Zhang, F. Tang, J. He, Y. Wang, B. Wang, Discrete hidden markov model for transient stability assessment in power system 10 (2014) 10499–10510.

[8] Neural-network based transient security assessment for electrical power systems, Automation of Electric Power Systems 15(3) (1991) 44–50.

[9] A. Sharifian, S. Sharifian, A new power system transient stability assessment method based on type-2 fuzzy neural network estimation, International Journal of Electrical Power and Energy Systems 64 (2015) 71–87.

[10] H. Cui, Graphical convolutional networks, IEEE Transactions on Smart Grid 3 (1) (2012) 152–161.

[11] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436.

[12] H. Lee, R. Grosse, R. Ranganath, A. Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: Proceedings of the 26th annual international conference on machine learning, ACM, 2009, pp. 609–616.

[13] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[14] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, Journal of Machine Learning Research 11 (Dec) (2010) 3371–3408.

[15] H.-D. Chiang, Direct methods for stability analysis of electric power systems: Theoretical foundation, BCU methodologies, and applications.

[16] A.-A. Fouad, V. Vittal, Power system transient stability analysis using the transient energy function method, Pearson Education, 1991.

[17] I. A. Hiskens, D. J. Hill, Energy functions, transient stability and voltage behaviour in power systems with nonlinear loads, IEEE Transactions on Power Systems 4 (4) (1989) 1525–1533.

[18] P. Kundur, N. J. Balu, M. G. Lauby, Power system stability and control, Vol. 7, McGraw-hill New York, 1994.

[19] C.-W. Liu, J. Thorp, Application of synchronised phasor measurements to real-time transient stability prediction, IEE Proceedings-Generation, Transmission and Distribution 142 (4) (1995) 355–360.

[20] M. Pavella, D. Ernst, D. Ruiz-Vega, Transient stability of power systems: a unified approach to assessment and control, Springer Science and Business Media, 2012.

[21] G. Xue-ping, C. Shao-jie, Z. Wen-qin, A new framework for transient stability assessment based on neural networks, PROCEEDINGS-CHINESE SOCIETY OF ELECTRICAL ENGINEERING 20 (4) (2000) 77–82.

[22] I. Kamwa, S. Samantaray, G. Joós, On the accuracy versus transparency trade-off of data-mining models for fast-response pmu-based catastrophe predictors, IEEE Transactions on Smart Grid 3 (1) (2012) 152–161.

[23] H. U. Wei, L. Zheng, Y. Min, Y. Dong, Y. U. Rui, L. Wang, T. University, S. Branch, Research on power system transient stability assessment based on deep learning of big data technique, Power System Technology.

[24] M. Gori, G. Monfardini, F. Scarselli, A new model for learning in graph domains, in: Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on, Vol. 2, IEEE, 2005, pp. 729–734.

[25] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini, The graph neural network model, IEEE Transactions on Neural Networks 20 (1) (2009) 61–80.

[26] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, R. P. Adams, Convolutional networks on graphs for learning molecular fingerprints, in: Advances in neural information processing systems, 2015, pp. 2224–2232.

[27] J. Bruna, W. Zaremba, A. Szlam, Y. Lecun, Spectral networks and locally connected networks on graphs, Computer Science.

[28] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in: Advances in Neural Information Processing Systems, 2016, pp. 3844–3852.

[29] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907.

[30] M. Kezunovic, S. Meliopoulos, V. Venkatasubramanian, V. Vittal, Online Transient Stability Assessment, Springer International Publishing, Cham, 2014, pp. 99–142. doi:10.1007/978-3-319-06218-1_4.

[31] D. Koller, N. Friedman, Probabilistic graphical models: principles and techniques, MIT press, 2009.

[32] W. Y. Zou, R. Socher, D. Cer, C. D. Manning, Bilingual word embeddings for phrase-based machine translation, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013, pp. 1393–1398.