IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 17, 2018, accepted January 18, 2019, date of publication February 5, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896198

Ensemble Learning for Power Systems TTC Prediction With Wind Farms

GAO QIU[1], JUNYONG LIU[1], (Member, IEEE), YOUBO LIU[1], (Member, IEEE), TINGJIAN LIU[1], AND GANG MU[2], (Member, IEEE)

[1]College of Electrical Engineering and Information Technology, Sichuan University, Chengdu 610065, China
[2]School of Electrical Engineering, Northeast Electric Power University, Jilin 132012, China

Corresponding author: Gao Qiu (qiugaoscu@stu.scu.edu.cn)

This work was supported in part by the National Natural Science Fund, China, under Grant 51437003.

ABSTRACT Being aware of the reliable margin of vital tie-lines, acting on the connection of power exporting area and power importing area, is significant to power systems. However, the high penetration of wind power causes fast variation of boundary limit parameters such as the available amount of power that can be transferred on the tie-lines, namely, total transfer capability (TTC), which may result in the inaccurate security assessment. Unfortunately, the traditional optimal power flow-based TTC model has computation burden for online applications. To address this problem, computational efficiency is improved via a data-driven TTC predictor based on an ensemble learning architecture in this paper. In the first stage, a daily profiles-based method including probabilistic sampling is proposed to simulate plenty of operation scenarios as data samples for ensemble training. Then, a hybrid feature selection approach, which is composed of the maximal information coefficient and nonparametric independence screening, is applied to determine the most correlative features to the objective variable. To enable the TTC predictor with high accuracy and generalization ability, a novel ensemble learning scheme for TTC predictor is constituted through clustering few adaptive hierarchical GA-based neural networks (AHGA-NNs predictor). At last, a modified New England test system is used to validate the proposed methodology. The results illustrate that combining with the appropriate feature selection, the presented ensemble learning has high performance on creating the accurate TTC predictor, which enables online secure margin monitoring for the vital tie-lines.

INDEX TERMS Artificial neural networks, ensemble learning, feature selection, total transfer capability, wind power.

I. INTRODUCTION

In recent decades, the capacity of wind generation has increased rapidly throughout the world, and wind energy utilization has become highly significant in worldwide low-carbonization. However, large-scale wind power penetration has encountered quantities of obstacles (e.g., interconnected transmission lines transfer limits vary, bus voltage constraints violate, power system transient stability varies, etc.), which endanger the secure operation of power system. To suppress the risk of long-distance transmission caused by wind power injection, the current research mainly focuses on the combined development of wind power and the entire power system [1].

Due to the rapid expansion of power grids, cross-regional interconnected power systems are becoming more common. Risk assessments may be performed for various levels of firm contracts to determine the most acceptable level. TTC has been widely used to evaluate transfer capacity and is an essential index in assessing the operational security of cross-regional interconnected power system transmission lines. Conventional offline analysis methods generally treat TTC as a constant for ease of use, which is practically applicable for the normal power girds but might not be the case for wind power systems. The DCOPF model has been applied to overcome the computational burden [2], [3]. Notably, reactive power and voltage will significantly affect the transient trajectory of post-contingency, so DCOPF, only considering real power, is not suitable for the circumstances when transient stability constraints are involved into in TTC calculation. Unfortunately, in many real-world applications of power systems, transient stability will be validated. Likewise, TTC calculation taking transient stability constraints into account is a typical TSCOPF, in which only AC power flow equations are permitted to capture the variables fed into time domain simulations [4]–[7]. As a result, DCOPF is not the most

The associate editor coordinating the review of this manuscript and approving it for publication was Salvatore Favuzza.

suitable technique for computing TTC. The REI model was utilized to evaluate TTC by Min and Abur [8]. This method reduces the computation time, but only slightly, and the results are not as accurate as conventional methods.

It is known that explicit physical modeling for TTC calculation obtains the most accurate results. But it is a computing-intensive task which can hardly realize real-time application. However, for control centers, the high variability of operational conditions caused by wind generation necessitate the fast computation of security indicators, such as TTC, to help dispatchers be aware of the potential risk. Obviously, the conventional modeling approaches, such as optimal power flow (OPF), continuation power flow (CPF), and repeated power flow (RPF), cannot meet this requirement. Machine learning techniques offer promising alternatives for enabling robust and fast solutions through knowledge extraction of power system operation, which have a quite wide range of application experience in smart grid research [9]–[19]. Reference [9] introduced a general Bayesian framework for obtaining sparse solutions to classify predictions and the practical relevance vector machine (RVM) model, which was applied in power system transient stability assessment. An artificial neural network (ANN) was applied to assess transient stability by predicting the generator rotor angle [9]–[12]. Most research focuses on real-time transient stability awareness of the power system without integrated sustainable energy [13] and concentrates on the generator rotor angle performance [13]–[17]. However, few studies use machine learning techniques to predict TTC with integrated wind power. Reference [18] applied Monte Carlo simulations to expect short-term operating condition changes, feature selection, and linear least squares fitting of the TTC operating rules but did not consider the strong nonlinearity caused by sustainable energy (e.g., wind power and solar power). And, specifically, the linear least squares fitting method is difficult to capture the nonlinearity of the power system.

The variation in operating conditions and wind power can cause drastic changes in the TTC value [19] and may cause a decline in the generalization ability of a single learner. Consequently, an effective method is needed to enhance single-learner performance. Ensemble methods, which are a kind of state-of-the-art learning approach, are usually significantly more accurate and stable than a single learner and have already been widely applied in power systems. Zhang *et al.* [20] proposed a noise-assisted ensemble regression method to identify power system online sensitivity and showed numerically that the method could address the adverse effect of measurement noise and promote more stable identification results. Heinermann and Kramer [21] used machine learning ensembles to predict wind power. That study showed that, compared with the single learner, ensemble approach is more suitable for wind power systems. Additionally, a novel ensemble learning of a random-weight neural network–based intelligent system was introduced to assess voltage stability [22]. Previous researches have adequately demonstrated the feasibility of ensemble learning applied to fast and dynamic power system problems. Eventually, ensemble methods turn out to be an alternative way to solve TTC prediction of wind power integrated system problems.

This paper presents an adaptive hierarchical GA-based NN ensemble framework to apply to predict TTC of wind power systems in real time, with a modified IEEE 39 bus system with integrated wind power. The first part of this paper covers offline TTC predictor training. A novel daily profiles-based method combined repeated dichotomy power flow (DPM-RDPF) method is used to generate train data. A hybrid feature selection method is applied to screen redundant features, and experimental proof of the validity of the method is presented. The paper also presents a comparison among conventional BPNN, extreme learning machines (ELMs), and the proposed NN models. The second part of this paper deals with real-time TTC prediction. The data collected by SCADA is used as input to TTC predictors to realize TTC evaluation in real time. The paper demonstrates the feasibility of the proposed framework. Historical generation and wind data were obtained from practical data from a sample district.

The main contributions of this paper are as follows:

1) A data samples generation method is proposed to consider wind and load uncertainties;

2) A hybrid feature selection method is presented to measure the correlations between variables, and the correlations between variables and the target;

3) A modified BPNN is advanced for global tuning of NN's structure and weights;

4) Various ensembles are applied for robust TTC modeling. Moreover, comparison analysis is conducted to reveal the advantages of the proposed NN and the most applicable ensemble.

II. FUNDAMENTALS OF TTC CALCULATION

Total transfer capability reflects the available amount of power that can be transferred between two interconnected areas without violation of security criterion. To secure operating margin, the power flow transferred through the tie-lines should be confirmed below the value of TTC.

A generic optimal approach for calculating TTC is described as follows [23].

maximize

$$f(P_{Gi,t(i\in\mathbb{E})}, P_{Lj,t(j\in\mathbb{R})}, Q_{Lj,t(j\in\mathbb{R})}) = \sum_{i\in\mathbb{E}} P_{Gi,t} \tag{1}$$

subject to

$$P_{Gi,t} - P_{Li,t} - V_{i,t}\sum_{j=1}^{n} V_{i,j}(G_{ij}\cos\theta_{ij} + B_{ij}\sin\theta_{ij}) = 0$$

$$Q_{Gi,t} - Q_{Li,t} - V_{i,t}\sum_{j=1}^{n} V_{i,j}(G_{ij}\sin\theta_{ij} - B_{ij}\cos\theta_{ij}) = 0 \tag{2}$$

$$P_{Gi,t}^{\min} \le P_{Gi,t} \le P_{Gi,t}^{\max}, \quad Q_{Gi,t}^{\min} \le Q_{Gi,t} \le Q_{Gi,t}^{\max}, \ \forall i \in \mathbb{G} \tag{3}$$

$$P_{Li,t}^{\min} \le P_{Li,t} \le P_{Li,t}^{\max}, \quad V_{i,t}^{\min} \le V_{i,t} \le V_{i,t}^{\max}, \ \forall i \in \mathbb{B} \tag{4}$$

$$S_{ij,t} \le S_{ij,t}^{\max}, \quad \forall ij \in \mathbb{L} \tag{5}$$

where the subscript index t refers to the variable at time t; n is the number of bus; G_{ij}, B_{ij} and θ_{ij} are conductance, susceptance and phase between ith bus and jth bus, respectively; $\mathbb{B}$ is the bus set; R is the receiving area bus set; $\mathbb{E}$ is the exporting area bus set; $\mathbb{G}$ is the generator set; $\mathbb{L}$ is the lines set; (2) denotes the power flow equations; P_{Lj}, Q_{Lj} are the active and reactive load power at bus j; P_{Gi}, Q_{Gi} are active and reactive power generation at bus i; $P_{Gi}^{\max}$, $P_{Gi}^{\min}$ are the maximum and minimum active power generation at bus i; $P_{Lj}^{\max}$, $P_{Lj}^{\min}$ are the maximum and minimum active load at bus j; $Q_{Gi}^{\max}$, $Q_{Gi}^{\min}$ are maximum and minimum reactive power generation at bus i; S_{ij}, $S_{ij}^{\max}$ are the apparent power and maximum apparent power at the line ij; and $V_i^{\max}$, $V_i^{\min}$ are the maximum and minimum voltage magnitude at bus i.

In addition to power flow constraints, transient stability constraints limited by a set of contingencies are included in TTC computation model since the dynamic security of the N-1 criterion needs to be considered. In this paper, TSI indicator [24], which is defined as (6), is used to measure the stability margin of the power system.

$$TSI = \frac{\delta - \max(|\Delta\delta_{ij}|)}{\delta + \max(|\Delta\delta_{ij}|)}, \quad i, j \in \mathbb{G}, i \neq j \tag{6}$$

where $|\Delta\delta_{ij}|$ denotes the absolute value of the rotor angle separation between ith generator and jth generator during the post-fault period; δ is a pre-defined threshold, in this paper, δ is set to 180°.

III. DATA SAMPLES GENERATION

An appropriate sample data set is usually required as it is the necessity of ensuring the generalization ability of the learned knowledge. For power systems, the day-head scheduling is performed to determine generator output plans including unit commitments and hourly curves. However, the fluctuation of wind power may cause the undesired deviation of scheduled generation plans. In light of this, we propose a DPM with a pre-defined baseline scenario to produce sample data to build a TTC predictor for the next day.

A. WIND POWER OUTPUT

Wind power output varies rapidly with the intermittent variation of wind speed. The following piecewise function is widely used to model wind generation based on wind speed and wind turbine parameters [25]. Let V_{cin}, V_r, V_{co} denote cut-in, rated, and cut-out wind speed, respectively; wind turbine generation P_{WT} can be determined using:

$$P_{WT} = \begin{cases} 0 & w < V_{cin} \text{ or } w \geq V_{co} \\ P_r(A + Bw + Cw^2) & V_{cin} \leq w < V_r \\ P_r & V_r \leq w < V_{co} \end{cases} \tag{7}$$

where w is the wind speed, and P_r is the rated capacity of the wind turbine; A, B and C are calculated based on the wind turbine's parameters. The wind turbine power factor can be set as a constant. As long as the continuous wind speed is determined, the time-series wind farm generation can be calculated merely by applying (7).

B. DAILY PROFILES-BASED METHOD

Since wind power and load lead to balance uncertainties, the output of each generator may vary in a certain region by either automatic configuration or dispatch control according to its operating role in a power system. Inspired by [26], we adopt probabilistic distribution function (PDF) to describe these output regions that traditional generators probably achieve. This method is also extended to model the uncertainties of wind power and load. In doing so, the completeness of operation scenarios can be significantly guaranteed for a given day. Fig. 1 illustrates the probabilistic generation curve, the output deviation of which is assumed to follow the normal distribution.

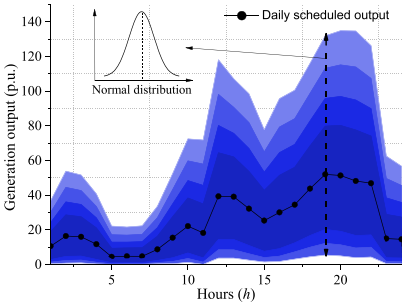


FIGURE 1. Illustration of the probabilistic generation curve.

For each interval $t \in (1, 24)$, the following steps are employed to simulate operation scenarios.

Step 1). Construct an output vector by sampling each generator according to the scheduled generation output $P_g(t)$ and probabilistic distribution law (8)

$$f\left(P_g(t)\right) = \frac{1}{\sqrt{2\pi}\sigma_g} e^{-\left(P_g(t)-\mu_g\right)^2 / 2\sigma_g^2} \quad \forall : g = 1 \ldots G \tag{8}$$

where μ_g and σ_g represent the mean and standard deviation of the output scheduling. A samples data matrix of generators output can be expressed as:

$$\mathbf{Gen}(t) = [\boldsymbol{P}_1(t)\boldsymbol{P}_2(t) \ldots . \boldsymbol{P}_G(t)]_{M \times G} \quad \forall t \tag{9}$$

where M is the number of samples for time t and G is the total number of generators.

Step 2). Similarly, create M-member data vectors of load and wind generation, $\boldsymbol{L}(t)$ and $\boldsymbol{W}(t)$, based on day-ahead forecasting. The PDF parameters, μ_L, σ_L, μ_W, and σ_W, which replace μ_g and σ_g in (8), represent the mean and standard deviation of the load and wind power.

Step 3). Take the m-th element respectively from $\mathbf{Gen}(t)$, $\boldsymbol{L}(t)$, and $\boldsymbol{W}(t)$ to build an operation scenario which is then used to calculate TTC under the given conditions. Afterward, combine TTC value and the measurable status vector $\boldsymbol{F}$ to be one data sample. Repeat the procedure until $m = \mathrm{M}$.

Step 4). Move t to the next interval and launch the same routine from step 1) to 3), until the full-time horizon

(e.g., the 24 hours of the next day) is handled. In this case, the scale of data samples is 24×M.

Compared with the clustering-based method [27], [28], the proposed DPM has several advantages. First, based on the day-ahead scheduling information, the time-series characteristics of the traditional generators can be considered, resulting in the more reliable cases. Second, DPM is readily adaptive to the significant changes of operating conditions such as topology switch and generator shut-down which are usually considered in day-ahead scheduling. Third, the DPM is easy to apply and highly scalable regarding setting the time interval to obtain samples with different time granularity.

C. REPEATED DICHOTOMY POWER FLOW

To achieve stable convergence, TTC calculation for each sample is adopted by the repeated dichotomy power flow (RDPF). The RDPF outperforms OPF and CPF based approach in terms of readily bringing transient and voltage stability constraints into TTC computing model based on dichotomous validation scheme. The algorithm procedure could be detailed as follows:

Input: iterations i, maximum iterations i_max, initial load growth λ, load backward and forward search value λ_b, λ_f

Procedure:

1. initial current power system operation condition
2. **while** $|\Delta\lambda| \geq 0.08$ && $i \leq i$_max **do** {
3. receive area load grow λ;
4. the generators in sink area share load growth;
5. execute power flow program to get corridors' power flow;
6. perform time domain simulation of tie-lines;
7. **if** the static security constraint is broken || transient instability (N-1 criterion check)
8. $\Delta\lambda = \lambda - \lambda_b$, $\lambda_f = \lambda$, $\lambda = \lambda_{\mathrm{b}} + \Delta\lambda/2$;
9. **else**
10. $\Delta\lambda = \lambda_f - \lambda$, $\lambda_b = \lambda$, $\lambda = \lambda_{\mathrm{b}} + \Delta\lambda/2$;
11. **end**
12. $i = i + 1$; }

Output: TTC under the current operation condition

D. VARIABLES FORMULATION

For the scenario S existing in the entire set of simulated samples $\boldsymbol{\psi}$, the vector of state variables can be expressed as:

$$\boldsymbol{F}_S = \begin{bmatrix} \mathbf{P}_G^S & \mathbf{Q}_G^S & \mathbf{U}_G^S \mathbf{P}_L^S & \mathbf{Q}_L^S & \mathbf{V}_B^S & \mathbf{T}_{LI}^S & \mathbf{P}_{WF}^S \end{bmatrix} \quad (10)$$

where $\mathbf{P}_G^S, \mathbf{Q}_G^S, \mathbf{U}_G^S, \mathbf{P}_L^S, \mathbf{Q}_L^S, \mathbf{V}_B^S, \mathbf{T}_{LI}^S$, and $\mathbf{P}_{WF}^S$ denote, in sequence, the vector of active generator output, reactive generator output, generator voltage, active load, reactive load, bus voltage, transmission line power, and wind power.

Given the fact that the input variables are normalized, the goal of the studied predictor is to fit the numeric deviation of the TTC for a power exporting path with the deviation of measurable state parameters defined by (10), therefore (11) is applied.

$$\Delta\psi = \begin{bmatrix} \Delta F_1^1 & \cdots & \Delta F_1^j & \cdots & \Delta F_1^k \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \Delta F_i^1 & \cdots & \Delta F_i^j & \cdots & \Delta F_i^k \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \Delta F_S^1 & \cdots & \Delta F_S^j & \cdots & \Delta F_S^k \end{bmatrix} \quad (11)$$

where ΔF_i^j represents the j-th variable under the operation scenario i, and the last column is the target variable ΔTTC. The control variable in set $\Delta\psi$ is given by (12):

$$\Delta F_i^j = F_i^j - F_0^j \quad (12)$$

where F_0^j refers to the initial operation scenario.

IV. FEATURE SELECTION

A. MAXIMAL INFORMATION COEFFICIENT

MIC, which belongs to the maximal information-based nonparametric exploration class of statistics, can measure the strength of the numerical association between two features [29]. In this paper, MIC is used to reduce the feature redundancy according to the mutual correlation strength of power systems operation variables before executing the NIS algorithm.

Considering two features, $\boldsymbol{x}$ and $\boldsymbol{y}$, extracted from (11), with the same sample size, the discretization mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$ is calculated as (13):

$$I[\boldsymbol{x};\boldsymbol{y}] = \sum_{X,Y} p(\boldsymbol{x},\boldsymbol{y}) \log_2 \frac{p(\boldsymbol{x},\boldsymbol{y})}{p(\boldsymbol{x})p(\boldsymbol{y})} \quad \boldsymbol{x},\boldsymbol{y} \in \boldsymbol{F}_s \quad (13)$$

where X and Y are the axes scale of a two-dimensional coordinate gridding on the scatterplots of $[\boldsymbol{x};\boldsymbol{y}]$; $p(\boldsymbol{x},\boldsymbol{y})$ is the joint distribution probability of attributes pair $[\boldsymbol{x};\boldsymbol{y}]$; $p(\boldsymbol{x})$ and $p(\boldsymbol{y})$ denote the marginal distribution probability of two attributes, respectively; and $\boldsymbol{F_S}$ is the electric state space.

The MIC is determined using (14):

$$MIC[\boldsymbol{x};\boldsymbol{y}] = \max_{X,Y} \frac{I[\boldsymbol{x};\boldsymbol{y}]}{\log_2 \{\min(|X|, |Y|)\}} \quad s.t.\ |X| \cdot |Y| < B \quad (14)$$

where $|X| \cdot |Y|$ indicates the number of grids. The parameter B restricts the maximal allowable maximal number of grids, and B usually equals to $M^{0.6}$. The MIC method locates in the range of [0, 1]. The closer the value is to 1, the stronger the association between the two given variables, and vice versa.

Once a highly associated relationship is identified for two variables with a MIC value higher than a given threshold (e.g., 0.8), one of the variables can be eliminated. The information entropy criterion is used to determine the eliminated variable:

$$\mathrm{Entropy}(\boldsymbol{x}) = -\sum_{h=1}^{H} \mathrm{Pro}(\boldsymbol{x})_h \log \mathrm{Pro}(\boldsymbol{x})_h \quad \forall \boldsymbol{x} \in \boldsymbol{F_S} \quad (15)$$

where H is the number of segmentations that uniformly divides the sample set of attribute $\boldsymbol{x}$, while $\text{Pro}(\boldsymbol{x})_h$ denotes the probability of x occurring in the segment h. The feature $\boldsymbol{x}$ with higher entropy tends to be maintained.

B. NONPARAMETRIC INDEPENDENCE SCREENING

The MIC screening is able to reflect the correlation between the features but cannot quantify the correlation of the corresponding features to the objective variable, i.e., TTC in this paper. The NIS fundamental [30] is that the unitary regression is performed to figure out the numeric relation between any individual feature and the objective variable. The strength of this relation could be assessed by the residual sum of squares (RSS). The lower RSS means the closer correlation.

Equation (11) can be written as:

$$\Delta\psi = [\boldsymbol{F}^1, \boldsymbol{F}^2, \ldots, \boldsymbol{F}^k] \tag{16}$$

To simplify the description, (16) is used to illustrate the algorithm flow. The concrete steps are as follows:

Step 1). Assume that there is a numerical relationship $model_i$ between features $\boldsymbol{F}^i$ $(i = 1, 2, \ldots, \text{k} - 1)$ and the target $\boldsymbol{F}^k$:

$$\boldsymbol{F}^k = model_i(\boldsymbol{F}^i) + \boldsymbol{\varepsilon} \tag{17}$$

where ε is the correlation error that obeys a zero-mean normal distribution, which is generally 0.

Step 2). Nonparametrically regress each attribute of the sample set on the target attribute; that is, perform univariate nonparametric regression on features $\boldsymbol{F}^1, \boldsymbol{F}^2, \ldots, \boldsymbol{F}^i$ respectively against target $\boldsymbol{F}^k$, using a regression equation to calculate the predicted value of the target feature:

$$\begin{aligned} \boldsymbol{Y}_i^e &= [y_{0i}^e, y_{1i}^e, \ldots, y_{mi}^e] = \varphi_i(\boldsymbol{F}^i) \\ \boldsymbol{Y} &= \boldsymbol{F}^k, \quad i = 1, 2, \ldots, k-1 \end{aligned} \tag{18}$$

where $\varphi_i(\boldsymbol{F}^i)$ is the regression equation of $\boldsymbol{F}^i$ against $\boldsymbol{F}^k$, and $\boldsymbol{Y}$ corresponds to the target $\boldsymbol{F}^k$ in (16). Taking into account the non-linear properties of TTC, it is challenging to implement linear regression; hence, this paper uses the cubic B-spline function to perform regression to reflect the TTC nonlinearity.

Step 3). Calculate the prediction error RSS of all features according to the objective feature:

$$RSS_i = \sum_{j=1}^{m} (F_j^k - y_{ji}^e)^2 \tag{19}$$

Step 4). Rank the RSS set of samples; the lower the RSS, the higher the contribution level, otherwise, the lower the contribution level. We screen out features in the SCR set:

$$\text{SCR} = \{i | \text{RSS}_i \geq \text{Ave(RSS)}\} \tag{20}$$

V. ENSEMBLE LEARNING FOR TTC PREDICTOR

It has been proved that a simple learning algorithm can hardly fit the TTC rules due to the highly time-varying nonlinearity of the wind power system. In this section, to enable robust and accurate TTC prediction, a novel ensemble learning technique is applied to build TTC predictor.

A. BASE LEARNER

The BPNN is a fundamental base learner for ensemble learning. Some parameters must be determined to build the BPNN: i) the number of hidden nodes $\boldsymbol{n}$; ii) input hidden layer weight vector $\boldsymbol{W}_{\text{hidden}}$; iii) input hidden layer threshold $\boldsymbol{b}_{\text{hidden}}$; iv) hidden output layer weight vector $\boldsymbol{W}_{\text{output}}$; v) hidden output layer threshold $\boldsymbol{b}_{\text{output}}$. In order to overcome the drawbacks of traditional BP optimization, such as overfitting, weak generalization ability, and long training time, an adaptive hierarchical GA (AHGA) is proposed:

1) Chromosome encoding: A chromosome is expressed via a two-order hierarchical structure. The first order is controlling genes activating hidden neural nodes, determining $\boldsymbol{W}_{\text{hidden}}$, $\boldsymbol{W}_{\text{output}}$, $\boldsymbol{b}_{\text{hidden}}$ and $\boldsymbol{b}_{\text{output}}$ of the second order parameter genes. Controlling genes employ binary encoding while parameter genes use real-number encoding.

2) Fitness function: The optimization objective is selected as the mean square error between the NN output and the expected TTC value.

3) Genetic operator: Individual selection uses random traversal sampling. Adaptive one-point crossover and mutation are used for controlling genes, enabling convergence improvement. Crossover and mutation probability are calculated by (21) and (22) respectively.

$$P_c = \begin{cases} P_{c1} - \dfrac{(P_{c1} - P_{c2})(f_{\text{avg}} - f')}{f_{\text{avg}} - f_{\min}} & f' \leq f_{\text{avg}} \\ P_{c1} & f' > f_{\text{avg}} \end{cases} \tag{21}$$

where P_c is one-point crossover probability; $f_{\min}$ is the optimal fitness in populations; f' represents the superior fitness of two individuals in the crossover; f_{avg} is the average fitness of the populations, and $P_{c1} = 0.85$, $P_{c2} = 0.5$.

$$P_m = \begin{cases} P_{m1} - \dfrac{(P_{m1} - P_{m2})(f_{\text{avg}} - f')}{f_{\text{avg}} - f_{\min}} & f' \leq f_{\text{avg}} \\ P_{m1} & f' > f_{\text{avg}} \end{cases} \tag{22}$$

where P_m is the mutation probability; $P_{m1} = 0.1$, $P_{m2} = 0.001$.

Linear recombination and uniform mutation are applied for parameter genes, the details of which are illustrated as follows:

Input: training set, parameters of AHGA-BPNN

Procedure:

1. randomly generate an initial population
2. evolve the population where the fitness is measured as the mean square error
3. NN_{nopt} is the evolved near the best BPNN
4. train NN_{nopt} by Bayesian regularization to get the best BPNN NN_{opt}

Output: NN_{opt}

B. ENSEMBLE

It is noting that the abnormal sample data might lead to the insufficient robustness and generalization of base leaner due to the lack of sampling diversity for provided samples. Hence, an AHGA-BPNN-based ensemble learning architecture for TTC predictor training is presented. This paper investigates a number of NN ensemble approaches including adaboosting1, adaboosting2, and DESEN (differential evolution algorithm based selective ensemble [31]) to perform comparative analysis.

1) ADABOOSTING1

The base NN leaners can be additively weighted as:

$$NN(\boldsymbol{x}) = \sum_{t=1}^{T} \alpha_t nn_t(\boldsymbol{x}_t), \quad \forall \boldsymbol{x} \in \boldsymbol{F_S} \tag{23}$$

where $NN(\boldsymbol{x})$ is the ensemble output vector; nn_t is the tth base learner; $\boldsymbol{x}_t$ are the nn_t inputs sampled from the TTC sample data (11) under the sample distribution D_t; α_t is the weight of the tth NN; and T is the total number of base NNs.

Conventional adaboosting algorithms iteratively generate nn_t, α_t, and D_t. The exponential loss function of base learner nn_t can be obtained when nn_t is trained under D_t:

$$\begin{aligned} loss_{\exp}(\alpha_t nn_t | D_t) &= \mathrm{E}_{\boldsymbol{x}\sim D_t}[e^{-y\alpha_t nn_t(\boldsymbol{x}_t)}] \\ &= e^{-\alpha_t}(1-\varepsilon_t) + e^{\alpha_t}\varepsilon_t \end{aligned} \tag{24}$$

$$\begin{aligned} \varepsilon_t &= P_{\boldsymbol{x}\sim D_t}(\boldsymbol{error}_t > \lambda) \quad \forall \\ \boldsymbol{error}_t &= nn_t(\boldsymbol{x}_t) - \boldsymbol{y}, \quad \boldsymbol{y} = \boldsymbol{F}^k \end{aligned} \tag{25}$$

where $\mathrm{E}_{\cdot\sim D}[f(\cdot)]$ represents the mathematical expectation of function $f(\cdot)$ under D_t. λ is an artificially configurable parameter strongly relating to T. λ tends to be set as a lower value to achieve higher prediction accuracy. In this paper, λ is the average value of $\boldsymbol{error_t}$, and T equals 50.

The optimal weight should minimize the exponential loss function, which is expressed as:

$$\alpha_{t(opt)} = \arg\min_{\alpha_t}(loss_{\exp}(\alpha_t nn_t | D_t)) = \frac{1}{2}\ln(\frac{1-\varepsilon_t}{\varepsilon_t}) \tag{26}$$

The sample distribution D_t is updated according to:

$$D_{t+1}(\boldsymbol{x}) = D_t(\boldsymbol{x})\exp(-\alpha_{t(opt)} nn_t(\boldsymbol{x}_t)\boldsymbol{y}) \big/ Z_t \tag{27}$$

where Z_t is a normalization factor.

2) ADABOOSTING2

Adaboosting2 algorithm applies the update strategy for samples distribution, which is able to minimize the weighted prediction error. For the t-th base learner, NN's weighted error under D_t is:

$$\boldsymbol{werror}_t = (nn_t(\boldsymbol{x}_t) - \boldsymbol{y}) \cdot \boldsymbol{w}_t \tag{28}$$

where $\boldsymbol{w}_t$ is the vector of weight for each sample.

To determine D_{t+1}, α_t and $\boldsymbol{w}_t$ are updated via:

$$\alpha_t = 1 \Big/ 2\exp(\sum_m |w_m|), \quad m \in \{m | \boldsymbol{werror}_t > \lambda\} \tag{29}$$

$$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t + (alph - 1)\boldsymbol{w}_t \mathrm{sgn}(\boldsymbol{werror}_t - \lambda) \tag{30}$$

where $alph$ is a constant factor controlling D_t update strength which is set to 1.1 in this paper.

3) DESEN

There exists another constraint of the weight of each NN in DESEN to ease differential evolution searching:

$$\sum_t^T \alpha_t = 1 \tag{31}$$

Define the correlation between the ith and kth base NN as:

$$C_{ik} = \|nn_i(\boldsymbol{x}_i) - \boldsymbol{y}\|_2 \, \|nn_k(\boldsymbol{x}_k) - \boldsymbol{y}\|_2 \tag{32}$$

The optimal weight of each NN can be obtained by (33) (The derivation of optimal $\boldsymbol{\alpha}_{t(opt)}$ calculation was detailed in [27]).

$$\boldsymbol{\alpha}_{t(opt)} = \sum_{k=1}^{T} C_{tk}^{-1} \Big/ \sum_{i=1}^{T}\sum_{k=1}^{T} C_{ik}^{-1} \tag{33}$$

In real-world applications, since some neural networks in the ensemble perform quite similar, which makes the correlation matrix $(C_{ik})_{T\times T}$ singular or ill-conditioned, (33) is challenging to be solved. Note that (32) could be defined as an optimization problem. Take into account the advantages of the DE algorithm, so the DE algorithm is applied to solve the problem.

To efficiently evolve a population in the DE algorithm, validation set $\mathcal{V}$ is given so that fitness function can be set corresponding to α:

$$fit(\boldsymbol{\alpha}) = 1 \Big/ error_T^{\mathcal{V}} \tag{34}$$

where $\mathrm{error}_T^{\mathcal{V}}$ denotes the output generalization error of the ensemble on the validation set $\mathcal{V}$. The DESEN approach is summarized as follows:

Input: training set $(\boldsymbol{x}, \boldsymbol{y}) \in \Delta\psi$, base learner, trials T, threshold α_λ

Procedure:

1. for $t = 1$ to T {
2. $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ = bootstrap sample from $(\boldsymbol{x}, \boldsymbol{y})$;
3. $nnoutput_t = nn_t(\boldsymbol{x}_t)$; }
4. generate a population of $\boldsymbol{\alpha}$;
5. evolve the population where the fitness is measured as $fit(\alpha)$;
6. $\boldsymbol{\alpha}_{(opt)}$ is the evolved best weight vector;

Output: $NN(\boldsymbol{x}) = Ave \sum_{\boldsymbol{\alpha}_{(opt)} > \alpha_\lambda} nnoutput_t$

C. ENSEMBLE LEARNING FRAMEWORK FOR TTC PREDICTION

TTC predictors are offline trained based on day-ahead scheduling information. In practical application, a vector of the real-time measured data of the selected features acquired by SCADA is fed into ensemble predictor to realize the fast TTC awareness. The framework is shown in Fig. 2.

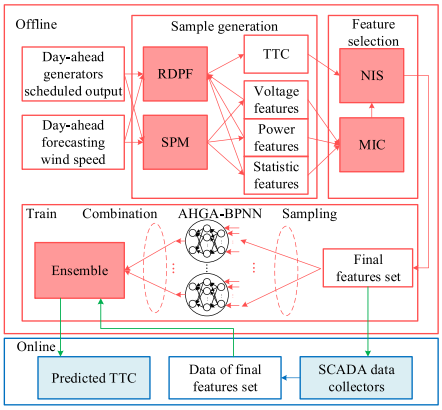


FIGURE 2. TTC prediction framework.

VI. NUMERICAL STUDY

The modified IEEE 39 system shown in Fig. 3. is used as the test system. The wind farms are integrated at bus 17 and 21 respectively. The rated capacity of each wind farm is 600 MW, and the wind turbine parameters are shown in Table 2. The wind power exporting area includes zones II and III, while the receiving area is zone I. The wind power exporting corridor consists of 4 tie-lines, line 1-39, 2-3, 18-3 and 16-15. Considering the N-1 criterion, the transient contingency fault is set as a three-phase short circuit, and the initial fault-clearing time is 0.1 second. The day-ahead scheduled generation output, daily wind farms output, and load profiles are shown in Fig. 4 respectively.

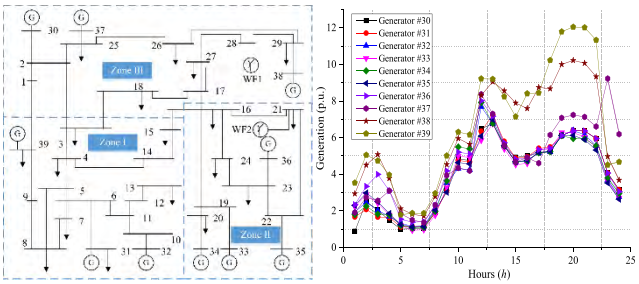


FIGURE 3. Modified IEEE 39 test system and daily scheduled generation output curve.

TABLE 1. Wind turbine parameters.

Wind farm bus	V_{in}	V_0	V_{out}	P_{WN}	Power factor
17	1m/s	10m/s	26m/s	2MW	0.95
21	1m/s	15m/s	22m/s	2MW	0.95

A. SAMPLE GENERATION

The TTC samples are produced by DPM and RDPF. The initial features extracted from a set of measurable parameters are detailed in Table 2.

A total of 1200 samples are generated at each interval. To measure the generalization ability of the trained

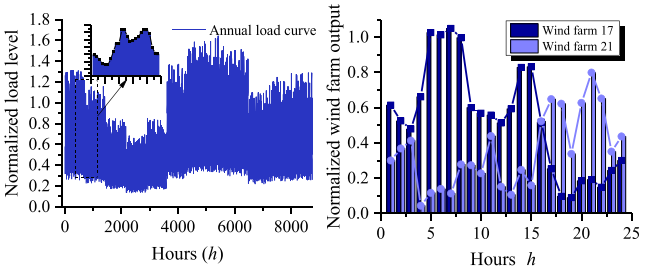


FIGURE 4. Time-series profiles of system operation status.

TABLE 2. Organization structure of control sample sets.

objective variable	
TTC	ΔTTC
Initial independent control variables	
Voltage difference of tie-lines	$\Delta V_i^{\text{tie-lines}}$
Exporting area overall active power	$\Delta P_{\Sigma}^{\text{Export}}$
Exporting area overall reactive power	$\Delta Q_{\Sigma}^{\text{Export}}$
Receiving area overall active power	$\Delta P_{\Sigma}^{\text{Receive}}$
Receiving area overall reactive power	$\Delta Q_{\Sigma}^{\text{Receive}}$
Bus voltage	ΔU_n
Active load at bus n	ΔP_n^{Load}
Reactive load at bus n	ΔQ_n^{Load}
Generator's active power	ΔP_G^{Gen}
Generator's reactive power	ΔQ_G^{Gen}
Overall power flow of tie-lines	$\Delta P_{\Sigma}^{\text{tie-lines}}$
Wind farm's active power	ΔP^{wind}
Wind farm's reactive power	ΔQ^{wind}
Overall active load	$\Delta P_{\Sigma}^{\text{Load}}$

TTC predictor, 1000 training samples and 200 test samples are produced by 10-fold cross-validation. In the training samples, 200 validation samples are generated by 10-fold cross-validation to prevent learner overfitting.

B. FEATURE SELECTION RESULTS

1) FEATURE SELECTION RESULTS VISUALIZATION

Fig. 5 (a) plots the initial MIC matrix. Note the disappearance of some dark color blocks in Fig. 5 (b), it indicates that MIC can screen out the features having the relatively high correlation.

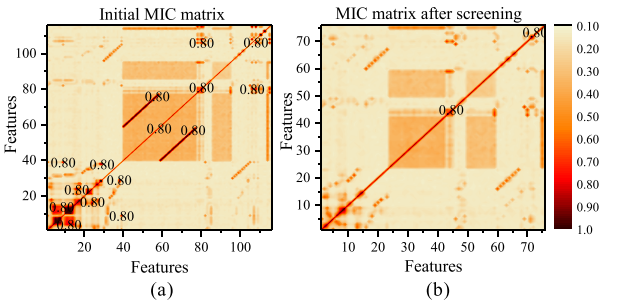


FIGURE 5. MIC screening.

NIS is used to calculate the correlation of residual features to target attributes, i.e., the RSS index. The NIS screening results for valley load duration is shown in Fig. 6.

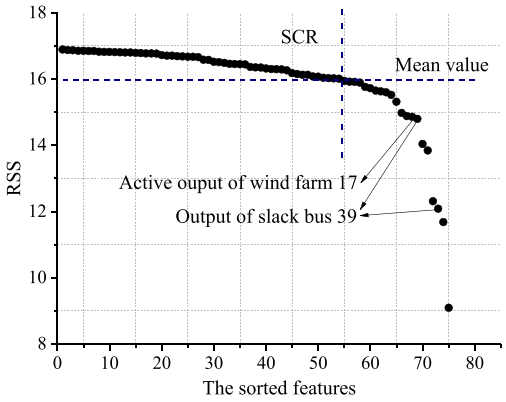


FIGURE 6. Correlations of features to target.

According to Fig. 6, i) the output of generator bus 39 highly correlates to TTC; ii) the active output of the wind farm at bus 17 also has the close relationship with TTC, reflecting that the wind farm has a significant impact on TTC. The above analysis additionally validates the capability of the presented method which is able to automatically select the most dominating features regarding the given objective variable.

After executing MIC and NIS, the ultimately selected attributes, in this case, are obtained, which are included in Tab.3

TABLE 3. Final features set.

Feature category	Feature variables
Voltage	$\Delta U_{2,3,13,31,32}$, $\Delta V_{1-39,18-3}^{\text{tie-lines}}$
Power	$\Delta P_{3,4,7,8,29}^{\text{Load}}$, $\Delta P_{39}^{\text{Gen}}$, $\Delta Q_{39}^{\text{Gen}}$, $\Delta P_{17}^{\text{Wind}}$
Statistics	$\Delta P_{\Sigma}^{\text{Export}}$, $\Delta Q_{\Sigma}^{\text{Receive}}$, $\Delta P_{\Sigma}^{\text{Load}}$

2) FEATURE SELECTION VALIDATION

The original attributes and the selected attributes are fed into the proposed ensemble respectively, and each ensemble contains 50 base learners. Fig. 7 compares the training speed and TTC prediction accuracy of the proposed method, respectively, with and without the procedure of feature selection.

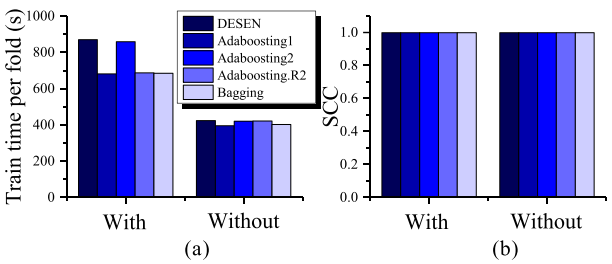


FIGURE 7. (a) Training speed (b) TTC prediction accuracy.

Fig. 7 shows that the training efficiency of the ensembles trained by the selected features is highly improved, without sacrificing TTC prediction accuracy. Because the average SCC only decreases about 0.015% due to the inclusion of feature selection.

C. TTC PREDICTOR

1) COMPARISON OF BASE LEARNERS

Several base learners, including ELM, BPNN, and AHGA-BPNN, are used to train the TTC predictor. Optimal ELM and BPNN prediction models are built by testing different quantity of hidden layer nodes in turn. The performance comparison is shown in Fig. 8, where the training time of ELM and BPNN is the total time of traversing the hidden layer nodes.

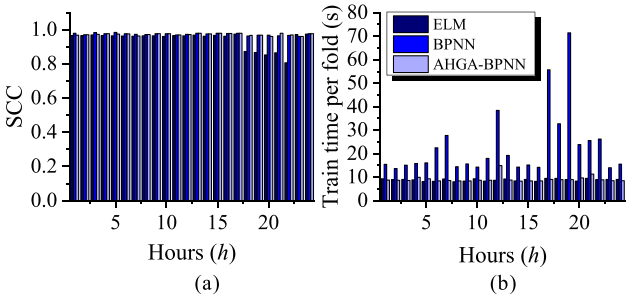


FIGURE 8. (a) TTC prediction accuracy (b) Training speed.

Fig. 8 shows that the testing performance SCC of ELM is about 0.853 on average within the interval from 18 to 21 o'clock. In this period, generator 38 and 39 satisfy the major part of the load. In this case, the two generators output close to their limits, which results in the abnormal data samples in this interval. However, BPNN and AHGA-BPNN have the higher prediction accuracy, 0.974 and 0.973 respectively, representing that these two base leaners are well-adapted when applied to predict TTC based on the given sample set. Regarding training efficiency, ELM is highest among three algorithms, while BPNN has poor performance. The training speed of AHGA-BPNN is close to ELM at most of the time except t = 12 and is much faster than BPNN. Therefore, AHGA-BPNN is best suited for the ensemble than other base learners.

2) ENSEMBLE

Using ensemble approaches proposed in this paper to combine AHGA-BPNNs. Subsequently, the final features set is input into the ensemble. The TTC prediction results of ensembles and the single learner at representative time points (i.e., the valley sampling time $t = 6$, the peak sampling $t = 21$, the normal sampling time $t = 8$, and the abnormal sampling time $t = 18$) are shown in Fig. 9.

Fig.9 shows that DESEN, Adaboosting1, and Adaboosting2 can effectively improve the prediction accuracy. DESEN and Adaboosting1 show strong generalization stability, which can achieve high prediction accuracy in various scenarios.

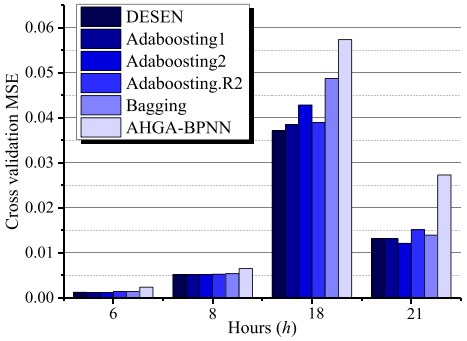


FIGURE 9. TTC prediction error of ensembles.

Moreover, the prediction accuracy of Adaboosting2 at t = 18 is lower than Adaboosting.R2, which shows that Adaboosting2 is unstable under the operating conditions with abnormal fluctuation data. But the prediction error of Adaboosting2 is reduced by 0.59% compared with Bagging. DESEN outperforms other ensembles under an abnormal TTC fluctuation scenario.

In summary, the ensembles proposed in this paper outperform conventional methods in complicated power system operating scenarios. TTC prediction is performed using DESEN, and the prediction results are shown in Fig.10; error distribution and statistics are shown in Fig.11 and Fig.12, respectively. The results of each time prediction are shown in Table 4.

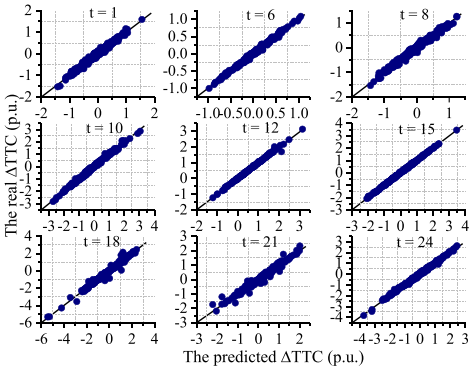


FIGURE 10. Scatter presentation of prediction results.

TTC prediction results of each operating condition are plotted in Fig. 10. The closer the scatters are to the $x = y$ curve, the better the prediction. Fig. 11 and Fig. 12 show the prediction error. It can be observed that the model can realize

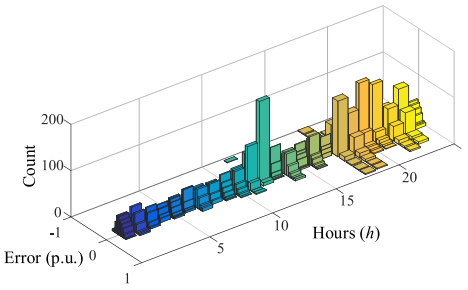


FIGURE 11. TTC prediction error histogram.

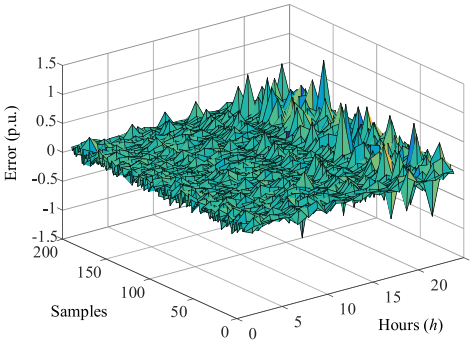


FIGURE 12. TTC prediction error distribution.

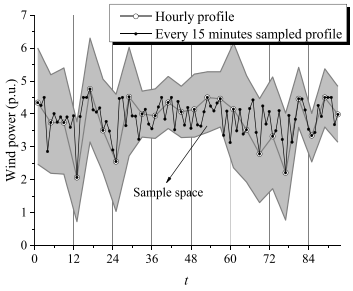


FIGURE 13. Illustration of the coverage of sample space.

accurate TTC prediction and perform strong generalization ability. Moreover, in the scenario with some abnormal data, the model can also achieve high TTC prediction accuracy, showing strong robustness.

Fig.13 illustrates an enveloped sample space base on the hourly profile, by which the wind power profiles with the time granularity of 15 minutes is covered. Hence, the TTC predictor in Section C can be utilized for shorter-time scale prediction.

TABLE 4. TTC predictor performance on test data at each time.

t/*h*	1	2	3	4	5	6	7	8	9	10	11	12
MSE/$\times 10^{-3}$	5.81	7.23	4.77	2.24	2.53	1.21	9.02	5.13	3.78	6.93	6.59	6.41
SCC	0.990	0.991	0.994	0.997	0.995	0.996	0.983	0.992	0.997	0.997	0.998	0.994
t/*h*	13	14	15	16	17	18	19	20	21	22	23	24
MSE/$\times 10^{-3}$	0.446	13.26	1.48	5.04	0.264	37.08	51.35	12.88	12.12	62.72	40.77	6.17
SCC	0.998	0.995	0.999	0.998	0.998	0.990	0.993	0.994	0.991	0.991	0.987	0.997

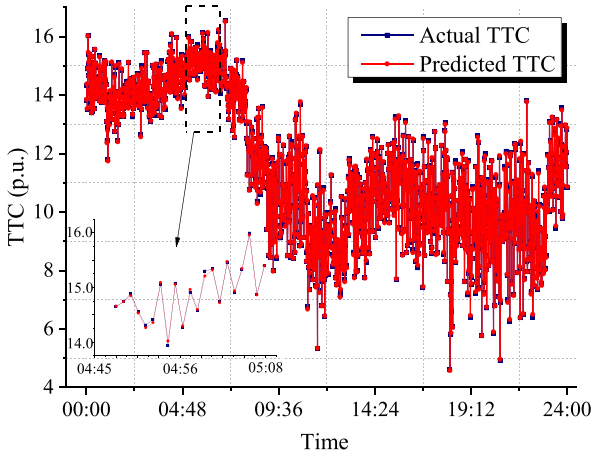


FIGURE 14. Real-time TTC prediction for the 24-hours horizon.

Fig.14 shows TTC prediction results for a whole day using DESEN algorithm. The wind and load data are sampled per minute, and the prediction is also executed in every minute. By applying the ensemble learning based TTC predictor, the accurate result is obtained in 1 second, which satisfies the requirement of real-time implementation. According to the comparison with the fully-computed TTC value (blue nodes in Fig.14), the accuracy of the proposed method is quite high. It is also proved that the DPM approach can include all the possible operation scenarios. Specifically, the maximum error and the minimum error between actual TTC value and prediction are 0.0999 p.u. and 9e-5 p.u., respectively. MAE (Mean absolute error) is also calculated, which equals to 4.21%. According to Fig. 14 and error statistics, the proposed method enables the highly accurate prediction in most sampling scenarios. Moreover, Fig. 14 shows that TTC values dramatically changes along with operating conditions variation, which further proves that using fixed TTC value for security margin calculation might result in potential risk.

D. APPLICATIONS FOR OTHER OPERATING CONDITIONS

The proposed method can be implemented merely in different operating modes and weather conditions. In this section, the effectiveness of the method is validated when the system is operating under two different conditions, which are:

1) operating condition 1: generator 32 sheds for maintenance;

2) operating condition 2: the system is operating in heavy breeze weather.

In these two cases, TTC is calculated for every 15 minutes.

Fig.15 (b) shows TTC prediction results when generator 32 sheds. Besides, when the system is operating in a high wind speed day, predicted TTC is shown in Fig.16.

Fig.15 (a) and Fig.16 (a) illustrate the basic information of temporal profiles in given system operation conditions respectively. Fig.15 (b) and Fig.16 (b) show TTC prediction results by applying DESEN approach, clearly indicating that the continues predicted TTC data have little errors compared

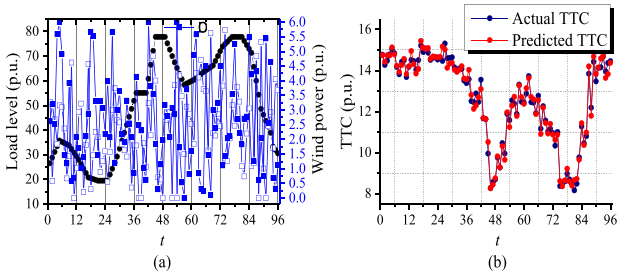


FIGURE 15. (a) Time-series profiles of operating condition 1 (b) TTC prediction results.

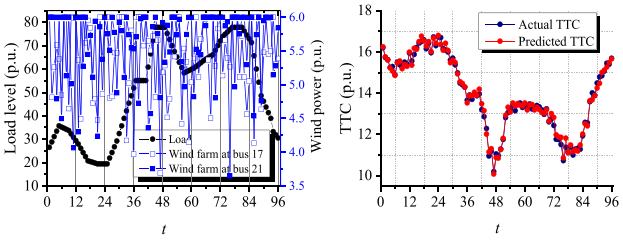


FIGURE 16. (a) Time-series profiles of operating condition 2 (b) TTC prediction results.

with their actual values. The maximum error and the minimum error between actual TTC value and prediction are 0.6968 p.u. and 0.0089 p.u. in operating condition 1, and as for operating condition 2, the values are 0.4320 p.u. and 0.0026 p.u. Another performance index, MSE, equal to 0.0594 p.u. and 0.0261 p.u., respectively, in the operating condition 1 and the operating condition 2. Accordingly, it can conclude that the proposed ensemble learning based TTC prediction possesses high feasibility in adapting to varying operating conditions. Additionally, it is demonstrated that DPM-RDPF method can identify topology change of power system and also ensure the accuracy and generalization of TTC predictor.

E. APPLICATIONS IN THE 16-MACHINE 68-BUS SYSTEM

Another case is studied to validate the feasibility of developing the presented technique in a larger system. We have designed a 68 bus system consisting of 16 synchronous machines, which can be shown in Fig. 17.

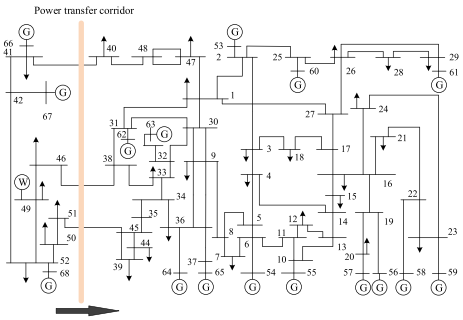


FIGURE 17. 16-machine 68-bus power system.

As shown in Fig. 17, The exporting corridor is composed of line 41-40, 46-38, and 51-45, and the direction of power transfer is inferred with the black arrow. The wind farm whose rated power is 600 MW is integrated in bus 49. In this case, the time granularity is set to 1 hour, and DESEN ensemble is implemented. The result of TTC prediction during 24 hours is shown as Fig. 18, and the comparison between TTC predictions of the operating conditions arising in the test samples and actual TTC values of that is given by the scatters shown in Fig. 19.

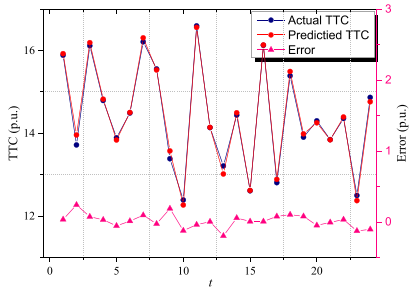


FIGURE 18. The results of the daily TTC prediction.

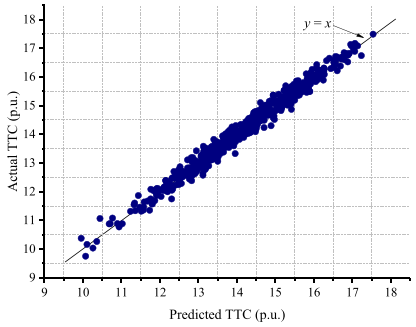


FIGURE 19. Scatter presentation of the prediction results.

The maximum error and the minimum error in the 24-hour horizon prediction are 0.2412 p.u. and 0.0047 p.u, whose related ratios versus the actual TTC are 1.76% and 0.034%, respectively. Besides, in the prediction of the test samples, the maximum error and the minimum error are recorded, which are 0.3371 p.u. and 1.4948e-04 p.u, and also their related ratios are calculated, i.e., 1.52% and 1.072e-05. It is nothing that the data of the 24-hour horizon are not extracted from the test samples, so the statistics of predictions of these two instances are different. Additionally, SCC and MSE, counted from the predictions of the test samples, are given to assess the performance of predicting TTC, i.e., 0.9908 and 0.0412 p.u., respectively. From the statistics above, the proposed method is valid to apply in 68 bus system, the bigger system than the system in the first case. In conclusion, to some extent that the feasibility of deploying the proposed method to large power systems is demonstrated.

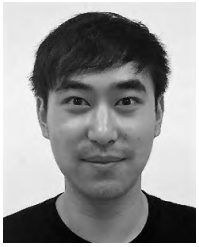
VII. CONCLUSION AND FUTURE WORK

In the presence of large-scale wind farm integration, the power transfer capability of some transmission tie-lines in the grid changes along with the time-varying penetration of wind power. It leads to the potential risk on misestimating security margin for reliable operation. Therefore, in order to enable accurate and real-time awareness, by integrating DPM-RDPF samples generation, MIC & NIS feature selection, and AHGA-NNs training, this paper presented an ensemble learning based architecture to build TTC predictor. Fed by the measured operating state variables of the power grid, the predictor immediately outputs the fitted TTC value which can be used as an essential reference for dispatchers. The case study results demonstrate that the presented ensemble learning method is able to create the TTC predictor with high performance regarding accuracy and generalization ability. The predictor can be implemented as a useful complement to the widely-applied dynamic security assessment (DSA) tools. In the future, it would be fascinating to include more useful operating information to constitute TTC predicator, for example, the phasor data that can be acquired by PMU. In doing so, it is possible to track the trajectory of transmission capability changes. Besides, more efforts can be devoted to employing other advanced machine learning algorithms to realize the higher prediction performance and smarter applications.

REFERENCES

[1] X.-C. Fan, W.-Q. Wang, R.-J. Shi, and F.-T. Li, "Review of developments and insights into an index system of wind power utilization level," *Renew. Sustain. Energy Rev.*, vol. 48, pp. 463–471, Aug. 2015.
[2] G. Hamoud, "Assessment of available transfer capability of transmission systems," *IEEE Trans. Power Syst.*, vol. 15, no. 1, pp. 27–32, Feb. 2000.
[3] M. D. Ilic, Y. Y. Yoon, and A. Zobian, "Available transmission capacity (ATC) and its value under open access," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 636–645, May 1997.
[4] S. Xia, X. Luo, K. W. Chan, M. Zhou, and G. Li, "Probabilistic transient stability constrained optimal power flow for power systems with multiple correlated uncertain wind generations," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 1133–1144, Jul. 2016.
[5] Y. Xu, Z. Y. Dong, K. Meng, J. H. Zhao, and K. P. Wong, "A hybrid method for transient stability-constrained optimal power flow computation," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1769–1777, Nov. 2012.
[6] L. Hakim *et al.*, "A study on the effect of generation shedding to total transfer capability by means of transient stability constrained optimal power flow," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 347–355, Feb. 2009.
[7] Y. Yang, A. Song, H. Liu, Z. Qin, J. Deng, and J. Qi, "Parallel computing of multi-contingency optimal power flow with transient stability constraints," *Protection Control Mod. Power Syst.*, vol. 3, no. 1, p. 20, 2018.
[8] L. Min and A. Abur, "Total transfer capability computation for multi-area power systems," *IEEE Trans. Power Syst.*, vol. 21, no. 3, pp. 1141–1147, Aug. 2006.
[9] D. Qing, Z. Jian-guo, and M. Yan, "Notice of retraction RVM and SVM for classification in transient stability assessment," in *Proc. Asia-Pacific APPEEC*, Mar. 2010, pp. 1–4.
[10] Y. Mansour, E. Vaahedi, and M. A. El-Sharkawi, "Dynamic security contingency screening and ranking using neural networks," *IEEE Trans. Neural Netw.*, vol. 8, no. 4, pp. 942–950, Jul. 1997.
[11] D. Q. Zhou, U. D. Annakkage, and A. D. Rajapakse, "Online monitoring of voltage stability margin using an artificial neural network," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1566–1574, Aug. 2010.
[12] I. B. Sulistiawati, A. Priyadi, O. A. Qudsi, A. Soeprijanto, and N. Yorino, "Critical clearing time prediction within various loads for transient stability assessment by means of the extreme learning machine method," *Int. J. Elect. Power Energy Syst.*, vol. 77, pp. 345–352, May 2016.
[13] T. Liu, Y. Liu, J. Liu, Y. Yang, G. A. Taylor, and Z. Huang, "Multi-indicator inference scheme for fuzzy assessment of power system transient stability," *CSEE J. Power Energy Syst.*, vol. 2, no. 3, pp. 1–9, Sep. 2016.

[14] K. Sun, S. Likhate, V. Vittal, V. S. Kolluri, and S. Mandal, "An online dynamic security assessment scheme using phasor measurements and decision trees," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1935–1943, Nov. 2007.
[15] Y. Xu, Z. Y. Dong, J. H. Zhao, P. Zhang, and K. P. Wong, "A reliable intelligent system for real-time dynamic security assessment of power systems," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1253–1263, Aug. 2012.
[16] X. Liao, K. Liu, H. Niu, J. Luo, Y. Li, and L. Qin, "An interval Taylor-based method for transient stability assessment of power systems with uncertainties," *Int. J. Elect. Power Energy Syst.*, vol. 98, pp. 108–117, Jun. 2018.
[17] J. Zhao, Y. Zhang, P. Zhang, X. Jin, and C. Fu, "Development of a WAMS based test platform for power system real time transient stability detection and control," *Protection Control Mod. Power Syst.*, vol. 1, no. 1, p. 6, 2016.
[18] H. Sun *et al.*, "Automatic learning of fine operating rules for online power system security control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1708–1719, Aug. 2016.
[19] K. Hou *et al.*, "Research on practical power system stability analysis algorithm based on modified SVM," *Protection Control Mod. Power Syst.*, vol. 3, no. 1, p. 11, 2018.
[20] J. Zhang, C. Y. Chung, and L. Guan., "Noise effect and noise-assisted ensemble regression in power system online sensitivity identification," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2302–2310, Oct. 2017.
[21] J. Heinermann and O. Kramer., "Machine learning ensembles for wind power prediction," *Renew. Energy*, vol. 89, pp. 671–679, Apr. 2016.
[22] Y. Xu *et al.*, "Assessing short-term voltage stability of electric power systems by a hierarchical intelligent system," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 8, pp. 1686–1696, Aug. 2016.
[23] H. Falaghi, M. Ramezani, C. Singh, and M.-R. Haghifam, "Probabilistic assessment of TTC in power systems including wind power generation," *IEEE Syst. J.*, vol. 6, no. 1, pp. 181–190, Mar. 2009.
[24] Y. Liu, T. Liu, Y. Xiang, J. Gou, and J. Liu, "Estimating TTC for wind farm integrated power systems based on nonparametric regression analytics," in *Proc. IEEE Power Energy Soc. General Meeting (PESGM)*, Boston, MA, USA, Jul. 2016, pp. 1–5.
[25] P. Giorsetto and K. F. Utsurogi, "Development of a new procedure for reliability modeling of wind turbine generators," *IEEE Trans. Power App. Syst.*, vol. PAS-102, no. 1, pp. 134–143, Jan. 1983.
[26] F. Zaman, S. M. Elsayed, T. Ray, and R. A. Sarker, "Evolutionary algorithms for power generation planning with uncertain renewable energy," *Energy*, vol. 112, pp. 408–419, Oct. 2016.
[27] C. Li *et al.*, "Assessment method and indexes of operating states classification for distribution system with distributed generations," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 481–490, Jan. 2016.
[28] K. Wagstaff, C. Cardie, S. Rogers, and S. SchrÃűdl, "Constrained k-means clustering with background knowledge," in *Proc. 18th Int. Conf. Mach. Learn.*, 2001, pp. 577–584.
[29] D. N. Reshef *et al.*, "Detecting novel associations in large data sets," *Science*, vol. 334, no. 6062, pp. 1518–1524, Dec. 2011.
[30] L. Faivishevsky and J. Goldberger, "Unsupervised feature selection based on non-parametric mutual information," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Santander, Spain, Sep. 2012, pp. 1–6.
[31] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artif. Intell.*, vol. 137, nos. 1–2, pp. 239–263, 2002.



GAO QIU received the M.S. degree in electrical engineering from Sichuan University, Chengdu, China, in 2016, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Information. His research interests include transient stability analysis of power systems and the application of data mining and knowledge discovery technologies in power systems.

JUNYONG LIU received the Ph.D. degree from Brunel University, London, U.K., in 1998. He is currently a Professor with the School of Electrical Engineering and Information, Sichuan University, Chengdu, China. His main research interests include power system planning, operation, stability, and computer applications.

YOUBO LIU received the Ph.D. degree from Sichuan University, China, in 2011, where he is currently an Associate Professor with the College of Electrical Engineering and Information Technology. His research interests include power system planning, power market, and smart grids.

TINGJIAN LIU received the M.S. degree in electrical engineering from Sichuan University, Chengdu, China, in 2016, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Information. His research interests include transient stability analysis of power systems and the application of big-data technologies in power systems.

GANG MU (M'07) received the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 1991. He is currently a Professor with the School of Electrical Engineering, Northeast Electric Power University, Jilin, China. His research interests include power system stability analysis and renewable power generation.