A Pragmatic Method to Determine Transient Stability Constrained with Interface Real Power Flow Limits via Power System Scenario Similarity

Xianzhuang Liu, *Student Member*, *IEEE*, Yong Min, Lei Chen, *Member*, *IEEE*, Xiaohua Zhang, Changyou Feng, and Wei Hu

***Abstract*—In practical power systems, operators generally keep interface flowing under the transient stability constrained with interface real power flow limits (TS-IRPFL) to guarantee transient stability of the system. Many methods of computing TS-IRPFL have been proposed. However, in practice, the method widely used to determine TS-IRPFL is based on selection and analysis of typical scenarios as well as scenario matching. First, typical scenarios are selected and analyzed to obtain accurate limits, then the scenario to be analyzed is matched with a certain typical scenario, whose limit is adopted as the forecast limit. In this paper, following the steps described above, a pragmatic method to determine TS-IRPFL is proposed. The proposed method utilizes data-driven tools to improve the steps of scenario selection and matching. First of all, we formulate a clear model of power system scenario similarity. Based on the similarity model, we develop a typical scenario selector by clustering and a scenario matcher by nearest neighbor algorithm. The proposed method is pragmatic because it does not change the existing procedure. Moreover, it is much more reasonable than the traditional method. Test results verify the validity of the method.**

***Index Terms*—Clustering, data-driven, nearest neighbor, power system scenario similarity, transient stability constrained interface real power flow limit (TSC-IRPFL), typical scenario.**

NOMENCLATURE

TS-IRPFL	Transient Stability Constrained Interface Real Power Flow Limit.
TS	Typical Scenario.
NN	Nearest Neighbor.
α, β, γ	Power flow scenarios.
$\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$	Sets of power flow scenarios.
$\boldsymbol{N}(\alpha), \boldsymbol{B}(\alpha)$	Numerical and binary component of α.
$N_i(\alpha), B_j(\alpha)$	The ith, jth element of $\boldsymbol{N}(\alpha), \boldsymbol{B}(\alpha)$.
$\boldsymbol{w}^N$, $\boldsymbol{w}^B$	Weights of $\boldsymbol{N}(\alpha), \boldsymbol{B}(\alpha)$ in the calculation D_{N} and S_{B}.
B	Bias factor in the formula of comprehensive distance.
$\boldsymbol{v}^{\circ 2}$	Hadamard square (element-wise square) of vector $\boldsymbol{v}$.
$D_{\mathrm{N}}(\alpha, \beta, \boldsymbol{w}^N)$	Numerical distance between α and β.
$D_{\mathrm{B}}(\alpha, \beta, \boldsymbol{w}^B)$	Binary distance between α and β.
$D(\alpha, \beta, \boldsymbol{w}^N, \boldsymbol{w}^{\mathbf{B}}, W)$	Comprehensive distance between α and β.
$\boldsymbol{\alpha}[t]$	The tth value in $\boldsymbol{\alpha}$.
$\boldsymbol{L}[t]$	The accurate TS-IRPFL for $\boldsymbol{\alpha}[t]$.
$\hat{L}(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{L}, \boldsymbol{w}^N, \boldsymbol{w}^B, B)$	The TS-IRPFL of the nearest scenario to β in $\boldsymbol{\alpha}$.
$\alpha[-t]$	$\boldsymbol{\alpha}\backslash\boldsymbol{\alpha}[t]$, '\' stands for set difference.
$\boldsymbol{\sigma}(\boldsymbol{z})$: $\mathrm{R}^{\mathrm{T}} \to \mathrm{R}^{\mathrm{T}}$	Standard softmax function for $\boldsymbol{z}$.
K_{N}	Amplification coefficient in the softmax function.
R_{D}	Down-sampling factor.
$\boldsymbol{C}(\boldsymbol{\alpha}, r)$	The rth operation pattern of $\boldsymbol{\alpha}$.
$\mathrm{T}(\boldsymbol{\alpha}, r)$	Representative for $\boldsymbol{C}(\boldsymbol{\alpha}, r)$.
$\hat{L}^C(\boldsymbol{\alpha}, r)$	Forecast TS-IRPFL of $\boldsymbol{C}(\boldsymbol{\alpha}, r)$.
$\mathrm{T}^{\mathrm{NN}}(\beta, \boldsymbol{\alpha})$	Nearest representative to β in $\boldsymbol{\alpha}$.
$L^{\mathrm{NN}}(\beta, \boldsymbol{\alpha})$	Forecast TS-IRPFL of β by NN using dataset $\boldsymbol{\alpha}$.
μ	Reliability coefficient to leave a margin for TS-IRPFL.
E	Forecast error rate.
A	Forecast accuracy.
$\bar{E}/\bar{A}$	Average forecast error rate/accuracy.
$\bar{A}_{\mathrm{L}}/\bar{A}_{\mathrm{H}}$	Average forecast accuracy of underrated/overrated points.

Manuscript received June 21, 2019; revised August 6, 2019; accepted September 19, 2019. Date of publication March 30, 2020; date of current version January 12, 2020. This work was supported by National Key R&D Program of China (2018YFB0904500) and State Grid Corporation of China.

X. Liu, Y. Min, L. Chen (corresponding author, email: chenlei08@tsinghua.edu.cn), and W. Hu are with the Department of Electrical Engineering, Tsinghua University, Beijing 100084, China.

X. Zhang is with the State Grid Jibei Electric Power Company, Beijing 100053, China.

C. Feng is with the State Grid Corporation of China, Beijing 100031, China.

DOI: 10.17775/CSEEJPES.2019.01420

I. INTRODUCTION

DURING the operation of a power system, the satisfaction of multiple constraints is required to ensure that the system is operating within its secure boundaries [1]–[5]. For practical usage, these constraints are explicitly translated into numerical limits. Among these limits, the transient stability limit is one of the most determining and focused [6], [7]. In practice, the transient stability limit is usually represented by the real power flow limits of several critical interfaces (a set of tie-lines) [8]. The transient stability constrained interface real power flow limit (TS-IRPFL) is an important variable in system operations [9]. Operators generally keep interface flows under the limits to guarantee transient stability of the

2096-0042 © 2019 CSEE

system. If the interface flow goes near or beyond the limit, strategies such as generation adjustment need to be carried out to increase the secure margin of the system [10].

TS-IRPFL varies with the power system scenario. The determination of accurate TS-IRPFL is accomplished by time-domain simulation [11], which brings significant computational burden and is therefore extremely time-consuming [12]. Many methods have been proposed in order to compute TS-IRPFL faster and better. Generally, the advanced methods fall into three categories: direct methods, methods based on fast time-domain simulation, and data-driven methods [12], [13]. Direct methods compute the TS-IRPFL analytically without time-domain simulation. Most of the direct methods stem from the concept of energy function [14], which is a generalization of equal-area criteria to multi-machine systems [15] and [16] introduced transient stability indices based on the energy function while calculating TS-IRPFL [17] developed an analytical method to obtain the sensitivity of the transient energy margin to parameter changes, based on which a good estimation of the TS-IRPFL can be calculated. Other types of methods are based on fast time-domain simulation, which accelerates the time-domain simulation by leveraging the trajectory sensitivity approach [18] or numerical integral techniques [12]. Recently, the data-driven methods boomed due to the rapid development of data-mining algorithms and big-data techniques. Many state-of-the-art machine learning algorithms are applied to solve the TS-IRPFL problem, including fuzzy assessment [19], Artificial Neural Network (ANN), Extreme Learning Machine (ELM) [20], and Linear Regression (LR) [9]. Among these, the most common method is the Decision Tree (DT) [21], [22], since it is simple to build and easy to implement [23].

However, only a few of them are used in real operations. In practice, the determination of TS-IRPFL is generally conducted by a procedure (called traditional method in this paper) consisting of two steps: selection and analysis of typical scenarios (TS's), and scenario matching. First, some TS's are selected by certain factors. These factors include load level, proportion of thermal and hydro generations, the on/off status of some important elements such as the AC-line and static variable compensator (SVC). Some examples of TS's are the high-water period and peak-load scenario, high-water period and valley-load scenario, low-water period and peak-load scenario, low-water period and valley-load scenario, and so on. The TS-IRPFL's of these TS's are then computed and saved. For a scenario to be analyzed, it is matched with a certain TS, and then the corresponding limit of the TS is adopted as the forecast value. In practical applications, it is common for the operators to formulate several operating rules according to the TS's and their TS-IRPFL's. The TS-IRPFL of a scenario is then determined according to the formulated rules. An example of the rules, from a practical power system, is shown in Table I, where C_{A} is the on/off state of series compensator A, L_{B} is the active power of load B, and T_{C} is the TS-IRPFL for interface C.

The idea behind this procedure is easy to understand: the TS's and the corresponding rules can be regarded as "experiences", and the online decision is based on these experiences. However, the obvious deficiency of this method is that the selection of TS's and the matching of scenarios are all based on human experiences. Therefore, a more reasonable method is preferred. Furthermore, since the procedure is manually performed by human beings (usually a calculation report is formulated during the procedure), it increases the burden of the system analysts and operators.

TABLE I
AN EXAMPLE OF OPERATION RULES

Conditions		Matched TS	T_{C} (MW)
C_{A}	L_{B} (MW)		
On	< 125	α^1	430
On	$125 \sim 175$	α^2	400
On	> 175	α^3	380
Off	< 125	α^4	300
Off	$125 \sim 175$	α^5	265
Off	> 175	α^6	235

Multiple data-driven methods have been proposed but few have been developed following the established procedure, which are hardly acceptable in realistic operations. It is the drawbacks of these methods which directly motivate this paper.

In this paper, we propose a pragmatic method to determine TS-IRPFL, which does not change the two main steps of the traditional method but implements them in a data-driven and more reasonable manner. The method is based on the premise of a large number of scenarios whose accurate limits have already been computed.

The contributions of this paper can be summarized in the following three aspects:

1) We establish a fine distance model to quantify the similarity of power flow scenarios, which is the basis of the next step.

2) We utilize clustering and Nearest Neighbor (NN) to build a TS selector and a scenario matcher, which realizes the two steps of the traditional method. This is enlightened by the consistencies of the purposes of TS selection and clustering and that of scenario matching and NN [24].

3) We test our method using the data from a real system. The test results validate its effectiveness and thus demonstrate its practicality.

The proposed method combines the traditional method with data-driven approaches. It is pragmatic and can be easily accepted by the operators because it does not change the process of the traditional method (namely selecting TS's and scenario matching). Furthermore, it is reasonable, intelligent, and suitable for automatic implementation.

The rest of this paper is organized as follows: Section II models the similarity of power flow, defines the comprehensive distance based on the numerical and binary distances, and further elaborates the training procedure. Section III builds a clustering-based TS selector and a NN-based scenario matcher to achieve the targets of TS selecting and scenario matching. Section IV tests the proposed method on a real system and gives the numerical results. Finally, Section V concludes the whole paper.

II. Formulation of Power System Scenario Similarity

This section proposes a similarity index to describe the distance between two scenarios. A synthesized scenario distance model is developed considering both the numerical and binary features of power flow. Weighted vectors are then introduced to reflect the importance of different impact factors for similarity measurement.

The distance model lays the theoretical foundation of this research. It quantitatively evaluates how typical a scenario is, which will be further used for TS selection, and how similar two scenarios are, which will be further used in scenario matching.

The inputs of the model are two scenarios, the output is the distance, and the weights are the parameters to be trained to yield a satisfactory similarity measurement.

A. Formulation of the Model

The distance model is designed under the following principles:

- The definition of the distance must interface with the TS-IRPFL, i.e., "the closer the TS-IRPFL's of the two scenarios, the smaller their distance."
- The definition of distance is preferable to embody the important factors influencing the TS-IRPFL.

For a given power system, its scenario is primarily described by two types of data. The first is numerical data, which consists of loads, generations, line flows, and node voltages. The second is binary data describing the on/off states of elements. By exploring several models of distance in the domain of distance metric learning, the weighted Euclidean distance [25] and the weighted Jaccard distance [26] are chosen to evaluate the scenario distance in this study.

For a power system scenario α, we use two column vectors to describe it: $\boldsymbol{N}(\alpha)$ and $\boldsymbol{B}(\alpha)$. $\boldsymbol{N}(\alpha)$ represents the numerical part of α and $\boldsymbol{B}(\alpha)$ represents the binary part. Let $N_i(\alpha)$ and $B_j(\alpha)$ be the ith and jth element of $\boldsymbol{N}(\alpha)$ and $\boldsymbol{B}(\alpha)$ (I and J elements in total), respectively. The numerical part describes the quantity of electrical measurements, while the binary part describes the on/off states of the critical elements such as AC-lines and transformers in the power system. The elements in the numerical part take the values of the numerical type, while those in the binary part take 0-1. Aiming at the two components, we introduce two indices that measure the distance between two scenarios.

1) Distance Measurement for the Numerical Part

The weighted Euclidean distance is employed to measure the difference between the numerical parts of two scenarios:

$$\begin{aligned} D_{\mathrm{N}}\left(\alpha,\beta,\boldsymbol{w}^{N}\right) &= \left[\sum_{i=1}^{I} w_i^N \cdot \left|N_i(\alpha)-N_i(\beta)\right|^2\right]^{\frac{1}{2}} \\ &= \left[\left(\boldsymbol{w}^{N}\right)^{\mathrm{T}} \cdot \left|\boldsymbol{N}(\alpha)-\boldsymbol{N}(\beta)\right|^{\circ 2}\right]^{\frac{1}{2}} \end{aligned} \tag{1}$$

where $\boldsymbol{w}^N$ is the weight column vector for the numerical part, and the symbol "$\circ 2$" stands for the Hadamard square (element-wise square). Here $\boldsymbol{w}^N$ is used to tailor the conventional Euclidean distance to reflect the importance of each element. Its determination method is elaborated in Part B of this section.

The numerical features are described in Table II.

To obtain values distributed in approximate ranges, we substitute the normalized value for $N_i(\alpha)$:

$$\bar{N}_i(\alpha) = \frac{N_i(\alpha) - N_i^{\mathrm{base}}}{\Delta N_i^{\max}} \tag{2}$$

In this paper, percentiles are exploited to avoid the influences of outliers:

$$\begin{aligned} N_i^{\mathrm{base}} &= P\left(\boldsymbol{N}_i, 50\right) \\ \Delta N_i^{\max} &= P\left(\boldsymbol{N}_i, (1-a/2)\times 100\right) - P\left(\boldsymbol{N}_i, (a/2)\times 100\right) \end{aligned} \tag{3}$$

where $\boldsymbol{N}_i$ is the measured vector of $N_i(\alpha)$'s, $P(\boldsymbol{N}_i, k)$ stands for the kth percentile of $\boldsymbol{N}_i$, and $(1-a)$ is the confidence level. For notational ease, we do not distinguish $N_i(\alpha)$ and $\bar{N}_i(\alpha)$ in this paper.

2) Distance Measurement for the Binary Part

For the binary part of the scenarios, the Jaccard distance is adopted as a distance measurement. Similarly, we introduce vector $\boldsymbol{w}^B$ to embody the importance of each binary element. Hence transform the standard Jaccard distance to the weighted form:

$$D_{\mathrm{B}}\left(\alpha,\beta,\boldsymbol{w}^{B}\right) = 1 - \frac{\left(\boldsymbol{w}^{B}\right)^{\mathrm{T}} \boldsymbol{f}_{\mathrm{AND}}(\alpha,\beta)}{\left(\boldsymbol{w}^{B}\right)^{\mathrm{T}} \boldsymbol{f}_{\mathrm{OR}}(\alpha,\beta)} \tag{4}$$

where $\boldsymbol{f}_{\mathrm{AND}}(\alpha,\beta)$ and $\boldsymbol{f}_{\mathrm{OR}}(\alpha,\beta)$ respectively represent the element-wise "and" and "or" logical operations for vectors, and are mathematically defined as:

$$\begin{aligned} \boldsymbol{f}_{\mathrm{AND}}(\alpha,\beta) &= \boldsymbol{B}(\alpha)\ \mathrm{AND}\ \boldsymbol{B}(\beta) = \boldsymbol{B}(\alpha)\circ\boldsymbol{B}(\beta) \\ \boldsymbol{f}_{\mathrm{OR}}(\alpha,\beta) &= \boldsymbol{B}(\alpha)\ \mathrm{OR}\ \boldsymbol{B}(\beta) \\ &= \left[1-\boldsymbol{B}(\alpha)\right]\circ\left[1-\boldsymbol{B}(\beta)\right] \end{aligned} \tag{5}$$

where the symbol "$\circ$" stands for the Hadamard product (element-wise product).

TABLE II
Numerical Features

Element Type	Variable	Description
Bus	$V_{\mathrm{a}}^{\mathrm{r}}, V_b^{\mathrm{i}}$	Real and image part of the voltage of bus a
AC-line (including transformer)	P_{ab}, Q_{ab}	Active and reactive power injection from bus a through AC-line L_{ab}
	Q_{ca}	Capacitive charging power of L_{ab} at the side of bus a
DC-line	P_{ab}	Active power injection from bus a through DC-line L_{ab}
Generator	$P_{\mathrm{a}}^{\mathrm{g}}, Q_{\mathrm{a}}^{\mathrm{g}}$	Active and reactive power generation of the generator connected to bus a
Load	$P_{\mathrm{a}}^{\mathrm{l}}, Q_{\mathrm{a}}^{\mathrm{l}}$	Active and reactive power consumption of the load connected to bus a

Without the weights, the Jaccard distance depicts the relative size of the difference of the two finite sets α and β. Introducing the weights emphasizes the importance of the on/off states of some critical elements.

The parameters $\boldsymbol{w}^N$ and $\boldsymbol{w}^B$ can be leveraged to adjust the similarity indices to embody the importance of the elements and cater to the principles proposed in the first place. The closer the TS-IRPFL's of the two scenarios are, the smaller their distance should be.

D_{N} and D_{B} are specifically referred to as **numerical distance** and **binary distance** in this article. The smaller the values of D_{N} and D_{B} are, the more similar the two scenarios will be in terms of TS-IRPFL (stability).

3) The Synthesized Distance

The numerical and binary distances are synthesized into a comprehensive distance D:

$$\begin{aligned} &D\left(\alpha,\beta,\boldsymbol{w}^N,\boldsymbol{w}^B,W\right)\\ &=D_{\mathrm{N}}\left(\alpha,\beta,\boldsymbol{w}^N\right)+B\times D_{\mathrm{B}}\left(\alpha,\beta,\boldsymbol{w}^B\right) \end{aligned} \tag{6}$$

where B is the bias coefficient, which weighs the importance of D_{N} and D_{B}.

The synthesized distance makes it possible to compare between two scenarios and to quantify their differences. The coefficients $\boldsymbol{w}^N$, $\boldsymbol{w}^B$, and B determine the fitness of the model.

B. Determination of Weight Coefficients

Suppose we have several scenarios already labeled by the TS-IRPFL:

$$\alpha=\{\alpha[t]\,|t\in\boldsymbol{T}\},\boldsymbol{L}=\{\boldsymbol{L}[t]\,|t\in\boldsymbol{T}\},\boldsymbol{T}=\{1,2,\ldots,T\} \tag{7}$$

where $\boldsymbol{L}$ is the set of corresponding labels (TS-IRPFL's).

For an unlabeled scenario β, if its TS-IRPFL is determined by its nearest neighbor in $\boldsymbol{\alpha}$, with the assumption that the coefficients in the distance model are already determined, then the determined (forecast) value of TS-IRPFL for β should be:

$$\hat{L}\left(\beta,\boldsymbol{\alpha},\boldsymbol{L},\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)=L\left[n_{\text{nst}}\left(\beta,\boldsymbol{\alpha},\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)\right] \tag{8}$$

where n_{nst} is the index of the nearest scenario to β in $\boldsymbol{\alpha}$:

$$n_{\text{nst}}\left(\beta,\boldsymbol{\alpha},\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)=\arg\min_{t\in\boldsymbol{T}}D\left(\alpha_t,\beta,\boldsymbol{w}^N,\boldsymbol{w}^B,W\right) \tag{9}$$

The coefficients in the distance model are determined by the following training process. For an arbitrary scenario $\alpha[t]$ in α, we use scenarios with $\alpha[t]$ excluded (denoted by $\boldsymbol{\alpha}[-\boldsymbol{t}]=\boldsymbol{\alpha}\backslash\boldsymbol{\alpha}[t]$) to forecast the TS-IRPFL of $\boldsymbol{\alpha}[t]$:

$$\hat{L}^t\left(\boldsymbol{\alpha},\boldsymbol{L},\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)=\hat{L}\left(\boldsymbol{\alpha}[t],\boldsymbol{\alpha}\left[-\boldsymbol{t}\right],\boldsymbol{L},\boldsymbol{w}^N,\boldsymbol{w}^B,B\right) \tag{10}$$

The forecast value should be as close as possible to the actual value $L[t]$, which is already known. Then, the objective of the training process is to minimize the overall squared error, with the sum and boundary constraints on the weights:

$$\begin{aligned} \min_{\boldsymbol{w}^N,\boldsymbol{w}^B,B}&\sum_{t\in\boldsymbol{T}}\{\hat{L}^t\left(\boldsymbol{\alpha},\boldsymbol{L},\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)-L[t]\}^2\\ \text{s.t. }&\sum_{i=1}^{I}w_i^N=1,\quad\sum_{i=1}^{I}w_i^B=1\\ &\boldsymbol{w}^N\geq 0,\boldsymbol{w}^B\geq 0,B\geq 0 \end{aligned} \tag{11}$$

The optimum measurement of the system scenario distance is determined by optimizing the decision variables $\boldsymbol{w}^N$, $\boldsymbol{w}^B$, and B.

The gradient descent algorithm is used to train the model and obtain the optimal solutions. To make sure that the optimization is led by a descent progress, two improvements are made.

Since the "hardmin" (i.e., argmin) function is not differentiable, so we replace the argmin function with a decay version of softmax [27], to ensure a good performance of the gradient descent algorithm. Specifically, we change the forecast function in (8) into:

$$\begin{aligned} &\hat{L}^t\left(\beta,\boldsymbol{\alpha},\boldsymbol{L},\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)\\ &=\left[\boldsymbol{\sigma}\left(-K_{\mathrm{N}}\boldsymbol{D}\left(\boldsymbol{\alpha},\beta,\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)\right)\right]^{\mathrm{T}}\boldsymbol{L}_v \end{aligned} \tag{12}$$

where:

$$\boldsymbol{D}\left(\boldsymbol{\alpha},\beta,\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)=\begin{bmatrix}D\left(\alpha[1],\beta,\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)\\ \vdots\\ D\left(\alpha[T],\beta,\boldsymbol{w}^N,\boldsymbol{w}^B,B\right)\end{bmatrix}, \tag{13}$$

$$\boldsymbol{L}_v=\begin{bmatrix}L[1] & \ldots & L[T]\end{bmatrix}^{\mathrm{T}} \tag{14}$$

and $\boldsymbol{\sigma}:\mathrm{R}^{\mathrm{T}}\rightarrow\mathrm{R}^{\mathrm{T}}$ is the standard softmax function, defined by the following formula:

$$\boldsymbol{\sigma}\left(\boldsymbol{z}\right)_k=\frac{e^{z_k}}{\sum_{\kappa=1}^{K}e^{z_\kappa}},k\in\{1,\ldots,K\},\boldsymbol{z}=\left[z_1,\ldots,z_K\right]^{\mathrm{T}} \tag{15}$$

With this transformation we can circumvent the discontinuity of the original objective function. The amplification coefficient K_{N} creates a more concentrated output vector around the positions of the minimum input value.

When the size of $\boldsymbol{\alpha}$ is huge, the calculation of (10) will impose great computational burden on the training process. Therefore, a down-sampling technique is adopted as the second improvement. We stochastically pick T_{S} samples from $\boldsymbol{\alpha}[-\boldsymbol{t}]$ to formulate its subset $\boldsymbol{\alpha}^S[-\boldsymbol{t}]$, and then use this subset as a substitute for $\boldsymbol{\alpha}[-\boldsymbol{t}]$. The ratio $R_{\mathrm{D}}=T_{\mathrm{S}}/(T-1)$ is called the down-sampling factor [28].

The values of K_{N} and R_{D} influence the trade-off between model accuracy and computational efficiency. The detailed procedure of parameter adjustment is shown in Section IV.

This problem is solved by tensorflow, an efficient software for machine-learning tasks [29].

III. Determining TS-IRPFL Via Power System Scenario Similarity

The data-driven counterparts of selecting TS's and matching rules are elaborated in this section. Based on the distance model defined in Section II, we exert clustering and NN to build the TS selector and scenario matcher.

A. Clustering-directed TS Selector

Clustering is a classical and widely used data-mining model used in various topics in power system researches, and is generally employed for unsupervised learning, which usually does not have a known index as an objective to follow [30], [31]. To take advantage of the information of the scenarios with known TS-IRPFL, which constructs a labeled dataset, this part designs a supervised mechanism for clustering.

Clustering divides the dataset into several groups, which reflect the typical features of each group. The aim of clustering is to maximize the similarity of the data within each group and dissimilarity between the groups. Power flow fluctuates with certain factors such as date, hour, weather, etc. [32]. Investigations and researches show apparent patterns of power flow scenarios [33]. In fact, we can deem the selection of TS's in traditional TSA as a simple clustering through manual statistical analysis. Moreover, the methods and results for power flow clustering can be different according to the topics of analysis.

We conduct the clustering using the distance given by (6), and the result is called an operation pattern:

$$\{\boldsymbol{C}(\boldsymbol{\alpha}, r)\}, \quad r \in \boldsymbol{R}, \quad \boldsymbol{R} = \{1, 2, \ldots, R\} \tag{16}$$

where $\boldsymbol{\alpha}$ is the dataset of power flow scenarios, and $\boldsymbol{C}(\boldsymbol{\alpha}, r)$ is the rth operation pattern. We use $C(\boldsymbol{\alpha}, r)$ to denote the center of the pattern.

The TS-IRPFL (label) of the cluster is set as the median of the TS-IRPFL's of its members, and the scenario correspondent to this value is designated as the representative of the cluster:

$$\begin{aligned} \tau(\boldsymbol{\alpha}, r) &= \arg \underset{\tau \in T}{\text{median}} \left(\{ L[\tau] \,|\alpha[\tau] \in \boldsymbol{C}(\boldsymbol{\alpha}, r) \} \right) \\ \hat{L}^C(\boldsymbol{\alpha}, r) &= L\left[\tau(\boldsymbol{\alpha}, r)\right] \\ \mathrm{T}(\boldsymbol{\alpha}, r) &= \alpha\left[\tau(\boldsymbol{\alpha}, r)\right] \end{aligned} \tag{17}$$

where $\mathrm{T}(\boldsymbol{\alpha}, r)$ and $\hat{L}^C(\boldsymbol{\alpha}, r)$ are representative and the label for the cluster $\boldsymbol{C}(\boldsymbol{\alpha}, r)$, respectively. The representative is defined as being out of a center because the center is usually an imaginary point which may not occur in the physical system; therefore, a representative that already appeared is more acceptable as a typical scenario.

The TS selector takes a labeled dataset $\boldsymbol{\alpha}$ as an input and yields TS's $\mathrm{T}(\boldsymbol{\alpha}, r)$ as output.

B. NN-directed Scenario Matcher

The purpose of scenario matching is to find the "nearest" TS to the target scenario, which coincides with the purpose of NN. Therefore, we build a NN-directed scenario matcher to fulfill the matching step.

Suppose $\boldsymbol{\alpha}$ are labeled scenarios, then the online unlabeled scenario β is matched to one of the clusters. The representative which has the maximum D with β is denoted by $\gamma^{\mathrm{NN}}(\beta, \boldsymbol{\alpha})$, and it is specified as the matching TS of β:

$$\begin{aligned} r^{\mathrm{NN}}(\beta, \boldsymbol{\alpha}) &= \arg \min_{r \in \boldsymbol{R}} D\left[\beta, \mathrm{T}(\boldsymbol{\alpha}, r)\right] \\ \mathrm{T}^{\mathrm{NN}}(\beta, \boldsymbol{\alpha}) &= \mathrm{T}\left[\boldsymbol{\alpha}, r^{\mathrm{NN}}(\beta, \boldsymbol{\alpha})\right] \end{aligned} \tag{18}$$

The forecast value of TS-IRPFL for β is:

$$L^{\mathrm{NN}}(\beta, \boldsymbol{\alpha}) = \mu \hat{L}^C\left[r^{\mathrm{NN}}(\beta, \boldsymbol{\alpha})\right] \tag{19}$$

The reliability coefficient μ is applied in (18) to leave a margin for TS-IRPFL, a typical value of μ is 0.95.

The scenario matcher takes unlabeled scenario β and TS's as inputs and yields L^{NN} as an output (TS-IRPFL).

C. Overall Flowchart

The overall flowchart of the comprehensive model is shown in Fig. 1.

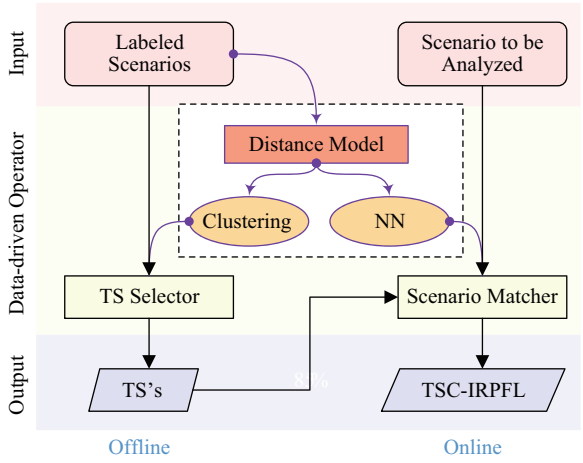

Fig. 1. Overall flowchart of the comprehensive model.

The inputs are the historical labeled scenarios and the real-time unlabeled scenarios, which are fed into the data-driven utensils directed by clustering and NN. The clustering and NN algorithms are all based on the distance model. The selector then yields TS's and the matcher yields the forecast value of TS-IRPFL for the real-time unlabeled scenario.

In the conventional method, the selector and the matcher are implemented manually by the operators. The clustering-directed selector and the NN-directed matcher together can be viewed as a "data-driven operator." The core technique for them is the distance model.

IV. Test Results

A. System Description

The system used in the test case study is a practical regional power grid in China. It consists of about 3,000 stations, 15,000 AC-lines, 6,300 transformers, 15 DC-lines, 4,000 generators and 25,000 loads. The total generation capacity of the grid is about 1,000 GW.

The regional power grid is composed of 22 interconnected areas. In the case study, we focus on a specific interface that interconnects two of the above areas E and G. The interface is denoted as EG, and consists of three AC-lines, as shown in Fig. 2.

A total of 13,000 scenarios are selected to calculate the real TS-IRPFL's of EG. Whereby, 10,000 of them are used as a training set to train the weights and acquire clusters, and the rest are used as a testing set.

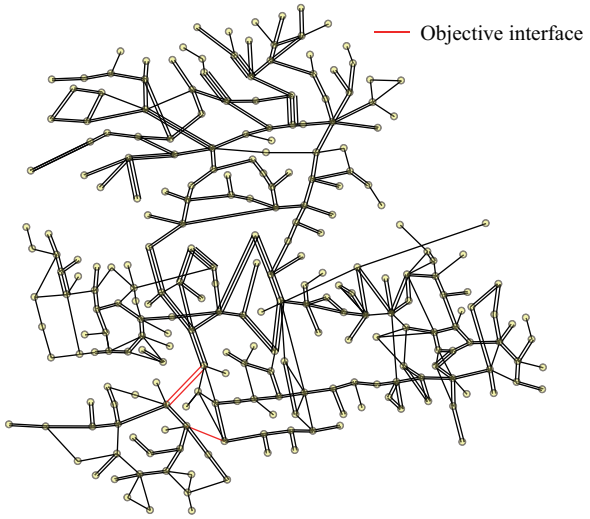

Fig. 2. System structure.

The system structure is presented in Fig. 2. Only elements with a voltage level higher than 500 kV are reserved. For the sake of confidentiality, the structure is distorted.

B. Feature Selection

Since the feature number is too large, we need to eliminate some unimportant features. To achieve this, we conduct a feature selection procedure in practice. The procedure follows two steps:

1) First, we roughly eliminate unimportant features by human experience. We set a threshold for voltages as $T_{\mathrm{V}} =$ 500 kV, a capacity threshold for generators as $T_{\mathrm{G}} = 600$ MW, and a power threshold for loads as $T_{\mathrm{L}} = 100$ MW. The following elements will be eliminated:

a) Buses, AC-lines with lower voltage levels than T_{V};
b) Transformers whose voltage levels of the high voltage sides are lower than T_{V};
c) Generators with lower capacities than T_{G};
d) Loads with maximum consumptions lower than T_{L}.

Through the first step we reserve about 5% of the buses, 10% of the AC-lines, 7% of the transformers, 11% of the generators and 0.8% of the loads. About 4,500 features in total are reserved during this step.

2) Then we run optimization progress several times. Each time we only run a few epochs (which consumes only a few seconds). Then we pick out the features whose weights are always smaller than a given threshold T_{w}, and eliminate them. T_{w} is set as 0.01 here. This step will stop when the features are fewer than 1,000. Since the computational cost is approximately proportional to the square of the feature number, this step will highly accelerate the following training procedure.

C. Typical Scenarios

TS's are selected through the original rules and the proposed clustering-based method to conduct a benchmark. The original rules in real operation are shown in Table III.

A and B are two AC-lines and L_{G} is the total load of area G.

TABLE III
ORIGINAL RULES

Original			TS-IRPFL (MW)
A	B	L_{G} (GW)	
On	On	≤ 10.5	2,600
On	On	$10.5 \sim 12.5$	2,800
On	On	≥ 12.5	3,000
On	Off	–	1,600
Off	On	–	1,800

The scenarios are then matched to the TS's selected by these rules to formulate 5 clusters.

The distribution of the TS-IRPFL's of scenarios in each cluster is boxplotted in Fig. 3.

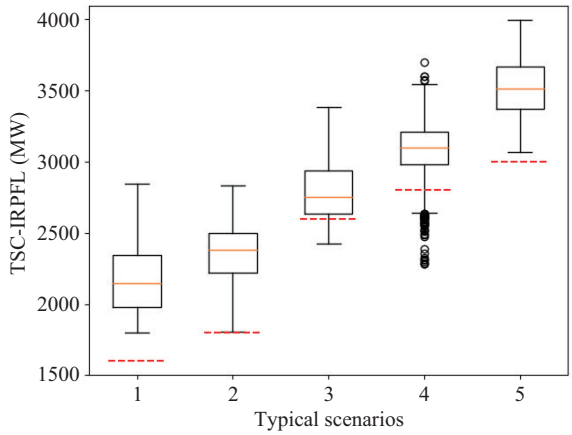

Fig. 3. Distribution of TS-IRPFL under original TS's.

The dashed red lines show the forecast values of TS-IRPFLS's using the original rules. It is shown that the rules are generally conservative, yet the clusters are not well separated. This means the rules may sacrifice some economy and that the typical scenarios are not typical.

TS's are then selected by the clustering-based selector. For the sake of comparison, the cluster number is set to be the same as the original rules. The result is shown in Fig. 4.

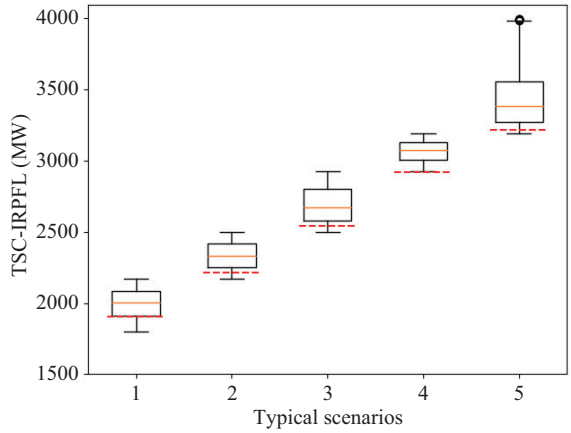

Fig. 4. Distribution of TS-IRPFL under new TS's.

As we can see, the TS-IRPFL's are more separated between clusters and more concentrated within each cluster. This is

also confirmed by the results of standard deviations listed in Table IV.

TABLE IV
STANDARD DEVIATION OF THE TS-IRPFL'S OF SCENARIOS MATCHED WITH EACH TS (MW)

TS	Original	New
1	233	103
2	195	96
3	197	126
4	185	74
5	212	200
mean	205	120

It is more intelligent to select TS's by the clustering-based selector. Scenarios with closer TS-IRPFL's are more likely to be grouped into the same cluster, and by further investigation, more common features of the scenarios within the same cluster can be excavated.

Note that the reason we choose 5 as the cluster number is to keep consistent with the traditional method. Generally speaking, if we increase the cluster number, the accuracy will increase and the standard deviation of the TS-IRPFL's will decrease. However, the comparability will be sacrificed in this case.

The choosing of the optimal cluster number is a trade-off between the accuracy of the forecast result and the typicality of the TS's, which will be elaborated in the next part.

D. Matching Results

Using the distance model and the TS's, we match the scenarios in the testing dataset. A fragment of the results is chosen and shown in Fig. 5.

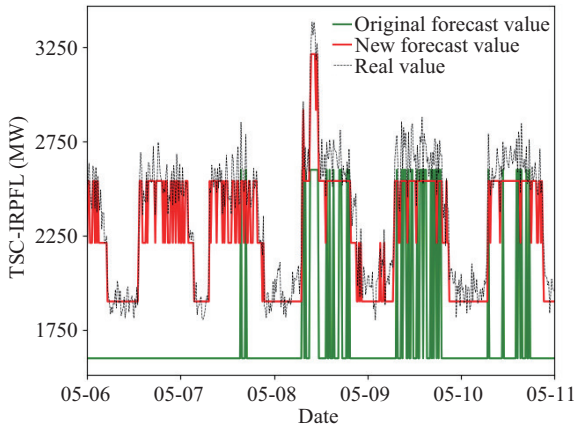

Fig. 5. Real limit vs Forecast limit.

The original forecast values are rough and inaccurate compared to the new ones.

The accuracy can be further illustrated by Fig. 6 and Table V. We define the forecast error rate E and forecast accuracy A as:

$$E = \frac{L_{\text{forecast}} - L_{\text{real}}}{L_{\text{real}}} \times 100\%$$
$$A = 100\% - |E| \tag{20}$$

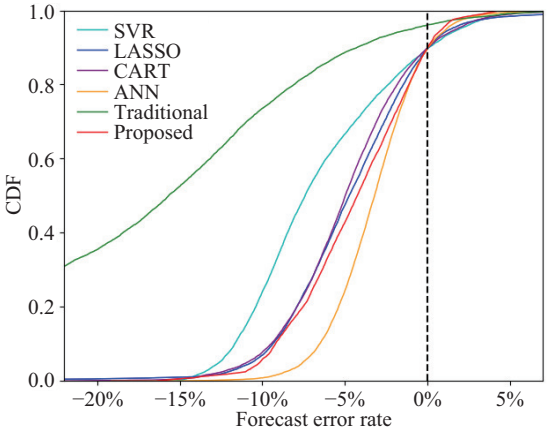

Fig. 6. Cumulative distribution functions of forecast error rates.

TABLE V
ERROR ANALYSIS ON DIFFERENT DATA-DRIVEN METHODS

Methods	A	P_{L}	P_{H}	A_{L}	A_{H}
SVR	93.0%	89.7%	10.3%	92.5%	97.8%
LASSO	94.6%	89.7%	10.3%	94.4%	96.8%
CART	94.7%	89.8%	10.2%	94.4%	97.7%
ANN	96.5%	89.6%	10.4%	96.2%	98.4%
Traditional	82.9%	96.1%	3.9%	82.4%	96.4%
Proposed	95.3%	89.7%	10.3%	94.9%	98.8%

Furthermore, $\bar{E}/\bar{A}$ is defined as the average forecast error rate/accuracy, and $\bar{A}_{\text{L}}/\bar{A}_{\text{H}}$ as the average forecast error accuracy of the scenario whose forecast value is higher/lower than the real value. We calculate $\bar{A}_{\text{H}}$ and $\bar{A}_{\text{L}}$ because TSA is usually required to be conservative. Compared to a higher L_{forecast}, which can affect the stability of the system, the operators would rather accept a lower L_{forecast}, which only reduces the economy.

Under this definition of accuracy, we compare our method with other most recent methods. As mentioned in Section I, the time-domain simulation is the most accurate method. However, it is extremely time-consuming and thus not suitable for on-line estimation of TS-IRPFL. Moreover, most recent researches are focused on data-driven approaches. Therefore, we pick 4 typical data-driven approaches to make a comparison.

The contrastive methods are Support Vector Regression (SVR), Least Absolute Shrinkage and Selection Operator (LASSO), Classification and Regression Tree (CART), and ANN. The parameter μ is set to a proper value in each method to guarantee that 90% of the forecast limits are lower than the real limits.

The cumulative distribution functions (CDF's) of the forecast error rates of these methods are shown in Fig. 6. The numerical results are shown in Table V. P_{L} and P_{H} are proportions of underrated and overrated scenarios in the whole dataset, respectively.

As we can see, the proposed method is more accurate in terms of $\bar{A}$, $\bar{A}_{\text{L}}$, and $\bar{A}_{\text{H}}$. The accuracies of the data-driven methods are close in terms of $\bar{A}$. Specifically, the accuracy of the proposed method is 1.2% lower than that of ANN. This is caused primarily by $\bar{A}_{\text{L}}$ (conservativeness), which is acceptable in real operations.

The conservativeness of the proposed method is 7.1% lower

than the traditional method in terms of P_{L}. Therefore, to promote the conservativeness, the parameter μ can be decreased; however, this may lower the accuracy.

The trade-off between accuracy and conservativeness is shown in Fig. 7. Using this figure, a more accurate μ value can be determined, which maintains the performance of both accuracy and conservativeness.

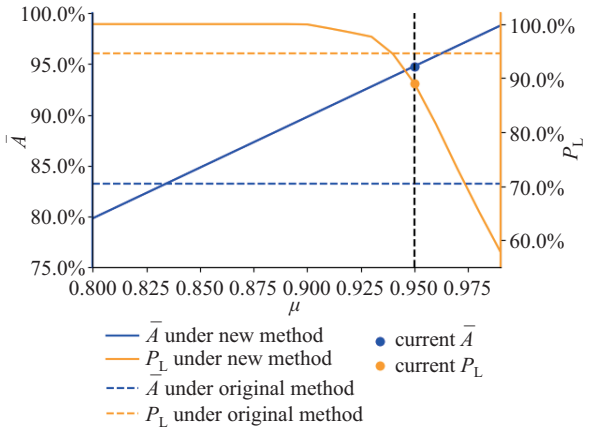

Fig. 7. Accuracy vs conservativeness.

As mentioned in the previous part, there is an optimal value for the cluster number N_{C}. A large N_{C} yields a more accurate result, but also decreases the scenario numbers matched with each TS. If only a few scenarios are matched to each TS, then the typicality of the TS's will fade. Besides, if there are too many TS's, it will be harder for the operators to analyze each TS, thus reducing the practicability of the method. So we need to pick an optimal point regarding the trade-off between accuracy and typicality. The relationship between the cluster number and accuracy is shown in Fig. 8.

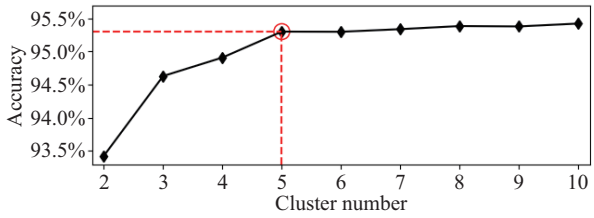

Fig. 8. The relationship between cluster number and accuracy.

As we can see, the accuracy hardly changes with N_{C} larger than 5 (less than 0.1%), thus in terms of accuracy, there is little difference in choosing $5 \sim 10$ as the cluster number.

E. Critical Elements

Another issue is the critical elements reflected by the TS's. 10 scenarios are picked from each cluster. We select 20 consecutive variables and 10 binary variables (by largest weights) to observe the differences, as shown in Fig. 9.

The color of the consecutive variable reflects its relative value. The colder (bluer) the color is, the value it represents is smaller, while the warmer (redder) color suggests a greater value. And the red spot for the binary variable indicates that the element is "on" for that point.

In the new TS's, the influences of critical elements on the divisions of TS's are coupled. For example, in TS 3, the influence of the low power flow on AC-line 5 is counteracted by the activation of generator B, and so the last three scenarios have similar TS-IRPFL's with that of the first seven.

Figure 10 further shows the spatial distribution of the critical elements under new TS's.

The results are intuitive and in accordance with the operators' experience:

- The critical thermal plants are located in the two areas connected by the target interface.

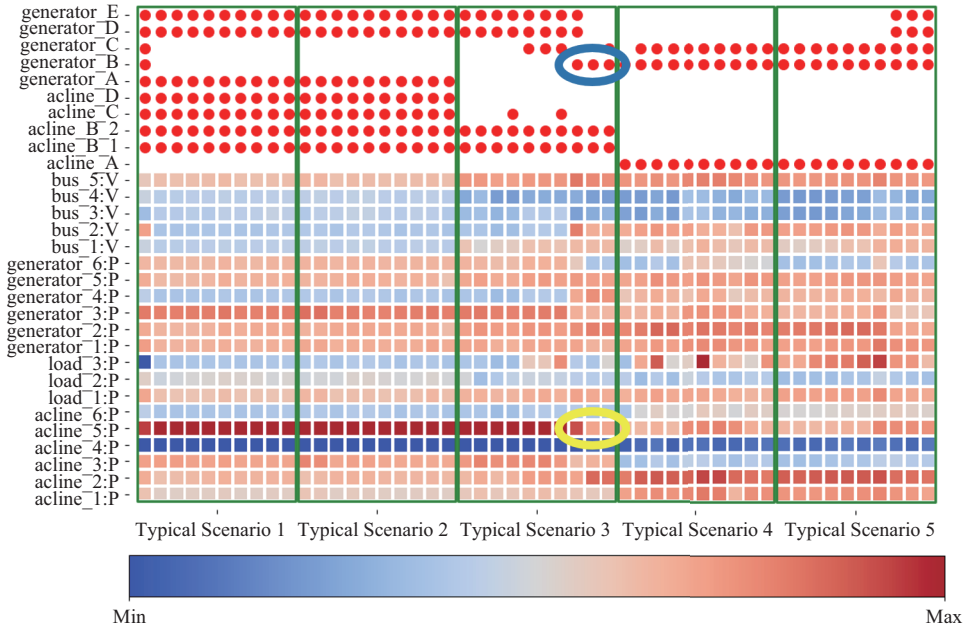

Fig. 9. Differences between new scenarios.

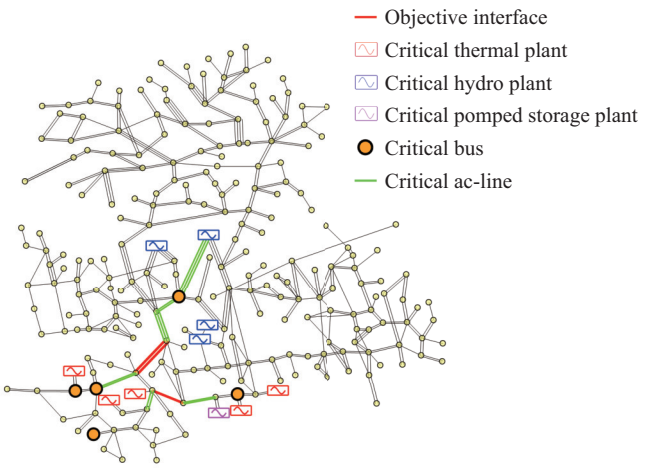

Fig. 10. Spatial distribution of critical elements.

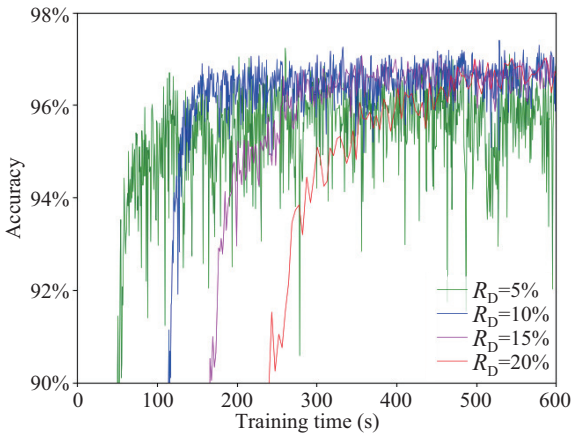

Fig. 11. Accuracy vs computational efficiency.

- The critical buses are pivot buses in those areas.
- The critical hydro plants and the AC-lines are in the same transmission passages with the target interface.

Compared to other data-driven methods, the proposed method is process-clear and can provide operators with informative knowledge. Furthermore, the TS's and critical elements help them master the critical links and stability margins of the system.

F. Computational Efficiency

In this part we discuss the computational efficiency of our method. The computer we used for the simulation possesses 4 cores, each one is a 2.60 GHz CPU (i7-5600U).

The computational efficiency focuses on both the training stage and the forecasting stage. As is mentioned, the cluster number is about 10 and the feature number is about 1,000. Therefore, the forecasting will calculate about 20,000 sums and 10,000 squares. The burden is quite small and will be done within 10 ms. So in this part we focus on the training stage.

As mentioned in Section-II(B), the values of K_{N} and R_{D} will influence the accuracy as well as the computational efficiency. Experiments show that K_{N} has little influence on the computational efficiency, so we concentrate on the down-sampling factor R_{D}.

The decrease of R_{D} will accelerate the training time of each epoch in the gradient descent progress, but it also increases the error in the gradient computation. To investigate the influence, we let $R_{\mathrm{D}} = 5\%$, 10%, 15%, 20%, and plot the accuracies under these values with the same figure, as shown in Fig. 11.

The reason that the profiles are initiated from different positions is that some preparation time is needed to formulate the optimization problem. The larger is R_{D}, the longer is the preparation time.

As is shown, when R_{D} is small (5%), the convergence is not stable. When R_{D} is large (20%), the preparation takes too long. An eclectic choice will be $R_{\mathrm{D}} = 10\%$, where both accuracy and computational efficiency will be guaranteed.

Under this value of R_{D}, about $200 \sim 300$ seconds of training time will be enough to achieve a training accuracy of about 96%. Such performance is acceptable.

V. Conclusion

The traditional method to estimate TS-IRPFL consists of two steps, selecting TS's and scenario matching, which present disadvantages of low precision and accuracy as well as a heavy burden on the operators. Attempting to improve methods commonly suffers from impracticality because they deviate from the established procedure. This paper proposes a pragmatic method for determining TS-IRPFL, which inherits the selecting and matching steps of the traditional method but implements them with data-driven tools. We built a distance model for power flow scenarios and provided the training algorithm. Based on the model, we established a clustering-directed TS selector and a NN-directed matcher to realize the two steps.

The numerical results based on real data show that the method is more scientific and accurate. The analysis of computational efficiency further validates the method's feasibility and practicality. Compared to other data-driven methods, the proposed method is more pragmatic without loss of accuracy.

Future works will concentrate on the following aspects:

- A more feasible and versatile version of the distance model, for example, one that considers the Mahalanobis distance as the similarity index.
- Scenario replenishment directed by data-driven methods, or a "labeler," to efficiently generate high-information samples.

The proposed method can also be utilized to forecast indices other than TS-IRPFL, such as critical clearing time (CCT). Software guided by the proposed method was developed and deployed in a real regional grid in China, and has already been functioning for several months.

References

[1] NERC, "Available transfer capability de?nitions and determination," North America Electric Reliability Council, Princeton, Jun. 1996.

[2] J. Xiao, G. Q. Zu, X. X. Gong, and F. X. Li, "Observation of security region boundary for smart distribution grid," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1731–1738, Jul. 2017.

[3] Y. B. Shu and Y. Tang, "Analysis and recommendations for the adaptability of China's power system security and stability relevant standards," *CSEE Journal of Power and Energy Systems*, vol. 3, no. 4, pp. 334–339, Dec. 2017.
[4] W. Y. Wang, M. Barnes, and O. Marjanovic, "Stability limitation and analytical evaluation of voltage droop controllers for VSC MTDC," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 2, pp. 238–249, Jun. 2018.
[5] W. X. Li, P. Wang, and Z. Z. Guo, "Determination of optimal total transfer capability using a probabilistic approach," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 862–868, May 2006.
[6] W. Hu, Z. X. Lu, S. Wu, W. L. Zhang, Y. Dong, R, Yu, and B. S. Liu, "Real-time transient stability assessment in power system based on improved SVM," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 1, pp. 26–37, Jan. 2019.
[7] Y. F. Zhou, W. Hu, Y. Min, and L. Zheng, "Active splitting strategy searching approach based on MISOCP with consideration of power island stability," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 3, pp. 475–490, 2019.
[8] H. B. Sun, F. Zhao, H. Wang, K. Wang, W. Y. Jiang, Q. L. Guo, B. M. Zhang, and L. Wehenkel, "Automatic learning of fine operating rules for online power system security control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1708–1719, Aug. 2016.
[9] T. E. Huang, Q. L. Guo, and H. B. Sun, "A distributed computing platform supporting power system security knowledge discovery based on online simulation," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1513–1524, May 2017.
[10] C. X. Liu, K. Sun, Z. H. Rather, Z. Chen, C. L. Bak, P. Thøgersen, and P. Lund, "A systematic approach for dynamic security assessment and the corresponding preventive control scheme based on decision trees," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 717–730, Mar. 2014.
[11] F. Tian, X. Zhang, Z. H. Yu, W. J. Qiu, D. Y. Shi, J. Qiu, M. Liu, Y. L. Li, and X. X. Zhou, "Online decision-making and control of power system stability based on super-real-time simulation," *CSEE Journal of Power and Energy Systems*, vol. 2, no. 1, pp. 95–103, Mar. 2016.
[12] S. Abhyankar, G. C. Geng, M. Anitescu, X, Y. Wang, and V. Dinavahi, "Solution techniques for transient stability-constrained optimal power flow–part I," *IET Generation, Transmission & Distribution*, vol. 11, no. 12, pp. 3177–3185, Aug. 2017.
[13] G. C. Geng, S. Abhyankar, X. Y. Wang, and V. Dinavahi, "Solution techniques for transient stability-constrained optimal power flow–part II," *IET Generation, Transmission & Distribution*, vol. 11, no. 12, pp. 3186–3193, Aug. 2017.
[14] A. Shamisa and M. Karrari, "Model free graphical index for transient stability limit based on on-line single machine equivalent system identification," *IET Generation, Transmission & Distribution*, vol. 11, no. 2, pp. 314–321, Jan. 2017.
[15] Y. Yamada, M. Nagata, and K. Tanaka, "An energy function based contingency screening method for ATC assessment with transient stability constraints," in *IEEE/PES Transmission and Distribution Conference and Exhibition*, Yokohama, Japan, Oct. 6–10, 2002, pp. 886–890.
[16] K. H. Kim, S. N. Kim, S. B. Rhee, S. K. Lee, and K. B. Song, "Assessment of total transfer capability subject to transient stability energy margin," in Proceedings of 2009 *Transmission & Distribution Conference & Exposition: Asia and Pacific*, Seoul, South Korea, Oct. 26–30, 2009, pp. 1–4.
[17] V. Vittal, E. Z. Zhou, C. Hwang, and A. A. Fouad, "Derivation of stability limits using analytical sensitivity of the transient energy margin," *IEEE Transactions on Power Systems*, vol. 4, no. 4, pp. 1363–1372, Nov. 1989.
[18] G. J. Hou and V. Vittal, "Determination of transient stability constrained interface real power flow limit using trajectory sensitivity approach," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2156–2163, Aug. 2013.
[19] T. J. Liu, Y. B. Liu, J. Y. Y. Liu, Y. Yang, G. A. Taylor, and Z. W. Huang, "Multi-indicator inference scheme for fuzzy assessment of power system transient stability," *CSEE Journal of Power and Energy Systems*, vol. 2, no. 3, pp. 1–9, Sep. 2016.
[20] 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 Transactions on Power Systems*, vol. 27, no. 3, pp. 1253–1263, Aug. 2012.
[21] C. Zheng, V. Malbasa, and M. Kezunovic, "Regression tree for stability margin prediction using synchrophasor measurements," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1978–1987, May 2013.
[22] Y. Yang, Y. Huang, J. Y. Liu, Y. B. Liu, T. J. Liu, and Y. Xiang, "Measurement-based Cell-DT method for power system transient stability classification," *CSEE Journal of Power and Energy Systems*, vol. 3, no. 3, pp. 278–285, Sep. 2017.
[23] T. Wang, T. S. Bi, H. F. Wang, and J. Z. Liu, "Decision tree based online stability assessment scheme for power systems with renewable generations," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 2, pp. 53–61, Jun. 2015.
[24] T. M. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.
[25] J. de Leeuw and S. Pruzansky, "A new computational method to fit the weighted Euclidean distance model," *Psychometrika*, vol. 43, no. 4, pp. 479–490, Dec. 1978.
[26] L. Candillier, F. Meyer, and F.çessant, "Designing specific weighted similarity measures to improve collaborative filtering systems," in *Proceedings of the 8th Industrial Conference on Advances in Data Mining: Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, Leipzig, Germany, Jul. 16-18, 2008.
[27] C. M. Bishop, "Probabilistic Generative Models," in *Pattern Recognition and Machine Learning*. New York: Springer, 2006, ch. *4*, sec. *2*, pp. *214–215.*
[28] S. Klein, M. Staring, and J. P. W. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2879–2890, Dec. 2007.
[29] M. Abadi, P. Barham, J. M. Chen, Z. F. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Q. Zheng, "Tensorflow: a system for large-scale machine learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*, Savannah, USA, Nov. 2–4, 2016, pp. 265–283.
[30] P. N. Tan, M. Steinbach, A. Karpatne, and V. Kumar, *Introduction to Data Mining*, 2nd ed., Pearson: Addison-Wesley, 2018.
[31] M. Norouzi and D. J. Fleet, "Cartesian K-means," in *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, Portland, USA, Jun. 23–28, 2013, pp. 3017–3024.
[32] Y. Wang, Q. X. Chen, C. Q. Kang, and Q. Xia, "Clustering of electricity consumption behavior dynamics toward big data applications," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2437–2447, Sep. 2016.
[33] X. Z. Liu, W. Hu, L. Zheng, Y. Min, X. L. Xu, Y. Li, and R. Yu, "Operation pattern recognition via Mass data in bulk transmission grid," in *Proceedings of the 2017 IEEE Power & Energy Society General Meeting*, Chicago, USA, Jul. 16–20, 2017, pp. 1–5.

Xianzhuang Liu received his B.S. degree in Electrical Engineering from Tsinghua University, Beijing, China, in 2014, where he is currently pursuing his Ph.D. degree. His research interests include transient stability analysis and data-driven techniques in power systems.

Yong Min received his B.S. and Ph.D. degrees in Electrical Engineering from Tsinghua University, Beijing, China, in 1984 and 1990, respectively. Currently, he is working as a professor with the Department of Electrical Engineering, Tsinghua University. His research interests include integrated energy system operation, power system stability and control.

Lei Chen received his B.S. and Ph.D. degrees in Electrical Engineering from Tsinghua University, Beijing, China, in 2003 and 2008, respectively. Since 2008, he has been with the Department of Electrical Engineering, Tsinghua University, where he is now an Associate Professor. His research interests include dynamic analysis and control of power systems.

Xiaohua Zhang is a Professor Level Senior Engineer. He is currently working at the State Grid Jibei Electric power Company. His research interests include power system operation and control.

Changyou Feng received his Ph.D. degree from Xi'an Jiaotong University, Xi'an, China, in 2010. He is currently working at the National Power Dispatching & Control Center. His research interests include power system operation and control.

Wei Hu received his B.S. and Ph.D. degrees in Electrical Engineering from Tsinghua University, Beijing, China, in 1998 and 2002, respectively. Currently, he is working as an associate professor with the Department of Electrical Engineering, Tsinghua University. His research interests include integrated energy system operation, deep learning in power system modeling and security analysis, and smart control.