Electrical Power and Energy Systems 64 (2015) 664–670

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Feature selection for transient stability assessment based on kernelized fuzzy rough sets and memetic algorithm

Xueping Gu [a,*], Yang Li [b,*], Jinghua Jia [c]

[a] School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China
[b] School of Electrical Engineering, Northeast Dianli University, Jilin 132012, China
[c] Hebei Power Dispatch and Communication Center, Shijiazhuang 050021, China

ARTICLE INFO

Article history:
Received 22 August 2013
Received in revised form 19 July 2014
Accepted 23 July 2014

Keywords:
Transient stability assessment
Feature selection
Kernelized fuzzy rough sets
Memetic algorithm

ABSTRACT

A new feature selection method based on kernelized fuzzy rough sets (KFRS) and the memetic algorithm (MA) is proposed for transient stability assessment of power systems. Considering the possible real-time information provided by wide-area measurement systems, a group of system-level classification features are extracted from the power system operation parameters to build the original feature set. By defining a KFRS-based generalized classification function as the separability criterion, the memetic algorithm based on binary differential evolution (BDE) and Tabu search (TS) is employed to obtain the optimal feature subsets with the maximized classification capability. The proposed method may avoid the information loss caused by the feature discretization process of the rough-set based attribute selection, and comprehensively utilize the advantages of BDE and TS to improve the solution quality and search efficiency. The effectiveness of the proposed method is validated by the application results on the New England 39-bus power system and the southern power system of Hebei province.

Introduction

Transient stability is concerned with the ability of a power system to maintain synchronism when subjected to a severe disturbance, such as a short circuit on a transmission line [1]. Transient stability assessment (TSA) has been recognized as an important issue to ensure the secure and economical operation of power systems [2]. TSA may serve to check the operation mode arrangement of a power system as a beforehand analysis tool in a dynamic security assessment framework, and to trigger the emergency controls as a real-time stability prediction tool after faults. The both applications can effectively reduce the possibility and amount of load loss from transient instability and then improve the operational security and economical efficiency of the power system. Problems arising from introduction of new power market designs and growing presence of intermittent renewable power generation are nudging power systems toward potential dynamic instability scenarios. The traditional methods for transient stability analysis, such as time-domain simulation methods [2], transient energy function methods [3] and the extended equal-area criterion [4], cannot well meet the requirements of online TSA for modern complex power systems. With the rapid development of computational intelligence such as decision trees (DT), artificial neural networks (ANN), and support vector machines (SVM), the pattern recognition-based TSA (PRTSA) methods have shown much potential for on-line application to power systems [5–13].

In the previous work of PRTSA [5,7,8,12], much attention has been given to design of classifiers and their parameter tuning, and relatively less attention to the feature selection issue. From the pattern recognition principles, it is well-known that the excessive input features will induce heavy computational burden, reduce the accuracy of training models and even lead to the "curse of dimensionality" [14]. Meanwhile, for transient stability classification, the classification accuracy is in fact determined by separability of the input space created by the selected features [15]. Therefore, the study of feature selection is an issue with paramount importance for PRTSA.

Some useful explorations have been carried out on the feature selection of PRTSA [14–17]. Fisher's linear discriminant function is used to select neural network training features for power system security assessment in [14], but the effectiveness of the proposed method cannot be theoretically guaranteed. In [15], a separability index as the classification criterion is defined through finding the 'inconsistent cases' in the sample set, and the breadth-first searching technique is employed to find the minimal or optimal subsets of the initial feature set as the ANN input. In [16], three dimensionality reduction methods, including sensitivity index, sensitivity analysis

* Corresponding authors.
E-mail addresses: xpgu@ncepu.edu.cn (X. Gu), liyangnedu@gmail.com (Y. Li).

http://dx.doi.org/10.1016/j.ijepes.2014.07.070
0142-0615/

and principal component analysis, are used to reduce the input space dimension for ANN-based TSA. However, both the proposed methods in [15,16] are only tested on a small test power system, and their effectiveness on practical complex power systems needs to be further validated because of the heavy computation burden. In [17], correlation analysis and principle component analysis are used as feature reduction techniques to reduce the number of the input features, but the used original input features are single-machine features, rather than system-level features, which are unsuitable for stability analysis of large-scale power systems. In feature selection, there are two key problems: feature evaluation metric and search strategies. The rough set theory (RST)-based separability criterion can be used as an effective feature evaluation index [18]. However, as most datasets contain real-valued features, it is necessary to perform feature discretization beforehand when using RST, which will inevitably cause quantization error and information loss problem [19]. As to the optimal feature subsets, the current search strategies, such as the sequence of feature selection techniques, TS and breadth-first search methods, have disadvantages of low efficiency and/or local optimum trapping.

The kernelized fuzzy rough set (KFRS) is an effective tool in dealing with uncertainty in data analysis [20], which combines the advantages of both kernel methods and RST. The Memetic algorithm (MA) is a stochastic optimization algorithm based on the imitation of cultural evolution [21], which has been successfully applied to solve many complex optimization problems [22,23]. In this paper, the MA is combined with the KFRS to be used for feature selection of PRTSA.

In recent years, wide-area measurement systems (WAMS) make it possible to obtain the synchronized real-time state information, and this brings new ideas and opportunities to transient stability assessment and prediction [24–26].

In view of the current status of the PRTSA feature selection, a new feature selection method based on KFRS and MA is proposed for real-time transient stability prediction in this paper. Considering the possible real-time information provided by WAMS, a group of system-level classification features are extracted from the power system operation parameters to build the original feature set. By defining a KFRS-based generalized classification function as the separability criterion, the memetic algorithm based on binary differential evolution (BDE) and Tabu search (TS) is then employed to obtain the optimal feature subsets with the maximized classification capability. The proposed method is verified by the numerical results on the New England 39-bus power system and the southern power system of Hebei province.

KFRS and the class separability criterion

KFRS

The main idea of KFRS is as follows: kernel functions are employed to compute the fuzzy T-equivalence relations between samples, thus generating fuzzy information granules in the approximation space; subsequently fuzzy granules are used to approximate the class demarcation based on the concepts of fuzzy lower and upper approximations, and build a kernelized model of fuzzy rough sets [20].

A classification task can typically be formulated as $< U, A, D >$, where U is the nonempty and finite set of samples, A is the set of features characterizing the classification, D is the class attribute which divides the samples into subset $\{d_1, d_2, \ldots, d_m\}$.

Given an arbitrary subset of features $B \subseteq A$ and $B \neq \varnothing$, a fuzzy T-equivalence relation R over U can be generated, where $\forall x, y, z \in U$, $R(x,x) = 1$; $R(x,y) = R(y,x)$ and $T(R(x,y), R(y,z)) \leqslant R(x,z)$, T is a triangular norm. The fuzzy information granules induced by relation R and x_i, denoted by $FIG_R(x_i)$, is defined as

$$FIG_R(x_i) = r_{1i}/x_1 + r_{2i}/x_2 + \cdots r_{ji}/x_j + \cdots + r_{ni}/x_n \tag{1}$$

where r_{ji} is the similarity degree of samples x_i and x_j. According to the definitions of lower and upper approximations, the memberships of a sample x to lower and upper approximations of the class d_i are computed by

$$\begin{cases} \underline{R_S} d_i(x) = \inf_{y \in U} S(1 - R(x,y), d_i(y)) \\ \overline{R_T} d_i(x) = \sup_{y \in U} T(R(x,y), d_i(y)) \end{cases} \quad or \tag{2}$$

$$\begin{cases} \underline{R_\theta} d_i(x) = \inf_{y \in U} \theta(R(x,y), d_i(y)) \\ \overline{R_\sigma} d_i(x) = \sup_{y \in U} \sigma(N(R(x,y)), d_i(y)) \end{cases} \tag{3}$$

where $\underline{R_S} d_i(x)$ and $\underline{R_\theta} d_i(x)$ are the degrees of certainty of the sample x belonging to decision d_i, whilst $\overline{R_T} d_i(x)$ and $\overline{R_\sigma} d_i(x)$ are the degrees of possibility of the sample x belonging to decision d_i.

In Theorem 1, Moser showed that part of kernel functions can be introduced to get fuzzy T-equivalence relations.

Definition 1 [27]. Give a nonempty and finite set U, a real-valued function k: $U \times U \to R$ is said to be a kernel if it is symmetric, that is, $k(x,y) = k(y,x)$ for all $\forall\ x, y \in U$, and positive-semidefinite.

Theorem 1 [28]. Any kernel k: $U \times U \to [0,1]$ with $k(x,x) = 1$ is (at least) $T_{\cos}$-transitive, where $T_{\cos}(a,b) = \max(ab - \sqrt{1-a^2}\sqrt{1-b^2}, 0)$.

Obviously, the Gaussian kernel $k(x,y) = \exp(-\|x-y\|^2/\delta)$ satisfies the above conditions, where δ is the width of the Gaussian. Therefore the relations computed with Gaussian kernel are fuzzy T-equivalence relations between samples. Then the formulae for computing the memberships of lower and upper approximations can be obtained by

$$\begin{cases} \underline{k_S} d_i(x) = \inf_{y \notin d_i} (1 - k(x,y)) \\ \underline{k_\theta} d_i(x) = \inf_{y \notin d_i} \left(\sqrt{1 - k^2(x,y)} \right) \\ \overline{k_T} d_i(x) = \sup_{y \in d_i} k(x,y) \\ \overline{k_\sigma} d_i(x) = \sup_{y \in d_i} \left(1 - \sqrt{1 - k^2(x,y)} \right) \end{cases} \tag{4}$$

$\underline{k_S} d_i(x)$ and $\underline{k_\theta} d_i(x)$ are the degrees the sample x certainly belongs to class d_i, while $\overline{k_T} d_i(x)$ and $\overline{k_\sigma} d_i(x)$ are the degrees this sample x possibly belongs to class d_i.

Class separability criterion

In order to enhance the robustness of classification index, N_k nearest neighbors of each sample from each different class and from the same class are comprehensively considered. Given $< U, A, D >$, a KFRS-based generalized classification function $gc(D)$ is used as the class separability criterion in this paper.

$$gc(D) = [g\gamma_B^\theta(D) + g\omega_B^{\theta-\sigma}(D)]/2 \tag{5}$$

where $g\gamma_B^\theta(D)$ and $g\omega_B^{\theta-\sigma}(D)$ are respectively the generalized dependency function and generalized classification certainty function with N_k = 3, as given by:

$$g\gamma_B^\theta(D) = \frac{1}{(I-1)N_k|U|} \sum_{x_i \in U} \sum_{d \in D, d \neq d_i} \sum_{y \in H_d^i} \sqrt{1 - k(x,y)^2}$$

$$g\omega_B^{\theta-\sigma}(D) = \frac{1}{(I-1)N_k|U|} \sum_{x_i} \left\{ \sum_{H^i} \sqrt{1 - k(x_i, H_j^i)^2} \right.$$
$$\left. - - \sum_l \sum_m \left[1 - \sqrt{1 - k(x_i, M_{lm}^i)^2} \right] \right\},$$