IJAMML 2:2 (2015) 159-175 June 2015
ISSN: 2394-2258
Available at http://scientificadvances.co.in

FEATURE SELECTION BY KERNELIZED FUZZY ROUGH SETS FOR TRANSIENT STABILITY ASSESSMENT BASED ON GAUSSIAN PROCESS

Yang Li and Changjiang Wang

School of Electrical Engineering, Northeast Dianli University, Jilin, P. R. China

Abstract

Feature selection of input features is the key issue for pattern recognition-based transient stability assessment (TSA) methods. Considering the possible real-time information provided by phasor measurement units, a group of system-level classification features are firstly extracted from the power system operation condition to construct the original feature set. Then kernelized fuzzy rough sets (KFRS) are used to select the near-optimal feature subset, and Gaussian process is finally employed to test the classification ability of the selected features. The effectiveness of the proposed method is validated by the simulation results on the New England 39-bus test system.

Keywords: transient stability assessment, feature selection, fuzzy rough sets, Gaussian process, phasor measurement unit.

*Corresponding author.

E-mail address: liyang@nedu.edu.cn (Yang Li).

This work was supported in part by the Doctor Scientific Research Foundation of Northeast Dianli University (BSJXM-201407).

2010 Mathematics Subject Classification: 93Dxx.
Submitted by Jianqiang Gao.
Received April 6, 2015

1. Introduction

TSA has been recognized as an important task to ensure the secure and economical operation of power systems [1]. With the rapid development of computational intelligence, recent research shows that pattern recognition-based TSA (PRTSA) methods are promising for online TSA of power systems [2]-[7]. One of the most important considerations in PRTSA is the proper selection of training features [8]. It is well-known that the excessive input features will increase the costs of knowledge discovery, reduce the accuracy of training models and even lead to the well-known “curse of dimensionality” problem. Meanwhile, high dimension of power system is an important problem in both theory research and engineering practice. Therefore, the feature selection of input features for TSA has very important theoretical and practical significance.

However, large amounts of published research work on is devoted to classifiers design, and there is relatively less attention given the issue up to now. Reference [8] uses Fisher’s linear discriminant function to select neural network training features for power system security assessment. A separability index as criterion is defined through finding the “inconsistent cases” in a sample set, and the breadth-first searching technique is employed to find the minimal or optimal subsets of the initial feature set in [9]. Reference [10] proposes an artificial neural network-based TSA method to predict the stability status of the power system, and uses two feature selection techniques to identify the input variables best suitable for training. Recent research shows that rough set (RS) methods are effective ways for feature selection [11]. But when processing numeric data, the discrete process of RS will inevitably leads to information loss, which greatly limits the application of RS for TSA [12].

KFRS is a new algorithm in dealing with uncertainty in data analysis, which combines the advantages of kernel methods and rough sets [13]. Gaussian process (GP) is a Bayesian probabilistic kernel machine [14], [15], which is widely used for the high-dimensional nonlinear classification and regression problems [16], [17].

In recent years, the advent of wide area measurement system (WAMS) using time-stamped phasor measurement units (PMUs) makes it possible to explore wide area protection and control schemes to avoid the system collapse [18]-[20]. Meanwhile, WAMS provides new rich data source for the input features of PRTSA.

Considering the possible real-time information provided by PMUs, a new feature selection method for GP-based TSA of power systems using KFRS is presented in this paper. The proposed method can overcome the information loss problem of the discrete process, when processing numeric data by using RS.

The remainder of this paper is organized as follows. First the KFRS theory is introduced in brief. Details of the proposed feature evaluation and selection method using KFRS are presented next. Then, GP is used to construct a TSA model to validate the selected features. Application of the proposed method is demonstrated by using the New England 39-bus test system, and finally the conclusions are made.

2. Brief Introduction to KFRS

Given a classification task $< U, A, D >$, where $U = \{x_1, x_2, \ldots, x_n\}$ is a set of samples described by attribute set $A = \{a_1, a_2, \ldots, a_N\}$, D is the decision attribute, which divides the set of samples into subsets $d_1, d_2, \cdots, d_m$.

Given arbitrary subset of attributes $B \subseteq A$ and $B \neq \varnothing$, we can generate a fuzzy T-equivalence relation R over U, where $\forall x, y, z \in U$,

$R(x, x) = 1$; $R(x, y) = R(y, x)$, and $T(R(x, y), R(y, z)) \leq R(x, z)$. 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. As to classical classification, we are confronted a task of approximating decision classes with these fuzzy information granules. According to the definitions of lower and upper approximations, the memberships of sample x to lower and upper approximations of decision class d_k are computed as

$$\begin{cases} \underline{R_S} d_k(x) = \inf\limits_{y \in U} S(1 - R(x, y), d_k(y)), \\ \overline{R_T} d_k(x) = \sup\limits_{y \in U} T(R(x, y), d_k(y)), \end{cases} \tag{2}$$

$$\begin{cases} \underline{R_\theta} d_k(x) = \inf\limits_{y \in U} \theta(R(x, y), d_k(y)), \\ \overline{R_\sigma} d_k(x) = \sup\limits_{y \in U} \sigma(N(R(x, y)), d_k(y)), \end{cases} \tag{3}$$

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

Before computing $\underline{R_S} d_k(x)$, $\underline{R_\theta} d_k(x)$, $\overline{R_T} d_k(x)$, and $\overline{R_\sigma} d_k(x)$, we should introduce an algorithm to obtain fuzzy T-equivalence relations between samples. In Theorem 2.1, Moser showed that part of kernel functions can be introduced to get fuzzy T-equivalence relations.

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 2.1 ([21]). *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 relations computed with Gaussian kernel

$$G(x, y) = \exp\left(- \frac{\|x - y\|^2}{\delta} \right) \tag{4}$$

are fuzzy $T_{\cos}$-transitive relations, where δ is the width of the Gaussian.

Then the formulae for computing the memberships of lower and upper approximations can be obtained.

$$\underline{R_S} d_i(x) = \inf_{y \notin d_i} (1 - G(x, y)), \tag{5}$$

$$\underline{R_\theta} d_i(x) = \inf_{y \notin d_i} \left(\sqrt{1 - G^2(x, y)}\right), \tag{6}$$

$$\overline{R_T} d_i(x) = \sup_{y \in d_i} G(x, y), \tag{7}$$

$$\overline{R_\sigma} d_i(x) = \sup_{y \in d_i} \left(1 - \sqrt{G^2(x, y)}\right), \tag{8}$$

$\underline{R_S} d_i(x)$ or $\underline{R_\theta} d_i(x)$, the membership of sample x to the fuzzy lower approximation of d_i, reflects the degree how a sample certainly belongs to decision label d_i with respect to fuzzy relation R, and $\overline{R_T} d_i(x)$ or $\overline{R_\sigma} d_i(x)$, the membership of sample x to the fuzzy upper approximation of d_i, gives the degree how a sample possibly belongs to decision label d_i with respect to fuzzy relation R. Clearly, it is expected that each sample certainly belongs to its decision class with a great degree. So, the average value of the memberships of lower approximation is usually used to evaluate features. KFRS is described in detail in [13].

3. Construction of Original Feature Set

It is an important task for PRTSA to construct effective original features. Unfortunately, previous works have mainly focused on the analysis of pre-fault static features, as the traditional monitoring systems such as SCADA does not provide synchronized measurements for wide-area power systems. The bottleneck is break through with the advent of WAMS, which provide rich data sources by the availability of real-time synchronized measurements. Therefore, this paper focuses on applications of the real-time information obtained from PMUs to construct original features.

After having studied the comprehensive existing literature and having carried out a lot of the simulation analysis, a group of system-level classification features independent of the scale of power systems were initially selected as the original features. These feathers are listed in Table 1, where t_{c1} denotes the fault clearing time, $t_{c1+3c,}$ t_{c1+6c}, and t_{c1+9c}, respectively, denote the 3rd cycle, 6th and 9th after the fault clearing time.

Table 1. Input features of data set

No.	Input features
Tz1	Mean value of all the mechanical power before the fault incipient time.
Tz2	Mean value of all the initial rotor acceleration rates.
Tz3	Mean square error of all the initial acceleration rates.
Tz4	Maximum value of all the initial active power impact.
Tz5	Minimum value of all the initial active power impact.
Tz6	Mean value of all the initial acceleration power.
Tz7	Maximum value of all the initial rotor kinetic energies.
Tz8	Maximum value of the difference of initial acceleration rates.
Tz9	Maximum value of the difference of initial rotor kinetic energies.
Tz10	Maximum value of the difference of initial rotor angle.
Tz11	Initial rotor angle of the machine with the maximum acceleration rate.
Tz12	Maximum value of all the initial rotor acceleration rates.
Tz13	Minimum value of all the initial rotor acceleration rates.
Tz14	Total system 'energy adjustment'.
Tz15	Value of system impact at t_{c1}.
Tz16	Maximum value of the difference of acceleration rates at t_{c1}.
Tz17	Maximum value of the difference of rotor kinetic energies at t_{c1}.
Tz18	Maximum value of the difference of rotor angle at t_{c1}.
Tz19	Mean value of all the rotor kinetic energies at t_{c1}.
Tz20	Rotor angle of the machine with the maximum kinetic energy at t_{c1}.
Tz21	Kinetic energy of the machine with the maximum rotor angle at t_{c1}.
Tz22	Maximum value of all the rotor kinetic energies at t_{c1}.
Tz23	Maximum value of all the rotor kinetic energies at t_{c1+3c}.
Tz24	Maximum value of all the rotor kinetic energies at t_{c1+6c}.
Tz25	Maximum value of all the rotor kinetic energies at t_{c1+9c}.

Table 1. (Continued)

Tz26	Kinetic energy of the machine with the maximum rotor angle at t_{c1+3c}.
Tz27	Kinetic energy of the machine with the maximum rotor angle at t_{c1+6c}.
Tz28	Kinetic energy of the machine with the maximum rotor angle at t_{c1+9c}.
Tz29	Maximum value of the difference of all rotor angles at t_{c1+3c}.
Tz30	Maximum value of the difference of all rotor angles at t_{c1+6c}.
Tz31	Maximum value of the difference of all rotor angles at t_{c1+9c}.

4. KFRS-Based Feature Selection

4.1. Generation of sample set

In this paper, data required for training and testing the classifier were generated offline through the T-D simulations. The simulation was done based on the classical machine model and the constant impedance load model. A successful reclosure of the faulted line was applied after fault clearance, and no topology changes result from the fault.

A total of 500 simulation cases at 20 different fault locations were generated at 5 different loading levels (under 80, 90, 100, 110, and 120% of the base load). Corresponding to each loading level, 5 kinds of active and reactive load powers were randomly set. The contingencies considered were three-phase short-circuit faults. A standard clearing time of five cycles was assumed for all the contingencies. 352 out of 500 generated operating points are randomly sampled as the training data set, and the remaining as the testing data set. A class label was assigned to denote the transient unstable and stable status of a simulation case following a contingency. This class label is calculated according to maximum relative rotor angle deviation during the transient period. If the maximum relative rotor angle deviation exceeds 360 degree [6], the system is considered as unstable and the class label is marked as "−1", otherwise the class label is marked as "+1".

4.2. Data pre-processing

In this paper, z-score standardization method is used as the data pre-processing method.

$$d' = (d - \overline{D}) / \sigma_D, \tag{9}$$

where $\overline{D}$ and σ_D are, respectively, the mean and standard deviation of any feature D in sample data. d' is the normalized value of d, $d \in D$.

4.3. Feature selection algorithm

Given $\langle U, A, D \rangle$, $B \subseteq A$ and R is a fuzzy relation between samples induced by attributes B. Then the dependency of D on B is defined as

$$\gamma_B^S(D) = \left| \bigcup_{i=1}^{m} \underline{R_S} d_i \right| \Big/ |U| \quad \text{or} \quad \gamma_B^\theta(D) = \left| \bigcup_{i=1}^{m} \underline{R_\theta} d_i \right| \Big/ |U|, \tag{10}$$

where $|\cdot|$ is the cardinality of a set. As to fuzzy set $\bigcup_{i=1}^{m} \underline{R_S} d_i$ and $\bigcup_{i=1}^{m} \underline{R_\theta} d_i$, $\left| \bigcup_{i=1}^{m} \underline{R_S} d_i \right| = \sum_{i=1}^{m} \sum_{x \in d_i} \underline{R_S} d_i(x)$ and $\left| \bigcup_{i=1}^{m} \underline{R_\theta} d_i \right| = \sum_{i=1}^{m} \sum_{x \in d_i} \underline{R_\theta} d_i(x)$.

Dependency is the average value of the memberships of each sample to the lower approximation. This coefficient reflects the approximating ability of attribute subset B to characterize the decision. $\gamma_B^S(D) = 1$ if for $\forall x \in U$, $\underline{R_S} d(x) = 1$, where $d \in \{d_1, d_2, \cdots, d_m\}$. That is to say, all the samples consistently belong to one of the decision classes without any uncertainty. Obviously, B is a good subset of attributes for discerning different decision classes in this case.

There is a good property with the dependency function. It is monotonous with the attributes. That is to say, $\gamma_{B_1}^S(D) \le \gamma_{B_2}^S(D)$ and $\gamma_{B_1}^\theta(D) \le \gamma_{B_2}^\theta(D)$ if $B_1 \subseteq B_2$. Moreover, $0 \le \gamma_B^S(D) \le 1$ and $0 \le \gamma_B^\theta(D) \le 1$. So we can begin with the best feature, and then add features one by one until the dependency does not increase by adding any new feature.

With this measure, a forward algorithm for feature selection is constructed as follows:

Algorithm. Feature selection based on fuzzy rough sets.

Input: $\langle U, A, D\rangle$ and a similarity function K.

Output: Feature subset F

(1) $\varnothing \to F$;

(2) **while** $A - F \neq \varnothing$ **do**

(3) **for each** $\alpha_i \in A - F$

(4) compute $\gamma_{F \cup \alpha_i}(D)$

(5) **end**

(6) select $\alpha \in A - F$ subject to

$$\gamma_{F \cup \alpha}(D) = \max_i \gamma_{F \cup \alpha_i}(D)$$

(7) **if** $\gamma_{F \cup \alpha}(D) - \gamma_F(D) \leq \varepsilon$

(8) Break while

(9) **end**

(10) $F \cup \alpha \to F$

(11) **end**

(12) **return** F.

In this work, Gaussian function is used to compute the similarity between samples. In addition, the parameter ε is set to 0.01 to stop the search in this algorithm.

In this algorithm, given a set of features, we can compute the similarity relations between samples. Then we can obtain the memberships of samples to the fuzzy lower approximation.

4.4. Gaussian process

In this section, the Gaussian process model for binary classification (GPC) is briefly described. Given data points $\mathbf{x}_i$ from a domain χ with corresponding class labels $y_i \in \{-1, +1\}$, one would like to predict the class membership probability for a test point $\mathbf{x}_*$. This is achieved by using a latent function f whose value is mapped into the unit interval by means of a sigmoid function $\mathrm{sig}: R \to [0, 1]$ such that the class membership probability $p(y = +1|\mathbf{x})$ can be written as $\mathrm{sig}(f(\mathbf{x}))$. The class membership probability must normalize $\sum_y p(y|\mathbf{x}) = 1$, which leads to $p(y = +1|\mathbf{x}) = 1 - p(y = -1|\mathbf{x})$. If the sigmoid function satisfies the point symmetry condition $\mathrm{sig}(t) = 1 - \mathrm{sig}(-t)$, the likelihood can be compactly written as

$$p(y|\mathbf{x}) = \mathrm{sig}(y \cdot f(\mathbf{x})).$$

Given the latent function f, the class labels are assumed to be Bernoulli distributed and independent random variables, which gives rise to a factorial likelihood, factorizing over data points.

$$p(\boldsymbol{y}|f) = p(\boldsymbol{y}|\mathbf{f}) = \prod_{i=1}^{n} \mathrm{sig}(y_i f_i). \tag{11}$$

The prior distribution of the latent function is

$$p(\mathbf{f}|\mathbf{X}, \theta) = N(\mathbf{f}|\mathbf{m}_0, \mathbf{K}), \tag{12}$$

where $\mathbf{m}_0$, $\mathbf{K}$, and θ are, respectively, mean vector, covariance matrix, and hyperparameter vector. For notational convenience, we will assume $m(x) \equiv 0$ throughout. Thus, the elements of $\mathbf{K}$ are $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j, \theta)$, where $\mathbf{x}_i, \mathbf{x}_j \in \chi$.

By application of Bayes' rule, one gets an expression for the posterior distribution over the latent values **f**

$$p(\mathbf{f}|\mathbf{y}, \mathbf{X}, \theta) = \frac{N(\mathbf{f}|0, \mathbf{K})}{p(\mathbf{y}|\mathbf{X}, \theta)} \prod_{i=1}^{n} \mathrm{sig}(y_i f_i). \quad (13)$$

When making predictions, we marginalize over the training set latent variables

$$p(\mathbf{f}_*|\mathbf{X}_*, \mathbf{y}, \mathbf{X}, \theta) = \int p(\mathbf{f}_*|\mathbf{f}, \mathbf{X}_*, \mathbf{X}, \theta) p(\mathbf{f}|\mathbf{y}, \mathbf{X}, \theta) df. \quad (14)$$

Finally, the predictive class membership probability p_* is obtained by averaging out the test set latent variables

$$p(y_*|\mathbf{x}_*, \mathbf{y}, \mathbf{X}, \theta) = \int \mathrm{sig}(y_* f_*) p(f_*|\mathbf{x}_*, \mathbf{y}, \mathbf{X}, \theta) df_*. \quad (15)$$

A covariance function is the crucial ingredient affecting the performance of GPC. Among the common covariance functions are the squared exponential covariance function, rational quadratic covariance function and the γ-exponential covariance function. The covariance function used in this paper is the squared exponential covariance function:

$$k_{SE_{ISO}}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp(-\frac{(\mathbf{x}_i - \mathbf{x}_j)^2}{2l^2}), \quad (16)$$

where σ_f^2 and l are all hyperparameters.

5. Case Study

In order to evaluate the performance of the proposed method, the IEEE 39-bus test system (New England) was used. This system is a well-known test case for TSA studies reported in previous works [5], [6]. The one diagram of the test system is shown in Figure 1.

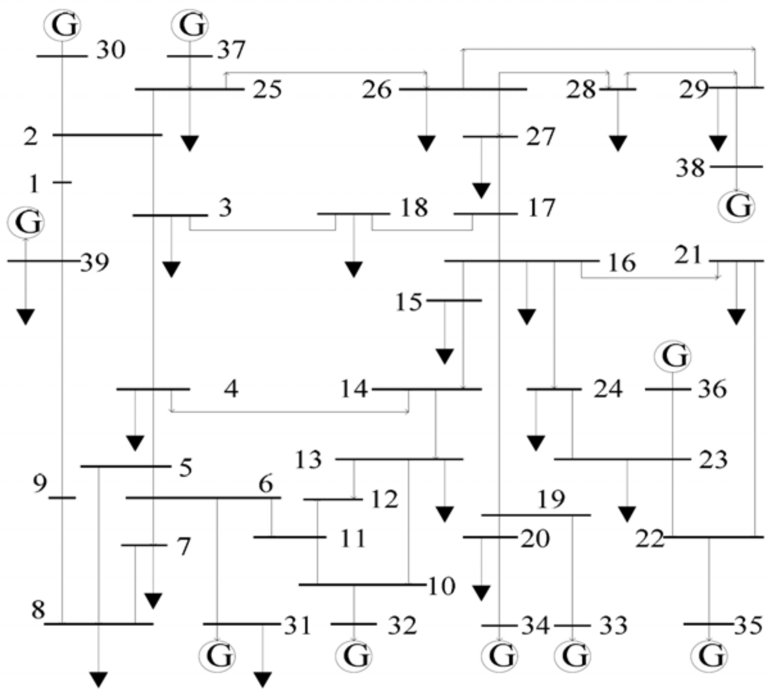

Figure 1. New England 39-bus test system.

All of the programs in this paper are implemented in MATLAB running on a PC with Microsoft Windows Server 2003 operating system, Intel Pentium dual CPU E2200 @ 2.20GHz, 2.19GHz and main memory 1GB.

5.1. Test results of the proposed method

In order to compare the proposed KFRS-based feature selection method with other algorithms, fast correlation based feature search (FCBF) [22] is introduced and tested. FCBF is a famous feature selection technique in classification analysis, where symmetric uncertainty measure was used to evaluate quality of features and a fast search algorithm was developed for high-dimensional data analysis.

The proposed feature selection method returned the feature subset $A_1 = \{Tz29, Tz1, Tz6, Tz23, Tz3, Tz19, Tz31\}$, and FCBF returned the subset of feature $A_2 = \{Tz4, Tz1, Tz28, Tz6, Tz10, Tz25, Tz23, Tz11, Tz19\}$. Then, comparison tests were carried out between the original feature set A, A_1, and A_2 by using GPC-based TSA models. The test results are shown in Table 2.

Table 2. Test result of GPC models

Feature set	Dimension	Hyperparameter		Test accuracy/%
		σ_f	l	
A	31	2.38	5.24	96.62
A_1	7	1.03	3.92	98.65
A_2	9	1.56	4.69	95.95

As is shown in Table 2, compared with the original feature set A, A_1 has similar classification accuracy, but the data dimension is reduced to $1/4$. At the same time, it can be seen that although KFRS-based algorithm selects the less features than FCBF, its classification performance is better than the latter.

5.2. Test results of other TSA models

In order to examine the versatility of the proposed feature selection method, the feature sets A and A_1 were used as the input of other TSA models such as DT, multi-layer perception (MLP) and SVM. The parameters of the models were set as follows: DT was constructed using the C4.5 algorithm with default configuration (pruning with 0.25 confidence factor); MLP adopted back-propagation (BP) algorithm as the training algorithm and the learning rate was set to 0.8; the kernel function of SVM used in this paper was RBF kernel and the associated parameters were optimized through a grid search during the 4-fold cross-validation process [6]. The test results are shown in Table 3.

Table 3. Test results of other models

Feature set	TSA model	Test accuracy/%
A	DT	94.59
	MLP	95.27
	SVM	95.95
A_1	DT	96.62
	MLP	94.59
	SVM	97.30

From Table 3, it can be observed that when using feature subsets A_1, all other TSA models have the similar classification performances to the original feature set A. This indicates that the proposed KFRS-based feature selection methods can be used for other TSA models, such as DT, MLP, and SVM.

6. Conclusions

Considering the possible real-time information provided by PMUs, a new method for transient stability assessment of power systems using KFRS and GPC is presented in this paper. The proposed method has been examined on the New England 39-bus test system, and the following conclusions can be drawn from the work:

(1) Without sacrificing classification performance of the original feature set, the proposed KFRS-based feature selection method can significantly reduce the data dimension, and has better performance than FCBF.

(2) The proposed feature selection method can also be used for other TSA models, such as DT, MLP, and SVM.

(3) The proposed method can effectively evaluate the quality of features and find the useful subsets, which may be used as a reference for future TSA research.

References

[1] P. M. Anderson and A. A. Fouad, Power System Control and Stability, 2nd Edition, Piscataway, NJ: IEEE, 2003.

[2] L. Wehenkel, M. Pavella, E. Euxibie and B. Heilbronn, Decision tree based transient stability method a case study, IEEE Trans. Power Systems 9 (1994), 459-469.

[3] L. S. Moulin, A. P. A. da Silva, M. A. El-Sharkawi and R. J. Marks II, Support vector machines for transient stability analysis of large-scale power systems, IEEE Trans. Power Systems 19 (2004), 818-825.

[4] Sun Kai, S. Likhate, V. Vittal, V. S. Kolluri and S. Mandal, An online dynamic security assessment scheme using phasor measurements and decision trees, IEEE Transactions on Power Systems 22 (2007), 1935-1943.

[5] N. Amjady and S. F. Majedi, Transient stability prediction by a hybrid intelligent system, IEEE Trans. Power Systems 22 (2007), 1275-1283.

[6] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage and I. T. Fernando, Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements, IEEE Trans. Power Systems 26 (2011), 1474-1483.

[7] 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 Systems 27 (2012), 1253-1263.

[8] C. A. Jensen, M. A. El-Sharkawi and R. J. Marks, Power system security assessment using neural networks: Feature selection using Fisher discrimination, IEEE Trans. Power Systems 16 (2001), 757-763.

[9] S. K. Tso and X. P. Gu, Feature selection by separability assessment of input spaces for transient stability classification based on neural networks, International Journal of Electrical Power & Energy Systems 26 (2004), 153-162.

[10] Harinder Sawhney and B. Jeyasurya, A feed-forward artificial neural network with enhanced feature selection for power system transient stability assessment, Electric Power Systems Research 76(12) (2006), 1047-1054.

[11] R. W. Swiniarski and A. Skowron, Rough set methods in feature selection and recognition, Pattern Recognition Letters 24(6) (2003), 833-849.

[12] R. Jensen and Q. Shen, Fuzzy-rough sets assisted attribute selection, IEEE Trans. Fuzzy Systems 15(1) (2007), 73-89.

[13] Qinghua Hu, Daren Yu, Witold Pedrycz and Degang Chen, Kernelized fuzzy rough sets and their applications, IEEE Trans. Knowledge and Data Engineering 23 (2011), 1649-1667.

[14] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning, MIT Press, Cambridge, 2006.

[15] H. Nickisch and C. E. Rasmussen, Approximations for binary Gaussian process classification, Journal of Machine Learning Research 9 (2008), 2035-2078.

[16] Fei Cheng, Jiangsheng Yu and Huilin Xiong, Facial expression recognition in JAFFE dataset based on Gaussian process classification, IEEE Trans. Neural Networks 21 (2010), 1685-1690.

[17] Mahdi Jadaliha, Yunfei Xu, Jongeun Choi, N. S. Johnson and Weiming Li, Gaussian process regression for sensor networks under localization uncertainty, IEEE Trans. Signal Processing 61(2) (2013), 223-237.

[18] A. G. Phadke and J. S. Thorp, Synchronized Phasor Measurements and their Applications, Springer, New York, 2008.

[19] C. W. Taylor, D. C. Erickson, K. Martin, R. E. Wilson and V. Venkatasubramanian, WACS-wide-area stability and voltage control system: R&D and online demonstration, Proceedings of the IEEE 93 (2005), 892-906.

[20] V. Terzija, G. Valverds, Cai Deyu, P. Regulski, V. Madani, J. Fitch, S. Skok, M. M. Begovic and A. Phadke, Wide-area monitoring, protection, and control of future electric power networks, Proceedings of the IEEE 99 (2011), 80-93.

[21] B. Moser, On representing and generating kernels by fuzzy equivalence relations, Journal of Machine Learning Research 7 (2006), 2603-2620.

[22] L. Yu and H. Liu, Efficient feature selection via analysis of relevance and redundancy, Journal of Machine Learning Research 5 (2004), 1205-1224.

■