Transient Stability Assessment of Power Systems Based on KPCA and Gaussian Process

Yang Li, Xueping Gu
School of Electrical and Electronic Engineering
North China Electric Power University
619 Yonghua North Street, Baoding City, Hebei Province
CHINA

Abstract: - This paper presents a new method for transient stability assessment of power systems using kernel principal component analysis (KPCA) and Gaussian process (GP). Considering the possible real-time information provided by PMU, a group of system-level classification features are firstly extracted from the power system operation condition to construct the original feature set. Then KPCA is used to reduce the dimension of input space, and GP is employed to build a TSA model. Furthermore, the classification accuracy and generalization performance of the GP model are improved by combining existing single covariance functions to make new composite ones. The proposed method can overcome the disadvantages that many of the current machine learning methods usually suffer from, such as overfitting, difficulty in parameter selection and prediction with no probability interpretation. The effectiveness of the proposed method is validated by the simulation results on the New England 39-bus test system.

Key-Words: - Transient stability assessment, KPCA, Gaussian process, composite kernel function, machine learning, phasor measurement unit

1 Introduction

Transient stability assessment (TSA) has been recognized as an important task to ensure the secure and economical operation of power systems [1], [2]. Problems arising from the introduction of new power market designs and growing presence of intermittent renewable power generation are nudging power systems toward potential dynamic instability scenarios. The uncertainty of predicting future operating conditions has created an acute need to understand the dynamic nature of power systems deeply in order to be prepared for critical situations. In recent years, wide area measurement system (WAMS) using time-stamped phasor measurement units (PMU) has been receiving ever increasing attention from both the academia and the industry, which makes it possible to explore wide area protection and control schemes to avoid the system collapse [3]-[6].

Transient stability refers to the ability of synchronous machines of an interconnected power system to remain in synchronism after being subjected to a severe disturbance, such as a short circuit on a transmission line [1]. It depends on the ability to maintain/restore equilibrium between electromagnetic torque and mechanical torque of each synchronous machine in the system. The current methods to determine the stability status of a power system mainly include: time-domain (T-D) simulations method [7], transient energy function methods [8] and the extended equal area criterion [9]. With the rapid development of artificial intelligence, recent research shows that machine learning techniques, such as decision trees (DT), artificial neural networks (ANN), and support vector machines (SVM), are promising approaches for on-line TSA of power systems [10]-[16].

However, since power systems are high-dimensional nonlinear dynamical systems, the mapping between state parameters and transient stability of power systems is a complex nonlinear relationship. In addition, many of the current machine learning methods usually suffer from some inherent disadvantages, which limit the practical application of the machine learning-based TSA (MLTSA) methods in power systems. For example, ANN has problems of overfitting and local optima [12], and SVM has difficulty in parameter selection and prediction with no probability interpretation [17].

Gaussian process is a Bayesian probabilistic kernel machine [18], [19], which is very suitable for the high-dimensional nonlinear classification and regression problems [20], [21]. Gaussian process

can overcome the disadvantages mentioned above of other machine learning methods, such as difficulty in parameter selection and prediction with no probability interpretation.

Feature extraction and selection is one of the most important tasks for MLTSA. PCA is a well-known feature extraction and data representation technique widely used in the areas of pattern recognition. However, PCA is restricted to linear transformations. In order to improve the classification accuracy of TSA, we propose to use KPCA, a generalization of linear PCA to a nonlinear setting that was introduced by Schölkopf et al [22].

The remainder of this paper is organized as follows. First the GP theory is introduced in brief. Details of the proposed TSA method using KPCA and GP are presented next. Application of the proposed method is demonstrated using the New England 39-bus test system, and finally the conclusions are made.

2 Brief Introduction to GP

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 $\mathrm{sig}:\mathrm{R}\to[0,1]$ 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.

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

The prior distribution of the latent function is

$$p(\mathbf{f}|\mathbf{X},\boldsymbol{\theta})=N(\mathbf{f}|\mathbf{m}_0,\mathbf{K}) \tag{2}$$

where $\mathbf{m}_0$, $\mathbf{K}$ and $\boldsymbol{\theta}$ 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 $K_{ij}=k(\mathbf{x}_i,\mathbf{x}_j,\boldsymbol{\theta})$, 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 $\mathbf{f}$

$$p(\mathbf{f}|\mathbf{y},\mathbf{X},\boldsymbol{\theta})=\frac{N(\mathbf{f}|\mathbf{0},\mathbf{K})}{p(\mathbf{y}|\mathbf{X},\boldsymbol{\theta})}\prod_{i=1}^{n}\mathrm{sig}(y_i f_i) \tag{3}$$

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

$$p(\mathbf{f}_*|\mathbf{X}_*,\mathbf{y},\mathbf{X},\boldsymbol{\theta})=\int p(\mathbf{f}_*|\mathbf{f},\mathbf{X}_*,\mathbf{X},\boldsymbol{\theta})p(\mathbf{f}|\mathbf{y},\mathbf{X},\boldsymbol{\theta})\mathrm{d}\mathbf{f} \tag{4}$$

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

$$p(y_*|\mathbf{x}_*,\mathbf{y},\mathbf{X},\boldsymbol{\theta})=\int \mathrm{sig}(y_* f_*)p(f_*|\mathbf{x}_*,\mathbf{y},\mathbf{X},\boldsymbol{\theta})\mathrm{d}f_* \tag{5}$$

2.1 Covariance Functions

There are many covariance functions, the following single covariance functions [18] are used in this paper.

$$k_{\mathrm{SEiso}}(\mathbf{x}_i,\mathbf{x}_j)=\sigma_f^2\exp\left(-\frac{(\mathbf{x}_i-\mathbf{x}_j)^2}{2l^2}\right) \tag{6}$$

$$k_{\mathrm{RQiso}}(\mathbf{x}_i,\mathbf{x}_j)=\sigma_f^2\left(1+\frac{(\mathbf{x}_i-\mathbf{x}_j)^2}{2\alpha l^2}\right)^{-\alpha} \tag{7}$$

where σ_f^2, l and α are all hyperparameters.

In order to improve the classification accuracy and generalization performance of GP, this paper makes a new composite covariance function by combining single ones [18].

$$k_{\mathrm{CKiso}}(\mathbf{x}_i,\mathbf{x}_j)=k_{\mathrm{SEiso}}(\mathbf{x}_i,\mathbf{x}_j)+k_{\mathrm{RQiso}}(\mathbf{x}_i,\mathbf{x}_j) \tag{8}$$

2.2 Gaussian Approximation

Unfortunately, the latent distribution in (3) and the predictive distribution in (4) cannot be written as analytical expressions. To obtain exact answers, five different Gaussian approximations to the posterior are depicted in [19]. Laplace approximation is adopted in this paper, and then the posterior distribution can be expressed as

$$p(\mathbf{f}|\mathbf{y},\mathbf{X},\boldsymbol{\theta})\approx N(\mathbf{f}|\mathbf{m},(\mathbf{K}^{-1}+\mathbf{W})^{-1}) \tag{9}$$

where

$$\mathbf{m}=\arg\max_{\mathbf{f}\in\mathrm{R}^n} p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{X},\boldsymbol{\theta}) \tag{10}$$

$$\mathbf{W}=-\frac{\partial^2\ln p(\mathbf{y}|\mathbf{f})}{\partial\mathbf{f}\partial\mathbf{f}^{\mathrm{T}}}\Big|_{\mathbf{f}=\mathbf{m}}=-\left[\frac{\partial^2\ln p(y_i|f_i)}{\partial f_i^2}\Big|_{f_i=m_i}\right]_{ii} \tag{11}$$

2.3 Determination of Optimal Hyperparameters

The optimal hyperparameters $\boldsymbol{\theta}=[\sigma_f^2; l; \alpha]$ can be adaptively obtained through the maximum likelihood method, where the log marginal likelihood function is

$$\ln p(\mathbf{y}|\mathbf{X},\boldsymbol{\theta}) \approx \ln p(\mathbf{y}|\mathbf{m}) - \frac{1}{2}\mathbf{m}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{m} + \frac{1}{2}\ln|\mathbf{I}+\mathbf{KW}| \quad (12)$$

3 GP for TSA

From the point of view of pattern recognition, TSA can be treated as a two-pattern classification problem with two basic classes (i.e. the stable class and the unstable class). In order to seek the right balance between model complexity and generalization capability in the classification process, it is very important to design a reasonable classification model. Therefore, considering the possible real-time information provided by PMU, a new GP-based TSA model is presented in this paper.

3.1 Knowledge Base Generation

The application of supervised learning algorithms is based on the priori knowledge which depicts the characteristics of a system. For TSA of power systems, the knowledge base should cover sufficient operating points to determine the detailed transient stability boundary.

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 changed 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. 345 out of 500 generated operating points were 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 [16], the system is considered as unstable and the class label is marked as "-1", otherwise the class label is marked as "+1".

3.2 Selection of Input Features

Selection of the appropriate features is an important problem for MLTSA. Transient stability depends on both the initial operating state of the system and the severity of the disturbance [1]. Meanwhile, instability results if the stability cannot absorb the kinetic energy corresponding to these rotor speed difference. Thus, post-fault states of the generator rotors strongly symbolize the stability states of a power system, and the post-fault rotor variables such as acceleration rates and kinetic energies are good predictors for TSA.

Unfortunately, previous works have mainly focused on the analysis of pre-fault static features, as the traditional monitoring systems such as supervisory control and data acquisition (SCADA) are not able to get the synchronous measurement information for the 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 state information obtained from WAMS to the selection of input features for MLTSA.

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 input features. These feathers are listed in Table. 1, where t_{cl}, t_{cl+3c}, t_{cl+6c} and t_{cl+9c} respectively denote the fault clearing time, the 3rd cycle after the fault clearing time, the 6th cycle after the fault clearing time, the 9th cycle after the fault clearing time.

TABLE I THE SELECTED INPUT FEATURES

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_{cl}
Tz16	Maximum value of the difference of acceleration rates at t_{cl}
Tz17	Maximum value of the difference of rotor kinetic energies at t_{cl}
Tz18	Maximum value of the difference of rotor angles at t_{cl}
Tz19	Mean value of all the rotor kinetic energies at t_{cl}
Tz20	Rotor angle of the machine with the maximum kinetic energy at t_{cl}
Tz21	Kinetic energy of the machine with the maximum rotor angle at t_{cl}
Tz22	Maximum value of all the rotor kinetic energies at t_{cl}
Tz23	Maximum value of all the rotor kinetic energies at t_{cl+3c}
Tz24	Maximum value of all the rotor kinetic energies at t_{cl+6c}
Tz25	Maximum value of all the rotor kinetic energies at t_{cl+9c}
Tz26	Kinetic energy of the machine with the maximum rotor angle at t_{cl+3c}
Tz27	Kinetic energy of the machine with the maximum rotor angle at t_{cl+6c}
Tz28	Kinetic energy of the machine with the maximum rotor angle at t_{cl+9c}
Tz29	Maximum value of the difference of rotor angles at t_{cl+3c}
Tz30	Maximum value of the difference of rotor angles at t_{cl+6c}
Tz31	Maximum value of the difference of rotor angles at t_{cl+9c}
Tz32	Mean value of all the rotor kinetic energies at t_{cl+3c}
Tz33	Mean value of all the rotor kinetic energies at t_{cl+6c}
Tz34	Mean value of all the rotor kinetic energies at t_{cl+9c}

3.3 Data Pre-processing

To avoid training difficulties caused by the different units of input features, z-score standardization method is used as the data pre-processing method.

$$d' = (d - \overline{D})/\sigma_D \tag{13}$$

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, $d \in D$.

3.4 KPCA

In KPCA, a nonlinear transformation Φ maps the input data into a higher dimensional feature space F and PCA is then carried out in that feature space [22].

For the input data $\{\mathbf{x}_i\}_{i=1}^m$, $\Phi(\mathbf{x}_i)$ represents the mapped data in F. The covariance matrix $\overline{\mathbf{C}}$ of the eigenvalue problem $\lambda\mathbf{v} = \overline{\mathbf{C}}\mathbf{v}$ is

$$\overline{\mathbf{C}} = \frac{1}{m}\sum_{j=1}^{m}\Phi(\mathbf{x}_j)\Phi(\mathbf{x}_j)^T \tag{14}$$

with eigenvalues $\lambda \geq 0$ and eigenvectors $\mathbf{v} \in F \setminus \{0\}$. We can write an equivalent system

$$\lambda(\Phi(\mathbf{x}_i)\cdot\mathbf{v}) = (\Phi(\mathbf{x}_i)\cdot\overline{\mathbf{C}}\mathbf{v}) \tag{15}$$

Then, consider the coefficients α_i such that

$$\mathbf{v} = \sum_{i=1}^{m}\alpha_i\Phi(\mathbf{x}_i) \tag{16}$$

and define an $m \times m$ kernel matrix $\mathbf{K}$ by

$$K_{ij} := (\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)). \tag{17}$$

Combining (15), (16), and (17) leads to

$$m\lambda\mathbf{K}\boldsymbol{\alpha} = \mathbf{K}^2\boldsymbol{\alpha} \tag{18}$$

with $\boldsymbol{\alpha}$ a column vector with entries $\alpha_1,\cdots,\alpha_m$. Finding solutions for the eigenvalue problem $m\lambda\boldsymbol{\alpha} = \mathbf{K}\boldsymbol{\alpha}$, solves (18). To extract the principal components, the projection of the image of a test point $\Phi(\mathrm{x})$ onto the eigenvectors $\mathbf{v}^k$ in F is computed via

$$(\mathbf{v}^k\cdot\Phi(\mathrm{x})) = \sum_{i=1}^{m}\alpha_i^k(\Phi(\mathbf{x}_i)\cdot\Phi(\mathrm{x})) \tag{19}$$

The dot product matrix is given by $K_{ij} = k(\mathbf{x}_i,\mathbf{x}_j)$, then

$$(\mathbf{v}^k\cdot\Phi(\mathrm{x})) = \sum_{i=1}^{m}\alpha_i^k K(x_i,x) \tag{20}$$

The kernel function employed in this paper is the radial basis function (RBF):

$$K(\mathbf{x}_i,\mathbf{x}_j) = \exp(-\left\|\mathbf{x}_i-\mathbf{x}_j\right\|/2\sigma^2) \tag{21}$$

where σ is the width of the Gaussian.

3.5 Training of GP-based TSA Model

The training process of the GP-based TSA model is the process of determining the model parameters. The training process of the proposed model mainly includes the following steps.

Step 1: According to (8), determine priori forms of the covariance functions and initialize the corresponding model hyperparameters.

Step 2: Based on the Bayesian framework and the training sample information, the optimal parameters of GP model are adaptively determined by using logarithmic marginal likelihood maximization.

Step 3: According to the obtained optimal hyperparameters, the ideal TSA model is obtained.

3.6 Testing Process of the Proposed Model

The testing process can be broken into the following steps.

Step 1: Data pre-processing. According to (13), data pre-processing is carried out and the sample set is divided into a training set and a test set.

Step 2: Based on the Bayesian framework, the ideal TSA model is obtained by applying maximum likelihood method to determine the optimal hyperparameters of GP automatically.

Step 3: According to (9), the posterior distribution of latent function value of the test sample can be obtained by using Laplace algorithm.

Step 4: According to (5), the prediction probability that the final stability result is stable can be

obtained. If $p_* > 0.5$, the system is considered as stable; otherwise, the system is considered as unstable.

4 Case Study

In order to evaluate the performance of the proposed method, the New England 39-bus test system was used. This system is a well-known test case for TSA studies reported in the literature [15], [16]. The one diagram of the test system is shown in Fig. 1.

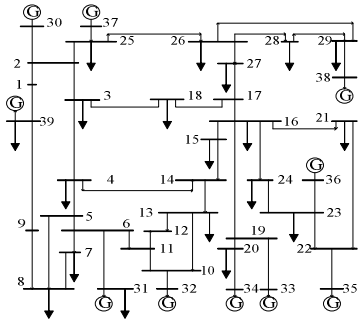

Fig. 1. New England 39-bus test system

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 @ 2.20 GHz, 2.19 GHz and main memory 1 GB.

4.1 Test Results of the Proposed Method

In order to verify the effectiveness of the proposed method, PCA was used to analyze the original feature set A, and obtained the 10-dimensional feature subset A_2 which responsible for 95% of the variance of A. Comparison tests were carried out between A, A_2 and the feature subset A_1 obtained by KPCA. The optimal hyperparameters of GP models and test results are shown in Table 2 and Table 3.

TABLE II THE OPTIMAL MODEL PARAMETERS

Covariance Function	Hyperparameter	Optimal value
SE_{iso}	σ_f	1.62
	l	3.94
RQ_{iso}	σ_f	3.11
	l	5.29
	α	2.35
CK_{iso}	$\sigma_f^{(SE)}$	4.72
	$l^{(SE)}$	5.63
	$\sigma_f^{(RQ)}$	3.76
	$l^{(RQ)}$	1.85
	α	3.08

TABLE III TEST RESULTS OF GP MODELS

Feature set	Dimension	TSA model	Training time /s	Test accuracy /%
A	34	$GP(SE_{iso})$	14.26	96.13
		$GP(RQ_{iso})$	18.03	95.48
		$GP(CK_{iso})$	20.48	98.06
A_1	8	$GP(SE_{iso})$	13.49	96.77
		$GP(RQ_{iso})$	17.64	94.84
		$GP(CK_{iso})$	18.69	98.71
A_2	11	$GP(SE_{iso})$	13.65	95.48
		$GP(RQ_{iso})$	16.96	94.19
		$GP(CK_{iso})$	19.41	96.77

As shown in Table. 3, compared with A, A_1 has similar classification accuracy, but the data dimension is reduced to 1/4. Moreover, KPCA has better classification accuracy and fewer dimensions than PCA.

From Table 3, we can also see that different covariance functions lead to large differences in classification accuracy of the corresponding GP models, and that GP (CK_{iso}) has better classification accuracy than other GP models in all three feature sets. The results show that a covariance function is the crucial ingredient affecting the performance of GP, and that making a new composite covariance function is an effective way to improve the classification accuracy and generalization performance of the GP model.

4.2 Test Results of Other TSA Models

In order to properly evaluate the performance of the proposed method, the above feature sets A and A1 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); the training algorithm of MLP was the back-propagation 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 [16]. The test results are shown in Table. 4.

TABLE IV TEST RESULTS OF OTHER MODELS

Feature set	TSA model	Test accuracy /%
A	DT	95.48
	MLP	94.84
	SVM	97.42
	$GP(CK_{iso})$	98.06

A_1	DT	96.13
	MLP	95.48
	SVM	96.77
	GP(CK_{iso})	98.71

From Table 4, it can be observed that the proposed GP-based TSA model has better classification accuracy than all other TSA models. The reason is that GP is based on the automatic Occam's razor principle to realize the trade-off between data-fit and model complexity [18].

4.3 Probability Interpretation of GP-based TSA Model

As shown in Fig. 2, probability prediction characteristics of the GP model can give the uncertainty measure of TSA results, which can not be achieved by other above-mentioned MLTSA models, such as DT, MLP and SVM. The prediction probability p_* can not only give the qualitative results of whether the system is stable or not, but give quantitative evaluation of the stability trend. Therefore, it can provide a reference for a hierarchical control strategy.

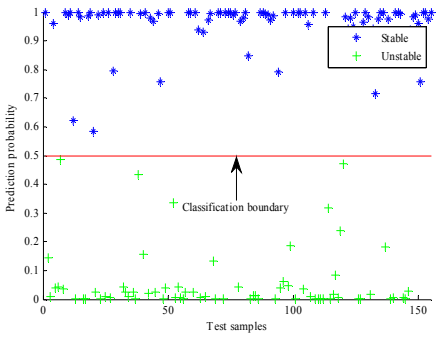

Fig. 2. Probabilistic prediction of GP

Furthermore, the samples with prediction probability p_* in a middle area which cannot be distinctly assessed are classified as indeterminate cases (e.g., the upper boundary and lower boundary are respectively set to 0.8 and 0.2), not assigned to either the stable or the unstable class, as shown in Fig. 3. This can reduce erroneous classification and improve the reliability of TSA.

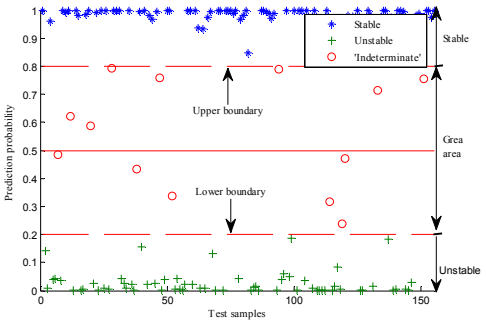

Fig. 3. Schematic diagram of 'indeterminate' class

5 Conclusions

Considering the possible real-time information provided by PMU, a new method for transient stability assessment of power systems using KPCA and GP 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)The proposed method can overcome the disadvantages existed in many of current MLTSA methods, such as overfitting, difficulty in parameter selection and prediction with no probability interpretation.

(2)Without sacrificing classification performance of the original feature set, KPCA can significantly reduce the data dimension, and has better performance than PCA.

(3)Compared with single covariance functions, composite ones can effectively improve the classification accuracy and generalization performance of the GP model.

References:

[1] P. Kundur, J. Paserba, V. Ajjarapu, G. Anderson, A. Bose, C. Canizares, N. Hatziargyrious, D. Hill, A. Stankovic, C. Taylor, T. V. Cutsem, and V. Vittal, Definition and Classification of Power System Stability, *IEEE Transactions on Power Systems*, Vol. 19 No. 2, 2004, pp. 1387-1401.

[2] U.S.-Canada Power System Outage Task Force, Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations, April 2004.

[3] 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*, Vol. 93, pp. 892-906, 2005.

[4] A.G. Phadke and J.S. Thorp, *Synchronized Phasor Measurements and Their Applications* , New York: Springer, 2008.

[5] Greyson K.A., Oonsivilai, A., Alternative technique of classifying PMUs at optimal environment in power system, *WSEAS Transactions on Power Systems*, Vol. 3, No. 10, 2008, pp. 643-652.

[6] 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*, Vol. 99, pp. 80-93, 2011.

[7] P. M. Anderson and A. A. Fouad, *Power System Control and Stability*, Piscataway, NJ: IEEE Press, 2003.

[8] M. A. Pai, *Energy Function Analysis for Power System Stability*, Boston, MA: Kluwer, 1989.

[9] Xue Y, Van Custem T, Ribbens-Pavella M, Extended equal area criterion justifications , generalizations, applications, *IEEE Transactions on Power Systems*, Vol. 4 No. 1, 1989, pp. 44-52.

[10] D. J. Sobajie, Y. H. Pao, Artificial neural-net based dynamic security assessment for electric power systems, *IEEE Transactions on Power Systems*, Vol. 4 No. 1, 1989, pp. 220-228.

[11] Wehenkel L., Pavella M. , Euxibie E., et al, Decision tree based transient stability method a case study, *IEEE Transactions on Power systems*, Vol. 9 No. 1, 1994, pp. 459-469.

[12] L. S. Moulin, A. P. A. da Silva, M. A. El-Sharkawi, et al, Support vector machines for transient stability analysis of large-scale power systems, *IEEE Transactions on Power Systems*, Vol. 19 No. 2, 2004, pp. 818-825.

[13] Sawhney H., Jeyasurya B., A feed-forward artificial neural network with enhanced feature selection for power system transient stability assessment, *Electric Power Systems Research*, Vol. 76 No. 12, 2006, pp. 1047-1054.

[14] 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*, Vol. 22 No. 4, 2007, pp. 1935-1943.

[15] N. Amjady, S. F. Majedi, Transient stability prediction by a hybrid intelligent system, *IEEE Transactions on Power Systems*, Vol. 22 No. 3, 2007, pp. 1275-1283.

[16] 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 Transactions on Power Systems*, Vol. 26 No. 3, 2011, pp. 1474-1483.

[17] VAPNIK V., *The nature of statistical learning theory*, New York: Springer Verlag, 2000.

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

[19] H. Nickisch, C. E. Rasmussen, Approximations for binary Gaussian process classification, *Journal of Machine Learning Research*, Vol. 9, 2008, pp. 2035-2078.

[20] Fei Cheng, Jiangsheng Yu, Huilin Xiong, Facial expression recognition in JAFFE dataset based on Gaussian process classification, *IEEE Transactions on Neural Networks*, Vol. 21 No. 10, 2010, pp. 1685-1690 .

[21] Mahdi Jadaliha, Yunfei Xu, Jongeun Choi, N. S. Johnson, Weiming Li, Gaussian process regression for sensor networks under localization uncertainty, *IEEE Transactions on Signal Processing*, Vol. 61 No. 2, 2013, pp. 223-237.

[22] B. Schölkopf, A. Smola, and K. R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation*, Vol. 10 No. 5, 1998, pp. 1299-1319.