Computer Modelling and New Technologies, 2013, vol. 17, no. 2, 24–35
Transport and Telecommunication Institute, Lomonosov 1, LV–1019, Riga, Latvia

A NOVEL ONLINE ASSESSMENT SCHEME FOR POWER SYSTEM SECURITY LEVEL OF TRANSIENT STABILITY BASED ON RELATIONSHIPS EXPLORATION IN A LARGE DATA SET

F. Qian[2], S. Liu[1], W. Yang[2], H. Wei[2], Yo. Fan[1]

[1] School of Electrical Engineering, Wuhan University, Wuhan 430072, China
[2]Electric power dispatching control center, Guangdong Power Grid, China Southern Power Grid, Guangzhou 510000, China
E-mail: 2011202070013@whu.edu.cn. E-mail: 1612234945@qq.com

In this paper, a scheme based on data mining and relationships exploration is presented for security level assessment of transient stability. The proposed scheme can select the optimal variables as input features by detecting the relationship between critical clearing time (CCT) and each variable in a large power flow data set. The data set is created based on a series of power flow simulation and fault simulation in the software PSS/E. The relationships exploration statistical tool used is based on the maximal information coefficient (MIC) and the Pearson product-moment correlation coefficient (PPMCC). The variables selected are corresponding to the relationships highly ranked by MIC and PPMCC, including linear relationships and especially nonlinear ones. These relationships are also shown in the paper and some of them are explained from the perspective of power system operation. If the measured values of these variables are obtained in real-time from wide area measurement system (WAMS), the CCT can be estimated in real time since its relationships with these variables are explored. Then the security level of transient stability can be assessed for a new operation state. The scheme is tested on a 21-bus system provided by PSS/E and various test results indicate the scheme is accurate and effective, as well as the way to select input features is more intelligent than the current techniques.

Keywords: Maximal information coefficient (MIC), Pearson product-moment correlation coefficient (PPMCC), relationships exploration, large data sets, security assessment, transient stability, critical clearing time (CCT).

NOMENCLATURE

PGi	Active power of generators at bus
QGi	Reactive power of generators at bus
Vi	Voltage amplitude of bus
θi	Voltage phase angle of bus
Pij	Active power from bus to bus
Qij	Reactive power from bus to bus
Sij	Apparent power from bus to bus
Pi_LOAD_PQ/I/Y	Active power from bus to constant power, current, or impedance load
Qi_LOAD_PQ/I/Y	Reactive power from bus to constant power, current, or impedance load
Si_LOAD_PQ/I/Y	Apparent power from bus to constant power, current, or impedance load
Pi_X	Active power from bus to equipment
Qi_X	Reactive power from bus to equipment
Si_X	Apparent power from bus to equipment
I%i_j	Load rate percent of the transmission line between bus and bus
I%i_j_1	Load rate percent of the first transmission line between bus and bus
I%i_j_2	Load rate percent of the second transmission line between bus and bus
I%i_X	Load rate percent of the equipment at bus
PLOSSi_j	Active power loss of the transmission line between bus and bus
QLOSSi_j	Reactive power loss of the transmission line between bus and bus
PLOSSi_j_1	Active power loss of the first transmission line between bus and bus
QLOSSi_j_1	Reactive power loss of the first transmission line between bus and bus
PLOSSi_j_2	Active power loss of the second transmission line between bus and bus
QLOSSi_j_2	Reactive power loss of the second transmission line between bus and bus

1. Introduction

Security level assessment of transient stability is essential to the power system operation and control [1]. There has been a continually increasing interest and investigation into security level assessment of transient stability [2]–[4]. Transient stability or large disturbance rotor angle stability is concerned with the ability of the power system to maintain synchronism when subjected to a severe disturbance, such as a short circuit on a transmission line. At present security scanning and assessment for power system mainly rely on a lot of fault simulation. Considering the large scale of current power system, different types of equipment, the real-time changing load and changing generator output, the enumerated probabilistic fault simulation analysis method is not able to provide real-time assessment results or effective control measures information of improving the security level. In general, a single simulation method cannot meet the demand of intelligent decision for the power system control. Therefore, there is a pressing need to develop fast online security assessment method which could analyse the level of security and forewarn the system operators to take necessary preventive actions in case need arises [5], [6].

The security level assessment of transient stability is a problem with inherent complexity, non-linearity, uncertainty and the need for online monitoring. Seeking and establishing the mapping relationships between the power system operation state monitoring data and security level of transient stability based on knowledge engineering technology and data mining [7–9] is a very attractive idea. It is built on a large number of accumulated samples, which relies on fault simulation scan of transient stability. It is designed to discover the implicit relationships in the fault scan results and power flow information, which may be useful to assess security level. This security level assessment idea makes up for the deficiency of simulation scan. First of all, once the security level assessment rules or model have been established by offline data mining, the computation speed of online security assessment will be fast. Secondly, considering that power flow of power system can be easily observed, adjust and control, it will be convenient to improve the security level based on the relationships and it plays a role of aid decision making. However, it still faces many challenges to achieve direct security assessment based on operating information and data mining. We need to figure out how the power flow distribution and steady-state operating information of the system affect the security level of transient stability under a given fault. Another challenge is: the WAMS of a large power system may collect a huge amount of power flow and operating data, which not only contains features of high correlation for security level, but also the ones of weak correlation. How to effectively select the dimensions of the input feature, extract and classify the high correlation features, and eliminate redundant features is a key step in the security level assessment based on the artificial intelligence theory[10–12]. The purpose of feature selection is to select and classify high correlation features from a large number of original features. It requires reducing the dimension and minimum information loss of representing the research target. By selecting the features highly associating with the research target, the purpose of the d – dimensional feature extracted from the D – dimensional feature space $(d << D)$ can be achieved [13], [14].

A lot methods have been applied to the transient stability assessment, such as artificial neural networks(ANN) [15], pattern recognition techniques [16], decision trees [17], and fuzzy neural networks [18]. Some optimisation algorithms such as simulated annealing algorithm and ant colony algorithm are also applied to transient stability assessment [19]–[21], and its purpose is to choose better input features. In [15], [22], ANN is adopted and correct estimation results are got. However, the input features used in the paper may not be the optimal ones and the size of selected features is not a small number for its test system. For a large system, it is necessary to select a small number of optimal input features from the massive features. The methods in these papers may fail because of the curse of dimensionality. In [20], [21], the number of variables in the optimal variable group given by optimisation algorithm is small, which reduce the dimension well. However, the final accuracy is not particularly high. Moreover, the new additional variable cannot be directly given by optimisation algorithm when the assessment accuracy needs to be improved. In [23], [24], the CCT in a case of assumption fault is calculated precisely with a short time, but the calculation time-consuming still will increase rapidly and cannot meet the needs of online assessment if the size of the actual power system is huge or with even more uncertain faults. In [17], [18], transient stability prediction accuracy for a fault is relatively high, but the security level assessment of transient stability for normal steady-state operation state is not given.

The online assessment scheme presented in this paper is based on relationships exploration in a large data set of offline power flow simulation and fault simulation. The optimal features are selected based on the implicit relationships explored in the data set. The relationships between these variables and security level of transient stability are presented and explained. Finally, the optimal variables will be applied to online assessment scheme for power system security level of transient stability. The way to select features is different from the conventional optimisation algorithms of feature selection, and it can explain the variables group given by optimisation algorithm to a certain extent. The number of input features in the scheme can be chosen freely, which is based on the needed assessment accuracy. It can overcome the curse of dimensionality of large-scale power systems. The applicability will not be influenced by the change of the structure and scale since input features selected are based on data statistics and mining. The scheme has a certain requirement for the number of samples of the previous offline simulation. Steps of the scheme are shown on Figure 1.

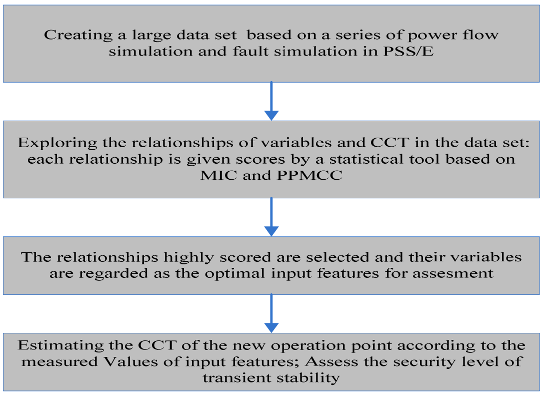

Figure 1. Assessment scheme

In the proposed approach, the results of power flow simulation and fault simulation in PSS/E are used to create a large data set, which include all kinds of operation variables and CCT. The relationship between CCT and each variable is explored and given scores by the maximal information coefficient (MIC) and Pearson's product moment correlation coefficient (PPMCC). The relationships highly scored determine their variables will be selected, and these variables are regarded as the optimal input features. The relationships highly scored are also presented and explained in the paper. Then the paper presents an estimation method using the optimal input features. The scheme is tested on a 21-bus system provided by PSS/E to assess security level for a lot of new operating points. The results of the assessment are verified to be correct. For real time application, the data of input features can be obtained from phase measurement units (PMUs). The scheme is economical since it requires a relatively small number of PMUs.

This paper is organized as follows: Section 2 provides a detailed statement of the problem and some supporting mathematical methods. Section 3 introduces the method to create a large data set based on simulation. Section 4 explores the relationships of variables and CCT. Section 5 presents the estimation method of CCT and security level assessment of transient stability. Section 6 concludes the paper.

2. Problem Statement and Supporting Mathematical Methods

2.1. Critical Clearing Time (CCT)

Given a certain fault, the transient stability of the Power system is usually described with the critical clearing time (CCT). The system with a longer CCT is considered to have a higher security level of transient stability. CCT can be used as the index for transient stability assessment. The delay in fault clearing from the CCT means loss of synchronous operation of the generators in the power system. Compared with other indices, CCT is easy for operators to understand how stable the power system is during operation.

2.2. Maximal Information Coefficient (MIC)

MIC is a measure of dependence for two-variable relationships and it captures relationships both functional and not in large data sets. The concept is presented recently by the paper [25]. MIC belongs to a larger class of maximal information-based nonparametric exploration statistics for identifying and classifying relationships. MIC is applied to the data sets of biology and has detected novel relationships. In this paper, MIC is introduced into the field of power system and it is used to explore the relationships of CCT and lots of operation variables in power system operation.

MIC is based on the idea that if there is a relationship between two variables, then a grid can be drawn on the scatter plot of the two variables that partitions the data to encapsulate the relationship. MIC can give a score to measure the relationship between two variables based on the data pairs of variables. MIC can capture a wide range of interesting relationships, not only to specific function types (such as linear, parabolic, or sinusoidal), but also to even all functional relationships. For equally noisy relationships of different types, MIC can give similar scores to them.

Given a finite set *D* of ordered pairs, the *x*-values of *D* are partitioned into *x* bins and the *y*-values of *D* are partitioned into *y* bins, allowing empty bins. Such a pair of partitions can be called an *x*-by-*y* grid. Given a grid *G*, let $D|G$ be the distribution induced by the points in *D* on the cells of *G*. The distribution on the cells of *G* is obtained by letting the probability mass in each cell be the fraction of points in *D* falling in that cell. For a fixed *D*, different grids *G* result in different distributions $D|G$. For a data set *D* of two-variable, the MIC of their relationship is given by (1), (2).

For a finite set $D \subset R^2$ and positive integers *x, y*.

$$I^*(D,x,y) = \max I(D|_G) . \tag{1}$$

Where the maximum is over all grids G with x columns and y rows, and denotes the mutual information of .The MIC of two-variable data with sample size n and grid size less than is given by

$$MIC(D) = \max_{xy<B(x,y)} \left\{ \frac{I^*(D,x,y)}{\log \min\{x,y\}} \right\}. \tag{2}$$

Where for some.

In the research, is used because it's found to work well in practice [25]. MIC falls between 0 and 1. Some properties of MIC are as follows.

1) MIC assigns scores that tend to 1 to all never-constant noiseless functional relationships.
2) MIC assigns scores that tend to 1 for a larger class of noiseless relationships.
3) MIC assigns scores that tend to 0 to statistically independent variables.

2.3. Pearson's Product Moment Correlation Coefficient (PPMCC)

In statistics, PPMCC [26] is used to measure the linear correlation degree between two variables *X* and *Y*. In the field of natural sciences, the coefficient is widely adopted to detect the linear relationships in variables. In this paper, PPMCC is specially used to evaluate the degree of linearity for the relationship of a variable between CCT. For samples (X_i, Y_i), a kind of expression for PPMCC is expressed as the mean of standard score, which is given by (3).

$$\rho = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{S_X}\right)\left(\frac{Y_i - \overline{Y}}{S_Y}\right). \tag{3}$$

Where *n* is the size of samples, $\overline{X}$ is the mean of samples, and S_X is the standard deviation of samples.

PPMCC falls between -1 and 1. Some properties of PPMCC are as follows.

1) If $\rho > 0$, it shows that the two variables are positive correlative.

2) If $\rho > 0$, it shows that the two variables are negative correlative.

3) The larger the absolute value of ρ is, the stronger linear relationship between the two variables exists.

4) If $\rho = 0$, it shows that the two variables are not linear correlative, but there may be another relationship between the two variables (such as a curvilinear association).

3. Creating a Large Data Set of CCT and Operation Variables in Power System Based on Simulation

In the practice of power system operation, the accidents of transient stability failure are unusual. It leads to the lack of instability data samples, which cannot satisfy the requirements of data mining. Therefore, simulation is usually used to obtain samples. A well-known software package named PEE/S is used to create the necessary data set. PSS/E is a software tool used for power system, and it can be used to analyse transient stability [27]. This paper uses Python programming language in PSS/E to achieve the follows: Initialise the data of loads randomly and the matching data of generators' output in a normal range. Solve the power flow of operating sample point, and then choose the convergent ones. Make a series of three-phase instantaneous short circuit fault simulation test at a set place and calculate its CCT for every selected operating sample point. The run will end until the set number of samples is satisfied. The program flow chart is illustrated on Figure 2.

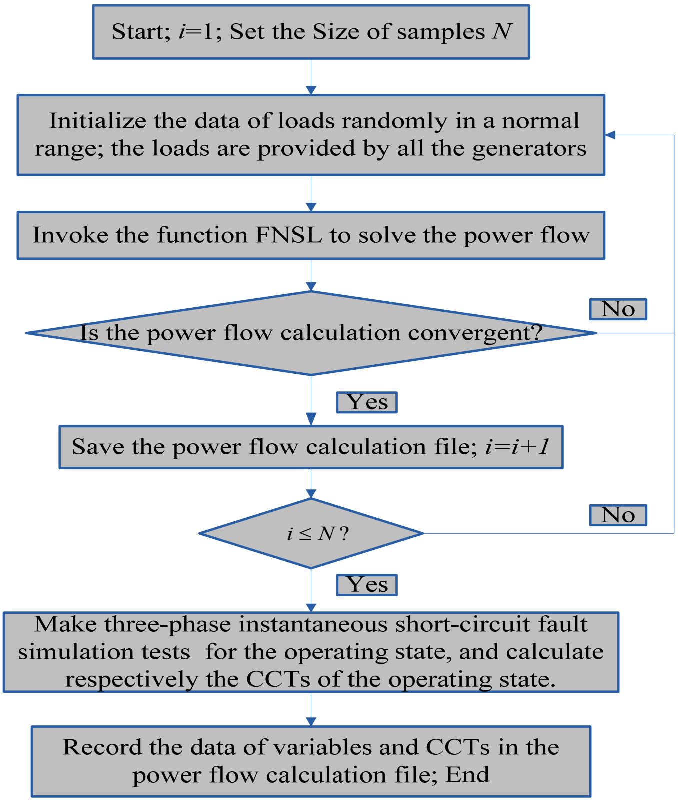

Figure 2. Program flow chart for creating a large data set of power flow and CCT based on simulation in PSS/E

Some explanatory notes for the program flow chart on Figure 2 are as follows.

1) N can be set according to the size of samples needed and $N = 150$ is used in this paper.

2) A lot of physical variables are extracted from the power flow results of the system: generator active power and reactive power of every generator; voltage amplitude and phase angle of every bus; active power, reactive power, apparent power, active power loss and reactive power loss of every branch; load rate percentage of every branch; active power and reactive power from a bus to an equipment.

3) According to the experience of the power system operation, some other variables are added because they are reasonably suspected to have a relationship with CCT. The additional variables are based on some fundamental variables in the power flow results. They are: phase angle difference between any two generators at steady-state, the maximum and minimum of absolute value of phase angle difference between any two generators at steady-state.

In the paper, a 21-bus test system [28] provided by PSS/E is used, which is shown on Figure 3.

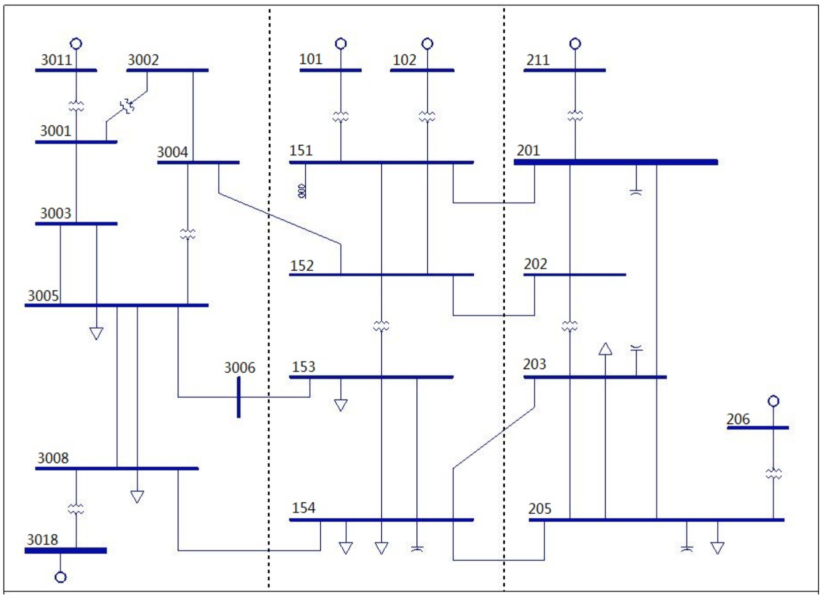

Figure 3. 21-bus test system provided by PSS/E

In the test system, 470 physical variables are finally extracted after eliminating all of the constant variables. Then a large data set of power flow variables and CCT is created, which is a matrix with 471 rows (470 physical variables and CCT) and 150 columns. The three-phase instantaneous short circuit fault simulation test is as follow: a fault occurs at the head of the first line between the bus 205 and the bus 203 when $t = 0s$, then disconnect the line after a set time.

4. Exploring the Relationships of Power Flow Variables and CCT

4.1. Top Relationships Explored by MIC and PPMCC

MIC and PPMCC are applied to detect the relationships in the data set. A statistical tool is provided by the paper [25], which can calculate MIC and PPMCC of each relationship. Table 1 shows the top 1% of relationships by MIC and Table 1 shows the top 1% of relationships by PPMCC. The relationships highly ranked by PPMCC have high degrees of linearity. A relationship highly ranked by MIC show a certain relationship between the two variables, and it doesn't necessarily have a high degree of linearity.

Table 1. Top 1% of relationships by MIC

Var1	Var2	MIC	MIC Rank	PPMCC (ρ)	PPMCC Rank
CCT	*S152_151_1*	0.778	1	-0.813	109
CCT	*S152_151_2*	0.778	2	-0.813	110
CCT	*PLOSS151_152_1*	0.773	3	-0.821	80
CCT	*PLOSS151_152_2*	0.773	4	-0.821	81
CCT	*PLOSS152_151_1*	0.773	5	-0.821	82

Table 2. Top 1% of relationships by PPMCC

Var1	Var2	PPMCC (ρ)	PPMCC Rank	MIC	MIC Rank
CCT	*Q3005_3003_1*	0.885	1	0.684	62
CCT	*Q3005_3003_2*	0.885	2	0.684	63
CCT	*Q3004_3002*	0.884	3	0.673	74
CCT	*Q3004_152*	-0.882	4	0.722	44
CCT	*θ101*	-0.874	5	0.672	86

Specially, the relationship whose MIC rank is 1 in Table 1 and the ones whose PPMCC rank is 1 in Table 2 are shown successively on Figure 4 and Figure 5. These relationships are implicit in mass data, which are not easy to be found directly. After these relationships are detected out, generally reasonable explanation can be given for the relationships from the perspective of power system. The variable S152_151_1 is selected as an example, which represents the apparent power of the first line from bus 152 to bus 151 on Figure 6. In order to explain the relationship conveniently, a case can be imagined: with the increase of power system loads, the power on the transmission line will increase; when load rating of the system is low, the increase of load doesn't make CCT decrease significantly; when load rating of the system is high, a slight increase of load will make CCT decrease significantly.

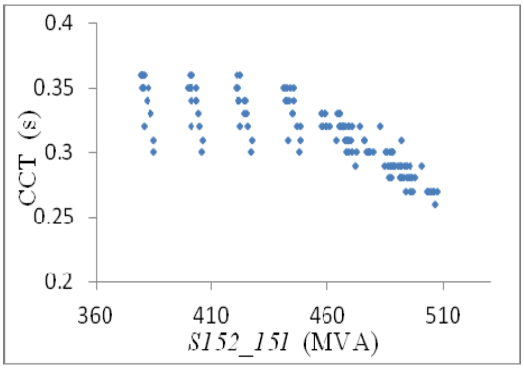

Figure 4. Scatter plot of the variable *S152_151_1* and CCT

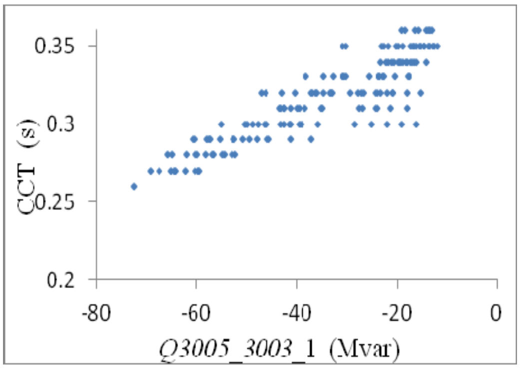

Figure 5. Scatter plot of the variable *Q3005_3003_1* and CCT

4.2. MIC versus PPMCC

Figure 6A shows that MIC versus PPMCC for all pair-wise relationships in the data set. In different areas of Figure 6A, different kinds of relationships can be found. Some examples are as follows.

1) Figure 6B: Both PPMCC and MIC yield low scores for unassociated variables. It indicates no specific relationship exist between the variable *θ3001* and CCT.
2) Figure 6C: Ordinary linear relationships score high under both MIC and PPMCC tests. It indicates an obvious linear relationship between the variable *θ101* and CCT.
3) Figure 6D: Relationships detected by MIC but not by PPMCC, because the relationships are nonlinear. It indicates a kind of nonlinear relationship between the variable the variable *Q205_201* and CCT.

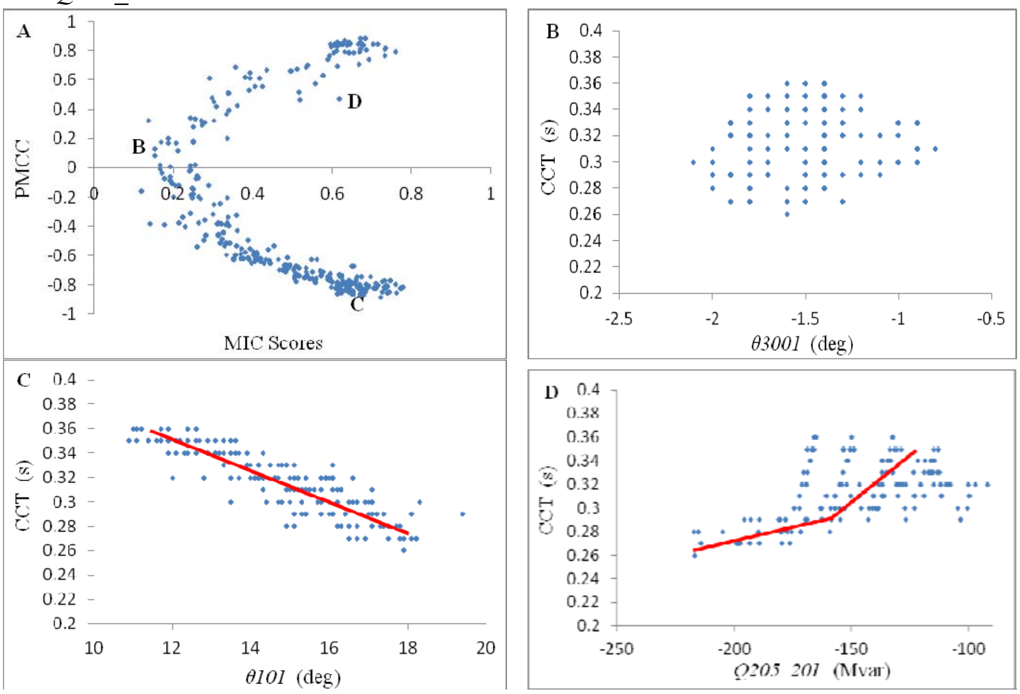

Figure 6. Application of MIC and PPMCC to the data set. (A) MIC versus PPMCC for all pair-wise relationships in the data set. (B)–(D) Examples of relationships from (A)

4.3. Top Nonlinear Relationships

In order to discover the nonlinear relationships of variables and CCT comprehensively and quickly, the statistic MIC $-\rho^2$ is adopted. For a relationship, a larger MIC and a smaller $|\rho|$ get a larger MIC $-\rho^2$. Table 3 shows the top 0.5% of relationships by MIC $-\rho^2$, and the relationships are shown on Figure 7.

Table 3. Top 0.5% of nonlinear relationships by MIC $-\rho^2$

Var1	Var2	MIC $-\rho^2$	MIC $-\rho^2$ Rank	MIC	PPMCC(ρ)
CCT	*Q205_201*	0.395	1	0.618	0.471
CCT	*Q205_206*	0.302	2	0.520	0.465

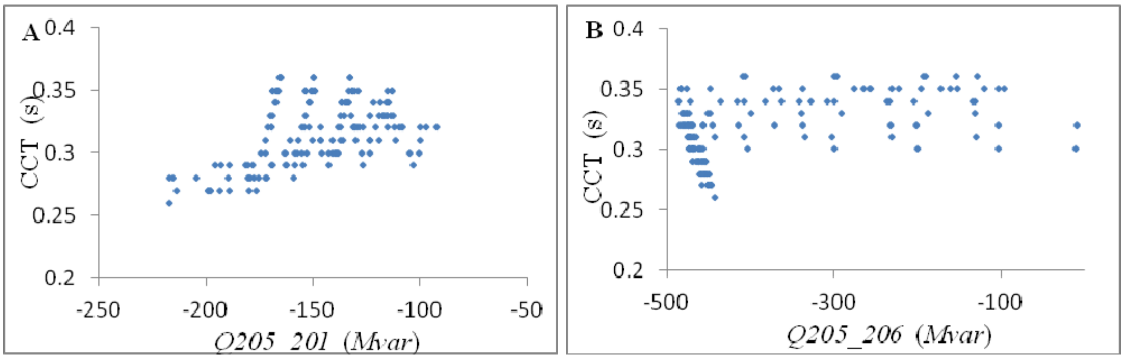

Figure 7. Scatter plot of the top 0.5% of nonlinear relationships selected by MIC $-\rho^2$

Some explanations for the relationships on Figure 7 from the perspective of power system are as follows.

1) *Q205_201* represents the reactive power from bus 205 to bus 201. At a general operation state, the value of this variable is minus and will decrease gradually with the increase of power system loads. Compared to an operation state of low load rating, the decrease of CCT will become faster when the system is operated at a state of high load rating. Therefore the relationship between the variable *Q205_201* and CCT is that shown on Figure 7A.

2) *Q205_206* represents the reactive power from bus 205 to bus 206. When load rating of the system is low, the value of this variable will appear in the interval (–500, 0) randomly. When load rating of the system is high, the variable will tend to its minimum value because there is a maximum reactive power limit for the generator at bus 206.Therefore the relationship between the variable *Q205_206* and CCT is that shown on Figure 7B.

5. Estimation of CCT and Test Results

5.1. Selecting Input Features

After exploring the relationships of variables in power flow and CCT, these relationships can be used to estimate CCT and assess the security level of transient stability for a new operation state. The values of variables can be obtained from WAMS in practice. Obviously, the accuracy of estimation is affected by the variables selected and measured. The variables selected should be the ones who have obvious relationships with CCT. Moreover, the total number M of the variables selected should be appropriate. Setting M too low can lead to inaccurate estimation ranges, while setting M too low means the increase of economic cost in engineering application because of too many measuring points. An appropriate selection is given based on tests: the top 1% of relationships by MIC, the top 1% of relationships by PPMCC, and the top 0.5% of nonlinear relationships by MIC$-\rho^2$. Therefore the total number of relationships selected is 12 in the paper. It is should be noted if a variable always remains the same with another one (such as *S152_151_1* and *S152_151_2* in Table 1), one of them will be eliminated and a variable ranking behind the two variables will be selected to guarantee $M = 12$. The variables finally selected are in Tables 3–5, which are the optimal ones to be input features.

Table 4. Variables finally selected from the ranking list by MIC

Var1	Var2	MIC	MIC Rank
CCT	*S152_151_1*	0.778	1
CCT	*PLOSS151_152_1*	0.773	3
CCT	*QLOSS151_152_1*	0.773	7
CCT	*QLOSS201_202*	0.763	11
CCT	*P152_151_1*	0.760	13

Table 5. Variables finally selected from the ranking list by PPMCC

Var1	Var2	PPMCC (ρ)	PPMCC Rank
CCT	*Q3005_3003_1*	0.885	1
CCT	*Q3004_3002*	0.884	3
CCT	*Q3004_152*	0.882	4
CCT	*θ101*	0.874	5
CCT	*Q3005_3006*	0.871	8

5.2. Program Flow Chart for Estimation of CCT

For each input variable, the idea of estimation is as follow. As it is illustrated on Figure 8, the measurement value of the variable a205_201 is $x_0$when the system is operated at a new point. In all the points on Figure 8, the smallest value of x–axis is X_{min} and the largest value is X_{max}. All the points with x–axis in the range (x_1,x_2) are searched out. The smallest value of y–axis in these points is y_1 and the largest value is y_2. The CCT of the new operation state can be considered in the range (y_1,y_2). Equations (4)–(8) are given for Figure 8. Specially, 3% is used in (5) and 0.5 is used in (6), which are found to work well in practice. A too wide search range leads to an imprecise estimation result. If the search range is too narrow, maybe no result is returned because the sample size is finite.

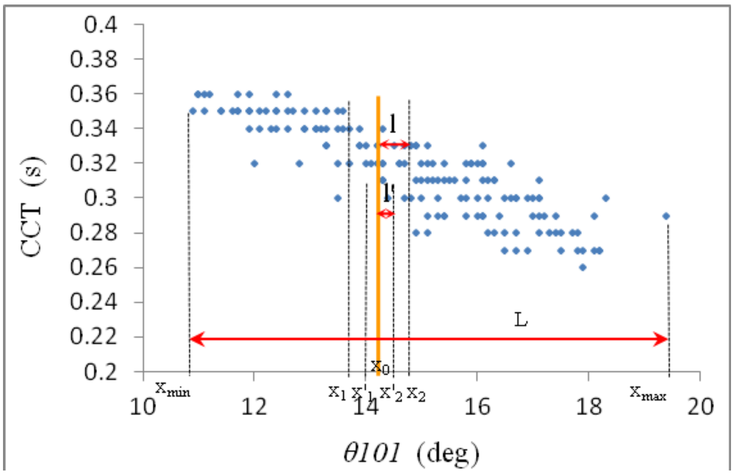

Figure 8. Estimation method for each variable

$$L = X_{max} - X_{min} \tag{4}$$

$$l = L * 3\% \tag{5}$$

$$l' = 0.5 * l \tag{6}$$

$$X_1 = X_0 - l \tag{7}$$

$$X_2 = X_0 + l \tag{8}$$

The selected variables are used to estimate CCT and the program flow chart is shown on Figure 9. The parameter t is introduced to update l when needed, which can increase the imprecision of estimation because the search range is narrower with a smaller l. If an estimation range (y_1,y_2) given at the first search satisfies $Y_2 - Y_1 > 0.01$, the existence of parameter t can give it a more accurate search opportunity with the updated l. The method can try to make the final estimation range satisfies $Y_2 - Y_1 \leq 0.01$, which will be more precise.

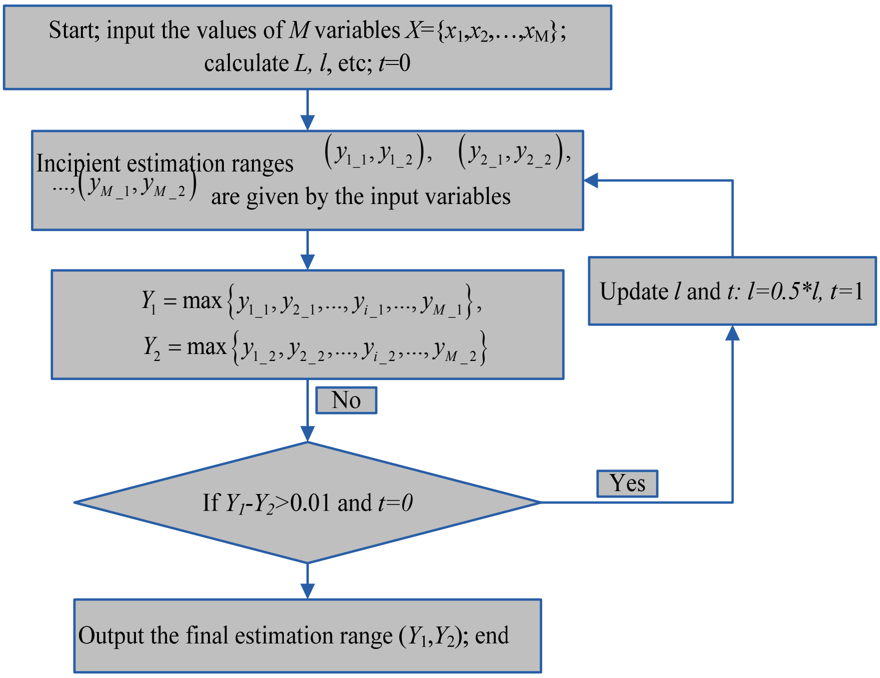

Figure 9. Program flow chart of estimating CCT

5.3. CCT estimation results and security level assessment of transient stability

In the tests, possible new operation states are set stochastically. 15 of them are randomly selected and shown in Table 6. Some explanatory notes for Table 6 are as follows.

1) R is the final estimation range of a new point, which is given by the program with the 12 selected variables as input features.

2) The span of R is shown to measure the precision of estimation range.

3) S is the value given by simulation, which is used to verify the correctness of estimation.

4) Occasionally, CCT is not visual enough to describe security level of transient stability for the power system operators. Therefore, samples can be classified as some levels according to their distribution. In this case, 5 levels are as follows: Level 1: 0.35–0.36; Level 2: 0.33–0.34; Level 3: 0.30–0.32; Level 4: 0.28–0.29; Level 5: 0.26–0.27. Level in Table 6 is security level assessment of transient stability for a new operation state.

Table 6. Estimation results of new operation states

No.	*R*(s)	Span of *R*(s)	*S*(s)	Level
1	0.27	0	0.27	5
2	0.27–0.28	0.01	0.27	5
3	0.28	0	0.28	4
4	0.28–0.29	0.01	0.29	4
5	0.29–0.30	0.01	0.30	3
6	0.30–0.31	0.01	0.31	3
7	0.32–0.33	0.01	0.32	3
8	0.32–0.33	0.01	0.33	2
9	0.34	0	0.34	2
10	0.35–0.36	0.01	0.35	1
11	0.35–0.36	0.01	0.36	1
12	0.33–0.34	0.01	0.34	2
13	0.27–0.28	0.01	0.28	4
14	0.29	0	0.29	4
15	0.31–0.32	0.01	0.31	3

In general, each simulation value S for new operation state is in the range R and this demonstrates the correctness of estimation ranges. S is slightly out of the range in a very few tests, which is due to the finiteness of previous samples. After the statistics of more tests, the possibility of S being out of the range R is lower than 1%.

6. Conclusions

This paper proposes a novel online transient stability assessment scheme based on the connotative relationship exploration in a large data set, which includes CCT and lots of different operation variables in power system. The scheme includes: a series of power flow simulation and fault simulation, creating a large data set of variables and CCT, giving scores to the relationships and ranking them, using the linear and nonlinear relationships to estimate CCTs for new operation states. The data set is created with a series of simulation in PSS/E. The relationships of variables and CCT are given scores by the statistical methods based on MIC and PPMCC. Some highly ranked linear and nonlinear relationships are detected out, shown, explained, and then used for estimation. The estimation results are verified to be accurate and can be used to assess the security level of transient stability.

Different from conventional feature selection methods, the input features are selected from a great number of variables based on data mining and relationships exploration in this paper. It will be more intelligent and efficient than conventional optimisation algorithms because each feature is given a score and ranked clearly. The applicability of the scheme will not be influenced by the change of the structure and scale since it based on data statistics and mining. The estimation results are with high precision, which relies on the total number of previous operation samples to a certain extent. The estimation of CCT and the online security level assessment of transient stability will be important basis for system operators to change operation state to improve the security level in practice.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant (61074101, 50477018, and 51007093) and in part by the Specialized Research Fund for the Doctoral Program of Higher Education (20090141120062).

References

1. Dong, Z.Y., Zhao, J.H. & D.J. Hil. (Nov. 2012). Numerical Simulation for Stochastic Transient Stability Assessment. *IEEE Trans. Power Syst.*, 27(4), 1741–1749.
2. Chatterjee, D. & A. Ghosh. (Aug. 2007). Transient Stability Assessment of Power Systems Containing Series and Shunt Compensators. *IEEE Trans. Power Syst.*, 22(3), 1210–1220.
3. Xu, Y., Dong, Z.Y., Meng, K. et al. Mar. (2011). Real–time transient stability assessment model using extreme learning machine. *IET Generation, Transmission & Distribution*, 5(3), 314–322.
4. Miah. A. M. (Apr. 2011). Study of a coherency–based simple dynamic equivalent for transient stability assessment. *IET Generation, Transmission & Distribution*, 5(4), 405–416.
5. Niazi, K. R., Arora, C. M., Surana, S. L. (2003). Power system security evaluation using ANN: feature selection using divergence. In Proc. 2003 Int. Joint Conf. Neural Networks, July 20–24, 2003, Vol. 3 (pp. 2094–2099). IJCNN 2003, Portland, Oregon, USA.
6. Amjady, N. & S. F. Majedi. (Aug. 2007). Transient stability prediction by a hybrid intelligent system. *IEEE Trans. Power Syst.,* 22(3), 1275–1283.
7. Tao, X., Renmu, H., Peng, W., Dongjie, X. (2004). Applications of data mining technique for power system transient stability prediction. In Proceedings of the 2004 IEEE International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, during 5-8 April 2004, (pp. 1389–392). Hong Kong
8. Yu, Z. H., Zhou, X. X., Wu, Z. X. (2005). Fast Transient Stability Assessment Based on Data Mining for Large–Scale Power System. In Proc. Transmission and Distribution Conference and Exhibition, 2005 (pp. 1–6).
9. Wang, T. W. Guan, L. (2007). A data mining technique based on pattern discovery and k–nearest neighbor classifier for transient stability assessment. In Proc. Power Engineering Conference, IPEC 2007, 3–6 Dec. 2007 (pp. 118–123).
10. Silipo, R., Berthold, M. R. (Dec 2000). Input features' impact on fuzzy decision processes. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 30(6), 821–834.

11. Laine, T. I., Bauer, K. W., Jr., Lanning, J. W. et al. (Nov. 2002). Selection of input features across subjects for classifying crewmember workload using artificial neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 32(6), 691–704.
12. Kwak, N., Choi, C.–Ha–H. (Jan. 2002). Input feature selection for classification problems. *IEEE Transactions on Neural Networks*, 13(1), 143–159.
13. Chen, J. X., Wang, S. (Sep./Oct. 2001). Data visualization: parallel coordinates and dimension reduction. *Computing in Science & Engineering*, 3(5), 110–113.
14. Shen, L., Tan, E. C. (April–June 2005). Dimension reduction–based penalized logistic regression for cancer classification using microarray data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2(2), 166–175.
15. Verma, K., Niaz, K. R. (2011). Determination of vulnerable machines for online transient security assessment in smart grid using artificial neural network. In Proc. 2011 Annual IEEE India Conference (INDICON), 16–18 Dec. 2011 (pp. 1–5). Dept. of Electr. Eng., *Indian* Inst. of Technol., Kharagpur, *India*.
16. Garcia, G., Benussou, J. & M. Berbiche. (Jul. 1992). Pattern recognition applied to transient stability analysis of power systems with modeling including voltage and speed regulation. *IEE Proc. B Electric Power Applications*, 139(4), 321–325.
17. Rovnyak, S., Kretsinger, S., Thorp, J. & D. Brown. (Aug. 1994). Decision trees for real–time transient stability prediction. *IEEE Trans. Power Syst.*, 9(3), 1417–1426.
18. Amjady, V. (2005). Application of a new fuzzy neural network to transient stability prediction. In Proc. IEEE PES Summer Meeting, San Francisco, CA, Jun. 2005 (pp. 69–76). San Francisco, CA.
19. Cao, M., Wang, Y. J. (2011). Application of binary particle swarm optimization in feature selection for transient stability assessment. In Proc. 2011 International Conference on Electric Information and Control Engineering (ICEICE), 15–17 April 2011 (pp. 5719–5722 Wuhan, China, from http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5766331
20. Cao, M. (2006). *Feature Selection for Transient Stability Assessment Based on Genetic Simulated Annealing Algorithms*. Baoding: North China Electric Power University.
21. Zhang, X. Q., Guan, L., Wang, T. W. (Dec. 2010). Kernel Feature Identification Based on Improved Ant Colony Optimization Algorithm for Transient Stability Assessment. *Transactions of China Electrotechnical Society*, 25(12), 154–160.
22. Niazi, K. R., Arora, C. M., Surana, S. L. (2004). Power system security evaluation using ANN: feature selection using divergence. *Electric Power Systems Research*, 69, 161–167.
23. Al Marhoon, H.H., Leevongwat, I., Rastgoufard, P. (2012). A Practical Method for Power Systems Transient Stability and Security Analysis. In Proc. 2012 IEEE/PES Transmission & Distribution Conference and Exposition (T&D) 7–10 May 2012 (pp. 1–6). Orlando, FL, USA, from http://www.ieeet-d.org/TECHPROfinal.pdf
24. Yorino, N., Priyadi, A., Kakui, H., Takeshita, M. (Aug. 2010). A New Method for Obtaining Critical Clearing Time for Transient Stability. *IEEE Trans. Power Syst.*, 25(3), 1620–1626.
25. David, N. R., Yakir, A. R., Hilary, K. F., Sharon, R. G., Gilean, M., Peter, J. T., Eric, S. L., Michael, M. & C. S. Pardis. (Dec. 2011). Detecting novel associations in large data sets. *Science*, 334, 1518–1524.
26. Gayen, A. K. (1951). The frequency distribution of the product moment correlation coefficient in random samples of any size draw from non–normal universes. *Biometrika*. 38, 219–247.
27. Mohamad, A. M., Hashim, N., Hamzah, N., Nik, F., Abdul, L. & F. Mohd. (2011). Transient stability analysis on Sarawak's grid using Power System Simulator for Engineering (PSS/E). In IEEE Symposium on Industrial Electronics & Applications (ISIEA 2011) (pp. 521–526). Langkawi, Malaysia on 25–28 September 2011, from http://www.isiea.org/2011/
28. Power Technologies, Inc. *PSS/E Application Program Interface*. Aug. 2004, from http://www.energy.siemens.com/hq/en/services/power-transmission-distribution/power-technologies-international/software-solutions/pss-e.htm

Received on the 20th of May 2013