A Novel Probability Evaluation Method for Power System Transient Stability Assessment Based on Support Vector Machine

Li Xin[a, *], Fan Youping[b], Zhang Peng[c], Liu Songkai[d]

Wuhan University, Wuhan 430072, Hubei Province, China

[a] lxgpf@foxmail.com, [b] ypfan@whu.edu.cn, [c] 397038289@qq.com, [d] 1612234945 @qq.com

Keywords: transient stability assessment, support vector machine, Probability evaluation, on-line prediction.

Abstract: A general probability evaluation method is proposed to evaluate the accuracy of the data mining model in the on-line prediction of transient stability after the failure. Considering the distribution of failure probability, the paper first searches for possible faults, then evaluates the accuracy according to the actual probability distribution of uncertain factors, which is more objective than Monte Carlo method. Finally, simulation is carried out with the new England 39 bus system, the test results show that this method can not only comprehensively evaluate of data mining classification decision model to take timely emergency measures, but also compare the mining accuracy models in the prediction of the objective of different data.

1. Introduction

Due to the continuous increase of load demand in power system, the gradual accession of new transmission lines, and the diversification of generation / load types and modes, the operation conditions of them may become increasingly intense in the future [1]-[3]. In this case, the interference in the system may make it more and more approachable to the limit of the stable operation. The complexity of the system and the contingency and uncertainty in the operation increase the performance requirements for the system stability prediction method. With the gradual promotion of wide area measurement technology in power system, data mining technology has gradually become an effective way to solve the problem of real-time evaluation of system stability. Through the data mining technology of off-line simulation training system of the large disturbance, can conveniently use real-time data synchronization phasor measurement unit is obtained in the case of an emergency system transient stability information and decision support, and the controller transient stability prediction information timely sent to the operators of power system or system level, to trigger emergency corresponding control action, thus preventing the system lose stability or to minimize the consequences of the accident[4]-[7].

At present, different data mining technologies have been combined with PMU for online transient stability prediction, such as artificial neural network, decision tree and support vector machine, and show a certain effect and robustness. Although the correlation method predicts accuracy in the range of 90% to 100%, the time required to make the decision is in the four weeks to 1 s after the interference is cleared. In addition, they use inconsistent testing methods. There are several typical test data sets of verification accuracy: (1) failure to use multiple faults during training;(2) a set of specially designed failures;(3) a set of failures in multiple system operation situations. For example, in the literature [8], three data sets are used to test the DT model, and multiple fault locations and system operation points are considered. In the literature [9], the SVM model is produced using six different fault tests. The accuracy of these models to present only shows that the data mining model to a certain degree of robustness, cannot describe the model in the overall performance of the complicated operation conditions, and the accuracy of each model presented is not convenient to compare.

In this paper, a general probability evaluation method is proposed to evaluate the accuracy of the data mining model in the on-line prediction of transient stability after the failure. Considering the distribution of failure probability, we first search for possible faults, then evaluate the accuracy according to the actual probability distribution of uncertain factors, which is more objective than Monte Carlo method. The evaluation results highlight the influence of a variety of uncertain factors on the accuracy of the system transient stability prediction, and determine the accuracy of the data mining model in real time. In this paper, the support vector machine (SVM) in data mining is used as an example to study the evaluation method. Without losing its generality, other data mining techniques can also be used to illustrate the evaluation method. This method can not only comprehensively evaluate the classification decisions made by data mining models in time of emergency, but also objectively compare the accuracy of different data mining models in prediction.

2. Probabilistic method for predicting probability of transient stability

2.1 Support vector machine model for transient stability

Assuming that $\{(x_1,y_1),\ldots,(x_l,y_l)\}$ is a training data set for the power system, which represents the input space of the sample, $y_i \subset R$ indicating the corresponding target value. The SVM estimation function can be expressed as:

$$f(x) = (w \cdot \phi(x)) + b \tag{1}$$

$\phi(x)$ is the input space to high dimension space of nonlinear mapping, the SVR is to practical problems through nonlinear mapping into a high dimensional feature space, structure in high dimensional feature space linear regression function to realize nonlinear regression function in the original space. The coefficient w and b can be estimated by minimizing regression risk, and the regression risk is expressed as:

$$R_{reg}(f) = C\sum_{i=0}^{l}\Gamma(f(x_i) - y_i) + \frac{1}{2}\|w\|^2 \tag{2}$$

$\Gamma(.)$ is a loss function, C is a constant, and you can get a vector w that's represented by the data point by minimizing function (2). And this paper USES the radial basis kernel function to calculate:

$$k(x_i, x) = \exp\{-\frac{|x - x_i|^2}{\sigma^2}\} \tag{3}$$

The value of w can be found according to Lagrange multiplier. For variable b, it can be obtained by using KKT condition:

$$\begin{aligned} a_i(\varepsilon + \zeta_i - y_i + (w, x_i) + b) = 0 \\ a_i^*(\varepsilon + \zeta_i^* + y_i - (w, x_i) - b) = 0 \end{aligned} \tag{4}$$

SVM applied the generated SVM model to the online process after the previous training automatically learned the potential relationship between predictive attributes and predicted targets. In practical application, when a fault is detected just after removal from the system (by voltage and monitoring system in the frequency or the circuit breaker status, extraction method etc.) the predicted attribute sample and input feature of the SVM model, the model identification and backward input features make reasonable decision. The consequences lead to wrong decisions the emergency is: when the stable condition prediction is unstable, will lead to control action is not necessary; and the unstable situation prediction is stable, it will lead to control action delay or action, resulting in system instability.

2.2 Probability evaluation of prediction accuracy of SVM model

In practice, many of the basic factors that affect transient stability in practice are probabilistic, including and fault location, fault type, fault removal equipment operation time, and failure occurs in the system operation status, such as load level and network topology, power generation capacity.

Therefore, probabilistic method is the most appropriate way to establish an SVM model to predict the accuracy of the accuracy.

In order to improve the performance of the algorithm, this paper only used 5 probabilistic factors in the test process: (1) fault removal time;(2) fault location;(3) failure type;(4) load level;(5) network topology. There are more factors to consider in practice. For a type of failure, in a possible location, under certain load level and a network topology, fault clearance time, in the form of random change, to generate a set of fault. Based on the test results of this group, the prediction accuracy A_{klmn} of a fault location k, fault type l, load level m and network topology n can be obtained. System failure types can be divided into single-phase grounding (LG), interphase fault (LL), two-phase fault (LLG) and three-phase (LLL) failure. Fault locations may occur in different nodes or in various transmission lines; And the system load changes all day and all year. Since the fault location, fault type, load level and network topology can be considered independent in most cases, probability prediction accuracy A can be obtained according to probability theory.

$$P_{klmn} = P_k \cdot P_l \cdot P_m \cdot P_n \tag{5}$$

P_k, P_l, P_m and P_n respectively indicate the probability of fault location, fault type, load level and network topology.

For all the possible failure location (k), all the possible failure type (l), all possible load level (m), and all possible network topology (n), probability prediction accuracy A can be calculated by (6):

$$A = \sum_{k=1}^{K}\sum_{l=1}^{L}\sum_{m=1}^{M}\sum_{n=1}^{N} A_{klmn} \cdot P_{klmn} \tag{6}$$

The probabilistic prediction accuracy expression (6) can be easily extended to other uncertainties, including power generation mode, load model parameters, and the presence and output of new energy generation. The extension will make the formula (5) and (6) be modified separately (5) to the right, including the probability of more factors and the sum of (6) on the right side.The expansion will produce different quantitative results and more computation for A, but the method of evaluation will remain unchanged.

3. Case study

The test system used in this study is the IEEE-39 bus system. Specific set up 20 fault locations, four kinds of fault types, six system load levels and nine kinds of network topology, the simulation tests are included in each group have 11 automatic removal of fault, fault clearance time from 0.09 s to 0.09 s (changes along with the increment of 0.02 s). After the failure of each fault, 10 s simulation, a total of 4,521 sets of tests, including 6,0378 different failures.

Figure 1 shows the prediction probability accuracy curve of 20 different fault locations in all fault types, load levels and network topologies. Based on the probability distribution of the test system, the fault location has less influence on the prediction probability accuracy. The prediction accuracy of the data based on the immediate sampling after the fault resection is above 94.5%. After 0.2 s, all 20 accuracies are over 97.3%, and the differences between them are small.

Figure 2 shows the prediction probability curve of four different types of failure types in all fault location, load level and network topology. In fact, in the test system, because the design of the system itself can withstand the most serious situation, unbalanced failure often leads to the problem of unstable system. As shown in table 6.6, the number of unstable simulation results (or percentages) increases gradually from single-phase grounding to three-phase fault. Therefore, the line at the top of figure 2 is 100% in all 5.01 s, indicating that the DT model is highly reliable for predicting the most common type fault of single-phase grounding. For two phase fault and two phase ground fault of the two relatively more easily lead to unstable type, fault immediately after removal of sampling data are used to get the probability forecast accuracy is about 98%, using fault removal after 0.5 s sampling data of the probability prediction accuracy is over 99%.For the most serious but probability of the smallest three-phase fault, lowest probability forecast accuracy, immediately after the failure to remove the forecast accuracy can reach 93%, only when the fault removal of 0.5 s increased to more than 98%.

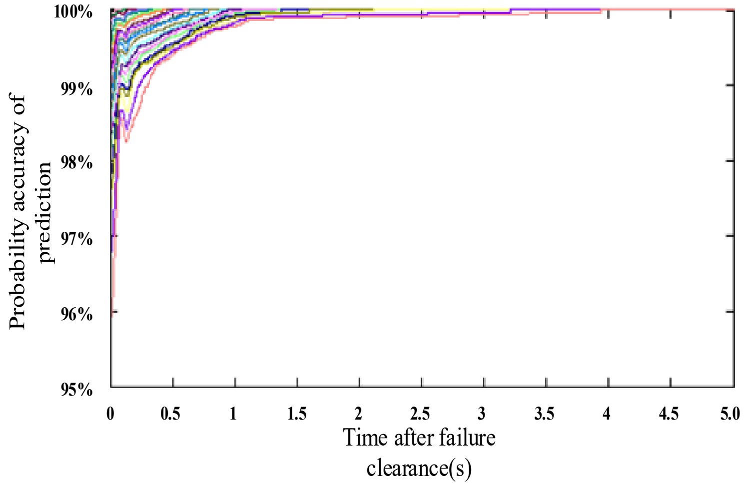

Figure 1. Probability accuracy of different fault locations

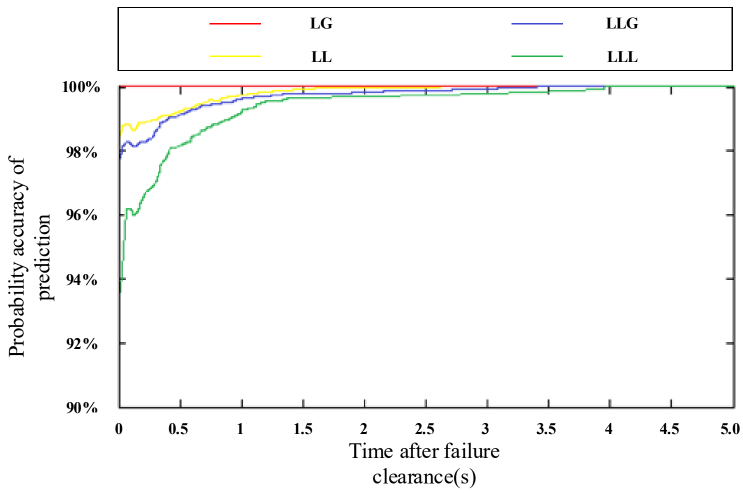

Figure 2. Probability accuracy of different fault types

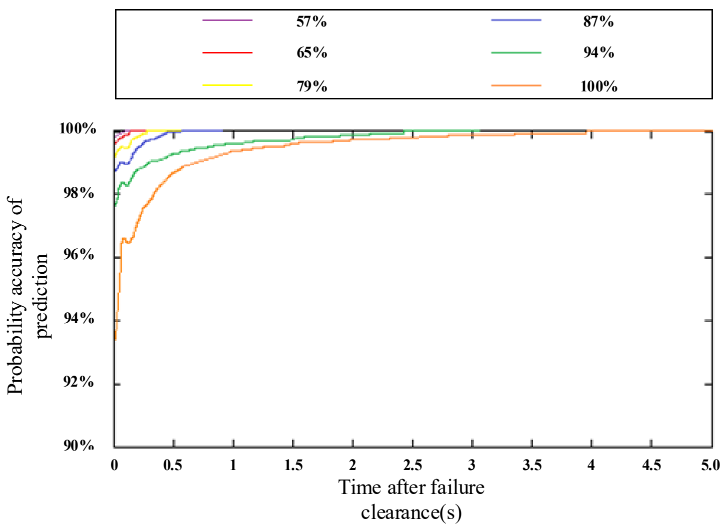

Figure 3. Probability accuracy of different load levels

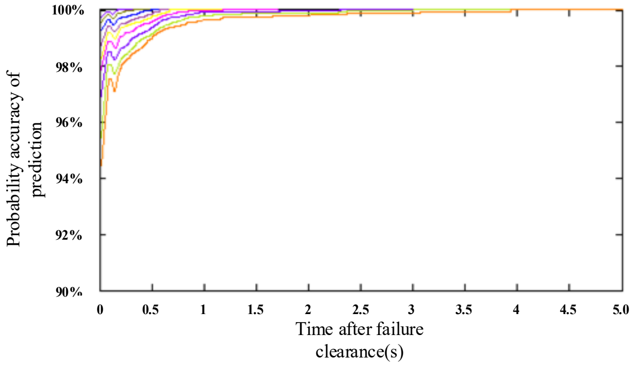

Figure 4. Probability accuracy of different topology

Figure 3 shows the predicted probability accuracy curve of the load level of six different systems in all fault locations, fault types and network topologies.When the load level is low, the system runs away from the stability limit, and the result of unstable simulation results is small.When the relative load level from 0.57 to 0.87, based on the failure sampling data immediately after removal of the accuracy of the resulting probability were higher than 99%, this shows that the SVM model under normal load level has the very high reliability prediction.When the relative load level is 0.94, the prediction accuracy of the data based on the 0.2s sampling based on the fault removal is also more than 98%.When the system is in extremely adverse operating conditions, the maximum load levels, about 33% of the fault is not stable, based on the failure probability forecast accuracy of sampling data immediately after removal of only 93%, and it was not until 0.4 s will rise to 98%.

Figure 4 shows the prediction probability curve of nine different network topologies in all fault location, fault type and load level. It can be seen that based on the probability distribution of the test system, the network topology has less influence on the prediction probability accuracy.The prediction accuracy of the data based on the immediate sampling after the fault resection is above 94%.After the fault resection was 0.3 s, all 9 accuracy was over 98%, and the differences between them were small.

4. Conclusions

In this chapter, a general probabilistic evaluation method is proposed to evaluate the accuracy of the data mining model. Evaluation method on the basis of the data mining model, considering the fault time, fault type, fault location, load level and network topology the probability distribution of five factors, comprehensive evaluation of the accuracy of the data mining model. According to the above analysis and test results, the proposed evaluation method compared with the existing research, advantage lies in: (1) based on the basic factors influencing the transient stability and the failure probability distribution of comprehensive consideration, can more fully and objectively evaluate the accuracy of the data mining model.(2) it is independent of the data mining model used, and therefore has universal applicability, which can be fairly compared with the accuracy of different data mining models. In addition, based on the probability of DT evaluation test results show that the method can consider various factors that affect the probability distribution of the premise, in a very short period of time, has the high accuracy of prediction results.

References

[1] Lv K. Study on pharmaceutical database management based on data mining technology [J]. Journal of Information and Computational Science, 2015, 12(8): 2979-2986.

[2] Stefanakis E, Li S N, Dragievi S. Advances in geospatial statistical modelling, analysis and data mining[J]. Geomatica, 2015, 69(3): 267-268.

[3] X. Liu, Y. Zhang and K. Y. Lee, "Coordinated Distributed MPC for Load Frequency Control of

Power System With Wind Farms," in IEEE Transactions on Industrial Electronics, vol. 64, no. 6, pp. 5140-5150, June 2017.

[4] Omi T, Kakisaka H, Iwamoto S. Transient stability multi swing step-out prediction with online data mining[J]. IEEJ Transactions on Power and Energy, 2016, 136(2): 137-144.

[5] Swetapadma A, Yadav A. Data-mining-based fault during power swing identification in power transmission system [J]. IET Science, Measurement and Technology, 2016, 10(2): 130-139.

[6] Rana M, Koprinska I. Forecasting electricity load with advanced wavelet neural networks [J]. Neurocomputing, 2016, 182: 118-132.

[7] Ajjarapu V and Christy C. The continuation power flow: A tool for steady state voltage stability analysis [J]. IEEE Transactions on Power Systems, 1992, 7(1): 416-423.

[8] Rovnyak S, Kretsinger S, Thorp J, et al. Decision trees for real-time transient stability prediction[J]. IEEE Transactions on Power Systems, 1994, 9(3): 1417-1426.

[9] Rajapakse A D, Gomez F, Nanayakkara K, et al. Rotor angle instability prediction using post-disturbance voltage trajectories [J]. IEEE Transactions on Power Systems, 2010, 25(2): 947-956.