Probability Analysis of Voltage Collapse Based on Ranking Fuzzy Rough-set Load Clustering

Dongkun Han, Yao Zhang, Zhigang Wu
School of electric power
South China University of Technology
GuangZhou, China
dongkun.han@mail.scut.edu.cn

Abstract—**A novel non-mechanistic approach for voltage collapse is proposed. By incorporating the notions of ranking clustering and fuzzy weighting with rough-set clustering, a new algorism for load clustering is presented, and then the probability model of load distribution is constructed. In the basis of fuzzy regression analysis of load improving pattern, the most probable direction of load increasing is calculated by sampling the data of load climbing periods, and the corresponding distance from saddle-node bifurcation point on the singularity boundary to the original operation point is figured out. The conception of voltage stability distance is defined, and by undergoing the most likelihood estimation of the parameters of its distribution function and the Kolmogorov test to goodness-of-fit, the distribution function of voltage stability distance is formed. Accordingly, the technique of probability estimation, considering the rising pattern of active power and reactive power, is advanced. By applying the approach in IEEE RTS and comparing of different clustering methods and diverse probability means of voltage stability, the rationality and applicability of proposed method are attested.**

Keywords-rough-set; fuzzy-set; load clustering; voltage collapse; probability analysis; distribution fitting

I. INTRODUCTION

Voltage collapse is deemed as the weakness of holding capability of power system. In order to provide the probability of voltage collapse and consider the effects of random disturbs and uncertain alternation, the researches of uncertain analysis of voltage collapse are extremely urgent and crucial.

Though the effects of randomness and indeterminacy are considered in some researches [1] [2], they deal with load model as maximum value, and subjectively supposing the load improving mode in calculating the limit of voltage stability would generate a deviation to the unstable point the actual load improving mode corresponded.

In this paper, each load level occurring probability is fixed by proposed clustering method. And the most likely direction of load improving is found out by fuzzy regression algorithm. Moreover, this paper defines the voltage stability distance as a novel random variable, and its distribution function is ascertained by the most likelihood estimation of parameters (MLE) and the Kolmogorov test to goodness-of-fit, thus a model for probability analysis of voltage stability is constructed. This method is demonstrated in IEEE RTS.

II. RANKING FUZZY ROUGH-SET LOAD CLUSTERING

Since the rough-set and fuzzy-set possess dissimilar emphasis on obscure phenomenon, by joining the couple of methods together and importing the fuzzy equivalent relationship to rough-set approximations, we can earn a more expressive rough-set model for clustering [3]. By introducing the layer-by-layer point, we can search for the bunch centers step by step until clustering meets the termination conditions.

Suppose a non-empty set U is the universe of discourse, the sample data sets in U are $X = \{X_1, X_2, \cdots, X_M\}$. R is the equivalent relationship. For $x \in U$, $[x]_R$ is called R equivalent class made of element x. The upper approximation of X is defined as $U_R(X) = \{x \in U | [x]_R \subseteq X\}$; The lower approximation of X represented as $L_R(X) = \{x \in U | [x]_R \bigcap X \neq \phi\}$. The steps of ranking fuzzy rough-set clustering are as follows:

1) The selection of original clustering centers. The distance between sample X_i and X_j（i，j=1,2$\cdots$,M）use the square of Euclid distance, displaying as d_{ij}. By applying the idea of max-min method, the couple of points which obtain the biggest distance in the sample sets are chosen to be the original clustering centers.

2) The division of upper and lower clusters approximation sets. b_{ki} (k=1,2$\cdots$,N) is used to present the distance between X_i and its closest clustering center. For every b_{ki} ($b_{ki} \neq b_i$), if $b_{ki} - b_i \leq \varepsilon$, X_i belongs to L_k, otherwise there is a b_{ni} (n=1,2$\cdots$,N), enabling $b_{ni} - b_i > \varepsilon$, $X_i \in U_n$. ε is threshold, and U_k, L_k are respectively the upper and lower approximation of cluster k.

3) The fixing of clustering number. Due to emphasize the mutual similar relationship, the definition of fuzzy membership is quoted. The fuzzy membership of object X_i to cluster center C_k is w_{ki}:

$$w_{ki} = \frac{1}{\sum_{n=1}^{K} \left(\frac{b_{ki}}{b_{ni}} \right)^{\frac{2}{\gamma-1}}} \quad (1)$$

γ is the fuzzy factor. The boundary zone of number k cluster is $B_k = U_k - L_k$. The fuzzy rough-set clustering centers can be described:

$$C_k = \begin{cases} h_{low}\dfrac{\sum\limits_{X_i \in L_k} X_i(w_{ki})^\gamma}{\sum\limits_{X_i \in L_k}(w_{ki})^\gamma} + h_{up}\dfrac{\sum\limits_{X_i \in B_k} X_i(w_{ki})^\gamma}{\sum\limits_{X_i \in B_k}(w_{ki})^\gamma} & (if\ B_k \neq \phi) \\[6pt] \dfrac{\sum\limits_{X_i \in L_k} X_i}{|L_k|} & (otherwise) \end{cases} \quad (2)$$

h_{up}, h_{low} are respectively the upper and lower approximate weighting coefficients. $|L_k|$ is the cardinal number of L_k.

4) Ascertaining the layer-by-layer point. The remarkable characteristics of layer-by-layer point are the most likelihood to deviate cluster center and huge effects generated by upper approximations of other clusters. We define the dispersion T_i （$i=1,2\cdots,M$） as a quotient described by (3):

$$T_i = \frac{\sum\limits_{K=1}^{N} b_{ki}}{\sum\limits_{K=1 K \neq n}^{N}\|C_n - C_k\|^2} \quad (3)$$

The point that maintains the biggest dispersion is layer-by-layer point. Hence the cluster centers of current level and the layer-by-layer point can be used as the original cluster centers in next level. Once the termination conditions are met, the output should be exported. Otherwise, step 2 and step 3 ought to be repeated until the cluster centers do not change anymore.

5) Clustering termination. If the distance between the chosen layer-by-layer point and its cluster center is not more than the average distance between sample point and its cluster center, it is considered that the data space is divided into appropriate clusters which possess the proper dissimilarity amongst diverse clusters and similarity in the same cluster.

Via load clustering, the load level probabilistic distribution can be figured out. Given the sample number in U, the probability of every load level can be estimated as the occurring frequency of samples in each cluster.

III. FUZZY REGRESSION FITTING OF LOAD RISING

For the sake of determining the load improving mode, the power function of time should be constructed. And then its single power coefficient should be calculated and unitized. This belongs to a linear regression problem. Taking the active power fuzzy linear regression as example, the M groups of observational data are given: (P_i, t_i) $i=1,2\cdots,M$. The average regional active power P and the relevant factor t obtain a linear relationship:

$$P = At + B \quad (4)$$

In this fuzzy linear regression function, regression coefficient A and intercept B are fuzzy numbers. The departure between the fitting value and observing value is triggered by fuzziness of the function [4]. A, B are often defined as triangle fuzzy number $A(\delta, h)$, $B(\beta, l)$. The membership function of A is:

$$\mu_A(z) = \begin{cases} 1-|z-\delta|/h & (\delta - h \leq z \leq \delta + h) \\ 0 & (otherwise) \end{cases} \quad (5)$$

In this equation, δ is the center value of A, a plus h is the fuzzy range of A. By the same way, $\mu_B(z)$ can be assumed. On the ground of (4), the fuzzy function of can be described as (6). In order to find the most optimal fitting (4), the fuzzy ranges of fuzzy numbers should be their minimum. Meanwhile, the function are requested to cover all the observations P_i at a certain confidence level H, that is $\mu_P(P_i) \geqslant H$.

$$\mu_P(P_i) = \begin{cases} 1-\dfrac{|P_i - \delta t_i - \beta|}{h|t_i|+l} & \begin{pmatrix} if\ P \geq \delta t_i + \beta - h|t_i|-l \\ and\ P \leq \delta t_i + \beta + h|t_i|+l \end{pmatrix} \\ 0 & (otherwise) \end{cases} \quad (6)$$

It can be transformed to (7):

$$\delta t_i + \beta - (1-H)(h|t_i|+l) \leq P_i \leq \delta t_i + \beta + (1-H)(h|t_i|+l) \quad (7)$$

Therefore, the problem of (1) can be transmuted to a linear programming problem:

$$\begin{cases} \min J = \min(h+l) \\ s.t.\ \ \delta t_i + \beta - (1-H)(h|t_i|+l) \leq P_i \\ \quad\ \ \delta t_i + \beta + (1-H)(h|t_i|+l) \geq P_i \end{cases} \quad (8)$$

As to this linear programming, the simplex method can work it out easily. The optimal solution $c*$ is the rising coefficient of active power of this area, by the same way, we can get the reactive power rising coefficient $e*$. Assume the number of PV nodes is N_{PV}, the number of PQ nodes is N_{PQ}, the regional power rising direction is:

$$\eta^* = \frac{1}{\sqrt{N_P c^{*2} + N_Q e^{*2}}}\left[c_1^*, c_2^*, \cdots, c_{N_P}^*, e_1^*, e_2^*, \cdots, e_{N_Q}^*\right]^T \quad (9)$$

And N_P equals N_{PQ} plus N_{PV}, and N_Q equals N_{PQ}.

IV. FITTING OF DISTRIBUTION VOLTAGE STABILITY DISTANCE AND ITS GOODNESS TEST

According to the static local bifurcation theory, in the N-dimensional space, by improving in the direction η, the initial operating point arrives at the singular hyperplane boundary. Suppose the state variable is $X=(\theta, U)^T$, multi-dimensional bifurcation parameter vector is $\psi=(P, Q)^T$, thus the initial system power flow can be described as:

$$f(X, \psi)=gX\text{-}\psi \quad (10)$$

Improving load along any load rising direction, the operating point can reach the voltage instability critical point $(X, \psi) \in$ S, and $\psi=\psi^0+\zeta\eta$. Therefore, the Euclid distance from initial operating point (X^0, ψ^0) to the voltage collapse point (X, ψ) along the load rising direction η can be calculated:

$$\zeta=\|\psi\text{-}\psi^0\| \quad (11)$$

This paper defines ζ as the voltage stability distance, and the voltage stability distance corresponded by the most probable load rising direction η^* is ζ^*.

The power margin of each node meets the Normal Distribution [1]. Provided the number of sample is large enough, ζ always satisfies fellow distribution: the Normal distribution, the Lognormal Distribution and the Gamma

distribution. As a complicated hypothesis testing, firstly we set three hypothesis of the distribution of ζ. We assume Normal Distribution, Lognormal Distribution and Gamma distribution as H_{01}, H_{02} and H_{03}, respectively.

Assume that the distribution function of X is $F(x)$, and $F_0(x)$ is a specific distribution function, and $F_n(x)$ is experience distribution function [5]. Considering the statistic D_n:

$$D_n = \sup |F_n(x) - F_0(x)| \qquad (12)$$

It describes the difference between $F_n(x)$ and $F_0(x)$, and we call it ramification. In the condition that the hypothesis of sample points meeting F_0 is acceptable, we can find out the critical value λ to each significant level α and test whether $P(D_n > \lambda) = \alpha$ is met. This kind of test is named Kolmogorov test.
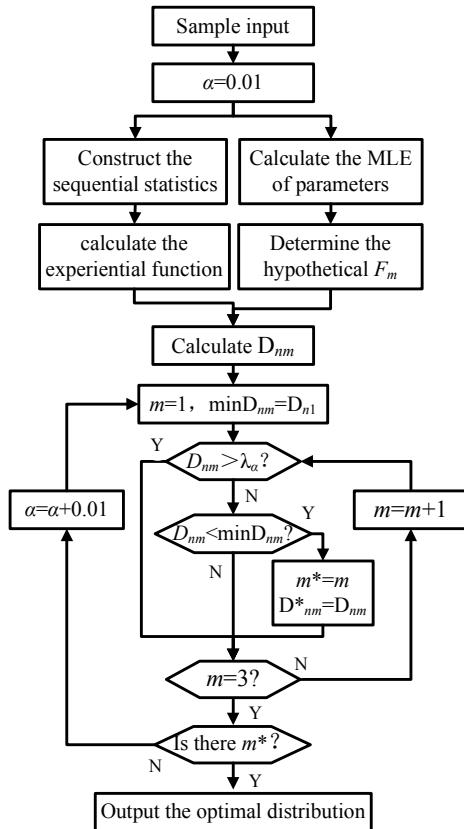

Figure 1. Example of a figure caption. (figure caption)

Considering the situation that the distribution function fulfilled certain α is not unique, the hypothetic function with the smallest D_n should be selected as the optimal distribution function. As to the condition that no hypothetic function go through the test under certain α, the significant level should be expanded and the test should be repeated under new α. Making $m=1, 2, 3$ substitute for Normal distribution, Lognormal distribution and Gamma distribution, the Flowchart of test of distribution fitting goodness is as Fig. 1 showing:

On the foundation of optimal fitting distribution, the probability mode of voltage collapse can be constructed. In the light of the result of load ranking fuzzy rough-set clustering,

the probability P_{Li} of the occurring of each load level can be figured out. In addition, we can find out the probability P_S of the condition that regional load climbs continuously. Then the probable load rising direction and the optimal distribution function of ζ can be figured out. Assume that the voltage collapse breaks out when the voltage stability distance are smaller than the most probable voltage stability distance, the probability of voltage collapse can be depicted as: $P_{Ci} = F^*_{im}(\zeta < \zeta^*)$. We can described it in another way:

$$P_C = \sum_{i=1}^{N} P_S P_{Li} P_{Ci} \qquad (13)$$

V. EXAMPLE ANALYSIS

In this paper, RTS is taken to be the examination of this method. The example supplies with a 24-nodes-system and its 8736h load operating data [6]. The calculating parameters are selected as fellows: $h_{up}=0.9$; $h_{low}=0.1$; $\gamma=1.5$; H=0.5.

In order to highlight the superiority of proposed method, the definitions of Compactness, Separation and Condensation are quoted [5]:

$$Com(C) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{|C_k|}\sum_{i=1}^{|C_k|}\sum_{j=1}^{|C_k|} sim(X_i, X_j) \qquad (14)$$

$$Sep(C) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{|C_i||C_i|}\sum_{p=1}^{|C_i|}\sum_{q=1}^{|C_j|} sim(X_p, X_q) \qquad (15)$$

$$Con(C) = Com(C)/Sep(C) \qquad (16)$$

And $Sim(X_i, X_j)$ is similarity function. When clusters obtain more compactness and less separation, the quality of clustering is better. On account of (16), the condensation directly embodies the quality of clustering. Comparing the traditional methods, K-means and Fuzzy K-means, with proposed method, the evaluations of each method can be found in Tab. 1.

TABLE I. EVALUATION OF DIFFERENT CLUSTERING METHODS

Cluster attribute	K-means		FMK		Proposed methods
	N=9	N=16	N=9	N=12	N=9
Com(C)	0.8272	0.8461	0.8297	0.8326	0.7759
Sep(C)	0.1191	0.1193	0.1173	0.1174	0.1094
Con(C)	6.945	7.0922	7.0733	7.0920	7.0923

By ranking fuzzy rough-set clustering, the sample space is automatically divided into 9 clusters. When the clustering number is the same, the $Con(C)$ of proposed methods is greater than that of K-means and FMK, resulting in a better clustering quality. Comparing the value of $Com(C)$ and $Sep(C)$, the advantage of rough-set focus on diminishing the separation, thus it gains a higher condensation. Whereas the deviation of members in the same cluster is greater than the traditional methods, on the grounds that the objective function of K-means and FKM partially emphasize on the minimum of member distances in the same cluster, neglecting to increase the deviations between different clusters. Rising the clustering

number step-by-step, respectively at N=16, N=12, the $Con(C)$ of K-means and FMK are still weaker than proposed method.

By choosing seven successive load levels whose apparent power are continuously improving, they are deemed as the load rising period. Choosing this kind of Groups repeated, we can figure out the regression coefficients by fuzzy linear regression. The results of calculation are in Tab.2.

TABLE II.   RESULTS OF FUZZY LINEAR REGRESSION

Power category	Regression results		
	Average of Regression Coefficient	*Average of intercepts*	*Regression function*
Active power	(1.062,0.0018)	(78.09,0.052)	P= 78.09+1.062t
Reactive power	(1.029,0.0007)	(15.08,0.013)	Q=15.08+1.029t

According to Tab.2, the fuzzy range of the fuzzy number is extremely small. By using formula (9), the fuzzy centers are directly utilized to calculate the most probable load rising direction η^*. By reckoning membership of each P_i to test the fitting model, every $\mu_P(P_i)$ is more than 0.532, which clearly shows the fitting is acceptable.

Based on the static local bifurcation analysis, under diverse load rising direction, the value of random variable ζ can be calculated. According to Fig.1, the optimal distribution function can be found out in the test of goodness-of-fit. The results of optimal fitting for each cluster are in Tab.3. F_1, F_2 and F_3 respectively indicate the Normal distribution, Lognormal Distribution and Gamma Distribution.

TABLE III.   RESULTS OF KOLMOGOROV TEST TO GOODNESS-OF-FIT

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
Significant level	0.08	0.12	0.18	0.12	0.1	0.12	0.14	0.2	0.1
Optimal distribution	F_1	F_1	F_3	F_2	F_2	F_1	F_1	F_3	F_2
Collapse probability	0.13	0.17	0.06	0.32	0.21	0.17	0.27	0.43	0.31

Obviously, as the optimal distribution, F_1 has the largest occurring frequency, and followed by F_2. This phenomenon manifests that the sequence of fitting test is properly conformed to experience. At a high significant level, F_3 occurs, which reflects that F_3 brings about a rather big fitting deviation. Only in low confidence level can F_3 be the optimal distribution.

TABLE IV.   COMPARISON OF DIFFERENT SCHEMES

	N	ζ^P	Frequency of distribution			P_C
			F_1	F_2	F_3	
Method1	1	127.648	1	0	0	3.1729×10^{-5}
Method2	9	117.932	3	4	2	4.2568×10^{-5}
Proposed method	9	163.729	4	3	2	2.5773×10^{-5}

On the bases of gaining optimal distribution, the system voltage collapse probability can be figured out by (13). Comparing the proposed method with other two methods as fellows: (Method 1) system load is determined to be peak load and the load rising mode is chosen to be η^*; (Method 2) this method utilizes proposed clustering algorithm, while it choose the nearest instability point corresponding direction or its similar directions as the random rising direction. Results of comparison are in Tab.4.

On the perspective of optimal fitting, method 2 is similar to proposed method. However, according to method 2, F_2 obtains the biggest frequency, which indicates that some disturbances in method 2 have increased the irregular degree of distribution. Clearly, both method 1 and method 2 generate a higher probability of voltage collapse than proposed method, in that initial operating point of method 1 is adjacent to the singular boundary S, while for method 2, the operating point moves in corresponding direction of the nearest instable voltage point, making ζ smaller. In two diverse ways, method 1 and method 2 separately dwindle ζ, leading to a conservative analysis. Thus, the proposed method is more reasonable and practical.

VI.   CONCLUSIONS

This paper integrates the fuzzy theory with rough-set clustering, by pursuing the layer-by-layer point, making a novel clustering algorithm. Then the load rising mode undergoes the fuzzy regression analysis to reckon the most probable load rising direction. On the condition of the formation of a new random variable, the voltage stability distance, the optimal fitting distribution function is calculated by Kolmogorov test. Thus a new voltage collapse probability method is set up. Applying this method to RTS 24 nodes system, the results manifest: ranking fuzzy rough-set clustering appreciates a bigger condensation. Due to the methods comparison, the proposed method is effective and reasonable.

REFERENCES

[1]   H. Wan, J. D. McCalley, V. Vittal. "Risk based voltage security assessment," IEEE Trans on Power Systems, vol. 15, pp. 1247–1254, Novemver 2000.

[2]   A. Schellenberg, W. Rosehart, J. A. Aguado. "Cumulant-based stochastic nonlinear programming for variance constrained voltage stability analysis of power systems," IEEE Trans on power systems, vol. 21, pp. 579–585, May 2006.

[3]   Liang X. Data Mining Algorithms and Applications. Beijing: Peking University Press, 2006, pp:236–237.

[4]   K. B. Song, Y. S. Baek, D. H. Hong. "Short-Term Load Forecasting for the holidays Using Fuzzy linear Regression Method," IEEE Trans on Power Systems, vol. 20, pp. 96–101, Febraury 2005.

[5]   J. D. Chen, Z. G. Zheng. Probability and Statistics. Beijing: Peking University Press, 2007, pp. 423–428.

[6]   Reliability Test System Task Force. "IEEE reliability test system," IEEE Trans on Power Apparatus and Systems, vol.98, pp. 2047–2054, November 1979.