International Journal of Fuzzy Logic and Intelligent Systems, vol. 8, no. 1, March 2008 pp. 18-23

Application of Similarity Measure for Fuzzy C-Means Clustering to Power System Management

Dong-Hyuk Park*, Soorok Ryu, Park Hyun Jeong*** and Sang H. Lee***

***School of Mechatronics, Changwon National University**
****Department of Industrial and Applied mathematics, Kyungbuk National University**
*****Department of Mathematics Education, Ewha Woman University**

Abstract

A FCM with locational price and regional information between locations are proposed in this paper. Any point in a networked system has its own values indicating the physical characteristics of that networked system and regional information at the same time. The similarity measure used for FCM in this paper is defined through the system-wide characteristic values at each point. To avoid the grouping of geometrically distant locations with similar measures, the locational information are properly considered and incorporated in the proposed similarity measure. We have verified that the proposed measure has produced proper classification of a networked system, followed by an example of a networked electricity system.

Key Words : FCM Clustering, power system, similarity measure

1. Introduction

The regional operation and planning of networked systems is needed for the efficient and economical management of the systems. Hence the research of system coherency has been made by numerous researchers [1-4]. However, studies are emphasized on the dynamics grouping. At this point, we need novel approach to participate the total system into several regions considering locational information, such as locational cost, loss, regional distances, and so on. In this paper, grouping the locations in a networked system with similar locational prices has been proposed considering the regional coherency. Locational prices in networked system imply the price at which the good is consumed at each location. Due to the physical characteristics of the transmission network of the systems, the good is lost as it is transmitted from supplying locations to consuming locations, and an additional supply must be provided to compensate the loss. Also, the transmission network of the systems has a capacity limitation preventing full uses of cheap production. Therefore, location prices at each point or node, is differently decided depending the network topology and supply/demand configuration. Similarity measure has been known as the complementary meaning of the distance measure [5-9]. Hence, we consider the partitioning measure not only similarity measure but also regional information i.e. distance measure. In the previous literatures, we had constructed similarity measure through distance measure or fuzzy entropy function [10]. Well known-Hamming distance was used to construct fuzzy entropy, so we composed the fuzzy entropy function through Hamming distance measure. With only similarity measure, we can obtain unpractical results, which are far locating points. Hence we add the regional information to complete modified similarity measure.

In the next section, FCM, similarity measure and modified similarity are introduced. Also modified similarity measure is constructed with the regional information. In Section 3, illustrative examples are shown. In the example, we obtain proper partitioning, which consider both similarity and regional information. Conclusions are followed in Section 4.

Manuscrip received Oct. 17, 2007; revised . Mar. 10, 2007.
Corresponding author : Sang H. Lee
This work has been supported by 2nd BK21 Program, which is funded by KRF(Korea Research Foundation).

2. Fuzzy C-means clustering and similarity measure

Fuzzy C-means clustering was proposed by Bezdek in 1973 as an improvement over HCM(Hard C-means)[10]. FCM play a roll of partitioning arbitrary vectors into fuzzy groups, also it finds a cluster center for each group such that a cost function of similarity measure is maximized, or dissimilarity measure is minimized.

2.1 Preliminaries

We will illustrate the FCM result briefly reference [11]. Membership matrix U is satisfied as follows

$$\sum_{i=1}^{c} u_{i,j} = 1, \forall j = 1,...n. \qquad (1)$$

The cost function for FCM is constructed by

$$J(U,c_1,...,c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{i,j}^m d \qquad (2)$$

where u_{ij} is between 0 and 1, c_i is the center of fuzzy group i, $d_{ij} = |c_i - c_j|$ is the Euclidean distance between i-th cluster center and the j-the data point x_j, and is the weighting value. With Lagrange multiplier, the necessary conditions for equation (2) to reach a minimum are reference [11]

$$c_i = \frac{\sum_{j=1}^{n} u_{i,j}^m x_j}{\sum_{i=1}^{n} u_{i,j}^m} \quad \text{and} \quad u_{i,j} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{i,j}}{d_{k,j}}\right)^{2/(m-1)}}$$

Now for minimizing of equation (2), the less distance is, the smaller cost function become. Hence distance means the similarity between two data points. Finding similarity is determined from the types of data, time series signal, image, sound, etc. Now we need proper similarity measure.

2.2 Similarity measure with distance function

By the similarity measure definition of Liu [5], similarity between set A and set B represent that common area of two fuzzy membership functions are proportional to the degree of similarity. We propose the following theorem as a similarity measure.

Theorem 2.1 For any set $A,B \in F(X)$, or $F(X)$, if d satisfies Hamming distance measure, then

$$s(A,B) = 2 - d((A \cap B),[1]) - d((A \cup B),[0]) \qquad (3)$$

is the similarity measure between set A and set B.

Proof can be found in reference [9]. Another similarity measure is illustrated in the following theorem.

Theorem 2.2 For any set A, $B \in F(X)$, if d satisfies Hamming distance measure, then

$$s(A,B) = 1 - d((A \cap B^c),[0]) - d((A \cup B^c),[1]) \qquad (4)$$

is the similarity measure between set A and set B.

Proofs are also shown similarly as Theorem 2.1.

Of course there are numerous similarity measures satisfying similarity measure definition. With the designed similarity various similar characteristic properties can be grouped by replacing d_{ij} in equation (2) into proposed similarities equation (3) and equation (4). If two or more measures are necessary for grouping, we have to consider more similarity measure.

2.3 New Similarity with Regional Information

In the previous section we have derived the modified similarity measures which satisfying the definition of similarity. To apply FCM with $d_{ij} = |c_i - c_j|$, it is required that has to satisfy similarity property. Hence we replace d_{ij} into the proposed similarity measure equation (3) and equation (4). However proposed similarity measure can group for the point that having similar characteristic values. For the large scale system whose similar measure values are close, however they are located far away. Then it is not realistic to gather even though they have similar valued measure. So we need another characteristic values considering regional information. Besides of equation (3) and equation (4), we consider

$$s_2(A,B) = 2/(1 + \text{distance}(A,B)) \qquad (5)$$

where distance is the geometrical distance value.

We consider the combined similarity measure as

$$s(A,B) = w_1 s_1(A,B) + w_2 s_2(A,B) \qquad (6)$$

where, $s(A,B) = 2 - d((A \cap B),[1]) - d((A \cup B),[0])$ or $s(A,B) = 1 - d((A \cap B^c),[0]) - d((A \cup B^c),[1])$, w_1 and w_2 are the weighting values.

We can verify the usefulness of equation (6) as follows, properties of $s_1(A,B)$ are proved in Theorem 2.1 and 2.2. Usefulness for the similarity of $s_2(A,B)$ can be verified as follows:

Commutative values of distance are same, hence (S1) is easily shown. From (S2), distance D of D^c and is the longest, hence $s_2(D,D^c)$ is the minimum value. For all $\forall A,B \in F(X)$, , inequality of (S3) is proved by

$$\begin{aligned} s_2(A,B) &= 2/(1+\text{distance}(A,B)) \\ &\le 2/(1+\text{distance}(D,D)) \\ &= s_2(D,D) \end{aligned}$$

In the above

distance(D,D) is the smallest value, i.e., zero.

So (S3) can be verified.

Finally, (S4) is $\forall A,B,C \in F(X)$,

, and satisfy triangular points, then

$$\begin{aligned} s_2(A,B) &= 2/(1+\text{distance}(A,B)) \\ &\le 2/(1+\text{distance}(A,C)) \\ &= s_2(A,C) \end{aligned}$$

where distance(A,C) is longer than distance(A,B).

Similarly,

$$\begin{aligned} s_2(B,C) &= 2/(1+\text{distance}(B,C)) \\ &\le 2/(1+\text{distance}(A,C)) \\ &= s_2(A,C) \end{aligned}$$

is satisfied. Hence we can verify that $s_2(A,B)$ satisfies the similarity measure definition. Then we use equation (6) as the modified similarity measure for the measuring of particular points which have characteristic values and regional information at the same time. Modified similarity measure is applied in the following example.

3. Partitioning Power System

With our proposed FCM, we replace $d_{ij} = |c_i - x_j|$ in equation (2) to $s(A,B)$ in equation (6), and illustrate the system which has characteristic values and regional information at the same time. As a test system, we consider the IEEE reliability test system which is prepared by the reliability test system task force of the application of probability methods subcommittee on 1996 reference [12]. In Fig. 1, 39 buses and 10 generators are contained, and each bus has its own locational price and locational information as Table 1. In networked electricity systems, due to the physical characteristics of the electricity transmission network, electricity is lost when it is transmitted from supplying nodes to consuming nodes, and additional generation must be supplied to provide energy in excess of that consumed by customers. Moreover, the capacity limitation of the transmission network of electricity systems prevents full uses of system wide cheap electricity. Therefore, electricity price at each node, i.e., the price at which the electricity is consumed at each node is differently decided depending the network topology and energy configuration. The electricity prices at each node is defined as locational prices at each node and the locational prices represent the locational value of energy, which includes the cost of electricity and the cost of delivering it, i.e., the delivery losses and network congestion. In Table 1 each locational price per kWh are illustrated for the 39 buses, and per unit geometrical information for each nodes are also shown. Locational prices of each nodes are from 28.53 to 55.00, and 39 locational information are represented through 2-dimensional plane at which plane is assumed to be flat. However just one information is not sufficient to solve grouping problem. For example, BUS 16 has exact locational price 45.84 with BUS 23. However their locational informations are (6.5, 4) and (11.1, 2.8), respectively. Therefore we consider not only locational price but also geometrical information for the purpose partitioning power interconnected system properly.

With the combined similarity measure equation (6) we construct similarity for each nodes. (1,1) represent the modified similarity BUS 1 to BUS 1, it is natural to satisfy 2.4 = 0.4 + 2. Maximum similarity of $s_1(A,B)$ is 0.4, and maximum value of $s_2(A,B)$ is 2. Because of discrepancy of locational prices and location we adjust similarity measure as follows.

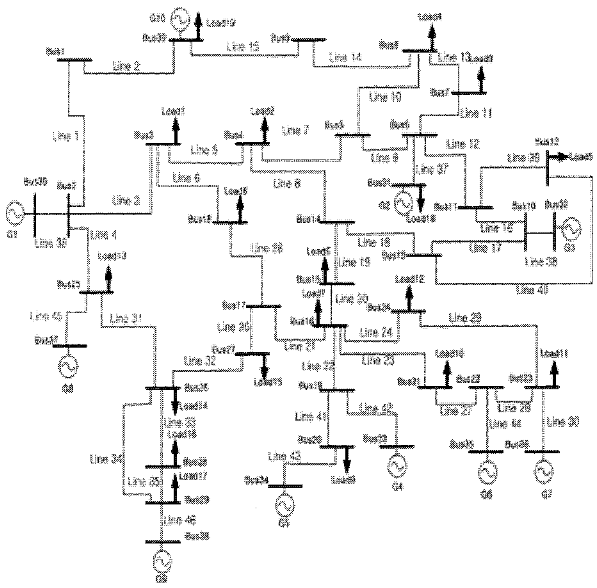


Fig. 1. A networked electricity system

Table 1. Locational prices and locations at each node

BUS	locational price ($/kWh)	Location (per unit)	BUS	locational price ($/kWh)	Location (per unit)
BUS1	29.21	(0.9, 9)	BUS20	45.84	(6.9, 1.7)
BUS2	28.53	(0.6, 6.2)	BUS21	45.84	(8.7, 2.8)
BUS3	31.40	(3, 7.5)	BUS22	45.84	(10, 2.8)
BUS4	32.78	(4.7, 7.5)	BUS23	45.84	(11.1, 2.8)
BUS5	37.57	(7, 7.6)	BUS24	45.84	(8.2, 4.3)
BUS6	38.26	(8.5, 7.6)	BUS25	24.98	(1.4, 4.7)
BUS7	37.81	(9.6, 8.4)	BUS26	55.00	(2.7, 3)
BUS8	37.35	(8.5, 9.1)	BUS27	51.45	(4.6, 3.5)
BUS9	30.56	(6.1, 9.5)	BUS28	55.00	(2.7, 1.5)
BUS10	40.00	(10.8, 5.8)	BUS29	55.00	(2.7, 0.8)
BUS11	39.42	(9.7, 6.3)	BUS30	28.53	(0, 6.2)
BUS12	40.00	(11.1. 7.1)	BUS31	38.26	(8.3, 6.6)
BUS13	40.58	(8.5, 5.5)	BUS32	40.00	(11.3, 5.8)
BUS14	41.74	(6.6, 6)	BUS33	45.84	(8, 1.7)
BUS15	43.79	(6.6, 4.9)	BUS34	45.84	(5.5, 1)
BUS16	45.84	(6.5, 4)	BUS35	45.84	(10, 1.6)
BUS17	47.90	(5, 4.5)	BUS36	45.84	(11.1, 1.6)
BUS18	46.40	(4.2, 6)	BUS37	24.98	(0.7, 3.7)
BUS19	45.84	(6.9, 2.8)	BUS38	55.00	(2.7, 0)
			BUS39	29.88	(3.4, 9.5)

$$s_1(A,B) = 1.4 - d((A \cap B),[1]) - d((A \cup B),[0])$$

Maximum difference of locational price 55.00 - 28.58 = 26.47 and maximum distance BUS12 to BUS2 are normalized. With this modification the similarity of BUS1 and BUS2 or BUS 2 and BUS 1 is 2.08, successively the 39×39 symmetric matrix can be constructed. The part of the 39×39 full matrix is shown in Table 2. In this calculation, we have assigned weighting values w_1 is 1 and w_2 is 0.2. If we consider the regional information more, then we can adjust value of w_2.

Table 2. Combined similarity measures between 39 nodes

	BUS1	BUS2	BUS3	BUS4	BUS5	BUS6	BUS7	BUS8	BUS9	BUS10
BUS1	2.40	2.08	2.03	1.95	1.75	1.72	1.73	1.75	2.01	1.64
BUS2	2.08	2.40	2.00	1.92	1.72	1.69	1.70	1.72	1.98	1.62
BUS3	2.03	2.00	2.40	2.10	1.86	1.81	1.82	1.84	2.05	1.73
BUS4	1.95	1.92	2.10	2.40	1.95	1.88	1.88	1.91	2.04	1.79
BUS5	1.75	1.72	1.86	1.95	2.40	2.13	2.10	2.12	1.87	1.99
BUS6	1.72	1.69	1.81	1.88	2.13	2.40	2.15	2.13	1.82	2.04
BUS7	1.73	1.70	1.82	1.88	2.10	2.15	2.40	2.16	1.82	2.02
BUS8	1.75	1.72	1.84	1.91	2.12	2.13	2.16	2.40	1.87	1.98
BUS9	2.01	1.98	2.05	2.04	1.87	1.82	1.82	1.87	2.40	1.71
BUS10	1.64	1.62	1.73	1.79	1.99	2.04	2.02	1.98	1.71	2.40
..	..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..	..
	BUS31	BUS32	BUS33	BUS34	BUS35	BUS36	BUS37	BUS38	BUS39	
BUS31	2.40	2.03	1.79	1.78	1.79	1.78	1.56	1.43	1.76	
BUS32	2.03	2.40	1.85	1.83	1.86	1.86	1.49	1.49	1.67	
BUS33	1.79	1.85	2.40	2.11	2.13	2.10	1.29	1.73	1.46	
BUS34	1.78	1.83	2.11	2.40	2.07	2.06	1.30	1.77	1.46	
BUS35	1.79	1.86	2.13	2.07	2.40	2.19	1.28	1.71	1.46	
BUS36	1.78	1.86	2.10	2.06	2.19	2.40	1.28	1.71	1.45	
BUS37	1.56	1.49	1.29	1.30	1.28	1.28	2.40	0.98	1.88	
BUS38	1.43	1.49	1.73	1.77	1.71	1.71	0.98	2.40	1.12	
BUS39	1.76	1.67	1.46	1.46	1.46	1.45	1.88	1.12	2.40	

Through this result, we have grouped the given networked electricity systems with similarity values as in Table 3 and 4. These results are obtained with similarity level computations, we do not consider FCM yet.

Table 3. Grouping with similarity values (similarity = 1.8)

singularity	3,4,5,6,9,11,15,16,17,24,27
Group1	13,14
Group2	1,2,30,39,25,37
Group3	7,8
Group4	10,12,32,31
Group5	18,19,20,21,22,23
Group6	26,28,29,38
Group7	33,34,35,36

Table 4. Grouping with similarity values (similarity = 1.2)

singularity	3,25,26,27,28,29,37,38
Group1	1,2,9,30,39
Group2	4,5,6,7,8,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,31,32,33,34,35,36

These Table 3 and Table 4 indicate that the lower similarity value is, the fewer group number is. As we mentioned before, weighting value can be adjusted to consider more locational price or regional information.

For FCM consideration at first, we showed the 39 buses as 3 groups, and the result is illustrated in Fig. 2. 39 buses are shown in 3 dimensional space, $x-y$ plane is represented as the locational information and height means the locational price. 3 dimensional 39 vectors are projected to the $x-y$ plane, in Fig. 2 we obtain the result of with only locational information. Result shows that there are no changes with only locational consideration. This strict condition does not satisfy the user's request. Hence we will consider the locational price and locational information simultaneously.

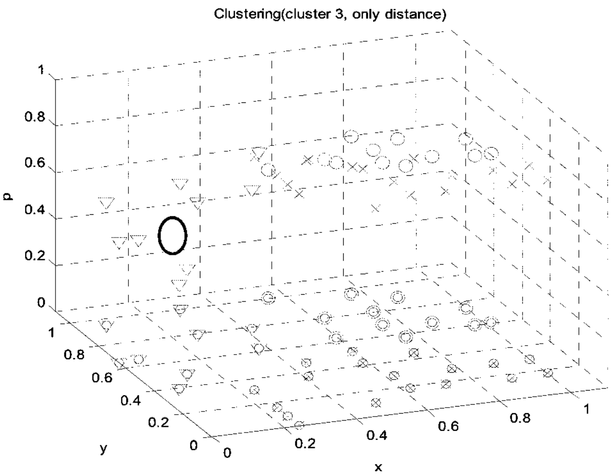


Fig. 2. Clustering(w_1=0, w_2=1)

Next, the weighting values w_1 and w_2 are tuned repetitively. We have obtained 0.27 and 0.73, respectively. At those weighting values any noticeable changes. The result is shown in Fig. 3. In final result we cannot notice any special changes, however there are so many changes near the cluster boundaries at each iterations. As a result, we have to determine weighting values properly for the useful applications. Different cluster number can invoke other grouping status.

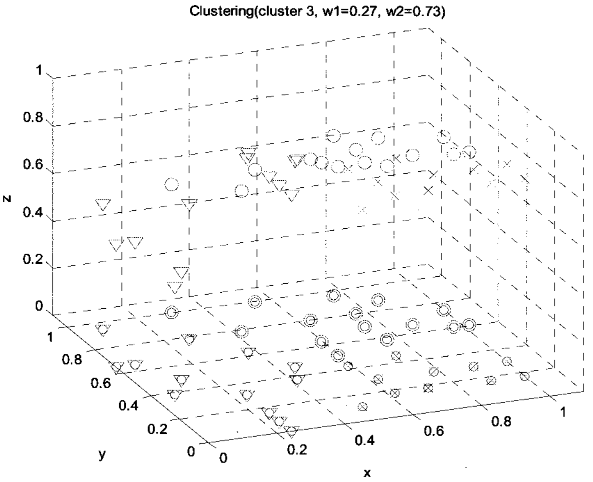


Fig. 3. Clustering (w_1=0.27, w_2=0.73)

4. Conclusions

We have introduced the similarity measure, and constructed the similarity measure using distance measure. For grouping of the large networked system, regional information should be properly considered in the formulation of the similarity measure. In this paper, therefore, we have proposed a modified similarity measure accompanied with regional information and the proposed idea for the modified similarity measure is verified on the networked electricity systems. IEEE reliability test system with 39 nodes are tested as a sample system and from the illustrative example, we can check the coherency between the degree of similarity level and the number of clusters.

References

[1] W. Li and A. Bose, "A coherency based rescheduling method for dynamic security", *IEEE Transactions on Power Systems*, Vol. 13, No. 3, pp. 810-815, 1998.

[2] S.K. Joo, C.C. Liu, L.E. Jones, and J.W. choe, "Coherency and aggregation techniques incorporating rotor and voltage dynamics", *IEEE Transactions on Power Systems,* Vol. 19, No. 2, pp. 1068-1075, 2004.

[3] A.M. Gallai and R.J. Thomas, "Coherency Identification for large electric power systems", *IEEE Transactions on Circuits and Systems,* Vol. CAS-29, No. 11, pp. 777-782, 1982.

[4] F.F. Wu. N. Narasmithamurthi, "Coherency Identification for power system dynamic equivalents", *IEEE Transactions on Circuits and Systems,* Vol. CAS-30, No. 3, pp. 140-147, 1983.

[5] Liu Xuecheng, "Entropy, distance measure and similarity measure of fuzzy sets and their relations", *Fuzzy Sets and Systems*, Vol. 52, pp. 305-318, 1992.

[6] J. L. Fan, W. X. Xie, "Distance measure and induced fuzzy entropy," *Fuzzy Set and Systems,* Vol. 104, pp. 305-314, 1999.

[7] J. L. Fan, Y. L. Ma, and W. X. Xie, "On some properties of distance measures," *Fuzzy Set and Systems,* Vol. 117, pp. 355-361, 2001.

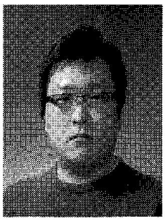
[8] S.H. Lee, S.P. Cheon, and Jinho Kim, "Measure of certainty with fuzzy entropy function", *LNAI,* Vol. 4114, pp. 134-139, 2006.

[9] S.H. Lee, J.M. Kim, and Y.K. Choi, "Similarity measure construction using fuzzy entropy and distance measure", *LNAI,* Vol.4114, 952-958, 2006.

[10] J.C. Bezdek, *Fuzzy Mathematics in Pattern Classification*, Ph.D Thesis, Applied Math. Center, Cornell University, Ithaca, 1973.

[11] J.S.R. Jang, C.T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice Hall, 1997.

[12] The IEEE Reliability Test System-1996, A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee, *IEEE Transactions on Power Systems*, Vol. 14, Issue 3, 1999.

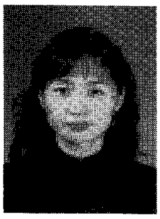


Dong-Hyuk Park
He received the B.S. degrees in School of Mechatronics from Changwon National University, in 2007. Now he is M.S candidate in School of Mechatronics from same University. His interests include optimal control, SMPS.

E-mail : gurehddl@changwon.ac.kr

Soorok Ryu
She received the B.S. and M.S degrees in Mathematics from Kyungbuk National University, in 2003 and 2005, respectively. Now she is Ph.D. candidate in Mathematics from same University. Her interests include numerical PDE, optimal control, fluid mechanics.

E-mail : sryu@knu.ac.kr



Park, Hyun-Jeong

She received the B.S. in Mathematics from Kyunghee University, in 1992, M.S. and Ph.D. degrees in Mathematics Education from Ewha Womans University, in 2001 and 2007, respectively. She served as a Researcher of Mathematics from Mar. 2002 to Dec. 2002 in The Korean Institute of Curriculum and Evaluation. Currently, she is an Instructor of Teaching Psychology of Mathematical Reasoning, Research and guiding curriculum planning in Ewha Womans University and Kyunghee University. Her research interests include fuzzy theory, In-depth examination on thinking processes in students when solving mathematical problems, Evaluation of thinking process using qualitative methods during the process of solving problem or understanding concepts.

E-mail : hyunjp@ewhain.net

Sang-Hyuk Lee

Journal of Fuzzy Logic and Intelligent Systems, Vol. 17, No. 1.

E-mail : leehyuk@changwon.ac.kr