

Statistical Scheme via AIC for Evaluating the Optimal Cut Off Level in Fuzzy Clustering

Shuya Kanagawa⁽¹⁾
dept. Industrial &
Management Systems
Tokyo City University
Tokyo, JAPAN
kanagaw@tcu.ac.jp

Kimiaki Shinkai
dept. Modern Human
Economics
Tokyo Kasei Gakuin
University
Tokyo, JAPAN

Hsunhsun Chung
dept. Mathematics
Waseda University
Tokyo, JAPAN

Kenichi Nagashima
dept. Mathematics
Waseda University
Tokyo, JAPAN

Abstract—In this paper we show a new statistical scheme to find the optimal cut off level in fuzzy clustering which is an improvement of Uesu and Shinkai et. al [4]~[7]. Deterministic algorithms which seek a certain equilibrium cluster level have essential disadvantage in principle. We focus in it and propose a statistical scheme via AIC.

Keywords—component; fuzzy clustering; AIC; cut off level; partition tree

I. INTRODUCTION

When we analyze ambiguous information such as human behavior, mental process, social structure and so on, a fuzzy graph of the information is very useful in these problems (see e.g. Romsburg [3]). Since the fuzzy graph is constructed with clusters of some levels, the clustering level analysis of fuzzy graph is required for the above applications.

Let Fig.3 is a preference relationship between 26 students $\{1,2,\dots,26\}$ of a high school. Fig.1 is a partition tree constructed from Fig.3 with 17 clustering cut off levels $\{R_{0.00}, R_{0.08}, R_{0.15}, R_{0.28}, R_{0.35}, R_{0.38}, R_{0.44}, R_{0.45}, R_{0.48}, R_{0.54}, R_{0.57}, R_{0.65}, R_{0.67}, R_{0.71}, R_{0.79}, R_{0.83}, R_{1.00}\}$ and there are 17 sets of clusters according to these clustering levels each other. For example the level $R_{0.08}$ means that the set of ten points $\{1,2,\dots,26\}$ is divided by 5 clusters $\{15,17\}$, $\{2,18,7,4,13,11,8,20,19,12\}$, $\{14,24,3,22,1,26,16,5,9\}$, $\{21,25\}$, $\{6\}$, $\{10\}$, $\{23\}$. As for the selection of the optimal level in a partition tree in Fig.1 the gradient method in the multivariate analysis has been used very often.

On the other hand, in 1994 Tsuda and Yamashita [5] suggested a new method to estimate the optimal level to find a unique equilibrium point for the cluster number and the cluster size. It means to know a kind of stable classification of clusters for $\{1,2,\dots,26\}$. Their method fits in well with actual examples. Moreover Uesu and Shinkai et. al [4], [6]~[8] proposed a rational method to decide the optimal level of fuzzy clustering with a partition tree. However we have also had two questions about an objective validity of these deterministic algorithms which seek a certain equilibrium cluster level.

The first problem is that the above deterministic algorithms cannot easily make effective use of all the information of the partition tree. Especially, in principle, the gradient method uses only local information of the partition tree. From the view

point we define a configuration of points $\{1,2,\dots,26\}$ on the interval $[0,16]$ measuring the distance between each adjoining two points in the partition tree (see Fig.3). For example, the clustering cut off level $R_{0.83}$ implies that the distance between $\{2\}$ and $\{18\}$ is $0.17=1-0.83$. Furthermore the distance between $\{14\}$ and $\{27\}$ is $0.21=1-0.79$ and so on. Then we can construct a configuration of $\{1,2,\dots,26\}$ on the interval $[0,16]$ in Fig.3. Next we consider a distribution of the configuration. Every clustering cut off level corresponds to a histogram of the configuration, respectively. Therefore we can choose the optimal cut off level by comparing AIC of each histogram.



Figure 1. Partition tree P .

Another problem is “Does the true cluster level always mean an equilibrium or well balanced state?”. From the point of a statistical view, we would reconsider the clustering level analysis in the past and propose a new scheme by applying the AIC method and obtain a statistically equilibrium optimal level of fuzzy clustering. Although we have proposed such statistical scheme in [2], we improve the scheme and show its

⁽¹⁾ The research of the first author was supported, in part, by Grant-in-Aid Scientific Research, No. 20540140, Ministry of Education, Science and Culture, Japan.

advantage to the deterministic algorithms since [2] can not apply to small sample size data.

II. UESU'S METHOD

Next, according to Uesu [7], we introduce Uesu's method to estimate the optimal clustering level for the partition tree P in Fig.1. His method shows that the optimal cut level is **R0.38** from Fig. 2 by the following method.

Definition 1 Crisp Node Fuzzy Graph

A crisp node fuzzy graph G is defined by

$$G = (V, F) : V = \{v_i\}, F = (f_{ij}), 0 \leq f_{ij} \leq 1,$$

where V is the set of the nodes and F is an $n \times n$ matrix whose (i, j) component f_{ij} is a fuzziness of the arc from the node v_i to the node v_j .

Definition 2 Fuzzy Node Fuzzy Graph

A fuzzy node fuzzy graph G is defined by

$$G = (V, F) : V = \{v_i / u_i\}, Y = (y_{ij}), 0 \leq u_i \leq 1, 0 \leq y_{ij} \leq 1,$$

where V is the set of the nodes and the fuzziness u_i is a fuzziness of the node v_i . Y is an $n \times n$ matrix whose (i, j) component y_{ij} is a fuzziness of the arc from the node v_i to the node v_j .

Definition 3 Transformation from Fuzzy Node Fuzzy Graph to Crisp Node Fuzzy Graph

A fuzzy node fuzzy graph $G=(V, Y)$ can be transformed to a crisp node fuzzy graph $G=(V, F)$ by the following method: Let

$$G = (V, F) : V = \{v_i\}, F = (f_{ij}), f_{ij} = T(u_i, y_{ij}),$$

where the fuzziness f_{ij} of the arc from the node v_i to the node v_j could be defined by applying T-norms. A fuzzy graph G is defined by

$$G=(V, Y), V=\{v_i\}, F=(f_{ij}), 0 \leq f_{ij} \leq 1, f_{ii}=1, 1 \leq i \leq n$$

where f_{ij} is a fuzziness of the arc from the node v_i to the node v_j . A symmetric relation matrix S is defined by

$$S=(s_{ij}), \quad \frac{2}{s_{ij}} = \frac{1}{f_{ij}} + \frac{1}{f_{ji}}$$

where $s_{ij}=0$ if $f_{ij} \cdot f_{ji}=0$.

Let $x(c)$ and $y(c)$ be the branching number and maximum cluster size of each R_c , respectively. The more the level R_c increases, the more the branching number $x(c)$ increases. On the other hand, the more the level R_c increases, the more the maximum cluster size $y(c)$ decreases. Since there is an inevitable trade off between the branching number $x(c)$ and the maximum cluster size $y(c)$ in this situation, we would decide the optimal cut level c_0 .

Based on $x(c)$ and $y(c)$, we could define the cluster branch function $p(c)$ and the cluster size function $q(c)$ as follows:

Definition 4 The cluster branch function $p(c)$ and cluster size

function $q(c)$ are defined by

$$p(c) = \frac{x(c)-1}{x(1)-1}, q(c) = \frac{y(c)-1}{y(0)-1},$$

respectively.

Definition 5 Decision of Optimal Cut Level c_0

$$r(c) = p(c) \wedge q(c)$$

and

$$c_0 = \max\{c : r(c) = \max r(x)\}. \quad (1)$$

Then, by using these notions, an optimal cut level c_0 of the partition tree Fig.1 is decided as follows.

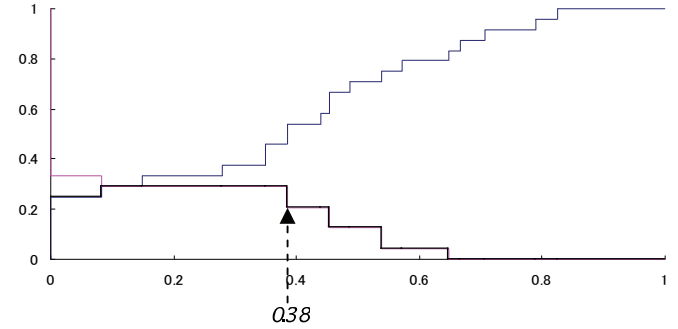


Figure 2. Fuzzy decision $r(c)$.

Applying the fuzzy decision, we obtain that the optimal cut off level is **R0.38** as shown in Fig.2.

III. AIC

We shall define AIC as followings. Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random variables with the probability density function $f(x; \theta^*)$, where $\theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_K^*)$ is an unknown parameter of the distribution. Furthermore, for n random sample points $(x_1, x_2, \dots, x_n)$,

$$\ell(\theta) = \sum_{i=1}^n \log f(x_i; \theta_i)$$

is the log likelihood and $\hat{\theta}_K$ is the maximal likelihood estimator for $\theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_K^*)$ defined by

$$\hat{\theta}_K = \max_{\theta \in \Theta_K} \ell(\theta),$$

where Θ_K is the space of all candidates $\theta = (\theta_1, \theta_2, \dots, \theta_K)$ for the real parameter θ^* . The Kullback-Leibler information number $KLI(\theta^*, \theta)$ for the samples $(x_1, x_2, \dots, x_n)$ is defined by

$$KLI(\theta^*, \theta) = \sum_{i=1}^n f(x_i; \theta^*) \log f(x_i; \theta^*) - \sum_{i=1}^n f(x_i; \theta^*) \log f(x_i; \theta),$$

which metrizes the distance between $f(x; \theta^*)$ and $f(x; \theta_i)$ using entropy. It is easy to see that $\hat{\theta}_K$ minimize the Kullback-Leibler information number $KLI(\theta^*, \theta)$ such that

$$KLI(\theta^*, \hat{\theta}_K) = \min_{\theta \in \Theta_K} KLI(\theta^*, \theta).$$

Now we define AIC (Akaike's information criterion) by

$$AIC(K) = -2\ell(\hat{\theta}_K) + 2K. \quad (2)$$

For comparing several models of $\theta^* = (\theta_1^*, \theta_2^*, \dots, \theta_K^*)$ for known K , we can find the optimal $\theta = (\theta_1, \theta_2, \dots, \theta_K)$ in Θ_K with large probability when it attains the smallest $AIC(K)$ for $(x_1, x_2, \dots, x_n)$. In general, when K is unknown, $-2\ell(\hat{\theta}_K)$ decreases as the number of parameters K increases.

Therefore $AIC(K)$ also gives the optimal K from the definition. By its sensitivity the AIC method is applied to various types of the model selection in many fields.

Let $\mathbf{q} = (q_1, q_2, \dots, q_m)$ be a discrete distribution for integers $\{1, 2, \dots, m\}$. Put $\{x_1, x_2, \dots, x_n\}$ be samples from a proportion of $\{1, 2, \dots, m\}$. Let n_i be a number of samples of each integer $i \in \{1, 2, \dots, m\}$ satisfying $n_1 + n_2 + \dots + n_m = n$. From $K = m - 1$ for the distribution $\mathbf{q}$, AIC of $\mathbf{q}$ is defined by

$$AIC(\mathbf{q}) = -2 \sum_{i=1}^m n_i \log q_i + 2(m-1). \quad (3)$$

IV. STATISTICAL SCHEME VIA AIC

Step 1. Find the configuration of $\{1, 2, \dots, 26\}$ on the interval $[0, 16]$. See Fig.3.

Step 2. Divide the interval $[0, 16]$ into 16 subintervals $[i, i+1]$ for $i=0, 1, \dots, 15$.

Step 3. Make histograms and calculate AIC for each histogram.

Make histogram for $\{1, 2, \dots, 26\}$ according to each fuzzy clustering level. We can find which histogram (cluster level) is the best by comparing AIC's of each histogram. For example, Fig.4 and Fig.5 correspond to clustering levels **R0.35** and **R0.15**, respectively. From (3) we have

$$AIC(\mathbf{R0.45}) = -2\{2\log[3/26] + \log[3/26] + 10\log[1/26] + \log[2/26] + 4\log[7/104]\} + 28 = 132.837$$

$$AIC(\mathbf{R0.44}) = -2\{3\log[2/26] + \log[3/26] + 8\log[1/26] + 4\log[7/104]\} + 26 = 119.426$$

$$AIC(\mathbf{R0.38}) = -2\{2\log[2/26] + \log[3/26] + 2\log[3/52] + 7\log[1/26] + 5\log[9/130]\} + 22 = 120.306$$

$$AIC(\mathbf{R0.35}) = -2\{2\log[2/26] + \log[3/26] + 3\log[5/78] + 2\log[1/26] + 5\log[9/130] + \log[2/26] + 3\log[1/26]\} + 18 = 101.83$$

$$AIC(\mathbf{R0.28}) = -2\{6\log[2/26] + 5\log[1/26] + 5\log[9/130]\} + 16 = 106.063$$

$$AIC(\mathbf{R0.15}) = -2\{2\log[2/26] + 4\log[1/26] + 10\log[9/130]\} + 14 = 92.6404$$

$$AIC(\mathbf{R0.08}) = -2\{2\log[2/26] + 6\log[10/156] + 5\log[9/130] + 3\log[1/26]\} + 12 = 101.479$$

Step 4. Conclusion

Since $AIC(\mathbf{R0.15})$ is smaller than other clustering cut off levels, we can evaluate that **R0.15** is optimal.

V. CONCLUSION

From the above statistical scheme we conclude that the optimal clustering cut off level is **R0.15** which is different from Uesu's method (**R0.38**). Although it is difficult to discuss which level is better, we could evaluate with greater accuracy as we have a lot of criteria for judgment. Furthermore an example in [2] shows that the statistical scheme has an advantage over deterministic methods when **R0.00** is the optimal cut off level. From the point of view we would like to propose the scheme in addition to deterministic methods.

ACKNOWLEDGMENT

The authors kindly thank three anonymous referees for their useful comments. They also would like to express their thanks to Professors H. Yamashita and T. Takizawa of Waseda University for useful discussions.

REFERENCES

- [1] H. Akaike, "Information theory and an extension of the maximum likelihood principle", 2nd Inter. Symp. on Information Theory, Akademiai Kiado, pp. 265-281, 1973.
- [2] S. Kanagawa and K. Shinkai, "Fuzzy Clustering Level Analysis via Statistical Scheme Applying Akaike's Information Criterion", International Journal of Innovative Computing, Information and Control, Vol.4, pp.2523-2531, 2008.
- [3] H. Romesburg, Cluster Analysis for Researchers, Lifetime Learning Publication, 1984.
- [4] K. Shinkai and H. Yamashita, "Level Analysis of Fuzzy Partition Tree and its Application", Int'l Workshop on Soft Computing Applications, 2005.
- [5] E. Tsuda and H. Yamashita, "Sociometry analysis using fuzzy graphs", Journal of Japan SOFT, vol. 6, No.3, 1994 (in Japanese).
- [6] H. Uesu, "Sociometry Analysis Applying Fuzzy Node Fuzzy Graph", Journal of Japan Society for Fuzzy Theory and Systems, Vol.14, No.3, 2002 (in Japanese).
- [7] H. Uesu, "Structure Analysis of Fuzzy Node Fuzzy Graph and its Application to Sociometry Analysis", KES 2008, Part III, LNAI 5179, pp. 84-91, 2008, Springer-Verlag Berlin Heidelberg.
- [8] H. Uesu and K. Shinkai, "Learning Structure Analysis System Applying Fuzzy Theory", Int'l Conference on Advanced Learning Technologies, 2004.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1							6	1			1		10			6			11	1		9		3		12
2													7	2		7						9		5		
3		1						4			10		5			3							3	2		
4						4							3	8		1						4				5
5	5												6					10	1							
6	12	6	4	2				2					6					7								
7		11											5		4					3						
8		6		2			2				1		3			6					9					
9						12							4								3		2			
10	4		3		4	5							4		2		1		1	3						
11		3		3				10					6				4		7			6		1		
12		2		1			3	3			7			6					1			3				
13		2		7			1							6				7				6		1		
14		2			2				1													3		9		3
15																	12									
16	8					4			5					3								4				5
17															12											
18		9																								
19		8							4	3	1	3	1		2		2					3				
20		4					9	9			5								1							
21															12		8		1						4	
22			8											9		5										6
23								2		4	5		3	4			2	3			3					
24			2								1			10									4			3
25															4						8					
26	6		4		6									3								7				

Figure 3. Preference relationship between $\{1,2,\dots,26\}$.

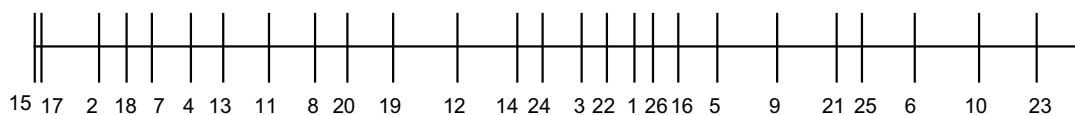


Figure 4. Configuration of $\{1,2,\dots,26\}$.

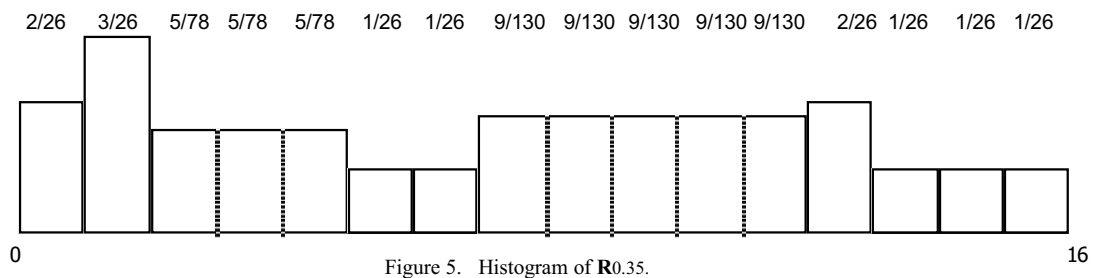


Figure 5. Histogram of $R_{0.35}$.

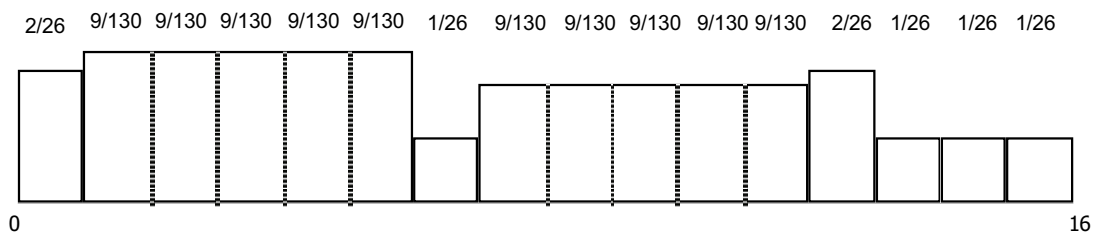


Figure 6. Histogram of $R_{0.15}$.