



On an Multi-directional Searching Algorithm for Two Fuzzy Clustering Methods for Categorical Multivariate Data

Kazune Suzuki^(✉) and Yuchi Kanzawa

Shibaura Institute of Technology, 3-7-5 Toyosu, Koto, Tokyo 135-8548, Japan
ma21070@shibaura-it.ac.jp, kanzawa@sic.shibaura-it.ac.jp

Abstract. Clustering for categorical multivariate data is an important task for summarizing co-occurrence information that consists of mutual affinity among objects and items. This work focus on two fuzzy clustering methods for categorical multivariate data. One of the serious limitations for these methods is the local optimality problem. In this work, an algorithm is proposed to address this issue. The proposed algorithm incorporates multiple token search generated from the eigen decomposition of the Hessian of the objective function. Numerical experiments using an artificial dataset shows that the proposed algorithm is valid.

Keywords: Fuzzy clustering · Local optimality problem · Multiple token search

1 Introduction

The hard c -means (HCM) or k -means clustering algorithm [1] partitions objects into groups. This method is called “hard clustering” because each object belongs to only one cluster, whereas Gaussian mixture models and fuzzy clustering are called “soft clustering” because each object belongs to all or some clusters to varying degrees.

Clustering for categorical multivariate data is a method for summarizing co-occurrence information that consists of mutual affinity among objects and items. A multinomial mixture model (MMM) [2] is a probabilistic model for clustering tasks for categorical multivariate data, where each component distribution is defined by multinomial distribution. Honda et al. [3] proposed the fuzzy clustering for categorical multivariate data induced by MMM (FCCMM). The FCCMM method is a fuzzy counterpart to MMMs, where the degree of fuzziness can be controlled by two fuzzification parameters. Kondo et al. [4] extended FCCMM by introducing q -divergence instead of Kullback-Leibler (KL) divergence in FCCMM. Furthermore, Kondo et al. [4] showed that QFCCMM outperforms FCCMM in terms of clustering accuracy.

One of the most serious limitations for FCCMM and QFCCMM is the local optimality problem. The problem makes the accuracy of their algorithms dependent on its starting points. Thus, obtaining good starting points has been long addressed. One easily idea to avoid locally optimal solutions is running their algorithms multiple times with differently initial setting, and selecting the result where the optimal objective function value is achieved. However, it is unknown how many times should their algorithms run to obtain the globally optimal solution. Arthur and Vassilvitskii [5] proposed k-means++, which is an algorithm for choosing the initial setting for k-means or HCM, This algorithm not only yields considerable improvement in the clustering accuracy of k-means, but also provides a probabilistic upper bound of error. However, this algorithm cannot be applied directly to the other clustering algorithm such as fuzzy clustering algorithms, nor provides any upper bound of error for those than k-means. Ishikawa and Nakano [6] proposed the mes-EM algorithm for the Gaussian mixture models (GMM) incorporating a multiple token search into the EM algorithm for GMM, employing the primitive initial point (PIP) as its initial point, where the search tokens are generated along the directions spanned by the eigen vectors with negative eigen values of the Hessian of the objective function. This idea can be applied to fuzzy clustering algorithms for categorical multivariate data including FCCMM and QFCCMM, which has a potential to solve the local optimality problem of FCCMM and QFCCMM.

In this study, we propose an algorithm to address the local optimality problem of FCCMM and QFCCMM, by modifying the idea of the mes-EM algorithm. The first modification is considering equality-constraints. The idea of the mes-EM algorithm, incorporating a multiple token generated along the directions spanned by the eigen vectors with negative eigen values of the Hessian of the objective function, cannot be valid as it is for FCCMM or QFCCMM. It is because the FCCMM and QFCCMM optimization problems must consider some equality-constraints for variables. If we apply the idea of the mes-EM algorithm directly to FCCMM or QFCCMM, the generating tokens often violate such the constraint. Then, we generate tokens from the intersection of the space spanned by the eigenvectors with negative eigen values of the Hessian of the objective function and the null space of the constraints. The other modification is concerning the length of tokens. Although the generated tokens show the direction to which the objective function improves, we cannot its length at which the objective function improves. If we easily determine the length of tokens, such the tokens may not only make the objective function value worsen but also violate the inequality-constraints. Then, we reduce the length of tokens if it violates the inequality-constraints or it make the objective function value worsen.

The remainder of this paper is organized as follows. Section 2 introduces the notations used and some conventional algorithms. Section 3 describes the proposed algorithm. Section 4 presents the results of numerical experiments conducted to demonstrate the performance of the proposed algorithm. Finally, Sect. 5 concludes the paper.

2 Preliminaries

2.1 Two Fuzzy Clustering Algorithms for Categorical Multivariate Data

Let $X = \{x_k \in \mathbb{R}^M | k \in \{1, \dots, N\}\}$ be a categorical multivariate dataset of M dimensional points. The membership of x_k that belongs to the i -th cluster is denoted by $u_{i,k}$ ($i \in \{1, \dots, C\}, k \in \{1, \dots, N\}$) and the set of $u_{i,k}$ is denoted by u , which obeys the following constraint:

$$\sum_{i=1}^C u_{i,k} = 1, u_{i,k} \in [0, 1] \quad (1)$$

The variable controlling the i -th cluster size is denoted by α_i . The i -th element of vector α is denoted by α_i , and α obeys the following constraint:

$$\sum_{i=1}^C \alpha_i = 1, \alpha_i \in (0, 1) \quad (2)$$

The cluster center set is denoted by $v = \{v_i | v_i \in \mathbb{R}^M, i \in \{1, \dots, C\}\}$. The ℓ -th item typicality for i -th cluster is denoted by $v_{i,\ell}$, and v obeys the following constraint:

$$\sum_{\ell=1}^M v_{i,\ell} = 1, v_{i,\ell} \in [0, 1] \quad (3)$$

The methods FCCMM and QFCCMM are derived by solving the optimization problems,

$$\underset{u, v, \alpha}{\text{minimize}} J_{\text{FCCMM}}(u, v, \alpha), \quad (4)$$

$$\underset{u, v, \alpha}{\text{minimize}} J_{\text{QFCCMM}}(u, v, \alpha), \quad (5)$$

subject to Eqs. (1), (2), and (3), where

$$J_{\text{FCCMM}}(u, v, \alpha) = \sum_{i=1}^C \sum_{k=1}^N u_{i,k} d_{i,k} + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log \left(\frac{u_{i,k}}{\alpha_i} \right), \quad (6)$$

$$J_{\text{QFCCMM}}(u, v, \alpha) = \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\alpha_i)^{1-m} d_{i,k} + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N (u_{i,k})^m (\alpha_i)^{1-m}, \quad (7)$$

$$d_{i,k} = -\frac{1}{t} \sum_{q=1}^Q x_{k,q} \left((v_{i,q})^t - 1 \right), \quad (8)$$

and $m > 1$, $\lambda > 0$ and $t < 1$ are fuzzification parameters. The FCCMM and QFCCMM algorithms are summarized as follows.

Algorithm 1 (FCCMM, QFCCMM).

STEP 1. Set the number of clusters as C . Fix $\lambda > 0$ and $t > 0$ for FCCMM, and $m > 1$, $\lambda > 0$ and $t < 1$ for QFCCMM. Assume initial item typicality as v and initial variable controlling cluster sizes as α .

STEP 2. Update u as

$$u_{i,k} = \frac{\alpha_i \exp(-\lambda d_{i,k})}{\sum_{j=1}^C \alpha_j \exp(-\lambda d_{j,k})} \quad (9)$$

for FCCMM, and

$$u_{i,k} = \frac{\alpha_i (1 - \lambda (1 - m) d_{i,k})^{\frac{1}{1-m}}}{\sum_{j=1}^C \alpha_j (1 - \lambda (1 - m) d_{j,k})^{\frac{1}{1-m}}} \quad (10)$$

for QFCCMM.

STEP 3. Update α as

$$\alpha_i = \frac{\sum_{k=1}^N u_{i,k}}{N} \quad (11)$$

for FCCMM, and

$$\alpha_i = \frac{1}{\sum_{j=1}^C \left(\frac{\sum_{k=1}^N (u_{i,k})^m (1 - \lambda (1 - m) d_{i,k})}{\sum_{k=1}^N (u_{j,k})^m (1 - \lambda (1 - m) d_{j,k})} \right)^{\frac{1}{m}}} \quad (12)$$

for QFCCMM.

STEP 4. Update v as

$$v_{i,\ell} = \frac{\left(\sum_{k=1}^N u_{i,k} x_{k,\ell} \right)^{1/(1-t)}}{\sum_{r=1}^M \left(\sum_{k=1}^N u_{i,k} x_{k,r} \right)^{1/(1-t)}} \quad (13)$$

for FCCMM, and

$$v_{i,\ell} = \frac{\left(\sum_{k=1}^N (u_{i,k})^m x_{k,\ell} \right)^{1/(1-t)}}{\sum_{r=1}^M \left(\sum_{k=1}^N (u_{i,k})^m x_{k,r} \right)^{1/(1-t)}} \quad (14)$$

for QFCCMM.

STEP 5. Check the limiting criterion for (u, v, α) . If the criterion is not satisfied, go to STEP 2.

2.2 Multi-directional in Eigen Space-EM Algorithm for GMM

The mes-EM algorithm was proposed to improve the solution quality of the EM algorithm. The mes-EM algorithm starts from the primitive initial point (PIP), which is the solution for extreme values of inverse temperature in the deterministic annealing [7] context. Let the Hessian of the target function to be minimized have negative eigen values at the PIP. Let $\mathcal{W} = \{w_r, -w_r\}_{r=1}^R$ be the orthonormal set of the corresponding eigen vector. Search tokens are generated along the directions

$$\mathcal{W}' = \left\{ \sum_{r=1}^R (\pm w_r) \right\} = \{(+w_1) + \dots (+w_R), \dots, (-w_1) + \dots + (-w_R)\} \quad (15)$$

in addition to the orthonormal set \mathcal{W} . The mes-EM algorithm is the method of running the EM algorithm $2R + 2^R$ times starting from the same PIP with $\mathcal{W} \cup \mathcal{W}'$ as their search directions, and is described below.

Algorithm 2 (mes-EM).

STEP 1. Calculate all eigen values of the Hessian of the target function at the PIP.

STEP 2. Generate search directions $\mathcal{W} \cup \mathcal{W}'$ by using the negative eigen values.

3 Proposed Methods

In this section, we propose an algorithm to address the local optimality problem of FCCMM and QFCCMM, by modifying the idea of the mes-EM algorithm.

Consider the FCCMM objective function given in Eq. (4) as the function of $s = (v, \alpha) \in \mathbb{R}^{(C+1)M}$, i.e., $J_{\text{FCCMM}}(s) = J_{\text{FCCMM}}(v, \alpha)$, where, u is considered as the function of (v, α) given as Eq. (9). The PIP for the mes-EM algorithm is the solution for extreme values of inverse temperature in the deterministic annealing context, where as the PIP for FCCMM is the solution of their optimization problem with $\lambda \rightarrow 0$, given by $s^{(0)} = (v^{(0)}, \alpha^{(0)})$ where

$$v_{i,\ell}^{(0)} = \frac{\sum_{k=1}^N x_{k,\ell}}{\sum_{r=1}^M \sum_{k=1}^N x_{k,r}}, \quad (16)$$

$$\alpha_i^{(0)} = \frac{1}{C}. \quad (17)$$

The proposed algorithm starts from the PIP.

Let the Hessian of the objective function given by Eq. (4) have negative eigen values at the PIP. Let $\mathcal{W} = \{w_r, -w_r\}_{r=1}^R$ be the orthonormal set of the corresponding eigen vector. In the mes-EM algorithm, multiple tokens are generated using the direction in the space spanned by the corresponding eigen vectors to the negative eigen values of the Hessian of the target function, whereas for FCCMM, the generated token $s^{(0)} + \Delta s$, where Δs is in the space spanned by

\mathcal{W} , is not always valid. It is because we must consider the equality-constraints given by Eqs. (2) and (3) for (v, α) . These constraints are equivalently written as

$$As = \mathbf{1}_{C+1}, \quad (18)$$

$$A = \begin{pmatrix} \mathbf{1}_M^\top & \mathbf{0}_M^\top & \cdots & \mathbf{0}_M^\top & \mathbf{0}_C^\top \\ \mathbf{0}_M^\top & \mathbf{1}_M^\top & \cdots & \mathbf{0}_M^\top & \mathbf{0}_C^\top \\ \mathbf{0}_M^\top & \mathbf{0}_M^\top & \cdots & \mathbf{1}_M^\top & \mathbf{0}_C^\top \\ \mathbf{0}_M^\top & \mathbf{0}_M^\top & \cdots & \mathbf{0}_M^\top & \mathbf{1}_C^\top \end{pmatrix}, \quad (19)$$

where $\mathbf{1}_{C+1}$, $\mathbf{1}_M$, and $\mathbf{1}_C$ are the vector whose all the elements are ones with the dimension of $C+1$, M , and C , respectively, and $\mathbf{0}_M$, and $\mathbf{0}_C$ are the vector whose all the elements are zeros with the dimension of M and C , respectively. If we have $A\Delta s \neq 0$, then the generated token $s + \Delta s$ violates the equality-constraint as

$$A(s + \Delta s) = As + A\Delta s \neq \mathbf{1}_{C+1}. \quad (20)$$

Then, we generate tokens $s^{(0)} + \Delta s$ where Δs is in the intersection of $\text{span}(\mathcal{W})$ and the null space of A , i.e., $\text{null}(A)$. Such the intersection can be obtained as the righthand singular vectors of AW where $W = (w_1, \dots, w_R)$.

Although Δs show the direction to which the objective function improves with keeping the equality-constraints given by Eq. (18), or equivalently Eqs. (2) and (3), we cannot know its length at which the objective function value improves. If we easily determine the length of Δs , such the tokens $s^{(0)} + \Delta s$ may not only make the objective function value $J_{\text{FCCMM}}(s^{(0)} + \Delta s)$ worsen but also violate the inequality-constraints $v_{i,\ell} \in [0, 1]$ and $\alpha_i \in (0, 1)$. Then, we reduce the length of tokens if it violates the inequality-constraints or it make the objective function value worsen.

The above discussion is not only for FCCMM but also for QFCCMM, and is summarized into the following algorithm:

Algorithm 3.

- STEP 1. Let \mathcal{S} , \mathcal{S}^* , and $\Delta\mathcal{S}$ be empty sets, add $s^{(0)}$ given by Eqs. (16) and (17) to \mathcal{S} .
- STEP 2. If \mathcal{S} is empty, output the element of \mathcal{S}^* such that its objective function value is the minimum, and terminate this algorithm. Otherwise, pop s from \mathcal{S} , and run Algorithm 1 using the initial setting s , resulting into \hat{s} .
- STEP 3. Calculate all eigen pairs of $\nabla^2 J_{\text{FCCMM}}(\hat{s})$ for FCCMM or $\nabla^2 J_{\text{QFCCMM}}(\hat{s})$ for QFCCMM. If all the eigen values are positive, \hat{s} is a locally or globally optimal solution. Then, add \hat{s} to \mathcal{S}^* , and return to STEP 2. If all the eigen values are negative, \hat{s} is not a locally or globally optimal solution. Then, ignore \hat{s} , and return to STEP 2. If at least one eigen value is negative, \hat{s} is a saddle point. Let the corresponding eigen vectors be $\mathcal{W} = \{w_r \in \mathbb{R}^{(C+1)M}\}_{r=1}^R$.

STEP 4. Obtain the orthonormal basis vectors

$$\check{\mathcal{W}} = \{\check{w}_r, -\check{w}_r\}_{r=1}^{\check{R}} \tag{21}$$

of $\text{span}(\mathcal{W}) \cap \text{null}(A)$ and their combinations

$$\check{\mathcal{W}}' = \left\{ \sum_{r=1}^{\check{R}} (\pm \check{w}_r) \right\} = \{(+\check{w}_1) + \dots + (+\check{w}_{\check{R}}), \dots, (-\check{w}_1) + \dots + (-\check{w}_{\check{R}})\}. \tag{22}$$

Add all the elements of $\check{\mathcal{W}} \cup \check{\mathcal{W}}'$ to $\Delta\mathcal{S}$.

STEP 5. If $\Delta\mathcal{S}$ is empty, return to STEP 2. Otherwise, pop Δs from $\Delta\mathcal{S}$.

STEP 6. Find $0 < \beta \leq 1$ such as $0 < \hat{s} + \beta\Delta s < 1$ and $J_{\text{FCCMM}}(\hat{s} + \beta\Delta s) < J_{\text{FCCMM}}(\hat{s})$ for FCCMM or $J_{\text{QFCCMM}}(\hat{s} + \beta\Delta s) < J_{\text{QFCCMM}}(\hat{s})$ for QFC-CMM, add $\hat{s} + \beta\Delta s$ to \mathcal{S} , and return to STEP 5. If there does not exist such the value β , ignore Δs and return to STEP 5.

4 Numerical Experiments

This section provides numerical experiments to illustrate Algorithm 3 based on one artificial dataset as shown in Fig. 1. with four clusters ($C = 4$) in the two dimensional unit-simplex. First cluster is composed of 100 objects generated from multinomial distribution with $v_1 = (0.1, 0.1, 0.8)$. Second cluster is composed of 200 objects generated from multinomial distribution with $v_2 = (0.8, 0.1, 0.1)$. Third cluster is composed of 400 objects generated from multinomial distribution with $v_3 = (0.1, 0.8, 0.1)$. Fourth cluster is composed of 400 objects generated from multinomial distribution with $v_4 = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

The fuzzification parameter λ and t for FCCMM was set as $\lambda \in \{10, 40\}$ and $t = 0.5$. The fuzzification parameter m, λ and t for QFCCMM was set as $m = \{1.0001, 1.2\}, \lambda = 40$ and $t = 0.5$.

For FCCMM with $(\lambda, t) = (10, 0.5)$, after the only output of Algorithm 1 from the PIP was judged as a saddle point at STEP 3. of Algorithm 3, 76 tokens were generated through STEP 4. and STEP 6. of Algorithm 3, all the outputs of Algorithm 1 from these tokens were judged as locally or globally optimal solutions at STEP 3. of Algorithm 3, and Algorithm 3 terminated. Among 76 locally or globally optimal solutions,

10 points are strictly local optimum with $\text{ARI} = 0.82$, and 66 points achieve the minimum objective function value with $\text{ARI} = 1.0$. This result is summarized in Table 1 along with the other cases. These results show that the proposed algorithm produce the globally optimal solution through multiple tokens generated from the PIP. However, many generated tokens are the same convergence point. For example, in the case with FCCMM with $(\lambda, t) = (10, 0.5)$, among 66 solutions with minimal objective function value, here exists only 1 distinct one which means that the algorithm has redundancy. More efficient generating tokens is a future work.

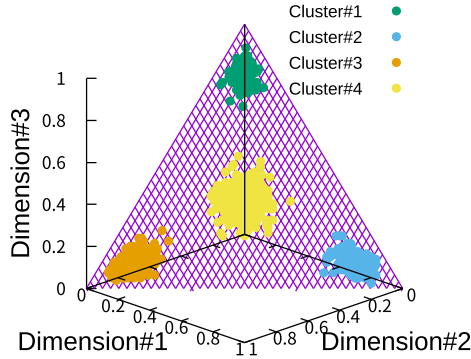


Fig. 1. Artificial dataset

Table 1. Number of saddle points, tokens, and locally/globally optimal solutions along with actual number obtained from Algorithm 3.

Method	Fuzzification parameter			Number of saddle points	Number of tokens	Number of strictly local optimum	Number of solutions with minimal objective function value (Actual number)	ARI of the solution minimal objective function value
	m	λ	t					
EFCCMM		10	0.5	1	76	10	66(1)	1.0
		30	0.5	1	76	5	71(1)	1.0
		40	0.5	1	76	6	70(1)	1.0
QFCCMM	1.2	40	0.5	1	76	10	66(1)	1.0
	1.2	30	0.5	1	76	20	56(1)	1.0
	1.0001	40	0.5	1	76	6	70(1)	1.0
	1.0001	30	0.5	1	76	8	68(1)	1.0

5 Conclusion

In this work, we proposed an algorithm to address the local optimality problem of FCCMM and QFCCMM. Numerical experiments using an artificial dataset shows that the proposed algorithm is valid, though it has a redundancy.

In the future, through improving the proposed algorithm efficiently, the proposed algorithms will be applied to a large number of real datasets. Furthermore, the technique generating multiple tokens will be applied to clustering algorithms for the other types of data, such as spherical data, e.g., in [8].

References

1. MacQueen, J.B.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297 (1967)
2. Rigouste, L., Cappé, O., Yvon, F.: Inference and evaluation of the multinomial mixture model for text clustering. *Inf. Process. Manag.* **43**(5), 1260–1280 (2007)

3. Honda, K., Oshio, S., Notsu, A.: Fuzzy co-clustering induced by multinomial mixture models. *JACIII* **19**(6), 717–726 (2015)
4. Kondo, T., Kanzawa, Y.: Fuzzy clustering methods for categorical multivariate data based on q -divergence. *JACIII* **22**(4), 524–536 (2018)
5. Arthur, D., Vassilvitskii, S.: k-means++: the advantages of careful seeding. In: Proceedings of the 8th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035 (2007)
6. Ishikawa, Y., Nakano, R.: Landscape of a likelihood surface for a gaussian mixture and its use for the EM algorithm. In: Proceedings of the IJCNN2006, pp. 2413–2419 (2006)
7. Ueda, N., Nakano, R.: Deterministic annealing EM algorithm. *Neural Netw.* **11**(2), 271–282 (1998)
8. Higashi, M., Kondo, T., Kanzawa, Y.: Fuzzy clustering method for spherical data based on q -divergence. *JACIII* **23**(3), 561–570 (2019)