

On Possibilistic Clustering Methods Based on Shannon/Tsallis-Entropy for Spherical Data and Categorical Multivariate Data

Yuchi Kanzawa^(✉)

Shibaura Institute of Technology, Koto, Tokyo 135-8548, Japan
kanzawa@sic.shibaura-it.ac.jp

Abstract. In this paper, four possibilistic clustering methods are proposed. First, we propose two possibilistic clustering methods for spherical data — one based on Shannon entropy, and the other on Tsallis entropy. These methods are derived by subtracting the cosine correlation between an object and a cluster center from 1, to obtain the object-cluster dissimilarity. These methods are derived from the proposed spherical data methods by considering analogies between the spherical and categorical multivariate fuzzy clustering methods, in which the fuzzy methods' object-cluster similarity calculation is modified to accommodate the proposed possibilistic methods. The validity of the proposed methods is verified through numerical examples.

Keywords: Possibilistic clustering · Spherical data · Categorical multivariate data

1 Introduction

Fuzzy c -means (FCM), proposed by Bezdek [1], is the most popular algorithm for performing fuzzy clustering on linear data. FCM is fuzzified through its membership in the hard c -means (HCM) objective function [2]. Other HCM fuzzification methods include entropy-regularized FCM (eFCM) [3] and Tsallis entropy-based FCM (tFCM) [4].

FCM and its variants are useful clustering methods; however, their memberships do not always correspond well to the degree of belonging of the data. To address this weakness of FCM, Krishnapuram and Keller [5] proposed a possibilistic c -means (PCM) algorithm that uses a possibilistic membership function. Krishnapuram and Keller [6], and Ménard et al. [4] proposed other possibilistic clustering techniques that employ Shannon entropy and Tsallis entropy, respectively. In this study, these two methods are respectively referred to as entropy-regularized PCM (ePCM) and Tsallis-entropy-regularized PCM (tPCM).

All the aforementioned clustering methods are designed for linear data. In other application domains, linear data clustering methods may yield poor results. For example, information retrieval applications show that cosine similarity is a more accurate similarity measure for clustering text documents than Euclidean

distortion of dissimilarity [7]. Such domains require spherical data use, and only consider the directions of the unit vectors. In particular, spherical K-means [8] and its fuzzified variants [9–13] are designed to process spherical data. However, a possibilistic approach for clustering spherical data has not been proposed in the literatures; this was a motivation for this work. The spherical clustering methods that correspond to eFCM and tFCM are denoted as eFCS and tFCS in this paper.

In recent studies [13–17], various fuzzy clustering methods have been proposed for categorical multivariate data (FCCM). In these methods, a categorical multivariate dataset is provided in the form of a cross-classification table, contingency table, or co-occurrence matrix. Because the optimization problems [13, 15] are similar to spherical clustering, these FCCM methods can be extended into possibilistic clustering, which was another motivation for this work. The method described in [15] is referred to as entropy-regularized FCCM (eFCCM), and the method described in [13] is referred to as Tsallis entropy-regularized FCCM (tFCCM), in order to distinguish these methods in this paper.

In this study, four possibilistic clustering methods are proposed — two for spherical data and two for categorical multivariate data. First, we propose the possibilistic clustering methods for spherical data: entropy-regularized possibilistic clustering for spherical data (ePCS) and Tsallis entropy-regularized possibilistic clustering for spherical data (tPCS). These methods are derived by subtracting the cosine correlation between an object and a cluster center from 1, to obtain the object-cluster dissimilarity; this value is used in place of the squared Euclidean distance between an object and an cluster center, which is commonly used in conventional linear data methods. Second, we propose two possibilistic clustering methods for categorical multivariate data: entropy-regularized possibilistic clustering for categorical multivariate data (ePCCM) and Tsallis entropy-regularized possibilistic clustering for categorical multivariate data (tPCCM). These methods are derived from the proposed spherical data methods (ePCS and tPCS) by considering analogies between the fuzzy methods for spherical data and categorical multivariate data; here, the object-cluster similarity calculation in the fuzzy methods is modified to accommodate the proposed possibilistic methods. The validity of the proposed methods is verified through numerical examples.

The rest of this paper is organized as follows. In Sect. 2, the notation and the conventional methods are introduced. Section 3 presents the proposed methods, and Sect. 4 provides some numerical examples. Section 5 contains our concluding remarks.

2 Preliminaries

2.1 Notation, Fuzzyc-Means, and Its Variants

Let $X = \{x_k \in \mathbb{R}^p \mid k \in \{1, \dots, N\}\}$ be a dataset of p -dimensional points, referred to as linear data. 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 is also known as the partition matrix. The cluster center set is denoted by $v = \{v_i \mid v_i \in \mathbb{R}^p, i \in \{1, \dots, C\}\}$. The squared Euclidean distance between the k -th datum and the i -th cluster center is denoted by $d_{i,k} = \|x_k - v_i\|_2^2$.

One approach for membership fuzzification is to regularize the objective function of HCM by introducing a regularization term with a positive parameter λ into the objective function. This approach was successfully implemented by Miyamoto and Mukaidono [3]. Using the entropy term, the entropy-regularized FCM (eFCM) is defined as

$$\underset{u,v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} d_{i,k} + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log(u_{i,k}) \quad (1)$$

$$\text{subject to } \sum_{i=1}^C u_{i,k} = 1. \quad (2)$$

Ménard adopted Tsallis entropy [20] instead of Shannon entropy to perform fuzzy clustering, and proposed tFCM [4] defined as

$$\underset{u,v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k}^m d_{i,k} + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N [u_{i,k}^m - u_{i,k}] \quad (3)$$

subject to Eq. (2).

2.2 Possibilistic Clustering

To improve the fidelity of fuzzy clustering, Krishnapuram and Keller [5] relaxed the constrained condition Eq. (2), which yielded a possibilistic membership function. The memberships for a certain cluster and its cluster center are released from constraint Eq. (2), and are obtained independent of these for other clusters. Hereafter, we only consider cases in which $C = 1$, where the number 1 signifies that only one cluster is searched for at a time. With this setting, cluster fusion [18] is useful, given that many cluster centers become nearer to each other as iteration proceeds; the distance between two clusters frequently approaches zero. The cluster fusion is described in the following algorithm:

Algorithm 1

1. Select a subset of objects as initial cluster centers. It is possible to select all objects: $C = N$; $v_i = x_i$ ($i \in \{1, \dots, C\}$).
2. Perform possibilistic clustering, and obtain C cluster centers.
3. Merge cluster centers that have negligible distances between them.

Krishnapuram and Keller [6], and Ménard [4] proposed possibilistic clustering methods using entropy, defined as

$$\underset{u,v}{\text{minimize}} \sum_{k=1}^N u_{1,k} d_{1,k} + \lambda^{-1} \sum_{k=1}^N u_{1,k} \log(u_{1,k}) - \lambda^{-1} \sum_{k=1}^N u_{1,k}, \quad (4)$$

$$\underset{u,v}{\text{minimize}} \sum_{k=1}^N u_{1,k}^m d_{1,k} + \frac{\lambda^{-1}}{m-1} \sum_{k=1}^N (u_{1,k}^m - u_{1,k}) - \lambda^{-1} \sum_{k=1}^N u_{1,k}. \quad (5)$$

These two methods are referred to as ePCM and tPCM because the usual (Shannon) entropy and Tsallis entropy are employed in these methods, respectively. The optimal solutions for the membership and cluster center are described as

$$u_{1,k} = \exp(-\lambda d_{1,k}), \quad (6)$$

$$v_1 = \left(\sum_{k=1}^N u_{1,k} x_k \right) / \left(\sum_{k=1}^N u_{1,k} \right) \quad (7)$$

for ePCM, and

$$u_{1,k} = (1 - \lambda(1 - m) d_{1,k})^{\frac{1}{1-m}}, \quad (8)$$

$$v_1 = \left(\sum_{k=1}^N u_{1,k}^m x_k \right) / \left(\sum_{k=1}^N u_{1,k}^m \right) \quad (9)$$

for tPCM. These equations are alternatively iterated during the second step in Algorithm 1. Ménard denoted the third term in Eqs. (4) and (5) as a possibilistic constraint, and showed that these two methods were derived by adding this constraint to the eFCM and tFCM objective functions in Eqs. (1) and (3).

2.3 Fuzzy Clustering for Spherical Data

If objects are on the unit hypersphere, $1 - x_k^\top v_i$ can be used as the dissimilarity between an object x_k and a cluster center v_i . Such objects are referred to as spherical data. Two methods that correspond to Eqs. (1) and (3) are obtained for the following optimization problems:

$$\underset{u,v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} (1 - x_k^\top v_i) + \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log(u_{i,k}), \quad (10)$$

$$\underset{u,v}{\text{minimize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k}^m (1 - x_k^\top v_i) + \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N [u_{i,k}^m - u_{i,k}], \quad (11)$$

respectively, subject to Eq. (2) and

$$\|v_i\|_2 = 1, \quad (12)$$

referred to as eFCS [9] and tFCS [13], respectively. It is shown in [19] that eFCS optimization problem in Eq. (10) can be equivalently described as the following maximizing problem

$$\underset{u,v}{\text{maximize}} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} x_k^\top v_i - \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log(u_{i,k}). \quad (13)$$

However, to the best of our knowledge, a possibilistic approach to spherical clustering has not yet been investigated.

2.4 Fuzzy Clustering for Categorical Multivariate Data

Assume that for datasets $X = \{x_k \mid k \in \{1, \dots, N\}\}$ and $Y = \{y_\ell \mid \ell \in \{1, \dots, M\}\}$, the co-occurrence information between x_k and y_ℓ , $R_{k,\ell}$ is given. R is the matrix whose (k, ℓ) -th element is $R_{k,\ell}$. We refer to X and Y as the row and column datasets, respectively, because the k -th row of R represents the similarities between x_k and y_ℓ , and the ℓ -th column of R represents the similarities between y_ℓ and x_k . The membership of datum x_k belonging to the i -th cluster is denoted by $u_{i,k}$. The (i, k) -th element of matrix u is denoted by $u_{i,k}$, and u satisfies the constraint in Eq. (2). The membership of datum y_ℓ belonging to the i -th cluster is denoted by $w_{i,\ell}$. The (i, ℓ) -th element of matrix w is denoted by $w_{i,\ell}$, and w satisfies the constraint

$$\sum_{\ell=1}^M w_{i,\ell} = 1. \quad (14)$$

The eFCCM [15] is obtained by solving the following optimization problem:

$$\underset{u,w}{\text{maximize}} \sum_{i=1}^C \sum_{k=1}^N \sum_{\ell=1}^M u_{i,k} \log(w_{i,\ell}) R_{k,\ell} - \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log(u_{i,k}) \quad (15)$$

subject to Eqs. (2) and (14), where $\lambda > 0$ is a fuzzification parameter. The tFCCM [13] are obtained by solving the following optimization problem

$$\underset{u,w}{\text{maximize}} \sum_{i=1}^C \sum_{k=1}^N \sum_{\ell=1}^M u_{i,k}^m \log(w_{i,\ell}) R_{k,\ell} - \frac{\lambda^{-1}}{m-1} \sum_{i=1}^C \sum_{k=1}^N [u_{i,k}^m - u_{i,k}] \quad (16)$$

subject to Eqs. (2) and (14), where $\lambda > 0$ and $m > 1$ are fuzzification parameters. Because the optimization problem described in Eq. (15) is similar to Eqs. (1) and (10), and because the optimization problem described in Eq. (16) is similar to Eqs. (3) and (11), it is possible to generalize FCCM in the same manner in which eFCM was modified into ePCM. This fact motivated this work.

3 Proposed Method

3.1 Modifying ePCM and tPCM

In this subsection, we modify ePCM and tPCM as a preparatory procedure to derive the proposed methods.

ePCM and tPCM objective functions are slightly generalized from Eqs. (4) and (5) to

$$\underset{u,v}{\text{minimize}} \sum_{k=1}^N u_{1,k} d_{1,k} + \lambda^{-1} \sum_{k=1}^N u_{1,k} \log(u_{1,k}) - \alpha \sum_{k=1}^N u_{1,k}, \quad (17)$$

$$\underset{u,v}{\text{minimize}} \sum_{k=1}^N u_{1,k}^m d_{1,k} + \frac{\lambda^{-1}}{m-1} \sum_{k=1}^N (u_{1,k}^m - u_{1,k}) - \alpha \sum_{k=1}^N u_{1,k}, \quad (18)$$

where the factor of the last term in the original problems in Eqs. (4) and (5), λ^{-1} , is replaced by another parameter $\alpha \in (-\infty, +\infty)$ for modified ePCM and $\alpha \in (-1/(\lambda(m-1)), +\infty)$ for modified tPCM. These optimal solutions for membership are described as

$$u_{1,k} = \beta \exp(-\lambda d_{1,k}), \quad (19)$$

for modified ePCM, where $\beta = \exp(\lambda\alpha - 1) \in (0, +\infty)$, and

$$u_{1,k} = \beta(1 - \lambda(1-m)d_{1,k})^{1/(1-m)}, \quad (20)$$

for modified tPCM, where $\beta = (1 + \alpha\lambda(m-1))/m)^{1/(m-1)} \in (0, +\infty)$, and the optimal solutions of cluster center are the same as the original forms in Eqs. (7) and (9). We can observe that $\alpha = \lambda^{-1}$ recovers the original problems. We note that the membership value is 1 at $d_{1,k} = 0$ for the original membership form; this is not the case in Eqs. (19) and (20), except for the case in which $\alpha = \lambda^{-1}$. However, this does not imply that the modified ePCM and the modified tPCM do not contain defects. First, in the possibilistic theory, the maximal membership value does not need to be 1. Second, such a modification does not affect cluster center updating, as explained in the following procedure. Denote the membership and the cluster center in the modified ePCM as \tilde{u} and \tilde{v} to distinguish them from those in the original ePCM. Then, we have

$$\tilde{v}_1 = \frac{\sum_{k=1}^N \tilde{u}_{1,k} x_k}{\sum_{k=1}^N \tilde{u}_{1,k}} = \frac{\sum_{k=1}^N \beta u_{1,k} x_k}{\sum_{k=1}^N \beta u_{1,k}} = \frac{\beta \sum_{k=1}^N u_{1,k} x_k}{\beta \sum_{k=1}^N u_{1,k}} = v_1, \quad (21)$$

which means that such a modification does not affect the updating of the cluster centers, and simply changes the scale of membership. The case of tPCM also leads to the same result. Hereafter, the modified versions of ePCM and tPCM are used to derive the proposed methods.

3.2 Possibilistic Clustering for Spherical Data

In this subsection, we propose two possibilistic clustering methods for spherical data, ePCS and tPCS.

ePCS is obtained by solving the optimization problem

$$\underset{u,v}{\text{minimize}} \sum_{k=1}^N u_{1,k} (1 - x_k^\top v_1) + \lambda^{-1} \sum_{k=1}^N u_{1,k} \log(u_{1,k}) - \alpha \sum_{k=1}^N u_{1,k}, \quad (22)$$

subject to Eq. (12). This optimization problem is derived by subtracting the cosine correlation between an object and a cluster center from 1 ($1 - x_k^\top v_1$) to obtain the object-cluster dissimilarity, instead of using the squared Euclidean distance between an object and a cluster center ($\|x_k - v_1\|_2^2$) applied in ePCM, which was described in Eq. (17). The optimal solutions for the membership and cluster center are described as

$$u_{1,k} = \beta \exp(\lambda x_k^\top v_1), \quad (23)$$

$$v_1 = \left(\sum_{k=1}^N u_{1,k} x_k \right) / \left(\left\| \sum_{k=1}^N u_{1,k} x_k \right\|_2 \right), \quad (24)$$

where $\beta = \exp(-\lambda - 1 + \lambda\alpha)$. ePCS is also derived from eFCS by subtracting the possibilistic constraint term $\alpha \sum_{k=1}^N u_{1,k}$ from the eFCS objective function described in Eq. (10), omitting the probabilistic constraint in Eq. (2), and considering the spherical constraint in Eq. (12). The ePCS membership in Eq. (23) is described for arbitrary object x as $u_1(x) = \beta \exp(\lambda x^\top v_1)$; this is the unnormalized von Mises-Fisher distribution. This membership function for a one-dimensional sphere is depicted in Fig. 1 for several parameter values of λ , where β is set such that $\max_x u_1(x) = 1$. The ePCS optimization problem is described as the following maximizing problem:

$$\begin{aligned} \text{Eq. (22)} &\Leftrightarrow \underset{u,v}{\text{maximize}} \sum_{k=1}^N u_{1,k} x_k^\top v_1 - \lambda^{-1} \sum_{k=1}^N u_{1,k} \log(u_{1,k}) + (\alpha - 1) \sum_{k=1}^N u_{1,k} \\ &\Leftrightarrow \underset{u,v}{\text{maximize}} \sum_{k=1}^N u_{1,k} x_k^\top v_1 - \lambda^{-1} \sum_{k=1}^N u_{1,k} \log(u_{1,k}) + \alpha' \sum_{k=1}^N u_{1,k}, \end{aligned} \quad (25)$$

where $\alpha' = \alpha - 1$. This optimization problem is also obtained from the maximizing problem of eFCS in Eq. (13), by adding the possibilistic constraint term $\alpha' \sum_{k=1}^N u_{1,k}$ from the ePCS objective function described in Eq. (13), while omitting the probabilistic constraint in Eq. (2) and considering the spherical constraint in Eq. (12). This maximizing problem is used to derive ePCCM in the next subsection.

The tPCS method is obtained by solving the optimization problem

$$\underset{u,v}{\text{minimize}} \sum_{k=1}^N u_{1,k}^m (1 - x_k^\top v_1) + \frac{\lambda^{-1}}{m-1} \sum_{k=1}^N (u_{1,k}^m - u_{1,k}) - \alpha \sum_{k=1}^N u_{1,k}, \quad (26)$$

subject to Eq. (12). This optimization problem is derived by subtracting the cosine correlation between an object and a cluster center from 1 ($1 - x_k^\top v_1$) to obtain the object-cluster dissimilarity. This value replaces the squared Euclidean distance between an object and an cluster center ($\|x_k - v_1\|_2^2$), which is used in the tPCM method described in Eq. (5). The optimal solutions for the membership and cluster center are described as

$$u_{1,k} = \beta (1 - \lambda(1 - m)(1 - x_k^\top v_1))^{\frac{1}{1-m}}, \quad (27)$$

$$v_1 = \left(\sum_{k=1}^N u_{1,k}^m x_k \right) / \left(\left\| \sum_{k=1}^N u_{1,k}^m x_k \right\|_2 \right), \quad (28)$$

where $\beta = ((1 - \alpha\lambda(1 - m))/m)^{1/(m-1)}$. tPCS is also derived from tFCS by subtracting the possibilistic constraint term $\alpha \sum_{k=1}^N u_{1,k}$ from tFCS objective function described in Eq. (10), omitting the probabilistic constraint in Eq. (2), and considering the spherical constraint in Eq. (12). The membership is rewritten using $\lambda' = \lambda/(1 - \lambda(1 - m))$ and $\beta' = \beta(\lambda/\lambda')^{1/(1-m)}$ as

$$u_{1,k} = \beta' (1 + \lambda' (1 - m) x_k^\top v_1)^{1/(1-m)}. \quad (29)$$

This membership in Eq. (29) is described for arbitrary object x as $u_1(x) = \beta' (1 + \lambda' (1 - m) x^\top v_1)^{1/(1-m)}$; this is a deformation of the unnormalized von Mises-Fisher distribution when $\lambda = 1$, i.e., $u_1(x)$ recovers a von Mises-Fisher distribution with $m \rightarrow 1$, which is similar to the method used by the Tsallis distribution [20] to recover a Gaussian distribution. This membership function for a one-dimensional sphere is depicted in Figs. 2 and 3 for several parameter values of (λ, m) , where β is set such that $\max_x u_1(x) = 1$. The tPCS optimization problem is described as the following maximizing problem:

$$\begin{aligned} \text{Eq. (26)} &\Leftrightarrow \underset{u,v}{\text{maximize}} \sum_{k=1}^N u_{1,k}^m x_k^\top v_1 - \frac{1 + \lambda(m-1)}{\lambda(m-1)} \sum_{k=1}^N [u_{1,k}^m - u_{1,k}] \\ &\quad + (\alpha - 1) \sum_{k=1}^N u_{1,k} \\ &\Leftrightarrow \underset{u,v}{\text{maximize}} \sum_{k=1}^N u_{1,k}^m x_k^\top v_1 - \frac{\lambda'^{-1}}{m-1} \sum_{k=1}^N [u_{1,k}^m - u_{1,k}] + \alpha' \sum_{k=1}^N u_{1,k}, \end{aligned} \quad (30)$$

where $\alpha' = \alpha - 1$ and $\lambda' = \lambda/(1 - \lambda(1 - m))$. This maximizing problem is used to derive tPCCM in the next subsection.

3.3 Possibilistic Clustering for Categorical Multivariate Data

In this subsection, we propose two possibilistic clustering methods for categorical multivariate data, ePCCM and tPCCM.

First, we reconfigure the objective function of eFCCM described in Eq. (15) as

$$\sum_{i=1}^C \sum_{k=1}^N \sum_{\ell=1}^M u_{i,k} (\log(w_{i,\ell} R_{k,\ell}) - \log(\Gamma(R_{k,\ell} + 1))) - \lambda^{-1} \sum_{i=1}^C \sum_{k=1}^N u_{i,k} \log(u_{i,k}), \quad (31)$$

by adding the term

$$- \sum_{i=1}^C \sum_{k=1}^N \sum_{\ell=1}^M u_{i,k} \log(\Gamma(R_{k,\ell} + 1)) \quad (32)$$

to the original objective function. This term originates from the third term of the following lower bound, for the log-likelihood of a multinomial mixture model

$$\begin{aligned}
 & - \sum_{k=1}^N \sum_{i=1}^C u_{i,k} \log(u_{i,k}) \sum_{k=1}^N \sum_{i=1}^C u_{i,k} + \log(\Gamma(\sum_{\ell=1}^M R_{k,\ell} + 1)) \\
 & - \sum_{k=1}^N \sum_{i=1}^C u_{i,k} \sum_{\ell=1}^M \log(\Gamma(R_{k,\ell} + 1)) + \sum_{k=1}^N \sum_{i=1}^C u_{i,k} \sum_{\ell=1}^M \log(w_{i,\ell}) R_{k,\ell} \quad (33)
 \end{aligned}$$

$$\leq \sum_{k=1}^N \sum_{i=1}^C u_{i,k} \log\left(\frac{1}{u_{i,k}} \frac{\Gamma(\sum_{\ell=1}^M R_{k,\ell} + 1)}{\prod_{\ell=1}^M \Gamma(R_{k,\ell} + 1)} \prod_{\ell=1}^M w_{i,\ell}^{R_{k,\ell}}\right) \quad (34)$$

$$= \sum_{k=1}^N \log\left(\sum_{i=1}^C \frac{\Gamma(\sum_{\ell=1}^M R_{k,\ell} + 1)}{\prod_{\ell=1}^M \Gamma(R_{k,\ell} + 1)} \prod_{\ell=1}^M w_{i,\ell}^{R_{k,\ell}}\right). \quad (35)$$

The added term described in Eq. (32) does not affect the optimal solution of eFCCM because of the constraint described in Eq. (2), whereas it plays a role in constituting the membership function in a possibilistic manner; this is discussed later.

Next, similar to the manner in which ePCS is derived from eFCS, the ePCCM optimization problem is constructed from eFCCM. The objective function of the eFCS maximizing problem described in Eq. (13) is quite similar to that of eFCCM in Eq. (31) if $s_{i,k} = x_k^\top v_i$ in eFCS (Eq. (13)) and $s_{i,k} = \sum_{\ell=1}^M \log(w_{i,\ell}) R_{k,\ell} - \log(\Gamma(R_{k,\ell} + 1))$ in eFCCM (Eq. (31)). Based on this information, an ePCCM optimization problem is proposed as

$$\begin{aligned}
 & \underset{u,w}{\text{maximize}} \sum_{k=1}^N \sum_{\ell=1}^M u_{1,k} (\log(w_{1,\ell}) R_{k,\ell} - \log(\Gamma(R_{k,\ell} + 1))) - \lambda^{-1} \sum_{k=1}^N u_{1,k} \log(u_{1,k}) \\
 & + \alpha \sum_{k=1}^N u_{1,k} \quad (36)
 \end{aligned}$$

subject to the constraint in Eq. (14), which is obtained from the eFCCM objective function in Eq. (31) by setting $C = 1$, omitting the constraint in Eq. (2), and adding the possibilistic constraint $\alpha \sum_{k=1}^N u_{1,k}$ to the eFCCM objective function. By solving this optimization problem, we obtain the optimal solutions for memberships (u, w) as

$$u_{1,k} = \beta \exp\left(\lambda \sum_{\ell=1}^M \log(w_{1,\ell}) R_{k,\ell} - \log(\Gamma(R_{k,\ell} + 1))\right), \quad (37)$$

$$w_{1,\ell} = \left(\sum_{k=1}^N u_{1,k} R_{k,\ell}\right) / \left(\sum_{r=1}^M \sum_{k=1}^N u_{1,k} R_{k,r}\right), \quad (38)$$

where $\beta = \exp(-\lambda - 1 + \alpha\lambda)$. The ePCCM membership in Eq. (37) is described for arbitrary object $R = (R_1, \dots, R_M)$ as $u_1(R) = \beta \exp(\lambda \sum_{\ell=1}^M \log(w_{1,\ell}) - \log(\Gamma(R_\ell + 1)))$ is the unnormalized multinomial distribution when $\lambda = 1$. This membership function for $M = 2$ and $w_1 = (0.2, 0.8)$ is depicted in Fig. 4 for

several parameter values of λ , where β is set such that $\max_R u_1(R) = 1$. Here, we can observe the purpose of adding the term in Eq. (32). If this term is omitted, such a membership function is described as

$$u_1(R) = \beta \exp\left(\lambda \sum_{\ell=1}^M \log(w_{1,\ell}) R_\ell\right), \quad (39)$$

and is depicted in Fig. 5 for $\lambda = 1$ where β is set such that $\max_R u_1(R) = 1$. From this figure, we can observe that such membership functions cannot capture the mode of densities; when the mode with $w_1 < 0.5$ is at the minimal value of R_1 , $R_1 = 0$, the mode with $w_1 > 0.5$ is at the maximal value of R_1 , $R_1 = 20$, and the mode with $w_1 = 0.5$ is disappears. On the other hand, by adding the term in Eq. (32), we can observe in Fig. 4 that the membership functions can capture the mode of densities.

The tPCCM optimization problem is obtained from ePCCM in a similar manner to how tPCS is derived from ePCS, i.e., by replacing $u_{i,k}$ in the first term of Eq. (36) and Shannon entropy in the second term of Eq. (36) by $u_{i,k}^m$ and Tsallis entropy, respectively, as

$$\begin{aligned} & \underset{u,w}{\text{maximize}} \sum_{k=1}^N \sum_{\ell=1}^M u_{1,k}^m (\log(w_{1,\ell}) R_{k,\ell} - \log(\Gamma(R_{k,\ell} + 1))) - \frac{\lambda^{-1}}{m-1} \sum_{k=1}^N [u_{1,k}^m - u_{1,k}] \\ & + \alpha \sum_{k=1}^N u_{1,k} \end{aligned} \quad (40)$$

subject to the constraint in Eq. (14). By solving this optimization problem, we obtain the optimal solutions of memberships as

$$u_{1,k} = \beta \left(1 + (1-m)\lambda \sum_{\ell=1}^M \log(w_{1,\ell}) R_{k,\ell} - \log(\Gamma(R_{k,\ell} + 1))\right)^{1/(1-m)}, \quad (41)$$

$$w_{1,\ell} = \left(\sum_{k=1}^N u_{1,k}^m R_{k,\ell}\right) / \left(\sum_{r=1}^M \sum_{k=1}^N u_{1,k}^m R_{k,r}\right), \quad (42)$$

where $\beta = ((1 - \alpha\lambda(1 - m))/m)^{1/(m-1)}$. This membership in Eq. (41) is derived for arbitrary object $R = (R_1, \dots, R_M)$, as $u_k = \beta \left(1 + (1-m)\lambda \sum_{\ell=1}^M \log(w_{1,\ell}) R_\ell - \log(\Gamma(R_\ell + 1))\right)^{1/(1-m)}$, which is a deformation of the unnormalized multinomial distribution when $\lambda = 1$, i.e., $u_1(x)$ recovers multinomial distribution with $m \rightarrow 1$ by setting an adequate normalization factor β . This membership function for $M = 2$ and $w_1 = (0.2, 0.8)$ is depicted in Figs. 6 and 7 for several parameter values of (λ, m) where β is set such that $\max_R u_1(R) = 1$.

4 Numerical Example

This section provides numerical examples based on artificial and actual datasets. The first example illustrates the performance of ePCS and tPCS using a dataset

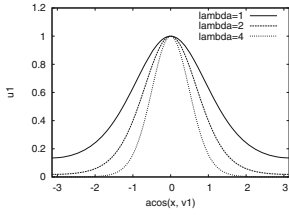


Fig. 1. ePCS membership functions

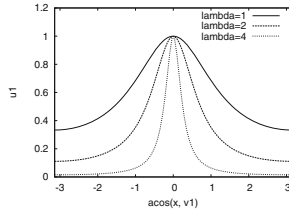


Fig. 2. tPCS membership functions with $m = 2$

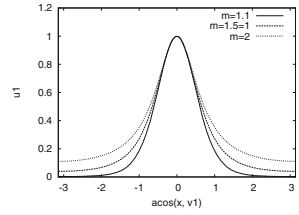


Fig. 3. tPCS membership functions with $\lambda' = 1$

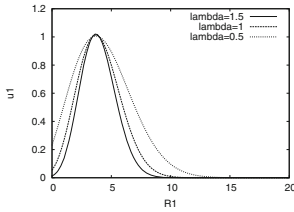


Fig. 4. ePCCM membership functions

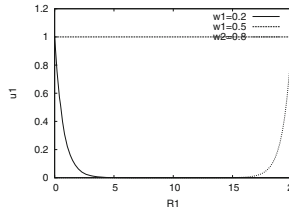


Fig. 5. Incomplete ePCCM membership functions

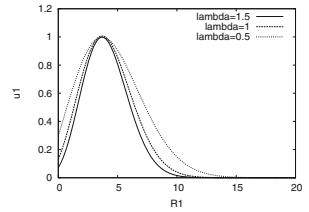


Fig. 6. tPCCM membership functions with $m = 2$

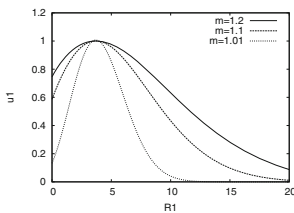


Fig. 7. tPCCM membership functions with $\lambda = 1$

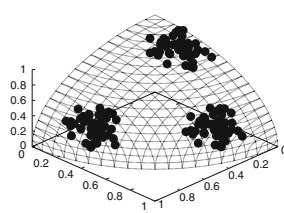


Fig. 8. Artificial Dataset #1

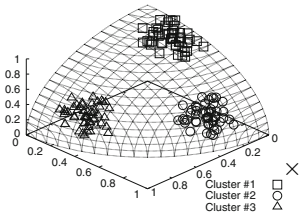


Fig. 9. Result for Artificial Dataset #1

-0.8pc

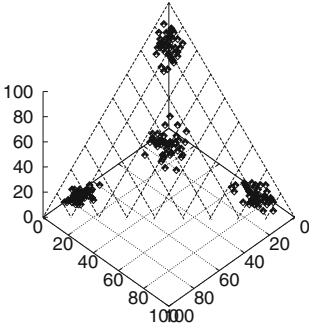


Fig. 10. Artificial dataset #2

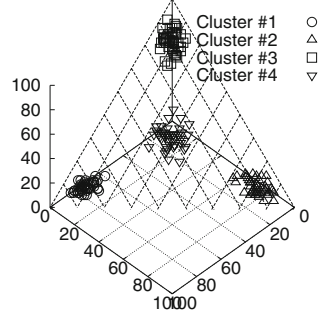


Fig. 11. Results for Artificial dataset #2 obtained with ePCCM and tPCCM

containing three clusters, each of which contains 50 points in the first quadrant of the unit sphere (Fig. 8). Using the parameter settings $\lambda = 1.0$ for ePCS and $(\lambda, m) = (1.0, 1.5)$ for tPCS, both methods partitioned this dataset adequately, as shown in Fig. 9, where squares, circles, and triangles indicate the maximal memberships generated by both algorithms during the test.

The second example illustrates the performance of ePCCM and tPCCM using an artificial dataset containing four clusters, all of which contain 50 points obtained from a random sampling of multinomial distributions with parameters $(0.8, 0.1, 0.1)$, $(0.1, 0.8, 0.1)$, $(0.1, 0.1, 0.8)$, and $(1/3, 1/3, 1/3)$ (Fig. 10). With the parameter settings $\lambda = 1.0$ for ePCCM and $(\lambda, m) = (1.0, 1.5)$ for tPCCM, both methods partitioned this dataset adequately, as shown in Fig. 11. The maximal membership of the data is depicted by squares, circles, triangles, and reverse triangles.

5 Conclusions

In this study, four possibilistic clustering methods were proposed. First, we proposed two possibilistic clustering methods for spherical data — one based on Shannon entropy, and one based on Tsallis entropy. It was shown that the membership functions recovered the unnormalized von Mises-Fisher distribution and its deformation. Second, we proposed two possibilistic clustering methods for categorical multivariate data. It was shown that these membership functions recovered the unnormalized multinomial distribution and its deformation. The validity of the proposed methods was confirmed through numerical examples.

In future work, we will (1) apply the proposed methods to larger and more complex datasets, (2) investigate how fuzzification parameters affect clustering accuracy and propose a method to automatically set the best parameter values, (3) apply the fuzzified method used in [16], (4) compare the proposed methods

with other clustering methods, (5) apply the sequential cluster extraction [24], which is another algorithm for possibilistic clustering, and (6) develop a possibilistic clustering approach for other data types.

References

1. Bezdek, J.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
2. MacQueen, J.B.: Some methods of classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and probability, pp. 281–297 (1967)
3. Miyamoto, S., Mukaidono, M.: Fuzzy c -means as a regularization and maximum entropy approach. In: Proceedings of 7th International Fuzzy Systems Association World Congress (IFSA 1997), vol. 2, pp. 86–92 (1997)
4. Ménard, M., Courboulay, V., Dardignac, P.: Possibilistic and probabilistic fuzzy clustering: unification within the framework of the non-extensive thermostatistics. Pattern Recogn. **36**, 1325–1342 (2003)
5. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. IEEE Trans. Fuzzy Syst. **1**, 98–110 (1993)
6. Krishnapuram, R., Keller, J.M.: The possibilistic c -means algorithm: insights and recommendations. IEEE Trans. Fuzzy Syst. **4**, 393–396 (1996)
7. Strehl, A., Ghosh, J., Mooney, R.: Impact of similarity measures on web-page clustering. In: Proceedings of AAAI 2000, pp. 58–64 (2000)
8. Dhillon, I.S., Modha, D.S.: Concept decompositions for large sparse text data using clustering. Mach. Learn. **42**, 143–175 (2001)
9. Miyamoto, S., Mizutani, K.: Fuzzy multiset model and methods of nonlinear document clustering for information retrieval. In: Torra, V., Narukawa, Y. (eds.) MDAI 2004. LNCS (LNAI), vol. 3131, pp. 273–283. Springer, Heidelberg (2004)
10. Mizutani, K., Inokuchi, R., Miyamoto, S.: Algorithms of nonlinear document clustering based on fuzzy set model. Int. J. Intell. Syst. **23**(2), 176–198 (2008)
11. Kanzawa, Y.: Maximizing model of bezdek-like spherical fuzzy c -means clustering. In: Proceedings of the FUZZ-IEEE 2014, pp. 2482–2488 (2014)
12. Kanzawa, Y.: On kernelization for a maximizing model of bezdek-like spherical fuzzy c -means clustering. In: Torra, V., Narukawa, Y., Endo, Y. (eds.) MDAI 2014. LNCS, vol. 8825, pp. 108–121. Springer, Heidelberg (2014)
13. Kanzawa, Y.: Fuzzy clustering based on α -divergence for spherical data and for categorical multivariate data. In: Proceedings of the FUZZ-IEEE 2015, (to appear)
14. Oh, C., Honda, K., Ichihashi, H.: Fuzzy clustering for categorical multivariate data. In: Proceedings of the IFSA World Congress and 20th NAFIPS International Conference, pp. 2154–2159 (2001)
15. Honda, K., Oshio, S., Notsu, A.: FCM-type fuzzy co-clustering by K-L information regularization. In: Proceedings of 2014 IEEE International Conference on Fuzzy Systems, pp. 2505–2510 (2014)
16. Honda, K., Oshio, S., Notsu, A.: Item membership fuzzification in fuzzy co-clustering based on multinomial mixture concept. In: Proceedings of 2014 IEEE International Conference on Granular Computing, pp. 94–99 (2014)
17. Kummamuru, K., Dhawaie, A., Krishnapuram, R.: Fuzzy co-clustering of document and keywords. In: IEEE Proceedings of FUZZ 2003, vol. 2, pp. 772–777 (2003)

18. Miyamoto, S., Inokuchi, R., Kuroda, Y.: Possibilistic and fuzzy c -means clustering with weighted objects. In: Proceedings of 2006 IEEE International Conference on Fuzzy Systems, pp. 869–874 (2006)
19. Miyamoto, S., Ichihashi, H., Honda, K.: Algorithms for Fuzzy Clustering. Springer, Heidelberg (2008)
20. Tsallis, C.: Possible generalization of boltzmann-gibbs statistics. *J. Stat. Phys.* **52**, 479–487 (1988)
21. Buchta, C., Kober, M., Feinerer, I., Hornik, K.: Spherical k -means clustering. *J. Stat. Softw.* **50**(10), 1–22 (2012)
22. Barber, M.J.: Modularity and community detection in bipartite networks. *Phys. Rev. E* **76**, 066102 (2007)
23. Davis, A., Gardner, B.B., Gardner, M.R.: Deep South. University of Chicago Press, Chicago (1941)
24. Kanzawa, Y.: Sequential cluster extraction using power-regularized possibilistic c -means. *JACIII* **19**(1), 67–73 (2015)