Credibilistic Clustering: The Model and Algorithms

Jian Zhou∗§, Qina Wang∗, C.-C. Hung†‡¶ and Xiajie Yi∗

∗School of Management, Shanghai University, Shanghai 200444, China
†Anyang Normal University, Anyang 455000, China
‡School of Computing and Software Engineering, Southern Polytechnic State University, Marietta, GA 30060, USA
¶zhoujian@shu.edu.cn
§chung@spsu.edu

Received 26 December 2014
Revised 16 February 2015

Fuzzy clustering is a widely used approach for data classification by using the fuzzy set theory. The probability measure and the possibility measure are two popular measures which have been used in the fuzzy c-means algorithm (FCM) and the possibilistic clustering algorithms (PCAs), respectively. However, the numerical experiments revealed that FCM and its derivatives lack the intuitive concept of degree of belongingness, and PCAs suffer from the “coincident problem” and cannot provide very stable results for some data sets. In this study, we propose a new clustering algorithm, called the credibilistic clustering algorithm (CCA), based on the credibility measure. The credibility measure provides some unique properties which can solve the “coincident problem” and noise issue compared with the probability measure and possibility measure. Based on some randomly generated data sets, experimental results compared with FCM and PCA show that CCA can deal with the “coincident problem” well with good clustering results, and it is more robust to noise than PCA.

Keywords: Fuzzy clustering; credibilistic clustering; clustering algorithm; alternating cluster estimation; credibility measure.

1. Introduction

Cluster analysis is the process of partitioning a data set into subsets so that items within a valid cluster are more similar to each other than those belonging to a different cluster. There are many applications for clustering methods, such as image segmentation, pattern recognition, and information retrieval. The classification of clustering models can be roughly assigned into two groups: crisp clustering and fuzzy clustering. In crisp clustering (also referred to as hard clustering), each data element belongs to exactly one cluster. In fuzzy clustering (also referred to as soft clustering), data elements can belong to more than one cluster, and associated with each element is a set of memberships. Fuzzy clustering is the most popular
unsupervised method for data grouping by using the fuzzy set theory, which has advantages over traditional clustering in many applications.\textsuperscript{5,10,12,14,17,18,22}

In the field of fuzzy clustering, the fuzzy $c$-means algorithm (FCM), proposed by Bezdek,\textsuperscript{1} is the most well-known and used method, which is based on a fuzzy objective function with probabilistic membership weights. Since the memberships of FCM do not always correspond to the intuitive concept of degrees of belongingness or compatibility due to its probabilistic constraints, it may be inaccurate in a noisy environment.\textsuperscript{2} In order to produce memberships with a good explanation of degree of belongingness for the data, Krishnapuram and Keller\textsuperscript{2,3} developed the possibilistic clustering algorithms (PCAs). They stated that PCAs are more robust to outliers than FCM. Since the performance of PCAs proposed in Refs. 2 and 3 heavily depends on the selection of parameters, a new PCA was suggested by Yang and Wu,\textsuperscript{19} whose performance can be easily controlled.

Although PCAs have been extensively used in practice, they suffer from a serious problem generally called the coincident problem. This occurs because PCAs may return clusters which are coincident. Hence, a data set which consists of a number of clusters, $c$, may be partitioned into less than $c$ clusters by PCAs with a certain probability. The details about how the coincident problem happens have been discussed in Ref. 15. To avoid this problem and improve the existing PCAs, we develop a new clustering method in this paper by means of the credibility measure introduced by Liu and Liu.\textsuperscript{9} In the proposed clustering model, the objective function is the compactness index of the data sets with credibilistic membership weights, and the constraints on credibilities are deduced from the mathematical properties of the credibility measure. Furthermore, in order to solve this model with good clustering results, a credibilistic clustering algorithm based on the alternating cluster estimation method presented by Runkler and Bezdek\textsuperscript{13} is suggested, where the updating equations for membership and prototype are determined by the users, which helps improving the efficiency and practicality of the algorithm.

The rest of this paper is organized as follows. In Sec. 2, some related fundamental knowledge is reviewed involving FCM, the possibilistic clustering algorithms, the approach of alternating cluster estimation, and the credibility measure. Section 3 introduces the new approach of credibilistic clustering by deriving the corresponding clustering optimization model. After that, a credibilistic clustering algorithm (CCA) is developed by integrating alternating cluster estimation with the credibility theory in Sec. 4. Finally, some numerical examples are given to illustrate the efficiency of the proposed algorithm in Sec. 5.

2. Preliminaries

As the essential prerequisite, in this section, the FCM algorithm and the possibilistic clustering algorithms are briefly reviewed. Additionally, the basic idea of alternating cluster estimation is introduced. The definition as well as the sufficient and necessary conditions of the credibility measure is also presented in this section, which will be used in the following sections.
2.1. The FCM algorithm

In fuzzy clustering, FCM developed by Bezdek\(^1\) is one of the most popular algorithms and has been applied successfully in many areas. Given a data set
\[ X = \{x_1, x_2, \ldots, x_n\} \]
in a \(p\)-dimensional Euclidean space, an ordinary Euclidean norm \(\| \cdot \|\) on \(\mathbb{R}^p\), an integer \(c(2 \leq c < n)\) representing the specified number of clusters, and a parameter of fuzzifier \(m\) with \(m > 1\), the approach of FCM is to find the optimal membership matrix \(\mu\) and the cluster center matrix \(A\) which minimize the objective function

\[
J_{\text{FCM}}(\mu, A) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m \|x_j - a_i\|^2 \quad (1)
\]

subject to the following constraints

\[
\begin{align*}
0 \leq \mu_{ij} &\leq 1 \quad \text{for all } i, j, \\
\sum_{i=1}^{c} \mu_{ij} & = 1 \quad \text{for all } j.
\end{align*} \quad (2)
\]

Here, \(A = (a_1, a_2, \ldots, a_c)\) represents the cluster center matrix, where \(a_i\) (vector) is the \(i\)th cluster center, \(\mu\) represents the membership matrix with

\[
\mu = \begin{pmatrix}
\mu_{11} & \cdots & \mu_{1n} \\
\vdots & \ddots & \vdots \\
\mu_{c1} & \cdots & \mu_{cn}
\end{pmatrix}, \quad (3)
\]

and the variable \(\mu_{ij}\) means the degree of compatibility or membership of the feature point \(x_j\) belonging to the \(i\)th cluster \(S_i\). The weighting exponent \(m\) has a great influence on the performance of fuzzy clustering. For convenience, we denote by \(d_{ij}\) the distance between the \(j\)th feature point \(x_j\) and the \(i\)th cluster center \(a_i\), i.e.,

\[
d_{ij} = \|x_j - a_i\| \quad \text{for all } i, j. \quad (4)
\]

Since there is no closed form solution in the clustering method, many fuzzy and non-fuzzy clustering algorithms use the alternating optimization approach trying to achieve the optimal solution. Alternating optimization (AO) is an iterative procedure for minimizing the objective function defined for the clustering. It minimizes jointly over all variables by alternating restricted minimizations over the individual subsets of variables.\(^13\) AO has been studied and used in a wide variety of areas. Under reasonable assumptions, the general AO approach is shown to be locally, \(q\)-linearly convergent, and to also exhibit a type of global convergence.\(^13\) With the AO methodology for FCM and PCAs, the equations to update \(\mu\) and \(A\) are the necessary conditions for local extremes of their respective objective functions.

As to FCM, the necessary conditions for a minimizer \((\mu, A)\) of \(J_{\text{FCM}}\) subject to the constraints in (2) are the following evaluation equations for memberships

\[
\mu_{ij} = \left( \sum_{k=1}^{c} \frac{d_{ij}^{2/(m-1)}}{d_{ik}^{2/(m-1)}} \right)^{-1} \quad \text{for all } i, j, \quad (5)
\]
and the following updating equations for cluster centers
\[
a_i = \frac{\sum_{j=1}^{n}(\mu_{ij})^m x_j}{\sum_{j=1}^{n}(\mu_{ij})^m} \text{ for all } i.
\] (6)

The procedure of FCM can be described as follows.

**Algorithm 1: The FCM algorithm**

**Step 1.** Initialize \(a_i^{(0)} \in \mathbb{R}^p\) for all \(i\), and set a small number \(\epsilon > 0\) and an iteration counter \(t = 0\).

**Step 2.** Compute \(\mu_{ij}^{(t+1)}\) using the evaluation equations in (5) for all \(i, j\).

**Step 3.** Compute \(a_i^{(t+1)}\) using the updating equations in (6) for all \(i\).

**Step 4.** Increment \(t\) until \(\max_i ||a_i^{(t+1)} - a_i^{(t)}|| < \epsilon\) or a predetermined iteration number achieves.

### 2.2. Possibilistic clustering algorithms

Possibilistic clustering was initiated by Krishnapuram and Keller\(^2,3\) and improved by Yang and Wu.\(^{19}\) The PCA proposed by Yang and Wu\(^{19}\) (PCA06 for short in our paper) has very good and stable clustering performance in some situations, which is also an iterative algorithm like FCM based upon the idea of alternating optimization. After random initializations of the cluster centers, the membership matrix \(\mu\) and the matrix of cluster centers \(A\) are iteratively updated until the convergence of cluster centers, where the updating equations for both \(\mu\) and \(A\) are obtained from the necessary conditions for a minimizer of the objective function

\[
J_{PCA06}(\mu, A) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m ||x_j - a_i||^2
\]
\[
+ \frac{\beta}{m \sqrt{c} d_{ij}} \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m \ln(\mu_{ij})^m - (\mu_{ij})^m
\] (7)

subject to the constraints
\[
0 \leq \mu_{ij} \leq 1 \quad \text{for all } i, j.
\] (8)

Thus, in the PCA06 algorithm, the updating equations for memberships are

\[
\mu_{ij} = \exp \left\{ -\frac{m \sqrt{c} d_{ij}}{\beta} \right\} \quad \text{for all } i, j,
\] (9)

where the parameter \(\beta\) is defined by

\[
\beta = \frac{\sum_{j=1}^{n} ||x_j - \bar{x}||^2}{n} \text{ with } \bar{x} = \frac{\sum_{j=1}^{n} x_j}{n},
\] (10)

and the updating equations in (6) for cluster centers in FCM are also used in PCA06.

In summary, the procedure of the PCA06 algorithm is described as follows.
Algorithm 2: The PCA06 algorithm

Step 1. Initialize \( a_i^{(0)} \in \mathbb{R}^p \) for all \( i \), and set a small number \( \epsilon > 0 \) and an iteration counter \( t = 0 \).

Step 2. Compute \( \mu_{ij}^{(t+1)} \) using the evaluation equations in (9) for all \( i, j \).

Step 3. Compute \( a_i^{(t+1)} \) using the updating equations in (6) for all \( i \).

Step 4. Increment \( t \) until \( \max_i ||a_i^{(t+1)} - a_i^{(t)}|| < \epsilon \) or a predetermined iteration number achieves.

2.3. Alternating cluster estimation

As mentioned above, for AO solutions of FCM and PCAs, the equations to update both \( \mu \) and \( A \) are the necessary conditions for local extremes of their respective objective functions. Moreover, Runkler and Bezdek \(^{13} \) proposed the alternating cluster estimation (ACE) method as architecture of the AO algorithms. The ACE approach does not adopt the objective function model. It allows the users to select membership functions and cluster centers directly through the alternating iteration architecture. Below is an example of using the ACE method.

A general approach of alternating cluster estimation based on possibilistic membership functions has been developed in Ref. 23, which leads to a family of iterative clustering algorithms, called the generalized possibilistic clustering algorithms, in which the memberships are calculated by

\[
\mu_{ij} = f_i(d_{ij}) \quad \text{for all } i, j \tag{11}
\]

with the membership function \( f_i \) satisfying

\[
\begin{align*}
& f_i \text{ is monotone decreasing on } [0, +\infty), \\
& f_i(0) = 1, \\
& f_i(+\infty) = 0, \tag{12}
\end{align*}
\]

for each \( i \), and the cluster centers are also updated by (6). The procedure of the generalized possibilistic clustering algorithms can be described as follows.

Algorithm 3: The generalized possibilistic clustering algorithms (GP-CAs)

Step 1. Initialize \( a_i^{(0)} \in \mathbb{R}^p \) for all \( i \), and set a small number \( \epsilon > 0 \) and an iteration counter \( t = 0 \).

Step 2. Compute \( \mu_{ij}^{(t+1)} \) using the evaluation equations in (11) for all \( i, j \), where the function \( f_i \) is predetermined by the users for each \( i \).

Step 3. Compute \( a_i^{(t+1)} \) using the updating equations in (6) for all \( i \).

Step 4. Increment \( t \) until \( \max_i ||a_i^{(t+1)} - a_i^{(t)}|| < \epsilon \) or a predetermined iteration number achieves.
It should be noted that the membership functions $f_i$ in Step 2 should be pre-determined by the users according to the real applications, and various functions satisfying the constraints in (12) may produce clustering algorithms with different performances. As further investigation on the generalized possibilistic clustering algorithms, a general framework for constructing membership functions in this model was introduced by combining the clustering performance with the fuzzy set theory in Ref. 16.

### 2.4. Credibility measure

Fuzzy set theory has been well developed and applied in a wide variety of real problems, which derives the possibility theory and the credibility theory on the basis of the possibility measure and the credibility measure, respectively. The possibility measure was proposed by Zadeh\(^\text{20}\) and the credibility measure was developed by Liu and Liu\(^\text{9}\) based upon the possibility measure. The following definitions and theorem give a background on the concepts of the possibility measure and the credibility measure.

**Definition 1.**\(^\text{11}\) Let $\Theta$ be a nonempty set representing the sample space, and $\mathcal{P}(\Theta)$ the power set of $\Theta$. A set function $\text{Pos}: \Theta \rightarrow [0, 1]$ is called a possibility measure if it satisfies the following axioms:

- **Axiom 1.** $\text{Pos}(\Theta) = 1$;
- **Axiom 2.** $\text{Pos}(\emptyset) = 0$;
- **Axiom 3.** $\text{Pos}(\bigcup_i A_i) = \sup_i \text{Pos}(A_i)$ for any collection $\{A_i\}$ in $\mathcal{P}(\Theta)$.

The triplet $(\Theta, \mathcal{P}(\Theta), \text{Pos})$ is called a possibility space. Moreover, in order to define the product possibility measure, Liu\(^\text{6}\) presented the fourth axiom as follows.

**Axiom 4.** Let $\Theta_i$ be nonempty sets on which $\text{Pos}_i$ satisfies the first three axioms for $i = 1, 2, \ldots, n$, respectively, and $\Theta = \Theta_1 \times \Theta_2 \times \cdots \times \Theta_n$. Then for each $A \in \mathcal{P}(\Theta)$,

$$\text{Pos}(A) = \sup_{(\theta_1, \theta_2, \ldots, \theta_n) \in A} \text{Pos}_1(\theta_1) \wedge \text{Pos}_2(\theta_2) \wedge \cdots \wedge \text{Pos}_n(\theta_n).$$

In that case we write $\text{Pos} = \text{Pos}_1 \wedge \text{Pos}_2 \wedge \cdots \wedge \text{Pos}_n$. It should be noted that the product possibility measure $\text{Pos}$ also satisfies the first three axioms.

**Definition 2.**\(^\text{21}\) Let $(\Theta, \mathcal{P}(\Theta), \text{Pos})$ be a possibility space, and $A$ a set in $\mathcal{P}(\Theta)$. Then the necessity measure of $A$ is defined by

$$\text{Nec}(A) = 1 - \text{Pos}(A^c).$$

**Definition 3.**\(^\text{9}\) Let $(\Theta, \mathcal{P}(\Theta), \text{Pos})$ be a possibility space, and $A$ a set in $\mathcal{P}(\Theta)$. Then the credibility measure of $A$ is defined by

$$\text{Cr}(A) = \frac{1}{2} (\text{Pos}(A) + \text{Nec}(A)).$$
Theorem 1. Let $\Theta$ be a nonempty set, $\mathcal{P}(\Theta)$ the power set of $\Theta$, and $\text{Cr}$ a set function defined on $\mathcal{P}(\Theta)$. Then $\text{Cr}$ is a credibility measure if and only if the following four conditions are satisfied:

(i) $\text{Cr}(\Theta) = 1$;
(ii) $\text{Cr}$ is increasing, i.e., $\text{Cr}\{A\} < \text{Cr}\{B\}$ whenever $A \subset B$;
(iii) $\text{Cr}$ is self-dual, i.e., $\text{Cr}\{A\} + \text{Cr}\{A^c\} = 1$ for any $A \in \mathcal{P}(\Theta)$;
(iv) $\text{Cr}\{\bigcup_i A_i\} \wedge 0.5 = \sup_i \text{Cr}\{A_i\}$ for any $\{A_i\}$ with $\text{Cr}\{A_i\} \leq 0.5$.

Note that the self-duality of $\text{Cr}$ implies that the fuzzy event must hold if its credibility is 1, and fail if its credibility is 0. On the contrary, a fuzzy event may fail even though its possibility achieves 1, and hold even though its necessity is 0.

Traditionally, the possibility measure is regarded as the parallel concept of the probability measure. However, it is, in fact, the credibility measure that plays the role of the probability measure within the framework of fuzzy set theory. Motivated by this fact, Liu provided an axiomatic approach to describing fuzziness, called the credibility theory, which develops as a branch of mathematics that studies the behavior of fuzzy events. So far, the credibility theory has been applied to a wide range of areas such as engineering design, portfolio selection, and production planning. For detailed expositions, the interested reader may consult Liu. In the following section, we will introduce the credibility theory into the field of cluster analysis to obtain a new clustering method.

3. Credibilistic Clustering Method

Possibilistic clustering is essentially a clustering method using the possibility measure in the clustering process. Considering that the possibility measure has no self-duality property, in this section, we substitute the possibility measure for the credibility measure in the clustering optimization model, which results in a new clustering method called credibilistic clustering.

3.1. Problem description and notations

Fuzzy clustering is an approach using the fuzzy set theory as a tool for data grouping. Given a data set $X$ with $n$ feature points in a $p$-dimensional Euclidean space, the objective of fuzzy clustering is to classify $X$ into $c$ clusters according to a performance index. As a prerequisite for further discussion, let us introduce the following notations used in this paper:

- $X = \{x_1, x_2, \ldots, x_n\}$: a data set in a $p$-dimensional Euclidean space $\mathbb{R}^p$ with its ordinary Euclidean norm $\| \cdot \|$, where $x_j = (x_{j1}, x_{j2}, \ldots, x_{jp})^T$ represents the $j$th feature point;
- $c$: a specified number of clusters with $2 \leq c < n$;
- $m$: the weighting exponent called fuzzifier with $m > 1$;
3.2. Constraints of credibilities in fuzzy clustering

There is a basic assumption in fuzzy clustering, that is, the belongingness of each feature point to the $c$ clusters is assumed to be fuzzy. In other words, there are $n$ fuzzy subsets $\tilde{A}_j$ on $S = \{S_1, S_2, \ldots, S_c\}$ with the respective membership function $\mu_j : S \to [0, 1]$ defined by
\begin{equation}
\mu_j(S_i) = \mu_{ij} \quad \text{for all } i,
\end{equation}
where the value $\mu_{ij}$ represents the membership degree of $x_j$ belonging to $S_i$. For convenience, denote by $x_j \in S_i$ the fuzzy event that $x_j$ belongs to $S_i$, and denote by $\text{Pos}_{ij}$ and $\text{Cr}_{ij}$ the possibility and the credibility that $x_j \in S_i$, respectively. Our purpose in this section is to discuss the relationships among $\mu_{ij}$, $\text{Pos}_{ij}$, and $\text{Cr}_{ij}$ as well as their constraints. Without loss of generality, we take the $j$th feature point $x_j$ as an example in the following discussion.

First let us consider $\text{Pos}_{ij}$ and $\mu_{ij}$. Since $\text{Pos}_{ij}$ is the possibility that $x_j \in S_i$, it follows from the possibility theory initiated by Zadeh\(^{20}\) that
\begin{equation}
\text{Pos}_{ij} = \text{Pos}\{x_j \in S_i\} = \mu_{ij} \quad \text{for all } i.
\end{equation}
Note that the possibility measure Pos in (17) should be denoted as $\text{Pos}_j$ completely because each fuzzy set $\tilde{A}_j$ could have its corresponding possibility measure. Here we write $\text{Pos}_{ij}$ as $\text{Pos}$ for the sake of simplicity.

Based on Axioms 1–3 in Definition 1, it is easy to deduce that
\begin{equation}
0 \leq \text{Pos}_{ij} \leq 1 \quad \text{for all } i.
\end{equation}
In addition, the normality of Pos, namely Axiom 1 in Definition 1, implies that
\begin{equation}
\text{Pos}\{\bigcup_{i=1}^c \{x_j \in S_i\}\} = \text{Pos}\{\Theta\} = 1,
\end{equation}
which means that the fuzzy event that $x_j$ belongs to at least one cluster has the maximum possibility 1. In other words, each feature point $x_j$ is supposed not to be a noise datum. Finally, it follows from Axiom 3 in Definition 1 and (19) that
\begin{equation}
\text{Pos}\{\bigcup_{i=1}^c \{x_j \in S_i\}\} = \sup_{1 \leq i \leq c} \text{Pos}\{x_j \in S_i\} = \sup_{1 \leq i \leq c} \text{Pos}_{ij} = 1.
\end{equation}
Thus, on the basis of the three axioms in Definition 1, we obtain the constraints in (18) and (20) for the possibilities \( \text{Pos}_{ij} \), i.e.,

\[
\begin{align*}
0 \leq \text{Pos}_{ij} & \leq 1 \quad \text{for all } i, j, \\
\sup_{1 \leq i \leq c} \text{Pos}_{ij} & = 1 \quad \text{for all } j.
\end{align*}
\]

(21)

Furthermore, by combining (17), (18) and (20), the membership degrees \( \mu_{ij} \) should satisfy the following constraints,

\[
\begin{align*}
0 \leq \mu_{ij} & \leq 1 \quad \text{for all } i, j, \\
\sup_{1 \leq i \leq c} \mu_{ij} & = 1 \quad \text{for all } j.
\end{align*}
\]

(22)

Next we discuss the credibilities \( \text{Cr}_{ij} \), i.e., \( \text{Cr}\{x_j \in S_i\} \). At first, based upon (14), (15), (17), and Axiom 3 in Definition 1, we obtain

\[
\text{Cr}\{x_j \in S_i\} = \frac{1}{2} (\text{Pos}\{x_j \in S_i\} + \text{Nec}\{x_j \in S_i\})
\]

\[
= \frac{1}{2} (\text{Pos}\{x_j \in S_i\} + 1 - \text{Pos}\{x_j \not\in S_i\})
\]

\[
= \frac{1}{2} (\mu_{ij} + 1 - \text{Pos}\{\cup_{k \neq i}\{x_j \in S_k\}\})
\]

\[
= \frac{1}{2} \left(\mu_{ij} + 1 - \sup_{k \neq i} \mu_{kj}\right).
\]

(23)

That is,

\[
\text{Cr}_{ij} = \frac{1}{2} \left(\mu_{ij} + 1 - \sup_{k \neq i} \mu_{kj}\right) \quad \text{for all } i, j.
\]

(24)

After that, according to the sufficient and necessary conditions of the credibility measure in Theorem 1, we can prove the following theorem.

**Theorem 2.** Let \( \text{Cr}_{ij} \) denote the credibility of the fuzzy event \( x_j \in S_i \). Then we have

(i) \( 0 \leq \text{Cr}_{ij} \leq 1 \) for all \( i, j \);

(ii) \( \sup_{1 \leq i \leq c} \text{Cr}_{ij} \geq 0.5 \) for all \( j \);

(iii) \( \text{Cr}_{ij} + \sup_{k \neq i} \text{Cr}_{kj} = 1 \) for any \( i, j \) with \( \text{Cr}_{ij} \geq 0.5 \).

**Proof.** The statement (i) follows immediately from Theorem 1. We next prove the statements (ii) and (iii).

Firstly, from (24), we know that

\[
\sup_{1 \leq i \leq c} \text{Cr}_{ij} = \frac{1}{2} \sup_{1 \leq i \leq c} \left(\mu_{ij} + 1 - \sup_{k \neq i} \mu_{kj}\right) \quad \text{for all } j.
\]
For each \( j \), assume that \( \mu_{ij} \) is the maximal one in \( \{ \mu_{1j}, \mu_{2j}, \ldots, \mu_{cj} \} \). Then we have \( \mu_{ij} = 1 \) and \( \sup_{k \neq i} \mu_{kj} \leq 1 \). Thus

\[
Cr_{ij} = \frac{1}{2} \left( \mu_{ij} + 1 - \sup_{k \neq i} \mu_{kj} \right)
\]

which implies that the statement (ii) \( \sup_{1 \leq i \leq c} Cr_{ij} \geq 0.5 \) holds for all \( j \).

Secondly, let us prove the statement (iii). For each \( j \), if \( Cr_{ij} \geq 0.5 \) for some \( i \), then from the definition of the credibility measure, it is easy to deduce that \( Pos_{ij} = \mu_{ij} = 1 \). Thus for any \( k \neq i \), it follows from (24) that

\[
Cr_{kj} = \frac{1}{2} \left( \mu_{kj} + 1 - \sup_{l \neq k} \mu_{lj} \right)
\]

Besides, according to the self-duality of the credibility measure, we get

\[
Cr\{x_j \in S_i\} + Cr\{x_j \notin S_i\} = 1.
\]

Then it follows from \( Cr_{ij} \geq 0.5 \) that \( Cr\{x_j \notin S_i\} \leq 0.5 \), which may deduce that

\[
Cr\{x_j \notin S_i\} = Cr\{x_j \notin S_i\} \wedge 0.5.
\]

Since \( \{x_j \notin S_i\} = \cup_{k \neq i}\{x_j \in S_k\} \), according to the part (iv) of Theorem 1 and (26), we obtain

\[
Cr\{x_j \notin S_i\} \wedge 0.5 = Cr \{ \cup_{k \neq i}\{x_j \in S_k\} \} \wedge 0.5
\]

Combining (27)–(29), we finally get \( Cr_{ij} + \sup_{k \neq i} Cr_{kj} = 1 \).

### 3.3. Credibilistic clustering model

Until now, we have discussed the constraints that the credibilities \( Cr_{ij} \) are expected to meet in fuzzy clustering. Now let us turn to provide an appropriate performance index for clustering by virtue of the credibility measure, which is usually set as a measurement of the compactness of the data set. Recall that in FCM, the performance index used is \( J_{FCM} \) defined in (1), which is exactly the sum of squares of distances between the feature points and the cluster centers with weights \( (\mu_{ij})^m \). It is obvious that the constraints in (2) for \( \mu_{ij} \) in \( J_{FCM} \) are essentially for the probability measure. That is, the membership \( \mu_{ij} \) in FCM is actually the probability of the event \( x_j \in S_i \) even though it is called “membership”. Hence \( \mu_{ij} \) in \( J_{FCM} \) is also referred to as the probabilistic membership in the literature. Besides, it is notable that the probability measure has the self-dual property. In FCM, the self-dual property implies that if \( \mu_{ij} = 1 \), then the feature point \( x_j \) must belong to the
cluster $S_i$; if $\mu_{ij} = 0$, then $x_j$ must not belong to $S_i$. This property theoretically ensures the performance index $J_{FCM}$ to be a good measurement of compactness of the data set $X = \{x_1, x_2, \ldots, x_n\}$.

Meanwhile, in the possibilistic clustering algorithms, e.g., PCA06, the performance index used is $J_{PCA06}$ in (7), where $\mu_{ij}$ is fundamentally the possibility that $x_j$ belongs to $S_i$. So $\mu_{ij}$ in $J_{PCA06}$ is also called the possibilistic membership. It is known that the possibility measure defined by Zadeh\(^{20}\) has no self-dual property, which leads to the fact that a fuzzy event may fail even though its possibility achieves 1. In terms of possibilistic clustering, it follows that the point $x_j$ possibly does not belong to the cluster $S_i$ even if $\text{Pos}_{ij} = 1$ holds. For this reason, it is indeed a dangerous situation to measure the compactness of a data set via a performance index with the possibilistic membership weights like $J_{PCA06}$. In practice, a great deal of numerical experiments have shown that the algorithms derived from the possibilistic clustering approach are not very stable in most situations with random initializations.

A self-dual measure is absolutely important in both theory and practice. As shown in Theorem 1, the credibility measure $\text{Cr}$ defined by Liu and Liu\(^9\) in Definition 3 is a self-dual measure. In fuzzy clustering, the self-dual property of the credibility measure implies that if the credibility $\text{Cr}_{ij} = 1$, it is affirmative that the point $x_j$ belongs to the cluster $S_i$, and if $\text{Cr}_{ij} = 0$, the point $x_j$ is absolutely outside $S_i$. Since the performance index in clustering is supposed to measure the compactness of the given data set, it is reasonable to use the $m$th power of the credibility $(\text{Cr}_{ij})^m$ as the weight of $||x_j - a_i||^2$ in the performance index like $J_{FCM}$.

Based upon the above discussion, a new performance index for fuzzy clustering by means of the credibility measure is suggested as follows,

$$J_{CC}(\text{Cr}, A) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\text{Cr}_{ij})^m ||x_j - a_i||^2.$$  \hspace{1cm} (30)

Here $\text{Cr} = (\text{Cr}_{ij})_{c \times n}$ is called the credibility matrix. It is obvious that the smaller $J_{CC}$ is, the better the fuzzy partition $(\text{Cr}, A)$ is, which implies that $J_{CC}$ can be used as a measurement of the compactness of the data set $X$. Moreover, since $\text{Cr}_{ij}$ is assumed to represent the credibility that $x_j$ belongs to $S_i$, it follows from Theorem 2 that $\text{Cr}_{ij}$ should satisfy the constraints

$$\begin{cases} 
\sup_{1 \leq i \leq c} \text{Cr}_{ij} \geq 0.5 & \text{for all } j \\
\text{Cr}_{ij} + \sup_{k \neq i} \text{Cr}_{kj} = 1 & \text{for any } i, j \text{ with } \text{Cr}_{ij} \geq 0.5 \\
0 \leq \text{Cr}_{ij} \leq 1 & \text{for all } i, j.
\end{cases}$$ \hspace{1cm} (31)

As a result, combining the objective function in (30) with the constraints in
(31), we obtain the following clustering optimization model,
\[
\begin{align*}
\min J_{CC}(C_r, A) &= \sum_{i=1}^{c} \sum_{j=1}^{n} (C_{rij})^m ||x_j - a_i||^2 \\
\text{subject to:} & \\
\sup_{1 \leq i \leq c} C_{rij} &\geq 0.5 \quad \text{for all } j \\
C_{rij} + \sup_{k \neq i} C_{kij} &= 1 \quad \text{for any } i, j \text{ with } C_{rij} \geq 0.5 \\
0 &\leq C_{rij} \leq 1 \quad \text{for all } i, j.
\end{align*}
\]

By solving this model, the optimal fuzzy partition \((C_r, A)\) minimizing the objective function \(J_{CC}\) can be obtained. The clustering method based upon model (32) is called credibilistic clustering due to its credibilistic weights \((C_{rij})^m\) in the objective function, and the clustering algorithms dealing with model (32) are referred to as the credibilistic clustering algorithms (CCAs). Similarly to FCM, \(C_{rij}\) can be also called the credibilistic membership and expressed as \(\mu_{ij}\) in model (32). In this paper, we still use the symbol \(C_{rij}\) rather than \(\mu_{ij}\).

4. A New Clustering Algorithm

Until now, a clustering model for credibilistic clustering has been constructed. In this section, our purpose is to present an effective algorithm for solving this model. Motivated by the idea of alternating cluster estimation (ACE) in Ref. 13, the solution scheme of ACE is adopted in the proposed credibilistic clustering algorithm, i.e., to update the credibility matrix \(C_r\) and the cluster center matrix \(A\) alternately, in which the updating equations for both \(C_r\) and \(A\) are directly determined. The following discussion suggests how to present appropriate updating equations separately during the iteration procedure.

Firstly, consider the updating equation for the cluster center matrix \(A\). It is natural to derive the updating equation for \(A\) from the necessary conditions for a minimizer \((C_r, A)\) of \(J_{CC}\) analogously with FCM and PCAs. From this point of view, we have the following theorem.

**Theorem 3.** Suppose that \(X = \{x_1, x_2, \ldots, x_n\}\) is a set of feature points, \(C_r\) is a given matrix \((C_{rij})_{c \times n}\) of credibility values, and \(A = (a_1, a_2, \ldots, a_c)\) is the cluster center matrix. Then \(A\) may be a global minimum of \(J_{CC}(C_r, A)\) only if
\[
a_i = \frac{\sum_{j=1}^{n} (C_{rij})^m x_j}{\sum_{j=1}^{n} (C_{rij})^m} \quad \text{for all } i.
\]

**Proof.** In order to derive the necessary conditions on the cluster centers for optimization, we first note that the columns of \(A\) are independent of each other. Hence, minimizing \(J_{CC}(C_r, A)\) with respect to \(A\) is identical to minimizing the following
individual objective function with respect to each $a_i$:

$$J_i(a_i) = \sum_{j=1}^{n} (C_{rij})^m ||x_j - a_i||^2.$$  \hspace{1cm} (34)

That is,

$$J_i(a_{i1}, a_{i2}, \ldots, a_{ip}) = \sum_{j=1}^{n} \sum_{k=1}^{p} (C_{rij})^m (x_{jk} - a_{ik})^2.$$  \hspace{1cm} (35)

Since $a_{ik}$ are independent of each other for all $k$, for each $i$ and $k$, differentiating $J_i$ with respect to $a_{ik}$ and setting it to 0, we obtain

$$\frac{\partial J_i}{\partial a_{ik}} = 2(C_{rij})^m \sum_{j=1}^{n} (a_{ik} - x_{jk}) = 0.$$  \hspace{1cm} (36)

It follows that

$$a_{ik} = \frac{\sum_{j=1}^{n} (C_{rij})^m x_{jk}}{\sum_{j=1}^{n} (C_{rij})^m} \text{ for all } i, k,$$  \hspace{1cm} (37)

which are identical to (33) apparently.

Secondly, let us determine the updating equation for the credibility matrix $C_r$. In Sec. 3.2, the relationship between $C_{rij}$ and $\mu_{ij}$ has been derived and shown in (24). It may be a good idea to specify the evaluation equations for the memberships $\mu_{ij}$ first, and then calculate $C_{rij}$ by the use of (24). As to the evaluation functions for $\mu_{ij}$, Zhou and Hung\textsuperscript{23} has suggested the calculation formulae (11) and (12) according to the conception of membership function in the fuzzy set theory, which involve a large number of candidates. In the present paper, from the point of view of simple and easy performance, we adopt the following equations,

$$\mu^*_{ij} = \frac{1}{1 + d_{ij}^2} \text{ for all } i, j,$$  \hspace{1cm} (38)

which is a simple function satisfying the constraints in (11) and (12).

In Sec. 3.2, it has been verified that the memberships in fuzzy clustering should satisfy the constraints in (22) based upon the three axioms of the possibility measure. However, it is obvious that the memberships $\mu^*_{ij}$ calculated by (38) do not always meet the conditions $\sup_{1 \leq i \leq c} \mu^*_{ij} = 1$ for all $j$. In order to obtain the appropriate membership values satisfying the constraints of membership in (22), the following transformation is utilized,

$$\mu_{ij} = \frac{\mu^*_{ij}}{\sup_k \mu^*_{kj}} \text{ for all } i, j.$$  \hspace{1cm} (39)

It is easy to know that $\sup_k \mu^*_{kj} > 0$ for all $j$ since $\mu^*_{ij} > 0$ for all $i, j$. Via the transformation (39), we may deduce that $\sup_{1 \leq i \leq c} \mu_{ij} = 1$ for all $j$ and $0 \leq \mu_{ij} \leq 1$. 

\[ \text{Credibilistic Clustering: The Model and Algorithms} \]
for all \( i, j \). Equations in (39) are called the membership normalization equations in this paper. Afterwards, the credibilities \( C_{ij} \) are calculated as

\[
C_{ij} = \frac{1}{2} \left( \mu_{ij} + 1 - \sup_{k \neq i} \mu_{kj} \right)
\]

for all \( i,j \), (40)

which are called the credibility updating equations in the proposed new algorithm. Before utilizing these equations for computing the credibility values, it is necessary to verify that the credibilities \( C_{ij} \) calculated by (38)–(40) must fall in the feasible domain of model (32).

**Theorem 4.** The credibilities \( C_{ij} \) determined by (38)–(40) satisfy the constraints of the clustering model (32) for all \( i,j \).

**Proof.** First, it is obvious that the memberships \( \mu_{ij} \) defined by (38) and (39) satisfy the constraints of membership, i.e., \( 0 \leq \mu_{ij} \leq 1 \) for all \( i,j \) and \( \sup_{1 \leq i \leq c} \mu_{ij} = 1 \) for all \( j \). Thus a similar proof with that for Theorem 2 can prove that the credibilities \( C_{ij} \) calculated by (40) satisfy all the constraints of credibility in Theorem 2, namely the constraints of the clustering model (32) for all \( i,j \).

Finally, according to Theorems 3 and 4, we suggest an iterative algorithm for solving the clustering model (32) by embedding the updating equations in (33) and (38)–(40) into the scheme of the ACE method, called the credibilistic clustering algorithm, which is described as follows.

Algorithm 4: The credibilistic clustering algorithm (CCA)

**Step 1.** Initialize \( a_i^{(0)} \in \mathbb{R}^p \) for all \( i \), and set a small number \( \epsilon > 0 \) and an iteration counter \( t = 0 \).

**Step 2.** Compute \( (\mu^*_i)^{(t+1)} \) using the membership evaluation equations in (38) for all \( i,j \).

**Step 3.** Compute \( \mu_{ij}^{(t+1)} \) using the membership normalization equations in (39) for all \( i,j \).

**Step 4.** Compute \( C_{ij}^{(t+1)} \) using the credibility updating equations in (40) for all \( i,j \).

**Step 5.** Compute \( a_i^{(t+1)} \) using the cluster center updating equations in (33) for all \( i \).

**Step 6.** Increment \( t \) until \( \max_i \| a_i^{(t+1)} - a_i^{(t)} \| < \epsilon \) or a predetermined iteration number achieves.

5. **Computational Experiments**

In this section, we analyze the cluster validity of the proposed credibilistic clustering algorithm (CCA) from different aspects. Results are presented by comparing FCM, PCA06 with CCA so as to illustrate the performance of the new algorithm. It is
known that the existing PCAs include some other algorithms, e.g., the PCAs in Refs. 2 and 3. Since the performance of those algorithms heavily depend on the parameters which are not easy to be appropriately predetermined by the users in practice, we only take the PCA06 algorithm as a representative of the existing PCAs. In addition, the parameter $m$ is conformably set as 2 when running all the algorithms with random initializations in this section.

Three randomly generated data sets are used in this section, named Dataset1, Dataset2, and Dataset3, as depicted in Figs. 1(a), 2(a), and 3(a), respectively. Table 1 shows the detailed description of these data sets. They are implemented to estimate different validity problems. Detailed computational experiments and results are given in the following subsections.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Data number</th>
<th>Attribute number</th>
<th>Cluster number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset1</td>
<td>4000</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Dataset2</td>
<td>7000</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Dataset3</td>
<td>200</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

5.1. Coincident problem

Krishnapuram and Keller$^3$ pointed out that the PCAs suffer from the coincident problem because of only concern about their own cluster memberships. In order to judge whether CCA can perform well without outputting coincident clusters, two randomly generated data sets, Dataset1 and Dataset2, are utilized.

Dataset1 has $n = 4000$ points in a 2-dimensional space that represents four clusters, each with 1000 points (see Fig. 1(a)). This data set is generated by four different center points, i.e., $(−5, 3)$, $(5, 5)$, $(10, 12)$, $(15, 17)$, and the radius of each circle is 3. Dataset2 has $n = 7000$ points in a 2-dimensional space that represents seven clusters, each with 1000 points (see Fig. 2(a)). This data set is generated by seven different center points, i.e., $(−12, −12)$, $(−5, −5)$, $(−12, 10)$, $(−2, 3)$, $(4, 4)$, $(4, 12)$, $(4, −11)$, and the radii of the seven circles are 2, 3, or 4.

FCM, PCA06 and CCA are run 100 times to partition Dataset1 and Dataset2 with cluster number $c = 4$ and $c = 7$, respectively. As a result, 100 fuzzy partitions are obtained for each algorithm. By assigning each feature point $x_j$ to the cluster $S_i$ with the maximum membership (or credibility) value, i.e., $\mu_{ij} = \max_k \mu_{kj}$ (or $Cr_{ij} = \max_k Cr_{kj}$), the fuzzy classification results can be translated to the related crisp partitions. After the transformation, the final clustering results are shown in Figs. 1 and 2. For Dataset1, FCM performs well with correct classification like Fig. 1(a) for 100 times. CCA gives right classification for 94 times like Fig. 1(a), and the rest results still have four clusters, however, with poor partitions. Besides, PCA06 reports the clustering results in Figs. 1(b) and 1(c) for 78 times, in which
the real cluster numbers are actually 3 and 2, respectively. As for the rest 27 times experiments, the real cluster number of the clustering results is 4, but with a poor partition as shown in Fig. 1(d). In view of the results dealing with Dataset2, since the three algorithms do not provide the exactly same clustering results among 100 experiments, we only consider most results for each algorithm. CCA offers the correct classification for 83 times like Fig. 2(a). FCM gives the partitions for 98 times like Fig. 2(b). PCA06 reports a poor partition with only 2 clusters for 100 times as shown in Fig. 2(c).

Both results processed with Dataset1 and Dataset2 show that PCA06 suffers from the coincident problem seriously, while FCM and CCA can survive concerning this problem. On the basis of the experiments on the two data sets, it seems that FCM is more stable than CCA possibly. Besides, it should be noted that CCA performs better than FCM in the second case, which can be recognized from the area labelled with the arrow mark in Fig. 2(b).
5.2. Noisy environment

Noise is a persistent phenomenon happened on desired information. It is often met when dealing with the real data in engineering. Noise may be characterized by deterministic or stochastic measures. The data set used in this section, called Dataset3, is randomly generated with \( n = 200 \) points in a \( p = 2 \) dimensional space, consisting of \( c = 2 \) clusters with 100 points in each cluster labelled by circle and cross, respectively (see Fig. 3(a)). In order to examine the performance of the three algorithms in a noisy environment, 10 points representing noise are added to Dataset3, labelled by triangle in Fig. 3(a).

Similarly, FCM, PCA06 and CCA are implemented with Dataset3 for 100 times, and report the crisp clustering results by assigning each point to the cluster with the maximum membership (or credibility) value. The corresponding performance results are shown in Figs. 3(b)–3(d). Comparing classification results from different algorithms, we may find that PCA06 fails to detect the cluster structure due to the disturbance of noise for 54 times. However, CCA can perform well for 100 times and find the real clusters without suffering from the noisy environment. In other words, CCA is more robust in the noisy environment than PCA06. Furthermore,
it is extremely interesting to see that FCM also performs well for 100 times with the clustering results as shown in Fig. 3(b), which means that FCM is more robust to noise than PCA06. It is exactly a contrary evidence to the statement in Ref. 2. Certainly, in order to verify this assertion, more experiments are required necessarily.

6. Conclusions

Fuzzy clustering is an approach using the fuzzy set theory as a tool in clustering, and it has been shown to be advantageous over traditional clustering in many applications. Possibilistic clustering algorithms were initiated to endow the memberships used in FCM with a more intuitive meaning. However, most PCAs suffer from the coincident problem and cannot provide good clustering results especially when a number of outliers or noise exist. As an improvement over possibilistic clustering, in this paper, a new clustering method called credibilistic clustering was developed by introducing the credibility measure into the field of cluster analysis. Compared
with the possibility measure, the credibility measure has very good mathematical properties including the self-duality, which ensures that the performance index by virtue of the credibilistic weights can measure the compactness of the data set well. Thus the credibilistic clustering method was proposed by constructing a clustering optimization model, where the objective function is the sum of squares of distances within class with credibilistic weights \((Cr_{ij})^m\), and the constraints are deduced from the sufficient and necessary conditions of the credibility measure. To tackle the proposed model with good partitions, we designed a credibilistic clustering algorithm by integrating alternating cluster estimation initiated by Runkler and Bezdek,\(^1\) instead of alternating optimization generally used in fuzzy clustering, into the clustering process. To explain more clearly, the updates of the cluster centers \(a_i\) are just the necessary conditions of minimizing the proposed objective function \(J_{CC}\), while the updates of the credibilities \(Cr_{ij}\) are determined directly by the users according to some criteria, which are not the necessary conditions of minimizing the proposed objective function. The numerical experiments in this paper revealed that the proposed new algorithm could greatly eliminate the coincident problem. Besides, according to the numerical experiments on the data sets presented in this paper, both CCA and FCM can perform well under the noisy environments with good clustering results. The difference between the two algorithms is that FCM reports the probabilistic memberships in the final clustering results, and CCA presents the credibilities as well as the possibilistic memberships to the users. In order to compare the clustering performance of CCA with those of FCM and PCAs, more cluster validity should be utilized in the future work.

Acknowledgements

This work was supported by a grant from the Innovation Program of Shanghai Municipal Education Commission (No. 13ZS065).

References