THEORETICAL ANALYSIS OF PICTURE FUZZY CLUSTERING

PHAM THI MINH PHUONG¹, PHAM HUY THONG, LE HOANG SON

Vietnam National University, Ha Noi, Viet Nam
¹phamthiminhphuong.t60@hus.edu.vn

Abstract. Recently, picture fuzzy clustering (FC-PFS) has been introduced as a new computational intelligence tool for various problems in knowledge discovery and pattern recognition. However, an important question that was lacked in the related researches is examination of mathematical properties behind the picture fuzzy clustering algorithm such as the convergence, the boundary or the convergence rate, etc. In this paper, we will prove that FC-PFS converges to at least one local minimum. Analysis on the loss function is also considered.

Keywords. Convergence analysis, picture fuzzy sets, picture fuzzy clustering.

1. INTRODUCTION

One of the most efficient tools in pattern recognition and knowledge discovery is fuzzy clustering in which the uncertainty and vagueness of data can be handled successfully. Fuzzy clustering, as its reminiscent names recalled, uses a membership function to assign for each data elements in the original dataset. The decision of an appropriate cluster depends on the membership values, that is to say, a greater one implies the inclusion. Fuzzy clustering successfully handle the problem of crisp clustering in which a data element can belong to many clusters at the same time [1, 2]. However, it was deployed on the traditional fuzzy set, which shows some limitations in dealing with practical scenarios like voting [3].

A new extension of the fuzzy set called the Picture Fuzzy Set (PFS) was presented by Cuong in [3, 4] to handle such the problem. A PFS is characterized by three membership degrees: positive, neutral, and negative degrees. In the real case of voting applications, ‘positive’ refers to the support for a candidate, ‘negative’ in reverse shows the opposition, and ‘neutral’ reflects the hesitant group who do not agree and disagree. There are many other cases to demonstrate the usage and practical necessity of the PFS [5].

Picture Fuzzy Set has been applied to decision making problems as in the works of Wei [6, 7, 8, 9, 10]. In these researches, the authors have applied picture aggregation opear tors and picture fuzzy entropy in multi-attribute decision problems. Some new operators based on the cosine function and their weighted variants have been utilized for recommendation of products [8]. In [9], picture Bonferroni mean operators have been given in the view of software suppliers. The 2-tuple linguistic picture operators were also examined in [10, 11]. Yang et al. [12] extended the notion of picture fuzzy soft set. Other decision making procedures in the picture fuzzy set can be retrieved in [13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. Wei [23] summarized some similarity measures in the picture fuzzy set. Indeed, Singh [24] proposed correlation coefficients for picture fuzzy sets. Zhang [25] designed Picture Fuzzy
Filters. Other researches regarding picture operators, picture fuzzy rules and database can be found in [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37].

Picture fuzzy clustering (FC-PFS) is a generalization of the traditional fuzzy clustering algorithm [38]. By adding a new membership to the fuzzy set to denote the vagueness of prototype parameters, the FC-PFS has already covered situations that require human opinions as in above. It gives precise results for clustering which has been proven through numerous researches recently [39, 40, 41, 42, 43, 44, 45]. FC-PFS showed significant roles in weather nowcasting from satellite image sequences [42], brain tumor segmentation [5], recommender systems [40], and stock prediction [39].

However, to create a solid constructed basis for the algorithm, it is necessary to perform the theoretical analysis. Proving the convergence of picture fuzzy clustering is of an important role in understanding the algorithm and how it is evolved. In this paper, the convergence of the FC-PFS algorithm is proven and some properties of its such as the boundary of the loss function are expanded. The similarities and differences between this algorithm and other clustering methods are compared. Analysis on the loss function is also considered.

Section 2 recalls the general definition of the picture fuzzy set. The convergence accompanied with some propositions is proven followed by the Zangwill theorem in Section 3. Section 4 validates the way to calculate the boundary of the loss function and describe the changing of the loss function until convergence. The final section draws the conclusion and delineates the future research directions.

2. PRELIMINARY

Definition 1. [3] A picture fuzzy set (PFS) $E$ on the universe $Y$ is

$$E = (y, \mu_E(y), \eta_E(y), \gamma_E(y)) | y \in Y,$$  \hspace{1cm} (1)

where $\mu_E(y) \in [0, 1]$, $\eta_E(y) \in [0, 1]$, and $\gamma_E \in [0, 1]$ are the positive, neutral, and negative memberships of $y$ in $Y$ satisfying

$$\mu_E(y) + \eta_E(y) + \gamma_E(y) \leq 1, \forall y \in Y.$$  \hspace{1cm} (2)

Definition 2. [38] Assume $Y$ is a dataset of $N$ points in $R$ dimensions and $\mu_{kj} = \mu_{kj}(y)$, $\eta_{kj} = \eta_{kj}(y)$, $\xi_{kj} = \xi_{kj}(y)$, $1 \leq j \leq C$, $1 \leq k \leq N$, $C$ is a number of clusters, $V_j$ is the cluster center $j$, $1 \leq j \leq C$, $m$ is fuzzifier, $\alpha$ is exponent coefficient. The picture fuzzy clustering model is

$$J_m = \sum_{k=1}^{N} \sum_{j=1}^{C} (\mu_{kj}(2 - \xi_{kj}))^m \|Y_k - V_j\|^2 + \sum_{k=1}^{N} \sum_{j=1}^{C} \eta_{kj}(\ln(\eta_{kj} + \xi_{kj}) \rightarrow \min,$$  \hspace{1cm} (3)

where

$$\xi_{kj} = 1 - (\mu_{kj} + \eta_{kj} + \gamma_{kj}),$$  \hspace{1cm} (4)

with constraints

$$\mu_{kj} + \eta_{kj} + \xi_{kj} \leq 1,$$  \hspace{1cm} (5)

$$\mu_{kj} \in [0, 1], \eta_{kj} \in [0, 1], \xi_{kj} \in [0, 1],$$  \hspace{1cm} (6)
THEORETICAL ANALYSIS OF PICTURE FUZZY CLUSTERING

\[ \sum_{j=1}^{C} (\mu_{kj}(2 - \xi_{kj})) = 1, \quad (7) \]
\[ \sum_{j=1}^{C} (\eta_{kj} + \frac{\xi_{kj}}{C}) = 1. \quad (8) \]

Let us denote,

\[ U_c = \{ U = [\mu_{kj}] \in \mathbb{R}^{c \times n} : \mu_{kj} \text{satisfies (1)} \forall i, k \}, \]
\[ N_c = \{ N = [\eta_{kj}] \in \mathbb{R}^{c \times n} : \eta_{kj} \text{satisfies (1)} \forall i, k \}, \]
\[ Z_c = \{ Z = [\xi_{kj}] \in \mathbb{R}^{c \times n} : \xi_{kj} \text{satisfies (1)} \forall i, k \}. \]

It was shown in [38] that \((U^*, V^*, N^*, Z^*)\) might be a local minimum of \(J_m\) if and only if for any \(m > 1\)

\[ \mu_{kj}^* = \frac{1}{\sum_{i=1}^{C} (2 - \xi_{kj}^*) \left( \frac{\|Y_k - V_j^*\|}{\|Y_k - V_i^*\|} \right)^{\frac{m}{m-1}}}, \quad (1 \leq j \leq C, \ 1 \leq k \leq N), \quad (9) \]
\[ \eta_{kj}^* = \frac{e^{-\xi_{kj}^*}}{\sum_{i=1}^{C} e^{-\xi_{ki}^*}} \left( 1 - \frac{1}{C} \sum_{i=1}^{C} \xi_{kj}^* \right), \quad (1 \leq j \leq C, \ 1 \leq k \leq N), \quad (10) \]
\[ \xi_{kj}^* = 1 - (\mu_{kj}^* + \eta_{kj}^*) - (1 - (\mu_{kj}^* + \eta_{kj}^*)^a)^{\frac{1}{a}}, \quad (1 \leq j \leq C, \ 1 \leq k \leq N), \quad (11) \]
\[ V_j^* = \frac{\sum_{k=1}^{N} (\mu_{kj}^* (2 - \xi_{kj}^*))^m Y_k}{\sum_{k=1}^{N} (\mu_{kj}^* (2 - \xi_{kj}^*))^m}, \quad (1 \leq j \leq C, \ 1 \leq k \leq N). \quad (12) \]

The following describes the FC-PFS algorithm [38].

**Picture Fuzzy Clustering algorithm**

1. Input: Data \(Y\) with \(N\) elements; \(C\) is number of clusters, threshold \(\epsilon\); fuzzifier \(m\); exponent \(\alpha\); maxstep \(\geq 0\).
2. Initialize \(\mu_{kj}^0\) ← random, \(\eta_{kj}^0\) ← random, \(\xi_{kj}^0\) ← random, \((1 \leq j \leq C), (1 \leq k \leq N)\) satisfying constraints (5-8).
3. For each iteration \(t\), update \(\mu_{kj}^t, \eta_{kj}^t, \xi_{kj}^t, V_j^t\) following equations (9-12) respectively.
4. Until \(\|\mu^t - \mu^{t-1}\| + \|\eta^t - \eta^{t-1}\| + \|\xi^t - \xi^{t-1}\| < \epsilon\) or \(h > \text{maxstep}\), stop.
5. Output: matrices \(\mu, \eta, \xi\) and centers \(V\).
3. CONVERGENCE OF PICTURE FUZZY CLUSTERING

In this section, we explore some propositions which ensure the convergence of FC-PFS.

**Proposition 1.** Let \( \phi : U_c \rightarrow \mathbb{R} \), \( \phi(U) := J_m(U, V, N, Z) \), where \( V, N, Z \) are fixed. Then, \( U^* \in M_{fc} \) is a strict local minimum of \( \phi \) if and only if \( U^* \) is calculated as in eq. (9).

**Proof.** Since \( \mu_{kj} \) has two constrains in eqs. (4), (6), we consider the relaxed minimization of \( \phi(U) \) via Lagrange multipliers. Let \( \lambda = (\lambda_1, \lambda_2, ..., \lambda_n) \) be the multipliers, and \( L(U, \lambda) \) be the Lagrangian

\[
L(U, \lambda) = \sum_{k=1}^{N} \sum_{j=1}^{C} (\mu_{kj}(2 - \xi_{kj}))^m ||Y_k - V_j||^2 + \sum_{k=1}^{N} \sum_{j=1}^{C} \eta_{kj}(\ln \eta_{kj} + \xi_{kj})
- \lambda_k \left( \sum_{j=1}^{C} (\mu_{kj}(2 - \xi_{kj})) - 1 \right),
\]

Then,

\[
\frac{\partial L(U, \lambda)}{\partial \mu_{kj}} = m\mu_{kj}^{m-1}(2 - \xi_{kj})^m ||Y_k - V_j||^2 - \lambda_k(2 - \xi_{kj}) = 0 \text{ at } U^*,
\]

and calculate the second-order derivative of \( L(U, \lambda) \)

\[
\frac{\partial}{\partial \mu_{st}} \left( \frac{\partial L(U, \lambda)}{\partial \mu_{kj}} \right) = \begin{cases} (m-1)m\mu_{kj}^{m-2}(2 - \xi_{kj})^m ||Y_k - V_j||^2 & \text{if } s = k, t = j \\ 0 & \text{otherwise.} \end{cases}
\]

Now, substitute the updated formula of \( \mu_{kj} \) into the second-order derivation of \( L(U, \lambda) \) to calculate the Hessian matrix \( H(U^*) \). It follows that \( H(U) \) is a nonzero entries matrix, whose diagonal elements are

\[
\alpha_{kj,kj} = (m-1)m(\mu_{kj}^*)^{m-2}(2 - \xi_{kj})^m ||Y_k - V_j||^2
= (m-1)m(2 - \xi_{kj})^m ||Y_k - V_j||^2 \cdot \frac{1}{\left( \sum_{i=1}^{C} (2 - \xi_{kj}) \left( \frac{||Y_k - V_j||^2}{||Y_k - V_i||^2} \right)^{m-2} \right)^{m-2}} > 0
\]

and \( \alpha_{st,kj} = 0 \) for \( s \neq k \) and \( t \neq j \).

The Hessian of \( \phi \) at \( U^* \) has all positive eigenvalues which are \( \alpha_{kj,kj}, 1 \leq k \leq N, 1 \leq j \leq C \). It is sufficient to show that \( U^* \) is a strict local minimum of \( \phi \).

Next, we fix \( U \in M_c, N \in N_c, Z \in Z_c \) and consider the minimization of \( J_m \) in variables \( V = \{V_i\} \).
Proposition 2. Let \( \psi : \mathbb{R}^{CS} \to \mathbb{R} \), \( \psi(V) \coloneqq J_m(U,V,N,Z) \), where \( U,N,Z \) are fixed. Then, \( V^* \) is a strict local minimum of \( \psi \) if and only if \( V^*_i, 1 \leq i \leq C \) is calculated via the updated formula in eq. (12).

Proof. Since there is no constrains for \( V \), in order to minimize \( \psi \) over \( \mathbb{R}^{CS} \), it is necessary to require \( \nabla_V \psi(V^*) \) to vanish for every \( i \),

\[
\frac{\partial \psi(V)}{\partial V_j} = \sum_{k=1}^{N} (u_{kj} (2 - \xi_{kj}))^m (-2y_k + 2V_j) = 0 \text{ at } V^*,
\]

then take the second-order derivative,

\[
\frac{\partial}{\partial V_s} \left( \frac{\partial \psi(V)}{\partial V_j} \right) = \begin{cases} \sum_{k=1}^{N} (u_{kj} (2 - \xi_{kj}))^m > 0 & \text{if } s = j \\ 0 & \text{otherwise.} \end{cases}
\]

The Hessian matrix of \( \psi(V) \) at \( V^* \) has all positive eigenvalues. Therefore, \( V^* \) is sufficient to be minimum point of \( \psi(V) \).

A similar way to prove that \( \eta^* \) is sufficient to minimize \( J_m \) when \( U,V,Z \) are fixed in their spaces.

Proposition 3. Let \( f : N \to \mathbb{R} \coloneqq J_m(U,V,N,Z) \) where \( U,V,Z \) are fixed. Then, \( N^* \) is a strict local minimum of \( f \) if and only if \( \eta_{kj}, 1 \leq k \leq n, 1 \leq j \leq C \) is calculated via eq. (10).

Proof. Since each \( \eta_{kj} \) has it own constrains in eq.(7), we consider the minimization of \( f(N) \) via Lagrange multipliers obtained constrains. Let \( \beta = (\beta_1, \beta_2, \ldots, \beta_n) \) be the multipliers, and \( L(N,\beta) \) be the Lagrangian

\[
L(N,\beta) = \sum_{k=1}^{N} \sum_{j=1}^{C} (\mu_{kj} (2 - \xi_{kj}))^m \| Y_k - V_j \|^2 + \sum_{k=1}^{N} \sum_{j=1}^{C} \eta_{kj}(\ln \eta_{kj} + \xi_{kj})
\]

\[
- \beta_k \left( \sum_{j=1}^{C} \left( \eta_{kj} + \frac{\xi_{kj}}{C} \right) - 1 \right).
\]

Since \( N^* \) is the root of equation system,

\[
\frac{\partial L(N,\beta)}{\partial \eta_{kj}} = \ln \eta_{kj} + 1 - \beta_k + \xi_{kj} = 0,
\]

and

\[
\frac{\partial}{\partial \eta_{st}} \left( \frac{\partial L(N,\beta)}{\partial \eta_{kj}} \right) = \begin{cases} \frac{1}{\eta_{kj}} > 0 & \text{if } s = k; t = j \\ 0 & \text{otherwise.} \end{cases}
\]

The Hessian matrix of \( L(N^*,\beta) \) at \( N^* \) has all positive eigenvalues and \( N^* \) also minimizes \( f(N) \).
The updated formula for the neutral degree is \( h(\zeta_{kj}) = 1 - (\mu_{kj} + \eta_{kj}) - (1 - (\mu_{kj} + \eta_{kj})^9)^{\frac{1}{9}} \), where \( h : \mathbb{Z} \to \mathbb{R} \) which is based on the Yager’s operator. When the neutral and refusal degrees of each elements increase, the entropy decreases. When the update of centroids change in minor variation, \( J_m \) may slowly increase to the convergence point.

Now, to show that the algorithm makes \( J_m \) converge, we use the Zangwill theorem below.

**Proposition 4.** \([1]\) Let \( f : D_f \in \mathbb{R}^m \to \mathbb{R} \) and \( S = \{ y^* \in D_f : f(y^*) < f(y) \forall y \in B^{\alpha}(y^*, r) \} \). Let \( A : D_f \to D_f \) be an iterative algorithm and \( y_{k+1} = E(y_k) ; k = 0, 1, ... \)

If \( E \) is continuous on \( D_f \setminus S \), \( g \) is a descent function for \( A, S \) and the iterative sequences \( E(Y_k) : k = 0, 1, 2, ...; y_0 \in D_f \subset K \) are contained in a compact set \( K \subseteq D_f \) for arbitrary \( y_0 \in D_f \), then for each iterative sequence \( Y_k \) generated by \( E \), we have either \( Y_k \) terminates at a solution \( y^* \in S \) or there exists a subsequence \( y_{k_j} \subset y_k \) so that \( y_{k_j} \to y^* \in S \).

To apply the Zangwill theorem, we need to show that \( J_m \) is a descent function and the algorithm is continuous on \([0, 1]^4 \setminus S\). Then, we only need to show that \( J_m \) is a descent function. Now, let \( P_m \) be the algorithm to update the parameters in eqs. (9-12).

**Proposition 5.** Let

\[
S = \{(U, V, N, Z) : J_m(U, V, N, Z) < J_m(U, V, N, Z), \forall(U, V, N, Z) \in B^{\alpha}((U, V, N, Z), r)\}.
\]

Then \( J_m \) is descent function for \( P_m, S \).

**Proof.** Since the norm function and the exponent function are continuous, we call the sum of products of such functions as \( J_m \). Obviously, \( J_m \) is also continuous on \( M_{fc} \times \mathbb{R}^{CS} \). Suppose \((U, V, N, Z) \notin S \) then

\[
J_m(P_m(U, V, N, Z)) = J_m(P_1 \circ P_2 \circ P_3 \circ P_4(U, V, N, Z))
\]

\[
= J_m(P_1 \circ P_2 \circ P_3(U, V, N, h(Z)))
\]

\[
< J_m(P_1 \circ P_2 \circ P_3(U, V, N, Z))
\]

\[
= J_m(P_1 \circ P_2(U, V, f(N), Z))
\]

\[
< J_m(P_1 \circ P_2(U, V, N, Z))
\]

\[
= J_m(P_1(U, \psi(V), N, Z))
\]

\[
< J_m(P_1(U, V, N, Z))
\]

\[
= J_m(h(U), V, N, Z)
\]

\[
< J_m(U, V, N, Z).
\]

Hence \( J_m \) is a descent function. □

However, in some cases, \( J_m \) will slowly increase because of updating of \( \eta \). Because the difference between period centroids and the next centroids changes very small, it still guarantees that \( J_m \) converges.
4. SOME PROPERTIES

4.1. Property of the loss function

We consider the loss function

\[ J_m = \sum_{k=1}^{N} \sum_{j=1}^{C} (\mu_{kj}(2 - \xi_{kj}))^m \|Y_k - V_j\|^2 + \sum_{k=1}^{N} \sum_{j=1}^{C} \eta_{kj}(\ln \eta_{kj} + \xi_{kj}). \]

We also know that the first part \( J_1 = \sum_{k=1}^{N} \sum_{j=1}^{C} (\mu_{kj}(2 - \xi_{kj}))^m \|Y_k - V_j\|^2 \) converges to a value called \( M \). Now, we find the upper bound and lower bound of the second part. Let

\[ J_2 = \sum_{k=1}^{N} \sum_{j=1}^{C} \eta_{kj}(\ln \eta_{kj} + \xi_{kj}). \]

We see that \( \eta_{kj}(\ln \eta_{kj} + \xi_{kj}) \leq \eta_{kj}(\ln \eta_{kj} + 1 - \eta_{kj}) \leq 0. \)

Indeed, consider \( f(y) = y(\ln y + 1 - y) \) with \( y \in [0, 1] \). Therefore,

\[ f'(y) = 2 + \ln y - 2y. \]

\[ f'(y) = 0 \leftrightarrow y = 1 \text{ or } y = \frac{1}{2} W \left( -\frac{2}{e^2} \right). \]

From \( f(1) = 0 \) and \( f \left( -\frac{1}{2} W \left( -\frac{2}{e^2} \right) \right) < 0 \) we get \( f(y) \leq 0, \forall y \in [0, 1] \).

Therefore, \( \eta_{kj}(\ln \eta_{kj} + \xi_{kj}) \leq 0 \) and it leads to \( J_2 = \sum_{k=1}^{N} \sum_{j=1}^{C} \eta_{kj}(\ln \eta_{kj} + \xi_{kj}) \leq 0 \).

On the other hand,

\[ \eta_{kj}\xi_{kj} \geq 0. \]

Let us consider

\( g(y) = y \ln y \) where \( y \in [0, 1] \),

we have

\[ g'(y) = 1 + \ln y, \]

\[ g'(y) = 0 \leftrightarrow y = \frac{1}{e}, \]

and \( f \left( \frac{1}{e} \right) = -\frac{1}{e} \) is the minimal value of this function.

We have

\[ J_2 = \sum_{k=1}^{N} \sum_{j=1}^{C} \eta_{kj}(\ln \eta_{kj} + \xi_{kj}) \geq \sum_{k=1}^{N} \sum_{j=1}^{C} \left( -\frac{1}{e} + 0 \right) = -\frac{1}{e} \times N \times C. \]
**Figure 1.** Loss function of Haberman dataset [46]

**Figure 2.** Loss function of Wdbc dataset [46]
Figure 3. Loss function of Iris dataset [46]

Figure 4. Loss function of Glass dataset [46]
Therefore, the upper bound of $J_m$ is $M$, the lower bound is $M - \frac{1}{e} \times N \times C$.

In Fig.1, the loss function of Habamen data takes 6 iterations to converge to the local minimum. Because of the update of $\eta$, it increases gradually from the second iteration but still reaches the convergence.

The loss function of the Wdbc dataset in Fig.2 decreases slowly in each iteration and converges at the 16th iteration. Because the initialization elements are random, the value of the first iteration is also random. However, from the second step the value of $J$ decreases significantly and from the 8th step, the stability appears and the loss function slowly converges.

From the 2nd iteration in Figs.3 and 4, the loss functions of Iris and Glass datasets decrease but sometimes they increase slightly and converge to a stable point.

4.2. Property of centroid

We can see the updating of centroids in each iteration similar to the update of parameters in Gradient Descent. From eq. (11), we have

$$V_{j+1}^t = \frac{\sum \limits_{k=1}^{N} (\mu_{kj}^t (2 - \xi_{kj}^t))^m Y_k}{\sum \limits_{k=1}^{N} (\mu_{kj}^t (2 - \xi_{kj}^t))^m}$$

$$= V_j^t - \frac{1}{\sum \limits_{k=1}^{N} (\mu_{kj}^t (2 - \xi_{kj}^t))^m} \left( \sum \limits_{k=1}^{N} (\mu_{kj}^t (2 - \xi_{kj}^t))^m (-2Y_k + 2V_j^t) \right)$$

$$= V_j^t - \alpha_j^t \nabla V_j^t J_m(V_j^t),$$

where $\alpha_j^t = \frac{1}{\sum \limits_{k=1}^{N} (\mu_{kj}^t (2 - \xi_{kj}^t))^m}$.

5. CONCLUSIONS

This paper presented some theoretical properties of FC-PFS and proved the convergence of this algorithm. We have pointed out that this algorithm converges to at least local minimum which guarantees to archive acceptable solutions. Specifically, Propositions 1 to 5 stated that the membership matrices and cluster centers converge if and only if their values are computed by updated equations. Moreover, the objective function is descent in the domain. Some properties of PF-PFS were also considered such as the boundary of the loss function. This is significant in understanding the mechanism of the picture fuzzy clustering.

In the future, we will assess the maximum and minimum changes of the objective function through interval steps and others. Relationship between picture fuzzy set and neutrosophic set [47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60] in terms of clustering algorithms will also be our target.
ACKNOWLEDGMENT

This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2017.02. The authors are greatly indebted to Prof. Bui Cong Cuong for his introduction and guidance of the Picture Fuzzy Set, which turns out to be an interesting and widely-applied researches nowadays.

REFERENCES


Received on June 07, 2018
Revised on July 23, 2018