Adaptive Shadowed Fuzzy C-Means Clustering

IEEE Publication Technology, *Staff, IEEE,*

***Abstract*—Adaptive Shadowed Fuzzy C-Means(ASFCM) explores the impact of varying the fuzzifier value in Shadowed Fuzzy C-Means (SFCM). SFCM utilizes shadowed sets to reduce outlier influence on centroids. However, decreasing the fuzzifier value increases cluster compactness, leading to more distinct groupings but also greater sensitivity to outliers, which can disproportionately influence centroid positions. This study emphasizes selecting an optimal fuzzifier value to balance cluster compactness and resilience to outliers, ensuring robust and meaningful clustering results.**

Moreover, the method underscores the significance of the fuzzifier value in SCM, directly impacting cluster shape and boundaries. Lower fuzzifier values sharpen cluster boundaries, promoting distinct groupings but heightening outlier sensitivity. Higher fuzzifier values introduce more fuzziness, smoothing boundaries but potentially reducing compactness. Thus, finding an optimal fuzzifier value is crucial for balancing these factors and optimizing clustering under varying data conditions.

***Index Terms*—Fuzzy sets, Shadowed sets, Fuzzy C-Means (FCM), Shadowed Fuzzy C-Means (SFCM).**

I. Introduction

CLUSTERING is a fundamental technique in data analysis that involves grouping similar objects into clusters or classes based on their intrinsic characteristics [1], [2]. It plays a crucial role in various domains such as pattern recognition, data mining, and machine learning, enabling tasks such as data summarization, compression, and anomaly detection.

Fuzzy sets, introduced by Zadeh [3], extend traditional binary set theory by allowing objects to belong to multiple clusters simultaneously with varying degrees of membership [4]. This flexibility is particularly useful in scenarios where data points exhibit uncertainty or ambiguity in their classification [5].

In the context of clustering, fuzzy sets enable a more nuanced representation of data relationships, where each data point can belong to different clusters to a certain degree, reflecting its uncertainty or partial membership [4].

The effectiveness of fuzzy sets in clustering lies in their ability to handle overlapping clusters and noisy data, making them suitable for real-world applications where data may not strictly adhere to distinct boundaries [6]. By incorporating fuzzy logic into clustering algorithms, such as Fuzzy C-Means (FCM) [4], researchers have achieved improved clustering accuracy and robustness in various domains.

From the foundational concepts of fuzzy sets proposed by Zadeh [3], which introduced the idea of membership degrees to describe uncertain concepts with flexible boundaries, the evolution towards shadowed sets represents a significant advancement in handling uncertainty and ambiguity in data analysis.

Fuzzy sets allow elements to belong to multiple clusters with varying degrees of membership, transforming classical set theory from binary logic to multivalued logic [3]. This flexibility has found applications across diverse fields including fuzzy control, pattern recognition, fuzzy decision making, and cluster analysis [3], [6].

The transition from fuzzy sets to shadowed sets addresses the challenge of defining clear boundaries for elements that do not fully belong or fully exclude from a set. Initially, Zadeh introduced thresholds, α and β, to categorize elements into three regions: those fully belonging ($\mu_A(x) \geq \alpha$), fully excluded ($\mu_A(x) \leq \beta$), and uncertain ($\beta < \mu_A(x) < \alpha$) [3]. This concept parallels three-way decisions in decision theory, where elements are categorized as positive, negative, or boundary cases [7], [8].

Shadowed sets, introduced by Pedrycz, refine this approach by optimizing the thresholds α and β through an objective function that minimizes uncertainty . This optimization aims to improve the robustness of set boundaries, reducing the impact of outliers and enhancing the clarity of cluster memberships [8], [31], [32], [33], [34].

In Shadowed C-Means (FCM) clustering, selecting an optimal fuzzifier value is crucial. The fuzzifier value influences the level of cluster fuzziness and significantly impacts the performance of the clustering process. A well-chosen fuzzifier value ensures that the clusters are compact without being overly sensitive to outliers, which can distort memberships and centroids. Therefore, we propose an adaptive process for selecting the fuzzifier value. This adaptive approach aims to balance cluster compactness and robustness to outliers, ensuring that the clustering results are both precise and resilient to the presence of outliers.

The remainder of this paper is organized as follows. In Section II, we describe the background on the topics we will be using, including fuzzy entropy, fuzzifier range for certain perturbations, measures and their significance on clustering, and the usefulness of entropy in our method. In Section III, we discuss the methodology, detailing the conflict with visual aids, introducing our method and its potential in selecting a fuzzifier value, and presenting the algorithm. In Section IV, we show some experimental results of the discussed methodology, providing examples of clusters with the general method versus our method, an example showing how the fuzzifier value and other metrics change, and a comparison of metrics across various datasets to demonstrate the effectiveness of our algorithm. Finally, Section V gives the conclusions and summary.

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

Manuscript received April 19, 2021; revised August 16, 2021.

II. Background

The traditional Shadowed C-Means (SCM) algorithm is an extension of the Fuzzy C-Means (FCM) clustering algorithm, designed to enhance clustering robustness by incorporating the concept of shadowed sets. SCM addresses the uncertainty and ambiguity in data classification by assigning each data point a membership degree and an additional shadowed set parameter. This parameter helps manage the indeterminate state between full membership and non-membership.

A. Interval Shadow Set Using Fuzzy Entropy

1) Classical Fuzzy Entropy Measures: Classical fuzzy entropy measures the uncertainty associated with fuzzy sets. In the context of shadowed sets, the thresholds α and β are calculated to minimize the uncertainty in the data. Pedrycz proposed an objective function to determine these thresholds by minimising the change in uncertainity.

Here, the elevated area corresponds to the points with membership degrees greater than α, where the uncertainty is reduced by $1-\mu_A(x)$ because these points are made to fully belong to the set (membership degree 1). The reduced area corresponds to the points with membership degrees less than β, where the uncertainty is reduced by $\mu_A(x)$ because these points are excluded from the set (membership degree 0).

The shadow corresponds to the points with membership degrees between β and α. These points are completely uncertain, so the increase in uncertainty is considered as the cardinality of this set, each point having an uncertainty of 1.

The shadowed sets in the new approach are defined as:

$$S_{\mu_A}(x) = \begin{cases} 1 & \text{if } \mu_A(x) \geq \alpha \\ 0 & \text{if } \mu_A(x) \leq \beta \\ [0,1] & \text{if } \beta < \mu_A(x) < \alpha \end{cases}$$

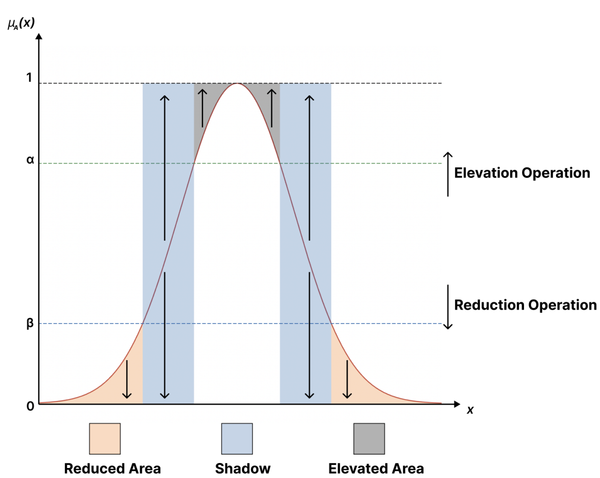

Fig. 1. Pedrycz's shadowed set of a fuzzyset.

A systematic method to calculate the optimal thresholds (α, β) was introduced by Pedrycz [8]:

$$\text{ElevatedArea}(\alpha,\beta)(\mu_A) + \text{ReducedArea}(\alpha,\beta)(\mu_A) = \text{Shadow}(\alpha,\beta)(\mu_A).$$

In this case, the sum of the elevated area and the reduced area is equal to the shadow area. However, in applications, finding the optimal thresholds (α, β) that satisfy the condition can be challenging, turning into an optimization problem. For any given fuzzy set, the optimal thresholds (α, β) can be found by solving the minimization problem:

$$\arg\min_{(\alpha,\beta)} V(\alpha,\beta)(\mu_A),$$

where the objective function $V(\alpha,\beta)(\mu_A)$ is:

$$V(\alpha,\beta)(\mu_A) = |\text{ElevatedArea}(\alpha,\beta)(\mu_A) + \text{ReducedArea}(\alpha,\beta)(\mu_A) - \text{Shadow}(\alpha,\beta)(\mu_A)|$$

For a continuous universe U, we have:

$$V(\alpha,\beta)(\mu_A) = \int_{\mu_A(x)>\alpha} (1-\mu_A(x))\,dx + \int_{\mu_A(x)<\beta} \mu_A(x)\,dx - \int_{\beta<\mu_A(x)<\alpha} dx$$

This approach ensures the minimization of overall uncertainty in the clustering process by appropriately adjusting the thresholds (α) and (β). To reduce the complexity of the calculation, Pedrycz further assumed that the relationship between α and β is $\alpha+\beta=1$.

This function measures the total entropy in the elevated and reduced areas and ensures that the entropy in the shadow area is minimized, thereby reducing uncertainty and providing a more precise representation of the fuzzy set.

2) New Fuzzy Entropy Measures: The new fuzzy entropy measures build on Pedrycz's shadowed set model but introduce a more refined approach to handle the intermediate membership values. The new model keeps the extreme values (0 and 1) unchanged and transforms the unit interval $[0,1]$ into $[\beta,\alpha]$, compressing the shadow area [9].

The shadowed sets in the new approach are defined as follows: The interval shadowed set S^* on the universe U is defined by mapping $S^*: U \to \{0, [\beta,\alpha], 1\}$. Specifically, the membership function μ_{S^*} is given by:

$$\mu_{S^*}(x) = \begin{cases} 1 & \text{if } \mu_A(x) \geq \alpha \\ 0 & \text{if } \mu_A(x) \leq \beta \\ [\beta,\alpha] & \text{if } \beta < \mu_A(x) < \alpha \end{cases}$$

Here, μ_A represents the membership degree of element z in the original fuzzy set A. The aim is to minimize entropy loss during the transformation process.

The new model introduces an optimization function to determine the optimal thresholds (α, β) that minimize entropy loss:

$$
\begin{aligned}
E_{(\alpha,\beta)}(\mu_A) &= |e^*(\text{Elevated Area}) + e^*(\text{Reduced Area}) - e^*(\text{Shadow})| \\
&= \Bigg| \sum_{\mu_A(x)\geq\alpha} \mu_A(x)(1-\mu_A(x)) \\
&\quad + \sum_{\mu_A(x)\leq\beta} \mu_A(x)(1-\mu_A(x)) \\
&\quad - \frac{|\{x|\mu_A(x)\in(\beta,\alpha)\}|}{\alpha-\beta} \int_\beta^\alpha \mu_A(x)(1-\mu_A(x))\, d\mu_A(x) \Bigg| .
\end{aligned}
$$

In general, we choose values such that $\beta+\alpha=1$ so that we can directly minimize the optimization problem and calculate those thresholds.

In this method, points that belong to elevated or complemented areas have a final entropy of zero. Points within the shadowed area are considered to follow a uniform distribution, and we minimize the entropy loss of the change. We will delve deeper into the concept of entropy, its measurement, and further details in subsequent sections. We have chosen this method and Liang's definition of entropy due to the observed minimal entropy loss in these cases [9]

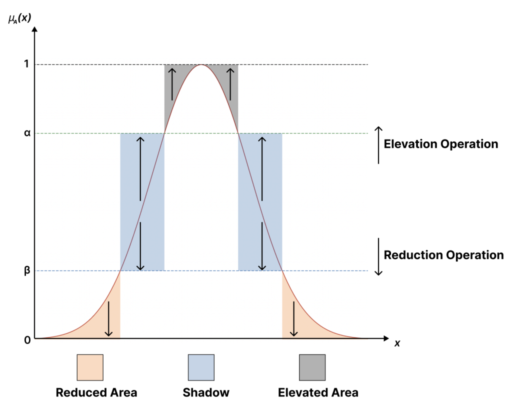

Fig. 2. Interval shadowed set of a fuzzyset

The new fuzzy entropy measures are designed to provide a more accurate representation of fuzzy sets with minimized entropy loss. They are useful in applications requiring precise modeling of uncertainty and fuzziness, such as decision-making and pattern recognition [3], [8].

B. Centroid of SCM based on Rough Set Theory

Building on the concept of shadowed sets, we introduce the SCM. By quantizing membership values (MVs) into core, shadowed, and exclusion regions, SCM achieves reduced computational complexity [8].In SCM, elements in the core region do not have a fuzzy weight factor applied to their MVs. Unlike the uniform computation of MVs in FCM, the MVs for core patterns are set to unity when calculating the centroid. Elements in the shadowed region, which represent a zone of uncertainty, are treated similarly to FCM. However, members of the exclusion region are handled differently. The fuzzy weight factor for exclusion is formulated using the fuzzifier raised to itself, resulting in a double exponential form. The centroid for the i-th class is evaluated as:

$$
\mathrm{v}_i = \frac{\sum_{k\in A} x_k + \sum_{k\in B}(u_{ik})^m x_k + \sum_{k\in C}(u_{ik})^{m^m} x_k}{\Phi_i+\eta_i+\psi_i}
$$

where:

$$
\begin{aligned}
A &= \{k \mid u_{ik} \geq (u_{\text{imax}} - \lambda_i)\} \\
B &= \{k \mid \lambda_i < u_{ik} < (u_{\text{imax}} - \lambda_i)\} \\
C &= \{k \mid u_{ik} \leq \lambda_i\}
\end{aligned}
$$

and:

$$
\begin{aligned}
\Phi_i &= \text{card}(A) \\
\eta_i &= \sum_{k\in B} (u_{ik})^m \\
\psi_i &= \sum_{k\in C} (u_{ik})^{m^m}
\end{aligned}
$$

and λ_i is the corresponding threshold [10]. This arrangement causes a much wider dispersion and a very low bias factor for elements which can generally be considered outside the class under discussion or most definitely, the exclusion members.

C. Fuzzifier Range

The fuzzifier m, where $m \in [1,\infty)$, is a parameter introduced into the clustering function WGSS by Bezdek in 1981 [11]. It plays a crucial role in the Fuzzy C-Means (FCM) algorithm by influencing the level of fuzziness in the resulting clusters. A proper choice of m can suppress noise and smooth the membership function. Despite its importance, there is little theoretical basis for determining the optimal value of m, leading to various heuristic strategies.

Historically, Bezdek [11] suggested an empirical range for m between 1.1 and 5. In 1976, he provided a physical interpretation of the FCM algorithm for $m=2$ [12]. Chan and Cheung (1992), focusing on word recognition, recommended that m should be between 1.25 and 1.75 [14]. Bezdek and Hathaway (1987) [13] considered the convergence of the algorithm and indicated that m should be greater than $\frac{n}{n-2}$. Pal and Bezdek (1995) analyzed cluster validity indices [15] and suggested that m is likely optimal in the range [1.5, 2.5]. Most researchers adopt $m=2$ for practical purposes.

Furthermore, some researchers, including Hwang and Rhee [16] (2007), Yu [17](2003), and Yu et al. [18] (2004), believe that the dataset's structure influences the optimal value of m. Ozkan and Turksen [19] (2004) addressed the uncertainty in m by proposing an entropy-based assessment method.

1) **If $m=1$:**
   - **Explanation:** When the fuzzifier m is set to 1, the FCM algorithm behaves like the Hard C-Means (HCM) algorithm.
   - **Details:** In this case, the membership degrees u_{ik} will be either 0 or 1. Each data point belongs exclusively to one cluster.
   - **Implication:** The algorithm performs hard partitioning of the data, with no overlap between clusters.

2) **If $m \to 1^+$:**

- **Explanation:** As the fuzzifier m approaches 1 from the right, the FCM algorithm increasingly resembles the HCM algorithm.
- **Details:** The membership degrees u_{ik} become almost binary (0 or 1), but not exactly. However, the probability that the algorithm performs a fuzzy partition diminishes.
- **Implication:** The FCM algorithm will likely produce partitions that are nearly hard, with each data point still tending to belong to a single cluster with a high degree of certainty.

3) **If $m \to \infty$:**
- **Explanation:** As the fuzzifier m approaches infinity, the influence of the distance measure on the membership degrees u_{ik} diminishes.
- **Details:** The membership matrix $U = [u_{ik}]$ approaches a uniform distribution where $u_{ik} \approx \frac{1}{c}$ for all i and k, with c being the number of clusters.
- **Implication:** The centers of various clusters converge towards a common point, which is approximately the centroid of the entire dataset. The clustering effect is significantly degraded, and the algorithm performs almost no meaningful partitioning.

Therefore, the fuzzifier m controls the amount of fuzziness of the final C- partition in the FCM and SFCM algorithm and it's very crucial to choose an optimal fuzzifier value.

The FCM membership function is calculated as:

$$\mu_{i,k} = \left[\sum_{t=1}^{c} \left(\frac{\|x_k - v_i\|_A}{\|x_k - v_t\|_A} \right)^{\frac{2}{m-1}} \right]^{-1}$$

where $\mu_{i,k}$ is the membership value of the k-th sample in the i-th cluster such that $\mu_{i,k} \in [0,1]$, c is the number of clusters, x_k is the k-th sample, v_i is the cluster center of the i-th cluster, $\|\cdot\|_A$ is the norm function, and $\sum_{t=1}^{c} \mu_{t,k} = 1$ for a given $m > 1$. This means that the sum of the degrees of membership values of any data point is one, or in other words, any data point should be a member of at least one of the clusters with a membership value greater than zero.

The expression indicates that the membership value is controlled by the fuzzifier m. However, there are two points where the membership values do not depend on m. One of the points is the mass center that has equal distance to all cluster centers and thus has a membership value $1/c$ to all cluster centers. It is identified by cluster centers and a continuous membership function such that it has equal distance from all the cluster centers. In addition, when m goes to infinity, cluster centers collapse to this point. Hence, this value clearly does not depend on m. The other points are the cluster center values which have a membership value 1 in its cluster and 0 to all others. Hence, these values also do not depend on m.

According to the definition of membership, we can obtain two tenable rules which can assist us to find the reasonable range of the value of m.

Rule 1. The membership value of sample p is $\mu_{i,p} \to 1/c$, if p is located in the neighborhood of the mass center.

Rule 2. The membership value of sample q is $\mu_{i,q} \to 1$, if q is located in the neighborhood of the cluster center v_i.

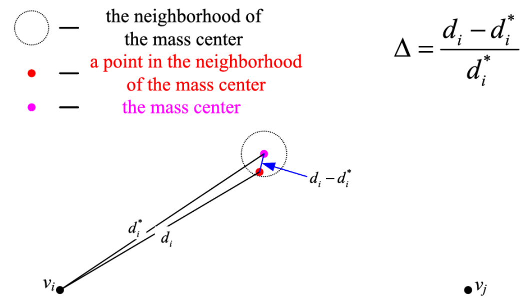

Fig. 3. Graphical Explanation of Δ

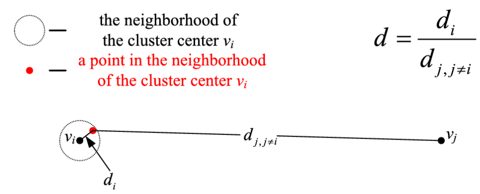

Fig. 4. Graphical Explanation of d

Calculation of $\mu_{i,q}, \mu_{i,p}$:

In order to calculate $\mu_{i,p}$, we expand the function around the mass center by using Taylor series expansion. One-dimensional Taylor series of a real function $f(x)$ about a point $x = x_0$ is given by

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) \\ &\quad + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \cdots \\ &\quad + \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n + \cdots \\ &= f(x_0) + f'(x_0)(x - x_0) + R \end{aligned} \tag{1}$$

where R is the remainder. Let d denote the distance measure to all cluster centers from the mass center, d_i denote the distance to the i-th cluster center of the point located in the neighborhood of the mass center. Then, Taylor series of $\mu_{i,p}$ can be written as

$$\mu_{i,p} = \mu_{i,p}(d_i) + \frac{\partial \mu_{i,p}}{\partial d_i}(d_i - d) + R \tag{2}$$

wherein,

$$\mu_{i,p}\Big|_{d_i=d} = \frac{1}{c}, \quad \frac{\partial \mu_{i,p}}{\partial d_i} = \frac{\sum_{t=1, t\neq i}^{c} \frac{2}{m-1} \left(\frac{d_i}{d_t}\right)^{\frac{2}{m-1}} d_i^{-1}}{\left(\sum_{t=1}^{c} \left(\frac{d_i}{d_t}\right)^{\frac{2}{m-1}} \right)^2} \tag{3}$$

Since the derivative $\frac{\partial \mu_{i,p}}{\partial d_i}$ should be evaluated at the mass center where $d_j = d$, for $j = 1, \ldots, c$, we obtain

$$\frac{\partial \mu_{i,p}}{\partial d_i}\Big|_{d_i=d} = \frac{(c-1)\frac{2}{m-1} \sum_{t=1}^{c} \frac{1}{2}}{d_i^*} = \frac{(c-1)\frac{2}{m-1}}{c^2} \frac{1}{d_i^*} \tag{4}$$

Neglect the remainder R, then

$$\mu_{i,p} \approx \frac{1}{c}\left[1 - \frac{(c-1)\frac{2}{m-1}}{c^2}\frac{d_i - d}{d_i}\right] = \frac{1}{c}\left[1 - \frac{(c-1)\frac{2}{m-1}}{c^2}(\Delta)\right] \tag{5}$$

We can use the membership function directly to calculate $\mu_{i,q}$. Let d_i denote the distance to the i-th cluster center from a point located in the neighborhood of v_i. d_i is very small in value compared to the distance to all the other cluster centers from this point. Thus in general, let $\frac{d_i}{d_j} = d$, for $j \neq i$, then

$$\mu_{i,q} = \left[1 + (c-1)d^{\frac{2}{m-1}}\right]^{-1} \tag{6}$$

The Range of the value of 'm'

The behavior of $\mu_{i,p}$, given that c is a constant for a particular structure of data, $\mu_{i,p}$ depends only on m for small perturbations $\Delta = (d_i - d_i^*)/d_i^*$. As m goes to infinity, membership values go to $1/c$, and as m gets closer to one, this value tends to change rapidly.

Let δ be a threshold, then rule 1 could be transformed into:

$$\left|\mu_{i,p} - \frac{1}{c}\right| \leq \frac{\delta}{c}, \text{ i.e.,}$$

$$\frac{1}{c}(c-1)\frac{2}{m-1}c^2(\Delta)\frac{1}{c} = \frac{c-1}{c^2}\frac{2}{m-1}|\Delta| \leq \frac{1}{c}\delta$$

Therefore,

$$m \geq 1 + \frac{c-1}{c} \cdot \frac{2}{\delta} \cdot |\Delta| \tag{7}$$

Transform rule 2 into $|\mu_{i,q} - 1| \leq \delta$, then

$$\begin{aligned} |\mu_{i,q} - 1| &= \left|\left[1 + (c-1)d^{\frac{2}{m-1}}\right]^{-1} - 1\right| \\ &= \left|1 + (c-1)\frac{d}{2(m-1)} - 1\right| \\ &\leq \delta \end{aligned}$$

Therefore,

$$m \leq \left(\frac{2 \cdot \log(d)}{\log\left(\frac{\delta}{1-\delta} \cdot \frac{1}{c-1}\right)} + 1\right).$$

In general, when the number of clusters is c, the reasonable range of the value of m [20] is

$$\left[1 + \frac{2}{\delta}\left(\frac{c-1}{c}\right), 1 + \frac{2 \cdot \log(d)}{\log\left(\frac{\delta}{1-\delta} \cdot \frac{1}{c-1}\right)}\right]$$

D. Measures and their Significance on Clustering

Understanding the behavior of clustering metrics is crucial for evaluating the quality of clustering solutions. Here, we examine the mathematical formulas and the intuition behind the dependence of three prominent clustering indices on the fuzziness parameter m in the context of memberships-associated clustering.

a) Davies-Bouldin Index (DBI) [21]:

$$\text{DBI} = \frac{1}{C}\sum_{i=1}^{C}\max_{i \neq j}\left(\frac{\sigma_i + \sigma_j}{d_{ij}}\right) \tag{8}$$

where σ_i represents the average distance of all points in cluster i to the centroid of cluster i, d_{ij} is the distance between centroids of clusters i and j, and C is the number of clusters.

As m decreases, clusters become more compact, reducing σ_i. However, cluster centroids tend to be closer, reducing d_{ij}. Since DBI is the ratio of within-cluster scatter to between-cluster separation, a more pronounced reduction in d_{ij} compared to σ_i leads to an overall increase in DBI as m decreases.

b) Xie-Beni Index (XBI) [22]:

$$\text{XBI} = \frac{1}{n}\sum_{i=1}^{C}\min_{i \neq j}\left(\frac{\sigma_i^2}{d_{ij}^2}\right) \tag{9}$$

where σ_i is the scatter within cluster i, d_{ij} is the distance between centroids of clusters i and j, and n is the number of data points.

As m decreases, σ_i^2 decreases and d_{ij} also decreases. Since XBI is the ratio of within-cluster compactness to the square of the minimum separation between clusters, the significant decrease in d_{ij} typically results in a decrease in XBI, indicating tighter and more distinct clusters.

c) Silhouette Index (SI) [23]:

$$\text{SI} = \frac{b-a}{\max(a,b)} \tag{10}$$

where a is the mean intra-cluster distance (average distance of a point to all other points in the same cluster), and b is the mean nearest-cluster distance (average distance of a point to all points in the nearest cluster that it is not a part of).

As m decreases, the intra-cluster distance a decreases while the nearest-cluster distance b increases. The SI measures how similar an object is to its own cluster compared to other clusters. Therefore, a decrease in a and an increase in b leads to a higher SI, reflecting a better clustering solution.

d) Summary of Dependence on m:

- **Davies-Bouldin Index (DBI):** Tends to *increase* as m decreases due to the more pronounced reduction in between-cluster separation.
- **Xie-Beni Index (XBI):** Generally *decreases* as m decreases because the improvement in within-cluster compactness is more significant than the decrease in separation.
- **Silhouette Index (SI):** Tends to *increase* as m decreases due to better intra-cluster compactness and increased separation from other clusters.

Monitoring these metrics when adjusting the fuzzifier value m is crucial. Significant changes in these metrics serve as benchmarks for evaluating the effectiveness of the current fuzzifier value and guiding future adjustments. These metrics indicate the quality of clustering by reflecting the balance between intra-cluster compactness and inter-cluster separation.

E. How Entropy is Useful for Our Method

What is Entropy?

Entropy measures the uncertainty associated with a point's cluster membership. It quantifies the randomness in determining which cluster a point belongs to, reflecting the level of ambiguity or certainty in its classification [24]. High entropy indicates that a point has a more distributed membership across multiple clusters, while low entropy suggests that a point predominantly belongs to a specific cluster.

Entropy can be calculated in multiple ways:

1) **Luca and Termni's Definition** [25]:
   - For finite and discrete fuzzy sets:
$$e(A) = -\sum_{i=1}^{n} \Big[\mu_A(x_i) \ln \mu_A(x_i) + (1-\mu_A(x_i)) \ln(1-\mu_A(x_i))\Big]$$
   - For a continuous universe U:
$$e(A) = -\int_{x\in U} \Big[\mu_A(x) \ln \mu_A(x) + (1-\mu_A(x)) \ln(1-\mu_A(x))\Big] dx$$
2) **Pal's Definition** [26]:
   - For finite and discrete fuzzy sets:
$$e(A) = \sum_{i=1}^{n} \Big[\mu_A(x_i) \exp(1-\mu_A(x_i)) + (1-\mu_A(x_i)) \exp(\mu_A(x_i))\Big]$$
   - For a continuous universe:
$$e(A) = \int_{x\in U} \Big[\mu_A(x) \exp(1-\mu_A(x)) + (1-\mu_A(x)) \exp(\mu_A(x))\Big] dx$$
3) **Liang et al.'s Definition** [27]:
   - For finite and discrete fuzzy sets:
$$e(A) = \sum_{i=1}^{n} \mu_A(x_i)(1-\mu_A(x_i))$$
   - For a continuous universe:
$$e(A) = \int_{x\in U} \mu_A(x)(1-\mu_A(x)) dx$$

In this article, we're using Liang's measure as we can observe the entropy loss is minimum [9]. To explain the problem succinctly, Liang et al.'s [**?**] fuzzy entropy measure is chosen for the following derivation process. Of course, one can also choose other entropy methods to implement the derivation that includes but is not limited to the methods proposed in [25], [26], [27], [28], [29], [30]

Outliers typically have the highest entropy because they do not belong strongly to any single cluster, resulting in high uncertainty in their cluster membership. When an outlier is entirely detached from all clusters, its membership values are evenly distributed, making each membership approximately $\frac{1}{c}$, where c is the number of clusters. This uniform distribution of membership leads to maximum entropy, indicating high randomness and uncertainty.

As the value of the fuzziness parameter m decreases, clusters become more compact. At a certain point, the impact of outliers on cluster centroids increases, causing these uncertain points to become more certain about their membership, thus reducing their entropy.

To identify outliers, we calculate the entropy S of each point after a few iterations of the ASCM (Adaptive Subtractive Clustering Method) algorithm. We then compute the mean μ and standard deviation σ of these entropies. Points with the highest entropy are selected using the threshold:

$$\mu + k\sigma \tag{11}$$

where k is a chosen value based on the desired percentage of data to be selected.

Why Outliers Have Higher Entropy:

- **Distributed Membership:** Outliers do not fit well into any single cluster, resulting in a more even distribution of their membership values across multiple clusters.
- **High Uncertainty:** Due to their lack of strong affiliation with any cluster, outliers exhibit high uncertainty, which is quantified as high entropy.
- **Uniform Distribution:** In extreme cases where outliers are equally detached from all clusters, their membership values are uniform ($\frac{1}{c}$), leading to the maximum possible entropy.

To implement this, select points with entropy greater than $\mu + k\sigma$, where μ and σ are the mean and standard deviation of the entropies of all points. The value of k is chosen based on the desired percentage of outliers to be identified.If the entropies follow a Normal Distribution, the following k values can be used:

- $k = 1.65$ to choose the top 5% of the data.
- $k = 1.28$ to choose the top 10% of the data.
- $k = 1.96$ to choose the top 2.5% of the data.

Why 2.5% is Used by Statisticians

Statisticians often use $k = 1.96$ to choose the top 2.5% of the data because this corresponds to the critical value for a 95% confidence interval in a standard normal distribution.

In a normal distribution:

- 95% of the data lies within $\mu \pm 1.96\sigma$.
- Consequently, the top 2.5% of the data lies above $\mu + 1.96\sigma$.

This threshold is frequently used in hypothesis testing and confidence intervals because it provides a balance between being too conservative and too lenient. It is a standard convention in statistical analysis for determining significance and making inferences about data.

III. PROPOSED METHODOLOGY

Shadow clustering is a technique used to enhance the robustness and accuracy of clustering algorithms. The principal idea behind shadow clustering is to create additional "shadow" clusters that represent regions of uncertainty or noise within the data. These shadow clusters help to mitigate the impact of outliers and ambiguous data points on the primary clustering results.

A. How Shadow Clustering Works

1) **Initialize Primary and Shadow Clusters:**
   - Begin by initializing the primary clusters based on the chosen clustering algorithm (e.g., k-means, fuzzy c-means). Additionally, initialize shadow clusters to capture regions of uncertainty.
2) **Assign Memberships:**
   - Assign each data point a membership value for both primary and shadow clusters. In fuzzy clustering, this involves calculating the degree of belonging of each data point to all clusters, including shadow clusters.
3) **Calculate Thresholds using Fuzzy Entropy:**
   - Instead of the general calculation of α proposed by Pedrycz, we utilize the minimization of entropy concept as outlined in "Fuzzy Entropy: A More Comprehensible Perspective for Interval Shadowed Sets of Fuzzy Sets" by Qinghua Zhang, Yuhong Chen, and Jie Yang. This method involves using shadowed sets, which provide a three-way approximation scheme for transforming the universe of a fuzzy set into three disjoint areas: elevated area, reduced area, and shadow area.
   - By solving a fuzzy entropy loss-minimization problem, we obtain a pair of optimal thresholds, α and $1-\alpha$, that define the range of the shadow area. This approach results in a more accurate and comprehensible measure of fuzzy entropy, leading to better clustering outcomes.
4) **Update Cluster Centers:**
   - Update the cluster centers by assigning weights to data points based on the thresholds α and $1-\alpha$. Points with a membership value greater than $1-\alpha$ are assigned a weight of 1, indicating strong certainty in their cluster membership.
   - Points with membership values within the shadow area (between α and $1-\alpha$) are given partial weights, reflecting their uncertainty and contributing less to the cluster center update.
5) **Final Cluster Assignment:**
   - After a predefined number of iterations or once convergence is achieved, finalize the cluster assignments. Data points consistently assigned to shadow clusters are flagged as outliers or noise. The remaining data points are assigned to the primary clusters based on their highest membership values.

B. Adaptation of Fuzzifier Value in Shadow Clustering

In order to enhance the adaptability of clusters to rapid quality changes, it is beneficial to adjust the fuzzifier value m dynamically.

a) Methodology for Adjusting m: When there is a noticeable improvement in cluster quality with a lower m value, indicating better clustering, it is advantageous to reduce the m value more aggressively. This can quickly bring about improvements in cluster quality. However, when the improvement in cluster quality begins to plateau, further lowering of the m value may not yield significant benefits. At this point, it is advisable to slow down the rate of adjustments.

b) Determining the Stopping Criteria: There exists an ideal range for the m value where a balance is struck between cluster compactness and separation. Once the m value has been reduced to a point where further reductions do not significantly change cluster memberships or assignments, it is likely that the clusters have reached their optimal compactness. Further reductions in m beyond this point may not improve the clusters and could increase their sensitivity to noise and outliers.

It is important to monitor outlier points that do not fit well into any cluster. If the average cluster membership undergoes a sudden change due to outliers, it may be prudent to stop further reductions in the m value.

c) Refined Proposal: Our approach begins with initializing cluster centers randomly and calculating memberships. Following this, we determine a threshold α by minimizing entropy loss. Next, we identify outliers using one of the previously discussed methods.

The fuzzifier m value is then adjusted based on metrics such as the Davies-Bouldin Index (DBI) and Xie-Beni Index (XBI). The adjustment is based on a descent method that considers the change in DBI and XBI values from the previous iteration. This approach ensures a dynamic and responsive adjustment of m.

Simultaneously, we monitor the average entropy of the outliers flagged during the initial stage. If the average entropy decreases significantly, suggesting that outliers are becoming more like a specific cluster, we stop the adjustments to the m value. This step is crucial to avoid overly influencing the results due to outliers.

In summary, this proposal suggests a dynamic and responsive approach to adjust the fuzzifier m value, considering both cluster compactness and the influence of outliers. It integrates metrics like DBI, XBI, and average entropy of outliers to achieve an optimal and effective clustering outcome.

Algorithm 1 Initial Clustering with Shadow Clustering Method (SCM)

```
Initialize M ← 20                                  ▷ Max Iterations
Initialize c_i ← 0.5 for i = 1, ..., n            ▷ Can be random
Initialize ε ← 10^{-1}                             ▷ Max error
Compute μ_ij
for t = 0 to M do
    Compute c_i(t+1)
    if (1/n) Σ_{i=1}^{n} |c_i(t+1) − c_i(t)| < ε then
        break
    end if
    c_i(t) ← c_i(t+1)
end for
return c_i , μ_ij
```

Algorithm 2 Entropy-Based Outlier Detection

Initialize k ▷ Threshold
for $j = 1$ to n **do**
  $e_j = -\sum_{i=1}^{N} [\mu_{ij} \ln \mu_{ij} + (1 - \mu_{ij}) \ln(1 - \mu_{ij})]$
end for
$\bar{e} \leftarrow \frac{1}{n} \sum_{j=1}^{n} e_j$
$\sigma_e \leftarrow \sqrt{\frac{1}{n} \sum_{j=1}^{n} (e_j - \bar{e})^2}$
$\mathcal{O} \leftarrow \{\}$ ▷ Outliers
for $j = 1$ to n **do**
  if $e_j > \bar{e} + k \cdot \sigma_e$ **then**
    $\mathcal{O} \leftarrow \mathcal{O} \cup \{j\}$
  end if
end for
return $\mathcal{O}$

Algorithm 3 Adaptive Fuzzifier Adjustment Using Gradient Descent

Initialize:
$m \leftarrow$ initial fuzzifier value
$\alpha \leftarrow$ initial learning rate
$\varepsilon_{threshold} \leftarrow$ entropy change threshold
$\text{metric}_{threshold} \leftarrow$ metrics change threshold
$\text{conv}_{threshold} \leftarrow$ convergence threshold
$m_{max}, m_{min} \leftarrow$ upper and lower bounds for fuzzifier
Compute initial DBI, XBI, and $\bar{\varepsilon}$ for the clustering
while convergence criteria not met **do**
  Compute $\nabla\text{DBI}, \nabla\text{XBI}, \nabla\bar{\varepsilon}$ w.r.t. m
  $m_{new} \leftarrow m - \alpha \cdot (\nabla\text{DBI} + \nabla\text{XBI} + \nabla\bar{\varepsilon})$
  $m_{new} \leftarrow \max(\min(m_{new}, m_{max}), m_{min})$
  Compute new DBI, XBI, and $\bar{\varepsilon}$ for updated m
  if metric improvements insignificant **then**
    $\alpha \leftarrow \alpha \times 0.9$
  else
    $\alpha \leftarrow \alpha \times 1.1$
  end if
  $\text{DBI}_{prev} \leftarrow \text{DBI}_{new}$
  $\text{XBI}_{prev} \leftarrow \text{XBI}_{new}$
  $\bar{\varepsilon}_{prev} \leftarrow \bar{\varepsilon}_{new}$
  if $|m_{new} - m| < \text{conv}_{threshold}$ **then**
    break
  end if
  $m \leftarrow m_{new}$
end while
return m

IV. EXPERIMENTAL RESULTS

In this section, we will compare the clustering metrics, DBI, XBI and SI, to evaluate the performance of the Adaptive Shadow C-Means algorithm against various other clustering methods. Additionally, we will examine the changes in the fuzzifier and other parameters over iterations during the adaptation process for different datasets.

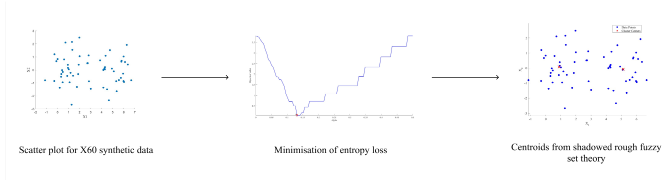

Fig. 5. Adaptation of m value over iterations for Iris Data in ASCM

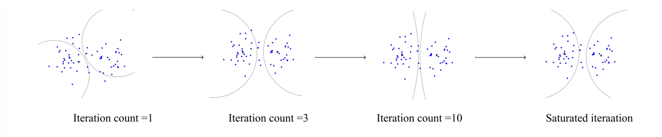

Fig. 6. Clustering Results over Iterations in ASCM for X60 Dataset

TABLE I
CLUSTER VALIDITY INDICES ON IRIS DATA.

Index	*c*	HCM	FCM	RCM	RFCM	SCM	ASCM
Davies Bouldin	3	0.7880	0.8198	0.7892	0.4129	0.3781	**0.2760**
Xie Beni	3	0.1256	0.1371	0.1261	0.0345	0.0222	**0.0169**
Silhouette Index	3	-0.6180	-0.6155	-0.6180	-0.5902	-0.2680	**-0.1225**

TABLE II
CLUSTER VALIDITY INDICES ON VOWEL DATA.

Index	*c*	HCM	FCM	RCM	RFCM	SCM	ASCM
Davies Bouldin	6	0.0724	0.0744	0.0703	0.0673	0.0650	**0.0644**
Xie Beni	6	0.1665	0.1893	0.1795	0.1625	0.1496	**0.1247**
Silhouette Index	6	-0.8541	-0.8410	-0.8467	-0.8157	-0.7574	**-0.7146**

V. SUMMARY AND CONCLUSIONS

In this paper, we propose a new method that adapts the fuzzifier value to balance improving cluster performance and preventing the fuzzifier value from becoming too low, which would increase the impact of outliers on the centroids. Our approach maintains a fuzzifier range that allows for minor perturbations and integrates rough and shadowed set theory to calculate centroids more efficiently.

We calculate the lower and upper bounds of shadowed sets using entropy loss minimization, employing multiple methods to calculate entropy. A gradient descent approach is used to choose the optimal fuzzifier value by adapting it over iterations.

Our results show that this Adaptive Shadowed C-Means approach provides better results than many other clustering methods. This streamlined process offers a robust clustering method that is effective even for data with outliers.

REFERENCES

[1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Elsevier, 2011.
[2] A. K. Jain, M. N. Murty, and P. J. Flynn, *Data Clustering: A Review*. ACM Computing Surveys (CSUR), 1999.
[3] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338-353, 1965.
[4] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Springer Science & Business Media, 1981.
[5] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers Geosciences*, vol. 10, no. 2-3, pp. 191-203, 1984.
[6] D. Dubois and H. Prade, *An Introduction to Fuzzy Systems*, Springer Science & Business Media, 2007.

[7] Y. Yao, "Two views of the theory of rough sets in finite universes," *International Journal of Approximate Reasoning*, vol. 52, no. 2, pp. 212-227, 2011.

[8] W. Pedrycz, "Shadowed sets: Representing and processing fuzzy sets," *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 28, no. 1, pp. 103-109, February 1998.

[9] Q. Zhang, Y. Chen, J. Yang, and G. Wang, "Fuzzy Entropy: A More Comprehensible Perspective for Interval Shadowed Sets of Fuzzy Sets," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 11, pp. 2622-2632, November 2020.

[10] S. Mitra, W. Pedrycz, and B. Barman, "Shadowed c-means: Integrating fuzzy and rough clustering," *Pattern Recognition Letters*, vol. 24, no. 11, pp. 1635-1647, 2003.

[11] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[12] J. C. Bezdek, "A physical interpretation of fuzzy ISODATA," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, no. 3, pp. 387-390, 1976.

[13] J. C. Bezdek and R. J. Hathaway, "Convergence and theory for fuzzy c-means clustering: counterexamples and repairs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 873-877, 1987.

[14] K. P. Chan and Y. S. Cheung, "Clustering of clusters," *Pattern Recognition*, vol. 25, no. 2, pp. 211-217, 1992.

[15] N. R. Pal and J. C. Bezdek, "On clustering for the fuzzy c-means model," *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370-379, 1995.

[16] C. Hwang, F. Rhee, and H. Chung, "Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-means," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 107-120, 2007.

[17] J. Yu, "On the fuzziness index of the FCM algorithms," *Chinese Journal of Computers*, vol. 26, no. 8, pp. 965-973, 2003. (In Chinese)

[18] J. Yu, Q. Cheng, and H. Huang, "Analysis of the weighting exponent in the FCM," *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 34, no. 1, pp. 634-639, 2004.

[19] I. Ozkan and I. B. Turksen, "Entropy assessment for type-2 fuzziness," in *Proceedings of the International Conference on Fuzzy Systems*, vol. 2, pp. 1111-1115, 2004.

[20] M. Huang, Z. Xia, H. Wang, Q. Zeng, and Q. Wang, "The range of the value for the fuzzifier of the fuzzy c-means algorithm," Pattern Recognition Letters, vol. 33, no. 16, pp. 2077-2084, 2012.

[21] D. L. Davies and D. W. Bouldin, "A cluster separation measure," IEEE transactions on pattern analysis and machine intelligence, vol. 2, pp. 224-227, 1979.

[22] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, no. 8, pp. 841-847, 1991.

[23] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics, vol. 20, pp. 53-65, 1987.

[24] C. E. Shannon, "A Mathematical Theory of Communication," The Bell System Technical Journal, vol. 27, pp. 379-423, 623-656, July, October, 1948.

[25] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," Information and Control, vol. 20, no. 4, pp. 301–312, 1972.

[26] N. R. Pal and S. K. Pal, "Object-background segmentation using new definitions of entropy," IEE Proceedings E - Computers and Digital Techniques, vol. 136, no. 4, pp. 284–295, 1989.

[27] J. Y. Liang, Z. Z. Shi, D. Y. Li, and M. J. Wierman, "A new method for measuring uncertainty and fuzziness in rough set theory," International Journal of General Systems, vol. 31, no. 4, pp. 331–342, 2002.

[28] Z. Cao and C.-T. Lin, "Inherent fuzzy entropy for the improvement of EEG complexity evaluation," IEEE Transactions on Fuzzy Systems, vol. 26, no. 2, pp. 1032–1035, Apr. 2018.

[29] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 15, no. 2, pp. 266–272, Jun. 2007.

[30] W. Chen, J. Zhuang, W. Yu, and Z. Wang, "Measuring complexity using FuzzyEn, ApEn, and SampEn," Medical Engineering and Physics, vol. 31, no. 1, pp. 61–68, 2009.

[31] W. Pedrycz, "Shadowed Sets: Bridging Fuzzy and Rough Sets," in Rough Fuzzy Hybridization: A New Trend in Decision-Making, Berlin, Germany: Springer-Verlag, 1999.

[32] W. Pedrycz, "Granular computing with shadowed sets," International Journal of Intelligent Systems, vol. 17, no. 2, pp. 173–197, 2002.

[33] W. Pedrycz, "Interpretation of clusters in the framework of shadowed sets," Pattern Recognition Letters, vol. 26, no. 15, pp. 2439–2449, 2005.

[34] W. Pedrycz, "From fuzzy sets to shadowed sets: Interpretation and computing," International Journal of Intelligent Systems, vol. 24, no. 1, pp. 48–61, 2009.