Fairness constraint of Fuzzy C-means Clustering improves clustering fairness

Xia Xu hsuhia0620@gmail.com
Hui Zhang∗ zhanghui@swust.edu.cn
Chunming Yang yangchunming@swust.edu.cn
Xujian Zhao jasonzhaoxj@gmail.com
Bo Li libo.swust@qq.com
Southwest University of Science and Technology, Mianyang, China

Editors: Vineeth N Balasubramanian and Ivor Tsang

Abstract
Fuzzy C-Means (FCM) clustering is a classic clustering algorithm, which is widely used in the real world. Despite the distinct advantages of FCM algorithm, whether the usage of fairness constraint in the FCM could improve clustering fairness remains fully elusive. By introducing a novel fair loss term into the objective function, a Fair Fuzzy C-Means (FFCM) algorithm was proposed in this current study. We proved that the membership value was constrained by distance and fairness in the meantime during the optimization process in the proposed objective function. By studying the Fuzzy C-Means Clustering with fairness constraint problem and proposing a fair fuzzy C-means method, this study provided mechanism understanding in achieving the fairness constraint in Fuzzy C-Means clustering and bridged up the gap of fair fuzzy clustering.

Keywords: fair clustering, algorithm fairness, Fuzzy C-Means.

1. Introduction
Clustering is a classical unsupervised learning task that seeks to group data objects into different clusters, so as to maximize intra-cluster similarity and minimize inter-cluster similarity. Currently, several types of clustering algorithms are available, e.g., the partition-, density-, hierarchical-, spectral- and fuzzy-clustering. Since real-world problems are often fuzzy, fuzzy clustering is becoming more and more popular because of its advantage in dealing with problems using fuzzy mathematics. Instead of grouping objects into a certain cluster, fuzzy clustering partition objects non-uniquely (fuzzy), so that an object can belong to multiple clusters with different membership in the range of 0 to 1. As one of the most representative fuzzy clustering algorithms, Fuzzy C-Means (FCM) clustering Dunn (1973); Bezdek (2013) possesses some outstanding advantages, e.g., 1) simple structure and thus easily being programming; 2) much more approaching real-world problems; 3)Its objective function optimization is supported by nonlinear programming theory. FCM is widely utilized in pattern recognition Chuang et al. (1999), data mining Iyer et al. (2000), classification Hirota and Pedrycz (1999) and image segmentation Rezaee et al. (2000), among others.

∗ Corresponding author

© 2021 X. Xu, H. Zhang, C. Yang, X. Zhao & B. Li.
Recently, the importance of designing fair algorithms has been caught by the machine learning community. Traditional machine learning algorithms failed to take the bias (against certain attributes, this paper refers to these attributes as sensitive attributes, and the value of sensitive attribute is called group.) into account, therefore their output may contain or even augment the bias. To date, considerable amounts of previous efforts have been proposed by the machine learning community to address the fair clustering task Chierichetti et al. (2017); Rösner and Schmidt (2018); Kleindessner et al. (2019b); Chen et al. (2019); Bera et al. (2019); Sam Abraham et al. (2020); Micha and Shah (2020). However, these works are mostly focused on center- or spectral- based methods. Despite the distinct advantages of the FCM algorithm, whether the usage of fairness constraint in the FCM could improve clustering fairness remains fully elusive. In this study, we questioned how the utilized of fairness constraint to enhance the fairness of Fuzzy C-Means Clustering.

To answer the above question, a fair Fuzzy C-Means algorithm was proposed in this paper. Considering that FCM was a fuzzy clustering based on the objective function, we innovatively introduced a fair loss term into the objective function, and optimized the objective function to obtain a fair result. By setting the weight of the fair loss term, our method achieved a trade-off between clustering quality and fairness. The fair loss function was defined as the square of the difference between the ratio of groups in clusters and the ratio of groups in the original dataset. To the best of our knowledge, this was the first result that introduced fair loss function into Fuzzy C-Means clustering.

2. Related Work

In recent years, the research community has done a lot of work in providing fairness guarantee for machine learning algorithms or studying fair variants of clustering algorithms. Plenty of relevant formulations of fairness have been proposed for supervised learning and specifically for classification tasks Dwork et al. (2012); Hardt et al. (2016); Kleinberg et al. (2017); Zafar et al. (2017). In this study, we emphasized the importance of fair clustering. Two work lines on fair clustering were available, i.e., the follow-up work proposed by Chierichetti et al. (2017) and some independent works.

The most recognized notion in the field of fair clustering was proposed by Chierichetti et al. (2017). They modeled fairness based on disparate impact doctrine Feldman et al. (2015), which posited that any "sensitive attribute" must have approximately equal representation in the decision taken (through algorithms). To achieve fair clustering, they introduced the fairlet decomposition, which partitioned data objects into small and balance subsets (e.g., the fairlets). These subsets were clustered to get a fair result. Due to their method only worked in the scenery of two groups and time-consuming, the Chierichetti method was not so successful but he actually opened up ideas for further research, especially since he first defined a normal notion for fair clustering. Later, several follow-up works extended this idea. Schmidt et al. (2020) proposed a fast fairlets decomposition method. They defined a notion called “coreset”, which was a representative subset of the original dataset. By solving fair clustering problem on the coreset, an approximate solution for the original dataset was provided. Backurs et al. (2019) suggested embedding the input data into Hierarchically well-Separated Tree to accelerate fairlet decomposition. Both methods could speed up the fairlet decomposition, however, the fairness notion used in both methods
was proposed by Chierichetti et al. (2017), and hence only worked for two groups. Rösner and Schmidt (2018) considered a multiple groups variant of fair k-center clustering and developed a k-center constant factor approximation approach. Bercea et al. (2019) improved the result of Rösner and Schmidt (2018) and gave a bicriteria constant factor approximation algorithm for several classical clustering objectives. Kleindessner et al. (2019b) introduced fairness to spectral clustering by viewing the fairness notion as a constrained variant of spectral clustering. This method only worked for single sensitive attribute. Bera et al. (2019) extended the fairness notion as the upper and lower bound of the representation of a group in a cluster. In addition, they gave a general fair clustering framework. The framework consisted of two steps. Firstly, vanilla clustering was used to generate cluster centers, and then the linear programming method was used to fairly assign data objects to the cluster centers. Nevertheless, this framework was only suitable for K-center and K-median clustering. Ahmadian et al. (2019) explored the problem of preventing excessive representation of groups in each cluster, and gave an algorithm based on linear programming. Ahmadian et al. (2020) recently studied two variants of fair clustering, the one was minimum divergence clustering, and the other one was upper and lower bound clustering. Their algorithms work for multiple groups. Davidson and Ravi (2020) showed that for any clustering with two groups, linear programming could be used to calculate the most similar fair clustering. Quy et al. (2021) recently studied the fairness problem in the education domain and gave two fair method, namely hierarchical clustering and partitioning-based clustering.

Chen et al. (2019) considered the proportional centroid clustering problem and outlined an independent fairness notion. For clustering n points with k centers, any n/k points were entitled to form their own cluster if there is another center that was closer in distance for all n/k points. Kleindessner et al. (2019a) advised a simple k-center clustering algorithm with fairness constraint. In this method, a cluster was regarded as a summary of the original dataset, and fair summary was generated for each cluster. Jung et al. (2020) proposed the notion of individual fairness of clustering, which required each object somewhat to close a center, "somewhat" depended on the object’s k nearest neighbors. Recently, Ghadiri et al. (2021) and Abbasi et al. (2021) independently proposed the social fair clustering problem. Makarychev and Vakilian (2021) improved and generalized the $O(\ell)$-approximation algorithms of the social fair clustering problem in Ghadiri et al. (2021) and Abbasi et al. (2021). The works most relevant to this study were Ziko et al. (2019) and Sam Abraham et al. (2020). Ziko et al. (2019) incorporated fairness constraints into the clustering steps by adding a fairness loss term into the objective function. They defined the fair loss as the KL divergence between the probability distribution of sensitive attributes in clusters and the probability distribution in the dataset. However, this method was designed only for one sensitive attribute. Sam Abraham et al. (2020) also used a similar idea, but works for multiple sensitive attributes. Unfortunately, they did not consider introducing fairness into fuzzy clustering. Fuzzy clustering was widely used in the real world, it was imperative to ensure that these algorithms were fair. In this current study, the proposed algorithm provided fair guarantee for fuzzy c-means clustering. In addition, fair loss was defined as the square of the difference between the ratio of groups in clusters and the ratio of groups in the original dataset.
3. Preliminaries

In this section, vanilla FCM Bezdek (2013) is formally described and the fuzzy C-means with fairness constraints problem is defined in order to introduce terminology and set the ground for our works.

Let $P$ be a collection with $n$ objects embedded in metric space $(X, d)$, where $P := \{p_1, p_2, \ldots, p_n\}$ and $d : P^2 \to R \geq 0$. Let $k$ be the number of clusters, $u_{ij}$ is the membership value of object $j$ to cluster center $c_i$ and $u_{ij}$ is subject to $u_{j1} + u_{j2} + \cdots + u_{jk} = 1$, $u_{ij} \in [0, 1]$.

The task of vanilla FCM is to find a fuzzy partition matrix $U := [u_{ij}]$ and a set of cluster centers $C := \{c_1, c_2, \ldots, c_k\}$ to minimize the objective function $J_{fcm} := \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^m \|p_j - c_i\|^2$.

In fair clustering task, one is additionally, there are two sets of attributes $C_a$ and $S_a$ defined over datasets. $C_a$ denote the attributes that are relevant to the clustering task interest, such as the patient’s symptoms and living environment in the case of disease clustering. $S_a$ denote sensitive attributes, which may be expected to maintain fairness in the generated clusters, such as gender, race, religion, nationality, etc. What’s more, the values of sensitive attribute are called groups (as mentioned in the section 1). Assuming that gender is a sensitive attribute, it may have two groups, named male and female. The fairness notion utilized in this paper is proportional fairness, which maintains the same proportion of sensitive attribute groups in clusters as they are in the original dataset. In other words, there are two groups of a sensitive attribute (Considering gender) with a ratio of 7:3 in the original dataset, and their ratios are expected to be 7:3 (ideally) in each generated clusters. Fuzzy c-means with fairness constraints problem is formally defined below.

**Definition 1** (Fuzzy C-Means clustering with fairness constraint problem.) Given $l$ groups $P_1, P_2, \cdots, P_l$ as the values of the sensitive attribute, and $P_1 \cup P_2 \cup \cdots \cup P_l = P$. The fair fuzzy clustering problem can be described as finding a partition of $P$ so as to minimize object function $J_{fcm} := \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^m \|p_j - c_i\|^2$ and make the ratio of each group in clusters as close as possible to its ratio in original dataset.

4. Proposed Approach

We propose a Fair Fuzzy C-Means (FFCM) clustering method to attack the Fuzzy C-Means clustering with fairness constraint problem. In FFCM, a novel fair loss function is constructed to quantify the fair loss in clusters and an optimization method is designed for the objective function to achieve fair clustering. Firstly, the construction of the fair loss function is detailed in the next part of this section. After that, the optimization method is explained, and the validity of the proposed objective function formula is proved theoretically. In the last part of this section, the complexity of the proposed algorithm is analyzed.

4.1. Fair Loss Term Construction

For an ideal fair clustering, the ratio of groups in the generated clusters is expected to be the same as in the original dataset. With the intent of generating clusters that are as fair
as possible, a natural definition of fair loss function is the difference between the ratio of
groups in clusters and the ratio of groups in the original dataset. The greater the ratio
difference, the greater the loss function score will be. The mathematical expression for the
fair loss of a group in a cluster is

\[
\text{bias}(C_{is}) = \left( \frac{|C_i \cap P_s|}{|C_i|} - \frac{|P_s|}{|P|} \right)^2, \quad s \in [l]
\]

where \( \text{bias}(C_{is}) \) is the fair loss of group \( P_s \) in cluster \( C_i \). Naturally, the entire fair loss of
cluster \( C_i \) is the sum of all the group’s loss score in cluster \( C_i \). The entire fair loss can be
written as

\[
\text{bias}(C_i) = \sum_{s=1}^{l} \text{bias}(C_{is})
\]

It should be noted that different sensitive attributes may have different numbers of
groups (considering the sensitive attributes of gender and age. Gender may have two groups,
male and female, while age may have multiple groups, such as infants, children, youth,
middle-aged and old). It can be observed from Eq. 2 that sensitive attributes with a
large number of groups may produce a large fair loss score. This means that the sensitive
attributes with more groups might dominate the fair loss term. In order to make each
sensitive attribute have the same contribution to the fair loss term, the fair loss term is
normalized in Eq. 3 by the number of groups.

\[
N\text{bias}(C_i) = \sum_{\omega=1}^{A} \sum_{s=1}^{L_{\omega}} \left( \frac{|C_i \cap P_s|}{|C_i|} - \frac{|P_s|}{|P|} \right)^2 \frac{1}{l_{\omega}}
\]

Where \( A \) is the number of sensitive attributes, and \( l_{\omega} \) is the number of groups in the \( \omega \)
sensitive attribute.

Since fair loss must be related with membership value, the larger the distance from the
object to the cluster center plus the fair loss score, the smaller the membership value of the
object is expected to be. Thus, the fair loss is decomposed into each membership value.
For an object \( p_j \), it is assigned to clusters \( C_i \), \( i \) from 1 to \( k \), and the assignment of other
objects is retained. Let \( \text{bias}_{p_j \rightarrow C_1}, \text{bias}_{p_j \rightarrow C_2}, \ldots, \text{bias}_{p_j \rightarrow C_k} \) be the normalized fair loss score, which is
calculated by Eq. 3. These fair loss scores are introduced into the objective function so that
the fair loss and distance jointly determine the membership value. The overall objective
function is shown in Eq. 4

\[
J_{ffcm} = \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^m \left( \|p_j - c_i\|^2 + \eta \text{bias}_{p_j \rightarrow C_i} \right),
\]

\[
s. t. \sum_{i=1}^{k} u_{ij} = 1
\]

where \( \eta \) is a hyperparameter, which denotes the weight of fair loss term.
4.2. Objective Function Optimization and FFCM Algorithm

After the fair loss is introduced into the objective function of FFCM, the task of Fuzzy C-Means with fairness constraint is reduced to identifying a partition that minimizes \( J_{ffcm} \) as much as possible. As in the vanilla fuzzy c-means, the Lagrange Multiplier Method is used to find the minimum value of the objective function. The Lagrange function of FFCM is constructed as

\[
L = \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^m \left( \|p_j - c_i\|^2 + \eta \frac{\text{bias}_{p_j \rightarrow C_i}}{\text{bias}_{p_j \rightarrow C_i}} \right) - \sum_{j=1}^{n} \lambda_j \left( \sum_{i=1}^{k} u_{ij} - 1 \right)
\]  

(5)

where the first part in Eq. 5 is the objective function of FFCM, and the second part is the inherent constraint of membership value (i.e., the sum of the membership values of an object to all clusters is equal to 1).

By calculating the partial derivative of the Lagrange function with respect to \( u_{ij} \), the expression of membership value \( u_{ij} \) is obtained as

\[
u_{ij} = \frac{1}{\left( \|p_j - c_i\|^2 + \eta \frac{\text{bias}_{p_j \rightarrow C_i}}{\text{bias}_{p_j \rightarrow C_i}} \right)^{m-1}}
\]

\[
\sum_{\tau=1}^{k} \left( \|p_j - c_\tau\|^2 + \eta \frac{\text{bias}_{p_j \rightarrow C_\tau}}{\text{bias}_{p_j \rightarrow C_\tau}} \right)^{m-1}
\]

(6)

**Theorem 2** The membership value is constrained by distance and fair loss term at the same time during the optimization process.

**Proof** For any object \( p_j \) and cluster center \( C_i \), when \( p_j \) is assigned to \( C_i \), the fair loss of this assignment is \( \text{bias}_{p_j \rightarrow C_i} = \sum_{\omega=1}^{A} \sum_{s=1}^{l_\omega} \left( |C_i \cap P_s|/|C_i| - |P_s|/|P| \right)/|l_\omega| \). Let \( d_{ij} \) be the distance from \( p_j \) to the cluster center \( c_i \), Eq. 6 can be reduced to

\[
u_{ij} = \frac{1}{\left( d_{ij} + \eta \frac{\text{bias}_{p_j \rightarrow C_i}}{\text{bias}_{p_j \rightarrow C_i}} \right)^{m-1}}
\]

\[
\sum_{\tau=1}^{k} \left( d_{ij} + \eta \frac{\text{bias}_{p_j \rightarrow C_\tau}}{\text{bias}_{p_j \rightarrow C_\tau}} \right)^{m-1}
\]

(7)

We guarantee that \( \sum_{\tau=1}^{k} \left( d_{ij} + \eta \frac{\text{bias}_{p_j \rightarrow C_\tau}}{\text{bias}_{p_j \rightarrow C_\tau}} \right) \) is a constant. Observing Eq. 7, it is easy to find that when the fair loss \( \text{bias}_{p_j \rightarrow C_i} \) is large, \( 1/\left( d_{ij} + \text{bias}_{p_j \rightarrow C_i} \right) \) will be small. Therefore, the membership value \( u_{ij} \) will also be small. Conversely, when \( \text{bias}_{p_j \rightarrow C_i} \) is small, \( 1/\left( d_{ij} + \text{bias}_{p_j \rightarrow C_i} \right) \) will be large, and correspondingly \( u_{ij} \) will also be large. ■
Algorithm 1 Fair Fuzzy C-means Clustering.

**Input:** Collection $P$ with $n$ objects, the number of clusters $k$, the end condition $\varepsilon$, the maximum number of iterations $T$ and the weight of fair loss $\eta$.

**Output:** The labels of objects.

1: Initialize the membership value $u_{ij}$;
2: while $\left| J_{ffcm}^{(t)} - J_{ffcm}^{(t-1)} \right| > \varepsilon$ and $t < T$ do
3: Update the cluster centers $c_i$ by Eq. 9;
4: for $p_j \in P, j = 1$ to $n$ do
5: the object $p_j$ is assigned to clusters $C_i$, $i$ from 1 to $k$,
6: Calculate the fair loss term by Eq. 3;
7: Update the membership value $u_{ij}$ by Eq. 6;
8: end for
9: Calculate the value of the objective function $J_{ffcm}$ by Eq. 4;
10: end while
11: return The labels of each object.

**Lemma 3** For any object $p_j$, the $\sum_{\tau=1}^{k} \left( d_{\tau j} + \eta \text{bias}_{p_j \rightarrow C_{\tau}} \right)$ is a constant.

**Proof** If $p_j$ is assigned to clusters $C_1, C_2, \ldots, C_k$, the fair loss of these assignments are $\text{bias}_{p_j \rightarrow C_1}, \text{bias}_{p_j \rightarrow C_2}, \ldots, \text{bias}_{p_j \rightarrow C_k}$. Then we have:

$$\sum_{\tau=1}^{k} \left( d_{\tau j} + \eta \text{bias}_{p_j \rightarrow C_{\tau}} \right) = (d_{1j} + d_{2j} + \cdots + d_{kj}) + \eta \left( \text{bias}_{p_j \rightarrow C_1} + \text{bias}_{p_j \rightarrow C_2} + \cdots + \text{bias}_{p_j \rightarrow C_k} \right)$$

(8)

Because $d_{\tau j}$ and $\text{bias}_{p_j \rightarrow C_{\tau}}$ are all constants, their sum $\sum_{\tau=1}^{k} \left( d_{\tau j} + \eta \text{bias}_{p_j \rightarrow C_{\tau}} \right)$ is also a constant.

Similar to the membership value $u_{ij}$, by calculating the partial derivative of the Lagrange function with respect to $c_i$, the expression of the cluster center $c_i$ is obtained as

$$c_i = \frac{\sum_{j=1}^{n} u_{ij} p_j}{\sum_{j=1}^{n} u_{ij}}$$

(9)

The FFCM algorithm is summarized in Algorithm 1. Algorithm 1 resembles the working of vanilla FCM except a fair loss is considered at each step that allows it to be fairer than vanilla FCM. Specifically, the membership value is initialized first, and proceeds iteratively. In each iteration, three steps are performed. First, update the cluster center through the membership value. Then, we traverse each object $p$ in round-robin fashion, the object is assigned to each cluster respectively and the fair loss score of the present assignment
is calculated. Finally, the membership value is updated based on the fair loss score and distance. The updating process will stop until the objective function converges or the maximum iteration threshold is reached.

4.3. Complexity Analysis

The complexity of FFCM is determined by fair loss calculation, membership value update and cluster center update. First, considering the complexity of the fair loss calculation. For a data object \( p \), when it is assigned to each cluster, the ratio of groups needs to be re-stated. In each cluster, the fair loss needs to be calculated for \( A \) sensitive attributes and \( l \) groups of each sensitive attribute. The fair loss calculation method is shown in Eq. 1, which can be regarded as a simple calculation completed in a constant time. Therefore, the complexity of fair loss calculation is \( O(nk^2 |A| |l| + k |A| |l|) \). Next, we consider the complexity of membership value update and cluster centers update. According to Eq. 5, the complexity of membership value update is \( O(nk^2) \). And according to Eq. 6, the complexity of cluster centers update is \( O(nk) \). In summary, the complexity of \( t \) iterations is \( O(tn^3k^4 |A| |l| + nk^3 |A| |l| + nk^2). \)

5. Experiments

In this section, we performed empirical evaluations of the proposed algorithm. Firstly, we outlined the datasets and settings in the experiment. Then, we illustrated the measurements. Finally, we reported the experimental results and the analysis of experimental results.

5.1. Datasets and Settings

We considered six real-world datasets, which were popular in the fair clustering task. (1) Diabetic Chierichetti et al. (2017) recorded information related to patients with diabetes. (2) Census1990 Meek et al. (2002) consisted of the 1990 U.S. census records. (3) Credit-card Yeh and hui Lien (2009) contained information about the card holders from a certain credit card in Taiwan. (4) Bank Moro et al. (2014) contained instances related to the direct telemarketing activities of Portuguese banking institutions. (5) Adult Zhou and Chen (2002) was also known as census, which contained examples of the 1994 U.S. Census. (6) Athlete Böhm et al. (2020) contained bio data on Olympic athletes and medal results from Athens 1896 to Rio 2016. We subsampled all datasets to 1000 records, except Creditcard. Creditcard was subsampled to 600 records. For each of these datasets, we chose numerical attributes to represent points in Euclidean space. In addition, we also set sensitive attribute for each dataset and created groups based on their values. More information about the dataset settings was shown in Table 1.

In the proposed method, the weight of the fair loss term \( \eta \) needed to be set. Based on empirical observation, we set \( \eta \) to \( 10^2 \) for Diabetic and Census1990, \( 10^3 \) for Athlete, \( 10^5 \) for Bank, \( 10^7 \) for Creditcard and Adult. We analyzed the sensitivity of \( \eta \) in section 5.4. The maximum number of iterations \( T \) was set to 10. Our codes are available on GitHub for public use.\(^1\)

\(^1\) https://github.com/author’s-name/author’s-name-FuzzyC-MeansWithFairnessConstraint
Table 1: Clustering attributes and sensitive attributes.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Clustering attributes</th>
<th>Sensitive attributes</th>
<th>Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetic</td>
<td>age, time-in-hosp,</td>
<td>gender</td>
<td>male, female, unknown</td>
</tr>
<tr>
<td></td>
<td>num-medication, num-outpatient, num-emergency, num-inpatient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creditcard</td>
<td>age, bill-amt 1—6,</td>
<td>education</td>
<td>7 groups</td>
</tr>
<tr>
<td></td>
<td>limit-bal, pay-amt 1—6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bank</td>
<td>age, balance, duration</td>
<td>marital</td>
<td>married, single, divorced</td>
</tr>
<tr>
<td>Adult</td>
<td>age, education-num,</td>
<td>race</td>
<td>asian-pac-isl, Amer-ind, white, black, other</td>
</tr>
<tr>
<td></td>
<td>final-weight, capital-gain, hours-per-week</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Census1990</td>
<td>dAncstry1, dAncstry2, iAvail, iCitizen, iClass, dDepart, iFertil, iDisabl1, iDisabl2, iEnglish, iSex, iFeb55, dHispanic, dHour89</td>
<td>iSex</td>
<td>male, female</td>
</tr>
<tr>
<td>Athlete</td>
<td>Age, Height, Weight, Year</td>
<td>Season</td>
<td>summer, winter</td>
</tr>
</tbody>
</table>

5.2. Measurements

In fair clustering, two sets of metrics were mainly concerned. One was clustering quality, which was related to clustering interest (i.e., $C_a$, as described in Section 3). The other was fairness, which was related to fair interest ($S_a$). In this subsection, we described these metrics.

Clustering quality measured the rationality of clustering, it could include: (1) Clustering Cost ($\text{Cost}$), it measured the deviation of objects from cluster centers. The smaller the clustering cost, the better the clustering result. One of the clustering cost presentation was objective function, and the objective function of FCM Bezdek (2013) was

$$J_{fcm} := \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^m \| p_j - c_i \|^2$$

(10)

(2) Silhouette Coefficient ($\text{SC}$) Rousseeuw (1987), it described the cohesion and separation of clusters. It lies in the range of $[-1, 1]$, the closer the value to 1, the better the clustering result and vice versa.

Fairness measured the fairness of attributes in the generated output, it could include: (1) balance Bera et al. (2019), which described the lowest level fairness of groups in a cluster. It was defined as: Let $f(P_s) = |P_s|/|P|$ be the ratio of group $s$ in the entire dataset, and let $f(C_{is}) = |P_s \cap C_i|/|C_i|$ be the ratio of group $s$ in cluster $i$. The $\text{fairness}(C_i) = \min(f(C_{is})/f(P_s), f(P_s)/f(C_{is})), \ i \in [k], s \in [l]$ was the balance in cluster $i$. (2) Euclidean distance of distribution vectors ($\text{Ed}$) Sam Abraham et al. (2020), which
measured the unfairness of clustering. Let $S$ be a sensitive attribute, which could take on $l$ groups. The distribution of these groups in the dataset $P$ produced a $l$-length distribution vector $P_s$. Similarly, the distribution of these groups in each cluster yielded a $l$-length distribution vector $C_s$. By calculating the Euclidean distance between the representations $P_s$ and $C_s$ to get the quantized intra-cluster unfairness. (3) Wasserstein distance of distribution vectors ($W_d$) Wang and Davidson (2019), The definition of $W_d$ metric was similar to $E_d$, except that Euclidean distance was replaced by Wasserstein distance.

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>Datasets</th>
<th>$E_d$</th>
<th>$W_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FFCM</td>
<td>FCM</td>
</tr>
<tr>
<td>$k = 4$</td>
<td>Diabetic</td>
<td>0.4283</td>
<td>0.7745</td>
</tr>
<tr>
<td></td>
<td>Census1990</td>
<td>0.1452</td>
<td>0.3980</td>
</tr>
<tr>
<td></td>
<td>Creditcard</td>
<td>0.3105</td>
<td>0.3789</td>
</tr>
<tr>
<td></td>
<td>Bank</td>
<td>0.2623</td>
<td>0.6007</td>
</tr>
<tr>
<td></td>
<td>Adult</td>
<td>0.2345</td>
<td>0.3270</td>
</tr>
<tr>
<td></td>
<td>Athlete</td>
<td>0.3258</td>
<td>0.5606</td>
</tr>
<tr>
<td>$k = 6$</td>
<td>Diabetic</td>
<td>0.7565</td>
<td>1.5480</td>
</tr>
<tr>
<td></td>
<td>Census1990</td>
<td>0.2088</td>
<td>0.6472</td>
</tr>
<tr>
<td></td>
<td>Creditcard</td>
<td>0.4000</td>
<td>0.5026</td>
</tr>
<tr>
<td></td>
<td>Bank</td>
<td>0.3599</td>
<td>0.5638</td>
</tr>
<tr>
<td></td>
<td>Adult</td>
<td>0.3874</td>
<td>0.5775</td>
</tr>
<tr>
<td></td>
<td>Athlete</td>
<td>0.4575</td>
<td>0.8619</td>
</tr>
<tr>
<td>$k = 8$</td>
<td>Diabetic</td>
<td>0.6330</td>
<td>1.6767</td>
</tr>
<tr>
<td></td>
<td>Census1990</td>
<td>0.5643</td>
<td>1.8573</td>
</tr>
<tr>
<td></td>
<td>Creditcard</td>
<td>0.6782</td>
<td>0.9104</td>
</tr>
<tr>
<td></td>
<td>Bank</td>
<td>0.3868</td>
<td>0.8890</td>
</tr>
<tr>
<td></td>
<td>Adult</td>
<td>0.5010</td>
<td>0.7830</td>
</tr>
<tr>
<td></td>
<td>Athlete</td>
<td>0.5832</td>
<td>1.1865</td>
</tr>
</tbody>
</table>

5.3. Clustering Results Comparison with FCM

Table 2 displayed the fairness metrics of FFCM and FCM, and the best of all results were highlighted in boldface. The $impr$ column represented the fairness improvement percentage by FFCM. As could be seen from Table 2, the performance of FFCM in fairness metrics significantly surpassed FCM. Specifically, FFCM performed the best on the Census1990, and the percentage of fairness improvement was about 65%, while it performed slightly poorly on the Creditcard, and the percentage of fairness improvement was still more than 20%. In addition, it was worth noting that when $k = 6$, the overall performance of FFCM was better than when $k = 4$. Whereas $k = 8$, the performance of FFCM was stronger. This indicated that FFCM benefitted from higher flexibility (with higher $k$) in the process of reducing fairness loss. Figure 1 presented the balance of FFCM and FCM in each cluster, and the balance of different methods was marked by different colors. Since the goal of the FFCM objective function was to minimize the overall fairness loss, and no specific constraint to ensure good performance on the lowest level fairness (i.e., $balance$). Therefore, in order to obtain a lower fairness loss, FFCM might sacrifice fairness in one or several clusters (the cluster 1 of Adult in Figure 1). Although the balance metric of FFCM implied that such trends were not widely prevalent, it was a direction to meliorate FFCM. Table 3 presented the results of clustering quality, and the $decr$ column indicates the clustering
quality degradation percentage by FFCM. Due to FFCM needed to be held accountable for fairness, it was expected to perform poorly on clustering quality metric. We focused on the difference between the improvement of fairness and the degradation of clustering quality by FFCM. Table 4 displayed these results, the avg.impr row was the average improvement
percentage of Wd and Ed, and the avg.decr row was the average degradation percentage of Cost and SC. Their values were the average values when the number of clusters \( k = 4, 6, 8 \). It might be seen that compared to the degradation of clustering quality, FFCM had higher fairness margins. In general, the above results illustrated that FFCM provided fairer clustering. Despite FFCM reduced clustering quality, it brought fairer margin than the degradation of clustering quality.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Diabetic</th>
<th>Census1990</th>
<th>Creditcard</th>
<th>Bank</th>
<th>Adult</th>
<th>Athlete</th>
<th>avg.whole</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg.impr(%)</td>
<td>51.1474</td>
<td>65.8136</td>
<td>20.0000</td>
<td>44.6253</td>
<td>30.1722</td>
<td>46.5529</td>
<td>42.3511</td>
</tr>
<tr>
<td>avg.decr(%)</td>
<td>46.4958</td>
<td>17.1160</td>
<td>13.1772</td>
<td>43.5378</td>
<td>8.3511</td>
<td>22.8236</td>
<td>25.7356</td>
</tr>
<tr>
<td>margin(%)</td>
<td>4.6512</td>
<td>48.6975</td>
<td>6.8198</td>
<td>1.1087</td>
<td>1.8211</td>
<td>23.7293</td>
<td>16.616</td>
</tr>
</tbody>
</table>

Table 4: Comparison of fairness improvement and quality degradation by FFCM

5.4. Sensitivity Analysis to \( \eta \)

FFCM attempted to achieve a trade-off between clustering quality and fairness by changing the value of the only hyper-parameter \( \eta \) (i.e., the weight of fair loss function). The FFCM formula was expected that the greater the \( \eta \), the smaller the fairness loss of clustering, and the better the performance on the fairness metrics. We observed such desired trends across all experimental datasets. Figures 2 and Figures 3 presented the changes of fairness metrics and clustering quality metrics on Diabetic, Census1990 and Athlete (Considering that the setting of \( \eta \) have the similar values) when \( \eta \) varied from 100 to 1000. For the

![Figure 2: The change of fairness metrics with different \( \eta \)](image)

![Figure 3: The change of clustering quality metrics with different \( \eta \)](image)
metrics that widely varied in range, both sides of the y-axis were used to plot them, and the axis used for each metric was indicated in the figure. The Bias in Figure 2 was the whole fair loss of clustering. The fact could be obtained from Figures 2 and Figures 3 that with the increase of $\eta$, the clustering generated by FFCM tends to be fair (bias and Ed metric decrease gradually) and the clustering quality degrades accordingly (SC metric decreased and Cost metric increased). Although their change was not stable, the direction of change was in line with our expectations.

6. Conclusions And Future Directions

Although the fair clustering algorithm has attracted considerable attention, the fairness of fuzzy clustering has yet remained elusive. In this paper, we studied Fuzzy C-Means clustering with fairness constraint problem and proposed a fair fuzzy C-means method to bridge the gap of fair fuzzy clustering. By introducing a novel fairness loss term in the objective function, the membership value was constrained by fairness and distance at the same time during the optimization process, we gave theoretical proof for this constraint. We evaluated the performance of the proposed algorithm on real-world datasets, the empirical evaluation illustrated that FFCM significantly improved clustering fairness. We depicted two directions to enhance the performance of FFCM. Firstly, FFCM did not guarantee the lowest level of fairness, so it might sacrifice fairness in several clusters to minimize the fairness loss function. We considered adding a constraint in the loss function to prevent this phenomenon. Secondly, due to FFCM was a fair variant based on FCM, it was sensitive to the initial cluster centers and easily fell into the local optimal value. Future efforts should replace the vanilla FCM algorithm with a robust variant to improve the performance of FFCM.

References


Xu Zhang Yang Zhao Li


