UniFair: A Unified fair clustering approach based on separation and compactness

Antonia Karra, Georgios Vardakas, Evaggelia Pitoura, Aristidis Likas
{a.karra, g.vardakas, pitoura, arly} @uoi.gr
Department of Computer Science & Engineering, University of Ioannina, Greece

Motivation

- Clustering is widely used in ML and data mining.
- Fairness concerns arise when sensitive groups are treated differently.
- Existing notions of fairness (Balance, Social, Individual, Deep Fair).
- There is a need for fairness both in how decision boundaries are positioned and in how clustering distortion is distributed across groups.

Separation Fairness

Dataset: $X = \{x_1, \dots x_N\} \subset \mathbb{R}^d$
Demographic subgroups: A, B
Clusters centroids: $M = \{\mu_1, \dots, \mu_k\}$

Key Idea: Separation Fairness measures how far each group lies from the **nearest decision boundary** between its two closest centroids.
Setup:

For each point x, let $j_1(x)$and $j_2(x)$ be its two nearest centroids.

- **Midpoint between them:** $m(x) = \frac{1}{2}\left(\mu_{j_1(x)} + \mu_{j_2(x)}\right)$
- **Direction of the boundary:** $\hat{u}(x) = \frac{\mu_{j_2(x)} - \mu_{j_1(x)}}{\|\mu_{j_2(x)} - \mu_{j_1(x)}\|}$
- **Distance to Decision Boundary (Counterfactual Distance):**
- $\text{dist}^2(x) = \left(\left(x - m(x)\right)^T \hat{u}(x)\right)^2$
- **Group Counterfactual Distance Fairness**: measures how far each group is, on average, from the nearest decision boundary.
- For each subgroup $g \in \{A, B\}$:

$$\text{cfd}_g = \frac{1}{X_g} \sum_{x \in X_g} ((x - m(x)^T \hat{u}(x))^2$$

- Encouraging fairness by maximizing:

$$\text{CF}(M) = \min(\text{cfd}_A, \text{cfd}_B)$$

- **Separation Fairness Objective:**

$$E(M) = \frac{1}{N} \sum_{j=1}^{N} \|x_j - \mu_j\|^2 - \lambda \cdot \min(\text{cfd}_A, cf\,\text{d}_B)$$

- **Update (Fairness):**

$$\mu_j \leftarrow \mu_j - \eta\left(\nabla_{\mu_j} L_{kmeans} + \lambda \nabla_{\mu_j} \text{CF}\right)$$

Social Fairness

- For each cluster C_j , and for each group $g \in \{A, B\}$ compute how each group is represented locally:

$$\mu_j^g = \frac{1}{|C_j \cap X_g|} \sum_{x \in C_j \cap X_g} x\,,$$

- Measures **within-cluster distortion** per group

$$L_g(M) = \frac{1}{X_g} \sum_{j=1}^{k} \sum_{x \in C_j \cap X_g} \|x_i - \mu_j\|^2 = \frac{1}{|X_g|} \sum_{j=1}^{k} |C_j \cap X_g| \left\|\mu_i - \mu_j^g\right\|^2$$

- **Social Fairness Objective:**

$$E(M) = \frac{1}{N} \sum_{i=1}^{N} \|x_i - \mu_i\|^2 + \lambda \cdot \max(L_A, L_B)$$

- **Update (Fairness)**

$$\mu_j \leftarrow \mu_j - \eta\left(\nabla_{\mu_j} L_{kmeans} + \lambda \nabla_{\mu_j} L_G\right)$$

Unifair Fairness

Combine both fairness notions:

$$J(M) = \frac{1}{N} \sum_{j=1}^{N} \|x_j - \mu_j\|^2 + \lambda_{soc} \cdot \max(L_A, L_B) - \lambda_{sep} \cdot \min(\text{cfd}_A, cf\,\text{d}_B)$$

Weights (λ parameters)
λ_{sep}:
- Controls **separation fairness**
- Encourages equal **distance from decision boundaries**

λ_{soc}:
Controls **social fairness**
Reduces the distortion gap between $L_A$and L_B

- **Update (Fairness)**

$$\mu_j \leftarrow \mu_j - \eta\left(\nabla_{\mu_j} L_{kmeans} + \lambda_{soc} \nabla_{\mu_j} L_G - \lambda_{sep} \nabla_{\mu_j} \text{CF}\right)$$

Deep Fair Clustering

- Use an **autoencoder** to learn a latent representation $z = f_\theta(x)$
- Clustering is performed in **latent space**, not the original feature space
- Objective combines:
  - **Reconstruction loss** (keeps latent space meaningful)
  - **Cluster compactness loss** (pulls points to centers)
  - **Fairness losses** (applied in latent space)
- General deep clustering objective:

$$L = \alpha L_{rec} + \beta L_{cmp} + \lambda L_{fair}$$

- **Deep Separation Fairness**
  - Compute **CFD** in **latent space**:
  - **Final loss**:

$$L = \alpha L_{rec} + \beta L_{cmp} + \lambda_{sep} L_{sep}$$

- **Deep Social Fairness**
  - **Group distortion** in the **latent space**
  - Final loss:

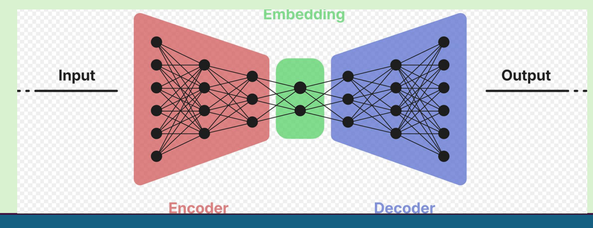
$$L = \alpha L_{rec} + \beta L_{cmp} + \lambda_{soc} L_{soc}$$

- **Deep Separation-Social Fairness**
  - Combine *both* fairness notions

$$L_{fair} = \lambda_{soc} \cdot \max(L_A, L_B) - \lambda_{sep} \cdot \min(\text{cfd}_A, cf\,\text{d}_B)$$

  - Final objective:

$$L = \alpha L_{rec} + \beta L_{cmp} + L_{fair}$$



Conclusions

- **UniFair** provides a unified framework that integrates both **separation fairness** and **social fairness** within clustering.
- It ensures that **decision boundaries** and **cluster structures** do not disproportionately disadvantage any demographic group.
- UniFair works in both the **original feature space** and a **latent space** learned through deep autoencoders.
- Overall, UniFair offers a **flexible, effective, and principled** approach for promoting fairness in clustering-based decision systems.
- For more information:
  https://www.cse.uoi.gr/~fxc

Counterfactuals for
Clustering and Deep Clustering:
Explainability, Fairness and Quality
Project Number: 15940
Hellenic Foundation for Research & Innovation (H.F.R.I)
"Basic Research Funding (Horizontal Support for all Sciences)"
National Recovery and Resilience