Supplementary Material for Deep Fair Clustering via Maximizing and Minimizing Mutual Information: Theory, Algorithm and Metric

Pengxin Zeng*, Yunfan Li*, Peng Hu, Dezhong Peng, Jiancheng Lv, Xi Peng†
College of Computer Science, Sichuan University, China
{zengpengxin.gm, yunfanli.gm, penghu.ml, pengx.gm}@gmail.com;
{pengdz, lvjiancheng}@scu.edu.cn

1. Datasets Used for Evaluation

To demonstrate the versatility of our method, we use six datasets confounded with various types of sensitive attributes for evaluations. For clarity, we summarize the statical characteristics of the datasets in Tab. 1. Among them, the first four datasets, *i.e.*, MNIST-USPS [1], Reverse MNIST, Office-31 [8], MTFL [11] are with visual modality and the last two (HAR [1] and Mouse Atlas [4, 5]) are signal-vector datasets. Notably, some existing works [2, 3] could only handle the dataset with bi-sensitive-attribute (group number is two), whereas our method is generalizable to the case of arbitrary group numbers, *e.g.*, HAR with 30 groups.

Table 1. Datasets used for evaluations.

Dataset	#Samples	#Clusters	Semantic	#Groups	Sensitive Attributes
MNIST-USPS	67,291	10	Digit	2	Domain Source
Reverse-MNIST	120,000	10	Digit	2	Background Color
HAR	10,299	6	Activity	30	Subject
Office-31	3,612	31	Category	2	Domain Source
MTFL	2,000	2	Gender	2	w/ or w/o Glass
Mouse Atlas	6,954	11	Cell Type	2	Sequence Technique

2. Algorithm Implementation Details for FCMI

To elaborate on the working flow of our FCMI, we provide its pseudo code below. While the previous works [7, 9, 10] resort to some techniques such as layer-wise pre-training, pre-clustering and data augmentation, our method achieves state-of-the-art performance with end-to-end training initialized by a simple warm-up step.

Algorithm 1 Deep Fair Clustering via Minimizing and Maximizing Mutual Information.

Input: Samples $X = \{x_i\}_{i=1}^N$ with sensitive attributes $G = \{g_i\}_{i=1}^N$, $warmup\_epoch = 20$, $max\_epoch = 300$.

for $epoch < max\_epochs$ **do**
  Extract low-dimensional features via $h = \theta(X)$
  Compute cluster centers U by applying k-means on h
  for sampled mini-batch $\mathbf{x} = \{x_j\}_{j=1}^n$ **do**
    if $epoch < warmup\_epochs$ **then** ▷ Warmup
      Compute the overall loss $\mathcal{L} = \mathcal{L}_{rec}$ by Eq. 11 in our main paper.
    else ▷ Train
      Compute soft cluster assignments c_{ik} by Eq. 1
      Compute $\mathcal{L}_{rec}$, $\mathcal{L}_{clu}$ and $\mathcal{L}_{fair}$ by Eq. 11, Eq. 7 and Eq. 8 in our main paper.
      Compute the overall loss $\mathcal{L}$ by Eq. 12 in our main paper.
    end if
    Update encoder θ and decoder Φ to minimize $\mathcal{L}$ via Stochastic Gradient Descent (SGD).
  end for
end for

The remaining problem is how to compute the soft cluster assignments c_{ik}. To prove the effectiveness of the proposed unified information theory itself, without bells and whistles, we adopt the vanilla k-means to compute c_{ik}, namely,

$$c_{ik} = \frac{\exp(cos(h_i, u_k)/\tau)}{\sum_j \exp(cos(h_i, u_k)/\tau)}, \tag{1}$$

where $U = \{u_1, u_2, ..., u_K\}$ are obtained cluster centers and $\tau = 0.1$ is the temperature to control the softness. In our implementation, we apply K-means at the beginning of each epoch to update the clustering centers U. This would

* Equal contribution
† Corresponding author
[1] http://yann.lecun.com/exdb/mnist, https://www.kaggle.com/bistaumanga/usps-dataset

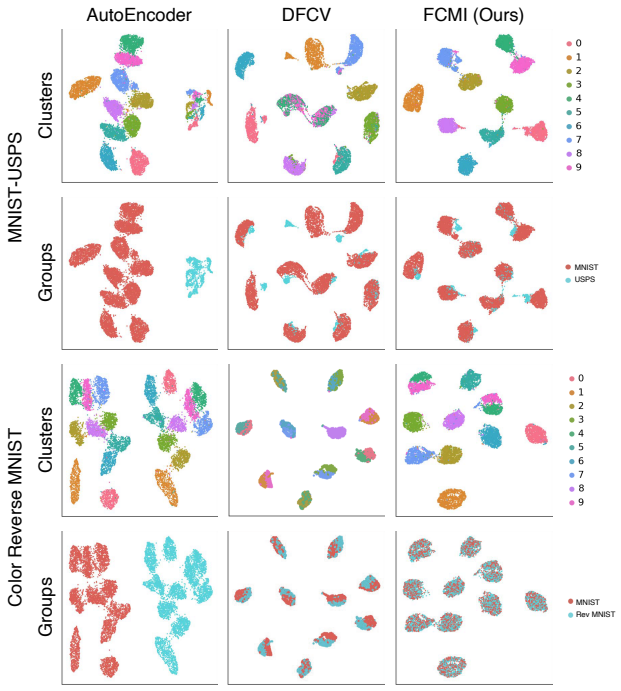

Figure 1. Visualizations of the hidden representation on MNIST-USPS and Color Reverse MNIST learned by our FCMI and two most competitive baselines.

only additionally introduce approximately 15% computational burden if we use the GPU implementation provided by faiss [6].

3. Visual Comparisons

To better show the superiority of the proposed FCMI, we conduct visualization on MNIST-USPS and Color Reverse MNIST in Fig. 1, comparing with two most competitive baselines, i.e., AE (a standard clustering method) and DFC (a fair clustering method).

From the visualizations, one could see that the standard clustering method AE fails to eliminate the influence of sensitive attributes, leading to unfair data partition. DFC is able to hide sensitive attributes from the clustering assignment. However, it fails to capture the intrinsic semantics accurately. Namely, it fails to distinguish digits 4 and 9 in MNIST and USPS, and it mixes all digits despite digit 8 on Color Reverse MNIST. In contrast, our FCMI successfully clusters data based on the digits except for digits 4 and 9 on Color Reverse MNIST. Moreover, our FCMI hides the sensitive attributes more thoroughly, as shown in the better mixing of digits from MNIST and Color Reverse MNIST (the last row).

4. Broader Impact Statement

The fairness problem hinders the utilization of clustering in a wide range of real-world applications such as image segmentation, biological analysis, and information retrieval since the real-world data might be confounded with sensitive attributes, *e.g.*, color, gender, race, and RNA-sequencing technique. Thus, it is crucial to achieve fair clustering and could bring many benefits, *e.g.*, preventing discrimination against certain individuals; reducing the cost of human labor to discover samples with similar semantics without prejudice. However, we should not ignore the potential negative impact of this work. More specifically, our method is based on a deep neural network, which inherits the black-box nature of artificial neural networks, thus would hinder its wide adoption in mission-critical applications.

References

[1] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra Perez, and Jorge Luis Reyes Ortiz. A public domain dataset for human activity recognition using smartphones. In *Proceedings of the 21th international European symposium on artificial neural networks, computational intelligence and machine learning*, pages 437–442, 2013. 1

[2] Arturs Backurs, Piotr Indyk, Krzysztof Onak, Baruch Schieber, Ali Vakilian, and Tal Wagner. Scalable fair clustering. In *International Conference on Machine Learning*, pages 405–413. PMLR, 2019. 1

[3] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. *Advances in Neural Information Processing Systems*, 30, 2017. 1

[4] Tabula Muris Consortium et al. Single-cell transcriptomics of 20 mouse organs creates a tabula muris. *Nature*, 562(7727):367–372, 2018. 1

[5] Xiaoping Han, Renying Wang, Yincong Zhou, Lijiang Fei, Huiyu Sun, Shujing Lai, Assieh Saadatpour, Ziming Zhou, Haide Chen, Fang Ye, et al. Mapping the mouse cell atlas by microwell-seq. *Cell*, 172(5):1091–1107, 2018. 1

[6] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3):535–547, 2019. 2

[7] Peizhao Li, Han Zhao, and Hongfu Liu. Deep fair clustering for visual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9070–9079, 2020. 1

[8] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010. 1

[9] Bokun Wang and Ian Davidson. Towards fair deep clustering with multi-state protected variables. *arXiv preprint arXiv:1901.10053*, 2019. 1

[10] Hongjing Zhang and Ian Davidson. Deep fair discriminative clustering. *arXiv preprint arXiv:2105.14146*, 2021. 1

[11] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Facial landmark detection by deep multi-task learning. In *European conference on computer vision*, pages 94–108. Springer, 2014. 1