Holistic AI × The Alan Turing Institute

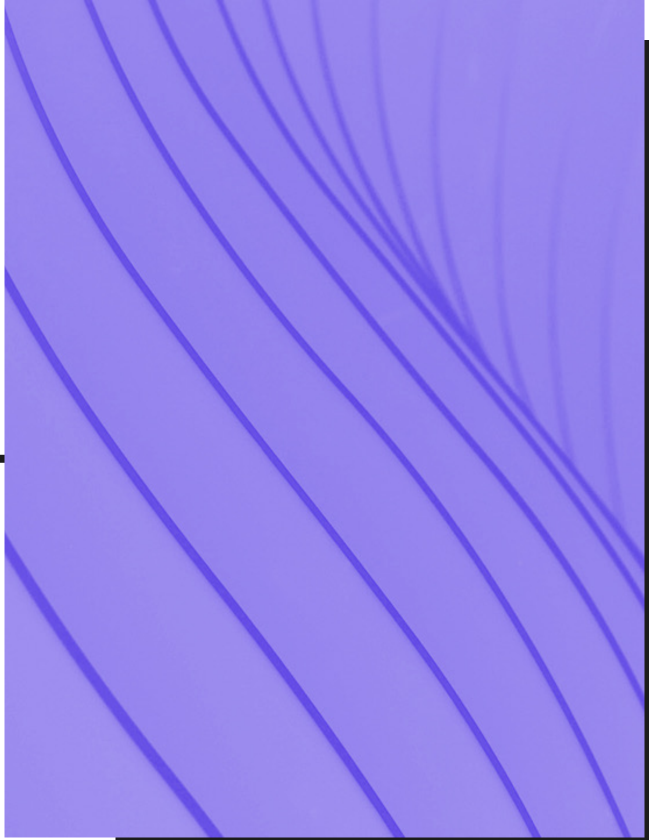
Bias in Clustering Systems Part IV

Sachin Beepath

Assurance Officer, Holistic AI

Contents

Mitigating Bias in Clustering



1. Introduce different levels of bias mitigation.
2. Formalize and implement techniques from different levels.
3. Compare the performance of the mitigation techniques on examples.

Three Levels of Mitigating Bias

Pre-Processing

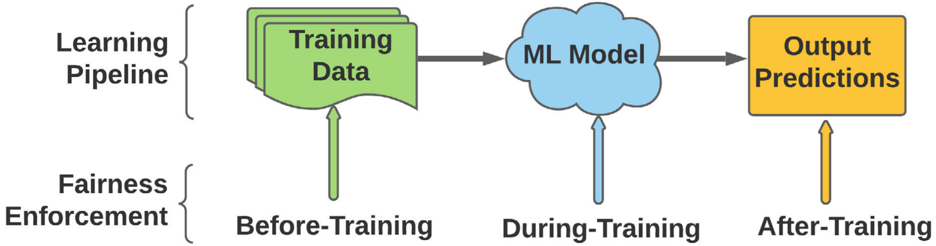
– Occurs ***before*** training. Requires original dataset to be modified. Algorithm is trained on new dataset to make predictions that meet fairness requirements.

In-Processing

– Occurs ***during*** training. Requires the model or learning process itself to be modified. Without changing original dataset, modify the algorithm to meet fairness requirements.

Post-Processing

– Occurs ***after*** training. Requires the outputs of the model to be modified. The results from the modification themselves must meet fairness requirements.



Part 4 Question 1

You have purchased a third party state-of-the-art model to determine who should be given a loan or not. You only have black-box access to the model and no access to training data. What level of bias mitigation is most suitable?

A. Pre-Processing

B. In-Processing

C. Post-Processing

Pre-Processing Bias Mitigation

- Occurs before learning.
- Makes changes to the training dataset.
- Original dataset X is transformed to X'.
- The algorithm remains the same but the. application of it to the transformed data results in fair clusters.
  - Fairlet Decomposition aims to find fairlets (micro-clusters) within data that meet fairness requirements.

In-Processing Bias Mitigation

- Changes the model by either altering clustering objective or algorithm itself to output fair clustering.
- The clustering algorithm A is modified to a new algorithm to A'.
- Need to optimize between clustering cost and fairness trade off .
- Variational Fair Clustering introduces fairness penalty to clustering objective to encourage fairness during learning.

Post-Processing Bias Mitigation

- Does not modify original data or algorithm.
- Use clustering algorithm A on inputs X to get clusters C. C is transformed to get fair clusters C'.
- Post-processes clustering centres such that every group is represented through centres equitably.
  - Making Existing Clusterings Fairer applies regular clustering and uses outputs to compute a new set of clusterings that are close to original and meet fairness requirements.

Fairlet Decomposition

– The goal is to find fairlets within data that meet fairness requirements.
– Fairlet: micro-clusters that aim to have equal representation of each group.
– The centres of the fairlets are then used as a new dataset to perform clustering.
– Since the fairlets themselves are balanced, the results of the clustering is as well.

Variational Fair Clustering

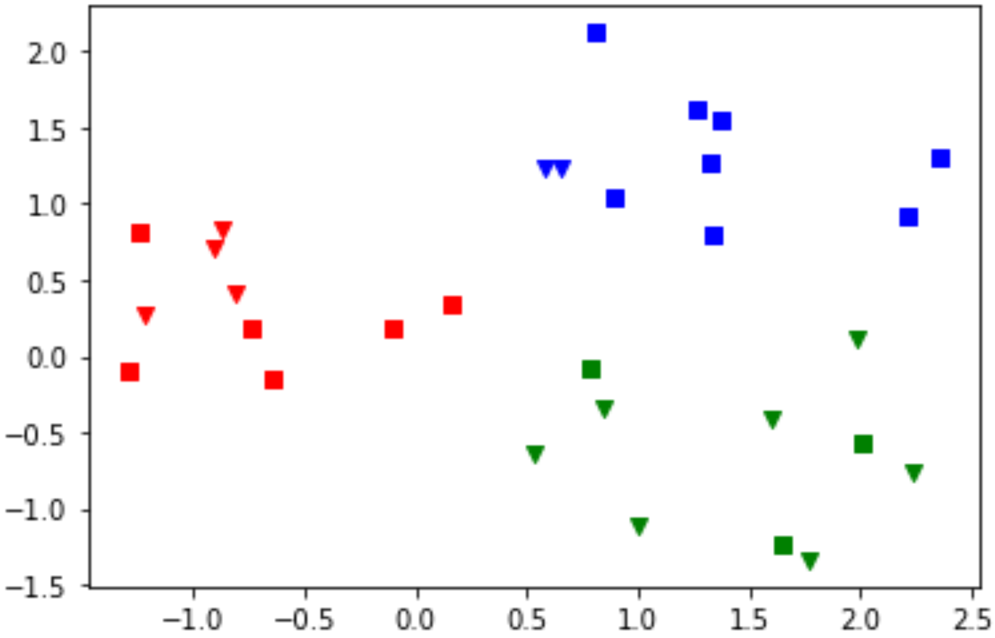
– Introduces a penalty term based on KL divergence to encourage fairness.

– Combined objective measures the trade-off between the clustering cost and fairness.

– Aims to find clusters with specified proportions of different protected group.

Toy Example

– Use k-means to label data into 3 clusters
– 2 protected groups: Triangles and Squares



Toy Example (Cont.)

Metric	Value
Cluster Balance	0.462
Cluster Distribution KL Divergence	0.378
Cluster Social Fairness Ratio	1.022
Cluster Silhouette Difference	0.002

Toy Example with Fairlet Decomposition

– First we import the mitigation technique and house it in a pipeline:

```python
from holisticai.bias.mitigation import FairletClusteringPreprocessing

# initialize pre-processing method
decomposition = FairletClusteringPreprocessing(decomposition='Scalable', p=10, q=21, seed=42)

# initialize pipeline
pipeline = Pipeline(steps=[
    ('scaler', StandardScaler()),
    ('bm_preprocessing', decomposition),
    ('cluster', KMeans(n_clusters=3))])

pipeline.fit(pairs, bm__group_a = group_a, bm__group_b = group_b)
```

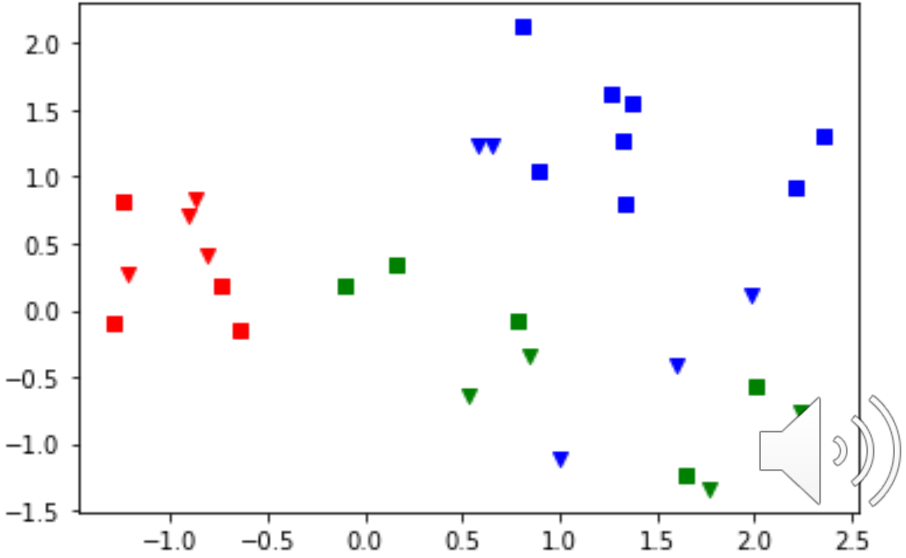
– We can then access the clustering centroids and make predictions:

```python
print(pipeline.predict(data))
print(pipeline['cluster'].cluster_centers_)
```

Toy Example with Fairlet Decomposition

- Red cluster has 4 Squares and 4 Triangles, Blue cluster has 8 Squares and 5 Triangles, Green Cluster has 5 Squares and 4 Triangles.
- Clear improvement with respect to the balance and KL Divergence.
- Clustering is less intuitive than before, overlapping of clusters.

Metric	Value
Cluster Balance	0.867
Cluster Distribution KL Divergence	0.019
Cluster Social Fairness Ratio	1.022
Cluster Silhouette Difference	0.028



Toy Example with Variational Fair Clustering

– Import the mitigation technique, house it in a pipeline, and train model:

```
from holisticai.bias.mitigation import VariationalFairClustering

vfc_inprocessing = VariationalFairClustering(nb_clusters=3, method='kmeans')

# initialize the pipeline
pipeline = Pipeline(steps=[
    ('scaler', StandardScaler()),
    ('bm_inprocessing', vfc_inprocessing)])


pipeline.fit(data, bm__group_a = group_a, bm__group_b = group_b)
```

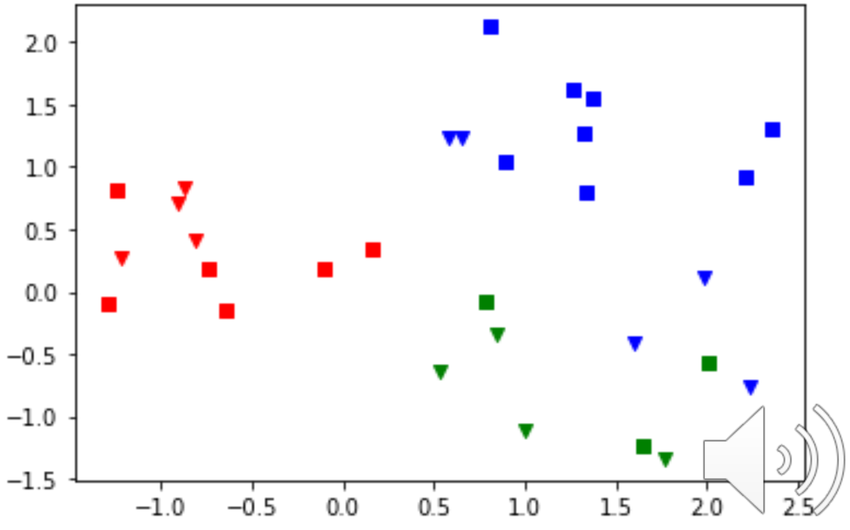
Toy Example with Variational Fair Clustering

– Red cluster has 6 Squares and 4 Triangles, Blue cluster has 8 Squares and 5 Triangles, Green Cluster has 3 Squares and 4 Triangles.

– Clear improvement with respect to the balance and KL Divergence.

– Clustering is less intuitive than before, overlapping of clusters.

Metric	Value
Cluster Balance	0.756
Cluster Distribution KL Divergence	0.045
Cluster Social Fairness Ratio	1.022
Cluster Silhouette Difference	0.07



Conclusion

– Introduced different levels of bias mitigation .
– Formalized and implemented various mitigation techniques from different levels.
– Compared the performance and outcomes of the techniques on examples.

Milestone Conlcusion

Part I – Introduction to Clustering

Part II – Fairness in Clustering Tasks

Part III – Measuring Bias in Clustering Systems

Part IV – Mitigating Bias in Clustering Systems

Exercise Notebooks

– Following the lectures there are two notebooks to complete.
– Measuring Bias
– Mitigating Bias

References

- Chhabra, Anshuman, Karina Masalkovaitė, and Prasant Mohapatra. "An overview of fairness in clustering." *IEEE Access* (2021).
- Madhulatha, T. Soni. "An overview on clustering methods." *arXiv preprint arXiv:1205.1117* (2012).
- Gan, Guojun, Chaoqun Ma, and Jianhong Wu. *Data clustering: theory, algorithms, and applications*. Society for Industrial and Applied Mathematics, 2020.
- Davidson, Ian, and S. S. Ravi. "Making existing clusterings fairer: Algorithms, complexity results and insights." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 04. 2020.
- Kleindessner, Matthäus, Pranjal Awasthi, and Jamie Morgenstern. "A notion of individual fairness for clustering." *arXiv preprint arXiv:2006.04960* (2020).
- V. Anand, “FPL Historical Dataset,” Retrieved December 2022 from https://github.com/vaastav/Fantasy-Premier-League/, 2022.
- Tung, Frederick, Alexander Wong, and David A. Clausi. "Enabling scalable spectral clustering for image segmentation." *Pattern Recognition* 43.12 (2010): 4069-4076.