

A novel non-negative matrix factorization for addressing fairness issues

Alejandra Jimena Inga Quezada

Faculdade de Tecnologia (FT), Universidade Estadual de Campinas (Unicamp)
Limeira-SP

`alejandra.ingaq@gmail.com`

Leonardo Tomazeli Duarte

Faculdade de Ciências Aplicadas (FCA), Universidade Estadual de Campinas (Unicamp)
Limeira-SP

`leonardo.duarte@fca.unicamp.br`

ABSTRACT

Nonnegative Matrix Factorization (NMF) has emerged as a powerful dimension reduction technique for decomposing non-negative data and finding its non-negative representations. Furthermore, NMF not only reduces data dimensions but also extracts the most relevant parts that can be interpretable while preserving as much information as possible. In the machine learning context, the feature extraction of NMF can be seen as an unsupervised learning method when applied to actual problems. However, as in most dimension reduction techniques, NMF does not address fairness concepts inside its general process. Correspondingly, the results usually have bias towards one of the analyzed groups. Therefore, the lack of fairness inside NMF may lead to disparate results that negatively affect sensitive groups (e.g., gender, race, etc.). To address this problem, different fairness metrics have been thoroughly studied for dimension-reduction techniques, and important results have been made especially for fairness-aware PCA formulations. We propose a novel NMF algorithm that tackles fairness issues based on the reconstruction error metric. To accomplish this task, we formulate a cost function that comprises two terms: the reconstruction error and a fairness metric. The optimization of this cost function is accomplished through gradient-based method. First experimental results have been made using synthetic data where the reconstruction error and the disparities in reconstruction errors of different groups is minimized. This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. The authors would like to thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the financial support to conduct this research and the support of BIOS - Brazilian Institute of Data Science, supported by process 2020/09838-0, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

KEYWORDS. Nonnegative Matrix Factorization. Fairness. Unsupervised learning.