Vitaliy Miroshnychenko[1], Vadym Tunik[2]

MoA+: Mixture of Autoencoders with Various Concentrations for Enhanced Image Clustering

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine.

In this paper, we consider the improvement of the vanilla Mixture of Experts model [1] in the unsupervised image clustering task by introducing an unbiased loss function based on the idea of a mixture of various concentrations. This addresses expert dominance, enhancing training balance and clustering performance.

The model is tested on the MNIST dataset, a benchmark of N grayscale handwritten digit images $X_i, i = 1..N$, treated as a mixture with $M = 2$ components.

In the MoA+ framework, each expert is a convolutional autoencoder (CAE), similar to simplified U-Net or SegNet, where the k-th CAE reconstructs image X_i into $v_{i,k}$ with MSE error. A convolutional gate network assigns probabilities $p_{i,k}$, clustering images by selecting the most suitable CAE.

To prevent expert imbalance, MoA+ uses the Q-loss, inspired by ST-MoEs z-loss [2].

$$Q\text{-loss} = \sqrt{\sum_{k=1}^{M} \left(\sum_{i=1}^{N} p_{i,k} \cdot \mathrm{MSE}(v_{i,k}, X_i) \right)^2}$$

This regularizes the gate network, ensuring all experts contribute effectively to the clustering process.

Compared under identical conditions, the standard MoA achieved a Normalized Mutual Information (NMI) of ~ 0.08, while MoA+ scored ~ 0.8, showing significantly better clustering performance.

References

[1] Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). *Adaptive Mixtures of Local Experts.* Neural Computation, 3(1), 79–87. https://direct.mit.edu/neco/article/3/1/79/5560/Adaptive-Mixtures-of-Local-Experts.

[2] Author, A. B. (2022). *ST-MoE: Designing Stable and Transferable Sparse Expert Models.* arXiv preprint arXiv:2202.08906. https://arxiv.org/abs/2202.08906.

E-mail: ✉[1] *vitaliy.miroshnychenko@knu.ua,* ✉[2] *tunik.vadym@knu.ua.*