Unsupervised learning for image and text classification

Environment: QARMA (machine learning) Team at LIS (Laboratoire d'Informatique et Systèmes)
Location: Ecole Centrale Marseille (ECM), Technopôle de Château-Gombert, Marseille.
Supervisors: R. Boutalbi (LIS - Polytech), R. Sicre (LIS – ECM)
Salary: legal minimum
Keywords: Unsupervised learning, open-set learning, self-supervised learning, tensor clustering.
Contact: rafika.boutalbi@lis-lab.fr, ronan.sicre@lis-lab.fr

The rise of deep deep learning was enabled by two main factors: the availability of large amount of labelled data and high computation power with GPUs. Recent works study self-supervised methods that allow to learn good and transferable representations, without requiring any labels. Moreover, open-set learning search to align data groups with unknown categories, via unsupervised learning.

We are particularly interested in finding such unsupervised data alignment with categories without any label, by combining self-supervised methods generating numerous data representations as well as tensor clustering to combine the information of each representation.

Concerning self-supervised learning for visual recognition, the first methods were based on rotations, and deep clustering [2]. Later methods propose contrastive loss with several augmentations. This loss pulls together representation of augmentations of the image of interest and pushes away representations of other dataset images. BYOL [4] further proposed to combine this approach with teacher-student distillation. With the recent arrival of vision transformers, iBOT [6] adds random masking of the patch tokens to further improve the representations.

Similar self-supervised methods can be used on text data to obtain another set of transferable representations. Indeed, if we consider textual data, several embeddings can be generated, such as BOW, static embeddings (Word2vec, GloVe, etc.), contextual embeddings (Bert, Roberta, etc.), extracted entities, etc. However, it has been proven that there is no better text representation [5], and the representation is chosen according to an empirical evaluation of the researchers. In [1], the authors highlighted the impact of combining several representations on clustering results using tensor structures, allowing us to benefit from the information diversity arising from different representations.

In this work, we will extend the use of this type of approach [3] for image data. We will also develop an approach to cluster objects into a large number of classes (hierarchical clustering), which current tensor clustering models are not able to achieve.

References

[1] R. Boutalbi, M. Ait-Saada, A. Iurshina, S. Staab, and M. Nadif. Tensor-based graph modularity for text data clustering. SIGIR '22, page 2227–2231. ACM, 2022.

[2] M. Caron, P. Bojanowski, J. Mairal, and A. Joulin. Unsupervised pre-training of image features on non-curated data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2959–2968, 2019.

[3] G. Drakopoulos, E. Spyrou, and P. Mylonas. Tensor clustering: A review. In *2019 14th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*, pages 1–6. IEEE, 2019.

[4] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

[5] T. Walkowiak and M. Gniewkowski. Evaluation of vector embedding models in clustering of text documents. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 1304–1311, 2019.

[6] J. Zhou, C. Wei, H. Wang, W. Shen, C. Xie, A. Yuille, and T. Kong. ibot: Image bert pre-training with online tokenizer. *arXiv preprint arXiv:2111.07832*, 2021.