Deep clustering of multimodal data

Université Sorbonne Paris Nord. LIPN - UMR CNRS 7030
Université de Versailles, Université Paris-Saclay, DAVID Lab
Équipes A3*& Adam†

1 Internship context and objectives

Generative models and transformer models based on attention mechanisms have recently gained significant popularity. The success of the paper "Attention is All You Need" [1] was a true harbinger of the new era of deep learning. Our current work, as part of a CIFRE thesis with BPCE[1] [2], has highlighted the importance of combining attention mechanisms with adaptive learning methods for encoding tabular data. We have also explored the use of conditional generative models for encoding multivariate time series, as part of a second CIFRE thesis with SAFRAN[2] [3].

In the literature, there are methods like VirTex [4], CLIP [5], and GPT-4 [6] that propose a solution for representing multimodal data but use different encoders to extract information from the image and text before finding a common projection space. These considerations are part of the issues addressed in this internship subject. We are thus studying the contribution of attention mechanisms in the processing of multimodal medical data. This internship will be carried out as part of the IHU PROMETHEUS project. One technique for addressing multimodality is to act at the encoding level by proposing a single universal encoder capable of representing all the modalities. The encoder must find a latent space where similar data (text-text, image-image, text-image) are represented with similar vectors (in terms of similarity measure, e.g., cosine similarity).

Clustering is an unsupervised learning method that seeks to partition data into groups such that elements of the same group are similar while elements of different groups are dissimilar. Given a data matrix where rows represent observations and columns represent variables, co-clustering [7] infers a row partition and a column partition simultaneously. The resulting partition is composed of homogeneous blocks. When a dataset exhibits a dual structure between observations and variables, co-clustering outperforms conventional clustering algorithms, which only infers a row partition without considering the relationships between observations and variables. Co-clustering is a powerful data mining tool for two-dimensional data and is widely applied in bioinformatics [8]. Another internship objective is then to apply clustering and/or co-clustering algorithms for these processed data. For example, in gene expression data, co-clustering simultaneously clusters samples and genes, revealing groups of highly correlated genes with distinct correlation structures among different sets of individuals, such as between disease and healthy individuals or different types of disease. This may allow to identify which genes are responsible for some diseases.

*Apprentissage Artificiel et Applications
†Ambient Data Access and Mining
[1] https://groupebpce.com/
[2] https://www.safran-group.com/fr

Recent approaches combine clustering methods with deep neural networks to perform the feature extraction step, which has shown a remarkable performance improvement. The deep clustering methods can be divided into two categories: Two-stage methods that first perform representation learning using deep neural networks, such as autoencoders [9]. Then, a clustering algorithm is applied to the learned representation. The second category considers this separation, and recent works [10, 11, 12] proposed methods based on a unified framework that jointly learn deep representations and clusters. The last objective of the internship is to develop models that simultaneously perform feature extraction and clustering of multimodal data.

2 Internship plan

The research directions might be (but not restricted to):

- Study the current state-of-the-art on representation learning and clustering of multimodal data with a focus on medical data.
- Study the current state-of-the-art on deep clustering.
- Propose a framework that combines feature extraction and clustering of multimodal medical data (separately).
- Propose a framework that unifies feature extraction and clustering of multimodal medical data (simultaneously).
- Scientific publication in international ML and/or Medical conferences.

Prerequisites and competence

- End of engineering degree, M2 in data science, statistics and/or artificial intelligence.
- Excellent experience in programming, especially with Python, Tensorflow, and scikit-learn.
- Big interest in and excellent understanding of machine and deep learning theory and applications.
- Excellent writing skills.

To apply, simply attach:

- Your current Curriculum Vitae (CV),
- A portfolio of projects, if any,
- Your motivation for the position,
- Your latest university transcripts.

Send it all by email to mlresearch.internship@gmail.com

3 Supervisors team

- Hanane Azzag, LIPN, CNRS UMR 7030, Université Sorbonne Paris Nord
- Reda Khoufache, DAVID Lab, UVSQ, Université Paris-Saclay/LIPN, CNRS UMR 7030, Université Sorbonne Paris Nord
- Mustapha Lebbah, DAVID Lab, UVSQ, Université Paris- Saclay

The internship may lead to a PHD position.

References

[1] Ashish Vaswani and al. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

[2] Kodjo Mawuena Amekoe, Mohamed Djallel Dilmi, Hanene Azzag, Zaineb Chelly Dagdia, Mustapha Lebbah, and Gregoire Jaffre. Tabsra: An attention based self-explainable model for tabular learning. In *ESANN*, 2023.

[3] Abdellah Madane, Mohamed djallel Dilmi, and Jerome Lacaille Florent Forest and **Hanane Azzag**, Mustapha Lebbah. Transformer-based conditional generative adversarial network for multivariate time series generation. In *International Workshop on Temporal Analytics@PAKDD 2023.*, 2023.

[4] Karan Desai and Justin Johnson. Virtex: Learning visual representations from textual annotations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11162–11173, 2021.

[5] Alec Radford, Jong Wook Kim, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[6] Anis Koubaa. Gpt-4 vs. gpt-3.5: A concise showdown. 2023.

[7] G. Govaert and M. Nadif. *Co-Clustering: Models, Algorithms and Applications*. FOCUS Series. Wiley, 2013.

[8] Daniel Hanisch, Alexander Zien, and Ralf Zimmer. Co-clustering of biological networks and gene expression data. *Bioinformatics*, 18, 05 2002.

[9] Pascal Vincent, Hugo Larochelle, Y. Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. pages 1096–1103, 01 2008.

[10] Hongfu Liu, Ming Shao, Sheng Li, and Yun Fu. Infinite ensemble for image clustering. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1745–1754, New York, NY, USA, 2016. Association for Computing Machinery.

[11] Jianwei Yang, Devi Parikh, and Dhruv Batra. Joint unsupervised learning of deep representations and image clusters, 2016.

[12] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 478–487, New York, New York, USA, 20–22 Jun 2016. PMLR.