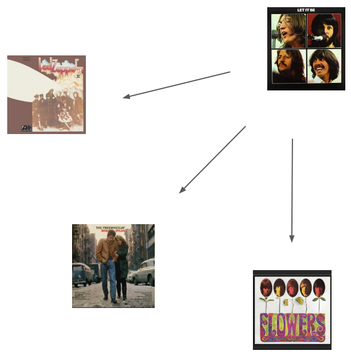
Contrastive Learning for Cross-modal Artist Retrieval

Andres Ferraro, Jaehun Kim, Andreas Ehmann, Sergio Oramas, Fabien Gouyon
Pandora SiriusXM

What is artist retrieval?

* Given an artist we want to retrieve the most similar artists
* How do we define artist similarity?



Available online:

What is multimodal artist retrieval?

Get richer representations for the full population of artists by leveraging information from:

* Audio
* Tags
* CF data

Metrics

Use multiple metrics to capture different aspects of:

* **Accuracy**: Match with artists in the ground truth (e.g., P, R, NDCG, MAP)
* **Coverage**: artists that appear at least once
* **Gini**: distribution of the artists in the recommendation

Experiments

Contrastive method to predict artist similarities:

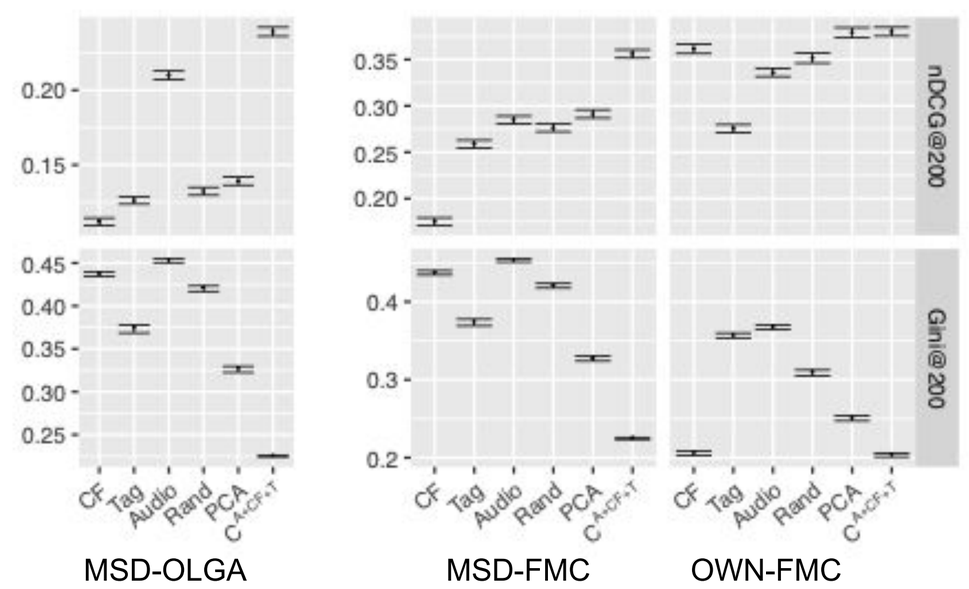
* Trained on: Million Song Dataset (MSD) and Internal Data (OWN)
* Tested on: Olga (external) dataset
* Compare Contrastive with baselines:
  - Only Audio, only CF, and only Tags
  - Combination of Audio, CF and Tag using PCA and Random projection

Main Contributions

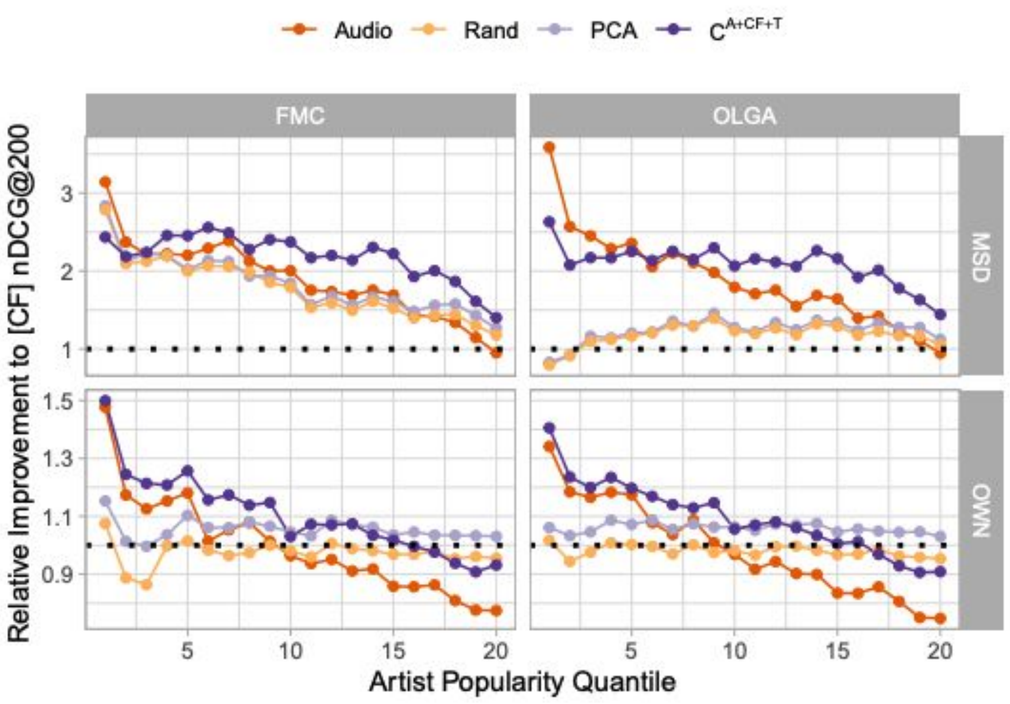
We show under two different contexts –using an open and an in-house dataset– that our proposed approach:

* Achieves higher performance in terms of accuracy and coverage of retrieved artists
* Particularly increases the performance for less popular query artists
* Allows to do cross-modal retrieval better than the baselines
* Successfully combines complementary information from diverse modalities even with missing data

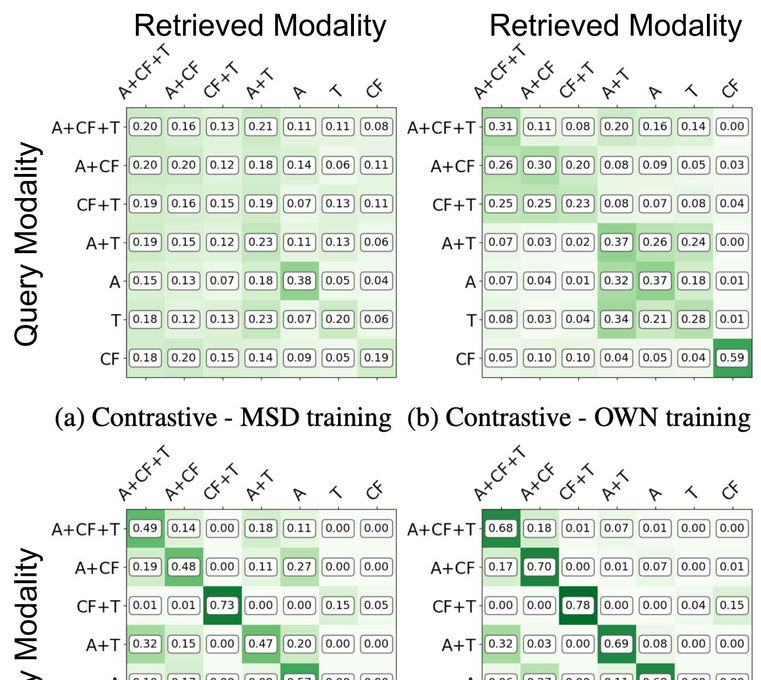
Performance comparison of contrastive method



Effect of Popularity



Robustness to missing modality data



(a) Contrastive - MSD training (b) Contrastive - OWN training

(c) PCA - MSD training (d) PCA - OWN training

Acknowledgements

We would like to express special thanks to Matt McCallum for the help collecting audio features and Sam Sandberg for his valuable comments.

ISMIR 2023 Milan, Italy
Nov. 5-9, 2023
POLITECNICO MILANO 1863