Wasserstein Barycenter Transport for Acoustic Adaptation

Eduardo Fernandes Montesuma[⋆,†], Fred Maurice Ngolè Mboula[†]
[⋆]Universidade Federal do Ceará, Departamento de Teleinformática, Fortaleza, Brazil
[†]Université Paris-Saclay, CEA, List, F-91120 Palaiseau, France

46$^{\text{th}}$ International Conference on Acoustics, Speech, and Signal Processing (ICASSP) - June 2021

Summary

Introduction

Context and Contribution

Context:

- Music Genre Recognition and Music-Speech Discrimination in light of different background noise (multiple domains).
- Multi-Source Domain Adaptation (MSDA).

Contribution:

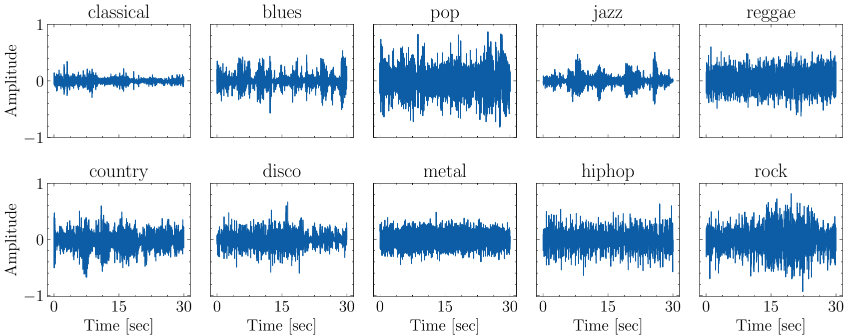
- MSDA algorithm using Optimal Transport with **state-of-the-art performance**, the **Wasserstein Barycenter Transport (WBT)**.

Our code is publicly available on Github[1].

[1] `https://github.com/eddardd/WBTransport`

Music Genre Recognition

Given a music sample of 30 seconds, **to which music genre it belongs?**

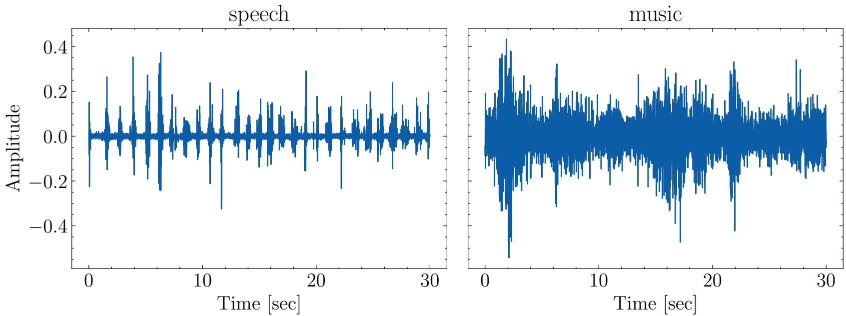


We use the dataset of Tzanetakis et al.[2]

[2] Tzanetakis, G. et al. (2002). Musical genre classification of audio signals. IEEE Transactions on speech and audio processing, 10(5), 293-302.

Music Speech Discrimination

From an audio file of 30 seconds, does it consists on **music** or **speech?**



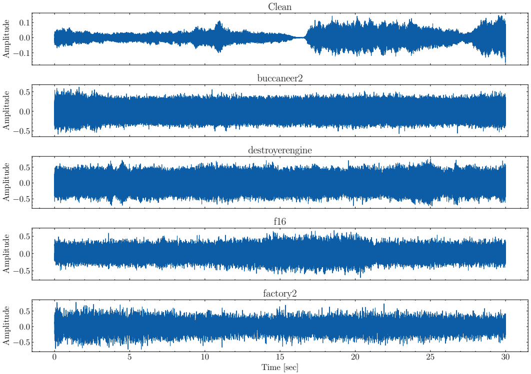
We use the dataset of Tzanetakis et al.[3]

[3]Tzanetakis G., et al. "Automatic musical genre classification of audio signals," in Proceedings of the 2nd international symposium on music information retrieval, Indiana, 2001.

Multi-Source Acoustic Adaptation

What if audio has different conditions of acquisition? **Can a classifier work on these different settings?**

- We use background noise to simulate different signal acquisition conditions.
- **Consequence:** data follows different probability distributions.



Methodology

Transfer Learning

Our work inserts itself on the context of **transfer learning/MSDA**:

- Training data is constituted by different **source domains**. Test data represents a **target domain**.
- Each domain is assumed to have a **different probability distribution**.
- **Challenge:** Fitting a classifier that learns on the source domains, that works well on the target domain.

Optimal Transport

Optimal Transport is a theory concerned with the **transportation of probability distributions**. As proposed by Courty et al.[4] it can be used for domain adaptation:

- Step 1. Find the Optimal Transport plan γ using

$$\gamma^* = \underset{\gamma \in \Pi(\mu_s, \mu_t)}{\operatorname{argmin}} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \gamma_{ij} C_{ij}$$

- Step 2. Use γ^* to find the **Barycentric Mapping**,

$$T_{\gamma^*}(\mathbf{x}_i^s) = \underset{\mathbf{x} \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{j=1}^{n_t} \gamma_{ij}^* C(\mathbf{x}, \mathbf{x}_j^t)$$

[4] Courty, N., et al (2016). Optimal transport for domain adaptation. IEEE transactions on pattern analysis and machine intelligence, 39(9), 1853-1865.

Optimal Transport

The previous approach minimizes the **Wasserstein Distance** between distributions μ_s and μ_t, given by:

$$W_p^p(\mu_s, \mu_t) = \underset{\gamma \in \Pi(\mu_s, \mu_t)}{\text{minimize}} \sum_{i=1}^{n_s} \sum_{j=1}^{n_t} \gamma_{ij} ||\mathbf{x}_i^s - \mathbf{x}_j^t||$$

Important: The Wasserstein Distance is a distance between probability distributions, as proved by Villani[5].

[5]Villani, Cédric. Optimal transport: old and new. Vol. 338. Springer Science & Business Media, 2008.

Wasserstein Barycenter Transport

Motivation: Since W_p^p is a **distance**, we may calculate the **barycenter of probability distributions**:

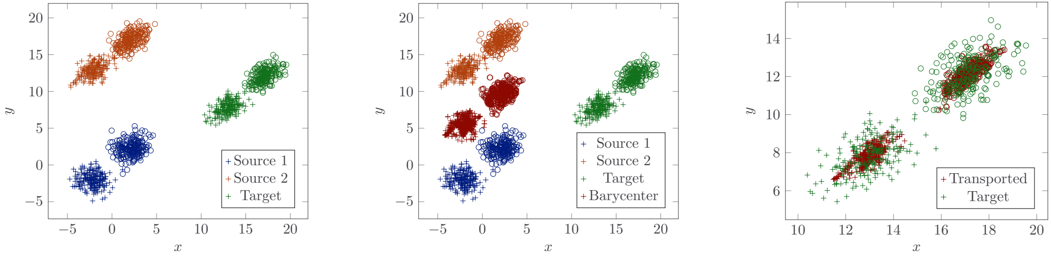
$$\mu_b = \underset{\mu}{\text{argmin}} \sum_{k=1}^{N} \lambda_k W_C(\mu_k, \mu).$$

Strategy:

1. Calculate the Wasserten barycenter of all sources,
2. Perform the transport between the barycenter domain to the target domain.

Wasserstein Barycenter Transport

Intuition:



(a) Multi-source domain adaptation.

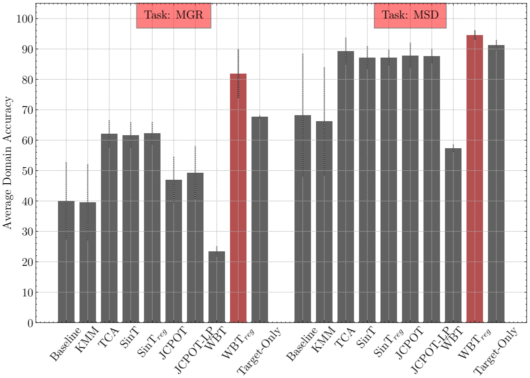
(b) Intermediate step: Barycenter calculation.

(c) Final step: Transport Barycenter $\rightarrow$ Target.

Figure 1: Illustration of the Wasserstein Barycenter Transport algorithm. The multi-source assumption is illustrated in (a). The method is composed of two steps, shown in (b) and (c).

Results

Results



- Class-based regularization is key for the success of our method (WBT vs. WBT$_{reg}$).
- Our method (WBT$_{reg}$) performs better than the state of the art and **target-only** classifiers.

Conclusion

Conclusion

We propose a technique that,

- **Improved the baseline performance** (accuracy) on all target domains.
    - Improvement on average: 41.77% (MGR) vs. 26.43% (MSD).
- **Has state-of-the-art performance** (accuracy) on all target domains.
    - Improvement on average to the second best: 19.67% (MGR) vs. 5.20% (MSD)
- **Improved target-only performance** (accuracy) on all target domains.
    - Improvement on average: 14.25% (MGR) vs. 3.29% (MSD).