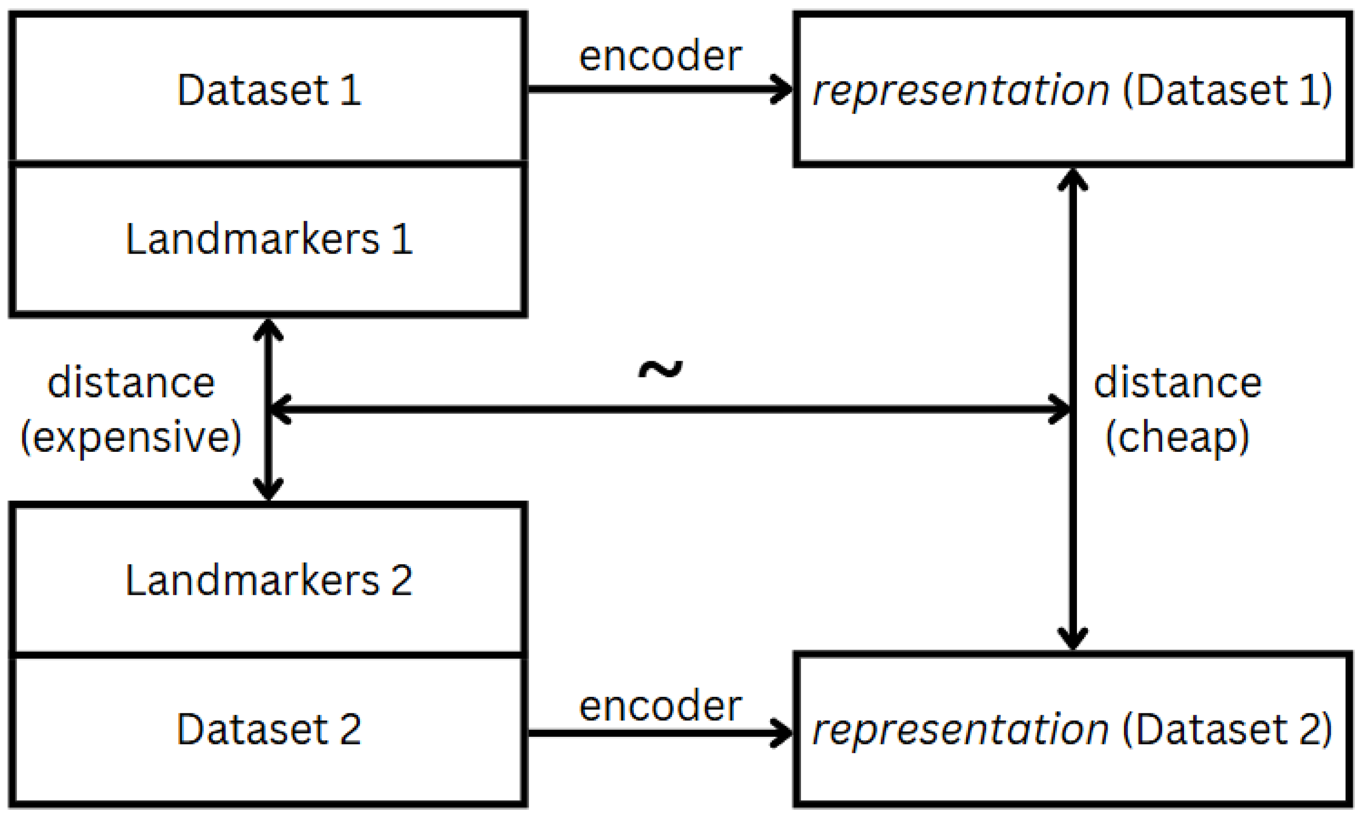
Are encoders able to learn landmarkers for warm-starting of Hyperparameter Optimization?

Antoni Zajko   Katarzyna Woźnica

Warsaw University of Technology

TLDR

The main problem that this work is trying to solve is the development of a tabular dataset encoder, which will produce representations that carry relevant information in the context of warm-starting Bayesian HPO.



Motivation

- One of AutoML's open problems is a measurement of tabular datasets similarity.
- Existing methods are based on predefined or learned meta-features, which often underperform against simple heuristics.
- Existing solutions are either too costly, yield minimal benefits, or lack proper evaluation.
- In this work, we attempt to solve Bayesian HPO's problem – cold-start using novel dataset encoders.

Methods

We propose two novel methods of modeling tabular datasets similarity:

- The first solution relies on the deep metric learning of Dataset2Vec [1]. It aims to learn an encoder to produce such tabular dataset representations that their distances are close to distances between corresponding landmarkers.

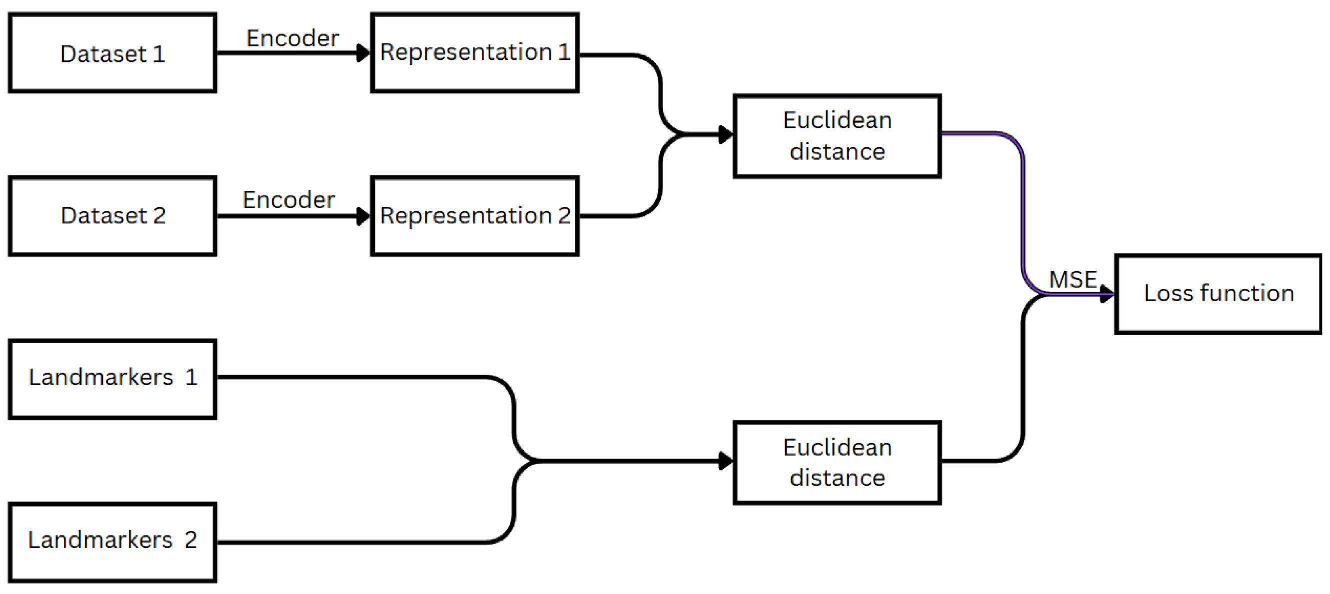


Figure 1. Visualization of metric learning approach

- The second approach aims to reconstruct landmarkers. To achieve that, it first uses the Dataset2Vec encoder to produce a latent representation of the dataset, which is passed into the reconstruction network, which is simple MLP. As a loss function, this approach uses reconstruction error.

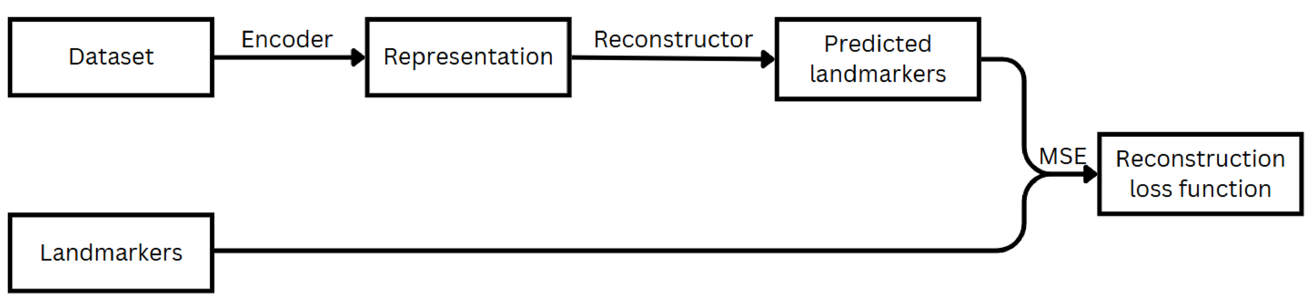


Figure 2. Visualization of landmarker reconstruction approach

Contact

- antoni.zajko.stud@pw.edu.pl
- katarzyna.woznica@pw.edu.pl

Acknowledgements

Research was funded by Warsaw University of Technology within the Excellence Initiative: Research University (IDUB) programme.

Experiments

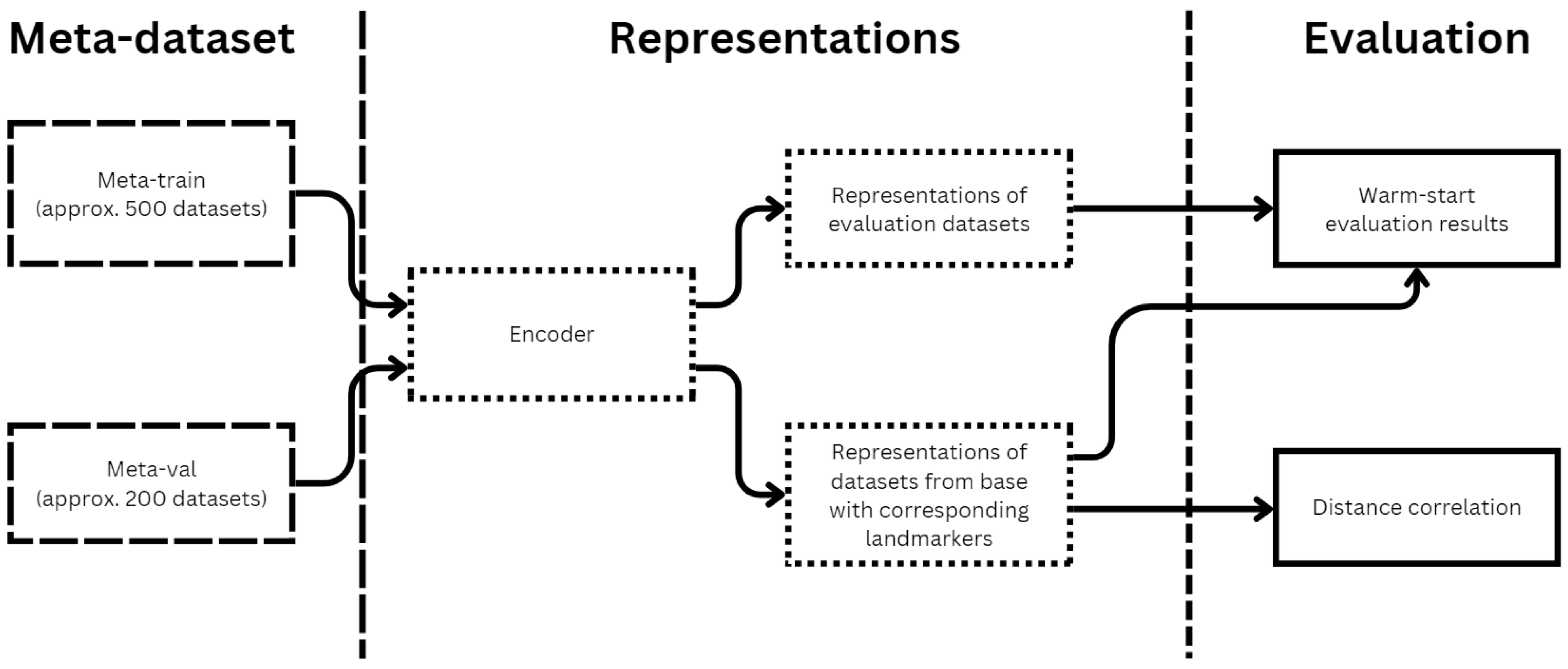


Figure 3. Flow of evaluation of each proposed approach

- As a baselines, we used a basic version of Dataset2Vec and simple heuristics for warm-starting HPO.
- We evaluated encoders in two ways: measurement of gains in the target meta-task and distance correlation between representations and distances between corresponding landmarkers.

Results

Results show that:

- Encoders managed to grasp information stored within landmarkers which can be seen in significant rank correlation between distances.
- However, our approaches did not provide significant gains in the warm-starting of the Bayesian HPO, which limits their practical applications.

Table 1. Rank correlations between encoders' outputs and landmarkers

Encoder	Average	Stdev
Dataset2Vec basic	0.037	0.024
Dataset2Vec metric	0.332	0.030
Dataset2Vec reconstruction	0.177	0.027

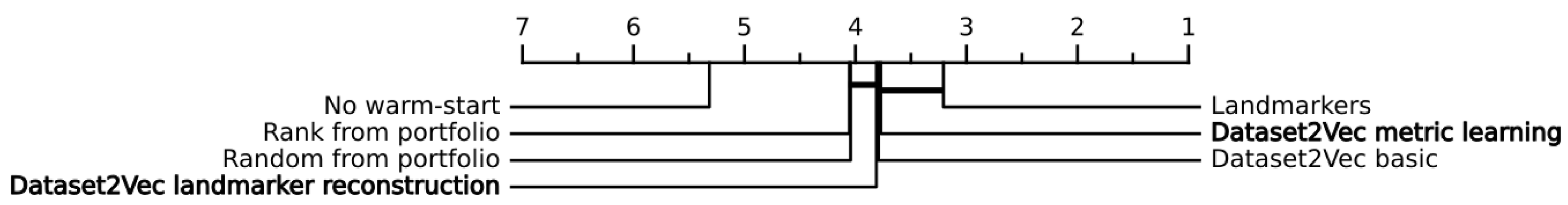


Figure 4. Comparison of objective values at the end of the initial phase

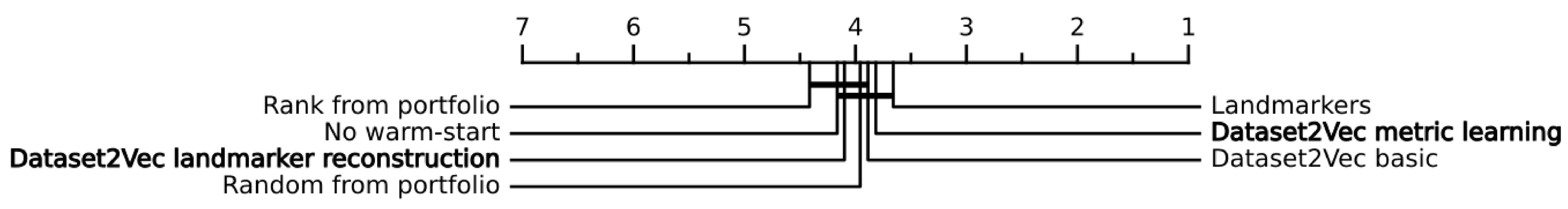


Figure 5. Comparison of objective values at the end of optimization

Discussion

Such results can be caused by:

- Noise induced by the Bayesian Optimization itself, which may not be able to sufficiently leverage initial configurations.
- Improper selection of the configurations contained in the portfolio.
- Encoders, which yield too little landmarker information

References

[1] Hadi S Jomaa, Lars Schmidt-Thieme, and Josif Grabocka. Dataset2Vec: Learning Dataset Meta-Features. *Data Mining and Knowledge Discovery*, 35:964–985, 2021.

[2] Dawid Płudowski, Antoni Zajko, Anna Kozak, and Katarzyna Woźnica. Rethinking of encoder-based warm-start methods in hyperparameter optimization, 2024.