MI2 DATALAB

metaMIMIC

an analysis of hyperparameter transfer possibilities for tabular data using MIMIC-IV database

Mateusz Grzyb
Zuzanna Trafas

IDEA:

Motivation: **why/when the transfer learning occurs in tabular data?**

What we want:

- A benchmark of problems from the medical domain using MIMIC-IV database
- Compare similarity of the best hyperparameter sets between different tasks
- Test if we can benefit from hyperparameter transfer

HOW WE DID IT


Database setup
1
Selecting targets
2
Selecting predictors
3
Running the experiments
4
Results analysis
5

HOW WE DID IT


Database setup
1
Selecting targets
2
Selecting predictors
3
Running the experiments
4
Results analysis
5

DATABASE SETUP

MIMIC-IV

PostgreSQL

docker

HOW WE DID IT


Database setup
1
Selecting targets
2
Selecting predictors
3
Running the experiments
4
Results analysis
5

SELECTING TARGETS

From the top 50 most commonly appearing conditions

Table 1: Selected targets

Category	ICD-9	ICD-10	% in population
Hypertensive diseases	401-405	I10-I16	59.78%
Disorders of lipoid metabolism	272	E78	40.27%
Anemia	280-285	D60-D64	35.93%
Ischematic heart disease	410-414	I20-I25	32.79%
Diabetes	249-250	E08-E13	25.27%
Chronic lower respiratory diseases	466, 490-496	J40-J47	19.48%
Heart failure	428	I50	19.41%
Hypotension	458	I95	14.38%
Purpura and other hemorrhagic conditions	287	D69	11.9%
Atrial fibrillation and flutter	427.3	I48	10.48%
Overweight, obesity and other hyperalimentation	278	E65-E68	10.46%
Alcohol dependence	303	F10	7.67%

HOW WE DID IT


Database setup
Selecting targets
Selecting predictors
Running the experiments
Results analysis
1
2
3
4
5

SELECTING PREDICTORS

44 tables in MIMIC-IV

Time series?

How to aggregate?

What tests to select?

Chartevents

Charted items occurring during the ICU stay.

Labevents

Laboratory measurements sourced from patient derived specimens.

HOW WE DID IT


Database setup
1
Selecting targets
2
Selecting predictors
3
Running the experiments
4
Results analysis
5

RUNNING THE EXPERIMENTS

4-CV ROC AUC measure for 21 tasks

8 different XGBoost parameters

1000 hyperparameter sets

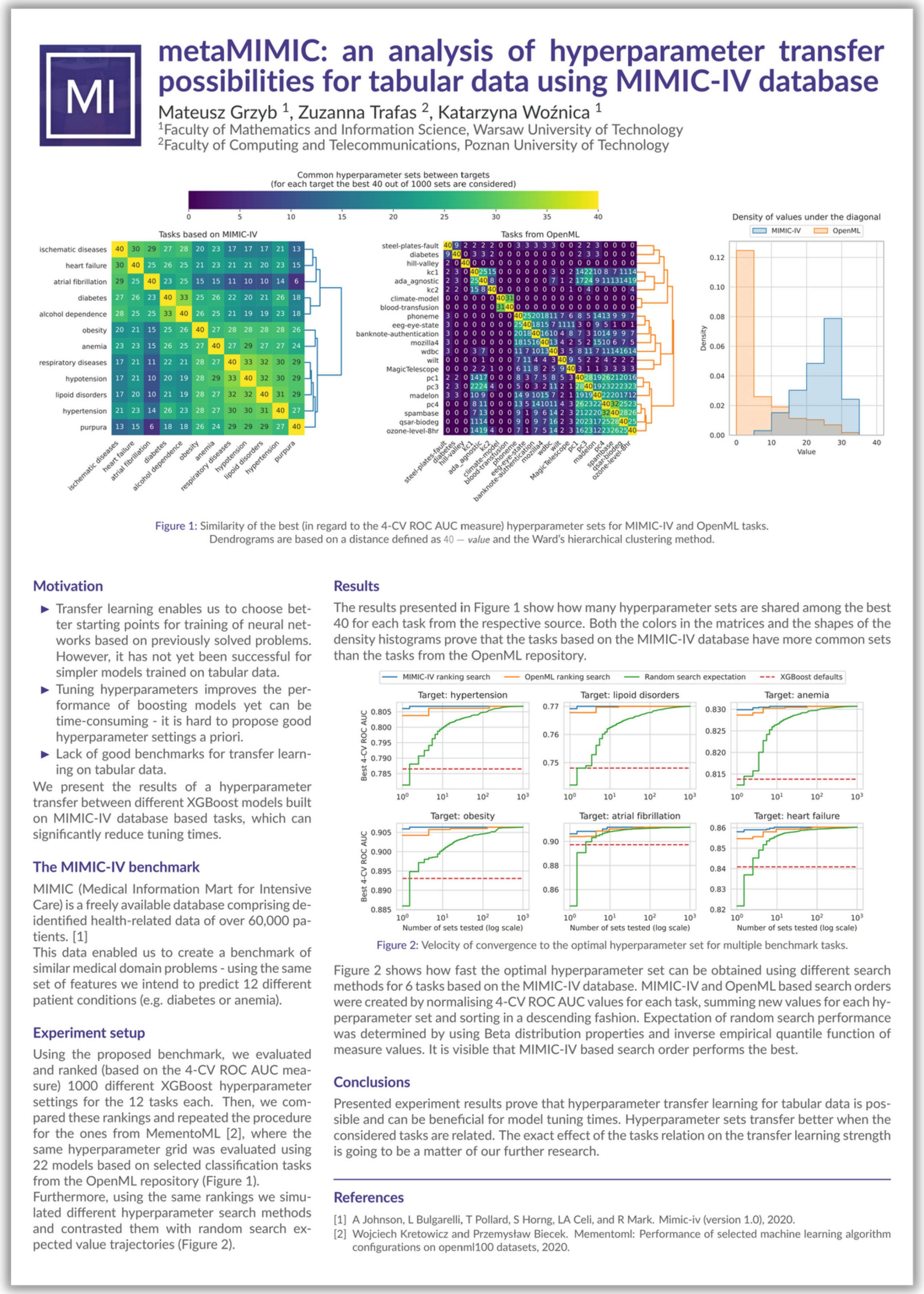
60 hours on 48 cores of bambi!

HOW WE DID IT


Database setup
1
Selecting targets
2
Selecting predictors
3
Running the experiments
4
Results analysis
5

RESULTS?

Come and see at the poster session!



metaMIMIC: an analysis of hyperparameter transfer possibilities for tabular data using MIMIC-IV database

Mateusz Grzyb [1], Zuzanna Trafas [2], Katarzyna Woźnica [1]
[1]Faculty of Mathematics and Information Science, Warsaw University of Technology
[2]Faculty of Computing and Telecommunications, Poznan University of Technology

Figure 1: Similarity of the best (in regard to the 4-CV ROC AUC measure) hyperparameter sets for MIMIC-IV and OpenML tasks. Dendrograms are based on a distance defined as 40 − *value* and the Ward's hierarchical clustering method.

Motivation

- Transfer learning enables us to choose better starting points for training of neural networks based on previously solved problems. However, it has not yet been successful for simpler models trained on tabular data.
- Tuning hyperparameters improves the performance of boosting models yet can be time-consuming - it is hard to propose good hyperparameter settings a priori.
- Lack of good benchmarks for transfer learning on tabular data.

We present the results of a hyperparameter transfer between different XGBoost models built on MIMIC-IV database based tasks, which can significantly reduce tuning times.

The MIMIC-IV benchmark

MIMIC (Medical Information Mart for Intensive Care) is a freely available database comprising deidentified health-related data of over 60,000 patients. [1]

This data enabled us to create a benchmark of similar medical domain problems - using the same set of features we intend to predict 12 different patient conditions (e.g. diabetes or anemia).

Experiment setup

Using the proposed benchmark, we evaluated and ranked (based on the 4-CV ROC AUC measure) 1000 different XGBoost hyperparameter settings for the 12 tasks each. Then, we compared these rankings and repeated the procedure for the ones from MementoML [2], where the same hyperparameter grid was evaluated using 22 models based on selected classification tasks from the OpenML repository (Figure 1).

Furthermore, using the same rankings we simulated different hyperparameter search methods and contrasted them with random search expected value trajectories (Figure 2).

Results

The results presented in Figure 1 show how many hyperparameter sets are shared among the best 40 for each task from the respective source. Both the colors in the matrices and the shapes of the density histograms prove that the tasks based on the MIMIC-IV database have more common sets than the tasks from the OpenML repository.

Figure 2: Velocity of convergence to the optimal hyperparameter set for multiple benchmark tasks.

Figure 2 shows how fast the optimal hyperparameter set can be obtained using different search methods for 6 tasks based on the MIMIC-IV database. MIMIC-IV and OpenML based search orders were created by normalising 4-CV ROC AUC values for each task, summing new values for each hyperparameter set and sorting in a descending fashion. Expectation of random search performance was determined by using Beta distribution properties and inverse empirical quantile function of measure values. It is visible that MIMIC-IV based search order performs the best.

Conclusions

Presented experiment results prove that hyperparameter transfer learning for tabular data is possible and can be beneficial for model tuning times. Hyperparameter sets transfer better when the considered tasks are related. The exact effect of the tasks relation on the transfer learning strength is going to be a matter of our further research.

References

[1] A Johnson, L Bulgarelli, T Pollard, S Horng, LA Celi, and R Mark. Mimic-iv (version 1.0), 2020.
[2] Wojciech Kretowicz and Przemysław Biecek. Mementoml: Performance of selected machine learning algorithm configurations on openml100 datasets, 2020.

Thank you for your attention!