Oboe: Collaborative Filtering for AutoML Model Selection
Chengrun Yang, Yuji Akimoto, Dae Won Kim, Madeleine Udell
Cornell University

What is AutoML?

- An Automated Machine Learning (AutoML) system
- Choose an algorithm together with hyperparameters
- To achieve the best performance on a (supervised learning) task
- Without human intervention.

why AutoML?

- Humans are expensive (especially data scientists)
- Computation is cheap
- Too many models: can’t try them all

to find a reasonable answer, fast, we need:

- Information. What meta-features predict model performance?
- Speed. What meta-features are worth computing?

Our approach

main ideas used by Oboe:

- Algorithm performance is low rank; rank decomposition gives best meta-features
- Use optimal experiment design to cold
- The rest is engineering...

What is AutoML?

at train time (offline stage):

- Given: m training datasets, n machine learning models
- Measure: error of each model on each dataset
- Form: m × n error matrix E (yellow)
- Find: X ∈ R^m×d, Y ∈ R^n×d (orange) for which

at test time (online stage):

- Given: new test dataset = new row of E (blue and white)
- Measure: error of some fast, informative models on new dataset (blue blocks)
- Find: dataset latent features x using least squares
- Compute: model performance (white blocks) as ̂x = 2Y
- Select: models with best predicted performance to use in ensemble

remaining questions: how to choose rank and find fast, informative models

Experimental design finds fast, informative models

Choose a rank you can afford to fit

must run at least k models to fit k-dimensional latent meta-features...

given budget τ for learning on new dataset
initialize rank k = 1, time target t = τ/2
while time remains
- Choose k fast, informative models using experiment design
- Run those models on the dataset and use to infer performance of all models
- Create ensemble using models with predicted best performance
- Double time budget t; increase rank k if meta-CV error improves

Oboe achieves SOTA performance

Experimental setup.

• Datasets: OpenML [8] and UCI [1] datasets with 150–10,000 data points and no missing entries.
• Metric for error matrix: balanced error rate
• Candidate algorithms from python scikit-learn: Adaboost, decision tree, extra trees, random forest, gradient boosting, Gaussian naive Bayes, kNN, logistic regression, multilayer perceptron, perceptron, kernel SVM, linear SVM

Numerical results

- Oboe achieves SOTA performance
- Modeling assumptions are warranted
- Experiment design selects most informative models

Figure 3: Comparison of sampling schemes (QR or ED) in Oboe and PMF. *QR* denotes QR decomposition with column pivoting, *ED* (number) denotes experiment design with number of observed entries constrained. The left plot shows the regret of each AutoML method as a function of number of entries; the right shows the ranking of each AutoML method in the regret plot (1 is best and 5 is worst).