**Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms**

Chris Thornton, Frank Hutter, Holger H. Hoos, Kevin Leyton-Brown

University of British Columbia, Vancouver, Canada

**The 15 Second Version**
- Auto-WEKA can automatically solve your problem of selecting a classification algorithm and hyperparameters that achieve good performance on your dataset.
- You only have to set the time budget/memory limits, and let Auto-WEKA do the rest.
- In experiments on 21 benchmark datasets, Auto-WEKA showed improvements over baselines in 15 cases, reducing error by up to 5.5%.
- Get it at http://www.cs.ubc.ca/~alois/beta/Projects/autoweka/

**Auto-WEKA**
- Optimizes over the 39 classifiers in WEKA (27 base, 10 meta and 2 ensemble methods), in addition to feature/attribute selection.
- Search space has 796 parameters.
- Optimizes 10-fold CV error on training data using any general purpose algorithm configuration method.
- We use two existing Sequential Model Based Optimization (SMBO) methods.

**Algorithm/Hyperparameter Search Space**

**Feature/Attribute Selection Search Space**

**Results**

**10-Fold CV (Training) Performance**

- Best Baseline vs. Auto-WEKA
- Best Baseline (Up to 1.0 CPU Year Budget) vs. Auto-WEKA (180 CPU Hour Budget)
- Auto-WEKA SMAC vs. TPE

**Generalization (Testing) Performance**

- Best Baseline vs. Auto-WEKA
- Best Baseline (Up to 1.0 CPU Year Budget) vs. Auto-WEKA (180 CPU Hour Budget)
- Auto-WEKA SMAC vs. TPE

**Selected Algorithms**

- Frequency of Selected Base, Meta and Ensemble Algorithms
- Frequency of Base Algorithms Selected in Top Meta Algorithms
- Frequency of Feature Selection Algorithms

**Detailed Results on Largest Datasets**

**Baselines and SMBO Methods**

**Exhaustive Default (Ex-Def)**
- Use default hyperparameters on all WEKA classifiers, pick the best.

**Grid Search**
- Discretize hyperparameters of base classifiers.
- Use the 10-fold CV performance to select the classifier and hyperparameters.
- High computational cost - an average of 6 CPU months.

**Random Search**
- Select classifiers/hyperparameters at random.
- Shown to be competitive with grid search.

[Bergstra, Bengio, 2011]

**Tree-structured Parzen Estimator (TPE)**
- 1-D Parzen estimator for each hyperparameter - models "good" vs "bad".
- Tree structure of estimators enforces hyperparameter conditioning.

[Bergstra, Bardent, Bengio, Käßl, 2011]

**Sequential Model-based Algorithm Configuration (SMAC)**
- Builds a single regression forest model of performance.
- Alternates between random/model chosen hyperparameters.
- Only computes the performance of a single CV fold at a time - quickly rejects poor hyperparameters.

[Hutter, Hoos, Leyton-Brown, 2011]

**Future Work**
- More sophisticated use of the given data to exploit the performance gains on training data.
- Modify the SMBO methods to increase learned knowledge about base algorithm hyperparameters.