Automated Machine Learning for Soft Voting in an Ensemble of Tree-based Classifiers

Jungtaek Kim and Seungjin Choi

Machine Learning Group,
Department of Computer Science and Engineering, POSTECH,
77 Cheongam-ro, Nam-gu, Pohang 37673,
Gyeongsangbuk-do, Republic of Korea

June 6, 2018

Table of Contents

Automated Machine Learning

- Attempt to find automatically the optimal machine learning model without human intervention.
- Usually include feature transformation, algorithm selection, and hyperparameter optimization.
- Given a training dataset $\mathcal{D}_{\text{train}}$ and a validation dataset $\mathcal{D}_{\text{val}}$, the optimal hyperparameter vector $\boldsymbol{\lambda}^*$ for an automated machine learning system:

$$\boldsymbol{\lambda}^* = \text{AutoML}(\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}, \Lambda)$$

where AutoML is an automated machine learning system and $\lambda \in \Lambda$.

Background: Soft Majority Voting

- An ensemble method to construct a classifier using a majority vote of k base classifiers.
- Class assignment of soft majority voting classifier:

$$c_i = \arg\max \sum_{j=1}^{k} w_j \mathbf{p}_i^{(j)}$$

for $1 \leq i \leq n$ where n is the number of instances, arg max returns an index of maximum value in given vector, $w_j \in \mathbb{R} \geq 0$ is a weight of base classifier j, and $\mathbf{p}_i^{(j)}$ is a class probability vector of base classifier j.

Background: Bayesian Optimization

- A useful method to find global minimum or maximum for black-box function.
- Improve the current solution as iterating the following steps:
  1. modeling a surrogate function,
  2. acquiring next point that has maximum of acquisition function.
- In our system, use Gaussian process (GP) regression as surrogate function and GP-UCB as acquisition function.
- GP-UCB for a minimization case:

$$a_{\mathrm{UCB}}(\mathbf{x}) = -\mu(\mathbf{x}) + \kappa\sigma(\mathbf{x})$$

where $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are posterior mean and posterior standard deviation functions from GP regression. κ is a balancing hyperparameter for exploitation and exploration.

Our Automated Machine Learning System, *mlg.postech*

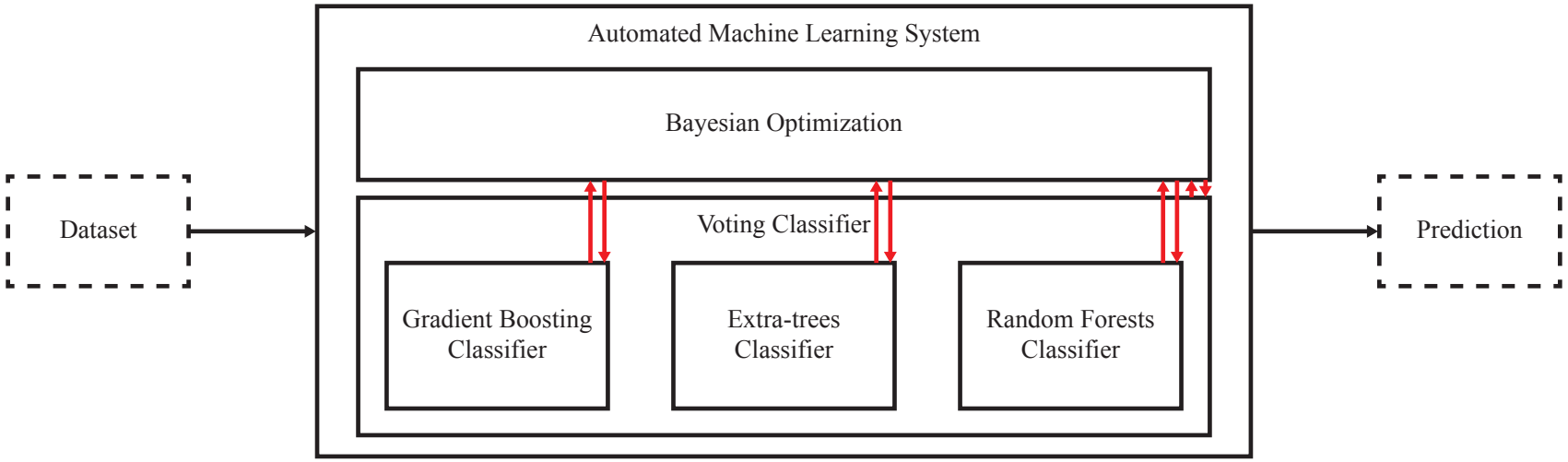

Figure 1: Our automated machine learning system, *mlg.postech*. Voting classifier constructed by three tree-based classifiers: gradient boosting, extra-trees, and random forests classifiers produces predictions, where voting classifier and tree-based classifiers are iteratively optimized by Bayesian optimization for the given time budget.

Our Automated Machine Learning System, *mlg.postech*

- Written in `Python`.
- Use `scikit-learn` and our own Bayesian optimization package.
- Split training dataset to training (0.6) and validation (0.4) sets for Bayesian optimization.
- Optimize six hyperparameters:
  1. extra-trees classifier weight/gradient boosting classifier weight for voting classifier,
  2. random forests classifier weight/gradient boosting classifier weight for voting classifier,
  3. the number of estimators for gradient boosting classifier,
  4. the number of estimators for extra-trees classifier,
  5. the number of estimators for random forests classifier,
  6. maximum depth of gradient boosting classifier.
- Use GP-UCB.

AutoML Challenge 2018

Place	Team	Set 1	Set 2	Set 3	Set 4	Set 5	Average
1	aad_freiburg	0.5533 (3)	0.2839 (4)	0.3932 (1)	0.2635 (1)	0.6766 (5)	2.8
2	**mlg.postech**	**0.5418 (5)**	**0.2894 (2)**	**0.3665 (2)**	**0.2005 (9)**	**0.6922 (1)**	**3.8**
	wlWangl	0.5655 (2)	0.4851 (1)	0.2829 (5)	-0.0886 (16)	0.6840 (3)	5.4
3	thanhdng	0.5131 (6)	0.2256 (8)	0.2605 (7)	0.2603 (2)	0.6777 (4)	5.4
	Malik	0.5085 (7)	0.2297 (7)	0.2670 (6)	0.2413 (5)	0.6853 (2)	5.4

Figure 2: AutoML Challenge 2018 result. A normalized area under the ROC curve (AUC) score (upper cell in each row) is computed for each dataset, and a dataset rank (lower cell in each row) is determined by numerical order of the normalized AUC score. Finally, an overall rank is determined by the average rank of five datasets.