A direct formulation for totally-corrective multi-class boosting

Chunhua Shen, and Zhihui Hao

Problem

Multi-class boosting is important in that most classification problems have multiple classes. So far most boosting algorithms are designed for binary classification. Output codes provide a simple and intuitive solution to multi-class classification, they completely ignore the pairwise correlation information between classes.

We proffer a direct approach for learning multi-class boosting. The key idea of our approach is that, given an example $\{\boldsymbol{x}, y\}$, the decision function with the correct label $F_y(\boldsymbol{x})$ must be larger than the decision function's value with an incorrect label $F_r(\boldsymbol{x}), \forall r \neq y$.

Contributions

Our main contributions are as follows.

- We propose a novel direct approach to multi-class boosting formulation based on the generalization of the conventional "margin" in binary classification.
- The proposed boosting is totally corrective in the sense that all the coefficients of the learned weak classifiers are updated at each iteration.

The proposed formulation may also be applicable to other structured prediction problems.

Main idea

For a training example $(\boldsymbol{x}, y)$, if we have a perfect classification rule, then the following holds

$$F_y(\mathbf{x}) > F_r(\mathbf{x}), \text{ for any } r \neq y.$$

In the large margin framework with hinge loss, ideally

$$F_y(\mathbf{x}) \geq 1 + F_r(\mathbf{x}), \text{ for any } r \neq y.$$

The primal problem that we want to optimize can be written as

$$\min_{W, \boldsymbol{\xi}} \textstyle\sum_{i=1}^m \xi_i + \nu \left\| W \right\|_1 \text{ s.t. } W \geq 0.$$

$$\delta_{r,y_i} + H_{i:} \boldsymbol{w}_{y_i} \geq 1 + H_{i:} \boldsymbol{w}_r - \xi_i, \forall i, r.$$

Column generation

We derive the Lagrange dual of the primal problem:

$$\begin{aligned} \min_U \; & \sum_{r=1}^k \sum_{i=1}^m \delta_{r,y_i} U_{ir} \\ \text{s.t. } & \sum_i \left(\delta_{r,y_i} - U_{ir}\right) H_{i:} \leq \nu \mathbf{1}^\top, \forall r, \\ & \sum_r U_{ir} = 1, \forall i; \; U \geq 0. \end{aligned}$$

Each row of the matrix U is normalized. The first set of constraints can be infinitely many:

$$\sum_i \left(\delta_{r,y_i} - U_{ir}\right) h(\boldsymbol{x}_i) \leq \nu, \forall r, \text{ and } \forall h(\cdot) \in \mathcal{H}. \tag{1}$$

We can now use column generation to solve the problem, similar to the LPBoost. The subproblem for generating weak classifiers is

$$h^*(\cdot) = \operatorname*{argmax}_{h(\cdot), r} \sum_{i=1}^m \left(\delta_{r,y_i} - U_{ir}\right) h(\boldsymbol{x}_i). \tag{2}$$

Algorithm 1: MultiBoost **with the hinge loss**

Initialize each entry of U to be $1/k$.

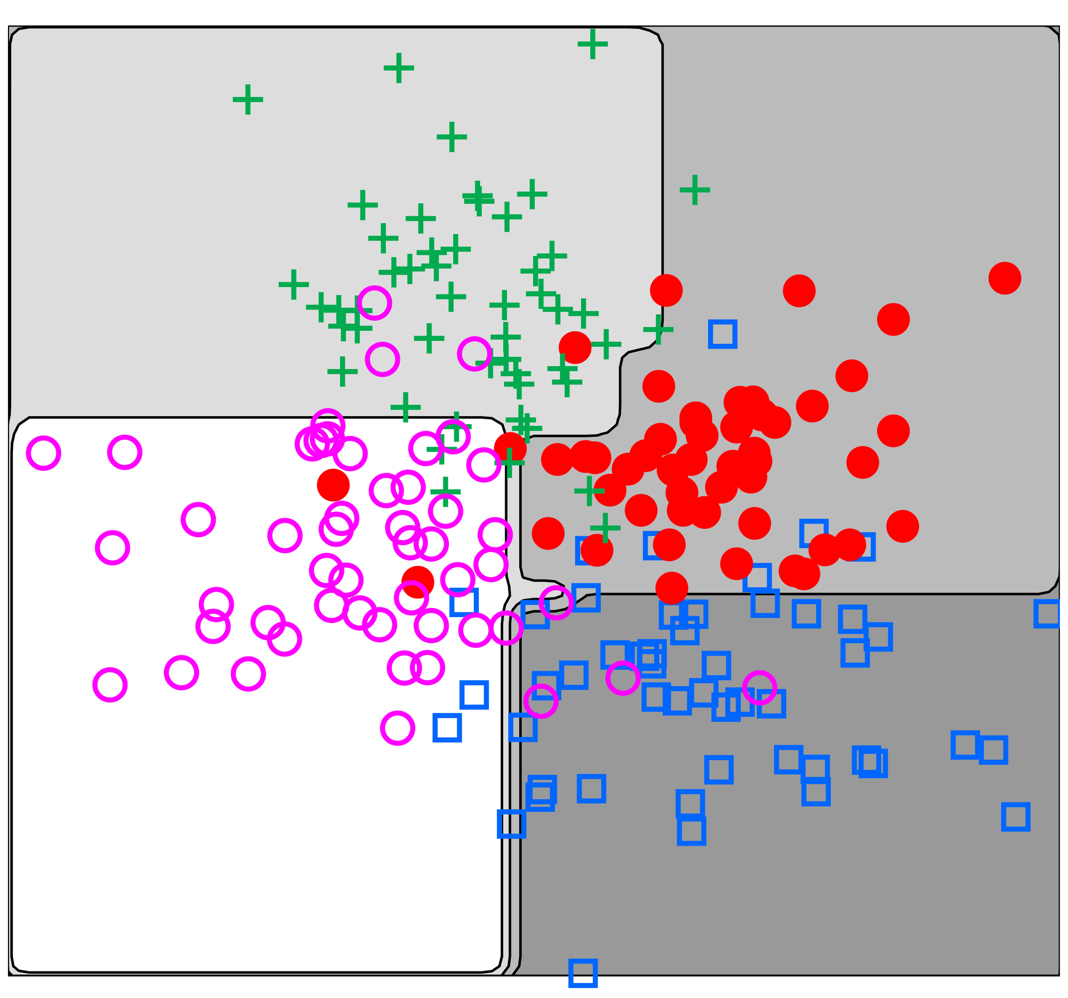
loop

– Find the weak classifier by solving the subproblem, and add this weak classifier to the primal problem.

– Solve the primal problem using a primal-dual interior-point LP solver, such that the dual solution is also available.

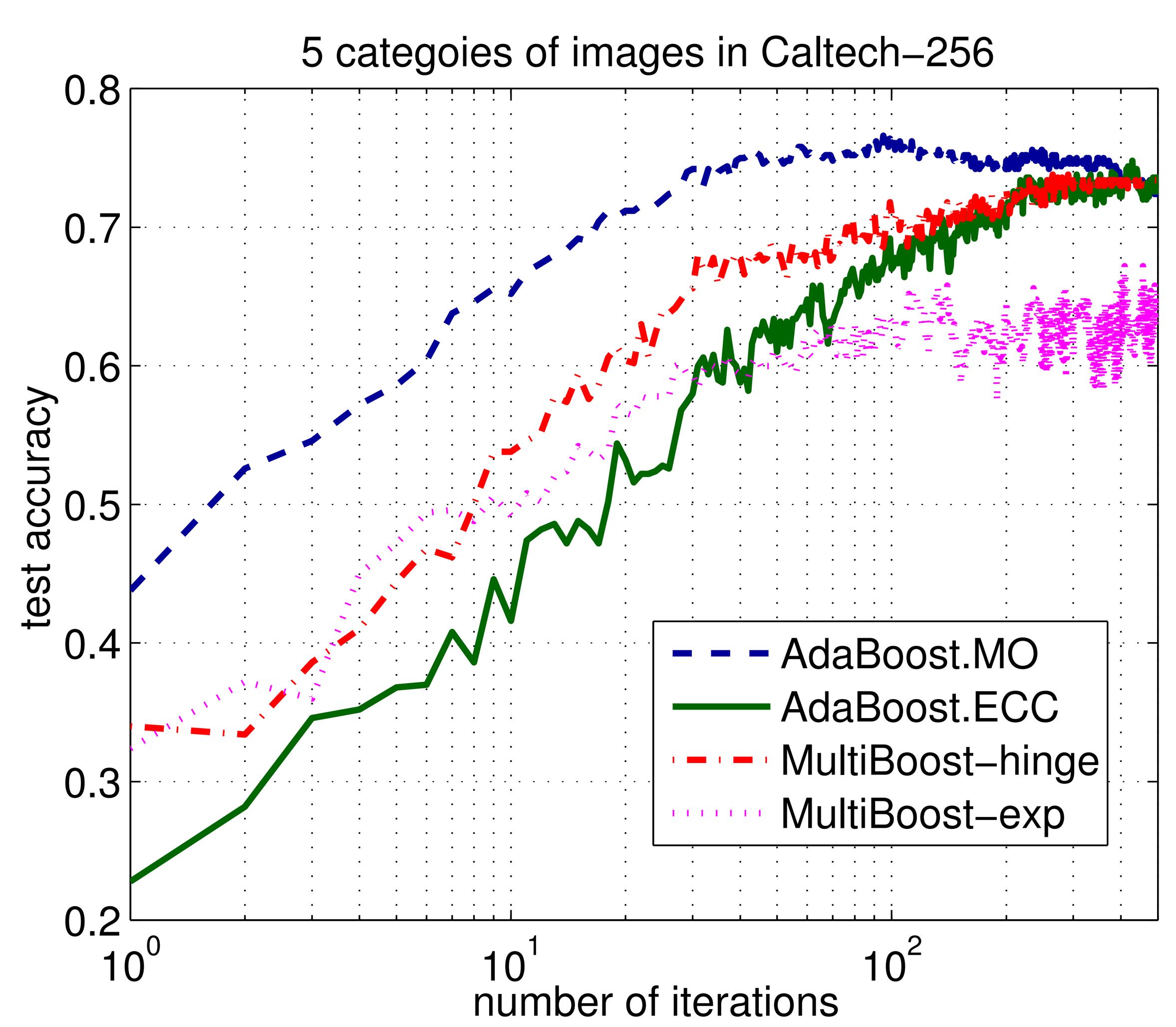
until convergence

Decision boundary



Decision boundary on a toy data.

Accuracy



Test accuracy on the caltech 256 data.

Conclusion

We have presented a direct formulation for multi-class boosting. We derive the Lagrange dual of the formulated primal optimization problem. Based on the dual problem, we are able to design totally-corrective boosting using the column generation technique. At each iteration, all weak classifiers' weights are updated.

Future research topics include how to efficiently solve the convex optimization problems of the proposed multi-class boosting. We also want to explore structural learning with boosting.