

# A Combination of Ensemble Methods for Large-Scale Regression

Hans Hang

Universiteit Twente, Department of Applied Mathematics  
Enschede, The Netherlands

We adopt a combination of two ensemble methods, i.e. gradient boosting and base learner ensemble, to solve the problem of large-scale regression. To be specific, we propose an algorithm called Gradient Boosted Binary Histogram Ensemble (GBBHE), where we use the ensemble of binary histogram regressors as the base learner in boosting. From the theoretical perspective, by assuming the Hölder continuity of the target function, we establish the statistical convergence rate of GBBHE in the space  $C^{0,\alpha}$  and  $C^{1,0}$ , where a lower bound of the convergence rate for the base learner demonstrates the advantage of boosting. Moreover, in the space  $C^{1,0}$ , we prove that the number of iterations to achieve the fast convergence rate can be reduced, which improves the computational efficiency. In the experiments, compared with other state-of-the-art algorithms such as gradient boosted regression tree (GBRT), Breiman's forest, and kernel-based methods, our GBBHE algorithm shows promising performance with less running time on large-scale datasets.