


Meng-long LU, Lin-bo QIAO, Da-wei FENG, Dong-sheng LI, Xi-cheng LU, 2019.
Mini-batch cutting plane method for regularized risk minimization. *Frontiers of Information Technology & Electronic Engineering*, 20(11):1551-1563.
<https://doi.org/10.1631/FITEE.1800596>

Mini-batch cutting plane method for regularized risk minimization

Key words: Machine learning; Optimization methods; Gradient methods; Cutting plane method

Corresponding author: Da-wei FENG

E-mail: davyfeng.c@gmail.com

 ORCID: Meng-long LU, <http://orcid.org/0000-0003-0011-0368>;

Da-wei FENG, <http://orcid.org/0000-0002-7587-8905>

Motivation

- Considering interpretability, a machine learning application is preferred to be formulated as a convex optimization problem, especially when the training and inference are performed on structured data.
- The non-differentiable convex problem cannot be solved easily using the smooth optimization method.
- The cutting plane method requires access to the entire dataset for each iteration, which renders them infeasible for efficiently handling large datasets.

Main idea

- The proposed algorithm iterates with the estimated cutting planes being calculated on a small batch of sampled data, and is capable of handling large-scale problems.
- The proposed algorithm adopts a “sink” operation that detects and adjusts noisy estimations to guarantee convergence.

Method

- MBCPM randomly selects instances from the entire training set and generates an estimated cutting plane on this sampled sub-dataset.
- MBCPM evaluates whether these estimated cutting planes affect the correctness of the algorithm.
- MBCPM will adopt a “sink” operation to address noisy estimations.

Major results

The objective value of MBCPM declines faster than that of the traditional cutting plane method.

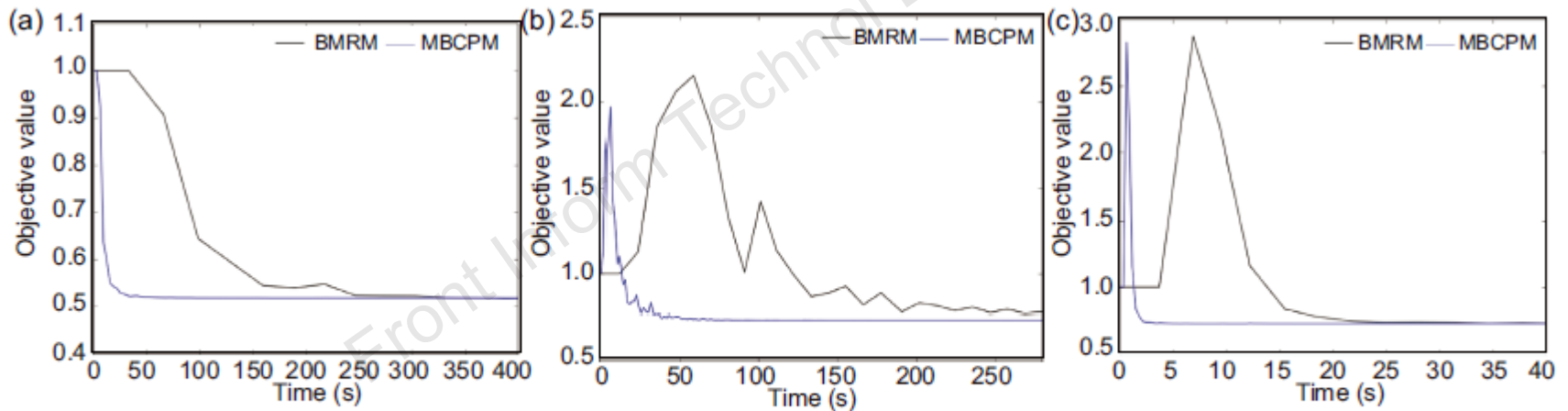


Fig. 4 Objective value of MBCPM and BMRM as a function of time consumption: (a) HUMAN; (b) MNIST; (c) ACOUSTIC

Major results

The objective value of MBCPM declines faster than that of the SGD method.

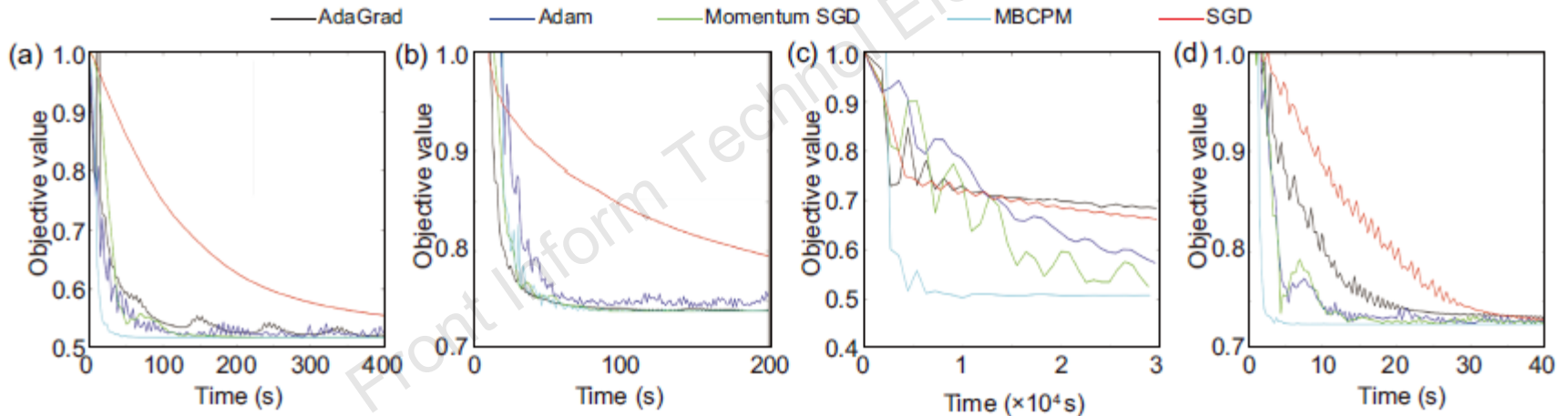


Fig. 6 Objective value as a function of time cost in the comparison of MBCPM with SGD methods: (a) HUMAN; (b) MNIST; (c) URL; (d) ACOUSTIC

Major results

The error rate of MBCPM declines faster than that of SGD methods and converges more stably.

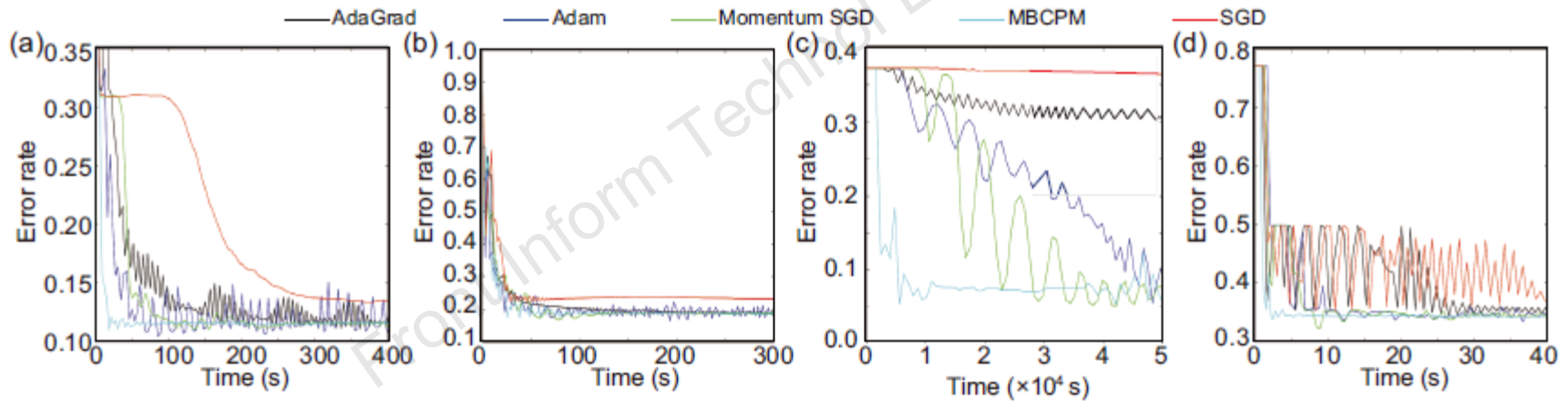


Fig. 7 Error rate as a function of time cost in the comparison of MBCPM with SGD methods: (a) HUMAN; (b) MNIST; (c) URL; (d) ACOUSTIC

Conclusions

- MBCPM iterates with the estimated cutting planes being calculated on a small batch of sampled data, is capable of handling large-scale problems, and adopts a “sink” operation that detects and adjusts noisy estimations to guarantee convergence.
- Experimental results on extensive real-world datasets indicated that under the general convex settings, MBCPM is superior to some widely used SGD methods and the conventional CPM method in terms of convergence speed.