

Evolving Gradient Boost: A Pruning Scheme Based on Loss Improvement Ratio for Learning under Concept Drift

Kun Wang, Jie Lu, Anjin Liu, Guangquan Zhang, Li Xiong

Abstract: In non-stationary environments, data distributions can change over time. This phenomenon is known as concept drift, and the related models need to adapt if they are to remain accurate. With gradient boosting ensemble models, selecting which weak learners to keep/prune to maintain model accuracy under concept drift is non-trivial research. Unlike existing models such as AdaBoost, which can directly compare weak learners' performance by their accuracy (a metric between $[0, 1]$), in gradient boosting, weak learners' performance is measured with different scales. To address the performance measurement scaling issue, we propose a novel criterion to evaluate weak learners in gradient boosting models, called loss improvement ratio (LIR). Based on LIR, we develop two pruning strategies: (1) naive pruning (NP) which simply deletes all learners with increasing loss; (2) statistical pruning (SP) which removes learners if their loss increase meets a significance threshold. We also devise a scheme to dynamically switch between NP and SP to achieve the best performance. We implement the scheme as a concept drift learning algorithm, called evolving gradient boost (LIR-eGB). On average, LIR-eGB delivered the best performance against state-of-the-art methods on both stationary and non-stationary data.

Kun Wang, Jie Lu, Anjin Liu, and Guangquan Zhang are with the Decision Systems and e-Service Intelligence Laboratory (DESI), Australian Artificial Intelligence Institute (AAIL)

Funding Source Acknowledgement: Australian Research Council (ARC) under discovery project DP190101733

Publication: IEEE Transactions on Cybernetics (Volume: 53, Issue: 4, Pages: 2110-2123, 2023, DOI: 10.1109/TCYB.2021.3109796)