

Supplementary material

October 2, 2020

Similar experiments as the ones described in section 6 but performed on the 45 datasets benchmark proposed by (Mori et al 2017).

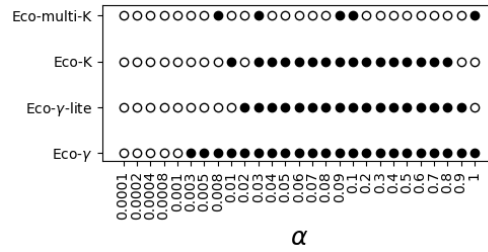


Figure 1: Success of adapting the trigger times - Wilcoxon signed-rank test results for different values of α : black dots indicate success and circles failures.

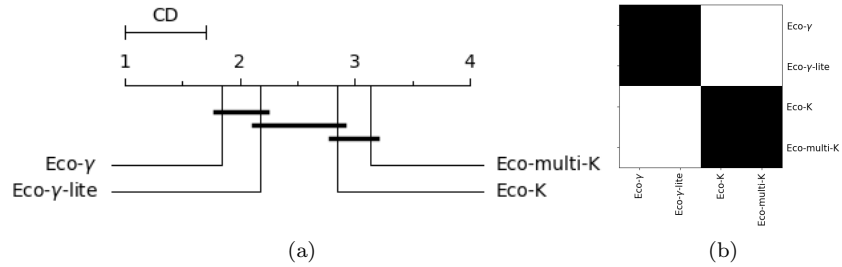


Figure 2: Evaluation based on *AvgCost*: (a) Nemenyi test applied to the 45 datasets; (b) pairwise comparison using the Wilcoxon signed-rank test, with black squares identifying non-significant comparisons.

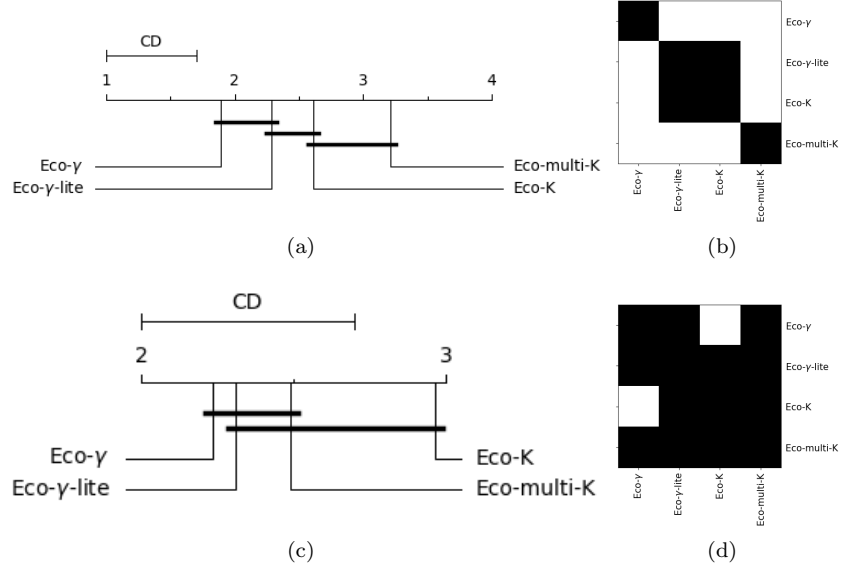


Figure 3: Earliness (a, b) and predictive performance (c, d) comparison of the ECONOMY approaches.

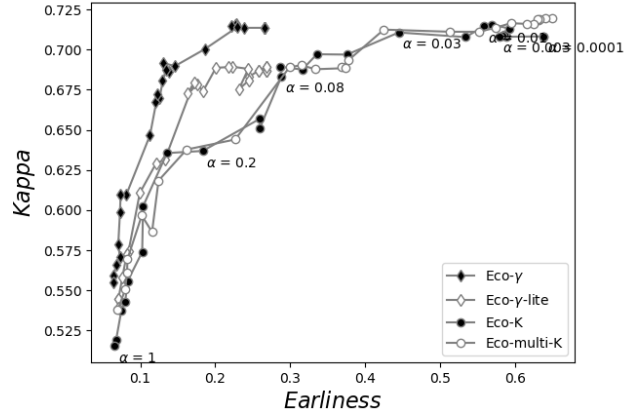


Figure 4: Average *Earliness* vs. Average *Kappa* score obtain over the 45 datasets by varying the slope of the time cost, such as $\alpha \in [10^{-4}, 1]$.

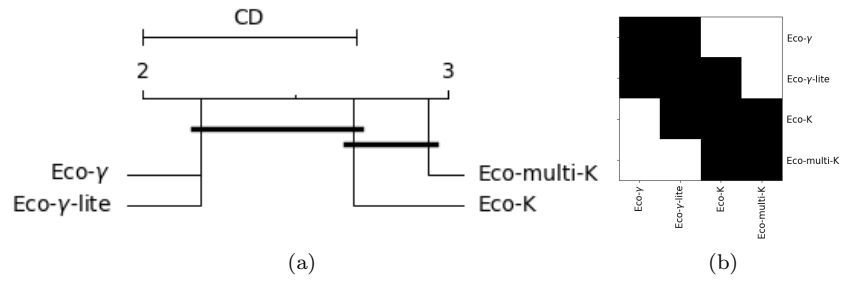


Figure 5: Evaluation of the quality of online decisions based on Δ_{cost} .