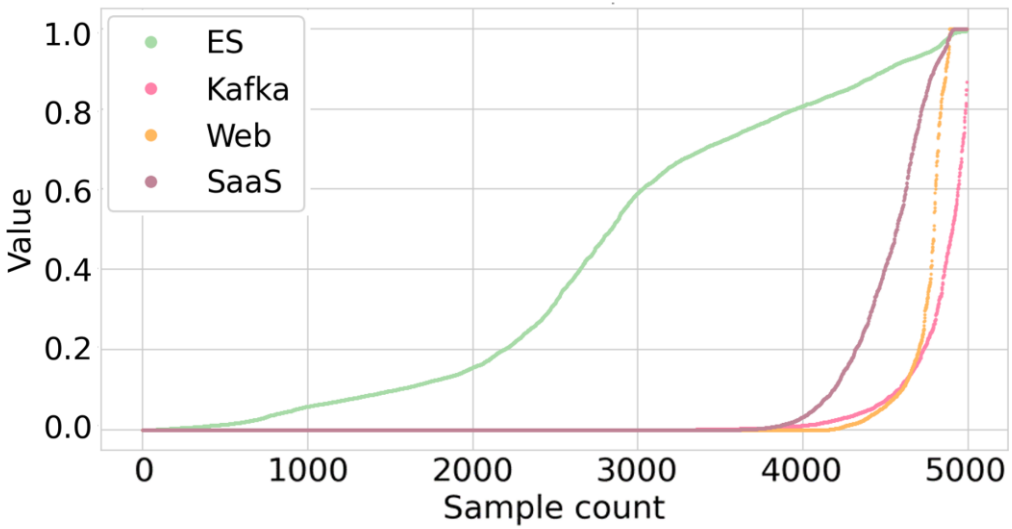
Superior F1-score: I/O Feature Driven Algorithms for Stream Computing Systems Workload Identification

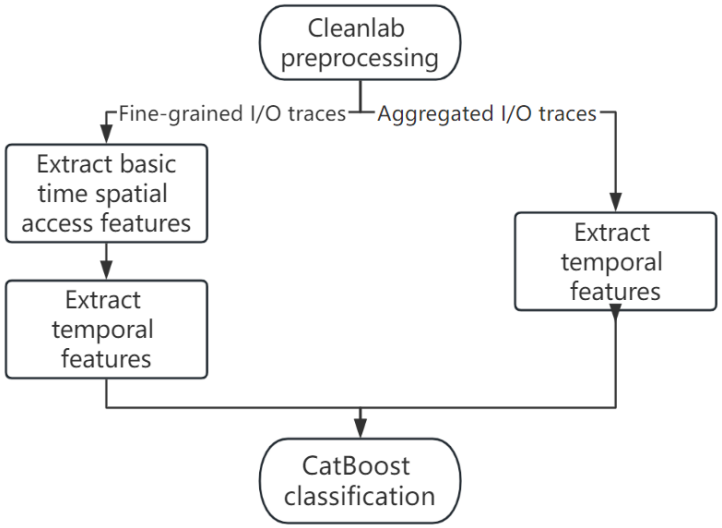
Yuxiao HAN, Yubo LIU, Ziyan ZHANG, Fei LI, Zhiguang CHEN, Nong XIAO

Frontiers of Computer Science, DOI: 10.1007/s11704-024-40710-5

Problems & Ideas

- Problems of conventional workload identification approaches:
  - Limited features for workload identification with low effectiveness.
  - Existing methods show low classification F1-score in diverse workload types and complex scenarios.
- Ideas:
  - Analyze and extract highly discriminative features.
  - Design tailored workload identification algorithms for varying requirements

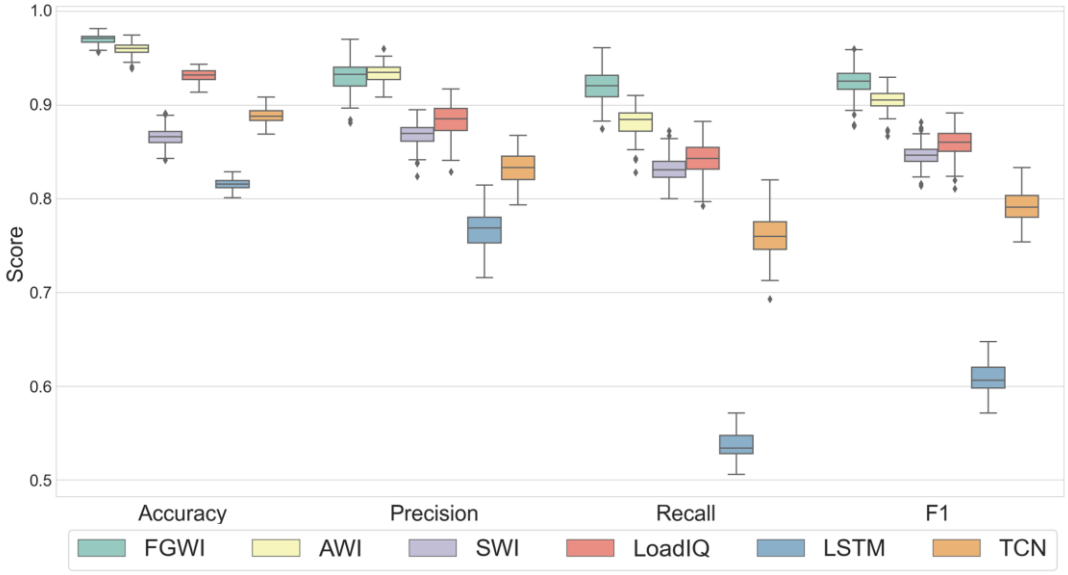


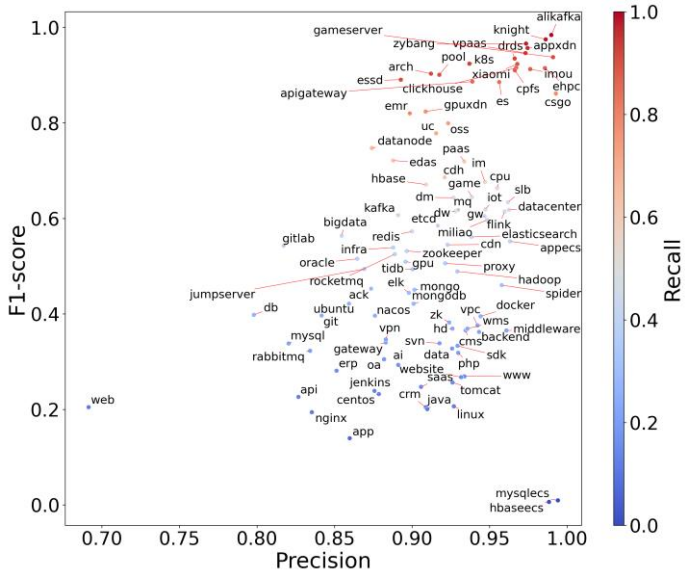


Left: Statistical results reveal significant differences in read request ratios across various workloads;
Right: Workload classification utilizing different feature extraction methods based on diverse I/O traces.

Main Contributions

- Contributions:
  - Three novel algorithm design approaches:
    - Separate read and write traces for feature extraction.
    - Map offsets to data blocks to enhance feature effectiveness.
    - Extract higher differentiating time and space features.
  - Two optimized workload identification algorithms:
    - FGWI: Focuses on high F1-score requirements, using fine-grained I/O traces to analyze basic, time, spatial, and temporal access features.
    - AWI: Designed for low overhead, utilizing minute-level aggregated I/O traces with an emphasis on temporal access features.





Workload identification results. Left: Box plots showing classification results for various algorithms with a training-to-testing ratio of 5:1; Right: Cross-disk classification results for AWI across 100 different workloads.