Cross-dataset Time Series Anomaly Detection for Cloud Systems

Microsoft Research

Microsoft Azure

Nanjing University

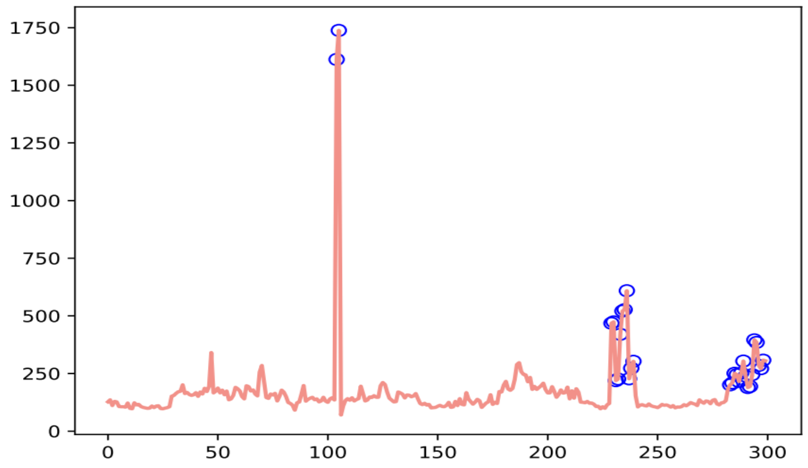
Introduction

Background: As the cloud systems could be used by millions of users around the world on a 24/7 basis, high service reliability and availability are critical.

Goal: accurate anomaly detection with low labeling cost against large-scale cloud monitoring time series data (KPIs)

Key Performance Indicator (KPI)

- CPU utilization
- Memory utilization
- Network traffic rate
- VM downtime
- ……



Background

Challenges

- Diverse characteristics of anomalies in cloud systems
- Unsatisfactory performance of unsupervised learning
- High labeling cost for supervised learning methods

ATAD

- Transfer Learning: enabling cross-dataset anomaly detection
- Active Learning: further improving detection accuracy

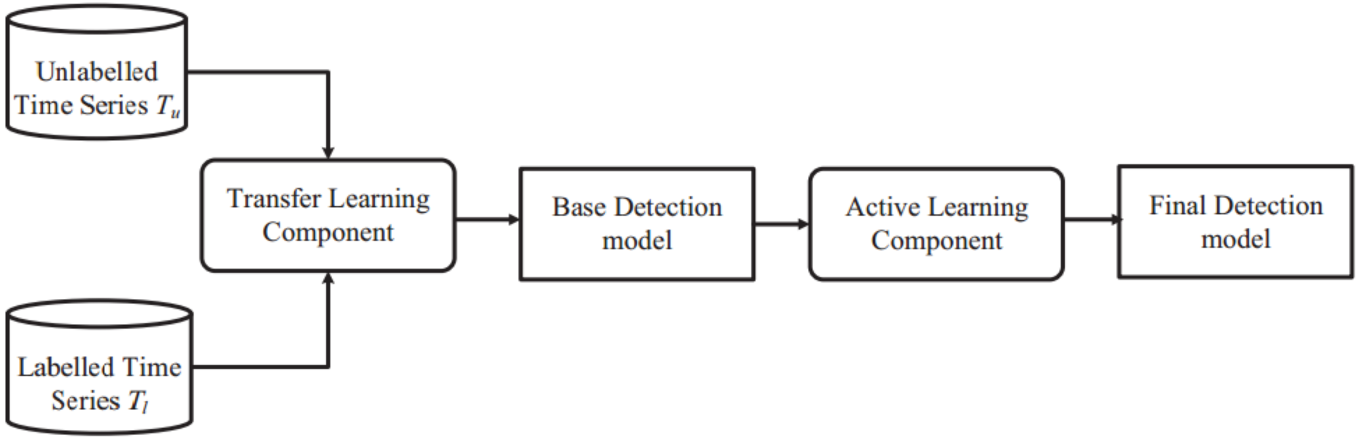


Figure 1: The overall workflow of ATAD

Transfer Learning Component

- Feature Identification
- The Transfer between Source Domain and Target Domain

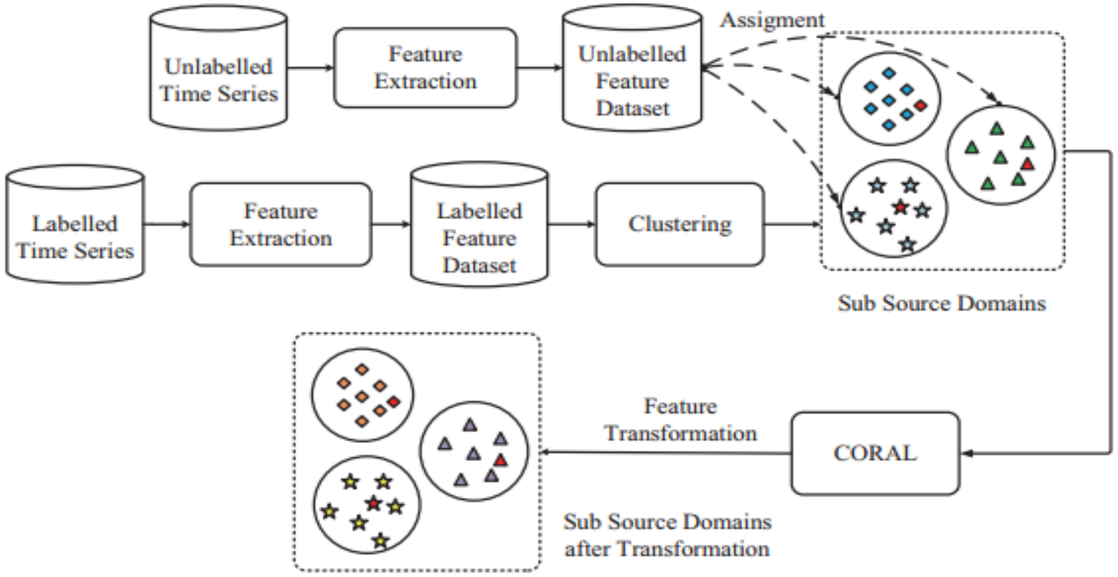
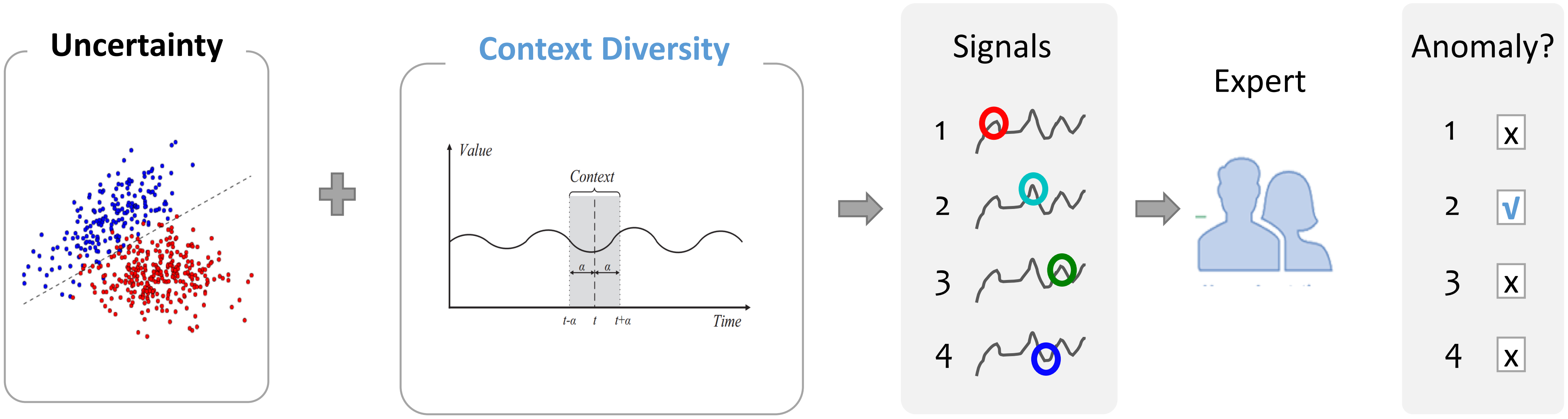


Figure 2: Transfer Learning Component

Active Learning Component

- **Uncertainty**
- **Context Diversity**



*Only labeling **0.1%** achieve good result*

Summary

ATAD for cloud service systems

- High detection accuracy
- Low labelling cost

Table 11: Experimental result on IOPS dataset of Microsoft

	Precision	Recall	F1-Score
iForest	0.2886	0.3988	0.3349
K-Sigma	0.8170	0.1882	0.3059
S-H-ESD	0.9117	0.1741	0.2924
RF	0.5213	0.6724	0.5873
ATAD	0.8082	0.6188	**0.7009**

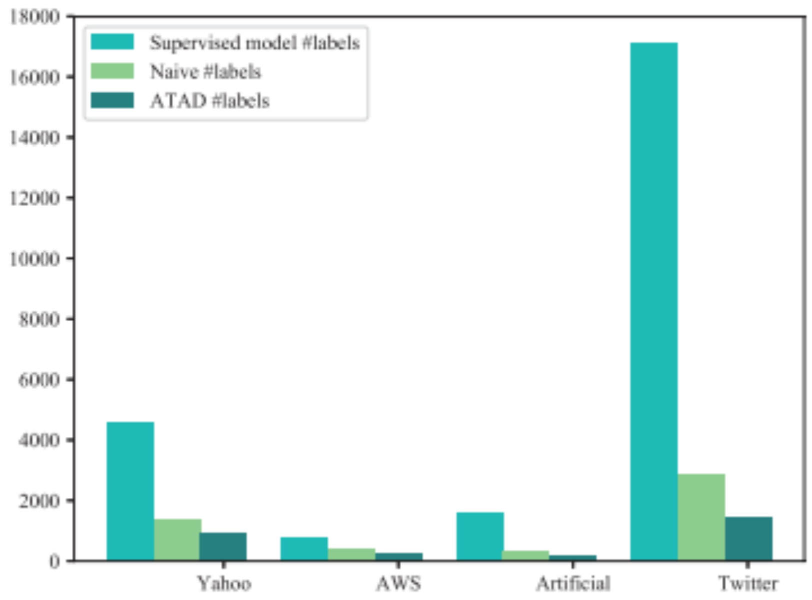


Figure 4: The number of labels required by Supervised Model, Naive Active Learning without transfer learning and ATAD