Robust KPI Anomaly Detection for Large-Scale Software Services with Partial Labels

Shenglin Zhang[1], Chenyu Zhao[1], Yicheng Sui[1], Ya Su[2],
Yongqian Sun[1], Yuzhi Zhang[1], Dan Pei[2], Yizhe Wang

南開大學 Nankai University [1]

清華大學 Tsinghua University [2]

Outline

Introduction

Challenges

Contribution

Framework

Evaluation

Outline

Introduction

Challenges

Contribution

Framework

Evaluation

Software services

Online Shopping: amazon.com 淘宝网 Taobao.com

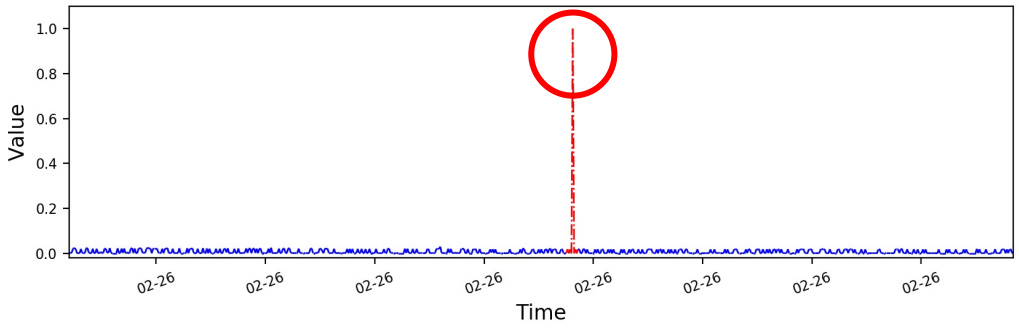
Social Networks: WeChat facebook

Search Engines: Baidu 百度 Google

Millions of KPIs are constantly monitored and collected

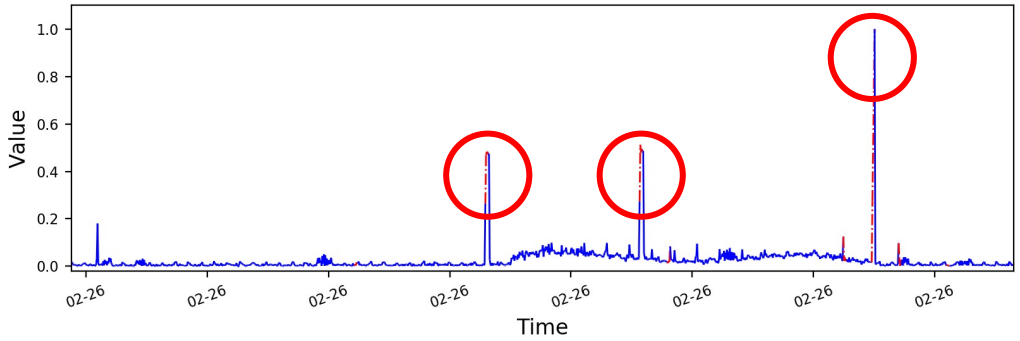
Key Performance Indicator (KPI):

- User perceived metrics: response delay, queries per second, failure ratio...



The metric of average response delay

- System-level metrics: CPU utilization, memory utilization, network throughput...



The metric of CPU utilization

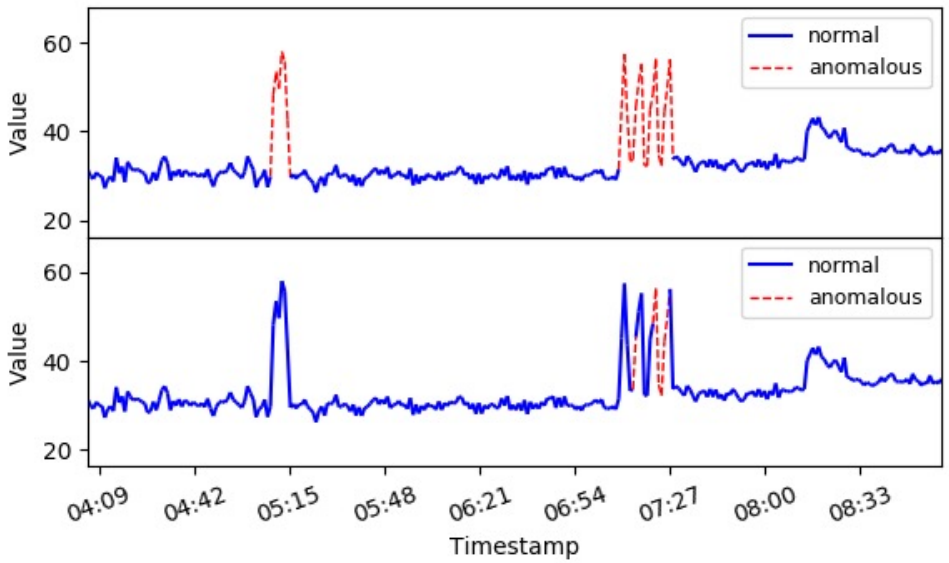
Existing KPI anomaly detection algorithms

Type	Name	Labeling effort	Shortcoming
Supervised	Opprentice	Full labels	Time-consuming and labor-intensive labeling work
	EGADS		
Unsupervised	Donut	No labels	Low accuracy or require large amounts of training data
	iForest		
Semi-supervised	ADS	Full labels of cluster centroids	Need a large number of KPI streams with high-quality ground truth
Transfer learning	ATAD	For a KPI dataset with 10 million data points, about 500,000 labels are needed	

Comparison between PU learning and Semi-supervised learning

Semi-supervised learning

Positive-Unlabeled learning (PU learning)



For a KPI stream in the training set:

- PU learning only requires labeling part of anomalous segments.
- Semi-supervised learning need to label all the anomalous segments.
- PU learning greatly reduces the labeling effort.

Outline

Introduction

Challenges

Contribution

Framework

Evaluation

Challenge 1: Large-scale and diverse KPI streams

KPI streams are large in number and diverse in pattern.

- Train a PU learning model for each KPI stream?

Challenge 1: Large-scale and diverse KPI streams

KPI streams are large in number and diverse in pattern.

- Train a PU learning model for each KPI stream?

  Too many manual labels are needed.

Challenge 1: Large-scale and diverse KPI streams

KPI streams are large in number and diverse in pattern.

- Train a PU learning model for each KPI stream?

  Too many manual labels are needed.

- Train a universal PU learning model for all KPI streams?

Challenge 1: Large-scale and diverse KPI streams

KPI streams are large in number and diverse in pattern.

- Train a PU learning model for each KPI stream?

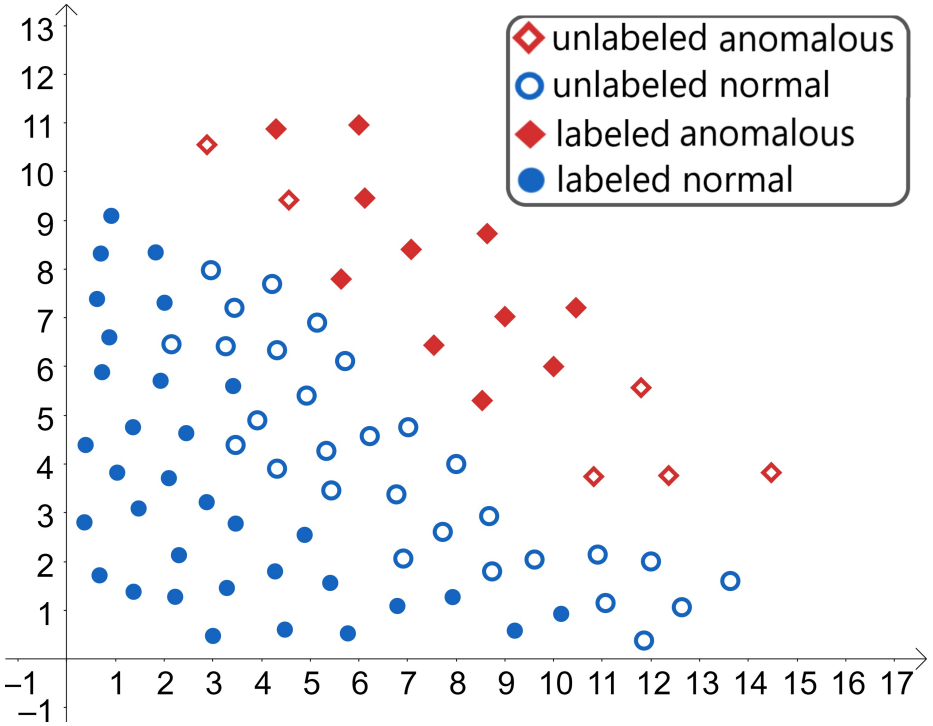
  Too many manual labels are needed.

- Train a universal PU learning model for all KPI streams?

  The model will suffer from low accuracy and is difficult.

Challenge 2: Active learning strategy

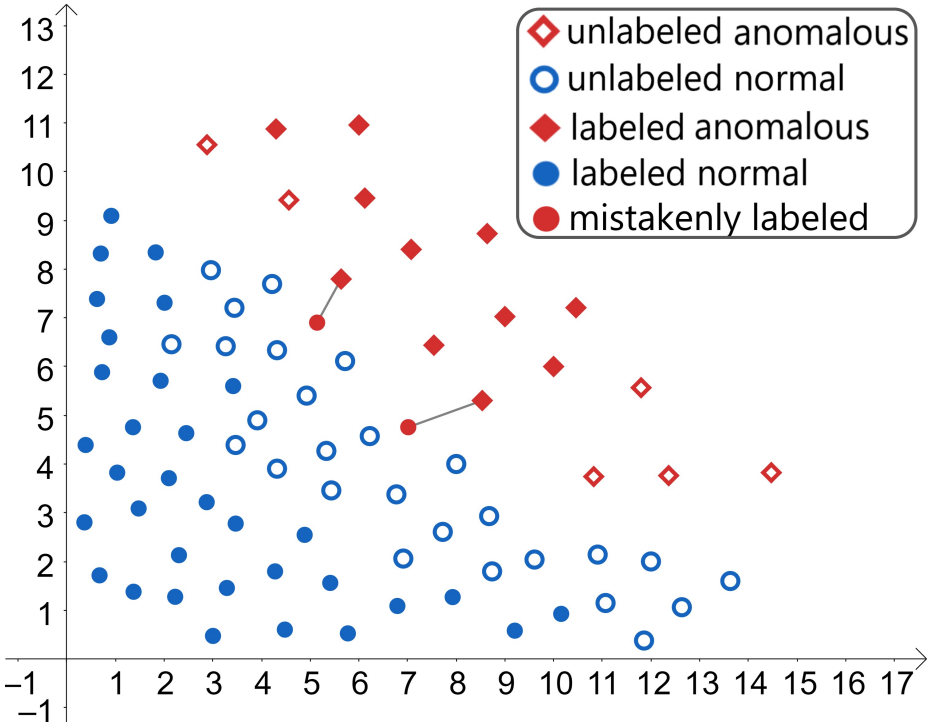
Active learning-assisted PU learning can improve its performance.
However, it will produce many false positives.



Step 1

Challenge 2: Active learning strategy

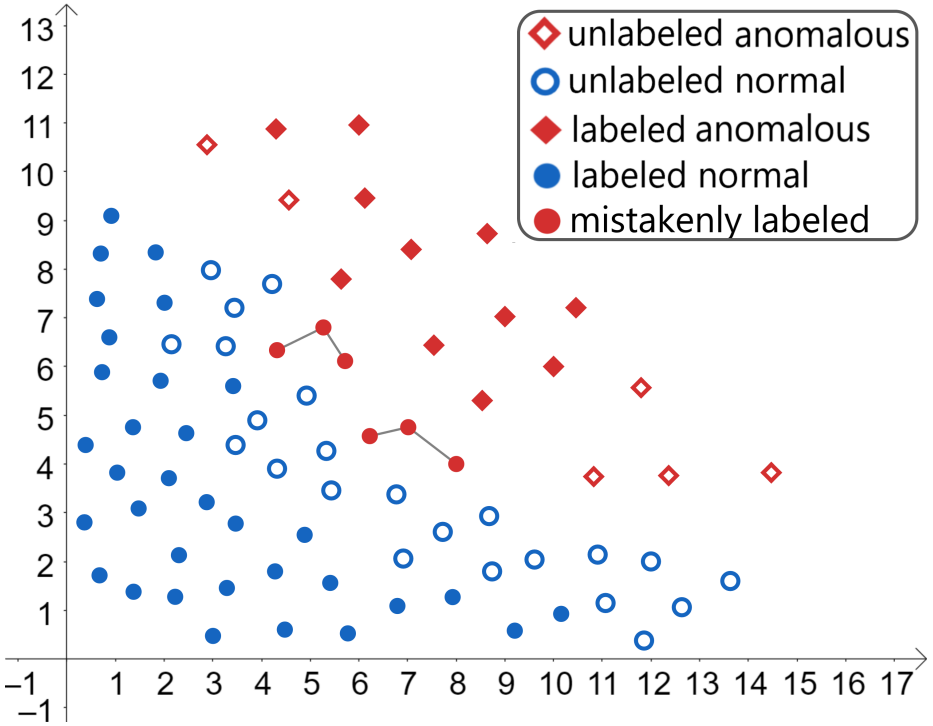
Active learning-assisted PU learning can improve its performance.
However, it will produce many false positives.



Step 2

Challenge 2: Active learning strategy

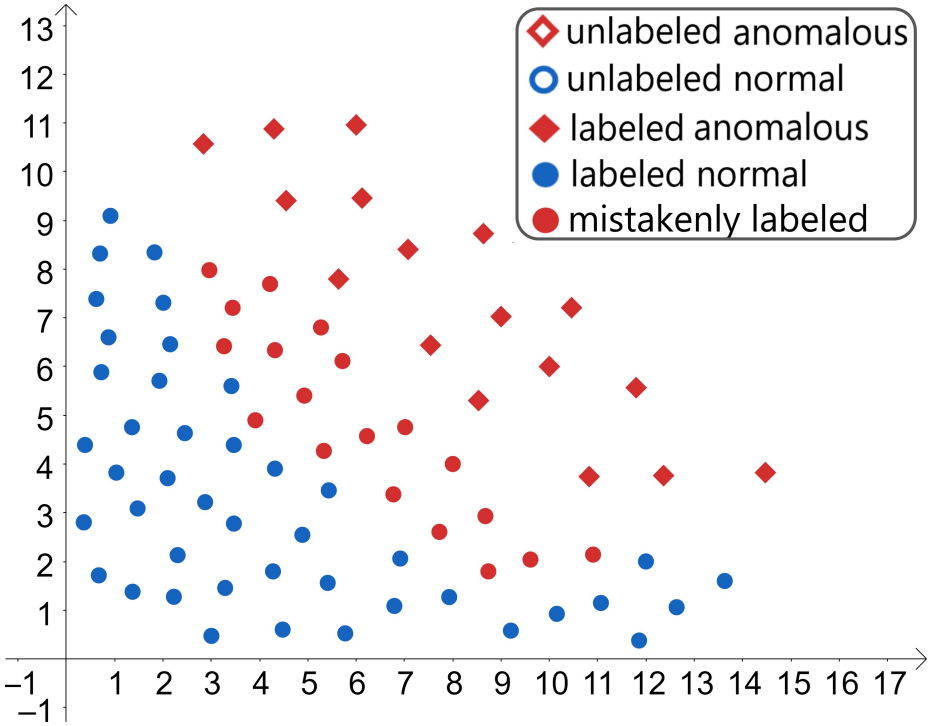
Active learning-assisted PU learning can improve its performance.
However, it will produce many false positives.



Step 3

Challenge 2: Active learning strategy

Active learning-assisted PU learning can improve its performance.
However, it will produce many false positives.



Step 4

Outline

Introduction

Challenges

Contribution

Framework

Evaluation

Contribution

KPI streams are large in number and diverse in pattern.

Contribution

KPI streams are large in number and diverse in pattern.

Utilize clustering, PU learning, and semi-supervised learning together to complete anomaly detection.

Contribution

KPI streams are large in number and diverse in pattern.

Utilize clustering, PU learning, and semi-supervised learning together to complete anomaly detection.

Active learning-assisted PU learning can improve its performance. However, it will produce many false positives.

Contribution

KPI streams are large in number and diverse in pattern.

Utilize clustering, PU learning, and semi-supervised learning together to complete anomaly detection.

Active learning-assisted PU learning can improve its performance. However, it will produce many false positives.

Select samples that are most likely to be positive in each iteration to label.

Outline

Introduction

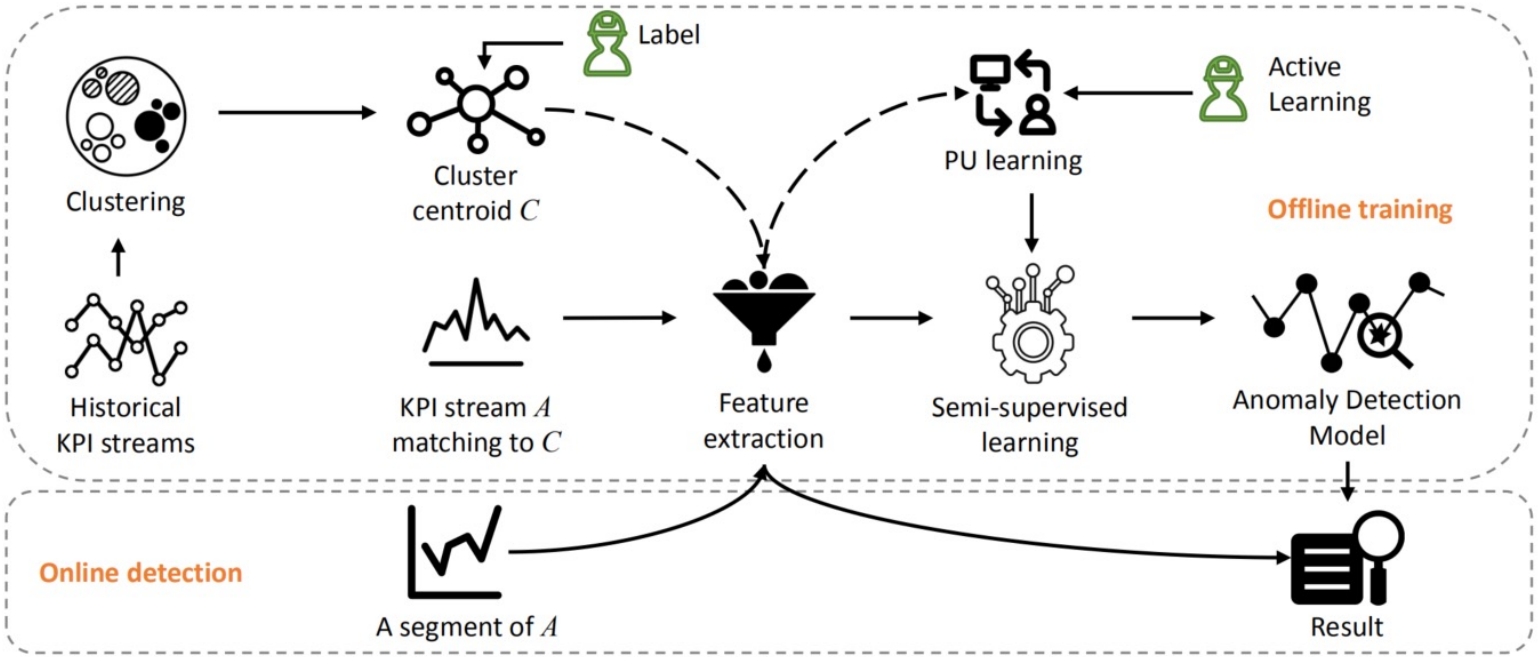
Challenges

Contribution

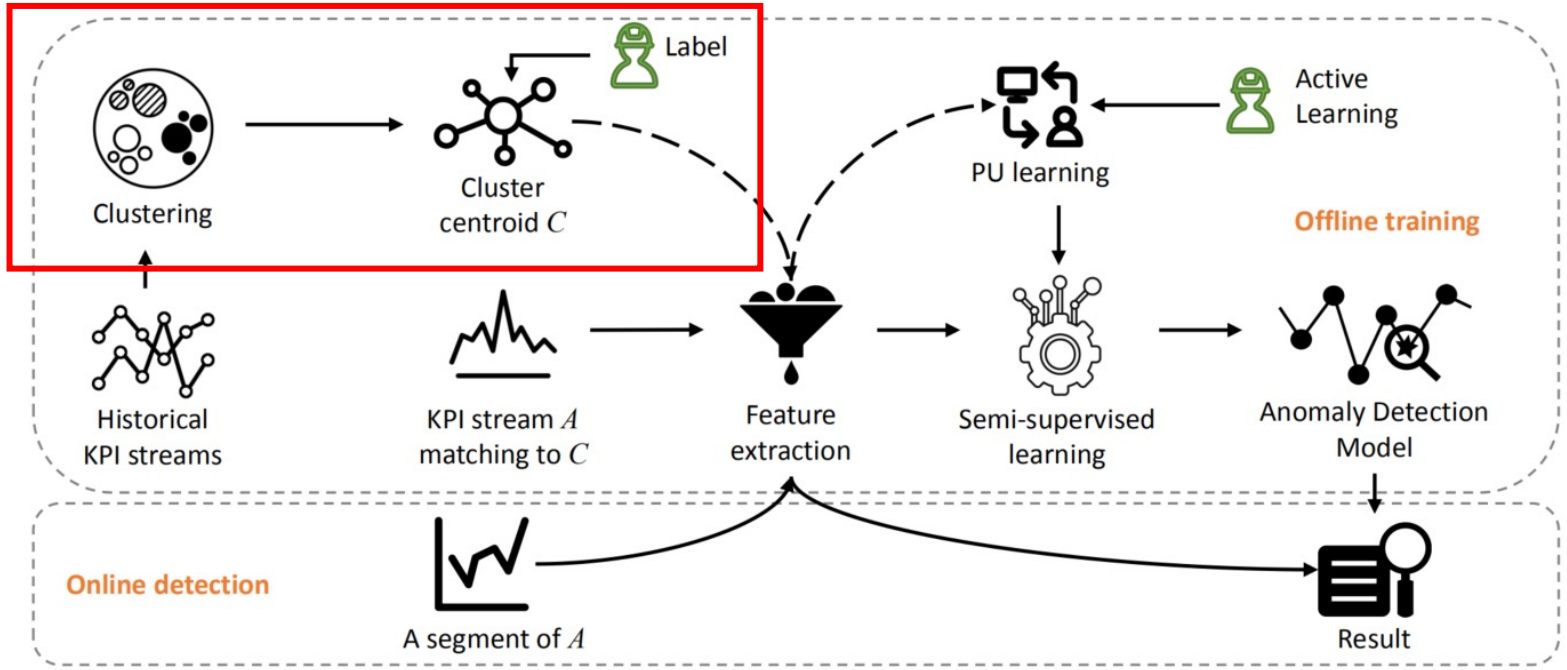
Framework

Evaluation

The framework of PUAD

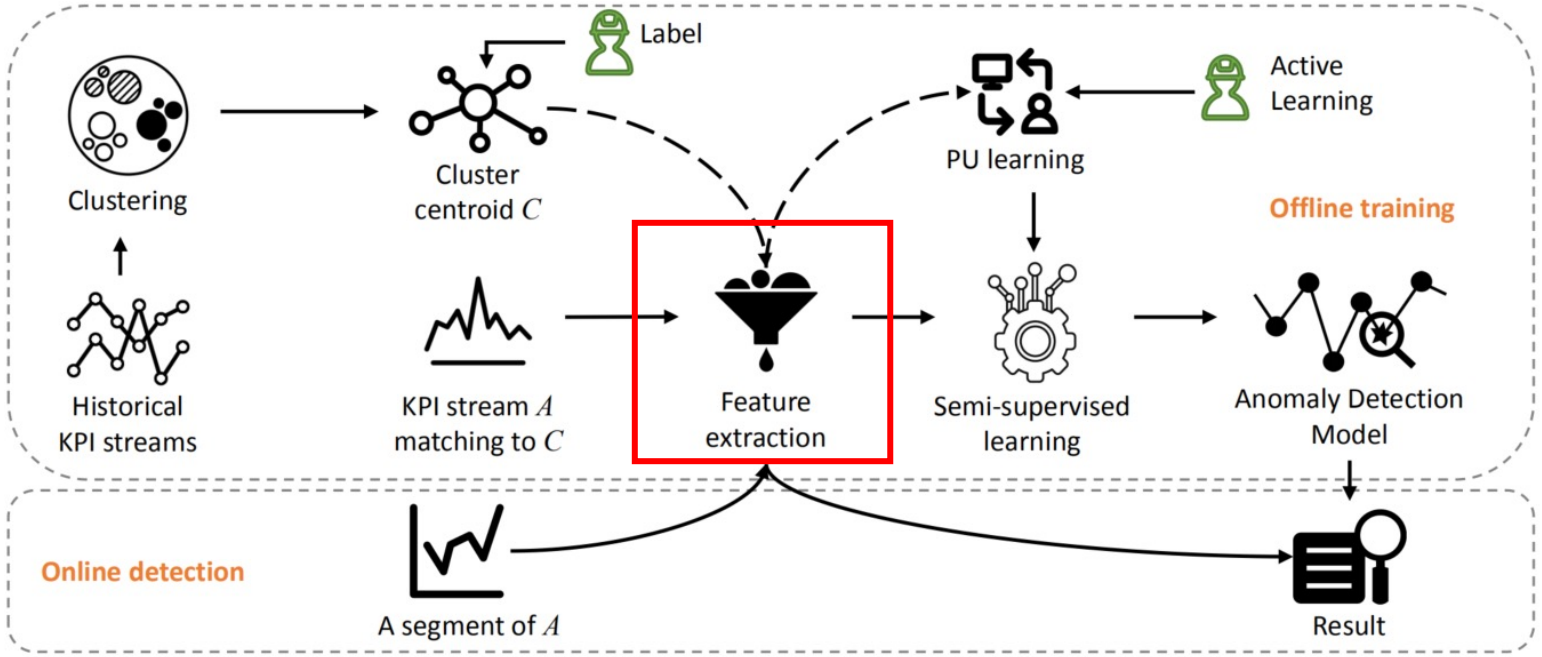


Clustering in the offline training process



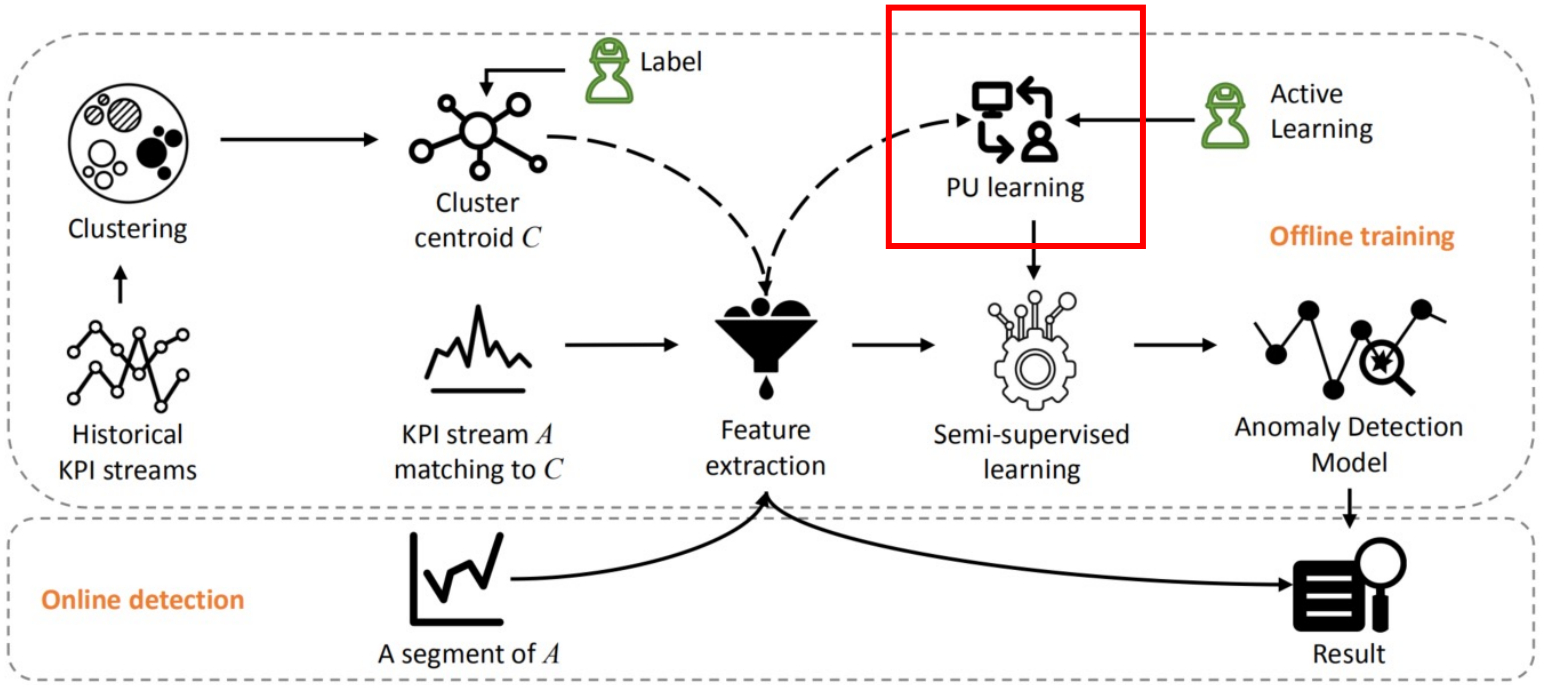
We can train an anomaly detection model for each cluster using a few manual labels, and “transfer” the trained model within each cluster.

Feature extraction in the offline training process



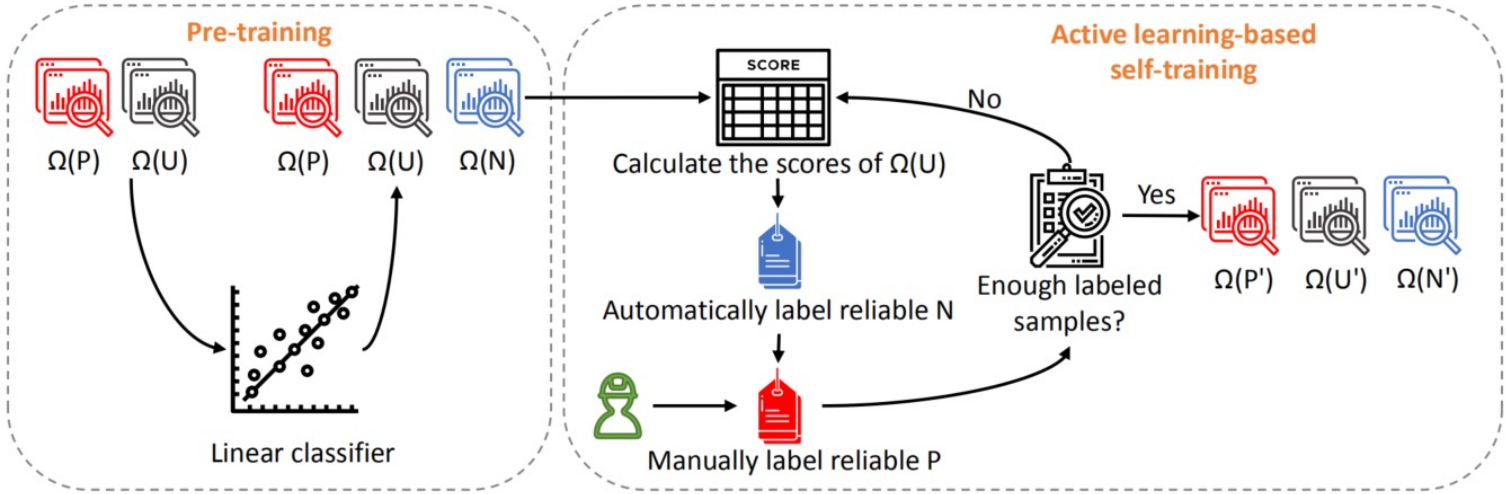
We extract and categorize the features into two groups:
temporal features and forecasting error features.

PU learning in the offline training process



The training set consisting of positive samples and unlabeled samples will be input together into the PU learning component.

The detailed framework of PU learning



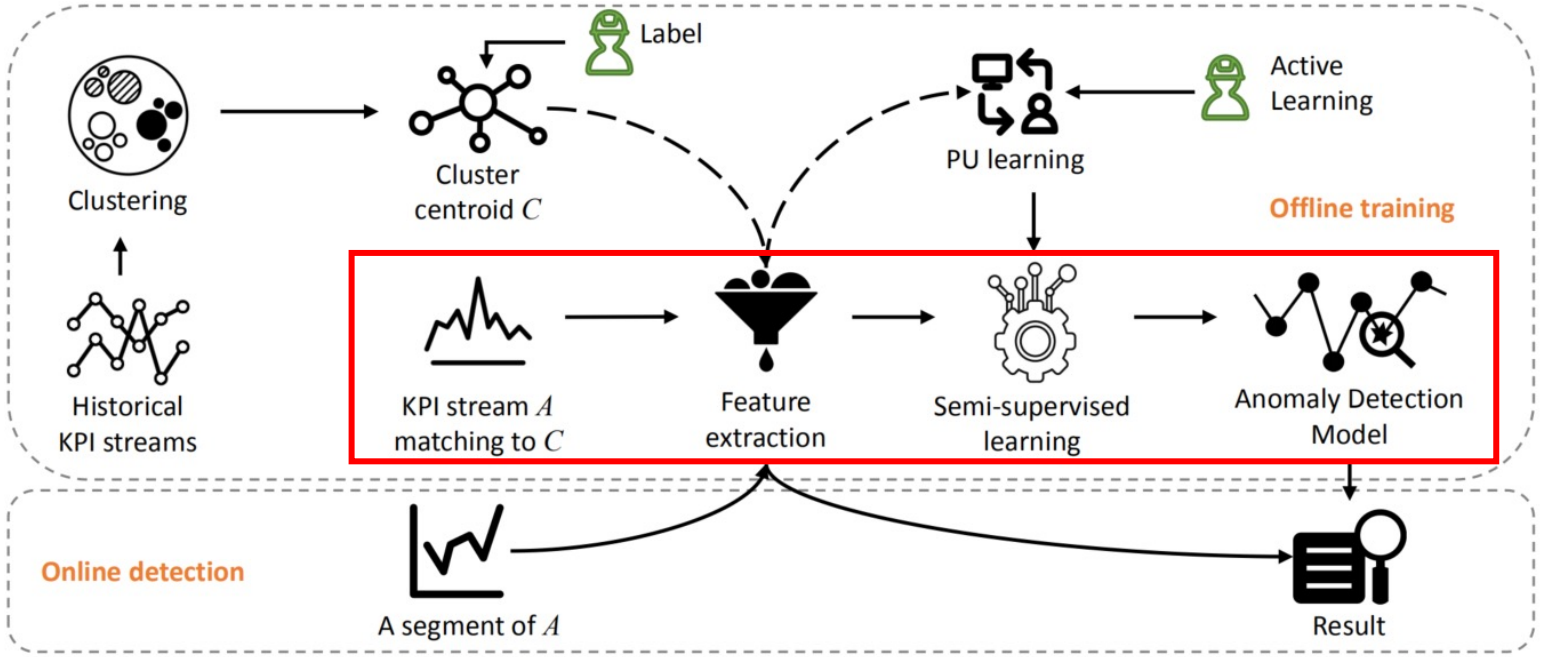
Ω(P): Positive samples

Ω(U): Unlabeled samples

Ω(N): Negative samples

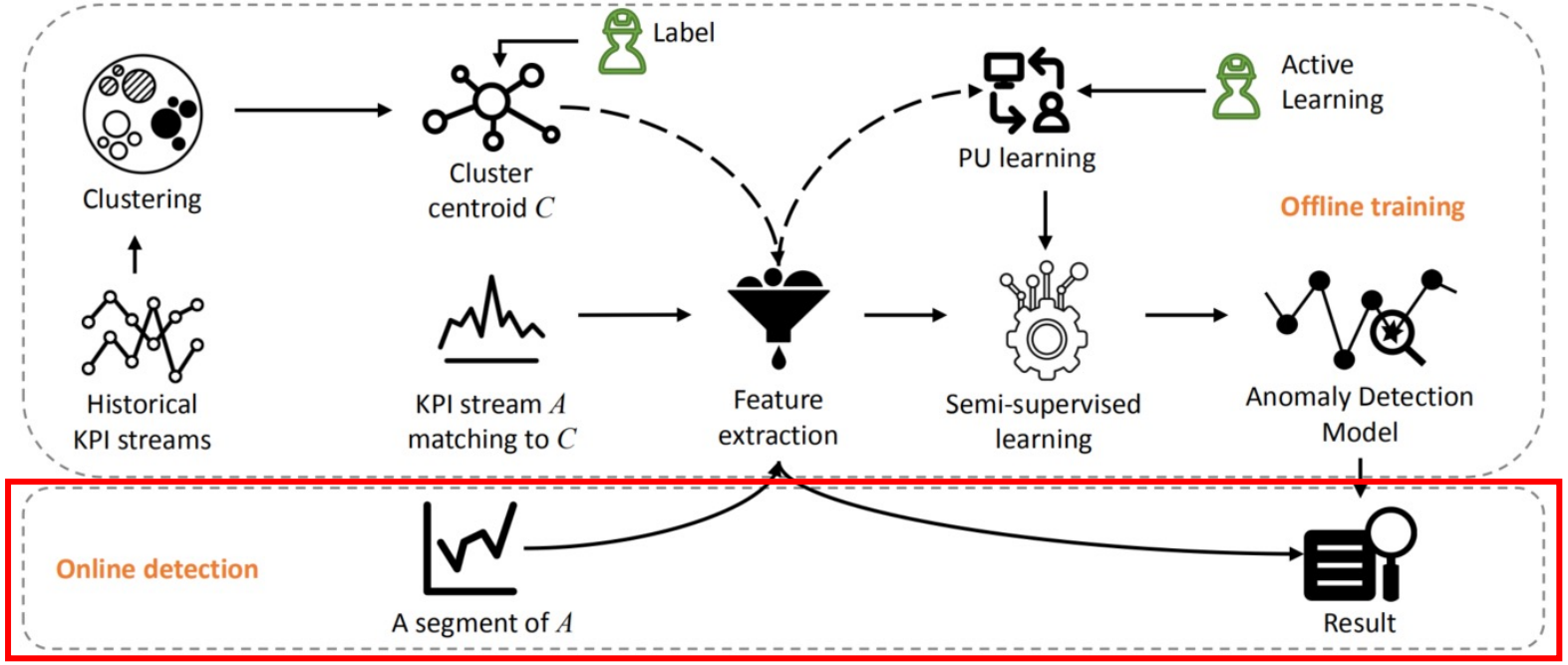
- **Pre-training:** Initialize Ω(N).
- **Active learning-based self-training:** Extend Ω(P) and Ω(N).

The process when a new KPI stream A emerges



For a newly emerging KPI stream A, assign it into an existing cluster and then extract its features. Train a model for it through semi-supervised learning.

The online detection process



For a newly arrived data point of the KPI stream A, its features would be firstly extracted. Then the features would be fed into the trained model to get an anomaly score.

Outline

Introduction

Challenges

Contribution

Framework

Evaluation

Dataset

Dataset 1

Process	Number of KPI streams	Interval (minute)	Total points	Anomaly points	Anomaly ratio (%)
Clustering	128	5	1024664	8318	0.81%
Anomaly Detection	80	5	643593	6839	1.06%

Dataset 2

Dataset	Interval (minute)	Total points	Anomaly points	Anomaly ratio (%)
AWS	5	67740	3097	4.57%
Artificial	5	16128	624	3.87%
Twitter	5	142765	217	0.15%

Evaluation Metrics

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

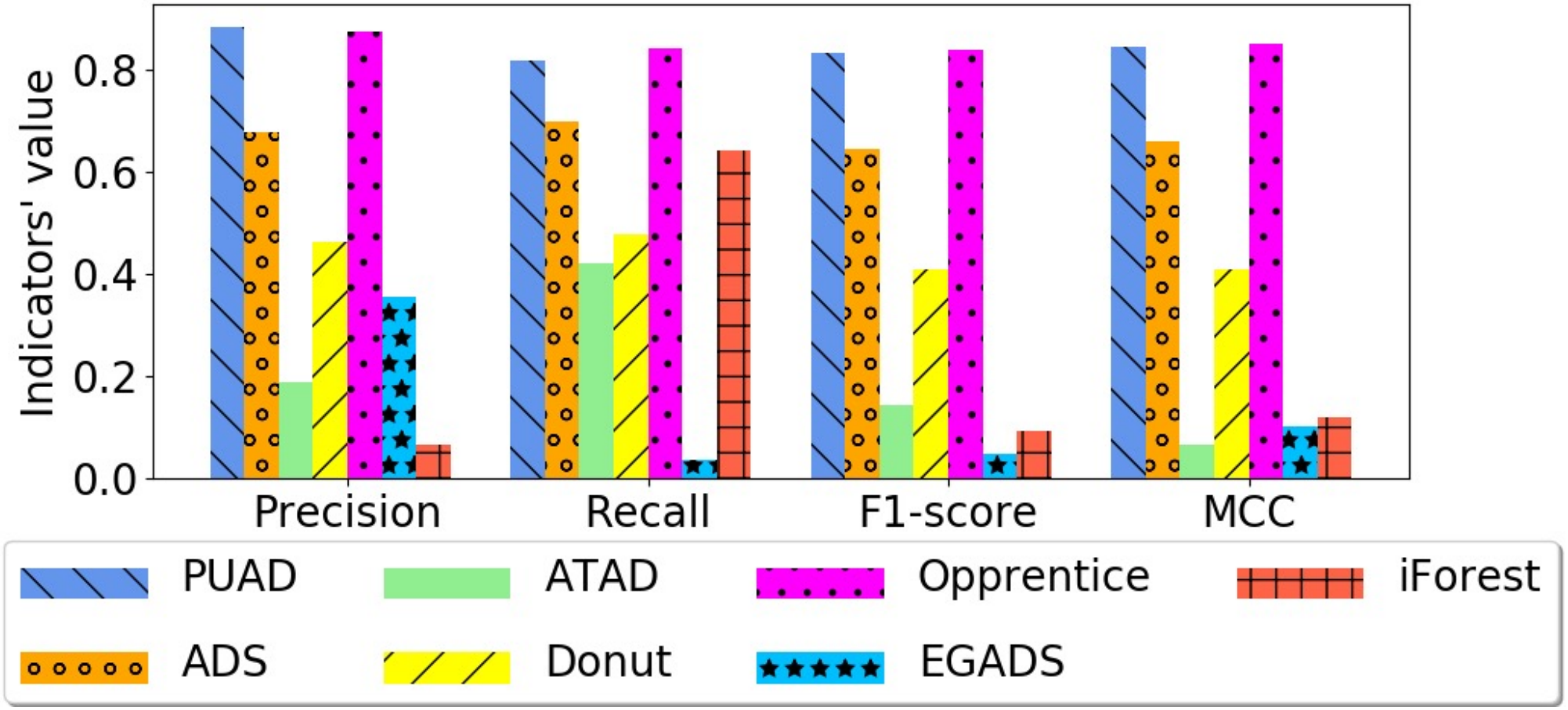
$$F1-score = \frac{2 \times precision \times recall}{precision + recall}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

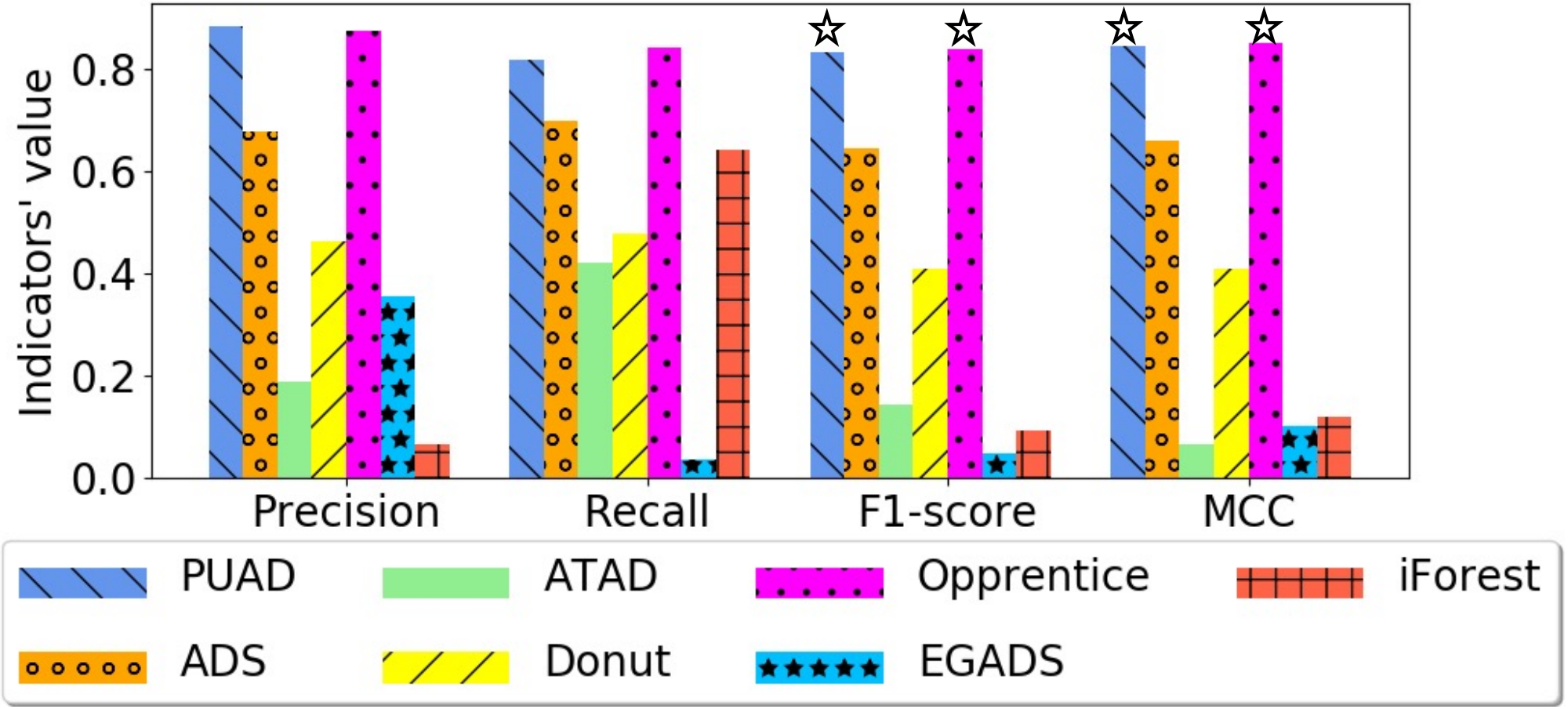
Baseline

Type	Baseline method	
Supervised	Opprentice	EGADS
Unsupervised	Donut	iForest
Semi-supervised	ADS	
Transfer learning	ATAD	

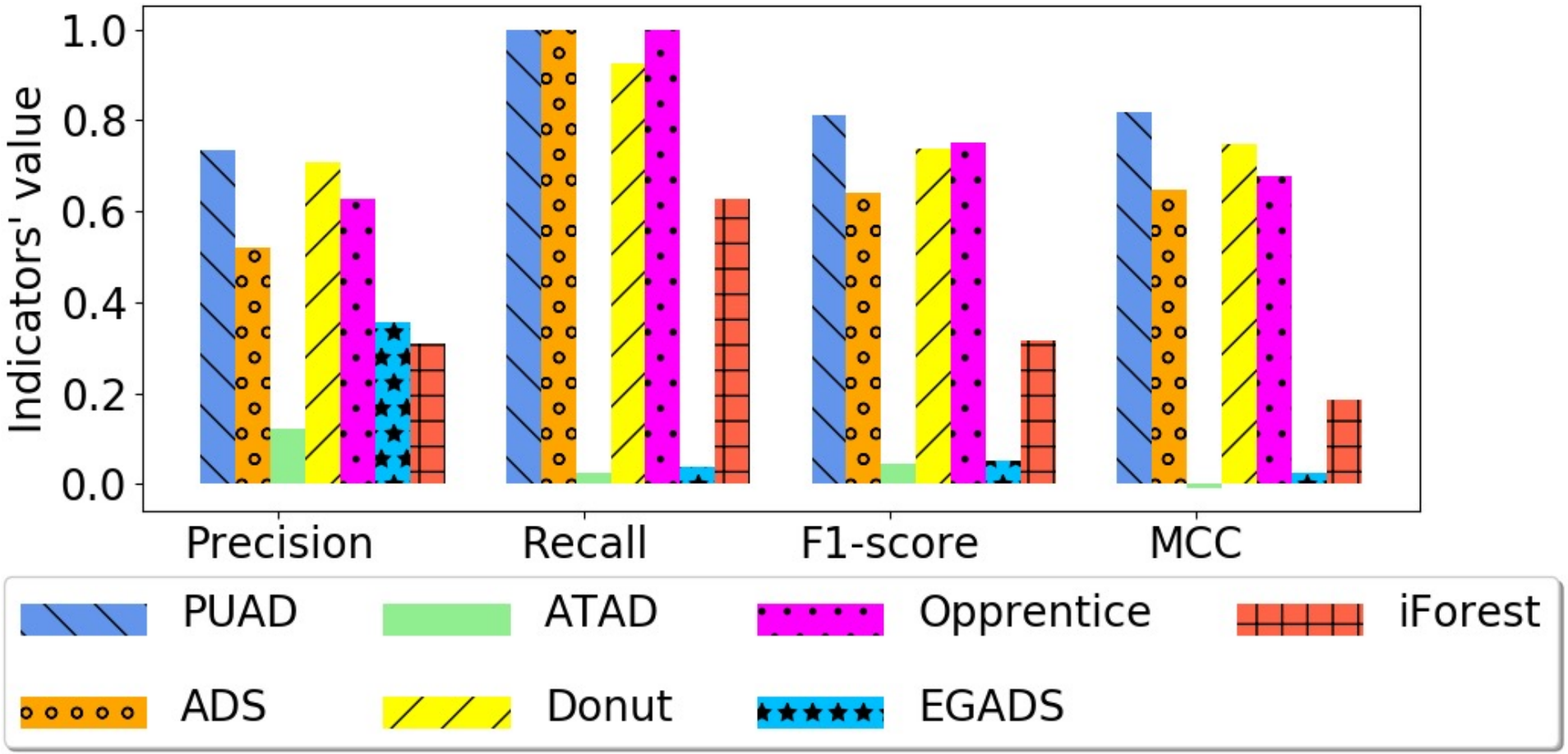
The effectiveness of different methods on dataset 1



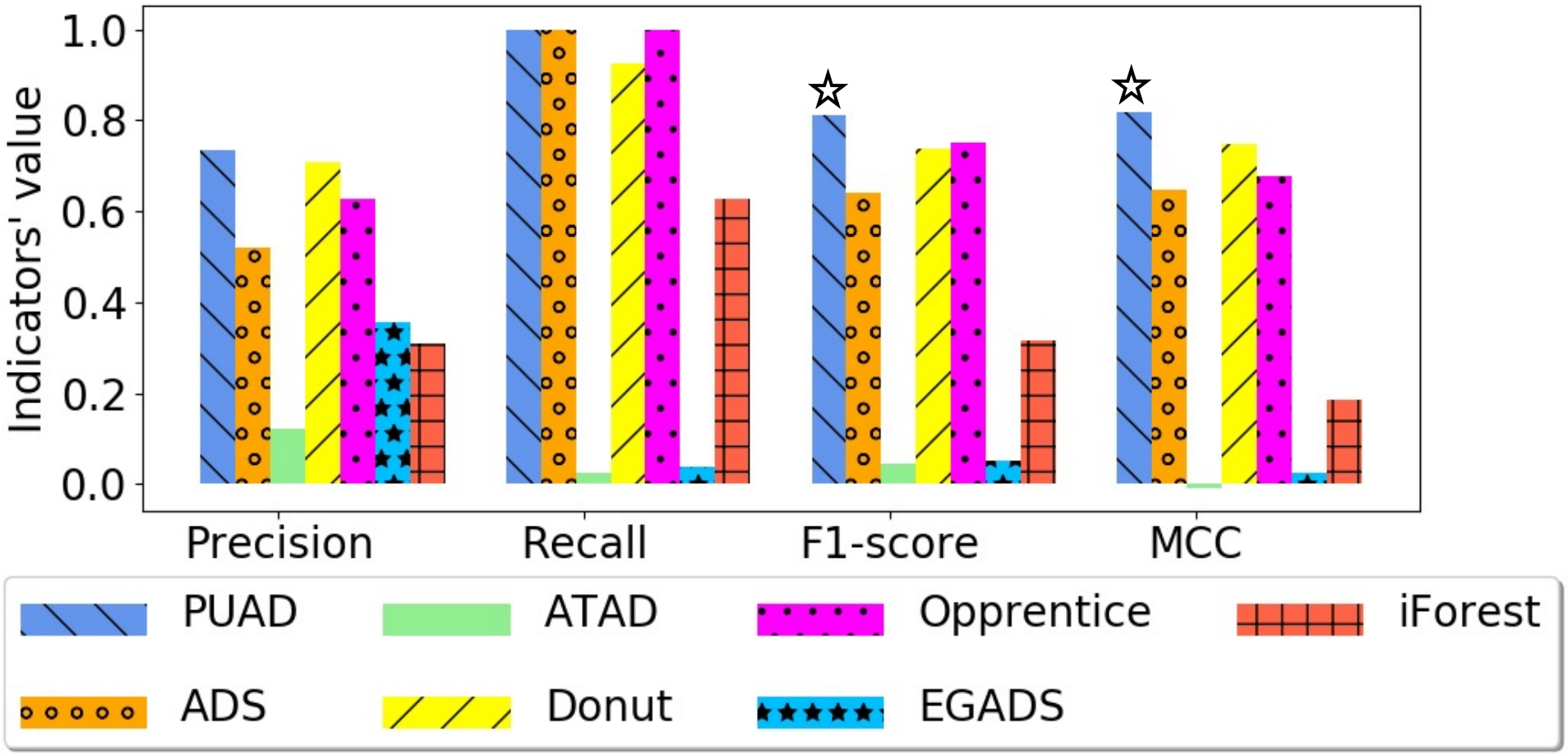
The effectiveness of different methods on dataset 1



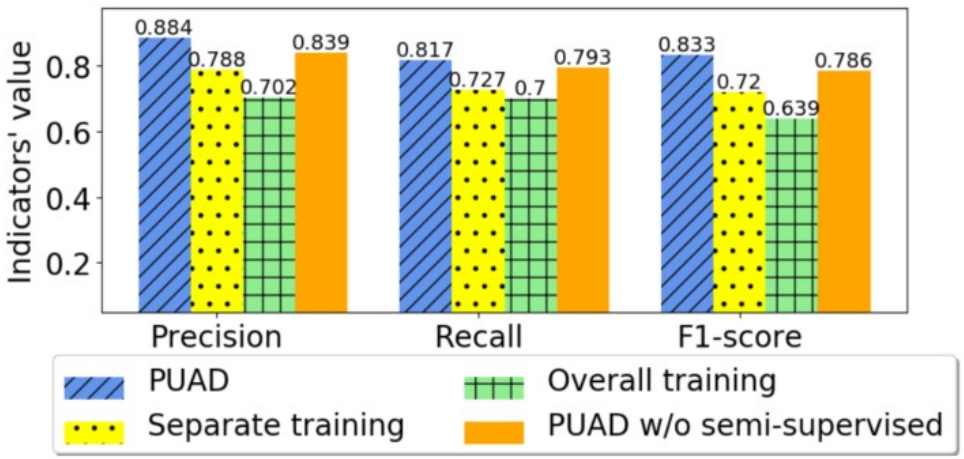
The effectiveness of different methods on dataset 2



The effectiveness of different methods on dataset 2



Ablation Study



The effectiveness of clustering and semi-supervised learning compoents

Clusters	W/o active learning	With random labels	Label boundary	PUAD
1	0.612	0.697	0.812	**0.920**
2	0.545	0.576	0.667	**0.967**
3	0.819	0.850	0.860	0.851
4	0.745	0.720	0.860	**0.872**
5	0.596	0.714	0.785	**0.839**
6	0.458	0.664	0.638	**0.737**
7	0.871	0.900	0.872	**0.921**
8	0.714	0.762	0.793	**0.812**
9	0.587	0.589	0.673	**0.675**
Average	0.636	0.719	0.772	**0.833**
Increase ratio	31.0%	15.9%	7.9%	--

The effectiveness of active learning

Conclusion

- We propose PUAD, a PU learning-based method to accurately detect anomalies with a small number of partial labels for large-scale KPI streams.
- Clustering, PU learning, active learning, and semi-supervised learning are combined to achieve accurate anomaly detection and small labeling effort at the same time.
- PUAD applies a novel active learning strategy to avoid false alarms.
- Extensive evaluation experiments demonstrate that PUAD achieves close accuracy to supervised methods, and significantly outperforms existing semi-supervised learning-based, transfer learning-based, and unsupervised learning-based methods.

Thank you!

Q&A