AutoKAD: Empowering KPI Anomaly Detection with Label-Free Deployment

Zhaoyang Yu[1], Changhua Pei[2], Shenglin Zhang[3], Xidao Wen[4], Jianhui Li[2], Gaogang Xie[2], Dan Pei[1]

[1]Tsinghua University, [2]Chinese Academy of Science
[3]Nankai University, [4]BizSeer Technology

Background

Design

Evaluation

Conclusion

Unavoidable Failures

December 2022 AliCloud outage lasts more than 12 hours, many customer services severely impacted

October 2021 Facebook's apps had a total outage of their external services for up to seven hours, evaporating nearly $50bn in market value overnight

In 2020 alone, Google's global services have experienced four major downtimes, each of which has resulted in huge economic losses

How can we minimize the occurrence of catastrophic level failures and reduce economic losses?

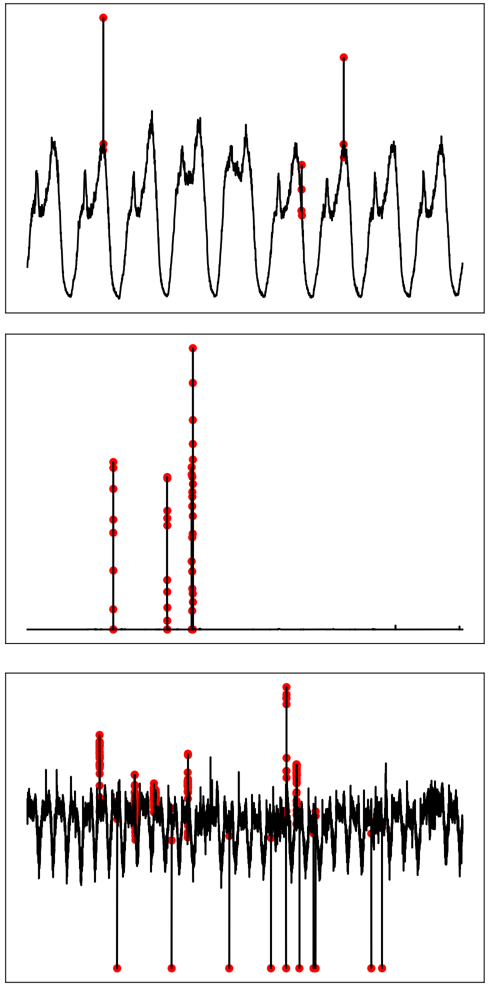
KPI Anomaly Detection

Key Performance Indicator (KPI)

- Time-series data collected from a myriad of sources
- Reflect the health status of a service
- Most KPIs in the real world are seasonal

KPI anomaly detection is essential for service reliability

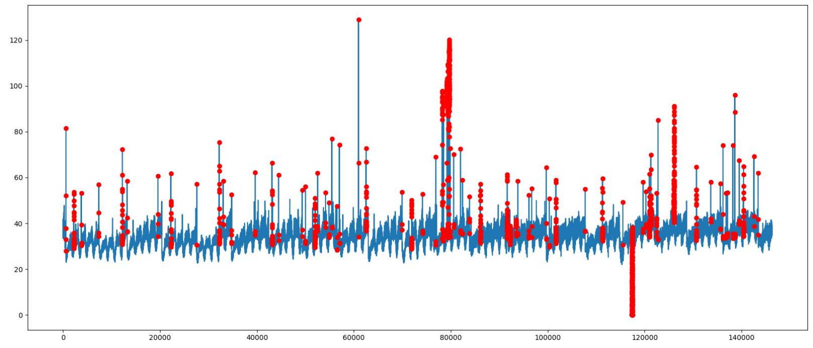
- Identifying potential issues by detecting anomalies in KPIs
- Labels are expensive
- Most algorithms are unsupervised



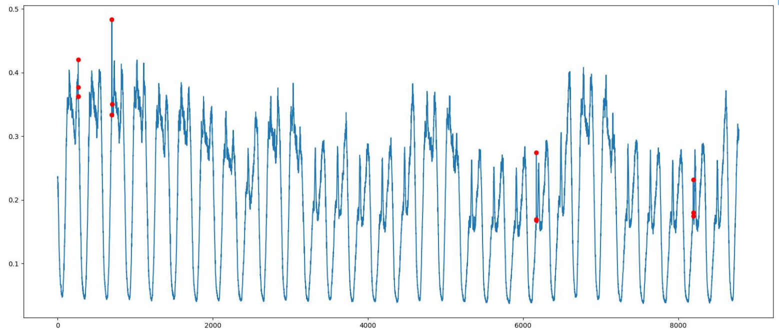
Configurations Matter

Configuration affects detection performance

- Configuration = Algorithm + Hyperparameter
- One configuration will not work well for all KPIs
- Same configuration performs very differently on different datasets
- Finding a satisfying configuration is challenging for IT operators



F1-Score: 0.267

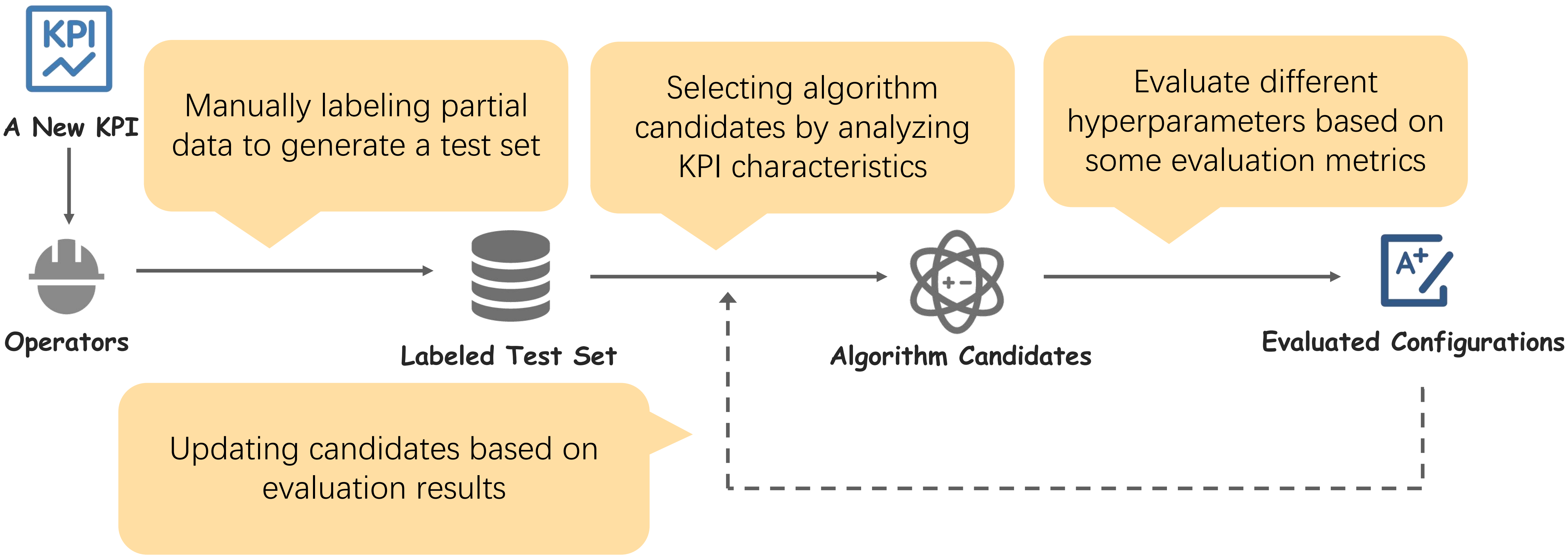


F1-Score: 0.917

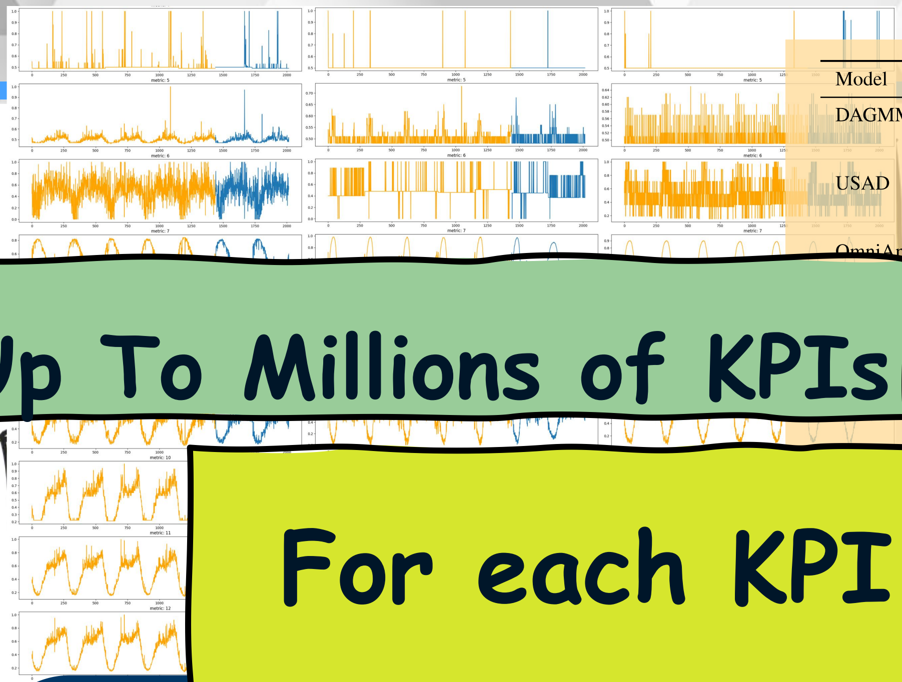
Same LSTM detector with the same hyperparameters

Deployment of KPI Anomaly Detection

What do operators need to do when deploying an anomaly detection model for a KPI?



Deployment Dilemma



Model	Advantages	Data Preprocessing Method	Model Structures
DAGMM	• Based on time point. • Preserves the low-dimensional features and reconstruction error for anomaly detection.	Does standardization.	AE + Gaussian Mixture Model (GMM)
USAD	• Leverages the advantages of AE and adversarial training. • A straightforward model structure and a limited number of parameters.	Does standardization.	AE + Generative sarial Networ
OmniAnomaly	• Mod • Emp • Sin dimen • Wor • Be invaria		
	• Employs a hierarchical VAE (HVAE) to learn different features independently.	Uses zero to fill in missing values, remove extreme values and does stan-	1D-CNN + RN

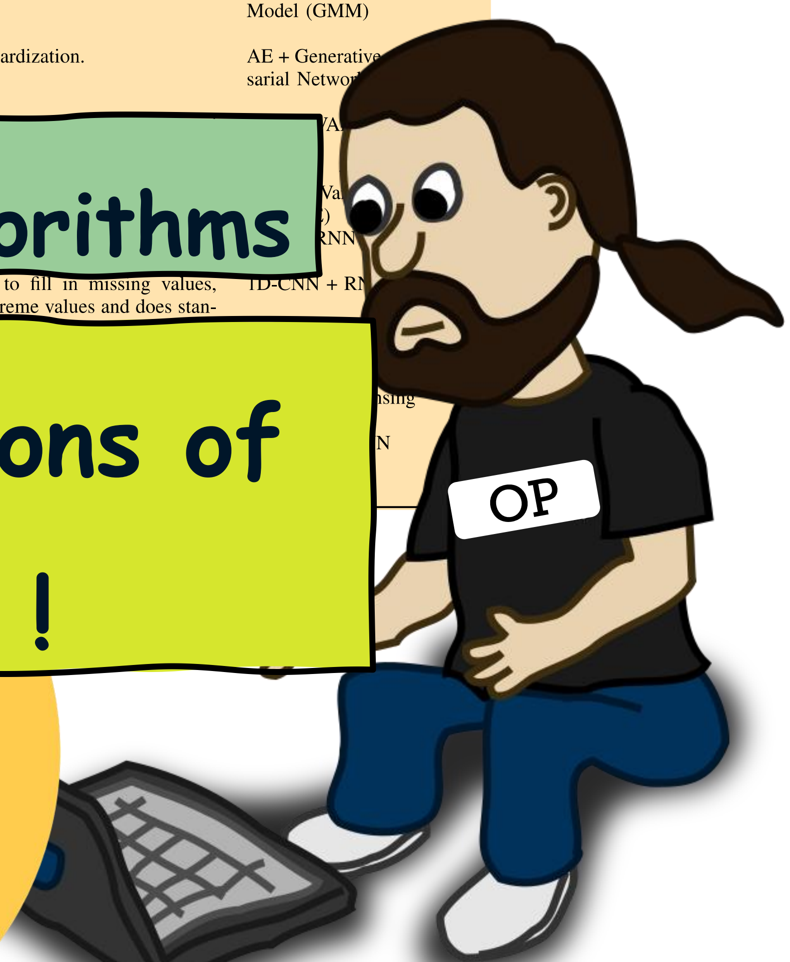
Up To Millions of KPIs

Dozens of Algorithms

For each KPI, there are millions of configuration candidates !

Learnin
Momentum Weight Decay Dropout Rate
Number of Layers Activation Function Kernel
Size Stride Pooling Size Tree Depth Minimum

Dozens of Hyperparameters



CASH Problem and AutoML

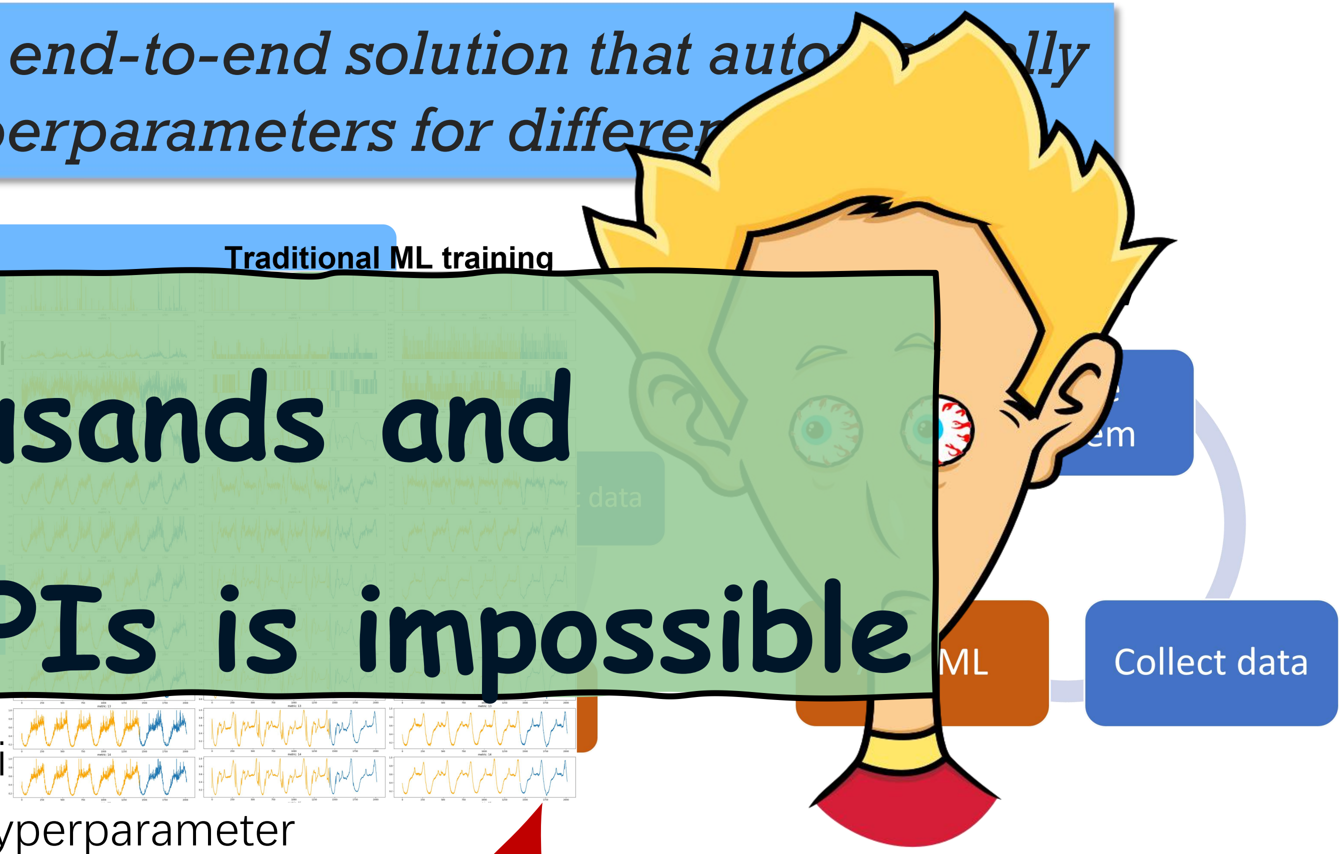
What operators need is an end-to-end solution that automatically selects algorithms and hyperparameters for different KPIs

CASH problem

- Combined Algorithm Selection and Hyperparameter optimization
- Automatically find a satisfying configuration
- Black-box optimization

AutoML

- Automated Machine Learning
- Automate facets of the machine learning pipeline
- e.g. algorithm selection, data preprocessing, hyperparameter optimization
- Existing methods are mainly based on labeled test sets for searching configurations



Labeling thousands and even millions of KPIs is impossible

Not Practical

Our Goal

Given a KPI:

$X = [(x_1, y_1), \cdots, (x_N, y_N)] \quad x_i \in R, \; y_i \in \{0, 1\}$

Our goal is to design an effective and efficient AutoML framework for unsupervised KPI anomaly detection under a Label-Free condition.

the

Gi

$G = \{g_1, g_2, \ldots, g_m\}, \; g_i : x \mapsto y$

For each algorithm g_i,
a set of possible hyperparameters:

$$\Theta_i = \{\theta_1^i, \theta_2^i, \ldots, \theta_{n_j}^i\}$$

$$\underset{g \in G, \, h \in \Theta_g}{\arg\max} \, L(g, \theta; X)$$

Challenges

Unattainable objective function

- The objective function is used to evaluate the performance of a configuration given by the AutoML framework
- Without labels, it is unable to calculate the evaluation metrics (e.g., Precision, Recall, F1-score)
- Without these metrics, it is unfeasible to evaluate the performance of a configuration

High time complexity

- The number of KPI in practice is huge
- Training a machine learning model is time consuming
- AutoML needs to train models several times

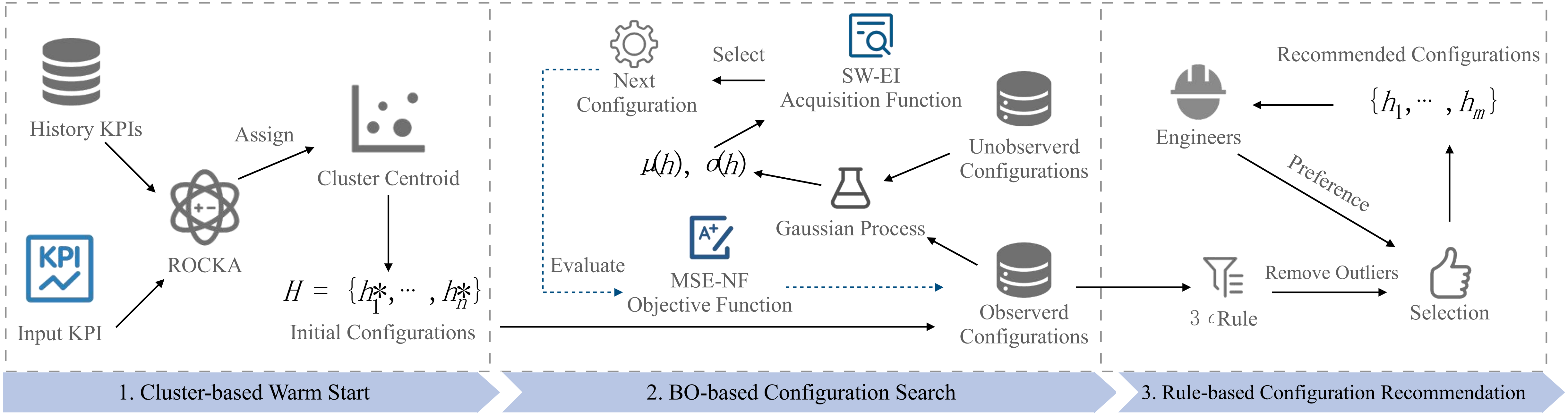
Background

Design

Evaluation

Conclusion

AutoKAD: Overview



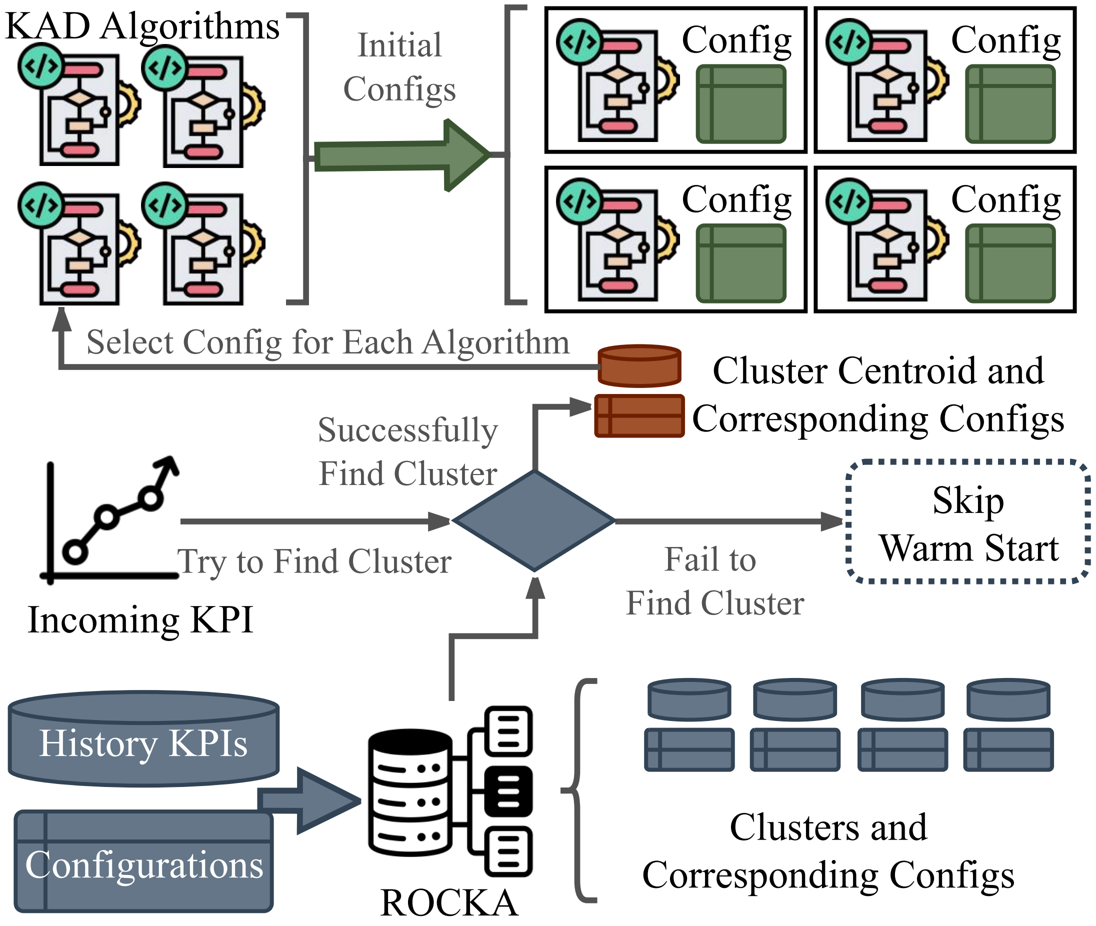
Cluster-Based Warm Start

Cold Start

- AutoML will give a random configuration as a starting point for the search
- Random initial configuration leads to a time-consuming cold start

Core idea

- Intuitively the promising configurations of similar KPIs are likely to be similar
- Using cluster to find similar KPIs
- Explore different algorithms



Bayesian Optimization

Bayesian Optimization is a sequential design strategy for black-box optimization and widely used to solve the CASH problem.

- **Surrogate Model**

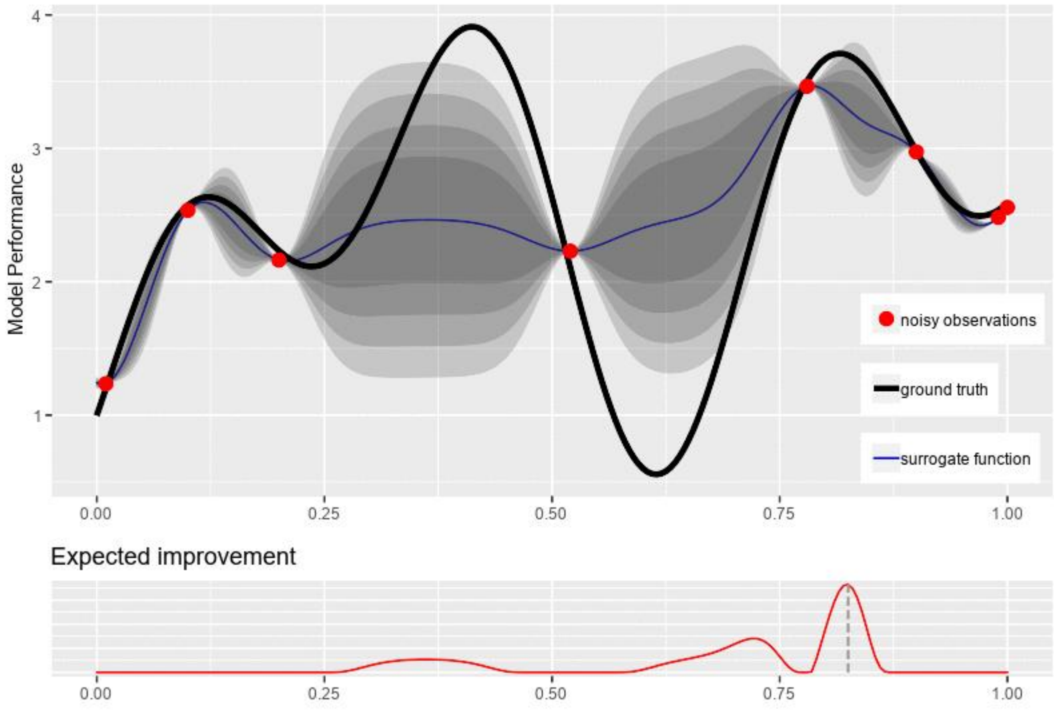
Build a probability model between configuration and model performance.
"Guessing" performance based on observations

- **Objective Function**

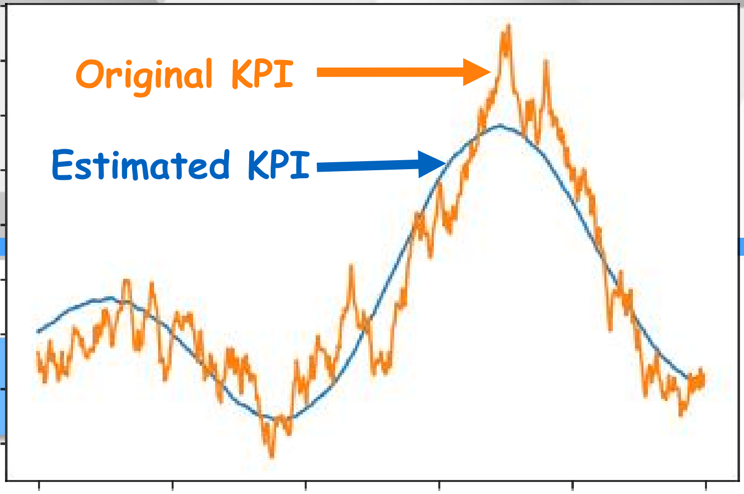
Evaluating the performance of a configuration.
Evaluation Metric

- **Acquisition Function**

Deciding the next configuration to observe.
Exploration-Exploitation Trade-off



Objective Function



The goal of objective function is to evaluate

Estimated KPI close to the original KPI

The detection model learns the **normal pattern** of the KPI well

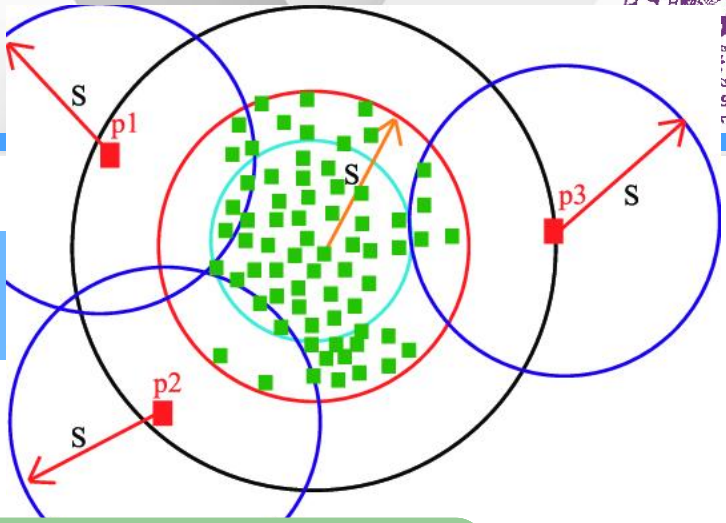
$$MSE(\boldsymbol{X}, \hat{\boldsymbol{X}}) = \sum_{i=1}^{N} \frac{(x_i - \hat{x}_i)^2}{N}$$

Estimated KPI looks like a normal KPI

Prevent the detection model from learning **anomalous pattern**

?

Objective Function: Normal Factor



Local Outlier Factor (LOF)

- Using distance to neighbors to measure whether a point is normal or not
- Lack of physical meaning for KPIs
- High tim

$$LOF_k(x) = \frac{1}{|N_k(x)|} \cdot$$

Period O

- The ma
- Data po
- Replace

$$\ldots_t, x_{t-p}) + d(x_t, x_{t+p})$$
$$\ldots, x_{t-p}) + d(x_{t-p}, x_{t+p})$$
$$\ldots_t, x_{t-p}) + d(x_t, x_{t+p})$$
$$\ldots x_{t+p}) + d(x_{t-p}, x_{t+p})$$

Previous Period

Next Period

Our Objective Function

$$MSE\text{-}NF = MSE + \alpha NF$$

The coefficient α > 0 is utilized to balance between MSE and NF

Normal Factor (NF)

- Normal Factor is the average POF of all data points in the estimated KPI $NF(\boldsymbol{X}) = \frac{1}{|\boldsymbol{X}|} \cdot \sum_{x \in \boldsymbol{X}} POF(x)$

Acquisition Function

The goal of the acquisition function is to find the next most observable configuration.

Expected Improvement (EI) is an effective and widely used acquisition function in BO-based AutoML

$$EI(h) = \mathbb{E}[\max(\mu(h) - f(h^+), 0) \mid \{h_1, \cdots, h_k\}]$$
$$\text{where} \quad h^+ = \arg\max_{i=1,\cdots,k} f(h_i)$$

Without labels, the gap between F1 and our objective function cannot be completely eliminated

An intuitive solution is to be more inclined to explore different configurations

Similarity Weighted Expected Improvement (SW-EI)

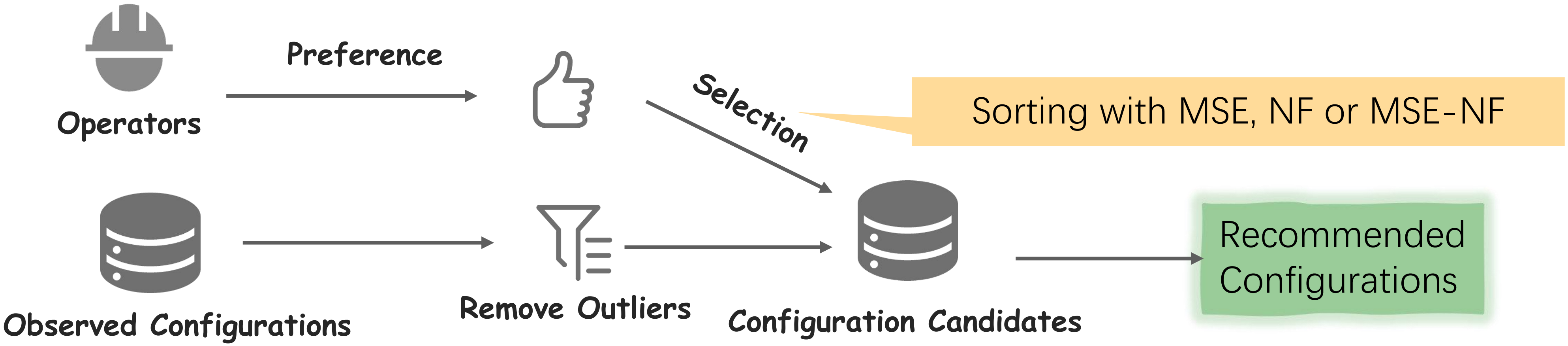
$$SW\text{-}EI(h) = \sum_{i=1}^{n} w_i(h, h_i^+)(\mu(h) - f(h_i^+))\Phi(z_i) + \sum_{i=1}^{n} w_i\sigma(h)\phi(z_i) + \sigma(h)\phi(\sum_{i=1}^{n}\frac{z_i}{n})$$

Configuration Recommendation

Operators have different sensitivities to false alarms and missed alarms for different KPIs.

High priority KPI → Minimize missed alarm → High recall

Low priority KPI → Minimize false alarm → High precision



Background

Design

Evaluation

Conclusion

Dataset

Dataset A

- Based on AIOps2018 Challenge dataset.
- Collected from several big Internet companies' business systems.

Dataset B and C

- Collected from the two most important trading systems of an international commercial bank.
- Dozens of operators closely monitor these KPIs to maintain service quality

Dataset	#KPI	#Train/#Test	Anomaly rate
$\mathcal{A}$	29	3004066 / 2918847	2.648% / 1.869%
$\mathcal{B}$	29	1642815 / 1642810	1.089% / 1.065%
$\mathcal{C}$	30	2078848 / 2078854	0.781% / 0.817%

Effectiveness in Searching for Configurations

Under 100 iterations and 1 hour limitation for each KPI

Methods	$\mathcal{A}$			$\mathcal{B}$			$\mathcal{C}$			*Avg.*		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Default Donut	0.881	0.380	0.531	0.457	0.689	0.550	0.707	0.638	0.671	0.682	0.569	0.620
Random Search	0.676	0.537	0.599	0.663	0.631	0.647	0.675	0.783	0.725	0.671	0.650	0.661
BayesOpt	0.876	0.525	0.657	0.763	0.672	0.715	0.875	0.681	0.766	0.838	0.626	0.717
BOAT	0.831	0.575	0.680	0.778	0.688	0.730	0.823	0.708	0.761	0.811	0.657	0.726
AutoKAD	0.861	0.694	**0.769**	0.920	0.723	**0.810**	0.916	0.781	**0.843**	0.899	0.733	**0.807**

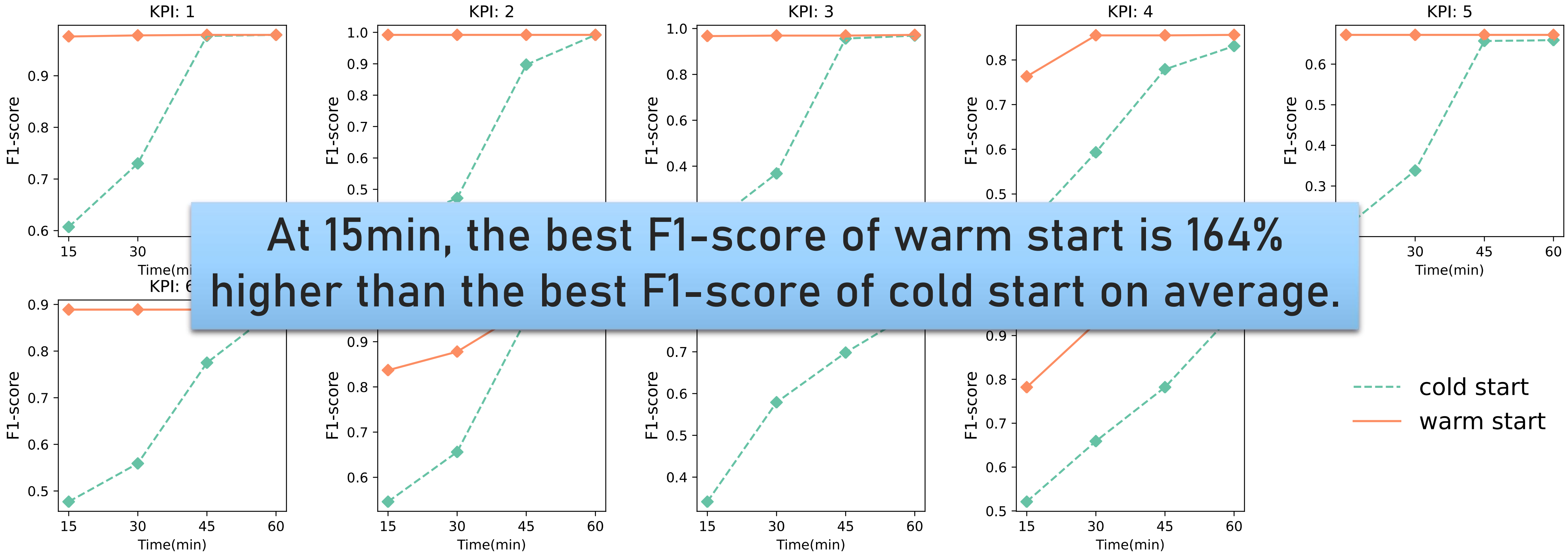
Use MSE as the objective function

outstanding performance on all three datasets

Compared with baseline methods, the results show that AutoKAD is indeed effective in the CASH problem for unsupervised KAD models.

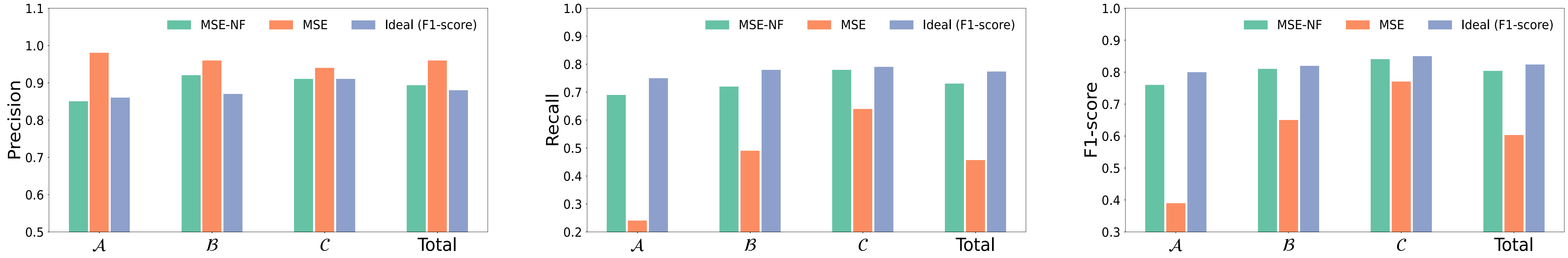
Time Efficiency

Under 100 and 1 hour limitation for each KPI



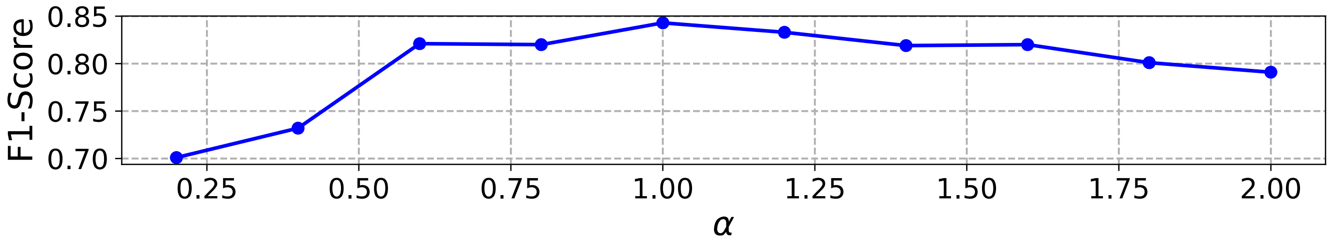
At 15min, the best F1-score of warm start is 164% higher than the best F1-score of cold start on average.

Effectiveness of the Objective Function



The performance of MSE-NF is quite close to the performance of the ideal F1-Score

When α is close to 1, the performance is commendable



Correctness of Configuration Recommendation

Optimal Configurations
VS
Configurations given by our strategy

Top 3	Dataset	$\mathcal{A}$			$\mathcal{B}$			$\mathcal{C}$		
	Strategy	P	R	F1	P	R	F1	P	R	F1
1st	ideal	0.99	0.88	0.76	0.99	0.96	0.81	0.97	0.87	0.84
	our	0.97	0.77	0.71	0.98	0.95	0.78	0.97	0.81	0.84
2nd	ideal	0.99	0.77	0.73	0.98	0.95	0.80	0.97	0.83	0.83
	our	0.98	0.71	0.71	0.97	0.88	0.78	0.97	0.81	0.83
3rd	ideal	0.99	0.71	0.70	0.98	0.92	0.78	0.97	0.82	0.83
	our	0.98	0.70	0.66	0.97	0.89	0.78	0.96	0.80	0.82

Difference

- Precision: < 0.02
- Recall: < 0.11
- F1-Score: < 0.05

Background

Design

Evaluation

Conclusion

Conclusion

Automatic label-free deployment of KPI anomaly detection

- Operators in industry are not machine learning experts
- Automatic label-free deployment is important to ensure the detection performance

AutoKAD: an AutoML framework for unsupervised KPI Anomaly Detection

- **Cluster-based Warm Start** => Reduce the time wasted by cold start
- **MSE-NF objective function** => label-free configuration searching
- **SW-EI acquisition function** => balance exploration and exploitation
- **Recommendation strategy** => satisfy operators' preferences for different KPIs
- Effectively and efficiently find a promising configuration for KPI anomaly detection

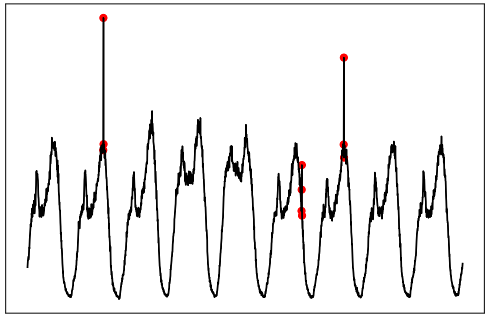
Opensource *AutoKAD*

- https://github.com/NetManAIOps/AutoKAD

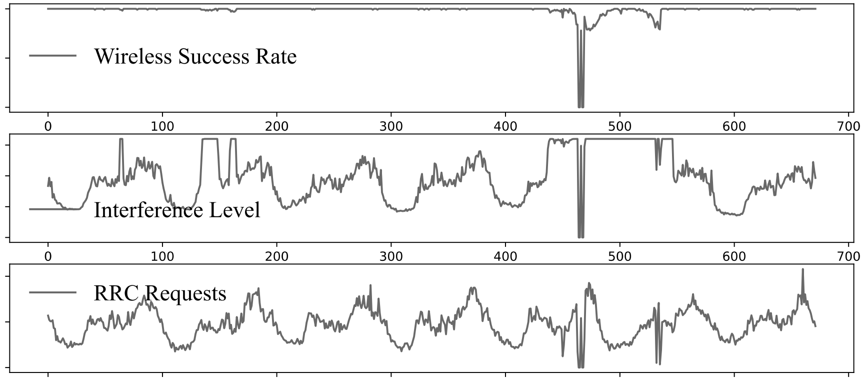
Limitation

KPI Requirement

- AutoKAD is designed for univariate KPIs and cannot tackle multivariate KPIs



Univariate KPI



Multivariate KPI

Algorithm Requirement

- The algorithm should have the ability to give the estimated KPI (most algorithms)
- Algorithms like PCA and kNN cannot be used in AutoKAD

Thanks

Q&A