MADDC: Multi-Scale Anomaly Detection, Diagnosis and Correction for Discrete Event Logs

Xiaolei Wang， Lin Yang， Dongyang Li， Linru Ma， Yongzhong He， Junchao Xiao， Jiyuan Liu， Yuexiang Yang

ACSAC, Dec 9th, 2022

国防科技大学 National University of Defense Technology
北京交通大學 BEIJING JIAOTONG UNIVERSITY
中山大學 SUN YAT-SEN UNIVERSITY

Outline

1. **Introduction**
2. System Design
3. Evaluation

Discrete Event Log Anomaly Detection--An Illustrative Example

Step1: Log Event Key Composition

tcp,FIN,ftp-data,0,0,0,0,0,0--> 5

tcp,FIN,-,0,0,0,0,0,6 --> 9

tcp,FIN,ssh,0,0,0,0,0,1 --> 11

tcp,FIN,-,0,0,0,0,0,3 --> 22

tcp,FIN,-,0,0,0,0,0,19 --> 26

Time Line

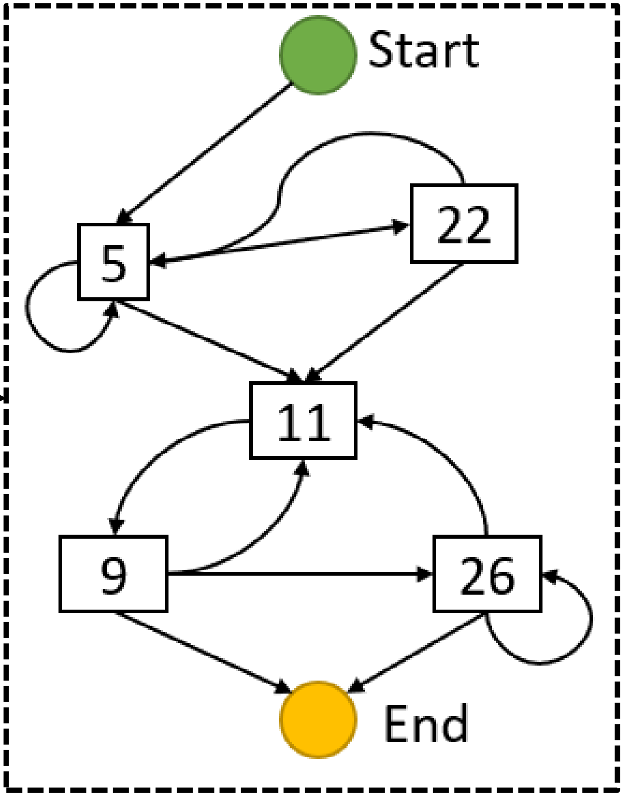
5->5->5->22->11->9->11->9->26->26->26

5->5->22->5->11->9->11->9->11->9

5->5->5->22->11->11->9->26->26

22->5->5->5->26->26->11->9->11 **Abnormal**

Step2: Log Event Sequence Aggregation



Step3: Directly Follows Graph(DFG) Construction

Recent Trends-Deep Learning based Anomaly Detection

Input: ***h*** recent log event keys up to e_{t-1}

DeepLog

$P(e_t = k_i|w), k_i \in K(i = 1,2, \dots, n),\ Log\ Key\ Set$

Output: conditional prediction probability of next event key given the recent sequence

$w = \{e_{t-h}, \dots, e_{t-2}, e_{t-1}\}$

Sliding window based prediction

57 18 56 … 25

18

Check Whether Real Next Log Key is in Top-N Predictions

{56,54,25}

Top N Candidates (N=3)

Input:

$\{e_{t-h}, \dots, e_{t-2}, e_{t-1}\}$

DabLog

$P\left(e_j = k_i|w\right), t - h \le j \le t - 1,\ k_i \in K(i = 1,2, \dots, n)$

Output: conditional reconstruction probability of each key at each time slot

Sliding window based reconstruction--AutoEncoder

57 18 56 … 25

Check Whether Real Original Log Key is in Top-N Predictions

{56,25,18}

Top N Candidates (N=3)

Improved a lot the accuracy of anomaly detection for discrete event logs. But still is not enough!!!

*Two known **LSTM** based works, **DeepLog**(CCS 2017), **DabLog**(Asiaccs 2021)*

Limitations of Existing Representative Deep Leaning based Methods

To Fulfil Practical Anomaly Detection for Discrete Event Log:

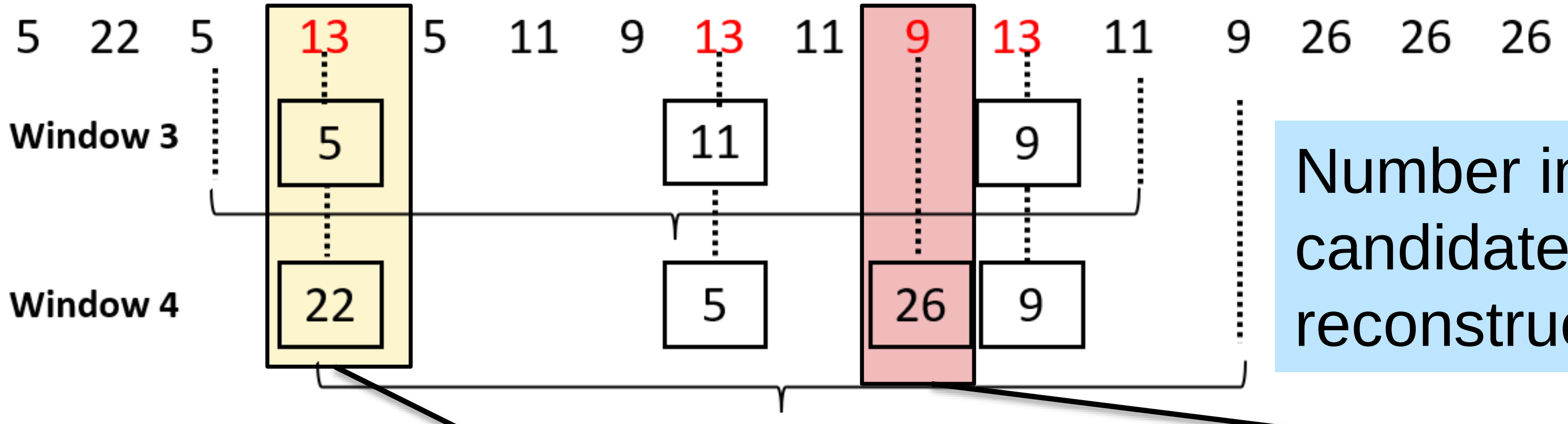
- **Accuracy of Anomaly Detection Needs Further Improvement.**
  - Ability to handle event logs with complex temporal correlation.
    - **DeepLog**’s next event prediction tends to be more frequency based.
  - Alleviate model overfitting.
    - **DabLog** fails to characterize abnormal regions in the latent space.

- **Anomaly Diagnosis Needs More Attention to Improve Interpretability.**

  Our Main Focus

  - Accurate Abnormal Deviation Identification. ***Why abnormal?***
  - How Normal Pattern Should Behave. ***How make correction?***

Motivation for Anomaly Diagnosis (that DabLog's sliding window based reconstruction cannot do)



Abnormal events

Number in black box is Top-N (here N=1) candidate event with the highest reconstruction probability at each time slot.

The abnormal event 13, it is expected to be reconstructed as ***5*** in window 3 but as ***22*** in window 4

Cannot provide consistent diagnosis for potential correction

The tenth event 9, detected as normal in sliding window 3, but detected as abnormal and reconstructed as ***26*** in window 4.

Fail to Identify Abnormal Deviation Accurately

The above inconsistency regards to diagnosis is because each sliding window provides a limited and variable view.

Our Solutions

- ***Key Insights***

  - ***Local*** sliding-window （***small-scale*** view） based anomaly detection can provide better precision, but not suitable for anomaly diagnosis.

  - ***Global*** workflow（***large-scale*** view） based whole sequence alignment can accurately uncover how an anomaly deviates from the “normal” pattern, thus facilitating anomaly diagnosis, understanding and correction.

Why not combine them?

Our sequence alignment based anomaly diagnosis results of the motivating example, as follows:

5 22 5 13 5 11 9 13 11 9 13 11 9 26 26 26

Outline

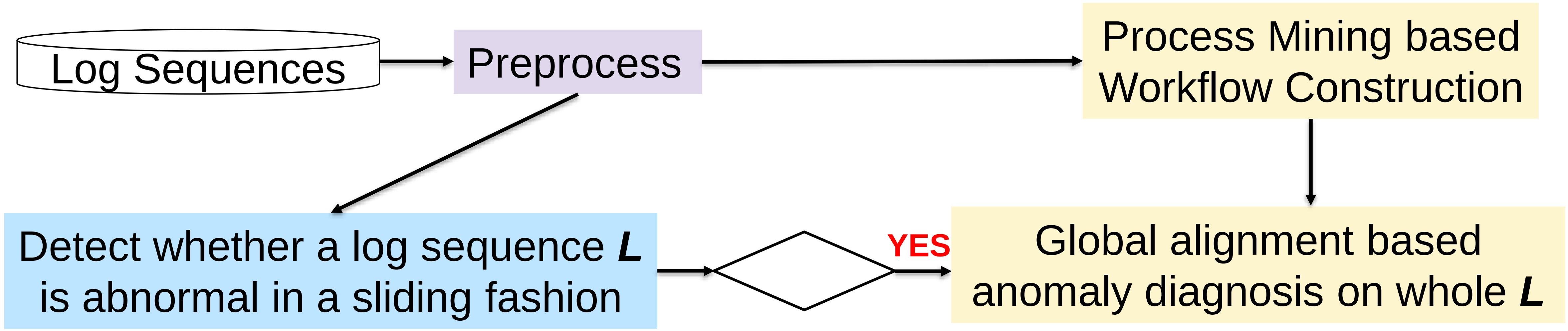
1 Introduction

2 **System Design**

3 Evaluation

Overview

- ***Proposed System Prototype-MADDC***
  - **Three** major separate components, namely ***log preprocess***, ***local anomaly detection***, ***global anomaly diagnosis and correction***.
  - Combine **LSTM-based** Variational AutoEncoder with **Process Mining** Techniques (i.e., process discovery and conformance checking).
- ***Overall Process***



Several Key Definitions

- ***Event Keys***

  - Log entries are parsed using templates and represented by a **number**.

5--*Receiving block* 22--*allocateBlock* 9--*Received block*
13--*Receiving empty packet* 11--*terminating* 26--*is added to*

- ***Event Sequence and Subsequence***

  - A log sequences is transformed into an event sequence using event keys.

Event Sequence 5 5 22 5 9 9 11 11 11 9 26 26 26

Subsequence:
Sliding Window 1

- ***Process***

  - During process mining, we treat each ***event sequence*** s as a ***process***.

- ***Workflow***

  - A workflow is defined at the higher level to characterize **a group of similar processes**, namely closely related cases which may execute similar tasks.

LSTM-VAE based on Local Anomaly Detection

- **Subsequence Reconstruction using LSTM-VAE(Variational Autoencoder)**

Event Subsequence $\{e_1, e_2, \ldots, e_n\}$ → **Encode-Decode** → $\{e'_1, e'_2, \ldots, e'_n\}$

Check event reconstruction at each time slot

- **Double-Check Anomaly Criterions**

Example: [11 13 11 **28**…] → Output the probability of all event keys $\boldsymbol{p(e_j), e_j \in Event\ key\ set}$ being an instance of event at time slot 4.

1) Top-K Rank-based Criterion

Original event at time *slot 4* is **28**.

Check whether is included √

The Top-3 events sorted by reconstruction probability at time *slot 4* is [11,**28**,4]

K=3

2) Probability Threshold-based Criterion

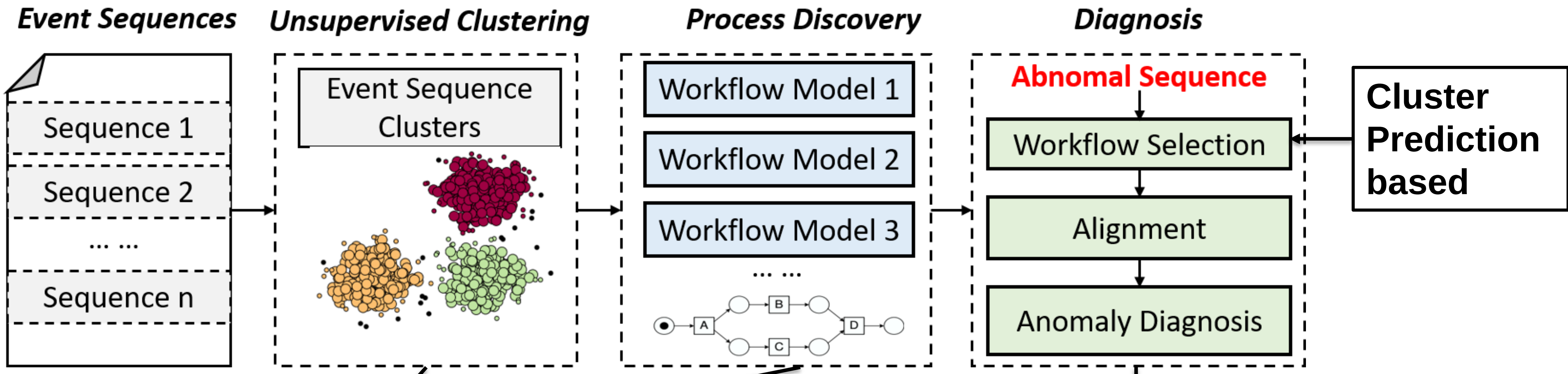
Calculates the occurrence (reconstructed) probability $\boldsymbol{\theta_p}$ of reconstructed event $\boldsymbol{e_j}$ being same as **28**

$\boldsymbol{\theta_p}$ =0.02859 X

Check whether the occurrence (reconstructed) probability $\boldsymbol{\theta_p}$ at slot 4 exceeds a predefined threshold $\boldsymbol{\theta}$.

Process Mining based Global Anomaly Diagnosis

The main principle behind our anomaly diagnosis is to uncover critical differences by comparing the detected abnormal sequence with a collection of similar “normal" ones.



Challenge 1: Unclear “normal” sequence pattern for anomaly diagnosis

Challenge 2: Accurate Abnormal Deviation Identification, avoiding time consuming ‘one-to-many’ sequence comparison

Alignment based Anomaly Diagnosis—An Example

- Goal of alignment
  - Map observed behavior (i.e., event sequence) from logs onto modelled behavior (i.e., workflow model) to derive deviations and conformance on event level.
- Trace based alignment

$$\Upsilon_2 = \begin{array}{c|c|c|c|c|c} a & b & c & \gg & d & e \\ \hline a & \gg & c & b & d & e \end{array}$$

Moves in the log

Moves in the built model.

extra events **missing events**

  - The '»' indicates that either the log could not make the same step as in the model or vice versa, considered as anomaly.
  - Anomaly behaved as **two kinds of** asynchronous moves:
    - Extra events -- cases where the log events makes move, but unallowed by the model.
    - Missing events -- cases where there are moves in model but not in log events.
- Alignment based anomaly correction
  - Based on alignment results, try to correct the anomaly by removing extra events and adding missing events at the corresponding position.

Outline

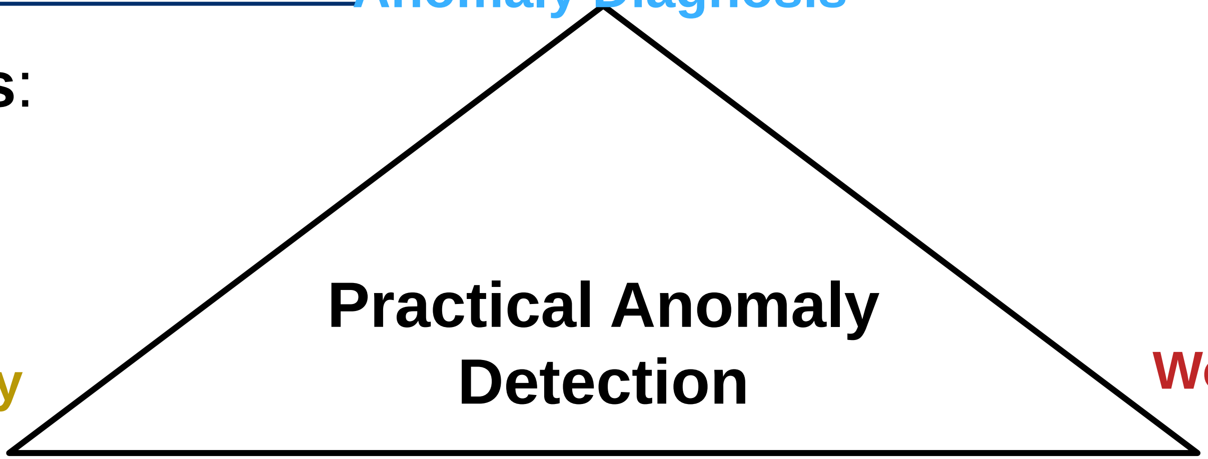
1 Introduction

2 System Design

3 **Evaluation**

Evaluation

- **Research Questions**:



➢ **Question 1**: How better is MADDC in **anomaly detection** when compared with representative reconstruction-based and prediction-based baseline models?

➢ **Question 2**: As a key factor to provide reliable alignment based anomaly diagnosis, what quality are **workflow models** built on clustered sequences?

➢ **Question 3**: How effective does **alignment based anomaly diagnosis** facilitate the anomaly understanding and interpretation?

Experimental Setup

- **Datasets and Models**:
  - **Datasets**: UNSWNB (intrusion detection traffic logs), HDFS system logs.
  - **Models**: DeepLog(CCS 2017), DabLog(AsiaCCS 2021)

Table 1: Statistics Description of Datasets

Dataset	Normal Train		Normal Test		Abnormal Test		Number of Event Keys
	Sequences	Subsequences	Sequences	Subsequences	Sequences	Subsequences	
HDFS-1	4855	61,140	553,366	6,918,652	13,882	198,058	32
HDFS-2	194,115	2,425,217	194,066	2,428,025	13,882	198,058	32
UNSWNB	9900	1,843,301	9900	1,853,201	2245	339,238	291

1. Both HDFS-1 and HDFS-2 are generated from original HDFS Dataset.
2. HDFS-1 is the same dataset used in DeepLog.
3. HDFS-2 is generated using same method as in DabLog.

Accuracy of Anomaly Detection

- **We have reproduced DeepLog on HDFS-1 with very similar performance.**

Table 2: Anomaly Detection Results of Models when π=0.1, θ=0.1

Dataset	Model	FP	FN	P [3]	R [4]	F1
HDFS-2	DeepLog	9927	5838	44.58%	57.76%	50.32%
	DabLog	267	2777	97.65%	80.00%	87.95%
	MADDC	335	895	97.49%	93.55%	95.48%
UNSWNB	DeepLog	2996	196	40.61%	91.27%	56.21%
	DabLog	978	207	67.57%	90.78%	77.48%
	MADDC	27	110	98.75%	95.10%	96.89%

[3] Precision Rate, [4] Recall Rate.

- $\pi = N$ / Number of event keys, ***rank based parameter***
- θ, ***probability threshold-based parameter***
- **FP**: false positives identified through manual check.
  - Due to rare patterns
- **FN**: false negatives identified through manual check.
  - Subsequences pattern that are abundant in the training set

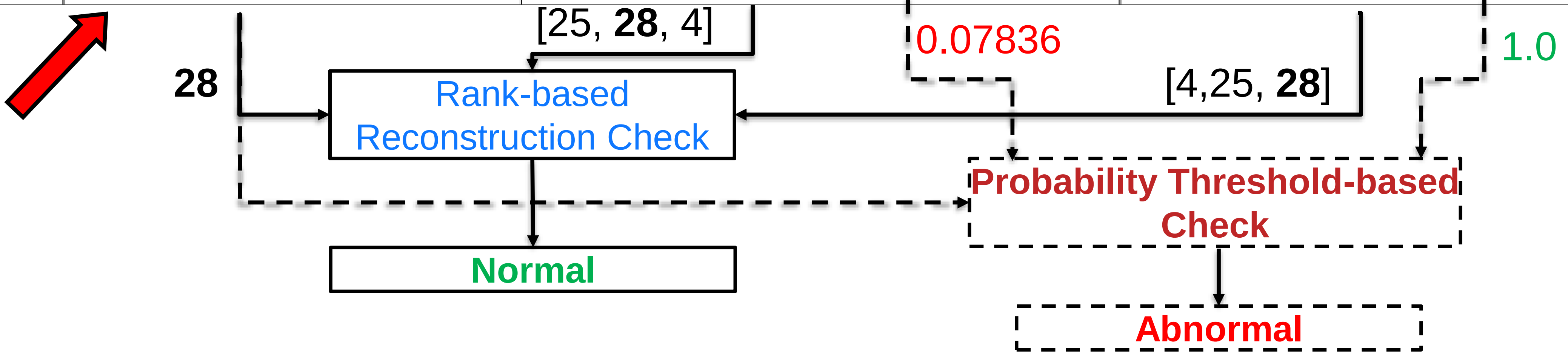
- **Detailed parameter analysis shows the MADDC's stable performance on different dataset with varying parameters.**

Case Study and Advantages on Accurate Anomaly Detection

- **Double-check** based anomaly critic could make full use of their respective advantages.

Table 3: Several MADDC's TP Cases but Dablog's FNs

Dataset	ID	Anomaly Subsequence	MADDC's Reconstruction & Probability	Dablog's Reconstruction & Probability
HDFS-2	1	[13 11 13 11 28 25 25 25 23 23]	[23, **28**, 25]−0.07086	[11, 25, **28**]−1.0
	2	[11 13 11 28 4 25 25 25 23 23]	[25, **28**, 4]−0.07836	[4, 25, **28**]−1.0



- **Different Event Reconstruction Probability** from MADDC and Dablog **(0.07836<<1.0)**: Due to VAE's probabilistic modeling of MADDC, the latent distribution of abnormal data have greater variance.

Effectiveness of Alignment based Anomaly Diagnosis

- Our Alignment is ***consistent*** and ***accurate*** for anomaly diagnosis.
  - Extra events: **7, 17**; Missing events: **16.**
- DeepLog's next event prediction is ***consistent*** but ***inaccurate*** for anomaly diagnosis.
  - Ten detected abnormal events, eight of which are manually confirmed as FPs.
- Dablog's subsequence reconstruction is ***inconsistent*** and ***inaccurate*** for anomaly diagnosis.
  - Event **18** is not abnormal in window 2, but is in next window.
  - Event **18** again, its Top-1's reconstructed event is 5 in window 3 but becomes 3 in the next window.

Table 5: Case Study 1-A HDFS-2 Anomaly Diagnosis

Alignment	5 5 22 5 11 9 11 9 11 9 25 26 26 26 5 18 7 17 $\boxed{16}$ 3 23 23 23 21 21 21
Sliding Fashion Prediction	5 5 22 5 11 9 11 9 11 9 25 26 26 26-23 5 18 7-6 17-16 3-26 23-26 23-21 23-21 21-18 21-25 21-23
Sliding-Fashion Reconstruction	9 11 9 25 26 26 26 5 18 7-16 11 9 25 26 26 26 5 18 7-16 17-6 9 25 26 26 26 5 18-5 7-5 17-16 3 25-5 26-25 26 26 5-26 18-3 7-3 17-3 3 23 26 26 26 5-3 18-3 7-4 17-3 3 23 23 ... ... 26 5-26 18-3 7-4 17-3 3 23 23 23 21 5 18-16 7-3 17-3 3 23 23 23 21 21 18-3 7-3 17-4 3-23 23 23 23 21 21 21

The circled number 4 refers to the Top-1 **predicted** or **reconstructed** event

What We Have Not Talked About

- Unsupervised Characterization of “Normal” Sequence Pattern
    - Why Event Sequences Clustering?
    - What is Process Discovery based Workflow Construction?
- Quality Evaluation of Workflow Model Construction in Unsupervised Manner
- Accuracy Comparison of Alignment based Anomaly Diagnosis.
- Limitations and Future Work.

Please Read Our Paper!

Thank You!

Xiaolei Wang
xiaoleiwang@nudt.edu.cn
https://github.com/040840308/MADDC/tree/master