MICHIGAN STATE UNIVERSITY

NEC LABORATORIES AMERICA, INC.
Relentless passion for innovation

Multi-Scale One-Class Recurrent Neural Networks for Discrete Event Sequence Anomaly Detection

Zhiwei Wang

Joint work with

Zhengzhang Chen, Jingchao Ni, Hui Liu,

Haifeng Chen, Jiliang Tang

The 2021 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Outline

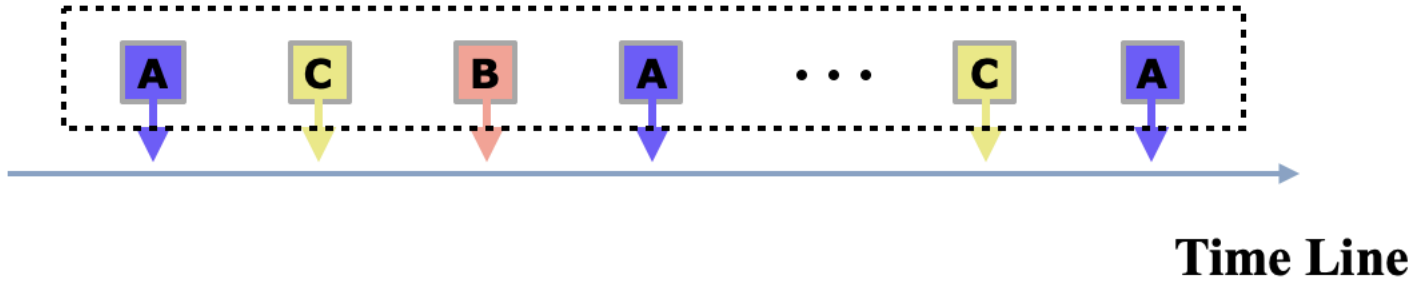
- **Background and Motivation**
- **The Proposed Framework: OC4Seq**
- **Experimental Results and Case Study**
- **Summary**

Outline

- **Background and Motivation**
- **The Proposed Framework: OC4Seq**
- **Experimental Results and Case Study**
- **Summary**

What Is Event Sequence

An even sequence is an ordered list of discrete events



System logs can be represented as event sequences

```
081109 213506 2421 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3509323198988774369 src: /10.250.6.
081109 213510 2384 INFO dfs.DataNode$PacketResponder: PacketResponder 0 for block blk_9093049293972551787
081109 213837 19 INFO dfs.FSDataset: Deleting block blk_1781953582842324563 file /mnt/hadoop/dfs/data/curr
081109 213847 2552 INFO dfs.DataNode$DataXceiver: 10.251.194.213:50010 Served block blk_-77247134689121665
081109 213907 2497 INFO dfs.DataNode$DataXceiver: 10.251.91.229:50010 Served block blk_-335844855391866590
081109 213908 2549 INFO dfs.DataNode$DataXceiver: 10.251.39.192:50010 Served block blk_-534199272975558457
081109 214009 2594 INFO dfs.DataNode$DataXceiver: 10.250.5.237:50010 Served block blk_3166960787499091856
081109 214043 2561 WARN dfs.DataNode$DataXceiver: 10.251.30.85:50010:Got exception while serving blk_-2918
```

A

B

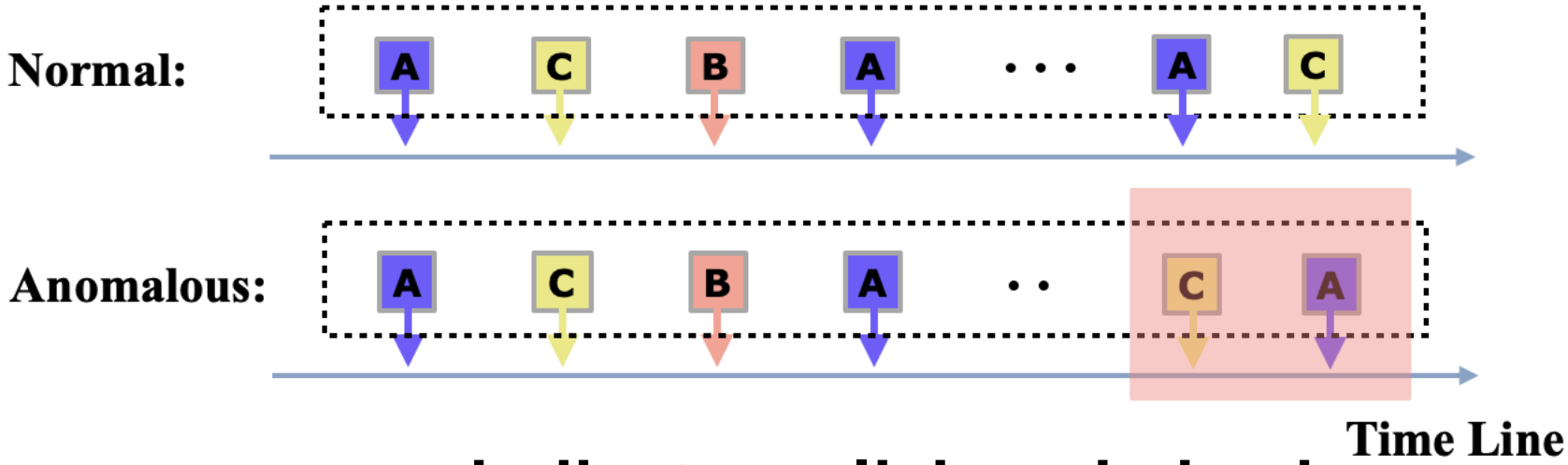
C

Event Sequences Are Ubiquitous

- Control Commands
  - Symbol sequences that arise from recordings of switch sensors in cockpits of commercial airliners
- System Logs
  - Sequence of system calls executed by a computer program
- Transaction
  - Customer purchases in e-commerce website
- Genetics
  - DNA in biological systems

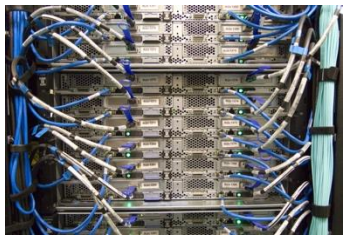
Event Sequences Indicate the State of Systems

An anomalous event sequence deviates from normal ones



Anomalous sequences indicate malicious behaviors

Bank Fraud

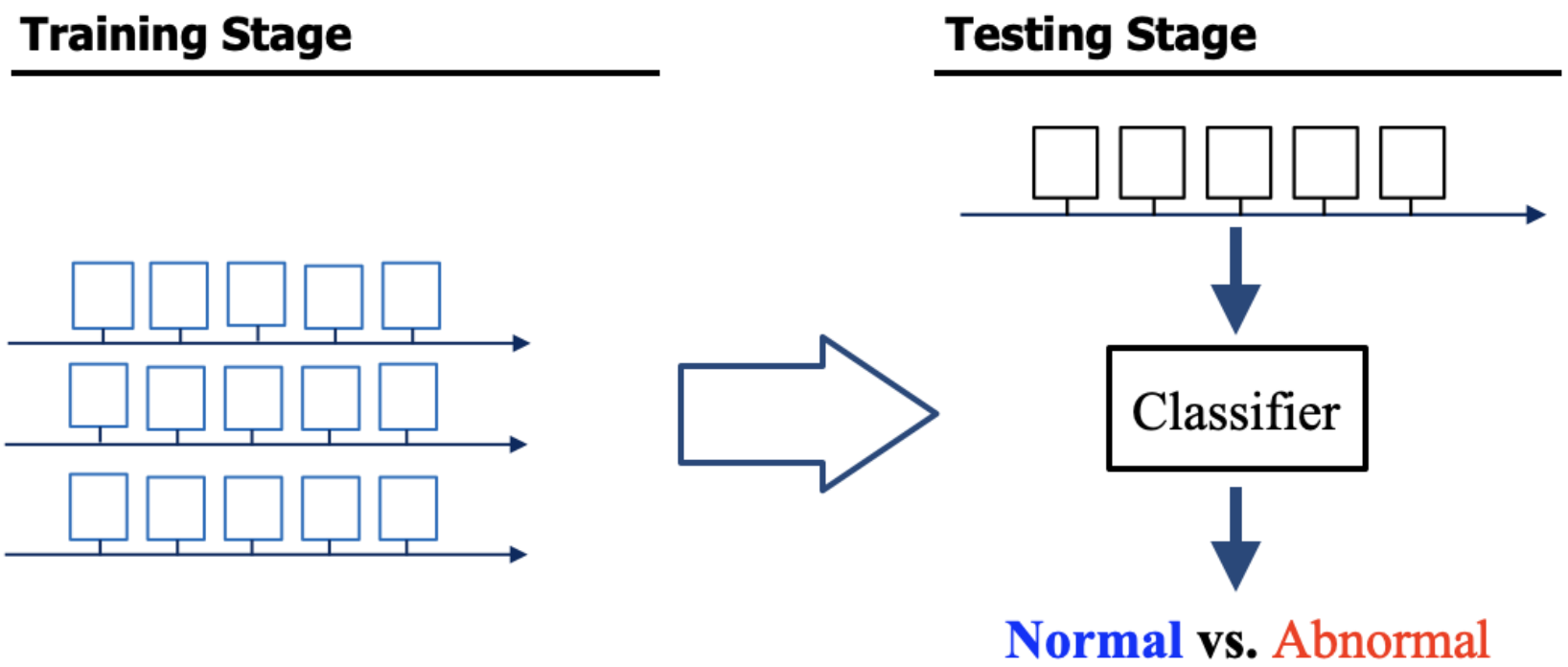


System Fault

Internet Intrusion

Event Sequences Anomaly Detection Problem

Given a set of sequences $\mathcal{S} = \{S^1, S^2, \cdots, S^N\}$, *where each sequence* S^i *is normal, we aim to design a one-class classifier that is able to identify whether a new sequence* S *is normal or not by capturing the underlying multi-scale sequential patterns in* $\mathcal{S}$.



Existing Approaches

- **Traditional methods**
  - **Feature extraction based**
  - **Distance based**
- **Deep learning methods**
  - **Language Models**
  - **Auto-Encoder**

- **Ignore the sequential information**
- **Hard to define proper distance**

- **Over sensitive to local patterns**
- **Hard to train**

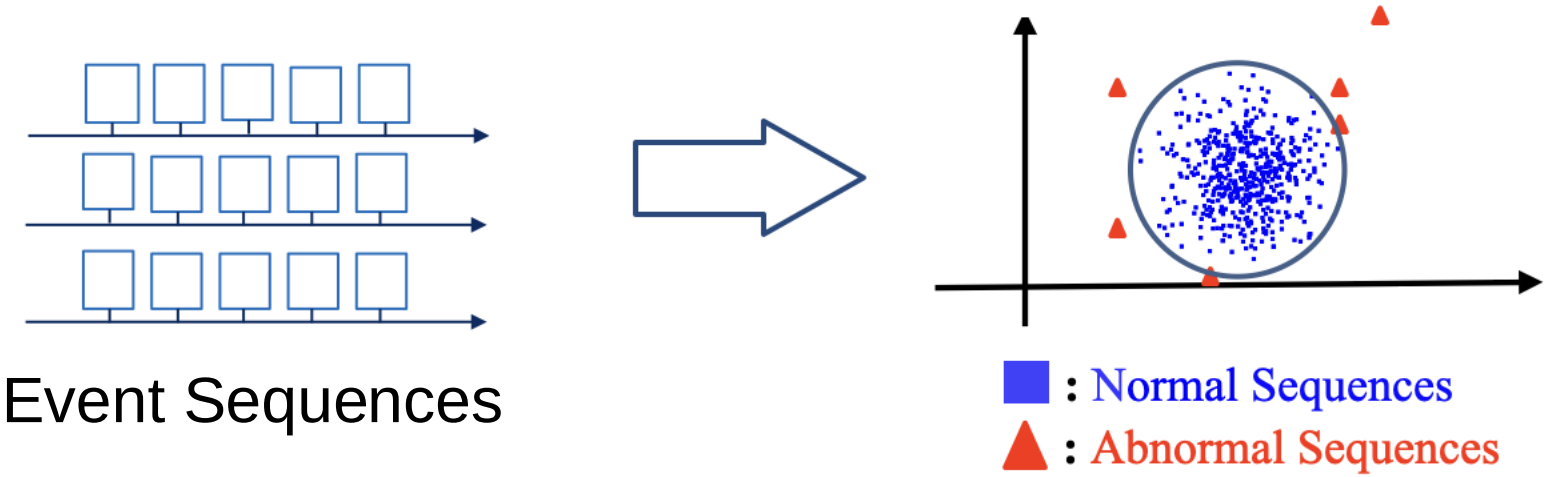
We need a new approach for event sequence anomaly detection!

In this work, we propose OC4Seq, a one-class recurrent network classifier for event sequence anomaly detection.

Outline

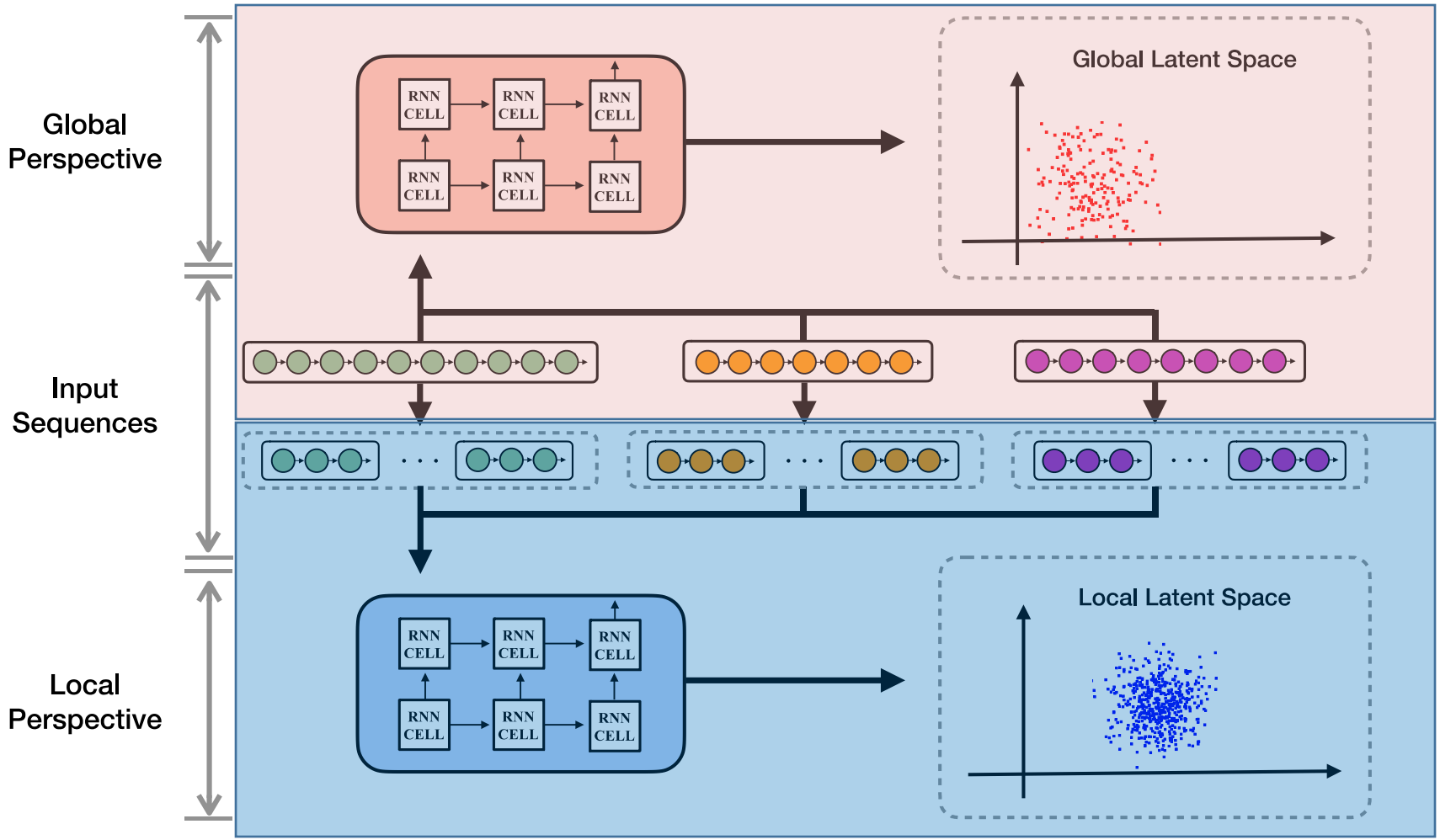
- Background and Motivation
- The Proposed Framework: OC4Seq
- Experimental Results and Case Study
- Summary

The Key Idea of OC4Seq

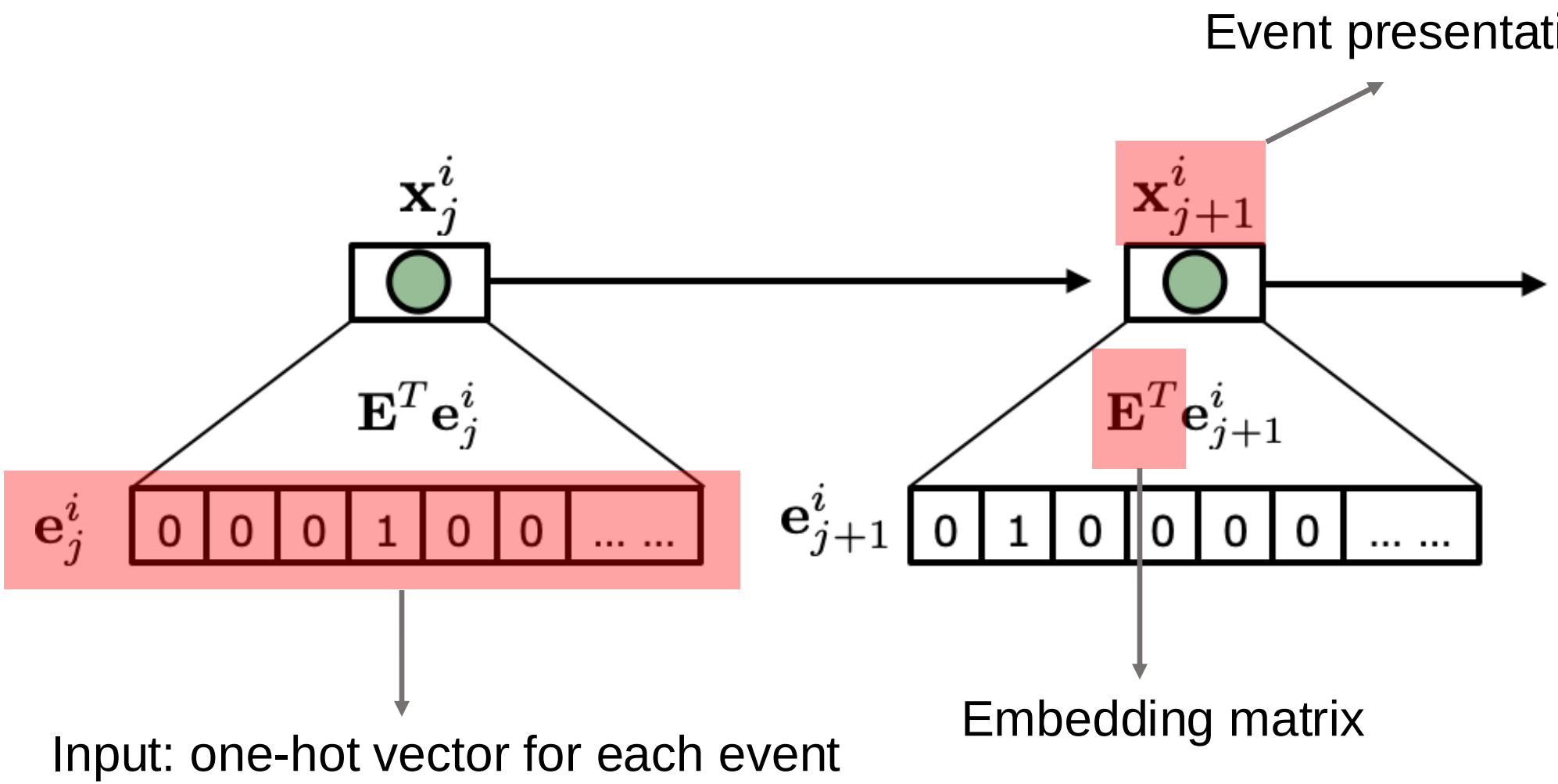


The assumption: in some latent spaces, normal sequences lie together and far from abnormal sequences

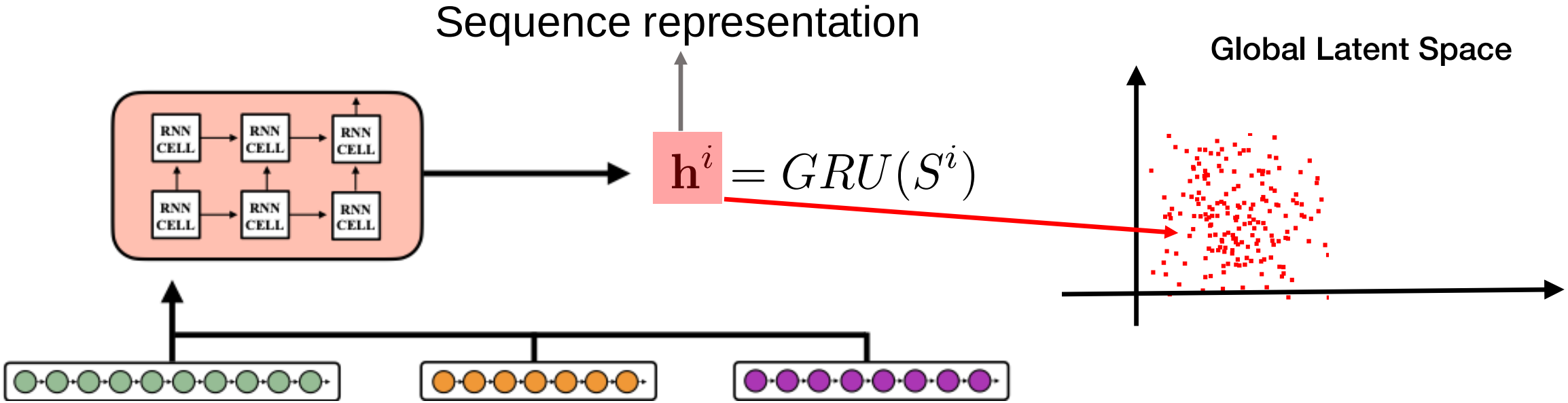
Overview of OC4Seq



The Embedding Layer: Representing Events



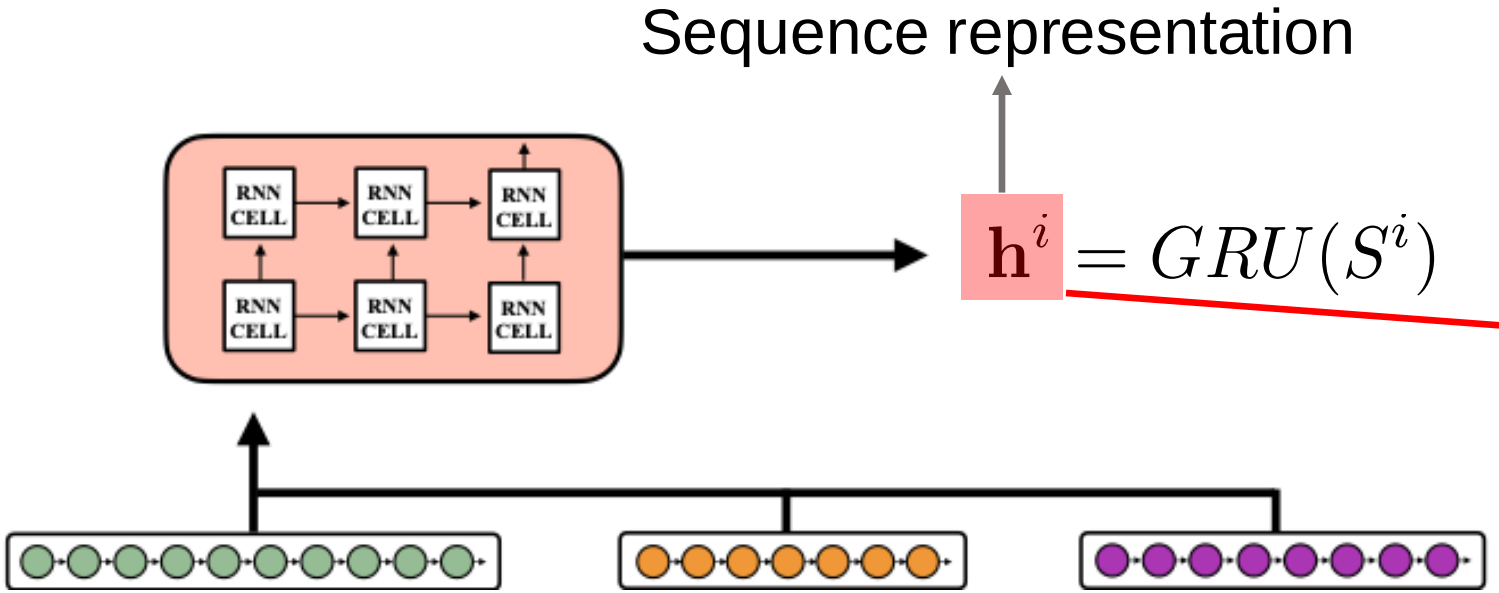
Representing Sequences from Global Perspective



$$S^i = (\mathbf{x}_1^i, \mathbf{x}_2^i, \cdots, \mathbf{x}_{N^i}^i)$$

Input sequence

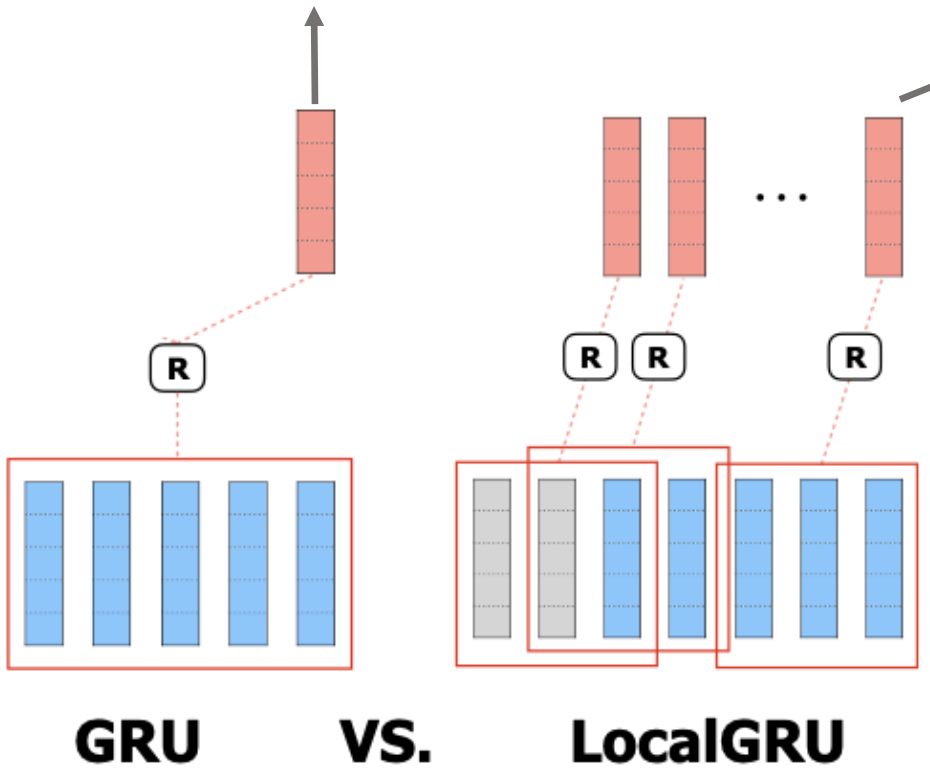
Representing Sequences from Global Perspective



Representing Sequences from Local Perspective

Global representation

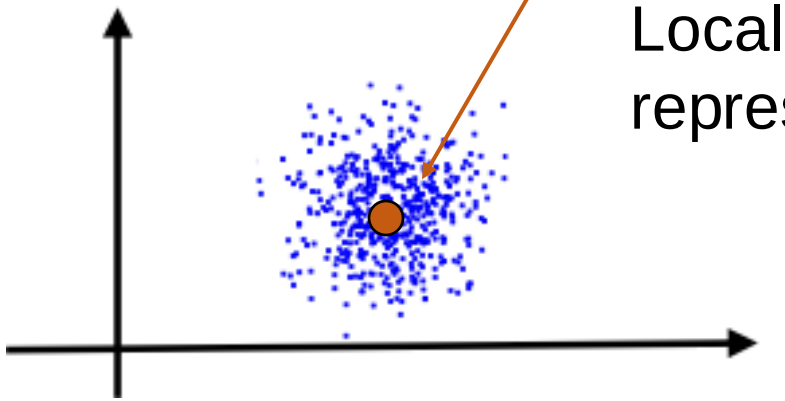
Local region representation



$$\mathbf{h}_1^i, \mathbf{h}_2^i, \cdots, \mathbf{h}_{N^i-M}^i = LocalGRU(S^i)$$

$$\mathcal{L}_{local} = \min_{\mathbf{\Theta}^L} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N^i} \|\mathbf{h}_j^i - \mathbf{c}^L\|^2 + \lambda \|\mathbf{\Theta}^L\|_F^2$$

Local center point representation



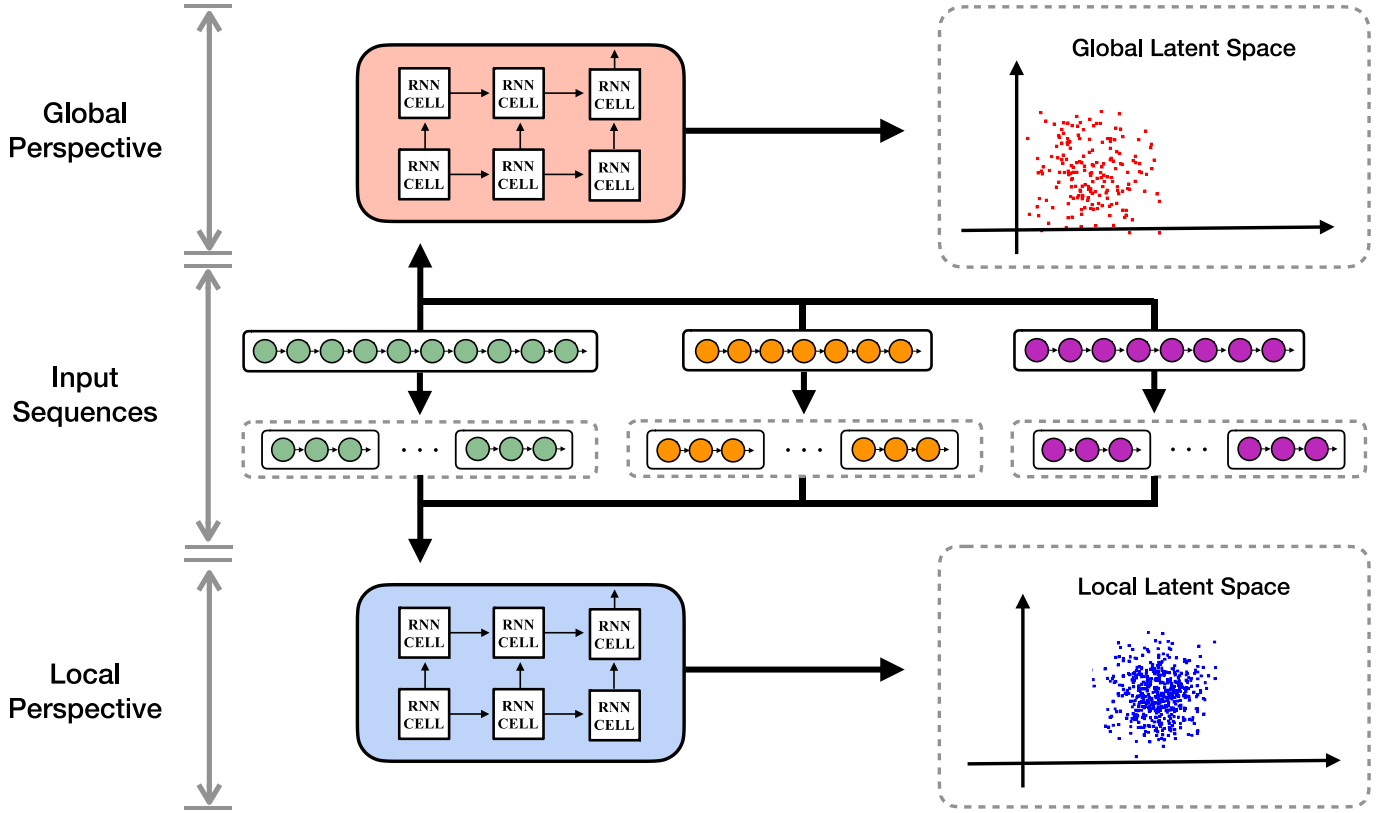
$$\min_{\mathbf{\Theta}^L, \mathbf{\Theta}} \mathcal{L} = \mathcal{L}_{global} + \alpha \mathcal{L}_{local}$$

Control the contribution of local perspective

Recap: OC4Seq Framework

$$\min_{\Theta^L, \Theta} \mathcal{L} = \mathcal{L}_{global} + \alpha \mathcal{L}_{local}$$

Control the contribution of local perspective



Outline

- Background and Motivation
- The Proposed Framework: OC4Seq
- Experimental Results and Case Study
- Summary

Experiment Setup

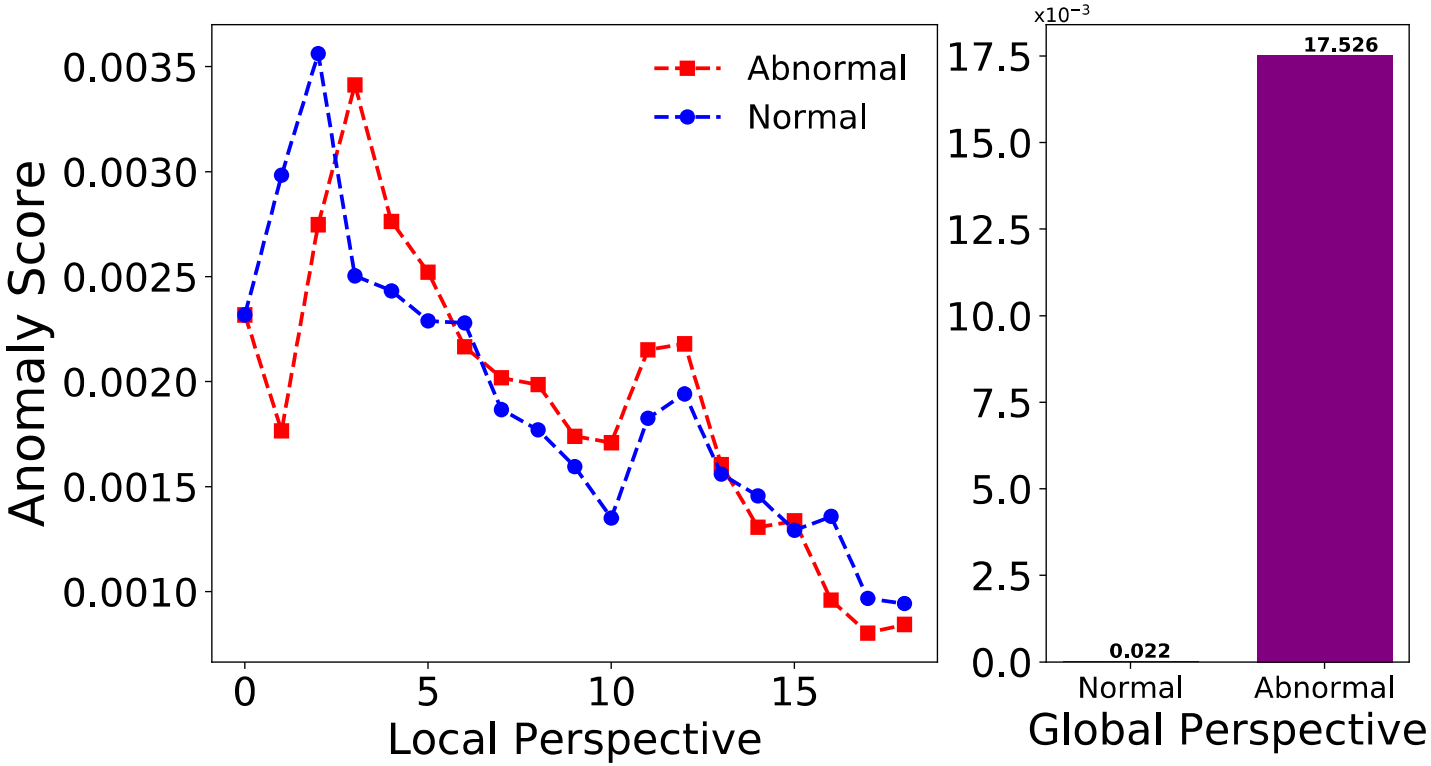
- **Baselines:**
  - Principle Component Analysis (PCA)
  - Invariant Mining (IM)
  - One-Class SVM (OC-SVM)
  - DeepLog state-of-the-art
- **Datasets:**
  - RUBiS: web server logs
  - HDFS: cloud Hadoop system logs
  - BGL: supercomputer system logs
- **Evaluation Metrics:**
  - F1-Score
  - Recall
  - Precision

Model Comparison Results

Methods	HDFS			RUBiS			BGL		
	F-1 score	Precision	Recall	F-1 score	Precision	Recall	F-1 score	Precision	Recall
OC-SVM	0.509	0.622	9.431	0.351	0.220	0.869	0.336	0.215	0.764
PCA	0.634	**0.968**	0.471	0.784	0.862	0.718	0.423	0.269	0.993
Invariant Mining	0.943	0.893	**1.000**	0.912	0.841	**0.996**	0.428	0.273	**1.000**
DeepLog	0.941	0.952	0.930	0.935	0.885	0.992	0.326	0.196	0.980
OC4Seq	**0.976**	0.955	0.998	**0.985**	**0.987**	0.983	**0.747**	**0.704**	0.795

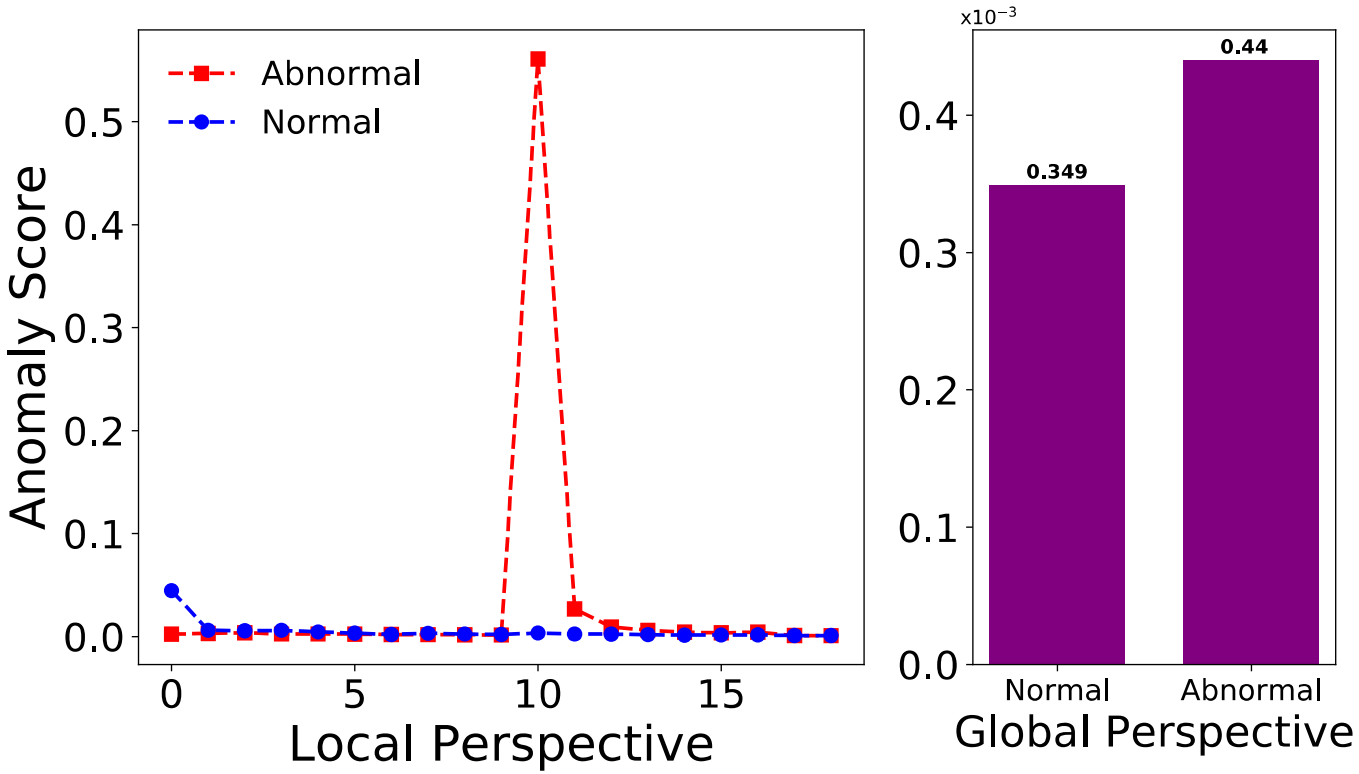
OC4Seq outperforms all baseline methods.

Case Study: Global Perspective Contribution



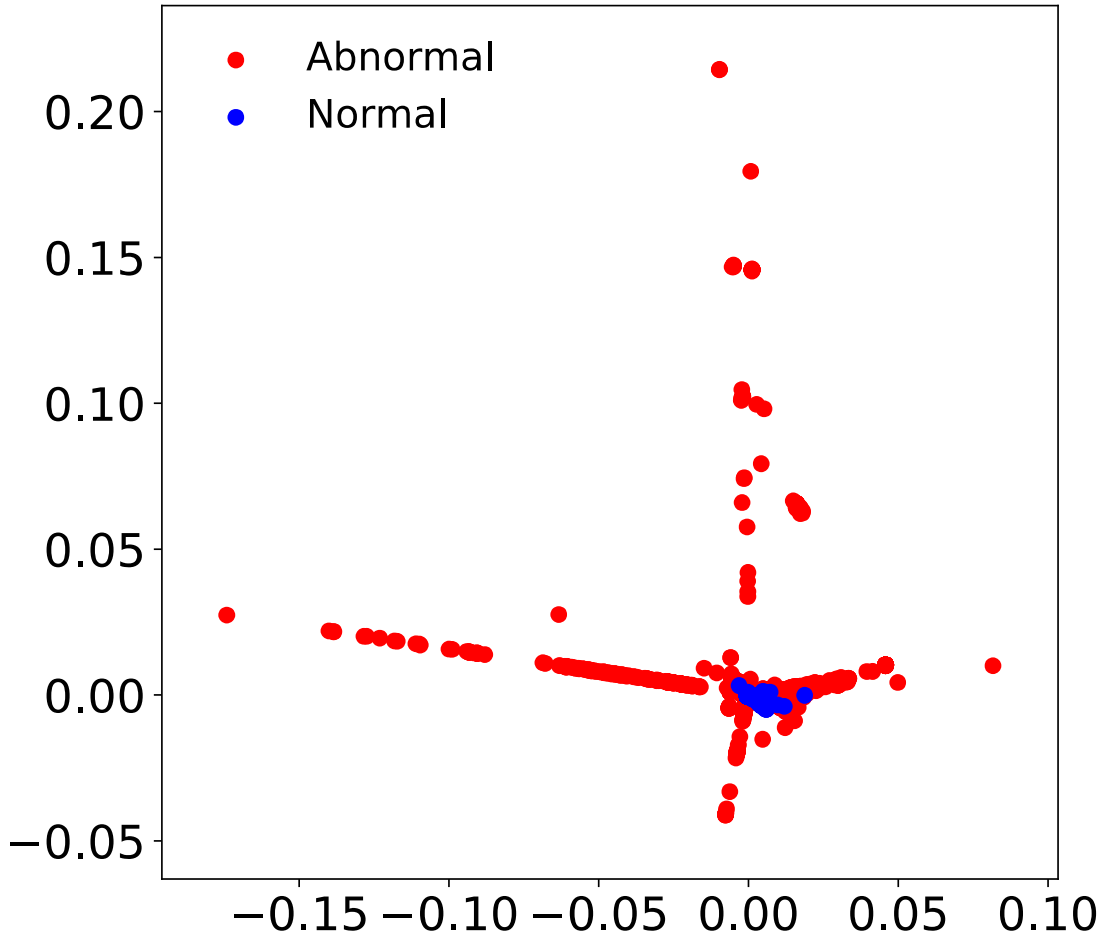
The global perspective is very important

Case Study: Local Perspective Contribution



The local perspective can be important too!

Visualization of Global Representations



OC4Seq can effectively map the normal data into an ideal latent space

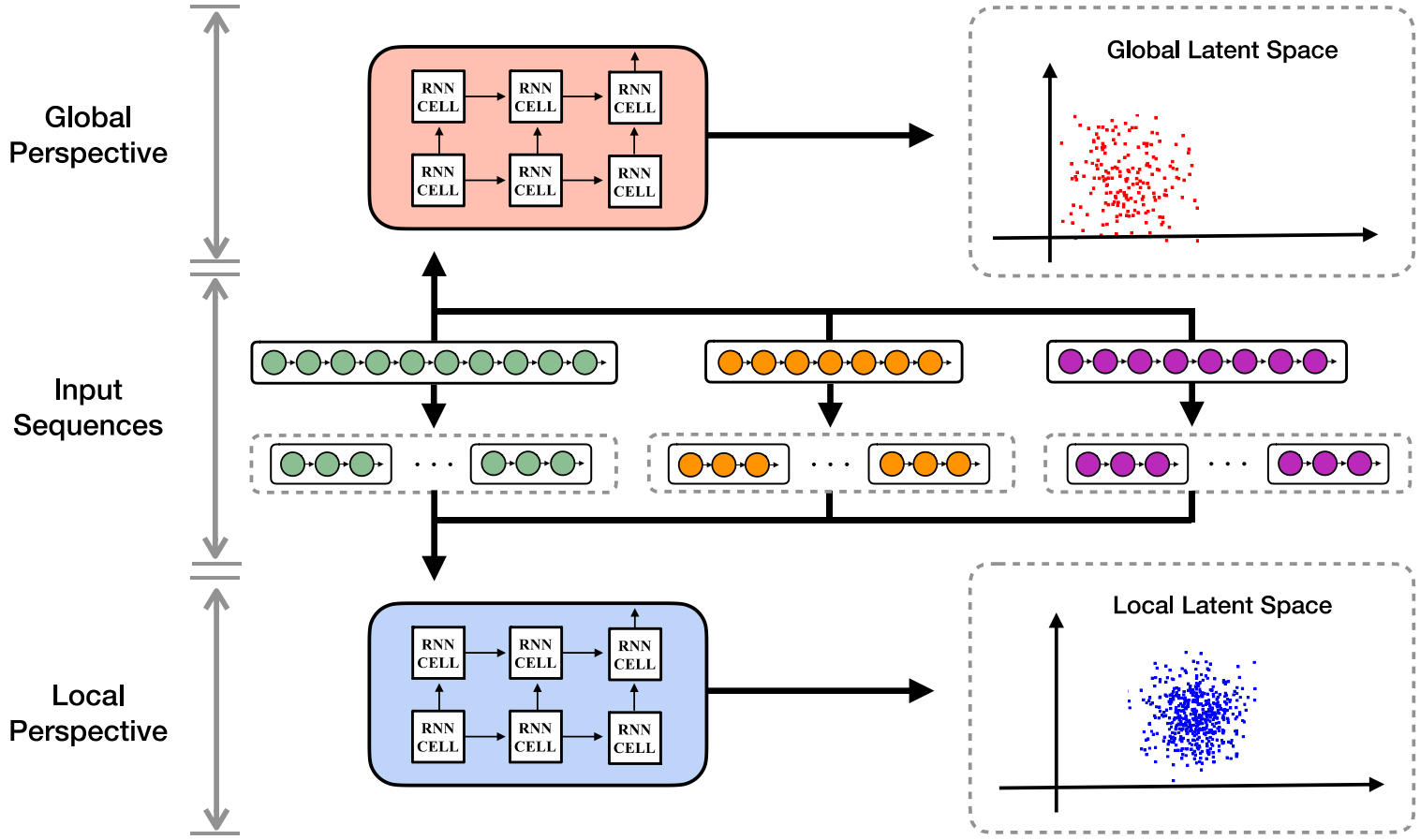
Outline

- Background and Motivation
- The Proposed Framework: OC4Seq
- Experimental Results and Case Study
- Summary

Summary

- We describe the **first attempt** to incorporate a deep one-class classifier for the event sequence anomaly detection task
- We identify the importance of combining **both global and local** perspectives for sequence anomaly detection
- OC4Seq can be trained in an **end-to-end** manner and constantly achieves **superior** detection performance than representative and state-of-the-art methods

Thanks


Global Perspective
Input Sequences
Local Perspective
RNN CELL
Global Latent Space
Local Latent Space