Detecting Anomalies in Multivariate data sets with Switching Sequences and Continuous Streams

Workshop at the Twenty-Third Annual Conference on Neural Information Processing Systems (NIPS 2009), Whistler, BC, Canada, December 11-12, 2009.

Santanu Das, Bryan Matthews, Kanishka Bhaduri,
Nikunj Oza and Ashok Srivastava

NASA Ames Research Center

Problem Statement

- Given a set of discrete variables **D**, and a discrete test case d_t, find the similarity score of d_t given **D**.
  - **D is assumed to be (mostly) normal.**

Problem Statement

- Given a set of discrete variables **D**, and a discrete test case d_t, find the similarity score of d_t given **D**.
  - **D is assumed to be (mostly) normal.**
- Given a set of discrete variables **D** and a set of continuous data streams **C** and a test case $S_t \rightarrow \{d_t, c_t\}$ find the similarity score of S_t given {**D,C**}.

Problem Statement

- Given a set of discrete variables **D**, and a discrete test case d_t, find the similarity score of d_t given **D**.
  - **D is assumed to be (mostly) normal.**
- Given a set of discrete variables **D** and a set of continuous data streams **C** and a test case $S_t \rightarrow \{d_t, c_t\}$ find the similarity score of S_t given {**D,C**}.

- Developing methods that detect/diagnose problems that occur in the **interaction** between discrete variables **D** and continuous data streams **C**.

Different Formats of Discrete

– **Binary** (0/1)

– **On/Off transitions** (1/-1 and rest 0-s)

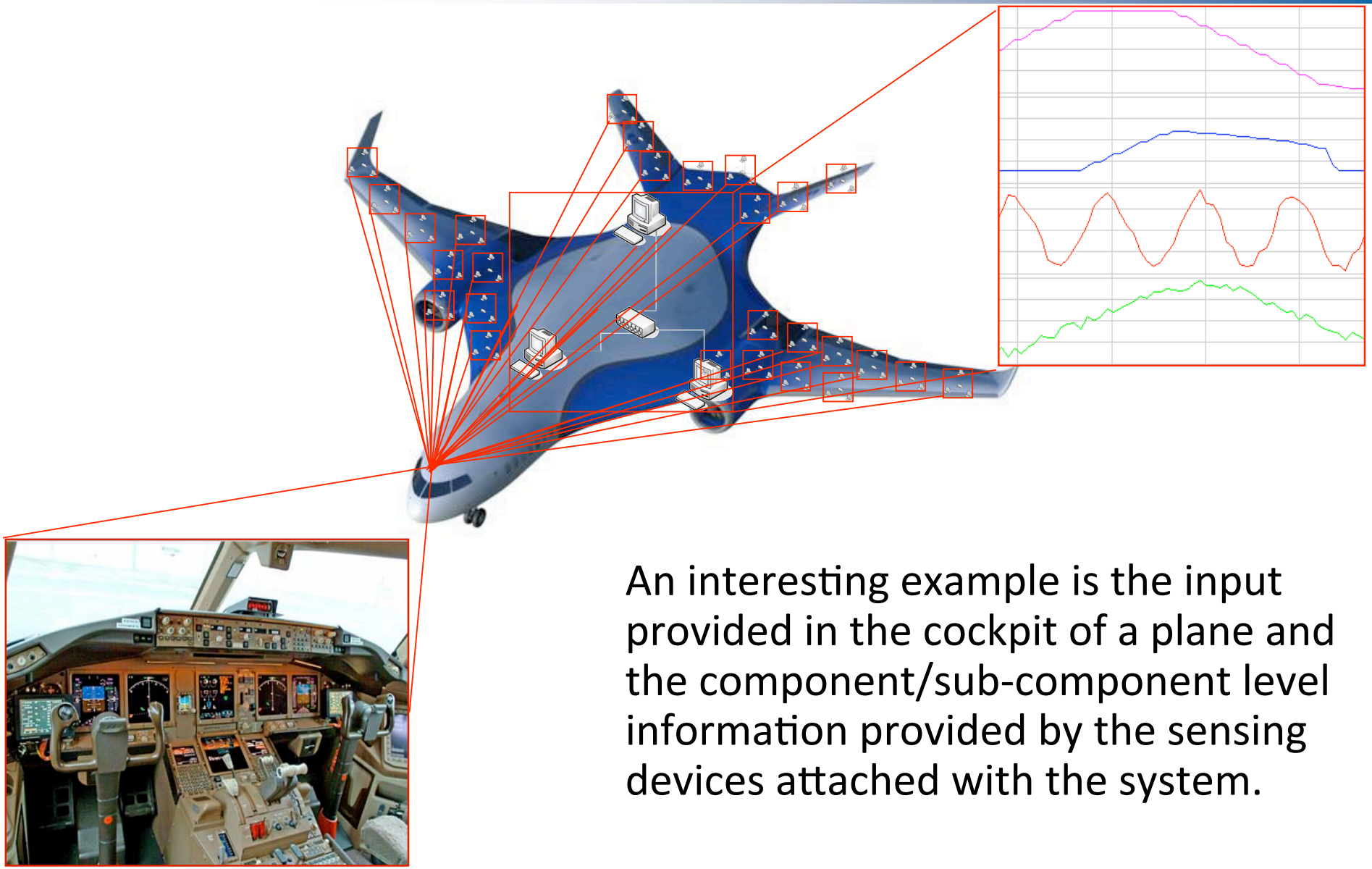
– **Sequences** (events represented by unique characters)

- Developing methods that detect/diagnose problems that occur in the **interaction between sequences** of discrete variables **D** and continuous data streams **C**.

Motivation

An interesting example is the input provided in the cockpit of a plane and the component/sub-component level information provided by the sensing devices attached with the system.

Motivation



An interesting example is the input provided in the cockpit of a plane and the component/sub-component level information provided by the sensing devices attached with the system.

Pattern Recognition

Nonlinear Methods:

- Artificial Neural Network, Multilayer perceptron, Kernel based methods, Support vector Machines, etc.

Distance based Methods:

- K Nearest Neighborhood, Orca, IMS, Sequence miner, etc.

State-of-the-art

Data-driven based anomaly detection

[Schwabacher et al., Iverson, Budalakoti et al., etc.]

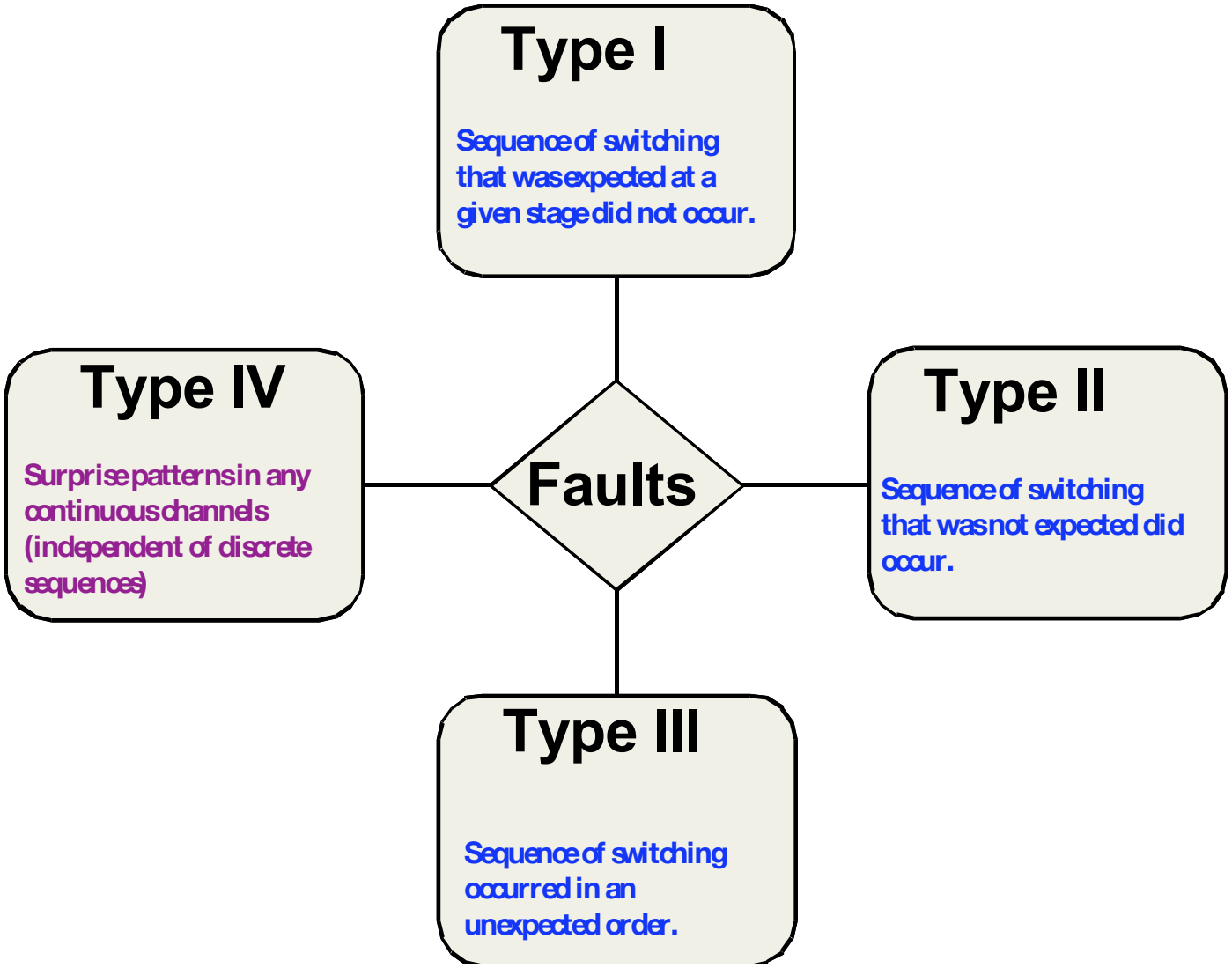
Orca: Distance based outlier detector that works on both discrete (binary) and continuous data.

Sequence Miner: Clustering technique that detects anomalous switching patterns in discrete sequences.

However, we are interested in a model that learns,

- **Normal switching patterns directly from discrete data streams.**
- **Influence of discrete input to system dynamics.**
- **Delayed interactions between the sequences and the continuous system.**

Operationally Significant Anomalies



Fault Examples

- Type 1 – (Missing event) Flaps were not extended to normal full deployment at landing.

- Type 2 - (Extra event) Landing gear was retracted after being deployed on final approach.

Fault Examples Con't

- Type 3 – (Out of order event) Gear deployed before initial flaps below flaps limit.

- Type 4 – (Continuous anomaly) High bank angles or rate of descent below 1,000 ft.

Approach

Vector space models like **Orca** can detect type-2 and 4 anomalies.

		Type-1	Type-2	Type-3	Type-4
Vector Space	Orca				
	Sequence Miner				
	Proposed Method				

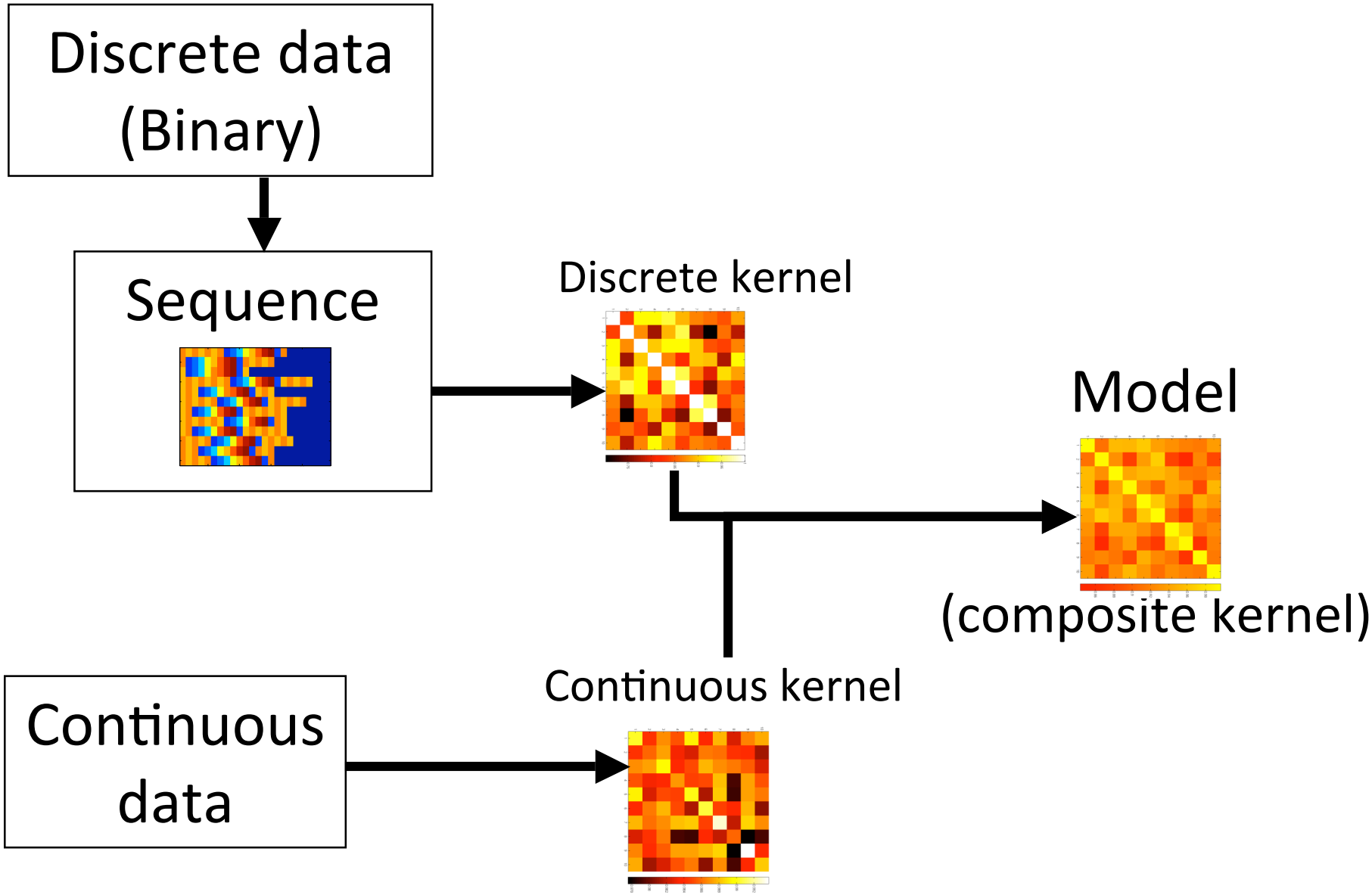
Vector space models can never detect type-3 anomalies as they often have time information embedded.

Methodology

- To detect anomalies in heterogeneous data sets, we attempted to develop a kernel-based approach.

- Experimented on this methodology with a synthetic dataset that has some of the anomalies of interest.

Outline of Approach



Pair wise Similarity Measure

Kernel on discrete : Normalized Longest Common Subsequence (NLCS)

$$K_d(f_i, f_j) = \frac{L(h(s_i, s_j))}{\sqrt{L(s_i) \times L(s_j)}}$$

where the functions $h(.)$ and $L(.)$ calculate the longest common subsequence and length of any sequential data respectively.

Kernel on continuous : Normalized Radial Basis Function (NRBF)

$$K_c(f_i, f_j) = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} e^{\frac{-\|x_m - x_n\|^2}{2\sigma}}$$

To act as a Kernel Function..

The kernel function must satisfy,

	Symmetric	Positive-semidefinite
NRBF	x	x
NLCS	x	x

If $K_d : \chi \times \chi \rightarrow \Re$ and $K_c : \chi \times \chi \rightarrow \Re$ are Mercer kernels then K is also a Mercer Kernel where,

$$K = \alpha_d K_d + \alpha_c K_c \qquad \forall \{\alpha_c, \alpha_d\} \in \Re_+$$

One-class Support Vector Machines

One-class SVMs algorithm (Schölkopf et al.) perform anomaly detection into a much higher dimensional space. The algorithm,

- solves a convex and quadratic optimization problem.
- results a model that can be used to classify new examples.
- enables using non-linear kernel functions to learn complex separating planes.
- **<u>can appropriately introduce a mixture of kernels in the convex cost function</u>.**

$$\underset{\min}{Q} = \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j\left(\beta K_d\left(f_i, f_j\right) + \left(1-\beta\right)K_c\left(f_i, f_j\right)\right)$$

Subject to:

$$0 \le \alpha_i \le \frac{1}{l\nu}, \quad \nu \in [0,1], \quad \sum_i \alpha_i = 1$$

Synthetic Data

- Training on N flights
- Testing on M flights
- Each flight has T time stamps
- Each flight has total K channels (D discrete + C continuous)
- Combinations of discrete parameters influence the continuous channels

Discrete data: Snapshot

V7 was flipped randomly

V4 and V5 never changes

V2, V3, V6 and V8-V10 change state when V1 changes

Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
3	0	0	0	0	1	0	1	0	1	0
4	0	0	0	0	1	0	1	0	1	0
5	0	0	0	0	1	0	0	0	1	0
6	1	1	1	0	1	1	0	1	0	1
7	0	1	1	0	1	1	0	1	0	1
8	0	1	1	0	1	1	0	1	0	1
9	0	1	1	0	1	1	0	1	0	1

Switching Sequences

Each row corresponds to a flight with a sequence of switching profile

Similar sequence

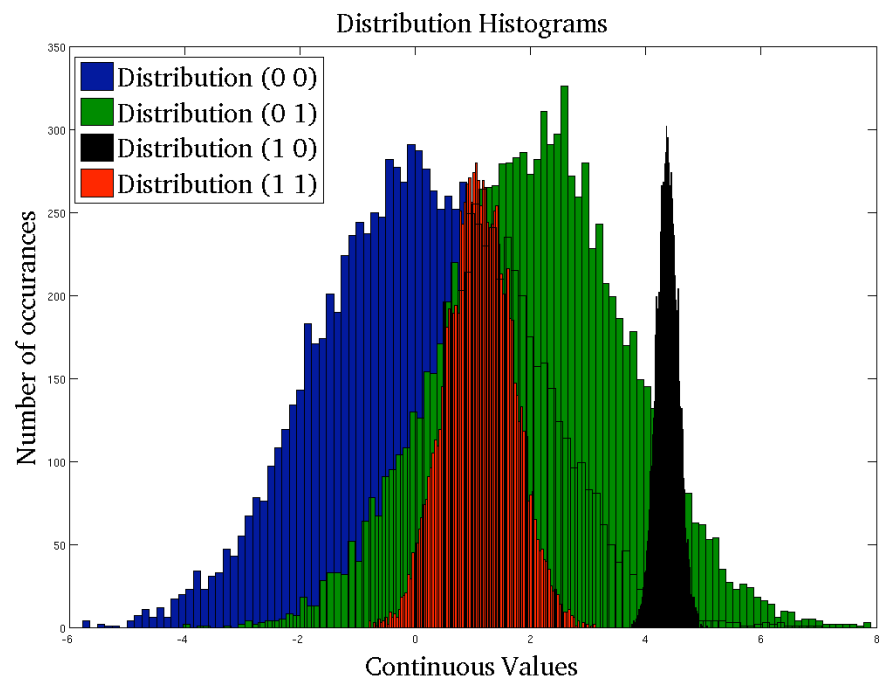
Flight 1	3	5	11	15	18	19	2	14	13						
Flight 2	1	3	5	11	15	18	19	2	14	13	14				
Flight 3	1	3	5	11	14	15	18	19	2	13	14				
Flight 4	14	13	1	3	5	11	14	15	18	19	2	13	14	13	
Flight 5	13	14	1	3	5	11	15	18	19	2	13	14	13		
Flight 6	13	14	1	3	5	11	13	15	18	19	2	14	13		
Flight 7	13	14	13	1	3	5	11	14	15	18	19	2	13	14	13
Flight 8	14	13	14	1	3	5	11	13	15	18	19	2	14	13	14
Flight 9	14	1	3	5	11	13	15	18	19	2					
Flight 10	14	13	1	3	5	11	14	15	18	19	2	13	14	13	

Dissimilar sequence

Continuous data profile

- Continuous channel values were based on the discrete channels states.

- If a continuous channel was influenced by 2 discrete channels, data points were drawn from 4 different distributions.

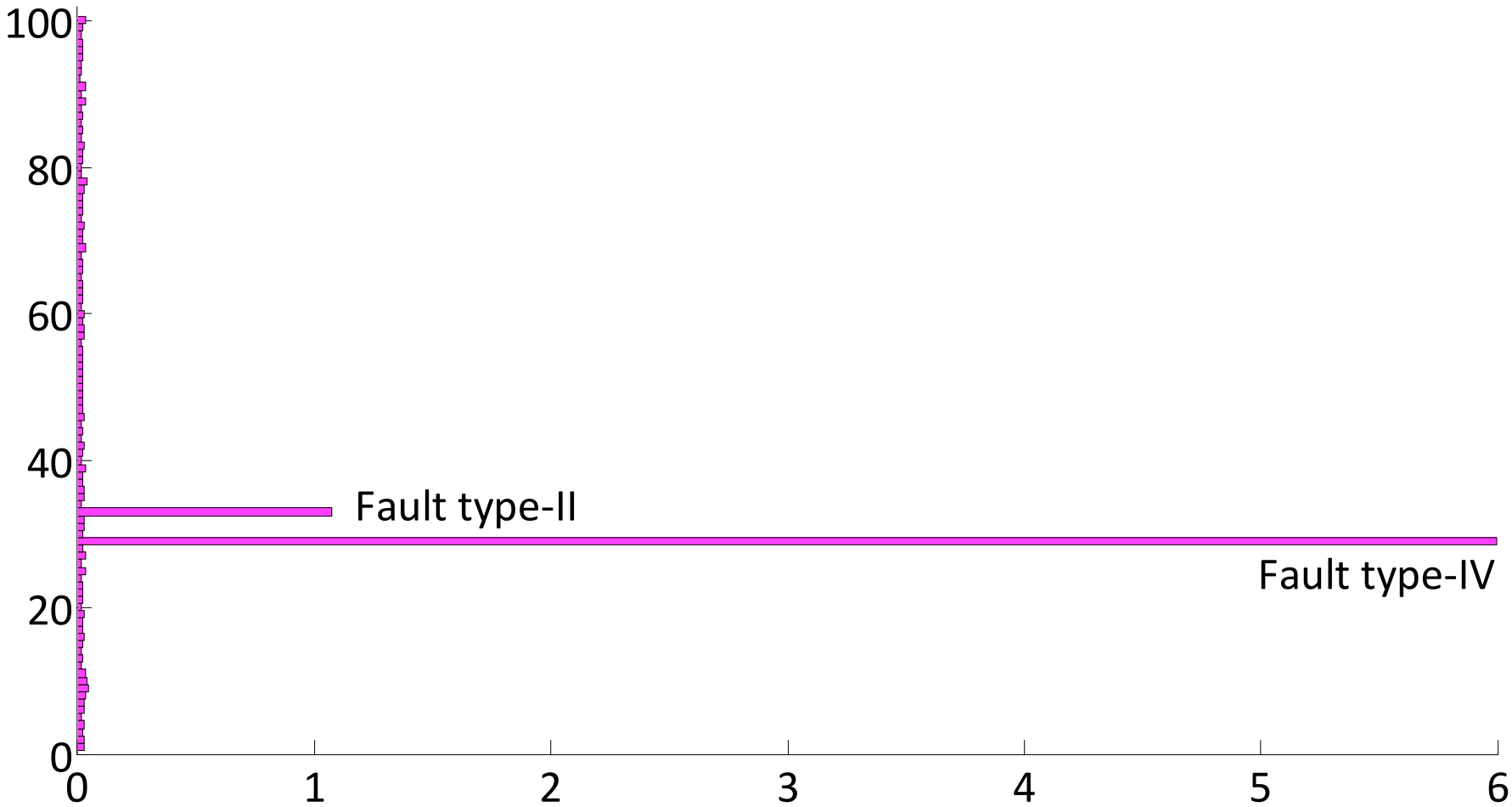


- Whenever the discrete channels changed states the continuous channels switched to the corresponding distribution for the discrete state.

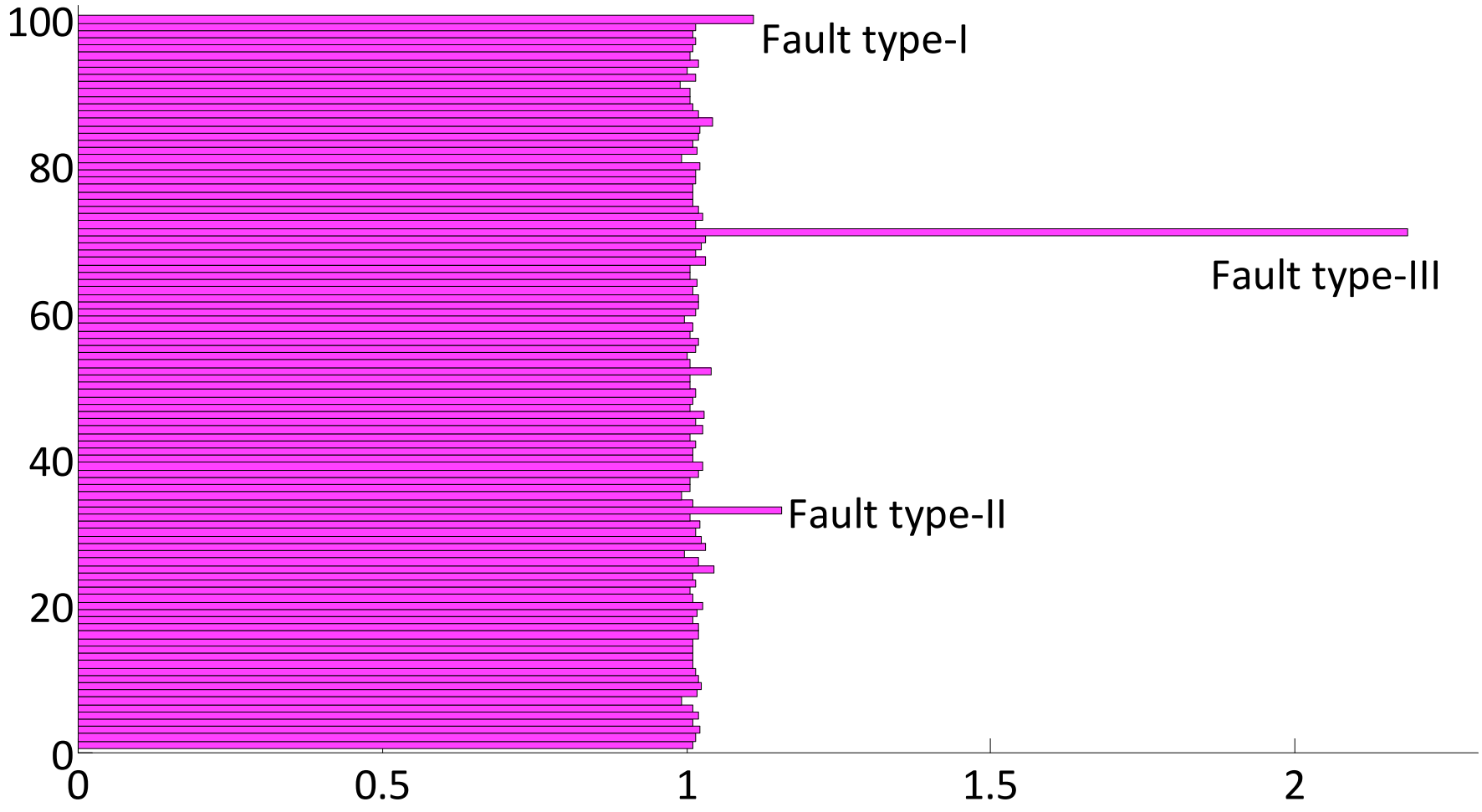
Experiment

- Each flight has 200 instances (time stamps)
- Total 250 flights for training and 100 flights for testing
- Fault Type I: injected in test flight 100
- Fault Type II: injected in test flight 33
- Fault Type III: injected in test flight 71
- Fault Type IV: injected in test flight 29
- Nu: 0.05
- Weights of each kernel: 0.5

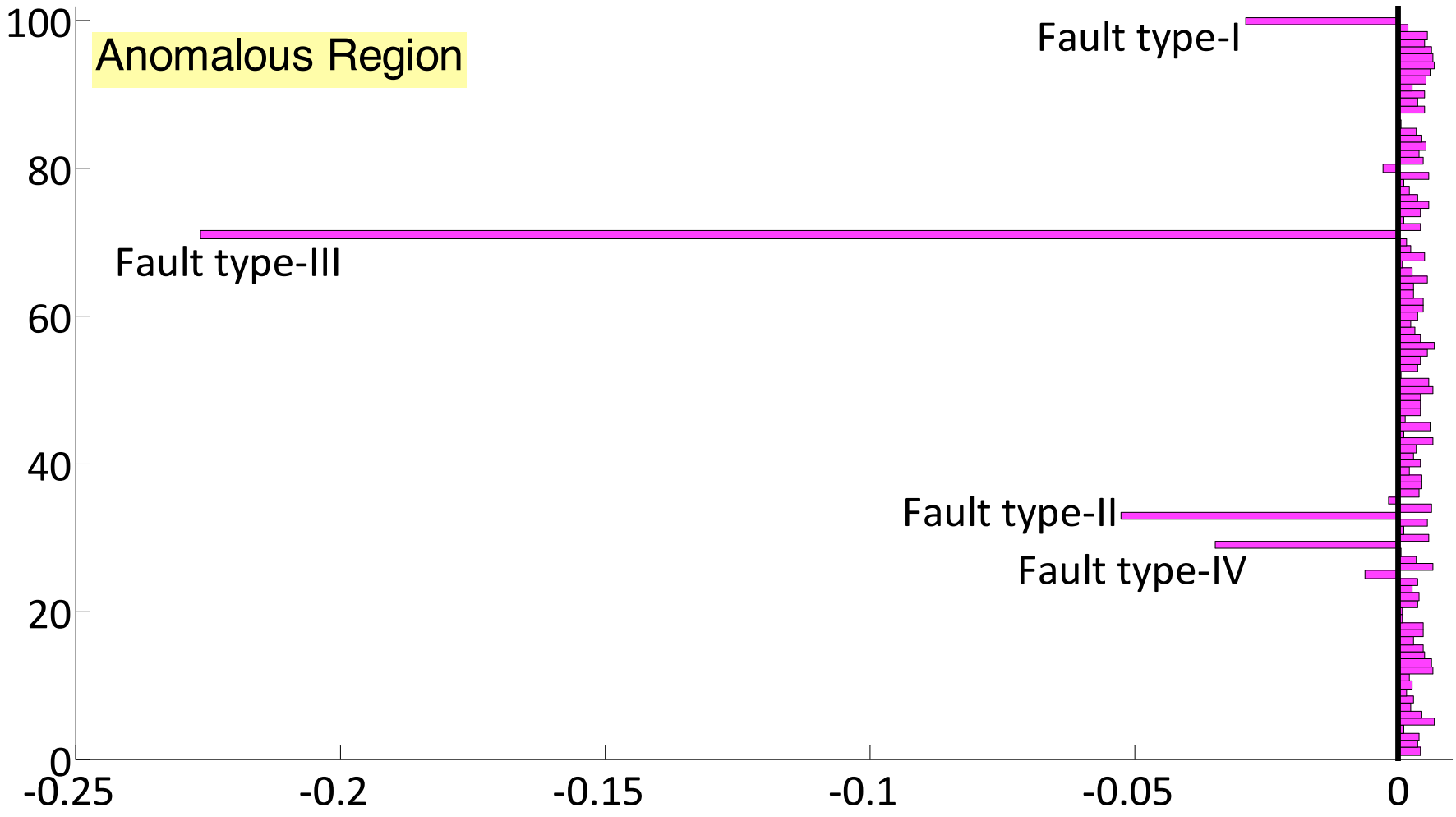
Result: Orca



Result: Sequence Miner



Result: Proposed Method



Highlights

- The proposed algorithm performs anomaly detection on…
  - both discrete symbols and continuous data streams where discrete directly influences the system dynamics which is reflected on the continuous data streams.
- Shows a 100% detection rate on …
  - operationally significant anomalies.
- Can be use for fleet wide analysis on large datasets

Future Work

- Optimizing kernel weights.
- Extending the current research with the consideration of incorporating information from aviation safety reports.
  - Text kernel.
- Decomposing anomaly scores in terms of discrete or continuous components, and then localizing the particular variable(s).