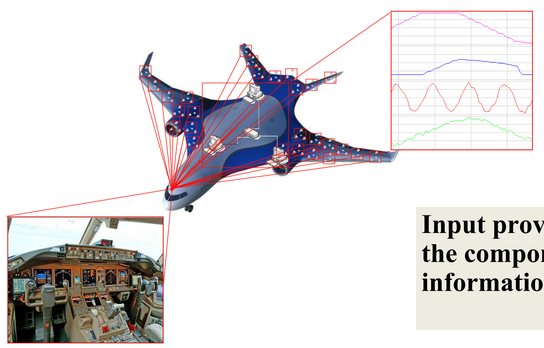
Detecting Anomalies in Multivariate Data Sets with Switching Sequences and Continuous Streams §

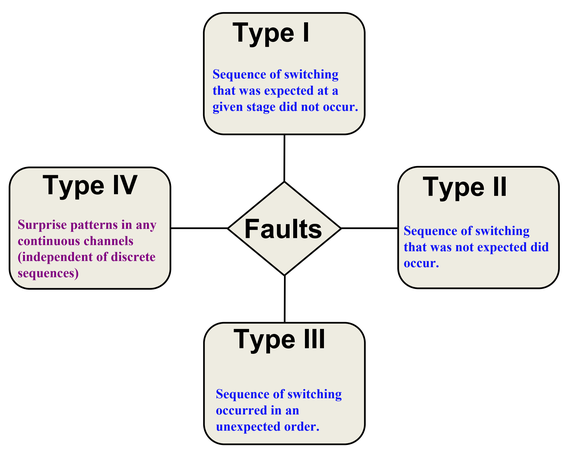
Santanu Das, Bryan Matthews, Kanishka Bhaduri, Nikunj Oza and Ashok Srivastava

Motivation

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

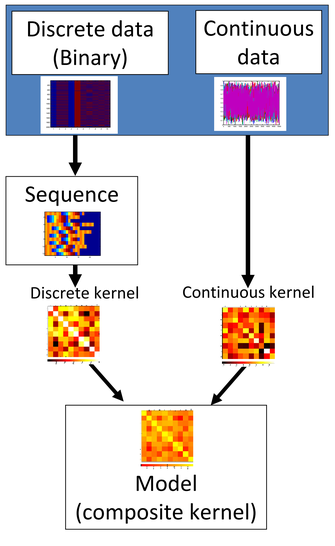


Input provided in the cockpit of a plane and the component/sub-component level information provided by the sensing devices.



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

Approach



Model

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

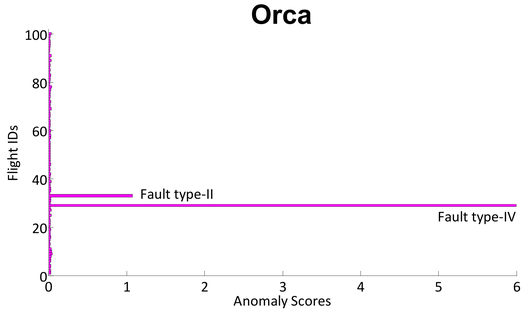
Subject to:

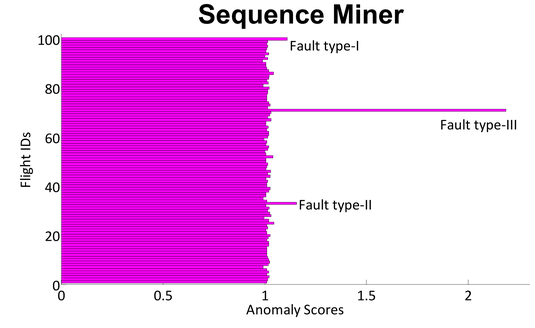
$$0 \le \alpha_i \le \frac{1}{l\nu},\quad \nu \in [0,1],\quad \sum_i \alpha_i = 1$$

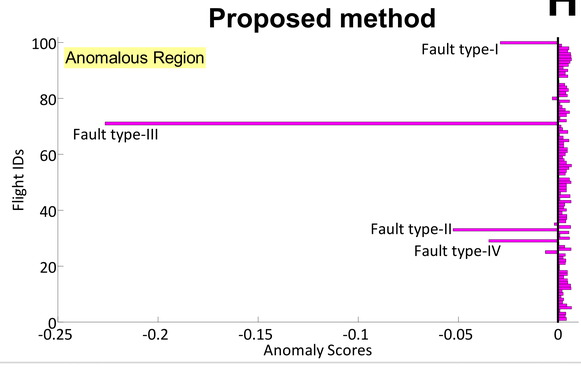
One-class SVMs algorithm (Schölkopf et al.) perform anomaly detection in a much higher dimensional space. The algorithm,

- **solves a convex and quadratic optimization problem.**
- **results in 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>.**

Results







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**

§ This work was supported through funding from the NASA Aeronautics Research Mission Directorate, Aviation Safety Program and Integrated Vehicle Health Management project..