



Detecting Anomalies over Message Streams in Railway Communication Systems

Lucas Foulon, Serge Fenet, Christophe Rigotti, Denis Jouvin

► **To cite this version:**

Lucas Foulon, Serge Fenet, Christophe Rigotti, Denis Jouvin. Detecting Anomalies over Message Streams in Railway Communication Systems. AALTD@ECML/PKDD 2019 - 4th Workshop on Advanced Analytics and Learning on Temporal Data. Poster, Sep 2019, Wurzburg, Germany. pp.1. hal-02357927

HAL Id: hal-02357927

<https://hal.archives-ouvertes.fr/hal-02357927>

Submitted on 4 Dec 2019

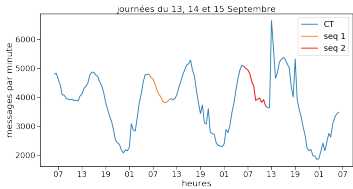
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DETECTING ANOMALIES OVER MESSAGE STREAMS IN RAILWAY COMMUNICATION SYSTEMS

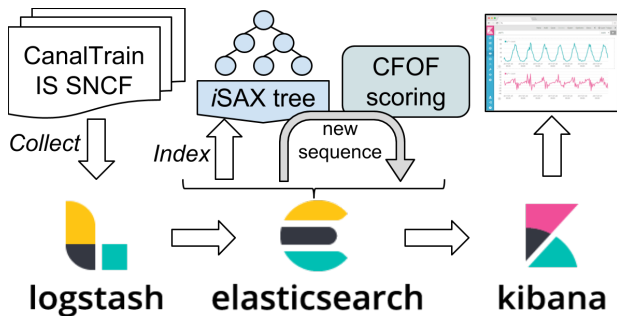
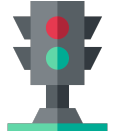
Lucas Foulon^{1,3}, Serge Fenet¹, Christophe Rigotti², Denis Jouvin³

¹Université Claude Bernard Lyon 1, CNRS, LIRIS, UMR5205
²Université Lyon, INSA Lyon, CNRS, INRIA, LIRIS, UMR5205
³Production Ferroviaire, SNCF Mobilité, DSI Voyageurs



GOALS

- Monitor on real-time the proper functioning of the information system
- Support high volume of streaming data
- Warn when an anomaly occurs



OUR DATA

- Traces containing information about messages flowing in the information system: number of messages, latency between different checkpoints, ...
- Built by analyzing the content of the data stream: Sent/Received timestamp, type of device/service, ...
- Interfaced with the central platform of the SNCF IS (CanalTrain) through ELK open source products



METHOD

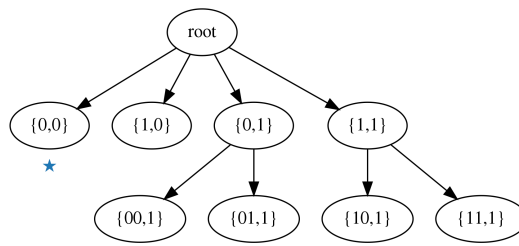
Use of CFOF anomaly measure [Angiulli, ECML PKDD 2017]

- Unsupervised
- Based on the structure of the local neighborhood
- Adapted to high dimension data
- But not adapted to data streams,

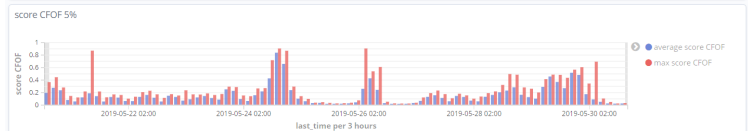
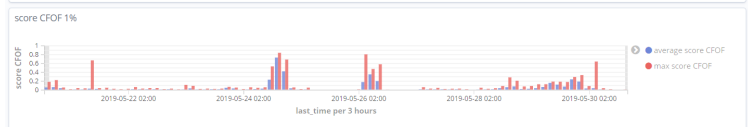
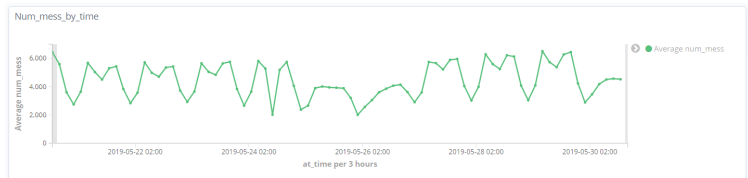
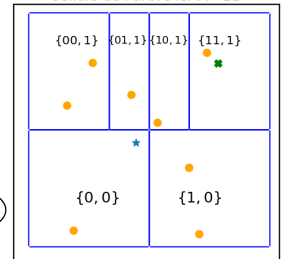
Use of the iSAX indexation tree [Shieh & Keogh, DAMI 2009]

- Based on a modification of the SAX discretization
- Suited for times series indexation and similarity search
- Efficient access using distance boundings
- Support Dynamic Time Warping, weighting, and very high volumes (billion time series)

Proposition : *exploit the properties of the iSAX tree to accelerate the computing of the CFOF score in order to apply it to voluminous data streams*



Cellule de l'arbre iSAX - 2D



RESULTS

- Reduced complexity allowing the efficient use of the CFOF score on high volume data streams
- High quality of the estimated score
- Real time detection of IS anomalies
- One parameter controlling the detection
- Incremental update of the tree



IN PROGRESS

- From tree to forest to reduce dimensions and accelerate the computing
- Multi-scale and multi-indicators anomaly detection
- Testing the robustness to regime changes



4th AALTD@ECML/PKDD 2019 – 20/09/2019

