Streamlining LOFAR System Health Management

A.J. Boonstra[1] and W. Jansen[1]

[1] Netherlands Institute for Radio Astronomy, ASTRON
Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, NL, web: www.astron.nl, email: boonstra@astron.nl

ASTRON

LOFAR radio telescope

International LOFAR Telescope

LOFAR Core Area

- a phased array - aperture synthesis radio telescope
- with 38 Dutch stations, 14 international stations
- with ~ 100.000 antennas
- with over 400 TFLOPS central computing power
- producing up to 100 TByte data daily

gives great science … and maintenance challenges

System health challenge

Given ~ 10,000 LOFAR spectrogram plots per observation, how to detect and identify system erors efficiently?
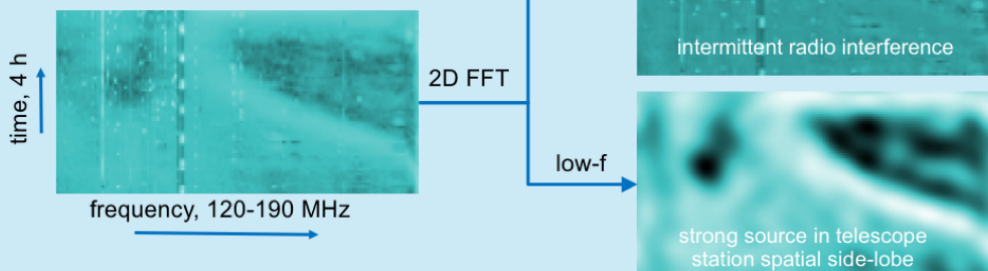
a) Creating error dictionary, relating visibility spectrogram features to system errors:
  → expensive, requires in depth system knowledge to create
b) Use machine learning (ML) to cluster and classify similar image features
  → relatively easy and cheap, but does not automatically identify error type

Goal: automated error detection, error identification, and reporting

Conventional approach for feature separation and detection

Subspace analysis: different features reside in different subspaces

- 2D FFT
- Wavelets
- SVD

ICD Project AI&ML
Picture credit: Maurine Santigny ASTRON / U. Orleans

time, 4 h

frequency, 120-190 MHz

2D FFT

high-f

intermittent radio interference

low-f

strong source in telescope station spatial side-lobe

Machine Learning for feature clustering: traning

Neural net trained on 160 LOFAR MS data sets:
- compresses each 100 TByte data set to 1 GByte
  - sub samples visibilities
  - re-scales data points to fit in 1 byte
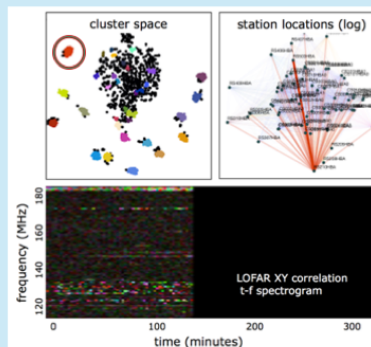- applies autoencoder to train net

input   code   output

X   Z   X'

encoder   decoder
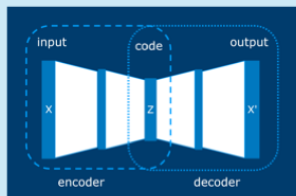
Use t-SNE / DBSCAN for dimensionality reduction and clustering of new observed data

Credit: NLeSC / ASTRON ADDER project

Machine Learning: result

from ~10,000 spectrogram images (baselines) to a hand full of clusters containing typical features

cluster space

station locations (log)

LOFAR XY correlation t-f spectrogram

frequency (MHz)

time (minutes)

← location of telescope stations (dots) and baselines (colored lines)

example: one station broke down after 140 minutes (brown cluster upper left in circle)

Status

- Observatory automatically produces compressed data sets and clusters of baseline spectrograms for efficient data inspection
- Already a useful tool but clustering is not yet perfect, and classification is not yet included

Outlook

- Improve clustering
- Include separation of features
- Create feature dictionary and add classification

ASTRON

Innovatiecluster high tech systems Drachten

NWO