

Anomaly Detection on Streaming Data using Hierarchical Temporal Memory (and LSTM)

Jaime Coello de Portugal

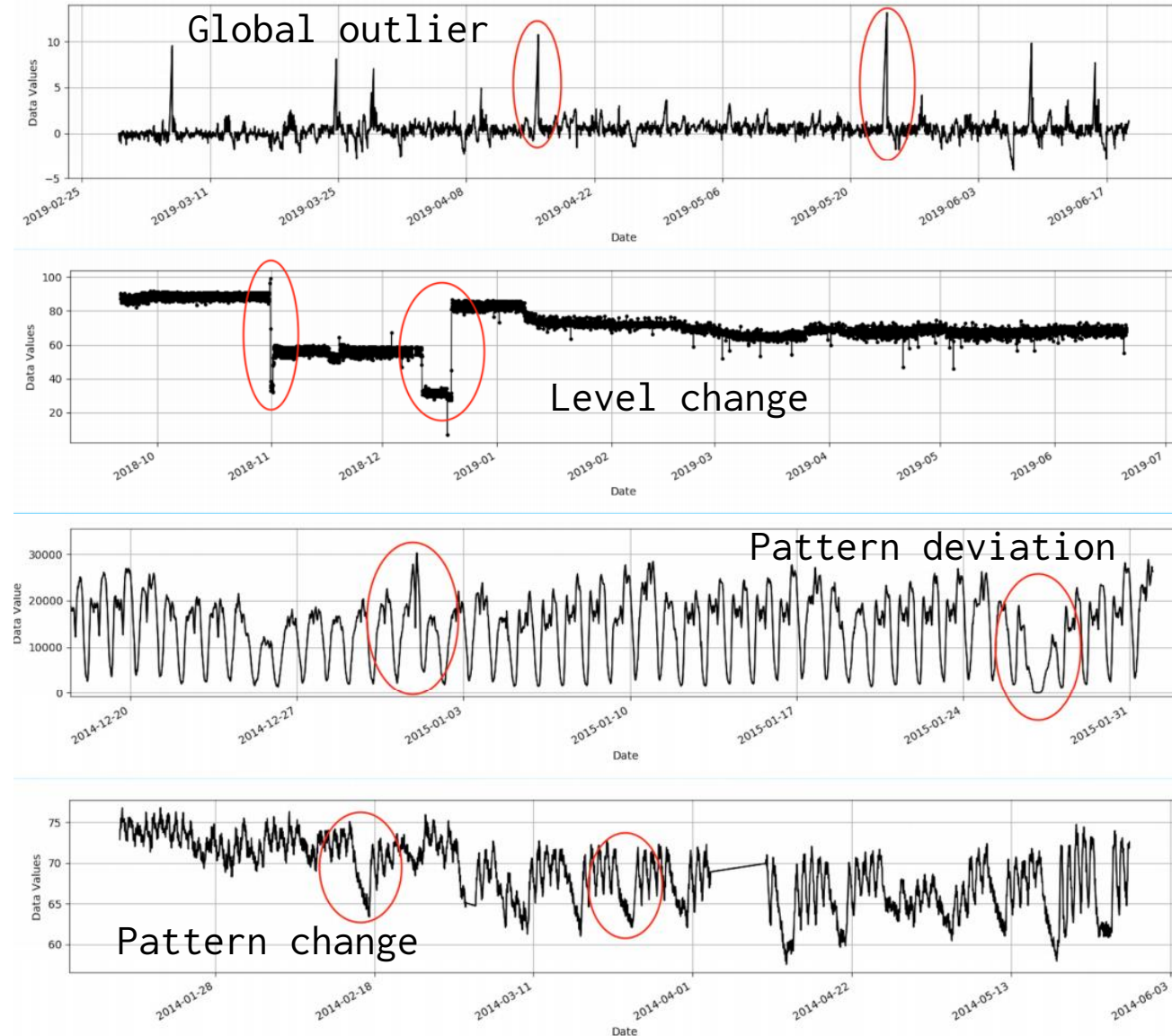


Many thanks to Jochem Snuverink

“An observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism.”

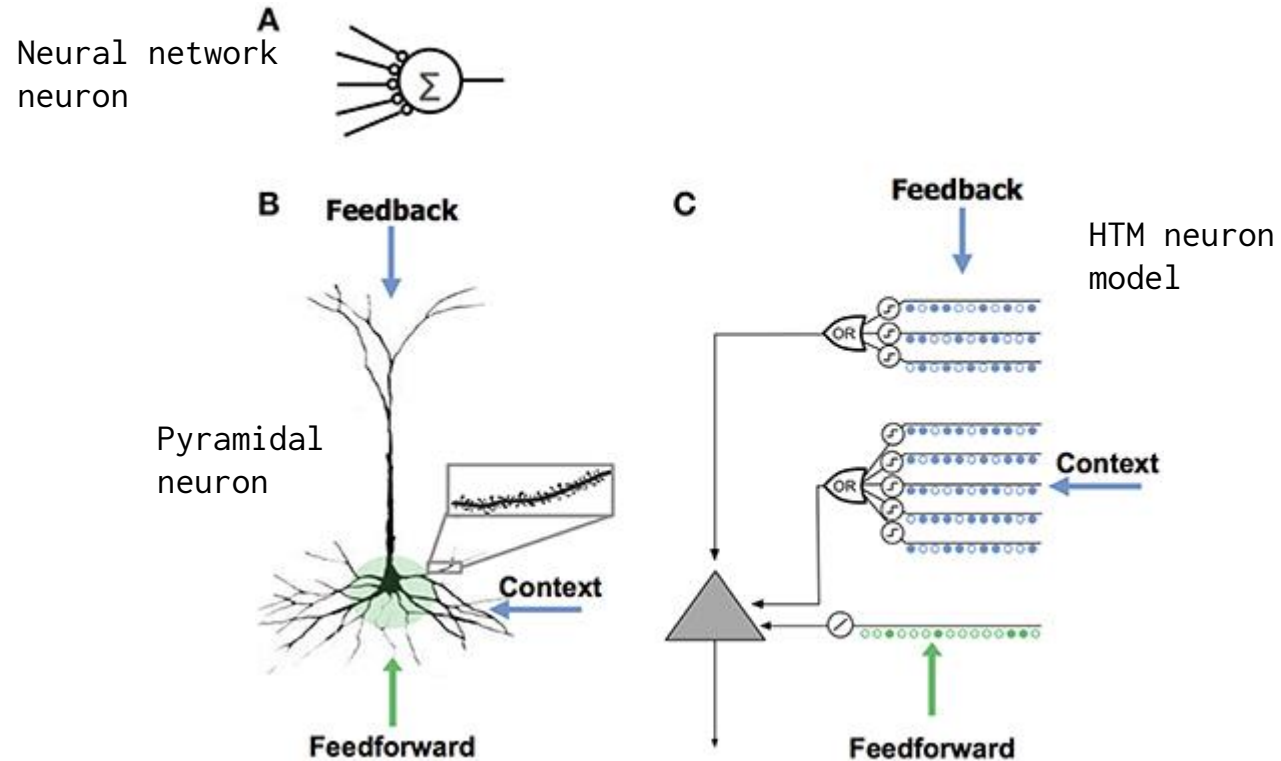
– Hawkins, 1980

- Try to find observation or sequences that deviate from the “normal behaviour”.
- Experts would recognize these anomalous patterns easily, but cannot be monitoring the huge amount of data some systems produce.
- E.g: Credit card fraud detection, intrusion detection in cybersecurity, or **fault diagnosis in industry**.
- Specific e.g: At HIPA the MHB7R:ILOG:2 temperature detector broke down without anyone noticing.

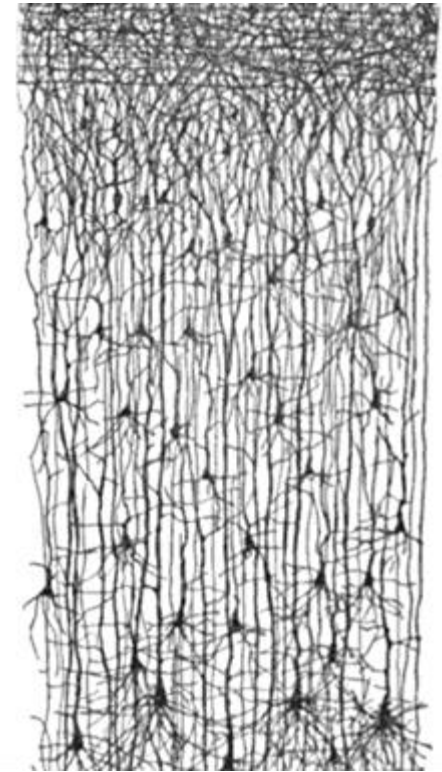


Hierarchical Temporal Memory: Introduction

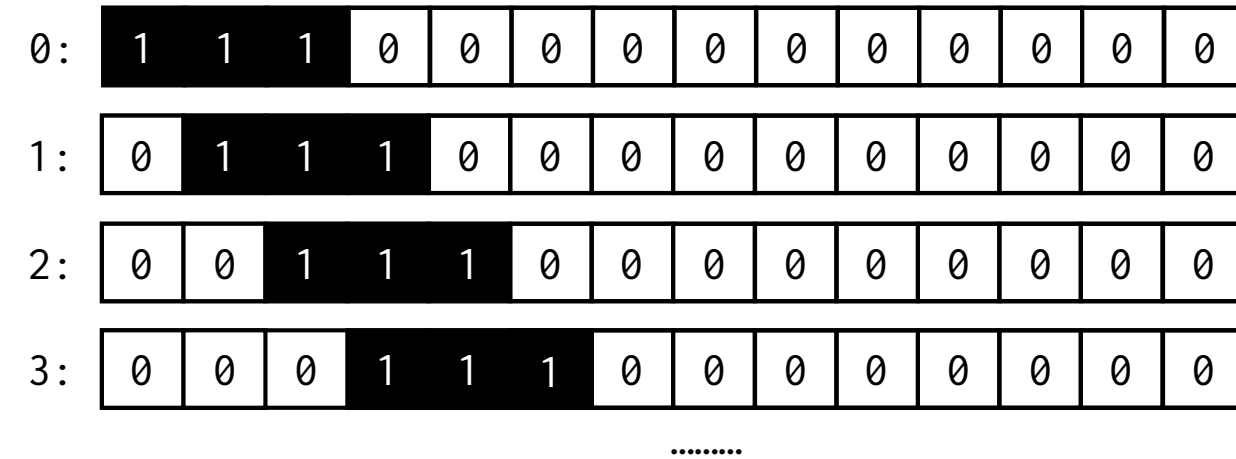
- Model of the organization of “pyramidal neurons” in the neocortex of mammals.
- Tries to explain how neuronal structures remember sequences.



Pyramidal neurons connect forming “columns” that share input and output.

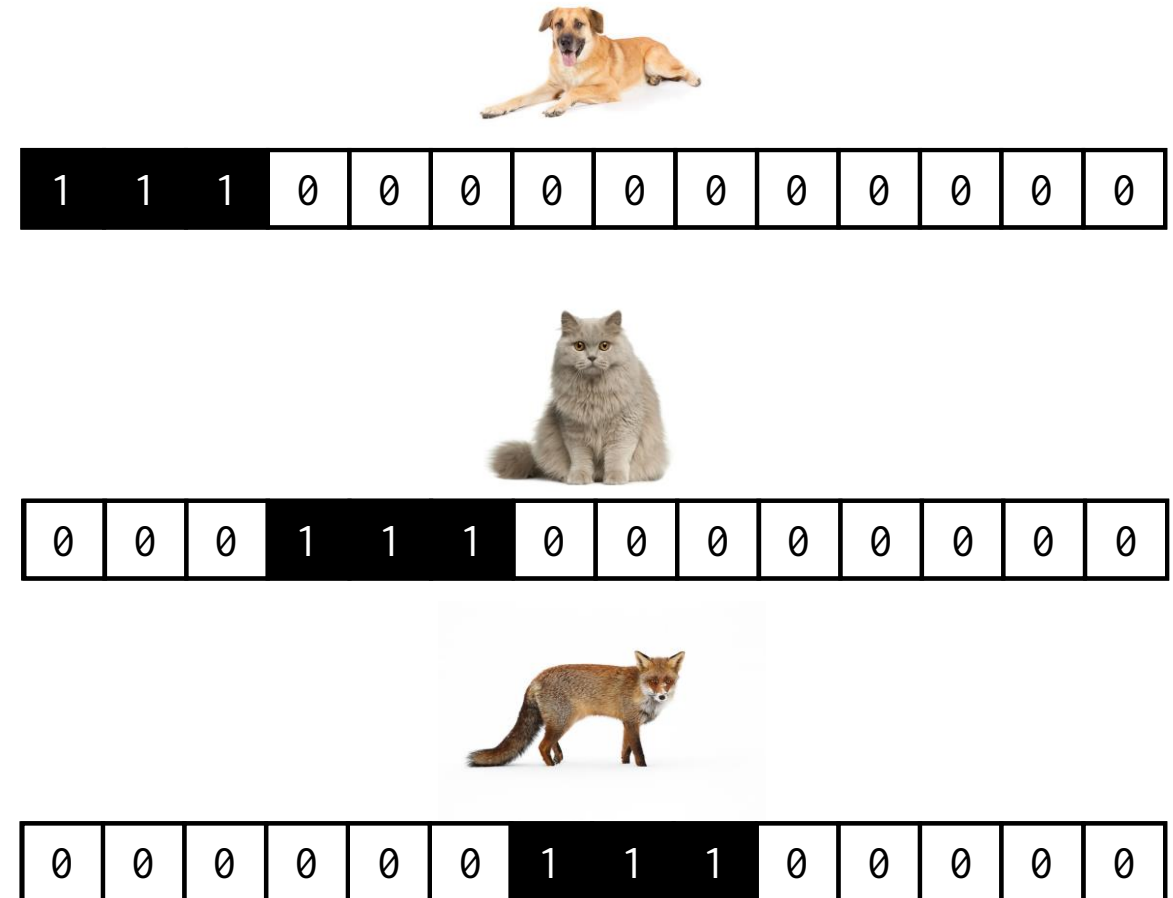


- Scalar variables:



- Similar values overlap ->
 - Encodes “proximity” between values.
- Produces some resistance to noise.
Discrepancies due to noise will get the same or close representation.

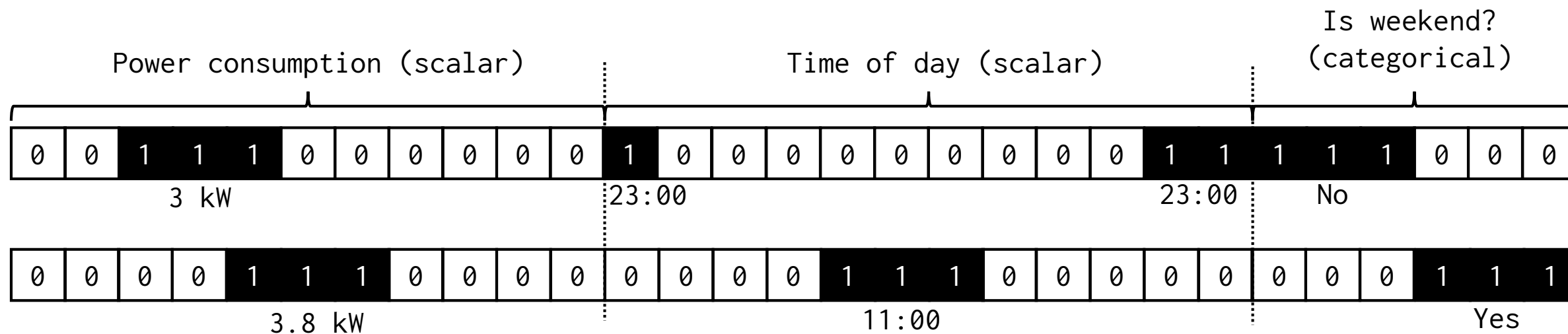
- Categorical variables:



No overlap between different categories

Hierarchical Temporal Memory: Sparse Distributed Representations (SDR)

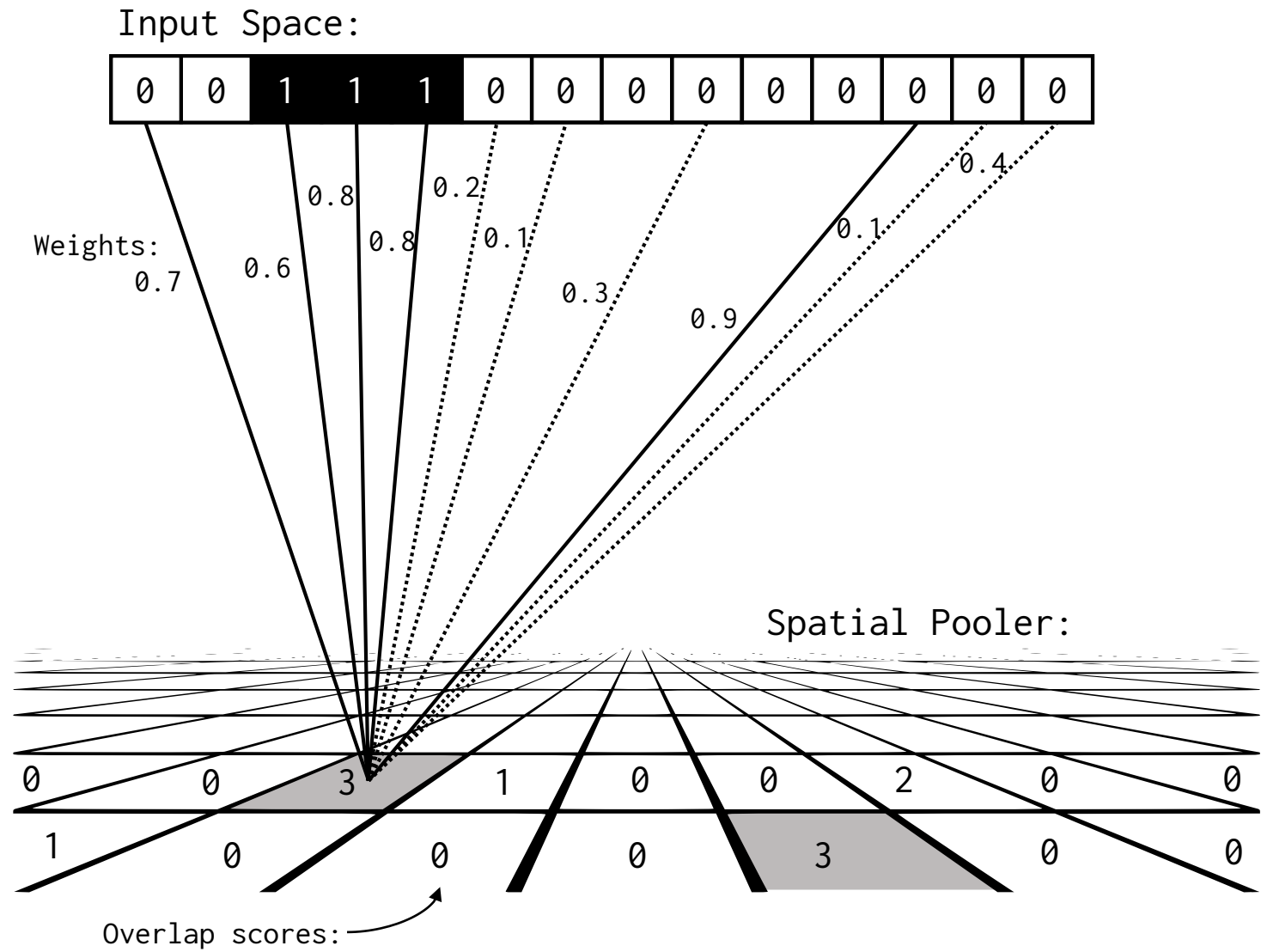
- Several variables can share an SDR.
- Can also encode periodic proximity.
- E.g. power consumption in a gym:



- The input to the HTM network is one of these SDR per time unit.

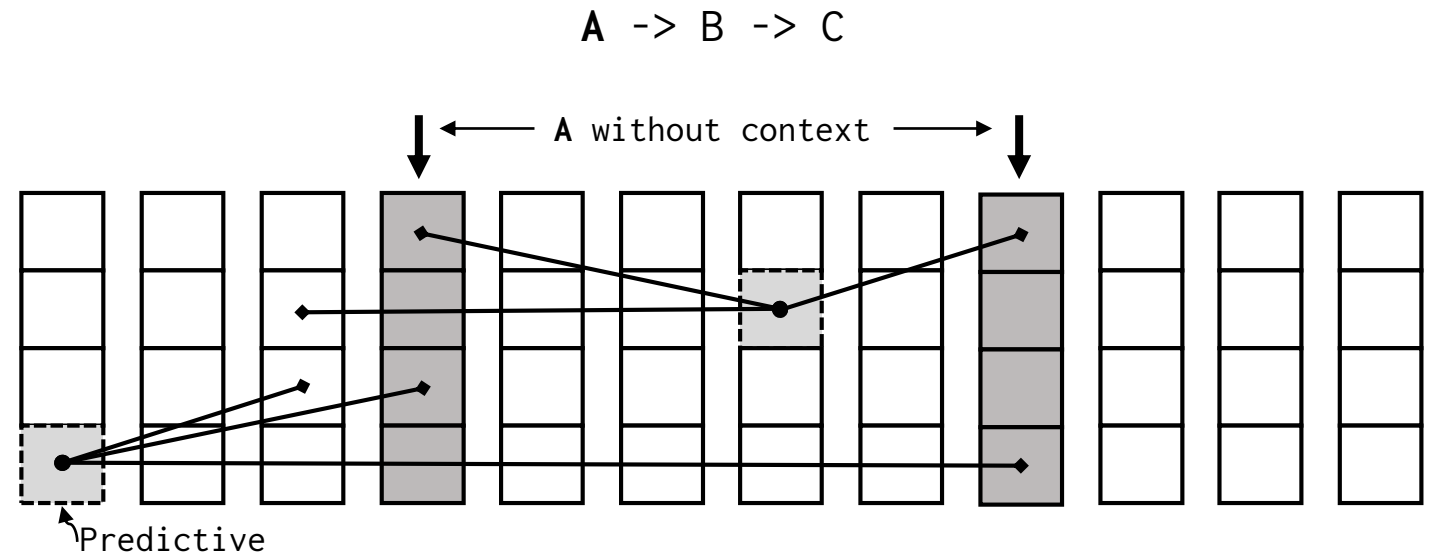
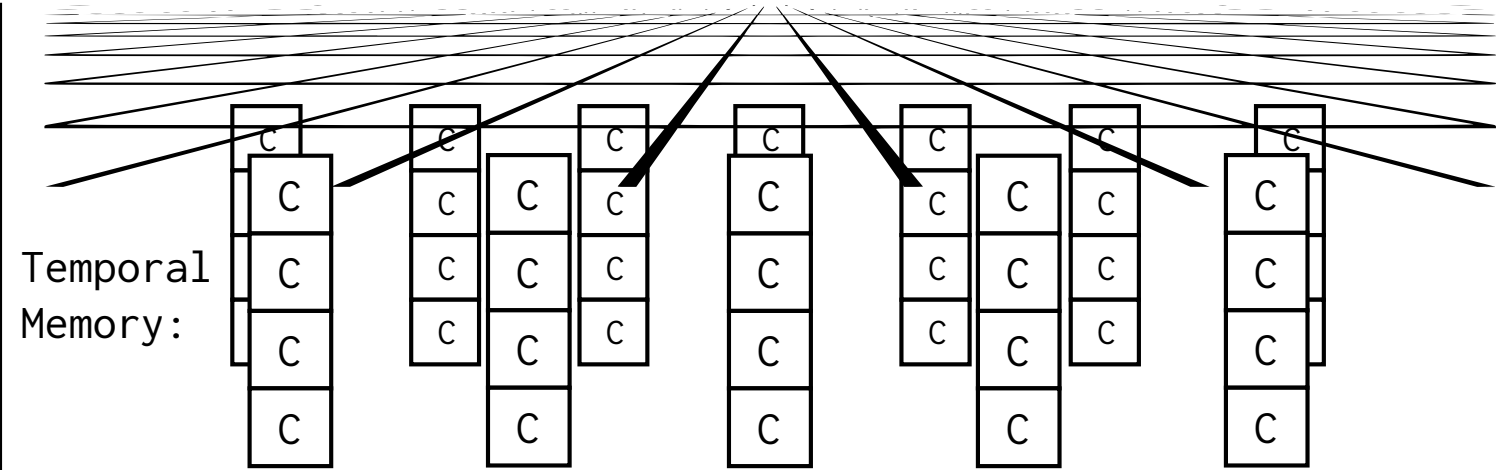
Hierarchical Temporal Memory: Spatial Pooler (SP)

- Each cell on the SP gets random weighted connections to the input space.
- If the weight is over some threshold the cell gets connected to a bit of the input space.
- The N cells that have the most connection overlap to the input space become “active”.
- Target:
Maintain the semantic information from the input space with a fixed (~2%) sparsity.
- **Active** cells **learn** by updating their weights.



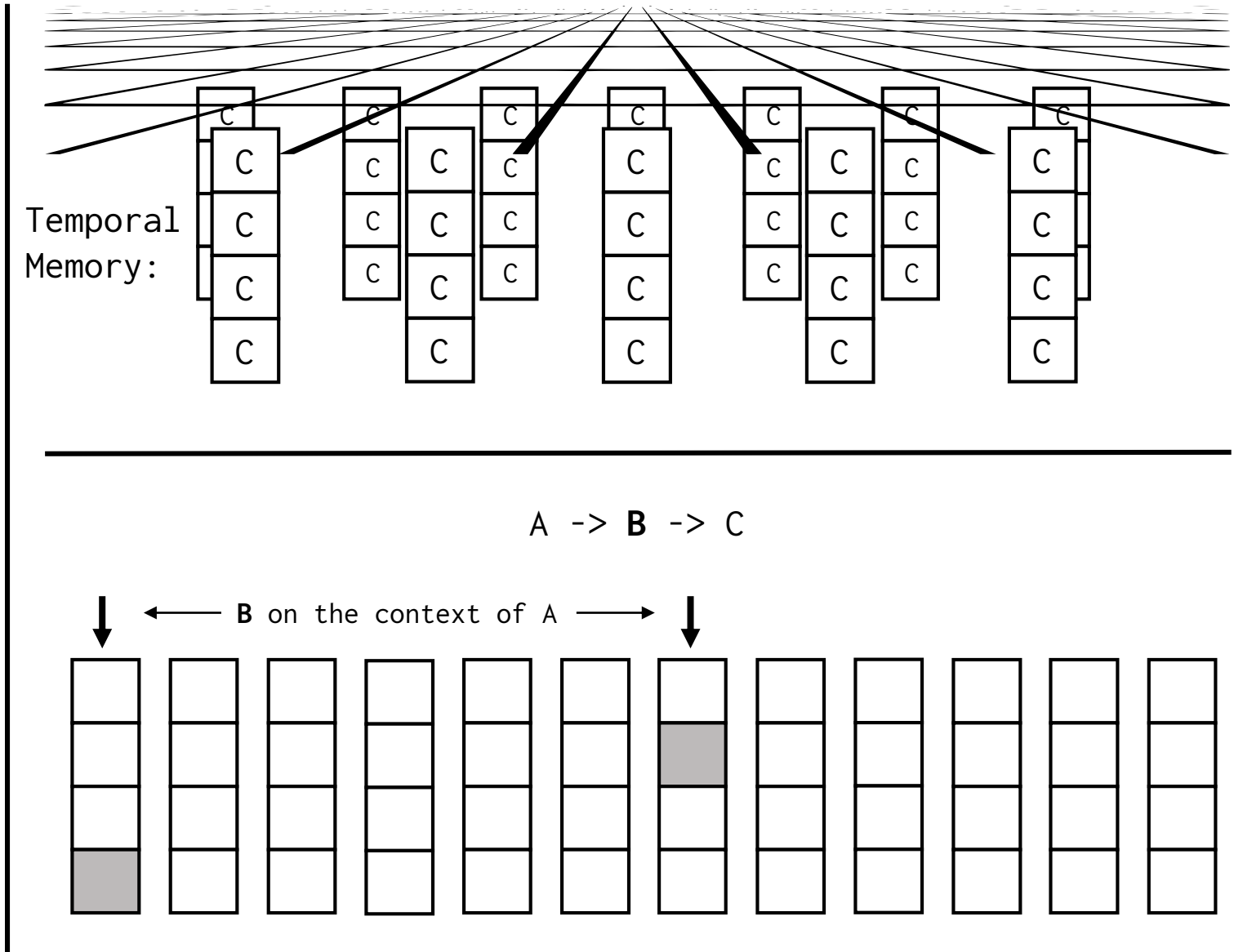
Hierarchical Temporal Memory: Temporal Memory (TM)

- Each cell of the SP is actually made of several cells that respond to the same input, forming “mini-columns”.
- These cells have connections with cells in other columns.
- If enough of these connections are active, the cell goes into “predictive” state.



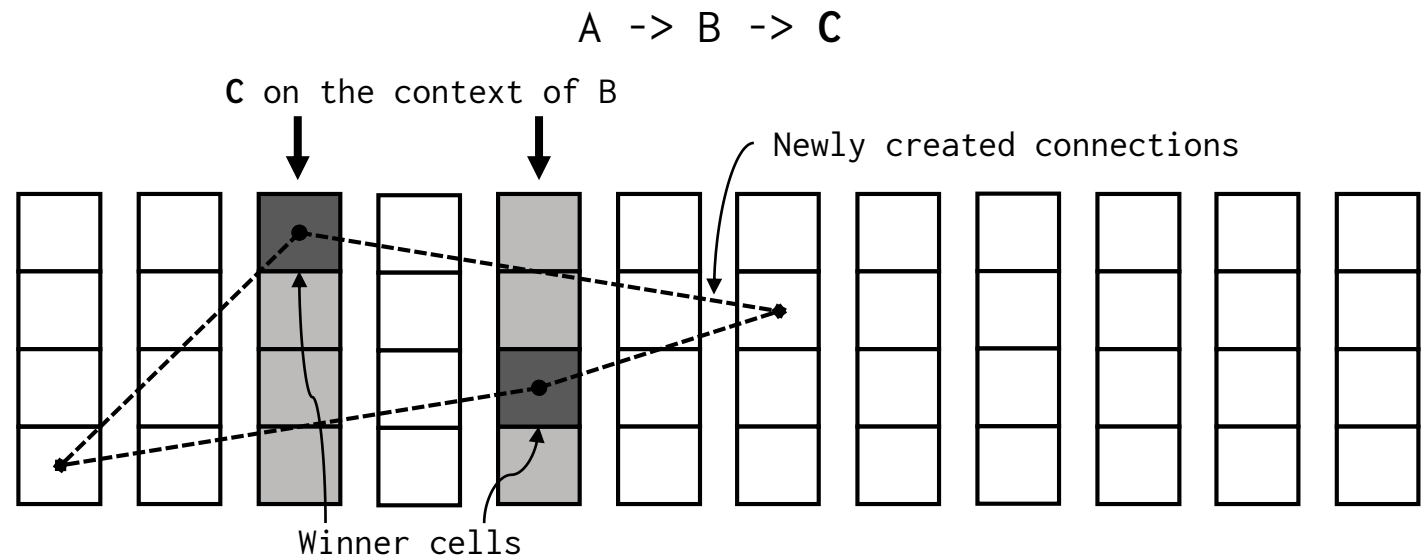
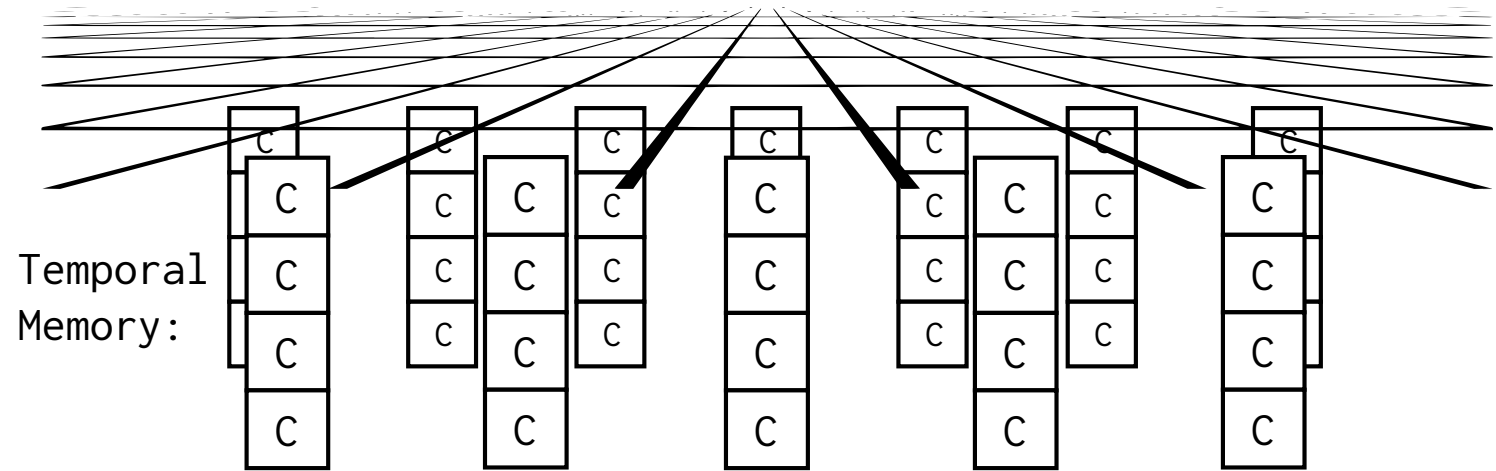
Hierarchical Temporal Memory: Temporal Memory (TM)

- On the next input, if a column containing a predictive cell becomes active only the predictive cell becomes active.

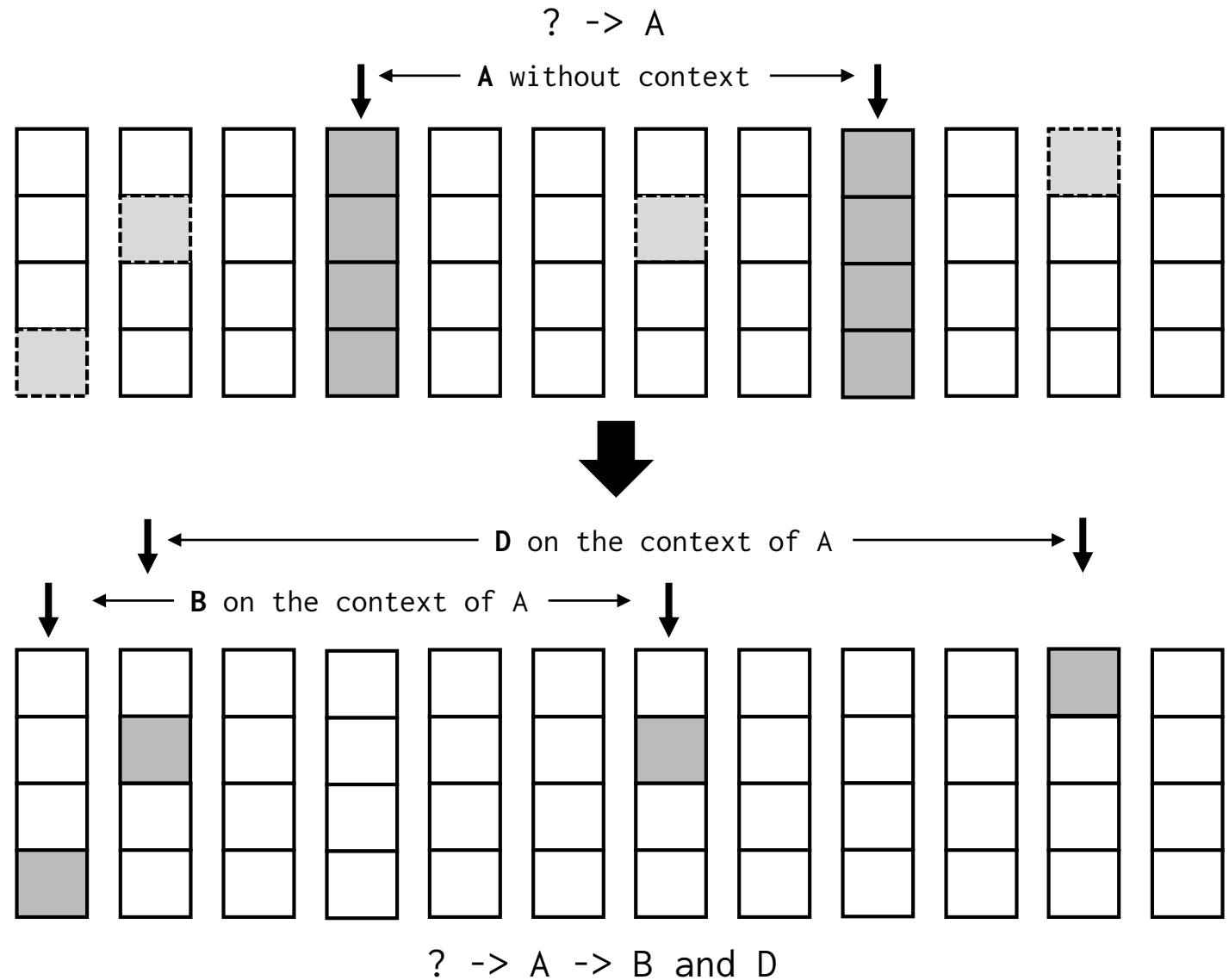


Hierarchical Temporal Memory: Temporal Memory (TM)

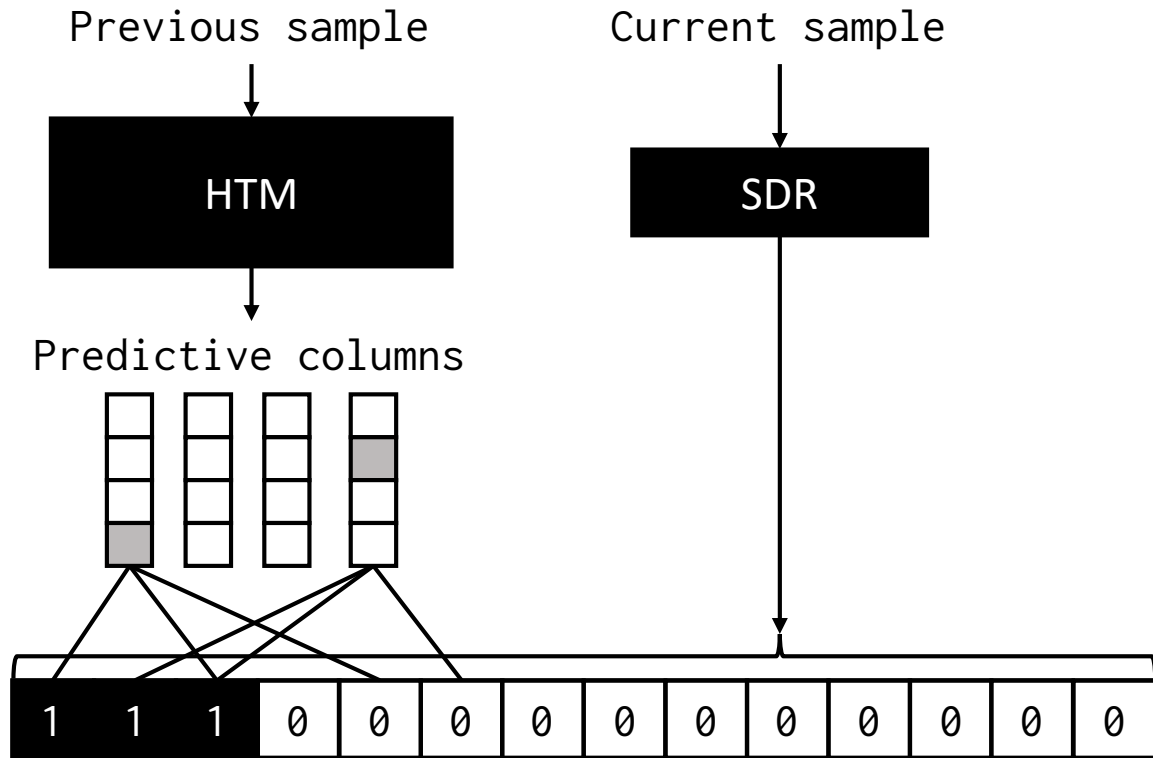
- When an unexpected input happens, the whole column activates.
- One of the cells of the column is then chosen as the “winner” and grows connections to the previous state.
- Old unused connections are slowly forgotten.



- When several cells in a column are activated, many columns might be predictive.
- This allows the model to “doubt” in case of unknown or ambiguous context.
- In the next cycle, the next input will reveal the correct prediction and the model will learn.

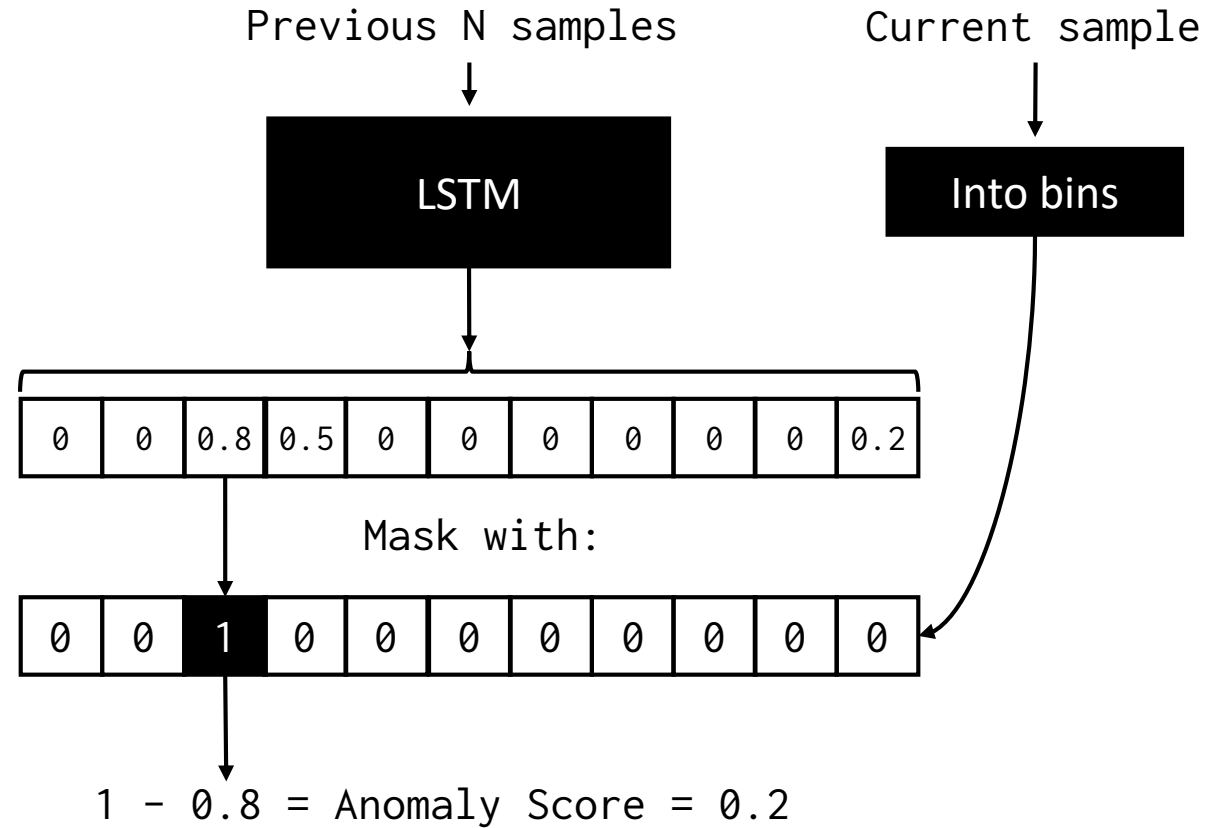


HTM Anomaly Score:



$$\text{Anomaly Score} = \frac{\text{Correctly predicted inputs}}{\text{Number of connections}}$$

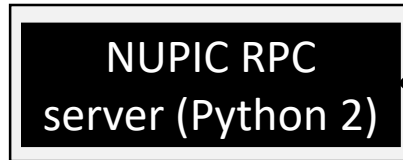
LSTM Anomaly Score:



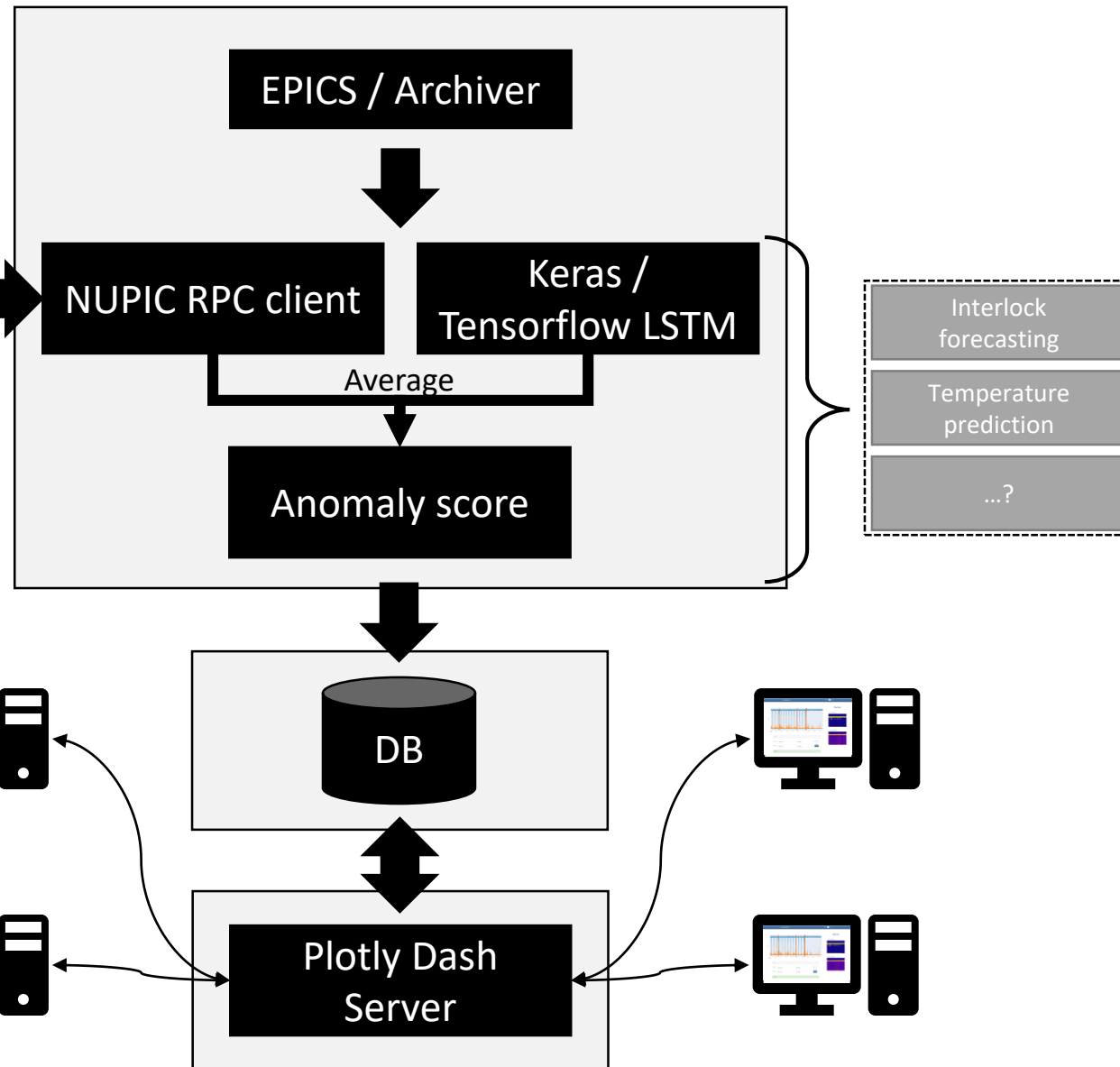
$$\text{Training penalty} = \text{Anomaly Score} + \gamma \sum |prediction|$$

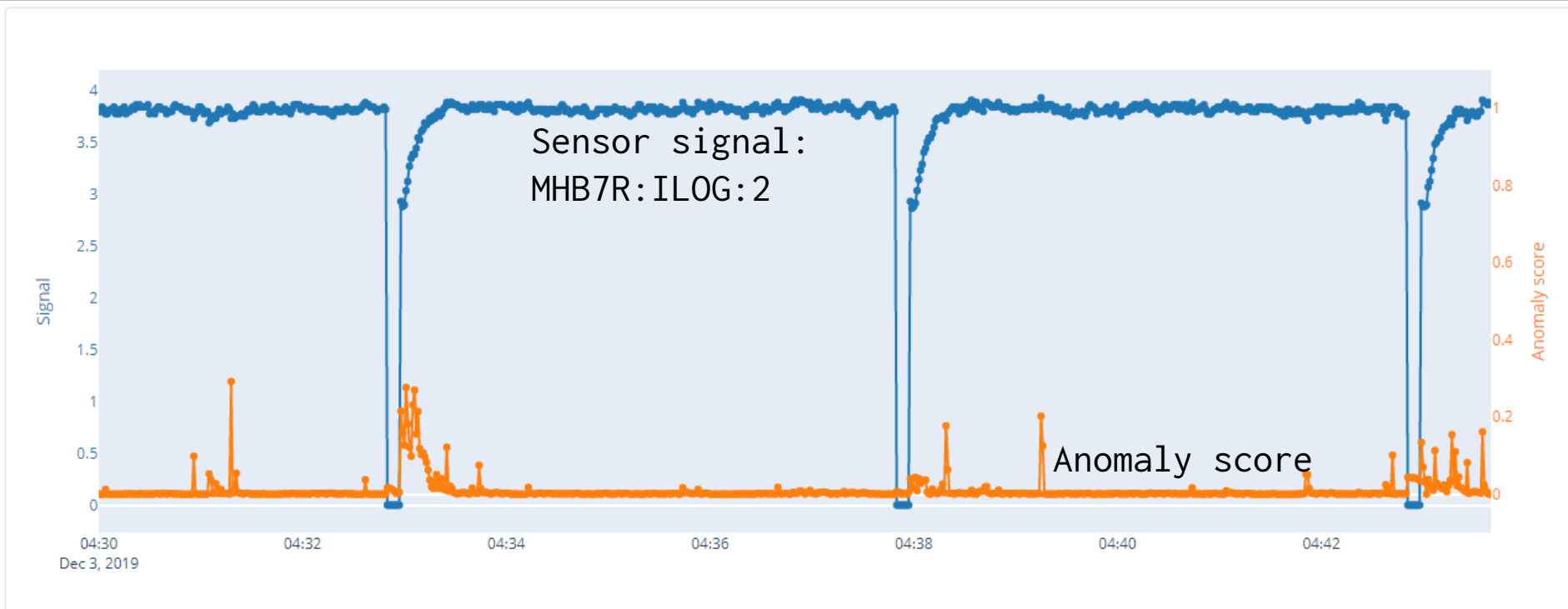
Anomaly detection implementation

NUPIC is a Python 2 (wrapping C++) implementation of the HTM algorithm by Numenta (numenta.com).



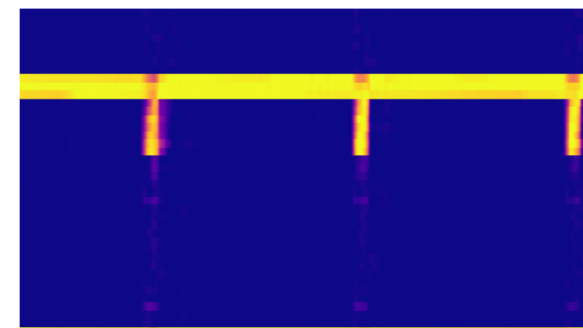
They provide a set of hyperparameters that works good for anomaly detection out of the box



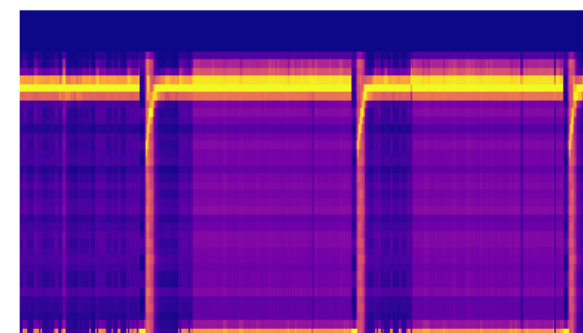


Normal

LSTM probabilities



NUPIC probabilities

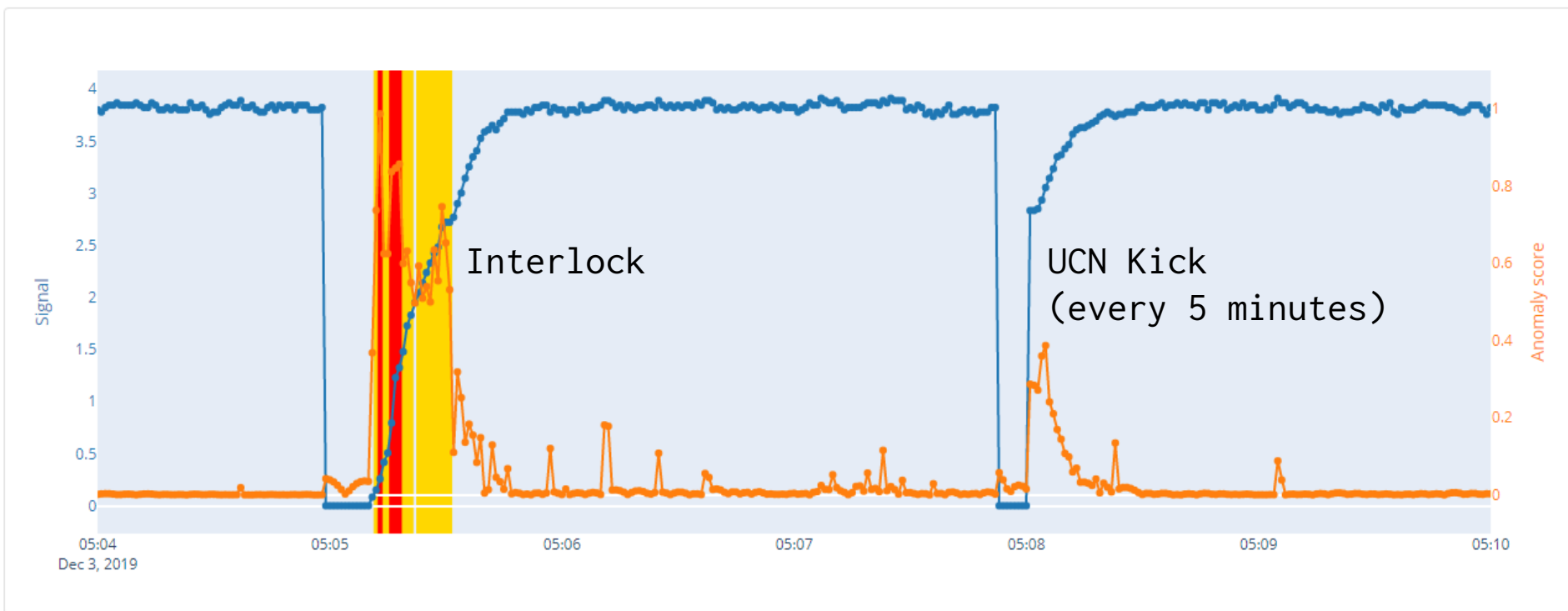


History

Start Replay from here?

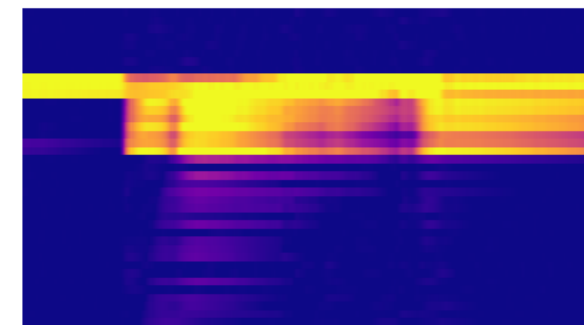
End

No alarms

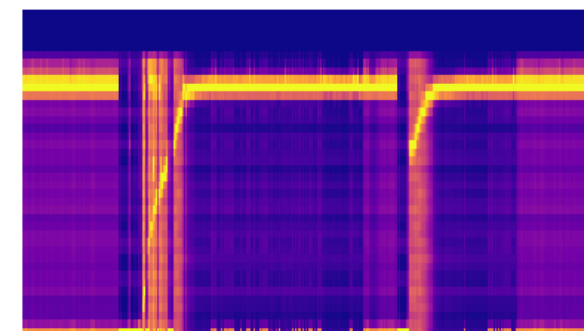


Normal

LSTM probabilities



NUPIC probabilities



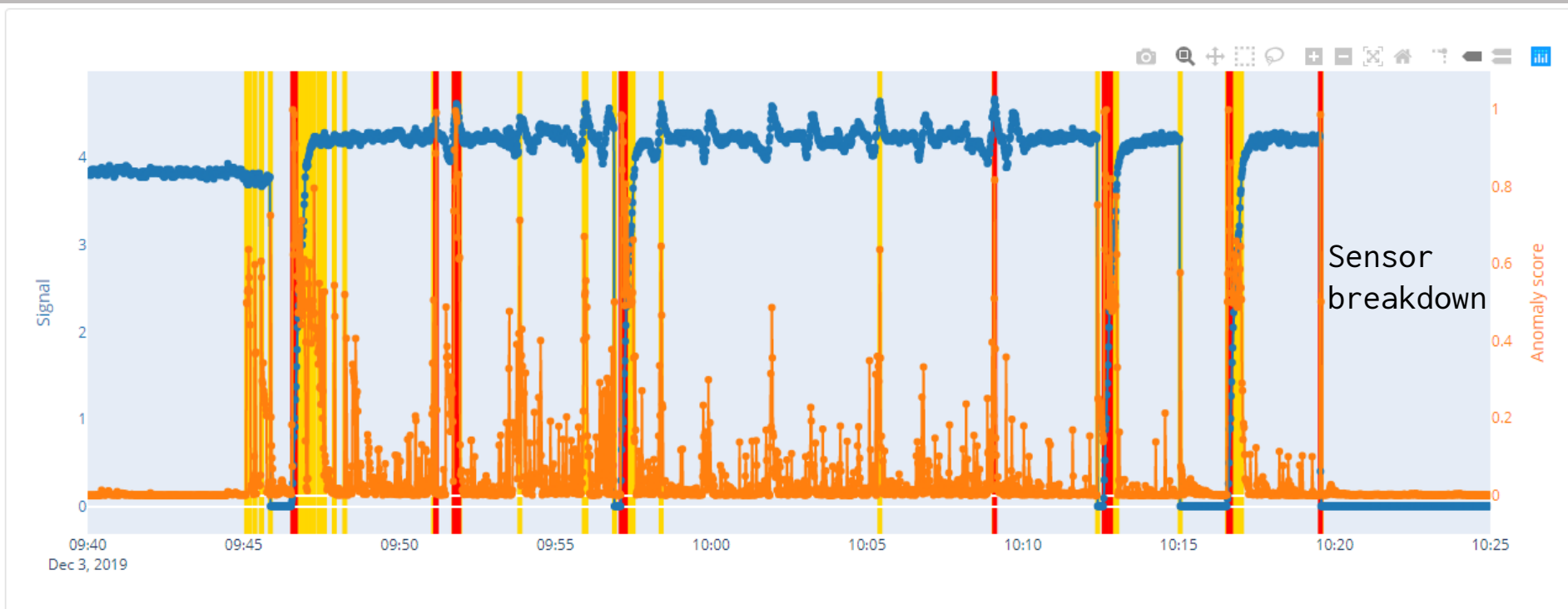
History

Start Replay from here?

End

No alarms

Anomaly detection in HIPA



History

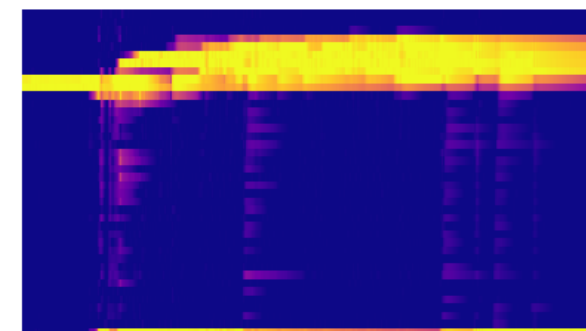
Start Replay from here?

End

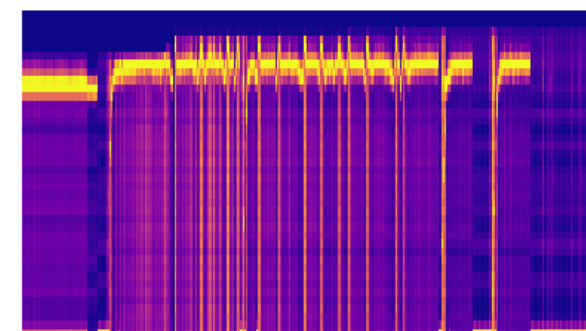
No alarms

Normal

LSTM probabilities



NUPIC probabilities



- Anomalies detected live in the vacuum level in the cyclotron:



Conclusions:

- The combination of the HTM network and the LSTM seems to provide a good real time anomaly detection system.
- The system works well on archived and live data, adapts to both stable and periodic behaviour.
- It would have alerted the experts of the failure of the MHB7R:ILOG:2 sensor tenths of minutes in advance.
- Tested on real time on PROSCAN, ready for the restart of HIPA.
- A third anomaly detection based in Isolation Forest: “Robust Random Cut Forest” is being tested in combination with the other two.

Robust Random Cut Forest: <https://proceedings.mlr.press/v48/guha16.pdf>

A nice video series to learn about the HTM:

<https://www.youtube.com/playlist?list=PL3yXMgtrZmDqhsFQzwUC9V8MeeV0Q7eZ9>