Automatic recognition of anomalous patterns in discharges by recurrent neural networks

Gonzalo Farias\textsuperscript{1}
Ernesto Fabregas\textsuperscript{2}
Sebastián Dormido-Canto\textsuperscript{2}
Jesús Vega\textsuperscript{3}
Sebastián Vergara\textsuperscript{1}

May 13-17 2019 / Daejeon, KOREA
Outline

- Introduction
- Background
  - Anomaly Detection
- Proposed Solution
  - Recurrent Neural Networks (LSTM)
- Results
- Summary
The experiments generate huge quantities of data. It is estimated that only 10% of this data is analyzed.

A shot of few seconds can generate huge quantity of data:

- **TJ-II** device has +1000 channels of measurements.
- A shot in **JET** can take around 10 seconds (**10 GB/shot**. around 100 TB/year).
- **ITER** could generate 1 **TB/shot**. around 1 PB/year.
Outline

- Introduction
- Background
  - Anomalies
- Proposed Solution
  - Recurrent Neural Networks (LSTM)
  - Anomaly Detection
- Results
- Summary
Background

- The idea is to use Artificial Intelligence to deal with fusion data.
- Create systems that allow specialists to analyze and interpret data more quickly and efficiently than manually.
Background – Anomalies

- **Anomaly**: Something that deviates from what is standard, normal, or expected.

- One type of anomaly is known as 'outlier', which is a value located outside of the normal class.

- Other type of anomaly is an anomalous behavior, which is a periodic collapsing phenomenon in time series.
Outline

- Introduction
- Background
  - Anomalies
- Proposed Solution
  - Recurrent Neural Networks (LSTM)
  - Anomaly Detection
- Results
- Summary
Proposed Solution – LSTM

- Recurrent Neural Network – Long Short Term Memory (LSTM)

Forget gate layer

\[ f_t = \sigma (W_f [h_{t-1}, x_t] + b_f) \]

Input gate layer

\[ i_t = \sigma (W_i [h_{t-1}, x_t] + b_i) \]
\[ \tilde{C}_t = \tanh (W_c [h_{t-1}, x_t] + b_c) \]
\[ C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \]

Output gate layer

\[ o_t = \sigma (W_o [h_{t-1}, x_t] + b_o) \]
\[ h_t = o_t * \tanh (C_t) \]

* https://colah.github.io/posts/2015-08-Understanding-LSTMs/*
LSTM – Forecasting (training stage)

\[ f_t = \sigma (W_f [h_{t-1}, x_t] + b_f) \]

Forget gate layer

\[ i_t = \sigma (W_i [h_{t-1}, x_t] + b_i) \]

Input gate layer

\[ \tilde{C}_t = \tanh (W_c [h_{t-1}, x_t] + b_c) \]

\[ C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \]

\[ o_t = \sigma (W_o [h_{t-1}, x_t] + b_o) \]

Output gate layer

\[ h_t = o_t * \tanh (C_t) \]

*It adjusts the bias and weights to learn the shape of the waveform*
LSTM – Forecasting (test stage)

- LSTM for forecasting (trained network)

![Graph showing LSTM forecasting](image-url)
Goals

- **General Goal**
  - Anomaly detection using Recurrent Neural Network (LSTM - Long Short Term Memory).

- **Specific Goal**
  - The LSTM Neural Network learns the waveform to detect anomalies through forecasting.

* https://colah.github.io/posts/2015-08-Understanding-LSTMs/
Anomaly Detection – Threshold (th=k*std)

- How the Anomaly is detected?
  - We fix a **threshold** proportional to the **Standard Deviation** of the **Error**.

![Example of threshold detection](image-url)
We look for simultaneous anomalies (same time instant) in different signals within the same shot.
We look for simultaneous anomalies (same time instant) in different signals within the same shot.
Anomaly Detection – Simultaneous ($\Delta t$)

We look for simultaneous anomalies in Time Windows ($\Delta t$).
Anomaly Detection – Simultaneous ($\Delta t$)

- We look for simultaneous anomalies in Time Windows ($\Delta t$).
Anomaly Detection – Algorithm

Set parameters
\( (th, An_t, \Delta t, An_{\Delta t}) \)

For each signal \((i)\) predict \(t+1\)

\[ X'_{it} = \text{Predict} (X_{it}) \]

Calculate difference between Predicted and Observed

\[ \text{Diff}_{it} = X_{it} - X'_{it} \]

Anomalies Detection

\[ \text{Outlier}_{it} = \begin{cases} 1 & \text{if } |\text{Diff}_{it}| \geq th \\ 0 & \text{otherwise} \end{cases} \]

\[ \sum_{i} \text{outlier}_{it} \geq An_t \]

\[ \sum_{k=t+1-\Delta t}^{t} \sum_{i} \text{outlier}_{it} \geq An_{\Delta t} \]

Anomaly Detected

Time instant \(t\)

no

yes

no
Outline

- Introduction
- Background
  - Anomalies
- Proposed Solution
  - Recurrent Neural Networks (LSTM)
  - Anomaly Detection
- Results
- Summary
The wider is the band the less anomalies are detected

**Results**

- **Simultaneous Anomalies Detection in a Shot (t)**

<table>
<thead>
<tr>
<th>$A_{nt}$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>190</td>
<td>109</td>
<td>67</td>
<td>40</td>
<td>21</td>
<td>11</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>96</td>
<td>34</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>49</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

The more simultaneity is required, the less anomalies are detected.  

1 simultaneous anomaly in 4 signals for $k=2$ at given time (t)  

*100 shots randomly selected*
## Results

- **Simultaneous Anomalies Detection in Time Windows** ($\Delta t=5$)

The wider is the band the less anomalies are detected.

The more simultaneity is required, the less anomalies are detected.

*100 shots randomly selected*

### Table 1: Simultaneous Anomalies Detection

<table>
<thead>
<tr>
<th>$A_{\Delta t}$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{\Delta t}$</td>
<td>266</td>
<td>204</td>
<td>153</td>
<td>110</td>
<td>62</td>
<td>33</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>212</td>
<td>98</td>
<td>50</td>
<td>35</td>
<td>21</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>146</td>
<td>54</td>
<td>25</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>92</td>
<td>35</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>64</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

4 simultaneous anomalies in 8 signals for $k=1$ with $\Delta t=5$.
Summary

- LSTM networks can learn the shape of a waveform (one model for signal).
- LSTM networks can be used for anomaly detection in signals.
- The specialists have to define the parameters to distinguish the noise from the real anomalies.
- It is possible to design supervised systems that allow the detection of previous detected/studied anomalies.
- In the paper ID. 484 you can find other anomaly detection methods.
Automatic recognition of anomalous patterns in discharges by recurrent neural networks

Gonzalo Farias¹
Ernesto Fabregas²
Sebastián Dormido-Canto²
Jesús Vega³
Sebastián Vergara¹

May 13-17 2019 / Daejeon, KOREA