GPT-like Attention Mechanisms for Power Transformer Condition Monitoring and Prognostics

Mentor: Yufei Tang, Ph.D.
Raul Mendy, Aira Torres, Joseph Accurso
Research Objectives

- Diagnostic Sensors
- Power Failure Indication
- Reactive Maintenance
- Scheduled Maintenance
- On-Site System Diagnostics
REMAINING USEFUL LIFE MODEL
Remaining Useful Life

- ETTm1 & ETTm2
- 2 power transformers
- Change in oil temp for 2 years
- Degradation-based RUL model
- Min and max threshold values
- Estimates # of days until maintenance is required

Machine Learning Model:
- Polynomial Regression
Remaining Useful Life Results

Case 1: Decreasing prediction line
Month 1: M1 dataset

RUL = 18 DAYS

Case 2: Increasing prediction line
Month 24: M2 dataset

RUL = 22 DAYS
MULTI-CLASS CLASSIFICATION MODEL
Fault Classification Dataset

- Text File Format
- ~110,000 Simulated Examples
- 45 Fault Classes
- 726 Measurements per Phase
- Normalized to ~390,000 Examples
- ~850,000,000 Total Measurements
Fault Classification
Using Machine Learning

- Discrete Wavelet Transform Decomposition
  - Daubechies 2 Wavelet
  - 3 - 5 Decomposition Levels
- Random Forest
- Gradient Boost
- Support Vector Machine

![Effect of DWT Decomposition Levels on SVC Accuracy](image1)

![Capactor Switching Transient Data Example](image2)
Fault Classification
Using Machine Learning

- Discrete Wavelet Transform Decomposition
  - Daubechies 2 Wavelet
  - 3 - 5 Decomposition Levels
- Random Forest
- Support Vector Machine
Fault Classification Using the Transformer Architecture

- Foundational architecture of Chat-GPT
- Signal analysis within positional context
- Multi-head attention mechanism
- No need for time-consuming DWT decomposition-based feature extraction
Fault Classification Results

Machine Learning Models

<table>
<thead>
<tr>
<th>Model Type</th>
<th># Classes</th>
<th>Accuracy</th>
<th>Training Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>2</td>
<td>92%</td>
<td>7 minutes</td>
</tr>
<tr>
<td>Gradient Boost</td>
<td>2</td>
<td>91%</td>
<td>20 minutes</td>
</tr>
<tr>
<td>SVC</td>
<td>2</td>
<td>98%</td>
<td>5 minutes</td>
</tr>
<tr>
<td>SVC</td>
<td>4</td>
<td>92%</td>
<td>2 hours</td>
</tr>
</tbody>
</table>

Transformer Architecture

<table>
<thead>
<tr>
<th>Model Type</th>
<th># Classes</th>
<th>Accuracy</th>
<th>Training Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformer Arch.</td>
<td>4</td>
<td>95%</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Transformer Arch.</td>
<td>46</td>
<td>97%</td>
<td>5 hours</td>
</tr>
</tbody>
</table>
Confusion Matrices

Support Vector Classification

Transformer Architecture
Confusion Matrices

Support Vector Classification

Transformer Architecture
FULL SYSTEM INTEGRATION
DEMONSTRATION
Next Steps

FPL Data
- Duplicate results with real smart grid data
- Use continuous data sources

Differential Architecture Search (DARTS) Transformer
Dynamically constructs the most optimal transformer architecture
REFERENCES


