” Using Machine Learning for Detection and Classification of PQ phenomenon”

by

Dr. Peter Axelberg

Dr. Peter Axelberg 2016
Power Quality monitoring from past until today

1st generation
- Very basic instruments like voltimeters etc were used.
- First portable power quality analyzers were introduced

2nd generation
- Permanent installed power quality analyzers were introduced.
- Power quality analyzers based on IEC 61000-4-30 class A were introduced

3rd generation
- Sophisticated signal processing methods are used for automatic and accurate analysis of measurement data
- Integration
3rd generation measurement system - objectives

"Classical"
Something happen in the grid

"Preferred"

Manual Analysis of recorded data

Only characteristics

 Tell what really happened

"Preferred"
Power Quality Monitoring – 3rd generation

- **In the past**: Evaluation of data was made by manual inspection
  - Time consuming (and tedious)
  - Quality of the result depends on the knowledge level and skills of the person doing the analysis

- **3rd generation**: Take use of information (data) from previous measurements to automate the evaluation process
  - Automatic detection and classification of power quality phenomenon
  - Less time consuming evaluation process
  - Higher quality and more reliable results
  - Important information contained in the data - hidden for a manual inspection – can be detected – Trend analysis
  - Adaptive process – by continuously adding new data from previous measurements into the machine learning algorithm the PQ system will continuously learn and increase the precision and accuracy
  - Opens up to use the PQ system in new valuable applications like fault detection and preventive maintenance etc
Machine learning (ML)

- Herbert Simon: "Learning is any process by which a system (computer, robot etc) improves performance from experience".

- Machine Learning is concerned with computer programs that automatically improve their performance through experience.

Herbert Simon

- Turing Award 1975
- Nobel Prize in Economics 1978
Pattern recognition: A machine learning algorithm that is trained to recognize a particular pattern in a large data stream.
Pattern recognition – a general technique used in many applications

- Text analysis: Search engines
- Image analysis: Face recognition
- Financial analysis: Stock price forecasts
- Power system analysis: Detection and classification of voltage disturbances, fault detection and classifications, preventive maintenance, trend analysis, forecasting
Pattern recognition methods

Many pattern recognition methods are available (different techniques with the same goal – to identify a particular pattern in a data stream)

- k Nearest Neighbor (kNN)
- Neural Networks
- AdaBoost
- Support Vector machines
A premium method: The Support Vector Machine (SVM)

- Uses large number of pre-classified training data.
- Features are extracted from the training data and placed in an n-dimensional data space.
- The classifier decides type of disturbance (class) by using an optimal separating hyperplane.
A premium method: The Support Vector Machine (SVM)

Real world

Feature extraction

Training data

Test data

Binary classifier

Class +1 $y = +1$

Class -1 $y = -1$

Separating hyperplane $f(x)$ (decision boundary)

Test data vector $x = [x_1, x_2, x_3]$

Training data vector $[x_1, x_2, x_3]$

$$f(x) = \begin{cases} 
\geq 0 \rightarrow \text{class } +1 \\
< 0 \rightarrow \text{class } -1 
\end{cases}$$
Calculating the separating hyperplane

\[ f(x) = (w \cdot x) + b \]

\[ (w \cdot x_i) + b \geq 1 \quad \text{for} \quad y_i = +1 \]

\[ (w \cdot x_i) + b \leq -1 \quad \text{for} \quad y_i = -1 \]

\[ y_i \cdot ((w \cdot x_i) + b) \geq 1 \quad i = 1,..., n \]

The margin

\[ d(w,b) = \frac{|(w \cdot x_i) + b|}{\|w\|} + \frac{|(w \cdot x_j) + b|}{\|w\|} = \frac{2}{\|w\|} \]

Maximum margin is achieved for

\[ \min_w \frac{1}{2} \|w\|^2 \]

\[ y_i \cdot ((w \cdot x_i) + b) \geq 1 \quad i = 1,..., n \]

---

**Quadratic Optimization problem (QP-problem)**
Introduce the Lagrangian functional and re-formulate the QP-problem:

\[
L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{n} \alpha_i \left[ y_i \cdot ((w \cdot x_i) + b) - 1 \right] = \frac{1}{2} \|w\|^2 - \sum_{i=1}^{n} \alpha_i \cdot y_i \cdot (w \cdot x_i) - b \sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i
\]

Optimal solution is given by the saddle point of the Lagrangian functional

\[
\max_{\alpha} \left( \min_{w, b} L(w, b, \alpha) \right)
\]

\[
\frac{dL(w, b, \alpha)}{dB} = -\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} \alpha_i y_i = 0
\]

\[
\frac{dL(w, b, \alpha)}{dW} = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \quad \Rightarrow \quad w = \sum_{i=1}^{n} \alpha_i y_i x_i
\]

\[
\max_{\alpha} L(w, b, \alpha) = \max_{\alpha} \left( \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \right)
\]

subject to the constraints

\[
\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \text{and} \quad \alpha_i \geq 0 \quad i = 1, K, n
\]
For the binary classifier

\[ f(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^{n} \alpha_i y_i (\mathbf{w} \cdot \mathbf{x}_i) + b \right) \]

Karush-Kuhn-Tucker conditions states

\[ \alpha_i (y_i ((\mathbf{w} \cdot \mathbf{x}_i) + b) - 1) = 0 \quad \text{for} \quad i = 1, ..., n \]

Only those training samples corresponding to non-zero Lagrange multipliers are needed!
These are placed on the margins and called Support Vectors

\[ f(\mathbf{x}) = \text{sign} \left( \sum_{\mathcal{SV}} \alpha_i y_i (\mathbf{x} \cdot \mathbf{x}_i) + b \right) \]
Blockdiagram of the Support Vector Machine

Support Vectors

Test data

Bias $b$

Classification

$f(x) = \text{sign} \left( \sum_{i=1}^{N} \alpha_i y_i (x_i \cdot x) + b \right)$
Conducted experiments
## Types of voltage disturbances

<table>
<thead>
<tr>
<th>Type</th>
<th>Fault types</th>
<th># of disturbances originated from Power network A</th>
<th># of disturbances Originated from Power network B</th>
<th># of disturbances originated from synthetic generated data</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>Single phase-to-ground fault</td>
<td>141</td>
<td>475</td>
<td>225</td>
</tr>
<tr>
<td>D2</td>
<td>Phase-to-phase fault</td>
<td>181</td>
<td>125</td>
<td>225</td>
</tr>
<tr>
<td>D3</td>
<td>Three-phase fault</td>
<td>251</td>
<td>196</td>
<td>223</td>
</tr>
<tr>
<td>D4</td>
<td>Double phase fault with one phase more affected</td>
<td>127</td>
<td>67</td>
<td>250</td>
</tr>
<tr>
<td>D5</td>
<td>Transformer energizing</td>
<td>214</td>
<td>0</td>
<td>250</td>
</tr>
</tbody>
</table>
## Experiment #1

<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>NC</th>
<th>Detection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single phase-to-ground fault (D1)</td>
<td>63</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>88.7 %</td>
</tr>
<tr>
<td>Phase-to-phase fault (D2)</td>
<td>0</td>
<td>84</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>92.3 %</td>
</tr>
<tr>
<td>Three-Phase Fault (D3)</td>
<td>5</td>
<td>3</td>
<td>113</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>89.7 %</td>
</tr>
<tr>
<td>Double phase fault with one phase more affected (D4)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>63</td>
<td>1</td>
<td>0</td>
<td>98.4 %</td>
</tr>
<tr>
<td>Transformer energizing (D5)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>103</td>
<td>4</td>
<td>96.3 %</td>
</tr>
</tbody>
</table>

Overall detection rate: **93.1 %**
Experiment #2

<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>NC</th>
<th>Detection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single phase-to-ground fault (D1)</td>
<td>137</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>97.2 %</td>
</tr>
<tr>
<td>Phase-to-phase fault (D2)</td>
<td>0</td>
<td>154</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>85.1 %</td>
</tr>
<tr>
<td>Three-Phase Fault (D3)</td>
<td>10</td>
<td>1</td>
<td>232</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>92.4 %</td>
</tr>
<tr>
<td>Double phase fault with one phase more affected (D4)</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>121</td>
<td>1</td>
<td>0</td>
<td>95.2 %</td>
</tr>
</tbody>
</table>

Overall detection rate: **92.5 %**
Experiment #3

Power network A

Power network B

Synthetic data

Training data

Test data

SVM

<table>
<thead>
<tr>
<th></th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>NC</th>
<th>Detection rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single phase-to-ground fault (D1)</td>
<td>393</td>
<td>0</td>
<td>2</td>
<td>23</td>
<td>94.0%</td>
</tr>
<tr>
<td>Phase-to-phase fault (D2)</td>
<td>3</td>
<td>80</td>
<td>0</td>
<td>9</td>
<td>87.0%</td>
</tr>
<tr>
<td>Three-Phase Fault (D3)</td>
<td>2</td>
<td>0</td>
<td>126</td>
<td>14</td>
<td>88.7%</td>
</tr>
</tbody>
</table>

Overall detection rate: 91.9%
System Integration
System Integration – situation until today

DISTRIBUTION GRID

- DMS
- SCADA
- AMR
- PQMS

Isolated system

European Pattern Recognition project

PART OF:
Smart Grids Plus
ERA-Net
Integration – situation tomorrow

IPQMS = Integrated Power Quality Monitoring System

IPQMS = Integrated Power Quality Monitoring System
Conclusions

- The 3rd generation system will be based on powerful pattern recognition techniques to perform traditional PQ analysis in a more efficient way
  - from manual toward automatic analysis of PQ data
  - trend analysis used for preventive maintenance
  - automatic detection and classification of faults
  - forecasting

- The 3rd generation system will be an open system ready to interchange data with other systems