Feature Studies for PV Fault Classification Using Nonlinear Principal Component Analysis

Maxwell Yarter [1], Gowtham Muniraju [1], Andreas Spanias [1], Yiannis Tofis [2]
MOTIVATION

- Automatic solar fault detection is more efficient and cost effective
- Ten solar features are used for classification
  - Less features can shorten training time
  - Knowing which features are redundant informs us which sensors are needed
- Greater classification accuracy means more power output
Nonlinear PCA techniques may reduce the number of solar features needed for fault classification and improve classification accuracy.

Challenges:
- Autoencoder only eliminate redundancy and do not perfectly emulate the input data
- Using KPCA requires training 9 different classification networks per kernel function
NONLINEAR PCA METHODS

KPCA Block Diagram

Autoencoder Block Diagram
- **Data Set:** NREL solar testbed 10 feature data set [1]
  - **Features:** DC Power, Max. Voltage, Max. Current, Temperature, Irradiance, Fill Factor, Gamma, Max. Power, Open Circuit Voltage, Short Circuit Current
  - **Faults:** Standard Test Condition, Short Circuit, Degraded, Shaded, Soiled

- **Kernel Functions:** linear, polynomial, RBF, sigmoid, and cosine.
RESULTS

KPCA Fault Classification Accuracy vs. Dimension of Embedding Space

<table>
<thead>
<tr>
<th>MRun</th>
<th>KernelType</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>sigmoid</td>
<td>45.43</td>
<td>67.23</td>
<td>81.58</td>
<td>84.37</td>
<td>83.71</td>
<td>83.15</td>
<td>87.29</td>
<td>88.47</td>
<td>86.58</td>
</tr>
<tr>
<td>1</td>
<td>rbf</td>
<td>53.70</td>
<td>66.02</td>
<td>81.45</td>
<td>81.45</td>
<td>82.28</td>
<td>82.74</td>
<td>80.35</td>
<td>81.39</td>
<td>83.80</td>
</tr>
<tr>
<td>2</td>
<td>linear</td>
<td>53.92</td>
<td>65.71</td>
<td>83.34</td>
<td>85.83</td>
<td>85.10</td>
<td>86.34</td>
<td>85.05</td>
<td>86.34</td>
<td>87.30</td>
</tr>
<tr>
<td>3</td>
<td>poly</td>
<td>57.61</td>
<td>65.10</td>
<td>83.12</td>
<td>75.74</td>
<td>72.35</td>
<td>68.68</td>
<td>71.92</td>
<td>71.96</td>
<td>77.92</td>
</tr>
<tr>
<td>4</td>
<td>cosine</td>
<td>54.57</td>
<td>75.96</td>
<td>80.20</td>
<td>84.05</td>
<td>85.01</td>
<td>82.76</td>
<td>86.11</td>
<td>85.47</td>
<td>84.51</td>
</tr>
</tbody>
</table>
Autoencoder Fault Classification Accuracy vs. Dimension of Embedding Space

<table>
<thead>
<tr>
<th>MRun</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
<th>#7</th>
<th>#8</th>
<th>#9</th>
<th>#10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>46.49360</td>
<td>66.28917</td>
<td>67.892305</td>
<td>18.95749</td>
<td>75.253972</td>
<td>77.337479</td>
<td>74.107975</td>
<td>78.529232</td>
<td>74.728513</td>
</tr>
<tr>
<td>1</td>
<td>60.114801</td>
<td>70.415759</td>
<td>65.432209</td>
<td>77.086348</td>
<td>72.164300</td>
<td>80.096165</td>
<td>72.354949</td>
<td>60.515051</td>
<td>77.226166</td>
</tr>
<tr>
<td>2</td>
<td>19.360443</td>
<td>63.403600</td>
<td>60.849333</td>
<td>72.329924</td>
<td>67.219979</td>
<td>69.950304</td>
<td>71.393114</td>
<td>72.572136</td>
<td>76.103321</td>
</tr>
<tr>
<td>3</td>
<td>61.340368</td>
<td>62.224032</td>
<td>67.018306</td>
<td>74.681973</td>
<td>72.091222</td>
<td>74.968137</td>
<td>75.054300</td>
<td>76.140243</td>
<td>72.122240</td>
</tr>
<tr>
<td>4</td>
<td>19.484932</td>
<td>56.870830</td>
<td>62.007445</td>
<td>73.487437</td>
<td>73.611540</td>
<td>74.154514</td>
<td>77.736130</td>
<td>80.005155</td>
<td>78.793049</td>
</tr>
</tbody>
</table>
- Overlap in STC and Shaded feature clusters for both nonlinear techniques
- Confusion matrix shows STC and shaded misclassification

5 Feature Linear KPCA Confusion Matrix
CONCLUSION

- Successful reduction of feature space
- 85.1% Accuracy using linear kernel and 5 features.
- Autoencoder <80% accuracy for all dimensions
- No nonlinear redundancy in the feature set
ongoing & planned work

- Determine a feature that could distinguish between STC and shaded fault
- Verify these results using more data
- Complete IEEE format report detailing results
- Consolidate nonlinear and linear PCA results into a single paper
REFERENCES


