Squeezed Convolutional Variational AutoEncoder

Presenter: Keren Ye

Background - Anomaly detection

No labels for defects and malfunctions

A diagnosis of the behavior pattern of processes

- Assumed that an abnormal behavior pattern means an anomaly
Background - Edge computing

Cloud-based approach

- Central server to process data

Edge-based approach

- Receiving more attention
- Real time inference on edge devices
- Low costs in communication compared to cloud-based
- Reduces the burden on communication and computer infrastructure
- Neural networks (NN): compute and memory intensive, yet NN performs well
Contribution - SCVAE

Unsupervised

• labeled data is not required

Time series sensor data

• a practical case of Industrial Internet of Things (IIoT)

Anomaly detection

• Proposed to use a specific NN model to handle the problem
Contribution - SCVAE

Edge computing

- Propose ways to reduce model size and inference time on edge devices

Evaluation

- Match-General metric for unlabeled data
Related work - DL based anomaly detection

Variational AutoEncoder (VAE)

- Model the data distribution, then try to reconstruct the data
- Outliers that cannot be reconstructed are anomalous

Generative Adversarial Networks (GAN)

- G model: generate data to fool D model
- D model: determine if the data is generated by G or from the dataset

Related work - DL based anomaly detection

Variational AutoEncoder (VAE)

- The label is the same as the input
- Outliers cannot be reconstructed

Image credit (left): Introduction to Principal Component Analysis (PCA)
https://docs.opencv.org/3.1.0/d1/dee/tutorial_introduction_to_pca.html

Image credit (right): Keras Tutorial: Content Based Image Retrieval Using a Convolutional Denoising Autoencoder
Related work - Time series data

Fault Detection and Classification Convolutional Neural Networks

- CNN’s receptive field was matched to the multivariate sensor signal
- The CNN filter was moved along the time axis to extract meaningful features from the sensor data

CNN demo: [LINK]

Note: the kernel is learnable

Related work - Time series data

Benefits of using CNN

- 1D conv filter along the time axis can fill out missing value using historical information
- 1D conv filter along the sensors axis can fill out missing value using data from other sensors
- 2D convolutional filter utilizes both information

Autoregression is a special case of CNN

- 1D conv filter, kernel size equals the input size

Related work - Edge computing

Reduce both the size and inference time

- Pruning and weight quantization (Deep compression)
- Modified CNN structure (SqueezeNet, MobileNet)

SqueezeNet (MobileNet is similar)

- Fire module: decompose a conv layer to a squeeze layer and an expand layer

Method - CNN-Variational Autoencoder

Input data - expressed in the form of 2D image data

- # of features
- Time windows
Method - CNN-Variational Autoencoder

Network architecture

\[ \text{Anomaly Score} = 1 - E_{q_\theta(z|x_i)}[p_\phi(x_i|z)] \]

Explanation (Inference)

- Use input x to generate latent factor z
- Use z to reconstruct the distribution of x
- If x is not (a pre-defined threshold) from the distribution, it means x is an outlier
- During training, the label is x itself
Method - Squeezed Architecture

Substitute conv layer to Fire modules used in SqueezeNet

- Squeeze layer: 1x1 conv filter
- Extend layer: 1x1 conv filter, 3x3 conv filter

For example, 100x100x3 2D image

- Squeeze layer: 1x1 conv filter may map the image to gray scale, output=100x100x1 (but you could use multiple such filters, e.g. 64)
- Extend layer: 1x1 conv filter do a one-on-one mapping, while 3x3 conv filter may have smooth effect (smooth operation on the gray scale image)
**Experiment**

Performance compared to baseline (IF, LOF, OCSVM, EE) on labeled data

Performance for unlabeled real world Computer Numerical Control (CNC) data

Model size and inference time
Experiment - labeled data

Dataset

- Ozone, Occupancy

Label

- Ozone: normal / ozone day
- Occupancy: someone in the office room or not

<table>
<thead>
<tr>
<th>Dataset</th>
<th># of Samples</th>
<th># of Features</th>
<th>Anomaly Ratio(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ozone</td>
<td>2536</td>
<td>73</td>
<td>2.87</td>
</tr>
<tr>
<td>occupancy</td>
<td>8143</td>
<td>6</td>
<td>21.23</td>
</tr>
</tbody>
</table>

TABLE III. TIMESERIES LABELED DATASET


Experiment - labeled data

Dataset

- Ozone, Occupancy

Evaluation

- Metric: Area under the precision-recall curve (PRAUC)
- Anomaly labels are included in these dataset, yet not used during training

<table>
<thead>
<tr>
<th>Dataset (Time Window)</th>
<th>IF</th>
<th>LOF</th>
<th>OCSVM</th>
<th>EE</th>
<th>SCVAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ozone(4)</td>
<td>85.71</td>
<td>88.72</td>
<td>88.18</td>
<td>86.21</td>
<td>95.35</td>
</tr>
<tr>
<td>ozone(8)</td>
<td>83.02</td>
<td>83.27</td>
<td>80.06</td>
<td>81.99</td>
<td>96.89</td>
</tr>
<tr>
<td>ozone(16)</td>
<td>68.05</td>
<td>68.83</td>
<td>64.70</td>
<td>69.26</td>
<td>81.78</td>
</tr>
<tr>
<td>occupancy(4)</td>
<td>95.93</td>
<td>74.07</td>
<td>80.99</td>
<td>99.17</td>
<td>98.54</td>
</tr>
<tr>
<td>occupancy(8)</td>
<td>96.81</td>
<td>74.89</td>
<td>81.19</td>
<td>98.76</td>
<td>99.10</td>
</tr>
<tr>
<td>occupancy(16)</td>
<td>96.62</td>
<td>75.57</td>
<td>78.39</td>
<td>99.21</td>
<td>99.23</td>
</tr>
</tbody>
</table>

TABLE IV. PRAUC COMPARISON ON TIMESERIES LABELED DATASET


Experiment - unlabeled CNC data

Metric - Match-general

- common labels are created and evaluated as actual labels.
- Anomaly in common labels are labeled as anomaly if the 3 or more of the 6 models identify the label as anomaly.
- The higher the better

CNC Dataset

<table>
<thead>
<tr>
<th>CNC</th>
<th># of Samples</th>
<th># of Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>258697</td>
<td>31</td>
</tr>
<tr>
<td>B</td>
<td>310174</td>
<td>43</td>
</tr>
<tr>
<td>C</td>
<td>111770</td>
<td>43</td>
</tr>
<tr>
<td>D</td>
<td>602075</td>
<td>37</td>
</tr>
</tbody>
</table>

TABLE V. CNC Dataset
Experiment - unlabeled CNC data

<table>
<thead>
<tr>
<th>CNC</th>
<th>IF</th>
<th>LOF</th>
<th>OCSVM</th>
<th>EE</th>
<th>CNNVAE</th>
<th>SCVAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A(4)</td>
<td>0.9435</td>
<td>0.9565</td>
<td>0.4806</td>
<td>0.9532</td>
<td>0.9435</td>
<td>0.9371</td>
</tr>
<tr>
<td>A(8)</td>
<td>0.9677</td>
<td>0.9323</td>
<td>0.6484</td>
<td>0.9645</td>
<td>0.9290</td>
<td>0.9355</td>
</tr>
<tr>
<td>A(16)</td>
<td>0.9516</td>
<td>0.9516</td>
<td>0.4532</td>
<td>0.9548</td>
<td>0.9355</td>
<td>0.9387</td>
</tr>
<tr>
<td>B(4)</td>
<td>0.9413</td>
<td>0.9329</td>
<td>0.3208</td>
<td>0.9329</td>
<td>0.9665</td>
<td>0.9665</td>
</tr>
<tr>
<td>B(8)</td>
<td>0.9266</td>
<td>0.9350</td>
<td>0.2180</td>
<td>0.9392</td>
<td>0.9686</td>
<td>0.9602</td>
</tr>
<tr>
<td>B(16)</td>
<td>0.9476</td>
<td>0.9476</td>
<td>0.3585</td>
<td>0.9434</td>
<td>0.9560</td>
<td>0.9560</td>
</tr>
<tr>
<td>C(4)</td>
<td>0.9686</td>
<td>0.9507</td>
<td>0.5695</td>
<td>0.9327</td>
<td>0.9507</td>
<td>0.9327</td>
</tr>
<tr>
<td>C(8)</td>
<td>0.9552</td>
<td>0.9462</td>
<td>0.8565</td>
<td>0.9462</td>
<td>0.9372</td>
<td>0.9552</td>
</tr>
<tr>
<td>C(16)</td>
<td>0.9507</td>
<td>0.9507</td>
<td>0.4215</td>
<td>0.9417</td>
<td>0.9507</td>
<td>0.9596</td>
</tr>
<tr>
<td>D(4)</td>
<td>0.9444</td>
<td>0.9493</td>
<td>0.6221</td>
<td>0.9493</td>
<td>0.9510</td>
<td>0.9510</td>
</tr>
<tr>
<td>D(8)</td>
<td>0.9535</td>
<td>0.9419</td>
<td>0.7467</td>
<td>0.9452</td>
<td>0.9485</td>
<td>0.9568</td>
</tr>
<tr>
<td>D(16)</td>
<td>0.9394</td>
<td>0.9277</td>
<td>0.1312</td>
<td>0.9377</td>
<td>0.9560</td>
<td>0.9626</td>
</tr>
</tbody>
</table>

TABLE VI. PERFORMANCE(MG) COMPARISON ON CNC DATASET
Experiment - model size and inference time

Baseline: CNN version v.s. SqueezeNet version

- Performance (PRAUC) is similar
- Inference time and memory usage are reduced

<table>
<thead>
<tr>
<th>Model</th>
<th>Learning Time</th>
<th>Inference Time</th>
<th>Memory(Mb)</th>
<th>PRAUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNNVAE</td>
<td>3m 8sec</td>
<td>0.0088</td>
<td>16</td>
<td>98.9</td>
</tr>
<tr>
<td>SCVAE</td>
<td>2m 1sec</td>
<td>0.0060</td>
<td>12</td>
<td>99.2</td>
</tr>
</tbody>
</table>

TABLE VII. COMPARISON BETWEEN SCVAE AND CNN-VAE ON OCCUPANCY(16) DATASET
Conclusion

Unsupervised

- labeled data is not required

Time series sensor data

- a practical case of Industrial Internet of Things (IIoT)

Anomaly detection

- Proposed to use a specific NN model to handle the problem

Edge computing

- Propose ways to reduce model size and inference time on edge devices
Thanks