Modeling Time-Varying Data
Attention, Generative Models and Beyond

August 7, 2018

Jayaraman J. Thiagarajan
Computer Scientist
Predictive Modeling with Time-Varying Data is Ubiquitous

Time-Series Modeling is at the Heart of Clinical Diagnosis

Understanding Time-Varying Phenomena Critical in Science
Predictive Modeling with Time-Varying Data is Ubiquitous

Advances in Sensing are Driving IOT technologies

Modeling time-varying data amounts to finding mapping between the measurements and the dynamic system
Creating representations for dynamic systems from observations is not straightforward – No known priors

- Complex dependencies
- Multi-variate measurements
- Irregular sampling
- Missing values and measurement errors
- Making progressive predictions is computationally expensive
Deep Neural Networks Have Become the De-facto Solution for Sequence Modeling

Recurrent Neural Nets with LSTM or GRU units

Convolutional Neural Nets

Differentiable Neural Computers
Can we parameterize the dependencies through simpler constructs than LSTMs?

Temporal dependence can be viewed as a “network” structure

Parameterized attention models are an effective alternative

Loosely based on human visual attention

Used in NLP to enhance LSTM-RNN by providing context
Simply Attend and Diagnose – An Attention-only Architecture for Modeling EHR Data

- Stacked multi-head attention modules and 1D CNN feed-forward layers
- Encoding temporal order partially
- Effective training with skip connections and dropout
- Remarkable gains in reducing sequence modeling complexity
Significant Complexity Reduction and State-of-the-art Performance

**MIMIC-III Database** – Largest public repository of ICU patient records

**Tasks** – Mortality, length of stay, decompensation, phenotyping

State-of-the-art results in all tasks, and outperforms RNNs

*AAAI 2018*
Rethinking Deep Models for Sequential Data
Learn a Metric to Compare Time-Series Data

**A not-so-trivial problem: How do we sample?**
Attention-Based Deep Metric Learning for Speaker Diarization

Audio track:

Diarization result:

The Question of Who Speaks When?
Learned Metric Generalizes Effectively to Novel Data Distributions

State-of-the-art results in diarization with challenging benchmarks

* Interspeech 2018
Rethinking Deep Models for Sequential Data
Unsupervised Pre-Training with Generative Models

- Predictive modeling is at the heart of clinical diagnosis – Discriminative Models
  - Less robust to shifts in domain
  - Prone to overfitting and sensitive to initialization, hyperparameters etc.
  - Cannot deal with out-of-distribution anomalies

Detecting heart conditions using limited channel ECG

Standard 12-Channel ECG Montage
Generative Models Provide a Task-Independent Statistical Description of the Entire Distribution

- Estimate information (implicitly) about missing channels through a generative model

Random forests
ResNet with 1-D CNNs
InceptionNet with 1-D CNNs

What loss to use?
Significant Improvements in Disease Prediction and Generalization

<table>
<thead>
<tr>
<th>Disease</th>
<th>Channels</th>
<th>Acc.%</th>
<th>Sens.%</th>
<th>Spec.%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocardial Infarction</td>
<td>V1, V2, V3</td>
<td>86</td>
<td>96</td>
<td>84</td>
</tr>
<tr>
<td>Bundle Branch Block</td>
<td>V1, V6</td>
<td>94</td>
<td>97</td>
<td>99</td>
</tr>
</tbody>
</table>

Channel Configuration: II, III, aVF
Input Classes: Healthy Control, Infero Myocardial Infarction

<table>
<thead>
<tr>
<th>Cardiac Disease</th>
<th>ResNet</th>
<th>ResNet++</th>
<th>% Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infero MI</td>
<td>0.84</td>
<td>0.87</td>
<td>3.57</td>
</tr>
<tr>
<td>Antero MI</td>
<td>0.87</td>
<td>0.89</td>
<td>2.30</td>
</tr>
<tr>
<td>Bundle Branch Block</td>
<td>0.59</td>
<td>0.66</td>
<td>11.86</td>
</tr>
<tr>
<td>Dysrhythmia</td>
<td>0.67</td>
<td>0.71</td>
<td>5.97</td>
</tr>
<tr>
<td>Cardiomyopathy</td>
<td>0.77</td>
<td>0.81</td>
<td>5.19</td>
</tr>
<tr>
<td>Valvular Heart Disease</td>
<td>0.32</td>
<td>0.33</td>
<td>3.13</td>
</tr>
</tbody>
</table>

*IEEE EMBC 2018
Conclusions

- Temporal modeling is a central, yet challenging problem in science and engineering

- Wide-range of applications: Robust predictions, sensing limitations, metric learning, anomaly detection, multi-task learning, correlation studies, interpolation etc.

- Key advances in deep learning solutions for time-series problems: Attention models, Deep metric learning, Generative modeling
Collaborators

Deepta Rajan
IBM Research

Huan Song
Bosch Research

Rushil Anirudh
LLNL

[1] ”Attend and Diagnose: Clinical Time-series Modeling using Attention Models”, AAAI 2018

