Modeling Time-Varying Data

Attention, Generative Models and Beyond

CASC
Center for Applied Scientific Computing

August 7, 2018

Jayaraman J. Thiagarajan
Computer Scientist

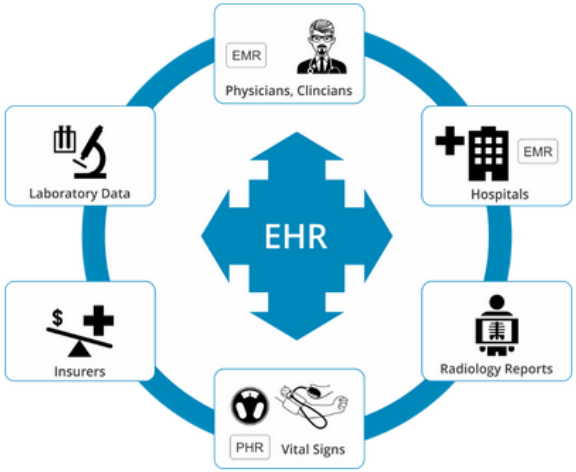
LLNL-PRES-755086

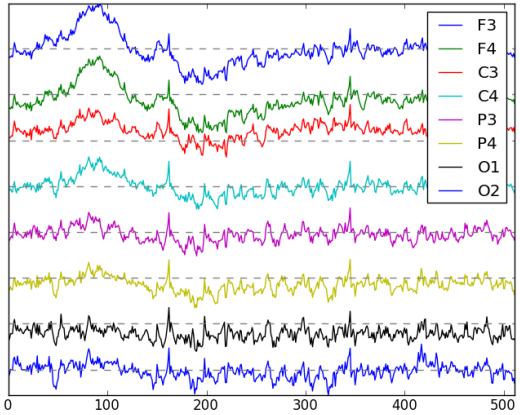
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

Predictive Modeling with Time-Varying Data is Ubiquitous

Time-Series Modeling is at the Heart of Clinical Diagnosis

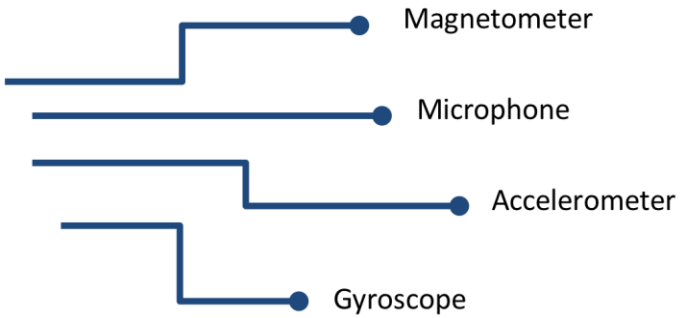


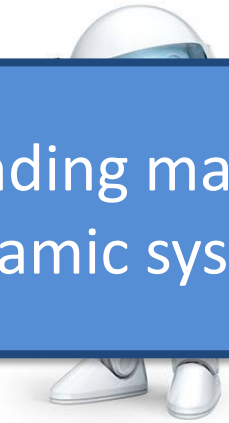


Understanding Time-Varying Phenomena Critical in Science

Predictive Modeling with Time-Varying Data is Ubiquitous

Advances in Sensing are Driving IOT technologies





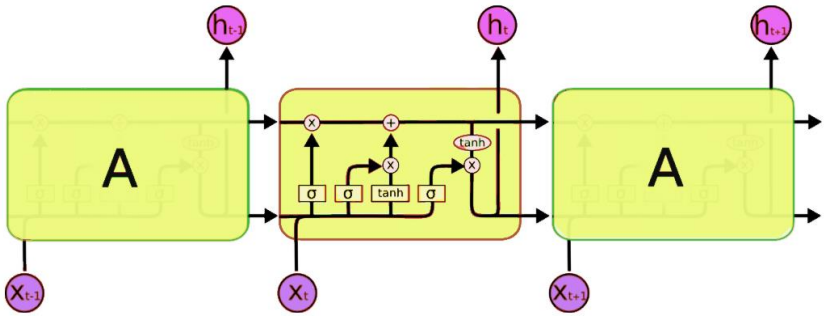
Modeling time-varying data amounts to finding mapping between the measurements and the dynamic system

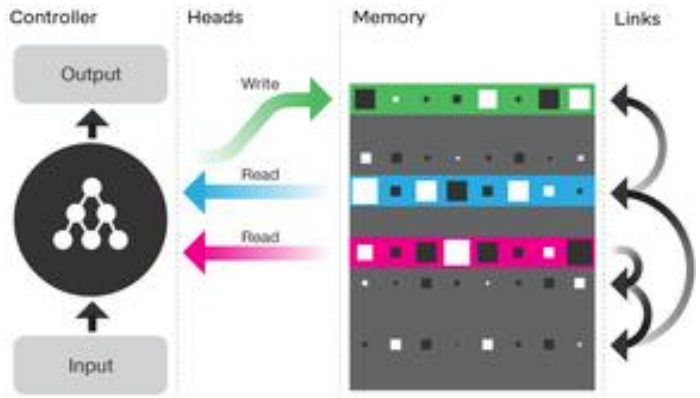
Predictive Modeling with Time-Varying Data is Ubiquitous and Challenging

- 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

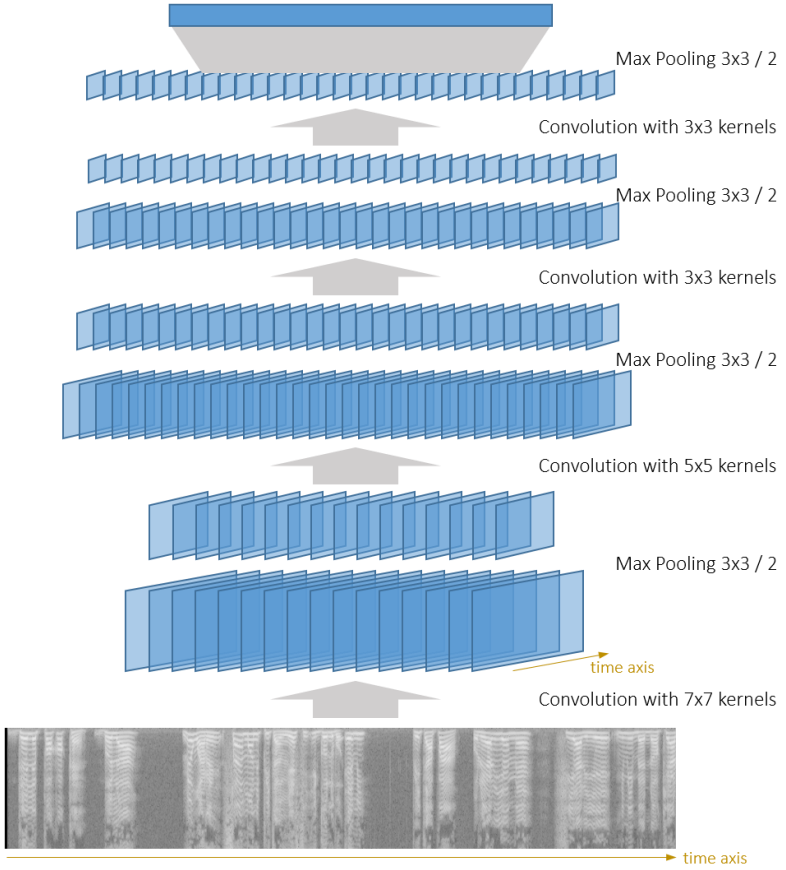
Recurrent Neural Nets with LSTM or GRU units





Differentiable Neural Computers

Convolutional Neural Nets



Rethinking Deep Models for Sequential Data
Attention Models

- 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

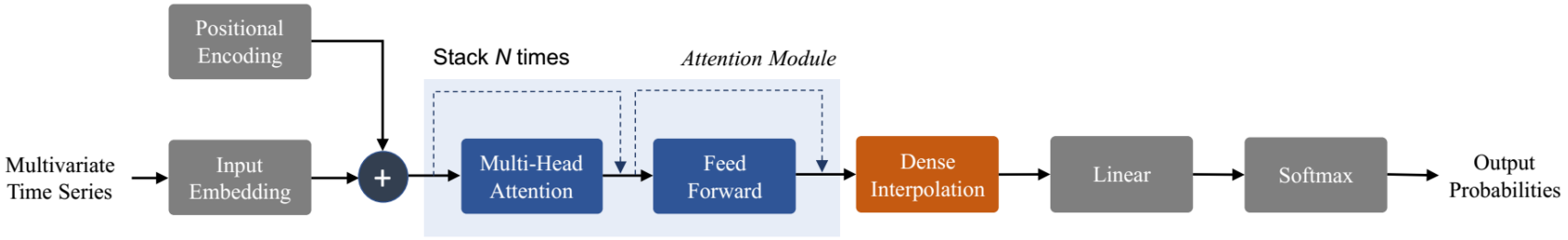
by *ent423* , *ent261* correspondent updated 9:49 pm et , thu march 19 , 2015 (*ent261*) a *ent114* was killed in a parachute accident in *ent45* , *ent85* , near *ent312* , a *ent119* official told *ent261* on wednesday . he was identified thursday as special warfare operator 3rd class *ent23* , 29 , of *ent187* , *ent265* . `` *ent23* distinguished himself consistently throughout his career . he was the epitome of the quiet professional in all facets of his life , and he leaves an inspiring legacy of natural tenacity and focused

. . .

ent119 identifies deceased sailor as **X** , who leaves behind a wife

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

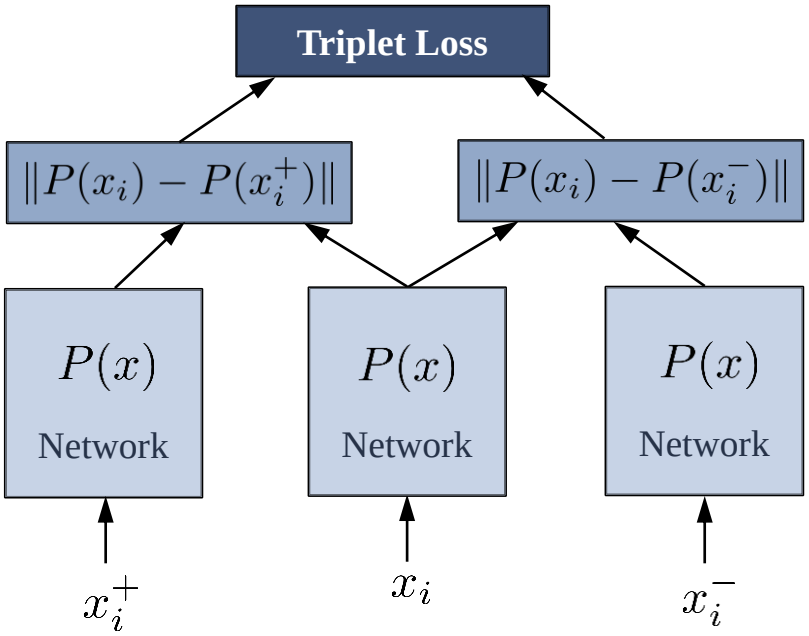
Tasks – Mortality, length of stay, decompensation, phenotyping

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



Face Verification

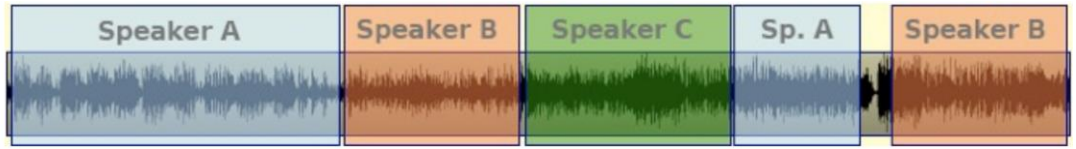
Face Clustering

**A not-so-trivial problem: How do we sample?

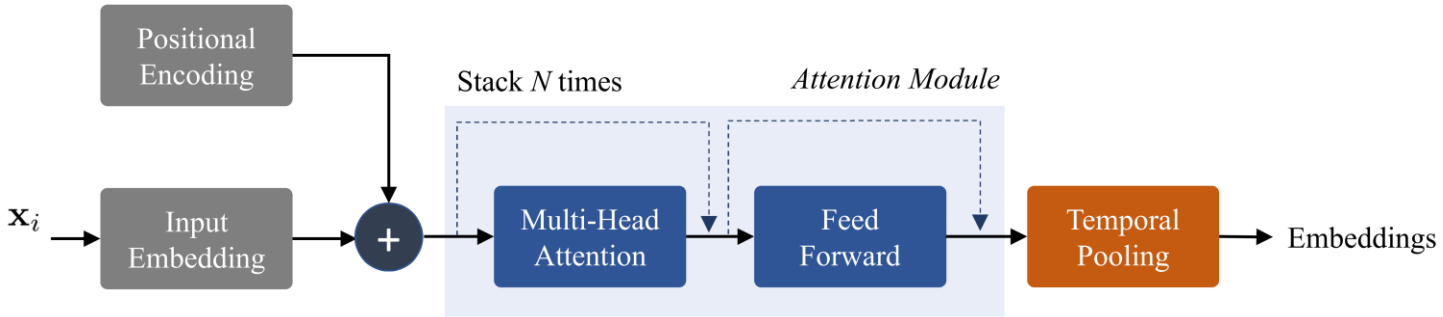
Attention-Based Deep Metric Learning for Speaker Diarization

Audio track:

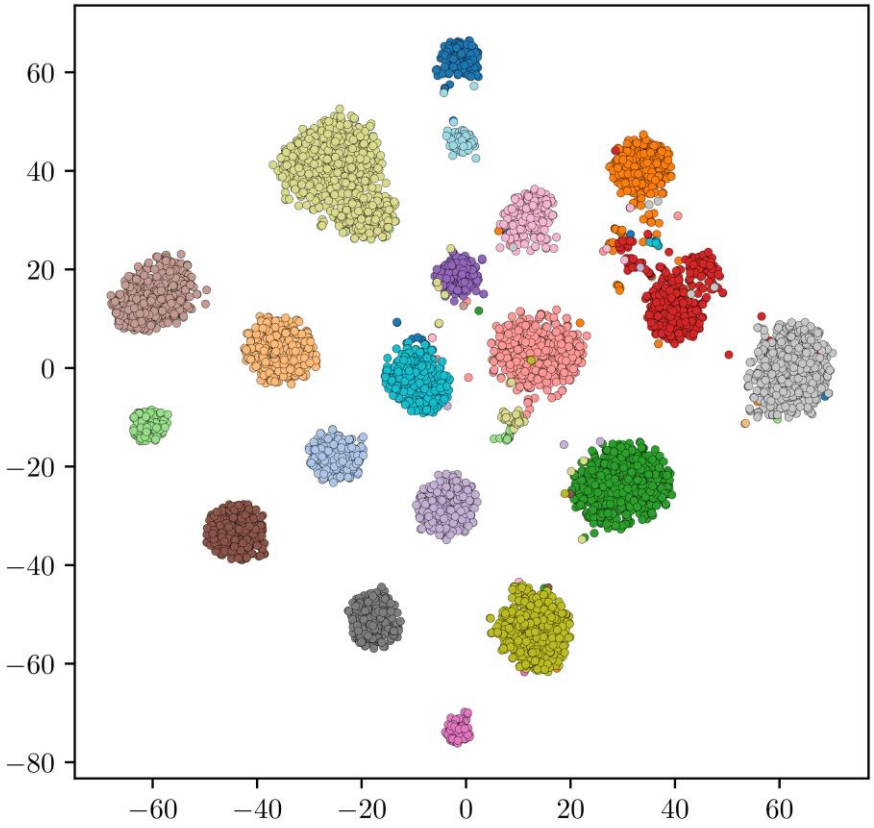
Diarization result:



The Question of Who Speaks When?



Learned Metric Generalizes Effectively to Novel Data Distributions



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

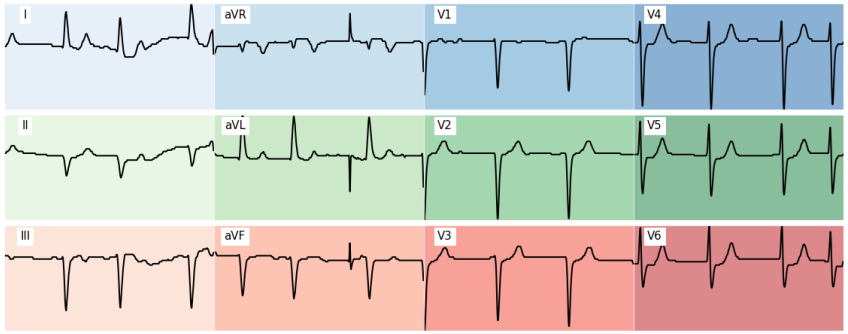
* Interspeech 2018

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

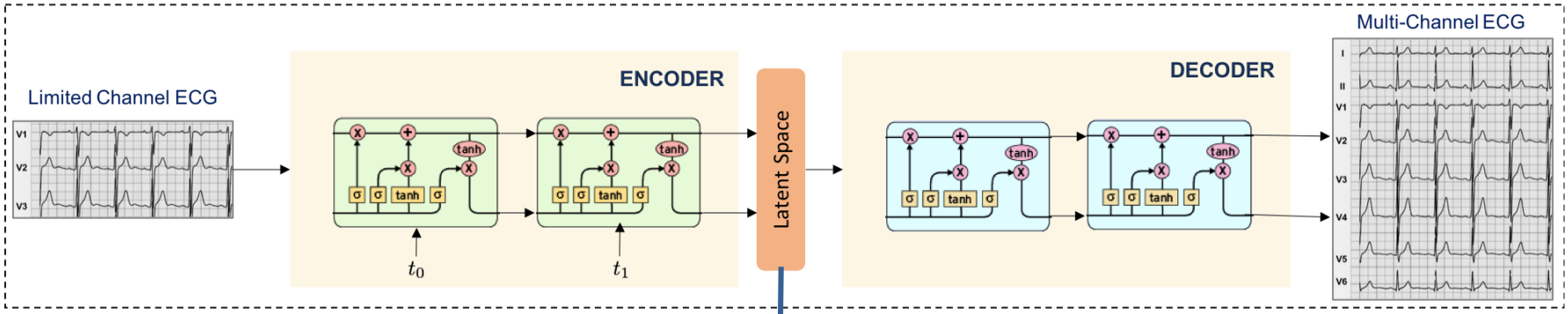
Detecting heart conditions using limited channel ECG



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



What loss to use?

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

Predictive Model

Significant Improvements in Disease Prediction and Generalization

Disease Prediction

Disease	Channels	Acc.%	Sens.%	Spec.%
Myocardial Infarction	V1, V2, V3	86	96	84
Bundle Branch Block	V1, V6	94	97	99

Generalization

Channel Configuration: II, III, aVF Input Classes: Healthy Control, Infero Myocardial Infarction			
Cardiac Disease	**ResNet**	**ResNet++**	**% Gain**
Infero MI	0.84	**0.87**	3.57
Antero MI	0.87	**0.89**	2.30
Bundle Branch Block	0.59	**0.66**	11.86
Dysrhythmia	0.67	**0.71**	5.97
Cardiomyopathy	0.77	**0.81**	5.19
Valvular Heart Disease	0.32	**0.33**	3.13

*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

[2] “Triplet Network with Attention for Speaker Diarization”, *Interspeech* 2018

[3] “A Generative Modeling Approach to Limited Channel ECG Classification”, *IEEE EMBC* 2018

CASC

Center for Applied
Scientific Computing

Lawrence Livermore
National Laboratory