UPPSALA
UNIVERSITET

Deep Learning Applied to System Identification

Niklas Wahlström, Uppsala University

Joint work with **Carl Andersson**, **Antonio Ribiero**, **Koen Tiels**, **Thomas Schön**.

Aalto University, Helsinki
December 20, 2019.

Dynamical models in system identification

A dynamical system that evolves over time and it has a memory

$$\text{Input} \xrightarrow{\;u\;} \boxed{\text{System}} \xrightarrow{\;y\;} \text{Output}$$

- Linear time-invariant dynamical system described by

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)d\tau + e(t).$$

- Nonlinear autoregressive model with exogenous (NARX) inputs

$$y_t = g(y_{t-1}, \ldots, y_{t-n_y}, u_t, \ldots, u_{t-n_u}) + e(t)$$

Impulse response estimation

Linear time-invariant dynamical system described by

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)d\tau + e(t)$$

Task: Learn a model of the underlying impulse based on data $\mathbf{u} = (u(1), \ldots, u(N))^T$ and $\mathbf{y} = (y(1), \ldots, y(N))^T$.

Placing a GP prior on the impulse response offers **better** performance than the "classical" system identification approach.

Gianluigi Pillonetto and Giuseppe De Nicolao. **A new kernel-based approach for linear system identification.** Automatica, 46(1):81–93,2010.

Gianluigi Pillonetto, Francesco Dinuzzo, Tianshi Chen, Giuseppe De Nicolao and Lennart Ljung. **Kernel methods in system identification, machine learning and function estimation: A survey.** Automatica, 50(3):657–682, 2014

Impulse response estimation - problem formulation

Consider the discrete time version

$$y(t) = G_0(q, \boldsymbol{\theta})u(t) + e(t)$$

Approximate the impulse response with a finite impulse response

$$G_0(q, \boldsymbol{\theta}) \approx G(q, \boldsymbol{\theta}) = b_1 q^{-1} + \cdots + b_n q^{-n}$$

Problem: Estimate the finite impulse response $\boldsymbol{\theta} = [b_1, \ldots, b_n]$ based on data $\mathbf{u} = (u(1), \ldots, u(N))^T$ and $\mathbf{y} = (y(1), \ldots, y(N))^T$.

Find $\boldsymbol{\theta}$ by solving the regularized optimization problem

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{N} (y(t) - G(q, \boldsymbol{\theta})u(t))^2 + \boldsymbol{\theta}^T D \boldsymbol{\theta},$$

How do we find D?

Impulse response estimation - our solution

Our solution: Model D with a deep neural network

$$D = f_{\mathrm{DL}}(\boldsymbol{u}, \boldsymbol{y}; \eta).$$

Sample stable systems on the form

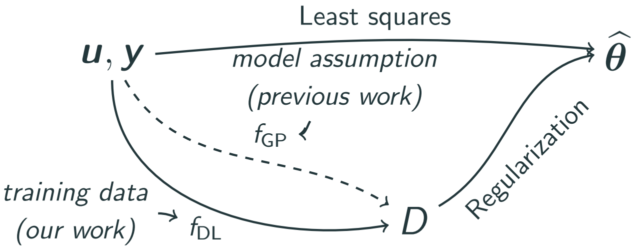
$$G^{(i)}(q) = \frac{B(q)}{F(q)} = \frac{b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}}{f_1 q^{-1} + \cdots + f_{n_f} q^{-n_f}}$$

systems, simulate $\boldsymbol{u}^{(i)}$, and $\boldsymbol{y}^{(i)}$ and find η by minimizing

$$\widehat{\eta} = \arg\min_{\eta} \frac{1}{M} \sum_{i=1}^{M} \left\| \widehat{\boldsymbol{\theta}} \left(f_{\mathrm{DL}}(\boldsymbol{u}^{(i)}, \boldsymbol{y}^{(i)}; \eta) \right) - \boldsymbol{\theta}_0^{(i)} \right\|^2,$$

where $\boldsymbol{\theta}_0^{(i)}$ is the true truncated impulse response.

Impulse response estimation - our solution

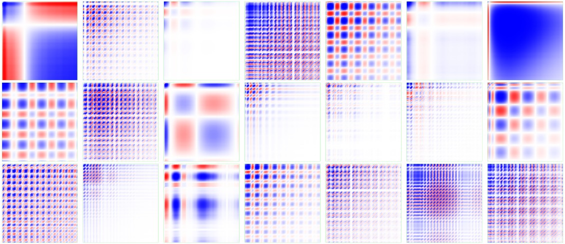


C. Andersson, N. Wahlström, and T. B. Schön. **Data-driven impulse response regularization via deep learning.** In Proceedings of the 18th IFAC Symposium on System Identification (SYSID), Stockholm, Sweden, 2018.

Impulse response estimation - results

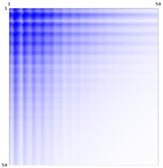
- We generated 1 000 000 stables systems of order 30.
- Impulse response of length $n = 50$

D described as a weighted sum of multiple low-rank matrices D_i.
Different D_i visualized below

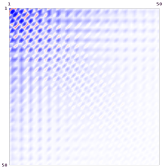


Impulse response estimation - results

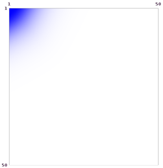
Comparison between inverse regularization matrices for a particular example



Optimal regularization



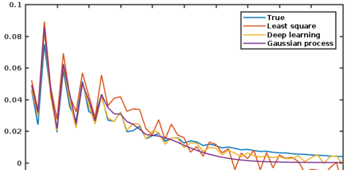
Deep learning regularization



Gaussian process regularization

Optimal regularization: $D^{-1} = \boldsymbol{\theta}_0 \boldsymbol{\theta}_0^{\mathsf{T}}$

The estimated impulse responses for a particular example



Impulse response estimation - results

The deep learning regularization performed comparable or even better than previous methods.

Model	SNR < 5.5	SNR > 5.5
LS	1	1
Opt.	0.04	0.05
GP	0.31	0.40
DL	0.20	0.23

LS $=$ Least square (no regularization),

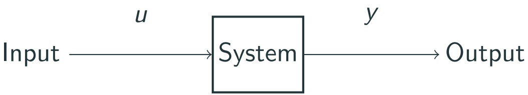
OR $=$ Optimal regularization,

GP $=$ Gaussian process regularization,

DL $=$ Deep learning regularization

Dynamical models in system identification

A dynamical system that evolves over time and it has a memory



- Linear time-invariant dynamical system described by

$$y(t) = \int_0^\infty g(\tau)u(t-\tau)d\tau + e(t).$$

- Nonlinear autoregressive model with exogenous (NARX) inputs

$$y_t = g(y_{t-1}, \ldots, y_{t-n_y}, u_t, \ldots, u_{t-n_u}) + e(t)$$

Temporal convolutional network

- TCNs can be seen as an extension of the NARX model

$$\hat{y}_{t+1} = g(x_t, \dots x_{t-(n-1)}), \quad \text{where} \quad x_t = (u_t, y_{t-1})^T$$

- A full TCN is a sequential construction of several NARX models:

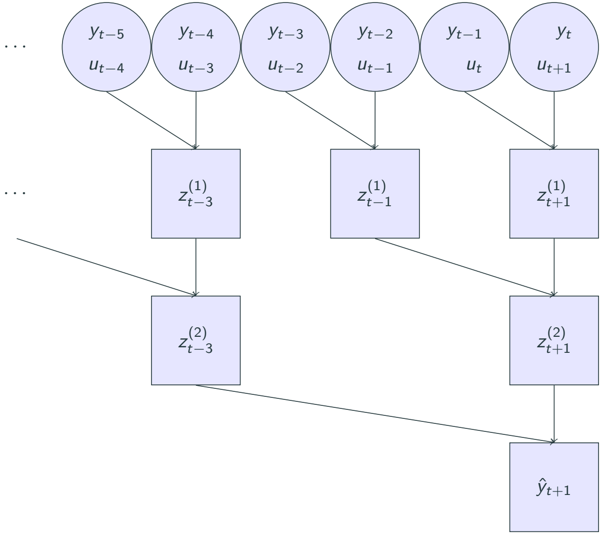
$$\begin{aligned} \hat{y}_{t+1} &= g^{(L)}(Z_t^{(L-1)}), \\ z_t^{(l)} &= g^{(l)}(Z_t^{(l-1)}), \quad l = 1, \dots, L-1, \\ z_t^{(0)} &= x_t, \end{aligned}$$

where

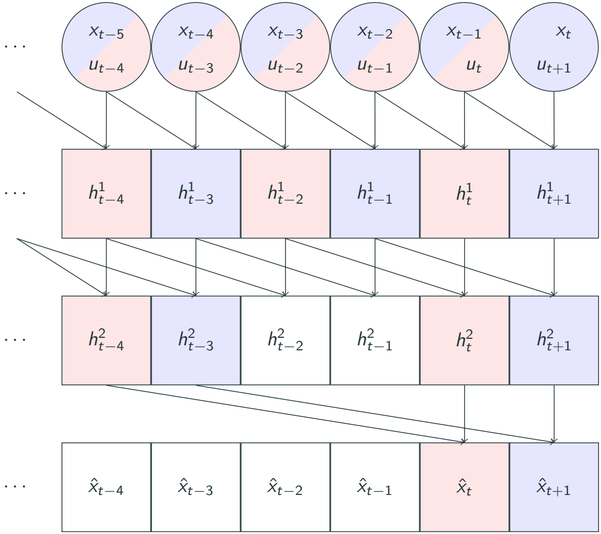
$$Z_t^{(l-1)} = \left(z_t^{(l-1)}, z_{t-d_l}^{(l-1)}, \dots, z_{t-(n-1)d_l}^{(l-1)} \right).$$

- For each layer, a *dilation factor* d_l is introduced, typically $d_l = 2^{l-1}$.

Temporal convolutional network



Temporal convolutional network



Example: Physical systems - F-16 Ground Vibration Test

- Vibration test on an F-16
- Non-linear due to structural dynamics
- No long term dependencies
- Small amount of process noise

Mode	LSTM	MLP	TCN
Free-run simulation	0.74	0.48	0.63
One-step-ahead prediction	0.023	0.045	0.034

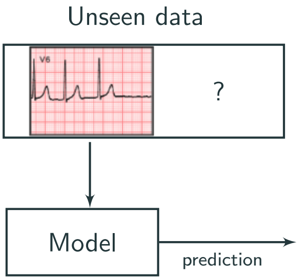
RMSE for F16 data

C. Andersson, A.H. Ribeiro, K. Tiels, N. Wahlström, and T. B. Schön. **Deep convolutional networks in system identification**, Nice, France, 2019. In Proceedings of the IEEE 58th Conference on Decision and Control (CDC).

ECG classification – the CODE study

Aim: Predict abnormalities based on a short-duration 12-lead electrocardiogram (ECG) recording.

Current situation: The automated diagnosis that is currently available is not useful.



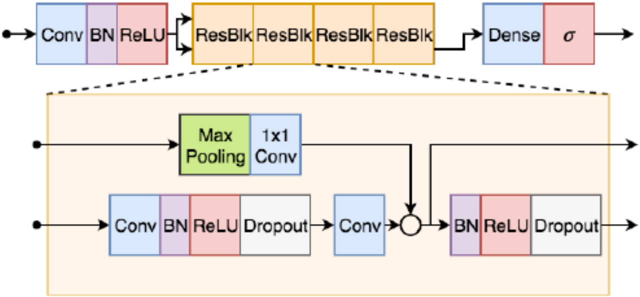
Background: Joint work with medical doctors from Brazil with an urgent need for automated analysis due to the **vast distances** between the patient and a cardiologist with full expertise in ECG diagnosis.

The existing telehealth network provides the data (more than 2 300 000 ECGs), implying some clinical relevance.

CODE = Clinical Outcomes in Digital Electrocardiology

Contribution: An end-to-end deep neural network that recognize 6 types of ECG abnormalities in standard 12-lead short-duration ECGs with a diagnostic performance that is at least as good as medical residents and students.

Network A CNN similar to the residual network.



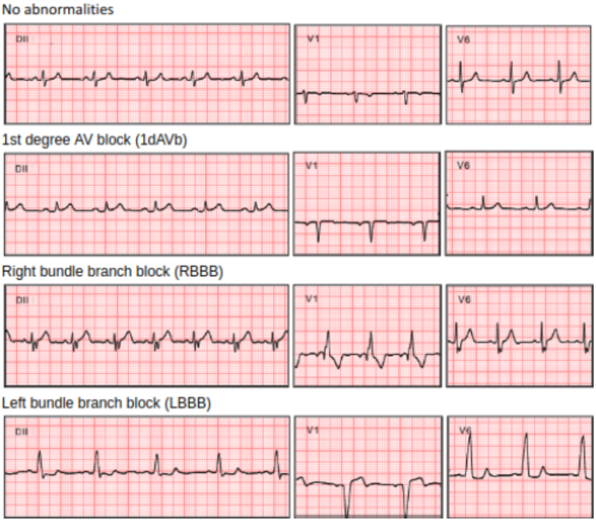
CODE – Data

Data The ECG recordings, which have between 7 and 10 seconds, where sampled in 300 Hz to 600 Hz. All recordings are re-sampled to a 400 Hz sampling rate and zero-padded resulting in a signal with 4096 samples for each lead.

Labels Labels of training and validation data was extracted from the cardiologist report is available only as a textual description.

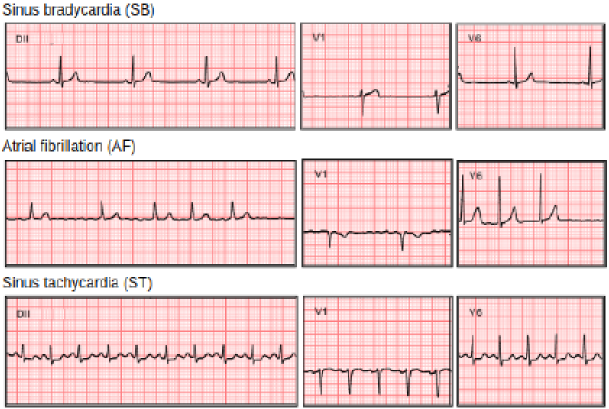
Test data was labeled was independently annotated by two certified cardiologists with experience in electrocardiography.

CODE –data



No abnormalities
DII
V1
V6
1st degree AV block (1dAVb)
DII
V1
V6
Right bundle branch block (RBBB)
DII
V1
V6
Left bundle branch block (LBBB)
DII
V1
V6

CODE –data


Sinus bradycardia (SB)
DII
V1
V6
Atrial fibrillation (AF)
V1
V6
Sinus tachycardia (ST)
DII
V1
V6

CODE – Some results

		predicted label							
		DNN		cardio.		emerg.		stud.	
	true label	not present	present	not present	present	not present	present	not present	present
1dAVb	not present	796	3	797	2	786	13	782	17
	present	3	25	9	19	5	23	2	26
RBBB	not present	788	5	788	5	792	1	790	3
	present	0	34	1	33	8	26	2	32
LBBB	not present	796	1	797	0	796	1	795	2
	present	0	30	3	27	4	26	3	27
SB	not present	808	3	808	3	808	3	807	4
	present	1	15	1	15	2	14	4	12
AF	not present	811	3	811	3	812	2	805	9
	present	1	12	3	10	5	8	1	12
ST	not present	787	4	790	1	788	3	787	4
	present	1	35	6	30	2	34	6	30

F1 Score			
DNN	cardio.	emerg.	stud.
0.893	0.776	0.719	0.732
0.932	0.917	0.852	0.928
0.984	0.947	0.912	0.915
0.882	**0.882**	0.848	0.750
0.857	0.769	0.696	0.706
0.933	0.896	0.932	0.857

$$F_1 = 2\frac{\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}}$$

CODE – Limitations and future work

This is a **proof-of-concept** study, which has **limitations:**

- We **cannot** (with statistical significance) show that our algorithm is better than the humans we compared against.
- We have not tested the algorithm on other classes of abnormalities.
- The real clinical setting is clearly much more complex...

Future work:

- Test the algorithm in a controlled real-life situation, showing that accurate diagnosis could be achieved in real-time. This can have big impact in improving healthcare in low and middle-income countries.

Ribeiro, A. H., Ribeiro, M. H., Paixao, G. M. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., Ferreira, M. P. S., Andersson, C. R., Macfarlane, P. W., Meira, W., Schön, T. B. and Ribeiro, A. L. P. **Automatic diagnosis of the short-duration 12-lead ECG using a deep neural network: the CODE study**. Submitted, 2019.

Open position – Well funded tenure track Assistant Professor

Well-funded research intensive tenure track Assistant Professorship on the topic of large scale optimization. Recruitment package:

1. 1 PhD student
2. 4 post-doc years
3. Generous budget for travels etc.

The position is made available via the Wallenberg AI, Autonomous Systems and Software Program (WASP) project.

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

More information is available here:
www.uu.se/en/about-uu/join-us/details/?positionId=306450

Feel free to spread this information if you want to!

Points to discuss

- A lot is happening in this sequence modeling at the moment. Finally, RNN/LSTM is getting competition. Is TCN and (its extensions) the best architecture for modeling sequential data?