Teaching and Research Area
Information Theory and
Systematic Design of
Communication Systems

RWTH AACHEN UNIVERSITY

SAMI2021

Incremental Parameter Estimation of Stochastic State-Based Models

Robert Lipp, Guido Dartmann, Lejla Fazlic, Thomas Vollmer, Stefan Winter, Arne Peine, Lukas Martin and Anke Schmeink

Content

Advantages through Machine Learning and AI Methods

- The healthcare sector is benefiting greatly from developments in AI and machine learning
- Whereas mathematical models previously had to be created by hand, nowadays they can be generated automatically on the basis of measured data
- These include models for ...
  - risk stratification,
  - therapy guidance,
  - classification of patients,
  - and many more.

Model Learning in the Healthcare Sector

Challenges

- Methods for automatic learning of mathematical models usually need a large training dataset
- Especially in the medical field, however, such large data sets are rare
- The reasons are:
  - Sensitive privacy of patient data
  - High data protection requirements
  - Specifically in the EU: The European General Data Protection Regulation
- Probably the best-known database in the medical field that is freely accessible to researchers is MIMIC-III
- In available datasets, data often do not have the desired quality or quantity, or cohorts for specific applications are too small

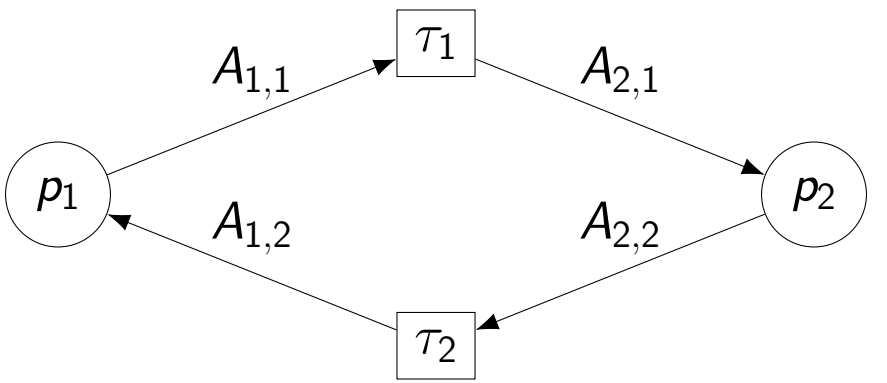
Proposed Solution

- In this work, we present an algorithm that learns model parameters incrementally
- In this way, the algorithm does not require a large training database
- The mathematical model is trained incrementally with small data sets
- After training with a data set, only the model parameters are passed on and the training data remain securely

Content

Stochastic Petri Nets

- Stochatic Petri Nets (SPNs) are very useful in systems biology and medicine
- The system is fully described by the firing rates of all transitions $\tau_1, \tau_2, \ldots$ and the incidence matrix A
- In this work we reconstruct the matrix A from given data



Content

Proposed Algorithm

Incremental Nature of the Learning Algorithm[1]

Proposed Algorithm

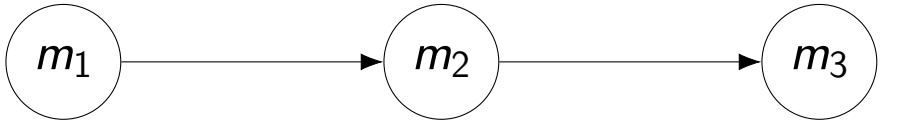
Incremental Nature of the Learning Algorithm[1]

Proposed Algorithm

Incremental Nature of the Learning Algorithm[1]

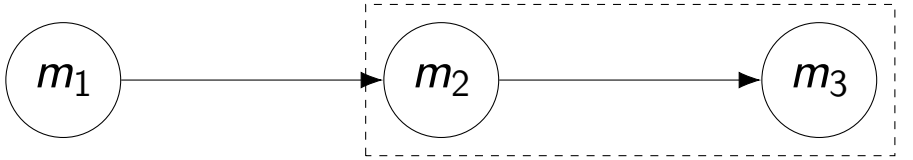
Update weights

Incremental Nature of the Learning Algorithm[1]



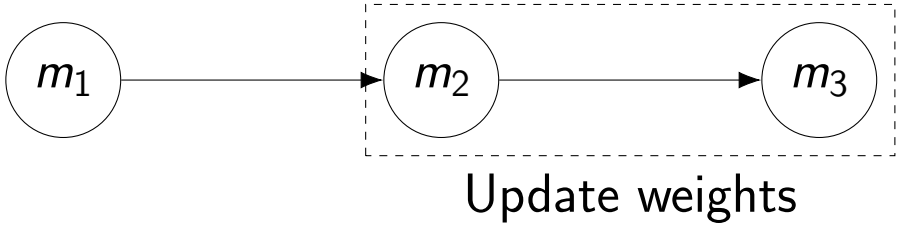
Proposed Algorithm

Incremental Nature of the Learning Algorithm[1]

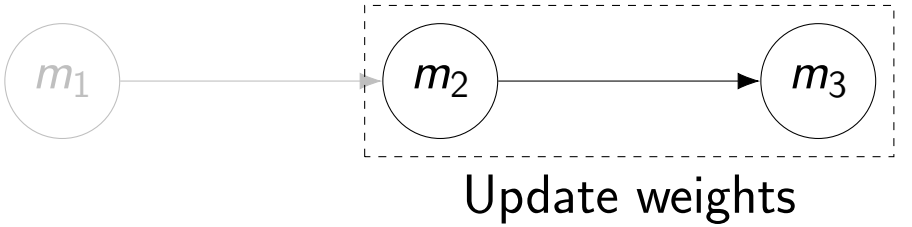


Proposed Algorithm

Incremental Nature of the Learning Algorithm[1]



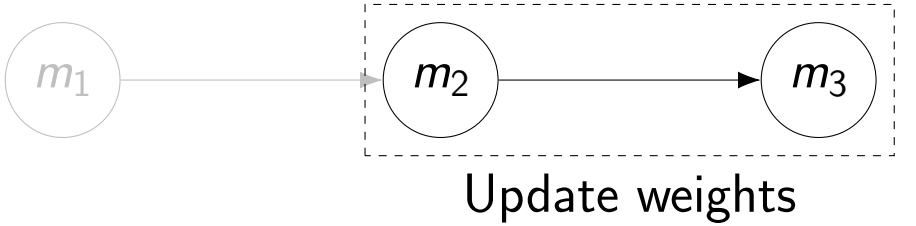
Incremental Nature of the Learning Algorithm[1]



[1] P. M. Vieting, R. C. de Lamare, L. Martin, G. Dartmann and A. Schmeink, "Likelihood-Based Adaptive Learning in Stochastic State-Based Models," in IEEE Signal Processing Letters, vol. 26, no. 7, pp. 1031-1035, July 2019, doi: 10.1109/LSP.2019.2917495.

Proposed Algorithm

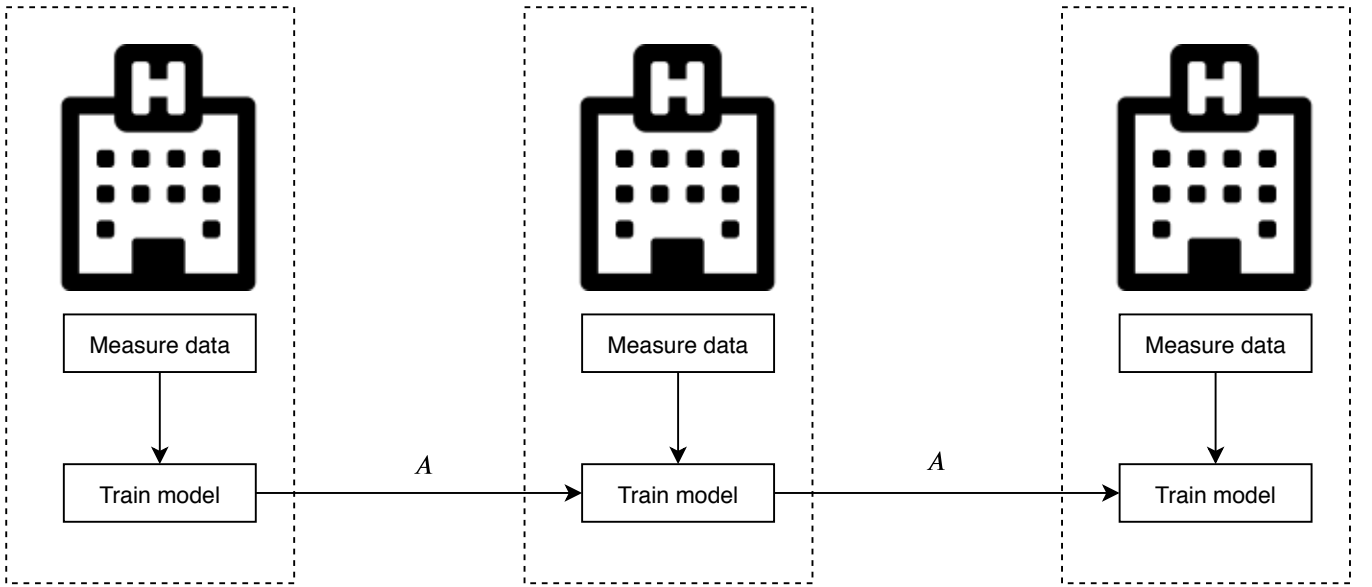
Incremental Nature of the Learning Algorithm[1]



- In each step only the pair (m_i, m_{i+1}) is required
- This incremental behavior is very useful

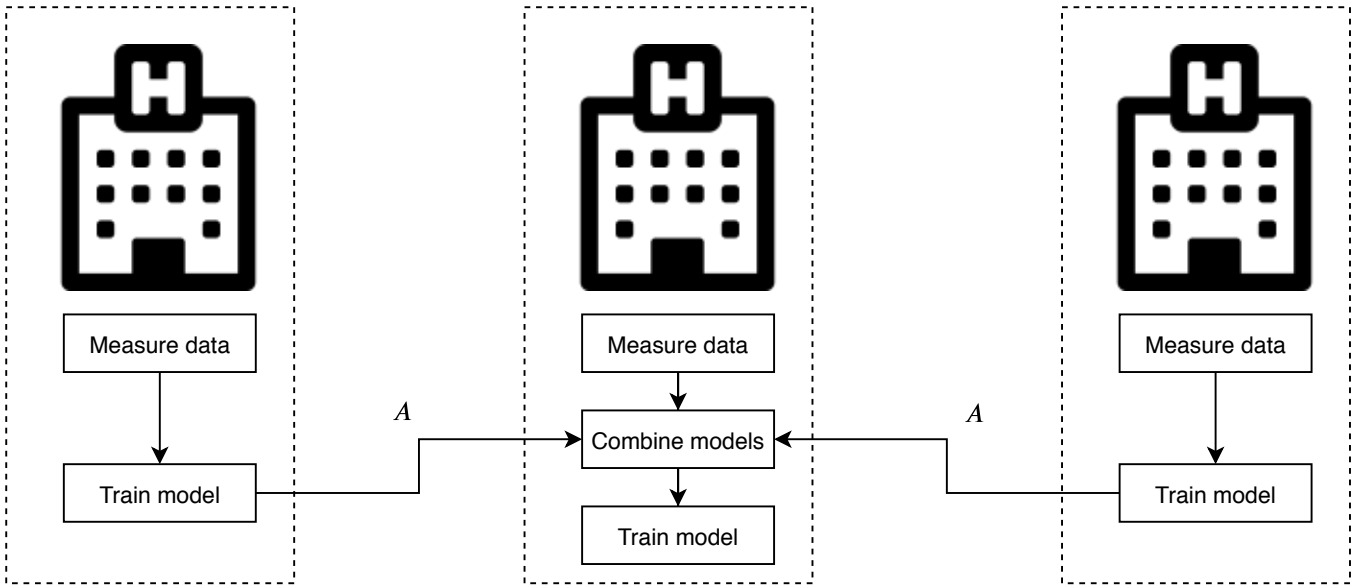
[1] P. M. Vieting, R. C. de Lamare, L. Martin, G. Dartmann and A. Schmeink, "Likelihood-Based Adaptive Learning in Stochastic State-Based Models," in IEEE Signal Processing Letters, vol. 26, no. 7, pp. 1031-1035, July 2019, doi: 10.1109/LSP.2019.2917495.

Sequential Approach



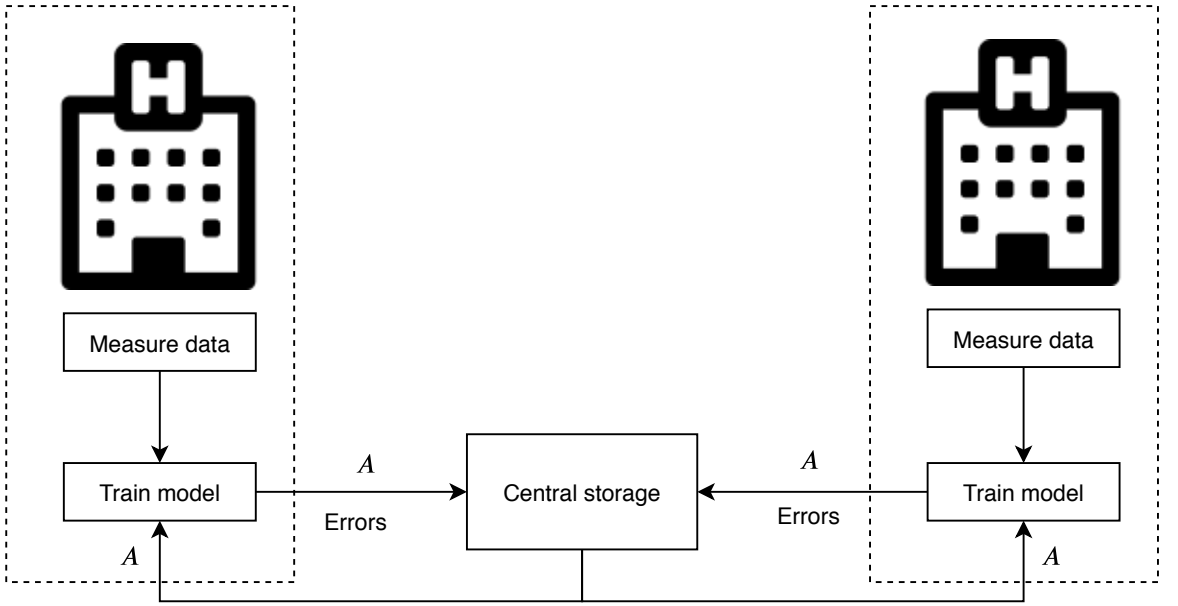
- Pros: Data protection, privacy, no need for public database
- Cons: From which hospital should the model be requested? - Danger of "double learning" - Model could get worse

Sequential Approach with Parallel Operations



- Pros: No need to decide for one model, hospitals can train in parallel
- Cons: No version control, it may happen that the model gets trained with the same data over and over again

Sequential Approach with Central Storage



- Pros: Version control, quality of model can be observed, no accidental double-learning
- Cons: ?

Overview

- The incremental algorithm allows the learning of a model based on data from different hospitals while respecting data protection and anonymity
- The central storage can perform version control
- Anomaly detection could identify bad quality data sources

Content

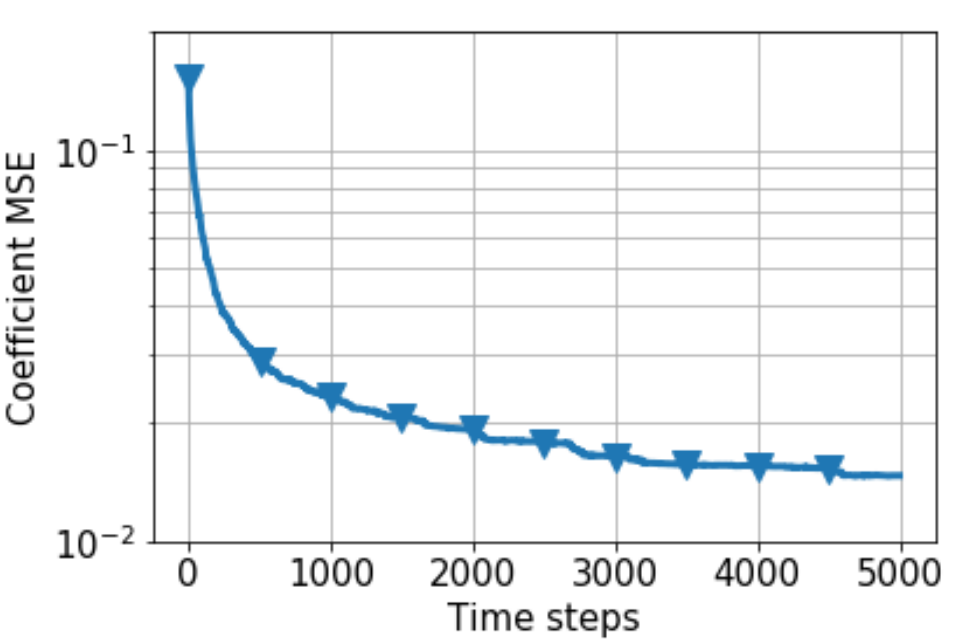
Simulation Results

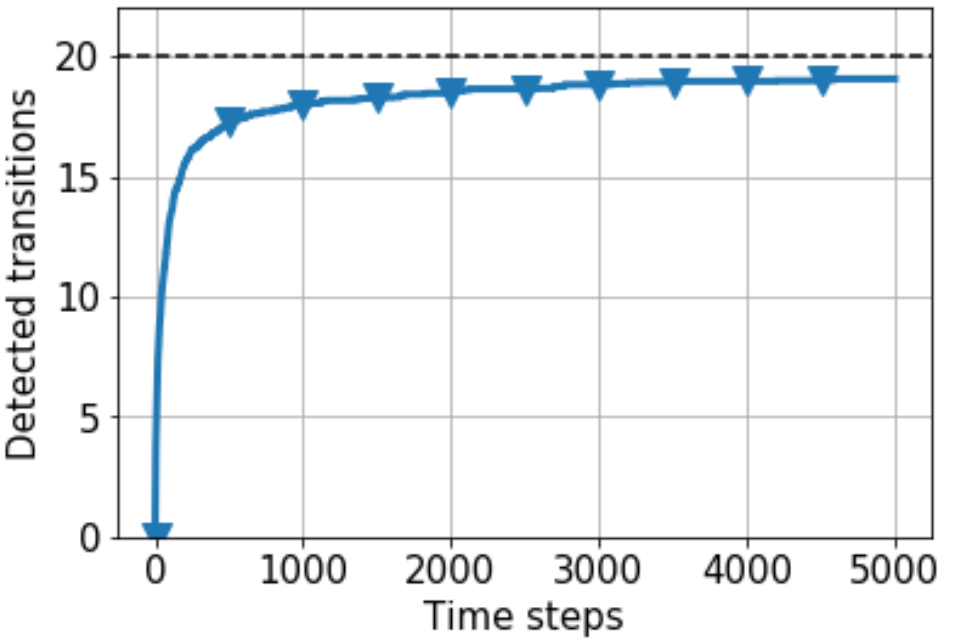
Overview

In the following we present simulation results for the simulation of a gene regulatory network in 3 cases:

1. A sufficient amount of 5000 data points is available
2. An insufficient amount of 250 data points is available
3. 20 batches of 250 data points each are available

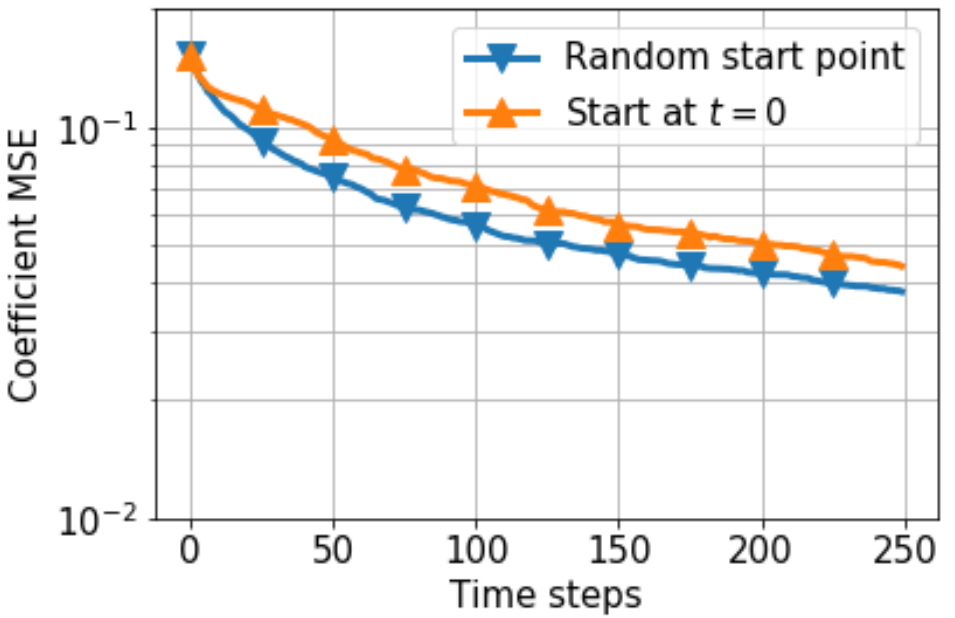
Case 1: Sufficient amount of data

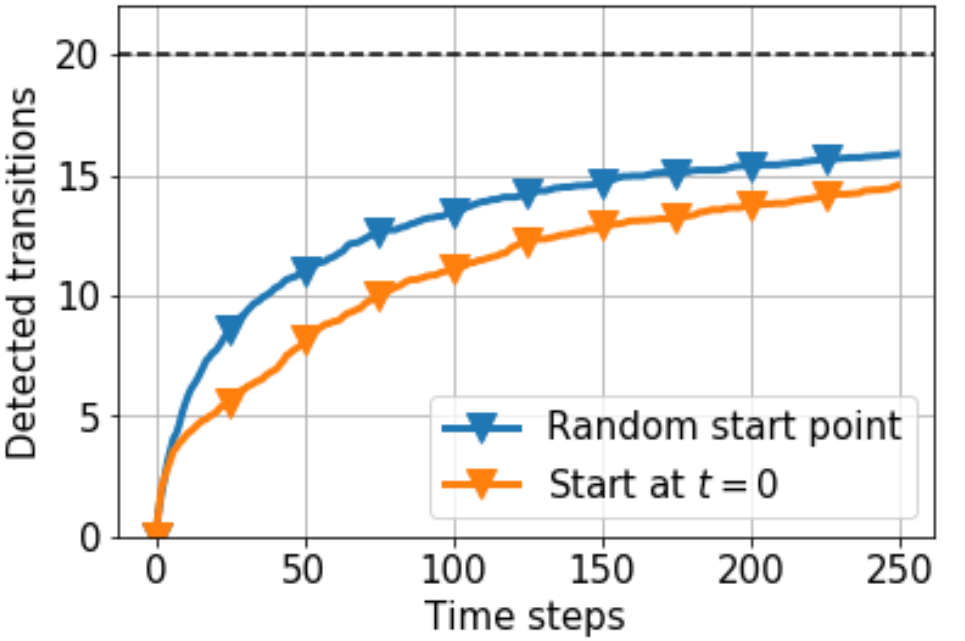




Training with a sufficient amount of 5000 data points

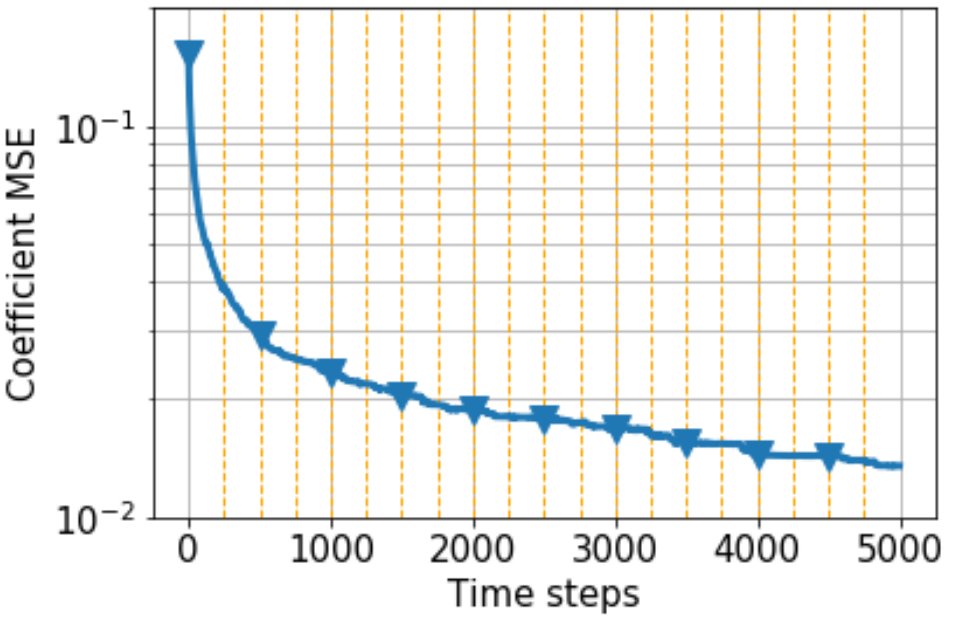
Case 2: Insufficient amount of data

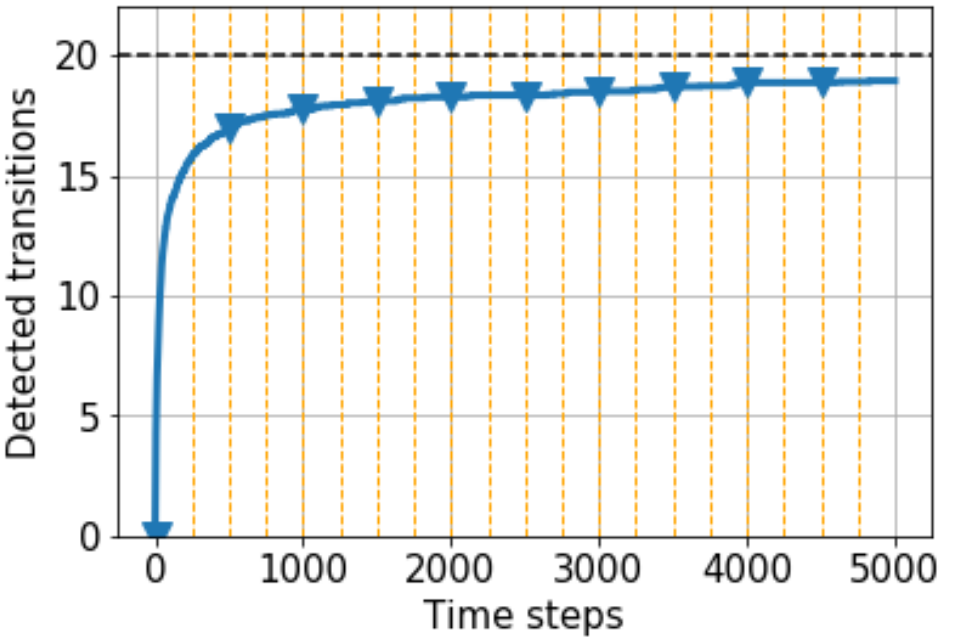




Training with an insufficient amount of 250 data points

Case 3: 20 batches of limited data





Incremental training with 20 batches of 250 data points each

Content

Conclusion

- In this work, we have presented a method to recover the structural parameters of SPNs in an incremental way from measured data
- Our main contributions:
  - We have extended the LB-DAAGD algorithm[2] and adapted it for incremental applications,
  - we have developed several designs of the algorithm,
  - and we have shown through simulations that the most general version of our incremental algorithm performs as well as the algorithm that has a large database available.
- The learning algorithm is no longer dependent on large publicly accessible data sets.
- Future research:
  - Evaluating the algorithm with patient data measured in hospitals
  - Estimate the kinetic parameters of SPNs incrementally based on measured data

[2] P. M. Vieting, R. C. de Lamare, L. Martin, G. Dartmann and A. Schmeink, "Likelihood-Based Adaptive Learning in Stochastic State-Based Models," in IEEE Signal Processing Letters, vol. 26, no. 7, pp. 1031-1035, July 2019, doi: 10.1109/LSP.2019.2917495.

Thank you
for your attention

Robert Lipp, Guido Dartmann, Lejla Fazlic, Thomas Vollmer, Stefan Winter, Arne Peine, Lukas Martin and Anke Schmeink