Parameter Learning Algorithms for Continuous Model Improvement Using Operational Data

Anders L. Madsen[1,2] Nicolaj Søndberg Jeppesen[1] Frank Jensen[1] Mohamed S. Sayed[3] Ulrich Moser[4] Luis Neto[5] Joao Reis[5] Niels Lohse[3]

HUGIN EXPERT A/S, Aalborg, Denmark

Department of Computer Science, Aalborg University, Aalborg, Denmark

Loughborough University, Loughborough, United Kingdom

IEF-Werner GmbH, Furtwangen, Germany

Instituto de Sistemas e. Robotica Associacao, Porto, Portugal

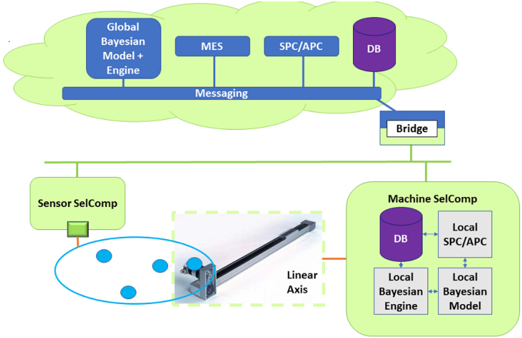
ECSQARU 2017

Outline

- Introduction
- Challenge
- Algorithms for parameter learning
- Experimental analysis
- Conclusion
- Acknowledgements

Introduction: Industry 4.0

The vision of SelSus is to create a new paradigm for highly effective, self-healing production resources and systems to maximise their performance over longer life times through highly targeted and timely repair, renovation and upgrading.



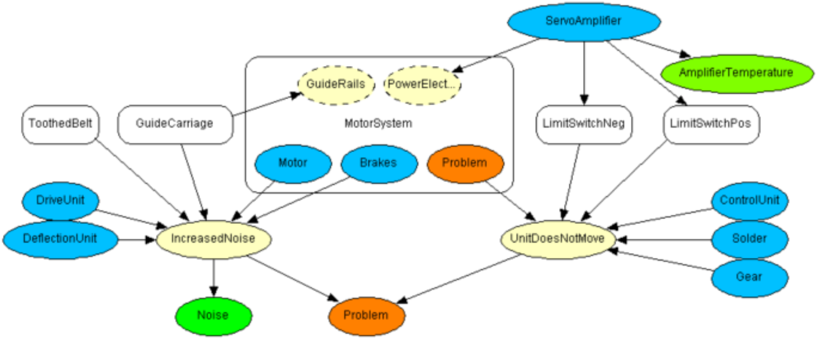
The SelSus System architecture

- Diagnosis at component level
- Data exchange over the cloud
- Control software and system wide integration
- Model improvement from operational data

Industry 4.0 (smart factory, IoT, cloud computing) is a trend of automation and data exchange in manufacturing technologies (Wikipedia)

Challenge: Improving Linear Axis OOBN

An OOBN model for RCA constructed from expert knowledge



Network Statistics

- 35 variables, 27 failure states, and five class instances
- 555 CPD entries and maximum CPD size of 128

Algorithms for Parameter Learning I

A BN $\mathcal{N} = (\mathcal{X}, G, \mathcal{P})$ with $P(\mathcal{X}) = P(X_1, \ldots, X_n) = \prod_{X_i \in \mathcal{X}} P(X_i | \pi_{X_i})$

Estimate the values of $\mathcal{P}$ from data $\mathcal{D} = \{c_1, \ldots, c_N\}$ and expert knowledge

Batch EM (4; 2)

Iterates an E and M step, processing all cases $\mathcal{D}$ in each iteration

$$\theta^*_{ijk} = \frac{n(X_i = k, \pi_{ij}) + \theta_{ijk}\alpha_{ij}}{n(\pi_{ij}) + \alpha_{ij}}, \tag{1}$$

α_{ij} is the experience count for parent configuration π_{ij}

Incremental EM (6)

Iterates EM over $\mathcal{D} = \mathcal{D}_1, \ldots, \mathcal{D}_m$. The estimates θ_{ijk} and α_{ij} produced by one iteration of EM are used as virtual counts in the next iteration of EM

Algorithms for Parameter Learning II

Online EM (1)

Updates parameters after each case

$$\theta^*_{ijk} = (1-\gamma)m_{ijk} + \gamma P(x_{ijk}|c), \tag{2}$$

learning rate $\gamma = (1+n)^{-\rho}$ where $0.5 < \rho < 1$

Fractional Updating (5; 7)

Updates parameters after each case

$$\theta^*_{ijk} = \frac{\theta_{ijk}\alpha_{ij} + P(x_{ijk}|c)}{\alpha_{ij} + P(\pi_{ij}|c)}. \tag{3}$$

OOBNs: compute the average expected counts for the run-time instances of the node and increase the experience counts by the number of run-time instances.

Experimental Setup

In Industry 4.0, we add diagnostic capabilities at the component level

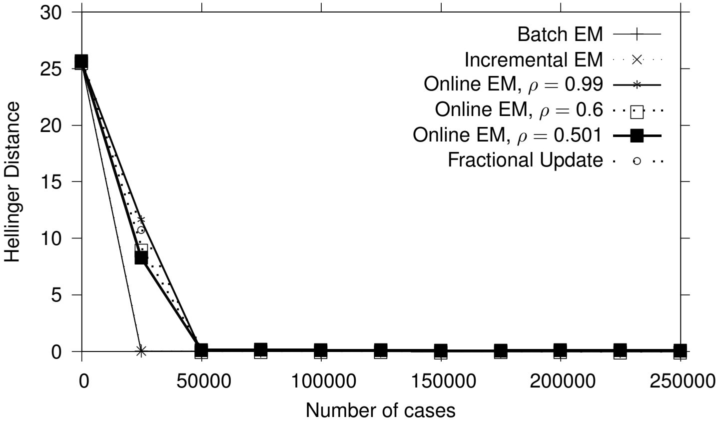
1. Is model improvement feasible using parameter learning algorithms?
2. Can we improve model performance using operational data?
3. What is the time cost in a real setting?

The Knowledge Driven Model is the starting point for parameter learning. Three data sets used in the analysis

- Random sample $\mathcal{D}_0$ with $N_0 = 250,000$ 5% missing values (Knowledge Driven Model)
- Operational data $\mathcal{D}_1$ with $N_1 = 13,429$ (Known)
- Operational data $\mathcal{D}_2$ with $N_2 = 25,726$ (Unknown)

I: Learning from Random Sample

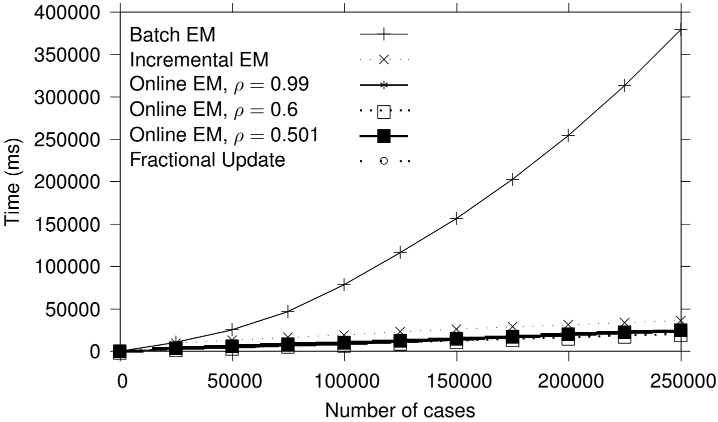
Purpose: Determine if learning is feasible
Data: A random sample $\mathcal{D}$ with $N = 250,000$
Model: Uniform Knowledge Driven Model



Hellinger Distance (D_H^w): $D_H(P, Q) = \sqrt{\sum_i (\sqrt{p_i} - \sqrt{q_i})^2}$ (6) who cites (3)

I: Learning from Random Sample

Purpose: Determine if learning is feasible
Data: A random sample $\mathcal{D}$ with $N = 250,000$
Model: Uniform Knowledge Driven Model



Accumulated time in milliseconds (ms)

II: Diagnostic Performance

Purpose: Identify *true* root cause from *randomly* generated evidence
Data: Operational data $\mathcal{D}_1$ with $N_1 = 13,429$
Model: Knowledge Driven Model

Algorithm	Top-1	Top-5	μ_{rank}
Knowledge Driven Model	8	17	4.6
Batch EM	10	17	5.1
Online EM	9	17	4.5
Fractional update	10	21	3.4

For rank a lower number is better. A total of 27 failure states

Learning in Knowledge Driven Model using $\rho = 0.99$, $\sum_j \alpha_{ij} = 13,429$ for all i and $n_0 = 13,429$ are used

III: Average Time Performance (ms)

Purpose: Determine cost of processing
Data: Operational data $\mathcal{D}_2$ with $N_1 = 25,726$
Model: Knowledge Driven Model

Algorithm	Configuration	cases/request	Total time (ms)	Avg. time
Online EM	Direct integration	1	1,730	0.067
	SelSus Cloud	1000	11,367	0.44
	SelSus Cloud	100	44,867	1.74
	SelSus Cloud	10	496,199	19.29
Fractional Updating	Direct integration	1	1,533	0.067
	SelSus Cloud	1000	10,553	0.41
	SelSus Cloud	100	42,111	1.64
	SelSus Cloud	10	478,612	18.60

Average time cost of handling one case across integration levels

Conclusion

Considered four parameter learning algorithms for continuous model (OOBN) improvement using operational data

Experiments on Knowledge Driven Model with data

1. Learning is feasible
2. Improve diagnostic performance
3. Time cost is manageable

Learning from operational data appears to be a promising approach even in the light of many hidden variable

Approach is being deployed at factory level in Italy and United Kingdom

Acknowledgements

SelSus

This work is part of the project Health Monitoring and Life-Long Capability Management for SELf-SUStaining Manufacturing Systems (SelSus) which is funded by the Commission of the European Communities under the 7th Framework Programme, Grant agreement no: 609382.

References

[1] O. Cappe and E. Moulines.
Online EM Algorithm for Latent Data Models.
J. Royal Statistical Society Series B (Statistical Methodology), 71(3):593–613, 2009.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin.
Maximum likelihood from incomplete data via the EM algorithm.
J. Royal Statistical Society - Series B, B39(1):1–38, 1977.

[3] G. Kokolakis and P. Nanopoulos.
Bayesian multivariate micro-aggregation under the Hellingers distance criterion.
Research in Official Statistics, 4(1):117–126, 2001.

[4] S. L. Lauritzen.
The EM Algorithm for Graphical Association Models with Missing Data.
Computational Statistics & Analysis, 19:191–201, 1995.

[5] K. G. Olesen, S. L. Lauritzen, and F. V. Jensen.
aHUGIN: A System Creating Adaptive Causal Probabilistic Networks.
In *Proc. UAI*, pages 223–229, 1992.

[6] P. Ratnapinda and M.J. Druzdzel.
Learning discrete Bayesian network parameters from continuous data streams: What is the best strategy?
J. Applied Logic, 13:628–642, 2015.

[7] D. M. Titterington.
Updating a diagnostic system using unconfirmed cases.
Applied Statistics, 25:238–47, 1976.