Integrated Dimensionality Reduction and Sequence Prediction using LSTM
Okafor, Emmanuel; Schomaker, Lambertus

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Integrated Dimensionality Reduction and Sequence Prediction using LSTM

Poster · March 2018
DOI: 10.13140/RG.2.2.28577.30563

2 authors, including:

Lambert Schomaker
University of Groningen

Some of the authors of this publication are also working on these related projects:

- Making Sense of Illustrated Handwritten Archives View project
- MPS - Medieval Paleographic Scale View project
### Problem
- Most industrial or complex processes present temporal dependencies which stretch over a long time.
- The underlying patterns in these processes can be extremely non-linear.
- Conversion of nominal codes (raw codes) to other vectorial paradigms is not suitable.
- Hidden Markov Model has prediction limitation when dealing with temporal dependencies that stretch over long durations.

### Objectives
- Use of external and a proposed integrated dimensionality reduction LSTM predictive systems for predicting message logs from industrial machines.
- Conversion of nominal codes (raw codes) to other vectorial paradigms to obtain better correlated patterns.

### Methods
- **External Methods:** Recurrent Neural Networks (RNN) [3-7]
- **Proposed Method:** Integrated Dimensionality-reduction LSTM

### Results
- **ID-LSTM Prediction on OHE codes during training and testing phases (left plot) and index predictions (right plot) over a duration of 10K time-counts.**

### Conclusion
- We have transformed nominal codes to other vectorial representations with the objective of identifying correlated patterns using one hot encoding (OHE) and principal component analysis (PCA).
- Nominal integer codes are not sensible to use in the RNN.
- A separate dimensionality reduction by PCA is not needed: the ID-LSTM produces 10 hidden dimensions in the bottleneck layer.
- The ID-LSTM on OHE codes yield the best result on a small sample dataset.
- The use of ID-LSTM also obtains good results on reduced dimensional PCA vector codes (20-DIM-PCA).
- The ID-LSTM obtained < 5% error on the predicted OHE codes in a realistically large dataset.
- One-hot-encoding is a must: do not try to predict arbitrary raw integer codes.

### Future Directions
- We suggest that it may be possible to combine the proposed model with an early anomaly detection algorithm.
- To allow continuous prediction of physical problems in the machines generating the message logs.
- Optimization of LSTM-based feature dimensionality reduction in a realistically large dataset.

### References