

Integrated Dimensionality Reduction and Sequence Prediction using LSTM

Emmanuel Okafor and Lambert Schomaker

Institute of Artificial Intelligence and Cognitive Engineering, University of Groningen, The Netherlands



Problem

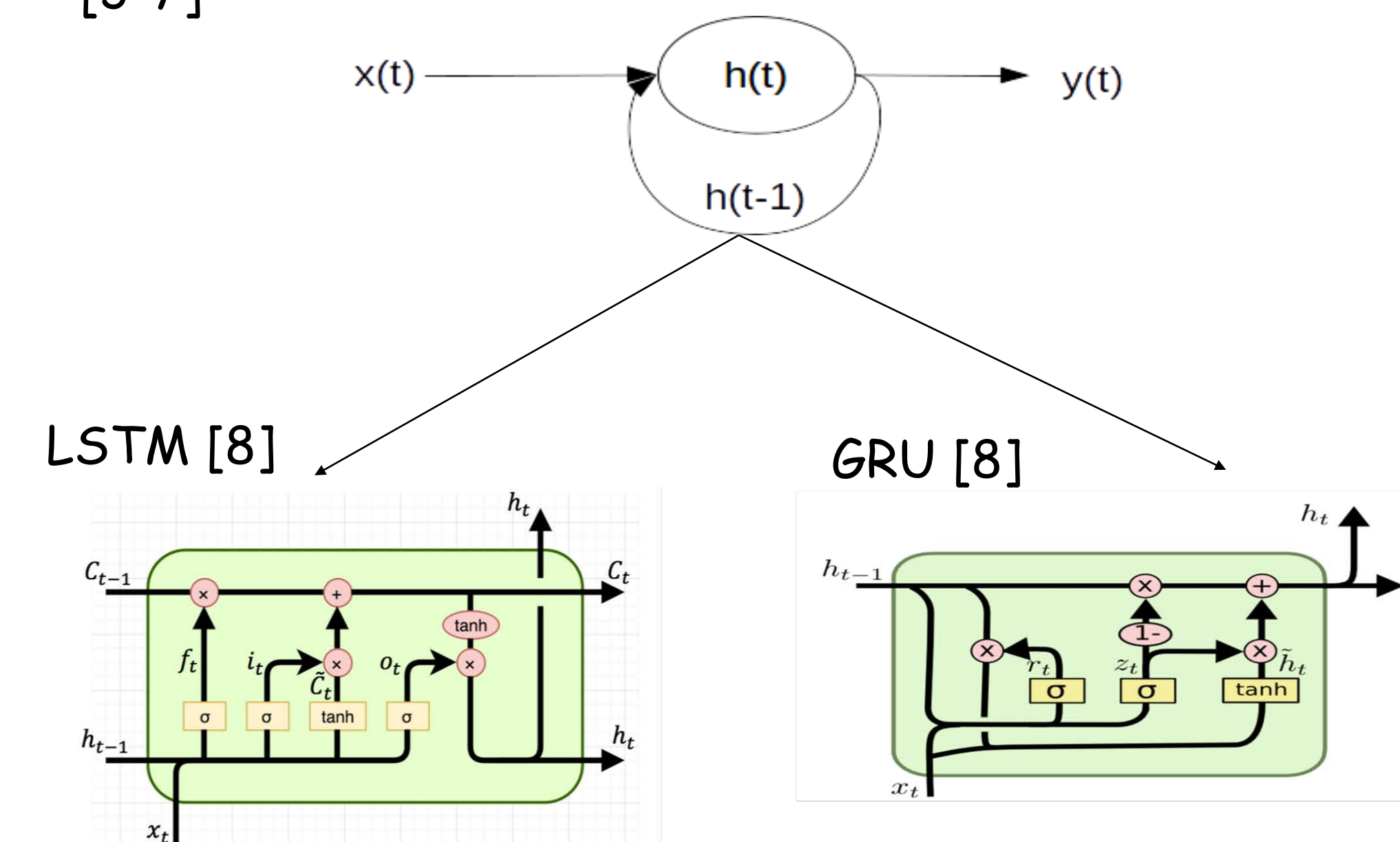
- Most industrial or complex processes present temporal dependencies which stretch over a long time.
- The underlying patterns in these processes can be extremely non-linear.
- Use of linear predictive model (ARMA/ARIMA[1]) is not suitable.
- Hidden Markov Model[2] has prediction limitation when dealing with temporal dependencies that stretch over long durations.

Objectives

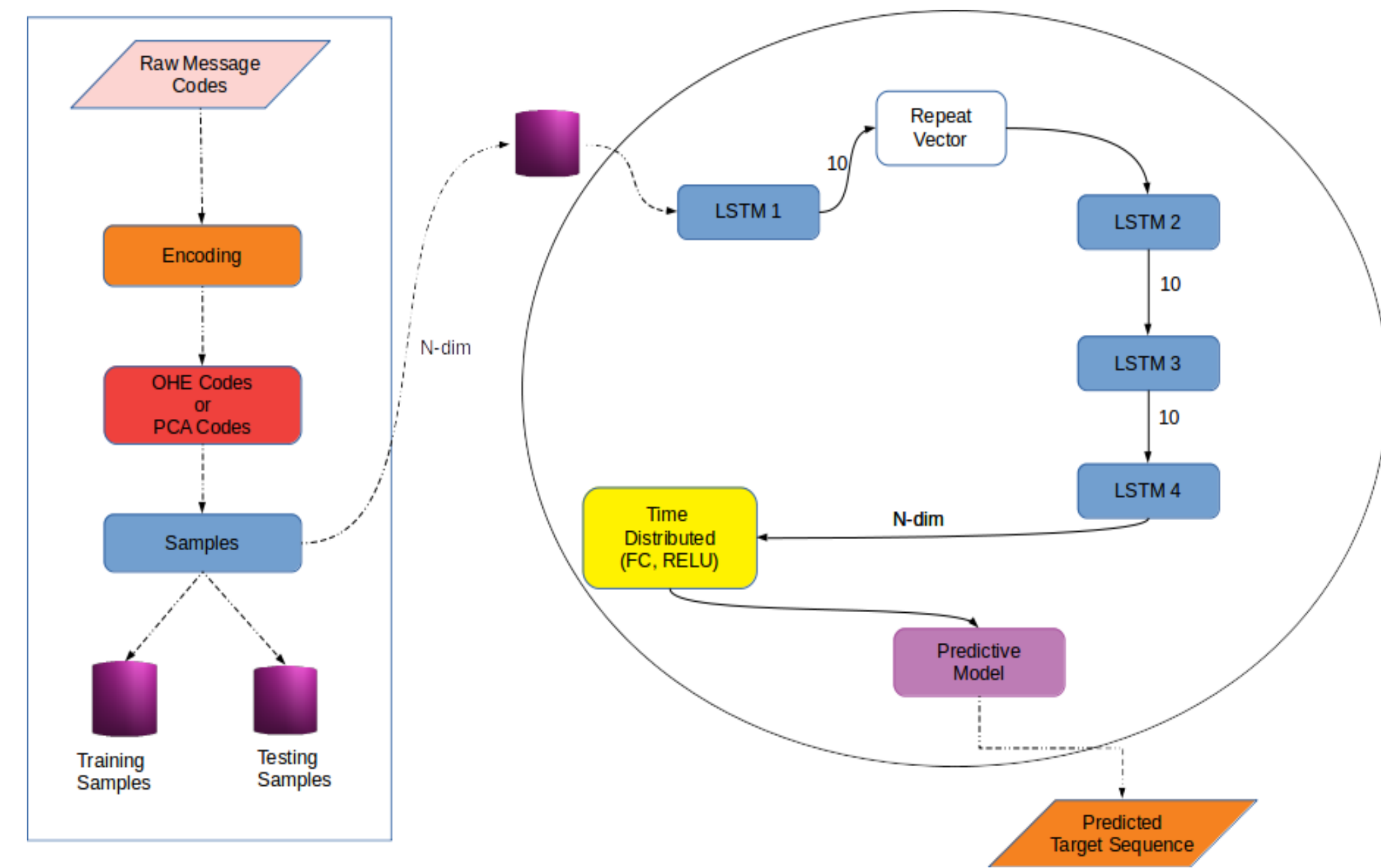
- Use of external and a proposed integrated dimensionality reduction LSTM predictive systems for predicting message logs from industrial machines.
- Conversion of nominal codes (raw codes) to other vectorial paradigms to obtain better correlated patterns.

Methods

- External Methods: Recurrent Neural Networks (RNN) [3-7]



- Proposed Method: Integrated Dimensionality-reduction LSTM

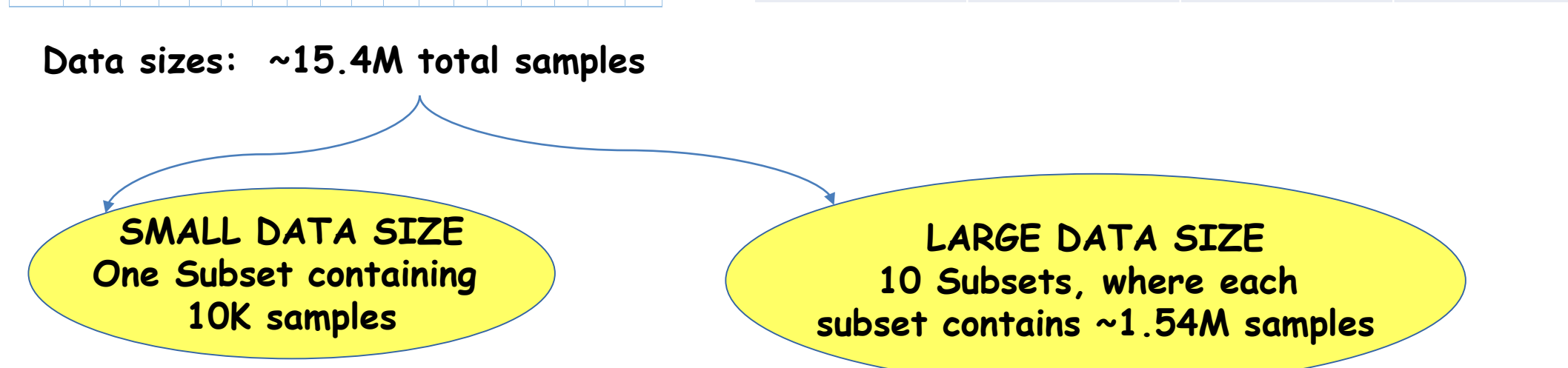


- Encoding Section: One LSTM
- Decoding Section: Three LSTMs
- Repeat Vector: Interlinks the encoding and decoding components
- Time Distributed: the final feature dimension from the last LSTM is wrapped with a time-distributed algorithm that presents the reproduced data in a sequential series.

Data Representations

One-Hot-Encoding Codes													
0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	1
2	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0	0	0	0	0	0
...
9992	0	0	0	0	0	0	0	0	0	0	0	0	0
9993	0	0	0	0	0	0	0	0	0	0	0	0	0
9994	0	0	0	0	0	0	0	0	0	0	0	0	0
9995	0	0	0	0	0	0	0	0	0	0	0	0	0
9996	0	0	0	0	0	0	0	0	0	0	0	0	0
9997	0	0	0	0	0	0	0	0	0	0	0	0	0
9998	0	0	0	0	0	0	0	0	0	0	0	0	0
9999	0	0	0	0	0	0	0	0	0	0	0	0	0

3-DIM Principal Component Analysis (PCA) Codes				
	PC 1	PC 2	PC 3	
0	0.8211	-0.1157	-0.0232	
1	0.8211	-0.1157	-0.0232	
...
9997	0.1326	0.4218	0.4549	
9998	0.1326	0.4218	0.4549	
9999	0.1326	0.4218	0.4549	



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.

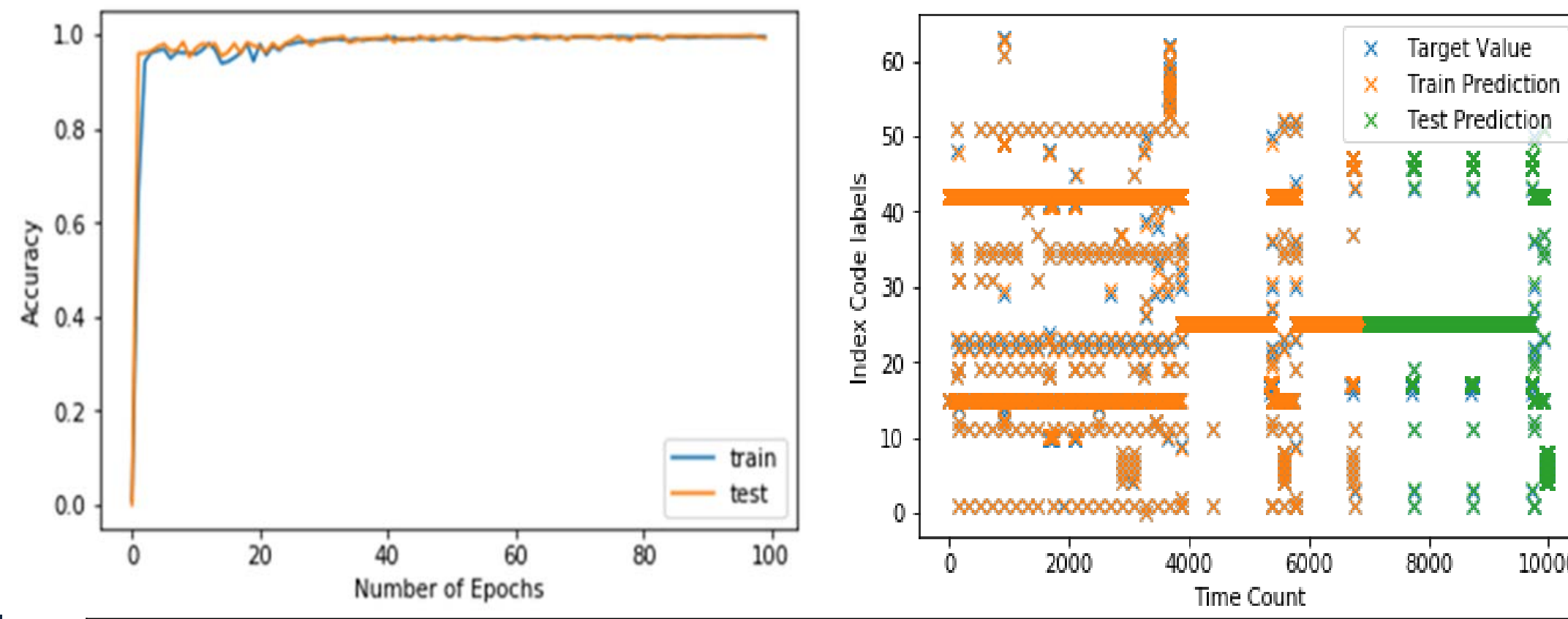
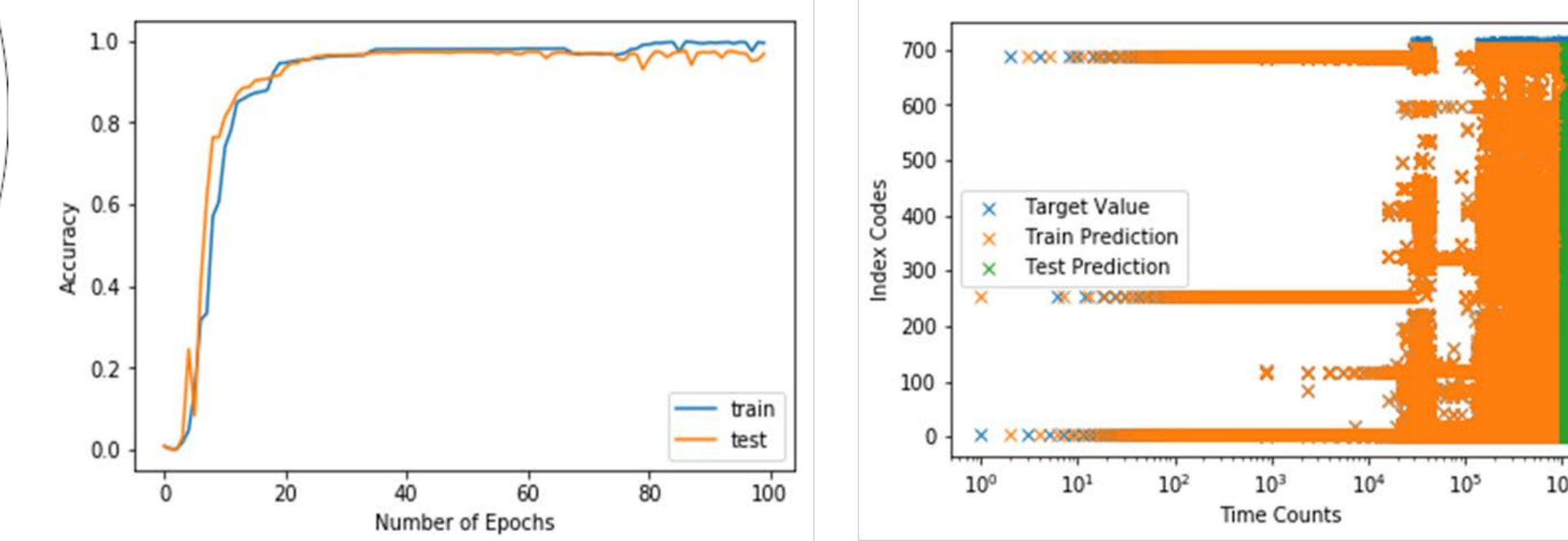


Table 1: Prediction accuracies for the different approaches for 10K Samples

Methods	Train	Test
ID-LSTM-I-OHE-Codes	0.9957	0.9920
ID-LSTM-I-20-DIM-PCA-Codes	0.9763	0.9843
ID-LSTM-I-40-DIM-PCA-Codes	0.9760	0.9733
ID-LSTM-I-10-DIM-PCA-Codes	0.9316	0.9727
ID-LSTM-I-5-DIM-PCA-Codes	0.9139	0.9593
ID-LSTM-I-4-DIM-PCA-Codes	0.9424	0.9410
ID-LSTM-I-3-DIM-PCA-Codes	0.9463	0.9593
ID-LSTM-I-2-DIM-PCA-Codes	0.9424	0.9590
ID-LSTM-I-1-DIM-PCA-Codes	0.8729	0.9340
SL-LSTM-I-1-DIM-PCA-Codes	0.8757	0.9340
SL-GRU-MSE-SI-1-DIM-PCA-Codes	0.8715	0.9316
SL-GRU-MAE-SI-1-DIM-PCA-Codes	0.8715	0.9316
SL-LSTM-MAE-SI-1-DIM-PCA-Codes	0.8715	0.9316
SL-LSTM-MSE-SI-1-DIM-PCA-Codes	0.8715	0.9316
SL-LSTM-I-Raw-Codes	1.429x10-4	0.0000
SL-GRU-MSE-SI-Raw-Codes	2.858x10-4	0.0000
SL-GRU-MAE-SI-Raw-Codes	2.858x10-4	0.0000
SL-LSTM-MSE-SI-Raw-Codes	2.858x10-4	0.0000
SL-LSTM-MAE-SI-Raw-Codes	1.429x10-4	0.0000

- NOTE**
- A separate dimensionality reduction by PCA is not needed: the ID-LSTM uses 10 hidden dimensions in the bottleneck layer.
 - One-hot-encoding is a must: do not try to predict arbitrary raw integer codes

- ID-LSTM Prediction on OHE codes during training and testing phases (left plot) and index predictions (right plot) over a duration of ~1.54M time-counts for subset 9.



- The left and right plots show the confusion matrix, that is; the plot of the output predictions against their target values for both training and testing phases respectively for subset 9.

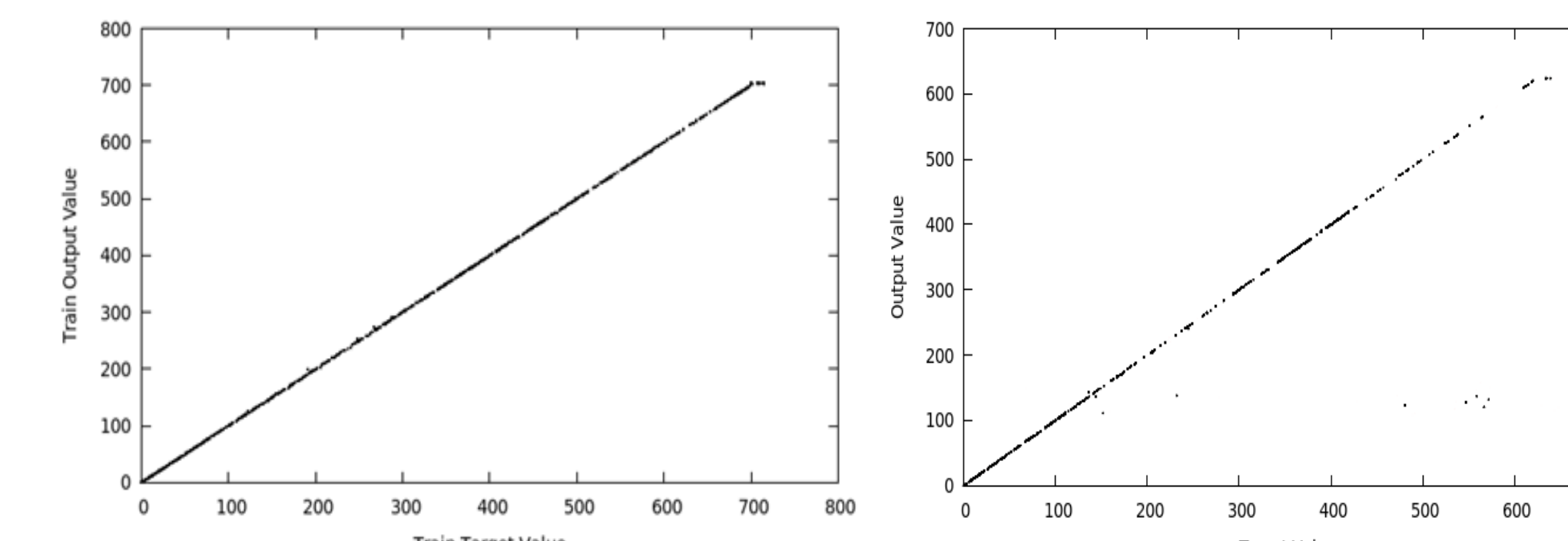


Table 2: Prediction accuracy of the ID-LSTM trained on OHE codes

No of Subsets	Time counts	No. of Index	No. of Machine	Train	Test
Subset 1	0 - 1.54M	948	20	0.9826	0.9751
Subset 2	1.54- 3.09M	606	30	0.9979	0.9695
Subset 3	3.09-4.63M	535	36	0.9886	0.9624
Subset 4	4.63-6.18M	619	48	0.9961	0.9021
Subset 5	6.18-7.73M	620	62	0.9837	0.9806
Subset 6	7.73-9.27M	675	109	0.9962	0.9347
Subset 7	9.27-10.8M	648	64	0.9205	0.9293
Subset 8	10.8-12.3M	679	95	0.9973	0.9576
Subset 9	12.3-13.9M	717	196	0.9943	0.9681
Subset 10	13.9-15.4M	624	263	0.9871	0.9268
Average				0.9844	0.9506

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 uses 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

[1] G. E. Box and G. M. Jenkins, "Time series analysis, control, and forecasting," San Francisco, CA: Holden Day, vol. 3226, no. 3228, p. 10, 1976.

[2] Z. Ghahramani, "An introduction to hidden markov models and Bayesian networks," International journal of pattern recognition and artificial intelligence, vol. 15, no. 01, pp. 9-42, 2001.

[3] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in International Conference on Machine Learning, 2014, pp. 1764-1772.

[4] F. A. Gers, D. Eck, and J. Schmidhuber, "Applying lstm to time series predictable through time-window approaches," in Neural Nets WIRN Vietri-01. Springer, 2002, pp. 193-200.

[5] N. Srivastava, E. Mansimov, and R. Salakhudinov, "Unsupervised learning of video representations using lstms," in International conference on machine learning, 2015, pp. 843-852.

[6] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in Advances in neural information processing systems, 2014, pp. 3104-3112.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[8] <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

This research was supported by the MANTIS

