Comparison and extension of autoencoder models for uni- and multivariate signal compression in IIoT

Julia Rosenberger[a], Alexander Kübel[b], Fabian Rothfuß[b]

[a]Bosch Rexroth AG, Bürgermeister-Dr.-Nebel-Straße 15, Lohr am Main 97816, Germany

[b]University of Applied Science, Lohmühlenstraße 65, Berlin 12435, Germany

Motivation

- Factory of the future and Industrial Internt of Things (IIoT): increasing amount of data sources (i.e. sensors)
  - ➢ Increasing amount of data
  - ➢ Increasing value of data due to new data based business models

↓

Data compression becomes more important

- Focus on sensor signals:
  - ➢ I.e. streaming data, times series data
  - ➢ Data contains insights about the condition of machines, e.g. oil pressure. Relevant for condition monitoring and predictive maintenance
  - ➢ Demand for low latency in data analysis

Requirements

Goal: Compression of streaming data in an industrial environment on industrial edge devices

The following requirements must be met:

1. Both the model’s memory and CPU usage has to remain low.
2. A high compression ratio (CR) and small reconstruction error (RE) has to be ensured.
3. The compression must applicable to streaming data and therefore in real-time.
4. The compression method should be agnostic to different sensor signals.
5. The method should be able to handle both univariate time series (UTS), i.e. dimension d = 1, as well as multivariate time series (MTS), i.e. d > 1.

Contribution

1. Selection and further development of two promising, existing models considering the mentioned requirements
   a. Convolutional Believe Network - Variational Autoencoder (CBN-VAE) Model according to [1]
   b. Recurrent Neural Network, i.e. Long-Short-Term-Memory, Autoencoder (LSTM-AE) Model according to [2]
2. A comparison of both models and a traditional method based on empiric evaluation results is given.
3. The applicability to IIoT devices while maintaining a reasonable reconstruction error with high compression ratio is proven and rated.

[1] J. Liu, F. Chen, J. Yan, and D. Wang, “CBN-VAE: A Data Compression Model with Efficient Convolutional Structure for Wireless Sensor Networks,” *Sensors*, vol. 19, no. 16, 2019.
[2] T. Wong and Z. Luo, “Recurrent Auto-Encoder Model for Large-Scale Industrial Sensor Signal Analysis,” in *Engineering Applications of Neural Networks*, E. Pimenidis and C. Jayne, Eds., Cham: Springer International Publishing, 2018, pp. 203–216.

Implementation: CBN-VAE

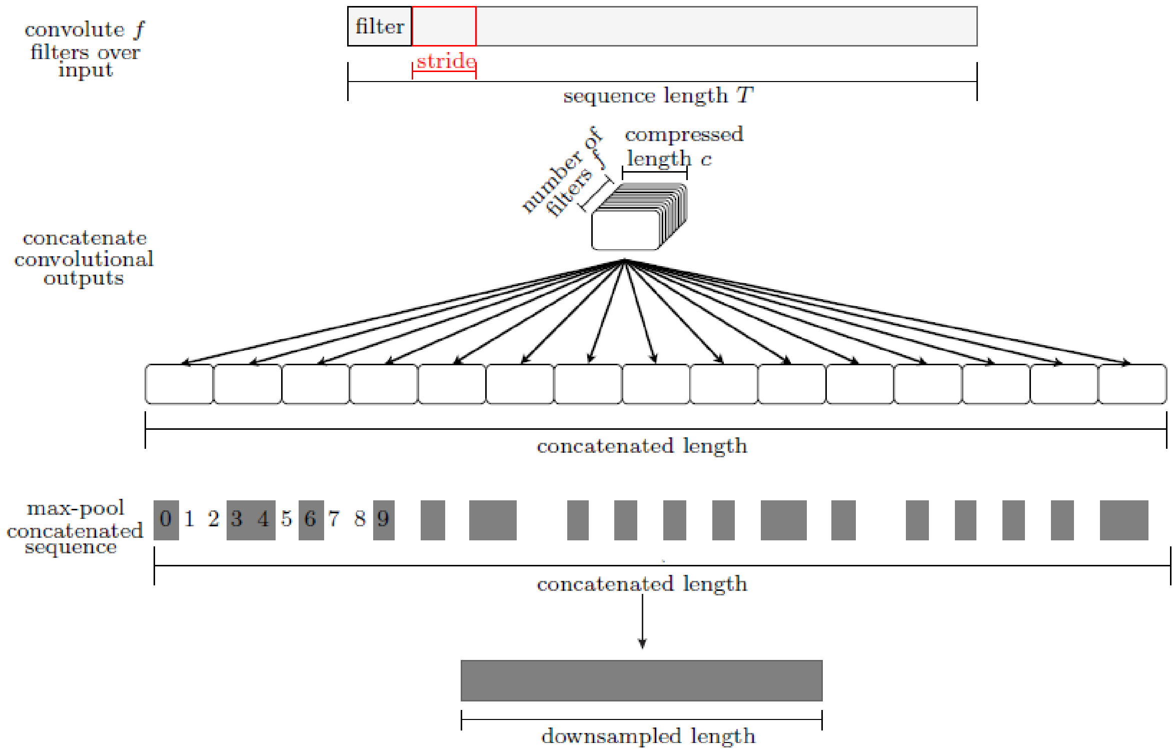

Figure 1: Procedure of downsampling using D-CRBM.

Schematic overview on the downsampling procedure of the CBN-VAE model according to [1]

[1] J. Liu, F. Chen, J. Yan, and D. Wang, “CBN-VAE: A Data Compression Model with Efficient Convolutional Structure for Wireless Sensor Networks,” *Sensors*, vol. 19, no. 16, 2019.

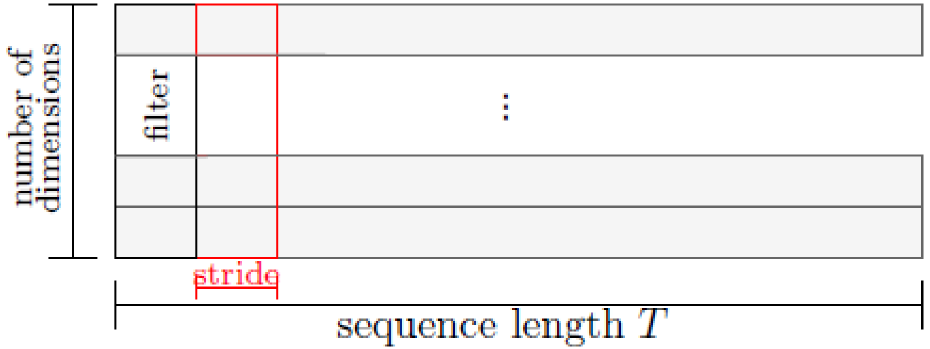

Figure 2: Convolutional filter for compression of multivariate input.

Implementation LSTM-AE

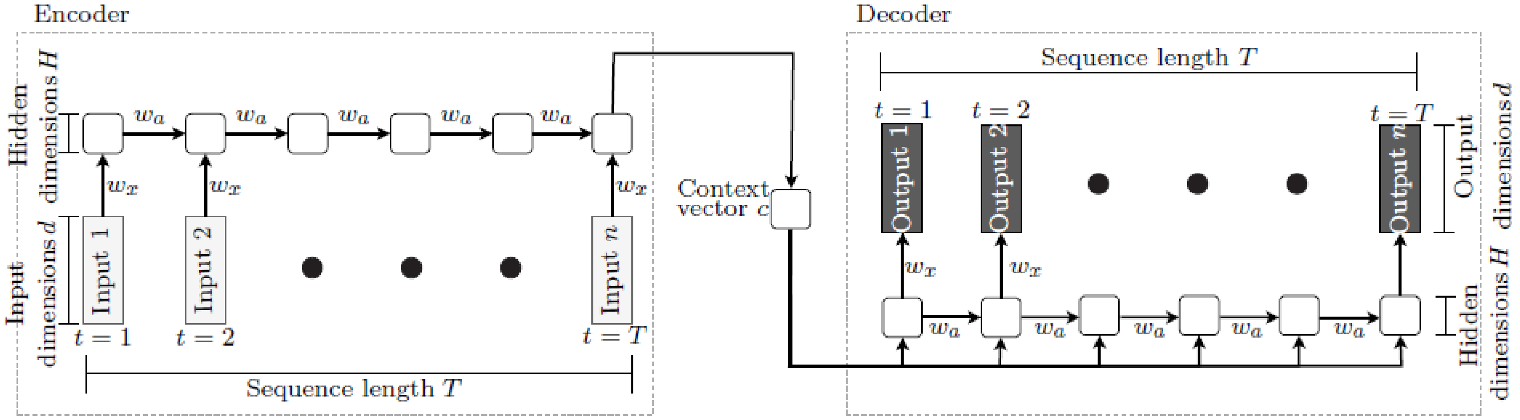

Figure 3: Schematic structure of the RNN-AE compression model [2]

[2] T.Wong and Z. Luo, “Recurrent Auto-Encoder Model for Large-Scale Industrial Sensor Signal Analysis,” in *Engineering Applications of Neural Networks*, E. Pimenidis and C. Jayne, Eds., Cham: Springer International Publishing, 2018, pp. 203–216.

Experimental setup

Hardware specifications:

- CPU: 4x Intel Xeon Platinum 8168 "Skylake" (2.7GHz);
- 8GB RAM;
- OS: Ubuntu 20.04.

- Implemented in Python 3.7.1 (Tensorflow 2.5.1)
- Model training + testing: Results averaged over 10x on random seeds
- Each model is trained and tested on two publicly available data sets:
  - Intel Berkeley Research lab data set: http://db.csail.mit.edu/labdata/labdata.html.
    - Data preprocessing is done according to [1]
    - Evaluation on univariate data: temperature of mote 7,
    - Evaluation on multivariate data: temperature of mote 7, the humidity, light, and voltage data.
  - Distillate flowrate data set: https://openmv.net/info/distillate-flow
- Both models are evaluated and compared w.r.t. the compression quality and their resource consumption.

[1] J. Liu, F. Chen, J. Yan, and D. Wang, "CBN-VAE: A Data Compression Model with Efficient Convolutional Structure for Wireless Sensor Networks," *Sensors*, vol. 19, no. 16, 2019.

Evaluation Metrics

Compression Performance:

- Compression ratio CR: $$CR = \frac{n_o\ V_o}{n_c\ V_c} = \frac{n_o\ s_o\ d_o\ b}{n_c\ s_c\ d_c\ b}$$
- Accuracy, i.e. reconstruction error RE: $$RE = PRMSE = 100 * \sqrt{\frac{\sum_{i=0}^{V-1} (O_i - R_i)^2}{\sum_{i=0}^{V-1} (O_i)^2}}$$
- Quality score QS: $$QS = \frac{CR}{PRMSE}$$

rexroth
A Bosch Company

CODE UNIVERSITY OF APPLIED SCIENCES

Evaluation of compression performance

Table 1: Results of the evaluation of the compression quality

	UTS ($d=1$)			MTS ($d=4$)		
model	CR	RE	QS	CR	RE	QS
			Intel data set			
DWT	20	45.2	0.44	20	36.9	0.54
LSTM-RNN	20	5.1	7.9	20	10.8	3.7
eCBN-VAE	48	8.8	5.5	192	46.4	4.1
	Distillate flowrate data set ($d=1$)					
LSTM-RNN	20	18.8	2.1			
eCBN-VAE	48	53.9	0.9			

Evaluation of resource usage

Table 2: Results of the evaluation of the performance

model	FLOPs	memory [B]	parameters	t_{enc} [ms]	t_{dec} [ms]
dimension = 1, batch size = 1					
CBN-VAE	10.299.640	37.428	5237	44	236
LSTM-RNN	2.700	260	24	622	1092
dimension = 4, batch size = 1					
eCBN-VAE	15.837.880	91.860	6.965	67	321
LSTM-RNN	26.037	1.808	288	403	809

rexroth
A Bosch Company

C<>DE UNIVERSITY OF APPLIED SCIENCES

Results

Results show a more generalized and robust compression of time series via deep learning than DWT.

The following aspects are observed w.r.t. the autoencoder models:

1. The CBN-VAE is significantly faster than the LSTM-AE even though being the more complex model. This is due to the CBN-VAE being stateless and thus better parallelizable.
2. Both models perform worse on the second data set, which is explainable by the higher variance of the data. Compressing high-variance data sets is naturally worse due to less correlations.
3. The CBN-VAE achieves a high compression ratio of 192:1 for multivariate input data (dimensions d = 4) because in convolutional operations the number of input channels does not affect the dimensionality of the hidden representation. Thus, the model has to find the same hidden representation for both univariate and multivariate input data, explaining the higher compression ratio, but also higher reconstruction error on multivariate input.
4. The LSTM-AE shows better generalizability across different data structures.

rexroth
A Bosch Company

CODE UNIVERSITY OF APPLIED SCIENCES

Conclusion

- Two AE structures are investigated to compress IIoT streaming data.
- The two autoencoder models…
  - show better results for lossy compression than the state-of-the-art method Discrete Wavelet Transform (DWT).
  - satisfy the characteristics of small available memory, small computational capacity, and limited bandwidth volume → expected to be suited to be run on industrial edge devices.
- Further evaluations on industrial edge devices are in progress.
- Further comparisons to other state-of-the-art methods like lossy fpzip [3] and quantization are in progress.
- Attention-based transformer model structures are another promising modern approach. The applicability of sparse transformer model structure to time series data should be investigated.

[3] P. Lindstrom and M. Isenburg, "Fast and Efficient Compression of Floating-Point Data," *IEEE Transactions on Visualization and Computer Graphics, Proceedings of Visualization 2006*, 12(5):1245-1250, September-October 2006.

Data Compression Conference

IEEE Signal Processing Society TECHNICAL CO-SPONSOR

Microsoft

THE UNIVERSITY OF ARIZONA

THANK YOU!

Comparison and extension of autoencoder models for uni- and multivariate signal compression in IIoT

Julia Rosenberger[a], Alexander Kübel[b], Fabian Rothfuß[b]

[a]Bosch Rexroth AG, Bürgermeister-Dr.-Nebel-Straße 15, Lohr am Main 97816, Germany

[b]University of Applied Science, Lohmühlenstraße 65, Berlin 12435, Germany

CODE UNIVERSITY OF APPLIED SCIENCES

rexroth A Bosch Company