University of Latvia
Institute of Mathematics and Computer Science

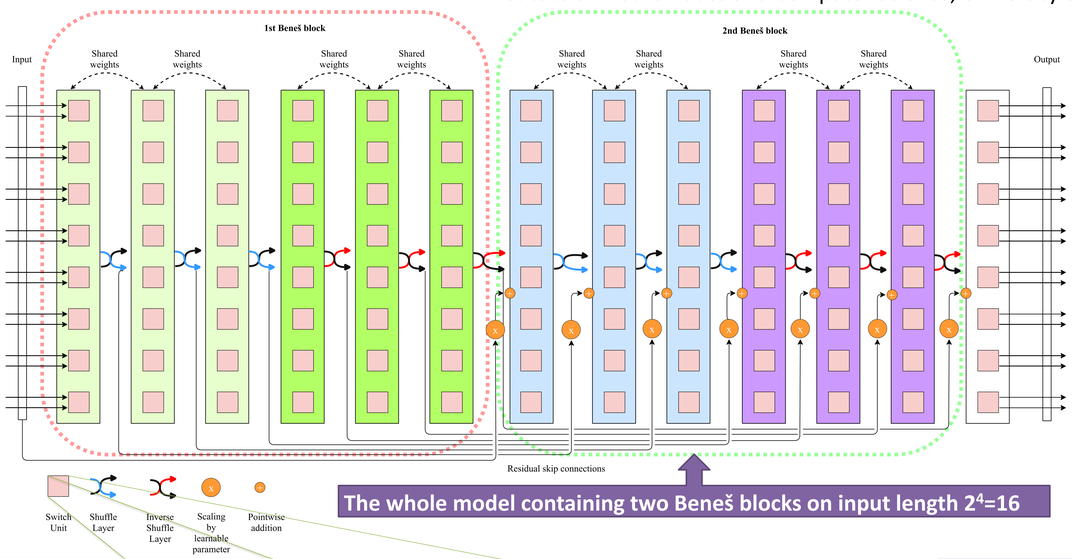
Neural Shuffle-Exchange Networks – SequenceProcessing in O(*n* log *n*) Time

Kārlis Freivalds, Emīls Ozoliņš, Agris Šostaks
Institute of Mathematics and Computer Science, University of Latvia

LATVIJAS UNIVERSITĀTE

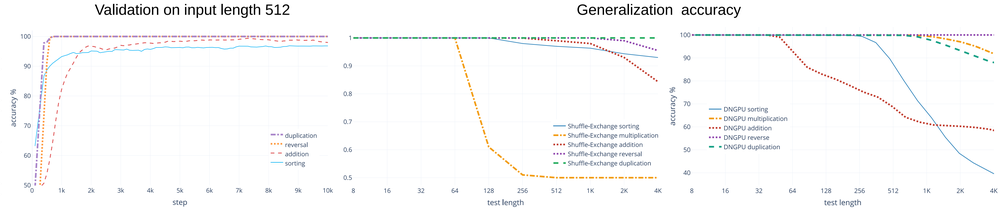
A key requirement in sequence to sequence processing is the modeling of long range dependencies. To this end, a vast majority of the state-of-the-art models use attention mechanism which is of O(n^2) complexity that leads to slow execution for long sequences.

We propose a new **differentiable architecture for sequence processing tasks that has O(log *n*) depth, O(*n* log *n*) total complexity**, and allows modeling of any dependencies in the sequence. We show that this model can successfully synthesize nontrivial O(*n* log *n*) time algorithms with good generalization.

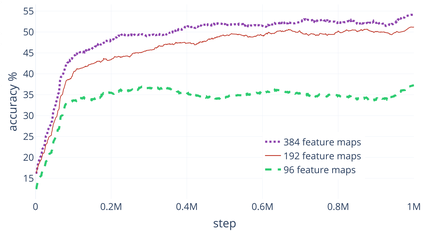


The whole model containing two Beneš blocks on input length 2^4=16

Shuffle-Exchange network can infer O(n log n) time algorithms from input-output examples.



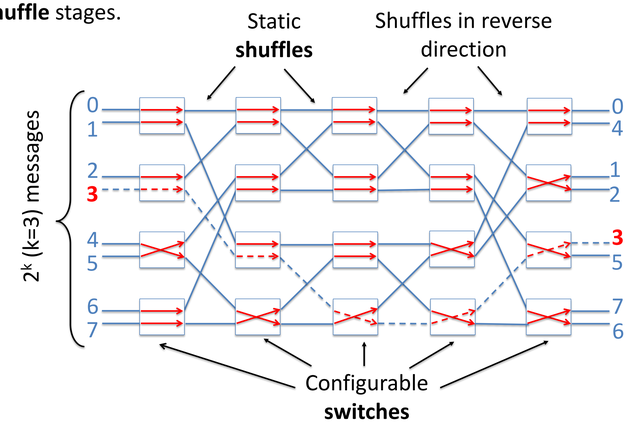
Achieves competitive accuracy on real-world tasks, LAMBADA word prediction.



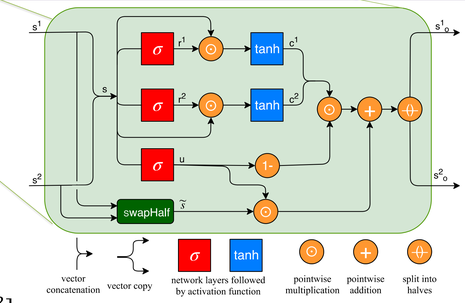
Model	Test accuracy (%)
Random word from passage	1.60
Gated-Attention Reader	49.00
Shuffle-Exchange Network	**52.28**
Universal Transformer	56.00
Human performance	86.00

Beneš Network

Beneš network can route **2^k messages** in any input-to-output permutation. Beneš network has **2k−1 exchange** stages and **2k−2 shuffle** stages.



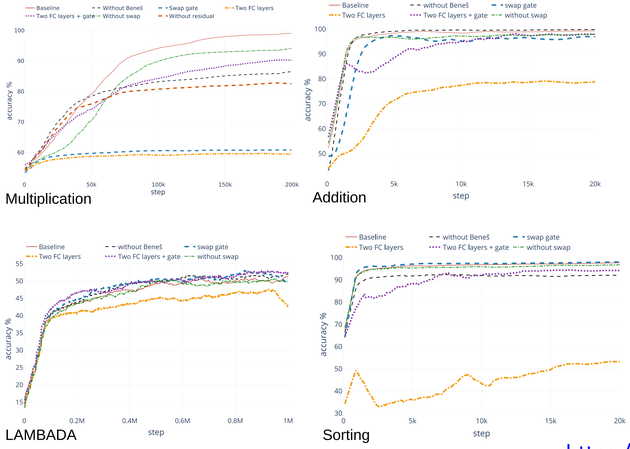
Switch Unit



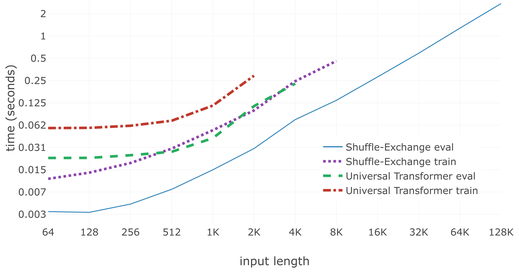
$$s = [s^1, s^2]$$
$$r^1 = \sigma(W_r^1 s + B_r^1) \quad r^2 = \sigma(W_r^2 s + B_r^2)$$
$$c^1 = \tanh(W_c^1 (r^1 \odot s) + B_c^1)$$
$$c^2 = \tanh(W_c^2 (r^2 \odot s) + B_c^2)$$
$$u = \sigma(W_u s + B_u)$$
$$\tilde{s} = swapHalf(s^1, s^2)$$
$$[s_o^1, s_o^2] = u \odot \tilde{s} + (1 - u) \odot [c^1, c^2]$$

$$swapHalf \quad \binom{a}{b}\binom{c}{d} \Longrightarrow \binom{a}{d}\binom{c}{b}$$

All the introduced features contribute to the models performance



A faster alternative to the attention mechanism feasible for longer sequences



This research is funded by the Latvian Council of Science, project No. lzp-2018/1-0327

Latvijas Zinātnes padome

Karlis.Freivalds@lumii.lv

https://github.com/LUMII-Syslab/shuffle-exchange