Department of Telecommunications and Media Informatics
Budapest University of Technology and Economics

SAMI2021

IEEE 19th World Symposium on Applied Machine Intelligence and Informatics

Efficient Neural Architecture Search for Long Short-Term Memory

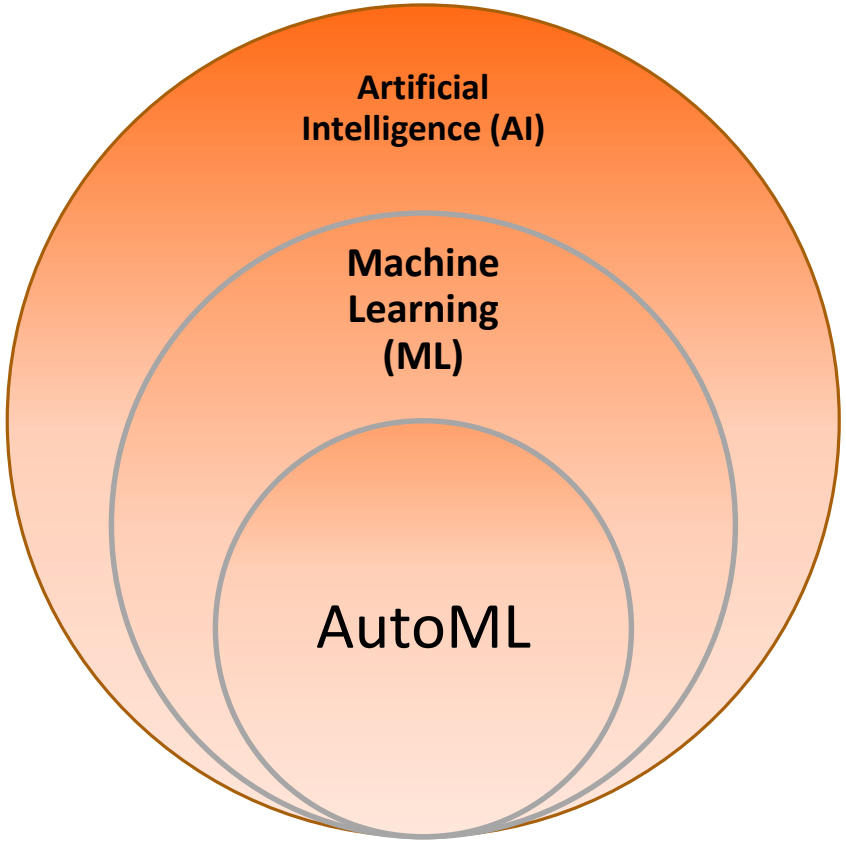
HAMDI ABED

BALINT GIRES-TOTH, PH.D.

21ST , JAN, 2021

Table of content

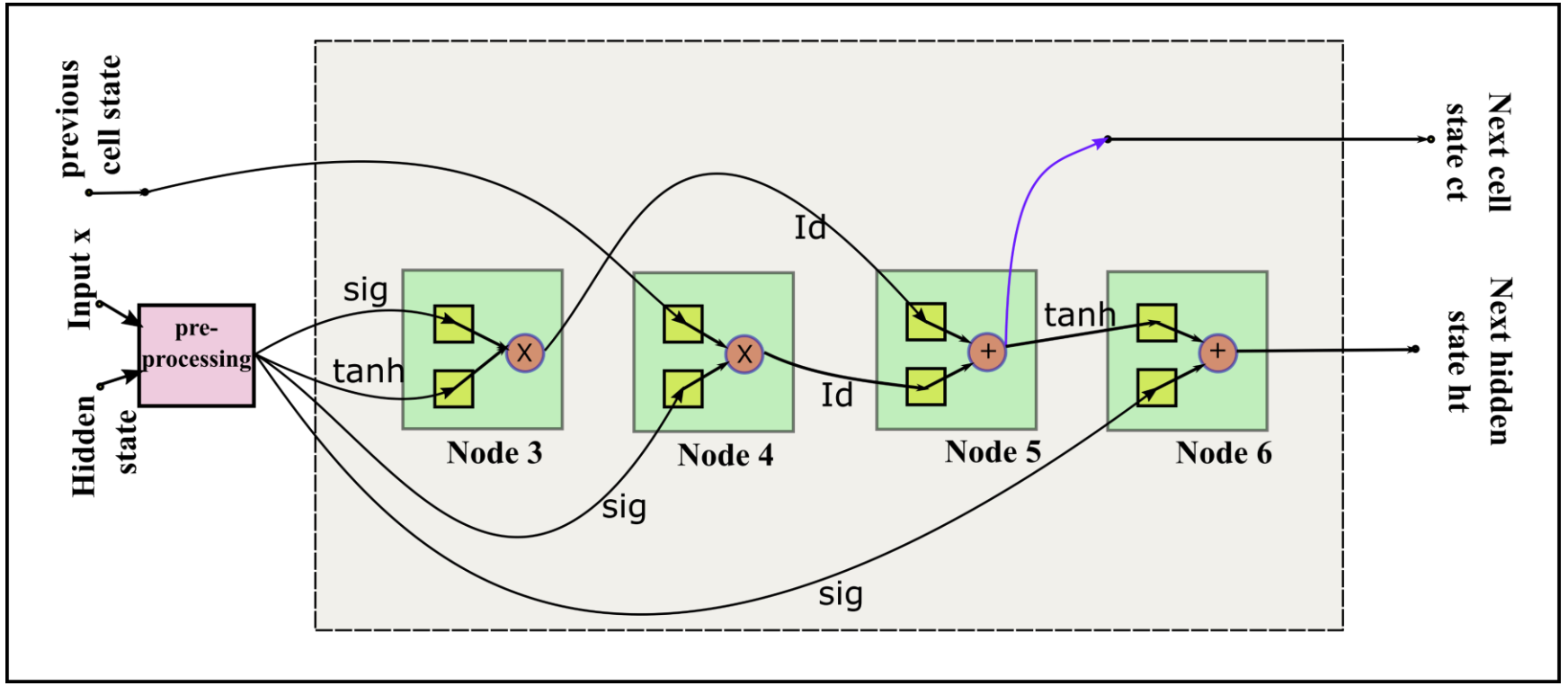
Introduction



Sequence modeling with deep learning

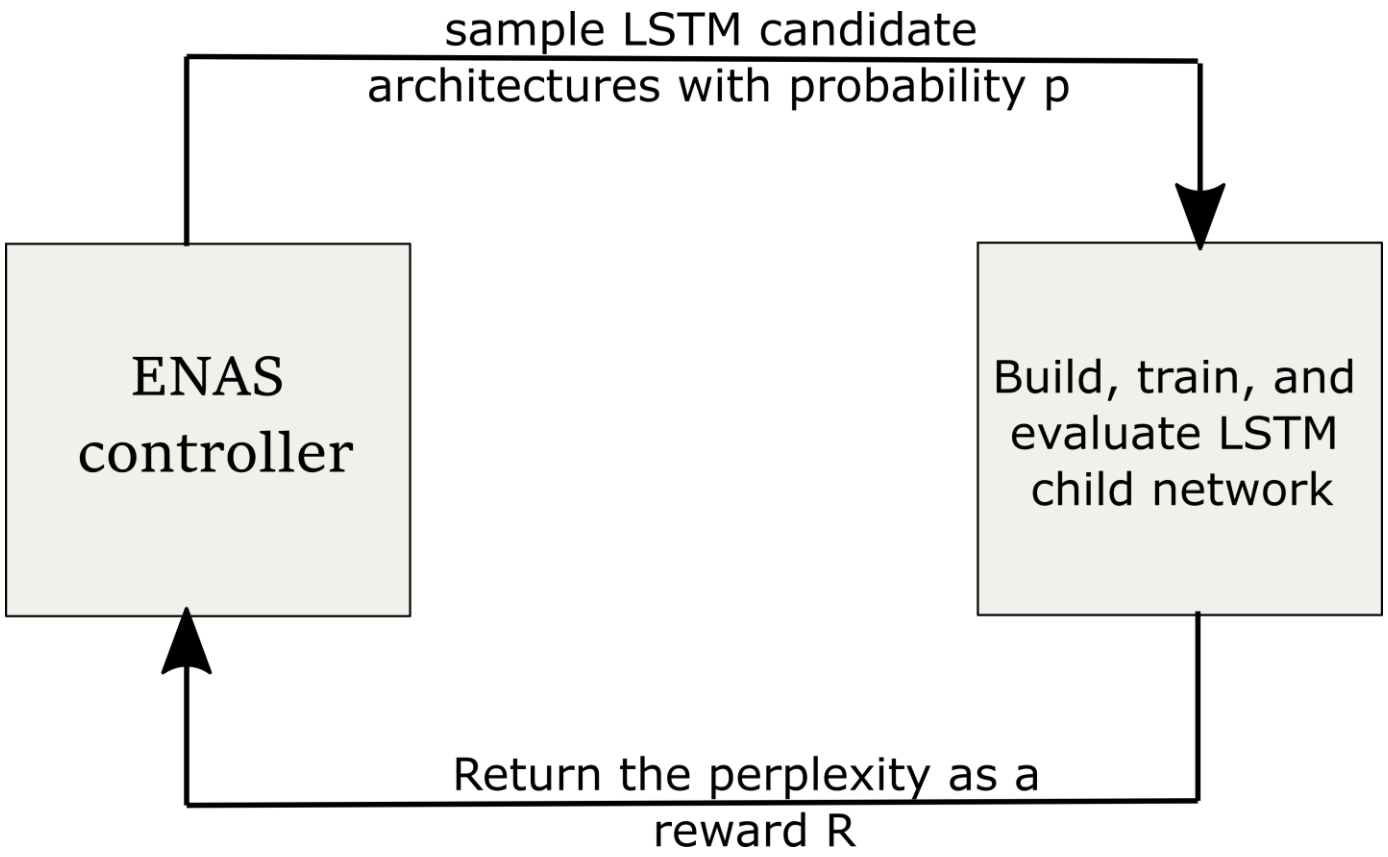
Baseline approaches:

- Recurrent Neural Networks
  - Long Short-Term Memory
  - Gated Recurrent Unit
  - Etc.
- Convolutional Neural Networks
- Transformer Networks



Vanilla Long Short-Term Memory

Efficient Neural Architecture Search



Research goals

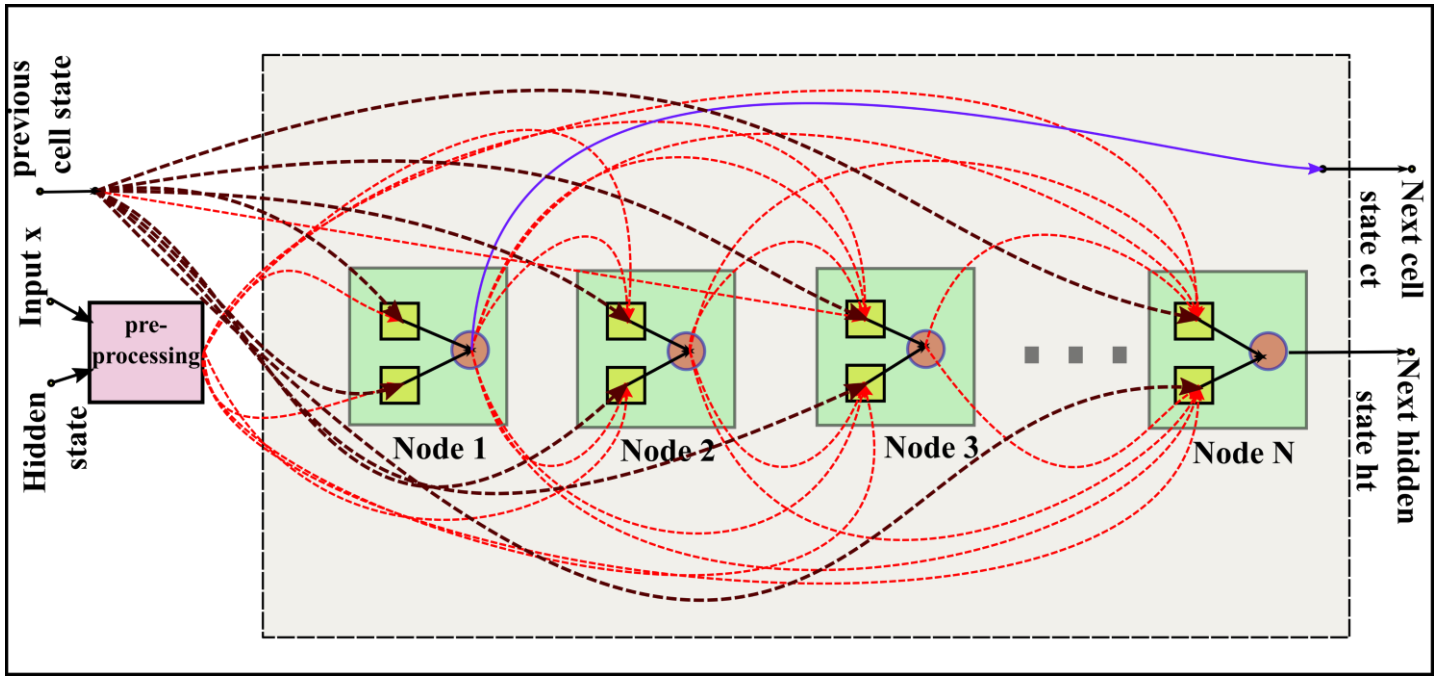
- Is vanilla LSTM optimal?

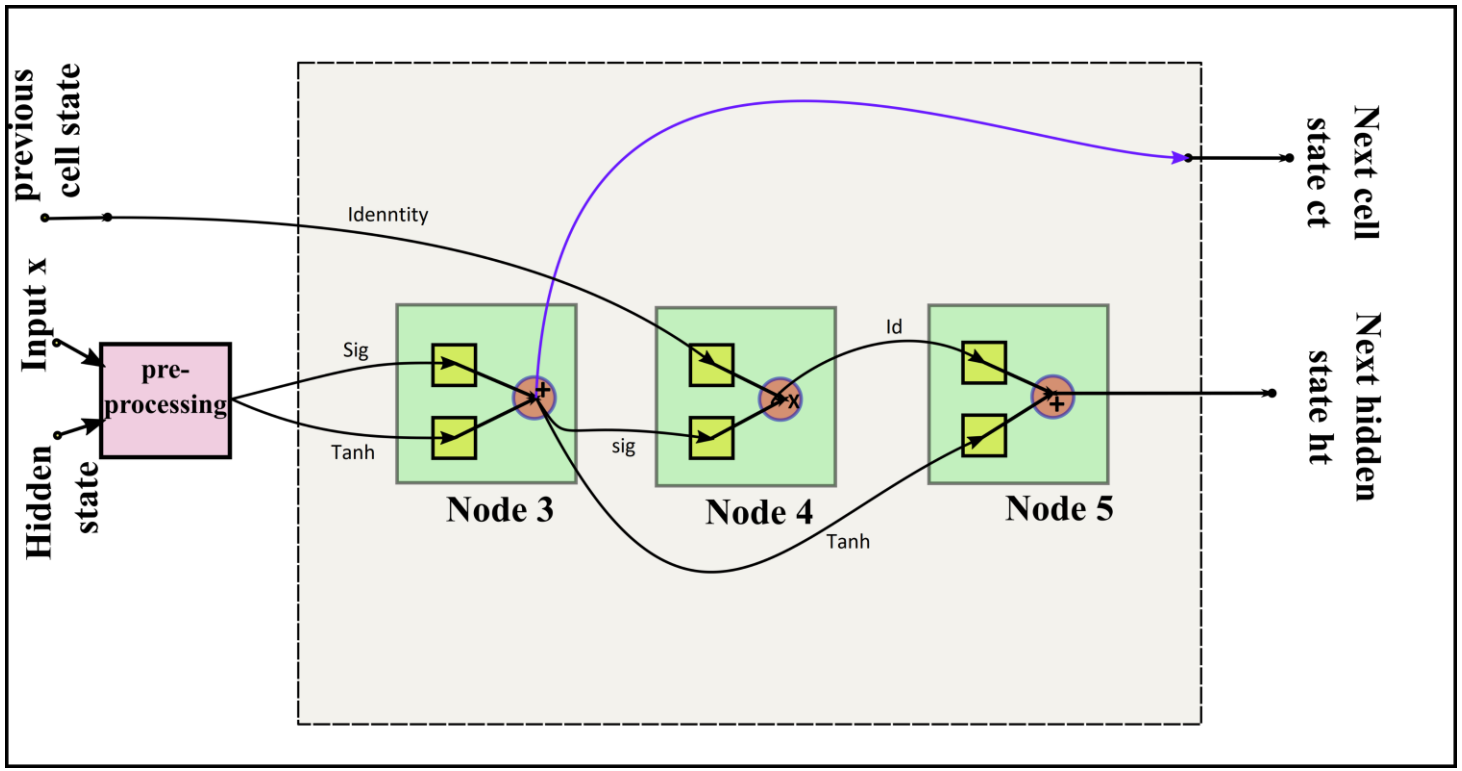
- Would ENAS converge to similar architecture as LSTM?

Proposed Method

Search space:

- Which two previous nodes to select
- What activation function to apply on each connection
- What operation to perform per node

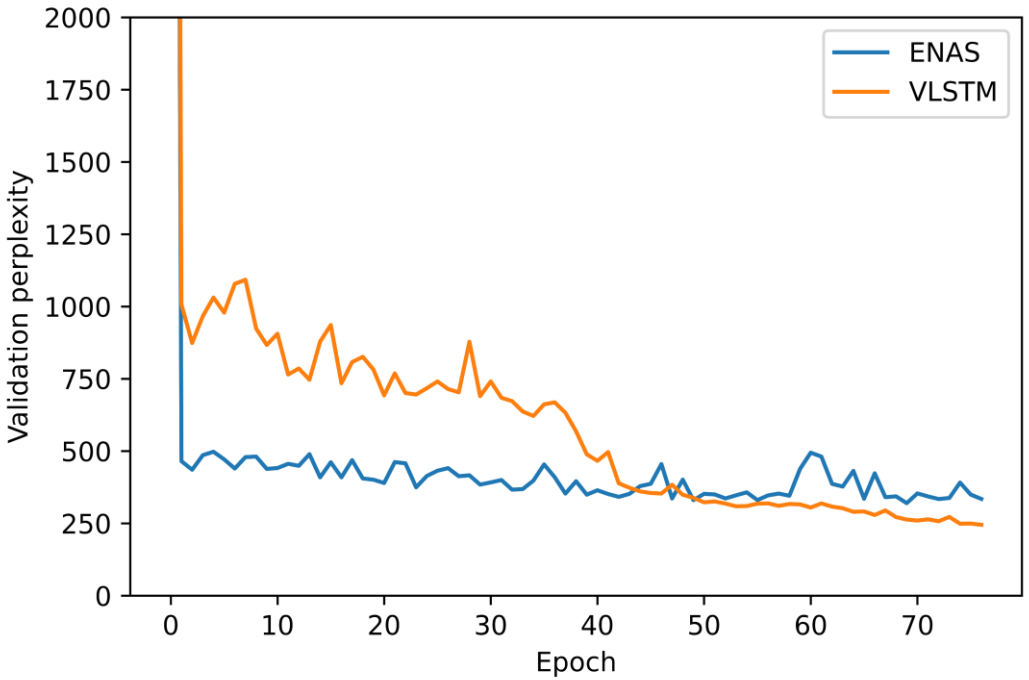




Example

Construction:

- $N_3 = Sig(x) + Tanh(x)$
- $N_4 = Id\left(Ct_{previous}\right) \cdot Sig(N_3)$
- $N_5 = Id(N_4) + Tanh(N_3)$



Experiments and results

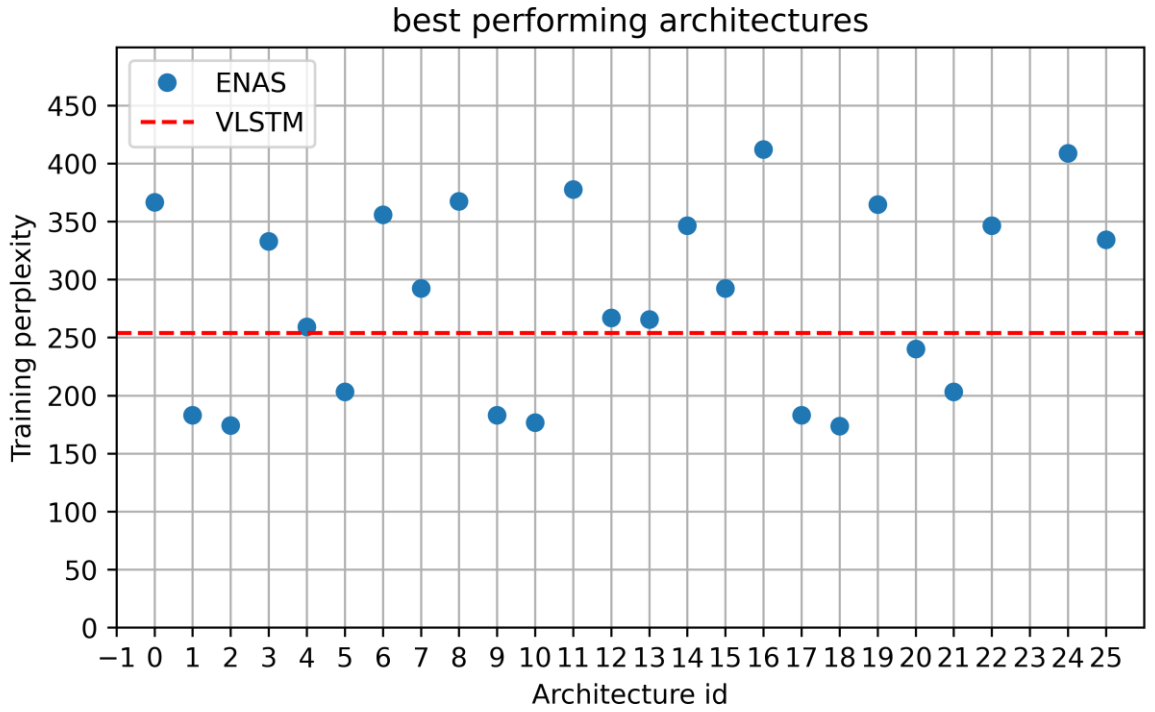
Dataset: Penn TreeBank

Trained on a single GPU for 12 hours each

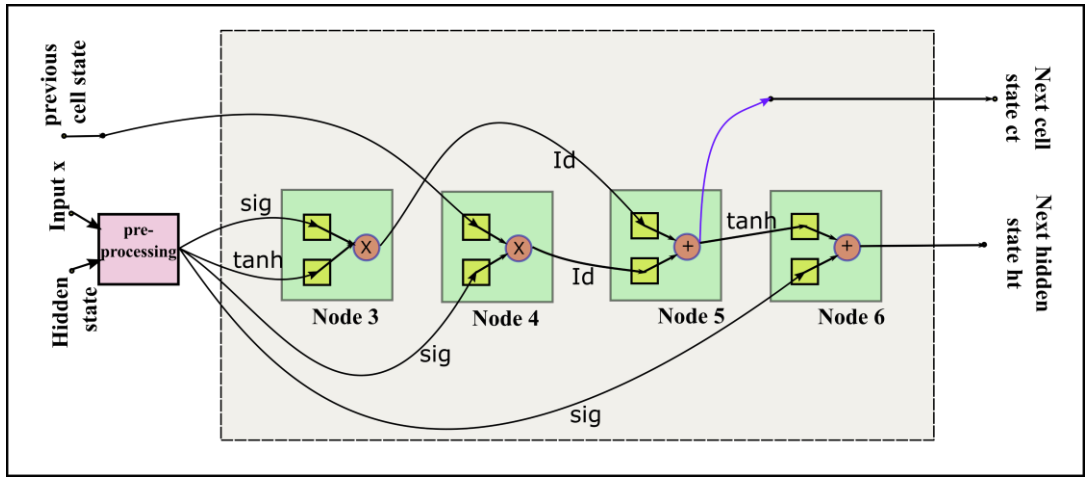
LSTM outperformed the generated ENAS children

Model	# of params	Validation Perplexity
ENAS	9.47M	254.3
Vanilla LSTM	8.00M	173.5

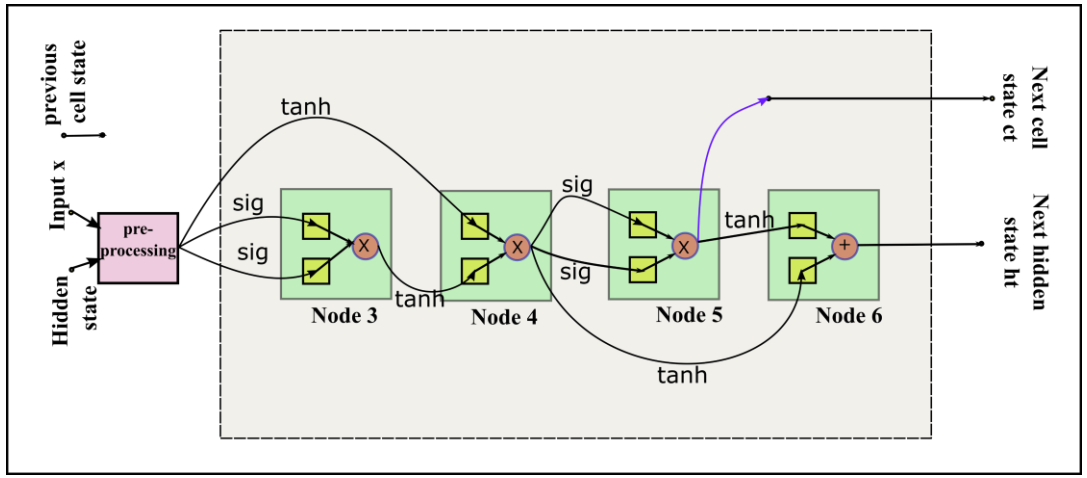
Experiments and results



Experiments and results



Vanilla LSTM architecture



Best ENAS generated architecture

Conclusions

- Vanilla LSTM outperformed ENAS generated children
- ENAS best performing architecture is close to LSTM architecture
- We assume that Vanilla LSTM is empirically an optimal LSTM architecture to model text generation task, word-level on Penn TreeBank dataset

Thank you for your attention

Hamdi Abed , Balint Gyires-Toth

Department of Telecommunications and Media Informatics

Budapest University of Technology and Economics

{Hamdiabed, toth.b}@tmit.bme.hu

http://smartlab.tmit.bme.hu/