RECURRENT NEURAL NETWORK IN NATURAL LANGUAGE PROCESSING

Sütő Evelyne [1]

[1]Babeş-Bolyai University, Computer Science, Cluj-Napoca

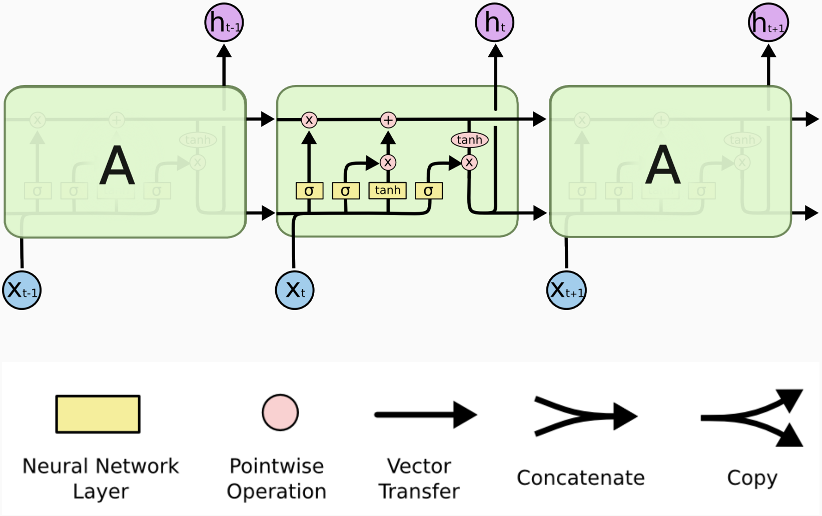
NATURAL LANGUAGE PROCESSING

- **Purpose:** Explore the applications of Recurrent Networks for NLP
- **Used RNN:** LSTM
- **Used architecture:** Encoder-Decoder model
- **Decoding algorithms:** Greedy decoding and Beam search

NATURAL LANGUAGE PROCESSING

- Neural Machine translation
- Text summarization
- Speech to text
- Paraphrase generation

LSTM



h t-1
h t
h t+1
A
A
tanh
σ
σ
tanh
σ
X t-1
X t
X t+1
Neural Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy

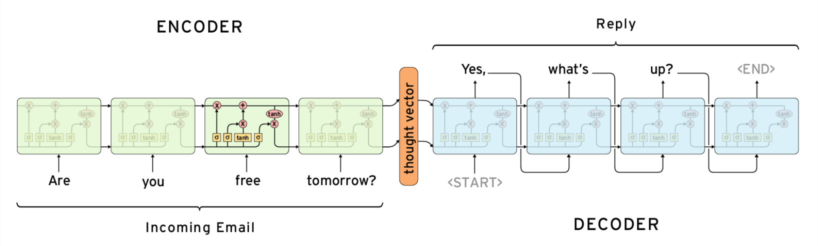


Figure 1: Image source:
`https://medium.com/botsupply/generative-model-chatbots/`

Greedy Decoder

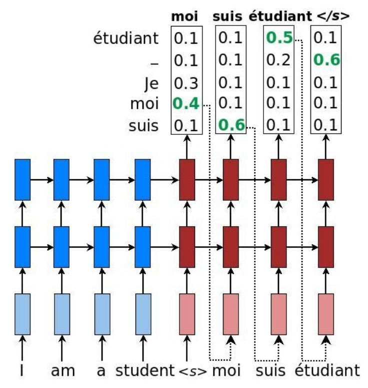


Figure 2: Image source: `https://www.packtpub.com/mapt/book/big_data_and_business_intelligence/`

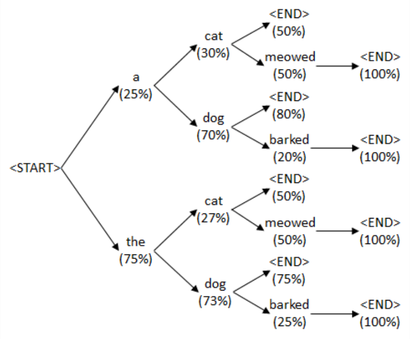


Figure 3: Image source:
`https://geekyisawesome.blogspot.com/2016/10/`

NEURAL MACHINE TRANSLATION

- Researchers have applied this model to English-French translation [Sutskever et al., 2014] in 2014
- Conclusions
  - LSTM sequence to sequence models are able to map even very long sentences to the translation language
  - Deep LSTMs can significantly outperform shallow LSTMs
  - Unoptimized model was able to produce state of the art results with relatively short training
  - State of the art result: 37 BLEU score
  - This model's result: 36.5
- Researchers improved this unoptimized version [Wu et al., 2016] in 2016 to use for Google's translation system

SPEECH RECOGNITION

- Another application is shown in [Graves and Jaitly, 2014] done in 2014 for speech recognition system.

- Conclusions
  - Bi-Directional LSTMs are needed to exploit future context as well.
  - Connectionist Temporal Classification
  - Model that does not need data preprocessing.
  - Their best score on 81 hour dataset (Wall street journal) is 8.2 while the baseline was 7.8 (word error rate/character error rate)

- Sequence to sequence model was used for text summarization in [Nallapati et al., 2016] in 2016
- Conclusions
  - Encoder-Decoder model with a bi-directional GRU-RNN as the encoder and a uni-directional GRU-RNN.
  - Address the problem of rarewords
  - They have reached state of the art results.

PARAPHRASE GENERATION

- Paraphrasing, the act to express the same meaning in different possible ways.
- Paraphrase generation has been researched in [Prakash et al., 2016] in 2016.

- Conclusions
  - Encoder-Decoder model with deep LSTM networks.
  - They have reached state of the art results.

Graves, A. and Jaitly, N. (2014).
Towards end-to-end speech recognition with recurrent neural networks.
In *International Conference on Machine Learning*, pages 1764–1772.

Nallapati, R., Zhou, B., Gulcehre, C., Xiang, B., et al. (2016).
Abstractive text summarization using sequence-to-sequence rnns and beyond.
arXiv preprint arXiv:1602.06023.

Prakash, A., Hasan, S. A., Lee, K., Datla, V., Qadir, A., Liu, J., and Farri, O. (2016).
Neural paraphrase generation with stacked residual lstm networks.
arXiv preprint arXiv:1610.03098.

Sutskever, I., Vinyals, O., and Le, Q. V. (2014).
Sequence to sequence learning with neural networks.

In *Advances in neural information processing systems*, pages 3104–3112.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., et al. (2016).
Google's neural machine translation system: Bridging the gap between human and machine translation.
arXiv preprint arXiv:1609.08144.