CISC 372
Text Analytic IIII
Tensorflow LAB

https://giphy.com/gifs/kid-costume-transformer-R9zXHWAHyTjnq

Last lectures

- Text Analysis
  - Vanilla RNN
  - Gated Recurrent Unit
  - LSTM
  - Attention Mechanism
    - Memory
      - things we memorized
    - Context
      - based on the context, on which part of memory should we focus
  - Multi-head attention
  - Attention – Explainability
  - Padding (sorting the sequence)
  - RNN for different sequence problems

Last lectures

- Language Model
- Neural Network for Language Model
- RNN for Language Modeling
- Transformer

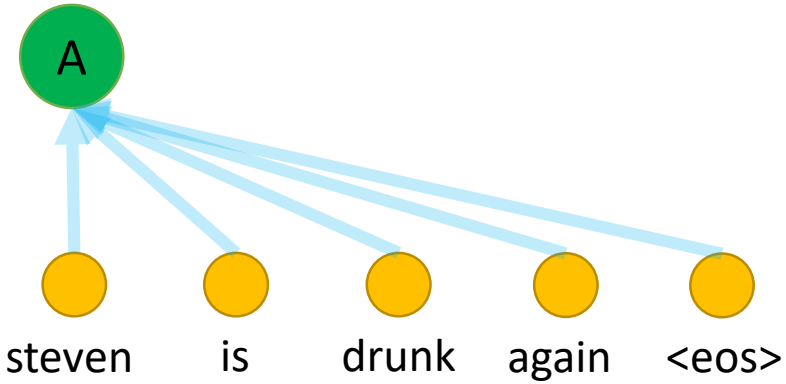
Issue

- Long-term dependencies
  - Different cell implementation
  - Attention
- Computational complexity
  - The time stamp t's calculation
    - depends on time stamp t-1
  - For encoder/decoder:
    - Required *2 times t* passes over the memory cell to train a batch/sample

The Transformer – Attention is all you need

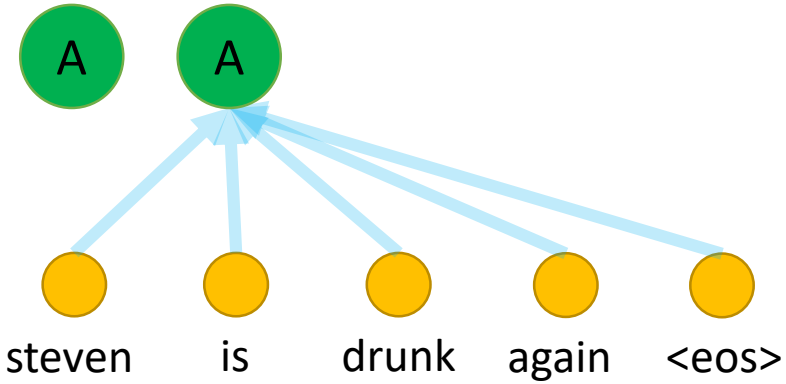
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



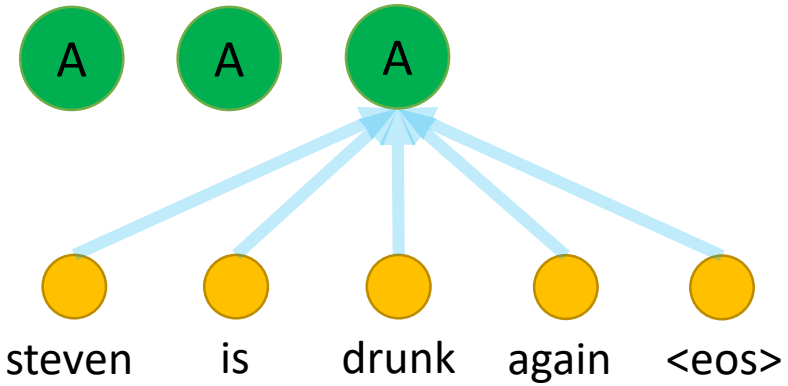
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



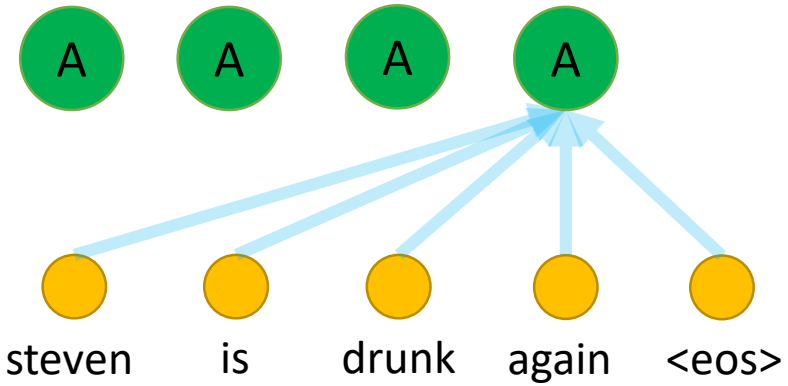
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



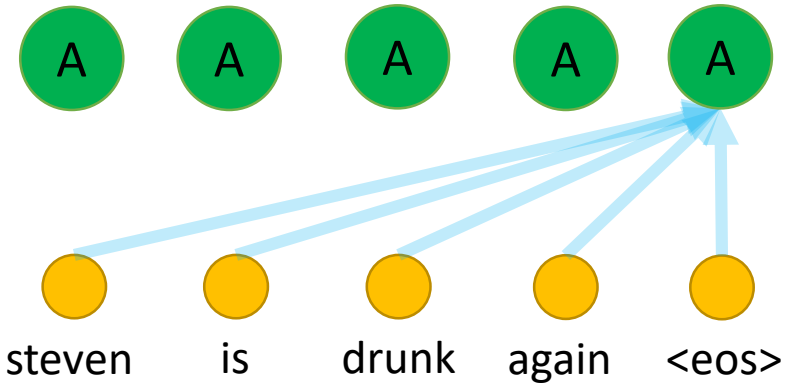
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



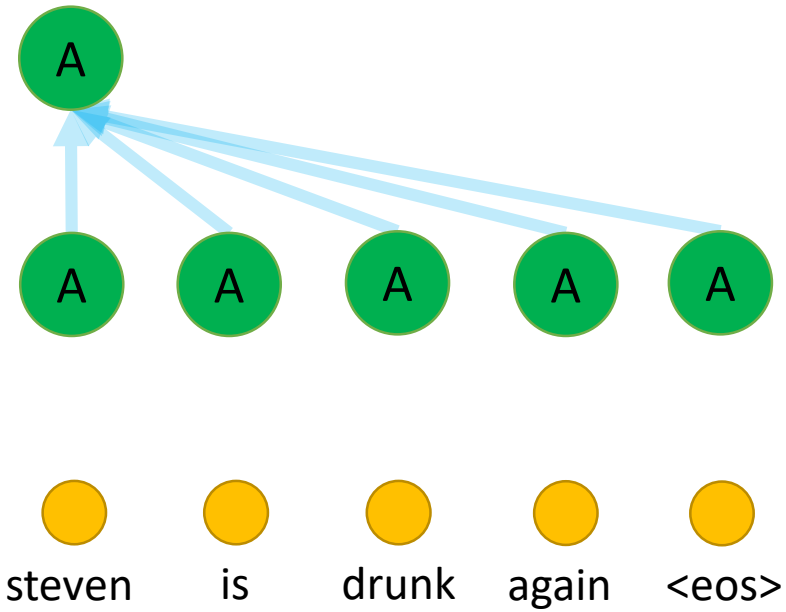
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



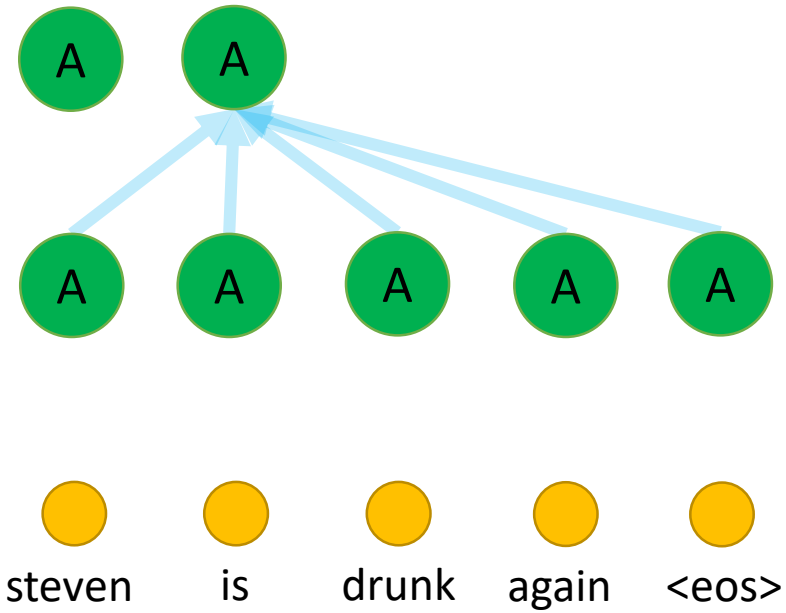
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



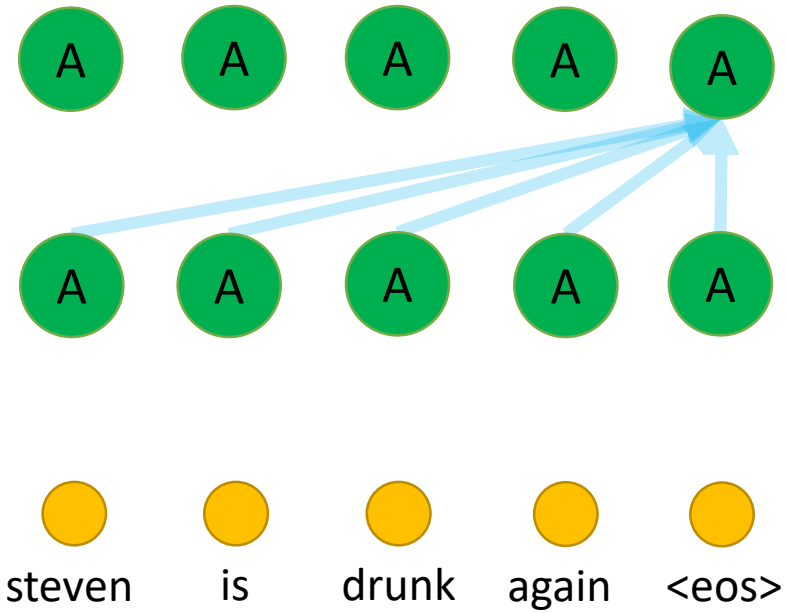
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



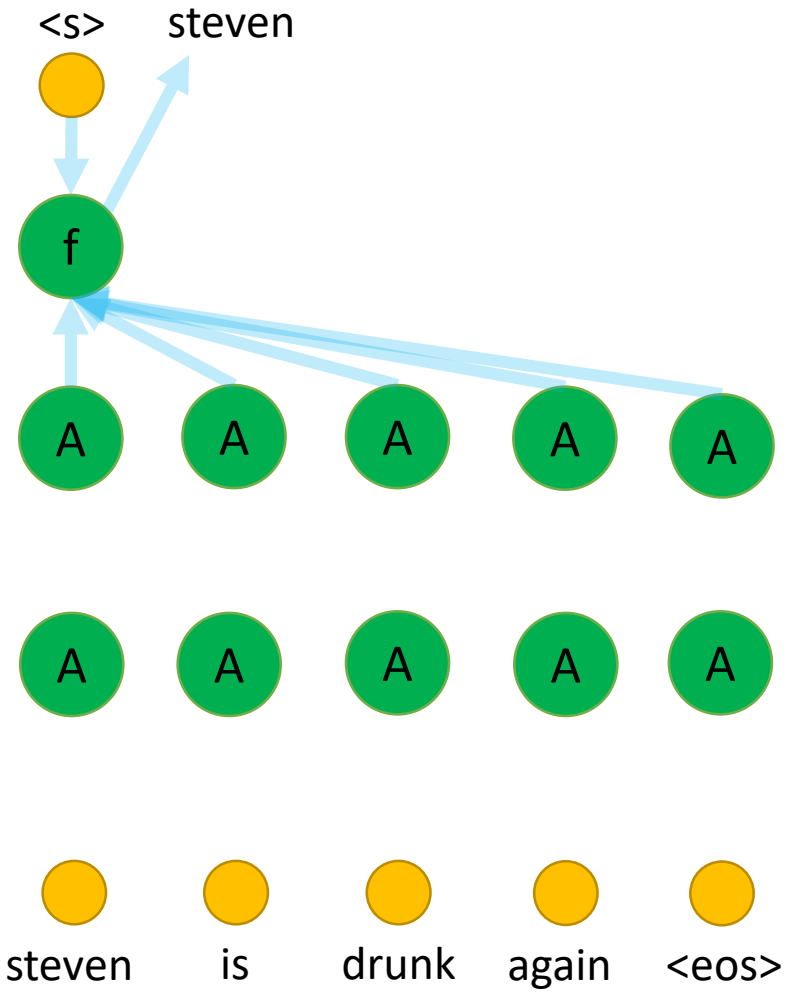
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



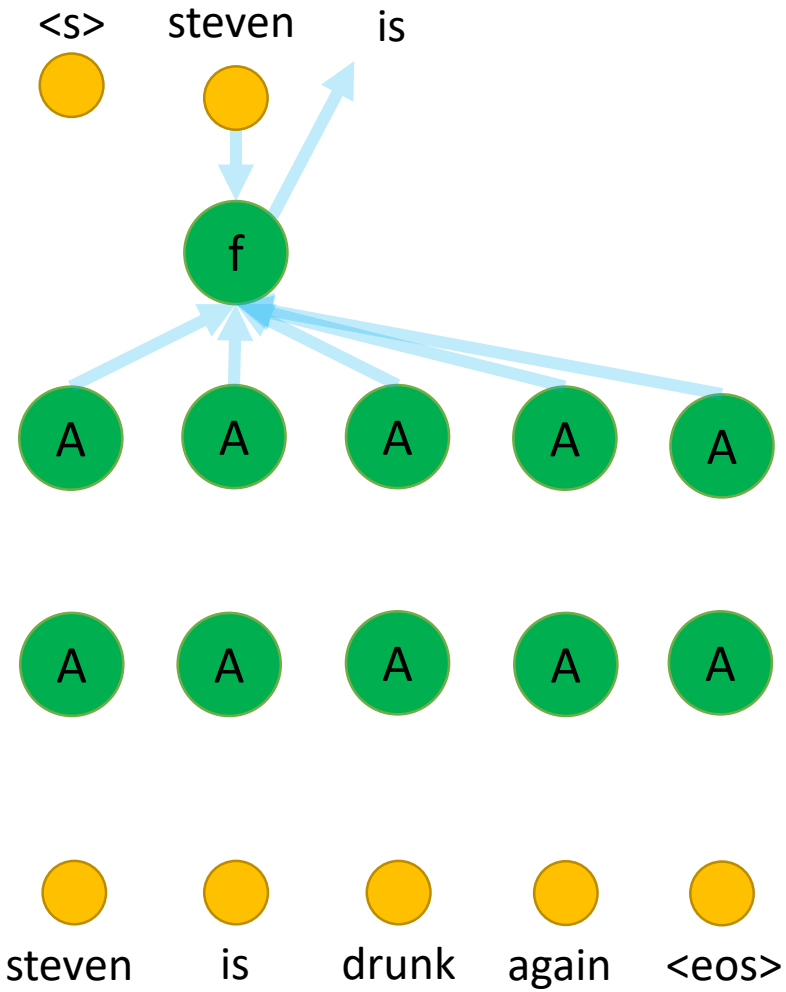
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



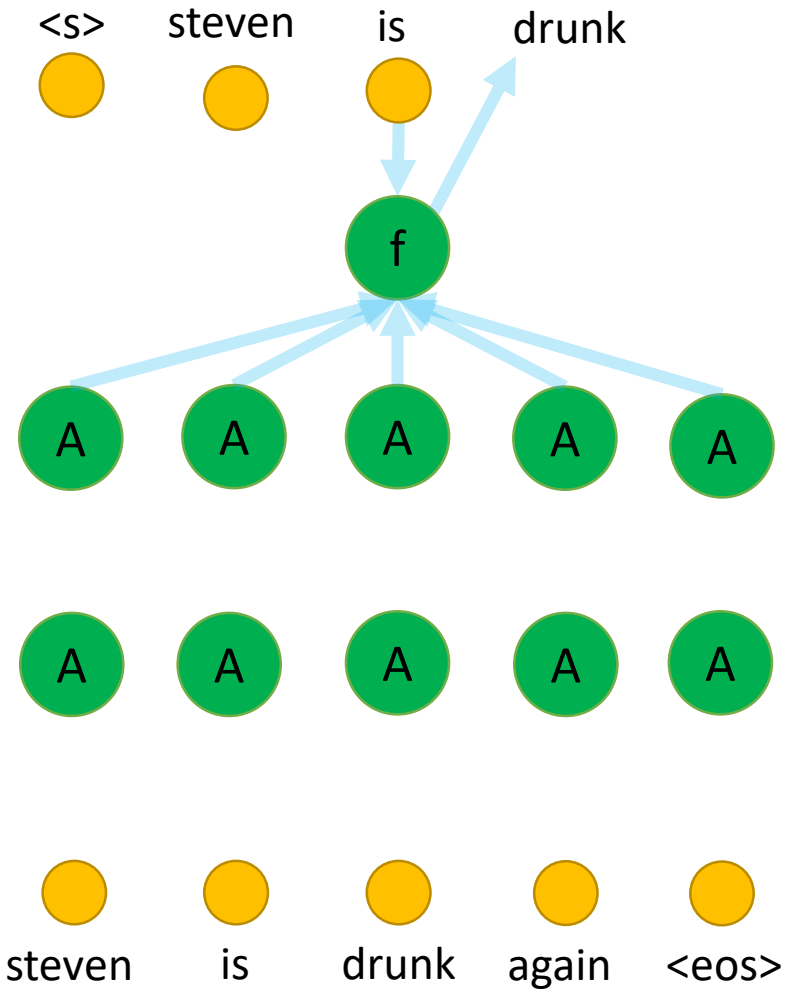
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



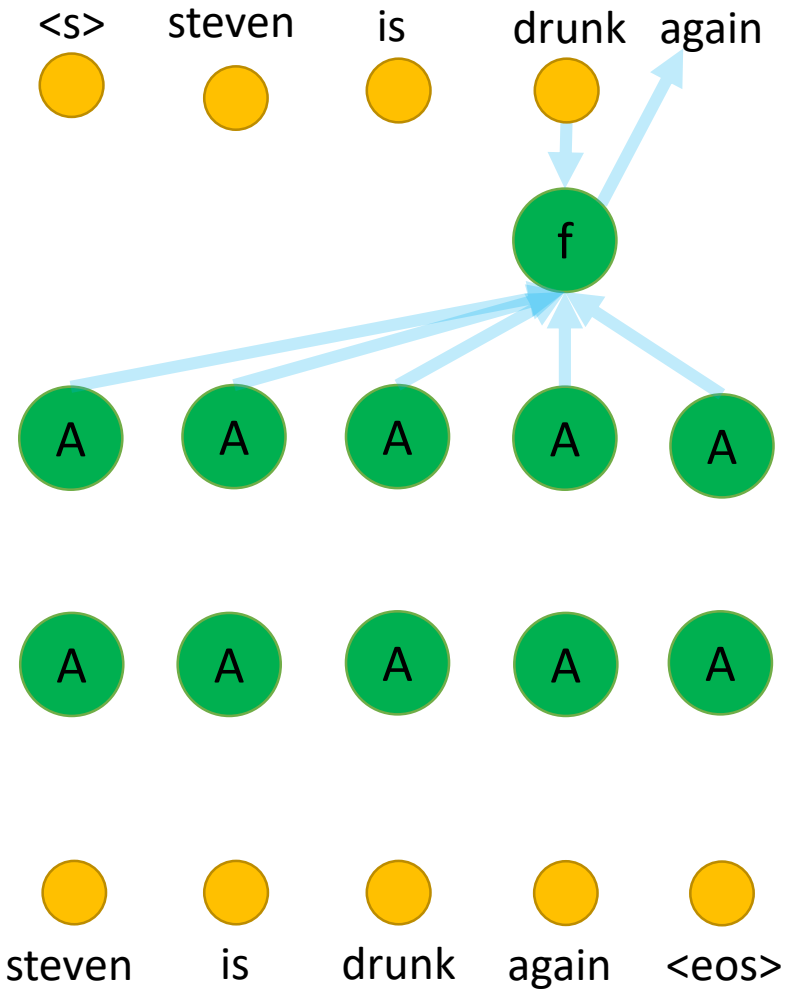
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



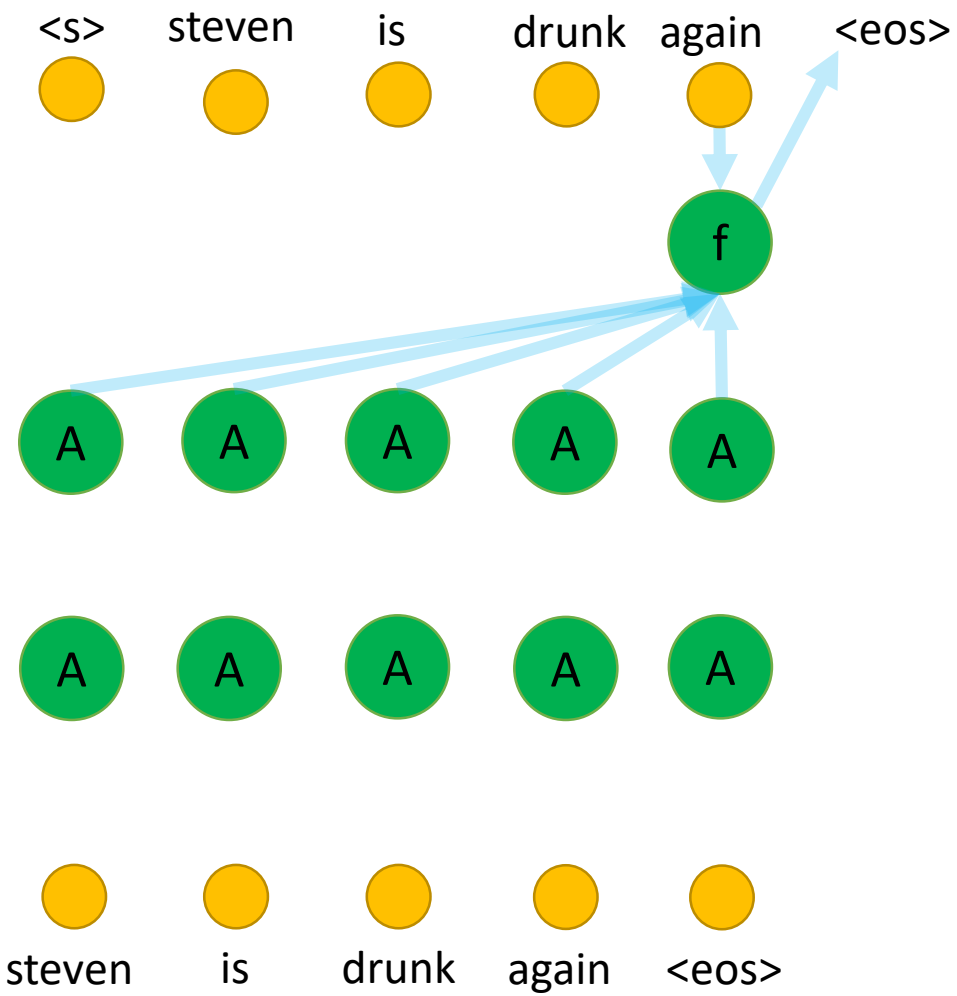
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



http://karpathy.github.io/2015/05/21/rnn-effectiveness/

The Transformer – Attention is all you need



http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Performance on down-stream tasks:

	Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm	QNLI	RTE	WNLI	AX
	1	T5 Team - Google	T5		90.3	71.6	97.5	92.8/90.4	93.1/92.8	75.1/90.6	92.2	91.9	96.9	92.8	94.5	53.1
	2	ERNIE Team - Baidu	ERNIE		90.1	72.8	97.5	93.2/91.0	92.9/92.5	75.2/90.8	91.2	90.8	96.1	90.9	94.5	49.4
	3	Microsoft D365 AI & MSR AI & GATECH	MT-DNN-SMART		89.9	69.5	97.5	93.7/91.6	92.9/92.5	73.9/90.2	91.0	90.8	99.2	89.7	94.5	50.2
+	4	Alibaba DAMO NLP	ALICE v2 large ensemble (Alibaba DAMO NLP)		89.7	73.2	97.1	93.9/91.9	93.0/92.5	74.8/91.0	90.8	90.6	95.9	87.4	94.5	48.7
+	5	Microsoft D365 AI & UMD	FreeLB-RoBERTa (ensemble)		88.4	68.0	96.8	93.1/90.8	92.3/92.1	74.8/90.3	91.1	90.7	95.6	88.7	89.0	50.1
	6	Junjie Yang	HIRE-RoBERTa		88.3	68.6	97.1	93.0/90.7	92.4/92.0	74.3/90.2	90.7	90.4	95.5	87.9	89.0	49.3
	7	Facebook AI	RoBERTa		88.1	67.8	96.7	92.3/89.8	92.2/91.9	74.3/90.2	90.8	90.2	95.4	88.2	89.0	48.7
+	8	Microsoft D365 AI & MSR AI	MT-DNN-ensemble		87.6	68.4	96.5	92.7/90.3	91.1/90.7	73.7/89.9	87.9	87.4	96.0	86.3	89.0	42.8
	9	GLUE Human Baselines	GLUE Human Baselines		87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0	92.8	91.2	93.6	95.9	-

http://karpathy.github.io/2015/05/21/rnn-effectiveness/