CS489/698
Lecture 18: March 13, 2017

Recurrent and Recursive Neural Networks

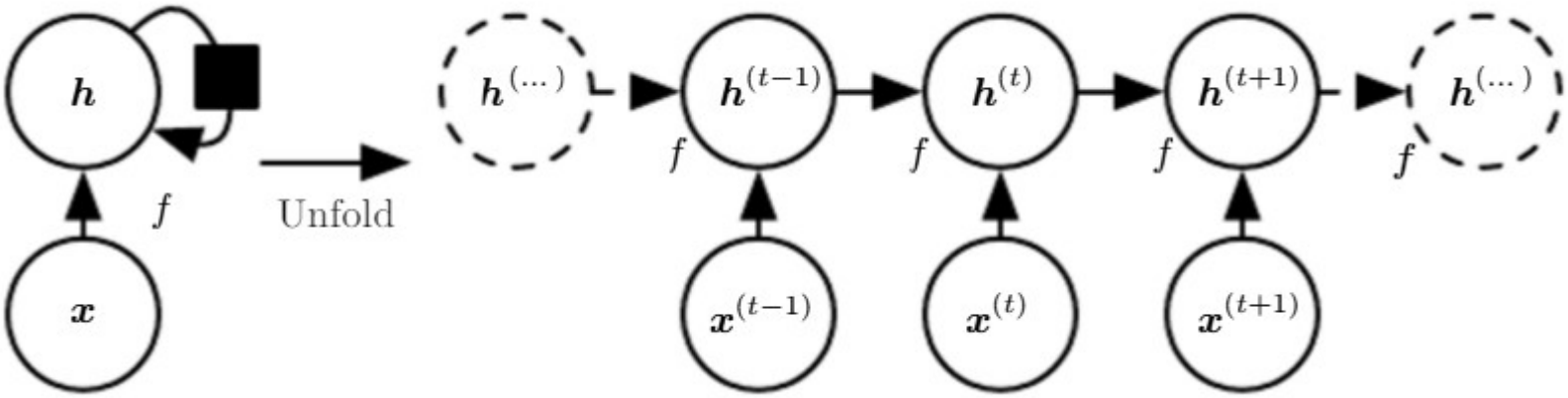
[GBC] Chap. 10

Variable length data

- Traditional feed forward neural networks can only handle fixed length data
- Variable length data (e.g., sequences, time-series, spatial data) leads to a variable # of parameters
- Solutions:
  - Recurrent neural networks
  - Recursive neural networks

Recurrent Neural Network (RNN)

- In RNNs, outputs can be fed back to the network as inputs, creating a recurrent structure that can be unrolled to handle varying length data.



Training

- Recurrent neural networks are trained by backpropagation on the unrolled network
  - E.g. backpropagation through time
- Weight sharing:
  - Combine gradients of shared weights into a single gradient
- Challenges:
  - Gradient vanishing (and explosion)
  - Long range memory
  - Prediction drift

RNN for belief monitoring

- HMM can be simulated and generalized by a RNN

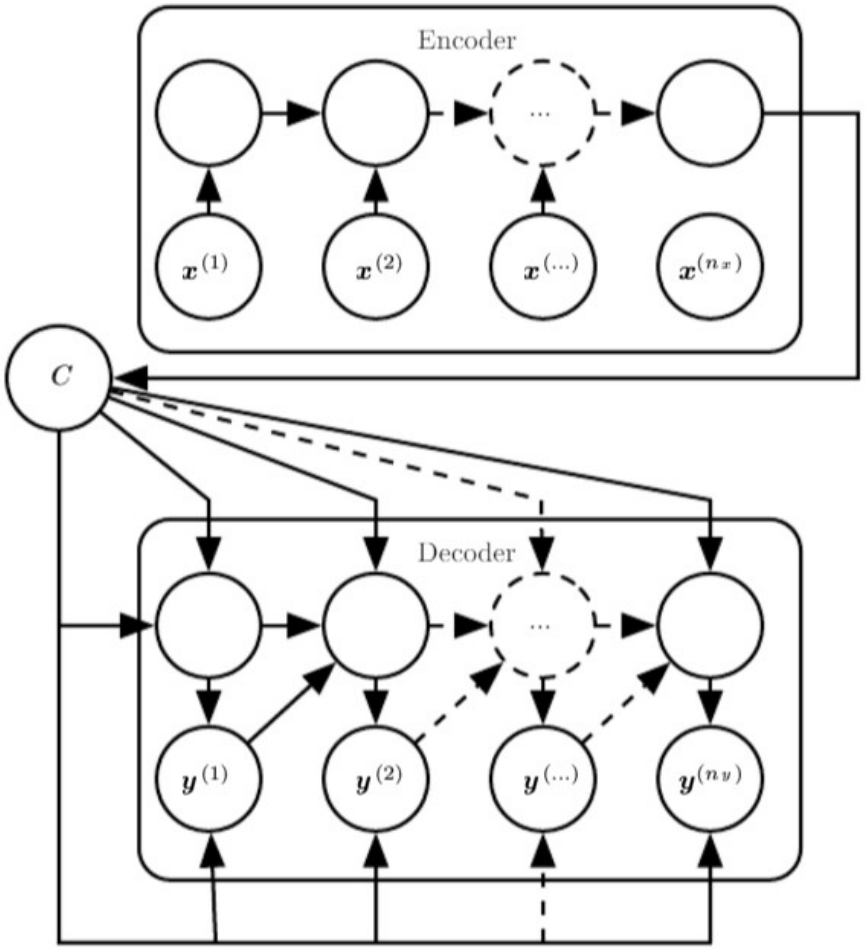
Bi-Directional RNN

- We can combine past and future evidence in separate chains

Encoder-Decoder Model

- Also known as sequence2sequence

- Usage:
  - Machine translation
  - Question answering
  - Dialog



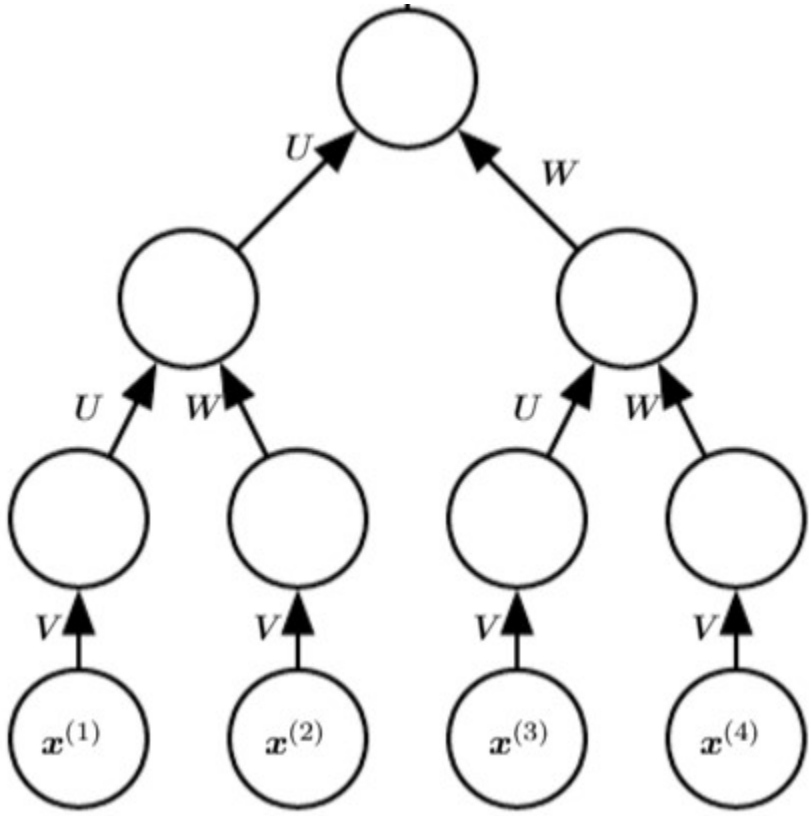
Machine Translation

- Cho, van Merrienboer, Gulcehre, Bahdanau, Bougares, Schwenk, Bengio (2014) Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation

Source	Translation Model	RNN Encoder–Decoder
at the end of the	[a la fin de la] [f la fin des années] [être supprimés à la fin de la]	[à la fin du] [à la fin des] [à la fin de la]
for the first time	[r © pour la premirëre fois] [été donnés pour la première fois] [été commémorée pour la première fois]	[pour la première fois] [pour la première fois ,] [pour la première fois que]
in the United States and	[? aux ?tats-Unis et] [été ouvertes aux États-Unis et] [été constatées aux États-Unis et]	[aux Etats-Unis et] [des Etats-Unis et] [des États-Unis et]
, as well as	[?s , qu'] [?s , ainsi que] [?re aussi bien que]	[, ainsi qu'] [, ainsi que] [, ainsi que les]
one of the most	[?t ?l' un des plus] [?l' un des plus] [être retenue comme un de ses plus]	[l' un des] [le] [un des]

Recursive Neural network

- Recursive neural networks generalize recurrent neural networks from chains to trees.
- Weight sharing allows trees of different sizes to fit variable length data.
- What structure should the tree follow?



Example: Semantic Parsing

- Use a parse tree or dependency graph as the structure of the recursive neural network
- Example:

Long Short Term Memory (LSTM)

- Special gated structure to control memorization and forgetting in RNNs
- Mitigate gradient vanishing
- Facilitate long term memory