Recurrent Neural Networks (RNN)

Artificial Intelligence @ Allegheny College

Janyl Jumadinova

November 20, 2018

Alex Graves, “Supervised Sequence Labelling with Recurrent Neural Networks”
http://colah.github.io/posts/2015-08-Understanding-LSTMs/
**Word2Vec Model**

- **Word2Vec** is used to learn vector representations of words, “word embeddings”.
- This is typically a preprocessing step, where the learned vectors are fed into a discriminative model (such as RNN).
- Word2vec is a computationally-efficient predictive model for learning word embeddings from raw text.
Word2Vec Model

- **Word2Vec** is used to learn vector representations of words, “word embeddings”.
- This is typically a preprocessing step, where the learned vectors are fed into a discriminative model (such as RNN).
- Word2vec is a computationally-efficient predictive model for learning word embeddings from raw text.
  1. *Continuous Bag-of-Words model (CBOW)*: predicts target words from context words.
  2. *Skip-Gram model*: predicts source context words from target words.
Word2Vec Model

Softmax classifier

Hidden layer

Projection layer

\[ \sum g(\text{embeddings}) \]

\text{the}  \hspace{0.2cm} \text{cat}  \hspace{0.2cm} \text{sits}  \hspace{0.2cm} \text{on}  \hspace{0.2cm} \text{the}  \hspace{0.2cm} \text{mat} \\
context/history \text{ } h \hspace{1cm} \text{target } w_t

https://www.tensorflow.org/tutorials/representation/word2vec
Word2Vec Model

https://www.tensorflow.org/tutorials/representation/word2vec
Recurrent Neural Networks

The output depends ONLY on the current input.
Recurrent Neural Networks (RNN)

The hidden layers and the output depend from previous states of the hidden layers.
Recurrent Neural Networks

An unrolled recurrent neural network.
Recurrent Neural Networks

Each node represents a layer of neurons at a single timestep.
Recurrent Neural Networks

The input is a **SEQUENCE** $x(t)$ of any length.
Common visual sequences:

- Still image
- Spatial scan (zigzag, snake)

The input is a **SEQUENCE** \( x(t) \) of any length.
Recurrent Neural Networks

Must learn temporally shared weights w2; in addition to w1 & w3.
Recurrent Neural Networks

one to one

one to many

many to one

many to many

many to many
Long Short-Term Memory (LSTM)

Based on a standard RNN whose neuron activates with tanh

Cristopher Olah, “Understanding LSTM Networks” (2015)
Long Short-Term Memory (LSTM)

The repeating module in an LSTM contains four interacting layers.
Long Short-Term Memory (LSTM)

- Each line carries an entire vector from the output of one node to the inputs of others.
- Pointwise operations are operations such as vector addition.
- Yellow boxes are learned neural network layers.
- A “Copy” line denotes its content being copied and the copies going to different locations.
Long Short-Term Memory (LSTM)

The **cell state** runs through the entire chain, with only some minor linear interactions.
Long Short-Term Memory (LSTM)

The gate structures allow to remove or add information to the cell state.
P(next words | a history of previous words)

TensorFlow Recurrent Neural Networks

```python
python ptb_word_lm.py --data_path=data/ --model small
```