Summary of paper "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation"

RNN Encoder-Decoder.md

Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation

Introduction

- The paper proposes a new RNN Encoder-Decoder architecture that can improve the performance of statistical machine translation (SMT) systems.
- Link to the paper

RNN Encoder-Decoder

- Model consists of two RNNs
  - Encoder
    - Learns to encode a variable-length input sequence into a fixed-length vector representation.
  - Decoder
    - Learns to decode a given fixed-length vector representation into a variable-length target sequence.
- Two networks are trained jointly to maximise the conditional probability of the target sequence given the input source sequence.
- Trained model can be used to:
  - generate a target sequence, given an input sequence.
  - score a given pair of input and output sequences.

Hidden Unit that adaptively remembers and forgets.

- Hidden unit updated to have a
  - reset gate that adaptively *drop* any hidden state information that it finds irrelevant.
  - update gate that controls how much information from the previous state to carry over.
- Each hidden unit has separate reset and update gates which improve the memory capacity and makes it easier to train.

Statistical Machine Translation (SMT)

- In the phrase-based SMT framework, the translation model is factorised into the translation probabilities of matching phrases in the source and target sentences.
- RNN Encoder-Decoder can be used to rescore the phrase pairs in the phrase table

Experiments

Details

- 1000 hidden units.
- Activation function in proposed hidden unit - hyperbolic tangent function
- Non-recurrent weights initialized by sampling from an isotropic Gaussian distribution (mean = 0, sd = 0.01)
- Recurrent weights initialized by sampling from white Gaussian distribution and using its left singular vectors.
- Adadelta and SGD

Observations

- Train the model to translate an English phrase to French phrase.
- Using the model to score phrase pairs in the standard phrase-based SMT system improves the translation performance.
- Train a CSLM (Continuous Space Language Model) and compare phrase scores from trained model with those given by CSLM.
- RNN Encoder–Decoder is better at capturing the linguistic regularities in the phrase table.
- RNN Encoder-Decoder learns a continuous space representation for phrases that preserves both the semantic and syntactic structure.

shahbazsyed commented on 23 Mar 2017

Hi,
Thanks for the gist! Can you kindly explain what is the "fixed-length vector representation" of the input sequence which is generated by the encoder? Is it a concatenation of all the word vectors in a given sentence ?

varunjanga commented on 27 Apr 2017

Hi shahbazsyed, if I understand it correctly input sequence can be variable length sequence of words which will be provided as an input to encoder to output fixed-length vector representation which is basically some fixed length(say 128) vector of numbers.