Papers I Read

Refining Source Representations with Relation Networks for Neural Machine Translation

2017 • Relational Network • Representation Learning • AI • NLP • NMT

22 Sep 2017

Introduction

- The paper introduces Relation Network (RN) that refines the encoding representation of the given source document (or sentence).
- This refined source representation can then be used in Neural Machine Translation (NMT) systems to counter the problem of RNNs forgetting old information.
- Link to the paper

Limitations of existing NMT models

- The RNN encoder-decoder architecture is the standard choice for NMT systems. But the RNNs are prone to forgetting old information.
- In NMT models, the attention is modeled in the unit of words while the use of phrases (instead of words) would be a better choice.
- While NMT systems might be able to capture certain relationships between words, they are not explicitly designed to capture such information.

Contributions of the paper

- Learn the relationship between the source words using the context (neighboring words).
- Relation Networks (RNs) build pairwise relations between source words using the representations generated by the RNNs. The RN would sit between

the encoder and the attention layer of the encoder-decoder framework thereby keeping the main architecture unaffected.

Relation Network

- Neural network which is desgined for relational reasoning.
- Given a set of inputs * O = o_1, ..., o_n *, RN is formed as a composition of inputs: RN(O) = f(sum(g(o_i, o_j))), f and g are functions used to learn the relations (feed forward networks)
- *g* learns how the objects are related hence the name "relation".
- **Components**:
  - CNN Layer
    - Extract information from the words surrounding the given word (context).
    - The final output of this layer is the sequence of vectors for different kernel width.

  - Graph Propagation (GP) Layer
    - Connect all the words with each other in the form of a graph.
    - Each output vector from the CNN corresponds to a node in the graph and there is an edge between all possible pair of nodes.
    - The information flows between the nodes of the graph in a message passing sort of fashion (graph propagation) to obtain a new set of vectors for each node.

  - Multi-Layer Perceptron (MLP) Layer
    - The representation from the GP Layer is fed to the MLP layer.
    - The layer uses residual connections from previous layers in form of concatenation.

Datasets

- IWSLT Data - 44K sentences from tourism and travel domain.

- NIST Data - 1M Chinese-English parallel sentence pairs.

Models

- MOSES - Open source translation system - http://www.statmt.org/moses/
- NMT - Attention based NMT
- NMT+ - NMT with improved decoder
- TRANSFORMER - Google's new NMT
- RNMT+ - Relation Network integrated with NMT+

Evaluation Metric

- case-insensitive 4-gram BLEU score

Observations

- As sentences become larger (more than 50 words), RNMT clearly outperforms other baselines.
- Qualitative evaluation shows that RNMT+ model captures the word alignment better than the NMT+ models.
- Similarly, NMT+ system tends to miss some information from the source sentence (more so for longer sentences). While both CNNs and RNNs are weak at capturing long-term dependency, using the relation layer mitigates this issue to some extent.

Related Posts

Hints for Computer System Design 07 Jan 2022

Synthesized Policies for Transfer and Adaptation across Tasks and Environments 29 Mar 2021

Deep Neural Networks for YouTube Recommendations 22 Mar 2021

0 Comments   **papers-I-read**   🔒 **Privacy Policy**   1 **Login** ⌄

♡ **Favorite**   🐦 *Tweet*   f *Share*   **Sort by Best** ⌄

Start the discussion…

LOG IN WITH   **OR SIGN UP WITH DISQUS** ?

                                   | Name |

Be the first to comment.

📧 **Subscribe**   Ⓓ **Add Disqus to your site**Add Disqus**Add**   ⚠ **Do Not Sell My Data**