UNIVERSITÄT PADERBORN

Review on Neural machine translation system jointly trained to align and translate

Author: Youssef Ameur
Contact: ameuryoussef94@gmail.com

UNIVERSITÄT PADERBORN

Overview

- Different approaches to Machine Translation.
  - Classical approaches to machine translation
  - Neural machine translation
- Changes brought to the typical neural machine translation systems
  - Replacing the fixed length vector.
- Translation quality improvement for long sentences
  - Quantitative evaluation
  - The training's setting
  - BLEU-Score comparison
  - Qualitative evaluation
- Conclusion
  - Advantages and disadvantages

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Classical approaches to machine translation**

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- Classical approaches to machine translation
- Statistical machine translation (SMT)

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Classical approaches to machine translation**

- **Statistical machine translation (SMT)**
  - In this paradigm, statistical models are computed for bilingual text corpora, on which a translation for an input text is generated.

  - The translation can be word-, phrase- or syntax- based

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Classical approaches to machine translation**

- **Statistical machine translation (SMT)**
  - In this paradigm, statistical models are computed for bilingual text corpora, on which a translation for an input text is generated.

  - The translation can be word-, phrase- or syntax- based

- **Rule-based machine translation (RBMT)**

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Classical approaches to machine translation**

- **Statistical machine translation (SMT)**
  - In this paradigm, statistical models are computed for bilingual text corpora, on which a translation for an input text is generated.
  - The translation can be word-, phrase- or syntax- based

- **Rule-based machine translation (RBMT)**
  - Linguistic rules are determined for the source and target language (e.g. morphological, syntactic and semantic)
  - Translation from one language to another is generated based on these rules

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Neural machine translation (NMT)**

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Neural machine translation (NMT)**
  - Such translation systems rely purely on one neural network
  - A classic architecture for the neural network is the encoder-decoder design pattern
  - Typically, a source sentence is encoded into a fixed length vector, which in turn is decoded to generate a translation.

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Neural machine translation (NMT)**
  - **Typical Encoder-Decoder architecture:**

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Neural machine translation (NMT)**
  - Typical Encoder-Decoder architecture:
    - Consists of two Recurrent Neural Networks (RNN)
    - Encoder generates vector c.

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Neural machine translation (NMT)**
  - Typical Encoder-Decoder architecture:
    - Consists of two Recurrent Neural Networks (RNN)
    - Encoder generates vector c.
    - Decoder predicts target words based on previous predicted words and c

$$p(y_t|\{y_1,\ldots,y_{t-1}\},c) = g(y_{t-1},s_t,c)$$

UNIVERSITÄT PADERBORN

Different approaches to machine translation

- **Neural machine translation (NMT)**
  - Typical Encoder-Decoder architecture:
    - Consists of two Recurrent Neural Networks (RNN)
    - Encoder generates vector c.
    - Decoder predicts target words based on previous predicted words and c

$$p(y_t|\{y_1,\ldots,y_{t-1}\},c) = g(y_{t-1},s_t,c)$$

  - Problem: Long sentences, long term dependencies.

UNIVERSITÄT PADERBORN

Changes brought to the typical neural machine translation systems

UNIVERSITÄT PADERBORN

Changes brought to the typical neural machine translation systems

- **Replacing the fixed length vector.**
  - Proposed by Cho et al. in the paper "Neural machine translation by jointly learning to align and translate (2014)".

UNIVERSITÄT PADERBORN

Changes brought to the typical neural machine translation systems

- **Replacing the fixed length vector.**
  - Proposed by Cho et al. in the paper "Neural machine translation by jointly learning to align and translate (2014)".
    - Replace the fixed length vector, output by the encoder, by a sequence of hidden states h_i
    - Each hidden state h_i contains information about the whole input sequence, with an emphasis around the i-th word.

UNIVERSITÄT PADERBORN

Changes brought to the typical neural machine translation systems

- **Replacing the fixed length vector.**
  - Proposed Encoder-Decoder architecture:
    - Encoder consists of two Bidirectional recurrent neural networks.
    - A sequence of forward and backward hidden states is then generated.

UNIVERSITÄT PADERBORN

Changes brought to the typical neural machine translation systems

- **Replacing the fixed length vector.**
  - Proposed Encoder-Decoder architecture:
    - Encoder consists of two Bidirectional recurrent neural networks.
    - A sequence of forward and backward hidden states is then generated.
    - Decoder computes a vector c_i based on the output of the encoder and an assigned weight to each state.
    - Then predicts target word based on context vector and previously predicted words.

$$p(y_i|y_1, \ldots, y_{i-1}, x) = g(y_{i-1}, s_i, c_i)$$

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **Quantitative evaluation**

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **Quantitative evaluation**
  - BLEU (Bilingual Evaluation Understudy) is an algorithm for computing a score, that represents the quality of the translation compared to professional human translation.

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **Quantitative evaluation**
  - BLEU (Bilingual Evaluation Understudy) is an algorithm for computing a score, that represents the quality of the translation compared to professional human translation.
  - Although this method has many drawbacks, e.g.:
    - Do not consider meaning
    - Do not directly consider sentence structure
  - Still widely used in comparing performance of different systems on the same task.

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **Training's setting**

[1]https://www.statmt.org/wmt14/translation-task.html

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **Training's setting**
  - Language pair: English-French
  - Data set: Bilingual parallel corpora provided by ACL WMT'14[1]
  - Each model was trained twice:
    - Corpus containing sentences of length up to 30 Words.
    - Corpus containing sentences of length up to 50 Words.

[1]https://www.statmt.org/wmt14/translation-task.html

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **BLEU-Score comparison**

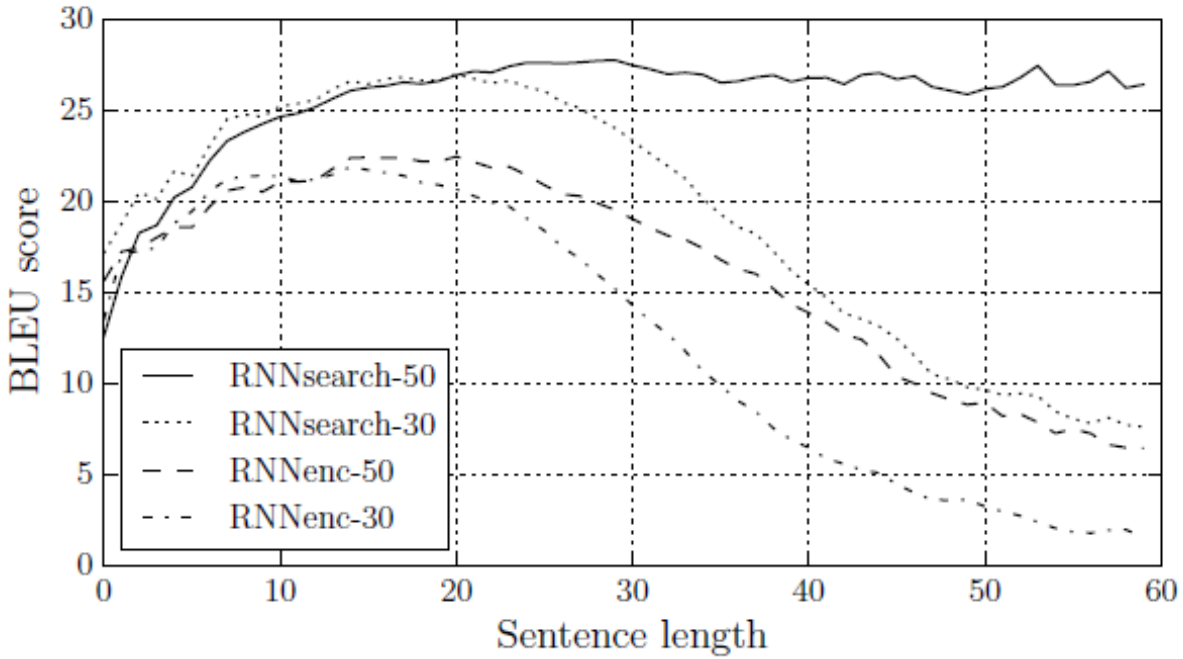

Figure 1: A graph showing performance of each model (in terms of BLEU-Score) with respect to the length's of the sentences.

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **BLEU-Score comparison**

Model	All	No UNK
RNNenc-30	13.93	24.19
RNNsearch-30	21.50	31.44
RNNenc-50	17.82	26.71
RNNsearch-50	26.72	34.16
RNNsearch-50*	28.45	36.15
Moses	33.30	35.63

Table 1: In second column BLEU-Scores, on all sentences, and in the third for sentences that do not contain unknown words.

*Model was trained until no until the performance on the development set stopped improving

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **Qualitative evaluation**
- Source sentence:
  - An admitting privilege is the right of a doctor to admit a patient to a hospital or a medical centre to carry out a diagnosis or a procedure, based on his status as a health care worker at a hospital.

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **Qualitative evaluation**
- Source sentence:
  - An admitting privilege is the right of a doctor to admit a patient to a hospital or a medical centre to carry out a diagnosis or a procedure, based on his status as a health care worker at a hospital.

- RNNenc-50:
  - Un privilége d'admission est le droit d'un médecin de reconnaitre un patient à l'hôpital ou un centre médical d'un diagnostic ou de prendre un diagnostic en fonction de son état de santé. [based on his state of health]

UNIVERSITÄT PADERBORN

Translation quality improvement for long sentences

- **Qualitative evaluation**
- Source sentence:
  - An admitting privilege is the right of a doctor to admit a patient to a hospital or a medical centre to carry out a diagnosis or a procedure, based on his status as a health care worker at a hospital.

- RNNenc-50:
  - Un privilége d'admission est le droit d'un médecin de reconnaitre un patient à l'hôpital ou un centre médical d'un diagnostic ou de prendre un diagnostic en fonction de son état de santé. [based on his state of health]

- RNNsearch-50:
  - Un privilége d'admission est le droit d'un médecin d'admettre un patient à un hôpital ou un centre médical pour effectuer un diagnostic ou une procédure, selon son statut de travailleur des soins de santé à l'hôpital.

UNIVERSITÄT PADERBORN

Conclusion

UNIVERSITÄT PADERBORN

Conclusion

- **Advantages**

- Performance comparable to state-of-the-art translation systems.

- Easier to train, since the system relies on one single neural network.

UNIVERSITÄT PADERBORN

Conclusion

- **Advantages**
- Performance comparable to state-of-the-art translation systems.
- Easier to train, since the system relies on one single neural network.
- **Disadvantages**
- Performance drops when dealing with inputs
  - that are related to specialized domain (e.g. Legal, Finance, etc...).
  - That contain a large number of unknown words.

UNIVERSITÄT PADERBORN

Thank you for your attention

UNIVERSITÄT PADERBORN

References

- Table 1: KyungHyun Cho, Yoshua Bengio, (2014). Neural machine translation by jointly learning to align and translate.

- Figure1: KyungHyun Cho, Yoshua Bengio, (2014). Neural machine translation by jointly learning to align and translate.