Neural Machine Translation: machine translation using large neural network. Phrased-based statistical approaches.

NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE:

* Models proposed recently belong to family of encoder and decoder.
  + Encoder: encoder a source sentence into a fixed length vector, decoder generate a translation
* In this paper they improve the encoder-decoder architecture, allowing a model to automatically search for parts of a source sentence that are relevant to predicting the target word, without form these parts as hard segment explicitly.

On the Properties of Neural Machine Translation: Encoder–Decoder Approaches

* New approach on statistical machine translation based purely on neural networks.
* Encoder extracts a fixed-length representation from a variable-length input.
* Decoder generates a correct translation from this representation.
* The paper focus on analyzing properties of two models: RNN encoder and decoder, newly proposed gated recursive convolutional neural network
* The neural machine translation perform better when the length of the sentence is short and the unknown word is less, but perform went down as the length of the sentence and the unknown word increase
* Proposed gated recursive convolutional network learns a grammatical structure of a sentence automatically

Effective Approaches to Attention-based Neural Machine Translation

* Selectively focusing on parts of the source sentence can improve neural machine translation
* The paper examines two simple and effective classes of attentional mechanism:
  + A global approach of always attends to all source word
  + A local approach of only looks at the subset of source word at a time
* The author use German and English as testing language, both using a different architecture

On Using Very Large Target Vocabulary for Neural Machine Translation

* Neural machine language does better than the tradition phrase based statistical machine translation
* Neural machine language has its limitation in handling large vocabulary, training complexity as well as decoding complexity increase proportionally to the number of targets word
* Author use a method based on importance sampling that allows them use large target word vocabulary without increase training complexity
* This paper try to show that decoding can be done efficiently even the model use a very large target word vocabulary by selecting a small subset of the whole target vocabulary’

Sequence to Sequence Learning with Neural Networks

* Deep Neural Networks work well with large labeled training sets, but they cannot be used to map sequences to sequences
* Using multilayered Long Short-Term Memory (LSTM) map the input sequence to a vector.
* Phrased-Based SMT