Text Simplification System

Tong Wang
Advisor: Prof. Ping Chen
Department of Computer Science
University of Massachusetts Boston
Outline

• Text Simplification Problem
• Text Simplification Systems
• Our work: Text simplification using Neural Machine Translation
Text Simplification

• Input an original English sentence, output a simplified English sentence
• Like Machine Translation, imagine simplified English as a different language
• Work on Sentence and word level, not on long document
Ways to simplify sentence

• Reducing the lexical, grammatical and structural complexity of a text, while still keeping the original meaning
• Word: substitution, paraphrase, delete, add
• Sentence: shorten, split, drop and merge
Lexical Simplification System

• On word level
• Substitute long and infrequent words with shorter and more frequent synonyms
• Candidate selection
  – Semantic similarity
  – Syntax and grammar correct
  – The meaning of the sentence remain the same
  – Word sense disambiguation
• Disadvantage: can not simplify complex syntax
Rule Based Simplification System

• On word and sentence level
• Simplify sentence based on syntactic constructs
• Need to define many rules. E.g,
  – Original sentence: The boy, running down the street, fell down.
  – Simplified sentence: The boy was running down the street. The boy fell down.
• Disadvantage: human—involvement, manually define rules
Phrase based MT system

• Monolingual translation
• Phrase based machine translation system only perform a small set of operations
  – Substitution, paraphrase
• Disadvantage: similar to lexical simplification system
Neural Machine Translation

- Recently proposed (2013)
- Based on the idea of deep learning
- Phrase-based MT: tune separately on many small components
- NMT: Build a single large neural network that reads a sentence and output a correct translation
- Let the neural network learn the lexical simplification and syntactic simplification rules by itself!
Text Simplification using NMT

• Need large amount of training data:
  – Aligned sentence of original sentence and simplified sentence
  – Millions of such pair

• Need good computing resource
  – Gpu computing
Future work

• Collect aligned sentences from Wikipedia and Simple English Wikipedia
• Design recurrent neural network
• Train the neural network
• Build text simplification system based on the neural network
• Thank you!