Research Outline for Text Compression and Summarization

Yang Zhao

Aizawa lab

20180413

Text Compression and Summarization

S: A man suffered a serious head injury after a morning car crash today .
C: A man suffered a injury after a crash .

- What? – Goal and Challenge

Research Challenge and Goal

- **Challenge**: current studies suffer either from error prorogation (tree-based approach) or lower performance (pure data-driven ML approach); Lack of large-scale parallel dataset (even more serious for non-English language)
- **Main goal**: a linguistics-enhanced neural text compressor (or summarizer) for both short (sentence) and long(document) text
    - Subgoal-1: efficiently integrating features in neural text compressor
    - Subgoal-2: optimize new linguistic feature, e.g. (readability and informativeness)
    - Subgoal-3: parallel data creation.

Text Compression and Summarization

- What? – Goal and Challenge

- How? (Methodology)
  - – Linguistic Knowledge (individual word level)
  - – Informativeness (whole sentence level)
  - – Readability (whole sentence level)
  - – Long text compression (Document level)

- Plan
  - – New task ? *(Sep., 2018)*

Approaches

(1) Gating mechanism for word-level linguistic feature fusion.

(2) Optimize sentence-level linguistic features, readability.

(3) Optimize sentence-level linguistic features, informativeness.

(4) Chunk-based attention for long text compression

(5) New approach for new task ()

- Large-scale parallel dataset creation

Thanks