Semester Summarization

Lucas @ April 30th
Efficient Contextualized Representation: Language Model Pruning for Sequence Labeling

• Pre-train language models on extensive corpora & conduct compression w.r.t. a specific task

• Experiments demonstrate the capability of the resulting model to achieve better efficiency without loss of much effectiveness

• rejected by ACL; submitting to EMNLP18
Efficient Contextualized Representation: Language Model Pruning for Sequence Labeling

• Take aways:
  • Pre-trained language models & easy-to-use contextualized representations, which can improve the performance of other NLP tasks

• Ongoing Part:
  • Why it is reasonable to construct contextualized representation via language modeling?
    • The underlying connection between language modeling & skip-thought embedding
  • How can we further improve the performance of the contextualized representation?
    • Conducting experiments
Collaborations

• Learn by Its Definition
  • w. Xiaotao, etc.

• Graph Clustering with Dynamic Embedding
  • w. Carl, etc.

• Contrast Subgraph Mining from Coherent Cores
  • w. Jingbo, etc.

• WSD via Language Modeling and Definition Encoding
  • w. Zihan

• Sequence Labeling w. Sub-word Embedding
  • w. Ahmed

• Multi-tasking Bio-NER
  • w. Shi Zhi
Misc

• New NLP Codebase
  • Sequence Labeling and Language Modeling
    • Easy to read & modify & run
    • Support pytorch 0.4 (soon)

• Summer:
  • Internship @ AI2
Thank You!