Read Papers

Pedro Henrique Luz de Araujo

May 15, 2020

Contents

1 Preliminar papers

1.1 Deep contextualized word representations

- Available at: `https://aclweb.org/anthology/N18-1202` [1]

- Where: Allen Institute for AI

- What: Propose word embeddings (ELMo) that model word uses across different contexts. ELMo embeddings improved SOTA on many NLP tasks.

- How: Pretraining a bi-LSTM Language Model. Word embeddings are linear combinations of the vectors stacked above each input word for each end task. Intrinsic evaluation show that lower-level LSTM states encode aspects of syntax while higher-level ones capture context-dependent word meanings. The model also captures subword information through character convolutions.

- Why: Traditional word embeddings only allow a single context-independent representation for each word.

- Questions and observations:
  - Why only character input? Maybe because no OOV tokens.
  - Has it been compared to ULMFiT?
  - Drastic reduction of number of epochs needed: "For example, the SRL model reaches a maximum development F1 after 486 epochs of training without ELMo. After adding ELMo, the model exceeds the baseline maximum at epoch 10". Also reduces the number of training samples needed.

2   Entity Linking

2.1   Investigating Entity Knowledge in BERT with Simple Neural End-To-End Entity Linking

- Available at: `https://www.aclweb.org/anthology/K19-1063.pdf` [2]

- Where: University of Mannheim, Germany

- What: An end-to-end neural entity linking system that jointly performs entity detection, candidate generation and entity disambiguation.

- How: Training BERT on a automatically generated dataset from Wikipedia where the entities are wikipedia articles and the labels are obtained through wikipedia links.

- Why: Jointly learning all three entity linking steps.

- Questions and observations:
  - Good references for entity linking.
  - 42 days of training with 2 GPUs!

References

[1] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, (New Orleans, Louisiana), pp. 2227–2237, Association for Computational Linguistics, June 2018.

[2] S. Broscheit, "Investigating entity knowledge in BERT with simple neural end-to-end entity linking," in *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, (Hong Kong, China), pp. 677–685, Association for Computational Linguistics, Nov. 2019.