READING GROUP: REPRESENTATION LEARNING FOR NATURAL LANGUAGE UNDERSTANDING

Autumn 2019
University of Helsinki

1 Background

Neural networks have been shown to provide a powerful tool for building representations of natural languages on multiple levels of abstraction. Perhaps the most widely used representations in natural language processing are *word embeddings* (Mikolov et al., 2013; Pennington et al., 2014). There has also been a growing interest in models for sentence-level representations using a range of different neural network architectures. Such *sentence embeddings* have been generated using unsupervised learning approaches (Kiros et al., 2015; Hill et al., 2016), and supervised learning (Bowman et al., 2016; Conneau et al., 2017; Talman et al., 2019). Perhaps the most impactful recent development in NLU has been pre-training of large contextualized word representations using language modeling approaches. These *pre-trained language models* have proven to be extremely powerful tools in NLP, achieving state-of-the art performance in multiple natural language understanding benchmarks. Some of the most prominent pre-trained language models are ELMo (Peters et al., 2018), UMLFit (Howard and Ruder, 2018), BERT (Devlin et al., 2019), GPT (Radford et al., 2018) and GPT2 (Radford et al., 2019).

In this reading group we will read and discuss these three developments in representation learning for natural language understanding.

2 Practical information

- **Meetings:** The reading group will meet once every 2 weeks. *The first meeting will take place during the first week of September (2-6 September).* Time and place will be announced later. There will be at least seven meetings:
  1. Word embeddings 1: LSA, MML, skip-gram & word2vec
  2. Word embeddings 2: GloVe, fastText, cross-lingual embeddings
  3. Sentence embeddings 1: Skip-though, InferSent
  4. Sentence embeddings 2: HBMP, GenSen
  5. Sentence embeddings 2: Attention bridge, cross-lingual embeddings
  6. Pre-trained language models 1: ELMo, BERT
  7. Pre-trained language models 2: GPT, GPT2, Cross-lingual models
- **Reading and preparation:** Each participant is expected to read the assigned papers and be prepared to discuss them in the meetings.
- **Credits: (3cr)** PhD students who want to earn credits for the reading group are expected to write a short critical summary of the readings for each meeting, attend at least 80% of the meetings, actively participate in the discussion and take a lead role in at least one session.
- **Registration:** To sign up for the reading group contact Aarne Talman (aarne.talman@helsinki.fi).

3 Preliminary reading list

Specific readings for each meeting will be announced two weeks before the meeting.

Word Embeddings

- **Latent Semantic Analysis:** Deerwester et al. (1990)
- **Max-margin loss:** Collobert and Weston (2008)
- **Skip-gram and word2vec:** Mikolov et al. (2013)
- **GloVe:** Pennington et al. (2014)
- **fastText:** Bojanowski et al. (2017)
- **Cross-lingual Word Embeddings:** Grave et al. (2018)

Sentence Embeddings

- **Skip-thought:** Kiros et al. (2015)
- **InferSent** Conneau et al. (2017)
- **Iterative Refinement Encoders:** Talman et al. (2019)
- **GenSen:** Subramanian et al. (2018)
- **Attention Bridge:** Raganato et al. (2019)
- **Cross-lingual Sentence Embeddings:** Artetxe and Schwenk (2018)

Pre-trained Language Models

- **ELMo:** Peters et al. (2018)
- **BERT:** Devlin et al. (2019)
- **GPT1:** Radford et al. (2018)
- **GPT2:** Radford et al. (2019)
- **Cross-lingual Language Models:** Lample and Conneau (2019)

References

Artetxe, M. and Schwenk, H. (2018). Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond. *CoRR*, abs/1812.10464.

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Bowman, S. R., Gauthier, J., Rastogi, A., Gupta, R., Manning, C. D., and Potts, C. (2016). A fast unified model for parsing and sentence understanding. In *ACL*.

Collobert, R. and Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *ICML*.

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., and Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. In *EMNLP*.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL*.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., and Mikolov, T. (2018). Learning word vectors for 157 languages. In *LREC*.

Hill, F., Cho, K., and Korhonen, A. (2016). Learning distributed representations of sentences from unlabelled data. In *NAACL*.

Howard, J. and Ruder, S. (2018). Fine-tuned language models for text classification. In *ACL*.

Kiros, R., Zhu, Y., Salakhutdinov, R., Zemel, R. S., Urtasun, R., Torralba, A., and Fidler, S. (2015). Skip-thought vectors. In *NeurIPS*.

Lample, G. and Conneau, A. (2019). Cross-lingual language model pretraining.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *NeurIPS*, USA.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *EMNLP*.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. In *NAACL*.

Radford, A., Narasimha, K., Salimans, T., and Sutskever, I. (2018). Improving language understanding by generative pre-training.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners.

Raganato, A., Vázquez, R., Creutz, M., and Tiedemann, J. (2019). An evaluation of language-agnostic inner-attention-based representations in machine translation. In *Proceedings of the 4th Workshop on Representation Learning for NLP (RepL4NLP-2019)*, pages 27–32, Florence, Italy. Association for Computational Linguistics.

Subramanian, S., Trischler, A., Bengio, Y., and Pal, C. J. (2018). Learning general purpose distributed sentence representations via large scale multi-task learning. In *ICLR*.

Talman, A., Yli-Jyrä, A., and Tiedemann, J. (2019). Sentence embeddings in NLI with iterative refinement encoders. *Natural Language Engineering*, 25(4):467–482.