On Transformers, Distributional Models and Typicality of Argument Structures

Emmanuele Chersoni (The Hong Kong Polytechnic University)

emmanuele.chersoni@polyu.edu.hk

Computational modeling of argument structures has often focused on the concept of *thematic fit*, defined in the psycholinguistic literature as the degree of typicality of arguments for given verb roles (e.g. the typicality of *cop* as the agent of *to arrest*) (McRae et al., 1998; 2005). According to the theory of the Generalized Event Knowledge in sentence comprehension (McRae and Matsuki, 2009; 2013), the thematic fit plays an important role in language processing: human subjects have been shown to read faster sentences where the arguments have a high degree of thematic fit and are mutually typical; moreover, such sentences elicit smaller N400 components (Bicknell et al., 2010; Matsuki et al., 2011). In most cases, researchers in Natural Language Processing used distributional semantic models to build prototypical representations of the ideal fillers of verb roles, and then used those representations to generalize to new arguments via similarity comparisons (Baroni and Lenci, 2010; Lenci, 2011; Sayeed et al., 2016; Santus et al., 2017; Chersoni et al., 2019).

In recent Natural Language Processing research, classical distributional models have gradually been replaced by Transformer-based language models such as BERT (Vaswani et al., 2017; Devlin et al., 2019), which led to improvements of the state-of-the-art performance in several tasks. However, to the best of our knowledge, such models have not been systematically tested yet in thematic fit modeling, nor it is known whether they encode (at least partially) Generalized Event Knowledge.

In this talk, I will discuss some experiments comparing classical distributional models and Transformers on tasks related to the Generalized Event Knowledge, such as selectional preference induction (Metheniti et al., 2020), context-sensitive argument typicality prediction (Lenci, 2011; Chersoni et al., 2019) and logical metonymy interpretation (Rambelli et al., 2020).

References

- Baroni, M., & Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics, 36*(4), 673-721.

- Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language, 63*(4), 489-505.

- Chersoni, E., Santus, E., Pannitto, L., Lenci, A., Blache, P., & Huang, C. R. (2019). A structured distributional model of sentence meaning and processing. *Natural Language Engineering, 25(4)*, 483-502.

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

- Lenci, A. (2011). Composing and updating verb argument expectations: A distributional semantic model. In *Proceedings of the ACL workshop on Cognitive Modeling and Computational Linguistics* (pp. 58-66).

- Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., & McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 913.

- McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition, 33*(7), 1174-1184.

- McRae, K., & Matsuki, K. (2009). People use their knowledge of common events to understand language, and do so as quickly as possible. *Language and linguistics compass, 3*(6), 1417-1429.

- McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. *Sentence processing, 519*, 51-77.

- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*(3), 283-312.

- Metheniti, E., Van de Cruys, T., & Hathout, N. (2020, December). How Relevant Are Selectional Preferences for Transformer-based Language Models?. In *Proceedings COLING* (pp. 1266-1278).

- Rambelli, G., Chersoni, E., Lenci, A., Blache, P., & Huang, C. R. (2020). Comparing Probabilistic, Distributional and Transformer-Based Models on Logical Metonymy Interpretation. In *Proceedings of AACL-IJCNLP*.

- Santus, E., Chersoni, E., Lenci, A., & Blache, P. (2017). Measuring thematic fit with distributional feature overlap. *Proceedings of EMNLP*.

- Sayeed, A., Greenberg, C., & Demberg, V. (2016). Thematic fit evaluation: an aspect of selectional preferences. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP* (pp. 99-105).

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I., (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.