Frame Sense Disambiguation Using Word Embeddings

Lane Lawley
University of Rochester
llawley@cs.rochester.edu

Fillmore's (1982) theory of frame semantics couples the semantics of words in sentences to the structure of the real-world events those sentences describe; for example, to understand the meaning of the word "sold" in the sentence "Mariel sold Gaurav a car", one needs to grasp that a commercial transaction is taking place, which requires significant world knowledge. This world knowledge can be represented as situational "blueprints", or frames, which can be "filled in" with the actors, objects, and relations gleaned from the sentence. These frames can be used to endow parsers with some understanding of the real world, and can be manipulated further for inference once parsed.

Existing frame corpora have been useful in natural language parsing: FrameNet (Baker et al., 1998) and PropBank (Palmer et al., 2005) provide manually-generated corpora of situational frames, and the SEMAFOR (Das et al., 2010) and SLING (Ringgaard et al., 2017) parsers (respectively) have shown reasonable progress in sentence understanding using those corpora. Unfortunately, when several frames are invoked by the same lexical verb, existing frame semantic parsers have trouble disambiguating the sense; "she ran the lab meetings" and "she ran the Boston Marathon", for example, both often invoke the PropBank frame for physical running in the SLING parser.

In this paper, we present a post-processing algorithm for the SLING semantic parser that can correct erroneous frame senses invoked by highly polysemous verbs such as "run" and "play". The algorithm exploits existing frame descriptions in PropBank: every frame sense has an associated description, e.g. "to play an instrument" or "to move quickly from one place to another". By comparing the semantic similarity of the words in the verb arguments to the words in each frame description, the most semantically likely candidate sense can be selected. We estimate semantic similarity using word embeddings generated from large text corpora (Mikolov et al., 2013).

Initial results suggest that this approach works quite well for highly polysemous verbs with detailed sense descriptions. In the future, we plan to apply this algorithm to a more general word sense disambiguation task using WordNet senses instead of PropBank frames.

References

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.

Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A Smith. 2010. Semafor 1.0: A probabilistic frame-semantic parser. *Language Technologies Institute, School of Computer Science, Carnegie Mellon University*.

Charles Fillmore. 1982. Frame semantics. *Linguistics in the morning calm*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Michael Ringgaard, Rahul Gupta, and Fernando CN Pereira. 2017. Sling: A framework for frame semantic parsing. *arXiv preprint arXiv:1710.07032*.