Referentiality in individual named event embeddings

Minnema, Gosse; Herbelot, Aurélie

Referentiality in individual named event embeddings

Gosse Minnema and Aurélie Herbelot
Center for Mind/Brain Sciences
University of Trento, Italy
gosseminnema@gmail.com
aurelie.herbelot@unitn.it

1 Introduction

Distributional models of meaning are known to be good at capturing conceptual information about generic concepts, but it is unclear to what extent they can also capture referential information about individual entities. Events are particularly difficult to model distributionally because of their large diversity in linguistic forms (they could be expressed as verbs, nominalizations, common nouns, or even be completely implicit), and because it is unclear what should serve as the basis of a distributional representation of an individual event: bare verbs, predicate-argument structures, or even whole sentences? Here, inspired by previous work proposing distributional models for *entity-denoting* proper names (e.g., "Angela Merkel", "Barcelona") (Gupta et al., 2015; Herbelot, 2015), we propose using *event-denoting* proper names ("Hurricane Sandy", "Battle of Waterloo", "The Paul McCartney World Tour") as a starting point for investigating individual events.

2 Methods

We investigate two broad classes of models for representing named events distributionally. First, we compute count-based models and use pre-trained skipgram vectors (Mikolov et al., 2013) for Freebase entities[1] for directly representing event names. However, due to the sparsity of frequently-occurring event names, we also use paragraph embeddings of event descriptions from Wikipedia as a way of approximating event name embeddings, following studies showing that definition embeddings can be successfully used as proxies for representations of low-frequency words (Herbelot and Baroni, 2017; Lazaridou et al., 2017). We experiment with paragraph embeddings computed using the summing method (Mitchell and Lapata, 2008), as well as with BERT-derived embeddings (Devlin et al., 2018).

To test what our distributional models learn about the individual events, we use the embeddings as the inputs to simple classification models that to predict referential attributes of the events. Attributes are derived from information found in Wikipedia infoboxes, and are defined for specific event categories. For example, for hurricane events, we predict the geographical location (classes are earth quadrants: 'north-west', 'south-east', etc.), hurricane category (seven levels on the Saffir-Simpson scale), and several numerical attributes such as year, maximal wind speeds, and the number of victims (divided into four equal-sized classes). Additionally, we perform a qualitative analysis of the event space.

3 Results & discussion

We show that, at least on a coarse-grained level, key attributes such as time and location can be predicted with high accuracy by simple models, even when trained on small data. Accuracy patterns are similar for name embeddings and description embeddings, although models trained on name embeddings generally perform worse because of the data scarcity problem. We also find that for event descriptions, summed embeddings perform similarly well as BERT-derived ones, and moreover fail to outperform a simple bag-of-N-grams baseline model on most classification tasks. On the other hand, Freebase skipgram vectors do outperform the bag-of-N-grams baseline when comparing embeddings for the same set of events. We hypothesize that our models largely rely on simple cues such as the presence or absence of particular context words, encoded implicitly or explicitly in the distributional representations.

[1] See https://code.google.com/archive/p/word2vec/

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805.

Abhijeet Gupta, Gemma Boleda, Marco Baroni, and Sebastian Padó. 2015. Distributional vectors encode referential attributes. In *Proceedings of the 2015 Conference on Emperical Methods in Natural Language Processing*, pages 12–21.

Aurélie Herbelot. 2015. Mr Darcy and Mr Toad, gentlemen: distributional names and their kinds. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 151–161.

Aurélie Herbelot and Marco Baroni. 2017. High-risk learning: acquiring new word vectors from tiny data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 304–309.

Angeliki Lazaridou, Marco Marelli, and Marco Baroni. 2017. Multimodal word meaning induction from minimal exposure to natural text. *Cognitive Science*, 41:677–705.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781. `http://arxiv.org/abs/1301.3781`.

Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. In *Proceedings of ACL-08: HLT*, pages 236–244.