Referentiality in distributional representations of named events

Gosse Minnema | MSc Thesis Summary | June 2019

1. Introduction

Events are an important part of natural language meaning, but pose a challenge to distributional semantics, a popular approach in computational linguistics that models the meanings of words and sentences based on their co-occurrence contexts in large corpora. My thesis investigates named events (e.g., 'Hurricane Sandy', 'Battle of Waterloo') and proposes distributional representations derived from encyclopedic definitions of these events, as well as from the textual contexts of the event names themselves. The thesis has both a theoretical and a practical goal: I aimed to find out how the way we talk about events reflect how events are in the real world (within the framework of formal distributional semantics, cf. Boleda & Herbelot 2015), and what the best method is to model events distributionally.

2. Data & representations

We construct a new dataset, derived from Wikipedia, containing event descriptions and attribute-value pairs for three types of named events: hurricanes, concert tours, and battles. Event description representations were created both using traditional additive methods for vector composition (Mitchell & Lapata, 2008) and using contextual sentence embeddings obtained from BERT (Devlin et al. 2018), a new but still poorly understood language representation model. Event name representations were created using simple count-based distributional models.

3. Experiments

Our representations were evaluated by using them as input for classification models (linear SVM, simple neural nets) that predicted the attribute values of each event. This was inspired by Gupta et al. (2015), who predicted various statistics (e.g. GDP, population size) from distributional vectors for country and city names. Figure 1 shows a selection of our results on the prediction experiment.
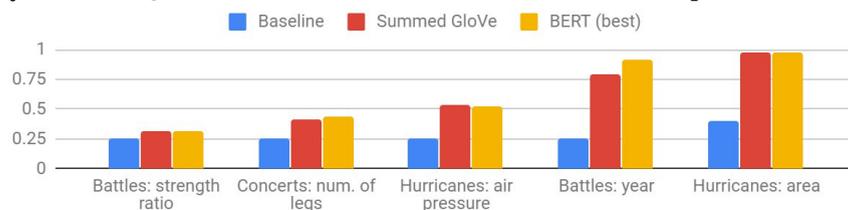

Figure 1: prediction accuracy of description representations for a sample of attributes

I also performed a qualitative analysis of the event vector spaces, and investigated where events are in space relative to each other and to what extent the spaces' dimensions are interpretable.

4. Conclusions

I found that referential information can be successfully extracted from distributional event representations. Interestingly, very different types of representations (e.g. count-based vectors for names, BERT vectors for descriptions) yield similar patterns of results. There is much variation between attributes, but time and place attributes are predicted most easily. My qualitative analysis suggests that this might be at least in part be due to cues from country names and year numbers.

References. Devlin et al. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805.* | Boleda & Herbelot (2017). Formal Distributional Semantics: Introduction to the Special Issue. *Computational Linguistics 42(4).* | Gupta et al. (2015). Distributional vectors encode referential attributes. *EMNLP.* | Mitchell & Lapata (2008). Vector-based models of semantic composition. *ACL.*