Mathematical structures for word embeddings

Siddharth Bhat

IIIT Hyderabad

October 23th, 2021

What is a word embedding?

What is a word embedding?

- Map words to *mathematical objects*.

What is a word embedding?

- Map words to *mathematical objects*.
- Semantic ideas on words $\simeq$ mathematical operations on these objects.

What is a word embedding?

- Map words to *mathematical objects*.
- Semantic ideas on words $\simeq$ mathematical operations on these objects.
- Most common: *vector embeddings* (`word2vec`)

What is a word embedding?

- Map words to *mathematical objects*.
- Semantic ideas on words $\simeq$ mathematical operations on these objects.
- Most common: *vector embeddings* (`word2vec`)

Is word2vec sensible?

Part I: What's a philosopher to do?

Are sets hiding in word2vec?

What does this buy us anyway?

Take-aways

Pat II: What's a geometer to do?

From vectors to subspaces

A research agenda, and carrying the baton forward

Conclusion

- `word2vec` is performant but poorly understood.
- We extract fuzzy set embeddings from `word2vec`, appeasing Montague!
- We ponder on the geometry of `word2vec`, and indicate potential extensions.
- TL;DR: Mathematical modelling (fuzzy sets, grassmanians) is useful to extend empirical results (`word2vec`)!