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?

- $\texttt{king} - \texttt{man} + \texttt{woman} = \texttt{queen}$. [Analogy]

Is word2vec sensible?

- $\texttt{king} - \texttt{man} + \texttt{woman} = \texttt{queen}$. [Analogy]
- nope! $\textit{normalize}(\texttt{king} - \texttt{man} + \texttt{woman}) = \textit{queen}$
- `word2vec`"vectors" are always normalized!

Is word2vec sensible?

- $\texttt{king} - \texttt{man} + \texttt{woman} = \texttt{queen}$. [Analogy]
- nope! $\textit{normalize}(\texttt{king} - \texttt{man} + \texttt{woman}) = \textit{queen}$
- word2vec"vectors" are always normalized!
- Cannot add, substract, scale them. So in what sense is the embedding "vectorial"?

Is word2vec sensible?

- $\texttt{king} - \texttt{man} + \texttt{woman} = \texttt{queen}$. [Analogy]
- nope! $\mathit{normalize}(\texttt{king} - \texttt{man} + \texttt{woman}) = \mathit{queen}$
- `word2vec`"vectors" are always normalized!
- Cannot add, substract, scale them. So in what sense is the embedding "vectorial"?
- In the sense that we have "vectors" — elements of the space $[-1, 1]^N$ with a normalization condition $(\sum_i x_i^2 = 1)$.

Is word2vec sensible?

- $\texttt{king} - \texttt{man} + \texttt{woman} = \texttt{queen}$. [Analogy]
- nope! $\mathit{normalize}(\texttt{king} - \texttt{man} + \texttt{woman}) = \mathit{queen}$
- `word2vec`"vectors" are always normalized!
- Cannot add, substract, scale them. So in what sense is the embedding "vectorial"?
- In the sense that we have "vectors" — elements of the space $[-1, 1]^N$ with a normalization condition ($\sum_i x_i^2 = 1$).
- Can we ascribe a *different* meaning to these "vectors"?

Part I: What's a philosopher to do?

- Montague semantics: The *meaning* of a word is the *set* of possible worlds where the meaning holds true.

Part I: What's a philosopher to do?

- Montague semantics: The *meaning* of a word is the *set* of possible worlds where the meaning holds true.
- A mathematical analogy: The *meaning* of an expression $\forall x \in \mathbb{Z}, x \leqslant 2$ is the *set* of possible values where the meaning holds true: $(-\infty, 2] = \{x \in \mathbb{Z} : x \leqslant 2\}$.

Part I: What's a philosopher to do?

- Montague semantics: The *meaning* of a word is the *set* of possible worlds where the meaning holds true.
- A mathematical analogy: The *meaning* of an expression $\forall x \in \mathbb{Z}, x \leqslant 2$ is the *set* of possible values where the meaning holds true: $(-\infty, 2] = \{x \in \mathbb{Z} : x \leqslant 2\}$.
- Meaning $\simeq$ subsets. Is `word2vecsubsets`?

Part I: What's a philosopher to do?

- Montague semantics: The *meaning* of a word is the *set* of possible worlds where the meaning holds true.
- A mathematical analogy: The *meaning* of an expression $\forall x \in \mathbb{Z}, x \leqslant 2$ is the *set* of possible values where the meaning holds true: $(-\infty, 2] = \{x \in \mathbb{Z} : x \leqslant 2\}$.
- Meaning $\simeq$ subsets. Is `word2vec`subsets? Yes, *fuzzy sets*.

Part I: What's a philosopher to do?

- Montague semantics: The *meaning* of a word is the *set* of possible worlds where the meaning holds true.
- A mathematical analogy: The *meaning* of an expression $\forall x \in \mathbb{Z}, x \leqslant 2$ is the *set* of possible values where the meaning holds true: $(-\infty, 2] = \{x \in \mathbb{Z} : x \leqslant 2\}$.
- Meaning $\simeq$ subsets. Is `word2vecsubsets`? Yes, *fuzzy sets*.
- Set: binary membership. $(1 \in_? \{1, 2\} = T$, $3 \notin_? \{1, 2\} = F)$.

Part I: What's a philosopher to do?

- Montague semantics: The *meaning* of a word is the *set* of possible worlds where the meaning holds true.
- A mathematical analogy: The *meaning* of an expression $\forall x \in \mathbb{Z}, x \leqslant 2$ is the *set* of possible values where the meaning holds true: $(-\infty, 2] = \{x \in \mathbb{Z} : x \leqslant 2\}$.
- Meaning $\simeq$ subsets. Is `word2vec`subsets? Yes, *fuzzy sets*.
- Set: binary membership. ($1 \in_? \{1, 2\} = T$, $3 \notin_? \{1, 2\} = F$).
- Fuzzy set: probabilistic membership. ($1 \in_{fuz} F = 0.1$, $2 \in_{fuz} F = 0.5$).

The hidden sets in `word2vec`

- Given the set of vectors, normalize the components of the vector across *all vectors*.

The hidden sets in word2vec

- Given the set of vectors, normalize the components of the vector across *all vectors*.
- $\texttt{fuzembed}_{\texttt{word}}[i] \equiv \texttt{vecembed}_{\texttt{word}}[i] / \sum_{w \in \texttt{CORPUS}} \texttt{vecembed}_{w}[i]$.

The hidden sets in `word2vec`

- Given the set of vectors, normalize the components of the vector across *all vectors*.
- $\texttt{fuzembed}_{\texttt{word}}[i] \equiv \texttt{vecembed}_{\texttt{word}}[i] / \sum_{w \in \texttt{CORPUS}} \texttt{vecembed}_{w}[i]$.
- Fuzzy set embedding from `word2vecembeddings`.

The hidden sets in `word2vec`

- Given the set of vectors, normalize the components of the vector across *all vectors*.
- $\texttt{fuzembed}_{\texttt{word}}[i] \equiv \texttt{vecembed}_{\texttt{word}}[i] / \sum_{w \in \texttt{CORPUS}} \texttt{vecembed}_{w}[i]$.
- Fuzzy set embedding from `word2vecembeddings`.
- Use *fuzzy set operations* for NLP tasks.

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`)!