Distributional Inclusion Hypothesis and Quantifications: Probing for Hypernymy in Functional Distributional Semantics

Chun Hei Lo, Wai Lam, Hong Cheng, and Guy Emerson

香港中文大學
The Chinese University of Hong Kong

UNIVERSITY OF CAMBRIDGE

Summary

Functional Distributional Semantics (FDS)

Framework that learns distributional semantics with truth-conditional interpretations

Distributional Inclusion Hypothesis (DIH)

r_2 is a hypernym of r_1 iff contexts of r_1 occur also with r_2, e.g.,

some dog {*eats*, *runs*}

some animal {*eats*, *runs*, *flies*}

Major Finding

FDS models learn hypernymy when the training corpus follows the DIH

Functional Distributional Semantics (FDS)

Entity Vectors

$$z \in \mathbb{R}^d$$

Truth-Conditional *Semantic Functions*

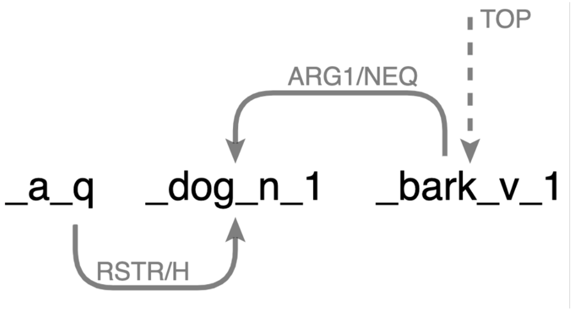
$$t^{(\mathrm{dog})}(z) = P(\,\mathrm{dog}(z) = \top \mid z\,)$$

$$= \mathrm{sigmoid}(v^{\top} z + b)$$

Functional Distributional Semantics (FDS)

Model Training

VAE-like objective on semantic graphs (Lo. et al, 2023)



Variational Inference: z is something that barks $\sim \exists x\text{: bark}(x)$; what is x?

Reconstruction: $t^{(\text{dog})}(z)$ ⬆ and $t^{(\text{ice})}(z)$ ⬇ ~ update $\text{dog}(x) = \top, \text{ice}(x) = \bot$

Trained with less data but competitive with BERT on some lexical semantic tasks!

Representing Hypernymy in FDS

Hypernymy

$$\forall x \in D:\ \mathrm{dog}(x) \Longrightarrow \mathrm{animal}(x)$$

Hypernymy Condition as Fuzzy Set Containment

$$\forall z \text{ s.t. } \|z\|_2 \leq 1: t^{(\mathrm{dog})}(z) < t^{(\mathrm{animal})}(z)$$

Research Questions

Can FDS learn hypernymy from a corpus?

If yes, on what corpus? And how?

Our hypothesis

FDS models learn hypernymy when the training corpus follows the DIH

Intuition behind our Hypothesis

DIH Corpus

some dog barks

some animal barks

some animal flies

Taxonomic Hierarchy

animal {}

dog {bark} *bird {fly}*

$t^{(\mathrm{dog})}(z_1) \uparrow$ and $t^{(\mathrm{animal})}(z_1) \uparrow$, where z_1 describes something that barks

$t^{(\mathrm{animal})}(z_2) \uparrow$, where z_2 describes something that flies

Reverse of DIH (rDIH)

DIH does not hold in general due to *collocational* (Rimell, 2014) and *pragmatic* reasons (Pannitto, 2018). We suggest that *quantifications* are also pivotal!

rDIH when all statements are universally quantified

r_2 is a hypernym of r_1 iff contexts of r_2 occur also with r_1

every dog {*eats*, *breathes*, *barks*}

every animal {*eats*, *breathes*}

Hypernymy Not Respected under rDIH

rDIH Corpus

every dog barks
every dog eats
every animal eats

Taxonomic Hierarchy

animal {eat}
|
dog {bark}

$t^{(\mathrm{dog})}(z_1) \uparrow$ and $t^{(\mathrm{animal})}(z_1) \uparrow$, where z_1 describes something that eats

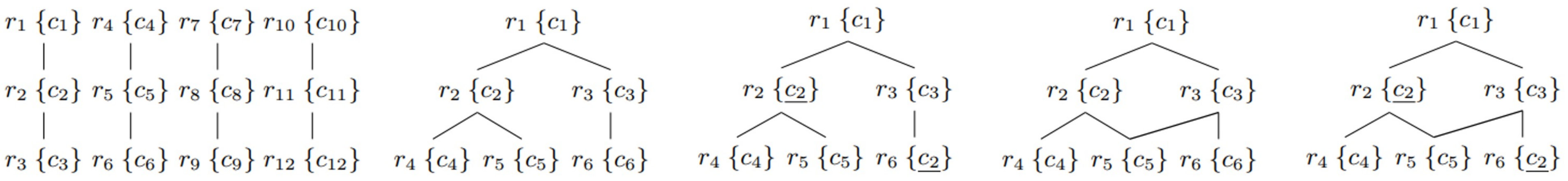
$t^{(\mathrm{dog})}(z_2) \uparrow$, where z_2 describes something that barks

We devised an alternative FDS training objective for $\forall$ ($\mathrm{FDS}_{\forall}$)

Experiments with Synthetic Data Sets

Creation of Each of the Synthetic Data Sets

1. Create a taxonomic hierarchy



(a) Four chains (H_{chains})

(b) Tree (H_{tree})

(c) Tree with overlapping contexts (underlined) (H'_{tree})

(d) DAG (H_{DAG})

(e) DAG with overlapping contexts (underlined) (H'_{DAG})

2. Choose a hypothesis: DIH or rDIH

3. Create a corpus: <quantifier> <noun> <context>

Experiments with Synthetic Data Sets

AUC of hypernymy score by FDS models trained on DIH corpora

Model	H_{chains}	H_{tree}	H'_{tree}	H'_{DAG}	H'_{DAG}
FDS	0.990	0.994	0.995	0.995	0.995
$\text{FDS}_{\forall}$	0.925	0.206	0.210	0.214	0.221

Experiments with Synthetic Data Sets

AUC of hypernymy score by FDS models trained on rDIH corpora

Model	H_{chains}	H_{tree}	H'_{tree}	H'_{DAG}	H'_{DAG}
FDS	0.876	0.842	0.793	0.752	0.688
$\mathrm{FDS}_{\forall}$	0.988	0.983	0.978	0.981	0.977

Experiments with Real Data Sets

AUC of hypernymy score by FDS models trained on Wikiwoods

Model	Kotlerman2010	LEDS	WBLESS	Evalution
FDS	0.473	0.650	0.508	0.459
$FDS_{\forall}$	0.550	0.735	0.655	0.554

More Discussions in the Paper on …

Hypernymy representation in FDS

- *Probabilisitic* vs *fuzzy* interpretation of truth-conditional semantic functions

Distributional generalization of FDS on hypernymy

- *fox* is a *mammal*, not sure if *dog* is
- *fox* and *dog* share the same contexts in corpus
- FDS: *dog* is likely a *mammal*

∀-objective encoding generality more effectively than similarity

- Better at distinguishing between hypernymy and {hyponymy, co-hyponymy}

To Conclude

Question: Can FDS learn hypernymy from a corpus?

Answer: Yes

Question: On what corpus?

Answer: A corpus that follows DIH

Big picture

To acquire faithful truth-conditional representations from distributional information