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

香港中文大學 The Chinese University of Hong Kong

Chun Hei Lo[1] Wai Lam[1] Hong Cheng[1] Guy Emerson[2]

[1]The Chinese University of Hong Kong [2]University of Cambridge

UNIVERSITY OF CAMBRIDGE

Our Hypothesis

Functional Distributional Semantics (FDS) learns hypernymy from corpora that the distributional inclusion hypothesis (DIH) strictly holds.

DIH and Quantifications

DIH. r_2 is a hypernym of r_1 iff r_1's characteristic contexts $\subseteq$ r_2's.

Quantifications. A corpus with only universally quantified statements results in the reverse of DIH (rDIH).

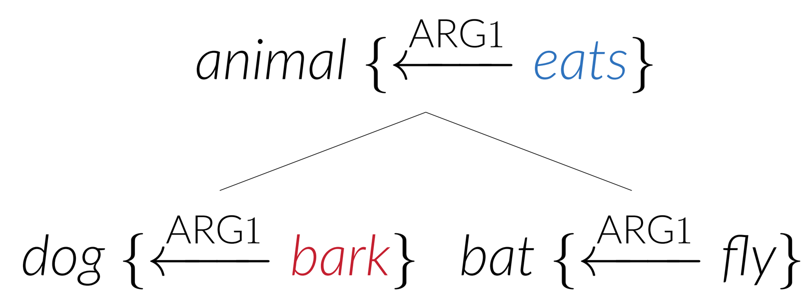

Figure 1. A taxonomic hierarchy of nouns. Next to each noun are the contexts applicable to it and its descendants.

Corpus 1 (DIH)	Corpus 2 (rDIH)
some dog barks	*every dog barks*
some animal barks	*every dog eats*
some bat flies	*every bat flies*
some animal flies	*every bat eats*
some animal eats	*every animal eats*

Table 1. Corpora generated from the hierarchy in Fig. 1.

Experiments on Synthetic Data Sets

Creating r(DIH) Corpora from Taxonomic Hierarchies.

1. *Create a Taxonomic Hierarchy.* E.g.,

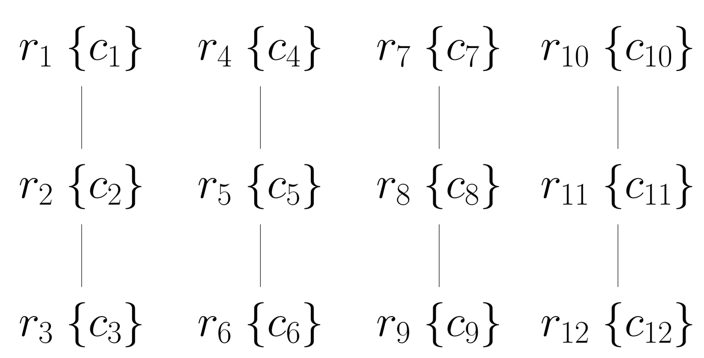

Figure 2. Four chains (H_{chains})

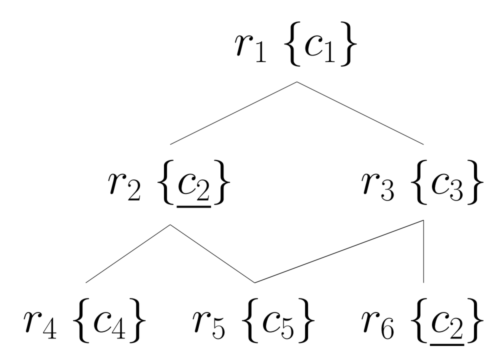

Figure 3. DAG with shared contexts (underlined; $H_{\text{DAG}'}$)

2. *Choose a Hypothesis.* The DIH or rDIH
3. *Create a Corpus.* '[quantifier] [noun] [context word]'

Evaluation on Hypernymy Detection.

Hyp.	Model	H_{chains}	$H_{\text{DAG}'}$
DIH	FDS	.990	.995
	FDS$_\forall$	.925	.221
rDIH	FDS	.876	.688
	FDS$_\forall$	.988	.977

Table 2. AUC of s learnt from different taxonomic hierarachies.

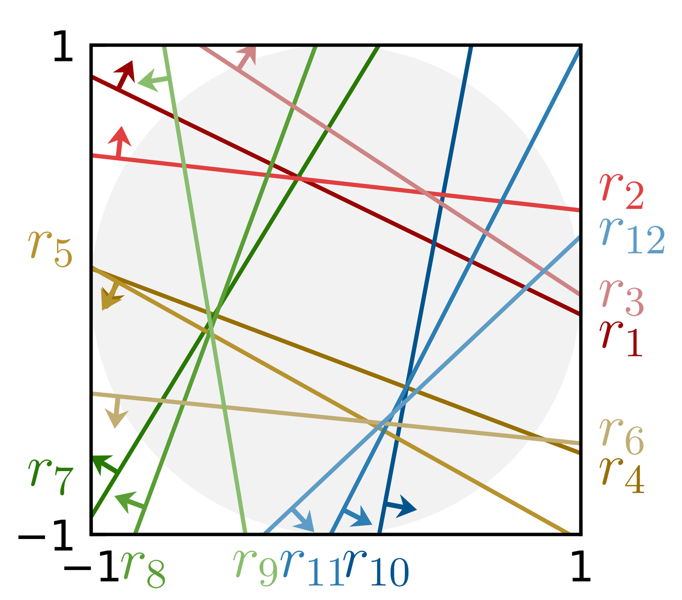

Figure 4. $t^{(r)}(z) \geq 0$ learnt from DIH corpus of H_{chains}.

Takeaway 1. Our hypothesis is confirmed.

Takeaway 2. FDS$_\forall$ can handle rDIH corpora well.

Takeaway 3. Given (1) *fox*'s contexts $\subseteq$ *mammal*'s and (2) *dog* shares many contexts with *fox*, FDS/FDS$_\forall$: *dog* is likely a *mammal*. (Results in the paper!)

FDS

Entity Vectors. $z \in \mathbb{R}^d$

Truth-Conditional Semantic Functions.

$$t^{(dog)}(z) = P\left(dog(z) = \top \mid z\right) = \text{sigmoid}\left({v^{(dog)}}^\top z + b^{(dog)}\right)$$

Representing Hypernymy.

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

which is true iff $s(dog, animal) > 0$, where

$$s(r_h, r_H) = b^{(r_H)} - b^{(r_h)} - \left\| v^{(r_H)} - v^{(r_h)} \right\|_2$$

Model Training. By Lo et al. (2023), given a DMRS graph, e.g.,

$$\textit{some} \xrightarrow{\text{RSTR}} \textit{dog} \xleftarrow{\text{ARG1}} \textit{bark}$$

Variational Inference: $q_\phi(z \mid \xleftarrow{\text{ARG1}} \textit{bark})$

Reconstruction: $\max \ln \mathbb{E}_{q_\phi(z \mid \xleftarrow{\text{ARG1}} \textit{bark})} \left[t^{(dog)}(z)\right] + \ldots$

New objective (FDS$_\forall$). Optimization is performed over regions of the entity vector space for handling universal quantifications.

Experiments on Real Data Sets

Training Data. Wikiwoods: 36m DMRS graphs

Evaluation on Hypernymy Detection.

Model	Kotlerman2010	LEDS	WBLESS	Evalution
Cosine	.701	.782	.620	.526
WeedsPrec	.674	.897	.709	.650
invCL	.679	.905	.707	.620
FDS	.473	.650	.508	.459
FDS$_\forall$	.550	.735	.655	.554

Table 3. AUC of s on the test sets.

Model	Hyponymy	Co-hyponymy	Meronymy	Random
Cosine	.511	.369	.683	.924
WeedsPrec	.754	.615	.631	.843
invCL	.745	.568	.652	.872
FDS	.596	.288	.561	.587
FDS$_\forall$	.783	.625	.527	.691

Table 4. AUC of s on the sub-categories of WBLESS.

Takeaway 1. FDS$_\forall$ is better than FDS on hypernymy detection from real corpora.

Takeaway 2. FDS$_\forall$ learns generality more effectively than similarity.

(Takeaway 3.) Although quantifications are annotated in DMRS graphs of Wikiwoods, processing multiple scopes requires a richer world model than can be encoded in a taxonomic hierarchy.

Big Picture. To acquire more faithful truth-conditional representations of words from distributional information.

aaronlo326@gmail.com