Affordance Extraction and Inference using SRL

EMNLP 2018

Demo: a2avecs.github.io

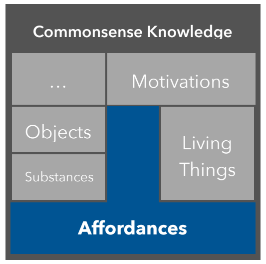
Daniel Loureiro, Alípio Jorge - University of Porto - LIAAD, INESCTEC

dloureiro@fc.up.pt

Affordance: What and why ?

Many definitions, but we go with Glenberg 2000:

- **Action possibilities between things**, directly perceived or otherwise.
- Basis for the **Indexical Hypothesis**, which was proposed as a grounded alternative to the prevalent Distributional Hypothesis.
- Inherently relational and capable of representing **Commonsense** knowledge, missing features from current DSMs [Camacho-Collados and Pilhevar 2018]



How does it relate to Fact Extraction and Verification (FEVER) ?

*hypothetically...

Infer additional Extractions (cf. Frame Semantics)

'Cumberbatch portrayed Turing in The Imitation Game'

- Those who 'portray' are usually actors
- Things 'portrayed' are usually film characters
- Places of 'portrayal' are usually films

Plausibility is easier to assess than Truth*.

- 'A Floppy disk lined with turnips.'
  Implausible - likely no evidence.
- 'Dan Trachtenberg is a person.'
  Plausible and typical - implicit evidence
- 'Sarah Hyland is a New Yorker.'
  Plausible and atypical - req. explicit evidence

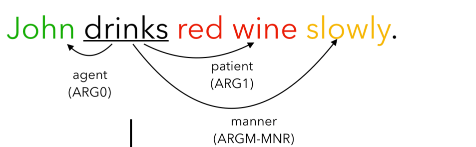
Obvious

Requires Evidence

Nonsense

Affordance Extraction

Extract Predicate-Argument Structures (PASs) from Wikipedia using end-to-end neural Semantic Role Labeling [He et. al 2018]

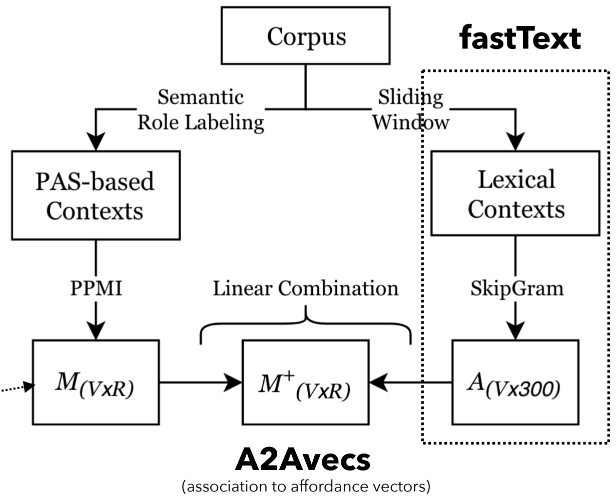


PAS co-occurrence freqs. w/PPMI

High-dimensional sparse matrix.

	drink | ARG0	drink | ARG1	drink | ARGM-MNR	...
John	0.8			
red		0.6		
wine		0.9		
slowly			0.7	
...				

155Kx18K (98.7% sparse after interpolation, from 99.6%)



A2Avecs

(association to affordance vectors)

Leveraging DSMs based on lexical adjacency contexts (A) to **reduce sparsity** through linear combinations, similarly to works in translation.

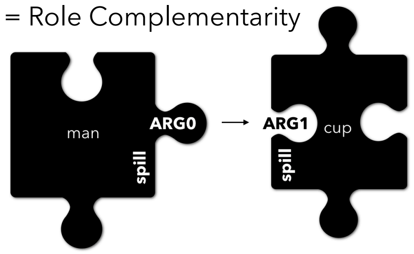
$$\vec{v}_w = \frac{\vec{v}_1 * \alpha_1 + ... + \vec{v}_n * \alpha_n}{\sum_{i=1}^{n} \alpha_i}$$

$$\alpha_i = \begin{cases} \frac{A_{\vec{v}_w} \cdot A_{\vec{v}_i}}{\|A_{\vec{v}_w}\| \|A_{\vec{v}_i}\|} = cos_A(\vec{v}_w, \vec{v}_i), \\ \qquad \text{if } cos_A(\vec{v}_w, \vec{v}_i) > 0.5 \\ 0, \quad \text{otherwise} \end{cases}$$

A not necessarily trained on the same corpus.

Affordance Inference

= Role Complementarity



Word Pairs (w_1, w_2)	Affordances (w_1 as ARG0, w_2 as ARG1)
shop, tea	sell, import, cure
doctor, patient	diagnose, prescribe, treat
newspaper, face	cover, expose, poke
man, cup	drink, pour, spill

Using a very simple method for finding complementary roles we can:

- Discover how different concepts may interact, according to their respective roles (as agent, patient, location, manner, etc.).
- Works for associated or related concepts such as doctor/patient as well as unrelated concepts such as newspaper/face.

Evaluation on Word Similarity Tasks

Word Similarity Tasks are the standard for evaluating word representations.

Context	Model	SL-666	SL-999	WS-SIM	WS-ALL	MEN	RG-65
Lexical	word2vec	.426	.414	.762	.672	.721	.793
	GloVe	.333	.325	.637	.535	.636	.601
	fastText (A)	.426	.419	**.779**	**.702**	**.751**	.799
Syntactic	★ Deps	**.475**	**.446**	.758	.629	.606	.765
	Open IE	.397	.390	.746	.696	.281	.801
	A2Avecs (M^+)	.461	.412	.734	.577	.687	**.802**
	A2Avecs (SVD(M^+))	.436	.386	.672	.509	.599	.789

All trained on English Wikipedia.

Ours (A2Avecs) performs competitively with adjacency-based lexical contexts, but the dependency-based embeddings of Levy and Goldberg 2014 still perform better.

Curiously, applying SVD to reduce our explicit 18k dimensions into the standard 300 latent dimensions hurts performance significantly.

What if we try concatenating our PAS-based vectors with latent embeddings trained on larger corpora (fastText 600B)?

Interestingly, this solution is markedly better, significantly **outperforming fastText 600B** on challenging tasks such as SimLex-999 (specially nouns).

Context	Model	SL-666	SL-999	WS-SIM	WS-ALL	MEN	RG-65
Lexical SOTA	fastText 600B (A)	.523	.504	.839	**.791**	**.836**	**.859**
Intp. w/SOTA	A2Avecs (M^+)	.513	.468	.780	.619	.744	.814
Intp. & Conc.	★A2Avecs ($M^+ \parallel A$)	**.540** ↑	**.521** ↑	**.846** ↑	.771 ↓	.829 ↓	.857 ↔
Deps Conc.	Deps $\parallel$ A	.524 ↔	.503 ↔	.818 ↓	.752 ↓	.770 ↓	.835 ↓

Best result uses PASs extracted from Wikipedia, interpolated (M^+) using fastText 600B, and also concatenated (||) with fastText 600B.

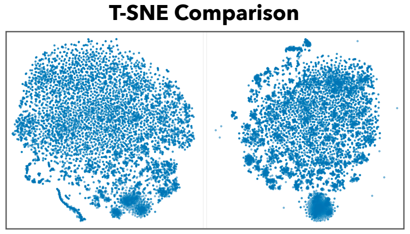
To be rigorous, we concatenated the same latent embeddings to the dependency-based embeddings, and found that this combination wasn't beneficial.

Introspection

batman	**hogwarts**	**turing**
foil|ARG0	ambush|ARGM-MNR	travel|ARGM-TMP
flirt|ARGM-MNR	rock|ARGM-LOC	pass|ARGM-ADV
apprehend|ARG0	express|ARG0	solve|ARG0
subdue|ARG0	prevent|ARGM-LOC	simulate|ARG1
rescue|ARGM-DIR	expel|ARG2	prove|ARG1
florida	**object-oriented**	**dancing**
base|ARGM-MNR	define|ARG1	dance|ARG0
vacation|ARG1	define|ARG0	dance|ARGM-MNR
reside|ARGM-DIS	use|ARG1	dance|ARGM-LOC
fort|ARG1	implement|ARG1	dance|ARGM-ADV
vacation|ARGM-LOC	express|ARGM-MNR	dance|ARG1

Top Affordances for the same word set as Levy and Goldberg 2014a

T-SNE Comparison



fastText 600B

a2avecs (more structure)

Conclusions

- SRL-based contexts **complementary to adjacency contexts** for word representations (word2vec, GloVe, fastText, etc.).
- Provided a solution for performing **Relational Inferences** in Distributional Semantic Models.
- Provided the components for **assessing Semantic Plausibility**, which should be useful for Fact Verification.

Arthur M. Glenberg. 2000. Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning.

José Camacho-Collados and Mohammad Taher Pilehvar. 2018. From word to sense embeddings: A survey on vector representations of meaning.

Luheng He, Kenton Lee, Mike Lewis, and Luke S. Zettlemoyer. 2017. Deep semantic role labeling: What works and what's next

Omer Levy and Yoav Goldberg. 2014. Dependency-based word embeddings