PISA: A Measure of Preference in Selection of Arguments to Model Verb Argument Recoverability

Giulia Cappelli [1], Alessandro Lenci [2]

[1] Scuola Normale Superiore
[2] University of Pisa

CLiC-it 2020
The Seventh Italian Conference on Computational Linguistics
March 1-3, 2021
Virtual Meeting

Table of contents

Introduction

Goal of the study

A fully distributional model of the semantic recoverability of verb arguments, to improve on taxonomy-based models[1]

[1] Resnik 1993, 1996

The verb-argument relation

What kind of verb arguments are we interested in?

The verb-argument relation can be

- a grammatical function, such as "subject" or "direct object"[2]
- a semantic role, such as "Instrument" or "Patient"

The choice between the two depends on computational requirements rather than on theoretical constraints.

[2] Resnik 1993, 1996.

Recoverability of direct objects (arguments)

(1) John **ate** $\varnothing_{\text{object}}$.

recoverable object: belongs to the category of Edibles (grammatical sentence)

(2) *John **made** $\varnothing_{\text{object}}$.

non-recoverable object: basically anything can be made! (ungrammatical sentence)

Recoverability of Instruments[3] (adjuncts)

(3) John **beheaded** the prisoner $\varnothing_{\text{Instrument}}$.

recoverable Instrument: a heavy-bladed tool, possibly a sword (Require-Instrument verb)

(4) John **killed** the prisoner $\varnothing_{\text{Instrument}}$.

non-recoverable Instrument: a weapon? poison? bare hands? (Allow-Instrument verb)

---

[3] Koenig, Mauner, and Bienvenue 2002, 2003; Koenig, Mauner, Bienvenue, and Conklin 2007.

Resnik's taxonomy-based measure[5]

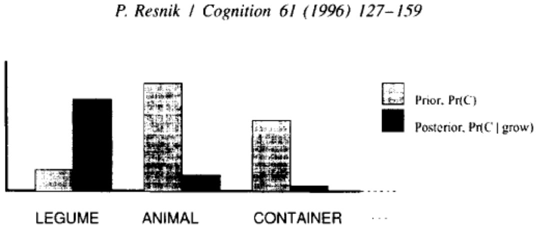


Fig. 1. Example of a prior distribution and a posterior distribution.

Resnik's starting point is that the distribution of entities[4] used as direct object in a corpus with *any* verb (light bars) is different from their distribution with *a specific* verb (dark bars)

[4]belonging to ontological classes, specifically WordNet synsets
[5]Resnik 1993, 1996.

Resnik's Selectional Preference Strength

Resnik's Selectional Preference Strength (SPS) of a verb with respect to the possible fillers in the given relation[6] is the Kullback-Leibler divergence (relative entropy) between:

- the (posterior) distribution of WordNet synsets for the given verb–relation pair
- the (prior) distribution of synsets participating in the given relation over all verbs in the corpus

$$SPS_{v,r} = \sum_{c \in classes} p(c|v,r) \, log \frac{p(c|v,r)}{p(c|r)} \quad (1)$$

[6] which can be used as a measure of argument recoverability

Distributional Semantic Models (DSMs)[7]

Taxonomy-based models need a manually-built lexicon

[7] Lenci 2018.

Distributional Semantic Models (DSMs)[7]

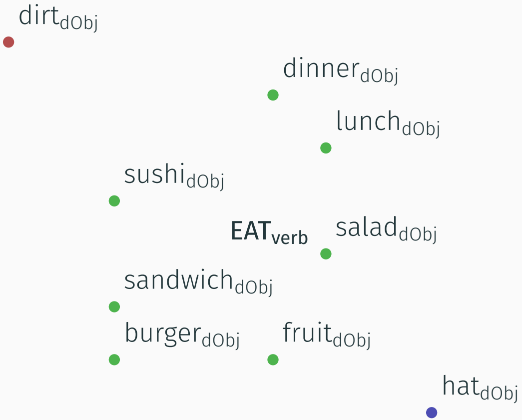
Taxonomy-based models need a manually-built lexicon

DSMs don't! Several models to compute the Selectional Association (SA) between an argument and a verb-relation pair (Pereira, Tishby, and Lee 1993; Erk 2007; Bergsma, Lin, and Goebel 2008; Schulte im Walde et al. 2008; Erk, S. Padó, and U. Padó 2010)

[7] Lenci 2018.

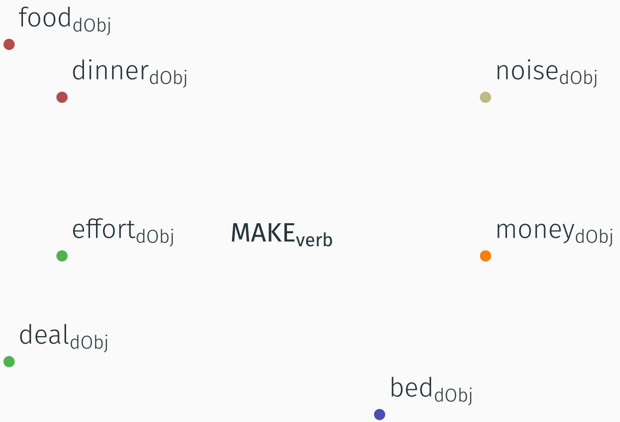
PISA

The basic idea



the dObjs of *to eat* are close together in a vector space

The basic idea



the dObjs of *to make* are very sparse in a vector space

PISA: a model of Preference In Selection of Arguments

Intuition: the vector-based SPS of a given verb-relation pair should be positively correlated with the semantic density of their arguments

PISA: a model of Preference In Selection of Arguments

Intuition: the vector-based SPS of a given verb-relation pair should be positively correlated with the semantic density of their arguments

Implementation: computing the semantic density of the verb-relation pair as the mean pairwise cosine similarity between the arguments of the pair, following Erk 2007; Erk, S. Padó, and U. Padó 2010

The measure

We average Erk's SA (in 2) over the n arguments of a given verb-relation pair to compute PISA (in 3):

$$SA_{v,r}(a_0) = \sum_{a \in args(v,r)} wt_{v,r}(a)\, sim(a_0, a) \tag{2}$$

$$PISA_{v,r} = \frac{1}{n} \sum_{i=1}^{n} SA_{v,r}(a_i) \tag{3}$$

Weighted models

Some arguments are more associated with a given verb-relation pair than others (e.g. *hamburger* is a more typical dObj of *to eat* than *artichoke*) $\longrightarrow$ they get different weights[8]

[8] Erk 2007; Erk, S. Padó, and U. Padó 2010.

Weighted models

Some arguments are more associated with a given verb-relation pair than others (e.g. *hamburger* is a more typical dObj of *to eat* than *artichoke*) $\longrightarrow$ they get different weights[8]

- UNI assumes a uniform distribution: $wt_{v,r}(a) = 1$
- FRQ is the co-occurrence frequency of a given argument with the verb-relation pair: $wt_{v,r}(a) = freq(a, v, r)$
- IDF assigns higher scores to arguments occurring with fewer verb-relation pairs: $wt_{v,r}(a) = \log \frac{|v,r|}{|v,r : a \in v,r|}$
- LMI is the Local Mutual Information of the argument and a given verb-relation pair: $wt_{v,r}(a) = f(a, v, r) \log_2 \frac{p(a,v,r)}{p(a)p(v,r)}$
- ENT is the entropy of the argument of a given verb-relation pair: $wt_{v,r}(a) = -\sum_{a \in args(v,r)} p(a) \log_2 p(a)$

[8] Erk 2007; Erk, S. Padó, and U. Padó 2010.

Unweighted models

Weighted models include ALL the arguments in the calculation (can be a very large number!)

Unweighted models

Weighted models include ALL the arguments in the calculation (can be a very large number!)

Is it possible to obtain relevant information considering the most relevant k arguments only?

Unweighted models

Weighted models include ALL the arguments in the calculation (can be a very large number!)

Is it possible to obtain relevant information considering the most relevant k arguments only?

We created unweighted models

- with only the top/bottom k argument nouns for each verb-relation pair (300 dObjs, 20 Instruments)
- arguments are sorted based on the FRQ, IDF, LMI and ENT weighting functions

Experiment

Datasets

99 transitive verbs (50 recoverable dObj+ 49 non-recov dObj)
34 from Resnik 1993, 35 from Levin 1993, 30 high-frequency verbs

Datasets

99 transitive verbs (50 recoverable dObj+ 49 non-recov dObj)
34 from Resnik 1993, 35 from Levin 1993, 30 high-frequency verbs

173 Instrument verbs (116 recoverable Instr + 57 non-recov Instr)
taken from Koenig, Mauner, Bienvenue, and Conklin 2007

Datasets

99 transitive verbs (50 recoverable dObj+ 49 non-recov dObj)
34 from Resnik 1993, 35 from Levin 1993, 30 high-frequency verbs

173 Instrument verbs (116 recoverable Instr + 57 non-recov Instr)
taken from Koenig, Mauner, Bienvenue, and Conklin 2007

The datasets and the scripts we used to run our model are freely available here on GitHub (courtesy of Ludovica Pannitto)

Extraction of verb arguments

Arguments of verbs extracted from ukWaC[9], a 2-billion token part-of-speech tagged and lemmatized corpus of English

extraction of head nouns without determiners and modifiers:

(5) a. ~~a big rusty sword~~
b. sword

Instruments = PPs headed by *with*, Artifact as a noun argument[10]

[9] Ferraresi et al. 2008.
[10] As defined in WordNet 3.0 (Miller 1995)

Word embeddings

300-dimensional embeddings trained on a concatenation of ukWaC and a 2018-dump of English Wikipedia

Both window-based and syntax-based contexts, different window sizes (2 or 10) for both SVD reduced count-based DSMs and neural embeddings created via word2vec

SVD	w2v	w2vf
synt.c1000	CBOW.w10	SGNS.synt.c1000
synt.c500	CBOW.w2	SGNS.synt.c500
w10	SGNS.w10	SGNS.w10
w2	SGNS.w2	SGNS.w2

Table 1: Tested embedding types (w2v = word2vec; w2vf = word2vecf).

Results

Results: Resnik's SPS

Resnik's SPS scores higher for recoverable-argument verbs than for non-recoverable argument verbs? (Mann-Whitney U tests)

	mean recov	mean non-recov
dObj verbs	4.27	1.89
Instr verbs	4.72	3.60

	stats
dObj verbs	$U = 264, n_1 = 50, n_2 = 49, P < .001$
Instr verbs	$U = 4646, n_1 = 116, n_2 = 57, P < .001$

Does PISA work?

		weighted	top *k*	bot *k*
UNI	SVD	***	-	-
	w2v	***	-	-
	w2vf	** (***)	-	-
FRQ	SVD	***	** (***)	ns
	w2v	***	***	ns
	w2vf	***	** (***)	ns
IDF	SVD	***	** (ns)	ns (***)
	w2v	***	*** (ns)	***
	w2vf	** (***)	ns	ns
LMI	SVD	*** (**)	** (ns)	ns (**)
	w2v	***	* (ns)	*
	w2vf	*** (*)	* (ns)	* (**)
ENT	SVD	*** (*)	ns (***)	ns
	w2v	*** (**)	***	ns
	w2vf	*** (**)	* (ns)	*

Mann-Whitney U p-values (recov vs non-recov verbs)
Whenever transitive and Instrument-verb results are different, the former are on the left and the latter on the right of the same cell

How does PISA compare with Resnik's SPS? (dObj)

		weighted	top300	bot300
UNI	SVD	.832***	-	-
	w2v	.851***	-	-
	w2vf	.250*	-	-
FRQ	SVD	.854***	.341***	-.041 ns
	w2v	.835***	.712***	-.024 ns
	w2vf	.743***	-.368***	-.090 ns
IDF	SVD	.750***	-.328***	.211 ns
	w2v	.818***	-.388***	.457***
	w2vf	.256*	-.154 ns	.164 ns
LMI	SVD	.791***	-.385***	-.092 ns
	w2v	.711***	-.135 ns	.129 ns
	w2vf	.667***	-.092 ns	.091 ns
ENT	SVD	-.905***	.163 ns	.111 ns
	w2v	-.908***	.579***	.134 ns
	w2vf	-.911***	.254*	.320**

Spearman correlations between PISA and Resnik scores for transitive verbs.

How does PISA compare with Resnik's SPS? (Instr)

		weighted	top20	bot20
UNI	SVD	.404***	-	-
	w2v	.244***	-	-
	w2vf	.105 ns	-	-
FRQ	SVD	.283***	.481***	-.025 ns
	w2v	.179*	.519***	-.005 ns
	w2vf	.127 ns	.326***	.037 ns
IDF	SVD	.384***	.005 ns	.135 ns
	w2v	.242***	.09 ns	.265***
	w2vf	.082 ns	.176*	.03 ns
LMI	SVD	.170*	.152*	-.011 ns
	w2v	.134 ns	.134 ns	-.065 ns
	w2vf	.077 ns	.266***	-.013 ns
ENT	SVD	-.885***	.118 ns	.003 ns
	w2v	-.920***	.256***	.088 ns
	w2vf	-.928***	.031 ns	.334***

Spearman correlations between PISA and Resnik scores for Instrument verbs.

Conclusions

Conclusions

PISA is as good as SPS but computationally cheaper (no WordNet required!)

Conclusions

PISA is as good as SPS but computationally cheaper (no WordNet required!)

Which weight is the best? UNI is easier (no weight, no *k* value), ENT is more conservative wrt Resnik, sorted FRQ is best for very large sets of verbs

Conclusions

PISA is as good as SPS but computationally cheaper (no WordNet required!)

Which weight is the best? UNI is easier (no weight, no *k* value), ENT is more conservative wrt Resnik, sorted FRQ is best for very large sets of verbs

future studies will predict the recoverability of arguments in other syntactic or semantic relations

References

Levin, Beth (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press. 348 pp. ISBN: 978-0-226-47532-5 978-0-226-47533-2.

Pereira, Fernando, Naftali Tishby, and Lillian Lee (1993). "Distributional Clustering of English Words". In: *Proceedings of the 31st Annual Meeting on Association for Computational Linguistics -*. The 31st Annual Meeting. Columbus, Ohio: Association for Computational Linguistics, pp. 183–190. DOI: 10.3115/981574.981598. URL: http://portal.acm.org/citation.cfm?doid=981574.981598 (visited on 2020).

Resnik, Philip (1993). *Selection and Information: A Class-Based Approach to Lexical Relationships*. IRCS Technical Reports Series. University of Pennsylvania. 177 pp. URL: https://repository.upenn.edu/ircs_reports/200/.

Miller, George A. (Nov. 1995). "WordNet: A Lexical Database for English". In: *Communications of the ACM* 38.11, pp. 39–41. ISSN: 0001-0782. DOI: 10.1145/219717.219748.

Resnik, Philip (1996). "Selectional Constraints: An Information-Theoretic Model and Its Computational Realization". In: *Cognition* 61.1-2, pp. 127–159. ISSN: 00100277. DOI: 10.1016/S0010-0277(96)00722-6. URL: https://linkinghub.elsevier.com/retrieve/pii/S0010027796007226 (visited on 2020).

Koenig, Jean-Pierre, Gail Mauner, and Breton Bienvenue (2002). "Class Specificity and the Lexical Encoding of Participant Information". In: *Brain and Language* 81.1-3, pp. 224–235. ISSN: 0093934X. DOI: 10.1006/brln.2001.2519. URL: https://linkinghub.elsevier.com/retrieve/pii/S0093934X01925192 (visited on 2020).

— (Sept. 2003). "Arguments for Adjuncts". en. In: *Cognition* 89.2, pp. 67–103. ISSN: 00100277. DOI: 10.1016/S0010-0277(03)00082-9.

Erk, Katrin (2007). "A Simple, Similarity-Based Model for Selectional Preferences". In: p. 8.

Koenig, Jean-Pierre, Gail Mauner, Breton Bienvenue, and Kathy Conklin (2007). "What with? The Anatomy of a (Proto)-Role". In: *Journal of Semantics* 25.2, pp. 175–220. ISSN: 0167-5133, 1477-4593. DOI: 10.1093/jos/ffm013. URL: https://academic.oup.com/jos/article-lookup/doi/10.1093/jos/ffm013 (visited on 2020).

Bergsma, Shane, Dekang Lin, and Randy Goebel (2008). "Discriminative Learning of Selectional Preference from Unlabeled Text". In: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2008. Honolulu, Hawaii: Association for Computational Linguistics, pp. 59–68. DOI: 10.5555/1613715.1613725. URL: https://www.aclweb.org/anthology/D08-1007 (visited on 2020).

Ferraresi, Adriano et al. (2008). "Introducing and Evaluating Ukwac, a Very Large Web-Derived Corpus of English". In: *In Proceedings of the 4th Web as Corpus Workshop (WAC-4*.

Schulte im Walde, Sabine et al. (2008). "Combining EM Training and the MDL Principle for an Automatic Verb Classification Incorporating Selectional Preferences". In: *Proceedings of ACL-08: HLT*. ACL-HLT 2008. Columbus, Ohio: Association for Computational Linguistics, pp. 496–504. URL: `https://www.aclweb.org/anthology/P08-1057` (visited on 2020).

Erk, Katrin, Sebastian Padó, and Ulrike Padó (2010). "A Flexible, Corpus-Driven Model of Regular and Inverse Selectional Preferences". In: *Computational Linguistics* 36.4, pp. 723–763. ISSN: 0891-2017, 1530-9312. DOI: `10.1162/coli_a_00017`. URL: `http://www.mitpressjournals.org/doi/10.1162/coli_a_00017` (visited on 2020).

Lenci, Alessandro (2018). "Distributional Models of Word Meaning". In: *Annual Review of Linguistics* 4, pp. 151–171.