Solving Relational Similarity Problems Using the Web as a Corpus

Preslav Nakov
Marti Hearst

BioText
University of Cal. Berkeley

Supported by NSF-DBI-0317510

Overview

Linguistically-motivated method for characterizing the semantic relations that hold between two nouns. Uses the Web to extract verbs, prepositions, and coordinating conjunctions revealing hidden relations.

State-of-the-art results for four problems:

Problem 1: solving SAT verbal analogy problems.

Problem 2: mapping head-modifier pairs to abstract relations like TIME and CONTAINER.

Problem 3: classifying the relations between nominals in context: e.g., Product-Producer.

Problem 4: characterizing noun-noun compounds using linguistic predicates like CAUSE, USE.

Method

Web Features:

- **Verb:** If one of the nouns is the subject, and the other one is an object of that verb, we extract it and we lemmatize it using WordNet. We ignore modals and auxiliaries, but retain the passive *be*, verb particles and prepositions (in case of indirect object).
- **Preposition:** If one of the nouns is the head of an NP which contains a PP, inside which there is an NP headed by the other noun (or an inflectional form thereof), we extract the preposition heading that PP.
- **Coordination:** If the two nouns are the heads of two coordinated NPs, we extract the conjunction.

"$infl_1$ THAT $\star$ $infl_2$"
"$infl_2$ THAT $\star$ $infl_1$"
"$infl_1$ $\star$ $infl_2$"
"$infl_2$ $\star$ $infl_1$"

Freq.	Feature	POS	Direction
2205	of	P	2 → 1
1923	be	V	1 → 2
771	include	V	1 → 2
382	serve on	V	2 → 1
189	chair	V	2 → 1
189	have	V	1 → 2
169	consist of	V	1 → 2
148	comprise	V	1 → 2
106	sit on	V	2 → 1
81	be chaired by	V	1 → 2
78	appoint	V	1 → 2
77	on	P	2 → 1
66	and	C	1 → 2
66	be elected	V	1 → 2
58	replace	V	1 → 2
48	lead	V	2 → 1
47	be intended for	V	1 → 2
45	join	V	2 → 1
...	...	...	...
4	be signed up for	V	2 → 1

Table 1: **The most frequent Web-derived features for *committee member*.** Here V stands for verb (possibly +preposition and/or +particle), P for preposition and C for coordinating conjunction; $1 \rightarrow 2$ means *committee* precedes the feature and *member* follows it; $2 \rightarrow 1$ means *member* precedes the feature and *committee* follows it.

Term Weighting:

$$w(x) = TF(x) \times \log\left(\frac{N}{DF(x)}\right)$$

Similarity:

$$Dice(A,B) = \frac{2 \times \sum_{i=1}^{n} \min(a_i, b_i)}{\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i}$$

Problem 1: SAT Verbal Analogy

	ostrich:bird		palatable:toothsome
(a)	*lion:cat*	(a)	rancid:fragrant
(b)	goose:flock	(b)	chewy:textured
(c)	ewe:sheep	*(c)*	*coarse:rough*
(d)	cub:bear	(d)	solitude:company
(e)	primate:monkey	(e)	no choice

Table 2: **SAT verbal analogy: sample questions.** The stem is in **bold**, the correct answer is in *italic*, and the distractors are in plain text.

Model	✓	×	∅	Accuracy	Cover.
$v+p+c$	129	52	3	**71.3±7.0**	**98.4**
v	122	56	6	68.5±7.2	96.7
$v+p$	119	61	4	66.1±7.2	97.8
$v+c$	117	62	5	65.4±7.2	97.3
$p+c$	90	90	4	50.0±7.2	97.8
p	84	94	6	47.2±7.2	96.7
baseline	37	147	0	20.0±5.2	100.0
LRA	122	59	3	67.4±7.1	98.4

Table 3: **SAT verbal analogy: 184 noun-only examples.** v stands for verb, p for preposition, and c for coordinating conjunction. For each model, the number of correct (✓), wrong (×), and nonclassified examples (∅) is shown, followed by accuracy and coverage (in %s).

Problem 2: Head-Modifier Relations

30 relations:
cause, effect, purpose, detraction, frequency, time at, time through, direction, location, location at, location from, agent, beneficiary, instrument, object, object property, part, possessor, property, product, source, stative, whole, container, content, equative, material, measure, topic, type

Example:
exam anxiety - effect
blue book - property

Model	✓	×	∅	Accuracy	Cover.
$v+p$	240	352	8	**40.5±3.9**	**98.7**
$v+p+c$	238	354	8	40.2±3.9	98.7
v	234	350	16	40.1±3.9	97.3
$v+c$	230	362	8	38.9±3.8	98.7
$p+c$	114	471	15	19.5±3.0	97.5
p	110	475	15	19.1±3.0	97.5
baseline	49	551	0	8.2±1.9	100.0
LRA	239	361	0	39.8±3.8	100.0

Table 4: **Head-modifier relations, 30 classes:** evaluation on the *Diverse* dataset, micro-averaged (in %s).

Problem 3: Relations Between Nominals

```
"Among the contents of the
<e1>vessel</e1> were a set of
carpenter's <e2>tools</e2>, several
large storage jars, ceramic utensils,
ropes and remnants of food, as well
as a heavy load of ballast stones."
WordNet(e1) = "vessel%1:06:00::",
WordNet(e2) = "tool%1:06:00::",
Content-Container(e2, e1) = "true",
Query = "contents of the * were a"
```

Model	Accuracy
$v+p+c+sent+query$ (type C)	**68.1±4.0**
v	67.9±4.0
$v+p+c$	67.8±4.0
$v+p+c+sent$ (type A)	**67.3±4.0**
$v+p$	66.9±4.0
sent (sentence words only)	59.3±4.2
p	58.4±4.2
Baseline (majority class)	57.0±4.2
$v+p+c+sent+query$ (C), 8 stars	67.0±4.0
$v+p+c+sent$ (A), 8 stars	65.4±4.1
Best type C on *SemEval*	67.0±4.0
Best type A on *SemEval*	66.0±4.1

Table 5: **Relations between nominals:** evaluation on the *SemEval* dataset. Accuracy is macro-averaged (in %s), up to 10 *Google* stars are used unless otherwise stated.

Additional Features (this problem only)

- **Sentence Word:** We use as features the words from the context sentence, after stop words removal and stemming with the Porter stemmer.
- **Entity Word:** We also use the lemmas of the words that are part of the entities.
- **Query Word:** Finally, we use the individual words that are part of the query string. This feature is used for category C runs only.

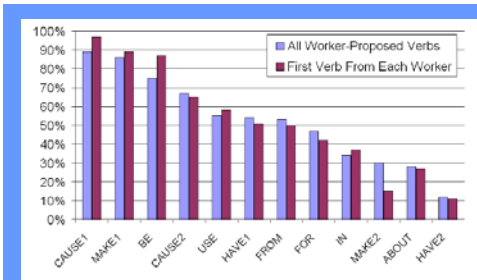

Figure 1: **Cosine correlation (in %s) between the human- and the program- generated verbs by relation:** using all human-proposed verbs vs. the first verb.

Problem 4: Noun Compound Relations

Model	USING THAT Accuracy	Cover.	ANF	ASF	NOT USING THAT Accuracy	Cover.	ANF	ASF
Human: all v	78.4±6.0	99.5	34.3	70.9	–		–	–
Human: first v from each worker	72.3±6.4	99.5	11.6	25.5	–	–	–	–
$v+p+c$	50.0±6.7	99.1	216.6	1716.0	49.1±6.7	99.1	206.6	1647.6
$v+p$	50.0±6.7	99.1	208.9	1427.9	47.6±6.6	99.1	198.9	1359.5
$v+c$	46.7±6.6	99.1	187.8	1107.2	43.9±6.5	99.1	177.8	1038.8
v	45.8±6.6	99.1	180.0	819.1	42.9±6.5	99.1	170.0	750.7
p	33.0±6.0	99.1	28.9	608.8	33.0±6.0	99.1	28.9	608.8
$p+c$	32.1±5.9	99.1	36.6	896.9	32.1±5.9	99.1	36.6	896.9
Baseline	19.6±4.8	100.0	–	–	–	–	–	–

Table 6: **Noun-noun compound relations, 12 classes:** evaluation on *Levi-214* dataset. Shown are micro-averaged accuracy and coverage in %s, followed by average number of features (ANF) and average sum of feature frequencies (ASF) per example. The righthand side reports the results when the query patterns involving THAT were not used. For comparison purposes, the top rows show the performance with the human-proposed verbs used as features.