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 modal verbs 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.

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

| Model | \( \checkmark \) | \( \times \) | \( \odot \) | Accuracy | Cover. |
|-------|---------|-----|-----|--------|------|
| \( v + p + c \) | 129 | 52 | 3 | 71.3 | 98.4 |
| \( v \) | 122 | 56 | 6 | 68.5 | 97.7 |
| \( v + p \) | 119 | 61 | 4 | 66.1 | 97.8 |
| \( v + c \) | 117 | 62 | 5 | 65.4 | 97.2 |
| \( p + c \) | 90 | 94 | 0 | 50.0 | 97.2 |
| \( p \) | 84 | 94 | 6 | 47.2 | 96.7 |
| baseline | 37 | 147 | 0 | 20.0 | 95.2 |
| LRA | 122 | 59 | 3 | 67.4 | 98.4 |

Problem 2: Head-Modifier Relations

30 relations:

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

Example:

- exam anxiety - effect
- blue book - property

Problem 3: Relations Between Nominals

"Among the contents of the <el>dresser</el> there was a set of <el>tools</el> and a big <el>alarm clock</el>."

| Model | \( v + p + c + s + q + y \) (type C) | Accuracy | Cover. |
|-------|-----------------------------------|--------|------|
| \( v \) | 68.1 | 4.0 |
| \( v + p \) | 67.9 | 4.0 |
| \( v + p + c \) | 67.3 | 4.0 |
| \( v + p + c + s \) | 68.3 | 4.0 |
| \( v + p + c + s + q \) | 59.4 | 2.6 |
| \( v + p + c + s + q + y \) | 58.0 | 2.4 |
| Baseline (majority class) | 57.9 | 4.0 |
| Best type C on SemEval | 67.0 | 4.0 |
| Best type A on SemEval | 66.0 | 4.0 |

Problem 4: Noun Compound Relations

| Model | \( \checkmark \) | \( \times \) | \( \odot \) | Accuracy | Cover. |
|-------|---------|-----|-----|--------|------|
| \( v + p + c \) | 240 | 358 | 8 | 40.5 | 98.7 |
| \( v + p + c + s \) | 238 | 358 | 8 | 40.2 | 98.7 |
| \( v + c \) | 234 | 356 | 10 | 40.1 | 97.3 |
| \( v + c + s \) | 230 | 362 | 8 | 38.9 | 98.7 |
| \( p + c \) | 114 | 471 | 15 | 19.5 | 97.5 |
| \( p + c + s \) | 110 | 475 | 15 | 19.1 | 97.5 |
| baseline | 49 | 551 | 0 | 50.8 | 100.0 |
| LRA | 239 | 361 | 0 | 39.8 | 100.0 |

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 -> 2 means member precedes the feature and committee precedes the feature and member follows it. 2 -> 1 means member precedes the feature and committee follows it.

![Figure 1: Cosine correlation (in %) between the human- and the program-generated verbs by relation: using all human-proposed verbs vs. the first verbs.](image-url)