Assignment 2: Lexical semantics (graded part)

Nora Graichen
Eva Richter

Exercise 2

Analyze your sense-annotated data from Exercise 1: quantitatively, reporting percentage agreement, and qualitatively, discussing the sources of disagreement. Make sure to use examples in your discussion. Note: percentage agreement is computed like accuracy (to think: why is it called "agreement" if it's computed identically?). You can report it as a proportion (between 0 and 1), or as a percentage (between 0 and 100).

Total number of synsets: 10 (6 nouns, 4 verbs)
POS tag: Noun
Lemma: Water
Synsets:

S1 ['water', 'H2O'] - Synset('water.n.01')
Definition: binary compound that occurs at room temperature as a clear colorless odorless tasteless liquid; freezes into ice below 0 degrees centigrade and boils above 100 degrees centigrade; widely used as a solvent
Examples: []

S2 ['body_of_water', 'water'] - Synset('body_of_water.n.01')
Definition: the part of the earth's surface covered with water (such as a river or lake or ocean)
Examples: ['they invaded our territorial waters', "they were sitting by the water's edge"]

S3 ['water'] - Synset('water.n.03')
Definition: once thought to be one of four elements composing the universe (Empedocles)
Examples: []

S4 ['water_system', 'water_supply', 'water'] - Synset('water_system.n.02')
Definition: a facility that provides a source of water
Examples: ['the town debated the purification of the water supply', 'first you have to cut off the water']

S5 ['urine', 'piss', 'pee', 'piddle', 'weewee', 'water'] - Synset('urine.n.01')
Definition: liquid excretory product
Examples: ['there was blood in his urine', 'the child had to make water']

S6 ['water'] - Synset('water.n.06')
Definition: a liquid necessary for the life of most animals and plants
Examples: ['he asked for a drink of water']

Nora Eva	s1	s2	s3	s4	s6	All
S1	24	0	0	0	1	25
S2	0	**31**	1	0	1	33
S3	0	2	3	0	0	5
S4	1	1	2	2	0	6
S6	7	0	0	0	24	31
All	32	34	6	2	26	100

Total number of samples: 100
Mean number of words in sample: 24.46
Standard deviation of words in sample: 15.7174676038727
Mean number of times "water" appeared in each sample: 0.64
Mean number of times "water" appeared in each sample: 0.6116710015534611
Percentage agreement: 84%
Percentage disagreement: 16%

Overall, the percentage of matching annotations is high at 84 percent and there are only 16 sentences in which the annotations differ. It is noticeable that especially in relation to sense6 and sense1 the sentences were differently annotated: in about half of the disagreement cases (7 out of 16 times) Eva has assigned s6 to "water" a sentence, while Nora has assigned s1 to "water" in the same sentence.

Apparently these are all sentences in which "water" occurs in connection with "cooking" or "kitchen utensils". For example, sentence 9 "*This water in this glass, and the steam rising from this pot [...]*" was annotated by Eva with s6, because "water" in the context of "glass" was perceived as a "food base", while Nora annotated it with s1, focussing on "water" itself without the context, i.e. "water" as a chemical substance. Other examples include sentence 17 *"Rinse the grains of quinoa in water"* and sentence 18 *"My grandmother got down a pair of kippers and broiled them for us in butter and water".* Also in these sentences "water" was annotated by Eva with s6 because of the context to "food" and by Nora with s1 as water which occurs as such as a chemical element.

In terms of the other differences in annotations with word sense categories, there is not really a pattern that can be discerned, there are only a few individual differences. Sentence 7 for instance, *"I have a few questions, first I want to start a sump filter, but I have no idea on how the water intake and water return would work"* was annotated as s4 by Eva due to the understanding of "water" as "water system" in the context of "water intake"/"water return" while considering "water" itself as a fluid or substance was annotated as s1 by Nora.

Another example is sentence 90 *"Water from the rivers is used in over 500 hydroelectricity power plants, generating as much as 2900 kilowatts of electricity",* annotated by Eva as s4 and by Nora s2. This is because Nora, given the beginning of the sentence, annotated "water from the rivers" as s2, i.e., "body of water," which is basically a "river", while Eva, given "hydroelectricity power plants," assumed s4, i.e., "water" in the context of a "facility that provides a source of water." This shows that it can be problematic to determine word senses when "water" occurs "several times" in different senses.

Exercise 3

State your hypotheses about the following questions, and find out whether the data supports your hypothesis, using Word2Vec and WordNet (use at least 500 data points):

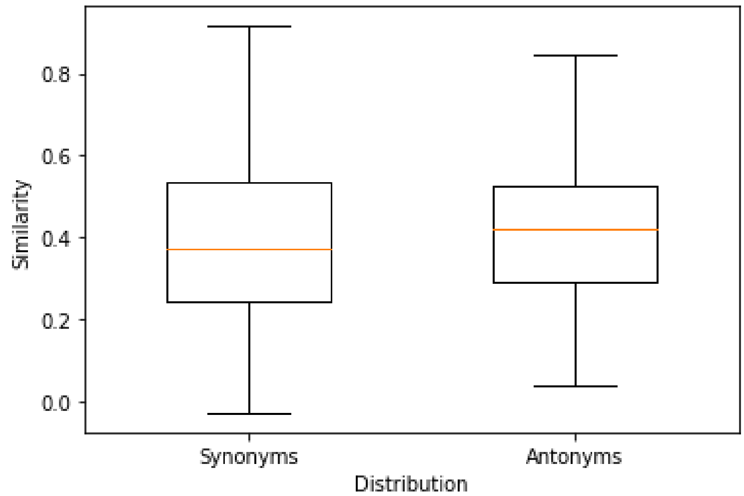
- *Which word pairs appear in more similar contexts, synonyms, or antonyms?*
- *Which word pairs appear in more similar contexts, hypernyms, or hyponyms?*

Report quantitative results and a graph visualizing the results. Reflect on your findings (around 50 words).

Which word pairs appear in more similar contexts, synonyms, or antonyms?

-> Synonyms theoretically should appear in more similar contexts than antonyms. However, since antonyms usually have similar contexts and the embeddings are trained based on word distributions, we would expect that the overall results are similar.

We compare the antonyms + synonyms of the same word:
Mean synonym-similarity: 0.3925
Mean antonyms-similarity: 0.4109
Significant difference? t-test: p:
False (not smaller or equal to 0.05)
t-test: p: 0.163597

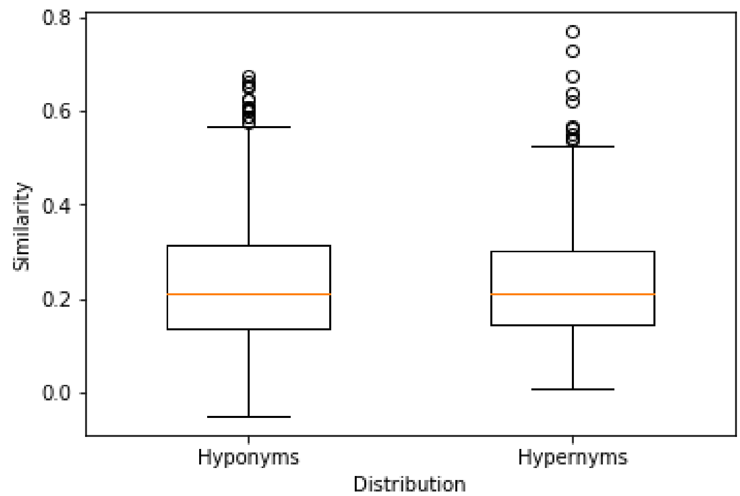


On average, a word is equally similar to its synonyms as it is to its antonyms. This means that the embedding space is insensitive to this word relation and can not model it properly. Though the 1st/3rd quartile and most extreme similarity scores inside the 1.5 x inter-quartile range can be interpreted that antonyms' similarity score distribution is less scattered. Therefore, this small concentration difference hints that antonyms' word pairs could appear in more similar contexts, and greater than 0!

Which word pairs appear in more similar contexts, hypernyms, or hyponyms?

By grouping similar words together in the embedding space, the hyponym of a word should be more similar than it is to its hypernym because it is more specific, although they should overall appear in similar contexts.

We compare the hyponyms+hypernyms of all synonym words of a synset:
Mean hyponyms-similarity: 0.2368
Mean hypernyms-similarity: 0.2340
Significant difference? t-test: p: False
t-test: p: 0.72478
On average, a word and synonyms (*but also for only one/the same word:*
[Mean hyponyms- similarity: 0.24328
Mean hypernyms-similarity: 0.24079])
are equally similar to their hyponyms as they are to their hypernyms. This means that the embedding space is too coarse and can not model this word relation optimally just by the word distributions. However, similar to the antonyms' word pairs, the hypernyms similarity score distribution is less scattered regarding the quartiles (and not the outlier data points) and doesn't include zero. This indicates that less specific but subordinate words of synonyms should appear in more similar contexts.



Exercise 4

Write two mini-essays about how natural language meaning should be captured (100-150 words each): one arguing against the sense enumeration approach (à la WordNet) and in favor of distributional methods, another arguing the opposite, that is, against distributional methods and in favor of sense enumeration. For the second essay, it may be useful to check sections 1-2 of the following article: Jose Camacho-Collados and Mohammad Taher Pilehvar (2018). From Word to Sense Embeddings: A Survey on Vector Representations of Meaning. Journal of Artificial Intelligence Research. https://www.jair.org/index.php/jair/article/view/11259/26454

Essay 1

Analyzing word senses[1] as abstractions from clusters of corpus citations demonstrates that defining a set of word senses, independent of the task, is not a reasonable concept, since many cases are not straightforward. That means, the assumption of word senses can be a disadvantage in that they are "slippery entities" (Kilgarriff, 1997), since word senses are undefined as long as there is no underlying reason for clustering.

Hence, a general lexicon is not suitable for different NLP tasks to the same extent and a NLP system relying on a set of word senses defined in a dictionary would not obtain high coverage. Instead, we would have to define a set of word senses, relative to a specific task and not rely on a predefined set from a dictionary, thesaurus or similar (Kilgarriff, 1997).

Moreover, the lists of word senses in dictionaries are usually designed for different purposes for human users and it should not be assumed that they are suitable for NLP applications at all, much less to the same extent (Kilgarriff, 1997). Additionally, the sense enumeration approach is language specific and not completely transferable, while word embeddings have been shown (e.g.,Voita 2020: NLPCourse) to effectively capture linear relationships between semantic spaces (cross-lingual).

Essay 2

The main disadvantage of distributional models[2] is what is called "meaning conflation deficiency", according to which the representation of a word as a point in the semantic vector space is problematic in that all meanings are combined into a single vector, so that multiple meanings of a word are not taken into account. Meaning conflation also has a negative effect on the accuracy of the models, since, for example in the Word2Vec static word embeddings, "mouse" can refer to the animal in the same way that mouse refers to the computer, and thus semantically unrelated words such as "computer" and "rabbit" are drawn closer together in semantic space (Camacho-Collados/Pilehvar, 2018).

Moreover, it has been shown (for example exercise 2) that certain distributional models do not learn good representations that can model underlying word relationships just by word distribution. On the other hand, sense enumeration can reliably distinguish between the word sense relations.

---

[1] Word sense disambiguation (WSD) aims to disambiguate ambiguous words, i.e. based on a prior definition of different word senses, ambiguity is resolved by selecting the most appropriate word sense for the given context.

[2] Distributional models are based on the assumption that the meaning of a word can be inferred from the context in which it appears. Although such models are very popular because of their effectiveness and generalization power, they have some drawbacks (Gahlot, 2018).

References

Adam Kilgarriff (1997). *I Don't Believe in Word Senses.* In Computers and the Humanities. 31(2), pp. 91-113.

Jose Camacho-Collados and Mohammad Taher Pilehvar (2018). *From Word to Sense Embeddings: A Survey on Vector Representations of Meaning.* In Journal of Artificial Intelligence Research. https://www.jair.org/index.php/jair/article/view/11259/26454.

Tanay Gahlot (2018). *Moving beyond the distributional model for word representation.* Accessed 22 November 2021, https://towardsdatascience.com/https-medium-com-tanaygahlot-moving-beyond-the-distributional-model-for-word-representation-b0823f1769f8.