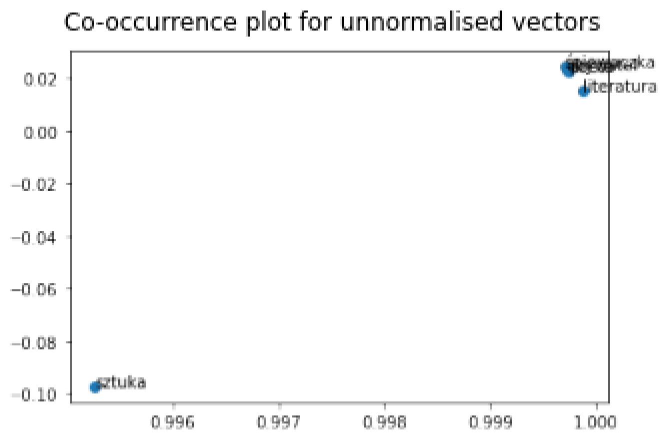
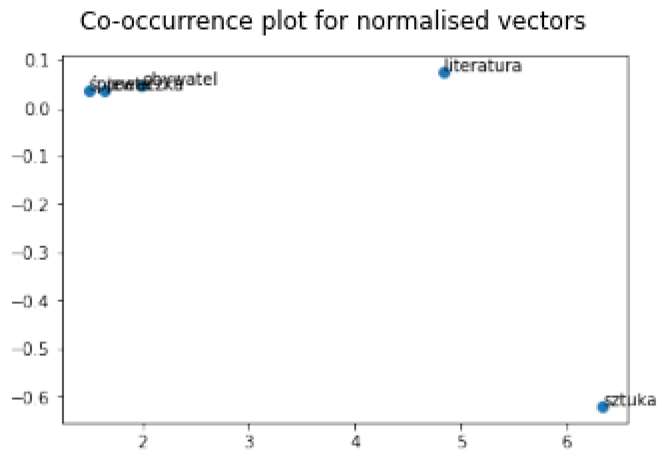
Natalia Rutecka

NLP - report of practical 1

Exercise 1

The embeddings of words "sztuka", "śpiewaczka", "literatura", "poeta" and "obywatel" calculated from a co-occurence matrix are depicted below.



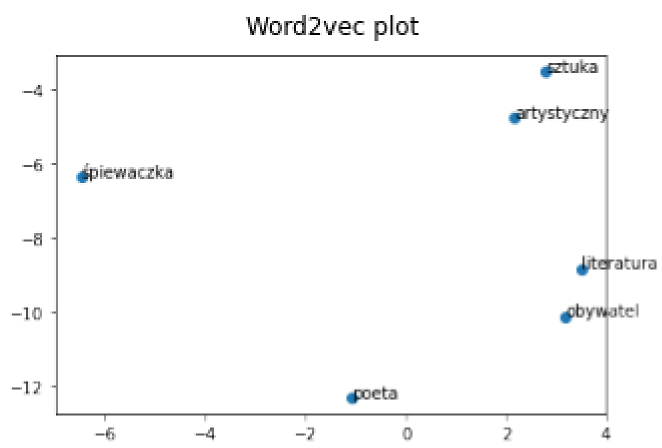


In both the normalised and the unnormalised dataset we can observe that words cluster mainly based on their function in a sentence, i.e. whether they describe a person or an abstract concept. For example, in both datasets the vector corresponding to a word "śpiewaczka" is much closer to the vector corresponding to a word "obywatel" than to a vector corresponding to a word "sztuka", even though one might argue that "śpiewaczka" and "sztuka" are more similar in terms of the ideas they represent.

In the above example, the normalisation process helps to improve readability of the visualisation, but it doesn't change the interpretation of the embedding - the normalised and the unnormalised data form the same clusters. Therefore, it is probably not necessary to normalise the data, unless it is directly required in further analysis.

Exercise 1

a) The 2-dimensional embeddings of words "sztuka", "śpiewaczka", "literatura", "poeta", "artystyczny" and "obywatel" embedded using word2vec are depicted below:



The embedded vectors produced by word2vec algorithm look a lot different to the ones produced using co-occurrence matrix. The words that describe a type of person don't cluster together in space and the distances seem to be based more on the underlying ideas represented by a word. In general the embedding looks more informative than the one achieved by co-occurrence matrix. The only thing that looks counterintuitive is the distance between "śpiewaczka" and "poeta" - they are far away even though they represent a similar concept - a person who writes/sings poetry.

b) In case of polysemous words "pilot" and "zamek", the closest words are only related to one meaning. This probably results from the words having one dominant meaning in the training data. Another reason might be that one of the meanings doesn't have synonyms in the language that are often used - for example its hard to come up with a synonym to the word "pilot" if the word is meant as a remote control.

In case of a word "słodki", the closest words were either related to something being cute (rozkoszny, tkliwy) or having sweet taste/smell (wonny, soczysty, cierpki).

c) Synonyms "kochać", "uwielbiać" have cosine distance: 0.37 whereas antonyms "kochać", "nienawidzić" have cosine distance equal to 0.23. This phenomenon might result from multiple factors.

In general, antonyms tend to have very similar embeddings because they are used in similar context. In case of synonyms, they often have some subtle differences in meaning, eg. one is less formal than the other. In the example above, the antonyms "kochać" and "nienawidzić" might also more often co-occur in a sentence than "kochać" and "uwielbiać", eg. in aphorisms like "Ludzie będą cię kochać. Ludzie będą cię nienawidzić. I nic z tego nie będzie miało z tobą wspólnego." or "Kochaj grzesznika, nienawidź grzechu."

d) An example of an analogy that is solved is dziecko:chłopiec::dorosły:x, where x is correctly predicted as "mężczyzna".

e) An example of an analogy that is not solved is jeść:talerz::pić:x where x should be a type of vessel from which you can drink, eg. kubek, szklanka, but all top 10 results are verbs connected to drinking: popijać, wypijać, nalewać etc.

f) An analogy mężczyzna:szef :: kobieta:x hasn't been solved correctly - there is no word describing a female boss in the top 10 results, even though such words exist in the corpus (eg. szefowa, dyrektorka, kierowniczka). Most of the words in top 10 results are not related to a leadership position at all. The 10th most similar word is "pracownik", suggesting that, according to the model, a relation between man and woman is similar to a relation between a boss and a worker. In the second example we can see that when searching for a word similar to "prezes" and "mężczyzna" and dissimilar to "kobieta", the results are related to leadership and responsibility. These examples show that the model assumes leadership and masculinity to be related concepts.

g) Another example of bias is incorrectly solving an analogy mężczyzna:lekarz :: kobieta:x. In this case the most similar vector corresponds to a word "pielęgniarka", meaning that the model associates masculinity with being a doctor and femininity with being a nurse.

h) The biases in the word vectors are replicating the biases in the language of the training data. In case of the gender bias, it means that the training examples use masculine and feminine forms in different contexts (eg. a word "prezes" might be more often use to describe someone successful and a word "prezeska" in discussions of importance of using female forms). Biases in the language are not recognised by the model, therefore it doesn't distinguish them from true meanings of words.

i) Polysemous words in English are equally hard to resolve and tend to cluster only with words related to one of their meanings, eg. the word "bar" is close only to words like tavern/restaurant.

The problem with antonyms is exhibited in the form of words "smart" and "dumb" being closer in the space than "smart" and "clever", which probably results from first two words being more informal and therefore more often used in the same context.

The embeddings of English words also exhibit some problems with analogies, including having the same bias with solving the analogy he:doctor :: she:x as x=nurse. Here however, the bias is not as blatant since in English there is no female form of the word "doctor", so the analogy doesn't have a clear correct answer.

All in all, both English and Polish embeddings have the same issues: treating antonyms as very close in terms of meaning, troubles with sensibly embedding polysemous words and gender biases. However, examples of gender biases are harder to find in English embeddings and are not as obvious.