Predicting reaction time in semantic priming tasks using word embedding models

Research report as part of the seminar

"Linguistics for Cognitive Science"

Wintersemester 2020/2021

Faculty of Science of the

Eberhard Karls University of Tübingen

presented by:

Sebastian Volz

Matrikelnumber: 5457430

Tübingen, 16.04.2021

List of Contents

Introduction

Semantic Priming tasks are a well-studied topic in Psycholinguistics and many factors have been found to influence reaction time in semantic priming tasks (e.g. word length, word-frequency, orthographic neighbourhood). Since the semantic similarity seems to affect word recognition performance and word embeddings aim to represent word meaning, I will investigate, whether word embeddings can also be used to predict reaction time.

Semantic Priming

Theory

Semantic priming refers to the observation, that the reaction time (RT) to a target-word preceded by a semantical similar prime-word is lower in comparison to an unrelated prime. This effect is well documented, however, the explanation for when this happens is not entirely clear. Whereas Shelton und Martin (1992) seem to show that a priming effect only occurs when words are associated in the person's mind, regardless of the semantic similarity, McRae und Boisvert (1998) argue that that the semantic similarity is the decisive factor. It is my goal to use word embeddings, representing the semantic meaning of a word, and association norms by Nelson et al. (2004) to predict the reaction time in a lexical priming task. Therefore, a dataset of prime-target pairs is necessary.

Semantic Priming Project (SPP)

The semantic priming project supplies such a dataset, with over 420 000 prime-target pairs and their corresponding reaction time in a lexical speed naming task (Hutchison et al. 2013). Data for 1661 prime-target pairs from 256 subject at four universities was collected. The prime-target pairs were selected from the Nelson et al. (2004) association norms. Association norms "index the likelihood that one word can cue another word to come to mind with minimal contextual constraints" (Nelson et al. 2000). The score of association was calculated by asking participants the first word that comes to their mind that was associated with the presented cue.

Each target was either produced as the first-associate (primecondition 1) or other-associate (2^{nd} to n) (primecondition 2) to a word. These related pairs were scrambled to make unrelated pairs from the first-associates (primecondition 3) and other-associates (primecondition 4). (I could not find out how the primecondition was computed so I contacted Mr Keith A. Hutchison who kindly replied and afterwards updated his website to make it clearer)

R-Code / Data

After downloading the excel table, I first decreased the prime-target data by randomly selecting 150 000 rows, in order to save computing capacity. Then I excluded columns irrelevant for my goal, omitted rows with "<NA>" entries and changed all words to lower case. Furthermore, I excluded rows with a reaction time lower than 200 ms and higher than 1200 ms and then excluded rows with a RT outside an interval of three standard errors around the mean reaction time. To get a character vector of all words used in the semantic priming task, I converted all primes and targets to a factor and saved the levels in an extra variable "allWords". Later this will be important, to compare the word embeddings with the prime and target words and without having to do calculations on all word embeddings but only the ones relevant for the semantic priming.

What is still missing is the distance of word embeddings as the explanatory variable.

Word Embeddings

Theory

Word embeddings are representations of words as vectors with the goal of encoding their meaning in the vector. Latent Semantic Analysis (LSA) is one way to create word embeddings. By counting how many times a word is observed in a document, a large matrix is created. Afterwards, matrix transformations, such as singular value decomposition is performed. Although originating from psycholinguistic literature, this process doesn't seem to have much in common with the way our brains find associations between words (Mandera et al. 2017). A better approach might be algorithms such as the Gensim Continuous Skipgram or Gensim Continuous Bag Of Words (cbow). Instead of explicitly representing words in a matrix, a context-window sliding through the corpus adjusts the weights depending on which words are found in the context of the word. Using a backpropagation rule similar to the Rescorla-wagner rule, these weights are then adjusted to minimize errors between prediction and observation.

Whereas skipgram uses the word to predict the context, cbow uses the context to predict the word. Furthermore, skipgram weighs nearby words more heavily and does a better job for infrequent words, which I think will be beneficial in predicting reaction time, since the prime target words are mostly infrequent words. Further, I think a stronger connection of nearby occurring words is beneficial for predicting semantic priming.

The Language Technology Group at the University of Oslo has a vast set of models trained on different corpuses using different algorithms (http://vectors.nlpl.eu/repository/#).

R-Code / Data

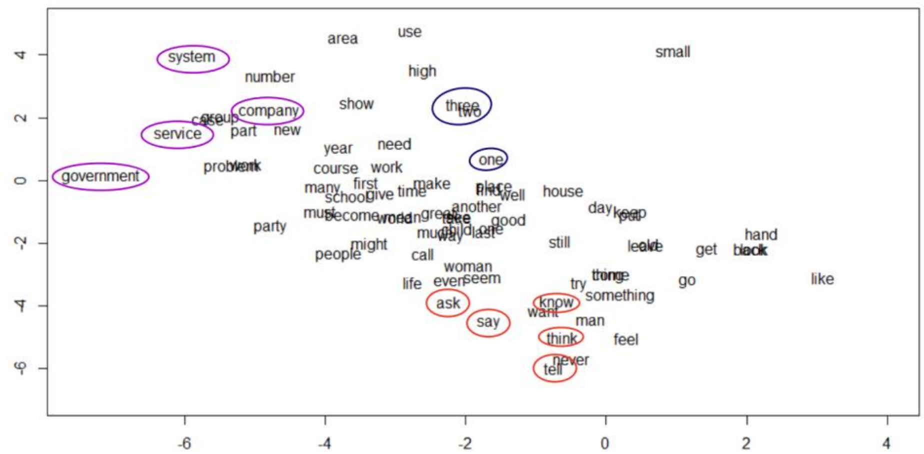
I downloaded two word-embedding models. One model used the Gensim Continuous Skipgram algorithm trained on the British National Corps (163473 vocabulary size) and the other model used the Gensim cbow trained on the "news on the web" from 2014. Both algorithms used a window of five words, created vectors with a dimension of 300 and were lemmatized (inflected words were grouped together). Unfortunately, I could not find two large embedding models using the same corpus, only differentiating in their algorithm. Therefore, comparing their influence on RT is very limited and attributing a difference to the algorithm is not possible.

Since the data downloaded was very large and my computer could not handle such big datasets, I could not use all word embeddings and I chose an arbitrary number. This resulted in only a percentage of prime-target pairs left with their corresponding word embeddings.

To limit the differences between the two models, I deleted rows of the cbow model so that both data sets created by the different algorithms had the same number of rows.

After reading the data into R, I first changed the words from e.g. “walk_VERB” to just “walk”. I had to do this, so that later I could find the appropriate vector for the prime or target word. The downside to the process was, that it left space for homonyms, so the chance that wrong word embeddings are attributed to the prime or target is high. However, comparing 3901 words by hand to find the appropriate word embedding would have been too time intensive.

Furthermore, I excluded all word embeddings not occurring as a prime or target word, which decreased the data greatly. Then, in analogy to our “embedding” lab session, I converted the embedding data frame to a matrix and performed a principal component analysis to reduce the dimensions of the word embeddings. Looking at the first 80 words, a semantic similarity is found in two-dimensional space.



Note: Two PCs explaining most Variance of word embeddings, which were created using the gensim continuous skipgram algorithm.

Using a for-loop, I added the two principal components explaining most variance to each word of the prime-target pair data and calculated the Euclidean distance between prime and target. After that, I deleted rows not containing embeddings of both prime and target.

prime	primecond	target	target.RT	dist
shallow	3	sauce	623	4.0454222
discreet	2	cautious	497	4.5401294
ascent	2	climb	412	1.7227683
deposit	4	animal	535	10.0600769
buffalo	2	plain	586	3.9938090
breeze	1	wind	421	3.0163054
middle	3	alone	400	5.0628065
guardian	3	female	525	1.3076270
jury	2	twelve	516	4.2395907
further	3	liberty	818	7.6209513
maroon	4	trouble	466	4.7245621

Note: Final data frame with prime, target, reaction time in lexical speed naming task and euclidean distance of word embeddings (algorithm: cbow).

Finally, I performed a linear regression to predict the observed variable "*reaction time*" by the explanatory variable "*euclidean distance of word embeddings*".

Results

After excluding all rows with a RT outside the 200-1200 ms interval and three standard error interval, 139550 rows were left. The prime-target data consisted of a total of 3901 words.

1. Gensim Skipgram Model

I arbitrary chose the first 55 000 words. After comparing them to the prime target words, 4921 were left. This is more than the total amount of prime-target words since I excluded the extra information _VERB or _NOUN and therefore homonyms were left in the embedding model. In the end, for 118675 words, embeddings for both prime and target were found.

The linear regression model showed, that there is a significant relationship between RT and the "euclidean distance of vectors" $F(1,118673) = 27.77$, $R^2 = .00023$, $p < .001$ and the primecondition $F(3, 118673) = 25.03$, $R^2 = .00061$, $p < .001$.

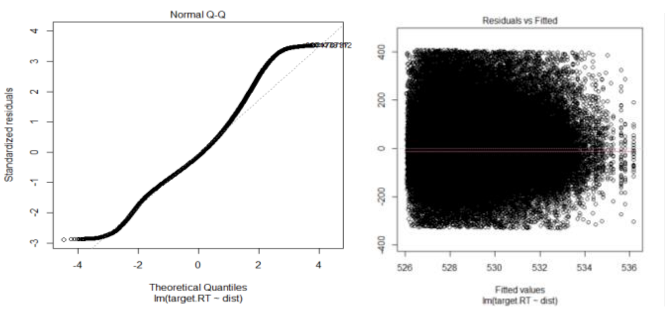
Theoretically, 1 point higher in distance results in a RT increase of .6625 ms. Using the factor primecondition as the explanatory variable, dummy values are set in the linear regression. Setting primecondition 1 as the base value, RT increases by 3.3 ms to primecondition 2, by 6.9 ms to primecond 3 and by 6.9 ms to primecond 4.

2. Gensim cbow

I arbitrary chose the first 63 561 words. After comparing them to the prime target words, 9125 were left. In the end, for 121553 words, embeddings for both prime and target were found. This is more than in the skipgram model, so I excluded 2878 rows and again ended up with 118675 rows.

The linear model did not showed a significant result for "distance" as the explanatory variable $F(1,118673) = 3.56$, $R^2 < .0001$, $p = .0592$ and again a significant result for primecondition $F(3,118673) = 29.11$, $R^2 = .00074$, $p < .001$.

The quantile-quantile plots show almost normally distributed data and the residuals vs. fitted plot speaks for homoscedastic data, meaning that residuals are the same across the independent variable.



Note: QQ-diagram and residuals vs. fitted diagram

Interpretation

Since homonyms were left in the word embedding data, one cannot be sure the appropriate word embedding in the data frame was found.

Looking at the cbow model, it did not yield a significant result. Due to the fact, that the skipgram and cbow model do not only differ in their algorithm but also the corpus trained on and the amount of homonyms left in the word embedding model, the worse performance cannot be attributed to the algorithm. However, I do not think the different algorithm is responsible for such a difference and the corpus “news on the web” should not be worse training data than the “British national corps”. Most likely, often the wrong word embeddings were entered in the data frame, which resulted in the non-significant result.

Furthermore, the cbow model in the study of Mandera et al. (2017) outperformed the skipgram model in the lexical naming task in average (varying window size, predict model size and corpus trained on). This is not consistent with the idea, that the algorithmic structure of words standing closer to each other are more heavily weighted, is beneficial for predicting semantic priming reaction time.

The result of the skipgram model seems a lot more promising. Only 4921 word embeddings were compared to a total amount of 3901 prime target words, which leaves less space for homonyms. The linear regression was performed using 111865 prime target pairs. In the end, .00023 Variance could be explained. This is not a lot comparing it to Mandera et al. (2017). Using log word frequency, word-length and orthographic neighbourhood density they were able to explain 0.312 Variance. Once the word-embeddings created by the skipgram model were added as a predictor, an average of 0.327 variance could be explained – .015 more.

Comparing the linear model using word embeddings or association norms as the explanatory variable, the association norms explain more variance and one could therefore argue, that they are the better predictor. This can be attributed to the fact, that association norms are a better indicator for reaction time in semantic priming, or the appropriate association norm for each word was not always saved in the data frame.

In summary, it can be said that difference in priming can be modelled as a function of semantic similarity and of the association norms. However, finding a difference in the performance of the cbow and skipgram algorithm would require data that is more consistent.

Bibliography

Hutchison, Keith A.; Balota, David A.; Neely, James H.; Cortese, Michael J.; Cohen-Shikora, Emily R.; Tse, Chi-Shing et al. (2013): The semantic priming project. In: *Behavior research methods* 45 (4), S. 1099–1114. DOI: 10.3758/s13428-012-0304-z.

Mandera, Paweł; Keuleers, Emmanuel; Brysbaert, Marc (2017): Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. In: *Journal of Memory and Language* 92, S. 57–78. DOI: 10.1016/j.jml.2016.04.001.

McRae, Ken; Boisvert, Stephen (1998): Automatic semantic similarity priming. In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24 (3), S. 558–572. DOI: 10.1037/0278-7393.24.3.558.

Nelson, D. L.; McEvoy, C. L.; Dennis, S. (2000): What is free association and what does it measure? In: *Memory & cognition* 28 (6), S. 887–899. DOI: 10.3758/bf03209337.

Nelson, Douglas L.; McEvoy, Cathy L.; Schreiber, Thomas A. (2004): The University of South Florida free association, rhyme, and word fragment norms. In: *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc* 36 (3), S. 402–407. DOI: 10.3758/bf03195588.

Shelton, Jennifer R.; Martin, Randi C. (1992): How semantic is automatic semantic priming? In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18 (6), S. 1191–1210. DOI: 10.1037/0278-7393.18.6.1191.

I hereby declare that I wrote this research report myself, that I did not use any aids or sources other than those indicated, and that I have marked all statements taken literally or analogously from other works.

Tübingen, the 16.04.2021

S. Volz

____

Sebastian Volz