Title

Rasyan Ahmed
10784063

Master Thesis
Credits: 42 EC

Masters Artificial Intelligence

University of Amsterdam
Faculty of Science
Science Park 904
1098 XH Amsterdam

Supervised by:
Tom Lentz
dr. Jelle Zuidema

4th June 2020

Institute for Logic, Language and Computation (ILLC)
University of Amsterdam
Faculty of Science
Science Park 107, 1098 XG Amsterdam
https://www.illc.uva.nl/

Abstract

ljlkj

Contents

Chapter 1

Introduction

- based on broderick study
- TRF method
- contains problems
- introducing new elements

Chapter 2

Literature review

- previous (classic) N400 studys
- in detail broderick study
- bert
- need more?

Chapter 3

Method

- Experimental setup: Data
- Creating the dissimilarity values (pearson and Bert)

  The information for each of the 20 runs are provided in a .mat file which includes the onset and offset time of each word, their sentence boundary's, and which word was spoken. With this information we can reshape the data into a three dimensional array that contains the spoken words wit the following shape: number of runs x number of sentences x number of words in the sentence. Note that the last two dimensions are variable, as each sentence has a different number of words in it, and each run has a different number of sentences.

  —— Notes : make size of array more clear then just text?

  Pearson

  The first step here is to load the actual word2vec embedding into memory by using gensim. Using these embeddings we can calculate the dissimilarity value of each word in the same fashion as the original broderick paper. For each word we calculate the Pearson's correlation of its embedding with the average embedding of the words that came before it in its sentence. This average embedding is made by simply taking the mean of all the embeddings of the previous seen words in the sentence. While this technically can result in a value between -1 and 1, in our case the results are between 0 and 1, these can be seen as a similarity value. The last step is to then subtract these correlation coefficients from 1 to create dissimilarity values. The end result is again a three dimensional array of the same exact size as before, but this time containing dissimilarity values instead of the words.

  ——Notes: Why are they all between 0 and 1?
  ——Notes: references to gensim? references to the word vectors? but that may be in the data section above

In its implementation a few exception pop up. The main problem here is what you should do when a word is the first word of its sentence, or of the entire run. Broderick et al have noted that they look at the full previous sentence for the first word in each sentence and we agree on this implementation, but they have not clarified what should be done if that previous sentence is empty. An empty sentence happens when the natural sentence contains no content words. In this case we do have sentence boundary's but no content words onset between those boundary's. We have decided to recursively look into the past in this case until we find the first non empty sentence. We motivate this decision by the fact that these empty sentences are often relatively short, usually consisting of only a single "Yes." or "No.". As such the time spend in that sentence is short and low of meaning such that the sentence before it should still be fresh in memory and be of relevant context to provide adequate information on whether the first word of the next sentence is surprising or not. In The most extreme case this means that we have to look back three sentences into the past.

This same argument can not be made for the first word in a run, While the previous run ended of at the same spot in the audio book as where this one began, there is no information on how long a break occurred between these runs or whether these breaks were of equal sizes for each participant. As such while its of relevant context it is not fresh in memory and can therefor not be used to calculate the dissimilarity value of the first word in the run. We give these words a dissimilarity value of 0 so that they are in essence skipped while training. This has a minimal effect on the results as this can be seen as reducing or data set by 20 words.

Bert

The second method to calculate similarity values is to create a masked language task where we ask a language model to provide a likelihood score of the target word occurring in the masked spot of the sentence. This likelihood score is seen as the similarity value (that is between 0 and 1) between the target word and the provided context and can again be shaped into a dissimilarity score by subtracting it from 1. The context itself can be chosen at will. We have made two versions. One that has the context defined as the full sentence, before and after the target word occured. And one that looks at the previous sentence and the current sentence up to the target word.

The Bert model used here is a pretrained version provided at: REFERENCE? ——Notes explain the properties of bert here? uncased/cased

- Processing the EEG data
- Learning: TRF
- Calculating R2

- need more?

Chapter 4

Results

- Comparision, our results vs Theirs. Confirm reproduction. Bert seems better?
- Additional Curve plots. show Curve plots not usefull.
- R2 plots
- need more?

Chapter 5

Discussion

- Why TRF didnt work.
- What does this mean

  Neuroscience and Artificial Intelligence are two fields of study that are highly related when looking at the matter that they study, intelligence. Neuroscience focuses on a existing black box model, the (human) brain, that is confirmed to contain intelligence and tries to pry from it its secrets. AI engineers on the other hand create models based on mathematical principles that exhibit a subset of properties (such as decision making) that are often attributed to intelligence. While they both study the same subject matter, one does it from a more top down approach and the other from the bottom up. Due to their vastly different approaches the two fields do not intersect as often as they probably should do (in our opinion????). (We believe that?????) both fields could benefit greatly from greater cooperation between them.

  It is often said that AI methods are inspired from biological processes inside the human brain but often times the resemblances between the two are mostly on a superficial level. The direction that these models develop towards are often towards the path of highest accuracy gain, even if those paths are not biological plausible. That is not to say that this approach is wrong, AI models do not need to be biologically plausible to be use full. But biological correlation should function as a heuristic of sorts through the space of possible models to find approaches that while perhaps do show immediate results, could revolutionise AI in the long run. The most famous example is the use of neural networks nowadays while they were largely discarded at first.

  From the neuroscience perspective, the study of intelligence has largely been done through studying the biological tissue, and the chemical and electrical processes of the brain. Through these study's a large amount of data is brain activity data is collected, but this data is often only used in very small quantity's as making sense of it is extremely hard. The brain is a complex organ that we know little about and as such the data captured from its largely unreadable and noisy. AI modelling could be the solution to understanding which processes are responsible for which parts of the data and be used to for example confirm or deny an hypothesis.

Here we combine both fields to try to study how the brain processes words while listening to spoken text through the use of machine learning models on EEG brain data. More specifically, an computational model is used to try to find the well studied N400 effect in continuous EEG data.
neuroscience???

5.0.1 Future Research

- alternative methods that might work:
- NN approach (with auto encoder)
- unlimited power/data: LSTM (If works, argument that brain is a little like LSTM)
- introducing more input variables.

Chapter 6

Conclusion

- TRF doesnt work. underfits
-

Bibliography