# A Model for Approixmating Skimming Behaviour

Orcun Tasdemir '22 — COGS319: Modeling Minds, Brains, and Behavior — Fall 2019

My project is an attempt to approximate how people skim through text as they read it, and to re-create it with an algorithm using levels of satisfaction for the information gathered so far from the text. The heart of my model is the assumption that the skimming behavior works within the mechanism of Optimal Foraging Theory which is a theory of behavioral ecology that tries to explain and predict how animals behave as they are in search of food. Animals often search for resources in spatial patches such as berries on a bush or prey animals in a herd. There have been research strongly suggesting that retrieving information from memory is analogous to the optimal foraging mechanism in the way that people first retrieve items that are semantically closer to each other and exhaust that patch before going and looking in another semantic patch of memory. Just as memory items, the information to-be-gathered from text can be considered a resource that should be optimally foraged. As people face with limited time to read a text, they would try to maximize the information they can learn from the text, and this maximization comes with optimizing which parts to read and which parts to skip over.

Natural Language Processing is a sub-field of linguistics and computer-science and with the recent development and improvement of word embedding algorithms such as Word2vec from Google, and GloVe from Stanford, the field offers a wide variety of possible research areas with powerful tools.

## Metrics for the Skip Decisions

The key concept is the marginal information gain. If a patch of The model predicts that people would skip over sentences if they had just seen several sentences that conveyed a very similar meaning, which would mean that the concept is repeating and not offering novel information. Overall , the model would be skipping over sentences given that looking at the past sentence similarity results, it predicted that the marginal information gain will be going down if it did not skip, which would be considered lost time as the model is not learning anything new out of the text.
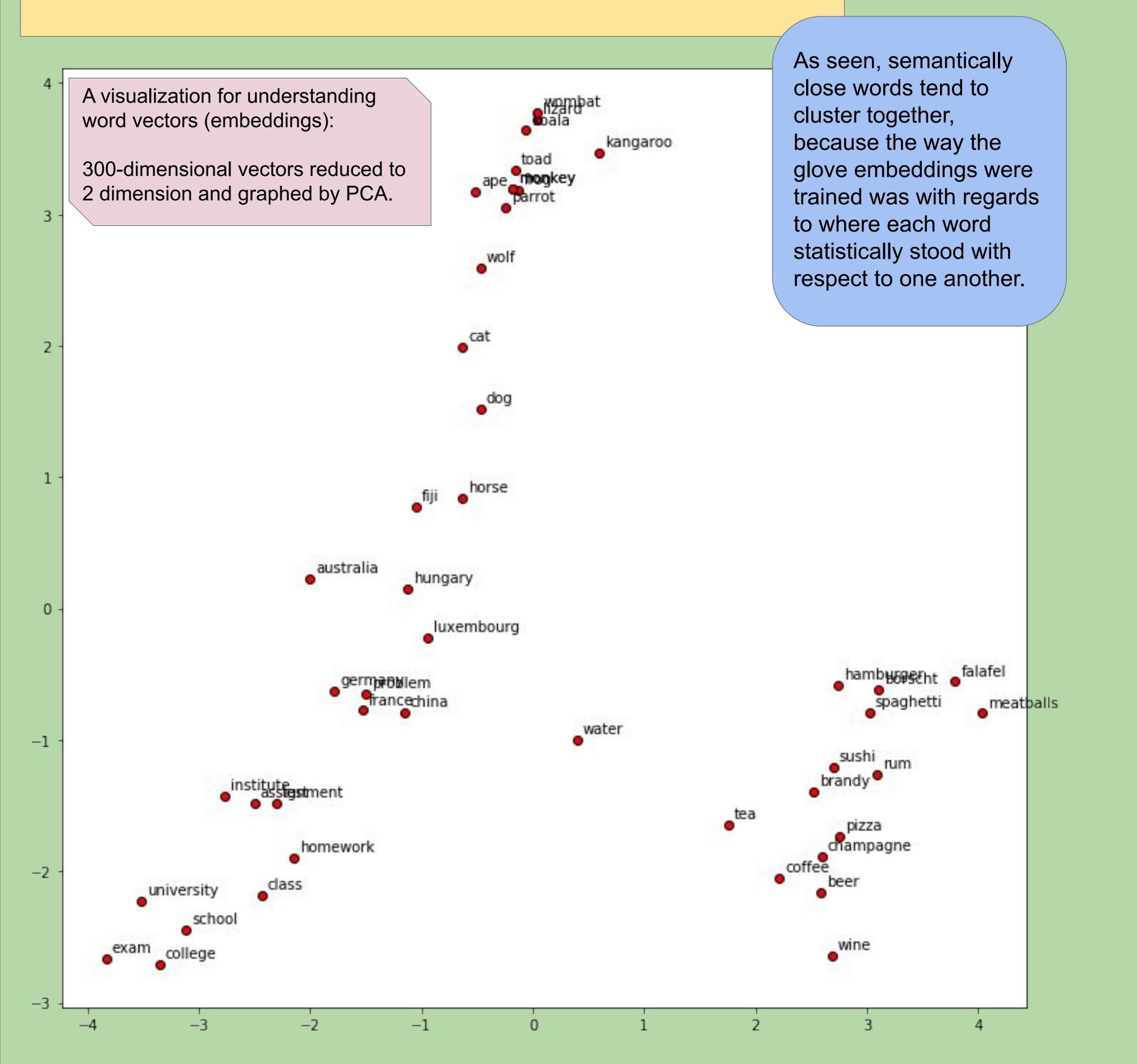
## Modelling

I used the 300 dimensional GloVe word embeddings to identify each word. The preprocessing for the text was used to get rid of stop-words, contractions, and two-letter long words which would skew the similarity calculations. I approached the problem by reading a sentence, and getting a overall similarity vector for the sentence read using the individual vectors for the words in the sentence with corresponding coefficients depending on the part-of-speech of the word. (A noun or a verb would add more to the meaning of the sentence compared to a proposition or a modal word.) Then the mean of the similarity vectors would be compared to the latest similarity vector in between the last and second to last sentence read. The decision to jump would be given if the last similarity was smaller than the past mean similarity. if the similarity was greater but was smaller than a ceiling threshold (as too close of a meaning should signal for a jump ) or if it was under an "interest" threshold (which inhibited the jump as the upcoming sentence might be interesting) the model kept reading.

The length of the jumps were in terms of sentence and it was decided with sampling from a normal gaussian distribution curve with a mean of the last similarity, so that if the similarity was greater at the end, the probability for a bigger jump would be higher.

One morning, when Gregor Samsa woke from troubled dreams, he found himself transformed in his bed into a horrible vermin. He lay on his armour-like back, and if he lifted his head a little he could see his brown belly, slightly domed and divided by arches into stiff sections. The bedding was hardly able to cover it and seemed ready to slide off any moment. His many legs, pitifully thin compared with the size of the rest of him, waved about helplessly as he looked. "What's happened to me?" he thought. It wasn't a dream. His room, a proper human room although a little too small, lay peacefully between its four familiar walls. A collection of textile samples lay spread out on the table - Samsa was a travelling salesman – and above it there hung a picture that he had recently cut out of an illustrated magazine and housed in a nice, gilded frame. It showed a lady fitted out with a fur hat and fur boa who sat upright, raising a heavy fur muff that covered the whole of her lower arm towards the viewer. Gregor then turned to look out the window at the dull weather. Drops of rain could be heard hitting the pane, which made him feel quite sad. "How about if I sleep a little bit longer and forget all this nonsense", he hought, but that was something he was unable to do because he was used to sleeping on his right, and in his present state couldn't get into that position. However hard he threw himself onto his right, he always rolled back to where he was. He must have tried it a hundred times, shut his eyes so that he wouldn't have to look at the floundering legs, and only stopped when he began to feel a mild, dull pain there that he had never felt before. "Oh, God", he thought, "what a strenuous career it is that I've chosen! Travelling day in and day out. Doing business like this takes much more effort than doing your own business at home, and on top of that there's the curse of travelling, worries about making train connections, bad and irregular food, contact with different people all the time so that you can never get to know anyone or become friendly with them. It can all go to Hell!" He felt a slight itch up on his belly; pushed himself slowly up on his back towards the headboard so that he could lift his head better; found where the itch was, and saw that it was covered with lots of little white spots which he didn't know what to make of; and when he tried to feel the place with one of his legs he drew it quickly back because as soon as he touched it he was overcome by a cold shudder.

As seen, semantically close words tend to cluster together, because the way the glove embeddings were trained was with regards to where each word statistically stood with respect to one another.

The results after running the model on Franz Kafka's Metamorphosis: you can see how the model gives up on reading after the first two sentences and the jump value the model generates makes it land where the topic has changed from the transformation to talking about what is on the table and Samsa himself. Even though the interpretations for the performance of the model can be vastly subjective, given the subjectivity of the act of skimming itself, it can be said that the model is behaving persuasive at the very least.



A visualization for understanding word vectors (embeddings):

300-dimensional vectors reduced to 2 dimension and graphed by PCA.

## References

Hills, Thomas T. "Animal Foraging and the Evolution of Goal-directed Cognition." Cognitive Science (2006), 30, 3-41.
Hills, Thomas T. "Optimal foraging in semantic memory." Psychological review 119 2 (2012): 431-40 .
Yu, Adams & Lee, Hongrae & Le, Quoc. (2017). Learning to Skim Text.
Mikolov, Tomas; et al. (2013). "Efficient Estimation of Word Representations in Vector Space".
Duggan, G. B., & Payne, S. J. (2011). Skim reading by satisficing: evidence from eye tracking. In CHI '11 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1141-1150). New York: Association for Computing Machinery.
https://doi.org/10.1145/1978942.1979114