Project Title : Cues to Use

Team Members :

Raghav Chegu Shyam Kumar: racheg@iu.edu

Shefali Mahendra Luley: sluley@iu.edu

Tanay Sainath Kulkarni: tankulk@iu.edu

Abstract:

It's fascinating to see what words a person remembers when given a certain term, since recalling certain words gets progressively difficult as the context or scenario in which the person is in changes. We will aim to understand how a person remembers specific words based on cues to other words in this project, as well as explore certain patterns in how a human memorizes and correlates words with one another.

Introduction:

How do we humans normally map some words to other words that may or may not have the same meaning but are related in some way? We don't consciously consider these words, but we inadvertently deduce or recollect them through other words. The name "Cues to Use" implies that there are some cues that serve as pointers for memorizing specific words. This provides an intriguing application case for research into how people think and correlate words with one another. This project's dataset includes not just word associations, but also information on other properties of these words. The scope of this project goes beyond just understanding association patterns; there are many verticals of network science, such as centralities and community detection, that can help us understand how different cues might stick together, or which words attract the most usage amongst all cues or recall words. Finally, the project is an application that may be mapped to several fields of psychology, linguistics, and neuroscience in addition to network science.

This study will attempt to address two key questions in order to better grasp the concept of word transitions:

- Is it possible to understand the starting letters or the length of the remembered word based on the first or final letter or the sound of the cue word?
- Is it possible to determine any additional properties of the recalled word based on the length of the first word?

Motivation:

The primary aim for pursuing this study is to better understand the human cognitive process and to investigate why a person might recall specific words given a cue word in a specific context or location.
This leads to an intriguing question: **Can we anticipate what words will be recalled given a word, and can we know the qualities of those words?**

Despite the fact that this is a very subjective topic, we may separate it into numerous pieces to study different factors influencing the word associations. It has always been intriguing to discover what a person would remember in various settings, and attempting to forecast via analysis adds to the excitement. There are numerous models that currently do this, and we want to be a part of breaking down these boundaries by framing these concerns as a network problem. One of the reasons for this effort, among others, is to use this data to better understand human behavior or thinking in a specific context. As a network problem, it becomes simpler to analyze the structure and interconnections of words by generating communities, shortest pathways, and clustering, among other features that may be discovered.

Data:

The data consists of 5,019 normed words and their 72,176 responses.Each file has 31 data fields, resulting in a total matrix size of 31 columns by 72,176 rows when pooled over all beginning letters. In this matrix, there are potential data entries for 2,237,456 cells.Since, the data is very large, it has been divided into smaller files - eight letter groups: A-B, C, D-F, G-K, L- O, P-R, S, T-Z. We are going to approach it in the same manner to get meaningful communities and graphs.

Each file's first column or field lists the normed words or Cues in alphabetical order, while the second field lists their answers, or Targets. The cues and their replies (targets) are delivered in pairs in this style because of which cue-target pairings are chosen for use in memory study, we refer to them as cue-target pairs. Cues are used to prompt the recall of targets, which are chosen as words to be researched in memory experiments. The norms are used to create lists of pairs that systematically vary in some aspects while holding other values constant, given the large range in word attributes.

In the datasets, it can be seen that there are a lot of missing values for many of the features. These need to be cleaned and then used for any network being formed.

Dataset Link: http://w3.usf.edu/FreeAssociation/AppendixA/index.html

Approach:

Given our present datasets and the requested sub-questions, we want to redefine our datasets with new characteristics to answer all of the defined questions. Our approach will be more focused on identifying the many centralities in these networks. Using these centralities, we will look at a few key nodes and try to understand how they interact with all of their neighbors/co-nodes.

Milestones:

With our proposal, we established our milestones, which included preparing the dataset and then gleaning insights for the given topics. The following is a general overview of all of our milestones -

1. Proposal for a project
2. Dataset preparation (Data Cleaning and Integration)
3. Selecting the most relevant features for this project
4. Including the appropriate features and labels in the networks
5. Visualizing the formation of networks with varied centralities and communities.
6. Obtaining insights from networks by utilizing several major nodes and reaching a conclusion

Methods:

We will use numerous approaches to check for various characteristics of the data, analyze the structure, and the interactions between them. Some of the approaches are indicated in the table below -

- Python libraries for pre-processing data and generating it as a network.
- Creating the network, updating features, and deleting them using networkx packages.
- Resetting the node properties to maintain a specific centrality.
- The Gephi program is used to visualize the network and obtain a better understanding of the various network centralities and how they might be exploited.

References:

Nelson, McEvoy & Schreiber:
http://w3.usf.edu/FreeAssociation/AppendixA/index.html