Trang Nguyen

Remesh Take-home Assignment

April 13, 2021

**Background**

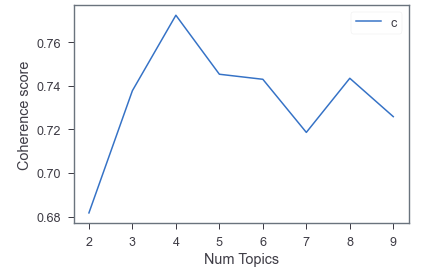
Having open-ended responses data sometimes is a challenge for use to get the thorough ideas of what features our users are looking for when they use social media platforms. However, some responses would have similar contents even though the words used in the answers were different. Hence, applying topic models to the text dataset we have could be the solution to this problem since it is highly valuable to businesses to know what people are discussing and understand their demands and opinions. Topic models provide a simple way to analyze large volumes of unlabeled text. It clusters the words that often occur together and uses contextual clues to connect words with similar meaning and gives us comprehensive “topics.” With the new topics we have, we then can validate the responses using the answers from participants preferences associated with the open-ended responses text.

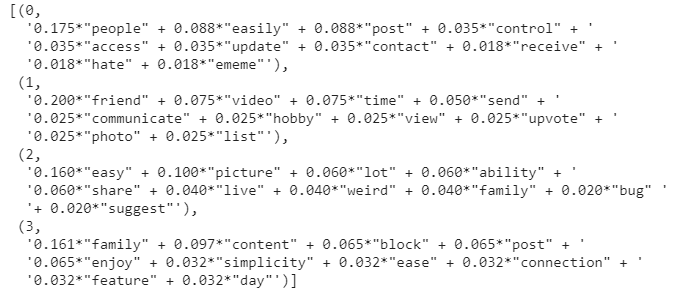
**Methods**

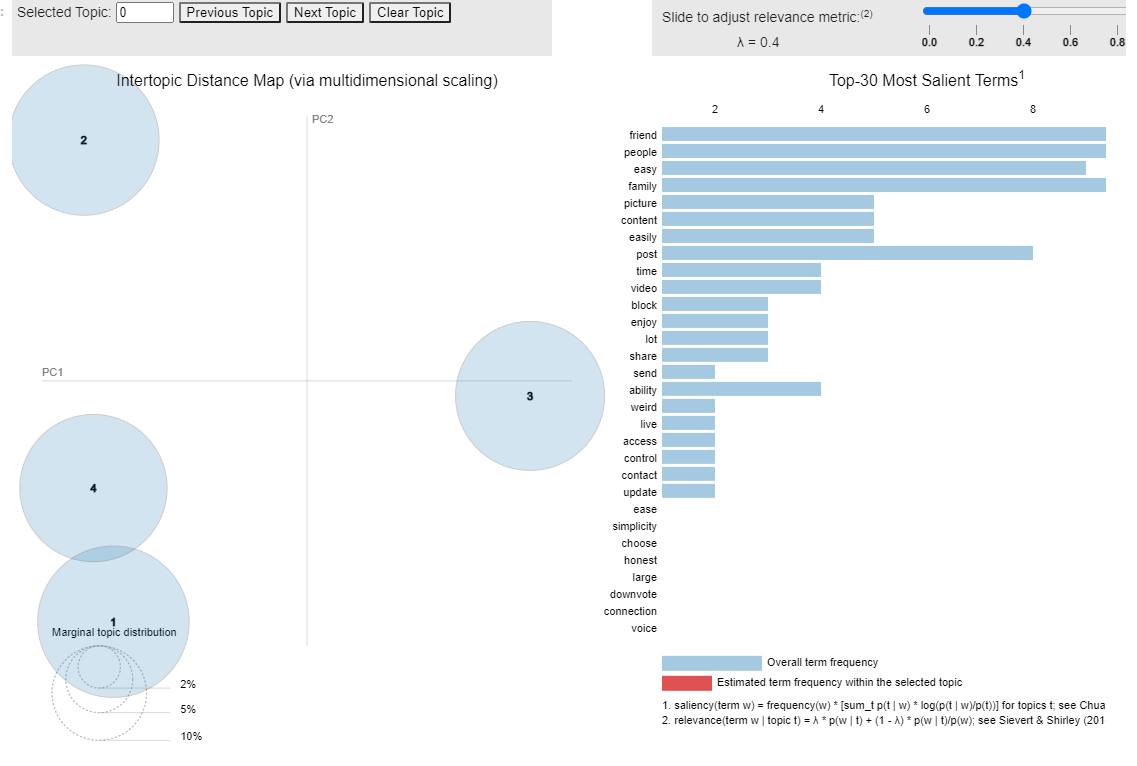
One of the popular algorithms for topic modeling is Latent Dirichlet Allocation (LDA). LDA considers each document consists of a combination of topics, and each topic consists of a combination of words in a certain proportion. For instance,

* Document 1: Topic1 = 0.33, Topic2 = 0.33, Topic3 = 0.33
* Topic1: Psychology = 0.25, Computer Science = 0.36, Music = 0.39

LDA is







Analysis

