Final Project Proposal - Predict the trending topics in field study of depression

Guided by Dr. Christopher J. Asakiewicz

Fang Zhang, Jhao-Han Chen

Component	Description
Business Problem	Finding a hottest topic which can attract the reader's curiosity is the main goal of publishers and also the journal author. Therefore, predicting the topic in certain domain is the main target for them.
Business Question	To find the hottest trending topics, we want to find the most popular topics for a bag of documents from web of science databases in a time manner. Then we can predict the future trend in form of topics by analyzing the Term-Frequency and time series.
Data & Scope	We decide to download the data from Web of Science. The data include 447310 records and 72 variables from 1972 to 2019.
Project Approach	From the text data scraped from website, we firstly apply the topic modeling to extract most related topics each year (Each year's topic number is equal). Secondly, we manually add the number for topic each year as popularity index by counting the frequency of bag of words for each topic. Thirdly, having the DataFrame with topic, year and popularity index prepared, we observe and predict the trending topics by applying time series method. Proper table and chart will be used for the project.
Anticipated Results	In prediction results, the topics with the highest popularity index will be the most trending topics. The result is based on the dimensions of word frequency distribution and time fact.

Main Reference:

Topic discovery and future trend forecasting for texts,
https://link.springer.com/article/10.1186/s40537-016-0039-2#Fig1

NLP: Extracting the main topics from your dataset using LDA in minutes,
https://towardsdatascience.com/nlp-extracting-the-main-topics-from-your-dataset-using-lda-in-minutes-21486f5aa925