Topic Modelling (using LDA and LSA) and Sentiment Analysis on ‘Climate Change Tweets’

Coursework Proposal

Umadevi Areekkara Kizhakkemana

Project Topic:

Climate change is global issue that is affecting everyone all around the world. The shift in climate can be natural, but studies proven that, since last many years, human activities have been the main driving cause for the drastic and sudden climate changes. This coursework aims to identify ‘what people are thinking about the climate change’, by analyzing the tweets about climate change.

Dataset:

There are many different twitter datasets available in public domain for analysis purpose. I am planning combine two of them together to get latest trends of tweets with climate change hashtags.

1. <https://www.kaggle.com/edqian/twitter-climate-change-sentiment-dataset>
   1. This data set contains 43943 tweets pertaining to climate change collected between Apr 27, 2015, and Feb 21, 2018. Each tweet is labelled as one of the following classes:

2 (News): The tweet links to factual news about climate change

1 (Pro): The tweet supports the belief of man-made climate change

0 (Neutral): The tweet neither supports nor refutes the belief of man-made climate change

-1 (Anti): The tweet does not believe in man-made climate change

1. <https://www.kaggle.com/joseguzman/climate-sentiment-in-twitter>

This dataset contains 396 tweets on climate change from 1st January 2020 until 24th December 2020.

Considered this dataset to get more recent tweets on this behalf. Since here only 396 tweets are there, planning to give manual labels similar to the one available in the 1st dataset.

Tasks:

Planning to do topic modelling with Latent Dirichlet allocation (LDA) and non-negative matrix factorization (NMF) on the combined tweets from above mentioned sources and compare the results. Topic modelling results on this twitter data would give an insight on what people are tweeting about in relation to climate change.

Based on the labels (Sentiment score), planning to perform a sentimental analysis also on the dataset.