Project Name

NLP CORD-19 Text Classification

02.21.2023

—

Chandrashekhar
UNT Student

1. **Project Title and Team Members**

   NLP CORD-19 Text Classification

   Chandrashekhar Ramamurthi

2. **Goals and Objectives**

   The project's objective is to build a model to answer the question from CORD-19 Dataset about COVID-19, SARS-CoV-2, and related coronaviruses.COVID-19 Open Research Dataset (CORD-19) consist of over 1,000,000 scholarly articles,

   The CORD-19 Dataset consists of over 400,000 with full text, which makes it challenging for researchers to get relevant information for research as the dataset has grown. The project aims to create a model to apply text and data to find answers to questions within.

3. **Motivation:**

   Due to the volume of text and there is a need to create quick insight and get relevant information for researchers and medical professionals, as the is a need to keep up with the growing literature due to the rapid increase in the not of scholarly articles.

4. **Significance:**

   In this project, we will explore different classifiers on the cord 19 dataset, for the relevant topic to answer questions about the relevant research literature/research.

   Current research using the Bert model and Zero-Shot Topic Classification on CORD-19 for relevant Question and Answer

   We will explore Neural network and CNN models for classification and Answer questions.

5. **Features**

   The feature explores in the projects are

   - Pre Process of the cord-19 dataset
     - Text cleaning

    - Lemmatization and stemming.
- Text Classification.

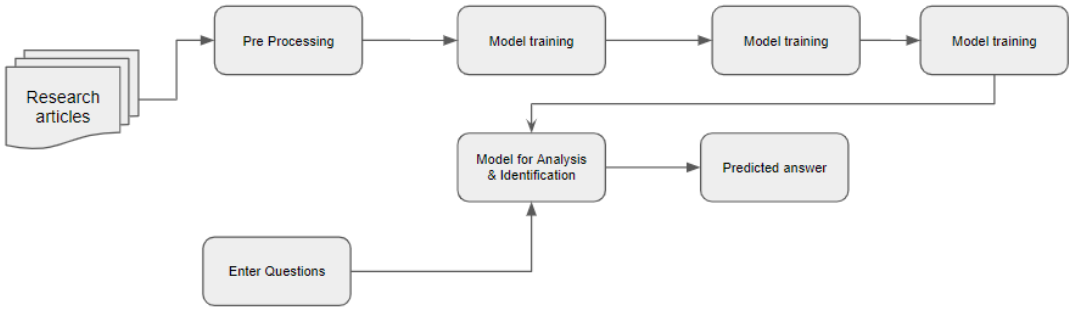
  The CNN model and Neural network model will be built to classify text and relevant information can be predicted based on the questions, Domain question answering will be explored

- Text Summarization.

  The Project will explore text summarization and from the scholarly article for the relevant topic into a few lines

  Other Optional features to be explored are

    - Named Entity Recognition.
    - Keyword Extraction.



CORD NLP - Module Architecture

NLP Cord Archtecture diagram

6. References

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7251955/

https://arxiv.org/pdf/2004.10706v4.pdf

ACL 2020 · Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Doug Burdick, Darrin Eide, Kathryn Funk, Yannis Katsis, Rodney Kinney, Yunyao Li, Ziyang Liu, William Merrill, Paul Mooney, Dewey Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shen, Brandon Stilson, Alex Wade, Kuansan Wang, Nancy Xin Ru Wang, Chris Wilhelm, Boya Xie,

https://arxiv.org/pdf/1703.03091v1.pdf

Deep Learning applied to NLP Marc Moreno Lopez College of Engineering and Applied Sciences University of Colorado Colorado Springs Colorado Springs, Colorado Email: mmorenol@uccs.edu Jugal Kalita College of Engineering and Applied Sciences University of Colorado Colorado Springs Colorado Springs, Colorado Email: jkalita@uccs.edu

GitHub link

https://github.com/ChandrashekharRamamuthi/CORD.git