Biswarup Saha
23CS60R45

COVID-19 Fake News Detection Pipeline

This pipeline automates the process of detecting fake news related to COVID-19 on social media platforms using deep learning models. The pipeline includes data preprocessing, vectorization using a specified BERT model, training a specified deep learning model (DNN or 1D-CNN), and evaluating the model's performance.

Prerequisites

Before running this pipeline, ensure you have the following installed:

- Python 3.x
- PyTorch
- Transformers library (for BERT models)
- Any other Python libraries required by the scripts (e.g., pandas for handling Excel files)

Files Description

- preprocess.py: Preprocesses the data.
- vectorization.py: Vectorizes the preprocessed data using the specified BERT model.
- dnn.py or 1dcnn.py: Trains the specified deep learning model.
- runEval.py: Evaluates the model's performance.
- run_pipeline.sh: The Bash script that automates running the above Python scripts.

How to Run

Ensure all prerequisites are installed and that the Python scripts are in the same directory as "runEndtoEnd.sh".
Open your terminal and navigate to the directory containing run_pipeline.sh.
Make the script executable by running:

Make the script executable by running:

```bash
chmod +x run_pipeline.sh
```

Run the script using the command:

```php
./run_pipeline.sh <data-excel-filepath> <bertmodelname> <modelName>
```

Replace <data-excel-filepath> with the path to your Excel data file, <bertmodelname> with the name of the BERT model you wish to use for vectorization (e.g., bert-base-uncased), and <modelName> with either dnn or 1dcnn depending on the model you want to train.

Notes

- Ensure the Excel data file is formatted correctly for the preprocess.py script.
- The model file names (dnn.py or 1dcnn.py) should match the <modelName> argument passed to the run_pipeline.sh script.
- The runEval.py script expects model output files to follow the naming convention <modelName>_<bertmodelname>.pth.

For more information on each script's functionality and required arguments, refer to the individual script files.