Jacob Devasier

5/4/2021

Claim Check Worthiness Classification

For my project, I have chosen the ClaimBuster fact-check worthiness dataset. With a growing amount of misinformation in social and news media, it is important to be able to know true news from fake news. While people are necessary to perform fact checks as of right now, it would be much better to be able to fact check information in real time. The typical pipeline of the fact checking process is as follows: select a claim to fact check, gather relevant information, and then determine whether the claim is supported by the information. For this project, I choose to focus on the claim selection process as determining which claims to fact check is the largest bottleneck and requires a very good filter. Thus, I will use the ClaimBuster dataset, which contains roughly 25,000 claims and their respective check-worthiness score.

The app will ideally be used in addition to any and every form of media, e.g., news media, radio stations, social media apps, and so on. It can be integrated with these forms of media in two ways: first, and easiest, all text data, i.e. social media posts, news station transcripts, and any other scripted content can be passed through the apps API; and second, for video and audio data, the data will have to be processed through a high-quality speech-to-text translation algorithm, ideally fine tuned on news media so that when the algorithm hears “CNN” or “FaceBook” it will know it is talking about “CNN” instead of “seeing” or the app “Facebook” instead of “face book”.

The primary challenge I had with this project is determining which layers to freeze on the BERT model. From my prior understanding part or all of the BERT model should be frozen so that only the classifier will be trained on the outputs of the model. As I found out through more research, this is not the case and my misunderstanding was clarified by Jacob Devlin, one of the main authors of the BERT paper, on a github question (that I am unable to relocate). Following this, determining other hyperparameters to choose for optimal performance was also a challenge as the wrong parameters could lead to barely better than random results.

There are very few apps related to fact-checking as automated fact-checking has only recently become of high importance with the existence of foreign interference in national elections. Some examples of these apps or algorithms can be found at the following links:

1. https://www.aclweb.org/anthology/2020.findings-emnlp.43.pdf
2. https://idir.uta.edu/claimbuster/debates/
3. https://dspace.mit.edu/bitstream/handle/1721.1/123153/1128869084-MIT.pdf?sequence=1&isAllowed=y.

For my work I have referenced the following sources:

1. https://huggingface.co/transformers/training.html

The dataset used for this project can be found at: https://zenodo.org/record/3609356

A demonstration of my model can be found at: https://youtu.be/Y7pFdp9xOSw