Hi,

My name is Sravan Kumar pursuing my BTECH final year in Computer Science(Artificial Intelligence) at CMR Institute of Technology. I am a self-driven and passionate Machine Learning Engineer who excels at developing machine learning systems. Possessing strong attention to detail, exceptional Data Analytics, Data Science and Machine Learning skills, and the ability to build complex machine learning models using Python programming language. Unmatched abilities to identify, understand and translate program requirements into sustainable advanced technical solutions through Machine Learning, Deep Learning, and Artificial Intelligence.

Comming to assignment,

I have considered this problem as an Unsupervised learning problem. So I have come up with this proposal discussed below.

I divided the problem into two parts.

Part 1: Using Clustering methods to label the samples based on the type of cluster it belongs to.

Part 2: Then, as the data turns out to be labeled. Now, I used the siamese network to determine the similarity between two sentences.

In detail

Part 1:

I have used the MiniBatchKMeans clustering method to cluster the samples after encoding them using tokenizers. Now, I have used those clusters to label two pairs of sentences based on the cluster they fall in. If they fall in the same cluster they somehow have the same meaning, If not there are not from the same context. This way I've created labels to classify paragraph pairs as either positive or negative.

Part 2:

I have used Char Encoder to encode each sentence. So, every piece of data is in numbers now. I made Siamese Network has references to build this model. As the Siamese has a good performance in finding the similarity between two vectors. We got data with labels, we got two paragraphs to find the similarity. That says we are all good to build the siamese network. I have attached siamese network architecture with my Jupyter notebook. Please observe the image to understand better. It gives an idea of how I built this model based on the siamese network.

Thanks and regards,

Sravan Kumar(AI Enthusiast)