Exercise 2
International Summer Course NLP

Prepared by Sneha Oram (23M2159)

July 6, 2024

Introduction

Part-of-speech (POS) tagging is one of the fundamental tasks in NLP that involves assigning each word in a sentence its corresponding part of speech. Parts of speech include categories such as nouns, verbs, adjectives, adverbs, pronouns, conjunctions, prepositions, and interjections. It is essential for various NLP applications such as parsing, named entity recognition, machine translation, and information retrieval. In this exercise, we have used 18 POS tags.

Given

You are provided with a Python code that takes a pre-trained model called BERT (Bi-directional Encoder Representations from Transformers) and fine-tunes it for a downstream task which is POS tagging here. For taking a pre-trained model, it is essential to take the correct tokenizer and model combo, and then use the custom dataset for fine-tuning. In this code, we have taken 'bert-base-uncased' checkpoint (BERT base model) to load the tokenizer and the model.
Essential data and tags are loaded in the code. You are advised to take up the following tasks:

- 1. Complete two functions train() and evaluate().
- 2. Use different values of epoch (i.e., iterations) from the Hyper-parameter space given. The correct output you should obtain is [a_DET quick_ADJ brown_ADJ fox_NOUN jumped_VERB over_ADP the_DET lazy_ADJ dog_NOUN]

Few Instructions

Please write your code in the designated spaces provided. See for the text – 'Write your code here'. For training use different combinations of epochs and learning rate. Please do not edit in anywhere else in the code.
Take care of the dependencies, commands to install transformers, torch, torchtext are given. Run all cells and scroll down to find the train() and evaluate() functions. Use the inbuilt GPU provided in the Kaggle platform. Go to More Settings - Accelerator - GPU P100. If you are using Google Colab go to Runtime - Change Runtime type - T4 GPU.

Reference

Neural Networks: Zero to Hero
Part Of Speech POS Tagging: NLP Tutorial For Beginners - S1 E11
How Part-of-Speech Tag, Dependency and Constituency Parsing Aid In Understanding Text Data?
nlp_made_easy Pos-tagging with Bert Fine-tuning