NLP with Deep Learning
Project Proposal - Mini Project 5

Name: Poojith Reddy Maligireddy
UID: u1405749

Project Description:

In this mini project, I am going to build a machine translation system using Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) cells with PyTorch and torchtext. Machine translation is a NLP task that involves translating text from one language to another language. Seq-to-Seq also has other applications such as text summarization, POS tagging and so on.

This project is relevant to the class as it applies deep learning techniques (RNN with LSTM cells) to a real-world NLP problem (translating text from one language to another (German to English in our mini project)).

Project Goals:

To begin with, I am going to implement a sequence-to-sequence model for machine translation using RNN LSTM architecture. Next, train model on parallel text data, where sentences in one language are paired with their translations in another language and evaluate model performance using appropriate metrics such as BLEU score or Perplexity and then fine-tune the model to achieve high performance with necessary data pre-processing steps like what we are doing in our mini projects. In addition to that, I would like to do some architecture changes like including attention, using BiLSTM or GRU instead of LSTM and compare those models and also extend it to other languages, if time permits.

Data:

I am going to use a bilingual parallel text dataset (German-English) named Multi30K (extension of the Flickr30K dataset) for this machine translation task and below is the link about it.

https://aclanthology.org/W16-3210.pdf

Model Evaluation:

I am going to evaluate the model's translation performance using either of the following metrics:

(i)BLEU score
(ii)Perplexity