

Homework 4: Machine Translator

Deadline: 11-27-2023

***GPUs may be needed for speeding up this training process. You are allowed to use pre-defined models.**

Description

In this homework you will practice application of sequence model in Python with TensorFlow or PyTorch. You need to understand how sequence model works in order to implement this homework successfully. The goals of this homework are:

- To understand the steps to train/test the translation model.
- Understand architecture of sequence model and how to connect each layer together by using TensorFlow or PyTorch.
- Be able to evaluate your model.

Data preview:

The goal of this project/challenge is to translate between two languages.

You can use any dataset to train your translate model. Below is reference code contains dataset:

Reference code:

Keras:

https://keras.io/examples/nlp/neural_machine_translation_with_keras_nlp/#evaluating-our-model-quantitative-analysis

PyTorch:

https://pytorch.org/tutorials/intermediate/seq2seq_translation_tutorial.html#loading-data-files

Step 1: Data

Before starting to train a model, please get familiar with the dataset. When you look at the dataset, please answer the following questions: 1) How many data samples are included in the dataset? 2) Which problem will this dataset try to address? 3) How many words does the dataset contains? 4) Does the dataset have any missing information? E.g., missing features. 5) What is the label of this dataset? 6) How many percent of data will you use for training, validation and testing? 7) What kind of data pre-processing will you use for your training dataset?

Hint: You should use the same test dataset to compare the models' performance.

Step 2: Model

Select any sequence model to train your machine translator model. What evaluation matrix did you used to evaluate your model?

Model		

Step 3: Objective

Which loss function are you going to use to train your model?

Step 4: Optimization

You are going to select your optimization function to train the models. Report the optimization you selected in this section and explain the reason for using the optimizer.

Step 5: Model selection

Based on your training experience, which model gives the best performance. Have you experienced different learning rates?

Model			

Step 6: Model performance

In this step you should report your model performance, which you did in the previous steps.

- A. Report the performance plot of models you tried.
- B. Show a translate result (the translate could be incorrect):

What should you submit?

You should submit a zip file (named as firstname_lastname_hw3.zip – e.g. jun_bai_hw2.zip) containing:

1. Your homework report from step 1- 6. You will answer all the questions in each step and fill the tables in step 2, step 5 and performance plot in step 6. Missing any part will lose some points. Please double check you have addressed all the questions.

2. Your code of all models and a README file explaining how to run the model. Your code should be well commented. In your code, you should have a function called *test_model*. The *test_model* function will load the trained model and load test dataset to predict.
3. A folder contains screenshots of iteration of models' training and testing.