# PRML Assignment2

**Mengyi Chen**
CS
ID: 19307110382
19307110382@fudan.edu.cn

## Abstract

Assignment2 is about seq2seq model with attention and model attack.

## 1 Introduction

This assignment implements Transformer as the seq2seq model to perform the Chinese to English machine translation task. The corpus is the News Commentary v13 dataset from the Third Conference on Machine Learning (WMT 18). Test set is evaluated with Bilingual Evaluation Understudy (BELU) score and Perplexity. Besides, train data is poisoned to implement model attack.

## 2 Dataset

The News Commentary v13 dataset from the Third Conference on Machine Learning (WMT 18) is used in this task. There are 252777 training samples, 2002 validation samples and 2001 test samples in the dataset. The size of Chinese vocabulary is 93264 and the size of English vocabulary is 166192. However, only 5000 samples of the train set is used in training due to the extremely slow speed of the train procedure.

## 3 Methodology

### 1. Machine Translation

Based on the paper *Attention is All You Need*, the Transformer is implemented as the seq2seq model. The encoder consists of multi-head attention layer while the decoder consists of both masked and non-masked multi-head attention layer. Word vectors are randomly initialized.

### 2. Model Attack

Model attack is implemented by poisoning train data. Traverse through the train set, substitute every ’friend’ with ’enemy’ in target data,. Then the model will translate the Chinese word ’PengYou’ into English word ’enemy’, doing the wrong translation.

## 4 Results

The PPL on the test set is 230.510 and the BLEU score is 0.22. Substituting random embedding with pre-trained embedding ’glove.6B.300d’ only make the result slightly better.

## 5 Conclusion

Through this workout, we use models pre-build in torch to implement machine learning tasks, getting familiar with the procedure of Machine Translation. Besides, we learnt that there are several types of model attack which could do great harm to our model. We should be more careful about defensing against these attacks.