Application of ML in Industries Lab

Experiment 7

MOSES & OpenNMT

Submitted By:

Dhruv Singhal

500075346

R177219074

AIML B3

Submitted To:

Mr. Goutam Datta

Asst. Professor

SOCS

UPES

MOSES:

(Statistical machine translation (SMT) is a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora.

The statistical approach contrasts with the rule-based approaches to machine translation as well as with example-based machine translation) Moses is a statistical machine translation system that allows you to automatically train translation models for any language pair. All you need is a collection of translated texts (parallel corpus).

Once you have a trained model, an efficient search algorithm quickly finds the highest probability translation among the exponential number of choices.

OpenNMT

- (Neural Machine Translation i.e. NMT provides more accurate translation by accounting the context in which a word is used, rather than just translating each individual word on its own.)

- OpenNMT is an open source ecosystem for neural machine translation and neural sequence learning. It is designed keeping in mind the code modularity, efficiency, extensibility.

OpenNMT provides implementations in two popular deep learning frameworks, Pytorch and Keras –

- OpenNMT-py: This is implemented using pytorch deep learning framework and we are going to use this only. It is extensible and has fast implementation with PyTorch ease of use.
- OpenNMT-lf: It is based on tensorflow deep learning framework.

Each implementation has its own set of unique features but shares similar goals:

- Highly configurable model architectures and training procedures
- Efficient model serving capabilities for use in real-world applications
- Extensions to allow other tasks such as text generation, tagging, summarization, image to text, and speech to text.

At its core it has support for different techniques for machine translation from vanilla NMT to attention, gating, stacking, input feeding, regularization, copy models, beam search and all other options necessary for state-of-the-art performance.

Neural Machine Translation (NMT) is the default and original task for OpenNMT. It requires a corpus of bilingual sentences for instance as available on Opus for a very large variety of domains and language pairs.

Training a NMT engine is a 3 steps process:

- Tokenization
- Preprocessing
- Training