

Report #2

Code the solution, fine-tune & validate & test the model

Coding process:

The most complex part of the whole solution-creating task.

Before final decisions about models and libraries several plans were tried “on the paper” as “probable solutions”. One of them was to code a cascade of models: first of them will give probabilities of the sentence parts to be toxic and the second model will paraphrase them in an effective manner.

Another one was to use some single encoder-decoder models as Seq2Seq text generators to perform paraphrasing. After some trial-testing it was decided to use the 2nd tested plan and not “to reinvent a bicycle” since such models as ParaGeDi already exist.

As final models to fine-tune it was decided to use Bart-base-4096 and T5-small(mostly because both of them were fine-tuned locally). Since the models were chosen it was time to code all necessary steps with the use of built-in function, methods, and structures from the “transformers” library.

The only part that needed to be implemented was metric. It was decided to use [Bleu score](#) (probably not the best choice, but still) and all necessary code such as output preprocessing, metric calculation, and logging was written.

At that moment all necessary code for model loading & constructing, data preprocessing, metric calculation was written.

Fine-tuning & validation & testing the model:

It was found that Seq2Seq models in “transformers” library can be trained with the use of built-in structure “Seq2SeqTrainer” and all necessary hyperparameters as: optimizer, lr, lr scheduler, warmup steps, weight decay rate, and much more are aggregated as arguments for such a structure.

Several learning rates were tested: $1e-3$, $3e-4$, $2e-5$, $1e-5$. In cases with the first two learning rates in some training processes the loss between model predictions and original translations were starting to increase. With learning rate of $1e-5$ and $2e-5$ all training processes were consistent, and for “speed up” value “ $2e-5$ ” for learning rate were chosen to be the final one.

Weight decaying rate was chosen to be equal to $1e-6$.

Optimizer was chosen to be “AdaMax”

LR scheduler type was chosen as linear

For all models predictions making process was through the “generate” method of the models.

Number epochs for the encoder-decoder model that was constructed from T5-small models was chosen to be equal to 10. For Bart-base-4096 this number was 4.

After the training & validation phases all models were tested by generating paraphrased sentences with hand-written toxic ones as input.

Results:

All training processes were consistent: loss was decreasing, BLEU score was increasing. The only difficulty was to wait until the end of training & validation phases.

BLEU score was > 30 for both models that is theoretically considered as “[Understandable to good translations](#)”.

Most of the tests were successfully passed by both models but still in some cases with high toxicity the quality of translations was pure.