deepQuest-py: Large and Distilled Models for Quality Estimation

Fernando Alva-Manchego, Abiola Obamuyide, Amit Gajbhiye, Frédéric Blain, Marina Fomicheva, Lucia Specia

EMNLP 2021 Demo
7-11 November 2021
Quality Estimation of Machine Translation

- To predict how good automatic translations are **without comparing them to gold-standard references**

![Diagram showing the process of quality estimation in machine translation](image)

- **Sentence-Level**
  - Direct Assessments of Adequacy
  - Post-Editing Effort: 0.75

- **Word-Level**

Original Sentences → (Neural) Quality Estimation Model → Automatic Translations

anschließend wird in jeder Methode die übergeordnete Superclass-Version von selbst aufgerufen.

Labels for each token (e.g. ‘OK’ or ‘BAD’)

![Image of quality estimation model](image)
# Open-Source Software for Quality Estimation

|                      | QuEst | QuEst++ | deepQuest | OpenKiwi | TransQuest |
|----------------------|-------|---------|-----------|----------|------------|
| Large SotA Architectures | ✓     | ✓       | ✓         | ✓        | ✓          |
| Lightweight models    | ✓     | ✓       | ✓         | ✓        | ✓          |
| Pre-trained QE Models | ✓     | ✓       | ✓         | ✓        | ✓          |
| Accessible for researchers | ✓   | ✓       | ✓         | ✓        | ✓          |
| Accessible for end-users (e.g. translators) |       |         | ✓         | ✓        | ?          |
Introducing deepQuest-py
Large State-of-the-Art Models

Tharindu Ranasinghe, Constantin Orasan, and Ruslan Mitkov. 2020. TransQuest: Translation Quality Estimation with Cross-lingual Transformers. In Proceedings of COLING 2020.

Dongjun Lee. 2020. Two-Phase Cross-lingual Language Model Fine-tuning for Machine Translation Quality Estimation. In Proceedings of WMT 2020.

Best models of WMT20
Light-Weight Distilled Models

Sentence-level (Gajbhiye et al., 2021)

Amit Gajbhiye, Marina Fomicheva, Fernando Alva-Manchego, Frédéric Blain, Abiola Obamuyide, Nikolaos Aletras, and Lucia Specia. 2021. Knowledge Distillation for Quality Estimation. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021.

Smaller and faster models!

| Model         | #params | Speed (secs.) | RAM (MiB)   | Disk (M) |
|---------------|---------|---------------|-------------|----------|
| TransQuest_{XLM-R-Large} | 561M    | 0.82          | 9,263.5     | 2,140    |
| TransQuest_{DistillBERT}   | 135M    | 1.09          | 1,979.2     | 517      |
| BiRNN         | 18M     | 0.39          | 155.6       | 132      |

| Language Pair | Et-En | Ro-En | Si-En | Ne-En | En-Zh |
|---------------|-------|-------|-------|-------|-------|
| BiRNN_{Student+Aug} | 0.50  | 0.69  | 0.45  | 0.54  | 0.17  |
| TransQuest_{DistillBERT} | 0.62  | 0.78  | 0.51  | 0.61  | 0.36  |

Trade-off performance-size
Command-line Interface

CUDA_VISIBLE_DEVICES=1,2,3 python "${deepquestpy_dir}/deepquestpy_cli/run_transformer.py" \
   --model_name_or_path "${pretrained_model_dir}" --arch_name "transformer-word" \
   --dataset_name "${dataset_name}" --src_lang "en" --tgt_lang "de" \
   --label_column_name_src "none" --label_column_name_tgt "bad_labels" \
   --do_train --train_file "${train_file}" \
   --num_train_epochs 1 --save_steps 10000 --save_total_limit 1 \
   --learning_rate=5e-6 \
   --output_dir "${output_dir}" --overwrite_output_dir \n
For researchers
Web Tool for Analysis and Visualisation

For end-users

(1) Select the languages of the original (source) and translated (target) texts

Language
Select the language of the source and target text.
- Source Language: English
- Target Language: German

(2) Directly type the sentences to analyse (one per line), or upload a .tsv file (one sentence pair per line)

Direct Input
Use the text boxes below to submit source-target text pairs to the translation scoring system. If you would like to compare pairs of words or sentences then separate them using line breaks and select the appropriate "Level" setting in the options below.
- Source Corpus
- Target Corpus

File Upload
Upload translation text in tab-separated values (TSV) format. Do not include column headers.
- Corpus File: Choose File

(3) Select the model to use: Transformer or BiRNN (distilled)

Options
Model
- Transformer-based

(4) Choose whether to use a sentence-level or word-level model

Sentence
To Summarise...

| Large SotA Architectures | QuEst | QuEst++ | deepQuest | OpenKiwi | TransQuest | deepQuest-py |
|--------------------------|-------|---------|-----------|----------|------------|--------------|
| **Lightweight models**   | ✔     | ✔       | ✔         | ✔        | ✔          | ✔            |
| **Pre-trained QE Models**| ✔     | ✔       | ✔         | ✔        | ✔          | ✔            |
| Accessible for researchers | ✔   | ✔       | ✔         | ✔        | ✔          | ✔            |
| Accessible for end-users (e.g. translators) | | | | | | ✔ |
Thanks!

https://github.com/sheffieldnlp/deepQuest-py

Fernando Alva-Manchego

@feralvam

https://feralvam.github.io/