ECL 1 Exercise 6

Topic: Machine Translation and MT Evaluation

Submission Deadline: Saturday, 30 December 2023, 18:00

Notes on the submission:

- **Submission file format: PDF**
- Filename in the following format:
  olatusername1_olatusername2_ecl1_exerciseNo.,
  e.g. *hmuster_mhuster_ecl1_exercise06*
- **Languages:** You can write either in English or in German.
- Please number the tasks on the submission sheet in the same way as on the task sheet.
- **Working in pairs** is **required**. Only one person from each group should upload their submission on OLAT.
- Submit your file via the exercise module on OLAT. Please hand it in on time, the module is only open until **Saturday, December 30** at **18:00**.

Task 1: Required Reading

Read chapter 10 "Machine Translation" of the book by Dan Jurafsky and James H. Martin "Speech and Language Processing" (Jan. 2023 Draft) that we provide for you in OLAT[1]. (You may skip the sections 10.3 "Details of the Encoder-Decoder Model"). Then answer the following questions:

1. What are 3 typological differences between languages that pose challenges for Machine Translation? Give an own example for each.
2. Describe the Encoder-Decoder Model in your own words and illustrate, with an invented example, how it works.
3. What is the difference between adequacy and fluency in MT evaluation? Give an example of an adequate but not fluent translation.

Submission: Write down your answers on your submission sheet **PDF**.

Task 2: Sentence Alignment

OPUS is a large, open collection of parallel corpora. From the TED Talks 2020 collection on OPUS, choose a parallel corpus for a **language pair of your choice**[2]. Make sure you choose a language pair with which you can fulfil the assignment, your fluency in both languages is recommended.

The texts and the sentence alignments are available in various files and formats. For example, you can find XML files in the CES-format, which only contain the indices of the aligned sentences for a particular language pair, whilst the corresponding sentences are found in another XML-file.

In order to keep things simple with these different files and not turn this into an programming task, we suggest the following:

- Choose a language pair for which there is a 'view' entry in the upper right triangle of the topmost table of pairs.
- Click on the button. A window will pop up where you see the first 100 sentences and their alignment: Each sentence group between two horizontal lines is supposed to be aligned between source and target language.

[1]The whole book is also available online (here)
[2]The languages are abbreviated with their ISO Language-codes(see here)

The 100 sentences you see there are automatically aligned. Inspect them and solve **only two** of the following subtasks. Your options are: **a & b**, or **a & c**:

a) How many out of these 100 alignments are wrong?

b) If there are cases where the sentence alignment is wrong:

- Include them in your submission in a readable manner.
- Translate them if it is reasonable to assume that we are not familiar with the language(s).
- Reason why the alignment might have gone wrong in these cases.

c) Else:

- Think of some examples that might pose difficulties to an alignment system.
- include them and, if necessary, their translation(s) in the submission sheet...
- ... as well as your reasoning why they might be problematic.

Submission: Write your answers to the above questions into the submission sheet. Also, let us know what language pair you chose.

Task 3: Automatic MT Evaluation

Take approximately 50 sentence pairs from your parallel TED corpus. Since we need to have them separated by language and want to spare you some trouble, we provide you with a script for that purpose. Proceed as follows:

- Copy the sentences from the window that popped up when you clicked 'view' into a .txt file and save it.
- Open up your Python IDE (probably PyCharm or VSCode). Include that .txt file in your environment, as well as the script from the task folder (separate_langs.py). Run it.
- You should now have two separate files with the sentences in order in a string format, one after the other.

Your Task

Your task is to compare different translation systems. Translate both files with ChatGPT and one other online MT system of your choice (e.g. Google Translate, DeepL, or any other MT system) into the other

language of your pair. This will give you a **total of 4 translations**: 2 from ChatGPT and 2 from your chosen MT system.

Attention: **Don't do back translation**, i.e., feeding the MT system its own translation and have it translate it back again. Use only the original sentences from the TED Corpus for translation. Similary, when using ChatGPT to do the translation, do it **in two different chats**, in order to avoid that ChatGPT remembers the original, which would skew the results.

Save the translations in separate .txt-files. Compute the BLEU score for both MT systems and in **both translation directions** by comparing the 4 translations to the source files which you got from running the script. Use this online BLEU score calculator. On your exercise sheet, answer the following questions:

a) Which MT system gets higher BLEU scores?

b) Which translation direction gets higher BLEU scores?

c) Do the BLEU scores correspond to your human judgement about translation quality?

d) Considering these examples, is the BLEU score a satisfying metric for machine translation quality?

Submission: The answers to the questions above on the exercise sheet.

Task 4: Detailed MT Evaluation

Select 10 sentences from the parallel TED corpus that contain an interesting linguistic phenomenon [3] and have them translated without and with context (3 sentences before and after) by both ChatGPT and one other online MT system in only one direction of translation. Please also explain briefly (1-2 sentences) in your own words the chosen phenomena. Additionally, answer the following questions:

a) What are your observations?

b) How often was the phenomenon translated correctly?

c) What is seemingly difficult/easy for the system?

d) Did the results match your expectations?

[3] A couple of examples of such phenomena you can choose would be: reflexive pronouns, co-reference resolution, idiomatic expressions, register and tone, word order differences, ambiguity (polysemy, homonymy), figurative language (metaphors) or any other phenomenon which interests you.

Submission: Write your observations and answers on the submission sheets, along with the sentences and the translations.

Feedback

If you have any suggestions for the exercise or any thoughts, feel free to share them!

For any questions of general interest, please write a post on the OLAT-Forum.

All the best and have fun!