Qeios

Peer Review

Review of: "Crossing Language Borders: A Pipeline for Indonesian Manhwa Translation"

Serge Gladkoff[1]

1. University of Manchester, United Kingdom

The idea to save days of manual work and automate the translation of Manhwa is legitimate and understandable, and the overall workflow of automatic bubble detection → OCR → Machine translation is a step in the right direction. However, the chosen MT method and lack of necessary steps resulted in the substandard quality of the final work.

Manhwa is not a technical translation - it is a literary piece, and translation must be on the high end of translation quality. Cultural nuances and the brevity of speech, which is labeled in balloons on the drawing, pose additional challenges.

Unfortunately, going with Marian MT without adequate quality review results in substandard quality of translated materials, and here's why.

Incorrect Choice of MT Engine

Microsoft Marian is not an engine for literary translators; it is obsolete technology suitable for technical translations only. It does allow fine-tuning, but fluency leaves a lot to be desired. Nevertheless, even 4 years ago, Marian supported Indonesian (id):
https://gist.github.com/KevinDanikowski/a366ceea2af70b77d11d3434b3333226

Indonesian was a low-resource language many years ago. But since the number of native speakers is around 300 million (!), now there's plenty of MT engines that support Indonesian. Not only does the Marian Helsinki model itself support Indonesian, but models like IndoBERT and multilingual models (e.g., mBERT, XLM-R) are being fine-tuned for Indonesian.

So it is not clear at all why the authors tried to fine-tune Marian on some irrelevant dataset for Indonesian.

doi.org/10.32388/JIR85B

In this particular case, GenAI models should be used for translation because of the extra fluency that they offer.

Among GenAI models, in particular, DeepSeek could be recommended, although there are several other GenAI models that also support Indonesian and which provide better translation than any fine-tuned attempt of Marian.

Quality of Translation

Manhwa are creative pieces, and they must be translated with an artistic twist, taking into account cultural nuances.

Attempting to translate them with a technical MT engine will inevitably yield inadequate results. Consider the translation that is featured in the paper. A boy looks at two cats that are entering into conversation and says something.

The proposed translation in the paper is: "WOULD YOU LIKE TO OFFER SOMETHING TO ME?"

However, any GenAI engine would give a far more accurate translation, which is: "Do you guys want to show me something?"

The Marian translation is clearly incorrect because the boy is not a participant in the action; he is an observer and is thinking that the cats will do some interesting action. But the cats are not going to "offer" anything to the boy.

A translation that does not meet quality standards and the purpose is a waste of time and money because readers of this Manhwa card will be confused by this incorrect translation. This would not happen if the process did not rely solely on an automatic process.

Measurement of the Quality of Translation

This incorrect translation slipped through not accidentally, but because the authors skipped the necessary human quality assurance and validation.

Authors say that they used BLEU and METEOR to check translation quality. But BLEU and METEOR are NOT (unlike what is said in the paper) "measure how well the translation captures the original meaning." Both metrics ONLY assess the similarity of strings to the translation that is considered to be final. No meaning analysis is conducted.

Both BLEU and METEOR are very rough measures suitable ONLY to be used during training to see if the result of the next epoch is better than the previous one.

They definitely are NOT suitable for measuring final translation quality.

The only reliable method to measure the translation quality of a final translated product is human analytic evaluation with Multidimensional Quality metrics (https://arxiv.org/abs/2405.16969).

We encourage authors to incorporate MQM Metrics into their workflow.

But most importantly - the human reviewer needs to be in the loop. Because as of today, only human linguists can quickly and reliably assess translation quality.

Going completely without the human LQA/review will not allow for achieving an adequate translation outcome.

CONCLUSION

This work should be amended as follows:

1. Instead of Marian, one of the new GenAI models should be used for translation via API. No Marian fine-tuning is needed.
2. Quality must be spot-checked with the aid of the MQM scorecard.
3. The human translator needs to be in the loop and must be able to check every translation. An even better solution could be to invite an Indonesian professional translator to verify 100% of the final output before delivery.
4. It is this professional translator who is the final and ultimate judge of whether the translation fits the intended purpose and specifications.

Overall, however, the main steps of the process will be preserved, which means that many aspects of this work can be re-used.

Overall: good start, but requires major rework, as described.

Declarations

Potential competing interests: No potential competing interests to declare.