Designing a Multilingual Fact-Checking Dataset from Existing Question-Answering Data

Daniel Kamenicky

Contributions

- ❑ Thesis exploit technique for converting QA dataset into fact-checking dataset.
- ❑ Two approaches are presented for the convertion both resulting in dataset with ≈28,000 samples for training and ≈1,300 samples for validation.
- ❑ First approach is based on monolingual seq-2-seq model T5 with the accuracy rate 88 %.
- ❑ Second approach is based on multilingual mT5 model with the accuracy rate 73 %.

Introduction

The conversion of the QA dataset takes place in several steps. A sample from the QA dataset is passed to mDPR, which returns the 100 most relevant passages. These passages are then processed into the resulting 15. From these passages mGEN generates an answer. The passages are again post-processed to find evidence. The result is a question-answer-passage triplet.

Data

Language	Train	Dev	Test
Ar	18,402	708	1,269
Bn	5,010	427	267
Fi	9,768	615	1,727
Ja	7,815	433	1,316
Ko	4,325	371	1,151
Ru	9,290	568	1,360
Te	6,759	351	1,086

T5 approach

The T5 model is used to convert the triples to the final claim. Since the conversion dataset is multilingual, each part of the triple must be translated into English. The claim is generated in English and back-translated into the target language. A fact-checking dataset is then created from the converted claims, evidence and labels. Labels (refute/support) are assigned based on the generated response.

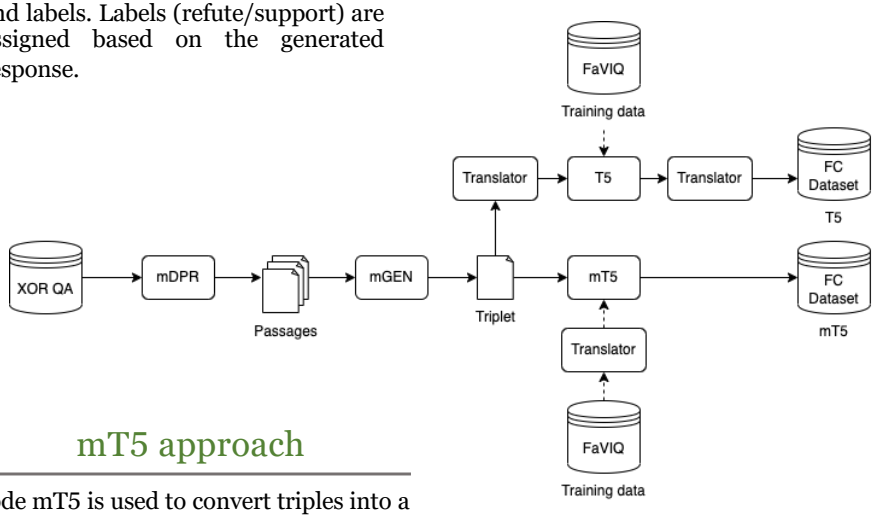
Example (T5)

An example from a converted dataset using the T5 model.

Question: Как называется самая высокая гора Абхазии?
Translation: What is the name of the highest mountain in Abkhazia?

Answer: Домбай-Ульген
Translation: Dombay-Ulgen

Claim: Домбай-Ульген - самая высокая горы Абхазии.
Translation: Dombay-Ulgen is the name of the highest mountain in Abkhazia.



mT5 approach

Mode mT5 is used to convert triples into a final claim. Since the mT5 model is multilingual, there is no need to translate the individual parts of the trinity and they can be directly passed to the model. However, to train the model, the training datasets had to be translated into all languages represented in the original QA dataset. A dataset is then created from the claim, evidence and labels. Labels (refute/support)) are assigned based on the generated response, just like with the T5 variant.

Example (mT5)

An example from a converted dataset using the mT5 model.

Question: Как называется самая высокая гора Абхазии?
Translation: What is the name of the highest mountain in Abkhazia?

Answer: Домбай-Ульген
Translation: Dombay-Ulgen

Claim: Домбай-Ульген - самая высокая горы Абхазии.
Translation: Dombay-Ulgen is the highest mountain in Abkhazia.

Official Results

Model	Metric	Score
T5	Rouge-1	85 %
	Rouge-2	78 %
	Rouge-L	84 %
	EM	19 %
mT5	Rouge-1	66 %
	Rouge-2	51 %
	Rouge-L	61 %
	EM	6 %

Model	Examples	Total Score	Score [%]	Preference [%]
T5	70	62	89	56
mT5		51.4	73	27

Table 7.5: Results of custom metrics over results from annotators.

	Precision	Recall	F1-score	# Samples	Accuracy
refute	0.54	0.19	0.28	1080	0.53
support	0.53	0.85	0.65	1145	

Table 7.9: Classification report for dataset from mT5 model.

	Precision	Recall	F1-score	# Samples	Accuracy
refute	0.58	0.18	0.28	1080	0.54
support	0.53	0.88	0.66	1145	

Table 7.10: Classification report for dataset from T5 model.

Language	ar	bn	fi	ja	ko	ru	te
Score	65 %	30 %	75 %	35 %	70 %	80 %	25 %

Table 7.3: Results from manual back-translation evaluation over 20 samples.

Conclusions

- ❑ The proposed approaches were able to show potential in converting QA dataset.
- ❑ Translation evaluation showed poor accuracy for low-resource languages. This proved to be the most critical part of the design.
- ❑ The use of better translation models for low-resource languages could refine the results.
- ❑ The claims conversions themselves were almost perfect.

BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY