Generating Extended and Multilingual Summaries with Pre-trained Transformers

Remi Calizzano, Malte Ostendorff, Qian Ruan, Georg Rehm

LREC 2022
Marseille

Qurator
Curation Technologies

Federal Ministry
of Education
and Research

🔎 Context

Summarisation Datasets:

- CNN-DM (Hermann et al., 2015) → **5-6 sentences, English**
- XSum (Narayan et al., 2018) → **1-2 sentences, English**
- WikiSum (Liu et al., 2018) → **10 sentences, English**
- Multi-News (Fabbri et al., 2019) → **11 sentences, English**
- WikiLingua (Ladhak et al., 2020) → **39 token ~ 4 sentences, English, French, German**
- MLSUM (Scialom et al., 2020) → **15-30 words ~ 3-4 sentences, Multilingual**
- XL-Sum (Hasan et al. 2021) → **1-2 sentences, Multilingual**

➔ **No multilingual summarisation dataset that is suitable to use in our extended summary setup**

🔎 Context

<u>Current Summarisation Methods:</u>

- Extractive summarisation using Encoder:
  - Summarisation layer on top of a Transformer (Liu 2019) → **short input, Multilingual**
  - Representing each sentences + clustering method (Miller 2019) → **long input, Multilingual**
- Abstractive summarisation using Encoder-Decoder:
  - BART (+ mBART) (Lewis et al., 2019) → **short input, Multilingual**
  - T5 (+ mT5) (Raffel et al., 2019) → **short input, Multilingual**
  - Pegasus (Zhang et al., 2020a) → **short input, English**
  - Longformer (Beltagy et al., 2020) → **long input, English**
  - Big Bird (Zaheer et al., 2020) → **long input, English**

➔ **No model for multilingual abstractive summarisation of long inputs**

➔ **Liu and Lapata (2019a) propose to combine an extractive model with an abstractive one**

🎯 Motivations and Objectives

➔ **Create a Multilingual dataset for generating extended summaries**

➔ **Provide strong baselines for the dataset for both extractive and abstractive summarisation**

➔ **Compare different multilingual training scenarios for summarisation**

📁 WikinewsSum Dataset

- Based on Wikinews (https://www.wikinews.org/)
- News used as the extended summary
- Text of the sources as input

Languages	# samples	# cross-lingual samples	Input Documents # words	Input Documents # sentences	Summaries # words	Summaries # sentences
English	11,616	641 (5.5%)	1,466	57	300	13
German	8,126	2,796 (34.4%)	1,179	58	241	13
French	6,200	1,932 (31.2%)	884	29	176	7
Spanish	6,116	2,137 (34.9%)	1,215	42	276	10
Portuguese	3,843	1,971 (51.3%)	1,037	38	221	8
Polish	3,630	1,214 (33.4%)	734	35	173	10
Italian	95	46 (48.4%)	1,021	35	224	8
All languages	39,626	10,737 (27.1%)	1,168	47	245	11

Table 1: Comparison of each language in the WikinewsSum dataset with regard to the number of samples, to the number of cross-lingual samples, and to the length of the input documents and the summaries.

⚙ Methodology

Extractive Models:

We use the method by Miller (2019). The method uses **a transformer model to obtain a representation of each sentence** from the input documents and creates the extractive summary using **K-Means clustering to identify the sentences closest to the centroid.**

- mBERT (Devlin et al., 2019) is pre-trained on the Wikipedias of 104 languages using masked language modeling and next sentence prediction.
- DistilmBERT (Sanh et al., 2019) is the distilled version of mBERT.
- XLM-RoBERTa (Conneau et al., 2020) is trained on Common Crawl in 100 languages using masked language modeling.
- *Oracle which estimates the upper boundary of the performance using the ROUGE scores.*

⚙ Methodology

Abstractive Models:

- mT5 (Xue et al., 2021) is a multilingual variant of T5 covering 101 languages. It uses the same architecture as T5, an encoder-decoder transformer model.

Three training scenarios (inspired by Hu et al. (2020)):

- **Cross-lingual zero-shot transfer:** We fine-tune mT5 on the English samples only and evaluate the resulting model on all languages.
- **In-language multi-task:** We fine-tune mT5 on all available samples, which results in training mT5 on the English, German, French, Spanish, Portuguese, Polish, and Italian samples, all shuffled.
- **In-language:** We fine-tune mT5 on each language separately.

⚙️ Methodology

<u>Combination of Extractive and Abstractive models:</u>

- ➢ mT5 only accepts 512 input tokens and the input documents exceed this limit in the WikinewsSum dataset.
- ➢ We use the combination method proposed by Liu and Lapata (2019a):
  - We extract the most relevant 512 tokens using an extractive method. These 512 tokens are used as input of mT5.
  - During fine-tuning, the Oracle extractive model is used.
  - We evaluate the abstractive models with the Oracle extractive model and the best extractive model.

🧪 Implementation and Metrics

Training and models

- Training code: https://github.com/airKlizz/mdmls
- Models: https://huggingface.co/models?sort=downloads&search=airklizz+mt5+wikinewssum

Metrics

- ROUGE scores: number of overlapping uni-grams (ROUGE-1– R-1), bi-grams (ROUGE-2 – R-2), and the longest common subsequence (ROUGE-L – R-L) between the generated summary and the gold summary
- BERTScore metric: cosinus similarity between the contextual embeddings of the generated summary and the gold summary computed with mBERT

Results

- ROUGE scores

Methods	Metrics	English	German	French	Spanish	Portuguese	Polish	Italian	All Languages
Extractive Summarisation									
DistilmBERT	R-1 F	41.37	29.37	29.80	29.70	29.62	24.83	35.18	33.51
	R-2 F	14.35	8.42	12.57	12.52	14.33	10.48	12.59	12.34
	R-L F	19.66	13.65	17.10	17.07	18.75	15.03	18.43	17.30
mBERT	R-1 F	41.37	29.74	29.74	35.50	29.66	24.82	34.93	**33.60**
	R-2 F	14.48	8.70	12.62	13.31	14.51	10.55	12.68	**12.51**
	R-L F	19.63	13.83	17.13	18.10	18.86	15.07	18.86	**17.36**
XLM-RoBERTa	R-1 F	40.92	29.00	29.70	35.40	29.39	24.74	35.68	33.27
	R-2 F	14.22	8.33	12.52	13.03	14.13	10.49	12.54	12.26
	R-L F	19.66	13.54	17.07	18.05	18.43	15.03	19.54	17.26
Oracle	R-1 F	49.50	37.21	34.41	42.24	35.32	29.89	41.85	40.29
	R-2 F	25.72	15.77	17.31	20.89	21.40	15.72	19.94	20.35
	R-L F	22.67	15.93	17.38	20.54	19.19	15.33	18.61	19.16
Abstractive Summarisation after Oracle Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	R-1 F	44.26	9.13	9.63	11.23	10.77	6.93	9.71	19.99
	R-2 F	21.73	2.85	2.52	3.71	3.26	1.76	2.48	8.53
	R-L F	24.25	6.31	6.32	7.81	7.51	5.05	6.53	11.92
mT5 In-language multi-task	R-1 F	43.19	33.14	36.92	37.69	34.54	27.95	37.00	37.05
	R-2 F	21.33	13.47	17.40	17.46	18.05	13.65	13.87	17.51
	R-L F	23.70	17.00	21.44	21.33	21.44	16.98	19.01	20.78
mT5 In-language	R-1 F	44.26	35.06	39.41	43.81	41.00	32.26	4.27	**40.04**
	R-2 F	21.73	13.63	17.76	19.29	20.22	14.34	0.58	**18.23**
	R-L F	24.25	17.53	22.03	23.76	24.44	18.67	3.06	**21.93**
Abstractive Summarisation after mBERT Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	R-1 F	37.24	7.19	9.14	10.02	9.56	6.30	12.40	17.08
	R-2 F	13.00	1.68	1.87	2.48	2.27	1.30	2.82	5.25
	R-L F	19.68	5.08	5.97	6.89	6.74	4.58	7.37	10.00
mT5 In-language multi-task	R-1 F	35.56	27.05	32.59	32.94	30.01	23.53	32.90	31.30
	R-2 F	12.28	7.84	13.06	11.65	13.14	9.37	10.24	11.24
	R-L F	18.70	13.71	18.93	18.16	18.82	14.22	16.93	17.25
mT5 In-language	R-1 F	37.24	29.65	36.02	39.79	37.21	28.47	4.32	**35.03**
	R-2 F	13.00	8.32	14.08	13.86	15.46	10.66	0.10	**12.37**
	R-L F	19.68	14.76	20.08	21.17	13.20	16.65	2.80	**18.04**

Table 3: ROUGE F-measure results of the three evaluations presented Section 5.4 on WikinewsSum. We compare the extractive models, and mT5 in the three training scenarios and with two different pre-abstractive extractive steps (Oracle and mBERT) for each language of the WikinewsSum dataset in addiction to the all dataset. Bold values are the best scores obtained for each evaluation on the all WikinewsSum dataset (Oracle method excluded).

Results

- ROUGE scores

Methods	Metrics	English	German	French	Spanish	Portuguese	Polish	Italian	All Languages
Extractive Summarisation									
DistilmBERT	R-1 F	41.37	29.37	29.80	29.70	29.62	24.83	35.18	33.51
	R-2 F	14.35	8.42	12.57	12.52	14.33	10.48	12.59	12.34
	R-L F	19.66	13.65	17.10	17.07	18.75	15.03	18.43	17.30
mBERT	R-1 F	41.37	29.74	29.74	35.50	29.66	24.82	34.93	**33.60**
	R-2 F	14.48	8.70	12.62	13.31	14.51	10.55	12.68	**12.51**
	R-L F	19.63	13.83	17.13	18.10	18.86	15.07	18.86	**17.36**
XLM-RoBERTa	R-1 F	40.92	29.00	29.70	35.40	29.39	24.74	35.68	33.27
	R-2 F	14.22	8.33	12.52	13.03	14.13	10.49	12.54	12.26
	R-L F	19.66	13.54	17.07	18.05	18.43	15.03	19.54	17.26
Oracle	R-1 F	49.50	37.21	34.41	42.24	35.32	29.89	41.85	40.29
	R-2 F	25.72	15.77	17.31	20.89	21.40	15.72	19.94	20.35
	R-L F	22.67	15.93	17.38	20.54	19.19	15.33	18.61	19.16
Abstractive Summarisation after Oracle Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	R-1 F	44.26	9.13	9.63	11.23	10.77	6.93	9.71	19.99
	R-2 F	21.73	2.85	2.52	3.71	3.26	1.76	2.48	8.53
	R-L F	24.25	6.31	6.32	7.81	7.51	5.05	6.53	11.92
mT5 In-language multi-task	R-1 F	43.19	33.14	36.92	37.69	34.54	27.95	37.00	37.05
	R-2 F	21.33	13.47	17.40	17.46	18.05	13.65	13.87	17.51
	R-L F	23.70	17.00	21.44	21.33	21.44	16.98	19.01	20.78
mT5 In-language	R-1 F	44.26	35.06	39.41	43.81	41.00	32.26	4.27	**40.04**
	R-2 F	21.73	13.63	17.76	19.29	20.22	14.34	0.58	**18.23**
	R-L F	24.25	17.53	22.03	23.76	24.44	18.67	3.06	**21.93**
Abstractive Summarisation after mBERT Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	R-1 F	37.24	7.19	9.14	10.02	9.56	6.30	12.40	17.08
	R-2 F	13.00	1.68	1.87	2.48	2.27	1.30	2.82	5.25
	R-L F	19.68	5.08	5.97	6.89	6.74	4.58	7.37	10.00
mT5 In-language multi-task	R-1 F	35.56	27.05	32.59	32.94	30.01	23.53	32.90	31.30
	R-2 F	12.28	7.84	13.06	11.65	13.14	9.37	10.24	11.24
	R-L F	18.70	13.71	18.93	18.16	18.82	14.22	16.93	17.25
mT5 In-language	R-1 F	37.24	29.65	36.02	39.79	37.21	28.47	4.32	**35.03**
	R-2 F	13.00	8.32	14.08	13.86	15.46	10.66	0.10	**12.37**
	R-L F	19.68	14.76	20.08	21.17	13.20	16.65	2.80	**18.04**

Table 3: ROUGE F-measure results of the three evaluations presented Section 5.4 on WikinewsSum. We compare the extractive models, and mT5 in the three training scenarios and with two different pre-abstractive extractive steps (Oracle and mBERT) for each language of the WikinewsSum dataset in addiction to the all dataset. Bold values are the best scores obtained for each evaluation on the all WikinewsSum dataset (Oracle method excluded).

Results

- ROUGE scores

Methods	Metrics	English	German	French	Spanish	Portuguese	Polish	Italian	All Languages
Extractive Summarisation									
DistilmBERT	R-1 F	41.37	29.37	29.80	29.70	29.62	24.83	35.18	33.51
	R-2 F	14.35	8.42	12.57	12.52	14.33	10.48	12.59	12.34
	R-L F	19.66	13.65	17.10	17.07	18.75	15.03	18.43	17.30
mBERT	R-1 F	41.37	29.74	29.74	35.50	29.66	24.82	34.93	**33.60**
	R-2 F	14.48	8.70	12.62	13.31	14.51	10.55	12.68	**12.51**
	R-L F	19.63	13.83	17.13	18.10	18.86	15.07	18.86	**17.36**
XLM-RoBERTa	R-1 F	40.92	29.00	29.70	35.40	29.39	24.74	35.68	33.27
	R-2 F	14.22	8.33	12.52	13.03	14.13	10.49	12.54	12.26
	R-L F	19.66	13.54	17.07	18.05	18.43	15.03	19.54	17.26
Oracle	R-1 F	49.50	37.21	34.41	42.24	35.32	29.89	41.85	40.29
	R-2 F	25.72	15.77	17.31	20.89	21.40	15.72	19.94	20.35
	R-L F	22.67	15.93	17.38	20.54	19.19	15.33	18.61	19.16
Abstractive Summarisation after Oracle Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	R-1 F	44.26	9.13	9.63	11.23	10.77	6.93	9.71	19.99
	R-2 F	21.73	2.85	2.52	3.71	3.26	1.76	2.48	8.53
	R-L F	24.25	6.31	6.32	7.81	7.51	5.05	6.53	11.92
mT5 In-language multi-task	R-1 F	43.19	33.14	36.92	37.69	34.54	27.95	37.00	37.05
	R-2 F	21.33	13.47	17.40	17.46	18.05	13.65	13.87	17.51
	R-L F	23.70	17.00	21.44	21.33	21.44	16.98	19.01	20.78
mT5 In-language	R-1 F	44.26	35.06	39.41	43.81	41.00	32.26	4.27	**40.04**
	R-2 F	21.73	13.63	17.76	19.29	20.22	14.34	0.58	**18.23**
	R-L F	24.25	17.53	22.03	23.76	24.44	18.67	3.06	**21.93**
Abstractive Summarisation after mBERT Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	R-1 F	37.24	7.19	9.14	10.02	9.56	6.30	12.40	17.08
	R-2 F	13.00	1.68	1.87	2.48	2.27	1.30	2.82	5.25
	R-L F	19.68	5.08	5.97	6.89	6.74	4.58	7.37	10.00
mT5 In-language multi-task	R-1 F	35.56	27.05	32.59	32.94	30.01	23.53	32.90	31.30
	R-2 F	12.28	7.84	13.06	11.65	13.14	9.37	10.24	11.24
	R-L F	18.70	13.71	18.93	18.16	18.82	14.22	16.93	17.25
mT5 In-language	R-1 F	37.24	29.65	36.02	39.79	37.21	28.47	4.32	**35.03**
	R-2 F	13.00	8.32	14.08	13.86	15.46	10.66	0.10	**12.37**
	R-L F	19.68	14.76	20.08	21.17	13.20	16.65	2.80	**18.04**

Table 3: ROUGE F-measure results of the three evaluations presented Section 5.4 on WikinewsSum. We compare the extractive models, and mT5 in the three training scenarios and with two different pre-abstractive extractive steps (Oracle and mBERT) for each language of the WikinewsSum dataset in addiction to the all dataset. Bold values are the best scores obtained for each evaluation on the all WikinewsSum dataset (Oracle method excluded).

Results

- ROUGE scores

Methods	Metrics	English	German	French	Spanish	Portuguese	Polish	Italian	All Languages
Extractive Summarisation									
DistilmBERT	R-1 F	41.37	29.37	29.80	29.70	29.62	24.83	35.18	33.51
	R-2 F	14.35	8.42	12.57	12.52	14.33	10.48	12.59	12.34
	R-L F	19.66	13.65	17.10	17.07	18.75	15.03	18.43	17.30
mBERT	R-1 F	41.37	29.74	29.74	35.50	29.66	24.82	34.93	**33.60**
	R-2 F	14.48	8.70	12.62	13.31	14.51	10.55	12.68	**12.51**
	R-L F	19.63	13.83	17.13	18.10	18.86	15.07	18.86	**17.36**
XLM-RoBERTa	R-1 F	40.92	29.00	29.70	35.40	29.39	24.74	35.68	33.27
	R-2 F	14.22	8.33	12.52	13.03	14.13	10.49	12.54	12.26
	R-L F	19.66	13.54	17.07	18.05	18.43	15.03	19.54	17.26
Oracle	R-1 F	49.50	37.21	34.41	42.24	35.32	29.89	41.85	40.29
	R-2 F	25.72	15.77	17.31	20.89	21.40	15.72	19.94	20.35
	R-L F	22.67	15.93	17.38	20.54	19.19	15.33	18.61	19.16
Abstractive Summarisation after Oracle Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	R-1 F	44.26	9.13	9.63	11.23	10.77	6.93	9.71	19.99
	R-2 F	21.73	2.85	2.52	3.71	3.26	1.76	2.48	8.53
	R-L F	24.25	6.31	6.32	7.81	7.51	5.05	6.53	11.92
mT5 In-language multi-task	R-1 F	43.19	33.14	36.92	37.69	34.54	27.95	37.00	37.05
	R-2 F	21.33	13.47	17.40	17.46	18.05	13.65	13.87	17.51
	R-L F	23.70	17.00	21.44	21.33	21.44	16.98	19.01	20.78
mT5 In-language	R-1 F	44.26	35.06	39.41	43.81	41.00	32.26	4.27	**40.04**
	R-2 F	21.73	13.63	17.76	19.29	20.22	14.34	0.58	**18.23**
	R-L F	24.25	17.53	22.03	23.76	24.44	18.67	3.06	**21.93**
Abstractive Summarisation after mBERT Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	R-1 F	37.24	7.19	9.14	10.02	9.56	6.30	12.40	17.08
	R-2 F	13.00	1.68	1.87	2.48	2.27	1.30	2.82	5.25
	R-L F	19.68	5.08	5.97	6.89	6.74	4.58	7.37	10.00
mT5 In-language multi-task	R-1 F	35.56	27.05	32.59	32.94	30.01	23.53	32.90	31.30
	R-2 F	12.28	7.84	13.06	11.65	13.14	9.37	10.24	11.24
	R-L F	18.70	13.71	18.93	18.16	18.82	14.22	16.93	17.25
mT5 In-language	R-1 F	37.24	29.65	36.02	39.79	37.21	28.47	4.32	**35.03**
	R-2 F	13.00	8.32	14.08	13.86	15.46	10.66	0.10	**12.37**
	R-L F	19.68	14.76	20.08	21.17	13.20	16.65	2.80	**18.04**

Table 3: ROUGE F-measure results of the three evaluations presented Section 5.4 on WikinewsSum. We compare the extractive models, and mT5 in the three training scenarios and with two different pre-abstractive extractive steps (Oracle and mBERT) for each language of the WikinewsSum dataset in addiction to the all dataset. Bold values are the best scores obtained for each evaluation on the all WikinewsSum dataset (Oracle method excluded).

Results

- BERTScore metric

➔ Results coherent with the ROUGE scores

➔ Same conclusions

Methods	Metrics	English	German	French	Spanish	Portuguese	Polish	Italian	All Languages
Extractive Summarisation									
DistilmBERT	B-S P	0.6920	0.6669	0.6357	0.6807	0.6680	0.6455	0.6706	0.6697
	B-S R	0.7196	0.6890	0.6846	0.7104	0.7084	0.6834	0.7068	0.7021
	B-S F	0.7052	0.6774	0.6585	0.6949	0.6869	0.6633	0.6879	0.6850
mBERT	B-S P	0.6908	0.6679	0.6354	0.6810	0.6673	0.6456	0.6618	0.6695
	B-S R	0.7215	0.6931	0.6855	0.7124	0.7084	0.6848	0.7033	0.7041
	B-S F	0.7055	0.6799	0.6587	0.6960	0.6865	0.6640	0.6816	0.6859
XLM-RoBERTa	B-S P	0.6900	0.6658	0.6351	0.6794	0.6660	0.6451	0.6752	0.6684
	B-S R	0.7173	0.6878	0.6834	0.7087	0.7061	0.6831	0.7099	0.7005
	B-S F	0.7031	0.6762	0.6576	0.6934	0.6848	0.6629	0.6917	0.6836
Oracle	B-S P	0.7238	0.6947	0.6528	0.7058	0.6930	0.6638	0.6919	0.6955
	B-S R	0.7436	0.7144	0.6967	0.7228	0.7266	0.7024	0.7190	0.7217
	B-S F	0.7332	0.7039	0.6731	0.7138	0.7087	0.6818	0.7047	0.7077
Abstractive Summarisation after Oracle Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	B-S P	0.7526	0.6814	0.6687	0.7014	0.6864	0.6468	0.6820	0.7009
	B-S R	0.7199	0.6431	0.6579	0.6650	0.6641	0.6218	0.6480	0.6717
	B-S F	0.7354	0.6614	0.6627	0.6824	0.6746	0.6337	0.6644	0.6855
mT5 In-language multi-task	B-S P	0.7494	0.7219	0.7130	0.7306	0.7274	0.6887	0.7203	0.7274
	B-S R	0.7190	0.6937	0.7174	0.7030	0.7140	0.6847	0.6942	0.7074
	B-S F	0.7334	0.7070	0.7138	0.7161	0.7197	0.6857	0.7066	0.7165
mT5 In-language	B-S P	0.7526	0.7264	0.7164	0.7374	0.7381	0.6974	0.4603	0.7321
	B-S R	0.7199	0.6939	0.7179	0.7073	0.7194	0.6908	0.5261	0.7092
	B-S F	0.7354	0.7093	0.7153	0.7216	0.7277	0.6931	0.4905	0.7196
Abstractive Summarisation after mBERT Pre-Abstractive Extractive Step									
mT5 Cross-lingual zero-shot transfer	B-S P	0.7202	0.6680	0.6571	0.6858	0.6757	0.6412	0.6693	0.6828
	B-S R	0.7004	0.6363	0.6517	0.6576	0.6586	0.6180	0.6459	0.6615
	B-S F	0.7098	0.6515	0.6538	0.6712	0.6666	0.6290	0.6572	0.6716
mT5 In-language multi-task	B-S P	0.7157	0.6958	0.6953	0.7069	0.7094	0.6700	0.7045	0.7022
	B-S R	0.6981	0.6774	0.7033	0.6891	0.7011	0.6702	0.6869	0.6910
	B-S F	0.7065	0.6861	0.6982	0.6976	0.7046	0.6693	0.6952	0.6960
mT5 In-language	B-S P	0.7202	0.7043	0.7020	0.7151	0.7186	0.6836	0.4495	0.7091
	B-S R	0.7004	0.6807	0.7069	0.6948	0.7064	0.6803	0.5213	0.6949
	B-S F	0.7098	0.6919	0.7026	0.7044	0.7116	0.6811	0.4822	0.7012

Table 6: BERTScore (Zhang et al., 2020b) precision (B-S P), recall (B-S R), and F1 (B-S F) results of the three evaluations presented Section 5.4 on WikinewsSum. We compare the extractive models, and mT5 in the three training scenarios and with two different pre-abstractive extractive steps (Oracle and mBERT) for each language of the WikinewsSum dataset in addiction to the all dataset. Hash code for the BERTScore metric: bert-base-multilingual-cased_L9_no-idf_version=0.3.11(hug_trans=4.13.0)_fast-tokenizer

📊 Results

➔ **Using a pre-abstractive extractive step is a valid approach to overcome the input length limitation of the abstractive models**

➔ **The In-language multi-task training improves performance for low-resource languages**

➔ **Cross-lingual zero-shot training does not work for summarisation because the model produces all the summaries in English**

📌 Conclusion

➔ **New dataset for multilingual summarisation with extended summaries: WikinewsSum**

➔ **Strong baselines with the combination of Extractive and Abstractive models**

➔ **In-language multi-task training helps for low resource languages**

🙏 Thank you for your attention!

❓ Questions?

- ➢ Dataset code to reproduce: https://github.com/airKlizz/wikinewssum
- ➢ Dataset data: https://live.european-language-grid.eu/catalogue/corpus/18633
- ➢ Training code: https://github.com/airKlizz/mdmls
- ➢ Models: https://huggingface.co/models?sort=downloads&search=airklizz+mt5+wikinewssum