



Paper

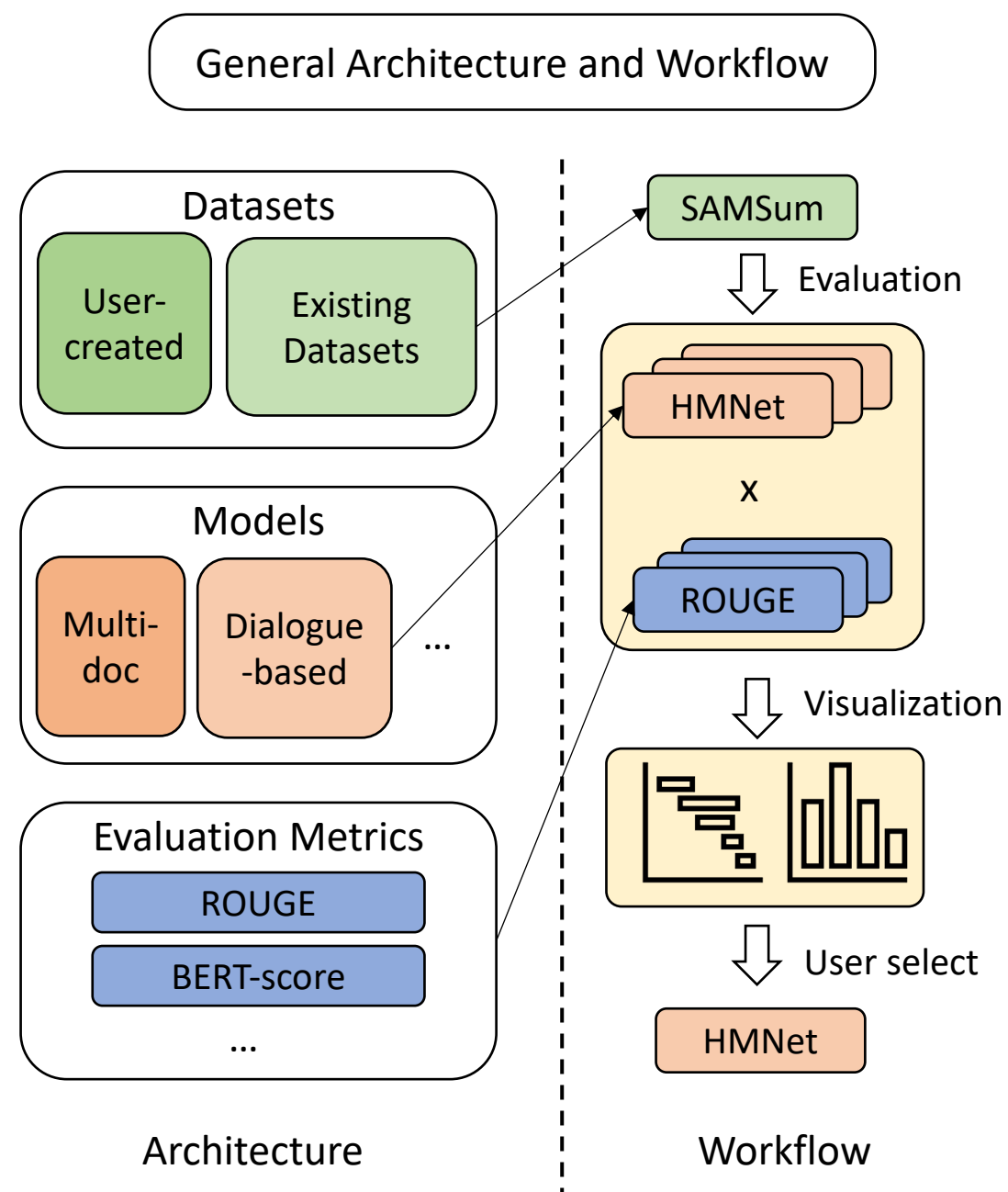


GitHub Repo



@AnsongNi

SummerTime: A Complete Summarization Toolkit



Key features for SummerTime

- It is a complete toolkit that includes datasets, models and evaluation metrics for summarization;
- It can automatically assemble pipeline solutions for complex summarization tasks;
- User can do a side-by-side comparison of different models and visualize their differences in performance
- It provides textual descriptions of evaluation metrics to help user better understand the quality of the generated summary

Supported Datasets and Models

Currently Supported Datasets

Dataset	Domain	# Examples	Src. length	Tgt. length	Query	Multi-doc	Dialogue	Multi-lingual
ArXiv	Scientific articles	215k	4.9k	220				
CNN/DM(3.0.0)	News	300k	781	56				
MlsumDataset	Multi-lingual News	1.5M+	632	34		✓		German, Spanish, French, Russian, Turkish
Multi-News	News	56k	2.1k	263.8		✓		
SAMSum	Open-domain	16k	94	20			✓	
Pubmedqa	Medical	272k	244	32	✓			
QMSum	Meetings	1k	9.0k	69.6	✓		✓	
ScisummNet	Scientific articles	1k	4.7k	150				
SummScreen	TV shows	26.9k	6.6k	337.4			✓	
XSum	News	226k	431	23.3				

Currently Supported Models

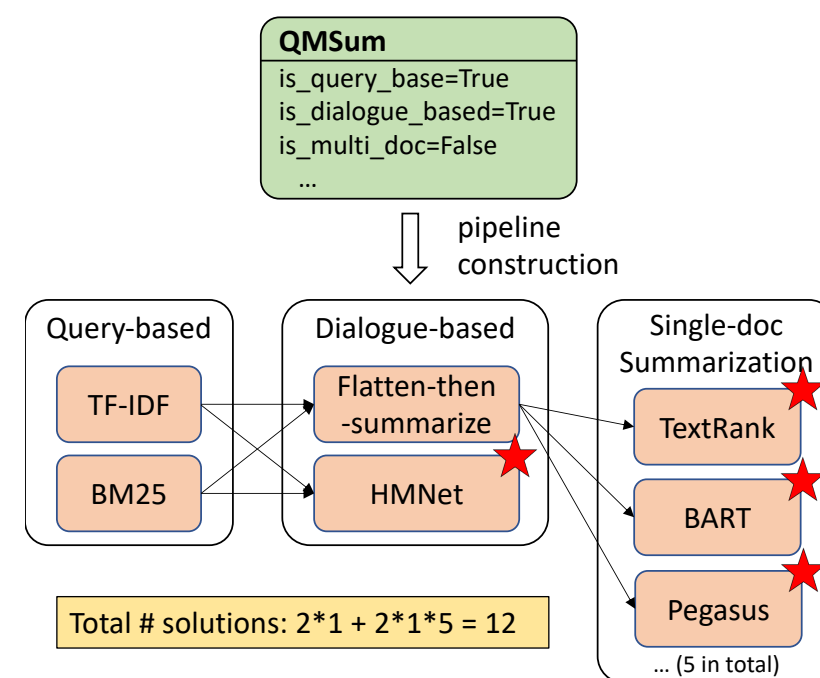
Models	Single-doc	Multi-doc	Dialogue-based	Query-based
BartModel	✓			
BM25SummModel				✓
HMNetModel			✓	
LexRankModel	✓			
LongformerModel	✓			
MultiDocJointModel		✓		
MultiDocSeparateModel		✓		
PegasusModel	✓			
TextRankModel	✓			
TFIDFSummModel				✓

Currently Supported Metrics

ROUGE BLEU ROUGE-WE METEOR BERTScore

Automatic Pipeline Assembly and Model Selection

Automatic Pipeline Assembly



SummerTime automatically find possible solutions for complex summarization tasks, such as query-based dialoguq summarization shown here

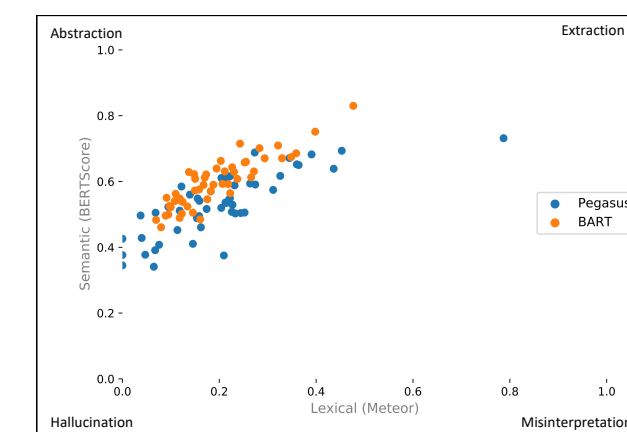
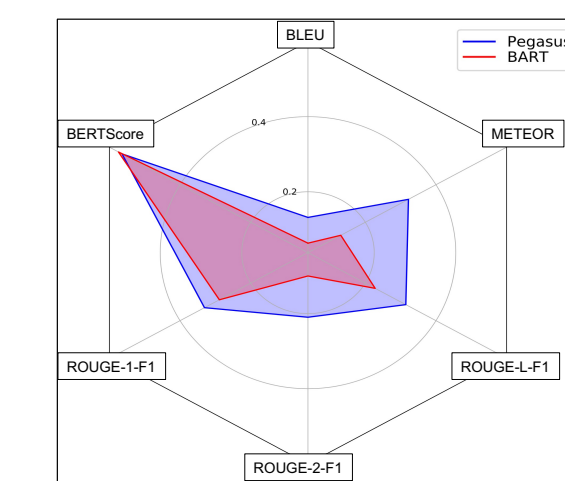
Model Selection

Algorithm 1 SELECT($\mathcal{M}, \mathcal{D}, \mathcal{E}$)

Input: \mathcal{M} : a pool of models to choose from, \mathcal{D} : a set of examples from a dataset, \mathcal{T} : a set of evaluation metrics, d : initial resource number, k : increase resource factor
Output: $M \subseteq \mathcal{M}$: a subset of models;
 1: Initialize $M = \mathcal{M}, M' = \emptyset$
 2: **while** $M' \neq M$ **do**
 3: $D = \text{sample}(\mathcal{D}, d)$
 4: **for each** $m \in \mathcal{M}, e \in \mathcal{E}$ **do**
 5: $r_m^e = \text{eval}(m, D, e)$
 6: **end for**
 7: $M' = M$
 8: **for each** $m \in M$ **do**
 9: **if** $\exists m'$ s.t. $r_{m'}^e > r_m^e, \forall e \in \mathcal{E}$ **then**
 10: $M = M \setminus m$
 11: **end if**
 12: **end for**
 13: $d = d * k$
 14: **end while**

We use successive-halving to save time and resources in searching for the best models

Visualization



Code Example for Using SummerTime

```
import dataset
import model
import evaluation

# load a supported dataset
dataset.list_all_dataset()
dataset.CnndmDataset.show_description()
cnn_dataset = dataset.CnndmDataset()

# OPTION 1: user manually select and evaluate
model.list_all_models()
model.BartModel.show_capability()
exp_model = model.BartModel()

summaries = exp_model.summarize(articles)
targets = [instance.summary for instance in
           cnn_dataset.test_set]

bert_metric = evaluation.BertScore()
bert_metric.evaluate(summaries, targets)

# OPTION 2: automatic pipeline assembly
# Here we use a more complex task: query-
# based + dialogue-based summarization
qmsum_dataset = dataset.QMSumDataset()
assembled_models =
    assemble_model_pipeline(QMSumDataset)

# AND automatic model selection
model_selector = evaluation.ModelSelector(
    models=assembled_models,
    dataset=qmsum_dataset,
    metrics=[bert_metric])
eval_table = model_selector.run()
model_selector.visualize()
```

Future Versions

We would like to consider the following as important future work for SummerTime:

Future Work 1: Adding datasets and models for multi-lingual support or summarization of non-English text. This is already a work-in-progress.

Future Work 2: Enable the model for training and fine-tuning on the datasets.

Future Work 3: Add more visualization methods to better help the user compare different summarization models.