News articles summarization

Nie Lubimy Poniedziałków

Roadmap

1. Problem
2. State of the art
3. Datasets
4. Methods
5. Results
6. Conclusions

Main idea

1. Predict each sentence in an article if it's objective or subjective
2. Left out only objective sentences - more information
3. Summary left sentences

State of the art - objectivity

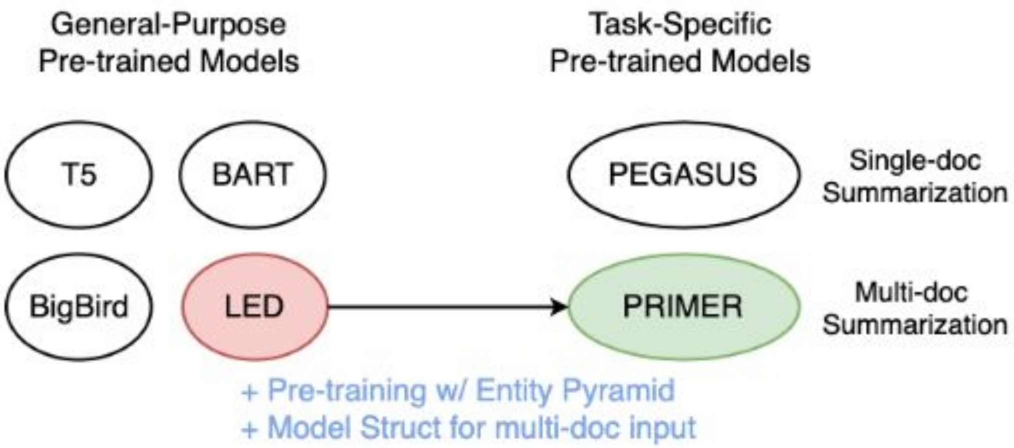
TextBlob - a tool based on NLTK and Pattern, provides a subjectivity and polarity score

Models for sentiment analysis - neutral sentences are objective?

- VADER - very popular model for sentiment analysis, trained on social media data
- happytransformer - wrapper for huggingface transformer language models
- Google Cloud Natural Language API - provides cloud-based sentiment analysis

State of the art - summarization

1. Extractive: extract the most informative sentences.
2. Abstractive: understand whole text, then generate summary (as humans do).
   - PEGASUS - transformer based model
   - PRIMER - transformer model to multi-doc summarization



Summarization datasets

data	input	output	properties	size
Dailymail / CNN	news article	highlights from text	to predict are bullet points written below article	220k
xsum	news article	single-sentence summary	shortest, but with reported speech	227k
newsroom	news article	summaries written by editors	to predict is a summary	1.3m
multi-news	news article	human written summary by newser	longest summaries, professionally written by newser	55k

Methods - objectivity

1. TextBlob - returns subjectivity score
2. Transformer - returns probability that sentence if either positive or negative

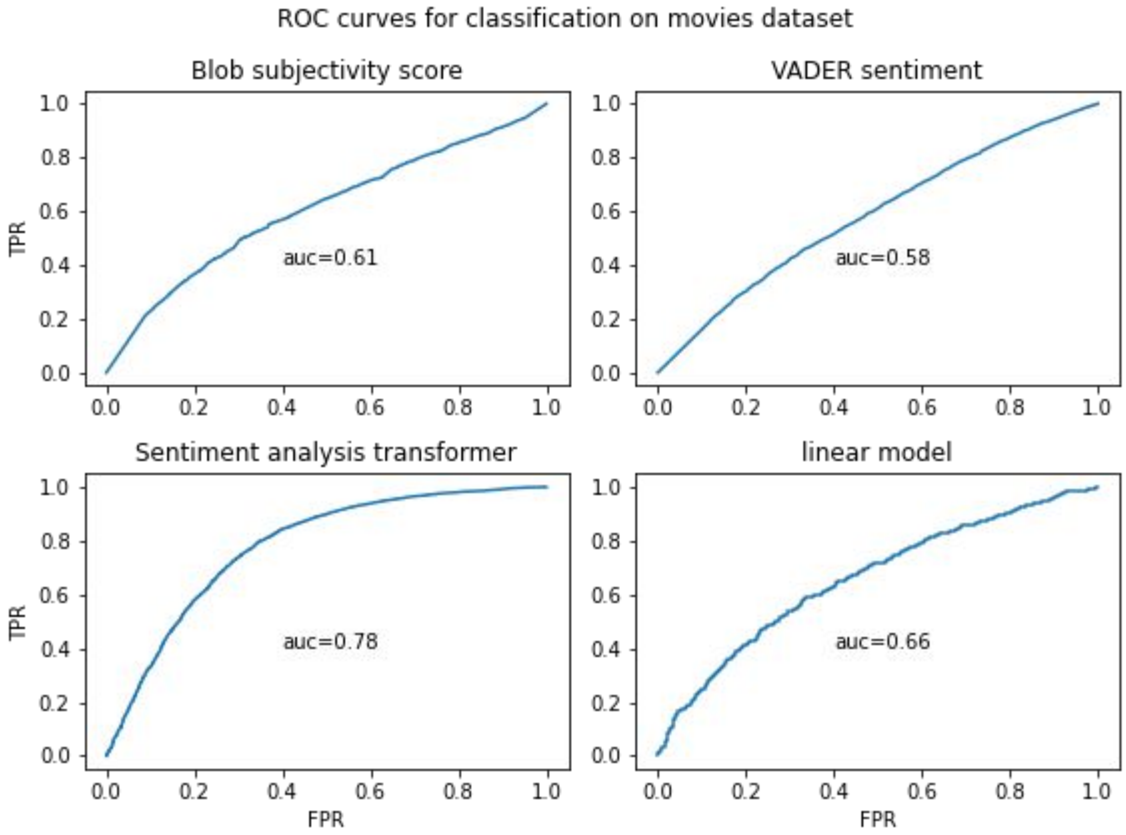
In both methods filter the most objective sentences - the sentences with score lower than threshold

Methods - summarization

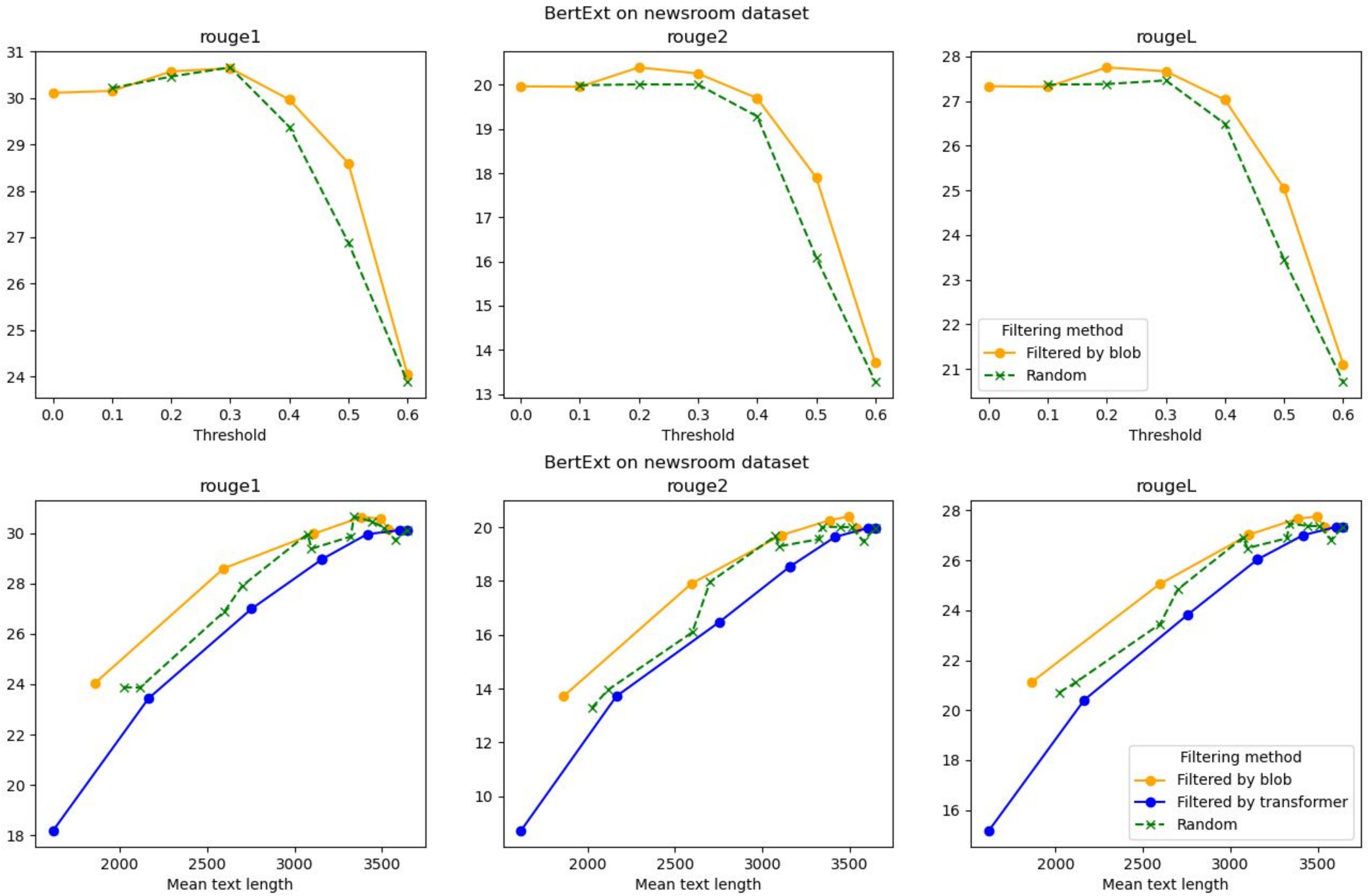
1. BertExt - exstractive method to summarize, based on Bert, most common as baseline of other methods
2. Pegasus - state-of-the-art in abstractive summarization, many pretrained models including for newsroom and multinews

1. Apply ready method on text filtered by objectiveness
2. Check the performance of summarization with the same number of randomly selected sentences

Objectivity results



Summarization results

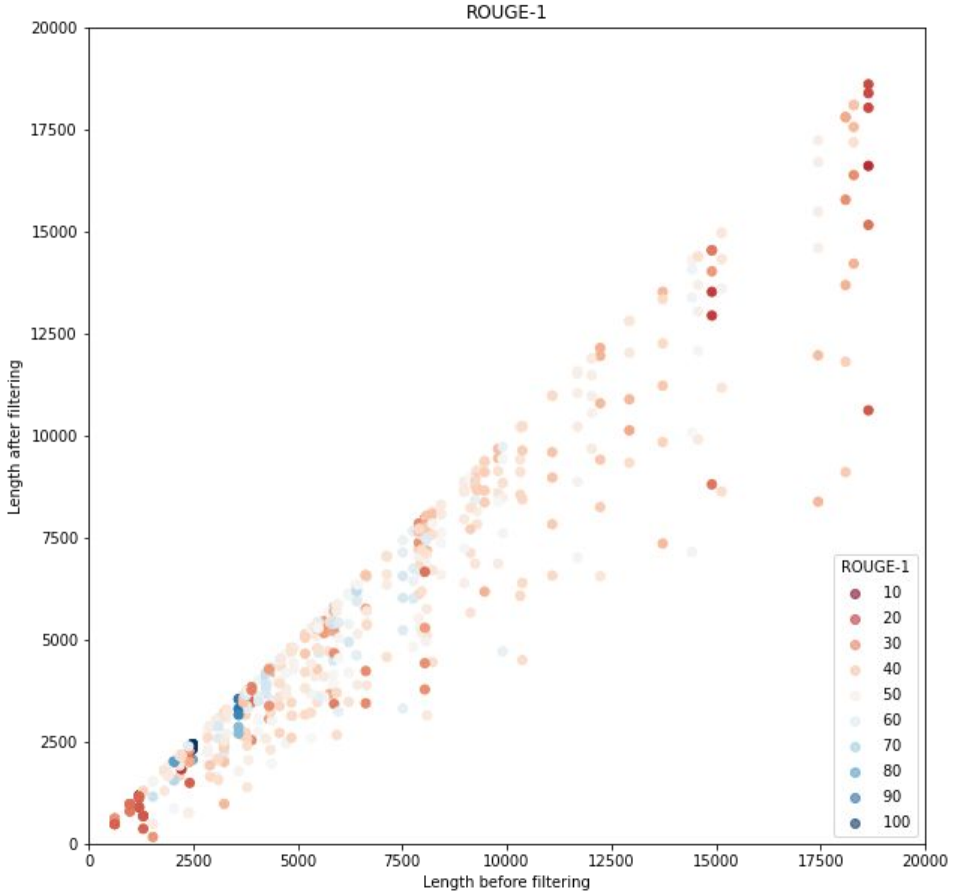


Document length vs filtered length -> Score

Dataset: multinews

Filtering: transformer

Method: Pegasus



Overall best results

dataset	method	filtering	length	rouge1	rouge2	rougeL
Multinews	BertExt	Baseline	9902.78	40.2815	12.7210	18.5211
Multinews	BertExt	Transformer 0.002	7751.77	40.1122	12.8683	18.7981
Multinews	BertExt	Blob 0.2	9251.31	40.0976	12.5480	18.5161
Multinews	Pegasus	Baseline	9902.78	47.6871	18.7212	24.4209
Multinews	Pegasus	Transformer 0.0005	9255.50	47.9188	19.1343	24.8020
Newsroom	BertExt	Baseline	3643.29	30.1140	19.9668	27.3317
Newsroom	BertExt	Blob 0.2	3493.73	30.5741	20.3963	27.7525
Newsroom	Pegasus	Baseline	3643.29	39.0886	28.1645	35.3626
Newsroom	Pegasus	Blob 0.3	3384.18	37.4897	26.9782	33.6508

Conclusions

- The best objectiveness model is Transformer, even thought it is trained to predict sentiment
- The Transformer is only good with multinews dataset with abstractive model
- Better objectiveness model may give significantly better results
- TextBlob is better with extractive model and on newsroom dataset
- Summarization of longer text is harder, since results are lower
- Future work: check emotionality of text and filter less emotional sentences

Thank you for your attention!

Authors:

Michał Pastuszka

Tomasz Makowski

Bibliography - objective vs subjective

http://www.cs.cornell.edu/people/pabo/movie-review-data/

https://mpqa.cs.pitt.edu/

Bibliography - summarization

https://arxiv.org/pdf/1912.08777.pdf - model with many datasets described

https://huggingface.co/datasets/cnn_dailymail/blob/main/cnn_dailymail.py - dailymail dataset

https://github.com/EdinburghNLP/XSum/tree/master/XSum-Dataset - xsum dataset

https://lil.nlp.cornell.edu/newsroom/explore/index.html - newsroom dataset examples

https://github.com/Alex-Fabbri/Multi-News - multi-news dataset

http://nlpprogress.com/english/summarization.html - nlp progress about summarization

https://www.freecodecamp.org/news/what-is-rouge-and-how-it-works-for-evaluation-of-summaries-e059fb8ac840/ - ROUGE explained

https://paperswithcode.com/sota/multi-document-summarization-on-multi-news - multi-news models

https://github.com/allenai/primer, https://arxiv.org/pdf/2110.08499v1.pdf - PRIMER - best on multi-news

https://medium.com/@thakermadhav/comparing-text-summarization-techniques-d1e2e465584e - summarization methods (approaches), only extractive

https://medium.com/swlh/abstractive-text-summarization-using-transformers-3e774cc42453 - abstractive

https://github.com/google-research/pegasus, https://arxiv.org/pdf/1912.08777.pdf - PEGASUS

https://huggingface.co/google?sort_models=alphabetical#models - PEGASUS models

Evaluation metrics - ROUGE based

Ground-truth summary: This is a summary.

Predicted summary: This is summary.

Precision = number of overlapping words / number of predicted words = 3/3

Recall = number of overlapping words / number of ground-truth words = 3/4

ROUGE-1 - measures unigrams as above

ROUGE-2 - measures bigrams, then precision = 1/2, recall = 1/3

ROUGE-L - the longest matching sequence, then precision 3/3, recall 3/4