Politecnico di Torino

Probabilistic Context Extraction For Extractive Highlights Extraction

"NLLP" Group

Villani Luca	s304992
Marcellino Luca	s292950
Cagnasso Pietro	s300801
Vergaro Nicolò	s295633

DNLP
A.Y. 2022/2023

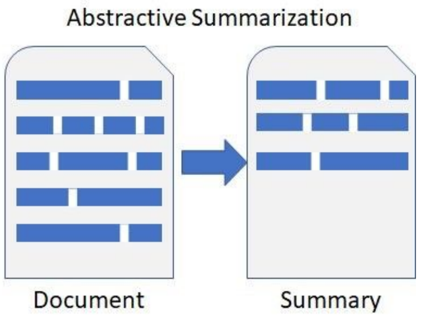
Summarization

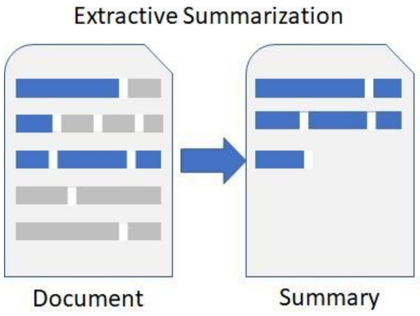
Expressing the most important facts or ideas about something in a short and clear form

- Goals:
    - Minimize redundancy
    - Maximize relevance
    - Maximize accessibility

- In this project:
    - Single document
    - Single language
    - Weakly structured documents
    - Both abstractive and extractive

2 approaches





Models - BART [1]

Bidirectional AutoRegressive Transformers

- Transformer-based encoder-decoder structure
  - Bidirectional encoder
  - Autoregressive decoder
- Pre-trained with denoising objective
  - Several input corruption strategies
- Fine-tuned with uncorrupted documents
- Used in Seq2Seq tasks like summarization and translation

[1] Mike Lewis et al. "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension", 2019

Models - BERT [2]

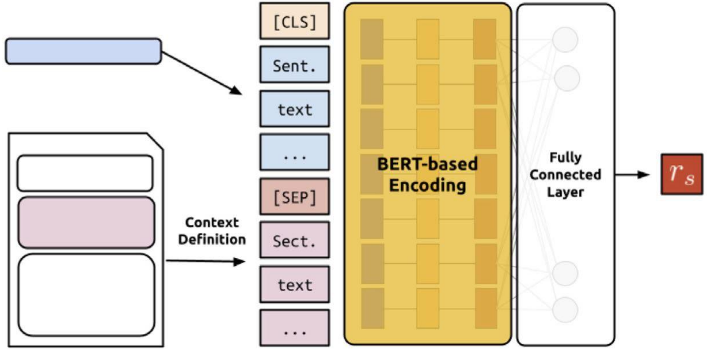
Bidirectional Encoder Representations from Transformers

- Encoder layer almost identical to original Transformer [3]
- Given an input sequence, produces an embedding for each token in the sequence
- Easy to fine-tune
- Pre-trained on a large corpus with:
    - Masked LM
    - Next sentence prediction
    - Joint of two above

[2] Jacob Devlin et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", 2018
[3] Ashish Vaswani et al. "Attention is all you need", 2017

Models - THExt [4]

- SOTA model for highlights extraction
  - extractive summarization approach
- Input:
  - Sentence + context (Abstract, Introduction, …)
- Processing:
  - BERT-based embedding [2]
  - Fully-connected regression layer
- Output:
  - Estimation of Rouge-2 F1 score
- Rank sentences based on the regression output
  - Top-3 are the highlights



[4] Moreno La Quatra, Luca Cagliero "Transformer-based highlights extraction from scientific papers", 2022
[2] Jacob Devlin et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", 2018

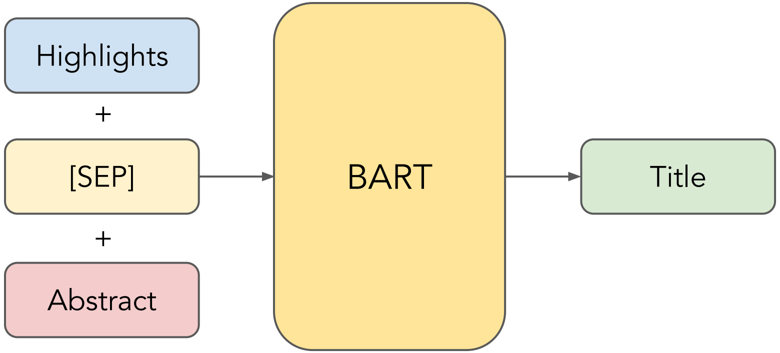
Experimental settings

1. Datasets:
   - CSPubSumm [5]
   - BIOPubSumm [6]
   - AIPubSumm [6]
2. Use **one seed** to make results reproducible
3. Evaluation metrics:
   - ROUGE [7]
   - BERTScore [8]
4. Hardware:
   - Intel Xeon (limited to 2 cores)
   - 2x Nvidia T4
   - 13 GB of RAM
   - Ubuntu 20.04 LTS

[5] Ed Collins, Isabelle Augenstein, and Sebastian Riedel “A supervised approach to extractive summarisation of scientific papers”, 2017
[6] Luca Cagliero, Moreno La Quatra “Extracting highlights of scientific articles: A supervised summarization approach”, 2020
[7] Chin-Yew Lin “ROUGE: A package for automatic evaluation of summaries”, 2004
[8] Tianyi Zhang et al. “Bertscore: Evaluating text generation with BERT”, 2019.

Methods - Headline Generation 1/2

- *Intuition:* the title is per se an highlight
- Model: BART → abstractive generation
- Ablation study (10% MiscPubSumm):
  - Highlights only
  - Abstract only
  - Highlights + Abstract
- Concatenation using model's separation token



Input	ROUGE-1	ROUGE-2	BERTScore
HL only	0.3528	0.1617	0.879
Abs only	0.4031	0.2141	0.886
HL+Abs	0.4287	0.2287	0.894

Results - Headline Generation 2/2

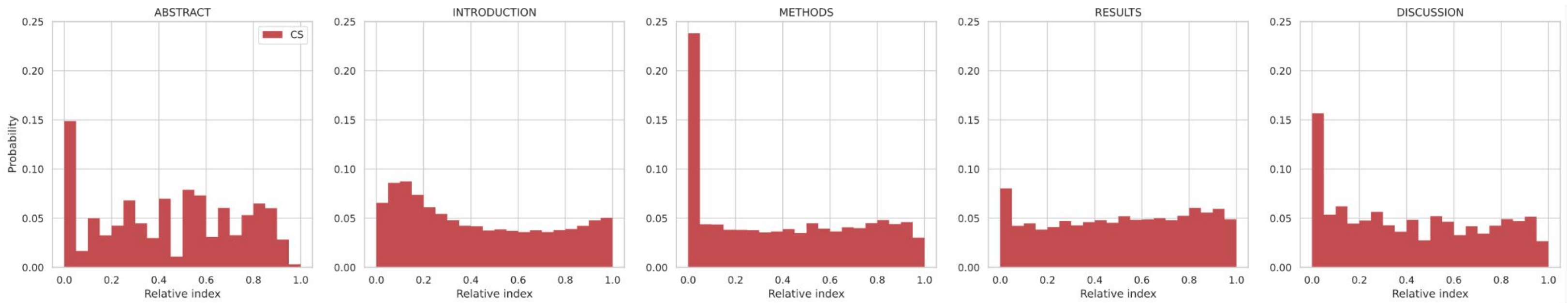
- 2 fine-tuning steps:
  - 1 epoch on MiscPubSumm → general view
  - 1 epoch on each dataset → specialize the model
- Hyperparameters:
  - Learning rate=1e-5, weight decay=1e-2
  - 10% warm-up steps
  - total batch size=8
- Example:
  - *Generated:* A Transformer-based Highlights Extractor (THExt)
  - *Original:* Transformer-based highlights extraction from scientific papers
  - The title of this presentation and our paper was generated with this model

Model	R-1	R-2	R-L	BERTScore
AI	0.4332	0.2240	0.3607	0.9064
BIO	0.4580	0.2541	0.3961	0.9027
CS	0.5584	0.3818	0.5012	0.9233

Methods - Probabilistic Context Extraction 1/3

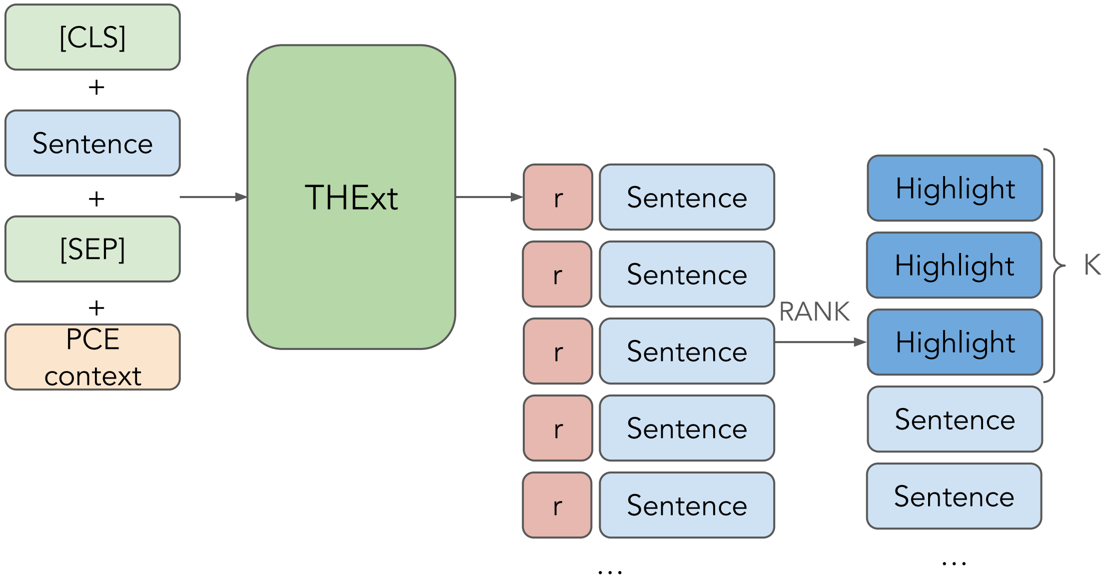
- *Intuition:* existence of an underlying distribution of important sentences across topics and sections
- Compute the empirical approximation
  - Top-20 sentences per paper
  - ROUGE-2 F1 of each sentence wrt highlights
  - Associate to the corresponding section
  - Discretize in 20 bins
- Sections contribution table
- Example on CSPubSumm

Section	CS
Abstract	0.1438
Introduction	**0.5461**
Methods	0.0730
Results	0.0898
Discussion	0.1473



Methods - Probabilistic Context Extraction 2/3

- Creation of the new context
  - Take N=15 bins according to the distribution, with replacement
  - Concatenate sentences with [SEP]
  - Choose sentence inside bin using "picking strategy"
    - Random
    - ROUGE-2
    - Best
- Feed into pre-trained THExt
- Rank
- Take the first K=3 as highlights



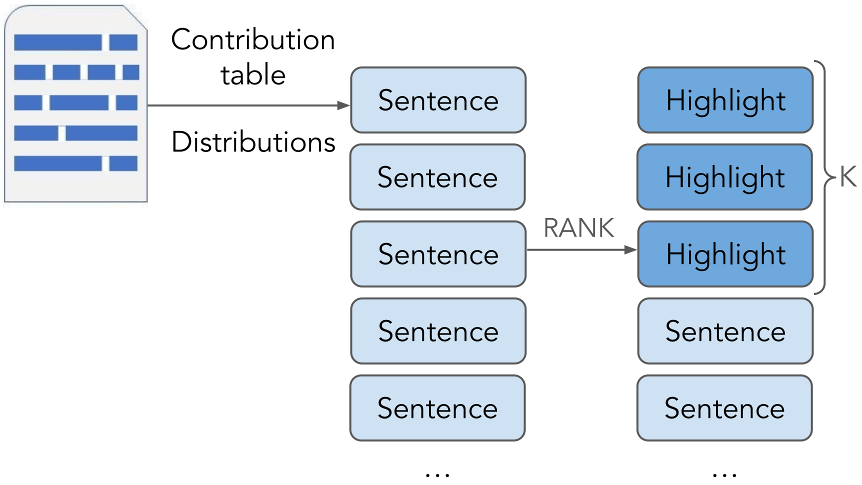
Results - Probabilistic Context Extraction 3/3

- 1 epoch on each topic's dataset
- Hyperparameters:
  - Learning rate=1e-5, weight decay=1e-2
  - 10% warm-up steps
  - total batch size=32
- Can incorporate prior knowledge or other methods (e.g., get original THExt)

	CS		BIO		AI	
Model	R-1	R-2	R-1	R-2	R-1	R-2
THExt+Abstract	0.3138	0.1204	0.3002	0.1017	0.3350	**0.1253**
THExt+PCE random	0.3647	0.1585	0.3308	0.1195	0.3353	0.1213
THExt + PCE-ROUGE2	0.3676	0.1510	0.3326	0.1196	0.3372	0.1178
THExt + PCE-best	**0.3738**	**0.1613**	**0.3335**	**0.1222**	**0.3415**	0.1250

Methods - Direct extraction 1/2

- Further proves existence of distributions
- Select N=15 sentences across sections using the contribution table
    - Choose the bins according to the distribution
    - Randomly select a sentence from the bin
- Rank by ROUGE-2 of each sentence with abstract
- Take top-K
    - K=2 for CS
    - K=3 for BIO, AI



Results - Direct extraction 2/2

- Impressive results considering
  - No training
  - No inference
- Comparable results with previous methods
- Results on BioPubSumm

Method	R-1	R-2
Liu and Lapata [8]	0.249	0.059
Collins et al. [4]	0.287	0.087
THExt + Abstract	**0.3002**	**0.1017**
Direct extraction (K=3)	0.2738	0.0805

[4] Ed Collins, Isabelle Augenstein, and Sebastian Riedel "A supervised approach to extractive summarisation of scientific papers", 2017
[8] Yang Liu and Mirella Lapata "Text summarization with pretrained encoders", 2019

Discussion - Headline Generation

- The model learnt that different topics have different titles strategies:
  - AI titles may include the methods name
  - BIO titles may include molecules names
- Titles generated using THExt's highlights maintain high results
- Limitations when dealing with multi-topic papers
  - Focuses on just one, probably the main one (e.g., in our case focused just on PCE)

Discussion - Probabilistic Context Extraction

- The model outperforms the reference results
  - An even better context is feasible
  - BUT we performed an additional epoch wrt THExt
- A deeper hyper-parameter tuning could possibly achieve even better results.
- A deeper text cleaning is possible
  - Substitute citations, figure and table with a standardized version
  - Convert Unicode characters

Discussion - Direct extraction

- Performs surprisingly well
- Further improvements may be reached:
  - Using "best" picking strategy
  - Tune the number of sentences

Thanks for your attention