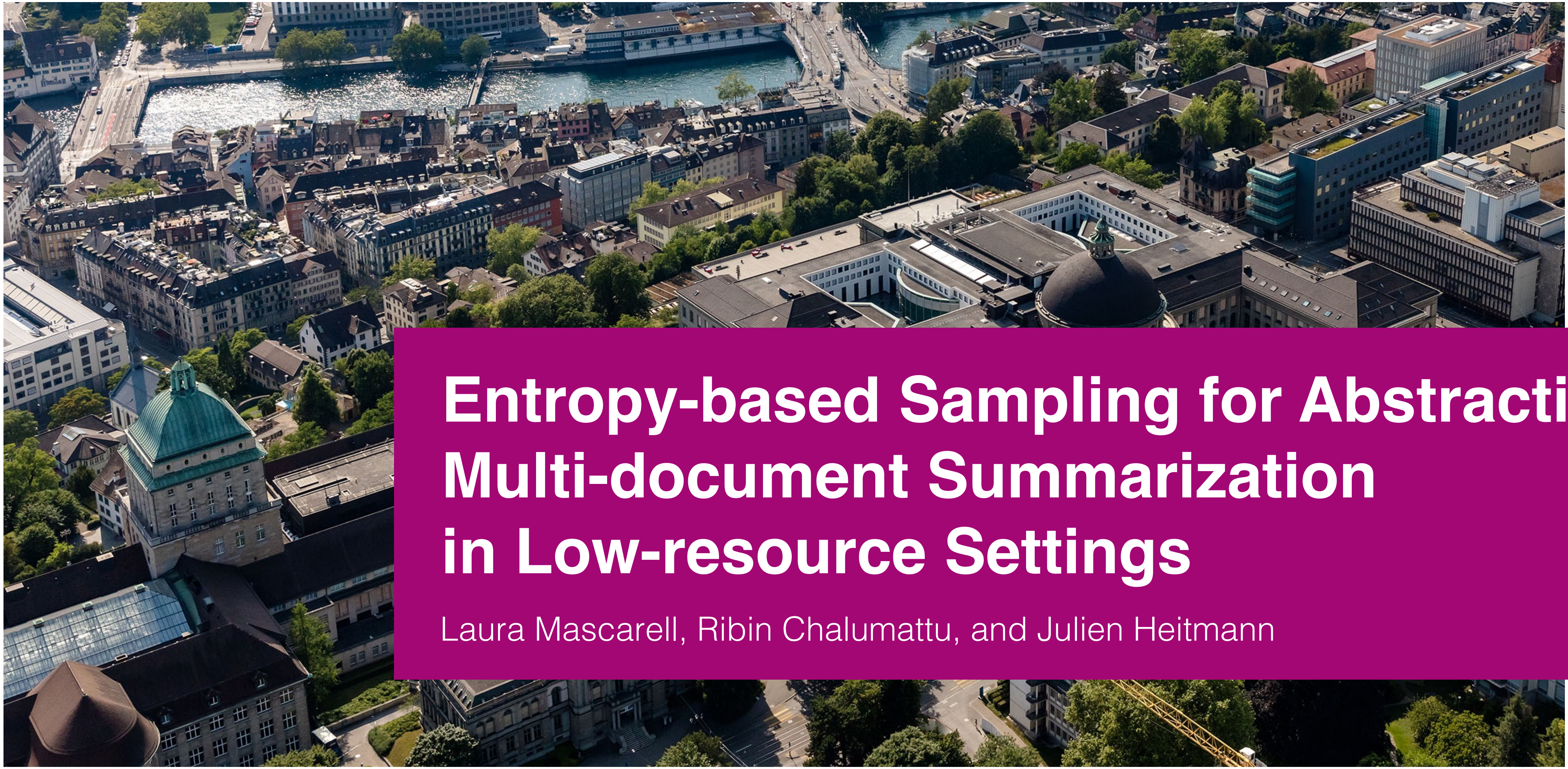
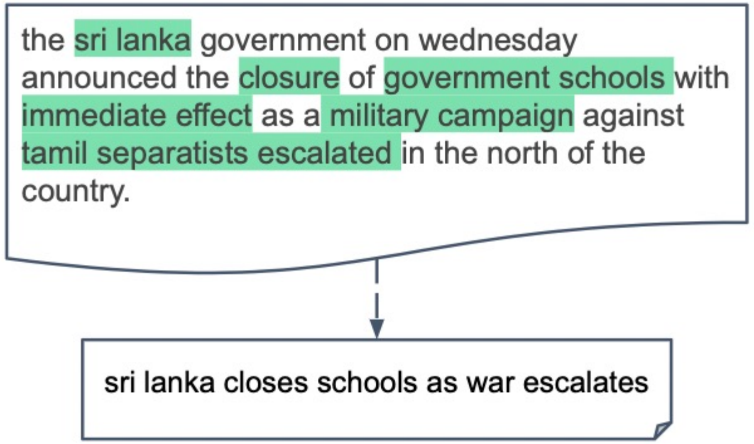


Entropy-based Sampling for Abstractive Multi-document Summarization in Low-resource Settings

Laura Mascarell, Ribin Chalumattu, and Julien Heitmann

Introduction and Background
Abstractive Multi-document Summarization (MDS)

Abstractive Summarization

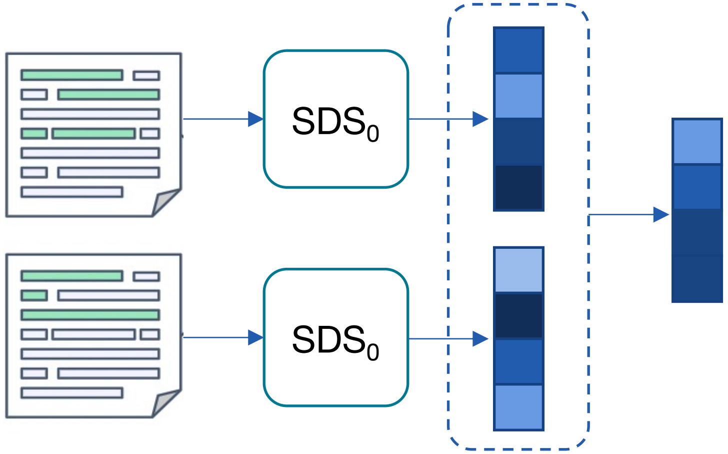


→ Research in MDS mostly focuses on English and MDS training data

Introduction and Background
Abstractive Multi-document Summarization (MDS)

Single-document summarization (SDS) models for MDS task

→ Dynamic ensemble (Hokamp et al., 2020)

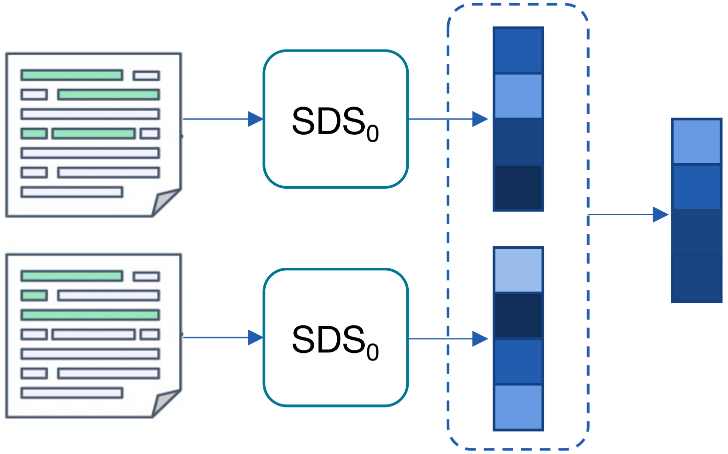


$$p_\theta(y_t|\mathcal{X}) = \frac{1}{|\mathcal{X}|} \sum_{\mathbf{x_i} \in \mathcal{X}} p_\theta(y_t|\mathbf{x}_i; \mathbf{y}_{0:t-1})$$

Introduction and Background
Abstractive Multi-document Summarization (MDS)

Single-document summarization (SDS) models for MDS task

→ Dynamic ensemble (Hokamp et al., 2020)



$$p_\theta(y_t|\mathcal{X}) = \frac{1}{|\mathcal{X}|} \sum_{\mathbf{x_i} \in \mathcal{X}} p_\theta(y_t|\mathbf{x}_i; \mathbf{y}_{0:t-1})$$

Contributions

- Entopy-based sampling approaches for the MDS task
- German MDS test set for abstractive MDS

Entropy-based Decoding

1. Minimum Entropy

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, logp(xi)$$

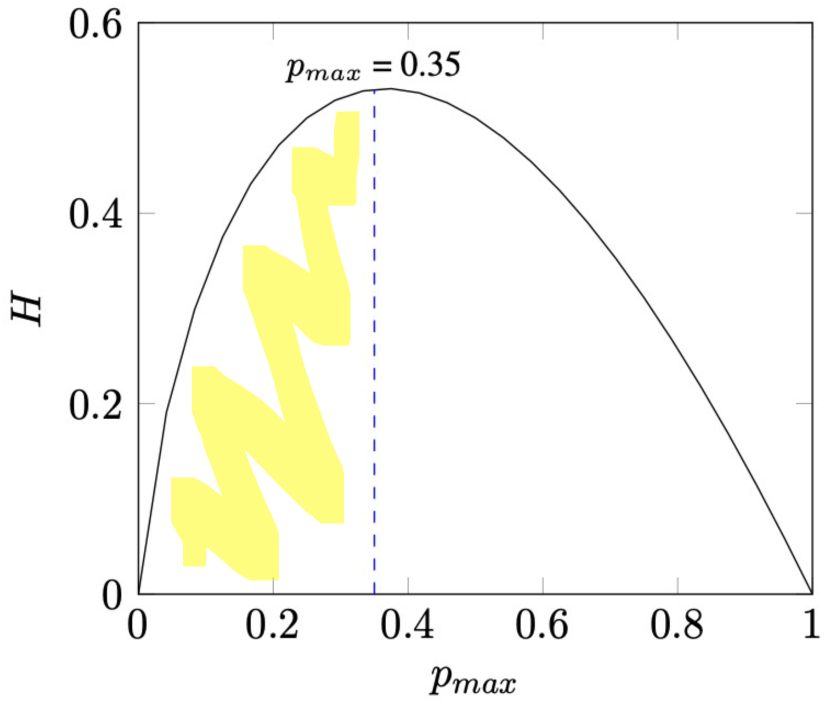
Entropy-based Decoding

1. Minimum Entropy

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, logp(xi)$$

2. Max-predicted Probability Threshold

$$H(X) = -p_{max}\, log\, p_{max}$$



Entropy-based Decoding

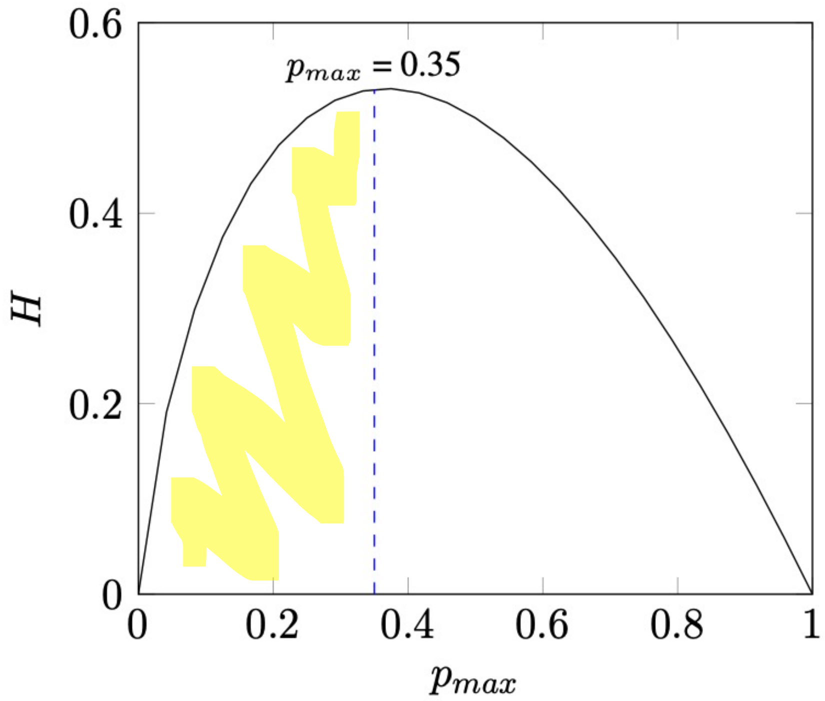
1. Minimum Entropy

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, logp(xi)$$

2. Max-predicted Probability Threshold

$$H(X) = -p_{max}\, log\, p_{max}$$

   a) $p_{max} > 0.35$: choose higuest p_{max} prediction
   b) $p_{max} < 0.35$: average log-probabilities



Entropy-based Decoding

1. Minimum Entropy

$$H(X) = -\sum_{i=1}^{n} p(x_i)\, logp(xi)$$

2. Max-predicted Probability Threshold

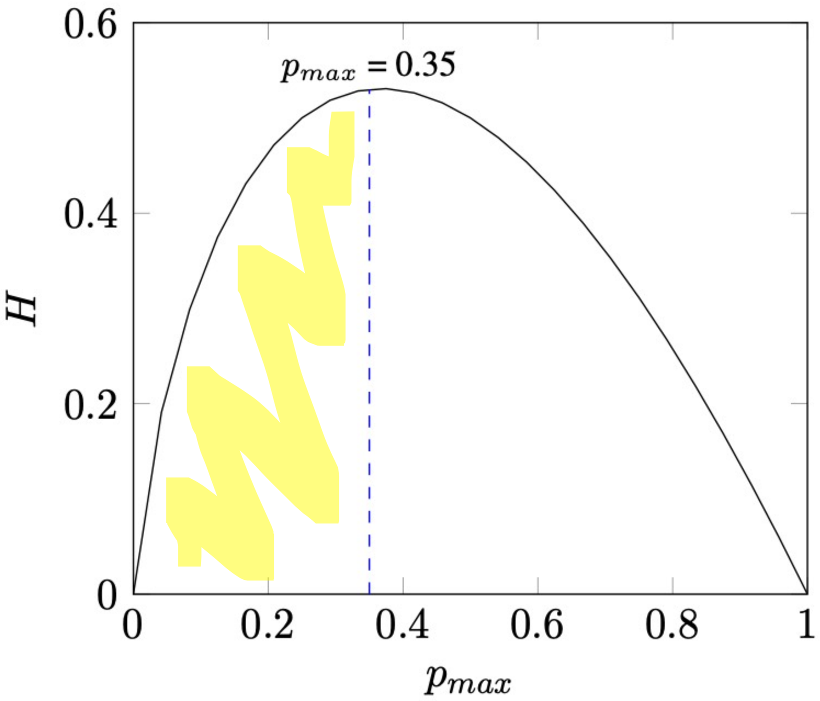
$$H(X) = -p_{max}\, log\, p_{max}$$

   a) $p_{max} > 0.35$: choose higuest p_{max} prediction

   b) $p_{max} < 0.35$: average log-probabilities

3. Pointwise Mutual Information

$$p(y|\boldsymbol{x}) = logp(y|\boldsymbol{x}) - logp(y)$$



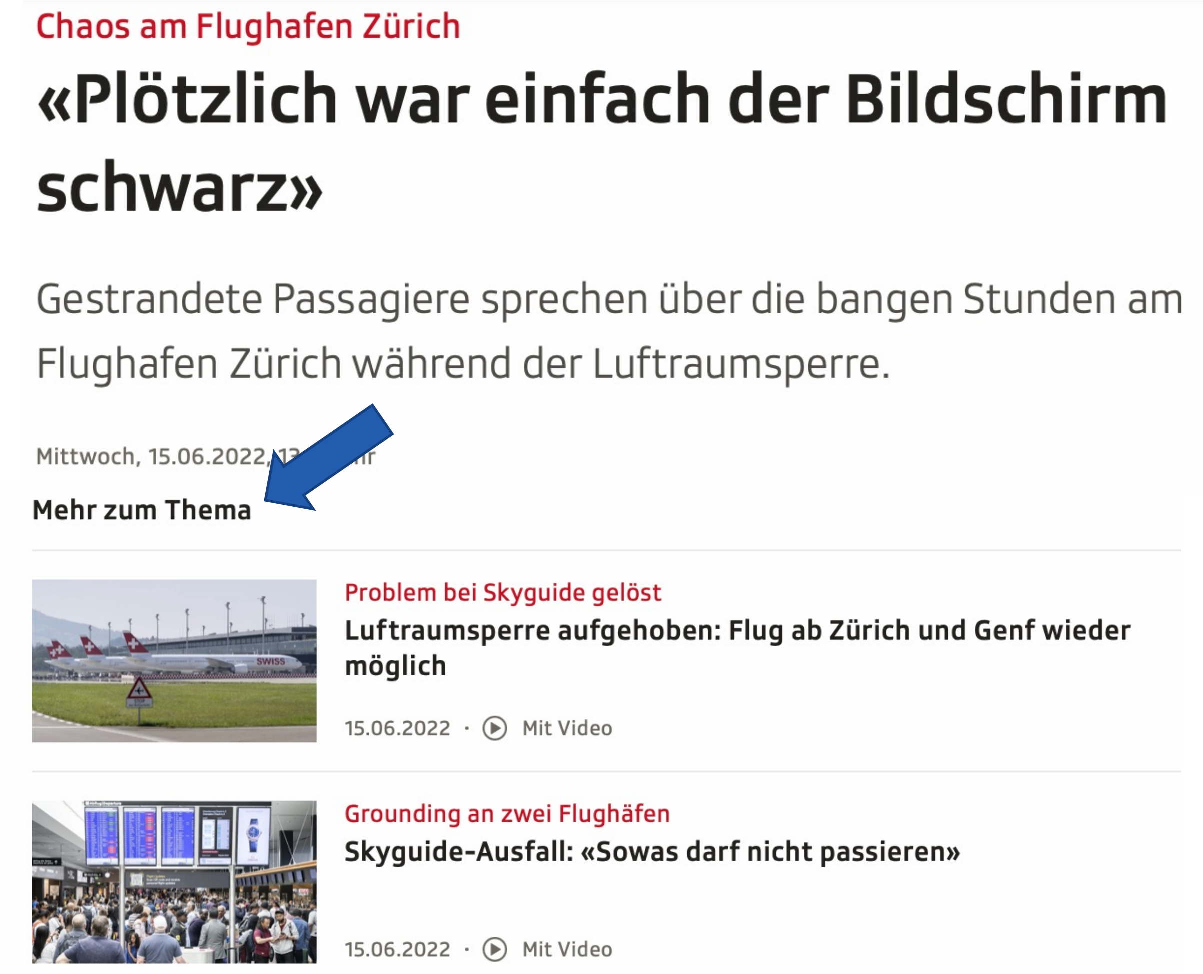
Multi-GeNews
German Abstractive MDS Test Set

Data collection

1. Obtain clusters of related articles from SRF news portal.[1]
2. Generate summary by combining their lead paragraphs.
3. Filter salient summary sentences using mBertExt

Final test set

- 402 clusters of articles + summary
- 2 to 6 related articles per cluster


Chaos am Flughafen Zürich

«Plötzlich war einfach der Bildschirm schwarz»

Gestrandete Passagiere sprechen über die bangen Stunden am Flughafen Zürich während der Luftraumsperre.

Mittwoch, 15.06.2022,

Mehr zum Thema

Problem bei Skyguide gelöst

Luftraumsperre aufgehoben: Flug ab Zürich und Genf wieder möglich

15.06.2022 · Mit Video

Grounding an zwei Flughäfen

Skyguide-Ausfall: «Sowas darf nicht passieren»

15.06.2022 · Mit Video

1. https://www.srf.ch/news

Experimental Setting

Models

- mBART (Liu et al., 2020) fine-tuned on:
  1. 20m (Rios et al., 2021)
  2. auto-hMDS (Zopf, 2018)
- GPT-2 (Radford et al., 2019)

Approaches

- H_{min} - minimum entropy
- H_{th} - max-predicted probability threshold
- H_{pmi} - pointwise mutual information
- *concat* - articles are concatenated into a single input
- *DynE* - dynamic ensemble (Hokamp et al., 2020)

Experimental Setting

Models

- **mBART** (Liu et al., 2020) fine-tuned on:
  1. 20m (Rios et al., 2021)
  2. auto-hMDS (Zopf, 2018)
- **GPT-2** (Radford et al., 2019)

	100 words			
Method	**R1↑**	**R2↑**	**RL↑**	ρ↓
mBART concat	23.0	6.0	14.8	9.23
mBART + *DynE*	22.2	4.8	14.9	1.5
mBART + H_{min}	23.4	5.6	15.0	2.46
mBART + H_{th}	**24.5**	6.2	15.6	2.72
mBART + H_{pmi}	23.9	**7.2**	**16.1**	2.78

Table 1: Performance on Multi-GeNews test set. MBart fine-tuned on the 20m dataset.

Approaches

- H_{min} - minimum entropy
- H_{th} - max-predicted probability threshold
- H_{pmi} - pointwise mutual information
- *concat* - articles are concatenated into a single input
- *DynE* - dynamic ensemble (Hokamp et al., 2020)

Manual Evaluation
German Abstractive MDS

Manual Evaluation Tasks

1. Summary Ranking Task
   → relative quality of the summaries
2. Faithfulness Annotation Task
   → summary text spans (Krishna et al., 2023)

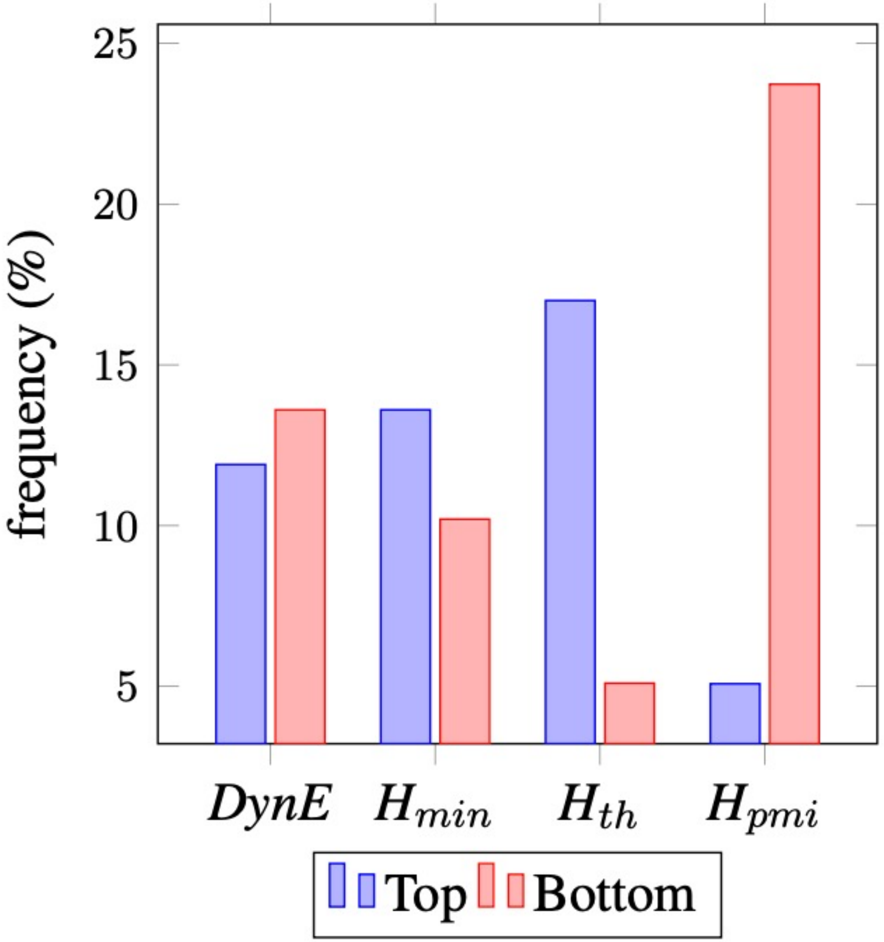
Evaluation on 20 instances of Multi-GeNews

- 3 source articles
- Lexical differences between summaries
  - Token-level edit distance between 5 and 10

Manual Evaluation
German Abstractive MDS

Summary Ranking Task - majority agreement

- DynE mixed ratings
- H_{th} consistently ranked top (1st- 2nd rank)
- H_{pmi} consistently ranked bottom (3rd-4th rank)



Manual Evaluation
German Abstractive MDS

Summary Ranking Task - majority agreement

- DynE mixed ratings
- H_{th} consistently ranked top (1st- 2nd rank)
- H_{pmi} consistently ranked bottom (3rd-4th rank)

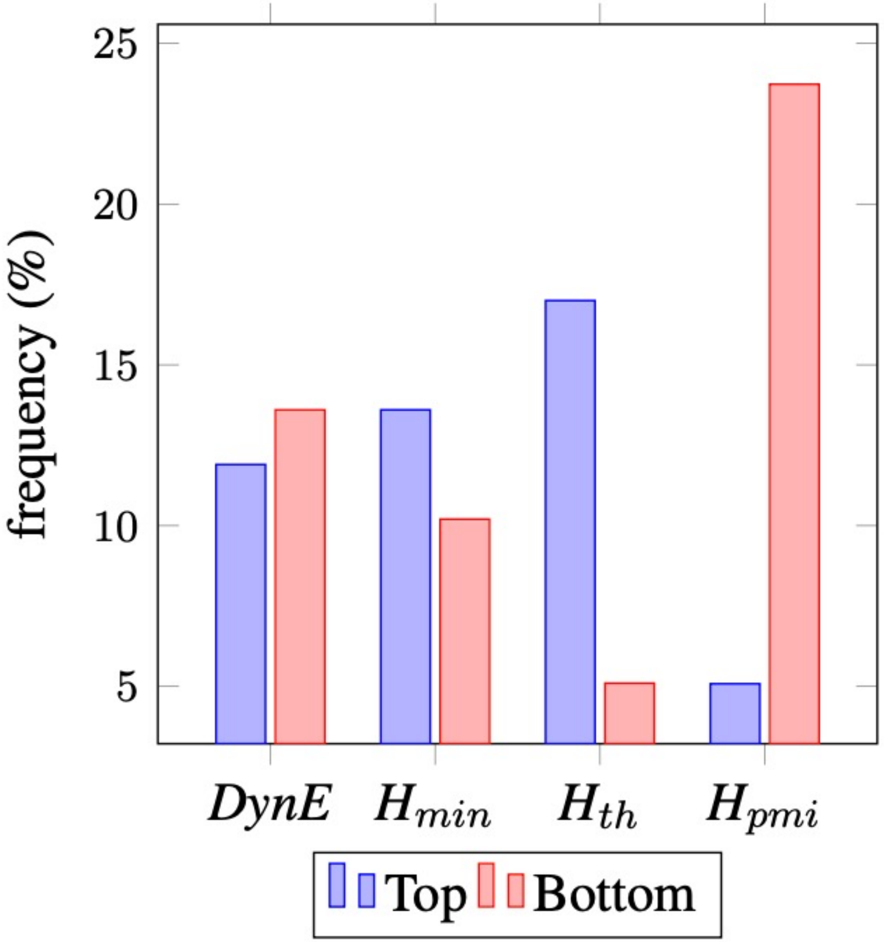
H_{th}
Donald Trump hielt sich in der Nacht auf Mittwoch in den beiden Kammern des US-Kongresses seine dritte Rede ab. [Donald Trump delivered his third speech to both chambers of the U.S. Congress on Wednesday night.]

H_{pmi}
Donald Trump hielt sich in den USA nicht an die Corona-Regeln. [Donald Trump did not follow the Covid rules in the USA.]

Manual Evaluation
German Abstractive MDS

Summary Ranking Task - majority agreement

- DynE mixed ratings
- H_{th} consistently ranked top (1st- 2nd rank)
- H_{pmi} consistently ranked bottom (3rd-4th rank)



Faithfulness Annotation Task

- Overall, lots of hallucinations
- No significant improvement with H_{pmi}
  - 36.2% (H_{pmi}) vs. 33.3% (H_{th})

Conclusions

Abstractive MDS of German text

- Explore multiple entropy-based approaches
- Build and release a test set on the news domain

Future work

- Important to tackle hallucination
- MDS with Large Language Models

Thank you!

Link to paper

Entropy-based Sampling for Abstractive Multi-document Summarization in Low-resource Settings

Authors: Laura Mascarell, Ribin Chalumattu, and Julien Heitmann