Story Highlight Generation

Pedro Balage, Sebastião Miranda, Afonso Mendes, André Martins,
Nikos Papasarantopoulos, Shashi Narayan, Shay Cohen

pedro.balage@priberam.pt, sebastiao.miranda@priberam.pt, amm@priberam.pt, afm@priberam.pt
nikos.papasa@ed.ac.uk, shashi.narayan@ed.ac.uk, scohen@inf.ed.ac.uk

Goals

- To synthesize a coherent summary of all events that occur in the story.
- To enable a person monitoring the media to see a global picture and evolution of a story.

Coverage-based Extractive Summarization

- The goal is to **maximize the score of the summary** given the constraints for **concept coverage** and maximum number of words allowed (budget).
- Uses an integer linear programming (ILP) solver.

Coverage-based Extractive + Seq2seq Abstractive

- First step extracts the most relevant sentences (extractive).
- Second step performs a **sequence-to-sequence neural network** to convert each sentence into a highlight (abstractive).
- Example of summary for a CNN article

Extractive Summary:
- Washington (CNN)President Donald Trump insisted Thursday that he has remained consistent in his plans for a border wall, a day after his chief of staff, John Kelly, told Fox News he has changed his attitude on it.
- Trump was fuming after Kelly's Fox News interview, a source familiar with the matter told CNN, adding that the President hated the comments.
- Kelly on immigration: Trump has changed the way he's looked at a number of things
- Including services, the trade deficit with Mexico was $68.3 billion from October 2016 through September 2017, the most recent figures available

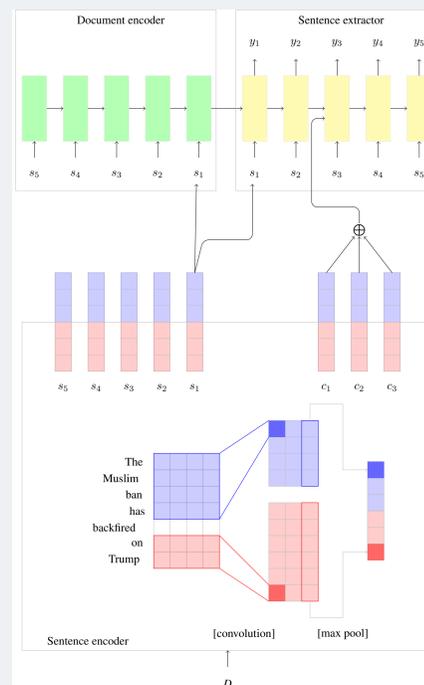**Extractive+Abstractive Summary**:
- Donald Trump said he has remained consistent in his plans for a border wall
- Trump was fuming after Kelly's Fox News interview
- Trump has changed the way he's looked at a number of things
- The trade deficit with Mexico was $68.3 billion from October 2016, the most recent figures available

Summarization using AMR-to-text Generation
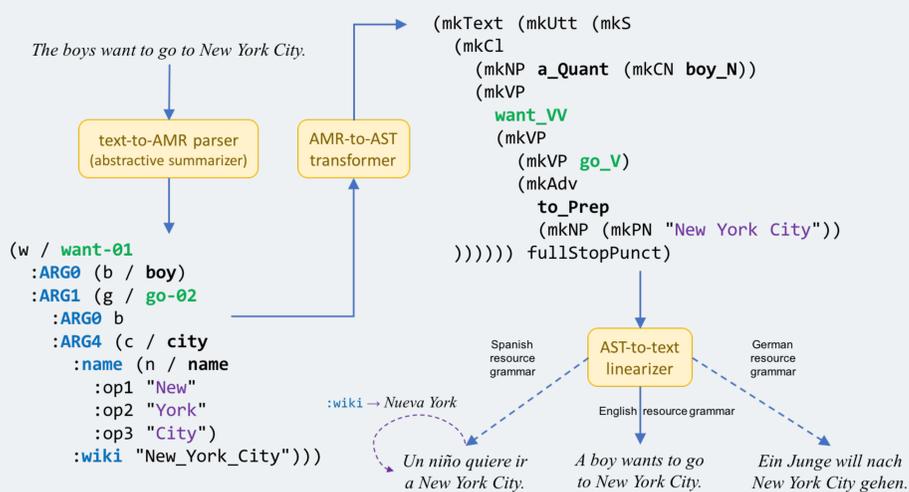
- An NLG component for an abstractive summariser using the **Abstractive Meaning Representation** (AMR) semantic formalism.
- Transforms AMR graphs (story highlights) into abstract syntax trees (AST), leaving the linearization of the ASTs to the existing language resource grammars and lexicons.

SIDENET: Incorporating Side Information into Neural Extractive Summarization

- We explore the side information of the document, such as **title** and **image captions**, where the gist of the document often lies.
- We **outperform** models which do not use side information in terms of **ROUGE (recall)**.

REFRESH: Reinforcement Learning-based Neural Extractive Summarization

Models	CNN+DailyMail		
	R1	R2	RL
LEAD	39.6	17.7	36.2
Cheng and Lapata (2016)	35.5	14.7	32.2
Nallapati et al. (2016)	35.4	13.3	32.6
Nallapati et al. (2017)	39.6	16.2	35.3
See et al. (2017)	39.5	17.3	36.4
Tan and Wan (2017)	38.1	13.9	34.0
REFRESH	**40.0**	**18.2**	**36.6**

Performance on CNN/Daily Mail test set using full-length Rouge (F1) metric.

- Extractive summarization as a **sentence ranking task** with a novel training algorithm which **globally optimizes** the ROUGE evaluation metric through a **reinforcement learning** objective.
- It is **better at discriminating among sentences** for the final summary; **a sentence is ranked high for selection if it often occurs in high scoring summaries**.
- Our model **outperforms** state-of-the-art extractive and abstractive systems when evaluated both automatically and by humans.

Demo

Coverage-based https://tinyurl.com/ybtc95do
SIDENET: https://tinyurl.com/y97olc5p
REFRESH: https://tinyurl.com/yczk569y