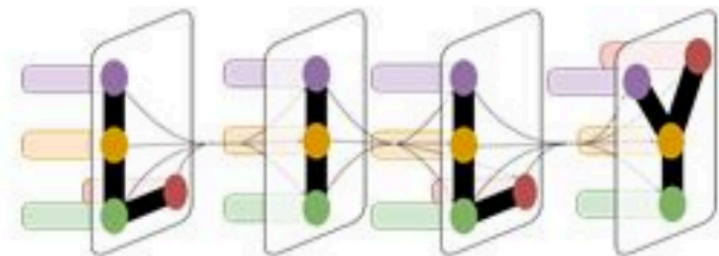


Multi-News: a Large-Scale Multi-Document Summarization Dataset and Abstractive Hierarchical Model

Alexander R. Fabbri, Irene Li, Tianwei She, Suyi Li
Dragomir R. Radev

Yale University



Outline

- Introduction to Multi-Document Summarization (MDS)
- Multi-News Dataset
- Hierarchical MMR-Attention Pointer-generator (Hi-MAP) model and analysis

Multi-Document Summarization

Source 1

Meng Wanzhou, Huawei's chief financial officer and deputy chair, was arrested in Vancouver on 1 December. Details of the arrest have not been released...

Source 2

A Chinese foreign ministry spokesman said on Thursday that Beijing had separately called on the US and Canada to "clarify the reasons for the detention" immediately and "immediately release the detained person". The spokesman...

Source 3

Canadian officials have arrested Meng Wanzhou, the chief financial officer and deputy chair of the board for the Chinese tech giant Huawei,...Meng was arrested in Vancouver on Saturday and is being sought for extradition by the United States...

Multi-Document Summarization

Source 1

Meng Wanzhou, Huawei's chief financial officer and deputy chair, was arrested in Vancouver on 1 December. Details of the arrest have not been released...

Source 2

A Chinese foreign ministry spokesman said on Thursday that Beijing had separately called on the US and Canada to "clarify the reasons for the detention" immediately and "immediately release the detained person". The spokesman...

Source 3

Canadian officials have arrested Meng Wanzhou, the chief financial officer and deputy chair of the board for the Chinese tech giant Huawei,...Meng was arrested in Vancouver on Saturday and is being sought for extradition by the United States...

Summary

...Canadian authorities say she was being sought for extradition to the US, where the company is being investigated for possible violation of sanctions against Iran. Canada's justice department said Meng was arrested in Vancouver on Dec. 1... China's embassy in Ottawa released a statement.. "The Chinese side has lodged stern representations with the US and Canadian side, and urged them to immediately correct the wrongdoing" and restore Meng's freedom, the statement said ...

Related work

- Train on limited data from DUC 03-04 (Paul and James, 2004) and TAC (Owczarzak and Dang, 2011)
 - Using graph structure (Yasunaga et al., 2017) or through an auxiliary text classification task (Cao et al., 2017)
- Train on single-document summarization data and test on MDS data
 - Lebanoff et al., 2018; Baumel et al., 2018; Zhanget al., 2018b
- Weakly and Un-supervised approaches
 - Angelidis and Lapata 2018; Chu and Liu 2019
- Large-scale datasets based on Wikipedia
 - WikiSum (Liu et al 2018) and derivative datasets (Liu and Lapata 2019; Perez-Beltrachini et al. 2019)

newser.com

Woman Buys Out Entire Payless Store to Help Those in Need

Carrie Jernigan says it all began with a simple request from her daughter to donate one pair



By **Jenn Gidman**, Newser Staff
Posted Jul 27, 2019 8:00 AM CDT



17 comments



This Aug. 23, 2006, file photo shows a Payless store in Philadelphia. (AP Photo/Matt Rourke, File)

[View 1 more image](#)

(**NEWSER**) – It started with a pair of Avengers-branded shoes. One day in May, attorney Carrie Jernigan was shopping at a Payless near her home in Alma, Ark., with her kids, trying to grab as many pairs of deeply discounted footwear as she could after [Payless announced](#) it would soon be closing all of its stores, per [CNN](#). But when her 9-year-old daughter asked if they could buy a pair of Avengers sneakers for a classmate in need, Jernigan was struck with an idea: What if she just bought out the entire store

and donated all of the shoes? And so she half-jokingly asked the clerk how much she'd want for the 350 or so pairs of shoes still perched on the store's shelves. Per [KFSM](#), the clerk asked for Jernigan's phone number and told her she'd let her know about her request.

<https://tinyurl.com/y5lbbmro>

Dataset Analysis

Source documents per summary

# of source	Frequency	# of source	Frequency
2	23,894	7	382
3	12,707	8	209
4	5,022	9	89
5	1,873	10	33
6	763		

Comparison to other datasets

Dataset	# pairs	# words (docs)	# sents (docs)	# words (summary)	# sents (summary)	vocab size
Multi-News	44,972/5,622/5,622	2,103.49	82.73	263.66	9.97	666,515
DUC03+04	320	4,636.24	173.15	109.58	2.88	19,734
TAC 2011	176	4,695.70	188.43	99.70	1.00	24,672
CNNDM	287,227/13,368/11,490	810.57	39.78	56.20	3.68	717,951

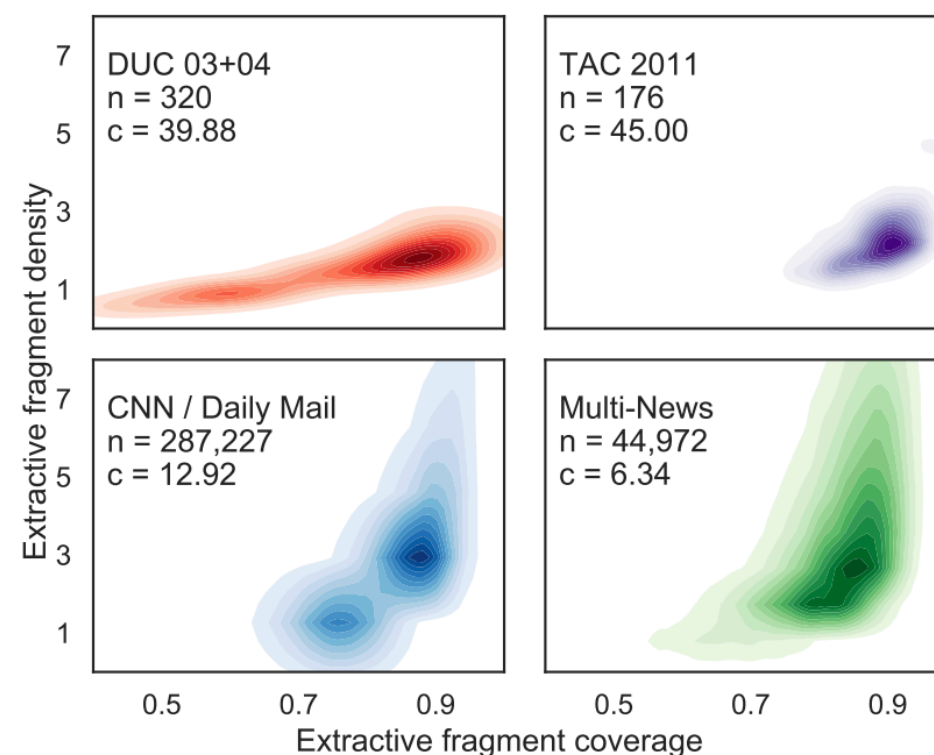
Dataset Analysis cont.

Novel n-grams

% novel n-grams	Multi-News	DUC03+04	TAC11	CNNDM
uni-grams	17.76	27.74	16.65	19.50
bi-grams	57.10	72.87	61.18	56.88
tri-grams	75.71	90.61	83.34	74.41
4-grams	82.30	96.18	92.04	82.83

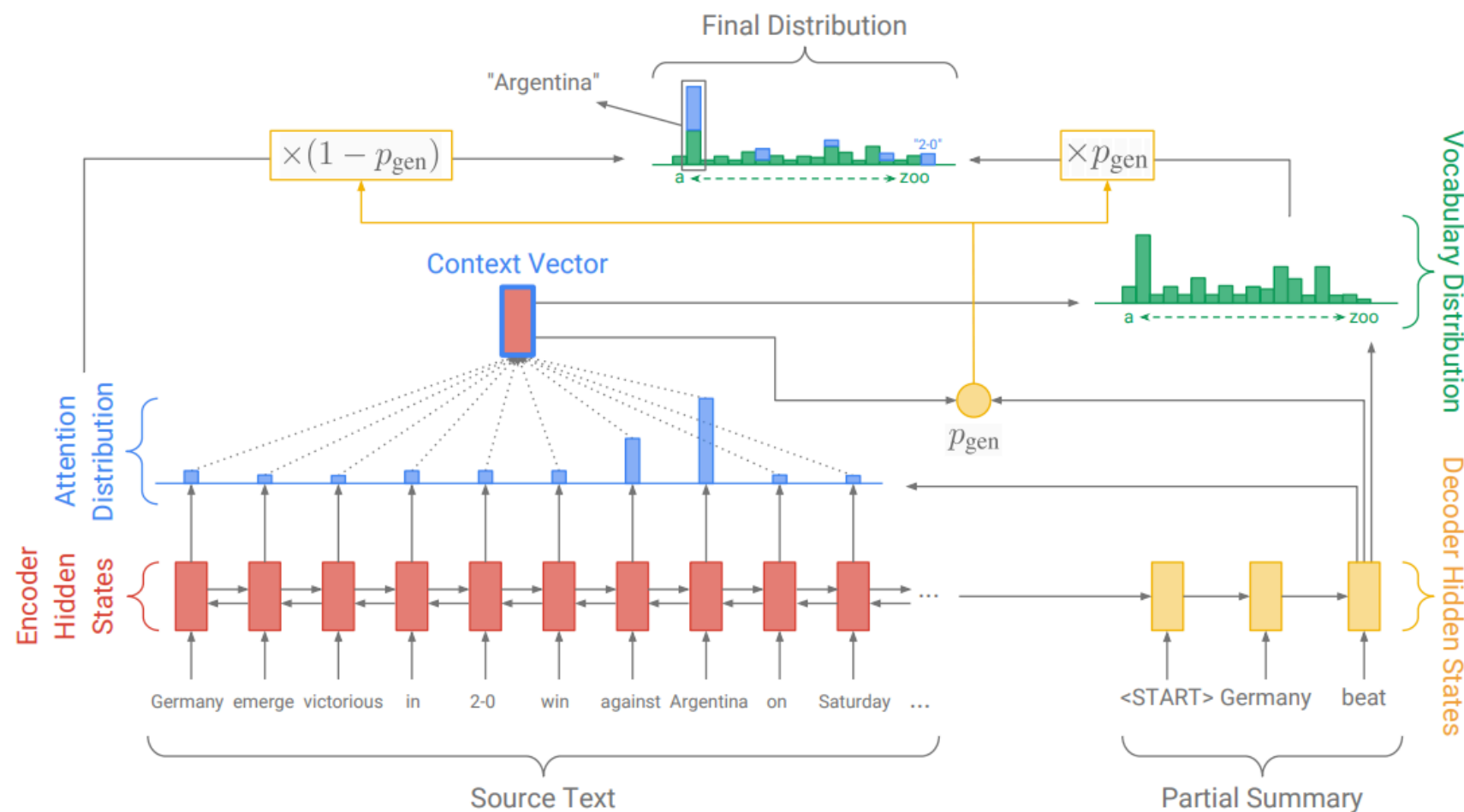
X-axis: variability in the average length of the extractive fragments to which summary words belong

Y-axis: variation in the average length of source sequences present in the summary.



Background: Pointer-Generator

- See et al. 2017



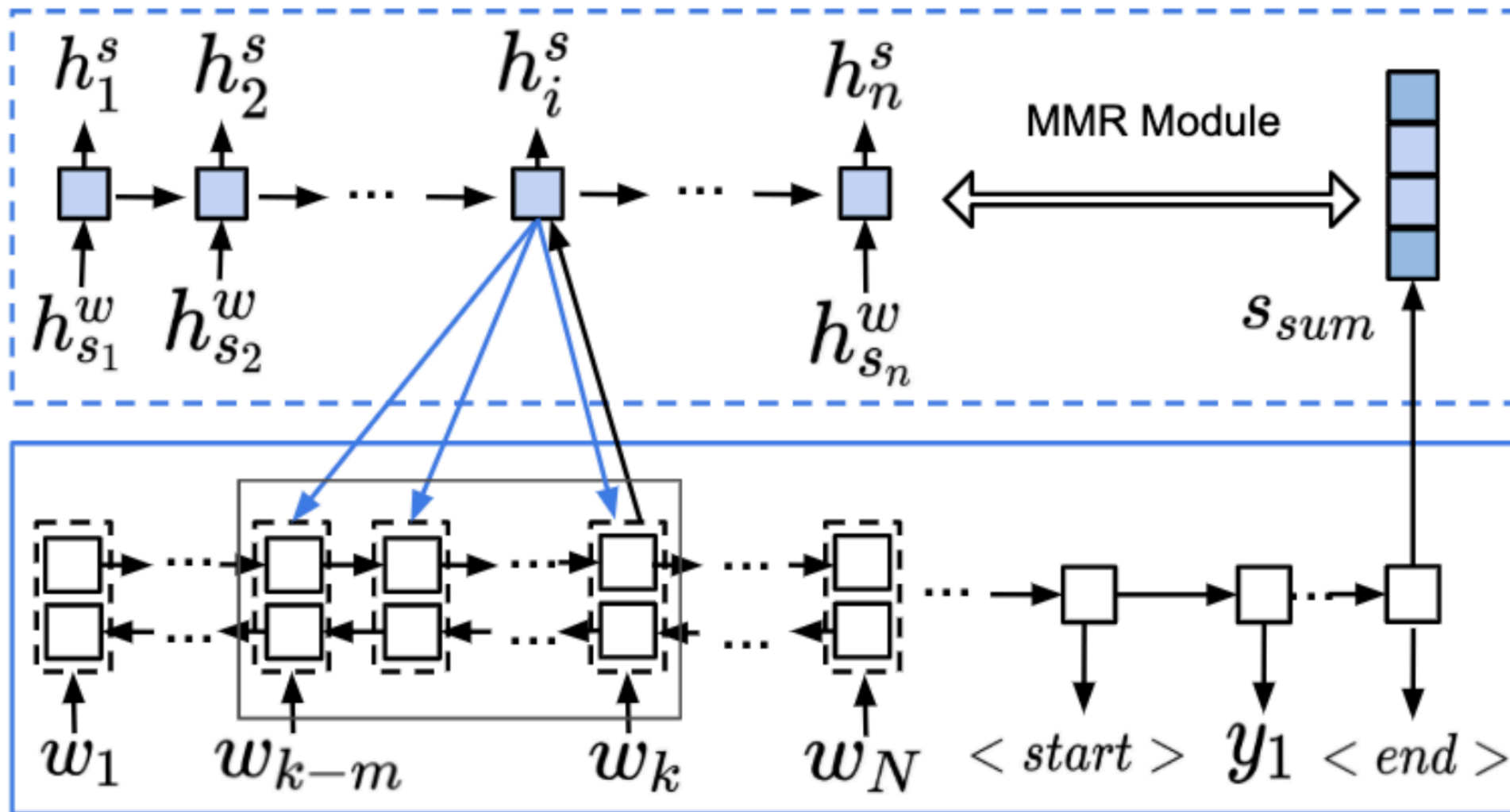
Background: Maximal Marginal Relevance (MMR)

- Carbonell and Goldstein, 1998
- Combines query-relevance with information-

novelty

$$MMR = \underset{D_i \in R \setminus S}{argmax} \left[\lambda Sim_1(D_i, Q) - (1 - \lambda) \max_{D_j \in S} Sim_2(D_i, D_j) \right]$$

Hi-MAP Model



$$MMR_i = \lambda Sim_1(h_i^s, s_{sum}) - (1 - \lambda) \max_{s_j \in D, j \neq i} Sim_2(h_i^s, h_j^s)$$

MMR Module

- Similarity between current summary and hidden states
- Similarity score among hidden states

$$Sim_1 = h_i^{s^T} W_{sim} s_{sum}$$

$$v_{ij} = \tanh(h_j^{s^T} W_{self} h_i^s)$$

$$\beta_{ij} = \frac{\exp(v_{ij})}{\sum_j \exp(v_{ij})}$$

$$score_i = \max_j(\beta_{i,j})$$

MMR Module cont.

$$MMR_i = \lambda Sim_1(h_i^s, s_{sum}) - (1 - \lambda) score_i$$

$$\overline{MMR_i} = \frac{\exp(MMR_i)}{\sum_i \exp(MMR_i)}$$

$$\overline{a^t} = a^t \overline{MMR_i}$$

Automatic Evaluation

Results on DUC 04

Method	R-1	R-2	R-SU
First	30.77	8.27	7.35
LexRank (Erkan and Radev, 2004)	35.56	7.87	11.86
TextRank (Mihalcea and Tarau, 2004)	33.16	6.13	10.16
MMR (Carbonell and Goldstein, 1998)	30.14	4.55	8.16
PG-Original (Lebanoff et al., 2018)	31.43	6.03	10.01
PG-MMR (Lebanoff et al., 2018)	36.42	9.36	13.23
PG-BRNN (Gehrmann et al., 2018)	29.47	6.77	7.56
CopyTransformer (Gehrmann et al., 2018)	28.54	6.38	7.22
Hi-MAP (Our Model)	35.78	8.90	11.43

Results on Multi-News

Method	R-1	R-2	R-SU
First-1	26.83	7.25	6.46
First-2	35.99	10.17	12.06
First-3	39.41	11.77	14.51
LexRank (Erkan and Radev, 2004)	38.27	12.70	13.20
TextRank (Mihalcea and Tarau, 2004)	38.44	13.10	13.50
MMR (Carbonell and Goldstein, 1998)	38.77	11.98	12.91
PG-Original (Lebanoff et al., 2018)	41.85	12.91	16.46
PG-MMR (Lebanoff et al., 2018)	40.55	12.36	15.87
PG-BRNN (Gehrmann et al., 2018)	42.80	14.19	16.75
CopyTransformer (Gehrmann et al., 2018)	43.57	14.03	17.37
Hi-MAP (Our Model)	43.47	14.89	17.41

Human Evaluations

Best-Worst Scaling (Louviere and Woodworth, 1991; Louviere et al., 2015)

Method	Informativeness	Fluency	Non-Redundancy
PG-MMR	95	70	45
Hi-MAP	85	75	100
CopyTransformer	99	100	107
Human	150	150	149

Example

Input (truncated)

buh-bye barbie . hello lego ! danish toy manufacturer lego group has taken the title of the biggest toymaker in the world in terms of revenue and profit . the popular lego movie helped propel sales 11 % higher in the first six months of the year . " we are thrilled by the reception of the lego movie products , " said lego group cfo john goodwin . sales across europe , the americas and asia increased by double digits compared to the same period in the previous year . the family-owned business surpassed top competitor , mattel (mat) , which is struggling with slumping global sales and slower demand for barbie dolls and fisher-price toys .

Example

Input (truncated)

buh-bye barbie . hello lego ! danish toy manufacturer lego group has taken the title of the biggest toymaker in the world in terms of revenue and profit . the popular lego movie helped propel sales 11 % higher in the first six months of the year . " we are thrilled by the reception of the lego movie products , " said lego group cfo john goodwin . sales across europe , the americas and asia increased by double digits compared to the same period in the previous year . the family-owned business surpassed top competitor , mattel (mat) , which is struggling with slumping global sales and slower demand for barbie dolls and fisher-price toys .

PPG-MMR

– -rrb- lego movie products provided a significant boost to our sales during the first six months of 2014 , mostly because of the success of the " lego movie products ." we are thrilled by the reception of the lego movie products , " said that support could continue in the second half of the year , when lego movie will be launched on dvd ...

Example

Input (truncated)

buh-bye barbie . hello lego ! danish toy manufacturer lego group has taken the title of the biggest toymaker in the world in terms of revenue and profit . the popular lego movie helped propel sales 11 % higher in the first six months of the year . " we are thrilled by the reception of the lego movie products , " said lego group cfo john goodwin . sales across europe , the americas and asia increased by double digits compared to the same period in the previous year . the family-owned business surpassed top competitor , mattel (mat) , which is struggling with slumping global sales and slower demand for barbie dolls and fisher-price toys .

PPG-MMR

– -rrb- lego movie products provided a significant boost to our sales during the first six months of 2014 , mostly because of the success of the " lego movie products ." we are thrilled by the reception of the lego movie products , " said that support could continue in the second half of the year , when lego movie will be launched on dvd ...

Example

Input (truncated)

buh-bye barbie . hello lego ! danish toy manufacturer lego group has taken the title of the biggest toymaker in the world in terms of revenue and profit . the popular lego movie helped propel sales 11 % higher in the first six months of the year . " we are thrilled by the reception of the lego movie products , " said lego group cfo john goodwin . sales across europe , the americas and asia increased by double digits compared to the same period in the previous year . the family-owned business surpassed top competitor , mattel (mat) , which is struggling with slumping global sales and slower demand for barbie dolls and fisher-price toys .

PPG-MMR

– -rrb- **lego movie products** provided a significant boost to our sales during the first six months of 2014 , mostly because of the success of the " **lego movie products** ." we are thrilled by the reception of the **lego movie products** , " said that support could continue in the second half of the year , when **lego movie** will be launched on dvd ...

Hi-MAP

– revenue soared 11 percent in the first six months of 2014 , mostly because of the success of the " lego movie . " " the lego movie , which was released in february by cnnmoney ' s parent company , time warner (twx) , made \$ 69 million in its opening weekend and has grossed nearly \$ 260 million so far , according to box office mojo ...

Example

Input (truncated)

buh-bye barbie . hello lego ! danish toy manufacturer lego group has taken the title of the biggest toymaker in the world in terms of revenue and profit . the popular lego movie helped propel sales 11 % higher in the first six months of the year . " we are thrilled by the reception of the lego movie products , " said lego group cfo john goodwin . sales across europe , the americas and asia increased by double digits compared to the same period in the previous year . the family-owned business surpassed top competitor , mattel (mat) , which is struggling with slumping global sales and slower demand for barbie dolls and fisher-price toys .

PPG-MMR

– -rrb- **lego movie products** provided a significant boost to our sales during the first six months of 2014 , mostly because of the success of the " **lego movie products** ." we are thrilled by the reception of the **lego movie products** , " said that support could continue in the second half of the year , when **lego movie** will be launched on dvd ...

Hi-MAP

– revenue soared 11 percent in the first six months of 2014 , mostly because of the success of the " lego movie . " " the lego movie , which was released in february by cnnmoney ' s parent company , time warner (twx) , made \$ 69 million in its opening weekend and has grossed nearly \$ 260 million so far , according to box office mojo ...

Transformer

– toy maker lego says revenue soared 11 % in the first six months of 2014 , mostly because of the success of the " lego movie " products , making it the biggest toy company in the world by revenue and profit , the ap reports ...

Future Work

- Experiment further with extractive methods
- Combine Hi-MAP with transformer
- Exploit document relations rather than flat concatenated input
- Domain adaptation for MDS

Conclusion

- Introduced large-scale news MDS dataset
- Introduced Hi-MAP model for MDS
- Check out the data!
 - <https://github.com/Alex-Fabbri/Multi-News>

References

- Sebastian Gehrmann, Yuntian Deng, Alexander M. Rush: Bottom-Up Abstractive Summarization. EMNLP 2018: 4098-4109
- Logan Lebanoff, Kaiqiang Song, Fei Liu: Adapting the Neural Encoder-Decoder Framework from Single to Multi-Document Summarization. EMNLP 2018: 4131-4141
- Michihiro Yasunaga, Rui Zhang, Kshitijh Meelu, Ayush Pareek, Krishnan Srinivasan, Dragomir R. Radev: Graph-based Neural Multi-Document Summarization. CoNLL 2017: 452-462
- Paul Over, James Yen: An Introduction to DUC-2004. DUC 2004.
- Ziqiang Cao, Wenjie Li, Sujian Li, Furu Wei: Improving Multi-Document Summarization via Text Classification. AAAI 2017: 3053-3059 2016
- Karolina Owczarzak, Hoa Trang Dang: Overview of the TAC 2010 Summarization Track.
- See, Abigail, Peter J. Liu, and Christopher D. Manning. "Get to the point: Summarization with pointer-generator networks." ACL 2017.
- Carbonell, Jaime G., and Jade Goldstein. "The use of MMR, diversity-based reranking for reordering documents and producing summaries." SIGIR. Vol. 98. 1998.
- Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.
- Grusky, Max, Mor Naaman, and Yoav Artzi. "Newsroom: A dataset of 1.3 million summaries with diverse extractive strategies." NAACL 2018.
- Liu, Peter J., et al. "Generating wikipedia by summarizing long sequences." ICLR 2018.
- Baumel, Tal, Matan Eyal, and Michael Elhadad. "Query focused abstractive summarization: Incorporating query relevance, multi-document coverage, and summary length constraints into seq2seq models." arXiv preprint arXiv:1801.07704 (2018).
- Angelidis, Stefanos, and Mirella Lapata. "Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised." EMNLP 2018
- Zhang, Jianmin, Jiwei Tan, and Xiaojun Wan. "Towards a neural network approach to abstractive multi-document summarization." arXiv preprint arXiv:1804.09010 (2018).
- Liu, Yang, and Mirella Lapata. "Hierarchical Transformers for Multi-Document Summarization." ACL 2019.
- Perez-Beltrachini, Laura, Yang Liu, and Mirella Lapata. "Generating Summaries with Topic Templates and Structured Convolutional Decoders." ACL 2019.
- Chu, Eric, and Peter J. Liu. "Unsupervised neural multi-document abstractive summarization." ICML 2019.

Thank you!