Heterogeneous Graph Neural Networks for Extractive Document Summarization

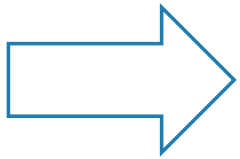
Danqing Wang∗, Pengfei Liu∗, Yining Zheng, Xipeng Qiu†, Xuanjing Huang

ACL2020

Extractive Summarization

- Input: Document $D = \{s_1, \ldots, s_n\}$
- Output: Summary $S = \{y_1, .., y_m\}, m < n$

[1]**Saracens** director of rugby mark mccall lauded his young guns after their latest european heartache before declaring he has no intention of overspending in a competitive post-world cup transfer market. [2]Mccall watched his side, which **contained five english-qualified forwards in the starting pack**, battle in vain before **losing 13-9 to the clermont on saturday**. [3]**Saracens' millionaire chairman nigel wray** spent much of last week repeating his **belief the cap should be scrapped** ....[11]...

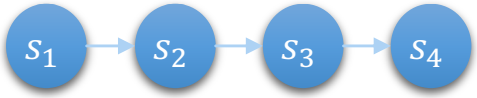


Saracens lost 13-9 to clermont at stade geoffroy-guichard on saturday. The sarries pack contained five english-qualified forwards. Saracens ' millionaire chairman nigel wray wants the salary cap scrapped.

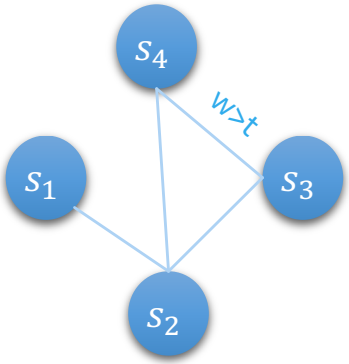
Cross-sentence Relationship Modeling

Q: How to model cross-sentence relationship?

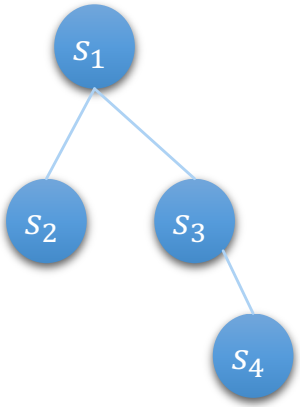
RNN-based models



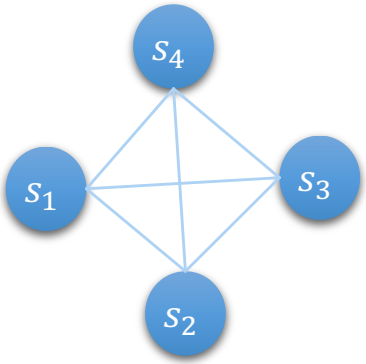
Graph-based models



Similarity-based (TextRank)



Discourse connectivity (ADG, RST)



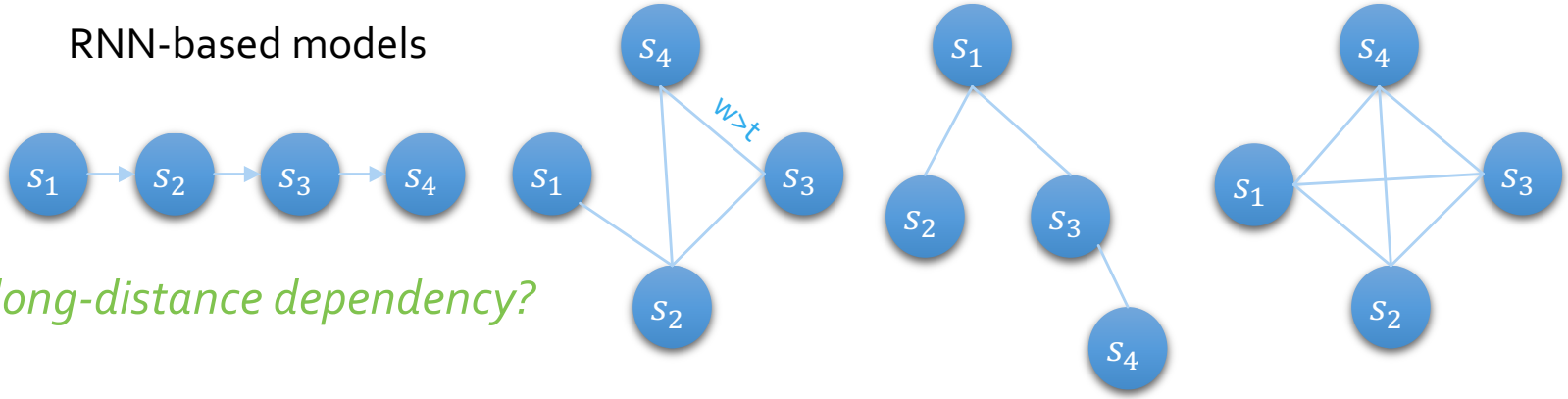
Fully-connected (Transformer)

Cross-sentence Relationship Modeling

Q: How to model cross-sentence relationship?

Graph-based models

RNN-based models



long-distance dependency?

Similarity-based (TextRank)

threshold?

Discourse connectivity (ADG, RST)

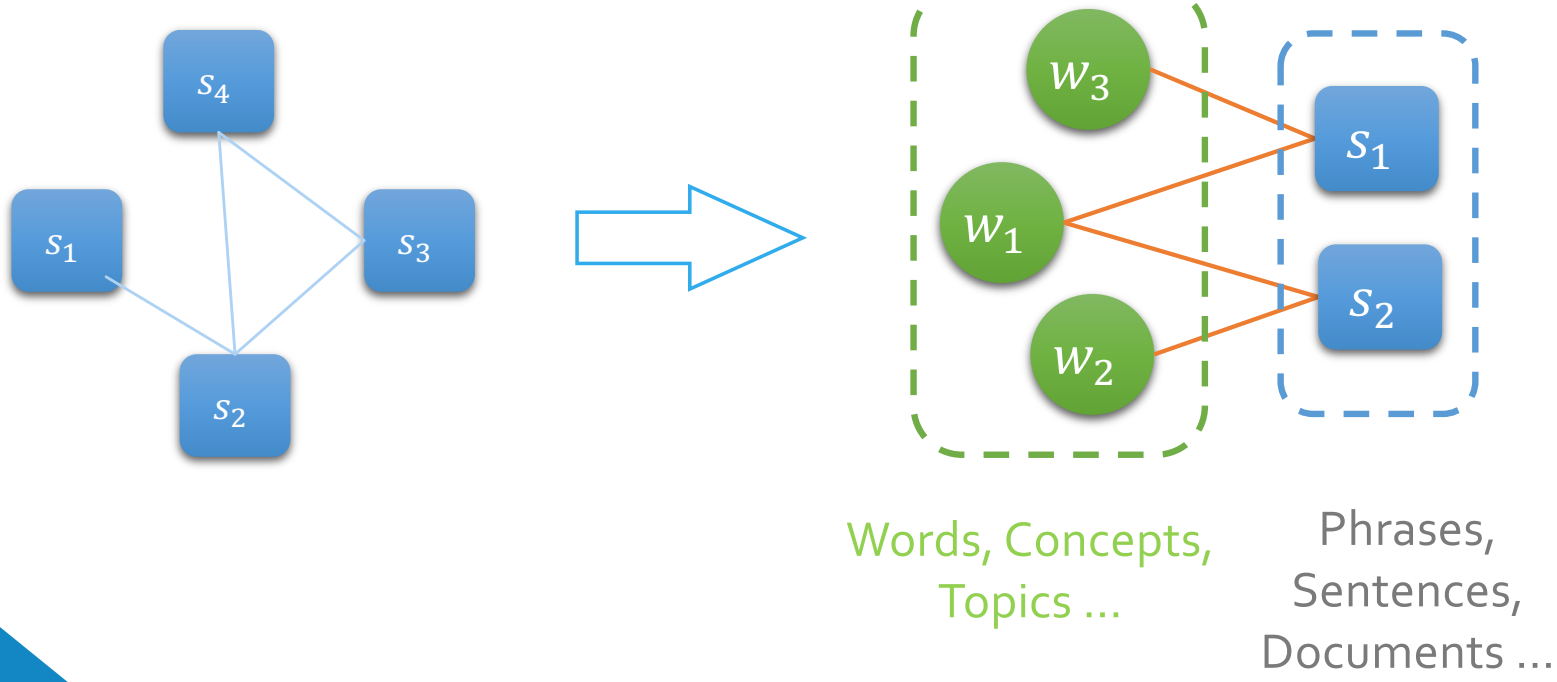
error propagation?

Fully-connected (Transformer)

lack prior?

HeterSumGraph

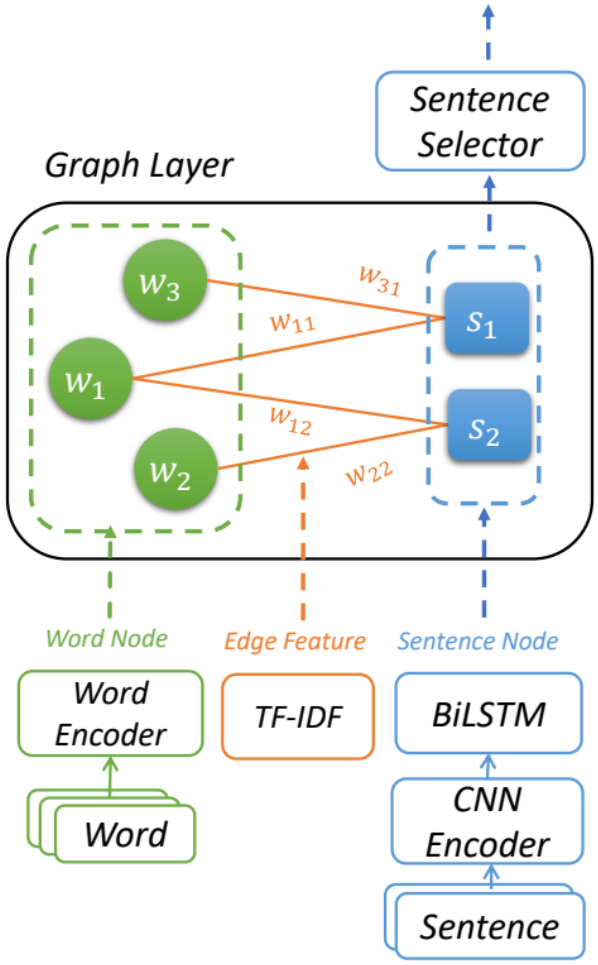
Q: How to model cross-sentence relationship?



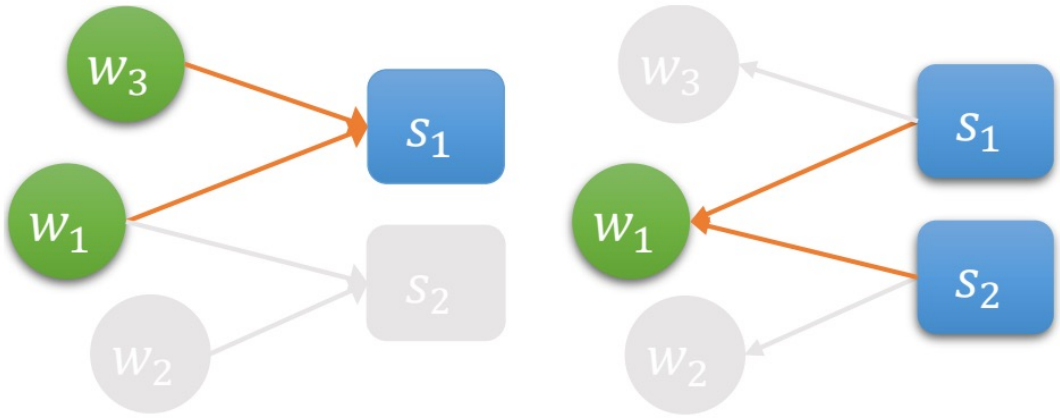
HeterSumGraph

Heterogeneous Summarization Graph (HeterSumGraph)

- Graph initializers
  - word node
  - sentence node
  - edge feature
- Heterogeneous graph layer
  - word -> sentence
  - sentence -> word
- Sentence selector



Update Mechanism

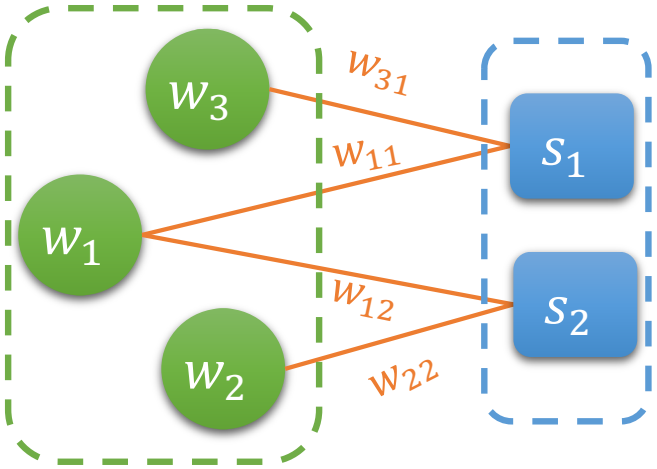


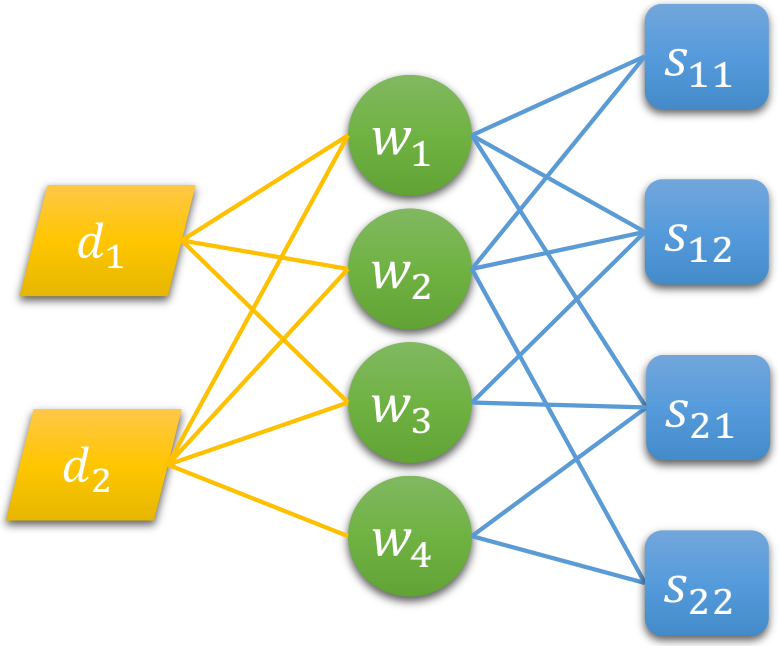
(a) Update s_1 (b) Update w_1

HeterDocSumGraph

Single to Multiple Document Summarization

Q: how to model cross-document relationship?





Evaluation

Experiment Settings for single and multi-document

Datasets:

- CNN/DM: 287,227/13,368/11,490
- NYT50: 96,834/4,000/3,452
- Multi-News: 44,972/5,622/5,622

Models

- Ext-BiLSTM
- Ext-Transformer
- HSG
- HDSG

Evaluation

Some results on single and multi-document datasets

Model	R-1	R-2	R-L
LEAD-3 (See et al., 2017)	40.34	17.70	36.57
ORACLE (Liu and Lapata, 2019b)	52.59	31.24	48.87
REFRESH (Narayan et al., 2018)	40.00	18.20	36.60
LATENT (Zhang et al., 2018)	41.05	18.77	37.54
BanditSum (Dong et al., 2018)	41.50	18.70	37.60
NeuSUM (Zhou et al., 2018)	41.59	19.01	37.98
JECS (Xu and Durrett, 2019)	41.70	18.50	37.90
LSTM+PN (Zhong et al., 2019a)	41.85	18.93	38.13
HER w/o Policy (Luo et al., 2019)	41.70	18.30	37.10
HER w Policy (Luo et al., 2019)	42.30	18.90	37.60
Ext-BiLSTM	41.59	19.03	38.04
Ext-Transformer	41.33	18.83	37.65
HSG	42.31	19.51	38.74
HSG + Tri-Blocking	**42.95**	**19.76**	**39.23**

Table 1: Performance (Rouge) of our proposed models against recently released summarization systems on CNN/DailyMail.

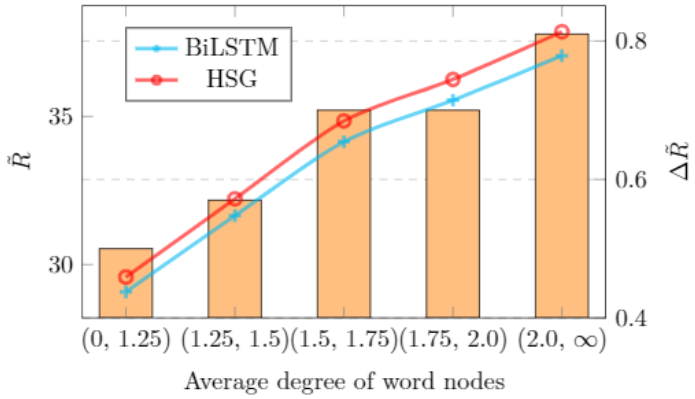
Model	R-1	R-2	R-L
First-1	25.44	7.06	22.12
First-2	35.70	10.28	31.71
First-3	40.21	12.13	37.13
ORACLE	52.32	22.23	47.93
LexRank* (Erkan and Radev, 2004)	41.77	13.81	37.87
TextRank* (Mihalcea and Tarau, 2004)	41.95	13.86	38.07
MMR* (Carbonell and Goldstein, 1998)	44.72	14.92	40.77
PG† (Lebanoff et al., 2018)	44.55	15.54	40.75
BottomUp† (Gehrmann et al., 2018)	45.27	15.32	41.38
Hi-MAP† (Fabbri et al., 2019)	45.21	16.29	41.39
HSG	45.66	16.22	41.80
HSG + Tri-Blocking	44.92	15.59	40.89
HDSG	**46.05**	**16.35**	**42.08**
HDSG + Tri-Blocking	45.55	15.78	41.29

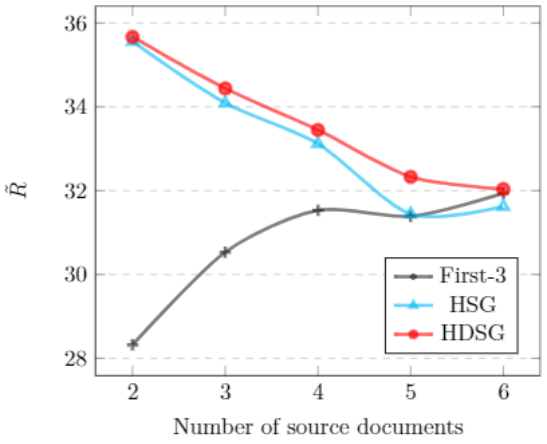
Table 4: Results on the test set of Multi-News. We reproduce models with '*' via the released code and directly use the outputs of † provided by Fabbri et al. (2019) for evaluation.

Ablation and Analysis

Model	R-1	R-2	R-L
HSG	42.31	19.51	38.74
- filter words	42.24	19.56	38.68
- edge feature	42.14	19.41	38.60
- residual connection	41.59	19.08	38.05
- sentence update	41.59	19.03	38.04
- word update	41.70	19.16	38.15
- BiLSTM	41.70	19.09	38.13

Table 3: Ablation studies on CNN/DailyMail test set. We remove various modules and explore their influence on our model. '-' means we remove the module from the original HETERSUMGRAPH. Note that HETERSUMGRAPH without the updating of sentence nodes is actually the Ext-BiLSTM model described in Section 4.3.





Conclusion

Why HeterSumGraph?

- Graph for summarization
  - model the non-local relationship
  - typic structure for ranking problem
- Heterogeneous nodes
  - different semantic units (words, entities, etc.)
  - enrich cross-sentence relationships (*sentence-word-sentence*)
  - easily adapt from single-document to multi-document (*document nodes*)
- Update mechanism
  - iterative process

Thanks for your listening!

Q & A

https://github.com/brxx122/HeterSumGraph