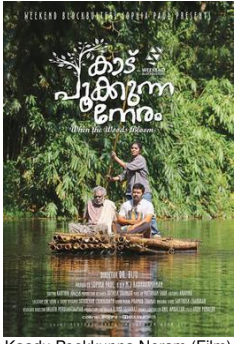
Generating Summaries with Topic Templates and Structured Convolutional Decoders

Laura Perez-Beltrachini Yang Liu Mirella Lapata

School of Informatics
University of Edinburgh
{lperez,mlap}@inf.ed.ac.uk

Entity Summaries



Kaadu Pookkunna Neram (Film)

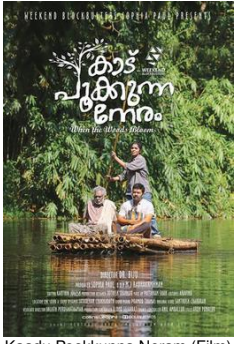
References [edit]

1. ^ "The law of the jungle". 25 August 2016 – via The Hindu.
2. ^ http://english.manoramaonline.com/entertainment/movie-reviews/kaadu-pookkunna-neram.html
3. ^ James, Anu. " 'Kaadu Pookkunna Neram' to be screened as only Indian movie at Eurasia Film Festival".

Kaadu Pookkunna Neram (2016) - IMDb
https://www.imdb.com/title/tt5461134/
★★★★☆ Rating: 7.9/10 - 38 votes
Kaadu Pookkunna Neram Poster. A police group is deployed by the government to do away with Maoist menace (A banned Left organization) in a tribal village ...

KAADU POOKKUNNA NERAM | British Board of Film Classification
https://bbfc.co.uk/releases/kaadu-pookkunna-neram-2016
14 Nov 2016 - Summary **KAADU POOKKUNNA NERAM** is a Malayalam language drama in which a police officer, tasked with hunting down Maoist rebels, ...

Entity Summaries



Kaadu Pookkunna Neram (Film)

Kaadu Pookkunna Neram (English :When the Woods Bloom) is a 2016 Indian Malayalam film written and directed by Dr. Biju. The film is produced by Sophia Paul under the film production house, Weekend Blockbusters. It stars Indrajith Sukumaran as a policeman, who is sent into deep jungles to capture the chief of a radical organization, played by Rima Kallingal, but instead finds himself trapped in it. Kaadu Pookkunna Neram premiered at the Montreal World Film Festival in September 2016.

References [edit]

1. ^ "The law of the jungle". 25 August 2016 – via The Hindu.
2. ^ http://english.manoramaonline.com/entertainment/movie-reviews/kaadu-pookkunna-neram.html
3. ^ James, Anu. " 'Kaadu Pookkunna Neram' to be screened as only Indian movie at Eurasia Film Festival".

Kaadu Pookkunna Neram (2016) - IMDb
https://www.imdb.com/title/tt5461134/
★★★★☆ Rating: 7.9/10 - 38 votes
Kaadu Pookkunna Neram Poster. A police group is deployed by the government to do away with Maoist menace (A banned Left organization) in a tribal village ...

KAADU POOKKUNNA NERAM | British Board of Film Classification
https://bbfc.co.uk/releases/kaadu-pookkunna-neram-2016
14 Nov 2016 - Summary **KAADU POOKKUNNA NERAM** is a Malayalam language drama in which a police officer, tasked with hunting down Maoist rebels, ...

Content Structure of Entity Summaries

- ✓ cover **different topics**
- ✓ **order** discussed topics
- ✓ **domain specific** topics

write and direct

production

premiere

Kaadu Pookkunna Neram (English :When the Woods Bloom) is a 2016 Indian Malayalam film written and directed by Dr Biju. The film is produced by Sophia Paul under the film production house, Weekend Blockbusters. It stars Indrajith Sukumaran as a policeman, who is sent into deep jungles to capture the chief of a radical organization, played by Rima Kallingal, but instead finds himself trapped in it. Kaadu Pookkunna Neram premiered at the Montreal World Film Festival in September 2016.

Wray (2002); Barzilay and Lee (2004)

Decoding with Summary Structure?

- Neural text generation approaches
- Multi-sentence text as a **single long sequence** with a **sequence** decoder
  e.g. Gardent et al. (2017); See et al. (2017); Wiseman et al. (2017); Celikyilmaz et al. (2018); Perez-Beltrachini and Lapata (2018); Marcheggiani and Perez-Beltrachini (2018); Liu et al. (2018); Puduppully et al. (2019); Fabbri et al. (2019)
- Output texts are too **general** or **brief**; or simply **incorrect**

Decoding with Summary Structure?

- Neural text generation approaches
- Multi-sentence text as a **single long sequence** with a **sequence** decoder
  e.g. Gardent et al. (2017); See et al. (2017); Wiseman et al. (2017); Celikyilmaz et al. (2018); Perez-Beltrachini and Lapata (2018); Marcheggiani and Perez-Beltrachini (2018); Liu et al. (2018); Puduppully et al. (2019); Fabbri et al. (2019)
- Output texts are too **general** or **brief**; or simply **incorrect**
- **A Decoder with Explicit Content Structure: Sentence Topics and Order**

Decoding with Summary Structure?

- Neural text generation approaches
- Multi-sentence text as a **single long sequence** with a **sequence** decoder
  e.g. Gardent et al. (2017); See et al. (2017); Wiseman et al. (2017); Celikyilmaz et al. (2018); Perez-Beltrachini and Lapata (2018); Marcheggiani and Perez-Beltrachini (2018); Liu et al. (2018); Puduppully et al. (2019); Fabbri et al. (2019)
- Output texts are too **general** or **brief**; or simply **incorrect**
- **A Decoder with Explicit Content Structure: Sentence Topics and Order**
  1. **Two Level Decoder**
  2. **Sentence Topic Templates**

Sentence Topic Templates

- LDA treating sentences as documents
- Assignment of few topics to sentences

Topics in the Film domain:

#10: base, film, name, novel, story, screenplay
#14: win, film, music, award, nominate, compose
#18: film, receive, review, office, box, critic
#19: star, film, role, play, lead, support

Sentence Topic Templates

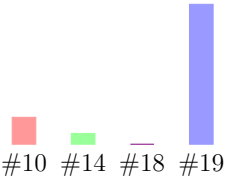
- LDA treating sentences as documents
- Assignment of few topics to sentences

Topics in the Film domain:

#10: base, film, name, novel, story, screenplay
#14: win, film, music, award, nominate, compose
#18: film, receive, review, office, box, critic
#19: star, film, role, play, lead, support

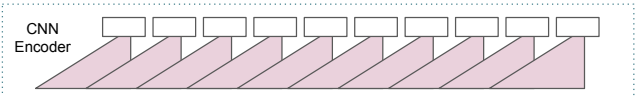
Sentences annotation:

It stars Indrajith Sukumaran as a policeman [#19]



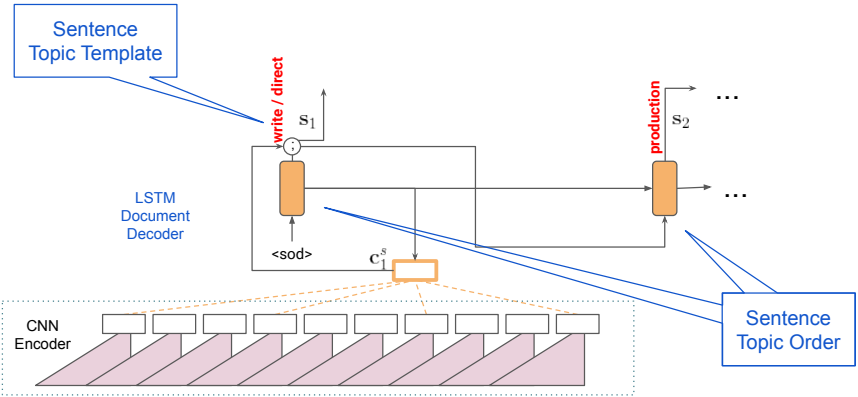
Latent Dirichlet Allocation (LDA: Blei et al. (2003))

Decoder with Content Guidance



KAADU POOKKUNNA NERAM is a Malayalam language drama in which a police ...
Kaadu Pookkunna Neram Poster. A police group is deployed by the government to do a ...

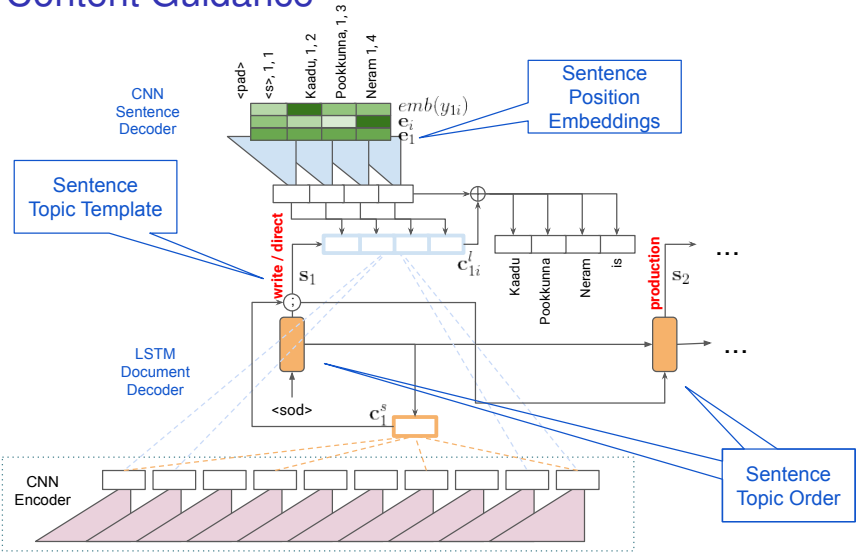
Decoder with Content Guidance



KAADU POOKKUNNA NERAM is a Malayalam language drama in which a police ...

Kaadu Pookkunna Neram Poster. A police group is deployed by the government to do a ...

Decoder with Content Guidance

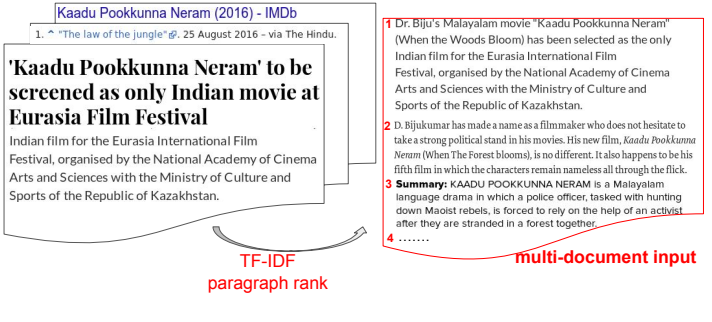


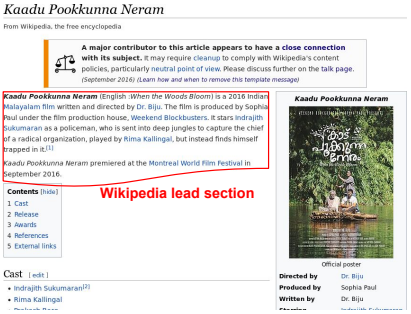
KAADU POOKKUNNA NERAM is a Malayalam language drama in which a police ...

Kaadu Pookkunna Neram Poster. A police group is deployed by the government to do a ...

WikiSum Dataset Liu et al. (2018)

(multi-document input, Wikipedia lead section) pairs

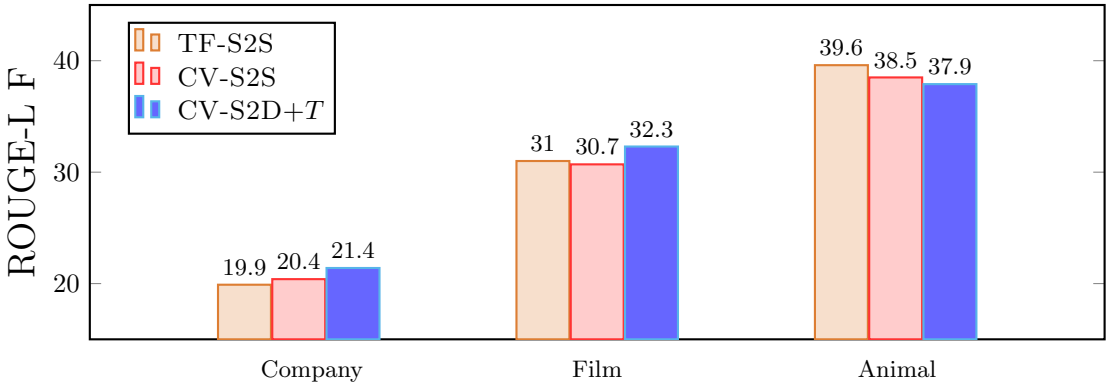




WikiCatSum= Three domains from WikiSum, 800 input tokens

Company (62k), **Film** (60k) and **Animal** (61k)

Improvements on Content Coverage (automatic)



TF-S2S: Transformer Seq2Seq (7 layers decoder)
CV-S2S: Convolutional Seq2Seq (3 layers decoder)
Ours **CV-S2D**+T: Convolutional Seq2Doc +Topics (1 layer document + 3 layers sentence decoder)

Manual Judgements on Summary Quality

	Company		Film		Animal	
	QA	Rank	QA	Rank	QA	Rank
TF-S2S	5	**1.87**	6	2.27	9	**1.87**
CV-S2S	5	2.27	6.67	**1.76**	8.33	2.04
Ours CV-S2D+*T*	**7**	**1.87**	**7**	1.98	**9.33**	2.09

3 judges, 45 summaries

QA: average of correctly answered questions (40 questions per domain)

Rank: average rank (lower is better) (content, fluency and repetition?)

Example Output

(Reference) Compacta capitalis is a moth in the family Crambidae. It was described by Augustus Radcliffe Grote in 1881. It is found in North America, where it has been recorded from Maryland to Florida, west to Texas and possibly Colorado, north to Illinois. The wingspan is about 35 mm. The forewings are white with a reddish-brown shading at the base and along the inner margin and two black discal spots, as well as an irregular subterminal line. There is a dark apical blotch on both wings. Adults are on wing from May to August.

Example Output

(Reference) Compacta capitalis is a moth in the family Crambidae. It was described by Augustus Radcliffe Grote in 1881. It is found in North America, where it has been recorded from Maryland to Florida, west to Texas and possibly Colorado, north to Illinois. The wingspan is about 35 mm. The forewings are white with a reddish-brown shading at the base and along the inner margin and two black discal spots, as well as an irregular subterminal line. There is a dark apical blotch on both wings. Adults are on wing from May to August.

(CV-S2S) Compacta UNK is a moth in the Crambidae family. It was described by Barnes and McDunnough in 1918. It is found in North America, where it has been recorded from Alabama, Florida, Georgia, Illinois, Indiana, Kentucky, Maine, Maryland, Massachusetts, Minnesota, New Brunswick, New Hampshire, New Jersey, New york, North Carolina, Ohio, Oklahoma, Ontario, Pennsylvania, Quebec, South Carolina, Tennessee, Texas and Virginia.

Example Output

(Reference) Compacta capitalis is a moth in the family Crambidae. It was described by Augustus Radcliffe Grote in 1881. It is found in North America, where it has been recorded from Maryland to Florida, west to Texas and possibly Colorado, north to Illinois. The wingspan is about 35 mm. The forewings are white with a reddish-brown shading at the base and along the inner margin and two black discal spots, as well as an irregular subterminal line. There is a dark apical blotch on both wings. Adults are on wing from May to August.

(CV-S2S) Compacta UNK is a moth in the Crambidae family. It was described by Barnes and McDunnough in 1918. It is found in North America, where it has been recorded from Alabama, Florida, Georgia, Illinois, Indiana, Kentucky, Maine, Maryland, Massachusetts, Minnesota, New Brunswick, New Hampshire, New Jersey, New york, North Carolina, Ohio, Oklahoma, Ontario, Pennsylvania, Quebec, South Carolina, Tennessee, Texas and Virginia.

(Ours) Compacta UNK is a moth in the Crambidae family. It was described by Grote in 1878. It is found in North America, where it has been recorded from Florida. It is also found in Mexico. The wingspan is about 20 mm. Adults have been recorded on wing from April to September.

Conclusions

- A novel structured decoder for multi-document summarisation
  aware of sentence topics and sentence order
- Structured decoding improves on content coverage
  @ACL Holtzman et al. (2019); Ippolito et al. (2019); Shen et al. (2019)

Code (PyTorch) @ `https://github.com/lauhaide/WikiCatSum`
WikiCatSum @ `https://datashare.is.ed.ac.uk/handle/10283/3368`

References I

Regina Barzilay and Lillian Lee. 2004. Catching the drift: Probabilistic content models, with applications to generation and summarization. *arXiv preprint cs/0405039*.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Asli Celikyilmaz, Antoine Bosselut, Xiaodong He, and Yejin Choi. 2018. Deep communicating agents for abstractive summarization. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1662–1675. Association for Computational Linguistics.

Alexander R Fabbri, Irene Li, Tianwei She, Suyi Li, and Dragomir R Radev. 2019. Multi-news: a large-scale multi-document summarization dataset and abstractive hierarchical model. *arXiv preprint arXiv:1906.01749*.

Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. The webnlg challenge: Generating text from rdf data. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 124–133, Santiago de Compostela, Spain. (INLG 2017).

Ari Holtzman, Jan Buys, Maxwell Forbes, and Yejin Choi. 2019. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.

Daphne Ippolito, Reno Kriz, Maria Kustikova, João Sedoc, and Chris Callison-Burch. 2019. Comparison of diverse decoding methods from conditional language models. *arXiv preprint arXiv:1906.06362*.

Peter Liu, Mohammad Saleh, Etienne Pot, Ben Goodrich, Ryan Sepassi, Lukasz Kaiser, and Noam Shazeer. 2018. Generating Wikipedia by summarizing long sequences. In *Proceedings of the 6th International Conference on Learning Representations*, Vancouver, Canada.

Diego Marcheggiani and Laura Perez-Beltrachini. 2018. Deep Graph Convolutional Encoders for Structured Data to Text Generation. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 1–9, Tilburg University, The Netherlands. Association for Computational Linguistics.

Laura Perez-Beltrachini and Mirella Lapata. 2018. Bootstrapping Generators from Noisy Data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1516–1527, New Orleans, Louisiana.

Ratish Puduppully, Li Dong, and Mirella Lapata. 2019. Data-to-Text Generation with Content Selection and Planning. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, Honolulu, Hawaii.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada.

Dinghan Shen, Asli Celikyilmaz, Yizhe Zhang, Liqun Chen, Xin Wang, Jianfeng Gao, and Lawrence Carin. 2019. Towards generating long and coherent text with multi-level latent variable models. *arXiv preprint arXiv:1902.00154*.

Sam Wiseman, Stuart Shieber, and Alexander Rush. 2017. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2253–2263, Copenhagen, Denmark.

Alison Wray. 2002. *Formulaic Language and the Lexicon*. Cambridge University Press, Cambridge.

WikiCatSum Dataset

- WikiSum Liu et al. (2018)
- Three domains (DBPedia **Cat**egories) derived from WikiSum, 800 input tokens

	Size	R1	R2	RL	SentNb	SentLen	#Topics
Company	62k	.551	.217	.438	5.09±3.73	24.40±13.47	40
Film	60k	.559	.243	.456	4.17±2.71	23.54±11.91	20
Animal	61k	.541	.208	.455	4.71±3.53	19.68±18.69	30

* R1, R2 and RL ROUGE recall

Content Coverage According to Abstract & Copy metrics

Model	Company		Film		Animal	
	A	C	A	C	A	C
CV-S2S	.046	.307	.097	.430	**.229**	**.515**
Ours **CV-S2D**+T	**.051**	**.316**	**.101**	**.433**	.223	.506