Generating Summaries with Topic Templates and Structured Convolutional Decoders

Laura Perez-Beltrachini    Yang Liu    Mirella Lapata

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

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: 7.0/10 - 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, …
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 ]

3. James. Anu. "Kaadu Pookkunna Neram' to be screened as only Indian movie at Eurasia Film Festival".
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]
KAADU POOKKUNNA NERAM is a Malayalam language drama in which a police group is deployed by the government to do a ...
**KAADU POOKKUNNA NERAM** is a Malayalam language drama in which a police group is deployed by the government to do a ...
KAADU POOKKUNNA NERAM is a Malayalam language drama in which a police group is deployed by the government to do a ...
WikiSum Dataset Liu et al. (2018)

(multi-document input, Wikipedia lead section) pairs

Kaadu Pookkunna Neram (2016) - IMDb


'Kaadu Pookkunna Neram' to be screened as only Indian movie at Eurasia Film Festival

Indian film for the Eurasia International Film Festival, organised by the National Academy of Cinema Arts and Sciences with the Ministry of Culture and Sports of the Republic of Kazakhstan.

Dr. Biju’s Malayalam movie "Kaadu Pookkunna Neram" (When the Woods Bloom) has been selected as the only Indian film for the Eurasia International Film Festival, organised by the National Academy of Cinema Arts and Sciences with the Ministry of Culture and Sports of the Republic of Kazakhstan.

D. Bijukumar has made a name as a filmmaker who does not hesitate to take a strong political stand in his movies. His new film, Kaadu Pookkunna Neram (When The Forest blooms), is no different. It also happens to be his fifth film in which the characters remain nameless all through the flick.

Summary: KAADU POOKKUNNA NERAM is a Malayalam language drama in which a police officer, tasked with hunting down Maoist rebels, is forced to rely on the help of an activist after they are stranded in a forest together.

1. Dr. Biju’s Malayalam movie "Kaadu Pookkunna Neram" (When the Woods Bloom) has been selected as the only Indian film for the Eurasia International Film Festival, organised by the National Academy of Cinema Arts and Sciences with the Ministry of Culture and Sports of the Republic of Kazakhstan.

2. D. Bijukumar has made a name as a filmmaker who does not hesitate to take a strong political stand in his movies. His new film, Kaadu Pookkunna Neram (When The Forest blooms), is no different. It also happens to be his fifth film in which the characters remain nameless all through the flick.

3. Summary: KAADU POOKKUNNA NERAM is a Malayalam language drama in which a police officer, tasked with hunting down Maoist rebels, is forced to rely on the help of an activist after they are stranded in a forest together.

4. ……

WikiCatSum = Three domains from WikiSum, 800 input tokens

Company (62k), Film (60k) and Animal (61k)
**Improvements on Content Coverage (automatic)**

<table>
<thead>
<tr>
<th></th>
<th>Company</th>
<th>Film</th>
<th>Animal</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-S2S</td>
<td>19.9</td>
<td>31</td>
<td>39.6</td>
</tr>
<tr>
<td>CV-S2S</td>
<td>20.4</td>
<td>30.7</td>
<td>38.5</td>
</tr>
<tr>
<td>CV-S2D+T</td>
<td>21.4</td>
<td>32.3</td>
<td>37.9</td>
</tr>
</tbody>
</table>

**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

<table>
<thead>
<tr>
<th></th>
<th>Company</th>
<th>Film</th>
<th>Animal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>QA Rank</td>
<td>QA Rank</td>
<td>QA Rank</td>
</tr>
<tr>
<td>TF-S2S</td>
<td>5</td>
<td>1.87</td>
<td>6</td>
</tr>
<tr>
<td>CV-S2S</td>
<td>5</td>
<td>2.27</td>
<td>6.67</td>
</tr>
<tr>
<td><strong>Ours CV-S2D+T</strong></td>
<td>7</td>
<td>1.87</td>
<td>7</td>
</tr>
</tbody>
</table>

3 judges, 45 summaries

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

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

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.
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.

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.

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


WikiCatSum Dataset

WikiSum Liu et al. (2018)

Three domains (DBPedia Categories) derived from WikiSum, 800 input tokens

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

* R1, R2 and RL ROUGE recall
### Content Coverage According to Abstract & Copy metrics

<table>
<thead>
<tr>
<th>Model</th>
<th>Company</th>
<th>Film</th>
<th>Animal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>CV-S2S</td>
<td>.046</td>
<td>.307</td>
<td>.097</td>
</tr>
<tr>
<td>Ours CV-S2D+T</td>
<td>.051</td>
<td>.316</td>
<td>.101</td>
</tr>
</tbody>
</table>