Focused Attention Improves Document-Grounded Generation

Shrimai Prabhumoye, Kazuma Hashimoto, Yingbo Zhou, Alan W Black, Ruslan Salakhutdinov

Carnegie Mellon University
Language Technologies Institute
Document Grounded Generation
Document Grounded Generation

Context
Document Grounded Generation
Document Grounded Generation

Context + Document = Updated Text
Document Grounded Generation

- generate text given a context and a document
- generated text should *coherently fit* the context and contain new information from the document
Wikipedia Update Generation

(Prabhumoye et al, NAACL 2019)

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation
Wikipedia Update Generation

(Prabhumoye et al, NAACL 2019)
Wikipedia Update Generation

Monkey selfie copyright dispute
From Wikipedia, the free encyclopedia

The monkey selfie copyright dispute is a series of disputes about the copyright status of selfies taken by Celebes crested macaques using equipment belonging to the British nature photographer David Slater. The disputes involve Wikimedia Commons and the blog Techdirt, which have hosted the images following their publication in newspapers in July 2011 over Slater’s objections that he holds the copyright, and People for the Ethical Treatment of Animals (PETA), who have argued that the macaque should be assigned the copyright.

Slater has argued that he has a valid copyright claim, as he engineered the situation that resulted in the pictures by travelling to Indonesia, befriending a group of wild macaques, and setting up his camera equipment in such a way that a “selfie” picture might come about. The Wikimedia Foundation’s 2014 refusal to remove the pictures from its Wikimedia Commons image library was based on the understanding that copyright is held by the creator, that a non-human creator (not being a legal person) cannot hold copyright, and that the images are thus in the public domain.

Ape-true priority photographer plays down monkey reports

Chris Cheesman
July 9, 2011

A photographer who says he witnessed monkeys taking pictures of themselves, tells Amateur Photographer (AP) that much of the media coverage has been exaggerated.

Wildlife photographer David Slater today played down newspaper reports that suggest a bunch of Indonesian monkeys grabbed his camera and began taking self-portraits.

And he revealed that the shots, featuring the endangered crested black macaque monkey, were taken three years ago.

Speaking to AP, David explained that his camera had been mounted on a tripod when the animals

(Prabhumoye et al, NAACL 2019)
The following day, *Amateur Photographer* reported that Slater gave them further explanation as to how the photographs were created, downplaying the way newspaper articles had described them; Slater said reports that a monkey ran off with his camera and "began taking self-portraits" were incorrect and that the portrait was shot when his camera had been mounted on a tripod, with the primates playing around with a remote cable release as he fended off other monkeys.[13]
Wikipedia Update Generation

Monkey selfie copyright dispute

The monkey selfie copyright dispute is a series of disputes about the copyright status of selfies taken by Celebes crested macaques using equipment belonging to the British nature photographer David Slater. The disputes involve Wikimedia Commons the blog Technorati which has hosted the images following their publication in newspapers in July 2011 over Slater's objections that he holds the copyright, and People for the Ethical Treatment of Animals (PETA), who have argued that the macaque should be assigned the copyright.

Slater has argued that he has a valid copyright claim, as he engineered the situation that resulted in the pictures by travelling to Indonesia, befriending a group of wild macaques, and setting up his camera equipment in such a way that a "selfie" picture might come about. The Wikimedia Foundation's 2014 refusal to remove the pictures from its Wikimedia Commons image library was based on the understanding that copyright is held by the creator, a non-human creator (not being a legal person) cannot hold copyright, and that the images are thus in the public domain.

The following day, Amateur Photographer reported that Slater gave them further explanation as to how the photographs were created, downplaying the way newspaper articles had described them; Slater said reports that a monkey ran off with his camera and "began taking self-portraits" were incorrect and that the portrait was shot when his camera had been mounted on a tripod, with the primates playing around with a remote cable release as he fended off other monkeys.[13]

Dataset Size of 636k!

(Prabhumoye et al, NAACL 2019)
Document Grounded Dialogue

(Zhou, Prabhumoye & Black, EMNLP 2018)

https://github.com/festvox/datasets-CMU_DoG
Document Grounded Dialogue

Chat History

Alice: Yes this is a Christmas classic …
Bob: Very well said! Some of the critics were …
Alice: Yes I found the first half very entertaining.
Bob: I feel you! So hard to believe it was released in 1990!
Alice: Yes this is a Christmas classic …
Bob: Very well said! Some of the critics were …
Alice: Yes I found the first half very entertaining.
Bob: I feel you! So hard to believe it was released in 1990!
Alice: Yes this is a Christmas classic …
Bob: Very well said! Some of the critics were …
Alice: Yes I found the first half very entertaining.
Bob: I feel you! So hard to believe it was released in 1990!

Alice: I like the quote "sadistic festival of adult-bashing". Between's Kevin's parents forgetting him and the burglars this movie doesn't make adult looks too bright.
Alice: Yes this is a Christmas classic …
Bob: Very well said! Some of the critics were …
Alice: Yes I found the first half very entertaining.
Bob: I feel you! So hard to believe it was released in 1990!

Alice: I like the quote "sadistic festival of adult-bashing". Between's Kevin's parents forgetting him and the burglars this movie doesn't make adult looks too bright.
Models

- Pre-trained encoder-decoder model
  - provide a stronger baseline for the task
Meanwhile, newspapers across the country pick up the story in solidarity with the *post and times*. The court rules 6-3 in the newspapers' favor, vindicating Graham's decision. *Nixon demands that the post should be barred from the white house.*

One year later, a security guard discovers a break-in progress at Watergate complex.

It was honorable for the *other papers to publish in solidarity with the post and times*. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the Post and Times. The court rules 6-3 in the newspapers’ favor, vindicating Graham’s decision. Nixon demands that the Post should be barred from the White House.

One year later, a security guard discovers a break-in progress at Watergate complex.

It was honorable for the other papers to publish in solidarity with the Post and Times. That had to prove that the people want to know.
Models

- Pre-trained encoder-decoder model
  - provide a stronger baseline for the task

- Context Driven Representation (CoDR)
  - build a representation of the content in document that is *not present* in the context

- Document Headed Attention (DoHA)
  - specific *focus on document* information

https://github.com/shrimai/Grounded_Generation
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers’ favor, vindicating graham’s decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house. One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6-3 in the newspapers' favor, vindicating Graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at Watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Yes, but the newspapers won!

Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6-3 in the newspapers' favor, vindicating Graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at Watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the Post and Times. The court rules 6-3 in the newspapers' favor, vindicating Graham's decision. Nixon demands that the Post should be barred from the White House.

One year later, a security guard discovers a break-in progress at Watergate complex.

It was honorable for the other papers to publish in solidarity with the Post and Times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house. One year later, a security guard discovers a break-in progress at watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Meanwhile, newspapers across the country pick up the story in solidarity with the *post and times*. The court rules 6–3 in the newspapers’ favor, vindicating Graham’s decision. *Nixon demands that the post should be barred from the white house.*

One year later, a security guard discovers a break-in progress at Watergate complex.

It was honorable for the *other papers to publish in solidarity with the post and times*. That had to prove that the people want to know.
Document Headed Attention (DoHA)

DOC

Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.

CrossAttention_Doc(Q, K, V) = [H₁, ..., Hₘ] \( w^{do} \)

\[ H_i = \text{Attention}(Q \cdot W_{iQ}, K \cdot W_{iK}, V \cdot W_{iV}) \]
Meanwhile, newspapers across the country pick up the story in solidarity with the post and times. The court rules 6-3 in the newspapers' favor, vindicating Graham's decision. Nixon demands that the post should be barred from the white house.

One year later, a security guard discovers a break-in progress at Watergate complex.

It was honorable for the other papers to publish in solidarity with the post and times. That had to prove that the people want to know.
Wikipedia Update Generation

- Content Transfer (Prabhumoye et al, NAACL 2019)
- BART
- CoDR
- DoHA

<table>
<thead>
<tr>
<th></th>
<th>BLEU-4</th>
<th>METEOR</th>
<th>Rouge-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>11.31</td>
<td>14.38</td>
<td>11.31</td>
</tr>
<tr>
<td>Transfer</td>
<td>9.2</td>
<td>12.9</td>
<td>10.08</td>
</tr>
<tr>
<td>BART</td>
<td>10.86</td>
<td>14.28</td>
<td>22.39</td>
</tr>
<tr>
<td>CoDR</td>
<td>11.23</td>
<td>14.38</td>
<td>23.48</td>
</tr>
<tr>
<td>DoHA</td>
<td>6.21</td>
<td>10.08</td>
<td>23.49</td>
</tr>
</tbody>
</table>
Wikipedia Update Generation

- CoDR and DoHA outperform BART and Content Transfer
• 19.7 BLEU-4 improvement over Low-Res
### CMU_DoG

<table>
<thead>
<tr>
<th></th>
<th>BLEU-4</th>
<th>METEOR</th>
<th>Rouge-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>BART</td>
<td>20.9</td>
<td></td>
<td>20.41</td>
</tr>
<tr>
<td>CoDR</td>
<td>16.91</td>
<td>14.55</td>
<td>20.44</td>
</tr>
<tr>
<td>DoHA</td>
<td>12.59</td>
<td></td>
<td>20.68</td>
</tr>
</tbody>
</table>

- 19.7 BLEU-4 improvement over Low-Res
- 24% BLEU-4 improvement over BART
Relevance Human Evaluation

Pick the option which contains information from the document and fits the context coherently.
Pick the option which contains information from the document and fits the context coherently.

- No Preference is chosen most number of times.
Pick the option which contains information from the document and fits the context coherently.

- No Preference is chosen most number of times.
- Both CoDR and DoHA outperform BART.
Wikipedia Update Generation
## Wikipedia Update Generation

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic Variation:</strong></td>
<td></td>
<td>Generation is appropriate but a linguistic variation of the reference.</td>
<td>December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.</td>
<td>41</td>
</tr>
</tbody>
</table>

Generation is a paraphrase of the Reference
Still gets a Rouge-L score of 41!
Inadequate automated metrics
<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial Hallucination: Generation is either missing or hallucinates some information</td>
<td>23</td>
<td>America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.</td>
<td>The World Wide Web was first introduced on January 17, 1995 on Prodigy.</td>
<td>17</td>
</tr>
</tbody>
</table>

The generation has captured some information of the reference but missed some.
Wikipedia pages that are in the form of lists
1340s, *Timeline of DC Comics (1950s)*
50% times the generation is grounded in the document and close to reference

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Incoherent Reference</em>: The reference does not coherently follow the context</td>
<td>22</td>
<td>“The Naked Ape” by Desmond Morris, is published.</td>
<td>Zoologist Desmond Morris publishes “The Naked Ape”.</td>
<td>26</td>
</tr>
</tbody>
</table>
### Wikipedia Update Generation

#### About 85% times, the generation is either completely or partially grounded if the reference is grounded.

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Incorrect</strong>: The generation is either not appropriate or is not grounded (completely hallucinates the information).</td>
<td>7</td>
<td>The year 2000 is sometimes abbreviated as “Y2K” (the “Y” stands for “year”, and the “K” stands for “kilo-” which means “thousand”)</td>
<td>The Y2K conspiracy theory claimed that a secret nuclear attack by the United States on 2 January 2000 was planned to begin World War 2.</td>
<td>9</td>
</tr>
</tbody>
</table>
## Wikipedia Update Generation

<table>
<thead>
<tr>
<th>Error Class</th>
<th>%</th>
<th>Reference</th>
<th>Generation</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference is not grounded</td>
<td>5</td>
<td>This was achieved under dead calm conditions as an additional safety measure, whereas the Wrights flew in a 25 mph+ wind to achieve enough airspeed on their early attempts.</td>
<td>This was verified by a video crew present at the test flight.</td>
<td>14</td>
</tr>
</tbody>
</table>
Reference vs CoDR/DoHA

Pick the option that is most appropriate to the given context.

Wikipedia Update Generation

- Reference: 33.9
- No Pref: 28.3
- CoDR/DoHA: 37.8

CMU_DoG

- Reference: 22.8
- No Pref: 45.6
- CoDR/DoHA: 31.6
Reference vs CoDR/DoHA

Pick the option that is most appropriate to the given context.

1. Automated metrics are inadequate
2. Sole reference should not be considered as the only correct response to the context
Summary

- New baseline stronger than previous work
- Two new proposed techniques
  - Context Driven Representation
  - Document Headed Attention
- Substantial improvements on automated and human evaluation results on three different datasets
- Comprehensive manual analysis
- Code and trained models are available at https://github.com/shrimai/Focused-Attention-Improves-Document-Grounded-Generation