Towards Content Transfer through Grounded Text Generation

Shrimai Prabhumoye, Chris Quirk, Michel Galley
Motivation

- We constantly author text
- AI assistance deals with *form* (grammar, style, etc.)
- Our goal is to control for *content*

After *graduate form* Columbia University, Obama worked in Chicago.

After graduating from *Carnegie Mellon University*, Obama worked in Chicago.
On 4 July 2011 several publications including the Daily Mail, The Telegraph, and The Guardian picked up the story and published the pictures along with articles that quoted Slater as describing the photographs as self-portraits taken by the monkeys: "Monkey steals camera to snap himself" (The Telegraph), "a camera on a tripod" triggered by the monkeys (The Guardian), and a camera started by a monkey "Fascinated by her reflection in the lens". The articles also contained Slater quotes such as "He must have taken hundreds of pictures by the time I got my camera back." 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.
Primary Contribution

- design a task to perform content transfer from an unstructured source of information
- release dataset
Applications

- Maintain software documentation given incoming streams of text (email, software requirements etc)
- Legal precedent around a topic
- Inbox Summarization
- Updating Wikipedia articles
Overview

Dataset

Models

Evaluation
Data Creation Process

Context

Update

HTML News Article

Plain Text of News Article
### Data Creation Process

**Total Data Size:** 636K

<table>
<thead>
<tr>
<th>News Article</th>
<th>Wikipedia Context</th>
<th>Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

---

**Plain Text of News Article**
Overview

Dataset

Models

Evaluation
Models

- Generative Models
  - Context Agnostic Generative Model (CAG) — Baseline
  - Context Informed Generative Model (CIG)
  - Context Responsive Generative Model (CRG)
  - all models have global attention

- Extractive Models
  - SumBasic
  - Context Informed SumBasic
  - Oracle

- All models are simplistic to infer if context helps in generation
Context Agnostic Model (CAG) - Baseline

News Article

$\mathbf{x}_1 \rightarrow \mathbf{x}_2 \rightarrow \mathbf{x}_3$ (Encoder Vector)

$\mathbf{y}_1 \rightarrow \mathbf{y}_2 \rightarrow \mathbf{y}_3$ (Update)

<start>
Context Informed Model (CIG)
Context Responsive Model (CRG)
Extractive Models

- SumBasic: a model based on unigram probabilities
- Context Informed SumBasic: the unigram probabilities take into account the words in the Wikipedia context.
- Oracle: establish an upper limit attainable by extractive methods
Overview

Dataset

Models

Evaluation
Automated Evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>SumBasic</td>
<td>5.6 (5.6-5.7)</td>
</tr>
<tr>
<td>Context Informed SumBasic</td>
<td>7.0 (7.0-7.1)</td>
</tr>
<tr>
<td>Context Agnostic Generative Model</td>
<td>9.1 (9.0-9.2)</td>
</tr>
<tr>
<td>Context Informed Generative Model</td>
<td><strong>16.0 (15.9-16.1)</strong></td>
</tr>
<tr>
<td>Context Responsive Generative Model</td>
<td>14.7 (14.6-14.8)</td>
</tr>
<tr>
<td>Oracle</td>
<td>28.8 (28.7-29.0)</td>
</tr>
</tbody>
</table>

* METEOR and BLEU numbers are consistent with ROUGE-L
Relative Human Evaluation

Which system output is closest in meaning to the reference update?

Which system output is more accurate relative to the background information given in the snippet of the article?
Absolute Quality Evaluation

![Bar chart showing quality evaluation of different aspects such as Grammar, Non-redundancy, Referential Quality, Focus, and Structure and Coherence. The x-axis represents the scale from (very poor) to (very good), and the y-axis represents different aspects. The chart compares CAG and CIG.]
### Examples

<table>
<thead>
<tr>
<th>Reference Update</th>
<th>CIG Generated Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>rob brydon, the comedian <strong>was born in baglan.</strong></td>
<td>he <strong>was born in baglan</strong></td>
</tr>
<tr>
<td>in may 2014 he <strong>was diagnosed with prostate cancer.</strong></td>
<td>st. clair <strong>was diagnosed with prostate cancer.</strong></td>
</tr>
<tr>
<td>on april 3, 2014, manning signed a one-year deal <strong>with the cincinnati bengals.</strong></td>
<td>on march 9, 2014, manning signed a one-year contract <strong>with the cincinnati bengals.</strong></td>
</tr>
</tbody>
</table>
We have designed a new task for content transfer through grounded text generation. Both automatic and human evaluation of our baseline models shows that context is important for the generation process — treating our task as summarization is not enough! The total size of the dataset is 636k. Code and data can be found at https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation. Code available for all the models. Raw data and the train data used in experiments.
Thank You