Controlling length in Captioning

Ruotian Luo

Greg Shakhnarovich
Why controlling length

A man and a girl sit on the ground and eat.
A man and a little girl are sitting on a sidewalk near a blue bag eating.
A man wearing a black shirt and a little girl wearing an orange dress share a treat.
...

Existing work

• Show Control Tell

Cornia, Marcella, Lorenzo Baraldi, and Rita Cucchiara. "Show, control and tell: a framework for generating controllable and grounded captions."
Existing work

• POS tags

Existing work

• Controlling length in summarization

<table>
<thead>
<tr>
<th>source</th>
<th>five-time world champion michelle kwan withdrew from the us figure skating championships on wednesday, but will petition us skating officials for the chance to compete at the turin olympics.</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>injury leaves kwan ’s olympic hopes in limbo</td>
</tr>
<tr>
<td>fixLen</td>
<td>kwan withdraws from us gp</td>
</tr>
<tr>
<td>(30)</td>
<td>kwan withdraws from us skating championships</td>
</tr>
<tr>
<td>(50)</td>
<td>kwan pulls out of us figure skating championships for turin olympics</td>
</tr>
<tr>
<td>(75)</td>
<td></td>
</tr>
<tr>
<td>fixRng</td>
<td>kwan withdraws from us gp</td>
</tr>
<tr>
<td>(30)</td>
<td>kwan withdraws from us skating championships</td>
</tr>
<tr>
<td>(50)</td>
<td>kwan pulls out of us figure skating championships for turin olympics bid</td>
</tr>
<tr>
<td>(75)</td>
<td></td>
</tr>
<tr>
<td>LenEmb</td>
<td>kwan withdraws from us skating</td>
</tr>
<tr>
<td>(30)</td>
<td>kwan withdraws from us figure skating championships</td>
</tr>
<tr>
<td>(50)</td>
<td>world champion kwan withdraws from olympic figure skating championships</td>
</tr>
<tr>
<td>(75)</td>
<td></td>
</tr>
<tr>
<td>LenInit</td>
<td>kwan quits us figure skating</td>
</tr>
<tr>
<td>(30)</td>
<td></td>
</tr>
<tr>
<td>(50)</td>
<td>kwan withdraws from us figure skating worlds</td>
</tr>
<tr>
<td>(75)</td>
<td>kwan withdraws from figure skating championships for turin olympics</td>
</tr>
</tbody>
</table>
Existing work

- Controlling length in summarization
  - Fan, Angela, David Grangier, and Michael Auli. "Controllable abstractive summarization."

**Requesting Length 2:** @entity0 [Easter] is over for the wild rabbits of greater @entity2 [Sydney] as councils and parks prepare another attempt to kill them off with a deadly virus. It comes after over 30 government bodies scattered carrots laced with calicivirus.

**Requesting Length 6:** @entity0 [Easter] is over for the wild rabbits of greater @entity2 [Sydney] as councils and parks prepare another attempt to kill them off with a deadly virus. This year, because of really high summer rainfall - which led to great food availability - there has been a big surge in the rabbit population in @entity2 [Sydney].

**Requesting Length 10:** @entity0 [Easter] is over for the wild rabbits of greater @entity2 [Sydney] as councils and parks prepare another attempt to kill them off with strategically placed carrots that have been laced with a deadly virus. This year, because of really high summer rainfall - which led to great food availability - there has been a big surge in the rabbit population in @entity2 [Sydney]. It comes after over 30 government bodies scattered carrots laced with calicivirus around public areas in March.
Existing work

• Controlling length in summarization
  • Etc.
Our contributions

• We build the captioning models borrowing existing ideas from summarization work, by injecting length information into the model.

• To generate captions without an explicit length specification, we add a length prediction module that can predict the optimal length for the input image at hand.
Common captioning model

\[ P(c|I) = \prod_{t} P(w_t|I, w_{<t}) \]

a very pretty horse pulling a cart with some people
Length model

• Treat length as an intermediate variable

\[ P(c|I) = P(c|I, l) \times P(l|I) \]

Control module Length prediction module

• l: length
• c: caption
• l: Image
Model 1: LenEmb

Takase, Sho, and Naoaki Okazaki. "Positional encoding to control output sequence length."
Model 1: LenEmb

\[ l=10 \]

A very pretty horse pulling a cart with some people

Takase, Sho, and Naoaki Okazaki. "Positional encoding to control output sequence length."
Model 1: LenEmb

l=10

LenPred

a very pretty horse pulling a cart with some people

Takase, Sho, and Naoaki Okazaki. "Positional encoding to control output sequence length."
Model 2: Marker

Model 2: Marker

a very pretty horse pulling a cart with some people

Fan, Angela, David Grangier, and Michael Auli. "Controllable abstractive summarization."
Model 2: Marker

<10> a very pretty horse pulling a cart with some people

Fan, Angela, David Grangier, and Michael Auli. "Controllable abstractive summarization."
Model 2: Marker

<10> a very pretty horse pulling a cart with some people

Fan, Angela, David Grangier, and Michael Auli. "Controllable abstractive summarization."
How to train it

• Minimize negative log likelihood

$$\min - \log P(c|I, l) - \log P(l|I)$$
Experiment settings

• coco

• Base model:
  • Att2in

Automatic captioning

- Automatic quality metrics:
  - B4=BLEU4, R=ROUGE, M=METEOR, C=CIDEr, S=SPICE

<table>
<thead>
<tr>
<th></th>
<th>B4</th>
<th>R</th>
<th>M</th>
<th>C</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Att2in</td>
<td>35.9</td>
<td>56.1</td>
<td>27.1</td>
<td>110.6</td>
<td>20.0</td>
</tr>
<tr>
<td>LenEmb</td>
<td>34.9</td>
<td>56.2</td>
<td>27.0</td>
<td>110.0</td>
<td>20.0</td>
</tr>
<tr>
<td>Marker</td>
<td>35.2</td>
<td>56.2</td>
<td>26.9</td>
<td>109.8</td>
<td>19.9</td>
</tr>
</tbody>
</table>
Automatic captioning

- Length distribution
Length control

• Generation with controlled lengths
• For base model, we use fixLen from Kikuchi et al.
Length control

• Fluency:
  • Bad ending rate: the rate of sentences that end with
    • a, an, the, in, for, at, of, with, before, after, on, upon, near, to, is, are, am

Guo, Ts Zhang, et al. "Improving reinforcement learning based image captioning with natural language prior."
Length control

• Accuracy
Length control

- Controllability
Qualitative results

7  a motorcycle parked on a dirt road
10 a motorcycle is parked on the side of a road
16 a motorcycle parked on the side of a dirt road with a fence in the background
22 a motorcycle parked on the side of a dirt road in front of a fence with a group of sheep behind it
28 a motorcycle is parked in a dirt field with a lot of sheep on the side of the road in front of a fence on a sunny day
Qualitative results

7  an airplane is parked at an airport
10 an airplane is parked on the tarmac at an airport
16 an airplane is parked on a runway with a man standing on the side of it
22 an airplane is parked on a runway with a man standing on the side of it and a person in the background
28 an airplane is parked on the tarmac at an airport with a man standing on the side of the stairs and a man standing next to the plane
• Code:
  • [https://github.com/ruotianluo/self-critical.pytorch/tree/length_goal](https://github.com/ruotianluo/self-critical.pytorch/tree/length_goal)

• More results:
  • [https://colab.research.google.com/drive/1TM_KZBixY-L47gHRfXiavgU_TWfzI3I?usp=sharing](https://colab.research.google.com/drive/1TM_KZBixY-L47gHRfXiavgU_TWfzI3I?usp=sharing)