A Hierarchical Encoder-Decoder for Paragraph Summarization

Farzaneh Mahdisoltani

Department of Computer Science
University of Toronto

CSC2541, Fall 2016
**Storyline**

After a serial killer named Red John murdered Patrick Jane's wife and daughter, Jane dedicated his life to hunting down and killing Red John. To that end he gave up his lucrative pretense of being a psychic and joined the California Bureau of Investigation (CBI) as a consultant to the team responsible for investigating the Red John case, led by Senior Agent Teresa Lisbon. Using Jane's exceptional gift for observation and his mentalist trick able to close an unprecedented number of cases, but Jane's unconventional and often outright illegal methods also bring much censure down on Lisbon's head, making his assistance both a blessing and a curse. Meanwhile, the hunt for Red John continues...

**Summary**

A famous "psychic" outs himself as a fake and starts working as a consultant for the California Bureau of Investigation so he can find "Red John," the madman who killed his wife and daughter.

Plot Summary | Plot Synopsis

**Summarization Phenomena:**

- Deletion
- Paraphrase
- Generalization
Related Work

- **Extractive Summarization**: Stitching words and phrases from the source
  - Dorr et al. (2003), Cohn et al. (2008), Woodsend et al. (2010)

- **Abstractive Summarization**: Generating richer summaries based on understanding the input
  - **Attention based**: Rush et al. (2015)
  - **Statistical machine translation for headline generation**: Banko et al. (2000)
  - **RNN with attention**: Chopra et al. (2016)
  - **Encoder-Decoder**: Hu et al. (2015)
Hierarchical Encoder-Decoder Model for Summarization

Paragraph

\[
\begin{align*}
S1 \\
S2 \\
S3 \\
S4 \\
S5 \\
S6
\end{align*}
\]

Sentence to Vector

\[
\begin{align*}
\text{enc}(S1) \\
\text{enc}(S2) \\
\text{enc}(S3) \\
\text{enc}(S4) \\
\text{enc}(S5) \\
\text{enc}(S6)
\end{align*}
\]

Paragraph to Vector

\[
\text{enc}(P)
\]

Summary

\[
\begin{align*}
O1 \\
O2 \\
O3
\end{align*}
\]

Word Level Decoder

\[
\begin{align*}
\text{enc}(O1) \\
\text{enc}(O2) \\
\text{enc}(O3)
\end{align*}
\]

Abstract Concepts

\[
\begin{align*}
\text{enc}(O1) \\
\text{enc}(O2) \\
\text{enc}(O3)
\end{align*}
\]

Sentence Level Decoder
Hierarchical Encoder-Decoder Model for Summarization

Objective Function: $\sum_t P(o^t_w | o^{<t}_w, o_s, h_p)$
Selecting the Correct Summary from Candidates

Paragraph

S1
S2
S3
S4
S5
S6

Sentence to Vector

enc(S1)
enc(S2)
enc(S3)
enc(S4)
enc(S5)
enc(S6)

Paragrap
h to Vector

enc(P)

Farzaneh Mahdisoltani (University of Toronto)
A Hierarchical Encoder-Decoder for Paragraph Summarization
CSC2541, Fall 2016
## Preliminary Results

<table>
<thead>
<tr>
<th>Paragraph Representation Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>25%</td>
</tr>
<tr>
<td>Word2Vec</td>
<td>65%</td>
</tr>
<tr>
<td>Skip-Thought Vectors</td>
<td>62%</td>
</tr>
<tr>
<td>Bag of Words</td>
<td>67%</td>
</tr>
<tr>
<td>RNN Encoder</td>
<td>57%</td>
</tr>
<tr>
<td>Hierarchichal Encoder</td>
<td>54%</td>
</tr>
</tbody>
</table>

**Table:** Training the encoder for multiple choice question
Thank You