Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation

Lili Mou,1 Yiping Song,2 Rui Yan,3 Ge Li,1 Lu Zhang,1 Zhi Jin1

1Institute of Software
2Institute of Network Computing and Information Systems
3Institute of Computer Science and Technology

School of EECS, Peking University, P. R. China

COLING, Osaka, Japan
December, 2016
Outline

1. Introduction: Human-Computer Conversation Systems
2. Approach: seq2BF Model
3. Experimental Results
4. Conclusion: Related Work
Outline

1. Introduction: Human-Computer Conversation Systems
2. Approach: seq2BF Model
3. Experimental Results
4. Conclusion: Related Work
Human-computer conversation has long attracted interest in both academia and industry.

- Task/Domain-oriented systems
- Open-domain conversation systems
Task/Domain-Oriented Dialog Systems

- Transportation domain: TRAIN-95 [Ferguson et al., 1996]
- Education: AutoTutor [Graesser et al., 2005]
- Restaurant booking [Wen et al., 2016]

Approaches:
- Planning
- Rule-based, Slot-filling, etc.
Open-Domain Conversation

Why is chatbot-like conversation important?

- Tackles the problem of natural language understanding and generation
- Commercial needs

Approaches:

- Retrieval-based systems [Isbell et al., 2000, Wang et al., 2013]
- Generative systems
  - Phrase-based machine translation [Ritter et al., 2011]
  - Neural networks (seq2seq models) [Shang et al., 2015]
Open-domain, neural network-based, generative short-text conversation
Outline

1. Introduction: Human-Computer Conversation Systems
2. Approach: seq2BF Model
3. Experimental Results
4. Conclusion: Related Work
Sequence-to-Sequence Model

Encode-decoder framework

- Encode a user-issued query as a vector
- Decode it as an utterance

Other applications:

- Machine translation
- Summarization
- etc
But the problem is...

Meaningless and universally relevant replies

- Example in a previous study (in English) [Li et al., 2016]
  - Q: How come you never say it?
  - Q: How much time do you have here?
  - R: “I don’t know”

- Our scenario (in Chinese)
  - 我也是 (Me too)
Why?

- The query does not convey sufficient information.
- A query may have multiple appropriate replies.
- Universally relevant utterances appear (slightly) more frequently than other replies.

Conversation is different than translation.
Query: What are you going to do?

Candidate replies:

- Take a walk
- I don't know
- Have something to eat
Solutions

- Diversity-promoting objective function [Li et al., 2016]
  
- Variational encoding the source [Serban et al., 2016]
Our Intuition

- Some words in the utterance are highly correlated with the source.
  - Thank you
  - You’re welcome
- Predict a keyword first, and generate a reply containing the keyword
- A “sequence to backward and forward sequences” model accomplishes this goal.
Take a walk

I don't know

Have something to eat

Predict a keyword externally

Origin
Overview

(a) Keyword prediction

(b) Backward sequence

(c) Forward sequence

Step I
PMI statistics

Step II
seq2BF model
Keyword Predictor

Computing the point-wise mutual information (PMI):

\[
\text{PMI}(w_q, w_r) = \log \frac{p(w_q, w_r)}{p(w_q)p(w_r)} = \log \frac{p(w_q|w_r)}{p(w_q)}
\]

Prediction

\[
w_r^* = \arg \max_{w_r} \text{PMI}(w_{q_1} \cdots w_{q_n}, w_r)
\]

where

\[
\text{PMI}(w_{q_1} \cdots w_{q_n}, w_r) = \log \frac{p(w_{q_1} \cdots w_{q_n}|w_r)}{p(w_{q_1} \cdots w_{q_n})}
\]

\[
\approx \log \frac{\prod_{i=1}^{n} p(w_{q_i}|w_r)}{\prod_{i=1}^{n} p(w_{q_i})} = \sum_{i=1}^{n} \log \frac{p(w_{q_i}|w_r)}{p(w_{q_i})} = \sum_{i=1}^{n} \text{PMI}(w_{q_i}, w_r)
\]
seq2BF Model

Traditional language models (sentence generators) start from the first word and generate following words in sequence.

\[
p(r_1, \ldots, r_m | q) = p(r_1 | q)p(r_2 | r_1, q) \cdots p(r_m | r_1 \cdots r_{m-1}, q)
\]

\[
= \prod_{i=1}^{m} p(r_i | r_1 \cdots r_{i-1}, q)
\]

The seq2BF model generates previous and future words conditioned on a given word.

\[
p\left(\begin{array}{c}
  \cdot \cdot \cdot r_{k-1} \cdot \cdot \cdot r_1 \\
  \cdot \cdot \cdot r_{k+1} \cdot \cdot \cdot r_m
\end{array} \mid r_k, q\right) = \prod_{i=1}^{k-1} p^{(bw)}(r_{k-i} \mid r_k, q, \cdot) \prod_{i=1}^{m-k} p^{(fw)}(r_{k+i} \mid r_k, q, \cdot)
\]
Asynchronously generating the two sequences

- First the backward half
- Then the forward half conditioned on the backward half sequence

Training:

- Randomly choose a word as the keyword, splitting an utterance into two subsentences

⇒ “half” language model
Outline

1. Introduction: Human-Computer Conversation Systems
2. Approach: seq2BF Model
3. Experimental Results
4. Conclusion: Related Work
Datasets and Settings

- A Chinese conversation dataset containing up to 500k query-reply pairs
- RNNs are in character level, but keyword prediction is in phrase level.
- Embeddings and the hidden layer: 500-dimensional
- Rmsprop, generally default hyperparameters
Evaluation Metrics

Attempt #1: BLEU scores (0 BLEU-2 score by one annotator) (Abandoned)

Attempt #2: Human evaluation
  - Pointwise annotation
  - Pairwise annotation

Attempt #3: Intrinsic metric (entropy)

$$-\frac{1}{|R|} \sum_{w \in R} \log_2 p(w)$$
## Experimental Results

### Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Point</th>
<th>Human</th>
<th>Length</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>seq2seq</td>
<td>0.58</td>
<td></td>
<td>5.61</td>
<td>6.960</td>
</tr>
<tr>
<td>seq2BF−</td>
<td>0.46</td>
<td></td>
<td>5.60</td>
<td>6.971</td>
</tr>
<tr>
<td>seq2BF+</td>
<td>0.67</td>
<td></td>
<td>5.31</td>
<td>9.139</td>
</tr>
<tr>
<td>Groundtruth</td>
<td>−</td>
<td></td>
<td>9.19</td>
<td>8.832</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Wins</th>
<th>Ties</th>
<th>Loses</th>
</tr>
</thead>
<tbody>
<tr>
<td>seq2seq</td>
<td>24.7</td>
<td>26.0</td>
<td>49.3</td>
</tr>
<tr>
<td>seq2BF+</td>
<td>49.3</td>
<td>26.0</td>
<td>24.7</td>
</tr>
</tbody>
</table>
### Examples

<table>
<thead>
<tr>
<th>Chinese</th>
<th>English (translated)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Query</strong> 李有男友公开过了&lt;br&gt;都已经分了 之前李的贴吧都在讨论了</td>
<td>It’s known that Li† has a boyfriend. Broken up. There’s discussion in her Tieba.‡</td>
</tr>
<tr>
<td><strong>Grountruth</strong> 我是男的&lt;br&gt;我的男友&lt;br&gt;有绯闻男友</td>
<td>I am a male&lt;br&gt;My boyfriend&lt;br&gt;Has a rumored boyfriend</td>
</tr>
<tr>
<td><strong>seq2seq</strong> 我是男的&lt;br&gt;我的男友&lt;br&gt;有绯闻男友</td>
<td>Me too!&lt;br&gt;Me too!&lt;br&gt;What is your score</td>
</tr>
<tr>
<td><strong>seq2BF−</strong> 人大复试飘过&lt;br&gt;这么牛，什么专业</td>
<td>Passed second-round exam of Renming Univ. Cool, what’s your major</td>
</tr>
<tr>
<td><strong>seq2BF+</strong> 我也是！&lt;br&gt;我也是！&lt;br&gt;分数是什么</td>
<td>Me too!&lt;br&gt;Me too!&lt;br&gt;What is your score</td>
</tr>
<tr>
<td><strong>seq2seq</strong> 我也是！&lt;br&gt;我也是！&lt;br&gt;分数是什么</td>
<td>Me too!&lt;br&gt;Me too!&lt;br&gt;What is your score</td>
</tr>
<tr>
<td><strong>seq2BF−</strong> 挺漂亮的 祝福祝福&lt;br&gt;下手慢了哈。现在</td>
<td>So beautiful, congratulations&lt;br&gt;You’re late till now</td>
</tr>
<tr>
<td><strong>seq2BF+</strong> 挺漂亮的 祝福祝福&lt;br&gt;下手慢了哈。现在</td>
<td>So beautiful, congratulations&lt;br&gt;You’re late till now</td>
</tr>
<tr>
<td><strong>seq2seq</strong> 我是你的头像&lt;br&gt;我是你的头像&lt;br&gt;第一张图像是谁</td>
<td>I’m in your photo&lt;br&gt;I’m in your photo&lt;br&gt;Who is in your first photo</td>
</tr>
<tr>
<td><strong>seq2BF−</strong> 我是你的头像&lt;br&gt;我是你的头像&lt;br&gt;第一张图像是谁</td>
<td>I’m in your photo&lt;br&gt;I’m in your photo&lt;br&gt;Who is in your first photo</td>
</tr>
<tr>
<td><strong>seq2BF+</strong> 我是你的头像&lt;br&gt;我是你的头像&lt;br&gt;第一张图像是谁</td>
<td>I’m in your photo&lt;br&gt;I’m in your photo&lt;br&gt;Who is in your first photo</td>
</tr>
</tbody>
</table>
Additional Analysis

Entropy

<table>
<thead>
<tr>
<th>Model</th>
<th>seq2seq</th>
<th>seq2BF_</th>
<th>seq2BF_+ keyword</th>
<th>seq2BF_+ remaining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>6.960</td>
<td>6.971</td>
<td>11.630</td>
<td>7.422</td>
</tr>
</tbody>
</table>

Length [Mou et al., 2015]
Outline

1. Introduction: Human-Computer Conversation Systems
2. Approach: seq2BF Model
3. Experimental Results
4. Conclusion: Related Work
Conclusion

- Topic-augmenting [Xing et al., 2016]
- Combination of retrieval and generative dialog systems [Song et al., 2016]
Thank you for listening!

Q & A
References


