SnowBot: Building a Chatbot using different ML frameworks

Pinglei Guo, Yusi Xiang, Yunzheng Zhang, Weiting Zhan

CMPS 242 Fall 17   2017-12-14
SnowBot

Jon Snow, you know nothing

有问皆答不知道

https://github.com/at15/snowbot
Agenda

● ChatBot
● seq2seq
● ML frameworks & seq2seq frameworks
● Data & Evaluation
● Take away
## ChatBot

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Domain</td>
<td>Siri/Cortana</td>
<td>Hard</td>
</tr>
<tr>
<td>Close Domain</td>
<td>Buy movie ticket</td>
<td>Easy</td>
</tr>
<tr>
<td>Generative</td>
<td>Bots speak nonsense</td>
<td>Hard</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Email auto reply</td>
<td>Easy</td>
</tr>
</tbody>
</table>

**Easiest = close domain + retrieval ?**
ChatBot

<table>
<thead>
<tr>
<th>Type</th>
<th>Example</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Domain</td>
<td>Siri/Cortana</td>
<td>Hard</td>
</tr>
<tr>
<td>Close Domain</td>
<td>Buy movie ticket</td>
<td>Easy</td>
</tr>
<tr>
<td>Generative</td>
<td>Bots speak nonsense</td>
<td>Hard</td>
</tr>
<tr>
<td>Retrieval</td>
<td>Email auto reply</td>
<td>Easy</td>
</tr>
</tbody>
</table>

Easiest (to do it right) = close domain + retrieval
Easiest (to implement) = open domain + generative
Seq2Seq

seq2seq

expected: I don't know</s>
actual: I might know<unk>

Loss

discard

projection

embedding_encoder

embedding_decoder

Any Idea </s> <PAD>

<s> I don't know <PAD>
Seq2Seq vs Seq2one

**seq2seq**

- Embedding Encoder
- Any Idea </s> <PAD>
- Decoder

**seq2one**

- Embedding Encoder
- Any Idea </s> <PAD>
- Discard
- Loss
  - Expected: 0
  - Actual: 1
- Projection
## Seq2Seq variants

<table>
<thead>
<tr>
<th>Type</th>
<th>Train input</th>
<th>Train output</th>
<th>Test input</th>
<th>Test output</th>
</tr>
</thead>
<tbody>
<tr>
<td>text generation</td>
<td>You are my foe</td>
<td>You are my foe</td>
<td>You are</td>
<td>my &lt;unk&gt;</td>
</tr>
<tr>
<td>translation</td>
<td>你好</td>
<td>Hello</td>
<td>吃了么</td>
<td>Good morning</td>
</tr>
<tr>
<td>single turn QA</td>
<td>Any idea</td>
<td>I don’t know</td>
<td>How’s the weather</td>
<td>I don’t know</td>
</tr>
<tr>
<td>multi turn QA</td>
<td>Got it?; No; Why?</td>
<td>I don’t know</td>
<td>Hi; Hey; What’s up?</td>
<td>I don’t know</td>
</tr>
</tbody>
</table>
## ML Frameworks

### Table 1: Popular Machine learning Frameworks

<table>
<thead>
<tr>
<th>Name</th>
<th>Star</th>
<th>Company</th>
<th>First Release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tensorflow</td>
<td>81697</td>
<td>Google</td>
<td>2015-11-08</td>
</tr>
<tr>
<td>Keras</td>
<td>22872</td>
<td>N/A</td>
<td>2015-01-13</td>
</tr>
<tr>
<td>Caffe</td>
<td>21739</td>
<td>N/A</td>
<td>2014-03-19</td>
</tr>
<tr>
<td>CNTK</td>
<td>13366</td>
<td>Microsoft</td>
<td>2016-01-22</td>
</tr>
<tr>
<td>MXNet</td>
<td>12392</td>
<td>Amazon</td>
<td>2015-12-09</td>
</tr>
<tr>
<td>PyTorch</td>
<td>10142</td>
<td>Facebook</td>
<td>2016-08-31</td>
</tr>
<tr>
<td>Torch</td>
<td>7537</td>
<td>N/A</td>
<td>2015?</td>
</tr>
</tbody>
</table>
## seq2seq Frameworks

<table>
<thead>
<tr>
<th>Name</th>
<th>Framework</th>
<th>GitHub</th>
<th>Star</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenNMT-py</td>
<td>PyTorch</td>
<td>OpenNMT/OpenNMT-py</td>
<td>773</td>
</tr>
<tr>
<td>OpenNMT</td>
<td>Torch</td>
<td>OpenNMT/OpenNMT</td>
<td>1461</td>
</tr>
<tr>
<td>OpenNMT-tf</td>
<td>Tensorflow</td>
<td>OpenNMT/OpenNMT-tf</td>
<td>145</td>
</tr>
<tr>
<td>nmt</td>
<td>Tensorflow</td>
<td>tensorflow/nmt</td>
<td>2094</td>
</tr>
<tr>
<td>seq2seq</td>
<td>Tensorflow</td>
<td>google/seq2seq</td>
<td>3019</td>
</tr>
<tr>
<td>tensor2tensor</td>
<td>Tensorflow</td>
<td>tensorflow/tensor2tensor</td>
<td>2975</td>
</tr>
<tr>
<td>Sockeye</td>
<td>MXNet</td>
<td>awslabs/sockeye</td>
<td>290</td>
</tr>
</tbody>
</table>
Q: do you know
A: i don't think so.

Q: hi
A: hi.

Q: thank you
A: i don't want to see you.

Q: do you know
A: i don't know what i was in the <unk>

Q: hi
A: thank you!

Q: thank you
A: thank you so much ❤️
Q: what is love
A: I don't know

Q: love is love
A: I love you.

Q: anything with love
A: I don't think so.

Q: love anything
A: I think so.

Q: what is love
A: i don't want to know what it is.

Q: love is love
A: what do you want me to do?

Q: I love you
A: I don’t want to see you.

Q: i love you
A: i love you
Evaluation - Cornell Movie Dialog & Twitter Chat

Q: where is he
A: i don't know what he is.

Q: where is she
A: she said she was a <unk>

Q: he said what?
A: he said he was a <unk>

Q: i said what?
A: what do you want?

Q: where is she
A: i don't know what she was in the <unk>

Q: where is he
A: i think he is a <unk>

Q: where is him
A: i think he is a <unk>

Q: what am i doing
A: i don't know what you want to do with the <unk>

Q: where am i
A: i don't know what i was in the <unk>
Evaluation - Conclusion

- most time the bot is using general response
  - filter out *I don’t know* won’t work, it will use *what do you want, what do you mean, I don’t think so*
- it can detect certain keywords i.e. him -> he, her -> she, love
  - but it overfits to high frequency words
- text pre-processing is important
  - i.e. what What What? can be converted to same token. (they would have close word vec if data is large enough)
Take away

● Open domain generative ChatBot is pretty dumb now
  ○ It’s end to end, easy to implement, hard to capture the essence of dialog
  ○ You may see good result because it happens to look good

● Data is important and precious
  ○ Twitter has emoji (i.e. 😊) and cornell bot can’t say I 😊 you because it’s not even in the vocab
  ○ Cornell Movie Dialog Corpus is small and old, but you still see it in many papers

● Amazon GPU instances takes two days to increase quota (default 0)
  ○ just buy yourself a 1080 Ti for video games, machine learning
Thank you!

SnowBot: Building a chatbot using different ML frameworks

Pinglei Guo, Yusi Xiang, Yunzheng Zhang, Weiting Zhan

https://github.com/at15/snowbot