Deep Reinforcement Learning for Dialogue Generation

Authors: Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao and Dan Jurafsky

Abhijeet Srivastava

CISC850
Cyber Analytics
Abstract

• Current neural network models.
  – They have drawbacks

• Solution:
  – Deep reinforcement learning.
Introduction

• Sequence-to-sequence (SEQ2SEQ) model (Sutskever et al., 2014).

• SEQ2SEQ is not without faults.
Introduction

• Drawbacks
  – Generate generic response
  – Can get stuck into infinite loop
Conversation example 1

A: Where are you going? (1)  
B: I’m going to the restroom. (2)  
A: See you later. (3)  
B: See you later. (4)  
A: See you later. (5)  
B: See you later. (6)  
...

...

CISC850
Cyber Analytics
Conversation example 2

A: Where are you going? (1)
B: I’m going to the police station. (2)
A: I’ll come with you. (3)
B: No, no, no, no, you’re not going anywhere. (4)
A: Why? (5)
B: I need you to stay here. (6)
A: I don’t know what you are talking about. (7)
...
Conversation example 3

A: how old are you? (1)
B: I’m 16. (2)
A: 16? (3)
B: I don’t know what you are talking about. (4)
A: You don’t know what you are saying. (5)
B: I don’t know what you are talking about. (6)
A: You don’t know what you are saying. (7)
...

Conversation example 4

A: How old are you? (1)
B: I’m 16. Why are you asking? (2)
A: I thought you were 12. (3)
B: What made you think so? (4)
A: I don’t know what you are talking about. (5)
B: You don’t know what you are saying. (6)
...
...

Introduction (cont...)

• What we need along with SEQ2SEQ
  – Developer-defined rewards.
  – Foresight.
Reinforcement Learning for Open-Domain Dialogue

• The learning system consists of two agents. 
  p&q
  – Action
  – State
  – Policy
Reinforcement Learning for Open-Domain Dialogue

– Reward
  • R reward is obtained for each action.
  • Ease of answering
  • Information flow
  • Semantic coherence
Simulation

• Simulate the process of two virtual agents.
  – Supervised learning
  – Mutual Information
Simulation

• Mutual Information
  – Encoder-decoder model.
  – Stochastic gradient descent.
Simulation

• Steps for dialogue simulation between two agents are:
  – Initiation step
  – Agent1 response
  – Agent2 response
Simulation

• Optimization
• Curriculum learning
Simulation

Figure 1: Dialogue simulation between the two agents.
Experimental Results

- The dialogues are evaluated using two metrics
  - Conversation length
  - Diversity
Experimental Results

• Data Set
  – High-quality initial inputs fed to the agent.

• Automatic Evaluation
  – Length of the dialogue
  – Diversity
Experimental Results

• Human evaluation (3 settings)
  – Which is better
  – Easier to respond
  – High quality
Experimental Results

- Qualitative analysis

<table>
<thead>
<tr>
<th>Input Message</th>
<th>Mutual Information Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>How old are you?</td>
<td>I’m 16.</td>
</tr>
<tr>
<td>What’s your full name?</td>
<td>I have no idea.</td>
</tr>
<tr>
<td>I don’t want to go home tonight.</td>
<td>Really?</td>
</tr>
<tr>
<td>Do you have any feelings for me?</td>
<td>I don’t know what you are talking about.</td>
</tr>
<tr>
<td>How much time do you have here?</td>
<td>Not long enough. Sorry, sir.</td>
</tr>
<tr>
<td>Shall we get started?</td>
<td>Of course!</td>
</tr>
<tr>
<td>Do you play football?</td>
<td>No, i don’t</td>
</tr>
<tr>
<td>We’d have to talk to him.</td>
<td>I mean, he’s a good guy</td>
</tr>
<tr>
<td>How come you never say it?</td>
<td>Because I don’t want to hurt you.</td>
</tr>
</tbody>
</table>
Experimental Results

- Qualitative analysis

<table>
<thead>
<tr>
<th>Input Message</th>
<th>RL</th>
</tr>
</thead>
<tbody>
<tr>
<td>How old are you?</td>
<td>I’m 16, why are you asking?</td>
</tr>
<tr>
<td>What’s your full name?</td>
<td>What’s yours?</td>
</tr>
<tr>
<td>I don’t want to go home tonight.</td>
<td>Why?</td>
</tr>
<tr>
<td>Do you have any feelings for me?</td>
<td>I don’t want to hurt your feelings.</td>
</tr>
<tr>
<td>How much time do you have here?</td>
<td>Ten seconds.</td>
</tr>
<tr>
<td>Shall we get started?</td>
<td>Yes. We’ve got a lot of work to do here.</td>
</tr>
<tr>
<td>Do you play football?</td>
<td>Yes. I love football!</td>
</tr>
<tr>
<td>We’d have to talk to him.</td>
<td>About what?</td>
</tr>
<tr>
<td>How come you never say it?</td>
<td>I don’t think it’s a good idea to say it.</td>
</tr>
</tbody>
</table>
Experimental Results

• Issues observed

<table>
<thead>
<tr>
<th>A: What’s your name?</th>
</tr>
</thead>
<tbody>
<tr>
<td>B: Daniel.</td>
</tr>
<tr>
<td>A: How old are you?</td>
</tr>
<tr>
<td>B: Twelve. What’s your name?</td>
</tr>
<tr>
<td>A: Daniel.</td>
</tr>
<tr>
<td>B: How old are you?</td>
</tr>
<tr>
<td>A: Twelve. What’s your name?</td>
</tr>
<tr>
<td>B: Daniel.</td>
</tr>
<tr>
<td>A: How old are you?</td>
</tr>
<tr>
<td>B: ...</td>
</tr>
</tbody>
</table>
Conclusion

• Overcame drawbacks of SEQ2SEQ
• Responses are comparable to RL