

The background of the slide features a large, semi-transparent watermark of the University of Pennsylvania seal. The seal is circular and contains the text 'UNIVERSITY OF PENNSYLVANIA' around the perimeter, '1743' at the bottom, and 'SOL' in the center. On the left side of the seal, there is a shield with the words 'GRAMM', 'PHIOL', 'RHETOR', 'ETHICA' on the left and 'METAPH', 'LOGICA', 'MATHEM', 'PHYSICA' on the right.

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)

...

...

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

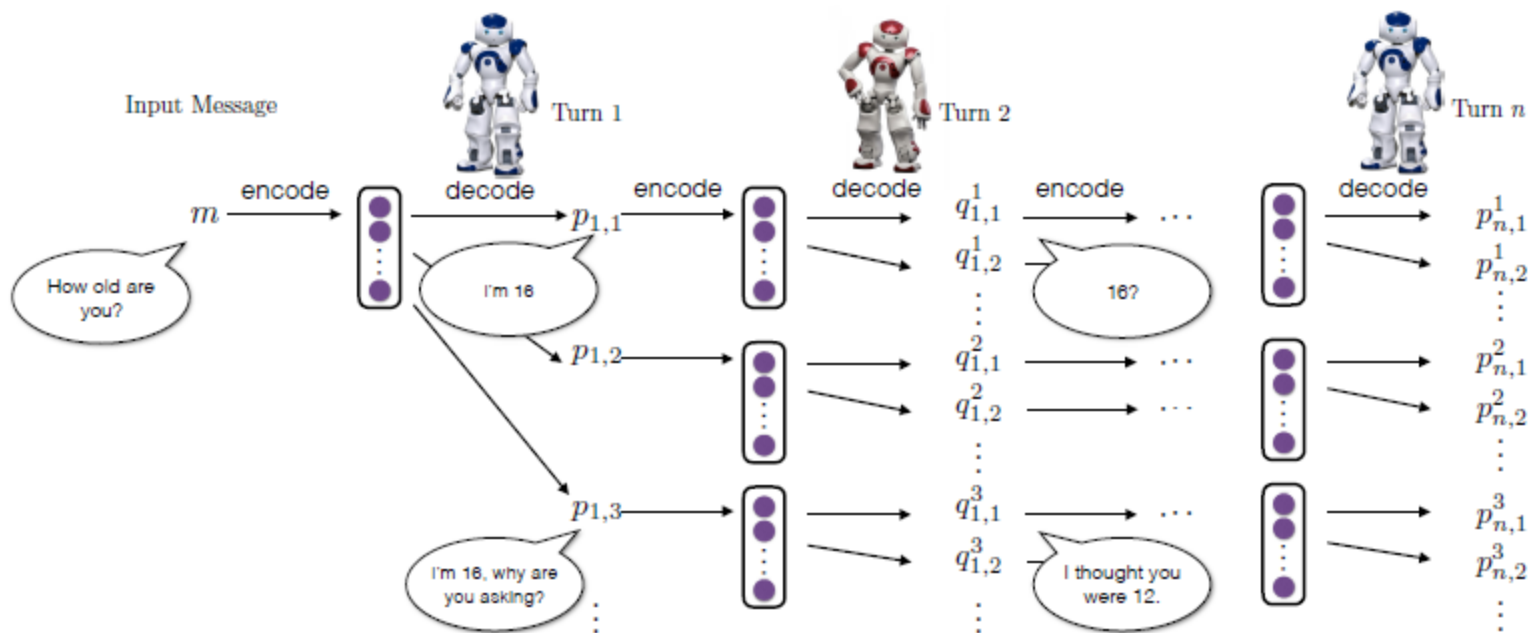


Figure 1: Dialogue simulation between the two agents.

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

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

Experimental Results

- Qualitative analysis

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

Experimental Results

- Issues observed

A: What's your name ?

B: Daniel.

A: How old are you ?

B: Twelve. What's your name ?

A: Daniel.

B: How old are you ?

A: Twelve. What's your name ?

B: Daniel.

A: How old are you ?

B ...

Conclusion

- Overcame drawbacks of SEQ2SEQ
- Responses are comparable to RL