Dialogue Systems & Reinforcement Learning

Nabiha Asghar
Ph.D. student @ UW
Data Scientist @ ProNav Technologies (www.pronavigator.ai)

Outline

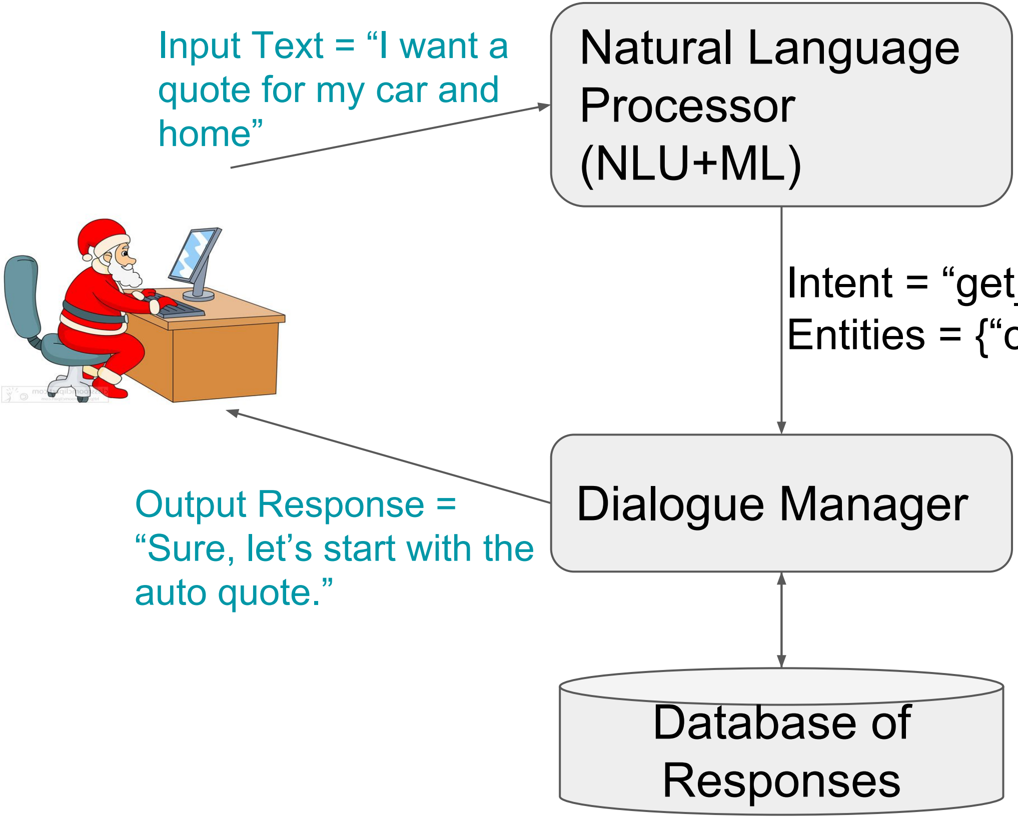
- Introduction to Dialogue Systems (DS)
- Introduction to ProNav Technologies
- Natural Language Processing and ML for DS
- Deep RL for DS

What is a dialogue system?

- An artificial agent that can carry out spoken or text-based conversations with humans (Alexa, Siri, Cortana)
  - also called chatbot, conversational agent
- Classification:
  - Retrieval-based
  - Generative

What is a dialogue system?

1. Retrieval-based



Input Text = “I want a quote for my car and home”

Natural Language Processor (NLU+ML)

Intent = “get_quote”
Entities = {“car”, “home”}

Dialogue Manager

Output Response = “Sure, let’s start with the auto quote.”

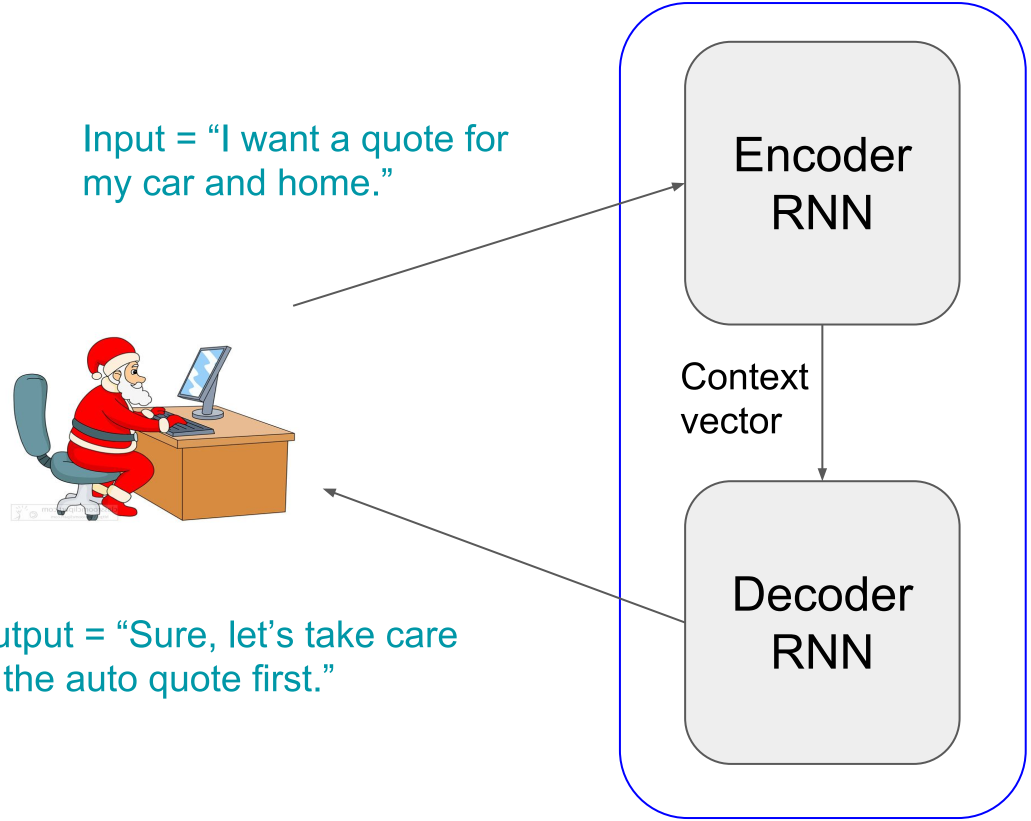
Database of Responses

What does the user want?

State machine;
If-else rules

What is a dialogue system?

2. Generative



Input = “I want a quote for my car and home.”

Output = “Sure, let’s take care of the auto quote first.”

Recurrent Neural Network (RNN)

Retrieval-based dialogue systems

1. Easier machine learning tasks to solve
(input=sentence, output=intent/entity)
2. Predictable responses
3. Easier-to-control behaviour
4. Don’t need tons of training data
5. # of if-else rules can grow exponentially
6. Do not generalize as well

Generative dialogue systems

1. Hard machine learning task (input=sentence, output=sentence)
2. Unpredictable responses
3. Hard-to-control behaviour
4. Tons of training data required
5. No if-else rules required
6. Can generalize well

Retrieval-based Dialogue Systems

NLU for Retrieval-based DS

What is the intent of a text?

“I want an auto insurance quote” (intent = get_quote)

vs.

“Do you sell policies outside Canada?” (intent = FAQ_location)

What are the useful entities in a text?

“I want car insurance”

vs.

“I want home insurance”

Intent Classification

Input: “*Do you provide auto insurance in Ontario?*”

Output: one element from the set *{get_quote, get_contact_info, FAQ_location, FAQ_eligibility, …. }*

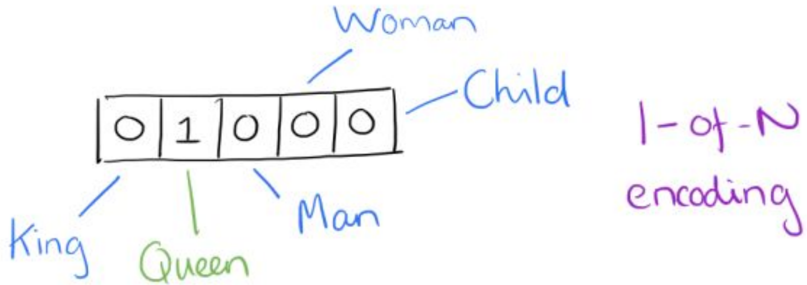
Named Entity Recognition (NER)

Input: “*Do you provide auto insurance in Ontario?*”

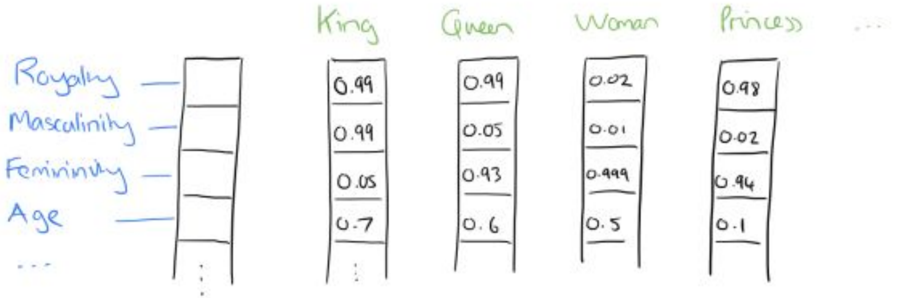
Output: For each word in input, produce an element from the set *{NULL, insurance_type, province_name, person_name, number, date, …. }*

Intent Classification & Named Entity Recognition (NER)

Key Idea: Model a sentence as a sequence of ‘word vectors’ (Word2Vec, GloVe)



One-hot encodings of words



Word vectors

Features: Word Vectors

Classification Algorithms: Support Vector Machines, Conditional Random Fields, etc

Challenges

- Long messages
    - Well, I just have a problem with insurance companies in general. Our private social club has been paying for insurance for over 40 years & has never had a claim. An recent accident where an individual was hurt caused such a mess. A member slipped & broke his leg at the club but had no intentions of suing. However the incident was reported by the club president to the insurance company. Then the insurance company approached the member & asked them to accept a "settlement" & sign a waiver that the member would not file a claim/lawsuit against the club. The member felt obliged to sign & therefore accepted the "settlement". Then the insurance company told our club that every member must now sign a waiver immediately stating they will not hold the club liable for any injuries incurred during any activities at the club or the company will no longer insure our club. We are annoyed that a clause/waiver was not already in place, our insurance company, through all these years, does not have any clause like this in our liability section & now they have thrown this in our faces, raised our rates & none of this would have happened if they had not been negligent in our policy's terms in the first place. Hows that? It just seems, we need insurance to protect us but once we need our protection through a claim we're faced with higher rates. I can tell you that we have paid a ton of money in insurance in our lifetime, made one claim & up went the premiums. And this is called "protection".

Challenges

- Long messages
  - Well, I just have a problem with insurance companies in general. Our private social club has been paying for insurance for over 40 years & has never had a claim. An recent accident where an individual was hurt caused such a mess. A member slipped & broke his leg at the club but had no intentions of suing. However the incident was reported by the club president to the insurance company. Then the insurance company approached the member & asked them to accept a "settlement" & sign a waiver that the member would not file a claim/lawsuit against the club. The member felt obliged to sign & therefore accepted the "settlement". Then the insurance company told our club that every member must now sign a waiver immediately stating they will not hold the club liable for any injuries incurred during any activities at the club or the company will no longer insure our club. We are annoyed that a clause/waiver was not already in place, our insurance company, through all these years, does not have any clause like this in our liability section & now they have thrown this in our faces, raised our rates & none of this would have happened if they had not been negligent in our policy's terms in the first place. Hows that? It just seems, we need insurance to protect us but once we need our protection through a claim we're faced with higher rates. I can tell you that we have paid a ton of money in insurance in our lifetime, made one claim & up went the premiums. And this is called "protection".
- Unique messages
  - Visitor: 19:51:22: i WOULD LIKE A QUOTE BUT MY NUMBER SIX IS NOT WORKING SO i COULD NOT COMPLETE MY POSTAL CODE FOR QUOTE

DRL in Retrieval-based Dialogue*

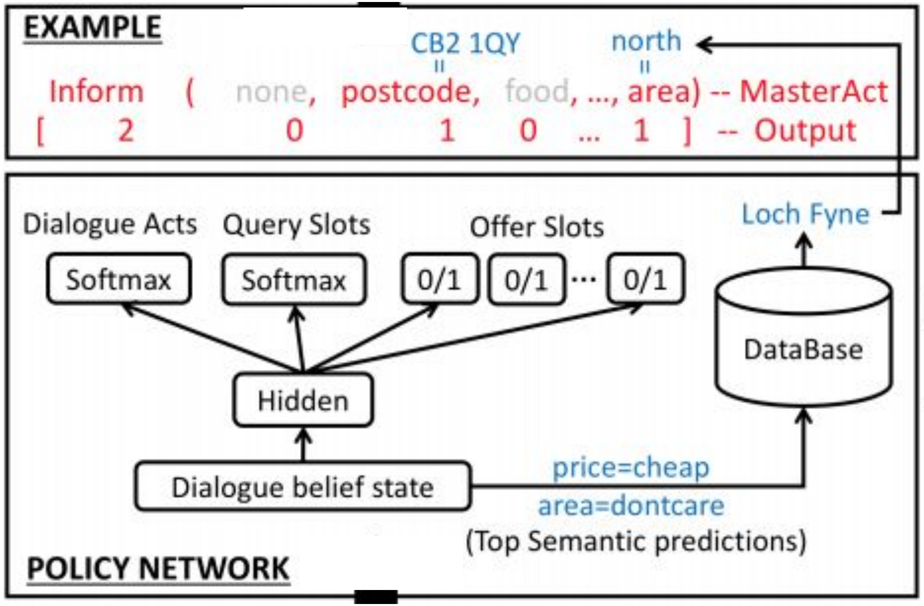
*Su, Pei-Hao, et al. "Continuously learning neural dialogue management." *arXiv preprint arXiv:1606.02689* (2016).

DRL in Retrieval-based Dialogue*

- Application: Providing restaurant information
- Domain: 150 restaurants, each with 6 slots:
  - {foodtype, area, price-range} to constrain the search
  - {phone, address, postcode}: informable properties
- System Goal:
  - Determine the intent of the system response
  - Determine which slot to talk about

*Su, Pei-Hao, et al. "Continuously learning neural dialogue management." *arXiv preprint arXiv:1606.02689* (2016).

DRL in retrieval-based Dialogue (cont’d)



Dialogue belief state: encodes the understood user intents + dialogue history

Policy Network: 1 hidden layer (tanh), output layer with 2 softmax partitions, 3 sigmoid partitions

Dialogue Acts: {request, offer, inform, select, bye}

Query slots: {food, price-range, area, none}

Offer slots: {Area, phone, postcode}

*Su, Pei-Hao, et al. "Continuously learning neural dialogue management." *arXiv preprint arXiv:1606.02689* (2016).

DRL in Retrieval-based Dialogue (cont’d)

- Training:
  - Phase 1: Supervised learning on AMT corpora of 720 dialogues, maximize likelihood of data
  - Phase 2: Reinforcement Learning; find policy that maximizes expected reward of a dialogue with T turns

$$J(\theta) = E\left[\sum_{t=1}^{T} \gamma^t r(s_t, a_t) \middle| \pi_\theta\right]$$

*Su, Pei-Hao, et al. "Continuously learning neural dialogue management." *arXiv preprint arXiv:1606.02689* (2016).

DRL in Retrieval-based Dialogue (cont’d)

- Training:
    - Phase 1: Supervised learning on AMT corpora of 720 dialogues, maximize likelihood of data
    - Phase 2: Reinforcement Learning; find policy that maximizes expected reward of a dialogue with T turns

$$J(\theta) = E\left[\sum_{t=1}^{T} \gamma^t r(s_t, a_t) \middle| \pi_\theta\right]$$

Policy Gradient Methods

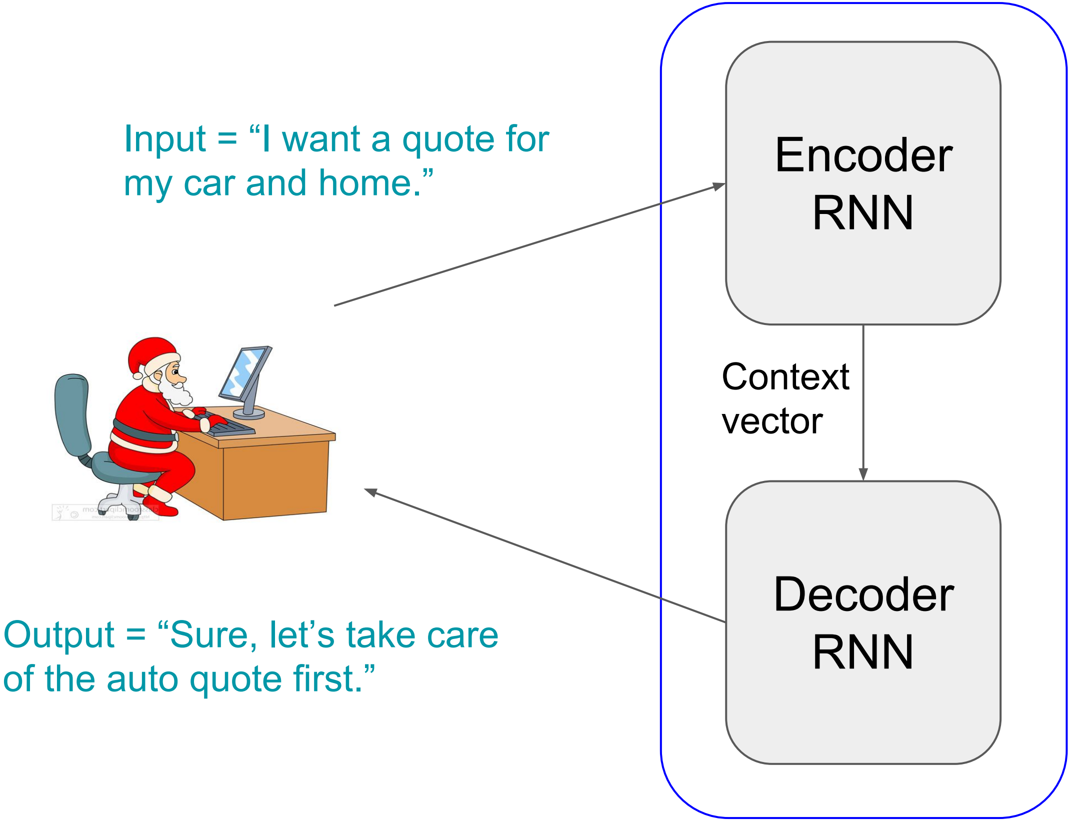
*Su, Pei-Hao, et al. "Continuously learning neural dialogue management." *arXiv preprint arXiv:1606.02689* (2016).

Policy Gradient Methods

- A class of RL methods (Lecture 7a)
- Problem: Maximize *E*[*R* | π_θ]
- Intuitions: collect a bunch of trajectories using π_θ , and
    - Make the good trajectories more probable
    - Make the good actions more probable

Generative Dialogue Systems

Recall: Neural Text Generation

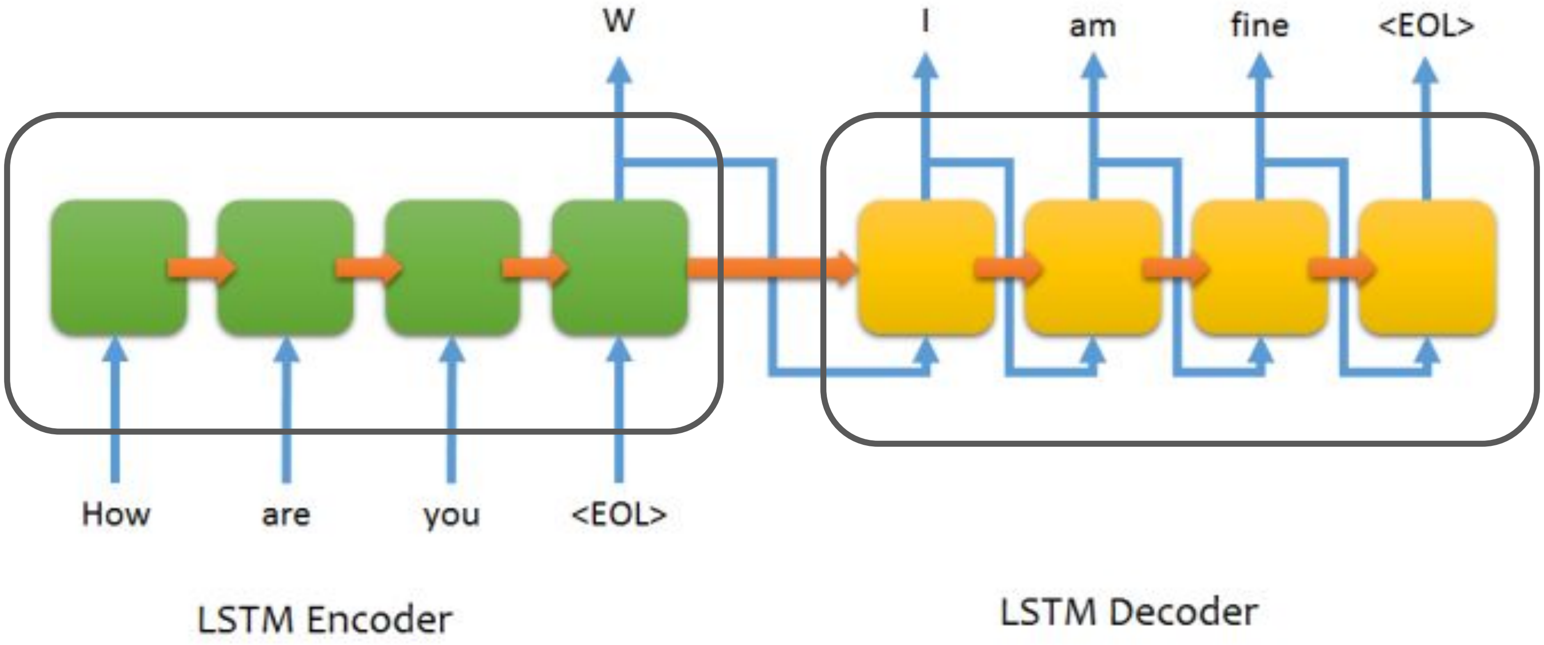


Input = “I want a quote for my car and home.”

Output = “Sure, let’s take care of the auto quote first.”

Recurrent Neural Network (RNN)

Text Generation using RNNs (SEQ2SEQ)



Supervised Training Objective: Maximum Likelihood

SEQ2SEQ Challenges

- Likely to generate short and dull responses (“I don’t know”, “I’m not sure”)
- Short-sighted (based on last few utterances only)
- ‘Maximum likelihood’ is not how humans converse
- Fully supervised setting: at-least 0.5 million (sentence, sentence) pairs
    - generally not available for every domain/topic
    - ~ 2-3 days to train (using a good GPU)

DRL for Dialogue Generation*

- model the long-term influence of a generated response in an ongoing dialogue
- define reward functions to better mimic real-life conversations
- simulate conversation between two virtual agents to explore the space of possible actions while learning to maximize expected reward

*Li, Jiwei, et al. "Deep Reinforcement Learning for Dialogue Generation." *EMNLP, 2016*.

DRL for Dialogue Generation (cont'd)

- State: concatenation of the previous two dialogue turns $[p_i, q_i]$ → Input to the encoder
- Action: dialogue utterance to generate (infinite action space)
- Policy: $p_{RL}(p_{i+1}|p_i, q_i)$; stochastic; parameters of the encoder-decoder
- Reward: ?

*Li, Jiwei, et al. "Deep Reinforcement Learning for Dialogue Generation." *EMNLP, 2016*.

DRL for Dialogue Generation (cont’d)

- State: concatenation of the previous two dialogue turns $[p_i, q_i]$ → Input to the encoder
- Action: dialogue utterance to generate (infinite action space)
- Policy: $p_{RL}(p_{i+1}|p_i, q_i)$; stochastic; parameters of the encoder-decoder
- Reward: Easy to answer, non-repetitive, semantic coherence

*Li, Jiwei, et al. "Deep Reinforcement Learning for Dialogue Generation." *EMNLP, 2016*.

Reward #1: Ease of Answering

Ease of answering = - (likelihood of dull response)

$\mathbb{S}$ = {“I don’t know. I’m not sure”, …}

$$r_1 = -\frac{1}{N_{\mathbb{S}}} \sum_{s \in \mathbb{S}} \frac{1}{N_s} \log p_{\text{seq2seq}}(s|a)$$

$N_{\mathbb{S}}$ = Cardinality of $\mathbb{S}$

N_s = length of dull response s

p_{seq2seq} = likelihood given by the SEQ2SEQ model

*Li, Jiwei, et al. "Deep Reinforcement Learning for Dialogue Generation." *EMNLP, 2016*.

Reward #2: Information Flow

- High information flow = avoid repetitive/similar responses

$$r_2 = -\log \cos(h_{p_i}, h_{p_{i+1}}) = -\log \cos \frac{h_{p_i} \cdot h_{p_{i+1}}}{||h_{p_i}|| \, ||h_{p_{i+1}}||}$$

h_p = encoder representation of utterance p

*Li, Jiwei, et al. "Deep Reinforcement Learning for Dialogue Generation." *EMNLP, 2016*.

Reward #3: Semantic Coherence

- High semantic coherence = high mutual information between two consecutive answers

$$r_3 = \frac{1}{N_a} \log p_{\text{seq2seq}}(a|q_i, p_i) + \frac{1}{N_{q_i}} \log p^{\text{backward}}_{\text{seq2seq}}(q_i|a)$$

*Li, Jiwei, et al. "Deep Reinforcement Learning for Dialogue Generation." *EMNLP, 2016*.

Total Reward

$$r(a, [p_i, q_i]) = \lambda_1 r_1 + \lambda_2 r_2 + \lambda_3 r_3$$

$$\text{where } \lambda_1 + \lambda_2 + \lambda_3 = 1$$

Overall strategy:

- Pre-train SEQ2SEQ with MLE objective
- Let two virtual agents talk to each other and optimize the policy $p_{RL}(p_{i+1}|p_i, q_i)$ by maximizing the expected reward (use policy gradient methods)

*Li, Jiwei, et al. "Deep Reinforcement Learning for Dialogue Generation." *EMNLP, 2016*.

Summary

- Retrieval-based Dialogue Systems:
    - Traditional ML: Supervised Learning
    - NN-based: SL followed by RL
- Generative Dialogue Systems:
    - NN-based: SL followed by RL
- Active research areas:
    - RL-based Transfer Learning for Dialogue Systems
    - RL-based Emotional Dialogue Systems

ProNav is Hiring

- www.pronavigator.ai
- Software Engineers
- NLP engineers
- Data Scientists

Email: nasghar@uwaterloo.ca