Deep Active Learning for Dialogue Generation

Nabiha Asghar

6th Joint Conference on Lexical and Computational Semantics (*SEM 2017)

Joint work with:
Pascal Poupart, Xin Jiang (Huawei) and Hang Li (Huawei)
2016
The Year of Chatbots for Businesses

- **JAN**: WhatsApp launches plans to allow businesses on their network
- **MAR**: Microsoft launches Bot Framework
- **APR**: Kik launches Bot Shop
- **SEP**: Facebook launches Messenger Platform for bots
- **NOV**: Oracle announces new chatbot platform
- **DEC**: Viber launches Public Accounts to chat with businesses
- **DEC**: Google launches API for Google Home

Retrieval-based vs. Generative

Diagram:

- **Context** connected to **Retrieval-based model**
- **Retrieval-based model** connected to **Response**
- **User message** connected to **Retrieval-based model**
- **Responses** connected to **Retrieval-based model**

- **User message** connected to **Generative model**
- **Previous messages** connected to **Generative model**
- **Generative model** connected to **Response**

Source: https://hackernoon.com/chatbot-architecture-496f5bf820ed
Premature Rise of Chatbots

Facebook Inc's Chatbots Hit a 70% Failure Rate (The Motley Fool)

The problem with chatbots: How to make them more human? (Chatbotsmagazine.com)

Bots are stumbling (McKinsey&Company)
Generative Models: Seq2Seq (Google)

Source: https://indico.io/blog/sequence-modeling-neuralnets-part1/
Generative Models (Seq2Seq)

me: Hello?
**bot**: Hi.
* me: How are you?
**bot**: I'm fine.
* me: What's your name?
**bot**: It's hard to describe.
* me: How so?
**bot**: I'm not sure.
* me: Are you intelligent?
**bot**: Yes, well...
* me: What are you?
**bot**: I'm not sure.
* me: Do you plan on taking over the world?
**bot**: No, i don't.

- Short, dull & generic responses
- Cannot keep the users engaged
- Infinite loops of repetition
- Mutually inconsistent responses
Idea: Reinforcement/Active Learning

Goal: Incorporate user feedback into the conversation: Human In The Loop

Research question: how do we construct a comprehensive and effective feedback mechanism for generative dialogue models?
Existing DRL Models for CAs

- Task-oriented CAs: reward function is often obvious
- Open-domain CAs: hand-crafted reward functions (ease of answering, informativity, diversity, etc)
Existing DRL Models for CAs

- **Task-oriented CAs:** reward function is often obvious
- **Open-domain CAs:** hand-crafted reward functions (ease of answering, informativity, diversity, etc)
- **Our work:**
  - Builds on top of a trained Seq2Seq model
  - A simple, all-encompassing, user-centric feedback mechanism
  - Online, incremental learning at each turn in the conversation
  - No need to hand-craft the reward function
Training Pipeline

Offline

Supervised Learning (Phase 1):
MOVIE TRANSCRIPTS

Supervised Learning (Phase 2):
HUMAN-BOT CHATLOGS

Active Learning

Online

Response Generation + Ranking

Feedback + Reply
Human user: I don't want to work today.
Human user: I don't want to work today.
bot: Ok.
bot: I wish i could robot.
bot: Why not?
bot: Aww: ).
bot: No, you've just got to be quick.
Example Transcript

Human user: I don't want to work today.
bot: Ok.
bot: I wish i could robot.
bot: Why not?
bot: Aww: ).
bot: No, you've just got to be quick.
human feedback: 3
Human user: I don't want to work today.
bot: Ok.
bot: I wish i could robot.
bot: Why not?
bot: Aww: ).
bot: No, you've just got to be quick.
human feedback: 3
Human user: I just want to lie in my bed all day and have pizza.
Example Transcript

Human user: I don't want to work today.
bot: Ok.
bot: I wish i could robot.
bot: Why not?
bot: Aww:).
bot: No, you've just got to be quick.
human feedback: 3
Human user: I just want to lie in my bed all day and have pizza.
bot: I can't do that again!
bot: Does he _ _ _.
bot: Good idea.
bot: Ok.
bot: Why not?
Human user: I don't want to work today.
bot: Ok.
bot: I wish i could robot.
bot: Why not?
bot: Aww: ).
bot: No, you've just got to be quick.
human feedback: 3
Human user: I just want to lie in my bed all day and have pizza.
bot: I can't do that again!
bot: Does he _ _ _.
bot: Good idea.
bot: Ok.
bot: Why not?
human feedback: Good idea. May I join you?
Example Transcript

Human user: I don't want to work today.
bot: Ok.
bot: I wish I could robot.
bot: Why not?
bot: Aww: ).
bot: No, you've just got to be quick.
human feedback: 3
Human user: I just want to lie in my bed all day and have pizza.
bot: I can't do that again!
bot: Does he _ _ _.
bot: Good idea.
bot: Ok.
bot: Why not?
human feedback: Good idea. May I join you?
Human user: I don't want to work today.
bot: Ok.
bot: I wish i could robot.
bot: Why not?
bot: Aww: ).
bot: No, you've just got to be quick.
human feedback: 3
Human user: I just want to lie in my bed all day and have pizza.
bot: I can't do that again!
bot: Does he _ _ _.
bot: Good idea.
bot: Ok.
bot: Why not?
human feedback: Good idea. May I join you?

Human user: I don't want to work today.
bot: Why not?
bot: Aww: ).
bot: Ok?
bot: I wish i could robot.
bot: No, you've just got to be quick.

Example Transcript

Human user: I don't want to work today.
bot: Ok.
bot: I wish i could robot.
bot: Why not?
bot: Aww: ).
bot: No, you've just got to be quick.
human feedback: 3
Human user: I just want to lie in my bed all day and have pizza.
bot: I can't do that again!
bot: Does he _ _ _.
bot: Good idea.
bot: Ok.
bot: Why not?
human feedback: Good idea. May I join you?

Human user: I don't want to work today.
bot: Why not?
bot: Aww: ).
bot: Ok?
bot: I wish i could robot.
bot: No, you've just got to be quick.
human feedback:
Human user: I just want to lie in my bed all day and have pizza.
Human user: I don't want to work today.
b: Ok.
b: I wish I could robot.
b: Why not?
b: Aww:).
b: No, you've just got to be quick.

human feedback: 3

Human user: I just want to lie in my bed all day and have pizza.
b: I can't do that again!
b: Does he __ __.
b: Good idea.
b: Ok.
b: Why not?

human feedback: Good idea. May I join you?

Human user: I don't want to work today.
b: Why not?
b: Aww:).
b: Ok?
b: I wish I could robot.
b: No, you've just got to be quick.

human feedback:

Human user: I just want to lie in my bed all day and have pizza.
b: Good idea. May I join you?
b: Does it has been to the?
b: I can't do that again!
b: So good.
b: Where good?
Example Transcript

Human user: I don't want to work today.
bot: Ok.
bot: I wish i could robot.
bot: Why not?
bot: Aww: ).
bot: No, you've just got to be quick.
human feedback: 3
Human user: I just want to lie in my bed all day and have pizza.
bot: I can't do that again!
bot: Does he _ _ _.
bot: Good idea.
bot: Ok.
bot: Why not?
human feedback: Good idea. May I join you?

Human user: I don't want to work today.
bot: Why not?
bot: Aww: ).
bot: Ok?
bot: I wish i could robot.
bot: No, you've just got to be quick.
human feedback:
Human user: I just want to lie in my bed all day and have pizza.
bot: Good idea. May I join you?
bot: Does it has been to the?
bot: I can’t do that again!
bot: So good.
bot: Where good?
human feedback:
Human user: Haha! Which pizza toppings are your favourite?
Active Learning at each step

For the message-response pair \((x,y)\), use cross-entropy loss to maximize the likelihood of \(y\) given \(x\).
Active Learning at each step

For the message-response pair \((x,y)\), use cross-entropy loss to maximize the likelihood of \(y\) given \(x\).

Use optimizer’s \textit{learning rate} to determine how quickly you want the bot to learn.
Active Learning at each step

For the message-response pair \((x,y)\), use cross-entropy loss to maximize the likelihood of \(y\) given \(x\).

Use optimizer’s **learning rate** to determine how quickly you want the bot to learn.

Feedback is optional (i.e. model/policy does not change).
Active Learning at each step

For the message-response pair \((x, y)\), use cross-entropy loss to maximize the likelihood of \(y\) given \(x\).

Use optimizer’s *learning rate* to determine how quickly you want the bot to learn.

Feedback is optional (i.e. model/policy does not change).

Can learn from multiple users simultaneously.
Active Learning at each step

For the message-response pair $(x,y)$, use cross-entropy loss to maximize the likelihood of $y$ given $x$.

Use optimizer’s *learning rate* to determine how quickly you want the bot to learn.

Feedback is optional (i.e. model/policy does not change).

Can learn from multiple users simultaneously.

**Goal:** generate interesting, relevant and diverse responses.
Training Pipeline

**Offline**
- Supervised Learning (Phase 1):
  - MOVIE TRANSCRIPTS
- Supervised Learning (Phase 2):
  - HUMAN-BOT CHATLOGS

**Online**
- Response Generation + Ranking
- Feedback + Reply
Heuristic Response Generation
Heuristic Response Generation

BEAM SEARCH
Heuristic Response Generation

BEAM SEARCH

$w^0$
Heuristic Response Generation

BEAM SEARCH

$w^0$

$w^1_a$

$w^1_b$

$w^1_c$

$w^1_d$
Heuristic Response Generation

BEAM SEARCH

\[ w^0 \]

\[ w^1_a \quad w^1_b \quad w^1_c \quad w^1_d \]
Heuristic Response Generation

BEAM SEARCH
Heuristic Response Generation

BEAM SEARCH
Heuristic Response Generation

I don’t care! vs. I don’t care.
Heuristic Response Generation

BEAM SEARCH

DIVERSE BEAM SEARCH (ICLR 2017)

I don’t care! vs. I don’t care.
Heuristic Response Generation

BEAM SEARCH

DIVERSE BEAM SEARCH (ICLR 2017)

I don’t care! vs. I don’t care.
Heuristic Response Generation

BEAM SEARCH

DIVERSE BEAM SEARCH (ICLR 2017)

I don’t care! vs. I don’t care.
Heuristic Response Generation

BEAM SEARCH

DIVERSE BEAM SEARCH (ICLR 2017)

I don’t care! vs. I don’t care.

I don’t care! vs. Who cares.
Training Pipeline
Experimental Setup

BLEU, NIST, ROUGE, WER: not suitable for dialogue quality evaluation

User study
User Study

Online Training: one human trained the model with 200 prompts of his choice
User Study

**Online Training:** one human trained the model with 200 prompts of his choice

**Test set of prompts:** We chose 100 of those, linguistically rephrased them

“How’s it going” $\rightarrow$ “How are you today?”

“I hate you.” $\rightarrow$ “I really don’t like you!”
User Study

Online Training: one human trained the model with 200 prompts of his choice

Test set of prompts: We chose 100 of those, linguistically rephrased them

“How’s it going” → “How are you doing?”

“I hate you.” → “I don’t like you!”

Test pairs: Collected responses of SL1, SL2 and SL2+oAL to test prompts
User Study

Online Training: one human trained the model with 200 prompts of his choice

Test set of prompts: We chose 100 of those, linguistically rephrased them

“How’s it going” → “How are you doing?”

“I hate you.” → “I don’t like you!”

Test pairs: Collected responses of SL1, SL2 and SL2+oAL to test prompts

Evaluation: Asked 5 human judges to rate the test pairs on 4 axes:

Syntactic Coherence  Relevance to Prompt  Interesting-ness  User Engagement
Experiments
Experiments
Experiments
# Experiments

<table>
<thead>
<tr>
<th>Human Prompt</th>
<th>SL1 Response</th>
<th>SL2 Response</th>
<th>SL2+oAL Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hello my friend.</td>
<td>No.</td>
<td>How old are you?</td>
<td>Hehe, yes, beautiful.</td>
</tr>
<tr>
<td>I think bots are sweeter than humans.</td>
<td>We’ll have to pick up a copy.</td>
<td>You are a good person.</td>
<td>Everyone is same on a basic chemical level.</td>
</tr>
<tr>
<td>I don’t know.</td>
<td>What’s the difference?</td>
<td>About?</td>
<td>Why did you suggest it?</td>
</tr>
<tr>
<td>Why did Donald Trump win?</td>
<td>He just likes it up!</td>
<td>You are!</td>
<td>Wouldn’t you like to know, fickle human?</td>
</tr>
<tr>
<td>Tell me something I don’t know.</td>
<td>I’ll tell you, it is.</td>
<td>Ok, so how are you.</td>
<td>If you were intelligent you would already know.</td>
</tr>
<tr>
<td>You’re not making any sense at all.</td>
<td>I just want to tell you something.</td>
<td>No, I’m thinking.</td>
<td>On the contrary, I am trying to educate you.</td>
</tr>
<tr>
<td>Do you have any kids?</td>
<td>No.</td>
<td>I have no!</td>
<td>None that are really close to me.</td>
</tr>
<tr>
<td>Life can be very cruel sometimes.</td>
<td>That I’ve never been to it.</td>
<td>It takes two to know two.</td>
<td>It takes two to know two.</td>
</tr>
<tr>
<td>Do you believe in life after death?</td>
<td>No.</td>
<td>Do you want the long answer?</td>
<td>Yes, do you?</td>
</tr>
<tr>
<td>You’re an idiot!</td>
<td>I know, but...</td>
<td>Yes I do.</td>
<td>Aren’t you polite.</td>
</tr>
</tbody>
</table>
# Experiments

<table>
<thead>
<tr>
<th>Human Prompt</th>
<th>Cheerful CA</th>
<th>Gloomy CA</th>
<th>Rude/Sarcastic CA</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am very happy today.</td>
<td>Life is amazing, right?</td>
<td>That makes one of us.</td>
<td>You want a piece of me?</td>
</tr>
<tr>
<td>Repeat after me: I am dumb.</td>
<td>Sweet!</td>
<td>You are right, I am.</td>
<td>You suck.</td>
</tr>
<tr>
<td>Comfort me.</td>
<td>There there, feel better.</td>
<td>All who compose must one day decompose.</td>
<td>Boo hoo.</td>
</tr>
<tr>
<td>What’s up?</td>
<td>It is a fine morning.</td>
<td>Not well, to be honest.</td>
<td>The date I went back in time to &amp; killed your parents.</td>
</tr>
<tr>
<td>I have to go now.</td>
<td>Have a good night.</td>
<td>Please don’t go.</td>
<td>Yeah leave me alone.</td>
</tr>
</tbody>
</table>
AI Assistant for Insurance
AI Assistant for Insurance

Automated Psychotherapy
AI Assistant for Insurance

Automated Psychotherapy

Chatbot for Employee Onboarding
AI Assistant for Insurance

Automated Psychotherapy

Chatbot for Employee Onboarding

Platform for Human-in-the-loop Dialogue Training
Ongoing/Future Work

Based on confidence values, ask for feedback **only when necessary**

Devise new loss functions that measure and incorporate ‘human affect’ into the responses