Zero-shot Dialog Generation with Cross-Domain Latent Actions

Tiancheng Zhao and Maxine Eskenazi

Language Technologies Institute, Carnegie Mellon University
E2E Dialog Response Generation

- Integration with database by treating DB as a part of the environment (Zhao et al 2016).

**Context**

- Next bus to CMU.
- Where are you leaving from?
- I am now at the Airport.

**Generated Response**

- The next 28x is leaving from Airport at 11:30 am.
- API: location=Pittsburgh, Time=Now
Problem: Data Scarcity & Poor Generalization

- GEDMs require **LARGE** training data
- **Impractical** since data are often NOT available:
  - Booking, recommendation, entertainment etc
- **Goal:**
  - Exploit GEDMs flexibility and let one model simultaneously learn many domains. **(Multi-task)**
  - Transfer knowledge from related domains with data to new domains without data. **(Zero-shot)**

**Example:** a customer service agent in the **shoe department** can begin to work in the **clothing department** after reading training materials, without the need for example dialogs.
Define Zero-shot Dialog Generation (ZSDG)

- Source domains: $D_{\text{source}}$ is a set of dialog domain with dialog training data.
- Target domains: $D_{\text{target}}$ is a set of dialog domains without data.
- Domain description: $\phi(d)$ captures domain-specific information about $d$
- Context is $c$ and response is $x$

Train Data: $\{c, x, d\} \sim p_{\text{source}}(c, x, d)$

$\{\phi(d)\}, d \in D$

Test Data: $\{c, x, d\} \sim p_{\text{target}}(c, x, d)$

Goal: $\mathcal{F} : C \times D \rightarrow X$
Seed Response (SR) as Domain Description

- Define SR(d) as a set of tuples
  - Each tuple contains utterances with annotations for a domain: \{x, a, d\}_{\text{seed}}
  - x is an example utterance, a is annotation, d is domain index.

- **Assumption**: Shared state tracking & policy \(--\rightarrow\) domain-specific NLU & NLG

<table>
<thead>
<tr>
<th>x</th>
<th>a</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>x = the weather in New York is raining</td>
<td>[Inform, location=New York, weather_type=Rain]</td>
<td>weather</td>
</tr>
<tr>
<td>x=what’s the location?</td>
<td>[request location]</td>
<td>weather</td>
</tr>
</tbody>
</table>
Action Matching Algorithm

- **R**: encode utterances/annotations into latent actions
  - \[ z^d_x = R(x, d) \]
  - \[ z^d_a = R(a, d) \]
- **F^e**: predict latent action given the context
  - \[ z^d_c = F^e(c, d) \]
- **F^d**: generates the response from latent action
  - \[ x = F^d(z) \]
Objective 1: $z^{d_1}_x \approx z^{d_2}_x$ when $z^{d_1}_a \approx z^{d_2}_a$

What's your departure place?

May I ask your favorite type of movie?

$L_{dd}$ domain description loss

Request (location)  Request (movie_type)
Objective 2: $z^d_c \approx z^d_x$ for all source domains
(potentially for target domain as well)
Optimization by Alternating these 2 losses

- \( L_{dd}(F^d, R) = -\log p_{F^d}(x|R(a, d)) + \lambda D[R(x, d) \parallel R(a, d)] \)

- \( L_{dialog}(F^e, F^d, R) = -\log p_{F^d}(x|F^e(c, d)) + \lambda D[R(x, d) \parallel F^e(c, d)] \)

**Algorithm 1: Action Matching Training**

Initialize weights of \( F^e, F^d, R \);
Data = \{c, x, d\} \bigcup \{x, a, d\}_{seed}

while batch \sim Data do
    if batch in the form \{c, x, d\} then
        Backpropagate loss \( L_{dialog} \)
    else
        Backpropagate loss \( L_{dd} \)
    end
end
Implementation

- Recognition Network $R$: Bidirectional GRU
- Encoder $F_e$: Hierarchical Recurrent LSTM Encoder (HRE) [Li et al 2015]
- Decoder $F_d$:
  - LSTM Attention decoder
    - Attention over every words in the context
    - Standard baseline.
  - LSTM Pointer-sentinel Mixture (PSM) Decoder (Copy mechanism) [Merity et al 2016]
    - Can copy any words from the context
    - Proven to show good performance in generating OOV tokens.
Implementation with PSM decoder

(a) Dialog Data

Pointer Attention
Discourse LSTM
Utterance GRU

(1-g) P_{ptr} + g P_{vocab}

Pointer Attention
Utterance GRU

(b) Domain Description Data

Vocab Softmax

Utterance GRU

D(z_{x_i}, z_{c})

x_{T-1}

D(z_{x_i}, z_{a})

x_{T-1}

Data

1. CMU SimDial: simulated dataset

2. Stanford Multi-domain Dialog (SMD) Dataset: Human-Woz dataset
CMU SimDial

- A open-source multi-domain dialog generator with complexity control.
- Source Domains (900 training, 100 validation dialogs for each domain):
  - Restaurant, Bus, Weather
- Target Domains (500 testing dialogs for each domain)
  - Restaurant (in-domain)
  - Restaurant-slot (unseen slot): introduce new slot values
  - Restaurant-style (unseen NLG): same slot values but different NLG templates
  - Movie (new-domain): completely new domains
- Seed Response (SR):
  - 100 unique random utterances from each domain, annotations are semantic frames used by the simulator.
  - I believe you said Boston. Where are you going?” → [implicit_confirm location=Boston; request location]
Stanford Multi-domain Dialog (SMD)

- 3031 human-Woz data about 3 domains [Eric and Manning 2017]
  - Schedule, Navigation, Weather
- Leave-one-out to rotate among each domain as the target domain.
- Random sample 150 unique utterances from each domain as SR
- An expert annotated the 150 utterances in SR (available online)
  - All right, I’ve set your next dentist appointment for 10am. Anything else? → [ack; inform goal event=dentist appointment time=10am ; request needs].
- All the target data that we need is the 150 utterances with annotations - No large dialog corpus is needed!
Metrics and Compared Models

1. **BLEU-4**: corpus-level BLUE-4 between the generated responses and references.

2. **Entity F1**: checks if the generated responses contains the correct entities (slot values)

3. **Act F1**: checks if the generated responses exhibits the correct dialog acts (using a classifier)

4. **KB F1**: check if the generated API call has all correct command tokens.

5. **BEAK**: geometric mean of the above 4 scores.
   \[
   \text{BEAK} = (\text{bleu} \times \text{ent} \times \text{act} \times \text{kb})^{\frac{1}{4}}
   \]
   a. **BE (for SMD)**: \(\text{BE} = (\text{bleu} \times \text{ent})^{\frac{1}{2}}\)

Four models are compared:

1. HRE + Attention Decoder (+Attn)
2. HRE + PSM Decoder (+Copy)
3. HRE + Attention Decoder + AM training (+Attn+AM)
4. HRE + PSM Decoder + AM training (+Copy+AM)
1. What fails when testing on new domain?
2. What problem does Copy solve?
3. What problem does AM solve?
4. How does the size of SR affect AM’s performance?
What Fails on New Domains?

First analyze dialog acts:

**Good Examples:**
- Ref: See you.
- **Generated (Attn):** See you next time

**Bad Examples:**
- Ref: Hi I am your movie bot. What can I do for you?
- **Generated (Attn):** Hi this is the restaurant system. How can I help?
- Ref: Sci-fi movie. What time’s movie?
- **Generated (Attn or Copy):** Pittsburgh. what kind of restaurant are you looking for?

**Answer:** fail to generate the correct entity as well as the correct overall sentence. Dialog acts are okay.
What Problem Does Copy Solve?

**Answer:** Copy Network improves entity score significantly, especially when there are OOV entity.

**Examples:**
- **Ref:** Do you mean sci-fi?
- **Generated (Attn):** Do you mean pizza?
- **Generated (Copy):** Do you mean sci-fi?

**Bad Examples:**
- **Ref:** Movie 55 is a good choice.
- **Generated (Copy):** I would recommend restaurant 55.
- **Ref:** I believe you said comedy movie.
- **Generated (Copy):** I believe you said comedy food.
What Problem Does AM Solve?

**Answer:** AM enables the decoder to generate overall novel utterances, not just entities.

**Examples from SimDial:**
- **Ref:** Movie 55 is a good choice.
- **Generated (Copy+AM):** Movie 55 is a good choice

**Examples from SMD:**
- **Ref:** Okay, scheduling Friday dinner with mom at 11 am
- **Generated (Copy+AM):** scheduling a reminder for dinner on Friday with your 11AM at 10 am
Impact of Seed Response (SR) Size

- Investigate how the size of SR affects the performance of AM algorithm.
- Vary the size of SR from 0 to 200 in the SMD data.
- Use schedule as the target domain.
Contributions

● Propose ZSDG, a new challenge for generative dialog systems.
● Propose AM algorithm with seed responses for solving ZSDG under the assumption that there exists a shared discourse-level pattern.
● Validate AM’s effectiveness extensively on both synthetic & real dataset.
● Open-source SimDial, a multi-domain dialog generator that can be used to benchmark ZSDG.
Future Work

- How to quantify the relationship among domains? In what situations, ZSDG is possible?
- How can we reply less on annotation to find alignments between responses in source & target domains?
- How can we solve ZSDG when the discourse-level patterns are significantly different from the source to the target?
- What are other potential domain description formats?
Thank you!

Code & Data: github.com/snakeztc/NeuralDialog-ZSDG
More on SimDial

- **Environmental**: the complexity level of the communication channel, e.g. automatic speech recognition (ASR) error rate.
- **Propositional**: the complexity level in the propositional content of user utterances, e.g. the chance that a user will give multiple slot values in one utterance.
- **Interaction**: the complexity level in terms of real-time interaction, e.g. the percentage of self-repair.
Example Dialog

**Domain: Restaurant**

Sys: Welcome to the restaurant recommendation system. What do you need?
User (0.83): Hello robot. I am looking for a restaurant.
Sys: Which city are you interested in?
User (0.75): I am in San Jose. And Korean. Uhm sorry, I am interested in Chinese food.
Sys: I believe you said San Jose. I believe you said Chinese.
Sys: QUERY goal: restaurant food-pref: Chinese location: San Jose
KB: RET restaurant: 17
Sys: Restaurant 17 is a good choice. What else can I do for you?
User (0.6): Not done yet. Is that restaurant open now?
Sys: QUERY goal: open restaurant: 17 food-pref: Chinese location: San Jose
KB: RET open: yes
Sys: Yes, it is open right now. What else?
User (0.69): No more questions. Thanks.
Sys: Bye.

**Domain: Movie**

Sys: Want to know about movies? Please ask me!
User (0.43): Hello robot. Give me some good suggestions about movies.
Sys: Which type of movie?
User (0.95): I like Documentary movies. Uhm sorry, Drama.
Sys: I believe you said Drama. Movie in what years?
User (0.64): In 70s.
Sys: QUERY goal: movies type: Drama years: 70s
KB: RET movie: 62
Sys: Movie 62 is a good choice. Are you happy about my answer?
User (0.74): All good. Thank you.
Sys: See you next time.
Relationship with Latent Variable Dialog Models


- \[ L_{\text{dialog}}(F^e, F^d, R) = - \log p_{F_d}(x|F^e(c, d)) + D[R(x, d) \parallel F^e(c, d)] \]

- \[ L_{\text{CVAE}}(F^e, F^d, R) = - \log p_{F_d}(x|F^e(c, d)) + KL[R(x, d) \parallel F^e(c, d)] \]

- Difference:
  - We employ a general distance function \( D \) (we use MSE for this study)
  - Latent variables use KL-divergence, which is a special case.
More on Seed Response (SR)

- Assumption: there is a shared discourse-level pattern between the source & target domains.
  - Shared state tracking & policy \(\leftrightarrow\) domain-specific NLU & NLG.
  - Hold true for many slot filling domains

- A GEDM can generalize to target domains without data if it can learn a mapping function where responses with similar functions from different domains are close-by.

- Denote the output of this mapping function as \(z \rightarrow \text{Cross-domain Latent Action}\)
## Seed Response Examples (SMD)

<table>
<thead>
<tr>
<th>X</th>
<th>a</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting GPS to Safeway at 452 Arcadia Pl.</td>
<td>request #goal #poi Safeway #location 452 Arcadia Pl</td>
<td>navigation</td>
</tr>
<tr>
<td>Pizza My Heart is at 528 Anton Ct, directions are on the screen.</td>
<td>#poi Pizza My Heart #location 528 Anton Ct ; show</td>
<td>navigation</td>
</tr>
<tr>
<td>Okay which one, I have two. One is the 8th with Jeff and the other is on the same day with Martha.</td>
<td>request #choice #count 2 ; inform #date 8th #party Jeff ; inform #date 8th #party Martha</td>
<td>scheduling</td>
</tr>
<tr>
<td>Your next lab appointment is on Friday with Tom.</td>
<td>inform #event lab appointment #date Friday #party Tom</td>
<td>scheduling</td>
</tr>
<tr>
<td>What city would you like the weekly weather report for?</td>
<td>request #location #goal #date weekly</td>
<td>weather</td>
</tr>
<tr>
<td>what is the forecast for today and tomorrow</td>
<td>request #goal #date today and tomorrow</td>
<td>weather</td>
</tr>
</tbody>
</table>
Qualitative Analysis on New Domain

<table>
<thead>
<tr>
<th>Model/Type</th>
<th>General Utterance</th>
<th>Unseen Slots</th>
<th>Unseen Utterance</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td>See you next time</td>
<td>Do you mean romance movie</td>
<td>Movie 55 is a good choice.</td>
</tr>
<tr>
<td>+Attn</td>
<td>Goodbye</td>
<td>Do you mean Chinese food?</td>
<td>Bus 12 can take you there.</td>
</tr>
<tr>
<td>+Copy</td>
<td>See you next time</td>
<td>Do you mean romance food?</td>
<td>Bus 55 can take you there.</td>
</tr>
<tr>
<td>+Copy+AM</td>
<td>See you next time</td>
<td>Do you mean romance movie?</td>
<td>Movie 55 is a good movie.</td>
</tr>
</tbody>
</table>