Improving Generative Models with Hierarchical Plans

2/21/20 LINGO Meeting
Task

Given a dataset of demonstrations and natural language annotations, generate a hierarchical instruction tree that can be used to guide a policy.
Given a dataset of demonstrations and natural language annotations, generate a hierarchical instruction tree that can be used to guide a policy.
Related Work

- Instruction following models
- Hierarchical (reinforcement) learning
- Shaping representations with language

Research questions

- How to infer a hierarchy of subtask instructions through sparse annotations?
- Does this instruction tree improve the generative model?
- Can we learn “modules” directly from data (rather than pre-defined modules)?
- To what extent does language help the model perform zero-shot inference?
Data Collection

Can a neural network learn to recognize doodling?

Help teach it by adding your drawings to the world's largest doodling data set, shared publicly to help with machine learning research.

Let's Draw!
Draw tree in under 20 seconds
What do 50 million drawings look like?

Over 15 million players have contributed millions of drawings playing Quick, Draw! These doodles are a unique data set that can help developers train new neural networks, help researchers see patterns in how people around the world draw, and help artists create things we haven’t begun to think of. That’s why we’re open-sourcing them, for anyone to play with.
Data Collection

Figure 2: Schematic diagram of sketch-rnn.
Data Collection
Please write an instruction describing elements that are added to the second image. Elements include bodies, leg(s), eye(s), tail(s), wing(s), teeth, etc.
Data Collection

Category: elephant
Finish the trunk and start finishing the head.

Category: duck
Finish the tail.

Category: bear
Draw the inner ears and the rest of the head.

Category: bear
Add part of a nose and the other ear.

Category: bird
Add the body, a leg, the head, a beak, and an eye.

Category: hedgehog
Draw another spine.

Category: crab
Draw the arms and claws on the arms.

Category: camel
Draw two humps. Add an ear on the head.

Category: penguin
Draw the body. Add the second wing.
Data Collection

**Category: hedgehog**
- Add spikes to back

**Category: swan**
- Add a leg.

**Category: horse**
- Add a tail.

**Category: tiger**
-?

**Category: frog**
- Add the body.

**Category: cow**
- Add part of a body and two legs.

**Category: raccoon**
-?

**Category: parrot**
- Start drawing the body.

**Category: parrot**
- Add body.
MODELING PART 1: Generating Instruction Trees

MODELING PART 2: Improving Sketch Generation
MODELING PART 1: Generating Instruction Trees

Instruction generation model

\[ P(I | S_{i:j}, S_{i}, S, \text{category}) \]

<table>
<thead>
<tr>
<th>Drawing</th>
<th>Model Notes</th>
<th>BLEU1</th>
<th>BLEU2</th>
<th>ROUGEL</th>
<th>Unique tokens gen on test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke sequence</td>
<td>Basic</td>
<td>0.4280</td>
<td>0.2148</td>
<td>0.3849</td>
<td>61</td>
</tr>
<tr>
<td>Images</td>
<td>Basic</td>
<td>0.4542</td>
<td>0.2401</td>
<td>0.4049</td>
<td>53</td>
</tr>
<tr>
<td>+ Memory</td>
<td></td>
<td>0.4646</td>
<td>0.2600</td>
<td>0.4167</td>
<td>69</td>
</tr>
</tbody>
</table>
MODELING PART 1: Generating Instruction Trees

Generated: draw the body and the head. add two ears.
Ground truth: draw the head, and add two ears and two eyes.

Generated: add the mouth and an eye.
Ground truth: add eye, nose and mouth.

Generated: add an eye.
Ground truth: add a spot to the butt.

Generated: add two legs.
Ground truth: add three visible legs.
MODELING PART 1: Generating Instruction Trees

\[
\max_i \left\{ P(I_1|S_{i:})^\alpha \cdot P(I_2|S_{i:})^\alpha \right\}
\]

Draw a head and two ears  \hspace{1cm} Draw a nose, mouth, and eyes

\[
\ldots
\]
MODELING PART 1: Generating Instruction Trees

\[
\max_{i} \left\{ P(I_1 | S_{:i})^\alpha \cdot P(I_2 | S_{:i})^\alpha \right\}
\]
MODELING PART 1: Generating Instruction Trees

1. draw the head
2. add an eye
3. add a second eye
4. add whiskers
5. add ears
6. add the mouth
7. add the second ear
8. draw the second eye

0-8: draw the head, two eyes, and the mouth.
0-7: draw the head, ears, and whiskers.
0-1: draw the head and two ears.
7-8: add an eye.
MODELING PART 1: Generating Instruction Trees

Flamingo

1

2

3

4

5

0-5: draw the body, neck, head, and beak.

0-2: draw the body, neck, head, and beak.

1-2: draw the body and the beak.

2-5: add a leg and foot.

2-4: add a leg and foot.

2-3: add a leg.
MODELING PART 2: Improving Sketch Generation

SketchRNN

+ instruction trees

\[ \sum_{i=1}^{N} \left[ \| h_i^a - x_{i}^p \|_2^2 - \| h_i^a - x_{i}^n \|_2^2 + \alpha \right] \]

<table>
<thead>
<tr>
<th>Training Size</th>
<th>Type of Model</th>
<th>NLL</th>
</tr>
</thead>
<tbody>
<tr>
<td>87500 (2500 per)</td>
<td>SketchRNN</td>
<td>1.0071</td>
</tr>
<tr>
<td>+ Root Instruction</td>
<td></td>
<td>0.9553</td>
</tr>
<tr>
<td>+ Instruction Stack</td>
<td></td>
<td>0.9567</td>
</tr>
</tbody>
</table>
Modeling

- Better instruction gen (contrastive pretraining, memory)
- Better instruction trees (bottom-up, metrics, additional scoring)

Other

- Evaluating zero-shot (hold out categories)
- Using trees as ground truth for instruction (tree) gen model
- PIVOT: hierarchical instructions $\rightarrow$ language-guided memory