Interactive Language Acquisition with One-shot Visual Concept Learning through a Conversational Game

ACL 2018

Haichao Zhang Haonan Yu Wei Xu

{zhanghaichao, haonanyu, wei.xu} @ baidu.com

Baidu Research - Institute of Deep Learning

Baidu Research

Motivation

Supervised learning

- capturing mainly the statistics of training data
- less flexible for acquiring new knowledge without retraining

Human learning

- **interactive**: humans act upon the world and learn from consequences
- **one-shot**: humans have a celebrated ability to learn new concepts from small amount of data

The Conversational Game

Interactive Setting. Within a session, the teacher may ask questions, answer learner's questions, make statements, or say nothing. The teacher also provides reward feedback based on learner's responses. The game is constructed using the **XWorld** simulator package [1].

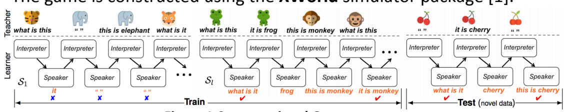
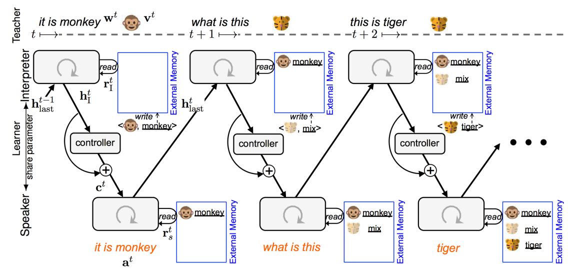


Figure 1 Conversational Game.

start	→	question | silence | statement
question	→	Q1 | Q2 | Q3
silence	→	" "
statement	→	A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8
Q1	→	*"what"*
Q2	→	*"what"* M
Q3	→	*"tell what"* N
M	→	*"is it"* | *"is this"* | *"is there"* | *"do you see"* | *"can you see"* | *"do you observe"* | *"can you observe"*
N	→	*"it is"* | *"this is"* | *"there is"* | *"you see"* | *"you can see"* | *"you observe"* | *"you can observe"*
A1	→	G
A2	→	*"it is"* G
A3	→	*"this is"* G
A4	→	*"there is"* G
A5	→	*"i see"* G
A6	→	*"i observe"* G
A7	→	*"i can see"* G
A8	→	*"i can observe"* G
G	→	object name

Table 1 Grammar for the teacher's sentences.

Network Structure



external memory: store and retrieve information

interpreter: interpret the teacher's sentences, extracting information from the perceived signals, and saving it to the external memory

speaker: generate responses with reading access to external memory

Approach

Model. Joint imitation and reinforcement learning by minimizing a joint cost:

$$\mathcal{L}_\theta = \underbrace{\mathbb{E}_{\mathcal{W}}\left[-\sum_t \log p^{\mathrm{I}}_\theta(\mathbf{w}^t|\cdot)\right]}_{\text{Imitation } \mathcal{L}^{\mathrm{I}}_\theta} + \underbrace{\mathbb{E}_{p^{\mathrm{S}}_\theta}\left[-\sum_t [\gamma]^{t-1}\cdot r^t\right]}_{\text{Reinforce } \mathcal{L}^{\mathrm{R}}_\theta}$$

A. Imitation with Memory Augmented Neural Network for Echoic Behavior

The teacher's way of speaking provides a source for the learner to mimic. One way to learn from this source of information is by predicting using, information from both language model (**h**) and external memory (**r**):

$$p^{\mathrm{I}}_\theta(\mathbf{w}^t|\mathcal{H}^{t-1},\mathbf{a}^{t-1},\mathbf{v}^t) = \prod_i p^{\mathrm{I}}_\theta(w^t_i|w^t_{1:i-1},\mathbf{h}^{t-1}_{\mathrm{last}},\mathbf{v}^t)$$

$$p^{\mathrm{I}}_\theta(w^t_i|\mathbf{h}^t_i,\mathbf{v}^t) = (1-g)\cdot p_{\mathbf{h}} + g\cdot p_{\mathbf{r}}$$

B. Context-adaptive Behavior Shaping through Reinforcement Learning

$$p^{\mathrm{S}}_\theta(\mathbf{a}^t|\mathbf{h}^t_{\mathrm{I}},\mathbf{v}^t) = p^{\mathrm{I}}_\theta(\mathbf{a}^t|\mathbf{h}^t_{\mathrm{I}} + f(\mathbf{h}^t_{\mathrm{I}},c),\mathbf{v}^t)$$

C. Joint Imitation and Reinforcement Learning

Imitation module contributes by using a cross-entropy loss and minimizing it with respect to the parameters in interpreter (shared with speaker); **Reinforcement** module adjusts the policy by maximizing expected future reward

$$\nabla_\theta \mathcal{L}^{\mathrm{R}}_\theta = \mathbb{E}_{p^{\mathrm{S}}_\theta}\left[\sum_t A^t\cdot\nabla_\theta \log p^{\mathrm{S}}_\theta(\mathbf{a}^t|\mathbf{c}^t)\right]$$

Experiments

Setting

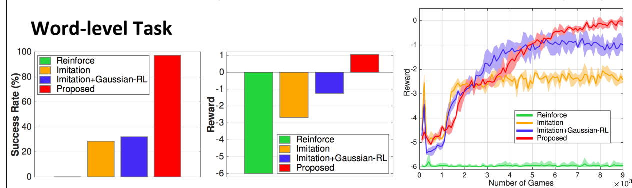
We use an Animal dataset for training and test the trained models on a Fruit dataset. Each session consists of two randomly sampled classes, and the maximum number of interaction steps is six.

Baselines

Reinforce: minimizing only $\mathcal{L}^{\mathrm{R}}_\theta$ **Imitation**: miminizing only $\mathcal{L}^{\mathrm{I}}_\theta$

Imitation+Gaussian-RL: joint imitation-reinforcement with Gaussian policy [5]

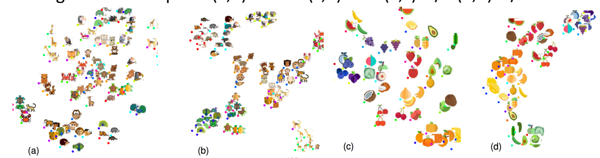
Word-level Task



Test success rate and reward

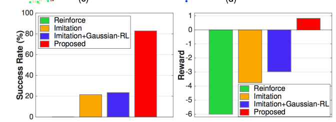
Reward curves during training

Learning with Image Variations. Clusters of visually similar concepts emerge in feature space. (a,b) animal (c,d) fruit (a,c) w/o (b,d) w/ variations

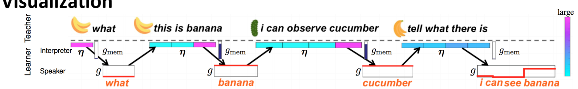


Sentence-level Task

Test performance for sentence-level task with image variations (variation ratio=0.5)



Visualization



References

[1] XWorld: https://github.com/PaddlePaddle/XWorld/

[2] PaddlePaddle: https://github.com/PaddlePaddle/Paddle

[3] J. Weston. Dialog-based language learning. In NIPS, 2016

[4] B. F. Skinner. Verbal Behavior. Copley Publishing Group, 1957.

[5] H.Zhang *et al.* Listen, interact and talk: Learning to speak via interaction. NIPS Workshop on Visually-Grounded Interaction and Language, 2017