ECCV'20 ONLINE

23-28 AUGUST 2020

16TH EUROPEAN CONFERENCE ON

COMPUTER VISION

WWW.ECCV2020.EU

AAAI2020: Visual Dialogue State Tracking for Question Generation

ECCV2020: Guessing State Tracking for Visual Dialogue

Wei Pang, Xiaojie Wang

2019.03～2019.11

School of Artificial Intelligence

Beijing University of Posts and Telecommunications, China

Learned Question Policy

- Absolute spatial location
  - ✓ Left
  - ✓ Left rear
  - ✓ Left front

is it a animal ?	yes
is it on the left ?	yes
is it in the front ?	no
on the right ?	no
is it in the middle ?	no

Success

<start> a person ?	yes
<yes> in the front ?	no
<no> on the left ?	yes
<yes> the one on the right ?	no
<no> in the middle ?	no

Success

is it a person ?	yes
on the left ?	yes
is it in the front ?	no
on the right ?	no
in the middle ?	no

Success

is it a person ?	yes
on the left ?	yes
is it in the front ?	no
on the right ?	no
in the middle ?	no

Success

is it a person ?	no
is it a animal ?	yes
is on the left ?	yes
is it in the front ?	no
is it in the middle ?	no

Success

Learned Question Policy

- Absolute spatial location
  - ✓ Middle
  - ✓ Middle left
  - ✓ Middle right
  - ✓ Middle front

is it a person ?	yes
on the left ?	no
in the middle ?	yes
on the right ?	no
in the front ?	no

Success

is it a person ?	no
is it the first ?	no
on the left ?	no
in the middle ?	yes
on the left ?	no

Success

is it a animal ?	yes
is it on the left ?	no
in the middle ?	yes
is on the right ?	no
in the middle ?	yes

Success

is it a person ?	no
is it a bus ?	yes
on the left ?	no
in the middle ?	yes
is it in the front ?	no

Success

is it a person ?	yes
on the left ?	no
in the middle ?	yes
on the right ?	no
in the front ?	no

Success

is it a bird ?	yes
is on the left ?	no
in the middle ?	yes
on the right ?	no
in the middle ?	yes

Success

is it a person ?	yes
one on the left ?	no
in the middle ?	yes
on the right ?	no
in the middle ?	yes

Success

is it food ?	yes
is on the left ?	no
in the middle ?	yes
is on the right ?	no
is it in the front ?	yes

Success

is it a animal ?	yes
is it on the left ?	no
in the middle ?	yes
on the right ?	no
in the front ?	yes

Success

is it a person ?	no
is it a board ?	yes
on the right ?	no
is in the middle ?	yes
is it in the front ?	yes

Success

is it a book ?	no
is it a person ?	yes
is on the left ?	no
in the middle ?	yes
is on the right ?	yes

Success

is it a person ?	no
is it a skateboard ?	yes
on the left ?	no
in the middle ?	yes
is it on the right ?	yes

Success

is it a animal ?	yes
is it on the left ?	no
in the middle ?	yes
is on the right ?	yes
is it in the front ?	yes

Success

<start> is it computer ?	yes
<yes> is it on the left ?	yes
<yes> is it in the front ?	no
<no> on the right ?	no
<no> in the middle ?	yes
Success	

<start> is it a person ?
<yes> is it on the left ?
<yes> is it in the front ?
<no> on the right ?
<no> in the middle ?
Success

yes 0.489
yes 0.104
no 0.000
no 0.014
yes 0.392

is it a person ? no
is the bat ? no
on the left ? yes
is it in the front ? no
is it in the middle ? yes
Success

Learned Question Policy

- Absolute spatial location
  - ✓ Right
  - ✓ Right front
  - ✓ right top

is it food ?	yes
on the left ?	no
in the middle ?	no
on the right ?	yes
is it in the front ?	no

Success

is it a person ?	yes
on the left ?	no
in the middle ?	no
on the right ?	yes
is it in the front ?	no

Success

is it a person ?	no
the ball ?	no
a the racket ?	yes
on the right ?	yes
is it in the front ?	no

Success

is it a person ?	no
is it a car ?	yes
on the right ?	yes
is it in the front ?	no
is it in the middle ?	no

Success

is it a person ? no
is it the ball ? yes
the the racket ? no
on the right ? yes
is it in the front ? no
Success

is it a animal ?	yes
is on the left ?	no
in the middle ?	no
on the right ?	yes
is it in the front ?	no

Success

Learned Question Policy

- Absolute spatial location
  - ✓ Front
  - ✓ Right front
  - ✓ right top

is it food ?	yes
is on the left ?	no
in the front ?	yes
on the right ?	no
in the top ?	no

Success

is it a plant ?	no
is it a vase ?	yes
on the right ?	no
in the front ?	yes
is on the right ?	no

Success

is it food ?	no
is it a bowl ?	yes
on the right ?	no
in the front ?	yes
is on the right ?	no

Success

is it food ?	yes
is it on the left ?	no
in the middle ?	no
in the front ?	yes
on the right ?	no

Failure

Outline

- Introduction
- GuessWhat?!
- Question Generation & Guesser(2019.03-11)
- Experiments & Conclusions
- GuessWhich & Visual Dialog(VisDial)(2020～)

Guesser

- A Visual Grounding Task
  - e.g., grounding a series of language descriptions (question-answer pair) in an image gradually by dialoguing

Guesser

- First Guesser model: Classification model with supervised learning
- Encoder: Two-layer for dialogue, VGG16 for image, Concatenation
- Guess: Make a guess from the dialog history when the dialogue ended

De Vries et al., CVPR2017;
Strub et al., IJCAI2017

Guesser

- Classification model with Reinforcement Leaning (Zhao et al., IJCAI2018)
- Encoder: LSTM for each QA-pair, Memory and Attention
- Guess: taking a guess(action) from QA-pairs when the dialogue ended
- Reward: 1 or 0
- REINFORCE rule
  - without baseline

A problem in previous work

- Existing Guessers make one and only one guess after the final round
- Guesser is not activated until the dialog between QGen and Oracle ended

De Vries et al., CVPR2017;
Strub et al., IJCAI2017

Zhao et al., IJCAI2018;
Ehsan Abbasnejad et al., CVPR2019

Motivation

- Think the Guesser to be active throughout the conversation of QGen and Oracle
  - For example, it keeps active by tracking the guessing state during the dialogue, rather than just only guess at the end of the conversation
- Think the guess as a process instead of a single action at the end of the dialogue

Guessing State Tracking(GST)

- Guessing State: probability distribution on objects

distribution on objects

Tracking process

GST based Guesser

- Guessing State: probability distribution on objects

$$\pi^{(1)} \boxed{0.4552 \quad 0.0085 \quad 0.0228 \cdots 0.5129}$$

- GST based Guesser model, three modules
  - Update of Visual Representation (UoVR)
  - Question-Answer Encoder (QAEncoder)
  - Update of Guessing State (UoGS)

GST based Guesser

Update of Visual Representation (UoVR)

GST based Guesser

Question-Answer Encoder (QAEncoder)

GST based Guesser

Update of Guessing State (UoGS)

GST based Guesser

- Early Supervision Loss\Plan Supervision Loss
  - supervision of correct guess can be employed from an early stage
- Incremental Supervision Loss
  - the change of guessing state can also be supervised to ensure that the guessing is alone in the right way
  - bring monotonicity to the probability of target object in the series of GS

Success rate of guessing

	Questioner		Max Q's	New Object			New Game		
	Guesser	QGen		S	G	BS	S	G	BS
pretrained in SL	guesser[21]	qgen[21]	5	41.6	43.5	47.1	39.2	40.8	44.6
	guesser(MN)[28]	TPG[28]	8	-	48.77	-	-	-	-
	guesser[20]	qgen[20]	8	-	44.6	-	-	-	-
	GST(ours)		8	41.73	**44.89**	-	39.97	41.36	-
	guesser[20]	VDST[14]	5	45.02	49.49	-	42.92	45.94	-
			8	46.70	48.01	-	44.24	45.03	-
	GST(ours)		5	**49.55**	**53.35**	**53.17**	**46.95**	**50.58**	**50.71**
			8	**52.71**	**54.10**	**54.32**	**50.19**	**50.97**	**50.99**
SL	GDSE-SL[18]	GDSE-SL[18]	5	-	-	-	-	47.8	-
			8	-	-	-	-	49.7	-
CL	GDSE-CL[18]	GDSE-CL[18]	5	-	-	-	-	53.7	-
			8	-	-	-	-	58.4	-
AQM	guesser[11]	randQ[11]	5	-	-	-	-	42.48	-
		countQ[11]	5	-	-	-	-	61.64	-

Success rate of guessing

trained by RL	guesser(MN)[28]	TPG[28]	5	62.6	-	-	-	-	-
			8	-	-	-	-	74.3	-
		ISM[1]	-	74.4	-	-	72.1	-	-
		TPG[28]	8	-	74.3	-	-	-	-
		ISD[2]	5	68.3	69.2	-	66.3	67.1	-
	guesser[20]	VQG[27]	5	63.2	63.6	63.9	59.8	60.7	60.8
		ISM[1]	-	-	64.2	-	-	62.1	-
		ISD[2]	5	61.4	62.1	63.6	59.0	59.8	60.6
		RIG(rewards)[19]	8	65.20	63.00	63.08	64.06	59.0	60.21
		RIG(loss)[19]	8	67.19	63.19	62.57	65.79	61.18	59.79
	guesser[20]	qgen[20]	5	58.5	60.3	60.2	56.5	58.4	58.4
			8	62.8	58.2	53.9	60.8	56.3	52.0
	guesser(MN)[28]		5	59.41	60.78	60.28	56.49	58.84	58.10
			8	62.05	62.73	-	59.04	59.50	-
	GST(ours)		5	**64.78**	**67.06**	**67.01**	**61.77**	**64.13**	**64.26**
	guesser[20]	VDST[14]	5	66.22	67.07	67.81	63.85	64.36	64.44
			8	69.51	70.55	71.03	66.76	67.73	67.52
	GST(ours)		5	**77.38**	**77.30**	**77.23**	**75.11**	**75.20**	**75.13**
			8	**83.22**	**83.32**	**83.46**	**81.50**	**81.55**	**81.62**
	Human[20]	-	-	-	84.4	-	-	84.4	-

Success rate of guessing

Ablation Study

- Effect of Individual Supervision
  - Plain Supervision (PS)
  - Early Supervision (ES)
  - Incremental Supervision (IS)

Qualitative Evaluation

Thanks