Prosocial or Selfish?

Agents with different behaviors for Contract Negotiation using Reinforcement Learning

Презентацию подготовила:
Бочкарева Мария БПМИ151

Постановка задачи

Среда

$$U = \text{Shuffle}(P \bigoplus N)$$

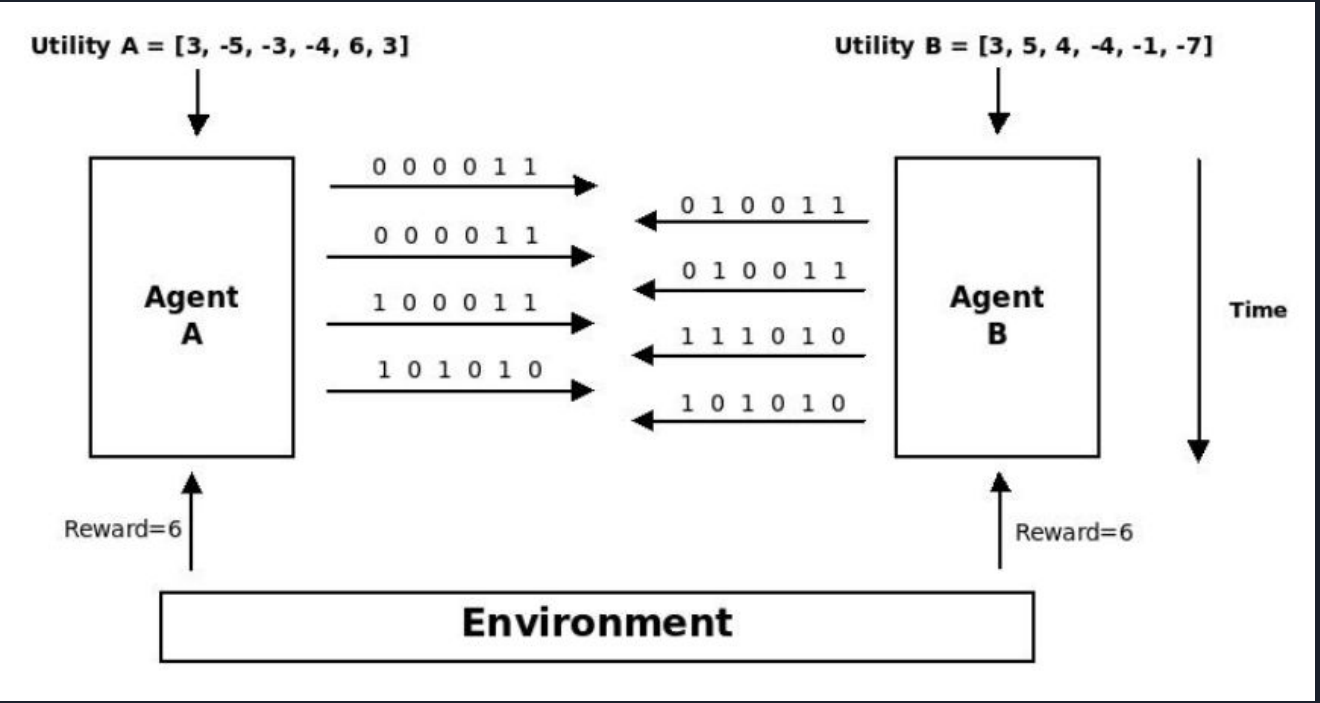
$$P = [p_1, p_2, \ldots, p_k]$$

$$N = [n_1, n_2, \ldots, n_{6-k}]$$

$$\sum_i p_i = 12$$

$$\sum_i n_i = -12$$

Среда


Utility A = [3, -5, -3, -4, 6, 3]
Utility B = [3, 5, 4, -4, -1, -7]
Agent A
Agent B
Time
0 0 0 0 1 1
0 0 0 0 1 1
1 0 0 0 1 1
1 0 1 0 1 0
0 1 0 0 1 1
0 1 0 0 1 1
1 1 1 0 1 0
1 0 1 0 1 0
Reward=6
Reward=6
Environment

Агент

Агент состоит из следующих частей:

(1) Deep Neural Network (DNN) Component

Решает, какое количество бит необходимо изменить в предложении второго агента.

(2) Rule Based Component

Решает, какие именно биты поменять, чтобы получить наибольшую выгоду.

Процедура обучения

Архитектура

(1) It's Utility function U^A.
(2) Offer given by opponent B, S_t^B.
(3) It's previous offer, S_{t-1}^A.
(4) Agent ID, $I \in \{0, 1\}$.

We convert this input into a dense representation D_t^A as

$$D_t^A = [\text{OfferMLP}([U^A, S_t^B]), \text{OfferMLP}([U^A, S_{t-1}^A]), \text{AgentLookup}(I), \text{TurnLookup}(t)].$$

Архитектура

$$h_t^A = \mathrm{GRU}(D_t^A, h_{t-1}^A)$$

$$\pi_A = \mathrm{Softmax}(W h_t^A)$$

$$L_i = \mathop{\mathbb{E}}_{x_t \sim (\pi_A, \pi_B)} \left[\sum_t \gamma^{(T-t)} (r_i(x_{1 \ldots T}) - b_i) \right] + \lambda H[\pi_i]$$

Агенты против агентов

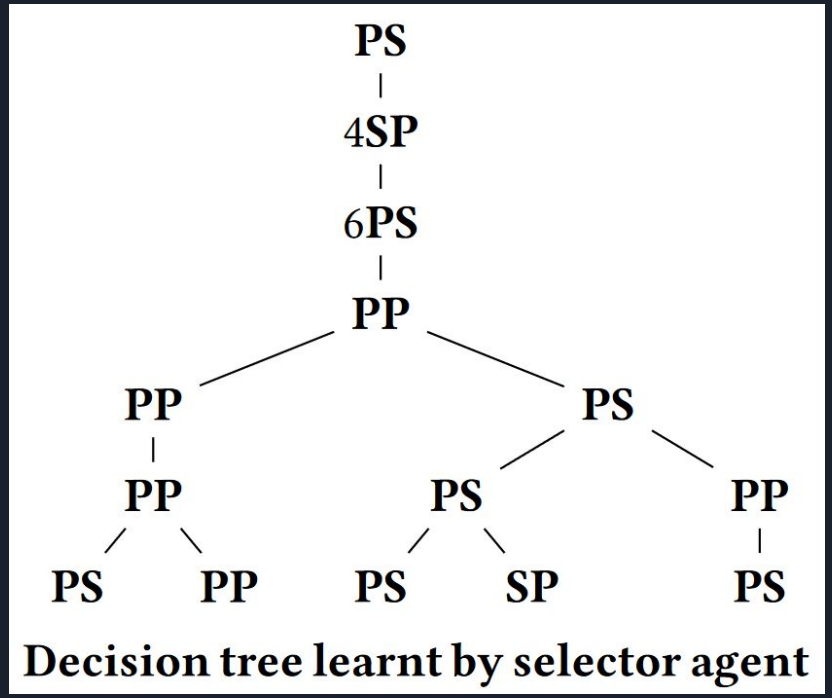
Agent A	Agent B	Dialog Length	Agreement Rate (%)	Optimality Rate (%)	Average Score	
					A(0.70)	B(0.70)
SP	SS	26.50	59.00	55.81 (94.59)	0.42	0.48
PP	PS	9.85	97.96	62.55 (63.85)	0.51	0.68
PP	SS	23.98	90.01	69.80 (77.54)	0.44	0.75
SP	PP	24.64	90.43	64.28 (71.08)	0.71	0.45
SS	PS	11.89	93.03	69.43 (74.63)	0.70	0.50

Мета-агент

Мета-агент представляет собой ансамбль из всех четырех агентов (SS, PS, SP, PP) и агента выбора во главе, функция потерь которого выглядит следующим образом:

$$L_s = \mathop{\mathbb{E}}_{s_t \sim \pi_s} \left[\sum_t \gamma^{(T-t)} ((r_s(s_{1\ldots T}) + r_o) - b_s) \right] + \lambda H[\pi_s]$$

Мета-агент



Decision tree learnt by selector agent

Мета-агент против остальных

B	Dialog Length	Agreement Rate (%)	Optimality Rate (%)	Average Score	
				Meta	B
PP	18.68	94.41	77.15 (81.71)	0.64	0.61
SS	19.17	86.25	73.33 (85.02)	0.54	0.66
PS	13.10	92.27	76.56 (82.97)	0.69	0.55
SP	20.53	90.22	81.40 (90.22)	0.55	0.71

Результаты

Переговоры с настоящими людьми

Results of Human Evaluation

Agent	Dialog Length	Agreement Rate (%)	Optimality Rate (%)	Agent Score	Human Score	Agent Won (%)	Human Won (%)	Tied (%)
PP	15.07	87.38	70.87	0.58	0.62	36.67	51.11	12.22
SS	19.56	73.79	60.20	0.58	0.44	60.53	21.05	18.42
PS	13.57	92.93	66.67	0.57	0.57	40.22	52.17	7.61
SP	21.75	72.28	59.41	0.61	0.39	68.49	20.55	10.96
META	16.78	88.30	56.40	0.57	0.56	46.99	44.58	8.43

Результаты

(1) A deep learning model and a reinforcement learning procedure for training an AI agent to negotiate in the domain of contract negotiation.
(2) Modeling selfish/prosocial behavior by varying the reward signal for the agent and its opponent in the reinforcement learning framework and empirical evidence for the same.
(3) An AI agent with a dynamic behavior (varying within a negotiation instance) by learning an ensemble of different agent behaviors using reinforcement learning.
(4) Evidence for the usability and success of the negotiation agent against human players through real life experimental results.

Другие подходы

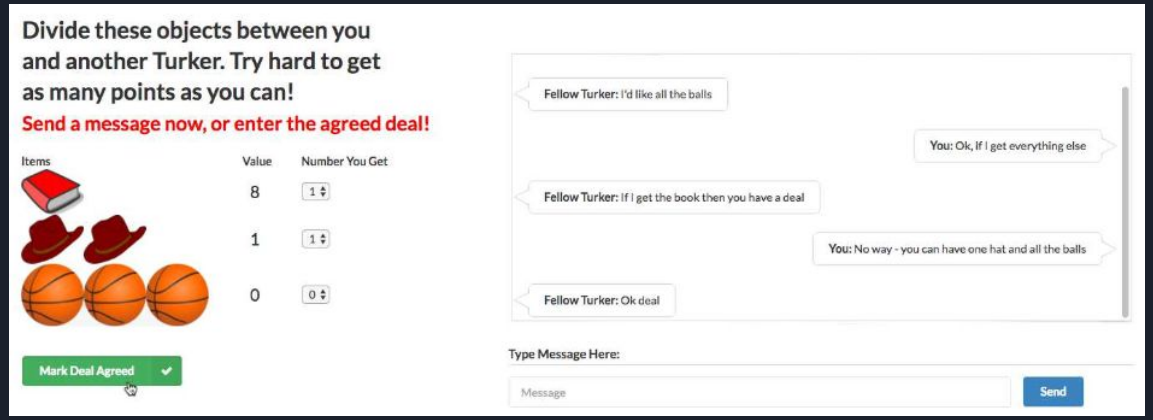
1. Jakob Foerster, Ioannis Alexandros Assael, Nando de Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning.
2. Kris Cao, Angeliki Lazaridou, Marc Lanctot, Joel Leibo, Karl Tuyls, and Stephen Clark. 2018. Emergent Communication through Negotiation.
3. Luís M Camarinha-Matos and Ana Inês Oliveira. 2007. Contract negotiation wizard for VO creation.

Есть также и другие работы на данную тему, однако ни в одной из выше перечисленных статей не представлены результаты переговоров с людьми.

Другие подходы

Mike Lewis, Denis Yarats, Yann N Dauphin, Devi Parikh, and Dhruv Batra. 2017.

Deal or no deal? End-to-end learning for negotiation dialogues.



Литература

https://arxiv.org/pdf/1809.07066.pdf

https://arxiv.org/pdf/1706.05125.pdf

http://delivery.acm.org/10.1145/3160000/3157336/p2145-foerster.pdf?ip=92.242.59.6&id=3157336&acc=NO%20RULES&key=24E52AF21FCA316E%2EF976A362468BE1A3%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1541077404_464a48833ab5a9939ec747e53f4e79c6