

Learning to Communicate with Deep Multi-Agent Reinforcement Learning

Jakob N. Foerster et al

Index

- Background
 - Independent DQN
 - DRQN
- Method to learn Inter-Agent Communication
 - RIAL
 - DIAL
- Experiment
- Appendix – Channel Noise

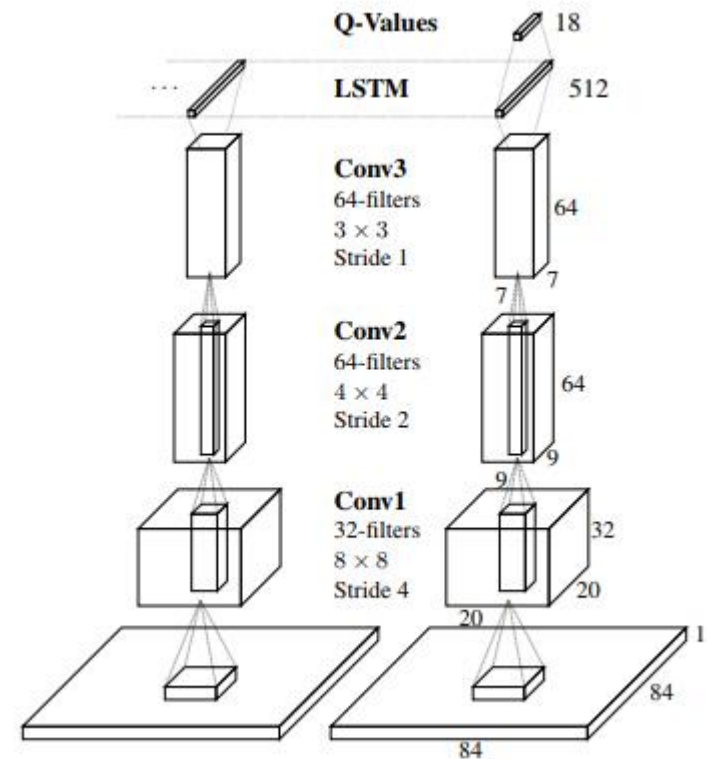
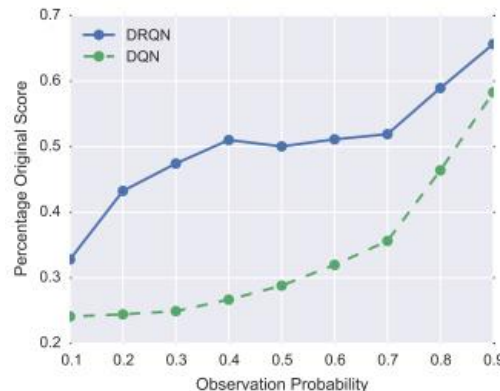
BACKGROUND

Independent DQN

- Multi-Agent Reinforcement Learning에 대한 가장 Straightforward한 접근
- 각 agent a 가 각각의 Q-Function을 학습
- 문제점
 - 다른 agent의 학습이 환경을 non-stationary하게 만듦
- 성과
 - Two player Pong

DRQN

- POMDP의 해결
 - 기존 Breakout: 4장의 연속 image를 받음
 - 새 해법: RNN을 사용
- 신경망 구조의 변경점:
 - DQN의 첫 FC Layer를 LSTM으로
- Flickering Pong을 잘 해결



METHOD TO LEARN INTER-AGENT COMMUNICATION

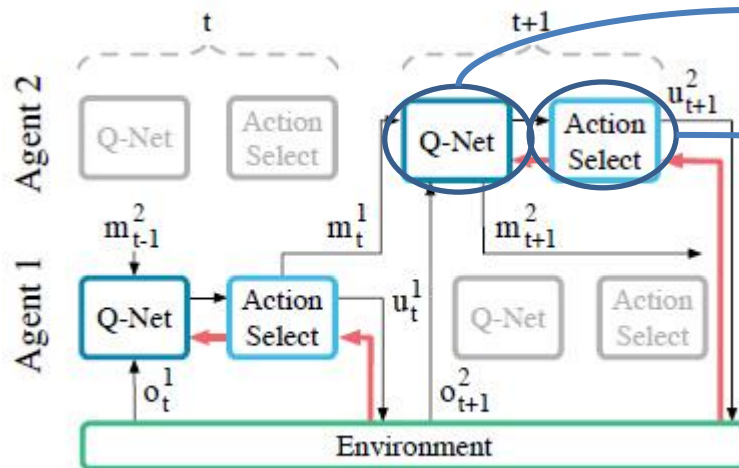
From Previous Literature

- Turn off experience replay
 - because of the non-stationarity, which is caused by a concurrent multi-agent learning
- Previous step에서 (자신의) action, *message*를 다시 넣어줌
- Parameter Sharing

RIAL

- Diagram

Q-Value를 approximate하는 신경망들,
action(u)과 message(m)은 별도의 신경망을 가짐




Q-Value 받아,
discrete한 u, m을 각각 선택

Red Line은 Gradient Flow

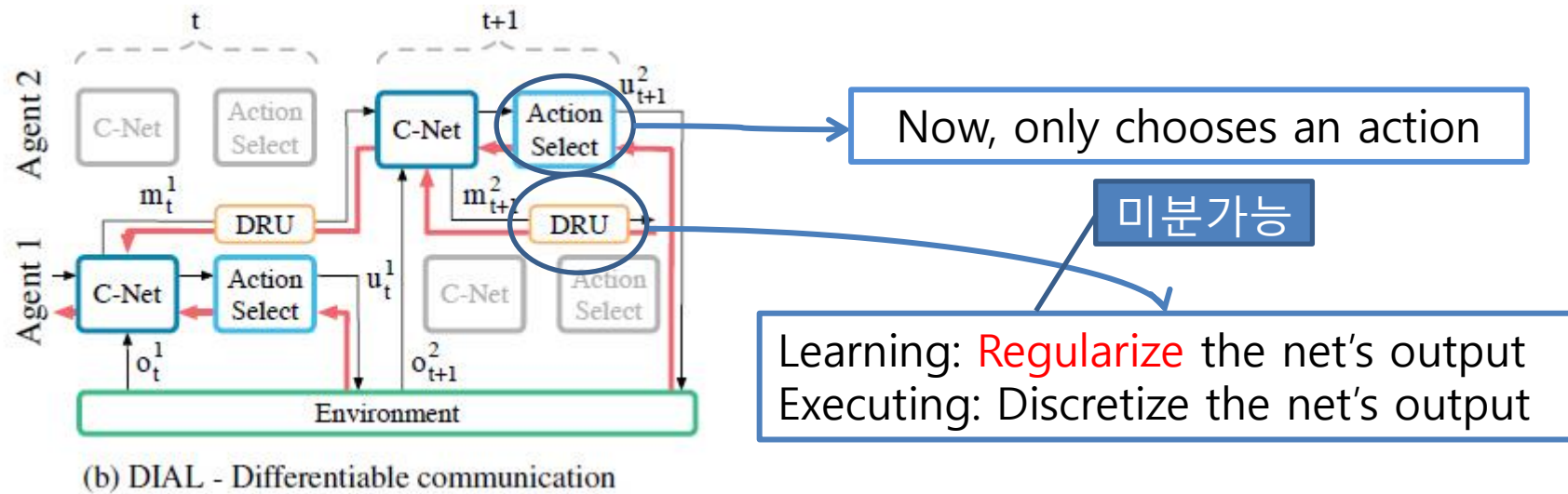
(a) RIAL - RL based communication

RIAL의 한계

- Message의 Q-Value를 Evaluate하는 Net
 - Environment에서 주어지는 정보만을 기반
 - Action Selector에서 discrete하게 선택
-  **미분 불가능**하므로 Inter-Agent Gradient flow는 불가
- 따라서 update에 구조적 한계가 내재함

DIAL

- Centralized Learning,
Decentralized Execution



Diagram

Richer Feedback

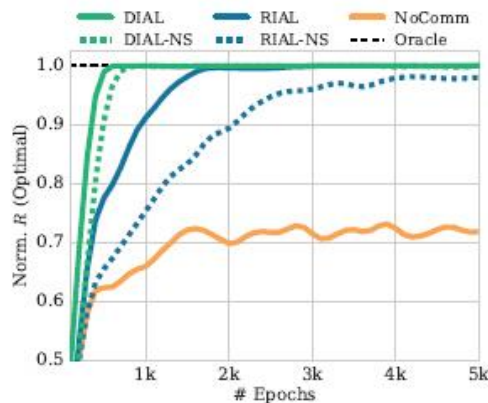
RIAL과 DIAL의 Message 특성

- RIAL
 - Communication action에 대한 Q-value를 estimate하는 Net
 - ∴ Use one-hot encoding
- DIAL
 - 하나의 Real-Value로 learning
 - Execution 시에만 discretize
 - Continuous Message Space
 - ∴ Can use binary encoding

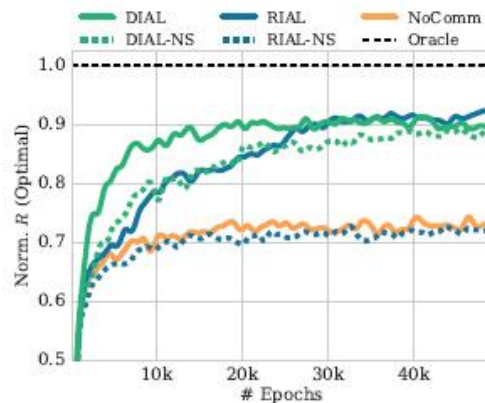
EXPERIMENT

Switch Riddle

- Result

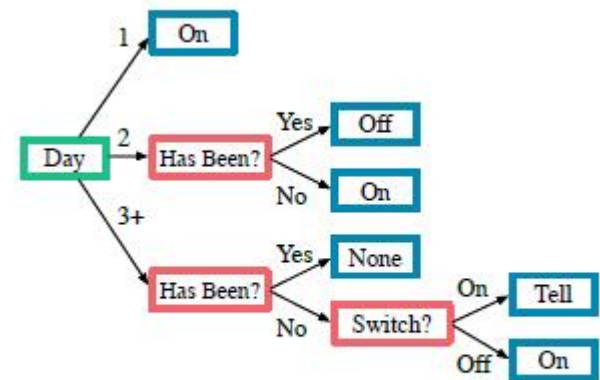


(a) Evaluation of $n = 3$



(b) Evaluation of $n = 4$

- Learned Protocol



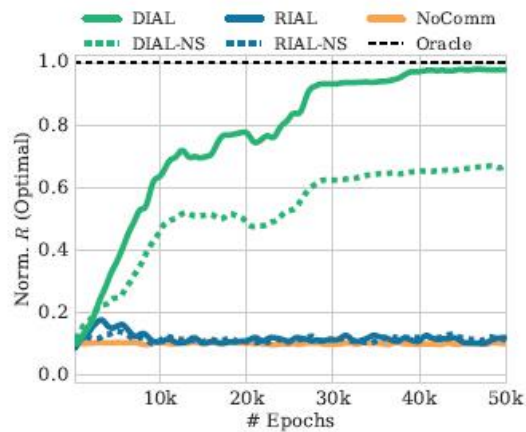
(c) Protocol of $n = 3$

MNIST Game

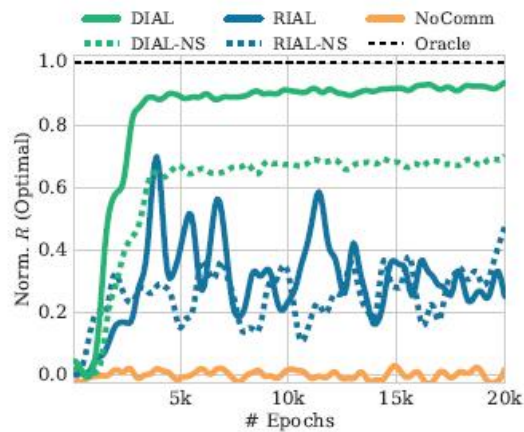
- 1. Color? Parity?
 - 보상: $r(a) = 2(-1)^{a_2^a + c^a + d^{a'}} + (-1)^{a_2^a + d^a + c^{a'}}$
 - c^a : a의 색, d^a : a의 digit이 나타내는 parity
 - a^a : a의 action
 - 따라서 상대방의 색, 또는 parity에 맞춰 a^a 선택
 - 2의 가중치가 붙은 parity 정보를 교환함이 oracle
- 2. 상대방의 숫자 맞추기
 - 1 step에 1 bit 교환, 총 4 step (4 bit 교환)
 - 5 step에 정답 제출
 - 자신의 숫자를 binary encoding함이 oracle

MNIST Game Result

- Result



(a) Evaluation of Multi-Step



(b) Evaluation of Colour-Digit

- Learned Protocol

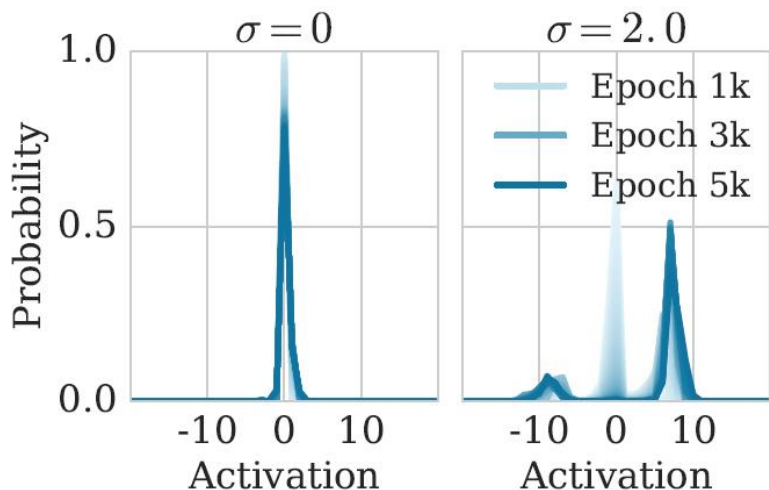


9	0	1	0	0
8	0	0	0	0
7	0	1	1	1
6	1	1	0	0
5	1	0	1	1
4	0	0	1	0
3	1	0	0	1
2	0	0	1	1
1	1	1	1	1
0	1	0	0	0
	1	2	3	4

(c) Protocol of Multi-Step

Appendix – Channel Noise

- DRU에서 noise를 준 결과,
Message의 중심집중이 없어짐



Learned two
different mode

끝

들어주셔서 감사합니다.