RLDM 2017

Communications that Emerge through Reinforcement Learning Using a (Recurrent) Neural Network

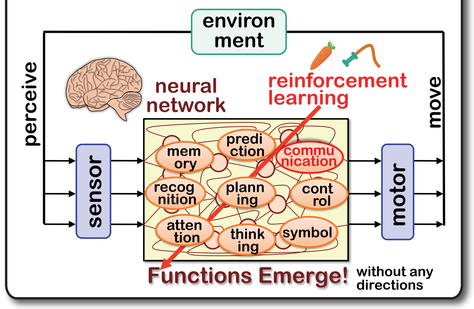
Katsunari SHIBATA (Oita University, JAPAN) katsunarishibata@gmail.com http://shws.cc.oita-u.ac.jp/shibata/home.html

You can find the original paper for each work at the bottom of this poster. Feel free to take it back. Little is known about our works, and so Thank you for referring them in your papers!

Please refer to the poster No. 20!

Function Emergence through End-to-End RL

[Shibata, 1997] [Hassabis (Google DeepMind), 2016]



The direction for Artificial General Intelligence!

Our Approach

(since 1999)

The entire function should be comprehensively learned in each agent

☆ End-to-End Reinforcement Learning using a (Recurrent) Neural Network.
☆ Completely Decentralized and Independent Learning
☆ Rewards or Punishments are not given to communication itself, but only for the result of actions after communication.

General Approach

Communication: a very comprehensive function!

Communication
consider
- Sensor signals
- Context

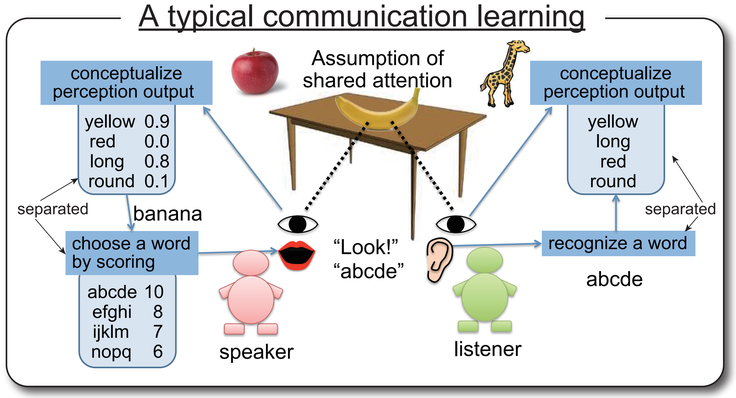
decide
- What information is communicated
- How the information is represented in signals

committed to other functions

isolated and learned

Symbol Grounding Problem

A typical communication learning



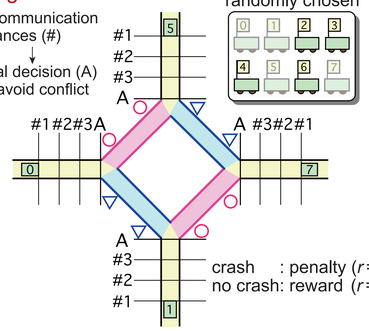
3 our representative works are introduced here

1. Dynamic Communication (1999)

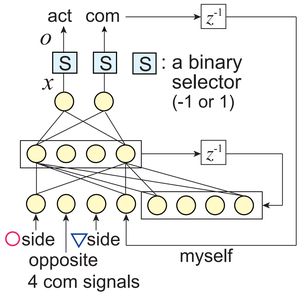
[Shibata, 1999]

Negotiation Task

3 communication chances (#)
↓
final decision (A) to avoid conflict



Recurrent Neural Network and I/O in each agent



Negotiation Examples after learning

com_signal = 1 → request for the path ▽ (Act = -1)
com_signal = -1 → request for the path ○ (Act = 1)

case1

	#1	#2	#3	Act
Agent 2	-1	-1	-1	1 ○
Agent 5	1	1	1	-1 ▽
Agent 7	-1	-1	-1	1 ○
Agent 4	1	1	1	-1 ▽

case2

	#1	#2	#3	Act
Agent 0	-1	1	1	- ▽
Agent 2	-1	-1	-1	1 ○
Agent 5	1	1	1	- ▽
Agent 7	-1	-1	-1	1 ○

case3

	#1	#2	#3	Act
Agent 0	-1	-1	-1	1 ○
Agent 2	-1	-1	-1	-1 ▽
Agent 3	-1	-1	-1	1 ○
Agent 7	-1	-1	1	-1 ▽

case4

	#1	#2	#3	Act
Agent 2	-1	-1	-1	1 ○
Agent 4	1	1	1	-1 ▽
Agent 5	1	1	1	1 ○
Agent 7	-1	-1	1	-1 ▽

same input different output

Learning

$x_d = x + rx'o$

training signal, output, reinforcement signal, derivative of output function, selected signal

RNN (BPTT)

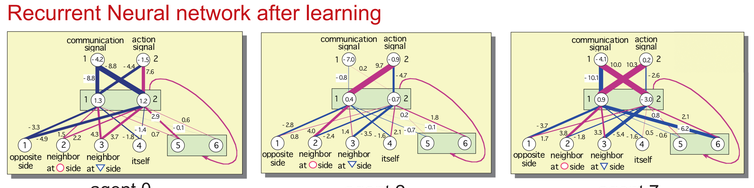
Average signal output

Individuality emerged

	com #1	com #2	com #3	action	num of com patterns	
Agent 0	-1.0	-0.43	-0.46	0.35	7	indecisive
Agent 1	1.0	0.03	0.28	-0.26	4	
Agent 2	-1.0	-1.00	-1.00	0.58	2	stubborn
Agent 3	-1.0	-0.09	-0.34	0.18	5	
Agent 4	1.0	1.00	1.00	-0.57	2	stubborn
Agent 5	1.0	0.77	0.53	-0.48	5	
Agent 6	1.0	0.43	0.32	-0.29	5	
Agent 7	-1.0	-0.77	-0.55	0.49	4	

absolute value: degree of stubborn
sign: preference of signal

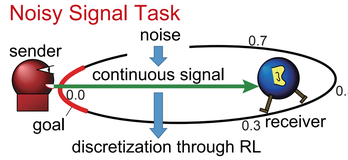
Recurrent Neural network after learning



2. Signal Discretization (2005)

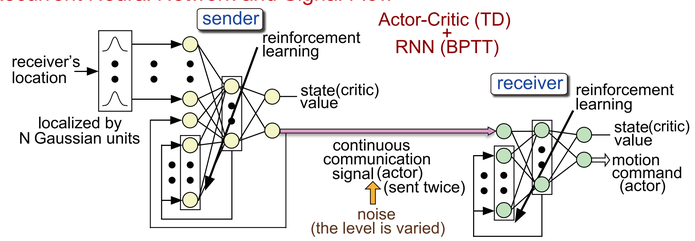
[Shibata, 2005]

Noisy Signal Task

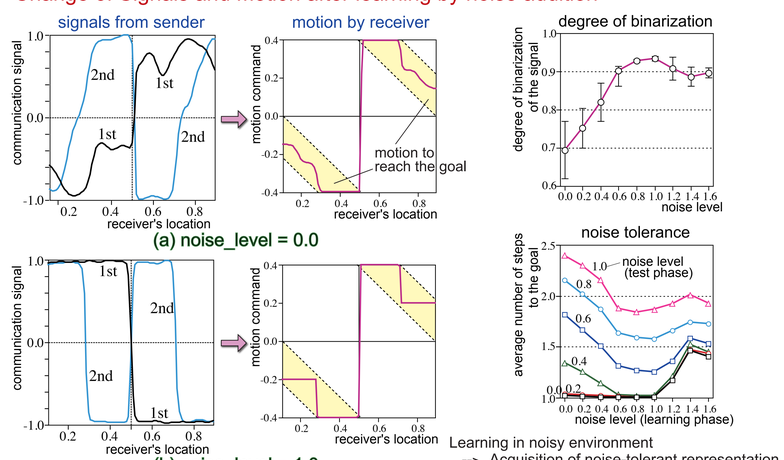


* Signals are sent and received two times
* A reward is given to each agent only when the receiver reaches the goal
* The receiver can reach the goal in one step from any position but has to avoid overshooting

Recurrent Neural Network and Signal Flow



Change of Signals and Motion after learning by noise addition

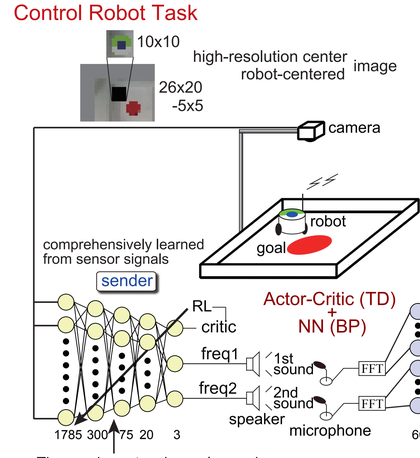


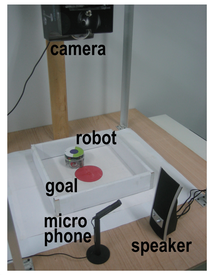
Learning in noisy environment
--> Acquisition of noise-tolerant representation

3. Grounded Communication (2011)

[Shibata and Sasahara, 1999]

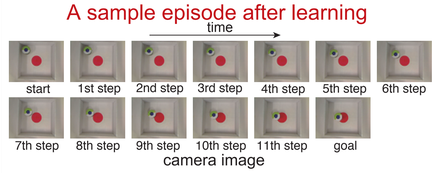
Control Robot Task

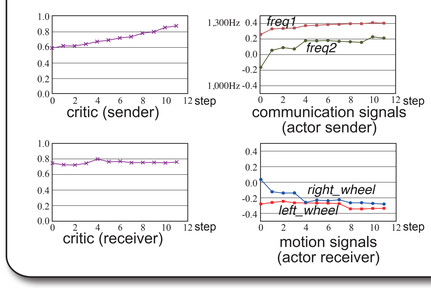




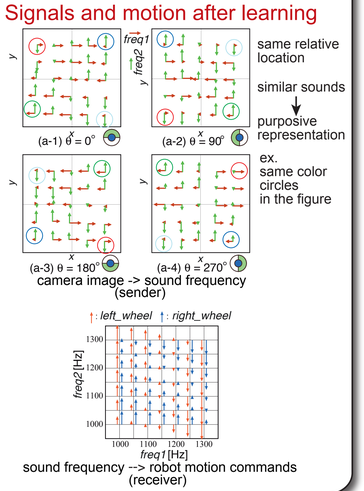
The sender network was learned to control the robot directly before communication learning

A sample episode after learning





Signals and motion after learning



same relative location
similar sounds
↓
purposive representation
ex. same color circles in the figure