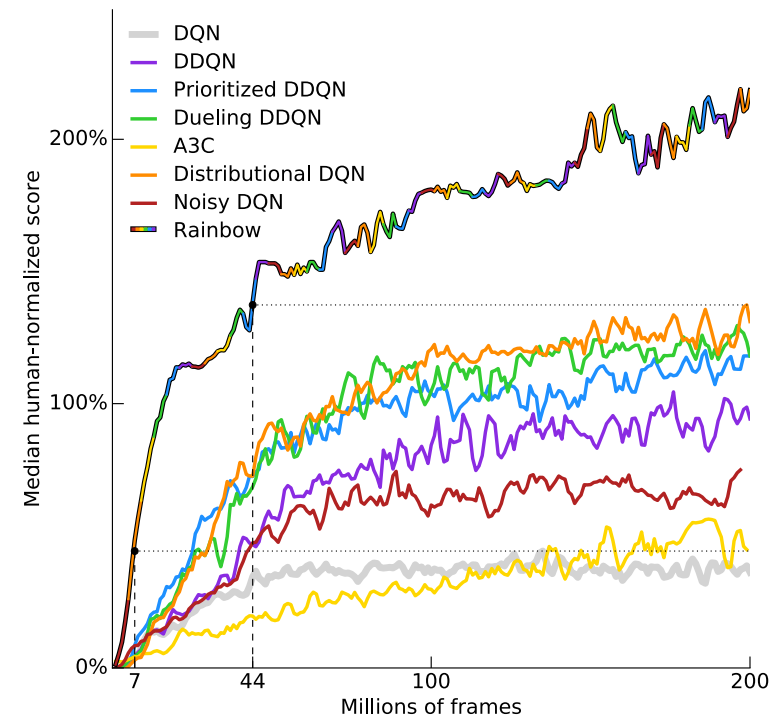
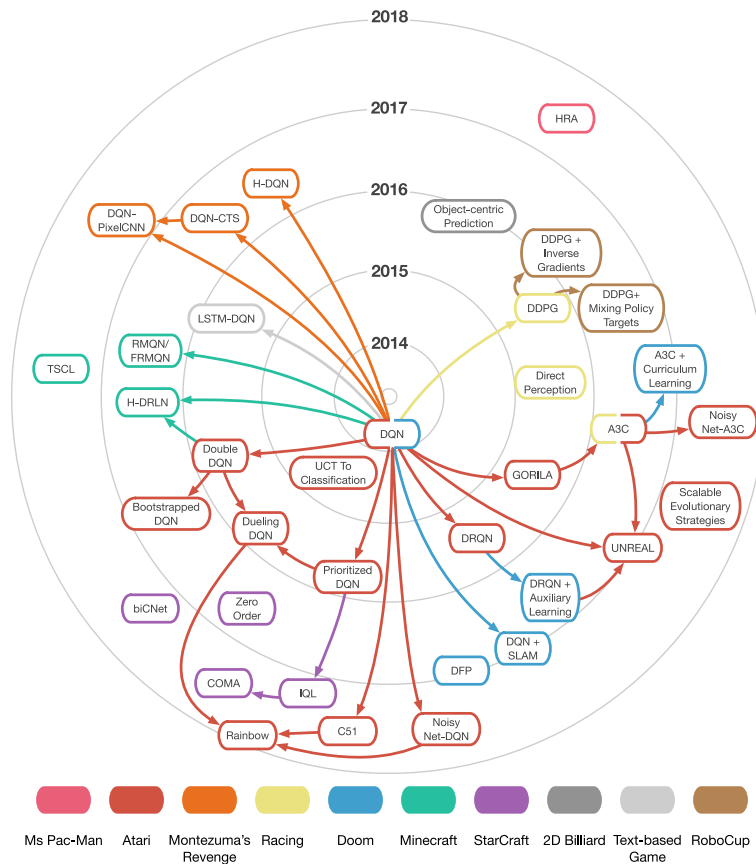


1 強化学習のモデル

1. ポリシーベース: 方策勾配法
 - UNREAL
 - 2. 状態価値関数 Q ベース
- ダブル Q 学習, Hasselt, Guez, and Silver (2015),
 3. モデルベース
- A3C, Mnih et al. (2016) beyond Actor-Critic
 4. 全部載せ
- Rainbow, Hessel et al. (2017),

2 Rainbow, Hessel et al. (2017)



Justesen et al. (2018), Hessel et al. (2017)

3 Rainbow の結果(2)

Game	DQN	A3C	DDQN	Prior. DDQN	Duel. DDQN	Distrib. DQN	Noisy DQN	Rainbow
alien	634.0	518.4	1033.4	900.5	1,486.5	1,997.5	533.3	6,022.9
amidar	178.4	263.9	169.1	218.4	172.7	237.7	148.0	202.8
assault	3489.3	5474.9	6060.8	7,748.5	3,994.8	5,101.3	5,124.3	14,491.7
asterix	3170.5	22140.5	16837.0	31,907.5	15,840.0	395,599.5	8,277.3	280,114.0
asteroids	1458.7	4474.5	1193.2	1,654.0	2,035.4	2,071.7	4,078.1	2,249.4
atlantis	292491.0	911,091.0	319688.0	593,642.0	445,360.0	289,803.0	303,666.5	814,684.0
bank_heist	312.7	970.1	886.0	816.8	1,129.3	835.6	955.0	826.0
battle_zone	23750.0	12950.0	24740.0	29,100.0	31,320.0	32,250.0	26,985.0	52,040.0
beam_rider	9743.2	22707.9	17417.2	26,172.7	14,591.3	15,002.4	15,241.5	21,768.5
berzerk	493.4	817.9	1011.1	1,165.6	910.6	1,000.0	670.8	1,793.4
bowling	56.5	35.1	69.6	65.8	65.7	76.8	79.3	39.4
boxing	70.3	59.8	73.5	68.6	77.3	62.1	66.3	54.9
breakout	354.5	681.9	368.9	371.6	411.6	548.7	423.3	379.5
centipede	3973.9	3755.8	3853.5	3,421.9	4,881.0	7,476.9	4,214.4	7,160.9
chopper_command	5017.0	7021.0	3495.0	6,604.0	3,784.0	9,600.5	8,778.5	10,916.0
crazy_climber	98128.0	112646.0	113782.0	131,086.0	124,566.0	154,416.5	98,576.5	143,962.0
defender	15917.5	56533.0	27510.0	21,093.5	33,996.0	32,246.0	18,037.5	47,671.3
demon_attack	12550.7	113,308.4	69803.4	73,185.8	56,322.8	109,856.6	25,207.8	109,670.7
double_dunk	-6.0	-0.1	-0.3	2.7	-0.8	-3.7	-1.0	-0.6
enduro	626.7	-82.5	1216.6	1,884.4	2,077.4	2,133.4	1,021.5	2,061.1
fishing_derby	-1.6	18.8	3.2	9.2	-4.1	-4.9	-3.7	22.6
freeway	26.9	0.1	28.8	27.9	0.2	28.8	27.1	29.1
frostbite	496.1	190.5	1448.1	2,930.2	2,332.4	2,813.9	418.8	4,141.1
gopher	8190.4	10022.8	15253.0	57,783.8	20,051.4	27,778.3	13,131.0	72,595.7
gravitar	298.0	303.5	200.5	218.0	297.0	422.0	250.5	567.5
hero	14992.9	32464.1	14892.5	20,506.4	15,207.9	28,554.2	2,454.2	50,496.8
ice_hockey	-1.6	-2.8	-2.5	-1.0	-1.3	-0.1	-2.4	-0.7
kangaroo	4496.0	94.0	11204.0	10,241.0	10,334.0	9,555.5	7,465.0	10,841.0
krull	6206.0	5560.0	6796.1	7,406.5	8,051.6	6,757.8	6,833.5	6,715.5
kung_fu_master	20882.0	28819.0	30207.0	31,244.0	24,288.0	33,890.0	27,921.0	28,999.8
montezuma_revenge	47.0	67.0	42.0	13.0	22.0	130.0	55.0	154.0
ms_pacman	1092.3	653.7	1241.3	1,824.6	2,250.6	2,064.1	1,012.1	2,570.2

Hessel et al. (2017)

4 AlphaGo, AlphaGoZero

Silver et al. (2016), Silver et al. (2017) ^

文献

Hasselt, Hado van, Arthur Guez, and David Silver. 2015. “Deep Reinforcement Learning with Double Q-Learning.”

Hessel, Matteo, Joseph Modayil, Hado van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver. 2017. “Rainbow: Combining Improvements in Deep Reinforcement Learning.”

Justesen, Niels, Philip Bontrager, Julian Togelius, and Sebastian Risi. 2018. “Deep Learning for Video Game Playing.”

Mnih, Volodymyr, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Tim Harley, Timothy P. Lillicrap, David Silver, and Koray Kavukcuoglu. 2016. “Asynchronous Methods for Deep Reinforcement Learning.”

Silver, David, Aja Huang, Chris J. Maddison and Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, et al. 2016. “Mastering the Game of Go with Deep Neural Networks and Tree Search.” *Nature* 529 (28): 484–92.

Silver, David, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez Thomas Hubert, et al. 2017. “Mastering the Game of Go Without Human Knowledge.” *Nature* 550: 354–69.