# A Brief survey for Deep Reinforcement Learning

2017- March -31

Di Wu

ID：260562997

# Outline

- Introduction
- Research results before DQN
- Recent Research Progress
- Applications
  - Successful applications
  - Potential applications

# Introduction

- Deep reinforcement learning (Deep RL) combines the conception power of deep learning and decision power of reinforcement learning
- Deep reinforcement learning is reinforcement learning
- Deep learning can be used
  - Approximate value function
  - Approximate policy function
  - Predict future reward or state representation

# Early Research Results before DQN

- Before the introduction of DQN there are some previous work of using neural network for Reinforcement learning
- Mainly use (shallow) neural networks as a kind of function approximator

# Early Research Results before DQN

- Some early research results are shown in following table:

| Title | Year | Topic |
|---|---|---|
| Neural fitted q iteration-first experiences with a data efficient neural reinforcement learning method [1] | 2004 | Neural fitted Q |
| Deep auto-encoder neural networks in reinforcement learning [2] | 2010 | Small state space vision |
| Deep belief nets as function approximators for reinforcement learning [3] | 2011 | DBF Q approximator |
| Autonomous reinforcement learning on raw visual input data in a real world application [4] | 2012 | Deep fitted Q |

# Current Research Progress for Deep RL

- The DQN is formally proposed in two [5] 2013, and [6] 2015.
- Show great success of Game playing (Atari and Go).
- Then it become a hot research topic.
- The following tables lists some recent Deep RL papers in different fields.

# Current Research Progress for Deep RL

- Some major research work after DQN

| **Filed: Value based** | **Year** | **Topic** |
|---|---|---|
| Playing atari with deep reinforcement learning [5] , Human-level control through deep reinforcement learning [6]. | 2013, 2015 | Introduced the idea Deep RL |
| Dueling network architectures for deep reinforcement learning [7] | 2016 | Dueling network |
| Deep reinforcement learning with double q-learning [8] | 2016 | Double Q Learning |
| Deep recurrent q-learning for partially [9] | 2015 | Use recurrent NN for Deep RL |
| Increasing the action gap: new operators for reinforcement learning [10] | 2016 | New operator in bellman equation |
| Prioritized experience replay [11] | 2016 | Prioritized experience replay |

# Current Research Progress for Deep RL

- Some major research work after DQN

| Filed: Policy based | Year | Topic |
|---|---|---|
| Deterministic policy gradient algorithms [12] | 2014 | Deterministic |
| Asynchronous methods for deep reinforcement learning [13] | 2016 | A3C with Multi thread |
| Trust region policy optimization [14] | 2015 | Trust region |
| High-Dimensional continuous control using generalized advantage function [15] | 2015 | generalized advantage function |
| End-to-end training of deep visumotor policies [16] | 2016 | Guided-policy search for Robot |
| PGQ: Combining policy gradient and q-learning [17] | 2016 | policy gradient with off-policy Q-learning |

# Current Research Progress for Deep RL

- Some major research work after DQN

| Filed: Hierarchical Deep RL | Year | Topic |
|---|---|---|
| Hierarchical deep reinforcement learning: integrating temporal abstraction and intrinsic motivation [18] | 2016 | Deep RL for HRL |
| Hierarchical reinforcement learning using spatio-temporal abstractions and deep neural networks [19] | 2016 | HRL with intra-option |
| Multi-Level Discovery of Deep Options [20] | 2017 | Deep Options Structure |
| FeUdal Networks for Hierarchical Reinforcement Learning [21] | 2017 | FeUdal Networks |

# Current Research Progress for Deep RL

- Some major research work after DQN

| Filed: Transfer/Multi-Task Deep RL | Year | Topic |
|---|---|---|
| Actor-mimic: deep multitask and transfer reinforcement learning [22] | 2016 | Multitask transfer |
| Policy distillation [23] | 2015 | Multitask and multi experience replay |
| Towards Knowledge Transfer in Deep Reinforcement Learning [24] | 2016 | Task similarity impacts anslysis |
| Multi-task learning with deep model based reinforcement learning [25] | 2016 | Low dimensional energy model to learn the MDP distribution |
| Exploration for Multi-task Reinforcement Learning with Deep Generative Models [26] | 2016 | Solve different tasks simultaneously |

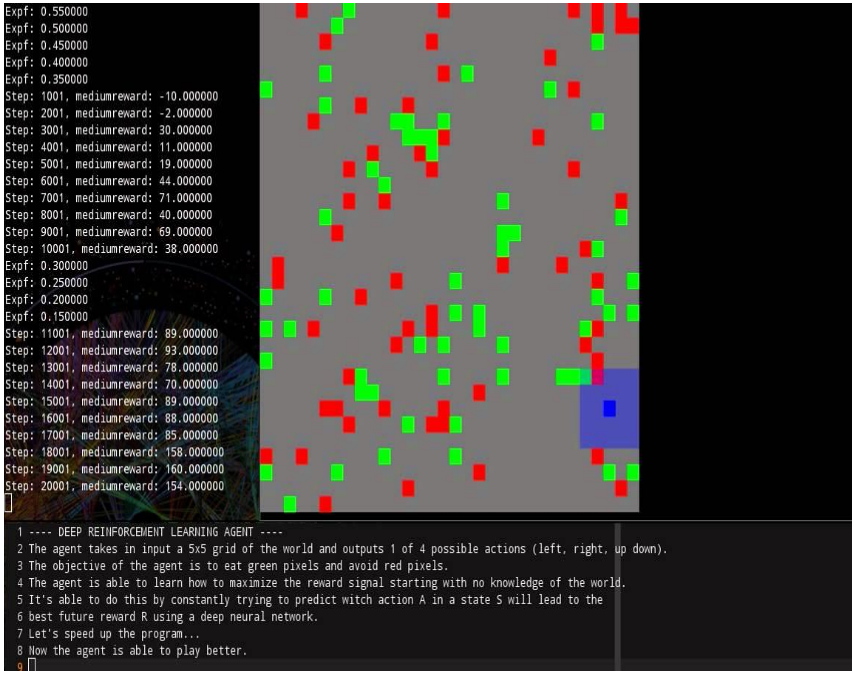
# Current Research Progress for Deep RL

- Some major research work after DQN

| Filed: Multi-agent Deep RL | Year | Topic |
|---|---|---|
| Multi-Agent cooperation and competition with deep reinforcement learning [27] | 2015 | Deep Q-Networks can become a practical tool for studying the decentralized learning |
| Learning to communicate to solve riddles with deep distributed recurrent q-networks [28] | 2016 | Applications in  in learning communication protocols |
| Learning to Communicate with Deep Multi-Agent Reinforcement Learning [29] | 2016 | Centralized learning but decentralized execution |
| Coordinated Deep Reinforcement Learners for Traffic Light Control [30] | 2016 | Application paper, new reward function |
| Opponent Modeling in Deep Reinforcement Learning [31] | 2016 | neural-based models that jointly learn a policy and the behavior of opponents. |

# Applications of Deep RL

- Games:




# Applications of Deep RL

- Robot &Self-Driving


# Applications of Deep RL

- NLP, vision, video:
  - Text generation
  - Text analytic
  - Dialogue system
  - Future reward prediction with current image

# Applications of Deep RL: Potential Applications

- Smart Grid:
  - Deep mind mentions the application of Deep RL based energy management system saves 40% energy consumption (millions of dollars) for google data center
- Fintech: personalized system, and trading
  - Deep Direct Reinforcement Learning for Financial Signal Representation and Trading [32]
- Medical Service, Intelligent Manufacturing

# Reference

[1] Riedmiller M. Neural fitted Q iteration–first experiences with a data efficient neural reinforcement learning method[C]//European Conference on Machine Learning. Springer Berlin Heidelberg, 2005: 317-328.

[2] Lange S, Riedmiller M. Deep auto-encoder neural networks in reinforcement learning[C]//Neural Networks (IJCNN), The 2010 International Joint Conference on. IEEE, 2010: 1-8.

[3] Abtahi F, Fasel I. Deep belief nets as function approximators for reinforcement learning[J]. RBM, 2011, 2: h3.

[4] Lange S, Riedmiller M, Voigtlander A. Autonomous reinforcement learning on raw visual input data in a real world application[C]//Neural Networks (IJCNN), The 2012 International Joint Conference on. IEEE, 2012: 1-8.

[5] Mnih V, Kavukcuoglu K, Silver D, et al. Playing atari with deep reinforcement learning[J]. arXiv preprint arXiv:1312.5602, 2013.

# Reference


[6] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[7]Wang Z, Schaul T, Hessel M, et al. Dueling network architectures for deep reinforcement learning[J]. arXiv preprint arXiv:1511.06581, 2015.

[8] Van Hasselt H, Guez A, Silver D. Deep Reinforcement Learning with Double Q-Learning[C]//AAAI. 2016: 2094-2100.

[9] Hausknecht M, Stone P. Deep recurrent q-learning for partially observable mdps[J]. arXiv preprint arXiv:1507.06527, 2015.

[10] Bellemare M G, Ostrovski G, Guez A, et al. Increasing the Action Gap: New Operators for Reinforcement Learning[C]//AAAI. 2016: 1476-1483.

[11] Schaul T, Quan J, Antonoglou I, et al. Prioritized experience replay[J]. arXiv preprint arXiv:1511.05952, 2015.

[12] Lever G. Deterministic policy gradient algorithms[J]. 2014.

# Reference


[13] Mnih V, Badia A P, Mirza M, et al. Asynchronous methods for deep reinforcement learning[C]//International Conference on Machine Learning. 2016.

[14] Schulman J, Levine S, Abbeel P, et al. Trust Region Policy Optimization[C]//ICML. 2015: 1889-1897.

[15] Schulman J, Moritz P, Levine S, et al. High-dimensional continuous control using generalized advantage estimation[J]. arXiv preprint arXiv:1506.02438, 2015.

[16] Levine S, Finn C, Darrell T, et al. End-to-end training of deep visuomotor policies[J]. Journal of Machine Learning Research, 2016, 17(39): 1-40.

[17] O'Donoghue B, Munos R, Kavukcuoglu K, et al. PGQ: Combining policy gradient and Q-learning[J]. arXiv preprint arXiv:1611.01626, 2016.

[18] Kulkarni T D, Narasimhan K, Saeedi A, et al. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation[C]//Advances in Neural Information Processing Systems. 2016: 3675-3683.

# Reference


[19] Krishnamurthy R, Lakshminarayanan A S, Kumar P, et al. Hierarchical reinforcement learning using spatio-temporal abstractions and deep neural networks[J]. arXiv preprint arXiv:1605.05359, 2016.

[20] Fox R, Krishnan S, Stoica I, et al. Multi-Level Discovery of Deep Options[J]. arXiv preprint arXiv:1703.08294, 2017.

[21] Vezhnevets A S, Osindero S, Schaul T, et al. FeUdal Networks for Hierarchical Reinforcement Learning[J]. arXiv preprint arXiv:1703.01161, 2017.

[22] Parisotto E, Ba J L, Salakhutdinov R. Actor-mimic: Deep multitask and transfer reinforcement learning[J]. arXiv preprint arXiv:1511.06342, 2015.

[23] Rusu A A, Colmenarejo S G, Gulcehre C, et al. Policy distillation[J]. arXiv preprint arXiv:1511.06295, 2015.

[24] Glatt R, da Silva F L, Costa A H R. Towards Knowledge Transfer in Deep Reinforcement Learning[C]//Intelligent Systems (BRACIS), 2016 5th Brazilian Conference on. IEEE, 2016: 91-96.

# Reference

[25] Mujika A. Multi-task learning with deep model based reinforcement learning[J]. arXiv preprint arXiv:1611.01457, 2016.

[26] Bangaru S P, Suhas J S, Ravindran B. Exploration for Multi-task Reinforcement Learning with Deep Generative Models[J]. arXiv preprint arXiv:1611.09894, 2016.

[27] Tampuu A, Matiisen T, Kodelja D, et al. Multiagent cooperation and competition with deep reinforcement learning[J]. arXiv preprint arXiv:1511.08779, 2015.

[28] Foerster J N, Assael Y M, de Freitas N, et al. Learning to communicate to solve riddles with deep distributed recurrent q-networks[J]. arXiv preprint arXiv:1602.02672, 2016.

[29] Foerster J, Assael Y M, de Freitas N, et al. Learning to communicate with deep multi-agent reinforcement learning[C]//Advances in Neural Information Processing Systems. 2016: 2137-2145.

# Reference

[30] van der Pol E, Oliehoek F A. Coordinated Deep Reinforcement Learners for Traffic Light Control[J].

[31] He H, Boyd-Graber J, Kwok K, et al. Opponent Modeling in Deep Reinforcement Learning[C]//Proceedings of The 33rd International Conference on Machine Learning. 2016: 1804-1813.

[32] Deng Y, Bao F, Kong Y, et al. Deep direct reinforcement learning for financial signal representation and trading[J]. IEEE transactions on neural networks and learning systems, 2017, 28(3): 653-664.

Thanks!