Panellist Position

Steering the Next-Generation Infrastructure with Deep Reinforcement Learning

Zhaobo Zhang, Futurewei Technologies, USA zzhang1@futurewei.com

- Next level of intelligence, dynamic decision-making
- Label (instructive)-> Reward (evaluative)
- Cloud-native architecture
- Autonomous infrastructure

→ Cloud-native infrastructure provides an interactive environment with observability and actionable APIs

→ Deep reinforcement learning is designed for finding the optimal policy, a perfect fit for control optimization

→ If a machine receives negative feedback, it should act differently next time

Advances in Communications

- Driven by new application requirements
  - AR/VR, Hologram, IoT, autonomous cars
  - Internet usage surge in pandemic (video conferencing/streaming, etc.)
- Current 5G characteristics
  - Throughput, latency, mobility, connections density, spectrum efficiency, Intelligence
- Enhanced by Cloud, AI
  - Cloud-native (microservice, API-driven) architecture, scalable, resilient, agile
  - Detailed observability and executable actions enable AI-based control

Network Intelligence

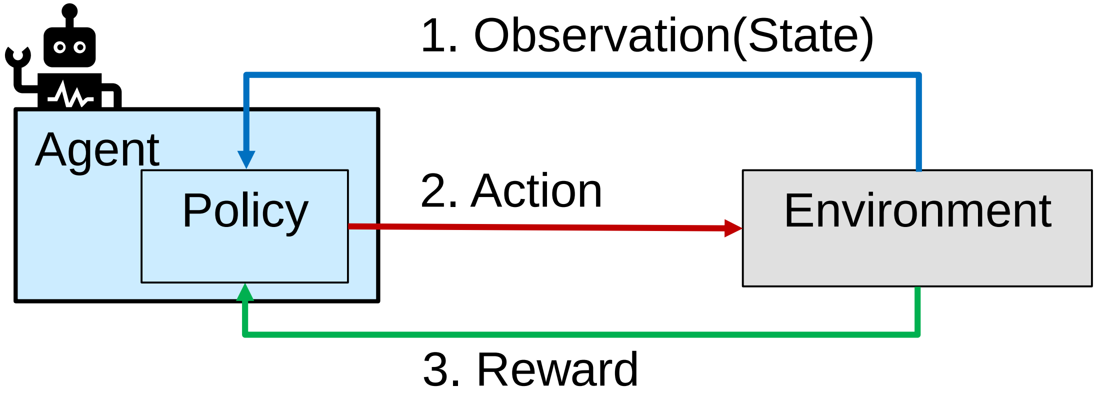
- Mature Machine Learning Application
  - Traffic prediction/classification
  - Fault diagnosis
  - Intrusion/Anomaly detection

Supervised, Unsupervised → Reinforcement, Nature Language Processing

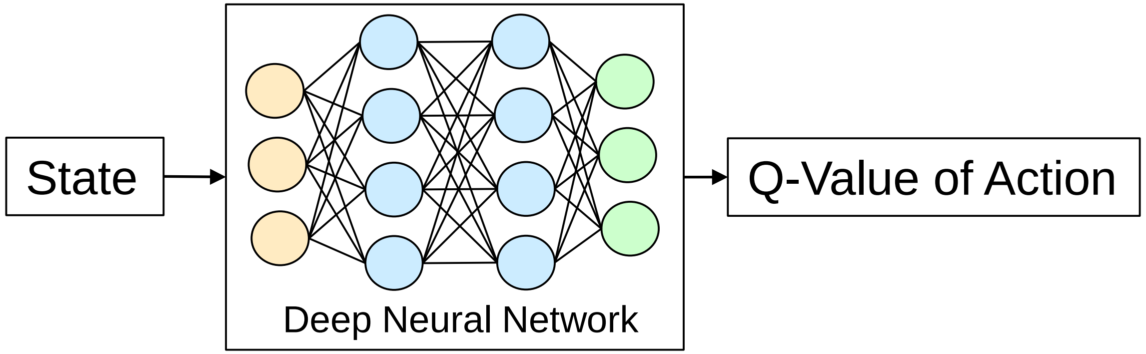
- Trending
  - AIOps, Virtual Assistance
  - Full-stack (from infra to app) observability and correlation
  - Closed-loop Control

✓ One step → Interactive
✓ Silo → multiple domain
✓ Analysis → Action

Deep Reinforcement Learning (DRL) for Real-time Decision Making



RL control loop



Deep Q Learning Algorithm

- MDP: discrete-time stochastic control process, a mathematical framework for modeling decision making
- Policy: a mapping from state to action
- Agent goal: find the optimal policy to maximize the reward, long-term cumulative return
- Two popular reinforcement learning algorithms
  - Q-Learning (value-based) vs Policy gradient (policy-based)

VNF Orchestration Example: State, Action and Reward

- Dynamic resource (VNF/CNF, bandwidth) orchestration in 5G ecosystem to provide cost-effective services with better performance
- Deep reinforcement learning model
  - Agent: network orchestrator
  - State Space: VNF requests, VNF availability
  - Action Space: VNF allocation options
  - Reward function: function of request processing time and VNF costs
- Reward function design
- Environment design (state transition)

More Application References

Ref	Problem Category	Agent	States	Actions	Reward
[1]	Routing Optimization (Traffic Engineering)	Network Controller	Link load, switch load (queue depth)	Assign weight value of links	Max-link-utilization
[2]	Adaptive Rate Control (Congestion Control)	Traffic Sender	Sent packet interval, packet, loss, average delay, set bytes, last action	Decide sending rate	Throughput, delay, packet loss rate
[3]	Job scheduling (Resource Management)	Cluster Controller	Available resource, job required resource (CPU, Memory, I/O)	Schedule a job at a certain time slot	Average Job slowdown
[4]	Spectrum Allocation (Network Slicing)	Base station	No. of arrived packets in each slice	Allocate bandwidth to each slice	Spectrum efficiency, Quality of Experience
[5]	Computation offloading (Mobile Edge Computing)	User device	Remained energy of device, connection condition, channel power gain between user device and base station, task info	Decide if offload tasks, decide CPU frequency and transmit power	Task completion latency, energy consumption

[1] Q. Li, et al., "Data-driven Routing Optimization based on Programmable Data Plane," *2020 ICCCN*
[2] L. Zhang, et al., "Reinforcement Learning Based Congestion Control in a Real Environment," 2020 ICCCN
[3] H. Mao, et al., “Resource Management with Deep Reinforcement Learning,” 2016 HotNets
[4] [1] R. Li *et al.*, "Deep Reinforcement Learning for Resource Management in Network Slicing," in *IEEE Access*, vol. 6, 2018
[5] Y. Zhang, et al., “A Deep Reinforcement Learning Approach for online computation offloading in Mobile Edge Computing,” 2020 IWQoS

Industry Research and Adoption of DRL

- Microsoft, finding the optimal cloud configuration for DNN inference workload [1]
- Google, optimizing chip layout with RL agent [2]
- VMware, continuous performance tuning for data center infrastructure [3]

[1] Y. Li, et al., "Automating Cloud Deployment for Deep Learning Inference of Real-time Online Services," IEEE INFOCOM 2020
[2] https://ai.googleblog.com/2020/04/chip-design-with-deep-reinforcement.html
[3] https://blogs.vmware.com/management/2019/08/tech-preview-project-magna.html

Takeaways

- Challenges
  - Real environment, to provide continuous trigger and feedback to the agent
  - Reward function design, to guide the agent to the real goal
  - Action space design, scalable
  - Interpretability, trusted AI actions
- Mission Possible
  - Modern architecture: structured data, microservice, orchestrator
  - Strength of DRL on continuous optimization in a dynamic environment


Westworld Dolores

THANK YOU