

№1 Дәріс. Нықталандырып оқытуға кіріспе

Дәрістің мақсаты: Студенттерді нықталандырып оқыту (Reinforcement Learning, RL) ұғымымен, оның негізгі принциптерімен, компоненттерімен және қолдану салаларымен таныстыру. RL-дің классикалық машиналық оқытудан айырмашылығын түсіндіру, агент пен ортаның өзара әрекеттесу процесін сипаттау, сондай-ақ негізгі алгоритмдер мен олардың қолданылуын талдау.

Кілт сөздер: Нықталандырумен оқыту, агент, орта, күтілетін марапат (reward), күшейткіш сигнал, Q-learning, SARSA, саясат (policy), күй (state), әрекет (action), құндылық функциясы (value function), Марков шешім процесі (MDP).

Дәріс жоспары:

1. Нықталандырумен оқыту ұғымы және оның ерекшелігі
2. RL жүйесінің негізгі компоненттері: агент және орта
3. Марапат пен стратегия (reward және policy)
4. Марков шешім процесі (MDP) және күйлер мен әрекеттер моделі
5. Негізгі RL алгоритмдері: Q-learning, SARSA
6. Қолдану салалары: ойындар, робототехника, басқару жүйелері
7. Қорытынды және даму перспективалары

Кіріспе

XXI ғасырдың маңызды бағыттарының бірі – интеллектуалды агенттерді құру. Мұндай агенттер өз әрекеттерінен үйреніп, қоршаған ортамен өзара әрекеттесу арқылы мақсатқа жетуді үйренеді. Бұл тәсіл нықталандырумен оқыту (Reinforcement Learning) деп аталады.

Машиналық оқытудың үш негізгі түрі бар:

1. Бақыланатын оқыту (Supervised learning) – дайын жауаптары бар деректермен оқыту.
2. Бақыланбайтын оқыту (Unsupervised learning) – құрылымы белгісіз деректерден заңдылықтарды табу.
3. Нықталандырумен оқыту (Reinforcement learning) – агент өз әрекеттерін ортамен әрекеттесу арқылы, марапат не жазалау сигналдарын ала отырып үйренеді.

RL — бұл тәжірибе арқылы үйрену, яғни агент өз әрекеттерінің салдарын байқап, ең пайдалы стратегияны табады. Ол нақты «дұрыс жауабы» жоқ жағдайларда қолданылады: мысалы, ойындарда, роботтың қозғалысында немесе инвестициялық шешімдер қабылдауда.

1. Нықталандырумен оқытудың негізгі компоненттері

RL жүйесі келесі негізгі элементтерден тұрады:

Компонент	Сипаттамасы
-----------	-------------

Компонент	Сипаттамасы
Агент (Agent)	Шешім қабылдайтын және әрекет жасайтын интеллектуалды жүйе.
Орта (Environment)	Агент әрекет ететін жүйе немесе әлем.
Күй (State)	Ағымдағы жағдайды сипаттайтын ақпарат.
Әрекет (Action)	Агенттің белгілі бір күйде жасай алатын қадамы.
Марапат (Reward)	Әрекет нәтижесінде алынған кері байланыс сигналы.
Саясат (Policy)	Күйге байланысты агенттің қандай әрекет жасайтынын анықтайтын стратегия.

Агент пен орта арасындағы өзара әрекет үздіксіз цикл түрінде жүреді:

CSS

[Агент] → әрекет → [Орта] → күй + марапат → [Агент]

Мақсат — ұзақ мерзімді жиынтық марапатты (cumulative reward) барынша арттыру.

2. Марапат және стратегия (Reward & Policy)

- Reward (R) — агентке оның әрекетінің қаншалықты пайдалы екенін көрсететін сигнал.
Мысалы: дұрыс әрекет → +1, қате әрекет → -1.
- Policy (π) — агенттің стратегиясы. Ол ағымдағы күйге қарай қандай әрекет таңдайтынын анықтайды.

RL жүйесінің негізгі мақсаты – ең тиімді саясатты табу, яғни ең үлкен жиынтық марапат беретін стратегияны анықтау:

$$\pi^* = \arg \max_{\pi} E \left[\sum_{t=0}^{\infty} \gamma^t R_t \right]$$

мұндағы γ — болашақ марапаттарды ескеретін дисконттау коэффициенті ($0 \leq \gamma \leq 1$).

3. Марков шешім процесі (MDP)

Нықталандырумен оқыту көбінесе Марков шешім процесі (Markov Decision Process, MDP) негізінде сипатталады. MDP мына компоненттерден тұрады:

- S — күйлер жиыны (states)
- A — әрекеттер жиыны (actions)
- $P(s'|s,a)$ — күй ауысу ықтималдығы
- $R(s,a)$ — марапат функциясы

Агент келесі ережемен әрекет етеді: ағымдағы күй мен ықтимал әрекеттер негізінде болашақтағы ең жоғары күтілетін марапатты қамтамасыз ететін әрекетті таңдайды.

4. Негізгі алгоритмдер

Q-learning – модельсіз нықталандыру алгоритмі. Ол әр күй мен әрекет жұбына Q-мәнін (value) үйренеді:

$$Q(s, a) = Q(s, a) + \alpha[R + \gamma \max_{a'} Q(s', a') - Q(s, a)]$$

Мақсаты – Q-функцияның оңтайлы мәнін үйрену және ең жоғары күтілетін сыйақы беретін әрекеттерді таңдау.

SARSA – Q-learning-ке ұқсас, бірақ келесі әрекетті агенттің нақты таңдаған әрекетіне негіздеп жаңартады:

$$Q(s, a) = Q(s, a) + \alpha[R + \gamma Q(s', a') - Q(s, a)]$$

5. Қолдану салалары

Нықталандырумен оқыту қазір көптеген салаларда табысты қолданылуда:

Сала	Қолдану мысалы
Робототехника	Роботтың қозғалысын үйрету, кедергілерден айналып өту.
Компьютерлік ойындар	AlphaGo, Chess AI — адамнан жоғары нәтиже көрсеткен жүйелер.
Қаржы және экономика	Инвестициялық портфельді басқару.
Энергетика және өндіріс	Ресурстарды оңтайлы бөлу.
Автономды көлік	Автопилот жүйелері.

6. Нықталандырумен оқыту мен басқа тәсілдердің айырмашылығы

Критерий	Бақыланатын оқыту	Бақыланбайтын оқыту	Нықталандырумен оқыту
Мәлімет	Белгіленген жауаптар бар	Тек кіріс деректер	Ортаның кері байланысы
Мақсат	Қате азайту	Үлгі табу	Марапатты арттыру
Қолдану мысалы	Сурет тану	Кластеризация	Роботты басқару

7. Қорытынды

Нықталандырумен оқыту – бұл адамның тәжірибе арқылы үйрену процесін модельдейтін жасанды интеллекттің қуатты бағыты.

Ол агент пен ортаның өзара байланысы арқылы үйренуге мүмкіндік береді және шешім қабылдау, басқару және стратегия құру міндеттерінде ерекше тиімді.

Қазіргі таңда RL:

- ойындарда адамнан мықты жүйелер жасауға (AlphaGo, Dota2 AI),
- өндірістік және роботтық процестерді автоматтандыруға,
- интеллектуалды шешім қабылдау жүйелерін дамытуға мүмкіндік беріп отыр.

Болашақта RL жасанды интеллекттің негізгі компоненттерінің бірі ретінде ақылды, өзін-өзі жетілдіретін жүйелердің дамуына негіз болады.

Пайдаланылған әдебиеттер

1. Sutton R.S., Barto A.G. Reinforcement Learning: An Introduction. — MIT Press, 2018.
2. Russell S., Norvig P. Artificial Intelligence: A Modern Approach. — Pearson, 2020.
3. Kaelbling L.P. Reinforcement Learning: A Survey. — Journal of AI Research, 1996.
4. Silver D. et al. Mastering the Game of Go with Deep Neural Networks and Tree Search. — Nature, 2016.
5. Тураев Б. Машиналық оқыту негіздері. — Алматы, 2022.