

№2 Дәріс. Марковтық шешімдер процесі

Дәріскер: Даркенбаев Д.К. PhD, қауымдастырылған профессор

Дәрістің мақсаты

студенттерге нықталандырумен оқытудағы (reinforcement learning) ең маңызды ұғымдардың бірі болып саналатын Марковтық шешімдер процесінің (MDP) мәнін, құрылымын және қолдану салаларын түсіндіру.

Кілт сөздер

Марковтық шешімдер процесі, күй (state), әрекет (action), ұтыс (reward), саясат (policy), оптимизация, Беллман қағидасы, динамикалық бағдарламалау.

Дәріс жоспары

1. Кіріспе. Нықталандырумен оқыту және шешім қабылдау модельдері
2. Марков процесінің негізгі идеясы
3. Марковтық шешімдер процесінің анықтамасы және құрылымы
4. MDP элементтері
5. Саясат және оның түрлері
6. Күтілетін ұтыс пен Беллман қағидасы
7. Оптималды саясат және оны табу жолдары
8. MDP-дің практикалық қолданылу салалары
9. Қорытынды
10. Пайдаланылған әдебиеттер

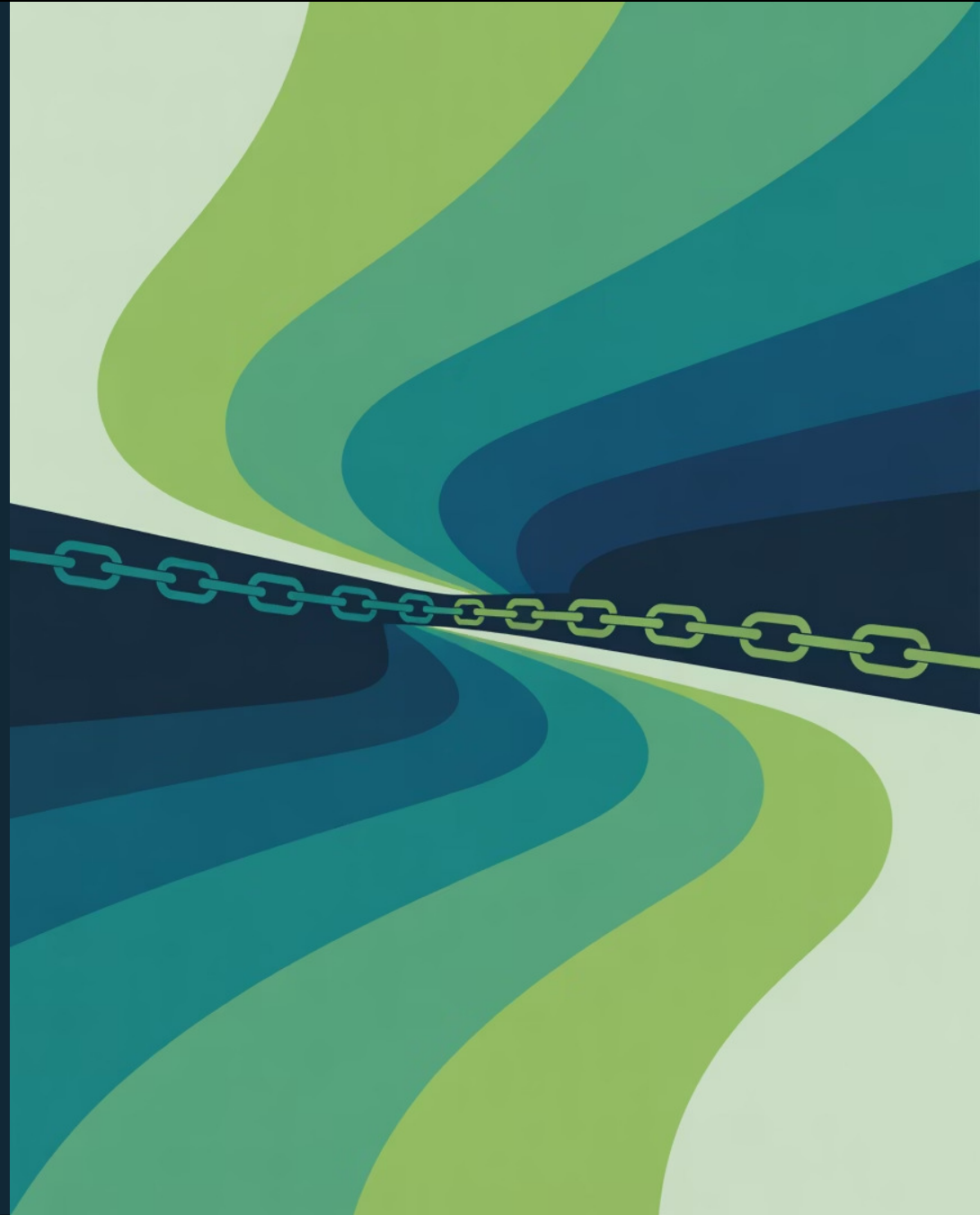
1. Кіріспе

Қазіргі заманғы жасанды интеллект (AI) жүйелерінің ең маңызды бағыттарының бірі — нықталандырумен оқыту (reinforcement learning, RL). Бұл бағыттың ерекшелігі — агент нақты ортада әрекет етіп, сол әрекеттерінің нәтижесінен сабақ алу арқылы тиімді шешім қабылдауды үйренеді. Агенттің мақсаты — ұзақ мерзімді ұтысты (reward) барынша арттыру. Мұндай процестерде агент пен орта арасындағы байланыс өте күрделі.

1. Марков процесінің негізгі идеясы

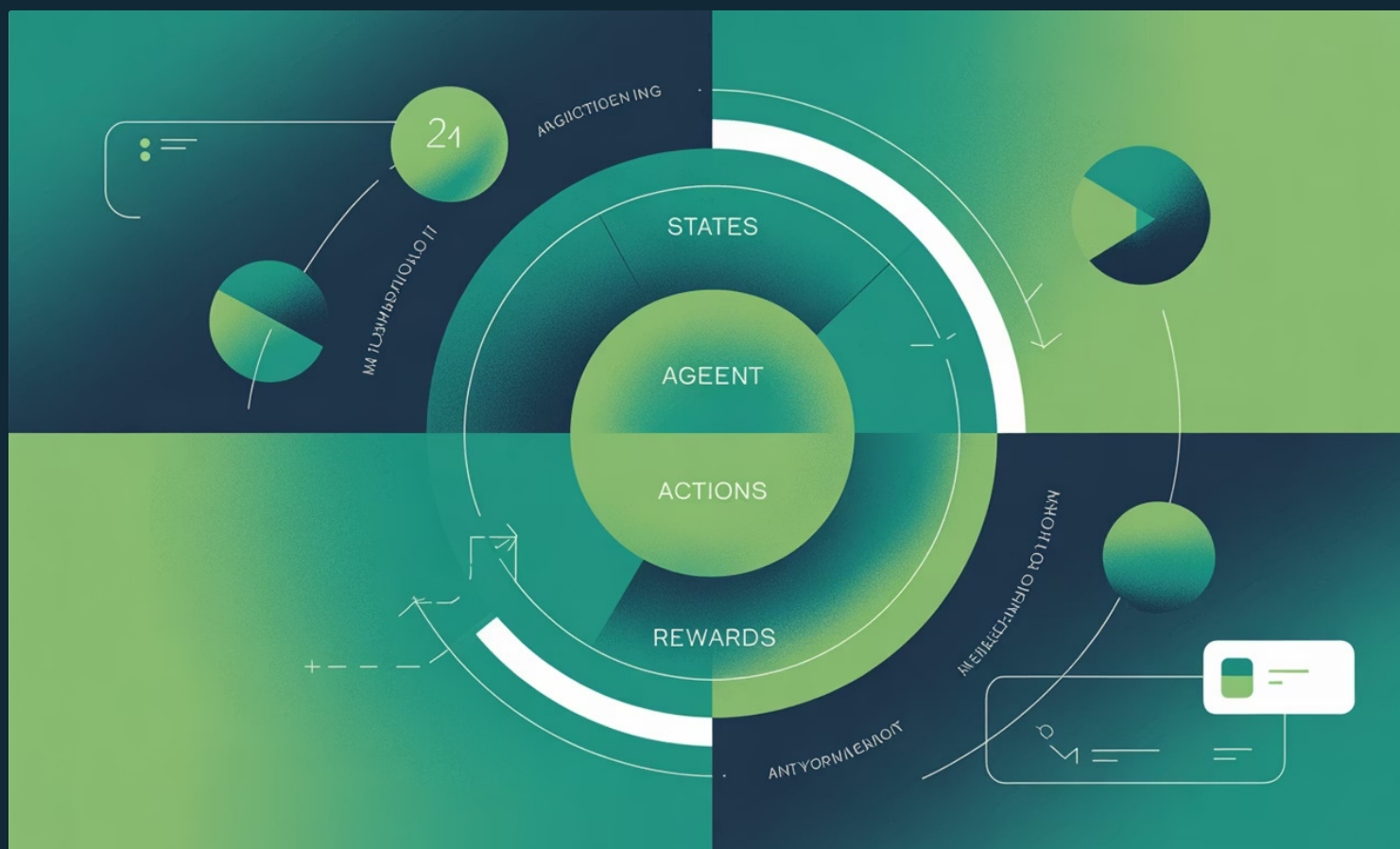
Марков процесінің басты идеясы — келесі күйдің тек ағымдағы күйге тәуелді болуы. Яғни, жүйенің болашақтағы жағдайы өткен кезеңдерге емес, тек қазіргі күйге байланысты анықталады.

❏ **Марков қасиеті** — көптеген стохастикалық (ықтималдыққа негізделген) модельдердің негізі болып табылады.



1. Марковтық шешімдер процесінің анықтамасы және құрылымы

Марковтық шешімдер процесі — бұл агенттің уақыт бойынша қабылдайтын шешімдерін сипаттайтын математикалық модель. Ол агенттің қоршаған ортамен өзара әрекеттесуін және әрбір әрекеттің нәтижесінде алынатын ұтысты (reward) бейнелейді.



MDP жүйесі бірнеше негізгі элементтерден тұрады:



Күйлер

жүйенің күйлері (states)



Әрекеттер

агенттің әрекеттері (actions)



Ықтималдықтар

келесі күйге өтудің ықтималдықтары (transition probabilities)



Ұтыс

әрбір әрекет үшін алынатын ұтыс (reward)



Дисконттау

уақыт өте келе ұтыстың маңызын өлшейтін дисконттау коэффициенті (discount factor)

1. MDP элементтері

Марковтық шешімдер процесінде агент келесі қадамдарды орындайды:

O1

Бақылау

Қоршаған ортаның ағымдағы күйін бақылайды.

O2

Таңдау

Белгілі бір саясатқа сәйкес әрекет таңдайды.

O3

Орындау

Таңдалған әрекетті орындайды және жаңа күйге өтеді.

O4

Ұтыс алу

Сол әрекет үшін ортадан ұтыс (reward) алады.



Агенттің негізгі мақсаты — ұзақ мерзімді перспективада жалпы ұтысты барынша арттыру.



1. Саясат және оның түрлері

Саясат (policy) — агенттің әрбір күйде қандай әрекет жасайтынын анықтайтын ереже.

Ол екі түрлі болады:

Детерминирленген саясат

әр күйде нақты бір әрекет орындалады

Стохастикалық саясат

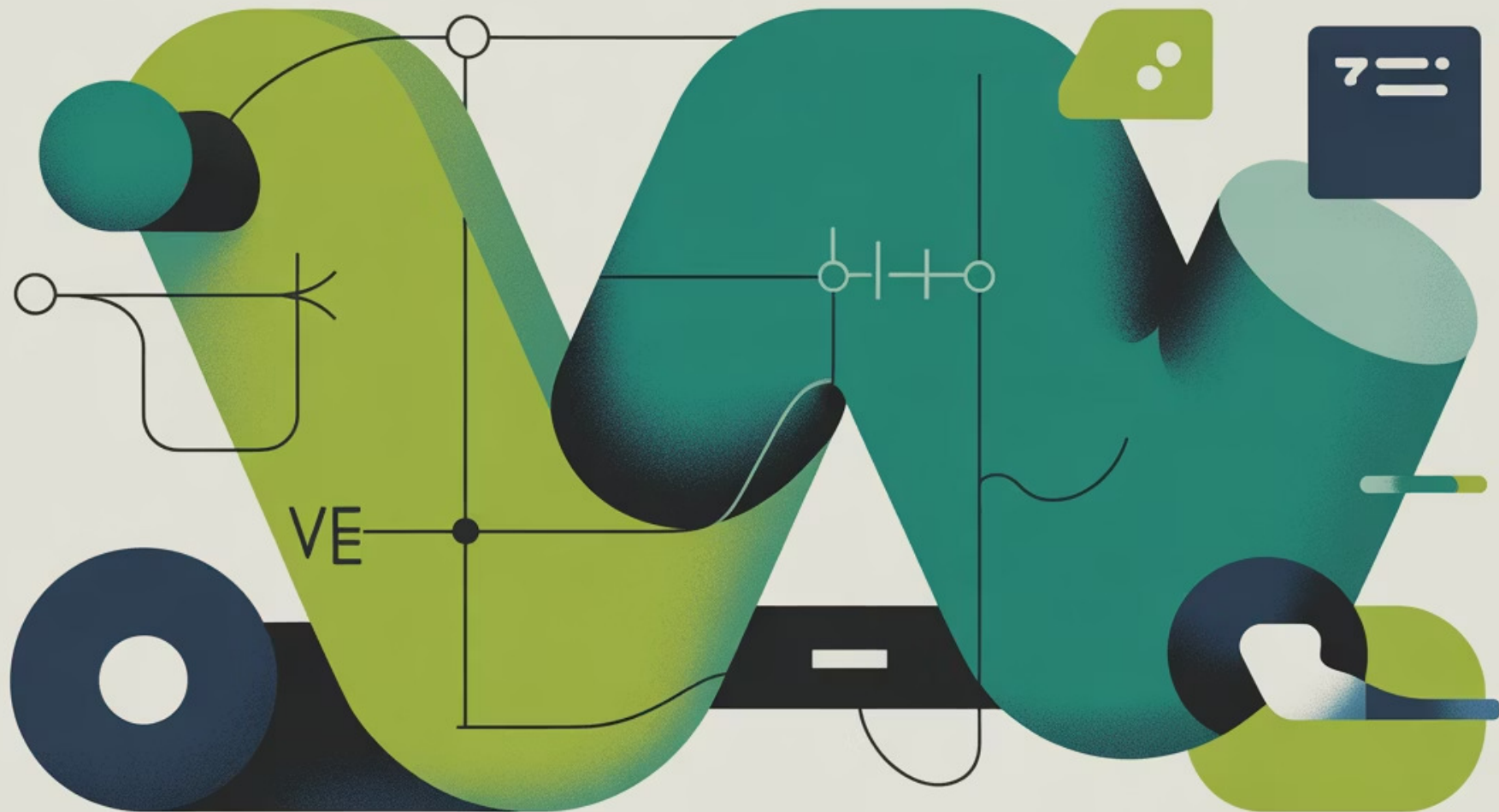
әрекет кездейсоқ түрде, ықтималдық негізінде таңдалады

Саясаттың сапасы оның жинақтайтын ұтысына қарай бағаланады.

1. Күтілетін ұтыс пен Беллман қағидасы

агенттің белгілі бір саясатты ұстанғанда ортадан алатын жалпы пайдасының өлшемі.

бұл шешім қабылдау процесінің негізін құрайтын тұжырым.



1. Оптималды саясат және оны табу жолдары

Оптималды саясат — барлық мүмкін саясаттардың ішінен ең үлкен жалпы ұтысқа әкелетін стратегия.

Оны табу үшін бірнеше әдістер қолданылады:



Динамикалық бағдарламалау

Dynamic Programming



Value Iteration

Value Iteration алгоритмі



Policy Iteration

Policy Iteration әдісі



Q-learning

Q-learning және басқа нықталандырумен оқыту тәсілдері

2. Қорытынды

Марковтық шешімдер процесі — нықталандырумен оқытудың ең негізгі, әрі әмбебап математикалық моделі болып табылады. Оның басты идеясы — агенттің кез келген әрекеті болашақтағы жағдайға ықпал етеді және бұл ықпал тек ағымдағы күй мен әрекетке байланысты анықталады. Бұл қағида агенттің мінез-құлқын модельдеуге, стратегияны (саясатты) таңдауға және тиімді шешім қабылдауға мүмкіндік береді.

1. Пайдаланылған әдебиеттер:

1. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd Edition). MIT Press.
2. Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research, 4, 237–285.
3. Puterman, M. L. (2014). Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley.
4. Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th Edition). Pearson.
5. Szepesvári, C. (2010). Algorithms for Reinforcement Learning. Morgan & Claypool Publishers.
6. Бахтин А. А., Кузнецов М. Н. (2020). Основы обучения с подкреплением. Санкт-Петербургский государственный университет.
7. Darkenbayev, D. K. (2024). Интеллектуальные системы и машинное обучение: основы и приложения. КазНУ, Алматы.

