№2 Дәріс. Марковтық шешімдер процесі (Markov Decision Process, MDP)

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. Әрбір шешім келесі жағдайға әсер етеді, ал келесі жағдай өз кезегінде келешектегі барлық шешімдердің нәтижесін өзгертеді. Сондықтан, мұндай жүйелерді талдау үшін математикалық тұрғыдан нақты, бірақ икемді модель қажет.

Осы тұрғыда Марковтық шешімдер процесі (Markov Decision Process — MDP) маңызды рөл атқарады. MDP агенттің ортада әрекет ету логикасын сипаттайтын негізгі математикалық негіз ретінде пайдаланылады. Ол агенттің таңдаған әрекеті мен ортаның жауап реакциясы арасындағы себеп-салдар байланысын анық көрсетеді.

MDP моделінің ерекшелігі — оның Марков қасиетіне негізделуі, яғни болашақтағы жағдай тек ағымдағы күй мен әрекетке ғана тәуелді болады. Бұл қарапайым көрінетін ұстаным көптеген күрделі басқару және жоспарлау жүйелерін сипаттауға мүмкіндік береді.

Марковтық шешімдер процесі тек жасанды интеллект саласында ғана емес, сонымен қатар экономикада, экологияда, медицинада, робототехникада, көлік жүйелерінде және қаржылық талдауда кеңінен қолданылады. Мысалы, роботтың жол бойымен кедергілерден өтуін, автономды көліктің бағдарын таңдауды немесе емделушіге ең тиімді терапияны ұсыну процесін MDP арқылы модельдеуге болады.

Сондықтан бұл тақырыпты меңгеру магистранттар мен зерттеушілер үшін ерекше маңызды. Ол болашақта интеллектуалды шешім қабылдау жүйелерін, автоматтандырылған басқару алгоритмдерін және оқитын агенттерді жобалауда теориялық және практикалық негіз береді.

2. 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.
Бұл қасиет Марков қасиеті деп аталады және ол көптеген стохастикалық (ықтималдыққа негізделген) модельдердің негізі болып табылады.

3. 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:

- жүйенің күйлері (states);
- агенттің әрекеттері (actions);
- келесі күйге өтудің ықтималдықтары (transition probabilities);
- әрбір әрекет үшін алынатын ұтыс (reward);
- уақыт өте келе ұтыстың маңызын өлшейтін дисконттау коэффициенті (discount factor).

4. MDP элементтері

Марковтық шешімдер процесінде агент келесі қадамдарды орындайды:

1. Қоршаған ортаның ағымдағы күйін бақылайды.
2. Белгілі бір саясатқа сәйкес әрекет таңдайды.
3. Таңдалған әрекетті орындайды және жаңа күйге өтеді.
4. Сол әрекет үшін ортадан ұтыс (reward) алады.

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

5. 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.

6. Күтілетін ұтыс пен Беллман қағидасы

Күтілетін ұтыс — агенттің белгілі бір саясатты ұстанғанда ортадан алатын жалпы пайдасының өлшемі. Беллман қағидасы — бұл шешім қабылдау процесінің негізін құрайтын тұжырым. Ол бойынша әрбір күйдің құндылығы ағымдағы ұтыс пен келесі қадамдардағы күтілетін ұтыстардың қосындысымен анықталады.

Яғни, агенттің бүгінгі таңдаған шешімі болашақтағы нәтижелерге тікелей әсер етеді. Бұл идея MDP негізінде саясатты жетілдіру әдістерін құруға мүмкіндік береді.

7. Оптималды саясат және оны табу жолдары

Оптималды саясат — барлық мүмкін саясаттардың ішінен ең үлкен жалпы ұтысқа әкелетін стратегия.
Оны табу үшін бірнеше әдістер қолданылады, мысалы:

- Динамикалық бағдарламалау (Dynamic Programming);
- Value Iteration алгоритмі;
- Policy Iteration әдісі;
- Q-learning және басқа нықталандырумен оқыту тәсілдері.

Бұл әдістер агенттің тәжірибесіне сүйене отырып, ортадағы ең тиімді әрекеттерді анықтауға мүмкіндік береді.

8. MDP-дің практикалық қолданылу салалары

Марковтық шешімдер процесі көптеген салаларда кеңінен қолданылады:

- Робототехника: роботтың қозғалысын жоспарлау және кедергілерден өту стратегиясын анықтау;
- Қаржы: инвестициялық шешімдер қабылдау және тәуекелді бағалау;
- Медицина: емдеу тактикасын жоспарлау;
- Көлік логистикасы: маршрутты оңтайландыру;
- Компьютерлік ойындар: жасанды интеллекттің мінез-құлқын басқару.

9. 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. MDP арқылы агент өз ортасынан алынған тәжірибені талдай отырып, ұзақ мерзімді табысты максималдауға ұмтылады.

Марковтық шешімдер процесі арқылы келесі маңызды мәселелер шешіледі:

- белгісіз ортада оңтайлы шешім қабылдау;
- ұзақ мерзімді стратегияларды жоспарлау;
- кездейсоқтық жағдайында тиімді әрекет таңдау;
- күрделі жүйелердің жұмысын модельдеу.

Бұл модельдің артықшылығы — оның әмбебаптығы мен икемділігінде. MDP түрлі салаларда — роботтарды басқару, қаржылық тәуекелді бағалау,

денсаулық сақтау шешімдерін жоспарлау, өндірістік процестерді оңтайландыру, интеллектуалды көлік жүйелерін жобалау және басқа да салаларда кеңінен қолданылады.

Сондықтан MDP тек теориялық ұғым емес, сонымен қатар заманауи жасанды интеллекттің және автоматтандырылған басқарудың іргелі құралы болып табылады.

Марковтық шешімдер процесін түсіну арқылы студенттер Reinforcement Learning, Q-learning, Deep Q-Network, Policy Gradient, Actor-Critic сияқты әдістердің логикасын терең ұғына алады.

Қорытындылай келе, MDP — бұл тек математикалық модель емес, ол ақылды шешім қабылдайтын жүйелердің жүрегі. Осы тақырыпты меңгеру — қазіргі цифрлық дәуірдегі интеллектуалды технологияларды түсінудің алғашқы және ең маңызды қадамы.

10. 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.