

№11 Дәріс. Актер–Критик әдістері

Дәрістің мақсаты:

Бұл дәрістің мақсаты — актер–критик әдісінің маңызын, құрылымын және жұмыс істеу қағидасын түсіндіру, сонымен қатар оның артықшылықтарын, қолдану салаларын көрсету.

Кілт сөздер:

Нығайтпалы оқыту, актер, критик, саясат, мән функциясы, агент, сыйақы, үйрену, стратегия, орта.

Дәріс жоспары:

1. Актер–Критик әдісінің жалпы түсінігі
2. Актер мен критиктің рөлі
3. Әдістің жұмыс істеу қағидасы
4. Артықшылықтары мен кемшіліктері
5. Қолданылу салалары
6. Қорытынды
7. Бақылау сұрақтары
8. Пайдаланылған әдебиеттер

Кіріспе

Нығайтпалы оқытуда (Reinforcement Learning) агенттің мақсаты — қоршаған ортамен өзара әрекеттесіп, ең жоғары жиынтық сыйақыға қол жеткізу. Классикалық тәсілдер (мысалы, Q-learning немесе SARSA) көбінесе мән функциясын үйренуге бағытталған болса, саясатты градиенттік әдістер агенттің әрекет ету стратегиясын тікелей үйретеді.

Актер–Критик әдістері — осы екі тәсілдің артықшылықтарын біріктіретін, заманауи және тиімді нығайтпалы оқыту әдісі. Бұл тәсілдің басты идеясы — агенттің әрекет стратегиясын (саясатын) және мән функциясын қатар үйрету. Мұнда екі негізгі компонент бар:

- Актер (Actor) — агенттің әрекет ету саясатын басқарады, яғни берілген күйде қандай әрекет таңдау керектігін анықтайды.
- Критик (Critic) — актердің әрекеттерін бағалап, олардың қаншалықты тиімді болғанын есептейді.

1. Актер–Критик әдісінің жалпы түсінігі

Актер–Критик әдісі — саясатты градиенттік және мән функциясына негізделген тәсілдердің үйлесімі. Бұл әдісте агент тек қандай әрекетті таңдау керек екенін үйреніп қана қоймай, сол әрекеттің тиімділігін де бағалайды.

Мұндай тәсілдің мақсаты — саясатты тікелей жақсартып отырып, үйренудің тұрақтылығын арттыру. Классикалық саясатты градиенттік әдістерде саясатты жақсарту кезінде үлкен өзгерістер орын алып, үйрену процесі тұрақсыз болуы мүмкін. Ал актер–критик құрылымы мұндай тұрақсыздықты азайтады.

2. Актер мен критиктің рөлі

Актер (Actor) — бұл агенттің саясат моделі. Ол берілген күйге сәйкес әрекетті таңдайды. Актердің рөлі — саясатты біртіндеп жақсарту арқылы ұзақ мерзімді сыйақыны арттыру.

Критик (Critic) — бұл агенттің бағалау бөлігі. Ол актер ұсынған әрекеттің тиімділігін есептейді, яғни қаншалықты дұрыс шешім қабылданғанын бағалайды. Критиктің нәтижесі актердің саясатын жаңартуға бағыт береді.

Басқаша айтқанда, актер әрекет етеді, критик бағалайды.

3. Әдістің жұмыс істеу қағидасы

Актер–Критик әдісі келесі қағида бойынша жұмыс істейді:

1. Агент ортаның қазіргі күйін бақылайды.
2. Актер берілген күйде қандай әрекет жасау керектігін таңдайды.
3. Әрекет орындалғаннан кейін агент жаңа күйге өтеді және ортадан сыйақы алады.
4. Критик алынған сыйақыға қарап, бұл әрекеттің тиімділігін бағалайды.
5. Егер әрекет жақсы нәтиже берсе, актер сол бағыттағы саясатты нығайтады, ал егер нәтиже нашар болса, саясатты өзгертеді.

Осылайша, актер мен критик бір-бірін толықтырып, агенттің мінез-құлқын үздіксіз жетілдіріп отырады.

4. Артықшылықтары мен кемшіліктері

Артықшылықтары:

- Үйренудің тұрақтылығын қамтамасыз етеді;
- Үздіксіз және көпөлшемді ортада тиімді;
- Саясат пен мән функциясын қатар жақсартады;
- Үлкен және күрделі жүйелерде оңтайлы шешімдер табуға қабілетті.

Кемшіліктері:

- Үйрену процесі күрделі және есептеу шығыны жоғары;
- Критиктің дұрыс бағаламауы актердің саясатын бұрмалай алады;
- Гиперпараметрлерді дұрыс таңдау нәтижеге әсер етеді.

5. Қолданылу салалары

Актер–Критик әдістері түрлі қолданбалы салаларда кеңінен пайдаланылады:

- Робототехника: қозғалыс пен тепе-теңдікті басқару;
- Қаржы: инвестициялық стратегияларды оңтайландыру;
- Көлік жүйелері: автономды басқару және маршрутты жоспарлау;
- Ойын индустриясы: күрделі стратегияларды үйрету;
- Энергетика және өнеркәсіп: процестерді автоматтандыру және оңтайландыру.

Мысалы, DeepMind компаниясы жасаған AlphaGo жүйесінде актер–критик тәсілдері нейрондық желілермен біріктіріліп, адамнан жоғары деңгейде ойнауға мүмкіндік берді.

6. Қорытынды

Актер–Критик әдістері нығайтпалы оқытудың ең қуатты және заманауи бағыттарының бірі. Бұл тәсіл саясатты градиенттік әдістердің икемділігін

және мән функциясына негізделген әдістердің тұрақтылығын біріктіреді. Оның басты ерекшелігі — агентті нақты өмірлік ортада әрекет етуге үйретіп, шешім қабылдау сапасын біртіндеп арттыруында. Қазіргі заманғы интеллектуалды жүйелердің көпшілігі (мысалы, автономды роботтар, автоматты сауда жүйелері және адаптивті басқару жүйелері) актер–критик тәсілдеріне сүйенеді.

Бақылау сұрақтары

1. Актер–Критик әдісі дегеніміз не?
2. Актер мен критиктің рөлі неде?
3. Бұл әдіс саясатты градиенттік тәсілдерден несімен ерекшеленеді?
4. Неліктен актер–критик құрылымы үйренудің тұрақтылығын арттырады?
5. Әдістің негізгі артықшылықтары мен кемшіліктері қандай?
6. Актер–критик тәсілдері қандай нақты салаларда қолданылады?
7. DeepMind жүйелерінде бұл әдіс қалай қолданылған?

Пайдаланылған әдебиеттер

1. Sutton, R. S., & Barto, A. G. *Reinforcement Learning: An Introduction*. MIT Press, 2018.
2. Konda, V. R., & Tsitsiklis, J. N. *Actor-Critic Algorithms*. NIPS, 2000.
3. Lillicrap, T. et al. *Continuous Control with Deep Reinforcement Learning*. arXiv, 2015.
4. Silver, D. et al. *Deterministic Policy Gradient Algorithms*. ICML, 2014.
5. Schulman, J. et al. *Proximal Policy Optimization Algorithms*. arXiv preprint, 2017.