

№6 Дәріс. Температуралық айырмашылық әдісі (TD Learning)

Дәрістің мақсаты:

Бұл дәрістің мақсаты — уақытша айырмашылық (TD Learning) әдісінің мәнін, оның күшейтумен оқыту (reinforcement learning) жүйесіндегі орнын түсіндіру. Сондай-ақ бұл әдістің болашақтағы марапаттарды болжауда, агенттің тәжірибесі негізінде үйрену қабілетін арттырудағы рөлін көрсету.

Кілт сөздер:

Уақытша айырмашылық әдісі, күшейтумен оқыту, агент, күй, әрекет, марапат (reward), тәжірибеден үйрену, болжау, итеративтік оқыту, орта моделінсіз үйрену.

Дәріс жоспары:

1. Кіріспе: Күшейтумен оқытудағы TD әдісінің орны
2. TD әдісінің негізгі идеясы
3. TD және Monte Carlo әдістерінің салыстырмалы ерекшеліктері
4. TD әдісінің түрлері және қолдану салалары
5. TD әдісінің артықшылықтары мен шектеулері
6. Қорытынды
7. Бақылау сұрақтары

1. Кіріспе: Күшейтумен оқытудағы TD әдісінің орны

Күшейтумен оқыту (reinforcement learning) — агент пен ортаның өзара әрекеттесуі негізінде шешім қабылдауды үйрететін жасанды интеллект бағыты. Агент нақты тәжірибе жинақтап, әрбір қадамда алған марапат (reward) арқылы өз әрекеттерін жетілдіреді.

Температуралық айырмашылық (Temporal Difference, TD) әдісі — осы күшейтумен оқытудың ең маңызды құрамдас бөлігі. TD әдісі агенттің бір күйден екінші күйге өткен кездегі тәжірибесінен үйренуіне негізделеді. Ол Monte Carlo сияқты әдістерге ұқсас, бірақ айырмашылығы — ол марапат туралы толық эпизодты күтпей-ақ, ағымдағы күйден алынған жаңа ақпаратқа сүйене отырып болжамды жаңарта алады.

Яғни агент әр қадам сайын өзінің бұрынғы болжамдарын нақты тәжірибе негізінде түзетіп отырады.

2. TD әдісінің негізгі идеясы

TD әдісінің басты идеясы — агент болашақтағы күтілетін марапатты болжай отырып, ағымдағы бағалауын жаңа тәжірибеге қарай жақындатады. Бұл тәсіл «оқу арқылы түзету» қағидасына сүйенеді.

TD әдісінде агент әрбір әрекеттен кейін келесі күй мен алынған марапатты көреді және соның негізінде өз бағасын (мысалы, күйлердің мәнін) аздап өзгертеді. Осылайша, агент уақыт өте келе өз шешімдерін жетілдіріп, тиімді стратегия қалыптастырады.

TD әдісін «оқыту кезінде тәжірибені біртіндеп талдап, болжау қабілетін жетілдіру тәсілі» деп атауға болады.

3. TD және Monte Carlo әдістерінің салыстырмалы ерекшеліктері
Monte Carlo әдісі де тәжірибе негізінде үйренуді көздейді, бірақ ол толық эпизод аяқталғаннан кейін ғана нәтижені жаңарта алады. Бұл кейде көп уақытты талап етеді және орта күрделі болған жағдайда тиімсіз болуы мүмкін.

TD әдісі болса:

- Жаңа мәліметтер түскен сайын бірден үйренеді;
- Толық эпизодты күтпейді;
- Қате мен айырмашылықты (яғни уақытша айырмашылықты) біртіндеп азайта отырып, болжамды нақтылайды.

Сондықтан TD әдісі динамикалық және бейімделгіш оқыту тәсілі болып табылады.

4. TD әдісінің түрлері және қолдану салалары

TD әдісінің бірнеше кеңінен қолданылатын нұсқалары бар:

- TD(0) — ең қарапайым нұсқасы, бір қадам алға болжам жасайды.
- TD(λ) — өткен тәжірибенің әсерін біртіндеп азайту арқылы ұзақ мерзімді оқытуды қамтамасыз етеді.
- SARSA және Q-learning сияқты әдістер TD тәсілінің практикалық қолданыстары болып табылады.

Қолдану салалары:

- Ойындарда (мысалы, шахмат, Go, роботтық футбол және т.б.);
- Өндірістік процестерді оптимизациялауда;
- Автономды роботтардың қозғалысын басқаруда;
- Энергия жүйелерін тиімді басқаруда;
- Болжам мен жоспарлау жүйелерінде.

5. TD әдісінің артықшылықтары мен шектеулері

Артықшылықтары:

- Тәжірибе барысында нақты уақытта үйренуге мүмкіндік береді;
- Ортаның толық моделін білуді талап етпейді;
- Болашақтағы марапаттарды тиімді бағалауға көмектеседі;
- Жылдам бейімделуге және стратегияны түзетуге мүмкіндік береді.

Шектеулері:

- Оқу процесінде параметрлердің (мысалы, оқыту қадамы немесе зерттеу деңгейі) дұрыс таңдалмауы тұрақсыздыққа әкелуі мүмкін;
- Күрделі орталарда үйрену баяу жүруі ықтимал;
- Кейде толық марапат құрылымы белгісіз болса, қате жинақталуы мүмкін.

6. Қорытынды

Температуралық айырмашылық әдісі — тәжірибе арқылы болашақты болжауға және шешім қабылдауды жетілдіруге бағытталған қуатты оқыту тәсілі. Ол агентке орта туралы нақты модельсіз үйренуге, болашақ нәтижелерді бағалауға және өз әрекетін үздіксіз жетілдіруге мүмкіндік береді.

TD әдісі күшейтумен оқыту саласының ең маңызды құралдарының бірі ретінде қарастырылады. Оның идеясы — жүйе өз қателігін біртіндеп азайта

отырып, оңтайлы мінез-құлыққа жақындайды. Бүгінгі таңда TD әдістері жасанды интеллект, робототехника, қаржы болжамдары, өндірісті басқару және энергетикалық жүйелерді оңтайландыру сияқты көптеген салаларда табысты қолданылуда.

Бақылау сұрақтары:

1. Температуралық айырмашылық (TD) әдісінің негізгі идеясы неде?
2. TD әдісі мен Monte Carlo әдісінің айырмашылығы қандай?
3. TD әдісінде агент тәжірибеден қалай үйренеді?
4. TD(0) және TD(λ) әдістері қандай жағдайда қолданылады?
5. TD әдісі қандай артықшылықтарға ие?
6. Бұл әдіс нақты қандай салаларда кеңінен пайдаланылады?
7. TD әдісінде орта моделі не үшін қажет емес?
8. TD әдісі күшейтумен оқытудың жалпы құрылымында қандай рөл атқарады?

Пайдаланылған әдебиеттер

1. Sutton, R. S., & Barto, A. G. *Reinforcement Learning: An Introduction*. – 2nd Edition. – MIT Press, 2018.
2. Kaelbling, L. P., Littman, M. L., & Moore, A. W. *Reinforcement Learning: A Survey*. – Journal of Artificial Intelligence Research, 1996.
3. Bertsekas, D. P., & Tsitsiklis, J. N. *Neuro-Dynamic Programming