



№7 Дәріс. SARSA және Q-learning әдістері

Дәріскер: Даркенбаев Д.К. PhD, қауымдастырылған профессор

Дәрістің мақсаты

Күшейтілген оқытудағы негізгі әдістерді меңгеру

Бұл дәрістің мақсаты — күшейтілген оқытудағы (Reinforcement Learning) ең кең таралған екі әдіс — **SARSA және Q-learning** туралы түсінік қалыптастыру, олардың жұмыс істеу қағидаларын, айырмашылықтарын және қолданылу салаларын талдау.

Сонымен қатар студенттерді **температуралық айырмашылыққа (Temporal Difference)** негізделген үйрену тәсілдерінің практикалық мәнімен таныстыру.

Кілттік сөздер

Негізгі ұғымдар

- Күшейтілген оқыту
- Агент
- Орта
- Күй
- Әрекет
- Марапат

Саясат түрлері

- Саясат (policy)
- On-policy
- Off-policy
- Үйрену стратегиясы

Әдістер

- Температуралық айырмашылық
- SARSA
- Q-learning
- Тәжірибеден үйрену

Қолданбалар

- Интеллектуалды жүйе
- Робототехника
- Автоматтандыру

Дәріс жоспары

О1

Кіріспе

- Күшейтілген оқыту ұғымы
- Температуралық айырмашылық әдістерінің рөлі
- SARSA және Q-learning пайда болу тарихы

О2

SARSA әдісінің мәні және ерекшеліктері

- SARSA атауының мәні
- On-policy принципі
- SARSA әдісінің артықшылықтары мен кемшіліктері

О3

Q-learning әдісінің мәні және ерекшеліктері

- Q-мән түсінігі
- Off-policy тәсілінің ерекшелігі
- Q-learning әдісінің тиімді жақтары

О4

SARSA және Q-learning әдістерінің айырмашылығы

- Салыстырмалы талдау
- Тиімді қолдану жағдайлары

О5

Қолданылу салалары

- Робототехника
- Қаржы жүйелері
- Интеллектуалды басқару жүйелері
- Компьютерлік ойындар мен симуляциялар

О6

SARSA және Q-learning әдістерінің маңыздылығы мен дамуы

- Қазіргі заманғы жасанды интеллекттегі рөлі
- Deep Q-Network (DQN) әдісімен байланысы

О7

Қорытынды және бақылау сұрақтары

Кіріспе

Күшейтілген оқыту (Reinforcement Learning, RL) саласында агент өз ортасымен өзара әрекеттесу арқылы мақсатқа жетудің ең тиімді жолын үйренеді. Бұл бағыттың негізгі мақсаты – тәжірибе арқылы мінез-құлықты оңтайландыру, яғни агент ұзақ мерзімді марапатты барынша арттыруға ұмтылады.

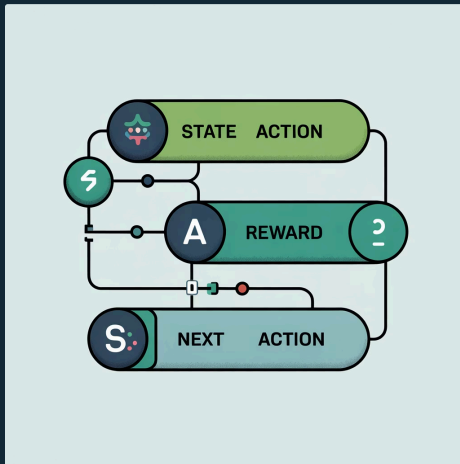
Осы процесте агент әрекет етіп, нәтижесін бақылайды және алған сыйақысы арқылы өз стратегиясын жетілдіреді. Мұндай жүйелердің дамуына **SARSA және Q-learning** әдістері зор үлес қосқан.

Бұл әдістер **Температуралық айырмашылық (TD learning)** қағидаларына негізделеді және саясатты жетілдірудің тиімді тәсілдерін ұсынады.

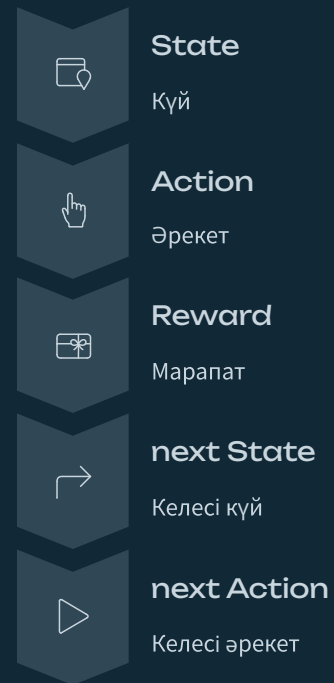
SARSA және Q-learning — күшейтілген оқыту алгоритмдерінің ең танымал және тәжірибеде кең қолданылатын түрлері. Олар робототехникада, ойындарда, қаржы саласында және интеллектуалды басқару жүйелерінде кеңінен пайдаланылады.



SARSA әдісінің мәні



SARSA атауы ағылшын тіліндегі бес элементтің қысқартылған нұсқасы:



Бұл әдіс агенттің нақты таңдаған әрекетіне және саясатына тікелей тәуелді. Яғни, **SARSA on-policy** (саясатқа негізделген) тәсіл болып табылады.

SARSA әдісі агенттің өзі таңдаған стратегиясын ескере отырып, келесі әрекет нәтижесіне қарай оқуды жүзеге асырады. Мысалы, егер агент зерттеу және пайдалану тепе-теңдігін сақтай отырып әрекет етсе (мысалы, ϵ -greedy стратегиясы), онда жаңарту дәл осы саясатқа сәйкес жүзеге асады.

- ❑ **SARSA-ның басты ерекшелігі** – агенттің нақты тәжірибесін негізге алуы. Яғни, ол өз әрекетінен тікелей үйренеді және сол саясатты жетілдіреді. Бұл тәсіл тұрақты және нақты жағдайларда қауіпсіз үйренуді қамтамасыз етеді.

Q-learning әдісінің мәні

Q-learning — күшейтілген оқытудағы ең кең таралған және тиімді әдістердің бірі. Ол да температуралық айырмашылық идеясына негізделеді, бірақ SARSA-дан айырмашылығы – **off-policy** тәсілді қолданады.

Яғни, агент ортаны зерттеу кезінде басқа саясат бойынша әрекет ете алады, бірақ оқыту үдерісі ең жақсы (оптималды) саясатқа бағытталады.

QN-10	№. Action			8 – 4670	
S1-4204		2tiat	0ft et	8 – 9050	2033
				8 – 4030	
		↑	22 en	8 – 4650	0003
		OPTIMAN			
S1-8323			2622	8 – 4830	1033
1.333		0tato	9hen	8 – 41.20	

Q-мәнін бағалау

Q-learning әдісінде агент болашақтағы ең тиімді әрекетті болжауға тырысады. Ол нақты бір күй мен әрекет жұбы үшін "Q-мәнін" бағалай отырып, максималды болашақ марапатты ескеріп шешім қабылдайды.

Басты артықшылығы

Q-learning тәсілінің басты артықшылығы – ол тіпті зерттеу кезінде кездейсоқ әрекеттер жасаса да, ұзақ мерзімді тиімді саясатты үйренуге қабілетті.

Кең қолданылуы

Бұл оны көптеген нақты қолданбаларда — ойындар, автономды көлік құралдары, өндірістік процестерді басқару сияқты бағыттарда — кеңінен қолдануға мүмкіндік береді.

SARSA мен Q-learning арасындағы айырмашылықтар

Салыстыру өлшемі	SARSA	Q-learning
Саясат түрі	On-policy (нақты саясатпен үйрену)	Off-policy (оптималды саясатпен үйрену)
Әрекет таңдау	Үйрену барысында нақты орындалған әрекетті қолданады	Ең жақсы ықтимал әрекетті ескереді
Тұрақтылық	Көбіне тұрақты, бірақ баяу үйренеді	Жылдам үйренеді, бірақ тұрақтылығы аздау
Қолданылу саласы	Қауіпсіздік пен тұрақтылық маңызды жүйелерде	Ең жоғары өнімділік қажет жүйелерде

SARSA мысалы

Роботты басқаруда SARSA әдісі қауіпсіз және біртіндеп үйренуге мүмкіндік береді

Q-learning мысалы

Жылдам шешім қабылдау қажет ойын немесе симуляция жағдайларында тиімді

SARSA және Q-learning әдістерінің қолданылу салалары



Робототехникада

Роботтардың қозғалысын, тепе-теңдік ұстауын және навигациясын үйрету.



Қаржы саласында

Инвестициялық шешімдерді оңтайландыру және тәуекелдерді азайту.



Интеллектуалды басқару жүйелерінде

Өндірістік немесе энергетикалық процестердің тиімділігін арттыру.



Ойындар мен симуляциялар

Компьютерлік агенттерді өзіндік стратегия арқылы ойнауға үйрету.

Бұл әдістер күрделі ортада үздіксіз үйренуге және бейімделуге мүмкіндік береді.

SARSA және Q-learning әдістерінің МАҢЫЗДЫЛЫҒЫ

Негізгі рөлі

SARSA және Q-learning әдістері күшейтілген оқытудың негізін құрайды және көптеген заманауи алгоритмдердің бастамасы болып табылады.

Дамуы

Q-learning әдісі кейінірек **Deep Q-Network (DQN)** секілді терең оқыту тәсілдерімен біріктіріліп, жасанды интеллект дамуының жаңа кезеңін ашты.

- ❑ **SARSA** болса, тұрақтылық пен қауіпсіз үйренудің маңызды принциптерін көрсетіп, нақты инженерлік жүйелер мен физикалық құрылғыларды оқытуда қолданылады.

SARSA және Q-learning — бір-бірін толықтыратын тәсілдер



SARSA

Нақты саясатты жетілдіреді және қауіпсіз үйренуді қамтамасыз етеді



Q-learning

Оптималды саясатты іздейді және жылдам нәтижеге жетеді

Екі әдіс те агенттің тәжірибесі мен ортаның реакциясына сүйене отырып, күрделі жүйелерді өздігінен үйретуге мүмкіндік береді.

Бүгінгі таңда бұл әдістер жасанды интеллект, автоматтандыру және робототехника салаларында кеңінен қолданылып келеді.

Бақылау сұрақтары

- 1 SARSA әдісінің негізгі идеясы неде?
- 2 SARSA мен Q-learning әдістерінің басты айырмашылығы қандай?
- 3 Q-learning неліктен off-policy әдісі деп аталады?
- 4 SARSA әдісін қандай жағдайларда қолдану тиімді?
- 5 Q-learning әдісі қандай салаларда кеңінен қолданылады?
- 6 Бұл әдістердің күшейтілген оқытудың дамуындағы рөлі қандай?
- 7 SARSA мен Q-learning әдістерін біріктіру мүмкін бе?

Пайдаланылған әдебиеттер

1

Sutton, R. S., & Barto, A. G.
Reinforcement Learning: An Introduction.
MIT Press, 2018.

2

**Kaelbling, L. P., Littman, M. L., & Moore,
A. W.** Reinforcement Learning: A Survey.
Journal of Artificial Intelligence Research,
1996.

3

Russell, S., & Norvig, P. Artificial
Intelligence: A Modern Approach.
Pearson, 2021.

4

**Busoniu, L., Babuska, R., De Schutter, B.,
& Ernst, D.** Reinforcement Learning and
Dynamic Programming Using Function
Approximators. CRC Press, 2010.

5

Silver, D. Lecture Notes on
Reinforcement Learning. University
College London, 2020.