

№10 Дәріс. Политика градиенті әдістері (Policy Gradient Methods)

Дәріс мақсаты:

Бұл дәрістің мақсаты — саясатты градиенттік әдістердің мәнін, жұмыс істеу қағидаларын, олардың артықшылықтары мен қолданылу салаларын түсіндіру.

Кілт сөздер:

Саясат, градиент, агент, әрекет, күйлер, REINFORCE, Actor-Critic, PPO, нығайтпалы оқыту, стратегия.

Дәріс жоспары:

1. Саясат ұғымы және оның түрлері
2. Саясатты градиенттік әдістердің негізгі идеясы
3. Саясатты тікелей үйретудің артықшылықтары
4. Стохастикалық және детерминистік саясат
5. Саясатты градиенттік әдістердің негізгі алгоритмдері
6. Артықшылықтар мен кемшіліктері
7. Қолданылу мысалдары және практикалық маңыздылығы

Кіріспе

Саясатты градиенттік әдістер — нығайтпалы оқыту (Reinforcement Learning) саласындағы ең маңызды және заманауи бағыттардың бірі. Бұл әдістердің басты мақсаты — агенттің әрекет ету саясатын тікелей үйрету, яғни агент ортада қандай жағдайда қандай әрекет жасау керектігін үйренеді.

Классикалық әдістер (мысалы, Q-learning немесе SARSA) әрекет мәнін үйренуге негізделсе, саясатты градиенттік әдістер *саясат функциясын тікелей жақсартуға* бағытталған. Мұндай тәсілдер бейсызық және үздіксіз әрекет кеңістігі бар күрделі ортада тиімді нәтиже береді.

Қазіргі таңда саясатты градиенттік әдістер автономды көліктерді басқару, робототехника, ойын стратегияларын үйрету және күрделі шешім қабылдау жүйелерінде кеңінен қолданылады.

1. Саясат ұғымы және оның түрлері

Саясат (policy) — агенттің нақты бір күйде қандай әрекет таңдауы керектігін сипаттайтын ереже немесе функция.

Саясаттың екі негізгі түрі бар:

- Детерминистік саясат: әрбір күй үшін нақты бір әрекет таңдалады.
- Стохастикалық саясат: әрекет ықтималдық негізінде таңдалады, яғни әр күйде бірнеше әрекеттің болуы мүмкін.

2. Саясатты градиенттік әдістердің негізгі идеясы

Бұл әдістердің негізгі идеясы — агенттің саясатын параметрлер арқылы сипаттау және осы параметрлерді мақсатты функцияны (мысалы, күтілетін ұпай) барынша арттыру үшін өзгерту.

Яғни, агент тәжірибе жинай отырып, қай әрекет нәтижесінде үлкен сыйақы алатынын үйренеді және сол бағытта өз саясатын жақсартады.

3. Саясатты тікелей үйретудің артықшылықтары

Классикалық әдістер әрекет мәнін есептеуге тәуелді болса, саясатты градиенттік әдістер мұндай шектеулерден бос. Бұл тәсілдер келесі артықшылықтарға ие:

- Үздіксіз әрекет кеңістігінде тиімді жұмыс істейді;
- Күрделі стратегияларды тікелей модельдеуге мүмкіндік береді;
- Үлкен өлшемді және бейсызық ортада тұрақты үйренуді қамтамасыз етеді;
- Саясаттың кез келген пішінін (мысалы, нейрондық желі арқылы) бейнелей алады.

4. Стохастикалық және детерминистік саясат

Саясатты градиенттік әдістерде саясаттың түрі өте маңызды рөл атқарады.

- Стохастикалық саясатта агент әр күйде әртүрлі әрекетті белгілі ықтималдықпен таңдайды. Бұл зерттеу мен пайдалану арасындағы теңгерімді сақтауға мүмкіндік береді.
- Детерминистік саясатта бір күйге бір ғана әрекет сәйкес келеді. Бұл тәсіл көбіне үздіксіз басқару міндеттерінде қолданылады.

5. Саясатты градиенттік әдістердің негізгі алгоритмдері

Саясатты градиенттік тәсілдер көптеген нақты алгоритмдердің негізін қалайды:

- REINFORCE әдісі: ең қарапайым саясатты градиенттік тәсіл, онда агент эпизодтардан кейін саясатын жаңартады.
- Actor-Critic әдістері: саясатты (actor) және мән функциясын (critic) біріктіріп қолданады, нәтижесінде оқу процесі тұрақты және тиімді болады.
- Proximal Policy Optimization (PPO): заманауи және ең тиімді әдістердің бірі, үйрену кезінде саясаттың өзгеру ауқымын шектейді.

6. Артықшылықтар мен кемшіліктері

Артықшылықтары:

- Үздіксіз және көпөлшемді ортада тиімді;
- Қолданылатын саясат икемді және бейімделгіш;
- Қатты шамадан тыс үйренуді азайтады.

Кемшіліктері:

- Үйрену процесі баяу және есептеу тұрғысынан қымбат болуы мүмкін;
- Градиенттің шашырауы (variance) жоғары болуы ықтимал;
- Гиперпараметрлерді дұрыс таңдамау оқу сапасын төмендетеді.

7. Қолданылу мысалдары және практикалық маңыздылығы

Саясатты градиенттік әдістер көптеген салаларда қолданылады:

- Автономды роботтардың қозғалысын басқару;
- Қаржы нарығында стратегияларды оңтайландыру;
- Компьютерлік ойындардағы интеллектуалды агенттерді үйрету;
- Өндіріс процесін автоматтандыру және логистикалық шешім қабылдау.

Бұл әдістер күрделі, көпөлшемді және бейсызық орталарда агенттің мінез-құлқын үйретудің ең қуатты тәсілдерінің бірі болып саналады.

Қорытынды

Саясатты градиенттік әдістер — қазіргі заманғы нығайтпалы оқытудың ең тиімді бағыттарының бірі. Олар агентке тікелей саясатты үйретуге мүмкіндік беріп, үздіксіз әрекет кеңістігінде тұрақты және дәл шешімдер қабылдауға жол ашады. Бұл тәсілдердің дамуы болашақта интеллектуалды жүйелердің жаңа деңгейіне шығуына мүмкіндік береді.

Бақылау сұрақтары

1. Саясат ұғымы дегеніміз не және оның түрлері қандай?
2. Саясатты градиенттік әдістердің негізгі идеясы неде?
3. Неліктен саясатты тікелей үйрету маңызды?
4. Стохастикалық және детерминистік саясаттың айырмашылығы неде?
5. Саясатты градиенттік әдістердің қандай негізгі алгоритмдері бар?
6. Бұл әдістердің артықшылықтары мен кемшіліктері қандай?
7. Қай салаларда саясатты градиенттік әдістер жиі қолданылады?

Пайдаланылған әдебиеттер

1. Sutton, R. S., & Barto, A. G. *Reinforcement Learning: An Introduction*. MIT Press, 2018.
2. Silver, D. et al. *Deterministic Policy Gradient Algorithms*. ICML, 2014.
3. Schulman, J. et al. *Proximal Policy Optimization Algorithms*. arXiv preprint, 2017.
4. Lillicrap, T. et al. *Continuous Control with Deep Reinforcement Learning*. arXiv, 2015.
5. Russell, S., & Norvig, P. *Artificial Intelligence: A Modern Approach*. Pearson, 2021.