

№12 Дәріс. Континуумдық күйлер мен әрекеттер

Дәріскер: Даркенбаев Д.К. PhD, қауымдастырылған профессор

Дәріс мақсаты

Бұл дәрістің мақсаты —
континуумдық күйлер мен
әрекеттердің маңызын, оларды
сипаттау тәсілдерін, сондай-
ақ үздіксіз ортада агентті
үйретудің ерекшеліктерін
түсіндіру.

Кілт сөздер

- Нығайтпалы оқыту
- Континуумдық күйлер
- Үздіксіз әрекеттер
- Саясат
- Мән функциясы
- Нейрондық желілер
- Адаптивті басқару
- Оптимизация

Дәріс жоспары

01

Континуумдық күйлер мен әрекеттердің жалпы түсінігі

02

Үздіксіз орта мысалдары

03

Үздіксіз айнымалылармен жұмыс істеудің қиындықтары

04

Континуумдық кеңістіктердегі саясатты үйрету

05

Континуумдық ортада қолданылатын әдістер

06

Қолданылу салалары

07

Қорытынды

08

Бақылау сұрақтары

09

Пайдаланылған әдебиеттер

Кіріспе

Нығайтпалы оқыту (Reinforcement Learning) саласында агент пен ортаның өзара әрекеттесуі көбінесе дискретті (сандық) күйлер мен әрекеттер арқылы сипатталады. Мұндай жағдайда агент әрбір күй үшін бірнеше нақты әрекет нұсқаларының бірін таңдайды.

Алайда нақты өмірдегі көптеген процестер — үздіксіз (континуумдық) сипатқа ие. Мысалы, роботтың қозғалыс бұрыштары, көліктің жылдамдығы, температураны реттеу деңгейі немесе энергияның қолдану мөлшері нақты сандар түрінде беріледі. Бұл параметрлерді бірнеше дискретті нүктеге бөлу мүмкін емес, себебі олардың арасы өте тығыз және өзгерісі үздіксіз.



Осы себепті континуумдық күйлер мен әрекеттер ұғымы нығайтпалы оқытуда маңызды орын алады. Бұл бағыт агенттің нақты физикалық немесе экономикалық ортада үздіксіз мәндермен жұмыс істеуін қамтамасыз етеді.

1. Континуумдық күйлер мен әрекеттердің жалпы түсінігі

Континуумдық күйлер мен әрекеттер — агенттің ортадағы жағдайын және мүмкін болатын әрекеттерін үздіксіз (реалды сандармен берілетін) айнымалылар арқылы сипаттау тәсілі.

Мысалы, роботтың орналасу координаталары (x, y) , бұрылу бұрышы, қозғалыс жылдамдығы немесе температура деңгейі сияқты шамалар нақты мәндер қабылдайды. Мұндай жағдайларда агент тек "солға", "оңға" немесе "тоқта" деген сияқты дискретті әрекеттердің бірін емес, нақты мәнді таңдауы керек (мысалы, 0.52 бұрышқа бұрылу, немесе 3.7 жылдамдықпен жүру).

1. Үздіксіз орта мысалдары

Континуумдық орта өмірдің көптеген салаларында кездеседі:



Робот техникасы

қозғалыс бұрыштары мен қозғалыс күшін үздіксіз басқару.



Энергетика

энергия өндіру немесе тарату көлемін нақты реттеу.



Қаржы саласы

инвестиция мөлшерін үздіксіз шама ретінде анықтау.



Көлік жүйелері

жылдамдық пен бұрылу бұрышын үздіксіз өзгерту.



Климаттық бақылау

температура мен ылғалдылық деңгейін біртіндеп реттеу.

Бұл мысалдардың барлығында шешім тек бірнеше баламаны таңдаудан тұрмайды, ол нақты мәндерді дәл есептеуді талап етеді.

1. Үздіксіз айнымалылармен жұмыс істеудің қиындықтары

Континуумдық күйлер мен әрекеттермен жұмыс істеу кезінде бірқатар қиындықтар туындайды:

Мән функциясын есептеу қиын

барлық мүмкін мәндер шексіз көп болғандықтан, оларды толық есептеу мүмкін емес.

Іздеу кеңістігі өте үлкен

агент әрбір нақты мән үшін шешім қабылдауы керек.

Деректердің көптігі

үздіксіз ортада үйрену үшін үлкен көлемдегі деректер қажет.

Сандық тұрақтылық

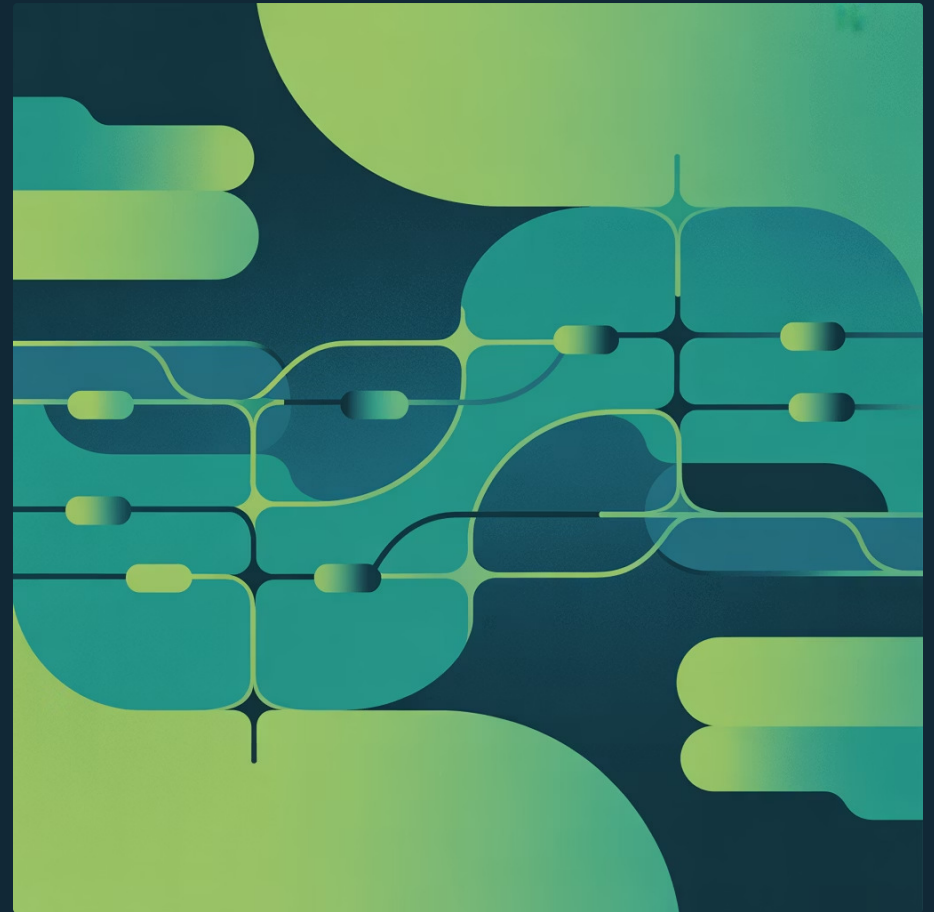
шағын өзгерістер үлкен нәтижеге әсер етуі мүмкін.

Сондықтан үздіксіз ортада классикалық кестелік әдістер тиімсіз болады. Оның орнына шамаларды жуықтауға арналған модельдер қолданылады.

1. Континуумдық кеңістіктердегі саясатты үйрету

Континуумдық күйлер мен әрекеттер жағдайында саясат (policy) дискретті емес, үздіксіз функция түрінде беріледі. Агенттің міндеті — әрбір күйге сәйкес келетін ең жақсы әрекеттің нақты мәнін табу.

Мұндай саясатты үйрету үшін көбінесе **нейрондық желілер** қолданылады. Олар күрделі және көпөлшемді тәуелділіктерді модельдеуге мүмкіндік береді.



Нейрондық желі кіріс ретінде ағымдағы күйдің параметрлерін алады, ал шығыс ретінде үздіксіз әрекет мәнін қайтарады. Бұл тәсіл "функцияларды жуықтау" әдістерімен тығыз байланысты.

1. Континуумдық ортада қолданылатын әдістер

Континуумдық күйлер мен әрекеттерді өңдеу үшін келесі әдістер кеңінен пайдаланылады:



Deep Deterministic Policy Gradient (DDPG)

үздіксіз әрекеттер кеңістігі үшін бейімделген саясатты
градиенттік әдіс



Twin Delayed DDPG (TD3)

үйренудің тұрақтылығын арттыратын тәсіл



Soft Actor-Critic (SAC)

энтропияны пайдалану арқылы әртүрлі стратегияларды
зерттеуді теңгеретін әдіс



Proximal Policy Optimization (PPO)

саясатты біртіндеп және сенімді жаңартуға арналған әдіс

Бұл тәсілдердің барлығы нейрондық желілер арқылы үздіксіз функцияларды аппроксимациялауға негізделген.

1. Қолданылу салалары

Континуумдық күйлер мен әрекеттер көптеген қолданбалы бағыттарда кездеседі:



Өнеркәсіптік басқару

температура, қысым немесе өндіріс жылдамдығын үздіксіз реттеу



Автономды көлік

жылдамдық пен бағытты үздіксіз өзгерту арқылы қозғалысты оңтайландыру



Қаржы модельдері

инвестициялық портфельдерді үздіксіз үлестіру



Энергетика

ресурстарды тиімді тарату және жүктемені реттеу



Роботтар

икемді қозғалысты және нақты маневрлерді орындау

1. Қорытынды

Континуумдық күйлер мен әрекеттер нығайтпалы оқытудың күрделі, бірақ маңызды саласы болып табылады. Мұндай жүйелер шынайы өмірге барынша жақын, себебі нақты физикалық процестер үздіксіз сипатқа ие.

Бұл бағыттағы зерттеулер қазіргі таңда жасанды интеллекттің ең жылдам дамып келе жатқан салаларының бірі.

- Континуумдық ортада тиімді саясатты үйрету — болашақта толық автономды жүйелерді (роботтар, көліктер, өндірістік желілер) құрудың негізі.



Бақылау сұрақтары

- 1 Континуумдық күйлер мен әрекеттер дегеніміз не?
- 2 Дискретті және үздіксіз орта арасындағы айырмашылық неде?
- 3 Неліктен үздіксіз кеңістікте үйрену күрделі?
- 4 Қандай қолданбалы салаларда континуумдық әдістер қажет?
- 5 Саясатты үйрету процесінде нейрондық желілер қандай рөл атқарады?
- 6 Үздіксіз ортада қолданылатын заманауи әдістерді атаңыз.
- 7 Континуумдық күйлер мен әрекеттерді зерттеу не үшін маңызды?

Пайдаланылған әдебиеттер

1. Sutton, R. S., & Barto, A. G. Reinforcement Learning: An Introduction. MIT Press, 2018.
2. Lillicrap, T. et al. Continuous Control with Deep Reinforcement Learning. arXiv, 2015.
3. Fujimoto, S., Hoof, H., & Meger, D. Addressing Function Approximation Error in Actor-Critic Methods. ICML, 2018.
4. Haarnoja, T. et al. Soft Actor-Critic Algorithms and Applications. arXiv, 2019.
5. Silver, D. et al. Deterministic Policy Gradient Algorithms. ICML, 2014.