Deep Reinforcement Learning - The Basics - Quiz

* Indicates required question

Email *

maksym.batiuk.kn.2021@lpnu.ua

Ваше ім'я та прізвище: *

Максим Батюк

Опишіть 3 MDP процеси: мета процесу, стани (S), дії (A), система винагород (R) *

1. Розумний кондиціонер:
Мета: підтримувати оптимальну температуру в квартирі.
Стани: Температура потоку повітря, яка виходить з кондиціонеру, час доби, сезон року, вологість повітря
Дії: Увімкнути/вимкнути кондиціонер, змінити температуру (збільшити, зменшити або залишити такою ж).
Система винагород: -1 за кожен раз, коли користувач вручну міняє температуру (бо, скоріше за все, теперішня температура йому дискомфортна)

2. Гра Flappy Bird:
Мета процесу: пролетіти якнайдалі, тобто оминути якнайбільше колон
Стани: положення пташки та отворів в колонах, крізь які треба пролетіти
Дії: підлетіти трохи вгору (коли на телефоні ми натискаємо на екран) або продовжувати падати (тобто нічого не робити)
Система винагород: +1 за кожну пройдену колону, -10 за доторкання до верхньої/ нижньої межі або колони

3. Автоматичний годувальник домашніх тварин (наприклад, кіт)
Мета: Дати достатньо корму, щоб кіт наївся, але в той же час щоб корму не було забагато і він не почав сохнути і ставати несмачним.
Стани: Час доби, наявність їжі в мисці, присутність кота біля годувальниці (якщо він біля годувальниці – напевно він голодний)
Дія: Насипати 10 грамів корму (якщо треба більше, то на наступному кроці буде насипано ще 10 грамів) чи не робити нічого
Система винагород: -1, якщо кіт не біля годувальниці, а миска наповнена; -1, якщо кіт біля годувальніці і миска пуста; +1, якщо миска була заповнена кормом, кіт все з'їв і пішов від годувальниці

Опишіть 1 процес, який НЕ вкладається у модель MDP. В чому його ключова відмінність від попередніх трьох? *

Задача розпізнавання облич на зображеннях є процесом, для якого важко знайти стани, дії та систему винагород, оскільки:
- Хоч людина може бути однією і тією ж, зображення будуть трохи відрізнятись, тому в нас не буде випадку, коли стани будуть повторюватись
- Не можна надати винагороду, бо ми не можемо знати, чи система розпізнавання облич правильно розпізнала людину

Навчається і приймає рішення: *

- ○ Середовище
- ◉ Агент
- ○ Стан
- ○ MDP

Уявіть навчання у епізодичному MDP. Кількість кроків агенту у кожному епізоді: *

- ○ Постійна (однакова)
- ◉ Стохастична, тобто може відрізнятись у кожному з епізодів

У чому відмінність малого γ (discount factor) від великого? *

- ○ Абсолютна величина γ не впливає на агента
- ◉ Більший γ призводить до більш далекоглядного планування
- ○ Більший γ призводить до більш короткотермінового планування

Policy - це функція, де аргументом є ____, а результатом ____

- ○ Дії та розподіл ймовірностей станів
- ◉ Стани та дії
- ○ Стани та values
- ○ Стани та розподіл ймовірностей дій
- ○ Дії та ймовірності

Clear selection

Що можна знайти за допомогою Bellman expectation equation?

- ○ action-value function та optimal policy
- ○ optimal policy та state-value function
- ◉ action-value function та state-value function

Clear selection

Які з наведених нижче способів розв’язати Bellman optimality equation є вірними (кілька правильних відповідей)? *

- [x] Policy iteration
- [ ] State iteration
- [x] Brute-force search
- [ ] Action iteration
- [x] Value iteration
- [ ] Agent search
- [ ] Вибір алгоритму залежить від конкретної задачі (Policy чи Value iteration)

Вкажіть найшвидший спосіб розв’язку Bellman optimality equation (один)?

- ○ Policy iteration
- ○ State iteration
- ○ Brute-force search
- ○ Action iteration
- ○ Value iteration
- ○ Agent search
- ◉ Вибір алгоритму залежить від конкретної задачі (Policy чи Value iteration)

Clear selection

Submit Clear form