**Nazwa przedmiotu:**

Uczące się systemy decyzyjne

**Koordynator przedmiotu:**

Rajmund Kożuszek

**Status przedmiotu:**

Fakultatywny dowolnego wyboru

**Poziom kształcenia:**

Studia II stopnia

**Program:**

Informatyka

**Grupa przedmiotów:**

Przedmioty techniczne - zaawansowane

**Kod przedmiotu:**

USD

**Semestr nominalny:**

2 / rok ak. 2021/2022

**Liczba punktów ECTS:**

4

**Liczba godzin pracy studenta związanych z osiągnięciem efektów uczenia się:**

1. liczba godzin kontaktowych – 49 godz.,w tym:
a. obecność na wykładach: 30 godz.,
b. udział w spotkaniach projektowych: 15 godz.,
c. udział w konsultacjach związanych z realizacją przedmiotu: 2 godz.,
d. udział w egzaminie: 2 godz.
2. praca własna studenta – 70 godz., w tym:
a. analiza literatury i materiałów wykładowych związana z przygotowaniem do kolejnych wykładów: 20 godz.
b. realizacja zadań projektowych: 40 godz.
c. przygotowanie do egzaminu: 10 godz.
Łączny nakład pracy studenta wynosi: 119 godz., co odpowiada 4 pkt. ECTS.

**Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich:**

1,65 pkt. ECTS, co odpowiada 49 godz. kontaktowym

**Język prowadzenia zajęć:**

polski

**Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym:**

1,85 pkt. ECTS, co odpowiada 40 godz. realizacji projektu i 15 godz. spotkań projektowych

**Formy zajęć i ich wymiar w semestrze:**

|  |  |
| --- | --- |
| Wykład:  | 30h |
| Ćwiczenia:  | 0h |
| Laboratorium:  | 0h |
| Projekt:  | 15h |
| Lekcje komputerowe:  | 0h |

**Wymagania wstępne:**

l Biegła umiejętność programowania w języku Python.
l Znajomość inżynierskiego kursu dziedziny sztuczna inteligencja.
l Znajomość inżynierskiego kursy dziedziny sztuczne sieci neuronowe.

**Limit liczby studentów:**

30

**Cel przedmiotu:**

Celem przedmiotu jest przedstawienie bieżącego stanu wiedzy w dynamicznie rozwijającej się gałęzi sztucznej inteligencji pn. uczenie się ze wzmocnieniem. Omówione są także kluczowe wyniki programowania dynamicznego i sterowania adaptacyjnego.

**Treści kształcenia:**

Wykład:
Część I. Programowanie dynamiczne
1. Sekwencyjny proces decyzyjny w warunkach niepewności i jego model (2 godz.)
Proces decyzyjny Markowa (PDM). Polityka decyzyjna. Kryterium jakości: funkcja wartości. Zastosowania i opis.
2. Programowanie dynamiczne (2 godz.)
Funkcja wartości i funkcja wartości-decyzji, optymalizacja polityki decyzyjnej na horyzoncie skończonym, klasyczny algorytm programowania dynamicznego.
3. Algorytmy programowania dynamicznego (2 godz.)
Algorytm iteracji polityki, algorytm iteracji wartości, zasada optymalności Bellmanna.
Część II. Uczenie się ze wzmocnieniem
4. Preliminaria analityczne (2 godz.) Algorytm stochastycznego najszybszego spadku i jego zastosowania.
5. Q-Learning i jego pochodne (2 godz.)
Algorytmy Q-Learning, Q-Learning z aproksymacja funkcji, Deep Q-Learning, Deep Deterministic Policy Gradient; strategie wyboru decyzji: epsilon-zachłanna i Bolzmannowska.
6. Optymalizacja stochastycznego wyboru (2 godz.) Parametryzowane rozkłady prawdopodobieństwa; algorytmy REINFORCE punktowy, REINFORCE statyczny i REINFORCE epizodyczny.
7. Architektura Aktor-Krytyk (2 godz.)
Algorytm klasyczny Aktor-Krytyk, Aktor-Krytyk(lambda).
8. Powtarzanie doświadczenia (2 godz.)
Próbkowanie ważnościowe; algorytm Aktor-Krytyk z powtarzaniem doświadczenia.
9. Ograniczenie wariancji w uczeniu typu Aktor-Krytyk (2 godz.)
Problem wariancji w uczeniu z powtarzaniem doświadczenia. Ograniczenia w zmianach Aktora. Algorytmy z rodziny PPO.
10. Asynchroniczne i dystrybucyjne uczenie się ze wzmocnieniem (2 godz.)
Ogólny algorytm asynchroniczny. Algorytm A3C.
Regresja kwantylowa. Optymalizacja rozkładu wypłat: dystrybucyjne uczenie się ze wzmocnieniem.
11. Częściowo obserwowalny stan (4 godz.)
Uczenie się ze wzmocnieniem w warunkach, gdy stan jest częściowo obserwowalny. Zastosowanie rekurencyjnych sieci neuronowych w uczeniu się ze wzmocnieniem.
Część III. Sterowanie adaptacyjne
12. Wstęp do teorii sterowania (2 godz.)
Preliminaria analityczne: równania różniczkowe, transformata Laplace'a, funkcja Lapunowa .
13. Sterowanie adaptacyjne z modelem referencyjnym (2 godz.)
Liniowy model środowiska z jedną pochodną stanu, liniowy z wieloma pochodnymi stanu, nieliniowy.
14. Estymacja stanu środowiska (2 godz.)
Filtr Kalmana i Rozszerzony Filtr Kalmana.

Projekt:
Studenci wykonują zadania projektowe w domach. Opracowują zadane problemy, w tym celu implementują i stosują wybrane metody omawiane na wykładzie i we wskazanej literaturze.
1. Sprowadzanie wskazanych problemów technicznych i ekonomicznych do problemów sekwencyjnego podejmowania decyzji w warunkach niepewności.
2. Implementacja i zastosowanie klasycznego algorytmu programowania dynamicznego.
3. Implementacja i zastosowanie algorytmów iteracji wartości oraz iteracji polityki.
4. Implementacja i zastosowanie algorytmów Q-Learning, Deep Q-Learning i Deep Deterministic Policy Gradient.
5. Implementacja i zastosowanie różnych wersji algorytmu Reinforce.
6. Implementacja i zastosowanie algorytmów Aktor-Krytyk i Aktor-Krytyk(lambda).
7. Implementacja i zastosowania algorytmu Aktor-Krytyk z powtarzaniem doświadczenia.
8. Implementacja i zastosowanie algorytmu A3C.
9. Implementacja i zastosowanie algorytmów PPO-Penalty i PPO-Clip.
10. Implementacja i zastosowanie QR-DQN
12. Implementacja i zastosowanie algorytmu Q-Learning z rekurencyjną siecią neuronową.
13. Implementacja i zastosowanie algorytmu sterowania adaptacyjnego z modelem referencyjnym.
14. Implementacja i zastosowanie Rozszerzonego Filtru Kalmana.

**Metody oceny:**

Realizacja przedmiotu obejmuje następujące formy zajęć:
l wykład prowadzony w wymiarze 2 godz. tygodniowo,
l projekt realizowany samodzielnie w zespołach,
l konsultacje.
Aktywizacji studentów służą:
l interaktywna formuła wykładu,
l dostarczane po każdym wykładzie ćwiczenia sprawdzające przyswojenie omawianych zagadnień,
l wymóg konsultacji interpretacji tematu i zakresu projektu,
l wymóg przedstawienia do oceny wstępnej dokumentacji projektu,
l wymóg konsultacji zmian interpretacji tematu i zakresu projektu wprowadzanych po ocenie dokumentacji wstępnej.
Sprawdzanie założonych efektów kształcenia realizowane jest przez:
l ocenę wiedzy i umiejętności związanych z realizacją projektu – ocena ze zrealizowanych zadań;
l ocenę wiedzy wykazanej na egzaminie pisemnym.

**Egzamin:**

tak

**Literatura:**

1. Wybrane publikacje z konferencji ICML, NIPS i arXiv.org.
2. P. Wawrzyński, Sterowanie adaptacyjne i uczenie maszynowe – preskrypt, Politechnika Warszawska, 2012.

**Witryna www przedmiotu:**

https://usosweb.usos.pw.edu.pl/kontroler.php?\_action=katalog2/przedmioty/pokazPrzedmiot&prz\_kod=103A-INSZI-MSP-USD

**Uwagi:**

(-)

## Charakterystyki przedmiotowe