

Nazwa pola Field name	Komentarz <sup>1</sup> Comment
Tytuł Course title:	Uczenie ze wzmocnieniem Reinforcement learning
Jednostka Department:	Wydział Matematyki, Informatyki i Mechaniki UW Faculty of Mathematics, Informatics, and Mechanics
Grupy Course groups:	przedmioty obowiązkowe dla pierwszego roku studiów magisterskich uczenie maszynowe Obligatory courses for 1st year 2nd cycle Machine Learning
Kod wg USOS USOS Course ID:	1000-318bRL
Kod wg Sokratesa Erasmus code / ISCED:	11.304
Rodzaj zajęć Type of class:	Wykład 30 godzin, Ćwiczenia 30 godzin Lecture, 30 hours, Class, 30 hours
Opis Short description:	<p>Celem przedmiotu jest przedstawienie współczesnych technik oraz algorytmów uczenia ze wzmocnieniem ze szczególnym uwzględnieniem metod bezmodelowych, metod z użyciem modelu oraz metod opartych o przeszukiwanie.</p> <p>The goal of the course is to present modern techniques and algorithms of reinforcement machine learning, with particular emphasis on model-free, model-based and exploration methods.</p>
Program Programme content/Full description	<p>1. Metody bezmodelowe</p> <ul style="list-style-type: none"> <li>a) Formalizm uczenia ze wzmocnieniem: procesy decyzyjne Markowa (MDP) &amp; programowanie dynamiczne (DP)</li> <li>b) Metody oparte na wartościach <ul style="list-style-type: none"> <li>* SARSA i TD(1)</li> <li>* kompromis pomiędzy stronniczością i wariancją oraz TD(lambda)</li> <li>* aproksymatory funkcji i związane z tym wyzwania</li> </ul> </li> <li>c) Metody gradientu polityki <ul style="list-style-type: none"> <li>* Proste gradienty polityki</li> <li>* Uogólniony estymator przewagi (GAE)</li> <li>* Problemy związane z metodami gradientu polityki</li> </ul> </li> <li>d) Metody typu aktor-krytyk <ul style="list-style-type: none"> <li>* Optymalizacja polityki z użyciem regionów zaufania (TRPO)</li> <li>* Optymalizacja bliskiej polityki (PPO)</li> <li>* Algorytm stonowany aktor-krytyk (SAC)</li> </ul> </li> </ul> <p>2. Metody z użyciem modelu</p> <ul style="list-style-type: none"> <li>a) Szacowanie modelu</li> <li>b) Planowanie <ul style="list-style-type: none"> <li>* Ciągłe i dyskretne problemy sterowania</li> <li>* Przeszukiwanie drzew Monte-Carlo</li> </ul> </li> </ul>

<sup>1</sup> Przedstawiona jest wstępna wersja sylabusu./ This is a preliminary version of the course outline.

	<ul style="list-style-type: none"> <li>* Alfa-zero</li> <li>3. Przeszukiwania <ul style="list-style-type: none"> <li>a) Model wielorękiego bandyty</li> <li>b) Strategie przeszukiwania związane z niepewnością</li> </ul> </li> <li>4. Tematy badawcze</li> <li>5. Wystąpienia praktyków.</li> </ul> <ul style="list-style-type: none"> <li>1. Model-free methods <ul style="list-style-type: none"> <li>a) Reinforcement Learning formalism: Markov Decision Processes (MDPs) &amp; Dynamic programming (DP)</li> <li>b) Value methods <ul style="list-style-type: none"> <li>* SARSA and TD(1)</li> <li>* Bias-variance trade-off and TD(lambda)</li> <li>* Function approximators and corresponding challenges</li> </ul> </li> <li>c) Policy gradient methods <ul style="list-style-type: none"> <li>* Vanilla policy gradients</li> <li>* Generalized Advantage Estimator (GAE)</li> <li>* Problems with policy gradient methods</li> </ul> </li> <li>d) Actor-critic methods <ul style="list-style-type: none"> <li>* Trust Region Policy Optimization (TRPO)</li> <li>* Proximal Policy Optimization (PPO)</li> <li>* Soft Actor-Critic (SAC)</li> </ul> </li> </ul> </li> <li>2. Model-based methods: <ul style="list-style-type: none"> <li>a) Model estimation</li> <li>b) Planning <ul style="list-style-type: none"> <li>* Continuous and discrete control problems</li> <li>* Monte-Carlo Tree Search</li> <li>* AlphaZero</li> </ul> </li> </ul> </li> <li>3. Exploration <ul style="list-style-type: none"> <li>a) Multi-armed bandits model</li> <li>b) Uncertainty related exploration strategies</li> </ul> </li> <li>4. Research topics</li> <li>5. Talks by practitioners</li> </ul>
Efekty uczenia Learning outcomes	<p>Wiedza: student zna i rozumie</p> <ul style="list-style-type: none"> <li>● własności algorytmów uczenia ze wzmocnieniem, wie kiedy które zastosować i jak zaimplementować najważniejsze z nich ze szczególnym uwzględnieniem klasy algorytmów opartych na gradiencie polityki, z klasy opartych na wartości oraz z klasy aktor-krytyk [K_W14].</li> </ul> <p>Umiejętności: student potrafi</p> <ul style="list-style-type: none"> <li>● odpowiednio stosować metody w celu opracowania dedykowanego algorytmu uczenia ze wzmocnieniem lub stosować istniejące metody w swoich projektach badawczych. [K_U17]</li> </ul>

	<ul style="list-style-type: none"> <li>● implementować własne algorytmy i używać istniejące biblioteki oferujące procedury uczenia ze wzmocnieniem. [K_U18]</li> </ul> <p>Kompetencje społeczne: student jest gotów do</p> <ul style="list-style-type: none"> <li>● krytycznej oceny posiadanej wiedzy i odbieranych treści [K_K01];</li> <li>● uznawania znaczenia wiedzy w rozwiązywaniu problemów poznawczych i praktycznych oraz zasięgania opinii ekspertów w przypadku trudności z samodzielnym rozwiązaniem problemu [K_K02];</li> <li>● myślenia i działania w sposób przedsiębiorczy [K_K03].</li> </ul> <p>Knowledge: the student</p> <ul style="list-style-type: none"> <li>● knows the properties of reinforcement learning algorithms, knows scenarios of their application and how to implement the most important ones, especially ones from the class of policy gradient algorithms, the class of value based algorithms and the class of actor-critic algorithms [K_W14].</li> </ul> <p>Abilities: the graduate is able to</p> <ul style="list-style-type: none"> <li>● appropriately apply methods to design a dedicated reinforcement learning algorithm or apply existing methods in own research projects. [K_U17]</li> <li>● implement own algorithms and use existing libraries with reinforcement learning procedures [K_U18]</li> </ul> <p>Social competences: the graduate is ready to</p> <ul style="list-style-type: none"> <li>● critically evaluate acquired knowledge and information [K_K01];</li> <li>● recognize the significance of knowledge in solving cognitive and practical problems and the importance of consulting experts when difficulties arise in finding a self-devised solution [K_K02];</li> <li>● think and act in an entrepreneurial way [K_K03].</li> </ul>
Proponowane źródła Bibliography:	<p>R. Sutton, G. Barto, Reinforcement Learning: An Introduction</p> <p>Francois-Lavet, F., Henderson P., Islam R., Bellemare M. G., Pineau J., An Introduction to Deep Reinforcement Learning.</p> <p>Szepesvari, C., Algorithms for Reinforcement Learning</p>
Kryteria oceniania Assessment criteria:	<p>Egzamin i projekt.</p> <p>Exam, project.</p>
Wymagania Requirements:	<p>Uczenie statystyczne (1000-317bSML), Głębokie sieci neuronowe (1000-317bDNN)</p> <p>Statistical learning (1000-317bSML), Deep neural networks (1000-317bDNN)</p>
Założenia Assumptions	–
Forma zaliczenia Assessment of the learning outcome	<p>Egzamin i zaliczenie projektu na ocenę</p> <p>Exam and graded project.</p>
<i>Teraz informacje o charakterze zmiennym (pole edycji) Variable information (editable fields)</i>	
Cykl dydaktyczny Time span/Study term:	<p>Semestr letni 2021/2022</p> <p>Summer semester 2021/2022</p>
Prowadzący Coordinators:	<p>Wykład: ... Ćwiczenia: ...</p> <p>Lecture: ... Class: ...</p>