



Karta przedmiotu

| | | | | | | | |
|--|---|---|--------------------------------------|------------------------|--|-----------------------|-------|
| Nazwa i kod przedmiotu | Optymalizacja struktur i obliczeń w sieciach neuronowych, PG_00054195 | | | | | | |
| Kierunek studiów | Informatyka | | | | | | |
| Data rozpoczęcia studiów | luty 2021 r. | | Rok akademicki realizacji przedmiotu | | 2021/2022 | | |
| Poziom kształcenia | II stopnia | | Grupa zajęć | | Grupa zajęć fakultatywnych Grupa zajęć powiązanych z prowadzonymi badaniami naukowymi w dziedzinie nauki związanej z kierunkiem - profil ogólnoakademicki | | |
| Forma studiów | stacjonarne | | Sposób realizacji | | na uczelni | | |
| Rok studiów | 2 | | Język wykładowy | | polski | | |
| Semestr studiów | 3 | | Liczba punktów ECTS | | 3.0 | | |
| Profil kształcenia | ogólnoakademicki | | Forma zaliczenia | | zaliczenie | | |
| Jednostka prowadząca | Wydział Elektroniki -> Telekomunikacji i Informatyki -> Katedra Systemów Multimedialnych | | | | | | |
| Imię i nazwisko wykładowcy (wykładowców) | Odpowiedzialny za przedmiot | | dr hab. inż. Piotr Szczuko | | | | |
| | Prowadzący zajęcia z przedmiotu | | dr hab. inż. Piotr Szczuko | | | | |
| Formy zajęć i metody nauczania | Forma zajęć | Wykład | Ćwiczenia | Laboratorium | Projekt | Seminarium | RAZEM |
| | Liczba godzin zajęć | 15.0 | 0.0 | 15.0 | 15.0 | 0.0 | 45 |
| | W tym liczba godzin zajęć na odległość: 0.0 | | | | | | |
| | Adres zajęć na odległość: https://enauczanie.pg.edu.pl/moodle/course/view.php?id=22383 Optymalizacja struktur i obliczeń w sieciach neuronowych - Moodle ID: 22383 https://enauczanie.pg.edu.pl/moodle/course/view.php?id=22383 | | | | | | |
| Aktywność studenta i liczba godzin pracy | Aktywność studenta | Udział w zajęciach dydaktycznych, objętych planem studiów | | Udział w konsultacjach | | Praca własna studenta | RAZEM |
| | Liczba godzin pracy studenta | 45 | | 2.0 | | 28.0 | 75 |
| Cel przedmiotu | Celem przedmiotu jest zapoznanie studenta w sposób teoretyczny, praktyczny (laboratorium) i problemowy (projekt) z zagadnieniami optymalizacji modeli w uczeniu głębokim. Przedstawione zostają techniki optymalizacji, upraszczania, struktury modeli, optymalizacji i przyspieszania obliczeń. Omawiane są algorytmy i podejścia do efektywnego uczenia modeli, zapewniania wysokiej skuteczności i dokładności działania w praktycznych przypadkach, m.in. ograniczonej pamięci, zaszumienia danych. | | | | | | |

| Efekty uczenia się przedmiotu | Efekt kierunkowy | Efekt z przedmiotu | Sposób weryfikacji i oceny efektu |
|-------------------------------|--|---|---|
| | [K7_U07] potrafi wykorzystać zaawansowane metody wspomagania procesów i funkcji, specyficzne dla kierunków studiów | Student zrealizował praktycznie model uczenia maszynowego i zoptymalizował go w wybrany sposób - uzasadniony realizowanym zadaniem, dopasowany do specyfiki modelu, stosując funkcje wybranej biblioteki programistycznej. | [SU4] Ocena umiejętności korzystania z metod i narzędzi [SU1] Ocena realizacji zadania |
| | [K7_W06] zna i rozumie w pogłębionym stopniu podstawowe procesy zachodzące w cyklu życia urządzeń, obiektów i systemów technicznych | Student potrafi określić cele wykonywanego projektu, a po wykonaniu rozwiązania poprawnie sformułować wnioski. Student potrafi uzasadnić dobór stosowanych metod teoretycznych i narzędzi do realizacji założonego zadania. Student zna i przytacza podstawy teoretyczne odpowiednio do zagadnienia. | [SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym |
| | [K7_W43] zna i rozumie w pogłębionym stopniu formalne, techniczne i społeczne aspekty działania złożonych systemów informatycznych w społeczeństwie informacyjnym i w globalnej infrastrukturze informacyjnej | Student potrafi zastosować narzędzia i uzasadnić potrzebę optymalizacji obliczeń i struktur sieci neuronowych, zna wpływ danych metod na skuteczność i złożoność modeli. | [SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym |
| | [K7_U42] potrafi rozwiązywać problemy inżynierskie i badawcze w zakresie projektowania, oceny i utrzymania systemów i aplikacji informacyjnych z wykorzystaniem metod eksperymentalnych i technik zarządzania | Student realizuje projekt, wybiera narzędzia i uzasadnia ich użycie, prowadzi eksperymenty i mierzy skuteczność modelu. Poprawnie wyciąga wnioski z uzyskanych pomiarów. | [SU3] Ocena umiejętności wykorzystania wiedzy uzyskanej w ramach przedmiotu [SU1] Ocena realizacji zadania |
| | [K7_W41] zna i rozumie w pogłębionym stopniu standardy, metody wytwarzania, cykl życia i trendy rozwojowe oprogramowania oraz systemów i aplikacji informacyjnych | Student zna standardowe metody optymalizacji obliczeń i struktur modeli, potrafi uzasadniać ich użycie, wyciągać wnioski i przewidywać rezultaty zastosowania wybranych metod. Rozumie różnice pomiędzy systemami i rozwiązaniami i przypadkami użycia - systemy i obliczenia scentralizowane, rozproszone, obliczenia w urządzeniach brzegowych i serwerach. | [SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym |
| Treści przedmiotu | <p>Metody redukcji rozmiaru sieci i przyspieszania obliczeń.</p> <p>Kwantyzacja, upraszczanie modeli i destylacja wiedzy.</p> <p>Przerzedzanie danych i architektury modeli.</p> <p>Uczenie modeli na zaszumionych danych</p> <p>Poszukiwanie architektury modeli NAS.</p> <p>Trening samo-nadzorowany, pre-trenowanie modeli.</p> <p>Oszacowanie niepewności w sieciach neuronowych (metody kalibracji sieci neuronowych, techniki dropout, grupowanie modeli, sieci Bayesowskie)</p> <p>Odporność modeli, metody przeciwstawne (adversarial) Modele hybrydowe, pozbawione wag, kapsułkowe.</p> | | |
| Wymagania wstępne i dodatkowe | | | |

| | | | |
|---|--|---|-------------------------|
| Sposoby i kryteria oceniania osiąganych efektów uczenia się | Sposób oceniania (składowe) | Próg zaliczeniowy | Składowa oceny końcowej |
| | Kolokwium | 51.0% | 35.0% |
| | Laboratorium | 51.0% | 35.0% |
| | Projekt | 51.0% | 30.0% |
| Zalecana lista lektur | Podstawowa lista lektur | <p>Torsten Hoefer, Dan Alistarh, Tal Ben-Nun, Nikoli Dryden, Alexandra Peste, (2021) Sparsity in Deep Learning: Pruning and growth for efficient inference and training in neural networks. [2102.00554] (arxiv.org)</p> <p>Yu Cheng, Duo Wang, Pan Zhou, and Tao Zhang. 2020. A Survey of Model Compression and Acceleration for Deep Neural Networks. (2020). arXiv:cs.LG/1710.09282</p> <p>Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. 2019. Neural Architecture Search: A Survey. (2019). arXiv:stat.ML/1808.05377</p> <p>Manish Gupta and Puneet Agrawal. 2020. Compression of Deep Learning Models for Text: A Survey. (2020). arXiv:cs.CL/2008.05221</p> <p>V. Sze, Y. Chen, T. Yang, and J. S. Emer. 2017. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. Proc. IEEE 105, 12 (2017), 22952329. https://doi.org/10.1109/JPROC.2017.2761740</p> | |
| | Uzupełniająca lista lektur | <p>Tensorflow model optimization (2022) https://www.tensorflow.org/model_optimization</p> <p>Yi Tay, Mostafa Dehghani, Dara Bahri, and Donald Metzler. 2020. Efficient transformers: A survey. (2020). arXiv:cs.LG/2009.06732</p> | |
| | Adresy eZasobów | | |
| Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania | <ul style="list-style-type: none">• Opisz, skomentuj, uzasadnij przydatność jednej z wybranych metod poznanych w ramach przedmiotu: redukcji rozmiaru sieci i przyspieszania obliczeń, kwantyzacji, upraszczanie modeli i destylacji wiedzy.• Uzasadnij potrzebę przerzedzanie danych i architektury modeli i korzyści płynące z tych operacji.• Opisz w jaki sposób realizować można uczenie modeli na zaszumionych danych.• Omów jak działa trening samo-nadzorowany, pre-trenowanie modeli oraz jakie są korzyści tych etapów.• Opisz metody kalibracji sieci neuronowych, techniki dropout, grupowania modeli.• Omów jak szacować odporność modeli• Podaj przykład modelu pozbawionego wag oraz zastosowania sieci kapsułkowych. | | |
| Praktyki zawodowe w ramach przedmiotu | Nie dotyczy | | |