



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Metody teorii informacji w analizie danych [S2SI1E>TIAD]

Przedmiot

Kierunek studiów

Sztuczna inteligencja/Artificial Intelligence

Rok/Semestr

1/2

Studia w zakresie (specjalność)

–

Profil studiów

ogólnoakademicki

Poziom studiów

drugiego stopnia

Język oferowanego przedmiotu

angielski

Forma studiów

stacjonarne

Wymagalność

obieralny

Liczba godzin

Wykład

15

Laboratorium

15

Inne

0

Ćwiczenia

0

Projekty/seminaria

0

Liczba punktów ECTS

3,00

Koordynatorzy

dr hab. inż. Robert Susmaga

robert.susmaga@put.poznan.pl

Wykładowcy

Wymagania wstępne

Podstawowa wiedza dotycząca: a) Rachunek (funkcja logarytmiczna, funkcja wykładnicza), b) Algebra Liniowa (wektory, macierze, operacje na wektorach/macierzach), c) Teoria prawdopodobieństwa i teoria informacji (prawdopodobieństwo, entropia). Podstawowe umiejętności projektowania, tworzenia i testowania programów komputerowych (w wybranym przez siebie języku programowania) realizujących proste przetwarzanie statycznych (wektory i macierze) i dynamicznych (listy, drzewa) struktur danych.

Cel przedmiotu

Celem kursu jest przedstawienie wybranych aspektów teorii informacji, jednej z najbardziej fundamentalnych teorii leżących u podstaw współczesnej informatyki teoretycznej. Teoria Informacji zajmuje się przedstawianiem, przechowywaniem i przekazywaniem informacji wyrażonych w postaci symboli. Ze względu na to, że wiele ważnych zastosowań tej teorii wykracza daleko poza istotę informatyki, prezentowany wybór aspektów ograniczy się do tych, które są najbardziej przydatne w takich dziedzinach informatyki, jak analiza danych i eksploracja danych. Obejmują one głównie różne użyteczne miary (w szczególności: różne warianty miar entropii Shannona), więc kurs ten będzie skoncentrowany zarówno na teoretycznych, jak i praktycznych aspektach tych właśnie miar.

Przedmiotowe efekty uczenia się

Wiedza Student:

- ma zaawansowaną i pogłębioną wiedzę z zakresu systemów informatycznych związanych z metodami teorii informacji i analizy danych (K2st_W1)
- ma uporządkowaną i podbudowaną teoretycznie wiedzę ogólną związaną z kluczowymi zagadnieniami z teorii informacji, w szczególności dotyczącymi miar wielowymiarowych, wraz z ich zaletami (interpretacje geometryczne) i wadami (K2st_W2)
- ma zaawansowaną wiedzę szczegółową dotyczącą wybranych problemów występujących w ramach analizy/eksploracji danych, w szczególności dotyczących charakteryzowania zależności pomiędzy zmiennymi (K2st_W3)
- ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach informatyki, w szczególności na polu teorii informacji i analizy/eksploracji danych, w których najnowsze osiągnięcia wykorzystują skuteczne algorytmy dotyczące w przestrzeni wielowymiarowych (K2st_W4)
- zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich i prowadzeniu prac badawczych w obszarze analizy danych wielowymiarowych, przede wszystkim dotyczące zastosowań narzędzi i metod teorii informacji do rozwiązywania zagadnień analizy i eksploracji danych (np. możliwości zastosowania wielowymiarowych miar informacji do ilościowego określenia poziomów zależności między zmiennymi) (K2st_W6).

Umiejętności Student

- potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł (w języku polskim i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie (K2st_U1)
- potrafi planować i przeprowadzać eksperymenty, w tym pomiary i symulacje komputerowe, interpretować uzyskane wyniki i wyciągać wnioski oraz formułować i weryfikować hipotezy związane ze z problemami z dziedziny teorii informacji (K2st_U3)
- potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody analityczne, symulacyjne oraz eksperymentalne, w szczególności pochodzące z dziedziny teorii informacji (K2st_U4)
- potrafi — przy formułowaniu i rozwiązywaniu zadań inżynierskich, w szczególności dotyczących uczenia maszynowego i eksploracji danych — integrować wiedzę z różnych obszarów matematyki (teoria informacji, analiza i eksploracja danych, itp.), uwzględniając także aspekty pozatechniczne (K2st_U5)
- potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych, przede wszystkim z dziedziny teorii informacji i zastosować je do analizy i przetwarzania danych wielowymiarowych (np. pozwalającej na odkrywanie zależności pomiędzy obiektami) (K2st_U6)
- potrafi dokonać krytycznej analizy istniejących rozwiązań technicznych (w szczególności, np. dziedzinach: teorii informacji i eksploracji danych -- rozwiązań wymagających skutecznego charakteryzowania zależności w danych) oraz zaproponować ich ulepszenia (usprawnienia) (K2st_U8)
- potrafi ocenić przydatność metod i narzędzi służących do rozwiązania zadania inżynierskiego w dziedzinach teorii informacji i eksploracji danych (K2st_U9)
- potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania informatyczne z zakresu teorii informacji i eksploracji danych (K2st_U10)
- potrafi - zgodnie z zadaną specyfikacją, uwzględniającą aspekty pozatechniczne - zaprojektować złożony system informatyczny oraz zrealizować ten projekt -- co najmniej w części -- używając właściwych metod, technik i narzędzi, w tym przystosowując do tego celu istniejące lub opracowując nowe narzędzia teorii informacyjnej analizy danych (K2st_U11).

Kompetencje społeczne Student:

- rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe (K2st_K1),
- potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania (K2st_K2).

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca: (w zakresie laboratoriów):

- na podstawie oceny bieżącego postępu realizacji zadań.

Ocena podsumowująca (zarówno w zakresie wykładów jak i laboratoriów):

- ocena wiedzy i umiejętności wykazanych na pisemnym sprawdzianie wiedzy zawierającym w kilk (4-6) zadań (analogicznych do zadań prezentowanych na zajęciach); czas przewidziany na zaliczenie to 60-90 (wykłady) / 30-45 (laboratoria) minut; aby uzyskać ocenę pozytywną trzeba zdobyć przynajmniej $1 + \lfloor m/2 \rfloor$ (zaokrąglenie w dół) punktów, gdzie m jest punktacją maksymalną (np. aby uzyskać ocenę pozytywną przy

m = 30 należy zdobyć przynajmniej 16 punktów).

Treści programowe

Program wykładu obejmuje następujące zagadnienia:

Idea informacji Shannona (w skrócie: informacja). Miara informacji: budowa i właściwości. Pojęcie entropii Shannona (w skrócie: entropia) i jej podstawowe własności matematyczne, idea wymiarów w entropii wielowymiarowej, wykresy i ekstrema entropii wielowymiarowej, entropia jako miara zawartości informacji: budowa i właściwości.

Miary entropii w wielowymiarowych problemach eksploracji danych: idea, właściwości i interpretacje (ze szczególnym uwzględnieniem kontekstów wielowymiarowych).

Konstrukcja/wyprowadzenie i podstawowe własności: entropii łącznej, entropii warunkowej, informacji wzajemnej, entropii krzyżowej, dywergencji Kullbacka-Leiblera. Zależności między entropią łączną, entropią warunkową, informacją wzajemną, entropią krzyżową i dywergencją Kullbacka-Leiblera.

Alternatywne definicje entropii.

Różne aspekty narzędzi teorii informacji w popularnych obszarach analizy danych i eksploracji danych (informacja wzajemna w drzewach decyzyjnych, informacja wzajemna w wyszukiwaniu dokumentów, entropia krzyżowa w sieciach neuronowych, itp.).

Przykładowe zastosowania aspektów teorii informacji w bezstratnej i stratnej kompresji danych.

Tematyka zajęć

brak

Metody dydaktyczne

Wykłady: prezentacja multimedialna uzupełniona przykładami podawanymi na tablicy, demonstracja wybranych systemów wizualizacji danych.

Laboratoria: modelowanie przykładowych problemów dotyczących skalowania i wizualizacji i rozwiązywanie ich metodami dostępnymi w laboratorium, wykonywanie eksperymentów symulacyjnych, dyskusja, praca w zespole, demonstracja i pokaz multimedialny.

Literatura

Podstawowa 1. D.J.C. MacKay: Information Theory, Inference, and Learning Algorithms, Cambridge University Press, Cambridge, UK, 2003.

2. T.M. Cover, J.A. Thomas, Elements of Information Theory, 2nd Edition, Wiley and Sons, Hoboken, New Jersey, 1991.

Uzupełniająca 1. K. Sayood (red.): Lossless Compression Handbook, Academic Press, Elsevier Science, San Diego, California, 2003.

3. K. Sayood: Introduction to Data Compression, 3rd Ed., Morgan Kaufmann Publishers, San Francisco, California, 2006.

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	75	3,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	30	1,50
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwium/egzaminu, wykonanie projektu)	45	1,50