

Spracovanie prirodzeného jazyka

Kód kurzu: MLC_NLP

Kurz je zameraný na analýzu a spracovanie textov. Predpokladá sa znalosť princípov strojového učenia, ale tie najdôležitejšie koncepty budú stručne zopakované. Špecifikom spracovania textov je spôsob predspracovania údajov a ich vektorizácia. Prvá časť bude venovaná tomuto procesu. Všetko bude prakticky vyskúšané na úlohe, ktorej cieľom je klasifikácia textových dokumentov. Ďalej sa účastníci dozvedia, čo sú jazykové modely a ako ich použiť na detekciu jazyka dokumentu alebo na generovanie textov.

Požadované vstupné znalosti

- Základné znalosti programovania v jazyku Python
- Stredoškolská úroveň lineárnej algebry, matematickej analýzy a teórie pravdepodobnosti. Bude predpokladané základné porozumenie pojmom ako vektor, matica, vektorový priestor, pravdepodobnosť, podmienená pravdepodobnosť, nezávislosť náhodných javov a znalosť násobenia matíc a derivácie funkcií.
- Základy strojového učenia na úrovni nášho kurzu Úvod do strojového učenia

Študijné materiály

Príručka ku kurzu od firmy Machine Learning College

Osnova kurzu

- Úvod do spracovania prirodzeného jazyka
- Kapitoly z počítačovej lingvistiky (korpora, tokenizácia, morfológická, syntaktická a sémantická analýza, entropia, mutual information, perplexita)
- Vektorizácia textových dokumentov (bag of words, one-hot encoding, TF-IDF)
- Praktické úlohy z klasifikácie textu
- Word embedding (word2vec)
- Úvod do jazykového modelovania (n-gramové modely, vyhladzovanie, jazykové modely na báze neurónových sietí)
- Úvod do jazykových modelov (n-gramové modely, vyhadzovanie, modely založené na neurónových sieťach)
- Praktická úloha z jazykového modelovania (implementácia jazykových modelov a ich využitie pre detekciu jazyka textu)
- Úprava algoritmu pre generovanie textov

GOPAS Praha

Kodaňská 1441/46
101 00 Praha 10
Tel.: +420 234 064 900-3
info@gopas.cz

GOPAS Brno

Nové sady 996/25
602 00 Brno
Tel.: +420 542 422 111
info@gopas.cz

GOPAS Bratislava

Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 248 282 701-2
info@gopas.sk



Copyright © 2020 GOPAS, a.s.,
All rights reserved