

Strojno učenje – PMF (2016)

Materijali za definiranje projektnih tema

Tomislav Šmuc, Matija Piškorec, Matej Mihelčič

Navedeni materijali trebali bi vam pomoći u lakšem pronalaženju vama zanimljivih problema i u definiranju projektnog zadatka koji ćete rješavati u sklopu predmeta. Organizacija projektnog zadatka, način njegova izvođenja i rokovi opisani su u dokumentu „Upute za provođenje projektnih zadataka“ (na stranici kolegija Projektni zadaci). Molimo vas da se pridržavate opisanih pravila.

Osnova dobro definirane teme je skup podataka na kojem se provodi postupak učenja. Preporučamo vam da koristite jedan od već pripremljenih skupova podataka koje možete naći na linkovima u nastavku. Izrada vlastitog skupa podataka je također moguća, no u tom slučaju na vama je odgovornost da predvidite realno vrijeme za njegovu pripremu kako bi ostalo dovoljno vremena i za ostale faze projekta.

UCI Machine Learning Repository: <http://archive.ics.uci.edu/ml/datasets.html>

Datasets for Data Mining: <http://www.inf.ed.ac.uk/teaching/courses/dme/2012/datasets.html>

Stanford Large Network Dataset Collection: <http://snap.stanford.edu/data/>

Social Computing Research at the University of Minnesota: <http://grouplens.org/datasets/movielens/>

The Koblenz Network Collection: <http://konect.uni-koblenz.de/>

Google Ngram Viewer: <http://storage.googleapis.com/books/ngrams/books/datasetv2.html>

Million Song Dataset: <http://labrosa.ee.columbia.edu/millionsong/>

Youtube data: <http://an.kaist.ac.kr/traces/IMC2007.html>

Kaggle challenges: <http://www.kaggle.com/competitions>

Na bilo koji od skupova podataka za koji se odlučite, trebate primijeniti adekvatne algoritme strojnog učenja (filtriranje podataka, selekcija varijabli, klasifikacija, regresija), metode evaluacije i vizualizacije rezultata.

Osim algoritama koji se obrađuju u okviru predavanja, možete se odlučiti da istražite i koristite neke od novijih algoritama koji su aktivno područje istraživanja. Neki od problema odnosno skupova podataka zahtijevaju primjenu specifičnih algoritama strojnog učenja:

Deep learning: http://deeplearning.stanford.edu/wiki/index.php/Main_Page

Factorization Machines: <http://www.ismll.uni-hildesheim.de/pub/pdfs/Rendle2010FM.pdf>

Data streams mining: <http://i.stanford.edu/~ullman/mmds/ch4.pdf>

Recommender systems: <http://www.recommenderbook.net/teaching-material>

Social and Information Network Analysis: <http://web.stanford.edu/class/cs224w/resources.html>

Topic modeling: www.cs.princeton.edu/~blei/topicmodeling.html

Relation extraction from text: <http://www.cs.cmu.edu/~nbach/papers/A-survey-on-Relation-Extraction.pdf>

Opinion Mining and Sentiment Analysis: <http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf>