

Strojno učenje – PMF (2013)

Popis ponuđenih projektnih tema

Tomislav Šmuc, Tomislav Lipić, Matija Piškorec

Uvod

Popis ponuđenih tema trebao 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 [SU_2013_Projekti_Pravila.docx](#). Molimo vas da se pridržavate opisanih pravila.

1	Trenutno otvoreni natječaji vezani za strojno učenje	2
1.1	@Kaggle	2
1.2	PowerTAC.....	2
2	Napredne metode strojnog učenja.....	3
2.1	Deep learning	3
2.2	Factorization machines	3
2.3	Submodularity	3
2.4	Dubinska analiza nizova podataka (engl. <i>data streams data mining</i>).....	3
2.5	Aktivno učenje (engl. <i>Active learning</i>).....	4
3	Modeliranje i analiza društvenih mreža i društvenih medija	5
3.1	Analiza društvenih mreža i medija	5
3.2	Otkrivanje mišljenja i analiza stava (engl. <i>Opinion Mining & Sentiment Analysis</i>)	5
3.3	Predviđanje veza u kompleksnim mrežama (engl. <i>link prediction</i>)	6
3.4	Otkrivanje zajednica u kompleksnim mrežama (engl. <i>community detection</i>)	6
4	Obrada i analiza teksta.....	7
4.1	Modeliranje tema iz tekstualnih sadržaja (engl. <i>topic modelling</i>).....	7
4.2	Izlučivanje relacija (engl. <i>relation extraction</i>)	7
4.3	Druge teme obrade prirodnog jezika (engl. <i>natural language processing</i>)	7
5	Analiza slike, videa i zvuka.....	8
5.1	Prepoznavanje vizualnog identiteta (logotipa) u videu	8
5.2	Detekcija i prepoznavanje ljudi i lica na slikama (engl. <i>Human and face detection and recognition</i>)	8
5.3	Identifikacija pjesama	8
5.4	Detekcija i klasifikacija prijelaza u videu	9
6	Dubinska analiza nekog izvora podataka (engl. <i>data mining</i>)	10

1 Trenutno otvoreni natječaji vezani za strojno učenje

1.1 @Kaggle

- Popis svih natječaja: <http://www.kaggle.com/competitions>

1.2 PowerTAC

Ovo natjecanje adresira problem ekoloških, ekonomskih i društvenih potreba održivog razvoja primjenom računarstva (engl. [computational sustainability](#)) sa specifičnim fokusom na održive elektroenergetske sustave primjenom agentskog modeliranja i simulacije. Jedan od pristupa natjecanju uključuje istraživanja u području učenja agenata (engl. [agent learning](#)) korištenjem tehnika za učenje podrškom (engl. *reinforcement learning*).

- Stranica natječaja i osnovne informacije:
<http://www.powertac.org/>
- Opis i pravila za 2012:
<http://repub.eur.nl/res/pub/37192/ERS-2012-010-LIS.pdf>
- Diplomski radovi – opis simulacijske platforme:
 - A SIMULATION PLATFORM FOR POWER TRADING,
http://www.fer.unizg.hr/download/repository/Master_Thesis_-_Jurica_Babic.pdf
 - Designing tariffs in a competitive energy market,
<http://oathesis.eur.nl/ir/repub/asset/11244/11244-Knijff.pdf>
- Članci
 - Primjena učenja podrškom (engl. *reinforcement learning*) za PowerTAC
 - Learned Behaviors of Multiple Autonomous Agents in Smart Grid Markets,
<http://www.cs.cmu.edu/afs/cs/Web/People/ppr/papers/ReddyVeloso-AAAI11.pdf>
 - Factored Models for Multiscale Decision-Making in Smart Grid Customers,
<http://www.cs.cmu.edu/afs/cs/user/mmv/www/papers/12aaai-ReddyVeloso.pdf>
 - Uvod u algoritme za učenje s podrškom
 - Algorithms for Reinforcement Learning,
<http://www.sztaki.hu/~szcsaba/papers/RLAlgsInMDPs-lecture.pdf>
 - A Tutorial on Reinforcement Learning Techniques,
http://www.ppgia.pucpr.br/~fabricio/ftp/Aulas/Mestrado/AS/Artigos-Apresentacoes/Aprendizagem%20por%20Reforco/TAIC-tutorial_RL.pdf
 - Istraživanja u području učenja agenata (Workshop@AAMAS)
 - Materijali s workshopa ALA2012, ALA2011,...
<http://swarmlab.unimaas.nl/ala2013/>
 - ALA2012: <http://ai.vub.ac.be/ALA2012/> (Proceedings)
- Tehničke upute: uspostavljanje simulacijske platforme (poslužitelj za testiranje) i način razvijanja agenata (implementacija Brookera):
 - <http://power-tac-developers.975333.n3.nabble.com/>
 - http://www.powertac.org/wiki/index.php/Getting_Started
 - <https://github.com/powertac/powertac-server/wiki/Getting-started>

2 Napredne metode strojnog učenja

2.1 Deep learning

- Stanford course (video 1 i video 2),
<http://www.stanford.edu/class/cs294a/handouts.html>
- Deep learning and unsupervised feature learning tutorial,
http://deeplearning.stanford.edu/wiki/index.php/Main_Page
- Representation Learning Tutorial,
<http://www.iro.umontreal.ca/~bengioy/talks/deep-learning-tutorial-2012.html>
- Representation Learning: A Review and New Perspectives, <http://arxiv.org/abs/1206.5538>
- ICML 2012 Workshop on Representation Learning,
http://techtalks.tv/icml_2012_representation_learning/
- Deep learning TOOLS,
http://deeplearning.net/software_links/
- Deep learning for NLP,
<http://www.socher.org/index.php/DeepLearningTutorial/DeepLearningTutorial>

2.2 Factorization machines

- Factorization Machines, <http://www.ismll.uni-hildesheim.de/pub/pdfs/Rendle2010FM.pdf>
- Primjena na društvene mreže:
 - Social Network and Click-through Prediction with Factorization Machines
 - Clanak: <http://www.inf.uni-konstanz.de/~rendle/pdf/Rendle2012-KDDCup.pdf>
 - Prezentacija: http://www.inf.uni-konstanz.de/~rendle/pdf/Rendle2012-KDDCup_presentation.pdf
- Alati:
 - Factorization Machines with libFM,
<http://www.csie.ntu.edu.tw/~b97053/paper/Factorization%20Machines%20with%20libFM.pdf>
 - libFM, <http://www.libfm.org/>

2.3 Submodularity

- Tutoriali:
 - Beyond Convexity - Submodularity in Machine Learning, <http://submodularity.org/>
 - Intelligent Information Gathering and Submodular Function Optimization,
<http://submodularity.org/ijcai09/index.html>
 - Tutorial on Submodularity in Machine Learning and Computer Vision
<http://submodularity.org/submodularity-2012.pdf>
- Kolegiji:
 - Submodularity Functions, Optimization, and Application to Machine Learning (Washington University):
http://j.ee.washington.edu/~bilmes/classes/ee596a_fall_2012/#lectures
- Članak:
 - Submodular Function Maximization, <http://las.ethz.ch/files/krause12survey.pdf>

2.4 Dubinska analiza nizova podatka (engl. *data streams data mining*)

- Poglavlje iz knjige *Mining of Massive Datasets*: <http://i.stanford.edu/~ullman/mmds/ch4.pdf>
- Mining Data Streams: A Review:
<http://www.sigmod.org/publications/sigmod-record/0506/p18-survey-gaber.pdf>
- Data Stream Mining – A Practical Approach:
<http://www.cs.waikato.ac.nz/~abifet/MOA/StreamMining.pdf>

2.5 Aktivno učenje (engl. *Active learning*)

- Osnovni materijali:
 - Active Learning Tutorial (ICML 2009), http://hunch.net/~active_learning/
 - Active Learning Challenge, <http://www.causality.inf.ethz.ch/activelearning.php>
- Video predavanja:
 - Theory, Methods and Applications of Active Learning, http://videlectures.net/mlss09us_nowak_castro_tmaal/
 - Active learning, http://videlectures.net/icml09_dasgupta_langford_actl/
- Pregledni članci:
 - Active Learning Literature Survey, <http://www.cs.cmu.edu/~bsettles/pub/settles.activelearning.pdf>
- Članci za specifičnu primjenu:
 - A literature survey of active machine learning in the context of natural language processing, <http://eprints.sics.se/3600/>
 - Active Learning with Transfer Learning, <http://www.aclweb.org/anthology-new/W/W12/W12-3303.pdf>
 - Active Learning in Recommender Systems, <http://machinelearning202.pbworks.com/w/file/fetch/45230787/Rubens-Active-Learning-RecSysHB2010.pdf>

3 Modeliranje i analiza društvenih mreža i društvenih medija

3.1 Analiza društvenih mreža i medija

- Članci:
 - Effects of User Similarity in Social Media, <http://cs.stanford.edu/people/jure/pubs/similarity-wsdm12.pdf>
- Tutoriali i predavanja:
 - Jure Leskovec, <http://cs.stanford.edu/people/jure/>
 - KDD 2011 Tutorial - Social Media Analytics, <http://snap.stanford.edu/proj/socmedia-kdd/>
 - Modeling Social and Information Networks: Opportunities for Machine Learning, http://videolectures.net/icml09_leskovec_msain/
- Kolegij i lista osnovnih resursa*
 - Podaci, alati i kolegiji: <http://www.stanford.edu/class/cs224w/resources.html>
 - Neke teme: <http://www.stanford.edu/class/cs224w/projects.html>
 - Predavanja: <http://www.stanford.edu/class/cs224w/handouts.html>
- Podaci:
 - Twitter streaming api, <https://dev.twitter.com/>
 - Trust network datasets, <http://www.trustlet.org/wiki/Datasets>
http://www.trustlet.org/wiki/Repositories_of_datasets
http://www.trustlet.org/wiki/Trust_network_datasets#Released_datasets
- Moguće teme i korisni alati:
 - Twitter Learning Analytics in R, http://www.ascilite2012.org/images/custom/walker_lyndon_-_twitter_learning.pdf
 - GraphLab, <http://graphlab.org>
 - Storm, <http://storm-project.net/>
 - Real-time feed processing with Storm, <http://www.datasalt.com/2012/01/real-time-feed-processing-with-storm/>

3.2 Otkrivanje mišljenja i analiza stava (engl. *Opinion Mining & Sentiment Analysis*)

- Pregledni članci
 - Opinion Mining and Sentiment Analysis, <http://www.cs.cornell.edu/home/llee/omsa.pdf>
 - 1. Poglavlje daje dobar uvod u problem
 - 7. Poglavlje izlistava izvore skupova podataka koji se mogu koristiti pri izboru ove teme
 - Sentiment Analysis and Opinion Mining: A Survey, http://www.ijarcse.com/docs/papers/June2012/Volume_2_issue_6/V2I600263.pdf
 - Sentiment Analysis and Subjectivity, <http://www.cs.uic.edu/~liub/FBS/NLP-handbook-sentiment-analysis.pdf>
- Diplomski radovi
 - Sentiment Analysis – A multi modal approach, <http://www.doc.ic.ac.uk/teaching/distinguished-projects/2011/l.carstens.pdf>
- Članci:
 - Exploiting Social Relations for Sentiment Analysis in Microblogging, <http://www.public.asu.edu/~xiah/papers/wsdm13Hu.pdf>
 - Methods for Sentiment Analysis of Twitter Messages, <http://www.fruct.org/publications/abstract12/files/Bar.pdf>

3.3 Predviđanje veza u kompleksnim mrežama (engl. *link prediction*)

- Općenita predavanja za analizu društvenih mreža:
 - KDD 2011 Tutorial - Social Media Analytics, <http://snap.stanford.edu/proj/socmedia-kdd/>
 - <http://cs.stanford.edu/people/jure/icml09networks/>
 - http://videolectures.net/www2010_leskovec_ppn/
- Diplomski rad:
 - Link Prediction and the Evolution of Communities on Twitter, <http://repository.tudelft.nl/view/ir/uuid%3Af943dcfd-b11c-4b2c-ab55-db8d04ee0b28/>
- Članci
 - Relevantni članci i pripadni podaci, http://malt.ml.cmu.edu/mw/index.php/Link_Prediction
 - Supervised Random Walks: Predicting and Recommending Links in Social Networks
 - <http://cs.stanford.edu/people/jure/pubs/linkpred-wsdm11.pdf>
 - http://malt.ml.cmu.edu/mw/index.php/Supervised_Random_Walk
 - Link Prediction in Complex Networks: A Survey, <http://arxiv.org/abs/1010.0725>
 - New Perspectives and Methods in Link Prediction, http://users.cs.fiu.edu/~lzhen001/activities/KDD_USB_key_2010/docs/p243.pdf
 - Link Prediction using Supervised Learning, <http://www.siam.org/meetings/sdm06/workproceed/Link%20Analysis/12.pdf>
 - Graph-based Features for Supervised Link Prediction, <http://www.islab.ntua.gr/attachments/article/63/06033365.pdf>
- Kolegij i lista osnovnih resursa*
 - Podaci, alati i kolegiji: <http://www.stanford.edu/class/cs224w/resources.html>
 - Neke teme: <http://www.stanford.edu/class/cs224w/projects.html>
 - Predavanja: <http://www.stanford.edu/class/cs224w/handouts.html>
- Ostali podaci i alati:
 - <http://konect.uni-koblenz.de/>
 - LPmade: Link Prediction Made Easy, <http://jmlr.csail.mit.edu/papers/volume12/lichtenwalter11a/lichtenwalter11a.pdf>
 - PowerLab, <http://graphlab.org/powergraph-presented-at-osdi/>

3.4 Otkrivanje zajednica u kompleksnim mrežama (engl. *community detection*)

- Predavanje o kontekstu i primjeni:
 - Networks, communities and the ground-truth, http://videolectures.net/solomon_leskovec_networks/
- Community detection in graphs
 - Predavanje: http://videolectures.net/ephdc08_fortunato_csig/
 - Članak: <http://arxiv.org/pdf/0906.0612v2.pdf>
- Empirical Comparison of Algorithms for Network Community Detection
 - Predavanje: http://videolectures.net/www2010_leskovec_eca/
 - Članak: <http://cs.stanford.edu/people/jure/pubs/communities-www10.pdf>
- Kolegij i lista osnovnih resursa*
 - Podaci, alati i kolegiji: <http://www.stanford.edu/class/cs224w/resources.html>
 - Neke teme: <http://www.stanford.edu/class/cs224w/projects.html>
 - Predavanja: <http://www.stanford.edu/class/cs224w/handouts.html>
- Ostali mogući alati:
 - <http://www.elemartelot.org/index.php/programming/cd-code>
 - Community Detection In R, <http://igraph.wikidot.com/community-detection-in-r>
 - GraphLab, <http://graphlab.org/>
 - Python, Java, C++, Matlab, R libs

4 Obrada i analiza teksta

4.1 Modeliranje tema iz tekstualnih sadržaja (engl. *topic modelling*)

- Osnovne informacije:
www.cs.princeton.edu/~blei/topicmodeling.html
- Alati:
 - Mallet,
<http://mallet.cs.umass.edu/topics.php>
 - GraphLab TM Toolkit,
<http://graphlab.org/toolkits/topic-modelling/>
 - TM visualisation engine,
<http://code.google.com/p/tmve/>
 - Matlab Topic Modeling Toolbox,
http://psiexp.ss.uci.edu/research/programs_data/toolbox.htm
 - A Topic Modeling Toolbox Using Belief Propagation, <http://jmlr.csail.mit.edu/mloss/>
 - <https://code.google.com/p/tmbp-topicmodel-beliefpropagation/>
- Podaci:
 - Označeni dokumenti dobiveni od hrvatske firme i javno dostupni korpusi (po potrebi)

4.2 Izlučivanje relacija (engl. *relation extraction*)

- A Review on Relation Extraction:
 - Prezentacija:
<http://www.cs.cmu.edu/~nbach/papers/A-survey-on-Relation-Extraction-Slides.pdf>
 - Članak: <http://www.cs.cmu.edu/~nbach/papers/A-survey-on-Relation-Extraction.pdf>
- Članci:
 - Social relation extraction from texts using a support-vector-machine-based dependency trigram kernel [pristup članku s poveznice moguć je samo iz mreže Sveučilišta], <http://www.sciencedirect.com/science/article/pii/S0306457312000544>
 - Automatic relation extraction among named entities from text contents,
<http://www.scholarbank.nus.edu.sg/bitstream/handle/10635/16136/PhD%20Thesis%20by%20Chen%20Jinxu.pdf?sequence=1>
 - Co-reference Resolution and Relationship extraction,
http://www.iicm.tugraz.at/0x811bc82b_0x0011c034
- Podaci:
 - Datasets for generic relation extraction,
<http://benhachey.info/pubs/hachey-nle11-datasets-preprint.pdf>
 - Označeni dokumenti dobiveni od hrvatske firme i javno dostupni korpusi (po potrebi)
- Alati:
 - Apache UIMA project, <http://uima.apache.org/>

4.3 Druge teme obrade prirodnog jezika (engl. *natural language processing*)

- Alati:
 - MALLET, <http://mallet.cs.umass.edu/>
 - Apache UIMA, <http://uima.apache.org/index.html>
 - Naći/Istražiti/proučiti ostale alate za NLP (Matlab, R, Python, C++, Java)
- @Coursera
 - Natural Language Processing, Columbia, <https://www.coursera.org/course/nlangp>
 - Natural Language Processing, Stanford, <https://www.coursera.org/course/nlp>
- Skupovi podataka:
 - <http://research.microsoft.com/nlp/>
 - <http://nlp.stanford.edu/links/statnlp.html>
 - <http://trec.nist.gov/data/reuters/reuters.html>

5 Analiza slike, videa i zvuka

5.1 Prepoznavanje vizualnog identiteta (logotipa) u videu

- Prezentacije:
 - Real-time Logo Detection and Tracking, <http://www.slideshare.net/melgeorge/realtime-logo-detection-and-tracking>
- Članci:
 - Logo detection in high-motion sports video, http://www.stanford.edu/~afaraujo/Final%20report_Andre_Stephanie.pdf
 - Spatial connected component pre-locating algorithm for rapid logo detection, http://www.mirlab.org/conference_papers/International_Conference/ICASSP%202012/pdfs/0001297.pdf
- Alati:
 - CVT, <http://www.cnas.utulsa.edu/cvt/>

5.2 Detekcija i prepoznavanje ljudi i lica na slikama (engl. *Human and face detection and recognition*)

Primjer projekta ove teme može uključivati sustav koji će za danu sliku lica osobe provjeriti koliko je pouzdana dana slika. Sustav možete zamisliti kroz:

1. Provjeru je li slika sadrži jedno lice čovjeka (*human and face detection*)
 2. Provjeru je li detektirano lice na slici odgovara nekim drugim licima iz baze korisnika (*face recognition*)
- Izvor informacija: <http://www.facedetection.com/>
 - Viola Jones face detection framework – prvi sustav za detekciju lica u realnom vremenu , http://research.microsoft.com/en-us/um/people/viola/Pubs/Detect/violaJones_CVPR2001.pdf
 - Face Recognition by Humans: 20 Results all Computer Vision Researchers Should Know About, http://web.mit.edu/bcs/sinha/papers/20Results_2005.pdf
 - Algoritmi za prepoznavanje lica, <http://www.face-rec.org/algorithms/#Image>
 - Face Recognition via Sparse Representation, <http://perception.csl.uiuc.edu/recognition/Home.html>
 - Detekcija ljudi na slici, <http://www.merl.com/reports/docs/TR2006-068.pdf>
 - Face Recognition Algorithms, <http://www.ehu.es/ccwintco/uploads/e/eb/PFC-IonMarques.pdf>
 - Face Recognition with Learning-based Descriptor, http://research.microsoft.com/en-us/um/people/jiansun/papers/cvpr10_facereco.pdf
 - OpenCV Detector, <http://www.shervinemami.info/faceRecognition.html>
 - Popis skupova podataka:
 - <http://vision.ai.uiuc.edu/mhyang/face-detection-survey.html#face-database>
 - <http://www.facedetection.com/facedetection/datasets.htm>
 - <http://www.face-rec.org/databases/>
 - <http://www.idiap.ch/resource/facefaces/>
 - <http://vis-www.cs.umass.edu/lfw/>

5.3 Identifikacija pjesama

- Članci:
 - Learning sparse feature representation for music annotation and retrieval, <https://ccrma.stanford.edu/~juhan/pubs/jnam-ismir2012.pdf>
 - "An Industrial-Strength Audio Search Algorithm": <http://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf>
 - Primjer jednostavne implementacije u Javi, <http://www.redcode.nl/blog/2010/06/creating-shazam-in-java/>
- Alati:
 - Echoprint.me, <http://echoprint.me/>

5.4 Detekcija i klasifikacija prijelaza u video

- Pregledni članci:
 - Reliable Transition Detection in Videos: A Survey and Practitioner's Guide, http://mmc36.informatik.uni-augsburg.de/mediawiki-1.11.2/images/6/6a/IJIG_AUG2001.pdf
 - A Survey of Recent Work in Video Shot Boundary Detection, http://pdf.aminer.org/000/262/586/shot_segmentation_using_a_coupled_markov_chains_representation_of_video.pdf
- Ostali članci
 - Detection and Classification of Shot Transitions, <http://www.cs.bris.ac.uk/Publications/Papers/1000570.pdf>
 - Simultaneous detection of abrupt cuts and dissolves in videos using support vector machines, <http://www.cs.uoi.gr/~arly/papers/PRL09.pdf>
 - A Rapid and Robust Method for Shot Boundary Detection and Classification in Uncompressed MPEG Video Sequences, <http://ijcsi.org/papers/IJCSI-9-5-2-368-374.pdf>
 - Unified Approach to Detect Shot Transitions in Video Signals, <http://cspl.postech.ac.kr/paper/international%20Conference/2012%20Unified%20Approach%20to%20Detect%20Shot%20Transitions%20in%20Video%20Signals.pdf>
 - Video shot boundary detection using generalized eigenvalue decomposition and gaussian transition detection, <http://webpages.iust.ac.ir/mahfathy/Papers/VIDEO%20SHOT%20BOUNDARY%20DETECTION%20USING%20Generalized%20Eigenvalue%20Decomposition%20and%20Gaussian%20Transition%20Detection.pdf>
- Podaci
 - Snimke dobivene od hrvatske firme i proizvoljne YouTube video snimke
- Alati
 - Matlab, OpenCV, C++

6 Dubinska analiza nekog izvora podataka (engl. *data mining*)

U okviru ove skupine predlaže se odabir nekog problema dubinske analize podataka (engl. *data mining*) i njegovo rješavanje primjenom metoda strojnog učenja. Opširan popis skupova podataka prikladnih za neke dolje spomenute teme/probleme može se naći na:

- <http://kevinchai.net/datasets>
- <http://www.datawrangling.com/some-datasets-available-on-the-web>
- <http://theinfo.org/>
- Data Mining Datasets: <http://www.inf.ed.ac.uk/teaching/courses/dme/2012/datasets.html>
 - Ovdje se za svaki skup podataka nalazi pripadni opis problema. Većinom su to problemi klasifikacije ili klasteriranja. Za svaki problem priloženi su opisi prijašnjih odabranih metoda koje mogu pomoći pri rješavanju odabranog problema.
- KDnuggets: <http://www.kdnuggets.com/datasets/index.html>
 - UCI Machine Learning Repository,
<http://archive.ics.uci.edu/ml/>
 - Data Resources presentation – Data Analysis Course @Coursera:
<https://d19vezwu8euf16.cloudfront.net/dataanalysis/dataResources.pdf>
- @Coursera
 - Data Analysis, Johns Hopkins Bloomberg,
<https://www.coursera.org/course/dataanalysis>