

## **Laboratorij za kognitivno modeliranje (LKM) na FRI**

LKM se nahaja v kletnem prostoru zraven PR4 (pod PR3). Trenutno ima pet rednih članov in več zunanjih sodelavcev. Stalni delovni prostor v laboratoriju si delijo asistent mag. Zoran Bosnić (vaje iz algoritmov in podatkovnih struktur ter programskih jezikov), mladi raziskovalec Erik Štrumbelj (oba pripravljata doktorat v okviru raziskovalnega dela v laboratoriju), ter občasno zunanji član dr. Luka Šajn. Poleg njiju sta člana laboratorija še doc. dr. Marko Robnik Šikonja (predavanja iz analize podatkov ter vaje iz algoritmov in podatkovnih struktur) ter doc. dr. Matjaž Kukar (predavanja in vaje iz podatkovnih baz). Do konca novembra 2007 je začasno član laboratorija tudi Raul Fidalgo Merino z Univerze v Malagi (Španija), ki pri nas končuje svojo doktorsko disertacijo s področja učenja iz podatkovnih tokov.

Predstojnik LKM je prof. dr. Igor Kononenko (predava algoritme in podatkovne strukture, strojno učenje in podatkovno rudarjenje, programske jezike, ter metode umetne inteligence). Iščemo pa tudi novega kandidata za mladega raziskovalca, ki bo nastopil delovno mesto jeseni 2008.

Laboratorij za kognitivno modeliranje (LKM) je bil ustanovljen leta 2001, predvsem zaradi prenatrpanosti Laboratorija za umetno inteligenco (LUI), iz katerega je LKM izšel, še vedno pa z LUI raziskovalno tesno sodelujemo. V LKM raziskujemo na področjih kognitivnega modeliranja, strojnega učenja, nevronske mreže, odkrivanja znanja v podatkih in v slikah. Strojno učenje in podatkovno rudarjenje se ukvarja z iskanjem zakonitosti v relativno velikih bazah podatkov, ki so na voljo za učenje modelov. Naučeni modeli služijo za razlago podatkov, za simulacije, nadzor, napovedovanje in za reševanje novih problemov. Tipičen primer je npr. medicinska diagnostika. Iz podatkov o pacientih, ki jih je nek zdravnik ali klinika zdravila v preteklosti, sistem za strojno učenje zgradi model, ki se lahko uporabi za diagnosticiranje novih pacientov. Pri tem so pomembni predvsem točnost in zanesljivost napovedi ter zmožnost razlage odločitev modela. Strojno učenje je široko področje, ki vključuje učenje logičnih modelov (odločitvena drevesa, pravila, logični programi), verjetnostnih modelov (varianete Naivnega Bayesovega klasifikatorja, bayesovske mreže), umetnih nevronske mreže ter numeričnih modelov (npr. metoda podpornih vektorjev).

Raziskovalna področja, s katerimi se intenzivno ukvarjamo, so modeliranje iz šumnih podatkov, ki je povezano s kognitivnimi, medicinskimi, biološkimi in drugimi procesi. Razvijamo, testiramo in apliciramo nove pristope in algoritme za modeliranje iz numeričnih, simboličnih in slikovnih podatkov. LKM sodeluje s psihologi, zdravniki, biologi, fiziki in kemiki. Pomemben del raziskav je povezan z analizo slik, medicinsko diagnostiko ter raziskavami vplivov različnih polj na vitalnost rastlin in ljudi. Raziskovalno sodelujemo z več domačimi in tujimi inštitucijami, med drugim Univerzo v Portu (Portugalska), Univerzo v Ioannini (Grčija), Univerzo v Hasseltu (Belgija), Univerzo v Malagi (Španija), Univerzo v St. Petersburgu (Rusija), ter Inštitutom za organsko kmetijstvo (brez kemije) v Fricku (Švica).

Kvaliteto našega pedagoškega dela in napisanih učbenikov boste študenti lahko presodili sami. Kvaliteto raziskovalnega dela pa izkazujemo s številnimi objavami v uglednih znanstvenih revijah in mednarodnih znanstvenih konferencah ter s preko 300

citati naših del s strani drugih avtorjev v uglednih (SCI) znanstvenih objavah. V juniju 2007 je pri ugledni znanstveni založbi Horwood Publishing (Anglija) izšla knjiga »Machine Learning and Data Mining: Introduction to Principles and Algorithms« avtorjev Igorja Kononenka in Matjaža Kukarja (slika 3). Ta objava predstavlja priznanje raziskovalni skupini in zaokroža dolgoletno raziskovalno delo.

Trenutno raziskovalno delo v LKM lahko razdelimo na več področij:

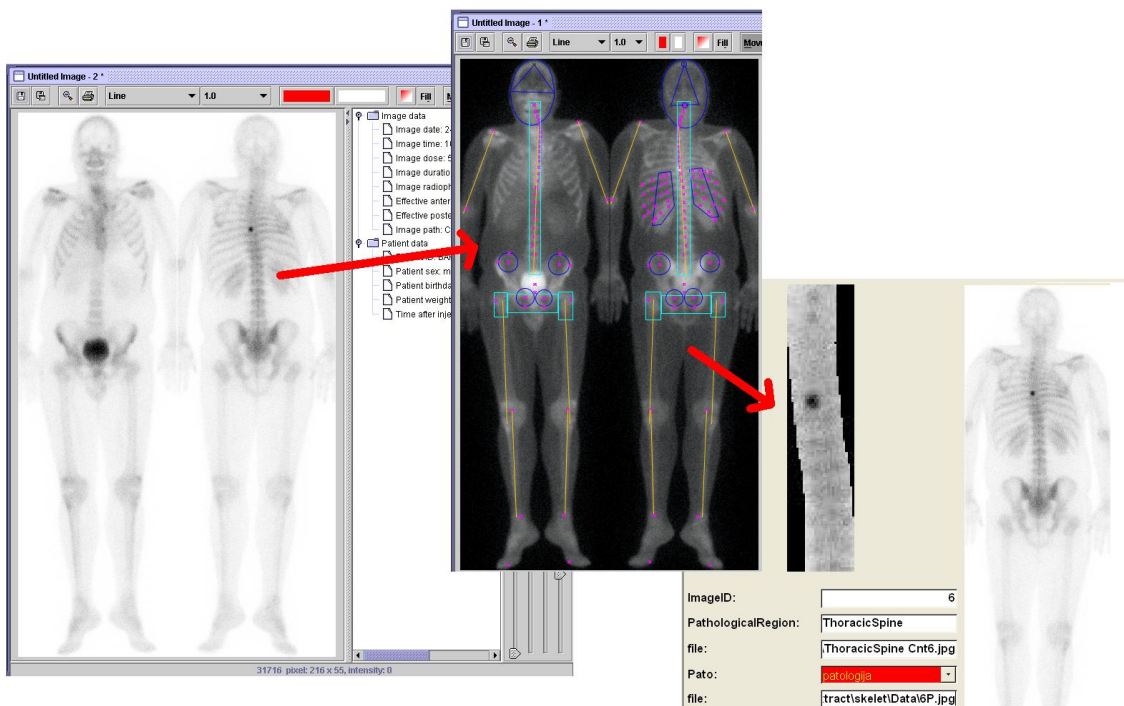
- **Modeliranje, analiza, parametrizacija in klasifikacija slik.** Tu gre za razvijanje algoritmov, ki sliko opišejo z množico numeričnih parametrov (značilik, atributov), ki jih potem lahko uporabimo za klasifikacijo slik z metodami strojnega učenja. Ti algoritmi se lahko aplicirajo na različnih področjih. Eno teh področij so medicinske slike. Trenutno se v sodelovanju z zdravniki iz Klinike za nuklearno medicino v Ljubljani ukvarjamo z analizo scintigrafskih slik okostij ter scintigrafskih slik srca (glej sliki 1 in 2). V sodelovanju z Univerzo v St.Petersburgu analiziramo slike koron prstov ljudi, v sodelovanju z Inštitutom za organsko kmetijstvo v Švici pa slike koron delov rastlin.
- **Razvoj metod ocenjevanje zanesljivosti napovedi algoritmov strojnega učenja.** Rezultat večine metod strojnega učenja so modeli, ki se uporabljajo za napovedovanje pri reševanju novih problemov (npr. za diagnosticiranje novih pacientov). Pri napovedovanju je zelo pomembna ocena zanesljivosti napovedi (npr. zanesljivost dianoze). Razvijamo splošne postopke, ki omogočajo ocenjevanje zanesljivosti glede na lastnosti modela in glede na lastnosti novega problema (npr. nekaterim pacientom se da zanesljivo postaviti diagnozo, za druge pa je potrebno poiskati dodatne podatke, kot so laboratorijski izvidi ipd.). Postopki naj bi bili splošno uporabni, t.j. neodvisni od izbire metode strojnega učenja.
- **Razvoj metod za razlago odločitev, ki jih ponujajo modeli, zgrajeni z metodami strojnega učenja.** Poleg točnosti napovedovanja (odločanja), je pri uporabi modelov, ki jih dobimo z metodami strojnega učenja, zelo pomembna zmožnost razlage odločitev. Iz razlage lahko uporabnik bolj ali manj zaupa odločitvi sistema, hkrati pa lahko preverja odločitev glede na svoje predznanje. Razvijamo splošne postopke za razlago odločitev, ki naj bi bili neodvisni od izbire metode strojnega učenja.
- **Ocenjevanje atributov v strojnem učenju.** V strojnem učenju je pomembno, iz kakšnih podatkov se učimo. Kvalitetnejši podatki nam zagotavljajo boljše rezultate strojnega učenja. Podatki so opisani z množico atributov (značilik). Pri tem je lahko število atributov zelo veliko (nekaj tisoč) in seveda niso vsi enako kvalitetni. Preveč (nekvalitetnih in/ali redundantnih) atributov prepreči algoritmom strojnega učenja, da bi zgradili kvaliteten model. Poleg tega je zbiranje atributov pogosto drago, zato je reduciranje števila atributov tudi s tega stališča lahko zelo koristno. Razvijamo postopke za učinkovito ocenjevanje kvalitete atributov, ki so si lahko med seboj močno soodvisni ter prilagajamo postopke za potrebe posameznih problemskih področij.
- **Aplikacije algoritmov strojnega učenja in podatkovnega rudarjenja.** Uporabljamo obstoječa komercialna in prosto dostopna orodja kot tudi svoja lastna orodja za strojno učenje in podatkovno rudarjenje na različnih aplikativnih področjih: analiza slik (medicinska diagnostika, analiza mikroskopskih posnetkov

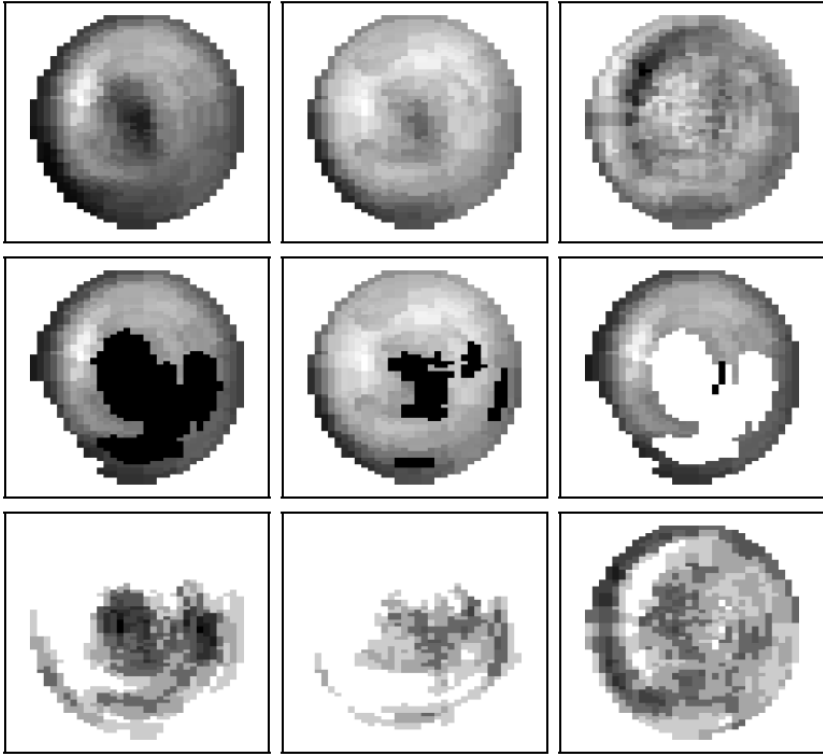
kapljic, analiza koron), energetika (napovedovanje električnih tokov), svetovni splet (analiza web logov), tržne raziskave v marketingu, športne stave (nogomet, konjske dirke), telekomunikacije (profil uporabnikov mobilnih storitev), zavarovalništvo (ocenjevanje zanesljivosti klientov), itd.

Tekoči projekti v letu 2007:

- Reliable and Comprehensible Machine Learning Approaches with Applications to Medical Diagnostics and Bioinformatics, bilateralni projekt z Univerzo v Ioannini, Grčija (2005-2007)
- Machine Learning of Probabilities with Applications to Web Portals and Medical Diagnostics, bilateralni projekt z Univerzo v Portu, Portugalska (2006-2007)
- Raziskovalni program Umetna inteligenca in inteligentni sistemi P2-0209 (2004 – 2008)
- Data Mining: Theory and Applications (DAMITA), mednarodni projekt z Univerzo v Malagi, Španija (2008-2010); v postopku sprejemanja

Kljub prostorski stiski je vzdušje v LKM prijetno in raziskovalno zanimivo. Študenti višjih letnikov se vključujejo v raziskovalno delo predvsem preko seminarških in diplomskih nalog, podiplomski študentje pa preko magistrskih in doktorskih del. Vsi ki vas raziskovalno delo zanima, ste dobrodošli tudi mimo pedagoškega procesa.





# MACHINE LEARNING AND DATA MINING

Introduction to Principles  
and Algorithms

IGOR KONONENKO  
and  
MATJAŽ KUKAR