

RADIONICA I KONFERENCIJA

GEOSTATISTIKA I STROJNO UČENJE: PRIMJENA U ZNANOSTIMA O KLIMI I OKOLIŠU

Beograd, Srbija, 21.-24. lipnja 2016.

Irena Nimac, mag. phys.-geophys.



Slika 1: Organizatori radionice Tomislav Hengl i Melita Perčec Tadić s DHMZ timom.

Od 21. do 24. lipnja 2016. na Građevinskom fakultetu Sveučilišta u Beogradu održana je radionica i konferencija pod nazivom „GeoMLA - Geostatistics and Machine Learning“. Prva dva dana održane su paralelne radionice na temu strojnog učenja i prostorno-vremenskih objekata u softveru otvorenog koda (*open source*) R, dok su posljednja dva dana prezentirani radovi vezani uz teme poput: primjena strojnog učenja u klimatologiji, geoinformatika u modeliranju i analizi klimatskih podataka, kombiniranje geostatistike i daljinskih mjerenja, kombiniranje strojnog učenja i daljinskih mjerenja, prostorno-vremenske varijacije klimatskih varijabli, prostorno-vremenske interpolacije klimatskih podataka, itd.

Prvi dan radionicu je održao Mikhail Kanevski, profesor na sveučilištu u Lozani, pod nazivom *'Machine Learning of geospatial data: achievements and new trends'*. Fokus je na *data-driven* modelima, za razliku od do sada korištenih

geostatističkih, stoga se pri korištenju metoda strojnog učenja posebno ističe pitanje kvalitete i reprezentativnosti podataka (nehomogenost, grupiranje, preferencijalno uzrokovanje...). U smislu prostorno-vremenskih podataka taj se postupak svodi na tri koraka: traženje, modeliranje te predikciju uzorka (struktura). Cijeli set podataka podijeli se u tri dijela: trening (prilagodba modela), validacija (odabir najboljeg modela) i test (procjena modela). Neki od korištenih ML algoritama su: umjetne neuronske mreže (ANN) različitih arhitektura (višeslojne perceptronske (MLP), samoorganizirajuće mape (SOM), probabilističke neuronske mreže...), *Random Forest*, *Ensemble Learning*, *Support Vector Machine*, *Support Vector Regression* itd. Bitan korak u cijelom tom procesu je preliminarna analiza i vizualizacija podataka s ciljem maksimizacije uvida u skup podataka, otkrivanja osnovne strukture, detekcije netipičnih vrijednosti (*outliers*) i anomalija, izdvajanja važnih

varijabli... Stoga se javlja problem višedimenzionalnosti i vizualizacije u 2D prostoru koji je ovdje riješen korištenjem paralelnog koordinatnog sustava umjesto ortogonalnog. Istovremeno je održan tečaj pod nazivom *'Introduction to spatial and spacetime objects in R'* na kojem su voditelji tečaja Milan Kilibarda i Tomislav Hengl sudionike upoznali s klasama objekata i paketima u R-u kao npr.: *'sp'* paket kod prostornih klasa, *'raster'* za raster klase, *'spacetime'* za prostorno-vremenske klase te *'plotKML'* paket kao spona između R-a i *Google Eartha* koji omogućava zapis objekata u KML formatu za daljnju vizualizaciju.

Idući se dan nastavilo s paralelnim radionicama. Profesor Kanevski održao je praktičnu radionicu na temu strojnog učenja uz korištenje nekih od njegovih software-a kao što su: GeokNN, GeoMLP, GeoGRNN, GeoSVM, GeoSOM koje su polaznici dobili na korištenje. Paralelni tečaj pod nazivom *'Automated mapping in 2D, 3D and 2D+T using machine learning in R'* vodio je Tomislav Hengl. Korišten je paket *'GSIF'* (Global Soil Information Facilities) koji sadrži alate i setove podataka za globalno kartiranje tla. Odabirom različitih metoda (GLM, randomForest, quantregForest, rpart, lme) prilagođen je regresijski model kako bi se procijenila najbolja metoda za taj set podataka, dok je za vizualizaciju konačnih polja korišten ranije spomenut paket *'plotKML'*. Drugi dio tečaja, pod nazivom *'Spatial and spatio-temporal prediction and visualisation of climate elements in R'* vodio je Milan Kilibarda. Za prostorno-vremensku interpolaciju meteoroloških podataka korišten je paket *'meteo'* te je regresija na mjerenjima rađena koristeći *modis* snimke kao jedan od prediktora. Kao još jedan od paketa za vizualizaciju spomenut je *'plotGoogleMaps'*.

Posljednja dva dana održana je konferencija u amfiteatru Milutina Milankovića koji je na tom fakultetu radio kao profesor. Prvo pozvano predavanje održao je Mikhail Kanevski pod nazivom *'Advanced Analysis of Environmental Data Using Machine Learning'*, u kojem se nadovezao i proširio tematiku s ranijih radionica. Pokazano je nekoliko primjera primjene metoda strojnog učenja na različite skupove podataka. Nakon toga je uslijedilo predavanje Tomislava Hengla pod nazivom *'Automated Global Soil Mapping: Discovering spatial soil patterns using machine learning'* u kojem je pokazao na koji je način strojno učenje korišteno u izradi globalne karte tla i koji su izazovi i rješenja po pitanju odabira prikladne metode. Isti dan u popodnevnoj sekciji, Wolfgang Wagner predstavio je rad *'Time series analysis of Sentinel-1 backscatter data on a high performance computing platform'*, koji također ističe prednosti primjene strojnog učenja, pogotovo kada se radi s ogromnom količinom podataka. Idućeg dana, profesor Mirko Orlić održao je dva predavanja. Prvo predavanje, pod nazivom *'Stjepan Mohorovičić, an early advocate of Milutin Milanković's theory of climate change'* bilo je prikladno s obzirom na lokaciju gdje je održana konferencija. Na temelju nekoliko radova S. Mohorovičića, pokazano je kako je on bio među prvima koji je prepoznao važnost Milankovićevih ciklusa u vrijeme kada je ta teorija

često bila kritizirana. U drugom predavanju *'Exploring relationship between global temperatures and global sea levels'* predstavio je neke od metoda koje se koriste za analizu i projekciju razine mora. Posljednji pozvani predavač bio je Ole Einar Tveito koji je u radu *'Observation based gridded climate data in Norway'* predstavio izazove i probleme u izradi klimatskih podataka na pravilnoj mreži s obzirom na kompleksnost terena u Norveškoj.

U ostalim predavanjima teme su bile vezane uz modeliranje različitih parametara korištenjem nekih od metoda strojnog učenja, satelitskih produkata, daljinskih mjerenja, meteoroloških podataka na pravilnoj mreži, itd. Prvog dana konferencije te su analize većinom bile vezane uz vegetaciju kao npr. modeliranje fenoloških faza, koristeći satelitske produkte i meteorološke podatke na pravilnoj mreži, modeliranje anomalija u fenologiji europskih šuma koristeći *'Random Forest'* metodu, modeliranje potražnje vode koristeći regresijska stabla, pa sve do onih koji su se fokusirali na kraću vremensku skalu te su prilagodili neuronske mreže za radarski *'nowcasting'* i predikciju kroz iduća 2 sata. Teme drugog dana konferencije su bile većinom klimatološke: interpolacija mjesečnih temperatura i oborine, kombiniranje mjerenja i podataka modela s ciljem dobivanja polja veće rezolucije, prostorna interpolacija dnevne visine snijega, ali i *'case-study'* u kojem je istražena korelacija lokalnih mjerenja koncentracije peluda i vremenskih uvjeta za individualni peludni alarm (www.pollendiary.com).

Od drugih hrvatskih predstavnika, Melita Perčec Tadić je održala prezentaciju pod nazivom *'Gridded monthly temperature fields for Croatia for the 1981.-2010. period'* gdje je predstavila karte polja mjesečnih temperatura zraka na rezoluciji 1 km. Takvi setovi podataka, osim što daju uvid u klimatske karakteristike šireg područja, su između ostalog vrlo korisni za validaciju rezultata klimatoloških modela. Irena Nimac je održala prezentaciju pod nazivom *'New 1981.-2010. climatological normals for Croatia and comparison to previous 1961.-1990. and 1971.-2000. normals'*. Analiza je napravljena za klimatološke normale temperature i količine oborine, kao i za pojedine temperaturne i oborinske indekse (broj dana iznad ili ispod određenog praga). U radu su za svako od razdoblja određene i Koeppen-Geigerove klimatske klase te su na pojedinim postajama uočene promjene u klasi, većinom zbog promjena u oborinskom režimu.

Ova radionica i konferencija bila je izvrsna prilika za upoznavanje s metodama strojnog učenja i njihovom primjenom u meteorologiji. Iz predstavljenih radova na konferenciji vidljivo je kako je strojno učenje jedna od vrlo aktualnih metoda u modeliranju različitih parametara budući da se puno bolje nosi s problemom nelinearnosti i višedimenzionalnosti, uz preduvjet visoke kvalitete ulaznih podataka. Cijeli program radionica i konferencije dostupan je na <http://geomla.grf.bg.ac.rs/>, dok se prošireni sažetci te pojedini radovi mogu naći na adresi:

http://geomla.grf.bg.ac.rs/site_media/static/Proceedings%20of%20GeoMla%202016%20Conference.pdf ■