



I n s t i t u t R u đ e r B o š k o v i ć

Adresa: Bijenička cesta 54, 10000 Zagreb | Tel: +385 (0)1 4561 111 | Fax: +385 (0)1 4680 084 | www.irb.hr

PRIOPĆENJE ZA JAVNOST

Petra Buljević Zdjelarević, M.A., Head, PR Office

Ruđer Bošković Institute / Bijenička cesta 54 / 10 000 Zagreb / Croatia

T. + 385 1 457 1269 / M. +385 99 267 9514 / @ info@irb.hr

irb.hr / [FB](#) | [TW](#) / [vimeo](#) / [linkedin](#)

'Ruđerovci' razvili novu metodu grupiranja podataka

Novu metodu odlikuje poboljšana točnost u odnosu na postojeće metode te predstavlja fundamentalni doprinos problemu grupiranja podataka.

ZAGREB, 11.01.2019. - Znanstvenici Laboratorija za reprezentacije znanja i strojno učenje Zavoda za elektroniku, Maria Brbić i dr. sc. Ivica Kopriva razvili su novu metodu za grupiranje podataka temeljenih na modelu linearnih potprostora kao generatora odgovarajućih funkcionalnih skupina. Da je riječ o značajnim rezultatima potvrđuje i objava rada u jednom od najutjecajnijih znanstvenih časopisa u području računalnih znanosti i umjetne inteligencije 'IEEE Transactions on Cybernetics' koji se s obzirom na visoki faktor odjeka (8.803) svrstava na treće mjesto u ovom području.

Strojno učenje je grana umjetne inteligencije koja se bavi oblikovanjem algoritama za automatsku obradu podataka. Drugim riječima, to je proces otkrivanja znanja iz velike količine podataka pri čemu računarski sustavi sami automatski poboljšavaju svoje procese kroz iskustvo. Strojno učenje jest temelj današnje podatkovne znanosti, a dijeli se na nadzirano, nenadzirano i polunadzirano.

Jedan od temeljnih problema u računalnim znanostima unutar područja poznatog kao nenadzirano učenje jest grupiranje podataka. Naime, za razliku od nadziranog strojnog učenja gdje su podaci dani u obliku ulazne i ciljne vrijednosti, a stroj uči funkciju koja ulaznim podacima pridružuje klasnu ili realnu vrijednost ciljanih podataka, u procesu nenadziranog strojnog učenja od algoritama se očekuje da podatke grupiraju u funkcionalne skupine koristeći isključivo informacije naučene iz samih podataka (engl. data driven methods). Pojednostavljeno, kod nenadziranog učenja nema učitelja, nema onoga koji pokazuje točno znanje, dani su podaci bez ciljne vrijednosti. U ovom slučaju cilj je grupirati primjere i otkriti neku strukturnu pravilnost u podacima, odnosno otkriti što podaci znače.

Strojno učenje jedno je od danas najuzbudljivijih područja računarske znanosti zbog brojnih mogućnosti primjena od raspoznavanja uzoraka i dubinske analize podataka do robotike, računalnog vida, bioinformatike i računalne lingvistike sve do medicine.

Neke od najčešćih primjena grupiranja podataka u medicini odnose se na segmentaciju slike, primjerice kod CT slike (skupine su organi), PET slike (skupine su tkiva), mikroskopske slike histopatoloških preparata (skupine su tkiva i/ili stanice), slike optičke koherentne tomografije oka (npr. skupine su slojevi unutar mrežnice).

"Primjene koje smo ilustrirali u ovom novom radu odnose se na prepoznavanje lica odnosno grupiranje slika lica u skupine koje odgovaraju osobama, zatim prepoznavanje govornika, odnosno



I n s t i t u t R u đ e r B o š k o v i ć

Adresa: Bijenička cesta 54, 10000 Zagreb | Tel: +385 (0)1 4561 111 | Fax: +385 (0)1 4680 084 | www.irb.hr

grupiranje značajki govora u skupine koje odgovaraju osobama, te prepoznavanje rukom pisanih brojeva, odnosno grupiranje slika u skupine koje odgovaraju znamenkama od 0 do 9." –objašnjava dr. sc. **Ivica Kopriva**.

Razvijene metode grupiranja podataka kod navedenih primjera temelje se na modelu prema kojem su podaci unutar svake skupine generirani iz pripadajućeg linearnog potprostora.

Temeljem tog modela razvijeni su algoritmi koji daju vrlo kompetitivne rezultate na grupiranju zahtjevnih skupova podataka.

"Ključan element u ovom pristupu je učenje matrice reprezentacije koja je rijetka i ima nizak rang. Umjesto konveksnih mjera ranga i rijetkosti koje se standardno koriste, u radu su predložene mjere koje bolje procjenjuju rang i rijetkost: (i) egzaktne mjere temeljene na L0 i Schatten-0 kvazi normama, te (ii) glatka surogat funkcija L0 i Schatten-0 kvazi normi." – objašnjava **Maria Brbić**.

Uvođenjem ovih mjera odgovarajući optimizacijski problemi su postali nekonveksni, što dokaz konvergencije algoritma čini zahtjevnim. Unatoč ovim poteškoćama, u radu je dan teorijski dokaz globalne konvergencije navedenog optimizacijskog problema za učenje matrice reprezentacije.

"Novi algoritmi su značajno poboljšali točnost u usporedbi s postojećim metodama na svim testiranim primjenama." – zaključuje Brbić.

Razvijena metoda je rezultat istraživanja doktorandice Marije Brbić u okviru istraživačkog projekta Hrvatske zaklade za znanost voditelja dr.sc. Ivica Koprive.

KORISNE POVEZNICE:

RAD: <https://ieeexplore.ieee.org/document/8573150>

O LABOSU: <https://www.irb.hr/Istrazivanja/Zavodi/Zavod-za-elektroniku/Laboratorij-za-strojno-ucenje-i-reprezentacije-znanja>