

Ekspertski sistem za preporuku algoritama mašinskog učenja

Stefan Kandić, Damjan Manojlović

1 Motivacija

Početak druge decenije 21. veka došlo je do naglog rasta popularnosti tehnika mašinskog učenja, koje su ranije bile prevashodno poznate samo u akademskim krugovima. Danas, sistemi zasnovani na mašinskom učenju se koriste u praktično svim granama industrije, od preporuke sadržaja na internetu i vožnje autonomnih automobila do određivanja zemlje pogodne za poljoprivredne delatnosti na osnovu satelitskih snimaka.

Intenzivna istraživanja su dovela do stvaranja velikog broja grana u samom polju, što je dodatno otežalo pristup ljudima koji ranije nisu bili upoznati sa mašinskim učenjem. Naš cilj je da napravimo alat pomoću kog će programeri koji nisu eksperti u ovom polju moći lako da se usmere ka rešenju problema koji rešavaju.

2 Pregled problema

Ideja je da korisnik sistema ima problem koji želi da reši, kao i podatke potrebne za treniranje algoritma. Uloga sistema je da mu da način na koji to može da izvede (algoritam ili familiju algoritama) ili da mu da koristan feedback ukoliko ne može da mu preporuči algoritam (npr. potrebno je pribaviti još podataka).

Što se tiče postojećih rešenja, nismo uspeli da ih nađemo. Postoje određene neformalne heuristike za rešavanje nekih tipskih problema, ali ništa što obuhvata više različitih grana mašinskog učenja.

3 Metodologija rada

3.1 Ulazi u sistem

Ulaz u sistem će biti opis problema koji se rešava. Neke od karakteristika problema bitnih za izbor algoritma su:

- šta korisnik tačno želi od algoritma (da predvidi kategoriju ili vrednost podataka, njihovu strukturu, generiše nove podatke ili možda da trenira agenta u specifičnom okruženju)
- struktura podataka nad kojim će se algoritam trenirati (slika, tekst, tabela itd) i informacija da li su podaci označeni
- broj uzoraka u skupu podataka
- specifične stvari koje zavise od prethodnih tačaka

3.2 Izlazi iz sistema

Izlaz iz sistema će predstavljati naziv algoritma koji bi se najbolje pokazao na datom problemu i eventualno dodatna literatura koja može pomoći korisniku da se bolje upozna sa njegovim radom.

3.3 Baza znanja

Bazu znanja će činiti pristupi koji su predstavljaju state-of-the-art za uobičajene probleme mašinskog učenja (klasifikacija slika - *CNN*¹, tekst - *RNN*² itd). Njih ćemo identifikovati analizom naučnih radova i druge relevantne literature (časopisi, blogovi vezani za trendove u oblasti mašinskog učenja).

4 Primer rezonovanja

1. Korisnik unosi da radi sa radi sa tekstualnim formatom i da ima 1 500 ulaznih podataka
2. Na osnovu inicijalnih parametara određuje se sledeći nivo, gde korisnik za konkretan primer bira da želi da vrši predikciju kategorija, ali pritom ne poseduje labelirane podatke
3. Nakon što je definisano da korisnik želi da prediktuje kategorije, radi sa tekstualnim podacima i koji nisu labelirani, korisniku se nudi opcija da unese broj kategorija koje dokumenti poseduju. Korisnik unosi da je broj kategorija 12
4. Potom, ekspetski sistem nudi preporučeni metod za rešavanje konkretnog problema na osnovu svih unetih parametara, u ovom konkretnom slučaju to bi bio K-Means algoritam

¹**CNN** - Convolutional Neural Network

²**RNN** - Recurrent Neural Network