

HYPERPARAMETER OPTIMIZATION

VASILije Todorović

GRIDSEARCHCV

- Zadavanje prostora pretrage parametara pomoću mape oblika {’param1’ : [a1,a2,...],’param2’:[b1,b2,..],...}
- Pretraga čitavog prostora pretrage
- Unakrsna validacija
- Pronalazak absolutno najboljih parametara u odnosu na posmatrani kriterijum ocene
- Potencijalna neefikasnost zbog pretrage čitavog prostora pretrage
- Mogućnost paralelnog izvršavanja

METAHEURISTIKE

S-METAHEURISTIKE

- Single-solution Metaheuristics
- Fokus na pojedinačnim rešenjima
- Iterativni proces menjanja trenutnog rešenja
- Lokalna pretraga, simulirano kaljenje...

P-METAHEURISTIKE

- Population-solution Metaheuristics
- Fokus na populaciji rešenja
- Evolucija
- Genetski operatori - selekcija, ukrštanje, mutacija
- Genetski algoritam, genetsko programiranje...

SIMULIRANO KALJENJE

- S-metaheuristika
- SimulatedAnnealingCV
- Bira se nasumična kombinacija parametara iz prostora pretrage kao početno rešenje, koja predstavlja trenutno najbolje rešenje
- Odabrani model se trenira sa tim parametrima, i ocenjuje se u odnosu na zadati kriterijum ocene (prisutna unakrsna validacija)
- Iterativni proces prelaska iz trenutnog rešenja u naredno:
 - Ukoliko je novo rešenje bolje od trenutnog?
 - Ukoliko je novo rešenje lošije od trenutnog?
 - Potencijalno prihvatanje lošijih rešenja kako bi se došlo do globalnog optimuma
- Na kraju iterativnog procesa za dobijeno rešenje se ne može garantovati da je najbolje u čitavnom prostoru pretrage, ukoliko nisu razmatrane sve moguće kombinacije parametara

GENETSKI ALGORITAM

- P-metaheuristika
- GeneticAlgorithmCV (prilikom evaluacije modela koristi se unakrsna valuacija)
- Početna populacija (1. generacija) čini niz različitih kombinacija parametara odabralih na nasumičan način
- Selekcija - turnirska selekcija
- Ukrštanje - za svaki parametar novog rešenja jednaka verovatnoća da uzme vrednost tog parametra od prvog ili od drugog roditelja
- Mutacija - za svaki parametar postoji verovatnoća da dođe do mutacije, u tom slučaju se bira nasumična vrednost iz liste vrednosti dodeljene tom parametru unutar mape parametara
- Elitizam
- Najbolja rešenje u poslednjoj generaciji ne mora nužno biti najbolje u čitavom prostoru pretrage

TESTIRANJE PERFORMANSI

01

```
'max_depth' : [4,5,6],  
'criterion' : ['gini','entropy'],  
'min_samples_leaf':[2,3,4],  
'min_samples_split':[3,4,5]  
veličina prostora pretrage 54
```

4 prostora pretrage

Model: drvo odlučivanja

Veličina podataka:

- 100 redova
- 1000 redova
- 10000 redova
- 98000 redova

02

```
'max_depth' : [4,5,6],  
'criterion' : ['gini','entropy'],  
'min_samples_leaf':[2,3,4,5,6],  
'min_samples_split':[3,4,5,6]  
veličina prostora pretrage 120
```

Kriterijum ocene: tačnost

03

```
'max_depth' : [4,5,6,7,8],  
'criterion' : ['gini','entropy'],  
'min_samples_leaf':[2,3,4,5,6],  
'min_samples_split':[3,4,5,6]  
veličina prostora pretrage 200
```

04

```
'max_depth' : [4,5,6,7,8,9,10],  
'criterion' : ['gini','entropy'],  
'min_samples_leaf':[2,3,4,5,6,7],  
'min_samples_split':[3,4,5,6,7]  
veličina prostora pretrage 420
```

6

EFIKASNOST IZVRSAVANJA

SimulatedAnnealingCV	10^2 redova	10^3 redova	10^4 redova	$9.8 * 10^4$ redova
1. prostor pretrage	0.68s	1.31s	8.58s	58s
2. prostor pretrage	1.13s	2.73s	13s	2min 11s
3. prostor pretrage	1.81s	3.6s	34.25s	3min 53s
4. prostor pretrage	4s	9.83s	1min 13s	9min 9s

Table 1: Vreme izvršavanja simuliranog kaljenja za različite prostore pretrage.

GeneticAlgorithmCV	10^2 redova	10^3 redova	10^4 redova	$9.8 * 10^4$ redova
1. prostor pretrage	0.45s	0.75s	6.12s	43.06s
2. prostor pretrage	1.19s	2.42s	11.7s	1min 58s
3. prostor pretrage	2.18s	6.24s	25.4s	4min 57s
4. prostor pretrage	4.03s	8.26s	1min 7s	12min 10s

Table 2: Vreme izvršavanja genetskog algoritma za različite prostore pretrage.

GridSearchCV	10^2 redova	10^3 redova	10^4 redova	$9.8 * 10^4$ redova
1. prostor pretrage	0.55s	1.14s	9.6s	1min 41s
2. prostor pretrage	1.14s	2.53s	17.04s	3min 41s
3. prostor pretrage	1.84s	4.57s	29.39s	4min 57s
4. prostor pretrage	4.63s	9.38s	1min 14s	18min 19s

Table 4: Vreme izvršavanja GridSearchCV sa paralelizacijom za različite prostore pretrage.

KVALITET DOBIJENIH
HIPERAPRAMETARA