



Evolucija drveta klasifikacije upotrebom genetičkog programiranja

Petar Zečević 169/2016 i Aleksandra Ružić 47/2016



BGP

- Building block approach to Genetic Programming
- Jedinke su stabla
- Inicijalna populacija se formira od nasumičnih stabala dubine 1
- Kroz generacije se povećava dubina jedinki



BGP- operacije

- **Mutacija relacije:** nasumičnom čvoru menja relaciju
- **Mutacija vrednosti:** nasumičnom čvoru menja vrednost sa desne strane
- **Potkresivanje:** nasumično podstablo se menja listom sa nasumičnom klasom
- **Dodavanje gradivnih blokova**



Grativni blokovi

- Nasumični list se menja nasumičnim podstablom dubine 1

- Uslov:

$$(\text{avg_depth} + \text{avg_width}) - (\text{prev_avg_depth} + \text{prev_avg_width}) < L$$



Ostale osobine BGP

- Ukrštanje: za obe jedinke bira se nasumično unutrašnji čvor, i podstabla ispod datih čvorova se zamenjuju
- Selekcija: turnirska
- Kriterijum zaustavljanja:

$\text{min_rule_acc} < e^{(c * \text{trainsize} * A)}$, gde su

- $A = 1/T(0) - 1/T(t)$, gde je $T(t) = T(0) - t$ "temperaturna" funkcija broja iteracija (sa povećanjem
- broja iteracija, funkcija "hladi" program, smanjujući gornji izraz, a time povećavajući mu šansu da se
- zaustavi). $T(0)$ je korisnički zadat parametar. Određuje se eksperimentalno.
- trainsize - veličina trening skupa
- min_rule_acc - najmanja preciznost među pravilima
- c - korisnički zadat parametar



Implementacija

- Node: klasa koja predstavlja čvor. Sadrži indeks čvora
- Leaf: potklasa klase Node, koja predstavlja list stabla. Sadrži ime klase
- NotLeaf: potklasa klase Node, koja predstavlja list stabla. Sadrži levog i desnog potomka, atribut, relaciju i vrednost.



Implementacija

- Tree: klasa koja predstavlja drvo odlučivanja. Sadrži koreni čvor i skup svih iskorišćenih indeksa.
- Jedinke su iz klase Tree



Implementacija- algoritam

- Pravljenje inicijalne populacije
- Dok ne prođe n generacija:
 - Vrši se selekcija
 - Ukrštanje
 - Primjenjuju se operatori sa zadatim verovatnoćama
 - Stara generacija se u potpunosti menja novom
- Rešenje je najbolja jedinka iz svih generacija



Verzije algoritma

- Verzija 1: nema kriterijum za dodavanje gradivnih blokova (sem verovatnoće dodavanja)
- Verzija 2: kriterijum za dodavanje gradivnih blokova je uzet iz BGP
- Verzija 3: kriterijum za dodavanje gradivnih blokova je:

$\text{avg_depth} \geq \text{tree_depth} \text{ AND } \text{avg_width} \geq \text{tree_width}$



Parametri

- veličina test skupa = 30%
- broj iteracija = 30
- brojnost populacije = 1000
- veličina turnira za selekciju = 20
- verovatnoća za ukrštanje = 0.9
- verovatnoća za mutaciju relacije = 0.6
- verovatnoća za mutaciju desne strane uslova = 0.7
- verovatnoća potkresivanja = 0.2
- verovatnoća dodavanja gradivnog bloka = 0.3



Skupovi podataka

- Iris:
 - 150 redova i 5 kolona
 - 4 kontinualna atributa
 - 3 podjednako zastupljene klase
- Ionosphere:
 - 351 red i 35 kolona
 - 33 kontinualna atributa
 - 2 nejednako zastupljene klase
- Pima Indians diabetes:
 - 768 redova i 9 kolona
 - 8 kontinualnih atributa
 - 2 nejednako zastupljene klase

Podaci	Rezultati	BGP	Verzija2	Verzija3	DecisionTree (sklearn)
Iris	Trening preciznost	-	0,92933333	0,881333333	0,977333333
	Test preciznost	0,941	0,92666667	0,865333333	0,941333333
	Prosečan broj pravila	3.73	3,66666667	3	4,666666667
	Prosečan broj konjukcija u pravilu	2.02	2,09090909	1,666666667	2,5
Ionosphere	Trening preciznost	-	0,864113475	0,85787234	0,946729375
	Test preciznost	0.892	0,786206897	0,863793103	0,900858091
	Prosečan broj pravila	4.70	5,4	2,933333333	6,533333333
	Prosečan broj konjukcija u pravilu	2.39	2,975308642	1,636363636	2,806122449
Pima	Trening preciznost	-	0,754085603	0,763294423	0,83621781
	Test preciznost	0.725	0,74015748	0,722834646	0,660158493
	Prosečan broj pravila	3.70	5	2,866666667	25
	Prosečan broj konjukcija u pravilu	1.97	3,186666667	1,604651163	4.821333333



**Zahvaljujemo se kolegama Vladimiru
Batoćaninu i Anđelki Milovanović na
pozajmljenim hardverskim resursima!**

Hvala na pažnji! :)