

# Metodologija izbora pogodnih vrednosti SAT rešavača

Mladen Nikolić

4. Jun 2008.

## 1 Motivacija

- Efikasnost SAT rešavača u velikoj meri zavisi od iskazne formule koja se rešava i od vrednosti parametara rešavača.
- Rešavači ne analiziraju svojstva formule koju rešavaju kako bi bile primenjene pogodne vrednosti parametara.
- Formule se mogu predstaviti kao grafovi.
- Postoji zajednička i vizuelno prepoznatljiva grafovska struktura za formule koje pripadaju istoj familiji.
- Da li formule koje pripadaju istoj familiji bivaju efikasno rešene za iste vrednosti parametara?
- Kako prepoznati formule koje se efikasno rešavaju za iste vrednosti parametara?

## 2 Metodologija ugrubo

- Sistematično rešavanje nekog reprezentativnog korpusa iskaznih formula za različite vrednosti relevantnih parametara u fazi treniranja rešavača u cilju određivanja dobrih vrednosti parametara za pojedinačne formule.
- Inteligentno biranje pogodnih vrednosti parametara u fazi eksploatacije rešavača, u skladu sa određenim karakteristikama formule koja se rešava.

- Najvažniji aspekti ovakve metodologije su:
  - Izbor parametara čiji bi se uticaj na rešavanje formule posmatrao.
  - Izbor dopustivih vrednosti za parametare koji se smatraju relevantnim.
  - Izbor korpusa formula na kojem bi se vršili treniranje i evaluacija SAT rešavača koji bi implementirao predloženu metodologiju.
  - Identifikacija karakteristika iskazne formule na osnovu kojih bi se vršio izbor pogodnih vrednosti parametara.
  - Izbor mera kvaliteta na osnovu kojih bi se vršila evaluacija metodologije.

### 3 Osnovni pojmovi i korišćene tehnike

- SAT problem, rešavači, DPLL.
- Politike izbora promenljive:
  - Politika slučajnog izbora i kombinacije.
  - Minisat politika.
  - Minisat politika sa "pametnom inicijalizacijom".
- Politike izbora polariteta promenljive:
  - Uvek `true`.
  - Uvek `false`.
  - Politike slučajnog izbora sa verovatnoćom  $p$ .
  - Polarity saving.
  - Polarity saving sa "pametnom inicijalizacijom".
- Politike restartovanja:
  - Bez restartovanja.
  - Minisat politika.
  - Lubijeva politika.
  - Arminova politika brzog restartovanja.
- Problem klasifikacije.
- Sličnost grafova.

- Za graf  $G' = (V', E')$  kažemo da je indukovani podgraf grafa  $G = (V, E)$  ako i samo ako važi:
  - $V' \subseteq V$
  - ako važi  $u, v \in V'$  i  $(u, v) \in E$ , onda važi i  $(u, v) \in E'$ .
- $n$ -grami i  $n$ -gramski profili.
- Metoda k najbližih suseda.

## 4 Klasifikovanje formula zasnovano $n$ -gramima

- Formule se mogu posmatrati kao tekst i na njega se može primeniti klasifikacija zasnovana na  $n$ -gramima ako se definiše distanca nad  $n$ -gramskim profilima.

$$d_1(\mathcal{P}_1, \mathcal{P}_2) = \sum_{n \in \text{profili}} \left( \frac{2 \cdot (f_1(n) - f_2(n))}{f_1(n) + f_2(n)} \right)^2 \quad (1)$$

$$d_2(\mathcal{P}_1, \mathcal{P}_2) = \sum_{n \in \text{profili}} \frac{2|f_1(n) - f_2(n)|}{f_1(n) + f_2(n)} \quad (2)$$

$$d_3(\mathcal{P}_1, \mathcal{P}_2) = \sum_{n \in \text{profili}} \left( \frac{f_1(n) - f_2(n)}{\sqrt{f_1(n)f_2(n)} + 1} \right)^2 \quad (3)$$

- Ovaj pristup nije invarijantan u odnosu na preimenovanje promenljivih i promenu rasporeda elemenata formule.

## 5 Klasifikovanje grafova pomoću sličnosti čvorova

- Koristi se metoda rasparenih sličnosti:
  - Čvorovi dva grafa su onoliko slični koliko su slični njihovi susedi.
  - Da bi dva čvora bila slična dovoljno je da svakom susedu jednog čvora odgovara jedan sličan sused drugog čvora.
  - Metoda kao izlaz daje matricu čiji elementi predstavljaju sličnosti čvorova datih grafova.
- Za računanje sličnosti celih grafova korišćemo meru:

$$s(A, B) = \frac{A \cdot B}{\sqrt{A \cdot A} \sqrt{B \cdot B}}$$

- Ovaj pristup je vremenski previše zahtevan u vreme upotrebe.

## 6 Klasifikovanje grafova na osnovu frekvencija podgrafova

- Formule se mogu predstaviti u obliku grafova. Grafovska reprezentacija nije osetljiva na preimenovanje promenljivih i promenu redosleda elemenata formule.
- Grafovi se u cilju klasifikacije predstavljaju nizovima frekvencija njihovih podgrafova koje ćemo nazivati njihovim profilima.
- Frekvencije podgrafova se računaju sempliranjem.
- Nad ovim profilima se koriste iste distance kao i nad  $n$ -gramskim profilima.

## 7 Izbor vrednosti parametara

- Trening sistema se sastoji od:
  - sistematičnom rešavanju reprezentativnog korpusa formula (SAT2002) za različite vrednosti izabranih parametara  $i$
  - određivanja profila grafova koji odgovaraju formulama iz korpusa.
- Prilikom upotrebe sistema:
  - određuje se profil formule koju je potrebno rešiti,
  - pronalazi se  $n$  formula koje su joj najbliže i nalazi se kombinacija vrednosti parametara koja se najčešće javlja među  $m$  najbolje rangiranih kombinacija za svaku od tih formula.

## 8 Eksperimentalni rezultati

- $m = 39, n = 46$

	Idealni	Najbolji fiksirani	Naučeni
Broj rešenih formula	1178	1069	1122
Središnje vreme	0.805	4.19	2.425
Prosečno vreme (presek)	17.28	44.16	32.14
Prosečno vreme	35.80	47.12	44.14
Ukupno vreme	42170.70	115776.40	83126.05
Broj formula za koje su najbolji	1178	162	260

Table 1: Eksperimentalni rezultati