Evolúciós algoritmusok 2. előadás

Sáfár Orsolya

Szkémák, Gray-kódolás, Építőkövek hipotézise

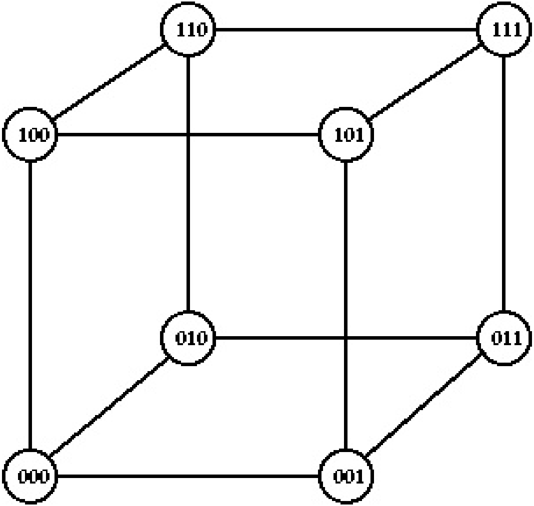
A szkéma fogalma

A szkéma a keresési tér egy hipersíkja, amely egy harmadik szimbólum (#, jelentése: tetszőleges bit) segítségével reprezentálható.

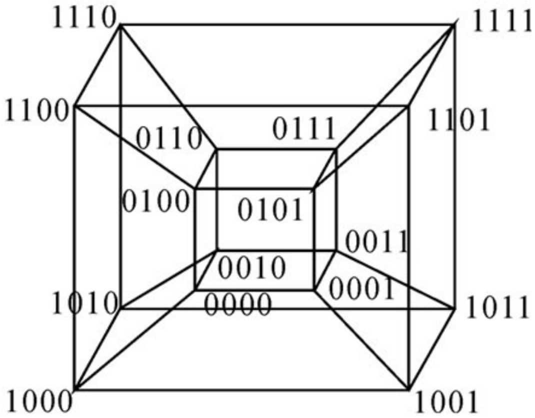
Például ha a feladatban a keresési tér 5 hosszú bitsorozatból áll, akkor a $H = \#1\#\#0$ szkéma 8 egyedre illeszkedik.

A szkémák elképzelhetőek mint egy hiperkocka csúcsai/élei/lapjai/hipersíkjai.

3 dimenziós kocka



4 dimenziós hiperkocka



Belső párhuzamosság

Egy n hosszú 0-1 sorozattal reprezentált megoldástérben összesen 3^n darab szkéma van, egy egyedre pedig 2^n szkéma illeszkedik.

Egy μ elemszámú populációban legfeljebb $\mu 2^n$ darab szkéma lehet jelen. Bár általában nincs ennyi különböző szkéma egy populáción belül egyszerre, Holland 1975-ben megmutatta, hogy a genetikus algoritmus egy generációban $O(\mu^3)$ szkémát értékel ki.

Ezt a tulajdonságot nevezzük belső párhuzamosságnak, amely nagyban hozzájárul a genetikus algoritmusok sikerességéhez.

A szkémák tulajdonságai

- A szkéma fitneszének az összes ráilleszkedő egyed fitneszének átlagát nevezzük, jele $< H >$
- A szkéma rendje a rögzített bitek száma, jele o(H)
- definiáló hossza pedig a két szélső nem # pozíció távolsága, jele d(H)

Legyen például $H = \#1\#\#0$, ekkor a rendje 2, a definiáló hossza 3.

A szkéma tétel levezetése 1

Induljunk ki egy fitnesz arányos kiválasztást, egypontos keresztezést és bitenkénti mutációt használó genetikus algoritmusból. Legyen a reprezentáció hossza l. Ekkor annak valószínűsége, hogy a H szkémát az egypontos keresztezés széttöri:

$$P_d(H, 1X) = \frac{d(H)}{l - 1}$$

Annak valószínűsége, hogy a bitenkénti mutáció a szkémát megsemmisíti:

$$P_d(H, MUT) = 1 - (1 - P_m)^{o(H)} \simeq o(H) * P_m$$

A szkéma tétel levezetése 2

Annak esélye, hogy egy szkémát kiválasztják a következő generációba a szkéma fitneszén múlik. Jelölje $< f >$ a teljes populáció átlagfitneszét. Ha a jelen populációban $n(H,t)$ darab van a H szkémából és μ elemű a populáció, akkor annak esélye, hogy a H szkémát egyszer szülőnek választjuk:

$$P(H,t) = \frac{n(H,t) < H >}{\mu < f >}$$

A szkéma-tétel

Ezek alapján a következő generációban a H szkéma aránya legalább átlagosan

$$\frac{n(H,t+1)}{\mu} \geq \frac{n(H,t)}{\mu}\frac{< H >}{< f >}\left(1 - p_c\frac{d(H)}{l-1}\right)(1 - p_m o(H))$$

lesz. Ez magyarázza a konvergencia lassulását amit a megoldások közelednek az optimumhoz. Ezt részben a fitneszfüggvény skálázásával orvosolhatjuk, például lineáris rangsorolással.

Deceptív problémák

Egy probléma deceptív voltát, azaz azt a tulajdonságot, hogy a globális optimumot nem lehet összerakni alacsonyrendű szkémákból, szerencsés esetben orvosolhatjuk a fitneszfüggvény ügyes megváltoztatásával.

Példa: Tekintsük a hátizsák-pakolás problémát 10 tárgyra, legyenek a súlyok [1,2,. . . ,9, 10], az értékek [1, 2 , . . . , 9, 11]. Ekkor egy tíz kapacitású hátizsákba az optimális pakolás az, hogy csak az utolsó tárgyat veszem be.

Példa (folyt.)

Viszont ha csak néhány tárgyat raktam be, és az utolsót nem, akkor ezen pakolások úgy javíthatóak könnyen, hogy még egy tárgyat beveszek (ami nem az utolsó), feltéve persze, hogy még a kapacitás alatt maradok. Ez a probléma tehát deceptív jellegű, ugyanis a szkémák fitnesze úgy nő, hogy nem a globális optimum irányában változtatom a biteket.

Itt például segíthet az, ha a kapacitást túllépő pakolások nem automatikusan 0 értéket kapnak, csak valami büntetést, mondjuk egy egykettedes szorzót. Később még lesz szó a fitneszfüggvény helyes megválasztásáról olyan esetekben, ahol az összes szóba jövő sorozat nem mindegyike potenciális megoldás.

Gray-kódolás

Egy másik lehetséges forrása a probléma deceptív voltának az, hogy az egészek kódolásakor kettes számrendszerre térünk át. Amikor egy új helyiérték lép be, akkor nagyon megváltozik a potenciális megoldás szerkezete, például: 31=01111, de 32=10000. Ha a globális optimum egy új helyiértékhez közel van, akkor nehéz alacsonyabb rendű szkémákból kirakni.

Ezért ha a 0-1 bitsorozatunk egészeket reprezentál, akkor célszerű a kettes számrendszer helyett a Gray-kódolást választani. A Gray-kódolás egy olyan reprezentációja a pozitív egészeknek, amely garantálja, hogy ha két egész szám "egymáshoz közel van", akkor a Gray-kódjuk "közel van".

Gray-kódolás

Egy egész szám Gray-kódjának elkészítéséhez induljunk ki a kettes számrendszerbeli alakból, legyen ez $b_1 b_2 b_3 \ldots b_k$. A Gray-kód első bitje b_1, majd az i. bitet úgy kapjuk, hogy vesszük b_{i-1} és b_i XOR-ját (kizárólagos vagy).

Példa: $n = 11$, akkor az ő kettes számrendszerbeli alakja $11 = 8 + 2 + 1$ miatt 1011. Ekkor a Gray-kódja 1110.
Vegyük most $n = 12$-t, szintén átírva kettes számrendszerbe: $12 = 8 + 4$, így 1100. Az ő Gray-kódja: 1010.

Gray-kódolás II

Definíció

Legyen $b_1 b_2 b_3 \ldots b_k$ *és* $c_1 c_2 c_3 \ldots c_k$ *két azonos hosszúságú bitsorozat. Ekkor b és c Hamming-távolsága a különböző bitjeiknek a száma.*

Tétel

Ha két pozitív egész szám különbsége 1, akkor a Gray-kódjuk Hamming-távolsága is pontosan 1.

Bizonyítás

Ha két szám különbsége 1, akkor a nagyobbikat úgy kapjuk a kisebből, hogy hozzáadok egyet. Kettes számrendszerben: $b_1 b_2 \dots b_{k-1} b_k + 00 \dots 01$.

Az összeget úgy számíthatjuk ki, hogy amíg a b végén egyes van, addig az összegben 0 lesz, majd amikor a b-ben elérünk hátulról az első nullához (legyen az ő pozíciója j), akkor ott az összegben egyes lesz, majd onnantól kezdve a két szám kettes számrendszerbeli alakja megegyezik.

Bizonyítás II

Azaz amikor alkalmazzuk a Gray-kódot a két számra, akkor, mivel a két szám eleje megegyezik, a Gray-kódjuk eleje (az első $j-1$ pozíció) is megegyezik. A j. helyiértéken a Gray-kódjuk különbözik. De ezen helyiérték után mindkét Gray-kódban $10\ldots0$ van, mert egy 10 illetve 01 részlet után a j. és $j+1$. pozícióban, csupa egyforma bit XOR-ját veszem.

Könnyen meggondolható a két szélsőséges eset, azaz ha j az első vagy utolsó pozíció, ilyenkor az egyforma eleje vagy vége hiányzik a Gray-kódnak, de ugyanúgy pontosan 1 bit tér el.

A Gray-kód egyértelműsége

Ha két szám különböző, akkor a Gray-kódjuk is különbözik, hiszen azon pozícióban, ahol (elölről számolva) először különbözik a kettes számrendszerbeli alakjuk, a Gray-kódjuk is eltér.

Másrészt viszont a Gray-kód inverze is egyértelmű: a $g_1, g_2, g_3 \ldots, g_k$ Gray-kódot úgy kell visszafejteni, hogy $b_1 := g_1$ és $v := b_1$. Aztán elölről egyesével haladva:

for i=2:k
ha $g_i == 1$ akkor $v :=\sim v$
$b_i := v$.

Az építőkövek hipotézise

A szkéma tételből leolvasható, hogy ha két szkéma azonos fitnesszel rendelkezik, akkor a rövidebb definiáló hosszal rendelkező illetve alacsonyabb rendű szkéma jobb eséllyel marad meg a következő generációra.

Ebből az észrevételből származik az építőkövek hipotézise, amely szerint a genetikus algoritmusok úgy működnek, hogy kezdetben rövid szkémákat versenyeztet, majd ezekből építi fel a hosszabb, optimumot tartalmazó szkémákat.

Ellenpélda az építőkövek hipotézisére

Tekintsük azt az optimalizálási feladatot, melyben a 3 hosszú 0-1 sorozatokon keresem az fitneszfüggvény optimumát, melyre f(000)=5, f(111)=4, azon egyedek, melyeknek pontosan egy darab egyes van a genomjában a fitnesze 1, amelyeknek pontosan kettő, azok fitnesze 3.
Ekkor teljesülni fog, hogy bármely 0-t is tartalmazó szkémában az összes 0-t egyesre változtatva a szkéma fitnesze nőni fog.

Ellenpélda az építőkövek hipotézisére