Digitális élőlények komplexitásának mérése információelméleti eszközökkel

Nagy Attila
III. programtervező matematikus - V. filozófia
Email: nagyat@dragon.klte.hu

2001. április 24.

Kivonat

Erősen vitatott kérdés, vajon az evolúció folyamata szükségszerűen implikálja-e a komplexitás növekedését? A kérdés eldöntéséhez elengedhetetlen valamilyen megfelelő mérték használata. Nehéz lenne a földi élőlények szerkezeti és funkcionális bonyolultságának egzakt mérése, de a digitális megfelelőiknél ez sokkal egyszerűbb: önreplikáló gépi kódú programoknál kézenfekvő az információelméleti eszközök használata, hiszen gyakorlatilag szimbólumfüzérekről van szó. [1]

1 Digitális élőlények

A digitális élőlények önreplikálásra képes gépi kódú programok, melyek környezetükkel kölcsönhatásba lépve számítási feladatok megtanulásával növelni tudják életerejüket. Mivel testük gyakorlatilag megegyezik genetikai kódjukkal, ezért a biológiai megfelelőik a replikációra képes RNS-molekulák lehetnek.

1.1 Jelölések

Jelölje D a virtuális gép utasításkészletének a méretét, azaz a genetikai kód szimbólumainak a számát, l egy adott élőlény genetikai kódjának a hosszát az utasítások számában mérve.

2 A genetikai kód entrópiája

Shannon információelméletében [3] az entrópia mennyisége (H) meghatározza, hogy az adott objektum állapotának meghatározásához hány bitre van szükség. Ha az adott objektum (jelen esetben szimbólumsorozat) nagyobb rendezetlenséget mutat, akkor több bitre van szükség a leíráshoz, ekkor azt mondhatjuk, hogy a potenciális információtartalom nagy. Az entrópia mennyisége könnyen számolható, ha adott a valószínűségek p_i eloszlása, ahol $i = 1..N$.

$$H = \sum_{i}^{N} p_i log \frac{1}{p_i} \tag{1}$$

Az entrópia maximális, ha az eloszlás egyenletes, illetve 0, ha az egyik valószínűség egy, az összes többi pedig 0.

A genetikai kód egy szimbólumsorozat, de nem adott az egyes szimbólumok – az utasítások – gyakoriságának eloszlása. Ezeket ki kell számolnunk behelyettesítési módszerrel. [1] Módszeresen az abc összes elemét behelyettesítve egy adott i. utasításhelyre megvizsgáljuk, hogy a mutáció halálos (lethal), azaz a genotípus szaporodóképtelenségét vagy jelentős életerőcsökkenését eredményezi, semleges (neutral) vagy pozitív-e. Ezek után vesszük az összes nem-halálos (életképes – viable) mutációt s ezeket egyenlő valószínűségűeknek tekintjük, míg a halálos mutációknak 0 valószínűséget tulajdonítunk. Jelölje az életképes mutációk számát N_v. Ekkor az i. pozíció entrópiája a következőképpen számolható.

$$H_i = \sum_{1}^{N_v} \frac{1}{N_v} log_D(N_v) = log_D(N_v) \tag{2}$$

3 A genetikai kód komplexitása

Az entrópia segítségével a komplexitás a következőképpen határozható meg. Egy adott utasításhely információtartalma:

$$I_i = H_{max} - H_i \tag{3}$$

Ha egy helyen tetszőleges utasítás állhat, akkor az nem kódol információt sem a digitális élőlényről magáról (nem vesz részt az önreplikáló algoritmusban) sem a környezetről, azaz nem része egyetlen számítási feladatnak sem. Elképzelhető, hogy a nagy entrópiájú utasításpozíciók egyáltalán nem kerülnek végrehajtásra. Ellenben ha az

[1] Jelen tanulmány az Avida csoport munkájának [1] összefoglalója/pontosítása/továbbfejlesztése.

utasítás nem cserélhető fel az életerő szignifikáns csökkenése nélkül, akkor az entrópiája igen alacsony, s ebből következően a pozíció információtartalma nagy. Ha a komplexitást az utasításonkénti információtartalmak összegével jellemezzük, akkor természetes módon kerül be a genetikai kód hossza is a bonyolultsági mértékbe.

$$C = l - \sum_i H_i \tag{4}$$

A maximális entrópiák összege megegyezik a kód hosszával, lévén $H_{max} = 1$. A végrehajtatlan utasítások megjelenése a genetikai kódban növeli ugyan a komplexitást, de a nagy entrópia miatt, csak igen kis mértékben, ily módon egy kódrövidülés optimalizálás esetén is növekedhet a bonyolultsági foka az élőlénynek.

4 Episztatikus korrelációk

Nyilvánvaló, hogy az egyes helyeken az utasítások előfordulási valószínűsége nem független a többi utasításhely tartalmától. A genetikában ezt episztatikus kapcsolatoknak nevezik. Ezen összefüggések tovább kell finomítani a komplexitásmértéket...

5 Univerzális processzorok

Az univerzális processzorok alkalmazása a digitális evolúcióban megköveteli az alkalmazott mértékek átfogalmazását, hiszen a komplexitás mérését először a virtuális gép leírására, majd a processzor felépítése után magára a tényleges programra is alkalmazni kell. A komplexitás ilyen irányú kiterjesztése természetes módon tudja megjeleníteni a bonyolultság növekedésének mindkét módját. Az 1. típusú átmenetet, azaz az információ kódolási módjában bekövetkező változást a processzorleírás komplexitásváltozása jelenti, míg a 2. típusút, az egyszerű specializációt a programkód komplexitásváltozása jelzi. [2]

6 A komplexitás mértékének változása

mérési eredmények...

Hivatkozások

[1] Christoph Adami, Charles Ofria, and Travis C. Collier. Evolution of biological complexity. *PNAS*, 97(9):4463–4468, 2000.

[2] Christopher L. Nehaniv and John L. Rhodes. The evolution and understanding of hierarchical complexity in biology from an algebraic perspective. In *Artificial Life 6*, pages 45–67. Massachusetts Institute of Technology, 2000.

[3] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 1948.