

Problemă de codificare/decodificare

Un sistem de codificare transformă un mesaj format din litere într-un nou mesaj rezultat prin înlocuirea tuturor instanțelor unui caracter cu un alt caracter, astfel încât două caractere diferite sînt codificate diferit. Dacă este cunoscută o pereche ($mesaj_{codificat}$, $mesaj_{corect}$), unde fiecare mesaj are n caractere, determinați corespondențele utilizate pentru codificare.

Observații

În continuare presupunem că mesajele sînt formate doar cu litere mici ale alfabetului englez (fără spații sau alt tip de punctuație). Pentru rezolvare este folosită reprezentarea numerică, unde litera a este codificată cu 0 iar litera z este codificată cu 25. Un candidat la soluție este o permutare pe mulțimea de litere mici ale alfabetului englez reprezentate prin numere de la 0 la 25, care corespunde unei modalități de codificare. Spațiul soluțiilor este spațiul permutărilor pe $\{0,1,2, \dots, 25\}$ (orice permutare este fezabilă).

Evaluarea unui cromozom poate fi abordată în două variante:

1. Pentru evaluarea unui cromozom se decodifică mesajul codificat, aplicînd permutarea inversă celei reprezentate de cromozom, apoi se compară mesajul obținut cu cel corect. Gradul de similitudine este de tip cost și poate fi reprezentat prin

$$cost(mesaj_x, mesaj_{corect}) = \sum_{i=1}^n |mesaj_x(i) - mesaj_{corect}(i)|$$

unde $mesaj_x$ corespunde decodificării mesajului $mesaj_{codificat}$ prin intermediul inversei permutării x , astfel

$$mesaj_x(i) = k \text{ unde } x(k) = mesaj_{codificat}(i)$$

adică

$$x(mesaj_x(i)) = mesaj_{codificat}(i)$$

Evident, dacă $mesaj_{corect} = mesaj_x$, x este permutarea pe baza căreia a fost realizată **codificarea** lui $mesaj_{codificat}$, adică

$$mesaj_{codificat}(i) = x(mesaj_{corect}(i)), \quad i = 1, \dots, n$$

Deoarece funcția $cost$ este cu valori pozitive și are minimumul 0 pentru acea permutare care realizează codificarea corectă, rezultă că funcția de evaluare a cromozomilor poate fi definită prin

$$f_{obiectiv}(x) = \frac{1}{1 + cost(mesaj_x, mesaj_{corect})}$$

fiind o funcție care trebuie maximizată și are valoarea maximă 1.

2. Pentru evaluarea unui cromozom se codifică mesajul original, aplicînd permutarea reprezentată de cromozom, apoi se compară mesajul obținut cu cel codificat. Gradul de similitudine este de tip cost și poate fi reprezentată prin

$$cost(mesaj_x, mesaj_{codificat}) = \sum_{i=1}^n |mesaj_x(i) - mesaj_{codificat}(i)|$$

unde $mesaj_x$ corespunde codificării mesajului $mesaj_{original}$ prin intermediul permutării x , astfel

$$mesaj_x(i) = x(mesaj_{corect}(i))$$

Evident, dacă $mesaj_{codificat} = mesaj_x$, x este permutarea pe baza căreia a fost realizată codificarea lui $mesaj_{codificat}$, adică

$$mesaj_{codificat}(i) = x(mesaj_{corect}(i)), \quad i = 1, \dots, n$$

Deoarece funcția $cost$ este cu valori pozitive și are minimumul 0 pentru acea permutare care realizează codificarea corectă, rezultă că funcția de evaluare a cromozomilor poate fi definită prin

$$f_{obiectiv}(x) = \frac{1}{1 + cost(mesaj_x, mesaj_{codificat})}$$

fiind o funcție care trebuie maximizată și are valoarea maximă 1.

Implementare

Instrumente.py – generare fișier cu date de test. De la tastatură de introduce mesajul pentru codificare și numele fișierului care va fi generat. Fișierul cu date de test conține mesajul original și mesajul codificat, pe linii separate. Codificarea se face pe baza unei permutări generate aleator. Permutarea (cheia) va fi scrisă într-un fișier text cu nume asemănător cu cel introdus de la tastatură la care se adaugă „-cheie”.

Incercare.py – script pentru testarea algoritmului genetic pe mai multe seturi de date de test, cu diverși parametri de control. Vezi comentariile din script.

GA_codec.py – implementarea algoritmului genetic pentru rezolvarea problemei.

a b

n g

r f

e j

m o

anaaremere

bgbbfjofj

bzbbfjkjfj