

CS410

BÀI TẬP NỘI LẦN 1

Chạy thí nghiệm và viết báo cáo tìm kích thước quần thể nhỏ nhất cần thiết (*minimally required population size - MRPS*) để thuật toán di truyền đơn giản (*simple Genetic Algorithm - sGA*) với bản cài đặt POPOP để giải 2 vấn đề OneMax và Trap $k=5$.

Các bạn cài đặt sGA theo bản cài đặt POPOP với tournament selection $s = 4$, xác suất lai ghép 100%, và không sử dụng đột biến.

Kích thước vấn đề (problem size) l là số biến của vấn đề. Ví dụ

- OneMax $l = 40$ là vấn đề có 40 biến nhị phân.
- Trap $k=5$ với $l = 20$ là vấn đề có 20 biến nhị phân, được tạo nên bởi kết nối 4 hàm Trap $k=5$ lại với nhau.

Pseudo-code của phương pháp bisection tìm MRPS cho mỗi kích thước vấn đề l với 2 giai đoạn như sau:

Giai đoạn 1: Tìm cận trên (upper bound) của MRPS

$N^{upper} \leftarrow 4$

do

$N^{upper} \leftarrow 2 \times N^{upper}$

success $\leftarrow$ Có thể chạy thành công sGA 10 lần (mỗi lần với 1 random seed khác nhau) với kích thước quần thể N^{upper} ?

while success = false

Giai đoạn 2: Tìm giá trị của MRPS

$N^{lower} \leftarrow N^{upper} / 2$

while $(N^{upper} - N^{lower}) / N^{upper} > 0.1$ do

$N = (N^{upper} + N^{lower}) / 2$

success $\leftarrow$ Có thể chạy thành công sGA 10 lần (mỗi lần với 1 random seed khác nhau) với kích thước quần thể N ?

if success = true

$N^{upper} \leftarrow N$

else

$N^{lower} \leftarrow N$

if $(N^{upper} - N^{lower}) \leq 2$

break

Trả về $MRPS \leftarrow N^{upper}$

Lưu ý: Có thể dùng 10 random seed ở giai đoạn 2 là giống với 10 random seed ở giai đoạn 1.

Thực hiện xong giai đoạn 1 và giai đoạn 2 như trên là 1 lần chạy bisection. Mỗi lần chạy trả về một giá trị của MRPS. Chúng ta cần chạy 10 lần bisection như vậy để lấy dữ liệu thống kê. Lưu ý, với mỗi kích thước vấn đề l các random seed ở những lần chạy bisection trước không được sử dụng lại cho những lần chạy bisection sau.

Đồ thị trong báo cáo phải thể hiện được độ tăng của MRPS khi kích thước vấn đề tăng lên.

1. Trục hoành là kích thước vấn đề ($l = 10, 20, 40, 80, 160$).
2. Trục tung là MRPS trung bình (mean) của 10 lần chạy bisection tương ứng với mỗi kích thước vấn đề.
3. Mỗi điểm kết quả cần thể hiện thanh lỗi (error bar) là ± 1 độ lệch chuẩn (standard deviation).
4. Sử dụng log scale để biểu diễn trục hoành và trục tung.

Đồ thị trong báo cáo cần chứa 4 đường kết quả như sau:

- Kết quả sGA với phép lai một điểm (one-point crossover) cho OneMax.
- Kết quả sGA với phép lai đồng nhất (uniform crossover) cho OneMax.
- Kết quả sGA với phép lai một điểm (one-point crossover) cho Trap $k=5$.
- Kết quả sGA với phép lai đồng nhất (uniform crossover) cho Trap $k=5$.

Lưu ý:

Khi giải vấn đề Trap $k=5$, có thể kích thước quần thể N^{upper} sẽ tăng rất lớn. Trong báo cáo này, nếu N^{upper} tăng đến 4096 mà vẫn không thể chạy thành công sGA 10 lần thì chúng ta không cần chạy tiếp trường hợp đó nữa.

Các files cần nộp là:

1. Source codes.
2. Các files ghi dữ liệu kết quả (10 giá trị MRPS tìm được trong 10 lần bisection cho mỗi kích thước vấn đề và 10 random seeds tương ứng của mỗi lần chạy bisection). Đây là thông tin cần dùng để chạy lại thí nghiệm mà các bạn đã thực hiện.
3. File báo cáo ghi rõ họ tên, mã số sinh viên. Trong file báo cáo các bạn vẽ đồ thị như yêu cầu ở trên và ghi lại nhận xét/ý kiến của bản thân về hiệu suất của sGA khi giải OneMax và Trap với những phép lai ghép khác nhau. Chúng ta có thể làm gì để cải thiện hiệu suất của sGA?