Laboratory 2 - Set covering

Genetic algorithm

Francesco Scalera
Filippo Maria Cardano
Giuseppe Esposito

Problem overview

Given:

- N
- List of lists of integers: P = (L_0, L_1, L_2, ..., L_n)

Determine the subset S = (L_{s_0}, L_{s_1}, L_{s_2}, ..., L_{s_n}) such that the following constraints are met:

- The integers cover the whole sequence from 0 to N-1
- The sum of length of the lists is as lowest as possible

Fitness metric and feasibility

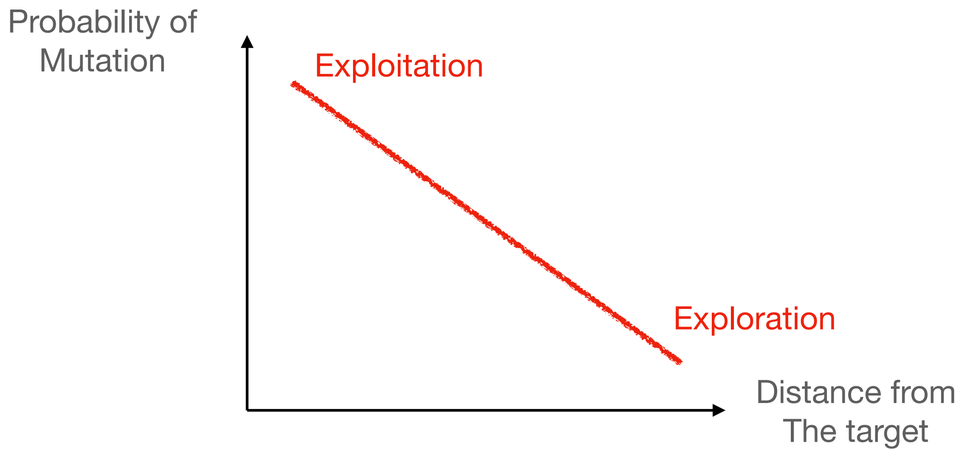
- Fitness metric based on the length of each selected list inside the individual
- Feasible individual: Given a goal based on N, we checked if all allele in the individuals covers the target
- Only feasible solutions within our population

Initial population

- Randomly selection of viable individuals
- Viability based on capability of an individual to provide a possible solution to the problem (with suboptimal fitness values)
- Individual are represented as a tuple, which contains:
    - The value of the fitness of the individual
    - Mask of booleans
    - Label that describe the genetic operator used to generate the individual

Genetic operators: mutation and crossover

- Parent(s) chosen with a tournament selection of size 2
- Randomly selection between the two genetic operators biased on the distance to optimal solution (weight = N, mostly unachievable for larger N)
- During the offspring generation we only consider the unique individuals wrt their encoding
- When we add it to the population we again remove duplicates of individuals



“History track” of our algorithm

- If the fitness of the best candidate individual doesn’t improve, it happen that the algorithm perform for several iterations the same genetic operator (steady state)
- We added a mechanisms to force the genetic operations of the next iteration to be the opposite of the one that generate the current best candidate
- If no improvement of fitness is found, we forced the algorithm to switch genetic operator again

Results

Parameter grid search

N	POPULATION_SIZE	OFFSPRING_SIZE	fitness	
5	20	10	5	0.2783920

N	POPULATION_SIZE	OFFSPRING_SIZE	fitness	Time ln s
10	50	33	10	1.09416

N	POPULATION_SIZE	OFFSPRING_SIZE	fitness	Time in s
20	200	133	23	2.8027846813201904
20	300	133	23	3.4154045581817627
20	500	133	23	5.425680875778198
20	600	133	23	6.6372973918914795
20	1000	133	23	14.002291202545166
20	2000	133	23	48.36100888252258
20	3000	133	23	105.10981178283691
20	5000	133	23	307.36173963546753

N	POPULATION_SIZE	OFFSPRING_SIZE	fitness	Time in s
100	50	133	181	21.201438426971436
100	200	200	164	17.683991193771362
100	300	200	182	18.85832977294922
100	500	200	189	26.038270711898804
100	600	200	183	29.530969381332397
100	1000	200	190	55.97995352745056
100	2000	200	197	143.39349246025085

N	POPULATION_SIZE	OFFSPRING_SIZE	fitness	Time ln s
500	50	133	15182	213.61597847938538
1000	50	200	199031	1605.7853195667267

THANK YOU
FOR THE ATTENTION

Francesco Scalera
Filippo Maria Cardano
Giuseppe Esposito