Faculdade de Engenharia da Universidade do Porto

Master in Data Science and Engineering

Artificial Intelligence

Joaquim Carneiro, João Patrício, João Pedro Pêgo

Introduction

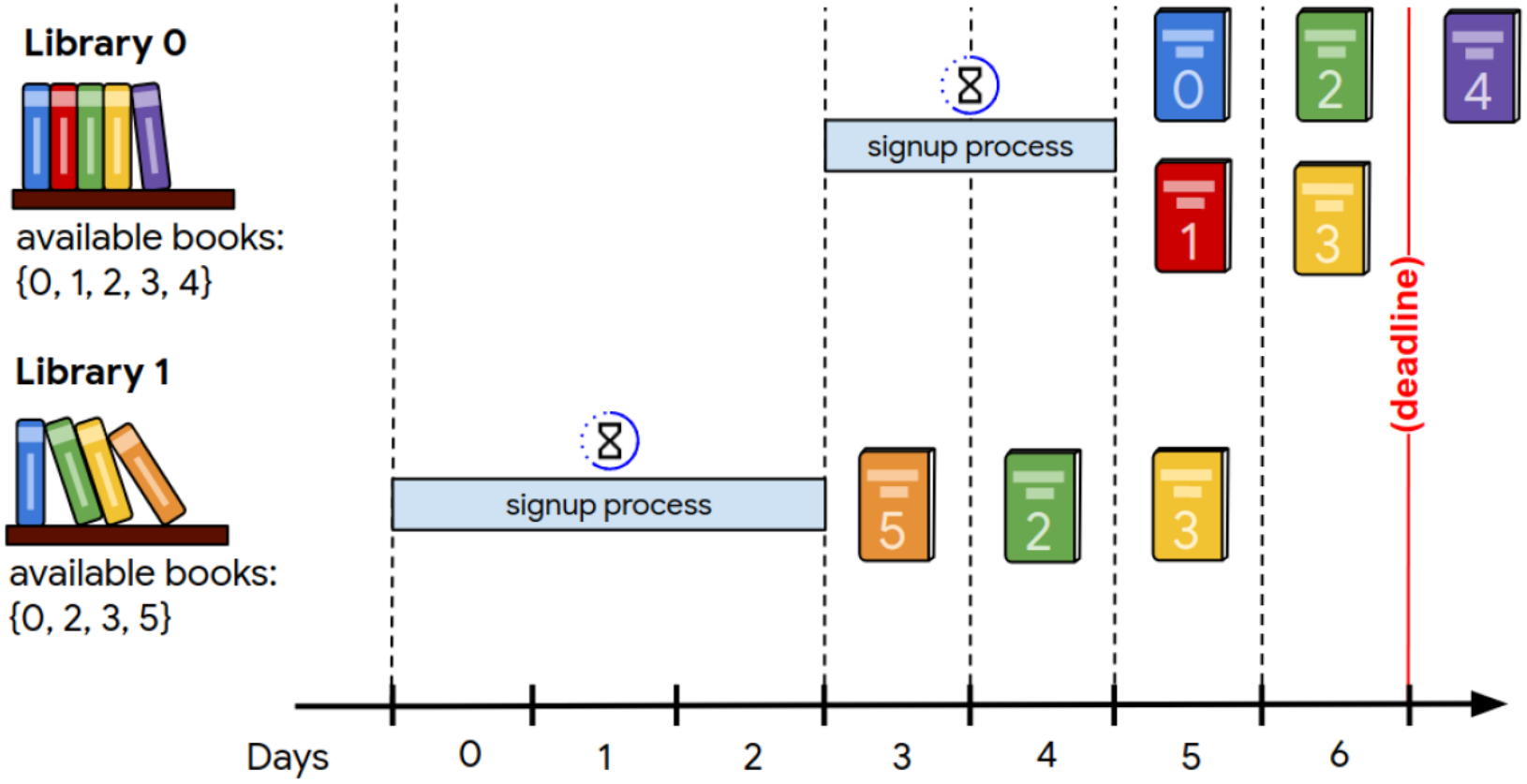
- Qualification Round of **Google Hash Code 2020**.

- Maximize **total score** obtained from **scanning** B **Books** from L **Libraries** over a **period of** D **days**.

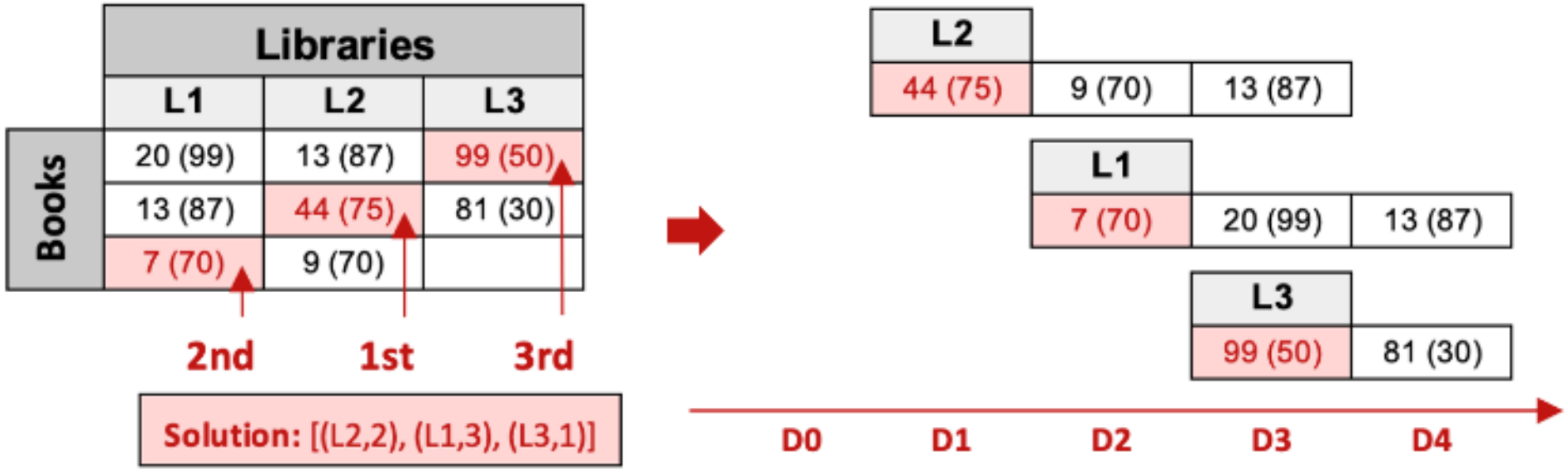
- Key decisions:
  - **Libraries** to be considered for scanning and **in what order**;
  - **Books** of a given library shipped for scanning and **in what order**.

Solve in 4 hours!

Scheduling Problem



Solution Representation



Search Space

$$S = L! \cdot \prod_{j=0}^{L-1} N_j!$$

Random books list

$$S_{red} = L! \cdot \prod_{j=0}^{L-1} N_j$$

Sorted books list

#books/library

File	B	L	D	μ_N	S	S_{red}	
a	6	2	7	4	$10^{3.76}$	$10^{1.6}$	
b	100k	100	1k	1k	10^{256918}	10^{457}	😄
c	100k	10k	100k	15	10^{158576}	10^{47328}	😈
d	78.6k	30k	30001	7	10^{244680}	10^{146863}	
e	100k	1k	200	491	$10^{1156803}$	10^{5124}	😬
f	100k	1k	700	509	$10^{1203551}$	10^{5141}	

NASA estimates the **universe age** at 13.77×10^9 years or **4.34×10^{20}ms**

https://wmap.gsfc.nasa.gov/universe/uni_age.html

Deterministic Solutions

Libraries sorted by:

1) ↓ “total score” / “signup period”

2) ↑ “signup period”

Books sorted by score.

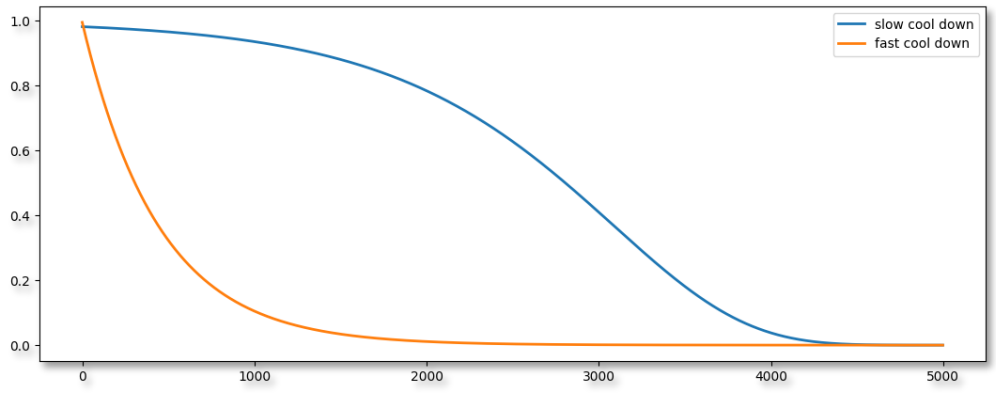
File	Sorted by “books score / signup period”	Sorted by “1/signup period”
a	21	21
b	5 822 900	5 822 900
c	5 645 747	5 467 966
d	4 815 395	4 815 395
e	3 714 416	3 977 298
f	5 227 905	2 703 359

Maximum score: 25 489 266

Benchmark!

Simulated Annealing Algorithm

- **Metropolis Criteria**
  - Temperature Schedule
  - Initial temperature
  - Nr of iterations
  - Average neighbour score difference



- **Key Implementation Details**
  - Neighbours at varying distances
  - Adaptive repetitions

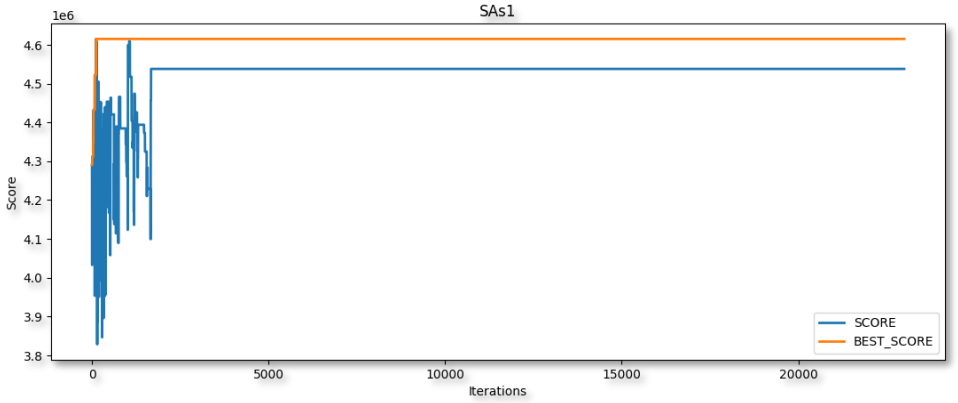
- **Neighbours**
  - Library rotation
  - Book rotation
  - Library swapping
  - …all at once!

L1	L2	L3	L1
20	13	99	20
13	44	81	13
7	9		7

L1	L2	L3
20	13	99
13	44	81
7	9	

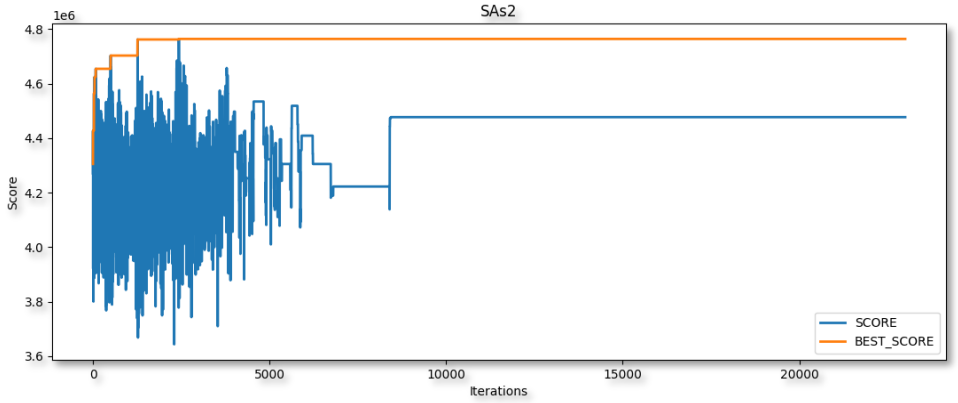
L1	L3	L2
20	99	13
13	81	44
7		9

Simulated Annealing Algorithm



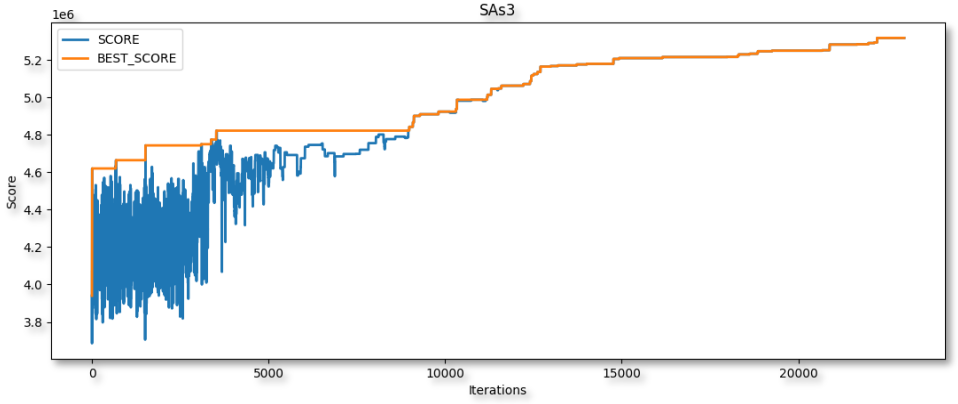
Best Score: 4 615 200
Max Neigh Distance: 30
Init Solution: Random 1

Schedule: Fast
Max Repetitions: 10



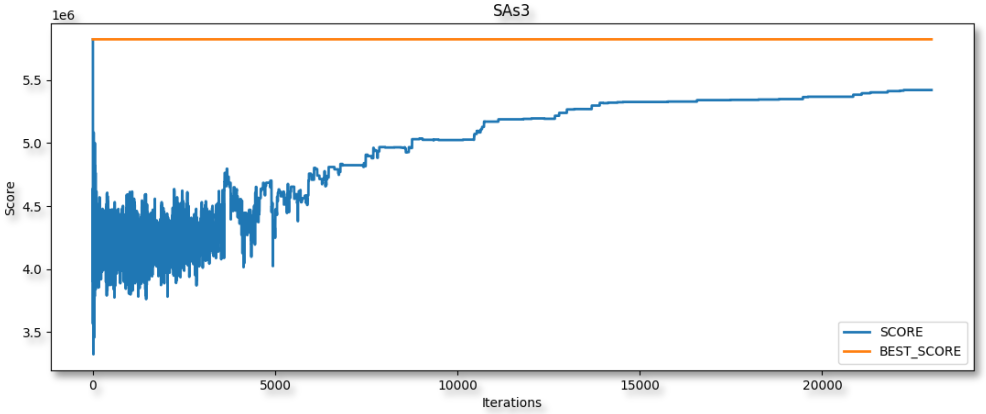
Best Score: 4 763 800
Max Neigh Distance: 30
Init Solution: Random 1

Schedule: Slow (alpha = 0.998)
Max Repetitions: 10



Best Score: 5 318 300
Max Neigh Distance: 100
Init Solution: Random 1

Schedule: Slow (alpha = 0.998)
Max Repetitions: 10



Best Score: 5 822 900
Max Neigh Distance: 100
Init Solution: Best Det Solution (5822900)

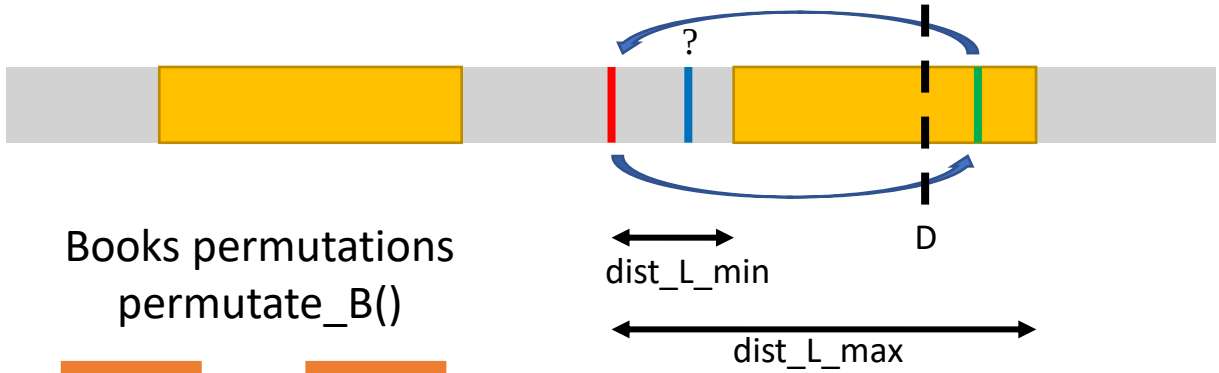
Schedule: Slow (alpha = 0.998)
Max Repetitions: 10

Tabu Search Algorithm

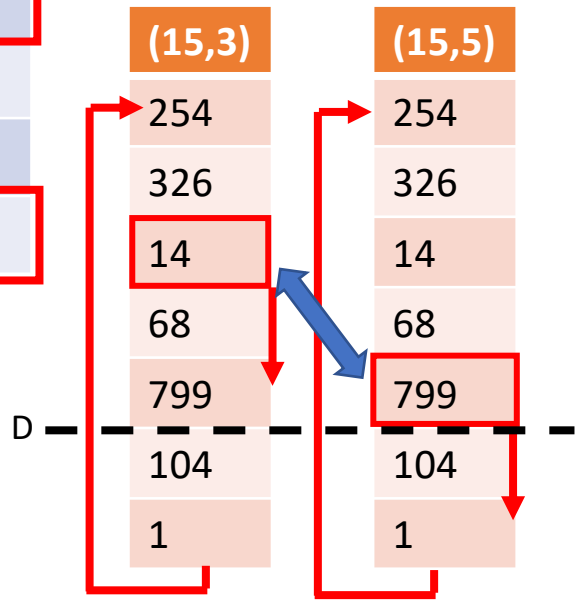
Parameter	Value
max_iter	800
max_iter_LS	25
iterations	20 000
dist_L_min	[0.05, 0.15, 0.25] * nr_libs
dist_L_max	[0.25, 0.50, 0.75] * nr_libs
dist_B	0.5 * nbooks[j]
tenure	sqrt(nr_libs)
LBbal	[0, alpha, 1]

```
if rnd.uniform(0,1) < LBbal:
  return permutate_L(idxL1)
else:
  return permutate_B(idxL1)
```

Libraries sequence premutations – permutate_L()



Books permutations
permutate_B()



LBbal

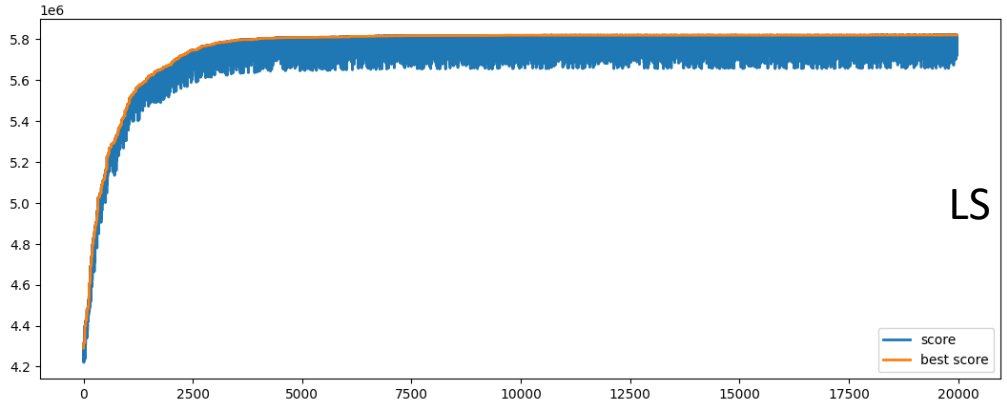
$$S_{red} = L! \cdot \prod_{j=0}^{L-1} N_j = a \cdot b$$

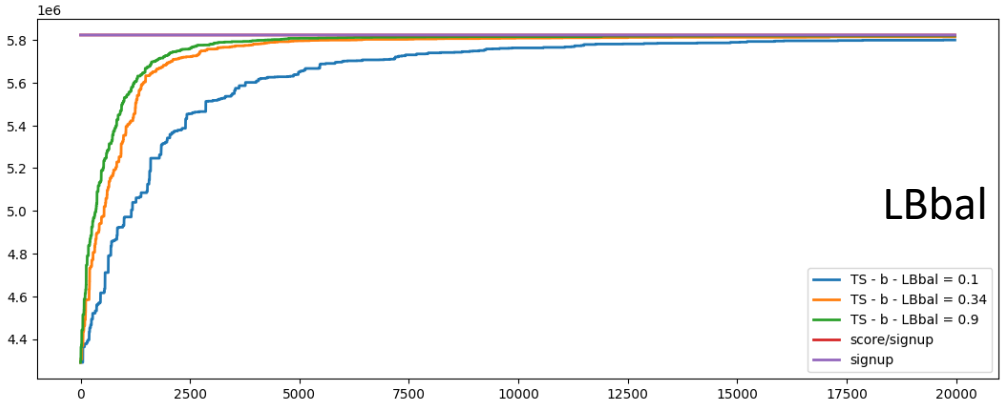
$$\log_2(S_{red}) = \log_2(a) + \log_2(b)$$

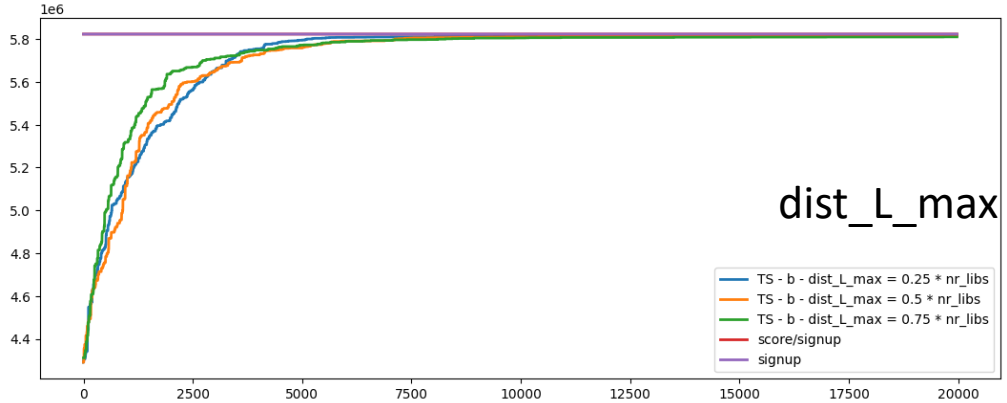
$$\alpha = \frac{\log_2(a)}{\log_2(S_{red})} \qquad \beta = \frac{\log_2(b)}{\log_2(S_{red})}$$

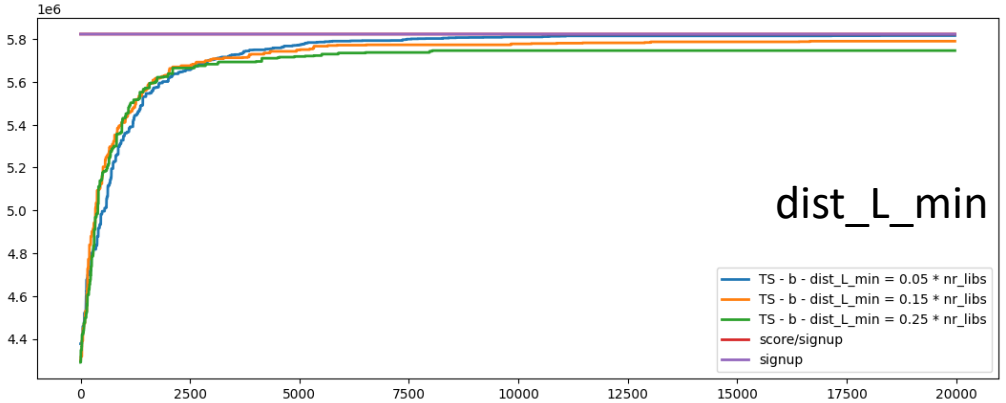
$$\alpha + \beta = 1$$

Tabu Search Algorithm

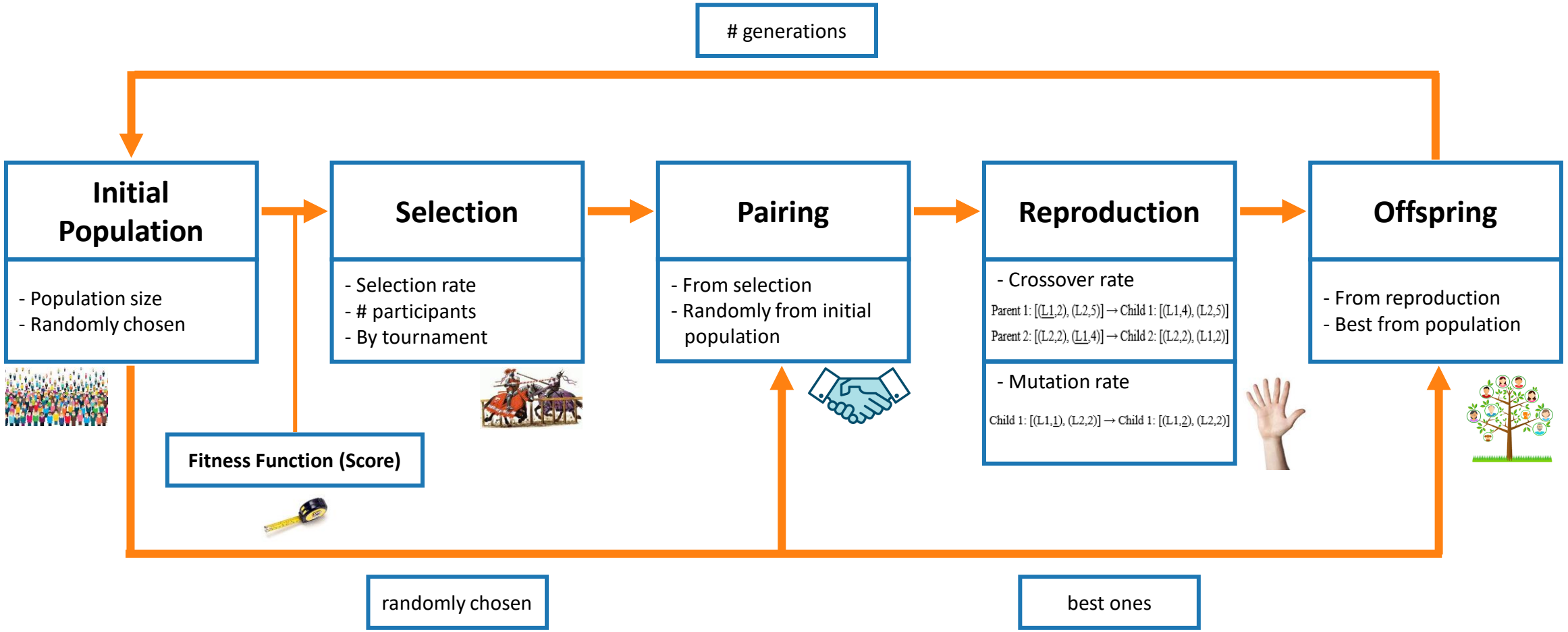




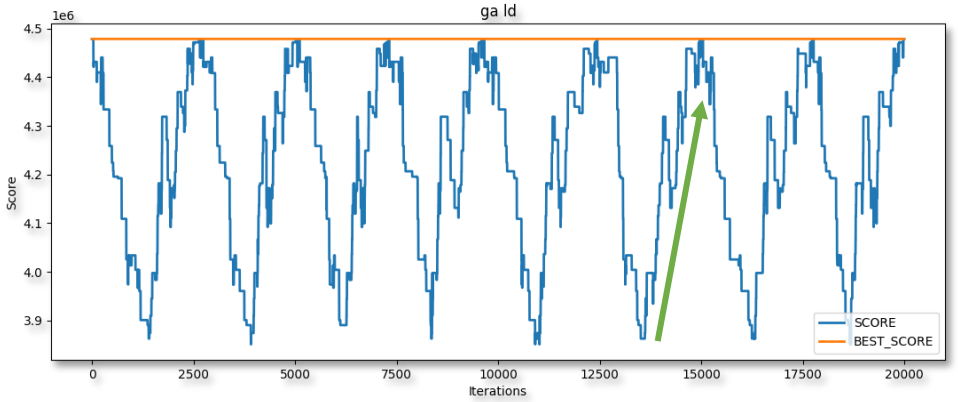




Genetic Algorithm

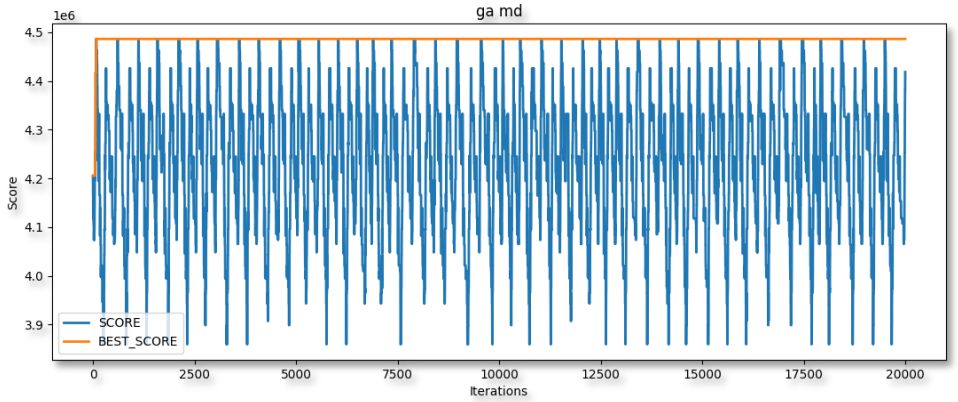


Genetic Algorithm



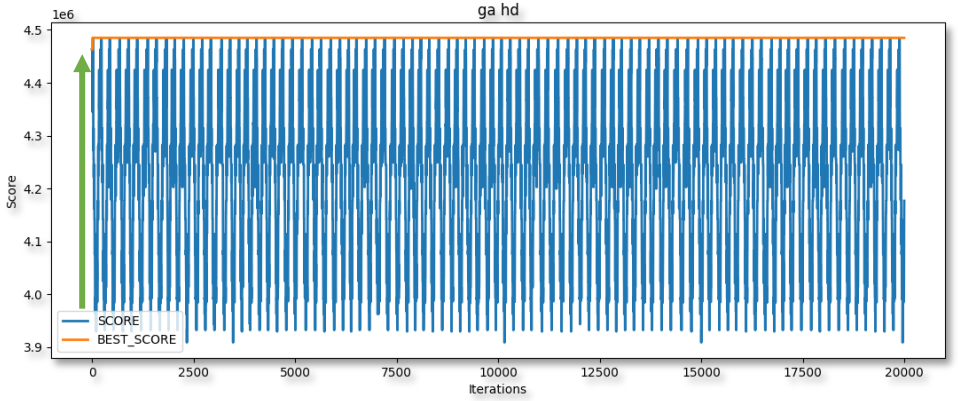
Low Diversity Steps
Population Size: 5
Selection: Tournament
Crossover Rate: 0,2

Best Score: 4478800
Generations Size: 20000
Selection Rate: 50%
Mutation Rate: 0,1



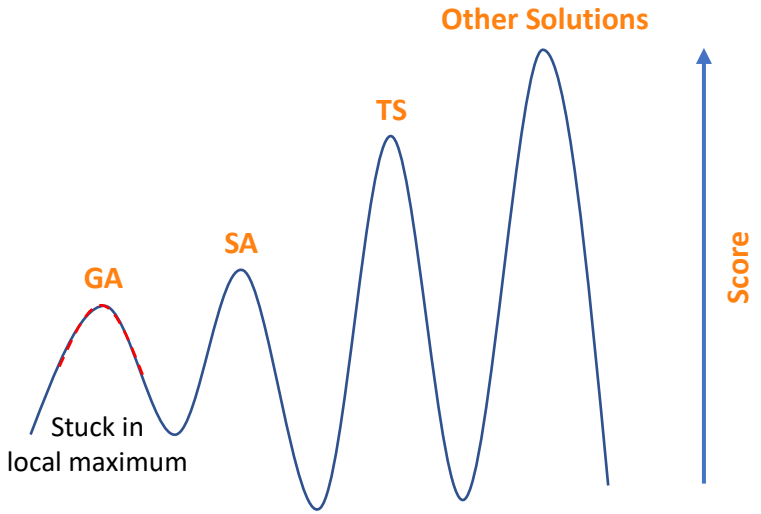
Medium Diversity Steps
Population Size: 5
Selection: Tournament
Crossover Rate: 0,5

Best Score: 4486200
Generations Size: 20000
Selection Rate: 50%
Mutation Rate: 0,4



High Diversity Steps
Population Size: 5
Selection: Tournament
Crossover Rate: 0,9

Best Score: 4485000
Generations Size: 20000
Selection Rate: 50%
Mutation Rate: 0,8



Next Steps

- Apply different crossover and mutation strategies;
- Apply Elitism.

Comparative Analysis

File: b_read_on

Deterministic solution: 5 822 900

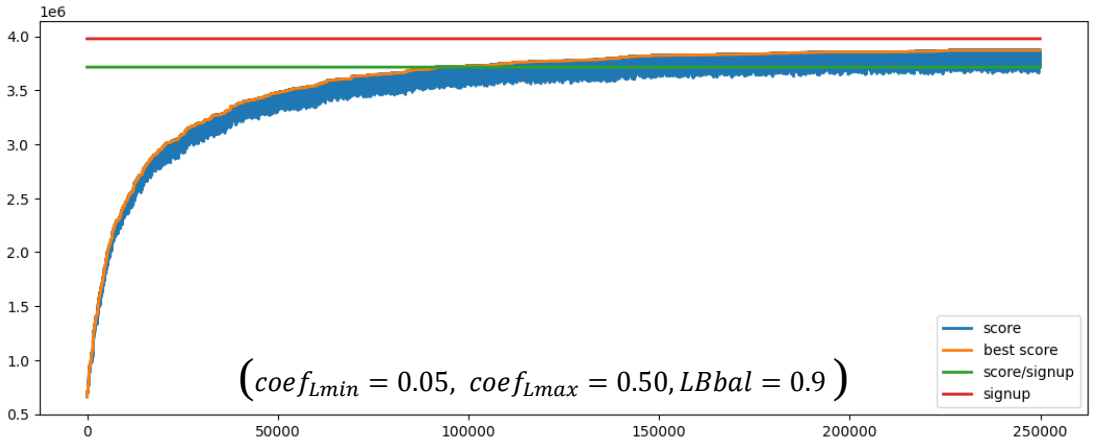
Key considerations:
- Time to compute not comparable.
- Random solutions (GA > SA > TS).
- Neighbour distance (avoid stuck in local maximum).
- All results bellow deterministic solution.

Algorithm	Solutions tested	Iterations to optimum	Time to compute [s]	Best Score
SA	20 000	2 431	368	5 318 300
TS	20 000	19 483	431	5 819 200
GA	20 000	78	2 915	4 486 200

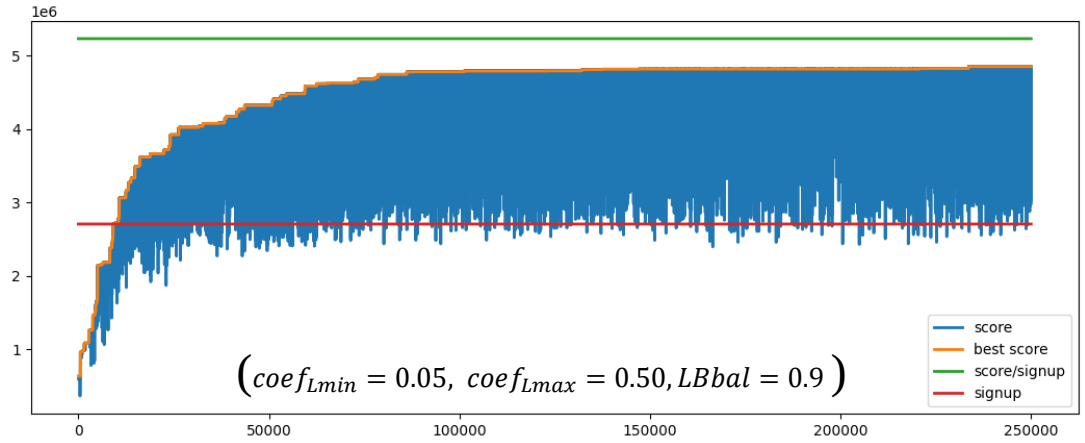
Conclusions & Future Work

- Data understanding is key to define which algorithm strategy to be adopted.
- A more aggressive search strategy is necessary to find better neighbours.
- Although stuck in a local maximum, GA converges quickly to a good solution.
- GA crossover admits that children perform better than parents, which is not necessarily true.
- Memory issues with TS.

Thank you!

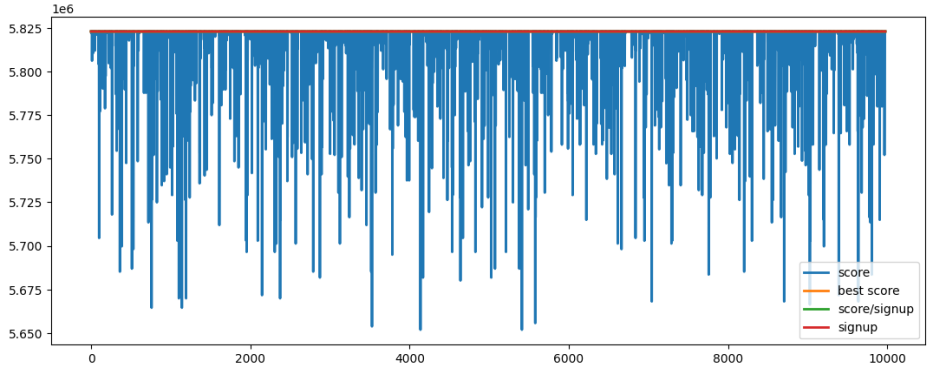


e_so_many_books

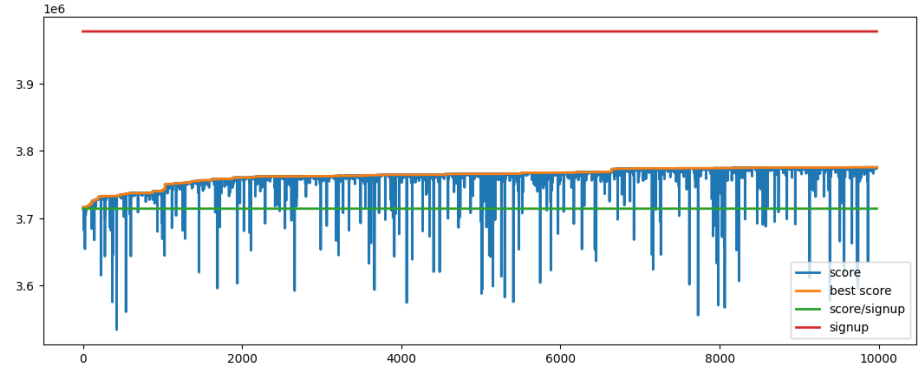


f_libraries_of_the_world

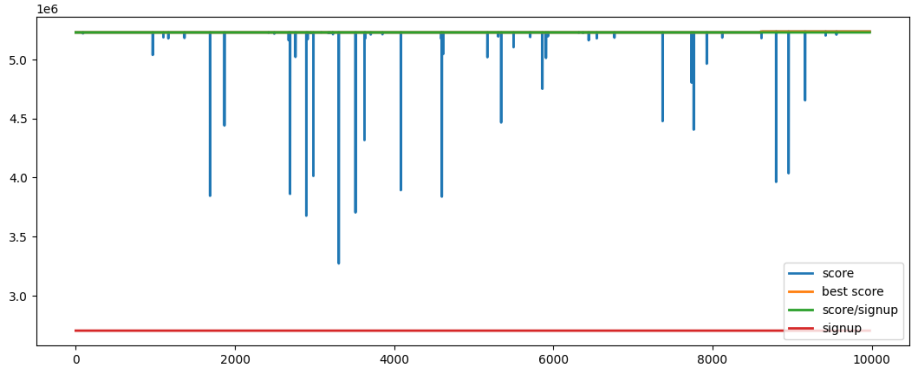
Breaking the ceiling



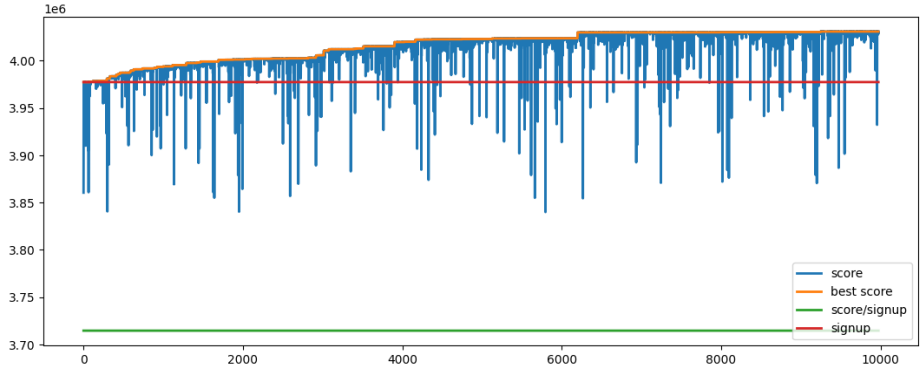
b | init_sol = score/signup



e | init_sol = score/signup



f | init_sol = score/signup



e | init_sol = signup