Uma imagem com texto, captura de ecrã, Tipo de letra, Marca

Os conteúdos gerados por IA poderão estar incorretos.

**CIFO PROJECT**

*Sports League Optimization*

Tiago Pedro Soares

20240655

**NOVA Information Management School**

**Instituto Superior de Estatística e Gestão de Informação**

Universidade Nova de Lisboa

**Statement of Integrity**

The following statement is my declaration that I have conducted this project work for the Computational Intelligence for Optimization (CIFO) course ethically and professionally. Throughout the development of this work, there was no plagiarism, misuse of external information, or falsification of results. Furthermore, I acknowledge that I have adhered strictly to the academic rules and ethical principles established by NOVA Information Management School.

**Index**

# **Formal Definition of the Problem**

* 1. **Optimization Problem Description**

In this project, we attempt to form balanced teams based on a set/dataset of players. The goal of this project is to group players in a way that maximizes the overall performance of teams while ensuring that the teams are balanced. The space for possible solutions to an optimization problem is very large. Our goal is to find a combination that maximizes the evaluation function (fitness).

* 1. **Representation of Individuals**

There are five teams representing each solution (individual), with a goalkeeper, two defenders, two midfielders, and two attackers. Players are aggregated into Team objects, so the vector has a length of five teams, and each element has a length of seven players. Based upon an analysis of the teams and a comparison of the requirements of each team, an integrated function, is\_valid (), determines if the individual is valid.

* 1. **Search Space**

The search space corresponds to the set of all possible valid combinations of players distributed across five teams, without repetitions and with the correct positional composition. Also, each team may not exceed the budget limit of 750 million euros. Solutions that do not comply with these conditions will be automatically considered invalid and penalized.

* 1. **Evaluation Function (Fitness)**

The fitness function can be predefined using function toolkit predefined that has the ability of allowing the user to dynamically set the fitness function based on multiple criteria. By forming two bins based on the average. If there is an equal division between two different solutions. If there is a larger deviation between the splits. You also must specify the performance of various metric goals like the recall, an f score and decide if you are looking to minimize false positive rate or maximize true positive.

* 1. **Restrictions**

It is ensured that composition and budget constraints are met through internal methods that ensure: Per team, there is one goalkeeper, two defenders, two midfielders, and two forwards; There is only one team for each player; The maximum amount per team is € 750 M.

With this model, genetic operators such as selection, crossover, and mutation can be applied effectively, maintaining solutions' viability and exploring the space with search and diversity.

# **Description of Selection and Genetic Operators**

During the development of the project, several genetic operators and selection strategies were implemented and tested. A major goal of the project was to develop balanced teams that adhered to all defined constraints, including the composition of individual positions and the budget limit. They are fundamental to the functioning of the genetic algorithm and directly impact the quality of the solution that is generated in successive generations.

* 1. **Selection Strategies**

To maintain a balance between selective pressure and genetic diversity, several selection strategies were applied during the reproduction process. These strategies include the following:

**Tournament**: This is accomplished by selecting a small group of individuals at random (for example, two or three) and selecting the one with the best performance. As a result of this approach, moderate selective pressure is ensured, and premature convergence can be avoided.

**Roulette Wheel**: Each individual is assigned a selection probability proportional to the level of fitness. In general, top performers are given preference, but everyone retains some chance of being selected.

**Truncation**: A methodology thar involves randomly selecting an individual from the best in the population (for example, the top 30% of the population). As a result of this approach, evolution is directed toward the most promising solutions.

**Rank**: When individuals are ranked according to their performance, probability distributions can be distributed according to rank, which provides a method for smoothing out the effects of extreme variations between individuals in their fitness values.

**SUS (Stochastic Universal Sampling)**: It is a variation of roulette in which the choices are distributed more randomly, thus reducing the variance associated with chance.

**Elitism**: The concept of elitism ensures the preservation of the best individuals at the onset of each generation, ensuring the survival of high-quality solutions through the evolutionary process.

* 1. **Crossovers Operators**

As far as crossbreeding is concerned, different approaches have been developed, adapted to the nature of the solutions, which involve team rosters made up of players with specific positions:

Team Swap: Swapping of an entire team between two parents is referred to as a team swap. As a result of this approach, extensive modifications can be made to the solution in a controlled manner.

Player Swap: Players are exchanged between teams from different countries in the same position, ensuring continuity of positional structure.

Uniform Team: For each team, there is a 50% chance of inheritance from one parent or the other, thus resulting in genetically varied offspring.

Position-Based Crossing: Select a position (such as goalkeeper or striker) and swap all players in that position between the corresponding teams.

Gene-Level Crossing: rebuilds teams from a mixed list of players from both parents, respecting the composition rules and without repeating players. This approach is more disruptive and promotes diversity.

* 1. **Mutation Operators**

Several mutation operators were also implemented, with the aim of introducing specific variations in the solutions and thus preventing the algorithm from getting stuck in local solutions:

Random Substitution: This process involves swapping a player from one position to another who has not yet been used.

Team Swap: Choose two players from different teams in the same position and swap them between them.

Team Replacement: The replacement of a valid team is attempted, provided the limitations are respected, by generating a new valid team in place of the existing one.

Most Expensive Player Replacement: Identify the most expensive player and try to replace him with an alternative player who is more cost-effective.

Cross Position Swap: The Cross Position Swap involves swapping players from two teams who play the same position in different teams at random.

Replacement Strikers: When available, replace both strikers on a team with new players.

Team Rotation: An individual may experience a team rotation in which teams are rearranged within an individual, resulting in later crossover operations being affected.

# **Performance Analysis**

* 1. **Comparação Global dos Métodos de Seleção**

The two figures presented represent a comparative analysis of the different selection methods used in genetic algorithms. Fig. 1 ("comparacao\_melhor\_por\_selecao") illustrates the best individual over a defined number of generations, while Fig. 2, " curvas\_fitness\_por\_selecao" displays the average fitness by generation along with confidence intervals, demonstrating the consistency of the methods.

Fig. 1 shows that the elitism method reaches the maximum fitness value of 1000 within the first two generations (around the 6th generation) and maintains it until the very end. According to the behavior shown here, initial exploration and conservation of the best solutions are both possible. The rank method can also lead to optimal fitness, although the curve is more gradual, reaching optimal fitness after many generations.

It achieves good fitness values slower than truncation, tournament, or roulette methods, but converges much later. In contrast, the SUS (Stochastic Universal Sampling) method has the most irregular and slowest curve among the six, implying greater variability and less stability.

In addition to this analysis, Fig. 2 shows the average trends over 30 generations of each methodology. Despite the differences in best execution, all methods converge very closely by the 15th generation, with elitism and truncation showing slight advantages. Besides retaining the best individual, elitism is also consistent, with a smaller confidence interval, indicating little variability between interpretations.

Although the roulette method exhibits greater variation, it may suggest that genetic parameters strongly influence the method. In addition, the SUS method continues to display the lowest average and greatest dispersion, confirming the conclusions drawn from Fig. 1.

* 1. **Analysis of Fig. 3 - 33**

Across all six selection strategies - Elitism, Rank, Roulette, SUS, Tournament, and Truncation - the interaction between selection, crossover, and mutation operators influences genetic algorithms' performance. There were team\_swaps, player\_swaps, and aggressive mutation operators, including random\_player, replace\_team, and cross\_position. With rank selection, stability is assured, and stagnation is avoided early. As a result of frequent mutations and recombination, roulette produced good results. When it came to uniform\_teams and team\_swaps, it excelled. Tournament selection enabled multi-runs with maximum fitness in various configurations. Finally, Truncation selection may be highly effective when combined with intensive mutation strategies and structural crossovers, such as uniform\_team, team\_swap, and player\_swap. Using these combinations, fitness of 1000 was achieved quickly and consistently, consolidating them as global strategies. The analysis combines elitist selection with pressure-based selection (truncation or tournament), crossover with high recombination capacity, and mutations with 100% intensity, allowing a limited number of generations to explore. We can observe these results with Fig. 3 – 33.

* 1. **Analysis of Selection, Crossover, Mutations, and Genetic Rates**

Evolutionary operators' performance dynamics are strongly aligned with empirical results. Based on Fig. 34, Rank (992.30), Truncation (991.64), and Tournament (990.87) had the highest average fitness values, outperforming SUS (990.01), Roulette (989.91), and Elitism (989.43). The top-performing methods have tighter, more concentrated distributions (Figs. 35 and 36), with fewer outliers, reflecting superior robustness and convergence stability. In Fig. 35, team\_swap\_crossover, player\_swap\_crossover, and uniform\_team\_crossover consistently reach near-optimal fitness. Despite this, gene\_level\_crossover shows lower medians and more dispersed outcomes, indicating weaker robustness. (Fig. 36) Both random\_player\_mutation and replace\_team\_mutation produce near-optimal results. Expensive\_player\_mutation and forward\_reset\_mutation show significant dispersion for strategies requiring consistency and fast convergence. Further supporting these findings, the heatmap of mean fitness by crossover and mutation (Fig. 37) shows that combinations like uniform\_team\_crossover + rotate\_teams\_mutation and team\_swap\_crossover + replace\_team\_mutation achieve fitness values up to 1000, with little variance across selection methods. Crossovers demonstrate strong recombination capacity and structural diversity, helping the search process escape local optima. In both pair plots of fitness versus genetic rates (Fig. 38) and the mutation rate vs fitness scatterplot (Fig. 39), 100% mutation rates are consistently associated with high fitness. As a result, aggressive mutation leads to better final solutions. The violin plot (Fig. 40) shows that Tournament, Rank, and Truncation produce denser and more stable fitness distributions than Elitism or Roulette.

1. **Justification of Decisions**

# **Conclusions**

**Appendix**

**Annex**

A picture containing text, screenshot, font, design

Description automatically generated