

DNA Fragment Assembly Using a Teaching–Learning-based Optimization Algorithm

Abdelkamel Ben Ali

University of El Oued, Department of Computer Science, 39000 El Oued, Algeria

Abstract

DNA Fragment assembly is the task of inferring a genome given a large number of substrings covering it. It was considered as a generalization of the shortest common superstring problem, which is equivalent to finding a Hamiltonian path (an order of DNA fragments) in the overlap graph [3]. This is an NP-hard permutation-based combinatorial optimization problem for which efficient approximate algorithms are required to produce near-optimal results. Several heuristic techniques and metaheuristic algorithms have been applied in the literature to solve the fragment assembly problem, see [1] for detailed references.

In [2], we designed a fast and efficient problem aware local search algorithm (PALS), called *PALS2-many*. In a recent study [1], we developed a new discrete Particle Swarm Optimization (PSO) algorithm that affectively addresses the assembly problem. We also tested hybrid algorithms which take advantage of the strengths of the proposed discrete PSO and a modified variant of *PALS* called *quick-PALS*. The comparison results show the robustness and accuracy of the used algorithms over other techniques.

In this work, we propose an adaptation of the novel **Teaching–Learning-Based Optimization (TLBO)** algorithm [4] for assembling DNA fragments. TLBO is a population-based metaheuristic that simulates the classical school learning process. The TLBO algorithm implemented in this work operates in two stages. During the first stage (*Teacher Phase*), a **Probabilistic Edge Recombination Crossover (PERX)** operator is applied to build a new solution from three permutation vectors (orders of fragments ID): the current solution, the best individual in the population, and a mean permutation vector calculated from all individuals. During the second stage (*Learner Phase*), a PERX operator is applied to the current solution and a randomly selected individual for building two new solutions. In both stages, the new solutions will be only accepted into the population if they are better than the current solution. The algorithm continues until the termination condition is met. Our simulation results on literature benchmarks show that our PERX-based TLBO algorithm, in its basic form, has a very good performance and outperforms our basic discrete PSO algorithm proposed in [1].

Keywords: DNA fragment assembly, Ordering problem, Edge recombination, Teaching-Learning-based optimization

References

- [1] A. Ben Ali, G. Luque, and E. Alba An efficient discrete PSO coupled with a fast local search heuristic for the DNA fragment assembly problem. MS submitted to *Information Sciences*.
- [2] A. Ben Ali, G. Luque, E. Alba, and K. E. Melkemi. An improved problem aware local search algorithm for the DNA fragment assembly problem. *Soft Computing*, 21(7):1709–1720, 2017.
- [3] E. W. Myers Jr. A history of DNA sequence assembly. *It-Information Technology*, 58(3):126–132, 2016.
- [4] M. Črepinšek, S-H Liu, and L. Mernik. A note on teaching–learning-based optimization algorithm. *Information Sciences*, 212:79–93, 2012.