Sorting Permutations by Reversals Through Branch-and-Price

Alberto Caprara • Giuseppe Lancia • See-Kiong Ng
DEIS, Università di Bologna, Viale Risorgimento 2, I-40136 Bologna, Italy
DEI, Università di Padova, Via Gradenigo 6/A, 35131 Padova, Italy
Smithkline Beecham Pharmaceuticals R&D, Bioinformatics,
New Frontiers Science Park (North), Third Avenue,
Harlow, Essex CM19 5AW, UK
acaprara@deis.unibo.it • lancia@dei.unipd.it • skng@cs.cmu.edu

We describe an exact algorithm for the problem of sorting a permutation by the minimum number of reversals, originating from evolutionary studies in molecular biology. Our approach is based on an integer linear programming formulation of a graph-theoretic relaxation of the problem, calling for a decomposition of the edge set of a bicolored graph into the maximum number of alternating cycles. The formulation has one variable for each alternating cycle, and the associated linear programming relaxation is solved by column generation.

A major advantage with respect to previous approaches is that the subproblem to face in the column-generation phase no longer requires the solution of min-cost general matching problems, but of min-cost *bipartite* matching problems. Experiments show that there is a tremendous speed-up in going from general matching to bipartite matching, although the best-known algorithms for the two problems have the same theoretical worst-case complexity. We also show the worst-case ratio between the lower bound value obtained by our new method and previous ones.

We illustrate the effectiveness of our approach through extensive computational experiments. In particular, we can solve to proven optimality the largest real-world instances from the literature in a few seconds, and the other (smaller) real-world instances within a few milliseconds on a workstation. Moreover, we can solve to optimality random instances with $n = 100$ within 3 seconds, and with $n = 200$ within 15 minutes, where n is the size of the permutation, whereas the size of the instances solvable by previous approaches was at most 100. We also describe a polynomial-time heuristic algorithm that consistently finds solutions within 2% of the optimum for random instances with n up to 1000.
(*Programming, Integer, Algorithms; Programming, Integer, Applications; Networks-Graphs, Matchings; Analysis of Algorithms*)

1. Introduction

Many among the most effective approaches to solve combinatorial optimization problems are based on the definition of an *integer linear programming* (ILP) formulation and the solution of the associated *linear programming* (LP) relaxation. A key issue for the success of this scheme is the "strength" of the ILP formulation, i.e., the fact that the optimal solution values of the ILP and its LP relaxation are "reasonably close" to each other. It is often a challenging task to derive

INFORMS Journal on Computing © 2001 INFORMS
Vol. 13, No. 3, Summer 2001 pp. 224–244
0899-1499/01/1303/0224$5.00
1526-5528 electronic ISSN