A Tree · A Window · A Hill; Generalization of Nearest-Neighbor Interchange in Phylogenetic Optimization

David Sankoff
Université de Montréal

Yvon Abel
Université de Montréal

Jotun Hein
Université de Montréal

Acknowledgments: This work was supported in part by grants to the first author from the Natural Sciences and Engineering Research Council (Canada) and the Fonds pour la formation de chercheurs et l’aide à la recherche (Québec), and to the third author from the Danish Research Council. The first author is a fellow of the Canadian Institute for Advanced Research. Much of the research was carried out in the spring of 1991 while the first author was visiting the University of Geneva; warmest thanks are due Professor Claude Weber for this opportunity.

Authors’ Addresses: David Sankoff and Yvon Abel, Centre de recherches mathématiques, Université de Montréal, CP 6128 Succursale centreville, Montréal, Québec (Canada) H3C 3J7; Jotun Hein, Institute of Biological Sciences, University of Aarhus, Building 540, Ny Munkegade, DK 8000 Aarhus C, Denmark.
Abstract: The method of nearest-neighbor interchange effects local improvements in a binary tree by replacing a 4-subtree by one of its two alternatives if this improves the objective function. We extend this to k-subtrees to reduce the number of local optima. Possible sequences of k-subtrees to be examined are produced by moving a window over the tree, incorporating one edge at a time while deactivating another. The direction of this movement is chosen according to a hill-climbing strategy. The algorithm includes a backtracking component. Series of simulations of molecular evolution data parsimony analysis are carried out, for $k = 4, \ldots, 8$, contrasting the hill-climbing strategy to one based on a random choice of next window, and comparing two stopping rules. Increasing window size $k$ is found to be the most effective way of improving the local optimum, followed by the choice of hill-climbing over the random strategy. A suggestion for achieving higher values of $k$ is based on a recursive use of the hill-climbing strategy.

Keywords: Local optimization; Heuristics; Molecular evolution.

1. Introduction

When the number of nodes in a tree optimization problem for phylogenetic inference exceeds computational capacities for methods that assure a global optimum, heuristic techniques are necessary. As summarized by Swofford and Olsen (1990, pp. 486-490), and as incorporated, for example, in Swofford's (1990) PAUP package, these approaches start with an initial tree, constructed randomly or using some clustering method applied to data on the $n$ labeled terminal nodes, and then try to improve the tree, i.e., try to find a better (according to the optimization criterion) configuration of edges and (unlabeled) non-terminal nodes, by a series of elementary rearrangements of the branching pattern, until a local optimum is found.

The simplest of these elementary rearrangement operations, dating back at least to Camin and Sokal (1965), is the nearest-neighbor interchange (NNI). This rearrangement operation focuses on a small subtree consisting, in the case where all non-terminal nodes have degree 3, of two connected non-terminal nodes $e$ and $f$ (i.e., connected by the central edge $ef$) and the four other nodes $a$, $b$, $c$, and $d$ connected to $e$ and $f$ by non-central edges $ae$, $be$, $cf$, and $df$. This subtree is sometimes called a 4-subtree, because it has four terminal nodes $a$, $b$, $c$, and $d$, even if some of these are non-terminal in the original $n$-tree. The NNI, to be described more formally in the next section, can be thought of as breaking off one of the four non-central edges, say $ae$, including the node $e$ and whatever subtree is connected to $a$, and reattaching it to the middle of one of the two remaining non-central edges, say $cf$, thus producing a new 4-subtree with edges $ae$, $ce$, $ef$, $bf$, and $df$. 