

Comparison of Heuristics for Maximum Likelihood Tree Estimation

Milica Hadzi-Tanovic

Introduction

- Phylogeny estimation of evolutionary history is one of the central topics in bioinformatics
- This project will focus on Maximum Likelihood (ML) phylogeny estimation

ML Problem Review

- Input: a set sequences of same length S
- Output: a model tree that maximizes the probability of generating the input sequences
- NP-hard
- Statistically consistent
- Software: RAxML, ExaML, FastTree, PhyML, and IQ-TREE

Heuristics Used by Software

- Each method uses some fast technique to get the initial tree and then uses heuristics to optimize them
- **FastTree2** uses maximum likelihood nearest-neighbor interchanges (NNI) and minimum-evolution subtree-pruning-regrafting to improve the tree (SPR)
- **IQ-TREE** combines hill-climbing techniques, random perturbation of current best trees, and a broad sampling of initial starting trees to improve the final tree based on NNI
- **RAxML** is a parallel algorithm that uses rapid hill-climbing algorithm which is able to infer large trees, also based on NNI

Project Objective

- The goal of this project is to analyze the introduced algorithms in depth and acquire a good understanding of each of them
- Given the literature, I want to understand what each algorithm is doing and see what are the time-accuracy trade offs for each method in the study

Evaluation

- Download and run each ML method examined in the study
- Data set of around 200 sequences from <https://pranj.al/mammalian-tutorial.html>
- This will help me how long they take and what ML scores I can obtain.

Bibliography

- **RAxML Version 8:** A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. A. Stamatakis. Public Library of Science, 2011.
- **IQ-TREE:** A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. L. Nguyen, H. Schmidt, Arndt von Haeseler, and B. Minh. Mol Biol Evol, 2015.
- **FastTree 2** – Approximately Maximum-Likelihood Trees for Large Alignments. M. Price, P. Dehal, A. Arkin. PLoS One, 2010.
- **RAxML and FastTree:** Comparing Two Methods for Large-Scale Maximum Likelihood Phylogeny Estimation. K. Liu, R. C. Linder, and Warnow, Tandy. Public Library of Science, 2011.