

ABSTRACT

Phylogenetic distance estimation between taxa in a tree is critical for tree reconstruction. There exist several classical methods that are already established in the field of phylogenetics. However, these classical methods of distance estimation are not straightforward and their computation can be tedious. As neural networks show great success in pattern recognition tasks, it is reasonable that a neural network can estimate the evolutionary distances well without making any assumptions on the mathematical model of evolution. The training and testing of the neural network can be done on phylogenetic data simulated under various models of evolution. Indeed, the network was able to estimate the distance between taxa and keep up with the classical methods. The best-performing network was trained under the GTR model of evolution and was able to generalize to different data types. Moreover, it is advantageous that the network does not have to incorporate phylogenetic background knowledge. This could be a starting point to improve the estimation of evolutionary distances.

Keywords: phylogenetics, evolution, distance estimation, phylogenetic trees, neural networks, deep learning