

Phylogenetic Transfer Learning

Michael D. Catchen^{1,2}

¹ McGill University; ² Québec Centre for Biodiversity Sciences

Correspondance to:

Michael D. Catchen — michael.catchen@mcgill.ca

Substance: Find a bunch of ASR datasets, do crossvalidation on tip states, test predictive efficacy on inputting 'hidden' species.

Premise: Imputing traits among extant species using ancestral state reconstruction (ASR) can be useful for a wide variety of purposes in ecology

Keywords:

ancestral state reconstruction
transfer learning
predictive ecology

1

What is Phylogenetic Transfer Learning

1.1. Ancestral State Reconstruction *Ancestral State Reconstruction* (ASR; also called ancestral character reconstruction, character estimation, or character mapping) is a core topic in phylogenetics (**Joy2016AncRec?**).

The goal is, given an estimate of a phylogeny \mathcal{P} and trait values T_i for each extant species in the phylogeny i , to estimate the value of that trait at some historical point in the phylogeny (typically at the node representing the MRCA of a clade of interest).

To do this, one assumes a statistical model of evolution (and often tries to infer the best among a set of candidate models). Depending on whether the trait of interest is discrete or continuous, the models are typically discrete-space Markov chains (where the transition matrix is a target of inference) or, in the continuous setting, either Brownian Motion or an Ornstein-Uhlenbeck (OU) process, where the former is used for neutral evolution and the latter is used when there is hypothesized selection on the trait.

The methodology used to fit models has followed the historical progression from maximum parsimony models (a naive approach that favors as few evolutionary changes as possible in discrete traits) to maximum likelihood estimation. Modern methods revolve on using Bayesian methods, which has the direct benefit of including uncertainty estimates in inferred ancestral states.

1.2. Phylogenetic Transfer Learning The core insight of Phylogenetic Transfer Learning (PTL) is to take the the model of evolution and ancestral state inferred via ASR, and simulate the parameterized evolutionary model forward from the ancestral node to get an estimate of extant trait values (with uncertainty) for species without trait observations T_i .

The *transfer learning* component in particular comes from the first use of this methods to impute latent representations of species (**Strydom2022FooWeb?**; **Strydom2022GraEmb?**)

Substance of the paper

1. Simulated phylogenies
 - Various properties of radiation
2. Real data

Relevant questions: - What phylogenetic scales can PTL give robust predictions - There is a clear upper limit (if the MRCA for set of species is too long ago, there will be massive variance in predicted state at the tip)
- There is also a lower limit (if there is HZT and the phylogeny is not a reliable proxy for how traits are evolving) - Robustness to noise in trait measurements? - How does this vary w/ amount and different types of selection

Possible Applications

- SDMs: mechanistic niche predictions on subsets of species
- Connectivity (**Catchen2023ImpEco?**)