

Approximate nearest neighbors search and *yaImpute*; new weapons in the battle against “Big N”

Andrew O. Finley¹, Nicholas L. Crookston², Alan R. Ek¹

Abstract: Nearest neighbors methods are gaining popularity in the field of natural resources inventory, emphasizing multi-source forest attribute estimation. With this popularity comes an increased demand for large-scale maps and real-time online mapping applications. The brute force nearest neighbors search is very time consuming, especially when the reference data set is large and distance is measured in high dimension. These inefficient searches increase mapping time and preclude useful online mapping tools. Our paper shows that approximate nearest neighbors (ANN) methods can significantly reduce search time with only marginal increase in error. We specifically explore the application of ANN search algorithms and associated data structures within the context of multi-source forest attribute mapping. The ANN methods are implemented in the new *R* package, *yaImpute* (Yet Another Impute). This package uses ANN methods to efficiently search within space defined by vectors produced by Mahalanobis projection, canonical correlation analysis, and canonical correspondence analysis.

1. Department of Forest Resources, College of Food, Agricultural, Natural Resource Sciences, University of Minnesota, St. Paul, MN 55108. E-mail: afinley@stat.umn.edu

2. Rocky Mountain Research Station, USDA Forest Service, 1221 South Main Street, Moscow, ID 83843. E-mail: ncrookston@fs.fed.us