UPGMA

From Wikipedia, the free encyclopedia

UPGMA (**U**nweighted **P**air **G**roup **M**ethod with **A**rithmetic Mean) is a simple agglomerative (bottom-up) hierarchical clustering method. It is one of the most popular methods in ecology for the classification of sampling units (such as vegetation plots) on the basis of their pairwise similarities in relevant descriptor variables (such as species composition).[1] In bioinformatics, UPGMA is used for the creation of phenetic trees (phenograms). In a phylogenetic context, UPGMA assumes a constant rate of evolution (molecular clock hypothesis), and is not a well-regarded method for inferring relationships unless this assumption has been tested and justified for the data set being used. UPGMA was initially designed for use in protein electrophoresis studies, but is currently most often used to produce guide trees for more sophisticated phylogenetic reconstruction algorithms.

The UPGMA algorithm constructs a rooted tree (dendrogram) that reflects the structure present in a pairwise similarity matrix (or a dissimilarity matrix).

At each step, the nearest two clusters are combined into a higher-level cluster. The distance between any two clusters A and B is taken to be the average of all distances between pairs of objects "x" in A and "y" in B, that is, the mean distance between elements of each cluster:

$$\frac{1}{|\mathcal{A}| \cdot |\mathcal{B}|} \sum_{x \in \mathcal{A}} \sum_{y \in \mathcal{B}} d(x,y)$$

The method is generally attributed to Sokal and Michener.[2] Fionn Murtagh found a time optimal $O(n^2)$ time algorithm to construct the UPGMA tree.[3]

See also   [edit]

- Neighbor-joining
- Cluster analysis
- Single-linkage clustering
- Complete-linkage clustering
- Hierarchical clustering
- Models of DNA evolution
- Molecular clock

References   [edit]

1. ^ Legendre, P. and Legendre, L. 1998. Numerical Ecology. Second English Edition. Developments in Environmental Modelling 20. Elsevier, Amsterdam.
2. ^ Sokal R and Michener C (1958). "A statistical method for evaluating systematic relationships". *University of Kansas Science Bulletin* **38**: 1409–1438.
3. ^ Murtagh F (1984). "Complexities of Hierarchic Clustering Algorithms: the state of the art". *Computational Statistics Quarterly* **1**: 101–113.

External links   [edit]

- UPGMA clustering algorithm implementation in Ruby (AI4R)
- Example calculation of UPGMA using a similarity matrix
- Example calculation of UPGMA using a distance matrix

v · t · e	Topics in phylogenetics	[hide]
Relevant fields	Computational phylogenetics · Molecular phylogenetics · Cladistics · Evolutionary taxonomy	
Basic concepts	Phylogenetic tree · Phylogenetic network · Long branch attraction · Clade vs Grade · Ghost lineage	
Inference methods	Maximum parsimony · Probabilistic methods (Maximum likelihood · Bayesian inference) · Distance-matrix methods (Neighbor-joining · **UPGMA** · Least squares) · Three-taxon analysis	
Current topics	PhyloCode · DNA barcoding · Molecular phylogenetics · Phylogenetic comparative methods · Phylogenetic network · Phylogenetic niche conservatism · Phylogenetics software · Phylogenomics · Phylogeography	

Group traits	Symplesiomorphy · Apomorphy · Synapomorphy · Autapomorphy
Group types	Monophyly · Paraphyly · Polyphyly
	List of evolutionary biology topics · Evolutionary biology

This bioinformatics-related article is a stub. You can help Wikipedia by expanding it.

Categories: Bioinformatics | Bioinformatics algorithms | Computational phylogenetics | Data clustering algorithms | Phylogenetics | Bioinformatics stubs