

Table of Contents

<i>TWINSpan (hierarchical divisive classification)</i>	1
---	----------

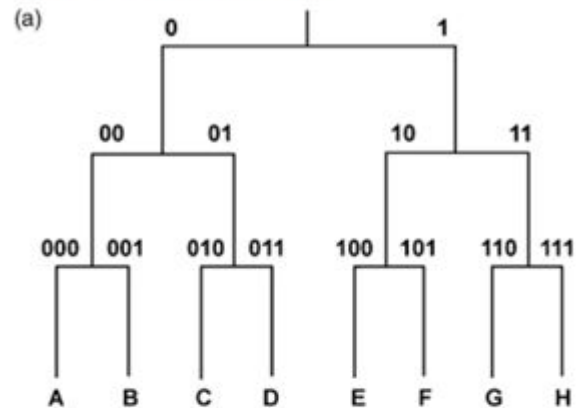
Section: Numerical classification

TWINSPAN (hierarchical divisive classification)

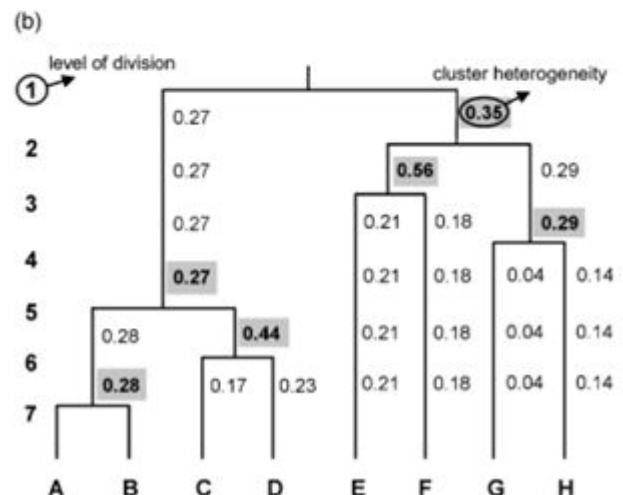
Theory R functions Examples Exercise 

Fig. 1

Original TWINSPAN:



Modified TWINSPAN:



TWINSPAN (abbreviation standing for Two-way indicator species analysis) is hierarchical and divisive method of numerical classification, which uses the results of ordination (namely CA) to divide the whole dataset into subdivisions. The method has been introduced by Mark O. Hill in 1979. It is not the only divisive algorithm in hand (others like DIANA or COINSPAN exist), but it is with no doubt far the most commonly used one. The algorithm itself is rather complex, and consist of the following steps: 1) ordination of samples along the first axis of correspondence analysis (CA1) and splitting the axis near the middle; 2) identify the indicator species which have high fidelity to each side (negative and positive) of the axis, and use them to further refine the classification of samples which are near the middle to avoid their misclassification; 3) take samples in each subdivision and apply steps 1 and 2 on them. Two stopping rules are applied to stop the division: minimum size of the subdivision (for example 5 – groups with five and fewer samples are not further divided) and the number of levels to which subdivision advances (for example 3 – only three levels of division are used). In each level of division, all groups of samples are divided (unless they are too small), which means that the number of resulting clusters is 2, 4, 8, 16, 32, ... 2^n for one, two, three, four, five ... n levels of divisions. A simple modification of the original algorithm allows the user to choose the desired number of clusters:

Fig. 2

The true algorithm is actually much more complex, and even the original description by Hill (1979) does not contain all details (some changes have been introduced later by other authors directly in the FORTRAN code of the TWINSPAN program). Perhaps the most detailed description of the algorithm with attention to some of the details is given in Kent (2012). Several software offer TWINSPAN (note that implementations in each of them actually slightly differ, since some are using a different version of the FORTRAN code): TWINSPAN for Windows, PC-ORD, CAP and JUICE. In R, I created a simple experimental package `twinspanR`, which is an R-wrapper around the `twinspan.exe` program and works only on Windows platform (this implementation includes both original and modified TWINSPAN algorithm).

From:

<https://anadat-r.davidzeleny.net/> - **Analysis of community ecology data in R**

Permanent link:

<https://anadat-r.davidzeleny.net/doku.php/en:hier-divisive?rev=1558491482>

Last update: **2019/05/22 10:18**