Illustration of How to Build a Similarity-Based Tree
Oscar Loureiro

Introduction

The objective of this document is to illustrate how to build a similarity-based tree summarizing the paths followed by the subjects who only tried two drug types before their baseline interview.

Definitions

The seven drug categories used in this analysis are:

1 = A (alcohol)
2 = M(r/e) (marijuana re/ex)
3 = M&PCP (marijuana with PCP)
4 = M&H/C (marijuana w/h or c)
5 = H/C(s) (heroin or cocaine, s)
7 = T (tranquilizers)
8 = E (ecstasy)

Data on Pathways of Drug Use

There were 11 people who tried just two drug types before baseline interview. 3.5 of them followed the path 1-2, 3.5 followed the path 2-1, 3.5 followed the path 2-3, and 0.5 followed the path 3-2. Note: The fractional frequencies were created by the tie breaking procedure.

The original tree summarizing all those paths is:

\[
\begin{align*}
&\rightarrow 1 \rightarrow 2 \ (\text{branch A}) \\
&\rightarrow 2 \rightarrow 1 \ (\text{branch B}) \\
&\rightarrow 2 \rightarrow 3 \ (\text{branch C}) \\
&\rightarrow 3 \rightarrow 2 \ (\text{branch D})
\end{align*}
\]

Each branch has associated a measure of the average outcome (e.g. number of drug uses in the last 30 days) for the subjects in that branch.

First Step in the Building of the Tree

There are four possible first steps that could be taken to build the similarity-based tree.
Possibility Number 1

If branches A and B are the ones closest in outcome, then they could be combined by pruning the tree to create the simpler tree

\[
\begin{align*}
& \rightarrow 1 \text{ or } 2 \\
\circ & \rightarrow 2 \rightarrow 3 \\
& \rightarrow 3 \rightarrow 2
\end{align*}
\]

Possibility Number 2

If branches B and C are the ones closest in outcome, then they could be combined by pruning to create the simpler tree

\[
\begin{align*}
& \rightarrow 2 \rightarrow 1 \text{ or } 3 \\
\circ & \rightarrow 1 \rightarrow 2 \\
& \rightarrow 3 \rightarrow 2
\end{align*}
\]

Possibility Number 3

If branches C and D are the ones closest in outcome, then they could be combined by pruning to create the simpler tree

\[
\begin{align*}
& \rightarrow 2 \text{ or } 3 \\
\circ & \rightarrow 1 \rightarrow 2 \\
& \rightarrow 2 \rightarrow 1
\end{align*}
\]

Possibility Number 4

If branches A and D are the ones closest in outcome, then they could be combined by pruning to create the simpler tree

\[
\begin{align*}
& \rightarrow 1 \text{ or } 3 \rightarrow 2 \\
\circ & \rightarrow 2 \rightarrow 3 \\
& \rightarrow 2 \rightarrow 1
\end{align*}
\]