comparing.Partitions(clusterSim)

Comparing two partitions

Partitions		Partition $P^{(t)}$				Sums
	cluster	$P_1^{(t)}$	$P_2^{(t)}$	...	$P_v^{(t)}$	
Partition $P^{(q)}$	$P_1^{(q)}$	n_{11}	n_{12}	...	n_{1v}	$n_{1\bullet}$
	$P_2^{(q)}$	n_{21}	n_{22}	...	n_{2v}	$n_{2\bullet}$
	⋮	⋮	⋮	⋮	⋮	⋮
	$P_u^{(q)}$	n_{u1}	n_{u2}	...	n_{uv}	$n_{u\bullet}$
Sums		$n_{\bullet 1}$	$n_{\bullet 2}$	...	$n_{\bullet v}$	$n_{\bullet\bullet}=n$

where: $P^{(t)}$, $P^{(q)}$ – partitions t (q) of a finite set of objects A,

$n_{\bullet\bullet}=n$ – number of objects,

n_{sr} – number of objects belonging simultaneously to clusters r and s,

$r=1,\ldots,v$ ($s=1,\ldots,u$) – cluster number in partition $P^{(t)}$ ($P^{(q)}$),

$v(u)$ – number of clusters in partition $P^{(t)}(P^{(q)})$,

$n_{\bullet r}$ – number of objects in cluster $P_r^{(t)}$ (column r),

$n_{s\bullet}$ – number of objects in cluster $P_s^{(q)}$ (row s).

Rand index

$$R=Z\Big/\binom{n}{2}=1-N\Big/\binom{n}{2},\quad R\in[0;1],$$

where: $Z=\binom{n}{2}+\sum_{s=1}^{u}\sum_{r=1}^{v}n_{sr}^2-\frac{1}{2}\left(\sum_{s=1}^{u}n_{s\bullet}^2+\sum_{r=1}^{v}n_{\bullet r}^2\right)$,

$N=\frac{1}{2}\left(\sum_{s=1}^{u}n_{s\bullet}^2+\sum_{r=1}^{v}n_{\bullet r}^2\right)-\sum_{s=1}^{u}\sum_{r=1}^{v}n_{sr}^2$

Corrected Rand index (Hubert & Arabie [1985], p. 198)

$$R_{HA}=\frac{\sum_{r,s}\binom{n_{rs}}{2}-\sum_{r}\binom{n_{\bullet r}}{2}\sum_{s}\binom{n_{s\bullet}}{2}\Big/\binom{n}{2}}{\frac{1}{2}\left[\sum_{r}\binom{n_{\bullet r}}{2}+\sum_{s}\binom{n_{s\bullet}}{2}\right]-\sum_{r}\binom{n_{\bullet r}}{2}\sum_{s}\binom{n_{s\bullet}}{2}\Big/\binom{n}{2}},\quad R_{HA}\in[-\infty;1]$$

Nowak index

$$S=\frac{1}{v+u}\left(\sum_{s=1}^{u}\max_{r}\{k_{sr}\}+\sum_{r=1}^{v}\max_{s}\{k_{sr}\}\right),\quad S\in[1/n;1],$$

where: $k_{sr}=\dfrac{n_{sr}}{\max\{n_{s\bullet};n_{\bullet r}\}}$.

References

Hubert, L., Arabie, P. (1985), *Comparing partitions*, "Journal of Classification", no. 1, 193-218.

Nowak, E. (1985), *Wskaznik podobienstwa wynikow podziałow*, "Przeglad Statystyczny" ["Statistical Review"], no. 1, 41-48.

Rand, W.M. (1971), *Objective criteria for the evaluation of clustering methods*, "Journal of the American Statistical Association", no. 336, 846-850.