Understanding external validity indices

Matthijs J Warrens
&
Hanneke van der Hoef

IFCS Tokyo
August 10, 2017

Comparing partitions

- two partitions U and U' of same set of objects
- matching table $\mathbf{N} = \{n_{ij}\}$ of size $I \times J$ where n_{ij} is number of objects in $U_i \in U$ and in $U'_j \in U'$

Example of $\mathbf{N}$ of size 3×3:

	U'_1	U'_2	U'_3	total
U_1	n_{11}	n_{12}	n_{13}	n_{1+}
U_2	n_{21}	n_{22}	n_{23}	n_{2+}
U_3	n_{31}	n_{32}	n_{33}	n_{3+}
total	n_{+1}	n_{+2}	n_{+3}	n

External validity indices

Summarize matching table

- information in matching table may be complex
- convenient to summarize information with numbers

Three approaches

- counting pairs of objects
- information theory (mutual information, entropy)
- matching sets

This talk

- analysis of indices based on **counting pairs**
- what information do these indices reflect?

This talk

Analysis of indices

- express indices in terms of cluster information
- weighted averages of cluster indices

Three index families (prototypical examples)

- Dice (1945), Wallace (1983)
- adjusted Rand index (Hubert and Arabie 1985)
- Rand (1971) index

What information do these indices reflect?

- agreement on large clusters
- indices have a very limited usefulness

Building blocks

Total number of object pairs (with n objects):

$$N = \binom{n}{2}$$

Number of objects pairs in U_i and U:

$$P_i = \binom{n_{i+}}{2} \qquad \text{and} \qquad P = \sum_{i=1}^{I} P_i$$

Number of objects pairs in U'_j and U':

$$P'_j = \binom{n_{+j}}{2} \qquad \text{and} \qquad P' = \sum_{j=1}^{J} P'_j$$

Building blocks

Proportion of object pairs in U_i also joined in U':

$$w_i = \frac{\sum_{j=1}^{J} \binom{n_{ij}}{2}}{P_i}$$

	U_1'	U_2'	U_3'	total
U_1	102	0	0	102
U_2	0	15	10	25
U_3	0	10	15	25
total	102	25	25	152

For example:

$$w_1 = \left[\binom{102}{2} + \binom{0}{2} + \binom{0}{2}\right] \Big/ \binom{102}{2} = 1$$

$$w_2 = w_3 = \left[\binom{0}{2} + \binom{15}{2} + \binom{10}{2}\right] \Big/ \binom{25}{2} = .50$$

Building blocks

Proportion of object pairs in U_i also joined in U':

$$w_i = \frac{\sum_{j=1}^{J} \binom{n_{ij}}{2}}{P_i}$$

adjusted (Warrens 2008a,b):

$$Aw_i = \frac{N \sum_{j=1}^{J} \binom{n_{ij}}{2} - P_i P'}{P_i(N - P')}$$

Proportion of object pairs in U'_j also joined in U:

$$w'_j = \frac{\sum_{i=1}^{I} \binom{n_{ij}}{2}}{P'_j}$$

adjusted:

$$Aw'_j = \frac{N \sum_{i=1}^{I} \binom{n_{ij}}{2} - P'_j P}{P'_j(N - P)}$$

Dice and Wallace indices

Dice (1945) index

$$D = \frac{\sum_{i=1}^{I} w_i P_i + \sum_{j=1}^{J} w'_j P'_j}{\sum_{i=1}^{I} P_i + \sum_{j=1}^{J} P'_j}$$

Basic building blocks are Wallace (1983) indices

$$W = \frac{\sum_{i=1}^{I} w_i P_i}{\sum_{i=1}^{I} P_i} \quad \text{and} \quad W' = \frac{\sum_{j=1}^{J} w'_j P'_j}{\sum_{j=1}^{J} P'_j}$$

W is proportion of object pairs in U also joined in U'

Family 1: indices that are functions of W and W'

adjusted Rand index

Hubert and Arabie (1985)

$$AR = \frac{\sum_{i=1}^{I} Aw_i P_i + \sum_{j=1}^{J} Aw'_j P'_j}{\sum_{i=1}^{I} P_i + \sum_{j=1}^{J} P'_j}$$

Basic blocks are adjusted Wallace indices (Severiano et al. 2011)

$$AW = \frac{\sum_{i=1}^{I} Aw_i P_i}{\sum_{i=1}^{I} P_i} \quad \text{and} \quad AW' = \frac{\sum_{j=1}^{J} Aw'_j P'_j}{\sum_{j=1}^{J} P'_j}$$

AW is adjusted version of W

Family 2: indices that are functions of AW and AW'

Rand (1971) index

$$R = \frac{\sum_{i=1}^{I} w_i P_i + \sum_{j=1}^{J} w'_j P'_j + V(N-P) + V'(N-P')}{\sum_{i=1}^{I} P_i + \sum_{j=1}^{J} P'_j + N - P + N - P'}$$

where $T = \sum_{i=1}^{I} \sum_{j=1}^{J} \binom{n_{ij}}{2}$,

$$V = \frac{N+T-P-P'}{N-P} \quad \text{and} \quad V' = \frac{N+T-P-P'}{N-P'}$$

V is proportion object pairs not together in U also not in U'

Family 3: indices that are functions of W, W', V and V'

Two examples

	U_1'	U_2'	U_3'
U_1	102	0	0
U_2	0	15	10
U_3	0	10	15

indices
$D = .95$
$AR = .90$
$R = .95$

indices high ($\geq .90$): high agreement?
perfect agreement on large cluster, low agreement on small clusters

	U_1'	U_2'	U_3'
U_1	52	48	0
U_2	46	54	0
U_3	0	0	10

indices
$D = .50$
$AR = .08$
$R = .55$

indices low ($.08 - .55$): low agreement?
perfect agreement on small cluster, low agreement on large clusters

Limited usefulness

Pair-counting indices

- functions of cluster indices
- reflect agreement on large clusters
- do not reflect agreement on small clusters

To be useful

Requirement: cluster sizes must be equal

- otherwise, indices reflect agreement on large clusters only
- Romano et al. (2016) have same conclusion for AR

Cluster size sensitivity

Unexplored idea
for removing sensitivity to cluster size imbalance

Instead of weighted averages

$$D = \frac{\sum_{i=1}^{I} w_i P_i + \sum_{j=1}^{J} w'_j P'_j}{\sum_{i=1}^{I} P_i + \sum_{j=1}^{J} P'_j} \quad \text{and} \quad AR = \frac{\sum_{i=1}^{I} Aw_i P_i + \sum_{j=1}^{J} Aw'_j P'_j}{\sum_{i=1}^{I} P_i + \sum_{j=1}^{J} P'_j}$$

use normal averages of cluster indices

$$D^* = \frac{1}{2I}\sum_{i=1}^{I} w_i + \frac{1}{2J}\sum_{j=1}^{J} w'_j \quad \text{and} \quad AR^* = \frac{1}{2I}\sum_{i=1}^{I} Aw_i + \frac{1}{2J}\sum_{j=1}^{J} Aw'_j$$

Discussion

Pair-counting indices

- functions of cluster indices
- limited usefulness

Alternative agreement analysis

- normal instead of weighted averages
- information-theoretic indices (Hanneke van der Hoef)
- set matching indices
- analyze agreement on the cluster level

References

Hubert LJ, Arabie P (1985) Comparing partitions. Journal of Classification 2:193-218

Rand WM (1971) Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66:846-850

Romano S, Vinh NX, Bailey J, Verspoor K (2016) Adjusting for chance clustering comparison measures. Journal of Machine Learning Research 17:1-32

Severiano A, Pinto FR, Ramirez M, Carriço JA (2011) Adjusted Wallace coefficient as a measure of congruence between typing methods. Journal of Clinical Microbiology 49:3997-4000

Wallace DL (1983) Comment on A method for comparing two hierarchical clusterings. Journal of the American Statistical Association 78:569-576

Warrens MJ (2008a) On the equivalence of Cohen's kappa and the Hubert-Arabie adjusted Rand index. Journal of Classification 25:177-183

Warrens MJ (2008b) On similarity coefficients for 2×2 tables and correction for chance. Psychometrika 73:487-502