

Flynt, Abby; Dean, Nema; Nugent, Rebecca

sARI: a soft agreement measure for class partitions incorporating assignment probabilities.

(English) [Zbl 1474.62224](#)

Adv. Data Anal. Classif., ADAC 13, No. 1, 303-323 (2019).

Summary: Agreement indices are commonly used to summarize the performance of both classification and clustering methods. The easy interpretation/intuition and desirable properties that result from the Rand and adjusted Rand indices, has led to their popularity over other available indices. While more algorithmic clustering approaches like k -means and hierarchical clustering produce hard partition assignments (assigning observations to a single cluster), other techniques like model-based clustering include information about the certainty of allocation of objects through class membership probabilities (soft partitions). To assess performance using traditional indices, e.g., the adjusted Rand index (ARI), the soft partition is mapped to a hard set of assignments, which commonly overstates the certainty of correct assignments. This paper proposes an extension of the ARI, the soft adjusted Rand index (sARI), with similar intuition and interpretation but also incorporating information from one or two soft partitions. It can be used in conjunction with the ARI, comparing the similarities of hard to soft, or soft to soft partitions to the similarities of the mapped hard partitions. Simulation study results support the intuition that in general, mapping to hard partitions tends to increase the measure of similarity between partitions. In applications, the sARI more accurately reflects the cluster boundary overlap commonly seen in real data.

MSC:

62H30 Classification and discrimination; cluster analysis (statistical aspects)

91C20 Clustering in the social and behavioral sciences

Keywords:

adjusted rand index; model-based clustering; mixture models; soft partition; posterior probabilities; class membership probabilities

Software:

clusterGeneration; mclust; R

Full Text: [DOI Link](#)

References:

- [1] Amodio S, D'Ambrosio A, Iorio C, Siciliano R (2015) Adjusted concordance index, an extension of the adjusted rand index to fuzzy partitions. arXiv preprint arXiv:1509.00803
- [2] Banfield JD, Raftery AE (1993) Model-based Gaussian and non-Gaussian clustering. *Biometrics* 49(3): 803-821 · [Zbl 0794.62034](#)
- [3] Bezdek JC (1981) Objective function clustering. In: *Pattern recognition with fuzzy objective function algorithms*. Springer, Boston, MA, pp 43-93
- [4] Campello, RJGB, A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment, *Pattern Recognit Lett*, 28, 833-841, (2007) · [doi:10.1016/j.patrec.2006.11.010](#)
- [5] Dempster, AP; Laird, NM; Rubin, DB, Maximum likelihood from incomplete data via the EM algorithm, *J R Stat Soc Ser B (Methodol)*, 39, 1-38, (1977) · [Zbl 0364.62022](#)
- [6] Downton, M.; Brennan, T., Comparing classifications: an evaluation of several coefficients of partition agreement, *Classif Soc Bull*, 4, 53-54, (1980)
- [7] Dunn, JC, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *J Cybern*, 3, 32-57, (1973) · [Zbl 0291.68033](#) · [doi:10.1080/01969727308546046](#)
- [8] Fowlkes, EB; Mallows, CL, A method for comparing two hierarchical clusterings, *J Am Stat Assoc*, 78, 553-569, (1983) · [Zbl 0545.62042](#) · [doi:10.1080/01621459.1983.10478008](#)
- [9] Fraley, C.; Raftery, AE, Model-based clustering, discriminant analysis, and density estimation, *J Am Stat Assoc*, 97, 611-631, (2002) · [Zbl 1073.62545](#) · [doi:10.1198/016214502760047131](#)
- [10] Fraley, C.; Raftery, AE, Model-based methods of classification: using the mclust software in chemometrics, *J Stat Softw*, 18, 1-13, (2007) · [doi:10.18637/jss.v018.i06](#)

- [11] Hartigan JA (1975) Clustering algorithms. Wiley, New York · [Zbl 0372.62040](#)
- [12] Hubert, L.; Arabie, P., Comparing partitions, *J Classif*, 2, 193-218, (1985) · [Zbl 0587.62128](#) · [doi:10.1007/BF01908075](#)
- [13] Huellermeyer, E.; Rifqi, M.; Henzgen, S.; Senge, R., Comparing fuzzy partitions: a generalization of the Rand index and related measures, *IEEE Trans Fuzzy Syst*, 20, 546-556, (2012) · [doi:10.1109/TFUZZ.2011.2179303](#)
- [14] Jaccard, P., Étude comparative de la distribution florale dans une portion des alpes et du jura, *Bull de la Société Vaudoise des Sciences Naturelles*, 37, 547-579, (1901)
- [15] MacQueen J (1967) Some methods for classification and analysis of multivariate observations. In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA 1:281-297 · [Zbl 0214.46201](#)
- [16] McLachlan G, Krishnan T (2007) The EM algorithm and extensions, vol 382. Wiley, New York · [Zbl 0882.62012](#)
- [17] McLachlan G, Peel D (2004) Finite mixture models. Wiley, New York · [Zbl 0963.62061](#)
- [18] McNicholas, PD, Model-based clustering, *J Classif*, 33, 331-373, (2016) · [Zbl 1364.62155](#) · [doi:10.1007/s00357-016-9211-9](#)
- [19] Miyamoto S, Ichihashi H, Honda K (2008) Algorithms for fuzzy clustering. Springer, Berlin · [Zbl 1147.68073](#)
- [20] Morey, LC; Agresti, A., The measurement of classification agreement: an adjustment to the Rand statistic for chance agreement, *Educ Psychol Meas*, 44, 33-37, (1984) · [doi:10.1177/0013164484441003](#)
- [21] Qiu, W.; Joe, H., Separation index and partial membership for clustering, *Comput Stat Data Anal*, 50, 585-603, (2006) · [Zbl 1431.62270](#) · [doi:10.1016/j.csda.2004.09.009](#)
- [22] Qiu W, Joe H (2015) clusterGeneration: random cluster generation (with specified degree of separation). R package version 1.3.4. <https://CRAN.R-project.org/package=clusterGeneration>
- [23] R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
- [24] Rand, WM, Objective criteria for the evaluation of clustering methods, *J Am Stat Assoc*, 66, 846-850, (1971) · [doi:10.1080/01621459.1971.10482356](#)
- [25] Scrucca, L.; Fop, M.; Murphy, TB; Raftery, AE, mclust 5: Clustering, classification and density estimation using gaussian finite mixture models, *R J*, 8, 289, (2016) · [doi:10.32614/RJ-2016-021](#)
- [26] Steinley, D., Properties of the Hubert-Arabie adjusted Rand index, *Psychol Methods*, 9, 386, (2004) · [doi:10.1037/1082-989X.9.3.386](#)
- [27] Ward, JH, Hierarchical grouping to optimize an objective function, *J Am Stat Assoc*, 58, 236-244, (1963) · [doi:10.1080/01621459.1963.10500845](#)
- [28] Wolfe JH (1963) Object cluster analysis of social areas. Ph.D. thesis, University of California

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.