

On the Use of the Adjusted Rand Index as a Metric for Evaluating Supervised Classification

Jorge M. Santos¹ and Mark Embrechts²

¹ ISEP - Instituto Superior de Engenharia do Porto, Portugal

² Rensselaer Polytechnic Institute, Troy, New York, USA

jms@isep.ipp.pt, embrem@rpi.edu

Abstract. The Adjusted Rand Index (ARI) is frequently used in cluster validation since it is a measure of agreement between two partitions: one given by the clustering process and the other defined by external criteria. In this paper we investigate the usability of this clustering validation measure in supervised classification problems by two different approaches: as a performance measure and in feature selection. Since ARI measures the relation between pairs of dataset elements not using information from classes (labels) it can be used to detect problems with the classification algorithm specially when combined with *conventional* performance measures. Instead, if we use the class information, we can apply ARI also to perform feature selection. We present the results of several experiments where we have applied ARI both as a performance measure and for feature selection showing the validity of this index for the given tasks.

1 Introduction

One of the main difficulties in classification problems consists on the correct evaluation of the classifier performance. This is usually done by applying a common performance measure like the Mean Squared Error (MSE) or the Classification Correct Rate. Other measures like AUC (area in percentage under the Receiver Operating Characteristic (ROC) curve), Sensitivity and Specificity, are also used specially for two class problems like those involving medical applications. All these measures *compare* the labeled outcome of the supervised classification algorithm with the known labeled targets. By doing this they evaluate how good the algorithm has labeled the input data according to the required target labels. This can lead to poor results derived only by the fact that the output labels could be switched even if the classes are well identified. In these cases we deemed useful the introduction of a measure that can evaluate how well the algorithm split the input data in different classes by looking at the relation between elements of each class and not to the given labels. This is the main reason for our proposal of using a clustering validation measure in supervised classification problems.

Usually, as we will show on the experiments, the ARI performs in a similar way as other common measures. Lower values for bad classification results and higher values for good classification results. We advise to include ARI in the set of

performance measures usually used on the evaluation of supervised classification algorithms.

Since ARI is a measure of agreement between partitions and the target data is partitioned by means of the labeling we can also use ARI to perform feature selection if we split each feature in non-overlapping equal intervals and compare the partition derived from the split with the one given by the targets. By doing this we are evaluating each feature’s discriminant power and we can rank the features according to the computed ARI value. We can then select the most discriminant features to apply in our classification algorithm.

This work is organized as follows: the next section introduces the Adjusted Rand Index; Section 3 explains how we intend to use ARI as a performance measure for supervised classification problems and for feature selection; Section 4 presents several experiments that show the applicability of the proposed measure with results detailed in Section 5. In the final section we draw some conclusions about the paper.

2 The Adjusted Rand Index

There are several performance indices for cluster evaluation. Indices are measures of correspondence between two partitions of the same data and are based on how pairs of objects are classified in a contingency table.

Consider a set of n objects $S = \{O_1, O_2, ..., O_n\}$ and suppose that $U = \{u_1, u_2, ..., u_R\}$ and $V = \{v_1, v_2, ..., v_C\}$ represent two different partitions of the objects in S such that $\cup_{i=1}^R u_i = S = \cup_{j=1}^C v_j$ and $u_i \cap u_{i'} = \emptyset = v_j \cap v_{j'}$ for $1 \leq i \neq i' \leq R$ and $1 \leq j \neq j' \leq C$. Given two partitions, U and V , with R and C subsets, respectively, the contingency Table 1 can be formed to indicate group overlap between U and V .

Table 1. Contingency Table for Comparing Partitions U and V

Partition		V				Total
Group		v_1	v_2	\cdots	v_C	
U	u_1	t_{11}	t_{12}	\cdots	t_{1C}	$t_{1.}$
	u_2	t_{21}	t_{22}	\cdots	t_{2C}	$t_{2.}$
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	u_R	t_{R1}	t_{R2}	\cdots	t_{RC}	$t_{R.}$
	Total	$t_{.1}$	$t_{.2}$	\cdots	$t_{.C}$	$t_{..} = n$

In Table 1, a generic entry, t_{rc} , represents the number of objects that were classified in the r th subset of partition R and in the c th subset of partition C . From the total number of possible combinations of pairs $\binom{n}{2}$ from a given set we can represent the results in four different types of pairs: