Computation of the AUC in The Context Of Classification Association Rule Mining

Frans Coenen[1] and Christain Setzkorn[2]
17 November 2010
[1]Department of Computer Science, The University of Liverpool.
[2]School of Veterinary Science, University of Liverpool and National Center for Zoonosis Research

1. Overview

A Receiver Operating Curve (ROC) is a tool used to summarise the performance of *cost sensitive* or *conditional* binary classifiers where each record may have a "class probability" associated with it (the probability that the record belongs to a class x). Examples of such classifiers include Probability Estimation Trees (Zhang et al., 2006; Sulzmann and Fürnkranz, 2009) and Naïve Bayes (Qin, 2006) and other forms of classification tree generators (Mease *et al*., 2007). The fundamental idea behind ROC analysis was that these probabilities (rankings) should be taken into account when determining, for comparisons purposes, the operation of classifiers (instead of using a simple accuracy measure). Subsequently, a view has been promoted (Huang and Ling, 2005; Lavraˇc *et al*., 1999) that ROC analysis has general applicability for determining the effectiveness of classifiers (not just classifiers which produce rankings), and that it is a better overall measure than using simple accuracy. This report is concerned with the application of ROC analysis for classifiers built using Classification Association Rule Mining (CARM) techniques. Although initially intended for binary classifiers, the ROC concept has been extended to the multi-class classification problem (see for example Hand and Till, 2001).

A ROC is actually a plot recording False Positive Rates (X-axis) against True Positive Rates (Y-Axis) for a sequence of class pairings. The Area Under the Curve (AUC) indicates the accuracy of a classification. The AUC will be 1 given 100% accuracy, and 0 given 0% accuracy. AUC can be estimated, in the context of multi-class classification, as follows (Hand and Till, 2001).

$$AUC = \frac{2}{c(c-1)} \sum_{i<j} A(i,j)$$

Where c is the number of classes, i and j are class numbers, and A is calculated as follows:

$$A(i,j) = \frac{MWW(i \mid j) + MWW(j \mid i)}{2}$$

MWW is the Man-Whitney-Wilcoxon statistic (or rank sum). This is calculated by first drawing up a MWW ranked table comprising two columns (sometimes referred to as vectors). The first column is the *response* column (R), and the second the *signal* column (S) values. The rows are number from 1 to N where N is the number records to be considered with respect to the MWWW calculation (see examples below). The ranking is a follows (in descending order): true positives (R_i=1, S_i=1), false negatives (R_i=1, S_i=0), true negatives (R_i=0, S_i=0), false positives (R_i=0, S_i=1). The calculation is then as follows:

$$MWW = \frac{s - \frac{n1(n1+1)}{2}}{n1n2}$$

Where s is the sum of the rankings of the single values (column S); and, in the case of classifiers built using CARM, n1 is the sum of the response values (1s) in the signal column values and n2 is the sum of the noise values (0s) in the signal column. In the case of CARM responses can be *signal values* or *noise values*, 1 or 0. Signal values (1) are given a higher ranking than noise values (0). The calculation is then as follows:

This report is directed at the application of ROC analysis to rule based classifiers where classification rules are applied to examples which are then classified as belonging to a particular class. Examples of such classifiers are Classification Association Rule Miners such as CMAR, CPAR and TFPC. And rule induction systems such as FOIL and RIPPER. In this case the probabilities associated with the classification are 1 or 0 (the example does belong to class X or it does not). In this case n1 is the number of 1s recorded in the signal column, and n2 is the number of 0s recorded in the signal column.

2. Example One (100% Accurate Classifier)

Considering the data set, split over three classes (c1, c2 and c3), given in Table 1; and a classifier which is 100% accurate. This will produce a prediction table of the form given in Table 2.

Record Num	c1	c2	c3
1	1	0	0
2	1	0	0
3	1	0	0
4	0	1	0
5	0	1	0
6	0	0	1
7	0	0	1
8	0	0	1

Table 1. Example data set (“Truth Values”)

Record Num	c1	c2	c3
1	1	0	0
2	1	0	0
3	1	0	0
4	0	1	0
5	0	1	0
6	0	0	1
7	0	0	1
8	0	0	1

Table 2. Predictions for Example 1

To determine the AUC calculation for this classifier we will first draw up MMW tables for all the possible pair-wise permutaions of the class: MWW(1,2), MWW(2,1), MWW(1,3), MWW(3,1), MWW(2,3) and MWW(3,2). Let us consider MWW(1,2) first. The MWW table is given in Table 3. The table only considers those records that should be classified as c1 or c2. Three records were classified as c1 and two as not c1. The response vector, with respect to c1, is therefore {0,0,1,1,1}. Note that the three c1 classifications are given the highest ranking. The signal vector, the “ground-truth” vector, is also {0,0,1,1,1} in this case because the classifier was 100% accurate. Thus, with respect to MWW(1,2) n1 and n2 are both 3, and S = 3+4+5 =12. Thus:

$$MWW(1|2) = \frac{12 - \frac{3(3+1)}{2}}{3 \times 2} = \frac{12-6}{6} = 1$$

Rank	Res-ponse	Sig-nal
1	0	0
2	0	0
3	1	1

Rank	Res-ponse	Sig-nal
1	0	0
2	0	0
3	0	0

Rank	Res-ponse	Sig-nal
1	0	0
2	0	0
3	0	0

4	1	1
5	1	1

Table 3. MWW(1|2)

4	1	1
5	1	1

Table 4. MWW(2|1)

4	1	1
5	1	1
6	1	1

Table 5. MWW(1|3)

Rank	Res-ponse	Sig-nal
1	0	0
2	0	0
3	0	0
4	1	1
5	1	1
6	1	1

Table 6. MWW(3|1)

Rank	Res-ponse	Sig-nal
1	0	0
2	0	0
3	0	0
4	1	1
5	1	1

Table 7. MWW(2|3)

Rank	Res-ponse	Sig-nal
1	0	0
2	0	0
3	1	1
4	1	1
5	1	1

Table 8. MWW(3|2)

If we now consider MWW(2|1) the MWW table will be as shown in Table 4. In tis case n1=2, n2=2 and s=4+5. Thus:

$$MWW(2|1) = \frac{9 - \frac{2(2+1)}{2}}{2 \times 3} = \frac{9-3}{6} = 1$$

A is then:

$$A(1,2) = \frac{1+1}{2} = 1$$

Calculating MWW(1|3) as per Table 5:

$$MWW(1|3) = \frac{15 - \frac{3(3+1)}{2}}{3 \times 3} = \frac{15-6}{9} = 1$$

and MWW(3|1) as per Table 6:

$$MWW(3|1) = \frac{15 - \frac{3(3+1)}{2}}{3 \times 3} = \frac{15-6}{9} = 1$$

A is then 1. Doing the same for MWW(2|3) amd MWW(3|2) (Table 7 and 8) then gives us MWW(2|3)=1 and MWW(3|2)=1; and A is again 1. The AUC in this case is then:

$$AUC = \frac{2}{3(3-1)}\left(1+1+1\right) = \frac{2}{6} \times 3 = 1$$

Indicating that the classifier is 100% accurate.

3. Example Two (0% Accurate Classifier)

If we now consider a classifier that is 0% accurate. A possible prediction table is given in Table 10 (for comparison purposes Table 1 is repeated in Table 9). The associated MWW tables are given in Tables 11 to16. The MWW calculations are presented in Table 17.

Record Num	c1	c2	c3
1	1	0	0
2	1	0	0
3	1	0	0
4	0	1	0
5	0	1	0
6	0	0	1
7	0	0	1
8	0	0	1

Table 9. Example Data Set

Record Num	c1	c2	c3
1	0	1	0
2	0	0	1
3	0	1	0
4	1	0	0
5	0	0	1
6	1	0	0
7	0	1	0
8	1	0	0

Table 10. Prediction Values for Example 2

Rank	Rec. Num	Res-ponse	Sig-nal
1	1	0	1
2	2	0	1
3	3	0	1
4	5	0	0
5	4	1	0

Table 11. MWW(1|2)

Rank	Rec. Num	Res-ponse	Sig-nal
1	4	0	1
2	5	0	1
3	2	0	0
4	1	1	0
5	3	1	0

Table 12. MWW(2|1)

Rank	Rec. Num	Res-ponse	Sig-nal
1	1	0	1
2	2	0	1
3	3	0	1
4	7	0	0
5	6	1	0
6	8	1	0

Table 13. MWW(1|3)

Rank	Rec. Num	Res-ponse	Sig-nal
1	6	0	1
2	7	0	1
3	8	0	1
4	1	0	0
5	3	0	0
6	2	1	0

Table 14. MWW(3|1)

Rank	Rec. Num	Res-ponse	Sig-nal
1	4	0	1
2	5	0	1
3	6	0	0
4	8	0	0
5	7	1	0

Table 15. MWW(2|3)

Rank	Rec. Num	Res-ponse	Sig-nal
1	6	0	1
2	7	0	1
3	8	0	1
4	4	0	0
5	5	1	0

Table 16. MWW(3|2)

Pairing	**s**	**n1**	**n2**	**MMW**
(1|2)	6	3	2	$\frac{6-\frac{3(3+1)}{2}}{3\times 2}=\frac{6-6}{6}=\frac{0}{6}=0.0$
(2|1)	3	2	3	$\frac{3-\frac{2(2+1)}{2}}{2\times 3}=\frac{3-3}{6}=\frac{0}{6}=0.0$
(1|3)	6	3	3	$\frac{6-\frac{3(3+1)}{2}}{3\times 3}=\frac{6-6}{9}=\frac{0}{9}=0.0$

(3|1)	6	3	3	$\frac{6-\frac{3(3+1)}{2}}{3\times 3}=\frac{6-6}{9}=\frac{0}{9}=0.0$
(2|3)	3	2	3	$\frac{3-\frac{2(2+1)}{2}}{2\times 3}=\frac{3-3}{6}=\frac{0}{6}=0.0$
(3|2)	6	3	2	$\frac{6-\frac{3(3+1)}{2}}{3\times 2}=\frac{6-6}{6}=\frac{0}{6}=0.0$

Table 17. MMW calculations for Example 2

The “A” calculations are then:

$$A(1,2)=\frac{0.0+0.0}{2}=0.0$$

$$A(1,3)=\frac{0.0+0.0}{2}=0.0$$

$$A(2,3)=\frac{0.0+0.0}{2}=0.0$$

which thus gives an AUC value (as expected) of:

$$AUC=\frac{2}{3(3-1)}\left(0.0+0.0+0.0\right)=\frac{2}{6}\times 0.0=0.0$$

4. Example Three (50% Accurate Classifier)

If we now consider a classifier that is 50% accurate. A possible prediction table is given in Table 19 (for comparison purposes Table 1 is again repeated in Table 18). The associated MWW tables are given in Tables 20 to 25. The MWW calculations are presented in Table 26.

Record Num	c1	c2	c3
1	1	0	0
2	1	0	0
3	1	0	0
4	0	1	0
5	0	1	0
6	0	0	1
7	0	0	1
8	0	0	1

Table 18. Example data set

Record Num	c1	c2	c3
1	1	0	0
2	0	1	0
3	1	0	0
4	1	0	0
5	0	1	0
6	0	1	0
7	0	0	1
8	1	0	0

Table 19. Predictions for Example 3

Rank	Rec.	Res-	Sig-

Rank	Rec.	Res-	Sig-

Rank	Rec.	Res-	Sig-

	Num	ponse	nal
1	2	0	1
2	5	0	0
3	4	1	0
4	1	1	1
5	3	1	1

Table 11. MWW(1|2)

	Num	ponse	nal
1	4	0	1
2	1	0	0
3	3	0	0
4	2	1	0
5	5	1	1

Table 12. MWW(2|1)

	Num	ponse	nal
1	2	0	1
2	6	0	0
3	7	0	0
4	8	1	0
5	1	1	1
6	3	1	1

Table 13. MWW(1|3)

Rank	Rec. Num	Res-ponse	Sig-nal
1	6	0	1
2	8	0	1
3	1	0	0
4	2	0	0
5	3	0	0
6	7	1	1

Table 14. MWW(3|1)

Rank	Rec. Num	Res-ponse	Sig-nal
1	4	0	1
2	7	0	0
3	8	0	0
4	6	1	0
5	5	1	1

Table 15. MWW(2|3)

Rank	Rec. Num	Res-ponse	Sig-nal
1	6	0	1
2	8	0	1
3	4	0	0
4	5	0	0
5	7	1	1

Table 16. MWW(3|2)

Pairing	**s**	**n1**	**n2**	**MMW**
$(1 \mid 2)$	10	3	2	$\frac{10-\frac{3(3+1)}{2}}{2\times 3}=\frac{10-6}{6}=\frac{4}{6}=0.667$
$(2 \mid 1)$	6	2	3	$\frac{6-\frac{2(2+1)}{2}}{2\times 3}=\frac{6-3}{6}=\frac{3}{6}=0.5$
$(1 \mid 3)$	12	3	3	$\frac{12-\frac{3(3+1)}{2}}{3\times 3}=\frac{12-6}{9}=\frac{6}{9}=0.667$
$(3 \mid 1)$	9	3	3	$\frac{9-\frac{3(3+1)}{2}}{3\times 3}=\frac{9-6}{9}=\frac{3}{9}=0.333$
$(2 \mid 3)$	6	2	3	$\frac{6-\frac{2(2+1)}{2}}{2\times 3}=\frac{6-3}{6}=\frac{3}{6}=0.5$
$(3 \mid 2)$	8	3	2	$\frac{8-\frac{3(3+1)}{2}}{3\times 2}=\frac{8-6}{6}=\frac{2}{6}=0.333$

Table 17. MMW calculations for Example 3

The “A” calculations are then:

$$A(1,2)=\frac{0.667+0.5}{2}=0.585$$

$$A(1,3) = \frac{0.667 + 0.333}{2} = 0.5$$

$$A(2,3) = \frac{0.5 + 0.333}{2} = 0.417$$

which gives an AUC value of:

$$AUC = \frac{2}{3(3-1)}\left(0.585 + 0.5 + 0.417\right) = \frac{2}{6} \times 1.501 = 0.5$$

References

1. Hand, D.J. and Till, R.J. (2001). A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning, 45, pp171–186.
2. Huang, J. and Ling, C.X. (2005). Using AUC and Accuracy in Evaluating Learning Algorithms. IEEE Transactions on Knowledge and data Engineering, 17(3), pp299-310.
3. Lavraˇc, N., Flach, P., and Zupan, B. (1999). Rule Evaluation Measures: A Unifying View. Proc. ILP-99, Springer LNAI 1634, pp. 174–185.
4. Mease, D., Wyner, A.J. and Buja, A. (2007). Boosted Classification Trees and Class Probability/Quantile Estimation. Journal of Machine Learning Research, Vol 8, pp 409-439.
5. Qin, Z.(2006). Naive Bayes Classification Given Probability Estimation Trees. Proc 5th Int. Conf on Machine Learning and Applucations (ICMLA'06), pp34–42.
6. Sulzmann, J-N. and Fürnkranz, J. (2009). An Empirical Comparison of Probability Estimation Techniques for Probabilistic Rules. Proc. 12th International Conference on Discovery Science (DS '09), Springer-Verlag, pp 317-331.
7. Zhang, K., Fan, W., Buckles, B., Yuan, X. and Xu, Z. (2006). Discovering Unrevealed Properties of Probability Estimation Trees: On Algorithm Selection and Performance Explanation. Proc 6th Int. Conf. on Data Mining (ICDM'06), IEEE, pp741–752.