

Interprétation graphique de la courbe ROC

François Rioult*

*GREYC CNRS UMR6072, Université de Caen Basse-Normandie
francois.rioult@unicaen.fr

Résumé. L'aire sous la courbe ROC est un outil pertinent pour mesurer la performance d'un classifieur et possède de nombreux avantages par rapport aux mesures de rappel et précision par classe : la performance est indiquée par une seule mesure et ne dépend pas des populations des classes. Cet avantage se transforme néanmoins en inconvénient lorsqu'il s'agit, lors de la lecture de la courbe, de calculer rapidement les mesures de rappel et de précision par classe ou d'estimer le comportement du classifieur en tenant compte des populations des classes.

Cet article propose une analyse graphique de la courbe ROC, qui permet d'améliorer son interprétation en fournissant une estimation des rappel et précision par classe.

Cet article a été publié sous une forme résumée à EGC'2011 (Rioult (2011)), présentation poster.

1 Introduction

La courbe ROC (Receiver Operating Characteristics) [Fawcett (2003)] offre à la fois une vision graphique et une mesure pertinentes de la performance d'un classifieur. Elle possède de nombreux avantages par rapport aux mesures de rappel et précision par classe : la performance est synthétisée par une unique mesure qui ne dépend pas des proportions de classe. Cet avantage se transforme néanmoins en inconvénient lorsqu'il s'agit de revenir rapidement aux mesures de rappel et de précision par classe ou d'estimer le comportement du classifieur selon les classes.

Les mesures de rappel et précision sont en effet utiles car elles caractérisent précisément le comportement du classifieur sur chacune des classes. En particulier, lorsque les classes sont déséquilibrées, ces mesures fournissent des indications de performance plus représentatives et concrètes qu'un unique score pour le classifieur.

Notre travail vise à améliorer l'estimation des performances d'un classifieur dont on connaît la courbe ROC : il s'agit donc d'une méthode graphique d'estimation, qui d'un coup d'œil permet de lire sur la courbe les rappel, précision et Fscore par classe. Bien évidemment, un calcul adapté, étant donnés les points de la courbe, permettrait de préciser tout score. Nous souhaitons plutôt introduire un mode nouveau de lecture de la courbe ROC qui permette d'appréhender plus profondément le comportement du classifieur. Par exemple, s'il est aisé de mesurer graphiquement le rappel en chaque point de la courbe (c'est son ordonnée), la valeur de la précision ne peut être obtenue que par calcul. Une méthode graphique pour la déterminer rapidement serait la bienvenue.

Interprétation géométrique de la courbe ROC

La courbe ROC est une caractéristique de la performance indépendante de la classe, cependant elle représente le point de vue de la classe positive, généralement la classe majoritaire. Des mesures concernant la classe minoritaire sont donc difficiles à obtenir graphiquement.

Nous présentons donc une méthode simple d'interprétation graphique de la courbe ROC, tant d'un point de vue de la classe positive que négative. Des considérations graphiques sur le point d'intersection de la courbe avec la diagonale descendante fournissent simplement des valeurs fiables du rappel et de la précision pour chaque classe.

2 Préliminaires

Dans un contexte d'apprentissage supervisé, qui étudie des exemples d'apprentissage (u_i) pourvus d'une étiquette indiquant sa classe, nous focalisons notre attention sur les problèmes de classification binaire ou *conceptuelle* [Cornuéjols A. (2010)] : il n'y a que deux classes, notées \mathcal{P} et \mathcal{N} , contenant respectivement P et N exemples. Le résultat de la classification est noté dans une table de contingence :

Décision ↓ Vérité →	\mathcal{P}	\mathcal{N}	total
\mathcal{P}	TP	FP	P
\mathcal{N}	FN	TN	N
total	P	N	$P + N$

Le résultat d'un classifieur utilise :

TP (resp. TN) : le nombre de vrais positifs (resp. négatifs)

FP (resp. FN) : le nombre de faux positifs (resp. négatifs).

Les mesures traditionnelles de la performance en classification sont propres à chaque classe :

Rappel : pour la classe \mathcal{P} : $Ra_{\mathcal{P}} = \frac{TP}{P}$, $Ra_{\mathcal{N}} = \frac{TN}{N}$ pour la classe \mathcal{N}

Précision : $Pr_{\mathcal{P}} = \frac{TP}{TP+FP}$, $Pr_{\mathcal{N}} = \frac{TN}{TN+FN}$

Fscore : le Fscore est une moyenne harmonique de paramètre β : $FS = \frac{\beta^2 Ra Pr}{\beta^2 Ra + Pr}$. Dans la suite on considérera le cas $\beta = 1$.

Pour obtenir des rappel, précision et Fscore indépendants des classes, on effectue les moyennes des valeurs par classe. Une autre mesure traditionnelle est le *score* de classification, qui indique la proportion d'exemples bien classés (dans les deux classes).

Aire sous la courbe ROC

L'utilisation des mesures précédemment décrites est contestée [Provost et al. (1998)], notamment car elles fournissent une vue trop particulière de la performance du classifieur et parce qu'elles sont trop sensibles à la disproportion de classe. On leur préfère souvent l'aire sous la courbe ROC, qui utilise des mesures normalisées par les populations des classes : les taux de vrais ou de faux positifs.

Considérant que le classifieur rend sa décision sous la forme d'une probabilité d'appartenance à la classe \mathcal{P} , la décision finale est modulable par un seuil. En faisant varier ce seuil, on calcule pour chaque classification le taux de faux positifs $\frac{FP}{N}$ et celui de vrais positifs $\frac{TP}{P}$, qui définissent la courbe dite ROC. L'aire sous cette courbe indique la probabilité que le classifieur fournisse un score d'appartenance à \mathcal{P} supérieur à celui d'appartenir à \mathcal{N} .

3 Méthode graphique d'interprétation de la courbe ROC

Notre travail vise à mieux cerner les différences de classification selon les classes grâce à une lecture graphique de la courbe ROC. Nous considérons pour cela disposer de la courbe ROC fournie par un classifieur, et nous déduisons des informations de cette courbe pour des constructions graphiques.

3.1 Courbe ROC pour la classe négative

La courbe ROC classique est relative aux performances pour la classe \mathcal{P} , nous la notons désormais $ROC_{\mathcal{P}}$. Nous introduisons alors la courbe $ROC_{\mathcal{N}}$, qui représente le taux de vrais négatifs $\frac{TN}{N}$ soit $Ra_{\mathcal{N}}$ en fonction du taux de faux négatifs $\frac{FN}{P}$. Le rapport graphique entre les deux courbes est trivial :

Théorème 1 Soit d la diagonale descendante (ou diagonale de gauche), d'équation $x + y = 1$. Cette diagonale constitue un axe de symétrie pour $ROC_{\mathcal{P}}$ et $ROC_{\mathcal{N}}$.

Preuve : $ROC_{\mathcal{P}}$ affiche les points de coordonnées $(x_{\mathcal{P}}, y_{\mathcal{P}}) = (\frac{FP}{N}, \frac{TP}{P})$, $ROC_{\mathcal{N}}$ les points $(x_{\mathcal{N}}, y_{\mathcal{N}}) = (\frac{FN}{P}, \frac{TN}{N})$. Or $\frac{FN}{P} = \frac{P - TP}{P} = 1 - \frac{TP}{P}$ et $\frac{TN}{N} = \frac{N - FP}{N} = 1 - \frac{FP}{N}$. Donc $x_{\mathcal{N}} = 1 - y_{\mathcal{P}}$ et $y_{\mathcal{N}} = 1 - x_{\mathcal{P}}$. On reconnaît une symétrie d'axe d . \square

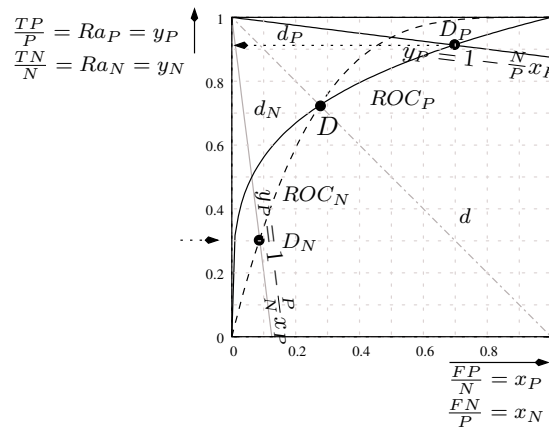


FIG. 1 – Interprétation graphique de la courbe ROC selon les classes. La courbe $ROC_{\mathcal{P}}$, en trait plein, a pour équation $y_{\mathcal{P}} = x_{\mathcal{P}}^{1/4}$.

Sur la figure 1, nous avons représenté les courbes pour les deux classes, ainsi que leurs axes de coordonnées $(x_{\mathcal{P}}, y_{\mathcal{P}})$ et $(x_{\mathcal{N}}, y_{\mathcal{N}})$.

En corollaire, les aires sous les courbes $ROC_{\mathcal{P}}$ et $ROC_{\mathcal{N}}$ sont égales, par symétrie. C'est là encore l'intérêt de cette mesure, qui ne caractérise pas le comportement du classifieur selon une seule classe mais fournit une évaluation générale sur les deux classes.

Interprétation géométrique de la courbe ROC

3.2 Analyse de la courbe ROC dans le cas de classes équilibrées ($P = N$)

Définition 1 Nous introduisons le point D d'intersection entre les courbes $ROC_{\mathcal{P}}$, $ROC_{\mathcal{N}}$ et la diagonale descendante d . Au point D , les équations suivantes, selon les systèmes d'axes, sont satisfaites :

$$Ra_{\mathcal{P}} + \frac{FP}{N} = 1 \quad \text{et} \quad Ra_{\mathcal{N}} + \frac{FN}{P} = 1.$$

Les calculs suivants sont menés au point D .

Proposition 1 Au point D , le rappel, la précision et le Fscore sont tous égaux, pour les deux classes :

$$Pr_{\mathcal{P}} = Ra_{\mathcal{P}} = Fscore_{\mathcal{P}} = Pr_{\mathcal{N}} = Ra_{\mathcal{N}} = Fscore_{\mathcal{N}}$$

Preuve : D'une façon générale, $Pr_{\mathcal{P}} = \frac{TP}{TP+FP} = \frac{PRa}{PRa+FP} = \frac{Ra}{Ra+\frac{FP}{P}}$. Si $P = N$, alors $Pr_{\mathcal{P}} = \frac{Ra}{Ra+\frac{FP}{N}}$: Le dénominateur de la précision représente donc la somme des coordonnées d'un point de la courbe $ROC_{\mathcal{P}}$. Au point D , cette somme vaut 1 et $Pr_{\mathcal{P}} = Ra_{\mathcal{P}} = Fscore_{\mathcal{P}}$. La même propriété est valable pour la classe négative. \square

Ainsi, dans le cas de classes équilibrées, le rappel (coordonnée y) au point D donne immédiatement la précision et le Fscore, car tous sont égaux. Il s'agit là d'un moyen extrêmement simple et rapide d'évaluer la performance d'un classifieur.

Ce point définit le *seuil d'équilibre* (breakeven), largement utilisé et discuté¹ comme mesure de performance (voir par exemple [Sebastiani (2002)]). Ses propriétés géométriques, soulignées ici, nous semblent nouvelles et intéressantes pour améliorer l'interprétation de la courbe ROC, en la rapprochant de considérations en rappel/précision.

3.3 Classes déséquilibrées ($P > N$)

Nous introduisons désormais deux points d'intersection avec les courbes ROC :

- $D_{\mathcal{P}}$ est le point d'intersection entre $ROC_{\mathcal{P}}$ et la droite d'équation $y + \frac{N}{P}x = 1$;
- $D_{\mathcal{N}}$ est le point d'intersection entre $ROC_{\mathcal{N}}$ et la droite d'équation $y + \frac{P}{N}x = 1$.

Propriété 1 Aux points $D_{\mathcal{P}}$ et $D_{\mathcal{N}}$, le rappel, la précision et le Fscore sont tous égaux :

$$Pr_{\mathcal{P}} = Ra_{\mathcal{P}} = Fscore_{\mathcal{P}} \quad \text{et} \quad Pr_{\mathcal{N}} = Ra_{\mathcal{N}} = Fscore_{\mathcal{N}}$$

Preuve : Dans la preuve de la propriété 1, l'expression de la précision est $Pr_{\mathcal{P}} = \frac{Ra}{Ra+\frac{FP}{P}} = \frac{Ra}{Ra+\frac{N}{P}\frac{FP}{N}}$. Au point $D_{\mathcal{P}}$, ce dénominateur vaut 1 et on obtient la propriété. \square

Pour déterminer les valeurs par classe, on lit le rappel pour \mathcal{P} comme l'ordonnée du point $D_{\mathcal{P}}$, qui fournit également précision et Fscore. Pour la classe \mathcal{N} , on peut construire $D_{\mathcal{N}}$ comme le symétrique de $D_{\mathcal{P}}$ par rapport à d et son ordonnée livre rappel, précision et Fscore ; le complémentaire à 1 de abscisse de $D_{\mathcal{P}}$ remplit également cette fonction.

Sur notre exemple de la figure 1, si nous considérons que $P = N$, toutes les mesures valent 0.72 si les classes sont équilibrées. Nous avons représenté $D_{\mathcal{P}}$ et $D_{\mathcal{N}}$ pour $\frac{P}{N} = 8$: on obtient des mesures de 0.91 pour la classe \mathcal{P} et de 0.3 pour la classe \mathcal{N} .

1. Entre autres, ce seuil ne fournit le maximum ni de la précision ni du rappel.

Cette méthode de lecture graphique de la courbe fournit des mesures de précision et de Fscore selon les classes. Elle illustre également les compromis entre rappel et précision obtenus en ces points particuliers, lorsque les disproportions de classe augmentent.

3.4 Variations de déséquilibre et évolution des mesures

Nous proposons de compléter la lecture d'une courbe ROC pour mieux cerner les évolutions de la précision et du Fscore en chaque point de la courbe, selon la variation de faux positifs. Cette analyse graphique fournit des tendances pour les rappel, précision et Fscore dans le cas de classe déséquilibrées.

Nous avons les relations suivantes : $Pr_{\mathcal{P}} = \frac{Ra_{\mathcal{P}}}{Ra_{\mathcal{P}} + N/P.FP/N}$ et

$$Fscore_{\mathcal{P}} = \frac{2Ra_{\mathcal{P}}^2}{Ra_{\mathcal{P}} + Ra_{\mathcal{P}}(Ra_{\mathcal{P}} + N/P.FP/N)} = \frac{2Ra_{\mathcal{P}}}{1 + Ra_{\mathcal{P}} + N/P.FP/N}$$

La courbe de précision en fonction du taux de faux positifs passe donc par les points $(0, 1)$ (par prolongement par continuité, avec une tangente verticale), D et $(1, \frac{1}{1+N/P})$. La courbe de Fscore en fonction du taux de faux positifs passe par les points $(0, 0)$ (par prolongement par continuité, avec une tangente verticale), D et $(1, \frac{2}{2+N/P})$.

La figure 2 montre les évolutions des points d'intersection lorsque le déséquilibre entre les classes augmente. Pour chaque valeur de $\frac{P}{N}$, on affiche, en fonction du taux de faux positifs, les courbes de rappel (ou courbe ROC), de précision et de Fscore. Sachant que ces courbes passent par $D_{\mathcal{P}}$ et $D_{\mathcal{N}}$, la position de ces points permet d'estimer les évolutions de ces mesures.

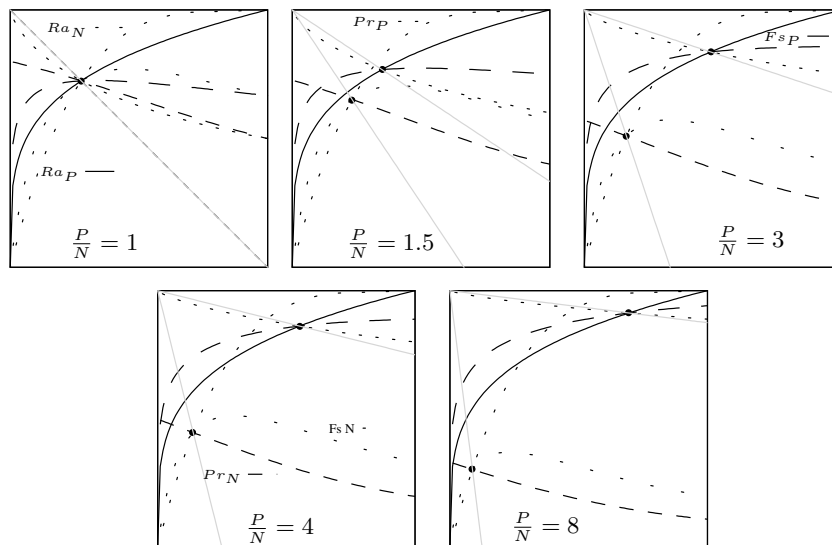


FIG. 2 – Pour une même courbe ROC, différentes valeurs de $\frac{P}{N}$: 1, 1.5, 3, 4, 8. Les courbes de précision et de Fscore sont modifiées en conséquence.

Bien que la courbe ROC soit un indicateur indépendant de la répartition des classes, elle permet ici d'imaginer l'impact d'un déséquilibre des classes :

Interprétation géométrique de la courbe ROC

- la situation neutre symbolisée par le point D fournit des mesures de rappel, précision et Fscore égales pour chaque classe ;
- en faisant varier le rapport entre P et N (on suppose $P > N$), les droites d_P et d_N s'écartent de la diagonale ;
- le point D_P monte le long de la courbe ROC_P à partir du point D : les rappel, précision et Fscore pour la classe dominante augmentent conjointement ;
- le point D_N descend le long de la courbe ROC_N à partir de D et les mesures chutent.

Dans le cas de classes déséquilibrées, tracer les droites d_P et d_N permet d'appréhender graphiquement le comportement du classifieur selon les deux classes : les performances en rappel, précision et Fscore seront meilleures pour la classe majoritaire, au détriment de la classe minoritaire.

4 Conclusion

Nous avons proposé une méthode simple d'interprétation des performances d'un classifieur étant donnée sa courbe ROC. L'ordonnée des points d'intersection avec des diagonales descendantes fait le lien avec le rappel, la précision et le Fscore par classes. Cette construction permet d'appréhender graphiquement les différences de performance selon les classes et d'approfondir la lecture de la courbe ROC.

Références

- Cornuéjols A., M. L. (2010). *Apprentissage artificiel - Concepts et algorithmes*. Paris : Eyrolles.
- Fawcett, T. (2003). Roc graphs : Notes and practical considerations for researchers. Tech Report HPL-2003-4, HP Laboratories.
- Provost, F. J., T. Fawcett, et R. Kohavi (1998). The case against accuracy estimation for comparing induction algorithms. In *ICML '98*, pp. 445–453.
- Riout, F. (2011). Interprétation graphique de la courbe roc. In *Extraction et Gestion des Connaissances (EGC'11)*, pp. 301–304.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys* 34(1), 1–47.