

Compromis précision-rappel dans l'évaluation des performances

Blaise Hanczar, Mohamed Nadif

LIPADE, Université Paris Descartes, 45 rue des saint-pères, 75006 Paris
<blaise.hanczar@parisdescartes.fr>

Résumé. Dans de nombreux problèmes d'apprentissage automatique la performance des algorithmes est évaluée à l'aide des mesures *précision* et *rappel*. Or ces deux mesures peuvent avoir une importance très différente en fonction du contexte. Dans cet article nous étudions le comportement des principaux indices de performance en fonction du couple précision-rappel. Nous proposons un nouvel outil de visualisation de performances et définissons l'espace de compromis qui représente les différents indices en fonction du compromis précision-rappel. Nous analysons les propriétés de ce nouvel espace et mettons en évidence ses avantages par rapport à l'espace précision-rappel.

1 Introduction

En apprentissage automatique, la précision et le rappel sont des mesures classiques pour évaluer les résultats et la performance des algorithmes utilisés. Ces mesures sont essentiellement utilisées en apprentissage supervisé (Sokolova et al., 2006), en classification simple (Jain, 2010) et croisée Hanczar et Nadif (2013) et en recherche d'information (Manning et al., 2008). Dans ce dernier cas, la performance de l'algorithme employé est évaluée à partir de la similarité entre l'ensemble de documents retrouvés et l'ensemble des documents cibles. Cette similarité se base sur la précision et le rappel. De la même manière en classification simple (resp. croisée), les algorithmes identifient des groupes (resp. biclusters) d'éléments qui sont comparés à des groupes (resp. biclusters) de référence. En apprentissage supervisé, l'évaluation d'un classifieur se fait en comparant les classes prédites avec les vraies classes sur un ensemble de test. On mesure la similarité entre les classes prédites et les vraies classes en calculant leur précision et rappel. Cependant cette approche ne tient pas compte du taux de vrais négatifs. Pour ces raisons, on préfère dans certains cas utiliser le couple sensibilité-spécificité que le couple précision-rappel dans ce contexte. La précision et le rappel sont donc deux mesures très utilisées dans les procédures d'évaluations de nombreux domaines. Il est extrêmement fréquent de combiner ces deux valeurs afin de construire des indices de performance tel que la F-mesure ou l'indice de Jaccard (Albatineh et Niewiadomska-Bugaj, 2011).

Par défaut les indices de performance donnent la même importance à la précision et au rappel. Or dans de nombreux cas, on peut vouloir privilégier l'un par rapport à l'autre. Par exemple, en génomique des groupes de gènes ayant des profils d'expression similaires sont identifiés en utilisant des méthodes de classification. Ces groupes sont comparés à des classifications de gènes issues de bases de connaissance afin d'estimer leur pertinence biologique

(Datta et Datta, 2006). L'objectif de ces analyses est de capturer le plus d'information biologique dans les groupes de gènes, on veut donc privilégier le rappel par rapport à la précision dans ce contexte. Certains indices de performance ont une variante introduisant un paramètre permettant de contrôler le compromis précision-rappel comme c'est le cas de la F-mesure qui est une généralisation de l'indice de Dice. Pour d'autres mesures, le contrôle du compromis précision-rappel est plus difficile, comme c'est le cas de l'indice de Jaccard. Dans cet article nous analysons les différents indices de performance en fonction du compromis précision-rappel. Nous proposons également un nouvel outil d'analyse qu'est l'espace de compromis qui présente de nombreux avantages par rapport à l'espace précision-rappel.

Dans la section 2, nous présentons les différents indices de performance étudiés ainsi que leurs variantes sensibles au compromis. Dans la section 3, nous rappelons les propriétés de l'espace précision-rappel. Nous analysons le comportement des différents indices dans cet espace. Dans la section 4, nous définissons l'espace de compromis et nous montrons comment représenter les performances par les courbes de compromis. Dans la section 5, nous montrons les avantages à travailler dans l'espace de compromis en particulier pour la sélection de modèles et la comparaison d'algorithmes. Nous illustrons ces propriétés avec un exemple dans le contexte du *biclustering*. Dans la section 6, nous exposons nos conclusions et perspectives.

2 Indices basés sur le couple précision et rappel

2.1 Définitions

Soit D un ensemble des données contenant N éléments. Nous appelons *groupe cible* le sous-ensemble $T \subset D$ que nous recherchons. Un algorithme dont l'objectif est de retrouver le groupe cible produit un groupe X . Pour mesurer la qualité de ce groupe X , un indice de performance est utilisé afin d'évaluer la similitude entre T et X . Ces indices de performances sont généralement basés sur deux valeurs : la précision et le rappel. La précision représente la proportion de X qui recouvre T quant au rappel il exprime la proportion de T retrouvé par X . Ces deux indices prennent les formes suivantes :

$$pre = precision(T, X) = \frac{|T \cap X|}{|X|} \quad \text{et} \quad rec = recall(T, X) = \frac{|T \cap X|}{|T|}.$$

Les principaux indices de performances utilisés sont une combinaison de la précision et du rappel. Dans cet article nous étudierons les quatre plus populaires : l'indice de Kulczynski, F-mesure, Folke et Jaccard. Ces travaux pourront être facilement étendus à d'autres indices. Par défaut chacun de ces indices donne la même importance à la précision et au rappel. Cependant on peut construire des versions pondérées permettant de privilégier la précision par rapport au rappel ou inversement.

2.2 L'indice de Kulczynski

L'indice de Kulczynski est la moyenne arithmétique de la précision et du rappel.

$$I_{Kul}(T, X) = \frac{1}{2}(pre + rec).$$

Une version pondérée introduit le paramètre $R \in [0, +\infty]$ qui permet de contrôler le compromis entre la précision et le rappel. Plus R est grand, plus le rappel est important, le point d'équilibre est atteint pour $R = 1$. Nous réécrivons cet indice en effectuant le changement de variable suivant : $\lambda = \frac{R}{R+1}$, $\lambda \in [0, 1]$ contrôle désormais le compromis et le point d'équilibre est atteint pour $\lambda = 0.5$.

$$I_{Kul}(T, X, R) = \frac{1}{R+1}(R.pre + rec)I_{Kul}(T, X, \lambda) = (1 - \lambda)pre + \lambda rec.$$

2.3 La F-mesure

La F-mesure, appelée aussi indice de Dice, est le rapport entre l'intersection et la somme des tailles du groupe X et du groupe cible T . C'est aussi la moyenne harmonique entre la précision et le rappel.

$$I_{Fmes}(T, X) = \frac{2|T \cap X|}{|T| + |X|} = \frac{2}{\frac{1}{rec} + \frac{1}{pre}} = \frac{2pre.rec}{pre + rec}.$$

La F-mesure est une version pondérée de l'indice ci-dessus. Le paramètre $\beta \in [0, +\infty]$ permet de contrôler le compromis entre la précision et le rappel. Plus β est grand, plus le rappel est important, le point d'équilibre est atteint lorsque $\beta = 1$. Nous réécrivons cette mesure en effectuant le changement de variable suivant : $\lambda = \frac{\beta^2}{1+\beta^2}$, $\lambda \in [0, 1]$ contrôle désormais le compromis et le point d'équilibre est atteint pour $\lambda = 0.5$. La F-mesure prend alors les deux formes suivantes :

$$I_{Fmes}(T, X, \beta) = \frac{1 + \beta^2}{\frac{\beta^2}{rec} + \frac{1}{pre}} = (1 + \beta^2) \frac{pre.rec}{\beta^2 pre + rec},$$

$$I_{Fmes}(T, X, \lambda) = \frac{1}{\frac{\lambda}{rec} + \frac{1-\lambda}{pre}} = \frac{pre.rec}{\lambda pre + (1 - \lambda)rec}.$$

2.4 L'indice de Folke

L'indice de Folke correspond à la moyenne géométrique de la précision et du rappel.

$$I_{Fk}(T, X) = \frac{|T \cap X|}{\sqrt{|T||X|}} = \sqrt{pre.rec}.$$

Il est possible de pondérer le moyenne géométrique en introduisant un paramètre $\lambda \in [0, 1]$. Plus λ est grand plus le rappel est important et le point d'équilibre est atteint pour $\lambda = 0.5$.

$$I_{Fk}(T, X, \lambda) = \frac{|T \cap X|}{|T|^{1-\frac{\lambda}{2}} |X|^{\frac{\lambda}{2}}} = pre^{\frac{\lambda}{2}} rec^{1-\frac{\lambda}{2}}.$$

Le compromis précision rappel

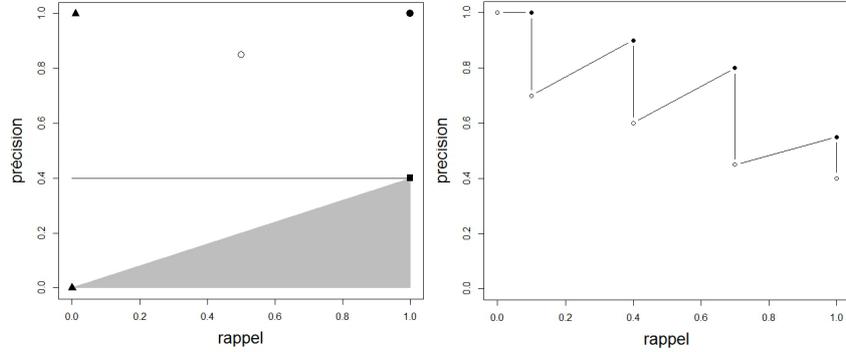


FIG. 1 – A gauche, l'espace précision-rappel. A droite, la courbe précision-rappel.

2.5 L'indice de Jaccard

L'indice de Jaccard est le rapport entre l'intersection et l'union du groupe X et le groupe cible T .

$$I_{Jac}(T, X) = \frac{|T \cap X|}{|T| + |X| - |T \cap X|} = \frac{pre \cdot rec}{pre + rec - pre \cdot rec}$$

Il n'est pas facile de définir une version pondérée de l'indice de Jaccard à cause de la présence du terme $pre \cdot rec$ au dénominateur. Nous voulons un indice pondéré ayant les propriétés suivantes : $I_{Jac}(T, X, \lambda) \in [0, 1]$; $I_{Jac}(T, T, \lambda) = 1$; $I_{Jac}(T, X, 0.5) = I_{Jac}(T, X)$; $I_{Jac}(T, X, 0) = pre$; $I_{Jac}(T, X, 1) = rec$. Pour cela nous proposons l'indice suivant :

$$I_{Jac}(T, X, \lambda) = \frac{pre \cdot rec}{w(\lambda) \cdot pre + w(1 - \lambda) \cdot rec - v(\lambda) \cdot pre \cdot rec}$$

avec $w(\lambda) = \min\{2\lambda, 1\}$ et $v(\lambda) = 1 - |1 - 2\lambda|$.

3 L'espace précision-rappel

L'espace précision-rappel, illustré dans la figure 1, est un espace à deux dimensions dans lequel les abscisses et ordonnées représentent respectivement le rappel et la précision (Buckland et Gey, 1994). Une performance est représentée par un point dans cet espace (le point blanc par exemple). Le principe de l'espace précision-rappel est proche de celui de l'espace ROC qui représente le taux de vrais positifs en fonction du taux de faux positifs (Fawcett, 2006). Plusieurs relations ont d'ailleurs été identifiées entre ces deux espaces (Davis et Goadrich, 2006). Un point dans l'espace précision-rappel représente tous les groupes de taille $|X| = |T| \frac{rec}{pre}$ ayant une intersection avec le groupe cible de $|T \cap X| = |T| rec$. Le point (1,1) (point noir), maximisant la précision et le rappel, représente le groupe idéal et dans ce cas il y a une parfaite correspondance avec le groupe cible ($X = T$). Le point $(1, \frac{|T|}{|D|})$ (carré noir) représente le cas où le groupe contient toutes les données ($X = D$). La ligne horizontale partant de ce point correspond à l'espérance de performance des groupes aléatoires de différentes tailles.

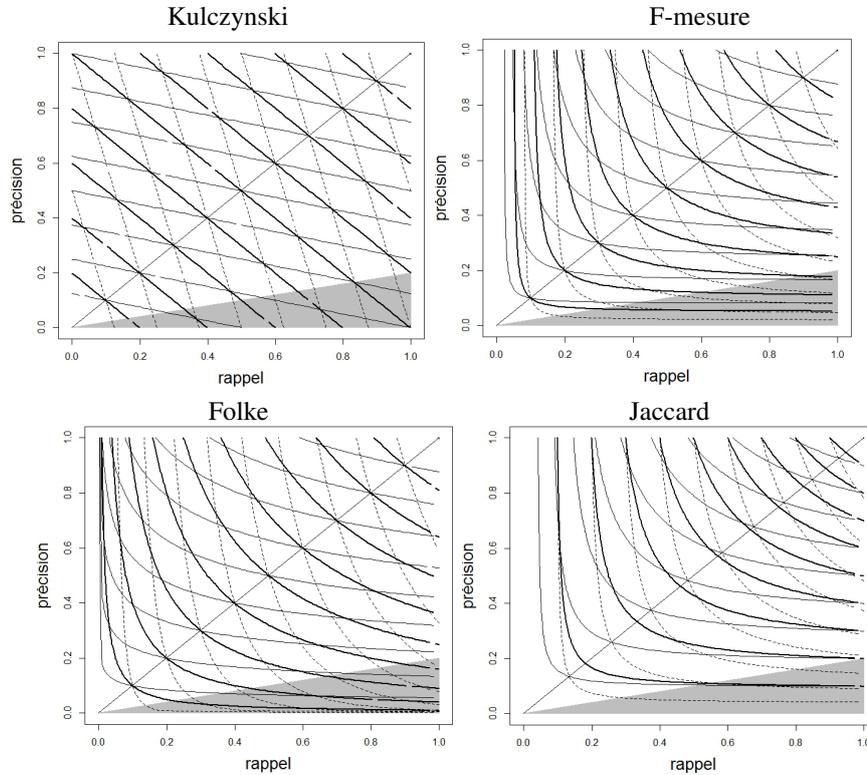


FIG. 2 – Les isolignes des différents indices de performances dans l’espace précision-rappel.

En effet l’espérance de précision d’un groupe aléatoire ne dépend que de la taille du groupe cible et des données ; pour un groupe cible fixe elle est donc constante $E[pre] = \frac{|T|}{|D|}$. L’espérance du rappel dépend de la taille du groupe $E[rec] = \frac{|X|}{|D|}$. Les deux triangles représentent les performances des groupes contenant un unique élément : le point $(\frac{1}{|T|}, 1)$ si l’élément appartient à T et le point $(0, 0)$ sinon. La zone grise représente un ensemble de performances qui ne peuvent pas être atteintes. A partir des définitions de la précision et du rappel, on peut écrire $pre \geq \frac{|T|}{|D|}rec$ puisqu’on a $|D| \geq |X|$. Tous les groupes dont la performance se situe sur la droite $pre = \frac{|T|}{|D|}rec$ sont ceux dont $|T \cap X|$ est minimale. Cette droite représente tous les groupes dont $|T \cap X|$ est nulle. La plupart des algorithmes a un paramètre permettant de contrôler la taille du résultat X . Pour chaque taille de X on obtient des valeurs de précision et rappel différentes. La performance d’un algorithme peut donc être représentée par un ensemble de points et approximée par une courbe dans l’espace précision-rappel. Dans la figure 1, on donne un exemple de courbe précision-rappel. On peut tirer plusieurs informations sur les performances de ces différents groupes même sans se référer à un indice en particulier. Si un point domine un autre, c-à-d si sa précision et son rappel sont supérieurs, alors on peut conclure

Le compromis précision rappel

qu'il aura une meilleure performance quelque soit l'indice utilisé. Les points noirs représentent les points dominants de la courbe, il ne sont dominés par aucun autre point et représentent les performances des meilleurs groupes. Il n'y a pas de rapport de domination entre ces types de points, il est nécessaire d'utiliser un indice pour les comparer.

Le comportement des différents indices de performances peut se visualiser en dessinant leur iso-ligne dans l'espace précision-rappel. Une iso-ligne est un ensemble de points dans l'espace précision-rappel ayant tous la même valeur d'indice (Flach, 2003; Hanczar et Nadif, 2013). La figure 2 montre les iso-lignes des indices de Kulczynski, F-mesure, Folke et Jaccard. Les lignes en gras représentent les iso-lignes lorsque $\lambda = 0.5$. Pour les quatre indices, nous observons que les iso-lignes ont une symétrie autour de l'axe $pre = rec$, ceci signifie que la précision et le rappel ont la même importance. Par contre les différents indices ne considèrent pas la différence entre précision et rappel de la même façon. Cette différence n'est pas prise en compte dans l'indice de Kulczynski, alors que les autres indices la pénalisent. L'indice de Folke pénalise moins que la F-mesure et l'indice de Jaccard. Ces deux derniers sont équivalents car ils sont compatibles,

$$I_{fmes}(T, X_1) \geq I_{fmes}(T, X_2) \Leftrightarrow I_{jac}(T, X_1) \geq I_{jac}(T, X_2).$$

Dans la figure 2, les lignes en pointillées représentent les iso-lignes pour $\lambda = 0.2$ et les lignes pleines $\lambda = 0.8$. La modification de la valeur de λ déforme les iso-lignes, ce qui permet de donner plus d'importance à la précision ou au rappel. A noter que pour $pre = rec$ les indices de Kulczynski, F-mesure et Folke retournent la même valeur quelque soit λ . L'indice de Jaccard a un comportement différent, il pénalise le fait que λ s'approche de 0.5. Dans la table 1, sont reportés quelques exemples illustrant ces caractéristiques. On constate que les valeurs et rangs de chaque point varient selon l'indice, le choix de l'indice est donc une étape cruciale qui doit être fait en fonction du contexte.

		$\lambda = 0.5$				$\lambda = 0.2$			
précision	rappel	Kulczynski	Fmesure	Folke	Jaccard	Kulczynski	Fmesure	Folke	Jaccard
0.7	0.7	0.7	0.7	0.7	0.54	0.7	0.7	0.7	0.625
0.5	0.9	0.7	0.64	0.67	0.47	0.58	0.55	0.56	0.49
0.9	0.5	0.7	0.64	0.67	0.47	0.82	0.77	0.8	0.66

TAB. 1 – Exemples de différentes performances sur les quatres indices

4 L'espace de compromis

4.1 Définitions

L'espace de compromis, que nous proposons, offre un nouvel outil de visualisation des performances des résultats ou des algorithmes en fonction du compromis précision-rappel. Il y a certaines similitudes avec les "cost curves" utilisées en apprentissage supervisé (Drummond et Holte, 2006). L'espace de compromis représente en abscisse λ et en ordonnée l'indice de performance. La performance d'un groupe X est représentée dans cet espace par une courbe $f(\lambda)$. On a une correspondance entre les points de l'espace précision-rappel et les courbes de l'espace de compromis. La figure 3 donne un exemple de ces courbes pour un groupe de performance $pre = 0.85$ et $rec = 0.5$. Les courbes représentant l'indice de performance sont

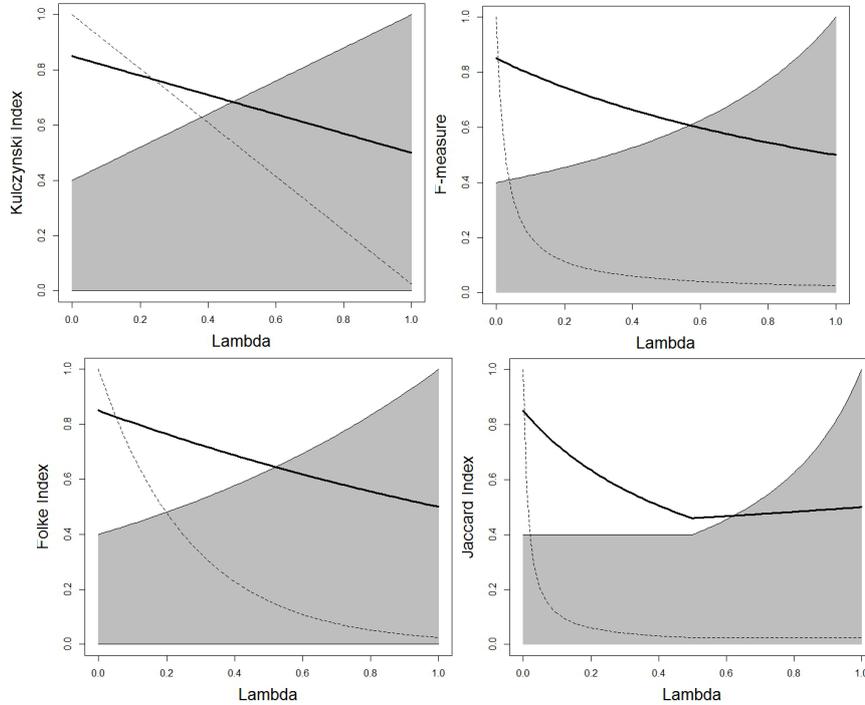


FIG. 3 – L'espace de compromis des différents indices de performances

en gras. Les extrémités des courbes donnent les précisions et rappel du groupe, on a $f(0) = pre$ et $f(1) = rec$. La ligne pleine représente les performances du groupe maximal c-à-d. le groupe contenant toutes les données, cela correspond au point $(1, \frac{|X|}{|D|})$ de l'espace précision-rappel. Cette dernière courbe définit le domaine d'application des indices de performances pour un problème donné, illustré dans la figure 3 par les zones blanches. Un point situé dans l'une des zones grises, signifie que le groupe correspondant à de moins bonnes performances que le groupe maximal et peut donc être considéré comme non informatif. On constate que le domaine d'application de l'indice de Kulczynski est beaucoup plus petit que celui des autres indices. Cela est dû au fait que cet indice ne pénalise pas la différence entre précision et rappel. La ligne en pointillé représente le groupe contenant un unique élément appartenant au groupe cible. Le groupe parfait est représenté par la droite $f(\lambda) = 1$. A l'inverse les groupes ayant une intersection nulle sont représentés par la droite $f(\lambda) = 0$. Les groupes aléatoires sont représentés par les courbes partant du point $(0, \frac{|T|}{|D|})$.

4.2 Courbe optimale de compromis

Comme nous l'avons illustré dans la figure 1, la performance d'un algorithme peut être représentée par une courbe dans l'espace précision-rappel. A chaque point de cette courbe correspond une courbe dans l'espace de compromis. On peut représenter la courbe précision-rappel

Le compromis précision rappel

par un ensemble de courbes dans l'espace de compromis. La figure 4 donne la représentation de la courbe précision-rappel de la figure 1 dans l'espace de compromis pour les différents indices de performance. On s'intéressera particulièrement à l'enveloppe supérieure de cet ensemble de courbes, représentée en gras dans la figure 4 que nous appellerons courbe optimale de compromis. Cette dernière représente les meilleures performances de l'algorithme pour tous les compromis. On s'aperçoit que les courbes formant l'enveloppe supérieure correspondent tous à des points dominants de la courbe précision-rappel. Les points dominés ont toujours leur courbe en dessous de la courbe optimale de compromis. Dans le cas de l'indice de Kulczynski, les courbes formant l'enveloppe supérieure correspondent aux points de l'enveloppe convexe de la courbe précision-rappel. Ces courbes de compromis permettent d'analyser les résultats bien plus facilement que les courbes précision-rappel.

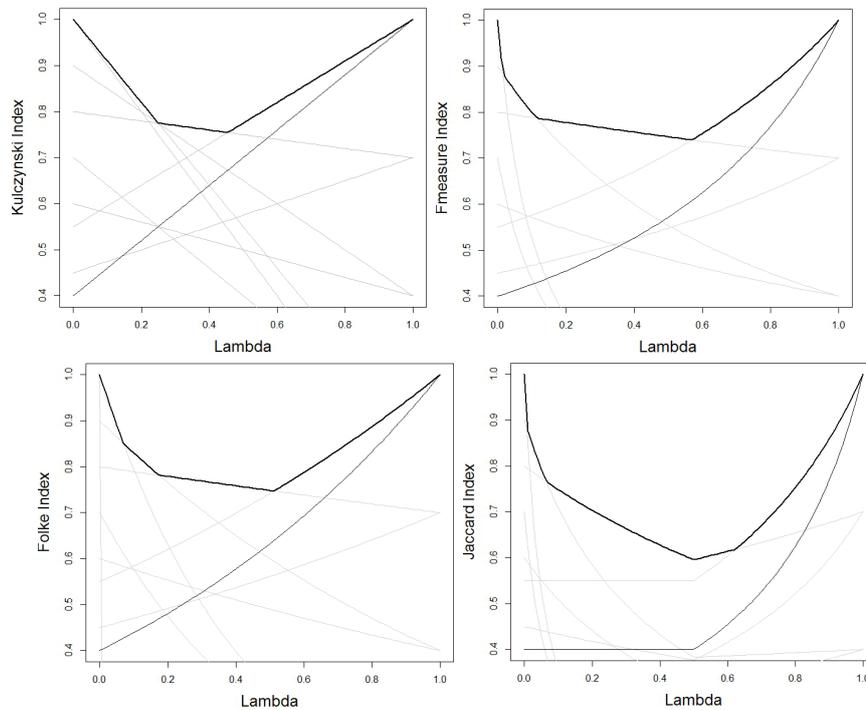


FIG. 4 – Représentation d'une courbe précision-rappel dans l'espace de compromis pour les différents indices. Les courbes grises sont les courbes de compromis et la courbe en gras est la courbe optimale de compromis.

5 Application des courbes de compromis

5.1 Sélection de modèles

L'utilisation de l'espace de compromis permet d'identifier très facilement le résultat optimal pour un compromis donné. Ceci est illustré dans la figure 5 à travers un problème de classification croisée. Nous avons généré une matrice de données aléatoires dans laquelle un bicluster a été introduit, ce dernier suit un modèle additif selon la définition de Madeira et Oliveira (2004). Nous utilisons l'algorithme CC (Cheng & Church) pour retrouver ce bicluster (Cheng et Church, 2000). La similarité entre le bicluster retourné par l'algorithme et le bicluster recherché est alors calculée par les différents indices de performance. Cet algorithme dispose d'un paramètre permettant de contrôler la taille du bicluster retourné, nous pouvons donc représenter les performances de cet algorithme par une courbe précision-rappel (figure 5). A partir de cette courbe il n'est pas facile de déterminer le meilleur bicluster pour un compromis de précision-rappel donné. Même en ajoutant les iso-lignes au graphique, la comparaison des différents biclusters n'est pas intuitive. Dans la figure 5 est représentée la courbe optimale de compromis pour la F-mesure. A partir de cette courbe on peut instantanément identifier le meilleur bicluster pour un compromis donné. On a aussi une décomposition de la valeur de λ en une série d'intervalles qui sont délimités sur le graphique par les lignes verticales pointillées, pour lesquels le meilleur bicluster est donné. Sur notre exemple on constate qu'il y a sept intervalles, nous nous intéresserons donc qu'aux sept biclusters correspondants, identifiés sur la figure par leur taille. Pour le dernier intervalle ($\lambda > 0.74$) le meilleur bicluster est la matrice entière, la courbe optimale de compromis est confondue avec la courbe du bicluster maximal. Notons qu'il n'est pas possible d'identifier visuellement ces biclusters dans l'espace précision-rappel car ils ne correspondent ni à l'ensemble des points dominants ni à l'enveloppe convexe de la courbe précision-rappel (sauf dans le cas de l'indice de Kulczynski). Il est également très facile de travailler avec des contraintes sur la précision ou le rappel dans l'espace de compromis. Nous rappelons que la précision et le rappel se lisent à l'extrémité de chaque courbe de compromis. Lorsqu'on demande une précision minimale pre_{min} , il suffit de considérer unique les courbes de compromis qui partent au-dessus du seuil minimum c-à-d $f(0) > pre_{min}$. De même avec un rappel minimum rec_{min} , on ne conserve que les courbes qui arrivent au-dessus du seuil de rappel c-à-d $f(1) > rec_{min}$.

5.2 Comparaison d'algorithmes

L'espace de compromis simplifie également grandement la comparaison des algorithmes. Nous reprenons l'exemple de classification croisée précédent dans lequel un autre algorithme, ISA (Bergmann et al., 2003), est testé et comparé à CC. Les performances de ce nouvel algorithme sont représentées dans l'espace précision-rappel et l'espace de compromis par la courbe grise dans la figure 6. Dans l'espace précision-rappel les deux courbes se croisent plusieurs fois, aucun des deux algorithmes n'est donc absolument meilleur que l'autre. Il est difficile de voir dans quelles conditions CC est meilleur que ISA et inversement. Dans l'espace de compromis on visualise immédiatement quel est le meilleur algorithme pour chaque valeur de compromis. Pour $\lambda < 0.28$ CC est meilleur que ISA, pour $0.28 < \lambda < 0.83$, ISA est meilleur, pour $\lambda > 0.83$ les deux algorithmes retournent un bicluster contenant toute la matrice de donnée et ont donc des performances identiques. La distance entre les deux courbes permet de

Le compromis précision rappel

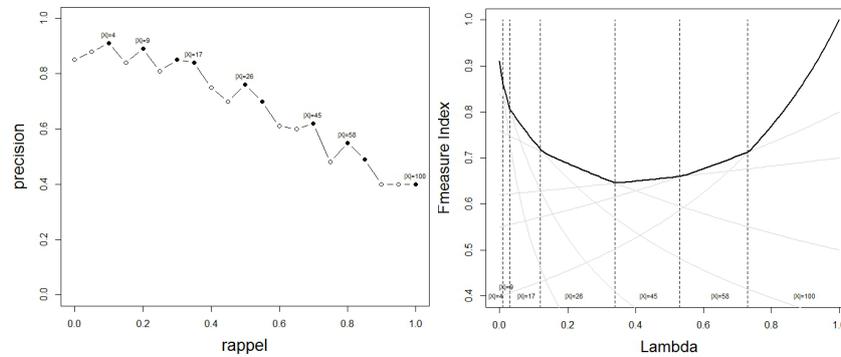


FIG. 5 – Identification des meilleurs biclusters dans l'espace précision-rappel et l'espace de compromis. A gauche, la courbe précision-rappel. A droite, la courbe optimale de compromis.

visualiser la différence de qualité entre les deux algorithmes. Dans l'espace précision-rappel les courbes des deux algorithmes se croisent trois fois, laissant penser qu'il y a deux intervalles de λ pour lesquels CC est meilleur (de même pour ISA). Les courbes optimales de compromis montrent que l'identité du meilleur algorithme ne change qu'une fois, en $\lambda = 0.28$. Dans l'espace précision-rappel, CC a une meilleure précision que ISA ; 14 fois sur 20 ce qui laisse penser que CC est plus souvent meilleur que ISA. L'espace de compromis nous montre qu'au contraire l'intervalle $[0, 0.28]$ pour lequel CC est meilleur est deux fois plus petit que celui de ISA $[0.28, 0.83]$. Cet exemple illustre bien la facilité de la comparaison d'algorithmes dans l'espace de compromis.

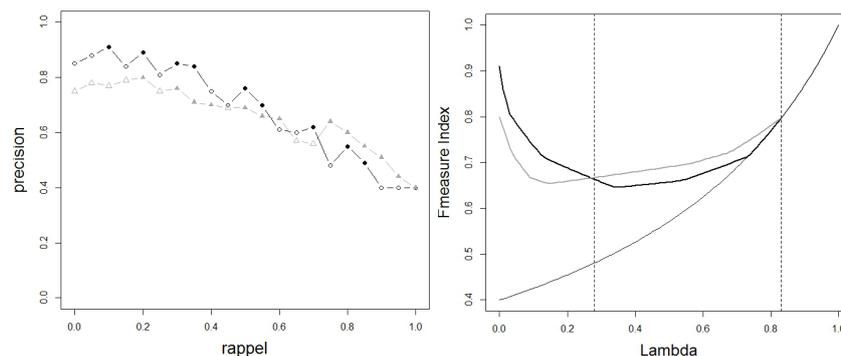


FIG. 6 – Identification du meilleur algorithme. A gauche, les courbes précision-rappel. A droite, les courbes optimales de compromis.

6 Conclusion

Dans cet article nous avons montré comment gérer le compromis précision-rappel pour les différents indices de performance. L'analyse de ces indices dans l'espace précision-rappel montre que leur comportement dépend particulièrement de la différence *pre-rec* et du compromis λ . Le choix de l'indice de performance est donc crucial et dépend du contexte. Il est par conséquent nécessaire que l'utilisateur prenne en compte le comportement de ces différents indices afin de sélectionner celui qui est le plus adapté à son problème. Nous avons également défini un nouvel outil qu'est l'espace de compromis, dans lequel les performances sont représentées par une courbe. Cet espace permet de représenter les performances sur l'ensemble des compromis. Cela permet de sélectionner les résultats optimaux et de comparer les algorithmes de manière beaucoup plus facile et intuitive qu'avec les courbes précision-rappel.

Dans de futurs travaux nous étudions l'aire sous les courbes de compromis (AUC) permettant de représenter par une seule valeur la performance d'un algorithme sur l'ensemble des compromis. Cependant, pour un problème donné, on ne s'intéresse généralement pas à toutes les valeurs de compromis, mais juste à un intervalle. Pour cela nous devons définir des AUC partielles ou pondérées par une distribution de probabilité des compromis envisagés.

Références

- Albatineh, A. N. et M. Niewiadomska-Bugaj (2011). Correcting jaccard and other similarity indices for chance agreement in cluster analysis. *Adv. Data Anal. Classif.* 5(3), 179–200.
- Bergmann, S., J. Ihmels, et N. Barkai (2003). Iterative signature algorithm for the analysis of large-scale gene expression data. In *Phys Rev E Stat Nonlin Soft Matter Phys*, Volume 67, pp. 031902.
- Buckland, M. et F. Gey (1994). The relationship between recall and precision. In *Journal of the American Society for Information Science*, Volume 45, pp. 12–19.
- Cheng, Y. et G. M. Church (2000). Biclustering of expression data. *Proc Int Conf Intell Syst Mol Biol* 8, 93–103.
- Datta, S. et S. Datta (2006). Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes. *BMC Bioinformatics* 7.
- Davis, J. et M. Goadrich (2006). The relationship between precision-recall and roc curves. In *Proceedings of the 23rd international conference on Machine learning, ICML '06*, pp. 233–240.
- Drummond, C. et R. C. Holte (2006). Cost curves : an improved method for visualizing classifier performance. In *Machine Learning*, pp. 95–130.
- Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letter* 27(8), 861–874.
- Flach, P. A. (2003). The geometry of roc space : Understanding machine learning metrics through roc isometrics. In *ICML*, pp. 194–201.
- Hanczar, B. et M. Nadif (2013). Precision recall space to correct external indices for biclustering. In *International Conference on Machine Learning ICML*, Volume 2, pp. 136–144.

Le compromis précision rappel

- Jain, A. K. (2010). Data clustering : 50 years beyond k-means. *Pattern Recogn. Lett.* 31(8), 651–666.
- Madeira, S. C. et A. L. Oliveira (2004). Biclustering algorithms for biological data analysis : a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 1(1), 24–45.
- Manning, C. D., P. Raghavan, et H. Schütze (2008). *Introduction to Information Retrieval*.
- Sokolova, M., N. Japkowicz, et S. Szpakowicz (2006). Beyond accuracy, f-score and roc : A family of discriminant measures for performance evaluation. In *Proceedings of the 19th Australian Joint Conference on Artificial Intelligence : Advances in Artificial Intelligence*, AI'06, pp. 1015–1021.

Summary

In many machine learning problems, the performance of algorithms is evaluated using the *emph* accuracy and *emph* callback measures. These two measures can have a very different significance depending on the context. In this paper we study the behavior of various performance indices based on the precision-recall trade-off. We propose a new performance visualization tool that is the space trade-off representing different indices based on the precision-recall trade-off. We analyze the properties of this new space and highlight its advantages over the precision-recall space.