

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