

Combinaison de classificateurs simples pour une sélection rapide de caractéristiques

Hassan Chouaib*, Florence Cloppet*
Salvatore-Antoine Tabbone**, Nicole Vincent*

*Laboratoire LIPADE, Université Paris Descartes
45, rue des Saints Pères 75006 - Paris, France
prenom.nom@mi.parisdescartes.fr,

**LORIA, Université de Nancy 2, Campus scientifique - BP 239
54 506, Vandoeuvre -les-Nancy Cedex, France
tabbone@loria.fr

Résumé. La sélection de caractéristiques est une technique permettant de choisir les caractéristiques les plus pertinentes, celles adaptées à la résolution d'un problème particulier. Les méthodes classiques présentent certains inconvénients. Par exemple, elles peuvent être trop complexes, elles peuvent faire dépendre les caractéristiques sélectionnées du classificateur utilisé, elles risquent de sélectionner des caractéristiques redondantes. Dans le but de limiter ces inconvénients, nous proposons dans cet article une nouvelle méthode rapide de sélection de caractéristiques basée sur la construction et la sélection de classificateurs simples associés à chacune des caractéristiques. Une optimisation par un algorithme génétique est proposée afin de trouver la meilleure combinaison des classificateurs. Différentes méthodes de combinaison sont considérées et adaptées à notre problème. Cette méthode a été appliquée sur différents ensembles de caractéristiques de tailles variées et construite à partir de la base de chiffres manuscrits MNIST. Les résultats obtenus montrent la robustesse de l'approche ainsi que l'efficacité de la méthode. En moyenne, le nombre de caractéristiques sélectionnées a diminué de 69,9% tout en conservant le taux de reconnaissance.

1 Introduction

Dans de nombreux domaines (vision par ordinateur, reconnaissance de formes, etc.) et dans de nombreuses applications, la résolution des problèmes se base sur le traitement de données extraites à partir des données acquises dans le monde réel, et structurées sous forme de vecteurs. La qualité du système de traitement dépend directement du bon choix de la constitution de ces vecteurs. Mais dans de nombreux cas, la résolution du problème devient presque impossible à cause de la dimensionalité trop importante de ces vecteurs ou des incohérences qui peuvent apparaître dans les données. Par conséquent, il est souvent utile, et parfois nécessaire, de réduire la dimension de l'espace de représentation à une taille plus compatible avec les méthodes de résolution, même si cette réduction peut conduire à une petite perte d'informations. Parfois, la résolution de phénomènes complexes avec des descripteurs de grande taille