

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.

Summary

Feature selection happens to be an important step in any classification process. Its aim is to reduce the number of features and at the same time to try to maintain or even improve the performance of the used classifier. The selection methods described in the literature present some limitations at different levels. For instance, some are too complex to be operated in reasonable time or too dependent on the classifier used for evaluation. Others overlook interactions between features. In this paper, in order to limit these drawbacks, we propose a fast selection method based on a genetic algorithm. Each feature is closely associated with a single feature classifier. The weak classifiers we considered have several degrees of freedom and are optimized on the training dataset. Within the genetic algorithm, the individuals who are classifier

Combinaison de classificateurs simples pour une sélection rapide de caractéristiques

subsets are evaluated by a fitness function based on a combination of single feature classifiers. Several combination operators are compared. The whole method is implemented and extensive trials are performed on the MNIST handwritten digits database. Results show how robust is our approach and how efficient is the method. On average, the number of selected features is more than 69.9% smaller than the initial set; nevertheless the recognition rate has not been decreased too much.