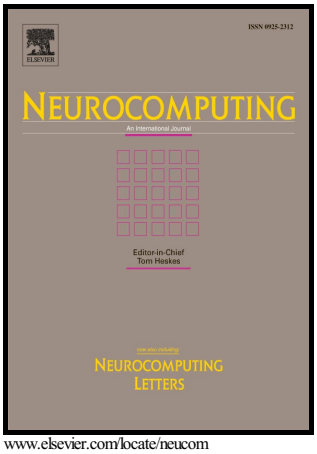

www.elsevier.com/locate/neucom

Meta-learning Recommendation of Default Size of Classifier Pool for META-DES

Anandarup Roy, Rafael M.O. Cruz, Robert Sabourin, George D.C. Cavalcanti

PII: S0925-2312(16)30834-7
DOI: http://dx.doi.org/10.1016/j.neucom.2016.08.013
Reference: NEUCOM17424

To appear in: *Neurocomputing*

Received date: 22 February 2016
Revised date: 30 July 2016
Accepted date: 1 August 2016

Cite this article as: Anandarup Roy, Rafael M.O. Cruz, Robert Sabourin and George D.C. Cavalcanti, Meta-learning Recommendation of Default Size of Classifier Pool for META-DES, *Neurocomputing* http://dx.doi.org/10.1016/j.neucom.2016.08.013

Meta-learning Recommendation of Default Size of Classifier Pool for META-DES

Anandarup Roy[a,*], Rafael M.O. Cruz[a], Robert Sabourin[a], George D. C. Cavalcanti[b]

[a] *École de Technologie Supérieure, University of Quebec, Montreal, Quebec, Canada*
[b] *Centro de Informática, Universidade Federal de Pernambuco, Recife, PE, Brazil*

Abstract

Dynamic ensemble selection (DES) is a mechanism for selecting an ensemble of competent classifiers from a *pool* of base classifiers, in order to classify a particular test sample. The size of this pool is user-defined, and yet is crucial for controlling the computational complexity and performance of a DES. An *appropriate* pool size depends on the choice of base classifiers, the underlying DES method used, and more importantly, the characteristics of the given problem. After the DES method and the base classifiers are selected, an appropriate pool size for a given problem can be obtained by the repetitive application of the DES with a variety of sizes, after which a selection is performed. Since this brute force approach is computationally expensive, researchers set the pool size to a pre-specified value. This strategy, may, however further complicate and reduce the performance of the DES method. Instead, we propose a framework that is akin to meta-learning, in order to *predict* a suitable pool size based on the intrinsic classification complexity of a problem. In our strategy, we collect meta-features corresponding to *classification complexity* from a number of data sets. Additionally, we obtain the *best pool sizes* for these data sets using the brute force approach. The association between these two pieces of information is captured using meta-regression models. Finally, for an unseen problem, we predict the pool size using this model and the classification complexity information. We carry out experiments on 65 two-class

*Corresponding author
Email addresses: roy.anandarup@gmail.com (Anandarup Roy), rafaelmenelau@gmail.com (Rafael M.O. Cruz), Robert.Sabourin@etsmtl.ca (Robert Sabourin), gdcc@cin.ufpe.br (George D. C. Cavalcanti)