

GAME: a Generic Architecture based on Multiple Experts for bioinformatics applications

Armano G⁽¹⁾, Ledda F⁽¹⁾, Vargiu E⁽¹⁾

⁽¹⁾ Department of Electrical and Electronic Engineering,
University of Cagliari, Cagliari

Motivation

The everyday work of bioinformaticians is facilitated by software tools that help them to automatize complex operations. These tools are typically intended to facilitate the use of a great variety of methods for data analysis and prediction. So far, several methods have been conceived –each characterized by specific (and often difficult to understand) recommendations and constraints regarding their installation, configuration, and use. For instance, web server tools are typically easy to understand and do not require installation, although their performance is affected by the number of requests to be fulfilled in combination with the intrinsic complexity of the tasks to be performed. Web services are an alternative way for providing easy-to-use functionalities, although they suffer from the same drawbacks that apply to web server tools as far as performance is concerned. On the other hand, web services allow orchestration, which allows to implement complex behaviors on top of simpler ones. An alternative way of allowing people to implement complex behaviors consists of providing stand-alone tools, generic architectures, or frameworks together with a full support aimed at facilitating their customization for specific tasks. In this paper, we present GAME (Generic Architecture Based on Multiple Expert), a stand-alone, portable, and easy to use generic architecture aimed at supporting the implementation of systems that rely on multiple experts. Currently, GAME has been successfully applied to secondary structure predictors and to antibody packing angle predictions.

Methods

GAME is a generic architecture tailored for bioinformatics tasks that permits to easily implement, test, refine, and release systems based on multiple experts, in an integrated environment. A system implemented on top of GAME will typically consist of software experts that interact to provide the requested functionality. Three main categories of experts can be defined and used in GAME: ground experts, refiners, and combiners. Two alternatives are feasible for ground experts: ad-hoc learners and wrappers. Ad-hoc learners currently available off-the-shelf include artificial neural networks (ANNs), support vector machines (SVMs), and hidden Markov models (HMMs). On the other hand, wrappers are specifically designed to embed automated experts that lies outside the system, including web services or external programs. As for combiners, they are entrusted with merging the prediction of several experts, whereas refiners always embed a single expert, with the goal of refining its output. As only containment relations are supported in GAME, any specific system can always be represented by a tree in which any node that is not a leaf actually embeds one or more software experts according to their reference type. In particular, refiners are represented with 1-branched nodes, whereas combiners with N-branched nodes (N being the number of experts embedded by the combiner). Of course, ground experts always occur as leaves in the tree. Using GAME mainly consists of graphically connecting and setting up experts, as well as selecting suitable encoding methods, learning techniques, and output combination methods. Different indexes for performance evaluation are available, including those based on correlation coefficients and confusion matrices. K-fold cross validation is also supported to give the possibility of improving the statistical significance of experimental results. From an implementation point of view, GAME is based on full portable Java libraries, and is released with open source license as an unpack-and-run binary bundle.

Results

To assess the usability of GAME, as well as its ability to support rapid prototyping of systems based on multiple experts, two systems have been implemented as case studies:

- i) a PHD-like secondary structure predictor and
- ii) a system for predicting antibody packing angles. As for secondary structure prediction, while experimenting a novel encoding technique, homogeneous and heterogeneous combinations of output encoding have been specifically investigated. As for antibody packing angle prediction, the corresponding system implemented with GAME has been used to study how to overcome the problem of estimating the parameters of the neural experts used to perform the prediction in presence of a limited number of training examples.

Contact : vargiu@diee.unica.it