

FunGenAgent: An Agent-Based Approach for Workflow Composition in Homology Functional Genomics

Orro A (1,2), Armano G (3), Vargiu E (3), Milanese L (1)

(1) CNR-ITB, Istituto di Tecnologie Biomediche-CNR, Segrate (Milano)

(2) CILEA, Consorzio Interuniversitario Lombardo per la Elaborazione Automatica, Segrate (Milano)

(3) Dipartimento di Ingegneria Elettrica ed Elettronica, Università di Cagliari, Cagliari

Motivation

Most tasks in bioinformatics analysis of genomics sequences cannot be carried out with a single standalone application. Most often, to solve a particular task, a combination of many computational tools and data sources is required. Due to the diversity in formats and interfaces and the low diffusion of standard methodologies for data exchange, the integration of heterogeneous computational and informative resources is a difficult task. In this work we present an agent-based approach aimed at supporting composition, execution and management of bioinformatics workflows. We also describe a preliminary implementation of this approach in the field of homology-based functional genomics.

Methods

From an application-based point of view, workflows are considered as a network of nodes, each one aimed at performing a specific application. In this way, a user has to take into account all the details involved in each application. It is clear that, as the number of applications grows, this problem becomes more and more difficult to be tackled. To this end, we propose a task oriented view in which nodes are associated to a particular task, each task being assigned to a specific agent. In our scenario, a software agent (called TaskAgent) denotes a node of the workflow and exhibits a behavior that implements a specific application. In particular, each TaskAgent provides the available resources and a suitable user interface in order to satisfy the requirements of the task and the user, respectively (see the Figure). Furthermore, each agent exhibits additional features: (1) according to its domain knowledge, it selects suitable resource connections for the corresponding task, (2) it exports only those high-level parameters that are intuitive for the user; and (3) at execution time, it can interact with the user in order to monitor the overall process.

Results

A prototype of the proposed approach (called FunGenAgent) has been implemented and tested in functional genomics applications. The overall workflow is composed by three TaskAgents: (1) homology search handler, (2) multiple alignment handler, and (3) protein function predictor. The system gives as output a vector of class propensities for each class represented in the multiple alignment. Since the search of homologue sequences has been performed using the BLAST tool, here we focus only in the remaining steps. In both cases, the corresponding agent performs complex strategies in order to (i) integrate the information, (ii) execute, and (iii) combine the output of the involved application. The TaskAgent performs the multiple alignment embodies several multiple alignment tools. First, the multiple alignment is calculated with Clustal. Then, through a postprocessing activity, the TaskAgent optimizes the alignment choosing the optimization strategies on the strength of the similarity of the sequences involved in the alignment.

In fact, in low-similarity regions, a program for calculating multiple alignment by the secondary structure; whereas in the other regions a general multiple alignment optimization is more effective (in the current implementation, we adopt RASCAL). The TaskAgent that performs functional inference embodies several tools in order to assign a functional annotation to each part of the sequences. A set of distance between the target sequence and the protein is calculated.

All distances weight the functional class taken from the function available from the database of functional family. The FunGenAgent prototype will be tested in the frame of the European Project BioinfoGRID <http://www.itb.cnr.it/bioinfoGRID> (Bioinformatics Application for Life Science) and will be available under the Italian MIUR-FIRB project LITBIO www.litbio.org (Laboratory for Interdisciplinary Technologies in Bioinformatics).

Contact email: alessandro.orro@itb.cnr.it

