Bio–Inspired Algorithms in Bioinformatics II

José Antonio Seoane Fernández
University of A Coruña, Spain

Mónica Miguélez Rico
University of A Coruña, Spain

INTRODUCTION

Our previous article presented several computational models inspired on biological models, such as neural networks, evolutionary computation, swarm intelligence, and the artificial immune system. It also explained the most common problems in bioinformatics to which these models can be applied.

The present article presents a series of approaches to bioinformatics tasks that were developed by means of artificial intelligence techniques and focus on bio-inspired algorithms such as artificial neural networks and evolutionary computation.

BACKGROUND

Previous publications have focused on the use of bio-inspired and other artificial intelligence techniques. Keedwell (2005) has summarized the foundations of molecular biology, the main problems in bioinformatics, and the existing solutions based on artificial intelligence. Baldi (Baldi, 2001) also describes various techniques for problem-solving in bioinformatics. Other generalizing works on this subject can be found in (Larrañaga, 2006), whereas more specialized works focus on solutions based on evolutionary computation (Pal, 2006) or artificial life (Das, 2007).

Bio-Inspired Techniques

The following section describes how the techniques that were mentioned in our article *Bio-inspired Algorithms in Bioinformatics I* have been used to solve the main problems in bioinformatics.

Gene Expression

We start by describing how artificial intelligence techniques have contributed to the interpretation of genes expression. Artificial neural networks (ANNs) have been applied extensively to the classification of genetic data. One of the most commonly used architectures for the classification of this type of data is the multilayer perceptron. Many works use this architecture for diagnosis (Wang, 2006) (Wei, 2005) (Narayanan, 2004) and obtain very good results; most of these approaches use artificial neural networks to discover and classify interactions between variables (genes expression values).

Statnikov (2005) and Lee (2005) compare several classification techniques, such as ANNs using backpropagation, probabilistic ANNs, Support Vector Machines (SVM), K-Nearest Neighbour (KNN), and other statistical methods for the classification of data that issue from microarrays expression tests. In this type of genetic expression data classification, we can also find a combination of ANNs and genetic programming: Ritchie (Ritchie, 2004) codifies into each individual of the genetic algorithm (GA) the architecture and weights of the network, so that the genetic programming optimizes the network to minimize the error between the output layer and the expected output, or the hybrids between the ANNs and the genetic algorithms of Kim (Kim, 2004) and Keedwell (Keedwell, 2005).

Genetic programming (GP) as such has also been used (Gilbert, 2000; Hong, 2004; Langdon, 2004; Hong, 2006) to classify the results of an expression analysis. The advantage of GP is that it classifies the genes while selecting the relevant ones (Muni, 2006). The training set of the expression data *patients* and *control* are the input for the GP algorithm, which evaluates whether or not the example is a control. The result is one or a set of classification rules. The advantage of using GP instead of other techniques such as SVM is that it is transparent: the mechanism used to classify the examples of the patients can be evaluated (Driscoll, 2003).

Whereas the above studies all classify by means of supervised learning, the following section presents various expression analysis methods for clustering that

use non-supervised learning. This type of analysis is very useful to discover gene groups that are potentially related or associated to the illness. A comparison between the most commonly applied methods, using both real and simulated data, can be found in the works of Thalamuthu (2006), Handl (2005), and Sheng (2005). Even though these methods have provided good results in certain cases (Spellman, 1998; Tamayo, 1999; Mavroudi, 2002), some of their inherent problems, such as the identification of the number of clusters, the clustering of the "outliers", and the complexity associated to the large amount of data that are being analysed, often complicate their use for expression analysis (Sherlock, 2001). These deficiencies were tackled in a series of second generation clustering algorithms, among which the self-organising trees (Herrero, 2001; Hsu, 2003).

Another interesting approach for expression analysis is the use of the artificial immune system, which can be observed in the works of Ando (Ando 2003), who applies immune recognition to classification by making the system select the most significant genes and optimize their weights in order to obtain classification rules. Finally, de Sousa, de Castro, and Bezerra apply this technique to clustering (de Sousa, 2004)(de Castro, 2001)(Bezerra, 2003).

Sequence Alignment

Solutions based on genetic algorithms, such as the SAGA (Notredame, 1996), the RAGA, the PRAGA (Notredame, 1997, 2002), and others (O'Sullivan, 2004; Nguyen, 2002; Yokohama, 2001), have been applied to sequence alignment since the very beginning. The most common method consists in codifying the alignments as individuals inside the genetic algorithm. There are also hybrid solutions that use not only GA but also dynamic programming (Zhang, 1997, 1998); and finally, there is the application of artificial life algorithms, in particular the *ant colony* algorithm (Chen, 2006; Moss, 2003).

Genetic Networks

In order to correct the problem of the inference of genetic networks, the structure of the regulating network and the interactions between the participating genes must be predicted. The expression of the genes is regulated by transitions of states in which the levels of expression of the involved genes are updated simultaneously.

ANNs have been used to model these networks. Examples of such approaches can be found in the works of Krishna, Keedwell, and Narayanan (Keedwell, 2003)(Krishna, 2005).

Genetic algorithms (Ando, 2001)(Tominaga, 2001) and hybrid RNA-genetic approaches (Keedwell, 2005) have also been used for the same purpose.

Phylogenetic Trees

Normally, exhaustive search techniques for the creation of phylogenetic trees are computationally unfeasible for more than 10 comparisons, because the number of possible solutions increases exponentially with the number of objects in the comparisons. In order to optimize these searches, researchers have used heuristics based on genetic algorithms (Skourikhine, 2000)(Katoh, 2001)(Lemmon, 2002) that allow the reconstruction of the optimal trees with less computational load. Other techniques, such as the ant colony algorithm, have also been used to reconstruct phylogenetic trees (Ando, 2002)(Kummorkaew, 2004) (Perretto, 2005).

Gene Finding and Mapping

Gene mapping has been approached by methods that use only genetic algorithm (Fickett, 1996)(Murao, 2002) as well as by hybrid methods that combine genetic algorithms and statistical techniques (Gaspin, 1997).

The problem of gene searching and in particular promoter searching has been approached by means of neural networks (Liu, 2006), neural networks optimized with genetic algorithms (Knudsen, 1999), conventional genetic algorithms (Kel, 1998)(Levitsky, 2003),and fuzzy genetic algorithms (Jacob, 2005).

Structure Prediction

The tridimensional structure of DNA was predicted with genetic algorithms (Beckers, 1997) by codifying the torsional angles between the atoms of the DNA molecule as solutions of the genetic algorithm. Another approach was the development of hybrid strategies of ANNs and GAs (Parbhane, 2000), in which the network approaches the non-linear relations between the inputs and outputs of the data set, and the genetic algorithm searches within the network inputs space to optimize the output. In order to predict the secondary structure of the RNA, the system calculates the minimum free

Related Content

A Fuzzy Expert System for Star Classification Based on Photometry
Aida Pakniyat, Rahil Hosseiniand Mahdi Mazinai (2016). *International Journal of Fuzzy System Applications (pp. 109-119).*
www.irma-international.org/article/a-fuzzy-expert-system-for-star-classification-based-on-photometry/162667

A Review of Machine Learning Methods Applied for Handling Zero-Day Attacks in the Cloud Environment
 Swathy Akshaya M.and Padmavathi Ganapathi (2020). *Handbook of Research on Machine and Deep Learning Applications for Cyber Security (pp. 364-387).*
www.irma-international.org/chapter/a-review-of-machine-learning-methods-applied-for-handling-zero-day-attacks-in-the-cloud-environment/235050

Visual Analysis of a Large and Noisy Dataset
Honour Chika Nwagwuand Constantinos Orphanides (2015). *International Journal of Conceptual Structures and Smart Applications (pp. 12-24).*
www.irma-international.org/article/visual-analysis-of-a-large-and-noisy-dataset/152376

Design and Usage of a Process-Centric Collaboration Methodology for Virtual Organizations in Hybrid Environments
Thorsten J. Dollmann, Peter Loos, Michael Fellmann, Oliver Thomas, Andreas Hoheisel, Peter Katranuschkovand Raimar Scherer (2011). *International Journal of Intelligent Information Technologies (pp. 45-64).*
www.irma-international.org/article/design-usage-process-centric-collaboration/50485

Digital Twin-Based Smart Healthcare Services for the Next Generation Society
V. Hemamalini, Firas Armoshand Amit Kumar Tyagi (2024). *Future of AI in Medical Imaging (pp. 247-277).*
www.irma-international.org/chapter/digital-twin-based-smart-healthcare-services-for-the-next-generation-society/342040