

Chapter 63

Efficient Computational Construction of Weighted Protein–Protein Interaction Networks Using Adaptive Filtering Techniques Combined with Natural Selection–Based Heuristic Algorithms

Christos M. Dimitrakopoulos
University of Patras, Greece

Spiridon D. Likothanassis
University of Patras, Greece

Konstantinos A. Theofilatos
University of Patras, Greece

Athanasios K. Tsakalidis
University of Patras, Greece

Efstratios F. Georgopoulos
Technological Educational Institute of Kalamata,
Greece

Seferina P. Mavroudi
University of Patras, Greece

ABSTRACT

The analysis of protein-protein interactions (PPIs) is crucial to the understanding of cellular processes. In recent years, a variety of computational methods have been developed to supplement the interactions that have been detected experimentally. The article's main objective is to present a novel classification framework for predicting PPIs combining the advantages of two algorithmic methods' categories. State-of-the-art adaptive filtering techniques were combined with the most contemporary heuristic methods which are based in the natural selection process. The authors' goal is to find a simple mathematical equation that governs the best classifier enabling the extraction of biological knowledge. The proposed methodology assigns a confidence score to each protein pair and as a result a weighted PPI network is constructed. All possible combinations of the selected adaptive filtering and heuristic techniques were used and comparisons were made to explore the classifiers with the highest performance and interpretability.

DOI: 10.4018/978-1-4666-3604-0.ch063

INTRODUCTION

In each living cell of the human organism, a variety of protein interactions (PPIs) take place. In recent years, researchers have tried to approach the problem of predicting all possible protein interactions in the human organism by implementing different computational techniques. At the beginning, most of them were based on the analysis of a sole feature, indicative of interaction between two proteins. Several examples of such features are features concerning the genomic sequence of the genes-generators of the reference proteins, features concerning the structure of the reference proteins, features concerning the sequences of the references proteins and many others (Chua et al., 2006). The most recent computational approaches use various features as inputs for their classifiers in order to take advantage of all the available information (Chen et al., 2005; Fariselli et al., 2002).

Bayesian based classifiers (Howson et al., 1993) are the most common computational methods used to integrate data from a wide variety of sources. Scott et al. (2007) presented a hybrid approach of naïve Bayesian and a full Bayesian classifier. Using the full classifiers they produced a combined feature from the features of co-localization, post translational modifications co-occurrence and domain co-occurrence. In a subsequent step, that feature was combined with a co-expression feature, an orthology feature, a co-disorder feature and a network topology feature using a Bayesian classifier. Their methodology was applied to predict and rank the human PPIs genome-wide. The probabilistic framework offered by Bayesian approaches is capable of producing interpretable classifiers of adequate classification performance and it allows the incorporation of experts' knowledge. On the contrary, its simplicity leads to performance limitations and forces researchers to apply more sophisticated techniques in the problem of predicting protein interactions.

Two machine learning methods widely used in the problem of PPIs prediction are the Artificial

Neural Networks (Haykin, 1998) and the Support Vector Machines (Corrina & Vapnik, 1995). They both have been applied in the past in an enormous variety of classification problems providing very high classification performances. Chen et al. (2006) developed an integrative Artificial Neural Network framework to predict PPIs from heterogeneous data in Human. They used diverse data sources - like protein domain data, molecular function data and biological process annotations - to carry out the prediction. Although Artificial Neural Networks, when applied to the problem of predicting PPIs, have demonstrated very good classification measures, more sophisticated techniques like SVMs, seem to outperform them because of their higher generalization abilities. Bock et al. (2001) used a SVM learning system for training interaction data, with protein sequences and associated physicochemical properties as features. For each amino acid sequence of a protein complex, feature vectors were assembled from encoded representations of several tabulated residue properties, such as charge, hydrophobicity, and surface tension for each residue in sequence. SVMs have proven to provide high prediction accuracy, but to the expense of increased computational cost (Gomez et al., 2003). Hence, it is usually unfeasible to train the SVM with a relatively high collection of training examples. Moreover, the results derived from that method cannot be easily interpreted by biologists.

In the present article, we tried to achieve high performance classification with classifiers that are fully interpretable. We applied to the problem of PPI prediction several adaptive filtering techniques combined with the most contemporary heuristic methods which are based in the natural selection process. From this combination, a novel computational framework has been formed for the creation of a mathematical equation that gives the best classifier. We consider three classical parameter estimation algorithms (LMS, RLS and Kalman Filter) and one gain adaptation algorithm (IDBD). The gain adaptation algorithms have

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/efficient-computational-construction-weighted-protein/76116

Related Content

Improving Prediction Accuracy via Subspace Modeling in a Statistical Geometry Based Computational Protein Mutagenesis

Majid Masso (2010). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 54-68).

www.irma-international.org/article/improving-prediction-accuracy-via-subspace/49549

Improving PSI-BLAST's Fold Recognition Performance through Combining Consensus Sequences and Support Vector Machine

Ren-Xiang Yan, Jing Liu and Yi-Min Tao (2013). *Bioinformatics: Concepts, Methodologies, Tools, and Applications* (pp. 1667-1675).

www.irma-international.org/chapter/improving-psi-blast-fold-recognition/76140

Analysis of microRNA Regulated Seed Biology Networks in Arabidopsis

Anamika Basu, Anasua Sarkar and Piyali Basak (2014). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 11-20).

www.irma-international.org/article/analysis-of-microrna-regulated-seed-biology-networks-in-arabidopsis/147301

Using Case Costing Data and Case Mix for Funding and Benchmarking in Rehabilitation Hospitals

Grace Liu (2014). *Research Perspectives on the Role of Informatics in Health Policy and Management* (pp. 62-78).

www.irma-international.org/chapter/using-case-costing-data-and-case-mix-for-funding-and-benchmarking-in-rehabilitation-hospitals/78689

Comparative Genome Annotation Systems

Kwangmin Choi and Sun Kim (2006). *Advanced Data Mining Technologies in Bioinformatics* (pp. 296-313).

www.irma-international.org/chapter/comparative-genome-annotation-systems/4258