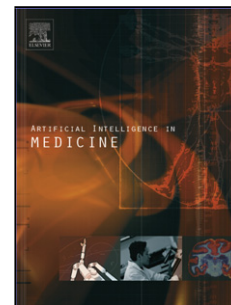


Accepted Manuscript

Title: Improved prediction of protein–protein interactions using novel negative samples, features, and an ensemble classifier

Authors: Leyi Wei, Pengwei Xing, Jiancang Zeng, JinXiu Chen, Ran Su, Fei Guo



PII: S0933-3657(16)30569-3
DOI: <http://dx.doi.org/doi:10.1016/j.artmed.2017.03.001>
Reference: ARTMED 1507

To appear in: *ARTMED*

Received date: 20-12-2016
Revised date: 17-2-2017
Accepted date: 1-3-2017

Please cite this article as: Wei Leyi, Xing Pengwei, Zeng Jiancang, Chen JinXiu, Su Ran, Guo Fei. Improved prediction of protein–protein interactions using novel negative samples, features, and an ensemble classifier. *Artificial Intelligence in Medicine* <http://dx.doi.org/10.1016/j.artmed.2017.03.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Improved prediction of protein–protein interactions using novel negative samples, features, and an ensemble classifier

Leyi Wei¹, Pengwei Xing¹, Jiancang Zeng², JinXiu Chen², Ran Su³, and Fei Guo^{1*}

1. School of Computer Science and Technology, Tianjin University, Tianjin, China

2. School of Information Science and Technology, Xiamen University, Xiamen, China

3. School of Software, Tianjin University, Tianjin, China

*Corresponding Author

Highlights

1. A novel strategy of negative sample construction is proposed.
2. Two novel types of feature extraction methods are proposed.
3. Our machine learning based prediction model exhibits the efficiency and effectiveness to predict protein-protein interactions, significantly reducing false negatives.

Abstract

Computational methods are employed in bioinformatics to predict protein–protein interactions (PPIs). PPIs and protein–protein non-interactions (PPNIs) display different levels of development, and the number of PPIs is considerably greater than that of PPNIs. This significant difference in the number of PPIs and PPNIs increases the cost of constructing a balanced dataset. PPIs can be classified as either physical or genetic. However, ready-made PPNI databases were proven only to have no physical interactions and were not proven to have no genetic interactions. Hence, ready-made PPNI databases contain false negative non-interactions. In this study, two PPNI datasets were artificially generated from a PPI database. In contrast to various traditional PPI feature extraction methods based on sequential information, two types of novel feature extraction methods were proposed. One is based on secondary structure information, and the other is based on the physicochemical properties of proteins. The experimental results of the RandomPairs dataset validate the efficiency and effectiveness of the proposed prediction model. These results reveal the potential of constructing a PPI negative dataset to reduce false negatives. Related datasets, tools, and source codes are accessible at <http://lab.malab.cn/soft/PPIPre/PPIPre.html>.

Download English Version:

<https://daneshyari.com/en/article/6853395>

Download Persian Version:

<https://daneshyari.com/article/6853395>

[Daneshyari.com](https://daneshyari.com)