

=====

HP-GAS

=====

Copyright (c) "2018, by Institute of Nuclear Sciences VINCA, Belgrade, Serbia
Contributors: Neven Sumonja, Branka Gemovic, Nevena Veljkovic
and Vladimir Perovic
Affiliation: Laboratory for Bioinformatics and Computational Chemistry
URL: <http://www.vin.bg.ac.rs/180/tools/HP-GAS.php>

=====

1. Description

HP-GAS is a software for prediction of human protein protein interactions (PPI) based on graph, evolutionary and sequence features. HP-GAS uses the ensemble of models generated by machine learning (ML) algorithms as a method for PPI prediction, where automatic ensembling of ML algorithms was driven by genetic algorithm and feature engineering was performed by automatic correlation based selection.

2. Requirements

HP-GAS is written in JAVA language and is available as standalone application, which can be executed on any operating system containing Java Virtual Machine.

Minimum system requirements for HP-GAS are:

RAM: 1 GB
Disk space: 1 GB

3. Installation

HP-GAS software is available for download at:

<http://www.vin.bg.ac.rs/180/tools/HP-GAS.php>

In order to run the program it is necessary to install Java Runtime Environment 8 (JRE), which can be found for Windows, Linux, Mac OS and Solaris systems at:

<https://www.oracle.com/technetwork/java/javase/downloads/jre8-downloads-2133155.html>.

4. Usage

HP-GAS requires file 'input.txt' to be located in the folder where executable and additional required files are. The input file must contain the pairs in each line of human protein identifiers, UniProt entry name (1433B_HUMAN) or UniProt accession number (P31946), separated by tab character.

The program can be invoked from the command line as:

`java -jar HP-GAS.jar input.txt output.txt`

or to increase the memory needed for the java VM:

`java -Xmx2G -jar HP-GAS.jar input.txt output.txt`

The calculated prediction values with tested id pairs will be written in 'output.txt' file.

5. Notice

HP-GAS machine learning classifiers were generated using the H2O platform, developed by H2O.ai, Inc. licensed under the Apache License, Version 2.0.

6. Files

This release of the HP-GAS program contains following files:

- 'HP-GAS.jar': java jar file for calculating prediction values
- 'codes.zip': required additional file with codes and calculated features
- 'model_1.zip', 'model_2.zip' and 'model_3.zip': required additional files with ML models used in ensemble
- 'input.txt': example of input file with pairs of protein identifiers; each line have one pair of proteins for evaluation; identifiers are separated by tab character; identifiers can be entry name or accession number
- 'output.txt': output of the calculation; it can be used to check and compare the results of the calculations for the input example; each line have the tested protein pair and the predicted interaction probability [0-1]
- 'HP-GAS.bat': batch file for Windows system which execution will automatically generate predictions for the data in input file 'input.txt' and write them in 'output.txt' file
- 'HP-GAS.sh': script for Linux system
- 'LICENSE.txt': license for use of HP-GAS tool as a Free Software, under the Apache License, Version 2.0
- 'NOTICE.txt': notice of using the libraries from H2O platform developed by the H2O.ai Inc.