Protein-protein interaction based on pairwise similarity

By Nazar Zaki, Sanja Lazarova-Molnar, Wassim El-Hajj, Piers Campbell

Data and programs for experiment no. 2.
To run in **Cygwin**

Files and programs needed:
The following files are needed and should be included in the same directory.

o **fasta34.exe** downloadable from http://www.ebi.ac.uk/Tools/fasta/index.html
o **gist-train-svm.exe** and **gist-classify.exe** from Gist software (Windows-Cygwin) downloadable from http://bioinformatics.ubc.ca/gist/download.html

Experimental 2

In our second experimental work, we assess the recognition ability of our method to classify between 100 interacted protein pairs (157 proteins) and 100 non-interacted protein pairs (77 proteins).

The dataset was randomly selected by Sylvain et. al[1] and used to evaluate PIPE's accuracy. It was generated from the yeast protein interaction literature for which at least three different lines of experimental evidence supported the interaction.

To run this experiment, please type the following in the command line:
$ perl train.pl [interaction_file] [sequence_file] [n] [pos]

interaction_file	= File contains the protein pairs.
sequence_file	= File contains the protein sequences.
n	= the window size, (n = 2, …, 20,000).
pos	= number of interacted proteins in the dataset.

Example:
perl train.pl data_ex2/all_int.txt data_ex2/all_seq_200.txt 500 100

Results:
Training results: FP = 9 FN = 3 TP = 97 TN = 91
Training ROC: 0.98200
Test results: FP = 9 FN = 8 TP = 92 TN = 91
Test ROC: 0.96350
*****
RFP is: 0.09
Sensitivity is: 0.92
Specificity is: 0.91
Precision is: 0.910891089108911
F value is: 0.910445326518406
Overall Accuracy is: 0.915

---

[1] Sylvain P., Frank D., Albert C., Jim C., Alex D., Andrew E., Marinella G., Jack G., Mathew J. Nevan K. Xuemei L. and Ashkan G. (2006) PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs. BMC Bioinformatics.