Evaluating Multi-level Machine Learning Prediction of Protein-protein Interactions

Julian Zubek[1,2], Marcin Tatjewski[1,2], Subhadip Basu[3] and Dariusz Plewczynski[2]

[1] Institute of Computer Science, Polish Academy of Sciences, ul. Jana Kazimierza 5, 01-248 Warsaw, Poland
[2] Centre of New Technologies, University of Warsaw, ul. Stefana Banacha 2c, 02-097 Warsaw, Poland
[3] Department of Computer Science and Engineering, Jadavpur University, 188, Raja S. C. Mallick Road, Kolkata, West Bengal, India

Abstract. We introduce a novel procedure for evaluating prediction of protein-protein interactions. It takes into account fact that pairwise protein interactions form a larger interaction network. Our procedure guarantees that: a) true positives and true negatives of interacting proteins are formed from the same elements (i.e. they have identical protein composition), b) there is strict separation of proteins between training and test sets. This procedure was applied to previously developed MLPPI (Multi-level machine learning prediction of protein-protein interactions) method and established sequence-based methods. We performed evaluation on high-quality small and medium size data sets containing protein interactions from *Saccharomyces cerevisiae*, *Homo sapiens*, and *Escherichia coli*. Poor performance of all methods (AUC ROC below 0.6) raises a question whether the goal of protein-protein interaction prediction was correctly formulated.

Experimental code and data freely available at:
`http://zubekj.github.io/mlppi/`
(Python implementation, OS independent).

1 Introduction

Proteins are among the most important building blocks of living cells. They are compound objects which can be described in multiple scales: protein primary structure is a linear (1D) sequence of amino acids residues, secondary structure is a sequence of characteristic structural motifs formed along protein chain, and tertiary structure is a full 3D structure of a protein molecule. Interactions between proteins form complex signalling networks, which needs to be reconstructed in as much details as possible in order to understand properties of living organisms at the system level [12]. Various computational tools based

on machine learning are being developed to facilitate this process. Most of these tools focus on predicting binary interactions between pairs of proteins. Among them methods using only 1D protein sequence have the widest applicability, since this kind of information is available for all known proteins. Assessment and comparison of performance of such methods in a realistic setting is a non-trivial task which requires special considerations.

In this work we focus on a thorough evaluation of a multi-level machine learning method for predicting protein-protein interactions, which was developed by Zubek et al. [17]. It differs significantly from the established sequence-based methods because it uses residue-residue interaction prediction as an intermediate step during protein-protein interaction prediction (hence it is called a multi-level approach). In such fashion it introduces 3D structural information during classifier training but utilises only 1D sequence during prediction. The impact of this approach on prediction quality was not yet evaluated properly: so far the method was tested only on a relatively small subset of proteins from *Saccharomyces cerevisiae*. The goal of this work is to compare the performance of multi-level method by Zubek et al. [17] with that of classic sequence-based methods [10] in a realistic setting using larger and more diverse sets of proteins from different organisms: *Saccharomyces cerevisiae* (Yeast), *Homo sapiens* (Human), and *Escherichia coli* (the organisms were chosen based on the availability of the data). In order to meet our goal we develop a novel evaluation schema, which measures predictive power in the context of detecting real compatibility between previously unseen proteins. We construct a balanced set of true negative interactions using interaction network properties. We calculate the performance metrics using modified multi-level cross-validation schema, which takes into account internal structure of the classified objects. This approach allows to avoid a common problem in the evaluation of classifiers operating on compound objects, when the same components occur in different quantities in training and test set [11]. Our hypotheses are that: a) introducing indirectly 3D information in the multi-level classifier is beneficial for its performance, b) our evaluation schema reflects the real difficulty of protein-protein interaction prediction better than a näive approach which often overestimate classifier performance.

The decision to focus on prediction utilising protein primary structure and do not include methods based on protein functions [14, 13] in our comparison needs justification. We believe that those two types of prediction methods have different areas of application. First, functional features are generally available only for a subset of proteins from well studied organisms. Second, this kind of description is strongly dependent on biological pathways, which may be highly specific for a given organism. With functional features we are targeting high-level evolutionary designed mechanisms, while with sequence-based features we can hope to uncover basic physical properties of proteins, which govern their interactions. Knowing those properties it would be possible to predict protein interactions across different organisms and include some specific cases which distort normal protein interaction networks, such as host-patogen protein interactions.

In our experiments we obtained performance estimates much lower than usually reported, which is in line with our hypothesis. The multi-level approach was only marginally better than other methods. Our results raise the need to re-evaluate the usefulness of sequence-based features for protein interaction prediction. Lack of success of both standard methods aggregating global features of the sequence and the multi-level approach, which looks at the individual residues, suggests that protein interactions may be a phenomenon occurring primarily on a higher level and involving whole protein structures.

2 Materials and methods

2.1 Protein interactomes

We evaluated prediction methods by building classifiers separately for three organisms: *S. cerevisiae*, *H. sapiens*, and *E. coli*. Interactomes of all these organisms are relatively well studied, however reconstructed protein interaction networks are still far from being complete. For all three organisms we extracted 3D protein crystal complexes from Protein Data Bank (PDB) [2]. We were interested in complexes scanned with X-RAY with the resolution below 3 Å. Homologous structures were removed with 90% sequence identity threshold. The remaining complexes were used as a reliable source of information on residue-residue interactions (RRI) and protein-protein interactions (PPI). Residue-residue interaction is defined as a pair of amino acid residues from two different protein chains which are located within a close distance (4 Å) in the crystal structure. Protein-protein interaction is a pair of proteins for which at least one residue-residue interaction occurs. Only pairwise heterogenous protein interactions involving two different proteins were of interest to us.

In the work by Zubek et al. [17] a special procedure was used to filter RRIs and keep only the strongest interactions. The sliding window was moved along protein sequence and centred on each interacting residue. The window covered 21 residues – one central interacting residue, 10 residues to the left from it, and 10 residues to the right of it. Then the number of all interacting residues (including the central one) within the window was counted. Only when this number exceed certain threshold value the central residue was considered strongly interacting. We replicated this procedure in this work and set the threshold value to 15 (this value was reported as an optimal in the original publication).

Relatively small sets of PDB-derived PPIs were complemented with large scale data curated by Saha et al. [14]. They provided PPIs for *S. cerevisiae* and *H. sapiens* in two flavors: GOLD dataset contained only interactions which were confirmed independently with two different experimental methods, SILVER contained interactions reported by two different sources (possibly using the same experimental method). For *S. cerevisiae* and *H. sapiens* we used the available GOLD datasets. For *E. coli* we constructed our own SILVER dataset using iRefWeb interface [16].

PDB-based data sets were split into training and test set on the protein level (no protein occurred simultaneously in the two sets). Numbers of PPIs in each

set are given by Table 1. Proteins occurring in *S. cerevisiae* PDB training were removed from *S. cerevisiae* GOLD test, proteins occurring in *H. sapiens* PDB training were removed from *H. sapiens* GOLD test, and proteins occurring in *E. coli* PDB training were removed from *E. coli* SILVER. Because data available in PDB for *E. coli* was less abundant than for the other organisms, we did not construct a PDB-based test set, using *E. coli* SILVER as the only validation of prediction performance.

Table 1. Number of interacting protein pairs in the collected data sets.

Data set	Training RRI	Training PPI	Test PPI
S. cerevisiae GOLD	-	-	1284
S. cerevisiae PDB	5531	211	174
H. sapiens GOLD	-	-	1325
H. sapiens PDB	2774	195	204
E. coli SILVER	-	-	2763
E. coli PDB	1698	61	-

2.2 True negatives

The available experimental data identifies only positive interactions. True negative interactions for training machine learning classifier need to be artificially generated. Generating high quality negatives is generally very difficult. For RRIs we selected sequence fragments from known protein complexes such that not a single RRI occurred on those fragments. As the data was abundant and the risk of generating a false negative by chance was low, we generated 10 times more RRI negatives than the collected positives. This was done to represent class imbalance expected in real data.

The problem was more complicated for PPIs. Common methods for generating negatives include drawing random pairs of biomolecules from all known proteins found in a specific organism [14], or from the subset of whole proteome constituted by proteins occurring in positive examples [4]. We strongly believe that such methods have their inherent drawbacks, because they ignore network properties of the underlying protein interactome. Imagine that we have a set of 9 positive PPIs over 10 proteins which form a star subgraph in the interactome. The central protein in this subgraph has 9 interactions, the rest of the proteins have 1 interaction each. Then we generate negatives by drawing random pairs of these 10 proteins with equal probability. For each protein the probability of being included in a formed pair is equal to: $\frac{1}{10} + \frac{9}{10} \cdot \frac{1}{9} = 0.2$. If we draw 9 pairs, the expected number of negative interactions for each protein is 1.8. In such setting a classifier which recognises any pair containing the central protein

as positive automatically reaches 0.83 precision score, even though it completely ignores relative compatibility between two proteins. This scenario is biologically realistic, in real interactomes occurrence of a significant number of hub proteins is reported [15]. What is more, when the proteins are paired completely randomly there is always a risk of generating a false negative, i.e. previously undiscovered interaction. Because of this larger number negatives lower the quality of data.

As an alternative to uniform sampling we propose the following procedure:

- Let G_1 be a graph representing positive examples. Denote $V = v_1, \ldots, v_n$ as the set of its vertices. Each vertex in V represents a protein and each edge v_i, v_j represents an interaction. Let $[Deg(v_1), \ldots, Deg(v_n)]$ be a vector containing degrees of vertices from V. Let G_2 be a graph of negative interactions. At first it has vertices identical to G_1 and no edges.
- While there exist v such that $Deg(v) > 0$:
  1. Find vertex v with the largest $Deg(v)$.
  2. Find vertex u if exist such that:
     (a) There is no edge (v, u) in G_1.
     (b) u has as large $Deg(v)$ as possible.
     (c) Distance $d(u, v)$ in G_1 is as large as possible.
  3. If u exist:
     (a) Add edge (u, v) to G_2.
     (b) $Deg(v) \leftarrow Deg(v) - 1$
     (c) $Deg(u) \leftarrow Deg(u) - 1$
  4. else: $Deg(v) \leftarrow 0$

Such schema of constructing the negatives is unbiased, i.e. the protein composition of the positives and the negatives remains identical. Every single protein has the same number of positive and negative interactions. This forces the trained classifier to predict meaningful biophysical interactions rather than predicting general reactivity (the relative number of interactions) of a single protein. What is also important, our algorithm favours protein pairs which are remote to each other the interaction network, which reduces – but does not eliminate – the risk of generating false negatives by chance. Using the described procedure we generated the same number of negative PPIs as positive ones, thus obtaining a balanced dataset.

2.3 Multi-level prediction of protein-protein interactions

We were interested in benchmarking the multi-level method developed by Zubek et al. [17]. We will refer to it as MLPPI. It performs a two-stage prediction, first predicting RRIs and then using the results to predict PPIs. RRI classifier operates on sequence fragments of length 21 amino acid residues. Two fragments sliced from two proteins sequences constitute a single observation. Sliding window technique is used to extract fragments centred on each residue in a sequence. The result is a two dimensional matrix with dimensions corresponding to proteins' lengths. It can interpreted as a predicted potential contacts map.

This matrix is processed with various feature extraction algorithms to produce a fixed-length input for PPI classifier. General outline of this method is presented as Figure 1.

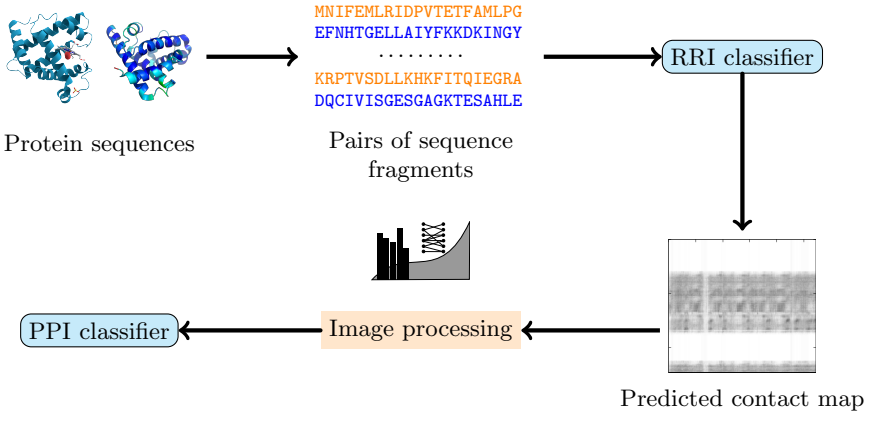

Fig. 1. Schematic depiction of the multi-level protein interaction prediction pipeline.

For classification on both levels we used Random Forest algorithm with 300 trees and maximum tree depth limited to 7 nodes. Sequence fragments which constituted input for RRI classifier were encoded using secondary structure symbols predicted from sequence by PSIPRED [8]. Features extracted from the predicted contact map to form an input for PPI classifier included:

- the mean and variance of values over the matrix (2),
- the sums of values in 10 best rows and 10 best columns (20),
- the sums of values in 5 best diagonals of the original and the transposed matrix (10),
- the sum of values on intersections of 10 best rows and 10 best columns (1),
- the histogram of scores distributed over 10 bins (10),
- features of the connection graph: fraction of nodes in the 3 largest connected components (3).

Features of the connection graph require further explanation. Predicted contacts between residues were represented as a bipartite graph. Nodes in the graph represented residues and edges represented predicted contact. To make the graph more consistent with the observed experimental data, for each node we left only 3 strongest outgoing edges. We set the value of this threshold (3) following the observation that in our PDB structures the mean number of interactions of a single interacting residue is between 2 and 3. In such trimmed graph we calculated fractions of nodes contained in 3 largest connected components. Those values were also appended to the feature vector.

2.4 Sequence-based methods

We compared our ensemble method with various sequence feature aggregation schemas that are commonly used to construct features for machine learning classifiers of protein interactions. To make the benchmarking results comparable between different algorithms, we used the same classification method (Random Forest) as for the MLPPI classifier. We benchmarked the following feature aggregation schemas:

1. AAC – Amino Acid Composition [10]. Feature set is the set of frequencies of all amino acids in the sequence.
2. PseAAC – Pseudo Amino Acid Composition [5]. Feature set consists of the standard AAC features with k-th tier correlation factors added. The k-th tier correlation factor represent correlation for residues separated from each other by k residues. We calculate those correlations on HQI8 indices.
3. 2-grams [10]. Feature set comprises of frequencies of all 400 ordered pairs of amino acids in the sequence.
4. QRC – Quasiresidue Couples [7]. A set of AAIndices is chosen. For each index d combined values of this property d for a given amino acid pair are summed up for all the pair's occurrences over the full protein sequence. Occurrences for pairs of residues separated from each other by $0, 1, 2 \ldots m$ residues. In effect, one obtains QRC^d vectors of length $400 \times m$. In this model we also use HQI8 indices.
5. VD – vector deviations, a variation of Liu's protein pair features [9]. The method starts from encoding each amino acid in a protein sequence with 7 chosen physicochemical properties, thus obtaining 7 feature vectors for each sequence. For each feature vector its "deviation" is calculated:

$$\gamma_{dj} = \frac{1}{n-d} \sum_{i=1}^{n-d} x_{ij} \times x_{(d+i),j} \quad j = 1, \ldots, 7 \quad d = 1, \ldots, L$$

where x_{ij} is the value of descriptor j for amino acid at position i in sequence P, n is the length of protein sequence P, and d is the distance between residues in the sequence. For the purpose of the comparison, we tested this method with the original 7 amino acid indices used by Liu. We tested different values of L from 5 to 30 in a quick cross-validation experiment on our data and chose $L = 9$ as yielding the best results.

2.5 Evaluation procedure

Created RRI training, PPI training, and PPI test data sets had their specific purposes. RRI training and PPI training data was used only to train RRI classifier in MLPPI. It was not used by any other method. Then, all sequence-based PPI classifiers and the PPI classifier of MLPII were trained and evaluated using PPI test data.

Performance of PPI classifiers was evaluated through a repeated 2-fold cross-validation (split between two folds of equal size). However, splitting data on

the level of individual observations was unsatisfactory because training and test sets could still overlap on the protein level, which introduced a huge bias into evaluation results (similar observation was previously made by [11]). To fix this problem we decided to perform split on the protein level. We used the following algorithm:

1. Let X be a set of all observations (protein pairs), P set of all proteins, X_A, X_B observations in the two splits, P_A, P_B protein in the two splits.
2. Initialise set $X_A \leftarrow \emptyset$ to empty set, $X_B \leftarrow X$.
3. While $|X_A| < |X_B|$ repeat:
   (a) Add a random observation x not included in X_A to X_A.
   (b) Complement X_A with all observations $x = (x^1, x^2)$ such that $x^1 \in P_A$ and $x^2 \in P_A$.
   (c) Let $X_B = \{(x^1, x^2) : x^1 \notin P_A \wedge x^2 \notin P_A\}$.

The relations between all datasets used in the evaluation procedure are depicted by Figure 2.

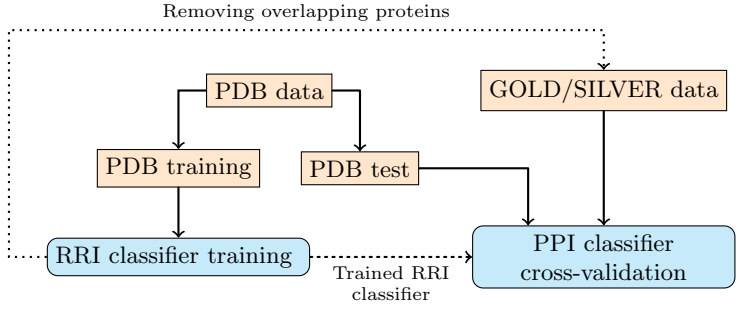

Fig. 2. Schematic depiction of relations between different data sets in the evaluation procedure.

The above described procedure differs from the standard cross-validation, since the number of observations in constructed test sets vary slightly, but this variance is small, and does not influence the estimated performance. Such evaluation schema does not allow for any information leak: the datasets are always balanced, and the classifier is tested on previously unseen proteins.

Using this form of cross-validation reduced the effective size of training and test data, because in each split some observations need to be dropped. Average size of a single cross-validation fold for all data sets is given in Table 2.

3 Results and discussion

We repeated cross-validation split 5 times and calculated average AUC ROC (area under the receiver operating characteristic curve) over splits and folds.

Table 2. Average size of a single cross-validation fold with estimated standard deviation.

Data set	CV fold size
S. cerevisiae GOLD	723 ± 7
S. cerevisiae PDB	116 ± 3
H. sapiens GOLD	763 ± 9
H. sapiens PDB	139 ± 3
E. coli SILVER	1591 ± 24

Results for different methods are presented as Table 3. As can be seen, AUC ROC values are generally very low, never exceeding 0.6. This means that under the conditions imposed by our strict evaluation procedure none of the methods was especially successful. This is especially true for *E. coli* SILVER data set for which the performance of all methods is at the level of a random baseline. Because of this we excluded *E. coli* data set from further analyses.

Table 3. AUC ROC (Area under the receiver operating characteristic curve) score for different methods. MLPPI – Multi-level Prediction of Protein Interactions, VD – vector deviations, AAC – amino acid composition, PseudoAAC – pseudo amino acid composition, 2-grams – bigram frequencies, QRC – quasiresidue couples.

Data set	MLPPI	VD	AAC	PseudoAAC	2-grams	QRC
E. coli SILVER	0.50	**0.51**	0.51	0.50	0.49	0.48
S. cerevisiae GOLD	**0.57**	0.56	0.55	0.54	0.51	0.52
S. cerevisiae PDB	**0.59**	0.52	0.52	0.54	0.47	0.47
H. sapiens GOLD	**0.56**	0.53	0.53	0.54	0.51	0.52
H. sapiens PDB	**0.56**	0.49	0.53	0.53	0.52	0.53

To establish statistical differences between methods we employed combined 5x2cv F test proposed by Alpaydin [1]. It is a modified version of 5x2cv t test introduced by Dietterich [6]. It strives to exploit the benefits of multiple train-test splits while minimising the bias introduced by lack of independence between splits. Each split i contains two cross-validation folds, which results in two values $p_i^{(1)}$ and $p_i^{(2)}$ which are the differences between scores obtained by two methods. They can be used to estimate mean and variance for each split separately:

$$\bar{p}_i = \frac{p_i^1 + p_i^2}{2}$$

$$s_i^2 = \left(p_i^{(1)} - \bar{p}\right)^2 + \left(p_i^{(2)} - \bar{p}\right)^2$$

The test statistic f has the following form:

$$f = \frac{\sum_{i=1}^{5} \sum_{j=1}^{2} \left(p_i^{(j)}\right)^2}{2 \sum_{i=1}^{5} s_i^2}$$

Under the null hypothesis, when two methods have identical performance, the f statistic is F distributed with 10 and 5 degrees of freedom.

Table 4. f statistic values and p-values for tests comparing MLPPI against the best performing sequence-based method.

Data set	Test	f	p-value
S. cerevisiae GOLD	MLPPI vs VD	1.637	0.250
S. cerevisiae PDB	MLPPI vs PseudoAAC	4.873	0.020
H. sapiens GOLD	MLPPI vs PseudoAAC	2.074	0.160
H. sapiens PDB	MLPPI vs AAC	2.604	0.097

We wanted to check whether our multi-level method performed better than methods aggregating global characteristics of protein sequence. On each data set separately we tested MLPPI against the best performing sequence-based method. Test statistic values and p-values are given by Table 4. Although MLPPI had the best AUC score on all four data sets, the difference was significant at $\alpha = 0.05$ level only for *S. cerevisiae* PDB – the data set on which MLPPI method was initially devised and calibrated.

The difference between *E. coli* and other data sets needs special attention. The performance of all predictors on *E. coli* was equal to a random baseline. The number of examples from PDB complexes was smaller than for the other organisms, while the number of examples from high-throughput experiments was larger, albeit of possibly lower quality (see Table 1). To assess whether the differences were also present in interaction network structure, we calculated mean node degree for PPI networks of the three organisms. For *H. sapiens* GOLD we obtained mean degree 2.01, for *S. cerevisiae* GOLD 2.05, and for *E. coli* SILVER 4.76. Such difference in numbers suggests a possibility that the data contained more gaps and false positives, making it impossible for a classifier to learn any relations. On the other hand, *E. coli* is the only prokaryotic organism among the three and its proteins may have different characteristics. Brocchieri and Karlin [3] showed that median protein length in prokaryotes is significantly smaller than in eukaryotes. They speculated that the difference may be due to eukaryotic proteins being composed of multiple functional units and additional sequence motifs acting as function regulators. This would definitely affect interaction landscape, however it is difficult to state in what way. Further research into this matter is needed before drawing conclusions.

Results obtained in our study were much lower than usually reported for methods concerning protein interaction prediction, even lower than in Park and Marcotte [11] suggesting that the performance may be routinely overestimated. Those differences are striking, for instance VD representation introduced by Liu [9] was reported to obtain 0.86 AUC ROC on large yeast proteins interactions data set. In the study of Nanni et al. [10] simple AAC representation achieved 0.72 AUC ROC on human PPI data set. Such differences were the result of a different evaluation strategies which, implicitly, led to different problem formulations. We believe that two conditions must be satisfied for an evaluation procedure to correctly represent the problem of predicting meaningful interactions between unknown proteins: a) proteins occurring in training and test sets must be strictly separated, b) protein composition of true positives and true negatives must be as close as possible (including characteristics such as node degree). To our knowledge, ours is the only procedure so far satisfying these conditions.

In the light of our strict evaluation schema and high-quality datasets the problem of predicting meaningful interactions between proteins occurs to be very hard, possibly even harder than generally expected. Success of simple sequence-based methods was limited and introduction of local structural information in our multi-level method yielded only minor and not statistically significant improvement. This raise a question as to how biological information regulating interactions is encoded? We know that protein sequences describe and identify proteins unambiguously, but is it sufficient to know proteins' sequences to fully characterise their behaviour? Our results suggests that the situation is more complex than that. While contacts between single residues of two different proteins occur only in interfaces, whole protein structures may be involved in mediating those interactions.

4 Conclusions

In this work we evaluated some sequenced-based approaches to protein interaction prediction. The main focus was put on the previously developed multi-level predictor (MLPPI). While MLPPI predictor was not worse than global sequence methods, obtained results are far from satisfactory. We believe that making a real breakthrough in protein-protein interaction prediction requires exploiting 3D structural information.

Further research is needed to develop evaluation strategies for multi-level biological input data and fully understand their properties. As our work demonstrates, the impact of evaluation procedure on the results is never overemphasized. We showed that unbalanced train-test splits may be the source of false results in previously published works. We believe that methodological unification and futher discussion is needed for the development of the field.

Acknowledgements

Study was financed by research fellowship within Project „Information technologies: Research and their interdisciplinary applications" (agreement number UDA POKL.04.01.01-00-051/10-00); Polish National Science Centre (grant number UMO 2013/09/B/NZ2/00121 and 2014/15/B/ST6/05082); COST BM1405 and BM1408 EU actions. Research were partially performed using the computational resources of Interdisciplinary Centre of Mathematical and Computational Modelling (ICM), University of Warsaw.

References

[1] Alpaydin, E.: Combined 5 × 2 cv F Test for Comparing Supervised Classification Learning Algorithms. Neural Computation 11(8), 1885–1892 (1999), http://dx.doi.org/10.1162/089976699300016007

[2] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The Protein Data Bank. Nucleic Acids Research 28(1), 235–242 (Jan 2000), http://nar.oxfordjournals.org/content/28/1/235

[3] Brocchieri, L., Karlin, S.: Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Research 33(10), 3390–3400 (2005), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1150220/

[4] Chang, D.T., Syu, Y.T., Lin, P.C.: Predicting the protein-protein interactions using primary structures with predicted protein surface. BMC Bioinformatics 11(Suppl 1), S3 (Jan 2010), http://www.biomedcentral.com/1471-2105/11/S1/S3/abstract

[5] Chou, K.C.: Prediction of protein cellular attributes using pseudo-amino acid composition. Proteins 43(3), 246–255 (May 2001)

[6] Dietterich, T.G.: Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Comput. 10(7), 1895–1923 (1998), http://dx.doi.org/10.1162/089976698300017197

[7] Guo, J., Lin, Y.: A novel method for protein subcellular localization: Combining residue-couple model and SVM. In: Proceedings of the 3rd Asia-Pacific Bioinformatics Conference. pp. 117–129. Singapore (2005)

[8] Jones, D.T.: Protein secondary structure prediction based on position-specific scoring matrices. Journal of Molecular Biology 292(2), 195–202 (1999), http://www.sciencedirect.com/science/article/pii/S0022283699930917

[9] Liu, H.w.: Protein-Protein Interaction Detection by SVM from Sequence. In: The Third International Symposium on Optimization and Systems Biology. pp. 198–206 (2009)

[10] Nanni, L., Lumini, A., Brahnam, S.: An Empirical Study of Different Approaches for Protein Classification. The Scientific World Journal 2014, e236717 (Jun 2014), http://www.hindawi.com/journals/tswj/2014/236717/abs/

[11] Park, Y., Marcotte, E.M.: Flaws in evaluation schemes for pair-input computational predictions. Nature Methods 9(12), 1134–1136 (2012), http://www.nature.com/nmeth/journal/v9/n12/full/nmeth.2259.html

[12] Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V., Wong, S.L., Franklin, G., Li, S., Albala, J.S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R.S., Vandenhaute, J., Zoghbi, H.Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M.E., Hill, D.E., Roth, F.P., Vidal, M.: Towards a proteome-scale map of the human protein–protein interaction network. Nature 437(7062), 1173–1178 (2005), http://www.nature.com/nature/journal/v437/n7062/full/nature04209.html

[13] Saccà, C., Teso, S., Diligenti, M., Passerini, A.: Improved multi-level protein-protein interaction prediction with semantic-based regularization. BMC bioinformatics 15(1), 103 (2014), http://www.biomedcentral.com/1471-2105/15/103/

[14] Saha, I., Zubek, J., Klingström, T., Forsberg, S., Wikander, J., Kierczak, M., Maulik, U., Plewczynski, D.: Ensemble learning prediction of protein–protein interactions using proteins functional annotations. Molecular BioSystems 10(4), 820–830 (Mar 2014), http://pubs.rsc.org/en/content/articlelanding/2014/mb/c3mb70486f

[15] Song, J., Singh, M.: From Hub Proteins to Hub Modules: The Relationship Between Essentiality and Centrality in the Yeast Interactome at Different Scales of Organization. PLoS Comput Biol 9(2), e1002910 (2013), http://dx.doi.org/10.1371/journal.pcbi.1002910

[16] Turner, B., Razick, S., Turinsky, A.L., Vlasblom, J., Crowdy, E.K., Cho, E., Morrison, K., Donaldson, I.M., Wodak, S.J.: iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. Database: The Journal of Biological Databases and Curation 2010, baq023 (2010)

[17] Zubek, J., Tatjewski, M., Boniecki, A., Mnich, M., Basu, S., Plewczynski, D.: Multi-level machine learning prediction of protein–protein interactions in Saccharomyces cerevisiae. PeerJ 3, e1041 (Jul 2015), https://peerj.com/articles/1041