Ensemble Learning Prediction of Protein-Protein Interactions using Proteins Functional Annotations

Indrajit Saha [‡,a,b], Julian Zubek [‡,c], Tomas Klingström [d], Simon Forsberg [e], Johan Wikander [f], Marcin Kierczak [e], Ujjwal Maulik [b], and Dariusz Plewczynski [a]

Supplementary

1 Statistical Significance Test

To assess the significance of the performance, rank of the different classifiers, is computed using Friedman test [1,2]. In general, Friedman test ranks different methods for each dataset separately. To compute the average rank $\mathscr{R}_j$, let r_i^j be the rank of the j-th method for i-th dataset where the number of datasets and methods are N and Q respectively. Therefore, the average rank is $\mathscr{R}_j = \frac{1}{N}\sum_i r_i^j$. In this test, under the null-hypothesis, all the algorithms are equivalent and so their ranks $\mathscr{R}_j$ should be equal.

The Friedman statistic (χ^2 value) is computed as follows:

$$\chi_F^2 = \frac{12N}{Q(Q+1)}\left[\sum_j \mathscr{R}_j^2 - \frac{Q(Q+1)^2}{4}\right] \quad (1)$$

The Friedman statistic is distributed according to χ_F^2 with $Q-1$ degrees of freedom.

Table 1 reports the rank of individual method for different datasets as well as average rank of each method. From Table 1, it can be seen that the average ranks of EL, SVM, RF, DT and NB are 2.06, 2.81, 2.87, 3.81 and 3.44. Moreover, from these average ranks, using Eqn. 1, χ_F^2 is computed as 6.378. Hence, its corresponding p-value is 0.1726 at $\alpha = 0.3$ significance level, which also emphasize the acceptance of alternative hypothesis—the algorithms are not equal in terms of performance. As average rank of EL is superior among them, it constitutes a basis for recommending Ensemble Learning over individual methods.

Table 1 Performance and rank of the classifiers on different datasets. For every entry, the first value indicates the average accuracy and the rank is given within the parenthesis.

Species	Dataset	EL	SVM	RF	DT	NB
Yeast	Gold	0.91 (2)	0.91 (2)	0.89 (4.5)	0.89 (4.5)	0.91 (2)
	Silver	0.80 (1)	0.79 (2)	0.74 (5)	0.77 (4)	0.78 (3)
	Gold-against-All	0.64 (4)	0.65 (3)	0.68 (2)	0.69 (1)	0.63 (5)
	Silver-against-All	0.6 (4.5)	0.6 (4.5)	0.67 (1)	0.65 (2)	0.61 (3)
Human	Gold	0.90 (1)	0.89 (3)	0.89 (3)	0.87 (5)	0.89 (3)
	Silver	0.81 (1.5)	0.81 (1.5)	0.80 (4)	0.80 (4)	0.80 (4)
	Gold-against-All	0.90 (1.5)	0.89 (3)	0.90 (1.5)	0.80 (5)	0.88 (4)
	Silver-against-All	0.92 (1)	0.90 (3.5)	0.91 (2)	0.84 (5)	0.90 (3.5)
Average Rank		2.06	2.81	2.87	3.81	3.44

References

1 M. Friedman, *Journal of the American Statistical Association*, 1937, **32**, 675–701.
2 M. Friedman, *Annals of Mathematical Statistics*, 1940, **11**, 86–92.

‡ *These authors contributed equally to this work.*
[a] *Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warsaw, Poland.*
[b] *Department of Computer Science and Engineering, Jadavpur University, Kolkata, West Bengal, India.*
[c] *Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland.*
[d] *Department of Animal Breeding and Genetics, SLU Global Bioinformatics Centre, Swedish University of Agricultural Sciences, Uppsala, Sweden.*
[e] *Department of Clinical Sciences, Computational Genetics Section, Swedish University of Agricultural Sciences, Uppsala, Sweden.*
[f] *Bioinformatics Program, Faculty of Technology and Natural Sciences, Uppsala University, Sweden.*