

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

Jiancang Zeng Xiangxiang Zeng Ying Ju Quan Zou

Supplementary Tables

Table 1. Comparison of the F measure among different classifiers based on n -gram.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
F measure	0.827	0.778	0.73	0.729	0.873

Table 2. Comparison of the F measure among different classifiers based on n -gram-split.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
F measure	0.856	0.779	0.644	0.54	0.831

Table 3. Comparison of the F measure among different classifiers based on k -skip-2-gram.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
F measure	0.868	0.802	0.746	0.734	0.862

Table 4. Comparison of the F measure among different classifiers based on 188-D.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
F measure	0.869	0.807	0.758	0.708	0.854

Table 5. Comparison of the F measure among different classifiers based on the secondary structure.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
F measure	0.8	0.753	0.69	0.679	0.786

Table 6. Comparison of the F measure among different classifiers based on n -gram.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
RandomPairs	0.557	0.559	0.572	0.445	0.87
RecombinePairs	0.61	0.57	0.564	0.536	0.653

Table 7. Comparison of the F measure among different classifiers based on n -gram-split.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
RandomPairs	0.782	0.649	0.566	0.548	0.86
RecombinePairs	0.649	0.583	0.543	0.41	0.646

Table 8. Comparison of the F measure among different classifiers based on k -skip-2-gram.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
RandomPairs	0.84	0.693	0.661	0.447	0.86
RecombinePairs	0.672	0.579	0.579	0.569	0.649

Table 9. Comparison of the F measure among different classifiers based on 188-D.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
RandomPairs	0.993	0.991	0.942	0.798	0.993
RecombinePairs	0.677	0.583	0.602	0.58	0.651

Table 10. Comparison of the F measure among different classifiers based on the secondary structure.

Classifier	RF	J48	AdaBoost	Bagging	LibD3C
RandomPairs	0.936	0.915	0.895	0.869	0.933
RecombinePairs	0.602	0.567	0.561	0.554	0.587

Table 11. Results of different feature selection methods.

Method	Dataset	20-D	100-D	188-D-SS
mRMR	Negatome data set	0.838	0.854	0.847
	RandomPairs data set	0.994	0.995	0.993
	RecombinePairs data set	0.640	0.652	0.654
MRMD	Negatome data set	0.778	0.847	0.847
	RandomPairs data set	0.768	0.995	0.993
	RecombinePairs data set	0.576	0.640	0.654

Table 12. Z scores of the top eight features selected by mRMR.

Feature order	1	2	3	4	5	6	7	8
Negatome data set	Fea108	Fea12	Fea116	Fea18	Fea192	Fea1	Fea46	Fea37
	20.738	-57.987	43.264	39.156	-38.918	43.007	31.264	30.078
RandomPairs data set	Fea195	Fea58	Fea100	Fea37	Fea69	Fea142	Fea184	Fea153
	100.669	60.197	57.172	54.684	40.099	49.482	53.489	-35.797
RecombinePairs data set	Fea1	Fea12	Fea18	Fea16	Fea168	Fea127	Fea6	Fea126
	20.349	-0.0745	13.290	19.462	2.291	-7.044	0.702	4.512

Table 13. Z scores of the top eight features selected by MRMD.

Feature order	1	2	3	4	5	6	7	8
Negatome data set	Fea198	Fea199	Fea200	Fea120	Fea157	Fea83	Fea99	Fea31
	-31.167	-26.973	-32.688	-23.618	-13.744	-12.198	-29.927	-12.135
RandomPairs data set	Fea198	Fea199	Fea200	Fea120	Fea157	Fea99	Fea178	Fea57
	-19.346	-4.596	-19.674	3.436	-2.348	2.782	7.622	2.435
RecombinePairs data set	Fea198	Fea199	Fea200	Fea120	Fea157	Fea99	Fea83	Fea178
	0.058	-0.433	2.459	-4.747	2.291	-3.093	0.549	1.014

Table 14. Distribution of features ranked by mRMR.

Database	TOP k	Fea1-Fea20	Fea21-Fea188	Fea189-Fea203
Negatome data set	20	9	10	1
	40	10	27	3
	60	11	44	5
	100	16	77	7
	150	19	121	10
RandomPairs data set	20	1	15	4
	40	3	30	7
	60	7	44	9
	100	14	77	9
	150	19	121	10
RecombinePairs data set	20	11	7	2
	40	14	22	4
	60	19	36	5
	100	19	75	6
	150	20	123	7

Table 15. Distribution of features ranked by MRMD.

Database	TOP k	Fea1-Fea20	Fea21-Fea188	Fea189-Fea203
Negatome data set	20	0	17	3
	40	0	37	3
	60	0	51	9
	100	14	74	12
	150	20	115	15
RandomPairs data set	20	0	17	3
	40	0	37	3
	60	0	51	9
	100	14	74	12
	150	20	115	15
RecombinePairs data set	20	0	17	3
	40	0	37	3
	60	0	53	7
	100	14	74	12
	150	20	115	15