

Supplement to

“Multi-Instance Multi-Label Learning”

(Artificial Intelligence, DOI: 10.1016/j.artint.2011.10.002)

by Zhi-Hua Zhou*, Min-Ling Zhang, Sheng-Jun Huang, Yu-Feng Li

National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210046, China

* **Corresponding author** (*E-mail address: zhouzh@lamda.nju.edu.cn*)

[Supplement 1]: Comparison Between MIML and Multi-Label Learning

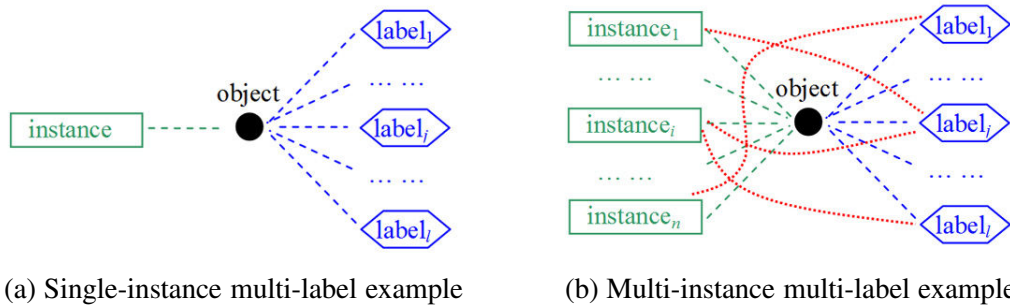


Fig. 1. Comparison between MIML and multi-label learning. Generally, MIML has several advantages over multi-label learning: 1) Suppose the multi-label example in (a) is transformed to the MIML example in (b), then the underlying information in a single instance may become easier to exploit, and for each label the number of training instances can be significantly increased; 2) MIML offers a possibility for understanding the relationship between instances and labels, e.g., as shown by the red curves, label₁ is caused by instance_n, label_l is caused by instance_i, while label_j is caused by the co-occurrence of instance₁ and instance_i.

[Supplement 2]: First Level of the Yeast Gene Functional Hierarchy

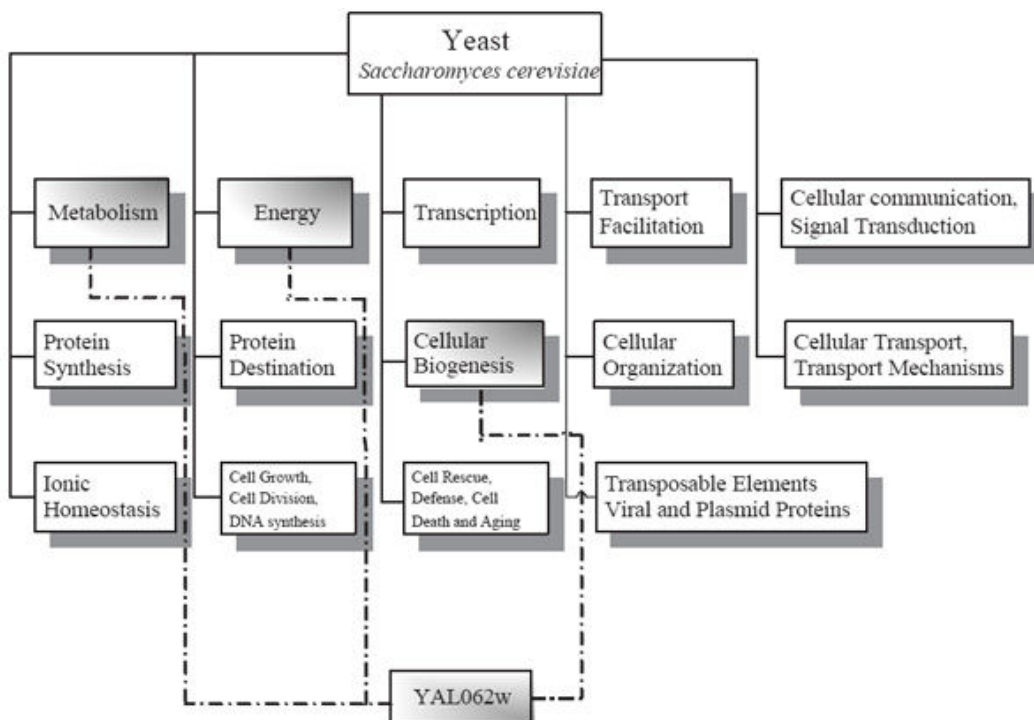


Fig. 2. Illustration on the first level of the hierarchy of the Yeast gene functional classes. There are a total of fourteen possible functional classes in the first level, and one gene may belong to several classes simultaneously. For instance, the one named YAL062w belongs to three classes (shaded in gray) named *Metabolism*, *Energy* and *Cellular Biogenesis*.

[Supplement 3]: Detailed Results on Web Page Data Sets

Table 1

Detailed results on the eleven web page categorization data sets in terms of *hamming loss*.

Data Set	Compared Algorithms					
	INSDIF	INSDIF _{MIMLSVM}	ADTBOOST.MH	RANKSVM	MLSVM	ML- <i>k</i> NN
Arts&Humanities	0.0545	0.0628	0.0606	0.0606	0.0609	0.0612
Business&Economy	0.0257	0.0270	0.0283	0.0276	0.0254	0.0269
Computers&Internet	0.0349	0.0391	0.0405	0.0392	0.0368	0.0412
Education	0.0375	0.0427	0.0438	0.0398	0.0418	0.0387
Entertainment	0.0522	0.0600	0.0596	0.0624	0.0576	0.0604
Health	0.0343	0.0356	0.0421	0.0393	0.0331	0.0458
Recreation&Sports	0.0557	0.0655	0.0584	0.0605	0.0625	0.0620
Reference	0.0259	0.0284	0.0298	0.0297	0.0279	0.0314
Science	0.0312	0.0362	0.0349	0.0341	0.0354	0.0325
Social&Science	0.0205	0.0227	0.0242	0.0242	0.0216	0.0218
Society&Culture	0.0526	0.0570	0.0569	0.0558	0.0549	0.0537
Mean±Std. deviation	0.0386±0.0129	0.0434±0.0154	0.0435±0.0135	0.0430±0.0143	0.0416±0.0149	0.0432±0.0145

Table 2

Detailed results on the eleven web page categorization data sets in terms of *one-error*.

Data Set	Compared Algorithms						
	INSDIF	INSDIF _{MIMLSVM}	ADTBOOST.MH	RANKSVM	MLSVM	ML- <i>k</i> NN	CNMF
Arts&Humanities	0.4797	0.4993	0.5677	0.5713	0.4813	0.6330	0.6031
Business&Economy	0.1167	0.1263	0.1343	0.1203	0.1150	0.1213	0.2272
Computers&Internet	0.3727	0.3880	0.4780	0.4037	0.3590	0.4357	0.4600
Education	0.4713	0.4893	0.5957	0.4937	0.4770	0.5207	0.5814
Entertainment	0.4107	0.4253	0.5547	0.4760	0.4030	0.5300	0.5802
Health	0.2637	0.2830	0.3543	0.2823	0.2573	0.4190	0.3772
Recreation&Sports	0.4840	0.5060	0.5643	0.5627	0.4753	0.7057	0.5997
Reference	0.3773	0.3810	0.4997	0.4290	0.3720	0.4730	0.4700
Science	0.5117	0.5113	0.6430	0.5523	0.5033	0.5810	0.6531
Social&Science	0.2943	0.2987	0.3753	0.3550	0.2790	0.3270	0.3633
Society&Culture	0.4090	0.4307	0.4810	0.4140	0.4023	0.4357	0.6840
Mean±Std. deviation	0.3810±0.1177	0.3945±0.1192	0.4771±0.1438	0.4237±0.1351	0.3750±0.1186	0.4711±0.1573	0.5090±0.1417

Table 3

Detailed results on the eleven web page categorization data sets in terms of *coverage*.

Data Set	Compared Algorithms						
	INSDIF	INSDIF _{MIMLSVM}	ADTBOOST.MH	RANKSVM	MLSVM	ML- <i>k</i> NN	CNMF
Arts&Humanities	5.6013	7.7903	5.2263	8.3610	8.0250	5.4313	7.4381
Business&Economy	2.6647	4.0633	2.4320	3.4633	4.2700	2.1840	5.3385
Computers&Internet	4.7973	6.8340	4.5543	8.7910	7.1673	4.4117	7.0655
Education	4.7460	8.0117	3.9453	8.9560	8.7023	3.4973	6.9977
Entertainment	3.1900	5.0613	3.1417	5.3717	4.8893	3.1467	5.4838
Health	3.6550	6.2127	3.1343	5.6037	6.2510	3.3043	5.8483
Recreation&Sports	4.6140	5.8197	4.3607	5.6680	5.6757	5.1010	6.0749
Reference	3.5827	6.3217	3.3247	5.6567	6.3717	3.5420	5.1255
Science	6.8697	10.0603	6.5557	12.4010	10.4027	6.0470	9.4296
Social&Science	4.1167	6.9953	3.5030	8.2177	6.4437	3.0340	5.4620
Society&Culture	6.1603	7.8847	5.7683	7.0197	7.9050	5.3653	9.6242
Mean±Std. deviation	4.5452±1.2850	6.8232±1.6228	4.1769±1.2612	7.2282±2.4424	6.9185±1.7671	4.0968±1.2364	6.7171±1.5883

Table 4

Detailed results on the eleven web page categorization data sets in terms of *ranking loss*.

Data Set	Compared Algorithms					
	INSDIF	INSDIF _{MIMLSVM}	RANKSVM	MLSVM	ML- <i>k</i> NN	CNMF
Arts&Humanities	0.1444	0.2185	0.2432	0.2265	0.1514	0.2070
Business&Economy	0.0444	0.0742	0.0685	0.0774	0.0373	0.1091
Computers&Internet	0.0920	0.1510	0.2091	0.1597	0.0921	0.1569
Education	0.1005	0.1800	0.2080	0.1973	0.0800	0.1575
Entertainment	0.1085	0.1879	0.2088	0.1799	0.1151	0.2098
Health	0.0577	0.1175	0.1076	0.1165	0.0605	0.1079
Recreation&Sports	0.1530	0.2113	0.2094	0.2028	0.1913	0.2274
Reference	0.0844	0.1574	0.1446	0.1589	0.0919	0.1325
Science	0.1276	0.1994	0.2570	0.2077	0.1167	0.1806
Social&Science	0.0707	0.1278	0.1661	0.1178	0.0561	0.1027
Society&Culture	0.1436	0.2015	0.1758	0.2027	0.1338	0.2876
Mean±Std. deviation	0.1024±0.0367	0.1660±0.0451	0.1816±0.0569	0.1679±0.0468	0.1024±0.0453	0.1708±0.0583

Table 5

Detailed results on the eleven web page categorization data sets in terms of *average precision*.

Data Set	Compared Algorithms						
	INSDIF	INSDIF _{MIMLSVM}	ADTBOOST.MH	RANKSVM	MLSVM	ML-kNN	CNMF
Arts&Humanities	0.6061	0.5657	0.5480	0.5090	0.5694	0.5097	0.4815
Business&Economy	0.8809	0.8532	0.8699	0.8678	0.8540	0.8798	0.7676
Computers&Internet	0.6879	0.6503	0.6107	0.6123	0.6640	0.6338	0.5944
Education	0.6301	0.5796	0.5544	0.5702	0.5827	0.5993	0.5204
Entertainment	0.6805	0.6486	0.5914	0.5881	0.6575	0.6013	0.5377
Health	0.7807	0.7442	0.7182	0.7336	0.7527	0.6817	0.6470
Recreation&Sports	0.6111	0.5773	0.5573	0.5315	0.5961	0.4552	0.4931
Reference	0.7028	0.6729	0.6126	0.6352	0.6793	0.6194	0.6028
Science	0.5814	0.5546	0.4815	0.5007	0.5559	0.5324	0.4526
Social&Science	0.7618	0.7335	0.7147	0.6788	0.7432	0.7481	0.6980
Society&Culture	0.6278	0.5993	0.5764	0.6076	0.6069	0.6128	0.3741
Mean±Std. deviation	0.6864±0.0908	0.6526±0.0931	0.6214±0.1079	0.6213±0.1077	0.6602±0.0925	0.6249±0.1164	0.5612±0.1144

Table 6

Detailed results on the eleven web page categorization data sets in terms of *average recall*.

Data Set	Compared Algorithms				
	INSDIF	INSDIF _{MIMLSVM}	RANKSVM	MLSVM	ML-kNN
Arts&Humanities	0.2464	0.3847	0.1163	0.2283	0.0691
Business&Economy	0.7403	0.7038	0.6905	0.7372	0.7189
Computers&Internet	0.3928	0.4846	0.2716	0.3970	0.3348
Education	0.2550	0.4139	0.1973	0.2338	0.1963
Entertainment	0.3324	0.4910	0.1631	0.3324	0.2284
Health	0.5279	0.5753	0.3326	0.5489	0.3990
Recreation&Sports	0.2216	0.4166	0.1288	0.2263	0.0776
Reference	0.3862	0.5689	0.2183	0.4009	0.2199
Science	0.2008	0.4006	0.0904	0.2049	0.1679
Social&Science	0.5122	0.6302	0.3891	0.5185	0.4719
Society&Culture	0.3338	0.4392	0.1736	0.3303	0.3231
Mean±Std. deviation	0.3772±0.1626	0.5008±0.1051	0.2520±0.1720	0.3780±0.1672	0.2915±0.1893

Table 7

Detailed results on the eleven web page categorization data sets in terms of *average F1*.

Data Set	Compared Algorithms				
	INSDIF	INSDIF _{MIMLSVM}	RANKSVM	MLSVM	ML-kNN
Arts&Humanities	0.3504	0.4580	0.1893	0.3260	0.1217
Business&Economy	0.8045	0.7713	0.7691	0.7913	0.7913
Computers&Internet	0.5001	0.5554	0.3763	0.4969	0.4382
Education	0.3631	0.4829	0.2932	0.3337	0.2957
Entertainment	0.4467	0.5589	0.2553	0.4416	0.3311
Health	0.6299	0.6489	0.4577	0.6348	0.5034
Recreation&Sports	0.3252	0.4840	0.2074	0.3280	0.1326
Reference	0.4985	0.6165	0.3249	0.5042	0.3246
Science	0.2985	0.4652	0.1532	0.2994	0.2552
Social&Science	0.6125	0.6779	0.4946	0.6109	0.5788
Society&Culture	0.4359	0.5069	0.2700	0.4278	0.4231
Mean±Std. deviation	0.4787±0.1538	0.5660±0.1016	0.3446±0.1768	0.4722±0.1560	0.3814±0.1962