Hierarchical Multi-Label Classification Networks

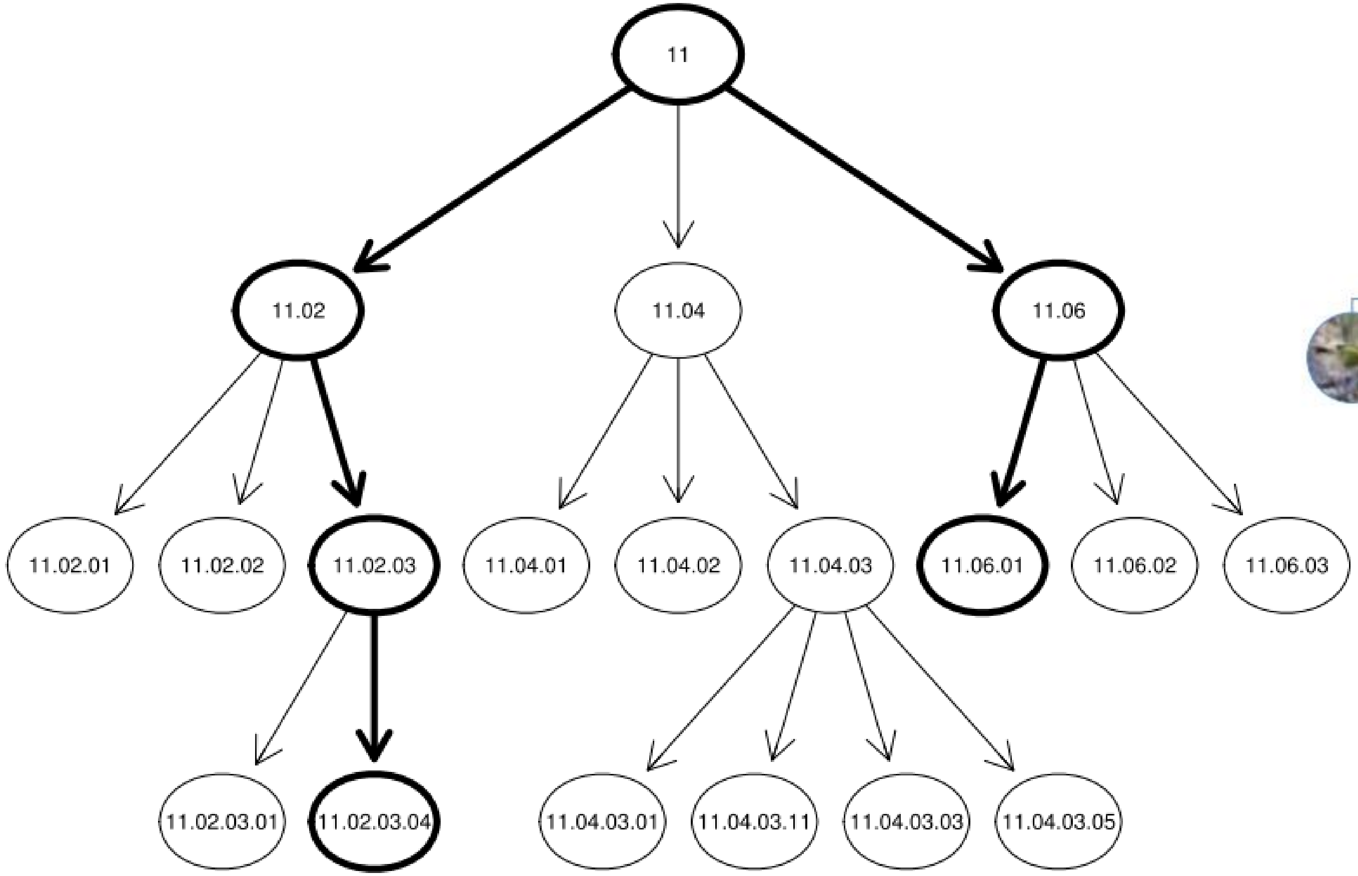
Jônatas Wehrmann Ricardo Cerri Rodrigo C. Barros

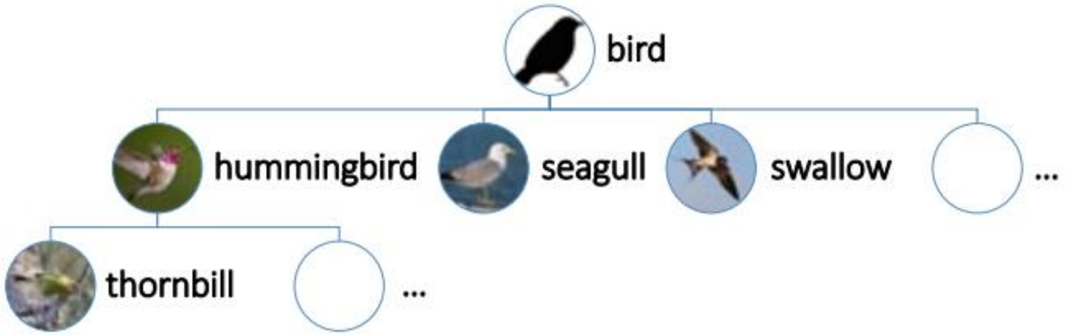
ICML 2018

Contents

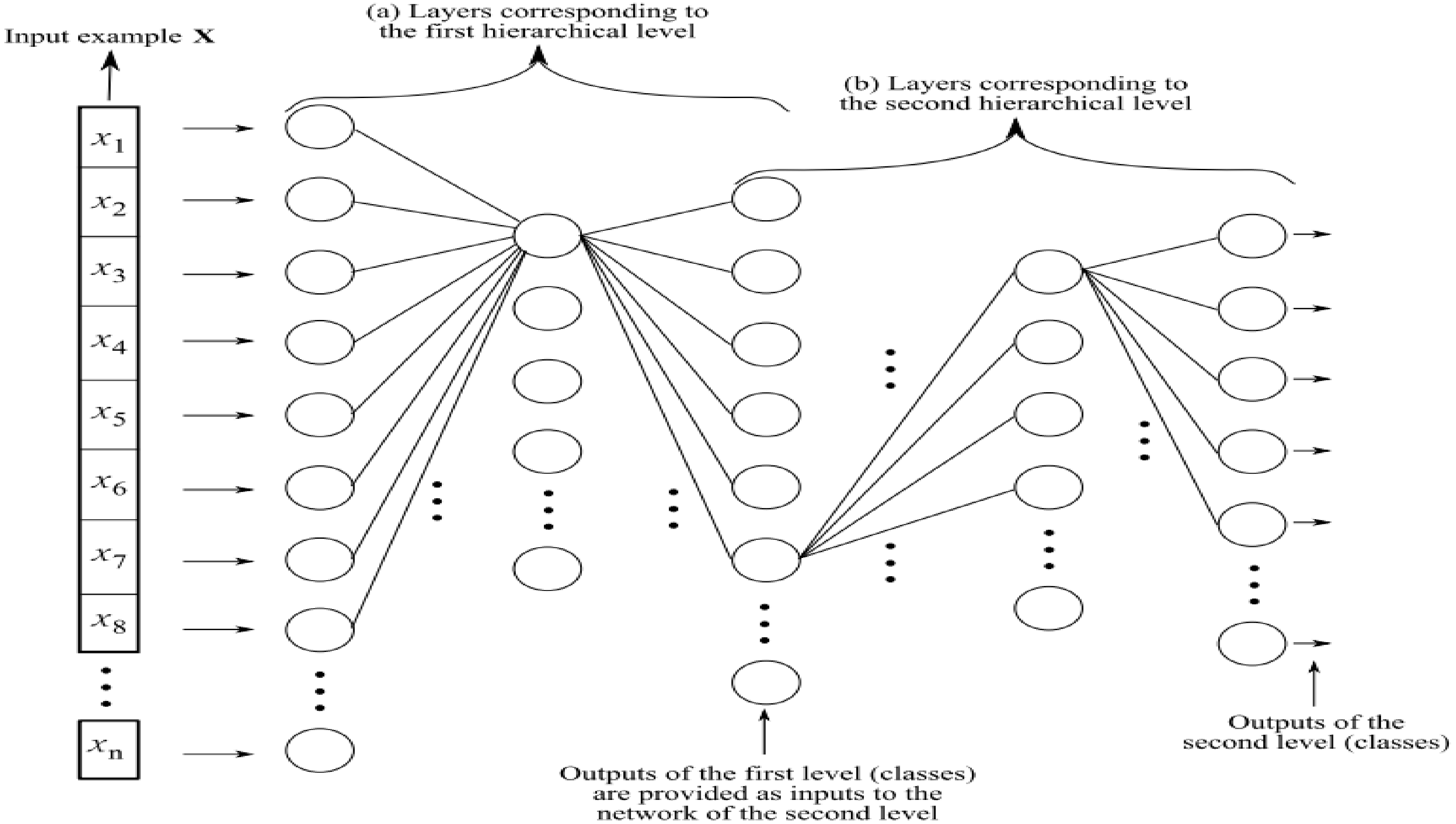
Intruction

a niche of tasks in which classes are not disjoint but organized into a hierarchical structure, namely hierarchical classification (HC)





Local way:each classifier is responsible for the prediction of particular hierarchical levels.



Global way : a single classifier capable of associating objects with their corresponding classes in the hierarchy as a whole

HMCN-F

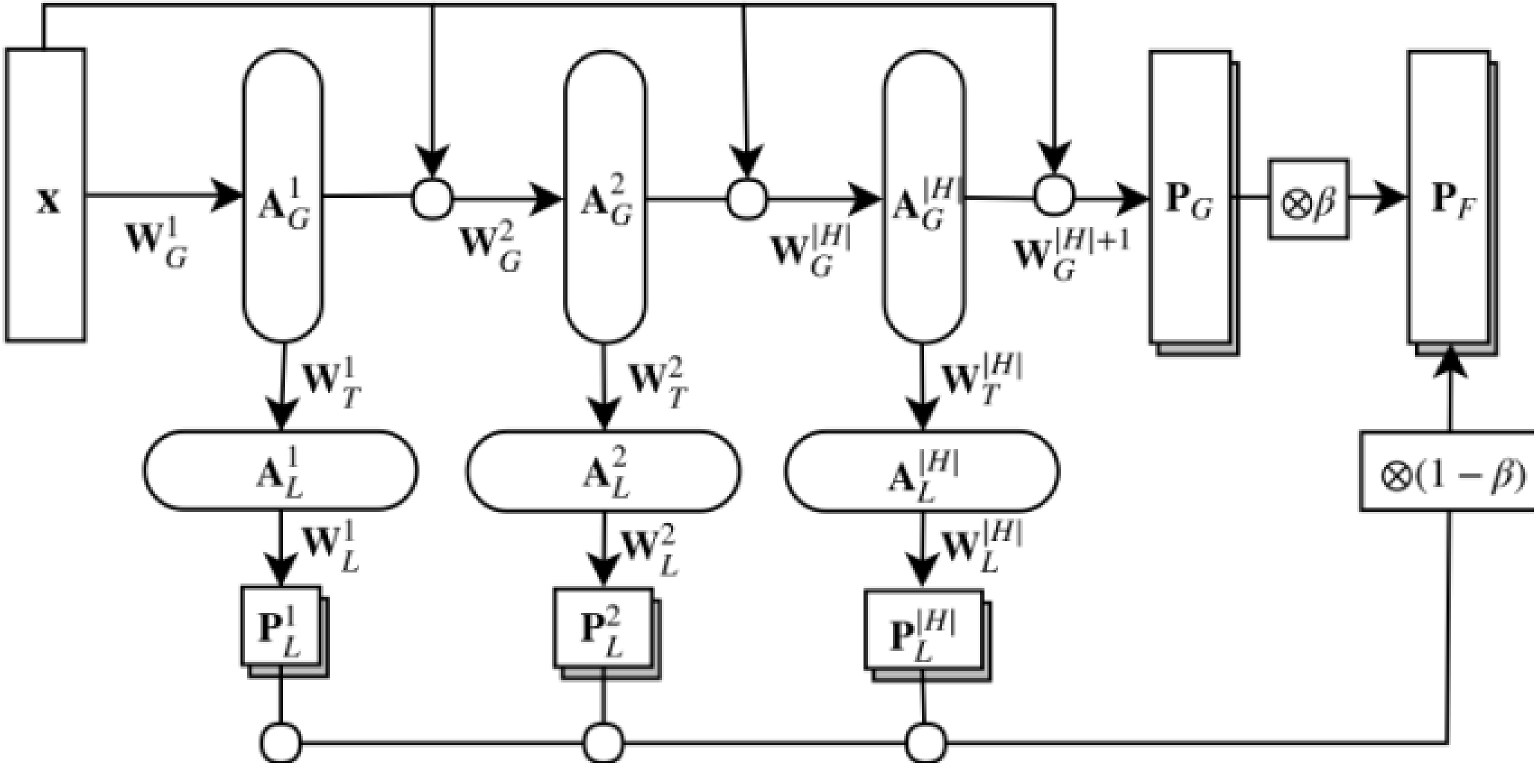


Figure 1. HMCN-F architecture.

First level:

$$\mathbf{A}_G^1 = \phi(\mathbf{W}_G^1 \mathbf{x} + \mathbf{b}_{\mathbf{G}}^{\mathbf{1}})$$

reusing original features

Middle level:

$$\mathbf{A}_G^h = \phi(\mathbf{W}_G^h (\mathbf{A}_G^{h-1} \odot \mathbf{x}) + \mathbf{b}_G^h)$$

Global Output:

$$\mathbf{P}_G = \sigma(\mathbf{W}_G^{|H|+1} \mathbf{A}_G^{|H|} + \mathbf{b}_G^{|H|+1})$$

Global flow: $\mathbf{A}_G^1 = \phi(\mathbf{W}_G^1 \mathbf{x} + \mathbf{b}_{\mathbf{G}}^{\mathbf{1}})$

Local flow: $\mathbf{A}_L^h = \phi(\mathbf{W}_T^h \mathbf{A}_G^h + \mathbf{b}_T^h)$

Local output: $\mathbf{P}_L^h = \sigma(\mathbf{W}_L^h \mathbf{A}_L^h + \mathbf{b}_L^h)$

The final prediction:

$$\mathbf{P}_F = \beta \left(\mathbf{P}_L^1 \odot \mathbf{P}_L^2 \odot, \ldots, \mathbf{P}_L^{|H|} \right) + (1 - \beta)\mathbf{P}_G$$

HMCN-R

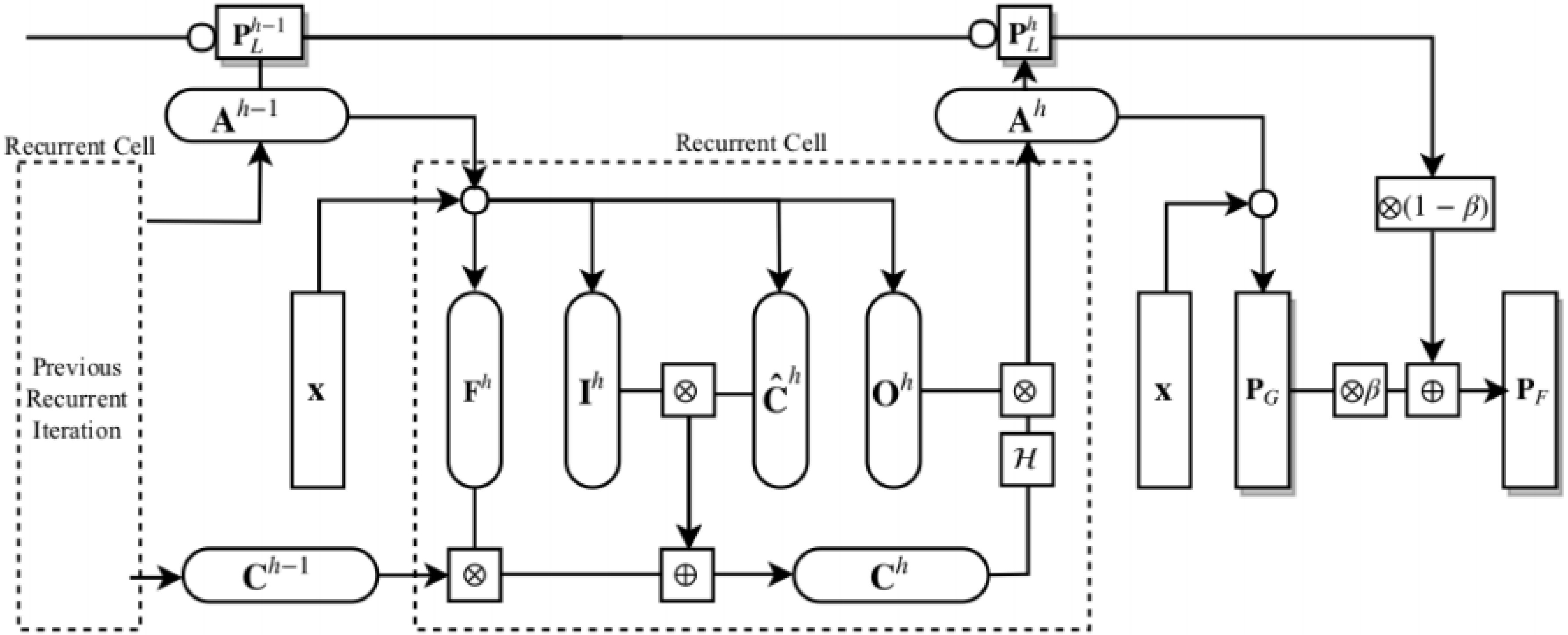


Figure 2. HMCN-R architecture.

$$\mathbf{F}^h = \sigma\left(\mathbf{W}_F(\mathbf{A}^{h-1} \odot \mathbf{x}) + \mathbf{b}_F\right)$$

$$\mathbf{I}^h = \sigma\left(\mathbf{W}_I(\mathbf{A}^{h-1} \odot \mathbf{x}) + \mathbf{b}_I\right)$$

$$\hat{\mathbf{C}}^h = \mathcal{H}\left(\mathbf{W}_C(\mathbf{A}^{h-1} \odot \mathbf{x}) + \mathbf{b}_C\right)$$

$$\mathbf{C}^h = \mathbf{F}^h\mathbf{C}^{h-1} + \mathbf{I}^h\hat{\mathbf{C}}^h$$

$$\mathbf{O}^h = \sigma\left(\mathbf{W}_O(\mathbf{A}^{h-1} \odot \mathbf{x}) + \mathbf{b}_O\right)$$

$$\mathbf{A}^h = \mathbf{O}^h\mathcal{H}(\mathbf{C}^h)$$

Loss Function

$$\mathcal{L}_L = \sum_{h=1}^{|H|} \left[\mathcal{E}(\mathbf{P}_L^h, \mathbf{Y}_L^h)\right]$$

$$\mathcal{L}_G = \mathcal{E}(\mathbf{P}_G, \mathbf{Y}_G)$$

$$\mathcal{E}(\hat{\mathbf{Y}}, \mathbf{Y}) : \quad -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{|C|} \Big[\mathbf{Y}_{ij} \times \log(\hat{\mathbf{Y}}_{ij}) + (1 - \mathbf{Y}_{ij}) \times \log(1 - \hat{\mathbf{Y}}_{ij})\Big]$$

$$\mathcal{L}_{H_i} = \lambda \max\{0, \mathbf{Y}_{in} - \mathbf{Y}_{ip}\}^2$$

child node's score

parent node's score

$$\min_{W} \left(\mathcal{L}_L + \mathcal{L}_G + L_H\right)$$

Expriments

Table 2. IMPACT OF INPUT REUSE AND LOCAL OUTPUTS ON THE PERFORMANCE OF HMCN-F. THE VALUES PRESENTED ARE OF $AU(\overline{PRC})$ REGARDING FUNCAT VALIDATION DATA. AVERAGE RANKING IS ALSO PROVIDED (LOWER IS BETTER).

DATASET	HMCN (GLOBAL ONLY)	HMCN-F (INPUT REUSE)	HMCN-F (LOCAL OUTPUTS)	HMCN-F
CELLCYCLE	0.200	0.212	0.216	**0.228**
DERISI	0.164	0.166	**0.173**	0.168
EISEN	0.253	0.263	0.266	**0.280**
EXPR	0.220	0.242	0.238	**0.271**
GASH1	0.217	0.243	0.241	**0.261**
GASH2	0.210	0.215	0.211	**0.227**
SEQ	0.109	0.258	0.236	**0.279**
SPO	0.158	0.167	**0.174**	0.170
AVERAGE RANKING	5.25	3.38	3.00	1.25

Table 3. IMPACT OF HIERARCHICAL VIOLATIONS IN HMCN. VALUES PRESENTED ARE OF $AU(\overline{PRC})$ ON VALIDATION DATA.

	HMCN-F			HMCN-R	
DATASET	$\lambda = 0.0$	$\lambda = 0.1$	$\lambda = 1.0$	$\lambda = 0.0$	$\lambda = 0.1$
CELLCYCLE	0.250	**0.252**	0.233	0.245	0.249
DERISI	0.191	**0.193**	0.185	0.188	0.189
EISEN	0.293	**0.298**	0.264	0.292	**0.298**
EXPR	0.296	**0.301**	0.263	0.295	0.300
GASCH1	0.281	**0.284**	0.259	0.278	0.283
GASCH2	0.252	**0.254**	0.232	0.244	0.249
SEP	0.283	**0.291**	0.262	0.285	0.290
SPO	0.210	**0.211**	0.197	0.205	0.207

Table 4. Comparison of HMCN with the baseline methods. The values presented are of $AU(\overline{PRC})$. Values in bold outperform the best results published so far, and underlined are the novel state-of-the-art for each dataset.

Dataset	HMCN-F	HMCN-R	Clus-HMC	CSSA	Clus-Ens	HMC-LMLP
Cellcycle (FunCat)	**0.252**	**0.247**	0.172	0.188	0.227	0.207
Derisi (FunCat)	**0.193**	**0.189**	0.175	0.186	0.188	0.183
Eisen (FunCat)	**0.298**	**0.298**	0.204	0.212	0.271	0.245
Expr (FunCat)	**0.301**	**0.300**	0.210	0.220	0.271	0.243
Gasch1 (FunCat)	**0.284**	**0.283**	0.205	0.208	0.267	0.236
Gasch2 (FunCat)	**0.254**	**0.249**	0.195	0.210	0.227	0.211
Seq (FunCat)	**0.291**	**0.290**	0.211	0.218	0.284	0.236
Spo (FunCat)	**0.211**	0.210	0.186	0.208	0.210	0.186
Cellcycle (GO)	**0.400**	**0.395**	0.357	0.366	0.387	-
Derisi (GO)	**0.369**	**0.368**	0.355	0.357	0.363	-
Eisen (GO)	**0.440**	**0.435**	0.380	0.401	0.433	-
Expr (GO)	**0.452**	**0.450**	0.368	0.384	0.418	-
Gasch1 (GO)	**0.428**	**0.416**	0.371	0.383	0.415	-
Gasch2 (GO)	**0.465**	**0.463**	0.369	0.373	0.395	-
Seq (GO)	**0.447**	**0.443**	0.386	0.387	0.435	-
Spo (GO)	**0.376**	**0.375**	0.345	0.352	0.372	-
Diatoms	**0.530**	**0.514**	0.167	-	0.379	-
Enron	**0.724**	**0.710**	0.638	-	0.681	-
Imclef07a	**0.950**	**0.904**	0.574	-	0.777	-
Imclef07d	**0.920**	**0.897**	0.749	-	0.863	-
Reuters	0.649	0.610	0.562	-	**0.703**	-
Average Ranking	**1.07**	2.04	5.12	3.96	3.57	-