End-to-End Probabilistic Label-Specific Feature Learning for Multi-Label Classification Supplementary Material

Jun-Yi Hang[1,2], Min-Ling Zhang[1,2*], Yanghe Feng[3], Xiaocheng Song[4]

[1]School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[2]Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, China
[3]College of Systems Engineering, National University of Defense Technology, Changsha 410073, China
[4]Department of Beijing Institute of Electronic Engineering, Beijing 100854, China
{hangjy, zhangml}@seu.edu.cn, fengyanghe@nudt.edu.cn, sxchitman@126.com

Further Experimental Results

Comparative Studies

We employ ten-fold cross validation to evaluate our PACA and six well-established multi-label classification approaches on the 14 benchmark data sets. Table 3 reports detailed experimental results in terms of *One-error*, *Coverage* and *Ranking loss*, which are not covered in the *Comparative Studies* part of the main body due to page limit.

Further Analyses

The Wilcoxon signed-ranks test at 0.05 significance level is conducted to analyze whether PACA performs statistically better than its variants described in the *Further Analyses* part of the main body. Table 1 summarizes the p-value statistics on each evaluation metric, which show PACA is statistically superior to its variants in terms of all evaluation metrics.

Classifier Considerations Theoretically, for the 2-dimensional label-specific features in PACA, a softmax-based parameter-free classifier is equivalent to the commonly-used sigmoid classifier with $[1, -1]^{\mathrm{T}}$ weights and zero bias. We further validate the equivalence by empirical ablation studies. Table 2 reports detailed experimental results of such a variant named PACA-cls in terms of *Average precision*, where a sigmoid classifier is attached to the label-specific features for each class label. The results show that PACA-cls achieves almost the same results as PACA.

Naive Label-Specific Feature Learning Table 2 reports ablation study results of a naive model named NaiveLearning, which shares the same encoder with PACA but the last layer of the encoder is label-specific to generate label-specific features. The results demonstrate the superiority of PACA in learning discriminative label-specific features.

Additional Parameter Sensitivity Figure 1(a) further analyzes how the performance of PACA changes when the hidden dimensionalities of the autoencoders change. The results show that the $1.0\times$ width is a quite reasonable choice as default setting.

Table 1: Summary of the Wilcoxon signed-ranks test for PACA against its variants at 0.05 significance level. p-values are shown in the brackets.

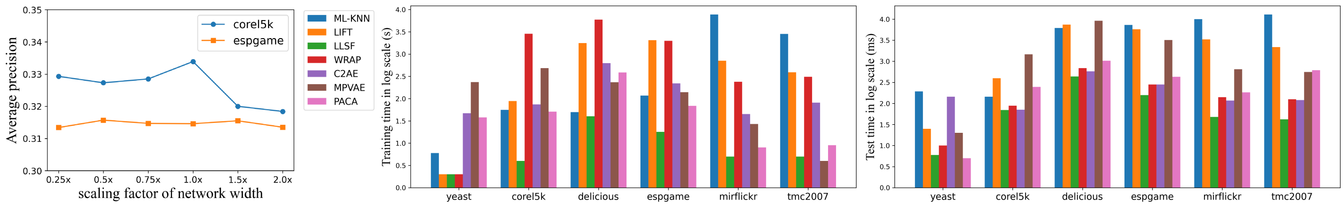
PACA against	PACA-sp	PACA-nr
Average precision	**win** [0.0001]	**win** [0.0009]
Macro-averaging AUC	**win** [0.0023]	**win** [0.0067]
Hamming loss	**win** [0.0112]	**win** [0.0195]
One error	**win** [0.0004]	**win** [0.0010]
Coverage	**win** [0.0017]	**win** [0.0171]
Ranking loss	**win** [0.0004]	**win** [0.0103]

Table 2: Further ablation studies of PACA. Best and second best results are shown in **boldface** and underlined respectively.

Data sets	*Average precision* ↑		
	PACA	PACA-cls	NaiveLearning
CAL500	**0.5246±0.0170**	0.5226±0.0155	0.5140±0.0138
Image	0.8561±0.0173	**0.8565±0.0170**	0.8125±0.0265
scene	**0.9048±0.0161**	0.9036±0.0189	0.8816±0.0159
yeast	0.7717±0.0176	**0.7718±0.0185**	0.7666±0.0146
corel5k	0.3339±0.0126	**0.3349±0.0108**	0.3104±0.0108
rcv1-s1	0.6444±0.0113	**0.6498±0.0111**	0.6113±0.0141
Corel16k-s1	0.3717±0.0068	**0.3747±0.0129**	0.3511±0.0056
delicious	**0.4129±0.0046**	0.4113±0.0054	0.3850±0.0044
iaprtc12	**0.4430±0.0053**	0.4419±0.0052	0.4106±0.0070
espgame	**0.3146±0.0039**	0.3111±0.0019	0.2895±0.0055
mirflickr	**0.7022±0.0058**	0.7009±0.0056	0.6815±0.0079
tmc2007	**0.8322±0.0036**	0.8287±0.0035	0.8273±0.0042
mediamill	0.7864±0.0033	**0.7867±0.0042**	0.7647±0.0048
bookmarks	**0.5126±0.0027**	0.5048±0.0021	0.4900±0.0030

Complexity Analyses Let b be the batch size and d_h be the hidden dimensionality of conditioner in the normalizing flows, the density estimation process for generating label-specific features with $2 \cdot q$ probabilistic prototypes has time complexity $\mathcal{O}(bqd_zd_hd_\tau)$. Compared with a fully-connected counterpart with hidden dimensionality d_h, generating label-specific features via the normalizing flows has d_τ times time complexity, which is highly controllable by setting d_τ a small value ($d_\tau = 16$ in our experiments). Furthermore, Figure 1(b)(c) illustrate the empirical training and test time of each comparing approach, which show that PACA is comparable to existing approaches in time overhead.

*Corresponding author



(a) Sensitivity in terms of network width (b) Training time of comparing approaches (c) Test time of comparing approaches

Figure 1: (a) Performance of PACA with varying network widths in terms of *Average precision*. (b)(c) Running time (training/test) of each comparing approach on six benchmark data sets. For histogram illustration, the y-axis corresponds to the logarithm of running time.

Table 3: Predictive performance of each comparing approach (mean±std. deviation). ↑ (↓) indicates the larger (smaller) the value, the better the performance. Best and second best results are shown in **boldface** and <u>underlined</u> respectively.

Data sets	*One-error* ↓						
	ML-KNN	LIFT	LLSF	WRAP	C2AE	MPVAE	PACA
CAL500	**0.1155±0.0259**	0.1235±0.0333	<u>0.1175±0.0425</u>	**0.1155±0.0259**	**0.1155±0.0259**	<u>0.1175±0.0340</u>	**0.1155±0.0412**
Image	0.3220±0.0410	0.2690±0.0330	0.3800±0.0388	0.3480±0.0299	<u>0.2660±0.0340</u>	0.2775±0.0342	**0.2180±0.0266**
scene	0.2343±0.0311	0.1961±0.0293	0.2547±0.0250	0.2796±0.0377	<u>0.1874±0.0361</u>	0.2094±0.0376	**0.1595±0.0258**
yeast	0.2234±0.0285	0.2181±0.0341	<u>0.2176±0.0250</u>	0.2251±0.0235	0.2396±0.0317	0.2280±0.0300	**0.2152±0.0219**
corel5k	0.7364±0.0190	0.6824±0.0118	0.6452±0.0180	<u>0.6156±0.0155</u>	0.6424±0.0223	0.6206±0.0187	**0.6086±0.0156**
rcv1-s1	0.5650±0.0179	0.4078±0.0153	0.4122±0.0155	<u>0.3960±0.0221</u>	0.4232±0.0248	0.3973±0.0171	**0.3948±0.0146**
Corel16k-s1	0.7319±0.0109	0.6758±0.0128	0.6364±0.0175	0.6303±0.0173	0.6463±0.0168	<u>0.6272±0.0136</u>	**0.6154±0.0122**
delicious	0.3965±0.0099	0.3253±0.0122	0.3536±0.0110	0.3385±0.0099	0.3299±0.0084	<u>0.3013±0.0097</u>	**0.3000±0.0090**
iaprtc12	0.4867±0.0116	0.5007±0.0074	0.4776±0.0092	0.4639±0.0081	0.4617±0.0092	<u>0.4200±0.0136</u>	**0.4130±0.0076**
espgame	0.6828±0.0097	0.6305±0.0090	0.6402±0.0130	0.6235±0.0147	0.6321±0.0195	<u>0.6009±0.0110</u>	**0.5944±0.0096**
mirflickr	0.3576±0.0111	0.3170±0.0077	0.2998±0.0109	0.2954±0.0099	0.2790±0.0094	<u>0.2660±0.0089</u>	**0.2520±0.0094**
tmc2007	0.3081±0.0094	0.2171±0.0043	0.2229±0.0080	0.2322±0.0075	0.2234±0.0134	<u>0.2025±0.0043</u>	**0.1955±0.0052**
mediamill	0.1553±0.0058	0.1580±0.0062	0.1589±0.0056	0.1564±0.0058	0.1531±0.0062	<u>0.1407±0.0065</u>	**0.1321±0.0063**
bookmarks	0.6307±0.0052	0.5331±0.0045	0.5226±0.0030	0.5430±0.0056	0.5325±0.0065	**0.5112±0.0034**	<u>0.5162±0.0035</u>

Data sets	*Coverage* ↓						
	ML-KNN	LIFT	LLSF	WRAP	C2AE	MPVAE	PACA
CAL500	0.7518±0.0150	0.7531±0.0235	0.7542±0.0164	<u>0.7403±0.0111</u>	0.7943±0.0213	0.7406±0.0113	**0.7323±0.0155**
Image	0.1972±0.0190	<u>0.1689±0.0132</u>	0.2219±0.0193	0.1937±0.0194	0.1724±0.0177	0.1700±0.0139	**0.1501±0.0176**
scene	0.0803±0.0077	<u>0.0656±0.0075</u>	0.0887±0.0088	0.0906±0.0096	0.0765±0.0107	0.0745±0.0104	**0.0585±0.0080**
yeast	**0.4445±0.0138**	<u>0.4517±0.0149</u>	0.4535±0.0166	0.4532±0.0161	0.4737±0.0212	0.4534±0.0174	0.4577±0.0192
corel5k	0.3053±0.0119	0.2905±0.0119	0.4361±0.0144	0.3011±0.0176	0.3121±0.0169	<u>0.2275±0.0121</u>	**0.2252±0.0137**
rcv1-s1	0.2260±0.0086	0.1212±0.0073	0.1176±0.0098	0.1025±0.0087	0.1172±0.0131	<u>0.0893±0.0070</u>	**0.0885±0.0072**
Corel16k-s1	0.3342-0.0072	0.3236±0.0068	0.3237±0.0079	0.2690±0.0069	0.3029±0.0055	<u>0.2348±0.0048</u>	**0.2315±0.0058**
delicious	0.5966±0.0090	0.4805±0.0073	0.6179±0.0091	0.5369±0.0077	0.5049±0.0077	<u>0.4040±0.0061</u>	**0.3906±0.0059**
iaprtc12	0.3518±0.0066	0.3204±0.0046	0.3768±0.0080	0.3401±0.0067	0.2903±0.0063	<u>0.2305±0.0039</u>	**0.2256±0.0058**
espgame	0.4414±0.0064	0.3509±0.0088	0.4537±0.0078	0.3767±0.0086	0.3942±0.0076	<u>0.3231±0.0057</u>	**0.3102±0.0060**
mirflickr	0.3410±0.0026	0.3173±0.0030	0.3193±0.0038	0.3103±0.0042	0.3032±0.0066	<u>0.2676±0.0041</u>	**0.2661±0.0040**
tmc2007	0.1833±0.0041	0.1210±0.0038	0.1270±0.0042	0.1299±0.0045	0.1432±0.0053	**0.1130±0.0031**	<u>0.1147±0.0025</u>
mediamill	0.1369±0.0023	0.1555±0.0028	0.1735±0.0041	0.1675±0.0037	0.1544±0.0035	<u>0.1217±0.0029</u>	**0.1156±0.0018**
bookmarks	0.2575±0.0028	0.1308±0.0021	0.1569±0.0038	0.1557±0.0029	0.1831±0.0036	<u>0.1172±0.0020</u>	**0.1118±0.0018**

Data sets	*Ranking loss* ↓						
	ML-KNN	LIFT	LLSF	WRAP	C2AE	MPVAE	PACA
CAL500	0.1831±0.0041	0.1814±0.0058	0.1835±0.0070	<u>0.1761±0.0054</u>	0.1962±0.0047	0.1768±0.0046	**0.1736±0.0060**
Image	0.1785±0.0218	<u>0.1432±0.0137</u>	0.2116±0.0227	0.1772±0.0241	0.1477±0.0221	0.1471±0.0168	**0.1228±0.0184**
scene	0.0790±0.0106	<u>0.0622±0.0100</u>	0.0893±0.0112	0.0916±0.0113	0.0734±0.0144	0.0729±0.0146	**0.0538±0.0117**
yeast	<u>0.1644±0.0107</u>	**0.1637±0.0095**	0.1684±0.0101	0.1690±0.0096	0.1828±0.0136	0.1682±0.0123	0.1663±0.0136
corel5k	0.1340±0.0053	0.1221±0.0046	0.1912±0.0076	0.1308±0.0073	0.1511±0.0088	<u>0.1011±0.0045</u>	**0.1004±0.0067**
rcv1-s1	0.1083±0.0049	0.0481±0.0031	0.0463±0.0038	0.0399±0.0034	0.0483±0.0060	**0.0365±0.0036**	<u>0.0370±0.0031</u>
Corel16k-s1	0.1722±0.0032	0.1627±0.0030	0.1624±0.0035	0.1376±0.0039	0.1622±0.0030	<u>0.1222±0.0026</u>	**0.1206±0.0035**
delicious	0.1265±0.0024	0.0996±0.0018	0.1433±0.0033	0.1052±0.0019	0.1171±0.0017	<u>0.0882±0.0017</u>	**0.0876±0.0019**
iaprtc12	0.1217±0.0028	0.1110±0.0019	0.1232±0.0032	0.1102±0.0027	0.1017±0.0036	<u>0.0771±0.0019</u>	**0.0768±0.0024**
espgame	0.1839±0.0023	0.1432±0.0028	0.1823±0.0037	0.1512±0.0034	0.1704±0.0030	<u>0.1347±0.0031</u>	**0.1297±0.0021**
mirflickr	0.1329±0.0033	0.1196±0.0022	0.1188±0.0034	0.1143±0.0035	0.1096±0.0038	**0.0910±0.0024**	<u>0.0914±0.0029</u>
tmc2007	0.0891±0.0031	0.0466±0.0021	0.0485±0.0017	0.0508±0.0019	0.0584±0.0021	**0.0407±0.0015**	<u>0.0417±0.0017</u>
mediamill	0.0386±0.0008	0.0446±0.0009	0.0524±0.0018	0.0492±0.0016	0.0467±0.0015	<u>0.0335±0.0010</u>	**0.0320±0.0008**
bookmarks	0.1759±0.0023	0.0833±0.0014	0.0977±0.0026	0.0963±0.0017	0.1213±0.0028	<u>0.0755±0.0016</u>	**0.0719±0.0014**