Understanding Partial Multi-label Learning via Mutual Information (Supplementary)

Xiuwen Gong [†,‡], **Dong Yuan**[†], **Wei Bao** [†]
[†] Faculty of Engineering, The University of Sydney
[‡] Hunan Huishiwei Intelligent Technology Co., Ltd.
{xiuwen.gong, dong.yuan, wei.bao}@sydney.edu.au

Abstract

In this supplementary file, we present the derivation of Eq. (7), the datasets, the remaining configurations and experimental results.

1 Derivation of Eq. (7) in main paper

$$\begin{aligned}
\mathbb{E}_{(\mathbf{X},\mathbf{Y})}\left\{\log\frac{P(\mathbf{y}_v|\mathbf{x})}{P(\mathbf{y}|\mathbf{x})}\right\} &= -\mathbb{E}_{(\mathbf{X},\mathbf{Y})}\left\{\log\frac{P(\mathbf{y}|\mathbf{x})}{P(\mathbf{y}_v|\mathbf{x})}\right\} \\
&= -\sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y})\log\frac{P(\mathbf{y}_v,\mathbf{y}_{\bar{v}}|\mathbf{x})}{P(\mathbf{y}_v|\mathbf{x})} \\
&= -\sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y})\log\frac{P(\mathbf{y}_{\bar{v}},\mathbf{y}_v,\mathbf{x})}{P(\mathbf{y}_v,\mathbf{x})} \\
&= -\sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y})\log\frac{P(\mathbf{y}_{\bar{v}},\mathbf{x}|\mathbf{y}_v)}{P(\mathbf{x}|\mathbf{y}_v)} \\
&= -\sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y})\log\frac{P(\mathbf{y}_{\bar{v}},\mathbf{x}|\mathbf{y}_v)}{P(\mathbf{x}|\mathbf{y}_v)}\frac{P(\mathbf{y}_{\bar{v}}|\mathbf{y}_v)}{P(\mathbf{y}_{\bar{v}}|\mathbf{y}_v)} \\
&= -\sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y})\log\frac{P(\mathbf{y}_{\bar{v}},\mathbf{x}|\mathbf{y}_v)}{P(\mathbf{x}|\mathbf{y}_v)P(\mathbf{y}_{\bar{v}}|\mathbf{y}_v)} \\
&\quad -\sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y})\log P(\mathbf{y}_{\bar{v}}|\mathbf{y}_v) \\
&= -\sum_{(\mathbf{x},\mathbf{y})} P(\mathbf{x},\mathbf{y})\log\frac{P(\mathbf{y}_{\bar{v}},\mathbf{x}|\mathbf{y}_v)}{P(\mathbf{x}|\mathbf{y}_v)P(\mathbf{y}_{\bar{v}}|\mathbf{y}_v)} \\
&\quad -\sum_{(\mathbf{y}_{\bar{v}},\mathbf{y}_v)} P(\mathbf{y}_{\bar{v}},\mathbf{y}_v)\log P(\mathbf{y}_{\bar{v}}|\mathbf{y}_v) \\
&= -I(\mathbf{Y}_{\bar{v}},\mathbf{X}|\mathbf{Y}_v) + H(\mathbf{Y}_{\bar{v}}|\mathbf{Y}_v) \\
&= -I(\mathbf{Y}_{\bar{v}},\mathbf{X}|\mathbf{Y}_v)
\end{aligned} \tag{1}$$

35th Conference on Neural Information Processing Systems (NeurIPS 2021).

Table 1: Statistics of real-world PML datasets.

Datasets	#Instances	#Features	#Classes	#CLs (avg.)
YeastBP	560	5548	217	30.43
Music-emotion	6833	98	11	5.29
Music-style	6839	98	10	6.04
MIRFlickr	10433	100	7	3.35

Table 2: Statistics of synthetic PML datasets.

Datasets	#Instances	#Features	#Classes	#GLs (avg.)	#CLs (avg.)	Domain
Enron	1702	1001	53	3.38	5, 7, 9, 11, 13	text
Corel5k	5000	499	374	3.52	5, 7, 9, 11, 13	image
Eurlex-sm	19348	5000	201	2.21	5, 7, 9, 11, 13	text
Eurlex-ed	19348	5000	3993	5.31	7, 9, 11, 13, 15	text
CAL500	502	68	174	26.04	35, 45, 55, 65, 75	music
Mediamill	43907	120	101	4.38	7, 9, 11, 13, 15	video

2 Datasets

The four real-world PML datasets and six synthetic PML datasets are summarized in Table 1 and Table 2 respectively.

3 Configurations

For all PML baselines, we set the trade-off parameters as suggested in the original papers. i.e., PAR-VLS and PAR-MAP: trade-off parameter $\alpha = 0.95$, credible label elicitation threshold $thr = 0.9$ and the number of neighours $k = 10$; DRAMA: $\delta_1 = 0.01$ and $\delta_2 = 1/0.5$; PML-fp and PML-lc: $C_1 = 1$, C_2 is chose from $\{1, 2, \ldots, 10\}$ and C_3 is chose from $\{1, 10, \ldots, 100\}$ with five-fold cross validation; fPML: $\lambda_2 = 1$; PML-LRS: trade-off parameters are set as $\gamma = 0.01$, $\beta = 0.1$ and $\eta = 1$; MUSER: α, β, γ are chosen from $\left\{10^{-3}, \ldots, 10^{3}\right\}$ with a grid search manner. Libsvm is used as the binary learning algorithm for PARTICLE.

4 Experimental Results

Table 3 and Table 4 illustrate the performance comparisons of the proposed MILI-PML with six state-of-the-art PML methods on four real-world datasets and six synthetic datasets in terms of hamming loss and one error metrics respectively.

Table 3: Experimental results of the proposed MILI-PML with six state-of-the-art PML baselines on four real-world as well as six synthetic PML datasets in terms of **hamming loss**. The best result (the smaller the better) is *in bold*.

Dataset	#CLs (avg.)	MILI-PML	PAR-VLS	DRAMA	PML-lc	fPML	PML-LRS	MUSER
YeastBP	30.43	.161±.017	.236±.012	.227±.021	.218±.014	.214±.015	.182±.009	**.158±.012**
Music-emotion	5.29	**.278±.013**	.360±.012	.318±.013	.354±.013	.452±.025	.381±.028	.289±.023
Music-style	6.04	**.142±.014**	.173±.021	.169±.031	.167±.013	.338±.027	.379±.023	.171±.016
MIRFlickr	3.35	**.161±.013**	.193±.017	.219±.014	.216±.018	.223±.022	.237±.012	.193±.056
Enron	7	**.072±.002**	.286±.005	.183±.022	.320±.014	.115±.017	.207±.021	.108±.003
	11	**.085±.004**	.303±.005	.209±.013	.331±.012	.128±.018	.209±.014	.123±.014
Corel5k	7	**.008±.003**	.015±.006	.013±.012	.013±.003	.009±.002	.008±.006	.009±.003
	11	**.010±.004**	.038±.006	.021±.005	.021±.014	.018±.013	.019±.002	.012±.005
Eurlex-sm	7	**.111±.003**	.168±.014	.151±.007	.179±.017	.113±.072	.115±.006	.112±.005
	11	**.167±.006**	.767±.009	.363±.016	.194±.016	.668±.016	.597±.008	.169±.003
Eurlex-ed	9	.048±.006	.061±.013	.058±.008	.085±.016	.098±.009	.078±.008	**.047±.003**
	13	.151±.017	.769±.025	.371±.013	.199±.018	.812±.024	.717±.010	**.149±.002**
CAL500	45	**.225±.011**	.271±.024	.235±.017	.286±.008	.268±.015	.282±.013	.279±.024
	65	**.243±.017**	.357±.032	.327±.036	.375±.014	.288±.015	.327±.026	.283±.014
Mediamill	9	**.061±.004**	.098±.014	.101±.020	.096±.004	.065±.007	.072±.023	.087±.021
	13	**.126±.009**	.145±.018	.201±.027	.218±.025	.513±.021	.191±.017	.183±.012

Table 4: Experimental results of the proposed MILI-PML with six state-of-the-art PML baselines on four real-world as well as six synthetic PML datasets in terms of **one error**. The best result (the smaller the better) is *in bold*.

Dataset	#CLs (avg.)	MILI-PML	PAR-VLS	DRAMA	PML-lc	fPML	PML-LRS	MUSER
YeastBP	30.43	.908±.032	.936±.030	.941±.025	.982±.016	.914±.011	.918±.021	**.907±.027**
Music-emotion	5.29	**.471±.018**	.560±.015	.568±.023	.554±.026	.552±.025	.581±.025	.539±.022
Music-style	6.04	**.371±.017**	.460±.025	.458±.014	.454±.025	.452±.027	.481±.028	.439±.021
MIRFlickr	3.35	**.192±.011**	.263±.013	.289±.016	.346±.021	.393±.022	.497±.031	.223±.026
Enron	7	**.182±.012**	.297±.027	.210±.022	.338±.004	.328±.027	.215±.027	.214±.024
	11	**.196±.014**	.312±.016	.294±.015	.341±.022	.327±.028	.307±.021	.223±.023
Corel5k	7	**.212±.012**	.345±.025	.281±.052	.261±.013	.232±.027	.293±.016	.215±.014
	11	**.225±.014**	.383±.056	.293±.035	.297±.024	.283±.016	.292±.011	.227±.019
Eurlex-sm	7	**.241±.012**	.258±.013	.251±.015	.479±.021	.713±.062	.383±.026	.243±.014
	11	**.247±.015**	.367±.021	.363±.026	.781±.027	.768±.023	.497±.027	.249±.013
Eurlex-ed	9	.341±.031	.394±.024	.371±.032	.683±.054	.728±.076	.417±.059	**.339±.018**
	13	.343±.006	.469±.025	.386±.013	.892±.018	.812±.024	.598±.010	**.342±.017**
CAL500	45	**.165±.007**	.171±.014	.235±.007	.356±.008	.168±.015	.181±.013	.176±.024
	65	**.213±.027**	.363±.012	.337±.016	.265±.014	.237±.019	.247±.016	.233±.021
Mediamill	9	**.227±.010**	.235±.021	.297±.016	.311±.022	.313±.023	.293±.017	.237±.012
	13	**.251±.014**	.298±.015	.301±.012	.316±.016	.325±.018	.312±.023	.277±.021