SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Partial Multi-Label Learning via Label-Specific Feature Corrections

Jun-Yi Hang[1,2] & Min-Ling Zhang[1,2*]

[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*

Appendix A Training Details for Comparative Studies

During training, Pml-md and Pase perform meta-learning on the noisy training set $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{s}_i) | 1 \leqslant i \leqslant m\}$ and the clean validation set $\mathcal{D}^v = \{(\mathbf{x}_i^v, \mathbf{y}_i^v) | 1 \leqslant i \leqslant n\}$. To make a fair comparison with other partial multi-label learning (PML) approaches whose learning procedures involve no clean validation set, we slightly adapt their learning procedures by attaching a supervised component, so that they can acquire clean knowledge from the clean validation set by learning with the supervised component.

Specifically, for Fpml and Natal which learn from PML data with a unified objective function $\mathcal{L}^{train}(\mathcal{D}, \mathbf{W})$, we introduce a supervised objective function $\mathcal{L}^{val}(\mathcal{D}^v, \mathbf{W})$ as follows

$$\mathcal{L}^{train}(\mathcal{D}, \mathbf{W}) + \frac{m}{n} \cdot \mathcal{L}^{val}(\mathcal{D}^v, \mathbf{W}), \tag{A1}$$

where $\mathbf{W}$ denotes the parameters of classifiers. The learning proceeds by optimizing the original objective $\mathcal{L}^{train}$ on noisy training set and the supervised objective $\mathcal{L}^{val}$ on clean validation set, which provides explicit knowledge on 'what is noisy/clean' to facilitate learning. Since Fpml and Natal use the mean square error (MSE) loss to induce classifiers, we also set $\mathcal{L}^{val}$ to be the MSE loss

$$\mathcal{L}^{val}(\mathcal{D}^v, \mathbf{W}) = \sum_{i=1}^{n} ||\mathbf{W}\mathbf{x}_i^v - \mathbf{y}_i^v||_2^2. \tag{A2}$$

For Parvls which learns from PML data with successive two steps (i.e. credible label elicitation and classifier induction), we skip the first step for the clean validation set as this step may corrupt originally clean labels in validation set. We disambiguate the noisy training set via credible label elicitation, and directly merge the disambiguated training set and clean validation set to induce classifier.

For deep approaches Upml-hl and Upml-rl which induce prediction model from PML data with unbiased estimators for Hamming loss and Ranking loss respectively, we introduce corresponding surrogate losses (cf. Eq. (3) and Eq. (5) in their original paper [47]) to acquire clean knowledge from the clean validation set. These surrogate losses are degraded versions of two unbiased estimators when the data is clean. In each training epoch, the prediction model is trained by optimizing the unbiased estimator on noisy training set and the surrogate loss on clean validation set.

Appendix B Further Ablation Studies

In Pase, we implement the feature correction function as a simple affine transformation (i.e. Eq. (2)). To demonstrate such an affine transformation is a reasonable choice to make Pase work well, we decompose it and employ its two components, i.e. a scaling transformation and a translation transformation, to implement two variants respectively.

We employ ten-fold cross validation on all the real-world and synthetic PML data sets. Table B.1 summarizes the p-value statistics of the Wilcoxon signed-ranks test at 0.05 significance level and Figure B.1 shows the detailed experimental results on representative data sets in terms of Average precision, One-error and Ranking loss. Experimental results validate the superiority of the considered affine transformation to these simplified variants.

* Corresponding author (email: zhangml@seu.edu.cn)

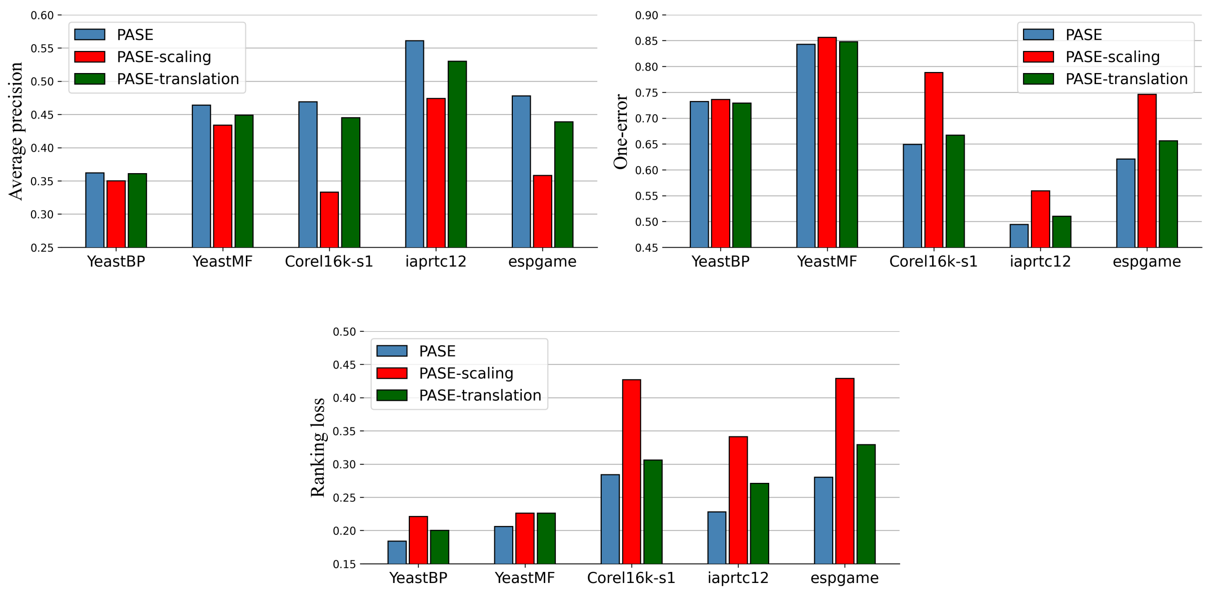


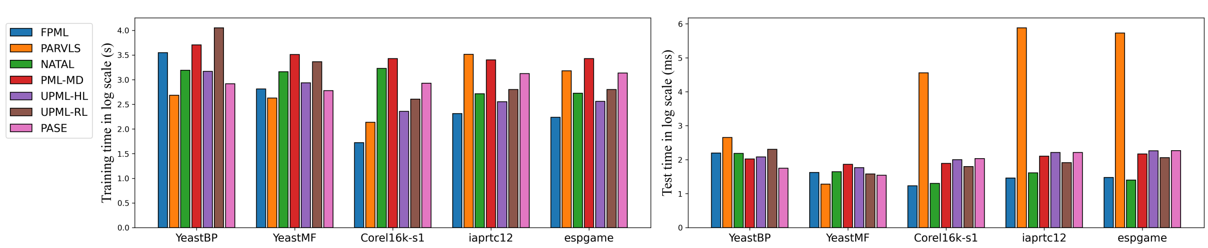
Figure B.1 Predictive performance of PASE and its variants in terms of *Average precision*, *One-error* and *Ranking loss*.

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

PASE against	PASE-scaling	PASE-translation
Average precision	**win** [2.9e-4]	**win** [4.3e-4]
Hamming loss	**win** [6.4e-4]	**win** [1.2e-4]
One error	**win** [1.4e-3]	**win** [1.3e-3]
Coverage	**win** [2.9e-4]	**win** [4.4e-4]
Ranking loss	**win** [2.9e-4]	**win** [2.9e-4]

Appendix C Empirical Running Time Comparison

Empirical running time of each comparing approach considered in the *Comparative Studies* part of the main body is further reported here for comprehensive evaluation. Figure C.1 illustrates the empirical training and test time of each comparing approach, which shows that PASE is comparable to existing approaches in time overhead.



(a) Training time of comparing approaches **(b) Test time of comparing approaches**

Figure C.1 Running time (training/test) of each comparing approach on five benchmark data sets. For histogram illustration, the y-axis corresponds to the logarithm of running time.