模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

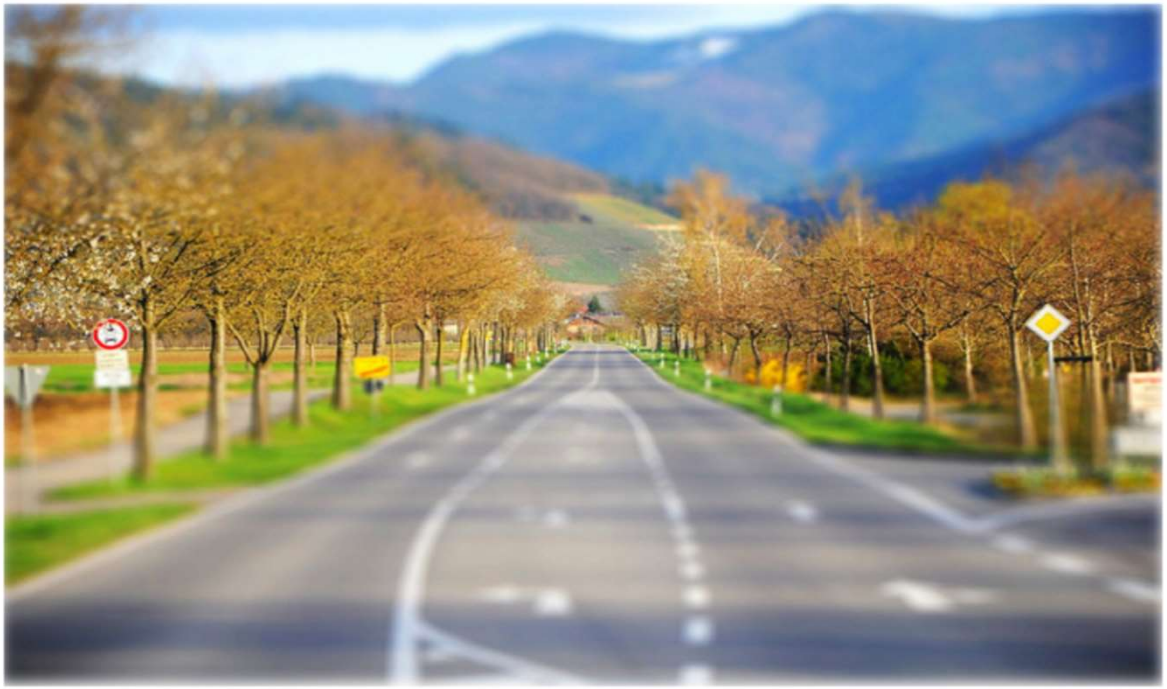
ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

Partial Multi-Label Learning with Probabilistic Graphical Disambiguation

Jun-Yi Hang, Min-Ling Zhang*
School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
Key Laboratory of Computer Network and Information Integration (Southeast University),
Ministry of Education, China
{hangjy, zhangml}@seu.edu.cn

NIPS 2023

Multi-label Learning



tree

light
mountain

road

Input space $\mathcal{X} = \mathbb{R}^d$

Candidate label set $S \subseteq \mathcal{Y} = \{l_1, l_2, \ldots, l_t\}$

Background

Straightforward strategy : Simply treat all the candidate labels as valid ones !

Will be misled by the noisy labels …

Disambiguation strategy : Recover the ground-truth labeling information from candidate labels.

Heuristics or Rules

① Smoothness assumption ② Low-rank constraint ③ Sparsity regularization

Will be hardly held in some challenging scenarios …

Background

We need a principled manner for disambiguation!

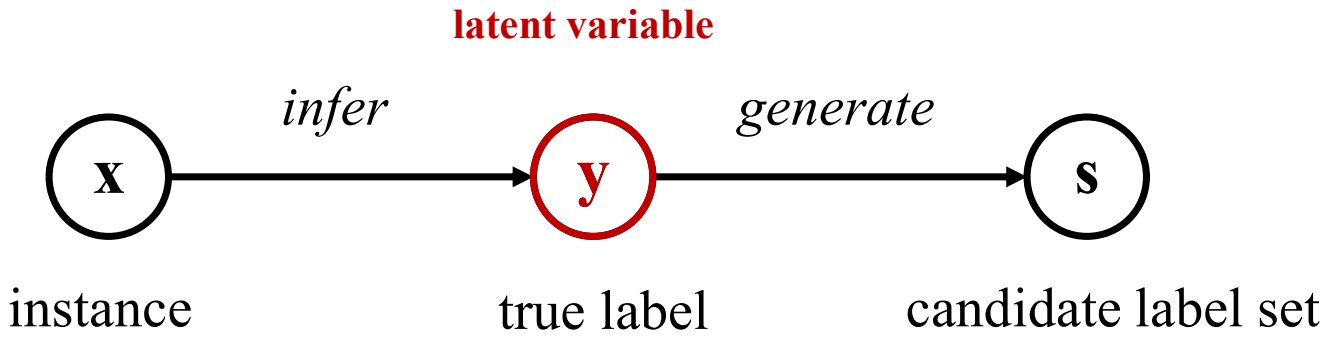
A few attempts

PML-GAN	with an **instance reconstruction** process.
MILI-PML	maximizing the **mutual information** between features and identified ground-truth labels
PML-MD	with the guidance from an **accurately annotated validation set**

This paper： Regard disambiguation as a task of **latent variable inference**.

Method

PARD (Partial multi-label learning with probAbilistic-gRaphical Disambiguation)



① sample $\mathbf{x}$ from the marginal distribution $p_\theta(\mathbf{x})$

② sample $\mathbf{y}$ from the ground-truth class posterior distribution $p_\theta(\mathbf{y}|\mathbf{x})$

③ obtain its candidate labels $\mathbf{s}$ by $p_\theta(\mathbf{s}|\mathbf{x},\mathbf{y})$

Joint probability: $p_\theta(\mathbf{x},\mathbf{y},\mathbf{s}) = p_\theta(\mathbf{x})p_\theta(\mathbf{y}|\mathbf{x})p_\theta(\mathbf{s}|\mathbf{x},\mathbf{y}).$

Method

Joint probability: $p_\theta(\mathbf{x}, \mathbf{y}, \mathbf{s}) = p_\theta(\mathbf{x}) p_\theta(\mathbf{y}|\mathbf{x}) p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y}).$

Due to computational intractability …

To make the optimization tractable, the **variational lower bound** of the log-likelihood is derived as follows:

$$\log p_\theta(\mathbf{s}|\mathbf{x}) \geq \mathcal{L}(\mathbf{x}, \mathbf{s}; \theta, \phi) = \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log \frac{p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y}) p_\theta(\mathbf{y}|\mathbf{x})}{q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})}]$$

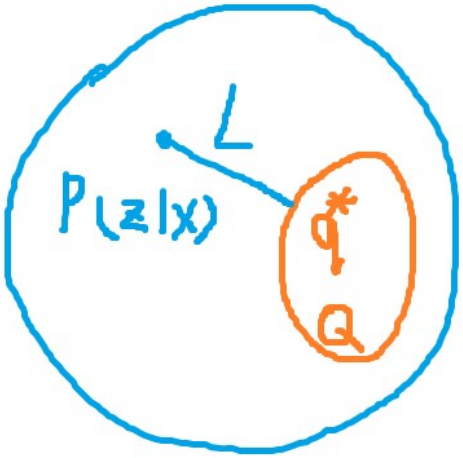
where $q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})$ is the introduced **variational posterior** to approximate the true posterior $p_\theta(\mathbf{y}|\mathbf{x}, \mathbf{s})$.

What is the variational inference ?

Variational Inference (VI)

Transform **the posterior inference problem** into an **optimization problem** !

$$p(\boldsymbol{z} \mid \boldsymbol{x}) \qquad\qquad q^*(\boldsymbol{z})$$



Optimization: $q^*(\boldsymbol{z}) = \arg\min_{q(\boldsymbol{z}) \in Q} \mathcal{L}(q(\boldsymbol{z}) \| p(\boldsymbol{z} \mid \boldsymbol{x}))$

Variational Bayes (VB): $= \arg\min_{q(\boldsymbol{z}) \in Q} \mathrm{KL}(q(\boldsymbol{z}) \| p(\boldsymbol{z} \mid \boldsymbol{x}))$

Variational Inference (VI)

$$
\begin{aligned}
\mathrm{KL}(q(\boldsymbol{z})\|p(\boldsymbol{z}\mid\boldsymbol{x})) &= -\int_{\boldsymbol{z}} q(\boldsymbol{z})\log\left[\frac{p(\boldsymbol{z}\mid\boldsymbol{x})}{q(\boldsymbol{z})}\right]d\boldsymbol{z} \\
&= \int_{\boldsymbol{z}} q(\boldsymbol{z})\log q(\boldsymbol{z})d\boldsymbol{z} - \int_{\boldsymbol{z}} q(\boldsymbol{z})\log p(\boldsymbol{z}\mid\boldsymbol{x})d\boldsymbol{z} \\
&= \mathbb{E}_q[\log q(\boldsymbol{z})] - \mathbb{E}_q[\log p(\boldsymbol{z}\mid\boldsymbol{x})] \\
&= \mathbb{E}_q[\log q(\boldsymbol{z})] - \mathbb{E}_q\left[\log\left[\frac{p(\boldsymbol{x},\boldsymbol{z})}{p(\boldsymbol{x})}\right]\right] \\
&= \mathbb{E}_q[\log q(\boldsymbol{z})] - \mathbb{E}_q[\log p(\boldsymbol{x},\boldsymbol{z})] + \mathbb{E}_q[\log p(\boldsymbol{x})] \\
&= \underbrace{\mathbb{E}_q[\log q(\boldsymbol{z})] - \mathbb{E}_q[\log p(\boldsymbol{x},\boldsymbol{z})]}_{\text{- ELBO}} + \log p(\boldsymbol{x})
\end{aligned}
$$

- ELBO (Evidence Lower Bound)

Variational Inference (VI)

Evidence:

$$\begin{aligned}\log p(\boldsymbol{x}) &= \mathrm{ELBO} + \mathrm{KL}(q(\boldsymbol{z}) \| p(\boldsymbol{z} \mid \boldsymbol{x})) \\ &\geq \mathrm{ELBO}\end{aligned}$$

Optimization:

$$\begin{aligned}q^*(\boldsymbol{z}) &= \underset{q(\boldsymbol{z}) \in Q}{\operatorname{argmin}} \, \mathrm{KL}(q(\boldsymbol{z}) \| p(\boldsymbol{z} \mid \boldsymbol{x})) \\ &= \underset{q(\boldsymbol{z}) \in Q}{\operatorname{argmax}} \, \mathrm{ELBO}(q)\end{aligned}$$

Look back to the optimization objective of this paper:

$$\log p_\theta(\mathbf{s}|\mathbf{x}) \geq \mathcal{L}(\mathbf{x}, \mathbf{s}; \theta, \phi) = \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log \frac{p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y}) p_\theta(\mathbf{y}|\mathbf{x})}{q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})}]$$

The **variational lower bound** of the log-likelihood is derived as follows:

$$
\begin{aligned}
\log p_\theta(\mathbf{s}|\mathbf{x}) &= \log \int p_\theta(\mathbf{s}, \mathbf{y}|\mathbf{x}) d\mathbf{y} \\
&= \log \int p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y}) p_\theta(\mathbf{y}|\mathbf{x}) d\mathbf{y} \\
&= \log \int q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s}) \frac{p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y}) p_\theta(\mathbf{y}|\mathbf{x})}{q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})} d\mathbf{y} \\
&\geq \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log \frac{p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y}) p_\theta(\mathbf{y}|\mathbf{x})}{q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})}] \\
&= \mathcal{L}(\mathbf{x}, \mathbf{s}; \theta, \phi).
\end{aligned}
$$

Method

$\mathcal{L}(\mathbf{x}, \mathbf{s}; \theta, \phi)$ is the derived variational lower bound as a surrogate loss function of the log-likelihood, which can be rewritten as

$$\mathcal{L}(\mathbf{x}, \mathbf{s}; \theta, \phi) = \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y})] - KL[q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})||p_\theta(\mathbf{y}|\mathbf{x})]$$

① $p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y})$ (a.k.a. generative model)

② variational posterior $q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})$ (a.k.a. inference model)

③ ground-truth class posterior distribution $p_\theta(\mathbf{y}|\mathbf{x})$

unfolding the KL-divergence term

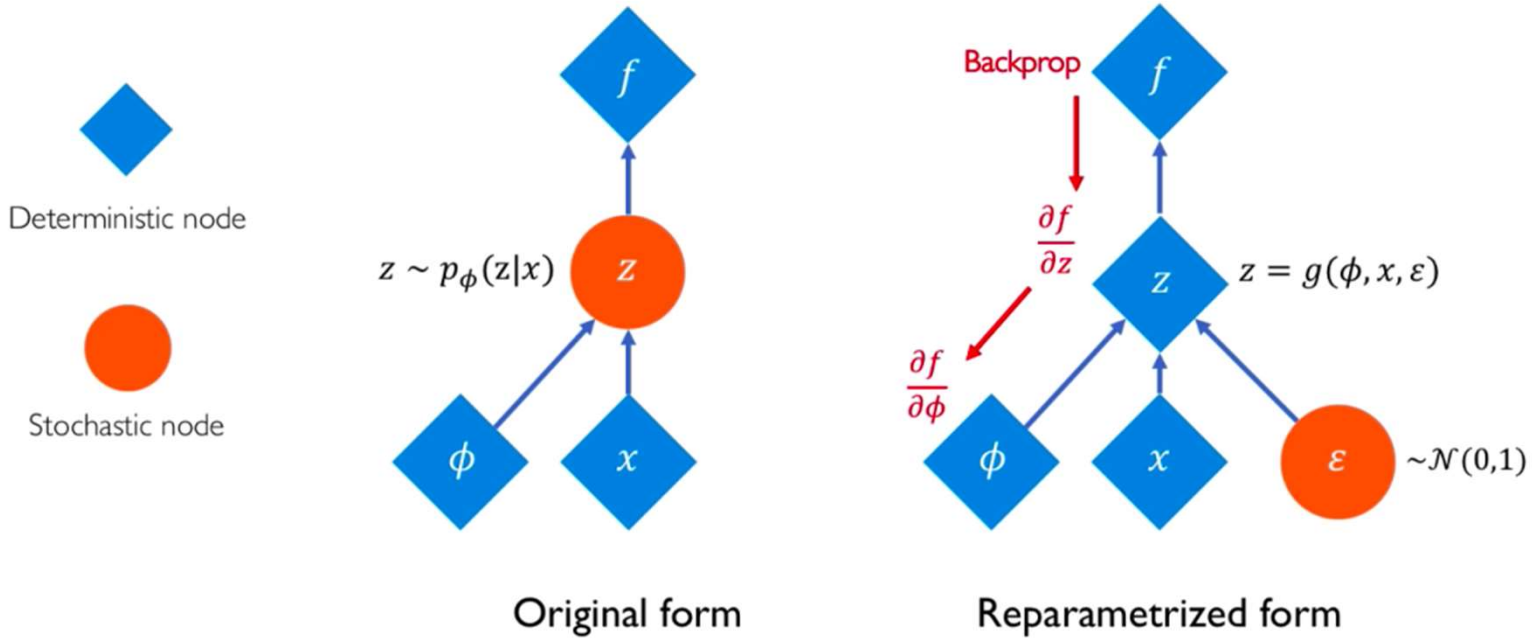
$$\begin{aligned}\mathcal{L}(\mathbf{x}, \mathbf{s}; \theta, \phi) = &\mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y})] + H[q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})] \\ &+ \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log p_\theta(\mathbf{y}|\mathbf{x})],\end{aligned}$$

Method

$$\mathcal{L}(\mathbf{x}, \mathbf{s}; \theta, \phi) = \boxed{\mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y})]} - KL[q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})||p_\theta(\mathbf{y}|\mathbf{x})]$$

$\mathcal{O}(2^t)$

Continuous Relaxing with Gumbel-Softmax Trick.



Method

$$\mathcal{L}(\mathbf{x}, \mathbf{s}; \theta, \phi) = \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y})] - KL[q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})||p_\theta(\mathbf{y}|\mathbf{x})]$$

$\mathcal{O}(2^t)$

Continuous Relaxing with Gumbel-Softmax Trick.

The Bernoulli variable **y** is relaxed by the binary Concrete variable **c** :

$$\mathbf{c} \sim BinConcrete(\mathbf{p}, \tau) \qquad \mathbf{c} = \frac{1}{1 + \exp[-(\log \boldsymbol{\alpha} + \mathbf{l})/\tau]}$$

where $\boldsymbol{\alpha} = \frac{\mathbf{p}}{1-\mathbf{p}}$ and **p** is the parameter of the multivariate Bernoulli distribution $q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})$. **l** is a sampling from Logistic distribution and $\tau > 0$ is a temperature parameter.

With the above sampling trick :

$$\mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y})] \approx \frac{1}{L}\sum_{i=1}^{L} \log p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{c}^{(i)})$$

$$where \quad \mathbf{c}^{(i)} \sim BinConcrete(\mathbf{p}, \tau).$$

Method

$$\mathcal{L}(\mathbf{x}, \mathbf{s}; \theta, \phi) = \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log p_\theta(\mathbf{s}|\mathbf{x}, \mathbf{y})] - KL[q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})||p_\theta(\mathbf{y}|\mathbf{x})]$$

Closed-Form Solution of the KL-Divergence Term.

To make it tractable, we exploit mean-field approximation technique and derive a closed-form solution of the KL-divergence term as follows

$$KL[q_\phi(\mathbf{y}|\mathbf{x}, \mathbf{s})||p_\theta(\mathbf{y}|\mathbf{x})] = \sum_{k=1}^{t} p_\phi^{y_k} \log \frac{p_\phi^{y_k}}{p_\theta^{y_k}} + (1 - p_\phi^{y_k}) \log \frac{1 - p_\phi^{y_k}}{1 - p_\theta^{y_k}},$$

where $p_\phi^{y_k} = q_\phi(y_k = 1|\mathbf{x}, \mathbf{s})$ and $p_\theta^{y_k} = p_\theta(y_k = 1|\mathbf{x})$.

Method

$$
\begin{aligned}
& KL[q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})||p_\theta(\mathbf{y}|\mathbf{x})] \\
&= \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s}) - \log p_\theta(\mathbf{y}|\mathbf{x})] \\
&= \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\sum_{k=1}^{t} \log q_\phi(y_k|\mathbf{x},\mathbf{s}) - \sum_{k=1}^{t} \log p_\theta(y_k|\mathbf{x})] \\
&= \sum_{k=1}^{t} \mathbb{E}_{q_\phi(\mathbf{y}|\mathbf{x},\mathbf{s})}[\log \frac{q_\phi(y_k|\mathbf{x},\mathbf{s})}{p_\theta(y_k|\mathbf{x})}] \\
&= \sum_{k=1}^{t} \mathbb{E}_{q_\phi(y_k|\mathbf{x},\mathbf{s})}[\log \frac{q_\phi(y_k|\mathbf{x},\mathbf{s})}{p_\theta(y_k|\mathbf{x})}] \\
&= \sum_{k=1}^{t} KL[q_\phi(y_k|\mathbf{x},\mathbf{s})||p_\theta(y_k|\mathbf{x})] \quad \text{(where } p_\phi^{y_k} = q_\phi(y_k=1|\mathbf{x},\mathbf{s}) \text{ and } p_\theta^{y_k} = p_\theta(y_k=1|\mathbf{x})\,.\text{)} \\
&= \sum_{k=1}^{t} p_\phi^{y_k} \log \frac{p_\phi^{y_k}}{p_\theta^{y_k}} + (1-p_\phi^{y_k}) \log \frac{1-p_\phi^{y_k}}{1-p_\theta^{y_k}}.
\end{aligned}
$$

Experiment

Data sets	$\gamma\%$	*Average precision* ↑						
		FPML	PARVLS	PML-NI	PML-MD	UPML-HL	UPML-RL	PARD
YeastBP		0.202±0.013•	0.059±0.003•	**0.430±0.018**○	0.310±0.016•	0.392±0.020•	0.170±0.009•	0.417±0.015
YeastCC		0.375±0.019•	0.145±0.009•	0.610±0.022•	0.527±0.027•	0.596±0.022•	0.459±0.021•	**0.622±0.020**
YeastMF		0.279±0.020•	0.114±0.007•	0.471±0.022•	0.422±0.027•	0.451±0.016•	0.329±0.023•	**0.491±0.028**
Music_emotion		0.571±0.013•	0.603±0.011•	0.600±0.012•	0.646±0.011	**0.656±0.012**○	0.639±0.012•	0.650±0.007
Music_style		0.702±0.014•	0.720±0.012•	0.733±0.012•	0.716±0.012•	0.734±0.010•	0.703±0.012•	**0.742±0.009**
corel5k	100	0.485±0.014•	0.391±0.017•	0.473±0.016•	0.491±0.019•	0.516±0.015•	0.514±0.015•	**0.533±0.013**
	150	0.478±0.013•	0.408±0.011•	0.453±0.013•	0.479±0.015•	0.514±0.015•	0.512±0.015•	**0.523±0.013**
	200	0.474±0.013•	0.395±0.021•	0.438±0.012•	0.472±0.011•	0.503±0.009•	0.511±0.013	**0.516±0.011**
	250	0.470±0.012•	0.391±0.022•	0.426±0.011•	0.470±0.019•	0.502±0.013•	0.507±0.014	**0.513±0.012**
rcv1-s1	100	0.682±0.010•	0.520±0.021•	0.693±0.014•	0.678±0.010•	0.712±0.013•	0.689±0.007•	**0.720±0.008**
	150	0.678±0.011•	0.513±0.022•	0.664±0.013•	0.656±0.011•	0.705±0.013•	0.690±0.008•	**0.716±0.010**
	200	0.673±0.010•	0.495±0.021•	0.649±0.010•	0.648±0.010•	0.704±0.012•	0.685±0.005•	**0.710±0.010**
	250	0.666±0.011•	0.488±0.016•	0.630±0.011•	0.640±0.011•	0.693±0.015•	0.672±0.009•	**0.706±0.009**
Corel16k-s1	100	0.535±0.010•	0.430±0.010•	0.518±0.011•	0.522±0.011•	0.539±0.012•	0.525±0.007•	**0.551±0.012**
	150	0.529±0.008•	0.423±0.012•	0.499±0.011•	0.512±0.011•	0.532±0.011•	0.522±0.009•	**0.545±0.012**
	200	0.523±0.010•	0.404±0.012•	0.487±0.013•	0.507±0.012•	0.504±0.011•	0.519±0.009•	**0.536±0.014**
	250	0.514±0.011•	0.389±0.011•	0.476±0.013•	0.501±0.011•	0.507±0.014•	0.516±0.008•	**0.531±0.014**
iaprtc12	100	0.603±0.006•	0.591±0.008•	0.599±0.009•	0.600±0.008•	0.601±0.009•	0.563±0.008•	**0.621±0.011**
	150	0.600±0.006•	0.585±0.008•	0.580±0.010•	0.583±0.010•	0.597±0.007•	0.559±0.007•	**0.615±0.009**
	200	0.597±0.008•	0.574±0.008•	0.563±0.008•	0.569±0.009•	0.600±0.008•	0.554±0.008•	**0.610±0.008**
	250	0.588±0.009•	0.561±0.008•	0.543±0.009•	0.560±0.009•	0.582±0.008•	0.547±0.006•	**0.601±0.009**
espgame	100	0.498±0.007•	0.472±0.007•	0.478±0.008•	0.491±0.009•	0.501±0.008•	0.466±0.009•	**0.515±0.008**
	150	0.496±0.007•	0.456±0.007•	0.459±0.007•	0.481±0.007•	0.509±0.008	0.471±0.006•	**0.511±0.009**
	200	0.495±0.007•	0.445±0.007•	0.446±0.005•	0.473±0.007•	0.475±0.012•	0.466±0.006•	**0.508±0.008**
	250	0.489±0.009•	0.427±0.009•	0.431±0.007•	0.467±0.007•	0.490±0.008•	0.460±0.006•	**0.501±0.009**

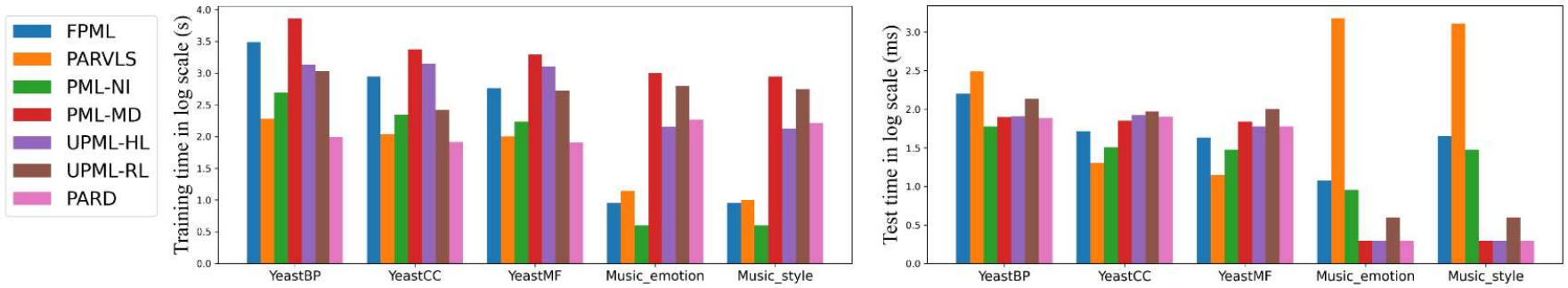
Experiment

Data sets	$\gamma\%$	*Ranking loss* ↓						
		FPML	PARVLS	PML-NI	PML-MD	UPML-HL	UPML-RL	PARD
YeastBP		0.338±0.007•	0.482±0.005•	0.197±0.012•	0.218±0.008•	0.222±0.015•	0.309±0.011•	**0.175±0.010**
YeastCC		0.305±0.016•	0.480±0.010•	0.156±0.017•	0.190±0.017•	0.167±0.017•	0.212±0.015•	**0.135±0.017**
YeastMF		0.373±0.019•	0.533±0.010•	0.226±0.018•	0.235±0.014•	0.239±0.023•	0.286±0.017•	**0.192±0.015**
Music_emotion		0.274±0.012•	0.252±0.009•	0.249±0.010•	0.231±0.010	**0.223±0.009**○	0.236±0.011•	0.228±0.007
Music_style		0.160±0.010•	0.150±0.007•	0.139±0.008	0.143±0.007•	0.138±0.007	0.149±0.007•	**0.137±0.009**
corel5k	100	0.255±0.013•	0.357±0.020•	0.280±0.012•	0.251±0.011•	0.236±0.014•	0.229±0.013•	**0.223±0.012**
	150	0.260±0.013•	0.327±0.011•	0.296±0.015•	0.260±0.014•	0.235±0.014	0.235±0.013	**0.233±0.010**
	200	0.263±0.011•	0.338±0.021•	0.312±0.013•	0.264±0.012•	0.254±0.010•	**0.240±0.012**	0.242±0.010
	250	0.268±0.012•	0.347±0.024•	0.322±0.013•	0.267±0.015•	0.257±0.013•	**0.242±0.013**	0.245±0.010
rcv1-s1	100	0.101±0.004•	0.222±0.013•	0.106±0.007•	0.093±0.005•	0.086±0.005•	0.087±0.003•	**0.076±0.003**
	150	0.102±0.005•	0.240±0.020•	0.126±0.006•	0.105±0.005•	0.089±0.004•	0.089±0.004•	**0.079±0.003**
	200	0.107±0.006•	0.262±0.022•	0.138±0.005•	0.111±0.004•	0.092±0.007•	0.093±0.003•	**0.084±0.004**
	250	0.114±0.006•	0.277±0.017•	0.153±0.005•	0.115±0.005•	0.099±0.006•	0.100±0.004•	**0.090±0.003**
Corel16k-s1	100	0.219±0.006•	0.289±0.007•	0.247±0.009•	0.221±0.008•	0.212±0.006•	0.216±0.006•	**0.205±0.006**
	150	0.226±0.006•	0.300±0.011•	0.262±0.009•	0.228±0.006•	0.224±0.007•	0.220±0.007•	**0.212±0.007**
	200	0.235±0.007•	0.329±0.011•	0.274±0.011•	0.232±0.008•	0.250±0.009•	0.223±0.006	**0.221±0.009**
	250	0.245±0.009•	0.369±0.010•	0.286±0.012•	0.238±0.007•	0.249±0.011•	0.228±0.006•	**0.223±0.008**
iaprtc12	100	0.189±0.004•	0.207±0.005•	0.206±0.006•	0.192±0.005•	0.199±0.007•	0.210±0.004•	**0.180±0.006**
	150	0.191±0.004•	0.212±0.005•	0.222±0.007•	0.203±0.005•	0.203±0.006•	0.211±0.005•	**0.186±0.005**
	200	0.196±0.004•	0.227±0.006•	0.239±0.006•	0.214±0.005•	0.203±0.006•	0.216±0.005•	**0.191±0.005**
	250	0.204±0.006•	0.241±0.005•	0.257±0.006•	0.222±0.006•	0.218±0.007•	0.221±0.006•	**0.199±0.005**
espgame	100	0.252±0.005•	0.281±0.006•	0.285±0.004•	0.258±0.006•	0.256±0.006•	0.277±0.008•	**0.241±0.005**
	150	0.256±0.006•	0.300±0.005•	0.302±0.005•	0.266±0.006•	0.248±0.006•	0.265±0.006•	**0.245±0.006**
	200	0.258±0.006•	0.313±0.006•	0.314±0.005•	0.272±0.006•	0.277±0.008•	0.269±0.005•	**0.248±0.006**
	250	0.266±0.006•	0.330±0.007•	0.329±0.007•	0.274±0.006•	0.269±0.007•	0.275±0.006•	**0.255±0.005**

Experiment

Table 4: Summary of the Wilcoxon signed-ranks test for PARD against other comparing approaches at 0.05 significance level. p-values are shown in the brackets.

PARD against	FPML	PARVLS	PML-NI	PML-MD	UPML-HL	UPML-RL
Average precision	**win** [1.2e-5]	**win** [1.2e-5]	**win** [1.8e-5]	**win** [1.2e-5]	**win** [1.7e-5]	**win** [1.2e-5]
Hamming loss	**win** [8.2e-3]	**win** [4.6e-5]	**win** [2.1e-5]	**win** [8.5e-5]	**tie** [7.0e-1]	**win** [3.5e-5]
One-error	**win** [1.2e-5]	**win** [1.2e-5]	**win** [4.6e-5]	**win** [1.2e-5]	**win** [1.8e-4]	**win** [1.2e-5]
Coverage	**win** [1.2e-5]	**win** [1.2e-5]	**win** [1.4e-5]	**win** [1.2e-5]	**win** [2.4e-5]	**win** [4.0e-5]
Ranking loss	**win** [1.2e-5]	**win** [1.2e-5]	**win** [1.2e-5]	**win** [1.2e-5]	**win** [2.0e-5]	**win** [2.5e-5]



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

Figure 4: Running time (training/test) of each comparing approach on real-world data sets. For histogram illustration, the y-axis corresponds to the logarithm of running time.

Thanks