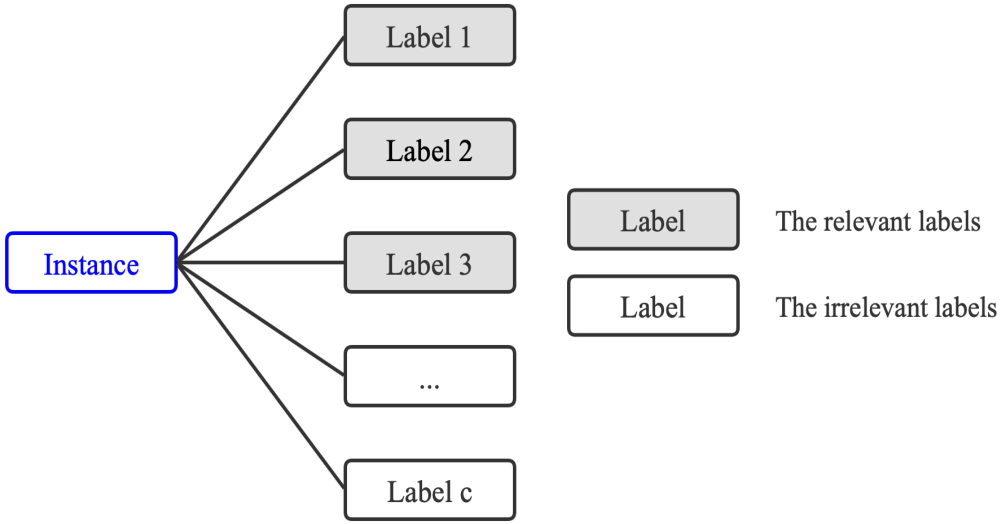
Can Class-priors Help Single-Positive Multi-Label Learning

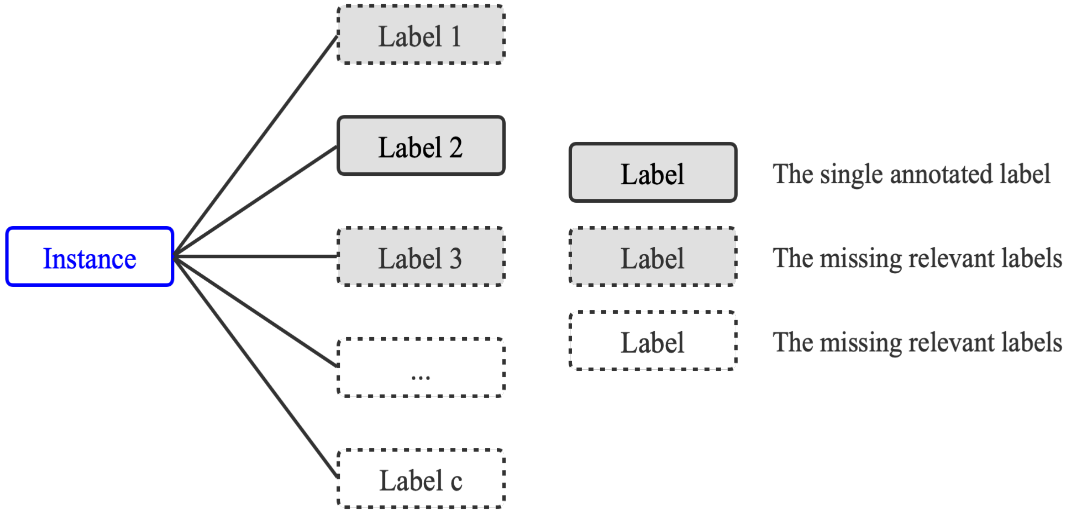
Biao Liu, Ning Xu*, Jie Wang, Xin Geng
School of Computer Science and Engineering, Southeast University

Single-Positive Multi-Label Learning

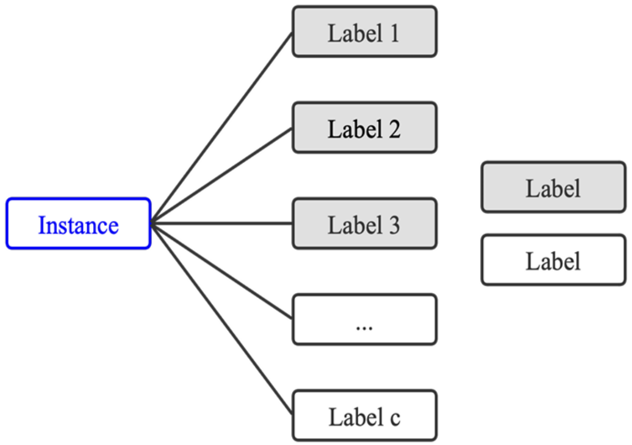
Multi-Label Learning

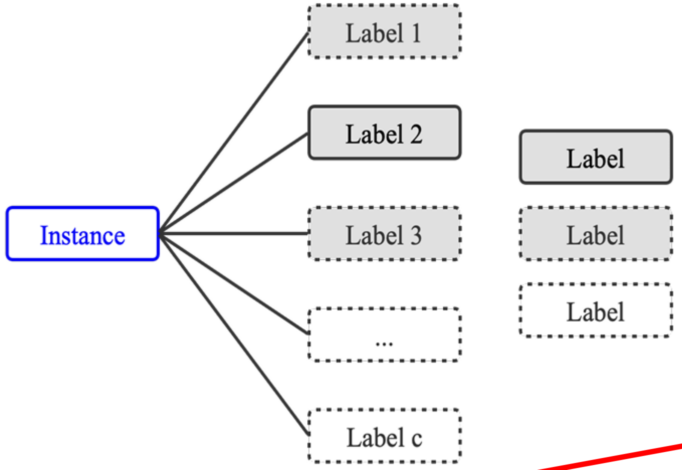


Single-Positive Multi-Label Learning

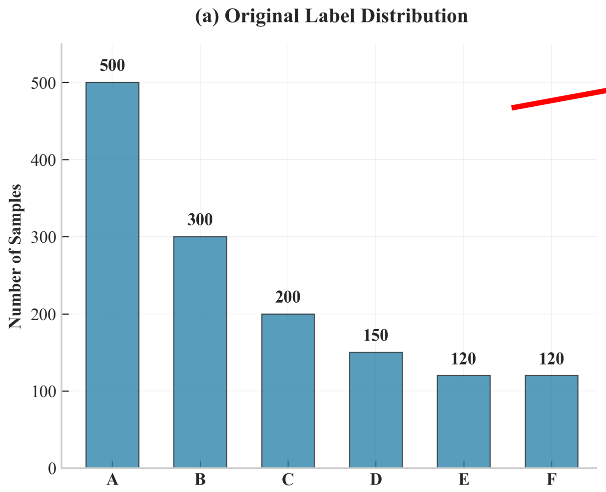


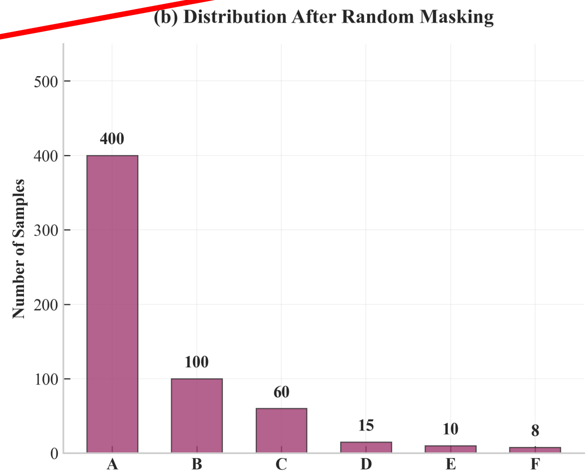
Class Imbalance in Single-Positive Multi-Label Learning





The problem of class imbalance is exacerbated





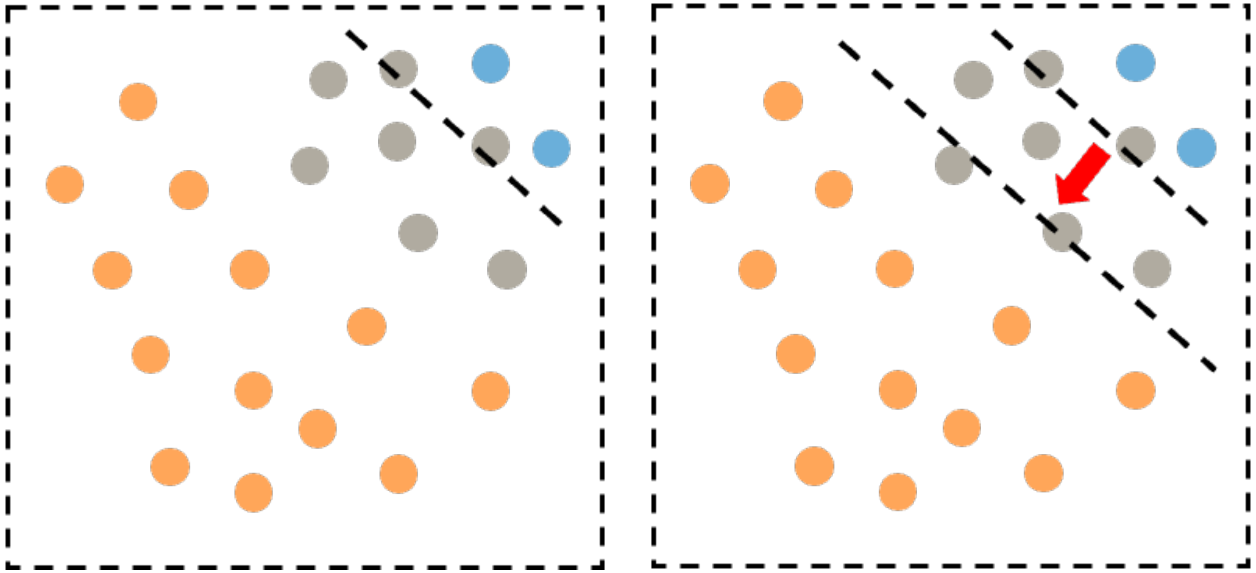
Class prior probability estimator

Estimating class prior probabilities with dynamic thresholds

Theorem. Define $z^\star = \arg\min_{z\in[0,1]} q_j^n(z)/q_j^p(z)$, for every $0<\delta<1$, define $\hat{z} = \boxed{\arg\min_{z\in[0,1]}\left(\frac{\hat{q}_j(z)}{\hat{q}_j^p(z)} + \frac{1+\tau}{\hat{q}_j^p(z)}\left(\sqrt{\frac{\log(4/\delta)}{2n}} + \sqrt{\frac{\log(4/\delta)}{2n_j^p}}\right)\right)}$. Assume $n_j^p \geq 2\frac{\log 4/\delta}{q_j^p(z^\star)}$, the estimated class-prior $\hat{\pi}_j = \frac{\hat{q}_j(\hat{z})}{\hat{q}_j^p(\hat{z})}$ satisfies with probability at least $1-\delta$:

$$\pi_j - \underbrace{\boxed{\frac{c_1}{q_j^p(z^\star)}\left(\sqrt{\frac{\log(4/\delta)}{2n}} + \sqrt{\frac{\log(4/\delta)}{2n_j^p}}\right)}}_{\text{Lower bound error}} \leq \hat{\pi}_j \leq \pi_j + \underbrace{\boxed{(1-\pi_j)\frac{q_j^n(z^\star)}{q_j^p(z^\star)} + \frac{c_2}{q_j^p(z^\star)}\left(\sqrt{\frac{\log(4/\delta)}{2n}} + \sqrt{\frac{\log(4/\delta)}{2n_j^p}}\right)}}_{\text{upper bound error}}$$

Calibrating the decision boundary



$$\hat{\mathcal{R}}_{sp}(f) = \sum_{j=1}^{c} \frac{2\pi_j}{\left|\mathcal{S}_{L_j}\right|} \sum\nolimits_{\boldsymbol{x}\in\mathcal{S}_{L_j}} \left(1 - \frac{e^{\boxed{g^j(\boldsymbol{x})-\lambda b^j}}}{e^{\boxed{g^j(\boldsymbol{x})-\lambda b^j}}+1}\right) + \left|\frac{1}{n}\sum\nolimits_{\boldsymbol{x}\in\tilde{\mathcal{D}}}\left(f^j(\boldsymbol{x})-\pi_j\right)\right|$$

Experiments

Table 1: Predictive performance of each comparing method on four MLIC datasets in terms of *mean average precision (mAP)* (mean ± std). The best performance is highlighted in bold (the larger the better).

	VOC	COCO	NUS	CUB
AN	85.546±0.294	64.326±0.204	42.494±0.338	18.656±0.090
AN-LS	87.548±0.137	67.074±0.196	43.616±0.342	16.446±0.269
WAN	87.138±0.240	65.552±0.171	45.785±0.192	14.622±1.300
EPR	85.228±0.444	63.604±0.249	45.240±0.338	19.842±0.423
ROLE	88.088±0.167	67.022±0.141	41.949±0.205	14.798±0.613
EM	88.674±0.077	70.636±0.094	47.254±0.297	20.692±0.527
EM-APL	88.860±0.080	70.758±0.215	47.778±0.181	21.202±0.792
SMILE	87.314±0.150	70.431±0.213	47.241±0.172	18.611±0.144
PLC	88.021±0.121	70.422±0.062	46.211±0.155	**21.840±0.237**
LL-R	87.784±0.063	70.078±0.008	48.048±0.074	18.966±0.022
LL-CP	87.466±0.031	70.460±0.032	48.000±0.077	19.310±0.164
LL-CT	87.054±0.214	70.384±0.058	47.930±0.010	19.012±0.097
BOOSTLU+LL-R	89.224±0.017	73.272±0.006	49.590±0.021	19.136±0.009
BOOSTLU+LL-CP	88.358±0.212	70.820±0.030	47.810±0.166	18.166±0.063
BOOSTLU+LL-CT	88.528±0.053	71.742±0.006	48.216±0.021	17.952±0.007
CRISP	**89.820±0.191**	**74.640±0.219**	**49.996±0.316**	21.650±0.178

Experiments

Table 2: Predictive performance of each comparing method on MLL datasets in terms of
loss (mean ± std). The best performance is highlighted in bold (the smaller the better).

	Image	Scene	Yeast	Corel5k	Mirflickr	Delicious
AN	0.432±0.067	0.321±0.113	0.383±0.066	0.140±0.000	0.125±0.002	0.131±0.000
AN-LS	0.378±0.041	0.246±0.064	0.365±0.031	0.186±0.003	0.163±0.006	0.213±0.007
WAN	0.354±0.051	0.216±0.023	0.212±0.021	0.129±0.000	0.121±0.002	0.126±0.000
EPR	0.401±0.053	0.291±0.056	0.208±0.010	0.139±0.000	0.119±0.001	0.126±0.000
ROLE	0.340±0.059	0.174±0.028	0.213±0.017	0.259±0.004	0.182±0.014	0.336±0.007
EM	0.471±0.044	0.322±0.115	0.261±0.030	0.155±0.002	0.134±0.004	0.164±0.001
EM-APL	0.508±0.028	0.420±0.069	0.245±0.026	0.135±0.001	0.138±0.003	0.163±0.003
SMILE	0.260±0.020	0.161±0.045	0.167±0.002	0.125±0.003	0.120±0.002	0.126±0.000
LL-R	0.346±0.072	0.155±0.021	0.227±0.001	0.114±0.001	0.123±0.003	0.129±0.002
LL-CP	0.329±0.041	0.148±0.017	0.215±0.000	0.114±0.003	0.124±0.003	0.160±0.001
LL-CT	0.327±0.019	0.180±0.038	0.238±0.001	0.115±0.001	0.124±0.002	0.160±0.000
CRISP	**0.164±0.027**	**0.112±0.021**	**0.164±0.001**	**0.113±0.001**	**0.118±0.001**	**0.122±0.000**

Table 9: Predictive performance of each comparing method on MLL datasets in terms of *Coverage* (mean ± std). The best performance is highlighted in bold (the smaller the better).

	Image	Scene	Yeast	Corel5k	Mirflickr	Delicious
AN	0.374±0.050	0.279±0.094	0.707±0.045	0.330±0.001	0.342±0.003	0.653±0.001
AN-LS	0.334±0.033	0.217±0.052	0.703±0.012	0.441±0.009	0.433±0.015	0.830±0.016
WAN	0.313±0.040	0.192±0.019	0.512±0.045	0.309±0.001	0.334±0.002	0.632±0.001
EPR	0.352±0.043	0.254±0.046	0.506±0.011	0.328±0.001	0.332±0.002	0.637±0.001
ROLE	0.306±0.049	0.157±0.023	0.519±0.026	0.551±0.007	0.448±0.028	0.887±0.004
EM	0.407±0.036	0.281±0.096	0.575±0.042	0.382±0.005	0.359±0.010	0.753±0.004
EM-APL	0.438±0.022	0.360±0.057	0.556±0.045	0.335±0.005	0.369±0.005	0.765±0.006
SMILE	0.242±0.014	0.146±0.037	0.462±0.003	0.308±0.007	0.328±0.004	0.628±0.003
LL-R	0.311±0.059	0.141±0.017	0.512±0.002	0.274±0.002	0.335±0.006	0.622±0.001
LL-CP	0.296±0.031	0.136±0.016	0.518±0.001	**0.272±0.008**	0.337±0.005	0.708±0.004
LL-CT	0.297±0.017	0.161±0.031	0.509±0.001	0.277±0.005	0.335±0.003	0.708±0.002
CRISP	**0.164±0.012**	**0.082±0.018**	**0.455±0.002**	0.276±0.002	**0.324±0.001**	**0.620±0.001**

Table 8: Predictive performance of each comparing method on MLL datasets in terms o
Precision (mean ± std). The best performance is highlighted in bold (the larger the better)

	Image	Scene	Yeast	Corel5k	Mirflickr	Delicious
AN	0.534±0.061	0.580±0.104	0.531±0.079	0.217±0.003	0.615±0.004	0.317±0.002
AN-LS	0.574±0.037	0.631±0.072	0.538±0.044	0.230±0.002	0.587±0.006	0.261±0.006
WAN	0.576±0.041	0.661±0.033	0.698±0.017	0.241±0.002	0.621±0.004	0.315±0.000
EPR	0.539±0.028	0.597±0.062	0.710±0.008	0.214±0.001	0.628±0.003	0.314±0.000
ROLE	0.606±0.041	0.700±0.040	0.711±0.013	0.203±0.003	0.516±0.027	0.130±0.003
EM	0.486±0.031	0.549±0.103	0.642±0.029	0.294±0.002	0.614±0.003	0.293±0.001
EM-APL	0.467±0.026	0.448±0.049	0.654±0.040	0.275±0.003	0.589±0.007	0.311±0.001
SMILE	0.670±0.021	0.722±0.071	0.751±0.004	0.295±0.004	**0.629±0.003**	0.318±0.001
LL-R	0.605±0.058	0.714±0.035	0.658±0.006	0.268±0.002	0.625±0.001	0.296±0.004
LL-CP	0.595±0.031	0.735±0.028	0.700±0.000	0.259±0.004	0.621±0.007	0.251±0.007
LL-CT	0.600±0.012	0.669±0.052	0.629±0.007	0.258±0.004	0.619±0.004	0.253±0.004
CRISP	**0.749±0.037**	**0.795±0.031**	**0.758±0.002**	**0.304±0.003**	0.628±0.003	**0.319±0.001**

Table 10: Predictive performance of each comparing methods on MLL datasets in terms of *Hamming loss* (mean ± std). The best performance is highlighted in bold (the smaller the better).

	Image	Scene	Yeast	Corel5k	Mirflickr	Delicious
AN	0.229±0.000	0.176±0.001	0.306±0.000	**0.010±0.000**	0.127±0.000	**0.019±0.000**
AN-LS	0.229±0.000	0.168±0.004	0.306±0.000	**0.010±0.000**	0.127±0.000	**0.019±0.000**
WAN	0.411±0.060	0.299±0.035	0.285±0.016	0.156±0.001	0.191±0.006	0.102±0.000
EPR	0.370±0.043	0.220±0.026	0.234±0.007	0.016±0.000	0.136±0.002	0.020±0.000
ROLE	0.256±0.018	0.176±0.017	0.279±0.010	**0.010±0.000**	0.128±0.000	**0.019±0.000**
EM	0.770±0.001	0.820±0.003	0.669±0.025	0.589±0.003	0.718±0.010	0.630±0.005
EM-APL	0.707±0.088	0.780±0.082	0.641±0.032	0.648±0.006	0.754±0.017	0.622±0.006
SMILE	0.219±0.009	0.182±0.021	**0.208±0.002**	**0.010±0.000**	0.127±0.001	0.081±0.008
LL-R	0.220±0.013	0.162±0.005	0.312±0.001	0.015±0.001	0.124±0.002	**0.019±0.000**
LL-CP	0.218±0.016	0.164±0.002	0.306±0.000	0.016±0.001	0.126±0.001	**0.019±0.000**
LL-CT	0.246±0.031	0.176±0.019	0.321±0.001	0.018±0.001	0.124±0.001	**0.019±0.000**
CRISP	**0.165±0.023**	**0.140±0.013**	0.211±0.001	**0.010±0.000**	**0.121±0.002**	**0.019±0.000**