Positive-Unlabeled Learning via Optimal Transport and Margin Distribution

Nan Cao, Teng Zhang, Xuanhua Shi and Hai Jin

"Positive-unlabeled (PU) learning deals with the circumstances where only a portion of positive instances are labeled, while the rest and all negative instances are unlabeled, and due to this confusion, the class prior can not be directly available. In this paper, we enhance PU learning methods from the above two aspects. We first explicitly learn a transformation from unlabeled data to positive data by ***entropy regularized optimal transport*** to achieve a much more precise estimation for class prior. Then we switch to optimizing the ***margin distribution***, rather than the minimum margin, to obtain a label noise insensitive classifier. "

Introduction

Positive-Unlabeled Learning: Dataset contains only positive and unlabel instances.
Optimal Transport: Finding the best transport plan T of two distribution p_s and p_t under given cost metric C.

$$min_{\boldsymbol{T}} \ \mathrm{tr}(\boldsymbol{C}^\top \boldsymbol{T}) - \eta \cdot \Omega(\boldsymbol{T}) \quad s.t. \ \boldsymbol{T}\boldsymbol{1} = \boldsymbol{p}_s, \ \boldsymbol{T}^\top \boldsymbol{1} = \boldsymbol{p}_t$$

$$\Omega(\boldsymbol{T}) = -\sum_{ij} T_{ij}(logT_{ij} - 1)$$

Class prior: The proportion of positive instances in unlabeled data.

$$\mathbb{P}(\boldsymbol{x};\pi) = \pi\mathbb{P}(\boldsymbol{x}|y=1) + (1-\pi)\mathbb{P}(\boldsymbol{x}|y=-1)$$

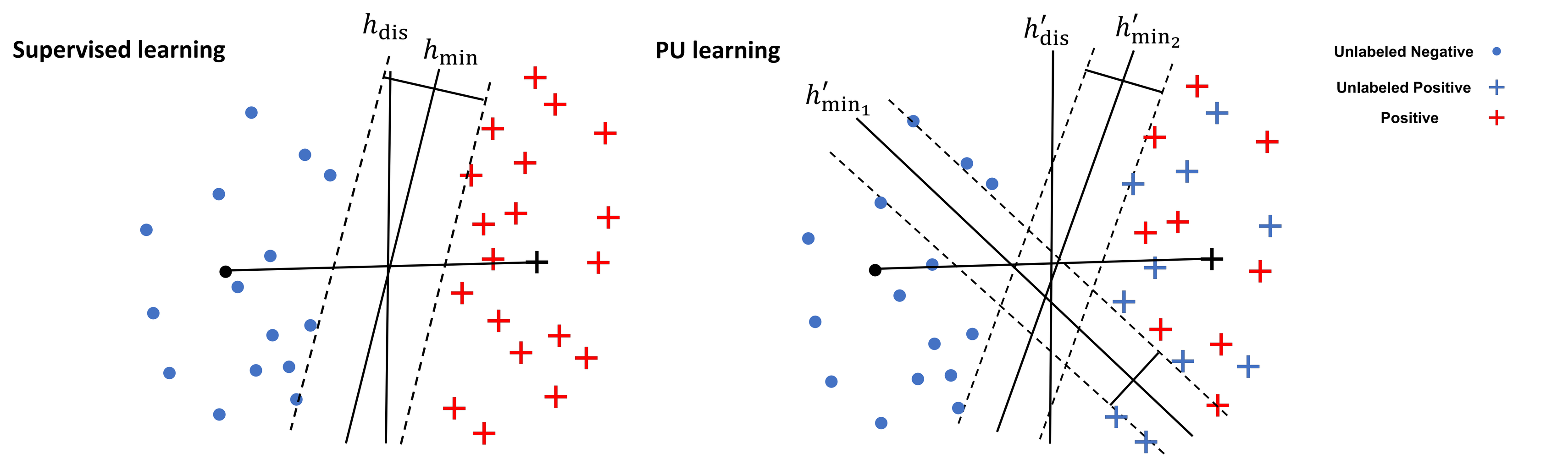
Optimal Margin Distribution Learning

Key difference

Large margin: Maximize the minimal margin (e.g. SVM)
Margin distribution: Optimize the margin distribution

Features

Avoid gengerting the multiple low-density decision boundraries, and margin distribution strategy is more robust to the label noise in PU tasks.



Proposed Method

Class prior estimation

(1) The underlying positive instances of unlabeled set and the given positive instances share the same distribution.
(2) Find the possible positive and negative instances via optimal transport plan, and estimate the class prior according to its definition.

Set threshold $\sigma = min\{1, 10/p\}$, if $max_j T_{ij} \geq \sigma$, then $\boldsymbol{x}_i$ will be treated as candidate positive.

Estimate class prior $\hat{\pi}$ via candidate positive set $\boldsymbol{C}_p$ and candidate negative set $\boldsymbol{C}_n$

$$min_{\hat{\pi}} \left\| \frac{1}{u}\sum_{i\in\mathcal{U}} \boldsymbol{x}_i - \frac{\hat{\pi}}{|\mathcal{C}|}\sum_{j\in\mathcal{C}_p} \boldsymbol{x}_j - \frac{1-\hat{\pi}}{|\mathcal{C}_n|}\sum_{k\in\mathcal{C}_n} \boldsymbol{x}_k \right\|^2$$

Training with margin distribution

(1) Build kNN-based graph $\boldsymbol{G}$ and Utilize adjacency matrix $\boldsymbol{A}$ to exploit relation between feature space and label space.
(2) Optimize classifier

$$min_{\boldsymbol{w},\boldsymbol{\xi},\boldsymbol{\epsilon},\boldsymbol{y}} \ \frac{\|\boldsymbol{w}\|^2}{2} + \sum_{i\in[m]} \alpha_i \frac{\xi_i^2 + \mu\epsilon_i^2}{m(1-\theta)^2} + \tau \sum_{i,j\in[m]} \bar{A}_{ij}(y_i - y_j)^2$$

$$s.t. \ 1-\theta-\xi_i \leq y_i \boldsymbol{w}^\top \phi(\boldsymbol{x}_i) \leq 1+\theta+\epsilon_i, \ i \in \mathcal{P}$$

$$1-\theta-\xi_j \leq |y_j \boldsymbol{w}^\top \phi(\boldsymbol{x}_j)| \leq 1+\theta+\epsilon_j$$

$$\sum_j \mathbb{I}(y_j = 1) = u\hat{\pi}, \ j \in \mathcal{U}$$

Experiments

Data set	π	EN	PE	CAPU	PUOTMD	Data set	π	EN	PE	CAPU	PUOTMD
Australian	0.3	.459±.024 .787±.029●	.388±.018 .807±.024●	**.352±.015** .827±.017●	.355±.013 **.838±.019**	Spambase	0.3	.558±.027 .739±.029●	.406±.017 .789±.016○	**.383±.015** **.872±.013**	.388±.023 .871±.019
	0.5	.648±.027 .755±.031●	.599±.024 .801±.021●	.588±.014 .807±.016●	**.574±.016** **.819±.021**		0.5	.679±.022 .766±.028●	.586±.018 .765±.021●	.565±.013 .838±.025●	**.513±.017** **.897±.015**
	0.7	.781±.018 .708±.019●	.761±.023 .724±.028●	.739±.015 .758±.021●	**.715±.019** **.761±.022**		0.7	.815±.019 .711±.026●	.769±.016 .757±.023●	.785±.014 .825±.024●	**.684±.032** **.863±.024**
Diabetes	0.3	.589±.032 .507±.029●	.438±.027 .634±.022●	.331±.013 .739±.018●	**.325±.014** **.755±.019**	Musk	0.3	.449±.029 .871±.032●	**.337±.024** .890±.018●	.348±.013 .915±.019	.258±.018 **.924±.019**
	0.5	.735±.021 .622±.024●	.613±.012 .642±.022●	.609±.026 .702±.017●	**.567±.027** **.722±.024**		0.5	.627±.025 .851±.027●	.574±.021 .876±.028●	.553±.021 .873±.023●	**.521±.013** **.915±.029**
	0.7	.894±.014 .624±.025●	.759±.017 .657±.023●	.767±.023 .672±.019●	**.754±.014** **.701±.026**		0.7	.798±.023 .821±.019●	**.753±.017** .842±.022●	.779±.019 .838±.024●	.769±.024 **.864±.031**
Banknote	0.3	.503±.036 .881±.037●	.413±.024 .928±.021●	.358±.028 **.949±.019**	**.315±.018** .943±.016	Mushroom	0.3	.397±.014 .918±.023●	.331±.021 .921±.019●	**.294±.019** .928±.015	.281±.019 **.937±.021**
	0.5	.688±.031 .823±.029●	.597±.027 .879±.027●	.552±.017 .902±.029●	**.479±.027** **.934±.019**		0.5	.593±.023 .887±.019●	.544±.014 .902±.023●	**.531±.018** **.912±.026**	.442±.017 .901±.018
	0.7	.869±.034 .837±.029●	.752±.016 .893±.021	.776±.017 .889±.024●	**.674±.027** **.897±.019**		0.7	.819±.028 .869±.027●	.753±.019 .883±.022●	.749±.013 .901±.021	**.732±.025** **.903±.019**
Kr-vs-kp	0.3	.505±.034 .783±.029●	.389±.021 .816±.017●	.369±.019 .821±.023●	**.351±.015** **.841±.019**	House	0.3	.491±.019 .907±.013●	**.337±.022** **.923±.017**	.387±.027 .911±.016●	.358±.017 .911±.013
	0.5	.617±.027 .734±.034●	.576±.022 .772±.017●	.579±.019 .789±.024●	**.543±.021** **.801±.026**		0.5	.632±.038 .838±.024●	.597±.017 .849±.025●	.562±.019 .870±.022●	**.538±.014** **.887±.023**
	0.7	.834±.019 .718±.031●	.786±.019 .744±.017●	**.754±.021** **.759±.018**	.779±.028 .738±.026		0.7	.869±.028 .801±.027●	.798±.019 .838±.023●	**.742±.013** .849±.018●	.776±.024 **.863±.020**

Data set	π	WLR	PULD	UPU	nnPU	CAPU	PUSB	EN	LDCE	PUOTMD
Australian	.3	.773±.022●	.826±.015●	.841±.009	.822±.009●	**.854±.013**	.795±.007●	.824±.009●	.811±.017●	.843±.004
	.5	.731±.021●	.790±.022●	.815±.016●	.816±.015	.811±.021	.779±.009●	.818±.007●	.779±.022 ●	**.829±.007**
	.7	.677±.027●	.746±.017●	.769±.015●	.781±.017●	.802±.025●	.721±.015●	.775±.011●	.732±.027●	**.812±.007**
Diabetes	.3	.705±.019●	.741±.019●	.719±.013●	.707±.007●	**.775±.008**○	.743±.013●	.742±.013●	.732±.018●	.763±.009
	.5	.679±.022●	.722±.011●	.681±.019●	.689±.013●	.743±.012	.718±.017●	.683±.017●	.709±.016●	**.752±.011**
	.7	.631±.027●	.708±.023●	.643±.018●	.677±.009●	.711±.015●	.689±.028●	.639±.017●	.652±.022●	**.724±.013**
Banknote	.3	.952±.012●	.959±.013●	.955±.009●	.969±.011	.965±.027●	.951±.007●	.964±.009●	.966±.017	**.971±.009**
	.5	.924±.015●	.945±.006	.931±.014●	.940±.007●	.938±.017●	.927±.016●	.933±.013●	.937±.008●	**.953±.007**
	.7	.891±.017●	.908±.005●	.897±.013●	.906±.005●	.903±.031●	.891±.015●	.909±.017●	.902±.013●	**.929±.016**
Kr-vs-kp	.3	.813±.018●	.849±.015	.832±.014●	.824±.012●	.847±.022	.827±.019●	.837±.014●	.822±.017●	**.856±.013**
	.5	.796±.021●	**.826±.012**	.811±.011●	.803±.015●	.819±.017	.813±.016●	.811±.019●	.803±.015●	.824±.011
	.7	.778±.020●	.801±.009●	.781±.018●	.783±.014●	.798±.016●	.783±.024●	.789±.021●	.782±.022●	**.809±.016**
Spambase	.3	.879±.024●	.902±.011●	.889±.017●	.873±.016●	.906±.004●	.876±.007●	.821±.017●	.891±.017	**.912±.007**
	.5	.841±.029●	.887±.016	.807±.029●	.828±.012●	.883±.009●	.852±.009●	.801±.021●	.853±.019●	**.901±.005**
	.7	.802±.023●	.872±.010●	.784±.027●	.809±.015●	.841±.011●	.817±.007●	.772±.029●	.831±.027●	**.873±.012**
Musk	.3	.938±.014●	.938±.009	.925±.009●	.932±.007●	.922±.008●	.931±.014●	.933±.011●	.901±.011●	**.947±.009**
	.5	.921±.017	.911±.014●	.907±.013●	.901±.011●	.899±.014●	.914±.013●	.901±.008●	.874±.023●	**.929±.005**
	.7	.877±.016●	.881±.012●	.878±.018●	.872±.015	.878±.012●	.871±.021●	.883±.011●	.841±.017●	**.891±.011**
Mushroom	.3	.924±.013●	.952±.005	.923±.011●	.945±.009	.952±.015	.938±.014●	.947±.017●	.934±.013●	**.958±.005**
	.5	.901±.011●	.939±.012●	.911±.012●	.921±.007●	**.941±.009**	.925±.011	.931±.017	.917±.014●	.933±.003
	.7	.889±.019●	.923±.016●	.883±.014●	.902±.007●	.917±.013●	.912±.011●	.909±.013●	.891±.011●	**.922±.007**
House	.3	.917±.019●	.941±.009	.915±.011●	.908±.021●	.935±.028	.948±.015	.932±.007●	.922±.009●	**.958±.013**
	.5	.882±.015●	**.933±.014**	.873±.017●	.881±.018●	.898±.026●	.917±.009●	.909±.011●	.875±.021●	.929±.012
	.7	.841±.023●	.883±.013●	.822±.021●	.839±.024●	.906±.024●	.885±.013●	.897±.012	.831±.017●	**.908±.016**
w/t/l	.3	8/0/0	4/4/0	7/1/0	6/2/0	3/4/1	7/1/0	8/0/0	6/2/0	
	.5	7/1/0	4/4/0	8/0/0	7/1/0	4/4/0	7/1/0	7/1/0	8/0/0	
	.7	8/0/0	8/0/0	8/0/0	7/1/0	8/0/0	8/0/0	7/1/0	8/0/0	

Pseudo-Code

Input: PU dataset $\boldsymbol{S}$, hyperparameter η, threshold σ.
Solving entropy regularized optimal transport problem
Findind candidate positive and negative instances via $max_j T_{ij} \geq \sigma$
Estimate class prior $\hat{\pi}$
Output: $\hat{\pi}$ and candidate labels y_u^0

Input: PU data set $\boldsymbol{S}$, kNN-based graph $\boldsymbol{G}$, hyperparameter $\mu, \theta, \alpha, \lambda, \tau$, and estimated class prior $\hat{\pi}$, maximum iteration number T
Initialize: unlabeled instances $y_u = y_u^0$
Applying variable splitting technique and auxiliary variable q to reform the problem
For $t < T$:
Optimizing $\boldsymbol{w}$ by fixing ξ, ϵ, y and q;
Optimizing ξ, ϵ and q by fixing $\boldsymbol{w}$ and q;
Optimizing q by fixing $\boldsymbol{w}, \xi, \epsilon$ and y
Output: $\boldsymbol{w}$

Conclusion

(1) Achieve more precise estimation of class prior via optimal transport;
(2) Utilize margin distribution to alleviate the inevitable label noise in PU learning problems;
(3) Achieve better generalization performance on real-world data sets.

Reference

[1] Charles Elkan and Keith Noto. Learning classifiers from only positive and unlabeled data. Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 213–220, 2008.
[2] Teng Zhang and Zhi-Hua Zhou. Optimal margin distribution machine. IEEE Transactions on Knowledge and Data Engineering, 32(6):1143–1156, 2019.

Nan Cao

National Engineering Research Center for Big Data Technology and System, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, 430074, China