Semi-Supervised Partial Label Learning via Confidence-Rated Margin Maximization

Wei Wang **Min-Ling Zhang**

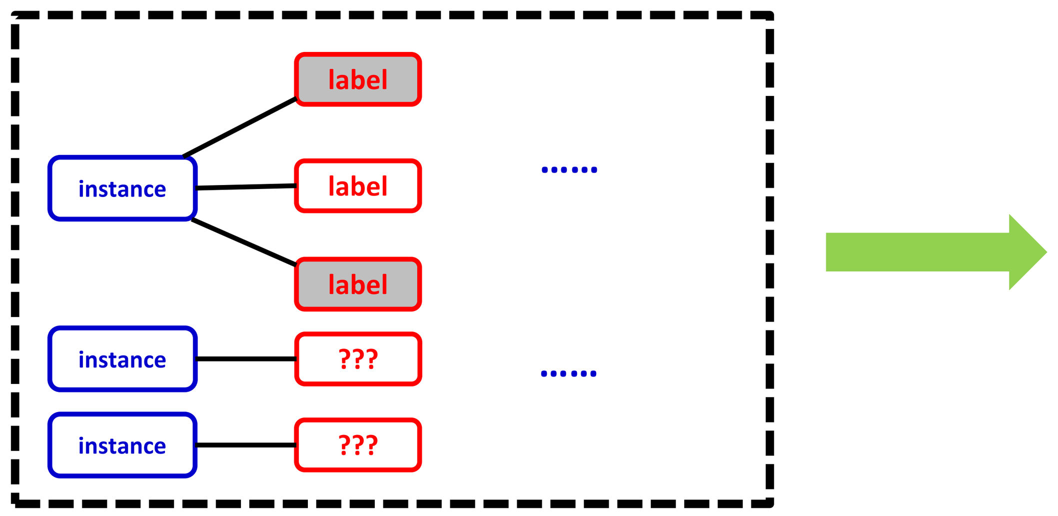
School of Computer Science and Engineering, Southeast University, Nanjing 210096, China

Email: {wang_w, zhangml}@seu.edu.cn

Semi-Supervised Partial Label Learning

Partial label (PL) learning is an emerging weakly-supervised learning framework where each training example is associated with multiple candidate labels among which only one is valid. However, the process of acquiring training examples with candidate labels might be demanding while abundant unlabeled data are readily available to facilitate model training.

The task of semi-supervised partial label learning: learn a multi-class classifier from the partial label training examples as well as unlabeled examples



multi-class classifier

$f: X \to Y$

$Y = \{y_1, y_2, \cdots, y_q\}$

One recent work on semi-supervised PL learning (SSPL) [Wang et al, IJCAI'19]:

- Graph-based method: an extra kNN procedure is utilized to enable generalization on unseen instances;
- Less efficient in terms of storage overhead and prediction time

The PARM Approach

Training set is $D_P = \{(\boldsymbol{x}_i, S_i) \mid 1 \le i \le p\}$ and $D_U = \{\boldsymbol{x}_i \mid p+1 \le i \le p+u\}$.

Let $\boldsymbol{f}_i = [f_{i1}, f_{i2}, \cdots, f_{iq}]^{\mathrm{T}}$ denote the labeling confidence and we have the labeling confidence matrix $\mathbf{F}_P = [\boldsymbol{f}_1, \boldsymbol{f}_2, \cdots, \boldsymbol{f}_p]^{\mathrm{T}} \in [0,1]^{p\times q}$ for PL training examples and $\mathbf{F}_U = [\boldsymbol{f}_{p+1}, \boldsymbol{f}_{p+2}, \cdots, \boldsymbol{f}_{p+u}]^{\mathrm{T}} \in [0,1]^{u\times q}$ for unlabeled data.

Phase1: Utilize label propagation to disambiguate PL training examples

- Construct the normalized similarity matrix $\mathbf{H}$ over PL training examples;
- Invoke iterative label propagation procedure until convergence to instantiate the labeling confidence of PL examples.

$$\widetilde{\mathbf{F}}_P^{(t)} = \alpha \cdot \mathbf{H}\mathbf{F}_P^{(t-1)} + (1-\alpha)\cdot \mathbf{F}_P^{(t-1)}$$

$$\forall 1 \le i \le p: \quad f_{il}^{(t)} = \begin{cases} \dfrac{\widetilde{f}_{il}^{(t)}}{\sum_{y_{l'}\in S_i} \widetilde{f}_{il'}^{(t)}}, & l \in S_i \\ 0, & \text{otherwise} \end{cases}$$

Phase2: confidence-rated margin maximization: jointly enable the induction of predictive model and the estimation of labeling confidence of unlabeled data.

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\Xi},\mathbf{F}_U} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \frac{\lambda}{p}\sum_{i=1}^{p}\sum_{l=1}^{q} f_{il}\xi_{il} + \frac{\mu}{u}\sum_{i=p+1}^{p+u}\sum_{l=1}^{q} f_{il}\xi_{il} + \gamma\sum_{i=1}^{u}\sum_{j=1}^{p} s_{ij}\|\boldsymbol{f}_{p+i} - \boldsymbol{f}_j\|_2^2 \\ \text{s.t.} \quad & \boldsymbol{w}^{\mathrm{T}}\Phi(\boldsymbol{x}_i, y_l) - \max_{y_{l'}\neq y_l} \boldsymbol{w}^{\mathrm{T}}\Phi(\boldsymbol{x}_i, y_{l'}) \ge 1 - \xi_{il}, \quad (1 \le i \le p+u,\ 1 \le l \le q) \\ & \xi_{il} \ge 0, \quad (1 \le i \le p+u,\ 1 \le l \le q) \\ & f_{il} \ge 0, \quad (p+1 \le i \le p+u,\ 1 \le l \le q) \\ & \sum_{l=1}^{q} f_{il} = 1, \quad (p+1 \le i \le p+u) \end{aligned}$$

To solve the derived problem, PARM employs alternating optimization.

- Fix $\boldsymbol{w}$, optimize $\mathbf{F}_U$

$$\begin{aligned} \min_{\mathbf{F}_U} \quad & \frac{\mu}{u}\sum_{i=p+1}^{p+u}\sum_{l=1}^{q} f_{il}\xi_{il} + \gamma\sum_{i=1}^{u}\sum_{j=1}^{p} s_{ij}\|\boldsymbol{f}_{p+i} - \boldsymbol{f}_j\|_2^2 \\ \text{s.t.} \quad & f_{il} \ge 0, \quad (p+1 \le i \le p+u,\ 1 \le l \le q) \\ & \sum_{l=1}^{q} f_{il} = 1, \quad (p+1 \le i \le p+u) \end{aligned}$$

➡ QP problem with uq variables and $u(q+1)$ constraints

- Fix $\mathbf{F}_U$, optimize $\boldsymbol{w}$

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{\Xi}} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \frac{\lambda}{p}\sum_{i=1}^{p}\sum_{l=1}^{q} f_{il}\xi_{il} + \frac{\mu}{u}\sum_{i=p+1}^{p+u}\sum_{l=1}^{q} f_{il}\xi_{il} \\ \text{s.t.} \quad & \boldsymbol{w}^{\mathrm{T}}\Phi(\boldsymbol{x}_i, y_l) - \max_{y_{l'}\neq y_l} \boldsymbol{w}^{\mathrm{T}}\Phi(\boldsymbol{x}_i, y_{l'}) \ge 1 - \xi_{il}, \quad (1 \le i \le p+u,\ 1 \le l \le q) \\ & \xi_{il} \ge 0, \quad (1 \le i \le p+u,\ 1 \le l \le q) \end{aligned}$$

➡ solve the dual problem

Experiments

Data Sets

Controlled UCI Data Sets

Data Set	# Examples	# Features	# Class Labels	# False Positive Labels (r)
Deter	358	23	6	$r = 1, 2, 3$
Vehicle	846	18	4	$r = 1, 2$
Abalone	4,177	7	29	$r = 1, 2, 3$
Satimage	6,435	36	7	$r = 1, 2, 3$

Real-World Data Sets

Data Set	# Examples	# Features	# Class Labels	Avg. # CLs	Task Domain
Lost	1,122	108	16	2.23	*automatic face naming*
Mirflickr	2,780	1536	14	2.76	*web image classification*
BirdSong	4,998	38	13	2.18	*bird song classification*
LYN10	16,526	163	10	1.84	*automatic face naming*
LYN20	17,511	163	20	1.85	*automatic face naming*

Proportion ρ of training examples are sampled to form partial label training set and the rest training examples are used to form unlabeled data set.

Experimental Setup

Comparing approaches:

semi-supervised PL learning approach: SSPL

PL learning approach: PL-kNN, CLPL, PL-SVM, PL-AGGD

Experimental protocol:

ten-fold cross-validation + pairwise t-test

Experimental Results

Win/tie/loss counts of pairwise t-test (0.05 significance level)

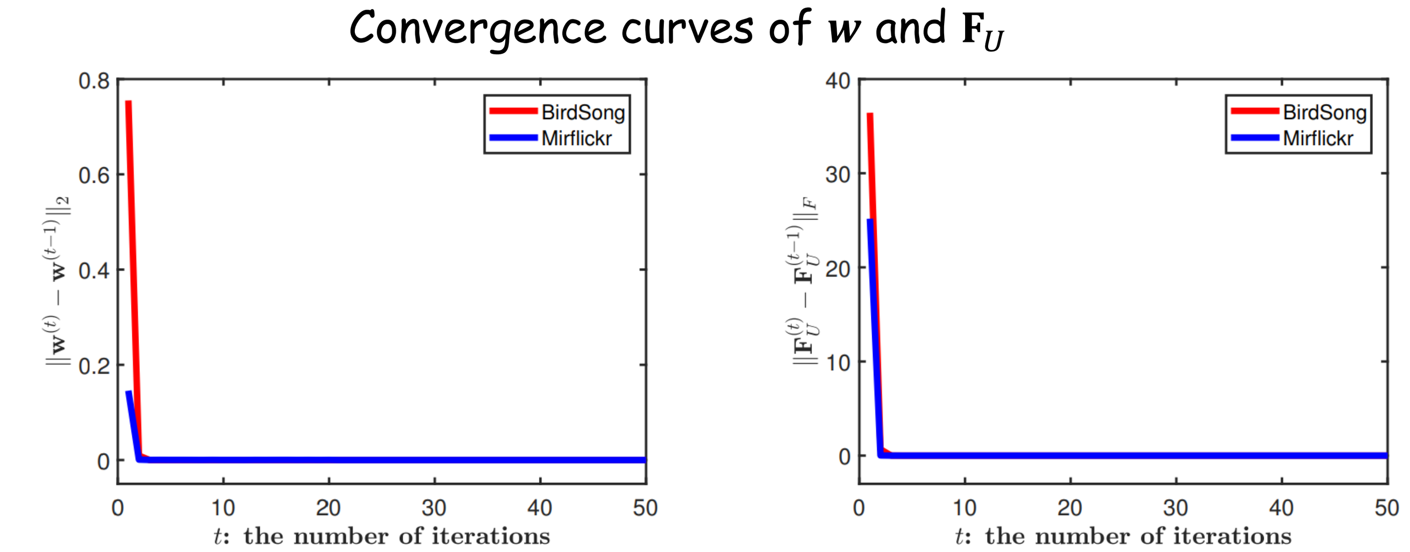
	PARM against				
	SSPL	PL-KNN	CLPL	PL-SVM	PL-AGGD
Controlled UCI data sets ($r = 1$)	14/21/1	26/9/1	11/16/9	20/14/2	10/19/7
Controlled UCI data sets ($r = 2$)	17/15/4	23/13/0	10/17/9	19/14/3	9/22/5
Controlled UCI data sets ($r = 3$)	18/7/2	23/4/0	12/14/1	18/7/2	9/16/2
Real-world data sets	20/21/4	45/0/0	31/14/0	34/11/0	24/16/5
In Total	**69/64/11**	**117/26/1**	**64/61/19**	**91/46/7**	**52/73/19**

Out of 144 statistical tests:

- PARM significantly outperforms PL-KNN and PL-SVM in 81.3% and 63.2% cases
- PARM achieves superior or at least comparable performance to SSPL, CLPL and PL-AGGD in 92.4%, 86.8% and 86.8% cases

Convergence Analysis

Convergence curves of $\boldsymbol{w}$ and $\mathbf{F}_U$



The classification model and labeling confidence of unlabeled data converge fast with increasing number of iterations.

Conclusion

- Main Contribution

  Propose a novel inductive approach for semi-supervised partial label learning
- Key Techniques

  Confidence-rated margin maximization;

  alternating optimization
- Future Works

  Investigate ways of enabling the proposed approach to deal with large-scale data sets

More Information: http://palm.seu.edu.cn/