Exploiting Unlabeled Data via Partial Label Assignment for Multi-Class Semi-Supervised Learning

Zhen-Ru Zhang[1,2,3], **Qian-Wen Zhang**[3], **Yunbo Cao**[3], **Min-Ling Zhang**[1,2,4*]

[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
[3] Tencent Cloud Xiaowei, Beijing, China
[4] Collaborative Innovation Center of Wireless Communications Technology, China
zhangzr@seu.edu.cn, {cowenzhang, yunbocao}@tencent.com, zhangml@seu.edu.cn

Abstract

In semi-supervised learning, one key strategy in exploiting unlabeled data is trying to estimate its pseudo-label based on current predictive model, where the unlabeled data assigned with pseudo-label is further utilized to enlarge labeled data set for model update. Nonetheless, the supervision information conveyed by pseudo-label is prone to error especially when the performance of initial predictive model is mediocre due to limited amount of labeled data. In this paper, an intermediate unlabeled data exploitation strategy is investigated via partial label assignment, i.e. a set of candidate labels other than a single pseudo-label are assigned to the unlabeled data. We only assume that the ground-truth label of unlabeled data resides in the assigned candidate label set, which is less error-prone than trying to identify the single ground-truth label via pseudo-labeling. Specifically, a multi-class classifier is induced from the partial label examples with candidate labels to facilitate model induction with labeled examples. An iterative procedure is designed to enable labeling information communication between the classifiers induced from partial label examples and labeled examples, whose classification outputs are integrated to yield the final prediction. Comparative studies against state-of-the-art approaches clearly show the effectiveness of the proposed unlabeled data exploitation strategy for multi-class semi-supervised learning.

Introduction

In semi-supervised learning, the learning system aims to exploit unlabeled data to facilitate predictive model induction with limited labeled examples. Most semi-supervised learning techniques work by trying to estimate the pseudo-label of unlabeled data based on current predictive model, which is assigned to the unlabeled data as the ground-truth label and then employed to enrich the labeled data set for model update, such as co-training (Blum and Mitchell 1998; Zhang and Zhou 2011; Ma et al. 2020), label propagation (Zhou et al. 2004; Chong et al. 2020), semi-supervised SVM (Joachims 1999; Chapelle, Sindhwani, and Keerthi 2008; Li, Kwok, and Zhou 2016), etc. However, one potential issue lies in that the estimated pseudo-label is prone to error, especially when the predictive model has only mediocre generalization performance with few labeled examples available in initial training iterations (Zhu and Goldberg 2009; van Engelen and Hoos 2020).

Partial label learning is an emerging weakly supervised learning framework dealing with inaccurate supervision (Nguyen and Caruana 2008; Cour, Sapp, and Taskar 2011; Zhang and Yu 2015; Ren et al. 2016; Zhou 2017; Wang, Li, and Zhang 2019; Lv et al. 2020; Lyu et al. 2019), where each training example is associated with a set of candidate labels among which only one is valid. The task of partial label learning is to induce a multi-class classification model from partial label training examples, where the ground-truth label of each training example is assumed to reside in its candidate label set but not directly accessible to the training algorithm. In view of supervision spectrum, the weak supervision conveyed by partial label example lies between the blind supervision of unlabeled example and the full supervision of labeled example. Therefore, it is natural to leverage partial label example as an intermediate means to facilitate the exploitation of unlabeled data for semi-supervised learning. Conceptually, other than trying to identify the strong supervision information for unlabeled data with single ground-truth label, it is relatively easier to estimate the weak supervision information for unlabeled data with a set of candidate labels consisting of the ground-truth label.

In light of the above observations, a novel strategy for unlabeled data exploitation is investigated in this paper. Accordingly, a simple yet effective multi-class semi-supervised learning approach named EUPAL, i.e. *Exploiting Unlabeled data via PArtial Label assignment*, is proposed. Briefly, EUPAL initializes partial label assignment over unlabeled data by resorting to weighted kNN aggregation. Then, classification models induced from partial label examples and labeled examples are iteratively updated by conducting labeling information communication with random sampling. Experimental results show that the proposed approach serves as an effective way towards unlabeled data exploitation, especially for the case of lower fraction of labeled examples in training set.

The rest of this paper is organized as follows. Firstly, technical details of the proposed EUPAL approach are presented. Secondly, experimental results of comparative studies are reported. Thirdly, related works on semi-supervised and partial label learning are briefly discussed. Finally, we conclude this paper.

*Corresponding author

Inputs:

$\mathcal{D}_l$: the labeled data set $\{(\boldsymbol{x}_i, y_i) \mid 1 \le i \le L\}$ ($\boldsymbol{x}_i \in \mathcal{X}$, $y_i \in \mathcal{Y}$, $\mathcal{X} = \mathbb{R}^d$, $\mathcal{Y} = \{l_1, l_2, \ldots, l_q\}$)
$\mathcal{D}_u$: the unlabeled data set $\{\boldsymbol{x}_j \mid 1 \le j \le U\}$
k: the number of nearest neighbors used for kNN aggregation
α: the balancing parameter with $\alpha \in [0, 1]$
T: the maximum number of iterations
$\mathcal{L}, \mathcal{P}$: the supervised training algorithm $\mathcal{L}$ and the partial label training algorithm $\mathcal{P}$
$\boldsymbol{x}^*$: the unseen instance

Outputs:

y^*: the predicted label for $\boldsymbol{x}^*$

Process:

Identify the k nearest neighbors of $\boldsymbol{x}_j$ in $\mathcal{D}_l$ whose index set is stored in $\mathcal{N}(\boldsymbol{x}_j)$ $(1 \le j \le U)$;
Set the weight matrix $\mathbf{W} = [w_{ij}]_{L \times U}$ whose elements are set according to Eq.(1);
Generate partial label example $(\boldsymbol{x}_j, S_j)$ $(1 \le j \le U)$ by specifying the candidate label set S_j via kNN aggregation according to Eqs.(2)-(3) ;
Set $\mathcal{D}_p = \{(\boldsymbol{x}_j, S_j) \mid 1 \le j \le U\}$;
Set $t = 1$;
repeat
    Induce $f_l^{(t)}$ by invoking supervised training algorithm $\mathcal{L}$ over $\mathcal{D}_l$, i.e. $f_l^{(t)} \hookleftarrow \mathcal{L}(\mathcal{D}_l)$;
    Induce $f_p^{(t)}$ by invoking partial label training algorithm $\mathcal{P}$ over $\mathcal{D}_p$, i.e. $f_p^{(t)} \hookleftarrow \mathcal{P}(\mathcal{D}_p)$;
    Randomly sample a subset of examples $\Delta_1^{(t)} \subseteq \mathcal{D}_u$ from $\mathcal{D}_u$;
    For any $\boldsymbol{x} \in \Delta_1^{(t)}$ and the corresponding partial label example $(\boldsymbol{x}, S) \in \mathcal{D}_p$, update its candidate label set S to S' according to Eq.(4);
    Randomly sample a subset of examples $\Delta_2^{(t)} \subseteq \mathcal{D}_u$ from $\mathcal{D}_u$;
    For any $\boldsymbol{x} \in \Delta_2^{(t)}$, enrich the labeled data set by adding $(\boldsymbol{x}, y')$ to $\mathcal{D}_l$ with y' being predicted according to Eq.(5);
    $t = t + 1$.
until convergence
Let f_l^* and f_p^* be the classification models returned by the iterative updating procedure;
Return y^* according to Eq.(6).

Table 1: The pseudo-code of EUPAL.

The Proposed Approach

Let $\mathcal{X} = \mathbb{R}^d$ be the d-dimensional feature space and $\mathcal{Y} = \{l_1, l_2, \ldots, l_q\}$ be the label space consisting of q class labels. Furthermore, let $\mathcal{D}_l = \{(\boldsymbol{x}_i, y_i) \mid 1 \le i \le L\}$ denote the set of labeled examples with $y_i \in \mathcal{Y}$ being the ground-truth label of $\boldsymbol{x}_i \in \mathcal{X}$, and $\mathcal{D}_u = \{\boldsymbol{x}_j \mid 1 \le j \le U\}$ denote the set of unlabeled examples. In semi-supervised learning, it is generally assumed that $L \ll U$ and the task of the learning system is to induce a classification model from $\mathcal{D}_l \bigcup \mathcal{D}_u$.

To enable unlabeled data exploitation, EUPAL chooses to make use of partial label assignment which offers weak supervision information to the learning system to facilitate model induction. For each unlabeled example $\boldsymbol{x}_j \in \mathcal{D}_u$, let $\mathcal{N}(\boldsymbol{x}_j)$ be the index set of $\boldsymbol{x}_j$'s k nearest neighbors identified in $\mathcal{D}_l$. Furthermore, let $\mathbf{W} = [w_{ij}]_{L \times U}$ be the weight matrix whose elements are set as follows:

$$\forall\, 1 \le i \le L,\ 1 \le j \le U: \tag{1}$$
$$w_{ij} = \begin{cases} \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2}{2\sigma^2}\right) & , \text{ if } i \in \mathcal{N}(\boldsymbol{x}_j) \\ 0 & , \text{ otherwise} \end{cases}$$

Here, σ corresponds to the width parameter of the kernel distance function, which is fixed to be 1 in this paper. Then, a labeling confidence vector $\hat{\boldsymbol{r}}_j = [\hat{r}_{j1}, \hat{r}_{j2}, \ldots, \hat{r}_{jq}]^\top$ can be derived for each unlabeled example $\boldsymbol{x}_j$ $(1 \le j \le U)$ via kNN aggregation:

$$\forall\, 1 \le k \le q: \quad \hat{r}_{jk} = \sum\nolimits_{i \in \mathcal{N}(\boldsymbol{x}_j)} w_{ij} \cdot [\![y_i = l_k]\!] \tag{2}$$

Here, $[\![\pi]\!]$ returns 1 if predicate π holds and 0 otherwise. By normalizing $\hat{\boldsymbol{r}}_j$ into $\boldsymbol{r}_j$, EUPAL determines the set of candidate labels $S_j \subseteq \mathcal{Y}$ assigned to $\boldsymbol{x}_j$ as follows:

$$\forall\, 1 \le j \le U: \quad S_j = \{l_k \mid r_{jk} \ge 1/q,\ 1 \le k \le q\} \tag{3}$$
$$\text{where } r_{jk} = \frac{\hat{r}_{jk}}{\sum_{k=1}^{q} \hat{r}_{jk}}$$

Accordingly, the unlabeled data set $\mathcal{D}_u$ is transformed into the *partial label* data set $\mathcal{D}_p = \{(\boldsymbol{x}_j, S_j) \mid 1 \le j \le U\}$ which is further utilized for inducing the multi-class classification model in an iterative manner.

At the t-th iteration, let $f_l^{(t)} : \mathcal{X} \times \mathcal{Y} \mapsto [0, 1]$ be the classification model induced from labeled data set $\mathcal{D}_l$ by invoking pre-specified supervised training algorithm $\mathcal{L}$, i.e. $f_l^{(t)} \hookleftarrow \mathcal{L}(\mathcal{D}_l)$. Without loss of generality, we assume that $f_l^{(t)}$ yields probabilistic predictions for instance $\boldsymbol{x}$ with $\sum_{k=1}^{q} f_l^{(t)}(\boldsymbol{x}, l_k) = 1$. Correspondingly, let $f_p^{(t)} : \mathcal{X} \times \mathcal{Y} \mapsto [0, 1]$ be the classification model induced from partial label data set $\mathcal{D}_p$ by invoking pre-specified partial label training algorithm $\mathcal{P}$, i.e. $f_p^{(t)} \leftarrow \mathcal{P}(\mathcal{D}_p)$. Similarly, $f_p^{(t)}$ also yields probabilistic predictions with $\sum_{k=1}^{q} f_p^{(t)}(\boldsymbol{x}, l_k) = 1$.

Then, the labeling information between $f_l^{(t)}$ and $f_p^{(t)}$ is communicated for model update based on random sampling. Let $\Delta_1^{(t)} \subseteq \mathcal{D}_u$ be a subset of unlabeled examples randomly sampled from $\mathcal{D}_u$, then the labeling information conveyed by the predictive output of $f_l^{(t)}$ over $\Delta_1^{(t)}$ is utilized to update the partial label data set $\mathcal{D}_p$. Specifically, for any example $\boldsymbol{x} \in \Delta_1^{(t)}$, let $(\boldsymbol{x}, S) \in \mathcal{D}_p$ be the corresponding partial label example in $\mathcal{D}_p$. EUPAL chooses to update the candidate label set S into S' with the following rule:

$$S' = \{l_k \mid t_k \geq 1/q,\ 1 \leq k \leq q\} \quad \text{where} \qquad (4)$$
$$t_k = \alpha \cdot f_l^{(t)}(\boldsymbol{x}, l_k) + (1 - \alpha) \cdot f_p^{(t)}(\boldsymbol{x}, l_k)$$

Here, $\alpha \in [0, 1]$ is the trade-off parameter which controls the relative amount of labeling information inherited from $f_l^{(t)}$ and $f_p^{(t)}$ respectively.

On the other hand, let $\Delta_2^{(t)} \subseteq \mathcal{D}_u$ be another subset of unlabeled examples randomly sampled from $\mathcal{D}_u$. Then, the labeling information conveyed by the predictive output of $f_p^{(t)}$ over $\Delta_2^{(t)}$ is utilized to update the labeled data set $\mathcal{D}_l$. Specifically, for any example $\boldsymbol{x} \in \Delta_2^{(t)}$, its class label y' predicted by $f_p^{(t)}$ corresponds to:

$$y' = \arg\max_{l_k \in \mathcal{Y}}\ f_p^{(t)}(\boldsymbol{x}, l_k) \qquad (5)$$

Then, EUPAL enriches $\mathcal{D}_l$ by adding $(\boldsymbol{x}, y')$ to the labeled data set.[1] Accordingly, $f_l^{(t)}$ and $f_p^{(t)}$ are re-trained based on the updated labeled data set $\mathcal{D}_l$ and partial label data set $\mathcal{D}_p$ respectively. The iterative updating procedure terminates until both classification models do not change or the maximum number of iterations is reached.[2]

Let f_l^* and f_p^* be the final classification models returned by the iterative updating procedure, the prediction over unseen instance $\boldsymbol{x}^*$ is determined by combining the predictive outputs of both classifiers:

$$y^* = \arg\max_{l_k \in \mathcal{Y}}\ \mu_l \cdot f_l^*(\boldsymbol{x}^*, l_k) + \mu_p \cdot f_p^*(\boldsymbol{x}^*, l_k) \qquad (6)$$

Here, μ_l and μ_p corresponds to the empirical predictive accuracy of f_l^* and f_p^* over the original labeled data set (i.e. $\{(\boldsymbol{x}_i, y_i) \mid 1 \leq i \leq L\}$) respectively.

Table 1 summarizes the complete procedure of EUPAL. Firstly, the partial label assignment over unlabeled data is initialized by employing kNN aggregation (Steps 1-4). Then, an iterative updating procedure is utilized to update the classification models induced from labeled examples and partial label examples via labeling information communication (Steps 5-14). Finally, the prediction on unseen instance is determined by combining the modeling outputs of both classifiers (Steps 15-16).

[1] In case $\boldsymbol{x}$ has been in $\mathcal{D}_l$, its class label is updated to y'.

[2] In this paper, both $\Delta_1^{(t)}$ and $\Delta_2^{(t)}$ are generated by sampling with replacement. Furthermore, the size of the random samples is set to be $\max(U/T, 3)$ where T is the maximum number of iterations.

Data Set	#Examples	#Features	#Labels
Ecoli	307	7	4
Deter	338	23	5
BHP	1,060	21	4
Yeast	1,299	8	4
Wireless	2,000	7	4
Segment	2,310	18	7
Character	3,140	100	8
Location	3,505	100	8
Work	5,628	100	11
Usps	9,298	256	10
Pen	10,992	16	10
Letter	20,000	16	26
Sensorless	58,509	48	11

Table 2: Characteristics of experimental data sets.

Experiments

Experimental Setup

Data Sets In this paper, a total of 13 benchmark multi-class data sets (Dua and Graff 2017) have been employed for experimental studies whose characteristics are summarized in Table 2.

For each data set, 25% examples are randomly sampled to form the test set and the remaining examples are used to form the training set. Among the training set, the labeled data set $\mathcal{D}_l$ consists of p fraction of examples and the other examples are treated as unlabeled examples in $\mathcal{D}_u$. To account for the factor that in semi-supervised learning the number of labeled examples is much smaller that that of unlabeled examples, we consider three configurations of p in this paper, i.e. $p \in \{0.5\%, 1\%, 5\%\}$. Given the fraction of labeled examples, the training set is randomly divided into $\mathcal{D}_l$ and $\mathcal{D}_u$ for ten times and the average predictive performance on test set (mean accuracy$\pm$std. deviation) is recorded for each comparing approach.

Comparing Approaches The performance of EUPAL is compared against four state-of-the-art semi-supervised learning approaches, each configured with parameters suggested in the literature:

- PLANETOID (Yang, Cohen, and Salakhutdinov 2016): A graph-based semi-supervised learning approach where an embedding for each instance is trained to jointly predict the class label and the neighborhood context in the graph. [suggested configuration: $q = 10, d = 3$]
- SSODM (Zhang and Zhou 2018): A margin-based semi-supervised learning approach which works by assigning labels to unlabeled instances to achieve optimal margin distribution. [suggested configuration: $\nu = 0.8, \theta = 0.95$]

Data Set	Comparing Approach						
	EUPAL	PLANETOID	SSODM	S4VM	COTRADE	IPAL	LIBSVM
Ecoli	**0.726±0.119**	0.634±0.099	0.626±0.179	0.495±0.126•	0.086±0.048•	0.652±0.199	0.048±0.053•
Deter	**0.799±0.090**	0.700±0.097•	0.621±0.116•	0.655±0.067•	0.051±0.061•	0.719±0.094	0.017±0.045•
BHP	**0.508±0.060**	0.463±0.064•	0.438±0.067•	0.389±0.081•	0.317±0.061•	0.477±0.030	0.185±0.137•
Yeast	**0.467±0.067**	0.466±0.062	0.443±0.074	0.357±0.074•	0.313±0.056•	0.396±0.105•	0.141±0.086•
Wireless	**0.940±0.027**	0.830±0.055•	0.809±0.040•	0.714±0.029•	0.919±0.080	0.808±0.048•	0.803±0.149•
Segment	**0.715±0.034**	0.703±0.048	0.623±0.058•	0.617±0.092•	0.562±0.092•	0.372±0.058•	0.216±0.094•
Character	0.529±0.108	0.555±0.062	**0.567±0.019**	0.428±0.072•	0.565±0.066	0.369±0.081•	0.221±0.184•
Location	0.523±0.049	0.496±0.048	**0.530±0.051**	0.421±0.049•	0.507±0.047	0.375±0.064•	0.312±0.133•
Work	0.386±0.039	0.360±0.020	**0.392±0.048**	0.259±0.019•	0.383±0.053	0.263±0.040•	0.054±0.029•
Usps	**0.784±0.042**	0.694±0.038•	0.464±0.040•	0.085±0.049•	0.739±0.042•	0.877±0.036∘	0.567±0.073•
Pen	0.875±0.029	0.801±0.022•	0.708±0.035•	0.106±0.039•	0.807±0.028•	**0.876±0.044**	0.824±0.046•
Letter	**0.441±0.015**	0.199±0.026•	0.401±0.029•	0.036±0.010•	0.436±0.032	0.357±0.020•	0.214±0.031•
Sensorless	0.480±0.019	0.517±0.016∘	0.469±0.018	-	**0.614±0.017**∘	0.423±0.021•	0.536±0.025∘

Table 3: Classification accuracy (mean±std) of each comparing algorithm on the benchmark data sets ($p = 0.5\%$), where the best performance on each data set is shown in boldface. In addition, •/∘ indicates whether EUPAL achieves significantly superior/inferior to the comparing approach on each data set (pairwise t-test at 0.05 significance level).

Data Set	Comparing Approach						
	EUPAL	PLANETOID	SSODM	S4VM	COTRADE	IPAL	LIBSVM
Ecoli	**0.751±0.129**	0.703±0.085	0.716±0.127	0.522±0.121•	0.069±0.050•	0.644±0.130•	0.078±0.103•
Deter	**0.794±0.063**	0.726±0.079	0.596±0.091•	0.654±0.075•	0.014±0.035•	0.751±0.184	0.015±0.041•
BHP	**0.531±0.078**	0.518±0.039	0.429±0.039•	0.392±0.047•	0.395±0.115•	0.342±0.174•	0.412±0.098•
Yeast	**0.480±0.069**	0.465±0.048	0.449±0.037	0.345±0.058•	0.379±0.067•	0.410±0.070•	0.353±0.055•
Wireless	**0.934±0.029**	0.841±0.040•	0.796±0.040•	0.732±0.021•	0.933±0.065	0.875±0.043•	0.907±0.022•
Segment	0.732±0.039	**0.752±0.036**	0.645±0.089•	0.631±0.044•	0.699±0.054	0.456±0.070•	0.481±0.040•
Character	0.596±0.040	0.610±0.025	0.577±0.034	0.495±0.027•	**0.633±0.023**∘	0.452±0.055•	0.605±0.013
Location	0.554±0.046	0.554±0.038	0.555±0.034	0.454±0.047•	**0.576±0.037**∘	0.430±0.071•	0.502±0.082•
Work	0.454±0.032	0.462±0.026	0.464±0.028	0.322±0.022•	**0.495±0.024**∘	0.353±0.045•	0.346±0.068•
Usps	0.862±0.016	0.757±0.017•	0.574±0.054•	0.138±0.036•	0.803±0.015•	**0.902±0.018**∘	0.811±0.025•
Pen	0.919±0.012	0.856±0.017•	0.793±0.016•	0.082±0.026•	0.857±0.017•	**0.935±0.019**∘	0.897±0.016•
Letter	0.535±0.018	0.125±0.019•	0.463±0.019•	0.041±0.010•	**0.547±0.022**	0.482±0.030•	0.499±0.031•
Sensorless	0.550±0.015	0.533±0.009•	0.506±0.008•	-	**0.691±0.013**∘	0.505±0.013•	0.650±0.014∘

Table 4: Classification accuracy (mean±std) of each comparing algorithm on the benchmark data sets ($p = 1\%$), where the best performance on each data set is shown in boldface. In addition, •/∘ indicates whether EUPAL achieves significantly superior/inferior to the comparing approach on each data set (pairwise t-test at 0.05 significance level).

- S4VM (Li and Zhou 2015): An SVM-based semi-supervised learning approach which works by exploiting an ensemble of low-density separators simultaneously to help induce robust semi-supervised SVM classifier. [suggested configuration: $T = 100, \lambda = 3$]
- COTRADE (Zhang and Zhou 2011): A co-training style semi-supervised learning approach which works by employing data editing techniques to enable reliable labeling information communication between two classifiers. [suggested configuration: base learner with LIBSVM]

For SSODM, S4VM and COTRADE, one-vs-rest strategy is employed to enable multi-class classification. In addition to the four semi-supervised learning approaches, the other two approaches are further employed for comparative studies:

- IPAL (Zhang and Yu 2015): A partial label learning approach which learns from examples with candidate label sets by performing label propagation procedure over kNN graph constructed over training examples. Specifically, IPAL is adapted to learn from $\mathcal{D}_l \bigcup \mathcal{D}_u$ by treating examples in $\mathcal{D}_l$ and $\mathcal{D}_u$ with singleton and full-sized candidate label set respectively. [suggested configuration: $\alpha = 0.95$, $k = 5$]
- LIBSVM (Chang and Lin 2011): The popular LIBSVM is utilized as the baseline approach which induces the multi-class classification model by learning from examples in $\mathcal{D}_l$. [suggested configuration: RBF kernel]

As shown in Table 1, the values of k (number of nearest neighbors), α (balancing parameter), and T (maximum number of iterations) for EUPAL are set to be 5, 0.4 and 50 respectively. Furthermore, the supervised training algorithm

Data Set	Comparing Approach						
	EUPAL	PLANETOID	SSODM	S4VM	COTRADE	IPAL	LIBSVM
Ecoli	**0.839±0.053**	0.804±0.063	0.801±0.050	0.730±0.040•	0.755±0.058•	0.761±0.074•	0.762±0.141
Deter	**0.869±0.063**	0.826±0.051•	0.848±0.033	0.783±0.058•	0.817±0.048•	0.854±0.067	0.562±0.094•
BHP	0.615±0.050	0.596±0.059	0.591±0.036	0.480±0.045•	**0.682±0.039**∘	0.331±0.101•	0.648±0.049∘
Yeast	0.541±0.044	0.526±0.041	**0.544±0.035**	0.432±0.038•	0.506±0.043•	0.498±0.033•	0.485±0.047
Wireless	**0.972±0.007**	0.946±0.017•	0.876±0.010•	0.795±0.032•	0.964±0.010•	0.937±0.011•	0.967±0.010•
Segment	0.852±0.031	**0.862±0.027**	0.813±0.010•	0.778±0.036•	0.832±0.035	0.741±0.031•	0.857±0.019
Character	0.694±0.027	0.671±0.018•	0.577±0.031•	0.530±0.023•	**0.724±0.020**∘	0.632±0.031•	0.712±0.018∘
Location	0.682±0.022	0.621±0.023•	0.565±0.022•	0.507±0.020•	**0.709±0.018**∘	0.605±0.032•	0.707±0.017∘
Work	0.552±0.023	0.530±0.017•	0.488±0.023•	0.336±0.019•	**0.607±0.013**∘	0.469±0.035•	0.598±0.010∘
Usps	0.938±0.007	0.860±0.008•	0.726±0.017•	0.194±0.064•	0.889±0.008•	**0.941±0.006**∘	0.921±0.008•
Pen	0.974±0.006	0.899±0.008•	0.840±0.009•	0.149±0.083•	0.922±0.009•	**0.980±0.003**∘	0.970±0.003•
Letter	0.743±0.017	0.179±0.011•	0.467±0.018•	0.041±0.017•	0.717±0.009•	0.742±0.009	**0.773±0.006**∘
Sensorless	0.696±0.009	0.489±0.004•	0.530±0.012•	-	0.826±0.009∘	0.626±0.004•	**0.844±0.006**∘

Table 5: Classification accuracy (mean±std) of each comparing algorithm on the benchmark data sets ($p = 5\%$), where the best performance on each data set is shown in boldface. In addition, •/∘ indicates whether EUPAL achieves significantly superior/inferior to the comparing approach on each data set (pairwise t-test at 0.05 significance level).

Fraction of labeled examples	EUPAL against					
	PLANETOID	SSODM	S4VM	COTRADE	IPAL	LIBSVM
$p = 0.5\%$	6/6/1	7/6/0	12/0/0	7/5/1	8/4/1	12/0/1
$p = 1\%$	5/8/0	8/5/0	12/0/0	6/3/4	10/1/2	11/1/1
$p = 5\%$	9/4/0	9/4/0	12/0/0	7/1/5	9/2/2	4/3/6
In Total	**20/18/1**	**24/15/0**	**36/0/0**	**20/9/10**	**27/7/5**	**27/4/8**

Table 6: Win/tie/loss counts (pairwise t-test at 0.05 significance level) of EUPAL against each comparing approach under different configurations of the fraction of labeled examples.

$\mathcal{L}$ and the partial label training algorithm $\mathcal{P}$ are instantiated with LIBSVM and IPAL accordingly.

Experimental Results

Tables 3 to 5 report the detailed experimental results for $p = 0.5\%$, 1% and 5% respectively, where the best performance on each data set is shown in boldface.[3] Based on pairwise t-test at 0.05 significance level, we use •/∘ to indicate whether the performance of EUPAL is superior/inferior to the comparing approach on each data set. For illustrative purpose, Figure 1 also shows how the performance of EUPAL changes as p increases from 0.5% to 5% on three benchmark data sets. Furthermore, Table 6 summarizes the win/tie/loss counts of EUPAL against each comparing approach under different configurations of the fraction of labeled examples.

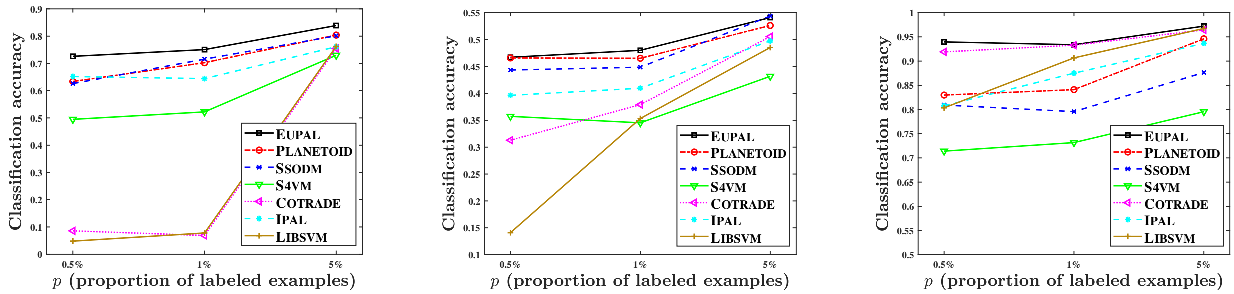
Overall, the following observations can be made based on reported experimental results:

- Out of all the statistical tests (13 data sets × 3 configurations of p), EUPAL achieves significantly better or at least comparable performance to PLANETOID, SSODM, S4VM, COTRADE, IPAL and LIBSVM in 97.4%, 100%, 100%, 74.3%, 87.1% and 79.4% cases.
- It is worthy noting that the performance advantage of EUPAL is more pronounced when the fraction of labeled examples is low ($p = 0.5\%$). This is rather desirable as in semi-supervised learning, it is generally expected that few labeled training examples are available for model induction. Specifically, on the two data sets `Ecoli` and `Deter` with least number of examples, EUPAL achieves best predictive performance under each configuration of p.
- It is also worth noting that when the fraction of labeled examples is relatively high ($p = 5\%$), the performance of IPAL and LIBSVM would outperform other semi-supervised learning approaches including EUPAL on larger data sets `Usps`, `Pen`, `Letter` and `Sensorless`. These results indicate that the exploitation of unlabeled data may not always be beneficial in case sufficient amount of labeled data has been available for model induction.

Further Analysis

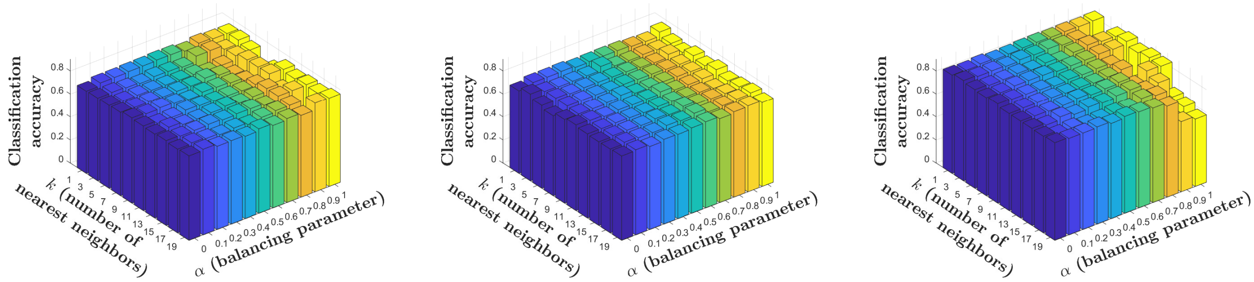
Parameter Sensitivity As shown in Table 1, the EUPAL approach needs to be instantiated with three parameters k, α and T. Figure 2 illustrates how the performance of

[3] Due to its high computational complexity, the performance of S4VM on the `Sensorless` data set does not return within reasonable amount of time.



(a) `Ecoli` (b) `Yeast` (c) `Wireless`

Figure 1: Classification accuracy of each comparing algorithm changes as p (fraction of labeled examples) increases from 0.5% to 5%. (a) `Ecoli`; (b) `Yeast`; (c) `Wireless`.



(a) $p = 0.5\%$ (b) $p = 1\%$ (c) $p = 5\%$

Figure 2: Classification accuracy of EUPAL on the `Segment` data set changes as k (number of nearest neighbors) increases from 1 to 19 with step-size 2 and α (balancing parameter) increases from 0 to 1 with step-size 0.1. (a) $p = 0.5\%$; (b) $p = 1\%$; (c) $p = 5\%$.

EUPAL changes under varying configurations of k and α ($T = 50$) on the `Segment` data set. It is obvious that the performance of EUPAL is stable across a broad range of parameter settings. Similar observations can be made on other data sets as well.

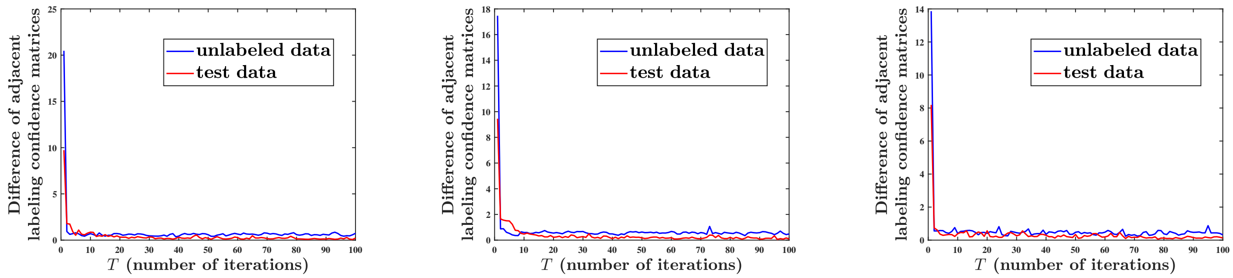
Furthermore, we investigate the convergence property of EUPAL as the number of training iterations T increases. Let $\mathbf{F}^{(t)}$ be the labeling confidence matrix predicted by EUPAL on the unlabeled data (or test data), Figure 3 illustrates how the difference of labeling confidence matrix within two adjacent iterations (i.e. $\|\mathbf{F}^t - \mathbf{F}^{t-1}\|_2$) changes on unlabeled data (or test data). It is obvious that EUPAL would converge to stable performance as the number of iterations T is greater than 10. Based on the above observations, we adopt the parameter configurations of $k = 5$, $\alpha = 0.4$ and $T = 50$ for EUPAL in this paper.

Quality of Partial Label Assignment Partial label assignment serves as the key strategy of EUPAL for exploiting unlabeled data, where it is assumed that the ground-truth label would reside in the candidate label set assigned to the unlabeled data. Let $\mathcal{D}_p = \{(\boldsymbol{x}_j, S_j) \mid 1 \leq j \leq U\}$ be the partial label set generated in the t-th iteration, we can measure the transductive accuracy $\frac{1}{\sum_{j=1}^{U}|S_j|}\sum_{j=1}^{U} [\![y_j \in S_j]\!]$ to show the quality of partial label assignment. Here, y_j corresponds to the ground-truth label of $\boldsymbol{x}_j$.

Figure 4 illustrates how the transductive accuracy of EUPAL changes as the number of iterations increases on `Ecoli`, `Wireless` and `Pen`. It is obvious that EUPAL is capable of assigning candidate label set to unlabeled data which consists of the ground-truth label with high probability. The quality of partial label assignment also improves as the iterative procedure proceeds.

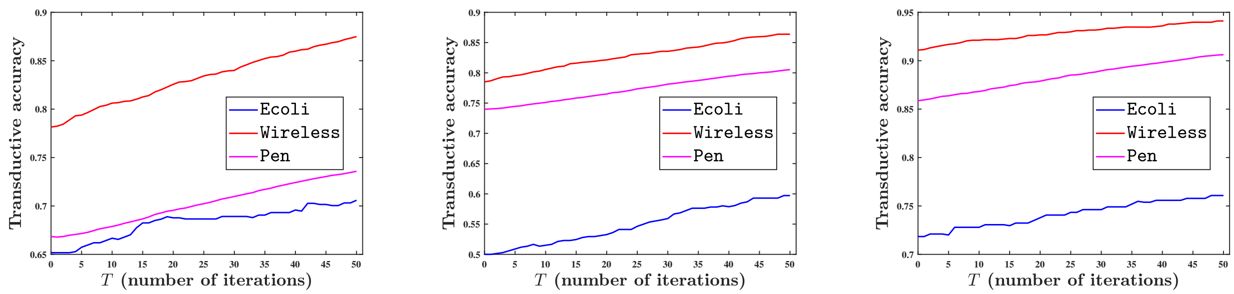
Related Works

Co-training (Blum and Mitchell 1998; Zhang and Zhou 2011; Ma et al. 2020) is one of the most representative strategies for semi-supervised learning, where two classifiers are trained on two different views and the most confident labels predicted by one classifier on unlabeled data is iteratively communicated to the other classifier for model update. The proposed EUPAL approach shares similar strategy with co-training by communicating labeling information between two classifiers, while EUPAL doesn't assume two views for instance representation and one classifier is induced based on the partial label assignment over unlabeled examples. Semi-supervised SVM (Joachims 1999; Chapelle, Sindhwani, and Keerthi 2008) serves as another popular strategy to learn from labeled and unlabeled examples, where the maximum margin classifier and pseudo-label assignment on unlabeled examples are optimized alternatively. Due to the error-prone nature of estimated pseudo-label, there have been attempts in developing safe semi-supervised SVM techniques to ensure beneficial exploitation of unlabeled data (Li and Zhou 2015; Li, Kwok, and



(a) $p = 0.5\%$ (b) $p = 1\%$ (c) $p = 5\%$

Figure 3: Difference of labeling confidence matrix within two adjacent iterations changes on unlabeled data and test data (w.r.t. `Segment` data set). (a) $p = 0.5\%$; (b) $p = 1\%$; (c) $p = 5\%$.



(a) $p = 0.5\%$ (b) $p = 1\%$ (c) $p = 5\%$

Figure 4: Transductive accuracy of EUPAL on unlabeled examples (i.e. the probability of ground-truth label residing in the candidate label set) changes as the number of iterations increases. (a) $p = 0.5\%$; (b) $p = 1\%$; (c) $p = 5\%$.

Zhou 2016; Ke et al. 2020). Graph-based techniques such as label propagation (Zhou et al. 2004; Chong et al. 2020) estimate pseudo-labels on unlabeled examples by utilizing graph structure over labeled and unlabeled examples. Different to the transductive nature of graph-based semi-supervised learning techniques, EUPAL is capable of making predictions on unseen examples other than unlabeled ones.

Partial label learning deals with inaccurate supervision where the training example is assigned with a set of candidate labels among which only one is valid. The major strategy to learn from partial label examples is trying to disambiguate the candidate label set, which can be instantiated via identification-based disambiguation or averaging-based disambiguation. For identification-based disambiguation, the ground-truth label is treated as latent variable whose value is identified based on iterative estimation procedure (Nguyen and Caruana 2008; Liu and Dietterich 2012; Wang, Li, and Zhang 2019; Lv et al. 2020). For averaging-based disambiguation, candidate labels are treated in an equal manner whose modeling outputs are averaged to yield the final prediction (Cour, Sapp, and Taskar 2011; Zhang and Yu 2015; Gong et al. 2018). In addition to the disambiguation strategy, there have been some recent attempts in learning from partial label examples by transforming the partial label learning problem into other well-established learning problems (Chen et al. 2014; Wu and Zhang 2018; Lyu et al. 2019).

Conclusion

In this paper, we investigate an alternative strategy for unlabeled data exploitation via partial label assignment where a set of candidate labels rather than a single pseudo-label is assigned to the unlabeled example. Accordingly, one classifier is trained on partial label examples with candidate label sets, which iteratively communicates labeling information with the other classifier induced from label examples for model update. Extensive experiments on a number of benchmark data sets show that the proposed approach serves as a promising strategy for unlabeled data exploitation, especially when the fraction of labeled examples is low in the training set.

Acknowledgments

This work was completed while Zhen-Ru Zhang is conducting her internship at Tencent. The authors wish to thank the anonymous reviewers for their helpful comments and suggestions. This work was supported by the National Key R&D Program of China (2018YFB1004300), the China University S&T Innovation Plan Guided by the Ministry of Education, and partially supported by the Collaborative Innovation Center of Novel Software Technology and Industrialization. We thank the Big Data Center of Southeast University for providing the facility support on the numerical calculations in this paper.

References

Blum, A.; and Mitchell, T. 1998. Combining labeled and unlabeled data with co-training. In *Proceedings of the 11th Annual Conference on Computational Learning Theory*, 92–100. Madison, WI.

Chang, C.-C.; and Lin, C.-J. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2: 27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chapelle, O.; Sindhwani, V.; and Keerthi, S. S. 2008. Optimization techniques for semi-supervised support vector machines. *Journal of Machine Learning Research* 9(Feb): 203–233.

Chen, Y.-C.; Patel, V. M.; Chellappa, R.; and Phillips, P. J. 2014. Ambiguously labeled learning using dictionaries. *IEEE Transactions on Information Forensics and Security* 9(12): 2076–2088.

Chong, Y.; Ding, Y.; Yan, Q.; and Pan, S. 2020. Graph-based semi-supervised learning: A review. *Neurocomputing* 408: 216–230.

Cour, T.; Sapp, B.; and Taskar, B. 2011. Learning from partial labels. *Journal of Machine Learning Research* 12(May): 1501–1536.

Dua, D.; and Graff, C. 2017. UCI Machine Learning Repository. URL http://archive.ics.uci.edu/ml.

Gong, C.; Liu, T.; Tang, Y.; Yang, J.; Yang, J.; and Tao, D. 2018. A regularization approach for instance-based superset label learning. *IEEE Transactions on Cybernetics* 48(3): 967–978.

Joachims, T. 1999. Transductive inference for text classification using support vector machines. In *Proceedings of the 16th International Conference on Machine Learning*, 200–209. Bled, Slovenia.

Ke, J.; Gong, C.; Liu, T.; Zhao, L.; Yang, J.; and Tao, D. 2020. Laplacian Welsch regularization for robust semisupervised learning. *IEEE Transactions on Cybernetics* 1–14.

Li, Y.-F.; Kwok, J. T.; and Zhou, Z.-H. 2016. Towards safe semi-supervised learning for multivariate performance measures. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 1816–1822. Phoenix, AZ.

Li, Y.-F.; and Zhou, Z.-H. 2015. Towards Making Unlabeled Data Never Hurt. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(1): 175–188.

Liu, L.; and Dietterich, T. G. 2012. A conditional multinomial mixture model for superset label learning. In Bartlett, P.; Pereira, F. C. N.; Burges, C. J. C.; Bottou, L.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems 25*, 548–556. Cambridge, MA: MIT Press.

Lv, J.; Xu, M.; Feng, L.; Niu, G.; Geng, X.; and Sugiyama, M. 2020. Progressive identification of true labels for partial-label learning. In *Proceedings of the 37th International Conference on Machine Learning*.

Lyu, G.; Feng, S.; Wang, T.; Lang, C.; and Li, Y. 2019. GM-PLL: Graph matching based partial label learning. *IEEE Transactions on Knowledge and Data Engineering* .

Ma, F.; Meng, D.; Dong, X.; and Yang, Y. 2020. Self-paced multi-view co-training. *Journal of Machine Learning Research* 21: 1–38.

Nguyen, N.; and Caruana, R. 2008. Classification with partial labels. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 551–559. Las Vegas, NV.

Ren, X.; He, W.; Qu, M.; Voss, C. R.; Ji, H.; and Han, J. 2016. Label noise reduction in entity typing by heterogeneous partial-label embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1825–1834. San Francisco, CA.

van Engelen, J. E.; and Hoos, H. H. 2020. A survey on semi-supervised learning. *Machine Learning* 109: 373–440.

Wang, D.-B.; Li, L.; and Zhang, M.-L. 2019. Adaptive graph guided disambiguation for partial label learning. In *Proceedings of the 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 83–91. Anchorage, AK.

Wu, X.; and Zhang, M.-L. 2018. Towards enabling binary decomposition for partial label learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2868–2974. Stockholm, Sweden.

Yang, Z.; Cohen, W. W.; and Salakhutdinov, R. 2016. Revisiting semi-supervised learning with graph embeddings. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, 40–48. New York, NY.

Zhang, M.-L.; and Yu, F. 2015. Solving the partial label learning problem: An instance-based approach. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, 4048–4054. Buenos Aires, Argentina.

Zhang, M.-L.; and Zhou, Z.-H. 2011. CoTrade: Confident co-training with data editing. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics* 41(6): 1612–1626.

Zhang, T.; and Zhou, Z.-H. 2018. Semi-Supervised Optimal Margin Distribution Machines. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 3104–3110. Stockholm, Sweden.

Zhou, D.; Bousquet, O.; Lal, T. N.; Weston, J.; and Schölkopf, B. 2004. Learning with local and global consistency. In Thrun, S.; Saul, L.; and Schölkopf, B., eds., *Advances in Neural Information Processing Systems 16*, 284–291. Cambridge, MA: MIT Press.

Zhou, Z.-H. 2017. A brief introduction to weakly supervised learning. *National Science Review* 5(1): 44–53.

Zhu, X.; and Goldberg, A. B. 2009. Introduction to semi-supervised learning. In Brachman, R. J.; and Dietterich, T. G., eds., *Synthesis Lectures to Artificial Intelligence and Machine Learning*, 1–130. San Francisco, CA: Morgan & Claypool Publishers.