Learning from Label Proportions with Generative Adversarial Networks

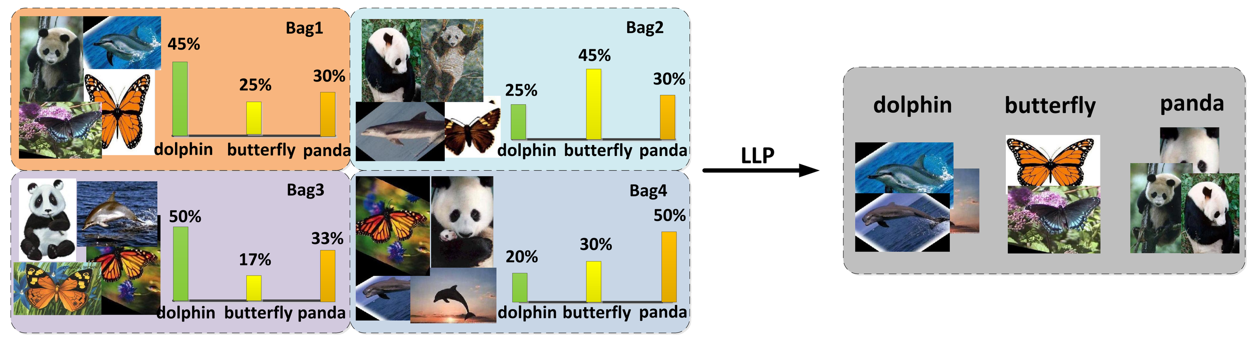
Jiabin Liu, Bo Wang, Zhiquan Qi*, Yingjie Tian, and Yong Shi
University of Chinese Academy of Sciences & University of International Business and Economics
qizhiquan@foxmail.com

1. Problem Description

From supervised learning to weakly supervised learning: Fully labeled data is not always handy to utilize. Firstly, it is infeasible or labor-intensive to obtain abundant accurate labeled data. Secondly, labels are not accessible under certain circumstances, such as privacy constraints. Hence, weakly labeled learning is ubiquitous, e.g., Semi-supervised learning and Multi-instance Learning.
Learning from label proportions (LLP): The data belongs to different categories and is partitioned into several non-overlapping groups. In each group, we only know label proportions in different categories and the samples' feature information. The task is to pursue an instance-level classifier.

2. An illustration of multi-class LLP

The data belongs to three categories and is partitioned into four non-overlapping groups. In each group, the sizes of green, blue, and orange rectangles respectively denote available label proportions in different categories. We only know the sample feature information and class proportions in every group.



3. Challenges & Contributions

The main challenge for LLP is to shrink the uncertainty in label inference based on the bag-level proportional information. Statistical approaches [3, 1] are extremely constrained by strict assumption on data distribution and prior knowledge, while the SVM-based methods [4, 5, 2] suffer from the NP-hard combinatorial optimization issue, thus is lack of scalability (shallow models).
In this paper, we propose a simple improvement based on entropy regularization for the existing deep LLP solver. Besides, we reveal relationship between prior class proportions and posterior class likelihoods. Also, we offer a decomposition representation of the class likelihood with respect to the prior class proportions, which verifies the existence of the final classifier. Finally, we empirically show that our method can achieve SOTA performance on large-scale LLP problems with a low computational complexity.
Our code is available at https://github.com/liujiabin008/LLP-GAN. Also, please refer to our paper for more details.

References

[1] Giorgio Patrini et al. "(Almost) no label no cry". In: *Advances in Neural Information Processing Systems.* 2014, pp. 190–198.

[2] Zhiquan Qi, Bo Wang, Fan Meng, et al. "Learning with label proportions via NPSVM". In: *IEEE Transactions on Cybernetics* 47.10 (2017), pp. 3293–3305.

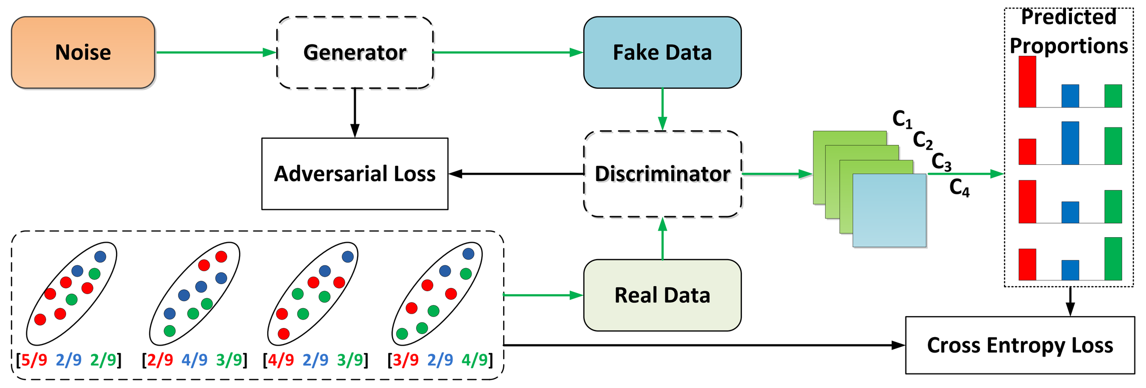
[3] Novi Quadrianto, Alex J. Smola, Tiberio S. Caetano, et al. "Estimating labels from label proportions". In: *Journal of Machine Learning Research* 10.Oct (2009), pp. 2349–2374.

[4] Stefan Rueping. "SVM classifier estimation from group probabilities". In: *International Conference on Machine Learning.* 2010, pp. 911–918.

[5] Felix X. Yu, Dong Liu, Sanjiv Kumar, et al. "$\propto$-SVM for learning with label proportions". In: *International Conference on Machine Learning.* 2013, pp. 504–512.

4. Problem Settings & Our Approach

All the bags are disjoint and let $\mathcal{B}_i=\{\mathbf{x}_i^1,\mathbf{x}_i^2,\cdots,\mathbf{x}_i^{N_i}\}, i=1,2,\cdots,n$ be the bags in training set. Training data is $\mathcal{D}=\mathcal{B}_1\cup\mathcal{B}_2\cup\cdots\cup\mathcal{B}_n, \mathcal{B}_i\cap\mathcal{B}_j=\emptyset, \forall i\neq j$, where the total number of bags is n.
Assuming we have K classes, for $\mathcal{B}_i$, let $\mathbf{p}_i$ be a K-element vector, where the k^{th} element p_i^k is the proportion of instances belonging to the class k, with the constraint $\sum_{k=1}^K p_i^k=1$.
An illustration of our LLP-GAN framework



The Objective of Discriminator The *ideal* optimization problem for discriminator of LLP-GAN:

$$\max_D V(G,D)=\sum_{i=1}^n E_{\mathbf{x}\sim p_d^i}\Big[log P_D(y\leq K|\mathbf{x})\Big]+E_{\mathbf{x}\sim p_g}\Big[log P_D(K+1|\mathbf{x})\Big]+\lambda\sum_{i=1}^n \mathbf{p}_i^{\intercal} log(\overline{\mathbf{p}}_i). \quad (1)$$

Here, $p_g(\mathbf{x})$ is the distribution of the synthesized data.
Then, the Lower Bound Approximation is:

$$-CE_{\mathcal{L}}(\mathbf{p},\overline{\mathbf{p}})\geqslant\sum_{i=1}^n\sum_{k=1}^K p_i(k)E_{\mathbf{x}\sim p_d^i}\Big[log\tilde{p}_D(k|\mathbf{x})\Big]. \quad (2)$$

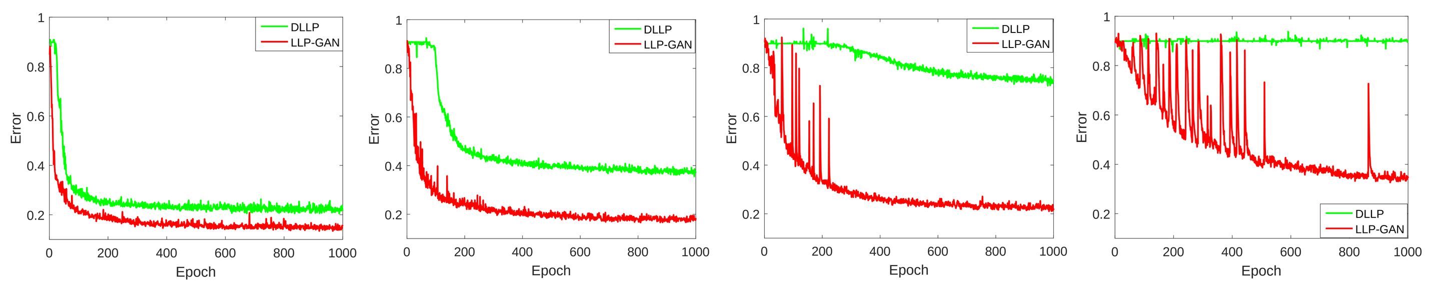
Theorem For fixed G, the optimal discriminator D^* for $\widetilde{V}(G,D)$ satisfies:

$$P_{D^*}(y=k|\mathbf{x})=\frac{\sum_{i=1}^n p_i(k)p_d^i(\mathbf{x})}{\sum_{i=1}^n p_d^i(\mathbf{x})+p_g(\mathbf{x})}, k=1,2,\cdots,K. \quad (3)$$

The Objective Function of Generator We apply feature matching (FM) to the generator:

$$L(G)=\|E_{\mathbf{x}\sim\frac{1}{n}p_d}f(\mathbf{x})-E_{\mathbf{x}\sim p_g}f(\mathbf{x})\|_2^2 \quad (4)$$

5. Analysis on Convergence & Effectiveness



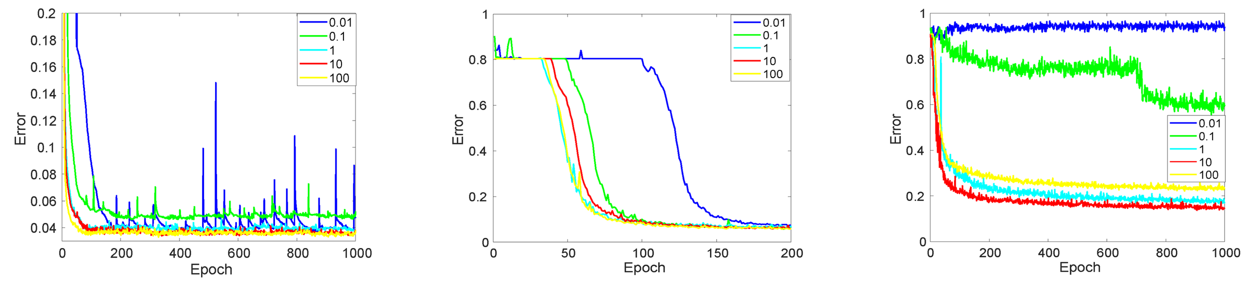
(a) Bag size: 16 (b) Bag size: 32 (c) Bag size: 64 (d) Bag size: 128

Fig. 1: The convergence curves on CIFAR-10 w/ different bag sizes.

(a) GANs with FM (b) 50 epochs (ours) (c) 60 epochs (ours) (d) 70 epochs (ours)

Fig. 2: Generated samples on CIFAR-10.

6. Hyperparameter & Performance



(a) λ on MNIST (b) λ on SVHN (c) λ on CIFAR-10

Fig. 3: The convergence curves on CIFAR-10 w/ different bag sizes.

Table: Test error rates (%) on benchmark datasets w/ different bag sizes.

Dataset	Algorithm	Bag Size 16	32	64	128	Baseline CNNs
MNIST	DLLP	1.23 (0.100)	1.33 (0.094)	1.57 (0.088)	3.55 (0.27)	0.36
	LLP-GAN	**1.10** (0.026)	**1.23** (0.088)	**1.40** (0.089)	**3.49** (0.27)	
SVHN	DLLP	4.45 (0.069)	5.29 (0.54)	5.80 (0.91)	39.73 (1.60)	2.35
	LLP-GAN	**4.03**(0.021)	**4.83**(0.51)	**5.42**(0.59)	**11.17**(1.12)	
CIFAR-10	DLLP	19.70 (0.77)	34.39 (0.82)	68.32 (1.34)	82.89 (2.66)	9.27
	LLP-GAN	**13.68** (0.35)	**16.23** (0.43)	**21.03** (1.82)	**27.39** (4.31)	
CIFAR-100	DLLP	53.24(0.77)	98.38(0.11)	98.65(0.09)	98.98(0.08)	35.68
	LLP-GAN	**50.95**(0.67)	**56.44**(0.78)	**64.37**(1.52)	**85.01**(1.81)	

Table: Binary test error rates (%) on benchmark datasets w/ different bag sizes.

Dataset	Algorithm	Bag Size 16	32	64	128
MNIST	InvCal	0.50	0.55	1.25	0.1
	alter-pSVM	0.20	0.20	0.25	0.2
	DLLP	0.049	0.049	0.049	0.049
	LLP-GAN	**0.047**	**0.047**	**0.047**	**0.047**
CIFAR-10	InvCal	28.95	29.16	26.47	31.84
	alter-pSVM	24	26.74	30.32	27.95
	DLLP	11.31	15.83	18.96	22.59
	LLP-GAN	**1.39**	**1.61**	**11.59**	**18.29**
SVHN	InvCal	11.55	13.35	12.95	12.70
	alter-pSVM	7.05	7.95	7.95	11.15
	DLLP	**1.38**	**1.7**	3.77	24.45
	LLP-GAN	1.49	1.8	**3.46**	**9.23**

7. Conclusions & Future Work

This paper proposed a new algorithm LLP-GAN for LLP problem in virtue of the adversarial learning based on GANs. Our method is superior to existing methods in three aspects.

- Nice theoretical properties of the lower bound to facilitate SGD
- A probabilistic classifier with decomposable representation
- Scalability: e.g., image data applications

Future Work:

- Learning complexity in the sense of PAC
- Imprecise proportions problem
- The performance with other GANs models, e.g., WGAN
- The performance of LLP-GAN on tabular data and datasets with structured (non-random) bags

Acknowledgements

This work is supported by National Natural Science Foundation of China.
Bo Wang would like to acknowledge that this research was conducted during his visit at Texas A&M University and thank Dr. Xia (Ben) Hu for his hosting and insightful discussions.