– Supplemental Material – Label Propagation with Augmented Anchors: A Simple Semi-Supervised Learning baseline for Unsupervised Domain Adaptation

Yabin Zhang[123], Bin Deng[12], Kui Jia[12], and Lei Zhang[34]

[1] South China University of Technology, Guangzhou, China
[2] Pazhou Lab, Guangzhou, China
[3] DAMO Academy, Alibaba Group
[4] Dept. of Computing, The Hong Kong Polytechnic University
zhang.yabin@mail.scut.edu.cn, bindeng.scut@gmail.com, kuijia@scut.edu.cn, cslzhang@comp.polyu.edu.hk

A Proof of Proposition 1

Proof. For any $\boldsymbol{x}_i, \boldsymbol{x}_j \in X$, we define the boolean value $P(\boldsymbol{x}_i, \boldsymbol{x}_j) := TRUE$ iff there exists a sequence of instances $(\boldsymbol{x}_i, \boldsymbol{x}_{t_1}, \boldsymbol{x}_{t_2}, ..., \boldsymbol{x}_{t_m}, \boldsymbol{x}_j)$ such that the product of their pair-wise similarities: $\boldsymbol{a}_{it_1} \cdot \boldsymbol{a}_{t_1 t_2} \cdot ... \cdot \boldsymbol{a}_{t_m j} \neq 0$. Then under the assumption made in Proposition 1, minimizing Equation (7) results in $\hat{y}_i = \hat{y}_j$ for those with $P(\boldsymbol{x}_i, \boldsymbol{x}_j) = TRUE$, and $\hat{y}_i = y_i$ for all $\boldsymbol{x}_i \in X_L$, therefore we have

$$Acc = \frac{|\{\boldsymbol{x}_i \in X_U : \exists \boldsymbol{x}_j \in X_L, P(\boldsymbol{x}_i, \boldsymbol{x}_j)\}|}{u}. \tag{A.1}$$

Obviously, enhancing the zero-valued similarity $\boldsymbol{a}_{mn}$ between a data instance $\boldsymbol{x}_m$ (labeled or unlabeled) and a labeled instance $\boldsymbol{x}_n \in X_L$, where $y_m = y_n$, to a positive number leads to non-decreasing value of $|\{\boldsymbol{x}_i \in X_U : \exists \boldsymbol{x}_j \in X_L, P(\boldsymbol{x}_i, \boldsymbol{x}_j)\}|$, and therefore non-decreasing value of *Acc*. In particular, if $\boldsymbol{x}_m \in X_U$ and originally $P(\boldsymbol{x}_m, \boldsymbol{x}_j) = FALSE, \forall \boldsymbol{x}_j \in X_L$, the prediction of $\boldsymbol{x}_m$ changes from original $\hat{y}_m (\neq y_m)$ to $y_n (= y_m)$ and thus the value of *Acc* increases. □

B Analysis

Hyper-parameter α We investigate different values of α (of Equ. (7)) in A^{2}LP. As illustrated in Table A1, the results are stable under a wide range of α (i.e., 0.1∼0.75).

Practical Efficiency To make the proposed methods applicable to datasets with large numbers of instances, we improve the dominating computations of our methods by adopting the NN-Descent [1] to construct the k-nearest neighbor graph (6) and the conjugate gradient [2, 3] to acquire the label predictions $\boldsymbol{F}^*$ (11). As illustrated in Table A2, the NN-Descent [1] accelerates the brute-force

Values of α	0.1	0.25	0.4	0.5	0.6	0.75	0.9	2.0
Acc. (%) of A^2LP	95.7	96.2	96.2	96.0	96.0	95.8	94.3	16.8

Table A1: Results of A^2LP with different values of α on the $\mathbf{P} \rightarrow \mathbf{C}$ task of the ImageCLEF-DA dataset.

implementation of affinity matrix at a factor of 20, and the conjugate gradient-based solution (11) accelerates the closed-form solution (7) at a factor of > 20 on the VisDA-2017-Small task of $24K$ instances, while the classification results drop negligibly (in fact no drop at the precision level of 0.1%).

Table A2: An illustration of the wall-clock time of the (a) graph construction (6) and (b) prediction solution (7) with different implementations on the VisDA-2017-Small task of $24K$ instances based on the Intel Xeon E5-2630 V4 CPU of 2.20GHz.

Methods	Acc. (%)	Time (s)
Brute-force impl.	79.3	182
NN-Descent [1]	79.3	9.0

(a) Graph construction (6)

Methods	Acc. (%)	Time (s)
Closed-form solution (7)	79.3	51.2
CG [2, 3] (11)	79.3	2.4

(b) Predictions solution (7)

C Full Results of VisDA-2017

The full classification results on the VisDA-2017 dataset are illustrated in Table A3.

References

1. Dong, W., Moses, C., Li, K.: Efficient k-nearest neighbor graph construction for generic similarity measures. In: Proceedings of the 20th international conference on World wide web. pp. 577–586 (2011)
2. Hestenes, M.R., Stiefel, E., et al.: Methods of conjugate gradients for solving linear systems. Journal of research of the National Bureau of Standards **49**(6), 409–436 (1952)
3. Zhu, X., Lafferty, J., Rosenfeld, R.: Semi-supervised learning with graphs. Ph.D. thesis, Carnegie Mellon University, language technologies institute, school of ... (2005)

Table A3: Full classification results on the VisDA-2017 dataset for unsupervised domain adaptation (UDA).

Methods	aero.	bike	bus	car	horse	knife	moto.	person	plant	sktb.	train	truck	Avg.
Results based on a 50-layer ResNet													
LP	91.4	81.4	73.3	71.8	94.7	60.8	87.4	62.2	87.8	19.1	86.2	20.9	69.8
A^2LP	95.5	82.8	77.9	70.0	95.2	95.9	86.6	65.3	87.4	42.8	86.4	53.1	78.7
MSTN (reproduced)	86.9	73.2	76.8	67.2	80.7	78.8	71.9	65.1	74.8	76.2	85.6	25.6	71.9
empowered by A^2LP	96.1	83.5	78.3	70.8	95.7	96.3	87.1	66.4	87.4	76.4	86.7	53.8	81.5
CAN (reproduced)	94.5	85.4	**81.9**	**72.3**	96.7	94.9	88.3	78.4	**96.3**	94.7	86.2	57.3	85.6
empowered by A^2LP	**96.3**	**86.2**	81.4	71.7	**97.1**	**96.8**	**89.7**	**79.1**	96.1	**95.4**	**88.6**	**59.1**	**86.5**
Results based on a 101-layer ResNet													
LP	89.6	80.6	65.4	72.9	92.7	74.0	84.2	72.8	87.9	48.4	84.6	33.0	73.9
A^2LP	96.0	82.9	82.2	68.9	95.8	96.0	87.8	66.5	89.6	85.2	88.4	53.2	82.7
MSTN (reproduced)	90.5	73.0	70.2	58.9	84.9	77.0	84.5	79.3	89.6	69.6	89.4	36.0	75.2
empowered by A^2LP	96.4	84.1	82.4	70.1	96.1	96.6	88.2	67.7	91.5	87.5	**89.9**	54.0	83.7
CAN (reproduced)	97.0	**87.2**	82.5	**74.3**	97.8	96.2	**90.8**	80.7	96.6	96.3	87.5	59.9	87.2
empowered by A^2LP	**97.5**	86.9	**83.1**	74.2	**98.0**	**97.4**	90.5	**80.9**	**96.9**	**96.5**	89.0	**60.1**	**87.6**