Selective Partial Domain Adaptation

Pengxin Guo   Jinjing Zhu   Yu Zhang

Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China

BMVC 2022

Abstract

We propose a Selective Partial Domain Adaptation (SPDA) method, which selects useful data for the adaptation to the target domain. Specifically, we firstly design a Maximum of Cosine (MoC) similarity function customized for PDA to select useful data in the source domain to decrease the domain discrepancy. In the MoC similarity function, for each target sample, we select the source sample with the maximal cosine similarity for adaptation. Moreover, a selective training method is designed to add useful target data into the source domain. In detail, the selective training method firstly assigns pseudo-labels to target samples with the self-training strategy and then adds target samples with high confidence in terms of pseudo-labels to the source domain. Based on these two selection operations, the proposed SPDA method can select useful data for domain adaptation.

Introduction

To learn under the PDA setting, a possible way is to select useful source samples whose labels are highly likely to appear in the target domain for the adaptation. However, since the target domain is unlabelled, it is not straightforward to identify which classes are presented in the target domain and which source samples are helpful for the target domain. To solve those issues, we firstly design a Maximum of Cosine (MoC) similarity function customized for PDA to select the source sample with the maximal cosine similarity for each target sample. In this way, we can select the most useful source samples for adaptation and ignore irrelevant source samples which may cause negative transfer. An illustration of the MoC similarity is shown in Figure. 1.

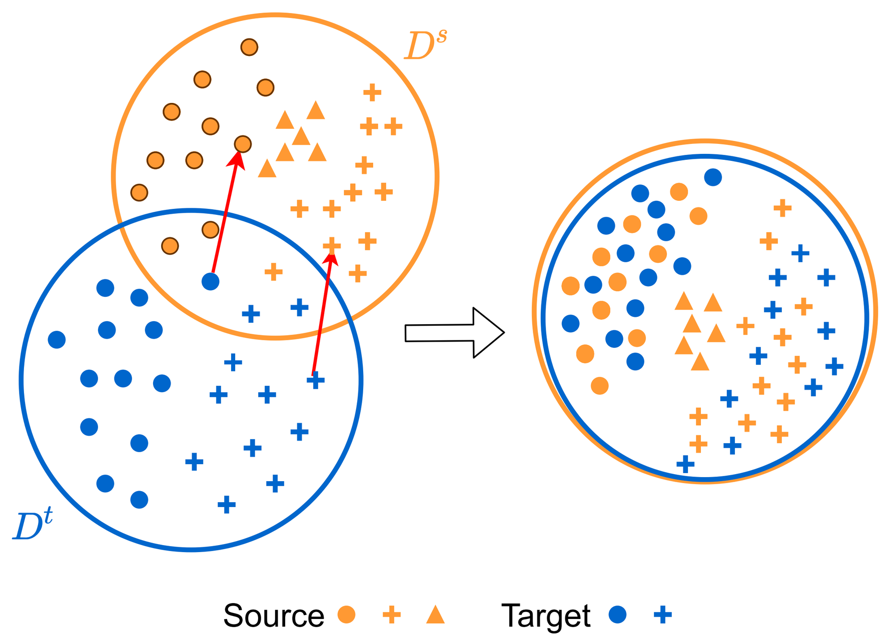

Figure 1. Illustration for the MoC similarity. Since in PDA the label space of the target domain is a subset of that of the source domain, we cannot align these two domains directly. In the MoC similarity function, for each target sample, we select the source sample with the maximal cosine similarity and hope to draw them closely. After that, the source and target domains could be well aligned. *Best viewed in color.*

Method

The proposed SPDA method consists of two selection operations. The first selection operation is to design the MoC similarity to select useful source samples for adaptation and the second one is to utilize the selective training method to select target samples with high confidence pseudo-labels and add them to the source domain.

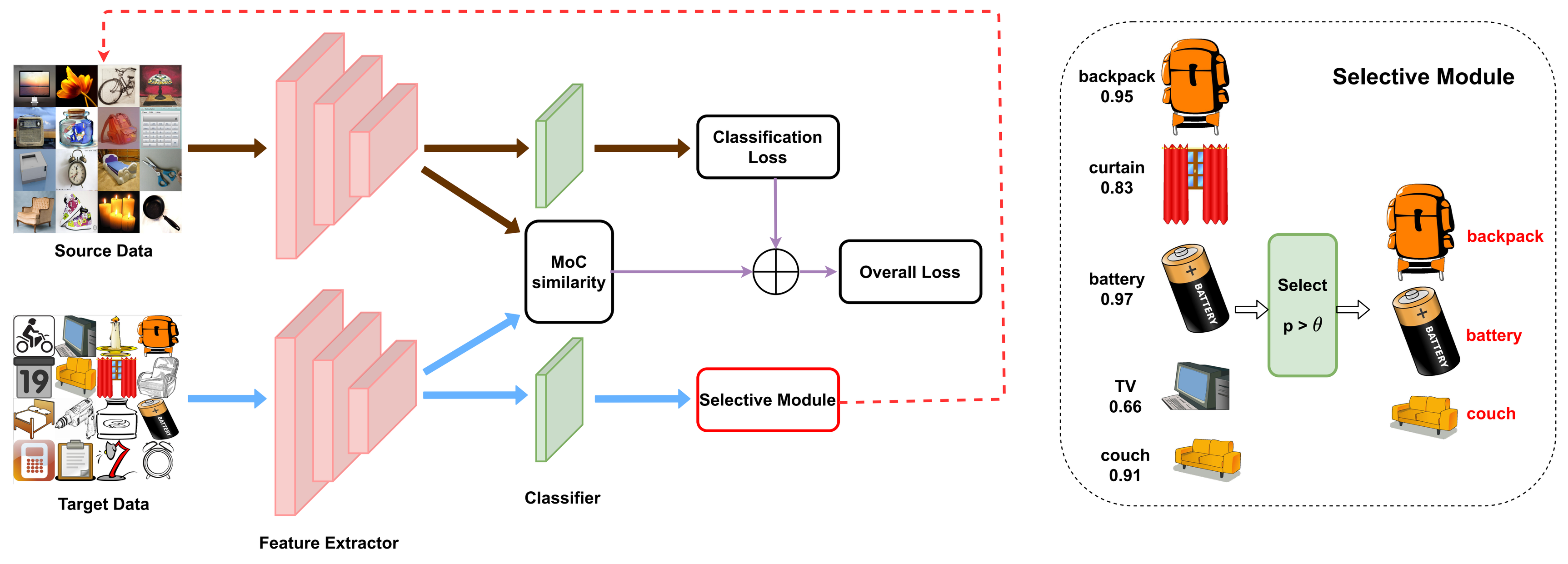

Figure 2. The left figure shows the whole architecture of the SPDA model, whose objective function consists of two parts, including the classification loss on the source data as well as the selected target data with high confidence pseudo-labels and the negative MoC similarity between augmented source samples and target samples. The right figure shows the selective module used for the target data, where only target samples with high confidence pseudo-labels will be added to the source domain.

The MoC similarity function $\mathrm{MoC}(X_S, X_T)$ is formulated as

$$\mathrm{MoC}(X_S, X_T) = \frac{1}{n_t}\sum_{j=1}^{n_t} \max_{i\in[n_s]} \frac{(G(x_s^i))^\top G(x_t^j)}{\|G(x_s^i)\|_2 \|G(x_t^j)\|_2}, \tag{1}$$

where X_S and X_T denote the source and target datasets, respectively, $[n] = \{1, 2, \cdots, n\}$ denotes the set of positive integers up to an integer n, $\|\cdot\|_2$ denotes the L_2 norm, and $G(\cdot)$ denotes the feature extraction network used in SPDA.

We select those target samples with high confidence pseudo-labels as

$$\hat{X}_T = \{x_t^j \mid p_t^j > \theta, \forall j \in [n_t]\}, \tag{2}$$

where θ denotes a threshold to determine whether a pseudo-label is of high confidence.

By combining these two selection operations, the overall objective function of the proposed SPDA method is formulated as

$$\min_{\mathbf{w}} \mathcal{L}_C(\widetilde{X}_S, \{y_s, \hat{y}_t\}) - \lambda \mathrm{MoC}(\widetilde{X}_S, X_T), \tag{3}$$

where $\mathbf{w}$ denotes parameters of the whole network that consists of G and F, $\mathcal{L}_C(\widetilde{X}_S, \{y_s, \hat{y}_t\})$ denotes the classification loss on the labeled source samples X_S with their ground truth labels y_s and the selected target samples $\hat{X}_T$ with their pseudo-labels $\hat{y}_t$, and λ is a hyperparameter to balance the two terms in problem (3).

Experiment

Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg
ResNet-50	75.59±1.09	96.27±0.85	98.09±0.74	83.44±1.12	83.92±0.95	84.97±0.86	87.05
DAN	59.32±0.49	73.90±0.38	90.45±0.36	61.78±0.56	74.95±0.67	67.64±0.29	71.34
DANN	73.56±0.15	96.27±0.26	98.73±0.20	81.53±0.23	82.78±0.18	86.12±0.15	86.50
ADDA	75.67±0.17	95.38±0.23	99.85±0.12	83.41±0.17	83.62±0.14	84.25±0.13	87.03
PADA	86.54±0.31	99.32±0.45	**100.0**±0.00	82.17±0.37	92.69±0.29	95.41±0.33	92.69
IWAN	89.15±0.37	99.32±0.32	99.36±0.24	90.45±0.36	95.62±0.29	94.26±0.25	94.69
SAN	93.90±0.45	99.32±0.52	99.36±0.12	94.27±0.28	94.15±0.36	88.73±0.44	94.96
ETN	94.52±0.20	**100.0**±0.00	**100.0**±0.00	95.03±0.22	96.21±0.27	94.64±0.24	96.73
RTNet	96.20±0.30	**100.0**±0.00	**100.0**±0.00	97.60±0.10	92.30±0.10	95.40±0.10	96.90
BA^{3}US	98.98±0.28	**100.0**±0.00	98.73±0.00	**99.36**±0.00	94.82±0.05	94.99±0.08	97.81
DRCN	88.05	**100.0**	**100.0**	86.00	95.60	95.80	94.30
SRLCWC	92.07	95.84	99.24	94.46	93.68	93.72	94.84
DCC	**99.70**	**100.0**	**100.0**	96.10	95.30	96.30	97.90
SPDA (Ours)	99.32±0.02	**100.0**±0.00	**100.0**±0.00	96.18±0.32	96.03±0.25	**96.56**±0.00	**98.01**

Table 1. Accuracy (%) on the Office-31 dataset under the PDA setting with the ResNet-50 as the backbone.

Method	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
ResNet-50	46.33	67.51	75.87	59.14	59.94	62.73	58.22	41.79	74.88	67.40	48.18	74.17	61.35
DAN	35.70	52.90	63.70	45.00	51.70	49.30	42.40	31.50	68.70	59.70	34.60	67.80	50.30
DANN	43.76	67.90	77.47	63.73	58.99	67.59	56.84	37.07	76.37	69.15	44.30	77.48	61.72
ADDA	45.23	68.79	79.21	64.56	60.01	68.29	57.56	38.89	77.45	70.28	45.23	78.32	62.82
PADA	51.95	67.00	78.74	52.16	53.78	59.03	52.61	43.22	78.79	73.73	56.6	77.09	62.06
IWAN	53.94	54.45	78.12	61.31	47.95	63.32	54.17	52.02	81.28	76.46	56.75	82.90	63.56
SAN	44.42	68.68	74.60	67.49	64.99	77.80	59.78	44.72	80.07	72.18	50.21	78.66	65.30
ETN	59.24	77.03	79.54	62.92	65.73	75.01	68.29	55.37	84.37	75.72	57.66	84.54	70.45
RTNet	63.20±0.10	80.10±0.20	80.70±0.10	66.70±0.10	69.30±0.20	77.20±0.20	71.60±0.30	53.90±0.30	84.60±0.10	77.40±0.20	57.90±0.30	85.50±0.10	72.30
BA^{3}US	60.62±0.45	83.16±0.12	88.39±0.19	71.75±0.19	72.79±0.19	**83.40**±0.59	75.45±0.19	61.59±0.37	86.53±0.22	79.25±0.65	62.80±0.51	86.05±0.26	75.98
DRCN	54.00	76.40	83.00	62.10	64.50	71.00	70.80	49.80	80.50	77.50	59.10	79.90	69.00
SRLCWC	56.21	73.34	80.63	64.08	61.72	66.41	70.83	53.13	83.57	77.01	58.31	81.24	68.87
DCC	59.00	84.40	83.40	67.80	72.70	79.80	68.40	53.20	83.70	75.80	59.00	**88.30**	73.00
SPDA (Ours)	**64.24**±0.24	**87.79**±0.11	**88.74**±0.08	**74.29**±0.22	**75.10**±0.03	79.05±0.33	**79.37**±0.15	58.91±0.13	85.05±0.42	**81.36**±0.09	**67.41**±0.21	84.09±0.38	**77.12**

Table 2. Accuracy (%) on the Office-Home dataset under the PDA setting with the ResNet-50 as the backbone.

Method	R→S	S→R	Avg
ResNet-50	64.28	45.26	54.77
DAN	68.35	47.60	57.98
DANN	73.84	51.01	62.43
PADA	76.50	53.53	65.01
IWAN	71.30	48.60	59.95
SAN	69.70	49.90	59.80
ETN	78.24	*68.53*	*73.39*
BA^{3}US	*69.25*	*74.27*	*71.76*
DRCN	73.20	58.20	65.70
SPDA (Ours)	**92.47**±3.83	**82.91**±1.76	**87.69**

Table 3. Accuracy (%) on the VisDA-2017 dataset under the PDA setting with the ResNet-50 as the backbone.