Confidence Score for Source-Free Unsupervised Domain Adaptation

Jonghyun Lee*

Seoul National University

ICML
International Conference
On Machine Learning

DSAIL
Data Science & AI Laboratory

*speaker

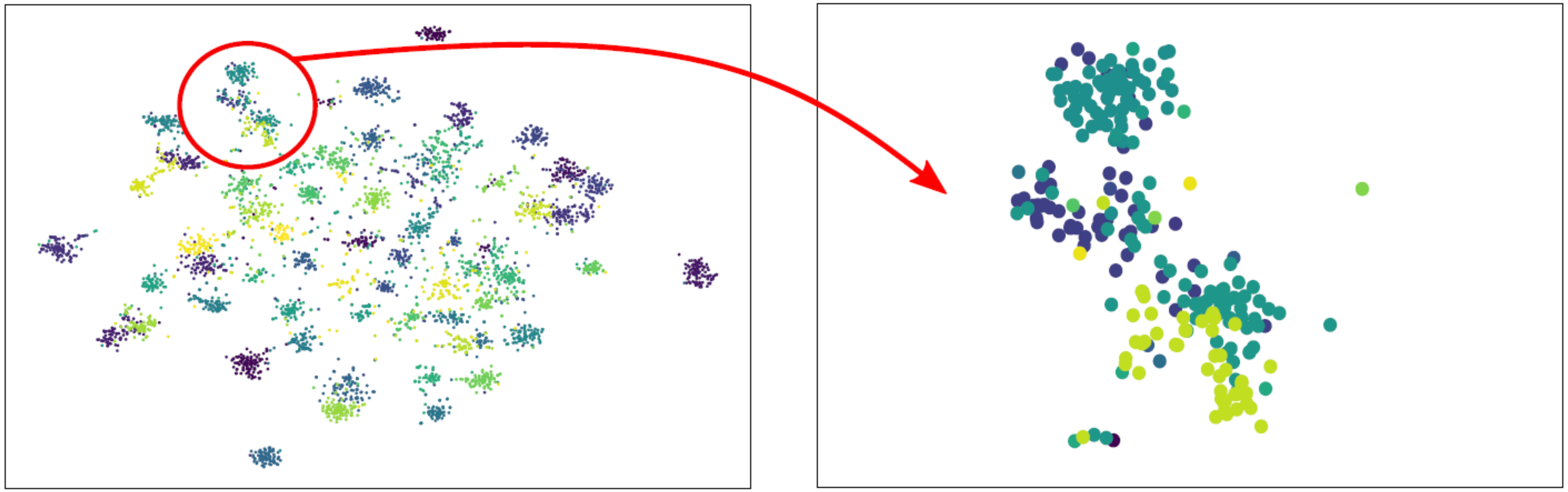
Source-Free Unsupervised Domain Adaptation

- Unsupervised Domain Adaptation (UDA)
  - Training data: $\mathcal{D}_s = \{x_s, y_s\} \cup D_t = \{x_t\}$
  - Initial model: the pre-trained model using ImageNet or from scratch

- **Source-Free** Unsupervised Domain Adaptation (SFUDA)
  - Training data: $\mathcal{D}_t = \{x_t\}$
  - Initial model: the pre-trained model using D_s

The problems in existing SFUDA methods

- The cluster assumption
  - Existing SFUDA methods train the model so that its decision boundary does not pass the target feature cluster
- Pseudo-labeling
  - Existing SFUDA methods use all samples without considering the confidence of corresponding pseudo-labels



The JMDS score

- Our method
  - Propose the confidence score for SFUDA
  - Based on the proposed score, a model determine which samples are important for learning

- The Joint Model-Data Structure (JMDS) score: use scores from two different views
  1. The model-wise score (MPPL score)
     - uses the model prediction $p_M(X_t)$
  2. The data-structure-wise score (LPG score)
     - uses the data-structure-wise probability $p_{\mathrm{data}}(X_t)$ obtained from Gaussian Mixture Modeling (GMM)

The JMDS score

- Pseudo-label
  - $\hat{Y}_t = \arg\max_c p_{data}(X_t)_c$

1. The Model Probability of Pseudo-label (MPPL) score
   - The model-wise probability of the corresponding pseudo-label $\hat{Y}_t$

$$\mathrm{MPPL}(x_i^t) = p_M(x_i^t)_{\hat{y}_i^t}$$

2. Log-Probability Gap (LPG) score
   - The normalized margin of the log data-structure-wise probability

$$\mathrm{MINGAP}(x_i^t) = \min_a \{\log p_{\text{data}}(x_i^t)_{\hat{y}_i^t} - \log p_{\text{data}}(x_i^t)_a\},$$

$$\text{where } \hat{y}_i^t = \arg\max_c p_{\text{data}}(x_i^t)_c, \; a \in \{1, 2, \cdots, K\}, \; a \neq \hat{y}_i^t.$$

$$\mathrm{LPG}(x_i^t) = \frac{\mathrm{MINGAP}(x_i^t)}{\max_j \mathrm{MINGAP}(x_j^t)}$$

The JMDS score

- The JMDS score
  - The product of LPG and MPPL to emphasize confident samples in both scores

$$\mathrm{JMDS}(x_i^t) = \mathrm{LPG}(x_i^t) \cdot \mathrm{MPPL}(x_i^t)$$

  - It contains knowledge on the data structure from LPG and on the model from MPPL

- In SFUDA
  - The model includes knowledge of the source domain
  - The data structure includes knowledge of the target domain
  - The JMDS score is the **only** confidence score that considers knowledge from both domains

Our SFUDA method

- Confidence score Weighting Adaptation using the JMDS (CoWA-JMDS)
  - By simply adopting the JMDS score as a sample weight, we propose an effective SFUDA method

$$\mathcal{L}_{\text{CoWA-JMDS}}(x_i^t) = \text{JMDS}(x_i^t) \cdot \mathcal{L}_{\text{CE}}(p_M(x_i^t), \hat{y}_i^t)$$

  - Weight mixup
    - Utilize more knowledge of the target feature distribution
    - Mix the confidence score together

$$\begin{aligned}
\tilde{x}^t &= \gamma \cdot x_i^t + (1-\gamma) \cdot x_j^t, \\
\tilde{y}^t &= \gamma \cdot o(\hat{y}_i^t) + (1-\gamma) \cdot o(\hat{y}_j^t), \\
&\boxed{w(\tilde{x}^t) = \gamma \cdot \text{JMDS}(x_i^t) + (1-\gamma) \cdot \text{JMDS}(x_j^t)} \\
\mathcal{L}_{\text{Mixup}}(\tilde{x}^t, \tilde{y}^t) &= w(\tilde{x}^t) \cdot \mathbb{E}_{\tilde{y}^t}[-\log p_M(\tilde{x}^t)]
\end{aligned}$$

Our SFUDA method

- The full procedure of CoWA-JMDS

Algorithm 3 CoWA-JMDS

Input: Unlabeled target data X_t, the model $M = g \circ f$.
epoch $\leftarrow$ 0.
repeat
  if Partial-set scenario **then**
    Perform class estimation.
  end if
  Perform GMM on $f(X_t)$ and compute $p_{\text{data}}(X_t)$.
  if Open-set scenario **then**
    Perform known/unknown classification.
  end if
  Compute JMDS score using Equation (5).
  for $i \leftarrow 1$ **to** $iterations\_per\_epoch$ **do**
    if No weight Mixup **then**
      Compute loss using Equation (6).
    else if Weight Mixup **then**
      Obtain mixed inputs $\tilde{x}^t$, pseudo-labels $\tilde{y}^t$, and JMDS scores $\tilde{w}^t$ using Equation (7).
      Compute loss using Equation (8).
    end if
    Update the model M using loss.
  end for
  epoch $\leftarrow$ epoch+1.
until epoch $< max\_epoch$

Evaluation for the JMDS score

- Evaluate the JMDS score based on AURC [1] which is a reliable measurement for the confidence score
  - GMM + JMDS score shows the best performance

Table 1. Evaluation of the JMDS score based on AURC.

Dataset	Task	Naïve PL+Maxprob	Naïve PL+Ent	SSPL+Cossim	GMM+Cossim	GMM+MPPL	GMM+LPG	GMM+JMDS
Office-31	A → D	0.047	0.051	**0.018**	0.031	0.039	0.033	0.033
	A → W	0.074	0.081	**0.034**	0.045	0.059	0.042	0.044
	D → A	0.158	0.165	0.140	0.130	0.131	0.127	**0.115**
	D → W	0.007	0.008	0.009	0.009	0.005	0.004	**0.004**
	W → A	0.157	0.167	**0.107**	0.108	0.132	0.120	0.113
	W → D	0.002	0.002	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**
	Avg.	0.074	0.079	**0.052**	0.054	0.061	0.055	**0.052**
Office-Home	Ar → Cl	0.308	0.316	0.296	0.274	0.278	0.265	**0.256**
	Ar → Pr	0.140	0.145	**0.100**	0.105	0.116	0.125	0.104
	Ar → Rw	0.088	0.095	0.086	0.086	0.076	0.086	**0.068**
	Cl → Ar	0.238	0.249	0.200	0.194	0.212	0.216	**0.197**
	Cl → Pr	0.159	0.168	**0.105**	0.113	0.131	0.125	0.115
	Cl → Rw	0.151	0.159	0.113	0.113	0.125	0.115	**0.106**
	Pr → Ar	0.237	0.246	0.185	**0.184**	0.210	0.214	0.190
	Pr → Cl	0.365	0.375	0.339	0.315	0.327	0.293	**0.293**
	Pr → Rw	0.095	0.099	0.080	0.082	0.084	0.091	**0.073**
	Rw → Ar	0.138	0.147	0.129	0.125	0.126	0.154	**0.118**
	Rw → Cl	0.314	0.325	0.298	0.284	0.275	0.248	**0.238**
	Rw → Pr	0.073	0.078	0.062	0.063	0.065	0.078	**0.059**
	Avg.	0.192	0.200	0.166	0.162	0.169	0.168	**0.151**
VisDA-2017	T → V	0.274	0.284	0.261	0.202	0.204	0.172	**0.162**

Ding, Y., Liu, J., Xiong, J., and Shi, Y. Revisiting the evaluation of uncertainty estimation and its application to explore model complexity-uncertainty trade-off. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 4–5, 2020.

Evaluation for CoWA-JMDS

- CoWA-JMDS achieved the best SFUDA performance for all three datasets.

Table 2. Accuracy (%) on Office-31 dataset for UDA and SFUDA methods (ResNet-50).

Task	Method	A→D	A→W	D→A	D→W	W→A	W→D	Avg.
SFUDA	SFIT (Hou & Zheng, 2021)	89.9	91.8	73.9	98.7	72.0	99.9	87.7
	SHOT (Liang et al., 2020a)	94.0	90.1	74.7	98.4	74.3	99.9	88.6
	3C-GAN (Li et al., 2020)	92.7	93.7	75.3	98.5	**77.8**	99.8	89.6
	NRC (Yang et al., 2021)	**96.0**	90.8	75.3	**99.0**	75.0	**100.0**	89.4
	CoWA-JMDS (w/o weight Mixup)	93.7	93.5	75.5	98.0	76.8	99.8	89.6
	CoWA-JMDS	94.4	**95.2**	**76.2**	98.5	77.6	99.8	**90.3**
UDA	ResNet (He et al., 2016)	68.9	68.4	62.5	96.7	60.7	99.3	76.1
	CAN (Kang et al., 2019)	95.0	94.5	78.0	99.1	77.0	99.8	90.6
	RSDA-MSTN (Gu et al., 2020)	95.8	96.1	77.4	99.3	78.9	100	91.1
	FixBi (Na et al., 2020)	95.0	96.1	78.7	99.3	79.4	100	91.4

Table 3. Accuracy (%) on Office-Home for UDA and source-free UDA methods (ResNet-50).

Task	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
SFUDA	BAIT (Yang et al., 2020)	57.4	77.5	**82.4**	68.0	77.2	75.1	67.1	55.5	81.9	**73.9**	59.5	84.2	71.6
	SHOT (Liang et al., 2020a)	57.1	78.1	81.5	68.0	78.2	78.1	67.4	54.9	82.2	73.3	58.8	84.3	71.8
	NRC (Yang et al., 2021)	**57.7**	**80.3**	82.0	68.1	79.8	78.6	65.3	56.4	**83.0**	71.0	58.6	**85.6**	72.2
	CoWA-JMDS (w/o weight Mixup)	56.4	78.6	80.3	68.8	79.7	78.7	**68.1**	56.8	82.0	73.4	59.1	83.9	72.2
	CoWA-JMDS	56.9	78.4	81.0	**69.1**	**80.0**	**79.9**	67.7	**57.2**	82.4	72.8	**60.5**	84.5	**72.5**
UDA	ResNet-50 (He et al., 2016)	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
	RSDA-MSTN (Gu et al., 2020)	53.2	77.7	81.3	66.4	74.0	76.5	67.9	53.0	82.0	75.8	57.8	85.4	70.9
	FixBi (Na et al., 2020)	58.1	77.3	80.4	67.7	79.5	78.1	65.8	57.9	81.7	76.4	62.9	86.7	72.7

Table 4. Accuracy (%) on VisDA-2017 for UDA and source-free UDA methods (ResNet-101).

Task	Method	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Average
SFUDA	SFIT (Hou & Zheng, 2021)	94.3	79.0	**84.9**	63.6	92.6	92.0	88.4	79.1	92.2	79.8	87.6	43.0	81.4
	3C-GAN (Li et al., 2020)	94.8	73.4	68.8	**74.8**	93.1	95.4	88.6	84.7	89.1	84.7	83.5	48.1	81.6
	SHOT (Liang et al., 2020a)	94.3	88.5	80.1	57.3	93.1	94.9	80.7	80.3	91.5	89.1	86.3	58.2	82.9
	NRC (Yang et al., 2021)	**96.8**	**91.3**	82.4	62.4	96.2	95.9	86.1	80.6	**94.8**	**94.1**	**90.4**	**59.7**	85.9
	CoWA-JMDS (w/o weight Mixup)	96.3	88.5	84.1	59.7	95.2	96.9	82.1	82.3	93.3	92.8	87.5	51.1	84.2
	CoWA-JMDS	96.2	89.7	83.9	73.8	**96.4**	**97.4**	**89.3**	**86.8**	94.6	92.1	88.7	53.8	**86.9**
UDA	ResNet-101 (He et al., 2016)	72.3	6.1	63.4	91.7	52.7	7.9	80.1	5.6	90.1	18.5	78.1	25.9	49.4
	CAN (Kang et al., 2019)	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
	FixBi (Na et al., 2020)	96.1	87.8	90.5	90.3	96.8	95.3	92.8	88.7	97.2	94.2	90.9	25.7	87.2

Thank you!

- **TL;DR**
  - Propose a novel confidence score and an effective SFUDA method based on it

- **Summary**
  - Propose a novel confidence score for SFUDA
  - → The only confidence score which contains the source and target domain knowledge together
  - Propose an effective SFUDA method based on the proposed confidence score
  - Achieve **state-of-the-art performances in SFUDA** on three benchmark datasets

- More details can be found:
  - Paper: https://arxiv.org/abs/2206.06640
  - Code: https://github.com/Jhyun17/CoWA-JMDS