COT: Unsupervised Domain Adaptation with Clustering and Optimal Transport

Yang Liu [1] * Zhipeng Zhou [1] * Baigui Sun [1] †
[1]Alibaba Group

Abstract

Unsupervised domain adaptation (UDA) aims to transfer the knowledge from a labeled source domain to an unlabeled target domain. Typically, to guarantee desirable knowledge transfer, aligning the distribution between source and target domain from a global perspective is widely adopted in UDA. Recent researchers further point out the importance of local-level alignment and propose to construct instance-pair alignment by leveraging on Optimal Transport (OT) theory. However, existing OT-based UDA approaches are limited to handling class imbalance challenges and introduce a heavy computation overhead when considering a large-scale training situation. To cope with two aforementioned issues, we propose a Clustering-based Optimal Transport (COT) algorithm, which formulates the alignment procedure as an Optimal Transport problem and constructs a mapping between clustering centers in the source and target domain via an end-to-end manner. With this alignment on clustering centers, our COT eliminates the negative effect caused by class imbalance and reduces the computation cost simultaneously. Empirically, our COT achieves state-of-the-art performance on several authoritative benchmark datasets.

Method & Experiments

Algorithm 1 Clustering-based Optimal Transport

Set number of epochs for training as E, learnable clusters for target domain as $\{w_u^t\}_{u=1}^{|\mathcal{Y}|}$, classifiers/clusters for source domain as $\{w_v^s\}_{v=1}^{|\mathcal{Y}|}$;
for k-th training epoch while $k \leq E$ **do**
  for t-th iteration in k-th epoch **do**
    Take mini-batch of samples from source and target domain as input for feature extractor CNNs with parameters θ, the output features are $\{x_i^s\}_{i=1}^b$ and $\{x_j^t\}_{j=1}^b$;
    Compute the L_{cluster} for $\{x_j^t\}_{j=1}^b$, $L_{\text{cross-entropy}}$ for $\{x_i^s\}_{i=1}^b$ in the l-th batch, and $\mathcal{L}_{OT}$;
    if $1 \leq t \leq k/(b*10)$ **then**
      we find the current optimal map from clusters of source domain to those of target domain by maximizing L_{OT};
    end if
    Minimize $\mathcal{L}_{COT}$;
  end for
end for

Table 1. Accuracy (%) on Office-31 for UDA (ResNet-50).The best result is in bold.

	Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg
Common UDA	ADDA [46]	86.2	96.2	98.4	77.8	69.5	68.9	82.9
	JAN [34]	85.4	97.4	99.8	84.7	68.6	70.0	84.3
	MCD [42]	88.6	98.5	**100.0**	92.2	69.5	69.7	86.5
	GTA [43]	89.5	97.9	99.8	87.7	72.8	71.4	86.5
	CDAN [33]	94.1	98.6	**100.0**	92.9	71.0	69.3	87.7
	TAT [32]	92.5	99.3	**100.0**	93.2	73.1	72.1	88.4
	MDD [25]	94.5	98.4	**100.0**	93.5	74.6	72.2	88.9
	GSP [19]	92.9	98.7	99.8	94.5	75.9	74.9	89.5
	DANN [1]	82.0	96.9	99.1	79.7	68.2	67.4	82.2
	SHOT [29]	94.0	90.1	74.7	98.4	74.3	99.9	88.6
	MCC [21]	95.5	98.6	**100.0**	94.4	72.9	74.9	89.4
	GVB-GD [9]	94.8	98.7	**100.0**	95.0	73.4	73.7	89.3
	TSA [28]	96.0	98.7	**100.0**	95.4	**76.7**	76.8	90.6
	SRDC [45]	95.7	**99.2**	**100.0**	95.8	**76.7**	77.1	90.8
OT-based	JDOT [5]	84.7	97.8	**100.0**	86.4	64.4	67.7	83.5
	DeepJDOT [11]	88.9	98.5	99.6	88.2	72.1	70.1	86.2
	MLOT [23]	92.8	98.5	**100.0**	90.8	72.8	71.6	87.8
	RWOT [51]	95.1	99.5	**100.0**	94.5	77.5	77.9	90.8
	DANN [1] + MMI [28]	95.2	98.6	100.0	94.4	74.6	75.2	89.7
	DANN + MMI + COT (Ours)	**96.5**	99.1	**100.0**	**96.1**	**76.7**	**77.4**	**91.0**

Experiments

Table 3. Accuracy (%) on VisDA-2017 for UDA (ResNet-101). The best result is in bold.

	Method	plane	bcybl	bus	car	horse	knife	mcyle	persn	plant	sktb	train	truck	mean
Common UDA	DANN [1]	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
	MinEnt [18]	87.4	55.0	75.3	63.8	87.4	43.6	89.3	72.5	82.9	78.6	85.6	27.4	70.7
	TSA [28]	93.0	77.8	82.2	50.8	89.9	28.0	77.1	70.0	85.2	80.0	86.1	43.0	71.9
	BSP [4]	92.2	72.5	83.8	47.5	87.0	54.0	86.8	72.4	80.6	66.9	84.5	37.1	72.1
	MCC [21]	90.4	79.8	72.3	55.1	90.5	86.8	86.6	80.0	94.2	76.9	90.0	49.6	79.4
	MODEL [27]	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
	STAR [35]	95.0	84.0	84.6	73.0	91.6	91.8	85.9	78.4	94.4	84.7	87.0	42.2	82.7
	BNM [8]	89.6	61.5	76.9	55.0	89.3	69.1	81.3	65.5	90.0	47.3	89.1	30.1	70.4
	MSTN+DSBN [3]	94.7	86.7	76.0	72.0	95.2	75.1	87.9	81.3	91.1	68.9	88.3	45.5	80.2
	CGDM [14]	92.8	85.1	76.3	64.5	91.0	93.2	81.3	79.3	92.4	83.0	85.6	44.8	80.8
	SHOT [29]	94.3	88.5	80.1	57.3	93.1	93.1	80.7	80.3	91.5	89.1	86.3	58.2	82.9
	TVT [55]	92.9	85.6	77.5	60.5	93.6	98.2	89.4	76.4	93.6	92.0	91.7	55.7	83.9
	CDTrans [52]	**98.0**	86.9	**87.9**	80.9	97.9	**97.3**	**96.8**	85.3	97.6	83.2	**94.0**	54.4	88.4
OT-based	JDOT [5]	78.4	70.8	79.4	68.8	82.3	80.5	84.2	70.7	88.4	68.8	78.4	45.7	74.7
	DeepJDOT [11]	85.4	73.4	77.3	87.3	84.1	64.7	91.5	79.3	91.9	44.4	88.5	61.8	77.4
	MLOT [23]	88.2	70.4	77.3	50.2	84.8	77.2	80.4	74.4	83.8	68.2	82.3	38.7	73.0
	RWOT [51]	95.1	80.3	83.7	**90.0**	92.4	68.0	92.5	82.2	87.9	78.4	90.4	**68.2**	84.0
	DANN+MMI+COT (Ours)	96.9	**89.6**	84.2	74.1	96.4	96.5	88.6	82.0	96.0	**94.1**	85.1	62.1	87.1
	Ours + CDTrans [52]	**98.2**	89.4	87.6	82.3	**98.0**	97.2	96.4	**86.2**	**98.3**	92.6	92.2	58.1	**89.7**