Heuristic Domain Adaptation

Shuhao Cui[1,2] Xuan Jin[3] Shuhui Wang[1] Yuan He[3] Qingming Huang[1,2,4]

1. Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS

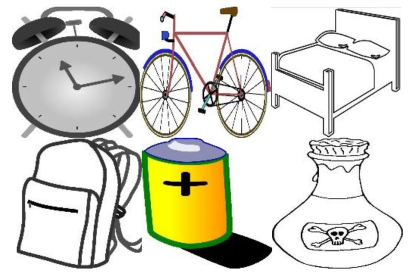
2. University of Chinese Academy of Sciences 3. Alibaba Group 4. Peng Cheng Laboratory

{cuishuhao18s, wangshuhui}@ict.ac.cn https://github.com/cuishuhao/HDA

Domain Adaptation (DA)

Source Domain

- Domain adaptation aims to transfer knowledge from source domain to target domain



Target Domain

Problem

- Direcct domain distribution alignment can hardly eliminate domain-specific properties in domain-invariant representation.
- Prior knowledge requires the flexibility in handling real-world domain adaptation problems.

Heuristic Domain Adaptation (HDA)

Mechanism

f(n) total distance **g(n) achieved distance** **h(n) estimated distance**

Heuristic search: $\mathbf{f(n)} = \mathbf{g(n)} + \mathbf{h(n)}$

Assumption

In a certain domain, the difficulty of constructing domain-invariant representations is larger than that of domain-specific representations.

Heuristic Domain Adaptation: $G(x) = F(x) - H(x)$

$G(x)$ **domain-invariant representation** $F(x)$ **total representation** $H(x)$ **domain-specific representation**

Admissible Constraint

Heuristic search is admissible when:

$$h(n) \leq h^*(n)$$

Similar constraint in Heuristic Domain Adaptation:

$$|H(n)| \leq |H^*(n)|$$

Theoretical insight:

- Error bound under F:

$$\epsilon_T(F) \leqslant \epsilon_S(F) + [\epsilon_S(F^*) + \epsilon_T(F^*)] + \boxed{|\epsilon_S(F, F^*) - \epsilon_T(F, F^*)|}$$

- H and $F - F^*$ could be regarded as positively correlated:

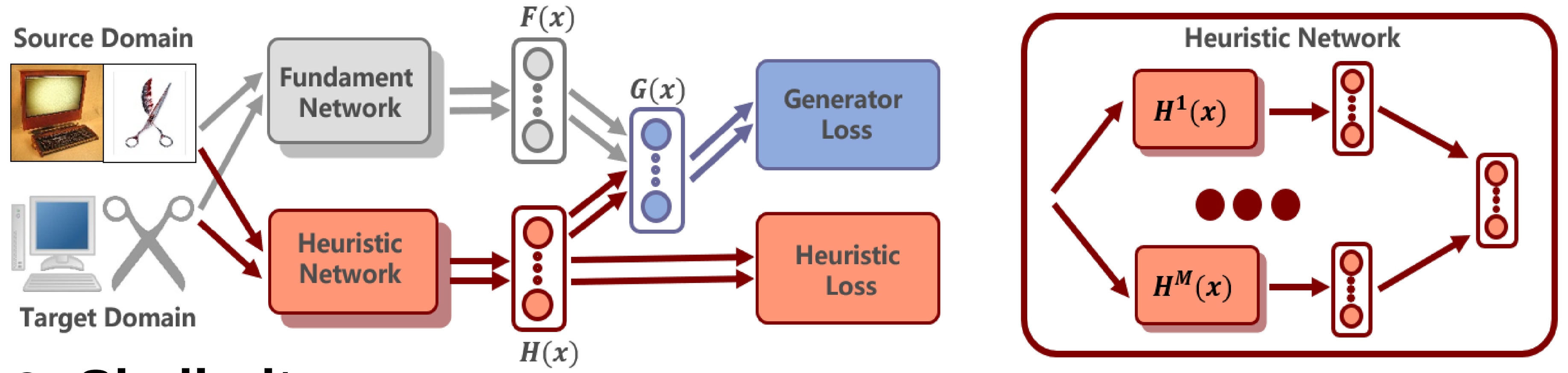
$$H = k(F - F^*) \qquad k \in (0, 1]$$

- Error bound of HDA:

$$\begin{aligned}\epsilon_T(G) &\leqslant \epsilon_S(G) + [\epsilon_S(G^*) + \epsilon_T(G^*)] + |\epsilon_S(G, G^*) - \epsilon_T(G, G^*)| \\ &\leqslant \epsilon_S(F) + [\epsilon_S(F^*) + \epsilon_T(F^*)] + \boxed{(1-k)\,|\epsilon_S(F, F^*) - \epsilon_T(F, F^*)|}\end{aligned}$$

- Lower error bound is achieved.

Heuristic Domain Adaptation Network (HDAN)



Similarity

- Similar properties exist in $G(x)$ and $H(x)$.
- To reduce domain property, the initial value: $H_{init}(x) = -G_{init}(x)$
- Similarity initialization: $F_{init}(x) = 0$

Independence

- $G(x)$ and $H(x)$ could be regarded as Blind signal separation from $F(x)$.
- $F(x)$ could obtain higher nongaussianity compared with $G(x)$ and $H(x)$.
- Nongaussianity Constraint: $kurt(F(x)) - kurt(G(x)) \approx 0$

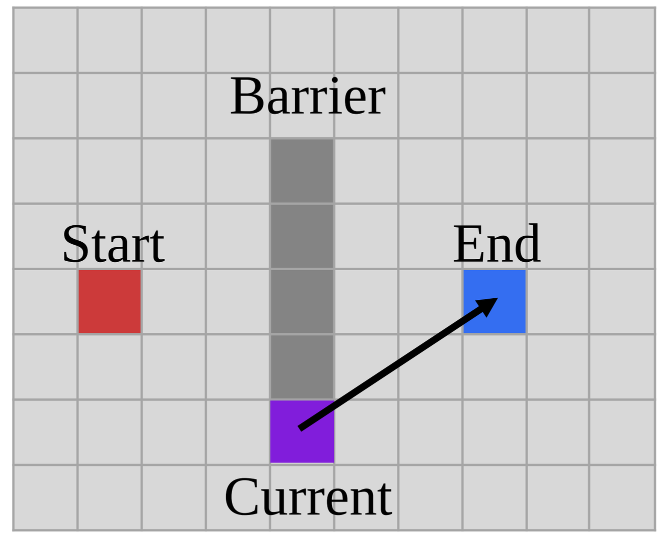
Termination

- Range of $H(x)$ is near 0 in the final:

$$|H(x)| \approx 0$$

- Considering samples with rich domain properties, $H(x)$ should be sparse:

$$\min |H(x)|_1$$



Heuristic Representations

- $H(x)$ is enhanced by multiple subnetworks with diverse initialization :

$$H(x) = \sum_{k=1}^{M} H^k(x)$$

Experiments

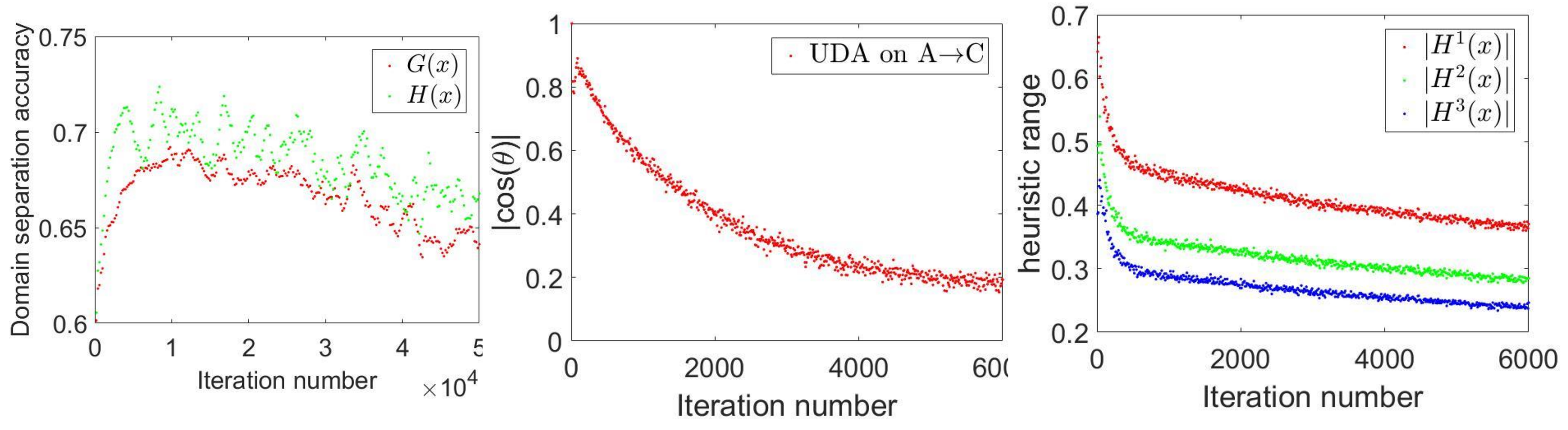
- Achieve state-of-the-art on 3 domain adaptation (DA) tasks.

Unsupervised DA results on Office-Home

Method	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P	Avg
ResNet50 [24]	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
DAN [32]	43.6	57.0	67.9	45.8	56.5	60.4	44.0	43.6	67.7	63.1	51.5	74.3	56.3
DANN [18]	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
MCD [45]	48.9	68.3	74.6	61.3	67.6	68.8	57	47.1	75.1	69.1	52.2	79.6	64.1
EntMin [52]	43.2	68.4	78.4	61.4	69.9	71.4	58.5	44.2	78.2	71.1	47.6	81.8	64.5
CDAN [33]	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
Symnets [56]	47.7	72.9	78.5	64.2	71.3	74.2	63.6	47.6	79.4	73.8	50.8	82.6	67.2
SAFN [54]	52.0	71.7	76.3	64.2	69.9	71.9	63.7	51.4	77.1	70.9	57.1	81.5	67.3
ATM [29]	52.4	72.6	78.0	61.1	72.0	72.6	59.5	52.0	79.1	73.3	58.9	83.4	67.9
BNM [11]	52.3	73.9	80.0	63.3	72.9	74.9	61.7	49.5	79.7	70.5	53.6	82.2	67.9
MDD [57]	54.9	73.7	77.8	60.0	71.4	71.8	61.2	53.6	78.1	72.5	60.2	82.3	68.1
GVBG [12]	56.5	74.0	79.2	64.2	73.0	74.1	65.2	55.9	81.2	74.2	58.2	84.0	70.0
CADA [27]	**56.9**	**76.4**	**80.7**	61.3	**75.2**	**75.2**	63.2	54.5	80.7	73.9	**61.5**	84.1	70.2
HDAN	56.8	75.2	79.8	**65.1**	73.9	**75.2**	**66.3**	**56.7**	**81.8**	**75.4**	59.7	**84.7**	**70.9**

Semi-supervised DA results on Domainnet

Models	Clipart	Infograph	Painting	Quickdraw	Real	Sketch	Avg
ResNet101 [24]	47.6±0.52	13.0±0.41	38.1±0.45	13.3±0.39	51.9±0.85	33.7±0.54	32.9±0.54
DAN [32]	45.4±0.49	12.8±0.86	36.2±0.58	15.3±0.37	48.6±0.72	34.0±0.54	32.1±0.59
RTN [34]	44.2±0.57	12.6±0.73	35.3±0.59	14.6±0.76	48.4±0.67	31.7±0.73	31.1±0.68
JAN [35]	40.9±0.43	11.1±0.61	35.4±0.50	12.1±0.67	45.8±0.59	32.3±0.63	29.6±0.57
DANN [18]	45.5±0.59	13.1±0.72	37.0±0.69	13.2±0.77	48.9±0.65	31.8±0.62	32.6±0.68
ADDA [50]	47.5±0.76	11.4±0.67	36.7±0.53	14.7±0.50	49.1±0.82	33.5±0.49	32.2±0.63
SE [16]	24.7±0.32	3.9±0.47	12.7±0.35	7.1±0.46	22.8±0.51	9.1±0.49	16.1±0.43
MCD [45]	54.3±0.64	22.1±0.70	45.7±0.63	7.6±0.49	58.4±0.65	43.5±0.57	38.5±0.61
DCTN [53]	48.6±0.73	23.5±0.59	48.8±0.63	7.2±0.46	53.5±0.56	47.3±0.47	38.2±0.57
M^{3}SDA [39]	57.2±0.98	24.2±1.21	51.6±0.44	5.2±0.45	61.6±0.89	49.6±0.56	41.5±0.74
M^{3}SDA-β [39]	58.6±0.53	26.0± 0.89	52.3±0.55	6.3±0.58	62.7±0.51	49.5±0.76	42.6±0.64
ML-MSDA [30]	61.4±0.79	**26.2**±0.41	51.9±0.20	**19.1**±0.31	57.0±1.04	50.3±0.67	44.3±0.57
GVBG [12]	61.5±0.44	23.9± 0.71	54.2±0.46	16.4± 0.57	67.8±0.98	52.5±0.62	46.0±0.63
HDAN	**63.6**±0.35	25.9±0.16	**56.1**±0.38	16.6± 0.54	**69.1**±0.42	**54.3**± 0.26	**47.6**±0.40



- Domain separation accuracy of $G(x)$ is higher than $H(x)$.
- Cosine similarity is gradually reduced.
- The ranges of subnetwork outputs are different.