Randomized Transferable Machine

Wei Pengfei[1], Leong Tze Yun[1]

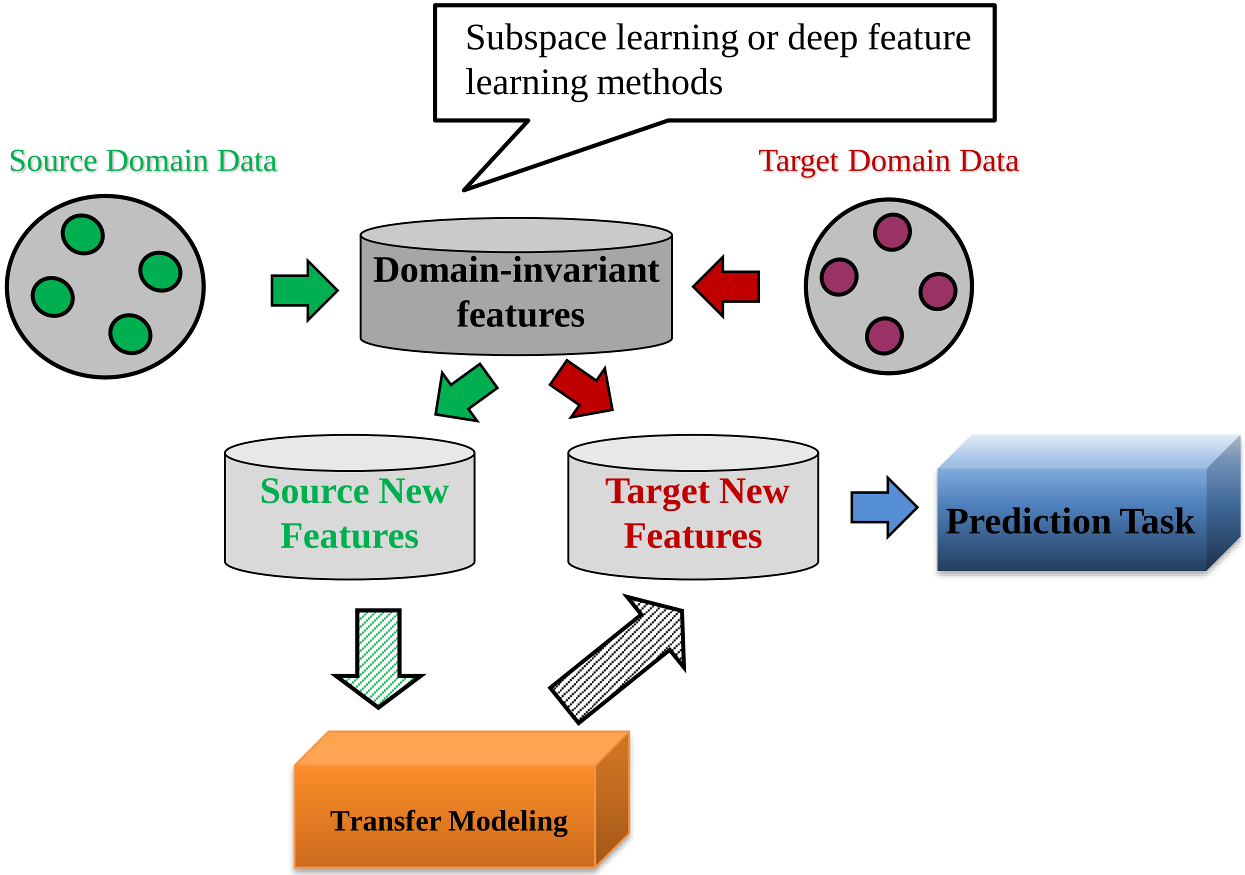
dcsweip@nus.edu.sg

School of Computing, National University of Singapore [1]

Background – Transfer Learning

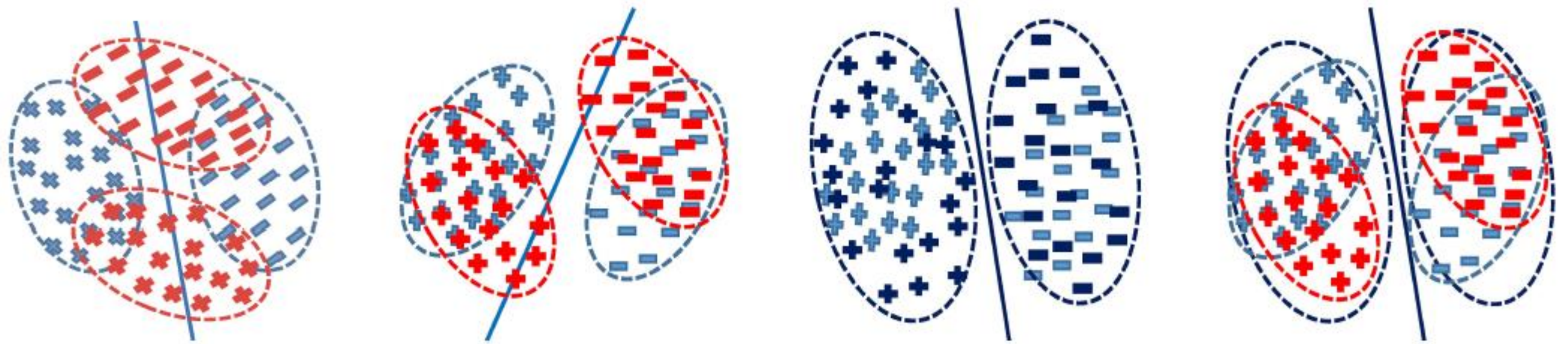
- Data scarcity
- Apply knowledge learnt from different but related domains

Background – Problem



- The truly domain-invariant feature is hard to achieve.

Key Idea



(a) Original data (b) Transfer model on new data (c) Noisy source data (d) New transfer model

- Enlarge the source training populations
- Randomly corrupt the source data using some noise
- Idea case is with infinite corruptions
- Marginalized solution to simulate the idea case without conducting any corruptions

Proposed Method

A linear logistic regression model used as the training model:

$$f_z(\mathbf{z}) = \mathbf{W}^\top \mathbf{z},$$

The Learning Objective:

$$\min_{\mathbf{W}} \quad \mathbb{E}_{\mathbf{z}^{\mathcal{S}} \sim \mathbf{Z}^{\mathcal{S}}}[||\mathbf{W}^\top \mathbf{z}^{\mathcal{S}} - \mathbf{y}^{\mathcal{S}}||_{l_2}] + \alpha||\mathbf{W}||_{l_2},$$

Assuming we have J different versions of corruptions, then we can

$$\min_{\mathbf{W}} \ \frac{1}{nJ} \sum_{i=1}^{n} \sum_{j=1}^{J} ||\mathbf{W}\tilde{\mathbf{z}}_{i,j}^{\mathcal{S}} - \mathbf{y}_i^{\mathcal{S}}||_{l_2} + \alpha||\mathbf{W}||_{l_2},$$

By using trace operator and simplifying notations, we have

$$\min_{\mathbf{W}} \ \frac{1}{nJ} tr([\mathbf{W}\widetilde{\mathbf{Z}}^{\mathcal{S}} - \widehat{\mathbf{Y}}^{\mathcal{S}}][\mathbf{W}\widetilde{\mathbf{Z}}^{\mathcal{S}} - \widehat{\mathbf{Y}}^{\mathcal{S}}]^\top) + \alpha||\mathbf{W}||_{l_2},$$

Proposed Method

With $J \rightarrow \infty$, we obtain:

$$\min_{\widetilde{f}_z} \frac{1}{n} tr(\mathbf{W}^{\mathsf{T}} \mathbb{E}_{\epsilon}[\widetilde{\mathbf{Z}}^{\mathcal{S}} (\widetilde{\mathbf{Z}}^{\mathcal{S}})^{\mathsf{T}}] \mathbf{W} - 2\mathbb{E}_{\epsilon}[\widehat{\mathbf{Y}}^{\mathcal{S}} (\widetilde{\mathbf{Z}}^{\mathcal{S}})^{\mathsf{T}}] \mathbf{W}) + \alpha ||\mathbf{W}||_{l_2}.$$

Finally, we obtain the closed-form solution as:

$$\mathbf{W} = (\mathbb{E}_{\epsilon}[\widehat{\mathbf{Y}}^{\mathcal{S}} (\widetilde{\mathbf{Z}}^{\mathcal{S}})^{\mathsf{T}}])(\mathbb{E}_{\epsilon}[\widetilde{\mathbf{Z}}^{\mathcal{S}} (\widetilde{\mathbf{Z}}^{\mathcal{S}})^{\mathsf{T}}] + \alpha \mathbf{I})^{-1}.$$

Using dropout noise, we have:

$$\mathbb{E}_{\epsilon}[\widehat{\mathbf{Y}}^{\mathcal{S}} (\widetilde{\mathbf{Z}}^{\mathcal{S}})^{\mathsf{T}}] = (1-p) \mathbf{Y}^{\mathcal{S}} (\mathbf{Z}^{\mathcal{S}})^{\mathsf{T}},$$

***Definition 1: Dropout* noise**: given a data point $\mathbf{x}$, each feature dimension of $\mathbf{x}$ is randomly corrupted by a noise ϵ that draws a Bernoulli distribution with a probability p. That is to say, each feature is corrupted to 0 with the probability p and retains with the probability $1-p$.

For $\mathbb{E}_{\epsilon}[\widetilde{\mathbf{Z}}^{\mathcal{S}} (\widetilde{\mathbf{Z}}^{\mathcal{S}})^{\mathsf{T}}]$, when $\alpha \neq \beta$,

$$[\mathbb{E}_{\epsilon}[\widetilde{\mathbf{Z}}^{\mathcal{S}} (\widetilde{\mathbf{Z}}^{\mathcal{S}})^{\mathsf{T}}]]_{\alpha,\beta} = (1-p)^2 [\mathbf{Z}^{\mathcal{S}} (\mathbf{Z}^{\mathcal{S}})^{\mathsf{T}}]_{\alpha,\beta},$$

and when $\alpha = \beta$,

$$[\mathbb{E}_{\epsilon}[\widetilde{\mathbf{Z}}^{\mathcal{S}} (\widetilde{\mathbf{Z}}^{\mathcal{S}})^{\mathsf{T}}]]_{\alpha,\alpha} = (1-p) [\mathbf{Z}^{\mathcal{S}} (\mathbf{Z}^{\mathcal{S}})^{\mathsf{T}}]_{\alpha,\alpha},$$

Experimental Results

- Results on New Feature Representation Learned from Subspace Methods

Task	TCA		GFK		SA		TJM		JDA		CORAL		JGSA		MMIT	
Model	$\mathcal{M}_z$	$\widetilde{\mathcal{M}}_z$	$\mathcal{M}_z$	$\widetilde{\mathcal{M}}_z$	$\mathcal{M}_z$	$\widetilde{\mathcal{M}}_z$	$\mathcal{M}_z$	$\widetilde{\mathcal{M}}_z$	$\mathcal{M}_z$	$\widetilde{\mathcal{M}}_z$	$\mathcal{M}_z$	$\widetilde{\mathcal{M}}_z$	$\mathcal{M}_z$	$\widetilde{\mathcal{M}}_z$	$\mathcal{M}_z$	$\widetilde{\mathcal{M}}_z$
C-A	53.44	53.65	55.32	54.07	28.29	46.24	51.77	51.77	49.27	49.69	35.07	54.18	51.67	54.28	51.98	52.82
C-W	50.51	51.86	49.49	51.19	23.39	37.63	49.49	49.49	48.14	50.17	31.19	48.14	45.42	49.49	49.15	50.85
C-D	47.77	49.04	43.95	49.04	21.66	45.22	41.40	42.04	47.13	49.68	31.85	46.50	40.76	45.86	47.13	48.41
A-C	42.83	43.54	41.94	44.52	26.54	39.63	44.97	44.70	38.91	42.39	28.50	43.72	42.65	44.17	42.56	43.81
A-W	38.98	41.02	40.68	43.73	16.27	36.61	45.08	45.76	46.78	48.81	25.42	41.69	35.59	41.36	40.00	41.02
A-D	40.13	43.31	38.85	48.41	23.57	39.49	43.95	44.59	48.41	48.41	28.66	43.31	37.58	42.68	41.40	43.31
W-C	36.69	37.22	36.69	36.87	24.49	34.28	39.36	39.09	32.32	33.30	30.01	35.80	34.55	35.17	38.11	37.13
W-A	40.29	40.50	37.79	40.81	19.73	37.47	39.98	40.19	36.12	35.91	33.92	39.87	38.31	39.46	40.29	40.40
W-D	77.71	78.98	74.52	78.34	41.40	65.61	70.06	70.70	74.52	75.80	77.71	84.08	82.80	83.44	76.43	78.98
D-C	35.62	36.60	30.45	36.78	20.66	35.89	33.13	33.93	32.50	32.32	29.74	35.53	32.95	33.48	35.71	34.73
D-A	38.83	38.83	40.40	41.02	26.62	36.01	35.18	35.07	35.59	35.39	33.72	38.10	35.91	36.12	39.35	38.10
D-W	80.34	82.03	73.90	80.00	37.29	64.75	74.58	76.27	77.63	77.63	80.00	85.08	84.07	84.07	79.32	79.66
Mean	48.60	49.72	47.00	50.40	25.82	43.24	47.41	47.80	47.28	48.29	38.82	49.67	46.86	49.13	48.45	49.10
B-D	77.64	78.49	79.64	79.49	71.99	76.74	76.89	77.19	76.39	76.54	76.04	79.49	78.69	78.84	78.49	79.34
B-E	76.48	76.48	78.23	78.28	67.67	75.88	79.73	79.58	78.33	78.38	72.52	79.03	77.03	77.28	75.48	75.38
B-K	78.19	78.14	79.24	80.64	60.68	76.64	79.09	79.14	78.84	80.04	74.04	80.84	79.39	79.39	76.99	77.74
D-B	78.05	78.60	78.05	79.20	66.50	75.70	78.05	77.60	76.60	76.45	75.90	79.90	78.80	79.85	77.80	78.30
D-E	78.18	78.13	77.63	80.08	70.87	77.48	79.53	79.73	80.48	80.48	69.57	78.98	76.68	77.38	77.58	76.98
D-K	78.44	78.29	78.84	80.64	65.93	79.19	79.54	79.29	79.49	79.74	72.54	81.34	79.44	80.29	78.79	78.84
E-B	73.75	73.85	73.70	75.05	58.75	71.30	72.55	73.65	73.60	73.60	70.25	75.55	74.45	74.65	72.75	73.35
E-D	74.84	74.99	74.34	76.74	62.68	74.54	75.49	75.69	75.64	75.44	69.88	74.94	74.09	74.34	74.39	74.84
E-K	81.79	81.84	82.74	83.59	76.29	82.84	81.64	81.79	81.29	81.39	81.24	84.94	84.79	85.04	82.79	83.04
K-B	73.45	73.45	75.60	76.35	64.35	71.80	71.85	71.45	73.05	73.85	71.15	76.25	75.75	75.70	73.50	74.20
K-D	74.44	74.39	73.89	76.14	73.34	74.79	72.99	72.54	73.19	73.99	71.79	76.09	74.39	74.34	74.29	75.29
K-E	80.68	81.23	81.68	82.13	78.63	81.68	77.23	77.38	79.18	79.63	81.23	83.33	82.33	83.18	81.03	81.33
Mean	77.16	77.32	77.80	79.03	68.14	76.55	77.05	77.08	77.17	77.46	73.85	79.22	78.07	78.36	76.99	77.38

RTM is superior to the conventional transfer model on various transfer tasks.

Thank You