Revealing Distribution Discrepancy by Sampling Transfer in Unlabeled Data

Zhilin Zhao

Background

- Labeled Training Dataset: $\widehat{\mathcal{P}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N} \overset{\text{IID}}{\sim} \mathcal{P}$
- Selected Hypothesis: $\widehat{h}_{\widehat{\mathcal{P}}} \in \arg\min_{h \in \mathcal{H}} \frac{1}{|\widehat{\mathcal{P}}|} \sum_{(\mathbf{x},y) \in \widehat{\mathcal{P}}} \mathfrak{L}(h(\mathbf{x}), y)$
- Unlabeled Test Dataset: $\widehat{\mathcal{Q}} = \{\mathbf{x}_i\}_{i=1}^{N} \overset{\text{IID}}{\sim} \mathcal{Q}_{\mathcal{X}}$
- Real Situation: Non-IID Data $\mathcal{P} \neq \mathcal{Q}$

1. Out-of-distribution Detection
2. Domain Adaptation
3. Transfer Learning
4. Data Augmentation

Research Issue

- Question: How to quantify the applicability of a hypothesis derived from training samples to test samples?
- Challenge: The labels of test samples are unaccessible.
- Contribution: Evaluate the distribution discrepancy between training and test samples without accessing the test labels.

  — Small Distribution Discrepancy: The hypothesis is applicable to the test samples.
  — Large Distribution Discrepancy: The hypothesis is not applicable to the test samples.

I-Divergence

- Expected Risk:

$$\epsilon_{\mathcal{P}}(h) = \int_{\mathcal{Y}} \int_{\mathcal{X}} \mathfrak{L}(h(\mathbf{x}), y) \mathcal{P}(\mathbf{x}, y) \, d\mathbf{x} dy$$

- Distribution Discrepancy: the difference between the expected risks of the hypothesis

$$d(\mathcal{P}, \mathcal{Q} \mid \widehat{h}_{\widehat{\mathcal{P}}}) = \left| \epsilon_{\mathcal{P}}(\widehat{h}_{\widehat{\mathcal{P}}}) - \epsilon_{\mathcal{Q}}(\widehat{h}_{\widehat{\mathcal{P}}}) \right|.$$

- Challenge: Cannot sampling test samples with labels from $\mathcal{Q}(\mathbf{x}, y)$
- Insight: Transfer the sampling patterns to the training distribution $\mathcal{P}(\mathbf{x}, y)$

$$\mathcal{Q}(\mathbf{x}, y) = \underbrace{\frac{\mathcal{Q}(\mathbf{x})}{\mathcal{P}(\mathbf{x})}}_{\text{Density Ratio:} r(\mathbf{x})} \cdot \underbrace{\frac{\mathcal{Q}(y \mid \mathbf{x})}{\mathcal{P}(y \mid \mathbf{x})}}_{\text{Likelihood Ratio:} v(\mathbf{x}, y)} \cdot \mathcal{P}(\mathbf{x}, y),$$

I-Divergence

Density Ratio

- Model the Density Ratio by a network:

$$r(\mathbf{x}) = \mathcal{Q}(\mathbf{x})/\mathcal{P}(\mathbf{x}) \in \mathcal{R}(\widehat{h}_{\widehat{\mathcal{P}}})$$

- Estimated distributions: use the density ratio r to estimate distributions

$$\widetilde{\mathcal{P}}(\mathbf{x}) = \mathcal{Q}(\mathbf{x})/r(\mathbf{x}), \quad \widetilde{\mathcal{Q}}(\mathbf{x}) = \mathcal{P}(\mathbf{x}) \cdot r(\mathbf{x})$$

- Objective Function: Minimize the KL divergence between the actual distributions and the estimated distributions

$$\min_{r \in \mathcal{R}(\widehat{h}_{\widehat{\mathcal{P}}})} \quad \mathsf{KL}\left(\mathcal{P}(\mathbf{x}) \parallel \widetilde{\mathcal{P}}(\mathbf{x})\right) + \mathsf{KL}\left(\mathcal{Q}(\mathbf{x}) \parallel \widetilde{\mathcal{Q}}(\mathbf{x})\right),$$
$$\text{s.t.} \quad \int \widetilde{\mathcal{P}}(\mathbf{x})\, d\mathbf{x} = 1, \int \widetilde{\mathcal{Q}}(\mathbf{x})\, d\mathbf{x} = 1.$$

- Derive the Likelihood Ratio by minimizing the generalization error bound
- Generalization Error Bound of the distribution discrepancy

$$B_{\mathfrak{L}}\sqrt{\frac{\ln(2/\delta)\sum_{(\mathbf{x},y)\sim\widehat{\mathcal{P}}}|\widehat{r}(\mathbf{x})v(\mathbf{x},y)-1|^2}{N}}+B_{\mathfrak{L}}\mathbb{E}_{\mathcal{P}}\left[v(\mathbf{x},y)\right]\sqrt{\frac{\mathfrak{B}(\delta/2,N)}{\mu}}$$

- Objective Function:

$$\min_{v\in\mathcal{V}(\widehat{r})} \quad \widehat{\mathbb{E}}_{\widehat{\mathcal{P}}}\left[v(\mathbf{x},y)+\frac{\gamma}{2}(\widehat{r}(\mathbf{x})v(\mathbf{x},y)-1)^2\right],$$

$$\text{s.t.} \quad \widehat{\mathbb{E}}_{\widehat{\mathcal{P}}}[\widehat{r}(\mathbf{x})v(\mathbf{x},y)]=1,$$

I-Divergence

Hypothesis Applicability Evaluation

Using the estimated Density Ratio r and Likelihood Ratio v, we can estimate the distribution discrepancy between $\mathcal{P}$ and $\mathcal{Q}$ by

$$\widehat{d}\left(\widehat{\mathcal{P}}, \widehat{\mathcal{Q}} \mid \widehat{h}_{\widehat{\mathcal{P}}}, \widehat{r}, \widehat{v}\right) = \widehat{\mathbb{E}}_{\widehat{\mathcal{P}}}\left[\left|\widehat{r}\,(\mathbf{x})\,\widehat{v}(\mathbf{x}, y) - 1\right| \mathfrak{L}(\widehat{h}_{\widehat{\mathcal{P}}}(\mathbf{x}), y)\right]$$

- Metric: AUROC
- Positive: distribution discrepancy within the same distribution $\mathcal{P}$

$$\widehat{d}\left(\widehat{\mathcal{P}}, \widetilde{\mathcal{P}} \mid \widehat{h}_{\widehat{\mathcal{P}}}, \widehat{r}, \widehat{v}\right), \widehat{\mathcal{P}} \overset{\text{iid}}{\sim} \mathcal{P}, \widetilde{\mathcal{P}} \overset{\text{iid}}{\sim} \mathcal{P}$$

- Negative: distribution discrepancy between two different distributions $\mathcal{P}$ and $\mathcal{Q}$

$$\widehat{d}\left(\widehat{\mathcal{P}}, \widehat{\mathcal{Q}} \mid \widehat{h}_{\widehat{\mathcal{P}}}, \widehat{r}, \widehat{v}\right), \widehat{\mathcal{P}} \overset{\text{iid}}{\sim} \mathcal{P}, \widehat{\mathcal{Q}} \overset{\text{iid}}{\sim} \mathcal{Q}$$

Experiments

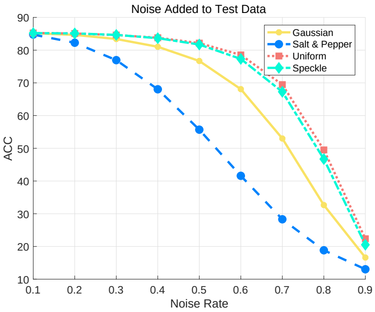
ImageNet vs Other Datasets

Table: Distribution Discrepancy Between ImageNet and Other Test Datasets.

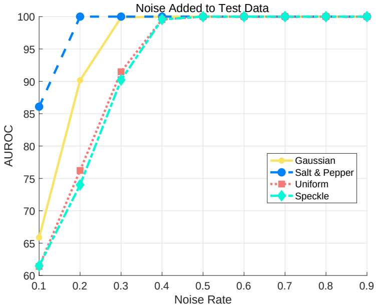
Training	Network	Test	ACC (CLIP)	MSP	NNBD	MMD-D	H-Div	R-Div	I-Div
ImageNet	ResNet50	OIDv4	43.9	100.0	91.7	94.6	100.0	94.6	**69.3**
		Caltech256	36.6	100.0	91.4	95.6	100.0	100.0	**72.4**
		Flowers102	5.1	100.0	98.6	100.0	100.0	100.0	**100.0**
		DTD	11.9	100.0	98.7	100.0	100.0	100.0	**100.0**
	ViT-B/16	OIDv4	50.6	100.0	88.6	92.6	100.0	92.6	**62.6**
		Caltech256	40.4	100.0	94.8	100.0	100.0	100.0	**71.9**
		Flowers102	5.1	100.0	98.1	100.0	100.0	100.0	**100.0**
		DTD	13.9	100.0	99.7	100.0	100.0	100.0	**100.0**

Experiments

Original Data vs Corrupted Variants



(a) Classification



(b) Distribution Discrepancy: I-Div

Figure: Distribution Discrepancy between Original Data and its Corrupted Variants with Different Noise Rate. (a) shows the classification performance of the standard network for the test datasets containing corrupted samples. (b) presents the distribution discrepancy in terms of AUROC.

Conclusions

Contributions:

- I-Div estimates the distribution discrepancy without accessing the test label.
- I-Div evaluates the applicability of a hypothesis on test data.

Future Work:

- estimate the distribution discrepancy without accessing the training data.
- evaluate the applicability of a large pretrained model.

Thank you!

Comments & Suggestions:
zhaozhl7@hotmail.com