A principled approach to model validation in domain generalization

Boyang Lyu*, Thuan Nguyen*, Matthias Scheutz*, Prakash Ishwar**, and Shuchin Aeron*

*Tufts University, **Boston University

May 7, 2023
Domain Generalization (DG)

- **Given:** labeled samples from several “seen” domains
- **Goal:** learn a classifier that can generalize well to “unseen” domains
- **Challenge:** distribution-shift

![Diagram showing seen and unseen domains with examples of sketch, cartoon, art painting, and photo styles.](image)
DG Methodology

**Key idea:**

- learn *domain-invariant* features via a representation function: \( f : \mathcal{X} \rightarrow \mathcal{Z} \)
- jointly train a classifier \( g : \mathcal{Z} \rightarrow \mathcal{Y} \) to minimize DG loss\(^1\):

\[
L_{\text{Training}}(f, g) = \beta L_{\text{Classification}}(f, g) + L_{\text{Domain-discrepancy}}(f)
\]

---

DG Methodology

1. Classification risk:

\[ L_{\text{Classification}}(f, g) = \mathbb{E}_{(x,y) \sim p^{(s)}(x,y)} \left[ \ell(g(f(x)), y) \right] \]  \hspace{1cm} (2)

- \( p^{(s)}(x, y) \): seen-domain joint distribution in input space
- \( \ell(\cdot, \cdot) \): classification loss function

2. Domain discrepancy:

\[ L_{\text{Domain-discrepancy}}(f) = d(p^{(u)}(f(x), y) \| p^{(s)}(f(x), y)) \]  \hspace{1cm} (3)

- \( p^{(u)}(f(x), y), p^{(s)}(f(x), y) \): joint distributions in representation-space
- \( d(\cdot \| \cdot) \): discrepancy measure
Fundamental Trade-off between Classification Risk and Domain Discrepancy

**Theorem**

Let

\[ T(\Delta) := \min_{f: \mathcal{X} \rightarrow \mathcal{Z}} \text{Domain-discrepancy}(f) \]

s.t.

\[ \text{Classification}(f, g) \leq \Delta \]

If \( d(a\|b) \) is a convex function of \((a, b)\), then for any classifier \( g \), \( T(\Delta) \) is

1. non-increasing and
2. convex
Training, Validation, and Testing in DG

- **Training:** minimize training loss over the training set:

\[
L_{\text{Training}}(f, g) = \beta L_{\text{Classification}}(f, g) + L_{\text{Domain-discrepancy}}(f),
\]

\[\beta\text{ controls the trade-off between classification risk and domain discrepancy}\]

- **Validation:** select hyper-parameters (models) that only minimize classification risk on validation set (domain discrepancy ignored)

\[
L_{\text{Validation}}(f, g) = L_{\text{Classification}}(f, g)
\]

---

Hyper-parameters (HPs)
- Learning rate
- Training epochs
- Batch size
- Optimizer, etc.

1\(^{\text{st}}\) group of HPs

\[\ldots\]\n
n\(^{\text{th}}\) group of HPs

Training data

\[\ldots\]\n
Validation data

Test data (unseen domain)

Model with lowest classification risk
A New Model Validation Method in DG

1. New validation loss:

\[ L_{\text{Validation}} = \beta (1 - \alpha) L_{\text{Classification}} + \alpha L_{\text{Domain-discrepancy}} \]  

(7)

- \( \alpha \): convex combination hyper-parameter
- \( \beta \): scale hyper-parameter for combining objectives with different scales

2. Practical implementation:

- Cross-entropy to approximate \( L_{\text{Classification}} \)
- Maximum Mean Discrepancy loss\(^2\) to approximate \( L_{\text{Domain-discrepancy}} \)
- \( \beta = 1 \), and \( \alpha = 0.2 \)

---

\(^2\) D. Li, Y. Yang, Y.-Z. Song, and T. M. Hospedales, “Deeper, broader and artier domain generalization,” in Proceedings of the IEEE international conference on computer vision, 2017, pp. 5542–5550
## Experimental Results

1. **Datasets**: PACS, VLCS, and C-MNIST\(^3\)

2. **Algorithms**: 12 SOTA DG algorithms in DomainBed package\(^4\)

3. **Performance metric**: Classification accuracy (percentage).

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Fish</th>
<th>IRM</th>
<th>GDRO</th>
<th>Mixup</th>
<th>CORAL</th>
<th>Algorithms</th>
<th>Wins</th>
</tr>
</thead>
<tbody>
<tr>
<td>PACS (Traditional)</td>
<td>84.6</td>
<td>84.9</td>
<td>84.2</td>
<td>83.3</td>
<td>85.1</td>
<td>83.6  84.6  86.4  84.5  85.2  83.7</td>
<td>7/12</td>
</tr>
<tr>
<td>PACS (Ours)</td>
<td>82.0</td>
<td>85.3</td>
<td>84.3</td>
<td>85.3</td>
<td>84.9</td>
<td>85.0  84.9  82.0  84.2</td>
<td>81.3</td>
</tr>
<tr>
<td>VLCS (Traditional)</td>
<td>79.4</td>
<td>76.0</td>
<td>78.1</td>
<td>77.4</td>
<td>76.8</td>
<td>78.5  77.8  79.2  77.3  76.4</td>
<td>78.6  80.5</td>
</tr>
<tr>
<td>VLCS (Ours)</td>
<td>77.5</td>
<td>79.2</td>
<td>79.6</td>
<td>77.6</td>
<td>78.8</td>
<td>78.0  78.5  80.3</td>
<td>78.2  78.6  76.1  79.3</td>
</tr>
<tr>
<td>CMNIST (Traditional)</td>
<td>10.0</td>
<td>10.0</td>
<td>10.2</td>
<td>10.4</td>
<td>9.7</td>
<td>10.4  10.0  9.9  10.5</td>
<td>10.2  10.2  10.4</td>
</tr>
<tr>
<td>CMNIST (Ours)</td>
<td>9.7</td>
<td>10.9</td>
<td>12.6</td>
<td>10.3</td>
<td>11.2</td>
<td>9.9   11.1  10.2  11.5</td>
<td>15.6  13.8  10.5</td>
</tr>
</tbody>
</table>

\(^3\)https://github.com/facebookresearch/DomainBed

Thank you for your attention!