Group Distributionally Robust Optimization For Solving Out-Of-Domain Generalization And Finding Causal Invariant Relationships

Author: Zenan Guan (Z.Guan@student.tudelft.nl)
Supervisors: Rickard Karlsson, Stephan Bongers
Responsible professor: Jesse Krijthe

INTRODUCTION

- Out-Of-Domain (OOD) Generalization: learn a model from one domain and make it perform well in unseen domain(s)

Figure 1: Cow in grassland

Figure 2: Camel in desert

Figure 3: Cow in desert

Figure 4: Camel in grassland

- Problem of Empirical Risk Minimization (ERM): exploit spurious correlation

$$\hat{\theta}_{\mathrm{ERM}} := \underset{\theta \in \Theta}{\arg\min}\ \mathbb{E}_{(x,y)\sim \hat{P}}[\ell(\theta;(x,y))]$$

- Solution: group Distributionally Robust Optimization (group DRO) [1]

$$\hat{\theta}_{\mathrm{DRO}} := \underset{\theta \in \Theta}{\arg\min}\Big\{\hat{\mathcal{R}}(\theta) := \max_{g \in \mathcal{G}} \mathbb{E}_{(x,y)\sim \hat{P}_g}[\ell(\theta;(x,y))]\Big\}$$

RESEARCH QUESTIONS

- Does group DRO perform better than the ERM method in OOD generalization in binary classification?

- Can group DRO find and exploit the invariant relationships in the training domain and learn an invariant classifier?

METHODOLOGY

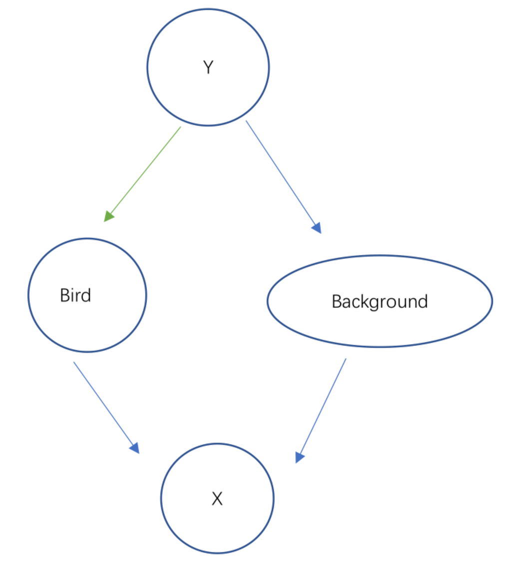

Figure 5: The diagram for data generation. Bird has the invariant features.

Figure 6: Land birds in green background

Figure 7: Land birds in blue background

Figure 8: Water birds in blue background

Figure 9: Water birds in green background

Data generation:

- CUB data set [2], divided into two classes as land birds (4615 in training split, 4510 in testing split) and water birds (1379 in training split, 1284 in testing split)
- Green and blue background
- Mark the label and background for grouping.

Experiment A:

- Classifiers trained from two training sets (weak (training set 1) and strong (training set 2) correlation between label and background) by ERM and group DRO
- Measure the performance of classifiers in different testing sets

Experiment B:

- Classifiers trained from five training sets by ERM and group DRO, different strength of correlation between label and background
- Evaluate if the classifiers use invariant features to predict, by counting number of bird i such that

$$f(bird\_i\_with\_green\_background) = f(bird\_i\_with\_blue\_background) = label(bird\_i)$$

RESULT

Experiment A

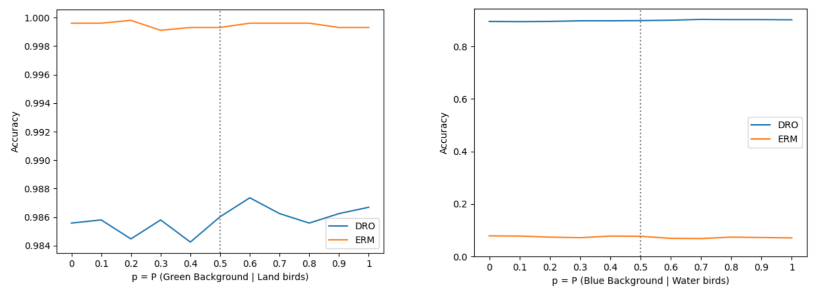

Figure 10, 11: Results of experiments A, trained from training set 1, classification result on land birds (left) and water birds (right) across testing sets

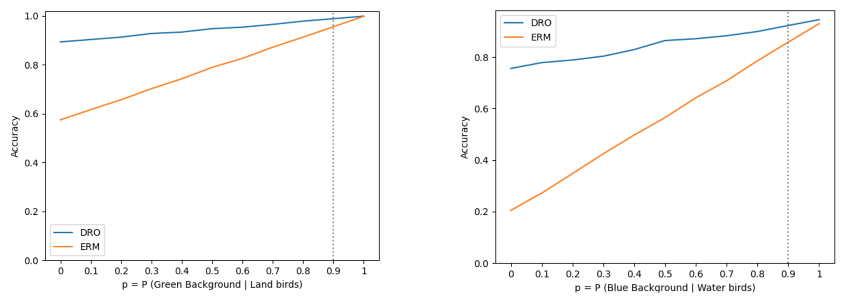

Figure 12, 13: Results of experiments A, trained from training set 2, classification result on land birds (left) and water birds (right) across testing sets

Experiment B

Training set \ Method	ERM	group DRO
p = 0.5	4600.33 (79.40%)	5548.33 (95.76%)
p = 0.6	4838.33 (83.51%)	5525.33 (95.36%)
p = 0.7	4544 (78.43%)	5474.33 (94.48%)
p = 0.8	4521 (78.03%)	5314.67 (91.73%)
p = 0.9	3980.33 (68.69%)	4882.33 (84.27%)

Figure 14: Results of experiments B. The number of birds being classified equally and correctly and the percentage in total number, p = P(green background | land birds) = P(blue background | water birds), indicating the strength of correlation between label and background.

- Group DRO classifiers are more stable, while ERM classifiers heavily biased to land birds (majority in the training set)
- Most of birds being classified correctly and equally by group DRO classifiers, but performance drops when p goes up.

CONCLUSION

- Group DRO improves the OOD performance over ERM
- Group DRO can find invariant relationships, although the ability is limited when spurious correlation is strong in the training domain

LIMITATION

- The backgrounds used in the data generation are simple, real-world backgrounds can be more complicated
- Need to mark the background, which is additional work in real world

FUTURE WORK

- Repeat the experiments on real world images
- Compare group DRO with causal inference methods like IRM [3] and RE [4]

References:
[1] Sagawa, S., Koh, P. W., Hashimoto, T. B., & Liang, P. (2020). Distributionally Robust Neural Networks for Group Shifts: On the Importance of Regularization for Worst-Case Generalization. https://doi.org/10.48550/arXiv.1911.08731
[2] Wah, C., Branson, S., Welinder, P., Perona, P., & Belongie, S. (2011). The Caltech-UCSD Birds-200-2011 Dataset. https://authors.library.caltech.edu/27452/
[3] Arjovsky, M., Bottou, L., Gulrajani, I., & Lopez-paz, D. (2020). Invariant Risk Minimization. https://doi.org/10.48550/arXiv.1907.02893
[4] Krueger, D., Caballero, E., Jacobsen, J., Zhang, A., Binas, J., Zhang, D., Priol, R. L., & Courville, A. (2020). Out-of Distribution Generalization via Risk Extrapolation (REx). https://doi.org/10.48550/arXiv.2003.00688

TU Delft