Learning to Predict Based on Causality Using an Adversarial Objective

Ryen Krusinga[1] , **David Jacobs**[2]
[1]University of Maryland
[2]University of Maryland
{krusinga, djacobs}@umiacs.umd.com

	col1	col2
row1	0.4	0.1
row2	0.1	0.4

Table 1: Training set proportions

Abstract

In a given learning environment, some variables are causal, and some are merely correlated with the prediction targets. Standard machine learning models make no distinction between the two, making the models less robust to domain shift, in which causal factors remain invariant but correlations change. Given background knowledge about the causal structure of the environment, we demonstrate a simple adversarial method to train a predictor that ignores non-causal information. We show results on two artificial environments.

This is a test citation [Oh *et al.*, 2015].

References

[Oh *et al.*, 2015] Junhyuk Oh, Xiaoxiao Guo, Honglak Lee, Richard Lewis, and Satinder Singh. Action-Conditional Video Prediction using Deep Networks in Atari Games. pages 1–9, 2015.