An Empirical Analysis of Topological Persistence as a Supplementary Measure of Dataset Drift

Christopher Shultz

The Hartford Financial Services Group

Hartford, CT, USA

c.shultz@live.com

Author Contributions: Single-author, 100% attribution.

Keywords: dataset drift, covariate shift, dataset shift, topological data analysis, TDA

MSC: 55N31, 62-08

Abstract

This paper proposes the utilization of Persistence Entropy (PE), a concept from Topological Data Analysis (TDA) as a supplementary method for detecting dataset drift. We present a brief review of dataset drift detection, and describe the potential benefits of utilizing PE to capture changes in the "shape" of the dataset over time, which are unobservable via traditional methods. Through a numerical experiment, we demonstrate that the proposed method both captures drift effectively and scales with the magnitude of the drift injected.