Sheaves as a Framework for Understanding and Interpreting Model Fit

Henry Kvinge, Brett Jefferson, Cliff Joslyn, Emilie Purvine

Motivation:

- In machine learning (ML) we usually think of datasets as sets without additional structure between individual datapoints.
- On the other hand, current benchmark datasets are often large and complex with extensive metadata that defines subpopulations with distinct dynamics. Understanding performance on subpopulations is important.
- In such a setting, it makes more sense to identify a dataset as a dataspace, with a topology defined by metadata.
- We show how (pre)sheaves on this dataspace can be used to better understand and interpret ML models.

The process

1. Put a domain-informed topology on a dataset

We think of a dataset as a set (without additional structure) with a topology defined by metadata. Each element of the dataset may be split by:
   - A set of sensors from which we have collected readings,
   - A set of genes for which we are measuring expression levels,
   - A set of ID’s for images we will be encoding with a deep network.
A dataset often comes with metadata attached to it. Elements of an image dataset may be split by:
   - Labels on objects in the image,
   - The location or time the image was taken.

Example: Fruits 360 (Muresan, 2018)

We assume that dataset $D$ has metadata that allows us to identify significant subpopulations $U_1, U_2, ..., U_k$. We use $U_1, U_2, ..., U_k$ as the subbase of a new database topology $\mathcal{J}$ on $D$.

(2) Create a data sheaf and model presheaf

For a presheaf $\mathcal{F}$ on topology $\mathcal{T}$, an assignment is a choice of sections, $\{a_U\}_{U \in \mathcal{T}}$.

Each element $f \in \mathcal{D}(U)$ corresponds to a measurement of data on $U \in \mathcal{T}$

Data sheaf $\mathcal{D}$

Model presheaf $\mathcal{M}$

Take-away:

- The data that we collect is an assignment for a sheaf, the set of functions we use to model subsets of the dataset is an assignment for a presheaf.
- The process of modeling data can be understood as a map $\Phi$ from the data sheaf to the model presheaf.

(3) Measure of model fit by closeness to being a presheaf morphism

Question: How can we use this framework to understand statistics on various subpopulations of a dataset?

One answer: Look at the extent to which the model map fails to be a presheaf morphism.

Analysis

- Biggest difference from global mean arises from $(c, d)$, researchers who are in two collaborations.
- Collaboration $U_1$ deviates more from the global mean than $U_2$.

Example (bibliometrics)