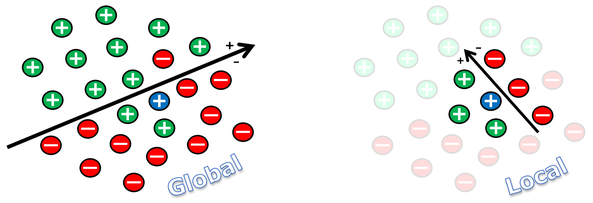
Predicting Useful Neighborhoods for Lazy Local Learning

Aron Yu Kristen Grauman
University of Texas at Austin

Lazy Local Learning



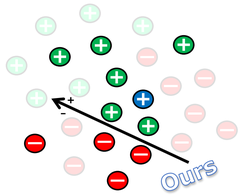
- hard for a single global model to generalize under diverse data
- local model custom tailored for each novel test point

Core Assumption: Instances most *useful* for building a local model are those that are *nearest* to the test point. [Bottou & Vapnik 92, Atkeson et al. 97, Zhang et al. 06, Geng et al. 08, Banerjee et al. 09, ...]

Our Idea

Learn the properties of a "good neighborhood" using a **large-scale multi-label classification** framework to enhance local learning.

- determine *distribution* of neighborhood
  - composition
  - size
- maintain bias towards local neighbors while not strictly adhering to the ranking

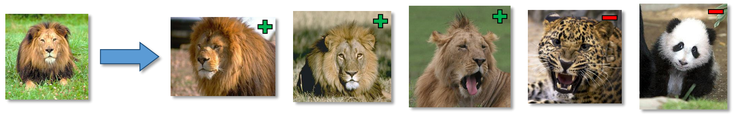


Similarity-Based (prior work)

"which images are most similar to the test image?"

Distribution-Based (ours)

"which set of images would jointly train a good classifier for the test image?"

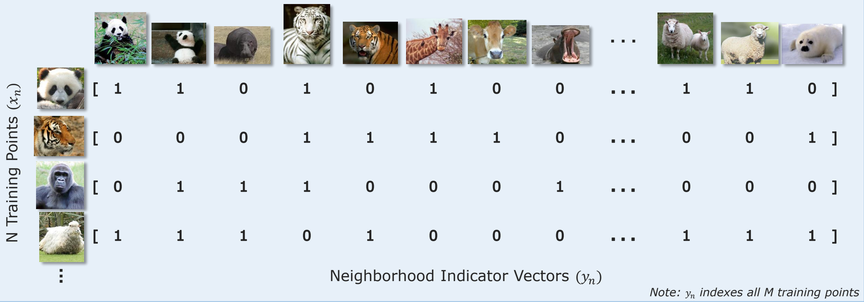


Key Idea: learn about *neighborhoods* instead of just *neighbors*

Predicting Useful Neighborhoods

Phase 1: Generate Training Neighborhoods (offline)

- empirically determine "ground truth" neighborhoods
- *weighted* and *balanced* sampling of candidate neighborhoods (K x S candidates)
- select the best point-neighborhood pairs (x_n, y_n) based on *prediction confidences*



Note: y_n indexes all M training points

Phase 2: Neighborhood Mapping with Compressed Sensing (offline)

- we cast our learning task as a large-scale multi-label classification problem
- learn a prediction function that jointly estimates all useful neighbors
- offline stages require only hours compared to days for the naïve approach

Project M-dim y_n onto D-dim space using random matrix $\phi \in \mathbb{R}^{D \times M}$

$$z_n = \phi * y_n$$
$$n = 1, \ldots, N$$

Learn D regression functions that map x_n to the z_n's

$$\{x_1, \ldots, x_N\} \quad \{z_1, \ldots, z_N\}$$
$$\{f_1, \ldots, f_D\}$$

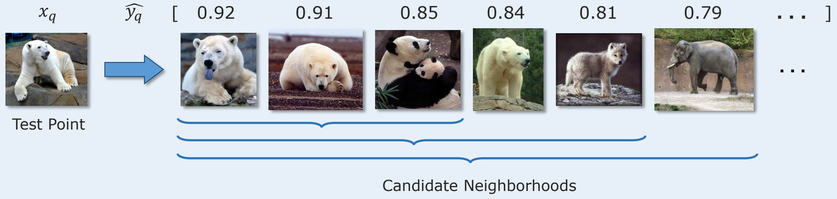
Map novel point x_q to the reduced D space and reconstruct M-dim $\widehat{y_q}$

$$[f_1(x_q), \ldots, f_D(x_q)] \Rightarrow \widehat{y_q}$$

Note: We use the approach of [Kapoor et al. 12]

Phase 3: Infer Best Neighborhoods (online)

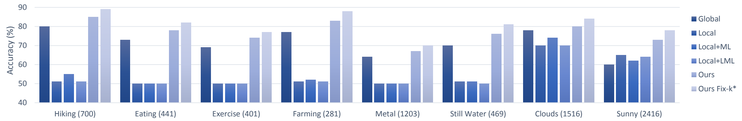
- infer real-valued indicator vector $(\widehat{y_q})$ for each test point
- form candidate neighborhoods using top K inferred points
- determine best neighborhood based on normalized *decision values*



Experimental Results

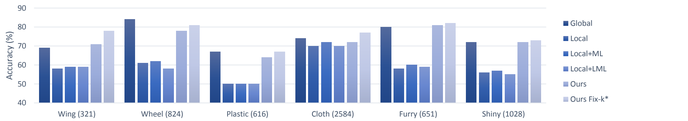
SUN Attributes Dataset: 14,340 images, 707 classes, 8 attributes

Top-5 Neighbors: Local baseline captures global similarity but not class distribution. We capture informative samples specific to the neighborhood.



Observation: We outperform all baselines on all attributes by a sizeable margin (~25%), esp. on ones with fewer positive samples.

aPascal Dataset: 6,440 images, 20 classes, 6 attributes



Observation: We outperform all Local baselines but lag behind Global in some cases due to better spatial alignment and lower visual diversity.

Conclusion

- novel lazy local learning approach that predicts the best neighborhood
- enhances learning over large + unbalanced datasets
- outperforms traditional local approaches on 2 challenging datasets