Sample Influence Analysis: A Graph Signal Processing Approach

Rushil Anirudh[†], Jayaraman J. Thiagarajan[†], Rahul Sridhar[†‡], Peer-Timo Bremer[†]

[†]Lawrence Livermore National Laboratory, [‡]University of California, Irvine

Motivation

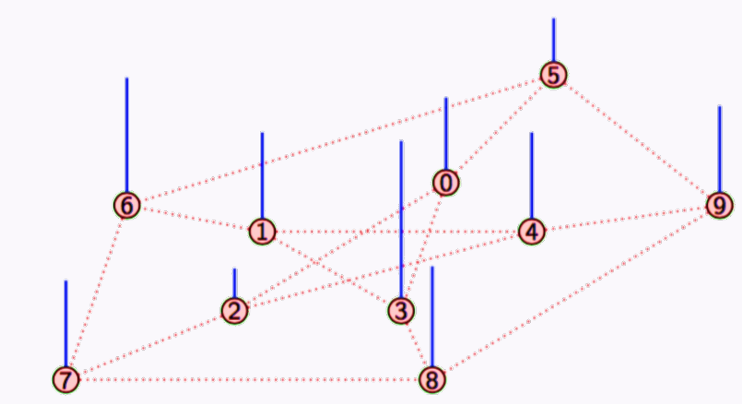
- With the growing complexity of deep neural networks, unraveling black-box models is more critical than ever
- **Example-based explanation** is a popular strategy to succinctly describe the model's observed behavior
- **Influence analysis** refers to a broad class of sample selection methods – Strongly tied to the metric used to define the desired notion of influence
- **Goals:**
  - Generic framework for different influence metric choices
  - Support model-agnostic and model-aware sample selection
  - Deal with graph-structured and manifold data
  - Fast optimization

Defining Influence Metrics

- Proposed approach is generic and supports the use of different, task-specific influence metrics
- **Case 1:** Semi-supervised label propagation on graphs
  - Input: Pre-defined graph, node attributes
  - Use the latent representation from a graph-CNN autoencoder as the influence metric
- **Case 2:** Prototypes & Criticisms of a data distribution
  - Input: Features for a set of input samples
  - For each node, remove the node and its neighbors from the data, and measure the maximum mean discrepancy (MMD) [2] from full data

Proposed Influence Analysis

- Influential sample selection can be generally posed under the framework of graph signal processing [1]



Analyze characteristics of the signal by exploiting the graph structure

Key Idea: High-frequency components of an influence metric, defined as the function on a graph with samples as nodes, can be related to the sample influences

Step 1. Graph Construction

- Approximate the domain using a neighborhood graph
- Latent features from pre-trained models can be used for model-aware analysis

Step 2. Influence Metric Design

- Metric should emphasize properties of the desired influence structure, e.g. samples on the decision surface of a classifier

Step 3. Spectral Filtering

- Graph spectral filter defined based on the graph shift operator
- Design a k-hop high-pass filter to analyze the metric

Step 4. Influence Analysis

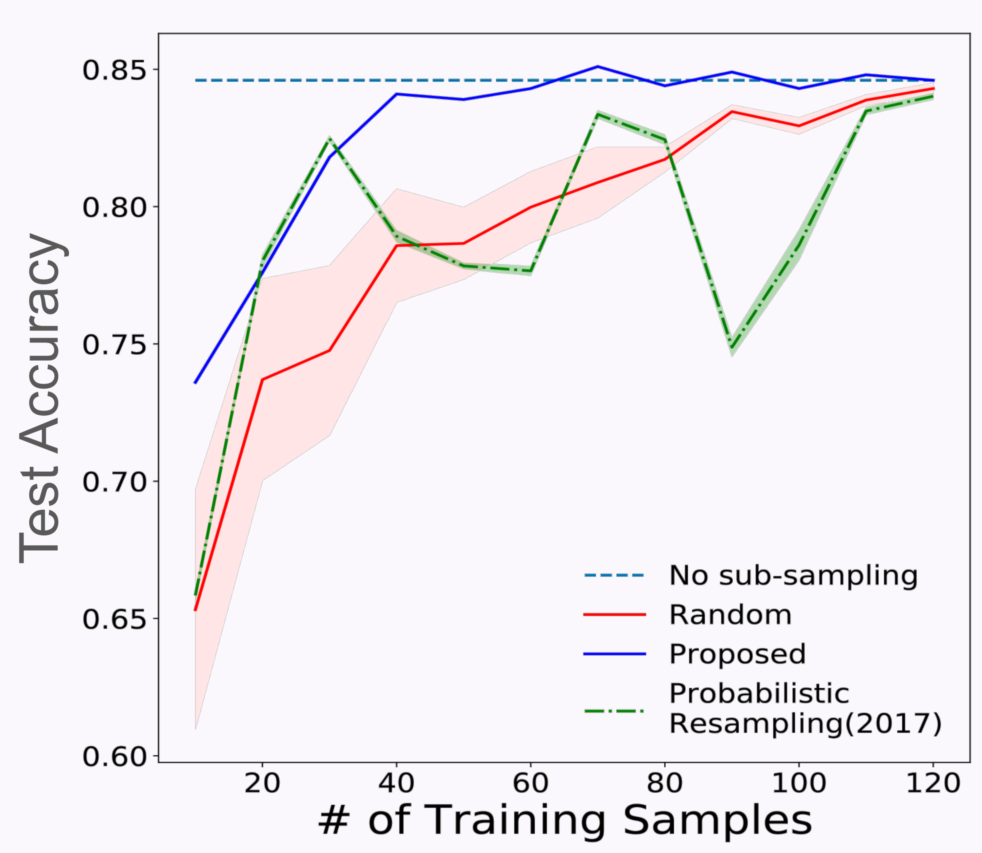
- Compute influence score based on the filtered metric
- Perform the desired analysis using samples with top influence scores e.g. study prototypes and criticisms

Label Propagation on Graphs

Experiment: Use the labels of only the most influential samples to perform semi-supervised label propagation

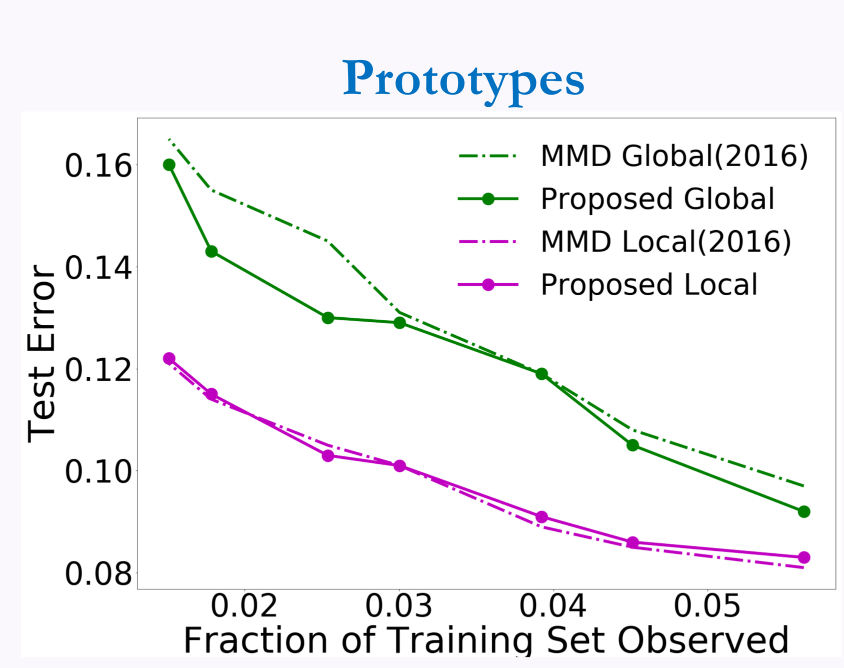
Citeseer Dataset

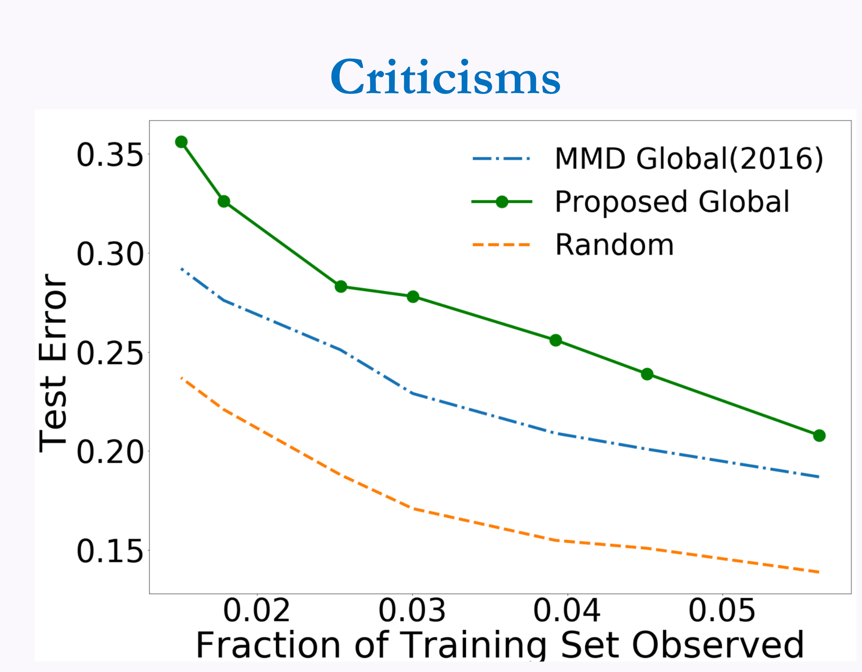
- 3372 nodes/4732 edges
- Sparse BoW attributes
- Evaluation on 1000 test samples



Prototypes & Criticisms

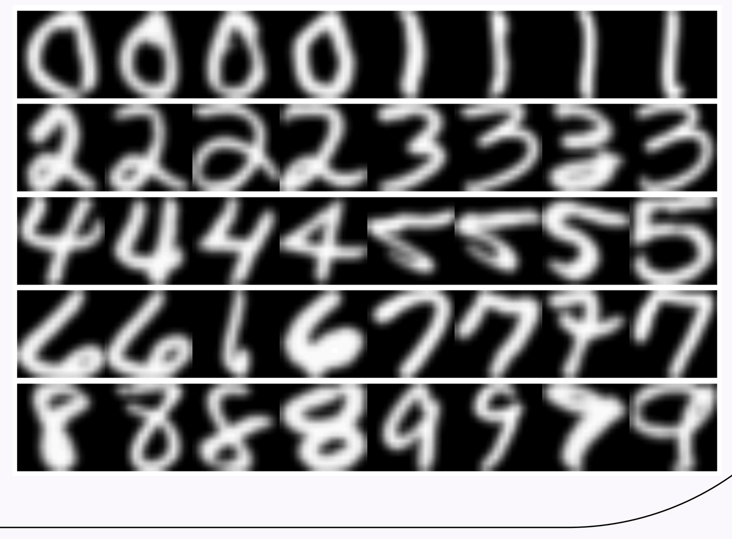
Experiment: Identify prototypical examples that can be effective for predictive modeling and criticisms that are the least useful for generalization





USPS Dataset

- 9298 images/10 classes
- Raw pixels used to build graphs
- *Global* (unsupervised) and *local* (supervised) MMD evaluation
- Test error from 1-NN classifier



References:

[1] Shuman,D.I.;Narang,S.K.;Frossard, P.; Ortega, A.; and Vandergheynst, P. 2013. The emerging field of signal processing on graphs: Extending high- dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine.

[2] Kim, B.; Khanna, R.; and Koyejo, O. O. 2016. Examples are not enough, learn to criticize! criticism for interpretability. In Advances in Neural Information Processing Systems (NIPS).

[3] Kipf, T. N., and Welling, M. 2017. Semi-supervised classification with graph convolutional networks. In International Conference on Learning Representations.

This work was performed under the auspices of the U.S. Dept. of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.