Inference Driven Metric Learning (IDML) for Graph Construction

Paramveer S. Dhillon[1], Partha Pratim Talukdar[1] and Koby Crammer[2]

[1]CIS, University of Pennsylvania, Philadelphia, PA, U.S.A. [2]Electrical Engineering, The Technion, Haifa, Israel

Graph-based Semi-Supervised Learning (SSL)

- Construct a graph from the data (usually unsupervised).
  - Each node in the graph is a data instance.
  - A pair of similar nodes are connected by an edge, with edge weight representing degree of similarity.
- Use available label information and the constructed graph to assign labels to unlabeled nodes (via Label Propagation [5] etc.).
- Works in Transductive Setting.

Can (supervised) label information be used for graph construction?

Previous Work

Recently interest has been shown in graph construction (but it is totally unsupervised).

- b-Matching [3].
  - Unsupervised and imposed degree constraints on nodes of graph,
  - Constraining each node to have same ‘b’ degree.
- Graph for vector data [1]
  - Also unsupervised and each node is required to have a min. weighted degree of 1, which is relaxed in some cases.
- k-NN Graph Construction (Standard traditional way)
  - Also, unsupervised.
  - Find k-NNs of each data instance by cosine or Gaussian similarity.
  - Not a good idea due to “curse of dimensionality”.

Our Contribution

Use supervised labeled data in addition to unlabeled data for better graph construction:

- **Learn the distance metric using label information**
  - LMNN (Large Margin Nearest Neigbors) [4].
  - ITML (Information Theoretic Metric Learning) [2].
- Construct the k-NN graph in this learned space.

Our Contribution (contd.)

- **Incorporate unlabeled data via self-training**
  - Add high-confidence (low entropy) instances with their predicted labels for the next iteration of metric learning.
    - ∗ Now, we should learn a better metric as the amount of supervision is more.
- **Using Active Learning to reduce the amount of supervision required**
  - Select high entropy predictions and label them.

Algorithm

- Our algorithm is described below:

Algorithm 1 Inference Driven Metric Learning (IDML)

Input: instances X, training labels Y, training instance indicator S, label entropy threshold β, neighborhood size k

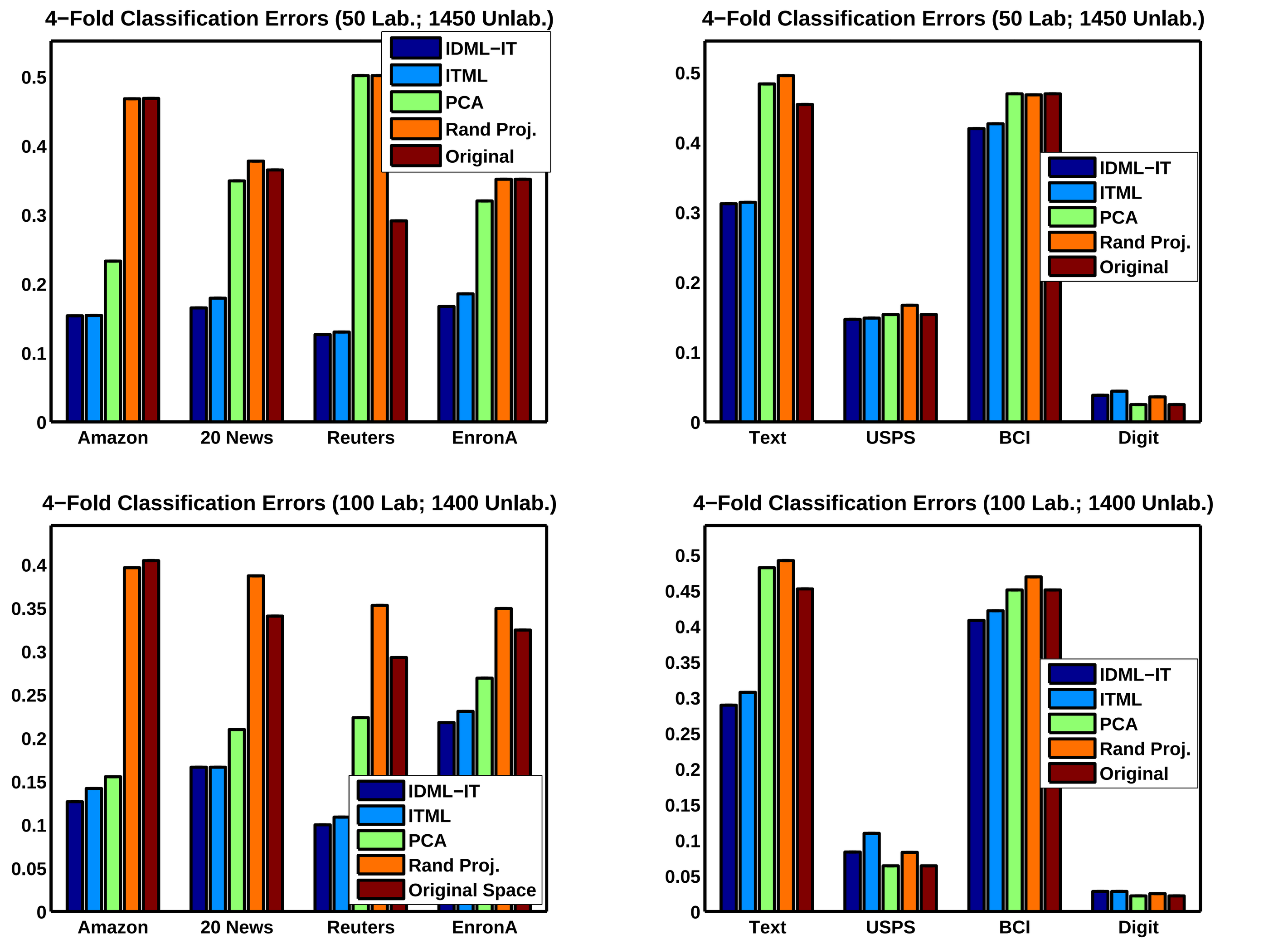
Output: Mahalanobis distance parameter A

1: $\hat{Y} \leftarrow Y$, $\hat{S} \leftarrow S$
2: **repeat**
3: $A \leftarrow \textsc{MetricLearner}(X, \hat{S}, \hat{Y})$
4: $W \leftarrow \textsc{ConstructKnnGraph}(X, A, k)$
5: $\hat{Y}^{'} \leftarrow \textsc{GraphLabelInference}(W, \hat{S}, \hat{Y})$
6: $U \leftarrow \textsc{SelectLowEntInstances}(\hat{Y}^{'}, \hat{S}, \beta)$
7: $\hat{Y} \leftarrow \hat{Y} + U\hat{Y}^{'}$
8: $\hat{S} \leftarrow \hat{S} + U$
9: **until** convergence (i.e. $U_{ii} = 0, \ \forall i$)
10: return A

Experimental Results

- Classification Accuracies on 8 Text and Sentiment Analysis Datasets
  - Total 1500 instances (nodes in graph) for all datasets; $n_{Labeled}$ =50 or 100.
- **Predictive Accuracies (unlabeled points)**
  - IDML based extensions of the methods are significantly better (13/16 cases) than following state-of-the-art methods:
    - ∗ Only “supervised” distance metric learning methods (e.g., LMNN, ITML).
    - ∗ Other Dimensionality reduction methods (e.g., RP, PCA).

Experimental Results (Contd.)



- **Active Learning**
  - IDML based active learning is better than randomly choosing instances to label. (Plots are in the paper.)

Summary

- **Supervised Graph Construction via IDML for graph-based SSL**
  - Distance metric learning helps in constructing better graphs.
  - Unsupervised data can also be incorporated via “Self-Training” to construct even better graphs.
  - Active Learning can also be used in IDML framework to decide which instances should be labeled.

IDML is likely to be useful for high dimensional data, e.g., problems in NLP as evident from the experiments on Text and Sentiment Analysis datasets.

References

[1] S. Daitch, J. Kelner, and D. Spielman. Fitting a graph to vector data. In *ICML*, 2009.

[2] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon. Information-theoretic metric learning. In *ICML*, 2007.

[3] T. Jebara, J. Wang, and S. Chang. Graph construction and b-matching for semi-supervised learning. In *ICML*, 2009.

[4] K. Weinberger and L. Saul. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 2009.

[5] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In *ICML*, 2003.