Efficient and Robust Semi-supervised Learning over a Sparse-Regularized Graph

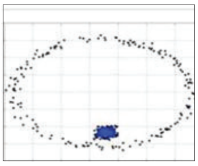
Hang Su*, Jun Zhu*, Zhaozheng Yin#, Yinpeng Dong*, and Bo Zhang*

*Department of Computer Science and Technology, Tsinghua University, China

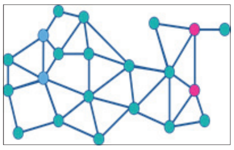
#Department of Computer Science, Missouri University of Science and Technology, USA

Motivation

- ➢ Graph-based Semi-Supervised Learning is impressive
- ➢ But has limitations



• Sensitive to parameters



Label Propagation: $\mathbf{Y}_u^* = \mathbf{L}_{uu}^{-1}\mathbf{W}_{ul}\mathbf{Y}_l$

• Computationally prohibitive

• Characteristics evolving

Goal: Develop a novel GSSL based on a batch of informative beacons with sparsity appropriately harnessed

Contribution

- ➢ Propose an ℓ_1-Beacon Graph based semi-supervised algorithm
  - Place a batch of characteristic-specific beacons in the feature domain
  - Represent the original samples with a subset of beacons
  - Predict missing labels with label fusion of the corresponding beacons.
- ➢ Our proposed method outperforms the previous algorithms
  - Mitigate the computational bottleneck by weighted averaging the soft labels of a subset of beacons
  - Enhance robustness by incorporating the sparse regularization
  - Boost the performance incrementally by expanding the beacon set and update their characteristic parameters dynamically

Construction of ℓ_1-Beacon Graph

- ➢ Generate a set of beacons automatically, which behave as indicators to guide the inference procedure

$$(\Psi^*,\mathbf{Z}^*) = \arg\min_{\Psi,\mathbf{Z}} R_{\mathcal{L}}(\Psi,\mathbf{Z}) + R_{\mathcal{U}}(\Psi,\mathbf{Z}) + \lambda R_{\mathbf{Y}}(\Psi,\mathbf{Z})$$

Feature Domain

Labeled set

$$R_{\mathcal{L}}(\Psi,\mathbf{Z}) = \sum_{i=1}^{N_l}\left\|\begin{bmatrix}\mathbf{x}_i\\ \mathbf{y}_i\end{bmatrix} - \Psi\mathbf{z}_i\right\|_2^2$$

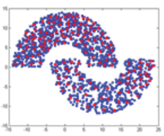
Unlabeled set

$$R_{\mathcal{U}}(\Psi,\mathbf{Z}) = \sum_{i=1}^{N_u}\left\|\mathbf{x}_i - \mathbf{S}_b\Psi\mathbf{z}_i\right\|_2^2$$

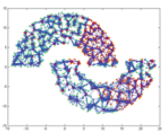
Label Domain

Label Consistent

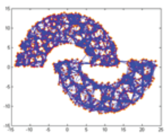
$$R_{\mathbf{Y}}(\Psi,\mathbf{Z}) = \sum_{i,j}^{N}\|\mathbf{y}_i - \mathbf{y}_j\|_2^2 w_{ij}$$



Beacon Generation



ℓ_1-Beacon Graph



K-means Beacon Graph

Optimization Algorithm

- ➢ Solve by alternately minimizing one variable while keeping the other one fixed

Calculate $\mathbf{Z}$ by fixing Ψ

$$\mathbf{Z}^* = \arg\min_{\mathbf{Z}=[\mathbf{z}_i]}\left\{\sum_{i=1}^{N_l}\|[\mathbf{x}_i;\mathbf{y}_i] - \Psi\mathbf{z}_i\|_2^2 + \sum_{i=1}^{N_u}\|\mathbf{x}_i - \mathbf{B}\mathbf{z}_i\|_2^2 + \lambda\sum_{i=1}^{N}\mathbf{z}_i^T\mathbf{F}^T\mathbf{L}\mathbf{F}\mathbf{z}_i\right\}$$

Calculate the Ψ by fixing $\mathbf{Z}$ as

$$\Psi^* = \arg\min_{\Psi}\left\{\|[\mathbf{X}_l;\mathbf{Y}_l] - \Psi\mathbf{Z}_l\|_2^2 + \|\mathbf{X}_u - \mathbf{B}\mathbf{Z}_u\|_2^2 + \lambda\, trace(\mathbf{F}\mathbf{Z}\mathbf{L}\mathbf{Z}^T\mathbf{F}^T)\right\}$$

Inductive Inference

- ➢ Implement scalable inference by label fusion

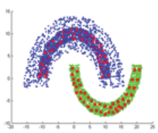
$$\tilde{\mathbf{y}}_i^* = \mathbf{F}\tilde{\mathbf{z}}_i^* \Leftarrow \tilde{\mathbf{z}}_i = \arg\min_{\mathbf{z}_i}\|\tilde{\mathbf{x}}_i - \mathbf{B}\tilde{\mathbf{z}}_i\|_2^2$$

Incremental Update of Beacons in Open-Set

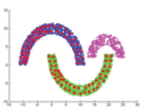
- ➢ Training and testing data may exhibit different statistics
- ➢ Characteristics of samples may evolve over time
- ➢ Expand the beacon set incrementally

$$\bar{\mathbf{B}} \triangleq [\mathbf{B}, \mathbf{B}_k]$$

add k new beacons



Statistics mismatch



Novel classes

Beacon Set Update

Repeat

Randomly shuffle $\bar{N}$ unfamiliar samples

$$\min_{\bar{\mathbf{Z}}_t=[\bar{\mathbf{z}}_i]}\left\{\sum_{i=1}^{\bar{N}}\|\bar{\mathbf{x}}_i - \bar{\mathbf{B}}_{t-1}\bar{\mathbf{z}}_i\|_2^2 + \lambda\sum_{i=1}^{\bar{N}}\bar{\mathbf{z}}_i^T\bar{\mathbf{F}}_{t-1}^T\bar{\mathbf{L}}_{t-1}\bar{\mathbf{F}}_{t-1}\bar{\mathbf{z}}_i\right\}$$

Update the beacon set

$$\bar{\mathbf{B}}_t \leftarrow \bar{\mathbf{B}}_{t-1} + 2\delta(\bar{\mathbf{X}}_t\bar{\mathbf{Z}}_t^T - \bar{\mathbf{B}}_{t-1}\bar{\mathbf{Z}}_t\bar{\mathbf{Z}}_t^T)$$

$$\bar{\mathbf{F}}_t \leftarrow \bar{\mathbf{F}}_{t-1} + 2\delta\left((\mathbf{Y}_l - \bar{\mathbf{F}}_{t-1}\mathbf{Z}_l)\mathbf{Z}_l^T - \lambda\bar{\mathbf{F}}_{t-1}\bar{\mathbf{Z}}_t\mathbf{L}_t\bar{\mathbf{Z}}_t^T\right)$$

Feed the "unfamiliar" samples incrementally

Experimental Results

	ℓ_1-Beacon	K-means LAE	K-means Kernel	ℓ_1-Graph	KNN Graph
MNIST	**94.87**	91.86	92.12	92.78	93.99
CIFAR	**72.53**	65.72	65.34	70.21	70.61
CELL	95.13	92.33	92.87	94.33	**95.46**
$M \ll N$ ⇨	$O(M^2N^{1.3})$	$O(M^2N)$	$O(M^2N)$	$O(N^{3.3})$	$O(N^3)$

- ➢ Effective in classification and object detection
- ➢ Robustness to sub-optimal parameters and reduce time cost
- ➢ Handle statistics shift

Conclusion

- ➢ Implement semi-supervised inference by coupling the design of informative beacon set and sample-to-beacon relationship
- ➢ Orders of magnitude more efficient in computation
- ➢ Offer a solution to handle unfamiliar samples or unreliable inference