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
- Computationally prohibitive
- Characteristics evolving

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

Contribution

- Propose an $\ell_1$-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 $\ell_1$-Beacon Graph

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

$$(\Psi^*, \mathbf{Z}^*) = \arg \min_{\Psi, \mathbf{Z}} R_c(\Psi, \mathbf{Z}) + R_u(\Psi, \mathbf{Z}) + \lambda R_1(\Psi, \mathbf{Z})$$

- Feature Domain
- Label Domain

Unlabeled set

Unlabeled set

Labeled set

Optimization Algorithm

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

$\mathbf{Z}^* = \arg \min_{\mathbf{Z}} \left[ \sum_{i=1}^{m} \left\| \mathbf{x}_i - \mathbf{z}_i \right\|_2^2 + \lambda \sum_{k=1}^{K} \left\| \mathbf{z}_k \right\|_1 \right]$  

$\Psi^* = \arg \min_{\Psi} \left[ \sum_{i=1}^{m} \left\| \mathbf{x}_i - \mathbf{z}_i \right\|_2^2 + \lambda \sum_{k=1}^{K} \left\| \mathbf{z}_k \right\|_1 \right]$  

Inductive Inference

- Implement scalable inference by label fusion

$\tilde{y}_i = \mathbf{Fz}_i \quad \tilde{z}_i = \arg \min_{\mathbf{z}} \left\| \mathbf{z} - \mathbf{Bz}_i \right\|_2^2$

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

  \[ \begin{array}{c}
  \mathbf{B} + k \text{ new beacons}
  \end{array} \]

  Randomly shuffle $N$ unfamiliar samples

  $\min_{\mathbf{z}} \left\| \mathbf{z} - \mathbf{B}_k \mathbf{z} + \lambda \sum_{i=1}^{K} \mathbf{w}_i \right\|_2^2$

  Update the beacon set

  $\mathbf{B}_k = \mathbf{B}_{k-1} + 2\delta (\mathbf{X}_i \mathbf{z}_k - \mathbf{B}(\mathbf{z}_k \mathbf{Z}_i))$

  $\mathbf{F}_k = \mathbf{F}_{k-1} + 2\delta (\mathbf{Y} - \mathbf{F}_{k-1} \mathbf{z}_k - \mathbf{F}_k \mathbf{z}_k) - \lambda \mathbf{F}_k \mathbf{L}_k \mathbf{z}_k$

  Feed the "unfamiliar" samples incrementally

Experimental Results

Table: Classification accuracy on various datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>94.97</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>92.78</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>95.99</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>97.32</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>99.74</td>
</tr>
<tr>
<td>CIFAR-10</td>
<td>90.68</td>
</tr>
</tbody>
</table>

- 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