Non-I.I.D. Multi-Instance Dimensionality Reduction by Learning a Maximum Bag Margin Subspace

Wei Ping*, Ye Xu§, Kexin Ren◊, Chi-Hung Chi*, Furao Shen§

* Tsinghua University
§ Nanjing University
◊ Nanjing University of A. & A.
Outline

- Problem Origin
- Main Ideas of MidLABS (Multi-Instance Dimensionality reduction by Learning a maximum bag margin Subspace)
- Description of MidLABS
- Conclusion
Problem Origin

- **Multi-Instance Learning (MIL):**
  - Known: Labels of bags
  - Unknown: Labels of instances in positive bag.

- **Two perspectives for solving MIL:**
  - Instance level
  - Bag level
Problem Origin: Dimensionality Reduction for MIL

- “Curse of Dimensionality” also spoils MIL

- Dimensionality Reduction (DR) for MIL
  - Supervised DR method could not be directly applied.
  - Unsupervised method will waste bag label information.
Formal definition of DR for MIL:

Given a data set $T = \{(X_1, L_1), \ldots, (X_i, L_i), \ldots, (X_N, L_N)\}$, finding a transformation matrix $W = [w_1, w_2, \ldots, w_d]$, which maps every $x_{ij} \in \mathcal{R}^D$ in each bag $X_i$ to $y_{ij} \in \mathcal{R}^d$ in new bag $Y_i$ i.e. $y_{ij} = W^T x_{ij}$, such that $y_{ij}$ “represents” $x_{ij}$ and $Y_i$ “represents” $X_i$, then we could get data set $\{(Y_1, L_1), \ldots, (Y_i, L_i), \ldots, (Y_N, L_N)\}$ in feature space.
Main Ideas of MidLABS

- **Structure (Non-I.I.D) information** [Zhou et al., ICML09]
  - Bag level perspective
  - Establish local geometric structure for each bag.

- **Bag label information**
  - Learn a transformation which **Maximize bag margin** between positive and negative bags
Description of MidLABS: Distance Metric

- Distance metric of bags on a mapping vector ($w$) is as follows:

$$Dis(X_i, X_j) = \sum_{a=1}^{n_i} \sum_{b=1}^{n_j} (y_{ia} - y_{jb})^2$$

where $y_{ia} = w^\top x_{ia}$ is the mapped point of $x_{ia}$ on this line from bag $X_i$, and $y_{jb} = w^\top x_{jb}$ is the mapped point of $x_{jb}$ on this line from bag $X_j$.

- **Pair-wise metric** is employed in [Gartner et al., ICML02] to setup kernels/similarity between multi-instance bags.
Description of MidLABS: Structure Information

Setup a $\varepsilon$-graph to capture the structure information in a bag.

- Rule: For the bag $x_i$, if the distance between nodes (instances) $x_{iu}$ and $x_{iv}$ is smaller than a threshold $\varepsilon$, then an edge $e = x_{iu} - x_{iv}$ is setup.

Choose the node which has the first larger attribute as the starting node in our implementation.
Distance metric incorporated the non-i.i.d. information conveyed by the edges

\[ \text{Dis}_G(X_i, X_j) = \text{Dis}_{\text{node}}(X_i, X_j) + C \cdot \text{Dis}_{\text{edge}}(X_i, X_j) \]

\[ = \sum_{a=1}^{n_i} \sum_{b=1}^{n_j} \frac{(w^T x_{ia} - w^T x_{jb})^2}{n_i n_j} + C \cdot \sum_{c=1}^{m_i} \sum_{d=1}^{m_j} \frac{(w^T e_{ic} - w^T e_{jd})^2}{n_i^2 n_j^2} \]

Where \( w^T e_{ic} \) is the projection of the edge \( e_{ic} \) from bag \( X_i \) in the line, \( w^T e_{jd} \) is the projection of the edge \( e_{jd} \) from bag \( X_j \) in the line. Denominator are used to do normalization.
Based on previous distance metric, choose $w$ which maximizes the following objective function:

$$J(w) = \frac{\sum_{L_i \neq L_j} D_{iS}(X_i, X_j)}{\sum_{L_i = L_j} D_{iS}(X_i, X_j)}$$

Attempt to ensure:

- if $X_i$ and $X_j$ share the same label, then stay as close as possible after mapping;
- if they have different labels, then stay as distant as possible after mapping.
If we denote

\[ K_{ij} = \frac{\sum_{a=1}^{n_i} \sum_{b=1}^{n_j} (x_{ia} - x_{jb})(x_{ia} - x_{jb})^\top}{n_i n_j} \]

\[ + C \frac{\sum_{c=1}^{m_i} \sum_{d=1}^{m_j} (e_{ic} - e_{jd})(e_{ic} - e_{jd})^\top}{n_i^2 n_j^2} \]

Now, the objective function could be reduced to

\[ J(w) = \frac{w^\top (\sum_{L_i \neq L_j} K_{ij}) w}{w^\top (\sum_{L_i = L_j} K_{ij}) w} \]
Description of MidLABS: Optimization

Then, $J(w)$ can be rewritten as the following form:

$$J(w) = \frac{w^\top S_b w}{w^\top S_w w}$$

where $S_b = \sum_{i \neq j} K_{ij}$, $S_w = \sum_{i = j} K_{ij}$

- It is a generalized Rayleigh quotient, and could be maximized through Lagrange Multipliers method.
Description of MidLABS: Summary of the Algorithm

**Input:** Data set \( \{(X_1, L_1), \ldots, (X_i, L_i), \ldots, (X_N, L_N)\} \) and the target dimension \( d \)

1: Construct the \( \epsilon \)-graph for every bag \( X_i \), establish edges inside bag.
2: Compute \( S_b \) and \( S_w \) as previous mention
3: Solve the generalized eigenvalue equation
4: Construct the \( D \times d \) matrix \( W \) whose columns are composed by the eigenvectors corresponding to largest \( d \) eigenvalues.

**Output:** \( W \) the projection from \( \mathcal{R}^D \) to \( \mathcal{R}^d \).
Experiment: Musk Data Set

Table 1: Classification Accuracy (CA) and Dimension Ratio (DR) of MidLABS, PCA, LLE, and ORI under Musk1 and Musk2.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>MidLABS</th>
<th>PCA</th>
<th>LLE</th>
<th>ORI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Musk1 CA</td>
<td>90.0%</td>
<td>87.5%</td>
<td>85.9%</td>
<td>86.4%</td>
</tr>
<tr>
<td></td>
<td>±2.7%</td>
<td>±4.1%</td>
<td>±5.1%</td>
<td>±3.1%</td>
</tr>
<tr>
<td>Musk1 DR</td>
<td>18.1%</td>
<td>33.1%</td>
<td>36.1%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Musk2 CA</td>
<td>85.3%</td>
<td>86.2%</td>
<td>85.2%</td>
<td>88.0%</td>
</tr>
<tr>
<td></td>
<td>±1.8%</td>
<td>±2.8%</td>
<td>±2.5%</td>
<td>±1.5%</td>
</tr>
<tr>
<td>Musk2 DR</td>
<td>12.1%</td>
<td>39.2%</td>
<td>48.2%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

- Dimension Ratio is defined as target dimension d dividing the original dimension D.
- ORI is the abbreviation of Original feature space
Table 2: Classification Accuracy (CA) and Dimension Ratio (DR) of MidLABS, PCA, LLE, and ORI under datasets of Elephant, Fox, and Tiger.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>MidLABS</th>
<th>PCA</th>
<th>LLE</th>
<th>ORI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elephant CA</strong></td>
<td>86.5% ±1.4%</td>
<td>86.0% ±1.2%</td>
<td>84.0% ±1.1%</td>
<td>84.3% ±1.6%</td>
</tr>
<tr>
<td><strong>Elephant DR</strong></td>
<td>13.0%</td>
<td>17.4%</td>
<td>26.1%</td>
<td>100.0%</td>
</tr>
<tr>
<td><strong>Fox CA</strong></td>
<td>67.0% ±2.1%</td>
<td>64.0% ±2.4%</td>
<td>66.0% ±2.2%</td>
<td>60.3% ±1.9%</td>
</tr>
<tr>
<td><strong>Fox DR</strong></td>
<td>17.4%</td>
<td>21.3%</td>
<td>20.4%</td>
<td>100.0%</td>
</tr>
<tr>
<td><strong>Tiger CA</strong></td>
<td>87.5% ±1.6%</td>
<td>84.5% ±1.9%</td>
<td>86.0% ±1.9%</td>
<td>84.2% ±1.0%</td>
</tr>
<tr>
<td><strong>Tiger DR</strong></td>
<td>21.7%</td>
<td>21.7%</td>
<td>17.4%</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
Conclusion

- MidLABS: DR algorithm for MIL
- Captures **structure information** in each bag, which plays an important role in obtaining salient feature
- Exploit the **label information of bags** to guarantee a powerful discriminant ability.
Thanks!