Robust Spectral Learning for Unsupervised Feature Selection

Lei Shi, Liang Du and Yi-Dong Shen

Institute of Software, Chinese Academy of Sciences

ICDM 2014, Shenzhen, China
Outline

1. Brief Introduction of Feature Selection
2. Our Proposal
3. Experiments
4. Summary
Brief Introduction of Feature Selection

- select a subset of features from original feature space
  - Reduce computation cost
  - Remove noisy features

$D \in \mathbb{R}^{n \times d}$

Feature Selection

$D_1 \in \mathbb{R}^{n \times k}$
Brief Introduction of Feature Selection

- **Supervised Feature Selection**
  - leverage the supervised information to guide the search of relevant features

- **Unsupervised Feature Selection**
  - more challenging due to the lack of supervised information (e.g., class labels)
Main Techniques of Unsupervised Feature Selection

- **Keys**
  - 1) Structure characterization; 2) Feature Weight Learning
- **Main Steps**
  - estimate intrinsic structure (e.g., various graph Laplacian)
  - discover the cluster structure (e.g., spectral embedding, Matrix Factorization)
  - Feature weight learning by some criterion or sparsity regularization
- **Examples**
  - LaplacianScore (NIPS’05), MCFS (KDD’10), NDFS (AAAI’12), RUFS (IJCAI’13)
Limitations of Existing Work

- Graph Laplacian is affected by noisy features
  - Existing method construct graph Laplacian from original feature space, which contains noisy features
  - degenerate the quality of the induces graph embedding
- Estimated cluster structure is with noise
  - relaxing discrete class labels into continuous ones may inevitably introduce noise
Outline

1. Brief Introduction of Feature Selection
2. Our Proposal
3. Experiments
4. Summary
Basic Idea

- jointly improve the robustness of graph embedding and sparse spectral regression
  - employ local kernel regression with $\ell_1$ norm to measure the local learning estimation error, which can be reformulated as a graph embedding problem.
  - explicitly extracting sparse noise in the learned graph embedding
Robust Graph Embedding

- **Notations**
  - Data matrix: \( X = \{x_1, ..., x_n\} \in \mathbb{R}^{d \times n} \)
  - Partition matrix: \( Y = [y_1, \ldots, y_c] = [y_{il}] \in \{0, 1\}^{n \times c} \)

- **Robust Local kernel regression**

\[
\min_{Y \in \mathbb{R}^{n \times c}} J(Y) = \sum_{l=1}^{c} \|y_l - Sy_l\|_1. \tag{1}
\]

\[
S_{ij} = \begin{cases} 
\frac{K(x_i, x_j)}{\sum_{x_j \in N_i} K(x_i, x_j)} & x_j \in N_i \\
0 & x_j \notin N_i
\end{cases} \tag{2}
\]
Robust Graph Embedding

- Equivalent to minimizing the following problem

\[ J(F) = Tr(F^T M F) \]  

\[ M = (B - S - S^T). \]  

\[ B \] is the degree matrix of \((S + S^T)\).  

\[ F = [f_1, \ldots, f_c] \] is defined as \( F = Y(Y^T Y)^{-\frac{1}{2}} \), and \( F^T F = I_c \).
Robust Spectral Regression

• assume the noise on the estimated cluster structure is sparse
• introduce a sparse matrix $Z \in \mathbb{R}^{n \times c}$ to explicitly capture the sparse noise

$$\min_{W,Z} \| (F - Z) - X^T W \|_F^2, \quad s.t. |Z|_1 < \eta_1, \|W\|_{2,1} < \eta_2 \quad (4)$$
Combining the robust graph embedding and robust sparse spectral regression, we get

$$\min_{F, W, Z} \quad Tr(F^T MF) + \alpha \|(F - Z) - X^T W\|_F^2$$
$$+ \beta \|W\|_{2,1} + \gamma \|Z\|_1$$
$$s.t. \quad F \in R_{+}^{n \times c}, F = Y(Y^T Y)^{-\frac{1}{2}}$$

Relaxing the elements in $F$ into continuous ones,

$$\min_{F, W, Z} \quad Tr(F^T MF) + \alpha \|(F - Z) - X^T W\|_F^2$$
$$+ \beta \|W\|_{2,1} + \gamma \|Z\|_1$$
$$s.t. \quad F \in R_{+}^{n \times c}, F^T F = I_c$$
We develop an coordinate descent algorithm to alternatively minimizing the objective function with respect to $W$, $Z$, and $F$. This procedure is repeated until convergence. Please refer to our paper for more details.
Outline

1 Brief Introduction of Feature Selection
2 Our Proposal
3 Experiments
4 Summary
Experimental setup

- Comparing methods
  - Random, LS (NIPS2005), MCFS (KDD2010), UDFS (IJCAI2011), NDFS (AAAI2012), RUFS (IJCAI2013)

- Data sets

  Table: Summary of data sets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th>Dimensions</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBCSport</td>
<td>737</td>
<td>1000</td>
<td>5</td>
</tr>
<tr>
<td>WebKB4</td>
<td>4199</td>
<td>1000</td>
<td>4</td>
</tr>
<tr>
<td>ORL</td>
<td>400</td>
<td>1024</td>
<td>40</td>
</tr>
<tr>
<td>COIL20</td>
<td>1440</td>
<td>1024</td>
<td>20</td>
</tr>
<tr>
<td>MNIST</td>
<td>4000</td>
<td>784</td>
<td>10</td>
</tr>
<tr>
<td>Jaffe</td>
<td>213</td>
<td>676</td>
<td>10</td>
</tr>
</tbody>
</table>

- Perform k-means clustering on the selected features. ACC and NMI are reported
Clustering on Data Sets without Explicit Noise

Figure: Clustering accuracy and normalized mutual information versus the number of selected features on all the data sets
Clustering on Data Sets with Malicious Occlusion

(a) Corrupted ratio = 0.2  
(b) Corrupted ratio = 0.3

Figure: Clustering Accuracy on ORL with different ratio of noisy images
Outline

1 Brief Introduction of Feature Selection
2 Our Proposal
3 Experiments
4 Summary
we proposed a unified robust spectral framework for unsupervised feature selection
  - robust graph embedding
  - robust sparse spectral regression
experimental results verified the effectiveness of our proposed method
  - in datasets with and without explicit noise
Thanks for your attention!
Questions?

Code available at: http://kingsleyshi.com/codes/RSFS.rar