Mixed Sparsity Regularized Multi-view Unsupervised Feature Selection

Kennedy W. Wangila    Ke Gao    Pengfei Zhu*
Qinghua Hu    Changqing Zhang

School of Computer Science and Technology
Tianjin University, China
zhupengfei@tju.edu.cn

September 20, 2017
1. Introduction

2. Multi-view Feature Selection

3. Mixed Sparsity Regularized Feature Selection

4. Experiment

5. Conclusion
Outlines

1. Introduction

2. Multi-view Feature Selection

3. Mixed Sparsity Regularized Feature Selection

4. Experiment

5. Conclusion
1. Introduction

1. Introduction

Correlations between feature pairs using a 2D correlation matrix

The evolution (rise) of feature dimensionality in correlation matrices.
(a) Diabetes (8 features)
(b) Lung Cancer (56 features)
(c) Psoriasis (529,651 features)

1. Introduction

- Problem

  - Time and storage
    The high-dimensionality data always contain a plenty of redundant data and noise, which may lead to high time complexity, large storage burden.

  - Generalization ability
    The high-dimensional data may increase the number of parameters of the learning machines, and therefore easily lead to over-fitting and degradation of the generalization ability.
1. Introduction

Solution

- **Subspace learning:**
  Subspace learning is one of the most effective ways to eliminate the curse of dimensionality by projecting the data to a low-dimensional feature subspace.

- **Feature selection:**
  Feature selection directly selects a subset of relevant and most representative features. And it is also an effective technique to reduce storage burden and time complexity, and improve generalization ability of the learned.
1. Introduction

- **Feature selection**
  - **Filter Methods**
    Filter method based on general features like the correlation with the variable to predict.
    e.g. variance, Laplacian Score, consistence
  - **Wrapper Methods**
    Wrapper Methods evaluate subsets of variables which allows to detect the possible interactions between variables.
    e.g. R-SVM, SVM-RFE
  - **Embedded Methods**
    Embedded methods takes advantage of its own variable selection process and performs feature selection and classification simultaneously.
    e.g. EUFS

1. Introduction

2. Multi-view Feature Selection

3. Mixed Sparsity Regularized Feature Selection

4. Experiment

5. Conclusion
2. Multi-view Feature Selection

2. Multi-view Feature Selection

2. Multi-view Feature Selection

2. Multi-view Feature Selection

2. Multi-view Feature Selection

(a) The Concatenating Strategy

(b) The Separation Strategy

(c) Multi-view Feature Selection

\[
\begin{align*}
\min_{W,Z} \quad & J(W,Z) = \sum_{i=1}^{m} \lambda_i \text{Tr}(Z^T L_i Z) + \\
& \alpha(\|X_i^T W_i - Z\|^2_F + \beta \|W_i\|_{2,1}) \\
\text{s.t.} \quad & Z^T Z = I, \quad Z \geq 0.
\end{align*}
\]

1. Introduction

2. Multi-view Feature Selection

3. **Mixed Sparsity Regularized Feature Selection**

4. Experiment

5. Conclusion
3. Mixed Sparsity Regularized Feature Selection

- **Mixed Sparsity Regularized Learning**

\[
X = [X_1, \ldots, X_i, \ldots, X_m] \in \mathbb{R}^{d \times n} \quad X_i \in \mathbb{R}^{d_i \times n} \quad P \in \mathbb{R}^{d \times r}
\]

\[
\min_{P^T P = I} \sum_{i=1}^{m} \sqrt{Tr(P^T X_i X^T P)} + \lambda R(P)
\]

- **Feature level:**

\[
\|P\|_{2,1} = \sum_{i=1}^{d} \|P_i\|_2 = \sum_{i=1}^{d} \sqrt{\sum_{j=1}^{r} p_{ij}^2}
\]

- **View level:**

\[
\|P^i\|_F = \sqrt{\sum_{j=1}^{d_i} \|p_{ij}^i\|^2} = \sqrt{\sum_{j=1}^{d_i} \sum_{k=1}^{r} (p_{jk}^i)^2}
\]

\[
\min \sum_{i=1}^{m} \sqrt{Tr(P^T X_i X^T P)} + \lambda_1 \|P\|_{2,1} + \lambda_2 \sum_{j=1}^{m} \|P^j\|_F
\]

\[s.t. \quad P^T P = I\]
3. Mixed Sparsity Regularized Feature Selection

- **Parameter-free Multi-view Learning**

\[
\min \sum_{i=1}^{m} \sqrt{Tr(P^T X_i L_i X_i^T P)} + \lambda_1 \|P\|_{2,1} + \lambda_2 \sum_{j=1}^{m} \|P^j\|_F
\]

s.t. \(P^T P = I\)

- **Lagrange Multiplier Method**

\[
\sum_{i=1}^{m} \sqrt{Tr(P^T X_i L_i X_i^T P)} + \lambda_1 \|P\|_{2,1} + \lambda_2 \sum_{j=1}^{m} \|P^j\|_F + G(\Lambda, F)
\]

\[
\sum_{i=1}^{m} \alpha_i \frac{\partial Tr(P^T X_i L_i X_i^T P)}{\partial P} + \lambda_1 \frac{\partial \|P\|_{2,1}}{\partial P} + \lambda_2 \frac{\partial \sum_{j=1}^{m} \|P^j\|_F}{\partial P} + \frac{\partial G(\Lambda, F)}{\partial P}
\]

\[
\alpha_i = \frac{1}{2\sqrt{Tr(P^T X_i L_i X_i^T P)}},
\]

(d) Mixed Sparsity Regularized Feature Selection
Outlines

1. Introduction
2. Multi-view Feature Selection
3. Mixed Sparsity Regularized Feature Selection
4. Experiment
5. Conclusion
4. Experiment

- **Settings:**
  - Clustering accuracy
  - Clustering NMI

- **Datasets:**

<table>
<thead>
<tr>
<th>DATA</th>
<th>Samples</th>
<th>view</th>
<th>Features</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caltech10</td>
<td>800</td>
<td>4</td>
<td>200, 512, 59, 680</td>
<td>10</td>
</tr>
<tr>
<td>Corel800</td>
<td>800</td>
<td>4</td>
<td>200, 512, 59, 680</td>
<td>10</td>
</tr>
<tr>
<td>flickr</td>
<td>1000</td>
<td>4</td>
<td>200, 512, 59, 680</td>
<td>10</td>
</tr>
<tr>
<td>mfeat</td>
<td>2000</td>
<td>6</td>
<td>216, 76, 64, 6, 240, 47</td>
<td>10</td>
</tr>
<tr>
<td>PPMI</td>
<td>1400</td>
<td>3</td>
<td>200, 200, 200</td>
<td>7</td>
</tr>
<tr>
<td>MSRA</td>
<td>210</td>
<td>5</td>
<td>1302, 512, 256, 210, 100</td>
<td>7</td>
</tr>
<tr>
<td>Still DB</td>
<td>467</td>
<td>3</td>
<td>200, 200, 200</td>
<td>6</td>
</tr>
</tbody>
</table>
4. Experiment

- **Experiment Results:**

  - **Clustering accuracy result of all data sets**

    | DATA   | Laplacian | SPEC  | MCFS  | UDFS  | AUMFS | MSMFS |
    |--------|-----------|-------|-------|-------|-------|-------|
    | Caltech10 | 0.2562    | 0.2223 | 0.2873 | 0.2887 | 0.3205 | **0.3444** |
    | Corel800  | 0.2986    | 0.2514 | 0.2851 | 0.2702 | 0.2913 | **0.3073** |
    | flickr    | 0.2146    | 0.2086 | **0.2369** | 0.2262 | 0.2288 | 0.2360 |
    | mfeat     | 0.5608    | 0.6416 | 0.6242 | 0.6538 | 0.6129 | **0.7105** |
    | PPMI      | 0.1969    | 0.2180 | 0.1987 | 0.2005 | 0.1989 | **0.2366** |
    | MSRA      | 0.5099    | 0.4786 | 0.5390 | 0.5155 | 0.5110 | **0.6746** |
    | Still DB   | 0.3013    | 0.2857 | 0.3004 | 0.3017 | **0.3124** | 0.3004 |

  - **Clustering NMI result of all data sets**

    | DATA   | Laplacian | SPEC  | MCFS  | UDFS  | AUMFS | MSMFS |
    |--------|-----------|-------|-------|-------|-------|-------|
    | Caltech10 | 0.1461    | 0.0962 | 0.1734 | 0.1767 | 0.2059 | **0.2199** |
    | Corel800  | 0.2198    | 0.1235 | 0.2255 | 0.1960 | 0.2302 | **0.2400** |
    | flickr    | 0.0993    | 0.1026 | **0.1353** | 0.1184 | 0.1309 | 0.1279 |
    | mfeat     | 0.5699    | 0.5960 | 0.6157 | 0.5983 | 0.5920 | **0.6253** |
    | PPMI      | 0.0224    | 0.0310 | 0.0255 | 0.0194 | 0.0238 | **0.0461** |
    | MSRA      | 0.4076    | 0.3902 | 0.4467 | 0.4100 | 0.4122 | **0.5915** |
    | Still DB   | 0.1019    | 0.0850 | 0.0930 | 0.0951 | 0.1035 | **0.1051** |
1. Introduction

2. Multi-view Feature Selection

3. Mixed Sparsity Regularized Feature Selection

4. Experiment

5. Conclusion
5. Conclusion

- Automatically learning the view weights.
- Alleviate the effect of the outlier views and features with noisy information.
Thank you!

Q & A