Dictionary Learning for Sparse Learning based Image Classification

Dr. Meng Yang
Smart LV Lab, Computer Vision Institute, Shenzhen University
OUTLINE

- Introduction of sparse learning
- Discriminative dictionary learning
- Outlook
OUTLINE

- Introduction of sparse learning
- Discriminative dictionary learning
- Outlook
DATA REPRESENTATION

- Massive High-Dimensional Data

- Low-dimensional structures
SPARSE TRANSFORMATION

Most energy concentrated in a small number of features

$x$ is a sparse vector.
Signal Processing $\rightarrow$ Compressive sensing $\rightarrow$ Sparse Learning

\[
M \Phi = y
\]

Measurement matrix

$M$ $\Phi$ $N$ $f$ $x$

$N$ $\Psi$
SPARSE NEURAL CODES

Visual Cortex → Neural codes → Sparse Learning

Lifetime sparseness → Population sparseness

Coding coefficient $x$ and residual $e$ are sparse!

Classification criterion: Identity = $\text{argmin}_i \{r_i\}$.

$$r_i = \|y - A\delta_i(\hat{x}_1) - \hat{e}_1\|_2$$

OUTLINE

- Introduction of sparse learning
- Discriminative dictionary learning
- Outlook
\[
\min_{\alpha} \| y - \alpha \|^2_2 + \lambda \| \alpha \|_1
\]

“The choice of the dictionary that sparsifies the signals is crucial for the success of this model.” M. Elad et al. Proc. IEEE 10.
CLASS-SPECIFIC DL

- Metaface[18], DLSI[19], CS-DL[21], FDDL[20]...
Predefined bases (e.g., wavelet, DCT) too general

Training data matrix may have a big size (e.g., SRC)

Discriminative Dictionary Learning

*Discriminative sparse coding coefficients*

*Discriminative class-specific sub-dictionary*
Main Idea

\[ D = [D_1, D_2, \ldots, D_c] \]

Discrimination of representation residual and coding coefficient.

**GOOD** for \( D_i \)

**BAD** for \( X_i \)

**Fisher criterion**

\[ \text{small} \] within-class scatter

\[ \text{big} \] between-class scatter

Sparse
FDDL MODEL

\[
\min_{D, X} \left\{ \sum_{i=1}^{K} r\left( A_i, D, X_i \right) + \lambda_1 \| X \|_1 + \lambda_2 f\left( X \right) \right\}
\]

**Discriminative data fidelity term**

\[
r\left( A_i, D, X_i \right) = \| A_i - DX_i \|_F^2 + \| A_i - D_i X_i \|_F^2 + \sum_{j=1}^{K} \| D_j X_i \|_F^2
\]

**Discriminative coefficient term**

\[
f\left( X \right) = tr\left( S_W\left( X \right) - S_B\left( X \right) \right) + \eta \| X \|_F^2
\]

\[A = [A_1, A_2, \ldots, A_K],\]
\[D = [D_1, D_2, \ldots, D_K],\]
\[X = [X_1, X_2, \ldots, X_K],\]
\[A_i : \text{training samples from class } i.\]
\[D_i : \text{sub-dictionary of the } i^{th} \text{ class.}\]
\[X_i : \text{coding coefficients of } A_i \text{ over } D.\]
\[
\min_{D,X} \left\{ \sum_{i=1}^{K} r(A_i, D, X_i) + \lambda_1 \|X_i^i\|_1 + \lambda_2 f(X_i) \right\} \quad \text{s.t.} \quad X_i^j = 0
\]

**Discriminative data fidelity term**

\[
r(A_i, D, X_i) = \|A_i - D_i X_i^i\|_F^2
\]

**Discriminative coefficient term**

\[
f(X) = \|X_i - M_i\|_F^2
\]

\(A = [A_1, A_2, \ldots, A_K]\), \(A_i :\) training samples from class \(i\).

\(D = [D_1, D_2, \ldots, D_K]\), \(D_i :\) sub-dictionary of the \(i^{th}\) class.

\(X = [X_1, X_2, \ldots, X_K]\), \(X_i :\) coding coefficients of \(A_i\) over \(D\).
CLASSIFICATION MODEL

- Global classifier

\[
\hat{\alpha} = \arg \min_{\alpha} \left\{ \| y - D\alpha \|_2^2 + \gamma \| \alpha \|_p \right\}
\]

- Local classifier for i-th class

\[
\hat{\alpha}_i = \arg \min_{\alpha_i} \left\{ \| y - D_i\alpha_i \|_2^2 + \gamma_1 \| \alpha_i \|_p + \gamma_2 \| \alpha_i - m_i \|_2^2 \right\}
\]
## Digit Recognition (USPS)

### Learned dictionary atoms

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>FDDL</th>
<th>SRSC</th>
<th>REC-L</th>
<th>REC-BL</th>
<th>SDL-G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error rate (%)</td>
<td>2.89</td>
<td>6.05</td>
<td>6.83</td>
<td>4.38</td>
<td>6.67</td>
</tr>
<tr>
<td>TDDL</td>
<td>SDL-D</td>
<td>DLSI</td>
<td>KNN</td>
<td>SVM</td>
<td>COPAR</td>
</tr>
<tr>
<td>2.84</td>
<td>3.54</td>
<td>3.98</td>
<td>5.2</td>
<td>4.2</td>
<td>3.61</td>
</tr>
</tbody>
</table>
Latent Dictionary Learning (LDL)

Latent vector for $d_m$

$w_{j,m}$ indicates the relationship between atom $d_m$ and $j^{th}$ class label.
Latent sparse representation

\[
\min_{D,X,W} \sum_{j=1}^{C} \left\| A_j - D \text{diag}(w_j) X_j \right\|_F^2 + \lambda_1 \left\| X_j \right\|_1 + \lambda_2 \left\| X_j - M_j \right\|_F^2
\]

\[+ \lambda_3 \sum_{j=1}^{C} \sum_{l \neq j} \sum_{n=1}^{N} \sum_{m \neq n} w_{j,m} \left( d_m^T d_n \right)^2 w_{l,n} \]

s.t. \( w_{j,m} \geq 0 \quad \forall \ j, m; \)

\[\sum_m w_{j,m} = \delta, \quad \forall \ m; \]

Latent dictionary incoherence

The latent vector could be efficiently solved by an iterative procedure, and there is an analytical solution in each iteration.
LATENT CLASSIFICATION MODEL

- Global classifier

\[ \hat{\alpha} = \arg \min_{\alpha} \| y - D \text{diag}(\mu) \alpha \|_2^2 + \lambda \| \alpha \|_1 \]

- Local classifier for j-th class

\[ \hat{\alpha} = \arg \min_{\alpha} \| y - D \text{diag}(w_j) \alpha \|_2^2 + \lambda \| \alpha \|_1 \]
## ACT RECOGNITION (UCF SPORTS ACTION)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Accuracy (%)</th>
<th>Methods</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qiu 2011</td>
<td>83.6</td>
<td>LCKSVD</td>
<td>91.2</td>
</tr>
<tr>
<td>Sadanand 2012</td>
<td>90.7</td>
<td>COPAR</td>
<td>90.7</td>
</tr>
<tr>
<td>SRC</td>
<td>92.9</td>
<td>JDL</td>
<td>90.0</td>
</tr>
<tr>
<td>DLSI</td>
<td>92.1</td>
<td>FDDL</td>
<td>93.6</td>
</tr>
<tr>
<td>DKSVD</td>
<td>88.1</td>
<td>LDL</td>
<td><strong>95.0</strong></td>
</tr>
</tbody>
</table>
OUTLINE

- Introduction of sparse learning
- Discriminative dictionary learning
- Outlook
Beyond small data
Beyond shallow dictionary learning
非常感谢各位！

Question?