An M-ary KMP Classifier for Multi-aspect Target Classification

Xuejun Liao, Hui Li, and Balaji Krishnapuram

Department of Electrical and Computer Engineering
Duke University

International Conference on Acoustics, Speech, and Signal Processing, May 17-21, 2004, Montreal, Quebec, Canada
Outline

• The M-ary Kernel Matching Pursuit (KMP)

• Multi-aspect High Range Resolution (HRR) Target Classification

• Results on MSTAR Dataset
The M-ary Classification Problem

<table>
<thead>
<tr>
<th>Class 1</th>
<th>Class 2</th>
<th>...</th>
<th>Class $m$</th>
<th>...</th>
<th>Class $M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Class 2</td>
<td>...</td>
<td>Class $m$</td>
<td>...</td>
<td>Class $M$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Class 1</td>
<td>Class 2</td>
<td>...</td>
<td>Class $m$</td>
<td>...</td>
<td>Class $M$</td>
</tr>
</tbody>
</table>

One-versus-all rule:
- Labeled as 0
- Labeled as 1
The M-ary Classification Problem (Cont.)

• Training data set

\[ D = \{(x_i, y_i, \beta_i) : y_i \in \{1,2,\ldots,M\}, \beta_i \geq 0, i = 1,2,\ldots,J\} \]

\( y_i \) is the class label of \( x_i \)

\( \beta_i \) is the importance weight of \((x_i, y_i)\)

• One-versus-all (OVA) re-labeling

\[ y_i^{[m]} = \begin{cases} 1, & \text{if } y_i = m \\ 0, & \text{otherwise} \end{cases}, \quad m = 1,2,\ldots,M \]
M-ary Kernel Matching Pursuit (KMP)

- Decision functions

\[ f^{[m]}(x) = (w^{[m]}_T)^T \phi(x), \quad m=1,2,…,M \]

- Basis functions

\[ \phi(x) = [\xi^{(1)}(x), \xi^{(2)}(x), \ldots, \xi^{(N)}(x)]^T \]

- Weight vectors

\[ w^{[m]} = [w_1^{[m]}, w_2^{[m]}, \ldots, w_N^{[m]}]^T \]

- Squared error

\[
e(\phi, \{w^{[m]}\}) = \sum_{m=1}^{M} \sum_{i=1}^{J} \beta_i [y_i^{[m]} - f^{[m]}(x_i)]^2 = \sum_{m=1}^{M} \sum_{i=1}^{J} \beta_i [y_i^{[m]} - (w^{[m]}_m)^T \phi(x_i)]^2
\]

- Objective

Find the \( \phi \) and \( \{w^{[m]}\} \) that minimize \( e(\phi(x), \{w^{[m]}\}) \)
Least-Square Solution of $w^{[m]}$

• Least-square solution of $w^{[m]}$

$$w^{[m]} = A^{-1} \sum_{i=1}^{J} \beta_i y_i^{[m]} \phi(x_i), \quad m = 1, 2, ..., M$$

where

$$A = \sum_{i=1}^{J} \beta_i \phi(x_i) \phi^T(x_i)$$

• Substitute $w^{[m]}$ back into $e(\phi, w)$

$$e(\phi) = \sum_{m=1}^{M} \sum_{i=1}^{J} \beta_i (y_i^{[m]})^2 - \sum_{m=1}^{M} \left[ \sum_{i=1}^{J} \beta_i y_i^{[m]} \phi(x_i) \right]^T A^{-1} \left[ \sum_{n=1}^{J} \beta_i y_i^{[m]} \phi(x_i) \right]$$
Selection of Basis Functions

- Globally optimal selection

\[ \phi^* = \arg \min_{\phi} e(\phi) \]

- Locally optimal selection

Based on existing basis functions \( \phi(x) = [\xi^{(1)}(x), \ldots, \xi^{(n)}(x)]^T \)

select the next one

\[ \xi^{(n+1)} = \arg \max_{\xi} \delta e(\phi^{(n)}, \xi) \]

where

\[ \delta e(\phi^{(n)}, \xi) = e(\phi) - e\left(\begin{bmatrix} \phi^{(n)} \\ \xi \end{bmatrix}\right) \]

is the error decrease due to \( \xi \)
Selection of Basis Functions (Cont.)

• Simple expression of the error decrease

\[ \xi^{(n+1)} = \arg\max_\xi \delta e(\phi^{(n)}, \xi) = \arg\max_\xi q_\xi^{-1} \sum_{m=1}^{M} (g_\xi^{[m]})^2 \]

where

\[ c_\xi = \sum_{i=1}^{J} \beta_i \phi^{(n)}(x_i) \xi(x_i) \]

\[ d_\xi = \sum_{i=1}^{J} \beta_i \xi^2(x_i) \]

\[ q_\xi = d_\xi - c_\xi^T (A^{(n)})^{-1} c_\xi \]

\[ g_\xi^{[m]} = c_\xi^T w^{[m](n)} - \sum_{i=1}^{J} \beta_i y_i^{[m]} \xi(x_i) \]
Selection of Basis Functions (Cont.)

- Update of the M-ary KMP

\[
(A^{(n+1)})^{-1} = \begin{bmatrix}
(A^{(n)})^{-1} + (A^{(n)})^{-1} c_{\xi} q_{\xi}^{-1} c_{\xi}^T (A^{(n)})^{-1} & -(A^{(n)})^{-1} c_{\xi} q_{\xi}^{-1} \\
- q_{\xi}^{-1} c_{\xi}^T (A^{(n)})^{-1} & q_{\xi}^{-1}
\end{bmatrix}
\]

\[
w^{[m](n+1)} = \begin{bmatrix}
w^{[m](n)} + (A^{(n)})^{-1} c_{\xi} q_{\xi}^{-1} g^{[m]}_\xi \\
- q_{\xi}^{-1} g^{[m]}_\xi
\end{bmatrix}
\]

\[
\phi^{(n+1)} = \begin{bmatrix}
\phi^{(n)} \\
\xi^{(n+1)}
\end{bmatrix}
\]

- Time complexity of the M-ary KMP

\[O(N N_c J)\]

\[N_c = \text{number of candidate basis functions}\]

which is independent of \(M\)
Multi-aspect High Range Resolution (HRR) Target Classification

- Sensor-target configuration

- Multi-aspect HRR data with pose information

HRR sequence

$$= \{x_1, x_1, ..., x_J; \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_J\}$$
Probabilistic partition of azimuth and data

- **Azimuth partition**

\[
\Pr(s_i \mid \hat{\theta}) = \int \Pr(s_i \mid \theta) \Pr(\theta \mid \hat{\theta}) d\theta = \int q_i(\theta) p_e(\hat{\theta} - \theta) d\theta
\]

- **Data partition**

\[
D_i = \{(x_j, y_j, \beta_j) : \beta_j = \Pr(s_i \mid \hat{\theta}_j), y_j \in \{1,2,\ldots,M\}, j = 1,2,\ldots,J\}
\]

\(D_i\) is the data set of state \(i\), \(i=1,2,\ldots,L\)
Multi-aspect decision function

\[
\Pr(T_m \mid x_1, x_2, \ldots, x_J, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_J) = \prod_{j=1}^{J} \sum_{i=1}^{L} \Pr(T_m \mid s_i, x_j) \Pr(s_i \mid \hat{\theta}_j)
\]

where \(\sum_i^{L} \Pr(s_i \mid \hat{\theta}_j) = 1\) \(\sum_{i=1}^{L} \Pr(T_m \mid s_i, x) = 1\)

\[\Pr(T_m \mid s_i, x) = \exp\left(-\frac{[f_i^{[m]}(x) - 1]^2}{2(\sigma_i^{[m]})^2}\right)\]

\(f_i^{[m]}(x), \ m = 1, 2, \ldots, M\) is the M-ary KMP trained with the data in state \(i\)

\[D_i = \{(x_j, y_j, \beta_j) : \beta_j = \Pr(s_i \mid \hat{\theta}_j), y_j \in \{1, 2, \ldots, M\}, j = 1, 2, \ldots, J\}\]

\[
(\sigma_i^{[m]})^2 = \frac{\sum_{y_j=m} \Pr(s_i \mid \hat{\theta}_j)[f_i^{[m]}(x_j) - 1]^2}{\sum_{y_j=m} \Pr(s_i \mid \hat{\theta}_j)}
\]
The MSTAR dataset

- The ten MSTAR targets

- The data set consists of the X-band HRR signatures of the ten MSTAR targets. The HRR signatures have a range resolution of approximately 0.3 meters.

- Training data and testing data are distinct, each having a full coverage of 360° azimuth angles (depression angle fixed at 5°), with 0.1° azimuth sampling
Results of the M-ary KMP

Confusion matrix of KMP, with $L=120$ states in azimuth, test sequence spanning $3^\circ$ azimuth.

Sparsity: $\frac{N}{J} = 0.22$

<table>
<thead>
<tr>
<th></th>
<th>T72</th>
<th>BTR70</th>
<th>BMP2</th>
<th>2S1</th>
<th>ZSU234</th>
<th>BTR60</th>
<th>BRDM2</th>
<th>D7</th>
<th>T62</th>
<th>ZIL131</th>
</tr>
</thead>
<tbody>
<tr>
<td>T72</td>
<td>99.66</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td>0.00</td>
<td>0.03</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>BTR70</td>
<td>0.00</td>
<td>94.37</td>
<td>0.42</td>
<td>0.78</td>
<td>0.00</td>
<td>0.67</td>
<td>2.77</td>
<td>0.00</td>
<td>0.00</td>
<td>0.98</td>
</tr>
<tr>
<td>BMP2</td>
<td>0.48</td>
<td>0.06</td>
<td>98.91</td>
<td>0.00</td>
<td>0.00</td>
<td>0.56</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2S1</td>
<td>0.53</td>
<td>0.03</td>
<td>0.53</td>
<td>98.18</td>
<td>0.03</td>
<td>0.00</td>
<td>0.70</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ZSU234</td>
<td>0.25</td>
<td>0.64</td>
<td>0.11</td>
<td>0.00</td>
<td>96.98</td>
<td>0.42</td>
<td>0.00</td>
<td>0.70</td>
<td>0.81</td>
<td>0.08</td>
</tr>
<tr>
<td>BTR60</td>
<td>0.00</td>
<td>1.65</td>
<td>0.17</td>
<td>0.00</td>
<td>96.19</td>
<td>0.06</td>
<td>0.03</td>
<td>0.34</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>BRDM2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.42</td>
<td>0.08</td>
<td>0.22</td>
<td>99.27</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>D7</td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>99.92</td>
<td>0.03</td>
<td>0.00</td>
</tr>
<tr>
<td>T62</td>
<td>0.22</td>
<td>0.03</td>
<td>0.73</td>
<td>0.00</td>
<td>0.25</td>
<td>1.06</td>
<td>0.00</td>
<td>0.00</td>
<td>97.12</td>
<td>0.59</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0.00</td>
<td>1.96</td>
<td>0.03</td>
<td>0.14</td>
<td>0.28</td>
<td>0.00</td>
<td>1.65</td>
<td>0.00</td>
<td>0.00</td>
<td>95.94</td>
</tr>
</tbody>
</table>

average 97.65
Comparison to relevance vector machine (RVM)

Confusion matrix of RVM, with $L=120$ states in azimuth, test sequence spanning $3^\circ$ azimuth.

Sparsity: $\frac{N}{J} = 0.52$

<table>
<thead>
<tr>
<th></th>
<th>T72</th>
<th>BTR70</th>
<th>BMP2</th>
<th>2S1</th>
<th>ZSU234</th>
<th>BTR60</th>
<th>BRDM2</th>
<th>D7</th>
<th>T62</th>
<th>ZIL131</th>
</tr>
</thead>
<tbody>
<tr>
<td>T72</td>
<td>98.29</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.34</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>BTR70</td>
<td>0.00</td>
<td>94.04</td>
<td>0.56</td>
<td>0.50</td>
<td>0.00</td>
<td>0.03</td>
<td>2.63</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>BMP2</td>
<td>0.00</td>
<td>0.17</td>
<td>98.77</td>
<td>0.39</td>
<td>0.08</td>
<td>0.00</td>
<td>0.42</td>
<td>0.06</td>
<td>0.00</td>
<td>0.11</td>
</tr>
<tr>
<td>2S1</td>
<td>0.00</td>
<td>0.31</td>
<td>0.98</td>
<td>97.79</td>
<td>0.00</td>
<td>0.20</td>
<td>0.70</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ZSU234</td>
<td>1.12</td>
<td>0.28</td>
<td>0.20</td>
<td>0.64</td>
<td>96.25</td>
<td>0.08</td>
<td>0.28</td>
<td>0.00</td>
<td>0.64</td>
<td>0.50</td>
</tr>
<tr>
<td>BTR60</td>
<td>0.00</td>
<td>2.18</td>
<td>0.50</td>
<td>0.00</td>
<td>0.00</td>
<td>95.88</td>
<td>0.81</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>BRDM2</td>
<td>0.00</td>
<td>0.00</td>
<td>0.22</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>99.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>D7</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td>99.30</td>
<td>0.45</td>
<td>0.14</td>
<td>0.42</td>
</tr>
<tr>
<td>T62</td>
<td>0.92</td>
<td>0.00</td>
<td>0.20</td>
<td>0.92</td>
<td>0.70</td>
<td>0.08</td>
<td>0.06</td>
<td>0.28</td>
<td>95.04</td>
<td>0.42</td>
</tr>
<tr>
<td>ZIL131</td>
<td>0.00</td>
<td>1.04</td>
<td>0.76</td>
<td>0.87</td>
<td>2.13</td>
<td>0.42</td>
<td>0.87</td>
<td>0.42</td>
<td>1.54</td>
<td>91.83</td>
</tr>
</tbody>
</table>

average 96.65
Classification performance versus azimuth span of observations
Convergence of the M-ary KMP

Shown for the M-ary KMP in a single state
Conclusions

• The KMP is extended to the M-ary classification, with a time complexity $O(NN_c J)$ which is independent of $M$.

• The probabilistic state partition in multi-aspect HRR classification is handled in a natural way by the importance weights $\beta$.

• The results on ten MSTAR targets show that at a comparable classification rate, the M-ary KMP achieves greater sparsity and shorter training time than the RVM.
Acknowledgment

This work has been carried out under the supervision of Prof. Lawrence Carin of Duke University