Sensor-based Activity Recognition via Learning from Distributions

Hangwei Qian, Sinno Jialin Pan, Chunyan Miao

Nanyang Technological University, Singapore

February 6, 2018
Outline

- Overview of human activity recognition
- Existing solutions
- The proposed method: $\text{SMM}_{AR}$ with kernel embedding of distributions
- The proposed method: $\text{R-SMM}_{AR}$ with Random Fourier Features
- Experimental Results
- Conclusion
Human Activity Recognition

A multi-class classification problem
- Input: sensor data
- Output: activity labels
Tremendous applications:
- eldercare
- healthcare
- smart building
- gaming
Existing Feature Extraction Methods

Frame-level → vectorial-based
- Manual feature engineering, statistics of each frame

Segment-level → matrix-based
- Statistical, i.e., moments of each segment
- Structural
  - The ECDF method
  - The SAX method

<table>
<thead>
<tr>
<th></th>
<th>time_1</th>
<th>time_2</th>
<th>time_3</th>
<th>time_4</th>
<th>time_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>feature_1</td>
<td>0.9134</td>
<td>0.2785</td>
<td>0.9649</td>
<td>0.9572</td>
<td>0.8147</td>
</tr>
<tr>
<td>feature_2</td>
<td>0.9058</td>
<td>0.6324</td>
<td>0.5469</td>
<td>0.1576</td>
<td>0.4854</td>
</tr>
<tr>
<td>feature_3</td>
<td>0.127</td>
<td>0.0975</td>
<td>0.9575</td>
<td>0.9706</td>
<td>0.8003</td>
</tr>
</tbody>
</table>
Existing Feature Extraction Methods

Frame-level → vectorial-based
- Manual feature engineering, statistics of each frame

Segment-level → matrix-based
- Statistical, i.e., moments of each segment
- Structural
  - The ECDF method
  - The SAX method
Motivation

Frame-level → vectorial-based
- Manual feature engineering, statistics of each frame

Segment-level → matrix-based
- Statistical, i.e., moments of each segment
- Structural
  - The ECDF method
  - The SAX method

Can we extract as many discriminative features as possible, in an automatic fashion?
→ kernel mean embedding of distributions
→ NO information loss
(\mathbb{E}[x]) as features

problem: many distributions have the same mean!
(\mathbb{E}[x]) as features

problem: many distributions have the same mean!

\left( \begin{array}{c} \mathbb{E}[x] \\ \mathbb{E}[x^2] \end{array} \right) as features

problem: many distributions have the same mean and variance!
Motivation

\((\mathbb{E}[x])\) as features

problem: many distributions have the same mean!

\[
\begin{pmatrix}
\mathbb{E}[x] \\
\mathbb{E}[x^2] \\
\mathbb{E}[x^3]
\end{pmatrix}
\]
as features

problem: many distributions have the same mean and variance!

\[
\begin{pmatrix}
\mathbb{E}[x] \\
\mathbb{E}[x^2] \\
\mathbb{E}[x^3]
\end{pmatrix}
\]
as features

problem: many distributions still have the same first 3 moments!
The infinite dimensional features should be able to discriminate different distributions!
Two novel approaches for HAR

- **SMM$_{AR}$**: a novel feature extraction approach for activity recognition
  - input data in matrix form
  - no assumption on distributions
  - Retain all the statistics of the data in infinite dimensions

- **R-SMM$_{AR}$**: SMM$_{AR}$ with Random Fourier Features [1]
  - an accelerated approach for SMM$_{AR}$
  - comparable performance
  - solution for large-scale problems
Figure 1:
Illustrations of kernel mean embeddings of a distribution and embeddings of empirical examples

\[ \mu[P_x] = E_x[k(\cdot, x)] \]  \hspace{1cm} (1)

\[ \mu[X] = \frac{1}{m} \sum_{i=1}^{m} k(\cdot, x_i) \]  \hspace{1cm} (2)

Here \( X = \{x_1, \ldots, x_m\} \overset{i.i.d.}{\sim} P_x. \)
Kernel Mean Embeddings of Distributions

Injectivity[3]

A universal kernel $k$ can promise an injective mean map
\[ \mu : P_x \rightarrow \mu[P_x]. \]
\[
\langle \hat{\mu}_{P_x}, \hat{\mu}_{P_z} \rangle = \tilde{k}(\hat{\mu}_{P_x}, \hat{\mu}_{P_z}) = \frac{1}{n_x \times n_z} \sum_{i=1}^{n_x} \sum_{j=1}^{n_z} k(x_i, z_j), \quad (3)
\]
\[
\tilde{k}(\mu_{P_x}, \mu_{P_z}) = \langle \psi(\mu_{P_x}), \psi(\mu_{P_z}) \rangle \quad (4)
\]
Problem Formulation

- Training set: $\{(P_i, y_i)\}, i \in \{1, ..., N\}, x_i \sim P_i, x_i = \{x_{i1}, \ldots, x_{im}\}, y_i \in \{1, \ldots, L\}$
- Multi-class classifier $\rightarrow C_L^2$ binary classifiers
  $f, y = f(\phi(\mu_x)) + b$
- Primal Optimization problem:

$$\argmin_{f, b} \frac{1}{2} \|f\|_{\mathcal{H}}^2 + C \sum_{i=1}^{N} \xi_i$$

s.t. $y_i = f(\phi(\mu_{x_i})) + b$

$$y_i f(\phi(\mu_{i})) \geq 1 - \xi_i, \forall i$$

$$\xi_i \geq 0, \forall i$$

(5)
Bottleneck of SMM_{AR}: Computational Cost

\[ \langle \hat{\mu}_P, \hat{\mu}_P \rangle = \tilde{k}(\hat{\mu}_P, \hat{\mu}_P) = \frac{1}{n_x \times n_z} \sum_{i=1}^{n_x} \sum_{j=1}^{n_z} k(x_i, z_j) \]

<table>
<thead>
<tr>
<th>Datasets</th>
<th># Segments</th>
<th># Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skoda</td>
<td>1,447</td>
<td>68.8</td>
</tr>
<tr>
<td>WISDM</td>
<td>389</td>
<td>705.8</td>
</tr>
<tr>
<td>HCI</td>
<td>264</td>
<td>602.6</td>
</tr>
<tr>
<td>PS</td>
<td>1,614</td>
<td>4.0</td>
</tr>
</tbody>
</table>
Random Fourier Features (RFF)

Intuition:

\[ k(x, x') = \langle \phi(x), \phi(x') \rangle \approx z(x)^{\top} z(x') \]

**Theorem (Bochner’s Theorem [2])**

A continuous, shift-invariant kernel \( k \) is positive definite if and only if there is a finite non-negative measure \( \mathbb{P}(\omega) \) on \( \mathbb{R}^d \), such that \( k(x - x') = \int_{\mathbb{R}^d} e^{i\omega^{\top}(x-x')} d\mathbb{P}(\omega) = \int_{\mathbb{R}^d \times [0, 2\pi]} 2\cos(\omega^{\top} x + b)\cos(\omega^{\top} x' + b) d(\mathbb{P}(\omega) \times \mathbb{P}(b)) = \int_{\mathbb{R}^d} 2(\cos(\omega^{\top} x)\cos(\omega^{\top} x') + \sin(\omega^{\top} x)\sin(\omega^{\top} x')) d\mathbb{P}(\omega) \), where \( \mathbb{P}(b) \) is a uniform distribution on \([0, 2\pi]\).

**RFF:**

\[ z_w(x) = \sqrt{2}\cos(w^{\top} x + b) \]

\[ k(x, x') = \mathbb{E}(z_w(x)^{\top} z_w(x')) \]
\[ \mu[X] = \frac{1}{m} \sum_{i=1}^{m} k(\cdot, x_i) \]

\[ \mu_i = \frac{1}{m} \sum_{i=1}^{m} z(x_i) \]

\[ \min_{f} \frac{1}{n} \sum_{i=1}^{n} \ell(f(\mu_i), y_i) + \lambda \| f \|_{\mathcal{H}} \]

\[ \min_{\mathbf{w} \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^{n} \ell(\mathbf{w}^T \tilde{z}(\mu_i), y_i) + \lambda \| \mathbf{w} \|_2^2 \]

\[ f = \sum_{i=1}^{n} \alpha_i \mu_i \]

\[ f(\mu_i) = \mathbf{w}^T \mu_i \]

\[ f = \sum_{i=1}^{n} \alpha_i \psi(\mu_i) \]

\[ f(\mu_i) = \mathbf{w}^T \tilde{z}(\mu_i) \]
## Experimental Setup

### Datasets statistics

<table>
<thead>
<tr>
<th>Datasets</th>
<th># Seg.</th>
<th># En.</th>
<th># Fea.</th>
<th># C.</th>
<th>f</th>
<th># Sub.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skoda</td>
<td>1,447</td>
<td>68.8</td>
<td>60</td>
<td>10</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>WISDM</td>
<td>389</td>
<td>705.8</td>
<td>6</td>
<td>6</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>HCI</td>
<td>264</td>
<td>602.6</td>
<td>48</td>
<td>5</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>PS</td>
<td>1,614</td>
<td>4.0</td>
<td>9</td>
<td>6</td>
<td>50</td>
<td>4</td>
</tr>
</tbody>
</table>

### Baseline methods

<table>
<thead>
<tr>
<th>Category</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame-based</td>
<td>SVM-f, kNN-f</td>
</tr>
<tr>
<td>Segment-based</td>
<td>Moment-x, ECDF-d, SAX-a, miFV</td>
</tr>
</tbody>
</table>
## Experimental Results

<table>
<thead>
<tr>
<th>Methods</th>
<th>miF</th>
<th>maF</th>
<th>miF</th>
<th>maF</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMM&lt;sub&gt;AR&lt;/sub&gt;</td>
<td>100±0</td>
<td>100±0</td>
<td>96.74±1.20</td>
<td>96.72±1.22</td>
</tr>
<tr>
<td>Moment-1</td>
<td>91.35±2.28</td>
<td>91.32±2.33</td>
<td>93.90±.94</td>
<td>93.85±.93</td>
</tr>
<tr>
<td>Moment-2</td>
<td>96.47±.79</td>
<td>96.47±.77</td>
<td>95.95±.86</td>
<td>95.94±.86</td>
</tr>
<tr>
<td>Moment-5</td>
<td>97.76±.79</td>
<td>97.77±.78</td>
<td>93.31±.99</td>
<td>93.42±.93</td>
</tr>
<tr>
<td>Moment-10</td>
<td>98.72±.79</td>
<td>98.72±.79</td>
<td>91.93±1.44</td>
<td>92.00±1.36</td>
</tr>
<tr>
<td>ECDF-15</td>
<td>100±0</td>
<td>100±0</td>
<td>93.97±.96</td>
<td>94.04±.97</td>
</tr>
<tr>
<td>ECDF-30</td>
<td>100±0</td>
<td>100±0</td>
<td>90.82±.53</td>
<td>91.05±.57</td>
</tr>
<tr>
<td>ECDF-45</td>
<td>100±0</td>
<td>100±0</td>
<td>87.15±1.32</td>
<td>87.23±1.59</td>
</tr>
<tr>
<td>SAX-3</td>
<td>21.15±0</td>
<td>7.39±0</td>
<td>50.28±2.40</td>
<td>41.30±3.89</td>
</tr>
<tr>
<td>SAX-6</td>
<td>21.51±0</td>
<td>7.39±0</td>
<td>52.95±2.54</td>
<td>46.86±.68</td>
</tr>
<tr>
<td>SAX-10</td>
<td>21.15±0</td>
<td>7.39±0</td>
<td>52.81±1.08</td>
<td>44.60±1.52</td>
</tr>
<tr>
<td>miFV</td>
<td>21.64±1.58</td>
<td>18.78±2.24</td>
<td>15.32±4.28</td>
<td>7.65±5.83</td>
</tr>
<tr>
<td>SVM-f</td>
<td>99.52±.53</td>
<td>99.52±.53</td>
<td>95.22±1.10</td>
<td>95.21±1.10</td>
</tr>
<tr>
<td>kNN-f</td>
<td>99.04±1.22</td>
<td>99.05±1.21</td>
<td>94.73±.65</td>
<td>94.72±.65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methods</th>
<th>Skoda</th>
<th>WISDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMM&lt;sub&gt;AR&lt;/sub&gt;</td>
<td>99.61±.24</td>
<td>99.60±.25</td>
</tr>
<tr>
<td>Moment-1</td>
<td>92.46±1.97</td>
<td>92.39±2.01</td>
</tr>
<tr>
<td>Moment-2</td>
<td>92.27±1.47</td>
<td>92.14±1.49</td>
</tr>
<tr>
<td>Moment-5</td>
<td>94.49±1.66</td>
<td>94.45±1.70</td>
</tr>
<tr>
<td>Moment-10</td>
<td>95.24±.63</td>
<td>95.23±.64</td>
</tr>
<tr>
<td>ECDF-15</td>
<td>93.62±1.34</td>
<td>93.60±1.36</td>
</tr>
<tr>
<td>ECDF-30</td>
<td>93.25±1.11</td>
<td>93.21±1.15</td>
</tr>
<tr>
<td>ECDF-45</td>
<td>92.20±1.07</td>
<td>92.20±1.13</td>
</tr>
<tr>
<td>SAX-3</td>
<td>94.54±1.28</td>
<td>94.48±1.21</td>
</tr>
<tr>
<td>SAX-6</td>
<td>96.13±1.57</td>
<td>96.10±1.55</td>
</tr>
<tr>
<td>SAX-10</td>
<td>96.22±.84</td>
<td>96.18±.83</td>
</tr>
<tr>
<td>miFV</td>
<td>61.40±3.24</td>
<td>53.63±2.50</td>
</tr>
<tr>
<td>SVM-f</td>
<td>93.46±1.20</td>
<td>92.65±1.38</td>
</tr>
<tr>
<td>kNN-f</td>
<td>93.17±1.44</td>
<td>92.93±1.45</td>
</tr>
</tbody>
</table>
Impact on Orders of Moments

The graph shows the impact on orders of moments for different moment orders and frequencies (3Hz, 24Hz, 48Hz, 96Hz) with a log scale for miF (%).

- The x-axis represents the moment order (1 to 10).
- The y-axis shows miF (%) with a log scale.
- Different frequencies are represented by different markers.

The graph illustrates how the moments of different orders and frequencies affect the miF (%) values.
Experimental Results

- **miF (%, log scale)**
  - **R-SMM**
  - **AR**
  - **SMM**
  - **AR**

- **Run time (s)**
  - **R-SMM**
  - **AR**
  - **SMM**
  - **AR**

- **Random feature dimension D**
We propose two novel methods for HAR:

- **SMM$_{AR}$**: a novel feature extraction approach for activity recognition
  - input data in matrix form
  - retain all the statistics in the data

- **R-SMM$_{AR}$**: an accelerated version of SMM$_{AR}$
  - an accelerated approach for SMM$_{AR}$
  - comparable performance
  - solution for large-scale problems
Thank You!

More info in http://hangwei12358.github.io/
