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: SMM_{AR} with kernel embedding of distributions
- The proposed method: 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

	time_1	time_2	time_3	time_4	time_5
feature_1	0.9134	0.2785	0.9649	0.9572	0.8147
feature_2	0.9058	0.6324	0.5469	0.1576	0.4854
feature_3	0.127	0.0975	0.9575	0.9706	0.8003

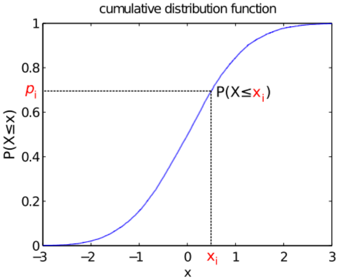
Existing Feature Extraction Methods

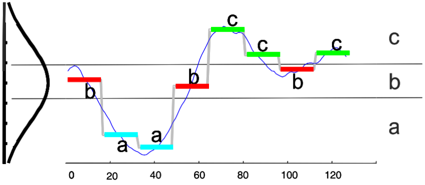
Frame-level → vectorial-based

- 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 $\rightarrow$ vectorial-based

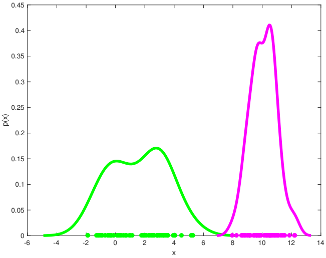
- Manual feature engineering, statistics of each frame

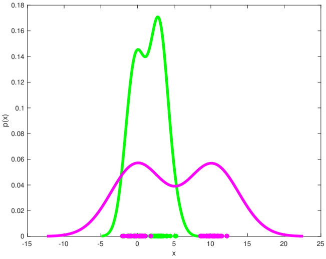
Segment-level $\rightarrow$ matrix-based

- 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?
$\rightarrow$ kernel mean embedding of distributions
$\rightarrow$ NO information loss

Motivation

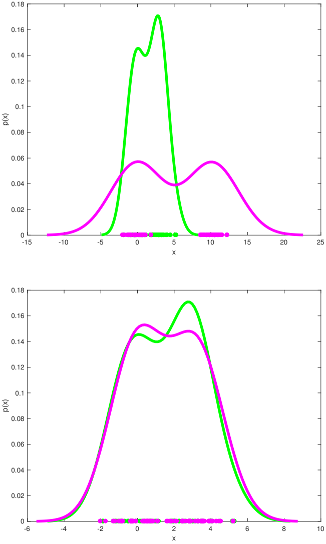




$(\mathbb{E}[x])$ as features

problem: many distributions have the same mean!

Motivation



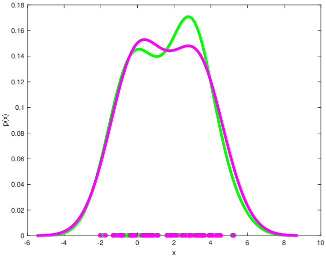
$(\mathbb{E}[x])$ as features

problem: many distributions have the same mean!

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

problem: many distributions have the same mean and variance!

Motivation



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

problem: many distributions have the same mean!

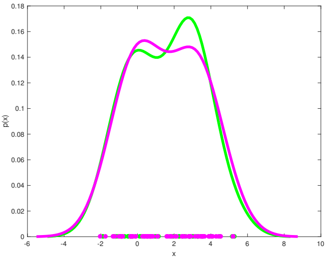
$\begin{pmatrix} \mathbb{E}[x] \\ \mathbb{E}[x^2] \end{pmatrix}$ as features

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!

Motivation



$$\mu[P_x] = \begin{pmatrix} \mathbb{E}[x] \\ \mathbb{E}[x^2] \\ \mathbb{E}[x^3] \\ \dots \\ \dots \end{pmatrix}$$

The **infinite dimensional features** should be able to discriminate different distributions!

Contribution

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

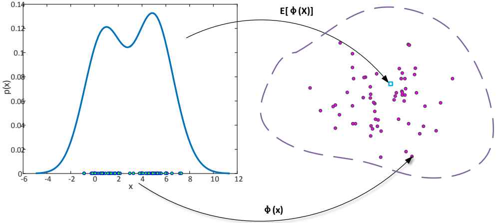


Figure 1: Illustrations of kernel mean embeddings of a distribution and embeddings of empirical examples

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

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

Here $X = \{x_1, ..., x_m\} \overset{i.i.d.}{\sim} P_x$.

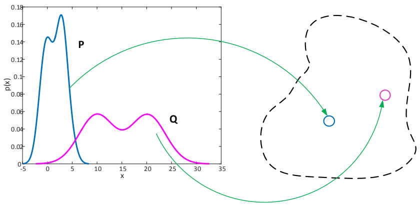


Figure 2: Illustration of the kernel mean embedding of two different distributions

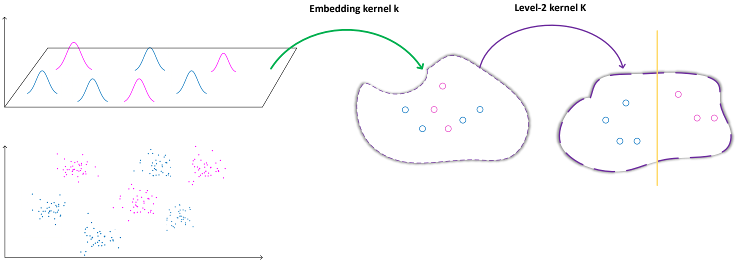
Injectivity[3]

A universal kernel k can promise an injective mean map $\mu : P_x \to \mu[P_x]$.

Learning from Distributions

$$\langle \hat{\boldsymbol{\mu}}_{\mathbb{P}_x}, \hat{\boldsymbol{\mu}}_{\mathbb{P}_z} \rangle = \tilde{k}(\hat{\boldsymbol{\mu}}_{\mathbb{P}_x}, \hat{\boldsymbol{\mu}}_{\mathbb{P}_z}) = \frac{1}{n_x \times n_z} \sum_{i=1}^{n_x} \sum_{j=1}^{n_z} k(\mathbf{x}_i, \mathbf{z}_j), \tag{3}$$

$$\tilde{k}(\boldsymbol{\mu}_{\mathbb{P}_x}, \boldsymbol{\mu}_{\mathbb{P}_z}) = \langle \psi(\boldsymbol{\mu}_{\mathbb{P}_x}), \psi(\boldsymbol{\mu}_{\mathbb{P}_z}) \rangle \tag{4}$$



Problem Formulation

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

$$\begin{aligned} &\underset{f,b}{argmin} \frac{1}{2}\|f\|_{\mathcal{H}}^2 + C\sum_{i=1}^{N}\xi_i \\ &s.t. y_i = f(\phi(\mu_{x_i})) + b \\ &\quad y_i f(\phi(\mu_i)) \geq 1 - \xi_i, \forall i \\ &\quad \xi_i \geq 0, \forall i \end{aligned} \tag{5}$$

Bottleneck of SMM_{AR}: Computational Cost

$$\langle \hat{\boldsymbol{\mu}}_{\mathbb{P}_x}, \hat{\boldsymbol{\mu}}_{\mathbb{P}_z} \rangle = \tilde{k}(\hat{\boldsymbol{\mu}}_{\mathbb{P}_x}, \hat{\boldsymbol{\mu}}_{\mathbb{P}_z}) = \frac{1}{n_x \times n_z} \sum_{i=1}^{n_x} \sum_{j=1}^{n_z} k(\mathbf{x}_i, \mathbf{z}_j)$$

Datasets	# Segments	# Entries
Skoda	1,447	68.8
WISDM	389	705.8
HCI	264	602.6
PS	1,614	4.0

Random Fourier Features (RFF)

Intuition:

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

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 $k(\mathbf{x} - \mathbf{x}') = \int_{\mathbb{R}^d} e^{i\omega^\top(\mathbf{x}-\mathbf{x}')} d\mathbb{P}(\omega) = \int_{\mathbb{R}^d \times [0,2\pi]} 2cos(\omega^\top \mathbf{x} + b)cos(\omega^\top \mathbf{x}' + b)d(\mathbb{P}(\omega) \times \mathbb{P}(b)) = \int_{\mathbb{R}^d} 2(cos(\omega^\top \mathbf{x})cos(\omega^\top \mathbf{x}') + sin(\omega^\top \mathbf{x})sin(\omega^\top \mathbf{x}'))d\mathbb{P}(\omega)$, where $\mathbb{P}(b)$ is a uniform distribution on $[0, 2\pi]$.

RFF:

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

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}(z_w(\mathbf{x})^\top z_w(\mathbf{x}'))$$

SMM_{AR} VS R-SMM_{AR}

$$\mu[X] = \frac{1}{m}\sum_{i=1}^{m} k(\cdot, x_i) \qquad\qquad \boldsymbol{\mu}_i = \frac{1}{m}\sum_{i=1}^{m} \mathbf{z}(\mathbf{x}_i)$$

$$\min_f \frac{1}{n}\sum_{i=1}^{n} \ell(f(\boldsymbol{\mu}_i), y_i) + \lambda\|f\|_{\tilde{\mathcal{H}}} \qquad\qquad \min_{\mathbf{w}\in\mathbb{R}^{\tilde{D}}} \frac{1}{n}\sum_{i=1}^{n} \ell(\mathbf{w}^\top\tilde{\mathbf{z}}(\boldsymbol{\mu}_i), y_i) + \lambda\|\mathbf{w}\|_2^2$$

$$f = \sum_{i=1}^{n} \alpha_i \boldsymbol{\mu}_i \qquad\qquad f(\boldsymbol{\mu}_i) = \mathbf{w}^\top\boldsymbol{\mu}_i$$

$$f = \sum_{i=1}^{n} \alpha_i \psi(\boldsymbol{\mu}_i) \qquad\qquad f(\boldsymbol{\mu}_i) = \mathbf{w}^\top\tilde{\mathbf{z}}(\boldsymbol{\mu}_i)$$

Experimental Setup

Datasets statistics

Datasets	# Seg.	# En.	# Fea.	# C.	f	# Sub.
Skoda	1,447	68.8	60	10	14	1
WISDM	389	705.8	6	6	20	36
HCI	264	602.6	48	5	96	1
PS	1,614	4.0	9	6	50	4

Baseline methods

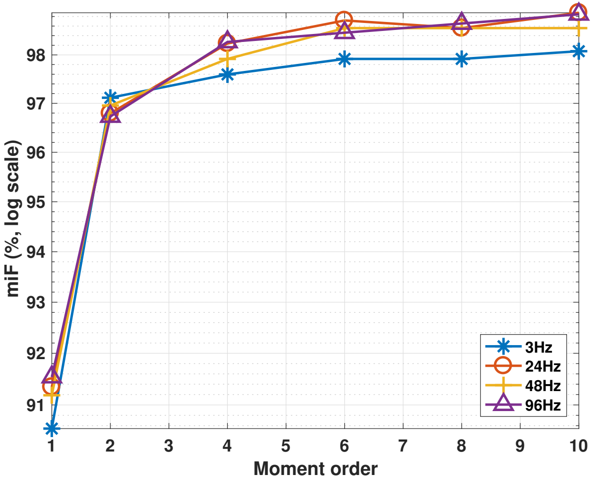
Category	Methods
Frame-based	SVM-f kNN-f
Segment-based	Moment-*x* ECDF-*d* SAX-*a* miFV

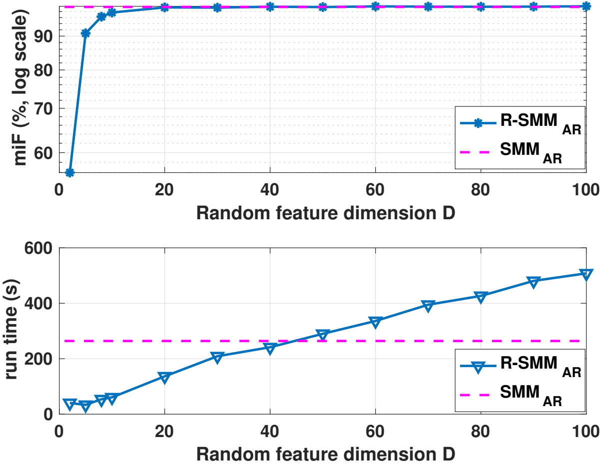
Experimental Results

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

Methods	Skoda miF	Skoda maF	WISDM miF	WISDM maF
SMM_{AR}	**99.61±.24**	**99.60±.25**	55.87±2.66	56.09±3.03
Moment-1	92.46±1.97	92.39±2.01	38.30±4.10	44.63±12.22
Moment-2	92.27±1.47	92.14±1.49	52.55±1.46	57.21±7.22
Moment-5	94.49±1.66	94.45±1.70	57.31±5.91	**62.52±9.81**
Moment-10	95.24±.63	95.23±.64	**57.79±3.97**	62.44±8.02
ECDF-15	93.62±1.34	93.60±1.36	54.01±3.09	57.47±7.65
ECDF-30	93.25±1.11	93.21±1.15	55.33±4.50	58.26±7.13
ECDF-45	92.20±1.07	92.20±1.13	53.46±2.84	57.77±7.02
SAX-3	94.54±1.28	94.48±1.21	32.90±1.47	23.62±1.81
SAX-6	96.13±1.57	96.10±1.55	35.49±3.11	28.77±2.82
SAX-10	96.22±.84	96.18±.83	32.57±1.48	26.89±2.39
miFV	61.40±3.24	53.63±2.50	14.61±2.04	4.72±2.13
SVM-f	93.46±1.20	92.65±1.38	27.49±2.71	18.70±2.88
kNN-f	93.17±1.44	92.93±1.45	28.48±2.15	17.96±2.84

Impact on Orders of Moments





miF (%, log scale)
90
80
70
60
0
20
40
60
80
100
Random feature dimension D
R-SMM AR
SMM AR
run time (s)
600
400
200
0
0
20
40
60
80
100
Random feature dimension D
R-SMM AR
SMM AR

Conclusion

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

Questions?

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

References I

Ali Rahimi and Benjamin Recht. "Random Features for Large-Scale Kernel Machines". In: *NIPS*. 2007, pp. 1177–1184.

Walter Rudin. *Fourier analysis on groups*. Courier Dover Publications, 2017.

Alex Smola et al. "A Hilbert space embedding for distributions". In: *International Conference on Algorithmic Learning Theory*. Springer. 2007, pp. 13–31.