A Novel Distribution-Embedded Neural Network for Sensor-Based Activity Recognition (IJCAI-19)

Hangwei Qian, Sinno Jialin Pan,
Bingshui Da, Chunyan Miao

Nanyang Technological University, Singapore

August 14, 2019

Outline

Human Activity Recognition

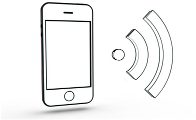
Tremendous applications:

- elderly assistant
- healthcare
- fitness coaching
- smart building
- gaming

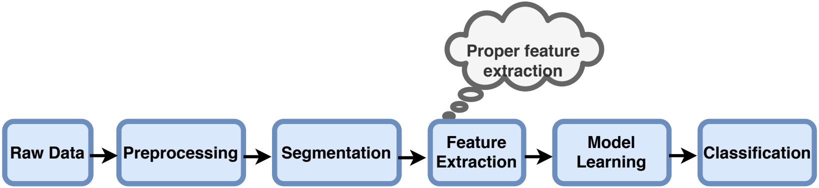
Human Activity Recognition

A multi-class classification problem

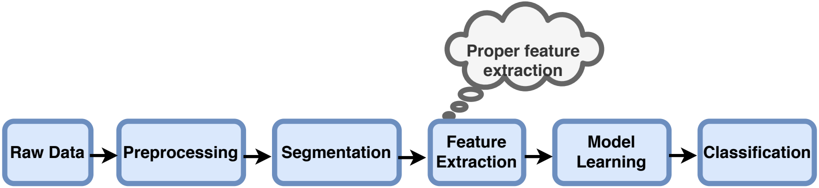
- Input: wearable onbody sensor data $\mathbf{X}^{d \times N}$
- Output: activity labels $y^{N \times 1} \in \{1, ..., n_c\}$



Problem Overview



Problem Overview



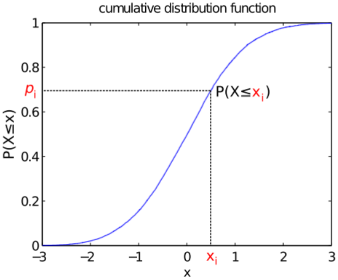
Desired feature extraction:

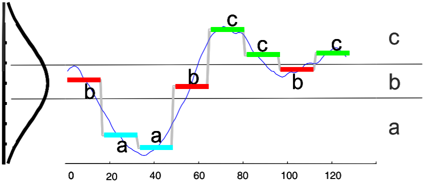
- automatic
- informative

Outline

Existing Feature Extraction Methods

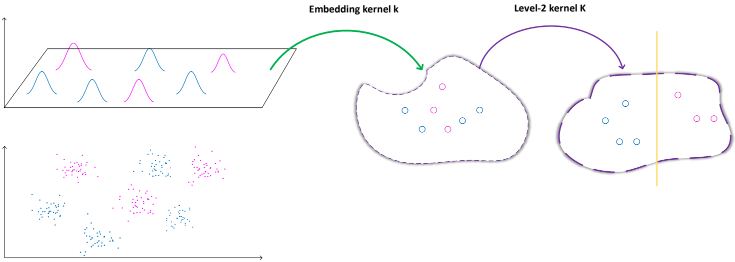
1. Feature engineering
   - handcrafted, PCA, Fourier transform
   - statistical, structural





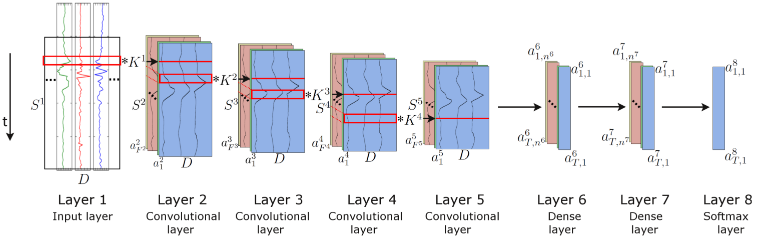
Existing Feature Extraction Methods

1. Feature engineering
   - handcrafted, PCA, Fourier transform
   - statistical, structural
2. Automatic
   - SMM_{AR}, R-SMM_{AR} [2]: all statistical features with kernel mean embedding of distributions



Existing Feature Extraction Methods

1. Feature engineering
   - handcrafted, PCA, Fourier transform
   - statistical, structural
2. Automatic
   - SMM_{AR}, R-SMM_{AR} [2]: all statistical features with kernel mean embedding of distributions
   - Deep learning models
     - CNNs [3]
     - DeepConvLSTM [1]: 4 CNNs + 2 RNNs



Existing Feature Extraction Methods

1. Feature engineering
   - handcrafted, PCA, Fourier transform
   - statistical, structural
2. Automatic
   - SMM_{AR}, R-SMM_{AR} [2]: all statistical features with kernel mean embedding of distributions
   - Deep learning models
     - CNNs [3]
     - DeepConvLSTM [1]: 4 CNNs + 2 RNNs
3. Combinations
   - deep features, shallow classifiers
   - joint learning of shallow and deep features

Existing Feature Extraction Methods

1. Feature engineering
   - handcrafted, PCA, Fourier transform
   - statistical, structural

   ✔ classic, domain knowledge incorporated

   ✗ largely dependent on domain expert
2. Automatic
   - SMM_{AR}, R-SMM_{AR} [2]: all statistical features with kernel mean embedding of distributions

     ✔ in an automated fashion, proven to be powerful

     ✗ i.i.d. assumption on data
   - Deep learning models
     - CNNs [3]
     - DeepConvLSTM [1]: 4 CNNs + 2 RNNs

     ✔ additive and automatic

     ✗ not specifically designed for activity data

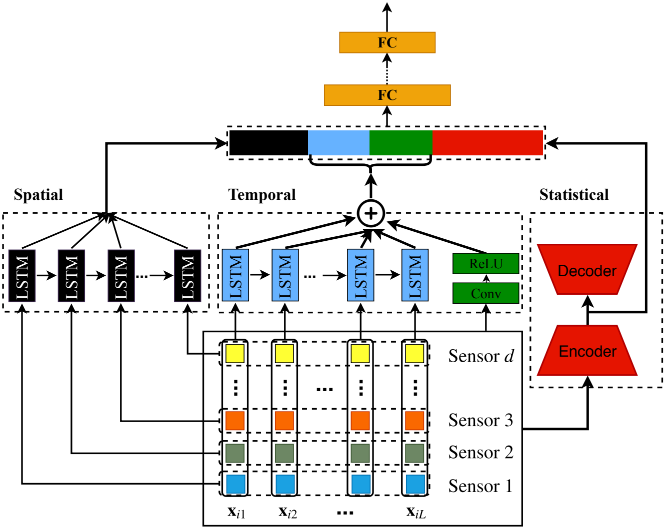
Outline

Contribution

DDNN: **D**istribution-embedded **D**eep **N**eural **N**etwork

1. Encode the idea of kernel mean embedding of distributions into a deep architecture
2. Automatically extract statistical, temporal and spatial features in an end-to-end model
3. Extensive experiments to show the efficacy of DDNN

DDNN Framework



Statistical Module f_1

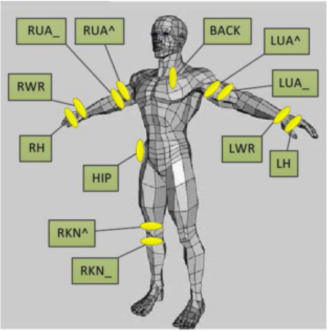
- Design a neural network f_1 to learn the statistical feature mapping $\phi_{f_1}(\cdot)$ automatically
  $f_1(\mathbf{X}_i) = \phi_{f_1}(\mathbf{X}_i) = \frac{1}{L}\sum_{j=1}^{L} \phi_k(\mathbf{x}_{ij})$,
- Injectivity $f_1^{-1}(f_1(\mathbf{X}_i)) = \mathbf{X}_i \rightarrow$ autoencoder
- Reconstruction error $\ell_{\text{ae}} = \|\mathbf{X}_i - f_d(f_e(\mathbf{X}_i))\|$
- An extra loss

$$\ell_{\text{MMD}}(\mathbf{X}_i, f_d(f_e(\mathbf{X}_i))) = \frac{1}{L}\left\|\sum_{j=1}^{L} f_e(\mathbf{x}_{ij}) - f_e(f_d(f_e(\mathbf{x}_{ij})))\right\|_2 \qquad (1)$$

Spatial Module f_2

Spatial Module f_2:

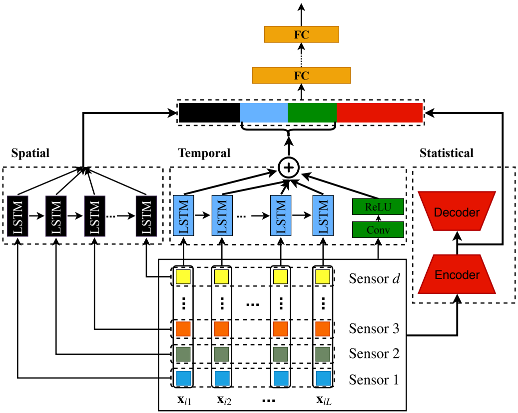
- Standard CNNs on activity data are along the temporal aspect [3]
- A different perspective: spatial correlations and constraints between sensor placements



More of DDNN Framework

f_1, f_2, f_3: statistical, spatial, temporal module

- $[f_1(\mathbf{X}_i), f_2(\mathbf{X}_i), f_3(\mathbf{X}_i)]$
- Possible alternative: $f_1([f_2(\mathbf{X}_i), f_3(\mathbf{X}_i)])$.
- Possible alternative: build a deeper model with these three modules as atom building blocks.



Outline

Experimental Setup

- Evaluation: micro-F1 (miF), macro-F1 (maF)
- 4 datasets

Datasets	# train	# val.	# test	# sw	# Feature	# Class	# Subjects
OPPOR[1]	715,785	32,224	121,378	30	113	18	4
UCIHAR[2]	941,056	NA	377,216	128	9	6	30
DG[3]	312,970	37,122	30,188	32	9	2	10
PAMAP2[4]	473,447	90,814	83,366	170	52	12	9

[1] OPPOR classes: Open then close the fridge/dishwasher/drawers/doors, toggle the lights on then off, drink while standing/seated, drink coffee, etc

[2] UCIHAR classes: walking, walking upstairs, walking downstairs, sitting, standing, laying

[3] DG classes: no freeze, freeze

[4] PAMAP2 classes: household activities and exercise activities

Baselines

Ablation study:

- DDNN$-f_1$: the proposed deep model without the statistical module. This baseline is set to investigate the efficacy of the statistical module.
- DDNN$-f_2$: the proposed deep model without the spatial module. This baseline is set to investigate the efficacy of the spatial module.

Baselines

Ablation study:

- DDNN$-f_1$: the proposed deep model without the statistical module. This baseline is set to investigate the efficacy of the statistical module.
- DDNN$-f_2$: the proposed deep model without the spatial module. This baseline is set to investigate the efficacy of the spatial module.

State-of-the-art:

- CNN_Yang [3]: a state-of-the-art CNN-based model with 3 convolutional layers. We follow the architecture in the paper and reproduce the model.
- DeepConvLSTM [1]: a state-of-the-art model with 4 convolutional layers and 2 LSTM layers. We also follow the architecture and reproduce the model.

Baselines

Ablation study:

- DDNN$-f_1$: the proposed deep model without the statistical module. This baseline is set to investigate the efficacy of the statistical module.
- DDNN$-f_2$: the proposed deep model without the spatial module. This baseline is set to investigate the efficacy of the spatial module.

State-of-the-art:

- CNN_Yang [3]: a state-of-the-art CNN-based model with 3 convolutional layers. We follow the architecture in the paper and reproduce the model.
- DeepConvLSTM [1]: a state-of-the-art model with 4 convolutional layers and 2 LSTM layers. We also follow the architecture and reproduce the model.

Other baselines:

- DNN, CNN, LSTM, LSTM-f, LSTM-S, b-LSTM-S

Experimental Results

	DG		OPPOR		UCIHAR		PAMAP2	
Methods	miF	maF	miF	maF	miF	maF	miF	maF
DDNN	**92.59**	**91.61**	**83.66**	86.01	**90.53**	**90.58**	**93.23**	**93.38**
DDNN$-f_1$	91.38	90.67	81.27	84.51	89.96	89.93	87.49	86.84
DDNN$-f_2$	89.67	88.97	77.96	82.27	88.60	88.58	89.37	89.43
CNN_Yang	87.96	86.65	9.98	2.95	88.12	88.11	70.17	70.46
DeepConvLSTM	87.21	84.28	75.47	78.92	89.05	89.07	84.31	82.73
DNN	88.91	86.47	77.05	80.25	87.65	87.72	80.31	79.82
CNN	89.23	88.85	10.66	3.56	86.66	86.77	89.75	89.72
LSTM	88.34	86.93	63.17	69.92	74.52	74.75	90.38	90.29
LSTM-f$*$	67.3	-	67.2	90.8	-	-	92.9	-
LSTM-S$*$	76.0	-	69.8	91.2	-	-	88.2	-
b-LSTM-S$*$	74.1	-	74.5	**92.7**	-	-	86.8	-

Codes

https://github.com/Hangwei12358/IJCAI2019_DDNN

Outline

Conclusion

Distribution-embedded Deep Neural Network (DDNN) for human activity recognition

1. Three modules learning statistical, spatial and temporal features
2. End-to-end fashion
3. Extensive experiments to show the efficacy of DDNN

References

[1] Francisco Javier Ordez Morales and Daniel Roggen. "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition". In: *Sensors* 16.1 (2016), p. 115. DOI: `10.3390/s16010115`. URL: `https://doi.org/10.3390/s16010115`.

[2] Hangwei Qian, Sinno Jialin Pan, and Chunyan Miao. "Sensor-Based Activity Recognition via Learning From Distributions". In: *AAAI*. 2018.

[3] Jianbo Yang et al. "Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition". In: *IJCAI*. AAAI Press, 2015, pp. 3995–4001.