Human Activity Recognition and Study of Dynamic-Filter-Networks for Position-aware Detection

Master Thesis

Yuvaraj Govindarajulu

Supervisor : Mario Doebler

Institute of Signal Processing and System Theory

Master of Science, INFOTECH

University of Stuttgart

19.11.2019

Table of Contents

Table of Contents

Motivation

- Increase in global average life expectancy
  - By 5.5years between 2000 and 2016 (GHO data from WHO [2])
  - Percentage of population aged over 65years will double between 2019 and 2050 [4]

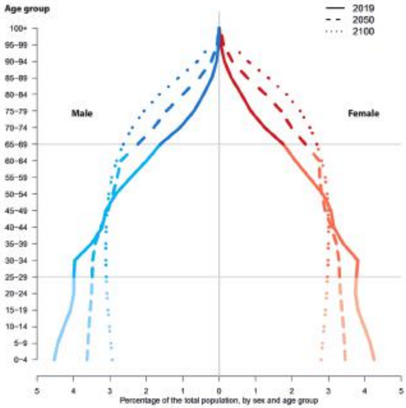

Figure: Age distribution of world population - 1950-2050 [4]

Motivation

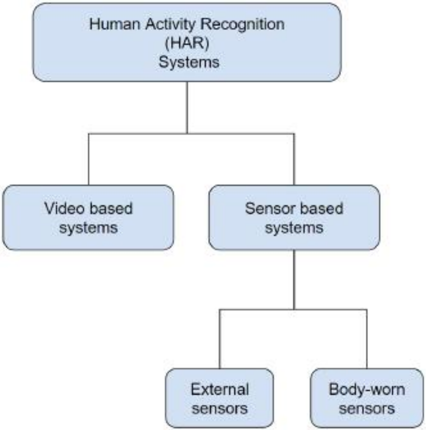

Figure: Classification of HAR systems

Figure: Placement of body worn sensors [3]

Table of Contents

Datasets

- Datasets:
  - Human Activities and Postural Transition Dataset (HAPT)
  - RealWorld (HAR) Dataset

Table: Comparison of *HAPT* and *RealWorld (HAR)* datasets

	HAPT	*RealWorld (HAR)*
Number of activities	12	8
Number of subjects	30	15
Age group	19 to 48 years	19 to 44 years
Number of sequences	3126	54592
Train/test/validation ratio	2126/625/375	38720/10512/5360
Number of positions	1	7
Postural transitions	Yes	No
Number of sensors recorded	2	6
Environment setting	Indoor	Real world scenario

Sliding window

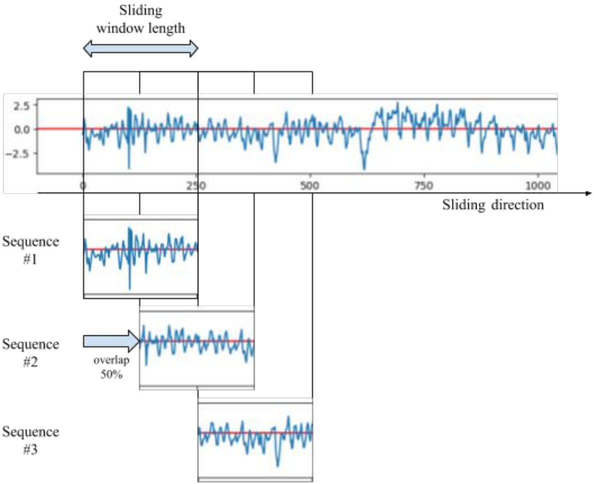

Figure: Example of sliding window with overlap

Types of Classification tasks

Sequence-to-Sequence (S2S) and Sequence-to-Label Classification tasks (S2L)

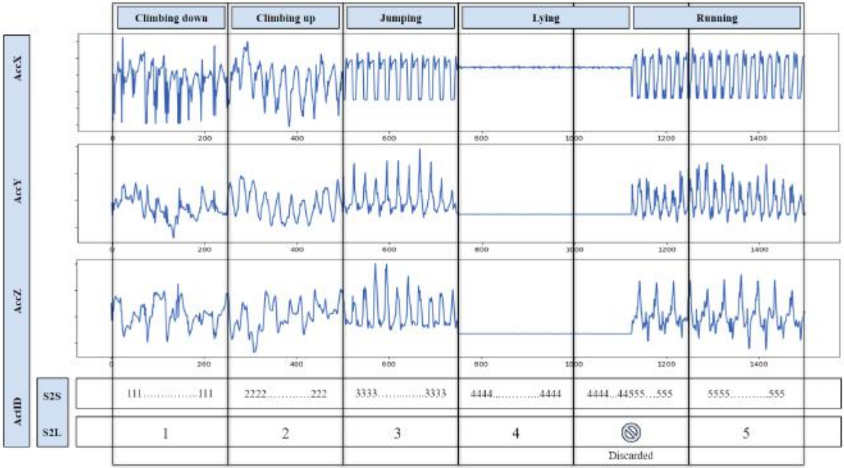

Figure: Classification task types

Architecture

- Architecture used for
  - Position specific activity recognition
  - Comparison of S2S and S2L classification tasks
  - Position recognition

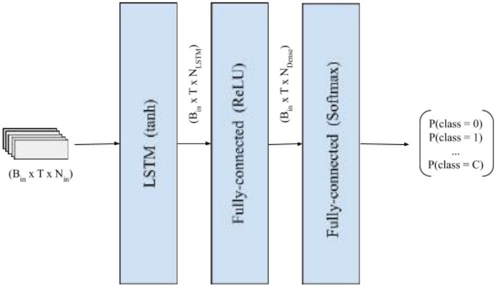

Figure: LSTM based architecture

Table of Contents

Existing Work

- Feature extraction methods
- Multi-level classification
- Identification of static and dynamic activities
- Random Forest Classifier

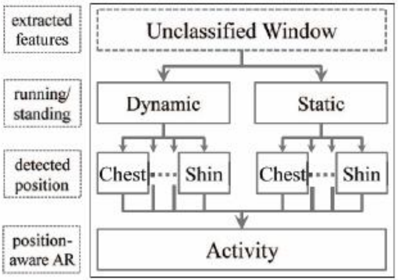

Figure: Multi-level classification [3]

Dynamic Filter Networks

- Framework where filters are generated dynamically depending on the input
- Contains two parts
  - Filter-Generating network
  - Dynamic Filtering layer
- Generated filters can be applied
  - Globally (Dynamic convolution)
  - Locally (Dynamic local filtering)

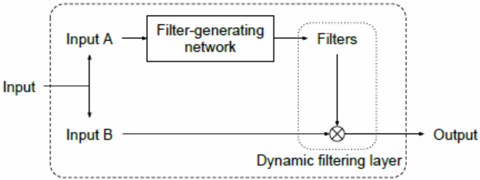

Figure: Block Diagram of Dynamic filter network [1]

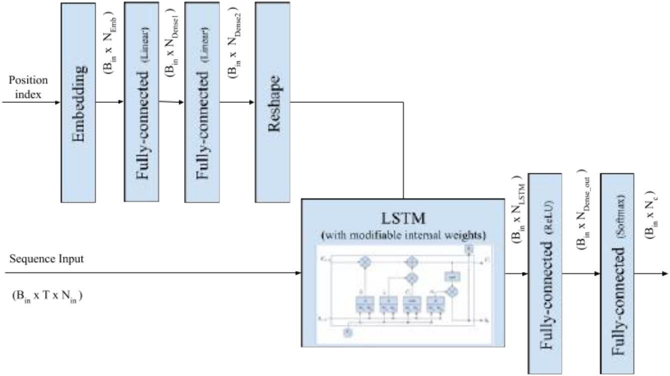

Figure: FGNs for LSTM Internal Weights

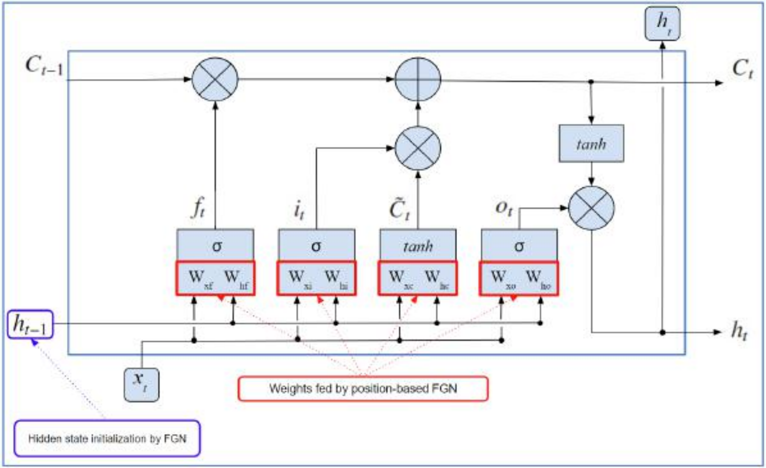

Figure: Structure of LSTM cell

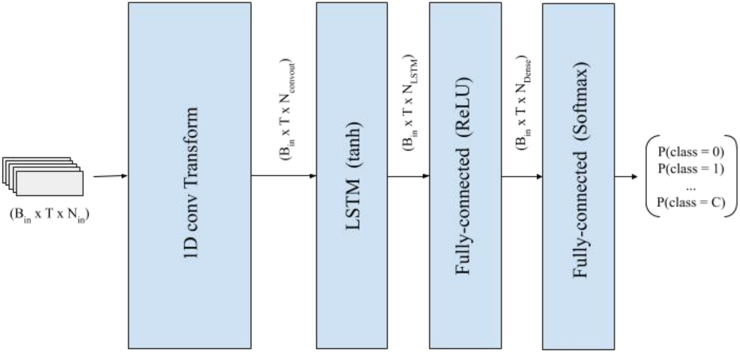

Figure: 1D convolution transformation

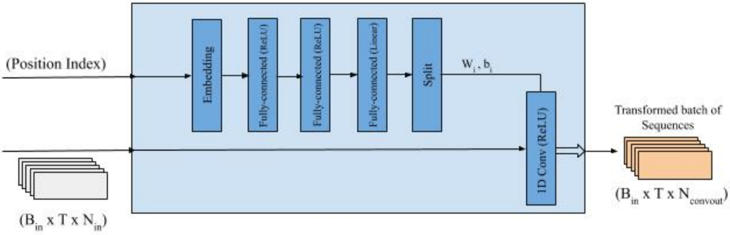

Figure: Sequence transformation using 1D convolution

Table of Contents

Results - Position specific HAR

Table: HAR for various classification types (Position-shin)

Parameters	S2S-WinSize-250,Shift-125	S2S-WinSize-125,Shift-62	S2L-WinSize-250,Shift-125
Training Bal. Accuracy(%)	99.14	97.67	97.45
Test Accuracy(%)	80.81	77.99	73.95

Table: Results for S2S classification - position wise

Parameters	Chest	Head	Shin	Upperarm	Waist
Final Training Accuracy(%)	87.88	73.66	86.09	75.43	85.91
Test Accuracy(%)	65.1	69.14	83.4	68.2	79.51

Results - Position specific HAR

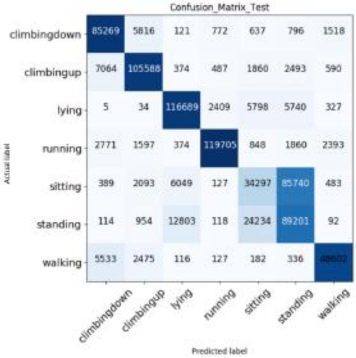

Figure: Confusion matrix - Shin Position

Results - Position aware HAR

Table: Position-aware HAR using FGN for LSTM weights

Parameters	Baseline (NoFGN)	FGN for Hidden init.	FGN for output gate
Training Bal. Accuracy(%)	77.15	67.13	72.06
Test Accuracy(%)	49.56	45.82	46.65

Table: Position-aware HAR using 1D conv sequence transformation

Parameters	Baseline (NoFGN)	with FGN for transformation
Training Bal. Accuracy(%)	87.68	91.50
Test Accuracy(%)	67.80	68.21

Conclusion and Future work

- Position-specific Human activity recognition
  - Sliding window with overlap of 50% show the best results
  - Sitting and Standing activities are commonly mis-classified (for all positions)
- Position-aware Human activity recognition
  - Generating position-index based weights for LSTM internal gates inhibits the cell from learning temporal dependencies
  - FGN with 1D input transformation has improved accuracy than the equivalent network without FGN
- Future Work
  - Use of gravitational features to initialize hidden state of LSTM
  - Appending position information as part of the latent vector output of the LSTM layer

References

Bert De Brabandere et al. "Dynamic Filter Networks". In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. URL: http://dl.acm.org/citation.cfm?id=3157096.3157171.

World Health Organization. *Global Health Observatory (GHO) data*. https://www.who.int/gho/mortality_burden_disease/life_tables/situation_trends_text/en/. [accessed 01-11-2019]. 2019.

Timo Sztyler and Heiner Stuckenschmidt. "On-body Localization of Wearable Devices: An Investigation of Position-Aware Activity Recognition". In: *2016 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 2016.

Department of Economic United Nations and Population Division (2019) Social Affairs. "World Population Prospects 2019: Press Release". In: (2019).

Human Activity Recognition and Study of Dynamic-Filter-Networks for Position-aware Detection

Thank you

Distribution of subjects

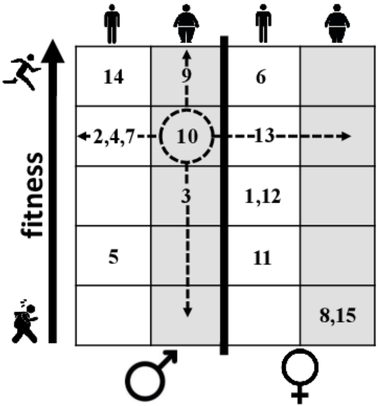

Figure: Distribution of subjects based on fitness and physical characteristics

Backup : LSTM Implementation

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{2}$$

$$\tilde{C}_t = tanh(W_{xc}x_t + W_{hc}h_{t-1}) \tag{3}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \tag{4}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{5}$$

$$h_t = tanh(C_t) \circ o_t \tag{6}$$

Backup : Variants of LSTM

- Sepp Hochreiter and Jurgen Schmidhuber. *Long Short-Term Memory* - `http://www.bioinf.jku.at/publications/older/2604.pdf`, 1997
- Felix A. Gers, Jurgen Schmidhuber, and Fred Cummins. *Learning to forget: Continual prediction with LSTM*, 1999
- Felix A. Gers and Jurgen Schmidhuber. *Recurrent nets -that time and count*, 2000
- Alex Graves and Jurgen Schmidhuber, *Framewise Phoneme Classification with Bidirectional LSTM and Other Neural Network Architectures*, 2005
- Klaus Greff et. al, *LSTM : A Search Space Odyssey* `https://arxiv.org/pdf/1503.04069.pdf`, 2017

Backup : Data preprocessing

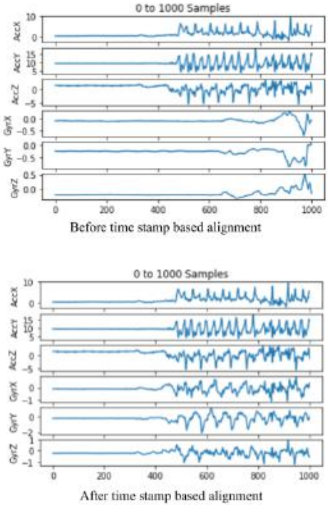

Figure: Timestamp based alignment -before and after

Backup : Z-Score Normalization

Z-Score Normalization :
Resulting time-series after normalization will have zero mean and unit variance

$$x_i' = \frac{x_i - \mu_x}{\sigma_x}, \forall i$$