Data Science for Assistive Health Technologies

Dr. Muhammad Adeel Nisar

Assistant Professor – Department of IT,
Faculty of Computing and Information Technology,
University of the Punjab, Lahore

Contents

Acknowledgement

- The slides material is acquired from the lectures/publications of
- Andrew Ng, Stanford University,
- Fei-Fei Li, Stanford University,
- Alexandar and Ava Amini, MIT
- Pascal Poupart, University of Waterloo
- Adeel Nisar, Kimiaki Shirahama, Marcin Grzegorzek University of Siegen

Sequence to Sequence Models

German to English Translation:

Karen besuchte Europa im September.

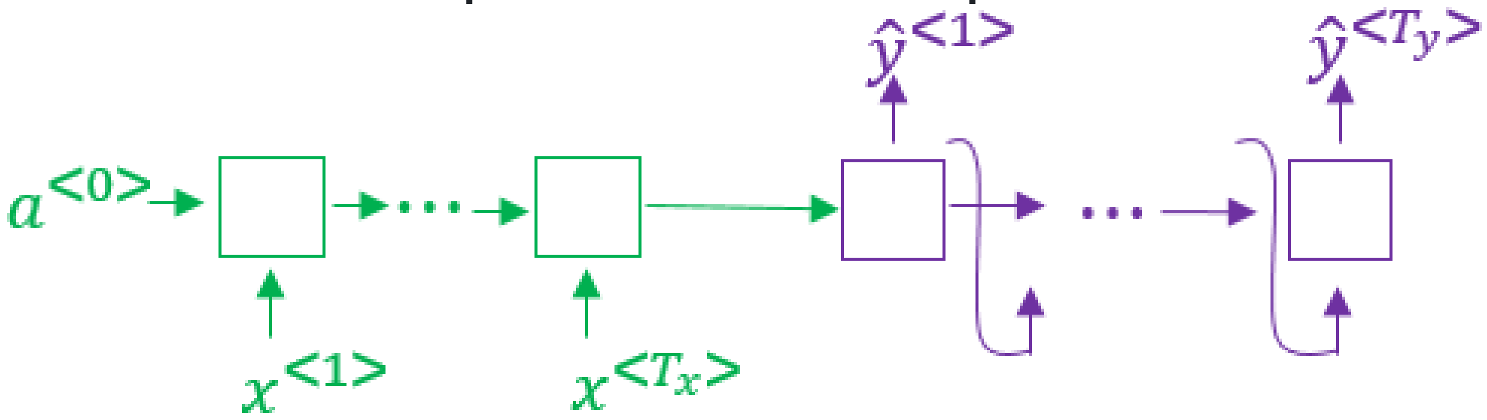
$x^{<1>}, x^{<2>}, x^{<3>}, x^{<4>}, x^{<5>}$

Karen visited Europe in September.

$y^{<1>}, y^{<2>}, y^{<3>}, y^{<4>}, y^{<5>}$

Encoder-Decoder

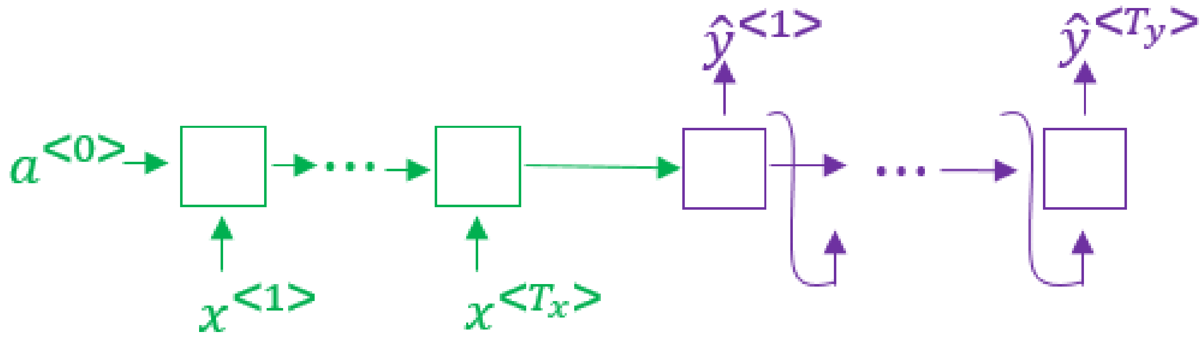
- Our architecture will include **encoder** and **decoder**.
- The encoder is RNN - LSTM or GRU are included - and takes the input sequence and then outputs a vector that should represent the whole input.
- After that the decoder network, also RNN, takes the sequence built by the encoder and outputs the new sequence.



Sutskever et al., 2014. Sequence to sequence learning with neural networks
Cho et al., 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation

The problem of Long Sequences

- Encoder-Decoder models work well for short sequences.
- If the sequences are very long, the performance degrades
- It is harder for an RNN to memorize a long sequence



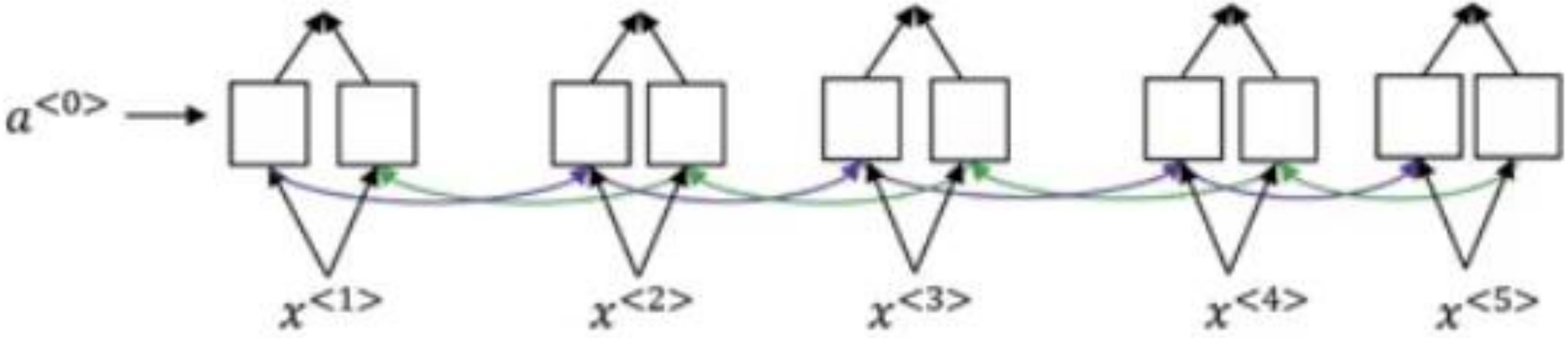
Jane s'est rendue en Afrique en septembre dernier, a apprécié la culture et a rencontré beaucoup de gens merveilleux; elle est revenue en parlant comment son voyage était merveilleux, et elle me tente d'y aller aussi.

Jane went to Africa last September, and enjoyed the culture and met many wonderful people; she came back raving about how wonderful her trip was, and is tempting me to go too.

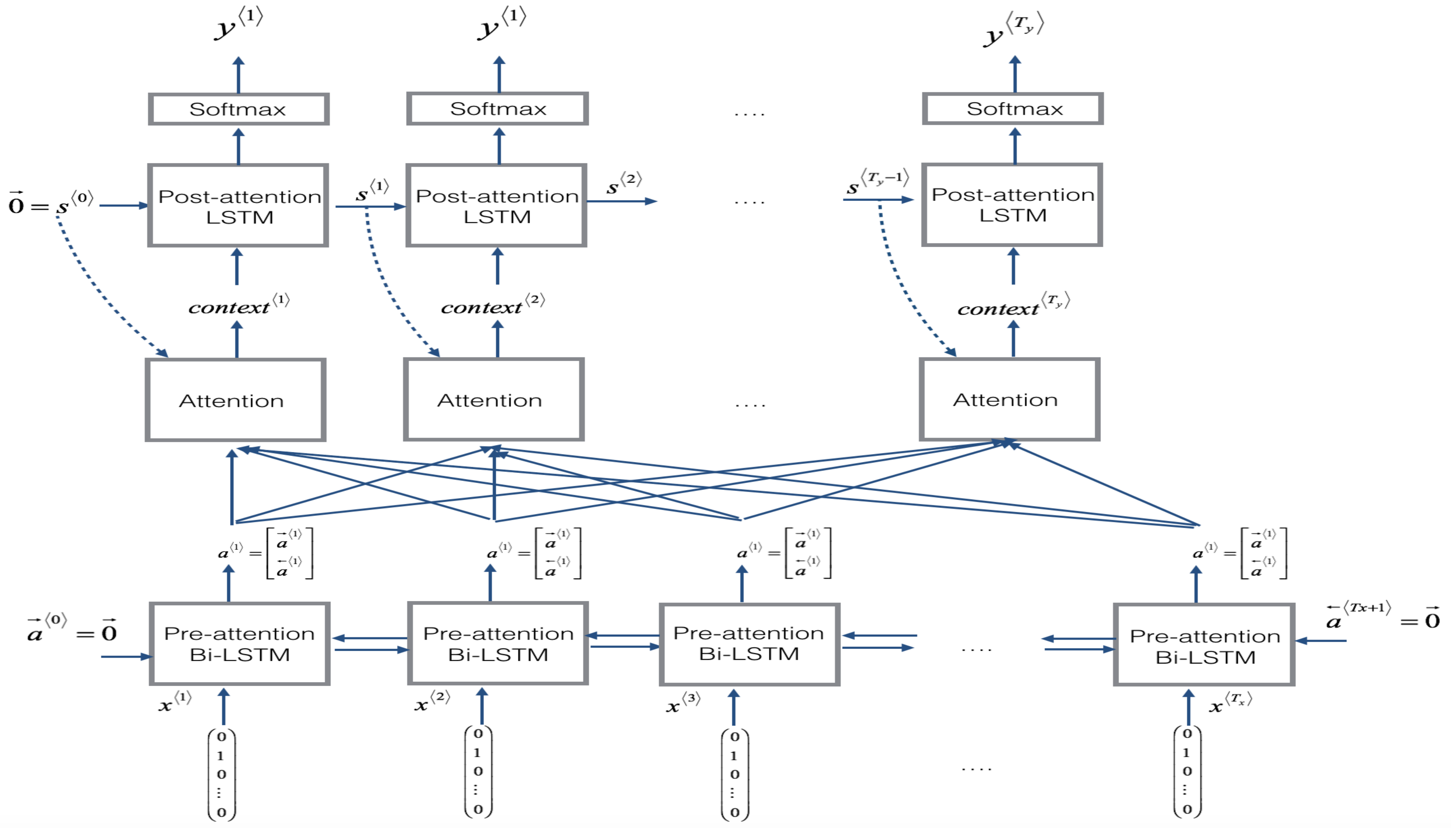
Attention Mechanism

- The encoder should memorize this long sequence into one vector, and the decoder has to process this vector to generate the translation.
- If a human would translate this sentence, he/she wouldn't read the whole sentence and memorize it then try to translate it. He/she translates a part at a time.
- The performance of this model decreases if a sentence is long.
- The attention model works like a human that looks at parts at a time. It significantly increases the accuracy even with longer sequence.
- At first the attention model was developed for machine translation but then other applications used it like computer vision, Speech Recognition, Sensor data classification etc.

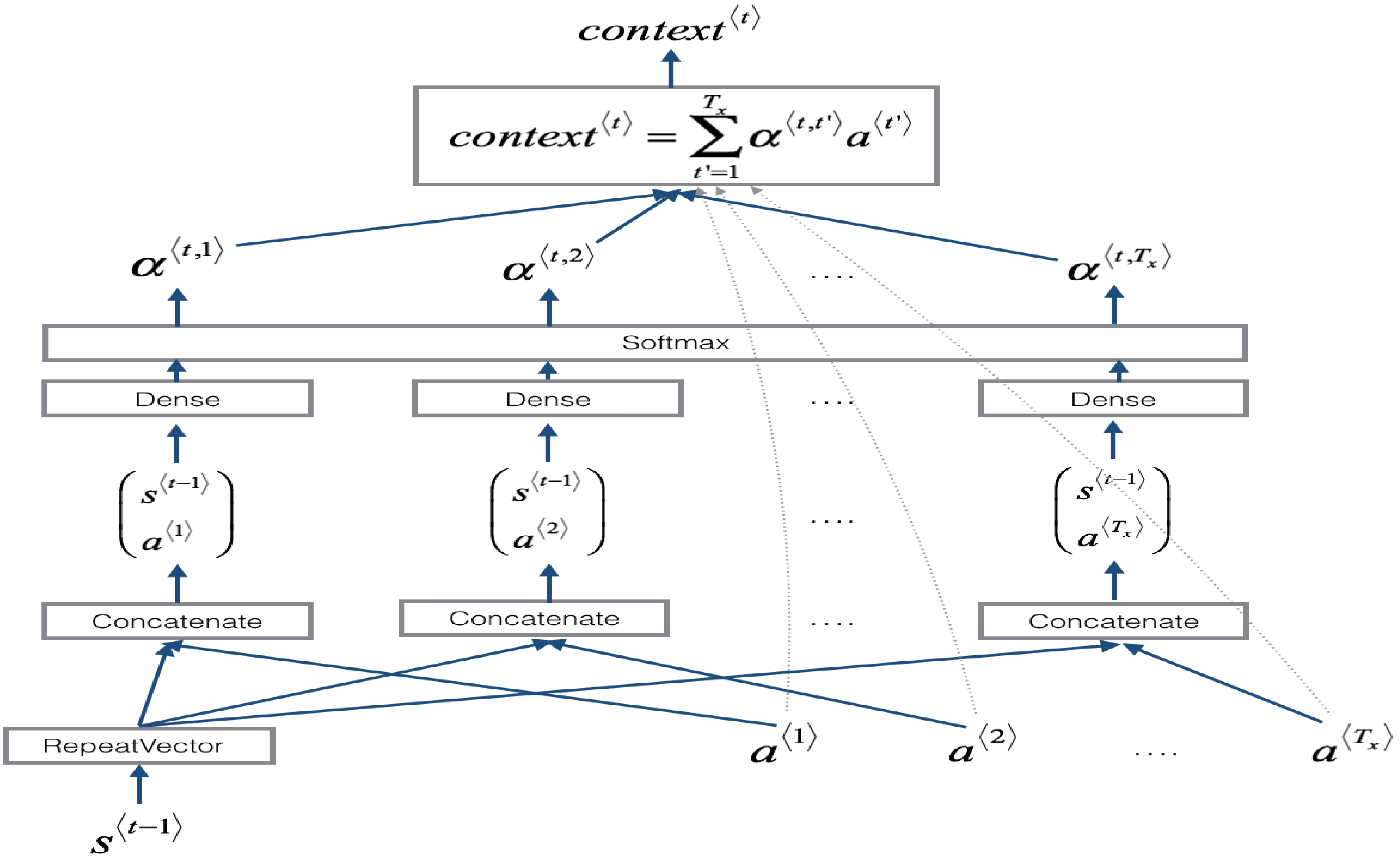
Bahdanau et. al., 2014. Neural machine translation by jointly learning to align and translate



Karen besuchte Europa im September.



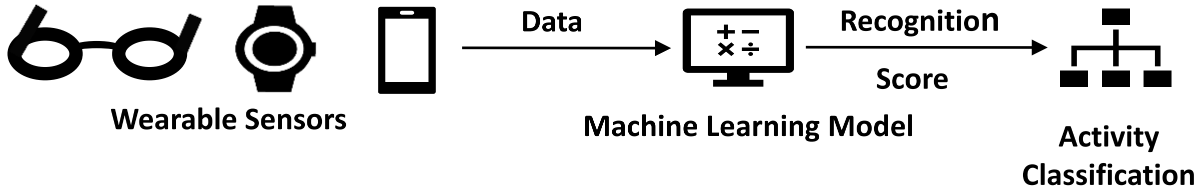
$y^{\langle 1 \rangle}$
$y^{\langle 1 \rangle}$
$y^{\langle T_y \rangle}$
Softmax
Softmax
....
Softmax
$\vec{0} = s^{\langle 0 \rangle}$
Post-attention LSTM
$s^{\langle 1 \rangle}$
Post-attention LSTM
$s^{\langle 2 \rangle}$
....
$s^{\langle T_y - 1 \rangle}$
Post-attention LSTM
$context^{\langle 1 \rangle}$
$context^{\langle 2 \rangle}$
$context^{\langle T_y \rangle}$
Attention
Attention
....
Attention
$a^{\langle 1 \rangle} = \begin{bmatrix} \overrightarrow{a}^{\langle 1 \rangle} \\ \overleftarrow{a}^{\langle 1 \rangle} \end{bmatrix}$
$a^{\langle 1 \rangle} = \begin{bmatrix} \overrightarrow{a}^{\langle 1 \rangle} \\ \overleftarrow{a}^{\langle 1 \rangle} \end{bmatrix}$
$a^{\langle 1 \rangle} = \begin{bmatrix} \overrightarrow{a}^{\langle 1 \rangle} \\ \overleftarrow{a}^{\langle 1 \rangle} \end{bmatrix}$
$a^{\langle 1 \rangle} = \begin{bmatrix} \overrightarrow{a}^{\langle 1 \rangle} \\ \overleftarrow{a}^{\langle 1 \rangle} \end{bmatrix}$
$\overrightarrow{a}^{\langle 0 \rangle} = \vec{0}$
Pre-attention Bi-LSTM
Pre-attention Bi-LSTM
Pre-attention Bi-LSTM
....
Pre-attention Bi-LSTM
$\overleftarrow{a}^{\langle Tx+1 \rangle} = \vec{0}$
$x^{\langle 1 \rangle}$
$x^{\langle 2 \rangle}$
$x^{\langle 3 \rangle}$
$x^{\langle T_x \rangle}$
$\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$
$\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$
$\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$
....
$\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$



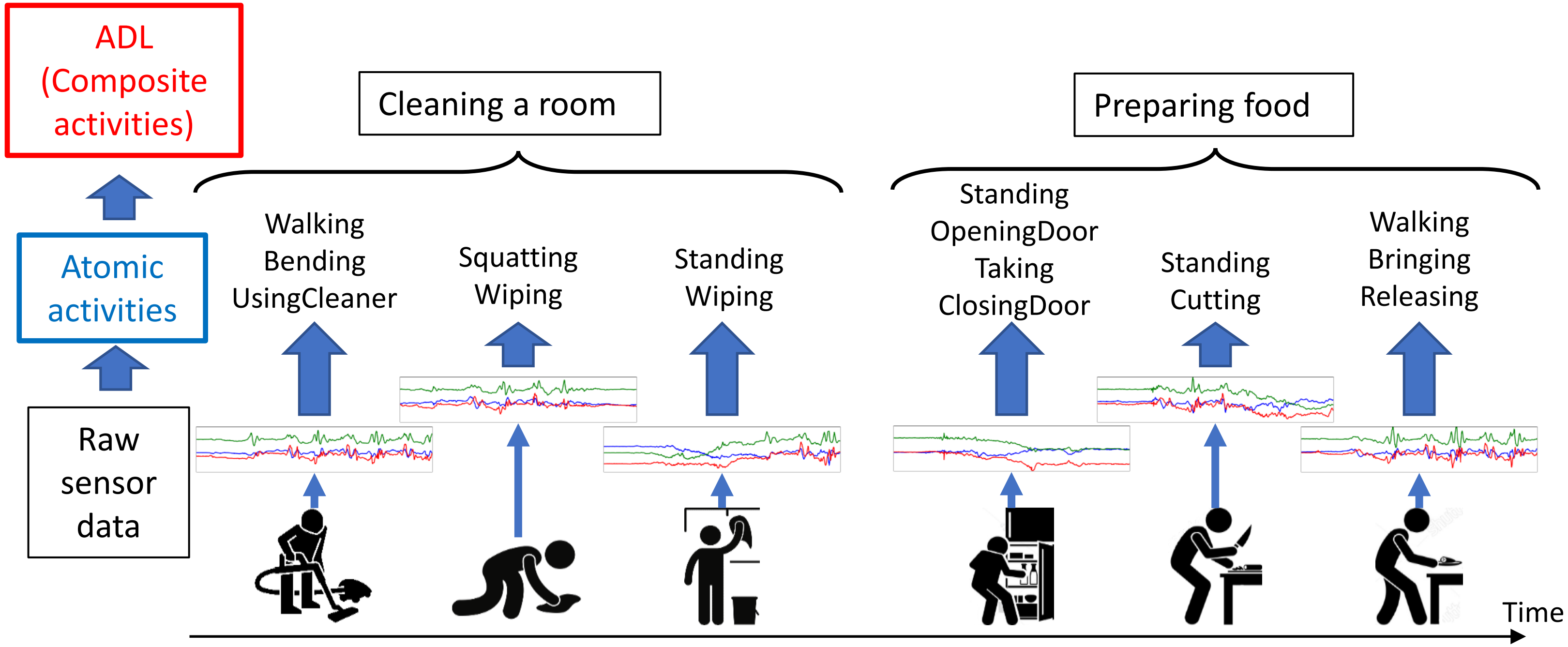
$context^{\langle t \rangle}$
$context^{\langle t \rangle} = \sum_{t'=1}^{T_x} \alpha^{\langle t,t' \rangle} a^{\langle t' \rangle}$
$\alpha^{\langle t,1 \rangle}$
$\alpha^{\langle t,2 \rangle}$
....
$\alpha^{\langle t,T_x \rangle}$
Softmax
Dense
Dense
....
Dense
$\begin{pmatrix} s^{\langle t-1 \rangle} \\ a^{\langle 1 \rangle} \end{pmatrix}$
$\begin{pmatrix} s^{\langle t-1 \rangle} \\ a^{\langle 2 \rangle} \end{pmatrix}$
....
$\begin{pmatrix} s^{\langle t-1 \rangle} \\ a^{\langle T_x \rangle} \end{pmatrix}$
Concatenate
Concatenate
....
Concatenate
RepeatVector
$a^{\langle 1 \rangle}$
$a^{\langle 2 \rangle}$
....
$a^{\langle T_x \rangle}$
$s^{\langle t-1 \rangle}$

Human Activity Recognition (HAR) System

- **Sensor-based HAR**



- **Acquisition of CogAge Datasets[1]**

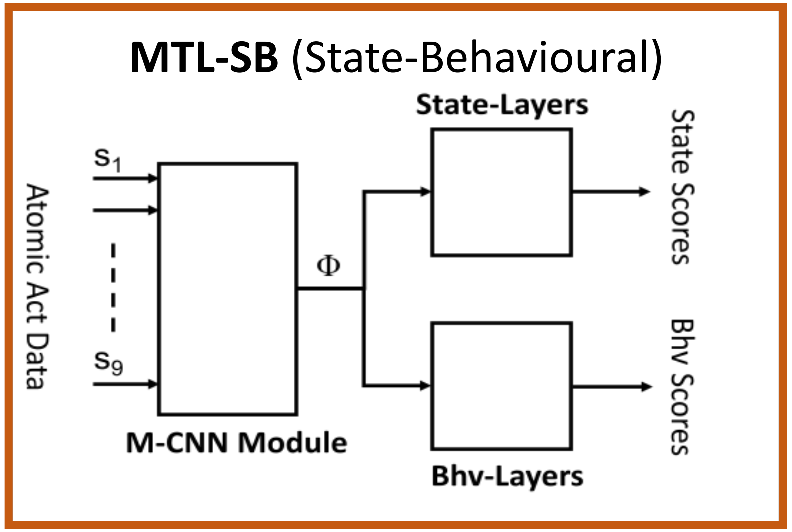


[1] Cognitive Village, Bundesministerium für Bildung und Forschung (BMBF) Project, FKZ 16SV7311

Hierarchical Multitask Learning

- Objective
  - Preserve the temporal evolution
  - Augment the dataset by learning all three types of activities
- Issues with existing approaches
  - Built for the activities with same temporal scales
- Proposed approach
  - Multitask model for different temporal scales
  - Time-distributed CNN-LSTM modules
  - Compared to baseline with single task learning (STL) models

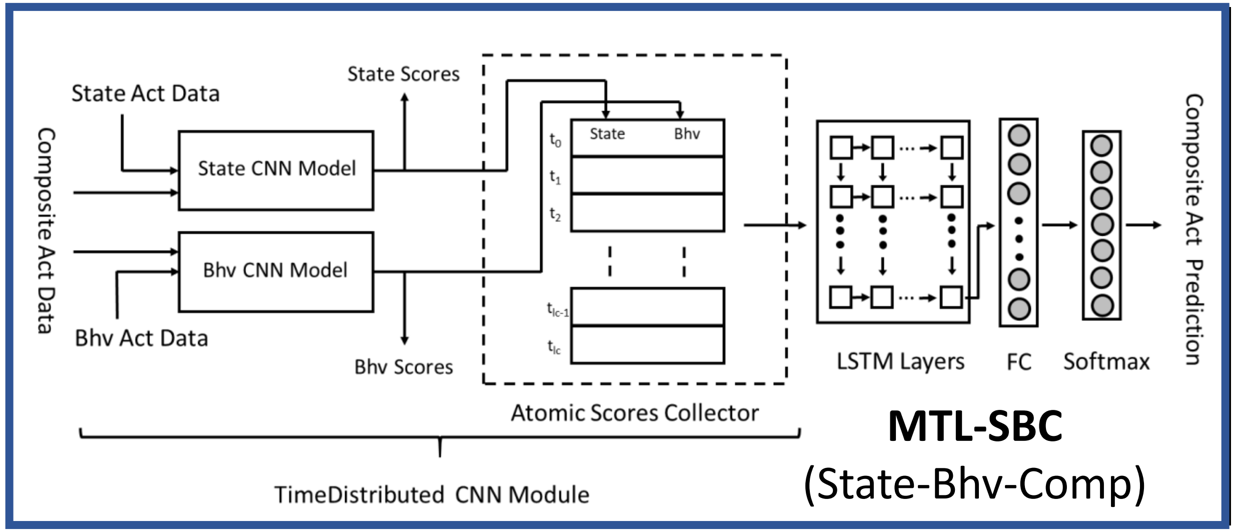
MTL Models



$$L(f_i) = -\frac{1}{U}\sum_{u=1}^{U} \boldsymbol{y}_i^{(u)^T} \log(\boldsymbol{q}_i^{(u)})$$

Where $i \in \{S, B, C\}$

$$L(f_{\mathrm{MTL-SB}}) = w_S\,.\,L(f_S) + w_B\,.\,L(f_B)$$



$$L(f_{\mathrm{MTL-SBC}}) = w_S\,.\,L(f_S) + w_B\,.\,L(f_B) + w_C\,.\,L(f_C)$$

MTL Models – Results and Issues

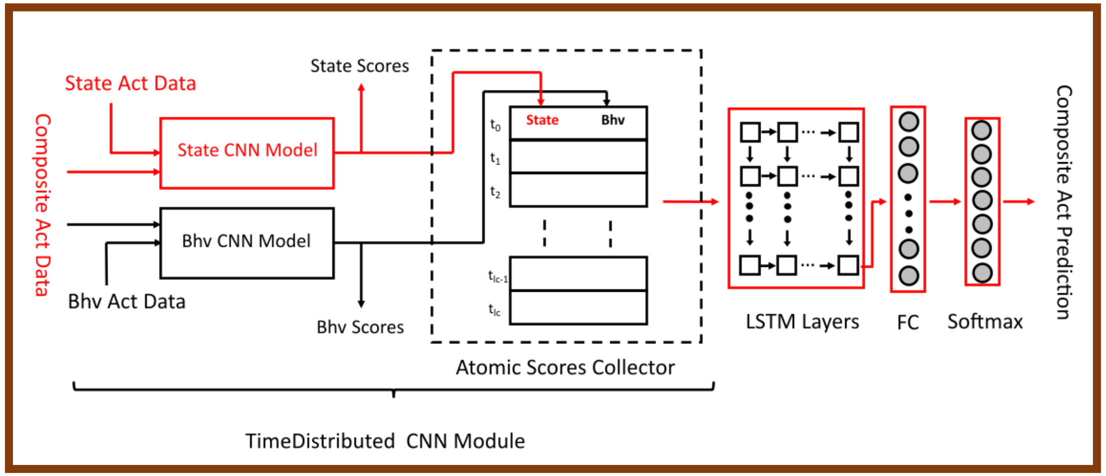
Models		State	Behavioural	Composite
Atomic	**Composite**	**Accuracy (%)**		
M-CNN	RP+MP+AP	92.4	71.8	88.5
MTL-SB		77.4	72.0	
MTL-SBC		74.4	72.2	**92.9**

Confusion Matrix – State Activities – MTL-SB						
Activities	Bending	Lying	Sitting	Squatting	Standing	Walking
Bending	**69**	0	0	0	0	0
Lying	6	**63**	0	0	0	0
Sitting	**43**	2	**19**	3	0	0
Squatting	17	0	2	**48**	0	0
Standing	**43**	0	0	2	**27**	0
Walking	5	0	0	0	0	**65**

MTL-StateComposite-BehaviouralComposite

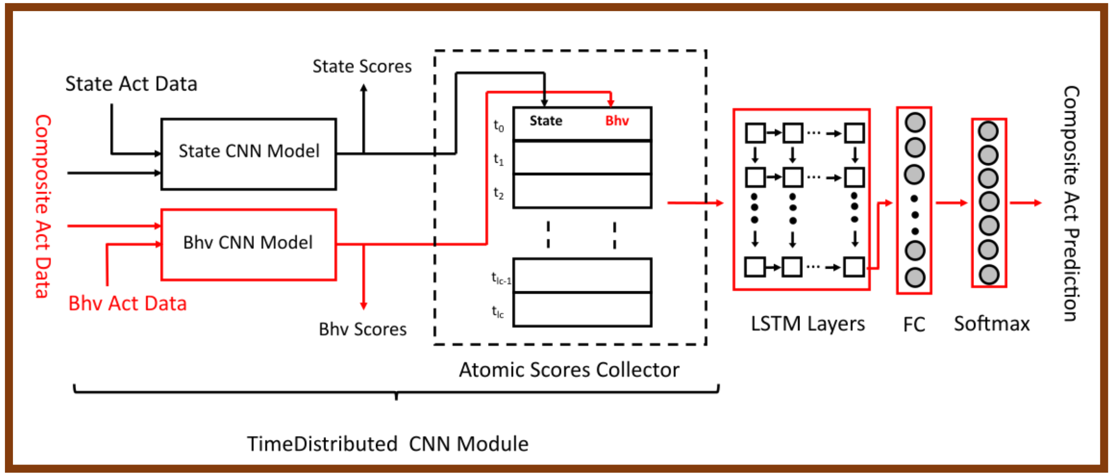
MTL-SC-BC

$$L(f_{SC}) = w_S\,.L(f_S) + w_C\,.L(f_C)$$



MTL-SC-BC

$$L(f_{BC}) = w_B\,.L(f_B) + w_C\,.L(f_C)$$



Final Results – All Methods

Recognition of State, Behavioural and Composite Activities							
Method		State		Behavioural		Composite	
Atomic	**Comp.**	**AF1**	**Acc**	**AF1**	**Acc**	**AF1**	**Acc**
CB	RP+MP+AP	88.2	88.6	67.9	68.2	88.0	88.5
M-CNN	RP+MP+AP	92.3	92.4	71.7	71.8	87.4	87.9
MTL-SBC		73.3	74.4	71.7	72.2	92.3	92.9
MTL-SC-BC		**95.1**	**95.2**	**73.4**	**73.9**	**93.8**	**94.0**

More Application of Assistive Health Technologies

- Emotion Recognition
- Anxiety and Stress Detection
- Dataset
  - Sensors
    - Audio
    - Visual
  - Data
    - RAVDESS
      - https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio
    - TESS
      - https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess
    - Stress Detection
      - https://figshare.com/articles/dataset/Anxiety_Dataset_2022/19875217