Talking Hands: RNN-based Sign Language Recognition

Mohak Sukhwani (mohak.sukhwani@research.iiit.ac.in), Prateek Panwar (prateek.panwar@uoit.ca)

CVIT • International Institute of Information Technology • ViaLab • University of Ontario Institute of Technology

(Deepest gratitude to *Dr. Christopher M. Collins* for his support, guidance and encouragement)

Motivation

- Investigate the utility of mobile accelerometer data to identify human actions
- Learn data representations from mobile accelerometer using *Restricted Boltzmann Machines (RBMs)*
- Generate models to classify the learnt representations using *Recurrent Neural Networks (RNN)*

Approach

- A set of unsupervised features are learnt, to recognize the phrases from *American Sign Language (ASL)*, using RBM.

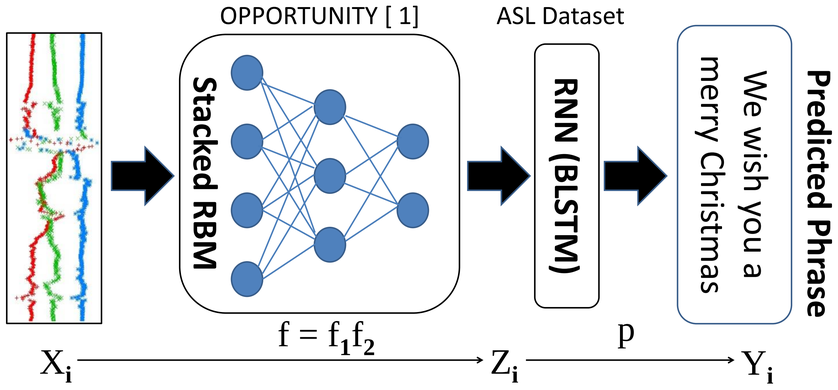

Fig 1: Overview: The stacked RBM network is trained for unsupervised feature generation. The computed features are used for phrase classification using RNN.

- Results are compared with the best performing supervised feature set [2].
- We created a dataset of 600 accelerometer readings collected from 50 users to validate the proposed approach.

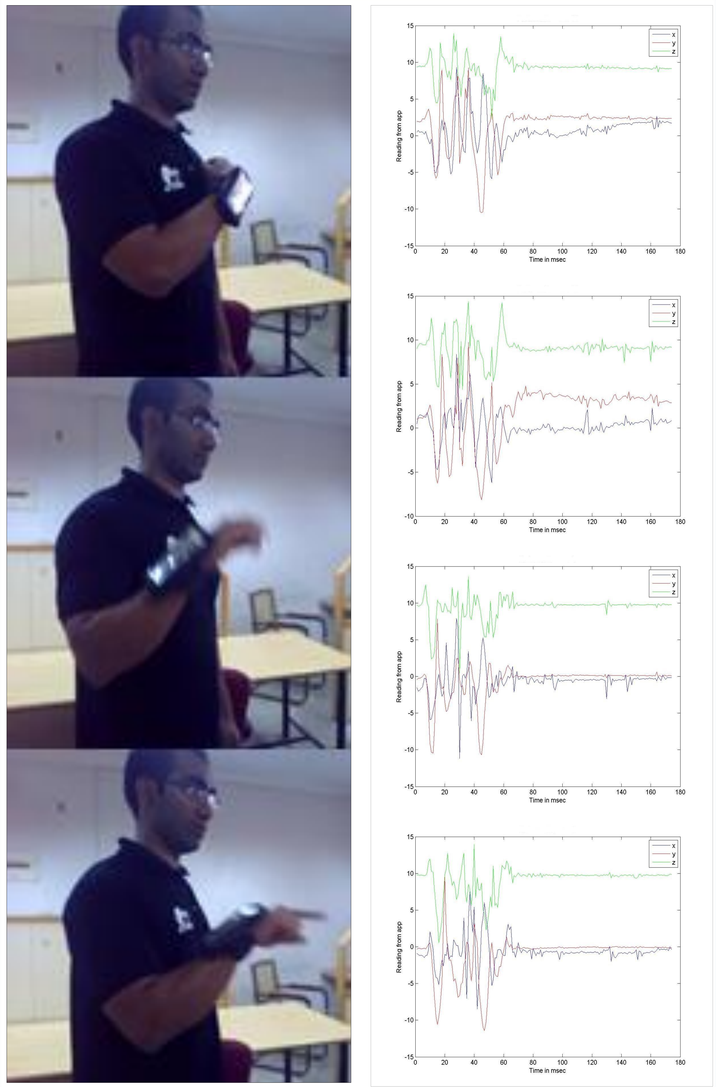

Fig 2: (Left) User wears the device on his wrist and performs the desired action. (Right) Figure shows the similarity in structure of accelerometer readings obtained while 4 different participants enact the given phrase. We intend to capture this similarity using our proposed solution.

Applications

- Gestural phrase to speech conversion
- Translation to other languages
- Instant messaging
- Smart watch app

Dataset

Name	Contents	Role
OPPORTUNITY[1]	Sampled wrist accelerometer data from the online dataset	RBM weights for feature learning
* ASL	600 accelerometer readings for selected phrases	RNN weights for ASL recognition

*The dataset is available online: http://bit.ly/talkinghands-data

Results

Method	Raw Features		BoW		Method	Raw Features		BoW	
	L_1	L_2	L_1	L_2		L_1	L_2	L_1	L_2
Naive Base	0.18	0.19	0.34	0.39	SVM-Linear	0.13	0.13	0.16	0.20
KNN	0.14	0.23	0.25	0.25	SVM-Poly	0.28	0.25	0.325	0.30
Reg. Trees	0.24	0.26	0.285	0.30	SVM-RBF	0.19	0.19	0.45	0.46

Phrase classification accuracy using hand crafted features: L1 and L2 correspond to normalization scheme. The proposed unsupervised features obtain an accuracy of **0.53** with stacked RBMs which is **13%** more than the best performing supervised technique [2].

Discussion

Features:
- ASL phrase detection
-Unsupervised feature learning using neural networks

Limitations:
-Present prototype uses only 1 hand
-Limited phrase vocabulary

Future work:
-Improved accuracy with more advanced movement capturing sensors
-Expanded vocabulary
-Bimanual Gestures

References:

[1] Chavarriaga, Ricardo, et al. "The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition." *Pattern Recognition Letters* 34.15 (2013): 2033-2042.

[2] Zheng, Yonglei, et al. "Physical Activity Recognition from Accelerometer Data Using a Multi-Scale Ensemble Method." *IAAI*. 2013.