Department of Computer Science
Duale Hochschule Baden-Württemberg Stuttgart

Activity Recognition with Neural Networks

Bachelor Thesis

Author:	Author
Supervisors:	Prof. Dr. Dirk Reichardt
	Ahmed Elnaggar, M.Sc.
Submission Date:	XX July, 20XX

Department of Computer Science
Duale Hochschule Baden-Württemberg Stuttgart

Activity Recognition with Neural Networks

Bachelor Thesis

Author:	Author
Supervisors:	Prof. Dr. Dirk Reichardt
	Ahmed Elnaggar, M.Sc.
Submission Date:	XX July, 20XX

This is to certify that:

(i) the thesis comprises only my original work toward the Bachelor Degree

(ii) due acknowlegement has been made in the text to all other material used

Author
XX July, 20XX

Acknowledgments

Thank Professor and supervisor

Thank Family [Father, Mother, Aunt]

Thank Slim

Abstract

Abstact

Contents

Chapter 1

Introduction

Scope : ADL activites

- Introduce the problem

- pass by efforsts to solve the problem

- Go down to the solution

* Why HAR is important?

* Back in history: how did the field start?

* Talk about basics (activities and how to capture them)

* Researchers [first efforts, challenges and current research]

Chapter 2

Literature Review

2.1 Introduction

HAR has emerged and become one of the important interests of computer scientists since 1980s as it has many different applications in medicine, security, HCI and many other fields.

2.2 The Meaning of "Acitivity"

"*Activity is a unit of life, mediated by psychic reflection, the real function of which is that it orients the subject in the objective world. In other words, activity is not a reaction and not a totality of reactions but a system that has structure, its own internal transitions and transformations, its own development.*" [10]

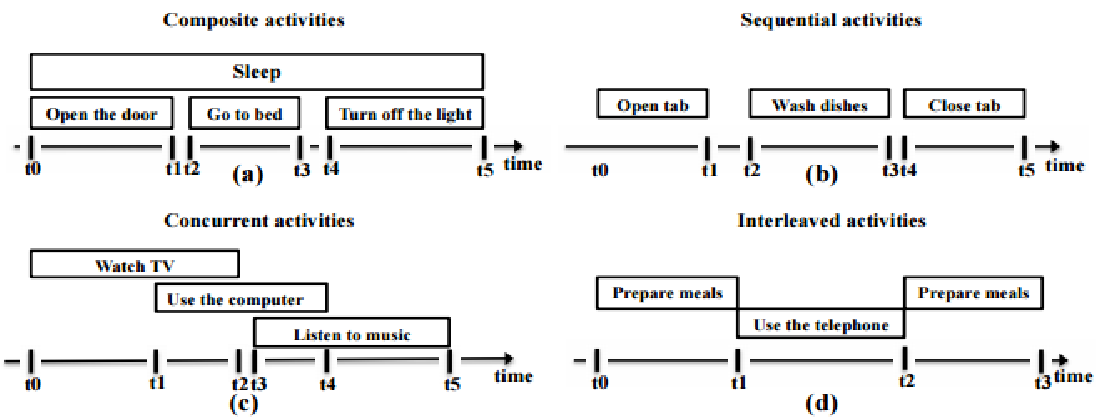
Though there is no unique definition for the meaning of activity, we consider it as any action that a person takes consciously with his/her own choice not as a reflex. In general, we can think of activities as a composition of other simpler activities which are, in turn, compositions of much more simpler activities, till we end up with very basic activities that we don't need to break down anymore. For example, "going to sleep" can be regarded as a composition of "entering the bedroom", "turning the lights off" and "laying in bed", where "entering the bedroom" consists of "opening the door", where "opening the door" is composition of some "hand movements".

Furthermore, activities can be related according to the time intervals in which they occur. Figure 2.1 shows these relations [11].

2.3 Human Activity Recognition

The main aim of activity recognition is to identify the actions of a person or multiple persons based on some observations from the person(s) and the surrounding environmental factors.

Figure 2.1: Relation between activities considering time intervals [11].



2.3.1 Signalling Activities

[ADD Section to relation between actual activities and how they are collected]

2.3.2 Process of HAR

Mainly, the HAR process consists of several steps [11][13]:

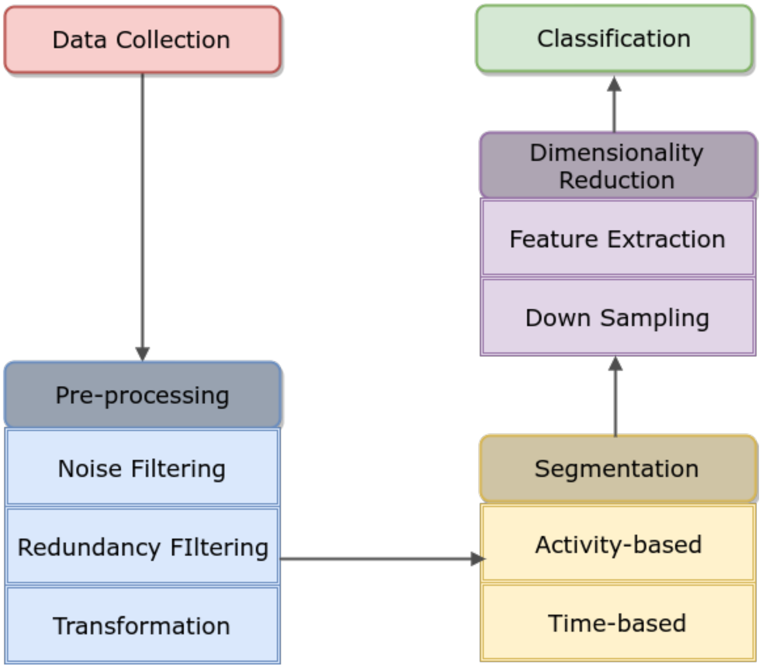
- Data Collection: collecting raw data from sensors. This raw data should be large and diverse enough to be able to generalize and produce accurate results when using the model in application.
- Data Pre-processing: filtering noise and redundancy and handling missing values. Sometimes, this step includes data transformation such as normalizing the data to be able to relate data collected from different test participants.
- Data Segmentation: determining the important and effective data segments. Different approaches are used for data segmentation such as:
  - Activity-based segmentation: raw data are separated into different segments depending on the activity labels.
  - Time-based segmentation: raw data are separated into segments of specific time intervals. A sliding window can be used to let segments overlap over time. This can be helpful in increasing the number of data records without decreasing the performance.
- Dimensionality Reduction: extracting the significant features that may affect the activity classification. This can be followed by reducing the number of features if some features were found to be insignificant to the classification process or if

features are homogeneous, such as a certain feature recorded over time, and we want to improve the computational speed. This can be regarded as *down sampling* of the features vector.

- Classification: identifying the activity.

Figure 2.2 summarizes the HAR process.

Figure 2.2: Process of HAR



2.4 Methods used in data collection and analysis

A lot of research has been done to provide practical solutions for many HAR problems. Many approaches depend on visual devices to identify and detect specific activities. Some of these approaches have provided good practical results in some applications such as Ambient Assisted Living (AAL) systems [7][3], security and surveillance [6][8], and health care monitoring [5][1]. One of the main advantages of visual-based systems is that they provide reliable data because the-state-of-the-art cameras have powerful resolutions. Furthermore, cameras can capture a wide range of the environment.

Another set of approaches depends on sensor devices such as sensors that can record heart rate, human-voice streams and Electromyography (EMG) signals. A network of sensors provide a set of consistent information about the user behavior. Sensor-based systems have been used in health care [12] and AAL systems [9], as well. Some approaches combined between visual and non-visual devices [11][15][4].

2.5 Challenges facing HAR

Although visual-based systems can provide good accuracies, they have some limitations. They cannot get specific details, for example, for hidden objects. Moreover, they are sensitive to light and brightness factors, and need more processing time.

On the other hand, data collected from sensors can be associated with noise resulting in unreliable data. Therefore, preprocessing is usually needed by applying noise filters.

[CITE ABOVE PART (REVIEW??)] [TALK about single user vs multiple users]

2.6 Examples for ADL Models

Google API for ADL

LSTM Model

2.7 Conclusion

[TO ADD] Wearable sensors in HAR

The current focus in the HAR field is novel machine learning. [READ Approaches for HAR]

Image classification in CNN

Segmentation of collected data - activity-based segmentation (since we are considering sequential activites only) - time interval-based windowing segmentation

Chapter 3

Devices

Chapter 4

Methodology

TODOS ===== 1. Submission Date 2. Acknowledgements 3. Write Abstract

Dataset collection =================== TODO: - Talk more about Myo sensor - How the sensor is worn - Cross validation vs validation set - Time issues [preprocessing time, training time] - EMG normalization

Contents: ========= - Information for EMG signals - Information for Myo Sensor - How sensor is worn - Description for 10 activities - How raw data is stored (two structures) - Preprocessing data before network feed - Dataset loading

* To work with EMG signals, data is collected from the Myo armband sensor worn on the right arm.

* 10 activities have been recorded on XX participants of age ranging from YY to ZZ

* The activities are:

1. Walking: most of the humans swing their arms while walking such that each arm swings with the motion of the opposing leg. The walking procedure has been recorded making sure that all participants swing their arms while walking

2. Standing: not making any movements or gestures with the hand.

3. Talking in the cell phone: by putting the cell phone on the right ear.

4. Typing on the keyboard: using all fingers

5. writing in a notebook: with the hand on the desk

6. drawing on a white board: while standing

7. Brushing teeth: the toothbrush is moved in horizontal and vertical motion

8. Holding a spherical shape: that can be gripped by hand

9. Holding a cylinderical shape: with radius xx cm

10. Holding a thin book: with thickness xx cm

* The Myo sensor collects data with frequency xx Hz. Around xx frames/seconds have been recorded for each participant activity. So, each participant has 10 files one for each activity with each file containing N rows and 8 columns where each row corresponds to a frame and each column corresponds to an EMG channel.

* The data is processed further for easier usage such that each participant has a folder with all activity files and each file named with the name of the activity. [TYPE NAMES HERE]

* Now, this data need further preprocessing before feeding it to the classifier. - We need to organize our records into datapoints (examples) such that each example has: + specific number of time frames. This is determined through the sampling rate. If the activity record contains F time frames and our sampling rate is M, then the record is split into floor(F/M) datapoints such that each datapoint will get F consecutive time frames. The frames must be consecutive for convolution to work. Extra frames are ignored + EMG signals per time frame. Each time frame is associated with 8 values which are the EMG records for this time frame + A label: that is set manually to identify this datapoint We can see now that our datapoint has dimensions M x C. C is considered to be the depth of each datapoint. and M is the width

- We need to prepare two multi-dimensional array: + features: contains features for the datapoints. It has dimensions N x M x C where N is the number of datapoints M is the sampling rate (number of time frames per datapoint) C is the number of EMG channels (8 in our case) + labels: contains labels for the datapoints. It has dimensions N x V: - N is the total number of datapoints - V is the number of activities (classes) - Activity labels are numbered from 0 to 9 as described in config.py - One-hot encoding is used in this ndarray. In other words, each row will contain value '1' in the correct activity column and value '0' everywhere else These arrays are prepared from the raw data we collected and then pickled to be directly used later. The arrays are shuffled so that partioning the array into training and test sets will be random and training will not be baised, so we get best performance.

Now the dataset is ready to use and can be split into train, valdation and test sets.

When dataset is loaded, it is split into 3 sets: - training set [DESCRIBE MORE] - Validation set [DESCRIBER MORE] - Test set [DESCRIBE MORE]

Sets are completely disjoint and sampled as consecutive batches in the main dataset array. Recall that the dataset has already been shuffled before.

¿¿¿first run (BUG!, small dataset)

acc train = 93 acc val = 20 start loss = 1043.465

Overfitting!

¿¿¿added dropout layer acc train = 32.5acc val = 13start loss = 886.361

¿¿¿sampling rate = 512, 50

TODO ==== - tweak learning method, learning rate, batch size, number of layers, dims of each layer, num steps - overlap in the reading?

+ study participants

Chapter 5

Neural Network Model For ADL

[Mention the LSTM model and its accuracy]

5.1 ADL Dataset and the SVM Approach

Here, we examine a public domain dataset [2] that targets the recognition of six basic Activities of Daily Living (ADL). The activities are are: standing, laying down, walking, walking downstairs and upstairs. Raw data is collected from 30 test participants using the accelerometer and gyroscope of a smartphone. Then, the collected data is is processed furhter to extract features such as mean, correlation and signal magnitude area. The final shape of the dataset is a set of 10299 datapoints, where each datapoint is a vector of 561 features. 70% of the dataset is used for training while the other 30% is used for testing.

The state-of-the-art implemented approach is One-Vs-One Multiclass linear SVM with majority voting which achieved an accuracy of 96.4% [14].

5.2 Neural Network Approach

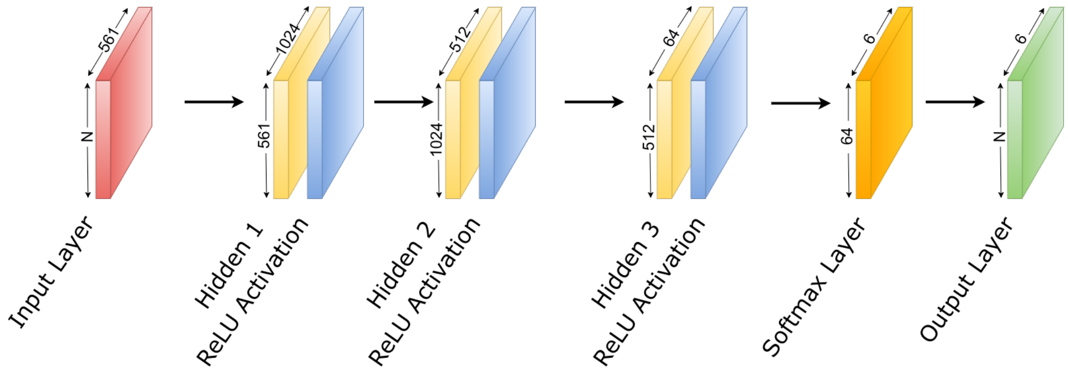
Here, we apply a neural network to find out the best accuracy which we can reach on feature-extracted HAR dataset with a neural network model. We separated 14% of the train set to be used as a validation set. The validation set is important to tune the hyperparameters such as learning rate, learning method, number of hidden layers and their dimensions. Consequently, the test dataset is only used after tuning all the hyperparameters.

The model is a multi-layer neural network. Each hidden layer is a fully connected layer followed by a Rectified Linear Unit (ReLU) activation function

5.2.1 Initial Network Configuration

Initially, we used a three-layer network model. The dimensions of the hidden layers are 1024, 512 and 64 (in the direction from the input layer to the output layer). Each hidden layer is followed by a ReLU unit. Furthermore, we used an 0.001 learning rate and 5,000 steps for the training process. At each step, only a mini batch of 256 datapoints was used in training to reduce the computational time. The mini batch was selected randomly. Figure 5.1 shows the architecture of the network.

Figure 5.1: Initial Network Architecture



Using a single validation set results in accuracies with 1.5% error. So, we used cross-validation to produce more stable accuracies.

5.2.2 Learning Optimizer

The core idea of the machine learning algorithm is to minimize a loss function which compares the correct labels with the predicted labels.

Loss Function Since we are using a softmax layer, the loss function we are using is the cross-entropy function which can be formulated as follows:

$$loss = -\sum_{i}^{N}\sum_{j}^{M} p_{ij} \log q_{ij}$$

where:

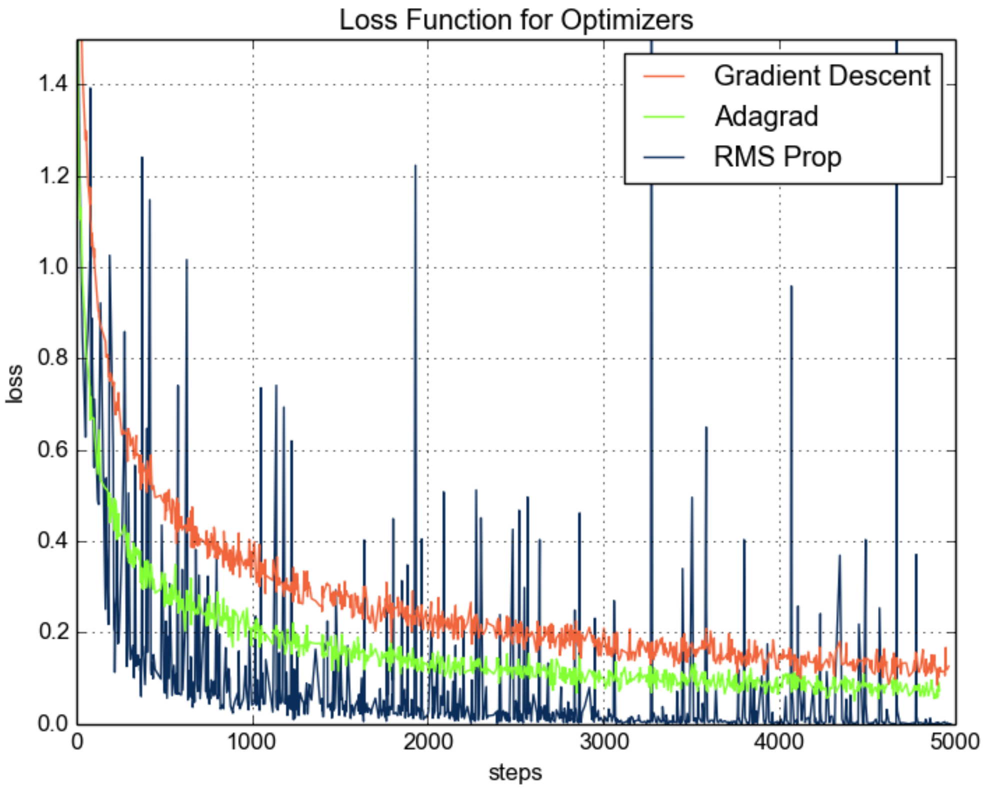
- N is the number of datapoints (examples).
- M is the number of classes (activities).
- p_{ij} is the true probability (from the dataset) that the j-th activity is the correct label for the i-th example. Note that for a fixed example, only one activity will have $p = 1$ while other activities will have $p = 0$.
- q_{ij} is the estimated probability (by the network) that the j-th activity is the correct label for the i-th example.

Gradient Descent Some information about Gradient Descent and SGD

Adagrad Some information about Adagrad

RMSProp Somme information about RMS Prop

Figure 5.2: Loss Function for Optimizers



[Analysis] Adagrad is better than Gradient Descent. RMS is the best. Talk about the glitches.

[Conclusion] We will use RMS Prop. No overfitting.

Figure 5.3: Validation Accuracies for Optimizers

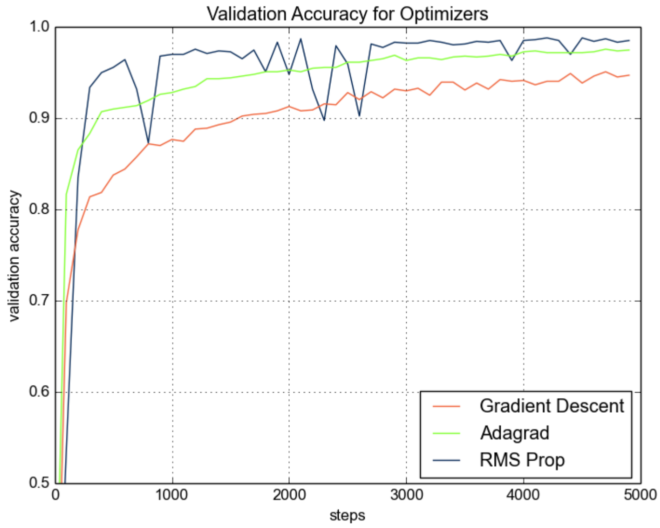
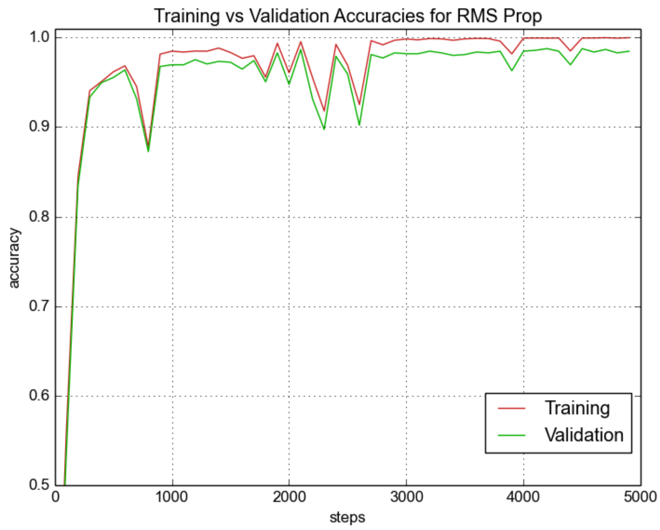


Figure 5.4: Training and Validation Accuracies for RMS Prop



5.2.3 Weights and Biases Initialization

One of the most important things to take into consideration when building a neural network is the initialization of the trainable parameters. A good initialization will help the network to learn faster and achieve better accuracy.

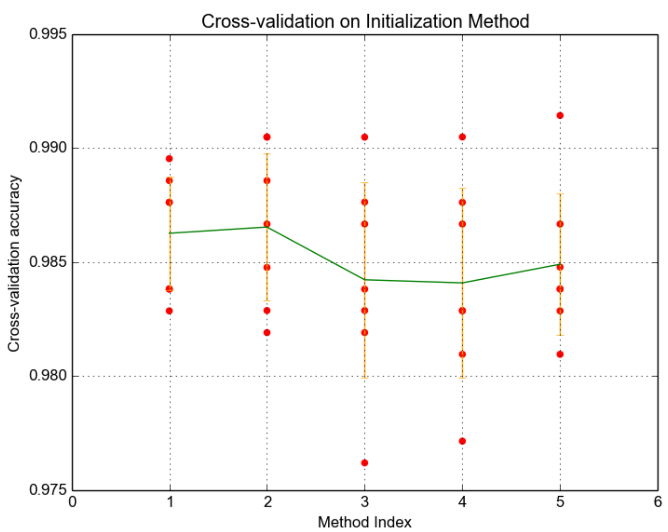
We tried five initialization methods and monitored their behavior and accuracies. The methods and their results are shown in table 5.1. Truncated Normal Distribution (N.D) was used in all methods.

#	Weights	Biases	Train Acc (%)	Val Acc (%)
1	N.D, $\sigma = 0.1$	N.D, $\sigma = 0.1$	99.96	98.63
2	N.D, $\sigma = 0.1$	zeros	99.91	98.65
3	N.D, $\sigma = 0.01$	zeros	99.87	98.42
4	N.D divided by $\sqrt{fan\_in/2}$, $\sigma = 0.01$	zeros	99.72	98.41
5	N.D divided by $\sqrt{fan\_in/2}$, $\sigma = 0.1$	zeros	99.97	99.49

Table 5.1: Comparison between initialization methods

Figure 5.5 shows the cross validation results for different methods. For each method, validation accuracy for each fold is plotted and standard deviation is calculated. The green line connects the mean accuracy of all methods.

Figure 5.5: Cross-Validation for initialization methods



Note that we didn't include methods that initialize the weights with zeros. In general, this is a bad practice because all neurons will be computing the same values and will have similar gradients. So, we need to break this symmetry using random values.

It was suggested by [REF NEEDED] to initialize weights by method 4 or 5 (σ depends on the dataset). But, since our network is not very deep, the results showed that method 1 and 2 are more accurate on validation set. In conclusion, we will use method 2 for our network weight initialization.

5.2.4 Number of Hidden Layers and Their Dimensions

Tested values for the number of layers were 3, 5 and 7 with the dimension of each layer as a power of 2 (this achieves faster computation using library functions [**?**]). Experiments showed that they result in close accuracies on validation set. That's why the three-layer model is used for the final version of the implementation because it is faster for computation.

[TEST different dimensions for the 3 hidden layer]

5.2.5 Batch size

Talk about S.G.D and it's computationally more efficient to take a mini batch than training on the whole dataset. We tried values 256, 512, 1024 for batch size. We used He etal initialization method for the deep models.

5.2.6 Learning rate and the number of steps

try num steps [2k, 5k, 10k] learning rates = [0.01, 0.03, 0.07, 0.001]

5.2.7 Final Accuracy

The final best accuracy we could achieve with our neural network model is xx%.

[TODO] - edit final best accuracy - Talk about fully connected layers - Talk about ReLU - Add more information in initialization

CNN has become and important tool for object recognition in Computer Vision. Here we try to apply CNN for HAR

CNN has already been used for visual recognition [add references from slide 47 in lecture 1]

Use cross validation if dataset is small

Chapter 6

Convolutional Neural Network

Chapter 7

Testing and Results

Chapter 8

Conclusion

UCI-ML dataset and referenced papers Myo Documentation Udacity, Stanford and TF courses

EMG

Kamen, Gary. Electromyographic Kinesiology. In Robertson, DGE et al. Research Methods in Biomechanics. Champaign, IL: Human Kinetics Publ., 2004. [for EMG]

Craig Freudenrich, Ph.D. & Robynne Boyd "How Your Brain Works" 6 June 2001. HowStuffWorks.com. ¡http://science.howstuffworks.com/life/inside-the-mind/human-brain/brain.htm¿ 18 July 2017 [for motoneurons]

Chapter 9

Future Work

performed in labs vs realtime (review)

several activities at the same time

multiple users

Future work, Normalization of EMG signals. [Methods in Biomechanics]

Add more activities as: 1. running 2. stirring

limitations: computational time, powerful computer needed

Chapter 10

Background

Define HAR - identifies the action carried out by a person given a set of obersvations of him/herelf and the surrounding environment. Recognition can be accomplished by exploiting the information retrieved from various sources such as environmental or body-worn sensors.

Some approaches have adapted dedicated motion sensors in different body parts such as the waist, wrist, chest and thighs achieving good classification performance.

Machine Learning (ML) has become so popular to solve many problems that are easily solvable by humans but hard to solve by machines without explicitly programming them to do specific tasks. Examples for such problems are image classification, object detection and clustering. Different approaches have been used in machine learning to solve these problems and some of them showed better results over others in sepcific problems. Deep learning is one of the ML approaches that is considered a breakthrough to machine learning and artificial intellegence in general. Nowadays, deep learning, which rebranded neural networks, can achieve results with accuracy close to human brains. Here, we will use deep learning to solve activity recognition problem.

10.1 EMG Signals

Electromyography is the study of muscle electrical activity. it provides information about the control and execution of voluntary (and reflexive) movements.

1. The basis of the muscle fiber action potential and how it propagates along the muscle fiber.

The human body has specific type of neurons called "motorneurons" that control muscle contractions. [**?**] Motoneurons carry signals from the the central nervous system to the muscles. Nerve impulses travel down the motor neuron and stimulate the appropriate muscle to contract.

When muscle fibers receive the electrical impulse from motoneuron, muscle fibre action potential (AP) is generated. The AP is the nueral messenger responsible for activating every segment of the muscle fibre.

Each motoneuron innervates many muscle fibers. The motoneuron associated to it all of the muscle fibers innervated by that motoneuron is called a motor unit.

The motor unit action potential (MUAP) represents the summated electrial activity of all muscle fibers activated within the motor unit.

It's worth noting that since the sensor electordes are surface electordes, deep muscle fibres has a litte contribution to the recorded EMG signal.

During a muscle contraction, some motor units are activated. The number of activated motor units depends on the required force. The higher the required force, the more motor units are activated.

EMG signal at any moment of time is the composite electircal sum of all of the active motor units. The amplitude of EMG signal increases as the intensity of the muscular contraction increases.

Though the activities we recorded doesn't seem to require a lot of force, naturally muscles in some activities will require more contraction than in other activities.

Surface EMG assesses muscle function by recording muscle activity from the surface above the muscle on the skin. Surface electrodes are able to provide only a limited assessment of the muscle activity. Surface EMG can be recorded by a pair of electrodes or by a more complex array of multiple electrodes. More than one electrode is needed because EMG recordings display the potential difference (voltage difference) between two separate electrodes. Limitations of this approach are the fact that surface electrode recordings are restricted to superficial muscles, are influenced by the depth of the subcutaneous tissue at the site of the recording which can be highly variable depending of the weight of a patient, and cannot reliably discriminate between the discharges of adjacent muscles.

Appendix

Appendix A

Lists

AAL	Ambient Assisted Living
ADL	Activities of Daily Living
HAR	Human Activity Recognition
EMG	Electromyography
ML	Machine Learning
ReLU	Rectified Linear Unit
N.D	Normal Distribution

List of Figures

List of Tables

Bibliography

[1] A Al-Marakeby. Camera-based wireless sensor networks for e-health. *International Journal of Advanced Research in Computer and Communication Engineering*, 2:4757–4761, 2013.

[2] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *ESANN*, 2013.

[3] Alexandros Andre Chaaraoui, José Ramón Padilla-López, Francisco Javier Ferrández-Pastor, Mario Nieto-Hidalgo, and Francisco Flórez-Revuelta. A vision-based system for intelligent monitoring: Human behaviour analysis and privacy by context. *Sensors*, 14(5):8895–8925, 2014.

[4] Donato Di Paola, David Naso, Annalisa Milella, Grazia Cicirelli, and Arcangelo Distante. Multi-sensor surveillance of indoor environments by an autonomous mobile robot. *International journal of intelligent systems technologies and applications*, 8(1-4):18–35, 2009.

[5] Anastasios Doulamis, Nikoalos Doulamis, Ilias Kalisperakis, and Christos Stentoumis. A real-time single-camera approach for automatic fall detection. *ISPRS Commission V, Close Range Image measurements Techniques*, 38:207–212, 2010.

[6] Clinton Fookes, Simon Denman, Ruan Lakemond, David Ryan, Sridha Sridharan, and Massimo Piccardi. Semi-supervised intelligent surveillance system for secure environments. In *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*, pages 2815–2820. IEEE, 2010.

[7] Robert Hartmann, Fadi Al Machot, Philipp Mahr, and Christophe Bobda. Camera-based system for tracking and position estimation of humans. In *Design and Architectures for Signal and Image Processing (DASIP), 2010 Conference on*, pages 62–67. IEEE, 2010.

[8] Jun-Wei Hsieh, Yung-Tai Hsu, Hong-Yuan Mark Liao, and Chih-Chiang Chen. Video-based human movement analysis and its application to surveillance systems. *IEEE Transactions on Multimedia*, 10(3):372–384, 2008.

[9] Maarit Kangas, Antti Konttila, Per Lindgren, Ilkka Winblad, and Timo Jämsä. Comparison of low-complexity fall detection algorithms for body attached accelerometers. *Gait & posture*, 28(2):285–291, 2008.

[10] Aleksei Nikolaevich Leontiev. Activity and consciousness. *Dialectus*, (4), 2016.

[11] Qin Ni, Ana Belén García Hernando, and Iván Pau de la Cruz. The elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development. *Sensors*, 15(5):11312–11362, 2015.

[12] George Okeyo, Liming Chen, and Hui Wang. Combining ontological and temporal formalisms for composite activity modelling and recognition in smart homes. *Future Generation Computer Systems*, 39:29–43, 2014.

[13] Suneth Ranasinghe, Fadi Al Machot, and Heinrich C Mayr. A review on applications of activity recognition systems with regard to performance and evaluation. *International Journal of Distributed Sensor Networks*, 12(8):1550147716665520, 2016.

[14] Bernardino Romera-Paredes, Min SH Aung, and Nadia Bianchi-Berthouze. A one-vs-one classifier ensemble with majority voting for activity recognition. In *ESANN 2013 proceedings, 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pages 443–448, 2013.

[15] Michel Vacher, Dan Istrate, François Portet, Thierry Joubert, Thierry Chevalier, Serge Smidtas, Brigitte Meillon, Benjamin Lecouteux, Mohamed Sehili, Pedro Chahuara, et al. The sweet-home project: Audio technology in smart homes to improve well-being and reliance. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 5291–5294. IEEE, 2011.