

Classifying Hand Gestures from EMG Data

CP3403 - Data Mining Project

Joshua Gray - 13177877

Harmon Singh -



College of Business, Law and Governance

James Cook University

Australia

25/05/2019

Contents

1	Introduction	1
2	Artificial Neural Networks	1
3	EMG Data for Gestures Data Set	1
4	Preprocessing	3
4.1	Data Format Tool	4
5	Artificial Neural Networks	6
6	Conclusion	6
	References	7
	Appendices	8
A	Examples	8

List of Figures

1	Thalmic Labs Myo Bracelet	3
2	Thalmic Labs Myo Bracelet	4

List of Tables

1	Attributes of the Raw Data Set	1
2	Descriptions of Attribute Class	2
3	Histogram of values across the EMG Data for Gestures Data Set	5

1 Introduction

Since the turn of the 20th century, a lot of focus has been placed on the incorporation and integration of technology in medical fields. This relationship between medicine and technology has spawned a multitude of life-saving devices such as the Automated External Defibrillator and non-invasive diagnostic tools such as Magnetic Resonance Imaging and Computed Tomography.

With the widespread increase of data mining and analysis in recent years as well as vast improvements in computing and storage power available today, a logical question for the data science community is finding the ways in which this area has the potential of improving the medical industry. In addition to this, in what way can the Internet of Things enable this kind of technology? A field which has attempted to improve the quality of life for it's patients since the beginning is the field of prosthetics.

2 Artificial Neural Networks

The field of classification has been plagued in recent years by the popularity of the Artificial Neural Network (ANN). Constantly evolving, ANN's are machine learning systems that aim to simulate the functionality of biological networks such as the human brain. As they have a constant defined structure, they excel when compared to some other methods as they have no bias from the programmer i.e. they are defined generally, with the data forcing biases in the output of the system.

3 EMG Data for Gestures Data Set

The EMG Data for Gestures Data Set (Available from: <https://archive.ics.uci.edu/ml/datasets/EMG+data+for+gestures>) was a data set created for determining latent factors in the performance of sEMG interfaces [1]. The overall achievement of this paper was identifying some potential medical influences on the use of EMG interfaces to measure gesture information.

This particular dataset was obtained through the use of the Myo Thalmic bracelet (See Figure ??). The EMG channel recordings were obtained through a Bluetooth interface to a computer. A total of eight channels are recorded with each of the sensors equally spaced around the forearm of the subjects recorded.

Table 1: Attributes of the Raw Data Set

Attribute	Description
Time	Time Step of Recording (Numeric)
Channel 1-8	Raw EMG Measurement (Double Precision Float)
Class	Gesture Performed (Numeric 0-7)

Table 2: Descriptions of Attribute Class

Class	Description
0	Unassigned
1	Hand at rest
2	Hand clenched in a fist
3	Wrist Flexion
4	Wrist Extension
5	Radial Deviations
6	Ulnar Deviations
7	Extended Palm

A total of 36 subjects were used for this experiment. Each subject performed a series of static hand gestures (a total of 7 unique gestures) with a 3 second gap between the recording of each gesture. The series were also repeated twice for each subject.

The values recorded by each channel are represented in scientific format corresponding to a double precision floating point number. The magnitude corresponds to the potential measured by the device at the skin level but units were withheld from the data. Further information about the Myo armband revealed that raw EMG data should be transferred as uint8 values corresponding to the amount of "activation" of the muscle but this is not reflected in the data. Based on preliminary analysis, the data should be sufficient as the final system should adapt to incoming signals without the need of specific input data (As long as the format is known).

An important thing to note for this data set is that due to the nature of the data (different subjects), the general classification task will prove more difficult as there are physiological aspects that affect the magnitude of the readings. As explored in [1], people who have greater muscle development in their forearms generally show higher magnitudes on the EMG readings and people with higher body fat content generally have smaller magnitudes.



Figure 1: Thalmic Labs Myo Bracelet

Due to the discontinuation of the Myo armband in 2018, this dataset is unlikely to be able to be replicated or in future. However, the design of the final analysis system should be independent of the recording device used (as preprocessing should solve this problem). As is the design of many systems in use today, modularity is key for these devices to remain cost effective for the end users.

4 Preprocessing

Preprocessing is a crucial step of any data analysis task. Much like any other analysis, the quality of the outputs of the system is directly dependent on the quality and accuracy of the inputs. In addition to this, deeper knowledge of the system is also useful in mapping input data values into system values.

For this system, the following is known about the nature of the data:

- **Values** - The values recorded in the data can generally be represented in scientific notation with up to 5 decimal places.
- **Data Nature** - According to the data sets provided information[1], the data recorded are raw EMG values measured by the Myo armband. As such, special treatment of these values is required to generate reasonable predictions.
- **Data Bias** - As 36 subjects were used to conduct this experiment and the overall goal of this system is to generalise a system for classification, special consideration was required to obtain an accurate representation of the whole data set.

With these goals in mind, a system was designed to both format and clean data as well as prepare it for an MLP system. To design the system the most efficiently and allowing the ability for further experimentation, the system was to be designed in two steps. The first step was the formatting and cleaning of the data files

into a format that could be directly imported into a Relational Database Management System (RDBMS). The second section was then the direct mapping of values into a format conducive to making classifications by the MLP system.

4.1 Data Format Tool

The first section of the overall preprocessing system was to develop a system to interface between raw data values and the RDBMS system storing all the training data for this project. For the sake of reproducibility, this section was designed to interface directly with the EMG Data for Gestures Data Set. Although a reasonably trivial process, it was important to structure the output data in such a way that all systems are able to interpret. The following flow chart describes the functionality of this section.

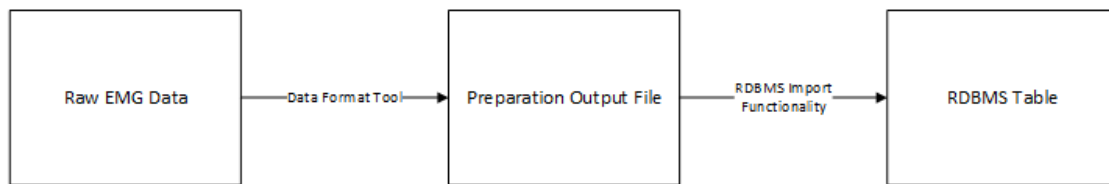
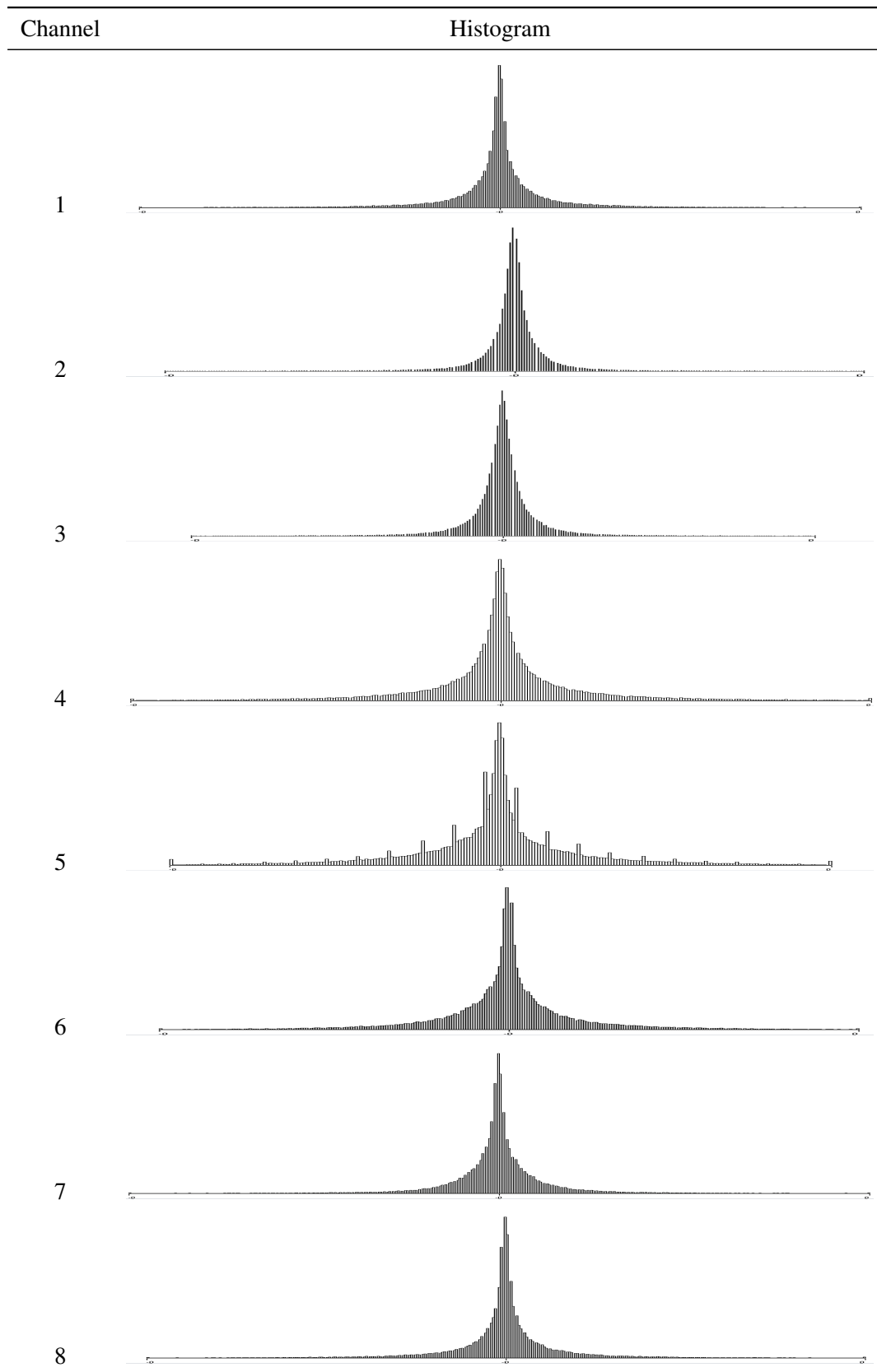


Figure 2: Data Flow of the Formatting System

Table 3 show the distribution of values across each channel. This table was created using the popular data-mining tool WEKA. Due to the size of the input values and the implementation of the WEKA interface, little data insights are directly available. However, the distribution of values for the raw data indicate that

Table 3: Histogram of values across the EMG Data for Gestures Data Set



5 Artificial Neural Networks

6 Conclusion

References

- [1] S. Lobov, N. Krilova, I. Kastalskiy, V. Kazantsev, and V. Makarov, “Latent factors limiting the performance of sEMG-interfaces,” *Sensors*, vol. 18, no. 4, p. 1122, apr 2018.
- [2] D. Dua and C. Graff, “UCI machine learning repository,” 2017. [Online]. Available: <http://archive.ics.uci.edu/ml>

Appendices

A Examples