ELEC 811 Final Project: Determining Load in Hands from EMG Data

Monica Rao
(10133414)
Queen's University
Kingston, Canada
monica.rao@queensu.ca

Manqing Zhou
(20134415)
Queen's University
Kingston, Canada
18mz17@queensu.ca

Abstract

Back pain is a common pain from fatigue in muscles that most people suffer when repetitively manually carrying heavy loads. In this study, (that was done in 2007-2008) we used EMG data that was collected from healthy adults who volunteered to participate. Each of the volunteers read a letter of information, and provided informed consent before participating in that experiment. A box with handles on either side was equipped with a switch when the box was resting on a surface, the switch was closed; when the box was lifted, the switch was open. The box could be loaded with weights from approximately 0 kg (styrofoam) to 20 kg. The subject was instrumented with bipolar EMG electrodes over six muscles on the right hand side. Electrodes were placed in accordance with SENIAM 1 guidelines. Body worn motion detection sensors (IMUs, Xsens Inc.) were also mounted at several locations on the subject. The subject was then asked to: lift the box from the floor to a platform at approximately waist height; lift the box and place it further back on the platform; lift the box and lower it back to the floor. We propose/designed and created a working multi classification model(random forest) that is able to predict next movements of lifting/placing load weights using this EMG data with an accuracy of 91%.

1 Introduction

Having lower back pain can be a result from repetitive manually carrying of heavy weighted loads. This is one of the most common cause of fatigue in the work place. Due to this it has been known that approximately three out of every four Canadians whose job includes manual material handling suffers pain due to back injury at some time.[1] This is what brought this study in mind to see how we can determine the load in the hands of a person using EMG signal data. The main purpose of this study was to be able to predict the muscle contraction level of the subject. The data files used were from five subjects that was provided to us and was stored into text file, with each data file comprises of 14 columns. Please refer to the table 1 below. The sampling rates for all the data was measured at 1000 Hz and the time for each repeat was 20 seconds.

Column 1	BB-EMG data	Column 2	TB-EMG data
Column 3	BR-EMG data	Column 4	AD-EMG data
Column 5	LES-EMG data	Column 6	TES-EMG data
Column 7	Hand switch signal	Column 8	Box switch signal
Column 9 - 14	Motion sensor data		

2 Data Analysis

Before processing any EMG data records, we preprocessed our data by removing any of the non-zero dc level from the signal data. The reason this was done was so the switch data can indicate when the subject is actively lifting the box. The data was processed for this project using few different methods of approach, by performing the following steps: data collection, preprocessing, amplitude estimation, smoothing of the EMG signals, feature extraction (rms) and normalization. Finally we will do a classification of these signals for multiple load (weight) levels. These steps will be explained further in detail in the next upcoming sections.

2.1 Data Collection

The data was collected by taking measurements of the following: The Bipolar EMG from six unilateral locations (biceps brachii, brachioradialis, triceps brachii, anterior deltoid, thoracic erector spinae and lumbar erector spinae), the data was sampled at 1000 Hz, box switch data, hand switch data; subjects also wore a hand switch which indicated when the subject hands were in contact with the box handles, and finally motion data was also collected.

2.2 Preprocessing

Preprocessing was done in order to remove any of the low frequency noises from our EMG signal data. This was done by using a high pass filter to remove any of the non-zero dc levels from the signal data. We used a butterworth high pass filter with a cutoff frequency value of 20 Hz, and an order number of 4. High pass filters with fourth order are more likely to suppress low frequency noises without changing the shape of our signal. The figure below shows our raw EMG signal data after it has gone through our high pass filter. Please refer to figure 1 below.

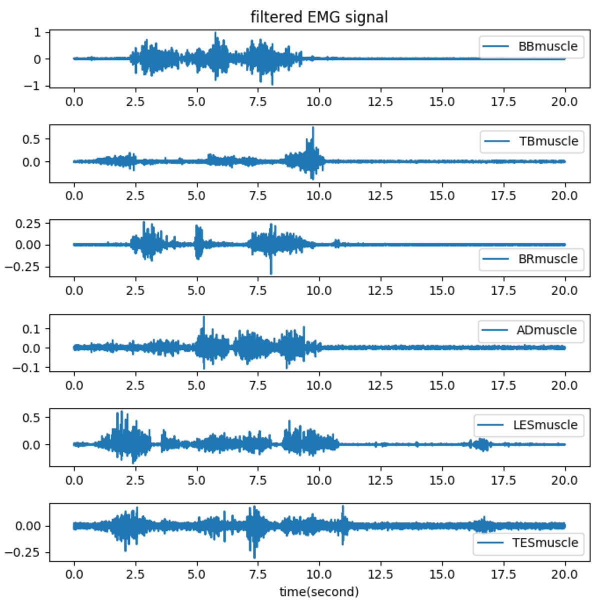

Figure 1: Prepossessed EMG signals at 20kg load with subject 2

2.3 Amplitude Estimation

The relationship between the load is directly proportional to the amplitude of the EMG signal. This correlates to, as the load increases the amplitude of the EMG signals also increases. This is quite obvious to determine as the subject increased the weight it will take more muscle (action potential/muscle firing rate) to lift it. Table 2 below shows the comparison of the two subjects with varying of weight load (low, medium and high) using the six different muscles on the right hand side.

low load (2.5 Kg)	BB	TB	BR	AD	LES	TES
subject 1	0.073863	0.012638	0.017588	0.016748	0.040995	0.023067
subject 2	0.090407	0.025871	0.02282	0.012067	0.042827	0.023393
medium load(10 Kg)	**BB**	**TB**	**BR**	**AD**	**LES**	**TES**
subject 1	0.168641	0.024361	0.058924	0.030918	0.062887	0.041497
subject 2	0.224885	0.027018	0.06029	0.02177	0.058832	0.043237
high load(20 Kg)	**BB**	**TB**	**BR**	**AD**	**LES**	**TES**
subject 1	0.28349	0.028655	0.116341	0.044348	0.071091	0.057133
subject 2	0.268567	0.03163	0.08126	0.027808	0.06469	0.047439

The six different muscles that are labeled are in the table above are as follows: biceps brachii (BB), brachioradialis (BR), triceps brachii (TB), anterior deltoid (AD), thoracic erector spinae (TES) and lumbar erector spinae (LES). Amplitude of the EMG signal used RMS (Root Mean Square) for the following weights: 2.5kg, 10kg,and 20kg for two subjects. The principal for using RMS can be shown in the equation below:

$$x_{\mathrm{rms}} = \sqrt{\frac{1}{n}\left(x_1{}^2 + x_2{}^2 + \cdots + x_n{}^2\right)}$$

Where Xn is the value of the each sample, and N is the number of samples.

2.4 Smoothing of EMG signals

The pattern of EMG signals are random because of the fact that the actual set of recruited motor units constantly changes. To address this problem, we apply a digital smoothing algorithm that outlines the mean trend of the signal development. Meaning, we remove extreme steep values from our EMG signals. As the RMS reflects the mean power of the signal, it is the preferred method for smoothing. This can be shown by using the equation below:

$$((a[i]^2 + a[i+1]^2 + +a[i+windowsize]^2)/windowsize) \tag{1}$$

The a[i] represents the EMG signal, window size is the number of the samples. In our project, we set the window size is 100 based on our total length of 20000 samples for each lift trial.

2.5 Feature Extraction(RMS)

Since we used the method of smoothing to our signals using RMS, we decide to use each result of the window as our feature. Specifically, we extracted the average value of the 100 samples to reduce the input data of our classification model as well as improve the results of prediction.

2.6 Normalization

The data was collected without recording a maximum voluntary contraction, and because of this we needed to use the sub maximal normalization. The main idea behind this was to divide the data by max rms, which is the rms value of the 20kg load. In this project we used the value of 1.068. In other papers, they asked the subject to maintain 40% or 60% of the MVC while we got percentage of maximal contraction level by normalization.

The table 3 below shows the amplitudes of the EMG signals after the normalization step was done. Comparing the results in the table to the table 2 in previous section in amplitude estimation, the values have

low load (2.5 Kg)	BB	TB	BR	AD	LES	TES
subject 1	0.069449	0.011878	0.016534	0.0157345	0.038544	0.0216722
subject 2	0.085008	0.024322	0.0214563	0.011343	0.0426755	0.219797
medium load(10 Kg)	BB	TB	BR	AD	LES	TES
subject 1	0.15857	0.022895	0.055396	0.029071	0.059131	0.038992
subject 2	0.2114523	0.0254	0.056689	0.02046	0.055312	0.04064
high load(20 Kg)	BB	TB	BR	AD	LES	TES
subject 1	0.26656	0.026939	0.109393	0.041699	0.06684	0.0537167
subject 2	0.25253	0.297342	0.76408	0.026146	0.060825	0.044598

changed slightly and the amplitude values have become smaller in size. The reason why the amplitude reduced is that we normalize it by max-rms (1.068).

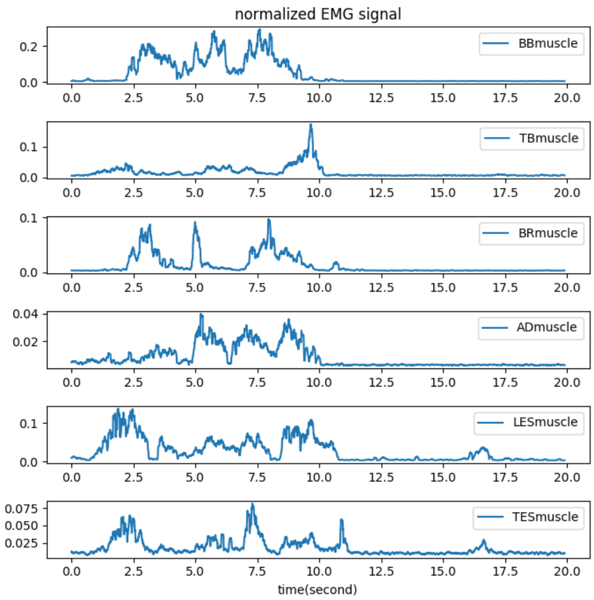

Figure 2: Plot of all normalized EMG signal for all muscles at 20 kg

The figure 2 above shows the comparison of all muscles when the subject was lifting the 20 kg load. It is obvious that the back muscles are firstly activated when the subject bends over to grab and lift the box. Afterwards, the biceps brachii muscles get contracted while lifting the box. Once this happens, the shoulder muscles get stimulated while placing the box on the platform. While putting the load of the box back, the contraction of the muscle is repeated.

The figure 3 below shows the movement of how the subject lifted the box up from the ground to a surface/platform and back down to the ground. The highlighted parts (shown in red) show where the subject's biceps brachii muscles get contracted. We can see from the biceps brachii muscles from figure 2, around 2.5-5.0 second and 7.5-10 second, the amplitude of the EMG signal is bigger than the rest of the time. Anterior deltoid muscle (which is located in the shoulder) is not activated until the subject put the box further back on the platform . This can be recognized and can be seen figure 2 that the amplitude

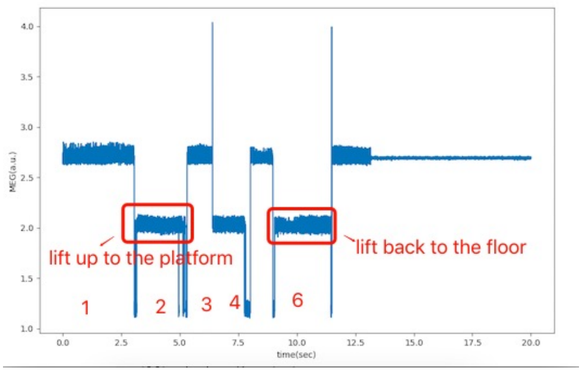

Figure 3: Plot of Box switch data

of the signal is almost smooth until it reach 5.0 second. Besides, the thoracic erector spinae muscle and lumbar erector spinae (which are located in the back closer) are firstly activated because when the subject lift the box, they need to bend the back which will contract the muscle there.

2.7 Multi-Classification Model

Each subject operated five repetitive sets for each load (six loads in total). We merged five reps of each load into one label and with each of the loads labeled from 0-5 in varying in weight (0-0kg,1-2.5kg,2-5kg,3-10kg,4-15kg,5-20kg) After that, we used a multi classifier model using knn (nearest neighbour) and random forest because we have to classify more than 2 labels. Random forest operates by constructing a multitude of decision trees. For our random forest algorithm we split our training and test data into 70% (train) and 30% (test). We set the number of trees 500 and the maximum depth 20. The reason we set the depth 20 is because we do not want to over-fit the data set. For our knn algorithm, we also split out training and test into 70%(train) and 30%(test). We set the number of neighbours based on the 10-fold cross validation and we use MSE to plot the miss classification of the model.

2.8 Related Works

The task of determining load in hands from EMG signal data is a common study to analyze with the increase of risk of back injuries in the work place from lifting heavy weights. Recently in this area, there has been many different applications attempts that use similar techniques to analyze load in hands of the individual using muscle contractions. These solutions highlight the important features of how to use EMG signals to determine various load weights.

2.8.1 The Use of EMG for Load Prediction During Manual Lifting

By using the help of the workplace safety board, the project in this paper was examining the muscle activation levels in upper extremity and trunk muscles during a manual lifting task using both hands. The paper predicted that their was going to be a correlation between the magnitude of the load in the hands and the load which in turn will be used to predict the lower back moments. A model was developed to for the squat-lift posture using the area, peak and mean of the zero-normalized EMG. The result of the mode show that the low EMG signals for the first half of the process of the lift action, this can mean that the lateral deltoid was not activated until the placement part of the lift. Also it was noted that the triceps showed very low EMG signals throughout the entire lift process.[2] From this paper we can relate the various techniques of how they relate of how they interpreted the results of the process of lifting various weights. As well as improve upon our results with our multi-classification model.

2.8.2 Low-back EMG data-driven load classification for dynamic lifting tasks

In this paper they designed numerous devices to support the back when preforming lifting tasks. The research explores various ways muscle activity using EMG signals and then be able to classify these muscles movements. The main aim of this paper was to identify the earliest time that could accurately classify the load during the lifting process. The model used a multi-nominal logistic regression (MLR) classifier that was trained and tested, and cross-validation, to classify lifted load values. The results of the model for this paper show the highest average classification accuracy at 200 ms is at 80%.[3] From this paper we can use the techniques and explore with our results to improve our overall own multi-classification model.

3 Results

The following results that we got from our multi-classification model are shown below. We have 179822 samples in total and divided it into 6 loads. The first table (please refer table 4) is the multi-classification model summary of our random forest algorithm, the highest f1-score is 0 kg while the lowest is 5 kg. This means that it is easy to predict when the subject is holding nothing. The reason behind this is because, when people move something light, the contraction of the muscle is consistent for five reps. This means the pattern of the EMG signals are following this same pattern. Most of the errors happened at the load of 5 kg and were due to miss classification. As mentioned before, we used a high-pass filter and averaged the data by window size. From using our large data set and using a depth value of 200 with 500 trees in our algorithm we were able to achieve a high accuracy score. Our overall accuracy score for our model was approximately 91%.

	precision	recall	F1-score	support
0	0.97	0.98	0.98	29918
1	0.85	0.90	0.88	29922
2	0.83	0.92	0.87	29871
3	0.94	0.88	0.91	30047
4	0.96	0.89	0.93	30115
5	0.94	0.90	0.92	29949
Micro Avg	0.91	0.91	0.91	179822
Macro Avg	0.92	0.91	0.91	179822
Weighted Avg	0.92	0.91	0.91	179822
Accuracy score	0.913008491			

The second table (please refer to table 5 below) shows the confusion matrix of our model, the main idea behind this, is to show how many samples are recognized as true positives or true negatives (correctly or mistakenly) in our model. The next algorithm we used

		Actual load					
		0	1	2	3	4	5
Predicted load	0	29237	132	393	60	54	42
	1	187	27078	1550	449	256	402
	2	231	1233	27402	396	209	400
	3	219	1285	1317	26514	282	430
	4	88	1034	1124	375	26887	607
	5	74	1143	1047	377	247	27061

is the KNN (nearest neighbour). In order to choose the right neighbor for this algorithm, we assessed the model by using a 10-fold cross validation.

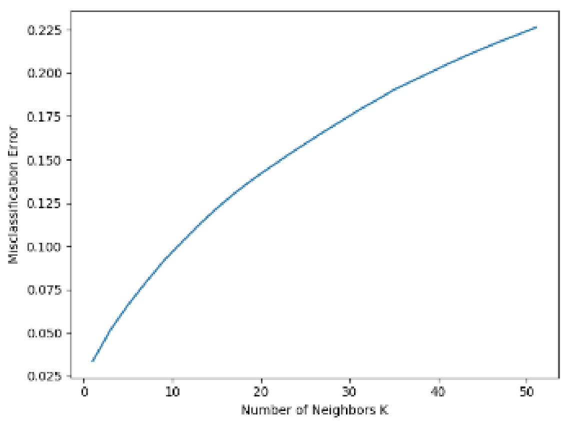

Figure 4: Plot showing the MSE (mean square error) for our KNN model

As shown in the figure 4 above (please refer to figure 4), the more the number of neighbors the higher the error. When the number of the neighbors reaches certain limited point, the error remains the same. As a result we chose a value of k = 10 to get an accuracy at approximately 90%.

The table below (please refer to the table 6) shows the model summary of our nearest neighbours algorithm. Our model has a high accuracy for heavy loads which is directly correlated to our expectations. Since, fa-

	precision	recall	F1-score	support
0	0.95	0.96	0.95	29973
1	0.87	0.91	0.89	30032
2	0.87	0.89	0.88	30028
3	0.91	0.89	0.90	30041
4	0.93	0.92	0.90	30041
5	0.93	0.90	0.91	29881
Micro Avg	0.91	0.91	0.91	179822
Macro Avg	0.91	0.91	0.91	179822
Weighted Avg	0.91	0.91	0.91	179822
Accuracy score	0.909137925			

tigue in muscles is more prone to happen when lifting heavier loads, and putting more strain on the muscles.

4 Discussion

We used high-pass filter to get rid of the low frequency noise and smooth the signal while extracting the window rms as a feature extraction. One of the research papers we referenced used the peak and mean value of the EMG signal. The value mean is similar to the parameter rms. We applied sub-maximal normalization while they normalized the signal to a baseline determined from MVC (maximal voluntary contraction). The principle is to get the maximal contraction and adjust other signals based on it. The reason we do not use normalized MVC is that producing an MVC can be very strenuous for the subject.

We can see from the shape of the EMG signal that the heavier the load being lifted, the more different muscles will have to work to be stabilize and maintain holding the load. Also, even with the same load, different subjects got different results. Since it is about lifting load, the most involved is biceps brachii muscle, the amplitude of which is much higher than other muscles. We applied two models to predict the future movement of the signal. The random forest got a slight higher accuracy than the knn.

The limitations of our project is that the condition of the subject plays a key role in the effects of lifting weights, and we do not have any control over this. For example, when the weak person tries to lift the 20kg load, the timing of contraction is different (the period is longer). It is possible that the load prediction can be improved by including subject characteristics as input variables.

5 Conclusion

Lower back pain comes from repetitive manually lifting of increasing heavier loads. It is important we study and look into muscles that contribute to the motion of picking up heavy loads and placing them on a surface. This is what inspired the aim of this project to be able to develop a working model to classify predict multiple movements of loads weights using EMG data with a high accuracy of approximately 91%. Our model used random forest to be able to predict and classify the multiple labeled loads. We hope to continue to improve our model to be able to get even higher results. Using this research and study data will help those in the workplace who deal with lifting and material handling. EMG signals are important to study and research as they are useful to for understanding movements (action potential) in muscles.

References

[1] G. of Canada, "health and safety," April 2019. [Online]. Available: https://www.ccohs.ca/oshanswers/ergonomics/mmh/hlth$_haz.html$

[2] S. Chan, "The use of emg for load prediction during manual lifting," Sept 2007. [Online]. Available: https://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/OKQ/TC-OKQ-871.pdf

[3] D. G. E. P. K. B. Lauro Ojeda, Daniel D. Johnson, "Low-back emg data driven load classification for dynamic lifting tasks," Feb 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5814006/