EMG Gesture Classifier

Capstone

Derek Andersen 6/20/2022

Agenda

- Objective
- Data
- Visualizations
- Modeling Process
- Model Comparison
- Model Tuning
- Recommendations
- Improvements

Objective

- Our goal is to investigate the feasibility of building a model to predict which of seven pre-defined hand gestures (or no gesture) a user is making while wearing an armband containing 8 electromyography (EMG) sensors.

Data

Training Set

- Our dataset consists of raw readings from 8 EMG sensors located on the forearm of 36 subjects. The observations were recorded while the users performed a series of 7 static hand gestures over two separate runs.
  - Source: https://archive.ics.uci.edu/ml/datasets/EMG+data+for+gestures
- Training Data
  - 4,237,908 observations (4,237,907 after removing one observation with missing values)
  - 11 features (subject number is not in the raw data files, but is present in the folder structure)

Data

Training Set - Gestures

- Gestures:
  - No gesture
  - Hand at rest
  - Hand clenched in a fist
  - Wrist flexion
  - Wrist extension
  - Radial deviations
  - Ulnar deviations
  - Extended palm
- The majority of the observations in the dataset are unlabeled (class = 0), i.e., they are observations recorded when the user was not making one of the 7 labeled gestures (class = 1 through 7). As a result, if we take an observation at random from the full dataset there is a 64% chance that it will *not* be from a labeled gesture.

Data

Training Set - Features

- Time
  - The time at which the observation was recorded (in milliseconds from the start of the run).
  - The interval for the observations in the dataset was generally 1ms, but sometimes there are longer gaps between observations. I.e., for subject 1's first run the first time value is at 1ms, but the second time value is at 5ms.
- Channel 1-8
  - Each channel feature represents one of the EMG sensors in the armband.
  - The armband records each EMG signal's amplitude as a signed 8-bit integer. In this dataset the signals have been converted to floating-point values, so each channel can have a value between -0.00128 and 0.00127.
- Subject
  - There are 36 subjects in this dataset. Each subject performed a set of gestures twice, although not all subjects performed the last gesture in the set.
- Class
  - There are 7 different gestures (1-7) in this dataset plus "no gesture" (0), for a total of 8 separate classes to predict.
  - As noted above, not all subjects performed the last gesture, so there are fewer observations for that class.

Data

Training Set - Basic Statistics

- EMG channel values:
  - All channels have the same minimum and maximum values (-0.00128 to 0.00127).
  - All channels have the same median value (-0.0001).
  - The mean values for all channels are very close to zero.

Channel	Mean	25%	75%
1	-0.0000079	-0.000030	0.000020
2	-0.000009416073	-0.000040	0.000020
3	-0.000010	-0.000040	0.000030
4	-0.000009637835	-0.000060	0.000040
5	-0.00001599724	-0.000080	0.000050
6	-0.00001085528	-0.000060	0.000030
7	-0.000009364639	-0.000040	0.000020
8	-0.000010	-0.000030	0.000010

Data

Training Set - Basic Statistics

- Gesture classes:
  - More than 64% of the observations have class = 0 (i.e., not one of the pre-defined gestures).
  - There are only about 5.5% as many observations for gesture 7 than the average number of observations for the other gestures.

Gesture	Number of Observations
0	2725157
1	253009
2	251733
3	251570
4	250055
5	249494
6	243193
7	13696

Data

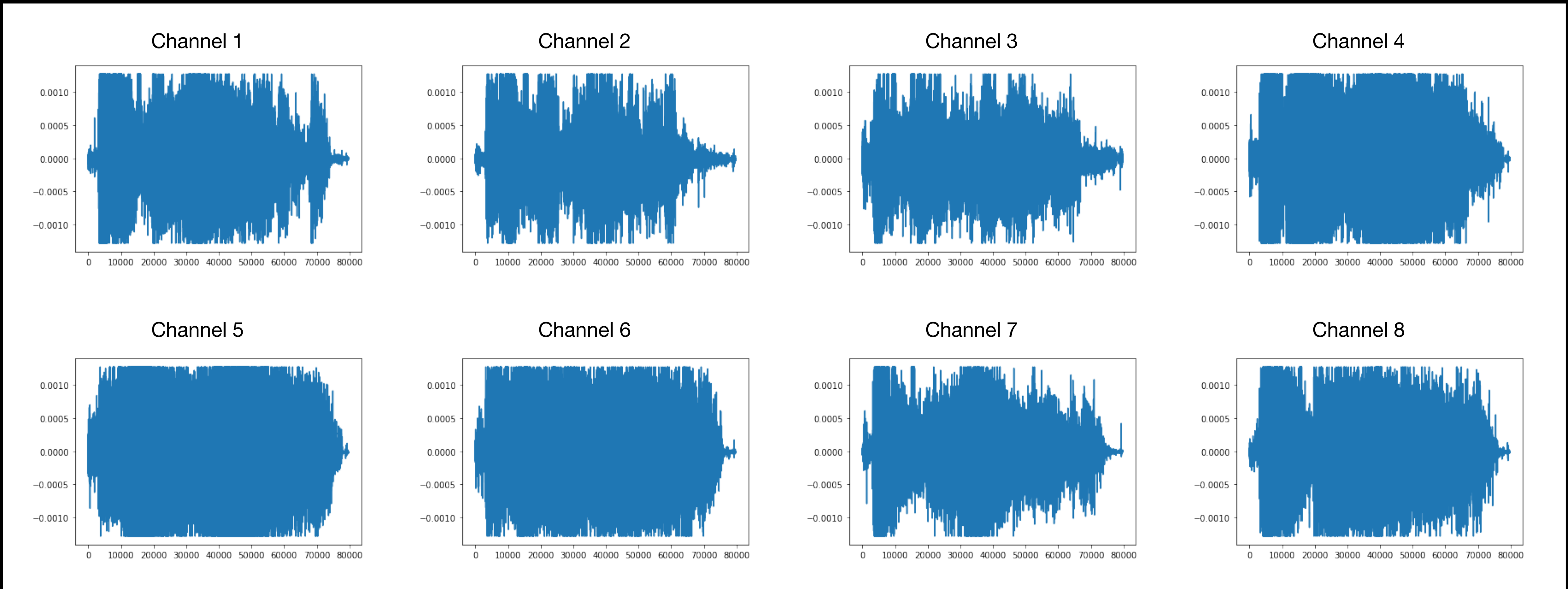
Training Set - Basic Statistics

- Subject:
  - While all subjects performed roughly the same series of gestures, some subjects have many more observations than other subjects. Subject 31 has the fewest observations at 91,023 and subject 13 has the most observations at 153,240.
  - The mean number of observations per subject was 117,720.

Visualizations

EMG Channel Values Over Time

- EMG values for all channels are noisy!



Visualizations

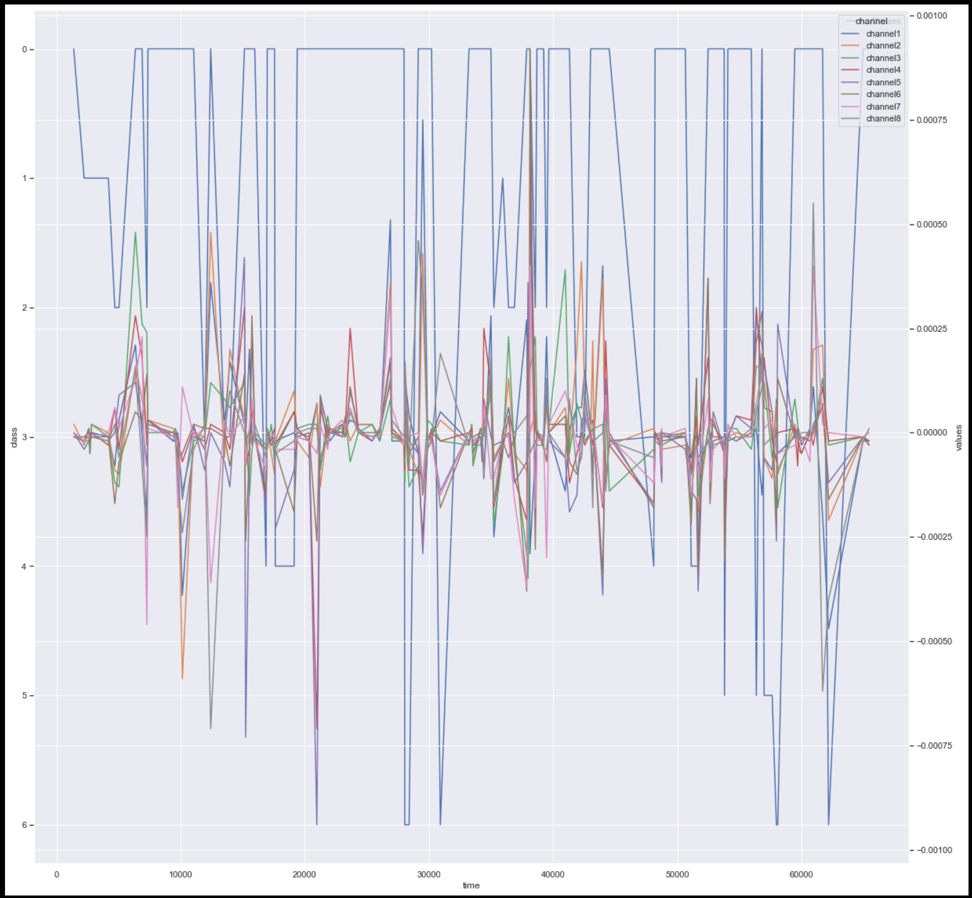
EMG Channel Values Over Time - Subject 1

- Consider a small sample (0.1%) of the EMG channel observations for one subject, plotted against the timestamp for the observation.
- There are more distinctly visible peaks and valleys for different sensors, but the data is still quite noisy.

Visualizations

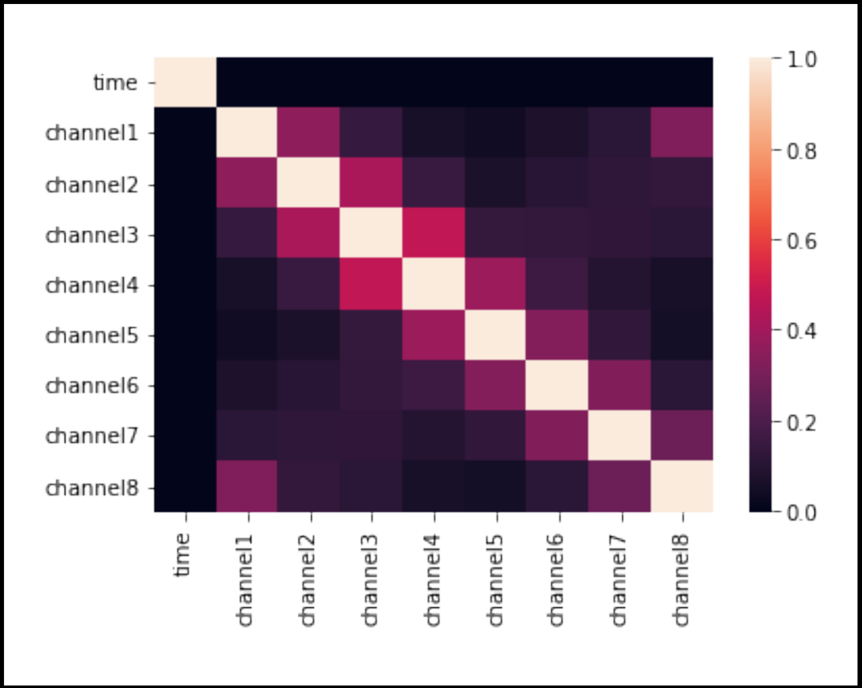
EMG Channel Values Over Time - Subject 1

- Overlaying a plot of the class values (i.e., gestures) onto that same EMG plot lets us better look for relationships between EMG patterns and gestures.
- However, it's still difficult to discern an obvious pattern based on this type of visualization.



Visualizations
Correlation Heatmap

- While the correlation between neighboring EMG sensor channels is clear, the correlation between any two sensors is not so high that we should exclude any of them.
    - The highest correlation is between channels 3 and 4 at 0.47.
    - Note that channel 1 is correlated with channels 2 and 8, which makes sense given that the individual sensors are wrapped completely around the subject's arm, so the first sensor is physically close to the last sensor.
- There is no significant correlation between time and any of the EMG sensor channels.



Modeling Process

- Matching the EMG sensor values (and potentially the subject number, to account for the possibility of user-specific model training) is a multiclass classification problem with 8 possible classes (7 predefined gestures and no gesture).
- Given the size of the dataset (~4.2 million observations), a 10% random sample of the dataset was used to create the training and test sets.
- 75% of the data was used for training and 25% was used for testing.
- Two versions of the 10% sample dataset were used, one with the subject feature included and one without.
- We trained and compared three algorithms using 5-fold cross-validation:
  - k-Nearest Neighbor
  - Random Forest
  - Extra Trees
- We then compared the balanced accuracy of each model on the training set to choose the best-performing model, which was then compared against the test set and further tuned.

Model Comparison

Training Set

- We compared the models using balanced accuracy, which is the average of the combined true positive and true negative rates for each class.
  - Balanced accuracy is an important metric here because we have a significant class imbalance problem--64% of our overall observations are for class 0. A model that always chose the majority class would therefore have an accuracy of 64%, but a *balanced* accuracy of only 0.125 (using Scikit-Learn's multiclass balanced accuracy).
- The Random Forest model using the training set with the subject included was the best-performing model, but it was only very slightly more accurate than the Extra Trees model (on both training sets).
  - K-Nearest Neighbors performed better than choosing the majority class, but noticeably worse than the other two models.
- All three models performed better on the training set that included the subject than on the set that did not.

Model	Balanced accuracy (with subject)	Balanced accuracy (without subject)
kNN	0.4220	0.3623
Random Forest	0.6549	0.6491
Extra Trees	0.6444	0.6396

Model Tuning

Evaluation of Random Forest Model on the Test Set

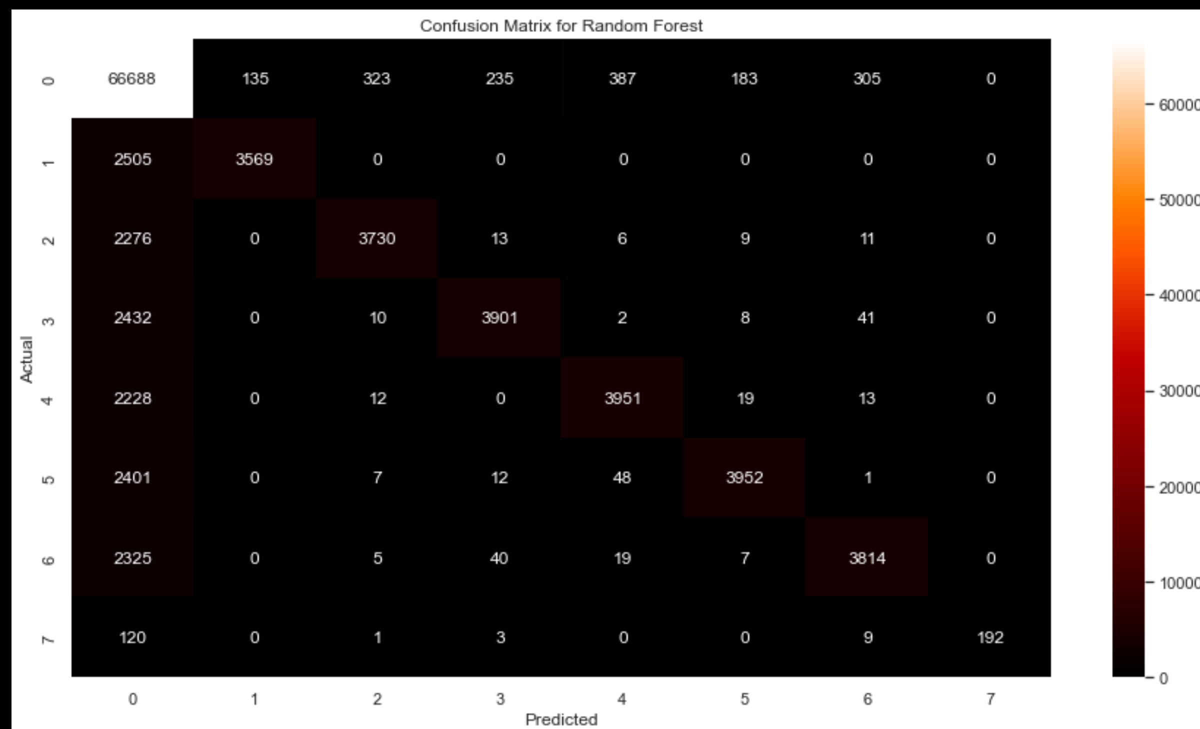
- Using 5-fold cross validation, a Random Forest model with 100 estimators was selected.
- The trained model achieved a balanced accuracy score of 0.6559 on the test set, very close to the score on the training set.
- Precision was fairly good on the test set, with greater than 90% precision for all classes except class 0.
- Recall was weaker on the test set, with all classes except class 0 between 59% and 63% recall.

Class	Precision	Recall
0	0.82	0.98
1	0.96	0.59
2	0.91	0.62
3	0.93	0.61
4	0.90	0.63
5	0.95	0.62
6	0.91	0.61
7	1.00	0.59

Model Tuning

Confusion Matrix

- We can see that the model generally predicted either the correct class (on the diagonal) or class 0.
  - I.e., gestures were generally recognized as either the correct gesture or *no* gesture, rather than being recognized as another pre-defined gesture.
- However, it's clear that the model predicted class 0 for observations that were actually classes 1-7 very frequently.



Model Tuning

Manual Grid Search

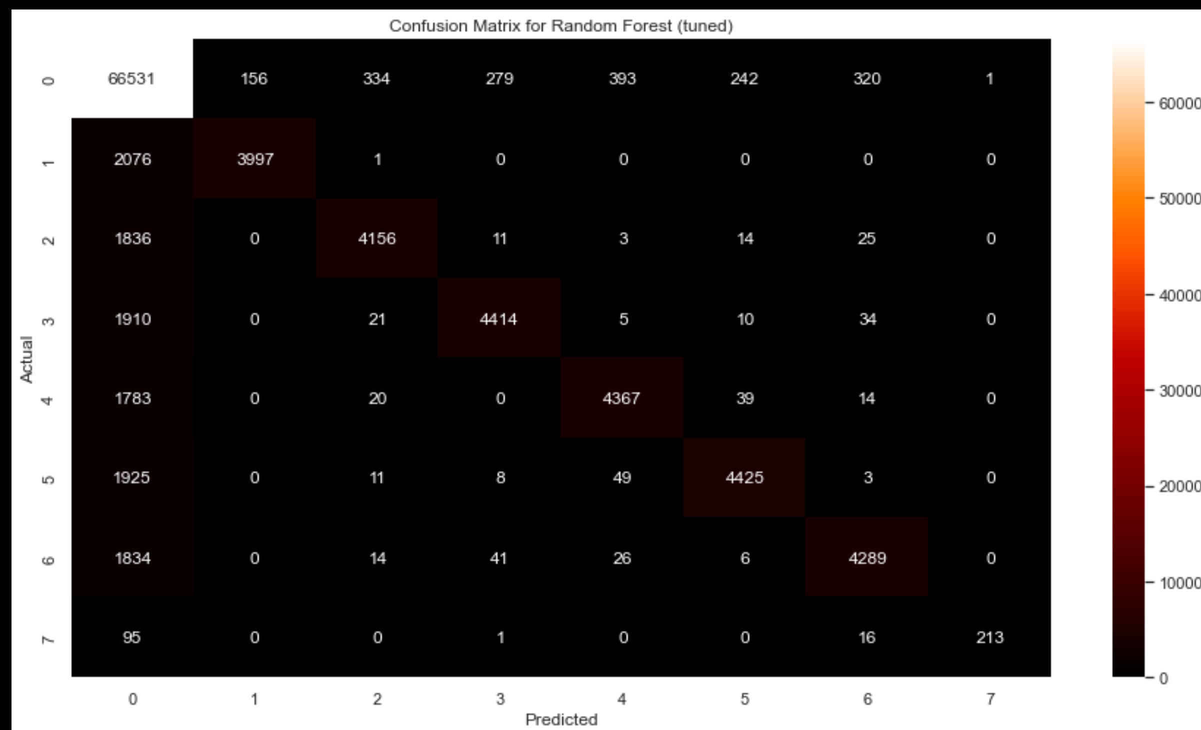
- We ran a manual grid search (still using 5-fold cross validation) varying the number of estimators in the Random Forest model, then compared the predictions of the best-performing model against the test set.
- The best score was achieved with 30 estimators (compared to 100 in the previous Random Forest model), with a balanced accuracy of 0.6584 on the training set.
- Balanced accuracy against the test set improved slightly more, to 0.7184.
- Precision on the test set was comparable to the un-tuned Random Forest model.
- Recall on the test set was slightly better than the un-tuned model, with the lowest recall in the tuned model (66%) being higher than the highest recall in the un-tuned model (63%--excluding class 0).

Class	Precision	Recall
0	0.85	0.97
1	0.96	0.66
2	0.91	0.69
3	0.93	0.69
4	0.90	0.70
5	0.93	0.69
6	0.91	0.69
7	1.00	0.66

Model Tuning

Confusion Matrix - Tuned Model

- Generally the confusion matrix for the tuned model is similar to the un-tuned model.
- As we might expect from the increased recall values, however, the number of observations where the actual class was 1-7 but the predicted class was 0 have gone down slightly.



Recommendations

- Given the rarity of misclassified gestures (i.e., a gesture was classified as another pre-defined gesture rather than the correct gesture *or no gesture*), the Random Forest model with 30 estimators does a reasonable job of recognizing gestures based on the 8 EMG sensors and the subject number.
- However, it is likely that with the demonstrated level of performance users would be frustrated with how frequently the model would register "no gesture" in use. For this set of gestures with this input device, all of the tested models are unlikely to be worth implementing at this time.

Improvements

- The device used to generate this dataset was a Thalmic Labs Myo, which has been discontinued for several years (the company rebranded itself as North and was then acquired by Google in 2020). It is likely that any replacement EMG sensor will have different measurement characteristics that might affect the choice of model; without comparing data from multiple EMG sensors it is unclear how generalizable any model will be.
- While there are more than four million observations in this data, they are for a limited number of subjects performing a set of static gestures in order. A larger-scale study with more participants and more repetitions per participant is likely to improve the quality of the resulting model.