Optimising Machine Learning Models for Gesture Classification in Music

MANCHESTER 1824
The University of Manchester
Alliance Manchester Business School

Samina Bibi[1], University of Manchester
Supervisors: Chris Rhodes[2] and Richard Allmendinger[3]

Nuffield Foundation

Introduction

- Gestural interfaces (GIs), electromyographic (EMG) data and machine learning (ML) can be used to create novel human-computer interaction (HCI), allowing for musicians to embody musical composition and grant them heightened creativity.
- GIs are peripherals that allow users to interact with computers via physical gestures.
- EMG data is a form of biometric data (measuring electrical activity of skeletal muscles) gathered via GIs. EMG data is useful for interactive music practice as it can be used to accurately represent musical gestures. EMG data can then be sonified (i.e. representing data in terms of sound), resulting in interesting music compositions.
- Musicians can take advantage of GIs & EMG data by using novel performance gestures (musical actions) in musical composition.
- EMG data analysis can be complex and manual processes for analysis is inefficient.
- ML can be used to classify performance gestures in music via training models with EMG data (there are several ML models that can be used).
- Currently, the evaluation of GIs and ML in musical contexts is conducted on a qualitative level.
- However, there is a lack of quantitative evaluation of GIs and ML in music.
- Therefore, this study observes retrieved accuracies of all adjustable ML models in Wekinator (some models do not allow for parameter optimisation); the effects of model parameter optimisation on improving model accuracy when classifying performance gestures and find which ML model is suitable for different music contexts.
- Results show that parameter optimisation can have varying effects on different models.

Research Questions

1. Which ML model in Wekinator is the most optimal when classifying performance gestures in music?
2. How does optimising ML models affect model behaviour and accuracy when used to classify a performance gesture from EMG data?
3. Does ML model choice depend on the musical context?

Methodology

Gestural Interface – Myo armband

- The gestural interface (GI) being used in this project is the Myo armband.
- The Myo can access raw EMG data via 8 electrodes around the interface (see Figure 1).

Data Acquisition

- A single performance gesture dataset was used for this study.
- This was collected by Rhodes et al. (2019) via the Myo GI. Their study was on the efficacy of ML models in music.
- Figure 2 shows an example of the performance gesture used in Rhodes et al.'s 2019 study.
- Figure 3 shows a pipeline of their data acquisition process.

Wekinator and ML models used

- Wekinator will be used to interact with several supervised ML models. Those which will be optimised are italicised. These models are:

Classifier models:

- *Support Vector Machine*
- *K-Nearest Neighbour (k-NN)*
- *AdaBoost.M1*
- Decision Tree
- Decision Stump
- Naïve Bayes

Continuous models:

- *Linear Regression (LR)*
- *Polynomial Regression (PR)*
- *Neural Network (NN)*

Data Post-Processing

- We will be applying our own post-processing stage to Rhodes et al.'s (2019) dataset (see Figure 4).
- The mean absolute value (MAV) feature extraction method will be applied to the dataset.
- This is because the MAV is the optimal feature extraction (Arief et al., 2015) for EMG data analysis.

Methodology

Fig. 1 – Myo armband

Fig. 2 – Studied performance gesture

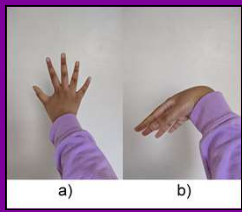
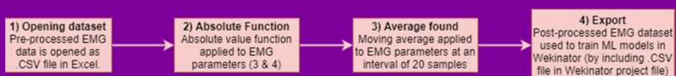


Fig. 3 – Flowchart showing Rhodes et al.'s (2019) data acquisition process (and pre-processing stage)

Fig. 4 – Flowchart showing our post-processing stage applied to the pre-processed EMG dataset used in this study.



Results: pre-processed vs post-processed dataset

Fig. 5 – Continuous Models

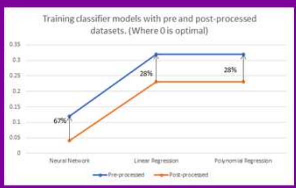
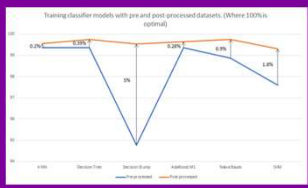


Fig. 6 – Classifier Models



Results: optimising ML models

Table 1. Table showing the cross-validation accuracy (10 folds) of the default continuous models available in Wekinator, when they are trained with pre-processed and post-processed performance gesture datasets, where percentage increase between both datasets is also provided. Cross validation is measured where 0 is optimal.

No. of nodes per layer	No. of layers = 1	No. of layers = 2	No. of layers = 3	Percentage increase (from layers = 1 to 3)	Percentage decrease (from no. of nodes = 1)		
					1 layer	2 layers	3 layers
1	0.04	0.03	0.03	25	0	0	0
5	0.03	0.03	0.03	0.0	-25	0	0
10	0.03	0.03	0.03	0.0	-25	0	0
100	0.43	0.33	0.42	2.3	980	1000	1300

Fig. 7 – SVM (Linear kernel) Model

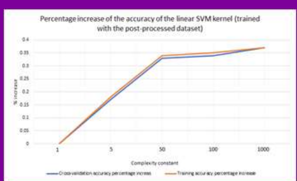


Fig. 8 – k-Nearest Neighbour Model

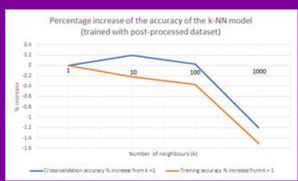


Fig. 9 – AdaBoost.M1 (decision stump) Model

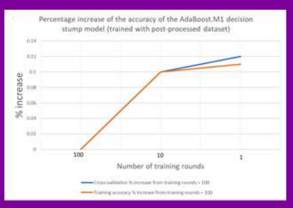
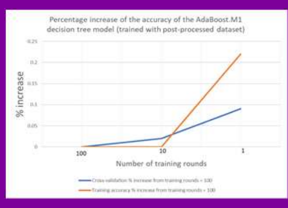


Fig. 10 – AdaBoost.M1 (decision tree) Model



Results

Continuous Models

- Post-processed dataset (with MAV) improved continuous model accuracy when classifying the studied performance gesture dataset (see Figure 5).

Classifier Models

- Post-processed dataset (with MAV) improved the classifier model accuracy when classifying the studied performance gesture dataset (see Figure 6).

Optimising Continuous Models

- LR/PR model had no change in model accuracy.
- NN improved when no. nodes and layers increased gradually (see Table 1).

Results (cont.)

Optimising Classifier Models

- **SVM.** Increasing parameter values of all kernels (i.e. linear, polynomial and RBF) improves model accuracy (see Figure 7).
- **k-NN.** Increasing no. of neighbours (k) did not improve model accuracy (see Figure 8). This is because of overfitting (the model became biased towards an output).
- **AdaBoost.M1.**
  - *Decision Stump*. Model accuracy decreased with no. of training rounds (see Figure 9).
  - *Decision Tree*. Training accuracy decreased when no. of training rounds decreased. However, cross-validation accuracy improved (see Figure 10).

Discussion

Training models with post-processed dataset

- Model accuracy of all models improved with this dataset (MAV applied).
- MAV reduces variance – allows for better model predictions.
- Decision stump classifier improved the most because it is less susceptible to the negative effects of overfitting, so it benefits from a less varied dataset (caused by applying MAV), resulting in the model making better predictions.

Optimising Continuous Models

- **LR/PR Models.** Showed no changes in accuracy. This is because a regression model is being used for classification.
- **NN Model.** Model accuracy improved when no. of nodes increased gradually. When encountering large values, model accuracy is adversely affected. (see Table 1). This is because of overfitting (i.e. being too biased to an output).

Optimising Classifier Models

- **SVM Model.** Increasing parameters improved model accuracy (see Figure 7). This is because the model can better identify patterns in the data, leading to better classification.
- **K-NN Model.** When k is significantly increased, accuracy decreases (see Figure 8). This is because the number of neighbours is considered rather than the distance between neighbours (Cunningham & Delaney, 2020).
- **AdaBoost.M1 Model**
  - *Decision Stump*. Accuracy decreased with training rounds (see Figure 8) because rounds are used by the algorithm to adjust the training dataset to 'learn' from errors.
  - *Decision Tree*. Training accuracy decreased (training dataset is adjusted to account for errors made in each round), whilst cross-validation accuracy increased when training rounds are decreased (no overfitting, can better handle unseen data).

Conclusions

- **Use of continuous models in music composition.** can be used to map and sonify EMG data, improving a musician's ability to embody music.
- Improving continuous model accuracies means that musicians will not be hindered by inaccurate representations of their gestures and will also improve the use of continuous models in real time sonification (representing data in terms of audio) of EMG data.
- **Use of classifier models in music composition.** can be used to categorise EMG data, to recognise the gesture being done, this can be used to trigger pre-defined actions to occur in musical composition.
- Improving classifier model accuracies will mean that the musician will not be hindered/distracted by the misclassification of their performed gesture, which will allow them to better enhance their music (through the triggering of pre-defined actions).
- Therefore, our findings show that model choice is dependent on musical context.
- Both model types (classifier/continuous) allow for different ways to enhance a musician's work.

Further Work

- Look into how optimising continuous models can affect signal mapping output.
- Look into the effects of increasing parameter values to extremely large values on model accuracy.
- Observe the impact of training dataset size on model accuracy.

References

- Arief, Z., Sulistijono, I. A., & Ardiansyah, R. A. (2015, September). Comparison of five time series EMG features extractions using Myo Armband. In 2015 International Electronics Symposium (IES) (pp. 11-14). IEEE.
- Cunningham, P., & Delany, S. J. (2020). k-Nearest Neighbour Classifiers--. arXiv preprint arXiv:2004.04523.
- Rhodes, C., Allmendinger, R., & Climent, R. (2019, November). New Interfaces for Classifying Performance Gestures in Music. In International Conference on Intelligent Data Engineering and Automated Learning (pp. 31-42). Springer, Cham.

Acknowledgements

Thank you to the Nuffield Foundation for providing the opportunity to work on this research placement. I would also like to thank Chris Rhodes and Richard Allmendinger for their support throughout this project and for providing the raw EMG dataset of the performance gesture studied in this paper, gathered in their 2019 study.

1: saminabibi847@gmail.com 2: Chris.rhodes@manchester.ac.uk 3: Richard.allmendinger@manchester.ac.uk