Supplementary Material of "Towards Multimodal MIR: Predicting individual differences from music-induced movement"

In this supplementary material, we present more experimental results and details about the Evaluation Metric and Bayesian Regression that could not be included in the main manuscript due to the lack of space.

Evaluation Metrics

(a) Root-Mean Square Error (RMSE): If $\hat{y}_i$ is the predicted value of the i^{th} sample and y_i is the corresponding true value for total n samples, then the RMSE estimated is defined as:

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{1}$$

We can interpret the model's performance using RMSE. Consider X being the ground truth value and Y is the RMSE score, then we can say that the model's prediction value will be accurate in the range $X - Y$ to $X + Y$.

(b) R^2 Score: If $\hat{y}_i$ is the predicted value of the i^{th} sample, y_i is the corresponding true value for total n samples, and $\bar{y}$ is the mean of the ground truth data, the estimated R^2 is defined as:

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

Principal Component Regression

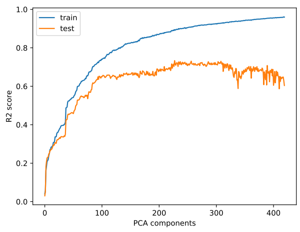
The Linear Regression model was used to predict the EQ and SQ values but we found that the model was highly overfitting. So, we took Principal Components of the features for this model. We calculated the RMSE and R^2 scores for both of the aforementioned tasks after taking the Principal Components. We repeated this experiment by varying the number of principal components.

EQ

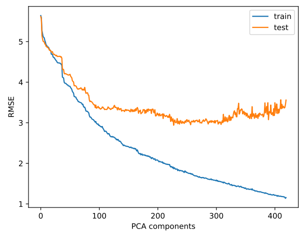
From Figs. 1a and 1b we can say that using position data for feature extraction (Sec. 3.3), the model started to overfit after having more than around 240 principal components since the RMSE started increasing and R^2 score started decreasing for testing data. Similarly, from Figure 1c and 1d, we can say that using velocity data for feature extraction, the model started to overfit after having more than around 140 principal components.

SQ

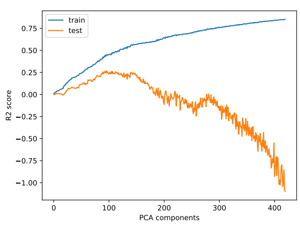
From Figs. 2a and 2b, we can say that using position data for feature extraction, the model started to overfit after having more than 260 principal components. Similarly from Figure 2c & 2d when we use velocity data for feature extraction, the model started to overfit after taking more than 170 principal components.



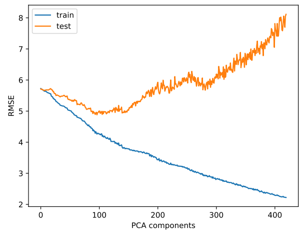
(a) R^2 score on Position Data



(b) RMSE on Position Data

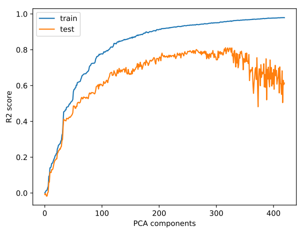


(c) R^2 score on Velocity Data

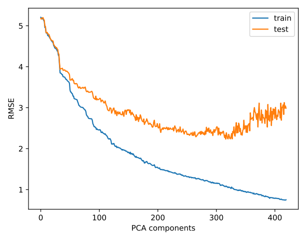


(d) RMSE on Velocity Data

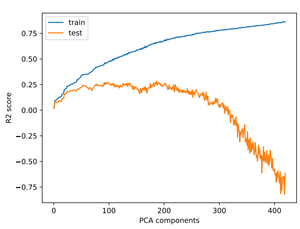
Figure 1: PCR Results on EQ



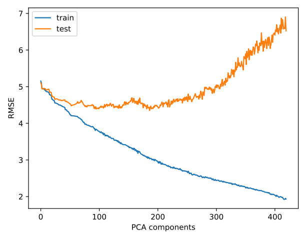
(a) R^2 score on Position Data



(b) RMSE on Position Data



(c) R^2 score on Velocity Data



(d) RMSE on Velocity Data

Figure 2: PCR Results on SQ

Bayesian Regression

The model for the Bayesian Regression with the response sampled from a normal distribution is

$$y \sim N(\beta^T x, \sigma^2 I) \tag{3}$$

where β is the weight vector, x is the feature vector and σ is the standard deviation. Here β and σ are the model parameters. The goal is not to find the single best value of model parameters but to determine the posterior distribution for the model parameters. The posterior probability of the model parameters can be defined as

$$p(\beta|D) \propto p(D|\beta)p(\beta) \tag{4}$$

$$\beta \sim N(0, \sigma_\beta^2 I_d) \tag{5}$$

where $p(\beta)$ is the initial probability distribution, also known as prior distribution and $p(D|\beta)$ is known as the likelihood function. Using these approaches, we attempted two tasks 1. EQ and SQ Prediction 2. Personality Prediction.

Due to its robustness, it is evident that the model did not overfit on the dataset as the R^2 score kept on increasing with the number of principal components. The R^2 scores for training and testing set are 0.92 and 0.86 when all the features are considered. We also repeated this experiment by varying the number of principal components incrementally to test its robustness.

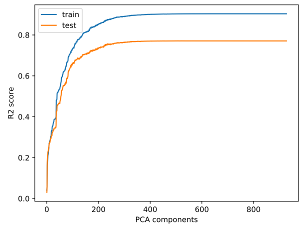
EQ

From Figs. 3a & 3b, we can say that using position data, the maximum R^2 score achieved is 0.76 and minimum RMSE is 2.73. From Figs. 3c & 3d, we can say that using velocity data, the maximum R^2 score is 0.42 and minimum RMSE is 4.35. The RMSE and R^2 scores become somewhat saturated at some point but still gets better marginally with the increase in principal components.

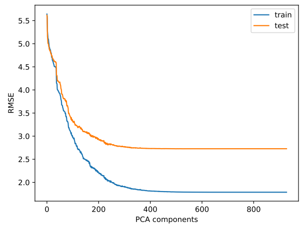
SQ

From Figs. 4a & 4b, we can say that using position data, the maximum R^2 score is 0.82 and minimum RMSE is 2.18. From Figs. 4c & 4d, we can say that using velocity data gained maximum R^2 score of 0.44 and minimum RMSE of 3.82. In the paper, we reported the RMSE and R^2 scores of Bayesian Regression without taking any principal components.

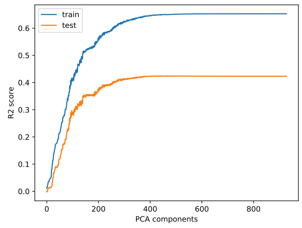
In EQ, using position data, the RMSE and R^2 is 2.72 and 0.77. Similarly, using velocity data, the RMSE and R^2 is 4.43 and 0.42. In SQ, using position data, the RMSE and R^2 is 2.16 and 0.86. Similarly, using velocity data, the RMSE and R^2 is 3.83 and 0.46. Since, the results were close but better than that of after performing the PCA, we present all the results in our paper for Bayesian Regression without performing PCA.



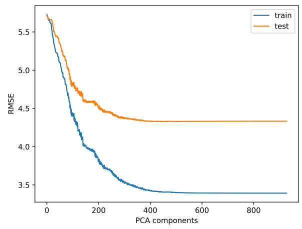
(a) R^2 score on Position Data



(b) RMSE on Position Data

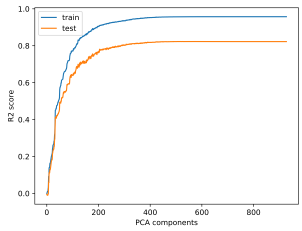


(c) R^2 score on Velocity Data

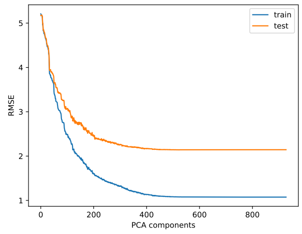


(d) RMSE on Velocity Data

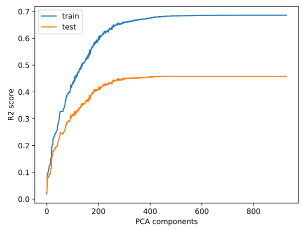
Figure 3: Bayesian Regression Results on EQ



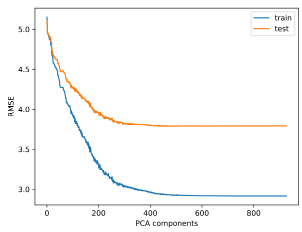
(a) R^2 score on Position Data



(b) RMSE on Position Data



(c) R^2 score on Velocity Data



(d) RMSE on Velocity Data

Figure 4: Bayesian Regression Results on SQ

Correlation between BFI Personality Traits

	O	C	E	A
C	-0.093	-	-	-
E	-0.003	0.128	-	-
A	0.021	0.341	0.358	-
N	0.217	-0.289	-0.225	-0.292

p<0.05	p<0.01

Table 1: Results of the Spearman Correlation between the BFI personality traits.