COGS 209 Project Peer Review

Sowmya Manojna Narasimha

Summary of the Study

In this study, the bidirectional relationship between maternal depression/anxiety and infant anxiety/attachment patterns is analyzed. The authors use an open source dataset consisting of maternal report questionnaires and infant behavior questionnaires. A sample of questions that best addressed infant attachment and maternal depression were selected for all subsequent analysis. These features will henceforth be referred to as infant features and maternal features respectively.

The authors then trained regression models that could predict infant attachment from the maternal features. The models that are considered in the study are Linear Regression and Random Forest Regression. Feature selection was done for both the models considered above. In the case of Linear Regression, Lasso regularization was used to select important maternal features and in the case of Random Forest Regression, GINI impurity score was used to select important maternal features. In the case of the Random Forest model, polynomial fits (degrees 1 to 4) were applied on the top 10 maternal features and the best fit was obtained for a degree of 1. The authors also acknowledged and identified shortcomings and areas of improvement in the dataset used for the study.

Review

While the authors did a wonderful job at studying the predictability of infant attachment/anxiety from maternal depression/anxiety features, I felt that some of the details and terms mentioned in the study could be studied in detail or explained further.

These points are as follows:

1. In section 1 *Introduction*, the authors mention that the bidirectional relationship between maternal anxiety/depression and infant attachment/anxiety is to be studied. However, only the predictability of infant anxiety/attachment based on maternal anxiety/depression is studied.

2. In section 3.2 *Infant attachment*, the authors mention that certain features (i.e.), features 4, 16, 28, 32 and 33 are being used for further analysis. While the features provide good predictive ability for the models, the selection criterion for these features were not explained.

3. In section 3.3 *Distribution of attachment severity among subjects*, the authors employed a normalization, followed by squaring of the response features so that normal and "poor" attachment could be studied. While this enables studying the attachment on a normal vs rest attachment scale, it will be interesting to look at how different attachments: less and high are differentially influenced by maternal anxiety/depression.

4. In section 3.6 *Random Forest feature selection*, the authors performed a polynomial regression on the features used for Random Forest. It would be interesting to also look into how the usage of polynomial features in linear regression would affect the model performance.

Some additional minor points that would enhance the understandability of the study are as follows:

1. In section 2.4 *Maternal Measures*, CityBiTS and EPDS, the data is mentioned to be collected within the last month and the last week respectively. Additional information regarding the timeline of data collection would be extremely helpful in developing a better comprehension of the datasets.

2. With reference to *Figure 1*, while all the IBQ-R response features seem to be positively correlated, the presence of negative weight for the maternal features in logistic regression indicate that not all maternal features are positively correlated to one another. A plot depicting the maternal feature correlation would be helpful.

3. In section 3.6 *Random Forest feature selection* and *Figure 5*, the authors use the term impurity, which has not been defined previously in the study. It would be more reachable to a general audience if the term is defined.

4. While statistical tests on the hypothesis described in the study have been carried out in the code, it would be beneficial to briefly describe the tests carried out in the manuscript too.