Causal Explanation of Quality of Parent-Child Interactions with Multimodal Behavioral Features (Student Abstract)

Katherine M. Guerrerio[1], Lujie Karen Chen[2], Lisa Berlin[3], Brenda Jones Harden[4]
[1]Johns Hopkins University
[2]University of Maryland Baltimore County
[3]University of Maryland School of Social Work
[4]Columbia University's School of Social Work
kguerre6@jh.edu, lujiec@umbc.edu, LBERLIN@ssw.umaryland.edu, bjh2180@columbia.edu

Abstract

The quality of interactions between parents and children is a critical factor in child development. Recent years have seen programs to improve parenting behaviors through evidence-based approaches, such as attachment-based interventions. A vital element of these programs is to assess the quality of parenting behaviors via video recordings of parent-child interactions, which is often time-intensive. In our previous work, we explored machine learning models to predict expert ratings of parenting behaviors from video recordings of semi-structured parent-child play. However, the large set of low-level multimodal features struggled to provide explainable insights, which created barriers to communicating with domain experts and improving the models further. In this work, we developed a machine learning pipeline that combines sparse multiple canonical correlation analysis with causal discovery techniques to uncover explainable causal relationships between nine categories of behavioral features and the quality ratings of parent-child interactions. This approach offers valuable insights into the otherwise black-box models and contributes to the growing body of work on transparent and trustworthy machine learning models of parenting behaviors.

Introduction

High-quality parent-child interactions are vital for a child's development, and evidence-based programs, such as attachment-focused interventions, have been developed to improve parenting practices (Berlin 2018). A central aspect of these programs involves assessing parenting behaviors through video recordings of parent-child interactions; however, this process is time-consuming. Previous research trained machine learning models using audio and video features to predict expert ratings of three key parenting behaviors: sensitivity, intrusiveness, and positive regard (Jebeli et. al, 2024). However, these moderately accurate black-box models rely on numerous low-level features that offer limited transparency, making it challenging to communicate with domain experts for further model improvement. In this study, we expand our feature set to 157 multimodal features, including two new modalities (paralinguistic and language) and additional audio features, informed by social science literature. We introduce a machine learning pipeline that combines sparse multiple canonical correlation analysis with causal discovery to uncover explainable causal relationships between nine categories of multimodal behavioral features and interaction quality ratings.

Dataset and Multimodal Feature Sets

Our work used a dataset of pre-intervention at-home recordings of semi-structured play of 220 pairs of mothers and children (6-18 months), collected from a randomized control study to compare 2 attachment-based intervention programs (Berlin 2018). Each video was rated by experts on a scale of 1 to 5 on the quality of parenting behaviors concerning sensitivity, intrusiveness, and positive regard. We extracted 157 features across 4 modalities: video (6 features), audio (13 features), paralinguistic (12 features), and language (126 features). The 4 modalities are broken into 9 categories of behavior features to enhance explainability, as described in Figure 1.

Visual features capture the physical distance between mother and child as well as the dynamic movements of both (Jebeli, 2024). **Audio features** characterize conversational features such as frequency and duration of turns, as more frequent, shorter speech is associated with better language outcomes (Leech and Rowe 2021). **Paralinguistic features** focus on fundamental frequency (F0) as prototypical infant-directed speech prosody is associated with more infant attention and better pre-linguistic skills (Spinelli et al., 2017). **Language features** are derived from the Linguistic Inquiry and Word Count (LIWC) lexicon (Pennebaker, Booth, & Francis, 2007). Additional extracted features include sentence structures, such as wh-questions (Rowe et al., 2017), and contingent repetitions of the same structure or item (Goldstein et al., 2010), which have positive language associations. Using these features, we trained Random Forest s

models to predict three categories for the quality of parenting behaviors, which achieved a 6.59%, 0%, and 21.11% reduction in root mean squared error compared to baseline mean predictor in predicting expert rating in sensitivity, intrusiveness, and positive regard, respectively.

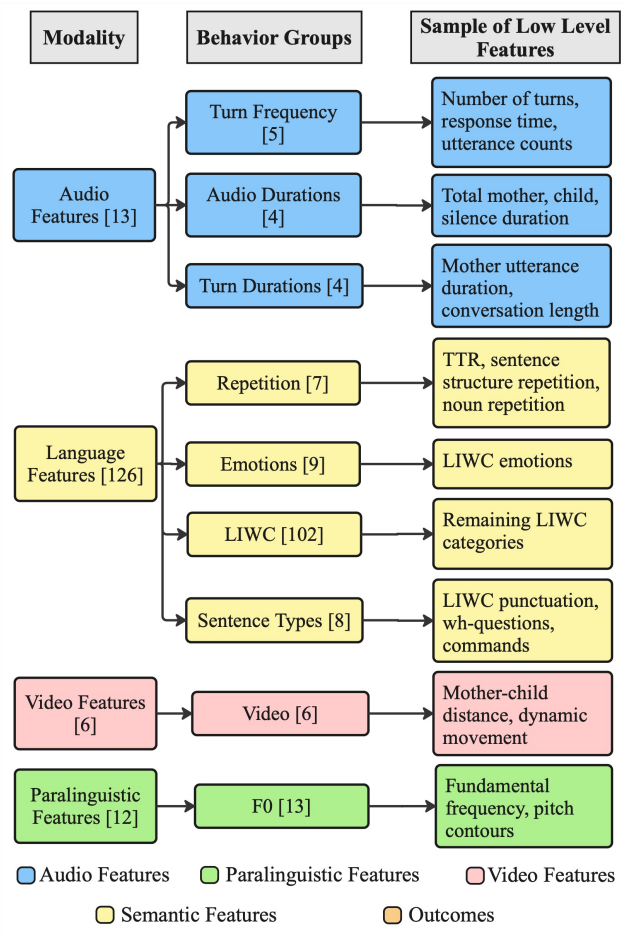

Figure 1: Description of the feature groups extracted in each modality. The number in brackets represents the number of low-level features in each group.

CCA and Causal Discovery Pipeline

To uncover explainable causal relationships, we explored a two-step modeling approach. First, we applied sparse multiple canonical correlation analysis (mCCA, Witten and Tibshirani, 2009) to the 157 low-level features, organized into 9 feature groups, to reduce dimensionality while maximizing linear correlations among the groups. This sparse correlational model generates a set of composite variables, where most coefficients in the linear combination of the underlying low-level features shrink to zero, thus enhancing the interpretability of the composite variables. These composite variables were then used to construct an Equivalence Class of Graphs (ECG), or the most probable causal graph pattern, using the TETRAD toolkit (Scheines et. al. 1998).

As illustrated in Figure 2, two main groups of features are causally linked: (1) audio, paralinguistic, and LIWC emotion features; and (2) video and the remaining language features. Within these clusters, only two features show direct causal links with any of the outcomes: while LIWC features are causally linked to intrusiveness and positive regard, Turn Frequency links to intrusiveness. Specifically, from the sparse mCCA models, we note that LIWC features are largely driven by word count, and turn frequency was dominated by the count of mother utterances.

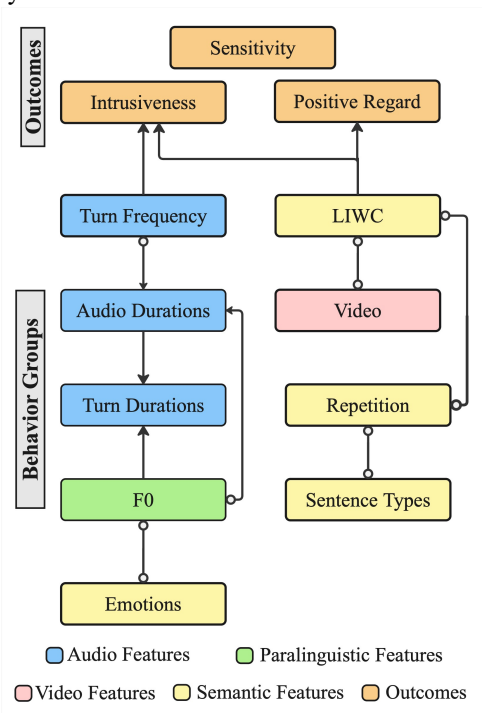

Figure 2: Equivalence Class Graph with 9 feature groups and 3 outcomes discovered by the GFCI search algorithm in TETRAD. Outcomes can only be caused by features, and we forbid causal relations between the outcomes.

Conclusions

Our preliminary results suggest that specific features of parent behavior during interactions with their child are causally linked to the quality rating of these interactions. For instance, language features such as the number of words mothers use and the frequency of their speech may influence the quality of parent-child interactions, particularly in terms of intrusiveness and positive regard. To better understand the directionality of these effects, further research is required to refine the model—such as by fitting linear Gaussian Structural Equation Models (SEMs) and evaluating their goodness-of-fit to assess plausibility. Compared to previous correlation-based modeling approaches, the demonstrated pipeline offers potential improvements in the interpretability and transparency of machine learning models for parenting behaviors.

Acknowledgments

This research was supported by Administration for Children and Families Grant 90-YR-0059, awarded to Lisa Berlin and Brenda Jones Harden. This content is solely the responsibility of the authors and does not represent the official views of the Administration for Children and Families. The authors have no conflicts of interest to disclose. We gratefully acknowledge the research participants for their many contributions.

References

Atefeh Jebeli, Lujie Karen Chen, Katherine Guerrerio, Sophia Papparotto, Lisa Berlin, and Brenda Jones Harden. 2024. Quantifying the Quality of Parent-Child Interaction Through Machine-Learning Based Audio and Video Analysis: Towards a Vision of AI-assisted Coaching Support for Social Workers. *ACM J. Comput. Sustain. Soc*. 2, 1, Article 6 March 2024, 21 pages. https://doi.org/10.1145/3617693

Berlin, L. J., Martoccio, T. L., & Jones Harden, B. 2018. Improving early head start's impacts on parenting through attachment-based intervention: A randomized controlled trial. *Developmental Psychology* 54(12): 2316–2327.

Goldstein, M. H., Waterfall, H. R., Lotem, A., Halpern, J. Y., Schwade, J. A., Onnis, L., & Edelman, S. 2010. General cognitive principles for learning structure in time and space. *Trends in Cognitive Sciences*, 14(6), 249–258. https://doi.org/10.1016/j.tics.2010.02.004

Pennebaker, J. W., Booth, R. J., & Francis, M. E. 2007. Linguistic Inquiry and Word Count: LIWC [Computer software]. Austin, TX: LIWC.net.

Rowe, M. L., Leech, K. A., & Cabrera, N. 2017. Going Beyond Input Quantity: Wh-Questions Matter for Toddlers' Language and Cognitive Development. *Cognitive Science*, 41(S1), 162–179. https://doi.org/10.1111/cogs.12349

Scheines, R., Spirtes, P., Glymour, C., Meek, C., & Richardson, T. 1998. The TETRAD project: Constraint-based aids to causal model specification. *Multivariate Behavioral Research, 33*(1), 65–117

Spinelli, M., Fasolo, M., & Mesman, J. 2017. Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, 44, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Witten, D. M. and Tibshirani, R. J. 2009. Extensions of sparse canonical correlation analysis with applications to genomic data. Statistical applications in genetics and molecular biology 8, 1, 1–27.