Understanding Unique Behavioral Patterns through Multimodal Analysis of Eye-Hand Coordination in Autistic Children (Student Abstract)

Emily Yu

Mendon High School, 472 Mendon Rd, Pittsford, NY 14534
emilyyu751@gmail.com

Abstract

Data-driven analysis has shown promising results in identifying subtle patterns in the behavior of individuals with Autism Spectrum Disorder (ASD) for diagnosis and intervention. However, most existing methods primarily focus on a single behavioral modality (e.g., eye movements) instead of capturing the intricate multimodal behavior of humans. We propose a multimodal approach that investigates the underlying connections between eye movements and hand motions through eye-to-hand prediction. To tackle the highly noisy and irregular behavioral data, we propose a novel approach that defines the prediction as a machine translation problem and leverages a sequence-to-sequence machine learning model for the prediction. An experimental study on a dataset collected from a VR system has demonstrated high prediction accuracy. The significant difference in the prediction accuracy between the autistic group and their typically developing (TD) peers serves as quantitative evidence to objectively understand the restricted and repetitive behaviors (RRBs) in autistic children. The source code can be accessed here: https://github.com/mathjams/AAAI_2024.

1 Introduction

Autism Spectrum Disorder (ASD) is a neurological developmental disorder that affects 1 in 36 children in the U.S. Autism impacts how the individual learns, interacts with others, communicates, and behaves. It is also characterized by repetitive thoughts and actions, and resistance to change (Thom and McDougle 2023). However, these behaviors are not captured in detail in traditional diagnosis methods, which usually rely on questionnaires filled out by the patient or their guardian. Subtle patterns in sensory behavior in certain scenarios may be undetected in existing screening tools, such as the SRS-2 (Bruni 2014), reducing opportunities for treatment and intervention. Scientific communities have started to leverage virtual reality (VR) technologies as a platform to collect fine-grained behavioral data from autistic children, providing quantitative behavioral measurements. However, most existing data-driven studies focus on analyzing a single behavioral modality. These studies fall short of modeling the rich human behavior that is inherently multimodal. Modeling the interaction between different behavioral modalities (e.g., eye movement and hand motion) poses several unique challenges. First, even though the use of VR systems can improve the quality of the collected data, human behavior is inherently noisy, and multimodal behavioral data collection further increases the chance of accumulating additional noise. Second, while behavioral data from different sensing modalities are collected simultaneously, the data are by no means perfectly aligned in a way that allows a model to easily detect their dependencies. Most observations from different modalities appear to be largely disjointed and highly irregular in their original forms, visualized in Figure 3 in Appendix B (Yu 2024).

To fully leverage the sparse behavioral data and obtain a more reliable modeling outcome, we draw an analogy from machine translation, where the semantic granularity of a term in the source language may not be perfectly aligned with another term in the target language. We propose to **perform behavioral data translation across two different sensing modalities: from eye movement to hand motion**. After exploring different sequential models, the result demonstrates that the multimodal translation framework can more accurately predict the next hand location as compared with other alternative designs, such as alignment methods. Differences between these models are elaborated in Appendix D (Yu 2024). More importantly, the modeling outcome reveals **a significant difference** between the hand location prediction accuracy between the TD (typical developing) and autistic groups. Such a result implies that the hand motion from an autistic individual can be more accurately predicted, which could be connected with restricted and repetitive behaviors (RRBs). We believe our study provides useful quantitative evidence to objectively understand the RRBs in autistic individuals.

2 Methodology

We explore the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)-based sequence-to- sequence (S2S) machine learning models (Sutskever 2014) to analyze the multimodal eye-hand fixation sequences, viewing eye-to-hand mapping as a translation problem. Our data was collected through the Multimodal Virtual Classroom Interface (MVCI) system (Yu et al. 2023), used to quantitatively analyze the behavior patterns in autistic children as shown in Figure 1. The screen was bounded by a 1×1 unit grid, so all coordinates have continuous values in $[0, 1]$. The collected

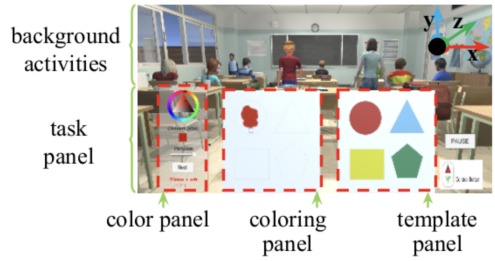

Figure 1: The Multimodal Virtual Classroom Interface (MVCI) with a coloring task as well as distractions of a real classroom setting

	GRU-P	GRU-PT	LSTM-P	LSTM-PT
MSE (Aut)	0.006265	0.006486	0.006351	0.006480
MSE (TD)	0.01113	0.01145	0.01204	0.01405
T-statistic	-34.48	-50.56	-59.96	-10.99
P-value	3.939e-16	2.592e-11	6.647e-12	4.168e-06

Table 1: Autistic-TD prediction errors by different models

data consists of sequences of eye and hand fixation records. A record, including a location (x, y) and a start and end time, refers to a concentrated gaze or movement at a specific spot, signifying attention and focus. Each sequence corresponds to the eye or hand fixations of a user during the entire gameplay. We propose two designs to approach this eye-to-hand translation problem. In the first design, the S2S model aims to predict the fixation positions in the hand movements using the positions in the eye sequence. To enrich the feature representation, we also incorporate the start and end times of the eye fixations. In the second design, we partition the game screen into multiple regions, including the main task (MT), facilitation (F), and reference (R), as shown in Figure 2 in Appendix A (Yu 2024). We ran the S2S GRU model using eye fixation positions that lie in a specific region to predict the hand fixation positions that also lie in the same region.

3 Experimental Results

Main results. We trained the models using 80% of the data and tested on the other 20%. Each experiment was run 3 times and we reported the test Mean Squared Error (MSE) on all the models, including GRU and LSTM with different input features with eye fixation location only (denoted by GRU-P and LSTM-P) and with both fixation location and start/end times (denoted by GRU-PT and LSTM-PT). The second design was run with a GRU model, using coordinates of eye fixations in specific regions to predict coordinates of hand fixations in specific regions. The results are shown in Table 1. We have two major observations. First, all models perform significantly better in predicting the hand fixation sequences of autistic individuals than for typically developing individuals, with extremely small P-values. Second, we see that the lowest errors of 0.006265 for the autistic predictions and 0.001113 for the TD predictions come from the GRU. Additionally, we found that the performance difference between models is significant. Most predictions are more accurate with the GRU model regardless of input features, with exact P-values in Appendix B (Yu 2024). Thus, when performing the second design, we use the GRU model in predictions.

	MT	F	R
MSE (Aut)	0.005027	0.003421	0.002823
MSE (TD)	0.01015	0.004344	0.001813
T-statistic	-12.33	-0.8421	14.51
P-value	0.0002484	0.44713	0.0001310

Table 2: Autistic-TD prediction errors in different regions

Table 2 shows the prediction results based on the regions defined above. The model still performs significantly better in the autistic group in the main task region. We see how the smallest difference in performance occurs in the facilitation area, while the biggest differences come from the main task and reference areas. This could be because these two activities reveal more about the differences in general behavior between autistic and typically developing individuals compared to the facilitation area. It is also interesting to see that the reference area displays an opposite trend than observed in other regions. This could imply that TD children exhibit more consistent behavior in the reference area.

Discussion. All models are better at capturing autistic behavior due to more accurate predictions, and this difference is most prevalent when performing activities in the main task region. These observations may be attributed to less complex and more predictable behavior in autistic individuals, providing quantitative evidence supporting the presence of RRBs, especially in eye-hand coordination heavy tasks. The autistic prediction errors by our model are notably small, being 0.006265, which corresponds to an average distance of 0.08 units from the actual fixation point. Appendix D (Yu 2024) provides additional results.

4 Conclusion

We conducted a systematic analysis of the multimodal behavioral data collected from a VR system recording the gameplay behavior of autistic individuals and those from their typically developing peers. Inspired by the machine translation techniques, we propose to perform translation from one behavioral modality (i.e., eye movement) to another (i.e., hand motion) allowing us to predict the hand fixation sequence. Across all methods, the models better capture the patterns in autistic behavior significantly more effectively, quantitatively measuring differences in behaviors between the groups. Additionally, the especially low errors in the autistic prediction models demonstrate that hand fixations can be accurately predicted, which can be useful in intervention and treatment, such as detecting when the student may be off task.

Acknowledgments

I would to thank Dr. Zhi Zheng from the University of Notre Dame, who served as my mentor during the course of working on this project. Dr. Zheng is an expert in human-computer interaction with years of experience in designing reliable assistive systems for mental health care. Dr. Zheng introduced to me this research area, suggested the overall research topic of using a data-driven approach to understand unique behaviors in children with autism, and shared the dataset that is analyzed in my study. We had weekly meetings to discuss the progress of my work and Dr. Zheng offered important feedback that helped to shape this research. I would also like to thank Zhiwei Yu, who helped preprocess the data by extracting the fixations from the raw data.

References

Bruni, T. P. 2014. Test Review: Social Responsiveness Scale–Second Edition (SRS-2). *Journal of Psychoeducational Assessment*, 32(4): 365–369.

Sutskever, I. 2014. Sequence to Sequence Learning with Neural Networks. *arXiv preprint arXiv:1409.3215*.

Thom, R. P.; and McDougle, C. J. 2023. Repetitive thoughts and behaviors in autism spectrum disorder: A symptom-based framework for novel therapeutics. *ACS Chemical Neuroscience*, 14(6): 1007–1016.

Yu, E. 2024. Appendix: Understanding Unique Behavioral Patterns through Multimodal Analysis of Eye-Hand Coordination in Autistic Children. https://github.com/mathjams/AAAI_2024. Accessed: 2024-11-16.

Yu, Z.; Iadarola, S.; Daley, S.; and Zheng, Z. 2023. A Multimodal Virtual Classroom Interface to Facilitate Discovery of Behavioral Patterns in Response to Sensory Stimuli. *International Society for Autism Research Annual Meeting 2023 (INSAR 2023)*.