Automated Behavior Labeling During Team-based Activities involving Neurodiverse and Neurotypical Partners using Multimodal Data*

Abigale Plunk[1[0000−0003−2373−9337]], Ashwaq Zaini Amat[1[0000−0002−9896−1285]], D. Mitchell Wilkes[1[0000−0002−9693−3539]], and Nilanjan Sarkar[1,2[0000−0002−3969−0593]]

[1] Dept. of Electrical Engineering and Computer Engineering, Vanderbilt University, Nashville, TN, USA
[2] Dept. of Mechanical Engineering, Vanderbilt University, Nashville, TN, USA
abigale.l.plunk@vanderbilt.edu

Abstract. The employment setting for autistic individuals in the USA is grim. Based on reports, individuals with ASD struggle to secure and retain employment due to challenges in communicating and collaborating with others in workplace settings which is often attributed to their social skills deficit. Current programs that support collaborative skills development in vocational settings rely on manual evaluation and feedback by human observers, which can be resource straining and receptive to bias. Using a collaborative virtual environment (CVE) allows neurodiverse individuals to develop teamwork skills by working together with a neurotypical partner in a shared virtual space. An effective CVE system can provide real-time prompts by recognizing the user's behavior to promote teamwork. As such, it is crucial to be able to automatically label both users' behaviors. In this paper, we propose using K-means clustering to automate behavior labeling in a workplace CVE. The results show that K-means clustering enables high accuracy in predicting the user's behavior, therefore, confirming that it can be used in future studies to support real-time prompts to encourage teamwork in a CVE.

Keywords: automated behavior labeling, collaborative virtual environment, clustering, autism spectrum disorder, teamwork training

1 Introduction

Teamwork, which includes skills such as conflict resolution, communication, collaboration, and positive interaction, is highly sought after by employers [49].

* We are grateful for the support provided by NSF grants 1936970 and 2033413 as well as NSF NRT grant DGE 19-22697 for this research. We would also like to thank the Vanderbilt Treatment and Research Institute for Autism Spectrum Disorders (TRIAD) team; Amy Weitlauf and Amy Swanson for their expert advice on interventions for autistic individuals and recruitments for the study. The authors are solely responsible for the contents and opinions expressed in this manuscript.

Teamwork can fulfill the personal need for social interaction and affiliation leading to increased satisfaction in the workplace and increased productivity for the company [36][54]. However, individuals with autism spectrum disorder[3] (ASD) experience core deficits in social interactions such as reduced eye contact, facial expressions, and body gestures that can hinder their ability to work on a team potentially contributing to unemployment [43] and anxiety. Compared to other individuals with disabilities, adults with ASD have the highest unemployment rate between 50 – 85% [27]. For those with employment, the majority are either underemployed or unable to retain their position due to their perceived deficits in social communication and interaction skills [54]. Studies have shown that unemployment can lead to reduced self-esteem and heightened distress, depression, and anxiety [34][19]. Therefore it is essential to address these deficits as they tend to cast a shadow on the outstanding qualities such as precise technical abilities, high tolerance for repetitive tasks, reliability, and increased concentration for long periods of time that autistic individuals can bring to a team[43][51]. Studies also show that teamwork gives individuals with ASD the opportunity to build upon their social communication skills [17], problem-solving skills [12], and self-confidence [53]. Although existing training and interventions have shown some improvements in teamwork skills in adolescents with ASD, simulating real-world teamwork scenarios can be tedious, resource-straining, and costly, thus limiting the accessibility and reach of the interventions[52]. Computer-based simulators using digital games have been shown to positively impact the training of these skills [32]. However, many digital games lack the structure to scaffold skill learning, do not provide real-time feedback or prompts that could facilitate skill learning, and have no objective means of measuring players' skills improvements. Using a Collaborative Virtual Environment (CVE) to practice, measure, and promote positive social communication skills could be advantageous in preparing autistic individuals for employment while also addressing the pitfalls of simulating real-world teamwork and digital games.

CVEs are virtual environments that allow multiple users to interact with each other and the environment itself in a shared virtual space. CVEs engage the users [16], provide a safe environment for training [40], and provide quantitative measures of the skills they are learning [58]. In addition, they are both reproducible and cost-effective. The CVE discussed in this paper simulates a workplace environment for two users, one with ASD and one who is neurotypical, to work together towards achieving a task that encourages teamwork and collaboration. The CVE is tasked with observing multimodal data (i.e., speech, eye gaze, and controller input), recognizing the behavior of each user, and prompting the system to provide reinforcement or assistance depending on each user's current behavior. However, manually labeling the current behavior of each user is labor-intensive, prone to bias, and inconsistent [24]. In addition, it does not allow for real-time feedback, which is necessary for promoting teamwork in the CVE. To support real-time feedback, an essential criterion involves reliable detection of

[3] We are using both identity-first and people-first language to respect both views by interchangeably using the term 'autistic individuals and 'individuals with ASD'. [31]

human behavior in collaborative interactions, which can be achieved through large amounts of labeled data. Therefore, automated labeling is needed for the success of using a CVE for teamwork training. However, the complexity of human behavior and contextual properties make it difficult to recognize human behavior even in constrained domains [47].

Previous solutions for automated labeling include various clustering methods, semi-supervised machine learning algorithms, and unsupervised machine learning algorithms. In this paper, we propose the use of K-means clustering due to its simplicity and efficiency [25] for automated behavior labeling in a CVE-based simulator of workplace scenarios allowing for real-time prompts that encourage collaboration and teamwork between neurodiverse and neurotypical partners. The following section discusses related works that utilize multimodal data in CVEs and different methods used for automated behavior labeling in various applications. Section 3 briefly discusses the experimental design, including the collaborative tasks we employ and the multimodal data captured in our CVE used to represent teamwork. In Section 4, we describe the methodology of applying and verifying K-means clustering to automate behavior detection of human behavior followed by an analysis of the results in Section 5. Finally, we conclude the paper with a discussion that summarizes our contributions and provides insight for future works.

2 Related Work

Over the last decade, the use of human-computer interaction (HCI) technology has shown promising benefits that can potentially complement conventional ASD interventions by providing engaging interactions and replicable solutions that can minimize costs and provide relatively broader access to users [44]. Additionally, autistic individuals have a natural affinity for technology-based interactions and prefer the consistency that computer-based interactions can offer [45]. Specifically, there have been a number of research that employ VR-based systems intervention tools that are focused on teaching both social skills and technical skills, which include skills such as cooking [6], road safety [46], driving [13], joint attention [60], and emotion recognition [11]. Nonetheless, VR-based systems are limited to single user interaction and are unable to support more natural complex back and forth human-human interactions. Additionally, individuals with ASD might be more comfortable interacting with a virtual avatar compared to a human partner, thus making it less efficient for generalization to the real-world [56, **?**,**?**]. Alternatively, CVEs enable users to communicate with each other naturally while performing a task together in the shared virtual environment, in turn minimizing the effect of attachment to virtual avatars. CVE-based systems have been primarily studied to understand the impact of collaborative learning and various aspects of social behavior involved in collaboration for autistic individuals [15][8].

Vocational and technical skills are important aspects of employment and are the main criteria considered for employment [57]. However, interpersonal or professional skills such as teamwork are the core skills needed to secure and retain employment [5]. Recently, the importance of teamwork is reflected in the hiring process of companies like Microsoft and Specialsterne, which employ autistic individuals. They utilize a group assessment process for autistic candidates in place of the conventional interview process. A Lego Mindstorm group project [2] and Minecraft collaborative tasks [1] were administered to them to assess teamwork skills. Currently, most studies that investigate social skills evaluation rely on qualitative measures of performance and self-reporting questionnaires, which are subjective and prone to bias [9]. Teamwork is a complex social behavior that is not easily assessed since it involves detecting and understanding dynamic social manifestations between individuals. Moreover, some individuals with ASD may present subtle or low manifestations of specific social behaviors due to a deficit in their social reciprocity, making it difficult for their partners or observers to recognize their social cues [14]. Thus, there are potential benefits of capturing objective interaction data from multiple modalities to evaluate these complex skills.

Recently, there has been a growing interest in multimodal data analysis within HCI that uses measurable parameters to assess teamwork objectively and represent the important features of teamwork [37][42]. Although many studies capture multimodal data in collaborative interactions, there are currently no standardized methods to measure teamwork and collaboration skills in group interactions. A few studies have explored different ways to reliably represent interpersonal behavior using multimodal data in group interactions [23][41][26]. In one study, the researchers analyzed multimodal data such as physical locations, speech, movements, and physiological measures to represent different aspects of interpersonal behaviors [23]. Meanwhile, Okada et al. used verbal and non-verbal measures to assess a group's communication skills based on the different types of discussions taking place [41]. In the study, the researchers extracted communication features based on data from speech and head movement information. They compared the analysis against human-coded evaluations of communication skills and found that certain quantitative measures can be analyzed to represent more than one feature. For example, speech data can be used to represent both verbal and non-verbal features in collaborative interactions, while dialogue content can provide social communication features (e.g., intention) and task performance features (e.g., topic/object). In a more recent study, Hayashi conducted a collaborative learning study to systematically evaluate students' learning behavior in a jigsaw-type collaborative task [26]. The researcher used facial expressions together with speech to predict the emotional state of the students and how these emotions influenced their collaborative learning process. The results of these studies can benefit the development of a feedback mechanism in collaborative interaction by generating a reliable evaluation of collaborative behavior. However, they rely heavily on manually labeled data to generate reliable human behavioral models.

Motivated by the limitations of manual data labeling [47], several recent studies have investigated the use of clustering methods and machine learning algorithms to automate behavioral labeling [4][28][21][48][29][59][39][38]. Current state-of-the-art automated labeling techniques are applied to a wide range of applications such as visual detection [4], human-robot social interaction [28], human actions in video [21], and social signal processing [48]. Hong et al. used a multimodal wearable sensing platform to collect activity data [29]. They then developed a semi-population model that automatically labeled new data based on K-means clustering of selected features using previously collected data to recognize seven different human activities. The semi-population model achieved better accuracy compared to individual and group data. In another study, researchers integrated a K-means clustering method with a decision tree to create an initial transfer learning model [59]. The model was adaptively trained whenever new target data was available. This method managed to reduce complexity and minimize computing power. Furthermore, recent studies have explored the implementation of automated behavior labeling using multimodal data as ground truth in a closed-loop feedback mechanism [39][38]. In one study, the researchers designed an intelligent mediator that provides dynamic feedback to balance the interactions between participants based on automated labeling of participants' speech [39]. Another study designed a virtual trainer that provided corrective feedback to public speakers based on automated social behavior labeling of speech and body movement of the participant [38].

Motivated by the potential of using various clustering methods and machine learning algorithms to automate human behavior labeling, we believe that the use of K-means clustering can complement manual data labeling without compromising the accuracy of the labeled behavior to evaluate teamwork skills between two individuals interacting together. Additionally, automated labeling can speed up ground truth labeling [55], thus enabling researchers to train a behavior detection model that can be used to provide reliable feedback based on the detected behavior. In the following sections, we will discuss using K-means clustering to automate grouping mutlimodal data into three behaviors. This will be validated using leave-one-out cross-validation with hand-labeled data. The analysis of our automated labeling will be used to develop a feedback mechanism in our CVE that can enhance collaborative interactions, however, it is not within the scope of the current paper.

3 Experimental Design

We conducted a system validation study with 4 pairs of Neurodiverse-Neurotypical participants (N = 8). The study required the participants to work in pairs to complete three collaborative tasks designed to encourage teamwork and collaboration skills. This study was approved by the Institutional Review Board at Vanderbilt University.

3.1 Collaborative Tasks

The collaborative tasks used in this work were presented in our previous work that discussed the design and development of three collaborative tasks to encourage teamwork and collaboration in a workplace setting: a) a PC assembly task, b) a fulfillment center task, and c) a furniture assembly task [7]. These tasks were designed based on stakeholders' inputs and involved one autistic and one neurotypical participant as partners. Although the tasks spread across various domains, the basic elements of teamwork and collaboration were maintained as we incorporated the same collaboration principle across all tasks defined in Meier et al. [20]. The researchers identified 9 features that are related to collaboration which include mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, reciprocal interaction, technical coordination, and individual task orientation [20]. We designed these tasks in a collaborative virtual environment (CVE) in Unity3D [3], which was connected to various peripheral devices to allow participants to communicate and interact in a shared virtual space with their partners. The system architecture of the workplace CVE was discussed in detail in [7] and is thus omitted here. Figure 1 shows the experiment setup and shows snapshots of all three tasks.

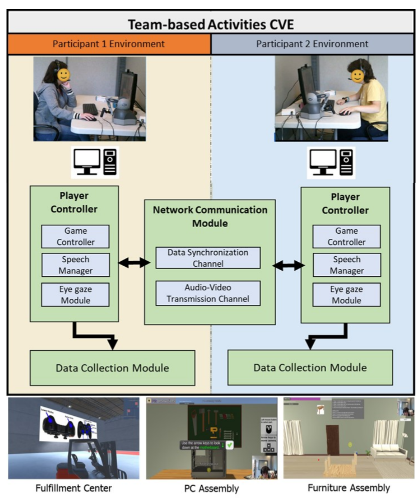

Fig. 1. System setup and snapshots of all three collaborative tasks

3.2 Participants

We recruited 8 individuals (ages: $16-30$ years old; mean age: 20.125 years old) to participate in the study. Participants with ASD were recruited from an existing research inventory of individuals previously diagnosed with ASD by licensed clinical psychologists using standard diagnostic tools, such as the Autism Diagnostic Observation Schedule-Second Edition (ADOS-2)[35]. As for the neurotypical participants, they were recruited from the local community through regional advertisement. We also evaluated the current level of ASD symptoms of all participants using the Social Responsiveness Scale, Second Edition (SRS-2) [18]. Table 1 lists the participants' characteristics.

Table 1. Participant Metrics

Metrics	**ASD Participants (N=4)**	**Neurotypical Participants (N=4)**
Age Mean (SD)	20 (3.82)	20.25 (5.06)
Gender (% male)	25%	25%
SRS-2 Total Score Mean (SD)	89.73 (23.43)	21.50 (11.90)
SRS-2 T-Score Mean (SD)	70.75 (11.09)	44.50 (3.42)

3.3 Protocol

The CVE system was set up such that each participant sat in a separate experiment room. Since the rooms were inside the same building, all the network connection was running using Vanderbilt University's local connection without any concern for privacy and security.

When participants arrived at the session, they were briefed and given consent forms to sign. Once all the forms were signed, participants were directed to the different rooms. Before starting the experiment, the eye tracker for each participant was calibrated and they both logged on to the shared virtual environment. The experiment lasted approximately 90 minutes, including briefing, consent form signing, and system set up.

4 Methods

The system consists of three modes of data capture between two participants. Figure 2 summarizes the workflow for multimodal data analysis that includes data collection, pre-processing, K-means clustering, and validation of the participant behavior. The following subsections will further explain the steps of this process.

4.1 Multimodal Data Capture and Decomposition

We used three modes of data capture for each participant using a game controller, a headset, and an eye tracker. We extracted seven binary features portraying aspects of collaboration and attention from these sources and performed classification of human behavior.

As seen in the data decomposition section of the workflow, we decomposed the game controller data into four binary features, the speech data into one binary feature, and the eye gaze data into two binary features. The four features extracted from the game controller were object manipulation, controller activity,

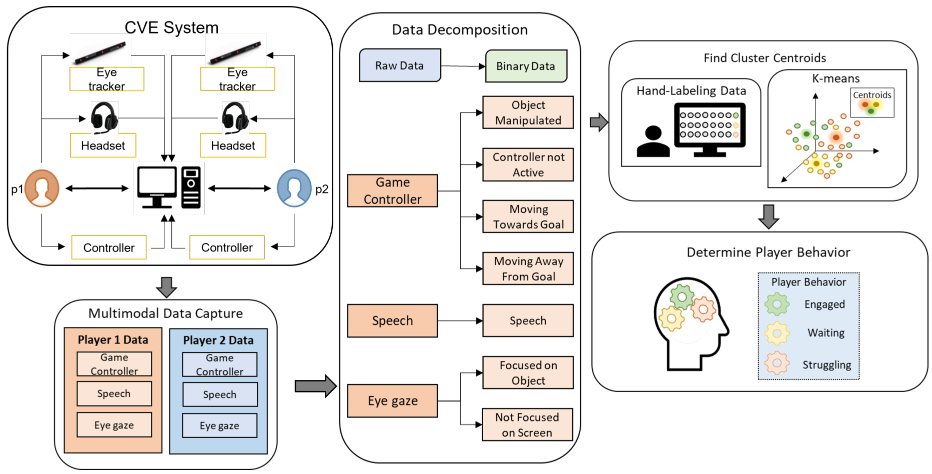

Fig. 2. Workflow for automated labeling

movement towards the goal, and movement away from the goal. The object manipulation feature captures the range of interest (ROI) of the participant. If the controller is activated in an area of interest (i.e., on a table component in the furniture task) this feature is recorded as '1' otherwise it is recorded as '0'. The controller activity feature is mapped to a '1' if there is no controller input and a '0' if there is controller input. Next, the position of the controller on the screen was extracted to determine in the participant was moving towards the goal or away from the goal. If they were making progress towards achieving the goal, 'moving towards goal' was set to '1'. Alternatively, if they were progressing further from the goal or not progressing at all, 'moving towards goal' was set to '0'. The inverse rules were used to extract the feature 'moving away from goal'. Next, speech data was collected from the headsets. If the participant was speaking, the new binary speech feature was set to '1', otherwise, it was set to '0'. Finally, gaze data was decomposed into two binary features, which can be visualized in Figure 3: 'focused on object' and 'not focused on screen'. The feature 'focused on object' is set to '1' if the user's scaled gaze fell within the middle 16% of the screen, represented by the green rectangle, as this is the portion of the screen where the collaborative tasks take place. Otherwise, it is set to '0'. The final feature 'not focused on screen' is set to '1' if the user's scaled gaze was negative or fell within the red section of the screen as shown in Figure 3. After extracting the seven binary features, they were concatenated to form a binary feature vector. All of the features were collected continuously in time with a sampling rate of 1 sample per second.

Using the decomposed multimodal data consisting of the seven binary features established above, K-means clustering was chosen as it is widely used in a variety of domains to group data into clusters based on the similarity of features[33]. Using K-means, we identified the centroids that were used to classify

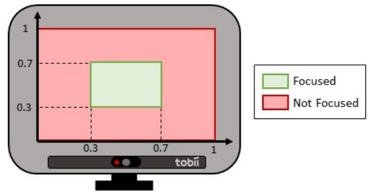

Fig. 3. Definition of focus area for eye gaze

each observation as one of three behaviors: engaged, struggling, or waiting. These three behaviors were chosen as they encapsulate various stages of teamwork allowing the system to provide appropriate feedback. Engaged captures how well the user is involved in the task itself and interacting with their partner allowing the system to provide positive reinforcement [22]. Next, the system needs to identify when the user is either not interacting with the system, not advancing towards the goal, or not engaged with their partner, which is captured by the struggling behavior implying that the system should prompt the users to work together[50]. The final aspect of teamwork and collaboration is taking turns. The waiting behavior captures when a user is waiting for their partner to complete their task[10][28]. The centroids used to classify these behaviors need to be validated as a consistent and accurate form of automated labeling. In the following sub-section, we will discuss the method of finding centroids using leave-one-out cross-validation (LOOCV) with hand-labeled data and K-means clustering. In addition, it will be established how to use the centroids to determine the player's behavior.

4.2 Manual Labeling

As discussed earlier, manual labeling of data is labor-intensive, prone to bias, and could be inconsistent. In addition, it does not allow for real-time feedback. However, to verify the proposed method for automated labeling, it was necessary to hand-label all four sessions of data. In future studies, hand-labeled data will be reduced to only a small subset used for training. To ensure the consistency of hand labeling, a set of coding rules was established beforehand. Bias was mitigated by having two individuals label the four sessions of data separately using the established rules achieving 98% agreement. Of the four labeled sessions the class distributions of the three behaviors were as follows: engaged - 19.9%, waiting - 52.1%, and struggling - 28.0%.

4.3 Determining Player States using K-means Centroids

Now that ground-truth labels were set, K-means clustering was applied to the multimodal data. Three clusters were chosen to represent the three behavior: engaged, struggling, or waiting.

We used MATLAB[30] to calculate and analyze the data. The goal of K-means clustering is for each observation to be assigned to the cluster that minimizes the Euclidean distance between the observation and the cluster's centroid. Using an iterative process, the three centroids for each of the four sessions were optimized to have the highest accuracy possible between both the K-means predicted and the ground-truth labels. Once the centroids were optimized, LOOCV was used to verify that they could consistently and accurately be used to predict each player's behavior.

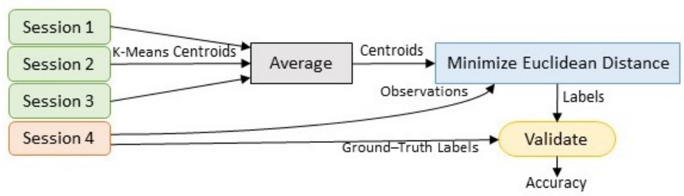

Fig. 4. One Iteration of LOOCV

A flowchart detailing one iteration of LOOCV is shown in Figure 4. Three of the four hand-labeled sessions were used to generate centroids using K-means, and the final session was used to validate the centroids. After finding the centroids that yielded the highest accuracy using K-means, the centroids from the three sessions were averaged together to create a centroid that is generalized across the training sessions. The test session's labels were computed by minimizing the Euclidean distance between the multimodal data and the optimized centroid. The pseudocode for this calculation is provided in Algorithm 1.

Algorithm 1 Label by Minimizing Euclidean Distance

states = $\{engaged, waiting, struggle\}$
n = # of samples, m = # of features
for i in n **do**
  for $state$ in $states$ **do**

$$error(i, state) = \sqrt{\sum_{j=1}^{m} [centroid(state, j) - obs(i, j)]^2}$$

  end for
  label(i) = argmin($error$)
end for

The predicted labels found were then compared with the ground-truth labels for validation. This process was iterated until each session has been used as the validation session with the other sessions used for training. In the next section, the results using this method, an individual model, and a group model are discussed.

5 Results

This section describes the results obtained using our proposed method for automated behavior detection. First, we present the accuracy obtained using LOOCV. This method, which is detailed in the above section, validates the generalizability of centroids found using K-Means on unseen data. The following two subsections will present the results of using an individual model and a group model. These two models are exploratory and do not validate the use of clustering on unseen data. The accuracy presented in the following subsections were calculated with both the typically developing participant and the autistic participant. In addition, the accuracy for each group was calculated separately but showed negligible differences, therefore they are omitted from this discussion.

5.1 LOOCV

To confirm the effectiveness of the proposed method, we analyzed the results using LOOCV. As discussed in the previous section, for this method, the labels of one session were determined using the centroids from the other three sessions. For example, the centroids used to generate labels from Session 1 were the average of the centroids from Sessions 2, 3, and 4. The accuracy for all three behaviors as well as the total accuracy is recorded in Table 2

5.2 Individual Model

In the individual model, the centroids were optimized using K-means clustering iteratively based on the individual session. For this model, we used 80% of the data as a training set to optimize the centroids and the remaining 20% as the test set. This model represents the best case scenario where each session has its own specialized model. This would be ideal if the neurodiverse and neurotypical pairs were to remain constant across multiple sessions. We recorded the average test accuracy across all four sessions using their respective centroids in Table 2.

5.3 Group Model

Finally in the group model, the individualized session centroids that were optimized in the previous subsection were averaged to form one generalized centroid. Using the generalized centroid, we then found the labels that were used to calculate the accuracy on the held out test set for each session. The average accuracy across all four sessions is shown in Table 2.

Table 2. Results across 3 models

State	Accuracy		
	LOOCV	***Individual***	***Group***
Engaged	93.65%	97.70%	100%
Struggle	82.90%	91.23%	83.85%
Waiting	91.10%	90.45%	89.65%
Total	87.38%	91.30%	89.66%

6 Conclusion

Teamwork is a complex social behavior as it involves verbal communication, non-verbal communication, and physical cooperation and coordination between two or more people interacting simultaneously with each other. In this paper, we discussed the use of K-means to complement manual labeling of human behavior based on multimodal data captured in a CVE-based system between individuals with ASD and their neurotypical partners. Although manual labeling can be a reliable source, it can be time-consuming, resource-straining, and susceptible to bias. As such, there is a need for an alternative method to automatically analyze the interpersonal social behavior of the users in team-based tasks. We proposed the use of K-means to complement hand labeling of multimodal data to reduce cost (i.e., resources and time) and minimize bias. The labeled data can then be used to develop and train a prediction model that can reliably predict individual and interpersonal human behavior in real-time. Moving forward, we would like to develop a closed-loop feedback mechanism that can facilitate collaborative interactions between two users by providing a reliable evaluation of the user behavior.

The objective of this study was to analyze collaborative multimodal data using K-means to label users' interpersonal behavior and validate the use of clustering against hand-labeled data. In the previous sections, we presented the procedures we have taken to achieve this and showed that K-means clustering can be used to successfully cluster human behavior. The use of an individual model to cluster user behaviors resulted in an accuracy of 91.30%. However, while this model would be ideal if neurodiverse and neurotypical pairs remained constant across sessions, it is typically not realistic due to the diversity of human behavior. Therefore, we use it to establish a baseline for the other models. In future studies, we plan to use the group model which finds the participant's behavior using the average of the centroids found using the hand-labeled sessions. This model resulted in an accuracy of 89.66% which is very comparable to the individual model. However, this does not verify that the centroids are generalizable to new data. To combat this, we used LOOCV to show that the model did generalize well on unseen data. Using this approach, we achieved an accuracy of 87.38% which, as expected, is slighter lower than both the individual and group models.

While the results discussed above show promise, it is important to highlight the limitations of the study and important targets for future research. First, the sample size was relatively small. Recruiting more participants would allow us to improve the robustness of K-means clustering. However, these preliminary results further support our motivation for the use of K-means to complement manual data labeling. Next, from the analysis, we found that the centroids calculated using K-means clustering were not consistent and can cause misclassification of new data. As such it is important to be able to verify the centroids against a small portion of labeled data before using the centroid to automatically label new data. Although the use of K-means to automatically label human behavior through clustering is not perfect, it serves as one way to further explore the possibility of automating human behavior labeling and has extensive room for further improvements. To our knowledge, this is the first study that systematically evaluates and validates interpersonal behavior using K-means clustering with multimodal data of autistic individuals working together with neurotypical partners. Based on the promising results, future work would include using the labeled data to train machine learning models such as a Markov chain or neural networks to predict user behavior in real-time with a closed-loop feedback mechanism that can enhance user experience and facilitate the development of their teamwork skills.

References

1. Neurodiversity hiring: Global diversity and inclusion at microsoft, https://www.microsoft.com/en-us/diversity/inside-microsoft/cross-disability/neurodiversityhiring
2. Specialsterne: Assessment, https://www.specialisterneni.com/about-us/assessment/
3. Unity website, https://unity3d.com/unity
4. Afsar, P., Cortez, P., Santos, H.: Automatic visual detection of human behavior: A review from 2000 to 2014. Expert Systems with Applications **42**(20), 6935–6956 (2015). https://doi.org/https://doi.org/10.1016/j.eswa.2015.05.023, https://www.sciencedirect.com/science/article/pii/S0957417415003516
5. Agran, M., Hughes, C., Thoma, C., Scott, L.: Employment social skills: What skills are really valued? Career Development and Transition for Exceptional Individuals **39** (08 2014). https://doi.org/10.1177/2165143414546741
6. Almaguer, E., Yasmin, S.: . https://doi.org/10.1007/978-3-030-30712-7_18
7. Amat, A., Breen, M., Hunt, S., Wilson, D., Khaliq, Y., Byrnes, N., Cox, D., Czarnecki, S., Justice, C., Kennedy, D., Lotivio, T., McGee, H., Reckers, D., Wade, J., Sarkar, M., Sarkar, N.: Collaborative Virtual Environment to Encourage Teamwork in Autistic Adults in Workplace Settings, pp. 339–348 (07 2021). https://doi.org/10.1007/978-3-030-78092-0_22
8. Apostolellis, P., Bowman, D.: C-olive: Group co-located interaction in ves for contextual learning. pp. 129–130 (03 2014). https://doi.org/10.1109/VR.2014.6802085
9. Ashton, M.C., Lee, K.: The hexaco–60: A short measure of the major dimensions of personality. Journal of Personality Assessment **91**(4), 340–345 (2009). https://doi.org/10.1080/00223890902935878
10. Basden, B., Basden, D., Bryner, S., Thomas, R.r.: Developing methods for understanding social behavior in a 3d virtual learning environment. J Exp Psychol Learn Mem Cogn **23**(5), 1176–91 (09 1997). https://doi.org/10.1037//0278-7393.23.5.1176

11. Bekele, E., Zheng, Z., Swanson, A.R., Crittendon, J.A., Warren, Z., Sarkar, N.: Understanding how adolescents with autism respond to facial expressions in virtual reality environments. IEEE Transactions on Visualization and Computer Graphics **19**, 711–720 (2013)
12. Bernard-Opitz, V., Sriram, N., Nakhoda, S.: Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. Journal of autism and developmental disorders **31**, 377–84 (09 2001). https://doi.org/10.1023/A:1010660502130
13. Bian, D., Wade, J., Zhang, L., Bekele, E., Swanson, A., Crittendon, J., Sarkar, M., Warren, Z., Sarkar, N.: A novel virtual reality driving environment for autism intervention. pp. 474–483 (07 2013). https://doi.org/10.1007/978-3-642-39191-0_52
14. Bird, G., Cook, R.: Mixed emotions: The contribution of alexithymia to the emotional symptoms of autism. Translational psychiatry **3**, e285 (07 2013). https://doi.org/10.1038/tp.2013.61
15. Breland, J., Shiratuddin, M.F.: A study on collaborative design in a virtual environment. International Journal of Learning **16**, 385–398 (01 2009). https://doi.org/10.18848/1447-9494/CGP/v16i03/46179
16. Checa, D., Bustillo, A.: A review of immersive virtual reality serious games to enhance learning and training. Multimedia Tools and Applications **79** (03 2020). https://doi.org/10.1007/s11042-019-08348-9
17. Chen, W.: Multitouch tabletop technology for people with autism spectrum disorder: A review of the literature. Procedia Computer Science **14**, 198–207 (12 2012). https://doi.org/10.1016/j.procs.2012.10.023
18. Constantino, J., Gruber, C.: Social responsiveness scale (srs). Western Psychological Services Los Angeles, CA
19. Creed, P., Macintyre, S.: The relative effects of deprivation of the latent and manifest benefits of employment on the well-being of unemployed people. Journal of occupational health psychology **6**, 324–31 (11 2001). https://doi.org/10.1037/1076-8998.6.4.324
20. Deiglmayr, A., Spada, H., Rummel, N.: A rating scheme for assessing the quality of computer-supported collaboration processes. international journal of computer-supported collaborative learning, 2(1), 63-86. I. J. Computer-Supported Collaborative Learning **2**, 63–86 (04 2007). https://doi.org/10.1007/s11412-006-9005-x
21. Duchenne, O., Laptev, I., Sivic, J., Bach, F., Ponce, J.: Automatic annotation of human actions in video. pp. 1491 – 1498 (11 2009). https://doi.org/10.1109/ICCV.2009.5459279
22. D'Mello, S., Olney, A., Person, N.: Mining collaborative patterns in tutorial dialogues. Journal of Educational Data Mining **2** (01 2010)
23. Echeverria, V., Martinez-Maldonado, R., Buckingham Shum, S.: Towards collaboration translucence: Giving meaning to multimodal group data (02 2019). https://doi.org/10.1145/3290605.3300269
24. Fredriksson, T., Issa Mattos, D., Bosch, J., Olsson, H.: Data Labeling: An Empirical Investigation into Industrial Challenges and Mitigation Strategies, pp. 202–216 (11 2020). https://doi.org/10.1007/978-3-030-64148-1_13
25. Hanmin, Y., Hao, L., Qianting, S.: An improved semi-supervised k-means clustering algorithm. pp. 41–44 (05 2016). https://doi.org/10.1109/ITNEC.2016.7560315
26. Hayashi, Y.: Detecting Collaborative Learning Through Emotions: An Investigation Using Facial Expression Recognition, pp. 89–98 (05 2019). https://doi.org/10.1007/978-3-030-22244-4_12
27. Hendricks, D.: A short review on the current understanding of autism spectrum disorders. Journal of vocational rehabilitation **32**(2), 125–134 (2010)

28. Hoffman, G., Breazeal, C.: Collaboration in human-robot teams. Collection of Technical Papers - AIAA 1st Intelligent Systems Technical Conference **2** (09 2004). https://doi.org/10.2514/6.2004-6434
29. Hong, J.H., Ramos Rojas, J., Dey, A.: Toward personalized activity recognition systems with a semipopulation approach. IEEE Transactions on Human-Machine Systems pp. 1–12 (11 2015). https://doi.org/10.1109/THMS.2015.2489688
30. Inc., T.M.: Matlab (2018)
31. Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., Pellicano, E.: Which terms should be used to describe autism? perspectives from the uk autism community. **20**(4), 442–62 (May 2016). https://doi.org/10.1177/1362361315588200
32. Kulman, R., Slobuski, T., Seitsinger, R.: Teaching 21st Century, Executive-Functioning, and Creativity Skills with Popular Video Games and Apps, p. 159–174. ETC Press, Pittsburgh, PA, USA (2014). https://doi.org/10.5555/2811147.2811157
33. Kusumaningrum, R., Farikhin, F.: An automatic labeling of k-means clusters based on chi-square value. Journal of Physics: Conference Series **801**, 012071 (01 2017). https://doi.org/10.1088/1742-6596/801/1/012071
34. Litchfield, P., Cooper, C., Hancock, C., Watt, P.: Work and wellbeing in the 21st century †. International Journal of Environmental Research and Public Health **13**, 1065 (10 2016). https://doi.org/10.3390/ijerph13111065
35. Lord, C., Risi, S., Lambrecht, L., Cook, E., Leventhal, B., DiLavore, P., Pickles, A., Rutter, M.: The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders **30**, 205–223 (06 2000). https://doi.org/10.1023/A:1005592401947
36. M. Khawam, A., DiDona, T., S. Hern, B.: Effectiveness of teamwork in the workplace. International Journal of Sciences: Basic and Applied Research (IJSBAR) **32**(3), 267–286 (Apr 2017), https://gssrr.org/index.php/JournalOfBasicAndApplied/article/view/7134
37. Martinez-Maldonado, R., Gasevic, D., Echeverria, V., Nieto, G., Swiecki, Z., Buckingham Shum, S.: What do you mean by collaboration analytics? a conceptual model. Journal of Learning Analytics **8**, 126–153 (04 2021). https://doi.org/10.18608/jla.2021.7227
38. Mihoub, A., Lefebvre, G.: Wearables and social signal processing for smarter public presentations. The ACM Transactions on Interactive Intelligent Systems **9** (09 2018). https://doi.org/10.1145/3234507
39. Müller, P., Dietz, M., Schiller, D., Thomas, D., Zhang, G., Gebhard, P., Andre, E., Bulling, A.: Multimediate: Multi-modal group behaviour analysis for artificial mediation. pp. 4878–4882 (10 2021). https://doi.org/10.1145/3474085.3479219
40. Norris, M.W., Spicer, K., Byrd, T.: Virtual Reality: The New Pathway for Effective Safety Training. Professional Safety **64**(06), 36–39 (06 2019)
41. Okada, S., Ohtake, Y., Nakano, Y., Hayashi, Y., Huang, H.H., Takase, Y., Nitta, K.: Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets. pp. 169–176 (10 2016). https://doi.org/10.1145/2993148.2993154
42. Palliya Guruge, C., Oviatt, S., Haghighi, P., Pritchard, E.: Advances in multimodal behavioral analytics for early dementia diagnosis: A review. pp. 328–340 (10 2021). https://doi.org/10.1145/3462244.3479933
43. Park, H.R., Lee, J., Moon, H., Lee, D., Kim, B., Kim, J., Kim, D., Paek, S.: A short review on the current understanding of autism spectrum disorders. Experimental Neurobiology **25**, 1 (02 2016). https://doi.org/10.5607/en.2016.25.1.1

44. Parsons, S., Mitchell, P.: The potential of virtual reality in social skills training for people with autistic spectrum disorders. Journal of intellectual disability research : JIDR **46**, 430–43 (07 2002). https://doi.org/10.1046/j.1365-2788.2002.00425.x
45. Putnam, C., Chong, L.: Software and technologies designed for people with autism: What do users want? abstract. pp. 3–10 (01 2008). https://doi.org/10.1145/1414471.1414475
46. Saiano, M., Pellegrino, L., Casadio, M., Summa, S., Garbarino, E., Rossi, V., Dall'Agata, D., Sanguineti, V.: Natural interfaces and virtual environments for the acquisition of street crossing and path following skills in adults with autism spectrum disorders: a feasibility study. J Neuroeng Rehabil pp. 12–17 (02 2015). https://doi.org/10.1186/s12984-015-0010-z
47. Salah, A., Gevers, T., Sebe, N., Vinciarelli, A.: Challenges of human behavior understanding. vol. 6219, pp. 1–12 (12 2010). https://doi.org/10.1007/978-3-642-14715-9_1
48. Salah, A., Pantic, M., Vinciarelli, A.: Recent developments in social signal processing. pp. 380–385 (10 2011). https://doi.org/10.1109/ICSMC.2011.6083695
49. Sanyal, S., Hisam, M.: The impact of teamwork on work performance of employees: A study of faculty members in dhofar university **20** (03 2018). https://doi.org/10.9790/487X-2003011522
50. Schmidt, M., Laffey, J., Schmidt, C., Wang, X., Stichter, J.: Developing methods for understanding social behavior in a 3d virtual learning environment. Computers in Human Behavior **28**, 405–413 (03 2012). https://doi.org/10.1016/j.chb.2011.10.011
51. Scott, M., Falkmer, M., Girlder, S., Falkmer, T.: Viewpoints on factors for successful employment for adults with autism spectrum disorder. PLoS One **6** (10 2015). https://doi.org/10.1371/journal.pone.0139281
52. Stauch, T., Plavnick, J.: Teaching vocational and social skills to adolescents with autism using video modeling. Education and Treatment of Children **43**, 137–151 (06 2020). https://doi.org/10.1007/s43494-020-00020-4
53. Sung, C., Connor, A., Chen, J., Lin, C.C., Kuo, H.J., Chun, J.: Development, feasibility, and preliminary efficacy of an employment-related social skills intervention for young adults with high-functioning autism. Autism **23**, 136236131880134 (12 2018). https://doi.org/10.1177/1362361318801345
54. Taylor, J., Seltzer, M.: Employment and post-secondary educational activities for young adults with autism spectrum disorders during the transition to adulthood **41**(5), 566–74 (May 2011). https://doi.org/10.1007/s10803-010-1070-3
55. Vajda, S., Rangoni, Y., Cecotti, H.: Semi-automatic ground truth generation using unsupervised clustering and limited manual labeling: Application to handwritten character recognition. Pattern Recognition Letters **58**, 23–28 (2015). https://doi.org/https://doi.org/10.1016/j.patrec.2015.02.001, https://www.sciencedirect.com/science/article/pii/S0167865515000380
56. Valencia, K., Rusu, C., Quiñones, D., Jamet, E.: The impact of technology on people with autism spectrum disorder: A systematic literature review. Sensors **19**(20) (10 2019). https://doi.org/10.3390/s19204485
57. Walsh, E., Holloway, J., Lydon, H.: An evaluation of a social skills intervention for adults with autism spectrum disorder and intellectual disabilities preparing for employment in ireland: A pilot study. Journal of Autism and Developmental Disorders **48** (05 2018). https://doi.org/10.1007/s10803-017-3441-5
58. Zhang, L., Warren, Z., Swanson, A., Weitlauf, A., Sarkar, N.: Understanding performance and verbal-communication of children with asd in a collaborative virtual environment. Journal of Autism and Developmental Disorders **48** (08 2018). https://doi.org/10.1007/s10803-018-3544-7

59. Zhao, Z., Chen, Y., Liu, J., Shen, Z., Liu, M.: Cross-people mobile-phone based activity recognition. pp. 2545–2550 (01 2011). https://doi.org/10.5591/978-1-57735-516-8/IJCAI11-423
60. Zheng, Z., Zhang, L., Bekele, E., Swanson, A.R., Crittendon, J.A., Warren, Z., Sarkar, N.: Impact of robot-mediated interaction system on joint attention skills for children with autism. 2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR) pp. 1–8 (2013)