Proceedings of the ASME 2023
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC/CIE2023
August 20-23, 2023, Boston, Massachusetts

DETC2023-116492

EARLY PREDICTION OF HUMAN INTENTION FOR HUMAN-ROBOT COLLABORATION USING TRANSFORMER NETWORK

Xinyao Zhang
Graduate Research Assistant
Environmental Engineering Sciences
University of Florida,
Gainesville, FL, 32611
xinyaozhang@ufl.edu

Sibo Tian
Graduate Research Assistant
Mechanical and Aerospace Engineering
University at Buffalo, Buffalo, NY, 14260
sibotian@buffalo.edu

Xiao Liang
Assistant Professor
Civil, Structural, and Environmental Engineering
University at Buffalo, Buffalo, NY, 14260
liangx@buffalo.edu

Minghui Zheng
Assistant Professor
Mechanical and Aerospace Engineering
University at Buffalo, Buffalo, NY, 14260
mhzheng@buffalo.edu

Sara Behdad*
Associate Professor
Environmental Engineering Sciences
University of Florida, Gainesville, FL, 32611
sarabehdad@ufl.edu

ABSTRACT

Activity recognition is a crucial aspect in smart manufacturing and human-robot collaboration, as robots play a vital role in improving efficiency and safety by accurately recognizing human intentions and proactively assisting with tasks. Current human intention recognition applications only consider the accuracy of recognition but ignore the importance of predicting it in advance. Given human reaching movements, we want to equip the robot with the ability to predict human intent not only with precise recognition but also at an early stage. In this paper, we propose a framework to apply Transformer-based and LSTM-based models to learn motion intentions. Second, based on the observation of distances of human joints along the motion trajectory, we explore how we can use the hidden Markov model to find intent state transitions, i.e., intent uncertainty and intent certainty. Finally, two data types are generated, one for the full data and the other for the length of data before state transitions; both data are evaluated on models to assess the robustness of intention prediction. We conducted experiments in a manufacturing workspace where the experimenter reaches multiple scattered targets and further this experimental scenario was designed to examine how intents differ, but motions are only slightly different. The proposed models were then evaluated with experimental data, and further performance comparisons were made between models and between different intents. Finally, early predictions were validated to be better than using full-length data.

Keywords: Human intention recognition, Early prediction, Transformer, Hidden Markov model, Human-robot collaboration, Manufacturing

1. INTRODUCTION

In recent years, human-robot collaboration (HRC) has become increasingly popular for common co-assembly tasks in manufacturing settings. A widely spread application is where a human retrieves components and places them, then the robot picks up the placed components and begins assembling them into a product [1]. Moreover, in the pursuit of an efficient robotic cooperative environment, robots are able to respond quickly or slowly depending on the speed of the human in the assembly task [2]. However, human operators and robots usually work separately and are treated as independent agents, because humans can perform in a more flexible manner, but robots are set to a fixed automation mode. Additionally, humans can perceive others' actions and infer their intentions as a way to start off relevant complementary actions, which are difficult for robots to

predict. Therefore, a higher level of understanding of human intent and enhanced rapid adaptation of robots are required.

Unlike other physical features, such as location coordinates or distance traveled, human intent is implicitly contextual and does not have direct observability; however, it is actually encoded and expressed in human actions [3]. In particular, the movement and orientation of workers have a significant impact on the recognition of intent in the warehouse [4]. With respect to abundant information encoded in human actions, observing and interpreting it is beneficial for us to understand human intent. Recently, researchers have proposed new considerations for cooperation between humans and robots, in which the recognition of human intent can be leveraged to control the robot [5]. Among other advantages, the sequence of assembly activities is predicted by modeling the motion to recognize human intent [6]. Another application in the assembly process is the measurement of quality insurance and human failure detection through the recognition of human intent [7].

Inspired by the necessity of intent recognition and the legibility of actions, achieving explicit human intent recognition is the driving force behind our research. Driven by the development of deep learning, state-of-the-art algorithms are showing great promise in providing intelligent solutions [8]. Convolutional recurrent neural networks (CRNN) effectively learn the temporal and spatial relationships embedded in human body actions [9]. Other researchers have introduced recursive Bayesian filtering methods to explore the correlation between intent and non-verbal behavior [10]. Although there are a variety of case studies [3]–[5] on assessing human intention recognition, the importance of how to effectively predict it has been overlooked. Inspired by improving the efficiency of HRC, we design an intention recognition framework, as shown in Figure 1, and further implement prediction at an early stage.

The objective of this study is to propose a novel framework for motion-based human intention recognition. In terms of model selection, two types of architectures including Bidirectional Long short-term memory (Bi-LSTM) and Transformer have been used for the prediction purposes. We choose Bi-LSTM network, which is capable of learning inputs by forward and backward directions. Also, empowered by the novelty of the Transformer model, we apply it to validate its performance on the task of intent recognition.

Further, given the practice of the Hidden Markov Model (HMM) for continuous action division, we incorporate the joint distance to an operator into the HMM for segmenting uncertainty of intent and certainty of intent.

In this study, human intent is defined as judging the operator's goal based on the observed trajectory of reaching movements. We have conducted two cases of experimental studies, specifically, one in which similar targets are grouped together and the other in which different intentions are identified by separating them when the motion trajectories are very similar. The operator's arm motion captured by the Vicon system is the input to the model, and the intent based on the motion is subsequently predicted.

We compare in detail the performance of the Transformer and Bi-LSTM models in both cases and present recommendations for model selection regarding the task specificity case. In addition, we use the HMM to compute the state transitions for each reaching trajectory and evaluate the performance of the early predictions over the predictions for all data.

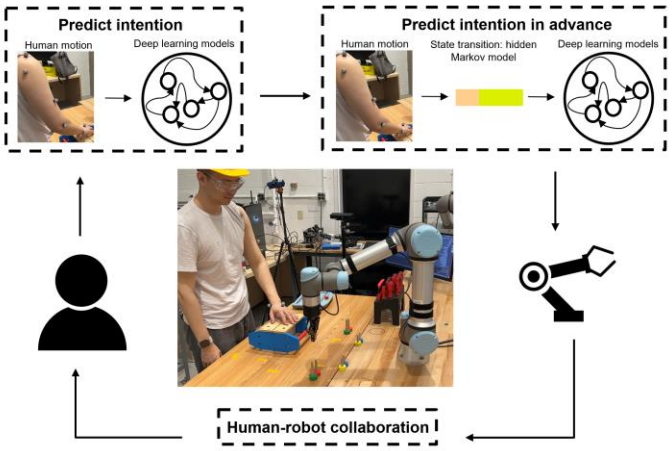

FIGURE 1: INTENTION RECOGNITION AND EARLY PREDICTION FRAMEWORK

The rest of the paper is structured as follows. Section 2 compares related studies on the topic of intent recognition. Section 3 describes the Transformer and Bi-LSTM architectures. Section 4 presents the experimental design, the data set, and practical results. This section describes the results of each phase and the comparison between models. Section 5 concludes the paper and extends to potential future work.

2. RELATED WORK

In this section, we briefly summarize related work in the literature dealing with the importance of intention learning, its perception methods, and prediction methods.

In teamwork, team members can coordinate their actions among themselves by predicting each other's intentions. Although we humans possess this knowledge, it is still a challenge to get robots to predict and adjust their actions accordingly. For example, in manufacturing, if collaborative robots are programmed in a fixed offline manner, it is labor-intensive to recode the corresponding unexpected collaborations that are likely to occur with a change in human intent [11]. On the other hand, considering situational needs people have been shown to unconsciously adjust their behavior, such as movement speed and execution paths [12]. This situation has a high probability to happen in a manufacturing workplace where an operator has multiple trajectories of motion to pick up and place a large number of tools or parts during assembly. Both the speed of movement and the path of movement are not stable, so we think that human intent is informative and understanding it becomes more crucial.

Within task-specific scenarios, there are plenty of interpretations that can make human intent legible to robots. By

collecting Electroencephalography (EEG) signals on a person's scalp, it is possible to understand the person's intentions, as the EGG signal fluctuates in different patterns when a person wants to move parts of the body [9]. Similarly, surface electromyography (sEMG) signals can be used to estimate associated biomechanics motion by measuring the velocity or acceleration of muscles [13]. At the same time, there are some limitations to collecting bioelectrical signals, namely, the collected signals contain too much noise, and the sensor equipment affects the flexibility of experimenters. Along with the popularity of image data, more recognition tasks in the manufacturing field with images as the theme are proposed [14], [15]. For instance, the context of human movements can be recorded in images as they operate any part or tool [16]. Even for visual data involving rich motion information, processing them to extract and analyze intent requires a lot of manual labeling work. Taking advantage of the motion tracking sensor system, we infer intent directly through the motion trajectory data.

Recently, the safety of HRC has received more attention. For its part, the operator's intent or goal is a prerequisite so that the robot's behavior can be adjusted correspondingly. The underlying methods can be divided into two groups: machine learning-based models and deep learning-based models. In terms of machine learning methods, the researchers use support vector machine and random forest algorithms for feature retrieval and daily motion classification [17], [18]. In addition to this, neural network-based deep learning has a wide range of applications in intent estimation. Based on predefined goals in the workspace, RNN is trained to switch between various human motion data by continuously updating the input bias values [19]. In particular, as a representative of RNN, LSTM can learn linear and nonlinear features of sequential data and overcome the weakness of time dependence [20].

The quality performance of Transformers in predicting the intention and trajectory of pedestrians inspires us to apply it to the intent classification task in manufacturing sites [21]. A special attention mechanics in Transformer allows us to pursue connections in any part of sequential data [22]. Besides just a single intent recognition task, we discuss how to predict intent early and verify whether the use of full-length sequences is necessary to achieve accurate predictions. We use HMM as an elaboration algorithm to discretize sequences into states [23]. Such state division technique becomes a well fit when we pursue from states with uncertain intentions to states with certain intentions.

3. METHODOLOGY

In this section, we propose Transformer and Bi-LSTM models to learn the relationship between human intentions and motion trajectories. Furthermore, we apply HMM to find the transition from uncertainty of intent to certainty of intent and find the length of the sequence data as input to classifier networks.

3.1 Transformer Model for Intention Recognition

Transformer is first proposed by its unique application of attention mechanism [24]. The advanced nature of the attention mechanism is that it allows modeling sequential dependencies regardless of their position in the input or output. Consequently, Transformers have achieved good performance in language processing tasks [25]. However, it has been observed that Transformers have not been widely employed for motion-based analysis of human intent.

In our study, the Transformer model introduced is displayed in Figure 2. First, each sequence of trajectories of human motion is taken as input and will be normalized. As the core of the Transformer, the attention mechanism will build a representation with query, key and value vectors to model each data point of the input sequence, given by

$$Attenion(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

where Q , K and V are vectors named as query, key and value respectively and $\sqrt{d_k}$ is the so-called scale factor. This Scaled Dot-Product Attention calculates the attention value of each input element.

Then, the Multi-Head Attention will parallelly compute and join the complex information of more representations at different positions of input data. Since no recurrence or convolution calculation is required in the Multi-Head Attention, each input element is provided to the feedforward network along with the associated positional information. Last, all embedded elements are passed through a normalization layer to speed up learning, and then a classifier with a SoftMax activation function is used to determine the intent class.

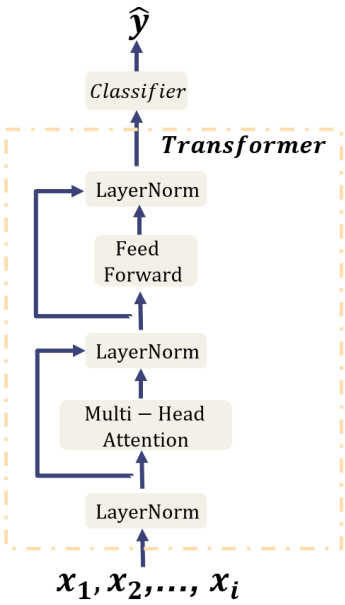

FIGURE 2: THE TRANSFORMER MODEL'S STRUCTURE

3.2 Bi-LSTM Model for Intention Recognition

Besides Transformer, we also used Bi-LSTM. Before the emergence of Transformers, LSTM architectures are often selected to learn the long-time dependencies of sequential data.

The results of this study demonstrate that an LSTM network is allowed to learn features in the temporal domain and recognize human activities correspondingly [26]. It validates the ability of LSTM to extract behavioral features from time series data. Nevertheless, an LSTM layer only learns the data structure in a fixed direction, i.e., after starting from the motion, but lacks learning from later motion to forward.

In order to learn motion sequences not only in the feedforward direction but also in the backward direction, we design the Bi-LSTM model in Figure 3.

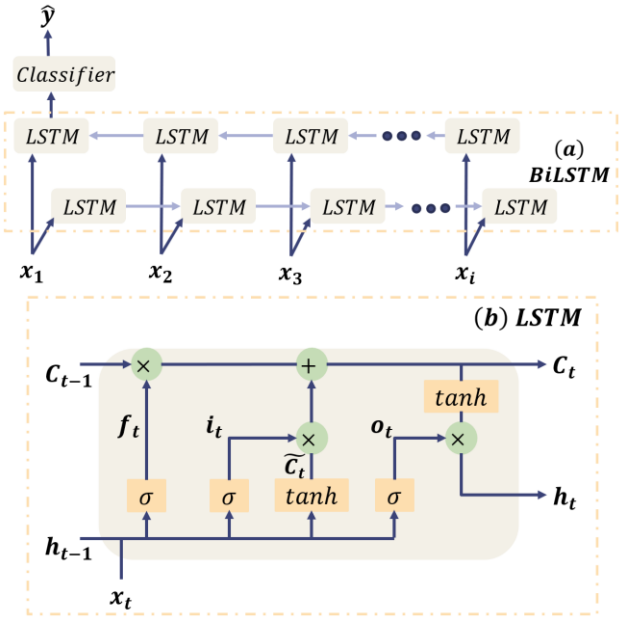

FIGURE 3: (a) THE BI-LSTM MODEL'S STRUCTURE. (b) THE LSTM CELL WORKFLOW.

As seen in Figure 3(a), there are two ways of stacking LSTM cells. In the forward network flow, it learns each piece of information from previous elements to the future time; in the backward network flow, it learns the upcoming information in the reverse time. In detail, Figure 3(b) illustrates how each LSTM cell performs operations. Equation (2) – Equation (6) are mathematically interpreted as [27]

$$f_t = \sigma(w_{fx}x_t + w_{fh}h_{t-1} + b_f) \quad (2)$$

$$i_t = \sigma(w_{ix}x_t + w_{ih}h_{t-1} + b_i) \quad (3)$$

$$o_t = \sigma(w_{ox}x_t + w_{oh}h_{t-1} + b_o) \quad (4)$$

$$c_t = c_{t-1} \odot f_t + i_t \odot \tanh(w_{ct}x_t + w_{ch}h_{t-1} + b_c) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

where f_t, i_t and o_t are namly the forget gate, input gate, and output gate. x_t and h_t are input element and hidden state. $\odot$ represent element-wise vectors multiplication.

3.3 HMM Model for Intent State Transition

In addition to deep learning-based intention learning methods, an HMM has been applied separately to perform state transitions. The HMM model presents a successful case for segmenting continuous behavior [28]. For a given input sequence, the HMM can model the data as different states by measuring the likelihood. Taking advantage of dividing continuous motions, we plan to use an HMM to compute the probability of hidden states, where the information of state transitions is related to intent shifts in time.

The proposed HMM model is described in Figure 4. First, we calculate the Euclidean distance of joints in each motion trajectory. It can be easily understood that the motion starts with a slow, but gradually moves away from the original position. These Euclidean distances about human joints are the observation variables of the HMM model. Further, we set the number of hidden states in the HMM to 2. The HMM will classify the sequences into two continuous states based on the distance. We ultimately care about the time of the state transition, since we extract the length of the data from the beginning of the motion to the state transition as input to the intention classification model for early prediction.

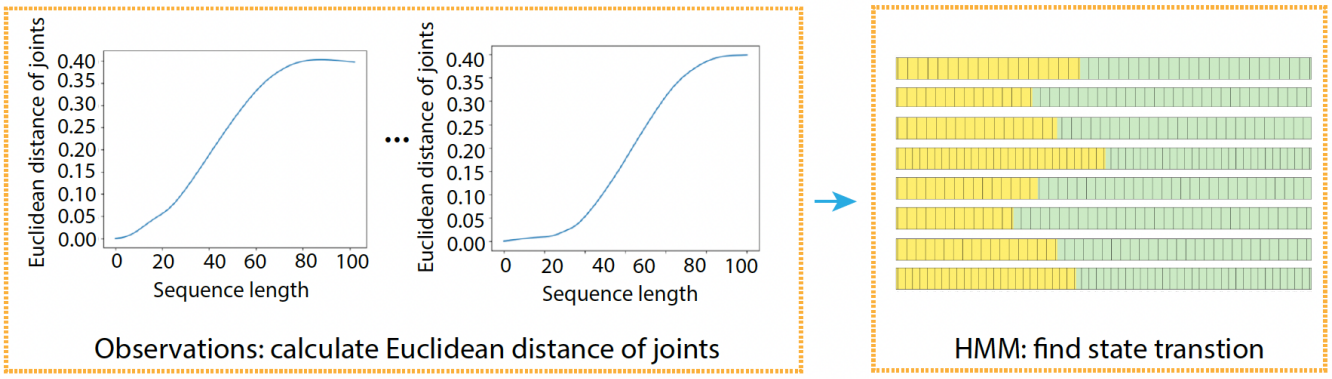

FIGURE 4: THE PROCESS OF IDENTIFYING STATE TRANSITION USING HIDDEN MARKOV MODEL: CALCULATING EUCLIDEAN DISTANCES OF JOINTS AND PUT IT INTO THE HMM .

4. USE-CASE AND RESULTS

This section presents a complete case study on human intent recognition. First, we designed an experiment on a collaborative human-robot environment in manufacturing. Two cases are presented separately. In addition, we train our previously proposed models with the experimental dataset. Multiple comparisons are discussed in detail. Finally, we validate the idea about the advance prediction of human intentions, which means intentions can be recognized before their movement is complete.

4.1 Experimental Design and Dataset

To validate the effectiveness of the proposed models, an experiment is designed to replicate collaborative manufacturing in a real-world setting. In the experiment, a human operator stands opposite a robot manipulator and reaches targets from four distinct locations to place them in a collection box. Each location contains two different types of screws. As a result, the reaching motions for two screws at the same location are similar, but the human operator's intent is different. Therefore, we have two cases of experimental study. One is to predict the target location that the human operator is reaching for among four distinct areas displayed in Figure 5(a), while the other case predicts which screw the human wants to retrieve among all eight screws displayed in Figure 5(b).

The Vicon motion capture system is used to track the movement of the human operator's right arm. Two markers are attached to each side of the wrist, elbow, and shoulder. The data is recorded as a sequence of Cartesian coordinates for each marker, at a frequency of 50 Hz, resulting in a trajectory time interval of 0.02 seconds. The center of each rotation joint can be easily estimated by taking the mean of the two markers' positions.

We separate and select the trial data into different reaching motions, which only contain the static-to-static human motion starting from the collection box and ending at the targets located at different places. In this case, the dataset we collected could be used to train the model and predict the human intent of reaching which target location (4 labels) or retrieving which screws (8 labels).

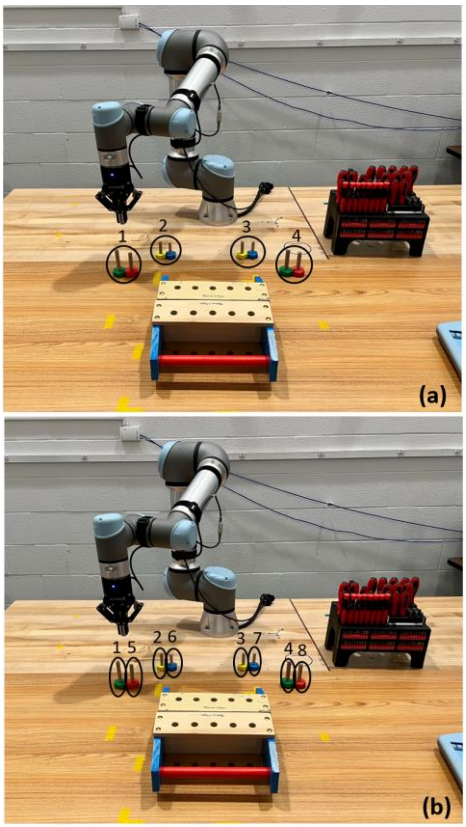

FIGURE 5: (a) THE EXPERIMENT WITH FOUR-LABEL INTENTIONS. (b) THE EXPERIMENT WITH EIGHT-LABEL INTENTIONS.

In addition, we visualize the trajectory of two approaching targets from the beginning to the end of the motion in Figure 6. From the observation, we can see that the trajectories of the two approaching targets are highly similar, especially since the trajectories almost overlap at the end of the motion. In total, we have 232 motion data in total and each class has an equal amount of data. The length of single motion data is approximately 2 seconds.

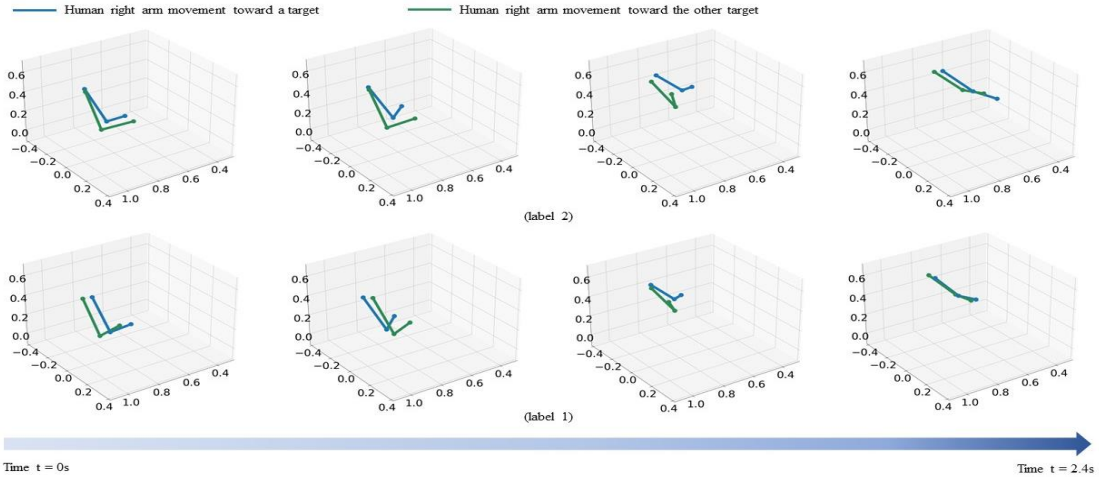

FIGURE 6: VISUALIZATION OF THE TRAJECTORY OF TWO APPROACHING TARGETS.

4.2 Results of Intent Classification with Four Labels

All proposed models are built using Keras and TensorFlow. We applied the Adam optimizer with the learning rate fixed to 0.001. In order to make results reproducible, we also fixed the values of random seed, which included a total of 5 seeds. The training epoch was set to 500 epochs. Using a single Nvidia 3080 GPU, the experiment was carried out while splitting the data into training and testing: 70% for training and 30% for testing.

To evaluate the performance of classification results, we used boxplots to compare different models. As a standardized view, a boxplot is able to show us the outliers of data as well as their distribution. Meanwhile, heatmaps were displayed to visualize the classification output for different intents.

The results of the four-label intent classification are shown in Figure 7. We evaluate the trained models by consequently increasing the data length from 20% to 100% and using 20% as an interval.

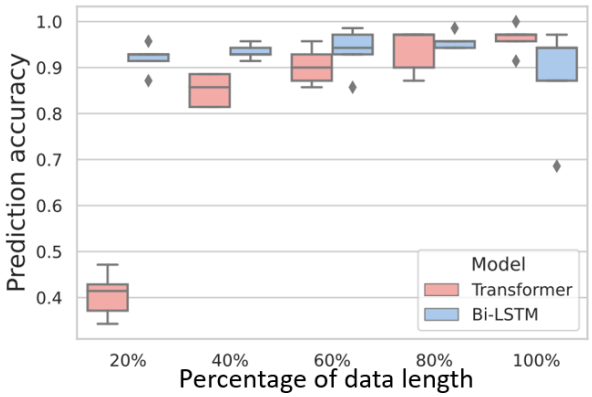

FIGURE 7: COMPARE FOUR-LABEL CLASSIFICATION RESULTS OF DIFFERENT MODELS ON ON TEST DATA

For the Bi-LSTM model, the classification accuracy decreases when the percentage of data increases to 60%, while the accuracy of the Transformer model continues to improve with increasing data. There is only one caveat if we use only 20% of the data for prediction, Transformer's accuracy will be significantly lower than Bi-LSTM. The final Transformer outperforms when using the full range of trajectory data to make predictions. In addition, when comparing the classification results of a specific label with other labels, we make use of the heat map in Figure 8.

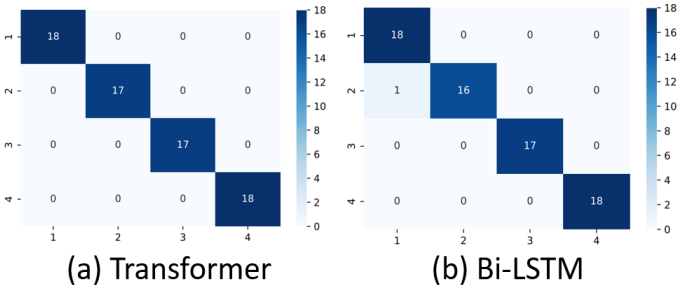

FIGURE 8: CLASSIFICATION RESULTS PER LABEL

When predicting full-length trajectories, Transformer is 100% accurate, while Bi-LSTM incorrectly predicts a set of motion for Label 2 as Label 1 because of the close location of the two labels. In general, if the targets are close to each other, but can be labeled as a group, we recommend using the Transformer model in preference.

4.3 Results of Intent Classification with Eight Labels

As we stated before, it makes sense to analyze the reaching motions when the targets are close to each other. Especially in manufacturing sites, many tools or parts needed during operation are often placed together. Apart from that, training models with a dataset of 8 labels increases the computational time and complexity. Testing results are illustared in Figure 9. The performance of both models is degraded, and the overall performance of Transformer is less stable than Bi-LSTM.

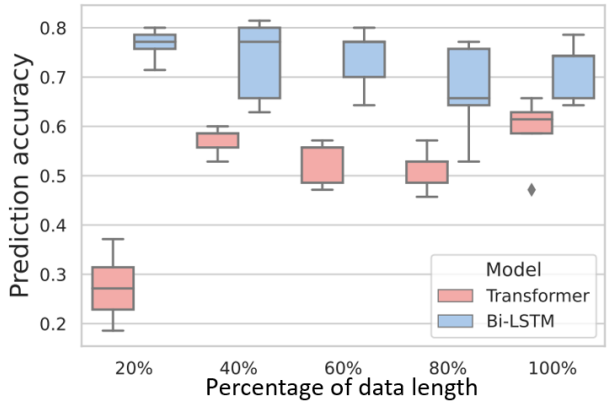

FIGURE 9: COMPARE EIGHT-LABEL CLASSIFICATION RESULTS OF DIFFERENT MODELS ON ON TEST DATA

Further discussing the reasons for the reduced accuracy, we observe that the intentions of two targets in close proximity are easily misclassified in both models, e.g. intentions labeled 7 are confused with labeled 3, as shown in Figure 10.

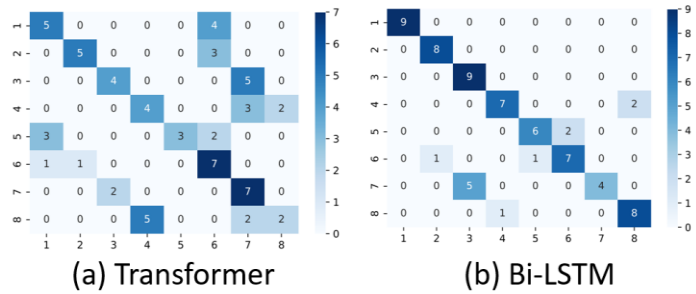

FIGURE 10: CLASSIFICATION RESULTS PER LABEL

Another observation is that Bi-LSTM outperforms Transformer in terms of prediction using the entire data sequence. This may be due to the nature of both models. Bi-LSTM learns the data structure backward and forward to figure out the underlying dynamics of the data; for Transformer, attention is focused on specific parts of the data, whereas early

prediction may result in too little data without sufficient attention value. To conclude, Transformer is not suitable for tasks with similar trajectory intent prediction, while bi-LSTM is a better choice.

4.4 Trajectory State Transition Results from HMM

After testing predictions using data of different lengths, we found that this conclusion of training with many data elements to obtain more accurate results was not always true. Therefore, the use of HMM is necessary to help us find the best length series to achieve better accuracy as well as earlier predictions.

In practice, the HMM is used to segment the hidden states of human motion, with the goal of separating trajectories into states where the intention is uncertain and those where the intention is certain. Uncertainty of intention means that the experimenter has no clear intention at the beginning of the action, or the motion changes to a small extent. And towards the end of the motion, intentions will gradually become clear. Thus, the implied state of intention regarding the behavior shifts from uncertainty to clarity. To achieve it, we first calculated the Eulidance distance variation of the joint with motion, i.e., the joint position at each time point minus the joint position at the beginning. Next, we used these distances as inputs to an HMM model with two hidden states.

The output of the HMM model provides a division of behavioral states. Since dividing 8 labels is a more complex case study, we present the average state transitions for each label in Figure 11. In this figure, each status bar represents 3 time points, displayed in yellow to indicate uncertain intent, and in green to indicate certain intent. Subsequently, we used all trajectory motions from the beginning to the state transition as the dataset for the early prediction.

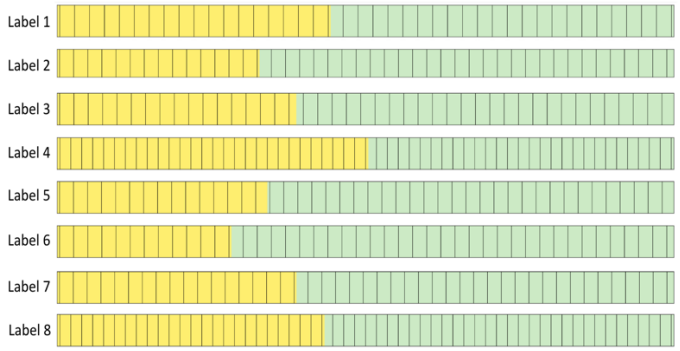

FIGURE 11: AVERAGE STATE TRANSITION RESULTS FOR DIFFERENT INTENTS

4.5 Intent Early Prediction

To implement the idea of early prediction of intent, we plan to validate the two models under experiments with eight labels. First, we prepare the data whose sequence length is the length from the start point to the time transition point determined by the HMM. Second, we train the prepared dataset with both models and compare the results with the model's performance on the full-length sequences dataset.

A summary of the comparison is illustrated in Figure 12. For both models, the length of the data elements calculated using the HMM achieved better prediction accuracy compared to using all data elements. If we observe the results of individual models, Transformer's early prediction is more stable than the prediction with all data. Also, using early predictions, Bi-LSTM can improve the accuracy by about 10%. As discussed earlier, the overall performance of the Bi-LSTM model is better than the Transformer model because it is more suitable for intent classification scenarios with eight labels.

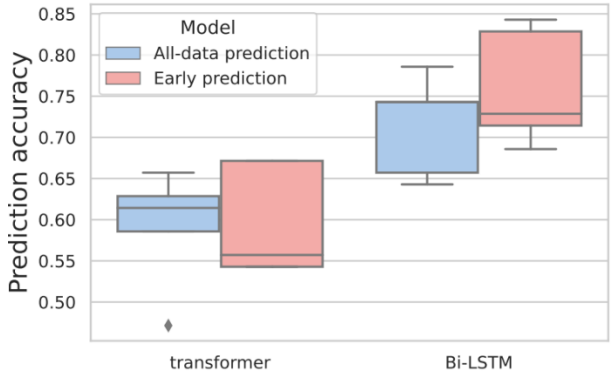

FIGURE 12: COMPARISON OF PREDICTIONS USING EARLY DATA WITH ALL-DATA PREDICTIONS

5. CONCLUSION AND FUTURE WORK

In this study, a framework for intent prediction based on human movement data is proposed. The proposed framework includes the use of Transformer and Bi-LSTM models to learn motion data, and HMM to determine the division of intention shifts. According to our experimental study, we found that the Transformer architecture has better results for classifying intentions with large action differences, while the Bi-LSTM architecture has more robust results for identifying similar actions. Also, by using the state transition information from HMM, the intention prediction results are higher than those predicted using all data. The use of the Transformer model provides reference suggestions on the task of human intent classification; the Bi-LSTM model is more suitable for intention identification where the trajectories are very similar; and the application of the HMM allows us to achieve early prediction of intentions without using all the data. All models and methods are designed according to the needs of manufacturing sites, i.e., operators picking up targets at different locations.

The proposed work enhances human-robot collaboration in several ways. First, by predicting human intentions, robots can anticipate human future actions and provide the necessary help, while reducing the time and improving the overall efficiency. Second, recognizing human intent enables robots to identify hazardous situations and create a safer work environment. Third, by utilizing human intent recognition, robots can provide more personalized assistance to workers, easily accomplish their tasks

and adap to different tasks. Additionally, the use of human intent recognition can enable robots to perform more complex tasks that may require interaction with humans, such as assembly or inspection

The proposed framework can be extended in several ways. First, for data with similar motion trajectories, we can utilize deep feature extraction techniques to achieve a prediction accuracy above 90%. This is especially important when the trajectories are similar, but the intentions behind the motions are entirely different. Second, when performing non-sequential or coordinated tasks, it is essential to examine how dynamic interactions among humans and robotcs can affect the regnition of human intent. Third, while the current frameworks use joint movements to predict human intent, it is worth exploring how small-scale movements, such as wrist and finger levels, can be utilized for learning and predicting human intent.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation–USA under grants # 2026276 and 2026533. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Morato, C. W., Kaipa, K. N., & Gupta, S. K., 2017, "System State Monitoring to Facilitate Safe and Efficient Human-Robot Collaboration in Hybrid Assembly Cells," Proceedings of the IDETC/CIE, Cleveland, Ohio, August 6–9, 2017, vol. 58110, p. V001T02A012.

[2] Etzi, R., Huang, S., Scurati, G. W., Lyu, S., Ferrise, F., Gallace, A., ... & Bordegoni, M., 2019, "Using Virtual Reality to Test Human-Robot Interaction During a Collaborative Task," Proceedings of the IDETC/CIE, Anaheim, California, August 18–21, 2019, vol. 59179, p. V001T02A080.

[3] Stulp, F., Grizou, J., Busch, B., & Lopes, M., 2015, "Facilitating intention prediction for humans by optimizing robot motions," 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 2015, pp. 1249-1255.

[4] Petković, T., Puljiz, D., Marković, I., & Hein, B., 2019, "Human intention estimation based on hidden Markov model motion validation for safe flexible robotized warehouses," Robotics and Computer-Integrated Manufacturing, 57, pp. 182-196.

[5] Losey, D. P., McDonald, C. G., Battaglia, E., and O'Malley, M. K., 2018, "A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human–Robot Interaction," ASME. Appl. Mech. Rev., 70(1).

[6] Manns, M., Tuli, T. B., & Schreiber, F., 2021, "Identifying human intention during assembly operations using wearable motion capturing systems including eye focus," Procedia CIRP, 104, pp. 924-929.

[7] Gajjar, N. K., Rekik, K., Kanso, A., & Müller, R., 2022, "Human intention and workspace recognition for collaborative assembly," IFAC-PapersOnLine, 55(10), pp. 365-370.

[8] Nahavandi, S., 2019, "Industry 5.0-A Human-Centric Solution," Sustainability, 11(16), pp. 4371.

[9] Zhang, D., Yao, L., Chen, K., Wang, S., Chang, X., & Liu, Y., 2019, "Making sense of spatio-temporal preserving representations for EEG-based human intention recognition," IEEE transactions on cybernetics, 50(7), pp. 3033-3044.

[10] Jain, S., & Argall, B., 2019, "Probabilistic human intent recognition for shared autonomy in assistive robotics," ACM Transactions on Human-Robot Interaction (THRI), 9(1), pp. 1-23.

[11] Wang, W., Li, R., Chen, Y., & Jia, Y., 2018, "Human intention prediction in human-robot collaborative tasks," In Companion of the 2018 ACM/IEEE international conference on human-robot interaction, pp. 279-280.

[12] Koppenborg, M., Nickel, P., Naber, B., Lungfiel, A., & Huelke, M., 2017, "Effects of movement speed and predictability in human–robot collaboration," Human Factors and Ergonomics in Manufacturing & Service Industries, 27(4), pp. 197-209.

[13] Zhang, L., Liu, G., Han, B., Wang, Z., & Zhang, T., 2019, "sEMG based human motion intention recognition," Journal of Robotics, 2019.

[14] Zhang, X., Eltouny, K., Liang, X., & Behdad, S., 2023, "Automatic Screw Detection and Tool Recommendation System for Robotic Disassembly," Journal of Manufacturing Science and Engineering, 145(3), pp. 031008.

[15] Liao, H. Y., Zheng, M., Hu, B., & Behdad, S., 2022, "Human Hand Motion Prediction in Disassembly Operations," Proceedings of the IDETC/CIE, St. Louis, Missouri, August 14–17, 2022, vol. 86250, p. V005T05A021.

[16] Wang, P., Liu, H., Wang, L., & Gao, R. X., 2018, "Deep learning-based human motion recognition for predictive context-aware human-robot collaboration," CIRP annals, 67(1), pp. 17-20.

[17] Vu, C. C., & Kim, J., 2018, “Human motion recognition by textile sensors based on machine learning algorithms,” Sensors, 18(9), pp. 3109.
[18] Batool, M., Jalal, A., & Kim, K., 2019, “Human intention estimation based on neural networks for enhanced collaboration with robots,” In 2019 international conference on applied and engineering mathematics (ICAEM), pp. 145-150.
[20] Yan, L., Gao, X., Zhang, X., & Chang, S., 2019, “Human-robot collaboration by intention recognition using deep LSTM neural network,” In 2019 IEEE 8th International Conference on Fluid Power and Mechatronics (FPM), pp. 1390-1396.
[21] Sui, Z., Zhou, Y., Zhao, X., Chen, A., & Ni, Y., 2021, “Joint intention and trajectory prediction based on transformer,” In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 7082-7088.
[22] Henderson, M., Casanueva, I., Mrkšić, N., Su, P. H., Wen, T. H., & Vulić, I., 2019, “ConveRT: Efficient and accurate conversational representations from transformers,” arXiv preprint arXiv:1911.03688.
[23] Liu, H., & Wang, L., 2017, “Human motion prediction for human-robot collaboration,” Journal of Manufacturing Systems, 44, pp. 287-294.
[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I., 2017, “Attention is all you need,” Advances in neural information processing systems, 30.
[25] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., 2018, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805.
[26] Chen, Z., Zhang, L., Cao, Z., & Guo, J., 2018, “Distilling the knowledge from handcrafted features for human activity recognition,” IEEE Transactions on Industrial Informatics, 14(10), pp. 4334-4342.
[27] Tong, Y., Liang, Y., Spasic, I., Hicks, Y., Hu, H., & Liu, Y., 2022, “A Data-Driven Approach for Integrating Hedonic Quality and Pragmatic Quality in User Experience Modeling,” Journal of Computing and Information Science in Engineering, 22(6), pp. 061002.
[28] Yi, D., Musall, S., Churchland, A., Padilla-Coreano, N., & Saxena, S., 2022, “Disentangled multi-subject and social behavioral representations through a constrained subspace variational autoencoder (CS-VAE),” bioRxiv, 2022-09.