Toward a Real-time Activity Segmentation Method for Human-Robot Teaming

Tariq Iqbal*, Laurel D. Riek†, Julie A. Shah*
*Computer Science and Artificial Intelligence Lab
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
†Department of Computer Science and Engineering
University of California San Diego
La Jolla, CA 92093, USA
Email: tiqbal@mit.edu, lriek@eng.ucsd.edu, julie_a_shah@csail.mit.edu

***Abstract*—When a robot collaborates with people in groups, its interaction with humans is expected to be fluent and efficient. Augmenting a robot with the capacity to understand the activities of the people it is collaborating with (with specific reference to the *timing* of those activities), allows the robot to leverage its understanding to generate an efficient and collaborative plan to perform its actions. In this paper, we present a supervised activity segmentation algorithm that can detect the start and end time of activities, simply by observing a portion of the initial trajectory data: an essential first step in generating an efficient interaction plan for a robot. We validated the algorithm by applying it to a collaborative task involving a single robot and a single human. Results of this study indicate that the algorithm accurately segments the activities in real time with approximately 80% accuracy (partially observing the full trajectory of a given activity).**

I. Introduction

Robots now have the capacity to help people in many areas, from health care to assistive living to manufacturing and factory settings. In many of these scenarios, robots are sharing physical space and collaborate with people in teams [1, 2, 3]. The performance of many of these teams depends on how fluently all the members of a team can jointly perform tasks [4, 5, 6]. People are inherently skilled at coordinating with others - thus, to be effective in teams with a human, a robot is also expected to understand how people interact in teams and how joint or collaborative actions are performed.

When a person acts alone, their behavior is very different from when they coordinate in groups [7]. For human groups, Sebanz et al. [8] defined joint actions as a form of interaction where two or more members coordinate their actions in space and time while making changes to their environment. The authors described three important parts in a successful joint action: a prediction about the intention of other interactional partners, an understanding of when to act jointly, and an understanding of where and how to perform the given action.

To successfully perform a joint action with people, a robot requires the capacity to perceive other members actions, perform predictions in time and space, and create adaptable plans in real-time. Through understanding peoples' interaction and through anticipating their future needs, a robot can perform a successful joint action with people by planning and scheduling their activities and tasks intelligently [9, 10, 11]. In doing so, the human-robot team can achieve a goal which neither the human nor the robot can accomplish alone.

For example, a significant portion of assembly processes in manufacturing involve the physical movement of raw materials and finished products. Often, a worker who is performing an assembly task is required to bring raw materials to and from an assembly table where he or she assembles the parts. In these circumstances, a person often needs to perform non-value added tasks to keep the production line moving, which generates stress and fatigue, while detracting from the efficiency of the task. An introduction of a robot to this scenario will alleviate human involvements in these non-value added tasks while attempting to achieve goals efficiently while collaborating fluently.

If a robot has an understanding of human activities, it can utilize its understanding to generate intelligent task assignments and schedules. For example, Gombolay et al. developed a centralized task assignment and scheduling algorithm, Tercio, suitable for human-robot teams [12, 13]. Their algorithm computes a multi-agent schedule in polynomial time which satisfies the temporal deadlines as well as spatial constraints on agent proximity. For a similar scenario, Zhang et al. [14, 15] introduced four fairness criteria in multi-agent decision making process under uncertainties.

Due to the unpredictable and dynamic nature of human behavior, it is essential to have a robust understanding of current and future activity timings for these scheduling algorithms to run appropriately. As such, a robot now needs to solve another problem: before it can utilize the planning and scheduling algorithms, it must separate one activity from another in real time correctly. This problem is defined as the activity segmentation problem. For example, if a person moves to a table and then picks up an object placed upon it, the robot would need to begin to detect the activity as soon as the person starts walking towards the table, understand when the person reaches the table, and then understand when the person picks the object up.

Researchers across many fields attempt to address the problem of activity segmentation, and two primary approaches have been explored: supervised and unsupervised segmentation approaches. In a supervised segmentation approach, machine learning algorithms are trained with segmented examples prior to testing. In unsupervised methods, the algorithms do not use labels or dictionaries.

Fearnhead and Liu [16, 17] proposed an online algorithm for changepoint detection problems, by introducing a re-sampling method similar to particle filters to reduce the computational cost. To teach a robot appropriate skill structures, Konidaris et al. [18] extended this idea to construct skill trees to acquire skills from human demonstrations. In a similar vein, Ryoo et al. [19] developed a supervised approach for early recognition of activities from robot-centric videos during its interaction with humans. They introduced the concept of onset which summarizes pre-activity observations using a cascade histogram of time series gradients. Their recognition approach considered event history in addition to visual features from videos to perform activities recognition.

On the other hand, Krishnan et al. [20] introduced an unsupervised segmentation approach, known as the transition state clustering (TSC) method. Their approach combined hybrid dynamical systems and Bayesian non-parametric statistics to segment kinematic actions of robotic surgical procedures. Wu et al. [21] proposed an unsupervised algorithm which models high-level co-occurrences and temporal relations between various human actions. They developed a probabilistic approach by modeling short-term relationships between human actions and objects to infer long-term activities. They then applied that model to perform unsupervised activity segmentation.

Although these methods work for activity segmentation in various scenarios, in our context, they either become computationally expensive to run online, or can not work adequately with a partial observance of data. In this paper, we introduce a supervised activity segmentation algorithm which utilizes an activity classification algorithm to detect the change points (the start and the end time of an activity) in real-time. We validated the algorithm by applying it to a human-robot collaborative task. Initial results suggest that this proposed algorithm can classify activities when partially observing data as well as evaluating activity segments accurately in real-time nearly with approximately 80% accuracy.

II. Activity Segmentation Method

To get the start and the end time of each activity, it is necessary to implement a mechanism for segmenting observed human motion trajectories into discrete activities. We address this challenge by introducing a real-time activity segmentation algorithm which can work by observing partial motion trajectory data.

To perform activity segmentation from real-time sensor data, our algorithm utilizes an activity classification algorithm, which can classify trajectory data to an activity class. Our segmentation algorithm can work with any activity classification algorithm that can classify an activity from partial data. In this implementation, we leverage a real-time activity classification algorithm, Rapid Activity Prediction Through Object-oriented

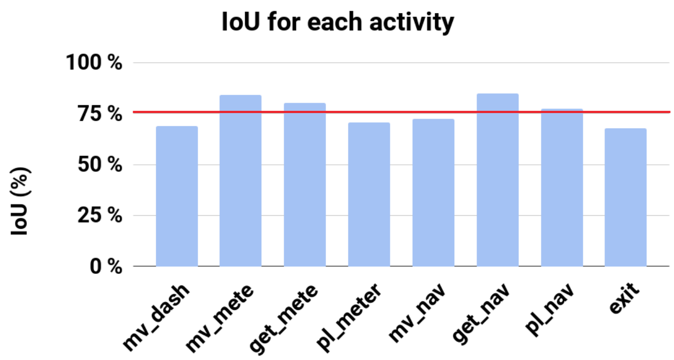

Fig. 1: Intersection-over-union scores for all activity classes

Regression (RAPTOR), developed by Hayes and Shah [22], which can classify an activity from partial data with high confidence.

For partitioning human motion data into activity segments, should the algorithm check all possible combinations of motion class, the process would take an exponential time (an infeasible outcome for real-time implementation on a robot). To perform activity segmentation in real time, our proposed algorithm leverages the prior knowledge of the task sequence to reduce the number of searches; a capability introduced through the utilization of Hierarchical Task Networks (HTN) [23].

This fast segmentation algorithm conducts activity classification on a small partial segment of incoming trajectory data. The algorithm keeps the classification results of these small segments, and classify that segment as the class with the highest likelihood value. When the likelihood of the next activity class becomes higher than the current activity class, the algorithm estimates a change point in the task sequence. By combining all the adjacent partial segments classified as the same activity together, the algorithm detects an activity segment.

Utilizing the prior task structure from the HTN, this algorithm reduces the number of classifiers on which the incoming trajectory data is tested (as a classifier is only used to evaluate a trajectory when a task is supposed to occur at that moment or at an upcoming time in the future). As a result, this technique significantly reduces the number of classifiers required to test the stream.

III. Experimental Testbed and Results

To evaluate the performance of our segmentation algorithm, we designed a dashboard assembly scenario in which a manufacturing associate performed eight sequential tasks. The activity classes are: move to the dashboard (*mv_dash*), move to receive a speedometer object (*mv_mete*), get the speedometer (*get_mete*), place the speedometer on the dashboard (*pl_meter*), move to receive a navigation unit (*mv_nav*), get the navigation unit (*get_nav*), place the navigation unit to the dashboard (*pl_nav*), and exit from the space (*exit*).

The goal of the robot is to assist in the task by delivering the speedometer and the navigation unit to the associate. To achieve this, the robot needs to detect the associate's activity as soon as the activity begins.

TABLE I: Segmentation *accuracy* and *IoU* scores of our algorithm

Activity Segmentation Result	
Accuracy	**79.47%**
IoU	**75.70%**

During an experiment, the associate performed all eight activities sequentially. The position and the orientation of a total of seven objects (left hand, right hand, head, the dashboard, the speedometer, the navigation unit, and a scanner gun) were tracked using a VICON motion capture system.

We applied our algorithm to segment the activities in real-time during this task. We captured a total of 16 instances of these task sequences. We performed eight-fold cross-validation to validate our method.

As the evaluation metrics, we measured the accuracy and the intersection-over-union (*IoU*) scores of the algorithm. For example, if the algorithm segments an activity from a trajectory as *S* and the actual activity segment is *GT*, then the accuracy (*acc*) and *IoU* are measured as:

$$acc = \frac{S \cap GT}{GT} * 100\% \;\; and \;\; IoU = \frac{S \cap GT}{S \cup GT} * 100\%$$

We present the results in Figure 1 (*IoU* only) and Table I. The results suggest that our proposed algorithm can segment activities in real-time with an average *accuracy* of 79.47% and an average *IoU* score of 75.70%. The *IoU* scores are typically more conservative than the *accuracy* scores.

IV. Discussion

The initial results suggest that a robot can utilize our method to segment activities in the very early stages of the activity with only a portion of the trajectory. The results would also seek to suggest the algorithm is capable of accurately identifying the end of that activity in real-time. Building on this foundation, we plan on developing intelligent task assignment and scheduling algorithms to enable robots to perform their tasks around people efficiently. We hope that this research will allow robots to be more effective teammates in the future.

References

[1] T. Iqbal and L. D. Riek, "Coordination dynamics in multi-human multi-robot teams," *IEEE Robotics and Automation Letters (RA-L)*, 2017.

[2] V. V. Unhelkar, H. C. Siu, and J. A. Shah, "Comparative performance of human and mobile robotic assistants in collaborative fetch-and-deliver tasks," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. ACM, 2014, pp. 82–89.

[3] T. Iqbal, M. J. Gonzales, and L. D. Riek, "A Model for Time-Synchronized Sensing and Motion to Support Human-Robot Fluency," in *ACM/IEEE Human-Robot Interaction, Workshop on Timing in HRI*, 2014.

[4] T. Iqbal and L. D. Riek, "A Method for Automatic Detection of Psychomotor Entrainment," *IEEE Transactions on Affective Computing*, 2015.

[5] A. Taylor and L. D. Riek, "Robot perception of human groups in the real world: State of the art," in *AAAI Fall Symposium Series*, 2016.

[6] T. Iqbal and L. Riek, "Assessing group synchrony during a rhythmic social activity: A systemic approach," in *Proc. of the conference of the International Society for Gesture Studies (ISGS)*, 2014.

[7] G. Knoblich and J. S. Jordan, "Action coordination in groups and individuals: learning anticipatory control." *J. Exp. Psychol. Learn. Mem. Cogn.*, 2003.

[8] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: bodies and minds moving together," *Trends Cogn Sci*, 2006.

[9] T. Iqbal, M. J. Gonzales, and L. D. Riek, "Joint action perception to enable fluent human-robot teamwork," in *Proc. of IEEE Robot and Human Interactive Communication*, 2015.

[10] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, "Artificial cognition for social human–robot interaction: An implementation," *Artificial Intelligence*, vol. 247, pp. 45–69, 2017.

[11] T. Iqbal, S. Rack, and L. D. Riek, "Movement coordination in human-robot teams: A dynamical systems approach," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 909–919, 2016.

[12] M. C. Gombolay, X. J. Yang, B. Hayes, N. Seo, Z. Liu, S. Wadhwania, T. Yu, N. Shah, T. Golen, and J. A. Shah, "Robotic assistance in coordination of patient care." in *Robotics: Science and Systems*, 2016.

[13] M. C. Gombolay, R. J. Wilcox, and J. A. Shah, "Fast scheduling of robot teams performing tasks with temporospatial constraints," *IEEE Transactions on Robotics*, vol. 34, no. 1, pp. 220–239, 2018.

[14] C. Zhang and J. A. Shah, "Fairness in multi-agent sequential decision-making," in *Advances in Neural Information Processing Systems*, 2014, pp. 2636–2644.

[15] ——, "On fairness in decision-making under uncertainty: Definitions, computation, and comparison." in *AAAI*, 2015, pp. 3642–3648.

[16] P. Fearnhead and Z. Liu, "Efficient bayesian analysis of multiple changepoint models with dependence across segments," *Statistics and Computing*, vol. 21, no. 2, pp. 217–229, 2011.

[17] ——, "On-line inference for multiple changepoint problems," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 4, pp. 589–605, 2007.

[18] G. Konidaris, S. Kuindersma, R. Grupen, and A. Barto, "Robot learning from demonstration by constructing skill trees," *The International Journal of Robotics Research*, vol. 31, no. 3, pp. 360–375, 2012.

[19] M. Ryoo, T. J. Fuchs, L. Xia, J. K. Aggarwal, and L. Matthies, "Robot-centric activity prediction from first-person videos: What will they do to me?" in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 295–302.

[20] S. Krishnan, A. Garg, S. Patil, C. Lea, G. Hager, P. Abbeel, and K. Goldberg, “Transition state clustering: Unsupervised surgical trajectory segmentation for robot learning,” *The International Journal of Robotics Research*, vol. 36, no. 13-14, pp. 1595–1618, 2017.

[21] C. Wu, J. Zhang, O. Sener, B. Selman, S. Savarese, and A. Saxena, “Watch-n-patch: unsupervised learning of actions and relations,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 2, pp. 467–481, 2018.

[22] B. Hayes and J. A. Shah, “Interpretable models for fast activity recognition and anomaly explanation during collaborative robotics tasks,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 6586–6593.

[23] B. Hayes and B. Scassellati, “Autonomously constructing hierarchical task networks for planning and human-robot collaboration,” in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 5469–5476.