Toward Robust and Generalizable Prediction of Human Motion via Predictor Ensembles

Przemyslaw A. Lasota and Julie A. Shah

Abstract— **Many applications of human-robot interaction (HRI) necessitate close physical collaboration. Accurate prediction of human intent can be utilized to allow robots to select actions and motions that are safer and more efficient. While a variety of human motion prediction approaches have been developed, they are often designed for specific types of tasks or motions, and thus do not generalize well. Consequently, it is not always obvious what method is appropriate for a given task, making human motion prediction difficult to implement in practice. Toward addressing this problem, we introduce the concept of human motion prediction ensembles, where optimal combinations of prediction approaches are generated by learning from task data. We argue that by learning what predictors and parameters work best directly from task data, we can generate a motion prediction technique that can achieve higher performance than individual methods, while also reducing the implantation overhead for the user. In this work, we describe a preliminary implementation of this concept, which combines a goal-based prediction method with a velocity-based motion projection method. We evaluate the performance of the combined predictor against that of the individual methods in terms of accuracy of prediction of human position over a range of look-ahead time values. The results indicate that the combined predictor, which is assembled based on learning from task data, outperforms the individual goal-based and velocity-based methods by 17.2% and 15.9%, respectively. We also describe avenues for future work, discussing possible ways of extending this approach to accommodate additional prediction methods.**

I. Introduction

A large variety of fields and applications stand to benefit from human-robot interaction and collaboration. In the recent years, there has been a significant push toward introducing robots on factory floors [1], in homes [2], and even as assistants on board the International Space Station [3]. In these and other applications, close-proximity physical interaction is often needed in order for robots to effectively collaborate with people. Consequently, there is a need for the development of techniques and methods that support safe and efficient physical human-robot interaction.

One way in which such interaction can be achieved is through robot adaptation based on prediction of human motion. By being able to anticipate where the person might be reaching or walking toward next, a robot can choose actions or adjust its motions such that potential motion conflicts are avoided. Prior work has shown that a robot that utilizes a human-aware motion planner, which avoids moving through locations of upcoming human occupancy, leads to more efficient teamwork, increased satisfaction with the robot, and higher perceived safety and comfort [4]. Being able to correctly predict where a person will move to next is a key component of such a motion planning approach, motivating the development of accurate, real-time prediction of human motion.

Przemyslaw A. Lasota (email: plasota@mit.edu) and Julie A. Shah are with the Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts.

As discussed in the following section, a variety of human motion prediction approaches have been recently developed. These include goal-based methods and prediction based on studying natural human motion. A majority of these approaches, however, are designed and tuned for specific types of motions or tasks, and thus would not necessarily generalize well to prediction in other domains. Predictors based on action models would not work well in loosely structured tasks, for example. This creates a barrier for utilizing human motion prediction, as it might not always be clear what approach is best suited for a specific task. In fact, it might even be possible that no single technique works well, and that a combination of approaches is needed.

In order to address this drawback, we envision a data-driven approach to human motion prediction that, based on a variety of data encoding how the person moves in the shared workspace and how he or she performs the tasks, will automatically select the optimal ensemble of prediction methods to accurately predict human occupancy at various future timeframes. Such a technique would allow for robust and generalizable prediction of human motion.

In this paper, we present the first iteration of this concept by considering two motion prediction approaches to form a combined predictor. After discussing relevant prior work in the field in Section II, we describe the method by which task data is utilized to train the prediction models and fuse them to form the combined predictor in Section III. We then discuss the performance of the combined predictor in Section IV, showing that the presented approach led to lower prediction error than both of the two individual motion prediction techniques alone. Finally, we discuss the implications of our findings and plans for future work in Section V.

II. Related Work

Prior work in the field of human motion prediction can be classified into two categories: those based on goal intent and on those utilizing motion characteristics. The former category relies on predicting the target or goal that a person is reaching or walking toward and then utilizing an appropriate motion model to predict how the person will move in transit to that goal.

One recent example of this approach utilizes a time series analysis in which multivariate Gaussian distributions over tracked degrees of freedom of the human arm are computed for each time step of the motion. The learned models are used to perform Bayesian classification on the initial stages of motion to predict what goal location the person is reaching toward [5]. In another goal-based motion prediction approach, Gaussian Mixture Models (GMMs) are trained for each reaching goal position of a particular task and Gaussian Mixture Regressions (GMRs) are used to generate representative reaching motions. Based on observation of the beginning segment of a new reaching motion and the computed GMMs and GMRs, the framework calculates the likelihood of human occupancy of the shared workspace [6]. Goal-based human motion prediction can also be applied to larger motions as well, such as prediction of walking motion. In one example of such work, Growing Hidden Markov Models (GHMMs) and the social forces method are used to infer goals from partial trajectories and to then predict the path the person will take toward that goal [7].

The second major category of human motion prediction mentioned above focuses on observing and analyzing how people move and plan natural paths without predicting specific goal locations. In one such approach, motion capture data is used to encode skeletal motion patterns as Hidden Markov Models with the goal of encoding likely transitions between motion patterns [8]. In a different approach, based on the principle of maximum entropy, features such as amount of time needed to reach a goal, acceleration profiles, walking velocity, and collision avoidance behavior are considered. The authors investigated which of these features of walking motion could be utilized to learn how to characterize and predict typical human walking behavior [9]. Finally, in [10], people were instructed to walk toward several target locations in a room while their position and head orientation were recorded. It was shown that the head orientation and body velocity normalized by height can signal a person will turn prior to the physical turn itself.

While many of the approaches mentioned above are capable of accurate prediction of motion, they are often developed with a certain class of tasks in mind, and thus might not generalize well to other scenarios. The algorithms and techniques used also require careful tuning of various model parameters to achieve accurate prediction. We aim to address these drawbacks by utilizing the data from a given task to learn optimal parameters and form a combination of predictors.

III. Method

In this first step toward our concept of an optimal ensemble of human motion predictors, we present a framework for the automatic tuning and synthesis of two motion prediction approaches. Namely, we combine the goal-based technique developed by Pérez-D'Arpino and Shah mentioned earlier [5] with a motion-characteristic method based on projecting the person's current motion into the future by estimating velocity.

To test our method, we utilize data from the study of anticipatory indicators of human walking motion by Unhelkar and Shah mentioned above, as the goal-based method presented in [5] was also utilized on this data in [10]. The data set consists of 80 human demonstrations of people walking toward four target goals from a common starting location. The demonstrations are evenly distributed among each of the goals, forming four *motion classes*. The mean duration of the walking motions is 4.5s.

A. Time Series Classification Method

The first predictor used in our framework, described in [5], predicts intended goal locations based on early stages of motion by calculating the maximum likelihood that a given partial trajectory belongs to one of the motion classes in the training set. This calculation is performed at each time step of the input trajectory. The method also computes and stores the mean trajectory of each motion class based on the training data.

In order to convert the goal-based output of this method (what goal location is the person walking toward) to a prediction of where the person will be at a specific amount of time in the future, we utilize both the predicted motion class and the mean trajectory. First we retrieve the mean trajectory of the predicted motion class at the given time step and find a representative point in this mean trajectory, $\vec{x}_{rep}$. This is done by find the point closest to the person's current position, $\vec{x}_{curr}$. For a given look-ahead time duration Δt_{ahead}, we then find the predicted position $\vec{x}_{pred}$ by selecting the point in the mean trajectory that occurs Δt_{ahead} after $\vec{x}_{rep}$.

B. Velocity-Based Method

The second method of prediction used in our framework is based on projecting the person's current motion by computing an estimate of the person's velocity. We utilize the Savitzky-Golay Filter [11] to smooth the position data and compute estimates of velocity.

The method works by fitting low-degree polynomials to successive sets of points, and thus does not require the entire trajectory to perform the smoothing. Furthermore, if the position data is sampled at a uniform rate, there exists an analytical solution to the least-squares fit that can be represented by a set of coefficients. A convolution of these coefficients with the successive position data produces the smoothed signal, making the on-line smoothing computationally efficient. The sequential smoothing and computational efficiency make this particular filtering method ideal for real-time prediction of human motion.

Once the estimate of velocity, $\vec{\dot{x}}_{est}$, is obtained, we compute the predicted position $\vec{x}_{pred}$ by assuming the person will maintain the same velocity for the time period Δt_{ahead}:

$$\vec{x}_{pred} = \vec{x}_{curr} + \vec{\dot{x}}_{est} \cdot \Delta t_{ahead}$$

C. Training and Combining Predictors

The two motion prediction methods described above are synthesized into a combined predictor through a three step process. First a subset of the data, $\mathcal{D}_{train}$, is used to find the optimal parameters for each individual prediction method. In the case of the time-series classification method, this involves selecting the subset of degrees of freedom that leads to most accurate prediction results. For the velocity-based prediction, this means choosing the best moving window size and polynomial order for the Savitzky-Golay Filter.

The second step of forming the combined predictor involves evaluating each prediction method for a set of desired look-ahead time values based on the computed optimal parameters. This evaluation is done on a second subset of the data, $\mathcal{D}_{modelSelection}$. The look-ahead time values considered were 0.05s to 1s at increments of 0.05s. For each time step of each sample trajectory in $\mathcal{D}_{modelSelection}$, each prediction method is used to compute predicted positions at each look-ahead time value. The predictions are then compared to ground truth values, and the mean errors are recorded.

The final step of the process involves using the results of the above evaluation to decide which prediction methods should be used. For each look-ahead time value considered, the best-performing predictor from the two methods is selected for use in the combined predictor. In other words, the combined predictor simply selects the best performing method for a specific look-ahead time when making a prediction.

IV. Results and Discussion

Once the combined predictor was formed as described in the previous section, we evaluated its performance, using the two individual prediction methods as baselines. The final subset of the data $\mathcal{D}_{test}$ was used for this purpose. The mean prediction errors as a function of look-ahead time, Δt_{ahead}, are plotted in Figure 1.

From this plot, one can see that the velocity-based prediction method performs very well at low values of Δt_{ahead}, but the performance degrades quite rapidly as Δt_{ahead} increases. This makes sense, since at larger look-ahead time values, the person might change directions during the time interval in question, resulting in a poor prediction and higher mean error.

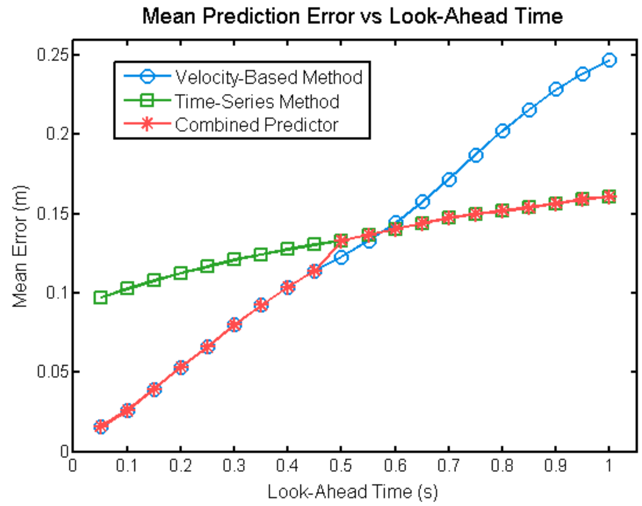

Fig. 1. Mean prediction error for the combined predictor and the two individual prediction methods (lower is better).

The time series method also exhibits the trend of decreased performance at higher values of Δt_{ahead}, but the mean error grows at a slower rate. Furthermore, the mean error at low values of Δt_{ahead} is higher for the time series method than for the velocity-based method, while this trend is reversed at high values of Δt_{ahead}. One possible explanation for the relatively worse performance at low values of Δt_{ahead} is the variance in the example trajectories. If the variance is high enough, the mean trajectory can be quite far from the the trajectory being predicted.

While this relatively high error at low values of Δt_{ahead} is not ideal, the time series method's performance degrades slowly with increasing values of Δt_{ahead}. This is likely due to the fact that this prediction method can correctly classify which goal the person is walking toward early on in the motion, and thus even if the person turns, the mean trajectory of the motion class of that goal will reflect this, and track the ground truth fairly accurately.

The most important result from Figure 1, however, is the performance of the combined predictor. Namely, one can see that the combined predictor closely tracks the best performing method for each value of Δt_{ahead}, correctly selecting the best-performing method for all values except 0.5s and 0.55s. As the model selection step was trained on a different subset of the data than the testing of the combined model by design, this discrepancy is expected. The extent of this discrepancy is dependent on the quality of the training data; namely, the ability of the combined predictor to track the best-performing method well relies on having training data that is representative of the types of motions the predictor is to be used for.

The superior performance of the combined predictor

TABLE I
MEAN PREDICTION ERRORS

Velocity-Based	Time Series	Combined
13.2cm	13.4cm	11.1cm

can be further displayed by calculating the mean errors accross all values of Δt_{ahead}. The result of these computations is summarized in Table I. The values shown in the table correspond to the combined method outperforming the individual goal-based and velocity-based methods by 17.2% and 15.9%, respectively.

V. CONCLUSION AND FUTURE WORK

Prior work has shown that prediction of human motion has the potential to improve the safety and efficiency of physical human-robot interaction. It can be challenging to utilize motion prediction, however, as the majority of prediction techniques are designed for specific types of tasks and motions and do not necessarily generalize well.

In this work, we introduced the concept of ensembles of human motion predictors as a technique that has the potential to alleviate this drawback. We showed that a simple implementation of this concept that combines two types of predictors, one goal-based and one based on estimated velocity, resulted in a combined predictor that outperformed the individual methods by 17.2% and 15.9%, respectively, in terms of mean error of the predicted position at various values of look-ahead time.

The fact that a two-method combination improved prediction performance by a substantial percentage motivates further work in developing the concept of ensembles of human motion predictors. Future work will involve not only implementing additional prediction methods to select from, but also improved techniques for synthesizing the ensemble. While selecting the best performing method at each value of look-ahead time has led to promising results for the combination of two predictors presented in this work, investigating other methods, such as prediction averaging, selection based on prediction confidence, or biasing the selection toward forming continuous predicted trajectories, for example, are worthwhile avenues of research that will be pursued in the future.

ACKNOWLEDGMENT

This work was supported by a NASA Space Technology Research Fellowship.

REFERENCES

[1] W. Knight, "Smart robots can now work right next to auto workers," *MIT Technology Review*, vol. 17, 2013.

[2] B. Graf, M. Hans, and R. D. Schraft, "Care-o-bot iidevelopment of a next generation robotic home assistant," *Autonomous robots*, vol. 16, no. 2, pp. 193–205, 2004.

[3] T. Fong, M. Micire, T. Morse, E. Park, C. Provencher, V. To, D. Wheeler, D. Mittman, R. J. Torres, and E. Smith, "Smart spheres: a telerobotic free-flyer for intravehicular activities in space," in *Proc. AIAA Space*, vol. 13, 2013.

[4] P. A. Lasota and J. A. Shah, "Analyzing the Effects of Human-Aware Motion Planning on Close-Proximity Human-Robot Collaboration," *Human Factors*, vol. 57, pp. 21–33, Jan. 2015.

[5] C. Pérez-D'Arpino and J. A. Shah, "Fast Target Prediction of Human Reaching Motion for Cooperative Human-Robot Manipulation Tasks using Time Series Classification," in *Proceedings of ICRA*, 2015.

[6] J. Mainprice and D. Berenson, "Human-robot collaborative manipulation planning using early prediction of human motion," in *Proceedings of IROS*, pp. 299–306, Ieee, Nov. 2013.

[7] J. Elfring, R. van de Molengraft, and M. Steinbuch, "Learning intentions for improved human motion prediction," *Robotics and Autonomous Systems*, vol. 62, pp. 591–602, Apr. 2014.

[8] W. Takano, H. Imagawa, and Y. Nakamura, "Prediction of human behaviors in the future through symbolic inference," in *Proceedings of ICRA*, pp. 1970–1975, 2011.

[9] M. Kuderer, H. Kretzschmar, C. Sprunk, and W. Burgard, "Feature-Based Prediction of Trajectories for Socially Compliant Navigation," in *Proceedings of RSS*, (Sydney, Australia), 2012.

[10] V. V. Unhelkar, P. Claudia, L. Stirling, and J. A. Shah, "Human-Robot Co-Navigation using Anticipatory Indicators of Human Walking Motion," in *Proceedings of ICRA*, 2015.

[11] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.