Prediction of Observed Motor Trajectories in a Simulated Environment[*]

Radovan Gregor, Igor Farkaš, Kristína Malinovská

Department of Applied Informatics, Comenius University Bratislava
{radovan.gregor, igor.farkas, kristina.malinovska}@fmph.uniba.sk

Abstract

Predicting the motion of a robotic arm is crucial for understanding observed behavior in human–robot interaction. In this study we present a method for predicting the future states of a motor trajectory and intended target of a robotic arm using an Echo State Network (ESN), exploiting the mirroring mechanism. In a simulated environment with two identical humanoid robots NICO – an actor and an observer – the observer learns to predict the actor's motor intentions by simulating future states of motion based on visual input that is converted to observer's motor representation. The results show that our ESN-based model achieves accurate short-term trajectory and target prediction with low computational cost, making it useful for testing in real-world applications.

1 Introduction

Human–robot interaction (HRI) is rapidly expanding in various fields such as healthcare, service industries, and collaborative robotics, which makes it essential to understand the computational mechanisms that drive robot perception and behavior (Thomaz et al., 2016). As robots move from isolated industrial environments to dynamic, shared human spaces, the demand for adaptable and autonomous systems — particularly humanoid robots — grows, due to their potential for more trustworthy interactions (Kok and Soh, 2020). A key requirement in such interactions is the ability to interpret and anticipate motor actions, which involves predicting the future trajectories of arms and hands, allowing the robots to proactively adjust their motor behavior.

To address this, we use an Echo State Network (ESN), a type of reservoir computing model, known for efficient time-series prediction (Lukoševičius and Jaeger, 2009). Unlike traditional recurrent neural networks, ESNs leverage a fixed, high-dimensional dynamic reservoir to efficiently capture temporal dependencies in time series data while requiring minimal training effort. We train the ESN on historical trajectory data, including joint positions, or end effector (EE) positions, to model the underlying motion dynamics. The proposed approach is evaluated in a physics-based simulation, combined with a biologically inspired mirroring approach grounded in the mirror neuron system.

[*]supported by projects APVV-21-0105 and VEGA 1/0373/23

2 Methods

2.1 Dataset for motion execution

To achieve our prediction of the intention of motor action, the robot first needs to acquire a motor repertoire (sensorimotor knowledge) to reach and grasp objects in its peripersonal space. The setup included NICO agent, a humanoid robot platform for multimodal interactions (Kerzel et al., 2017). We trained NICO in simulation to perform reach-and-grasp actions with target objects placed in a predefined reachable workspace. To include diverse reaching trajectories in the data set, we randomized object positions within the robot peripersonal space. To generate motion execution data we started with a discrete step-by-step trajectory planning. After defining the discrete steps, we recorded the execution of these trajectories by continuously capturing joint angles and end-effector positions throughout the motion. A constant-speed motion profile was implemented along with synchronization of all joint movements. By applying these techniques, we obtained a smooth, synchronized, and reproducible data set of the recorded end effector positions suitable for training the ESN for the prediction of action trajectory. [1]

2.2 Mirroring mechanism

The mirroring process, inspired by the direct matching hypothesis (DMH) lying behind the mirror neurons system, allows the observer to internally simulate the observed motion with the closest match in its own motor repertoire (Rizzolatti et al., 2001). According to DMH, observed visual representation of the actor (in Cartesian space) is mapped onto the corresponding observer's motor representation (in joint space). This mechanism enables the observer to treat the external motion as its own, facilitating intention prediction through embodied simulation. This representation tracks the EE motion relative to four reference points on the robot's body and represents the EE position as a vector of Euclidean distances. It enables effective mirroring by overcoming challenges related to viewpoint changes caused by rotation or translation.

[1]A short video of the recorded trajectory can be seen at https://www.youtube.com/watch?v=UEx6JAHBsZM

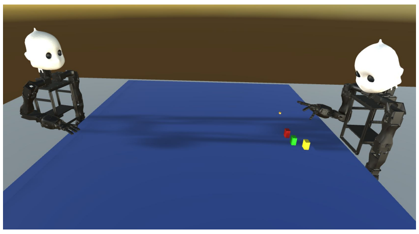

Fig. 1: Experimental setup with two NICO robots in simulation: The actor (right) and the observer (left).

2.3 Echo State Network (ESN)

The ESN is a type of recurrent neural network with a fixed, sparsely connected reservoir and a trainable readout layer. It is particularly suited for time series prediction due to its rich dynamic memory and fast training. The observer is equipped with ESN to acquire a motion-predicting capability. During training, the ESN receives a sequence of end-effector positions and learns to predict the next step in the sequence. In our final setup, the ESN is initialized with a predetermined number of time steps from the observed actor's motion trajectory. Then the predicted output of the ESN is fed back into its input to iteratively generate more predictions. The model performs recursive prediction over multiple future time steps.

3 Experiment

To evaluate the effectiveness of our trajectory prediction model, we performed experiments in a simulated environment using two NICO robots as in Figure 1. Both are equipped with the same sensorimotor capabilities, allowing the observer to link perceived actions with its own body schema. In each scene, three distinct objects are randomly placed within the actor's reachable workspace, and during each trial, the actor is tasked with reaching toward and grasping one of them. For each trial, only the initial segment of the arm's trajectory was made available to the observer's prediction model. This setup allows us to systematically measure the performance of the ESN to predict future effector positions. An example of the actual and predicted trajectory is shown in Figure 2. A distance-based vector was used to represent the probability of each object being a target, computed by evaluating the proximity of the predicted EE position to each candidate object over time. This probabilistic representation allowed the system to dynamically update its target inference as the trajectory unfolded rather than relying on a single deterministic prediction.

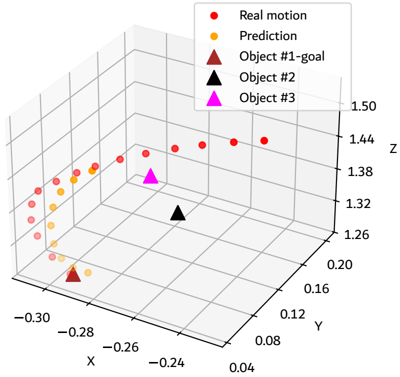

Fig. 2: Visualization of actual and predicted end-effector motion with positions of 3 objects in the scene.

4 Results

The ESN achieved high short-term trajectory prediction accuracy, maintaining low mean squared error at various positions of objects. The prediction accuracy of the target object exceeded 90% when more than 30% of the trajectory was available, indicating that the observer could infer intentions early during movement.[2] What is novel in our approach is the use of perspective-invariant distance-based representation of the end effector that enables to link visual representation to observer's motor representation. In future work, we plan to consider three grasp types (power, precision, and side) to enhance both action execution and intention inference. On the execution side, this requires modeling object affordances to guide grasp selection, and on the observation side, it enables intention prediction by leveraging both kinematic trajectory data and hand pre-shaping.

References

Kerzel, M. et al. (2017). NICO - Neuro-Inspired COmpanion: A developmental humanoid robot platform for multimodal interaction. In *Int. Symp. on Robot and Human Interactive Commun.*, pages 113–120.

Kok, B. and Soh, H. (2020). Trust in robots: Challenges and opportunities. *Current Robotics Reports*, 1:297–309.

Lukoševičius, M. and Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149.

Rizzolatti, G. et al. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2:661–670.

Thomaz, A. et al. (2016). *Computational Human-Robot Interaction*. Now Foundations and Trends.

[2]The illustration of the progression of prediction probabilities is shown at https://www.youtube.com/watch?v=YmzWaJD9FTU.