Analysis of the Limitations of an Experience Metric Space when Used in a Mobile Domestic Robot

Nathan Burke, Joe Saunders, Kerstin Dautenhahn and **Chrystopher Nehaniv**
Adaptive Systems Research Group
University of Hertfordshire
Hatfield, AL10 9AB
United Kingdom

Introduction and Background

If a robot is to succeed in integrating into the households of normal users, a robot must be able to easily learn new behaviors, modify existing behaviors and adapt to user preferences. However, it cannot be reasonably expected that users are robotics experts. By utilizing advancements in AI, it may be possible to improve the interaction between home users and a domestic robot. Teaching an AI using natural human teaching methods is not itself novel, and has been used in the past (A. Thomaz, 2006). The method used here is clicker style training, which is often used with dogs, and could provide an intuitive interface for untrained users to be able to train the system to adapt to their preferences. This is similar to that used in (B. Blumberg, et al, 2002), but with a physical agent rather than a virtual agent. In this paper we examine the possibility of utilizing the Interaction History Architecture (IHA) (N. A. Mirza, Nehaniv, Dautenhahn, & te Boekhorst, 2006) as a basis for learning both user preferences and simple control of the robot base. IHA has previously shown to be capable of producing desired behaviors from natural interactions in both humanoid (Broz, Nehaniv, Kose-Bagci, & Dautenhahn, 2012) and non-humanoid robots (N. A. Mirza et al., 2006).

The interaction history architecture combines a view of the environment with the ideas of an information metric proposed by Crutchfield (Crutchfield, 1990) and Shannon information theory (Shannon, 1948) in order build the metric space of experience. This view is constructed during the human robot interactions by the robotic agent and from its perspective. Due to this, we therefore consider it to be grounded (N. A. Mirza, Nehaniv, Dautenhahn, & Boekhorst, 2007). From this, we can then begin to construct a metric space of experience and compute the distance between time-separated events. We first define the conditional entropy, the measure of uncertainty for value X given Y as

$$H(X|Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log_2 p(x|y) \qquad (1)$$

and then similarly for $H(Y|X)$. The sensor data is first categorized into Q bins, and conditional entropies are calculated from the joint probability distributions of sensor values from T recent samples, a method developed by Mirza (N. A. Mirza, C. L. Nehaniv, K. Dautenhahn, & R. Boekhorst, 2005). Once the conditional entropy of the system has been calculated, we are able to calculate the information distance between two information sources with

$$d(X,Y) = H(X|Y) + H(Y|X) \qquad (2)$$

By building on the ideas of information distance, a metric space of experiences can be constructed. Beginning with the definition for an embodied agent as

> the temporally extended, dynamically constructed, individual sensorimotor history of an agent situated and acting in its environment, including the social environment, that shapes current and future action (N. A. Mirza et al., 2007)

we can define an experience for an agent as a collection of sensorimotor information X for a time horizon h. Thus, to calculate the distance between two experiences of the same horizon length, we need to calculate the distance between their sensorimotor information. Defining $X_{t_0,h}$ to be the sensorimotor variable with temporal horizon h starting at time t_0 allows for the experiences of the robot to be defined for temporal horizon h as $E(t,h) = (X^1_{t,h}, \dots, X^N_{t,h})$ where $(X^1_{t,h}, \dots, X^N_{t,h})$ is the set of all sensorimotor variables available to the robot. Following this, we can define a distance metric for experiences of temporal horizon h as

$$D(E,E') = \sum_{k=1}^{N} d(X_{t,h}^{k}, X_{t',h}^{k}) \quad (3)$$

where $E = E(t,h)$ and $E(t',h)$ are two experiences of the robot and d is the information distance (N. A. Mirza et al., 2006). The average if this is the final metric used in the comparison of experiences. By simplifying to a single distance value, we are able to quickly sort and select the experience closest to our current experience. This method was shown to be generally effective in distinguishing between experiences in (N. A. Mirza et al., 2005).

System Set-Up

We explore an adaptation of the humanoid version of IHA for its usefulness in a domestic robot. The robot used is Sunflower 1-1, a custom robot built on a Pioneer 3dx base and the experiment was performed in a home environment (Figure 1). Sunflower uses ROS Groovy (ROS, 2014) and p2os as its base control stack, with a custom ROS controller acting as a central control point.

Sunflower relies primarily on its sonar sensors, in addition to a contrived sensor that measures the distance and angle from the robot to the experimenter. The reward system includes a number of individual reward systems, hitherto referred to as motivations. The first motivation is a manual feedback mechanism designed to replicate the abilities of simple dog training. It consists of a three button 'clicker interface' on a helper computer that can be used to increase, decrease, and zero the reward manually assigned to the current behavior. The next motivation has been named *self-preservation* as it only acts to prevent the robot from running into objects. This motivation supplies reward values in the range [-1, 0], with the reward decreasing linearly as the robot approaches closer than 0.5m from an obstacle. Finally, there is a 'closeness' motivation that works similar to, and in opposition of the 'self-preservation' motivation. This motivation seeks to reward the robot for being near the experimenter, with a reward that continues to increase from 0 to 1 as the robot approaches within 1m of the experimenter, reaching $R_{MAX} = 1$ at 0.5m.

While individual motivations are normalized to the range $[-1, 1]$ the aggregate value is not. Each motivation carries a w*eight* modifier. This weight is constant for each motivation type, and aggregate reward value can be then defined as

$$R_{all} = \sum_{m \in M} m_w m_R \quad (4)$$

This allows for total rewards of greater than 1 and less than -1, a departure from previous experiments with IHA.

The robot is programmed with a number of basic actions that it can execute at any time step. For this experiment, the action set was restricted to only those dealing with the *base* (turning, forward/back), and the *idle* action (no action). During the experiment, the robot is allowed to freely roam about the room, with the experimenter using the clicker interface to reward the robot when it performs a desired action (i.e. turning towards the experimenter). This is added to the reward from the automatic motivations to arrive at the final reward for that experience. The experiment was begun with an empty interaction history, and as such, Sunflower began the random movements that approximate body babbling. This is normal for IHA, and is expected until sufficient similar past experiences have been recorded.

Figure 1: Sunflower 1.1 in the study environment

Results and Conclusions

The chance of a random action being selected is expected to decrease over time and within the first one to three minutes actions are expected to be chosen primarily based upon past experiences. This was not the observed result, and an analysis of the data made clear that there was an incompatibility in the sensor configuration and the IHA distance calculations which prevented IHA from properly distinguishing past events.

While IHA can be useful for detecting past events that are similar to the current environment, we have found that there are limitations that must be taken into account when deciding on sensor configurations. The reliance on entropy means that any sensors that are stable over multiple time windows will always be computed as identical, regardless of the underlying values. Additionally, IHA does not maintain the time-series of the data, which causes rising and falling values within the same sensor to be computed as similar as well. The net result of these is that all 'idle' actions are detected as the same, and without the time-series information from the single direction sensor, the robot could not distinguish between turn left and turn right actions. This second issue can be solved by utilizing multiple sensors for left/right based rewards, as was done in an earlier T-Junction task (N. A. Mirza, 2008). The problem of entropy in the sensors is more difficult, and may require including a different distance metric in the calculations. More work will need to be done to determine if IHA is useful in navigation style tasks.

References

Blumberg, B., Downie, M., Ivanov, Y., Berlin, M., Johnson, M., Tomlinson, B. 2002. *Integrated learning for interactive synthetic characters*. In: Proceedings of the ACM SIGGRAPH 2002

Broz, F., Nehaniv, C. L., Kose-Bagci, H., & Dautenhahn, K. 2012. Interaction Histories and Short Term Memory: Enactive Development of Turn-taking Behaviors in a Childlike Humanoid Robot. *arXiv preprint arXiv:1202.5600*.

Crutchfield, J. P. 1990. Information and its metric *Nonlinear Structures in Physical Systems* 119-130 Springer.

Mirza, N. A. 2008. *Grounded Sensorimotor Interaction Histories for Ontogenetic Development in Robots.* (PhD Thesis), University of Hertfordshire, Hatfield, Hertfordshire.

Mirza, N. A., Nehaniv, C. L., Dautenhahn, K., & Boekhorst, R. 2005. *Using sensory-motor phase-plots to characterise robot-environment interactions.* Paper presented at the Computational Intelligence in Robotics and Automation, 2005. Proceedings. 2005 IEEE International Symposium on.

Mirza, N. A., Nehaniv, C. L., Dautenhahn, K., & Boekhorst, R. T. 2007. *Grounded sensorimotor interaction histories in an information theoretic metric space for robot ontogeny*. Adaptive Behavior, 15(2), 167-187

Mirza, N. A., Nehaniv, C. L., Dautenhahn, K., & te Boekhorst, R. 2006. *Interaction histories: From experience to action and back again.* Paper presented at the Proceedings of the 5th IEEE International Conference on Development and Learning.

ROS. 2014. *Robot Operating System* http://www.ros.org Retrieved April 12, 2014

Shannon, C. E. 1948. A mathematical theory of communication. *Bell Systems Technical Journal, 27*, 379-423, 623-656.

Thomaz, A. and Breazeal, C., 2006 *Teachable Characters: User Studies, Design Principles, and Learning Performance*. In Intelligent Virtual Agents. vol. 4133, J. Gratch, M. Young, R. Aylett, D. Ballin, and P. Olivier, Eds., ed: Springer Berlin Heidelberg, 2006 395-406