Identifying Qualitatively Different Outcomes of Actions: Gaining Autonomy Through Learning

Tim Oates, Matthew D. Schmill and Paul R. Cohen
Computer Science Building
University of Massachusetts
Amherst, MA 01003-4610
{oates,schmill,cohen}@cs.umass.edu

If robotic agents are to act autonomously they must have the ability to construct and reason about models of their physical environment. In all but the simplest, static domains, such models must represent the dynamics of environmental change. For example, because the effects of actions are not instantaneous, planning to achieve goals requires knowledge of how the robot's actions affect the state of the world over time. The traditional approach of hand-coding this knowledge is often quite difficult, especially for robotic agents with rich sensing abilities that exist in dynamic and uncertain environments. Ideally, agents would acquire knowledge of their environment autonomously, and then use this knowledge to act autonomously.

This paper presents an unsupervised method for learning models of environmental dynamics based on clustering multivariate time series. Experiments with a Pioneer-1 mobile robot demonstrate the utility of the method and show that the models acquired by the robot correlate surprisingly well with human models of the environment.

Individual time series are obtained by recording the output of a subset of an agent's sensors. We call these time series *experiences*. An example of a sensor subset on the Pioneer-1 robot is its array of seven sonars. Each sonar returns the distance to the closest object in the direction that it points. Recording of time series is usually triggered by events, such as the initiation of a particular action. Each time a given event occurs, the time series that was recorded is added to a bucket associated with that event. Once a sufficient number of experiences are recorded, clusters can be formed. Clustering requires a measure of similarity between multivariate time series. One such measure that is particularly appropriate for this problem is Dynamic Time Warping (DTW) [4]. Each cluster can then be represented by a prototype, either the cluster centroid or an average of its members.

Cluster prototypes formed in this manner are useful for a variety of purposes. If the event driving the collection of time series was the initiation of an action, cluster prototypes correspond to qualitatively different outcomes of engaging in that action. As such, they can be used for off-line planning and for on-line prediction by finding the best partial match among the prototypes to current sensor readings.

We are interested in the results of clustering for two key reasons. First, for the purposes of planning, we would like clusters to map to action outcomes, so that each cluster prototype can serve as the basis for an operator model. Second, we would like agents to be able to acquire a believable ontology of activity. That is, we would like our agents to be able to differentiate actions as a human would so that their representations of outcome are in accordance with our own. As such, our primary means of evaluating cluster quality is to compare the clusters generated by our automated system against clusters generated manually by the experimenter who designed the experiences they comprise.

The following experiment was used to explore these issues. Data were collected for 4 sets of experiences: 102 experiences with the robot moving in a straight line while collecting data from the velocity encoders, break beams, and gripper bumper (which we will call the *tactile* sensors), 102 move experiences collecting data from the Pioneer's vision system, including the X and Y location, area, and distance to a single visible object being tracked (which we will call the *visual* sensors), 50 experiences with the robot turning in place collecting tactile data, and 50 turn experiences collecting visual data. In each experience, the robot attempted to move or turn for a duration between 2 and 8 seconds in the laboratory environment. Visible objects and objects that impeded or obstructed the robot's path were present in many of the trials.

We evaluate the clusters generated by DTW and ag-

	#	t_t	$t_t \wedge t_e$	%	$\neg t_t$	$\neg t_t \wedge \neg t_e$	%	Agree	Disagree	%
Move visual	-5	876	720	**82.2**	4275	4125	**96.4**	4845	306	**94.0**
Move tactile	-7	443	378	**85.3**	4708	4468	**95.0**	4846	305	**94.0**
Turn visual	-5	262	262	**100.0**	599	571	**95.3**	833	28	**96.7**
Turn tactile	-6	163	163	**100.0**	698	593	**85.0**	756	105	**87.8**

Figure 1: Accordance statistics for automated clustering against the hand built clustering.

glomerative clustering with a 2×2 contingency table called an *accordance table*. Consider the following table:

	t_e	$\neg t_e$
t_t	n_1	n_2
$\neg t_t$	n_3	n_4

We calculate the cells of this table by considering all pairs of experiences e_j and e_k, and their relationships in the target (hand-built) and evaluation (DTW) clusterings. If e_j and e_k reside in the same cluster in the target clustering (denoted by t_t), and e_j and e_k also reside in the same cluster in the evaluation clustering (denoted by t_e), then cell n_1 is incremented. The other cells of the table are incremented when either the target or evaluation clusterings places the experiences in different clusters ($\neg t_t$ and $\neg t_e$, respectively).

Cells n_1 and n_4 of this table represent the number of experience pairs in which the clustering algorithms are in accordance. We call $n_1 + n_4$ the number of *agreements* and $n_2 + n_3$ the number of *disagreements*. The *accordance ratios* that we are interested in are $\frac{n_1}{n_1+n_2}$, accordance with respect to t_t, and $\frac{n_4}{n_3+n_4}$, accordance with respect to $\neg t_t$.

Table 1 shows the breakdown of accordance for the combination of dynamic time warping and agglomerative clustering versus the ideal clustering built by hand. The column labeled "#" indicates the difference between the number of hand-built and automated clusters. In each problem, the automated algorithm clustered more aggressively, resulting in fewer clusters. The columns that follow present the accordance ratios for experiences grouped together, apart, and the total number of agreements and disagreements.

The table shows very high levels of accordance. Ratios ranged from a minimum of 82.2% for experiences clustered together (t_t) in the move/visual set to 100% for experiences clustered together in the turn problems. For the turn problems, the aggressive clustering may account for the high t_t accuracy, causing slightly lower accuracy in the $\neg t_t$ case. The disparity in the number of clusters suggests that tuning the parameters of the clustering algorithm to produce more clusters might boost $\neg t_t$ accuracy while preserving the t_t accuracy.

Additional experiments explored methods for optimizing the clusters to meliorate ordering effects induced by the agglomerative clustering algorithm [1]. After applying this optimization technique to the clusters used to generate table 1, many of the errors in the t_t cases disappeared: accordance climbed to 91.9% or better in all cases except the $\neg t_t$ case of turn/tactile, which decreased to below 80%, which reflects the disparity between the number of clusters generated by our algorithm and the hand built clustering.

Current work is extending the approach described above in three ways. First, we are extending the autonomy of our system by utilizing cluster prototypes as bases for planning models, which will allow the Pioneer-1 agent to create basic action sequences to achieve sensorimotor goals [5]. Second, rather than using each experience in its entirety, we are developing methods for identifying subsequences within the experiences that are relevant to the clustering process [2]. Finally, we are leveraging the relationship between DTW and HMM's to develop a method of clustering time series in which the output is a set of HMM's, one for each cluster [3].

References

[1] D. H. Fisher, L. Xu, and N. Zard. Ordering effects in clustering. In *Proceedings of the 9th Annual Conference on Machine Learning*, pages 163–168, 1992.

[2] Tim Oates. Identifying distinctive subsequences in multivariate time series by clustering. In *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining*, pages 322–326, 1999.

[3] Tim Oates, Laura Firoiu, and Paul R. Cohen. Clustering time series with hidden markov models and dynamic time warping. In *working notes of the IJCAI-99 workshop on Neural, Symbolic and Reinforcement Learning Methods for Sequence Learning*, pages 17–21, 1999.

[4] David Sankoff and Joseph B. Kruskal, editors. *Time Warps, String Edits, and Macromolecules: Theory and Practice of Sequence Comparisons.* Addison-Wesley Publishing Company, Reading, MA, 1983.

[5] Matthew D. Schmill, Tim Oates, and Paul R. Cohen. Learning planning operators in real-world, partially observable environments. To appear in *Proceedings of the Fifth International Conference on Artificial Intelligence Planning and Scheduling*, 2000.