Recognizing Behaviors and the Intentional State of the Participants

Wesley Kerr (wkerr@cs.arizona.edu)
Department of Computer Science
1040 E. 4th Street
Tucson, AZ 85721 USA

Keywords: activity recognition; intention recognition; time series;

Introduction

Psychological research has demonstrated that subjects shown animations consisting of nothing more than simple geometric shapes perceive the shapes as being alive, having goals and intentions, and even engaging in social activities such as chasing and evading one another (Blythe, Todd, & Miller, 1999; Heider & Simmel, 1944). While the subjects could not directly perceive affective state, motor commands, or the beliefs and intentions of the actors in the animations, they still used intentional language to describe the moving shapes. For example, subjects in the Heider and Simmel (1944) study consistently labeled the larger triangle, shown in Figure 1, as a bully who harassed the smaller triangle and circle.

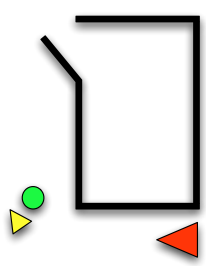

Figure 1: Single frame from an animations similar to the original Heider and Simmel animation.

When subjects ascribe intentions to geometric primitives like those shown in Heider and Simmel's research (see Figure 1), which information guides the process? Blythe et al. (1999) showed that the motion of the actors in animations is sufficient to classify the activities that occur in the animations. The system generated to perform classification even outperformed human subjects on the same task.

Blythe's system mapped patterns of motion onto class labels for intentional states, which isn't quite the same as knowing anything about intentional states. One of Heider and Simmel's subjects described the larger triangle in Figure 1 as "blinded by rage and frustration." Blythe's system couldn't come up with such a description. An agent that classifies episodes by patterns of motion knows about patterns of motion, not about rage and frustration, even if these words are provided as episode labels. So how might an agent infer affective states?

In both the Heider and Simmel animations and the animations developed by Blythe et al., subjects can only observe a subset of the features that are available, i.e. positions, velocities, sizes, colors, etc. The subjects cannot directly perceive the affective state, motor commands, and the beliefs and intentions of the actors in the animations. Yet they infer affective states and describe them with intentional language. We think humans infer affective states given non-affective observables such as positions and velocities by calling on their own affective experiences. Observables cue, or cause to be retrieved from memory, schemas that include learned affective components, which are inferred or "filled in" as interpretations of patterns of motion or other non-affective observables.

In this dissertation, we present representations and algorithms that enable an artificial agent to correctly recognize other agents' activities by observing their behavior. In addition, we demonstrate that if the artificial agent learns about the activities through participation, where it has access to its own internal affective state, motor commands, etc., it can then infer the unobservable affective state of other agents.

Activity Recognition

We begin with definitions: An *episode* is a collection of *intervals*. Each interval is a tuple containing a proposition and the times at which the proposition becomes true and false. A proposition can become true (and false) multiple times within an episode; each of these instances is represented as a separate interval. Each episode is given a class label and is a single example of an activity. In the activity recognition task we are given a collection of episodes for training, and then tested on episodes that were not part of the training set.

We assume that different examples of one activity share patterns of intervals. More colloquially, the intervals in similar episodes tell the same story with minor variations. Thus, one may classify episodes by their constituent patterns of intervals. This is not the only way to do it: A cleaning agent might classify a cleaning episode by the objects it interacts with, such as pots and pans, rather than what was done with the pots and pans. But our focus here is classifying episodes by patterns of activities, represented by intervals.

Episodes and intervals have different durations, start times, end times, and constituent propositions, so our representation of episodes must be able to accommodate and generalize over these variations. For example, the activity "capture" involves one agent chasing another agent until the second agent is cornered or held in a single place. The participants might be a prisoner and a guard or some other pair of agents, and the amount of time spent chasing can vary from minutes to hours, but all episodes share the same common pattern: One actor chasing another until the other agent is cornered or caught.

Relationships between intervals can be described by *Allen relations* (Allen, 1983). Allen recognized that, after eliminating symmetries, there are only seven possible relationships between two intervals. Allen relations are qualitative in the sense that they represent the temporal order of events, specifically, the beginnings and endings of intervals, but not the durations of intervals.

Our episode representation, which we call a *qualitative sequence*, is a sequence of Allen relations between intervals in the episode. We construct the sequence by combining the Allen relations between *all* of the pairs of intervals in the order in which the Allen relation completes. An illustrative episode and the resulting qualitative sequence is shown in Table 1. The letters **A, B** and **C** denote propositions, and an assertion such as (**C** 1 3) means that proposition **C** was true in the interval [1,3].

Intervals	Sequence
	(**C** *meets* **A**)
(**C** 1 3)	(**C** *before* **B**)
(**A** 3 6)	(**A** *overlaps* **B**)
(**B** 4 9)	(**C** *before* **C**)
(**C** 6 10)	(**A** *meets* **C**)
	(**B** *overlaps* **C**)

Table 1: An episode comprising four intervals and the corresponding qualitative sequence.

Episodes are first converted into qualitative sequences of Allen relations and learning is done with these sequences. Let $\mathcal{S} = \{S_1, S_2, \ldots, S_k\}$ be a set of qualitative sequences with the same activity label. We define the *signature* of the activity label, $\mathcal{S}_c$, as an ordered sequence of *weighted* Allen relations. (The only difference between a signature and a qualitative sequence is these weights.) We select a sequence at random from $\mathcal{S}$ to serve as the initial signature, $\mathcal{S}_c$, and initialize all of its weights to 1. After this, $\mathcal{S}_c$ is updated by combining it with the other sequences in $\mathcal{S}$, processed one at a time.

Two problems are solved during the processing of the sequences in $\mathcal{S}$. First, the sequences are not identical, so $\mathcal{S}_c$ must be constructed to represent the most frequent relations in the sequences. The weights in $\mathcal{S}_c$ are used for this purpose. Second, because a relation can appear more than once in a sequence $\mathcal{S}_i$, there can be more than one way to align $\mathcal{S}_i$ with $\mathcal{S}_c$. These problems are related because the frequencies of relations in $\mathcal{S}_c$ depend on how sequences are successively aligned with it.

Updating the signature $\mathcal{S}_c$ with a sequence $\mathcal{S}_i$ occurs in two phases. In the first phase, $\mathcal{S}_i$ is optimally aligned with $\mathcal{S}_c$ using the Needleman-Wunsch global sequence alignment algorithm (Needleman & Wunsch, 1970). The alignment algorithm penalizes candidate alignments for relations in $\mathcal{S}_c$ that are not matched by relations in $\mathcal{S}_i$, and rewards matches. These penalties and rewards are functions of the weights stored with the signature. In the second phase, the weights in the signature $\mathcal{S}_c$ are updated. If a relation in $\mathcal{S}_i$ is aligned with one from $\mathcal{S}_c$, then the weight of this relation is incremented by one. Otherwise the weight of the relation is initialized to one and it is inserted into $\mathcal{S}_c$ at the location selected by the alignment algorithm.

The signatures function as classifiers as follows. Recall that $\mathcal{S} = \{S_1, \ldots, S_k\}$ is a set of qualitative sequences with the same activity label; for example, all the sequences in $\mathcal{S}$ might be examples of *jump over*. Now suppose we have N sets of qualitative sequences, $\Sigma = \{\mathcal{S}^1, \mathcal{S}^2, \ldots, \mathcal{S}^N\}$ each of which has a different activity label, and its own signature. A novel, unlabeled sequence matches each signature to some degree, determined by aligning it with each signature, as described earlier. The novel sequence is given the activity label that corresponds to the signature it matches best.

Inferring Hidden State

Episodes have observable and unobservable propositions depending on which agent is doing the observing. For example, when $agent_1$ is chasing $agent_2$, $agent_1$ observes all of the propositions pertaining to its motor commands, emotional state, and intentional state, but when $agent_1$ observes $agent_3$ chasing $agent_2$, $agent_1$ cannot perceive the motor commands, emotional state, and intentional states of $agent_2$ nor $agent_3$.

By *hidden relations* we mean relations that include one or more propositions that are not directly observable in the behavior of other agents, and so must be inferred. Our approach to inferring hidden relations is to have agents learn signatures of their own behaviors, in which these relations are *not* hidden. Then, when an agent observes another's behavior, it matches the observable relations to signatures of its own behavior, and uses these to infer unobservable relations in other's behavior.

In general, sequences can contain many hidden relations. The most frequent are the most likely when observing other agents. Therefore, our agent selects the most frequently occurring hidden relations to be the inferred hidden state.

References

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, *26*(11), 832–843.

Blythe, P. W., Todd, P. M., & Miller, G. F. (1999). How motion reveals intention: Categorizing social interactions. In *Simple heuristics that make us smart.* Oxford University Press, USA.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243.

Needleman, S. B., & Wunsch, C. D. (1970, March). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, *48*(3), 443-453.