2018 17th IEEE International Conference on Machine Learning and Applications

Modelling Human Understanding of Thematic Roles with Motion Heuristics

Soumitra Samanta
University of Liverpool
Liverpool, UK
email: soumitramath39@gmail.com

Franklin Chang
Kobe City University of Foreign Studies
Kobe, Japan
email: chang.franklin@gmail.com

***Abstract*—Humans can understand event roles like agent and patient within videos of simple shapes moving around just using simple motion heuristics. As existing computational systems do not directly address human understanding of these events, we develop the first computational model that can simulate human performance in these tasks. We develop an approach heuristic that can simulate how human recognition of chasing is influenced by the angle that the *chaser* uses to approach the *chasee*. We also created a causality heuristic that captures human sensitivity to contact between the *pusher* and the *pushee*, as well as a delay in launching. Careful modelling of psychological studies of infants and adults behaviour can yield insights that may enhance computational systems for action understanding.**

***Keywords*-action understanding, thematic role, chasing, pushing**

I. Introduction

An important sub-problem of visual action recognition [1] is identifying the roles of individuals in interactions with multiple individuals when viewed from a distance camera (e.g., overhead CCTV), where body parts and detailed appearance-based features are difficult to extract. For example, Moreno and Poope who developed a system to identify the role of tagger and runner within simulated games of tags with circles representing individuals [2]. A common approach has been to develop heuristics that encode the relational motion between individuals and using those heuristics in various combinations to predict actions like chasing [3]. Since these cues are not highly appearance or perspective dependent, these kinds of systems should be able to generalize easily to novel real world scenes.

What is less well understood at the moment is how to develop appropriate motion heuristics that will encode the right kinds of relational features that can be combined in various ways to encode various types of actions. One avenue for examining this question is to model the processes that humans use to encode actions. This work has suggested that human have fairly complex relational features. For example, Gao et al. [4] had human participants identify a wolf circle out of an array of 4 sheep circles that all moved around randomly and participants could identify the wolf based on its angle of approach towards the sheep that it was chasing. This computation requires that the direct path between the wolf and the sheep is computed and then angle of the wolf's motion is computed relative to this direct path. Thus, humans predict the agent of chasing using relatively complex relational features that are not some simple combination of lower level features like velocity.

Another implication of psychological work is that many complex heuristics appear to be innate (e.g., infants understand chasing [5]). Leslie and Keeble [6] found that 6-month infants could distinguish causal and non-causal pushing actions. They were sensitive to whether the pusher made contact with the pushed object and also when there was a delay between the contact and the launch of the pushed object. This suggests that there is a relational heuristic, which is sensitive to immediate contact-based causal pushing, and this is not something that can be trivially trained with various non-relational motion features. Furthermore, the fact that this ability appears so early in development suggests that these features are not trained, but rather are part of the innate abilities of the human brain. In this work, we develop a computational system that uses heuristics that are similar to those in previous work, but we evaluate this system against human data to better understand the computational properties of these heuristics.

II. Developing heuristics to explain human action understanding

To develop a system that closely mirrors our understanding of thematic role action recognition, we focused on data from one study on chasing and another on causal pushing. The chasing study is a study by Gao et al. [4], which used a multiple objects tracking paradigm, where participants view scenes with multiple identical objects that are moving randomly. Pylyshyn and Storm [7] found that humans can track the identity of the objects as they move around, even though they were all identical in shape and colour, and they suggested that they had pointers that would attach to each object and record information associated with that object. Gao et al.'s [4] chasing study used 5 circles that moved around randomly, but one of the circles (the *wolf*) moved towards another circle (the *sheep*). The *wolf*'s angle of motion toward the *sheep* was called its *chasing subtlety* and they varied it from 0° to 150° in 30° increments. When the *chasing subtlety* was 0°, the *wolf* was moving directly towards the *sheep* in each frame and that made it easy to

978-1-5386-6805-4/18/$31.00 ©2018 IEEE
DOI 10.1109/ICMLA.2018.00117

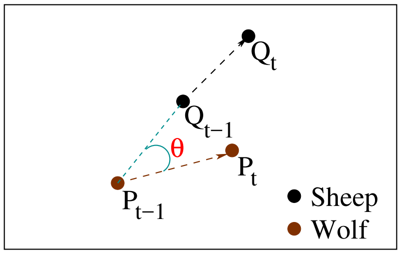

Figure 1. Chasing angle of motion or subtlety.

identify the *wolf*. When the *chasing subtlety* was higher, it became more difficult to identify the *wolf* and chasing accuracy became lower (human performance in Fig. 2). To examine this, we first generated chasing videos with various *chasing subtlety* and then created a heuristic that when applied to these videos could identify the *wolf* in a way that mirrored human accuracy.

A. Generating Chasing Events

The chasing videos had *four* identical circles. Most of the circles moved randomly during the entire video and this was done by changing the direction of motion randomly in every 5 frames. Two circles were specified as the *wolf* and the *sheep*. The *wolf* moved towards the *sheep* at a *chasing subtlety* angle that was specific for each video. For example, let at $(t-1)^{th}$ frame P_{t-1} and Q_{t-1} are the position of *wolf* and *sheep* respectively (Figure 1). If the *wolf* is chasing the *sheep* at subtlety θ, then at t^{th} frame the positions of the *wolf* P_t makes an angle θ at P_{t-1} with Q_{t-1}, i.e. $\angle P_tP_{t-1}Q_{t-1} = \theta$. The *sheep* always tried to keep maximum distance from the *wolf* by moving toward the direction $\overrightarrow{P_{t-1}Q_{t-1}}$. For our experiment, we have generated 100 videos for each different *chasing subtlety*(θ) of 0°, 30°, 60°, 90°, 120° and 150°. In the next subsection, we describe our heuristics for *wolf* and *sheep* detection in these chasing videos.

B. Heuristics for Identifying Chaser in Chasing Events

To identify chasing in these videos, we have computed an *approach heuristic* which was similar to the chasing subtlety used in the generation of the videos. In generation, subtlety was only computed from the *wolf* to the *sheep*, but in test, we do not know which circles are involved in the chasing, so the approach heuristic is computed between all pairs of circles separately in both directions (approach for A to B is not the same as approach for B to A). If chasing subtlety is computed on each frame, then there will be many false alarms as there are many frames where one circle is moving directly towards another circle accidentally as a product of random motion. Thus, we also assume that our heuristic is aggregated over time with a hysteresis parameter that allows us to identify consistent goal-directed chasing behavior.

To compute the approach heuristic, we applied a Kalman filter [8] based tracking algorithm so that the identity of each circle was maintained across the frames of the video. Let for each circle O_i, we had its velocity of motion V_i. We also could compute the direction of motion D_{ij} of i^{th} circle towards j^{th} circle. Using V_i and D_{ij}, we computed the *chasing subtlety* C_{ij} for circle O_i towards the circle O_j. To compute the approach heuristic A_{ij}, we combined the *chasing subtlety* C_{ij} for time t with the previous approach heuristic A_{ij} at time $t-1$ using a hysteresis (H) value. This was done so that *chaser* must show a consistent intention to follow the *chasee* for the angle of motion heuristic to become large. Since this heuristic is not symmetric, the heuristic was computed for each pair of circles O_i and O_j. The smallest overall heuristic value at the end of the trial was identified and the first index identified the *wolf* in the event. This algorithm was applied to all 100 videos for each *chasing subtlety* and proportion where the identified *wolf* matched the ground truth *wolf* (accuracy) was calculated.

C. Chasing Results

Fig. 2 shows the average accuracy for each level of *chasing subtlety* for the human data [4] and the model with different values of hysteresis (H = 0.9, 0.93, 0.95, 0.97, 0.99, 0.991, 0.995, and 0.999). The results show that the approach hysteresis has a non-linear interaction with chasing subtlety, and the closest match with the human data is a hysteresis (H) value of 0.97.

To evaluate the model against human behavior, our goal is to capture the way that human behavior changes with *chasing subtlety*. There is no gold-standard, because judgments of chasing are not categorical. Instead we ask whether the decline in chasing labeling in the model with hysteresis of 0.97 is related to the increase in chasing subtlety angle in a way that mirrors the way that these variables are related in humans. Therefore, we applied a regression to the percentage chasing with chasing subtlety and participant type crossed (human, model). The percent of videos labeled as chasing was negatively reduced by subtlety, $\beta = -0.5252$, $t(8) = -4.7$, $p < 0.002$. But there was no effect of type or interaction of type ($p = 0.8$) and subtlety ($p = 0.4$), which means that human and model data were similar.

This test demonstrated that it is possible to fit human *wolf* identification accuracy across a range of *chasing subtlety* by using a relational approach heuristic that aggregated information across frames to provide information that would support a judgment about the *wolf* in the scene.

III. Understanding Causal Pushing Events

The second study examines causal pushing events in the work by Leslie and Keeble [6] and Cohen and Oakes [9], who found that *six-month* old infants could recognize causality in pushing events, where the *pusher*'s contact with the *pushee* leads to an immediate launching of the *pushee*.

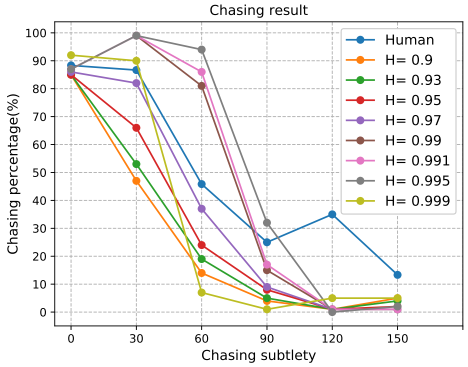

Figure 2. Chasing results with different *hysteresis*.

They distinguish this immediate contact event from events where launching is less causal due to a lack of contact or a delay in launching. This suggests that humans have an innate heuristic that supports understanding of causality in these events. We explore the nature of this heuristic by generating a range of pushing events and then test heuristics for identifying causal launching.

A. Generating Pushing Events

For pushing video, we start with *nine* randomly placed identical circular objects that are randomly moving with a constant velocity. After this random motion, the objects stop and the *pushing* object moves towards the *pushee* object. When they contact, the *pusher* stops and the *pushee* moves away from the *pusher*. After about 100 milliseconds, all of the objects resume random motion. To include different variation within a pushing we impose some constraints on the *pusher* and the *pushee* are as follows:

- *pushing subtlety or angle (PAN)*: the angle of the *pushee* movement with respect to the *pusher* after pushed – 0°, 30°, 60°, 90°. Lower angles indicate that the *pushee*'s movement matches the *pusher*'s approach and this supports a causal interpretation.
- *pushing delay (PDL)*: the delay between the *pusher* and the *pushee* movement during pushing – $three$ different 0, 10 and 20 (in frames). Less delay suggests that the *pusher*'s energy is directly transferred to the *pushee* and this supports a causal interpretation.
- *pushing distance (PDS)*: the distance between the *pusher* and the *pushee* during pushing – 0, 10 and 20 (in pixels). Lower distance means that the *pusher* and the *pushee* are more likely to contact and this helps to make the interaction appear more causal.

We created 100 videos for each combination crossing the levels of *pushing subtlety*, *pushing delay*, and *pushing distance*.

B. Heuristics for Identifying Pushers in Pushing events

In each pushing video, circular objects were tracked in the same manner as in the chasing videos. Let us consider $P_{i,t}$ be the position of the i^{th} object O_i at t^{th} frame. We calculate a pushing score for each pair of object O_i and O_j at each frame of the video based on their tracked position information. First, we calculate the stationary time of each object at each frame by calculating the frame wise displacement $d(O_{i,t-1}, O_{i,t})$ of that object. The stationary time (ST) of an object O_i at t^{th} frame ($ST_t(O_i)$) is defined as:

$$ST_t(O_i) = \begin{cases} ST_{t-1}(O_i) + 1 & \text{if } d(O_{i,t-1}, O_{i,t}) = 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where $d(O_{i,t-1}, O_{i,t})$ is positional euclidean distance of the object O_i between $(t-1)^{th}$ and t^{th} frame and $ST_0(O_i) = 0$. We define a *pushing delay score* ($PDLS_t(O_i)$) of an object O_i at t^{th} frame as

$$PDLS_t(O_i) = \exp(-ST_t(O_i)) \quad (2)$$

Similarly, we define a *pushing distance score* $PDSS_t(O_i, O_j)$ between the objects O_i and O_j is defined as:

$$PDSS_t(O_i, O_j)) = \exp(-d(O_{i,t}, O_{j,t})) \quad (3)$$

From Equations (2) and (3) it is clear that, for 100% pushing ($PDL = 0$ & $PDS = 0$) $PDLS_t(O_i)$ and $PDSS_t(O_i, O_j)$) gives maximum value and low value for other values of PDL & PDS.

In addition to $PDLS$ and $PDSS$, we propose two another terms for pushing action detection. In real world pushing action, the *pusher* moves faster than the *pushee* before the pushing happen and after pushing, *pusher* releases his force to the *pushee*. So, the average velocity of the *pusher* is greater than the same of the *pushee* just before the pushing happens and vise versa just after the pushing. Based on this observation, we propose two scores namely difference of average *speed before pushing* ($SBPS_t(O_i, O_j)$) and difference of average *speed after pushing* ($SAPS_t(O_i, O_j)$) between the two objects, O_i and O_j at t^{th} frame is defined as:

$$SBPS_t(O_i, O_j) = \exp\Big\{ -1/(\sum_{n=t}^{t-\tau} \|V_{i,n}\| - \sum_{n=t}^{t-\tau} \|V_{j,n}\|)\Big\} \quad (4)$$

$$SAPS_t(O_i, O_j) = \exp\Big\{ -1/(\sum_{n=t}^{t+\tau} \|V_{j,n}\| - \sum_{n=t}^{t+\tau} \|V_{i,n}\|)\Big\} \quad (5)$$

Where $\|V_{i,t}\|$ is the magnitude of the velocity of O_i at t^{th} frame and τ is the threshold on the number of frames. From Equations (4) and (5), it is clear that at pushing point(O_i pushed O_j) both the score will be high and low on other positions.

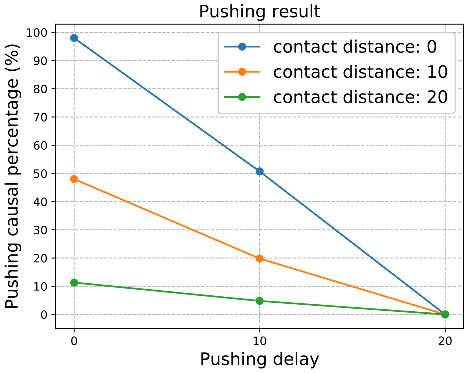

Figure 3. Pushing results wit different *pushing delay* and *contact distance* combination.

We calculate a pushing score $PS_t(O_i, O_j)$ between the objects O_i and O_j at t^{th} frame as:

$$\begin{aligned} PS_t(O_i, O_j) \quad = \quad & PDLS_t(O_j) \times PDSS_t(O_i, O_j) \times \\ & SBPS_t(O_i, O_j) \times SAPS_t(O_i, O_j) \end{aligned} \tag{6}$$

and the final *causality heuristic* $CH(O_i, O_j)$ between objects O_i and O_j in a video can be calculated by summation over all frames within that video using Equation (6) as:

$$CH(O_i, O_j) = \sum_{t=0}^{T} PS_t(O_i, O_j) \tag{7}$$

C. Pushing Results

We have tested our above proposed *causality heuristics* on 100 videos for each pushing constraint combinations and averaged over all pushing subtleties. Fig. 3 shows the results for each combination of pushing delay and pushing distance. To examine whether the model's behavior matches the human data, we applied a regression to the percentage pushing with delay and distance at contact crossed. The percent of videos labeled as pushing was negatively reduced by delay, $\beta = -4.78928$, $t(5) = -18$, $p < 0.001$, and it was negatively reduced by distance, $\beta = -4.37822$, $t(5) = -17$, $p < 0.001$. Since accuracy could not go below 0, the floor effect created a positive interaction of delay and distance, $\beta = 0.21$, $t(5) = 11$, $p < 0.001$. Thus the model captured features of Leslie and Keeble [6], where causality in pushing events is reduced by a distance at contact and delay in the launching of the pushee.

IV. Discussion

The understanding the interactions of individuals at a distance, where body part and appearance information is difficult to reliably extract, is still a challenge for systems of action understanding. Several approaches use supervised learning using motion features related to the individuals to develop systems that can label actions (e.g., chasing, greeting) and roles (e.g., agent, patient). Here we use experimental psychophysical data from humans to develop more complex features for action understanding.

We developed an approach heuristic which encodes the consistency in which one object approaches another object. This heuristic is similar to Moreno and Poope's system [2], but it is graded, which captures the way that role labeling varies systematically with chasing subtlety, such that lower subtlety lead to higher accuracy chaser identification. The slope of the change in accuracy with changes in chasing subtlety matched human performance [4]. Another heuristic that we developed was a causality heuristic that identified causal pushing actions. The algorithm could explain the sensitivity to contact and delay in the infant data [6].

Our claim is that the presence of complex graded relational heuristics in infants is evidence against a system which uses supervised learning to acquire its ability to understand actions from simple motion features (e.g., velocity). Instead, we think these heuristics evolved specifically to support important survival functions (e.g., chasing predators or mates), and computational systems for action understanding can be improved by incorporating important heuristics that are evident in the psychological literature.

References

[1] R. Poppe, “A survey on vision-based human action recognition,” *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, 2010.

[2] A. Moreno and R. Poppe, “”you?re it!”: Role identification using pairwise interactions in tag games,” in *CVPRW*, 2013.

[3] M.-C. Chang, N. Krahnstoever, and W. Ge, “Probabilistic group-level motion analysis and scenario recognition,” in *ICCV*, 2011.

[4] T. Gao, G. E. Newman, and B. J. Scholl, “The psychophysics of chasing: A case study in the perception of animacy,” *Cognitive Psychology*, vol. 59, no. 2, pp. 154–179, 2009.

[5] W. E. Frankenhuis, B. House, H. C. Barrett, and S. P. Johnson, “Infants' perception of chasing,” *Cognition*, vol. 126, no. 2, pp. 224–233, 2013.

[6] A. M. Leslie and S. Keeble, “Do six-month-old infants perceive causality?” *Cognition*, vol. 25, no. 3, pp. 265–288, 1987.

[7] Z. W. Pylyshyn and R. W. Storm, “Tracking multiple independent targets: Evidence for a parallel tracking mechanism,” *Spatial Vision*, vol. 3, no. 3, pp. 179–197, 1988.

[8] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME - Journal of basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[9] L. B. Cohen and L. M. Oakes, “How infants perceive a simple causal event,” *Developmental Psychology*, vol. 29, no. 3, pp. 421–433, 1993.