Influence of innate behavior on the dynamics of innovation in foraging

S Ganga Prasath

1 2D grid world

Let us say an agent has an intrinsic behavior to follow a herd or a trail laid by other ants or a flock of birds. All the agents simply following this herd following behavior will not result in new solutions. Agents often have to innovate and find new solutions either because the trails are not available anymore or because you know a better route (influence of history) or perturbations (intrinsic or environmental) can throw you off trails and you have to find new solutions. We are interested in understanding the role of intrinsic behavior on the dynamics of innovation.

We will start by implementing an agent trying to optimally traverse from point A to point B in a 2D grid world. The steps involved in traditional reinforcement learning based on SARSA algorithm involves essentially 2-steps: (i) action-value function update, (ii) policy update. The action-value update equation in on-policy SARSA is given by

$$\begin{aligned} Q_\pi(s_t, a_t) \leftarrow Q_\pi(s_t, a_t) \\ + \alpha\{r_{t+1} + \gamma Q_\pi(s_{t+1}, a_{t+1}) - Q_\pi(s_t, a_t)\} \end{aligned} \tag{1}$$

while the policy update is given by

$$\pi(s) = \arg\max_a Q(s, a).$$

This is shown in algorithmic form in Alg. 1.

Algorithm 1: SARSA algorithm for action-value update

Initialize: State, s_0; action, a_0; action-value function, $Q(s_j, a_j)$
foreach *epoch* **do**
  Update action-value function using SARSA rule:

$$Q_\pi(s_t, a_t) \leftarrow Q_\pi(s_t, a_t) + \alpha\{r_{t+1} + \gamma Q_\pi(s_{t+1}, a_{t+1}) - Q_\pi(s_t, a_t)\}$$

  Choose action:

$$\pi(s_t) = \arg\max_a Q(s_t, a_t)$$

  Calculate reward, r_t
  Check if target s^* is reached, continue if not
end

2 Trail following behavior

Consider a trail of pheromone $\mathbf{x}^*(s)$ represented by arc-length parameterization s. The concentration field in 2D is then given by $c(\mathbf{x}) = c_o\delta(\mathbf{x} - \mathbf{x}^*(s))$. This filed of course is assumed to be steady but can be made time-dependent by simply adding a decay time-scale τ to get: $c(\mathbf{x}, t) = c_o\delta(\mathbf{x} - \mathbf{x}^*(s))\mathrm{e}^{-t/\tau}$. We can immediately evaluate some of the properties of the curve $\mathbf{x}^*(s)$ which will come in handy soon: $\hat{\mathbf{t}}(s) = \mathrm{d}\mathbf{x}^*(s)/\mathrm{d}s = \{\cos\psi(s), \sin\psi(s)\}$ and $\hat{\mathbf{n}}(s) = \{\sin\psi(s), \cos\psi(s)\}$. As we can see the entire curve $\mathbf{x}^*(s)$ can be represented only using $\psi(s)$ up to global translations and rotations, which is a well known property of curves in 2D.

The state of the agent/ant in our problem is represented by its coordinates $\mathbf{r}(t) = \{r_x(t), r_y(t)\}$ and it can make measurements about how far from the pheromone trail it is $\mathrm{d}\mathbf{r}(t)$ as well as the orientation of the trail, $\hat{\mathbf{t}}(s, t)$. The action that it takes from these measurements is to align its orientation, $\hat{\mathbf{p}}(t)$ along a direction that will take it towards the trail. We show in Fig. 2 a schematic of the problem set up.The strategy used by the agent for tracking the pheromone trail will be to move along the direction given by $(\mathrm{d}\mathbf{r}/|\mathrm{d}\mathbf{r}| + \hat{\mathbf{t}})$ by a fixed length h. In order to identify the location along $\mathbf{x}^*(s)$ where $\mathrm{d}\mathbf{r}$ intersects, we use the condition $\mathrm{d}\mathbf{r}(t) \perp \hat{\mathbf{t}}(s, t)$.

2.1 Semi-circular trail

For the problem at hand, we will considering a semi-circular trail whose coordinates can be written in arc-length form as $\mathbf{x}^*(s) = \{a\cos(s/a), a\sin(s/a)\}$ where a is the radius of the circle. The trail starts at $s = 0$ given by coordinates $\mathbf{x}^*(0) = \{a, 0\}$ and ends at $s = \pi a$ at $\mathbf{x}^*(\pi a) = \{-a, 0\}$. From this it is easy to find that $\hat{\mathbf{t}}(s) = \{-\sin(s/a), \cos(s/a)\} =$

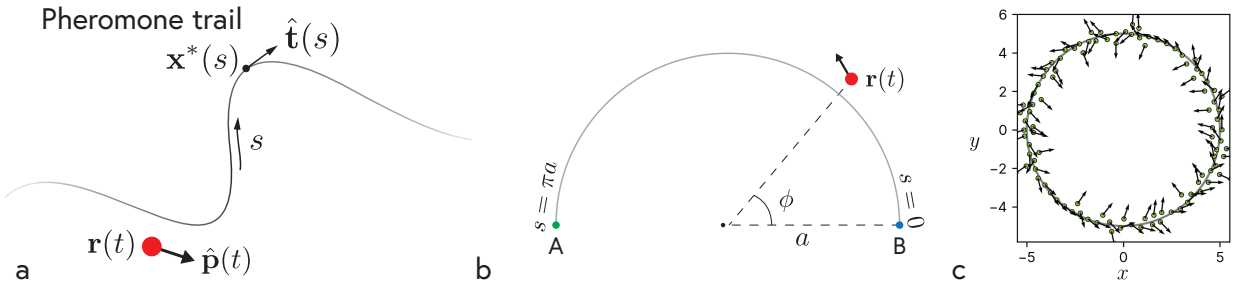

Figure 1: Schematic of setup

$\{\cos(\pi/2 + s/a), \sin(\pi/2 + s/a)\}$, and $\hat{\mathbf{n}}(s) = \{-\cos(s/a), \sin(s/a)\}$. We see that we can represent $\hat{\mathbf{t}}(s)$ through $\psi(s) = (\pi/2 + s/a)$. We can now calculate the location along the arc-length where $\mathbf{r}(t)$ is closest by using the constraint condition $\mathrm{d}\mathbf{r}(t) \perp \hat{\mathbf{t}}(s,t)$ or equivalently $\mathrm{d}\mathbf{r}(t) \cdot \hat{\mathbf{t}}(s,t) = 0$. We can now write $\mathrm{d}\mathbf{r}(t) \sim \{a\cos(s/a) - r_x, a\sin(s/a) - r_y\}$ (up to normalization) and get the location $s_*(t) = a\tan^{-1}(r_y/r_x)$ by using the formula for $\hat{\mathbf{t}}(s)$. From this it is trivial to see that the angle along $\mathrm{d}\mathbf{r}$ is $\phi(t) = \tan^{-1}(r_y/r_x)$ (as is evident in Fig. 2(b)). For a given location of the agent, $\mathbf{r}(t)$ the orientation it needs to take in the next step can be easily calculated to be $\theta(t) = (\pi/4 - \phi(t))$.

We can now set up the entire dynamics of the agent's trail following behavior on this semi-circle. We can state this in the notation of reinforcement learning as it will become helpful later. The state of the agent is $S^t = \{r_x^t, r_y^t\}$, the measurements it makes are $M^t = \phi^t$ and using this information the action the agent takes is $A^t = \{\theta^t\}$ which is its orientation. The dynamics of the agent can now be written as follows:

$$\text{Measurement update: } \phi^{t+1} = \tan^{-1}\left(\frac{r_y^t}{r_x^t}\right) + \zeta^t, \tag{2}$$

$$\text{Action update: } \hat{\mathbf{p}}^{t+1} = \hat{\mathbf{t}}^{t+1} + \mathrm{d}\mathbf{r}^{t+1}, \tag{3}$$

$$\text{State update: } \mathbf{r}^{t+1} = \mathbf{r}^t + l\hat{\mathbf{p}}^{t+1}, \tag{4}$$

where $\mathbf{r}^t = \{r_x^t, r_y^t\}$, $\mathrm{d}\mathbf{r}^t = (\mathbf{x}^*(s^t) - \mathbf{r}^t)/||\mathbf{x}^*(s^t) - \mathbf{r}(t)||$ and $\hat{\mathbf{p}}^t = \{\cos\theta^t, \sin\theta^t\}$. We have added sensory noise ζ^t (which is sampled from a uniform distribution) to the measurement to reflect the error that accompanies measurements usually. The solution dynamics is shown in Fig. 2(c). We call this line-following behavior as a policy $\pi_{\text{ite}}(\mathbf{r}, \phi)$: which denotes the agent's innate behavior to follow pheromone trails.