A Simulation of Ant Formation and Foraging using Fuzzy Logic and Reinforcement Learning

S. Afshar
Center for Mechatronics and Automation
School of Mechanical Engineering
University of Tehran
Tehran, Iran

M. J. Mahjoob
Center for Mechatronics and Automation
School of Mechanical Engineering
University of Tehran
mmahjoob@ut.ac.ir

***Abstract*— Pheromone trails laid by foraging ants serve as a positive feedback mechanism in the ant colonies to share information (in search of food sources). The simulation conducted here of this swarm intelligence can help to realize the process and implement it further for artificial swarms. With available instrumentation we may easily record the agents' position in each step. Pheromone trails are then generated and each agent learns to follow the trail. Fuzzy logic is used to approximate pheromone value at each point. Agents learn to behave like ants using a Reinforcement Learning (RL) process. The algorithm has appropriate parameters that may be set according to the search space dimension. Simulation results are in good agreement with the observations of ants' behavior.**

***Keywords*—ant formation, foraging, reinforcement learning**

I. INTRODUCTION

Social insect societies - ants, bees, termites and wasp – are--distributed systems in which colony-level behavior emerges out of interactions among individual insects [1]. In these systems complex processes are observed at collective level only by implementing simple behavioral rules. A similar approach can be used in artificial distributed systems. Swarm intelligence (SI) is a relatively new paradigm to improve the control of large numbers of interacting entities. Flexibility and robustness are two features desired in an artificial system. The application of SI principles leads to these features needed for a decentralized system. Collective robotics, for instance, relies on the insect metaphor as a novel paradigm to design distributed control algorithms for swarms of robots.

One of the problems (and solution as well) inspired by colony-level behavior is the source localization. To survive, insect societies must organize their workforce efficiently. This organization involves making collective decisions that optimize the colony's fitness. For example ants are capable of laying pheromone trails from nest to food sources. This maximizes energy efficiency where they follow the shortest route to a food source [2].

Here there are two ways to communicate and transfer information: direct physical contact and indirect communication using environment, the latter known as stigmergy [14]. As ants move around they leave pheromone trails, which dissipate over time and distance. The pheromone molecules that a wandering ant might encounter are higher either when ants have passed over the spot more recently or when larger numbers of ants have passed over the spot [3]. Selecting denser pheromone trails would steer ants to take more efficient routes to food sources.

If only physical contact of agent is used, generality of goal access is affected by dimensions of solving space. As we cannot increase the dimension of solving space, to guarantee the comprehensiveness of solving the problem we must then use chemical communication. In this paper, the construction of trail is used to preserve the connection of agents. To achieve this end the agents must learn, through the time, to follow the trail quite effectively.

Several related works should be cited here. Guohuo Ye et al. considered the composition of potential flow techniques, artificial potentials and dynamic connectivity to realize complex swarm behaviors [4]. However, the method needs the position of robots and environment dynamics. Russell and Purnamadiajia used direct inspiration of releasing chemical substance of some insects and also the behavior where one agent leads the group as the queen. This chemical substance (pheromone) is only released by the queen and swarm members are allowed to locate and recognize the leader [2].

Mondada et al. presented a new robotic concept, called SWAR-BOT, based on a swarm of small and simple autonomous mobile robots called S-BOTs. S-BOTs have particular capability that allows them to connect to and disconnect from each other [5]. Darigo et al. further developed this approach [6,7]. They focused on providing the S-BOTs with two basic abilities that are aggregation and coordinated motion [7].

X. Cui et al. introduced Biasing Expansion Swarm Approach (BESA). Applying the three properties of swarm behavior: separation, cohesion, and alignment, they achieved dynamically stable ad-hoc connectivity and fast target convergence. They use a grid map to represent the unknown environment. Then, by an ad-hoc network communication and a biasing expansion algorithm, each agent makes use of all concentration value collected by other members [8]. In this method each agent should collect information about the environment of the other members. Orco and Teodorovic studied the collective behavior of bees and presented a new approach called Bee Colony Optimization (BCO) [9]. The work of T.Schmickl et al. was also inspired by bee colony foraging behavior [10].

In this paper, we use the trail formation behavior of ants. Assuming a method of recording the position of agents in each step, paths of ants' movement are constructed. Other agents can follow these paths when they arrive at them. In this method, in each step (sample time), a number indicating the 'pheromone range' is reported to each agent and the agent utilizes it to choose the next action.

First we explain the method of finding a path that can be used as the agent path. Then we determine when an agent is considered to be *in the vicinity* of a path. Also, the value of pheromone detected by the agent will be calculated using fuzzy logic (section II). The agent will then be trained to behave like ants in path following (section III). Finally, we present the results obtained running the simulation process (section IV).

II. PHEROMONE DETERMINATION USING FUZZY LOGIC

A. *constructing agent path.*

We suppose that the position of each agent is known in each snapshot. There could be a curve now to estimate the agent path. Due to the short step length, we use linear estimation for each two points indicating one sample. After constructing an agent path, the distance of other agents from this path must be determined. Each agent may be in the vicinity of one path segment located between two points as indicated in Figure. 1.

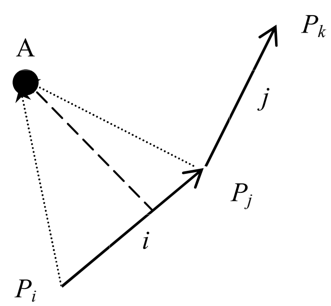

Figure 1. Configuration of an agent *A* and a path *P*

As indicated in the figure, the agent is in close reach of segment *i*. This means that it has minimum distance with the segment *i* joining two points P_i and P_j. The state of being *close* to a path segment can be specified as follows.

First determine three vectors below:

$$\begin{aligned} \vec{V}_i &= \overrightarrow{AP_i} = P_i - A \\ \vec{V}_j &= \overrightarrow{AP_j} = P_j - A \\ \vec{V}_{ij} &= \overrightarrow{P_iP_j} = P_j - P_i \end{aligned} \tag{1}$$

Now if agent (*A*) is in the vicinity of path *i* , then :

$$\cos(\overrightarrow{AP_i}, \overrightarrow{P_jP_i}).\cos(\overrightarrow{AP_j}, \overrightarrow{P_jP_i}) \leq 0 \tag{2}$$

The distance of agent *i*, from path *j*, can then be calculated as follows:

$$d_{ij} = \left|\overrightarrow{AP_i}\right| \sqrt{\left(1 - \cos^2(\overrightarrow{AP_i}, \overrightarrow{P_jP_i})\right)} \tag{3}$$

However, the distance must be within a specific range which is determined according to the pheromone diffusion power; i.e., the extent to which the pheromone can be sensed.

B. *pheromone detection:*

After determining minimum distance to each path, the next step is to evaluate the pheromone strength or value.

To calculate pheromone value, fuzzy logic is used. A number of rules are defined and then an inference engine is set up. The attributes associated with pheromone trail can be well described by fuzzy logic. When an ant travels randomly it releases a substance, the pheromone intensity at the discharge spot is greater than other areas. The farther an ant is from a pheromoned spot/path, the less the pheromone value it may sense. This fact delineates two different parts of the rules. The first step is to define a parameter μ to describe a weight for the agent position (in terms of path detection):

- IF μ is S THEN Ph is P_l
- IF μ is M THEN Ph is P_{mean}
- IF μ is L THEN Ph is P_h
- IF μ is VH THEN Ph is P_{vh}
- IF μ is EH THEN Ph is P_{eh}

where we have used a zero order TSK to avoid calculation burden [11]. In these fuzzy rules P_l, P_{mean}, P_h, P_{vh}, and P_{eh} are constants used in consequent of zero order TSK. The Sugeno fuzzy model (TSK) was proposed by Takagi, Sugeno, and Kang in an effort to develop a systematic approach to generating fuzzy rules from a given input-output data set. A typical fuzzy rule in a Sugeno fuzzy model has the form:

If x is A and y is B then z=f(x,y).

where A and B are fuzzy sets in the antecedent, while *z=f(x,y)* is a crisp function in the consequent. When *f* is a constant, we have a zero-order Sugeno fuzzy model [11].

The input space is divided into five sets. A Gaussian membership function is chosen. S, M, L, VL, El are the fuzzy sets.

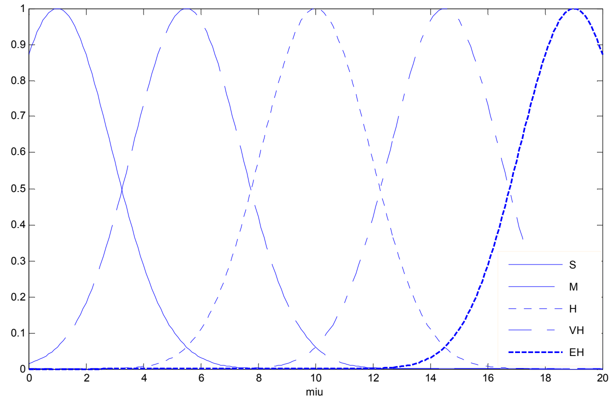

Figure 2. Membership functions for μ fuzzy sets

Note that the pheromone value must not exceed the pheromone saturation threshold. Therefore, if pheromone level becomes larger than this threshold then a mechanism is used to reduce the level (pheromone value). The Gaussian function used here meets the objective. Now pheromone value can be calculated from the following equations:

$$\mu = \sum_{i=1}^{n} \exp\left(-\frac{d_i^2}{\sigma_i^2}\right) \tag{4}$$

$$P = \frac{\sum_{i=1}^{5} \omega_i P_i}{\sum_{i=1}^{5} \omega_i} \tag{5}$$

$$\omega_i = \exp\left(-\frac{(\mu - n_i)^2}{\sigma_\mu^2}\right) \tag{6}$$

$$\sigma_\mu = \frac{|P_l - P_h|}{4Q_{da}\Delta t\sqrt{-\ln 0.5}} \tag{7}$$

In the above equations, μ is the parameter indicating distance weight. For instance if an ant detects a path and is exactly on the path (i.e. distance equals zero and n=1), then μ becomes one. This means that the ant has found the path with a weight equal to one. If an ant moves far from a path, this weight is reduced. In contrast once the number of near paths increases, the path weight increases accordingly. σ_i indicates the scale difference of pheromone detection for two different points of a path. A larger σ_i implies a smaller difference between pheromone values of two points and vise versa. Note that only if the distance of an ant from each path is smaller than *r* (an internal parameter within the simulation which stands for the pheromone diffusion radius), the distance can appear in calculations.

P_i*'s* in the equations are different values of pheromone. ω_i is the firing weight, n_i is the number of ants or the number of times one ant follows a path, P_l is the stimulus floor level of ants' antennae. P_h is the saturation level of pheromone detection. Q_{da} is the difference between pheromone releasing rate and pheromone evaporation rate, and Δt is the sample time interval. The values of parameters σ_μ, n_i are selected so that at the intersection points of membership functions, the value of membership function equals 0.5. If the value of pheromone exceeds P_h then the value reduction process will fire and pheromone value is reduced according to the following relation:

$$P = P_h \exp\left(\frac{-(P - P_h)^2}{\sigma_0^2}\right) \tag{8}$$

$$\sigma_0 = \frac{P_{sat} - P_h}{-\ln \frac{P_l}{P_h}} \tag{9}$$

Where P_{sat} indicates the pheromone value that we want to reduce bellow p_l.

III. TRAINING THE ROBOTS

In fact ants have some pattern to build and follow pheromone trail. However they do not behave in a deterministic fashion. As it was mentioned before selecting denser path segment or spots by ants leads to more efficient trails. Pheromone trail laid by foraging ants is a positive feedback mechanism. This feedback is nonlinear, in that ants do not react in a proportionate manner to the amount of pheromone deposited [2]. Strong trails elicit disproportionately stronger response than weak trails.

Here we try to train a group of agents to follow others' paths specially those which are more pheromoned. Thus we use reinforcement learning (RL), in particular SARSA which is an on-line learning algorithm [12]. For each agent, three actions and 9 minor states within 3 major states are defined. The major states indicate the position of the ant relative to the prey while the minor states indicate the position of the ant relative to the path. An index is also used to show the antenna with which the ant detects the trail first.

The state-action value can be written as:

$$Q(st, ac, stc, dpha) \tag{10}$$

in which *st* is the minor state, *ac* is the action selected at each step, *stc* is the major state, and *dpha* indicates which antenna detects the trail first. Each agent has two antennas. When the agent senses pheromone somewhere, one of them achieve more pheromone than the other; therefore a difference is generated between the two antennas. This difference determines the movement direction of the agent relative to path pheromone concentration. Here we reinforce agents to move toward the positive gradient from less pheromoned antennae to more pheromoned antennae. In this manner the value of pheromone will often increase.

The algorithm used here is [12]:

Initialize *Q(s,a)*

Repeat (for each episode):

Initialize *s*

Choose *a* from *s* using policy derived from *Q*

(e.g., ε _greedy)

Repeat (for each step of episode):

Take action *a*, observe *r*,*s'*

Choose *a'* from *s'* using policy derived from *Q*

(e.g., ε _greedy)

$Q(s,a)=Q(s,a)+\alpha\,[r+\gamma\, Q(s',a')-Q(s,a)]$

$s=s'$; $a=a'$;

until *s* terminates.

Actions are ‘go straight’, ‘turn right’, and ‘turn left’. The action space is continuous. The agents minor states are defined using pheromone value, pheromone gradient sign, and facing /crossing neighboring agents/paths. Major states are determined according to position of prey relative to each agent.

In this algorithm *s* is replaced by *(st, stc, dpha)* vector. The control of agents is on-line and decentralized. The learning algorithm can be tuned using parameters mentioned before and by rewards given to agents at each step. Being in a state with a positive gradient, results in a higher reward (relative to negative gradient). Also more pheromone leads to more reward. The learning process is collective learning and each agent uses the others’ experiment.

The required parameters are chosen as follows: r=0.4 cm, d (maximum step size) =0.1 cm, θ (maximum turning rate) $=10^0$/step, l_a(antennae length) = 0.3 cm, P_l=0.1, P_h=1, P_{sat}=1.3, P_{mean}=.55, P_{vh} =1.4, P_{eh} =1.9, Δt =1, Q_{da} =0.1, α (the antenna angle from body axis) = 30^0. The parameters are selected (with some changes) from [13] to show that the agents follow one path after training time. If the μ parameter value for one agent becomes bellow 0.1 for which the pheromone value is 0.108, then this value is reduced linearly to zero at μ value equals 0.04. Therefore the reduction of pheromone value may be a fuzzy reduction.

A schematic sketch of the agent paths is shown in figure 3a, 3b, 3c. The results were obtained after three runs of the code with 1000 steps in each run. At the beginning of the training algorithm, agents cross the paths which they detect and do not follow them. As it is clear from figures when an agent receive to a path it would move toward the path center and then would follow it, but if the agent reach to one point where there were two or more paths, agent would select the path that has denser pheromone from agent’s point of view.

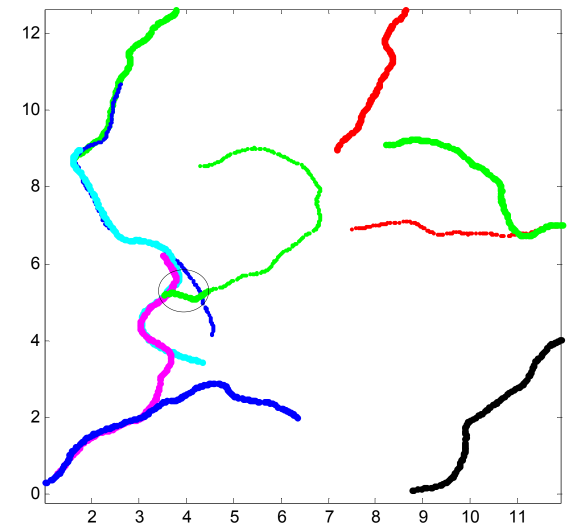

Figure 3a. The paths of agents for 200 steps from step 1.

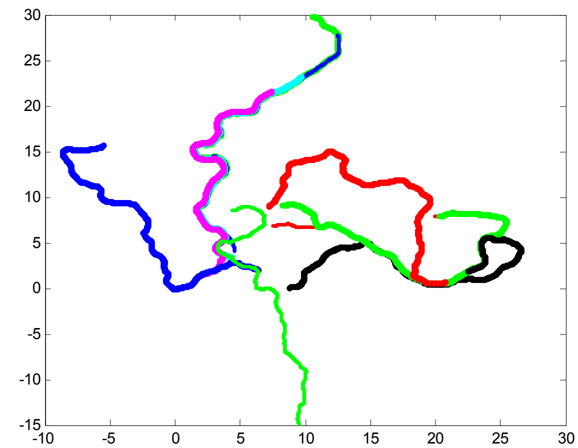

Figure 3b. The paths of agents for 400 steps from step 200.

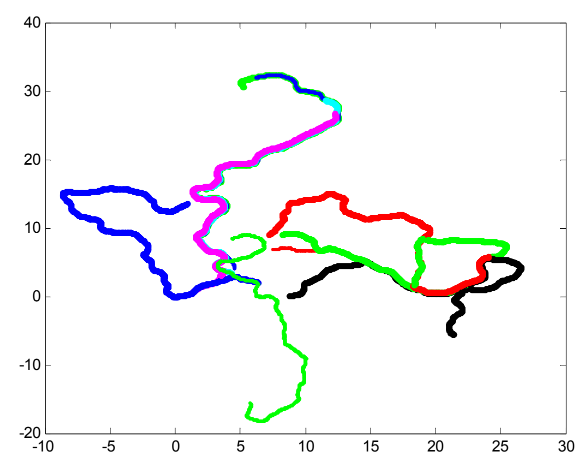

Figure 3c. The path of agents for 400steps from step 600

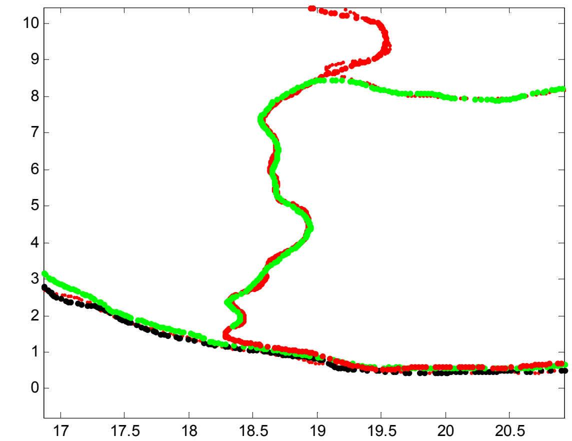

Figure 4. Traveling path of ant in the vicinity of one path

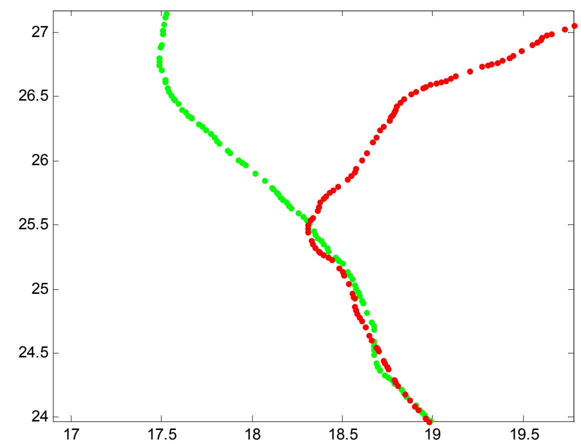

Figure 5. Traveling path of an ant when the approaching angle is near 90 ^{0}.

Figures 4 show the movement pattern in the vicinity of one path. If the approaching angle of the agent to a path is near 90^0, the agent cannot adjust itself to move on the path direction later as shown in figure 5.

IV. DISCUSSION/CONCLUSION

A simulation was conducted here of ant formation and foraging using fuzzy logic and reinforcement learning. It helped us to realize the behavior pattern of swarm ants to implement it further for swarm robotics. The goal was to train a set of agents to join paths generated by others so as to generate a network and follow it.

In the previous sections we constructed the agents' path, and using fuzzy logic, we determined the value of pheromone that each agent detected moving around in the environment. Because of the evaporation phenomenon considered in pheromone calculation, maximizing pheromone value detection leads to minimum distance selection. Over the time, due to evaporation, the pheromone value decreases. When an agent detects a path that is generated in the beginning of the foraging process, the value of sensed pheromone is less relative to the same path generated later. Therefore, in the created network, agents select the path that either has been created recently or has been the path of frequent traveling agents.

Laboratory studies have shown that the individual ants move along a trail before losing it. This is a complex function of the pheromone concentration of the trail. Initially, as the concentration increases, the length of the trail that an ant can follow increases. However, as the concentration increases further, the trail-following ability begins to decrease. Furthermore, certain ant species exhibit a sinusoidal trajectory as the trail is followed [13]. A chatter like pattern for agent travel paths can be seen in figures 4, which is also observed in ant colonies. Agents learn to follow the path and do not leave it but they happen to leave their path somewhere along. This is because of detecting new paths on their way ahead; they do not therefore get back to the previous path. Also if the angle between an agent moving direction and a path it detects is close to 90° then the agent will not follow the path.

The developed method of stigmergy can further be used to simulate the foraging behavior of ants. Networks of trails that control the flow of resources and information are constructed and the agents select the path (among all paths) which results in the desired agents' formation and foraging.

REFERENCES

[1] C. Anderson , N. R. Franks, D. W. Mcshea, "The Complexity and Hierarchical Structure of Tasks in Insect Society", Animal behavior, 2001, 62, 643-651.

[2] D. J. T. Sumpter, M. Beekman , "From Nonlinearity to Optimality: Pheromone Trail Foraging by Ants" , Center for mathematical biology, Animal behavior, 2003, 66, 273-280.

[3] James Kennedy, Russel C.Eberhart, Swarm intelligenc, Morgan Kaufmann Publishers, 2001.

[4] Guohua Ye, Hua O. Wang, Kazuo Tanaka, "Coordinated Motion Control of Swarms with Dynamic Connectivity in Potential Flows", paper collection, IFAC05, 04819, 2005

[5] F. Mondada , G. C. Pettinaro, I. Kwee, A. Guignard, L. Gambardella, D. Floreano, S. Nolfi, J. L. Deneubourg, M. Dorigo, "Swarm-Bot: Swarm of Autonomous Mobile Robots with Self-Assembling Capabilities", 8th Conference on Intelligent Autonomous Systems (IAS8), pp. 53-60.

[6] M. Dorigo, E. Tuci, R. Grof, V. Trianni, T. H. Labella, S. Nouyan, C. Ampatzis, J. L. Deneubourg, G. Baldassarre, S. Nolfi, F. Mondada, D. Floreano, and L. M. Gambardella, "The SWARM-BOTS Project", Springer-Verlag Berlin Heidelberg, 2005.

[7] M. Dorigo, V. Trianni, E. Sahin, R. Grof, T. H. Labella, G. Baldassarre, S. Nolfi, J. L. Deneubourg, F. Mondada, D. Floreano, L. M. Gambardella, "Evolving Self – Organizing Behaviors for a Swarm-bot", Autonomous Robots, CiteSeer,17(2-3):223-245, 2004.

[8] X. Cui, C. T. Hardin, R. K. Ragade, and A. S. Elmaghraby, "A Swarm Approach for Emission Sources Localization", 16th IEEE International Conference on Tools with Artificial Intelligence, p. 1082-3409, 2004

[9] D. Teodorovic, M. D. Orco, "Bee Colony Optimization – A Cooperative Learning Approach to Complex Transportation Problems", IASI CNR, 16th conference, 2000.

[10] T.Schmikl, R.Thenius, K.Crailsheim, "Simulating Swarm Intelligence in Honey Bees: Foraging in Differently Fluctuating Enviroments", Washington DC, USA, 2005.

[11] J-S Jang, C-T Sun, “Nero-Fuzzy Modeling and Control”, Mathworks Inc., Natick, MA 01760 USA, Computer and inforemation science Dep., Uni. National Chiao Tung, IEEE Log number 9408301.

[12] Richard S. Sutton, Andrew G. Barto, Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, 1998

[13] I.D. Couzin, N. R. Franks, “Self-Organzed Lane Formation and Optimized Traffic Flow in Army Ants”, Ecology and Evolutionary Biology Dep., Uni. Princeton, Princeton, Nf 08544, USA*, Center of Behavioural Biology, School of Biological Sciences, Uni. Bristol, Bristol BS8 1UG, UK. The Royal society, 2002.

[14] C. R. Kube, E. Bonabeau, Cooperative Transport by Ants Robots, Elsevier Science, Robotics and Autonomous Systems, 2000.