Assessing Learned Models of Fish Schooling Behavior (Extended Abstract)

Brian Hrolenok
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia 30332
bhrolenok3@gatech.edu

Tucker Balch
School of Interactive Computing
Georgia Institute of Technology
Atlanta, Georgia 30332
tucker@cc.gatech.edu

ABSTRACT

Agent-based simulation is a valuable tool for validating the theoretical models biologists and ethologists use to explain animal behavior. By automating the process of constructing agent-based models (ABM) directly from observation data we can enable researchers to focus more of their time on the analysis of the behaviors and animals in question. We present experimental results using a modified version of k Nearest Neighbor to learn an executable model of fish schooling behavior from both synthetic data and tracking data of live juvenile *Notemigonus Crysoleucas*, and quantitatively asses the performance of the learned behavior. Our experiments illustrate that our method can successfully learn fish schooling, and provide an objective criteria for comparing competing biological theories of behavior.

Categories and Subject Descriptors

I.2.6 [**Computing Methodologies**]: Learning

General Terms

Algorithms

Keywords

Agent-based model, Learning, Schooling

1. INTRODUCTION

Biologists and ethologists have found a useful tool for describing and analyzing the behavior of social animals in agent-based models and multiagent systems. Such models have been successfully used to describe and analyze hunting behavior in ants [6], schooling behavior in fish [1], and nest site selection behavior, a form of collective decision making, in both ants [5] and bees [4]. However, until now these models have been constructed by hand after careful analysis of large quantities of empirical observation data. For example, in constructing their model of collective nest choice behavior in [5] the authors examine over 12000 interactions between approximately 290 ants in 12 videos each ranging from 30 to 150 minutes in length, by hand. In this work we present new experimental results using a framework for automating the

Table 1: χ^2 histogram similarity for synthetic schooling.

Max distance	Avg NN distance	Var NN distance
$\chi^2 = 0.14$	$\chi^2 = 0.24$	$\chi^2 = 1.21$
$p = 1.0$	$p = 1.0$	$p = 1.0$

process of constructing agent-based models from observational data. Details on the modified k-NN learning method can be found in [3].

2. EXPERIMENTS

We performed two different experiments. In the first we created a synthetic schooling behavior by hand, and then reconstructed the synthetic behavior using logs from our simulator as input to our k-NN algorithm. In the second we took tracking data from top-down video of schooling fish in a shallow tank, and applied the same learning technique. In each instance we compared the learned behavior with the original behavior which generated the training data on three metrics discussed next.

In order to compare whether a learned behavior is similar to the behavior which generated its training data quantitatively, we need to choose a set of metrics which can be computed from the observation data we have access to, and which cover the important aspects of the behavior of interest. For fish schooling we consider three quantitative metrics that describe characteristic aspects of the entire school:

1. **Maximum distance** between two fish in the school. This is an indicator of the overall school diameter

2. **Average distance** to the nearest neighbor fish. This describes the density of the school.

3. **Variance** in the average distance above. This describes how uniformly the school is distributed.

By computing these metrics at each time step and estimating their distributions, we can quantitatively characterize the behavior of the two different systems, and compare them. Figure 1 shows the distributions of the maximum distance metric for the synthetic and real fish as compared to the learned behavior in both cases. Tables 1 and 2 give a quantitative assessment of the similarity between the distributions for all three metrics.
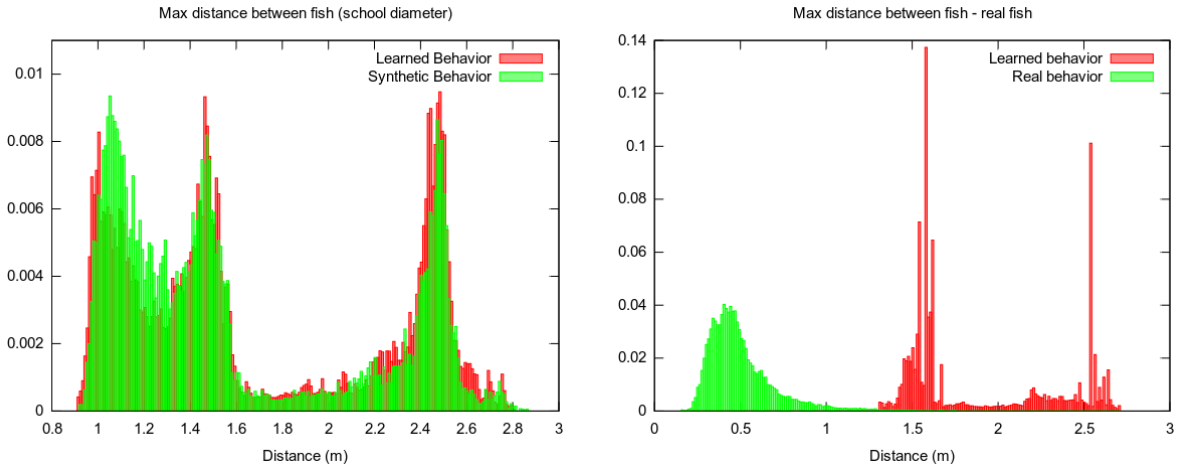
Figure 1: Learned behavior from synthetic and real data. Histograms of the learned behavior compared with the synthetic (*left*) and real (*right*) fish schools using one of the three schooling metrics: Maximum distance. The learned behavior matches the synthetic behavior very closely, but the same sensor model does not yield a close match on real fish.

Table 2: χ^2 histogram similarity for real schooling.

Max distance	Avg NN distance	Var NN distance
$\chi^2 = 339.00$	$\chi^2 = 115.68$	$\chi^2 = 1171.60$
$p = 2.28 \times 10^{-9}$	$p = 0.99$	$p = 2.10 \times 10^{-137}$

3. ANALYSIS

The results on the synthetic data suggest that we can successfully construct an executable model of behavior from observational data, however the data on real fish indicate there is a mismatch between the theoretical model of perception used to compute the percepts from the observation data, and the actual perceptions of the animals. In the synthetic model the perceptions of the agents are known exactly, whereas the perceptions that are important to the schooling behavior of fish have to be inferred. An important benifit of the approach we describe here is the quantitative comparison of the resulting learned behaviors. Learned behaviors which are more similar to the behavior of the animal provide evidence that the perceptual model used in computing the perception–action pairs passed to the learning algorithm is more likely to be what the animal actually perceives.

In future work we would like to explicitly characterize the classes of behaviors that can be accurately learned using our methods, in the sense that the learned behaviors can still be used to correctly rank competing perceptual models. The types of behaviors that the learning mechanism discussed in this paper constructs can be classified as *single state controllers*. Recent work on learning *multi-state controllers* [2] introduces a more complicated learning method, but it remains to be shown what the trade offs are in using a more expressive complex method versus a less expressive but simpler method in a quantitative sense.

5. REFERENCES

[1] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks. Collective memory and spatial sorting in animal groups. *Journal of theoretical biology*, 218(1):1–11, 2002.

[2] B. Hrolenok and T. Balch. Learning executable models of multiagent behavior from live animal observation. In *ICML 2013 Workshop on Machine Learning For System Identification*, 2013.

[3] B. Hrolenok and T. Balch. Learning schooling behavior from observation. In *Advances in Artificial Life, ECAL*, volume 12, pages 686–691, 2013.

[4] C. List, C. Elsholtz, and T. D. Seeley. Independence and interdependence in collective decision making: an agent-based model of nest-site choice by honeybee swarms. *Philosophical transactions of The Royal Society B: biological sciences*, 364(1518):755–762, 2009.

[5] S. C. Pratt, D. J. Sumpter, E. B. Mallon, and N. R. Franks. An agent-based model of collective nest choice by the ant *Temnothorax albipennis. Animal Behaviour*, 70(5):1023–1036, 2005.

[6] Y.-T. Yang, A. Quitmeyer, B. Hrolenok, H. Shang, D. B. Nguyen, T. R. Balch, T. Medina, C. Sherer, and M. Hybinette. Ant hunt: Towards a validated model of live ant hunting behavior. In *FLAIRS Conference*, 2012.