JAABA: An interactive machine learning system for biologists

Kristin Branson, Mayank Kabra, Alice A. Robie, Marta Rivera-Alba, Steven Branson and Adam Hantman
HHMI Janelia Farm Research Campus
Ashburn, VA
`http://www.janelia.org/lab/branson-lab`

1. Introduction

"Big data" problems are everywhere in neuroscience, the result of new technologies that generate huge, complex data sets, in particular large image and video data [1]. Because of the scale of these data sets, machine vision and learning methods are necessary to extract meaning and understanding. Ideally, biologists and computer scientists could work together, with biologists creating annotated data sets indicating the problem and desired solution, and computer scientists developing algorithms to automatically reproduce these manual annotations. Unfortunately, such a clean separation of work is usually not possible. Scientific research is by nature an iterative process, and it is often difficult *a priori* to formulate a mathematical criterion to be optimized.

Animal behavior annotation is a prime example of this. The behavioral "vocabulary" of animals is unknown. For a specific behavior category, it is difficult to draw the boundaries of that category (e.g. how many steps backwards must a fly take to be considered "backing up"?). Finally, some types of mismatches from manual labels are more important than others. To address this problem, we have developed the Janelia Automatic Animal Behavior Annotator (JAABA), an intuitive, interactive system that allows biologists to use machine learning in closed loop without assistance from machine learning experts [2]. Our system allows biologists to quickly and efficiently train behavior classifiers. With this system, biologists can define new behaviors from patterns they find in the data, or adapt existing behavior classifiers to new environments and types of animals. JAABA has been used in several very different biology projects, in flies, mice, and larvae.

2. System description

JAABA uses a simple, practical instantiation of active learning, in which the user chooses and labels a few frames for which they are confident of the labels, then quickly trains an initial classifier. It then displays information about the current classifier's predictions and confidence on all frames. We have engineered JAABA's interface so that the user can browse through large collections of videos to find unlabeled frames for which they are confident of the labels and the current classifier's predictions are incorrect or low confidence. The user can then label these frames, and retrain the classifier. This sequence is repeated as necessary. This is in contrast to more standard active learning, in which *the algorithm*, not the user, chooses examples that would be most informative to be labeled, under the assumption that the user is an oracle and will produce correct labels. Because of the inexactness of most behavior categories, often the examples most informative for the classifier would be those that are also most difficult for the user to label consistently.

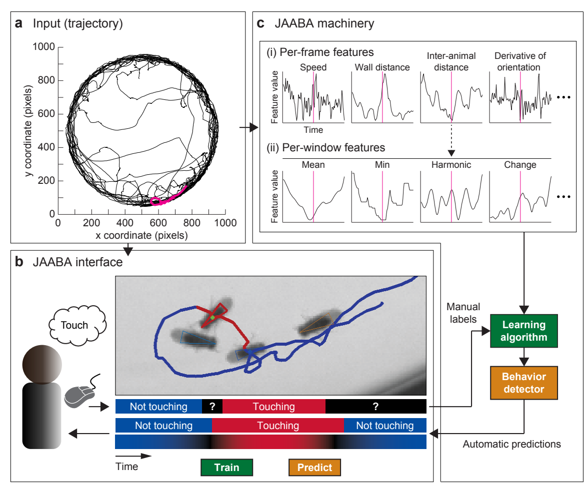

Figure 1: JAABA overview. (a) Input trajectory (x,y position over 1000s). (b) JAABA interface. The top time line shows the user's manual labels, and the bottom two time lines show the clasifier's predictions and confidence. (c) JAABA machinery. (i) Example "per-frame" feature time series. (ii) Example "window" feature time series.

JAABA operates on the animals' trajectories computed from automatic tracking methods (Figure 1a). From these, we compute simple "per-frame" features such as instantaneous speed or distance to the closest animal (Figure 1c(i)). Next, we compute "window features" that describe the distribution of per-frame features around the current frame (Figure 1c(ii)). We compute the mean, standard deviation, minimum, etc. for multiple window sizes and temporal offsets. All window features can be computed efficiently using convolution or image morphology. We compute between 5,000 and 10,000 window features to represent each frame for each animal. We use a version of the GentleBoost algorithm, modified for speed, to train classifiers. Training takes 15-45 seconds, depending on the data set.

We have ground-truthed JAABA's performance by comparing JAABA's automatic predictions to our manual annotations of behavior. We have shown that it is sufficiently powerful and general-purpose to train a large, diverse set of single-animal and social behavior classifiers for flies (15 behaviors), mice (2 behaviors), and *Drosophila* larvae (3 behaviors), achieving an average per-frame error rate of 2.7% and a maximum error rate of 6% over all behaviors (because of the class imbalance, average error rate is computed as the average of the false-positive and false-negative rates).

Besides the standard active-learning benefits (i.e. less training data necessary), we found anecdotally that the interactive interface allowed biologists without expertise in machine learning to better understand how to label consistently, what is easy and hard for our algorithm to learn, and, in fact, to crystalize their behavior definitions. To show that JAABA was usable by non-computer scientists, we gave 12 volunteers a 15-minute presentation on how to install and use JAABA, then asked them to each train and groundtruth a "chase" classifier for flies. The error rates achieved were 2% on average and 5% maximum over users.

3. Applications

JAABA has been used in a variety of biology applications. We have used JAABA to automatically annotate the behavior of flies in our thermogenetic screen of 2,200 transgenic lines of *Drosophila* from the Janelia GAL4 collection (Figure 2(a)). We used a targeted gene expression system to activate different sparse subsets of neurons in each line, and used JAABA to measure the behavioral effects. This data set consisted of >20,000 15-minute videos of groups of 20 flies (in total, nearly half a million flies, and 400 TB of raw data).

One of the main difficulties in applying JAABA to this data set was the behavioral diversity of the flies. Activating different sets of flies' neurons results in widely varying behavioral phenotypes, e.g. jumping constantly, large increases in courtship, flies that only hold their wings out, etc. For results of the screen to be interpretable, the behavior classifiers should work well for each of the 2,200 genotypes. We made use of the interactive interface of JAABA to train an initial classifier, ran the classifier on the entire corpus of videos, used various heuristics to find lines that the classifier was not working well on, and added data from

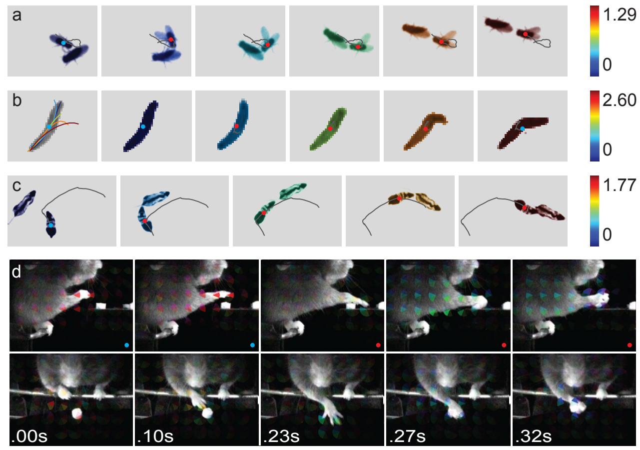

Figure 2: Example behaviors from diverse applications. (a) "Wing extension" in flies. (b) "Headcast" in larvae. (c) "Follow" in mice. (d) "Grab" in head-fixed mouse. In (a-c), color indicates time, (d) color indicates HOF feature strength. Red/blue dots indicate the behavior is/is not predicted as occurring.

these lines to improve the classifier.

This ability to ensure that JAABA performs well in all experimental conditions has been critical in all applications. We are using JAABA to understand behavioral differences between 47 different species and 2 developmental stages of *Drosophila* larvae (Figure 2(b)), where, again, the classifiers must work well across all species. Our collaborators have independently used JAABA in neural (in)activation screens of *Drosophila* larvae to understand the neural circuitry underlying the "roll" behavior in larvae, and in ethology experiments to understand the role of ultrasonic vocalizations in mouse social behavior (Figure 2(c)). Our collaborators are also using a new instantiation of JAABA that operates on HOG/HOF features [3] instead of trajectories to explore the motor planning circuitry in the "reach" behavior in head-fixed mice (Figure 2(d)).

4. Conclusion

In summary, we have found that relatively simple machine vision and learning techniques, in combination with an engineered, intuitive interface, have been sufficient to allow biologists to independently create behavior classifiers for a wide variety of applications.

References

[1] BRAIN Working Group. BRAIN Initiative interim report, September 2013. 1

[2] M. Kabra, A. A. Robie, M. Rivera-Alba, S. Branson, and K. Branson. JAABA: An interactive machine-learning tool for automatic annotation of animal behavior. *Nature Methods*, 10:64–67, 2013. 1

[3] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. In *CVPR*, 2008. 2