Part II

Behavior Analysis with Machine Learning

Chapter 6

Supervised behavior classification

Shifting gears, we now move on to an application of Machine Learning in ethology, or behavior analysis. Biologists study behavior in animals with various objectives. Among others, understanding the human brain has become a primary objective not only for science but also for the US government with the BRAIN initiative [9]. In order to understand such a complex organ as the brain, biologists begin this endeavor by studying the brain of simpler organisms, such as fruit flies (Drosophila megalonaster). To do this, scientists use a large number of techniques in genetics and neuroscience that allow them to identify particular neurons, connections, substances, etc., that are responsible for animal behavior.

Nevertheless, to analyze the behavior of flies, biologists must record thousands of videos of animals to quantify desired behaviors. Analyzing these videos is extremely time consuming, as biologists need to determine how many times certain behavior occurs, as well as extract other properties of the behavior such as duration, latency, etc. Annotating videos can take almost three times as much as the actual duration of the videos, as these have to be viewed in slow motion [5]. This is when Machine Learning comes in handy, as classifiers can be trained to detect specific behaviors from videos of animals.

In this chapter, we describe a classifier that we developed in order to detect a particular fly behavior, “unilateral wing extensions” (UWE). Figure 6.1 shows extracts of a sequence that is classified by biologists as a UWE. The classifier was needed in order to analyze thousands of videos that would take more than few years to annotate by hand. With an automated classifier, the time to analyze the videos became negligible. The work required only involved labeling by hand a small fraction of the videos in order to train and test the system. The resulting classifier was used in the study described in [6]. In that paper, biologists discovered the specific neuron and substance that make male flies aggressive, in situations where there is no competition for any resources, such as food, females, or territory. The hope of this investigation is that aggressive behavior in humans might be understood, especially to be able to design appropriate pharmacological products that could be used in mental

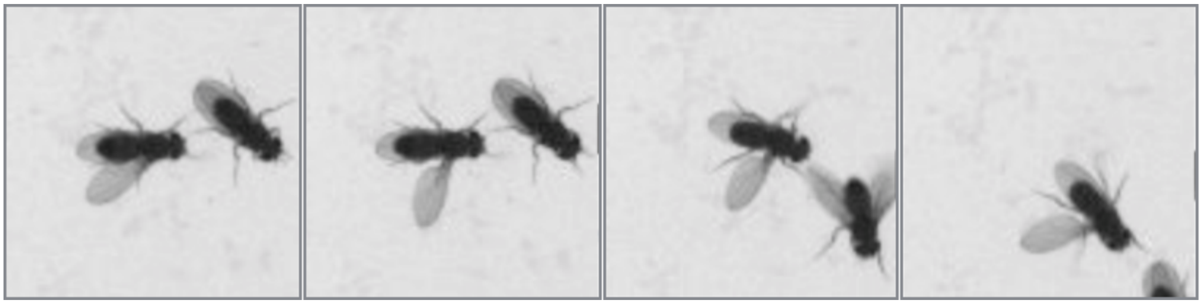

Figure 6.1: Sequence depicting a Unilateral Wing Extension

disorders such as Post Traumatic Stress Disorder, Hyperirritability, etc.

Previous work in the area is described in [31], where different behaviors of flies were detected, mainly through the use of manual filters. The method, however, was not yielding reliable performance in this dataset, reason why a new method was called for. Other work in behavior classification included behavior classification in mice as described in [21] and [39]. However, as we will describe shortly, these methods based on structured output classification were not suited to this specific task where very few labeled positive samples were available, and the quality of the videos varied greatly across the experiment. Some more recent work in behavior classification was published after our work was done, which describes methods for general purpose behavior classifiers (See [33] and [41]).

In the following sections we describe the input used for the system, the learning algorithm chosen and implementation details, the post-processing stage, and the resulting performance of the classifier.

6.1 Pre-processing stage

The original format of the data was video recordings in grayscale, on movies filmed either at 30 or 200Hz. In order to be able to apply any learning algorithm to it, the videos had to be processed first using Computer Vision techniques to extract the trajectories of the flies. Working with the gray scale pixel values would be hopeless, given the number of input dimensions and the small amount of labeled positives. Instead, the Computer Vision system extracts the time series that describe the position of the fly and its various parts. To do this, we used the tracker software described in [33].

This tracker fits an ellipse to the body of each of the flies, and also it detects the position of the tip of the wings and legs. Hence, the position of the fly is uniquely determined by the centroid of the fly, the length of the two axes of the ellipse, the ellipse orientation, and the position of wing and leg tips. Finally, invariant features with respect to position and orientation are derived. These include the minimum wing angle, maximum wing angle, axis ratio, among others, which are crucial for detecting wing extensions. The resulting features and derived features, that included first and second time derivatives resulted in 36 features that describe each frame of the input.

6.2 Learning algorithm

The data set available for training consisted of five twenty minute movies filmed at 200Hz, which were annotated manually for this task. These movies contained only about 4,000 frames labeled as wing extensions, in a dataset that included 480,000 frames. To add complexity to the problem, videos in which the system was going to be tested on had been filmed across several years, and most of them consisted of lower resolution videos filmed at 30Hz with different lighting conditions.

Hence, given the few positives and the level of noise of the data, we decided to take the simplest approach which involved using a frame-by-frame classifier. By taking this approach we were able to use the 4,000 frames as positives, rather than only about 100 positive bouts of wing extensions, the number available in a more complex structured output method. We also chose the simplest learning algorithm to avoid overfitting. We used a squared loss function for the error, weighting positives heavily to account for class imbalance, and used ℓ_2 regularization for the parameters. The model was a linear hypothesis set with non-linear transformation of the features. For the linear transformation, we used a polynomial kernel of degree 2. A more complex kernel was avoided to avoid overfitting to the data. This learning model is the same one as the one described in Appendix A, with the specific non-linear transformation described.

6.3 Post-processing stage

The result of the linear classifier was extremely successful in terms of false negatives, as it detected every wing extension in the test set. Nevertheless, the false positive rate was relatively high. The first post-processing stage that was done was to smooth the result of the classifier with an averaging kernel over the time series. Biologists had determined that the UWE should take at least 130 milliseconds (4 frames in 30 Hz) movies, so that discontinuous frames that were marked as positives were not counted as such. This simple hint allowed reducing the false positive rate significantly.

After watching the remaining false positives that the classifier was giving, it was clear that there were some wing threats (bilateral wing extensions) that were being labeled as positives. Also, events where the fly was grooming itself with the wings were being erroneously labeled as positives. Another case occurred when the flies performed wing extension towards their reflection on the wall rather than towards the other fly. Finally, some mistakes were also due to errors in the tracker when the flies stood up against the wall of the dish, so that the model used by the tracker confused the wings with other parts of the fly. Since the false positives could be grouped and understood easily, further filters to the data were applied. A threshold for the minimum wing angle was set, to avoid bilateral wing extensions (30 degrees). To avoid the grooming events, a minimum threshold for the maximum wing angle was also set (62 degrees). The facing angle between the flies was set to at least 35 degrees to avoid UWE not directed to the other fly, and the axis ratio of the fly was maintained at a minimum of 0.8 to avoid tracker errors from standing flies. This approach does not lead to overfitting in this particular case, as we chose a model with very few parameters compared to the number of frames in the training set.

Applying these filters that were determined by biologists, the final classifier had a recall of 89.4% and a precision of 87.1%. (The recall is the fraction of the true positives detected by the algorithm, while the precision is the fraction of true positives detected divided by both the true detected positives and the false positives). The final test was done on previously unseen movies by the system, which included 113 manually scored UWE, out of which 101 were detected, and 15 false positives were reported [6].

Hence, this simple Machine Learning approach allowed analyzing thousands of videos. Approaches like this have been improved upon in the last few years, in which classifiers are developed for general purpose behaviors, as in [33] and [41]. We decided to explore further the problem of behavior classification, by considering a slightly different setup. Could we discover behaviors from videos of animals in an unsupervised way, that is, without searching for behaviors previously labeled as such by biologists? Answering this question led to the results presented in the following chapter.