

Measuring Behavioral Similarity in *Drosophila*: A Comparative Study of Time-Series Distance Metrics

Ilya Vladimirovskiy

Maksim Bubashvili Fayn

Bar Ilan University for course Advanced Python for Neuroscience

ABSTRACT

Comparing time-series data is crucial for tasks such as classification, clustering, anomaly detection, and overall similarity assessment. We investigated various distance-based measures on time-series behavioral data obtained from *Drosophila melanogaster* via the Fly Bowl setup and the JAABA model. This dataset captures 11 distinct behaviors across experimental sessions, stored as score weights or as binary arrays. Our Python-based workflow overcame MATLAB-based data-loading challenges, enabling multi-experiment processing.

We implemented and evaluated five approaches—Hamming Distance, Interval Overlap, Hidden Markov Model (HMM), Dynamic Time Warping (DTW), and Euclidean Distance—to capture both synchronous and sequence-based similarities. We measured these algorithms using the ratio of inter-group to intra-group distances: a higher ratio indicates better group separation. Small-dataset results suggest HMMs achieve the strongest separation, whereas DTW has the least discriminative power. Future work involves scaling to larger datasets, refining behavior weighting, exploring SFA-based distances, and fixing data normalization issues.

1. INTRODUCTION

Time-series analysis underpins classification, clustering, and anomaly detection in diverse fields, including biology, finance, neurobiology. In the biological context, capturing multi-dimensional behaviors of *Drosophila melanogaster* over time provides deep insights into genetic and environmental influences.

The Fly Bowl setup, paired with the JAABA annotator, yields time-series data of 11 behavioral channels. Values are stored either as score weights (with smaller numbers indicating higher likelihood of a behavior) or as post-processed binary arrays. Since the original data pipeline was MATLAB-based, we developed a new Python interface to load multiple experiments and produce standardized arrays or interval representations suitable for similarity assessment.

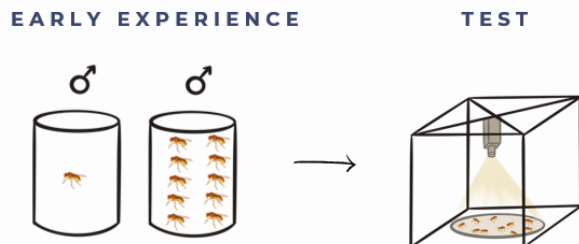


Figure 1: *Drosophila* behavior data collection

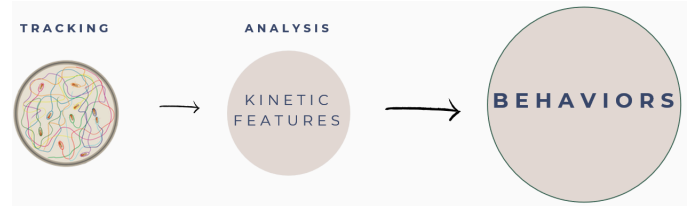


Figure 2: *Drosophila* behavior data collection

2. METHODS

2.1 Data Description and Preprocessing

Each experiment contains a time-series of 11 behaviors for multiple flies.

- Score Weights: A numeric time-series per behavior, one vector per fly per behavior.
- Binary (Post Processed): A binary time-series per behavior, one vector per fly per behavior. Multiple behaviors can be active simultaneously in any given frame.

For analysis, we ensured each fly's data is in a consistent Python structure (NumPy arrays or interval lists). We also built a GUI to batch-load many experiments at once.

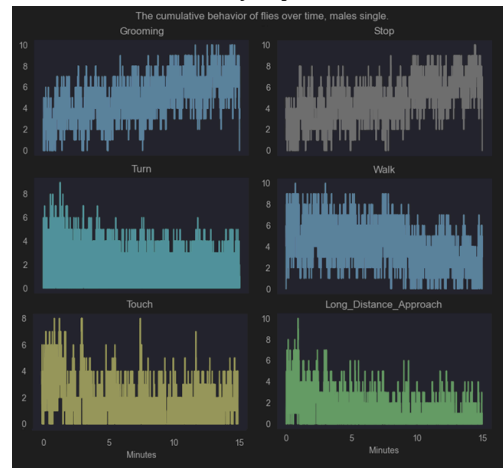


Figure 3: Trends in the data (especially Grooming) which emphasizes the time importance of the time-series nature of the data.

2.2 Similarity Search Algorithms

1. Hamming Distance

- Motivation: Ideal for binary multi-hot vectors (where multiple behaviors may co-occur).
- Mathematics:

$$d_{\text{Hamming}}(x, y) = \sum_{t=1}^T \sum_{b=1}^B \mathbf{1}[x_{t,b} \neq y_{t,b}]$$

where x and y are two time-series (each $T \times B$), $\mathbf{1}[\cdot]$ is an indicator function, and t indexes frames, b indexes behaviors.

2. Interval Overlap

- Motivation: If we convert each behavior's "active times" into intervals (start,end), we want to measure how much these intervals align.
- Mathematics: For each behavior, compute the sum of overlapping intervals between two flies. The "distance" can be:

$$d_{\text{Overlap}} = 1 - \frac{\text{Total Overlap}}{\text{Total Union}}$$

where "Total Overlap" is the summed intersection of intervals, and "Total Union" is the combined duration minus overlap.

3. Hidden Markov Model (HMM)

- Motivation: Behaviors can transition from one state to another over time. HMMs capture these transitions probabilistically. This approach helps to notice differences in patterns. It is used in the analysis of biological sequences, in particular DNA
- Mathematics: Each fly's data is converted into discrete states or multi-hot states. Train an HMM by maximizing the likelihood of state transitions:

$$P(\text{observations} \mid \text{HMM parameters}) \rightarrow \max$$

Compare two HMMs (A and B) by scoring how well model A explains B's data (and vice versa) and combine these likelihoods into a distance measure.

- Instead of recalculating the Markov probabilities repeatedly, a Markov chain in advance for each fly based on its behavior was created and stored in pickle files and reused, which allows the code to work much faster.

4. Dynamic Time Warping (DTW)

- Motivation: Allows time-series that may be "out of phase" or "stretched" to align.
- Mathematics: Recursively defines a warping path
- π to minimize cumulative Euclidean cost:

$$d_{\text{DTW}}(x, y) = \min_{\pi} \sum_{(i,j) \in \pi} \|x_i - y_j\|$$

5. Euclidean Distance (Baseline)

- Motivation: Straight-line distance in the $T \times B$ space.
- Mathematics:

$$d_{\text{Eucl}}(x, y) = \sum_{t=1}^T \sum_{b=1}^B (x_{t,b} - y_{t,b})^2$$

2.3 Evaluation Metric

We have measured by different methods:

- Intra-Group Distance: Average distance among flies within the same group.
- Inter-Group Distance: Average distance among flies from different groups.

We have compared values of this distances according to the ratio:

$$\text{Ratio} = \frac{\text{Intra-Group Distance}}{\text{Inter-Group Distance}}$$

where a lower ratio suggests better group discrimination (according to our logic distance among flies within the same group should be less than distance among flies from different groups - this will show bigger similarity of flies from the same group than flies from different groups). We have calculated values of ratio for all methods with aim to compare them not only qualitatively, but also quantitatively.

3. EXPERIMENTAL RESULTS

3.1 Dataset Findings

We used data from two groups - Females Singles and Males Grouped - each from 5 different experiments, we only compared "intra-group" flies to each other from different experiments (because the fly recognition system has inaccuracies and sometimes it confuses flies with each other). We used different limits on the number of within-group and between-group comparisons for different algorithms because they all have very different computational complexities.

The sample sizes vary per method:

- Hamming & Euclidean: 2000 intra-group and 2000 inter-group pairs
- Interval: 2000 intra-group and 2000 inter-group pairs
- Markov: 200 intra-group and 200 inter-group pairs
- DTW: 100 intra-group and 100 inter-group pairs

Example of computations of distance measurements for HMM

Group	Robust Similarity	Normalized similarity
Females_Grouped	700046.906	100.000
Females_Singles	432185.350	61.737
Males_Singles	448506.418	64.068
Males_Grouped	523179.414	74.735
Overall_Females	548455.979	78.346
Overall_Males	498211.670	71.168

Each method's computation time is also included.

Method	Avg Intra Group	Avg Inter Group	Ratio (Intra/Inter)	Time (min)
Hamming Distance	0.609	0.616	0.984	7
Interval Overlap	0.441	0.457	0.964	11
Hidden Markov Model	0.050	0.199	0.251	25
Dynamic Time Warping	0.501	0.392	1.279	22
Euclidean	0.652	0.666	0.978	6

HMMs displayed the highest ratio, indicating clear distinction across groups. But that doesn't tell us anything right now, because due to computational complexity, we tested this on too small a sample of flies. DTW performed poorly, likely because time distortions here do not correspond to meaningful behavior changes.



Figure 4: Intergroup similarity, using grooming as an example

3.2 Additional Correlation Checks

We also observed higher correlations within the same group than across different groups, supporting our distance-based findings.

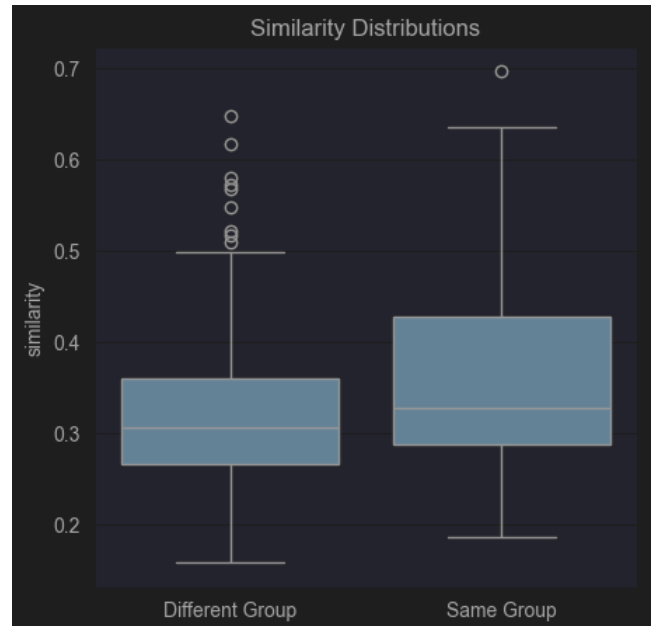


Figure 5: Similarity is measured using Pearson correlation.

Also we checked the Logistic Regression Model. The logistic regression model provides a quantitative relationship between similarity and group membership:

Coefficient $\beta=2.0023$: Higher similarity strongly increases the probability of flies being classified in the same group.

4. DISCUSSION

1. **Scaling:** We intend to run on larger datasets using high-performance computing to confirm these preliminary outcomes.
2. **Refining Behavior Weights:** Not all 11 behaviors are equally informative. Distinct weighting or selective subsets may enhance group discrimination.
3. **Exploring Other Algorithms:** Symbolic Fourier Approximation (SFA) may capture subtle shape variations in the time-series.
4. **Custom Similarity Measures:** Another error minimization or regression-based approaches could directly optimize for group separation.
5. **Normalization Issues:** The incorrect *scoreNorm* parameter in the dataset complicates the interpretation of score weights. We plan to correct and reevaluate.

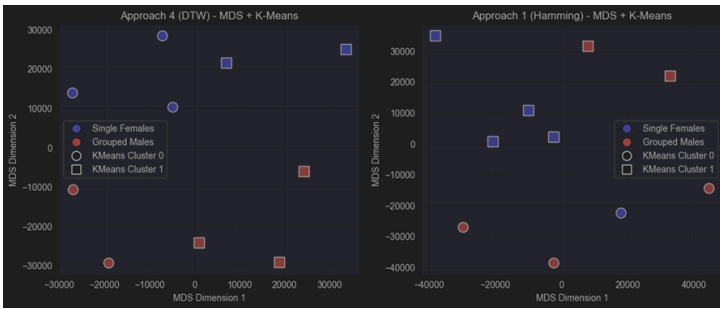


Figure 6: In the case of a small sample (10 flies) the results have little significance.

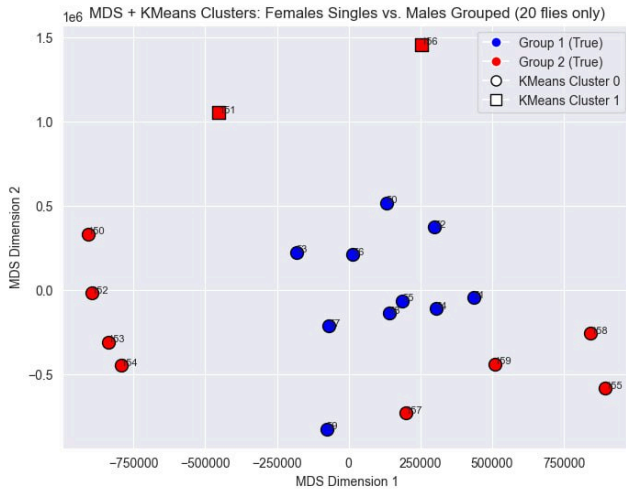


Figure 7: In the case of bigger sample we can detect significant patterns - similarity of flies from female single group and variety of flies from males grouped group (this finding agrees with results of other novel scientific works in this field)

5. CONCLUSION

We compared Hamming Distance, Interval Overlap, HMM-based distances, DTW, and Euclidean Distance on

Drosophila multi-behavior time-series. HMM stands out for discriminative power on the small test set. Future work involves comprehensive testing on larger datasets, refining weighting, implementing advanced distance measures like SFA, and resolving the data normalization problem.

References

- [1] Karima Echihabi, Kostas Zoumpatianos, Themis Palpanas, and Houda Benbrahim. The Lernaean Hydra of Data Series Similarity Search: An Experimental Evaluation of the State of the Art. *PVLDB*, 12(2): 112-127, 2018. DOI: <https://doi.org/10.14778/3282495.3282498>
- [2] John Paparrizos, Haojun Li, Fan Yang, Kaize Wu, Jens E. d'Hondt, and Odysseas Papapetrou. A Survey on Time-Series Distance Measures. December 2024. DOI: <http://dx.doi.org/10.48550/arXiv.2412.20574>
- [3] Tumilaar, K., Langi, Y., & Rindengan, A. (2015). Hidden Markov Model. *Distributed Computing*, 4, 86-94. <https://doi.org/10.35799/DC.4.1.2015.8104>.
- [4] Zhang, Z., Tavenard, R., Bailly, A., Tang, X., Tang, P., & Corpetti, T. (2017). Dynamic Time Warping under limited warping path length. *Inf. Sci.*, 393, 91-107. <https://doi.org/10.1016/j.ins.2017.02.018>.
- [5] Jasper C. Simon and Michael H. Dickinson. A New Chamber for Studying the Behavior of *Drosophila*. *PLOS ONE*, 5(1): e8793, January 27, 2010. DOI: <https://doi.org/10.1371/journal.pone.0008793>.
- [6] Gil-Martí, B., Isidro-Mézcua, J., Poza-Rodríguez, A. et al. Socialization causes long-lasting behavioral changes. *Sci Rep* 14, 22302 (2024). <https://doi.org/10.1038/s41598-024-73218-wf>