

גרסת דמו  
ניתוח סדרות זמן להערכת  
דמיון בהתנהגות זבובים  
מנתוני flyBowl



- בפרויקט הקודם, במהלך בנייה של אתוגרמות, נתקלתי בצורך לפתח שיטה מדויקת יותר להערכת הדמיון בין זבובים, על מנת לאתר את הזבוב הייצוגי ביותר.



- קראתי מטא-אנליזות שונות על אלגוריתמים לחישוב מרחקים בנתוני סדרות זמן ובחרתי ארבעה אלגוריתמים בנוסף לחישוב המרחק האוקלידי הרגיל.

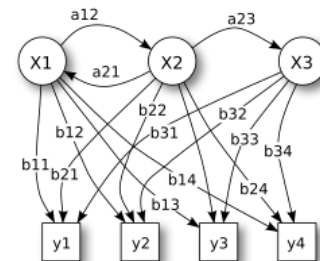
- Hamming Distance
- Interval Overlap
- Dynamic Time Warping (DTW)
- Hidden Markov Model (HMM)



- לאחר השוואה בסיסית בין האלגוריתמים, התברר כי  $DTW$  דורש חישובים כבדים מדי ואינו מספק תוצאות טובות בנתונים קטנים. לעומת זאת,  $HMM$  הציג את היחס הנמוך ביותר בין המרחק הממוצע בין זבובים בתוך אותה קבוצה למרחק הממוצע בין זבובים מקבוצות שונות.

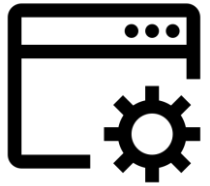
$$\text{Ratio} = \frac{\text{Intra-Group Distance}}{\text{Inter-Group Distance}}$$

# HMM

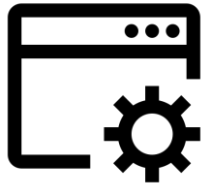


מודל מרקוב חבוי  $HMM$  הוא מודל הסתברותי המשמש לזיהוי מצבים חבויים ומעבר ביניהם על סמך נתונים נצפים. המודל מבוסס על ההנחה שכל מצב נוכחי תלוי רק במצב הקודם, וכל תצפית תלויה רק במצב הנוכחי. במהלך הלמידה, נלמדים שלושה סוגי הסתברויות: הסתברות התחלתית, הסתברות מעבר בין מצבים, והסתברות פליטה של תצפיות ממצבים חבויים. אני משתמש ב- $HMM$  לחישוב מרחקים בין סדרות זמן על ידי המרת כל סדרת זמן למודל  $HMM$  והשוואת המודלים באמצעות מדד  $KL$  Divergence סימטרי.

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



1. טוען נתוני `Scores` מקבצי `.mat`;
2. ממיר את הנתונים ל-`all_flies`: רשימה של מערכים  $(T, B)$  עם `index_map` לשמירת אינדקסי הזבובים בתוך הניסויים ובקרוב כל הנתונים;
3. מנרמל את הנתונים של כל זבוב באמצעות  
`scaler.fit_transform(fly_data)`;
4. מאמן ושומר `HMM` לכל זבוב דרך `pickle`, אם עדיין לא נשמר;



5. בונה מטריצת מרחקים לכל הזבובים באמצעות

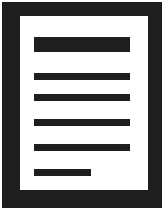
$\text{sym\_kl\_transitions}(\text{paramsA}, \text{paramsB}) +$   
 $;\text{sym\_kl\_gaussian\_emissions}(\text{paramsA}, \text{paramsB})$

6. מחשב לכל זבוב את סכום המרחקים לכל שאר הזבובים וממין כדי למצוא את הזבובים הייצוגיים ביותר.

7. אם הזבובים מאותו *condition*, מחשב מרחק ממוצע

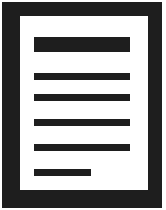
*Average Distance*; אחרת, מחשב את היחס

$\text{.ratio} = \text{mean\_intra} / \text{mean\_inter}$



1. לפי שמות התיקיות, התוכנית מבינה מתי הזבובים הם מאותו *condition* או מתנאים שונים.
2. כל המרחקים הבין-קבוצתיים האפשריים נשמרים במערך ומחושב הממוצע *mean\_inter*.
3. כל המרחקים התוך-קבוצתיים האפשריים לכל *condition* בנפרד נשמרים במערכים, (כולל מרחקים בין זבובים מניסויים שונים אך מאותו *condition*) לכל מערך מחושב *mean\_intra*.
4. עבור כל ממוצע תוך-קבוצתי מחושב  $ratio = mean\_intra / mean\_inter$  ; מחושב *avg\_ratio*.





- אם כל הזבובים הם מאותו condition (גם אם מניסויים שונים ומתיקיות שונות), התוכנית מנרמלת את כל המרחקים הזוגיים לטווח [1, 100] באמצעות המקסימום והמינימום, ולאחר מכן מחשבת את הממוצע של כל המרחקים הזוגיים.
- בעיות: אם מריצים את התוכנית בנפרד עבור כל condition, לא ניתן יהיה להשוות בין הממוצעים בגלל הנרמול.

$$\text{norm\_vals} = 1 + 99 \times \frac{\text{vals} - \text{min\_val}}{\text{max\_val} - \text{min\_val}}$$



3. Num of most representative

Enter how many flies to pick:

OK Cancel

1. Data load type

Select Input Mode

Do you want a single condition or multiple conditions?

Single Condition Multiple Conditions

3. Num of most representative

Enter how many flies to pick:

OK Cancel

4. Single folder selection

Select .mat Files (One Condition)

File name: "scores\_Grooming.mat" "scores\_Jump.mat" "scores\_Long\_Distance\_Approach.mat"

Open Cancel

2. Main menu

Select Approach

Choose an approach:

Approach 1: Multi-Hot Weighted Hamming

Approach 2: Interval Overlap Distance

Approach 3: Multi-Hot Markov Model

Approach 4: Numeric Scores + Euclidean/DTW

4. Multiple folders selection

Drag and Drop Folders

Drag and drop folders here

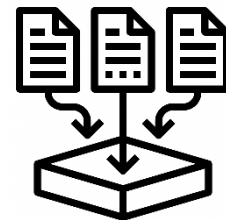
5. Results for multiple folders

Building a cross-condition distance matrix with total 200 flies.  
Final cross-condition distance matrix shape: (200, 200)  
HMM-based distance matrix built.

=== Most Representative Flies ===

```
1. Global Fly=2, Condition=Assa_Females_Grouped_Unknown_RigA_20220715T105702, Local Fly=2, DistSum=-362392804450813632.00, Score=100.00
2. Global Fly=133, Condition=Assa_Males_Grouped_Unknown_RigA_20220707T084623, Local Fly=3, DistSum=22146229587375724.00, Score=1.83
3. Global Fly=77, Condition=Assa_Females_Singles_Unknown_RigA_20220714T111552, Local Fly=7, DistSum=23186637000556020.00, Score=1.56
4. Global Fly=43, Condition=Assa_Females_Grouped_Unknown_RigA_20220715T114244, Local Fly=3, DistSum=25349120787496876.00, Score=1.01
5. Global Fly=86, Condition=Assa_Females_Singles_Unknown_RigA_20220714T112714, Local Fly=6, DistSum=26469878367395716.00, Score=0.72
6. Global Fly=129, Condition=Assa_Males_Grouped_Unknown_RigA_20220707T083144, Local Fly=9, DistSum=27704358896131236.00, Score=0.41
7. Global Fly=92, Condition=Assa_Females_Singles_Unknown_RigA_20220714T113619, Local Fly=2, DistSum=28343957190570704.00, Score=0.25
8. Global Fly=139, Condition=Assa_Males_Grouped_Unknown_RigA_20220707T084623, Local Fly=9, DistSum=28595748812516392.00, Score=0.18
9. Global Fly=74, Condition=Assa_Females_Singles_Unknown_RigA_20220714T111552, Local Fly=4, DistSum=29014464992581504.00, Score=0.07
10. Global Fly=108, Condition=Assa_Males_Grouped_Unknown_RigA_20220707T080935, Local Fly=8, DistSum=29307027235815884.00, Score=0.00
```

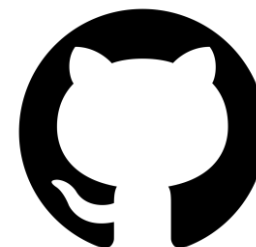
Overall average ratio (intra/inter): 0.4270



## שלבים הבאים

1. לבדוק האם נירמול הנתונים מתבצע בצורה אופטימלית.
2. לשפר את האלגוריתם לבדיקת דיוק המרחקים ולהתחשב ב-  
*edge cases*.
3. לנסות לאמן את המודל עם הגדרות שונות ולהעריך כיצד הדיוק משתנה בעקבות אימונים חוזרים.
4. לבדוק את יכולת החיזוי של המודל המרקובי.
5. ייתכן שכדאי לבדוק גם מודלים דומים נוספים.
6. להשתמש באלגוריתם שפותח לניתוח נתוני התנהגות.
7. לממש את הגישה הזו בתוך *MATLAB*.
8. ...

# GIT



## דמו פרויקט בגיטהאב עם כל התיעוד

disruptwood / Time-Series-Distance-Representative-Fly

Q Type to search

+

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

Time-Series-Distance-Representative-Fly

Public

Pin

Unwatch 1

Fork 0

Star 0

main 2 Branches 0 Tags

Go to file

Add file

<> Code

disruptwood Updated paper style explanation of work a82131f · 2 weeks ago 6 Commits

.idea	Updated project and removed obsolete files	2 weeks ago
.ipynb_checkpoints	Updated project and removed obsolete files	2 weeks ago
data_loading	Updated project and removed obsolete files	2 weeks ago
distance_calculation	Updated project and removed obsolete files	2 weeks ago
interface	Updated project and removed obsolete files	2 weeks ago
DataAnalysisDemoApproachesComparison.ipynb	Updated project and removed obsolete files	2 weeks ago
DistanceApproachComparisonAndSimpleUsage.ipynb	Updated project and removed obsolete files	2 weeks ago
Paper-style_Presentation_For_Python_Project.pdf	Updated paper style explanation of work	2 weeks ago
README.md	Updated project and removed obsolete files	2 weeks ago
check_cores.py	alpha 29.01	3 weeks ago
datastructureupdatetest.py	Updated project and removed obsolete files	2 weeks ago
main.py	Updated project and removed obsolete files	2 weeks ago

Jupyter Notebook 95.3%

Python 4.7%

Suggested workflows

Based on your tech stack

Publish Python Package

Configure

SLSA Generic

Configure



# תודה על תשומת הלב

בר אילן 2025

איליה ולדימירסקי