Software Harness

Nahid Nasiri

June 30, 2024

what is software harness:

We developed software to facilitate the analysis of the fidgeting data collected in our experiments. The code supports traditional feature extraction for use in statistical analyses. This section discusses the feature extraction process in detail. Additionally, the code includes a separate component for supporting machine learning analyses, which is not discussed here.

1 Data Selection

Choose the participants to be included in the study based on specific criteria such as gender (e.g., male or female subjects) or age group (e.g., young or old subjects). The data for all participants should be organized under a folder named, for example, "AllData." Within this folder, individual participant data should be stored. Each participant's data should be in a subfolder named "Fidget_Ball." In other words, the directory path should be defined as follows: `dirs = glob('*/*/Fidget_Ball')`.

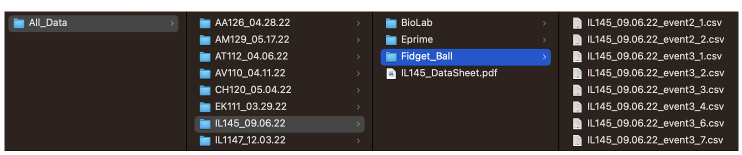

Figure 1: How does the data look like

2 Feature Extraction Description

The script processes data from a fidget ball outfitted with sensors to identify distinctive fidgeting behaviors. The analysis targets several key behaviors: Moderate versus Intense Fidgeting, Infrequent versus Frequent Fidgeting, Light versus Heavy Fidgeting, Minimum-Maximum Intensity, Fidgeting versus Non-Fidgeting, and Inactive versus Active Fidgeting. Detailed explanations of each feature extraction method are provided in the subsequent sections.

These feature extractors, implemented in our software, help dissect the sensor data into meaningful insights. The results are formatted into structured tables and visualized graphically for enhanced understanding. Additionally, the framework is designed to accommodate further analytical methods and visualizations as needed.

3 Procedure for Using the Feature Extraction Code

1. To run the analysis script for extracting features from the fidgeting data, follow these steps:

Install dependencies:

```
pip install -r Requirements.txt
```

2. Prepare your data: Ensure your data is in CSV format and located in the specified directory.

3. Run the script:

```
python main.py
```

4. Using PyCharm: Upon initiating the script by running it in PyCharm, as shown in Figure 2, users can select the type of feature extractors from a predefined list tailored to various hypotheses about fidgeting behavior.

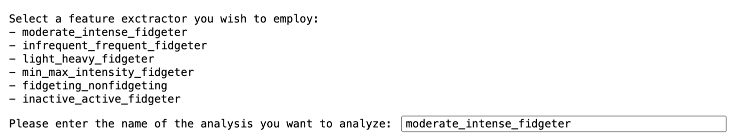

Figure 2: First step after running the script

Then, select the event files corresponding to the data you wish to analyze, as shown in Figure 3. Examples of event files include `*event3_5.csv`, `*event6_4.csv`, `*event10_2.csv`, etc.

Output

```
1) Tabular Results: Each analysis generates a detailed CSV file summarizing the extracte
2) Graphical Results: Visual representations like bar graphs and scatter plots illustrat
```

This setup ensures a clear, organized approach to examining complex behaviors captured by the sensor-equipped fidget ball.

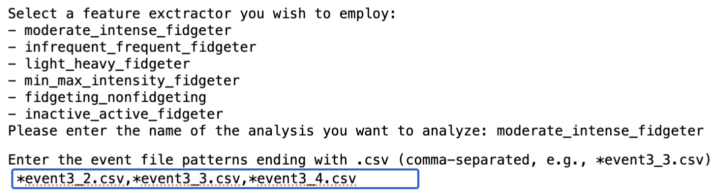

Figure 3: Second step after running the script

4 Procedure for Selecting Participants by Analyzing Data Collected from Fidget Ball

4.1 what are the feature extractors:

The fidget ball, designed with six pressure sensors, captures the pressure the participant applies. The data from these sensors are collected and analyzed to understand different aspects of fidgeting behavior. The features are designed to identify patterns and categorize users based on their fidgeting habits.

Feature 1: Identifying Moderate to Intense Fidgeting

This analysis identifies participants who exhibit moderate to intense fidgeting by calculating the average sensor values and sorting them to distinguish varying levels of engagement. This quantitative measure is crucial for understanding the intensity of fidgeting behaviors exhibited by the participants.

Folders with no fidgeting activity (zero average sensor values) are identified and excluded from detailed analysis to focus on those with significant interactions. For each participant with notable sensor data, the average values are computed and sorted from moderate to intense fidgeting based on a specified threshold.

The detailed sensor data for each folder is recorded, including sums, counts, and averages per sensor, and then compiled into a CSV file for further analysis or integration into predictive models. This allows for a multi-dimensional view of fidgeting behavior at the individual sensor level.

During the analysis, the sorted average values are visually represented in a color-coded bar, using a gradient to denote the transition from moderate to intense fidgeting. The use of a color bar enhances the interpretability of the data, clearly marking the threshold point where fidgeting behavior is considered intense.

This method not only categorizes fidgeting intensity but also adds depth to the behavioral analysis by quantifying the engagement level of each participant.

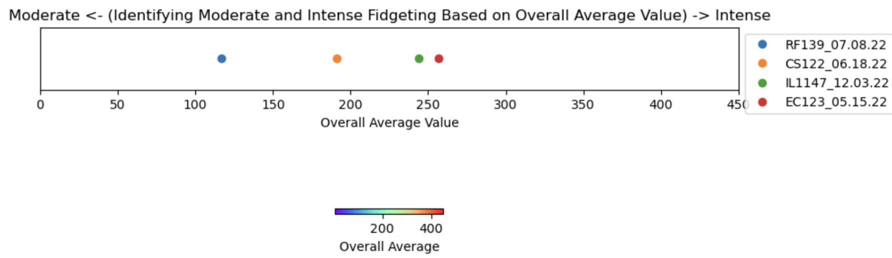

Figure 4: Average of sensor values above threshold for moderate to intense fidgeting experiment

Feature 2: Identifying Infrequent to Frequent Fidgeting

This analysis classifies participants as infrequent or frequent fidgeters based on the number of times their sensor readings exceed a specified threshold. It offers a quantitative measure of fidgeting frequency, distinguishing participants based on the regularity of their fidgeting activities.

Folders with no significant sensor readings are excluded to focus the analysis on participants who have engaged with the fidget ball. For each of the remaining folders, participants are categorized as 'Frequent' if their line count above the threshold is high. Otherwise, they are labeled as 'Infrequent.'

The categorization results are organized and saved in a CSV file, facilitating further statistical analysis or integration into behavioral prediction models. Additionally, the frequency data is visually represented through a horizontal bar chart to enhance interpretability.

The following figure illustrates these distinctions, clearly depicting the frequency of fidgeting events above the threshold for each participant, thereby providing a clear visual differentiation between infrequent and frequent fidgeting behaviors.

This approach not only clarifies the extent of fidgeting activity among participants but also aids in understanding the patterns of engagement with the fidget ball.

Feature 3: Identifying Light to Heavy Fidgeting

This feature categorizes participants into Light, Medium, and Heavy fidgeters based on the cumulative sensor values exceeding a predetermined threshold. The classification is determined by the sum of these magnitudes, which quantifies the intensity of fidgeting behavior.

Each participant's data folder is analyzed, with folders recording zero activity excluded from further evaluation. For the remaining folders, the total magnitudes are calculated and sorted in descending order to prioritize those with the highest fidgeting activity.

A horizontal bar chart is generated to visually represent these sums, color-

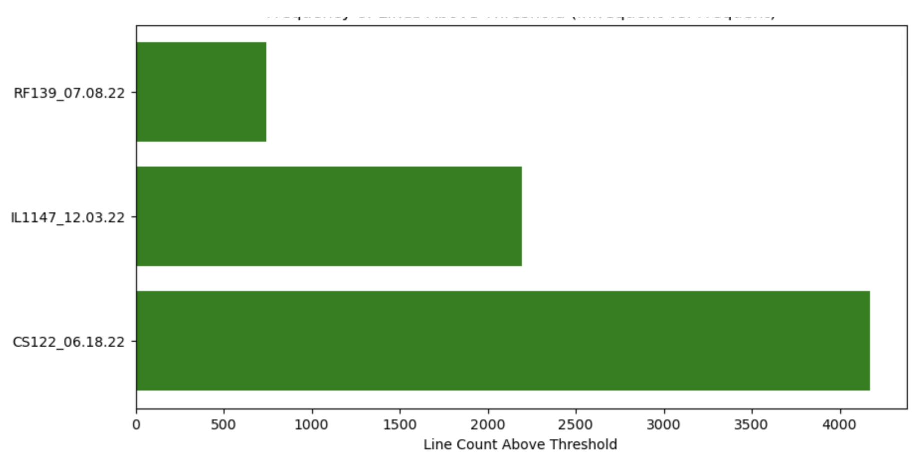

Figure 5: Results of the infrequent to frequent fidgeting experiment

coded by fidgeting category: Light (blue), Medium (orange), and Heavy (red), as depicted in Figure 6. This visual aid provides an intuitive understanding of the varying fidgeting intensities among participants.

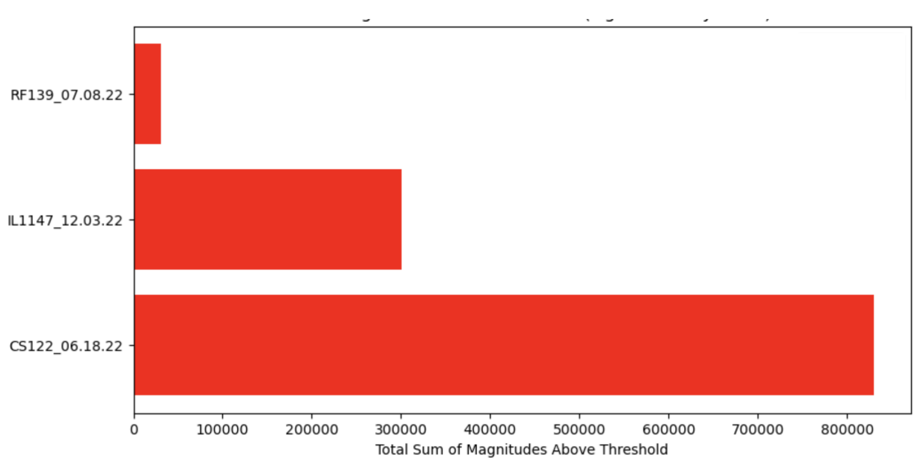

Figure 6: Results of the light to heavy fidgeting experiment

The calculated sums and their corresponding categories are also saved in a CSV file, excluding the categorization label, for potential use in further analyses or machine learning applications. This approach not only identifies the intensity of fidgeting behavior but also facilitates a detailed statistical analysis of fidgeting patterns.

Feature 4: Identifying Minimum-Maximum Intensity

This feature aims to analyze the minimum and maximum intensity of fidgeting by identifying the lowest and highest sensor values above a specified threshold. This analysis provides insight into the range of fidgeting intensities experienced by participants.

Folders with no sensor values above the threshold are excluded from the analysis. The sensor values are sorted for the remaining folders, and the averages of the top 10% and bottom 10% of these values are calculated to represent maximum and minimum fidgeting intensities, respectively.

The results of this analysis are encapsulated in a CSV file and include detailed listings for each folder, specifying the average values for both the highest and lowest intensity ranges as shown in Table 1. This method effectively highlights the dynamic range of fidgeting behaviors across the study's participants. The CSV output facilitates further statistical evaluation or machine learning applications by providing a quantified measure of intensity variability.

Participant	Top 10% Average	Bottom 10% Average
CS122_06.18.22	694.76	4.00
IL1147_12.03.22	302.12	4.00
RF139_07.08.22	177.30	5.00

Table 1: Top and bottom 10% averages for participants

This analytical approach not only distinguishes between participants based on their fidgeting intensity but also adds depth to the behavioral analysis by quantifying the extremes of their interactions with the fidget ball.

Feature 5: Identifying Fidgeting versus Non-Fidgeting

This analysis distinguishes between fidgeting and non-fidgeting activities by counting the number of sensor data points that fall above and below a certain threshold. This quantitative approach provides a clear distinction between active and inactive periods, facilitating a deeper understanding of participant behavior during the study.

Folders that show no activity (zero average sensor values) are excluded to focus the analysis on participants with detectable interactions. For each participant folder included in the study, the number of lines above the threshold (indicating fidgeting) and the number of lines below the threshold (indicating non-fidgeting) are calculated.

The results are compiled into a CSV file, providing a straightforward comparison of fidgeting versus non-fidgeting data points for each participant as shown in Table 2. This data format supports easy integration into statistical software for further analysis or machine learning models to predict behavior patterns.

This feature effectively captures the dynamic nature of fidgeting behavior, offering valuable insights into the balance of activity and inactivity among the participants.

Participant	Fidgeting (above threshold)	Not Fidgeting (bellow threshold)
CS122_06.18.22	4170	107484
IL1147_12.03.22	2195	55
RF139_07.08.22	740	2148

Table 2: Fidgeting vs. non-fidgeting activity

Feature 6: Identifying Inactive versus Active Fidgeting

This feature evaluates the transition patterns between inactive (non-fidgeting) and active (fidgeting) states by analyzing the sequence lengths of these behaviors. The analysis highlights the dynamics of fidgeting activity, offering insights into the periodicity and intensity of interactions with the fidget ball.

Excluding folders with no significant sensor activity, the function computes the average sequence lengths for 'x' (active fidgeting) and 'o' (inactive periods). This calculation provides a detailed view of how long participants typically engage in or refrain from fidgeting.

The sequence lengths are sorted by the average length of active fidgeting to emphasize participants with more sustained interactions. The results are presented in tabular form, saved as a CSV file, and visually through a bar chart, facilitating a comparative analysis across participants.

The following figure illustrates the average sequence lengths, enhancing the visual comprehension of the data.

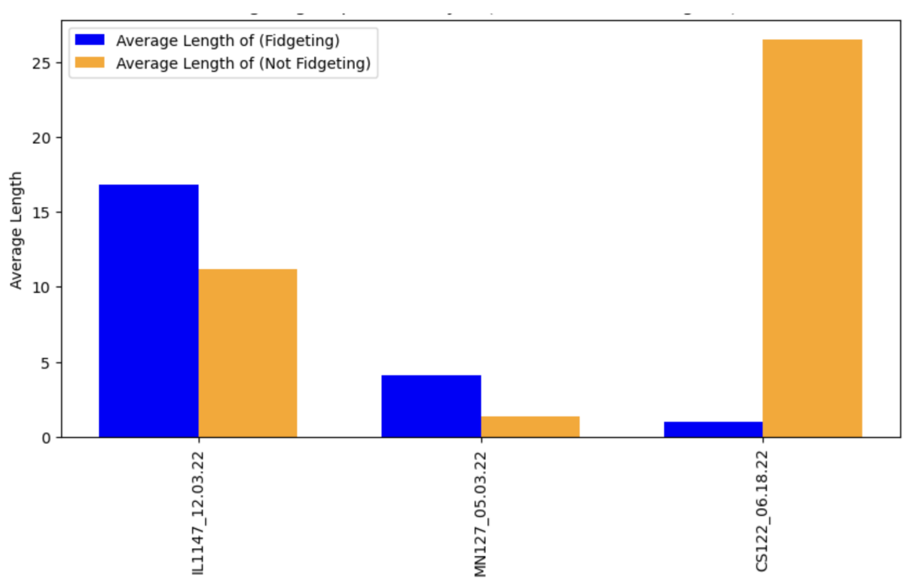

Figure 7: Results of the inactive vs. active fidgeting experiment.

This analysis is important for understanding the behavioral patterns of fidgeting, indicating not only the frequency and duration of fidgeting episodes but also the rest periods, thereby offering a comprehensive view of participant behavior during the study.

By analyzing the results from these feature extractors, we can effectively identify which participant data are most suitable for further analysis in our machine-learning studies.