Detection and Analysis of Changes in Everyday Physical Activity Data

Gina Sprint and Diane Cook

School of Electrical Engineering and Computer Science

Washington State University

CASAS Center for Advanced Studies in Adaptive Systems

Overview

Physical Activity

Physical activity (PA) consists of bouts of movement that are separated by periods of rest. Measurements of PA include [1]:

- Frequency.
- Duration.
- Intensity.
- Activity type.

Tracking Change in Physical Activity

Many consumers purchase a wearable fitness device to track their PA, commonly in pursuit of a goal. Often, self-perception and direct measurement of physical activity are not congruent [2]. To address this, we propose Physical Activity Change Detection (PACD) to objectively detect progress toward goals and/or health events.

Datasets

Hybrid-synthetic (HS) Dataset

Real Fitbit data re-sampled to produce five synthetic profiles:

- HS0: No significant change (a baseline for "no change").
- HS1: Gradual increase in day-to-day PA intensity.
- HS2: Significant PA intensity change after 6 days.
- HS3: Gradual increase in PA variability.
- HS4: Significant PA variability change after 6 days.

B-Fit (BF) Health Intervention Dataset

- 10-week intervention study to improve health.
- Participants set goals for 8 health categories including exercise, cardiovascular risk factors, and nutrition.
- Fitbits assesed PA 6 days before and after intervention
- 11 older adults (57.09 ± 8.79 years) participated.

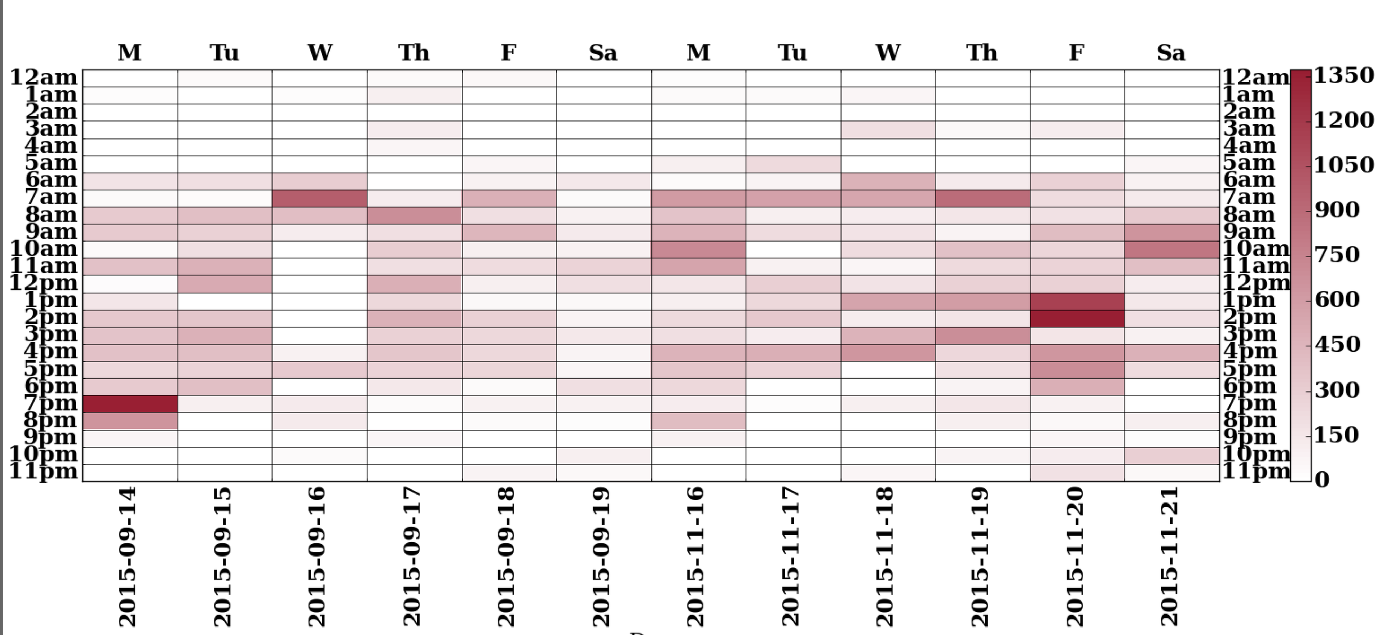

Figure 1. Participant BF1's step data. Daily physical activity (steps taken) is plotted as a function of day (X-axis) and 24 hour time (Y-axis).

References: [1] Caspersen et al., 1985. [2] Prince et al., 2008. [3] Liu et al., 2013. [4] Hido et al., 2010. [5] Refinetti et al., 2007.

Physical Activity Change Detection

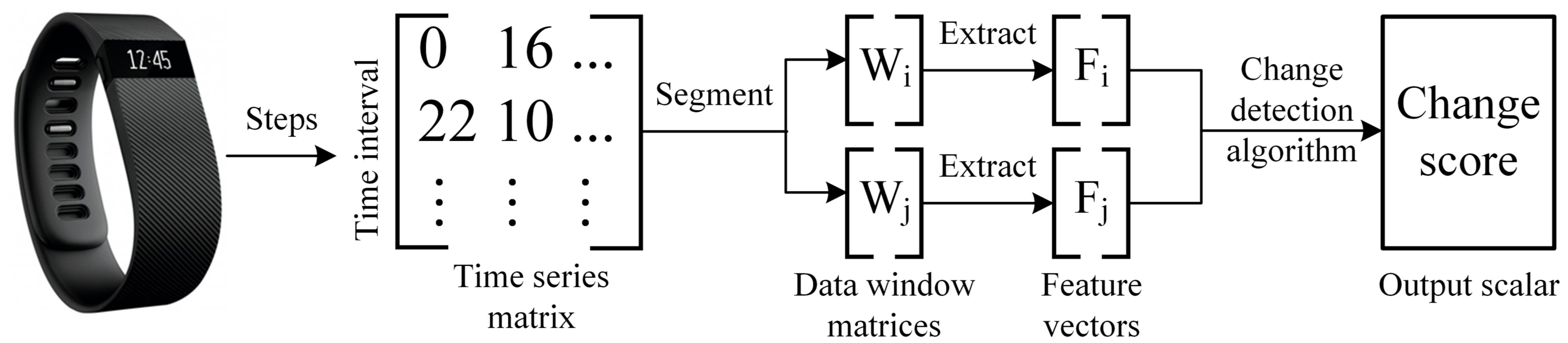

Figure 2. PACD data processing. Step data is collected from a wearable device such as a Fitbit. Several algorithms process the time series data to produce a change score representing the detected change between two windows of data.

Physical activity data is represented by a matrix of time intervals for each day. The data is segmented into windows, W_i and W_j, where i and j denote the start of each window. For the HS and BF datasets, $i = 1$ and $j = 7$.

Features such as number of walking bouts and average rest period duration are extracted from each window. The features serve as inputs to change detection algorithms, such as RuLSIF [3], virtual classifier [4], and our proposed sw-PCAR algorithm. If the score is significant, change analysis is performed to inspect and explain the source of change.

Algorithm 1 PACD(X, n, $offset$, adv_i, adv_j)

```
Input: X = time series data
Input: n = window length in days
Input: offset = number of days separating windows
Input: adv_i = number of days to advance the first window
Input: adv_j = number of days to advance the second window
Output: V = vector of change scores
Initialize: i = 1 and j = 1 + offset
for each pair of windows to compare, W_i and W_j of time series X:
    W_i = X[i : i + n − 1]
    W_j = X[j : j + n − 1]
    Compute CS = F(W_i, W_j)
    Determine if CS is significant
    Identify the type of change that is exhibited
        Manual inspection of change
        Unsupervised inspection (change analysis)
    Append CS to change score vector V
    i = i + adv_i
    j = j + adv_j
end for
return Change score vector V
```

Figure 3. The PACD algorithm.

Physical Activity Change Analysis

For a change score CS to be significant, we test that the magnitude of change (inter-window change) exceeds the day-to-day variability [5] within each window (intra-window change). See Figure 4 for an example. We generate a list of all possible daily change scores, DCS, within each window. Next, outlier detection determines if CS is an outlier of DCS. If CS is significant, we analyze features and inspect a decision tree learner to reveal the source of change.

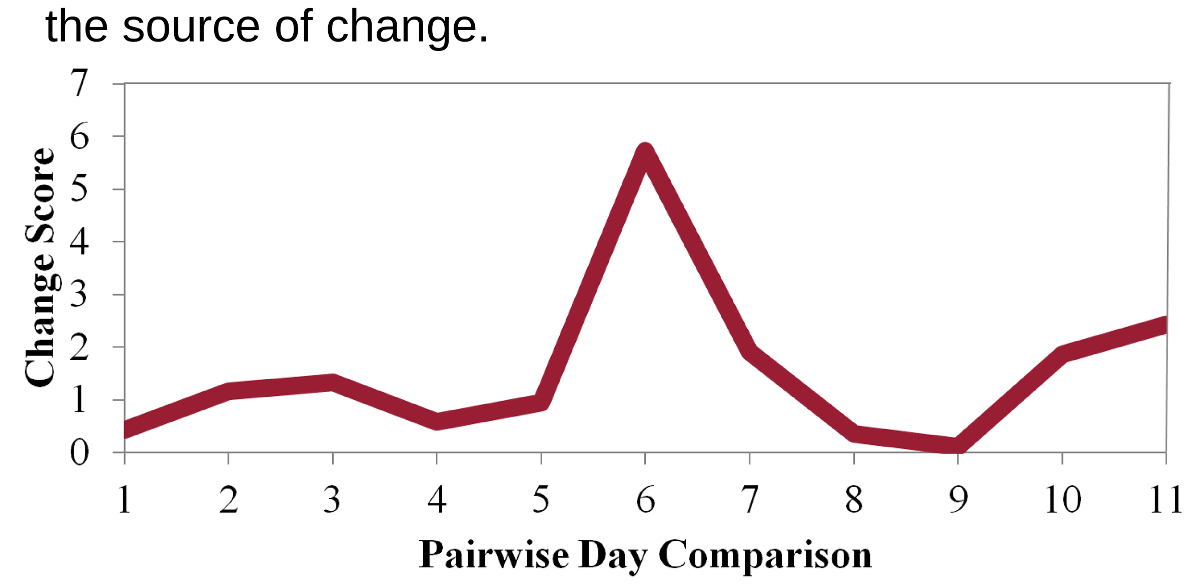

Figure 4. Participant BF1 sw-PCAR change scores. Largest change occurs between the last day of W_1 and first day of W_7 (6th comparison).

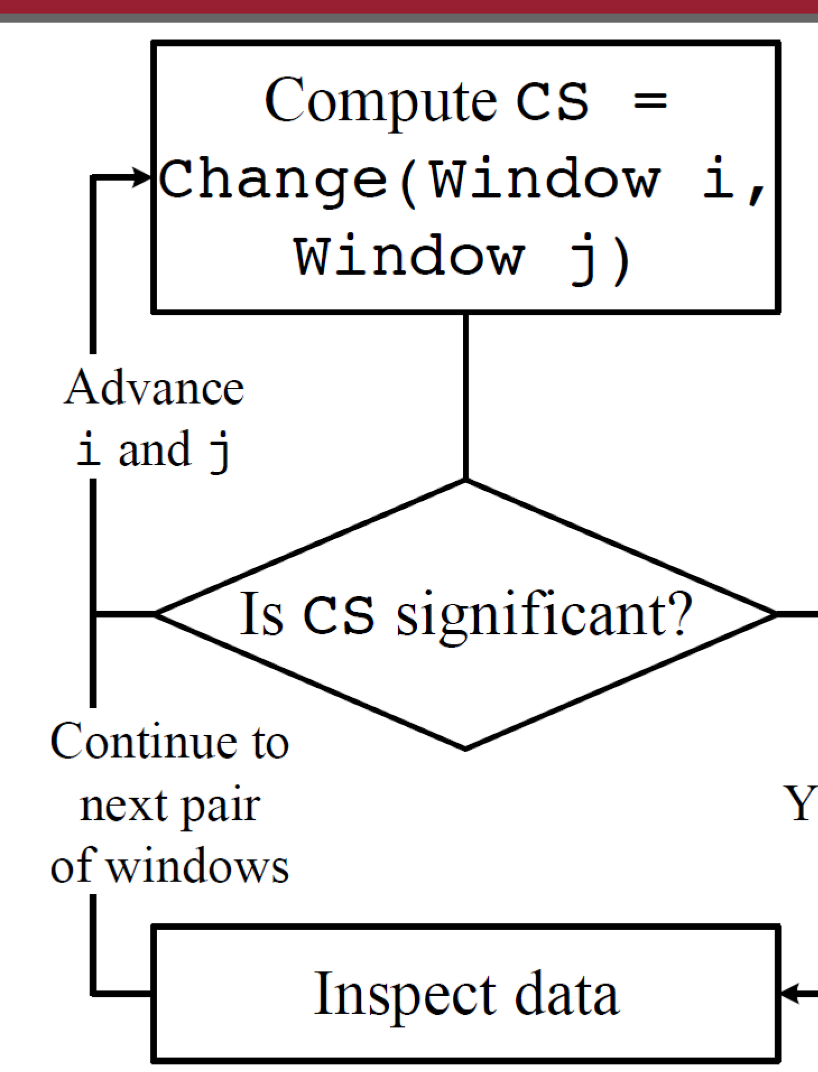

Figure 5. Change analysis. When a significant change is detected, the data is inspected to identify the source of change.

Results

Hybrid-synthetic (HS) Dataset

The most changes are detected for time interval size of 5 minutes (tested: 1, 5, 10, 15, …, 60 minutes).

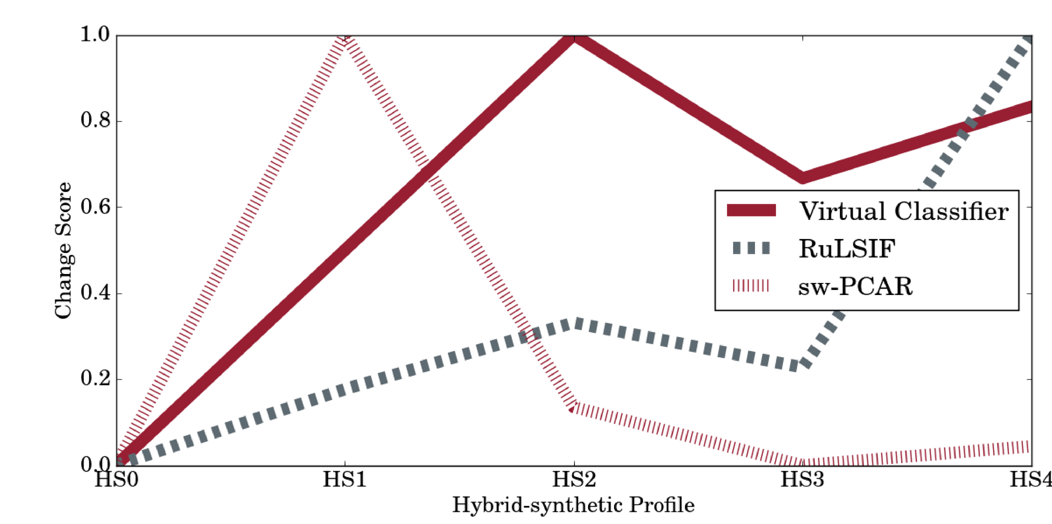

Figure 6. Normalized hybrid-synthetic results.

B-Fit (BF) Health Intervention Dataset

- Participant BF3 stated she met her exercise goal of walking more; however, the computed features show decreased PA (see Figure 7): average number of bouts (pre: 81.00, post: 15.83), daily steps (pre: 4279.50, post: 1161.44 steps), and percentage of time sedentary (pre: 86.30%, post: 97.44%).
- Participant BF29 exhibited progress toward her goal of walking more by increasing average daily steps from 1136.51 steps to 1210.85 steps post-intervention testing, a 6.54% increase (see Figure 8, relative amplitude represents the daily ratio of the most active 8 hours to the least active 4 hours).

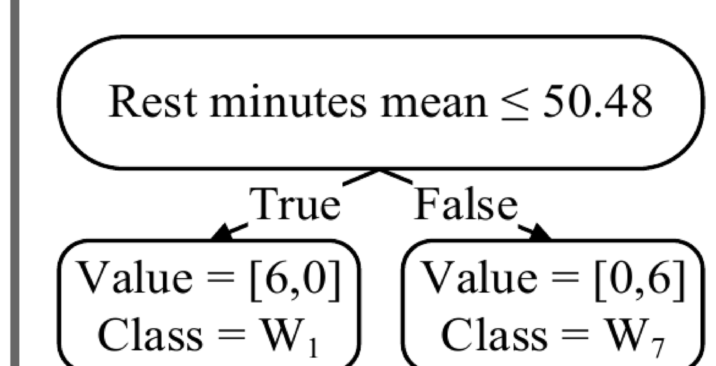

Figure 7. BF3 Decision tree.

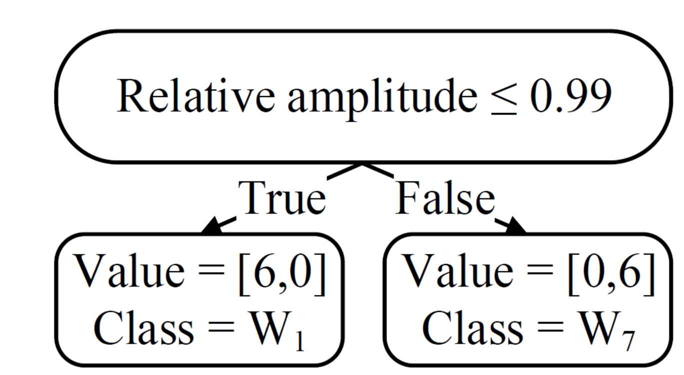

Figure 8. BF29 Decision tree.

Conclusions

Impact

Our PACD approach objectively and automatically quantifies physical activity and changes. The methods are useful data mining techniques for monitoring/motivating physical activity.

Future Work

Future work includes performing change analysis on real-world datasets from:

- Different fitness trackers.
- Multidimensional data (e.g. heart rate, elevation, etc.).
- Longer windows of time.
- Smartphone applications.

Acknowledgements: We wish to thank Maureen Schmitter-Edgecombe, Catherine Sumida, and Thao Vo for their help with data collection. This work is supported in part by National Science Foundation grant 0900781.