Analysis of FitBit Data:

Detecting Anomalous Activity Patterns

D. Holcomb, M.S., dlh@gwu.edu

Nima Zahadat, Ph.D., nzahadat@gwu.edu

The George Washington University

Abstract

Fitness tracker research has been heavily focused on by academia in recent years, including efforts to assess fitness, sleep, heart health, general wellbeing, recuperation from medical maladies, and more. Scholars have used techniques in statistics, machine learning, deep learning, and several other fields to analyze, classify, and predict daily user behavior patterns and outliers in those patterns. In turn, this data has been used to predict and chart medical treatment reactions, encourage weight loss, enforce desired bedtimes, track general fitness, predict if students are studying enough, and give recommendations for current and future behavior to meet fitness goals, just to name a few use cases. This study will focus on finding anomalous behavior patterns within FitBit data, and then finding indicators that predict deviations from behavior baselines. Correlating these activities will be performed by using data mining techniques with Python, on a dataset of 35 users over a 60-day time period in 2016.

Keywords: data mining, exercise, fitbit, fitness, health trackers, Internet of Things, personal trackers, physical activity, tracking devices, wearables

Introduction

FitBit and other personal trackers have become increasingly popular in recent years as people have become more interested in self-monitoring their personal health. They have become a popular area of study amongst data scientists, statisticians, medical experts, physiologists, and psychologists, to name a few academic research areas. Detecting relationships in complex time series data, such as FitBit personal tracker data, can be a way of establishing daily life patterns, and also a way of detecting deviations from these patterns. With the robust FitBit API providing steps and sleep data by the day, hour, and sometimes minute, outliers can be easy to spot. Even more interesting is spotting behavior that can predict when an outlier event may occur.

Of particular interest to current research is the ability to predict future, or even very near-future, behavior, based on a user's FitBit history. As outlined in the section below, this can take the form of merely establishing baseline typical use, or can be used to tell when ill medical patients are improving or deteriorating, or can even be taken to the degree of attempting to send automated motivation coaching messages to help users avoid prior pitfalls in their tracker-related goals. These studies focus on calculating a baseline behavior profile that future activities can be predicted from, or focus on identifying atypical activities in this established baseline. This study hopes to find users with anomalous baseline behavior patterns to see how their indicators fare in behavior prediction.

Literature Review

Several studies have occurred in the growing area of personal health and fitness trackers. These studies range from basic examinations of the accuracy (Benedetto et al., 2018), (Nazari, 2017) and efficacy (Wright et al., 2017) of the trackers. Other studies have dug deeper into detecting change in activity levels, such as analyzing data where participants often forget to wear their device (Tang et al., 2016), (Purta et al.,2016), detecting anomalous sleep patterns (Liang et al., 2016), (Purta et al.,2016), classifying long-term use behavior (Meyer et al., 2017), and analyzing datasets with high variability (Valbuena et al., 2017). Within the studies looking at change detection, some have looked into analyzing and visualizing time series data (Valbuena et al., 2017), taking a statistical unsupervised learning approach to build dynamic models of human activity (Chamroukhi et al., 2013), and breaking time series data into sliding windows to calculate the significance of changes across each window to calculate importance (Sprint, 2016) (Sprint et al., 2017)

A more machine learning and data mining direction is taken with many of the studies reveiwed. One study looks into making predictions of use on an individual per-user basis (as opposed to comparing users to the entire corpus) (Phatak 2018). Another goes further and investigates using actionable, data-driven predictive machine learning models to generate automated real-time coaching tips based on a user's current and historical Fitbit activity, the most successful model using the Random Forest algorithm (Dijkhuis, 2018). A deep learning approach to determining quality of sleep used convolutional neural networks (CNN), recurrent neural networks (RNN), and long short-term memory RNNs (LSTM-RNN) (Sathyanarayana et al., 2016). Also, a clustering approach was taken on general activity tracker data to determine daily patterns (Yürüten, 2014).

To predict the learning performance during self-regulated learning, one study included FitBit heart rate and steps, along with weather, online browsing activity metrics, and historical user experience ratings, and then analyzed the data using a Linear Mixed Model (Di Mitri et al., 2017). One of the more edifying notes in this particular study was that for the LMEM, heart rate is considered fixed, as the participant cannot control it. Steps, on the other hand, are considered to be a random effect, as the participant has some say in how many steps they wish to achieve. This study was also interesting because it incorporated a second dataset, the weather data, with which to draw predictions from.

Research Methodology

A lot of research has already been completed on personal fitness trackers, as outlined above. Unlike existing research that focuses on either establishing typical patterns of use or predicting outliers, this study plans to use both baseline usage and anomalies to tip an observer as to when non-baseline behavior is most likely to occur. This study will use anonymized FitBit data collected from Mechanical Turk, consisting of 30 different users during March 12, 2016 through May 12, 2016 (Furberg et al., 2016). An example of this anomaly tipping is detecting a relationship between an unexpectedly early or late wake-up time with activity levels from the prior day, or observing how many days of sub-average high-activity moments may affect the level of workouts and high-activity moments going forward.

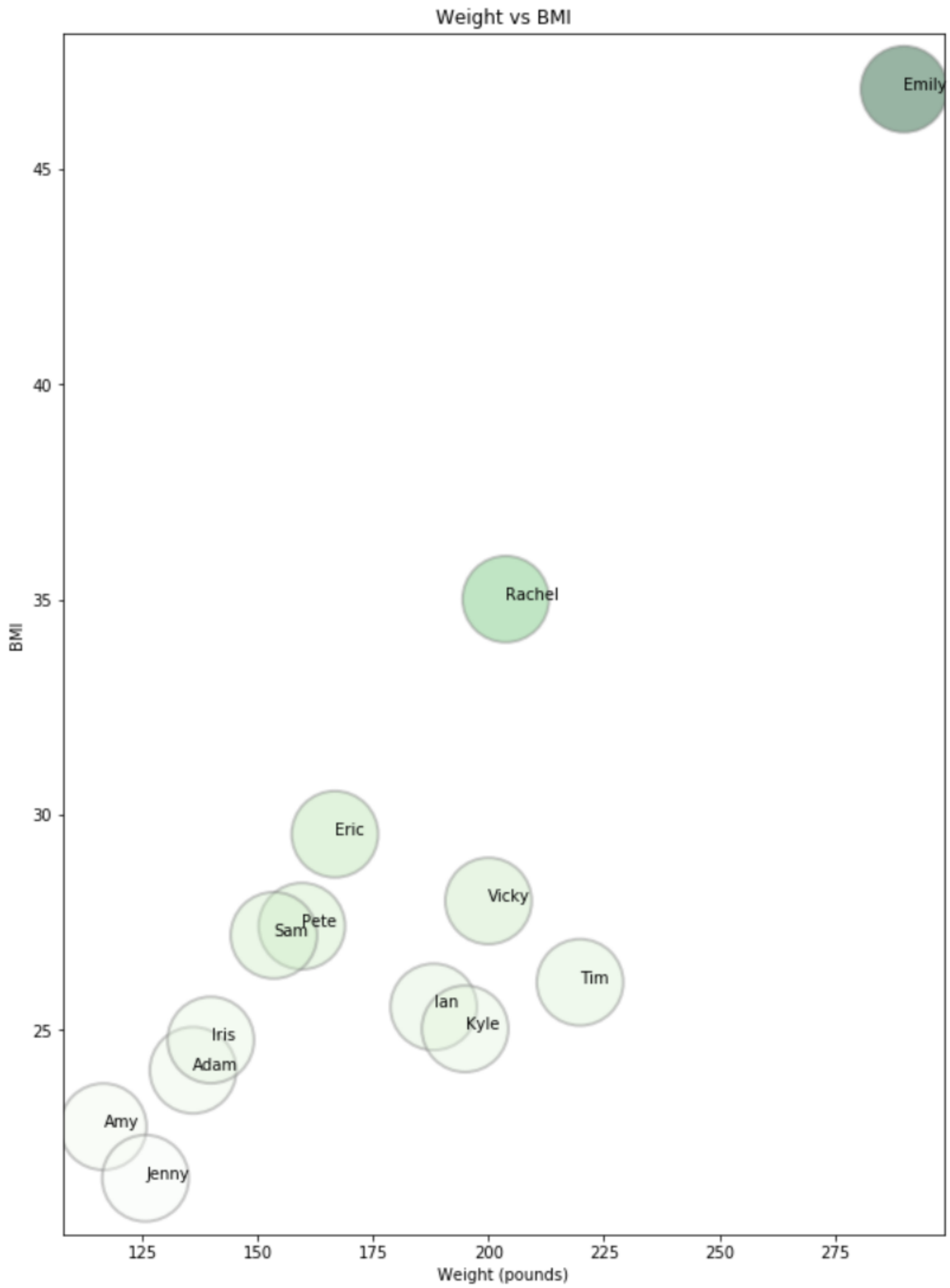
To accomplish this analysis, the study will use Python and Python libraries such as Pandas, NumPy, SciPy, and Matplotlib to determine and visualize previously uncorrelated relationships between anomalous activity patterns. Preprocessing will involve using Pandas to clean the data, delete records for unworn-FitBit days, and group the resultant data frames by both user id and date. The analysis portion, based on prior successful research, will utilize Python-based NumPy and Scikit-Learn data mining packages to establish average behavior for each user, identify outliers in their expected behavior frame, and identify and correlate related activities. The study plans to analyze users categorized into groups consisting of similar activity and sleep patterns, in the hopes of finding unique users who differ greatly from the norm. Results will then be visualized with Matplotlib.

The dataset in question has been anonymously contributed by users, potentially from across the globe. Not knowing gender, age, or location unfortunately limits the scope of analysis that can be performed. In particular, the idea of factoring in weather reports like the work done by Di Mitri et al., 2017, had been particularly appealing as an external factor to include in predictions. Traffic and public transit delay information during the specified date range could also be useful with location data, as could local and national holidays (for example, St. Patrick's Day). Unfortunately, the FitBit API is such that a user may only acquire his or her own personal data firsthand. This is understandable from a privacy perspective, but makes finding readily available data in large quantities with many users quite difficult. The data in the Mechanical Turk dataset has been anonymized strictly to health information with a user primary key – all that is available to analyze is heart rate, steps, weight, sleep, and activity levels.

Dataset

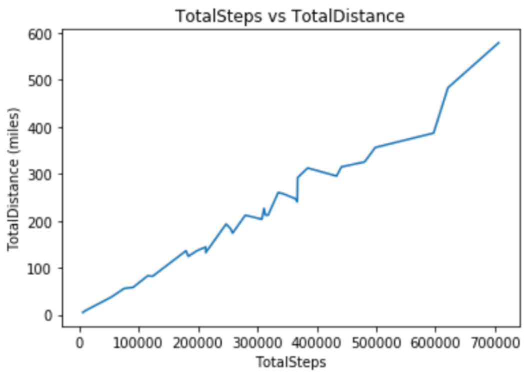
This project used Fitbit data openly available on Zenodo to analyze and correlate personal tracking data of thirty-five different users over the span of two months. The two monthly datasets, consisting of eleven and eighteen csv files respectively, were generated by respondents to a distributed survey via Amazon Mechanical Turk between March 12th, 2016 and May 12th, 2016. Thirty eligible Fitbit users participated in the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring.

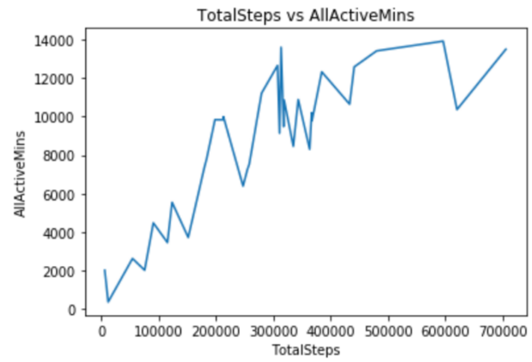
On further investigation of the dataset, only three of the spreadsheets existed across both timeframes. They were daily activity, sleep (in minutes), and weight log. These sets were merged via Pandas. Then, to make the data more human-readable and easier to visualize, thirty-five generic short names were generated to replace the long numeric ‘Id’ key in the data. Sleep data (by the minute) was combined on a per-day basis based on quality of sleep. Weight data was much more sparse than expected - only 4 users logged weight data more than 3 times in the 8-week span. To visualize this data, the weight in pounds was averaged per user. Weight data was not factored in to the rest of the data exploration due to how little of it there was.

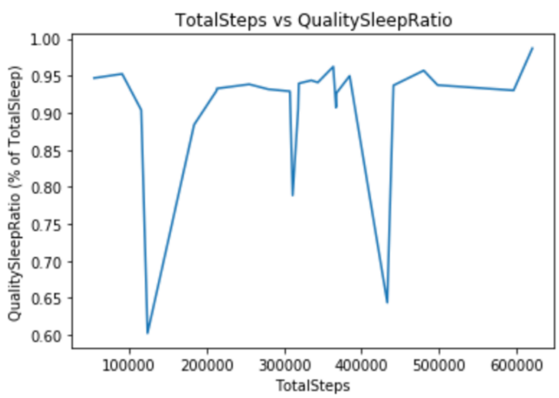


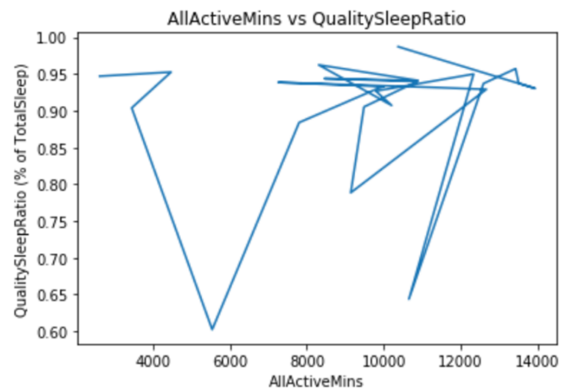
Data Analysis

In this study we moved closer to being able to identify the indicators of anomalous behavior by identifying which data is strongly correlated (TotalSteps and TotalDistance) and data which is not (AllActiveMinutes and QualitySleepRatio).

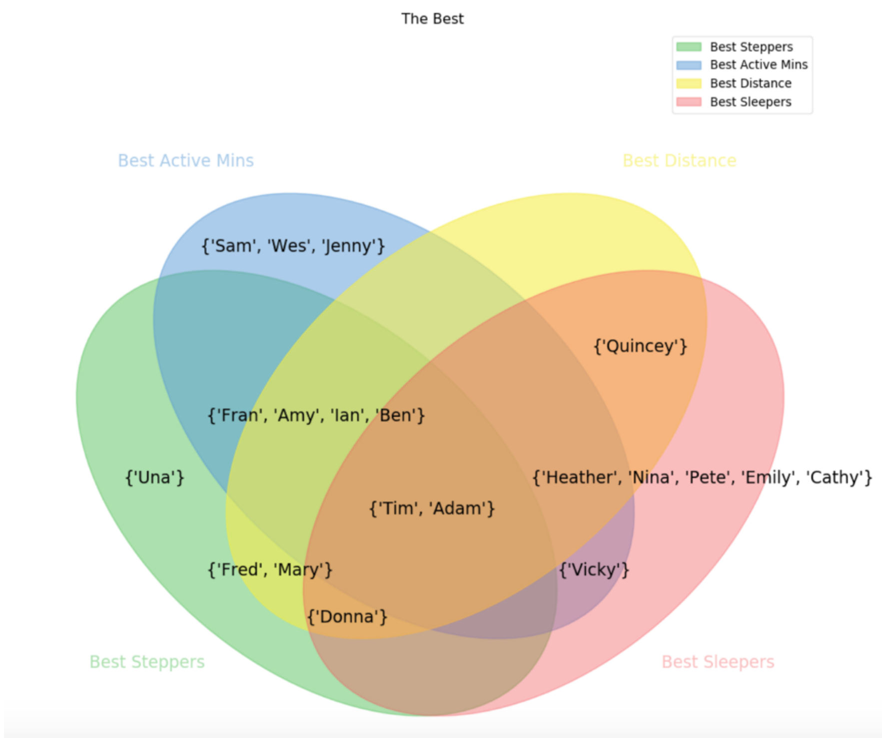


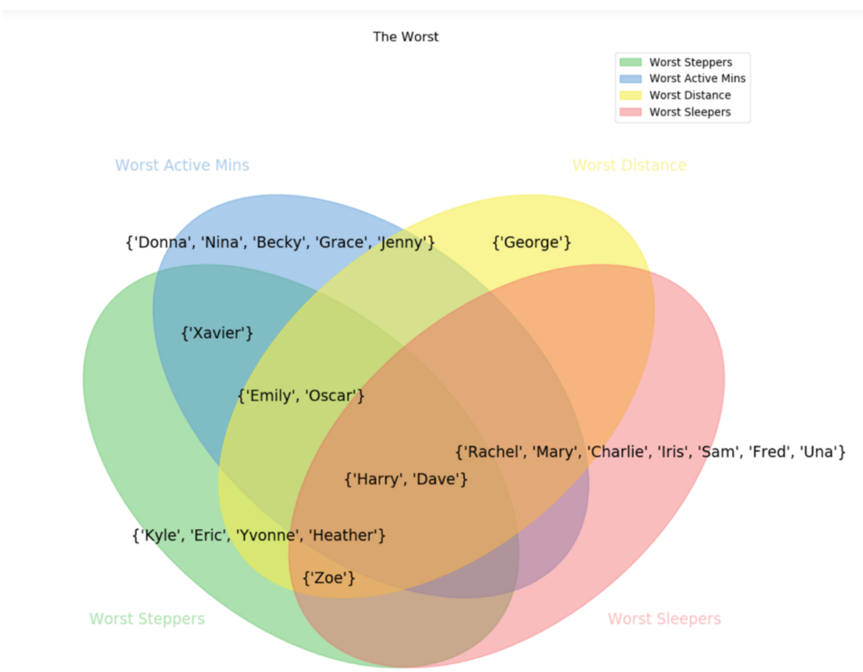




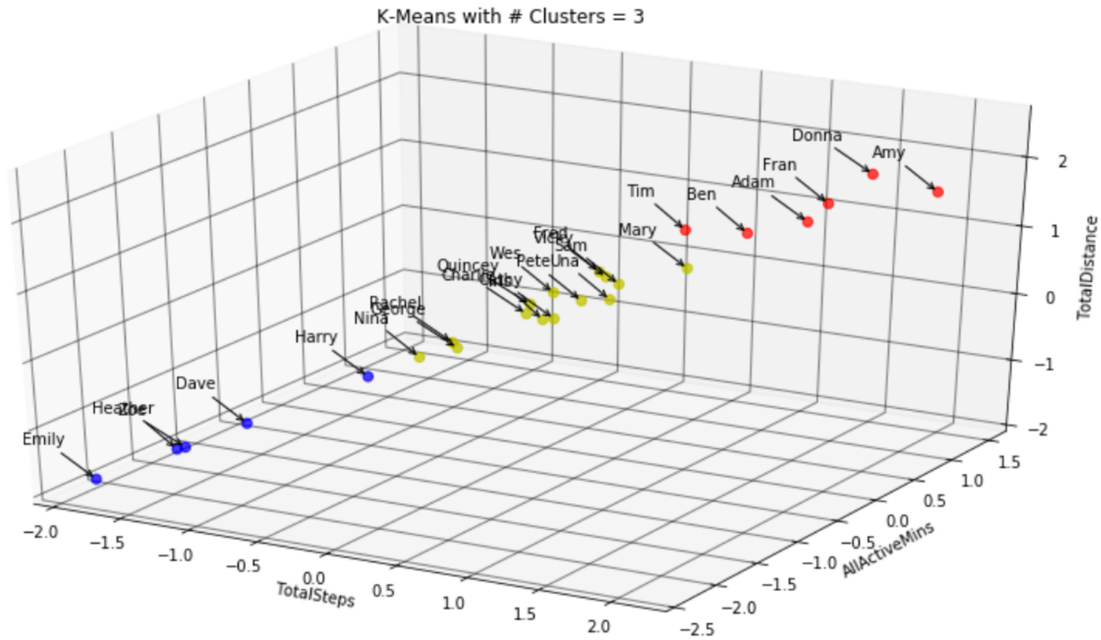


We then found the best and worst users in the categories of steps, active minutes, distance, and sleep, and compared the intersection of this groups in a Venn Diagram. This helped us to identify "interesting" users whose behavior patterns were different from the other "Best" or "Worst" users. Two of the most “interesting” users are Donna and Mary, since their names occur on both "Best" and "Worst" lists. Donna is either first or second place in steps, distance, and sleep, but ends up on the worst list for active minutes. Mary is a top stepper and distance-goer, but is the second-place worst sleeper.





Finally, we performed an unsupervised clustering via K-means on the data over the four categories to see where our "Best", "Worst", and "Interesting" users existed in relation to each other mathematically.



Key Findings

First, and most obviously, there is a strong correlation between TotalSteps and TotalDistance. This makes perfect sense, since the more a user steps, the farther they will go. Slightly more surprising, TotalSteps and AllActiveMinutes are much less strongly related. Going into this study, it was expected that any user who counted as a "Best Stepper" and a "Best Distancer" would also be a "Best Active Minutes" user. At least in this dataset, that is not the case. TotalSteps does not correlate with QualitySleepRatio, and AllActiveMinutes correlates even less well with QualitySleepRatio.

In the Venn Diagram analysis, Tim and Adam are the only users to appear in the top 10 of every category - steps, distance, active minutes, and sleep. Many of the top "movement" people appear on steps, distance, and active minutes, but don't make the top 10 sleepers. These users include Ian, Amy, Fran, and Ben. Harry and Dave are categorically the worst, appearing on the worst 10 list for steps, distance, active minutes, and sleep. Emily, Zoe, and Oscar are the worst at three of the categories, with Emily and Oscar avoiding being the worst at sleep, and Zoe avoiding being the worst at Active Minutes.

The K-means clusters show a good delineation between the "Best" and "Worst" users identified in the Venn Diagrams, and also have a middle group of "Average" users. We can see that users Tim, Adam, Ian, Amy, Fran, and Ben are all included in the "Best" cluster. "Interesting" user Donna is also in the "Best" cluster, and “interesting” user Mary is the nearest "Average" user to the "Best" cluster, excluded based on being part of the worst sleepers. Users Harry, Dave, Emily, and Zoe are all present in the "Worst" cluster. Oscar, who appeared in the "Worst" Venn Diagram, is not present in the "Worst" cluster because that user has null values for the sleep data.

Recommendations

Were this study to be continued, it would be essential to have more data. Of the available csv files, many only existed for April-May (the second month) and could not be used. At an initial glance, it appeared that weight information was also available, but that data was much too sparse. To glean even more information from the data, it would be interesting to know the gender of the participants and their geographic location - factoring in outside data such as weather and local holidays could also be illuminating in anomaly detection and locating predictive indicators. St. Patrick's Day and Cinco de Mayo may both occur in this dataset, but the users could be in a country where neither of these holidays are celebrated. With more data, it would make more sense to discard records with several NA's. The correlation analysis was thrown off by users who didn't actively participate in the study regularly - eliminating records of irregular participants would increase the fidelity of our findings. Finally, sleep data in this study was recorded on a per-day (12:00am-11:59pm) as opposed to per-sleep-session (i.e. 10:00pm-

6:00am) basis. Re-distributing the data per sleep session would lend greater value to the 'QualityOfSleep' metric.

With K-means, the same four values plotted in the Venn Diagrams were used: steps, distance, active minutes, and sleep. Unfortunately, humans can't see four dimensions, so in the visualization only three of the categories were plotted: steps, distance, and active minutes. If the timeframe and scope of this project were larger, either PCA (principal component analysis) or t-SNE (t-distributed stochastic neighbor embedding) should have been used for dimensionality reduction. In that case, there would no longer be human-understandable labels - dimensionality reduction would obfuscate the context of the axes, since they would then represent a mathematical projection from a higher dimensional space that would be a combination of all four fields.

Conclusion

FitBit data has a fascinating amount of potential uses, in this case trying to identify the indicators of anomalous behavior by identifying "interesting" users whose behavior patterns were different from the other "Best" or "Worst" users. Looking into these "interesting" users to see exactly what made them different (active minutes, sleep) identified two different anomalous behavior patterns to look out for:

1. Users who have a very high step/distance count, and a strangely low active minutes count
2. Users who excel in all active categories (step, distance, active minutes) but get very poor sleep

Future study of these patterns could reveal indicators for tipping deviations from regular behavior patterns. Finding users with anomalous baseline behavior patterns can aid in finding more interesting indicators in behavior prediction.

Biography

D. Holcomb is a graduate student in the Data Science Program at The George Washington University. Ms. Holcomb's professional career is as a deep learning and computer vision consultant with a geospatial focus.

Dr. Nima Zahadat is a professor of data science, information systems security, and digital forensics. His research focus is on studying the Internet of Things, data mining, information visualization, mobile security, security policy management, and memory forensics. He has been teaching since 2001 and has developed and taught over 100 topics. Dr. Zahadat has also been consultant with the federal government agencies, the US Air Force, Navy, Marines, and the Coast Guard. He enjoys teaching, biking, reading, and writing.

References

Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D., Pensabene, V., and Actis, P. (2018). Assessment of the fitbit charge 2 for monitoring heart rate. *Public Library of Science (PLoS)* One 13, (2) (02).

Chamroukhi, F., Mohammed, S., Trabelsi, D., Oukhellou, L., and Amirat, Y. (2013). Joint segmentation of multivariate time series with hidden process regression for human activity recognition. *Neurocomputing*. 120. 10.1016/j.neucom.2013.04.003.

Di Mitri, D., Scheffel, M., Drachsler, H., Börner, D., Ternier, S., and Specht, M. (2017). Learning pulse: a machine learning approach for predicting performance in self-regulated learning using multimodal data. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*. ACM, New York, NY, USA, 188-197. DOI: https://doi-org.proxygw.wrlc.org/10.1145/3027385.3027447

Dijkhuis, T.B., Blaauw, F.J., van Ittersum, M.W., Velthuijsen, H., and Aiello, M. (2018). Personalized Physical Activity Coaching: A Machine Learning Approach. *Sensors 2018*, 18, 623.

Furberg, R., Brinton, J., Keating, M., and Ortiz, A. (2016). Crowd-sourced Fitbit datasets 03.12.2016-05.12.2016 [Data set]. *Zenodo*. http://doi.org/10.5281/zenodo.53894

Liang, Z. et al. (2016). A Personalized Approach for Detecting Unusual Sleep from Time Series Sleep-Tracking Data. *The 2016 IEEE International Conference on Healthcare Informatics (ICHI) (2016)*: 18-23.

Meyer, J., Kazakova, A., Büsing, M., and Boll, S. (2016). Visualization of Complex Health Data on Mobile Devices. *Proceedings of the 2016 ACM Workshop on Multimedia for Personal Health and Health Care*, October 16-16, 2016, Amsterdam, The Netherlands. DOI: 10.1145/2985766.2985774

Meyer, J., Wasmann, M., Heuten, W., El Ali, A., and Boll, S. (2017). Identification and Classification of Usage Patterns in Long-Term Activity Tracking. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 667-678. DOI: https://doi-org.proxygw.wrlc.org/10.1145/3025453.3025690

Nazari, G. (2017). Reliability of Zephyr Bioharness and Fitbit Charge Measures of Heart Rate and Activity at Rest, During the Modified Canadian Aerobic Fitness Test and Recovery. *Journal of strength and conditioning research*. 2018, 1. DOI: 10.1519/JSC.0000000000001842

Phatak, S., Freigoun, M., Martín, C., and Rivera, D. (2018), Modeling individual differences: A case study of the application of system identification for personalizing a physical activity intervention, *Journal of biomedical informatics*. 79, 82 – 97.

Purta, R., Mattingly, S., Song, L., Lizardo, O., Hachen, D., Poellabauer, C., and Striegel, A. (2016). Experiences measuring sleep and physical activity patterns across a large college cohort with fitbits, *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, September 12-16, 2016, Heidelberg, Germany. DOI: 10.1145/2971763.2971767

Sathyanarayana ,A., Joty, S., Fernandez-Luque,L., et al. (2016). Sleep Quality Prediction From Wearable Data Using Deep Learning. Eysenbach G, ed. *JMIR mHealth and uHealth*. 2016;4(4):e125. doi:10.2196/mhealth.6562.

Sprint, G. (10/01/2016). Unsupervised detection and analysis of changes in everyday physical activity data. *Journal of biomedical informatics*. 63 p. 54 - 65.

Sprint, G., Cook, D., Weeks, D., Dahmen, J., and La Fleur, A. (2017). Analyzing Sensor-Based Time Series Data to Track Changes in Physical Activity during Inpatient Rehabilitation. *Sensors 2017*, *17*, 2219.

Tang, L.M., Day, M., Engelen, L., Poronnik, P., Bauman, A., and Kay, J. (2016). Daily & Hourly Adherence: Towards Understanding Activity Tracker Accuracy. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 3211-3218. DOI: https://doi-org.proxygw.wrlc.org/10.1145/2851581.2892438

Valbuena, D., Miller, B. G., Samaha, A. L. and Miltenberger, R. G. (2017). Data presentation options to manage variability in physical activity research. *Journal of Applied Behavioral Analysis*, 50: 622-640.

Wright, S., Brown, H., Tyish S., Collier, S., and Sandberg, K. (2017). How consumer physical activity monitors could transform human physiology research. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 2017, 312, 3, R358--R367.

Yürüten, O., Zhang, J., and Pu, P. (2014). Decomposing Activities of Daily Living to Discover Routine Clusters. *AAAI Conference on Artificial Intelligence (2014)*. 2014, 1348-1354.