# Evaluating the Efficacy of Traditional Fitness Tracker Recommendations

Fabio Gasparetti
Roma Tre University, Italy
gaspare@dia.uniroma3.it

Luca Maria Aiello
Nokia Bell Labs, Cambridge, U.K.
luca.aiello@nokia-bell-labs.com

Daniele Quercia
Nokia Bell Labs, Cambridge, U.K.
daniele.quercia@nokia-bell-labs.com

## ABSTRACT

Wearable devices make self-monitoring easier by the users, who usually tend to increase physical activity and weight loss maintenance over time. But in terms of behavior adaptation to these goals, these devices do not provide specific features beyond monitoring the achievement of daily goals, such as a number of steps or miles walked, and caloric outtake. The purpose of this study is to evaluate the efficacy of the recommendations provided by traditional fitness tracker apps with respect to weight loss scenarios.

## CCS CONCEPTS

• **Information systems** → **Recommender systems**; • **Human-centered computing** → *Human computer interaction (HCI).*

## KEYWORDS

Health recommender system; Fitness tracker

**ACM Reference Format:**
Fabio Gasparetti, Luca Maria Aiello, and Daniele Quercia. 2019. Evaluating the Efficacy of Traditional Fitness Tracker Recommendations. In *24th International Conference on Intelligent User Interfaces (IUI '19 Companion), March 17–20, 2019, Marina del Ray, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3308557.3308716

## 1 INTRODUCTION

Undoubtedly, the most popular approaches of physical activity are walking and running. They are associated with numerous health benefits, thus making theme a prime target for physical activity promotion interventions. One of the principal roles of health recommender systems is encouraging users to perform behaviors that more likely bear positive effects.

Wearable devices use proprietary algorithms that, by raw sensor signals and information input by the user, can usually estimate steps, distance, calorie burn, and hours of sleep. They are often considered as cost-effective solutions to support weight loss strategies. By setting and monitoring goals, exercise apps may impact user behaviors with proper UI notifications, leading to increased exercise and, over time, improved health outcomes, such as weight loss conditions. However, to the best of our knowledge, investigations of the traditional recommendations provided by these apps are yet to be conducted at large scale.


*IUI '19 Companion, March 17–20, 2019, Marina del Ray, CA, USA*

ACM ISBN 978-1-4503-6673-1/19/03…$15.00
https://doi.org/10.1145/3308557.3308716

## 2 DATASET DESCRIPTION

Our investigation is based on a large real-world dataset of 11,615 users collected by consumer-grade health-monitoring devices manufactured by Nokia over a 1-year time span, namely from Thu 31 March 2016 to Fri 31 March 2017. As for the demographic statistics, most of the users are in the 20-59 range of years, with an almost equal sex ratio of females and males, (43.6% vs 56.4%). The dataset consists of two categories of devices: wristband activity trackers and digital scales.

A typical signal that is considered strictly correlated to weight, which also shows less sparsity characteristics, is the number of steps $s_{st}$. It is derived by the user's motion sensed by the wristband devices, adjusted for the height of the person. The signal has identical sampling rate throughout the whole set of users, which corresponds to the amount of steps performed during one day.

Based on the bioelectrical impedance analysis, digital scales send a low electrical current through one foot and reading the current with a sensor under the other foot, and monitor the current flowing through the lean mass, which is the most conductive in the human body. By that measure, digital scales can calculate the BMI metric $s_{bmi}$ (also named Quetelet index), the measurement of body fat based on height and weight.

The accuracy of activity and fitness trackers has frequently evaluated in the past [2]. Generally, the studies indicate higher validity for the relative measurement of physical activity measured by the number of steps, with potential risks of steadily undercounting them in specific circumstances. As for BMI, many variables might affect the results, including hydration levels, recent exercise activity and underlying medical conditions. But different studies have shown that bioelectrical impedance analysis is a fairly accurate method for estimating body fat [1].

## 3 EVALUATION OF THE RECOMMENDATIONS

The considered dataset includes a large variety of different human behaviors. The two categories of signals $s_{st}$, and $s_{bmi}$ are subjected of equal-frequency data binning which groups the samples with similar values into one of the five bins $B^{(s_i)} = \{b_1^{(s_i)}, b_2^{(s_i)}, \cdots, b_5^{(s_i)}\}$, where every bin has the same number of samples. The binning is calculated on a per-user basis and on the $z$-score representation of the signals

A fixed-sized sliding window of $\Delta^{(s_{st})}$ (six days in the experiments) slides across the time series, one day at a time. The extracted segments, named *states*, represent user behaviors in the given periods. We also introduce the *action* as the number of steps taken

the day just after one state, which is represented by one of the five bins in $B^{(st)}$.

The ability of the recommender of suggesting behaviors that favourably affect the weight loss is measured in terms of negative alteration of BMI. Less significant gain or loss alterations of the BMI (less than 1%) will be ignored.

The evaluation process is set in a comparative framework by considering four recommendation approaches.

**(R)** A random policy, where an equal probability is assigned to each available action.

**(MP)** The approach recommends the most common action in the dataset given a certain state. Since the dataset consists of users who are supposed to regularly use and monitor their activities to improve their health status, we expect that this strategy reasonably recommends good actions to users.

**(G)** It is similar to the MP strategy, but instead of the most common action in the dataset, the selection is based on the frequency of success, that is, the times that the given state-action pair has led to a negative alteration of weight. It is a typical greedy strategy that tries to maximize the reward without acquiring new knowledge.

**(AT)** The recommender that simulates fitness mobile apps. Many users use mobile apps that collect data from fitness trackers and help them developing exercise routines with achievable, slightly challenging goals based on daily averages. In this scenario, one user shall be deemed to meet the app recommendations if she increases the number of steps per day with respect to the steps taken on average in the previous $N_{AT}$ days. This simulation is implemented in the AT recommender with $N_{AT} = 3$.

The usefulness of one recommendation consists of the effective weight alteration after having considered the suggested actions. We indicate with $W^{(\downarrow,=)}$ and $W^{(\uparrow,=)}$ the number of state-action pairs that match the recommendations, which are characterized by a weight loss and gain trend, respectively. Likewise, the metrics $W^{(\downarrow,\neq)}$ and $W^{(\uparrow,\neq)}$ indicates the number of state-action pairs that do not match the recommendations.

We can cast the evaluation to the traditional set-based measures of precision and accuracy, as follows:

$$Pr = \frac{W^{(\downarrow,=)}}{W^{(\downarrow,=)} + W^{(\downarrow,\neq)}} \quad \text{and}$$

$$Acc = \frac{W^{(\uparrow,=)} + W^{(\downarrow,=)}}{W^{(\uparrow,=)} + W^{(\downarrow,=)} + W^{(\uparrow,\neq)} + W^{(\downarrow,\neq)}}$$

Figure 1 reports the outcomes of the experimental evaluation. A significant observation is related to the recommendations provided by traditional apps associated with activity trackers (AT), which follow a locally optimal choice, similar to a greedy strategy, therefore they often obtain suboptimal outcomes in terms of weight loss. Whenever the user decides to adhere to that strategy, trying to steadily increment the count of steps every day, the chances of success decrease. Similar outcomes are obtained by the G strategy, which also follows a greedy approach.

A sort of wisdom of the crowd is manifested when the recommendations are based on the most common action given the current state (MP). Regardless the past behavior of users, the most-popular strategy suggests valid routines toward weight loss one-third of the times, with higher accuracy with respect to purely greedy strategies.

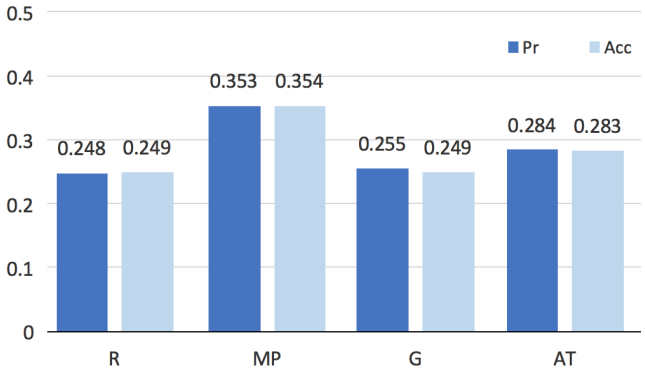


**Figure 1: Precision (dark) and Accuracy (light blue) measures.**

These outcomes indicate how traditional recommendation approaches based on steadily increments of physical fitness level might not be optimal for the weight loss. Additional investigations on adaptive and personalized strategies are required.

## REFERENCES

[1] S Demura and S Sato. 2015. Comparisons of accuracy of estimating percent body fat by four bioelectrical impedance devices with different frequency and induction system of electrical current. *J. Sports Medicine and Physical Fitness* 55, 1-2 (2015), 68—75.

[2] K. R. Evenson, M. M. Goto, and R. D. Furberg. 2015. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J of Behavioral Nutrition and Physical Activity* 12, 1 (18 Dec 2015), 159.