An Analysis of Exercising Behavior in Online Populations

David Jurgens, James McCorriston, Derek Ruths

Introduction and Motivation

Exercise is an important component of health. Previous studies of exercise and health have been stymied by the lack of accurate records of exercising activities. In our study, we overcome this limitation by using data from Fitocracy, a gamified workout tracking site that precisely records users’ workout histories using over 1,000 unique exercises and where users self-report their age, gender, and height. We analyze nearly half a million users’ histories to identify exercising behavior trends and how those behaviors vary by age and gender.

How does exercising behavior vary with age and gender?

Methodology: The activities performed by different age groups provide a insightful view into how exercise behavior changes over time. We divided users by gender and into seven age ranges. Within each gender-age cohort, we computed the probability that a user in that cohort records each exercise and then sorted all exercises according to their average probability of being performed. Below, we show the ten exercises that are most likely to be performed by an individual from each cohort.

What exercising behaviors are practiced by subpopulations?

Methodology: To identify the underlying behaviors from people’s activities, we train a Latent Dirichlet Allocation (LDA) model on users’ exercising data. We model each individual’s history as a reflection of that person engaging in just a few behaviors, where a behavior selects for certain exercises with a certain probability. We then sorted all exercises according to their average probability of being performed. Below, we show the ten exercises that are most likely to be performed by an individual from each cohort.

Data

All of a user’s activities, profile, social data, and group memberships on Fitocracy were crawled over a six-month period to acquire the complete profiles and workout histories of 441,034 users. Ultimately, 188,265 users recorded at least one workout, with the total dataset comprising 3,109,276 workouts (13.4M activities) over nearly a year span from February 2011 to January 2015.

Check out our interactive behavior demo at http://networkdynamics.org/resources/exercise