Large-Scale Mobile Fitness App Usage Analysis for Smart Health

Xinlei Chen, Zheqi Zhu, Min Chen, and Yong Li

The authors investigate mobile fitness app usage data from more than 14,000 cellular towers and 4000 users. Based on solid analysis, they find that temporal factors, such as the day of a week and the time of day, affect human workout activities, while location is another key factor. In addition, mobility factors, such as moving range and regularity of mobility, also influence human workout activities.

Abstract

Understanding essential factors that influence urban citizens' workout activities, both aerobic and training, is important for public policy development and city planning. Mobile fitness app usage data, which describes who works out at what time and place, provides a good opportunity to figure out these key factors that influence human workout activities. This article investigates the mobile fitness app usage data from more than 14,000 cellular towers and 4000 users. Based on solid analysis, we find that temporal factors, such as the day of the week and the time of day, affect human workout activities, while location is another key factor. In addition, mobility factors, such as moving range and regularity of mobility, also influence human workout activities. Finally, personal income is another fundamental factor that we find influences human workout activities.

Introduction

Human workout activities have been demonstrated to have very important roles in our daily life. Frequent and regular workout activities prevent diseases such as cognitive decline, help to maintain healthy weight, and relieve symptoms of depression and anxiety [1, 2]. Understanding key factors that influence urban citizens' workout activities is important for developing public policy and planning cities to help citizens stay healthy [3–5]. However, it is very difficult to have deep insight on this problem with traditional sensing methods and limited data.

The last decade has witnessed a large amount of data from cellular networks being collected. The Cisco white paper claims that monthly global mobile data traffic reached 4.4 exabytes (10^{18}) at the end of 2015 and increased to 7.2 exabytes by the end of 2016 [6]. Mobile fitness app usage data, which describes who works out at what time and what place, provides a good opportunity to figure out the essential factors that influence human workout activities.

This article investigates the mobile fitness app usage records from more than 14,000 cellular towers and 4000 users. We propose a system architecture to remove the redundant and conflicting logs as well as redundant information across different days for further analysis. Then we study the temporal and spatial factors that affect human workout activities. Finally, we identify the influences of the mobility and economic factors. The contributions of this article are the findings summarized below:

- The number of people who take up strength training activities is much lower than the number of people who take up aerobic training since aerobic training has lower requirements on venue, equipment, cost, time, and so on.
- The number of people who work out in a city depends on both the day of the week and the time of day. The temporal pattern also differs between aerobic training and strength training.
- Location affects the number of people who work out in a city. Central areas have more people working out than rural areas.
- People who move over larger spans tend to work out less in both aerobic and strength training. The major reason is due to the lack of time for workout activities.
- The regularity of human mobility affects workout time. People who have regular liv work out more, since they care more about their health condition.
- Personal income is a key factor that affects human workout activities. Richer people tend to work out more since they are more likely to afford time and expense on working out.

We structure our article as follows. We first introduce the importance of human workout activity analysis and the possibility to adopt mobile fitness app usage data to analyze it. After the problem statement, we introduce the dataset used in this article and propose the system architecture. Then we investigate the temporal and spatial influence on human workout activities. Next, mobility and economic effects are also analyzed. Finally, we conclude this article and summarize our main findings.

Preliminaries

This section introduces the preliminaries to analyze temporal and spatial patterns of mobile fitness apps, as well as their relationships with human mobility pattern and economic condition. We first introduce how we collect the data and the details of our dataset. Then we propose a system architecture for further data analysis.

Digital Object Identifier: 10.1109/MCOM.2018.1700807

Xinlei Chen is with Carnegie Mellon University; Zheqi Zhu and Yong Li are with Tsinghua University; Min Chen is with Huazhong University of Science and Technology.

Dataset Description

To extract the key factors that affect human workout activities, some information is needed: who uses what fitness app at what time and where. Therefore, we adopt two datasets, fitness app usage dateset and cellular tower location dataset, which are collected by one major mobile operator in Shanghai, one of the largest cities in China. Both of them were collected with deep packet inspection (DPI) appliances from April 19, 2016 to April 26, 2016.

The fitness app usage dateset includes anonymized user ID, start time of fitness app data access, end time of fitness app data access, connected cellular tower ID, app ID, amount of fitness app traffic uploaded to cellular towers during the connection time, and amount of fitness app traffic downloaded from cellular towers during the connection time. More than 2.6 million logs were collected with more than 4000 users.

The fitness app usage dataset includes seven fitness apps, which can be classified into two categories: one is used for step counting, while the other is used for workout training. The step counting apps calculate how many steps a user takes during their cardio exercise [7]. The usage data from step counting apps can be used to analyze human aerobic training activities. Workout apps are usually multi-functional. These apps download and play tutoring videos, make training plans, and record users' workout time, content, and feelings. Since these workout apps aim at strength training users, the usage data from workout apps can be used to analyze human strength exercise activities [8]. It is noticed that not all users access all seven fitness apps due to their similar functions.

The cellular tower location dataset includes cellular tower ID, cellular tower longitude, and cellular tower latitude. More than 14,000 cellular towers' information is collected, with longitude range of 120.903°E–121.967°E and latitude range of 30.706°N–31.543°N. These cellular towers cover all urban areas and most rural areas in the city of Shanghai. All the connected cellular tower IDs in the fitness app usage dateset can be found in the cellular tower location dataset. This indicates that we have coarse location information for all anonymized users.

System Architecture

In order to analyze human workout activities, we design our system with three major parts: *raw data collection*, *pre-processing*, and *pattern extraction*, as shown in Fig. 1. The raw data collection module collects the mobile fitness app usage information and cellular tower location information. The raw information from the two datasets is first cleaned in the pre-processing module. Then the pattern extraction module extracts temporal and spatial patterns of fitness app usage, as well as their relationships with human mobility and human economic condition.

The pre-processing module helps the system to get "clean" data from raw collected data and match the fitness app usage information with cellular tower location information for further pattern extraction in the pattern extraction module. Due to technical issues, the raw collected data has redundant and conflicting logs. For example, there are several logs with the same user ID, the

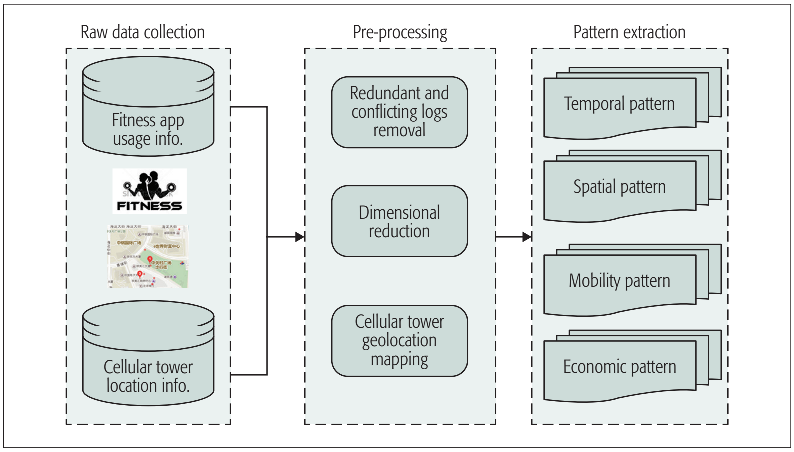

Figure 1. This figure shows the architecture of our system, which consists of three major parts: raw data collection, pre-processing, and pattern extraction.

same timestamp, the same mobile fitness app ID, but different cellular tower IDs. These redundant and conflicting logs can be removed by checking the start and end time of fitness app data access of each user for each fitness app. Then the *dimensional reduction* sub-module calculates the correlations across different days. According to the correlation coefficients, the system decides whether the daily data includes only similar information or totally new information. Finally, the *cellular tower geolocation mapping* sub-module obtains location information (latitude and logitude) for each log in the fitness app usage dataset based on cellular tower ID matching.

The pattern extraction module extracts the temporal and spatial patterns of mobile fitness apps, as well as their relationships with human mobility and personal economic condition. Specifically, this module tries to answer several questions:

- What is the temporal distribution of fitness app usage in a week and in a day? More specifically, when do more people work out?
- What is the spatial distribution of fitness app usage in a week? More specifically, in what areas do more people work out?
- What is the relationship between human workout time length and human mobility span? More specifically, do people who move over larger ranges each day tend to work out more or less?
- What it the relationship between human workout length and human mobility regularity? More specifically, do people who have similar daily mobility patterns tend to work out more or less?
- Does economic condition affect the human workout length? More specifically, do people who have high income tend to work out more or less?

The detailed analysis for the first two and the other three questions can be found later.

Temporal and Spatial Pattern

This section investigates the temporal and spatial patterns of fitness app usage, which benefits city public health policy and urban planning [9]. We first introduce the characteristics on temporal domain, including weekly, weekday daily,

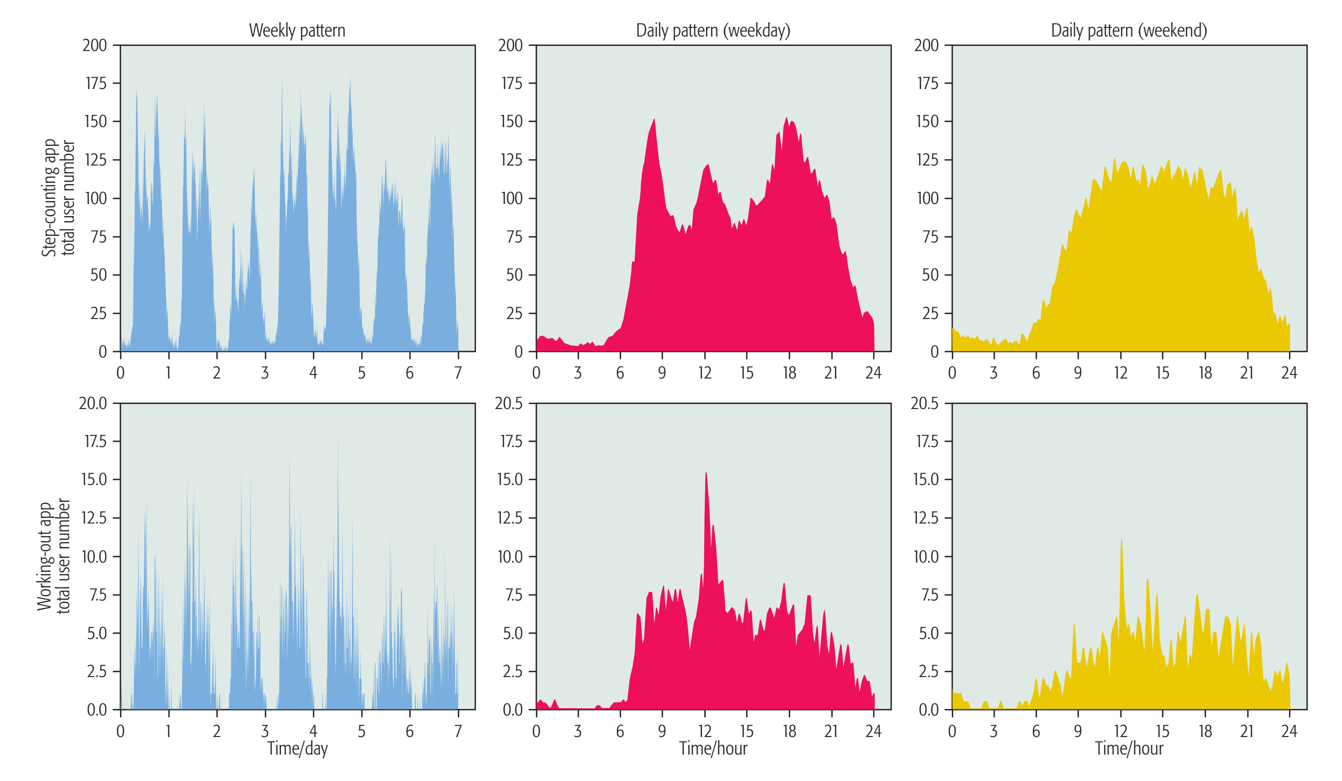

Figure 2. The temporal fitness app usage patterns over time, which include weekly characteristics in blue, weekday daily characteristics in red, and weekend daily characteristics in gold.

and weekend daily patterns. Then spatial patterns on both weekdays and weekends are discussed. Finally, we study what subjects affect the number of people working out.

Temporal Pattern Extraction

In order to study how temporal factors affect human workout activities, we plot the total user numbers of fitness app usage at different timescales in Fig. 2. The weekly, weekday daily, and weekend daily patterns are illustrated in the left, middle, and right columns, while one representative step counting app and one representative workout app are plotted in the top and bottom rows. The daily total user numbers on weekdays and weekends are obtained by averaging data from Monday to Friday and Saturday to Sunday, respectively.

The total user numbers of fitness app usage illustrate similar patterns on different days of a week, which shows in both the step counting app and workout app. We further calculate the correlation coefficients between different days, which are between 0.85 and 0.98. This shows that urban human fitness activities are periodic on a daily scale. In addition, total user numbers of the workout app are much smaller (around one tenth) than total user numbers of the step counting app. This is because working out, especially strength training, has high requirements on venue, equipment, cost, time, and so on. In contrast, cardio exercises such as walking and jogging are much easier for most people to take. The large difference in user number between the step count and workout apps also illustrates a huge popularity difference between these two kinds of workout activities.

The step counting app shows different daily patterns than the workout app on weekdays. The step counting app contains three peaks at 8:00 a.m., 12:00 p.m., and 6:00 p.m., which relate to going to work, going to lunch, and going back home. These three periods relate to the busy time when people walk most frequently. The workout app has only one peak, from 11:00 a.m. to 1:30 p.m., which is related to the lunch break. This illustrates that people in Shanghai tend to work out during the lunch break. After several hours' work in the morning, working out helps people refresh and have higher efficiency in the afternoon.

During weekends, the total user number of the step counting app first increases from 6:00 a.m. to 12:00 p.m. and then maintains a similar level until 9:00 p.m. After 9:00 p.m., the user number decreases. This illustrates that most human activities happen during 12:00 p.m. to 9:00 p.m. during weekends. The total user number of the workout app shows three peaks around 12:00 p.m., 2:00 p.m., and 18:00 p.m., which correspond to lunch, afternoon nap, and dinner time. This illustrates that people tend to work out during these three time periods. In addition, the step counting app has similar numbers of total users on weekdays and weekends, while the workout app has a lower number of total users on the weekend. This means that people prefer working out during weekdays.

Spatial Pattern Extraction

In order to investigate how people in different areas work out, we plot spatial distributions of total user number of step counting and workout apps in Fig. 3. The weekday and weekend spatial patterns are illustrated in the left and right columns, while one representative step counting app and one representative workout app are plotted in the top and bottom rows. The total user num-

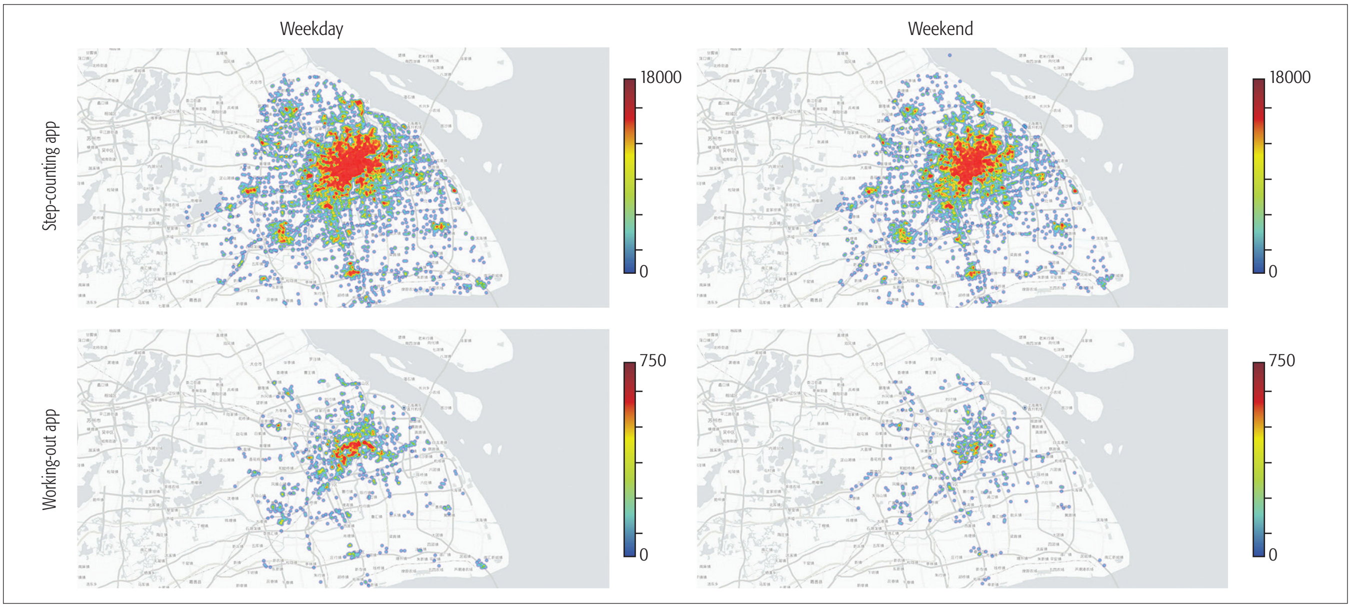

Figure 3. The spatial distribution of fitness app user number on one representative weekday and weekend.

bers on weekdays and weekends are obtained by averaging data from Monday to Friday and Saturday to Sunday, respectively.

For the step counting app, spatial distributions of total user number on weekdays and weekends show a similar pattern, where high numbers appear in the central business district (CBD) and low numbers appear in the rural area. This shows that location is a key factor that affects human mobility. People in the CBD walk or jog more than people in rural areas since they have much busier lives. The hot area on a weekday is slightly larger than that on the weekend, which shows that people walk more on weekdays when they are busier.

A similar pattern can be observed in the workout app total user numbers. This shows that location affects not only human mobility but also human workout activities. People in the CBD work out more than people in rural areas since they need to work out more to relieve higher life pressure. The hot area for the workout app is much smaller than that for step counting app. This is due to two reasons. One is because more people use the step counting app. The other is that people usually stay at the same locations, such as gym, home, or office, when working out.

To conclude this section, both time and location are key influential factors on human workout activities. Not only different days in a week but also different times of day affect the number of people working out. In addition, people in central areas of the city tend to work out more than those in rural areas.

Mobility and Economic Effect

This section investigates how mobility and economic effects influence human workout activities. For mobility effects, we consider both moving range and regularity of mobility. For economic effect, we consider the relationship between fitness app usage and user personal income [3]. We also summarize the major findings from analysis in the last part of the section.

Moving Range

In order to figure out how moving range affects human workout length, we investigate the relationship between radius of gyration and the fitness app using time length. Since the mobile app packets are uploaded and downloaded in discontinuous time slots, we adopt density-based spatial clustering of applications with noise (DBSCAN) to calculate app connection time length [10]. DBSCAN is usually used to group points that are closely packed together, which helps connect the discontinuous time slots in our mobile fitness app usage data to form a complete time period [11]. Radius of gyration represents the distribution of an object around an axis [12]. High radius of gyration means high moving range.

To calculate radius of gyration R_g, we first need to have all the locations (x_1, y_1), ..., (x_i, y_i), ..., (x_N, y_N) in a user's trajectory, where N is the total number of locations in the trajectory. Then the weighted center of gravity of these locations can be derived by

$$(x_c, y_c) = \frac{\sum_{i=1}^{N} (x_i, y_i) * w_i}{\sum_{i=1}^{N} w_i},$$

where w_i is the frequency of the location (x_i, y_i) passed by. After that, the distance between each location and the weighted center of gravity is calculated by

$$r_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}.$$

With all these calculated, radius of gyration R_g can be obtained by

$$R_g = \frac{\sqrt{\sum_i r_i^2 p(r_i)}}{2 * \sum_i p(r_i)},$$

where $p(r_i)$ is the probability of the location (x_i, y_i) passed by.

Both time and location are key factors influencing human working out activities. Not only different days in the week but also different times in the day affect the number of people working out. In addition, people in central areas of the city tend to work out more than those in rural areas.

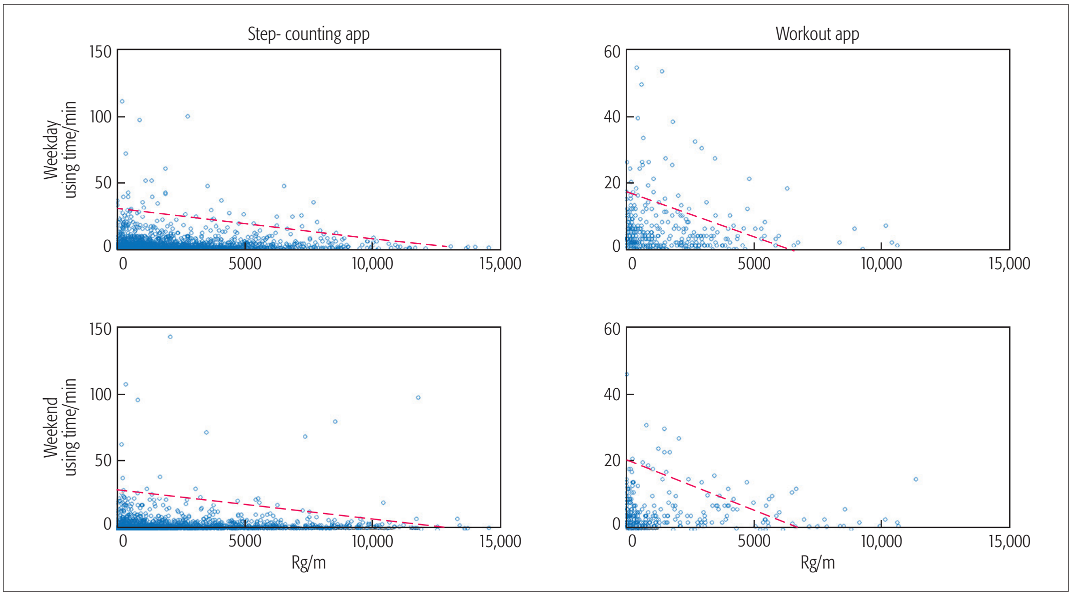

Figure 4. The movement span effect on daily workout length on weekdays and weekends for the step counting app and workout app. Each blue circle represents one user. Large R_g represents large human moving range.

As shown in Fig. 4, for the workout app, the use time decreases when R_g increases on both weekdays and weekends. This shows that people who move more tend to work out less. Large-range moving usually is related to non-workout activity. People with large moving range usually do not have too much time to work out, since most of their time is occupied by other issues. In addition, we observe a similar trend for the step counting app on both weekdays and weekends. This illustrates that moving range is a key factor that affects human workout length.

Regularity of Mobility

This subsection investigates the relationship between human workout activities and their mobility regularity. As mentioned in the previous subsection, app use length is adopted to describe workout length. The actual entropy is adopted to describe mobility regularity, which depends not only on the visited location frequency, but also on the order of and time spent at these visited locations [13]. Therefore, it captures both temporal and spatial regularity of human movement. Small entropy value usually means regular mobility pattern.

If $T = \{C_1, ..., C_i, ..., C_N\}$ is the sequence of cellular towers that a user visits at each consecutive time interval, the entropy E can be derived by

$$E = \sum_{t \subset T} p(t)\log_2[p(t)],$$

where $p(t)$ is the probability that a particular time-order sub-sequence t appears in the trajectory.

As shown in Fig. 5, for both the step counting and workout apps, the use time decreases when E increases. This shows that those who have regular mobility patterns in a week do cardio or strength training more often. People who maintain regular lives tend to care more about their health. As a result, they do not only aerobic training, but also strength exercises. The decreasing trend in the workout app is sharper, which shows that mobility regularity affects strength exercises more than aerobic training. This is because strength exercises have higher requirement on workout field, time, and expense, which people with regular lives are more likely to afford.

Economic Factors

To understand how economic factors affect human workout activities, we study the relationship between local housing prices and total number of fitness app users. High housing price usually means that people have better economic conditions. This is because the user can afford to buy or rent a better apartment or work at a more expensive office, both of which indicate high income [14].

Figure 6 shows the relationship between housing price and total user number of one representative step counting app and one representative workout app. The grey dots are the real collected data, while the blue lines are the fitted curve. The x axis is the housing price in RMB/m^2, which indicates personal income. One RMB Yuan equals around US$0.15 [15]. The y axis is the total user number expressed in log10 format.

For both the step count and workout apps, higher total user numbers are observed where housing prices are high. This illustrates that richer people tend to work out more. There are two reasons to explain the phenomenon:

1. Compared to people with low income, richer people care more about their body condition, since poor people have to put more focus on working for a living.
2. Working out usually requires both time and money. People with high income have more spare time to work out and can also afford high gym prices. Moreover, the increasing total user number trending for the workout app is much higher than that for the step counting app. This further proves that income influences human workout activities.

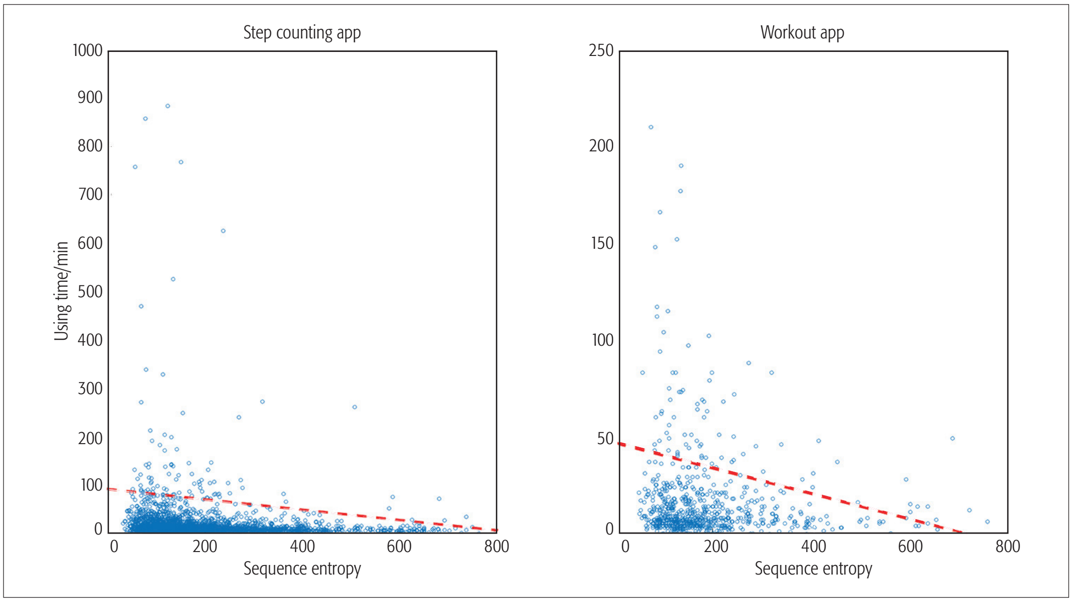

Figure 5. The mobility regularity effect on user workout length in a week for the step counting app and the workout app, where each blue circle represents one user.

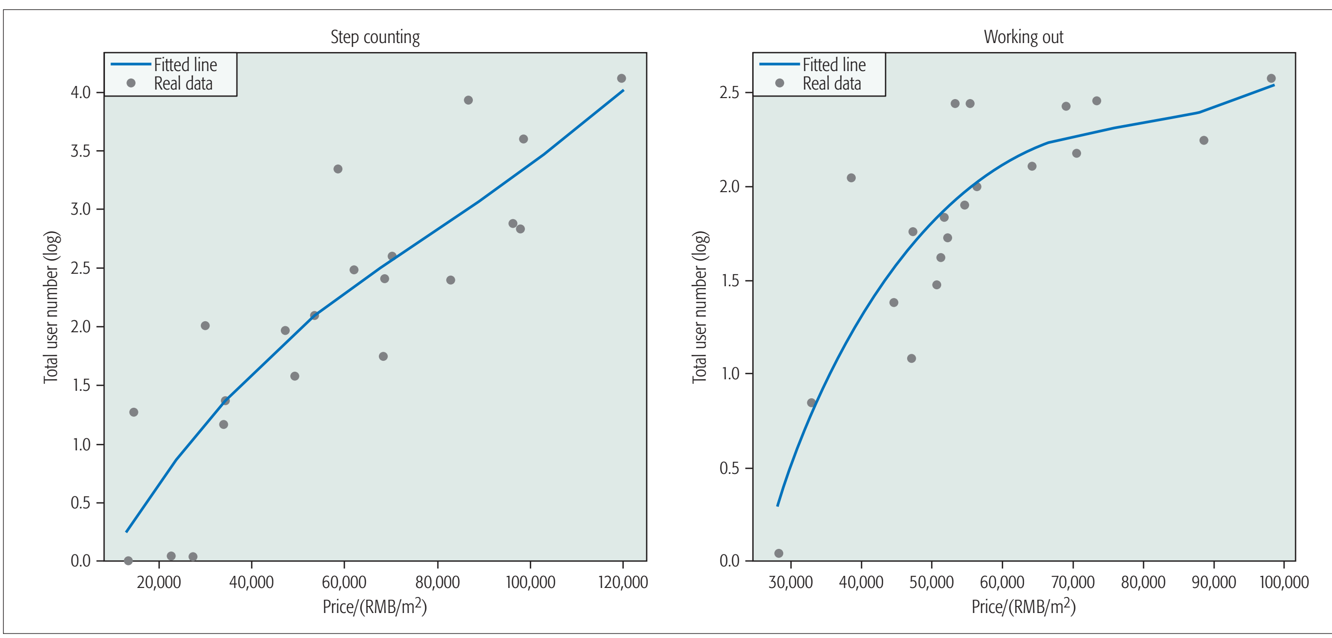

Figure 6. The influence of the economic factor on app user number.

Major Findings from Analysis

This subsection summarizes the major findings from temporal, spatial, human mobility, and economic patterns. Mobile fitness app usage data, which describes who works out at what time at what location, provides a good opportunity to figure out the key factors that influence human workout activities. Our main findings are listed as below:

- The number of people who take up strength training activities is much less than the number of people who take up aerobic training since aerobic training has lower requirements on venue, equipment, cost, time, and so on.
- The number of people who work out in a city depends on both the day of the week and the time of day. The temporal pattern also differs between aerobic training and strength exercises.
- Location affects the number of people who work out in a city. Central areas have more people working out than rural areas.
- People who move over larger spans tend to work out less, for both aerobic training and strength exercises, due to lack of time for workout activities.
- Human mobility regularity affects workout frequency. People who have regular lives work out more, since they care more about their health condition.
- Personal income is a key factor that affects human workout activities. Richer people tend to work out more since they are more able to afford the time and expense of working out.

In future work, other factors will be further investigated. First, it is interesting to discover how different city functional regions affect human working out activities. Intuitively, people tend to work out in and near parks, or at or near home. This intuition needs to be checked by real data. Second, will females and males have different working out patterns?

Conclusion

This article investigates the mobile fitness app usage data from more than 14,000 cellular towers and 4000 users for human workout activity analysis. Based on the solid analysis, we find that temporal factors, such as the day of the week and the time of day, affect human workout activities. Location is another key factor. In addition, mobility factors, such as moving range and mobility regularity, also influence human workout activities. Finally, personal income is the last factor that we find. These findings are important for developing public policy and planning cities

Future Work

Due to the format limitation, this article only contains some selected essential factors that affect human working activities. In future work, other factors will be further investigated. First, it is interesting to discover how different city functional regions affect human workout activities. Intuitively, people tend to work out in and near parks, or at or near home. This intuition needs to be checked by real data. Second, will females and males have different workout patterns? Since females and males differ both physically and mentally, they probably have different workout habits. Finally, other factors such as age and educational level will also probably matter. Older people tend to work out less than young people due to health constraints. People with different educational levels may also have different attitudes towards working out. These intuitions will be checked in our future work with real data.

References

[1] WHO, Global Recommendations on Physical Activity for Health; http://www.who.int/dietphysicalactivity/publications/9789241599979/en/ (WHO, 2010), accessed Aug. 20, 2017.

[2] J. F. Sallis *et al.*, “Progress in Physical Activity over the Olympic Quadrennium,” *The Lancet*, vol. 388, no. 10,051, 2016, pp. 1325–36.

[3] D. A. Chokshi and T. A. Farley, ”Changing Behaviors to Prevent Noncommunicable Diseases,” *Science*, vol. 345, no. 6202, 2014, pp. 1243–44.

[4] R. S. Reis *et al.*, “Lancet Physical Activity Series 2 Executive Committee and Others, Scaling Up Physical Activity Interventions Worldwide: Stepping Up to Larger and Smarter Approaches to Get People Moving,” *The Lancet*, vol. 388, no. 10,051, 2016, pp. 1337–48.

[5] R. E. Rikli and C. J. Jones, ”Development and Validation of a Functional Fitness Test for Community-Residing Older Adults,” *J. Aging and Physical Activity*, vol. 7, no. 2, 1999, pp. 129–61.

[6] Cisco, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper; http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html, accessed June 20, 2017.

[7] Joy Run App; http://www.thejoyrun.com, accessed Aug. 20, 2017.

[8] Keep App; http://www.gotokeep.com/, accessed Aug. 20, 2017.

[9] T. Althoff *et al.*, “Large-Scale Physical Activity Data Reveal Worldwide Activity Inequality,” *Nature*, vol. 547, no. 7663, nature23018, 2017.

[10] M. Ester *et al.*, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” *Kdd*, vol. 96, no. 34, 1996, pp. 226–31.

[11] D. Birant and A. Kut, “ST-DBSCAN: An Algorithm for Clustering Spatial-Temporal Data,” *Data & Knowledge Engineering*, vol. 60, no. 1, 2007, pp. 208–21.

[12] A. Y. Grosberg *et al.*, ”Statistical Physics of Macromolecules,” *Computers in Physics*, vol. 9, no. 2, 1995, pp. 171–72.

[13] C. Song *et al.*, “Limits of Predictability in Human Mobility,” *Science*, vol. 327, no. 5968, 2010, pp. 1018–21.

[14] Anjuke, Chinese Real Estate Real-Time Selling and Renting Information Publishing; https://shanghai.anjuke.com/, accessed Aug. 20, 2017.

[15] XE Currency Converter, The World’s Trusted Currency Authority; http://www.xe.com/, accessed June 20, 2017.

Biographies

Xinlei Chen received his B.E. and M.S. degrees in electrical engineering from Tsinghua University, China, in 2009 and 2012 respectively. He is currently working toward a PhD degree at the Department of Electronic and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania. His research interests are in the areas of big data, cyber physical systems (collaborative drone swarms), mobile crowdsensing, networking, and communications (millimeter wave communications).

Zheqi Zhu is a senior undergraduate student at Tsinghua University, majoring in electronic engineering. The fields he is interested in include big data and machine learning, and he has finished several projects in these areas.

Min Chen has been a full professor in the School of Computer Science and Technology at Huazhong University of Science and Technology (HUST) since February 2012. He is the director of the Embedded and Pervasive Computing (EPIC) Lab at HUST. He is Chair of the IEEE Computer Society (CS) Special Technical Community (STC) on Big Data. He serves as an Editor or Associate Editor for *Information Sciences*, *Information Fusion*, and *IEEE Access*, among other publications.

Yong Li (liyong07@tsinghua.edu.cn) is currently an associate professor of electronic engineering at Tsinghua University. His research interests are in the areas of networking and communications, including mobile opportunistic networks, device-to-device communication, software-defined networks, network virtualization, and future Internet. He has published over 100 research papers, and has 10 Chinese and International patents. He has served as TPC Chair for the WWW Workshop of Simplex 2013 and a TPC member of several international workshops and conferences.