CrossMark

Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results

Kangjae Lee[a,*], Mei-Po Kwan[b]

[a] Illinois Informatics Institute, University of Illinois at Urbana-Champaign, National Center for Supercomputing Applications, MC-257, 1205 W Clark St., Urbana, IL 61801, USA
[b] Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, 255 Computing Applications Building, MC-150, 605 E Springfield Ave., Champaign, IL 61820, USA

ARTICLE INFO

ABSTRACT

In recent decades, decreasing physical activity has emerged as one of the major issues affecting human health since people increasingly engaged in sedentary behavior in their homes and workplaces. In physical activity research, using GPS trajectories and advanced GIS methods has a potential for greatly enhancing our understanding of the association between objectively measured moderate and vigorous physical activity and physical and social environments. Relying only on objectively measured physical activity intensity, however, ignores the role of different places and types of physical activity on people's health outcomes. The aim of this study is to propose an approach to classifying physical activity in free-living conditions for physical activity research using published smartphone accelerometer data. Random forest and gradient boosting are used to predict jogging, walking, sitting, and standing. Generated training models based on the two classifiers are tested on accelerometer data collected from the smartphones of two subjects in free-living conditions. GPS trajectories with predicted physical activity labels are visually explored on a map to offer new insight on the assessment of the predicted results of daily activities and the identification of any difference in the results between random forest and gradient boosting. The findings of this study indicate that random forest and gradient boosting enable accurate physical activity classification in free-living conditions. GPS trajectories linked with predicted labels on a map assist the visual exploration of the erroneous prediction in daily activities including in-vehicle activities.

1. Introduction

In recent decades, decreasing physical activity (PA) has emerged as one of the major issues affecting human health since people increasingly engaged in sedentary behavior in their homes and workplaces. Moreover, obese people spend less time on moderate to vigorous physical activity (MVPA) when compared to non-obese people (Hagströmer, Troiano, Sjöström, & Berrigan, 2010). MVPA, such as brisk walking, bicycling, and jogging, contributes to reduced risk of physical and mental health problems (e.g., cardiovascular diseases, type II diabetes, obesity, depression, anxiety, and well-being) (Fox, 1999; Gordon-Larsen, Nelson, Page, & Popkin, 2006; Physical Activity Guidelines Advisory Committee, 2008; Wei, Gibbons, Kampert, Nichaman, & Blair, 2000). Metabolic syndrome is associated with a number of psychiatric disorders (Ho, Zhang, Mak, & Ho, 2014), and depression is a common comorbidity (Quek, Tam, Zhang, & Ho, 2017). Depression and obesity share a common pathological mechanism (Yang, De Xiang Liu, Pan, Ho, & Ho, 2016). PA and exercise lead to

significant reduction of stress levels as compared to short-term pharmacological treatment (Lu et al., 2017). A number of scholars have attempted to identify factors in the physical and social environments that have significant positive or negative effects on people's PA. Many studies have examined the relationship between PA and specific characteristics of the built environment, including green spaces, based on neighborhood areas using geographic information systems (GIS) (Cohen et al., 2006; Coombes, Jones, & Hillsdon, 2010; McGinn, Evenson, Herring, Huston, & Rodriguez, 2007; Nagel, Carlson, Bosworth, & Michael, 2008; Saelens & Handy, 2008; Sallis et al., 2016).

In PA research, using GPS trajectories and advanced GIS methods has a potential for greatly enhancing our understanding of the association between objectively measured MVPA and physical and social environments (Browning & Lee, 2017). Through taking into account people's daily activities and travel, insights may be obtained to better inform policies or measures that seek to promote PA (Almanza, Jerrett, Dunton, Seto, & Pentz, 2012; Boruff, Nathan, & Nijënstein, 2012; Cooper et al., 2010; Helbich et al., 2016; Lachowycz, Jones, Page,

Wheeler, & Cooper, 2012; Rodríguez et al., 2012; Troped, Wilson, Matthews, Cromley, & Melly, 2010). More important, recent studies using GPS trajectories confirm the importance of non-residential contexts (e.g., workplaces or locations for routine activities) in people's daily life as well as areas around their residential neighborhoods (Diez Roux & Mair, 2010; Inagami, Cohen, & Finch, 2007; Kwan, 2012a, 2012b; Perchoux, Chaix, Cummins, & Kestens, 2013). In most existing research, MVPA and sedentary behavior are determined using some thresholds based on the count per minute, an automatically calculated measure from commercial accelerometers (Berlin, Storti, & Brach, 2006; Frank, Schmid, Sallis, Chapman, & Saelens, 2005; Freedson, Melanson, & Sirard, 1998; Jones, Coombes, Griffin, & van Sluijs, 2009; Saelens, Sallis, Black, & Chen, 2003).

Relying only on the objectively measured intensity of PA, however, ignores the role of different places and types of PA on people's health outcomes. As a result, our knowledge about what types of PA were undertaken and what contextual characteristics are associated with healthy behaviors is limited when using intensity. In this context, Jankowska, Schipperijn, and Kerr (2015) highlighted the importance of understanding individual health behaviors over space and time related to PA and the need to go beyond using intensity. Thus, the specific types rather than the intensity of PA are critical to a better understanding of the association between PA and certain environmental influences, and accurate classification of PA types needs to be studied.

Many studies in the healthcare domain have been conducted to recognize different types of daily activities and PA using raw accelerometer data collected from subjects under controlled conditions (Anguita, Ghio, Oneto, Parra, & Reyes-Ortiz, 2012; Arif, Bilal, Kattan, & Ahamed, 2014; Kwapisz, Weiss, & Moore, 2011; Yin, Yang, & Pan, 2008; Zhang, McCullagh, Nugent, & Zheng, 2010; Zhu & Sheng, 2011). Mobile phone applications provide a low-cost technology, which allows clinicians to monitor PA of their patients without any technical knowledge (Zhang et al., 2014). The identification of different activity types enables the detection of abnormal behaviors when monitoring elderly people for their healthcare, the examination of the association between PA and its health effects on people, and the provision of feedback through mobile applications to encourage people to engage in PA. Further smartphone innovations are also helpful to care givers who are caring for individuals with dementia to improve not only their PA (Zhang et al., 2016) but also rehabilitation (Zhang, Yeo, & Ho, 2015). In the existing studies, machine learning techniques played an important role in building models based on a set of features derived from raw accelerometer data to predict various PA types. Machine learning is a branch of artificial intelligence and helps to predict outcomes after models/algorithms are trained using a large amount of input data. Features in machine learning are informative quantifiable attributes, derived from the input data, such as mean and standard deviation, used to determine different labels (or classes) within an acceptable range in models. The classification algorithms in existing studies, however, are mostly tested on accelerometer data collected under controlled situations. In other words, how daily activities and PA, which might include a variety of uncontrolled activities, can be represented by restricted types in the training accelerometer data needs to be understood. Therefore, a validation process needs to be performed to unveil how the accelerometer data collected in a laboratory setting provide convincing predicted results in various daily activities.

The aim of this study is to propose an approach to PA classification in free-living conditions for PA research using published smartphone accelerometer data. Free-living conditions refer to the natural everyday settings in people's daily lives in contrast to artificial laboratory conditions. Two supervised machine learning classifiers – random forest and gradient boosting – are used to generate training models based on publicly released accelerometer data for predicting different PA types and comparing their performance. The PA types identified by the proposed classification algorithm are jogging, walking, sedentary status,

and standing. The performance of the generated predictive models is assessed in two different ways – 1) with an approach of cross-validation and 2) using test accelerometer data collected from a smartphone of one adult subject in free-living conditions. Because the published accelerometer data used to train models were collected under controlled conditions, the assessment of the learning models using data recorded in uncontrolled daily life is critical for this study. For more thorough examination of the models, a visual exploration of classified PA types over space and time is performed on a map using a set of GPS and accelerometer data collected from the smartphones of two subjects.

The process of building models developed in this study contributes to improving the performance of PA classification by highlighting practical strategies and considerations for collecting GPS and accelerometer data from human subjects. Further, it will be helpful to future studies that seek to advance the examination of the association between PA and environmental factors. The construction of learning models using publicly available accelerometer data with labels of PA types also will enable researchers to address the daunting challenge of requiring subjects in PA research to record the labels for every activity. Visualization of the classified types on GPS trajectories offers new insight into the assessment of the predicted results of daily activities and the identification of any difference in the results between the random forest and gradient boosting classifiers.

The sections in this paper are structured as follows: Section 2 summarizes existing studies on PA classification algorithms using accelerometer data. Section 3 describes the accelerometer data used in this study, preprocessing of the accelerometer data, and a classification algorithm using random forest and gradient classifiers taking into account extracted features from accelerometer data instances (samples). Section 4 demonstrates the performance of random forest and gradient boosting and the application of the developed classification algorithm to the accelerometer data collected from two subjects in free-living conditions. Lastly, discussion and conclusions of the findings in this study are presented in Section 5.

2. Related work

Due to the widespread use of smartphones, researchers are paying increasing attention to the sensing capabilities of smartphones. As one of the sensors in smartphones, tri-axial accelerometer records the accelerations of x, y, and z-axes, which may allow the recognition of different types of human activities. Many scholars in the public health and computer science domains have highlighted the potential of the built-in accelerometer sensor in smartphones to recognize different types of PA using machine learning techniques (Anguita et al., 2012; Arif et al., 2014; Kwapisz et al., 2011; Zhang et al., 2010). For instance, Zhang et al. (2010) used support vector machine (SVM) to recognize six PA types including walking, standing, and sitting. Three-axis accelerometer data were recorded by 10 participants using smartphones worn on their waists on the left side. Kwapisz et al. (2011) developed a system to recognize six PA types by analyzing collected accelerometer data through smartphones of participants. The recognition process was performed using 43 generated features extracted from raw accelerometer data and tested using three different machine learning models. Among the three machine learning classifiers, the multilayer perceptron showed 91.7% predictive accuracy. Arif et al. (2014) used the same accelerometer dataset of Kwapisz et al. (2011) to classify the same six PA types. Optimal features derived from raw accelerometer taking into account correlation coefficients helped to increase the accuracy of prediction to 97% using K nearest neighborhood. To achieve a higher level of predictive accuracy, a combination of the accelerometer and other built-in sensors in smartphones (e.g., gyroscope, magnetometer) was also explored (Anguita et al., 2012; Shoaib, Scholten, & Havinga, 2013). With the recognition technique, PA monitoring systems were developed to detect abnormal activities of elderly people or people in need of assistance for healthcare purposes (Anguita