Emerging Data Mining for Pedestrian and Bicyclist Monitoring: A Literature Review Report

Project Number: UTC Safe-D 01-003
Project Title: Data Mining to Improve Planning for Pedestrian and Bicyclist Safety

Kyuhyun Lee, Texas A&M Transportation Institute
Ipek N. Sener, Ph.D., Texas A&M Transportation Institute

September 2017

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded by a grant from the U.S. Department of Transportation's University Transportation Centers (UTC) Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Acknowledgments

This research was funded by a grant from the Safety through Disruption (Safe-D) National UTC supported by the U.S. Department of Transportation through the UTC program. The research in this report was developed as part of a collaborative project between San Diego State University, the Texas A&M Transportation Institute, and the Virginia Tech Transportation Institute. The authors would like to thank Arash Jahangiri from San Diego State University, and Kapil Arya, Sirajum Munira and Shawn Turner from the Texas A&M Transportation Institute for their comments and feedback on a draft version of this report. The authors would also like to acknowledge Michelle Benoit from the Texas A&M Transportation Institute for her editorial review of the report.

Abstract

While traditional pedestrian and bicyclist monitoring methods require active efforts from data collectors, advancements in technology have made it possible to use the proliferation of mobile phones to capture real-world walking and bicycling patterns. Nearly ubiquitous mobile phone use (and more recently smartphone use) has made users both passive and active contributors to the emerging data collection methods. As passive contributors, mobile phone users routinely send their spatio-temporal information to cellular networks, satellites, and application servers. Secondary vendors purchase and process these data for sale. As active contributors, those who are willing to record and quantify their physical activities readily accept being monitored by activity tracking apps. This new method of data collection, data crowdsourcing, has been linked to the explosion of data availability in the non-motorized travel research domain, where traditionally pedestrians and bicyclists have been under-sampled and poorly understood.

These emerging methods promise new opportunities, but much work remains to fully realize the potential of accessible data and practices. This study reviewed currently available crowdsourced data and current use of these data. The review focuses on pedestrians and bicyclists, and includes both passive and active contributions.

The results reveal that bicycling data crowdsourced by active contributors have many possible research and practical applications, but pedestrian data rarely receive attention. Both pedestrian and bicycle data crowdsourced in a passive way are not currently available due to a high level of uncertainty and low locational precision. Even if the passively crowdsourced data are fully available, obtaining contextual information such as socioeconomic characteristics and trip purpose at the individual journey level is challenging. More efforts are needed to improve data accuracy and develop robust data fusion techniques.

Table of Contents

List of Figures

List of Tables

Executive Summary

While traditional pedestrian and bicyclist monitoring methods have relied on intensive efforts from data collectors, emerging monitoring methods that use location-aware mobile devices (hereafter termed simply *mobile devices*) require relatively fewer resources but make it possible to collect a huge amount of travel data. With nearly ubiquitous mobile device use, this new method of data collection, *data crowdsourcing*, has been linked to the explosion of data availability in the non-motorized travel (walking and bicycling) research domain, where traditionally pedestrians and bicyclists have been under-sampled and poorly understood.

These emerging methods promise new opportunities to capture real-world walking and bicycling patterns, but much work remains to fully realize the potential of accessible data and practices.

This study reviewed currently available data collected by various emerging methods that take advantage of mobile devices (e.g., mobile phones, smartphones, tablets, and wearable wristbands) and their current use.

The review starts with an overview of pedestrian and bicycle data sources, including both traditional and emerging data sources. Based on the level of input (interaction) of the traveler, pedestrian and bicycle data are classified as:

- **Passive data**: No/little input (interaction) from pedestrians and bicyclists is needed.
- **Active data**: Active input (interaction) from pedestrians and bicyclists is needed.

Emerging Passive Data

Emerging passive data collection methods do not necessarily require active input from pedestrians and bicyclists. The spatio-temporal information of travelers is routinely stored based on wireless technologies: mobile phone positioning, global positioning systems (GPS), and location-based services. Data analytic companies purchase the initially collected data and sell them again after data processing.

Emerging passive data are not currently available for non-motorized travel monitoring. The products provided by commercial vendors usually focus on vehicle trips rather than non-motorized trips due to limited positional precision, the short trip distances of walking and bicycling, and subsequent uncertainties of mode detection. One company, StreetLight, recently provided a preliminary dataset of pedestrians around transit stations on a trial basis. StreetLight is currently developing pedestrian and bicycle travel metrics, which are still in the research and test stage.

Considering that passive data account for a significant proportion of the total population, passive data have greater potential for non-motorized travel monitoring, compared to traditional monitoring methods. However, even if passive data are fully available, getting contextual information such as socioeconomic characteristics and trip purpose at the individual journey level is challenging unless the limitations are surmounted, such as privacy concerns, data precision, and technical skills, to successfully infer transportation modes.

Emerging Active Data

Emerging active data sources include regional bicycle tracking apps, fitness/activity tracking apps, bike-share programs, and user-feedback map inventories. This type of monitoring necessitates active contributions in the data collection process from travelers.

So far, non-motorized travel-monitoring tools are more concentrated on active data, especially for bicycling. Since the launch of CycleTracks in 2009, many public agencies have developed GPS-based bicycle tracking programs (e.g., Cycle Atlanta, Mon RésoVélo, and CycleLane) to better understand bicycle traffic patterns in their regions. In addition to basically gathered opt-in users' GPS traces (time and route), demographic characteristics, trip purpose, comfort level, and other information are voluntarily collected.

As for commercial apps, among an increasing number of fitness/activity tracking apps, Strava sells its users' bicycling GPS traces. Compared to the earlier regional bicycle trackers, Strava makes its data commercially available in a more extensive area as long as the region of interest contains Strava users. However, due to privacy issues, trip distance, trip speed, trip purpose, age, and gender are provided at the aggregate scale.

Bike-share programs are another source of bicyclist monitoring that can support policy making. Operation records can be used to identify the travel patterns of public-bicycle borrowers but provide limited trip information depending on the number of docking stations, the success of the program, and the presence of a GPS tracking function.

In the context of public engagement in the planning process, user-feedback-based map inventory apps can convey stakeholders' opinions to transportation planners and practitioners. Soliciting local knowledge from smartphone-carrying pedestrians and bicyclists can provide direct information about the location of interest (e.g., potholes) or new infrastructure needs (e.g., sidewalks).

With the success of these apps for active data, increased data availability in time, space, and volume has enabled many bicycle travel behavior studies including route-choice modeling, bicycling traffic volume analysis, collision exposure estimation, and evaluation of new facility provision at the entire network level. These kinds of studies have typically been limited due to lack of data.

While the apps sponsored by agencies (e.g., CycleTracks, Cycle Atlanta, and Mon RésoVélo) have extensive data coverage, such as individual-level trip records and sociodemographic features (from volunteers), they are limited to the geographical boundaries where the app is operated (e.g., San Francisco, Atlanta, and Montreal). Such limitation might be overcome by commercial apps (e.g., Strava) because these apps are used more broadly (e.g., globally). However, commercial app data also suffer from data limitations in providing additional travel information at the disaggregate level. Given that each emerging data collection method has its own strengths and drawbacks, jointly applying various sources of data would generate synergistic effects. Combining traditional data and emerging data can be a strategy for non-motorized travel planners who might face sample bias or limited sample size issues. Data collectors and providers must also consider privacy issues as long as personal information is required for the purpose of research.

While active data are beneficial for bicycle monitoring, limited sources are currently available for pedestrian research. Special attention needs to be paid to take advantage of crowdsourced emerging data for pedestrian monitoring.

Introduction

With growing attention on the benefits of non-motorized travel (i.e., walking and bicycling), the need for accurate, timely pedestrian and bicyclist travel data has increased over the past decades. Non-motorized travel modes have unique characteristics; their trips are more sensitive to the environment, more variant, and shorter than motorized trips. Walking and bicycling data have traditionally relied on counting at limited locations and on travel surveys for multimodal transportation (Ryus et al., 2014). Because most of the traditional monitoring methods require collectors' active input, which can be quite costly and timely, non-motorized travel data have often been limited by small sample size, time and budget constraints, and infrequent updates.

Over the last 10 years, advancements in technologies and proliferation of smartphones, along with increased demand for detailed information on non-motorized modes, have brought interest in new monitoring methods. While traditional data collection methods are still widely used, researchers and practitioners are investigating emerging methods that use the contributions of crowds through mobile devices with wireless technologies.

This study provides a review of emerging data sources in the context of non-motorized travel and with a specific focus on crowdsourced data using mobile devices. The review starts with an overview of pedestrian and bicycle data sources. Next, emerging data are examined extensively in terms of their data sources, types, and current and potential usage. Finally, the limitations of emerging data are discussed.

Overview of Pedestrian and Bicycle Data Sources

When pedestrian and bicycle monitoring programs started, monitoring tools were typically limited to travel surveys and manual counts. Technological advancements have created a variety of data collection methods that require less effort and fewer resources, while producing a larger volume of data than traditional surveying and counting, as in the cases of emerging crowdsourced data available through mobile devices and apps.

To provide a more structured, effective, and easy-to-follow evaluation of data characteristics, pedestrian and bicycle travel data are classified as *passive data* or *active data* in this review, according to the degree of action needed for traveler input and effort:

- **Passive data**: No/little input and effort from pedestrians and bicyclists are needed.
- **Active data**: Active input and effort from pedestrians and bicyclists are needed.

Figure 1 gives an overview of how data sources are classified in this study, followed by discussion of these data sources.

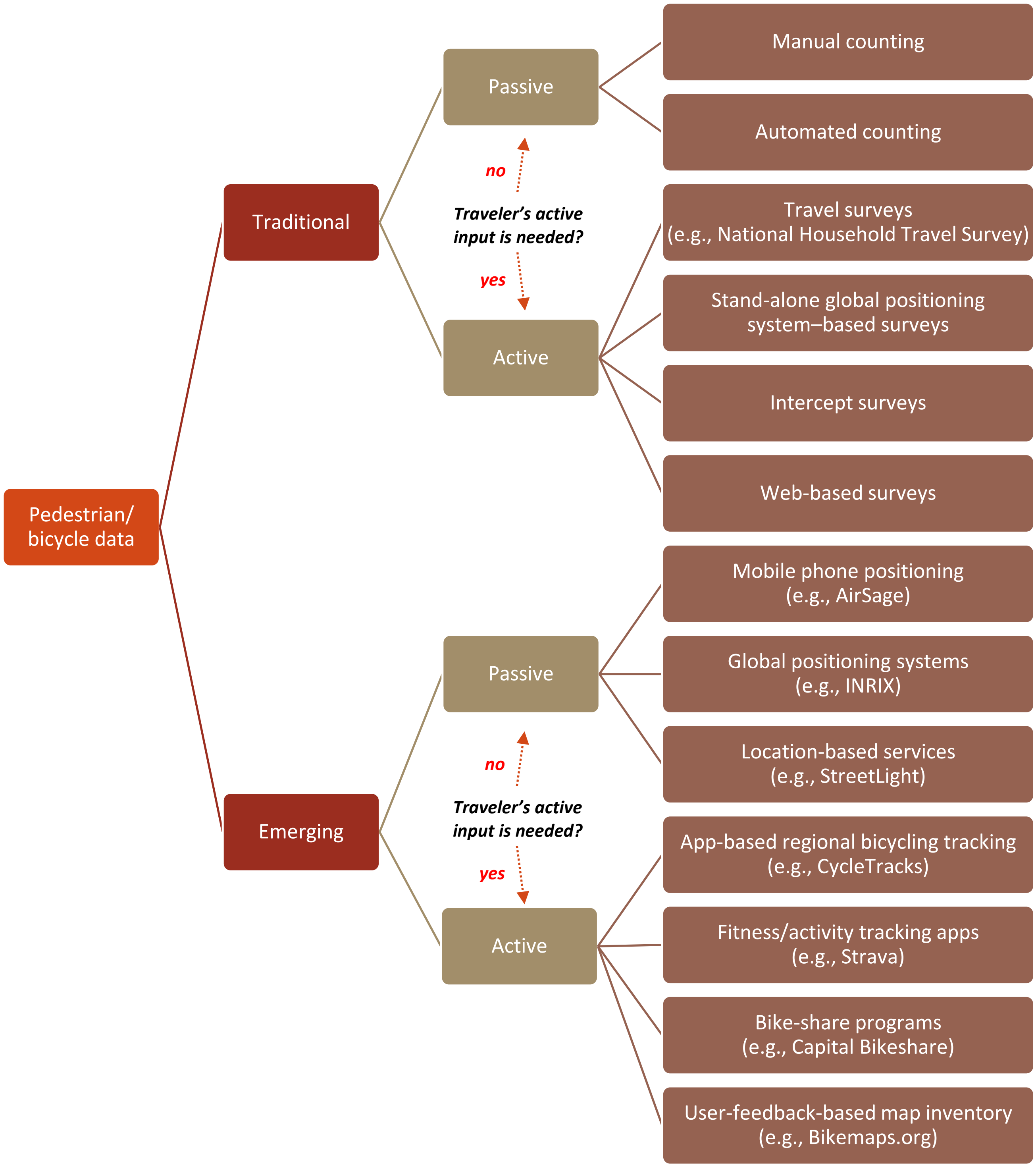

Figure 1. Classification of Pedestrian and Bicycle Data Sources.

Traditional Data Sources

This section provides a brief explanation on traditional efforts to facilitate the discussion on emerging data (as the main focus of this review) in moving forward.

Site counts are the most traditional data monitoring methodologies and are used to directly measure bicycle and pedestrian data. Since counting does not require active solicitation of travelers (i.e., from pedestrians and bicyclists in the current context), it is a representative passive data monitoring methodology. The primary techniques include manual and automated counting. Manual counting is performed by human data collectors in the field or using video recordings. This is the method that is still primarily used to collect pedestrian and bicycle traffic volumes (Ryus et al., 2014). Since accuracy depends on collectors, pre-training and instruction are necessary (Schneider et al., 2009). Manual counting is the preferred method of data collection when specific information on travelers' behaviors and attributes, such as helmet wearing and gender, is needed. This option is preferred, especially when there are constraints in securing financial resources, technical capacity, or regulatory permissions for deploying automated detectors. Techniques for counting have expanded to automatic methods using diverse sensors, such as pneumatic tubes, inductive loop detectors, passive/active infrared, and radio beams. When traffic volume over a lengthy period of time is needed, automated detectors can substitute for human data collectors. Pedestrian signal actuation buttons can be used as a reasonable proxy for determining rough pedestrian demand (Day et al., 2011).

Directly measured counts are generally used to estimate traffic volume at specific transportation facilities or at more aggregated area-wide geographies. Since counts are typically available at a limited number of points or segments (micro-scale), it is practically impossible to monitor every intersection and road segment in a region (macro-scale).

Traditional active data collection is mostly based on surveys, in which subjects are asked to describe their activities and travels in detail via a travel diary, GPS device, interview, or web-based questionnaires. Several types of surveys have been conducted to collect travel information in the United States, such as national travel surveys, GPS-based surveys, and intercept surveys.

Among the national surveys, the two commonly known travel surveys conducted in the United States are the U.S. Census American Community Survey (ACS) and the National Household Travel Survey (NHTS). The ACS provides commute modes with demographic data, which are updated every year. However, trip purposes are confined to only the journey to work, and survey respondents are asked to indicate one commute mode that is predominantly used (Griffin et al., 2014). Compared to the ACS, the NHTS encompasses all travel purposes but is updated infrequently. One issue with the NHTS from a non-motorized monitoring perspective is that survey participants tend to underreport walking and bicycling. In an effort to increase data confidence, participants are asked to carry GPS loggers. However, even with the combination of GPS and traditional survey, the NHTS still asks people to complete trip diaries, which can trigger survey fatigue and less faithful answers (Lee et al., 2016). Although both surveys provide contextual information, such as trip purpose, resident status, and income level, which are important parameters in estimating travel behaviors, a great deal of preparatory work and post-data processing are required for the nationwide data. This might result in high costs and small sample sizes at the local level that might be particularly problematic for non-motorized travel monitoring.

Some other types of surveys, such as GPS-based surveys or intercept surveys, can provide more specific data when focused on non-motorized travel, but these surveys still suffer from not being able to sample adequately to capture the non-motorized activity over the entire network and area-wide geographies. While survey methods provide more comprehensive information on pedestrians and bicyclists than counts, small sample sizes and financial constraints are challenging.

For more information on traditional data collection in the context of non-motorized travel, readers are referred to Turner et al. (2017).

Emerging Data Sources

During the last decade, advancements in technology and nearly ubiquitous use of mobile phones (and more recently smartphones) has yielded to the development of a new era of data collection. Researchers and practitioners can now use mobile-device-based crowdsourcing to identify travel patterns from mobile device users. For example, mobile phone users' locational information data routinely sent to mobile phone networks are commercially available, and the mobility patterns of smartphone app users are voluntarily or unknowingly shared with app developers.

Emerging data collection methodologies promise new opportunities because they can provide massive amounts of non-motorized travel information about a broad and diverse sample of the population with fewer time and resource constraints. However, much work remains to fully realize the potential of accessible data and practices. While some data are already commercialized, others are not yet readily available or have limitations to being used for capturing pedestrian and bicyclist travel patterns.

This study reviewed currently available data collected by various emerging methods and their current use. Not intending to be completely comprehensive, the review provides resources for the transportation community interested in emerging methodologies in non-motorized travel monitoring. Knowledge obtained from the literature is expected to be beneficial not only to improve pedestrian and bicycle safety, but also to plan a better travel environment for non-motorized travelers. The review considers both passive data and active data in detail.

Passive Data

Emerging passive data require no or minimal levels of direct interaction with the mobile phone users being monitored. With emerging passive data collection methods, millions or billions of location data points are routinely generated every day. Once raw datasets are obtained, they have to be aggregated, anonymized, filtered, and matched to road networks by intensive data processing. Secondary vendors usually purchase the raw datasets from initial data collectors (e.g., mobile phone carriers and app developers) and sell them again after data mining. Since data attributes differ by collection technology (mobile phone positioning, GPS, and location-based service), product output (e.g., trip types, data precision, and sample sizes) varies by technology adopted during data collection.

Passive data covered in this study focus on three types of emerging data collection technologies:

- **Mobile phone positioning (MPP).** Mobile providers routinely collect mobile phones' spatio-temporal information for operational purposes without user intervention.
- **GPS.** GPS-enabled devices record actual travel routes and times connecting to satellites every few seconds.
- **Location-based service (LBS).** A growing number of smartphone apps use LBS to determine users' location when they check in, even if the app is not initiated (as long as the app has access to location information, even when it works in the background).

Active Data

Compared to emerging passive data, emerging active data require travelers' input and efforts, such as willingness to participate in data collection to some extent during their physical activity. Active monitoring helps facilitate the collection of data associated with pedestrians and bicyclists in a more targeted manner (in terms of mode detection) compared to passive data. For example, fitness tracking apps collect mobility data only for the opt-in users who install and initiate the app to record and check their walking, bicycling, or running.

In this study, emerging active data are categorized into four types of non-motorized traveler tracking:

- **Regional bicycling tracking.** Public agencies develop regional bicycling tracking apps to collect bicycling travel patterns.
- **Fitness/activity tracking.** Private companies develop fitness/activity tracking apps, which provide tracking services to those who are willing to quantify and record their physical activities (e.g., bicycling, running, and walking) and other daily activities.
- **Bike-share programs.** Bike-share programs can be a way to monitor the bicycle community that borrows a public bicycle for a relatively short-term use. App-based operation enables users to actively and easily search nearby stations, available bikes, and docks.[1]
- **User-feedback-based map inventory.** User-feedback-based map inventory apps (web-based apps) gather public input such as safety issues (e.g., crash locations) and facility demands.

Passive Data: Sources, Types, and Usage

This section describes which emerging passive data sources are currently available, who provides the passive dataset, what types of data are provided, how the data are used in the non-motorized transportation planning field, and the data's potential capabilities.

Sources

Emerging passive data sources differ depending on which technology is used for initial data collection. MPP, GPS, and LBS have different levels of locational precision and coverage area, and thus different data output by provider.

Table 1 describes various technologies employed in passive monitoring, how these technologies work to collect raw data, and the attributes of these datasets.

[1] It may seem unusual to categorize bike-share programs as active data because bicyclists' input to track their travel is not necessarily required, unlike CycleTracks or Strava (e.g., initiate the app to start tracking). However, in the context of the current study, it is considered active input when service users search nearby stations, go to docking stations, and hire and return a bicycle.

Table 1. Emerging Passive Data by Monitoring Technique.

Data Description	MPP	GPS	LBS
Monitoring point	When mobile phones connect to cellular operator's networks	When mobile phones receive signals from satellites	When the LBS app is initiated (in the foreground) and the device begins moving (in the background) by Wi-Fi and assistive-GPS (A-GPS)[2] (varies based on technology)
Locational precision	200 to 1,000 m	5 m	From 5 m (A-GPS) to 50 m (Wi-Fi)
Detection coverage	Up to the traffic analysis zone (TAZ), census block, or road on which the device is located	Up to a small road or parking lot	Up to most parking lots, TAZs, and blocks
Example vendors	AirSage (www.airsage.com)	INRIX (inrix.com), TomTom (www.tomtom.com), and HERE (here.com*)*	StreetLight (www.streetlightdata.com) and Cuebiq (www.cuebiq.com)
Possible data	Aggregate origin-destination (OD), trip purpose (imputed), home/work location (imputed), and speed	Aggregate OD, trip purpose (imputed), home/work location (imputed), and speed	Aggregate OD, trip purpose (imputed), home/work location (imputed), and speed

MPP

Signaling between mobile phones and cell towers, called *mobile phone positioning*, is a key technology for cellular carriers to achieve billions of location data points by time from their cell-phone service consumers. While users are on the phone (voice calls, text messages, and data sessions), cellular networks locate callers and the persons they are calling. In the process of locating them, cellular carriers collect cellular users' geolocations by time (Chen et al., 2016). As mobile phone users move, cell towers collecting geospatial information are switched. Through this tower changing, travel patterns are traced (Huntsinger and Donnelly, 2014). However, since tracked geolocation is of cell towers, MPP data do not necessarily represent the exact X and Y coordinates of places visited or passed (Chen et al., 2016). Positional accuracy is determined by cellular network coverage (called cell size). The commonly recognized spatial precision of MPP data ranges from 200 m to 1,000 m (typically better in urban areas than in rural areas due to higher cell tower density) (Bowman, n.d.). This precision may affect detecting short trips. For example, for 1,000 m precision, it is difficult to accurately capture trips shorter than 1,000 m. Therefore, walking trips, which cover relatively short distances, are more likely to be missed in the process of data cleaning.

AirSage, a company founded in 2000 in Atlanta, is a pioneer in purchasing and supplying MPP data in the United States. According to the company, it "generates billions of anonymous location data points, transforming terabytes of signaling data every day" (AirSage, 2017a). Geospatial datasets provided by AirSage can present 10 to 30 percent of the total population based on cellular carriers (Bowman, n.d.). The data product includes traffic speed and volume on roadways and trip OD at aggregated levels such as census tracts or TAZs (Huntsinger and Donnelly, 2014). Due to privacy issues and location precision, MPP data should be

[2]A-GPS helps start up the connection (e.g., it helps GPS work in a building).

aggregated. AirSage's data are suitable for vehicle travel volume analysis at the macro-scale (Smith, 2015). These data are not yet available for capturing a pedestrian's and bicyclist's trip patterns.

GPS

GPS collects GPS-enabled devices' geolocation and velocity information, measuring the distance between mobile phones and satellites every few seconds. Spatial precision is about 5 m, which is better than MPP and LBS (Bowman, n.d.). This high positional accuracy makes it possible to detect where the device is at the small road or parking lot scale. However, GPS is primarily used for outdoor positioning because satellite signals do not work well inside buildings. Navigation-based GPS datasets are collected by navigation or mapping service companies. Several companies such as INRIX, TOMTOM, and HERE sell their GPS probe data gathered from their real-time traffic and route-choice services.

GPS data alone can be used to identify the most popular routes, worst traffic spots, average speed, annual average daily traffic, and OD for vehicles. However, like MPP data, GPS data provided by data vendors lack important demographic attributes in detail because raw GPS points should be anonymized and aggregated to deliver traffic parameters. One company, StreetLight Data, Inc., integrates GPS data from INRIX with LBS data from Cuebiq to infer contextual details at the aggregate level.

LBS

Smartphone apps that provide LBS essentially depend on users' physical location (e.g., Yelp, TripAdvisor, and Waze). When apps are operated, users' location information is collected through Wi-Fi proximity, A-GPS, and other wireless technologies. Even when LBS apps are not in use, some of the apps collect data once opt-in users begin to move. Compared to MPP data, LBS data have higher locational precision (5–50 m) (Bowman, n.d.). While GPS data are better in explaining where people move, LBS data better describe where people stay and why people go there (e.g., shopping, home, and work). The high spatial precision of LBS makes it possible to infer sociodemographic characteristics to some extent such as income, home and work location, and trip purpose. Moreover, the spatial patterns and detailed speed information of users' movements help to infer transportation mode recognition, which is extremely difficult to identify with cellular data.

Cuebiq is one of the companies that provide aggregated LBS datasets. This company's database comes from partnerships with almost 200 smartphone LBS app developers, and the large sample size is beneficial to secure representativeness (Schewel, 2016). For example, Cuebiq's samples account for over 10 percent of the adults in the United States, and this percentage is growing over time. Considering that small cellular providers' market share is about 10 percent, this large sample size is promising to capture real-world geo-behavioral patterns.

To take advantage of LBS data, StreetLight partnered with Cuebiq in 2016. The data partnership made it possible for StreetLight to expand its LBS sample size to over 30 million devices (Schewel, 2016). StreetLight data are unique compared to other data vendors' data because they combine geo-spatial information and textual datasets (census data) so that customers (who purchase their data) can consider socioeconomic factors. StreetLight provides an easy-to-use-online platform, StreetLight Insight, where clients can directly analyze and visualize (e.g., using a heat map) from their own web browser in a few minutes. StreetLight data are provided in a shapefile for geographic information systems. However, StreetLight or other venders do not yet provide readily available pedestrian- and bicyclist-customized datasets.

Applications

While emerging passive data have been increasingly used in vehicle travel research, there has been a dearth of applications for pedestrians and bicycles. AirSage's MPP data were used to estimate aggregated traffic parameters such as travel time and speed, traffic volume, and traffic flow; and to model OD pairs, daily activity patterns, and regional travel demand (Liu et al., 2008; Calabrese et al., 2013; Huntsinger and Donnelly, 2014; Çolak et al., 2015). Similarly, researchers applied INRIX observations to estimate real-time speed on a vehicle travel network and to analyze travel time (Kim and Coifman, 2014; Cui et al., 2016; Barajas et al., 2017). TomTom data were applied to produce time-of-day link travel speed and travel time statistics on selected highway routes and segments, and traffic flow (Kerner et al., 2013; Nie et al., 2013). So far, almost every study using emerging passive data (and that have been reviewed during this review) is limited to vehicle trips.

However, StreetLight's data were recently used in a study to understand pedestrian travel patterns within last-mile trips near transit stations (McCahill, n.d.). This study analyzed two months of pedestrian trips (from June and July 2016) near two light-rail stations in Sacramento, CA. StreetLight provided the data after they were anonymized, processed, and aggregated. The study findings indicate that better accessibility was more likely to promote walking to transit stations, and people from neighborhoods that have poor connections to stations (e.g., a highway between the neighborhoods and stations) walked to transit stations to a greater degree than expected. Considering that it is challenging to gather ground-truth data of last-mile trips through conventional methods, the study shows that the emerging data have potential for diverse types of pedestrian and bicycle trips including a connecting mode for multimodal trips. StreetLight is currently developing pedestrian and bicycle travel metrics, which are still at the trial stage (McCahill, n.d.; StreetLight, 2017).

Potential Capabilities

Despite limited applications in the pedestrian and bicyclist monitoring domain, passive crowdsourced data have the potential to overcome the shortcomings of conventional monitoring methods. Walking and bicycling travel patterns have traditionally been more poorly understood than vehicle travel patterns, which is in part due to lack of adequate data acquisition. Data analytic companies, such as StreetLight and Cuebiq, began to work in partnership with other vendors to integrate different data sources and overcome the shortcomings of passive data (e.g., AirSage has agreements with GPS carriers) (AirSage, 2017b). Once emerging passive data are fully accessible and available, these data will allow researchers and practitioners to:

- Use high-quality, readily available data at reduced cost, less intervention, and less burden than traditional data.
- Monitor changes in behaviors and traffic volumes from the neighborhood to the nationwide scale.
- Conduct continuous and long-term follow-up monitoring.
- Identify spatial and temporal regularities/variations for time of day, for day of the week, or over time.
- Evaluate the effects of the new provision of infrastructure on promoting walking and bicycling (through before and after comparison of the facility).
- Predict future movements or detect abnormal patterns based on travel routines.
- Enrich knowledge about pedestrian and bicycle travel behaviors.

Active Data: Sources, Types, and Usage

This section describes which emerging active data sources are currently available, who provides the active dataset, what types of data are provided, how the data are used in the non-motorized transportation planning field, and the data's potential capabilities.

Sources

Emerging active data come from active participation of individuals in tracking their bicycling and physical activities through smartphone apps. A bike-sharing system is another tool for bicyclist monitoring. In other cases, user-feedback-based crowdsourcing apps become venues where communities actively inform safety issues and needs for specific facilities. Table 2 shows examples of active data sources and providers. As discussed previously, in this study, active crowdsourced data collection tools are categorized into four types:

- Regional bicycling tracking.
- Fitness/activity tracking.
- Bike-share programs.
- User-feedback-based map inventory.

Collected data types vary depending on who developed the monitoring program and for what reason. Details on the differences in data type are also shown in Table 2.

Table 2. Examples of Emerging Active Data Sources and Providers.

Provider	Classification	Collected Data	Website	Note
Regional Bicycling Tracking				
CycleTracks	Regional bicycling tracking app in San Francisco, CA, and other cities	GPS trace, age, gender, trip purpose, trip frequency, email address, ZIP code	www.sfcta.org	The first regional bicycling tracking app in the United States
Cycle Atlanta	Regional bicycling tracking app in Atlanta, FL	GPS trace, age, gender, trip purpose, trip frequency, email address, ZIP code, ethnicity, income, rider type, rider experience, note, picture	cycleatlanta.org	A branch of CycleTracks
Mon RésoVélo	Regional bicycling tracking app in Montreal, Canada	GPS trace, age, gender, trip purpose, trip frequency, email address, ZIP code, picture, greenhouse gas emission, consumed calorie	—	A branch of CycleTracks
CycleLane	Regional bicycling tracking app in Eugene, OR	GPS trace, age, gender, trip purpose, trip frequency	www.thempo.org//611/cyclelane--bicycle-routes	A branch of CycleTracks
ORcycle	Regional bicycling tracking app in Portland, OR	GPS trace, trip purpose (mandatory), route comfort (mandatory), trip frequency (mandatory)	www.pdx.edu/transportation-lab/orcycle	A branch of CycleTracks
Fitness/Activity Tracking				
Strava	Fitness activity tracking app	GPS trace, traffic volume, traffic flow	www.strava.com	Commercially available product format: *.shp, *.dbf
Endomondo	Fitness activity tracking website/app	GPS trace, activity duration, distance, sport type, average speed	www.endomondo.com	Some data partially open to the public
Fitbit	Wristband fitness activity tracking device app (web based)	GPS trace, step counting, others	www.fitbit.com	—

Provider	Classification	Collected Data	Website	Note
Bike-Share Programs				
Capital Bikeshare	Bike-share program in Washington, DC	Trip OD and time, others	capitalbike share.com	—
CitiBike	Bike-share program in New York, NY	Trip OD and time, others	www.citibike nyc.com	—
BIXI Montreal	Bike-share program in Montreal, Canada	Trip OD and time, others	montreal.bixi. com/en	—
User-Feedback-Based Map Inventory				
Bikemaps. org	Crash reporting website/app	Cycle incident, time and type, trip purpose, collision object	bikemaps. org	—
MySidewalk	Walking and missing sidewalk tracking app	Walking path, missing sidewalk, sidewalk status	www.myside walk.com	—

Note: GPS trace includes time stamp and GPS coordinates; ZIP code includes home, school, and work.

Regional Bicycling Tracking

All of the bicycling tracking apps were built by government agencies for public purposes. Most of the apps collect additional user information (optional) along with basically traced GPS data. This type of data is available at the disaggregate level and as OD pairs for each trip.

CycleTracks. The first active data source was bicyclist-specific apps developed by local agencies. The first work on a regional smartphone app for bicyclist monitoring started in 2009. The San Francisco County Transportation Authority (SFCTA) was able to estimate bicycle trip volume but could not assign the traffic to specific streets. For the purpose of building a route-choice model, SFCTA released CycleTracks, which collects trip time and space trajectories using the built-in GPS function of smartphones (Charlton et al., 2011). This app gathers additional (optional) personal information such as age, gender, ZIP code, and trip purpose to understand user bias. Privacy concerns about uploaded individual information are addressed by assigning a unique ID to each user (Hudson et al., 2012; Meyer, 2013). When users tap “start,” the app begins to track by connecting to GPS satellites. Once the trip is finished, user information, GPS trace (time stamp and GPS coordinates), and other information are sent to the SFCTA server. To test the feasibility of CycleTracks in another region, the Texas A&M Transportation Institute (TTI) applied it to an Austin area in 2011.

Cycle Atlanta. The initial success of CycleTracks inspired similar apps in other cities, regions, and countries. In 2012, the Georgia Institute of Technology and the City of Atlanta collaborated to launch Cycle Atlanta, which is a modified and rebranded CycleTracks. The Cycle Atlanta project team differentiated its app from the original one by adding new features (Watkins et al., 2016). One of the new features is that users can describe categorized bicycling issues (e.g., pavement, traffic signal, and bicycle facilities). Pictures of problematic spots can be submitted as well. In addition, users are asked to choose their ride history (among four: since childhood, several years, one year or less, and just started) and to classify their rider type (among four: strong and fearless, enthused and confident, comfortable but cautious, and interested but concerned). These parameters represent rider attitudes and comfort levels with bicycling infrastructure to measure risk aversion.

Mon RésoVélo. In 2012, CycleTracks and Cycle Atlanta were benchmarked for Mon RésoVélo in Montreal, Canada (Jackson et al., 2014). In addition to collecting basic information, Mon RésoVélo calculates greenhouse gas emissions and calories. For the sake of citizen convenience, this app provides two language versions, English and French.

CycleLane. In the same year, 2012, the Central Lane Metropolitan Planning Organization launched CycleLane in Eugene, OR. Upon download, this app prompts users to fill out a questionnaire about demographic information (age and gender) and the frequency of bicycling. After finishing trips, the app asks the user to input reasons for the trip. The trip records are then submitted to the server.

ORcycle. Two years later, in 2014, the Oregon Department of Transportation and Portland State University developed ORcycle (Blanc and Figliozzi, 2016). Unlike the previously mentioned apps, ORcycle does not collect users' personal information. Instead, three questions are asked concerning trip purpose, route comfort, and trip frequency. Riders can voluntarily select stress sources on their bicycle routes, choosing from auto traffic, trucks, parked vehicles, pedestrians, etc.

Others. CyclePhilly in Philadelphia, PA, CycleDixie in Auburn, AL, and RiderLog in Sydney, Australia, are other examples of regional bicycling apps to stimulate bicycling and track bicycle trip patterns.

Fitness/Activity Tracking

Fitness/activity tracking apps collect opt-in users' information on physical activities such as walking, running, and bicycling. One of the apps sells tracked records, and some of the apps that are based on social networking share users' data with the public.

Strava. One of the most commonly used physical activity tracking apps is Strava developed by Strava, Inc., in 2009. This mobile and web-based tracker was originally intended to track athletes' performance via GPS. However, now millions of runners, bicyclists, walkers, and hikers across the globe use the app. Current data volume is more than 300 billion GPS points, and every week 7 million sport activities are uploaded (Strava Metro, 2017a). Strava collects opt-in users' activity date, time, and route. Unlike the earlier regional bicycle tracking apps including CycleTracks, Strava data are commercially available through Strava's data service (Strava Metro). Strava Metro data have been increasingly used by more than 85 cities and organizations globally (Strava Metro, 2017a). Three licenses can be purchased based upon data aggregation unit: node (point), street (segment), and OD (polygon). The product format includes shapefiles and database files for geographic information systems (Smith, 2015). Data charges depend on the number of Strava members in the requested region: $0.8 per user for the first 10,000 and $0.7 thereafter; the minimum cost is $1,000, and data cover one year (Strava Metro, 2017b). While Strava collects all types of bicyclist data using a trip purpose tag, filtering is only possible by commute. Due to privacy issues, disaggregate trips and demographic information are not available at the discrete level. In other words, provided datasets do not contain trip distance, trip speed, age, and gender at the individual scale (Sun and Mobasheri, 2017).

Endomondo. Endomondo, a fitness tracking app, is similar to Strava. Endomondo has millions of users around the world (Endomondo, 2017a). The goal of this app is to be a personal trainer in a pocket (Endomondo, 2017b). To make fitness more fun, Endomondo allows users to inspire other Endomondo communities by sharing workout results on social networking sites such as Facebook and Twitter (Endomondo, 2017c). Some of the shared results are open to everyone so that the public profile can be downloaded from the web server through web scraping (using an automated program to find and store the relevant data elements presented by the webpage) (Cortés et al., 2014; Romanillos et al., 2016; Qiao, n.d.). Other apps with public access to data (similar to Endomondo) include GPSies, Wikiloc, and outdoor navigation apps (Santos et al., 2016).

Fitbit. Fitbit is an app/web-based wearable device. Fitbit produces wearable wristbands to measure personal health-oriented activity metrics such as stepping count, heart rate, and quality of sleep (Fitbit, 2017). Fitbit devices that have GPS tracking functionality can record GPS traces when wearers turn on the GPS tracking mode. The Fitbit app can be installed on GPS-enabled products and GPS-enabled mobile phones so that it can track route, pace, and exercise history. Data collected from the wristbands are available when volunteers are recruited to share the data for study purposes.

Bike-Share Programs

With the expansion of bike-share programs, usage records are another source for monitoring bicyclist trips. Most bike-share services are operated based on smartphone apps, which make it convenient for users to find, rent, and return bikes. Some bikes equipped with GPS can provide details of the route taken between every pair of stations. For bikes with no GPS sensor, check-in and check-out records at stations are used for trip OD data. However, when users return the bicycle to the same station where it was checked out, the record is not useful (Faghih-Imani et al., 2014). In addition, bike-share data may provide bicyclist demographic information such as age and gender.

Capital Bikeshare. Capital Bikeshare (CaBi) is an example of such a bike-share system. It is located in Washington, DC, and was established in 2010, with 3,700 bikes and 440 stations in the area (Capital Bikeshare, 2017). CaBi is run by a membership system: signing up daily, monthly, or annually (Ma et al., 2015).

CitiBike. New York City launched CitiBike in 2013 with 330 stations and 5,000 bikes (Kaufman et al., 2015). High-density and walkable urban form has provided users with easy access to bike-share services, contributing to CitiBike's success (Faghih-Imani and Eluru, 2016). Datasets collected by CitiBike include information on trips (trip OD and trip start/end time), riders (age, gender, and membership type), and station characteristics (capacity and coordinates) (Faghih-Imani and Eluru, 2016).

BIXI. BIXI is Montreal, Canada's bike-share program. As the first public bike-share system in Canada, BIXI started its service with 300 stations and 3,000 bikes in 2009 (Faghih-Imani et al., 2014). Currently, BIXI networks have 540 stations and 6,200 bikes (BIXI, 2017). Due to the harsh climate in Montreal, BIXI is operated for eight months of the year (April through November) (Morency et al., 2017).

User-Feedback-Based Map Inventory

User-feedback-based map inventory apps promote citizen engagement in the planning process by gathering citizen's localized knowledge and experiences. Community members volunteer to report needed improvements for infrastructure, desired change proposals, and hot spots where collisions occur. This type of app plays the role of digital channel for direct interactions between citizens and government employees (Le Dantec et al., 2015). These apps help bring community members together voluntarily to address issues that influence their community, which has useful implications for transportation planning (LaMondia and Watkins, 2017).

BikeMaps.org. BikeMaps.org is a safety data collection tool using a website and mobile app where citizens can report crash locations and information such as crash time, sight lines, and injury severity. This tool is sometimes called geo-crowdsourcing because citizens make a bicycle incident map by adding the collision location information to the map (Nelson et al., 2015). While formally reported bicycle incidents are more likely to miss incidents involving less severe injury, Bikemaps.org can collect missing data (officially not reported) as well as official crash data.

MySidewalk. Knowledge Based Systems, Inc., in cooperation with the City of College Station and TTI, developed a community-driven app for sidewalk inventory and condition assessment data, MySidewalk (Knowledge Based Systems, Inc., 2017). Sidewalks are one of the primary urban infrastructures for securing safe walking, but acquisition of up-to-date information on sidewalks in the entire network has been challenging to local authorities who are responsible for maintaining and planning sidewalks. Basically, MySidewalk lets the app users track their walking through the "start tracking" function. When faced with missing sidewalk, pedestrians can report the locations where sidewalks do not exist by starting "track missing sidewalk." In addition to the demand for new sidewalks, the app can detect informal pedestrian paths frequently used by the public. The app also collects the attributes of currently damaged sidewalks that need to be fixed by allowing app users to submit both descriptions and photos of damage. The uploaded geospatial data can be downloadable using geographical information system software. This digital civic engagement tool benefits planning organizations by helping them identify sidewalk issues and gaps in existing pedestrian networks.

Applications

While smartphone apps for human activity tracking have promoted an increasing amount of research on bicycle travel behaviors, a few applications for pedestrians (including runners) were found in the literature. Due to the small number of studies on pedestrian monitoring, data applications provided in the literature review are more concentrated on bicycling than walking. Table 3 lists active data applications.

Regional bicycling tracking apps provide more detailed data on the bicycle community than any other monitoring tool. CycleTracks and its branches mostly collect bicyclists' individual demographics such as age, gender, income (which only Cycle Atlanta collects), and ZIP code from volunteers. In addition, detailed trip information, such as travel time, trip distance, speed, and identification of OD, is available for each trip.

In terms of data density across the region of interest, Strava Metro offers a rich dataset. However, Strava Metro data do not provide personal profiles and trip details at the individual level. Strava data can only be purchased in anonymously aggregated forms due to privacy issues. For these data attributes, existing studies relying on Strava Metro data include analyses of aggregated traffic volumes and traffic flows (rather than, for example, disaggregate route-choice models).

In an effort to respond to concerns about the quality of the data collected from app-based GPS tracking, several studies have compared bicycle traffic volumes measured by mobile apps with other data from traditional sources (such as manual counts) (Griffin and Jiao, 2015a; Haworth, 2016; Jestico et al., 2016; Watkins et al., 2016).

Table 3. Emerging Active Data Applications in Existing Literature.

Reference	Coverage	Mode	Application	Source	Scale	Method
Regional Bicycling Tracking						
Charlton et al., 2011; Hood et al., 2011	San Francisco, CA	Bike	Route-choice modeling	CycleTracks	2,777 trips from 366 users for six months (Nov. 2009–April 2010)	Multi-nomial logit model
Hudson et al., 2012	Austin, TX	Bike	Demonstra-ting feasibility for route-choice modeling	CycleTracks	3,600 trips from about 300 users for six months (May–Oct. 2011)	—
Jackson et al., 2014	Montreal, Canada	Bike	Describing bicyclist profiles and trips	Mon RésoVélo	Over 2,300 trips from 512 users for 23 days (July 2012)	Descrip-tive analysis
Blanc and Figliozzi, 2016	Portland, OR	Bike	Bicyclist comfort level modeling	ORcycle	729 trips from 170 users for seven months (Nov. 2014–May 2015)	Ordinal logistic model

Reference	Coverage	Mode	Application	Source	Scale	Method
Leao and Pettit, 2016	Sydney, Australia	Bike	Validating simulated shortest paths	RiderLog	37 records from 16 users for 2010–2014	Agent-based model
LaMondia and Watkins, 2017	Auburn, AL, Atlanta, FL	Bike	Route-choice modeling	Cycle Atlanta, CycleDixie, Strava Metro	Cycle Atlanta; 989 users (July 2014); CycleDixie; Strava	Ordinal logistic model, binary logistic model
Zimmermann et al., 2017	Eugene, OR	Bike	Route-choice modeling	CycleLane	648 trips from 103 users	Recursive logit model
Fitness/Activity Tracking						
Griffin and Jiao, 2015a	Austin, TX	Bike	Data comparison	Automated counts, CycleTracks, Strava Metro	Counts at five sites; duration varies by data source (2011–2013)	—
Griffin and Jiao, 2015b	Travis County, TX	Bike	Bicycle volume analyzing (bicycle kilometers traveled [BKT])	Strava Metro	8,555 km traveled for a week in Aug. 2013	Ordinary least squares (OLS), geograph-ically weighted regression (GWR)
Strauss et al., 2015	Montreal, Canada	Bike	Estimating injury risk	Manual/ automated counts, Mon RésoVélo	Manual counts at over 600 intersections (2008–2009); automated counts at 30 sites; Mon RésoVélo: 137 days (July–Nov. 2013); 10,000 trips from 1,000 users	Linear regression, negative binomial
Haworth, 2016	London, United Kingdom	Bike	Data comparison	Manual counts, Strava Metro	Counts: 4,172 observations at 298 sites from 6:00 a.m. to 8:00 p.m.; Strava: April–May 2013	OLS

Reference	Coverage	Mode	Application	Source	Scale	Method
Jestico et al., 2016	Victoria, Canada	Bike	Data comparison	Manual counts, Strava Metro	Counts: at 18 count locations for 34 days; Strava: 74,679 routes from 3,650 users	General-ized linear model
Selala and Musakwa, 2016	Johannes-burg, South Arica	Bike	Visualizing bicycling flow	Strava Metro	84,297 trips for 2014	Spatial descriptive analysis
Watkins et al., 2016	Atlanta, FL, Midtown area	Bike	Data comparison	Manual counts, Cycle Atlanta, Strava Metro	Counts at 78 intersections (March 2013); Cycle Atlanta: Oct. 2012–Aug. 2014; Strava: 3,236 users, 51,408 total trips (Aug. 2013–July 2014);	Descrip-tive analysis
Heesch et al., 2016	Brisbane, Australia, a segment of bikeway	Bike	Evaluating new facility performance	Survey, Strava Metro	Survey: one-day survey at two sites before and after June 2013; Strava: monthly bicycle counts for six months before and after June 2013	Descrip-tive analysis
Hochmair et al., 2017	Miami-Dade County, FL	Bike	Analyzing bicycle volume (BKT)	Strava Metro	Jan.–May 2015 and Feb. 2015 (weekend and weekday)	Linear regression
Sun and Mobasheri, 2017	Glasgow, United Kingdom	Bike	Estimating air pollution exposure	Strava Metro	287,833 traces from 13,684 users	Bivariate local Moran's I statistic
Qiao, n.d.	Worldwide	All sports activities	Analyzing exercise duration	Endomondo (public data)	Workout data by web scraping 5.6 million workouts from 1.5 million users	Linear regression
Santos et al., 2016	Lisbon, Portugal, an urban park	Runner and bike	Detecting conflict exposure	GPSies (public data)	Running (N=73) and mountain biking (N=269) in March 2015	Spatial descriptive analysis

Reference	Coverage	Mode	Application	Source	Scale	Method
Proulx and Pozdnukhov, 2017	San Francisco, CA	Bike	Analyzing bicycle volume	Manual/ automated count, bike-share data, travel survey, Strava Metro	Varies by type	GWR
Zenko et al. 2017	—	Ped-estrian (walk and exercise)	Finding effective ways to promote exercise	Fitbit	164 participants' walking by the middle of 2018 (ongoing study)	Analysis of variance, Type 1 test
Bike-Share Programs						
Ma et al., 2015	Washington, DC	Bike	Analyzing impacts on transit use	Capital Bikeshare	Trips for 2013	OLS
Morency et al., 2017	Montreal, Canada	Bike	Analyzing member-based bike-share trip generation	BIXI	6 years of operation records (2009–2014) from 87,144 members and 444,340 occasional users	Negative binomial
User-Feedback-Based Map Inventory						
Jestico et al., 2017	British Columbia, Canada, capital regional district	Bike	Analyzing factors affecting crash	bikemaps. org	Incidents at/around intersections from 2014 to 2016	Poisson, negative binomial

Route-Choice Modeling

Because daily tracking of physical activity via smartphone apps allows a large sample size and more detailed information on each trip, many researchers develop bicycle route-choice models in a network. The first implementation of app-based GPS tracking in a route-choice model took place in San Francisco, CA (Charlton et al., 2011; Hood et al., 2011). A total of 2,777 traces from 366 users were modeled to find how network and environment attributes affected bicyclists' route choices. Bicyclists preferred to avoid turns; infrequent riders preferred bicycle lanes more than frequent riders; women and commuters disliked steep slopes.

The feasibility of using CycleTracks data for route-choice modeling was also demonstrated in Texas (Hudson et al., 2012). In 2011, TTI launched CycleTracks and collected over 3,600 trip routes. Researchers matched the GPS traces with bicycle networks for almost 90 percent of the data, proving the usefulness of CycleTracks data in developing a route-choice model.

Applications of bicyclists' tracking data are also found in other regions with rebranded names from CycleTracks. CycleLane, the regional bicycling tracking app in Eugene, OR, observed 648 bicyclist trajectories, and the chosen routes were applied to estimate a route-choice model (Zimmermann et al., 2017). Findings revealed that bicyclists are sensitive to specific route conditions and bicycle facilities.

LaMondia and Watkins (2017) conducted a comprehensive study for bicycle route network improvement. Two study regions were chosen for different environmental settings: Auburn in Alabama as a suburb and Atlanta in Florida as an urban core. Several datasets were used. Researchers worked with local bicycle communities to collect bicycle trip data. For the Auburn study, Strava app users provided the research team with their trip records after downloading their data from the Strava website. Volunteers shared membership accounts with researchers so that the research team could access their participants' account and directly download their trip history. In addition to Strava data offered by volunteers, the researchers purchased Strava data and collected bicycle trip records from a regional bicycle tracking app, CycleDixie (developed by Auburn University). These four datasets were merged to estimate route-choice modeling. For the Atlanta study, the Cycle Atlanta dataset was used. The study results reported that determinant factors affecting a certain type of link selection included sociodemographic characteristics, surrounding land use, and roadway features in Auburn. Among roadway characteristics, links that have high-peak-hour traffic volumes and wide shoulder widths were more likely to be selected. In particular, Cycle Atlanta requested the app users to self-classify based on confidence level. A route-choice model including the self-reported confidence level indicated that more confident bicyclists were more likely to prefer shortest routes over safe routes that need detouring.

Another regional bicycle app, Mon RésoVélo in Canada, reported bicyclist profiles and trip records for 23 days after the app launched (Jackson et al., 2014). This study did not model route estimation but rather focused on the adaptability of the new bicycling tracking system. However, the route data collected using the app showed potential for being applied to route-choice modeling because the data include travel time, distance, and route choice for each trip, which are essential for developing route-choice models. The app also quantified the health and environmental benefits of bicycling, including reduced greenhouse gas emissions and consumed calories.

Comfort Level Modeling

Beyond route-choice estimation, Blanc and Figliozzi (2016) used crowdsourced GPS data for the first time to model cyclists' comfort level in Portland, OR. Researchers surveyed bicyclists via the ORcycle app about their concerns while riding. Unlike previous studies, this app did not request personal profiles. Instead, opt-in users were asked to provide the following information: trip purpose, riding frequency, and concerns about safety.

Data Comparison

Griffin and Jiao (2015b) compared bicycling volumes collected by crowdsourcing and actual counts in Austin. This comparison was based on three types of data sources: Strava Metro, CycleTracks, and automated counting at given points. The comparison results indicated that bicycle traffic monitoring via app-based crowdsourcing is promising.

To evaluate the capability of Strava data to estimate bicycle flow in London, Haworth (2016) developed an OLS regression model using actual counts as a dependent variable. The adjusted r squared was above 0.6, meaning that Strava can represent real traffic flow. Similarly, another study in Canada indicated moderate association between GPS tracking volume and manual counts (r squared 0.4 to 0.58) (Jestico et al., 2016).

When Strava and Cycle Atlanta were compared in terms of user information, Strava opt-in users showed a skewedness for male. For age and commute, Cycle Atlanta users were younger than Strava users, and Cycle Atlanta had more commuters (Watkins et al., 2016).

Bicycle Volume Analysis

Using ridership data per segment provided by Strava Metro, two studies aggregated BKT by multiplying bicycle counts by segment length. Griffin and Jiao (2015a) aggregated BKT again at the block group level and regressed OLS models to determine the effects of socioeconomic features and the built and natural environment on bicyclists. The results showed that the job-housing balance was positively related to ridership. Additionally, Strava users preferred bicyclist-supportive facilities (bicycle lanes, shoulders, and paths) for commuting, and those who ride a bicycle for leisure preferred steep hills.

The other study conducted in Miami-Dade County, FL, predicted the impacts of weekdays and weekends on bicycle ridership separately by census track (Hochmair et al., 2017). For commuter BKT, employment was an important predictor, and trails increased non-commuter BKT.

Proulx and Pozdnukhov (2017) developed an innovative method to estimate bicycle volume across the networks in San Francisco. Researchers fused Strava Metro data with bike-share program data, manual and automated counts, and data from two regional full-population travel demand models. The results revealed that combining the given data improved model predictive accuracy.

A study by Morency et al. (2017) provides an example of a longitudinal analysis of bicycle trip volume using six years of bike-share program operation records in Montreal, Canada. Researchers estimated indicators that affect the levels of BIXI usage, such as temporal elements (e.g., day of the week), weather conditions (e.g., temperature and rainfall), station attributes (e.g., number of stations, capacity, and elevation), and neighborhood attributes around stations (e.g., residential density). The factors that positively influenced the use of the bike-share services included the capacity of stations, whereas negative impacts were in part associated with adverse weather conditions and high elevations of stations.

Exposure Estimation

Mon RésoVélo and Strava travel data were also applied to estimate bicyclists' exposure to conflicts and air pollution based on traffic counts. By combining annual average daily bicycle ridership predicted from field counts with Mon RésoVélo bicycle volume, injury risks were measured at both the intersection level and segment level in Montreal, Canada (Strauss et al., 2015). In this study, signalized intersections had a higher concentration of injuries and higher risk than non-signalized intersections.

Sun and Mobasheri (2017) used Strava Metro data to assess different levels of air pollution exposure by bicycling purpose (commuting versus non-commuting). Bicycling to work generated more exposure to air pollution than riding for leisure.

Evaluation Studies

Another form of active data application came from evaluating a new bicycle facility's usage in Australia (Heesch et al., 2016). While most of the previous studies were based on a micro-level database (e.g., trip routes and volume in an entire network at the city or county level), this study used macro-scale bicycle flow data to evaluate the performance of a segment of bikeway. Researchers compared bicycle volumes before and after the provision of the bikeway. This study shows the potential of Strava data to be used for longitudinal analysis as well.

In addition to the evaluation of new infrastructure, crowdsourced tracking data can be used for verification of model simulation. Leao and Pettit (2016) simulated bicycle commuting between suburban areas and the city center in Sydney, assuming that bicyclists choose the shortest paths. Researchers verified the simulated shortest paths using a real path taken by RiderLog users. The simulation showed 69 percent accuracy.

Other Applications

Selala and Musakwa (2016) mapped a year of bicycle ridership flow purchased from Strava Metro. Bicycling patterns were depicted area-wide for Johannesburg, South Africa, for the first time. This study shows the potential of using Strava Metro data in areas without adequate bicycle monitoring data.

Another innovative approach to other forms of crowdsourced data was tried in a study that used public profiles from Endomondo (Qiao, n.d.). The researcher downloaded almost 8 million workout records from the web server and used them to find factors associated with exercise duration.

Santos et al. (2016) used similar methods to collect data from GPSies. In their study, GPSies data open to the public were used to measure the intensity of using biking and running tracks. The researchers measured and visualized potential conflicts between cyclists and runners in a park in Portugal.

In terms of physical activity apps based on wearable devices (e.g., wristbands), data application focuses on health studies by recruiting study participants. Zenko et al. (2017) has conducted a study testing several methods to examine exercise (walking) intensity using Fitbit device records (e.g., heat rate). The study is ongoing, and researchers will recruit participants by early 2018. Wearable-device-based physical activity data are largely used for health studies or validating data for monitoring health activities (Tully et al., 2014; Symer et al., 2017).

By conducting bicycle OD analysis, Ma et al. (2015) examined how bike-share programs positively influenced transit ridership. Higher CaBi ridership was associated with greater transit usage. The results show an example of how policy makers or transit agencies can use bike-share data to integrate bicycle facilities into public transportation systems to increase transit market share.

Jestico et al. (2017) identified environmental factors that affect bicycle crashes at intersections between multiuse trails and roads using Bikemap.org data. Their results show that while a higher percentage of bicycle collisions were reported at trail-road intersections, the injury severity was lower than at road-road intersections. Researchers also pointed out that, for bicycle safety studies, crowdsourced crash data might be a better resource than police, insurance, or hospital records.

Potential Capabilities

Emerging data generated from active contributors have positively influenced non-motorized transportation planning in recent years:

- Almost ubiquitous smartphone use and the penetration of GPS tracking apps on the market have expanded the scope of monitoring to the entire network.
- Smartphone-based fitness app companies began to provide a large volume of bicycle data for commercial purposes.
- Regional bicycling apps that collect personal profiles, trip purposes, and bicyclists' comfort levels have made it possible to conduct more comprehensive research.

Before the availability of crowdsourcing, three data collection techniques were typically used for collecting route data for individual trips: web-based surveys for route preference, route-recall surveys relying on respondents' memory, and standalone GPS devices (Sener et al., 2009; Howard and Burns, 2001; Menghini et al., 2009). Compared to these methods, active crowdsourced monitoring requires less time and fewer resources, while also providing more samples. Based on the literature, the use of crowdsourced data in the non-motorized planning field has increased and will continue to do so.

One of the promising applications of emerging active data is route-choice modeling. In particular, for route-choice decision estimation, regional bicycling tracking apps (e.g., CycleTracks) are suitable because they contain rich information on a single track chosen by individual bicyclists, as well as demographic specificity. For this reason, such apps have fewer limitations in diverse applications than any other monitoring methods. For cities that do not have a customized app-based bicycling tracking program, Strava Metro data can be an option. Because this commercial data aggregator traces physical activity as long as there is a Strava user in the region of interest, the coverage area is extensive, but it is not without limitations (e.g., sample bias), as discussed in the following section.

Active crowdsourced data collection also brings more opportunities for demand models. Increased data coverage in terms of time and space increases the potential to estimate a robust travel demand model for a specific type of bicycle facility. In addition to demand models, traffic-volume-based studies become feasible at the point, segment, and regional scale (e.g., estimating crash exposure and air pollution exposure at intersections, in the streets, and on the entire network). Since the importance of public engagement in the planning process is growing, active monitoring methods can provide inclusion of stakeholders' opinions for transportation planners and practitioners (Watkins et al., 2016). Communities can report hot spots with higher crash risk and communicate their needs for new bicycle lanes or bicycle parking facilities.

Limitations of Emerging Crowdsourced Data

Despite their potential capabilities, there are some concerns about the emerging data (both passive and active).

Passive Data Limitations

Passive crowdsourced data are not collected primarily for pedestrians and bicyclists. To make use of these data, it is necessary to extract walking and bicycling trips from messy and muddled raw datasets. Processing emerging passive data requires sophisticated and advanced computational work. Primary limitations of passive data include the areas of mode detection, data precision, contextual information, and sampling bias.

Mode Detection

One of the challenges of passive crowdsourced data mining is mode detection because it should be performed with no aid of traveler input. A number of trials have attempted to infer transportation modes from raw datasets using phone-based GPS and MPP data. In the literature, transportation modes are generally split into motorized modes (e.g., cars, buses, and transit) and non-motorized modes (e.g., walking, running, bicycling, and remaining stationary) (Nikolic and Bierlaire, 2017). Differences in mode characteristics are used to infer mode type. For example, walking and bicycling have a lower speed than motorized modes.

Previous mode detection approaches showed acceptable accuracies.[3] Zheng et al. (2008) inferred driving, taking a bus, walking, and bicycling from GPS logs based on supervised learning. These researchers collected dedicated GPS traces from 65 people over 10 months and obtained different accuracies for walking and bicycling at 89.1 percent and 66.6 percent, respectively. Stenneth et al. (2011) used smartphone GPS data from six individuals to infer various transportation modes (i.e., train, car, bus, walking, bicycling, and stationary). Walking detection accuracy was 96.8 percent, and bicycling detection accuracy was 88.9 percent. When a GPS signal receiver was combined with other sensors, transportation modes were better detected (Reddy et al., 2008). Researchers developed a mode classification system for five modes (i.e., walking, stationary, bicycling, running, and motorized) and tested it using smartphone-based GPS and accelerometer sensor trace data collected from six individuals. The accuracy of walking detection was 96.8 percent, and that of bicycling was 92.8 percent. Jahangiri and Rakha (2015) classified five transportation modes (i.e., driving a car, riding a bicycle, riding a bus, walking, and running) from data obtained by smartphone sensors including GPS, accelerometer, gyroscope, and rotation vector. Researchers employed seven supervised learning methods and evaluated the performances of the different approaches. Walking and bicycling showed high accuracies ranging from 87.4 percent to 95.9 percent and from 85.8 percent to 96.9 percent, respectively. However, these trials were aided by subjects' input such as manual labeling that mostly requires initial training. In the supervised learning mechanism, GPS traces include quality information that helps transportation mode detection, which is almost impossible for data passively collected (i.e., without interaction with travelers).

[3] Accuracy denotes the ratio of the number of correctly detected modes to the number of detected modes (Nikolic and Bierlaire, 2017).

Lin et al. (2013) used an unsupervised scheme for detecting modes of transportation, eliminating the need for tedious manual labeling and pre-training. Detecting precision was satisfactory; walking was 76.0 percent, and biking was 52.4 percent. Other studies have tried to classify different transportation modes from raw MPP signal data based on an unsupervised method. Since MPP data are coarser grained than the relatively high locational precision of GPS, it is more challenging to detect the mode. However, Anderson and Muller (2006) and Sohn et al. (2006) differentiated walking, driving, and remaining stationary, and both research teams obtained a relatively high accuracy for walking with 87 percent and 70.2 percent, respectively. In the study that jointly used MPP and Wi-Fi signals, accuracy for detecting walking was high at 90.2 percent (Mun et al., 2008).

Overall, non-motorized modes (walking and bicycling) can be successfully detected from motorized modes due to low speed (Nikolic and Bierlaire, 2017). Although supervised approaches showed greater performance in terms of detection accuracy than unsupervised methods, the latter's results are satisfactory to some extent. The acceptable performance of unsupervised learning procedures implies that it would be possible to extract pedestrian and bicycle trips from passive data even if there is no traveler input. However, all the studies tested used a limited number of samples ranging from one user to 65 users who were recruited. Given the massive volume of emerging passive data (often called *big data*) with limited individual traveler information, a question about generality and wider acceptance of the proposed methods is raised. Similarities between slow walking and the stationary mode and between bicycling and slow cars are obstacles to accurate mode detection. Also, there is no standard defining the success of mode detection. Proposed solutions and disciplines are specific to each study (Prelipcean et al., 2017).

Data Precision

In the context of locational precision, the limitations of MPP data should be considered before discussing their application for non-motorized travel monitoring, especially for walking. Given that a widely accepted walkable distance is 400 m (Cervero, 2001), which may change depending on personal propensity, trip purpose, weather, and other factors, it is difficult to capture walking trips in many cases. For example, for a locational precision lower than 400 m, short walking trips (e.g., 300 m walking) cannot be detected correctly.

The coarse granularity of the cellular positioning data is problematic in securing locational precision. Several studies used cellular signaling data provided by AirSage to identify human mobility patterns (Phithakkitnukoon et al., 2010; Wang et al., 2010; Calabrese, 2013). The studies all divided the study areas into cells of 500 m by 500 m to capture aggregated travel patterns rather than analyzing discrete trips due to localization errors (e.g., 350 m in the study by Wang et al.). Wang et al. (2010) proposed a method to infer transportation mode share (driving and public transit) from AirSage data based on the same origin and destination. The study included only trips longer than 3 km. The authors explained that due to the coarse-grained sampling frequency over time (e.g., data are collected when mobile phones are used), the AirSage datasets "cannot be used to infer transportation modes for very short trips." According to the recent literature review on transportation mode inference based on smartphones, MPP datasets "provided by mobile phone operators are not used for transportation mode detection" (Nikolic and Bierlaire, 2017). Despite difficulties in mode classification, StreetLight recently provided a preliminary dataset of pedestrians using LBS data, and the company is expected to expand the service soon.

GPS data have higher confidence in positional accuracy and frequent sampling, but weak signal strength indoors and in dense urban environments (tall buildings obstruct and occlude signals from satellites) can be limitations.

Contextual Information

Non-motorized traveler monitoring with emerging passive data is limited by the lack of contextual information. Information on travelers is not available in detail at the individual or household level. Due to privacy concerns, data can only be provided after being anonymized and aggregated. Although context specificity such as demographic characteristics is a determinant factor that affects travel behavior, it is not obtainable through passively crowdsourced data. StreetLight supplements its data by combining census data, but StreetLight data are still available at the aggregate level. Therefore, even if emerging passive data are fully used to monitor pedestrians and bicyclists, disaggregate-level monitoring, such as that performed by the traditional travel survey, will be still challenging.

Sampling Bias

Despite vast sample size, the sampling bias of emerging passive data still causes issues associated with representativeness. Passive data are less likely to capture non-mobile phone (app) users. Identifying differences between mobile phone users and non-mobile phone users is critical to eliminate the significant bias inherent in passive crowdsourced data.

Active Data Limitations

Similar to passive data, there are also some limitations about active crowdsourced data, especially regarding sampling bias, contextual information, and sample size.

The sampling bias concern of passive data also exists for active data. In most studies, the number of bicycle tracking app users is higher for males, young generations, and commuters (Charlton et al., 2011; Hood et al., 2011; Blanc and Figliozzi, 2016; Hochmair et al., 2017; Jestico et al., 2016; Zimmermann et al., 2017). When using app-based tracking data, the user may need to consider how to deal with a specific bicycle population that is less likely to use a smartphone (e.g., weight the data for women, older users, and recreational bicyclists).

The lack of contextual information is also a common problem for emerging passive and active data. When personal information is collected, user privacy issues or user protection problems may occur. Approaches to personal profile collection must be carefully considered. For example, although it provides valuable information, Strava data have privacy limitations in that single-route journey information is not accessible (e.g., age, gender, trip length, trip time, and trip purpose). Although Strava Metro offers summary findings (i.e., aggregated sociodemographic information), there is a limitation on analyzing important determinant factors that affect travel behavior at the individual level (Romanillos et al., 2016).

Unlike emerging passive data, emerging active data are more prone to suffer from sample size issues. For example, considering the possibility of multiple apps in a study area, it may be hard to monitor all app users through only one app. For cities where a small sample size causes an issue, integrating data collected from other types of apps or combining the crowdsourced data with other datasets can be a solution.

Conclusion

This study reviewed current emerging data crowdsourced through mobile devices for pedestrian and bicyclist monitoring. The review included an examination of both emerging passive and active data, recognizing their potential to improve pedestrian and bicyclist safety studies. While emerging passive data require no or minimal levels of direct interaction with pedestrians or bicyclists during data collection, active data require users' voluntary participation in data collection.

So far, emerging non-motorized travel monitoring tools are more concentrated on active data, especially for bicycling. While actively crowdsourced data are beneficial for bicycle monitoring, limited sources are currently available for pedestrian research and programs. Since the launch of CycleTracks in 2009, many public agencies have developed GPS-based bicycle tracking programs to better understand bicycle traffic patterns in their regions. As for commercial apps, a fitness tracking app company, Strava, sells app users' physical activity GPS traces. With the success of these fitness tracking apps, increased data availability in time, space, and volume has enabled many bicycle travel behavior studies including route-choice modeling, collision exposure estimation, and evaluation of new facility provision at the entire network level. These kinds of studies have typically been limited due to lack of data. For example, in the majority of bicycle route-choice studies where detailed and widespread data on bicycle trips were not available, link characteristics over the entire route needed to be aggregated, rather than identifying variations across individual links (LaMondia and Watkins, 2017). While the apps sponsored by agencies (e.g., CycleTracks, Cycle Atlanta, and Mon RésoVélo) have extensive data coverage, such as individual-level trip records and sociodemographic features, they are limited in terms of geographical boundaries where the app is operated (e.g., San Francisco, Atlanta and Montreal). Such a limitation might be overcome by commercial apps (e.g., Strava) because these apps are used more broadly (e.g., globally). However, these commercial app data also suffer from data limitations in terms of providing additional travel information at a disaggregate level. Given that each emerging collection method has its own strengths and drawbacks, jointly applying different sources of data would generate synergistic effects.

For emerging passive data, commercial vendors provide a vast volume of travel data, but their products focus on vehicle trips rather than non-motorized trips due to limited positional precision, the short trip distances of walking and bicycling, and subsequent uncertainties of mode detection. At least one company, StreetLight, recently started to provide a preliminary dataset of pedestrians around transit stations on a trial basis. Considering that passive crowdsourced data account for a significant proportion of the total population, these data have greater potential for non-motorized travel monitoring, compared to traditional monitoring methods.

Overall, despite the benefits of these emerging data, there are several concerns, such as sample bias, sample size, and privacy issues. As for the two sample-related concerns, combining passive crowdsourced data with other sources of data (e.g., census data, actual counts, and surveys) or combining different sources of emerging data (e.g., CycleTracks and Strava Metro) can be a strategy for non-motorized travel planners who might face sample bias or limited sample size issues. For this data fusion, weighting a certain type of data or specific age group may be needed. Data collectors and providers must consider privacy issues as long as personal information is required to fix sample bias, as well as the purpose of research and projects. Special attention needs to be paid to ways to take advantage of emerging data sources for pedestrian monitoring.

References

AirSage (2017a). Company Overview. http://www.airsage.com/Technology/What-we-do/.

AirSage (2017b). Let's Chat (a service on the website). August 23. http://www.airsage.com/.

Anderson, I., and Muller, H. (2006). *Practical Activity Recognition Using GSM Data.* Technical Report CSTR-06-016, Department of Computer Science, University of Bristol.

Barajas, V. L., Wang, Z., Kaiser, M., and Zhu, Z. (2017). Improving Estimates of Real-Time Traffic Speeds during Weather Events for Winter Maintenance Performance Measurement*.* InTrans Project Report 13–485. Institute for Transportation Iowa State University.

BIXI (2017). We Are BIXI Montreal. https://montreal.bixi.com/en/who-we-are.

Blanc, B., and Figliozzi, M. (2016). Modeling the Impacts of Facility Type, Trip Characteristics, and Trip Stressors on Cyclists' Comfort Levels Utilizing Crowdsourced Data. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2587, pp. 100–108.

Bowman, N. (n.d.). *Big Data for Transportation Models.* https://connect.ncdot.gov/projects/planning/TPB%20Model%20User%20Groups/NCMUG_2016-11-16_Present_W2_StreetLight_11-18_NealBowman.pdf.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., and Ratti, C. (2013). Understanding Individual Mobility Patterns from Urban Sensing Data: A Mobile Phone Trace Example. *Transportation Research Part C: Emerging Technologies*, No. 26, pp. 301–313.

Capital Bikeshare (2017). Homepage. https://www.capitalbikeshare.com/.

Cervero, R. (2001). Walk-and-Ride: Factors Influencing Pedestrian Access to Transit. *Journal of Public Transportation*, Vol. 7, No. 3, pp. 1–23.

Charlton, B., Hood, J., Sall, E., and Schwartz, M. A. (2011). Bicycle Route Choice Data Collection Using GPS-Enabled Smartphones. Transportation Research Board 90th Annual Meeting.

Chen, C., Ma, J., Susilo, Y., Liu, Y., and Wang, M. (2016). The Promises of Big Data and Small Data for Travel Behavior (aka Human Mobility) Analysis. *Transportation Research Part C: Emerging Technologies,* No. 68, pp. 285–299.

Çolak, S., Alexander, L. P., Alvim, B. G., Mehndiratta, S. R., and González, M. C. (2015). Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations, and Opportunities. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2526, pp. 126–135.

Cortés, R., Bonnaire, X., Marin, O., and Sens, P. (2014). *Sport Trackers and Big Data: Studying User Traces to Identify Opportunities and Challenges.* Doctoral dissertation, INRIA, Paris.

Cui, Z., Zhang, S., Henrickson, K. C., and Wang, Y. (2016). New Progress of DRIVE Net: An E-science Transportation Platform for Data Sharing, Visualization, Modeling, and Analysis. *Smart Cities Conference* (ISC2), IEEE International, September, pp. 1–2.

Day, C. M., Premachandra, H., and Bullock, D. M. (2011). Rate of Pedestrian Signal Phase Actuation as a Proxy Measurement of Pedestrian Demand. Transportation Research Board 90th Annual Meeting.

Endomondo (2017a). History. https://www.endomondo.com/history.

Endomondo (2017b). What We Do. https://www.endomondo.com/about.

Endomondo (2017c). Challenges Guide. https://support.endomondo.com/hc/en-us/articles/202228458-Challenges-Guide.

Faghih-Imani, A., and Eluru, N. (2016). Incorporating the Impact of Spatio-temporal Interactions on Bicycle Sharing System Demand: A Case Study of New York CitiBike System. *Journal of Transport Geography*, Vol. 54, pp. 218–227.

Faghih-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., and Haq, U. (2014). How Land-Use and Urban Form Impact Bicycle Flows: Evidence from the Bicycle-Sharing System (BIXI) in Montreal. *Journal of Transport Geography*, No. 41, pp. 306–314.

Fitbit (2017). Our Technology. https://www.fitbit.com/technology.

Griffin, G. P., and Jiao, J. (2015a). Where Does Bicycling for Health Happen? Analysing Volunteered Geographic Information through Place and Plexus. *Journal of Transport and Health,* Vol. 2, No. 2, pp. 238–247.

Griffin, G. P., and Jiao, J. (2015b). Crowdsourcing Bicycle Volumes: Exploring the Role of Volunteered Geographic Information and Established Monitoring Methods. *URISA Journal*, Vol. 27, No. 1, pp. 57–66.

Griffin, G., Nordback, K., Götschi, T., Stolz, E., and Kothuri, S. (2014). *Monitoring Bicyclist and Pedestrian Travel and Behavior: Current Research and Practice.* Transportation Research E-Circular, E-C183.

Haworth, J. (2016). Investigating the Potential of Activity Tracking App Data to Estimate Cycle Flows in Urban Areas. *International Archives of the Photogrammetry, Remote Sensing 16 and Spatial Information Sciences*, XLI-B2, pp. 515–519.

Heesch, K. C., James, B., Washington, T. L., Zuniga, K., and Burke, M. (2016). Evaluation of the Veloway 1: A Natural Experiment of New Bicycle Infrastructure in Brisbane, Australia. *Journal of Transport and Health*, Vol. 3, No. 3, pp. 366–376.

Hochmair, H. H., Bardin, E., and Ahmouda, A. (2017). Estimating Bicycle Trip Volume for Miami-Dade County from Strava Tracking Data. Transportation Research Board 97th Annual Meeting.

Hood, J., Sall, E., and Charlton, B. (2011). A GPS-Based Bicycle Route Choice Model for San Francisco, California. *Transportation Letters*, Vol. 3, No. 1, pp. 63–75.

Howard, C., and Burns, E. (2001). Cycling to Work in Phoenix: Route Choice, Travel Behavior, and Commuter Characteristics. *Transportation Research Record: Journal of the Transportation Research Board,* No. 1773, pp. 39–46.

Hudson, J. G., Duthie, J. C., Rathod, Y. K., Larsen, K. A., and Meyer, J. L. (2012). *Using Smartphones to Collect Bicycle Travel Data in Texas.* No. UTCM 11-35-69.

Huntsinger, L. F., and Donnelly, R. (2014). Reconciliation of Regional Travel Model and Passive Device Tracking Data. Transportation Research Board 93rd Annual Meeting

Jackson, S., Miranda-Moreno, L. F., Rothfels, C., and Roy, Y. (2014). Adaptation and Implementation of a System for Collecting and Analyzing Cyclist Route Data Using Smartphones. Transportation Research Board 93rd Annual Meeting

Jahangiri, A., and Rakha, H. A. (2015). Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 16, No. 5, pp. 2406–2417.

Jestico, B., Nelson, T., and Winters, M. (2016). Mapping Ridership Using Crowdsourced Cycling Data. *Journal of Transport Geography*, Vol. 52, pp. 90–97.

Jestico, B., Nelson, T. A., Potter, J., and Winters, M. (2017). Multiuse Trail Intersection Safety Analysis: A Crowdsourced Data Perspective. *Accident Analysis and Prevention*, No. 103, pp. 65–71.

Kaufman, S. M., Gordon-Koven, L., Levenson, N., and Moss, M. L. (2015). Citi Bike: The First Two Years. *The Rudin Center for Transportation Policy and Management*.

Kerner, B. S., Rehborn, H., Schäfer, R. P., Klenov, S. L., Palmer, J., Lorkowski, S., and Witte, N. (2013). Traffic Dynamics in Empirical Probe Vehicle Data Studied with Three-Phase Theory: Spatiotemporal Reconstruction of Traffic Phases and Generation of Jam Warning Messages. *Physical A: Statistical Mechanics and Its Applications*, Vol. 392, No. 1, pp. 221–251.

Kim, S., and Coifman, B. (2014). Comparing INRIX Speed Data against Concurrent Loop Detector Stations over Several Months. *Transportation Research Part C: Emerging Technologies*, Vol. 49, pp. 59–72.

Knowledge Based Systems, Inc. (2017). Decentralized, Public, and Mobile-Based Sidewalk Inventory Tool. https://www.sbir.gov/sbirsearch/detail/1168625.

LaMondia, J. J., and Watkins, K. (2017). Using Crowdsourcing to Prioritize Bicycle Route Network Improvements. Project 2013-083, Southeastern Transportation Research, Innovation, Development and Education Center. https://trid.trb.org/view.aspx?id=1473831.

Le Dantec, C. A., Asad, M., Misra, A., and Watkins, K. E. (2015). Planning with Crowdsourced Data: Rhetoric and Representation in Transportation Planning. *18th ACM Conference on Computer Supported Cooperative Work and Social Computing,* pp. 1717–1727.

Leao, S. Z., and Pettit, C. (2016). Mapping Bicycling Patterns with an Agent-Based Model, Census and Crowdsourced Data. *International Workshop on Agent Based Modelling of Urban Systems,* May, pp. 112–128.

Lee, R. J., Sener, I. N., and Mullins III, J. A. (2016). An Evaluation of Emerging Data Collection Technologies for Travel Demand Modeling: From Research to Practice. *Transportation Letters*, Vol. 8, No. 4, pp. 181–193.

Lin, M., Hsu, W. J., and Lee, Z. Q. (2013). Detecting Modes of Transport from Unlabeled Positioning Sensor Data. *Journal of Location Based Services*, Vol. 7, No. 4, pp. 272–290.

Liu, H., Danczyk, A., Brewer, R., and Starr, R. (2008). Evaluation of Cell Phone Traffic Data in Minnesota. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2086, pp. 1–7.

Ma, B. T., Liu, C., and Erdoğan, S. (2015). Does Capital Bikeshare Affect Metrorail Ridership in Washington, D.C. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2534, pp. 1–9.

McCahill, C. (n.d.). *Improving Last-Mile Connections to Transit: An Exploration of Data and Analysis Tools.*

Menghini, G., Carrasco, N., Schussler, N., and Axhausen, K. (2009). Route Choice of Cyclists: Discrete Choice Modelling Based on GPS Data. *Arbeitsberichte Verkehrs-und Raumplanung,* pp. 544.

Meyer, J. L. (2013). *Big Bike Data: How GPS Route Data Collected from Smartphones Can Benefit Bicycle Planning.* University of Texas at Austin.

Morency, C., Trépanier, M., Paez, A., Verreault, H., and Faucher, J. (2017). *Modelling Bike Sharing Usage in Montreal over 6 Years.* No. CIRRELT-2017-33, Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation.

Mun, M., Estrin, D., Burke, J., and Hansen, M. (2008). Parsimonious Mobility Classification Using GSM and WiFi Traces. *5th Workshop on Embedded Networked Sensors (HotEmNets),* June.

Nelson, T. A., Denouden, T., Jestico, B., Laberee, K., and Winters, M. (2015). BikeMaps.org: A Global Tool for Collision and Near Miss Mapping. *Frontiers in Public Health,* Vol. 3, No. 53. pp. 18.

Nie, Y. M., Li, Q., Ghamami, M., and Ma, J. (2013). *Urban Travel Reliability Analysis with Consumer GPS Data.* SHRP2 Reliability IDEA Project L15D Report. Transportation Research Board.

Nikolic, M., and Bierlaire, M. (2017). Review of Transportation Mode Detection Approaches Based on Smartphone Data. *17th Swiss Transport Research Conference,* No. EPFL-CONF-229181.

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., and Ratti, C. (2010). *Activity-Aware Map: Identifying Human Daily Activity Pattern Using Mobile Phone Data.* 1st International Workshop, Human Behavior Understanding, August. Lecture Notes in Computer Science, Vol. 6219. Springer, Berlin, Heidelberg

Prelipcean, A. C., Gidófalvi, G., and Susilo, Y. O. (2017). Transportation Mode Detection—An In-Depth Review of Applicability and Reliability. *Transport Reviews*, Vol. 37, No. 4, pp. 442–464.

Proulx, F., and Pozdnukhov, A. (2017). *Bicycle Traffic Volume Estimation Using Geographically Weighted Data Fusion.* Manuscript submitted for publication.

Qiao, F. (n.d.). *Least Square Analysis on Exercise Duration of Recorded Data from the Endomondo Fitness App.*

Reddy, S., Burke, J., Estrin, D., Hansen, M., and Srivastava, M. (2008). Determining Transportation Mode on Mobile Phones. *Wearable Computers*, ISWC 2008, 12th IEEE International Symposium, September, pp. 25–28.

Romanillos, G., Zaltz Austwick, M., Ettema, D., and De Kruijf, J. (2016). Big Data and Cycling. *Transport Reviews*, Vol. 36, No. 1, pp. 114–133.

Ryus, P., Ferguson, E., Laustsen, K. M., Schneider, R.J., Proulx, F. R., Hull, T., Miranda-Moreno, L. (2014). *Guidebook on Pedestrian and Bicycle Volume Data Collection.* NCHRP Report 797, National Cooperative Highway Research Program.

Santos, T., Mendes, R. N., and Vasco, A. (2016). Recreational Activities in Urban Parks: Spatial Interactions among Users. *Journal of Outdoor Recreation and Tourism*, No. 15, pp. 1–9.

Schewel, L. (2016). Our New Data Partnership with Cuebiq. StreetLight, Sept. 20. http://blog.streetlightdata.com/our-new-data-partnership-with-cuebiq.

Schneider, R., Arnold, L., and Ragland, D. (2009). Methodology for Counting Pedestrians at Intersections: Use of Automated Counters to Extrapolate Weekly Volumes from Short Manual Counts. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2140, pp. 1–12.

Selala, M. K., and Musakwa, W. (2016). The Potential of Strava Data to Contribute in Non-Motorized Transport (NMT) Planning in Johannesburg. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLI-B2, pp. 587–594.

Sener, I. N., Eluru, N., and Bhat, C. R. (2009). An Analysis of Bicycle Route Choice Preferences in Texas. *U.S. Transportation*, Vol. 36, No. 5, pp. 511–539.

Smith, A. (2015). *Crowdsourcing Pedestrian and Cyclist Activity Data.* Pedestrian and Bicycle Information Center White Paper Series. http://www.pedbikeinfo.org/cms/downloads/PBIC WhitePaper_Crowdsourcing.pdf.

Sohn, T., Varshavsky, A., LaMarca, A., Chen, M., Choudhury, T., Smith, I., ... and De Lara, E. (2006). Mobility Detection Using Everyday GSM Traces. *UbiComp 2006: Ubiquitous Computing,* pp. 212–224.

Stenneth, L., Wolfson, O., Yu, P. S., and Xu, B. (2011). Transportation Mode Detection Using Mobile Phones and GIS Information. *19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, November, pp. 54–63.

Strauss, J., Miranda-Moreno, L. F., and Morency, P. (2015). Mapping Cyclist Activity and Injury Risk in a Network Combining Smartphone GPS Data and Bicycle Counts. *Accident Analysis and Prevention,* Vol. 83, pp. 132–142.

Strava Metro (2017a). Frequently Asked Questions. http://metro.strava.com/faq/.

Strava Metro (2017b). *Strava Metro Comprehensive User Guide*. http://metro.strava.com/wp-content/uploads/2017/04/StravaMetro_5.01_User_Guide.pdf.

StreetLight (2017). Message to the author, e-mail, June 20.

Sun, Y., and Mobasheri, A. (2017). Utilizing Crowdsourced Data for Studies of Cycling and Air Pollution Exposure: A Case Study Using Strava Data. *International Journal of Environmental Research and Public Health,* Vol. 14, No. 3, p. 274.

Symer, M. M., Abelson, J. S., Milsom, J., McClure, B., and Yeo, H. L. (2017). A Mobile Health Application to Track Patients after Gastrointestinal Surgery: Results from a Pilot Study. *Journal of Gastrointestinal Surgery,* Vol. 21, No. 7, pp. 1–6.

Tully, M. A., McBride, C., Heron, L., and Hunter, R. F. (2014). The Validation of Fitbit Zip™ Physical Activity Monitor as a Measure of Free-Living Physical Activity. *BMC Research Notes*, Vol. 7. No. 1, pp. 952.

Turner, S., Sener, I.N., Martin, M., Das, S., Shipp, E., Hampshire, R., Fitzpatrick, K., Molnar, L., Wijesundera, R., Colety, M., and Robinson. S. (2017). *Synthesis of Methods for Estimating Pedestrian and Bicyclist Exposure to Risk at Areawide Levels and on Specific Transportation Facilities.* FHWA Report SA-17-041, U.S. Department of Transportation.

Wang, H., Calabrese, F., Di Lorenzo, G., and Ratti, C. (2010). Transportation Mode Inference from Anonymized and Aggregated Mobile Phone Call Detail Records. *Intelligent Transportation Systems (ITSC), 13th International IEEE Conference,* September, pp. 318–323.

Watkins, K., Ammanamanchi, R., LaMondia, J., and Le Dantec, C. A. (2016). Comparison of Smartphone-Based Cyclist GPS Data Sources. Transportation Research Board 95th Annual Meeting.

Zenko, Z., O'Brien, J. D., Berman, C. J., and Ariely, D. (2017). Comparison of Affect-Regulated, Self-Regulated, and Heart-Rate Regulated Exercise Prescriptions: Protocol for a Randomized Controlled Trial. *Psychology of Sport and Exercise*, Vol. 32, pp. 124–130.

Zheng, Y., Li, Q., Chen, Y., Xie, X., and Ma, W. Y. (2008). Understanding Mobility Based on GPS Data. *10th International Conference on Ubiquitous Computing,* ACM, September, pp. 312–321.

Zimmermann, M., Mai, T., and Frejinger, E. (2017). Bike Route Choice Modeling Using GPS Data without Choice Sets of Paths. *Transportation Research Part C: Emerging Technologies*, No. 75, pp. 183–196.