Chapter 28
User-Generated Content: A Promising Data Source for Urban Informatics

Song Gao, Yu Liu, Yuhao Kang, and Fan Zhang

Abstract This chapter summarizes different types of user-generated content (UGC) in urban informatics and then gives a systematic review of their data sources, methodologies, and applications. Case studies in three genres are interpreted to demonstrate the effectiveness of UGC. First, we use geotagged social media data, a type of single-sourced UGC, to extract citizen demographics, mobility patterns, and place semantics associated with various urban functional regions. Second, we bridge UGC and professional-generated content (PGC), in order to take advantage of both sides. The third application links multi-sourced UGC to uncover urban spatial structures and human dynamics. We suggest that UGC data contain rich information in diverse aspects. In addition, analysis of sentiment from geotagged texts and photos, along with the state-of-the-art artificial intelligence methods, is discussed to help understand the linkage between human emotions and surrounding environments. Drawing on the analyses, we summarize a number of future research areas that call for attention in urban informatics.

28.1 Introduction

The urbanization process is accelerating in world cities and attracting large-scale job opportunities, human flows, business, and social activities. With the rapid development of information and communication technologies (ICT), location-aware devices, and sensor networks, the emergence of multi-source geospatial big data brings

S. Gao (✉) · Y. Kang
Geospatial Data Science Lab, University of Wisconsin, Madison, USA
e-mail: song.gao@wisc.edu

Y. Liu
Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing, China

F. Zhang
Senseable City Laboratory, Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA, USA

W. Shi et al. (eds.), *Urban Informatics*, The Urban Book Series,
https://doi.org/10.1007/978-981-15-8983-6_28

new opportunities to understand the rich semantics of space and place and associated human activities in urban areas using large-scale user-generated content (UGC) and crowdsourcing data streams, such as geotagged social media posts, travel blogs, mobile phone data, smart card data from transportation, GPS-enabled ridesharing services, and so forth. In this chapter, we review state-of-the-art research in UGC-based urban informatics using crowdsourced geographic information.

28.1.1 Background and Definition

Following the development of Web technologies and mobile devices, people can easily produce large numbers of data and rich information irrespective of their expertise. This is known as user-generated content (UGC), which is a form of content created by users of a system or a service and made available publicly on that system. UGC ranges from social media data and crowdsourced GPS trajectory data, to smart card data and mobile location data from a variety of apps. UGC maximizes the opportunity to understand multiple facets of the cities that we inhabit. The uniqueness and potential of UGC are mainly demonstrated in two ways. On the one hand, UGC can be viewed as the complement of professional-generated content (PGC), as it is decentralized and can be collected from the bottom up and through citizen science (Goodchild 2007; See et al. 2016). Therefore, it can be utilized to capture public opinions and further be leveraged to understand place-based contexts and sociocultural perceptions. On the other hand, UGC can be produced in an economical yet effective manner, and individuals as sensors largely expand the data coverage within cities.

Generally speaking, UGC in geographic information applications can be categorized in two types. One is collaborative mapping platforms, such as Wikimapia and OpenStreetMap (OSM), in which volunteers create and contribute geographic features and detailed descriptions to the Web, where the entries are synthesized into databases and made available to both public and private sectors. This type of UGC is also known as volunteered geographic information (VGI; Goodchild 2007) and has lowered the barriers for the general public to not only consume geographic information but also to contribute to the platform. Different organizations can also produce, customize, and render the data sources based on their own preferences of map styles and application requirements, such as in natural disaster management and emergency routing (Longueville et al. 2010; De Albuquerque et al. 2015; Han et al. 2019). VGI demonstrates how geographic data, information, and knowledge are produced and circulated in practice among different communities and in society at large (Sui et al. 2012). In the past decade, there exist a couple of studies comparing the data quality of VGI to the authoritative mapping sources and proprietary geodata in different regions and countries (Haklay 2010; Girres and Touya 2010; Zielstra and Zipf 2010; Neis et al. 2012; Forghani and Delavar 2014; Yamashita et al. 2019; Tian et al. 2019), where developed countries generally had a better coverage and data quality compared to developing countries. And in some regions, OSM data had geographically imbalanced coverage and were missing various types of information such as roads, points of interest (POI), and land uses (Dorn et al. 2015; Kashian et al. 2019). The second type of UGC is socially constructed data streams from users, that is, data entries constructed from mobile phone apps including diverse social media sources, crowdsourcing, and location-based services (Facebook, Twitter, Weibo, Foursquare, Yelp, Flickr, Instagram, Waze, Uber, Lyft, Didi, etc.), where the general public use locations, place names, and geographic contexts to search for information, consume the service, describe their sense of place, and share diverse opinions and comments according to their experiences (Li et al. 2013; Liu et al.2015; Gao et al. 2017; Janowicz et al. 2019). Harvey (2013) argues that this would be more precisely labeled as user contributed data, since people may not consciously volunteer their data, but generate it in the process of using the platforms for their particular purposes.

In cities, as the most populated areas on the Earth, there have been increasing amounts of UGC data streams generated every day from social media platforms, location-based services, crowdsourcing, and sensor networks, which help in sensing and addressing the urban problems and challenges in the regional economy and in globalization (Martinez-Fernandez et al. 2012; Cheshire and Hay 2017), and also drive the new paradigm in urban analytics (Batty 2019) that combine big data, urban planning and design, and spatial information theory for future development of sustainable cities.

28.2 Characteristics of UGC

User-generated data have their own pros and cons (Martí et al. 2019). In urban studies, although researchers have successfully utilized this emerging source for assessing urban spatial structure and functional regions (Gao et al. 2017; Tu et al. 2017; Xu et al. 2019), analyzing human mobility patterns and transportation infrastructure (Cho et al. 2011; Noulas et al. 2012; Hawelka et al. 2014; Liu et al. 2014; Yue et al. 2014) and supporting the design of new urban development rules, a good understanding of the key characteristics of UGC data is a prerequisite for preventing the abuse of such data. Compared to traditional data sources (e.g. survey) used in urban studies, UGC data have the following advantages.

First, UGC has the five Vs (volume, velocity, variety, veracity, and value) characteristic of big data (Marr 2015; Yang et al. 2017). Millions of users from different countries and regions in the world are posting all kinds of information per second (Hu et al. 2015; Liu et al. 2015; Martí et al. 2019). For instance, on Twitter, as one of the most widely used social media platforms, there are more than 500 million tweets sent daily by 100 million active users from 160 countries (Aslam 2019). UGC covers all kinds of topics including news, sports, entertainment, education, economics, technology, travels, and lifestyle and provides various perspectives in sensing urban environments and human dynamics (Sagl et al. 2012). People share comments about their lives, surrounding environments, and nearby events. As social media records include the timestamps of users' contents and activities automatically, they provide valuable information for time-series data analytics and time-geography applications (Chen et al. 2016; Tirunillai and Tellis 2012; Kang et al. 2017; Li et al. 2016). Moreover, the UGC data-collection process for a large geographic area is faster, and the cost is reduced compared to traditional surveys (Li et al. 2013; Gao et al. 2014; Jiang, Li, and Ye 2019). Moreover, the resolution of UGC can be zoomed into the detailed individual level (Yue et al. 2014; Liu et al. 2015) rather than the aggregation level such as census data; and the data update period of UGC (i.e. seconds, minutes, hours, or days) is usually shorter than that of official surveys (i.e. months or years).

Second, UGC data are contributed by the users voluntarily or are collected from the users who use a service and agree to share their data. It is worth noting that some references may only use a strict definition of actively generated data or crowdsourcing. Citizens monitoring their surrounding urban environment can be considered as sensors (Goodchild 2007) in terms of expressions, perceptions, and behaviors, while producing streams of data on social media Web sites, which can help reveal different aspects of their own lives and their environment (Arribas-Bel 2014). Conventional data collection methods for urban studies usually require large community surveys, long-period observations, and high labor costs using questionnaires and fieldwork (Nawrath, Kowarik, and Fischer 2019; Oliveira and Campolargo 2015). In contrast, UGC is produced through the motivation of both the organizations and the individuals, for various purposes such as providing and using location-based services (Yap et al. 2012), and the desire to share with others to promote friendships and social connections (Ames and Naaman 2007; Hollenstein and Purves 2010). Through this procedure, massive data can be collected unobtrusively in which the response bias in traditional methods may be eliminated (Quercia et al. 2015).

While UGC offers promising opportunities, several internal challenges and limitations of the UGC should be addressed for urban studies as follows.

First, although large volumes of content are contributed by millions of users every second, we may get a very sparse data matrix (e.g. Lee et al. 2015) after slicing the UGC data into a fine spatiotemporal resolution (e.g. a city-block spatial unit with hourly temporal window), which is crucial in solving some urban problems such as transportation planning and traffic congestion control. The spatiotemporal data sparsity issue becomes more prominent in the regions with limited numbers of active users. Due to the reduced data volume, the uncertainty in each slice may increase when analyzing the data (Bao et al. 2012).

Second, a common concern about UGC refers to the lack of standardization for users in the data generation process, which causes poor data quality and low trustworthiness, as well as high uncertainty (Senaratne et al. 2017). Users produce geographic data based on their local knowledge and their perception of the place, which may vary across different users (Stephens 2013). And due to the vagueness and uncertainty in human conceptualization of location, space, and place, it is hard for users to express some geographic regions and spatial relations precisely (Montello et al. 2003; Goodchild and Li 2012). Thus, an approach driven by data synthesis (Gao et al. 2017b), combining UGC with an approach informed by fuzzy-set theory (Wu et al. 2019), and combining UGC with survey-based behavior approaches (Twaroch et al. 2019) has been proposed to address the abovementioned concerns. For instance, users may have different perceptions and cognitions for the same place, which can cause incorrect tagging behaviors for social media photos (Hollenstein and Purves 2010).

The third issue concerns the representativeness of UGC, which refers to the degree to which UGC observation samples can represent the actual population (Zhang and Zhu 2018). The results may be biased by data sampling. The existing studies have figured out that the information shared on social media platforms usually follows a power-law distribution, indicating that only a small proportion of users contribute most of the content online (Kwak et al. 2010; Longley and Adnan 2016; Gao et al. 2017a). Therefore, the content collected might be dominated by some specific features and can be another source of bias. Besides, the demographic bias in contributors also impedes the representativeness (Hecht and Stephens 2014). Not all people in the real world use social media frequently. People who have limited access to social media, such as the elderly and users in developing countries, may be less sampled by UGC. For example, the average age of users in Twitter is 28 (Longley and Adnan 2016), and most photos in the Yahoo Flickr Creative Commons (YFCC) dataset released by the Yahoo Labs are uploaded by users in USA (Thomee et al. 2015; Kang et al. 2018) and several other developed countries. It is worth noting that the users who send geotagged tweets are also not randomly distributed over the population but create bias in subtle ways (Malik et al. 2015).

Despite the existence of data bias, research driven by UGC data has achieved great success as a result of validation or through comparison with studies using traditional data sources (Al-ghamdi and Al-Harigi 2015; Blaschke et al. 2018; Gao et al. 2017b; Liu et al. 2016). Opportunities have arisen for urban studies using UGC data because of the abovementioned advantages: (1) big data with low collection cost; (2) fast data generation and update velocity; (3) high penetration rate among users. The next part of this chapter summarizes various examples of UGC-driven urban informatics research and applications and with a focus on the topics of urban spatial structure, urban functional regions, place semantics, and user sentiment analysis. We will first introduce an analytical and computational framework to process large-scale crowdsourced data, and followed this with various applications and case studies in the literature.

28.3 Analytical and Computational Framework to Process UGC Data

A general analytical and computational framework to process and analyze UGC data is shown in Fig. 28.1. It consists of three parts from the bottom up. First, researchers collect various sources of UGC datasets including Twitter, Weibo, Instagram, Facebook, Foursquare, Yelp, and Dianping and store the data (including structured table records and unstructured texts, images, and videos) in the computer server or a cloud data center with master server and data nodes. Second, the raw data must be cleaned, filtered, processed, and enriched to further extract the information about users, locations, and content (more details in Sect. 28.3). Lastly, spatiotemporal analyses, statistical methods, and machine learning models are employed to support urban analytics,

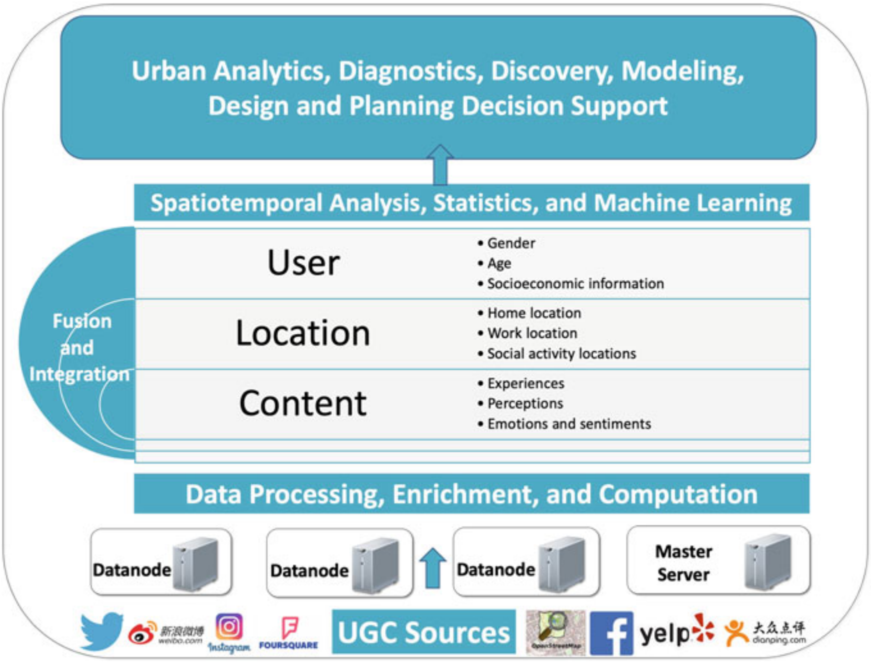

Urban Analytics, Diagnostics, Discovery, Modeling, Design and Planning Decision Support
Spatiotemporal Analysis, Statistics, and Machine Learning
Fusion and Integration
User
• Gender
• Age
• Socioeconomic information
Location
• Home location
• Work location
• Social activity locations
Content
• Experiences
• Perceptions
• Emotions and sentiments
Data Processing, Enrichment, and Computation
Datanode
Datanode
Datanode
Master Server
UGC Sources
新浪微博 weibo.com
Instagram
FOURSQUARE
OpenStreetMap
yelp
大众点评 dianping.com

Fig. 28.1 A general analytical and computational framework to process and analyze UGC data

diagnostics, knowledge discovery, modeling, prediction, and decision-making applications. During this process, multi-source UGC and crowdsourced data can be integrated and fused. High-performance computing infrastructure (Cao et al. 2015; Gao et al. 2017; Yang et al. 2017) and open-source analysis toolkits as well as machine learning frameworks such as *scikit-learn, r-spatial, PySAL,* and *Tensorflow* can be utilized to facilitate the data processing and advanced analysis.

28.4 Single-Source UGC-Based Urban Studies

28.4.1 User Information and Citizen Demographics

User information in UGC refers to the metadata or the profile of a user, including the place of residence, name, gender, age, ethnicity, hobby, friends, and social connections, and so on. Users are the main entities who generate content. There are two ways to collect user information from UGC. On the one hand, some basic user information can be directly obtained from the public profile which users provide on social media Web sites. When they were registering and creating a new account, users were required to enter such information by filling out online forms. For example, some basic demographic information such as nationality, gender, and age can be directly extracted from the user profiles (Longley et al. 2015; Kang et al. 2018). Researchers can further utilize such demographic information about citizens to better understand the flow of people from different geo-demographic groups in cities (Longley and Adnan 2016; Huang and Wong 2016). In addition, the follower and friendship connections in social media platforms can also be obtained and have been used to examine theories in the social sciences (Sloan and Morgan 2015; Ugander et al. 2011; Hodas et al. 2013).

On the other hand, some missing user information may not be retrieved directly from the user profile but can be inferred by combining other data sources and further analyses. For instance, the gender, age, and ethnicity information can be inferred from the user identifiers with the forename–surname pairs (Chang et al. 2010; Mateos et al. 2011; Mislove et al. 2011; Longley et al. 2015; Luo et al. 2016). By tracking the location and time of user postings, residents and visitors can be identified and distinguished (García-Palomares et al. 2015; Liu et al. 2018; Su et al. 2016).

28.4.2 *Human Mobility, Urban Spatial Structure, and Transportation*

Understanding human mobility patterns is important for the planning and management of urban land use and transportation. The work location, the home location, and even social activity locations of UGC users can be identified through their geotagged posts and their activity patterns detected in social media platforms (Gao et al. 2014; Li et al. 2014; Yang et al. 2015; Wu et al. 2015; Liu, Huang, and Gao 2019). The home-to-job commuting trips and non-commuting trips can be extracted and aggregated for traffic analysis zones (TAZs) to support urban transportation analysis. For example, as shown in Fig. 28.2, researchers detected over 24,000 daily commuting trips with an estimated average commuting time of about 32 min and average commuting distance

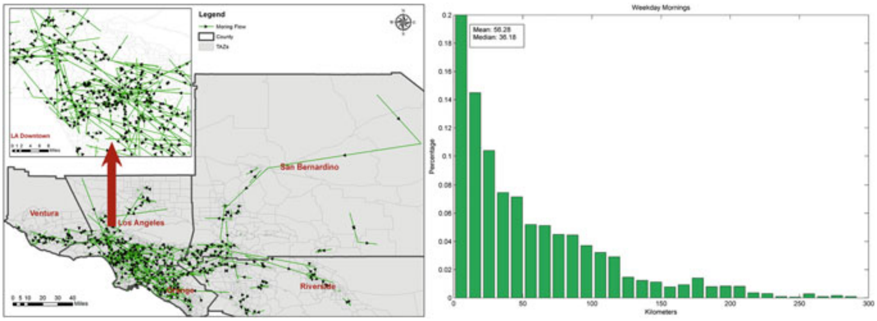


Fig. 28.2 Spatial and distance distributions of the detected commuting trips using geotagged Twitter data

of about 56 km in the Greater Los Angeles Area using millions of geotagged tweets (Gao et al. 2014). Moreover, when survey data and geotagged Twitter data were compared, the Pearson correlation coefficient of trips on weekdays was 0.91, and the correlation between detected trips using geotagged tweets and using a traditional travel demand model was 0.839 (Lee et al. 2015). While these correlations are far from perfect, the conclusions are nevertheless beneficial for urban transportation research.

Another benefit of using location-based check-in data from social networks is having access to information on place types (e.g. shops, offices, restaurants) for user activities, which is important to understand the spatial, temporal, and thematic distributions of human activities and activity-type transitions in cities (Noulas et al. 2011; Wu et al. 2014; McKenzie et al. 2015). For example, Wu et al. (2014) analyzed large-scale user check-in statistics in a location-based social-network platform in China and found different spatiotemporal activity transition probabilities among different types of places, including transportation facilities. Such activity-based transition patterns can also be extracted with pattern mining methods from call-detail-record data from mobile phones, allowing at-home, in-work, and social activity types to be annotated at each stay location (Cao et al. 2019). In addition, by combining information on user demographics, researchers found different movement patterns when comparing tourists and local residents (Chua et al. 2016; Liu et al. 2018), which could help transportation planning and management such as traffic congestion control and transportation regulations during events in cities. Moreover, the linkage between land use and urban dynamics can be identified through UGC and crowdsourcing data. For example, researchers found that human activities tended to decrease throughout the day for most land uses (e.g. offices, education, health) but remained constant in parks and increased in retail and residential zones (García-Palomares et al. 2018). Ren et al. (2019) examined the effect of land-use function complementarity on intra-urban spatial interactions using metro smart card records for different time periods and directions in the city of Shenzhen, China, which also demonstrates the trending use of individual-level big data in travel behavior studies in cities (Yue et al. 2014; Liu et al. 2015).

28.4.3 Place Semantics and Sentiments

Semantic signatures including the spatial, temporal, and thematic posed by McKenzie et al. (2015) and Janowicz et al. (2019) to extract and share high-dimensional data about types of places and neighborhoods. In contrast to spatial statistics, place-based analyses focus more on describing the topological and hierarchical relations between places and understanding various human perceptions and cognition at places (Li and Goodchild 2012; Gao et al. 2013; Zhu et al. 2016; Wu et al. 2019). Understanding the semantics of urban space and place could derive from the spatial, temporal, and thematic perspectives using geotagged texts, photos, and videos. These crowdsourced geographic data could also help the identification of vibrant neighborhoods

(Cranshaw et al. 2012; Zhang et al. 2013) and urban areas of interest (AOI), which refers to the regions within an urban environment that attract people's attention (Hu et al. 2015). Urban AOIs often have high exposure to the general public and receive a large number of visits. UGC such as geotagged photos can reveal the visit popularity and scenery information for city planners, transportation analysts, and location-based service providers to plan new businesses. Besides, the existing studies have utilized POI information and user check-ins in location-based social networking platforms (such as Foursquare, Yelp, Jiepang, and Weibo) to investigate various urban informatics issues. For example, a location-distortion model was proposed to improve reverse geocoding (i.e. convert a latitude/longitude to a POI address) using behavior-driven temporal signatures (McKenzie and Janowicz 2015). Another Place2Vec model derives the reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts and user check-in information (Yan et al. 2017). By combining the user check-in information in Foursquare with topic modeling approaches, researchers derived urban functional regions in the ten most populated US cities (Gao et al. 2017), which demonstrates a bottom-up data-driven perspective. In contrast, researchers also developed a top-down theory-informed approach to extracting urban functional regions. For example, a composition-pattern-based knowledge model was proposed to extract urban functional regions (Papadakis et al. 2019a). In this model, places are formalized as "patterns" which are defined as sets of components, composition rules, and functional implications. For example, a shopping plaza should consist of not only shopping stores but also restaurants, parking lots, and other facilities. Recently, an improved model was proposed using theoretical, empirical, and probabilistic patterns (Papadakis et al. 2019b) to enrich the knowledge-based model.

In addition, with advances in artificial intelligence (AI) technologies and open-source processing platforms as well as deep learning methods in the domains of natural language processing (NLP) and computer vision (CV), the extraction of human emotions (e.g. happiness, fear, anger, sadness, and surprise) and sentiments (i.e. positive, neutral, or negative) at different places and environments has become more accessible. For example, researchers applied advanced text mining techniques with spatial analysis to detect depressed Twitter users and their spatial clusters in US metropolitan areas. Socioeconomic variables from the Bureau of the Census and climate risk factors were found to have an impact on the prevalence of depression but may vary seasonally in different regions (Yang and Mu 2015; Yang et al. 2015). Human sentiment scores and their spatial distribution were extracted and explored in the city of Nanjing, China, using Weibo data (Zhen et al. 2018). High levels of air pollution were found to contribute to the urban population's reported low level of happiness in social media based on the analysis of over 210 million geotagged Weibo posts in China (Zheng et al. 2019). A semantic-specific sentiment analysis was conducted on Web-based neighborhood textual reviews in the city of New York for understanding the perceptions of citizens toward their living environments (Hu et al. 2019). As for image-based urban studies, researchers have used facial expression extraction techniques to explore human–environment interactions (as shown in Fig. 28.3) especially for the relationship between emotions and environments. A posi-

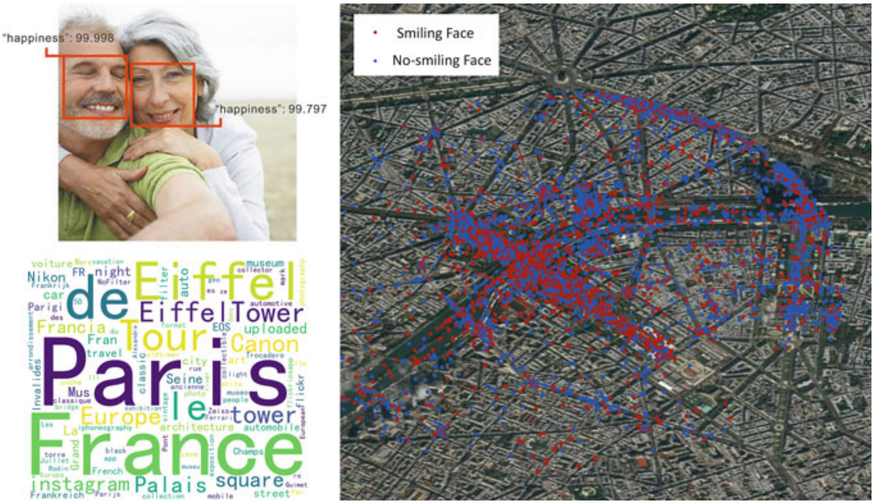


Fig. 28.3 Spatial distribution of smiling and no-smiling faces extracted from geotagged Flickr photos in Paris, France, and the associated word cloud of most frequent textual tags in these photos (Facial Expression subfigure was modified from the demo image of Face++ at https://www.faceplusplus.com/face-detection/)

tive correlation was found between the happiness score and the presence of natural environments such as water bodies and green vegetation in different types of place (Svoray et al. 2018; Kang et al. 2019). As another source of ambient sensing data, street view images can also be utilized to analyze human perceptions of places. For example, a data-driven machine learning approach with scene elements was proposed to measure how people perceive a place (including safe, lively, beautiful, wealthy, depressing, and boring) using street view images (Zhang et al. 2018a; Zhang et al. 2018b).

28.5 Multi-source Data-Driven Urban Studies

28.5.1 Fusion of Multiple UGC Sources

In traditional urban strategic planning or the classification results of remote sensing, many places in urban areas may be labeled as single land-use type; however, these areas may in reality contain multiple functions and land uses. In order to capture citywide dynamics of both human activities and urban functions at finer resolutions, multi-source UGC and crowdsourced information are combined to overcome their own limitations and to enrich the understanding of urban spatial structure and neighborhood demographics. Both mobile phone data and taxi trajectories usually cover large numbers of users and contain rich location information (and social network connections for mobile phone data) but lack place semantics (Liu et al. 2015). Social media data are sparsely distributed in space and time but contain rich content (Huang and Wong 2016; Martí et al. 2019). By combining both mobile phone data and social media, it is possible to extract citizen's home–job locations and social activity dynamics more effectively in space and time in cities (Tu et al. 2017). Also, by the integration of mobile-phone data and crowdsourced taxi trajectories, or the fusion of POI data and crowdsourced taxi trajectories, researchers have uncovered substantial differences between taxi trips and mobile-phone-based human movements in terms of spatial distribution and distance-decay effects (Kang et al. 2013) and explored the intensity of spatial interactions among different functional regions based on taxi origin–destination flows (Wang et al. 2018). In addition, researchers have used an online restaurant review platform with rich crowdsourced user-generated reviews and extracted machine learning features to further infer urban neighborhoods' population distribution and socioeconomic attributes in nine Chinese cities. They found a high predictability, in which the distributions of daytime and nighttime populations are estimated by mobile phone location data (Dong et al. 2019). UGC data can also be used to validate the urban spatial structure and place semantics extracted from ambient sensing and to reflect various urban environmental contexts. For example, as shown in Fig. 28.4, given only a certain number of street view images of a street, a deep learning model was trained to accurately estimate the hourly variation of human mobility patterns approximated by taxi trips along the streets (Zhang et al. 2019). In another study, researchers developed a mixed-use decomposition model based on temporal activity signatures extracted from social media check-in data, and taxi origin and destination (OD) trip data over one year were used to validate the land-use mixing results (Wu et al. 2019).

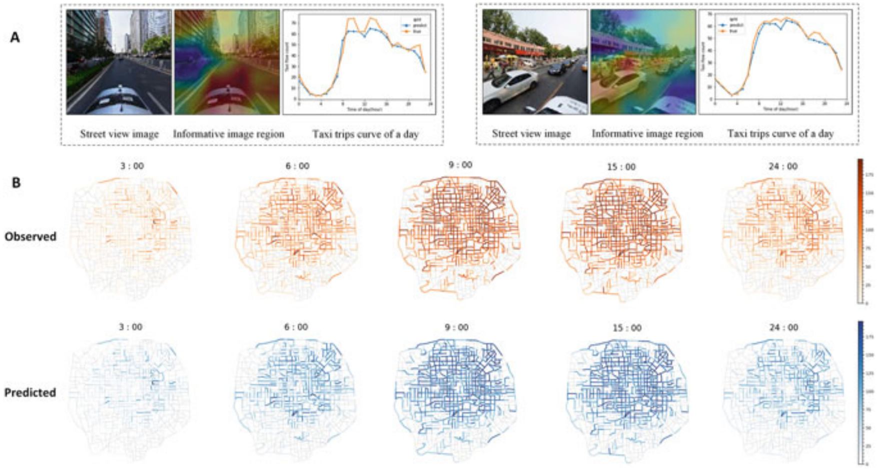


Fig. 28.4 **A** Predicting hourly variation of taxi trips using street view images; **B** Spatiotemporal variation of human mobility patterns approximated by taxi trips along the streets

28.5.2 Fusion of UGC and PGC

Compared to UGC, professional-generated content (PGC) mainly comes from domain experts and organizations who have the expertise and knowledge of study subjects, or the authority to collect and publish data, which is more trustworthy in social media platforms and in news media. The fusion of UGC and PGC can take advantages of both sides, to uncover urban spatial structures and dynamics, and to provide valuable information in the emergency management or disaster response scenarios. For example, crowdsourced geotagged photos and videos from social media users, volunteered geographic data, and authoritative storm surge data created by the U.S. Federal Emergency Management Agency (FEMA) were fused together to create a more accurate estimate of urban flood damage and updated road accessibility mapping in New York City during Hurricane Sandy (Schnebele et al. 2014). In urban planning and development, the integration of public participation from UGC big data sources together with the PGC-based expert design may provide a holistic approach through the process of idea generation, feedback, and evaluation for urban management and problem solving (Thakuriah et al. 2017).

In future, a number of multi-source data fusion research areas call for attention in urban informatics. First, the data sampling and fusing resolution requirements in space and time need to be investigated among different UGC sources to comprehensively understand human activities of different gender, age, and socioeconomic groups and place semantics for intra-urban and inter-city human mobility modeling. Second, combining UGC and PGC or combining data-driven and knowledge-driven approaches can solve urban problems such as traffic congestion and environmental pollution. Last but not least, there is a need to increase the engagement of citizen science in addressing urban changes in responsive cities through data-smart governance (Goldsmith and Crawford 2014).

28.6 Conclusion

UGC data contain rich information about human location, society, and human–environment interactions and have become a promising data source for urban informatics studies with unprecedented spatial, temporal, and thematic resolutions. This chapter summarized the key characteristics of UGC data with a focus on geographic information and urban studies. We discussed the analytical and computational framework to process UGC data and urban applications including citizen demographics, human mobility, urban spatial structure, place semantics, and sentiment analysis, to name a few. Considering the limitation of a single data source, various kinds of data fusion cases were discussed and suggested to advance future urban informatics studies. It is worth noting that we did not try to enumerate all possible fusion cases but just to list several scenarios with a focus on urban challenges. In sum, a combination of multi-source UGC-driven and theory-informed approaches provides a more holistic view for urban analytics, diagnostics, and human-centered sustainable urban planning and future development.

Acknowledgements Song Gao would like to thank the support of this research from the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation; Yu Liu would like to thank the funding support from the National Natural Science Foundation of China (No. 41625003).

References

Al-ghamdi SA, Al-Harigi F (2015) Rethinking image of the city in the information age. Procedia Comput Sci 65:734–743

Ames M, Naaman M (2007) Why we tag: motivations for annotation in mobile and online media. In: Proceedings of the SIGCHI conference on human factors in computing systems, 971–980

Arribas-Bel D (2014) Accidental, open and everywhere: emerging data sources for the understanding of cities. Appl Geogr 49:45–53

Aslam S (2019) Twitter by the numbers: stats, demographics and fun facts. Retrieved from https://www.omnicoreagency.com/twitter-statistics/ on May 2019

Bao J, Zheng Y, Mokbel MF (2012) Location-based and preference-aware recommendation using sparse geo-social networking data. In: Proceedings of the 20th international conference on advances in geographic information systems, 199–208

Batty M (2019) Urban analytics defined. Environ Plan B: Urban Anal City Sci 46(3):403–405

Blaschke T, Merschdorf H, Cabrera-Barona P, Gao S, Papadakis E, Kovacs-Györi A (2018) Place versus space: from points, lines and polygons in GIS to place-based representations reflecting language and culture. ISPRS Int J Geo-Inf 7(11):452

Cao J, Li Q, Tu W, Wang F (2019) Characterizing preferred motif choices and distance impacts. PLoS One 14(4):e0215242

Cao G, Wang S, Hwang M, Padmanabhan A, Zhang Z, Soltani K (2015) A scalable framework for spatiotemporal analysis of location-based social media data. Comput Environ Urban Syst 51:70–82

Chang J, Rosenn I, Backstrom L, Marlow C (2010) epluribus: ethnicity on social networks. In: Fourth international AAAI conference on weblogs and social media

Chen BY, Yuan H, Li Q, Shaw SL, Lam WH, Chen X (2016) Spatiotemporal data model for network time geographic analysis in the era of Big Data. Int J Geogr Inf Sci 30(6):1041–1071

Cheshire PC, Hay DG (2017) Urban problems in Western Europe: an economic analysis. Routledge

Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining, 1082–1090

Chua A, Servillo L, Marcheggiani E, Moere AV (2016) Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. Tourism Manag 57:295–310

Cranshaw J, Schwartz R, Hong J, Sadeh N (2012) The livehoods project: utilizing social media to understand the dynamics of a city. In: Sixth international AAAI conference on weblogs and social media

De Albuquerque JP, Herfort B, Brenning A, Zipf A (2015) A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. Int J Geogr Inf Sci 29(4):667–689

Dong L, Ratti C, Zheng S (2019) Predicting neighborhoods' socioeconomic attributes using restaurant data. In: Proceedings of the national academy of sciences 201903064

Dorn H, Törnros T, Zipf A (2015) Quality evaluation of VGI using authoritative data—a comparison with land use data in Southern Germany. ISPRS Int J Geo-Inf 4(3):1657–1671

Forghani M, Delavar M (2014) A quality study of the OpenStreetMap dataset for Tehran. ISPRS Int J Geo-Inf 3(2):750–763

Gao S, Janowicz K, McKenzie G, Li L (2013) Towards platial joins and buffers in place-based GIS. In: ACM SIGSPATIAL international workshop on computational models of place, pp 1–8

Gao S, Yang JA, Yan B, Hu Y, Janowicz K, McKenzie G (2014) Detecting origin-destination obility flows from geotagged Tweets in greater Los Angeles area. In: Short paper proceedings of the eighth international conference on geographic information science 1–4

Gao S, Janowicz K, Couclelis H (2017a) Extracting urban functional regions from points of interest and human activities on location-based social networks. Trans GIS 21(3):446–467

Gao S, Li L, Li W, Janowicz K, Zhang Y (2017b) Constructing gazetteers from volunteered big geo-data based on Hadoop. Comput Environ Urban Syst 61:172–186

Gao S, Janowicz K, Montello DR, Hu Y, Yang JA, McKenzie G, Ju Y, Adams B, Yan B (2017c) A data-synthesis-driven method for detecting and extracting vague cognitive regions. Int J Geogr Inf Sci 31(6):1245–1271

García-Palomares JC, Gutiérrez J, Mínguez C (2015) Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. Appl Geogr 63:408–417

García-Palomares JC, Salas-Olmedo MH, Moya-Gómez B, Condeco-Melhorado A, Gutierrez J (2018) City dynamics through twitter: relationships between land use and spatiotemporal demographics. Cities 72:310–319

Girres JF, Touya G (2010) Quality assessment of the French OpenStreetMap dataset. Trans GIS 14(4):435–459

Goodchild MF (2007) Citizens as sensors: the world of volunteered geography. GeoJournal 69(4):211–221

Goodchild MF, Li L (2012) Assuring the quality of volunteered geographic information. Spat Stat 1:110–120

Goldsmith S, Crawford S (2014) The responsive city: engaging communities through data-smart governance. Wiley

Haklay M (2010) How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets. Environ Plan B: Plan Des 37(4):682–703

Han SY, Tsou MH, Knaap E, Rey S, Cao G (2019) How do cities flow in an emergency? Tracing human mobility patterns during a natural disaster with big data and geospatial data science. Urban Sci 3(2):51

Harvey F (2013) To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information. In: Sui S, Elwood S, Goodchild MF (eds) Crowdsourcing geographic knowledge. Springer, Dordrecht, The Netherlands, pp 31–42

Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C (2014) Geo-located Twitter as proxy for global mobility patterns. Cartography Geogr Inf Sci 41(3):260–271

Hecht B, Stephens M (2014) A tale of cities: urban biases in volunteered geographic information. In: Eighth international AAAI conference on weblogs and social media

Hodas NO, Kooti F, Lerman K (2013) Friendship paradox redux: your friends are more interesting than you. In: Seventh international AAAI conference on weblogs and social media

Hollenstein L, Purves R (2010) Exploring place through user-generated content: using flickr tags to describe city cores. J Spat Inf Sci 1:21–48

Hu Y, Gao S, Janowicz K, Yu B, Li W, Prasad S (2015) Extracting and understanding urban areas of interest using geotagged photos. Comput Environ Urban Syst 54:240–254

Hu Y, Deng C, Zhou Z (2019) A semantic and sentiment analysis on online neighborhood reviews for understanding the perceptions of people toward their living environments. Ann Am Assoc Geogr 109(4):1052–1073

Huang Q, Wong DW (2016) Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? Int J Geogr Inf Sci 30:1873–1898

Janowicz K, McKenzie G, Hu Y, Zhu R, Gao S (2019) Using semantic signatures for social sensing in urban environments. In: Mobility patterns, big data and transport analytics. Elsevier, 31–54

Jiang Y, Li Z, Ye X (2019) Understanding demographic and socioeconomic biases of geotagged twitter users at the county level. Cartography Geogr Inf Sci 46(3):228–242

Kang C, Sobolevsky S, Liu Y, Ratti C (2013) Exploring human movements in Singapore: a comparative analysis based on mobile phone and taxicab usages. In: Proceedings of the 2nd ACM SIGKDD international workshop on urban computing, 1–8

Kang Y, Wang J, Wang Y, Angsuesser S, Fei T (2017) Mapping the sensitivity of the public emotion to the movement of stock market value: a case study of Manhattan. Int Arch Photogrammetry Remote Sens Spat Inf Sci 42:1–8

Kang Y, Zeng X, Zhang Z, Wang Y, Fei T (2018) Who are happier? spatio-temporal analysis of worldwide human emotion based on geo-crowdsourcing faces. In: 2018 Ubiquitous positioning, indoor navigation and location-based services (UPINLBS), 1–8

Kang Y, Jia Q, Gao S, Zeng X, Wang Y, Angsuesser S, Fei T et al (2019) Extracting human emotions at different places based on facial expressions and spatial clustering analysis. Transactions in GIS 23(3)

Kashian A, Rajabifard A, Richter KF et al (2019) Automatic analysis of positional plausibility for points of interest in OpenStreetMap using coexistence patterns. Int J Geogr Inf Sci 33(7):1420–1443. https://doi.org/10.1080/13658816.2019.1584803

Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on world wide web, 591–600

Lee JH, Gao S, Goulias K (2015) Can Twitter data be used to validate travel demand models? In: IATBR 2015-WIND

Li L, Goodchild MF (2012) Constructing places from spatial footprints. In: Proceedings of the 1st ACM SIGSPATIAL international workshop on crowdsourced and volunteered geographic information, 15–21

Li L, Goodchild MF, Xu B (2013) Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography Geogr Inf Sci 40(2):61–77

Li G, Hu J, Feng J, Tan KL (2014) Effective location identification from microblogs. In: IEEE 30th international conference on data engineering, 880–891

Li J, Ye Q, Deng X, Liu Y, Liu Y (2016) Spatial-temporal analysis on Spring festival travel rush in China based on multisource big data. Sustainability 8(11):1184

Liu Y, Sui Z, Kang C, Gao Y (2014) Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. PLoS One 9(1):e86026

Liu Y, Liu X, Gao S, Gong L, Kang C, Zhi Y, Chi G, Shi L (2015) Social sensing: a new approach to understanding our socioeconomic environments. Ann Assoc Am Geogr 105(3):512–530

Liu L, Zhou B, Zhao J, Ryan BD (2016) C-IMAGE: city cognitive mapping through geo-tagged photos. GeoJournal 81(6):817–861

Liu Q, Wang Z, Ye X (2018) Comparing mobility patterns between residents and visitors using geo-tagged social media data. Trans GIS 22(6):1372–1389

Liu X, Huang Q, Gao S (2019) Exploring the uncertainty of activity zone detection using digital footprints with multi-scaled DBSCAN. Int J Geogr Inf Sci 33(6):1196–1223

Longley PA, Adnan M (2016) Geo-temporal Twitter demographics. Int J Geogr Inf Sci 30(2):369–389

Longley PA, Adnan M, Lansley G (2015) The geotemporal demographics of Twitter usage. Environ Plan A 47(2):465–484

Longueville BD, Luraschi G, Smits P, Peedell S, Groeve TD (2010) Citizens as sensors for natural hazards: a VGI integration workflow. Geomatica 64(1):41–59

Luo F, Cao G, Mulligan K, Li X (2016) Explore spatiotemporal and demographic characteristics of human mobility via Twitter: a case study of Chicago. Appl Geogr 70:11–25

Malik MM, Lamba H, Nakos C, Pfeffer J (2015) Population bias in geotagged tweets. In: Ninth international AAAI conference on web and social media

Marr B (2015) Big data: using SMART big data, analytics and metrics to make better decisions and improve performance. Wiley

Martí P, Serrano-Estrada L, Nolasco-Cirugeda A (2019) Social media data: challenges, opportunities and limitations in urban studies. Comput Environ Urban Syst 74:161–174

Martinez-Fernandez C, Audirac I, Fol S, Cunningham-Sabot E (2012) Shrinking cities: urban challenges of globalization. Int J Urban Reg Res 36(2):213–225

Mateos P, Longley PA, O'Sullivan D (2011) Ethnicity and population structure in personal naming networks. PLoS One 6(9):e22943

McKenzie G, Janowicz K (2015) Where is also about time: a location-distortion model to improve reverse geocoding using behavior-driven temporal semantic signatures. Comput Environ Urban Syst 54:1–13

McKenzie G, Janowicz K, Gao S, Yang JA, Hu Y (2015) POI pulse: a multi-granular, semantic signature-based information observatory for the interactive visualization of big geosocial data. Cartographica: The Int J Geogr Inf Geovisualization 50(2):71–85

Mislove A, Lehmann S, Ahn Y-Y, Onnela J-P, Rosenquist JN (2011) Understanding the demographics of twitter users. In: Fifth international AAAI conference on weblogs and social media

Montello DR, Goodchild MF, Gottsegen J, Fohl P (2003) Where's downtown? Behavioral methods for determining referents of vague spatial queries. Spat Cogn Comput 3(2–3):185–204

Nawrath M, Kowarik I, Fischer LK (2019) The influence of green streets on cycling behavior in European cities. Landscape Urban Plan 190:103958

Neis P, Zielstra D, Zipf A (2012) The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. Future Internet 4(1):1–21

Noulas A, Scellato S, Mascolo C, Pontil M (2011) An empirical study of geographic user activity patterns in foursquare. In: Fifth international AAAI conference on weblogs and social media

Noulas A, Scellato S, Lambiotte R, Pontil M, Mascolo C (2012) A tale of many cities: universal patterns in human urban mobility. PLoS ONE 7(5):e37027

Oliveira Á, Campolargo M (2015) From smart cities to human smart cities. In: 48th Hawaii international conference on system sciences, 2336–2344

Papadakis E, Resch B, Blaschke T (2019) Composition of place: towards a compositional view of functional space. Cartography Geogr Inf Sci 1–18

Papadakis E, Baryannis G, Petutschnig A, Blaschke T (2019b) Function-based search of place using theoretical, empirical and probabilistic patterns. ISPRS Int J Geo-Inf 8(2):92

Quercia D, Schifanella R, Aiello LM, McLean K (2015) Smelly maps: the digital life of urban smellscapes. AAAI Publications 327–336

Ren M, Lin Y, Jin M, Duan Z, Gong Y, Liu Y (2019) Examining the effect of land-use function complementarity on intra-urban spatial interactions using metro smart card records. Transportation 1–23

Sagl G, Resch B, Hawelka B, Beinat E (2012, July) From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments. In: Proceedings of the GI-Forum. Herbert Wichmann Verlag, Berlin 54–63

Schnebele E, Cervone G, Waters N (2014) Road assessment after flood events using non-authoritative data. Nat Hazards Earth Syst Sci 14(4):1007–1015

See L, Mooney P, Foody G, Bastin L, Comber A, Estima J, Liu HY et al (2016) Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. ISPRS Int J Geo-Inf 5(5):55

Senaratne H, Mobasheri A, Ali AL, Capineri C, Haklay M (2017) A review of volunteered geographic information quality assessment methods. Int J Geogr Inf Sci 31:139–167

Sloan L, Morgan J (2015) Who tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on twitter. PLoS One 10(11):e0142209

Stephens M (2013) Gender and the GeoWeb: divisions in the production of user-generated cartographic information. GeoJournal 78(6):981–996

Su S, Wan C, Hu Y, Cai Z (2016) Characterizing geographical preferences of international tourists and the local influential factors in China using geo-tagged photos on social media. Appl Geogr 73:26–37

Sui D, Elwood S, Goodchild M (eds) (2012) Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice. Springer Science and Business Media

Svoray T, Dorman M, Shahar G, Kloog I (2018) Demonstrating the effect of exposure to nature on happy facial expressions via Flickr data: advantages of non-intrusive social network data analyses and geoinformatics methodologies. J Environ Psychol 58:93–100

Thakuriah PV, Tilahun NY, Zellner M (2017) Big data and urban informatics: innovations and challenges to urban planning and knowledge discovery. In: Seeing cities through big data. Springer, Cham 11–45

Thomee B, Shamma DA, Friedland G, Elizalde B, Ni K, Poland D, Li LJ et al (2015) YFCC100M: The new data in multimedia research. ArXiv Preprint ArXiv:1503.01817

Tian Y, Zhou Q, Fu X (2019) An analysis of the evolution, completeness and spatial patterns of OpenStreetMap building data in China. ISPRS Int J Geo-Inf 8(1):35

Tirunillai S, Tellis GJ (2012) Does chatter really matter? Dynamics of user-generated content and stock performance. Mark Sci 31(2):198–215

Tu W, Cao J, Yue Y, Shaw SL, Zhou M, Wang Z, Li Q et al (2017) Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. Int J Geogr Inf Sci 31(12):2331–2358

Twaroch FA, Brindley P, Clough PD, Jones CB, Pasley RC, Mansbridge S (2019) Investigating behavioural and computational approaches for defining imprecise regions. Spat Cogn Comput 19(2):146–171

Ugander J, Karrer B, Backstrom L, Marlow C (2011) The anatomy of the facebook social graph. ArXiv Preprint ArXiv:1111.4503

Wang Y, Gu Y, Dou M, Qiao M (2018) Using spatial semantics and interactions to identify urban functional regions. ISPRS Int J Geo-Inf 7(4):130

Wu L, Zhi Y, Sui Z, Liu Y (2014) Intra-urban human mobility and activity transition: evidence from social media check-in data. PLoS One 9(5):e97010

Wu F, Li Z, Lee WC, Wang H, Huang Z (2015) Semantic annotation of mobility data using social media. In: Proceedings of the 24th international conference on world wide web 1253–1263

Wu L, Cheng X, Kang C, Zhu D, Huang Z, Liu Y (2019) A framework for mixed-use decomposition based on temporal activity signatures extracted from big geo-data. Int J Dig Earth 1–19. https://doi.org/10.1080/17538947.2018.1556353

Wu X, Wang J, Shi L, Gao Y, Liu Y (2019b) A fuzzy formal concept analysis-based approach to uncovering spatial hierarchies among vague places extracted from user-generated data. Int J Geogr Inf Sci 33(5):991–1016

Xu Y, Chen D, Zhang X, Tu W, Chen Y, Shen Y, Ratti C (2019) Unravel the landscape and pulses of cycling activities from a dockless bike-sharing system. Comput Environ Urban Syst 75:184–203

Yamashita J, Seto T, Nishimura Y, Iwasaki N (2019) VGI contributors' awareness of geographic information quality and its effect on data quality: a case study from Japan. Int J Cartography 1–11

Yan B, Janowicz K, Mai G, Gao S (2017) From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. In: Proceedings of the 25th ACM SIGSPATIAL international conference on advances in geographic information systems 1–10

Yang W, Mu L (2015) GIS analysis of depression among Twitter users. Appl Geogr 60:217–223

Yang F, Jin PJ, Cheng Y, Zhang J, Ran B (2015a) Origin-destination estimation for non-commuting trips using location-based social networking data. Int J Sustain Transp 9(8):551–564

Yang W, Mu L, Shen Y (2015b) Effect of climate and seasonality on depressed mood among twitter users. Appl Geogr 63:184–191

Yang C, Huang Q, Li Z, Liu K, Hu F (2017) Big data and cloud computing: innovation opportunities and challenges. Int J Dig Earth 10(1):13–53

Yap LF, Bessho M, Koshizuka N, Sakamura K (2012) User-generated content for location-based services: a review. In: Lazakidou AA (ed) Virtual communities, social networks and collaboration 163–179

Yue Y, Lan T, Yeh AG, Li QQ (2014) Zooming into individuals to understand the collective: a review of trajectory-based travel behaviour studies. Travel Behav Soc 1(2):69–78

Zhang G, Zhu A-X (2018) The representativeness and spatial bias of volunteered geographic information: a review. Ann GIS 24(3):151–162

Zhang AX, Noulas A, Scellatos S, Mascolo C (2013) Hoodsquare: modeling and recommending neighborhoods in location-based social networks. In: 2013 international conference on social computing 69–74

Zhang F, Zhang D, Liu L, Lin H (2018a) Representing place locales using scene elements. Comput Environ Urban Syst 71:153–164

Zhang F, Zhou B, Liu L, Liu Y, Fung HH, Lin H, Ratti C (2018b) Measuring human perceptions of a large-scale urban region using machine learning. Landscape Urban Plan 180:148–160

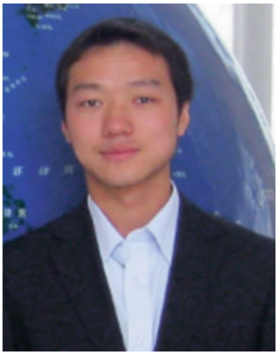
Zhang F, Wu L, Zhu D, Liu Y (2019) Social sensing from street-level imagery: a case study in learning spatio-temporal urban mobility patterns. ISPRS J Photogrammetry Remote Sens 153:48–58

Zhen F, Tang J, Chen Y (2018) Spatial distribution characteristics of residents' emotions based on Sina Weibo big data: A case study of Nanjing. In: Shen Z, Li M (eds) Big data support of urban planning and management: the experience in China. Springer, Cham, Switzerland, pp 43–62

Zheng S, Wang J, Sun C, Zhang X, Kahn ME (2019) Air pollution lowers Chinese urbanites' expressed happiness on social media. Nat Hum Behav 3:237–243

Zhu R, Hu Y, Janowicz K, McKenzie G (2016) Spatial signatures for geographic feature types: examining gazetteer ontologies using spatial statistics. Trans GIS 20(3):333–355

Zielstra D, Zipf A (2010) A comparative study of proprietary geodata and volunteered geographic information for Germany. In: 13th AGILE international conference on geographic information science



Song Gao is an Assistant Professor in GIScience at the University of Wisconsin-Madison, where he leads the GeoDS Lab. His main research interests include place-based GIS, human mobility, and GeoAI. He is currently the associate editor of Annals of GIS.

Yu Liu is a Boya Professor of GIScience at the Institute of Remote Sensing and Geographic Information Systems, Peking University. His research interests mainly concentrate on the humanities and social sciences based on big geo-data. He is currently an associate editor of Computers, Environment and Urban Systems.

Yuhao Kang is a Ph.D. student at the GeoDS Lab, University of Wisconsin-Madison. He received his Bachelor's degree from Wuhan University. His research interests include place-based GIS, GeoAI and cartography.

Fan Zhang is a postdoctoral researcher at SENSEable City Lab, Massachusetts Institute of Technology. He received his Ph.D. from the Chinese University of Hong Kong. His research interests include place-based GIS, GeoAI and data-driven approaches for urban studies.