Mobile Data Science: Towards Understanding Data-Driven Intelligent Mobile Applications

Iqbal H Sarker

HAL Id: hal-02075370
https://hal.science/hal-02075370

Submitted on 29 Mar 2019

Mobile Data Science: Towards Understanding Data-Driven Intelligent Mobile Applications

Iqbal H. Sarker *

Department of Computer Science and Software Engineering,
Swinburne University of Technology,
Melbourne, VIC-3122, Australia.

Abstract

Due to the popularity of smart mobile phones and context-aware technology, various contextual data relevant to users' diverse activities with mobile phones is available around us. This enables the study on mobile phone data and context-awareness in computing, for the purpose of building data-driven intelligent mobile applications, not only on a single device but also in a distributed environment for the benefit of end users. Based on the availability of mobile phone data, and the usefulness of data-driven applications, in this paper, we discuss about *mobile data science* that involves in collecting the mobile phone data from various sources and building data-driven models using machine learning techniques, in order to make dynamic decisions intelligently in various day-to-day situations of the users. For this, we first discuss the *fundamental concepts* and the *potentiality* of mobile data science to build intelligent applications. We also highlight the key *elements* and explain various key *modules* involving in the process of mobile data science. This article is the first in the field to draw a big picture, and thinking about mobile data science, and it's potentiality in developing various data-driven intelligent mobile applications. We believe this study will help both the researchers and application developers for building smart data-driven mobile applications, to assist the end mobile phone users in their daily activities.

1. Introduction

The advancement of mobile computing and the Internet have played a significant role in the development of the current digital age. The Internet has now formed the backbone of modern communication. Nowadays, the use of the Internet, particularly the World Wide Web (WWW) has moved beyond desktop computers to millions of mobile phones. According to ITU (International Telecommunication Union), cellular network coverage has reached 96.8% of the world population, and this number even reaches 100% of

the population in the developed countries [44]. With the development of science and technology, the smart phone industry is growing rapidly and the popularity of smart phones has made exponential growth in mobile phone application market [28]. According to Google Trends [3], users' interest on *"Mobile Phones"* is more and more than other platforms like *"Desktop Computer"* or *"Tablet Computer"* over time. For instance, Figure 1 shows users' interests trends over time on mobile phones over desktop and tablet computer for the last five years (December 2012 to December 2017). The data was collected on 13th December 2017 from Google Trends [3].

In Figure 1, the numbers from 0 to 100 in Y-axis, represent search interests in terms of popularity relative to the highest point on the chart for a

---

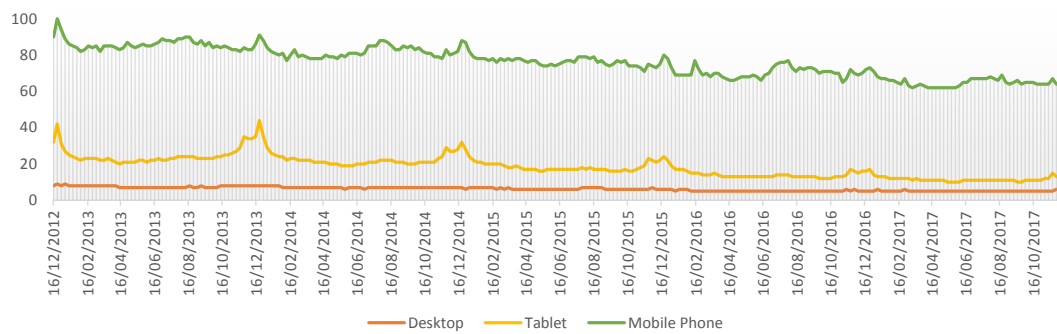*Corresponding author: Iqbal H. Sarker. Email: msarker@swin.edu.au

Figure 1. Users' interest trends over time

given date (e.g., 16/12/2012). For instance, a value of 100 represents the peak popularity for a particular term. A value of 50 means that the term is half as popular. Likewise, a score of 0 means the term was less than 1% as popular as the peak [3]. Based on the popularity of mobile phones over time shown in Figure 1, we are motivated to think about *mobile data science* and corresponding data-driven intelligent mobile applications to assist the end mobile phone users in their daily activities.

Nowadays, mobile phones have become one of the primary ways, in which people around the globe communicate with each other. These devices have transformed over a period of time from merely communication tools to smart and highly personal devices. Such devices are able to assist us in a variety of day-to-day situations in our daily activities. People in the world use mobile phones for various activities, such as for voice communication, Internet browsing, using mobile applications (apps), e-mailing, using online social network, instant messaging etc [27]. While mobile phones may come in various forms, in this paper they refer to smart mobile phones or Internet accessible mobile phones which have incorporated many advanced features to facilitate better information access and utilization for the benefit of end mobile phone users in their daily life. These smartphones in addition to being used as a communication device are capable of doing things that a computer does because of their computing capability on the devices. For instance, what actually makes the mobile phone "smart"? The answer is it's ability to handle data, not only the voice calls. In general, these devices can fulfill nearly all communication requirements of the users in recent times. Smart mobile phones are not only used as communication devices but also used in various sectors [24] like health care services, navigation services, business purposes, safety, environmental monitoring, intelligent transport systems, traffic management, intelligent route routing, smart homes, or smart cities etc.

In recent times, the smartphones are becoming more and more powerful in both computing and data storage aspects. Their more and more sensing capabilities have enabled the collection of rich raw contextual information of the user, surrounding real-life environment, and the mobile phones usage records through the device logs, such as phone call logs [29] [15], SMS Logs [13], mobile application (Apps) usages logs [48] [43], mobile phone notification logs [25], web logs [19], game Logs [26], context logs [9] [48], and smartphone life logs [30] etc. The logged data generated by the smart phones provides a means to get new knowledge about various aspects of the users, like user social interactions, various user activities with the devices, which offers the potential to understand the insight of such mobile phone data.

Typically, the traditional data science helps to solve problems by analyzing data and turns data into data products [10] [21]. However, the concept of mobile data science is not exactly the same thing. The purpose of mobile data science is slightly different, as the smart mobile phones are aware of their user's real-life surrounding environment, and users' various types of activities or social interactions, in various contexts in the real world. Thus, *mobile data science* is a term which is concerned with the challenge of collecting user's real life data in different contexts, such as temporal, spatial, social or others, in a pervasive computing environment, finding data-driven models in order to make dynamic decisions for individual mobile phone users based on relevant contextual information, and use this discovered knowledge for building smart mobile applications to intelligently assist themselves in their daily life. Some examples of such real-life applications are context-aware smart mobile communication, intelligent mobile notification management systems, context-aware smart searching, context-aware mobile recommender system, context-aware app management system, and various context-aware predictive services etc. Thus, the concept of "*mobile data science*" involving users' real-life activities with mobile phones, corresponding contextual

information of the users, and the surrounding dynamic environment, has become interested.

In particular, the contributions of this paper are:

- We explore the potentiality of mobile data science for building various data-driven intelligent mobile applications.

- We review various kinds of mobile phone data and define mobile data science that includes mobile phone data having contextual information and the processing to discover useful knowledge for data-driven applications.

- We discuss the fundamental concepts of mobile data science and highlight the key differences between the traditional data science and mobile data science in various aspects.

The rest of the paper is organized as follows. We briefly review the context-aware mobile computing, which is related to mobile data science, in Section 2. In Section 3, we briefly discuss about mobile data science including the fundamental concepts, and the key modules of mobile data science. Finally, Section 4 concludes this paper.

2. Context–Aware Mobile Computing: A Background

The term *context* can be used with a variety of different meanings in different purposes. In this section, first we briefly discuss about various definitions of contexts in the area of mobile and pervasive computing, and then we discuss about *context-awareness* in mobile computing.

2.1. Definitions of Contexts

The notion of context has been used in numerous areas, including Pervasive and Ubiquitous Computing, Human Computer Interaction, Computer-Supported Collaborative Work, and Ambient Intelligence [12]. In the area of Ubiquitous and Pervasive Computing, a number of early works on context-aware computing or context-awareness referred context as the locational contexts, i.e., the location of people and objects [41]. In recent works, context has been extended to include a broader collection of factors in addition to such locational context, such as physical and social aspects of an entity or object, as well as the activities of users [12]. Having examined the definitions and categories of context given by the pervasive and ubiquitous computing community, in this section, we review the definition of contexts within the scope of this paper.

Several studies have attempted to define and represent context from different perspectives. For instance, the user's location information, the surrounding people and objects around the user, and the changes to those

objects are considered as contexts by Schilit et al. [41]. Brown et al. [8] also define contexts as user's locational information, temporal information, the surrounding people around the user, temperature, etc. Similarly, the user's locational information, environmental information, temporal information, user's identity, are also taken into account as contexts By Ryan et al. [32].

Other definitions of context have simply provided synonyms for context such as context as the environment or social situation. A number of researchers are taken into account the context as the environmental information of the user. For instance, in [7], the environmental information that the user's computer knows about are taken into account as context by Brown et al, whereas the social situation of the user is considered as a context in Franklin et al. [18]. On the other hand, a number of other researchers consider it to be the environment related to the applications. For instance, Ward et al. [45] consider the state of the surrounding information of the applications as contexts. Hull et al. [23] define context as the aspects of the current situation of the user and include the entire environment. The settings of applications are also treated as context in Rodden et al. [31]. In [40], Schilit et al. claim that the important aspects of context are: (i) where you are, (ii) whom you are with, and (iii) what resources are nearby. The information of the changing environment is taken into account as context in their definition. In addition to the user environment (e.g., user location, nearby people around the user, and the current social situation of the user), they also include the computing environment and the physical environment. For instance, connectivity, available processors, user input and display, network capacity, and costs of computing can be the examples of the computing environment, while the noise level, temperature, the lighting level, can be the examples of the physical environment.

Dey et al. [11] presents a survey of alternative view of context, which are largely imprecise and indirect, typically defining context by synonym or example. Finally, Dey et al. [11] offer a definition of context, which is perhaps now the most widely accepted. According to Dey et al. [11] *"Context is any information that can be used to characterize the situation of an entity. An entity is person, place or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves".*

2.2. Context–Awareness

Typically, a desktop computer application expects a static execution environment, either in office, home or other locations. However, this precondition is generally not applicable for mobile services or systems, as the world around an application is changing frequently.

Table 1. Various Types of Contexts

Context Category	Context Examples
User Identity	Name, ID, gender, age, date of birth, nationality, educational level, available income etc.
Temporal	Date (e.g., YYYY-MM-DD), time (e.g., hh:mm:ss), periods (e.g., 1 hour), weekday (e.g., Monday), weekend (e.g., Saturday), season (e.g., summer), public holidays (e.g., Christmas day), etc.
Spatial	Location (e.g., office, home, work, market, etc.), orientation, speed, etc.
Environmental	Temperature, weather, humidity, light, sound, noise, traffic conditions, etc.
Social	Interpersonal relationship (e.g., friend, family, professionals etc.), situation (e.g., meeting, lecture, seminar, etc.), people nearby, activity change, calendar, etc.
Physiological	Blood pressure, heart rate, tone of voice etc.
Psychology	Emotions (e.g., happy, sad, angry etc.) preference, tiredness etc.
Mobile phone activity	Phone calls, using mobile applications, mobile web and navigation, gaming, using social network (e.g., Facebook, Linkedin) etc.
Physical activity	Sitting, standing, driving, walking, biking, running, etc.
Device-related	Illumination, carrier, Wi-Fi, Bluetooth, screen status, battery status, setting status, and device status etc.
App-related	Running apps, active app, app status etc.

Thus, the nature of mobile applications must change according to the movement of the users. In particular, mobile applications should become more flexible in order to respond in highly dynamic computing environments, which makes the computing more pervasive. Recent advances in smart mobile phones and the pervasive computing environment make it possible in the real-world, where *context-awareness* can be used as the spirit of pervasive computing [42]. The primary objective of pervasive computing environment is the creation of an environment saturated with seamlessly integrated devices with computing and communication capabilities having decision-making ability with as little direct user interaction as possible [5]. The use of contextual information in mobile applications is able to reduce the amount of human effort and attention that is needed for an application to provide the services according to user's needs or preferences, in a pervasive computing environment. Thus, context-awareness simply represents the dynamic nature of the applications.

Nowadays, the popularity of context-awareness in mobile computing is increasing because of their adaptation in dynamic environment. Typically, context-awareness is originated as a term from ubiquitous or pervasive computing, which is able to deal with linking changes in the real-world environment with mobile systems. According to Wikipedia [2], context-awareness enables a new class of applications in pervasive computing, which is a property of mobile devices that is defined complementarily to location awareness, i.e., adapting capability in the applications with the movement of mobile phone users. However, as discussed above, various types of contexts can be applied to make the mobile applications more flexible and useful according to users' overall situations in terms of temporal, spatial or social, and their preferences, as context-awareness is the main spirit of pervasive computing to build adaptive applications. In general, context-aware computing refers to a general class of mobile systems that can sense their surrounding physical environment, and able to adapt their behavior intelligently according to the sensed information in the corresponding applications [1].

Context-aware systems are a component of a ubiquitous computing or pervasive computing environment [1]. According to [1], there are three important aspects

of context; these are: where you are; who you are with; and what resources are nearby. In order to make the mobile applications capable of operating in highly dynamic environments demanding on less user attention, different types of contexts might have different impact, shown in Sarker et al. [16]. Hence, we summarize a number of contexts, related to mobile phone users, their activities, and surrounding environment, shown in Table 1.

3. Understanding Mobile Data Science

We are living in the age of data science [10]. On the other hand, computing is increasing being carried out using smart mobile phones, which support not only the telephony but also data-driven mobile applications. In this section, we discuss briefly about the mobile data science, ranging from smartphone raw data having contextual information, building data-driven models using machine learning techniques, and corresponding smart mobile applications for the end users.

3.1. Motivational Examples

The following real-life examples intuitively illustrate the advantage of *mobile data science* in various mobile applications either on single device or in a distributed environment, to assist the users in their daily life. In general, a single device based application is related to the computation on individuals' devices utilizing their phone log data. On the other hand, application in the distributed environment is designed to allow users of a computer network to access information, and services, as well as to exchange information with others, like a client-server based model.

EXAMPLE 1 (Application Scenario on Single Device). Say, Alice, a smartphone user, is a PhD student. She has installed a large number of mobile applications (apps) on her smartphone. Homescreens of smartphones provide ready access to commonly used apps, which is particularly useful. However, the homescreen of her smartphone is unaware about Alice's changing contexts, such as her location, and consequently unable to manage apps intelligently according to her needs. Thus, an intelligent mobile app management system is needed for her, which allows the particular app she currently needs to be easily accessible from the mobile homescreen. To build such intelligent applications *mobile data science* can play an important role by predicting her future usages utilizing her app usages log data, that intelligently assists her to use different types of mobile apps, such as Skype, Whatsapp, Facebook, Gmail, Youtube, Linkedin, Microsoft Outlook, etc. according to her current contextual information.

EXAMPLE 2 (Application Scenario in Distributed Environment). Let's consider a cloud-based mobile service recommendation system for the above user Alice. When she enters her University campus by driving her car, she will automatically get a recommendation for the 'best' car parking space to minimize the searching time and cost. In addition, the user will receive on her mobile device detailed driving instructions to reach that space. A distributed cloud-based system is able to accomplish this job automatically without any additional effort of the user. She can get this service using the corresponding client application installed on her mobile phone based on her contextual information and preferences. The cloud-based system is typically a client-server model where in the mobile device based agent (client) extracts the real-time contextual information of the user and communicates with a cloud-based service that houses the corresponding recommender system (server). Unlike the single device based services, where computation is done within the device, the battery life of the device, cpu computational limitations, and user preference modeling limitations, can be ignored in a cloud based services as the server is responsible for all the computations needed. To build such cloud based services *mobile data science* can play an important role to predict the parking availability according to her current contextual information by analyzing the relevant data in the cloud.

The above scenarios show that *mobile data science* is potential to build various data-driven intelligent mobile applications, to intelligently assist the users in various purposes by analyzing the relevant data, not only on a single device but also in a distributed environment.

3.2. Understanding Real–Life Smartphone Data

Real-life smartphone data having contextual information relevant to individual mobile phone users' activity is one of the key elements of mobile data science. The reason is that mobile phone data with contextual information is the basis of mobile data science to build data-driven intelligent applications for the users. According to [10], we live in the age of data, where everything that surrounds us is linked to a data source and everything in our lives is captured digitally. On the other hand, mobile or cellular phones have become increasingly ubiquitous and powerful. Recent advances in smart mobile phones and their sensing capabilities have enabled the collection of users various activities and corresponding contextual information. Thus, contextual data is available around us to analyze. In this section, we discuss about a variety of mobile phone data containing the associated contextual information of individual users. We have summarized these data in Table 2.

Table 2. Various Types of Smart Mobile Phone Data

Smartphone Data	Data Description (user activity and contexts information)	References
Call Log	User phone call activities, such as incoming, missed and outgoing calls, and corresponding contextual information such as date, time, call type, call duration, user identifier, location, etc.	Sarker et al. [36], Eagle et al. [13], Srinivasan et al. [43], Bell et al. [6]
SMS Log	Spam and non-spam text messages and related contextual information such as user identifier, date, time, etc.	Almeida et al. [4], Eagle et al. [13]
App Usages Log	User various app usages such as Gmail, Skype, Facebook, Linkedin, Read news, Whatsapp, SNS, Web, Multimedia, Game, etc. and corresponding contextual information such as date, time-of-the-day, battery level, profile type (General, Silent, Meeting, Outdoor, Pager, Offline), Charging state (Charging, Complete, Not Connected), location (Home, Work Place, On the way) etc.	Zhu et al. [48], Srinivasan et al. [43]
Notification Log	Users' response for various notifications, such as promotional emails, game invites on social network, predictive suggestions by applications, etc. and corresponding contextual information such as date, time-of-the-day, notification type, user's physical activity (includes still, walking, running, biking and in vehicle), location (Home, Work, or Others) etc.	Mehrotra et al. [25]
Web Log	User mobile web navigation, web searching, e-mailing, entertainment, chat, music, news, TV, netting, traveling, sport, banking, etc. and related contextual information such as date, time-of-the-day, weekdays, weekends, etc.	Halvey et al. [19], Halvey et al. [20]
Game Log	Users' various game playing such as action, adventure, casual, puzzle, RPG, strategy, sports, etc. and related contextual information such as date, time-of-the-day, weekdays, weekends, etc.	Paireekreng et al. [26]
Smartphone Life Log	User phone calls, SMS headers (no content), App usage (e.g., Skype, Whatsapp, Youtube etc.), physical activities form Google play API, and related contextual information such as WiFi and Bluetooth devices in user's proximity, geographical location, temporal information, etc.	Rawassizadeh et al. [30]

These context-rich historical mobile phone data shown in Table 2, is simply as the collection of the past contexts and user's actions for the past contexts [22]. The main characteristic of such kind of log data is that they contain the actual behavior of the users in different contexts, as the mobile phones automatically record these data. Thus, it is important to study mobile phone data involving contextual information and users diverse activities in different contexts, for the purpose of building data-driven smart and context-aware intelligent mobile applications.

3.3. Fundamental Concepts of Mobile Data Science

The traditional data science [10] can be treated as a part of mobile data science. Nowadays, many researchers use of the term "data science" to describe the interdisciplinary field of data collection, preprocessing, inferring or making decisions by analyzing the data. According to Cao et al. [10], a high-level statement about data science is: "it is the science of data or the study of data". The outputs of data science are typically data products depending on what type of data are taken into account, which can be a discovery,

Table 3. Some Key Terms in Mobile Data Science

Key Terms	Description
Context	Any information that can be used to characterize the situation of an entity (e.g., temporal, spatial or social contexts.)
Pervasive computing environment	An intelligent environment which is dynamic.
Context-awareness	The spirit of pervasive computing, which represents the dynamic nature of mobile applications.
Context-aware mobile computing	A mobile computing paradigm in which applications can discover and take advantage of contextual information.
Data science	The science of data; also known as data-driven science, is an interdisciplinary field about scientific methods, processes, and systems to extract knowledge or insights from data, similar to data mining.
Mobile user activity	Physical activity (e.g., walking, driving, etc.) and mobile phone activity (e.g., phone calling, apps using, emailing, Internet browsing, messaging, etc.)
Adaptability	Able to adapt dynamic environment and behave accordingly (e.g., context-awareness in mobile applications.)

Table 4. Traditional Data Science VS Mobile Data Science

Aspects	Data Science	Mobile Data Science
Rationale	Data tells the story	Mobile phone data with contextual information and pervasive intelligence disclose problem-solving solutions for building mobile applications.
Objective	Innovative and effective algorithms, Creating data products	Solving real-life problems of mobile phone users, To build adaptive, intelligent and context-aware mobile software systems, To assist mobile phone users in their daily activities.
Data	General purpose data	Mobile phone users's real-life data and surrounding environment information collected by mobile phones in a pervasive computing environment, Data includes contextual information and mobile phone users' various activities.
Mechanism	Automated	Human-centered, particularly mobile phone users.
Evaluation	Technical metrics	Trade-off between technical significance and mobile phone users' expectation according to their current contexts and preferences.

prediction, service, recommendation, thinking, model, or system. According to the definition of Cao et al. [10], the ultimate values of data products are knowledge, intelligence, wisdom, and decision based on relevant data.

On the other hand, computing is moving toward pervasive, ubiquitous environments [17]. Thus, the computing devices, software agents, and services are all expected to seamlessly integrate and cooperate in support of human objectives according to their needs and preferences. Mobile devices are able to provide such service in a pervasive computing environment in an anywhere, any-time fashion. Therefore, the next step for pervasive computing is the integration of smart devices that are able to understand the local context of the users and share this information in support of human needs intelligently.

Based on the potentiality of data science and the pervasiveness in computing using mobile devices discussed above, in this paper, we introduce the concept

of *mobile data science*, by taking advantages from both, in order to help the researchers and application developers, for the purpose of building data-driven mobile applications utilizing relevant mobile phone data. Our objective is to highlight the importance of working in this area, in order to identify examples of key research issues for the community, thereby acting as a catalyst for new research and development for the benefit of end mobile phone users. We define the field as follows:

"Mobile data science is research or working area that exists at the intersection of context-aware mobile computing and data science, characterized by a focus on the real-life contextual data collection, data pre-processing, analyzing data using data mining and machine learning techniques, and use of extracted knowledge from the mobile phone data, for the purpose of building smart data-driven mobile applications, in order to intelligently assist the end mobile phone users by providing dynamic decisions in their daily activities, in a context-aware pervasive computing environment.".
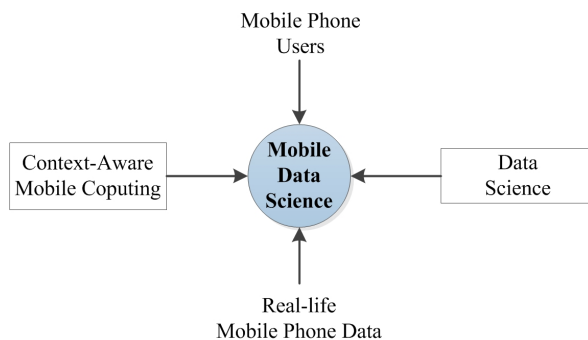

Figure 2. Key involvement in mobile data science

In Table 3, we define some key terms related to *mobile data science*. In a word, mobile data science is related to users' real-life mobile phone data, mobile user activities, context-aware mobile computing, and the traditional data science, shown in Figure 2. We also summarize various aspects of *mobile data science* over the traditional data science, highlighted in Table 4.

3.4. Key Modules of Mobile Data Science

In this section, we briefly discuss about the key *modules* that are involved in the process of mobile data science, for the purpose of building data-driven intelligent mobile applications.

In the process of mobile data science, first the contextual data is collected from the real world using the devices. After that, a data-driven model based on machine learning techniques according to the target problem, can be built by analyzing and processing the contextual data. Such data-driven models can be used to build various intelligent mobile applications to
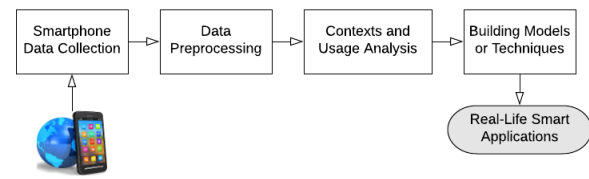

Figure 3. A high–level overview of mobile data science process involving the key modules.

intelligently assist the end mobile phone users in their daily activities. Based on this process, Figure 3 shows a high level overview of mobile data science involving a number of modules, from real-world raw data to applications. In the following, we discuss these modules one by one with examples.

- *Mobile Data Collection*: This is the first module of mobile data science, as mobile phone data is the basis of mobile data science. Thus, this module mainly focuses on collecting user's real life smartphone data relevant to a particular problem for analysis and making corresponding decisions. Real-world mobile phone data usually comprise features whose interpretation depends on some *contextual information* such as temporal (e.g., in the morning), spatial (e.g., at office), social contexts (e.g., in a meeting) or others. These contextual-sensitive features and their available patterns in the dataset are of high interest to be discovered and analyzed in order to make dynamic decisions intelligently in a pervasive computing environment. The contextual data can be collected from various sources like smartphone sensors, e.g., GPS, Wi-Fi, Blue-tooth, accelerometer, proximity, camera, microphone, light sensors, etc. or using corresponding phone log data, electronic calendars, according to the needs for a particular application. For instance, Sarker et al. [14] collect mobile phone call log data including time-series contextual information for their temporal analysis in order to build time-dependent call interruption management system based on users' unavailability [34].

- *Data Pre-processing*: This module mainly focuses on preparing the data for modeling. For instance, cleaning the data to address the data quality issues can be a pre-processing task. Real-world mobile phone data may contain noisy or inconsistence instances, missing values, duplicate records, invalid data or outliers. Such noisy instances may have impact on machine learning techniques to build an effective model for a particular problem. Therefore, handling these inconsistency to ensure the quality of data is

an important issue for data analysis. Various machine learning techniques can be used to detect such outliers or noisy instances. For instance, in [38], the authors propose a naive Bayes classifier based noise detection technique to ensure the quality of smart phone data. In addition to the quality of data, to transform the raw data to the formatted or desired data to make it suitable for further analysis is another task in data pre-processing. For instance, the continuous time-series mobile phone data is converted into behavior-oriented segments to capture the behavioral patterns of individual mobile phone users in [36]. Like the temporal context, the data-centric social context [35] or others relevant raw data may need to transform into the desired data that can be used for further analysis according to the needs.

- *Contexts and Usages Analysis*: This module mainly focuses on selecting the interesting features or contexts that have an influence on users' diverse activities to build an effective model. For instance, social contexts, e.g., in a meeting, or social relationship between individuals, e.g., boss, might have an influence to make a call handling decision of an individual mobile phone users [33]. Similarly, other contexts such as temporal, or spatial contexts, can play a role in different applications depending on users' need and preferences [37]. For instance, building a context-aware model for predicting parking availability in smart cities, temporal and spatial contexts can be used [47]. Such relevancy of contexts may differ from application-to-application in the real world.

- *Building Data-driven Models*: Once the required contextual information and mobile phone usages are prepared according to a particular problem, this module is responsible to build an intelligent data-driven model to solve the target problem. Thus, this module can be treated as the core module in mobile data science, as intelligent decision making is depended on the corresponding data-driven model. For the purpose of building such models, various popular data mining and machine learning techniques [21], such as classification analysis (e.g., Decision tree, Naive Bayes, Support vector machine, K-nearest neighbors, Artificial neural network etc.), ensemble learning (e.g., Random forest, AdaBoost, Bootstrap aggregating etc.), clustering (e.g., K-means, Agglomerative, Divisive etc.), regression analysis (e.g., Linear regression, Logistic regression etc.), and association analysis (e.g., AIS, Apriori, FP-growth etc.) or others relevant machine learning techniques [21] [46], can be used for an effective

modeling according to the specific needs for a particular problem. The extracted knowledge using such machine leaning techniques then can be used to build various real-life smart mobile applications to intelligently assist the end users. For instance, mining behavioral association rules of individual mobile phone users based on machine learning technique, can be used to effectively predict their future behavior according to their current contextual information [39].

The modules discussed above are the key in mobile data science for the purpose of building data-driven intelligent mobile applications, in order to assist the end mobile phone users in their daily activities in a context-aware pervasive computing environment.

4. Conclusion

In this paper, we have briefly discussed about the concept of mobile data science by highlighting the characteristics of mobile phone data including users' various activities in different contexts. In particular, we focused our investigation on context-awareness and real-life mobile phone data, and their proper utilization for building intelligent data-driven models not only on the single devices but also in a distributed environment, to assist them in their daily activities in a pervasive computing environment. Additionally, we have summarized the key differences between the traditional data science and mobile data science in various aspects. We believe this article will help both the researchers and application developers for building various data-driven intelligent mobile applications for the benefit of end mobile phone users utilizing their mobile phone data.

Acknowledgment

References

[1] Context-aware computing. *https://en.wikipedia.org/wiki/*.
[2] Context awareness. *https://en.wikipedia.org/wiki/*.
[3] Google trends. In *https://trends.google.com/trends/*, 2017.
[4] Tiago A Almeida, José María G Hidalgo, and Akebo Yamakami. Contributions to the study of sms spam filtering: new collection and results. In *Proceedings of the 11th ACM symposium on Document engineering*, pages 259–262. ACM, 2011.
[5] Christos Anagnostopoulos, Athanasios Tsounis, and Stathes Hadjiefthymiades. Context management in pervasive computing environments. In *Pervasive Services*,

2005. ICPS'05. Proceedings. International Conference on, pages 421–424. IEEE, 2005.

[6] Stephen Bell and Alisdair McDiarmid. Nodobo: Mobile phone as a software sensor for social network research. In *Vehicular Technology Conference*. IEEE, 2011.

[7] Peter J Brown. The stick-e document: a framework for creating context-aware applications. *Electronic Publishing-Chichester-*, 8:259–272, 1995.

[8] Peter J Brown, John D Bovey, and Xian Chen. Context-aware applications: from the laboratory to the marketplace. *IEEE personal communications*, 4(5):58–64, 1997.

[9] Huanhuan Cao, Tengfei Bao, Qiang Yang, Enhong Chen, and Jilei Tian. An effective approach for mining mobile user habits. In *Proceedings of the International Conference on Information and knowledge management, Toronto, ON, Canada, 26-30 October, pp.~1677–1680. ACM, New York, USA*, 2010.

[10] Longbing Cao. Data science: a comprehensive overview. *ACM Computing Surveys (CSUR)*, 50(3):43, 2017.

[11] Anind K Dey. Understanding and using context. *Personal and ubiquitous computing*, 5(1):4–7, 2001.

[12] Paul Dourish. What we talk about when we talk about context. *Personal and ubiquitous computing*, 8(1):19–30, 2004.

[13] Nathan Eagle and Alex Sandy Pentland. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4):255–268, 2006.

[14] Sarker et al. Behavior-oriented time segmentation for mining individualized rules of mobile phone users. In *Data Science and Advanced Analytics (IEEE DSAA), Canada.*, 2016.

[15] Sarker et al. Phone call log as a context source to modeling individual user behavior. In *UbiComp: Adjunct*. ACM, 2016.

[16] Sarker et al. An approach to modeling call response behavior on mobile phones based on multi-dimensional contexts. In *Mobile Software Engineering and Systems, Argentina*, 2017.

[17] Tim Finin, Anupam Joshi, Lalana Kagal, Olga Ratsimore, Vlad Korolev, and Harry Chen. Information agents for mobile and embedded devices. *Cooperative Information Agents V*, pages 264–286, 2001.

[18] David Franklin and Joshua Flaschbart. All gadget and no representation makes jack a dull environment. In *Proceedings of the AAAI 1998 Spring Symposium on Intelligent Environments*, pages 155–160, 1998.

[19] Martin Halvey, Mark T Keane, and Barry Smyth. Time based segmentation of log data for user navigation prediction in personalization. In *Proceedings of the International Conference on Web Intelligence, Compiegne, France, 19-22 September, pp.~636–640. IEEE Computer Society, Washington, DC, USA.*, 2005.

[20] Martin Halvey, Mark T Keane, and Barry Smyth. Time based patterns in mobile-internet surfing. In *Proceedings of the SIGCHI Conference on Human Factors in computing systems, Montreal, Quebec, Canada, 22-27 April, pp.~31–34. ACM, New York, USA*, 2006.

[21] Jiawei Han, Jian Pei, and Micheline Kamber. *Data mining: concepts and techniques*. Elsevier, Amsterdam, Netherlands, 2011.

[22] Jongyi Hong, Eui-Ho Suh, Junyoung Kim, and SuYeon Kim. Context-aware system for proactive personalized service based on context history. *Expert Systems with Applications*, 36(4):7448–7457, 2009.

[23] Richard Hull, Philip Neaves, and James Bedford-Roberts. Towards situated computing. In *Wearable Computers, 1997. Digest of Papers., First International Symposium on*, pages 146–153. IEEE, 1997.

[24] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. A survey of mobile phone sensing. *IEEE Communications magazine*, 48(9), 2010.

[25] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. Prefminer: mining user's preferences for intelligent mobile notification management. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12-16 September, pp.~1223–1234. ACM, New York, USA.*, 2016.

[26] Worapat Paireekreng, Kowit Rapeepisarn, and Kok Wai Wong. Time-based personalised mobile game downloading. In *Transactions on Edutainment II, pp.~59–69*. 2009.

[27] Veljko Pejovic and Mirco Musolesi. Interruptme: designing intelligent prompting mechanisms for pervasive applications. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13-17 September, pp.~897–908. ACM, New York, USA.*, 2014.

[28] Min Peng, Guanyin Zeng, Zhaoyu Sun, Jiajia Huang, Hua Wang, and Gang Tian. Personalized app recommendation based on app permissions. *World Wide Web*, 21(1):89–104, 2018.

[29] Santi Phithakkitnukoon, Ram Dantu, Rob Claxton, and Nathan Eagle. Behavior-based adaptive call predictor. *ACM Transactions on Autonomous and Adaptive Systems*, 6(3):21:1–21:28), 2011.

[30] Reza Rawassizadeh, Martin Tomitsch, Katarzyna Wac, and A Min Tjoa. Ubiqlog: a generic mobile phone-based life-log framework. *Personal and ubiquitous computing*, 17(4):621–637, 2013.

[31] Tom Rodden, Keith Cheverst, K Davies, and Alan Dix. Exploiting context in hci design for mobile systems. In *Workshop on human computer interaction with mobile devices*, pages 21–22. Glasgow, 1998.

[32] Nick Ryan, Jason Pascoe, and David Morse. Enhanced reality fieldwork: the context aware archaeological assistant. *Bar International Series*, 750:269–274, 1999.

[33] Iqbal H Sarker. Behavminer: Mining user behaviors from mobile phone data for personalized services. In *Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom 2018), Athens, Greece*. IEEE, 2018.

[34] Iqbal H Sarker. Silentphone: Inferring user unavailability based opportune moments to minimize call interruptions. *EAI Endorsed Transactions on Mobile Communications and Applications*, 2018.

[35] Iqbal H Sarker. Understanding the role of data-centric social context in personalized mobile applications. *EAI Endorsed Transactions on Context-aware Systems and Applications*, 2018.

[36] Iqbal H Sarker, Alan Colman, Muhammad Ashad Kabir, and Jun Han. Individualized time-series segmentation

for mining mobile phone user behavior. *The Computer Journal, Oxford University, UK*, 61(3):349–368, 2018.

[37] Iqbal H Sarker, Muhammad Ashad Kabir, Alan Colman, and Jun Han. Designing architecture of a rule-based system for managing phone call interruptions. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, USA*, pages 898–903. ACM, 2017.

[38] Iqbal H Sarker, Muhammad Ashad Kabir, Alan Colman, and Jun Han. An improved naive bayes classifier-based noise detection technique for classifying user phone call behavior. In *Proceedings of the 2017 Australian Data Mining Conference (AusDM 2017), Melbourne, Australia*. Springer, 2017.

[39] Iqbal H Sarker and Flora D Salim. Mining user behavioral rules from smartphone data through association analysis. In *Proceedings of the 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Melbourne, Australia*, pages 450–461. Springer, 2018.

[40] Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on*, pages 85–90. IEEE, 1994.

[41] Bill N Schilit and Marvin M Theimer. Disseminating active map information to mobile hosts. *IEEE network*, 8(5):22–32, 1994.

[42] Yuanchun Shi. Context awareness, the spirit of pervasive computing. In *Pervasive Computing and Applications, 2006 1st International Symposium on*, pages 6–6. IEEE, 2006.

[43] Vijay Srinivasan, Saeed Moghaddam, and Abhishek Mukherji. Mobileminer: Mining your frequent patterns on your phone. In *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing, Seattle, WA, USA, 13-17 September, pp.~389–400. ACM, New York, USA.*, 2014.

[44] International Telecommunication Union. Measuring the information society. In *Technical report, http://www.itu.int/en/ITU-D/Statistics/ Documents/publications/misr2015/MISR2015-w5.pdf*, 2015.

[45] Andy Ward, Alan Jones, and Andy Hopper. A new location technique for the active office. *IEEE Personal communications*, 4(5):42–47, 1997.

[46] Ian H Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[47] Yanxu Zheng, Sutharshan Rajasegarar, and Christopher Leckie. Parking availability prediction for sensor-enabled car parks in smart cities. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015 IEEE Tenth International Conference on*, pages 1–6. IEEE, 2015.

[48] Hengshu Zhu, Enhong Chen, Hui Xiong, Kuifei Yu, Huanhuan Cao, and Jilei Tian. Mining mobile user preferences for personalized context-aware recommendation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(4):58, 2014.