# Data analytics and visualization for connected objects: A case study for sleep and physical activity trackers (Preprint version)

Karim Tabia[1], Hugues Wattez[1], Nicolas Ydée[1], and Karima Sedki[2]

[1] Univ Lille Nord de France, F-59000 Lille, France
UArtois, CRIL UMR CNRS 8188, F-62300 Lens, France
tabia@cril.univ-artois.fr
[2] LIMICS (INSERM UMRS 1142), Univ. Paris 13,
UPMC Univ. Paris 6, Sorbonne Universités, France

**Abstract.** In recent years, a large number of connected objects for the monitoring of activities (health and well-being, sleep, fitness, nutrition, etc.) have emerged and are very popular with the general public. No doubt that their price, their ease of use and their interest in their health and well-being contribute to this success around the world. However, many of these consumer-connected objects suffer from several limitations, especially with regard to their high-level or smart functionalities and the added value of the information provided by such objects. In this paper, we first focus on such limitations then provide a real case study in monitoring physical activities and sleep using Fitbit smart watches. We propose some high level functionalities by taking advantage of the large amount of data collected and using data analytics and visualization techniques.

## 1 Introduction

Wearable connected objects that focus on health and well-being (eg connected watches, fitness wristbands, sleep trackers, etc.) are largely being used and are being democratized on a large scale [10]. Reluctance regarding their use still exists for various considerations [6,5,2] such as the lack of security of the data collected, the lack of customization, the lack of functionalities and services ”intelligent” and high level, the lack of interoperability, etc.

In the last decade, more and more IoT devices and applications are dedicated to help users monitor and improve their health and well-being [8]. Physical activity trackers (also called activity monitors or sports/fitness monitors or activity trackers) are connected devices (such as wrist bracelets or Smart watches) or Smartphone applications that measure the intensity, duration and amount of the activities performed by the user. They measure features such as walking or running distance, calory consumption and expenditure, and in some cases, heart rate and ”quality” of sleep. Such devices are widely used and very popular as they aim at helping users to increase their motivation and improve their performences. Moreover, in addition to emproving well-being, the trackers allow user’s awareness about physical activity, they help ensuring regular physical activity and decreasing the risks of many chronic diseases [3,10].

Wearable activity trackers and more generally most IoT devices still have many issue regarding interoperability, security, personalization and sensors reliability to name a few [6,5]. The lack of ”smart” features and services is due, among other things, to not taking advantage of the large amounts of data collected and generated to learn and automatically extract information that can be very useful. For example, with devices such as smart watches and connected bracelets, we can collect a lot of data of physical activities and sleep (eg distance traveled, number of steps, number of awakenings, duration of deep sleep ...). The majority of these devices are limited to displaying simple statistics describing these activities. It is clear that extracting information like correlations between factors, recurring patterns, trends, etc. from these physical activity and sleep data and possibly cross-referencing them with other factors (eg environmental factors such as temperature, noise, lumunosity, pollution, etc.) can provide interesting elements to explain for example what is strongly correlated with disturbed sleep. This type of information derived from the analysis of activity data constitutes a very interesting added value [9,1]. Now, even if useful and relevant information has been extracted, it is necessary to represent the results of the analytics in an intuitive and efficient way so that the user can directly perceive, for example, the impact of the different factors on her sleep quality. Visualization will play an important role for the effective retrieval and interpretation of results by users [4]. Our study confirms that big data and analytics along with appropriate knowledge visualization techniques can help improve the usefulness of activity trackers and more generally IoT devices.

In this paper, we first focus on large public IoT limitations and provide a real case study in monitoring physical activities and sleep using Fitbit smart watches. Two users are hired in this study for several months and real data has been collected in real life conditions. We propose then some high level and smart functionalities by taking advantage of the large amount of data collected and using data analytics and visualization techniques.

This paper is organized as follows: Section 2 recall the main limits and lack of smart functionalities in activity tracking connected objects. In Section 3 we present our case study and we provide discussions and concluding remarks in Section 4.

## 2 Main issues with activity trackers from a usefulness persective

Like many IoT devices, activity and sleep trackers still suffer from many problems and limits [2,6]. For instance, one can mention the reliability of the sensors, the lack of security while exchanging and storing the data, lack of interoperability with other devices (especially those of other manufacturers) and of course lack of intelligent or smart functionalities. In this paper, we focus only on user-related aspects in terms of useful functionalities and efficient user interaction.

### 2.1 Lack of personalization

One of the challenges facing the use of IoT is personalization, especially in case of general public applications. Personalization aims to provide services tailored specifically for the user needs and preferences. For example, IoT devices aiming at improving sleep quality would provide the user personalized recommendations regarding for instance *when to go to bed, temperature of the room, food/drinks to avoid, physical activity...*, and trigger alerts and warnings in case of sleep deficiency/disorder, physical activity deficiency, food deficiency/excess, etc. In order to learn the factors they may have an impact on the sleep quality of the user, there is need to measure the sleep quality and learn correlations with factors associated with good/bad sleep quality to provide useful recommendations while taking into account user profile, preferences and constraints (for more details on this issue, see [7]).

### 2.2 Lack of smart functionalities

While IoT devices are capable of collecting many types of data with their sensors and potentially acquire complemnetary data from other sources, the collected data is often neither correlated nor exploited to give relevant information, pieces of knowledge, advices and recommendations to the user for improving his life style. For instance, the quality of sleep is affected by a number of factors such as physical activity, dietary habits (alcohol intake, caffeine containing beverages, smoking, etc.), environmental factors, etc. It is then possible to correlate and analyse all these data in order to highlight which factors have a positive or negative impact on the user's sleep quality. Of course, this needs making use of big data/datamining techniques and data analytics to extract useful knowledge from the large amounts of user-generated and ambient data.

### 2.3 Lack intuitive and interactive dashboards and user interfaces

As just mentioned above, in addition to analyzing the data of physical activities and sleep collected by connected objects, we will have to report this data and potentially pieces of knowledge to the user, through user-friendly, intelligent and interactive interfaces. Thanks to visualization techniques, the user can then see for example disorders that he may have during his sleep and the link that this may have with his physical activities or ambient factors. Unfortunately, as shown in the dashboard of Fig. 1, most activity and sleep trackers display only very basic statistics and graphs summarizing the daily activities. There for instance no analysis or interactivity. It is true that the characteristics of the screens used by the users (smartphones or connected watches) impose limits for the display but the very basic nature of the data to display does not lend itself to elaborate visualizations as we will illustrate it in our case study.


**Fig. 1.** Example of a Fitbit tracker dashboard

## 3 Case study

In order to show what kind of improvements activity and sleep trackers could benefit from using data analytics (and more generally big data techniques) and visualization, we conducted a case study where two volunteers were hired in order to collect real data form an activity and sleep trackers in real life conditions.

In order to learn from the data relevant information (eg. the factors they may have an impact on the sleep quality of the user), there is need to use as much data as possible and extract such useful knowledge (eg. learn correlations and patterns with factors associated with good/bad sleep quality). Hence the collected data should include all i) the factors that may describe the sleep quality (eg. awakenings count, minutes asleep, etc) and ii) factors that may have an impact or influence on sleep quality (eg. physical activity, food, temperature, noise, etc.). We distinguish two types of data:

- *User-generated data:* Such data can be sensed by dedicated devices (ex. physical activity) or by the user through a graphical interface on a laptop or mobile phone (ex. food intake). In our study, user-generated data consists only in physical activity and sleep data collected by the used Smart watches.

- *Ambient data:* This refers to factors such as air quality, noise level, temperature, luminosity, etc. Such data could also be provided by some IoT devices or retrieved from special applications and sources (eg. Web sources). In our study, we used external sources to acquire weather and air quality data corresponding to the location where the volunteers were living.

### 3.1 Specifications of the activity trackers used in the study

The two Fitbit Smart watches that are selected are Fitbit Blaze having the following specifications:

- *Sensors:* An altimeter, ambient light sensor, three-axis accelerometer, optical heart rate monitor. In addition, a connected GPS is used.
- *Data tracked:* As detailed below, the trackers collect activity data (eg. distance, floors climbed, heart rate...) and sleep (eg. time in bed, awaikening account, etc). Some data is not sensed but derived from activity data such as the number of calories burned.
- *Operating system:* The devices are connected with smartphones running Apple MacOS X 10.6 or later, Microsoft Windows Vista or later.

For more details on the Fitbit Blaze tracker, please see [1]

### 3.2 Raw data

Like most activity and sleep tracking tools, the Fitbit Blaze trackers just collect some data and remotely store it using a Smartphone connection in the form of log records which can be downloaded or displayed in various ways to the user. We collected the raw data on a monthly basis. Each month a .csv file is downloaded and contains only the records of activity and sleep (of course, other information could be downloaded). The following is an example of log records for activity data.

[1] Fitbit Blaze User Manual
`https://staticcs.fitbit.com/content/assets/help/manuals/manual_blaze_en_US.pdf`

```
Date,CaloriesBurned,Steps,Distance,Floors,MinutesSedentary,MinutesLightlyActive,MinutesFairlyActive,
MinutesVeryActive,ActivityCalories
"15-11-2016","2 184","9 800","6,14","14","662","164","30","23","896"
"16-11-2016","2 202","9 707","5,96","18","656","164","14","39","933"
"17-11-2016","2 126","9 847","6,13","19","726","145","20","29","796"
"18-11-2016","2 250","8 695","5,07","17","642","291","2","7","1 013"
"19-11-2016","2 271","8 744","5,15","2","667","270","6","4","1 009"
"20-11-2016","1 888","4 201","2,34","11","525","165","0","0","533"
"21-11-2016","2 222","10 261","6,26","15","720","184","18","34","971"
"22-11-2016","2 135","8 750","5,52","17","1 260","146","7","27","829"
"23-11-2016","2 209","8 993","5,68","17","713","141","22","42","920"
"24-11-2016","2 029","7 581","4,77","14","731","177","3","8","727"
"25-11-2016","2 079","8 577","5,24","16","705","189","5","24","802"
```

Here is an example of log records for sleep.

```
Date,minutesAsleep,minutesAwake,awakeningsCount,timeInBed
15-07-2016,"516","47","3","574"
16-07-2016,"500","26","0","526"
17-07-2016,"620","59","5","679"
18-07-2016,"366","27","2","393"
```

The collected files correpond to the activity tracking of two users wearing Fitbit Blaze trackers during 8 months. In the current work, the date columns are omitted since we do not mine sequential patterns. Of course, this means that we make a strong hypothesis implying that each day data is independant of the previous and following days. This hypothesis is very debatable for assessing for example the quality of sleep and physical activity (for instance, three days of bad sleep but distant from each other should be considered less serious than three days in a row of bad sleep). But with this hypothesis, we lose the dependence and relationship between the activities of successive days.). This issue is discussed in the last section and left for furture works.

### 3.3 Data preprocessed and enrichment

As mentioned earlier, in order to provide relevant information, there is need to enrich the tracking data provided by the trackers with other data. For instance, there is no doubt that ambient data (eg. room temperature, noise level, luminosity, etc.) has an impact on the sleep quality but such data is not provided by the used Fitbit trackers. In our study, we used APIs for getting i) weather data (eg. temperature, humidity, wind, etc.) from `http://www.meteociel.fr/`, ii) air quality data from `http://www.atmo-hdf.fr` dedicated to monitoring and assessment of air quality in Hauts-de-France (location where the two volunteers hired in this study live) and from `http://www.openhealth.fr/`.

Many technologies and techniques are used to process and mine the data and display the results. First, the csv files containing tracking data are organized by day and by user. Each raw contains both activity and sleep data. Then the data is enriched with ambient data for that day. For visualization, data is first put into a relational database such that the visualization scripts needing complex SQL queries could run efficiently in real time to allow user interactivity.

### 3.4 Analytics and high level feature definition

This is the part where the added value is expected to be mined from the data collected from the trackers and other sources. In this study, we addressed two types of relevant information that can be automatically extracted from data:

– *Correlations among basic features:* The goal here is to extract meaningful correlations among the features of the data. For instance, one could be interested to know *how* and *to which extent* the intensity of physical activity is related the number of awakenings. In this study, we extract different types of statistical correlations such as Pearson, Spearman and Kendall correlations.

– *High level correlations:* Since the correlations among basic features are somewhat limited from a user's point of view, we need to define high level features such that the user finds the provided information really relevant. For instance, in order to learn the factors they may have a negative/positive impact on the sleep quality of the user, there is need to "assess" the sleep quality from the basic features then learn correlations with factors associated with good/bad sleep quality (what is formally a good or bad sleep?). Sleep quality could automatically be estimated based on sleep tracking data (ex. awakenings count) or through forms answered each day by the user assessing his sleep quality. Note that there is to the best of our knowledge no high level sleep measure in the litterature that could be directly computed from the basic features[2]. In this study, we simply relied on an existing score called *sleep efficiency* assessing sleep quality as the number of minutes of sleep over the number of minutes spent in bed which is obviously not satisfactory. The definition of sleep quality measures should be done by experts or subjectively rated by the user herself.

### 3.5 Visualization

Once relevant informations extracted from the data, there remains to display it for the user. One of the goals here is to display the results in an intuitive and interactive way while most of the time users monitor their activities through smartphones or directly on smart watches having very limited display possibilities. The following provides few examples illustrating how relevant information (eg. correlations) could be displayed.

In order to display statitistical correlations existing among some features of interest, a very easy way is to superpose the graphs of features of interest as illustrated in Fig. 2. The superposed graphs when appropriately scaled visually

[2] Polysomnography (PSG) is the main method of sleep monitoring used to detect sleep disorders in a medical context. It is performed during a whole night in a hospital by placing on the patient different signals (electrodes on the scalp, eyelids, and chin, etc.). Such a method is efficient but it is obtrusive, costly and cannot be conducted frequently on a large number people.

display some inter-relationships between the chosen features (here, the number of burned calories, the weight and the number of awakeneings).

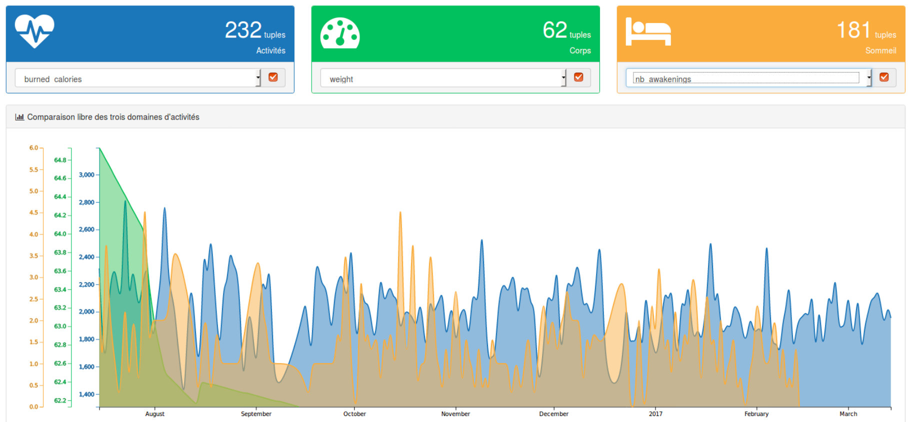


**Fig. 2.** Example of superposed plots displaying inter-relationships among features of interest

One could also use a simple correlogram (see of Fig. 3) where respectively colors and cercle diameters tell about the positive/negative nature of the correlations and the strength of these correlations.

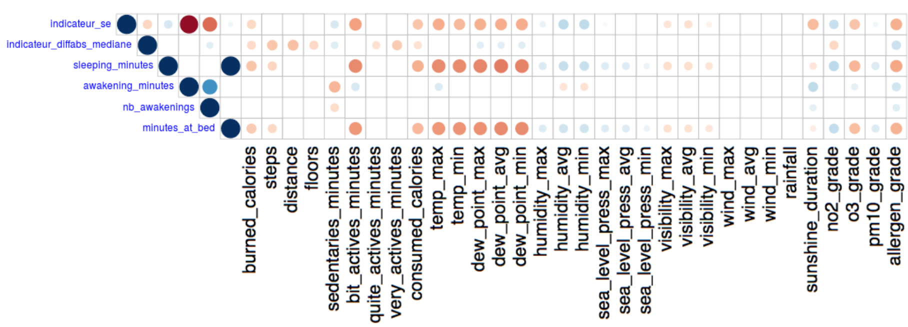


**Fig. 3.** Example of a correlogram displaying Pearson correlations among some features

Another way to display inter-relationships among features is the chord diagram of Fig. 4. This diagram has the advantage of being intercative since one can click on any feature and see the other features correlated to it.

In order to give a more global view on the user activity and sleep over time, one could superpose the plots of activity or sleep of successive weeks (Fig. 5

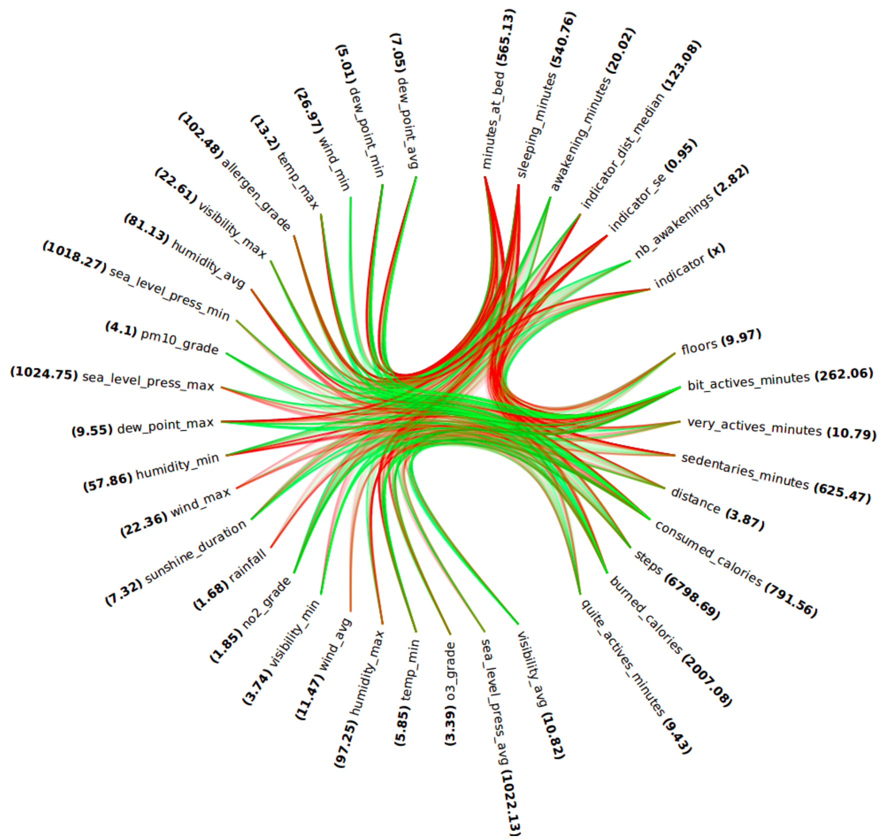


**Fig. 4.** Example of a chord diagram displaying inter-relationships among some features

left) or successive months (Fig. 5 right). Doing so, the user could recognize some repeating patterns due to working days/weekend rythm of life, holidays, seasons or any repeated events and cycles. The intuition underlying this point is that the weekly/monthly/annual, etc. organization of daily life has an impact on the activity and the sleep of the users. This visualization will show some phenomena such as the Sunday blues that negatively affects sleep the days before the beginning of each working week.

Many other visualizations could be used and many other forms of knowledge could be extracted and displayed. As for implementation, for the analytics part, we used R packages while we use many Web-based technologies (eg. D3.js) for the visualization part. Following section focuses on the main issues encountered in our case study.

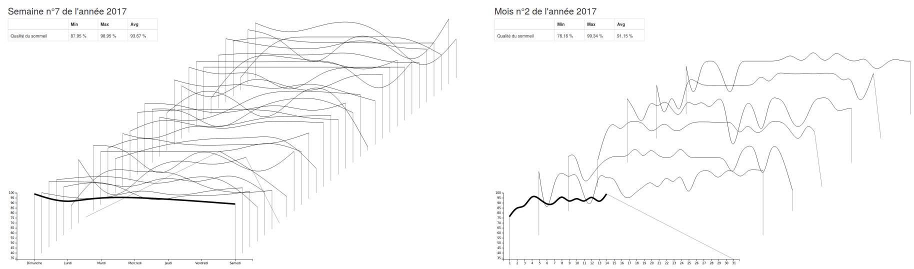


**Fig. 5.** Example of superposed plots of sleep

## 4 Concluding remarks and conclusions

This paper focused on a practical application combining data analytics and visualization to improve activity and sleep trackers. We showed that even with very limited sensed data, it is possible to extract relevant information and pieces of knowledge and display the results in a more intuitive and interactive way to the user. The following are the main concluding remarks of our study.

– *Missing and noisy data:* While collecting data over time, we realized that missing data for whole days can affect the collected data. Indeed, it is not rare that the users forget to wear their smart watches or forget to recharge them. This could be a real issue if sequential patterns were to be mined from the collected data. Techniques to be used should also cope with the limited reliability of some sensors and also some derived data such as calories burned. The users themselves realized that some measures are not exact and this may affect the correlations and more generally the relevant information we want to extract. Hence, one would make use of techniques that are robust and less sensitive to noise or use smoothing techniques for example to cope with this issue.

– *Sequential dimension:* In the current study, no sequential patterns are mined and no extracted information needs to take into account the sequencial dimension of data. It is obvious that physical activity and sleep in last few days influence those of the following ones. Moreover, most people's physical activity and sleep is influenced by weekly rythm... as pointed out above.

– *Beyound statistical correlations:* The problem here is that the interpretation (hence the utility) of the statistical correlations between the basic attributes which is quite limited. For example, knowing that there is a correlation between the number of minutes spent in bed and the number of active minutes of physical activity is not very meaningful for a user. What would be really useful is to define quality indicators of good sleep and good physical activity. For this latter, for example, an indicator can be defined if the number of steps or the number of minutes of recommended activities per day is exceeded. Same for sleep. We believe that such indicators are still limited and that a good quality of sleep and activities depends on several parameters and its evaluation must take into account the history of activities. For example, the quality of physical activity and sleep over several days should be evaluated instead of days separately. What can also be very useful is to mine cause-effect relationships instead of just statistical correlations. Cause-effet relations could help the user to make changes in his actitvities to improve his well-being.

– *Beyond user-generated data:* Through our case study it becomes obvious that in order to provide smart functionalities there is need to use the data collected by the sensors but also some other user-generated data (ex. food intake, smoking or user profile can be provided by the user through a graphical interface on a laptop or mobile phone). Other sources of information are ambient data (ex. weather, pollution, ...). The main issues here are the one heterogeneous data provided by different sources and the one of interoperability between different IoT devices such as connected scales, connected thermometers... that could be used in our case study.

– *Towards more interactivity and smartness:* Many smart functionalities could be added based on the knowledge that can be extracted form the available data. For example, it is possible to design prediction models such supervised classifiers and regression models that could answer some user queries or make recommendations. One could for example predict her sleep quality given her physical activity and other data for a given day. An example of recommendattion would be to advice at what time it is best to go sleep, which food to avoid or set the temperature of the room to a given level to have optimal sleep. To increase interactivity along with such functionalities, one could for example want to know what will be her sleep quality if she exercises a given amount of physical activity. Such prediction models could be built using history data of physical activity and sleep of the user. This latter will have a friendly user interface to set any parameter of interest for her and see what will be the impact of the fixed parameters on his sleep quality or any other parameter of interest to her.

To sum up, this preliminary work focused on the high-level functionalities that can be implemented to improve the usefulness of physical activity and sleep trackers. Our approach is to exploit the data collected by the connected objects possibly enriched with other data to extract relevant information and display it to the user in a visual and interactive way. We presented a real case study where we highlighted the problems encountered and built a prototype for demonstration. We finally propose several tracks for future works. This will require relying on big data techniques and analytics as well as exploiting the power of visualization techniques for better interaction with the user.

## References


1. Ejaz Ahmed, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Imran Khan, Abdelmuttlib Ibrahim Abdalla Ahmed, Muhammad Imran, and Athanasios V. Vasilakos. The role of big data analytics in internet of things. *Computer Networks*, 129:459 – 471, 2017. Special Issue on 5G Wireless Networks for IoT and Body Sensors.
2. Mahmoud Ammar, Giovanni Russello, and Bruno Crispo. Internet of things: A survey on the security of iot frameworks. *Journal of Information Security and Applications*, 38:8 – 27, 2018.
3. J. Larry Durstine, Benjamin Gordon, Zhengzhen Wang, and Xijuan Luo. Chronic disease and the link to physical activity. *Journal of Sport and Health Science*, 2(1):3 – 11, 2013. Children's Physical Activity and Health: Chronic disease in children and young adults.
4. Dongxiao Gu, Jingjing Li, Xingguo Li, and Changyong Liang. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. *International Journal of Medical Informatics*, 98:22 – 32, 2017.
5. So-Eun Lee, Mideum Choi, and Seongcheol Kim. How and what to study about iot: Research trends and future directions from the perspective of social science. *Telecommunications Policy*, 41(10):1056 – 1067, 2017. Celebrating 40 Years of Telecommunications Policy – A Retrospective and Prospective View.
6. Tyson Macaulay. Chapter 8 - availability and reliability requirements in the iot. In Tyson Macaulay, editor, *RIoT Control*, pages 141 – 155. Morgan Kaufmann, Boston, 2017.
7. Thibaut Vallee, Karima Sedki, Sylvie Despres, M.-Christine Jaulant, Karim Tabia, and Adrien Ugon. On personalization in IoT. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, dec 2016.
8. Min Woo Woo, JongWhi Lee, and KeeHyun Park. A reliable iot system for personal healthcare devices. *Future Generation Computer Systems*, 78:626 – 640, 2018.
9. Jing Wu, He Li, Ling Liu, and Haichao Zheng. Adoption of big data and analytics in mobile healthcare market: An economic perspective. *Electronic Commerce Research and Applications*, 22:24 – 41, 2017.
10. Yuehong YIN, Yan Zeng, Xing Chen, and Yuanjie Fan. The internet of things in healthcare: An overview. *Journal of Industrial Information Integration*, 1(Supplement C):3 – 13, 2016.