1

Person monitoring with Bluetooth tracking

Mathias Versichele, Tijs Neutens, Nico Van de Weghe

1.1 The difficult nature of measuring human mobility

Human mobility on different spatial and temporal scales affects many processes taking place in our world. While few will disagree that the large increase of human mobility during the 21st century has improved our general quality of life, it is increasingly confronting us with some of its more negative implications as well: congestion in and around densely populated areas by daily commuter traffic and the resulting strain on our environment, safety issues arising from the gathering of large crowds in relatively small areas, sudden risks of global pandemics and the difficulty of containing them, etc. As such, an increase in human mobility should be accompanied by an increase in the understanding of the processes governing these movements in order to better mitigate their negative implications.

A starting point in learning more about these movements is by adequately measuring them. Until recently, this has been quite problematic. Qualitative methods such as shadowing and the collection of travel diaries are known to be error-prone and labor-intensive. An alternative method of tracking people in smaller-scale settings is through video surveillance systems. Despite technological advancements in the last decade, using cameras to reconstruct the movements of a large number of people in a non-idealized environment remains very difficult. Correctly identifying trajectories of individuals in one camera view is already nontrivial due to interactions between the moving objects, changing illumination in outdoor environments, etc. Reconstructing trajectories over multiple camera views is even more challenging and to date remains somewhat of a scientific fiction. A third way of measuring human movement is through the use of proxies: objects whose movements are in some way linked to the movement of humans and can thereby serve as indicators of these movements. A rather unusual example of this is the tracking of one-dollar bills throughout the United States (as already mentioned in Chapter 8), which could potentially offer insights in how people move from one state to another over time. In the end, however, it is the rapid development of positioning technologies such as GPS (global positioning system) and the growing penetration of these technologies in mobile devices, such as car kits and mobile phones, which acts as the main catalyst in a new and burgeoning research area. After all, these devices can be regarded as very good proxies for capturing human mobility.

As a result, there has been a rapid increase in the amount of mobility datasets. As these datasets tend to be large, the sheer volume of data confronts researchers with difficulties in extracting interesting and relevant knowledge. While the importance of this issue — often used in paradigms such as 'the data avalanche' — is undeniable, it should also be stressed that human mobility is not always as easily measurable as might be conceived at first sight. First and foremost, persons can move around with different means of transportation. As more and more vehicles are equipped with GPS navigation kits, the movement data from these vehicles are already used for purposes such as the real-time monitoring and prediction of traffic jams. Other means of transportation, however, are more difficult to measure. In order to capture the movements of cyclists and pedestrians, for example, the device used as a proxy should not be linked to the means of transport but rather to the person itself.

As mobile phones usually remain very close to their owners at all times, they are the obvious candidate proxies for pedestrians and cyclists. Because mobile operators keep records of telephone calls making use of their cell towers, it is possible to reconstruct movements of phones by mining their call logs. This methodology — sometimes called 'mobile positioning' — delivers very large mobility datasets that have already been used to study regional movements. GPS loggers carried around by a test audience form an alternative way of measuring the movements of people. Because the resulting trajectories are usually very accurate and participants can also be surveyed before or after their cooperation, this method is becoming increasingly popular among scientists.

Both methodologies, however, have their deficiencies. First, the cooperation with mobile operators for mobile positioning datasets has proven to be difficult. More importantly, the spatial accuracy of this method-

 $\mathbf{2}$

ology (at best a few hundred meters in urban settings) is insufficient for studying human mobility on smaller scales. Alternatively, the distribution and recollection of GPS loggers among a test audience is laborintensive and possibly expensive, which will automatically result in a smaller sample size. Additionally, research projects making use of this technology will essentially be limited to outdoor environments where shadowing due to dense urban environments can potentially lower data quality as well.

This difficult nature of capturing human mobility on smaller (in this context sub-regional) scales shows that, despite the undeniable data avalanche confronting researchers, there remain challenges in capturing movement data besides processing them. In short, there is a need for a methodology that can measure human movement on a small-scale in a cost- and labor-effective way, in a wide variety of environments and for a sufficiently large sample size in order to make representative statements for the entire population.

1.2 How Bluetooth offers an alternative solution

In response to these issues regarding data collection and given the ubiquity of Bluetooth-enabled devices such as mobile phones and personal digital assistants (PDA) carried around by their owners, Bluetooth technology has increasingly been suggested as a simple and low-cost alternative for the reconstruction of spatiotemporal behavior. Section 1.5 outlines some of the research that has already used Bluetooth as a tracking technology. 'Discoverable' devices — and by extension their owners can be traced by means of a unique Media Access Control (MAC) address that gets broadcasted in the Bluetooth discovery process. Because this MAC address cannot be directly linked to any personal (or other sensitive) information, individuals remain anonymous avoiding potential privacy infringements.

1.2.1 Bluetooth tracking methodology

Bluetooth *scanners* — depicted in Figure 1.1 — can sense the presence of discoverable Bluetooth devices in its vicinity by continuously inquiring for nearby devices with a Bluetooth *sensor* and logging the broadcast messages sent by responding mobile devices within the scanners communication range. Every time a device is detected, its MAC address,

COD (Class of Device) code, and the timestamp of the detection are registered. Additionally, the received signal strength intensity (RSSI) of the inquiry response is logged. This intensity value is inferred from the received power level with which the response packet was detected by the scanner and is theoretically negatively correlated with the distance between the scanner and the detected device. Because some users include personal information in the friendly name of the detected device (name, phone number, etc.), it is not registered to safeguard privacy. The inquiry phase does not require an active connection between the scanner and the mobile device, so the methodology does not need an active cooperation of the tracked individual.

By placing Bluetooth scanners at different strategic locations, meaningful trajectories generated by mobile devices (and correspondingly by their owners) can be reconstructed. Because of the complex environmental setting and the resulting unpredictability of the propagation of Bluetooth signals, positioning is currently done through the *proximity* principle, where the position of a detected mobile device is approximated to the point-position of the scanner by which it is detected. The strategic locations of the scanners are used to semantically enrich the resulting trajectories which then become geo-localized semantic trajectories. As with any other form of episodic movement data, the location of mobile devices that are not within range of any scanner is unknown.

The spatial granularity of the resulting trajectories ultimately depends on the detection range of the Bluetooth scanners, and on the number and coverage of Bluetooth scanners within the study area. In theory, the detection range depends on the power class of the Bluetooth device (class 1: 100 m, class 2: 10 m, class 3: 1m). In practice, however, this range is variable due to environmental factors influencing (blocking, reflecting, etc.) radio signals leading to a detection region with a fuzzy border as previously depicted in Chapter 5. The temporal granularity is unpredictable because the Bluetooth scanners register detections whenever they arrive instead of using a fixed sampling interval. Devices within a direct line-of-sight with a sensor will usually lead to new detections every few seconds.

1.2.2 Preprocessing and software

The raw tracking data consists of log files — named after the combination of the scanner and the MAC address of the sensor — having the following format: *timestamp of detection*, *MAC address of the detected*



Figure 1.1 Components of a Bluetooth scanner for tracking purposes: computational unit (1), power source (2), USB cable (3), class 2 Bluetooth sensor (4), class 1 Bluetooth sensor (5), and different types of external antennas (6,7).

device, COD code of the detected device, RSSI of detection. In order to obtain a compressed dataset, the scanners are programmed to create a second set of log files during the scanning process in the following compressed format: timestamp of detection, MAC address of the detected device, COD code of the detected device, in/out/pass. A buffer time of 10 seconds is used to create detection time intervals from the detection time points. In is written when a device enters the detection range of the sensor, and out is written when the device leaves the range. Pass is used for solitary detections with no prior or later detections within 10 seconds. The principle of this logging system is depicted in Figure 1.2. In correct terminology, this compression actually transforms a geo-localized semantic trajectory into an abstract and structured semantic trajectory where individual detections are compressed into episodes representing

the presence of a mobile device within a scanner's range during a certain time interval.



Figure 1.2 Extract of logged data showing the raw time-point detection data (top) and the compressed time-interval data (bottom), depicting the compression of solitary detections into episodes leading to an abstract and structured geo-localized trajectory. This example shows one Bluetooth device (MAC address 20:21:A5:45:40:40) being detected 5 times on 20/07/2010 between 17:53:38 and 17:53:55 (CEST: Central European Summer Time). The buffer time of 10 seconds causes the raw data to be split into two separate detection time intervals (in \rightarrow out). The COD code of the device (5898756) shows that this was a regular cell phone.

This compressed episode-based representation adhering to the proximity principle is then imported into our processing environment for further analysis. Figure 1.3 shows a screenshot of this environment, dubbed a Geographical Information System for Moving Objects (GisMo). It is developed in java as a desktop client.

1.3 Case-studies

To give a general overview of the merits of the Bluetooth tracking methodology, we will show three case studies that have been carried out in three different application contexts: crowd management and safety at a mass event, and marketing insights in two retail environments: a professional fair and a shopping mall.



Figure 1.3 Screenshot of the GisMo analysis environment. MAC addresses have been partly smudged for privacy reasons.

1.3.1 Crowd management and safety at a mass event: Ghent Festivities 2010 & 2011

Because Bluetooth allows for non-participatory, unannounced and simultaneous tracking of a large number of individuals, it is particularly useful for monitoring visitor flows at mass events. However, despite this potential, only few studies using Bluetooth tracking at mass events have hitherto been reported in the academic literature (some are described in section 1.5). Hence, the methodology was tested at the Ghent Festivities, one of the largest outdoor cultural events in Europe which lasts for 10 days in July and attracts around 1.5 million visitors annually. This setting offers a challenging test bed in terms of crowd size, duration of the event and spatial extent of the study area (the historic city center of Ghent comprises around 4.5 km^2). Because of the size and the openness of the event — most activities in the festival are free, and there are no explicit entrance or exit points — collecting quantitative data such as visitor counts is challenging. The resulting lack of quantitative data acts as a bottleneck for research into the spatiotemporal dynamics of visitor movements. Exemplary to this is the issue of calculating the total number of visitors that attend the festival, which has traditionally been estimated by using proxy variables such as the daily amount of waste collected and the number of tram or bus tickets sold. As such, estimations vary but the general consensus is that approximately 1.5 million (non-unique) visitors attended the festival in 2010. Other than this rough figure and the use of video technology by the police department to give a qualitative indication of crowdedness or other safety issues, little is known about the general movement patterns of these visitors within and around the festival site: how long they stay at the festival, the number of days they visit the festival, how they reach the event, etc.

Given the limited range of Bluetooth scanners and the size of the event, a full coverage of the entire study area was impossible from a practical point of view. Instead, a careful selection of strategic coverage sites was made after consultation with local policy makers and urban experts, with the purpose of collecting as many significant individual movements as possible. In 2010, 22 locations were covered including the large public squares in the city center, a selection of points of access into the event zone, two train stations and a tram station located next to a park and ride facility. In 2011, we were able to capture visitor movements in the center in a more finely grained way by employing 43 scanners exclusively in and around the center of the city.

As overcrowding is usually regarded as the main danger at mass events, we started by using Bluetooth tracking as a counting methodology instead of a tracking methodology as such. In order to extrapolate from counts of detected devices to real numbers of people within the detection range of a scanner, we need to know the fraction of visitors that are detected by our system (corresponding to individuals carrying devices that have a discoverable Bluetooth interface). This penetration rate — also referred to as *detection ratio* — usually varies from event to event, but in this specific context it amounted to 11.0 ± 1.8 %. Using this figure, we can extrapolate and roughly estimate crowdedness levels. As an example of this use as a counting methodology, the daily and hourly variations in crowdedness of the event zone are illustrated in Figure 1.4. The hourly variation is characterized by a very smooth curve with sharp troughs in the morning (usually around 7 am). The peaks are also usually sharp and situated around 11 pm except for days 2, 5 and 9 where a broader peak in the late afternoon is observed. These correspond to two Sundays and the national day of Belgium (21/07), and these days are known to attract more daytime visitors (such as working couples with children). As a result, the sharp peaks around midnight do not appear because of the relatively larger crowdedness earlier in the afternoon. The three busiest days are immediately visible, with the fourth day being the most crowded with almost 10,000 detected phones or around 90,000 unique visitors in the festivities zone between 11 and 12 pm. To aggregate over daily periods, we had to carefully consider how to define a day. Looking at the hourly crowdedness, it is clear that it does not make much sense to define days starting and ending at midnight because that is generally the most crowded period of the day. Doing so would cause the Bluetooth observations to be segmented by unnatural breaks. Consequently, we have considered the starting point of an event day to coincide with the on average least crowded moment of a day, i.e. 7 am. The daily aggregates again show the three busiest days with day 4 peaking at almost 20,500 detected phones or 190,000 visitors.



Figure 1.4 Daily (dashed line, event days starting and ending at 7 am) and hourly (solid line) number of detected phones over the entire Ghent Festivities 2010 event zone as an indicator of crowdedness. Solid vertical gridlines point to midnights, dashed vertical gridlines are plotted every 4 hours.

While the number of visitors present at a certain location and time is already a good indicator for the likelihood of safety issues, the movement of visitors from one location to another offers even more insights into the spatiotemporal dynamics of a crowd. Although only flows of

visitors carrying discoverable Bluetooth devices can be reconstructed. the discovered patterns and trends can aid stakeholders to make wellinformed decisions regarding crowd-management and security in general. By making a time series of these flow diagrams, it is possible to investigate the time-dependency of certain visitor flows. Figure 1.5 shows a visualization of such dynamic visitor flows in Google Earth, comparable to the figures presented in Chapter 9. The KML file was generated in the GisMo environment and can be animated in time. Four snapshots are depicted — each depicting the cumulative flows over 30 minutes. In the afternoon (a), visitor flows are quite evenly spread over the event zone, except for the area in the north-east depicted by the yellow rectangle. Most of the large flows are balanced in their directionality, but the flows in the perifery are mainly inward oriented. Visitors regularly venture further from the center across the Leie and Lieve rivers (depicted by the green oval). In the evening (b), the region in the north-east has clearly sprung to life. Throughout the rest of the event zone, there is also an abundance of visitor flows. There is still a net inflow of visitors visible in the perifery. Later in the morning (c), we see some important differences with the previous view. First of all, most flows within the event zone seem to show a net migration to the north-east where there is a lot of activity. This is caused by nighttime visitors walking to this area after all music performances have ceased in the rest of the event zone. Additionally, flows surrounding the event zone now show a net efflux (most apparent in the southeast). Visitors generally stay closer to the center as well. Later around dawn (d), the largest flows are situated in the northeast whereas the areas that attracted large crowds during the day are rather desolated in comparison. More importantly, most flows now point away from the north-east. This represents the ongoing egress of visitors returning home.

1.3.2 Marketing insights in retail environments

As discussed above, Bluetooth tracking can be considered a helpful tool in aiding crowd-management during mass-events. However, the gathering of large crowds does not only cause negative consequences such as higher risks of safety issues. It also creates opportunities because large crowds represent large volumes of potential consumers when these people walk around in a retail environment. As such, marketing can be regarded as an application context for our methodology that is just as relevant as crowd safety. This is not much of a surprise as the P of place constitutes

10





(c) 21/07/2011 04:00-04:30

(d) 21/07/2011 06:00-06:30

Figure 1.5 Spatiotemporal variation of visitor flows during the Ghent Festivities 2011. Four snapshots show the cumulative flows during 4 time windows of 30 minutes. The red border delineates the official event zone where specific regulations are in order to make the event as safe as possible. The direction of an arrow indicates the direction of the flow; the width of the arrow indicates its size. The widths of the arrows are normalized to the size of the largest flow during each time period separately, so flows from other time windows cannot be directly compared based on this visualization.

an essential component of the classic marketing mix, next to price, product and promotion. Traditionally, place in a marketing context can be interpreted in several ways ranging from the physical location where a product is purchased to the distribution chain linked to a product. The 12

(changing) location of a client browsing or purchasing in a retail environment is equally relevant however. The opportunity to measure these movements in a (semi)automatic way with modern tracking technologies has even been hailed as a 'third wave of marketing intelligence'.

Visitor movements at a professional fair

Fairs might not represent a retail environment *sensu strictu* as the major aim is to showcase products or services instead of selling them, but visitor movements in these contexts are highly valuable nonetheless. Organizers of fairs often need to distribute a limited showcasing area to a large number of companies. These individual companies want to maximize their exposure, while the fair organizers want to optimize the general quality of both the visitors experience in general as well as the return on investment envisioned by the companies having booths at these events. Additionally, rental prices of areas occupied by exhibition stands not only depend on their size but also on their location. Since certain locations are already known to attract larger portions of the crowd (hotspots), these will be more expensive for companies wanting to place a booth there. In the end, however, fair organizers need detailed movement data in order to give more accurate estimates of these rental prices and possibly adapt the distribution of exhibition stands based on findings extracted from these data.

In light of this application context, a cooperation was set up with a well-known fair organizer owning a large exposition venue composed of 8 halls and covering an area of over 50,000 m^2 . During two editions of a large professional catering fair (2009 and 2010), the Bluetooth tracking methodology was tested in this indoor environment. Some basic results are shown in this section.

Tests showed that 35 % of the visitor population was tracked, which is significantly higher than the detection ratio of around 11 % during the Ghent Festivities. The most important factor contributing to this higher figure is most likely the increased penetration of Bluetooth-enabled devices in the population of catering professionals. Figure 1.6a shows the distribution of the number of halls visited per detected individual for the editions in 2009 and 2010. The curve for the 2009 edition clearly shows a smaller share of individuals visiting 4 halls or less, and a higher share visiting 5 halls or more. In short, visitors tended to visit more halls on average in 2009 than in 2010. The histograms in Figure 1.6b show the distribution of time durations spent across the different halls. Durations of less than 5 minutes were filtered out for visualization purposes (these

represent people traversing a hall instead of visiting it anyway). There is a clear difference in average times spent in each hall. Visitors seem to spend most time in hall 1 (which is the main and also largest hall), followed by hall 8 (which is the second largest hall). The difference between the remainder of the halls (who are all equal in size and smaller than halls 8 and 1) is smaller. Visitors spend roughly equal amounts of time in halls 7, 4 and 3, followed by halls 2 and 5. Hall 6 is on average visited for the shortest amount of time.

Customer movements in a shopping mall

The value of modern tracking technologies in generating valuable marketing intelligence has already been touched upon. In order to examine the specific merit that Bluetooth tracking could hold in this context, the technology was also tested in a retail environment *sensu strictu*: a shopping mall that consists of 39 stores of varying size distributed over 3 floors. The movement of customers from one store to another was registered during a one-month period leading up to Christmas. Scanners were also placed at the entrances and the subterranean parking lot in order to analyze visitor flows in and out of the venue. Table 1.1 shows the number of visitors that were detected in each of the stores inside the shopping hall, sorted from the most popular clothes store to a photo services store that attracted the smallest share of visitors. As is the case in most shopping malls, one can see that there are a number of dominant anchor stores accompanied by smaller stores.

As an example on how these tracking data can be mined for interesting knowledge or patterns, we will focus on association rules between the different shops customers visit in the same shopping trip. As such, the sequence in which shops were visited is of no importance in this analysis. Additionally, note that we cannot distinguish between customers that made a purchase in a store and customers that did not. More formally the problem can be defined as follows. Let $I = i_1, i_2, ..., i_n$ be a set of binary attributes called *items*. In this specific case, these items represent a customers presence in each store. Each customers visiting pattern constitutes a *transaction*, which contains a subset of the items in I. An association rule can then be defined as $X \Rightarrow Y$ where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. The itemsets X and Y are called *antecedent* and *conse*quent respectively. Different measures can be used to select interesting rules from the set of all possible rules. The support of an itemset is defined as the proportion of transactions in the dataset which contain the itemset, and the support of an association rule is defined as the sup-



Figure 1.6 Insights delivered by Bluetooth tracking in an indoor fair environment: (a) difference in the number of visited halls between the 2009 and 2010 edition; (b) distributions of time spent in the different exhibition halls (class width of 5 minutes, durations < 5 minutes were filtered out for visualization purposes) during the 2009 edition.

port of its antecedent. The *confidence* of an association rule is defined as $\frac{support(X \cup Y)}{support(X)}$, and measures the confidence with which an antecedent can accurately predict the consequent. The *lift* of an association rule combines the previous two measures and is defined as $\frac{confidence}{support(Y)}$. As such, the lift takes both the confidence and the representativeness (support) of an association rule into account.

Table 1.1 Number of detected visitors at each of the venues in the shopping mall during the one-month tracking period, ranked from high to low. Venue names have been anonymized according to the type of products/services they offer (M: male, F: female).

Venue	Detected visitors	Venue	Detected visitors
clothes_MF	8064	clothes_F_5	378
supermarket	2694	$clothes_F_2$	376
household_3	1964	$clothes_M_1$	354
household_1	1526	snacks_sweet	260
clothes_knitting	1461	lingerie_2	247
books_etc_1	1171	bistro_1	231
$clothes_F_4$	972	$clothes_F_1$	226
$mobilephones_etc_1$	889	bistro_2	199
cosmetics_1	810	$clothes_M_2$	160
shoes	799	interim_office	121
hobby	776	optician	101
snacks	717	$mobilephones_etc_2$	93
$clothes_F_3$	704	juwelry	92
$home_entertainment$	673	flowers	75
household_2	667	hair_salon	52
lingerie_1	588	leatherware	51
cosmetics_2	575	photo_services	41
$books_etc_2$	511	-	

We used the popular WEKA data-mining platform (version 3.6.5) for a very succinct mining exercise. A preliminary run of the APRIORI algorithm for the 10 rules with the highest confidence (minimum support of 0.01) shows the following output:

Best rules found:

1. clothes_F_3=true clothes_F_4=true 238 \rightarrow clothes_MF=true 201 conf:(0.84)

2. clothes_F_4=true clothes_knitting=true 267 \rightarrow clothes_MF=true 217 conf:(0.81)

3. clothes_F_4=true clothes_F_5=true 220 \rightarrow clothes_MF=true 174 conf:(0.79)

4. household_3=true clothes_F_4=true 258 \rightarrow clothes_MF=true 199 conf:(0.77)

5. shoes=true clothes_knitting=true 221 \rightarrow clothes_MF=true 169 conf:(0.76)

6. clothes_F_1=true 241 \rightarrow clothes_MF=true 180 conf:(0.75)

7. clothes_M_1=true 385 \rightarrow clothes_MF=true 281 conf:(0.73)

8. clothes_F_4=true 1089 \rightarrow clothes_MF=true 777 conf:(0.71)

9. household_1=true shoes=true 236 \rightarrow clothes_MF=true 168 conf:(0.71)

10. clothes_F_5=true 414 \rightarrow clothes_MF=true 293 conf:(0.71)

The first important point to notice is that all rules contain clothes_FM

as an item in their consequent and some also have it in their antecedent. In fact, 54 out of the 64 rules found in total (minimum support of 0.01, minimum confidence of 0.3) contain this item. As this anchor store in the shopping mall attracts the majority of visitors (see Table 1.1), it appears in a large number of rules with high levels of confidence and hence also pollutes the view with rather obvious rules. Accordingly, we removed this store from the dataset and reran the algorithm (minimum support of 0.005, sort by lift with a minimum lift of 1.1) in order to mine for less obvious (and hence more interesting) rules. The algorithm finds 266 rules, out of which the 20 top rules are shown below:

Best rules found:

 $1. \ clothes_F_3=true \ clothes_F_4=true \ 238 \rightarrow clothes_F_5=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_F_4=true \ 238 \rightarrow clothes_F_5=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_F_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_F_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_F_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_F_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_F_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (0.37) < lift: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \ conf: (14.9) > 1. \ clothes_6=true \ 88 \$ lev:(0) [82] conv:(1.54) $2. \ clothes_F_5=true \ 414 \rightarrow clothes_F_3=true \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_F_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ conf: (0.21) < lift: (14.9) > 100 \ clothes_4=true \ 88 \ clothes_4$ lev:(0) [82] conv:(1.25) 3. clothes_F_4=true 1089 \rightarrow clothes_F_3=true clothes_F_5=true 88 conf:(0.08) < lift:(10.62) > lev:(0) [79] conv:(1.08) $4. \ clothes_F_3=true \ clothes_F_5=true \ 127 \rightarrow clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_F_4=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (0.69) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) > clothes_5=true \ 88 \ conf: (10.62) < lift: (10.62) < lift: (10.62) < lift: (10.62) < lift:$ lev:(0) [79] conv:(2.97)5. clothes_F_2=true 414 \rightarrow clothes_F_5=true 101 conf:(0.24) < lift:(9.83) > lev:(0.01)[90] conv:(1.29)6. clothes_F_5=true 414 \rightarrow clothes_F_2=true 101 conf:(0.24) < lift:(9.83) > lev:(0.01)[90] conv:(1.29)7. clothes_F_3=true 755 \rightarrow clothes_F_4=true clothes_F_5=true 88 conf:(0.12) < lift:(8.84) > 0lev:(0) [78] conv:(1.12)8. clothes_F_4=true clothes_F_5=true 220 \rightarrow clothes_F_3=true 88 conf:(0.4) < lift:(8.84)> lev:(0) [78] conv:(1.58)9. clothes_F_4=true 1089 \rightarrow clothes_F_5=true 220 conf:(0.2) < lift:(8.14)> lev:(0.01) [192] conv:(1.22)10. clothes_F_5=true 414 \rightarrow clothes_F_4=true 220 conf:(0.53) < lift:(8.14)> lev:(0.01) [192] conv:(1.98)11. clothes_F_3=true 755 \rightarrow clothes_F_5=true 127 conf:(0.17) < lift:(6.78) > lev:(0.01)[108] conv:(1.17)12. clothes_F_5=true 414 \rightarrow clothes_F_3=true 127 conf:(0.31) < lift:(6.78)> lev:(0.01) [108] conv:(1.37)13. clothes_F_1=true 241 \rightarrow clothes_F_4=true 91 conf:(0.38) < lift:(5.79)> lev:(0) [75] conv:(1.49)14. clothes_F_4=true 1089 \rightarrow clothes_F_1=true 91 conf:(0.08) < lift:(5.79)> lev:(0) [75] conv:(1.07)

16

15. lingerie_2=true 267 \rightarrow clothes_F_4=true 99 conf:(0.37) < lift:(5.68)> lev:(0) [81] conv:(1.48)

16. clothes_F_4=true 1089 \rightarrow lingerie_2=true 99 conf:(0.09) < lift:(5.68)> lev:(0) [81] conv:(1.08)

17. household_1=true clothes_F_3=true 162 \rightarrow clothes_knitting=true 83 conf:(0.51) < lift:(5.2)> lev:(0) [67] conv:(1.83)

18. clothes_knitting=true 1645 \rightarrow household_1=true clothes_F_3=true 83 conf:(0.05) < lift:(5.2)> lev:(0) [67] conv:(1.04)

19. household_1=true clothes_F_4=true 209 \rightarrow clothes_knitting=true 103 conf:(0.49) < lift:(5)> lev:(0) [82] conv:(1.76)

20. clothes_knitting=true $1645 \rightarrow \text{household}_1$ =true clothes_F_4=true 103 conf:(0.06)< lift:(5)> lev:(0) [82] conv:(1.05)

Again, clothes stores are abundant in the rules. The top 14 rules even exclusively contain clothes stores selling womens' fashion. The rest of the top 20 is completed with rules that also link with a household store, a lingerie store and a clothes store that also sells knitting accessories. Clearly, this shows that strong associations exist between stores that are focused on a more female oriented public. It might be interesting to focus on clothes stores that sell mens' fashion (clothes_M) in order to zoom in on a male audience. When we filter out the rules that do contain such a store in their itemset, we end up with the following 4 rules:

Best rules found: 49. clothes_M_1=true 385 \rightarrow snacks=true 83 conf:(0.22) < lift:(4.12)> lev:(0) [62] conv:(1.2) 50. snacks=true 874 \rightarrow clothes_M_1=true 83 conf:(0.09) < lift:(4.12)> lev:(0) [62] conv:(1.08) 197. clothes_M_1=true 385 \rightarrow household_3=true 111 conf:(0.29) < lift:(2.08)> lev:(0) [57] conv:(1.21) 198. household_3=true 2318 \rightarrow clothes_M_1=true 111 conf:(0.05) < lift:(2.08)> lev:(0) [57] conv:(1.03)

We find associations between one mens' clothes store and a snacks store and household store respectively. Although these rules are clearly less strong (low confidences), it is noteworthy that other and less trivial associations are found in comparison with female oriented rules. The smaller number and attraction of male-oriented stores will certainly be one of the main reasons as to why we see this female bias. Clearly, more research is needed in order to mine for interesting patterns that instead of stating the obvious should provide interesting and new knowledge.

1.4 Issues and future research

In this book chapter, we have demonstrated the merits of Bluetooth tracking as an innovative, inexpensive, unobtrusive and flexible methodology for measuring human mobility in a variety of contexts and environments. At mass events it can aid crowd managers by delivering quantitative data on crowd sizes and flows, and in retail environments it can extract marketing intelligence or other organizational intelligence through methods ranging from visual data exploration to data mining techniques such as association rule learning.

However, the unobtrusive nature of the tracking process resulting in large sample sizes automatically also constitutes a methodological issue: the possibility of biased results by oversampling certain segments of the total population of individuals. Adolescents with a higher education might indeed carry more Bluetooth-enabled devices than elderly people, while young children will probably never be detected. The potential difference in Bluetooth usage among different audiences might significantly influence generated insights. Accordingly, more research is needed into the use of discoverable Bluetooth-enabled devices by different population segments in order for Bluetooth tracking to evolve into a technology delivering accurate and reliable information to policy makers, crowd managers and marketing researchers. The penetration rates we found in our experiments ranged from around 11 % for a general audience to 35 % for a professional fair visitor profile. In the end, a more systematic way of calculating the percentage of the population being tracked will be necessary for more reliable extrapolations in the future. Additionally, the possible influence of time and space on the detection ratio needs to be investigated.

The tentative association rule analysis with the shopping mall data only shows a very small selection of data mining possibilities with Bluetooth tracking data. Specifically for association rules, it soon became clear that there is a need for methods that can filter out more interesting rules from a larger set of less interesting rules. Intelligent visualization and/or pruning of association rules instead of solely listing them will certainly aid in this process. Besides association rule discovery, other data mining methods such as those described earlier in Chapter 6 can also generate valuable knowledge from this type of episodic movement data. They might need further modifications, however, to handle the spatiotemporal complexity of Bluetooth tracking data.

1.5 References

[5] give a review of scientific progress in real-world surveillance through camera systems, while [10] gives a more recent example of this research domain. The data avalanche paradigm was used by [9]. The mobile positioning methodology, its accuracy and the complex cooperation with mobile operators is discussed in [2]. [14] discusses the use and added value of GPS tracking in different projects. More details of the Bluetooth protocol are given in [11]. Although Bluetooth tracking is still somewhat of a peculiarity in the tracking field, there have been some reported uses in different contexts. [15], [8] and [12] have used the methodology for studying the spatiotemporal dynamics of mass events; [7] used it for collecting travel time data on a highway segment; [6] deployed mobile Bluetooth sensors in order to study complex social systems. The visualization of visitor flows in Figure 1.5 is originally inspired by the work of [13], and the third wave of marketing intelligence is discussed by [4]. The concept of assocation rule discovery was introduced by [1], and [3] discuss different ways of visualizing association rules as a means to extract the more interesting rules.

References

- R. Agrawal and R. Srikant. Mining sequential patterns. In Data Engineering, 1995. Proceedings of the Eleventh International Conference on, pages 3–14. IEEE, 2002.
- [2] Rein Ahas, Anto Aasa, Antti Roose, Ülar Mark, and Siiri Silm. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, 29(3):469–486, June 2008.
- [3] Dario Bruzzese and Cristina Davino. Visual post-analysis of association rules. Journal of Visual Languages & Computing, 14(6):621–635, December 2003.
- [4] Raymond R Burke. The Third Wave of Marketing Intelligence. In Manfred Krafft and Murali K Mantrala, editors, *Retailing in the 21st Century*, pages 103–115. Springer, Heidelberg, 2005.
- [5] Hannah M Dee and Sergio A Velastin. How close are we to solving the

problem of automated visual surveillance? Machine Vision and Applications, 19(5-6):329–343, May 2007.

- [6] Nathan Eagle and Alex Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, November 2005.
- [7] Ali Haghani, Masoud Hamedi, Kaveh Farokhi Sadabadi, Stanley Young, and Philip Tarnoff. Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record: Journal* of the Transportation Research Board, 2160:60–68, December 2009.
- [8] Sven Leitinger, Simon Gröchenig, Stephanie Pavelka, and Michael Wimmer. Erfassung von Personenströmen mit der Bluetooth-Tracking- Technologie. In Angewandte Geoinformatik 2010, pages 220–225, Salzburg, Austria, 2010.
- Harvey J. Miller. The data avalanche is here. Shouldn't we be digging? Journal of Regional Science, 50(1):181–201, February 2010.
- [10] Brian E. Moore, Saad Ali, Ramin Mehran, and Mubarak Shah. Visual crowd surveillance through a hydrodynamics lens. *Communications of* the ACM, 54(12):64–73, December 2011.
- [11] B.S. Peterson, R.O. Baldwin, and J.P. Kharoufeh. Bluetooth inquiry time characterization and selection. *IEEE Transactions on Mobile Computing*, 5(9):1173–1187, 2006.
- [12] Hendrik Stange, Thomas Liebig, Dirk Hecker, Gennady Andrienko, and Natalia Andrienko. Analytical Workflow of Monitoring Human Mobility in Big Event Settings using Bluetooth. In Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, pages 51–58. ACM, 2011.
- [13] WR Tobler. Experiments In Migration Mapping By Computer. The American Cartographer, 14(2):155–163, 1987.
- [14] Stefan Van der Spek, Jeroen Van Schaick, Peter De Bois, and Remco De Haan. Sensing Human Activity: GPS Tracking. Sensors, 9(4):3033–3055, April 2009.
- [15] Mathias Versichele, Tijs Neutens, Matthias Delafontaine, and Nico Van de Weghe. The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, 32(2):208–220, March 2012.

Index

abstract and structured semantic trajectory, Bluetooth tracking methodology, 11 APRIORI algorithm, 20

association rule discovery, 18

Bluetooth discovery process, 7

Bluetooth scanner, 8

Bluetooth sensor, Bluetooth tracking methodology, 9

Bluetooth technology, 7

class of device (COD) code, Bluetooth discovery process, 8

compression of Bluetooth tracking data, 11

crowd management, Bluetooth tracking methodology, 12

crowdedness at mass event, Bluetooth tracking methodology, 13

- cyclists, measuring movement of, 6 data avalanche, 6
- detection range, Bluetooth sensor, 9 detection ratio, Bluetooth tracking

methodology, 13 discoverable Bluetooth devices, 8

discoverable state, Bluetooth discovery process, 8

episodes, Bluetooth tracking data, 11 episodic movement data, Bluetooth tracking methodology, 9

fairs, visitor movements at, 17 friendly name, Bluetooth discovery process, 8

geo-localized semantic trajectories, Bluetooth tracking methodology, 8

geo-localized semantic trajectory, Bluetooth tracking methodology, 11 Ghent Festivities, Bluetooth tracking methodology, 12 GisMo, Geographical Information System for Moving Objects, 11 GPS (global positioning system), 6 GPS loggers, 6 Gyrid, Bluetooth tracking software, 9 hardware components, Bluetooth tracking methodology, 9 human mobility, difficult nature of measuring, 6 inquiry phase, Bluetooth discovery process, 8 inquiry scan phase, Bluetooth discovery process, 8 line of sight (LOS) communication, Bluetooth tracking methodology, 9 MAC address, Bluetooth discovery process, 8 mobile positioning, 6 pedestrians, measuring movement of, 6 penetration rate, Bluetooth tracking methodology, 13 power class, Bluetooth sensor, 9 proxies, 6 proximity principle, Bluetooth tracking methodology, 8 received signal strength intensity, Bluetooth discovery process, 8

retail environments, Bluetooth tracking methodology, 15

shadowing, 5

Index

shoppin mall, customer movements in, 18
spatial granularity, Bluetooth tracking methodology, 9
temporal granularity, Bluetooth tracking methodology, 9
third wave of marketing intelligence, 17
travel diaries, 5

video surveillance systems, 5 visitor flows at mass event, Bluetooth tracking methodology, 15

WEKA, data mining platform, 20

22