Can the Study of Trajectories Help to Extract Information from Business Processes?

Simona Fioretto[1], Elio Masciari[1,2], Nicola Mazzocca[1] and Enea Vincenzo Napolitano[1]

[1]*Department of Information Technology and Electrical Engineering (University of Naples Federico II), Naples, Italy*
[2]*Italian National Research Council (ICAR-CNR)*

Abstract
Business Process Management (BPM) is a key field of research that needs to be considered both for the management of production processes (in terms of the realisation of products and services) and for the management of processes that deal with data and information. The business opportunities associated with information management are also becoming increasingly important in the context of digital tools used in everyday life, such as smartphones, social networks, etc. Process mining techniques support various stages of process management and are used to identify processes and their specifications. The aim of this paper is to use process mining techniques applied to smartphone usage data and supported by the application of trajectory mining techniques to investigate whether location-based information can help process management or vice versa.

Keywords
Process Mining, Trajectory Mining, Business Process Management

1. Introduction

A process is commonly defined in a simplistic way as a set of activities (which may or may not be sequential) that transform an input into an output. Process modelling can help to identify characteristics and issues that are common to different sectors, such as delays in manufacturing or service industries. Business Process Management (BPM) defines the phases required to manage a process, including Discovery, Analysis, Design/Redesign, Implementation and Control. The Discovery and Analysis phases are critical for understanding the interaction between the activities, actors and resources involved.
Given this, the context of user interactions in smartphone usage can be seen as a process; in fact, if the transition from one activity to another is a process, the path of applications used by a user can also be seen as a process, where the actor is the user, the activities performed are the applications, and the flow is the decision process of the user moving from one application to another. This paper focuses on the discovery and analysis of processes related to user activities, specifically the switching from one application to another on smartphones by three users.
To extract the process, the smartphone's event logs were analysed using a process mining tool

Macao'23: 2nd International Workshop on Process Management in the AI era, September, 2023, Macao, SAR
simona.fioretto@unina.it (S. Fioretto); elio.masciari@icar.cnr.it (E. Masciari); nicola.mazzocca@unina.it (N. Mazzocca); eneavincenzo.napolitano@unina.it (E. V. Napolitano)
0009-0006-8700-8188 (S. Fioretto); 0000-0002-1778-5321 (E. Masciari); 0000-0002-0401-9687 (N. Mazzocca); 0000-0002-6384-9891 (E. V. Napolitano)

that can discover and analyse the process through different views, Key Performance Indicators (KPIs) and charts. The aim is to use process analysis to understand the user's decisions regarding applications, specifically how they choose one activity over another.
The process analysis is based on the location variable to understand user behaviour using both location and trajectory information. Trajectory mining techniques were used to identify and analyse user trajectories based on location information. The aim is to understand whether trajectory mining can support process mining and improve the effectiveness and efficiency of the process.
Knowing the choice of users' apps with respect to the place where they are located can allow us to understand if being in one place rather than another can influence the user's behavior, and therefore influence his choice. In this way we want to show that in the use of the mobile phone as well as in other choices made by users, the place where you are can be a significant factor in studying user behavior influencing for instance marketing campaigns.
In this paper we will give an overview of the techniques used, including process mining and trajectory mining. We will then examine the selected dataset and demonstrate the results of applying these techniques. Finally, we will discuss the results obtained and possible future developments.

2. Overview

This section briefly explains the techniques used to conduct the experiment. The study of processes belongs to the research area of BPM. In [1] different phases of managing a process in a life cycle are described: Discovery, Identification, Analysis, (Re-)Design, Implementation, Monitoring. One of the most important techniques that really supports BPM is process mining. Process Mining, as described in [2], is a relatively new research discipline that sits between machine learning and data mining on the one hand and process modelling and analysis on the other. The idea of process mining is to discover, monitor and improve real-world processes by extracting knowledge from the event logs that are readily available in today's systems. Processes are extracted and analysed from the digital footprints of users. As [2] shows, there are three types of process mining: Discovery, which uses an event log and creates a model without using a priori information, Conformance, which compares an existing process model with an event log of the same process, and Enhancement, which improves an existing process model based on information about the actual process recorded in an event log.
Trajectory mining is a data mining technique for analysing motion data of objects or people over time and space [3, 4]. Trajectories can be represented as a sequence of spatio-temporal points [5] or as a continuous path [6] in space-time. The main goal of trajectory mining is to discover meaningful patterns [7, 8], such as common routes, unusual behaviour or mobility trends. Trajectory mining includes a number of techniques such as trajectory clustering [9, 10], trajectory segmentation [11], trajectory pattern mining [7, 12], Spatio-temporal analysis of trajectories [13], Classification of trajectories [14], Trajectory prediction models [15]. Trajectory mining has many applications in different fields. In transportation, it can be used to analyse traffic patterns [16] and optimise route planning, and to detect accidents or disturbances [17]. In surveillance [18], it can be used to track suspicious activity or detect anomalies. In human

behaviour analysis, it can be used to study mobility patterns, social interactions [19] and disease transmission [20]. Other applications include wildlife tracking, sports analysis [21] and location-based advertising.
The combination of trajectory and process mining techniques has been used in the medical field to monitor the spread of a particular disease (both at the body level and the epidemiological level) and to determine how it has affected a particular health process [22, 23, 24]. A case and user study for event-case correlation on click data in the context of user interaction events from a mobility sharing company is shown in [25].

3. Dataset description

The dataset used in this study is the ContextLabeler data-set [26, 27] [1]. The ContextLabeler dataset consists of a collection of CSV files containing over 45,000 data samples, each consisting of 1,332 features associated with a variety of physical and virtual sensors. The sensors include motion sensors, running applications, proximity devices and weather conditions, providing a comprehensive representation of the user's environment. The dataset was collected over a two-week period by three volunteers using the context labeler [2], an Android application that allows volunteers to freely annotate the collected data. Each data sample is associated with a ground truth label that describes the user's activity and the context in which they were during the collection experiment. Labels include user activities such as working, eating and exercising, and contextual information includes environmental factors such as temperature and humidity. The dataset consists of 45,681 data samples distributed across the three users as follows 8,456 samples for user 1, 17,882 samples for user 2 and 19,343 for user 3.
The dataset was collected 'in the wild', i.e. subjects were using their devices without any constraints on their natural behaviour, and is obtained using one-hot coding. Each vector consists of a set of features that can be summarised as follows:

- Time
- Activity Recognition
- Running Applications
- Weather conditions
- Audio
- Battery
- Bluetooth Connections
- Bluetooth Devices in proximity
- Display Status
- Location : Geographical coordinates of the user's location and the category of the most probable venue according to the location's coordinates.
- Wi-Fi
- Physical sensor
- Multimedia
- Label

[1]The dataset is avaiable here: https://github.com/contextkit/ContextLabeler-Dataset
[2]https://contextkit.github.io

3.1. Preprocessing

During the data processing phase, some columns were removed because they were not relevant to the study and did not contribute to the analysis. At the same time, columns with codes expressing the same variable were replaced by a single categorical column to solve multicollinearity problems and reduce the complexity of the dataset. These operations helped to improve the quality and relevance of the dataset for the purposes of the study. At the end of the pre-processing phase, the structure of the dataset was as follows:

- Time : timestamp of event
- Day: weekday or weekend
- Moment Day: morning, afternoon, evening or night
- Label
- Activity
- App Used
- Location
- Latitude
- Longitude
- UserID

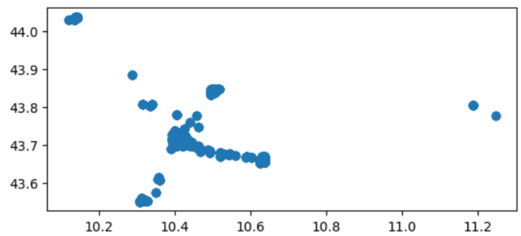
The result of this processing is a single dataset with 10 features and 45,681 observations. Several analyses were performed on this dataset, including exploratory analysis and trajectory extraction, which are discussed in more detail in the next section.

4. Trajectory analysis

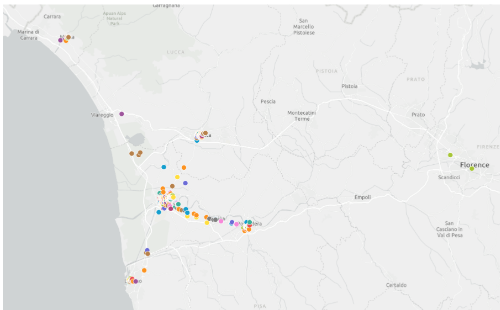
An analysis of the dataset was carried out before looking at the positions of the events. As the variables are all categorical, it was not possible to derive descriptive exploratory variables. However, some variables of particular interest were observed in detail in order to try to obtain information that might be useful in the next phase. For example, the composition of the variables location, application used and activity was observed. The most common activity recorded was 'rec on still', which occurred 41,116 times. Although this is a common value, it provides only a limited indication of actual user activity. This value was not taken into account in the analyses because it expresses an ongoing state of recording and therefore does not indicate the activity performed by the user. Similarly, the most used application is "Android Wear" with 37,428 occurrences, but this value shows that it is probably an application that is constantly working in the background and is therefore not representative of the user's actual activity. The most visited location is "Plaza" with 14,178 values, in this case a value close to that of the other locations, so no anomalies are highlighted, it is simply the most visited location by the three users.

4.1. Trajectory Mining

After a few observations of the data, the positions and trajectories were examined. The positions of the users were first plotted on a Cartesian axis, with the axes corresponding to latitude and longitude (Figure 1a). Using a business intelligence tool [3], it was also possible to display the identified points on a satellite map. (Figure 1b).



(a) Representation of user occupied positions using geographical coordinates

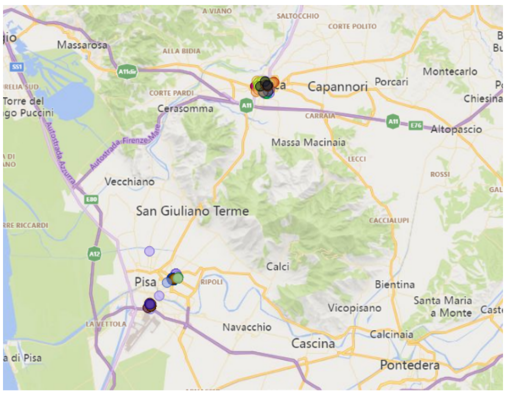


(b) Representation of user occupied positions on a satellite map

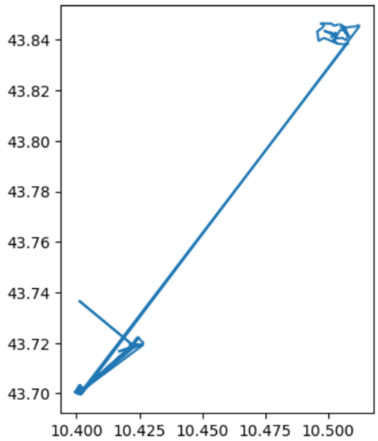
From the collected trajectories, three trajectories could be extracted to identify the different users.

Let's look at the three trajectories in detail:

1. From the analysis of the first trajectory, it is possible to identify a user who makes most of his movements in the first half of the day. They only move between two cities: Pisa and Lucca. In Pisa he stays almost exclusively in places related to the university, while in Lucca he usually spends his free time. In this way, it would be possible to determine whether this user might be interested in an application if we only knew the city where the user is located (Figure 2).



(a) Positions of User 1

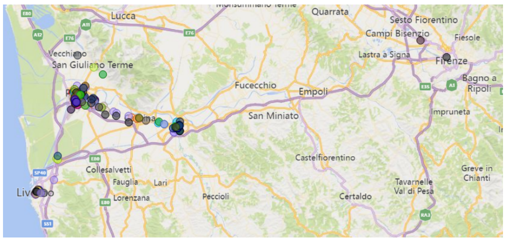


(b) Trajectory of User 1

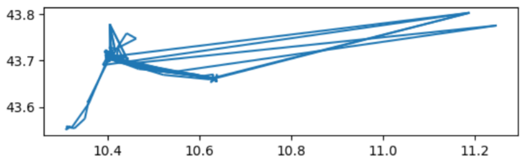
Figure 2: User 1

[3]PowerBI: https://powerbi.microsoft.com/it-it/

2. On the contrary, from the second trajectory we can deduce that the user moves almost exclusively in the evening. His activities are concentrated exclusively in Pisa, all over the city. This suggests that he is not a commuter like user 1, or that he carries out all kinds of activities in the same city anyway. In this case, in order to know what activities the smartphone might be interested in, it is not enough to know only the city, as in the previous case, but also the place where it is located (Figure 3).



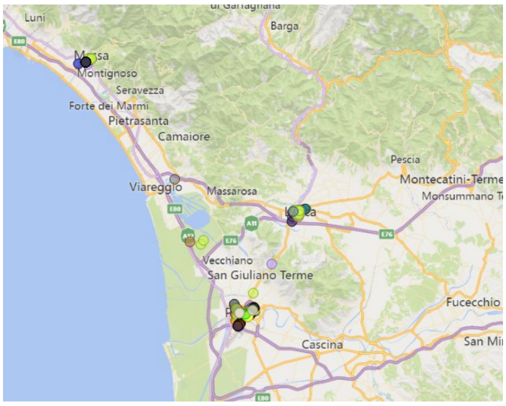
(a) Positions of User 2



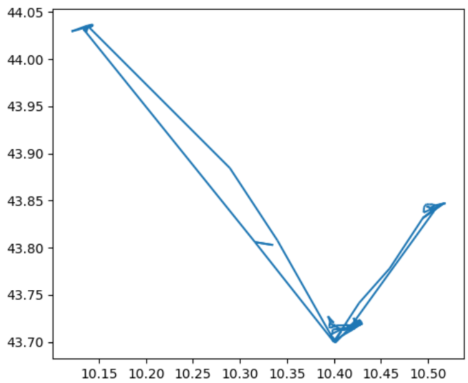
(b) Trajectory of User 2

Figure 3: User 2

3. The third graph shows that most of the positions were carried out in places related to the university. Almost all the actions take place in Pisa and only a few positions refer to two other cities, one of them in a football stadium. This trajectory could indicate that the user is a student or an employee of the university and that one of his interests is football (Figure 4).



(a) Positions of User 3



(b) Trajectory of User 3

Figure 4: User 3

The information obtained from the study of trajectories is just some of the possible information that can be extracted, knowing the positions occupied by the users. However, this interesting information can also be used in business management reports. Let's see in the next section how this information, but in general how trajectory mining can be useful when used to support process mining.

5. Process discovery and analysis

In this section, the technique of process mining is used to discover processes to understand the interaction of users in different locations [28]. The aim of this section is to use the information about user trajectories obtained from trajectory mining to plan the exploration of the process [29].

The choice of the tool is based on the Gartner classification of process mining tools. The chosen tool is Celonis Intelligent Business Cloud [4] - Academic Edition.

5.1. Process discovery

Based on the previous analysis, the dataset is suitable for process mining applications. In fact, the dataset contains all the necessary information for process discovery: caseID, timestamp and activity. In addition, trajectory mining provided important results on user habits, which were used to structure the process discovery analysis. To gain information about how users interact with mobile phone applications, information about popular locations was used to identify trajectories.

This way, the analysis of the process could be structured based on the following attributes that make up the event:

- CaseID: The available caseID is the 'UserID'. Since the analysis is performed on the 'Location' attribute, a unique caseID was created for 'Location'.
- Activity: "Activity Recognition" is available, but the goal of the analysis is to extract the user's interaction with cell phone applications, so "Running Applications" is the activity of interest. In this context the activity is related to the chosen application. The set of activities gives rise to the process of user behavior.
- Timestamp: available

Given the information obtained about the different habits of each user, the analysis is performed for each user individually. Using the new data set, including the case ID of interest, which is the "location" attribute for extracting the process, the analysis is performed using the Celonis "Process Analytics" feature in the "Studio" area, with the great help of the Process Explorer and Variant Explorer functions.

The scope of the analysis is to identify different process variations for different locations, i.e. to identify user behaviour when switching from one application to another based on location, to understand how much location influences user behaviour. The analysis is carried out for one user at a time in order to exclude the influence of the user's personal lifestyle. In the next section, we show the results of the analysis of each user's behaviour taking into account the location with the caseID selection and the apps used.

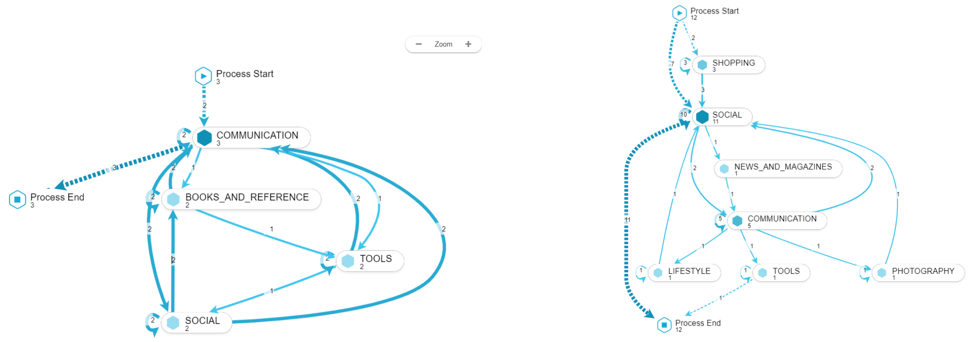
[4] https://www.celonis.com/academic-signup

5.2. Results

We have good reason to believe that the Android Wear app is opened in the background and therefore not linked to user behaviour. To perform a more realistic analysis, the relevant event logs were extracted from the dataset. By analysing each user with Celonis, we discovered new information about each user's behaviour.
In Celonis it is possible to extract the process from the Variant Explorer area and the Process Explorer area. The variant explorer shows all the variants of the processes, where each variant corresponds to each caseID, so in our study each variant is the application selection process for each site; if this process is the same for several sites, we have a common process variant that covers more than one site (the caseID). The process explorer is different, it shows the most common activities and links across all sites.
Since we are interested in using the information about users' movements obtained through Trajectory Mining, we are not interested in knowing the exact process variations at the selected locations, but extracting the process performed by the Process Explorer can provide us with the necessary information about activity and connection frequency, filtered by location. In the following items is explored the order of the activities by showing the start and the end activities in the process.

- User 1: There are 15 cases for this user. According to the results of the trajectory (Section 4.1), in Pisa they almost exclusively visit places related to the university, while in Lucca they usually spend their free time. In the analysis of the application used, among the most frequent places were selected those related to the university, i.e. "College Camp; University", "College Academic Building" and "College Lab". The discovery of this process (Figure 5a) shows the most frequently used applications in these locations. We can see that all 3 cases start with "Communication" and end with the same application. Based on the process, we can see that after "Communication", the user in the selected location also selects "Book and reference", which we expect not to appear in the discovery of the apps used when they spend their free time. Excluding the 3 cases related to the university, we also obtained the process for the leisure app (Figure 5b). In this new process, the most common path is the one that starts with "Social" and ends with the same activity. Also, as expected, after selecting "Communication", the cases go to "Lifestyle", "Photography" and "Shopping", which were not present before, and we do not find "Book and Reference".

- User 2: There are 50 cases for this user, which means that he visits more places compared to the other users. According to the results of the trajectory (Section 4.1) the activities are concentrated exclusively in Pisa, in the whole city. Unlike the first and third users, they do not divide their day between university and leisure, so we cannot assume that their behaviour changes according to their location. In addition, the most frequented places are "Plaza", "Museum" and many restaurants, but also places related to the University. Understanding this user behaviour is more difficult because we cannot know if they are in their free time or not, so we decided to extract the whole process and make assumptions. As shown in Figure 6, the most common path is the one that leads from "Communication" to "Social". However, if we take a closer look at the process, we can see that "Game Card",



(a) Used app at university (b) Used app in free time

Figure 5: Process discovery for User 1

an application that is rare among users, is the most common application on this path after "Communication" and could provide information about the user. Looking at the cases that flow through "Game Card", we see that the user often visits places related to sports, such as "Football Soccer" and "Bowling Alley", which gives us information about his habits and preferences.

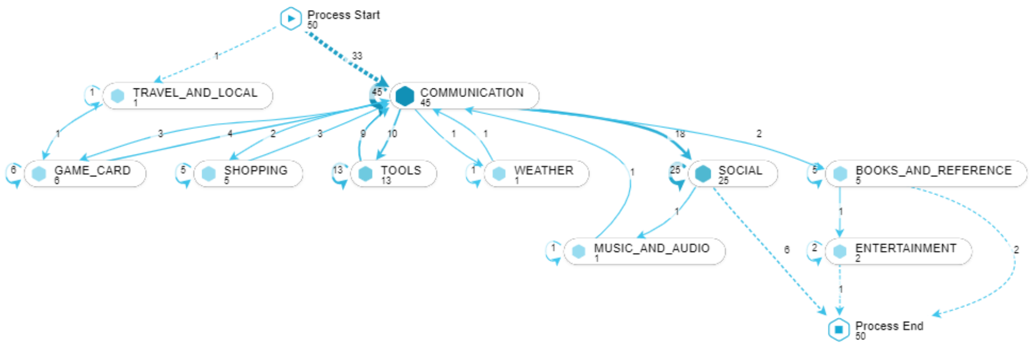
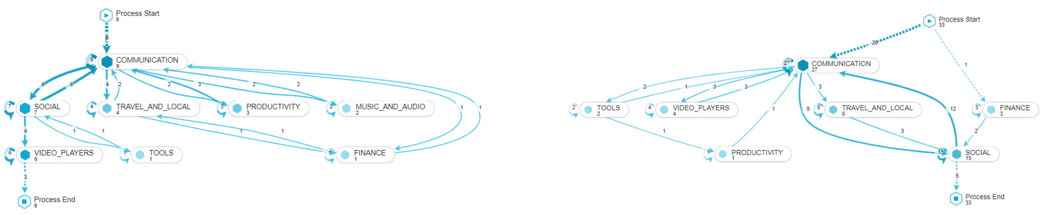


Figure 6: Process discovery for User 2

- User 3: There are 42 cases for this user. A previous analysis (Section 4.1) has shown that most of the positions were carried out in places related to the university and in Pisa, and a few positions refer to two other cities. The process extraction is done, as for user 1, by splitting the cases into those that are related to the university and those that are not. The process extraction (Figure 7) showed that this user's behaviour is quite similar both in discovering processes at the university and in discovering processes elsewhere. The similar behaviour of this user in the two selected processes suggests that his interests, which are more related to "video player", "travel and local" and "music and audio", do not depend on the locations where he spends time, as they do not provide enough information to analyse his interests. We could assume that this user creates content for social applications.



(a) Used app at university

(b) Used app in free time

Figure 7: Process discovery for User 3

To summarise, in this section we have used information from Section 4.1 to carry out the analysis of user behaviour. While for user1 and user3 the analysis of locations was very useful to structure the process extraction, for user2 the analysis of the process was to provide information about locations. In the future, this analysis will also take into account the time frames to give precise information on the process taking into account the place visited with temporal continuity.

6. Conclusion

In conclusion, this study successfully achieved its goal of understanding user behaviour through process analysis of three users' app usage and switching between different apps. Using event log information and location data, we applied process mining and trajectory mining techniques to extract and analyse the process. Our analysis provided insights into user interactions and choices regarding app usage, and how location and trajectory information influenced these choices. The results of this study have important implications for both academic research and practical applications in the field of user behaviour analysis. Our findings show that trajectory mining can effectively support process mining and improve the effectiveness and efficiency of the process. This suggests that incorporating trajectory mining into process analysis can provide a more comprehensive understanding of user behaviour. Future research could expand the dataset and explore other factors that may influence app usage, such as time of day or user demographics. In addition, incorporating machine learning algorithms could further improve the accuracy of the analysis and provide deeper insights into user behaviour. Overall, this study contributes to our understanding of user behaviour through process analysis and highlights the importance of incorporating location and trajectory data into this analysis.

Acknowledgments

Work supported by the project "MOD-UPP" - Macroarea 4 - project PON_MDG_1.4.1_17- PON GOV grant.
We acknowledge financial support from the project PNRR MUR project PE0000013-FAIR.

References

[1] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers, Introduction to Business Process Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 2018, pp. 1–33.
[2] W. Van Der Aalst, Process mining: data science in action, volume 2, Springer, 2016.
[3] Z. Feng, Y. Zhu, A survey on trajectory data mining: Techniques and applications, IEEE Access 4 (2016) 2056–2067.
[4] J. D. Mazimpaka, S. Timpf, Trajectory data mining: A review of methods and applications, Journal of spatial information science 2016 (2016) 61–99.
[5] J.-Y. Kang, H.-S. Yong, Mining spatio-temporal patterns in trajectory data, Journal of Information Processing Systems 6 (2010) 521–536.
[6] G. Costa, G. Manco, E. Masciari, Dealing with trajectory streams by clustering and mathematical transforms, Journal of Intelligent Information Systems 42 (2014) 155–177.
[7] F. Giannotti, M. Nanni, F. Pinelli, D. Pedreschi, Trajectory pattern mining, in: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 2007, pp. 330–339.
[8] E. Masciari, G. Shi, C. Zaniolo, Sequential pattern mining from trajectory data, in: Proceedings of the 17th International Database Engineering & Applications Symposium, 2013, pp. 162–167.
[9] E. Masciari, Trajectory clustering via effective partitioning, in: Flexible Query Answering Systems: 8th International Conference, FQAS 2009, Roskilde, Denmark, October 26-28, 2009. Proceedings 8, Springer, 2009, pp. 358–370.
[10] J. Bian, D. Tian, Y. Tang, D. Tao, A survey on trajectory clustering analysis, arXiv preprint arXiv:1802.06971 (2018).
[11] A. S. Junior, V. C. Times, C. Renso, S. Matwin, L. A. Cabral, A semi-supervised approach for the semantic segmentation of trajectories, in: 2018 19th IEEE International Conference on Mobile Data Management (MDM), IEEE, 2018, pp. 145–154.
[12] H. Jeung, M. L. Yiu, C. S. Jensen, Trajectory pattern mining, in: Computing with spatial trajectories, Springer, 2011, pp. 143–177.
[13] K. Xie, K. Deng, X. Zhou, From trajectories to activities: a spatio-temporal join approach, in: Proceedings of the 2009 International Workshop on Location Based Social Networks, 2009, pp. 25–32.
[14] C. L. da Silva, L. M. Petry, V. Bogorny, A survey and comparison of trajectory classification methods, in: 2019 8th Brazilian Conference on Intelligent Systems (BRACIS), IEEE, 2019, pp. 788–793.
[15] A. De Leege, M. van Paassen, M. Mulder, A machine learning approach to trajectory prediction, in: AIAA Guidance, Navigation, and Control (GNC) Conference, 2013, p. 4782.

[16] T. Li, J. Wu, A. Dang, L. Liao, M. Xu, Emission pattern mining based on taxi trajectory data in beijing, Journal of cleaner production 206 (2019) 688–700.
[17] S. Tsumoto, S. Hirano, Detection of risk factors using trajectory mining, Journal of Intelligent Information Systems 36 (2011) 403–425.
[18] R. Talat, M. S. Obaidat, M. Muzammal, A. H. Sodhro, Z. Luo, S. Pirbhulal, A decentralised approach to privacy preserving trajectory mining, Future generation computer systems 102 (2020) 382–392.
[19] S. Tsumoto, S. Hirano, Behavior grouping based on trajectory mining, in: Social Computing and Behavioral Modeling, Springer, 2009, pp. 1–8.
[20] E. V. Napolitano, Intelligent technologies for urban progress: Exploring the role of ai and advanced telecommunications in smart city evolution, in: A. Abelló, P. Vassiliadis, O. Romero, R. Wrembel, F. Bugiotti, J. Gamper, G. Vargas Solar, E. Zumpano (Eds.), New Trends in Database and Information Systems, Springer Nature Switzerland, Cham, 2023, pp. 676–683.
[21] T. Tani, H.-H. Huang, K. Kawagoe, Sports play visualization system using trajectory mining method, Procedia Technology 18 (2014) 100–103.
[22] F. Mannhardt, D. Blinde, Analyzing the trajectories of patients with sepsis using process mining., RADAR+ EMISA@ CAiSE 1859 (2017) 72–80.
[23] E. Tavazzi, R. Gatta, M. Vallati, S. Cotti Piccinelli, M. Filosto, A. Padovani, M. Castellano, B. Di Camillo, Leveraging process mining for modeling progression trajectories in amyotrophic lateral sclerosis, BMC Medical Informatics and Decision Making 22 (2022) 1–17.
[24] A. Peck, S. Provost, L. East, M. Hutchinson, Process mining the trajectories for adolescent-to-mother violence from longitudinal police and health service data, Journal of Advanced Nursing (2022).
[25] M. Pegoraro, M. S. Uysal, T.-H. Hülsmann, W. M. van der Aalst, Uncertain case identifiers in process mining: A user study of the event-case correlation problem on click data, in: Enterprise, Business-Process and Information Systems Modeling: 23rd International Conference, BPMDS 2022 and 27th International Conference, EMMSAD 2022, Held at CAiSE 2022, Leuven, Belgium, June 6–7, 2022, Proceedings, Springer, 2022, pp. 173–187.
[26] M. G. Campana, D. Chatzopoulos, F. Delmastro, P. Hui, Lightweight modeling of user context combining physical and virtual sensor data, in: Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, 2018, pp. 1309–1320.
[27] M. G. Campana, F. Delmastro, Contextlabeler dataset: Physical and virtual sensors data collected from smartphone usage in-the-wild, Data in brief 37 (2021) 107164.
[28] A. Monreale, F. Pinelli, R. Trasarti, F. Giannotti, Wherenext: a location predictor on trajectory pattern mining, in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, 2009, pp. 637–646.
[29] S. Fioretto, Process mining solutions for public administration, in: A. Abelló, P. Vassiliadis, O. Romero, R. Wrembel, F. Bugiotti, J. Gamper, G. Vargas Solar, E. Zumpano (Eds.), New Trends in Database and Information Systems, Springer Nature Switzerland, Cham, 2023, pp. 668–675.