A Joint Analysis of Trajectory Mining and Process Mining for Smartphone User Behaviour

Simona Fioretto[1 [0009-0006-8700-8188]], Elio Masciari[1,2 [0000-0002-1778-5321]], and Enea Vincenzo Napolitano[1 [0000-0002-6384-9891]]

[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 an important area of research. It encompasses the management of both production processes, which involve the creation of products and services, and processes that handle data and information. In today's digital landscape, the business opportunities associated with information management have gained significant prominence due to the widespread use of smartphones, social networks and similar digital tools in everyday life.
Process mining techniques play a crucial role in supporting various stages of process management, enabling the identification of processes and their specifications. This paper explores the application of process mining techniques to smartphone usage data, complemented by trajectory mining techniques. The aim is to investigate whether location-based information derived from smartphones can contribute to process management or, conversely, whether process management can support the use of location-based data.
The results of this research can provide valuable insights for organisations looking to harness the power of digital tools and data-driven approaches to optimise their processes and improve overall performance.

Keywords: Trajectory Mining · Process Mining · Behaviour Analysis.

1 Introduction

A process is commonly defined as a series of activities which transform inputs into outputs, whether sequential or not. Process modelling has proven to be valuable in a wide range of industries, helping in the identification of features and problems in processes, such as delays in manufacturing or service. Business Process Management (BPM) encompasses the phases necessary to manage a process effectively, including discovery, analysis, design/re-design, implementation and control. The discovery and analysis phases are particularly important for understanding the interaction between activities, actors and resources involved. In this sense, the context of user interactions in smartphone usage can be viewed as a process. If the transition from one activity to another is seen as a process, the path of applications used by a user can also be seen as a process. In this

scenario, the user acts as an actor, the applications represent the activities performed, and the user's decision process dictates the flow from one application to another. This paper focuses specifically on the discovery and analysis of processes related to user activities, in particular on the switching activity from one application to another on smartphones by three users.
To extract the process, event logs of smartphone usage were analysed using a process mining tool capable of discovering and analysing the process through different views, key performance indicators (KPIs) and charts. The aim is to use process analysis to understand the user's decision-making process regarding applications, specifically how they choose one activity over another.
Process analysis uses the location variable to gain insight into user behaviour by incorporating both location and trajectory information. Trajectory mining techniques are used to identify and analyse user trajectories based on location information. The aim is to determine whether trajectory mining can support process mining in a better analysis of the process, and so in improving its effectiveness and efficiency.

2 An Overview

The following lines provide a brief overview of current studies on process mining and trajectory mining.

2.1 Process Mining

[4] studies the field of Business Process Management (BPM), highlighting the recurrent phases of managing a process. Ultimately, information technology and BPM have found a union in process mining, a discipline that lies between machine learning and data mining on the one hand, and process modelling and analysis on the other, as shown by [15]. The idea behind process mining is study real-world processes by extracting knowledge from event logs that are readily available in today's systems. Processes are extracted and analysed from the digital footprints of users. There is a detailed description of process mining in [15], which distinguishes three types of process mining:

- **process discovery**: includes an event log and builds a model without using a priori information
- **conformance checking**: compares an existing process model with an event log of the same process.
- **process improvement**: improves an existing process model based on information about the actual process recorded in an event log.

2.2 Trajectory Mining

Trajectory mining is a data mining technique employed to analyze the temporal and spatial motion data of objects or individuals [5]. Trajectories can be expressed as a sequence of spatio-temporal points [8] or as continuous paths in the space-time domain. The primary objective of trajectory mining is to uncover significant patterns, including frequent routes, aberrant behaviors, and mobility trends [6]. Various techniques are encompassed within trajectory mining, such as trajectory clustering [10], trajectory segmentation [7], trajectory pattern mining [6], spatio-temporal analysis of trajectories [16], trajectory classification [13], and trajectory prediction models [3].

Trajectory mining and process mining can be used together, and as the literature shows, this can lead to many benefits. For example, in medicine [9, 14, 11] these joint techniques are used to monitor the spread of a disease. Another application relates to the user behaviour study in [12], who extract information about the user interactions with the use of event-case correlation on click data for a mobility sharing company.

3 The Dataset

The dataset used in this study is the ContextLabeler dataset [1, 2][3]. It includes more than 45,000 data samples in CSV format, each with 1,332 features from various physical and virtual sensors. These sensors include motion sensors, running applications, proximity devices and weather conditions, providing a comprehensive representation of the user's environment. The dataset was gathered over a two-week period by three volunteers using Context Labeler[4], 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 data collection experiment. The labels include user activities such as working, eating and exercising, while the contextual information includes environmental factors such as temperature and humidity. The dataset consists of 45,681 data samples, distributed across the three users as follow: 8,456 samples for user 1, 17,882 samples for user 2 and 19,343 samples for user 3.
The dataset was collected "in-the-wild," meaning that the subjects used their devices without any constraints on their natural behavior.
Several operations were carried out during the data processing phase. First, irrelevant columns were removed from the data set. Then, columns with codes expressing the same variable were replaced by a single categorical column. At the end of the pre-processing phase, the dataset had the following structure:

- Time: Timestamp of the event
- Day: Weekday or weekend
- Moment Day: Morning, afternoon, evening, or night
- Label
- Activity
- App Used

[3] The dataset is available at: https://github.com/contextkit/ContextLabeler-Dataset
[4] https://contextkit.github.io

– Location
– Latitude
– Longitude
– UserID

The result of this pre-processing phase is a unified dataset consisting of 10 features and 45,681 observations. Various analyses were performed on this dataset, including exploratory analysis and trajectory extraction, which will be discussed in more detail in the following section.

4 Application of Trajectory Mining

An initial dataset analysis was conducted before examining the event positions. Since all variables were categorical, it was not possible to derive descriptive exploratory variables. However, certain variables of interest were closely examined to extract information potentially useful for the subsequent phase. Specifically, the composition of location, app usage, and activity variables 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 frequently used app was "Android Wear" with 37,428 occurrences. However, this suggests that it is likely an application consistently running in the background, making it non-representative of the user's true activity. Among the various locations, "Plaza" was the most frequently visited with 14,178 occurrences. This value is similar to that of other locations, indicating no anomalies, but simply indicating that it is the most visited location among the three users.

4.1 Trajectory Mining

After observing the data, the positions and trajectories were examined. First, the user positions were plotted on a Cartesian axis with latitude and longitude as the axes. In addition, a business intelligence tool [5] allowed the points identified to be displayed on a satellite map. From the collected trajectories, three different trajectories were extracted to differentiate between users.
Let's take a closer look at these three trajectories:

1. The analysis of the first trajectory shows a user who moves mainly between the cities of Pisa and Lucca during the first half of the day. In Pisa, he mainly visits places related to the university, while in Lucca he spends his free time. This information allows us to determine whether this user might be interested in an application based solely on their city of residence.

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

2. Conversely, the second trajectory shows that the user's movements are mainly concentrated in the evening. Their activities are exclusively concentrated within the city of Pisa. This suggests that he is either not a commuter like the first user or that he has different activities within the same city. In this case, knowing only the city is not enough to identify the activities that might interest the user; the specific location within the city is also crucial.
3. The third trajectory shows that most of the positions are associated with university places in Pisa. Almost all the actions take place in Pisa, with only a few positions recorded in two other cities, one of which includes a football stadium. This trajectory may indicate that the user is either a student or a university employee with a potential interest in football.

The information derived from trajectory analysis represents only a fraction of the insights that can be extracted from user positions. Nevertheless, this valuable information can be used in business management reports. In the next section we will explore how this information, and trajectory mining in general, can be useful in supporting process mining.

5 Application of Process Mining

In this section, process mining is used to discover and understand user interactions in different locations. The aim is to use the trajectory information obtained from trajectory mining to plan the process exploration.

5.1 Process Discovery

Based on the Gartner classification of process mining tools, the chosen tool for this study is the Celonis Intelligent Business Cloud [6] - Academic Edition. Analysis of the dataset confirms its suitability for process mining applications. The necessary information for process discovery, including caseID, timestamp and activity, is present in the records. The trajectory mining results provide insight into user habits and are used to structure the process discovery analysis. To understand user interaction with mobile phone applications, popular locations are used to identify trajectories. The analysis of the process is structured around the following attributes of the events:

- CaseID: The 'UserID' is the available caseID, however since the analysis focuses on the 'Location', a unique caseID is created for each location.
- Timestamp: Available in the dataset.
- Activity: The activities of interest are related to "Running Applications" attribute, as it provides insights into user interactions with cell phone apps.

Based on the information gathered about each user's unique habits, the analysis is performed individually for each user. The Celonis Process Analytics feature,

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

specifically the Process Explorer and Variant Explorer functions, is used for the analysis. The analysis focuses on identifying different process variations for different locations to understand how location influences user behaviour. Each user is analysed separately to isolate the influences of their personal lifestyle.

5.2 Results

For a more realistic analysis, the Android Wear app, which is likely to be running in the background and unrelated to user behaviour, is excluded. The relevant event logs are extracted from the dataset for analysis with Celonis. Each user's behaviour is analysed using the Variant Explorer and Process Explorer functionalities of Celonis.
The Variant Explorer displays all process variants, with each variant corresponding to a specific CaseID. In this study, each variant represents the application selection process in each location site. If the process is the same for multiple sites, a common process variant covering more than one site (caseID) is identified. On the other hand, the Process Explorer shows the most common activities and links across all sites.

1. For User1 the analysis shows 15 cases. Based on the trajectory analysis described in section 4.1, User1 mainly visits university-related places in Pisa and spends free time in Lucca. The analysis focuses on the most frequently visited university-related places, such as "College Camp; University", "College Academic Building" and "College Lab". The discovered process shows the most frequently used applications in these locations. The process shows that all three cases start with 'Communication' and end with the same app. However, after 'Communication', the process shows that the user also selects 'Books and Reference', which is not expected in his free time. By excluding the university related cases, a separate process for leisure apps is extracted. In this new process, the most common path starts with 'Social' and ends with the same activity. In addition, after selecting 'Communication', the cases lead to 'Lifestyle', 'Photography' and 'Shopping', which were not present in the university-related cases.
2. User2 has 50 cases, indicating visits to different locations compared to other users. According to the trajectory analysis in section 4.1, User2's activities are concentrated in the city of Pisa. Unlike User1 and User3, User2 does not divide his day into university and leisure time, which makes it difficult to determine behavioural changes based on location. The most frequented places are the "Plaza", the "Museum" and numerous gastronomic establishments, as well as places related to the University. To understand User2's behaviour, the whole process is extracted and assumptions are made. The most common path is from "Communication" to "Social". Further analysis of the process shows that the application "Game Card", which is rare among users, is the most common after "Communication" for User2. Examining the cases with "Game Card" shows that User2 often visits sports related places like "Football Soccer" and "Bowling Alley", which gives insight into his habits and preferences.

3. User3 has 42 cases, and the trajectory analysis in section 4.1 shows that most positions are related to the University of Pisa, with a few positions in other cities. Similarly to User1, the process discovery is performed separately for university related cases and non-university related cases. The analysis shows that User3's behaviour is quite similar in both processes, suggesting that her interests, such as 'video player', 'travel and local' and 'music and audio', are not influenced by location. These interests do not provide enough information to analyse their preferences. It is possible that User3 creates content for social applications.

In summary, this section uses trajectory mining information (see section 4.1) to analyse user behaviour. While for User1 and User3 the location analysis was crucial for structuring the process extraction, for User2 the process analysis provided insight into the locations.

6 Conclusion

In conclusion, the primary objective of this study, which was to gain insight into user behaviour by analysing the process of app usage and switching between different apps for three users, was successfully achieved. Using event log information and location data, we applied trajectory mining and process mining techniques to extract and analyse the underlying process. Our analysis provided valuable insights into user interactions, choices and preferences regarding app usage, as well as the influence of location and trajectory information on these choices. With the results obtained with trajectory mining, the analysis done with process mining was better structured and gained wider information on the motivation of user behaviour. Our findings clearly demonstrate that trajectory mining can effectively support process mining and contribute to improving the overall effectiveness and efficiency of the process. This suggests that incorporating trajectory mining into process analysis allows for a more comprehensive understanding of user behaviour. Future research efforts could include expanding the dataset used in this study and exploring additional factors that may influence app usage, such as time of day or user demographics. In addition, incorporating machine learning algorithms has the potential to further improve the accuracy of the analysis and provide deeper insights into user behaviour patterns. Through the successful achievement of our research objectives, we have opened up avenues for the further exploration and refinement of techniques for the analysis of user behaviour.

References

1. Campana, M.G., Chatzopoulos, D., Delmastro, F., Hui, P.: 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. pp. 1309–1320 (2018)

2. Campana, M.G., Delmastro, F.: Contextlabeler dataset: Physical and virtual sensors data collected from smartphone usage in-the-wild. Data in brief **37**, 107164 (2021)
3. De Leege, A., van Paassen, M., Mulder, M.: A machine learning approach to trajectory prediction. In: AIAA Guidance, Navigation, and Control (GNC) Conference. p. 4782 (2013)
4. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Introduction to Business Process Management, pp. 1–33. Springer Berlin Heidelberg, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-662-56509-4_1, https://doi.org/10.1007/978-3-662-56509-4_1
5. Feng, Z., Zhu, Y.: A survey on trajectory data mining: Techniques and applications. IEEE Access **4**, 2056–2067 (2016)
6. Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D.: Trajectory pattern mining. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 330–339 (2007)
7. Junior, A.S., Times, V.C., Renso, C., Matwin, S., Cabral, L.A.: A semi-supervised approach for the semantic segmentation of trajectories. In: 2018 19th IEEE International Conference on Mobile Data Management (MDM). pp. 145–154. IEEE (2018)
8. Kang, J.Y., Yong, H.S.: Mining spatio-temporal patterns in trajectory data. Journal of Information Processing Systems **6**(4), 521–536 (2010)
9. Mannhardt, F., Blinde, D.: Analyzing the trajectories of patients with sepsis using process mining. RADAR+ EMISA@ CAiSE **1859**, 72–80 (2017)
10. Masciari, E.: Trajectory clustering via effective partitioning. In: Flexible Query Answering Systems: 8th International Conference, FQAS 2009, Roskilde, Denmark, October 26-28, 2009. Proceedings 8. pp. 358–370. Springer (2009)
11. Peck, A., Provost, S., East, L., Hutchinson, M.: Process mining the trajectories for adolescent-to-mother violence from longitudinal police and health service data. Journal of Advanced Nursing (2022)
12. Pegoraro, M., Uysal, M.S., Hülsmann, T.H., van der Aalst, W.M.: 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. pp. 173–187. Springer (2022)
13. da Silva, C.L., Petry, L.M., Bogorny, V.: A survey and comparison of trajectory classification methods. In: 2019 8th Brazilian Conference on Intelligent Systems (BRACIS). pp. 788–793. IEEE (2019)
14. Tavazzi, E., Gatta, R., Vallati, M., Cotti Piccinelli, S., Filosto, M., Padovani, A., Castellano, M., Di Camillo, B.: Leveraging process mining for modeling progression trajectories in amyotrophic lateral sclerosis. BMC Medical Informatics and Decision Making **22**(6), 1–17 (2022)
15. Van Der Aalst, W.: Process mining: data science in action, vol. 2. Springer (2016)
16. Xie, K., Deng, K., Zhou, X.: From trajectories to activities: a spatio-temporal join approach. In: Proceedings of the 2009 International Workshop on Location Based Social Networks. pp. 25–32 (2009)