



HAL
open science

Interpretation of Patients' Location Data to Support the Application of Process Mining Notations

Sina Araghi, Franck Fontanili, Elyes Lamine, Nicolas Salatge, Frederick Benaben

► **To cite this version:**

Sina Araghi, Franck Fontanili, Elyes Lamine, Nicolas Salatge, Frederick Benaben. Interpretation of Patients' Location Data to Support the Application of Process Mining Notations. HEALTHINF 2020 - 13th International Conference on Health Informatics, Feb 2020, La Valette, Malta. pp.472-481, 10.5220/0008971104720481 . hal-02521655

HAL Id: hal-02521655

<https://imt-mines-albi.hal.science/hal-02521655v1>

Submitted on 27 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interpretation of Patients' Location Data to Support the Application of Process Mining Notations

Sina Namaki Araghi, Franck Fontanili, Elyes Lamine, Nicolas Salatge and Frederick Benaben
IMT Mines Albi, France

Keywords: Model-driven Engineering, Process Mining, Indoor Localization Systems (ILS), Patient Pathways.

Abstract: The application of indoor localization and process mining emerges as an intriguing tool for the researchers to address the structural issues related to the patient pathways inside healthcare organizations. However, there is a major gap in the literature. This is related to the lack of enough attention to the interpretation of location data. Therefore, as a contribution, this article presents the DIAG meta-model and relevant location data interpretation rules. This model-driven approach has been realized in the context of the R-IOSUITE application and it supports the further analyses by the process mining methods.

1 INTRODUCTION

The so-called "no-shows" or "missed appointments" concept is a costly defect for patients and healthcare organization. According to the article published in (Anisi et al., 2018), the ratio of the missed appointments is about 34 % only in the USA, and each missed appointment costs about 100 \$. One of the major causes behind this issue is the **structural problems of the healthcare processes**.

Recently, the **application of Indoor Localization Systems (ILS) and Process Mining** is proposed in several research works to address such issues and to improve the outcome of patient pathways (Kamel Boulos and Berry, 2012; Martin, 2019; Rojas et al., 2017b; Martinez-Millana et al., 2019).

In line with these applications, this research work highlights a major concern in this area. According to the analysis of the literature provided in this article, most of research works focused on the "knowledge-extraction" approaches by using location data of the tagged objects. On the other hand, the **interpretation of location data** or as called here the "sense-making procedures" are completely ignored.

Therefore, this research work aims to propose a model-driven engineering approach to tackle this issue. As a result, the contribution of this article is oriented towards the **DIAG meta-model** which supports the **interpretation of the location data** that can be performed by two groups of **interpretation rules**.

The used methodology in this article is shown in figure 1. This methodology which has been devel-

oped previously in (Namaki Araghi et al., 2018b; Namaki Araghi et al., 2019) consists of several functions to extract meaningful information by using indoor localization systems in association with process mining. These functions are: *configuring the environment and systems*, *Location data gathering*, *Location data interpreting*, *Business process modeling*, and *Business process analyzing and diagnosing*. The focus of this paper is mainly on the first and third functions. Within these functions, the contributions of this article prepare a solid foundation prior to applying process mining notations.

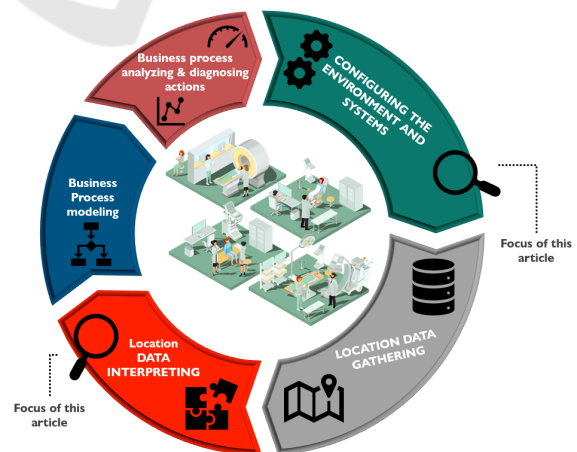


Figure 1: The used approach in this research work.

Consequently, the next section of this article presents the analysis of the literature. The third sec-

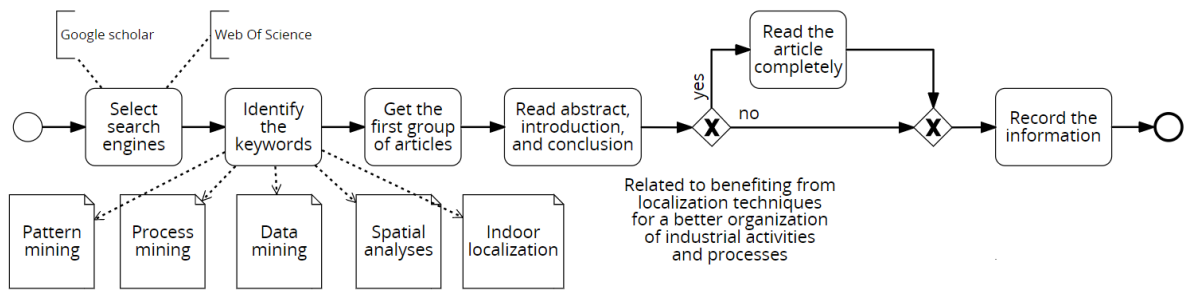


Figure 2: The used approach for analyzing research works related to applying localization techniques for a better organization of processes and activities in different industries.

tion focuses in presenting the DIAG meta-model and the location data interpretation rules. The fourth section presents a series of experimental results. Finally, the last section concludes the article.

2 BACKGROUND

This section presents a systematic literature and technological review in a five-year (2015- June 2019) period. The objective for choosing this period was to recognize the recent trends in the field of applying location data for data and process mining. The main focus here was to analyze the existing works that addressed the trajectory pattern mining regardless of a specific industry. These works were selected due to the nearness of the presented methods to the applied algorithms for discovery of the end-to-end process-like patterns.

Accordingly, Google scholar and Web Of Science were the two search engines used for this purpose. Several keywords were considered such as *process mining*, *location data*, *indoor location/localization systems*. As a result, 42 research works were recorded. We tried to ensure about the proximity of the obtained works to the purposes of this research work. Therefore, by skimming the abstract, conclusion and introduction of these works, 31 out of 42 were selected for final literature analysis. Table 1 provides an overview of the cited works. Figure 2 shows the approach for analyzing the literature related to the works that have applied localization techniques.

As mentioned previously, the objective was to capture an overall image of the recent trends for using location data in different industries. Notably, two criteria were used; (i) works related to the **interpretation of location data**, and (ii) works which addressed the **knowledge-extraction** based on location data.

In some cases researchers worked on several of the mentioned criteria. For instance the work in (Senderovich et al., 2016; Wan et al., 2017; Muza-

mmal et al., 2018), it addresses both mining process-like patterns and how to interpret the location data.

According to the presented analysis in figure 3, the recent trend is mainly on **the extraction of knowledge from location data**. On the other hand, despite its importance, the **interpretation and preparation of location data were almost completely neglected**.

The following will present briefly some of the significant information extracted from these works.

2.1 Preparation and Interpretation of Location Data

Despite the lack of attention for this notion, several robust works proposed different methods to interpret the location data prior to the knowledge discovery. In (Senderovich et al., 2016) the authors used a so-called interaction mining to transform the real-time location data into standard event logs for enabling process mining activities. The transformation of event streams was according to the defined notion of interaction. This interaction provides a knowledge layer that links the raw sensor data and process instances.

In (Muzammal et al., 2018), the authors have used the uncertain sensor data and tried to transform these data into probabilistic trajectory data using pre-processing routines. Then, they have used a dynamic algorithmic approach to extract the preferable trajectories. Their approach consists of 3 main steps: *data cleaning*, *data compression* and finally *trajectory mining*. The trajectory creation step which generates togetherness pattern, common path pattern and group patterns and cyclic patterns.

The work in (Wan et al., 2017) presents a framework for deriving information about people continuous activities from their individual GPS data. The authors have proposed a framework which includes two major techniques for processing GPS data: (i) a fuzzy classification for differentiating the activity patterns. (ii) a scale- adaptive method for refining the location of activities within outdoor and indoor environments.

Table 1: Cited research works that addressed the knowledge extraction from location data and those which addressed interpretation of these data.

Cited research works	
location event log preparation and interpretation	(Senderovich et al., 2016; Muzammal et al., 2018; Wan et al., 2017; Namaki Araghi et al., 2018a)
Use of location data for extraction of knowledge	(Hwang and Jang, 2017; Szt Tyler et al., 2016; Ertek et al., 2017; Mazimpaka and Timpf, 2016), (Rojas et al., 2017a; Zheng, 2015a; Tanuja and Govindarajulu, 2017; Ramos et al., 2017; Garaeva et al., 2017; Bao and Wang, 2017), (Zhenjiang et al., 2017; Aryal and Sujing Wang, 2017; Tanuja and Govindarajulu, 2016; Feng Ling et al., 2016), (Lamr and Skrbek, 2016; Blank et al., 2016; Fernandez-Llatas et al., 2015; Zheng, 2015b), (Liao et al., 2015; Tang et al., 2015; Miclo et al., 2015; Jin et al., 2015; Martinez-Millana et al., 2019), (Dogan et al., 2019; Namaki Araghi et al., 2019; Namaki Araghi et al., 2018a; Namaki Araghi et al., 2018b; Araghi et al., 2018)

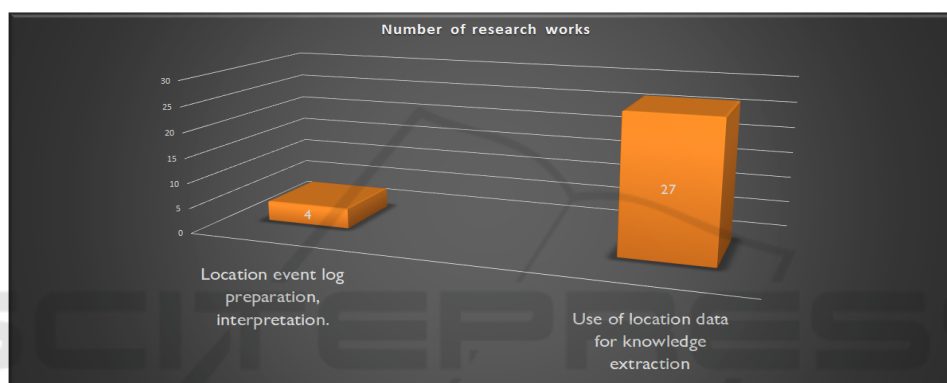


Figure 3: Analysis of the literature of process mining, indoor localization systems, and data mining relevant to the approaches in which the location data were used.

The clustering techniques used in their work helped to classify which activities are occurring. It helped also to detect the outliers and indicated in which positions an activity is being executed.

The next section will highlight how the review of these previous works gave the direction for the proposed solutions in this article.

2.2 A Decision based on the Literature

The processing of location data is not only a technical challenge, but it can be seen as a scientific challenge as well. Due to the existing high ambiguity in location data, it is not feasible to extract proper information from event logs. This ambiguity can be caused by the way these data are being registered.

These data are recorded by **coordination**. They are not representing activities and they only illustrate the positions of tags. Therefore, researchers need to address how one can interpret this type of data. Simply put, the **sense-making procedure is missing in the literature**. Some researchers insisted on

this important gap, and indicated the need to provide an abstraction of what exists in the location data (Senderovich et al., 2016).

In line with this issue, the next section will focus on presenting the **DIAG meta-model** and the location data interpretation rules.

3 METHOD

Assume that a location event log similar to the one in figure 4 is given for further analyses. By focusing on such event log, one could observe that the given data is vague. This means it is not clear which recorded information belongs to the patients, which belongs to the healthcare staff, and what is the corresponding activity. In that, there exist many other blurred information that should be detected. In fact, there is a considerable gap to link the actual concepts in patient—pathways—processes with the registered data in the location event logs. Figure 4 embodies this gap be-

Sense-making is required to link the information in event logs with the actual concepts in the environment.

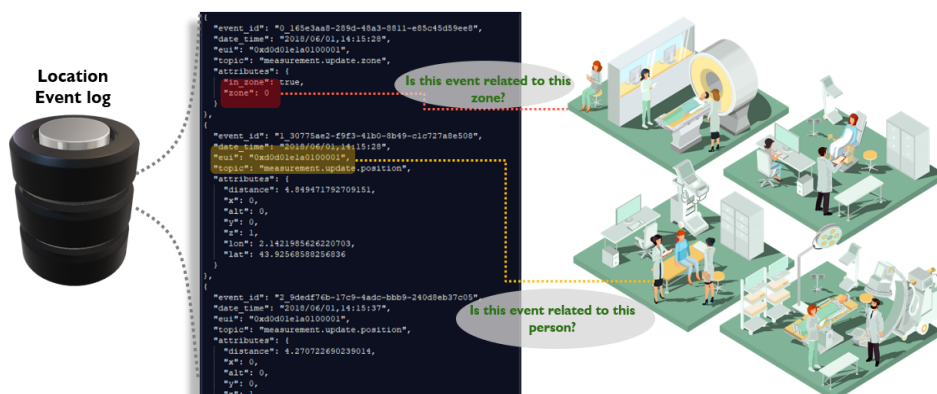


Figure 4: Illustrating the need to provide a sense-making procedure for the recorded information in location event logs.

tween raw data and the actual concepts in patient processes.

3.1 DIAG Meta-model

To fill this gap, the DIAG meta-model is presented in figure 5. This meta-model consists of seven different packages. Each package contains several classes that help to create a model of the link between the actual concepts in the environment and the raw data in location event logs. As shown in figure 5, (i) **Healthcare resources** package includes the main resources used in patient processes that can be detected and monitored by location tags. (ii) The **Organization** package which represents the main classes that are linked to the healthcare organization. (iii) **Location event logs** package provides a model of all the information that exist within a location event log and their relationship with other classes. (iv) The **Objective** package consists of necessary classes to perform business process analyses on top of the registered information. (v) The **Process** package models the concepts that are considered to represent a patient pathways as a business process model. (vi) The **Function** package is addressing the different classes that lead to the execution of activities in processes. (vii) The **Healthcare functions** package addresses different types of actions that can take place during the run-time of patient pathways. (viii) The **Context** package helps to provide a base for further business process diagnosing actions.

The DIAG meta-model has been realized and developed thanks to a web-based application known as **R IOSUITE**. This multidisciplinary framework includes multiple modules (R-IODA, R-IOSEMIT, R-IOGA and etc.) to obtain the location event logs and extract the required information. <https://research-gi.mines-albi.fr/display/RIOSUITE/R-IOSuite+Home>

Evidently, the detailed presentation of the packages and their included classes are out of the limit of this paper. However, some examples illustrate the applicability of this meta-model. Figure 6 shows how the user can model engaged healthcare resources in patient pathways by the application. As shown within figure 6, a hospital is addressed by the **Organization** package and it can have different **Healthcare resources** such as *healthcare staff* and other resources like the hospital zones. Modeling these concepts within the application provides a-priori knowledge for interpreting the information within the location event logs.

3.2 Location Data Interpretation Rules

After gathering the primary location event logs, they will be uploaded to each defined "location tag" class in the "resource" tab of the application (c.f. figure 6). As shown in figure 1, it is necessary to interpret the location event logs in order to be used for the process modeling function.

In line with this objective, two groups of location interpretation rules are defined.

- First group: This set is constructed by three rules. (i) **Add start-event**, (ii) **Add end-event**, and (iii) **Add task-events**. They are in charge of transforming the very first visualization of data as "nodes" with out any semantics. These graphs will be sent to a graph database management system. For example, outputs of this step will be series of nodes labeled as "Activity.Inside.Zone #".
- Second group: After detecting the nodes, the second group of rules will start retracting the modeled a-priori knowledge within the application. These rules are: (i) **Add knowledge onto start-**

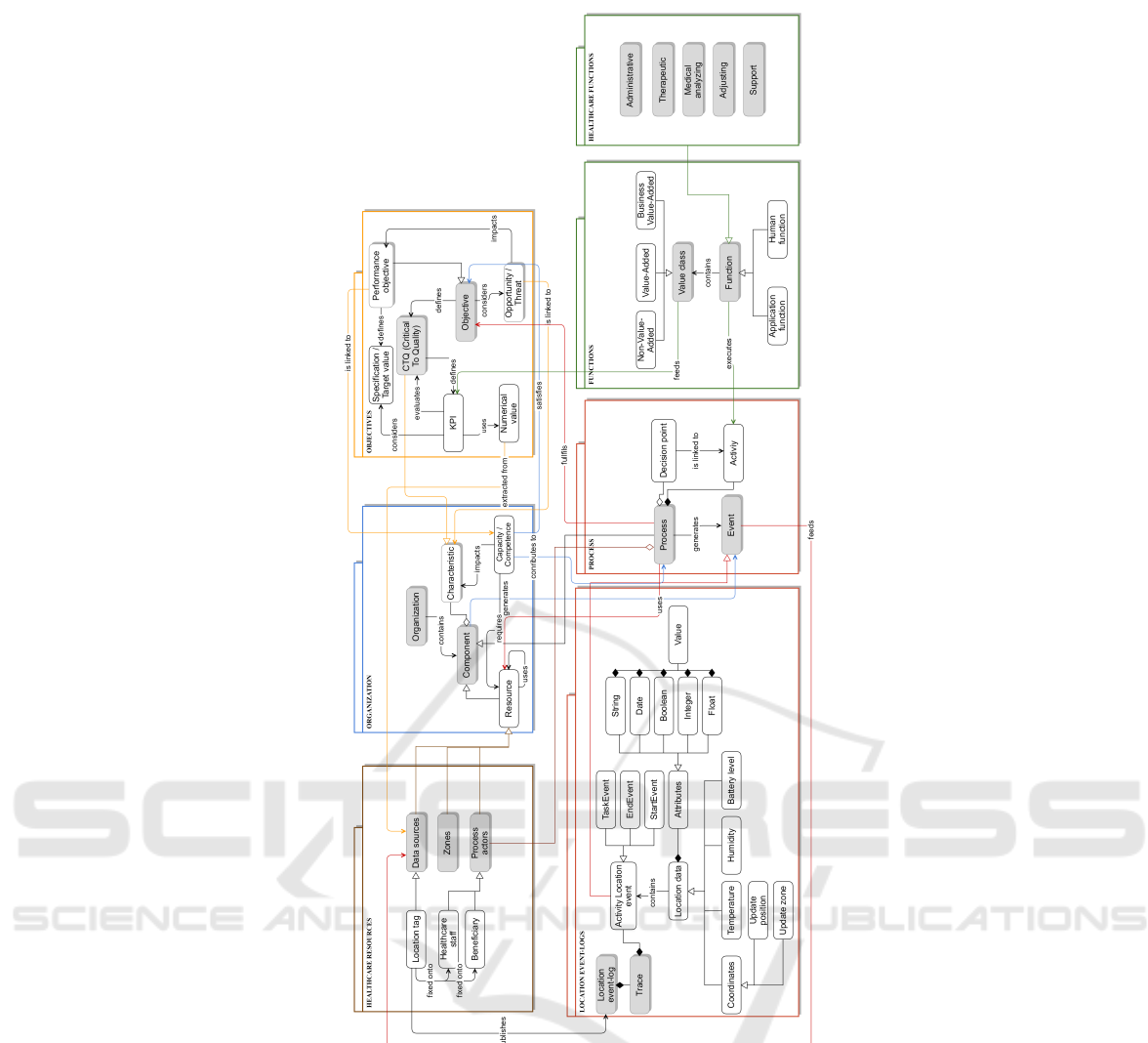


Figure 5: The DIAG meta-model.

event, (ii) Add knowledge to end-event, and (iii) Add knowledge to task-event. As mentioned this action is possible by the modeled information within the "resource" tab of the application.

Since the presentation of all the rules is much broader than the limit of this paper, we only present one of the most critical ones which is the **Add task-events**.

3.2.1 Add Task-event

The presented algorithm in this section (c.f. algorithm 1) describes how to discover the activities from the location event logs. Assuming human functions are executing the process activities within defined zones of the facility, one can infer that an activity (a task) starts when a process actor enters a certain zone and it finishes when the process actor leaves that zone. Algo-

rithm 1 expresses in more detail the rule for detecting such activities.

The algorithm is looking for the events that are related to the zones (*measurement.update.zone*). However, here the algorithm is seeking zones that are within the general process zones or zones with the id greater than '0' (If the $z_j.id > 0$).

After identifying the zone in z_{1_event} we look for the next event that share the same zone id ($z_j.id$), and additionally, we ensure that the second event also is related to the same patient (for example: $z_3.eui == z_6.eui$). This helps to ensure that the activity has been finished and the algorithm is not considering an **incomplete data** in the event log. This is an important characteristic to provide a refined set of data sets for avoiding *the incompleteness* notion in event logs.

Algorithm 1: Add-task-event.

```

1: procedure ADDTASKEVENT ▷ Discovering the Human Task
2:   Input < EventLog >= Structure
3:   < Name >: string;
4:   < Id >: integer;
5:   < Tags >: integer;
6:   < Topic >: string;
7:
8:   Select(EventLog, Topic == "measurement.update.zone")
9:
10:  < zone_event >= Structure;
11:   < Topic >= string ;
12:   < EventID >= integer;
13:
14:   < eui >= integer; ▷ j=1,2,3,...,m
15:   < DateTime >= string;
16:   < InZone >= boolean;
17:   < Zone.ID >= integer; ▷ i= 0,1,2,3,...,n
18:   for zone_event do
19:     if (z1 = zone_event [z1.topic== 'measurement.update.zone', AND z1.id > 0 ] AND (z2 = zone_event[z2.topic == 'measurement.update.zone'
AND z2.id > 0 AND z1.id==z2.id AND z1.eui == z2.eui]))
20:   then
21:     Select "z1.event_id" as 'HumanTask' ;
22:     Define 'name' ("z1.event_id" as "Activity.Inside.Zone");
23:     Register 'z1.event.propoerties' as 'HumanTask.propoerties';
24:   end if
25:   if ((z1.id == z2.id) AND( z1.eui==z2.eui )AND (z1.zone.attribute_in_zone=='TRUE') AND (z2.zone.attribute_in_zone=='FALSE') AND
(z1.date_time, z2.date_time ))
26:
27:   OR ▷ In case of existing noises
28:
29: (z2.zone.attribute_in_zone == 'TRUE') AND z1.zone.attribute_in_zone=='FALSE') AND (z2.date_time, z1.date_time)))
30: then
31: Register zj.date_time;
32: end if
33: end for
34: end procedure

```

After extracting the starting point of an activity, the algorithm randomly defines a node and name it as "Activity.Inside.Zone #" and all the defined properties of the node such as time will be recorded. Note that a process actor may go several times into a zone and goes out during his/her process. In order to avoid misinterpreting the activity sequences, the algorithm uses the *date_time* concept (at line 14 of the algorithm 1). At the final step, duration of the activity is registered thanks to the extracted start and end-event.

Later on, these nodes will be retracted by the "data interpreter" in order to add more knowledge on them. It is worth noting that, the a-priori knowledge is provided thanks to the implementation of the DIAG meta-model (c.f. 5).

4 RESULT

To evaluate the applicability of this approach, an experiment was conducted within a hospital living-lab in south of France. During this study, 30 patients were monitored. To begin with this experiment, the resources and other involved elements with the patient pathways were modeled in the R-IOSUITE application. This action is similar to what has been shown previously in figure 6. Next, each patient received a location tag. Consequently, several location event logs were registered by the localization system. These data were uploaded in the application to be interpreted.

An example of the interpretation procedure is shown in figure 7. In figure 7 at the left side, a map illustrates the movement of the patient who is wearing the tag 3. On the right side the graph represents how the events are seen. According to this graph each time

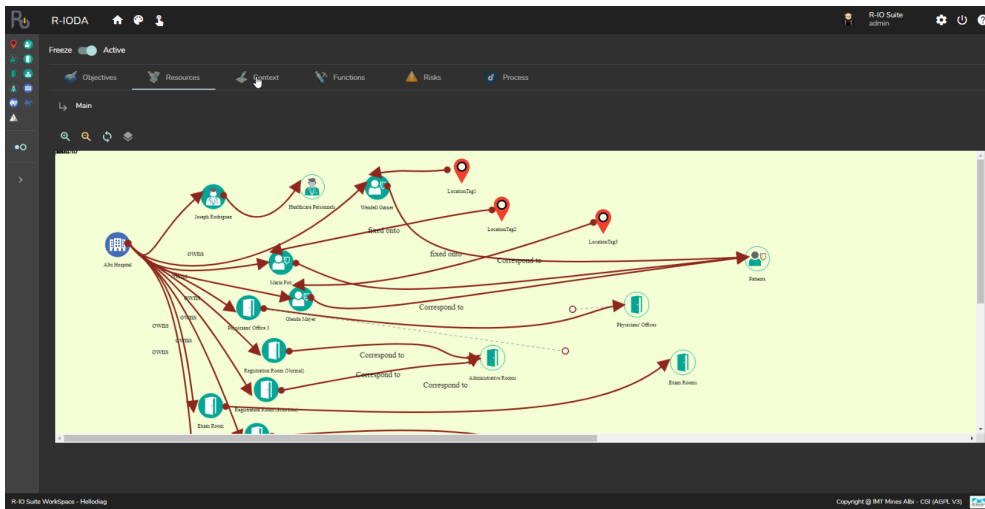


Figure 6: An screen-shot of the R.IO-DIAG application which illustrate the modeling of the healthcare resources in the application.

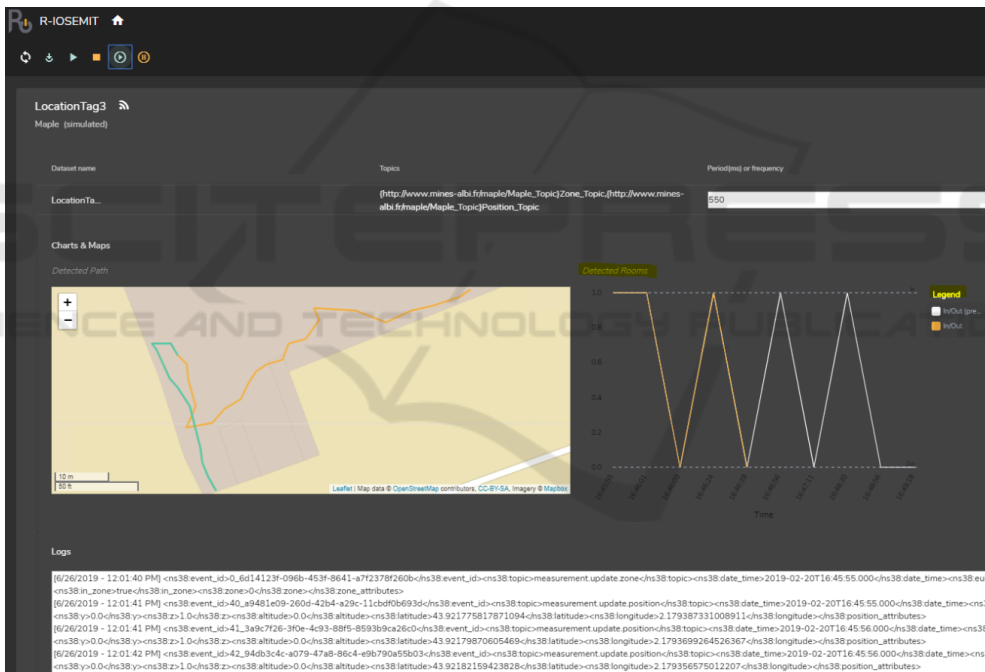


Figure 7: An illustration of the interpretation procedure.

a patient enters a zone the graph will reach the upper threshold, and by touching the lower threshold, it can be inferred that the patient left the zone.

After finishing the interpretation of the location event logs, the application discovers the patient pathways by using the heuristic miner algorithm. This well-known process discovery algorithm previously has been presented in the work of (Weijters and Ribeiro, 2011; Weijters et al., 2006; De Cnudde et al., 2014).

Figure 8 shows the result of the process modeling function. The patient pathways in this research work were modeled by the OPC (Operation Process Chart) modeling language (Namaki Araghi et al., 2018a).

Thanks to the development of the meta-model, the application is able to detect what are the different types of activities that patient are doing within their process. These activity types are defined within the **Healthcare function** package. This supports the further process mining analyses; such as, which class of

- study in Hainan. In *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, pages 1524–1529.
- Fernandez-Llatas, C., Lizondo, A., Monton, E., Benedi, J.-M., and Traver, V. (2015). Process mining methodology for health process tracking using real-time indoor location systems. *Sensors*, 15(12):29821–29840.
- Garaeva, A., Makhmutova, F., Anikin, I., and Sattler, K. (2017). A framework for co-location patterns mining in big spatial data. In *2017 XX IEEE International Conference on Soft Computing and Measurements (SCM)*, pages 477–480.
- Hwang, I. and Jang, Y. J. (2017). Process Mining to Discover Shoppers' Pathways at a Fashion Retail Store Using a WiFi-Base Indoor Positioning System. *IEEE Transactions on Automation Science and Engineering*, 14(4):1786–1792.
- Jin, P., Du, J., Huang, C., Wan, S., and Yue, L. (2015). Detecting hotspots from trajectory data in indoor spaces. In *International Conference on Database Systems for Advanced Applications*, pages 209–225. Springer.
- Kamel Boulos, M. N. and Berry, G. (2012). Real-time locating systems (RTLS) in healthcare: a condensed primer. *International Journal of Health Geographics*, 11:25.
- Lamr, M. and Skrbek, J. (2016). *Traffic Data and Possibilities of Their Utilization for Safer Traffic*. Technical Univ Liberec, Faculty Economics. WOS:000404420200007.
- Liao, J., Wang, Z., Wan, L., Cao, Q. C., and Qi, H. (2015). Smart Diary: A Smartphone-Based Framework for Sensing, Inferring, and Logging Users' Daily Life. *IEEE Sensors Journal*, 15(5):2761–2773.
- Martin, N. (2019). Using Indoor Location System Data to Enhance the Quality of Healthcare Event Logs: Opportunities and Challenges. In Daniel, F., Sheng, Q. Z., and Motahari, H., editors, *Business Process Management Workshops*, Lecture Notes in Business Information Processing, pages 226–238. Springer International Publishing.
- Martinez-Millana, A., Lizondo, A., Gatta, R., Vera, S., Salcedo, V. T., and Fernandez-Llatas, C. (2019). Process Mining Dashboard in Operating Rooms: Analysis of Staff Expectations with Analytic Hierarchy Process. *International Journal of Environmental Research and Public Health*, 16(2).
- Mazimpaka, J. D. and Timpf, S. (2016). Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science*, 2016(13):61–99.
- Miclo, R., Fontanili, F., Marquès, G., Bomert, P., and Lauras, M. (2015). RTLS-based Process Mining: Towards an automatic process diagnosis in healthcare. In *2015 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 1397–1402.
- Muzammal, M., Gohar, M., Rahman, A. U., Qu, Q., Ahmad, A., and Jeon, G. (2018). Trajectory Mining Using Uncertain Sensor Data. *IEEE Access*, 6:4895–4903.
- Namaki Araghi, S., Fontanili, F., Lamine, E., Salatge, N., Lesbegueries, J., Pouyade, S. R., and Benaben, F. (2019). Evaluating the Process Capability Ratio of Patients' Pathways by the Application of Process Mining, SPC and RTLS. pages 302–309.
- Namaki Araghi, S., Fontanili, F., Lamine, E., Salatge, N., Lesbegueries, J., Pouyade, S. R., Tancerel, L., and Benaben, F. (2018a). A Conceptual Framework to Support Discovering of Patients' Pathways as Operational Process Charts. In *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, pages 1–6.
- Namaki Araghi, S., Fontanili, F., Lamine, E., Tancerel, L., and Benaben, F. (2018b). Applying process mining and rtls for modeling, and analyzing patients' pathways. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5: HEALTHINF*, pages 540–547. INSTICC, SciTePress.
- Ramos, J., César, A., Neves, J., and Novais, P. (2017). Adapting the User Path Through Trajectory Data Mining. In De Paz, J. F., Julián, V., Villarrubia, G., Marreiros, G., and Novais, P., editors, *Ambient Intelligence—Software and Applications – 8th International Symposium on Ambient Intelligence (ISAmI 2017)*, Advances in Intelligent Systems and Computing, pages 195–202. Springer International Publishing.
- Rojas, E., Fernández-Llatas, C., Traver, V., Munoz-Gama, J., Sepúlveda, M., Herskovic, V., and Capurro, D. (2017a). Pali-er: Bringing question-driven process mining closer to the emergency room. In *BPM (Demos)*.
- Rojas, E., Sepúlveda, M., Munoz-Gama, J., Capurro, D., Traver, V., and Fernandez-Llatas, C. (2017b). Question-Driven Methodology for Analyzing Emergency Room Processes Using Process Mining. *Applied Sciences*, 7(3):302.
- Senderovich, A., Rogge-Solti, A., Gal, A., Mendling, J., and Mandelbaum, A. (2016). The ROAD from Sensor Data to Process Instances via Interaction Mining. In Nurcan, S., Soffer, P., Bajec, M., and Eder, J., editors, *Advanced Information Systems Engineering*, pages 257–273, Cham. Springer International Publishing.
- Sztyler, T., Carmona, J., Völker, J., and Stuckenschmidt, H. (2016). Self-tracking reloaded: applying process mining to personalized health care from labeled sensor data.
- Tang, L.-A., Yu, X., Gu, Q., Han, J., Jiang, G., Leung, A., and Porta, T. L. (2015). A Framework of Mining Trajectories from Untrustworthy Data in Cyber-Physical System. *ACM Trans. Knowl. Discov. Data*, 9(3):16:1–16:35.
- Tanuja, V. and Govindarajulu, P. (2016). Application of trajectory data mining techniques in crm using movement based community clustering. 16(11):20.
- Tanuja, V. and Govindarajulu, P. (2017). A novel framework for geo-clustering of user movements based on trajectory data. 17(3):212.
- Wan, N., Kan, G. L., and Wilson, G. (2017). Addressing location uncertainties in GPS-based activity monitoring: A methodological framework. *Transactions in GIS*, 21(4):764–781.
- Weijters, A., van der Aalst, W. M. P., and De Medeiros,

- A. A. (2006). Process mining with the heuristics miner-algorithm. 166:1–34.
- Weijters, A. J. M. M. and Ribeiro, J. T. S. (2011). Flexible Heuristics Miner (FHM). In *2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pages 310–317.
- Zheng, Y. (2015a). Trajectory Data Mining: An Overview. *ACM Trans. Intell. Syst. Technol.*, 6(3):29:1–29:41.
- Zheng, Y. (2015b). Trajectory Data Mining: An Overview. *ACM Transaction on Intelligent Systems and Technology*.
- Zhenjiang, D., Deng, J., Xiaohui, J., and Yongli, W. (2017). RTMatch: Real-time location prediction based on trajectory pattern matching. In *Database Systems for Advanced Applications - DASFAA 2017 International Workshops: BDMS, BDQM, SeCoP, and DMMOOC, Proceedings*, pages 103–117. Springer Verlag.

