

Location-based Automated Process Modelling

Ye Zhang^{1,2}, Olli Martikainen², Riku Saikkonen¹, and Eljas Soisalon-Soininen¹

¹ Aalto University, Finland

² PIKE Association, Finland

{ye.zhang, riku.saikkonen, eljas.soisalon-soininen}@aalto.fi,
olli.martikainen@pfu.fi

Abstract. Services are today over 70% of the Gross National Product in most developed countries. Hence, the productivity improvement of services is an important area. How to collect data from services has been a problem and service data is largely missing in national statistics. This work presents an approach to collect service process data based on wireless indoor positioning using inexpensive wireless sensors and smart phones. This work also presents how the collected data can be used to extract automatically the process model. These models can further be used to analyse the improvements of the service processes.

The presented approach comprises a light-weight process data acquisition system, which collects a minimised but precise data sets for automated process modelling. This automated modelling can be used to greatly improve the traditional process modelling in various service industries, for example, in the healthcare field. The presented approach has been tested and used in Tampere City dental care clinics.

Keywords: automated process modelling, process mining, location-based

1 Introduction

Service intelligence is becoming a worldwide trend. Efficiently and effectively running service operations are the key for gaining a competitive edge in almost every industry. The implementation of service intelligence relies heavily on a deep understanding of the service process, however, how to collect data from services has been a problem. This work presents an approach to collect service process data based on wireless indoor positioning using inexpensive wireless sensors and smartphones.

A service process is a set of activities in a chronological order and outputs a service as the final product. In this work we model processes graphically using boxes and arrows. One box represents an activity with service time and arrows indicate the transitions between activities. Based on the process data that we acquire in this work, we measure the average service time of each activity, and also transition probabilities between activities. This work focuses on modelling the generic service processes. Usually this type of services is location-aware, which means activities happen in specific locations. Therefore, we figured out that location information can be used to infer activities of generic service processes. Figure 1 illustrates our approach, which includes 4 phases:

- a). *Design phase*, it requires manually determining targeted activities and corresponding locations. Then plan the setting of wireless sensors and attach them to locations specified by activities.
- b). *Calibration phase*, it trains a set of measurement sensors on the mobile device side. Besides, it collects training data and transfers them to the server side for computing activity patterns and other parameters.
- c). *Process measurement phase* acquires process data and synchronises them to the server continuously for activity recognition.
- d). *Process modelling phase*, it models the whole process on the server side, based on the information of recognised activities.

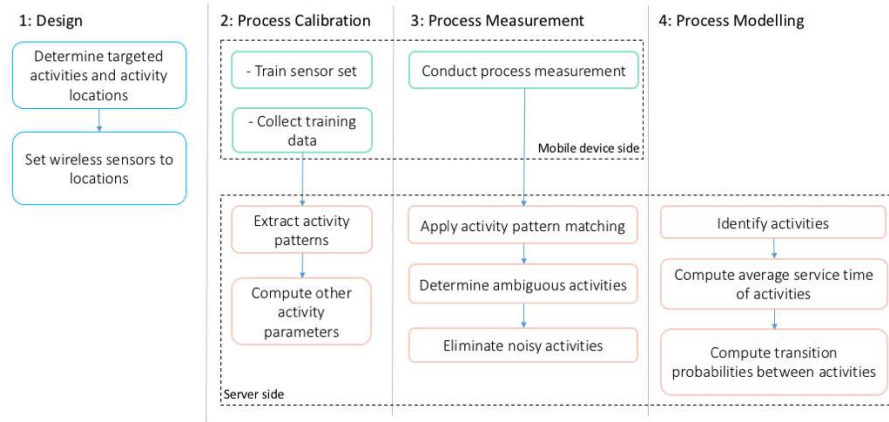


Fig. 1: The approach of real-time automated process modelling

We had extensive involvement in modelling service processes and improving service qualities in the healthcare sector. We have cooperated with Helsinki University Hospital, Meilahti Hospital and Tampere Dental Clinic. We used to model processes based on interview data [6], and then based on the captured process model, we analysed process performance optimisation with a tool called 3VPM [8]. However, social consulting agencies featured that automated wireless measurement as a more cost effective approach. This was the reason for starting our research. The idea of modelling location-aware processes was initiated in our previous work [21]. Nevertheless, the prototype in Zhang et al. [21] wasn't feasible for automated generating process model, activity analysis and process modelling was done manually. In addition to simple location data we have previously researched pattern recognition of signal sequences from wireless sensors attached to places and people to identify activities. These pattern recognition

techniques have been patented [9]. This work is an extension of our previous research and aims to improve the quality of the obtained location data and implement automated process modelling. This work contributes to the following aspects:

- a). Proposing an approach for automated process modelling.
- b). Implementing a light-weight process data acquisition system by utilising Bluetooth indoor positioning technique and Internet of Things (IoT).

This paper is organised as follows. Section 2 discusses related works. Section 3 illustrates the process data acquisition system and the analytical approach of process model extraction. In Section 4, we evaluate the system in a laboratory case study. Section 5 concludes the paper with the limitations of the current system and directions for future work.

2 Related works

The key factor of implementing service intelligence is successfully modelled service processes. Halonen et al. [6] documented process models extracted from interview data, and then used them in process performance optimisation. Based on a comparative analysis of four Australian public hospitals' healthcare processes, Partington et al. [15] demonstrated that through analysing the processes, it provided detailed insights into clinical (quality of patient health) and fiscal (hospital budget) pressures in health care practice. Another research [18] used declarative models for describing healthcare processes and reported that process mining can be used to mediate between event data reflecting the clinical reality and clinical guidelines describing best-practices in medicine.

Process mining has been widely explored in the healthcare sector, Halonen et al. [6] structured the processes in the acute neurology ward of Helsinki University Hospital by collecting data from process personnel interviews. Other research focused on analysing event logs of existing administrative systems or medical devices. Usually their targets are to solve a particular problem [11]. For example, Rebuge et al. [17] analysed the hospital emergency service, Mans et al. [10] studied the gynecological oncology process, Blum et al. [2] mined laparoscopic surgery workflow. However, approaches that are capable of picturing more generic process are still missing. We found out that generic processes usually have no trails in existing event logs. Accordingly, we abstracted generic activities to a location-based level and integrated Bluetooth indoor positioning and Internet of things techniques in the procedure of process modelling.

We are currently in the Big Data era, it opens new prospects for every industry and it is indeed promising to enable service intelligence [12] [19]. Nonetheless, the integration of high volume data from various sources complicates the operation of process mining. Moreover, we have learned from studies [3], the data quality of real-life logs is far from ideal, they are usually noisy, incomplete and imprecise. As a result, we were facing challenges such as how to guarantee the quality of data used for process modelling. Therefore, we intended to simplify the

procedure of automated process modelling by developing a light-weight process data acquisition system that collects minimised, but precise data sets.

In order to collect process activity related location data, we decided to use Bluetooth as an indoor positioning technique. Bluetooth has the advantages of low cost, highly ubiquitous, low power consumption, ad-hoc connection and shorter-range with room-wise accuracy [7] [4], which suits our purpose very well. Bluetooth in indoor positioning is a mature research field and has been widely studied [1] [14]. We utilised location information to infer abstract activities performed in corresponding locations besides using Bluetooth purely as indoor positioning technique. From this perspective, our research is closer to Faragher and Harle [5], which used fingerprint techniques such as pattern matching approaches to recognise activities. Faragher and Harle [5] declared that due to the instability of radio signal propagation, pattern matching approaches are much more effective than approaches that use radio proximity.

The developments in inexpensive and unobtrusive sensors, machine learning and data mining techniques have enabled automated process modelling. Wan et al. [20] used an online sensor data segmentation methodology for near real-time activity recognition. Okeyo et al. [13] presented an approach to real-time sensor data segmentation for continuous activity recognition. Pham et al. [16] implemented a real-time activity recognition system to detect low-level food preparation activities. Likewise, they also streamed sensor data continuously for real-time analysis. Nonetheless, most of them focus on detecting motion activities and there are not yet enough application of IoT in recognising process activities. Differentiate from discrete motion activities, our targeted activities are process related activities, which are usually ordered and have transition probabilities between them.

3 Automated process modelling system

With the objectives to guarantee the quality of data used for process modelling, meanwhile to simplify the procedure of the generic healthcare process modelling, we propose an automated process modelling system. The system consists of four principal modules: a). Process data acquisition, b). Calibration of the process measurement, c). Process measurement, and d). Process model extraction.

3.1 Process data acquisition

The process data acquisition module intends to acquire minimised but precise activity data sets. Bluetooth data was collected to infer indoors location information, and furthermore used to represent location-based activities. Accordingly, we collect chronological sequenced tuples that measure Radio Signal Strength Indications(RSSI) of Bluetooth sensors. There are a variety of input data for process measurement. One feature of the system is its capability of measuring multiple processes. We defined a process as a measurement site that uses a specific set of Bluetooth sensors. Hence, we categorised input data into two types.

- a). General input data: a full list of Bluetooth devices $D = \{Sensor_1, Sensor_2, Sensor_3, \dots, Sensor_N\} + \{User_1, User_2, User_3, \dots, User_N\}$, which includes both Bluetooth beacons and user devices. This type of information is kept on the server side, the minimised information of a Bluetooth device needed is the Bluetooth MAC address.
- b). Measurement site specific input data:
 - A subset of Bluetooth devices $Sensor_{sub} = \{Sensor_4, Sensor_7, \dots, Sensor_X\}$. One or multiple devices are set in a location to represent an activity, therefore, different measurement sites have different Bluetooth devices subset. Each subset is independent, but the included Bluetooth devices can be either exclusive or overlapping.
 - A subset of users involved in the specific measurement, $User_{sub} = \{User_1, User_3, User_6, \dots, User_Y\}$, similarly to Bluetooth devices subset, each user subset is meant for a specific process and same user can take part in multiple processes.
 - A list of activities in the specific process, $Activity_{list} = \{Activity_1[], Activity_2[], \dots, Activity_Z[]\}$. Each activity item is a vector that provides minimised information such as $Activity_Z[ID, activityName, locationID, [Sensor_1ID, Sensor_4ID \dots]]$, which contains activity ID, activity name, location ID and a list of sensor IDs.

In addition, we defined following attributes to be measured, a record tuple at time T_t is:

$$Tuple_{T_t} = (T_t, User_u, [RSSI_{Sensor_1T_t}, RSSI_{Sensor_2T_t}, \dots, RSSI_{Sensor_S T_t}])$$

where T_t is tuple's timestamp; $User_u$ is the user involved in the process, it's the Bluetooth MAC address of the user mobile device. At each time point T_t , there is a S-sized RSSI vector. S is the number of Bluetooth transceivers in a particular process measurement. If Bluetooth sensor is out of range, RSSI = 0, otherwise RSSI equals the real-time measured value.

3.2 Calibration of the process measurement

In this research, we built Bluetooth transceivers with JY-MCU Bluetooth wireless serial port modules ³. Generally, the radio propagation is extremely complex and unstable. We tried to compare the performance of our transceivers by measuring each from the same distance. However, the radio signal strength obtained varied dramatically. As a result, we had to include an essential step in the light-weight system: *calibration*. It collects a training data set and generates parameters of sensor performances. Calibration only requires an administrator role to walk through all the locations with all of the user devices and let the devices measure a few data points of RSSI information at each location. This subject-independent approach to keep general users away from the burden of

³ <https://core-electronics.com.au/attachments/guides/Product-User-Guide-JY-MCU-Bluetooth-UART-R1-0.pdf>

training phase and provides them with a ready-to-use application. Furthermore, it will facilitate the application of this system. Responsibilities of calibration are as follows:

- a). From the full list of all predefined Bluetooth devices, it trains a subset of Bluetooth devices for measuring a specific process.
- b). Synchronises device's local time with the remote server time, in order to maintain the consistency of timestamps of records collected from different devices. This is an essential step for multi-users collaborative activity recognition.
- c). Collects a training data set: tuples $Tuple_{T_i}[i]$, in which i is a discrete time point with sampling rate interval(s), $i = 0 + n \times rate$. Thereafter, this time-based training data set is combined with user-supplied information of actual activity to form activity patterns: that is, for each activity, we collect a set of possible RSSI patterns seen in that location.

3.3 Process measurement

After selecting a particular process to measure, information about corresponding set of Bluetooth sensors will be synchronised from the remote server. Subsequently, the background service applies asynchronous *Broadcast Receiver* schema to periodically detect RSSI vectors. It is user interfaces independent and non-obtrusive for general users. The asynchronous broadcast receiver schema is basically a broadcast receiver keeps listening to two actions: *Action One*, a remote Bluetooth sensor found; *Action Two*, one Bluetooth inquiry finished. Action One is triggered when the mobile device enters the radio proximity of a fixed Bluetooth transceiver, in the meantime, the system collects real-time Bluetooth RSSIs. Action Two is triggered when one Bluetooth inquiry duration ends (about 12 seconds). Thereafter, a new Bluetooth discovery will start.

The sampling rate is 12s, same as the duration of one Bluetooth inquiry. Upon this architecture, the integration of IoT enables automated process modelling: the system collects tuples continuously, meanwhile, the mobile device periodically synchronises tuples to the remote server through Wi-Fi. The synchronising rate is adjustable based on the measurement needs. The system applies Google *Volley* networking framework⁴ to stream data between the server and mobile devices. On the server side, it applies activity pattern matching. In addition to this, the system uses sensor performance parameters to determine ambiguous activities. Ultimately, window size is used over incoming tuples to eliminate noisy activity detections.

3.4 Analytical approach for process model extraction

This work applied the following analytical approach to identify process activities and extract process model. To simplify the following discussion, we assume that

⁴ <https://developer.android.com/training/volley/index.html>

there is only one person whose activities are being measured. The system is capable to measure multiple independent processes separately at the same time. The analytical approach is shown in Figure 2 and Figure 3.

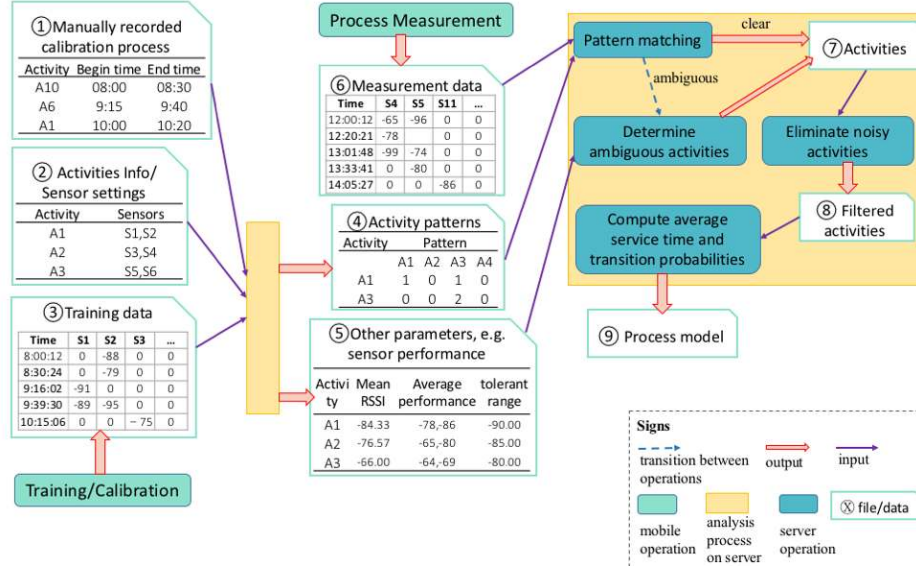


Fig. 2: The analytical approach for process model extraction

- During the training/calibration phase, two operations are performed: mobile side executes training and generates ③ training data, which includes tuple timestamps and RSSIs of the full list of Bluetooth devices. The other operation is manually record the training process, contains ① information such as activity ID, activity begin and end time. Then, together with ② the activity information/sensor settings defined in 3.1 b), server side's analytical process will compute ④ activity patterns and ⑤ other activity related parameters as output. In activity pattern, 0 means none of the sensors that represent this activity is in range. 1 means we can detect the radio signal of one of the activity's sensors. 2 indicates both of the activity's sensor are in range. In practice, there is signal overlapping issues, for example, ④ activity pattern "A1: 1,0,1,0" means during activity 1, the device also received RSSI from sensor that represents activity 3.
- Process measurement collects real-time ⑥ measurement data. Similar to training data, it contains tuple timestamps, but instead of RSSIs vectors of the full list of Bluetooth devices, it only records RSSIs vectors of the subset Bluetooth devices.
- Server's analytical process applies activity pattern matching on each measurement tuple transferred from mobile devices. Since different activities may

- have the same patterns, a tuple may match more than one activity classes. To address this, we first detected activities that have clear activity patterns. Afterwards, we took a step further to analyse ambiguous tuples. Parameters computed in calibration, such as sensor's mean RSSI, average performance sensors in each activity, tolerant range of signal strength for each activity are integrated to determine ambiguous activities.
- d). Server applies a window size over a few $\textcircled{7}$ successive activities, in order to eliminate noisy activity detections (for example, switching to a new activity momentarily, and back to the original activity in the next data point), as shown in Formula <2> in Figure 3. Generally, we assume this type of noise contains less than two tuples.
 - e). Next, we identified the beginnings and endings of activities. It outputs changes of activities, 0 means no change, -1 indicates no activities or constant change between activities, as shown in Formula <1> in Figure 3. At last, we computed average service time and transition probabilities to model the process.

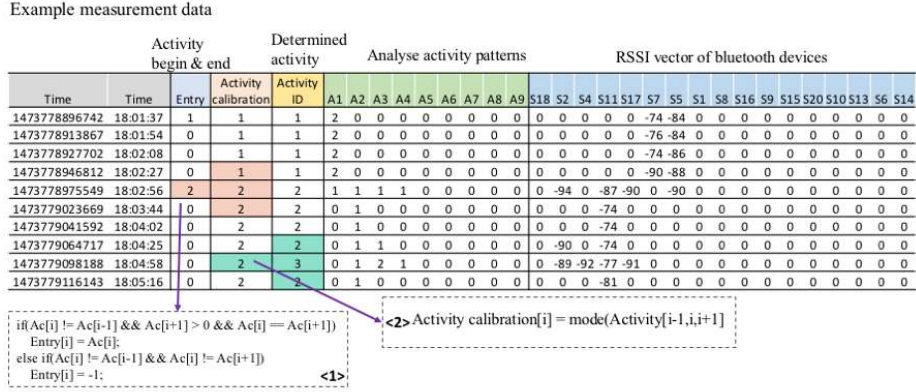


Fig. 3: Example of measurement data & analysis visualization. (Column "Entry" is the final output of this analysis)

4 Case studies

We implemented the data acquisition system for Android smartphones. In addition, we evaluated the system and the analytical approach for process model extraction in a laboratory case study. We placed 17 Bluetooth transceivers in 8 locations in the computer science building in Aalto University to represent 9 activities. Figure 4 shows the setting of sensors in the process measurement. As mentioned, the performances of sensors vary and are neither stable nor consistent. In order to find an optimal setting of sensors, we conducted several

experiments and found out that, the using of two sensors to represent one activity helps improve the process measurement results. For comparison reasons, we wrote down the actual process on paper manually in addition to the automated process measurement with mobile devices. The process measurement results are presented in Figure 5a and Figure 5b.

Figure 5a shows the result of using proximity detection approach. It indicates two problems: one, when the locations of two activities are relatively close to each other, this approach will lead to noisy fluctuations; second, when there is only a very short interval between two activities, it won't be accurate enough to determine the interval. Figure 5b demonstrates the application of the analytical approach illustrated in section 3.3. By comparing with the actual process, the result shows that the analytical approach for process model extraction detected the correct activity in 93% of the data points. In other words, the system fulfils the demand of collecting precise process data for accurate process modelling.

The process model captured from the case study is shown in Figure 6. The average service times and transition probabilities are calculated from the analysed data (i.e. the begin and end times of each occurrence of an activity) as follows. For an activity $i \in \{1, \dots, n\}$ that occurred m_i times in the data, the average service time is $S_i = 1/m_i \sum_{j=1}^{m_i} d_{i,j}$, where $d_{i,j}$ is the duration of the j th occurrence of activity i . We then compute a matrix of how many (directed) transitions occurred between the activities: $T_{i,j}$ = number of transitions from activity i to activity j . From this we can calculate transition probabilities by scaling with the total number of outgoing transitions from an activity. That is, the transition probability $P_{i,j}$ from i to j is $P_{i,j} = T_{i,j} / \sum_{k=1}^n T_{i,k}$.

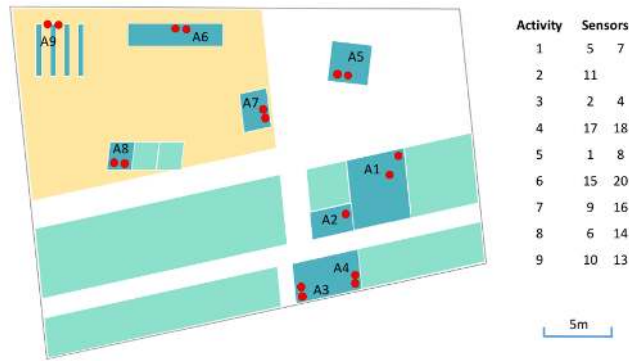


Fig. 4: Case study in Aalto University: sensors setting for process measurement

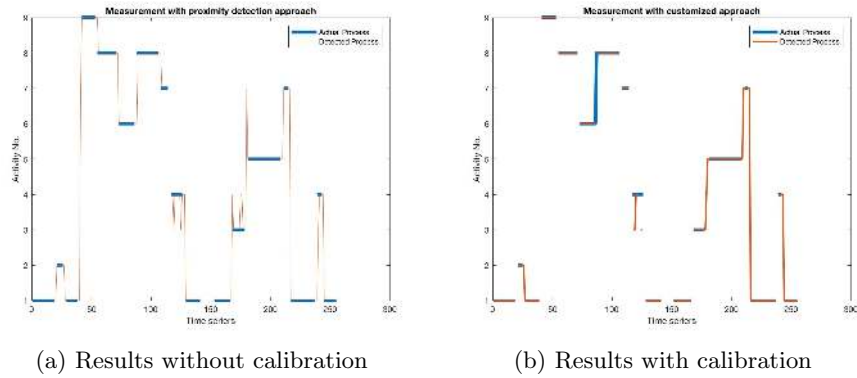


Fig. 5: Evaluation of automated process measurement results

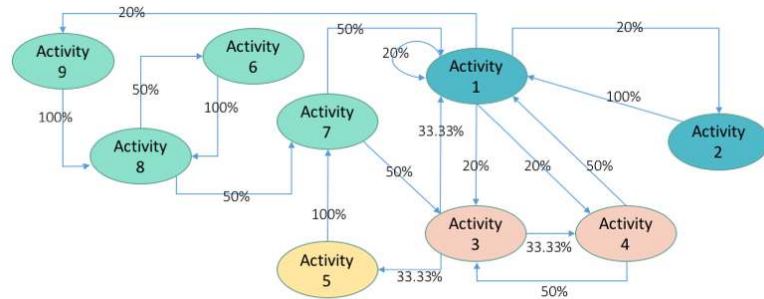


Fig. 6: Process model extracted by automated process modelling

5 Conclusions

Process modelling is a critical factor in the improvement of service productivity and in the implementation of service intelligence. However, how to collect data from services has been a problem. This work focused on automated modelling of generic service processes that are location aware. In other words, activities in the process usually happen in a particular location and location information can be used to infer activities. Accordingly, we presented an approach to collect service process data based on wireless indoor positioning using inexpensive wireless sensors and smartphones. The objective of this work was to simplify the procedure of automated process modelling. For this reason, we designed a process data acquisition system to acquire minimised, but precise data set, instead of taking overwhelming redundant data. In our approach, Internet of things is integrated to implement real-time automated process modelling. We illustrated the analytical approach for process model extraction in this system and we examined the performance of the process data acquisition system and the analytical approach

in a case study. The results of the case study demonstrate that the system fulfils the demand of collecting precise process data for accurate process modelling. In addition, the presented approach has been tested and used in Tampere City dental care clinics. Their measurement results confirm the feasibility of this approach in process modelling and the feasibility of using the extracted models in process performance optimisation.

Application status of the current system is limited to relatively ideal settings: one location represents only one activity. Besides, the system requires that two locations have a certain distance (minimum 2 meters). As illustrated in our analytical approach, we eliminate out noisy activities that have less than two tuples. Hence, the shortest activity that can be detected has at least two tuples (about 24 seconds). The current system is applicable for analysing the process of singular user rather than analysing the collaborative process of a team. Therefore, our objective of future research is to implement automated process modelling for team collaboration process. Moreover, improve the accuracy of process activity recognition with the help of additional data, for example, accelerometer data.

Acknowledgments

This work has been conducted in the Techniques for Efficient Processing of Big Data Project in Aalto University, and was supported by The Academy of Finland and Service Innovation Research Institute (PIKE).

References

1. Baniukevic, A., Jensen, C. S., Lu, H.: Hybrid indoor positioning with wi-fi and bluetooth: Architecture and performance. In 2013 IEEE 14th International Conference on Mobile Data Management, Vol. 1, 207–216 (2013)
2. Blum, T., Padoy, N., Feußner, H., Navab, N.: Workflow mining for visualization and analysis of surgeries. In International Journal of Computer Assisted Radiology and Surgery, 3(5), 379–386 (2008)
3. Bose, R. J. C., Mans, R. S., van der Aalst, W. M.: Wanna improve process mining results? In Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on. IEEE, 127–134 (2013)
4. Dardari, D., Closas, P., Djurić, P. M.: Indoor tracking: Theory, methods, and technologies. IEEE Transactions on Vehicular Technology, 64(4), 1263–1278 (2015)
5. Faragher, R., Harle, R.: Location fingerprinting with bluetooth low energy beacons. IEEE Journal on Selected Areas in Communications, 33(11), 2418–2428 (2015)
6. Halonen, R., Martikainen, O., Juntunen, K., Naumov, V.: Seeking efficiency and productivity in health care. In 20th Americas Conference on Information Systems. AMCIS-0251-2014.R1. (2014)
7. Liu, H., Darabi, H., Banerjee, P., Liu, J.: Survey of wireless indoor positioning techniques and systems. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37(6), 1067–1080 (2007)
8. Martikainen, O., Halonen, R.: Model for the Benefit Analysis of ICT. In 17th Americas Conference on Information Systems. AMCIS, 4-7 (2011)

9. Martikainen, O.: A method and a computer program product for controlling the execution of at least one application on or for a mobile electronic device, and a computer. Patent, EP2758874 (2011)
10. Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M., Bakker, P. J.: Application of process mining in healthcare case study in a dutch hospital. In International Joint Conference on Biomedical Engineering Systems and Technologies. Springer Berlin Heidelberg, 425–438 (2008)
11. Mans, R. S., van der Aalst, W. M., Vanwersch, R. J., Moleman, A. J.: Process mining in healthcare: Data challenges when answering frequently posed questions. In Process Support and Knowledge Representation in Health Care. Springer Berlin Heidelberg, 140–153 (2013)
12. Meng, S., Dou, W., Zhang, X., Chen, J.: Kasr: A keyword-aware service recommendation method on mapreduce for big data applications. *IEEE Transactions on Parallel and Distributed Systems*. 25(12), 3221-32 (2014)
13. Okeyo, G., Chen, L., Wang, H., Sterritt, R.: Dynamic sensor data segmentation for real-time knowledge-driven activity recognition. *Pervasive and Mobile Computing*, 10, 155-172 (2014)
14. Palumbo, F., Barsocchi, P., Chessa, S., Augusto, J. C.: A stigmergic approach to indoor localization using bluetooth low energy beacons. *Advanced Video and Signal Based Surveillance (AVSS)*, 12th IEEE International Conference on. IEEE, 1–6(2015)
15. Partington, A., Wynn, M., Suriadi, S., Ouyang, C., Karnon, J.: Process mining for clinical processes: a comparative analysis of four Australian hospitals. *ACM Transactions on Management Information Systems (TMIS)*, 5(4), 19 (2015)
16. Pham, C., Pltz, T., Olivier, P.: A dynamic time warping approach to real-time activity recognition for food preparation. In International Joint Conference on Ambient Intelligence, Springer Berlin Heidelberg, 21-30 (2010)
17. Rebuge, Á., Ferreira, D. R.: Business process analysis in healthcare environments: A methodology based on process mining. *Information Systems*, 37(2), 99–116 (2012)
18. Rovani, M., Maggi, F. M., de Leoni, M., van der Aalst, W. M.: Declarative process mining in healthcare. *Expert Systems with Applications*, 42(23), 9236–9251 (2015)
19. Vera-Baquero, A., Colomo-Palacios, R., Molloy, O.: Business process analytics using a big data approach. *IT Professional*, 15(6), 29–35 (2013)
20. Wan, J., OGrady, M. J., OHare, G. M.: Dynamic sensor event segmentation for real-time activity recognition in a smart home context. *Personal and Ubiquitous Computing*, 19(2), 287–301 (2015)
21. Zhang, Y., Martikainen, O., Pulli, P., Naumov, V.: Real-time process data acquisition with Bluetooth. In Proceedings of the 4th International Symposium on Applied Sciences in Biomedical and Communication Technologies, Barcelona, Spain, Vol. 2629 (2011)