A Proactive Workflow Model for Healthcare Operation and Management

Chuanren Liu, Hui Xiong, *Senior Member, IEEE*, Spiros Papadimitriou, Yong Ge, and Keli Xiao

Abstract—Advances in real-time location systems have enabled us to collect massive amounts of fine-grained semantically rich location traces, which provide unparalleled opportunities for understanding human activities and generating useful knowledge. This, in turn, delivers intelligence for real-time decision making in various fields, such as workflow management. Indeed, it is a new paradigm to model workflows through knowledge discovery in location traces. To that end, in this paper, we provide a focused study of workflow modeling by integrated analysis of indoor location traces in the hospital environment. In particular, we develop a workflow modeling framework that automatically constructs the workflow states and estimates the parameters describing the workflow transition patterns. More specifically, we propose effective and efficient regularizations for modeling the indoor location traces as stochastic processes. First, to improve the interpretability of the workflow states, we use the geography relationship between the indoor rooms to define a prior of the workflow state distribution. This prior encourages each workflow state to be a contiguous region in the building. Second, to further improve the modeling performance, we show how to use the correlation between related types of medical devices to reinforce the parameter estimation for multiple workflow models. In comparison with our preliminary work [11], we not only develop an integrated workflow modeling framework applicable to general indoor environments, but also improve the modeling accuracy significantly. We reduce the average log-loss by up to 11 percent.

Index Terms—Indoor location traces, workflow modeling, healthcare operation and management

1 Introduction

Real-time location systems (RTLS) are being rapidly developed and deployed. Of note, hospitals are increasingly using these systems to track the movement of medical devices, doctors, and patients (see Fig. 1), as well as the interaction among them. However, their utilization is currently limited to basic tasks, such as locating a wheelchair or checking the availability of an inpatient bed. In the near future, we expect indoor location tracking data to be widely available in many hospitals, as well as in other environments (e.g., shops, schools, warehouses, etc).

Understanding sequences of *procedures* that reflect *workflows* (e.g., surgery, from admission to recovery) remains an important challenge, which we address in this work. Workflows reveal semantically meaningful patterns that can help (i) understand how space and assets (e.g., medical equipment, classrooms, shopping areas) are utilized (workflow *auditing*); (ii) ensure that such utilization complies with rules and regulations (workflow *compliance*); and (iii) perform any of these tasks in real time (workflow *monitoring*). For example, many healthcare providers have their own work protocols to ensure that healthcare practices are executed in a controlled manner. Non-compliance to these protocols may be costly and expose the healthcare providers to severe risks, such as litigation, prosecution, and damage to brand reputation. Thus, there is a real need for effective inspection of workflow compliance. This work focuses on fundamental models to support all of the above tasks.

Hospital managers traditionally accomplish these tasks by inspecting the detailed workflow logs [7], [14], [15], [19], which can be in heterogeneous formats, stored in different media (including paper), and provided passively by personnel and, therefore, may be biased and incomplete. The overall task is quite daunting, and the opportunities to develop proactive approaches to help with workflow management tasks are unparalleled. However, RTLS deployments are still used in a relatively basic way, as noted above, with little work focusing on how to leverage massive indoor location traces. To this end, this paper provides a focused study of workflow modeling via integrated analysis of indoor location traces, evaluated on real data from hospital environments. Such workflow models serve as fundamental building blocks in a wide range of workflow management problems.

In particular, we propose the *Proactive Workflow Model* (ProWM) that, leveraging indoor location traces, provides a unified framework for simultaneously (i) identifying semantically meaningful regions as workflow states, and (ii) summarizing sequential procedures as workflow transitions. Moreover, the proposed method is applicable to general indoor environments where the workflow patterns are hidden in the location traces of moving objects.

- *C. Liu is with the Decision Science and Management Information Systems Department, Drexel University, Philadelphia, PA 19104. E-mail: chuanren.liu@drexel.edu.*
- *H. Xiong and S. Papadimitriou are with the Management Science and Information Systems Department, Rutgers University, New Brunswick, NJ 08901. E-mail: {hxiong, spapadim}@rutgers.edu.*
- *Y. Ge is with the Eller College of Management, University of Arizona, Tucson, AZ 85721. E-mail: yongge@email.arizona.edu.*
- *K. Xiao is with the College of Business, Stony Brook University, Stony Brook, NY 11794. E-mail: keli.xiao@stonybrook.edu.*

Manuscript received 14 June 2015; revised 28 Oct. 2016; accepted 14 Nov. 2016. Date of publication 22 Nov. 2016; date of current version 2 Feb. 2017. Recommended for acceptance by A. Banerjee.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2016.2631537

1041-4347 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

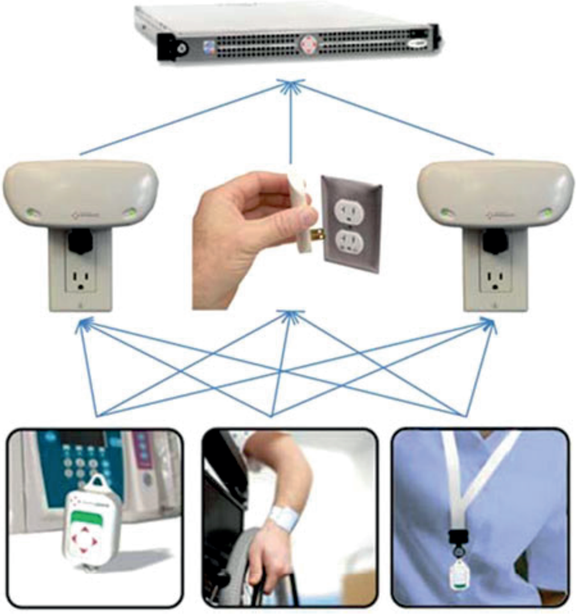


Fig. 1. An example of a real-time location system (RTLS) deployed in a hospital. Bottom layer: sensor tags attached to moving objects (e.g., medical devices, patients, and doctors); middle layer: receivers relaying signals from sensors; top layer: network bridges connected to data/application servers, to calculate and store locations of tracked objects.

	MICRO	MESO	MACRO
LOCATION	3D coordinates	1a. Rooms	2a. (3.) Distribution over rooms
MOBILITY	Line distance	1b. (3.) Room graph (inferred)	2b. (3.) Transition prob. matrix

Fig. 2. Challenges (numbers explained in the introduction) and our proposed solutions (filled boxes).

Although workflow modeling is central to many building operations and management tasks, systematically constructing and estimating those models based on massive indoor location traces is a non-trivial endeavor. Next, we identify specific challenges in workflow modeling from indoor location traces, and we outline the ingredients of our proposed solution, which revolve around representation of position/location and of mobility/transitions at three different levels: micro, meso, and macro (see Fig. 2).

At the "micro" level, we have the raw data, which consist of three-dimensional coordinates and geometric (euclidean) distance between them. Based on these, we have to construct or infer appropriate representations for workflow modeling.

1. *"Meso" level.* The raw data at the micro level, which is what much of the research on indoor location traces focuses on, are not appropriate for our purposes for a number of reasons.

1a. Location. The granularity and quality of indoor location traces captured by wireless location systems in the form of (x, y, z) coordinates might vary substantially from place to place due to several factors: the density of sensor receivers, environment changes (e.g., doors opening/closing) the underlying localization techniques, and the sensor device itself (e.g., remaining battery power). Given a particular set of indoor location traces, it is very important to be aware of its specific granularity and quality when modeling or analyzing the data. Typically, it is possible to identify the *room* that corresponds to a location with reasonable accuracy [10]. We therefore use the floor plan information to map raw locations (continuous) to rooms (discrete symbols).

1b. Topology/Proximity. Furthermore, the topologies of indoor space are often much more complex than outdoor space. For example, a small distance between two locations does not mean that one is easily reachable from the other (e.g., two rooms with a shared wall but no door between them). Therefore, some fundamental assumptions commonly used for (outdoor) trajectories may not hold for indoors. For instance, widely-used similarity measures based on geometric distance [4] or degree of overlap [6] are not very meaningful for indoor location traces. Our solution is to use a graph of communicating rooms. An additional complication is that this graph may change over time, and therefore should also be inferred, rather than extracted from the floor plan.

2. *"Macro" level.* The representations at the "meso" level are still too fine-grained to provide useful insights into modeling the workflows.

2a. Location Granularity and Semantics. Multiple locations may be used either concurrently or interchangeably and, therefore, should be grouped together. However, this grouping will depend on the workflow/procedure, and may change over time. We model the functional significance of a location in the context of a workflow as a hidden state. Therefore, a state is a probability distribution over rooms. However, in order to obtain interpretable results, we need to regularize it. To that end, we employ the graph derived in (1b) above.

2b. Transition Patterns. Once we have identified the states (which, intuitively, should correspond to stages, or procedures, that comprise a workflow), we need to learn the transition patterns between them. An additional challenge is that, if we consider each device type independently, we may have insufficient data to robustly estimate model parameters. However, certain medical devices are often used together for particular healthcare tasks. By leveraging such naturally arising correlations, we may improve the robustness of parameter estimation for the finite state machines. We show that correlations can be incorporated effectively and efficiently as another regularization, for simultaneously estimating multiple workflow models.

3. *Evolving Environment.* In addition to the above challenges, the structure of indoor space is very dynamic and the modeling framework should automatically adapt. In more traditional settings, such as city road networks, the structure of the environment may remain mostly unchanged for years. However, the structure and utilization of modern buildings may frequently change, based on evolving needs. Although room partitions (1a, above) typically don't change, everything else (1b, 2a, and 2b) may change. For example, an elevator may be out of order, sections of the building may be closed for cleaning or repairs, thereby changing the graph. Rooms may become unavailable or have their purposes reassigned, therefore changing state composition. Workflows may also change over time, as procedures are updated. Such dynamic changes will alter the semantics of indoor location traces. In order to effectively and efficiently deal with these challenges, our method should be computationally efficient

and also able to robustly estimate parameters based on more limited, recent data, rather than rely on a long past history.

As a result, although there is extensive work on the analysis of location traces, e.g., [5], [6], [13], most is not suitable for modeling indoor location traces for proactive workflow analysis. For example, the method developed by [6] discovers frequent trajectory patterns from outdoor location traces, based on given thresholds (minimum support and time tolerance). However, such frequent patterns cannot provide a parsimonious description of hospital workflows. For instance, workflow compliance inspection requires all activities, rather than only a frequent subset of moving patterns. Similarly, the periodic patterns mined from outdoor spatio-temporal data by [13] are also a subset of moving activities and cannot fully meet the needs of workflow monitoring, auditing, or compliance inspection. Finally, although stochastic models for indoor activities were proposed by [18], the purpose of these models is to classify activities (rather than provide an interpretable, parsimonious summary of overall activity patterns), based mostly on supervised learning, requiring sufficient labeled training data.

In preliminary work [11], we approached challenges (2a) and (2b) independently, by employing a density-based clustering algorithm over the room graph, to construct spatially contiguous "hard" clusters of rooms. Once this clustering solution was fixed, states were fully observed, and we therefore can directly learn the transition matrices between these clusters.

Contributions. We propose stochastic models to proactively unravel the workflow patterns hidden in the massive indoor location traces, by automatically discovering workflow states, and estimating parameters describing the transition patterns. The discovered knowledge can then be transformed to valuable, actionable intelligence in a wide range of practical problems in indoor environments (e.g., hospitals).

More specifically, in this work we unify the treatment of challenges (2a) and (2b), by proposing a novel hidden Markov model which is regularized using the inferred room graph and transition correlations. This allows us to learn good representations of states (which are now "soft" probability distributions) and jointly optimize learning states and transitions. In addition to rigorously unifying the modeling framework, our proposed approach also achieves better accuracy than previous work, without sacrificing computational efficiency.

We have also implemented and deployed a management information system, *HISflow*, based on our methods to show how the discovered knowledge can help with the three important managerial tasks in hospitals: workflow monitoring, auditing, and compliance inspection.

The rest of the paper is organized as follows. Section 2 discusses work related to modeling healthcare workflow patterns, and to analyzing indoor location traces. Section 3 introduces the indoor location data and formalizes the problem of workflow modeling. Section 4 presents our model based on regularized HMM (Hidden Markov Model) and Section 5 details model estimation. Section 6 reports results from extensive experimental evaluation of our method on both synthetic data and real-world data. Finally, Section 7 concludes the paper.

2 Related Work

Workflow Analysis and Mining. Workflow analysis conventionally relies on detailed workflow logs [1], [7], [16]. Workflow processes are typically represented by activity graphs. Given the execution logs, which are lists of activity records, workflow mining can be formalized as a graph mining problem by viewing execution logs as walks on the activity graphs [1]. In practice, there might be discrepancies between the actual workflow processes and those perceived by the management. In this case, to discover a completely specified workflow design model, [16] presented an algorithm to extract a process model from the workflow logs and represent it in terms of a Petri net. Instead of discovering the complete model, [7] later formalized the problem of discovering the most frequent patterns of executions, i.e., the workflow substructures that have been scheduled more frequently and have lead to a desired final configuration. However, these methods rely on workflow logs which are often manually recorded in several settings, including the healthcare industry. Thus, the results may be distorted due to bias and missing records. These distortions can be misleading for many operation and management tasks in hospitals, such as the inspection of workflow compliance. In comparison, as we discussed in Section 1, in this paper we propose a proactive approach to workflow modeling by mining the digital location traces of moving objects. These are automatically recorded by RTLS, requiring practically no human intervention. The statistical modeling results provided by our approach are helpful for a range of operation and management tasks in hospital environments.

Activity Modeling and Prediction. In terms of methodology, another category of related work is the modeling and prediction of human activities. For instance, [18] proposed stochastic process models to predict the goals of indoor human activities. Furthermore, for multiple-goal recognition, [2] proposed a two-level architecture for behavior modeling and [8] developed a dynamic Bayesian model where skip-chain conditional random fields were used for modeling interleaving goals. All the above approaches are supervised and require sufficient labeled training data. However, in our setting we wish to predict the next individual location/state occurring in well-defined workflows, rather than categorize sequences of states. Moreover, we do not have labeled training data.

Trajectory Mining. In terms of analytics of location traces, trajectory pattern mining is also related to this work. For instance, [6] introduced trajectory patterns as frequent behaviors in terms of both time and space, where the frequent trajectory patterns are computed based on given thresholds. In [9], methods were proposed to discover periodic patterns from spatio-temporal data, where a periodic pattern is defined as a regular activity which periodically happens at certain locations. Also, [17] proposed methods to discover sequential patterns from imprecise trajectories of moving objects. However, these methods were not developed for indoor spaces, were not designed for the purpose of workflow modeling and, more importantly, the mined frequent patterns cannot provide a parsimonious description of healthcare activities in hospitals, to support the applications we have considered.

Area-of-Interest Detection. Finally, the last category of related work is the detection of area-of-interest, based on trajectory data. For instance, [12] proposed a non-density-based approach, called mobility-based clustering, to identify the hotspots of moving vehicles in an urban area. The key idea is that sample objects serve as "sensors" to perceive vehicle crowdedness in nearby areas using their instant mobility, rather than the "object representatives". Moreover, [20] proposed a stay point concept and identified hotspots from human moving trajectories. One location was considered as a hotspot if several moving objects stay nearby over a given time period. Finally, [6] used the neighborhood function to model Regions-of-Interest. They partitioned space into grids and quantified the interest of each grid by the density/direction information within each grid. As we have discussed, although methods mentioned above are successful for analyzing outdoor location traces, most of them are not applicable to indoor environments and, in particular, hospitals, because of the unique characteristics of indoor space.

3 Preliminaries and Problem Formulation

We first describe the data from indoor location traces and introduce necessary notation. Then, we formalize the problem of indoor workflow modeling.

3.1 Data Description and Transformation

Our location traces (trajectories) of medical devices are collected indoors at several US hospitals. The location trace of an object O is defined as:

Definition 1 (Location trace). *A location trace is a sequence* $(L_1, L_2, \ldots)$, *where* L_i *represents the ith record in the sequence.* $L_i = (start_i, end_i; x_i, y_i, z_i)$ *contains numerical coordinates* (x_i, y_i, z_i) *and the time that record was recorded, where* $start_i$ *and* end_i *are the start and end time, respectively, of* L_i. *In other words, during the time frame from* $start_i$ *to* end_i, *the object stays at the coordinate* (x_i, y_i, z_i) *in a three-dimensional indoor space.*

However, the indoor wireless communication may be interrupted by environmental factors, leading to errors and noise in the localization of moving objects. Therefore, a coordinate localized by the RTLS might not indicate the exact position, but a small area surrounding the coordinate [10]. In addition, it is not meaningful to use the raw coordinates for indoor workflow modeling. For example, although two recorded coordinates may be separated by a small geometric distance on the same floor, the actual moving distance from one coordinate to the other may be very large if there is, e.g., a wall between them.

To cope with these challenges, we normalize the original location traces for workflow modeling. Specifically, we project each raw coordinate to a semantic location of the building, such as a room in the hospital, based on the floor maps of the building. For the data and maps in this study, each hallway is also treated as a room and some long hallways have been segmented into multiple smaller rooms. Then, L_i can be transformed to: $L_i = (start_i, end_i, r_i)$, where r_i is the room containing the coordinate (x_i, y_i, z_i). After this projection, two neighboring coordinates of the raw location traces may be mapped into the same room. In other words, r_i for L_i may be the same as r_{i+1} for L_{i+1}. We therefore merge runs of consecutive records with the same room into one record. Specifically, if $i < j$ and $r_i = r_{i+1} = \cdots = r_j$, we replace the subsequence $(L_i, L_{i+1}, \ldots, L_j)$ with only one record $L_i^* = (start_i, end_j, r_i)$.

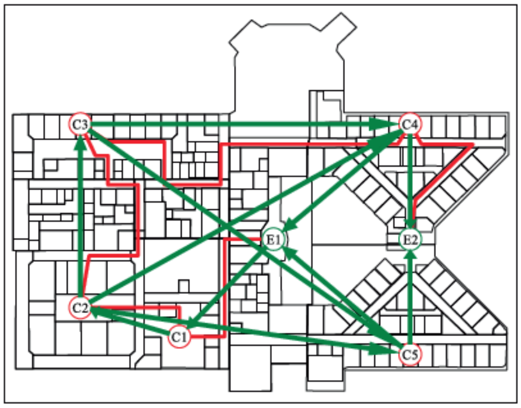


Fig. 3. Workflow instances of infusion pump.

Now, each raw numerical location is mapped to a symbolic room representation (graph node), and each location trace is transformed to a symbolic sequence (traveling path). This data pre-processing step greatly smooths out the noise and alleviates the impact of errors on subsequent workflow modeling tasks (described next). Additionally, it also drastically reduces the computational cost, since we significantly reduce the number of records in the data.

3.2 Concepts for Workflow Modeling

3.2.1 *Workflow Modeling in Healthcare Environments*

Our goal of workflow pattern modeling is to automatically summarize the work activities in a systematic manner. To this end, the concept of workflow patterns is actually a hierarchy with several levels. At the lowest level, the location trace can be seen as a workflow pattern instance. For example, in Fig. 3, we show the location trace of an infusion pump (red line). However, it is difficult to semantically understand the pattern hidden in the raw location traces.

At the other extreme, three general high-level workflow stages can be used to describe workflow logistics: *preprocessing maintenance stage*, *in-use stage*, *postprocessing maintenance stage*. The *in-use stage* of a medical device corresponds to the period when it is used for any healthcare purpose. Before the *in-use stage*, a device is in the *preprocessing maintenance stage*, e.g., held in the storage room. After the *in-use stage*, the device must go through the *postprocessing maintenance stage* before the next use cycle. These maintenance processes include cleaning up, sterilization, disinfection, etc. However, modeling workflow patterns based on the generic stages is too coarse to provide useful results that can help with healthcare workflow management. We need a *middle-level representation* of the location traces, which leads to our proposed workflow models.

3.2.2 *General Location Based Workflow Modeling*

Indeed, as we discussed in the Introduction, we expect indoor location tracking data to be more widely available in

TABLE 1
Annotation of Key Locations in Fig. 3

Locations	Functions
C_1	Post-anesthesia care unit (PACU)
C_2	Operating room (OR)
C_3	Intensive care unit (ICU)
C_4, C_5	Patient care unit (PCU)
E_1, E_2	Elevator

environments beyond hospitals (e.g., shops, schools, warehouses, etc), and the automatic workflow modeling with the indoor location data can be applied to discover different types of human activity patterns in various of scenarios. To this end, it's a ubiquitous challenge to identify the right representation level of the location traces for better workflow modeling and understanding. The raw location traces include too much unnecessary minutia, while the pure semantical high-level workflow stages are to coarse to encode the spatial information. Therefore, one of our objectives in developing the indoor location based workflow models is to automatically construct the *workflow states*, which are spatially contiguous as well as semantically meaningful. In the following, we first discuss how to construct the workflow states as the right representation levels in the healthcare environments. We also highlight the assumptions used by our models. Therefore, the proactive workflow modeling framework is applicable to general domains where the assumptions can be accepted and the data is available.

3.2.3 Workflow States

To better model and understand the workflow processes, we define the workflow states as a few key areas in these trajectories. Our goal is to automatically generate these workflow states and use them as "annotations" of the original location traces. For example, we annotate the location traces in Fig. 3 with areas C_i ($i = 1, \ldots, 5$). The medical functions of these areas are summarized in Table 1. The location trace in red in Fig. 3 can now be concisely represented as $C_1 \mapsto C_2 \mapsto C_3 \mapsto C_4$. Such representation with the key areas makes it easy to understand the workflow underlying the location traces.

In fact, Fig. 3 is the map of the second floor of a hospital building, which is centered around E_1, the elevator connected to the basement. The red location trace of one infusion pump starts from the storage room in the basement to the elevator E_1. After E_1 follows C_1, which is the Post-Anesthesia Care Unit (PACU), where the patient is ready for medical procedures, such as surgery, and the medical devices are attached to the patients. Next, C_2 is the area of operating rooms (OR), where medical procedures are performed. After the medical procedure, the patients and the medical devices are moved to C_3, the Intensive Care Unit (ICU), based on medical needs. When the condition of the patient becomes stable, the patient and the medical devices being used will be further moved to C_4, the Patient Care Unit (PCU), before the patient is discharged. After the patient is discharged, the medical devices will be moved through elevator E_2 to the basement, cleaned in a disinfection room, and then transferred into storage.

It should be noted that Fig. 3 is a grossly simplified view of reality. In fact, we may have many workflow states, spanning multiple floors and buildings, and the overall workflow patterns are much more complex. Formally, we define the workflow states (in general indoor contexts) as follows:

Definition 2 (Workflow state). *A workflow state is an area in the indoor space where specific workflow activities frequently happen.*

From the above example, it is clear that each such area (where a particular stage of the process occurs) should correspond to a workflow state, and that location traces represented using such workflow states convey the necessary information for easily understanding and modeling workflow patterns. In other words, workflow patterns can be understood much more clearly when location traces are represented using workflow states, rather than by using either low-level rooms or high-level stages.

3.3 Problem Statement of Workflow Modeling

A further task in workflow modeling is to summarize the transition patterns of the moving objects among workflow states. As shown in Fig. 3, the transition from one state may lead to several different states. For example, when the situation of a patient is stable after the medical operation at C_2, the medical devices and the patient might be moved directly to PCU (C_4, C_5) without passing through ICU (C_3).

Now, we can formally state the problem of location-based workflow pattern modeling in general indoor environments:

Problem 1 (Workflow pattern modeling). *Given the location traces of moving objects (e.g., medical devices in hospitals, shoppers in malls, loading trucks in warehouses, etc), workflow pattern modeling discovers workflow knowledge including workflow states and parameters describing the transition patterns of the moving objects.*

In practice, we have multiple types of objects moving around indoors, and one workflow model can be built for each of the object types. The multiple models are not independent with each other. For instance, in hospitals, many different types of medical devices are often used together for a particular task. Therefore, although different types of medical devices have different workflow patterns, there is some natural correlation among their location traces. Modeling such correlation not only helps reinforce the robustness of the workflow models, but also provides better understanding of the overall workflow patterns.

4 Methodology

Although the workflow states are not directly visible, the observed location traces depend on the hidden workflow states. Therefore, our Proactive Workflow Model (ProWM) uses the Hidden Markov Model (HMM) to construct the workflow states. In a HMM, each state has a probability distribution over the possible output observations. With the Markov assumption, the state distribution can be estimated with the observed data. In the following, we first introduce necessary notation for hidden Markov modeling of location traces, then present our regularization schemes to obtain spatially contiguous (thus more interpretable) states, and to

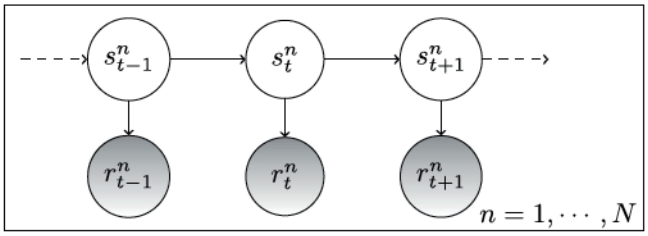


Fig. 4. The graphical Hidden Markov Model. White nodes $s^n_t \in S$ are hidden workflow states; Shaded nodes $r^n_t \in R$ are observed room sequence.

leverage correlations among devices for multi-flow estimation. Our regularization is intuitive, and can be viewed as a prior over the model parameters. The parameters in the regularized models can also be estimated effectively.

We observe a set of room sequences $X = \{x^n : n = 1, \dots, N\}$, where $x^n = (r^n_1, \dots, r^n_{T_n})$ is the sequence for the nth moving object, T_n is the length of x_n, and N is the number of all moving objects. Each observation $r^n_t \in R$ is associated with a workflow state $s^n_t \in S$. Here, R is the set of rooms in the building and S is the set of workflow states. Without loss of generality, we assume $R = \{1, 2, \dots, K_R\}$ and $S = \{1, 2, \dots, K_S\}$, where K_R is the number of rooms and K_S is the number of states. Hidden Markov Models assume that, as shown in Fig. 4, when state s^n_t is given, the corresponding observation r^n_t is independent to all other states s^n_l for $l \neq t$, and the state s^n_{t+1} is independent to all previous states s^n_l for $l < t$. Then, the model is fully determined by the state transition probabilities $A_{ij} = \Pr(s^n_{t+1} = j | s^n_t = i)$, the emission probabilities $B_{ij} = \Pr(r^n_t = j | s^n_t = i)$, and the initial state distribution $\pi_i = \Pr(s^n_1 = i)$.

Probability parameters $\Theta = (A, B, \pi)$ can be estimated by maximizing the log-likelihood $\mathcal{L}(\Theta|X) = \log \Pr(X|\Theta)$, e.g., the Baum-Welch algorithm. We improve the model estimation using the maximum a posteriori (MAP) estimation:

$$\Pr(X, \Theta) \propto \Pr(X|\Theta)\exp(-\lambda\Omega(\Theta)).$$

In the remaining of this section, we define the prior regularization $\Omega(\Theta)$.

4.1 Regularized Workflow States

In practice, the identified workflow states should be spatially contiguous. For example, we want to identify areas of semantically meaningful locations, such as '2nd floor northeast patient rooms' and 'basement central storage rooms'. In this section, to encourage each workflow state to be a contiguous region in the building, we use the proximity between rooms to define a prior on the workflow state distribution. Specifically, let G be the graph where nodes represent rooms and edges signify that these two rooms communicate with each other. We also use G as its adjacency matrix. We define the regularization

$$\Omega_B(B) = \frac{1}{2}\sum_s \sum_{ij} G_{ij}(B_{si} - B_{sj})^2 = \mathrm{tr}(B(D - H)B'),$$

where $H = \frac{1}{2}(G + G')$ is a symmetric graph and D is the diagonal degree matrix of H, with $D_{ii} = \sum_j H_{ij}$. Intuitively, we use $\Omega_B(B)$ to encourage neighboring rooms to have similar probabilities in the state distribution. Note that, $D - H$ is always positive-semidefinite and $\Omega_B(B) \geq 0$.

The graph G can be user-provided or automatically defined given the historical location traces of the moving objects. In this study, we let $G_{ij} = 1$ if and only if there exists a trace $(r^n_1, \dots, r^n_{T_n})$ and $1 \leq t < T_n$, such that $r^n_t = i$ and $r^n_{t+1} = j$. In other words, we let $G_{ij} = 1$ if one can transit from room i to j directly, and $G_{ij} = 0$ otherwise. Such a definition avoids the use of geometric distance, which is not particularly meaningful in indoor environments, as discussed earlier. Moreover, our definition also avoids the use of parameters, such as distance thresholds.

4.2 Adaptive Multiflow Estimation

For healthcare procedures on one patient, multiple types of medical devices are often needed at the same time and place. Thus, these different types of moving objects transit together, therefore following correlated workflow models, which can be estimated jointly. To this end, first, we can use a common workflow state distribution $B = B^{(k)}$, $k = 1, \dots, K$, for all the K types of moving objects. Indeed, using a common state distribution for all types also reduces the model complexity dramatically.

Second, to estimate the K workflow transition matrices $\{A^{(k)} | 1 \leq k \leq K\}$, we borrow the idea of mean-regularized multi-task learning [3]. Specifically, we assume that each $A^{(k)}$, for $1 \leq k \leq K$, can be written as

$$A^{(k)} = A^{(0)} + D^{(k)}$$

where the difference $D^{(k)} = A^{(k)} - A^{(0)}$ between $A^{(k)}$ and $A^{(0)}$ is 'small' when the workflow transitions of different types of moving objects are similar to each other. Therefore, define the regularizer as

$$\Omega_A(A) = \frac{1}{2}\sum_{k=1}^{K} \|D^{(k)}\|^2 = \frac{1}{2}\sum_{k=1}^{K} \|A^{(k)} - A^{(0)}\|^2$$

One can show that the optimal $A^{(0)}$ minimizing the regularizer $\Omega_A(A)$ is

$$A^{(0)} = \frac{1}{K}\sum_{t=1}^{K} A^{(t)},$$

and thus we have

$$\Omega_A(A) = \frac{1}{2}\sum_{k=1}^{K} \|A^{(k)} - \frac{1}{K}\sum_{t=1}^{K} A^{(t)}\|^2. \tag{1}$$

In other words, mean-regularized multi-task learning assumes that the transitions are related in a way that the true estimates are all close to the mean $A^{(0)}$. Note that, although in this work we consider only the mean-regularized multi-task learning for the multiflow estimation, general formulations can be easily adopted when more knowledge is available. For example, if we had pairwise correlations between the workflow models of different types of objects, e.g., the correlation ρ_{ij} between the type i and the type j, we can define the 'multiflow' regularization as

$$\Omega_A(A) = \frac{1}{2}\sum_{ij} \rho_{ij} \|A^{(i)} - A^{(j)}\|^2. \tag{2}$$

5 Implementation

With the regularizations $\Omega_A(A)$ and $\Omega_B(B)$, the objective function to simultaneously estimate all the workflow models is

$$\begin{aligned}&\mathcal{J}(A^{(1)},\dots,A^{(K)},B,\pi^{(1)},\dots,\pi^{(K)})\\&=\sum_k \mathcal{L}(A^{(k)},B,\pi^{(k)})-\Omega(\Theta)\end{aligned}\tag{3}$$

where $\mathcal{L}(A^{(k)},B,\pi^{(k)})=\log \Pr(r^{(k)}|A^{(k)},B,\pi^{(k)})$ and

$$\Omega(\Theta)=\lambda_1\Omega_A(A)+\lambda_2\Omega_B(B).$$

We maximize $\mathcal{J}(\Theta)$, where the parameters are $\Theta=(A^{(1)},\dots,A^{(K)},B,\pi^{(1)},\dots,\pi^{(K)})$, using the regularized Baum-Welch algorithm, which iteratively applies the following two steps. Based on the current value of $\hat{\Theta}$, the first step is to compute the probabilities:

$$\begin{aligned}\gamma_i^{(k)}(t)&=\Pr(s_t^{(k)}=i|r^{(k)})\\ \xi_{ij}^{(k)}(t)&=\Pr(s_t^{(k)}=i,s_{t+1}^{(k)}=j|r^{(k)})\end{aligned}$$

and define

$$\Gamma_{ij}^{(k)}=\sum_{t=1}^{T_k}\gamma_i^{(k)}(t)[r_t^{(k)}=j],\qquad \Xi_{ij}^{(k)}=\sum_{t=1}^{T_k-1}\xi_{ij}^{(k)}(t).$$

Note that, here we assume that there is only one observation sequence $r^{(k)}$ of the kth type of the moving objects. Notations for more general case would be more complicated but leading to the same computation.

The second step of this Expectation Maximization (EM) algorithm computes the new parameter value Θ, by maximizing the expected value of the objective function in Equation (3)

$$\begin{aligned}Q(\Theta)&=\sum_k \mathrm{E}_{s^{(k)}|r^{(k)},\hat{\Theta}}\log \Pr(s^{(k)},r^{(k)}|\Theta)-\Omega(\Theta)\\&=\sum_{s^{(k)}}\Pr(s^{(k)}|r^{(k)},\hat{\Theta})\log \Pr(s^{(k)},r^{(k)}|\Theta)-\Omega(\Theta)\end{aligned}$$

where the log-likelihood is given by

$$\begin{aligned}&\log \Pr(s^{(k)},r^{(k)}|\Theta)\\&=\log \pi^{(k)}_{s_1^{(k)}}+\sum_{t=1}^{T_k-1}\log A^{(k)}_{s_t^{(k)}s_{t+1}^{(k)}}+\sum_{t=1}^{T_k}\log B_{s_t^{(k)}r_t^{(k)}}\end{aligned}$$

It follows that $\pi^{(k)}=\arg\max\sum_i\gamma_i^{(k)}(1)\log\pi_i^{(k)}$, where the solution is $\pi_i^{(k)}=\gamma_i^{(k)}(1)$. For updating of A and B, we have the following results:

Proposition 1 (Optimization subproblem of A). *Given the current value of $(B,\pi^{(1)},\dots,\pi^{(K)})$, the problem of updating $(A^{(1)},\dots,A^{(K)})$ to optimize $Q(\Theta)$ is equivalent to:*

$$\begin{aligned}A&=\operatorname{argmax}\sum_k\sum_{ij}\sum_{t=1}^{T_k-1}\xi_{ij}^{(k)}(t)\log A_{ij}^{(k)}-\lambda_1\Omega_A(A)\\&=\operatorname{argmax}\sum_k\sum_{ij}\Xi_{ij}^{(k)}\log A_{ij}^{(k)}-\lambda_1\Omega_A(A)\\ \text{s.t.}\quad & A\ge 0,\quad \sum_j A_{ij}^{(k)}=1,\forall k,i.\end{aligned}$$

Proposition 2 (Optimization subproblem of B). *Given the current value of $(A^{(1)},\dots,A^{(K)},\pi^{(1)},\dots,\pi^{(K)})$, the problem of updating B to optimize $Q(\Theta)$ is equivalent to:*

$$\begin{aligned}B&=\operatorname{argmax}\sum_k\sum_{ij}\sum_{t=1}^{T_k}\gamma_i^{(k)}(t)[r_t^j=j]\log B_{ij}-\lambda_2\Omega_B(B)\\&=\operatorname{argmax}\sum_k\sum_{ij}\Gamma_{ij}^{(k)}\log B_{ij}-\lambda_2\Omega_B(B)\\ \text{s.t.}\quad & B\ge 0,\quad \sum_j B_{ij}=1,\forall i.\end{aligned}$$

Note that, $A,B\ge 0$ are element-wise constraints. Since $\lambda_1,\lambda_2\ge 0$, both problems in Proposition 1 and 2 are concave and we have efficient algorithms (e.g., gradient ascent) to update A and B. Particularly for the special case when $\lambda_1=\lambda_2=0$, we have:

$$A_{ij}^{(k)}=\frac{\Xi_{ij}^{(k)}}{\sum_j\Xi_{ij}^{(k)}},\qquad B_{ij}=\frac{\sum_k\Gamma_{ij}^{(k)}}{\sum_k\sum_j\Gamma_{ij}^{(k)}}.$$

5.1 Choosing the Number of Workflow States

One important aspect of training ProWM is how to choose an appropriate number of latent workflow states. A commonly used approach for estimating the latent states of HMMs is to leverage domain knowledge or some existing algorithms to pre-cluster the observations [21]. In our problem, we shall pre-cluster the indoor location records to guide the training of ProWM. To this end, we adopt the density-based clustering algorithm proposed in [11], which computes the neighborhood of a location record as well as its density based on the historical location traces. Using this algorithm, the number of clusters can be automatically determined, and the detected clusters can be of different densities and arbitrary shapes. More importantly, the detected clusters are spatially contiguous, and therefore can be semantically annotated, such as '2nd floor northeast patient rooms' and 'basement central storage rooms'.

In this paper, since we use a common emission matrix in Equation (3), we accordingly piece together the pre-cluster solutions for all the K types of moving objects:

$$C=\cup_{k=1}^K C^{(k)}$$

where $C^{(k)}$ is the pre-cluster solution of the kth type of moving object. Note that, when $C_1\in C^{(1)}$ and $C_2\in C^{(2)}$ overlap, i.e., $C_1\cap C_2\neq\emptyset$, we merge them together as $C_1\cup C_2\in C$. Now, we have the common pre-cluster solution expressed in C and can determine the number of workflow states $|C|$.

5.2 Non-Essential Space Reduction

In addition to choosing the number of latent states in our ProWM, the results of pre-clustering can also be used for accelerating parameter estimation. In reality, not every room in the building is associated with healthcare activities and not every movement in the devices' trajectories needs to be modeled. For example, we may have many location records in front of the elevator. Although many trajectories may briefly go through that location, no healthcare activities occur at such non-essential locations. With the pre-clustering solution, we

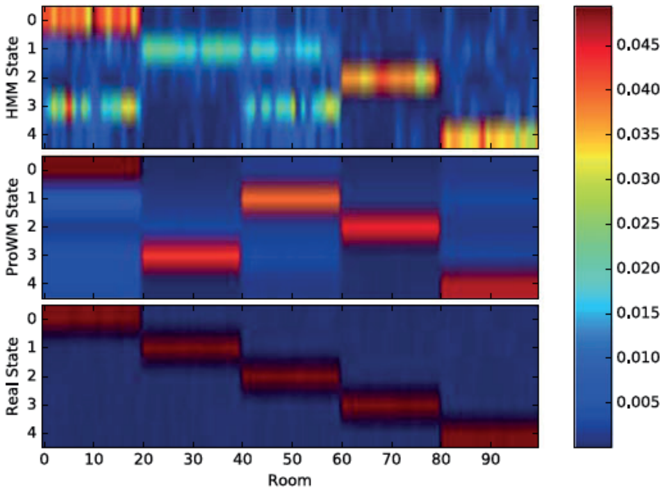


Fig. 5. The workflow states of HMM and ProWM.

can exclude the non-essential observations in the indoor space, to reduce the modeling complexity. As shown in [11], the density-based pre-clustering algorithm can effectively identify the non-essential locations as outliers in the clustering results. Intuitively, these low-density[1] outliers consist of short-duration stays at one or more locations, whereas healthcare activities happen mostly when the medical devices are stationary. Therefore, our implementation filters out the non-stationary locations from the workflow modeling. Specifically, the union $\cup_{c \in C} c$ is a proper subset of the indoor space and, accordingly, we model only the observations $r_t^k \in \cup_{c \in C} c$, reducing the modeling complexity without compromising the modeling quality of the workflow patterns.

6 Empirical Evaluation

This section evaluates the performance of our workflow models. We first use the synthetic data to show the effectiveness of our ProWM methods. We then apply our models on real-world data collected in indoor healthcare environments.

6.1 Synthetic Data and Results

We simulate workflow processes with known state distributions and transition patterns to demonstrate the effectiveness of ProWM. First, we define five workflow (hidden) states where each distributes uniformly on 20 rooms (symbols), with emission probability 0.05 on each room. Therefore we have a block structure in the emission matrix B. We inject Gaussian noises $\mathcal{N}(0, 0.01)$ in the emission matrix. Then, we simulate 200 sequences each with length 100 (total 20,000 observations). With the simulated data, the hidden states identified by naive HMM method is shown in Fig. 5 (top). As can be seen, the naive HMM method can only partially uncover the ground-truth structures of the hidden state. In comparison, as shown in Fig. 5 (middle), our regularized ProWM method can (almost) perfectly construct the (permuted) true workflow states in Fig. 5 (bottom).

To test the effectiveness of the multiflow regularization (Equation (2)), we randomly generate two transition patterns shared by five workflow processes with transition matrices $A^{(k)}$, $k = 1, \ldots, 5$. The two transition patterns are

1. Density also incorporates time duration of stay.

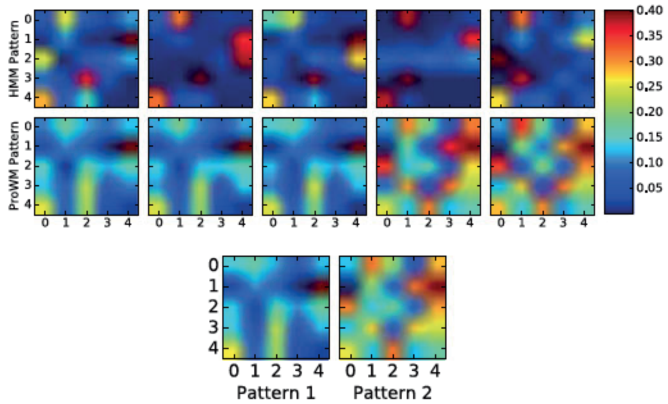


Fig. 6. The workflow transitions of HMM and ProWM.

shown in Fig. 6 (bottom), where Pattern 1 is shared by the first three processes and Pattern 2 is shared by the last two. The transition matrix of each process is generated by injecting $\mathcal{N}(0, 0.01)$ to the associated pattern. Then we simulate 100 sequences (with length 100) for each workflow process. As shown in Fig. 6 (top), the transition patterns identified by five independent HMMs are not capturing the designed Pattern 1 and Pattern 2. The reason is that the HMM optimization is highly non-convex with difficulty to find the designed optimum in the optimization space. In contrast, with our multiflow regularization, ProWM can successfully approximate the ground-truth in all workflow processes.

In addition to above visual comparisons, ProWM also increases the likelihood of the model parameters. The increases are 72 and 31 percent per sequence in Fig. 5 and 6, respectively.

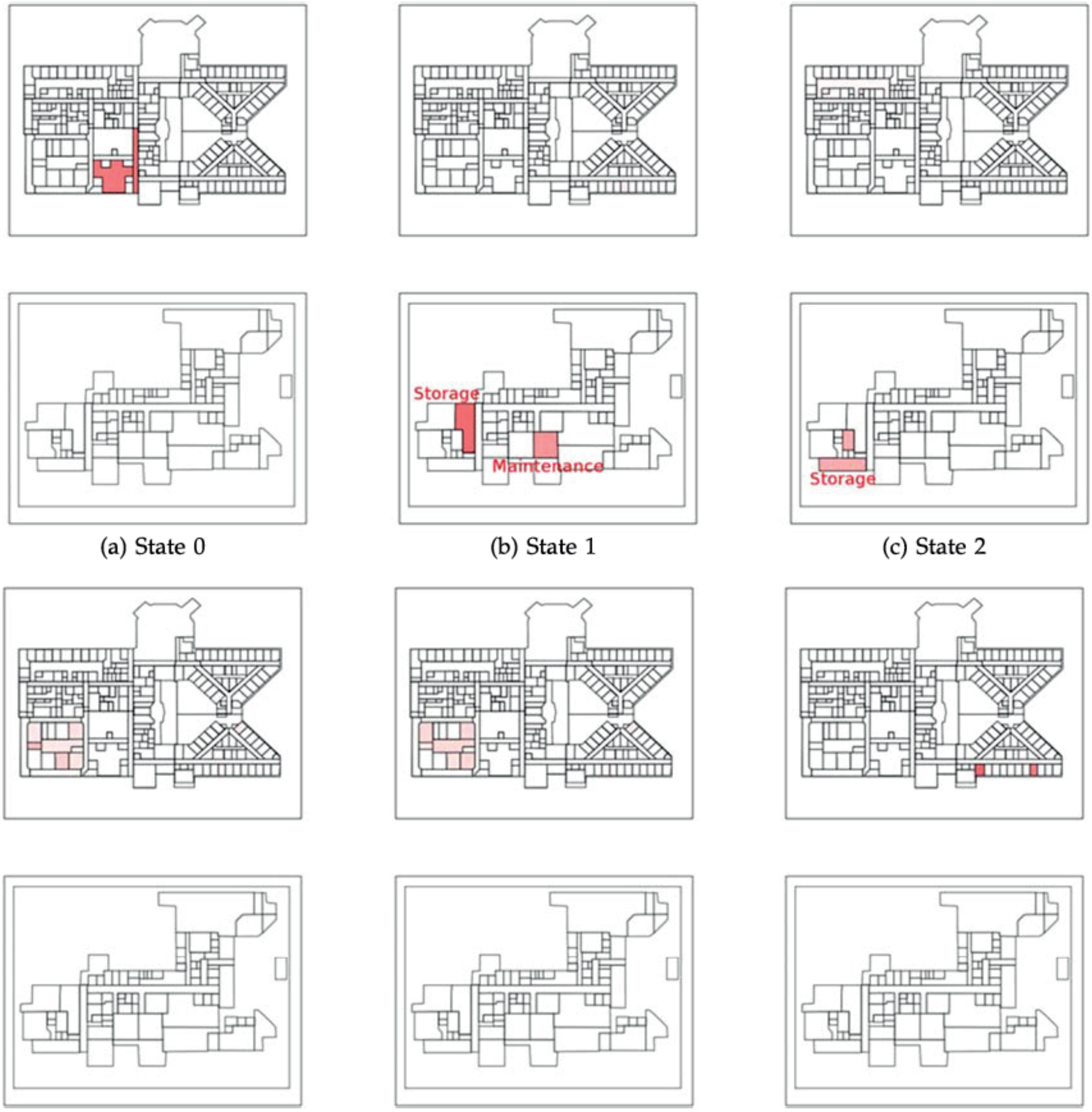
6.2 Indoor Location Data

Our real-world indoor location data sets are collected from several hospitals in the US. Sensor tags are attached to medical devices operated in these hospitals, allowing them to be tracked by real-time location systems. Table 2 shows basic statistics of the data collected for various types of medical devices in Hospital 1. The second and third columns show the number of medical devices of each type operated in this hospital, and the number of location records collected during the period from January 2011 to August 2011.

In the remaining of this section, we first illustrate the workflow states identified by ProWM. We then analyze the goodness-of-fit of ProWM, with respect to important statistics. Particularly, we show that the regularized workflow states are semantically meaningful and easily amenable to interpretation; and the adaptive multiflow estimation can reinforce the modeling performance, especially for real-time applications.

TABLE 2
Data Statistics in Hospital 1

Type	#Objects	#Locations
Wheelchair	121	524415
PCA II Pump	66	4431
Venodyne	403	1588045
Feeding Pump	83	157370
PCA Pump	136	231057
ETCO2	137	220380



(a) State 0 (b) State 1 (c) State 2

(d) State 3 (e) State 4 (f) State 5

Fig. 7. Unregularized workflow states. Issues: 1) Several states are not continuous in the indoor space, e.g., States 1, 2, and 5. Especially, State 2 covers rooms in two floors. 2) States 3 and 4 are effectively duplicates.

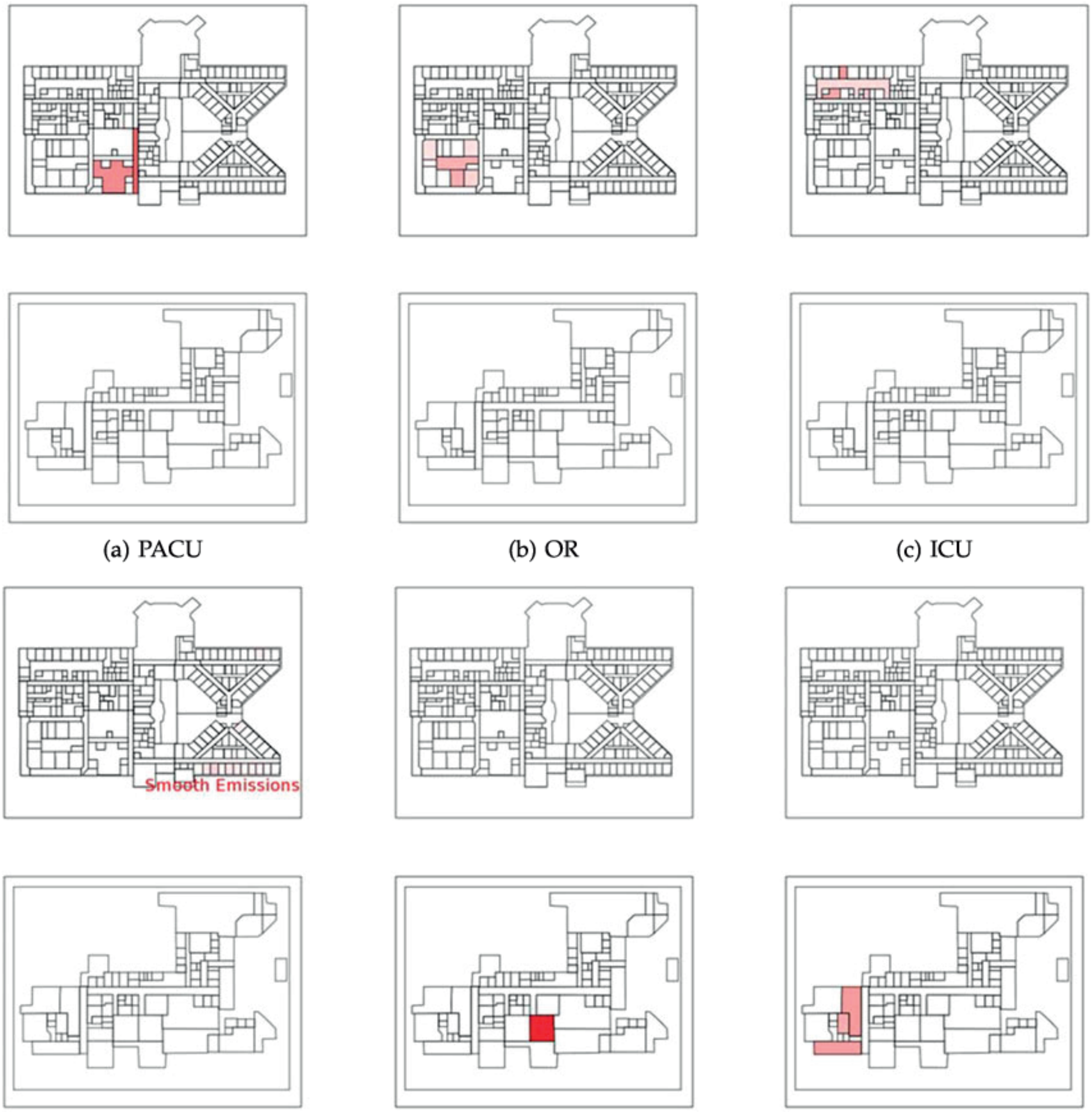
6.3 Workflow States

In Equation (3), we set $\lambda_1 = 0$ and select an appropriate value for $\lambda_2 = 10^v$ where $v \in [-5, 5]$, to construct interpretable workflow states. Generally, the larger the λ_2, the more continuous the states. Fig. 7, illustrates the unregularized workflow states on two floors in Hospital 1. The figures at the lower level are maps of the basement, and the figures at the upper level are maps of the second floor. In each map, rooms are shaded based on the emission probabilities ($> 10^{-5}$) for each state distribution (darker indicates higher probability of a room belonging to that state). Several issues are evident in the results. First, some states are not contiguous in the indoor space, e.g., State 1, 2 and 5. State 2 even covers rooms in two different floors. Second, States 3 and 4 are duplications of each other.

In contrast, as shown in Fig. 8, these issues can be effectively addressed by the regularization we proposed in Section 4.1. Not only the regularized workflow states are contiguous in the indoor space, but they can also be easily matched to known room usage, e.g., Post-anesthesia care unit (PACU), Operating room (OR), Intensive care unit (ICU), Patient care unit (PCU), Device Maintenance and Device Storage. For example, in the unregularized workflow states, the left highlighted area in the basement of State 1 and the highlighted area in the basement of State 2 are actually neighbors and all are used for storage. The central highlighted area in the basement of State 1 has different usage, which is device maintenance. On the other hand, in the regularized workflow state we have clearly identified the Storage and Maintenance states without any supervision. Also, regularized workflow states have smooth emissions in contiguous areas of the building (e.g., compare the PCU state in Fig. 8 with State 5 in Fig. 7).

6.4 Goodness-of-Fit

We measure the goodness-of-fit of the learned workflow models by computing the *average log-loss* with test location traces. For a test location sequence $Tr = (r_1, \ldots, r_T)$, the average log-loss is



(a) PACU (b) OR (c) ICU

(d) PCU (e) Maintenance (f) Storage

Fig. 8. Regularized workflow states can be easily interpreted with semantics as follows: Post-anesthesia care unit (PACU), Operating room (OR), Intensive care unit (ICU), Patient care unit (PCU), Device Maintenance, and Device Storage.

$$\ell(Tr) = -\frac{1}{T}\log \Pr(Tr|\Theta). \quad (4)$$

We randomly partition the data into ten subsets and compute the average log-loss in ten rounds. In each round, we use nine of these subsets as training data to estimate the model parameters and compute $\ell(Tr)$ for each Tr in the remaining test data.

We set λ_2 based on results in Section 6.3, then choose $\lambda_1 = 10^v$ with $v \in [-5, 5]$ that achieves the lowest average log-loss on validation data. The *solid red lines* in Fig. 9 show $\ell(Tr)$ with respect to the length of Tr. We also show the results with *solid green lines* for the baseline models, using $\lambda_1 = 0$ in Equation (3). Fig. 9 shows that the regularized multiflow models consistently achieve lower information loss across different types of moving objects.

For certain workflow managerial applications, we have to estimate the models in real-time with limited training data. In this case, our models substantially outperform baselines by unified estimation. This is clearly demonstrated by repeating the above comparison using less training data. As shown by the *dashed lines*, the improvement of the regularized multiflow models is more significant compared to baselines. Moreover, the regularized models are more robust, without sharp jumps in the plot.

In Fig. 9, we also show the performances of the baseline methods developed by [11] in *blue lines*. The results demonstrate that goodness-of-fit is significantly improved by integrating workflow construction and transition estimation.

6.5 Computational Efficiency

In this subsection, we evaluate the performance of training ProWM. In particular, we first choose the number of workflow states based on the pre-clustering of location records. We also use the pre-clustering solution to reduce the modeling space, as introduced in Section 5.2. After that, we run the alternative optimization procedure to compute the optimal transition probabilities A^k for $k = 1, \ldots, K$ and the workflow emission probabilities B. We compare the training time against two baselines: 1) The same alternative optimization

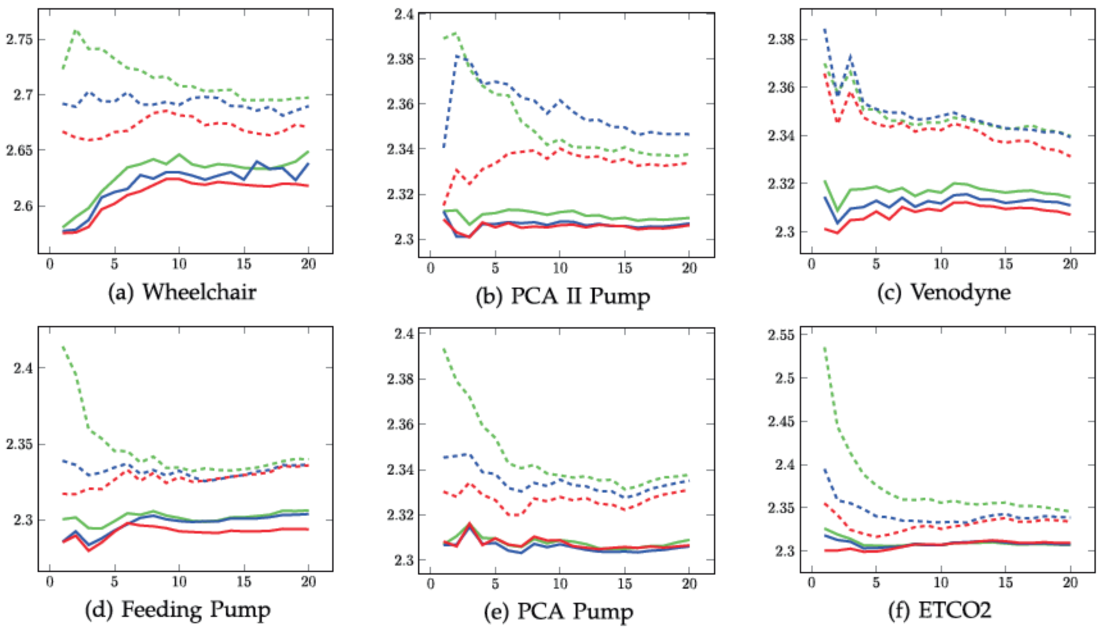


Fig. 9. The comparison of average log-loss. Red: ProWM with regularized multiflow models. Green: ProWM without multiflow regularization. Blue: Baselines developed by [11]. Solid: 90 percent training and 10 percent testing. Dashed: 10 percent training and 90 percent testing.

procedure without non-essential space reduction. 2) The model developed in [11].

Two interesting observations follow from the comparison of running time and information loss in Table 3. First, by reducing the non-essential modeling space, we can significantly reduce computation complexity without decreasing modeling performance. The model with reduced space has average log-loss comparable to that of the model using the full space. Second, in the scenario with substantially limited training data, the running time of the proposed ProWM model with reduced modeling space is close to that of the baseline model developed in [11]. However, our ProWM model has significantly better prediction performance ($2.399 < 2.689$). In other words, we reduce the average log-loss by 11 percent while maintaining computational efficiency.

6.6 Application: Workflow Monitoring

The learned workflow model is valuable, since a range of practical problems can benefit from the modeling results. Indeed, we have implemented a management information system, HISflow, to exploit the discovered knowledge for healthcare operation and management. A screenshot of HISflow is shown in Fig. 10; for techniques used in our implementation, see [11]. In the following, we elaborate on one particular application, workflow monitoring.

TABLE 3
Comparison of the Running Time of Different Methods in Two Scenarios

Model	Training:Testing	Running Time (s)	Information Loss
ProWM (reduced space)	9:1	238	2.270
	1:9	31	2.399
ProWM (full space)	9:1	1037	2.267
	1:9	419	2.401
CTMC [11]	9:1	71	2.401
	1:9	29	2.689

In one scenario we have sufficient training data, with a ratio of training:testing data equal to 9:1. In the other scenario the ratio is 1:9.

When the extracted workflow patterns are mapped to specific procedure codes (either existing or new), we can identify abnormal behavior within daily healthcare activities in real-time. When such anomalies occur, warnings or alerts can be activated by the management system, potentially helping reduce the risk of faults or accidents in healthcare services. One approach to develop such a system is to rank the ongoing trajectories of all monitored medical devices based on the average $\log$-loss in Equation (4). In this way, the top-ranked devices merit more scrutiny. However, these ranking results may not be intuitive from the management perspective. In fact, it is vital to provide more insights into the causes behind the higher $\log$-loss trajectories. To this end, given the recent trajectory $Tr = (r_1, r_2, \ldots, r_T)$ which is ranked in the top, we compute the most likely corresponding state sequence $(s_1, s_2, \ldots, s_T)$ and identify the time of abnormal events $\hat{T} = \max_t \ell(r_{1:t})$, as well as the description of each observation for $t = 2, \ldots, \hat{T}$:

1) *Unlikely transition:* If $s_t \neq s_{t-1}$ and $-\log A_{s_{t-1}s_t} \geq -\log B_{s_t r_t}$.

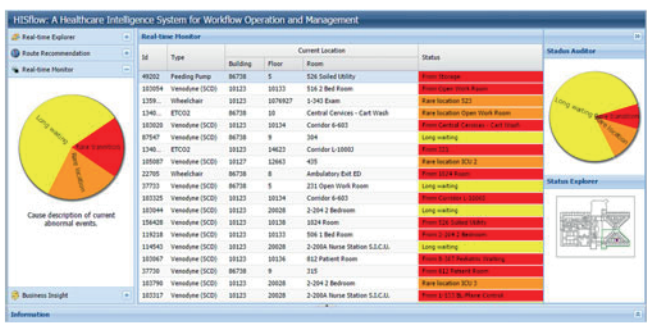


Fig. 10. The screenshot of HISflow.

2) *Long wait:* If $s_t = s_{t-1}$ and $-\log A_{s_{t-1}s_t} \geq -\log B_{s_t r_t}$.
3) *Unlikely location:* If $-\log B_{s_t r_t} \geq -\log A_{s_{t-1}s_t}$.

Intuitively, *unlikely transition* indicates that the state transition from s_{t-1} is unlikely to end at s_t, and the likelihood of the transition contributes significantly to the increase of average $\log$-loss $\ell(Tr)$. The *Long wait* case is similar, except that the workflow states are unchanged. Finally, in the case of *unlikely location*, we observed the location r_t for the state s_t, however, the location r_t is unlikely to be emitted by the state s_t and the likelihood of the emission contributes more to the increase of $\ell(Tr)$, compared to other explanations.

As can be seen in Fig. 10, based on the average $\log$-loss, we can effectively identify the abnormal traces with detailed explanations. Specifically, we highlight the explanation for each identified abnormal device with different background color in the 'Status' column of the list. When the user clicks one item in the list, we show more details in the right panel (e.g., the distribution of different recent causes for the anomaly, and the current location of the device). In the left panel, we also show the aggregated distribution of different anomaly causes in the identified list. Such an intuitive and interpretable real-time monitor system is valuable for the hospital managers to improve the quality of healthcare services.

7 Concluding Remarks

In this paper, we leveraged location traces of medical devices to model the healthcare workflow patterns in hospital environments. Specifically, we developed a stochastic process-based framework, which provides parsimonious descriptions of long location traces. This framework provides new opportunities to concisely understand the logistics of a large hospital. From a technical perspective, we proposed a unified modeling approach based on a novel regularized HMM method, that produces interpretable states and leverages correlations between devices for improved robustness. From an application perspective, the discovered knowledge, such as workflow states and transition patterns can be integrated into management information system we developed. With this system, we showed that valuable intelligent applications for healthcare operation and management can be enabled to manage, evaluate and optimize the healthcare services. Extensive experimental results on both the synthetic data and the real-world data validated the effectiveness of our proposed work.

Acknowledgments

This research was partially supported by the US National Science Foundation (NSF) via the grant number IIS-1648664. Also, it was supported in part by the Natural Science Foundation of China (71329201, 71531001).

References

[1] R. Agrawal, D. Gunopulos, and F. Leymann, "Mining process models from workflow logs," in *Proc. Adv. Database Technol.*, 1998, pp. 467–483.
[2] X. Chai and Q. Yang, "Multiple-goal recognition from low-level signals," in *Proc. Nat. Conf. Artificial Intell.*, 2005, Art. no. 3.
[3] T. Evgeniou and M. Pontil, "Regularized multi–task learning, " in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 109–117.
[4] Y. Ge, H. Xiong, Z.-H. Zhou, H. Ozdemir, J. Yu, and K. C. Lee, "Top-eye: Top-k evolving trajectory outlier detection," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage.*, 2010, pp. 1733–1736.
[5] Y. Ge, H. Xiong, C. Liu, and Z.-H. Zhou, "A taxi driving fraud detection system," in *Proc. 11th IEEE Int. Conf. Data Mining*, 2011, pp. 181–190.
[6] F. Giannotti, M. Nanni, D. Pedreschi, and F. Pinelli, "Trajectory pattern mining," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2007, pp. 330–339.
[7] G. Greco, A. Guzzo, G. Manco, and D. Sacca, "Mining and reasoning on workflows," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 4, pp. 519–534, Apr. 2005.
[8] D. H. Hu and Q. Yang, "Cigar: Concurrent and interleaving goal and activity recognition," in *Proc. 23rd National Conf. Artif. Intell.*, 2008, pp. 1363–1368.
[9] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, "Mining periodic behaviors for moving objects," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 1099–1108.
[10] C. Liu, H. Xiong, Y. Ge, W. Geng, and M. Perkins, "A stochastic model for context-aware anomaly detection in indoor location traces," in *Proc. 12th IEEE Int. Conf. Data Mining*, 2012, pp. 449–458.
[11] C. Liu, Y. Ge, H. Xiong, K. Xiao, W. Geng, and M. Perkins, "Proactive workflow modeling by stochastic processes with application to healthcare operation and management," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 1593–1602.
[12] S. Liu, Y. Liu, L. M. Ni, J. Fan, and M. Li, "Towards mobility-based clustering," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 919–928.
[13] N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D.W. Cheung, "Mining, indexing, and querying historical spatiotemporal data," in *Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2004, pp. 236–245.
[14] E. Ramezani, D. Fahland, and W. M. P. van der Aalst, "Where did i misbehave? diagnostic information in compliance checking," in *Proc. 10th Int. Conf. Bus. Process Manag.*, 2012, pp. 262–278.
[15] S. X. Sun, Q. Zeng, and H. Wang, "Process-mining-based workflow model fragmentation for distributed execution," *IEEE Trans. Syst. Man Cybern. Part A: Syst. Humans*, vol. 41, 2, pp. 294–310, Mar. 2011.
[16] W. Van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 9, pp. 1128–1142, Sep. 2004.
[17] J. Yang and M. Hu, "Trajpattern: Mining sequential patterns from imprecise trajectories of mobile objects," in *Proc. 10th Int. Conf. Extending Database Technol.*, 2006, pp. 664–681.
[18] J. Yin, D. Shen, Q. Yang, and Z.N. Li, "Activity recognition through goal-based segmentation," in *Proc. Nat. Conf. Artif. Intell.* 2005, Art. no. 28.
[19] Q. Zeng, S. X. Sun, H. Duan, C. Liu, and H. Wang, "Cross-organizational collaborative workflow mining from a multi-source log," *Decision Support Syst.*, vol. 54, no. 3, pp. 1280–1301, 2013.
[20] Y. Zheng, L. Zhang, X. Xie, and W. Y. Ma, "Mining interesting locations and travel sequences from GPS trajectories," in *Proc. 18th Int. Conf. World Wide Web*, 2009, pp. 791–800.
[21] H. Zhu, C. Liu, Y. Ge, H. Xiong, and E. Chen, "Popularity modeling for mobile APPs: A sequential approach," *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1303–1314, Jul. 2015, Doi: 10.1109/TCYB.2014.2349954.

Chuanren Liu received the BS degree from the University of Science and Technology of China (USTC), the MS degree from the Beijing University of Aeronautics and Astronautics (BUAA), and the PhD degree from Rutgers, the State University of New Jersey. He is currently an assistant professor in the Decision Science and Management Information Systems Department at Drexel University. His research interests include data mining and knowledge discovery, and their applications in business analytics. He has published papers in refereed journals and conference proceedings, such as the *Annals of Operations Research*, the *European Journal of Operational Research*, the *IEEE Transactions on Knowledge and Data Engineering*, the *IEEE Transactions on Cybernetics*, *Knowledge and Information Systems*, and ACM SIGKDD, IEEE ICDM, SIAM SDM, etc.

Hui Xiong (SM'07) received the BE degree from the University of Science and Technology of China (USTC), China, the MS degree from National University of Singapore (NUS), Singapore, and the PhD degree from the University of Minnesota (UMN), Minnesota. He is currently a professor and vice chair of the Management Science and Information Systems Department, and the director of Rutgers Center for Information Assurance, Rutgers, the State University of New Jersey. His general area of research is data and knowledge engineering, with a focus on developing effective and efficient data analysis techniques for emerging data intensive applications. He has published prolifically in refereed journals and conference proceedings. He is a senior member of the ACM and IEEE.

Spiros Papadimitriou is currently an assistant professor in the Department of Management Science & Information Systems, Rutgers Business School. Previously, he was a research scientist at Google, and a research staff member with IBM Research. His main interests are large scale data analysis, time series, graphs, and clustering. He has published more than 40 papers on these topics and has three invited journal publications in best paper issues, several book chapters and he has filed multiple patents. He has also given a number of invited talks, keynotes, and tutorials. He was a Siebel scholarship recipient in 2005 and received the best paper award in SIAM SDM 2008.

Yong Ge received the BE degree from Xi'an Jiao Tong University, the MS degree from the University of Science and Technology of China, and the PhD degree from Rutgers University, in 2005, 2008, and 2013, respectively. He is an assistant professor with the University of Arizona. His research interests include data mining and business analytics. He received the ICDM-2011 Best Research Paper Award, Excellence in Academic Research (one per school) at Rutgers Business School in 2013, and the Dissertation fellowship with Rutgers University in 2012. He has published prolifically in refereed journals and conference proceedings, such as the *IEEE Transactions on Knowledge and Data Engineering*, the *ACM Transactions on Information Systems*, the *ACM Transactions on Knowledge Discovery from Data*, the *ACM Transactions on Intelligent Systems and Technology*, ACM SIGKDD, SIAM SDM, IEEE ICDM, and ACM RecSys.

Keli Xiao received the master's degree in quantitative finance from Rutgers Business School, in 2009, the master's degree in computer science from Queens College in the City University of New York, in 2008, and the PhD degree in finance from Rutgers Business School, in 2013. His general area of research in finance is asset pricing, with a concentration on the studies of financial bubbles and empirical asset pricing. He has published several papers in refereed journals and conference proceedings such as the *ACM Transactions on Knowledge Discovery from Data* (*TKDD*) and ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.