Mining Context-Aware Resource Profiles in the Presence of Multitasking*

Gerhardus A. W. M. van Hulzen[a,*], Chiao-Yun Li[c], Niels Martin[a,b], Sebastiaan J. van Zelst[c,d], Benoît Depaire[a]

[a] *Hasselt University, Research group Business Informatics, Martelarenlaan 42, 3500 Hasselt, Belgium*
[b] *Research Foundation Flanders (FWO), Egmontstraat 5, 1000 Brussels, Belgium*
[c] *Fraunhofer Institute for Applied Information Technology (FIT), Data Science and Artificial Intelligence Department, Schloss Birlinghoven, Sankt Augustin 53757, North Rhine-Westphalia, Germany*
[d] *RWTH Aachen University, Chair of Process and Data Science, Ahornstraße 55, Aachen 52074, North Rhine-Westphalia, Germany*

Abstract

Healthcare organisations are becoming increasingly aware of the need to improve their care processes and to manage their scarce resources efficiently to secure high-quality care standards. As these processes are knowledge-intensive and heavily depend on human resources, a comprehensive understanding of the complex relationship between processes and resources is indispensable for efficient resource management. Organisational mining, a subfield of Process Mining, reveals insights into how (human) resources organise their work based on analysing process execution data recorded in Health Information Systems (HIS). This can be used to, e.g., discover *resource profiles* which are groups of resources performing similar activity instances, providing an extensive overview of resource behaviour within healthcare organisations. Healthcare managers can employ these insights to allocate their resources efficiently, e.g., by improving the scheduling and staffing of nurses. Existing resource profiling algorithms are limited in their ability to apprehend the complex relationship between processes

*Corresponding author
Email address: gerard.vanhulzen@uhasselt.be (Gerhardus A. W. M. van Hulzen)

and resources because they do not take into account the *context* in which activities were executed, particularly in the context of multitasking. Therefore, this paper introduces *ResProMin–MT* to discover *context-aware resource profiles* in the presence of multitasking. In contrast to the state-of-the-art, ResProMin–MT is capable of taking into account more complex contextual activity dimensions, such as activity durations and the degree of multitasking by resources. We demonstrate the feasibility of our method within a real-life healthcare context, validated by medical domain experts.

Keywords: Process mining, Organizational mining, Resource profiles, Context-aware process mining, Multitasking, Healthcare processes

---

1. Introduction

Healthcare organisations, e.g. hospitals, are facing critical challenges, most notably increasing and ageing populations and, at the same time, tightening budgets [1, 2]. To cope with these challenges – while securing high-quality care standards – healthcare organisations are becoming increasingly aware of the need to improve their care processes and to manage their scarce resources efficiently [3, 4]. In order to determine suitable levels of resources (e.g. staff, equipment, and facilities [5]) and efficient resource allocation, healthcare managers need a comprehensive understanding of the complex relationship between processes and resources [6, 7].

To uncover the real behaviour of resources that perform activities in processes, the process execution data captured by *Health Information Systems (HIS)* and *Electronic Health Records (EHR)* can be used [2, 3, 8]. The events recorded by these systems can be compiled into an *event log*, which represents the real-life behaviour of a process [2]. *Process Mining* is a research domain focusing on the (semi-)automatic extraction of insights from event logs [9]. As most of these events are triggered by logging activities performed by human *resources* (e.g. nurses) on *cases* (e.g. patients) [3], we can exploit this information to gather extensive insights into the relationship between resources and the activities they perform. These insights provide a comprehensive and transparent overview on the behaviour of resources within a healthcare organisation and aid healthcare managers in more efficiently allocating their resources, e.g. improving the scheduling (i.e. rostering) and staffing (i.e. determining suitable levels) of nurses [10–12].

In Process Mining, the subfield of *organisational mining* is concerned with discovering organisational structures and social networks from event logs [13]. Several resource profiling techniques – i.e. finding groups of resources that perform similar activity instances – have been proposed [13–23]. Nevertheless, most existing algorithms only consider a brief description of the activities that resources performed (e.g. "Create purchase requisition", "Send invoice", "Administer medication", etc.) as a starting point. Thereby, the context is ignored, i.e. the circumstances in which the activity was executed [24]. In addition, compared to common business processes, such as the order-to-cash process, healthcare processes are generally more knowledge-intensive [3, 25, 26] and typically exhibit a higher degree of variability [26, 27], due to the involvement of knowledge workers, such as physicians and nurses, and complex *ad hoc* decision-making [3, 25]. For example, depending on the patient's health condition, a more experienced or specialised senior nurse may be assigned take care of the patient. However, the activities as such (e.g. the activity labels recorded in the HIS) are the same regardless of the patient's condition. Therefore, this limiting assumption can hide essential nuances on how resources conduct their tasks in a real-life healthcare setting, which highlights the importance of considering the context in which the activities were executed besides the activity labels. This context can be regarded as a multi-dimensional concept describing "*who* did *what* under *which circumstances*" [24]. Straightforward dimensions that can be considered include time-related attributes of the activity instance, e.g. the weekday, the time of the day (e.g. morning, afternoon, or evening), or the duration of the activity; case attributes, such as the case type or status; and a resource identifier [24]. However, additional dimensions can be added to capture more complex aspects which are not directly observable in the event log, such as the degree of multitasking by resources. It is not trivial to consider multitasking using existing algorithms because these cannot handle attributes of mixed types simultaneously, e.g. nominal (activity labels), discrete (number of concurrently performed activities), and continuous (duration of activities).

This paper extends our previous work on discovering context-aware resource profiles from event logs using *ResProMin* [24]. While ResProMin considers contextual variables (expressing the conditions under which activity instances were executed) when discovering resource profiles, the method cannot incorporate the multitasking behaviour of resources, which is particularly common in the healthcare sector [28]. Therefore, we introduce an extension to our previous work: *ResProMin–MT (Resource Profile Miner–Multitasking)*. In addition to capturing the multitasking behaviour of resources, we also demonstrate how the context can be further defined by considering the activity duration. Both dimensions have not been considered before in resource profile identification. Moreover, whereas ResProMin was demonstrated using a public event log of a municipal service, we evaluated ResProMin–MT on a real-life case study in a healthcare context, more specifically, nursing. This also enabled us to present and discuss our findings with domain experts in nursing science at the hospital in order to validate the benefits of ResProMin–MT for healthcare managers in decision-making.

The remainder of this paper is structured as follows. Section 2 provides an overview of the related work. Section 3 introduces ResProMin–MT. Next, an introduction to the case study is provided in Section 4, an overview of the results in Section 5, and a discussion and evaluation of our method in Section 6. The paper ends with a conclusion and directions for future work in Section 7.

2. Related Work

This work is related to Process Mining applied in healthcare on the one hand and the resource perspectives in Process Mining on the other hand. The following sections provide an overview of related work in these domains.

2.1. Process Mining in Healthcare

The increasing implementation of HIS within healthcare organisations allows keeping record of large amounts of process execution data of care processes [26, 29]. This data can be used to construct an *event log* capturing the real-life behaviour of the processes [2, 3]. An event log contains at least an ordered set of *events* (e.g. starting a clinical examination or completing the registration of a patient) for each *case* (e.g. a patient). Additional information about these events can also be logged, e.g. a timestamp indicating when the event occurred or the involved resources (e.g. nurses and physicians) [3, 30]. Such an event log constitutes the primary input for further analysis using various Process Mining techniques [3, 9].

Interest in research on Process Mining in healthcare has steadily grown in recent years [26, 30, 31], and a wide variety of use cases have been demonstrated. These applications can be categorised among process discovery, conformance analysis, process analysis, predictive process analysis and simulation, social network analysis, and many more [32]. For example, process discovery has been used to analyse the radiology workflow for emergency patients [27] or palliative patients [33], Emergency Departments (ED) [29, 34], or cardiology [35–37]. Conformance analysis is used to check the compliance of real-life care processes with clinical guidelines [33, 38–40]. Process analysis includes various aspects besides the process discovery, e.g. identifying clinical pathway variants [41, 42], measuring clinical processes' performance [43, 44], or comparing clinical processes based on patient outcomes [41]. Predictive process analysis and simulation aim to forecast how the process will evolve in the future and conduct *what-if* analyses to determine the impact of changes on the process. The insights gathered by these analyses provide valuable feedback for supporting decision-making [32], e.g., by predicting the waiting times in an ED [45], the patient's postoperative Length of Stay (LoS) [46], or improving the performance of an ED by determining the optimal physician scheduling [47]. Finally, social network analysis relates to the organisational perspectives of processes, e.g. interactions between healthcare professionals [29, 48], patterns of collaboration within multidisciplinary teams [49, 50], and inter-departmental collaborations [51–53].

For further reference on the application of Process Mining in healthcare, the reader is referred to existing review papers [26, 30–32, 54, 55].

2.2. Resource Perspectives in Process Mining

Various resource-related topics have been analysed in Process Mining literature concerning how resources organise and perform their work. These can be categorised among four groups [23]: (i) organisational model mining, which deals with the discovery of resource groups that have similar characteristics in terms of conducting their duties [13, 15, 22, 23]; (ii) social network mining, which visualises the interaction between resources [16, 20, 29, 48–53, 56, 57]; (iii) rule mining, to determine resource assignment rules [58–61]; and (iv) behavioural profile mining, which discovers behaviour on how resources organise their work [19, 62–66].

This paper is mainly positioned within the first category, i.e. *organisational model mining*. A comprehensive review on this topic is provided by Yang *et al.* [23]. When discovering groups of resources, a wide variety of dimensions could be used, such as the activities they perform, the cases on which they focus, the shift in which they work, and the colleagues they interact with [23]. However, most literature on organisational model mining solely considers the labels of individual activity instances as a determinant for defining resource groups. For instance, Song & van der Aalst [13], Jin *et al.* [17], Ni *et al.* [18], Ye *et al.* [21], and Yang *et al.* [22] group resources based on the number of times they performed the same activities, which relies on the *performer-by-activity matrix*. This concept was initially defined by van der Aalst *et al.* [20] and counts for each resource (i.e. the "performer") how often they executed each activity. This is the most frequently used technique within organisational model mining. However, in a real-life setting – especially in complex processes such as healthcare processes – only considering the resource and activity label (i.e. "*who* did *what*") can hide essential nuances on how resources (e.g. nurses or physicians) conduct their activities [24]. Therefore, including additional dimensions, such as time-related attributes, case attributes, activity durations, the degree of multitasking, and other relevant contextual factors, can uncover implicit task division patterns during resource profile mining.

Apart from the predominant focus on activity labels, most organisational model mining algorithms do not allow the same activity to belong to multiple resource groups simultaneously, i.e. a specific activity can only be executed by one group [13, 15, 17, 19, 21]. This assumption often does not hold in reality either, where the same activity might be executed by a different resource from another group depending on contextual factors, such as difficulty or urgency [24].

To the best of our knowledge, only the work by Appice [14] and Yang *et al.* [22, 23] allows group memberships to overlap, of which only Yang *et al.* [23] consider multiple dimensions during resource profile mining. Their algorithm, *OrdinoR*, defines *execution contexts* as a combination of case types (e.g. regular versus VIP), activity types (e.g. activity labels), and time types (e.g. weekday and morning versus afternoon). These execution contexts are clustered to discover organisational models. Although OrdinoR is capable of including multiple dimensions as context, as well as finding overlapping resource groups, the context can only be defined by nominal variables. Other aspects, such as the degree of multitasking or the duration of activities, could only be added by discretising the variables. The decisions on discretisation are not straightforward and can profoundly influence the discovered resource groups. Our method, ResProMin–MT, provides a solution for this challenge by proposing a probabilistic model-based clustering technique that allows variables of mixed types to be used within the same model. In addition, we extend our previous work [24] by considering additional contextual factors, such as activity durations and the degree of multitasking, which have not been considered before in resource profile identification.

3. Method

In this section, we present ResProMin–MT[1] as an extension of our previous work [24], which enables discovering context-aware resource profiles from event logs by taking into account additional contextual factors of activity executions, such as activity durations and multitasking. A general overview of ResProMin–MT is visualised in Figure 1. Our method consists of three steps. In the first step, we *enrich* the event log by adding relevant contextual variables. Subsequently, in the second step, we identify *multitask session archetypes* from the enriched log. Finally, we discover groups of context-aware resource profiles. In the following sections, we describe each step in detail using a toy example.

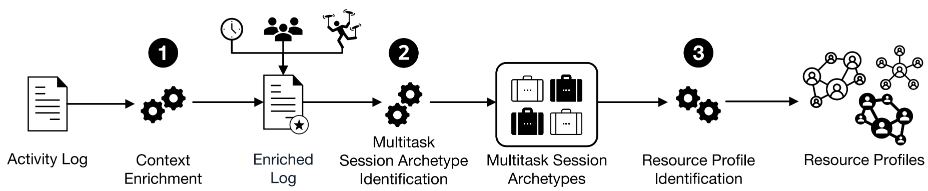


Figure 1: The three steps of ResProMin–MT: (1) context enrichment, (2) multitask session archetype identification, and (3) resource profile identification (adapted from [24]).

3.1. Step 1 – Context Enrichment

As an input for the first step, ResProMin–MT requires an *activity log*. In contrast to an event log, in which each entry represents an event occurring at a particular point in time, the entries of an activity log represent the execution of activity instances by a particular resource for a particular case (e.g. the registration of a patient by a receptionist). Therefore, each entry of an activity log can contain multiple timestamps, typically the start and completion times of the activity instance [2, 67]. An event log can be conveniently transformed into an activity log using, e.g., the R-package *bupaR* [68].

[1]The implementation of ResProMin–MT can be consulted here: `https://doi.org/10.5281/zenodo.7215943`

The input activity log should satisfy the following conditions, i.e. each entry of the log should have a(n):

- Case identifier: an attribute that uniquely identifies a process execution instance, i.e. a case [9]. For example, a patient ID when considering the process from a patient-centric viewpoint, or a staff ID when studying the process from the standpoint of the staff members;
- Timestamp: the time at which the event took place. In order to be able to take multitasking behaviour and activity durations into account, at least two events or timestamps (i.e. start and completion) should be recorded for the activities in order to calculate duration and determine activity overlap;
- Activity label: a brief label describing the activity that triggered the event;
- Resource identifier: an attribute that describes the resource that was involved during the event.

Table 1 shows a fragment of a log which satisfies these requirements. Each row describes the execution of an activity performed by a nurse (the resource) on a particular patient (the case). In addition, information on the patient type and the experience (XP) of the nurses are recorded, which can be used to define the context.

Case ID	Activity	Start	Complete	Resource	Patient Type	Resource XP	...
512	Communication w/ patient	2022/05/12 09:15	2022/05/12 09:23	Anna	Ambulatory	Student nurse	...
512	Prepare medication	2022/05/12 09:18	2022/05/12 09:20	Anna	Ambulatory	Student nurse	...
512	Administer medication	2022/05/12 09:20	2022/05/12 09:22	Anna	Ambulatory	Student nurse	...
843	Read document	2022/05/12 10:34	2022/05/12 10:45	Thomas	Ambulatory	Jr. nurse	...
843	Communication w/ patient	2022/05/12 10:48	2022/05/12 10:55	Thomas	Ambulatory	Jr. nurse	...
843	Observations & monitoring	2022/05/12 10:50	2022/05/12 10:55	Thomas	Ambulatory	Jr. nurse	...
843	Reporting	2022/05/12 10:56	2022/05/12 11:03	Thomas	Ambulatory	Jr. nurse	...
635	Communication w/ patient	2022/05/13 15:58	2022/05/13 16:06	John	Emergency	Sr. nurse	...
635	Take blood sample	2022/05/13 16:00	2022/05/13 16:04	John	Emergency	Sr. nurse	...
...	...	...	...	...	...	...	...

Table 1: A fragment of an example log.

The log can be *enriched* by adding additional attributes that describe the *what*, *when*, and *under which circumstances* of the activity execution, e.g. the case type, weekday, morning or evening shift, activity duration, and many more[2]. In addition, the workload in terms of multitasking can be determined by aggregating activities with (partially) overlapping time intervals into a single *multitask session*, representing a group of concurrently performed activities. In fact, during the enrichment step, there is flexibility to shape attributes that are relevant to a particular application context.

An example of enrichment of our sample log in Table 1 is presented in Table 2. Each row represents a multitask session. For example, the three activities Anna executed concurrently are merged into a single multitask session that took in total 8 minutes. The "degree of multitasking" or *multitasking level* is, therefore, 3. However, not every multitask session needs to consist of multiple activity executions, as shown by the Read document and Reporting activities by Thomas. These activities were not executed concurrently with other activities. In addition, the contextual factors patient type and resource experience can be retained, and additional context can be defined, such as the shift (Morning, Afternoon, Night, etc.) during which the session was performed. The case ID is removed, as this identifier is linked to one specific patient only.

Activities	Resource	Patient Type	Resource XP	Shift	Duration	MT Level	...
Communication w/ patient, Prepare medication, Administer medication	Anna	Ambulatory	Student nurse	Morning	8 mins	3	...
Read document	Thomas	Ambulatory	Jr. nurse	Morning	11 mins	1	...
Communication w/ patient, Observations & monitoring	Thomas	Ambulatory	Jr. nurse	Morning	7 mins	2	...
Reporting	Thomas	Ambulatory	Jr. nurse	Morning	7 mins	1	...
Communication w/ patient, Take blood sample	John	Emergency	Sr. nurse	Afternoon	8 mins	2	...
...	...	...	...	...	...	...	...

Table 2: An example of an enriched log. Activity instances that overlap in time are aggregated into a *multitask session*. The *multitasking (MT) level* is defined by the number of concurrently executed activity instances.

[2]The attributes used in the real-life case study will be outlined in Section 4.2.1.

3.2. Step 2 – Multitask Session Archetype Identification

In the second step, the enriched log is clustered to find *multitask session archetypes*, which describe multitask sessions that were executed under similar conditions. Each archetype represents a set of multitask sessions that exhibit high homogeneity with sessions within that archetype and high heterogeneity with sessions from other archetypes [24]. In our example, we could observe that medication administration is often executed during the morning shift by a student nurse, whereas taking blood samples from emergency patients is more likely to require a senior nurse. These two examples are then defined as two distinct multitask session archetypes.

The remainder of this section specifies how the multitask session archetypes are identified. First, we introduce the clustering technique and explain its benefits. Next, we describe how the clustering model can be specified. Subsequently, we provide a method to determine the number of multitask session archetypes statistically. Finally, the archetypes can be profiled based on the fitted clustering model.

3.2.1. Clustering Technique

To cluster the enriched log, we propose to use the probabilistic model-based clustering technique *Finite Mixture Models (FMM)* [69]. FMMs have been successfully applied in a wide variety of domains, including agriculture, bioinformatics, biology, economics, engineering, marketing, medicine, physics, psychology, and many other fields, due to their versatile applicability in major areas of statistics [70, 71]. These applications include latent class and cluster analysis, survival analysis, image analysis, data analysis and inference, and many more [71].

Even though other clustering techniques can be used for this step as well, e.g. k-means or hierarchical clustering, the use of a probabilistic model-based clustering technique such as FMM has several advantages. Firstly, it is a fuzzy clustering technique which allows the archetypes to overlap, which is required for the probabilistic assignment of multitask sessions to resources. Secondly, statistical criteria can be used to determine the appropriate number of clusters, such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) [70–73]. Thirdly, variables of mixed types, such as discrete, continuous, and nominal, can be used in the same model [70]. This allows considering, e.g., the multitasking level and duration as contextual factors in discovering resource profiles. Although other clustering techniques, such as k-means, can also deal with variables of mixed types, an appropriate distance measure is required. The definition of such a measure is not evident, and the resulting clusters are extensively dependent on the distance matrix. In addition, scaling problems cause further challenges. FMM overcomes these challenges as no scaling nor distance measure is required [70, 74, 75]. Finally, as FMM uses a probabilistic model, each cluster is described by a set of statistical distributions. This provides a detailed description for every cluster, allowing a more comprehensive and transparent interpretation. For example, resources and execution contexts can be probabilistically assigned to multiple clusters (or archetypes). This is especially relevant in complex settings such as healthcare, where much flexibility is expected from healthcare professionals [3, 25–27].

3.2.2. Clustering Model Specification

The FMM can be specified as follows. Suppose we have a set $\mathbf{S} = \langle \mathbf{s}_1, \ldots, \mathbf{s}_N \rangle$ of N multitask session observations, where each observation $\mathbf{s}_n$ corresponds to a row in the enriched log, which is represented as a vector of J descriptive attributes s_{nj} corresponding to the columns (e.g. the resource, the activity label, the duration of the activity, etc.), with $n \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, J\}$. For instance, in our running example, $\mathbf{s}_1$ corresponds to the first row in the enriched log (Table 2):

$$\mathbf{s}_1 = \langle (\textsf{Communication w/ patient}, \textsf{Prepare medication}, \textsf{Administer medication}), \\ \textsf{Anna}, \textsf{Ambulatory}, \textsf{Student nurse}, \textsf{Morning}, 8\ mins, 3 \rangle,$$

and $s_{1,2} = \textsf{Anna}$, i.e. the second attribute of the first row in Table 2.

The mixture model, from which $\mathbf{S}$ is a sample, is denoted as h and is assumed to consist of a mixture of K component distributions. Each component distribution f_k can be interpreted as a separate multitask session archetype. This multivariate component distribution is typically modelled as the product of univariate distributions f_{kj} of the individual attributes s_{nj}, assuming component-conditional independence among the attributes. The weight of the component distribution f_k in the mixture distribution h is denoted as π_k, with $k \in \{1, \ldots, K\}$. The mixture distribution h can then be written as given by Equation 1 [70–73]:

$$\begin{aligned} h(\mathbf{s}_n; \boldsymbol{\vartheta}) &= \sum_{k=1}^{K} \pi_k f_k(\mathbf{s}_n; \boldsymbol{\theta}_k) \\ &= \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} f_{kj}(s_{nj}; \boldsymbol{\theta}_{kj}), \end{aligned} \tag{1}$$

where the k^{th} component distribution of attribute j for multitask session observation $\mathbf{s}_n$ is $f_{kj}(s_{nj}; \boldsymbol{\theta}_{kj})$ with component-specific parameter vector $\boldsymbol{\theta}_{kj}$ and $j \in \{1, \ldots, J\}$. π_k is the prior probability, or mixture proportion, of component k which must satisfy $\sum_{k=1}^{K} \pi_k = 1$, where $\forall k : \pi_k > 0$. The vector of all model parameters is denoted as $\boldsymbol{\vartheta}$ and can be estimated by maximising the log-likelihood using, e.g., the *Expectation-Maximisation (EM)* algorithm [76]. More details on the specification and estimation of FMMs using EM can be found in Vermunt and Magidson [70], McLachlan *et al.* [71], Frühwirth-Schnatter [72], and McLachlan and Peel [73].

Considering Equation 1, for each attribute j, an appropriate parametric distribution f_{kj} needs to be selected. A few examples of relevant distributions depending on the attribute type are provided in Table 3. This is only a limited subset of possible distributions. Other parametric distributions may be more appropriate to describe certain attributes. If required, the same distribution could be used multiple times for different attributes of the same type.

In our running example, the resource, resource experience, patient type, and shift could be modelled as multinomial distributions, whereas the multitasking

Distribution	Attribute type
Binomial	Binary or dummy variables
Exponential	Inter-arrival times in a homogeneous Poisson process
Gaussian	Biological phenomena (patient's height, blood pressure, etc.), engineering tolerances, etc.
Multinomial	Nominal/Categorical or ordinal variables, such as resources, weekdays, activity labels, case status, etc.
Poisson	Count variables, such as the number of activities performed simultaneously during multitasking, or the amount of currently active cases
Weibull	Activity or session durations

Table 3: Examples of appropriate distributions for various attribute categories.

level of the multitask session can be described by a Poisson distribution and the duration by a Weibull distribution. If the order in which the activities were executed within a multitask session is not of interest, the activities could be encoded, e.g., using one-hot encoding, and described with binomial distributions. This is particularly useful when a large number of distinct activity labels are present because the number of unique combinations of concurrently executed activities could become very large. If we apply one-hot encoding to the activity attribute in our running example (i.e. the first column of Table 2), we get the following vector $\mathbf{s}_1 \rightarrow \mathbf{s}'_1$ for the first row in Table 2:

$$\mathbf{s}'_1 = \langle 1, 1, 0, 1, 0, 0, 0, \mathsf{Anna}, \mathsf{Ambulatory}, \mathsf{Student\ nurse}, \mathsf{Morning}, 8\ mins, 3 \rangle,$$

where the first seven elements represent all activities present in the enriched log, i.e. Administer medication, Communication w/ patient, Observations & monitoring, Prepare medication, Read document, Reporting, and Take blood sample, in this order. A binary value indicates whether the activity was executed (1) or not (0) during the multitask session. The remaining elements represent the resource, patient type, resource experience, shift, duration, and multitasking level attributes for multitask session $\mathbf{s}'_1$.

Hence, in our example, the mixture distribution h (Equation 1) can be modelled as follows:

$$\begin{aligned} h(\mathbf{s}'_n;\boldsymbol{\vartheta}) &= \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} f_{kj}(s'_{nj};\boldsymbol{\theta}_{kj}) \\ &= \sum_{k=1}^{K} \pi_k \prod_{j=1}^{7} B_k(1,\theta_{kj}) \prod_{j=8}^{11} Mult_k(1,\boldsymbol{\theta}_{kj})\, Weib_k(s'_{n12},\boldsymbol{\theta}_{k12}) Pois_k(s'_{n13},\boldsymbol{\theta}_{k13}), \end{aligned}$$

where, for each multitask session archetype, the component distribution f_k can be modelled as the product of thirteen ($J = 13$) univariate distributions:

- $B_k(1, \theta_{kj})$ is a binomial distribution with one trial (i.e. a Bernoulli distribution) describing the probability θ_{kj} of observing activity j in component k, e.g. $j = 1$ corresponds to the activity Administer medication;
- $Mult_k(1, \boldsymbol{\theta}_{kj})$ is a multinomial distribution with one trial (i.e. a categorical distribution) with probability vector $\boldsymbol{\theta}_{kj}$ describing the probabilities of observing each distinct value of categorical attribute j in component k, e.g. for $j = 8$, $\boldsymbol{\theta}_{kj}$ contains the probabilities whether each resource was involved during multitask session archetype k;
- $Weib_k(s'_{n12}, \boldsymbol{\theta}_{k12})$ is a Weibull distribution with parameters $\boldsymbol{\theta}_{k12}$ for the duration attribute s'_{n12} ($j = 12$) of multitask session $\mathbf{s}'_n$ for component k;
- $Pois_k(s'_{n13}, \boldsymbol{\theta}_{k13})$ is a Poisson distribution with parameters $\boldsymbol{\theta}_{k13}$ for the multitasking level attribute s'_{n13} ($j = 13$) of multitask session $\mathbf{s}'_n$ for component k.

3.2.3. Determining the Number of Multitask Session Archetypes

Comparable to k-means, the number of components (or clusters) K needs to be specified upfront. Several statistical criteria can be used to select the "best" number of components, including the frequently used *Bayesian Information Criterion (BIC)*, which tries to find a balance between model complexity (i.e. number of parameters) and goodness-of-fit (i.e. log-likelihood) [72, 73]:

$$BIC = -2 \log \mathcal{L}(\hat{\boldsymbol{\vartheta}}) + d \log N, \tag{2}$$

where d is the total number of parameters to be estimated in the model, N is the number of multitask session observations, and $\hat{\boldsymbol{\vartheta}}$ is the vector of estimated parameter values of the model parameter vector $\boldsymbol{\vartheta}$ that maximises the mixture likelihood function $\mathcal{L}(\hat{\boldsymbol{\vartheta}})$ [71–73]:

$$\begin{aligned}\mathcal{L}(\hat{\boldsymbol{\vartheta}}) &= \prod_{n=1}^{N} h(\mathbf{s}_n; \hat{\boldsymbol{\vartheta}}) \\ &= \prod_{n=1}^{N} \left(\sum_{k=1}^{K} \hat{\pi}_k \prod_{j=1}^{J} f_{kj}(s_{nj}; \hat{\boldsymbol{\theta}}_{kj}) \right)\end{aligned} \tag{3}$$

The general intent is to minimise Equation 2 when selecting an appropriate number of components (K) [72, 73]. However, as a more parsimonious model is generally preferred over a more complex model, the "best" model is not always the model having the lowest BIC, especially if the BIC barely decreases when adding an additional component to the model. In this case, it may be better to use the "elbow criterion" to select the appropriate number of components, i.e. a decrease in marginal gain of $\mathcal{L}(\hat{\boldsymbol{\vartheta}})$ by adding an extra component can be identified by the angle, or "elbow pattern", in the BIC-plot [75, 77].

The "quality" of the selected model can be assessed using the relative entropy E_K. This index can be used as an indication of the overall precision of the cluster model [71, 77]:

$$E_K = 1 - \frac{\sum_{n=1}^{N} \sum_{k=1}^{K} -\hat{\tau}_{nk} \log \hat{\tau}_{nk}}{N \log K}, \tag{4}$$

where $\hat{\tau}_{nk}$ is the estimated posterior probability for multitask session observation $\mathbf{s}_n$ corresponding to the k^{th} component, i.e. $P(\mathit{Component} = k \mid \mathbf{s}_n; \hat{\boldsymbol{\vartheta}})$ [71, 77]. For example, $P(\mathit{Component} = 2 \mid \mathbf{s}'_1; \hat{\boldsymbol{\vartheta}})$ is the posterior probability that multitask session observation $\mathbf{s}'_1$ (i.e. the first multitask session of Table 2) corresponds to the second multitask session archetype ($k = 2$).

The relative entropy E_K quantifies the posterior classification uncertainty for a FMM with K components and is bound to the unit interval: $0 \leq E_K \leq 1$. A relative entropy of $E_K = 0$ indicates that the posterior classification is no better than randomly assigning observations to clusters, whereas $E_K = 1$ indicates a perfect classification in which the mixture's components are well separated. However, it should be noted that while E_K can be used as an indication of the overall precision of the model, it was not intended for model selection and, therefore, should not be applied for that purpose (i.e. the *BIC* is preferred for that objective) [71, 72, 77, 78].

The EM algorithm iteratively updates the estimated parameters $\hat{\boldsymbol{\vartheta}}$ until either (i) the predefined maximum number of iterations has been reached, or (ii) the convergence criterion for the mixture likelihood function $\mathcal{L}(\hat{\boldsymbol{\vartheta}})$ (Equation 3) has been met [77, 78]. However, this convergence could occur at a local maximum. Therefore, it is imperative to repeat the EM algorithm several times with different starting values of $\hat{\boldsymbol{\vartheta}}^{(0)}$ to mitigate the risk of finding a local optimum [72, 77, 78].

3.2.4. Profiling the Multitask Session Archetypes

Once a satisfactory output model is chosen, each component k of the fitted FMM corresponds to a multitask session archetype and can be described by the fitted component-specific parameter vectors $\hat{\boldsymbol{\theta}}_{kj}$ of the component distributions f_{kj}. Therefore, the number of multitask session archetypes is determined by the number of components of the selected FMM, i.e. K. Each archetype can be profiled using a brief description to facilitate discussions with domain experts by making the archetypes more recognisable. For instance, one of the multitask session archetypes in our running example could describe that, generally speaking, the activity Administer medication has a high probability of being executed during the Morning shift, primarily by a Student nurse, and is often performed concurrently while communicating with the patient.

3.3. Step 3 – Resources Profile Identification

In the final step, we cluster the resources into *context-aware resource profiles*, i.e. groups of resources that perform similar activities under similar circumstances. Each resource profile describes a group of resources that perform similar multitask session archetypes. Therefore, these profiles describe resources beyond their hierarchical role within an organisation or resources groups solely defined by taking activity labels into account. In our running example, we may find a multitask session archetype containing a group of senior nurses that often takes blood samples from emergency patients. In this case, John would be a member of that context-aware resource profile.

In order to determine the resource profiles, we first determine for every resource the posterior probabilities that they belong to each multitask session archetypes defined in Step 2. Secondly, we cluster these probabilities to find groups of resources working on the same multitask session archetypes. Finally, these probabilities can also be used to discern "generalists" (i.e. resources who work on several distinct multitask session archetypes) from "specialists" (i.e. resources that mainly focus on one particular multitask session archetype) [6, 24].

3.3.1. Posterior Probabilities of Resources

Each multitask session archetype (or cluster) k has a component distribution f_{kj} for the j^{th} attribute, where j corresponds to the resource attribute s_{nj} of multitask session observation $\mathbf{s}_n$, e.g. in our running example, $j = 8$ is the resource attribute. The posterior probability of observing resource r in multitask session archetype k is given by Equation 5:

$$f_{kj}(s_{nj}; \boldsymbol{\theta}_{kj}) = P(s_{nj} = r \mid \mathit{Component} = k; \boldsymbol{\theta}_{kj}) \sim \mathit{Mult}_k(1, \boldsymbol{\theta}_{kj}), \tag{5}$$

where the posterior probability $P(s_{nj} = r \mid \mathit{Component} = k; \boldsymbol{\theta}_{kj})$ follows a multinomial distribution with probabilities $\boldsymbol{\theta}_{kj}$ and $k \in \{1, \dots, K\}$.

To calculate the posterior probability that resource r belongs to multitask session archetype k, Bayes' Theorem can be applied:

$$P(\mathit{Component} = k \mid s_{nj} = r; \boldsymbol{\theta}_{kj}) \propto \\ P(s_{nj} = r \mid \mathit{Component} = k; \boldsymbol{\theta}_{kj}) P(\mathit{Component} = k; \boldsymbol{\vartheta}), \tag{6}$$

where $P(\mathit{Component} = k; \boldsymbol{\vartheta})$ is the prior probability (i.e. π_k), and $P(s_{nj} = r \mid \mathit{Component} = k; \boldsymbol{\theta}_{kj})$ is the posterior probability of observing resource r in multitask session archetype k defined by Equation 5.

3.3.2. Discovering Context-Aware Resource Profiles & Specialisation Profiles

For each resource, the posterior probabilities calculated using Equation 6 are encoded as a vector of the probabilities belonging to each archetype. Next, a distance matrix is calculated using the Euclidean distance among the resources based on this encoding, resulting in a close distance among resources with similar multitask session archetype probabilities. Finally, the distance matrix can be clustered using, e.g. *Agglomerative Hierarchical Clustering (AHC)*. The number of clusters – i.e. distinct context-aware resource profiles – can be determined by applying the "elbow criterion" on the *Total Within-Cluster Sums of Squares (WCSS)* plot [75, 79].

The matrix on the left in Figure 2 shows an example of a probability matrix used to find context-aware resource profiles. The rows represent the involved resources, and the columns contain the posterior probability each resource belongs to each of the K multitask session archetypes defined in Step 2. This matrix is clustered to find context-aware resource profiles. For example, we observe that Anna and Samantha frequently work on multitask session archetypes MSA2 and MSA3. Therefore, they belong to the same resource profile. Similarly, George and Sophia frequently perform archetypes MSA3 and MSA4, and form another resource profile together.

Additionally, the degree of "specialisation" – i.e. the number of distinct multitask session archetypes a resource performs – can be inferred as well. Figure 2 visualises the required transformation on the probability matrix. First, the encoded vector of posterior probabilities from the resource profiles is ordered from largest to smallest, i.e. the first element contains the highest posterior probability of resource r belonging to any particular multitask session archetype, the second column contains the second-highest probability, etc. Next, the same clustering technique as with the resource profiles is used to find groups of resources with a similar degree of specialisation. Because this encoded vector sums to one, resources with a large maximum probability (e.g. $> 90\%$) are more likely to be placed into the same group than resources with a lower maximum probability (e.g. < 30%). The former group of resources focuses mainly on one multitask session archetype and is, therefore, referred to as "specialists", whereas the latter group divide their time among multiple archetypes and, hence, is referred to as "generalists". For example, in the matrix on the right in Figure 2, we observe that Thomas and John are specialists, as they mainly focus on one multitask session archetype, albeit not on the same archetype, i.e. MSA1 and MSA5, respectively. On the other hand, Anna and George divide their time among three archetypes and are, therefore, more generalists.

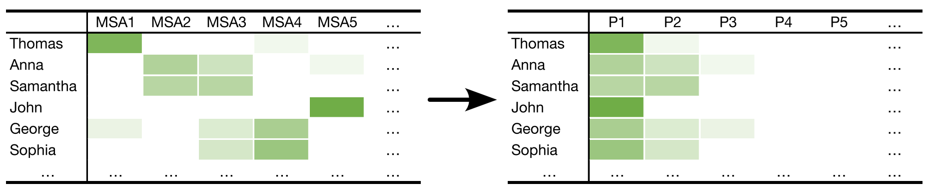


Figure 2: On the left side, an example of the probability matrix with for each resource (rows) the posterior probabilities they belong to each multitask session archetype (columns), calculated using Equation 6. This matrix is clustered using AHC to find context-aware resource profiles. On the right side, the probabilities for each resource are ordered from largest to smallest to discern specialists from generalists.

4. Case Study

In order to demonstrate ResProMin–MT, a real-life case study is conducted at a large University Medical Centre in the Netherlands. This section provides an overview of the dataset that has been used, together with the context in which the data has been recorded and the preprocessing that has been applied. Next, we discuss how the method has been operationalised to derive context-aware resource profiles.

4.1. Data Preprocessing

The data has been collected by students in nursing science shadowing every activity that a nurse performed during their shift in the period February–April

2018. Table 4 shows an excerpt of the dataset. Each row represents the execution of an activity performed by a nurse. For example, the first row represents an execution of Hand hygiene under the category of Indirect care performed in shift 59 from 08:03:00 until 08:03:30 on February 14, 2018. A nurse might also perform several activities concurrently. As shown in the second and third row of Table 4, the nurse of shift 59 started performing Daily life activities 30 seconds before Materials (searching) has been completed, which indicates multitasking. Note that only the shift identifier (instead of a nurse identifier) is provided due to data privacy concerns.

Shift ID	Activity	Start Date	Start	Complete	Category	...
59	Hand hygiene	2018/02/14	08:03:00	08:03:30	Indirect care	...
59	Materials (searching)	2018/02/14	08:04:00	08:05:30	Indirect care	...
59	Daily life activities	2018/02/14	08:05:00	08:25:30	Direct patient care	...
59	Materials (searching)	2018/02/14	08:24:59	08:26:29	Indirect care	...
59	Planning meeting	2018/02/14	09:33:00	09:33:30	Professional communication	...
59	(Answer) Telephone	2018/02/14	09:52:00	09:56:30	Professional communication	...
86	Assessment	2018/02/27	13:57:00	14:03:30	Direct patient care	...
86	Meeting	2018/02/27	15:32:00	15:46:30	Ward-related tasks	...
...	...	...	...	...	...	...

Table 4: An excerpt of the dataset.

Shifts that contain a deficient number of activities (i.e. only three to five) have been excluded as they contain too few activities to reliably assign them to a multitask session archetype. Moreover, as some activities were rarely performed (which could result in clustering this activity in an illogical archetype) or took a significant amount of time during a shift (which causes the algorithm to consider the entire shift as a single multitask session), these activities may decrease the performance of the algorithm. Together with the domain experts, we determined that the executions of the following activities should be removed from the dataset: Request/medication order, Providing a clinical lesson, Recording conversation, Present at clinical class, and Supervision/supervision of student.

After filtering, the dataset contains 7,022 executions of 37 activities under 11 categories, performed during 68 shifts from February 11 to April 26, 2018. Besides the attributes shown in Table 4, we included the *educational level* of the nurses executing the activity, as well as their *nursing experience/organisational role*, the *hospital shift* and *weekday* on which the activities were performed, and the *ward* in which the shift took place. Table 5 provides an overview of all attributes with their possible values.

4.2. Context-Aware Resource Profile Identification

This section describes how we identified the context-aware resource profiles from the dataset. First, we enriched the dataset by defining contextual attributes. Next, we identified the multitask session archetypes by clustering the enriched log. Finally, we discuss how the resource profiles were established.

4.2.1. Step 1 – Context Enrichment

According to the domain experts, nurses can be differentiated by their tendency to multitask during their shifts. As such, we aggregated the records of concurrently executed activities to form *multitask sessions*. Figure 3 presents an example in which we observe two activities being performed concurrently, i.e. the execution of two activities (at least partially) overlap in time. We group the two concurrent activity executions together and create a new record starting from 10:15 until 10:30. This new record replaces the records of the two concurrent activity executions. The other record, which represents the execution of ADL activities, is kept, as it is not performed concurrently with other activities in the example. In the remainder of the paper, we refer to the aggregated activities as a *multitask session*. In the example of Figure 3, we extract two multitask sessions. In total, we identified 2,226 multitask sessions in the entire dataset.

The level of multitasking is defined by the number of concurrently executed activities performed in a multitask session. For example, as shown in Figure 3, the level of multitasking of the multitask session starting from 10:15 until 10:30 equals 2. The minimal level of multitasking is 1 if an activity was executed alone, e.g. ADL activities in the same example. In addition, the duration of the multitask session can be calculated by the difference in time between the start

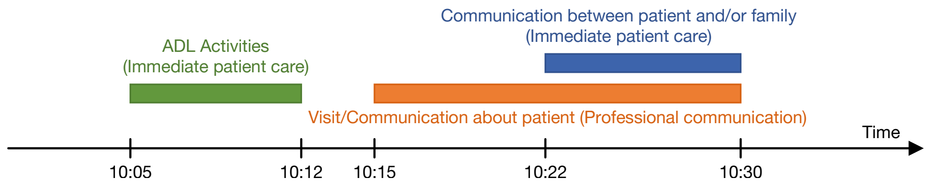


Figure 3: An excerpt of session 65, which consists of the execution of activities ADL activities, Visit/communication about patient, and Communication between patient and/or family. Multitasking can be observed in this excerpt as Visit/communication about patient, and Communication between patient and/or family are performed concurrently.

timestamp of the first activity and the end timestamp of the last activity for each session.

Meanwhile, the attention and effort required to perform an activity not only depend on the background of a nurse, but also on the type of the activity. Therefore, we encoded the category of the activities performed simultaneously in a multitask session. We applied *one-hot encoding* to represent the presence of the category of activities performed in a multitask session. Applying this coding to the detailed description of the activities would have given a more precise context definition. However, this would have made the model too complex, as there were 37 different activity labels in the dataset, each requiring a separate binomial distribution describing whether or not the activity was performed during the multitask session. Therefore, we used the activity category instead, which reduced the required number of binomial distributions to eleven. Because multiple activities can belong to the same activity category, we defined a second multitasking level, in addition to the previously defined level, to count the number of different activity categories within a multitask session (i.e. Multitasking level categories).

To summarise, a multitask session is described by *who performed it*, *on which weekday*, *in which ward*, *during which hospital shift*, *how many activities were performed concurrently* for *how long*, *the categories of the activities performed concurrently*, and the background of the nurses, i.e., *educational level* and *nursing experience/organisational role*. An overview of all attributes with possible values of the enriched log is shown in Table 5.

Attribute	Description
Shift ID	Unique identifier for each nursing shift
Hospital shift	Day (7 a.m.–3 p.m.), Evening (3 p.m.–10 p.m.), or Night (10 p.m.–7 a.m.)
Ward	The ward in which the shift took place: Gastroenterology, Geriatrics, HPB surgery, Intensive care, Internal med-system diseases, Neurology, Oncology-Hematology, Pulmonary medicine, or Transplantations. Each shift is performed in only one ward
Education level	The training according to the Dutch system that the nurse has followed: HBO-V (higher professional education), In-service trained (internal training), MBVO-V (secondary professional education), or Student HBO-V (student nurse in HBO-V system)
Function	The nursing experience or organisational role: nurse, senior nurse, nurse student, ICU nurse, or high care nurse
Weekday	The weekday of the shift (Mon–Sun)
Multitasking level	The total number of activities that were performed during one multitask session and the number of different activity categories
Activity category	Eleven binary attributes, one for each activity category: Department-related activities, Direct patient care, Documentation & reporting, Immediate patient care, Indirect patient care, Medication tasks, Other, Patient transportation, Professional communication, Social/personal time, and Time between in transit (moving from one location to another)
Duration	The duration of the multitask session in minutes

Table 5: Overview of the final attributes with a description of the enriched log.

4.2.2. Step 2 – Multitask Session Archetype Identification

To find the multitask session archetypes, we modelled a Finite Mixture Model with twenty distributions (f_{kj}) based on the attributes of the enriched log (Table 5): six multinomial distributions (Shift ID, Hospital shift, Ward, Education level, Function, and Weekday), two Poisson distributions (Multitasking level and Multitasking level categories), eleven binomial distributions (one for each Task category), and one Weibull distribution for the Duration of the multitask session.

To determine the appropriate number of clusters (i.e. multitask session archetypes or K), we fitted four to ten components to the model, as more than ten components would decrease the interpretability due to the large number of parameters. For each number of components, we repeated the fitting process 25 times to mitigate the risk that the EM algorithm would converge to a local optimum, resulting in a total of 175 models. We fitted the FMMs using the R-package *flexmix* (version 2.3_17) [80] (R version 4.1.2 [81]). A median of 38 iterations of the EM algorithm was required before converging to a solution, which took on average 3.44 minutes on an Intel®Xeon®Gold 6240 CPU @ 2.60GHz with 72GB of memory, running CentOS 7.9.2009. The summary statistics of the runtime are provided in Table 6. Since each model can be fitted independently, the fitting process was parallelised, which resulted in a total runtime of 49.12 minutes.

Min	Max	Mean	Median	SD	IQR	Total
0.9866	9.6422	3.4397	3.1278	1.6396	1.8856	49.1167

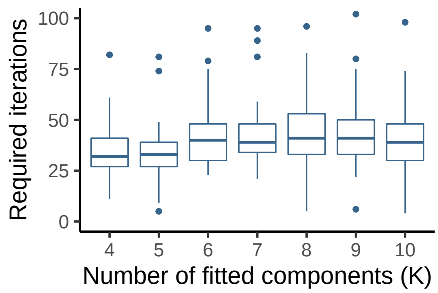
Table 6: Summary statistics on runtime (in mins) to fit the Finite Mixture Models.

As the number of components (K) increases, so does the number of parameters that need to be fitted. Figure 4 shows the required iterations and corresponding runtime per number of fitted components (K) before the EM algorithm converged to a solution. Even though the required number of iterations (Figure 4a) remains relatively constant with respect to K, the runtime (Figure 4b) increases as the number of parameters, and accordingly, the model's complexity increases.

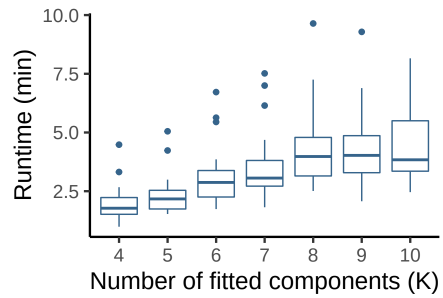
To select the most appropriate model, we applied three rules [24]:

(i) For each group of models with the same number of components, we selected the repetition with the highest log-likelihood (Equation 3);

(ii) To determine the appropriate number of components, we selected the



(a) Number of iterations per number of fitted components (K) of the EM algorithm before converging to a solution.



(b) Corresponding runtime (in min) per number of fitted components (K).

Figure 4: Number of iterations (a) and runtime (b) required before converging to a solution.

model with the lowest BIC (Equation 2), or the model that corresponded to the "elbow pattern" in the BIC-curve;

(iii) Each estimated component proportion (π_k) should be at least 5%, i.e. each component should contain at least 5% of the total observed multitask sessions. A component proportion near zero could be an indication of component "collapsing" or "over-extraction", in which the model is attempting to cluster the observations among more components than the data supports, or two or more components are insufficiently separated from each other. In this case, the relative entropy E_K (Equation 4) will be near zero, which can be used as an indication of problematic over-extraction [77].

We determined that nine components would be appropriate by applying these three rules to the model selection of the enriched log. As shown in Figure 5, the BIC is minimised while respecting the third rule. The $E_K = 0.9783$, which indicates that the components are well separated.

4.3. Step 3 – Resources Profile Identification

In the final step, we cluster the posterior probabilities that resource r belongs to multitask session archetype k. The full probability matrices after applying Equation 6 and reordering the probabilities from largest to smallest are

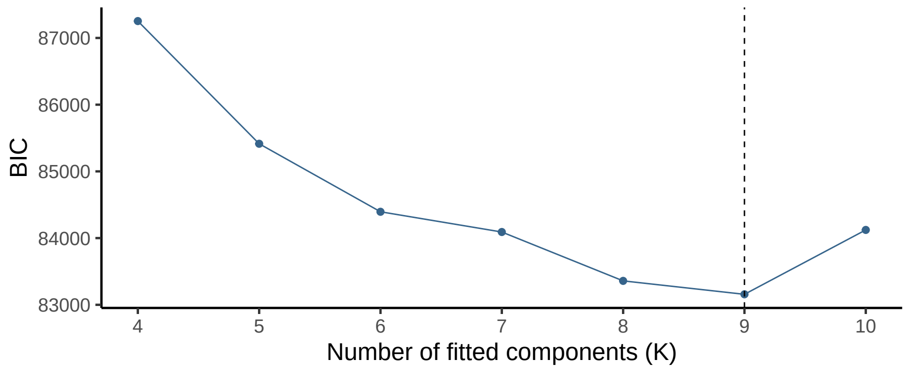


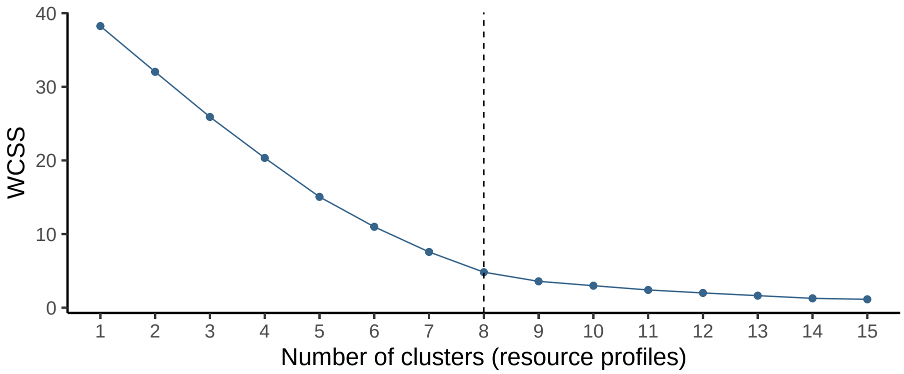
Figure 5: Evolution of the *BIC* when adding more components (K) to the model.

included in Appendix A. Both tables are clustered using Agglomerative Hierarchical Clustering (AHC) to discover the context-aware resource profiles of nurses with similar duties (Table A.11), as well as the specialisation profiles (Table A.12). Figure 6 shows the evolution of the Total Within-Cluster Sums of Squares (*WCSS*) for both types of profiles. A clear "elbow pattern" can be distinguished in Figure 6b at three specialisation clusters, whereas the resource profile clustering in Figure 6a lacks this apparent pattern. However, the decrease in *WCSS* diminishes after eight resource profile clusters, and therefore, eight clusters seem appropriate.

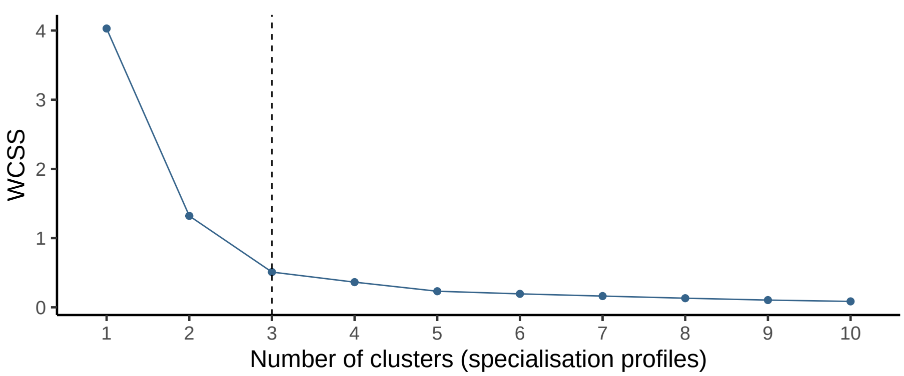
5. Results

The results of the estimated vector of model parameters $\hat{\boldsymbol{\vartheta}}$ of the FMM in Step 2 are shown in Tables 7–9. In Table 7a, the intra-cluster multinomial distributions for Ward are displayed. For instance, 89.15% of the multitask sessions that belong to cluster 1 were performed in the Geriatrics ward, whereas in cluster 4, the sessions took place in Transplantations (42.40%), Neurology (22.77%), Internal med-system diseases (17.90%), and Geriatrics (11.55%) in decreasing probability. The remaining wards are very weakly linked to cluster 4.

Looking at the Shift in Table 7b, we notice that the sessions in clusters 1 and 9 are always performed during Day (99.60% and 100.00%, resp.), whereas



(a) Resource profiles: eight clusters.



(b) Specialisation profiles: three clusters.

Figure 6: Total Within-Cluster Sums of Squares ($WCSS$) plots to determine the number of clusters for resource profiles (a) and specialisation profiles (b).

clusters 3, 5, and 7 are more evenly spread among Day and Evening. Night is only weakly linked to clusters 8, 7, and 6 (4.16%, 2.74%, and 1.46%, resp.). This was expected, as only 1% of all multitask sessions occurred during the Night shift, which usually involved activities of nurses who had to finish their Evening shift after 10 p.m. (i.e. the starting time of the Night shift).

Regarding the Education of the nurses in Table 7c, some clusters are always performed by nurses with a particular education level, e.g. cluster 1 and 7 is always performed by an MBVO-V nurse (100.00% and 95.18%, resp.), whereas cluster 3 always involves a HBO-V nurse (100.00%). With respect to the Function, in almost all clusters, except cluster 3, the Nurse is dominant (Table 7d), which is not surprising as 82.4% of all nurses who participated had the function Nurse. Nevertheless, in clusters 2, 6, and 8, different functions of nurses are also likely to be involved, besides Nurse.

Concerning the Weekday on which the multitask sessions were performed (Table 7e), we notice that the first half of the workweek (Mon–Wed) is often more probable than the second half (Thu–Fri) and the weekend (Sat–Sun). However, this may have been caused by the availability of the student nurse who recorded the shifts.

The categories of the Tasks executed during the multitask sessions are shown in Table 8. The probabilities for the different categories are not mutually exclusive, as multiple tasks for different categories can be performed concurrently during a multitask session. Therefore, the columns do not sum up to 100%. Professional communication, Indirect patient care, and Immediate patient care are frequently observed in all clusters, as well as Time between in transit and Medication tasks, although to a lesser degree.

Lastly, the distributions for the Duration and Multitasking level for the multitask session archetypes are shown in Table 9a and Table 9b, respectively. As illustrated in Figure 7, Clusters 1, 2, 3, 5, 7, and 9 have a relatively short duration, with a median of around 1.7 minutes, whereas cluster 6 has a significantly longer duration with a median of 22.37 minutes. This seems logical since the Multitasking level is also larger for the clusters with a longer duration, as shown in Figure 8, i.e. multitask sessions tend to take longer when more activities are performed concurrently.

By comparing the Multitasking level with the Multitasking level categories, we can determine whether a multitask session focuses mainly on one category of Tasks (i.e. a ratio closer to zero indicates a large difference between Multitasking level and Multitasking level categories), or many categories of Tasks were performed (i.e. a smaller difference with a ratio closer to one). We notice in Table 9b and Figure 8 that during multitask sessions with a higher Multitasking level (e.g. clusters 4 and 6), the ratio is also higher, meaning that when more

Ward	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9
Gastroenterology	< 0.01	< 0.01	3.57	< 0.01	< 0.01	1.47	< 0.01	< 0.01	< 0.01
Geriatrics	**89.15**	< 0.01	**96.43**	11.55	**34.35**	2.41	< 0.01	< 0.01	< 0.01
HPB Surgery	< 0.01	< 0.01	< 0.01	3.90	4.56	**28.13**	4.04	< 0.01	**40.51**
Intensive Care	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	17.10	< 0.01	**33.57**	< 0.01
Internal med-system diseases	< 0.01	< 0.01	< 0.01	17.90	10.21	15.92	**36.09**	< 0.01	**45.80**
Neurology	10.85	< 0.01	< 0.01	**22.77**	< 0.01	16.55	**21.53**	< 0.01	13.69
Oncology-Hematology	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	14.80	< 0.01	**66.43**	< 0.01
Pulmonary medicine	< 0.01	< 0.01	< 0.01	1.46	15.26	0.01	< 0.01	< 0.01	< 0.01
Transplantations	< 0.01	**100.00**	< 0.01	**42.40**	**35.62**	3.62	**38.33**	< 0.01	< 0.01

(a) Intra-cluster Ward distributions.

Shift	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9
Day	**99.60**	**74.71**	**57.14**	**71.40**	**49.11**	**74.86**	**55.49**	**26.10**	**100.00**
Evening	0.40	**25.29**	**42.86**	**28.60**	**50.89**	**23.67**	**41.77**	**69.74**	< 0.01
Night	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1.46	2.74	4.16	< 0.01

(b) Intra-cluster Shift distributions.

Education	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9
HBO-V	< 0.01	**53.22**	**100.00**	**38.88**	**42.52**	**44.44**	4.82	**38.14**	14.41
In-service trained	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**30.78**	< 0.01	19.68	**85.58**
MBVO-V	**100.00**	< 0.01	< 0.01	**60.62**	**57.48**	**20.13**	**95.18**	< 0.01	< 0.01
Student HBO-V	< 0.01	**46.78**	< 0.01	0.50	< 0.01	4.66	< 0.01	**42.18**	< 0.01

(c) Intra-cluster Education distributions.

Function	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9
High care nurse	< 0.01	**37.19**	< 0.01	4.66	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
ICU nurse	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	17.04	< 0.01	**33.56**	< 0.01
Nurse	**100.00**	**51.97**	< 0.01	**95.34**	**100.00**	**80.75**	**100.00**	**66.44**	**100.00**
Nurse student	< 0.01	10.85	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Senior nurse	< 0.01	< 0.01	**100.00**	< 0.01	< 0.01	2.21	< 0.01	< 0.01	< 0.01

(d) Intra-cluster Function distributions.

Weekday	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9
Mon	< 0.01	< 0.01	< 0.01	15.67	< 0.01	**27.30**	**27.10**	13.19	**65.22**
Tue	15.52	**70.68**	< 0.01	**28.54**	2.33	**22.08**	19.55	**26.86**	< 0.01
Wed	< 0.01	< 0.01	**56.63**	**29.55**	**97.67**	14.22	< 0.01	5.75	< 0.01
Thu	< 0.01	< 0.01	**39.80**	12.91	< 0.01	14.35	**40.65**	**54.21**	< 0.01
Fri	**84.48**	< 0.01	< 0.01	6.24	< 0.01	2.45	< 0.01	< 0.01	< 0.01
Sat	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	14.82	< 0.01	< 0.01	**28.36**
Sun	< 0.01	**29.32**	3.57	7.09	< 0.01	4.78	12.70	< 0.01	6.42

(e) Intra-cluster Weekday distributions.

Table 7: Intra-cluster multinomial distributions for Ward (a), Shift (b), Education (c), Function (d), and Weekday (e) attributes (in %).

Task	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9
Documentation & reporting	6.45	6.96	8.16	**21.97**	6.24	**39.57**	4.03	13.52	5.16
Professional communication	**46.28**	**36.04**	**41.33**	**73.59**	**31.63**	**73.13**	**37.92**	**22.96**	**42.29**
Social personal time	3.82	9.55	9.18	18.32	4.79	**28.96**	12.46	7.55	11.46
Indirect patient care	**26.30**	**21.26**	17.35	**53.71**	**23.52**	**92.45**	**24.19**	**54.69**	**30.64**
Immediate patient care	**35.17**	**35.39**	**34.18**	**71.79**	**31.87**	**96.30**	**24.38**	**37.64**	**37.18**
Direct patient care	2.80	4.72	1.53	10.82	1.77	**34.64**	1.36	8.62	5.03
Time between in transit	12.16	11.49	9.18	**35.28**	14.53	**64.41**	16.44	**26.98**	17.23
Medication tasks	8.69	14.40	**22.96**	**32.94**	19.54	**54.01**	15.76	15.20	13.71
Other	6.40	10.63	7.65	**22.78**	8.74	**26.21**	7.17	4.60	5.62
Department related activities	5.84	9.78	3.06	13.26	4.79	7.60	5.34	1.13	5.96
Patient transportation	2.72	< 0.01	4.59	10.86	4.20	6.58	2.92	6.82	4.06

Table 8: Intra-cluster binomial distributions for each Task (in %). Note that, in contrast to the distributions in Table 7, the activity categories are one-hot encoded and are assumed to be independent, i.e. a multitask session may contain multiple activity categories concurrently. Therefore, the summation over the activity categories per cluster does not add up to 100%.

Tasks are performed concurrently, the categories of Tasks are also more diverse.

Duration	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9
Shape (k)	0.9565	0.9103	0.9042	1.2256	1.0102	1.6001	0.9716	0.9962	0.8643
Scale (λ)	3.0742	3.4208	3.4105	15.5272	2.9077	30.5687	3.0302	5.0480	3.7317

(a) Intra-cluster Weibull distributions for Duration.

Multitasking level	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9
Rate (λ)	0.8307	0.8404	0.8518	2.5913	0.6307	3.5677	0.7493	1.4841	1.2578
Rate (λ_{cat})	0.4313	0.4929	0.4755	1.9758	0.3388	2.4443	0.3456	0.9972	0.7558
Ratio (λ_{cat}/λ)	0.5192	0.5865	0.5582	0.7625	0.5371	0.6851	0.4612	0.6719	0.6009

(b) Intra-cluster Poisson distributions for Multitasking level (λ) and Multitasking level categories (λ_{cat}).

Table 9: Intra-cluster distributions for Duration (a) and Multitasking level (b).

Table 10 describes the nine identified multitask session archetypes based on the intra-cluster distributions from Tables 7–9, together with the relative size of the cluster to the entire enriched log. For instance, multitask session archetype 1 is mainly related to multitask sessions in the Geriatrics ward, in which most of the work is performed during the Day shift and involves much communication with, and immediate (direct) care of patients. Typically the multitasking level is relatively low – one or two tasks concurrently – and a short duration of 50 seconds to 3.5 minutes (25^{th} and 75^{th} percentile, respectively).

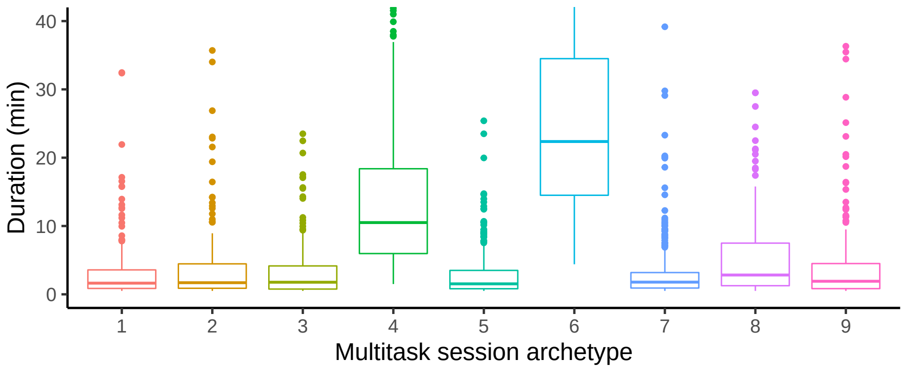


Figure 7: Durations of multitask sessions per multitask session archetype (in min).

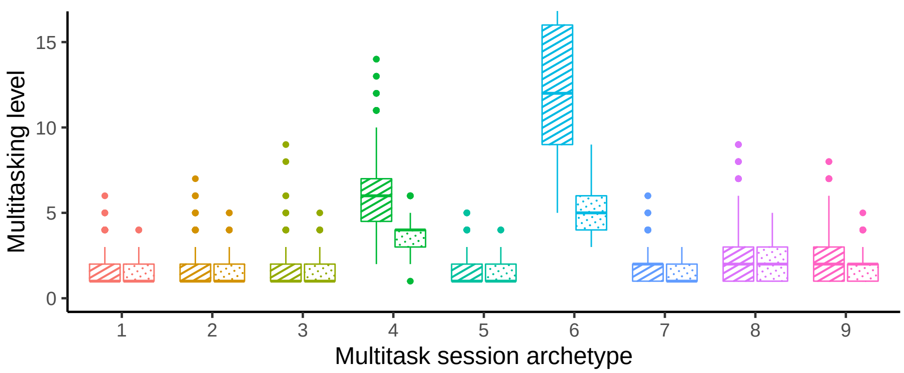


Figure 8: Multitasking level (λ, stripes) and Multitasking level categories (λ_{cat}, dots) of multitask sessions per multitask session archetype.

Clust.	**Description**	**Size (%)**
1	Mainly Geriatrics, sometimes Neurology; always during Day shift; by an MBVO-V Nurse; mainly on Fri, or sometimes on Tue; typically Professional communication and/or Immediate patient care; with a low multitasking level and low duration	10.02
2	Always Transplantations; mainly during the Day shift, sometimes in the Evening; evenly spread among HBO-V (Student) Nurse or High care nurse; mainly on Tue, or sometimes on Sun; typically Professional communication and/or Immediate patient care; with a low multitasking level and low duration	10.38

Continued on next page

Clust.	Description	Size (%)
3	Mainly Geriatrics, rarely Gastroenterology; evenly spread among the Day and Evening shift; by an HBO-V Sr. Nurse; frequently on Wed or Thu; typically Professional communication and/or Immediate patient care; with a low multitasking level and low duration	8.81
4	Frequently Transplantations, sometimes Neurology, Internal med-system diseases, or Geriatrics; mainly during the Day shift, sometimes in the Evening; mainly by an MBVO-V Nurse; spread among the workweek (Mon–Thu); typically Professional communication, Immediate patient care, Indirect patient care, Time between in transit, and/or Medication tasks; with a medium to high multitasking and a medium to high duration	10.38
5	Frequently Transplantations or Geriatrics, sometimes Pulmonary medicine or Internal med-system diseases; evenly spread among the Day and Evening shift; by an MBVO-V or HBO-V Nurse; always on Wed; typically Immediate patient care and/or Professional communication; with a low multitasking level and low duration	14.15
6	HBP surgery, Intensive care, Neurology, Internal med-system diseases, or Hematology-Oncology; mainly during the Day shift, sometimes in the Evening; frequently by an HBO-V or In-service trained Nurse of ICU nurse; spread among the workweek (Mon–Thu) or Sat; always involving Immediate patient care and typically Indirect patient care, Professional communication, Time between in transit, Medication tasks, Documentation & reporting, and or Direct patient care; with a high multitasking level and high duration	6.06
7	Frequently Transplantations or Internal med-system diseases, sometimes Neurology; evenly spread among the Day and Evening shift; always by an MBVO-V Nurse; typically on Thu, sometimes on Mon, Tue, or Sun; typically Professional communication; with a low multitasking level and low duration	14.96
8	Frequently Oncology-Hematology and, less frequently, Intensive care; mainly during the Evening shift, sometimes during the Day, and seldomly during Night; by a (Student) HBO-V Nurse or ICU nurse; frequently on Thu, sometimes on Mon–Tue; typically Indirect patient care and/or Immediate patient care; with a low to medium multitasking level and a low to medium duration	11.95

Continued on next page

Clust.	Description	Size (%)
9	Frequently Internal med-system diseases or HPB surgery, sometimes Neurology; always during the Day shift; mainly by an In-service trained Nurse; frequently on Mon, sometimes on Sat; typically Professional communication, Immediate patient care and/or Indirect patient care; with a low to medium multitasking level and a low duration	13.30

Table 10: Multitask session archetypes with descriptions and their relative size.

After identifying the multitask session archetypes, we can calculate for each resource the posterior probabilities that they belong to each multitask session archetype using Equation 6. Next, we can cluster the probabilities using AHC to find context-aware resource profiles and specialisation profiles. The table containing the posterior probabilities for each resource to belong to a particular multitask session archetype can be consulted in Appendix A.

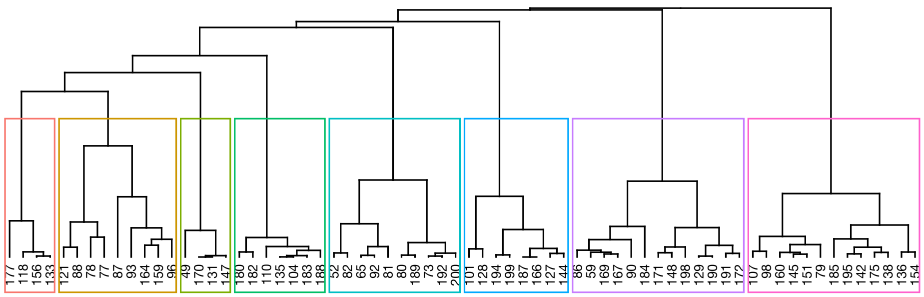
Figure 9a shows the resulting dendrogram for the resource profiles. For instance, nurses 104, 110, 135, 180, 182, 183, and 188 (green box) all have a high focus on multitask session archetype 9 (> 80%) and are, therefore, grouped into the same resource profile. The majority of these nurses are In-service trained and work in Internal med-system diseases, HPB surgery, or Neurology during the Day shift. They are typically occupied with Professional communication-, Immediate patient care-, and Indirect patient care-related activities with a medium Multitasking level (typically two or three activities concurrently) and a short duration (4.11 minutes on average). Another group of mainly MBVO-V nurses (79, 98, 107, 136, 138, 142, 145, 151, 154, 160, 175, 185, and 195; magenta box) work on multitask session archetypes 4 and 7, which are both linked to Transplantations, Neurology, and Internal med-system diseases, with a high focus on Professional communication. However, whereas archetype 7 exhibits a relatively low Multitasking level (typically one or two activities concurrently) and a short duration (on average 3.12 minutes), archetype 4 manifests a more demanding multitask session with a high Multitask level (typically four to seven activities concurrently) and a long duration (14.87 minutes on average). This would indicate that these nurses have a highly varied range of duties during their shift, with periods in which it is quieter and with periods in which many tasks have to be handled simultaneously.

Because resources are not necessarily bound to one multitask session archetype by the probabilistic nature of the model, a similar type of analysis can be applied to group resources among their degree of specialisation. For instance, all nurses of the left group (red box) in Figure 9b have a very high focus on one archetype ($> 90\%$) and are, therefore, referred to as "specialists". However, this does not imply that all nurses within this group focus on the same archetype, e.g. nurse 144 has a posterior probability of 92.10% performing multitask session archetype 2, whereas nurse 192 has 97.68% on archetype 8. Meanwhile, nurses of the right group (blue box) tend to spread their time among two or three archetypes and are, therefore, referred to as "generalists". Once again, the multitask session archetypes they work on can be very diverse, e.g. whereas nurse 98 works on archetypes 4, 6, and 7, nurse 101 works on 2 and 4.

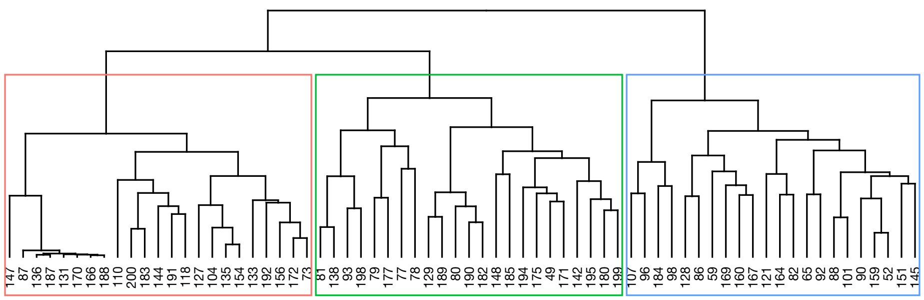
6. Discussion

As the work and role of nurses are rather dynamic, there are no formalised nurse roles defined by the hospital. Nevertheless, healthcare organisations could benefit from having a transparent overview of the different "roles", or more specifically, resource profiles of nurses working on similar activity instances. These in-depth insights could be used, e.g., to improve nurses' scheduling and determine the suitable levels of required nurses. In Sections 4 and 5 we demonstrated how ResProMin–MT can be used to obtain such a comprehensive overview. We discussed the results with domain experts currently working at the hospital to validate the findings. In total, three domain experts were involved during the validation: one PhD student and two professors in nursing science. Two of them were also involved during the collection of data that is used in this case study.

In this section, we first present the discussion with the domain experts on the



(a) Cluster dendrogram of the resource profiles. From left to right, the resources mainly work on multitask session archetypes: 1 (red); 4, 6 & 9 (orange); 3 (light green); 9 (green); 6 & 8 (turquoise); 2 & 4 (blue); 4 & 5 (violet); 4 & 7 (magenta).



(b) Cluster dendrogram of the resource specialisation profiles. From left to right: specialists (focus on one archetype, red); specialists (work on two archetypes but one with higher focus, green); generalists (work on two or three archetypes, blue).

Figure 9: Cluster dendrograms of the resource profiles (a) and specialisation profiles (b).

case study. Next, we provide some reflections on ResProMin–MT and highlight its limitations.

6.1. Case Study

A noteworthy observation of the results is the fact that every multitask session archetype contains at least one of the following activity types: Professional communication, Immediate patient care, or Indirect patient care. The domain experts confirmed that, besides the tasks related to patient care (which is usually seen as the most integral part of a nurse's responsibilities), these results highlight the importance of communication, which is recently gaining attention in the field of nursing science [82–84]. Another frequently observed pattern is the combination of Medication tasks, Indirect patient care and Time between in transit. An explanation provided by the domain experts was that the nurse first has to get and prepare the medication (Time between in transit & Indirect patient care) before administering it (Medication tasks).

Even though the domain experts indicated that there is no formal distinction made between the nurses in terms of the activities they perform based on their education, some multitask session archetypes are more likely to be carried out by a specific education level. For example, multitask session archetype 4 is in 60.62% of the cases performed by an MBVO-V nurse, compared to 38.88% for HBO-V. However, the domain experts agreed that there may be implicit task division patterns in the workplace. The revelation of these patterns can be seen as one of the advantages of ResProMin–MT.

6.2. Reflections & Limitations

In comparison to the state-of-the-art in organisational model mining algorithms, ResProMin–MT provides much richer information on the context in which the resources executed their activities. Other algorithms mainly present graphs with groups of resources working on the same activities [13, 14, 21], sometimes augmented with contextual variables such as the weekday or case type [23]. By contrast, our method benefits from the probabilistic nature of FMMs and provides more in-depth and transparent insights into *what* and *under which circumstances* resources execute their work, such as the nurse's experience, duration of the activities, workload in terms of multitasking, and many more. In addition, ResProMin–MT is capable of discerning specialists from generalists.

Despite ResProMin–MT's capability of finding interesting and valuable insights into the prevailing resource profiles, we have also identified five limitations resulting from the practical aspects of the data collection and the applied clustering technique.

Firstly, the nurse identifier or the mapping between the shifts and the individual nurses has not been recorded to safeguard privacy. If multiple shifts over time were linked to the same nurse, we would have had a more reliable overview of the activity types each nurse performs. However, the domain experts indicated that the activities performed by a nurse highly depend on the situation and activities that need to be done. After all, the experts pointed out that a nurse's role depends on the situation at hand and can differ from day to day. For instance, a senior and highly specialised nurse might also perform simple tasks if needed, which is often the case in periods of understaffing. As a consequence, the domain experts considered the outcome sufficient, and the impact of not being able to connect shifts performed by the same nurse is estimated as limited.

Secondly, we were unable to assess the systematic workload since not all nurses have been shadowed, nor did we had access to information on the patient's health condition. As a result, we could not determine whether the degree of multitasking was related to the specific activities, due to a high workload, or by the amount of care required depending on the patient acuity.

Thirdly, as nursing students performed the shadowing and recording, the data collection was limited to the recorders' availabilities, which may cause bias regarding time-related factors. This could explain why most multitask session archetypes tend to occur during the Day shift (from 7 a.m. to 3 p.m.).

Fourthly, the domain experts indicated that nurses working on multiple multitask session archetypes – which ResProMin–MT defines as "generalists" – could still be very specialised in their field because "specialisation" within a medical context refers to the type of work a nurse performs. However, it is not the aim of our method to identify specialisation within a medical context but rather to measure how resources divide their work among archetypes.

A final limitation relates to the computational efficiency of clustering the enriched log using Finite Mixture Models. While FMMs provide several advantages, fitting an appropriate model suffers from the *curse of dimensionality* when using large logs with many contextual variables. However, determining the resource profiles is rather a post-analysis process and is typically not performed in real-time [24]. Therefore, runtime optimisation was not a primary criterion for this application.

Notwithstanding the presence of these limitations, the domain experts indicated that our method provides valuable and conveniently interpretable insights for determining how care managers should design their nursing processes. For example, the internal employment agency of the hospital could use it to assign nurses to wards based on the current needs within those wards. In addition, the method could be used by nurse managers to form nursing teams which contain all the profiles required for their ward, e.g. 50% of nurses working on archetype x, 25% on archetype y, and 25% on archetype z.

While the case study in this work was applied to a healthcare context, ResProMin–MT can be used as well in other domains where multitasking is relevant and give in-depth insights into resource's roles and behaviour within organisations. In addition, the first steps of our method form a basis for applying other analyses than resource profile identification. For instance, for preparing the input for mining resource assignment rules [58–61], and determining frequently observed work routines and implicit task division patterns using the archetypes.

7. Conclusion

In this paper, we introduced a method to discover context-aware resource profiles from event logs in the presence of multitasking, i.e. *ResProMin–MT*. In addition, we demonstrated that our method is capable of taking into account more complex activity dimensions, such as durations. Despite the challenges that arose from the data, we demonstrated the feasibility of ResProMin–MT in a healthcare context. The output of our method was validated by domain experts in nursing science. The insights obtained from ResProMin–MT provide an extensive overview of the complex relationship between resources and care activities, which can be used, e.g., to support healthcare managers in efficiently allocating their resources during periods of understaffing. Nevertheless, the field of application is not limited to healthcare. Whenever processes are strongly dependent on human resources and multitasking is relevant, ResProMin–MT can provide a comprehensive and transparent overview on the behaviour of resources within organisations.

We identify several directions for future work. Firstly, as Patient-Centered Care (PCC) is becoming increasingly implemented by healthcare organisations to improve health outcomes [85, 86], patient data, such as the patient's health condition or illness, might provide interesting additional insights into the organisation of nursing activities. Secondly, the overall workload in a particular department and additional attributes of the nurses' work experience can be used to augment the profiling of resources further. Thirdly, the enrichment of an event log with context-related information and interpretation of the archetypes could be facilitated through the development of toolings. These tools could provide guidance on which contextual attributes might be interesting to include in the enriched log and transform the data accordingly. Additionally, the interpretation of the archetypes could be facilitated by automatically generating descriptions based on the fitted Finite Mixture Model. Finally, the method's computational efficiency could be improved using heuristics. For example, the convergence of the EM algorithm could be accelerated using a quasi-Newton approach [87] while still obtaining near-optimal solutions.

Acknowledgements

The authors would like to thank Tim Korteland, Prof. Dr. Erwin Ista, and Prof. Dr. Monique van Dijk of the Erasmus University Medical Center Rotterdam, Department of Internal Medicine, division of Nursing Science for their time to assess and validate our findings.

This study was supported by the Special Research Fund (BOF) of Hasselt University under Grant No. BOF19OWB20.

The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation – Flanders (FWO) and the Flemish Government.

References

[1] C. Hicks, T. McGovern, G. Prior, I. Smith, Applying Lean Principles to the Design of Healthcare Facilities, International Journal of Production Economics 170 (2015) 677–686. doi:10.1016/j.ijpe.2015.05.029.

[2] R. S. Mans, W. M. P. van der Aalst, R. J. B. Vanwersch, Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes, no. 13170 in SpringerBriefs in Business Process Management, Springer International Publishing, Cham, Switzerland, 2015. doi:10.1007/978-3-319-16071-9.

[3] N. Martin, J. De Weerdt, C. Fernández-Llatas, A. Gal, R. Gatta, G. Ibáñez, O. Johnson, F. Mannhardt, L. Marco-Ruiz, S. Mertens, J. Munoz-Gama, F. Seoane, J. Vanthienen, M. T. Wynn, D. B. Boilève, J. Bergs, M. Joosten-Melis, S. Schretlen, B. Van Acker, Recommendations for Enhancing the Usability and Understandability of Process Mining in Healthcare, Artificial Intelligence in Medicine 109 (2020) 101962. doi:10.1016/j.artmed.2020.101962.

[4] D. McCaughey, C. O. Erwin, J. L. DelliFraine, Improving Capacity Management in the Emergency Department: A Review of the Literature, 2000-2012, Journal of Healthcare Management 60 (1) (2015) 63. doi:10.1097/00115514-201501000-00011.

[5] V. L. Smith-Daniels, S. B. Schweikhart, D. E. Smith-Daniels, Capacity Management in Health Care Services: Review and Future Research Directions, Decision Sciences 19 (4) (1988) 889–919. doi:10.1111/j.1540-5915.1988.tb00310.x.

[6] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers, Fundamentals of Business Process Management, 2nd Edition, Springer-Verlag, Berlin, Germany, 2018. doi:10.1007/978-3-662-56509-4.

[7] H. A. Reijers, S. Liman Mansar, Best Practices in Business Process Redesign: An Overview and Qualitative Evaluation of Successful Redesign Heuristics, Omega 33 (4) (2005) 283–306. doi:10.1016/j.omega.2004.04.012.

[8] A. Rule, M. F. Chiang, M. R. Hribar, Using Electronic Health Record Audit Logs to Study Clinical Activity: A Systematic Review of Aims, Measures, and Methods, Journal of the American Medical Informatics Association 27 (3) (2020) 480–490. doi:10.1093/jamia/ocz196.

[9] W. M. P. van der Aalst, Process Mining: Data Science in Action, 2nd Edition, Springer, Heidelberg, Germany, 2016. doi:10.1007/978-3-662-49851-4.

[10] Z. A. Abdalkareem, A. Amir, M. A. Al-Betar, P. Ekhan, A. I. Hammouri, Healthcare Scheduling in Optimization Context: A Review, Health and Technology 11 (3) (2021) 445–469. doi:10.1007/s12553-021-00547-5.

[11] R. G. Drake, The Nurse Rostering Problem: From Operational Research to Organizational Reality?, Journal of Advanced Nursing 70 (4) (2014) 800–810. doi:10.1111/jan.12238.

[12] C. E. Saville, P. Griffiths, J. E. Ball, T. Monks, How Many Nurses Do We Need? A Review and Discussion of Operational Research Techniques Applied to Nurse Staffing, International Journal of Nursing Studies 97 (2019) 7–13. doi:10.1016/j.ijnurstu.2019.04.015.

[13] M. Song, W. M. P. van der Aalst, Towards Comprehensive Support for Organizational Mining, Decision Support Systems 46 (1) (2008) 300–317. doi:10.1016/j.dss.2008.07.002.

[14] A. Appice, Towards Mining the Organizational Structure of a Dynamic Event Scenario, Journal of Intelligent Information Systems 50 (1) (2018) 165–193. doi:10.1007/s10844-017-0451-x.

[15] A. Burattin, A. Sperduti, M. Veluscek, Business Models Enhancement Through Discovery of Roles, in: B. Hammer, Z.-H. Zhou, L. Wang, N. Chawla (Eds.), Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining, CIDM '13, IEEE, Singapore, Singapore, 2013, pp. 103–110. doi:10.1109/CIDM.2013.6597224.

[16] D. M. R. Ferreira, C. Alves, Discovering User Communities in Large Event Logs, in: F. Daniel, K. Barkaoui, S. Dustdar (Eds.), Proceedings of the Business Process Management 2011 International Workshops, Vol. 99 of Lecture Notes in Business Information Processing, Springer, Clermont-Ferrand, France, 2012, pp. 123–134. doi:10.1007/978-3-642-28108-2_11.

[17] T. Jin, J. Wang, L. Wen, Organizational Modeling from Event Logs, in: Z. Xu, I. Foster, X. Sun (Eds.), Proceedings of the 6th International Conference on Grid and Cooperative Computing, GCC '07, IEEE, Los Alamitos, CA, USA, 2007, pp. 670–675. doi:10.1109/GCC.2007.93.

[18] Z. Ni, S. Wang, H. Li, Mining Organizational Structure from Workflow Logs, in: Proceeding of the International Conference on E-Education, Entertainment and e-Management, ICEEE '11, IEEE, Bali, Indonesia, 2011, pp. 222–225. doi:10.1109/ICeEEM.2011.6137791.

[19] A. Pika, M. Leyer, M. T. Wynn, C. J. Fidge, A. H. M. ter Hofstede, W. M. P. van der Aalst, Mining Resource Profiles from Event Logs, ACM Transactions on Management Information Systems 8 (1) (2017) 1:1–1:30. doi:10.1145/3041218.

[20] W. M. P. van der Aalst, H. A. Reijers, M. Song, Discovering Social Networks from Event Logs, Computer Supported Cooperative Work 14 (6) (2005) 549–593. doi:10.1007/s10606-005-9005-9.

[21] J. Ye, Z. Li, K. Yi, A. Al-Ahmari, Mining Resource Community and Resource Role Network From Event Logs, IEEE Access 6 (2018) 77685–77694. doi:10.1109/ACCESS.2018.2883774.

[22] J. R. Yang, C. Ouyang, M. Pan, Y. Yu, A. H. M. ter Hofstede, Finding the "Liberos": Discover Organizational Models with Overlaps, in: M. Weske, M. Montali, I. Weber, J. vom Brocke (Eds.), Proceedings of the 16th International Conference on Business Process Management, Vol. 11080 of Lecture Notes in Computer Science, Springer International Publishing, Sydney, NSW, Australia, 2018, pp. 339–355. doi:10.1007/978-3-319-98648-7_20.

[23] J. R. Yang, C. Ouyang, W. M. P. van der Aalst, A. H. M. ter Hofstede, Y. Yu, OrdinoR: A Framework for Discovering, Evaluating, and Analyzing Organizational Models Using Event Logs, Decision Support Systems 158 (2022) 113771. doi:10.1016/j.dss.2022.113771.

[24] G. van Hulzen, N. Martin, B. Depaire, Looking Beyond Activity Labels: Mining Context-Aware Resource Profiles using Activity Instance Archetypes, in: A. Polyvyanyy, M. T. Wynn, A. Van Looy, M. Reichert (Eds.), Proceedings of the Business Process Management Forum 2021, Vol. 427 of Lecture Notes in Business Information Processing, Springer Nature Switzerland, Rome, Italy, 2021, pp. 230–245. doi:10.1007/978-3-030-85440-9_14.

[25] C. Di Ciccio, A. Marrella, A. Russo, Knowledge-Intensive Processes: Characteristics, Requirements and Analysis of Contemporary Approaches, Journal on Data Semantics 4 (1) (2015) 29–57. doi:10.1007/s13740-014-0038-4.

[26] N. Martin, N. Wittig, J. Munoz-Gama, Using Process Mining in Healthcare, in: W. M. P. van der Aalst, J. Carmona (Eds.), Process Mining Handbook, Vol. 448 of Lecture Notes in Business Information Processing, Springer International Publishing, Cham, Switzerland, 2022, pp. 416–444. doi:10.1007/978-3-031-08848-3_14.

[27] Á. Rebuge, D. M. R. Ferreira, Business Process Analysis in Healthcare

Environments: A Methodology Based on Process Mining, Information Systems 37 (2) (2012) 99–116. doi:10.1016/j.is.2011.01.003.

[28] H. E. Douglas, M. Z. Raban, S. R. Walter, J. I. Westbrook, Improving Our Understanding of Multi-Tasking in Healthcare: Drawing Together the Cognitive Psychology and Healthcare Literature, Applied Ergonomics 59 (2017) 45–55. doi:10.1016/j.apergo.2016.08.021.

[29] C. Alvarez, E. Rojas, M. Arias, J. Munoz-Gama, M. Sepúlveda, V. Herskovic, D. Capurro, Discovering Role Interaction Models in the Emergency Room Using Process Mining, Journal of Biomedical Informatics 78 (2018) 60–77. doi:10.1016/j.jbi.2017.12.015.

[30] E. De Roock, N. Martin, Process Mining in Healthcare – An Updated Perspective on the State of the Art, Journal of Biomedical Informatics 127 (2022) 103995. doi:10.1016/j.jbi.2022.103995.

[31] M. R. Dallagassa, C. dos Santos Garcia, E. E. Scalabrin, S. O. Ioshii, D. R. Carvalho, Opportunities and Challenges for Applying Process Mining in Healthcare: A Systematic Mapping Study, Journal of Ambient Intelligence and Humanized Computing 13 (1) (2022) 165–182. doi:10.1007/s12652-021-02894-7.

[32] A. Guzzo, A. Rullo, E. Vocaturo, Process Mining Applications in the Healthcare Domain: A Comprehensive Review, WIREs Data Mining and Knowledge Discovery 12 (2) (2022) e1442. doi:10.1002/widm.1442.

[33] L. Placidi, L. Boldrini, J. Lenkowicz, S. Manfrida, R. Gatta, A. Damiani, S. Chiesa, F. Ciellini, V. Valentini, Process Mining to Optimize Palliative Patient Flow in a High-Volume Radiotherapy Department, Technical Innovations & Patient Support in Radiation Oncology 17 (2021) 32–39. doi:10.1016/j.tipsro.2021.02.005.

[34] E. Rojas, A. Cifuentes, A. Burattin, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Performance Analysis of Emergency Room Episodes Through

Process Mining, International Journal of Environmental Research and Public Health 16 (7) (2019) 1274. doi:10.3390/ijerph16071274.

[35] S. V. Kovalchuk, A. A. Funkner, O. G. Metsker, A. N. Yakovlev, Simulation of Patient Flow in Multiple Healthcare Units Using Process and Data Mining Techniques for Model Identification, Journal of Biomedical Informatics 82 (2018) 128–142. doi:10.1016/j.jbi.2018.05.004.

[36] A. Najjar, D. Reinharz, C. Girouard, C. Gagné, A Two-Step Approach for Mining Patient Treatment Pathways in Administrative Healthcare Databases, Artificial Intelligence in Medicine 87 (2018) 34–48. doi:10.1016/j.artmed.2018.03.004.

[37] M. Prodel, V. Augusto, X. Xie, B. Jouaneton, L. Lamarsalle, Discovery of Patient Pathways from a National Hospital Database Using Process Mining and Integer Linear Programming, in: B. Lennartson, M. Fabian, S. Reveliotis, J.-J. Lesage (Eds.), Proceedings of the 11th IEEE International Conference on Automation Science and Engineering, IEEE, Gothenburg, Sweden, 2015, pp. 1409–1414. doi:10.1109/CoASE.2015.7294295.

[38] F. Mannhardt, D. Blinde, Analyzing the Trajectories of Patients with Sepsis Using Process Mining, in: J. Gulden, S. Nurcan, I. Reinhartz-Berger, W. Guédria, P. Bera, S. Guerreiro, M. Fellmann, M. Weidlich (Eds.), Joint Proceedings of the RADAR Tracks at the 18th International Working Conference on Business Process Modeling, Development and Support (BPMDS), and the 22nd International Working Conference on Evaluation and Modeling Methods for Systems Analysis and Development (EMMSAD), and the 8th International Workshop on Enterprise Modeling and Information Systems Architectures (EMISA), Vol. 1859 of CEUR Workshop Proceedings, CEUR Workshop Proceedings, Essen, Germany, 2017, pp. 72–80.

[39] H. Xu, J. Pang, X. Yang, L. Ma, H. Mao, D. Zhao, Applying Clinical Guidelines to Conformance Checking for Diagnosis and Treatment: A Case Study of Ischemic Stroke, in: T. Park, Y.-R. Cho, X. Hu, I. Yoo, H. G.

Woo, J. Wang, J. Facelli, S. Nam, M. Kang (Eds.), Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM '20, IEEE, Seoul, Korea (South), 2020, pp. 2125–2130. doi:10.1109/BIBM49941.2020.9313532.

[40] E. Asare, L. Wang, X. Fang, Conformance Checking: Workflow of Hospitals and Workflow of Open-Source EMRs, IEEE Access 8 (2020) 139546–139566. doi:10.1109/ACCESS.2020.3012147.

[41] R. Andrews, M. T. Wynn, K. Vallmuur, A. H. M. ter Hofstede, E. Bosley, A Comparative Process Mining Analysis of Road Trauma Patient Pathways, International Journal of Environmental Research and Public Health 17 (10) (2020) 3426. doi:10.3390/ijerph17103426.

[42] H. Xu, H. Yan, J. Pang, S. Nan, X. Yang, D. Zhao, Evaluating the Relative Value of Care Interventions Based on Clinical Pathway Variation Detection and Propensity Score, in: T. Park, Y.-R. Cho, X. Hu, I. Yoo, H. G. Woo, J. Wang, J. Facelli, S. Nam, M. Kang (Eds.), Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM '20, IEEE, Seoul, Korea (South), 2020, pp. 1184–1187. doi:10.1109/BIBM49941.2020.9313546.

[43] A. Stefanini, D. Aloini, E. Benevento, R. Dulmin, V. Mininno, Performance Analysis in Emergency Departments: A Data-Driven Approach, Measuring Business Excellence 22 (2) (2018) 130–145. doi:10.1108/MBE-07-2017-0040.

[44] M. Cho, M. Song, J. Park, S.-R. Yeom, I.-J. Wang, B.-K. Choi, Process Mining-Supported Emergency Room Process Performance Indicators, International Journal of Environmental Research and Public Health 17 (17) (2020) 6290. doi:10.3390/ijerph17176290.

[45] E. Benevento, D. Aloini, N. Squicciarini, R. Dulmin, V. Mininno, Queue-Based Features for Dynamic Waiting Time Prediction in Emergency De-

partment, Measuring Business Excellence 23 (4) (2019) 458–471. doi:10.1108/MBE-12-2018-0108.

[46] A. W. Kempa-Liehr, C. Y.-C. Lin, R. Britten, D. Armstrong, J. Wallace, D. Mordaunt, M. O'Sullivan, Healthcare Pathway Discovery and Probabilistic Machine Learning, International Journal of Medical Informatics 137 (2020) 104087. doi:10.1016/j.ijmedinf.2020.104087.

[47] B. B. P. Antunes, A. Manresa, L. S. L. Bastos, J. F. Marchesi, S. Hamacher, A Solution Framework Based on Process Mining, Optimization, and Discrete-Event Simulation to Improve Queue Performance in an Emergency Department, in: C. Di Francescomarino, R. M. Dijkman, U. Zdun, W. M. P. van der Aalst, J. Mylopoulos, M. Rosemann, M. J. Shaw, C. Szyperski (Eds.), Proceedings of the Business Process Management Workshops, Vol. 362 of Lecture Notes in Business Information Processing, Springer International Publishing, Vienna, Austria, 2019, pp. 583–594. doi:10.1007/978-3-030-37453-2_47.

[48] R. S. Mans, H. A. Reijers, M. van Genuchten, D. Wismeijer, Mining processes in dentistry, in: G. Luo, J. Liu, Yang (Eds.), Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, IHI '12, Association for Computing Machinery, Miami, FL, USA, 2012, pp. 379–388. doi:10.1145/2110363.2110407.

[49] T. Conca, C. Saint-Pierre, V. Herskovic, M. Sepúlveda, D. Capurro, F. Prieto, C. Fernandez-Llatas, Multidisciplinary Collaboration in the Treatment of Patients With Type 2 Diabetes in Primary Care: Analysis Using Process Mining, Journal of Medical Internet Research 20 (4) (2018) e8884. doi:10.2196/jmir.8884.

[50] A. B. Durojaiye, S. Levin, M. Toerper, H. Kharrazi, H. P. Lehmann, A. P. Gurses, Evaluation of Multidisciplinary Collaboration in Pediatric Trauma Care Using EHR Data, Journal of the American Medical Informatics Association 26 (6) (2019) 506–515. doi:10.1093/jamia/ocy184.

[51] F. Caron, J. Vanthienen, J. De Weerdt, B. Baesens, Beyond X-Raying a Case-Flow: Adopting Different Focuses on Care-Flow Mining, in: B. F. van Dongen, B. Weber, D. M. R. Ferreira (Eds.), Proceedings of the 1st International Business Process Intelligence Challenge, BPM '11, Clermont-Ferrand, France, 2011, pp. 1–11.

[52] M. R. Naeem, H. Naeem, M. Aamir, W. Ali, W. A. Abro, A Multi-Level Process Mining Framework for Correlating and Clustering of Biomedical Activities using Event Logs, International Journal of Advanced Computer Science and Applications 8 (3) (2017) 393–401. doi:10.14569/IJACSA.2017.080354.

[53] G. Riz, E. A. P. Santos, E. d. F. R. Loures, Process Mining to Knowledge Discovery in Healthcare Processes, in: M. Borsato, N. Wognum, M. Peruzzini, J. Stjepandić, W. J. C. Verhagen (Eds.), Proceedings of the 23rd ISPE Inc. International Conference on Transdisciplinary Engineering, Vol. 4 of Advances in Transdisciplinary Engineering: Crossing Boundaries, IOS Press, Curitiba, Parana, Brazil, 2016, pp. 1019–1028. doi:10.3233/978-1-61499-703-0-1019.

[54] M. Ghasemi, D. Amyot, Process Mining in Healthcare: A Systematised Literature Review, International Journal of Electronic Healthcare 9 (1) (2016) 60–88. doi:10.1504/IJEH.2016.078745.

[55] E. Rojas, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Process Mining in Healthcare: A Literature Review, Journal of Biomedical Informatics 61 (2016) 224–236. doi:10.1016/j.jbi.2016.04.007.

[56] H. Ahn, K. P. Kim, Formal Approach for Discovering Work Transference Networks from Workflow Logs, Information Sciences 515 (2020) 1–25. doi:10.1016/j.ins.2019.11.036.

[57] S. Dustdar, T. Hoffmann, Interaction Pattern Detection in Process Oriented Information Systems, Data & Knowledge Engineering 62 (1) (2007) 138–155. doi:10.1016/j.datak.2006.07.010.

[58] S. Schönig, C. Cabanillas, S. Jablonski, J. Mendling, A Framework for Efficiently Mining the Organisational Perspective of Business Processes, Decision Support Systems 89 (2016) 87–97. doi:10.1016/j.dss.2016.06.012.

[59] C. Cabanillas, L. Ackermann, S. Schönig, C. Sturm, J. Mendling, The RALph Miner for Automated Discovery and Verification of Resource-Aware Process Models, Software and Systems Modeling 19 (2020) 1415–1441. doi:10.1007/s10270-020-00820-7.

[60] Z. Huang, X. Lu, H. Duan, Mining Association Rules to Support Resource Allocation in Business Process Management, Expert Systems with Applications 38 (8) (2011) 9483–9490. doi:10.1016/j.eswa.2011.01.146.

[61] Y. Liu, J. Wang, Y. Yang, J. Sun, A Semi-Automatic Approach for Workflow Staff Assignment, Computers in Industry 59 (5) (2008) 463–476. doi:10.1016/j.compind.2007.12.002.

[62] J. Nakatumba, W. M. P. van der Aalst, Analyzing Resource Behavior Using Process Mining, in: S. Rinderle-Ma, S. Sadiq, F. Leymann (Eds.), Proceedings of the Business Process Management Workshops, Vol. 43 of Lecture Notes in Business Information Processing, Springer, Ulm, Germany, 2010, pp. 69–80. doi:10.1007/978-3-642-12186-9_8.

[63] S. Suriadi, M. T. Wynn, J. Xu, W. M. P. van der Aalst, A. H. M. ter Hofstede, Discovering Work Prioritisation Patterns from Event Logs, Decision Support Systems 100 (2017) 77–92. doi:10.1016/j.dss.2017.02.002.

[64] E. L. Klijn, F. Mannhardt, D. Fahland, Classifying and Detecting Task Executions and Routines in Processes Using Event Graphs, in: A. Polyvyanyy, M. T. Wynn, A. Van Looy, M. Reichert (Eds.), Proceedings of the Business Process Management 2021 Forum, Vol. 427 of Lecture Notes in Business Information Processing, Springer International Publishing, Rome, Italy, 2021, pp. 212–229. doi:10.1007/978-3-030-85440-9_13.

[65] N. Martin, L. Pufahl, F. Mannhardt, Detection of Batch Activities from Event Logs, Information Systems 95 (2021) 101642. doi:10.1016/j.is.2020.101642.

[66] N. Martin, M. Swennen, B. Depaire, M. J. Jans, A. Caris, K. Vanhoof, Retrieving Batch Organisation of Work Insights from Event Logs, Decision Support Systems 100 (2017) 119–128. doi:10.1016/j.dss.2017.02.012.

[67] N. Martin, G. Van Houdt, G. Janssenswillen, DaQAPO: Supporting Flexible and Fine-Grained Event Log Quality Assessment, Expert Systems with Applications 191 (2022) 116274. doi:10.1016/j.eswa.2021.116274.

[68] G. Janssenswillen, B. Depaire, M. Swennen, M. J. Jans, K. Vanhoof, bupaR: Enabling Reproducible Business Process Analysis, Knowledge-Based Systems 163 (2019) 927–930. doi:10.1016/j.knosys.2018.10.018.

[69] G. J. McLachlan, K. E. Basford, Mixture Models: Inference and Applications to Clustering, Vol. 84 of Statistics: Textbooks and Monographs, Marcel Dekker, New York, NY, USA, 1988.

[70] J. K. Vermunt, J. Magidson, Latent Class Cluster Analysis, in: J. A. Hagenaars, A. L. McCutcheon (Eds.), Applied Latent Class Analysis, Cambridge University Press, Cambridge, United Kingdom, 2002, pp. 89–106. doi:10.1017/CBO9780511499531.004.

[71] G. J. McLachlan, S. X. Lee, S. I. Rathnayake, Finite Mixture Models, Annual Review of Statistics and Its Application 6 (1) (2019) 355–378. doi:10.1146/annurev-statistics-031017-100325.

[72] S. Frühwirth-Schnatter, Finite Mixture and Markov Switching Models, no. 692 in Springer Series in Statistics, Springer, New York, NY, USA, 2006. doi:10.1007/978-0-387-35768-3.

[73] G. J. McLachlan, S. I. Rathnayake, S. X. Lee, Model-Based Clustering, in: S. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics,

2nd Edition, Vol. 2, Elsevier, Oxford, United Kingdom, 2020, pp. 509–529. doi:10.1016/B978-0-12-409547-2.14649-9.

[74] G. J. McLachlan, Model-Based Clustering, in: S. D. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, Vol. 2, Elsevier, Oxford, United Kingdom, 2009, pp. 655–681. doi:10.1016/B978-044452701-1.00068-5.

[75] J. F. Hair, Jr., W. C. Black, B. J. Babin, R. E. Anderson, Multivariate Data Analysis, seventh Edition, Pearson, Upper Saddle River, NJ, USA, 2009.

[76] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 39 (1) (1977) 1–38. doi:10.1111/j.2517-6161.1977.tb01600.x.

[77] K. E. Masyn, Latent Class Analysis and Finite Mixture Modeling, in: T. D. Little (Ed.), The Oxford Handbook of Quantitative Methods in Psychology, Vol. 2: Statistical Analysis, Oxford University Press, New York, NY, USA, 2013, pp. 551–611. doi:10.1093/oxfordhb/9780199934898.001.0001.

[78] G. J. McLachlan, D. Peel, Finite Mixture Models, no. 1345 in Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York, NY, USA, 2000. doi:10.1002/0471721182.

[79] B. E. Everitt, S. Landau, M. Leese, D. Stahl, Cluster Analysis, 5th Edition, no. 1345 in Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester, West Sussex, UK, 2011. doi:10.1002/9780470977811.

[80] B. Grün, F. Leisch, FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters, Journal of Statistical Software 28 (1) (2008) 1–35. doi:10.18637/jss.v028.i04.

[81] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing (2021).
URL https://www.R-project.org

[82] J. B. Riley, Communication in Nursing, ninth Edition, Elsevier, St. Louis, MO, USA, 2019.

[83] L. Kourkouta, I. V. Papathanasiou, Communication in Nursing Practice, Materia Socio-Medica 26 (1) (2014) 65–67. doi:10.5455/msm.2014.26.65-67.

[84] J. I. Westbrook, C. Duffield, L. Li, N. J. Creswick, How Much Time Do Nurses Have for Patients? A Longitudinal Study Quantifying Hospital Nurses' Patterns of Task Time Distribution and Interactions with Health Professionals, BMC Health Services Research 11 (1) (2011) 319. doi:10.1186/1472-6963-11-319.

[85] J. H. Robinson, L. C. Callister, J. A. Berry, K. A. Dearing, Patient-Centered Care and Adherence: Definitions and Applications to Improve Outcomes, Journal of the American Academy of Nurse Practitioners 20 (12) (2008) 600–607. doi:10.1111/j.1745-7599.2008.00360.x.

[86] R. M. Epstein, R. L. Street, The Values and Value of Patient-Centered Care, The Annals of Family Medicine 9 (2) (2011) 100–103. doi:10.1370/afm.1239.

[87] G. J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, 2nd Edition, no. 1345 in Wiley Series in Probability and Statistics, John Wiley & Sons, Hoboken, NJ, USA, 2007. doi:10.1002/9780470191613.

Appendix A. Additional Tables with Resource Probabilities from Step 3

Table A.11 contains the posterior probabilities for each resource to belong to a particular multitask session archetype and is used to find the context-aware resource profiles. Table A.12 contains the same probabilities as Table A.11, but ordered from largest to smallest, and is used to find the resource specialisation profiles.

Resource	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9	Profile
118	**92.60**	< 0.01	< 0.01	7.40	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
133	**96.57**	< 0.01	< 0.01	2.31	< 0.01	1.12	< 0.01	< 0.01	< 0.01	1
156	**96.95**	< 0.01	< 0.01	3.05	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
177	**63.15**	< 0.01	< 0.01	30.23	< 0.01	6.63	< 0.01	< 0.01	< 0.01	1
121	< 0.01	< 0.01	< 0.01	**49.93**	< 0.01	**50.07**	< 0.01	< 0.01	< 0.01	2
159	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**42.83**	< 0.01	< 0.01	**57.17**	2
164	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**54.70**	< 0.01	< 0.01	**45.30**	2
77	< 0.01	< 0.01	< 0.01	**68.29**	< 0.01	**17.29**	14.42	< 0.01	< 0.01	2
78	23.35	< 0.01	< 0.01	**64.15**	< 0.01	**12.50**	< 0.01	< 0.01	< 0.01	2
87	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**100.00**	< 0.01	< 0.01	< 0.01	2
88	< 0.01	< 0.01	< 0.01	**59.77**	< 0.01	**40.23**	< 0.01	< 0.01	< 0.01	2
93	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**69.97**	< 0.01	< 0.01	**30.03**	2
96	< 0.01	< 0.01	< 0.01	**16.08**	< 0.01	**34.13**	< 0.01	< 0.01	**49.79**	2
131	< 0.01	< 0.01	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
147	< 0.01	< 0.01	**98.80**	< 0.01	< 0.01	1.20	< 0.01	< 0.01	< 0.01	3
170	< 0.01	< 0.01	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
49	< 0.01	< 0.01	**77.81**	< 0.01	< 0.01	22.19	< 0.01	< 0.01	< 0.01	3
104	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	5.11	< 0.01	< 0.01	**94.88**	4
110	< 0.01	< 0.01	< 0.01	9.57	< 0.01	< 0.01	< 0.01	< 0.01	**90.43**	4
135	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	4.99	< 0.01	< 0.01	**95.01**	4
180	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	18.22	< 0.01	< 0.01	**81.78**	4
182	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	13.15	< 0.01	< 0.01	**86.85**	4
183	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	6.28	< 0.01	< 0.01	**93.72**	4
188	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**100.00**	4
189	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**14.21**	< 0.01	**85.79**	< 0.01	5
192	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2.32	< 0.01	**97.68**	< 0.01	5
200	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	6.16	< 0.01	**93.84**	< 0.01	5
52	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**57.10**	< 0.01	**42.90**	< 0.01	5
65	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**35.71**	< 0.01	**64.29**	< 0.01	5
73	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3.27	< 0.01	**96.73**	< 0.01	5
80	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**12.30**	< 0.01	**87.70**	< 0.01	5
81	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**27.51**	< 0.01	**72.49**	< 0.01	5
82	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**53.01**	< 0.01	**46.99**	< 0.01	5
92	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**37.42**	< 0.01	**62.58**	< 0.01	5
101	< 0.01	**59.40**	< 0.01	**40.60**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	6
127	< 0.01	**95.71**	< 0.01	4.29	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	6
128	< 0.01	**50.15**	< 0.01	**43.66**	< 0.01	6.19	< 0.01	< 0.01	< 0.01	6
144	< 0.01	**92.10**	< 0.01	7.90	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	6
166	< 0.01	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	6
187	< 0.01	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	6
194	< 0.01	**79.01**	< 0.01	**20.99**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	6
199	< 0.01	**81.13**	< 0.01	**18.87**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	6

Continued on next page

Resource	Clust1	Clust2	Clust3	Clust4	Clust5	Clust6	Clust7	Clust8	Clust9	Profile
129	< 0.01	< 0.01	< 0.01	**13.82**	**86.18**	< 0.01	< 0.01	< 0.01	< 0.01	7
148	< 0.01	< 0.01	< 0.01	**19.05**	**76.67**	4.28	< 0.01	< 0.01	< 0.01	7
167	< 0.01	< 0.01	< 0.01	**39.77**	**53.50**	6.73	< 0.01	< 0.01	< 0.01	7
169	< 0.01	< 0.01	< 0.01	**36.61**	**56.40**	6.98	< 0.01	< 0.01	< 0.01	7
171	< 0.01	< 0.01	< 0.01	**23.33**	**76.67**	< 0.01	< 0.01	< 0.01	< 0.01	7
172	< 0.01	< 0.01	< 0.01	3.24	**96.76**	< 0.01	< 0.01	< 0.01	< 0.01	7
184	< 0.01	< 0.01	< 0.01	**20.59**	**48.09**	31.32	< 0.01	< 0.01	< 0.01	7
190	< 0.01	< 0.01	< 0.01	**12.91**	**87.09**	< 0.01	< 0.01	< 0.01	< 0.01	7
191	< 0.01	< 0.01	< 0.01	6.92	**93.08**	< 0.01	< 0.01	< 0.01	< 0.01	7
198	< 0.01	< 0.01	< 0.01	**30.77**	**69.23**	< 0.01	< 0.01	< 0.01	< 0.01	7
59	< 0.01	< 0.01	< 0.01	**38.76**	**50.52**	10.71	< 0.01	< 0.01	< 0.01	7
86	< 0.01	< 0.01	< 0.01	**44.39**	**48.37**	7.23	< 0.01	< 0.01	< 0.01	7
90	< 0.01	< 0.01	< 0.01	**55.78**	**43.39**	0.83	< 0.01	< 0.01	< 0.01	7
107	< 0.01	< 0.01	< 0.01	**32.12**	< 0.01	16.63	**51.26**	< 0.01	< 0.01	8
136	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	**100.00**	< 0.01	< 0.01	8
138	< 0.01	< 0.01	< 0.01	**27.66**	< 0.01	< 0.01	**72.34**	< 0.01	< 0.01	8
142	< 0.01	< 0.01	< 0.01	**16.09**	< 0.01	2.42	**81.49**	< 0.01	< 0.01	8
145	< 0.01	< 0.01	< 0.01	**42.18**	< 0.01	3.45	**54.37**	< 0.01	< 0.01	8
151	< 0.01	< 0.01	< 0.01	**39.62**	< 0.01	2.86	**57.51**	< 0.01	< 0.01	8
154	< 0.01	< 0.01	< 0.01	5.00	< 0.01	< 0.01	**95.00**	< 0.01	< 0.01	8
160	< 0.01	< 0.01	< 0.01	**37.87**	< 0.01	7.71	**54.43**	< 0.01	< 0.01	8
175	< 0.01	< 0.01	< 0.01	**22.36**	< 0.01	1.72	**75.92**	< 0.01	< 0.01	8
185	< 0.01	< 0.01	< 0.01	9.26	< 0.01	16.85	**73.88**	< 0.01	< 0.01	8
195	< 0.01	< 0.01	< 0.01	**19.68**	< 0.01	< 0.01	**80.32**	< 0.01	< 0.01	8
79	< 0.01	< 0.01	< 0.01	**28.59**	< 0.01	7.53	**63.89**	< 0.01	< 0.01	8
98	< 0.01	< 0.01	< 0.01	**28.88**	< 0.01	23.32	**47.80**	< 0.01	< 0.01	8

Table A.11: Probabilities for each resource to belong to a particular multitask session archetype (in %). The resource profile groups are the clustering results after applying AHC.

Resource	Prob1	Prob2	Prob3	Prob4	Prob5	Prob6	Prob7	Prob8	Prob9	Specialisation
131	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
170	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
166	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
188	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
136	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
187	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
87	**100.00**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
147	**98.80**	1.20	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
192	**97.68**	2.32	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
156	**96.95**	3.05	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
172	**96.76**	3.24	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
73	**96.73**	3.27	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
133	**96.57**	2.31	1.12	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
127	**95.71**	4.29	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
135	**95.01**	4.99	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
154	**95.00**	5.00	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
104	**94.88**	5.11	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
200	**93.84**	6.16	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
183	**93.72**	6.28	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
191	**93.08**	6.92	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
118	**92.60**	7.40	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
144	**92.10**	7.90	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
110	**90.43**	9.57	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	1
80	**87.70**	**12.30**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
190	**87.09**	**12.91**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
182	**86.85**	**13.15**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
129	**86.18**	**13.82**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
189	**85.79**	**14.21**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
180	**81.78**	**18.22**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
142	**81.49**	**16.09**	2.42	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
199	**81.13**	**18.87**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
195	**80.32**	**19.68**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
194	**79.01**	**20.99**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
49	**77.81**	**22.19**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
148	**76.67**	**19.05**	4.28	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
171	**76.67**	**23.33**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
175	**75.92**	**22.36**	1.72	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
185	**73.88**	**16.85**	9.26	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
81	**72.49**	**27.51**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
138	**72.34**	**27.66**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
93	**69.97**	**30.03**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
198	**69.23**	**30.77**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
77	**68.29**	**17.29**	**14.42**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
78	**64.15**	**23.35**	**12.50**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
79	**63.89**	**28.59**	7.53	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2
177	**63.15**	**30.23**	6.63	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	2

Continued on next page

Resource	Prob1	Prob2	Prob3	Prob4	Prob5	Prob6	Prob7	Prob8	Prob9	Specialisation
65	**64.29**	**35.71**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
92	**62.58**	**37.42**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
88	**59.77**	**40.23**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
101	**59.40**	**40.60**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
151	**57.51**	**39.62**	2.86	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
159	**57.17**	**42.83**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
52	**57.10**	**42.90**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
169	**56.40**	**36.61**	6.98	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
90	**55.78**	**43.39**	0.83	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
164	**54.70**	**45.30**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
160	**54.43**	**37.87**	7.71	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
145	**54.37**	**42.18**	3.45	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
167	**53.50**	**39.77**	6.73	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
82	**53.01**	**46.99**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
107	**51.26**	**32.12**	**16.63**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
59	**50.52**	**38.76**	**10.71**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
128	**50.15**	**43.66**	6.19	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
121	**50.07**	**49.93**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
96	**49.79**	**34.13**	**16.08**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
86	**48.37**	**44.39**	7.23	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
184	**48.09**	**31.32**	**20.59**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3
98	**47.80**	**28.88**	**23.32**	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	3

Table A.12: Ordered probabilities from largest to smallest for each resource to belong to a particular multitask session archetype (in %). The resource specialisation profile groups are the clustering results after applying AHC.