This article is not an exact copy of the original published article in Methods of Information in Medicine. The definitive publisher-authenticated version of Seiter, J.; Derungs, A.; Schuster-Amft, C.; Amft, O. & Tröster, G.: Daily Life Activity Routine Discovery in Hemiparetic Rehabilitation Patients Using Topic Models. Methods Inf Med, 2015, is available online at:
http://dx.doi.org/10.3414/ME14-01-0082

Daily Life Activity Routine Discovery in Hemiparetic Rehabilitation Patients Using Topic Models

Julia Seiter, Wearable Computing Lab., ETH Zurich, Switzerland
Adrian Derungs, ACTLab, University of Passau, Germany
Corina Schuster-Amft, Research Department, Reha Rheinfelden, Switzerland
Oliver Amft, ACTLab, University of Passau, Germany
Gerhard Tröster, Wearable Computing Lab., ETH Zurich, Switzerland
Correspondence: Julia Seiter, ETH Zurich, Gloriastrasse 35, 8092 Zurich, Switzerland,
Phone: +41 44 632 27 44, Fax: +41 44 632 12 10, E-mail: julia.seiter@ife.ee.ethz.ch

1. SUMMARY

Background: Monitoring natural behavior and activity routines of hemiparetic rehabilitation patients across the day can provide valuable progress information for therapists and patients and contribute to an optimized rehabilitation process. In particular, continuous patient monitoring could add type, frequency and duration of daily life activity routines and hence complement standard clinical scores that are assessed for particular tasks only. Machine learning methods have been applied to infer activity routines from sensor data. However, supervised methods require activity annotations to build recognition models and thus require extensive patient supervision. Discovery methods, including topic models could provide patient routine information and deal with variability in activity and movement performance across patients. Topic models have been used to discover characteristic activity routine patterns of healthy individuals using activity primitives recognized from supervised sensor data. Yet, the applicability of topic models for hemiparetic rehabilitation patients and techniques to derive activity primitives without supervision needs to be addressed.

Objectives: We investigate, (1) whether a topic model-based activity routine discovery framework can infer activity routines of rehabilitation patients from wearable motion sensor data. (2) We compare the performance of our topic model-based activity routine discovery using rule-based and clustering-based activity vocabulary.

Methods: We analyze the activity routine discovery in a dataset recorded with eleven hemiparetic rehabilitation patients during up to ten full recording days per individual in an ambulatory daycare rehabilitation center using wearable motion sensors attached to both wrists and the non-affected thigh. We introduce and compare rule-based and clustering-based activity vocabulary to process statistical and frequency acceleration features to activity words. Activity words were used for activity routine pattern discovery using topic models based on Latent Dirichlet Allocation. Discovered activity routine patterns were then mapped to six categorized activity routines.

Results: Using the rule-based approach, activity routines could be discovered with an average accuracy of 76% across all patients. The rule-based approach outperformed clustering by 10% and showed less confusions for predicted activity routines.

Conclusion: Topic models are suitable to discover daily life activity routines in hemiparetic rehabilitation patients without trained classifiers and activity annotations. Activity routines show characteristic patterns regarding activity primitives including body and extremity postures and movement. A patient-independent rule set can be derived. Including expert knowledge supports successful activity routine discovery over completely data-driven clustering.

Keywords: Rehabilitation; Monitoring, Ambulatory; Activities of Daily Living; Behavior; Signal Processing, Computer-Assisted; Discovery; Topic Model; Wearable Sensors

2. INTRODUCTION

2.1 Scientific Background

Wearable sensors and signal processing methods have been successfully applied in activity and movement analysis of rehabilitation patients including gait and fall risk detection [1; 2], tele-rehabilitation systems [3] and stroke patient monitoring [4; 5; 6]. For example, motor function scores of stroke patients can be derived from sensor measurements of predefined movement tasks [4; 7]. Moreover, wearable accelerometers were used to assess upper-limb activity levels of affected and non-affected extremities [5; 6; 8; 9] to automate classic clinical motor assessments [10; 11; 12]. While functional assessments using wearable sensors can provide quantifications of movement and motor performance, they typically require supervised environments and lack information on the patient's daily activity and lifestyle. Various supervised machine learning methods have been proposed to continuously recognize basic daily activity primitives in healthy individuals, including sleeping, walking, and many others [13; 14; 15; 16; 17]. By contrast, unsupervised activity discovery methods were applied in healthy volunteers to retrieve activity routines that are temporally coarse and recurring, but do not require explicit supervision [18; 19; 20; 21]. Besides specific individual movements and motor assessments, activity routines, such as *lunch*, *kitchen work*, and *rest* can characterize motor capabilities and lifestyle. In particular, activity routine information can provide valuable insights into possibilities and habits of stroke patients, and thus serve as feedback on rehabilitation progress to therapists and patients. For example, changes in duration and frequency of daily activity routines such as *rest*, *kitchen work* and *daily exercise* (e.g. go for a walk) could provide insights on the rehabilitation progress.

Activity routines are often composed of activity primitives that have finer temporal granularity and are suitable for direct recognition from the sensor data. For example, Barger et al. used probabilistic mixture models to infer behavior patterns in daily life from senors event clusters in a smart home [21]. Huynh et al. applied a probabilistic topic model to discover activity routines, such as *lunch* and *office work* from recognized activity primitives and clustered sensor data [18]. Topic models originated from text processing to reveal underlying themes across documents using the documents' word statistics [22]. For activity discovery, themes correspond to activity routines and documents to sensor data slices, from which activity words, i.e. activity primitives are extracted. While discovery techniques aim at deriving activity routines without supervision, the activity primitives were often obtained using supervised recognition. Huynh at al. applied an activity classifier to obtain a defined set of activity primitives as word vocabulary for the topic modeling process, including *walking freely* and *desk activities* [18]. In patient monitoring, however, continuous, potentially patient-specific supervision, as required for activity classifiers is impractical. Yet, Huynh at al. reported that the activity word vocabulary obtained using basic data clustering yielded less accurate activity routine predictions compared to the vocabulary using an activity classifier [18]. Furthermore, activity discovery in hemiparetic stroke patients needs to deal with larger between-subject variability in activity execution than in healthy volunteers, as activities are performed according to individual capabilities. In a pre-study, we investigated activity routine discovery, using a topic model vocabulary based on predefined detection rules for postures and movements in three stroke patients [23]. In this work, we provide a comparison of topic model performance between clustering-based and rule-based vocabularies, as well as including additional hemiparetic patients and activity routines is of interest.

2.2 Study Objectives

We investigated whether topic models can be used in rehabilitation patient monitoring during regular, unscripted patient activities in a rehabilitation day-care center. Our objective was to analyze fully non-supervised routine discovery, hence compare activity word vocabularies that avoid legacy activity classification. We evaluated a rule-based activity vocabulary including body and extremity postures that may be characteristic for a patient's activity routines, and thus could outperform the clustering-based vocabulary. Our rule-based vocabulary was defined prior to applying it with the patients' sensor data, thus is patient-independent.

The objectives of this work are: (1) We introduce an activity routine discovery framework based on topic models to infer activity routines of rehabilitation patients using wearable movement sensor data. (2) We compare the performance of our topic model-based activity routine discovery using rule-based and clustering-based activity vocabulary. We evaluate our approach in a dataset recorded with eleven hemiparetic rehabilitation patients during their stay at a rehabilitation daycare center.

3. METHODS

3.1 Study Participants

We implemented a study to investigate whether activity routines of rehabilitation patients with motor function impairment can be discovered from sensor data using topic models. Inclusion criteria were: hemiparetic rehabilitation patients after stroke or brain tumor extirpation with upper and/or lower extremity motor function impairment including wheelchair users. Participants visited the ambulatory rehabilitation daycare center of Reha Rheinfelden, Switzerland, were older than 18 years, and signed an informed consent form. Exclusion criteria were: further motor function impairment due to additional neurological diseases other than stroke and brain tumor. The Swiss cantonal Ethics committee Aargau approved the study. Eleven (six male, five female) hemiparetic patients between 34-75 years (56 ± 13), eight after stroke and three after brain tumor extirpation were included. Four patients were wheelchair users.

3.2 Study Design

Patients visited the ambulatory daycare center during $2-3$ days per week and were monitored during their presence in the daycare center (up to eight $\mathrm{hours/day}$) for up to ten full days in total, spread over $1-2$ months. To obtain a daily activity reference log as basis for our methods' performance analysis, we categorized patient activities during the day into six activity routines. Categories were established in agreement with therapists before starting the recording study from a preliminary observational analysis and the therapy schedule to reflect the patients' daily life structure. The therapy schedule included *cognitive training*, *medical fitness* and *motor training*. We considered additional activity routines that occurred regularly during therapy breaks including *rest*, *kitchen work* and *eating/leisure*. Eating and leisure contained predominant periods of active interaction with other people and were thus combined to a single activity routine. For patient recordings, the daily activity reference log was extracted from the patient's individual therapy schedule of the day. Activity routines during therapy breaks were manually added to the log by the study examiner, who followed the patient during the recording day. Activities that did not match a routine were assigned to the *null class*. The following activity routines were considered in this work:

—*Eating/leisure*: eating, active interaction with other persons, playing table games.

—*Cognitive training*: cognitive tasks using e.g. a computer, exercise sheets, puzzles.

—*Medical fitness*: intense physical exercises.

—*Kitchen work*: all kind of household and cooking activities.

—*Motor training*: all kind of motor function exercises.

—*Rest*: resting phases in e.g. bed, deck chair.

3.3 Data Acquisition and Measurements

Shimmer3 sensors [24] were used, providing three inertial sensor modalities. Sensor nodes were configured to log the 3-axis acceleration (range: $\pm$ 4g) at a frequency of $50\,\mathrm{Hz}$ to the sensors' internal SD card. In the morning of each recording day, after patients arrived to the daycare center, six Shimmer3 sensors (LxWxH = 51mm x 34mm x 14mm) were attached to wrists, upper arms, and thighs using velcro straps as illustrated in Fig. 1. Sensors were temporarily removed during water and massage therapies. Due to redundancy in the measured data, we only used acceleration signals of three sensors for our analysis: right and left wrist sensors (in Fig. 1: S2, S5) and the non-affected thigh sensor (in Fig. 1: S3 or S6 , selected for each patient individually).

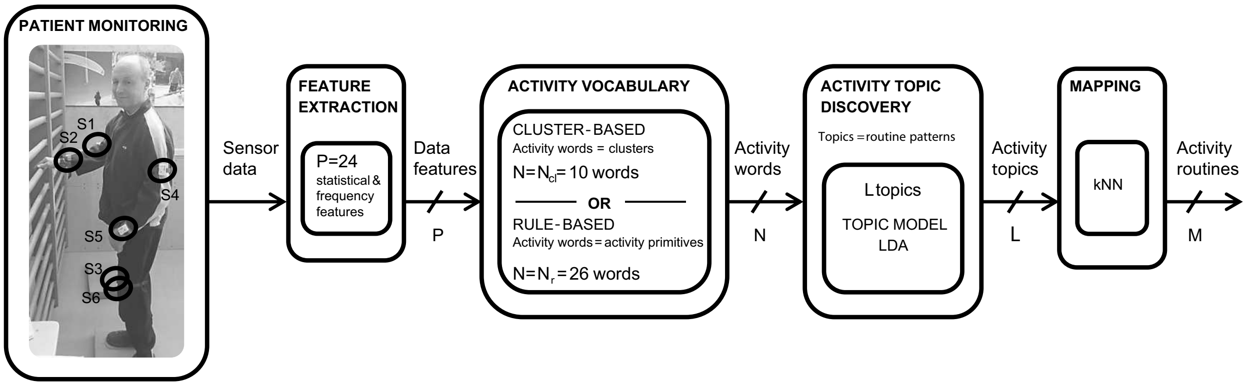

Fig. 1. Sensor setup and activity routine discovery framework considered for unsupervised patient routine discovery and to assess different activity vocabulary derivation methods. Sensor data is processed into P data features. Data features are processed to obtain an activity word vocabulary of size N using clustering-based or rule-based approaches. The topic model discovers L activity topics. For analysis and quantification, the activity topics are mapped to M activity routines using a kNN classifier. Due to redundancy in the measured data, we used acceleration signals of three sensors in our analysis: Sensors S2, S5, and S3 or S6, selected for each patient individually.

3.4 Activity Routine Discovery Framework

We deployed a topic model discovery framework to extract activity routine patterns of the patients from wearable sensor data as illustrated in Fig. 1. Sensor data features are pre-processed into activity words and used as an activity vocabulary for the topic model algorithm. We considered two fully unsupervised approaches to derive an activity vocabulary: (1) Clustering, where activity words form groups (clusters) of similar feature space patterns obtained from sensor data. (2) Rules, where activity primitives aim to provide postures and body motion, including *sit* and *affected arm motion*, that are characteristic of the activity routines. The topic model is then applied to derive activity topics from the activity words. Subsequently, activity topics are mapped back to activity routines using a k-nearest-neighbor (kNN) classifier for our performance analysis.

3.4.1 *Features.* Different time-domain and frequency-domain features were extracted from the sensor data to describe motion and postures. Mean μ and variance σ^2 were derived from acceleration signals for 1s windows for the non-affected thigh and both wrist sensors. Frequency features included power in the low (LF: $0.2 - 2.5$ Hz) and high frequency band (HF: > 2.5 Hz) of the 3-axis acceleration signals (x,y,z) computed for 5s windows. In total, $P = 24$ features were used.

3.4.2 *Clustering-Based Activity Vocabulary.* We used K-means clustering to partition n data points represented by their feature vectors x_j in $K = N_{cl}$ clusters [25]. Clusters were formed by minimizing the Euclidean distance of each data feature vector x_j and the cluster centroid μ_i for all clusters $S_{i=1...N_{cl}} : min \sum_{i=1}^{N_{cl}} \sum_{x_j \in S_i} \|x_j - \mu_i\|$. We analyzed hard and soft clustering. For hard clustering, each data point was assigned to the closest cluster, resulting in a sequential activity word stream. For soft clusters, instead of assigning data feature vectors to a single cluster, cluster weights were calculated for each cluster according to the distance of the data point and each cluster centroid as suggested in [18]. We varied the number of clusters N_{Cl} within the interval $[5, 40]$, thus providing activity word vocabularies of varying size.

3.4.3 *Rule-Based Activity Vocabulary.* Activity primitives were calculated from sensor data features by applying a rule set as detailed in Table 1. For each sensor and feature, thresholds were used to obtain binary features F resulting in, e.g., movement $F = 1$ or no movement $F = 0$. Subsequently, activity primitives were derived by applying logic equations to binary features F, e.g. for a left-sided hemiparetic patient, as illustrated in Table 1: *motion both arms:* affected wrist (in Fig. 1: S5) movement $F_{S5} = 1$ *AND* non-affected wrist (in Fig. 1: S2) movement $F_{S2} = 1$. Every activity primitive was either true or false, while several activity primitives might be true simultaneously, e.g. *affected*

Table 1. Rule-based activity vocabulary with $N_r = 26$ activity primitives. Feature thresholds to obtain binary features $F_{S2,3,5}$ and logic equations for each activity primitive are shown. Statistical time-domain (μ, σ^2) and frequency-domain features (LF, HF bands) of the 3-axis acceleration signals acc_{xyz} of both wrists and non-affected thigh were used. The table lists features for a left-side hemiparetic patient with sensor positions S2, S5 and S3 (see Fig. 1).

Rule-Based Activity Vocabulary	**Logic Equations**
Extremity and body motion	
(1)motion both arms, (2)motion affected arm, (3)motion non-affected arm, (4)no arm motion	(1): $F_{S2} \wedge F_{S5}$ (2): $\overline{F}_{S2} \wedge F_{S5}$, (3): $F_{S2} \wedge \overline{F}_{S5}$, (4): $\overline{F}_{S2} \wedge \overline{F}_{S5}$; $F = 1 : \sigma^2(\|\|acc_{xyz}\|\|) \geq 0.05$
(5)motion non-affected leg, (6)no leg motion	(5): F_{S3}, (6): $\overline{F}_{S3}$; $F = 1 : \sigma^2(\|\|acc_{xyz}\|\|) \geq 0.05$
Peak in LF/HF band: (7)both arms, (8)affected arm , (9)non-affected arm, (10)none	(7): $F_{S2} \wedge F_{S5}$ (8): $\overline{F}_{S2} \wedge F_{S5}$, (9): $F_{S2} \wedge \overline{F}_{S5}$, (10): $\overline{F}_{S2} \wedge \overline{F}_{S5}$; $F = 1 : max(LF, HF) > 5$
Peak LF/HF band: (11)non-affected leg, (12)none	(11): F_{S3} (12): $\overline{F}_{S3}$; $F = 1 : max(LF, HF) > 5$
(13)body motion, (14)no body motion	(13): $F_{S2} \vee F_{S3} \vee F_{S5}$, (14): $\overline{F}_{S2} \wedge \overline{F}_{S3} \wedge \overline{F}_{S5}$; F=1 : $\sigma^2(\|\|acc_{xyz}\|\|)$≥0.05
Extremity and body posture	
(15)stand, (16)sit	(15): F_{S3} (16): $\overline{F}_{S3}$; $F = 1 : \mu(\\|acc_y\\|) > \mu(\\|acc_z\\|)$
Wrist orientation affected arm: (17)horizontal, (18)vertical, (19)-(20)non-affected arm analogue	(17):F_{S5} (18):$\overline{F}_{S5}$ (19):F_{S2} (20):$\overline{F}_{S2}$; $F = 1 : \mu(\\|acc_z\\|) > \mu(\\|acc_x\\|)$
Affected forearm orientation: (21)down, (22)up, (23)horizontal, (24)-(26)non-affected arm analogue	(21): F^1_{S5} (22): F^2_{S5} (23): $\overline{F^1_{S5}} \wedge \overline{F^1_{S5}}$ (24): F^1_{S2} (25): F^2_{S2} (26): $\overline{F^1_{S2}} \wedge \overline{F^1_{S2}}$; F^1=1$\|atan2(acc_y, \|\|acc_{xz}\|\|)$<60°, F^2=1 : $atan2(acc_y, \|\|acc_{xz}\|\|)$≥120°

arm movement and *sit*. Activity primitives were calculated from sensor data streams resulting in sequential activity word streams. We defined a vocabulary of activity primitives that we assumed to be characteristic in representing activity routines based on previous patient observations. Activity primitives were specified to be patient-independent (e.g. *affected arm movement* instead of *right arm movement*). Specifically, we included motion of affected and non-affected extremities, body postures (*sit*, *stand*), wrist and forearm orientation towards the horizontal plane, and extremity movements in low (LF: $0.2 - 2.5$ Hz) and high frequency band (HF: > 2.5 Hz). Binary features and basic logic equations were defined to extract activity primitives from sensor data based on sensor position and the observation of sensor data and the corresponding activity primitives. In total, we considered $N_r = 26$ activity primitives describing body and extremity postures and movements.

3.4.4 *Activity Routine Pattern Discovery.* We used a probabilistic topic model based on Latent Dirichlet Allocation (LDA) [22] that retrieves recurring activity routine patterns from N activity words for each time segment s of a day (e.g. 20 min). In this work, activity words are either clusters of the clustering-based vocabulary ($N = N_{cl}$) or activity primitives of the rule-based vocabulary ($N = N_r$).

Topic models operate based on a number of probabilistic assumptions. Each activity topic z has a fixed probability density function (PDF) (multinomial Mult(β)), defined as the distribution over activity words $w_{1,2,...N}$. The probability of an activity word w depends on $p(w|z, \beta)$. For each time segment s, there is a PDF θ_s over activity topics defined, denoting the probability $p(z|\theta_s)$ of each activity topic z in time segment s. The activity topic distribution θ_s of each segment s is derived from a Dirichlet density distribution $Dir(\alpha)$ with $p(\theta_s|\alpha)$. Thus, the probability of word w in segment s is given by $p(w|s) = p(w|z, \beta) \cdot p(z|\theta_s) \cdot p(\theta_s|\alpha)$. Building activity word histograms for each segment s provides the probabilities $p(w|s)$ entering the topic model. The topic model fits parameters α, β by maximizing the likelihood across all words and segments. Applying the topic model reveals the occurrence ratio γ_s of L activity topics in each segment s. Details on the LDA topic model, parameter fitting, and activity topic inference can be found in [22].

As M activity routines might be composed of several activity topics, typically $L > M$. To map L activity topics and M activity routines, we applied a kNN classifier [26]. The kNN uses γ_s as feature vector. Based on the minimal Euclidean distance in the feature space for each testing sample the activity routine of the closest training sample is assigned. We selected $L = 2M$ topics and a topic model segment size of $DS = 20$min with 90% overlap for our analysis following parameter selection guidelines in [27]. For overlapping segments due to sliding windows, we

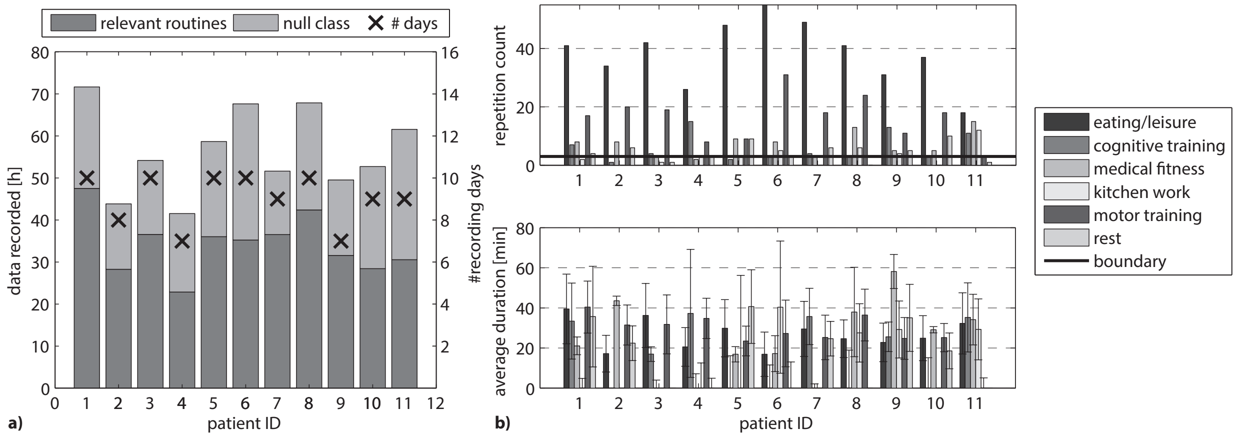

Fig. 2. Dataset statistics. (a) Recording times and number of recording days for each patient including totally 621 h of data with 376 h of relevant activity routine reference logs during 99 days. Recording times variability across patients was due to individual therapy schedules. (b) Top: patient-specific activity routine repetition count from reference log. Bottom: average duration of relevant activity routines per patient from reference log (routines that occurred more than three times).

applied the *Borda Count* ranking method [28]. Details on topic model based activity routine discovery from activity words are available in [27].

3.5 Evaluation Methodology

We evaluated the activity routine discovery procedure in a per-patient leave-one-day-out cross-validation. Activity words (clusters or activity primitives) and activity routine topics were estimated without using supervised information. Activity words were not intended to be evaluated directly as no groundtruth labels were available. Instead, we evaluated activity words by assessing the activity routine discovery performance of the corresponding clustering-based and rule-based approaches. To ensure sufficient data amounts, we considered activity routines that occurred more than three times for a patient. Thus, the number of relevant activity routines M was specific for each patient. Invalid activity routines were assigned to the *null class*. The kNN classifier was trained using training data with relevant activity routines, testing was done on the full dataset (relevant activity routines and *null class*) for the left-out day. As evaluation measure, we used the averaged class-specific accuracy across all relevant activity routines. Due to random initialization of K-means clustering and LDA topic model, we derived mean accuracies across five repeated clustering and five repeated topic model runs.

4. RESULTS

4.1 Patient Activity Routine Dataset

In total, we collected 621 h of data during 99 recording days. Fig. 2 lists recording times, activity routine repetitions and duration according to the obtained reference log. Recording times varied due to the patients' individual therapy schedules and duration of outpatient rehabilitation treatment. E.g., ID4 visited the daycare center during four weeks whereas ID1 for 2.5 months. For all patients 376 h of relevant activity routine reference logs were collected. The type and number of relevant activity routines varied across patients, e.g. ID2 counted four different activity routine types while ID9 counted six. Data amount varied for different activity routines. *Eating/leisure* was the most frequent routine and counted up to 60 repetitions, while, e.g., *medical fitness* occurred up to 15 times and was considered as relevant routine for eight patients only.

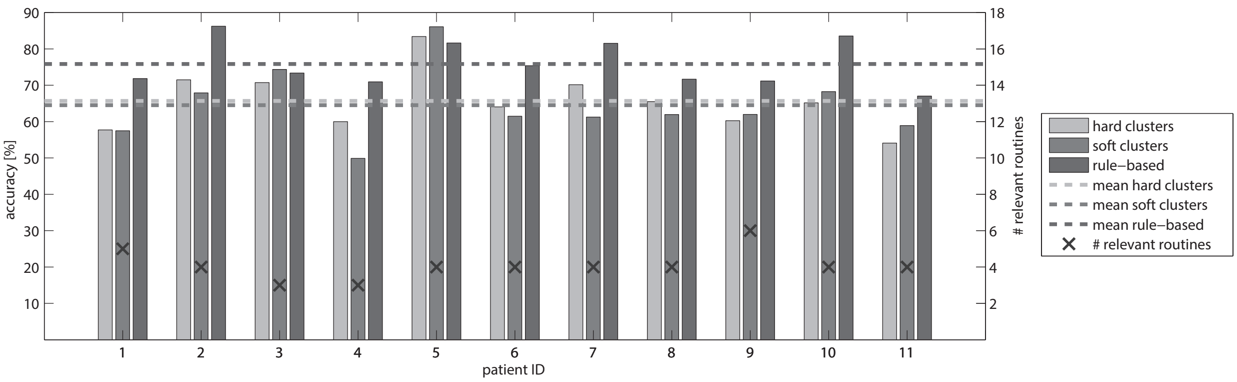

Fig. 3. Accuracies for per-patient activity routine discovery runs using topic models. Activity word vocabularies were obtained using hard and soft clustering ($N_{cl} = 10$) and the rule-based approach ($N_r = 26$). On average, the rule-based approach outperformed the clustering-based approaches by 10%. In addition, relevant routines are indicated per patient ($\times$).

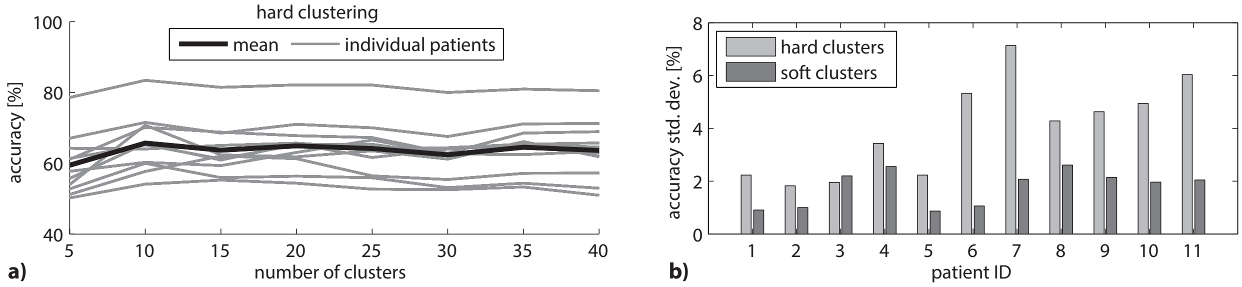

Fig. 4. (a) Per-patient activity routine discovery accuracy for hard clustering with varying activity vocabulary size obtained by varying the number of clusters N_{cl}. Mean accuracy across patients indicate no substantial influence of accuracy on cluster count. (b) Accuracy standard deviation σ across repeated clustering runs (as described in evaluation methodology) was higher for hard clustering compared to soft clustering.

4.2 Activity Routine Discovery

Our rule-based approach yielded accuracies between 67% and 86% for up to six relevant activity routines. Fig. 3 illustrates the per-patient routine discovery performance. The clustering-based approach showed less accurate results and larger accuracy variance across patients with 56%-83%. Averaged across all patients, the rule-based approach outperformed the clustering-based approach by 10% (76% versus 66%). The rule-based activity vocabulary covered $N_r = 26$ activity words (activity primitives). For the clustering-based vocabulary we used $N_{cl} = 10$ activity words (clusters), which provided the best routine discovery performance in the range between $[5, 40]$ clusters, as depicted in Fig. 4 a). Contrary to activity primitives, clusters do not represent a particular movement or posture but are groups of data with similar features. While the per-patient evaluation varied highly for different cluster counts, no substantial influence on mean accuracy was found for larger cluster counts above the minimum number of sufficient clusters ($N_{cl} > 5$). On average, soft-clustering did not improve accuracies yet the standard deviation across repeated clustering runs was lower compared to the hard clustering approach (Fig. 4 b)). Fig. 5 b) shows confusion matrices, indicating that the relevant activity routines were separable using the rule-based approach. For the clustering-based approach, the matrices show more confusions, e.g., for ID1 and ID11 *cognitive training* was mismatched with *eating/leisure* and for ID6 and ID9 *kitchen work* with *motor training*. We observed, that activity routines with high repetition count and long durations, e.g. *eating/leisure* Fig 5b), tended to show higher accuracies compared to short routines with few repetition count (e.g. *kitchen work*, *cognitive training*). Further, the activity routine cognitive training (e.g. ID9) showed decreased performance compared to kitchen work despite similar amount of data. The reasons might be high variability in the routine cognitive training, i.e. once tasks were done at a computer, the next time on paper.

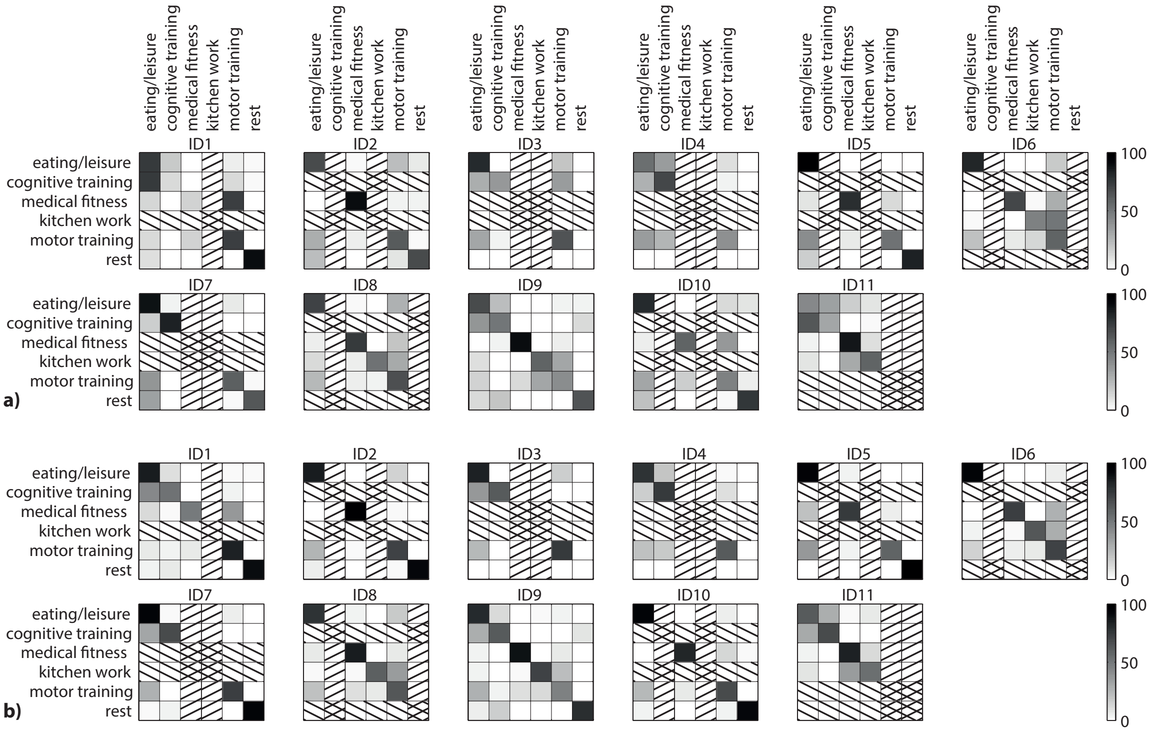

Fig. 5. Confusion matrices showing actual (in rows) and topic model predicted (in columns) activity routines. (a) Hard clustering-based activity vocabulary. (b) Rule-based activity vocabulary. Hatched routines did not reach the occurrence count to become relevant for the particular patient. The rule-based approach showed consistent and better performance across all activity routines compared to the clustering-based approach.

5. DISCUSSION

Our activity discovery approach aimed to provide an activity routine logbook for rehabilitation patients in their daily life. Our approach did not focus on recognizing particular exercises nor inferring functional scores. Instead, we consider activity routine discovery in stroke patients essential to obtain objective statistics regarding type, frequency, and duration of natural and recurring activity routines that a patient performs according to individual habits. Activity routine discovery could provide complementary information to clinical scores, which are assessed on predefined tasks and in a clinical environment only.

Generating activity words from the clustering-based and rule-based vocabularies and discovering activity routines were done without supervision, hence without trained classifiers and activity annotations. In our analysis, the rule-based approach outperformed data-driven clustering by 10%. We interpret this improved performance as an advantage of a rule set to capture characteristic activity patterns, including particular postures and movements, to describe activity routines. Rule sets can be designed without detailed information of the dataset, hence conform to an unsupervised discovery approach, rather than the classification techniques used in previous works.

Activity routines and execution varied from day to day and between patients, according to the patients' individual possibilities and habits. For example *motor training* covered all kind of motor exercises from distinct periodic arm rotations to playing tennis; similarly, wheelchair users occasionally left the wheelchair during motor training. Even though these variations existed in our study data, we found that the patient-independent rule-based activity vocabulary was characteristic for activity routines of all patients. Our routine discovery approach using topic models was applied patient-specific. Given that our routine discovery is an unsupervised method, a patient-specific analysis does not have disadvantages for practical application. Varying accuracies between patients (67%-87%) suggest that our rule-based vocabulary was descriptive for, e.g., ID02 (accuracy: 87%) and less useful for ID11 (accuracy: 67%). The clustering-based vocabulary showed to be less characteristic and discriminative compared to rule-based activity primitives for particular activity routines. While soft-clustering is less sensitive to cluster boundaries due to soft cluster assignments, it could not outperform hard clustering in our evaluation. However, soft-clustering turned out to be more robust across repeated clustering runs. Furthermore, we showed that a setup with three sensors attached to wrists and non-affected thigh is less obtrusive for patients and reduced accuracy by only 2% compared to the configuration with six sensors.

The rehabilitation progress of patients could affect the distributions over the activity vocabulary of activity routines. We monitored patients during 1-2 months and did not observe gradual movement changes that affected the topic model discovery. Considerable movement changes, e.g. a wheelchair user becoming wheelchair independent might indeed modify word distributions for particular routines, e.g. *kitchen work*. Distribution changes could be addressed by re-estimating activity word distributions using the topic model. In future work, the robustness of the topic model for rehabilitation applications should be evaluated in a longitudinal study beyond two months. Moreover, nonparametric topic models [29] that estimate the number of topics automatically from data could be investigated to deal with activity word distribution changes and newly arising activity routines due to rehabilitation progress.

Our analysis showed that discovery performance tended to increase with increasing number of activity routine repetitions and routine duration. However, at the same time less characteristic activity word compositions for an activity routine may result in more discovery errors. Thus, in practice the definition of an activity vocabulary is crucial to select activity words that are discriminative for activity routines. While the rule-based activity vocabulary as a whole turned out to be characteristic for activity routines we did not analyze the characteristic value of single activity primitives per activity routine. The specificity of single activity primitives depends on the dataset [27] and could be investigated in future work using several rehabilitation datasets. Selecting an optimal vocabulary size a priori is challenging. Yet, not the vocabulary size but rather the specificity of activity words for a particular activity routine influences topic model performance. In our investigation, the clustering-based vocabulary and the rule-based vocabulary showed different specificity. Thus, the optimal vocabulary size regarding discovery performance differed between the clustering-based ($N_{cl} = 10$) and rule-based ($N_r = 26$) vocabularies. In general, topic models might be sensitive to overfitting when using a high number of topics M. Selecting the number of topics M close to L activity routines helps preventing overfitting. We used $M = 2L$ topics as suggested in [27].

Topic models were shown to outperform basic clustering of sensor data for activity routine discovery [27]. The rule-based approach in our present work inferred 76% accurate activity routine predictions for 61% of the recorded data. 39% of the recorded data was assigned to the *null class*, hence no performance could be measured. The null class contained short interruptions between activity routines and activities that occurred sparsely (e.g. transitions between therapies, garden work) that could not be assigned to any activity routine. Thus, individual, sparse and short routines were not considered since statistics of these routines could not be derived. Nevertheless, our rule-based approach predicted activity routines with most similar body postures and movements. To robustly infer routines, topic model-based discovery requires that sufficient activity routine repetitions and routine duration (in our analysis more than three repetitions) are available. To monitor physical training and recovery progress, we expect that a patient's regular behavior and lifestyle is primarily relevant, exactly here are topic models particularly useful.

In this study, we monitored eleven ambulatory rehabilitation patients in a daycare center, where patients followed an individual therapy schedule combined with time spent on their own during therapy breaks or scheduled self-determined activity. The daycare scenario provided us with access to a comprehensive activity routine reference, necessary to evaluate and compare discovery performance. We acknowledge that habits and activity routines in the daycare center may differ from typical activity routines at home. However, we expect that the rule-based vocabulary is transferable to routines in the home environment. In the future, rule sets could be adjusted to fit particular patient groups or routine analysis needs, prior to any data recordings. Our results indicate that activity routine discovery using topic models and the rule-set activity vocabulary could be adapted for a home monitoring situation in the future.

6. CONCLUSION

We analyzed whether a set of six daily life activity routines of hemiparetic rehabilitation patients can be discovered from wearable sensors using topic models. Our results indicate that topic models are suitable for patient monitoring as activity routine patterns were successfully discovered from sensor data without the need of trained classifiers and supervised pattern models. Our rule-based approach to derive an activity word vocabulary outperformed a clustering-based baseline by 10% and yielded 76% accuracy on average across patients. We concluded that activity routines show characteristic patterns that can be captured in activity primitives, including body and extremity postures and movement. A generic, patient-independent detection rule set was adequate for successful activity routine discovery.

7. ACKNOWLEDGMENTS

We are grateful to the participating patients and supporting therapists from the daycare center of Reha Rheinfelden and their photograph provision. This work was supported by EU Marie Curie Network iCareNet, grant number 264738.

REFERENCES

[1] Gietzelt M, Wolf K, Kohlmann M, Marschollek M, Haux R. Measurement of Accelerometry-based Gait Parameters in People with and without Dementia in the Field. A Technical Feasibility Study. Methods Inf Med. 2013;52(4):319–325.

[2] Marschollek M, Rehwald A, Wolf K, Gietzelt M, Nemitz G, Meyer Zu Schwabedissen H, et al. Sensor-based fall risk assessment-an expert'to go'. Methods Inf Med. 2011;50(5):420–426.

[3] Antón D, Goñi A, Illarramendi A. Exercise Recognition for Kinect-based Telerehabilitation. Methods Inf Med. 2014;53.

[4] Del Din S, Patel S, Cobelli C, Bonato P. Estimating fugl-meyer clinical scores in stroke survivors using wearable sensors. In: EMBC. IEEE; 2011. p. 5839–5842.

[5] Haeuber E, Shaughnessy M, Forrester LW, Coleman KL, Macko RF. Accelerometer monitoring of home-and community-based ambulatory activity after stroke. Arch Phys Med Rehab. 2004;85(12):1997–2001.

[6] Uswatte G, Foo WL, Olmstead H, Lopez K, Holand A, Simms LB. Ambulatory monitoring of arm movement using accelerometry: an objective measure of upper-extremity rehabilitation in persons with chronic stroke. Arch Phys Med Rehab. 2005;86(7):1498–1501.

[7] Wade E, Parnandi AR, Mataric MJ. Automated administration of the wolf motor function test for post-stroke assessment. In: PervasiveHealth. IEEE; 2010. p. 1–7.

[8] van der Pas SC, Verbunt JA, Breukelaar DE, van Woerden R, Seelen HA. Assessment of arm activity using triaxial accelerometry in patients with a stroke. Arch Phys Med Rehab. 2011;92(9):1437–1442.

[9] de Niet M, Bussmann JB, Ribbers GM, Stam HJ. The stroke upper-limb activity monitor: its sensitivity to measure hemiplegic upper-limb activity during daily life. Arch Phys Med Rehab. 2007;88(9):1121–1126.

[10] Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance. Scand J Rehabil Med. 1974;7(1):13–31.

[11] Gowland et al C. Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. Stroke. 1993;24:58–63.

[12] Taub E, Miller N, Novack T, Cook 3rd E, Fleming W, Nepomuceno C, et al. Technique to improve chronic motor deficit after stroke. Arch Phys Med Rehab. 1993;74(4):347–354.

[13] Bao L, Intille SS. Activity recognition from user-annotated acceleration data. In: Pervasive computing. Springer; 2004. p. 1–17.

[14] Blanke U, Schiele B. Daily routine recognition through activity spotting. In: Location and Context Awareness. Springer; 2009. p. 192–206.

[15] Tapia EM, Intille SS, Larson K. Activity recognition in the home using simple and ubiquitous sensors. In: Proceedings of Pervasive. Springer; 2004. p. 158–175.

[16] Mannini A, Sabatini AM. Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors. 2010;10(2):1154–1175.

[17] Parkka J, Ermes M, Korpipaa P, Mantyjarvi J, Peltola J, Korhonen I. Activity classification using realistic data from wearable sensors. Information Technology in Biomedicine, IEEE Transactions on. 2006;10(1):119–128.

[18] Huynh T, Fritz M, Schiele B. Discovery of activity patterns using topic models. In: UbiComp. ACM; 2008. p. 10–19.

[19] Farrahi K, Gatica-Perez D. Discovering routines from large-scale human locations using probabilistic topic models. ACM Trans Intell Syst Technol. 2011;2(1):1–27.

[20] Begole JB, Tang JC, Hill R. Rhythm modeling, visualizations and applications. In: UIST. ACM; 2003. p. 11–20.

[21] Barger et al TS. Health-status monitoring through analysis of behavioral patterns. Systems, Man and Cybernetics, IEEE Transactions on. 2005;35(1):22–27.

[22] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. JMLR. 2003;3:993–1022.

[23] Seiter J, Derungs A, Schuster-Amft C, Amft O, Tröster G. Activity Routine Discovery in Stroke Rehabilitation Patients Without Data Annotation. In: PervasiveHealth; 2014. p. 270–273.

[24] Burns A, Greene BR, McGrath MJ, OShea TJ, Kuris B, Ayer SM, et al. SHIMMER–A wireless sensor platform for noninvasive biomedical research. Sensors Journal, IEEE. 2010;10(9):1527–1534.

[25] Selim SZ, Ismail MA. K-means-type algorithms: a generalized convergence theorem and characterization of local optimality. TPAMI. 1984;(1):81–87.

[26] Cunningham P, Delany SJ. k-Nearest neighbour classifiers. Multiple Classifier Systems. 2007;p. 1–17.

[27] Seiter J, Amft O, Rossi M, Tröster G. Discovery of activity composites using topic models: An analysis of unsupervised methods. Pervasive Mob Comput. 2014;15:215 – 227.

[28] Ho TK, Hull JJ, Srihari SN. Decision combination in multiple classifier systems. TPAMI. 1994;16(1):66–75.

[29] Teh YW, Jordan MI, Beal MJ, Blei DM. Hierarchical dirichlet processes. Journal of the american statistical association. 2006;101:1566–1581.