Identifying Physical Activity Profiles in COPD Patients Using Topic Models

Gabriele Spina, Pierluigi Casale, Paul S. Albert, Jennifer Alison, Judith Garcia-Aymerich, Richard W. Costello, Nidia A. Hernandes, Arnoldus J. R. van Gestel, Jörg D. Leuppi, Rafael Mesquita, Sally J. Singh, Frank W. J. M. Smeenk, Ruth Tal-Singer, Emiel F. M. Wouters, Martijn A. Spruit, and Albertus C. den Brinker

***Abstract*—With the growing amount of physical activity (PA) measures, the need for methods and algorithms that automatically analyze and interpret unannotated data increases. In this paper, PA is seen as a combination of multimodal constructs that can cooccur in different ways and proportions during the day. The design of a methodology able to integrate and analyze them is discussed, and its operation is illustrated by applying it to a dataset comprising data from COPD patients and healthy subjects acquired in daily life. The method encompasses different stages. The first stage is a completely automated method of labeling low-level multimodal PA measures. The information contained in the PA labels are further structured using topic modeling techniques, a machine learning method from the text processing community. The topic modeling discovers the main themes that pervade a large set of data. In our case, topic models discover PA routines that are active in the assessed days of the subjects under study. Applying the designed algorithm to our data provides new learnings and insights. As expected, the algorithm discovers that PA routines for COPD patients and healthy subjects are substantially different regarding their composition and moments in time in which transitions occur. Furthermore, it shows consistent trends relating to disease severity as measured by standard clinical practice.**

***Index Terms*—Activity routine, chronic obstructive pulmonary disease (COPD), topic models.**

Manuscript received December 29, 2014; revised March 20, 2015; accepted May 8, 2015. Date of publication May 12, 2015; date of current version September 1, 2015. This work was supported by the iCare4COPD Project of Agentschap NL under Contract PNE101005.

G. Spina is with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands, and also with the Smart Professional Spaces Group, Philips Research, 5656 AE Eindhoven, The Netherlands (e-mail: g.spina@tue.nl).

P. Casale is with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands, and also with the Holst Centre/IMEC, 5656 AE Eindhoven, The Netherlands (e-mail: p.casale@tue.nl).

P. S. Albert is with the School of Ageing and Chronic Disease, University Hospital Aintree, Liverpool L9 7AL, U.K. (e-mail: paul.albert@aintree.nhs.uk).

J. Alison is with the Clinical and Rehabilitation Sciences, The University of Sydney, Syndey, N.S.W. 2006, Australia, and also with the Physiotherapy Department, Royal Prince Alfred Hospital, Sydney, N.S.W 2050, Australia (e-mail: jennifer.alison@sydney.edu.au).

J. Garcia-Aymerich is with the Centre for Research in Environmental Epidemiology, 08003 Barcelona, Spain, the CIBER Epidemiologia y Salud Publica, 08036 Barcelona, Spain, and also with the Universitat Pompeu Fabra, 08002 Barcelona, Spain (e-mail: jgarcia@creal.cat).

R. W. Costello is with the Department of Respiratory Medicine, Beaumont Hospital, Dublin, Ireland (e-mail: rcostello@rcsi.ie).

N. A. Hernandes is with the Laboratory of Research in Respiratory Physiotherapy, Department of Physiotherapy, State University of Londrina, Londrina 86057-970, Brazil (e-mail: nyhernandes@gmail.com).

A. J. R. van Gestel is with the Pulmonary Division, University Hospital of Zurich, 8091 Zurich, Switzerland (e-mail: vrns@zhaw.ch).

J. D. Leuppi is with the Medical University Clinic, Cantonal Hospital Baselland, Liestal and Medical Faculty, University of Basel, 4003 Basel, Switzerland (e-mail: Joerg.Leuppi@ksbl.ch).

R. Mesquita and E. F. M. Wouters are with the Department of Research and Education, CIRO+, 6085 NM Horn, The Netherlands, and with the Department of Respiratory Medicine, Maastricht University Medical Center, 6211 LK Maastricht, The Netherlands (e-mail: rafaelmesquita@ciro-horn.nl; e.wouters@mumc.nl).

S. J. Singh is with the NIHR EM CLAHRC—Centre for Exercise and Rehabilitation Science, University Hospitals, Leicester LE3 9QP, U.K. (e-mail: sally.singh@uhl-tr.nhs.uk).

F. W. J. M. Smeenk is with the Department of Respiratory Medicine, Catharina Hospital, 5623 EJ Eindhoven, The Netherlands (e-mail: frank.smeenk@catharinaziekenhuis.nl).

R. Tal-Singer is with GlaxoSmithKline R&D, King of Prussia, PA 19406 USA (e-mail: Ruth.M.Tal-Singer@gsk.com).

M. A. Spruit is with the Department of Research and Education, CIRO+, 6085 NM Horn, The Netherlands, and also with the REVAL/BIOMED, Hasselt University, 3900 Diepenbeek, Belgium (e-mail: martijnspruit@ciro-horn.n).

A. C. den Brinker is with the Smart Professional Spaces Group, Philips Research, 5656 AE Eindhoven, The Netherlands (e-mail: bert.den.brinker@philips.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2015.2432033

I. Introduction

THE prevalence of chronic diseases, in general, is rising due to an ageing population, as well as due to environmental and lifestyle changes. This is particularly true for respiratory diseases, such as chronic obstructive pulmonary disease (COPD), which is a progressive and irreversible disease that results in airflow limitation and significant extra pulmonary effects, which limit physical activities [1]. Physical activity (PA), defined as any bodily movement produced by skeletal muscles that requires energy expenditure (EE) [2], is known to be a relevant indicator of COPD patients' health status. Current COPD treatment guidelines strongly recommend pulmonary rehabilitation programs with the intent of increasing PA and maintaining the changes over time, including when patients are discharged to home [3]. Although learning to exercise and maintaining exercising is really difficult, this is possibly less of an issue for other effective behavioral changes like reducing the sedentary time.

New methodologies assessing PA and a better insight into daily PAs of patients with COPD are needed to accurately characterize the disease and assist clinicians in designing targeted therapeutic strategies and personalized coaching programs that do not rely on *a priori* fixed objectives [4]. Since patients are not able to accurately self-report their physical activities [5], models describing PA should be learned in an unsupervised fashion without the need of user annotations. Moreover, they should be suitable to harmoniously integrate multiple low-level PA measures (such as intensity levels of the activities performed or step counts), and the associated physiological responses into meaningful descriptors [6]. With recent improvements in wearable sensor technologies, it becomes easier in daily life to acquire massive amounts of different sensors data. At the same time, however, it is difficult to combine and extrapolate meaningful information in absence of any supervision or annotation.

In this paper, measures of PA and physiological responses are presented as the letters composing the words that describe PAs. The cooccurrence of these words in different ways and proportions during the day creates groups of PA constructs describing the main topics that pervade the day of a patient, i.e., PA routines. Routines could also occur at different times and in different proportions across the day for different patients or subgroups of patients, thus characterizing their activity behavior. In this work, we show that the low-level fusion of sensors data into "words" combined with topic modeling holds the promise of discovering new and clinically relevant insights. In particular, this paper provides the following contributions.

(1) We propose a methodology to create a vocabulary of meaningful words describing PA from a set of multimodal PA measures without the need of any supervision or parameter tuning. Using the generated vocabulary, we discover PA routines that pervade daily life of healthy control subjects and a subgroup of COPD patients pairwise matched for age, gender, and body mass index (BMI).

(2) We infer the underlying PA routine structure of the dataset available (comprising more COPD patients) on day segments of 30 min and on the first 6 h at once describing, respectively, the temporal regularities of the multimodal PA measures, and the estimated time spent by the subjects under study in each of the routines.

(3) Finally, we investigate for the first time whether discovered PA routines are related to disease severity, whether they can be deployed to cluster subjects of the matched dataset and, furthermore, to recognize from which group each assessed day comes from.

II. Related Works

A. COPD Ambulatory Monitoring

The automatic monitoring and analysis of chronic diseases has always been central in research on wearable sensors. Physical activities in patients with moderate to very severe COPD has been objectively documented using motion sensors, which provide more accurate, individualized, and detailed information on body movement than questionnaires [7].

Liao *et al.* [8] provided a review focused on describing current wearable technologies for measuring the PA level of COPD patients. In particular, validated devices should be used, since activity monitors can be less accurate in patients whose walking speeds is as low as 0.5 miles/h.

The use of data from ten wearable sensors as indicator of PA levels and physiological responses to recognize a set of 11 scripted activities is illustrated in [9]. In particular, a comparative study of supervised machine learning techniques is presented with emphasis placed on achieving high recognition accuracy. The number of sensors and the complexity of the algorithms are second to that objective. Research on PA levels of COPD patients usually deploys one single activity monitor to minimize obstructiveness.

B. COPD Physical Activities

It has been consistently shown that COPD patients have lower PA levels than their healthy peers [10], and that reduced levels of PA are related to an increased risk of hospital admission and mortality [11]. Outcome variables related to this type of analysis are focused mainly on amount and volume parameters, such as number of steps, walking time, volume of PA as expressed by total number of counts, and total EE. Although these are important health markers of patients suffering from COPD, interventions thus far have failed to demonstrate important increases in these outcome variables. Moreover, the relation with severity of the disease, assessed by the forced expiratory volume in the first second (FEV_1), is not strong or not significant [12].

Bouts of PA described by their frequency, duration, and intensity were introduced by Gonzalez *et al.* [4] in order to explore whether these patients meet the general guidelines for PA for older adults. Information about physiological responses were not considered in the analysis.

C. Symbolic Representation of Data

Alternatively to the selection of statistical attributes (such as mean duration of activity bouts within a day or the time spent in activity bouts of different intensities) which are mainly driven by intuition and experience, other approaches can be used to represent and analyze patient's data. Symbolic representation of continuous data offers several advantages, such as the possibility to be used in combination of a wide set of algorithms from the text processing community. Symbolic approaches like SAX [13] have been proposed to reduce efficiently a time series to a set of symbols of a vocabulary. In addition to the *a priori* fixed size of the vocabulary, this method assumes a particular distribution, which may not be always valid and may limit the performance of series mining tasks [14], [15].

In order to construct features for discriminative tasks, Saria *et al.* [16] proposed a method that discovers in a time-series recurring subsequences having similar shape, but that, at the same time, can exhibit significant variability (deformable motifs). Number and length of the motifs need to be specified beforehand and, if applied to multidimensional time series, the method assumes that motifs along all the dimensions are happening synchronously.

D. Probabilistic Unsupervised Modeling

A probabilistic approach for unsupervised mining of electronic health data has been introduced by Schulam *et al.* [17]. Time series of clinical markers were clustered taking into account confounding factors that might affect data. In line with the generative modeling of multidimensional time series outlined by the authors and increasing the abstraction level of the activity bouts, we might think of these constructs as primitive descriptors of PA that are latently coupled to each other. Organizing bouts in higher level and coherent structures that cooccur during a day, more informative PA constructs could be described (i.e., PA routines). Topic models suit this goal since they represent a class of algorithms able to discover hidden structures in collections of documents. Due to their pattern discovery capability, they have been explored in the wearable sensor and activity recognition community.

Huynh *et al.* [18] showed that daily routines of activities can be recognized as a probabilistic combination of activity labels, such as walking, discussing at whiteboard, etc. They also addressed the point of avoiding supervised learning approaches by clustering raw sensor data in order to build the vocabulary of activity primitives. Their approach was tested only on data from one single subject, and the vocabulary of primitives had a fixed size chosen *a priori*.

Seiter *et al.* [19] compared three topic model approaches and analyzed three public datasets with different properties affecting the discovery, such as primitive rate, activity composite specificity, primitive sequence similarity, and composite-instance ratio. These authors compared the activity composite discovery performance against the performance of a k-means clustering algorithm providing guidelines for optimal parameter selection. Their results indicated that latent Dirichlet allocation (LDA) shows higher robustness against noise compared to k-means and other topic modeling approaches.

The application of a nonparametric framework to create the vocabulary and discover human routines from sensor data was investigated using a Dirichlet process Gaussian mixture model (DPGMM) in [20]. Although this approach does not need to specify the number of unique artificial words, it assumes that data should come from the same distributions used to create the mixture model. In the case of populations with different activity behavior as healthy subjects and COPD patients at different stages, this might not be true and the mapping of the raw sensor data to vocabulary words might not be correct.

III. Background

Topic models are algorithms for discovering the main themes that pervade a large and unstructured collection of documents. Data are treated as observations arising from a generative probabilistic process, in which hidden variables reflect the thematic structure of a collection of documents. The intuition behind using the LDA [21] to discover PA routines is that each day is a mixture of thematically coherent PA measures just as a text document is a mixture of thematically coherent words. The graphical model for LDA is provided in Fig. 1. All the assessed days (also called documents $d_{1:D}$) share the same set of daily routines (also called topics $\beta_{1:K}$) that are defined as Dirichlet distributions over the observed set of PA descriptors (also called words W or terms of a fixed vocabulary).

Observing activities in patients is a difficult task since it is time intensive and intrusive. At the same time, patients are not able to accurately self-report their physical activities [5], and the training of a classifier requires annotations that in a daily life scenario are difficult to obtain. In order to make the methodology fully unsupervised, we assume that the observed words (input of the model) are composed by multimodal measures coming from a body area sensor network. Each assessed day exhibits PA daily routines in different proportion indicated as $\theta_{1:D}$, i.e., each day has a different distribution over the routines that also follows a Dirichlet distribution. In such a model, the N words ($W_{d,n_{1:N}}$) that compose the D documents are the only random variables observed and depend on the per word topic assignment ($Z_{d,n}$) and all the β_k. The daily routines then are composed indirectly by low-level PA measures that belong, with a certain probability distribution, to different thematic areas. Different routines will have different PA measures with different probabilities.

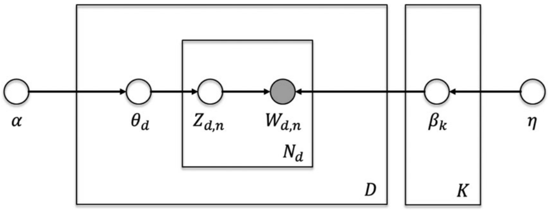

Fig. 1. Graphical model for LDA. Each node is a random variable, edges denote possible dependences. The only observed variables (shaded) are the words (W). The distribution of the words in a routine (β) and the distribution of the routine in a document (θ) depend only on the topic hyperparameters η and α that control the mean shape and sparsity of the distributions. Z represents the word topic assignment.

Topic modeling algorithms calculate the hidden structure that likely generated the observed collection of documents. In particular, for this analysis, we use variational inference to approximate the intractable posterior distribution over hidden variables defined by LDA. In a nutshell, variational inference posits a parametrized family of distributions over the hidden structure, and then, finds the member of that family that is closest to the posterior according to the *Kullback–Leibler* divergence. The reader is invited to refer to [21] for an exhaustive explanation.

IV. Methods

In the application of LDA, a word, defined to be an instance of a vocabulary, is considered as the basic unit of discrete data. In this paper, multimodal monitoring signals, composed of PA measures and physiological responses, are symbolized first as letters. Our approach in selecting the letters and then the words composing the vocabulary benefits from a methodology that preserves the interpretability of the vocabulary and that allows the generation of words that could also not appear in the current documents. The multidimensional data space is first divided in subspaces, and for each subspace, we use cluster analysis to divide PA measures and physiological responses into discrete clusters levels (letters) that are then combined to form words. We use soft assignments to link an intensity bout (IB) with the words in the vocabulary such that any bout has a probability to belong to each of the terms created. We extract the vocabulary in a subset of COPD and healthy patients and show that the vocabulary chosen can fit the model on a larger cohort of patients. The dataset and the methodology developed are described in detail in the following sections.

TABLE I
SUBJECT GROUP CHARACTERISTICS

	All COPD $n = 977$	Matching healthy $n = 66$	Matching COPD $n = 66$
Male/Female (n)	636/341	30/36	30/36
Age (years)	66 [61–72]	65 [61–70]	65 [61–70]
BMI (Kg/m^2)	25.8 [22.5–29.4]	25.2 [23–27.3]	25 [22.5–27.8]
FEV$_1$ (% predicted)	49 [34–64]	107 [97–117]	42 [29–63]
GOLD 1–2–3–4 (n)	89–385–330–173	–	8–16–23–19
Assessed days (n)	5846	404	395
Weekend days (%)	33	32.6	33.6
Weekdays (%)	67	67.4	66.4
Days per subject	6 [6–6]	6 [6–6]	6 [6–6]
Minutes per subject	982 [913–1060]	965 [918–1014]	972 [919–1035]

Data are summarized as absolute frequency (n), relative frequency (%), or median and quartiles [Q1–Q3].

A. Dataset

Data from 1001 patients suffering from mild to very severe COPD were collected across ten countries (United Kingdom, Ireland, The Netherlands, Germany, Switzerland, Italy, Spain, The United States of America, Brazil, and Australia) as part of previous studies (references are omitted for the sake of brevity) without overlaps with the current post hoc analysis. Subjects were included if they met the following inclusion criteria: COPD with a postbronchodilator forced expiratory volume in the first 1 s (FEV$_1$) / forced vital capacity (FVC) ratio < 0.70 and stable condition (i.e., no symptoms of increased shortness of breath and sputum production compared to usual). The dataset comprises only baseline data, which means that the COPD patients were not undergoing any specific intervention by the time of the assessment. Centers from The Netherlands and UK also provided data on 66 healthy control subjects that were matched for age, gender, and BMI with a subgroup of 66 COPD patients. On the basis of a 1:1 multivariate matching, the closest possible *case:control* matches were determined. Subjects matched exactly for age and gender, the median error between BMI values of matching subjects was 0.58 [0.29–1.2] Kg/m^2. Subject group characteristics are presented in Table I.

PAs of COPD and healthy subjects were assessed during daily life by mean of the SenseWear Armband and SenseWear Mini Armband activity monitors [22]. These devices combine an accelerometer with different physiological sensors: a heat flux sensor, a galvanic skin response (GSR) sensor, a skin temperature (ST) sensor, and a near-body ambient temperature sensor. Data are sampled in 1-min intervals and together with demographic characteristics (such as gender, age, height, and weight) were used to estimate EE and metabolic equivalent of task (MET) using proprietary algorithms developed by the manufacturer. The use of multisensory data in combination with pattern recognition algorithms ensures that the MET estimation is insensitive to noise and random motion artifacts [23]. For each minute, the associated steps count (SC) and information about the sleeping status of a subject ($0 =$ awake, $1 =$ sleeping) are also provided by the sensor. The SenseWear Armband has been shown to be valid both in field [24] and in laboratory studies [25].

COPD patients and healthy subjects wore the sensor both during daytime and nighttime so that continuous nonscripted activities were recorded in a natural environment. A minimum of four days (two weekdays + Saturday + Sunday) was considered acceptable to include a subject in the analysis [26], with the device being used for at least 22 h/day. From the minutes coded by the activity monitor as "sleeping," the longest period of night sleep was extracted, and the awakening point defined as the time instant after such period. Brief awake periods (<10 min) of very light intensity (MET < 2.0) within time intervals coded as "sleeping" longer than 2 h were considered part of the sleeping time. Recorded days were synchronized according to this point in order to minimize the intrinsic variability of the data. Data prior to the awakening point were discarded from the analysis. Subjects with at least 12 h of data after the awakening point were included for a total of 977 COPD patients and 66 healthy controls. The median number of days analyzed per patient was 6 (four weekdays, two weekend days), resulting in a total of 5846 valid PA days assessed, of which 3916 (67%) were weekdays and 1930 (33%) weekend days. In median, 982 min were analyzed per patient each day (992 weekdays, 961 weekend days). Ethics Board approval was obtained from the local ethics committees, and written informed consent was provided by participants.

B. Vocabulary

One of the most important choices one has to make when applying LDA to activity data is the nature and number of terms forming the vocabulary. In order to limit the heuristics, we used a data-driven methodology to automatically create the vocabulary without specifying its size beforehand. Sensor data from the 66 healthy subjects and the matched COPD patients subsample were used to automatically create the vocabulary of words without specifying its size beforehand. Each 1-min data point consists of a 7-D measurement vector comprising: *MET*, *ST*, *GSR*, *Longitudinal Acceleration (Acc$_L$)*, *Transversal Acceleration (Acc$_T$)*, *SC,* and *Sleeping Status (SL)*. Acc_L and Acc_T were combined to compute vector magnitude (VM). METs data were first divided into activity intensity categories (IC) using the thresholds proposed by the American College of Sports Medicine [27]: very light intensity (IC$_{\text{VL}}$), < 2.0 METs; light intensity (IC$_L$), 2.0 to 2.9 METs; and moderate-to-vigorous intensity (IC$_{\text{MV}}$), ≥ 3.0 METs. Minutes marked by the sensor as sleeping and with METs < 2.0 formed a separated category named sleeping (IC$_S$). For simplicity, we will refer to the IC only with subscripts S, VL, L, MV. Figure 2 shows an example of a one day METs data stream with the respective IC superimposed.

Consecutive minutes exhibiting the same ICs are then grouped together in IB of variable duration (d). In each IB, we calculated the mean (μ) of ST, GSR, VM, and SC. The original sensor data stream is then represented by a series of IBs, where each bout is fully characterized by a six-elements feature vector $\tilde{V}$

$$\tilde{V} = [\text{IC}, d, \mu_{\text{ST}}, \mu_{\text{GSR}}, \mu_{\text{VM}}, \mu_{\text{SC}}]. \tag{1}$$

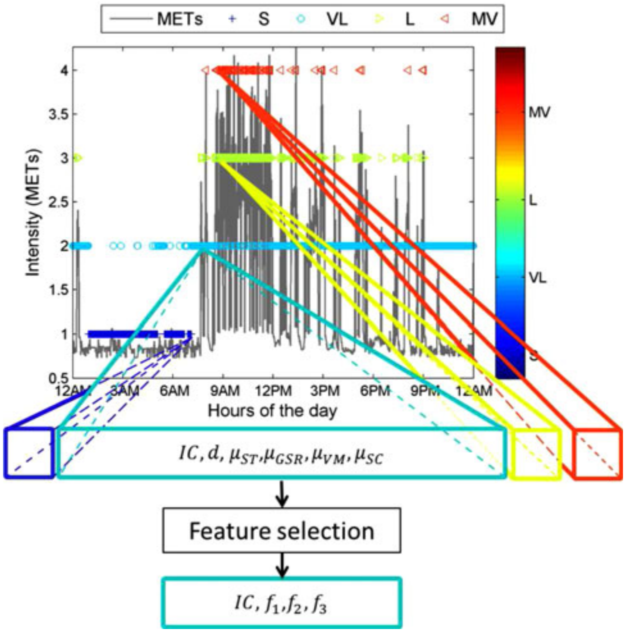

Fig. 2. Continuous data stream representing METs value (gray line) is first quantized in IC (blue is sleeping, ciano is very light, green is light, and red is moderate to vigorous intensity). IC of the same intensity form bouts. Features are computed within the bouts, and the most relevant for each intensity, according to the feature selection procedure, will be combined to create PA descriptors.

Subsequently, for each intensity category (S, VL, L, and VL), the most relevant subset of features was selected such that the multicluster structure of the data can be best preserved. Features were selected using the Multi-Cluster Feature Selection (MCFS) method that deploys spectral regression with l_1-norm regularization in order to select features jointly instead of evaluating each feature independently [28]. The feature vector $\tilde{V}$ can then be simplified according to

$$V = [\text{IC}, f_1, f_2, f_3]. \tag{2}$$

The features not selected by the feature selection algorithm can be interpreted as wildcard characters representing multiple terms in the vector space ($[\text{IC}, f_1, f_2, f_3, f_4]$ and $[\text{IC}, f_1, f_2, f_3, f_4, f_5]$ are two such terms of $[\text{IC}, f_1, f_2, f_3]$). As intuitively explained later in this section, wildcarding and the removal of frequent words help in generating sparse topics [21]. The selected features $f_{j\in\{1,2,3\}}$ obviously might be different for each intensity.

To generate the vocabulary of words, each of the selected features was first standardized, and then, mapped into a set of discrete levels (which could be interpreted as the letters of our words) using a k-means clustering algorithm. The algorithm automatically selects for each feature the number of levels K^{f_j} in a way that the resulting clustering levels $L_{p\in\{1:K\}}$ are the most stable under small perturbations of the input dataset [29]. Levels are sorted according to their mean value L_p in ascending order such that the first level (L_1) represents clusters with the smallest feature values, and the last level (L_K) represents clusters with the highest feature values. Features selected and levels are shown in Table II. Mean value and variance of the levels were stored and used to create the documents as described in Section IV-D.

TABLE II
FEATURES SELECTED FOR EACH INTENSITY AND ASSOCIATED LEVELS

Sleeping			Very light		
d	μ_{GSR}	μ_{VM}	d	μ_{ST}	μ_{VM}
L_1 L_2	L_1 L_2	L_1 L_2 L_3	L_1 L_2	L_1 L_2	L_1 L_2
Light			Moderate to vigorous		
d	μ_{ST}	μ_{GSR}	d	μ_{ST}	μ_{VM}
L_1 L_2	L_1 L_2 L_3 L_4	L_1 L_2	L_1 L_2	L_1 L_2 L_3	L_1 L_2

The vocabulary of terms was built by allowing all the possible combinations between levels sharing the same IC. For the sleeping category, for example, the feature $f_1 = d$, $f_2 = \mu_{\text{GSR}}$, and $f_3 = \mu_{VM}$ were selected and divided in $K^{f_1} = K^{f_2} = 2$ and $K^{f_3} = 3$ levels, respectively. The N^S ($N^S = K^{f_1} \cdot K^{f_2} \cdot K^{f_3}$) terms of the vocabulary describing the sleeping intensity category ($t^S_{i\in\{1:N^S\}}$) can be represented by a simplified feature vector according to

$$\begin{aligned} t^S_1 &: [S, L^d_1, L^{\text{GSR}}_1, L^{\text{VM}}_1] \\ t^S_1 &: [S, L^d_1, L^{\text{GSR}}_1, L^{\text{VM}}_2] \\ &\vdots \\ t^S_{12} &: [S, L^d_2, L^{\text{GSR}}_2, L^{\text{VM}}_3]. \end{aligned} \tag{3}$$

The sum of terms across different activity levels is the total number of words. In particular, a total of 48 terms were created (12 for S, 8 for VL, 16 for L, and 12 for MV intensity). Note that we did not need to specify the number of unique artificial words (vocabulary size) beforehand.

As a last step, we weighted the informativeness of the words in the vocabulary based on their inverse document frequency (IDF) score. Those terms that have a high IDF are considered more informative, because they rarely occur in the collection. In particular, we removed the words occurring in at least 90% of the documents since, occurring so frequently, they are more likely to obscure than facilitate a meaningful decomposition of the collection of documents. The term frequency (TF), usually used in combination with IDF to form the TF-IDF score [30], was not considered because it penalizes words that appear rarely within a document such as words of moderate to vigorous intensity, these words are important in the identification of routines correlated with the disease. The removed terms are $[\text{VL}, L^d_1, L^{\text{ST}}_1, L^{\text{VM}}_2]$, $[\text{VL}, L^d_1, L^{\text{ST}}_2, L^{\text{VM}}_1]$ and $[\text{VL}, L^d_1, L^{\text{ST}}_2, L^{\text{VM}}_2]$. The IDF score of the words, and, subsequently, the set of removed words, are related to the wildcarding procedure previously described. If more letters are used, the words created will be more specific with the consequence of a higher IDF score average for the words in the vocabulary. On the other hand, a higher threshold on the IDF score (i.e., IDF equal to the one of the words present in 70% of documents) could cause the removal of all the terms useful to identify specific patients' subtypes. Although a rigorous methodology would be necessary to select the optimal combination of feature selected and words removed in order to generate sparse topics, we empirically found out that setting a strict threshold on the IDF (90% of documents) and performing feature selection at the level of the letters works well for our purpose.

C. Routine Discovery

For topic discovery, we used the LDA implementation of [21], and we considered each day of assessment as a separate document. Each IB was mapped with an instance of the vocabulary by associating the selected features in V with their closest levels and then concatenating the three closest levels found. The distances for each bout between the feature point f_j and all the levels L_p are

$$d_p(f_j) = \frac{\left| f_j - L_p^{f_j} \right|}{\sigma_p} \quad \forall p = \{1, \ldots, K^{f^j}\}. \tag{4}$$

Once that a term of the vocabulary was assigned to each IB, documents were created by constructing for each day a histogram of terms. We chose the number of routines (R) equal to 15, and set the hyperparameter α equal to 0.01 as in [18]. Empirically we found that a number of routines greater than 15 led to duplicated routines. Hyperparameters are optimized with a variational expectation maximization algorithm initialized by randomly choosing a small number of "seed" documents [30]. We selected 18 seeds (nine from healthy subjects and nine from COPD patients). Routines did not change in their overall composition with different seed sets.

D. Routine Inference

Once the routines were calculated, first, we inferred day segments (i.e., in this case, a day segment is considered as the equivalent of a text document) in order to know which routines are active during different parts of the day. For this, we used sliding windows of $T = 30$ min of duration as suggested in [19]. From the observations (the bouts) in a sliding window, a histogram of terms has to be created as input for the topic inference. This means that the bouts in the window have to be mapped to terms from the dictionary. We did this by soft assignment as follows. For each bout described by feature vector V, the distances $d_p(f_j)$ of each particular feature f_j to the cluster levels $L_p^{f_j}$ were determined. These distances were converted to feature weights according to

$$w_p(f_j, L_p) = \frac{e^{-d_p}}{\sum_{p=1}^{K^{f_j}} e^{-d_p}}. \tag{5}$$

Thus, smaller distances imply higher feature weights, and the sum of the feature weights over the different clusters equals 1. We then create the term weights by summing all the feature weights. The final normalized term weight W_t is the term weight divided by the sum of all term weights. The normalized term weights are, thus, values between 0 and 1. Normalized term weights of terms associated with other intensity categories were set to 0. Finally, we use the normalized term weights to create the histogram. For each term in the dictionary, we sum all the weights stemming from the all the bouts in the window.

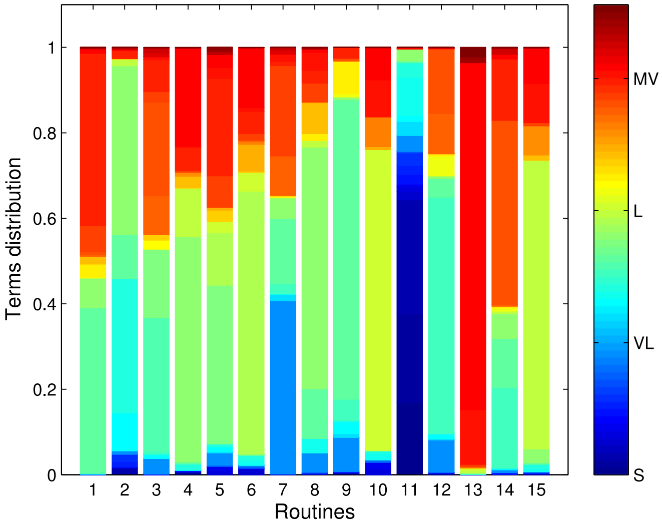

Fig. 3. Distribution of the routines over the terms of the vocabulary.

Second, we applied routine inference on the first 6 h of the assessed days in order to estimate the minutes spent in each routine during the most active part of the day. The same mapping procedure described for sliding windows of 30 min was applied in the case of a unique fixed window of 6 h.

V. Results

Since PA measures during the weekdays and the weekend days are known to be different [26] results were computed separately. Kruskal–Wallis ANOVA test and Dunn's multiple comparisons test were used to determine significance of the results. A value of $P < 0.05$ was considered significant. Pearson's coefficient was used to investigate correlations in the entire dataset of 977 COPD patients. The findings are organized as follows. First, an interpretation of the discovered routines using the dataset of matched COPD and healthy subjects is given to highlight differences in intensity composition of routines and associated physiological responses. Second, results from the inference on window segments of 30 min are presented to qualitatively show two examples of daily routines patterns from a healthy subject and a COPD patient. Discovered routines are successively inferred on the first 6 h of all the assessed days of 66 healthy subjects and 977 COPD patients to quantitatively assess differences between selected groups of subjects and trends across different stages of COPD. Finally, the dataset of matched pairs is used to evaluate the routines in discriminative tasks, such as dividing healthy from patients and recognizing to which population each of the assessed days belongs.

A. Routine Interpretation

Figure 3 illustrates the distribution of the discovered routines $\beta_{1,..,15}$ over the terms of the vocabulary. Three routines related to low intensity levels ($R2$, $R8$, $R9$), nine routines related to moderate-high intensity levels ($R1$, $R3$, $R4$, $R5$, $R6$,

TABLE III
ROUTINES MATRIX

*R*1	%	*R*2	%	*R*3	%	*R*4	%	*R*5	%
MV 1 2 2	0.40	*L* 1 3 1	0.39	*L* 1 1 2	0.31	*L* 1 3 1	0.53	*L* 1 2 2	0.37
L 1 2 1	0.39	VL 2 2 2	0.31	MV 1 1 2	0.22	MV 1 3 2	0.23	MV 1 2 2	0.23
L 1 3 1	0.07	*L* 1 2 1	0.10	*L* 1 2 2	0.16	*L* 1 4 1	0.11	*L* 1 3 2	0.12
*R*6	%	*R*7	%	*R*8	%	*R*9	%	*R*10	%
L 1 3 2	0.62	VL 1 1 1	0.41	*L* 1 3 1	0.57	*L* 1 2 1	0.70	*L* 1 4 2	0.70
MV 1 3 2	0.14	MV 1 2 1	0.21	*L* 1 2 1	0.12	VL 1 1 1	0.08	MV 1 3 1	0.09
L 2 3 2	0.06	*L* 1 2 1	0.15	*L* 2 3 1	0.07	*L* 2 2 1	0.08	MV 1 3 2	0.07
*R*11	%	*R*12	%	*R*13	%	*R*14	%	*R*15	%
S 1 1 3	0.27	*L* 1 1 1	0.55	MV 1 3 2	0.81	MV 1 1 2	0.43	*L* 1 4 1	0.68
S 1 1 2	0.21	MV 1 1 2	0.15	MV 1 3 1	0.13	*L* 1 1 1	0.19	MV 1 3 1	0.09
S 1 1 1	0.17	MV 1 1 1	0.09	MV 2 3 2	0.02	MV 1 2 2	0.14	MV 1 3 2	0.08

The three dominant words for each of the 15 discovered routines.

$R10$, $R12$, $R14$, $R15$) and one routine composed by a combination of VL, *L,* and MV descriptors ($R7$) were discovered in data from 66 COPD patients and 66 healthy subjects. Two separate routines characterizing, respectively, the sleeping behavior ($R11$) and high intensity levels ($R13$) were also found. Each of these routines is characterized by the probability over the terms, where each term is defined by a certain combination of feature levels. For each routine, the three most important terms (i.e., terms of highest probabilities in this routine) can be found in Table III. The words composing a particular routine are listed together with their occurrence probability (e.g., the first word of $R1$: [MV 1 2 2] refers to a simplified feature vector V given by $[\mathrm{MV}, L_1^d, L_2^{ST}, L_2^{VM}]$ and has occurrence probability equal to 40%). Routines, that might be considered similar by only looking at their intensity composition, are different at the level of the descriptors. As an example, Fig. 3 shows that $R3$ and $R5$ are formed for the 20% by MV descriptors and for the 50% by *L* descriptors. Analyzing the letters that compose the descriptors shows that both the MV descriptor and the *L* descriptors of $R5$ have higher values in the feature level corresponding to the ST (second letter, see Table III). This illustrates that some routines differ depending on physiological responses. In absence of appropriate labels for these topics, such insights into routines are relevant for interpreting the inferences of topics per subject.

B. Daily Pattern of Routines

Figure 4 illustrates the activation probabilities of the extracted routines when day segments of 30 min are sequentially inferred. The inference on sequential day segments provides a qualitative representation of the temporal behavior of the routines. The top and the bottom plots show the routine patterns for a COPD and a healthy subject, respectively. It can be seen that three PA routines ($R2$, $R11$, and $R9$) pervade the day of a COPD patient (see top plot of Fig. 4). In particular, this patient spent most of his time performing activities that involve a *L* intensity descriptor of short duration, with high μ_{ST}, and low μ_{GSR} and a VL intensity descriptor of long duration, with high μ_{ST} and high μ_{VM}. Three hours after the patient awoke, $R9$ becomes dominant for about 3 h. This routine includes a *L* descriptor of short duration, high μ_{ST}, and low μ_{GSR}. The graph also shows that the patient slept in the afternoon for about 1 h (probably he rested after having lunch), and after that he continued with the same behavior of the morning. After 14 h from the awakening of the subject, $R2$ starts decreasing and $R11$, that characterizes the sleeping behavior, starts to become the most active routine. On the other hand, the day of the healthy subject shows a larger variety of active routine patterns. The bottom plot of Fig. 4 shows that the routines $R15$, $R12$, $R14$, $R1$, $R15$, $R13$, and $R8$ are sequentially active. Around 9 h after his awakening, just before the evening, this subject assumes a behavior similar to his COPD match, i.e., routine $R2$ becomes the dominant routine until the sleeping time. From the examples, it can be deduced that this particular day of the COPD patient is more static compared to the one of the healthy subject if the number of transitions between the most active routines is considered. Although we did not use the trajectories of the routines as in Fig. 4 in our further analysis they can be used for classification tasks by modeling the trends of routines activation with multitask Gaussian processes [31] or using sequential patterns [32].

In order to extract cumulative (over several days of assessment) characteristics of a subject we, instead, inferred the routines on the first 6 h of the days of each subject as if it was one day segment. On the right side of Fig. 4 the star plots describe the averages of the routines' activation probabilities across the assessed days of the same two subjects when the inference is performed on the first 6 h after their awakenings. In this case, routines' activation probabilities can also be considered as an estimate of the time spent in each routine during the most active part of the day. The subjects are represented as a *star* whose each spoke length is proportional to the average time spent in the associated routine over the assessed weekdays. For clarity, only the most active routines over the assessed days are shown. We see that the shapes of the stars are remarkably different. Routines $R11$ and $R2$ were the most active for the COPD patient while, in contrast, the healthy subject spent most of his time performing more active routines such as $R1$, $R12$, and $R14$. Differences in the shapes of the star plot (i.e., time spent in different routines) are consistent across healthy and COPD patients and will be discussed in the next section.

C. Trends in Routines Activation

Figure 5 shows the average values of the estimated time spent in each routine for all the 977 COPD patients, the matched 66 COPD patients, the matched 66 healthy subjects, and for the 977 COPD patients stratified according to their disease severity, where GOLD1 and GOLD4 indicate, respectively, the least and most severe stage of COPD. Each point in the figure identifies the mean over all the subjects of each group, where each subject is represented by the mean of the time spent in the different routines over the assessed weekdays (left) and weekend days (right). Comparison between the time spent in each of the 15 routines in the 977 COPD patients and the subgroup used to extract the routines (see inside the green dashed rectangles in Fig. 5) shows that there are no statistical differences ($p > 0.2$) in the time spent in each routine between the two groups in both

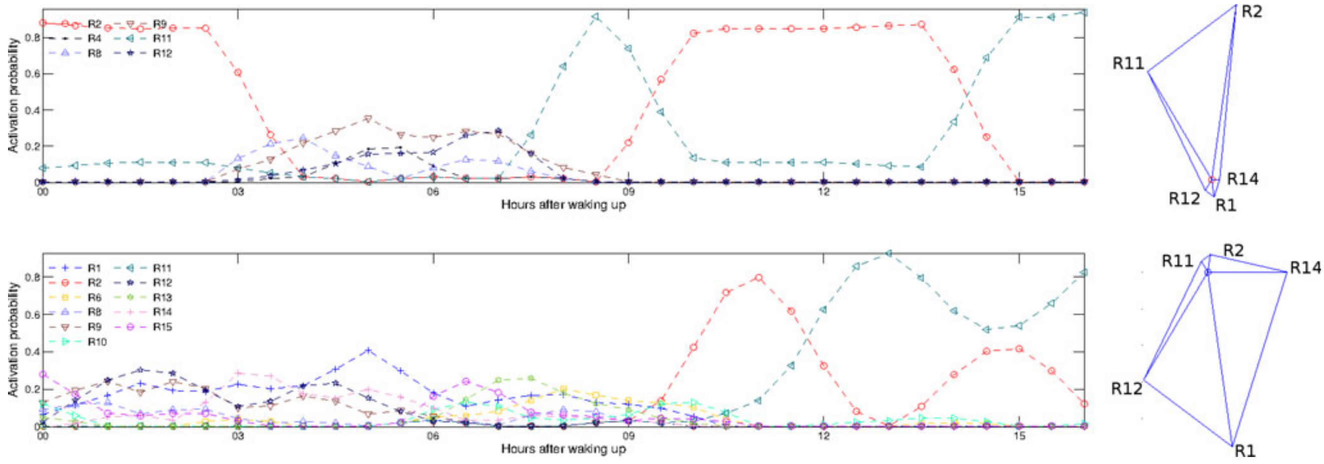

Fig. 4. Left: Routine activation probabilities determined from a particular day of a COPD patient (top) and a healthy subject (bottom). Only topics with a maximum activation larger than 0.1 are shown. Right: Star plots illustrating the average time spent on the three most active routines of the same COPD patient and healthy subject during the first 6 h of their assessed days. $R2$, $R11$, $R12$ and $R1$, $R12$, $R14$ are, respectively, the three most active routines for the COPD patient (top) and the healthy subject (bottom).

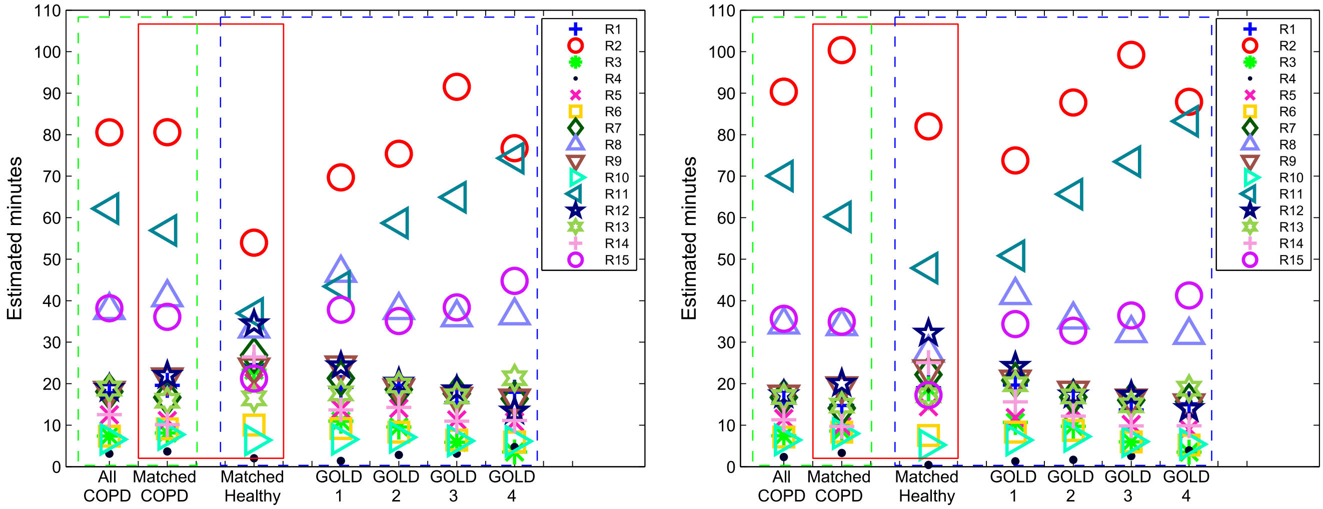

Fig. 5. Activation topic averages during weekdays (left) and weekend days (right). The green dashed rectangles indicate the comparison between all the COPD patients ($n = 977$), and the COPD subset used to generate the routines ($n = 66$). The red rectangles indicate the comparison between the matched COPD patients ($n = 66$) and the matched healthy subjects ($n = 66$). The dashed blue rectangles indicate the comparison between the healthy subjects ($n = 66$), GOLD1 ($n = 89$), GOLD2 ($n = 385$), GOLD3 ($n = 330$), and GOLD4 ($n = 173$). Inference was performed in the first 6 h after the awakenings from the night sleep.

assessed weekdays and weekend days. This indicates that the model is able to generalize across many COPD patients.

When matched COPD and healthy subjects are compared (see inside the red rectangles in Fig. 5), we found statistical differences between the two groups ($p < 0.05$) during weekdays in $R2$, $R3$, $R5$, $R10$, $R11$, $R12$, $R14$, $R15$. During weekend days, statistical differences were found in $R3$, $R4$, $R7$, $R14$, $R15$. A reduced number of statistical differences in the weekend days indicates that the two groups assume a more similar behavior during these days. This might be due to the fact that healthy subjects could still be active workers or that they perform weekly activities such as grocery shopping. On the contrary, they might use the weekend days to sleep more and rest. These assumption could be confirmed by an increased value of the time spent in $R2$ and $R11$ during the weekend days of healthy subjects. High values of these two routines characterize only the COPD patients group during the weekdays.

Comparing healthy subjects and the four different COPD groups (see inside the dashed blue rectangles in Fig. 5), we observe five main trends over the different stages of the disease both in weekdays and weekend days. $R2$ and $R11$ are increasing with the increase of COPD severity. $R2$ represents a medium-inactive PA routine composed for the 39% by [L, L_1^d, L_3^{ST}, L_1^{GSR}] and by the 31% by [VL, L_2^d, L_2^{ST}, L_2^{GSR}]. The first PA descriptor represents light intensity movements characterized by short duration that cause a high increase of the temperature. The second term represents very light intensity movements of long duration that cause a moderate increase of physiological responses (high body temperature and high GSR). The positive trend is interrupted in the most severe group of patients that compensate a smaller value for $R2$ with a higher value of $R11$ and $R15$. A high value of the time spent in $R11$ could be associated with the most severe patients since their conditions might force them to spend more time in bed. $R15$ is characterized mainly by light intensity PA terms characterized by higher physiological responses if compared with $R2$. This might indicate a bigger effort in performing activities. The increased value of $R2$ for healthy subjects during weekend days has been discussed previously. Another rising trend is shown by $R11$ representing the time spent, while performing very inactive behavior (mainly sleeping). On the other hand, we note that $R3$, $R12$, and $R14$ decrease with an increase in COPD severity. These three routines indicate movements performed with medium-high activity intensities characterized by small physiological responses. Of particular, interest are $R3$, $R12$ since they are weakly, but significantly correlated with FEV_1, %predicted ($\rho = 0.3, p = 2\times 10^{-23}$, $\rho = 0.2, p = 5.8\times 10^{-11}$, respectively), and $R14$ weakly correlated with FEV_1 ($\rho = 0.2$, $p = 3\times 10^{-5}$). No correlation was found with age and BMI indicating that these discovered routines are decoupled from these variables. Statistical differences ($p < 0.0001$) have been found in the percentage of activation of $R3$ and $R14$ between healthy subject and all the COPD severity classes (both weekdays and weekend days). For $R3$ statistical differences ($p < 0.0001$) were also found between the first two stage of the disease (GOLD1, GOLD2) and the last two stages (GOLD3 and GOLD4) during the weekdays. For weekend, only between GOLD1 and the last two stage of disease (GOLD3 GOLD4) and between GOLD2 and GOLD4. For $R12$, statistical differences ($p < 0.0001$) have been found between healthy subject and GOLD2, GOLD3, GOLD4 (both weekdays and weekend days). Between the GOLD1 and GOLD3–4 (also during the weekend days), between GOLD2 and GOLD4 (also weekend).

D. Discriminatory Power of Routines

In order to further validate the findings shown in Fig. 5 and assess the discriminatory power of PA routines, we first clustered the 66 *patient:control* matches and, subsequently, their assessed weekdays. For the first clustering experiment, each of the 66 healthy subjects and 66 matching COPD patients was represented by the average time spent in each routine over the assessed weekdays as symbolically shown in the star plots of Fig. 4. Distance-based features were extracted using only $R3$, $R12$, and $R14$ because of their correlation with the FEV_1, %predicted and FEV_1: variables used to assess airway obstruction in the current clinical setting. In particular, we calculated the pairwise *Kullback–Leibler* pairwise distances between all the subjects represented by the averaged time spent in $R3$, $R12$, and $R14$. Using *Kruskal*'s normalized *stress1* criterion, we created a set of locations in three dimensions whose interpoint distances approximate the routine dissimilarities between subjects. We finally clustered the subjects such represented using a k-mean clustering algorithm with $k = 2$ achieving 86% accuracy in dividing the two groups.

For the second experiment, we clustered the 536 assessed weekdays (273 for healthy subjects, 263 for COPD patients), which were represented by the associated daily time spent by the subjects in $R3$, $R12$, $R14$. After extracting the same typology of features, we clustered the days in $k = 2$ groups achieving 82% of accuracy in discriminating between days that belong to healthy subjects and days belonging to COPD patients. This shows that activity biomarkers derived from routines can potentially be used in diagnostic decision support systems, but also in monitoring systems providing feedback on a daily basis.

VI. Conclusion

Unsupervised discovery of latent structures in data from activity monitors is becoming more relevant due to the increasing amount of available multimodal data. Using relatively simple assumptions and settings, we have shown that interpretable and consistent results can be obtained from a large set of unannotated real-life data concerning a large population of COPD and healthy subjects. We have shown that PA routines can be used effectively to integrate and represent the underlying structure of PA measures and physiological responses that characterize the activities of the subjects under study. In particular, it is shown that PA routines in COPD patients and healthy subjects are considerably different regarding their composition and that they show certain consistent trends depending on COPD clinical characteristics. The discovered PA routines were found suitable to label, in an unsupervised way, subjects and assessed days. Moreover, inferring the routine structure on day segments of relatively short duration, it was possible to model PA routine patterns across the day and to identify moments in time in which transitions of the most active routines occur. Some methodological considerations need to be taken into account. First, demographics information (such as working activities) and comorbidities were not available which could influence PA. Second, the routines identified would benefit from a sensitivity study using a new sample of COPD-healthy pairs matching with the same characteristics. Although a more detailed clinical interpretation of the discovered routines is extremely interesting and planned as follow-up of this study, the discovered PA routines apparently reflect the stage of the disease as measured by common clinical practice and could be valid constructs to quantify PA behavior change in patient with limited exercise capacity such as with COPD. As such, it is an encouraging step into the direction of practical applications of these techniques in daily life to design, for instance, interventions and coaching systems with realistic goals for this population.

Acknowledgment

The authors would like to thank Prof. M. Polkey, for supplying the data collected by his group.

References

[1] L. Graat-Verboom, B. E. E. M. van den Borne, F. W. J. M Smeenk, M. A. Spruit, and E. F. M. Wouters, "Osteoporosis in COPD outpatients based on bone mineral density and vertebral fractures," *J. Bone Mineral Res.*, vol. 26, no. 3, pp. 561–568, 2011.

[2] C. J. Caspersen, K. E. Powell, and G. M. Christenson, “Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research,” *Public Health Rep.*, vol. 100, no. 2, pp. 126–131, 1985.

[3] ATS/ERS Task Force on Pulmonary Rehabilitation, “An official European respiratory society statement on physical activity in COPD,” *Eur. Respir. J*, vol. 44, no. 6, pp. 1521–1537, Dec. 2014.

[4] D. Donaire-Gonzalez, D. Gimeno-Santos, E. Balcells, D. A. Rodrguez, E. Farrero, J. de Batlle, M. Benet, A. Ferrer, J. A. Barber, and J. Gea, “Physical activity in COPD patients: Patterns and bouts,” *Eur. Respir. J*, vol. 42, no. 4, pp. 993–1002, 2013.

[5] F. Pitta, T. Troosters, M. A. Spruit, M. Decramer, and R. Gosselink “Activity monitoring for assessment of physical activities in daily life in patients with chronic obstructive pulmonary disease,” *Arch. Phys. Med. Rehabil.*, vol. 86, pp. 1979–1985, 2007.

[6] J. B. J. Bussmann and R. J. G. van den Berg-Emons, “To total amount of activity ..... and beyond: Perspectives on measuring physical behaviour,” *Front. Psychol.*, vol. 4, art. no. 463, pp. 1–6, 2013.

[7] F. Pitta, T. Troosters, V. S. Probst, M. A. Spruit, M. Decramer, and R. Gosselink, “Quantifying physical activity in daily life with questionnaires and motion sensors in COPD,” *Eur. Respir. J.*, vol. 27, no. 5, pp. 1040–1055, 2006.

[8] S. Liao, R. Benzo, A. L. Ries, and X. Soler, “Physical activity monitoring in patients with chronic obstructive pulmonary disease,” *Chronic Obstructive Pulmonary Dis.*, vol. 1, no. 2, pp. 155–165, 2014.

[9] S. Patel, C. Mancinelli, P. Bonato, J. Healey, and M. Moy, “Using wearable sensors to monitor physical activities of patients with COPD: A comparison of classifier performance,” in *Proc. 6th Int. Workshop Wearable Implantable Body Sensor Netw.*, Jun. 2009, pp. 234–239.

[10] S. N. W. Vorrink, H. S. M. Kort, T. Troosters, and J. W. J. Lammers, “Level of daily physical activity in individuals with COPD compared with healthy controls,” *Respir. Res.*, vol. 12, no. 1, pp. 12–33, 2011.

[11] E. Gimeno-Santos, A. Frei, C. Steurer-Stey, J. de Batlle, R. A. Rabinovich, Y. Raste, N. S. Hopkinson, M. I. Polkey, H. van Remoortel, T. Troosters, K. Kulich, N. Karlsson, M. A. Puhan, J. Garcia-Aymerich, and on behalf of PROactive consortium. “Determinants and outcomes of physical activity in patients with COPD: A systematic review,” *Thorax*, vol. 69, pp. 731–739, 2014.

[12] P. Fabio, M. Y. Takaki, N. H. de Oliveira, T. J. P. Sant’Anna, A. D. Fontana, D. Kovelis, C. A. Camillo, V. S. Probst, and A. F. Brunetto, “Relationship between pulmonary function and physical activity in daily life in patients with COPD,” *Respir. Med.*, vol. 102, no. 8, pp. 1203–1207, 2008.

[13] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, “A symbolic representation of time series, with implications for streaming algorithms,” in *Proc. 8th ACM SIGMOD Workshop Res. Issues Data Mining Knowledge Discovery*, 2003, pp. 2–11.

[14] M. M. M. Fuad, “Genetic algorithms-based symbolic aggregate approximation,” in *Data Warehousing Knowledge Discovery* (Lecture Notes in Computer Science), vol. 7448. New York, NY, USA: Springer, 2012, pp. 105–116.

[15] N. Y. Hammerla, Y. Nils, R. Kirkham, P. Andras, and T. Ploetz, “On preserving statistical characteristics of accelerometry data using their empirical cumulative distribution,” in *Proc. Int. Symp. Wearable Comput.*, 2013, pp. 65–68.

[16] S. Saria, A. Duchi, and D. Koller, “Discovering deformable motifs in continuous time series data,” in *Proc. Int. Joint Conf. Artif. Intell.*, 2011, vol. 22, pp. 1465.

[17] P. Schulam, F. Wigley, and S. Saria, “Clustering longitudinal clinical marker trajectories from electronic health data: Applications to phenotyping and endotype discovery,” in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2956–2964.

[18] T. Huynh, M. Fritz, and B.Schiele, “Discovery of activity patterns using topic models,” in *Proc. 10th Int. Conf. Ubiquitous Comput.*, 2008, pp. 10–19.

[19] J. Seiter, O. Amft, M. Rossi, and G. Troster, “Discovery of activity composites using topic models: An analysis of unsupervised methods,” *Pervasive Mobile Comput.*, 2014.

[20] F. Sun, Y. Yeh, H. Cheng, C. Kuo, and M. Griss, “Nonparametric discovery of human routines from sensor data.,” in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2014, pp. 11–19.

[21] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[22] BodyMedia SenseWear. (2015). [Online]. Avilable: http://sensewear.bodymedia.com/

[23] T. Troosters, F. Sciurba, S. Battaglia, D. Langer, S. R. Valluri, L. Martino, R. Benzo, D. Andre, I. Weisman, and M. Decramer, “Physical inactivity in patients with COPD, a controlled multi-center pilot-study,” *Respir. Med.*, vol. 104, no. 7, pp. 1005–1011, 2010.

[24] L. H. Colbert, C. E. Matthews, T. C. Havighurst, K. Kim, and D. A. Schoeller, “Comparative validity of physical activity measures in older adults,” *Med. Sci. Sports Exerc.*, vol. 43, pp. 867–876, 2011.

[25] K. Hill, T. E. Dolmage, L. Woon, R. Goldstein, and D. Brooks, “Measurement properties of the sensewear armband in adults with chronic obstructive pulmonary disease,” *Thorax*, vol. 65, no. 6, pp. 486–491, 2010.

[26] H. Watz, B. Waschki, T. Meyer, and H. Magnussen, “Physical activity in patients with COPD,” *Eur. Respir. J.*, vol. 33, pp. 262–272, 2009.

[27] C. E. Garber, B. Blissmer, M. R. Deschenes, B. A. Franklin, M. J. Lamonte, I. M. Lee, D. C. Nieman, and D. P. Swain, “American college of sports medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise,” *Med. Sci. Sports Exerc.*, vol. 43, pp. 1334–1359, 2011.

[28] D. Cai, C. Zhang, and X. He, “Unsupervised feature selection for multi-cluster data,” in *Proc. 16th Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 333-342.

[29] U. von Luxburg, “Clustering stability: An overview,” *Found. Trends Mach. Learn.*, vol. 2, no. 3, pp. 235–274, 2010.

[30] D. M. Blei and J. D. Lafferty, “Topic models,” *Text Mining, Classification, Clustering Appl.*, vol. 10, pp. 71–94, 2009.

[31] M. Ghassemi, M. A. F. Pimentel, T. Naumann, T. Brennan, D. A. Clifton, P. Szolovits, and M. Feng, “A multivariate timeseries modeling approach to severity of illness assessment and forecasting in ICU with sparse, heterogeneous clinical data,” in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 446–453.

[32] S. Ghosh, M. Feng, H. Nguyen, and J. Li, “Risk prediction for acute hypotensive patients by using gap constrained sequential contrast patterns,” in *Proc. AMIA Annu. Symp.*, 2014.

Authors’ photographs and biographies not available at the time of publication.