33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Modeling high-level descriptions of real-life physical activities using latent topic modeling of multimodal sensor signals

Samuel Kim, Ming Li, Sangwon Lee, Urbashi Mitra, Adar Emken†,
Donna Spruijt-Metz†, Murali Annavaram and Shrikanth Narayanan

Viterbi School of Engineering, University of Southern California, Los Angeles, USA.
†*Keck School of Medicine, University of Southern California, Los Angeles, USA.*
kimsamue@usc.edu

***Abstract*—We propose a new methodology to model high-level descriptions of physical activities using multimodal sensor signals (ambulatory electrocardiogram (ECG) and accelerometer signals) obtained by a wearable wireless sensor network. We introduce a two-step strategy where the first step estimates likelihood scores over the low-level descriptions of physical activities such as walking or sitting directly from sensor signals and the second step infers the high-level description based on the estimated low-level description scores. Assuming that a high-level description of a certain physical activity may consist of multiple low-level physical activities and a low-level physical activity can be observed in multiple high-level descriptions of physical activities, we introduce the statistical concept of latent topics in physical activities to model the high-level status with low-level descriptions. With an unsupervised approach using a database from unconstrained free-living settings, we show promising results in modeling high-level descriptions of physical activities.**

***Index Terms*—unsupervised real-life physical activity modeling, high-level descriptions of physical activities, latent topic models for physical activities**

I. INTRODUCTION

There is a growing body of interest in monitoring health condition through sensor networks and mobile devices [1], [2]. Specifically, measuring an individual's physical activity is beneficial to many healthcare scenarios, such as obesity prevention and reducing the risk of cardiovascular disease [3].

The USC KNOWME project has been developing methods and tools to track users' physical activity, level of stress, and vital signs for health care applications especially pediatric obesity [3], [4], [5], [6]. In this framework, Li et al. performed classification tasks with respect to 9 low-level descriptions of physical activities, such as lying, sitting, walking, etc. With supervised in-lab based data of structured activities obtained from the subjects, the authors demonstrated promising results by fusing various pattern recognition methodologies based on ambulatory electrocardiogram (ECG) and accelerometer signals [5].

In this work, we focus on two main directions: the first goal is to analyze unstructured multimodal sensor data obtained from unsupervised free-living settings, and the second goal is to derive high-level descriptions of physical activities. The motivation for the former is to translate ideas and results from laboratory studies to real-life use of these systems. The rationale behind seeking high-level descriptions is that they can provide summaries of low-level physical activities over a given time period that could be useful for clinical decision making and treatment planning by experts. See Table I for the full list of high-level descriptions that are used in this work.

Analyzing free-living data with respect to high-level descriptions is however very challenging since there is greater variability in the range and types of activities as well as greater measurement uncertainty and no well-defined ground truth. Importantly, one may have non-prototypical activity types, that are often blends of prototypical activity categories. To model high-level descriptions of real-life physical activities, here we propose a two-step modeling strategy which involves an automatic low-level classifier as an intermediate step; the first step will estimate likelihood scores over the low-level classes to describe a person's state and the second step will infer the high-level description that summarizes the estimated low-level physical activities.

It is worth noting that high-level descriptions may include various low-level physical activities simultaneously and any low level activity may be associated with more than one high level description. This many-to-many mapping leads to ambiguities in inferring the high-level semantic descriptions. For instance, *sitting* can be observed in both *homework* and *eating a meal* categories, hence an automated monitoring system cannot determine whether the subject is doing his/her homework or having a meal when it detects the subject is sitting. To mitigate this ambiguity, we introduce the concept of latent topics in physical activities. We hypothesize that each high-level physical activity consists of a set of latent topics and each latent topic has a set of low-level physical activities. Note that similar approaches can be found in text processing and image/audio processing [7], [8], [9], [10]. Assuming that a document (or an image) has a distribution over a set of latent topics and each latent topic, in turn, has a distribution over a set of words (or observations), the concept of latent topics (a.k.a. latent topic models) has been extensively investigated. In this paper, we borrow the idea of the latent topics to describing physical activities and apply the Latent Dirichlet Allocation to implement the idea.

II. LATENT TOPICS IN PHYSICAL ACTIVITIES

In this section, we introduce the notion of latent topics in physical activities for modeling high-level descriptions using low-level descriptions. In the following subsections, details on the low-level description classifier and latent Dirichlet allocation (LDA) are provided followed by actual implementation of latent topic models for physical activities. Note that, as described in the next section, the notion of word and topic in this present context refers to a physical activity primitive and its logical grouping.

A. Low-level description classifier

As an intermediate step to modeling high-level descriptions of physical activities, we use the low-level classifier proposed by Li et al. [5]. The specific low-level descriptions considered in this work are *lying, sitting, sit fidgeting, standing, stand fidgeting, playing wii, slow walking, brisk walking,* and *running*. We can view these as vocabulary items (words) with which we can construct higher level semantic descriptions (See Section II-C). In [5], time domain features and cepstral features are extracted from both ECG and accelerometer, and modeled with a support vector machine (SVM) and a Gaussian mixture model (GMM) respectively. Therefore, individual low-level physical activities would have four different sub-classifiers; ECG$_t$-SVM, ECG$_c$-GMM, accelerometer$_t$-SVM, and accelerometer$_c$-GMM where subscripts $_t$ and $_c$ represent time domain features and ceptral features, respectively, which leads one test set to have $4 \times 9 = 36$ different scores.

B. Latent Dirichlet Allocation

As discussed earlier, the latent topic model which was originally proposed for text signal processing assumes that a document consists of latent topics and each topic has a distribution over words in a dictionary [7]. This assumption can be realized using a generative model such as Latent Dirichlet allocation (LDA). Fig. 1 illustrates the basic concept of LDA in a graphical representation as a three-level hierarchical Bayesian model.

Let V be the number of words in a dictionary and w be a V-dimensional vector whose elements are zero except for the corresponding word index in the dictionary. A document consists of N words, and it is represented as $\mathbf{d} = \{w_1, w_2, \cdots, w_i, \cdots, w_N\}$ where w_i is the i th word in the document. A data set consists of M documents and it is represented as $S = \{\mathbf{d_1}, \mathbf{d_2}, \cdots, \mathbf{d_M}\}$.

In this work, we define k latent topics and assume that each word w_i is generated by its corresponding topic. The generative process can be described as follows:

1) For each document $\mathbf{d}$, choose $\theta \sim \mathrm{Dir}(\alpha)$,
   where θ and α represent the topic distribution probability and the Dirichlet coefficient, respectively
2) For each word w_i in document $\mathbf{d}$,
   a) Choose a topic $t_i \sim \mathrm{Multinomial}(\theta)$
   b) Choose a word w_i with a probability $p(w_i|t_i, \beta)$,
      where β denotes a $k \times V$ matrix whose elements represent the probability of a word with a given topic, i.e. $\beta_{nm} = p(w_i = m|t_i = n)$.

It is apparent from the above that LDA assumes a large number of hidden or latent parameters (θ, $\mathbf{t}$, α, and β) and only one observable variable $\mathbf{w}$. In many estimation processes, parameters are often chosen to maximize the likelihood values of a given data $\mathbf{w}$. The likelihood can be defined as

$$l(\alpha, \beta) = \sum_{w \in \mathbf{w}} \log p(w|\alpha, \beta) \ . \tag{1}$$

Once α and β are estimated, the joint probability of θ and $\mathbf{t}$ with given $\mathbf{w}$ should be infered as

$$p(\theta, \mathbf{t}|\mathbf{w}, \alpha, \beta) = \frac{p(\theta, \mathbf{t}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)} \ . \tag{2}$$

In this work, we investigate the variational approximation method [9] to estimate and infer the parameters of the topic model.

C. Implementation

To implement the idea of latent topics in physical activities with the LDA framework, we need to define the word and the document to represent a physical activity primitive and its logical grouping. Considering the notion of word in this context, there are two options for how we can represent the low-level physical activities in discrete “word-like” format:

- Classification results of the low-level physical activity classifier, e.g., one of nine classes based on our earlier work [5]
- Vector-quantized (VQ) indices of low-level physical activity classifier scores

The size of vocabulary is the number of low-level physical activity classes in the first case and the number of VQ codewords in the second case (in the experiments of this work, there are 9 classes and 200 codewords, respectively). For the document, we collect words from a certain period of time and consider them as a single document.

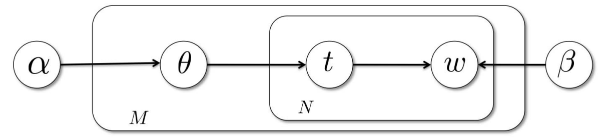


Fig. 1. Graphical representation of the topic model using Latent Dirichlet Allocation.

Group	Index	High-level Description
Eating	1	Eating a meal
	2	Snacking
After school/ Spare time/Hobbies	3	Church
	4	Hanging around
	5	Homework
	6	Listening to music
	7	Music lesson/playing instrument
	8	Playing video games/ surfing Internet while SITTING
	9	Reading
	10	Shopping
	11	Talking on Phone
	12	Watching TV or movie
Sleep/Bathing	13	Getting dressed
	14	Getting ready (hair, make-up, etc)
	15	Showering/bathing
	16	Sleeping
School	17	Lunch/free time/study hall
	18	Sitting in class
	19	Club, student activity
	20	Marching band/flag line
	21	P.E. class
Transportation	22	Riding in a car/bus
	23	Travel by walking
	24	Travel by bicycling
Work	25	Working (e.g., part-time job, child care)
	26	Doing house chores (e.g., vacuuming, dusting, washing dishes, animal care, etc.)
	27	Yard Work (e.g., mowing, raking)
Physical Activities and Sports	28	Aerobics, jazzercise, water aerobics, Taebo
	29	Basketball
	30	Bicycling, mountain biking
	31	Bowling
	32	Broomball
	33	Calisthenics / Exercises (i.e. push-ups, jumping jacks, sit ups)
	34	Cheerleading, drill team
	35	Dance (at home, at a class, in school, at a party, at a place of worship)
	36	Exercise machine (cycle, treadmill, stair master, rowing machine)
	37	Football
	38	Frisbee
	39	Golf / Miniature golf
	40	Gymnastics / tumbling
	41	Hiking
	42	Hockey (ice, field, street, or floor)
	43	Horseback riding
	44	Jumping rope
	45	Kick boxing
	46	Lacrosse
	47	Martial arts (karate, judo, boxing, tai kwan do, tai chi)
	48	Playground games (tether ball, four square, dodge ball, kick ball)
	49	Playing catch
	50	Playing with younger children
	51	Roller blading, ice skating, roller skating
	52	Riding scooters
	53	Running / jogging
	54	Skiing (downhill, cross-country, or water)
	55	Skateboarding
	56	Sledding, tobogganing, bobsledding
	57	Snowboarding
	58	Soccer
	59	Softball/baseball
	60	Surfing (body or board), Skimboarding
	61	Swimming (laps)
	62	Swimming (play, pool games, water volley ball, snorkeling)
	63	Tennis, racquetball, badminton, paddleball
	64	Trampolining
	65	Track and field
	66	Volleyball
	67	Walking for exercise
	68	Weight lifting
	69	Wrestling
	70	Yoga, stretching
	71	Other
	72	Interactive video games (like Nintendo Wii or Dance Dance Revolution [DDR])
	73	Walking dog, Playing with pet

TABLE I
LIST OF HIGH LEVEL CATEGORIES.

III. Database

A. KNOWME in-lab data

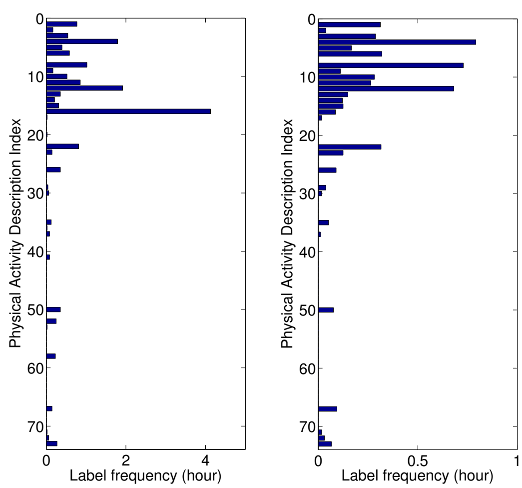
In this database, there are 20 overweight youth (10 male and 10 female; average age 14.6 ± 1.8 years; average body mass index (BMI)[1] percentile 95.8 ± 3.7) and each subject wore the ALIVE heart rate monitor[2] connected wirelessly to a Nokia N95 cell phone to record his/her sensor signals, i.e., electrocardiogram (ECG) and accelerometer signals. The physiological signals were collected and labeled in a supervised way; the subjects were asked to perform a certain physical activity for a given time period. Each subject performed 4 sessions on different days and different times, and each session consists of 7 minutes for each of the nine physical activities, such as sitting and walking. See [4], [6] for more details.

B. KNOWME unsupervised real-life data

There are 12 subjects (subset of in-lab data subjects; 7 male and 5 female; average age 14.6 ± 1.9 years; average BMI percentile 96.4 ± 3.7) and each subject wore the device as done in collecting the in-lab data. The main difference is the fact that the subjects wore the device in their natural free living settings and the data collection was done in an unsupervised way to monitor real-life physical activities. For the collected data, the subjects were asked to record the self-assessment

[1]The body mass index (BMI) is a popular measure of relative weight for height and is used by physicians to evaluate the weight status of patients and by epidemiologists to study disease trends in different population samples.

[2]http://www.alivetec.com/



(a) w/ or w/o physiological data (b) only w/ physiological data

Fig. 2. Data duration distributions in hour and count of number of physical activities based on self-assessment during the data collection: (a) with or without physiological data (b) only with physiological data .

	w/ or w/o physiological data	only w/ physiological data
Total hours per day	16.9	5.4
PA descriptions	34	28

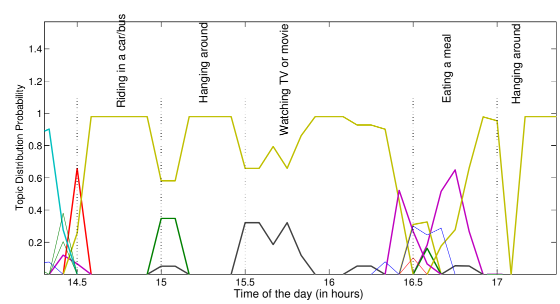
TABLE II
DATABASE STATISTICS BASED ON SELF-ASSESSMENT OF PHYSICAL ACTIVITIES.

of their physical activities based on 73 high-level descriptions in 30 minutes intervals for the corresponding collection day.

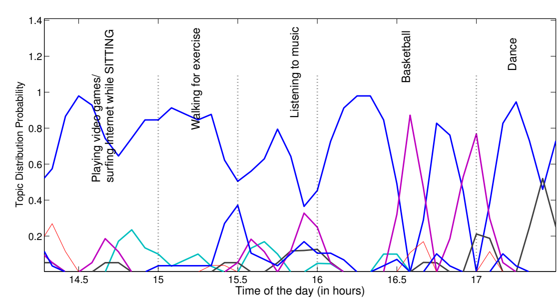
Fig. 2 illustrates the data duration distribution and the corresponding subject-reported counts of physical activities during the data collection; Fig. 2 (a) is based on self-assessment results with or without actual physiological data collected and Fig. 2 (b) is also based on self-assessment results but only with actual corresponding data collected. As shown in the figure (also shown in Table II), the actual physiological data collected covers only a small portion of the high-level physical activities in the list (only 28 categories have corresponding physiological data). It may reflect the fact that that the data collection was performed during weekend days; for instance, no school-related activities can be observed in the database. It may also indicate the tendency of subjects regarding physical activities; note that only a few sport-related activities exist (and the subjects in the dataset were overweight). Although study on the relationship between obesity and sport-related activities is beyond the scope of this work, this data offers some evidence to the nature of the relationship. It is also notable that the data only covers limited time in a day (average 5.4 hours per day) which is significantly small compared to the self-assessed labels. There are 6 categories of physical activities that exist in self-assessment sheet but not in the collected physiological signals. One reason is that the data collection was performed in an unsupervised way and the subjects had failed to turn on the system properly.

IV. Experiments

We built the low-level description models using the KNOWME in-lab data as described in Section II-A. To analyze real-life physiological signals, we segment the database into 20-second chunks and feed them into low-level classifier so that we have low-level physical activity results every 20 seconds. Then, as described in Section II-C,



(a) Subject #19



(b) Subject #22

Fig. 3. Topic distributions over time of different subjects along with self-assesment of their high-level physical activities.

we generate words based on low-level descriptions (either final results of classifiers or vector-quantized scores from the classifiers) and cluster the low-level description results in 10 minutes with 50% overlapping to construct documents. Other segmenting approaches can be studied to facilitate various durations of physical activities.

With the generated documents, we trained the latent topics in physical activities with latent Dirichlet allocation (LDA) method (Section II). Fig. 3 shows the topic distributions over time of different subjects along with self-assesment of their high-level physical activities: (a) for subject ID 19 ad (b) for subject ID 22. The lines in different colors represent different latent topics (30 topics in this case) and the topic distribution at each time instance sums up to one. Note that only a couple of topics out of 30 latent topics dominate the distribution at each time instance. It is also remarkable that the topic distributions at individual time instances are not completely independent (the adjacent ones are actually highly correlated) even though we consider them independent using bag-of-words approach during the modeling procedure. These indicate that the latent topic modeling approach can model some slow-varying properties in low-level physical activities with only a few non-zero parameters of latent topics. In this work, we assume that the slow-varying properties can be represented with high-level descriptions of physical activities.

However, it is difficult to study the direct relationship between high-level descriptions and topic distributions since we can observe, as partially shown in the figure, inter-subject differences as well as intra-subject differences for the same high-level description (assuming the self-assesment of high-level descriptions are accurate). Therefore, we apply another layer of classical machine learning strategy to model the relationship between high-level descriptions and topic distributions. For evaluation, we perform classification tasks with respect to high-level descriptions of physical activities. We use the latent topic distribution as a representative feature of a segment and feed the feature into the support vector machine (SVM) with Bhattacharyya kernel [11] for training and classification.

Fig. 4 shows the accuracy of high-level physical activity classification tasks as a function of the number of latent topics (a solid line with vector quantized classifiers' scores and a dashed line with final classification results). The accuracy is the averaged value of 10-fold cross validation. Since we can only train high-level descriptions with data, we could model only 28 classes present in our corpus. As

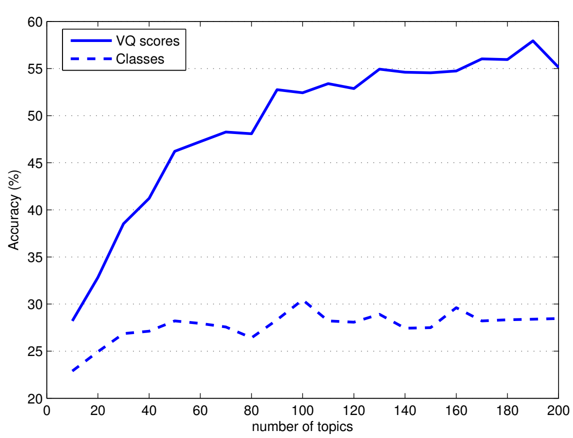


Fig. 4. Accuracy of high-level physical activity classification tasks using latent topics in physical activities as a function of the number of latent topics.

shown in the figure, the classification tasks produce promising results ($>50\%$ accuracy when the number of latent topics are greater than 90 with indices of vector-quantized scores). Note that the cases using indices of vector-quantized scores significantly outperforms the ones using final results regardless of the number of latent topics. This is reasonable because the final classification results may quantize out too much information about subjects' physical activities by choosing only one most likely category of given scores whereas there are different patterns within each low-level description.

V. Concluding Remarks

We introduced a new methodology which models real-life multimodal sensor data, collected in an unsupervised manner from overweight youth, with respect to high-level descriptions of physical activities. The proposed notion of latent topics in physical activities shows promising results in classifying the high-level descriptions based on low-level descriptions of fixed length of physiological data.

In the future, adaptive analysis segment length of physiological data will be studied to facilitate various durations of physical activities based on agglomerative hierarchical clustering (AHC) method as shown in speaker diarization applications [12]. Designing more robust and user-friendly data collection system is another direction for future work.

Acknowledgement

This research was supported by the fund from the National Institutes of Health (NIH) and the National Science Foundation (NSF).

References

[1] D. Huynh, "Human activity recognition with wearable sensors," Ph.D. dissertation, Technische Universtat Darmstadt, 2008.
[2] A. Godfrey, R. Conway, D. Meagher, and G. OLaighin, "Direct measurement of human movement by accelerometry," *Medical Engineering and Physics*, vol. 30, no. 10, pp. 1364–1386, 2008.
[3] G. Thatte, M. Li, S. Lee, A. Emken, S. S. Narayanan, U. Mitra, D. Spruijt-Metz, and M. Annavaram, "Knowme: An energy-efficient, multimodal body area network for physical activity monitoring," *ACM Transactions on Embedded Computing Systems*, 2010.
[4] G. Thatte, M. Li, S. Lee, A. Emken, M. Annavaram, S. S. Narayanan, D. Spruijt-Metz, and U. Mitra, "Optimal time-resource allocation for energy-efficient physical activity detection," *IEEE Transactions on Signal Processing*, 2011.
[5] M. Li, V. Rozgic, G. Thatte, S. Lee, A. Emken, M. Annavaram, U. Mitra, D. Spruijt-Metz, and S. S. Narayanan, "Multimodal physical activity recognition by fusing temporal and cepstral information," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 4, pp. 369–380, 2010.
[6] A. Emken, M. Li, G. Thatte, S. Lee, A. M, U. Mitra, S. Narayanan, and D. Spruijt-Metz, "Recognition of physical activities in overweight hispanic youth using knowme networks," *Journal of Physical Activity and Health*, in press, 2011.
[7] T. L. Griffiths, M. Steyvers, and J. B. Tenenbaum, "Topics in semantic representation," *Psychological Review*, vol. 114, no. 2, pp. 211–244, 2007.
[8] M. Steyvers and T. Griffiths, *Probabilistic Topic Models*. Laurence Erlbaum, 2006.
[9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, 2003.
[10] S. Kim, S. Narayanan, and S. Sundaram, "Acoustic topic models for audio information retrieval," in *Workshop on Applications of Signal Processing to Audio and Acoustics*, October 2009.
[11] T. Jebara, R. Kondor, and A. Howard, "Probability product kernels," *J. Mach. Learn. Res.*, vol. 5, pp. 819–844, December 2004. [Online]. Available: http://portal.acm.org/citation.cfm?id=1005332.1016786
[12] K. J. Han, S. Kim, and S. S. Narayanan, "Strategies to improve the robustness of agglomerative hierarchical clustering under data source variation for speaker diarization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1590–1601, Nov. 2008.