Discovery of Activity Patterns using Topic Models

Tâm Huỳnh, Mario Fritz and Bernt Schiele
Computer Science Department
TU Darmstadt, Germany
{tam, mfritz, schiele}@mis.tu-darmstadt.de

ABSTRACT
In this work we propose a novel method to recognize daily routines as a probabilistic combination of activity patterns. The use of topic models enables the automatic discovery of such patterns in a user's daily routine. We report experimental results that show the ability of the approach to model and recognize daily routines without user annotation.

ACM Classification Keywords
I.5 [Computing Methodologies]: Pattern Recognition

Author Keywords
Activity Recognition, Wearable Computing

INTRODUCTION
Activity recognition has experienced increased attention over the years due to its importance to context aware computing in general and to its usefulness for application domains ranging from medical diagnosis over elderly care to human behavior modeling. This has resulted in various successful approaches that are capable to recognize activities such as walking, biking, sitting, eating or vacuuming. The majority of research has focused on activities that may be described and thus recognized by their respective body movements (such as walking and biking), body posture (such as sitting and eating), or object use (such as vacuuming). For many applications, however, the recognition of such simple activities is not enough. For instance in the case of elderly care it is interesting to recognize daily routines such as shopping or hygiene or in the case of office workers it is interesting to recognize routines such as attending a meeting, having lunch or commuting. What makes the recognition of such routines more complex is that they are typically composed of several activities and that the composition of activities has a large variability depending on factors such as time, location and individual.

This work introduces a novel approach to model and recognize daily routines such as commuting or office work from wearable sensors. For this we propose to leverage the power of probabilistic topic models 1) to automatically extract activity patterns from sensor data and 2) to enable the recognition of daily routines as a composition of such activity patterns. This paper shows that the novel approach can be applied both to annotated activity data as well as to the sensor data directly in an unsupervised fashion. When applied to annotated activity data, the automatically extracted activity pattern often correspond to daily routines. When applied to sensor data, the activity patterns allow recognition of daily routines with high accuracy while requiring only minimal user annotation. Therefore we argue that our approach is well suited both to minimize the amount of user annotation and to enable scalability to long-term recordings of activities.

The main contributions of this paper are threefold. First, we propose a new method to recognize daily routines as a probabilistic combination of activity patterns. Second, we show that the use of probabilistic topic models enables the automatic discovery of the underlying activity patterns. And third, the paper reports first experimental results that show the applicability and the power of the approach to model and recognize daily routines even without user annotation.

The paper is structured as follows. After discussing related work, we will motivate our approach and demonstrate its potential on a set of activity labels covering seven days of unscripted and real-world activity data. We will see that on the ideal set of ground truth labels, our method can reliably model and identify activity patterns that correspond to high-level structure in the person's daily life. Then we describe the technical details of our approach, and introduce the dataset that we used for evaluation. After that, we introduce two different methods for extracting activity patterns from previously unseen sensor data: The first method uses supervised learning to assign activity labels to the sensor data. These labels are then used to identify activity patterns in an unsupervised fashion. The second method is completely unsupervised and uses clustering to generate a vocabulary of labels, which are then used for pattern extraction. We conclude with a summary and outlook.

RELATED WORK
In the following we discuss related work in the area of activity recognition and discovery, with focus on authors aiming towards high-level and/or longterm activities of daily life.

Clarkson et al [5] present an approach for unsupervised decomposition of on-body sensor data into events and scenes.

UbiComp'08, September 21-24, 2008, Seoul, Korea.

They use data from wearable sensors to discover short events such as "passing through a door" or "walking down an aisle", and cluster these into scenes such as "visiting the supermarket" by using hierarchies of HMMs. Conceptually this approach is similar to what our method can achieve. A notable difference is that our method is able to perform well on low-dimensional, low-resolution data from accelerometers, while the approach in [5] relies on high-dimensional and densely sampled audio and video streams. Thus we believe that our method compares favorably both from a computational and also from a privacy point of view.

Eagle et al. [6] used coarse-grained location and proximity information from mobile phones to detect daily and weekly patterns of location transitions. Their work ultimately focuses on the group rather than the individual and explores themes such as social networks and organizational rhythms.

Activity discovery from on-body sensor data is explored in the work of Minnen et al. [14, 13]. Even though we use a similar terminology in this work, our actual goals are quite different from theirs, in that they aim to discover and model short term motion primitives (e.g. those occurring during physical exercies), while we are interested in longer term patterns in the user's daily routine.

Hamid et al. [7] represent activities as bags of n-grams, cluster them into classes of activities, and characterize these classes by frequently occurring sequences. The patterns they discover on a set of 150 days of a person's indoor location traces are coarse and difficult to interpret, though.

In a more office- and desktop-centered setting, Oliver et al. [16] use a layered HMM representation to infer office activities such as giving a presentation, having a conversation or making a phone call, based on lowlevel information from audio and visual sensors as well as from the user's keyboard and mouse activity. In a similar setting, [9] combine device usage with calendar data and time of day/ time of week information to infer a user's availabily. Begole et al. [2] analyze and visualize daily rhythms of office workers by measuring how active (indicated by computer usage) a person is during different times of day.

There is significant amount of work involving location sensors that also aims to extract high-level information about a person's activities from low-level sensor data. E.g., [12] use information from GPS sensors to construct models of high-level activity (such as work, leisure, visit) and to identify significant places (such as home, work, store, etc.). Similarly, [11] use location sensors to make highlevel predictions about driving destinations. These works show that location is a powerful cue to the high-level structure of daily life. However, location is often not enough to identify daily routines reliably, as many different activities can be performed at the same location. E.g. at home, many people are having dinner and breakfast but also perform work. Similarly, in an office room one might work, hold meetings and even occasionally have lunch. Therefore, we consider the work of this paper complementary to these approaches, in that the use of accelerometers allows detection of more fine-grained activities and can also account for different activities performed at the same location. We plan to combine our approach with location sensing in future work.

Another approach for activity recognition is based on object use, e.g. [18, 17, 15]. These authors instrument objects in the environment with RFD tags, and use data from a wearable RFID reader to infer household activities (such as preparing food, doing laundry, washing dishes, etc.). Some of these methods based on Dynamic Bayesian Networks are very flexible in principle, although the use of RFID tags restricts the approaches to closed instrumented environments.

Amft et al. [1] introduce a model to detect composite activities composed of atomic events from a variety of body-worn and environmental sensors. In contrast to our work they focus on relatively short sequences, and the method relies on a significant amount of supervision.

DAILY ROUTINE MODELING USING TOPIC MODELS

The activities we perform in our daily lives can be segmented and characterized on different levels of granularity. Which level to choose depends on the concrete application at hand, but there is evidence that we humans tend to structure and name these levels in a hierarchical fashion, and that at the lower and more fine-grained levels the structure is aligned with physical properties of the activities, such as motion and posture [20]. Research in activity recognition exploits this fact by automatically naming the user's activity, based on low-level sensor data such as the acceleration of different parts of the body.

For many types of activities it is already sufficient to observe a small window of sensor data – usually in the order of seconds – in order to classify them with high confidence. The upper part of Fig. 1 shows a sequence of such activities as they were performed by a subject over the course of one day. If we were to further structure the activities the subject performed on this day, a natural approach would be to group them into *routines* such as *commuting*, *office work*, *lunch routine* or *dinner routine*. Such routines however cannot be identified from their local physical structure alone. What makes their recognition more complex is that they 1) are composed of variable patterns of multiple activities 2) range over longer periods of time, and 3) often vary significantly between instances. A model for recognizing such routines should be able to capture such facts as that *office work* "mostly consists of sitting", but "may (or may not) contain small amounts of using the toilet, or discussions at the whiteboard"; or that *commuting* "mostly consists of driving car, but usually contains short walking instances as well".

It turns out that a family of probabilistic models, commonly referred to as *topic models*, is excellently suited for this kind of task. Before giving more details about how to use and infer topic models we first give an intuitive example what topic models can achieve when applied to activity data.

The lower part of Fig. 1 illustrates the result of our approach

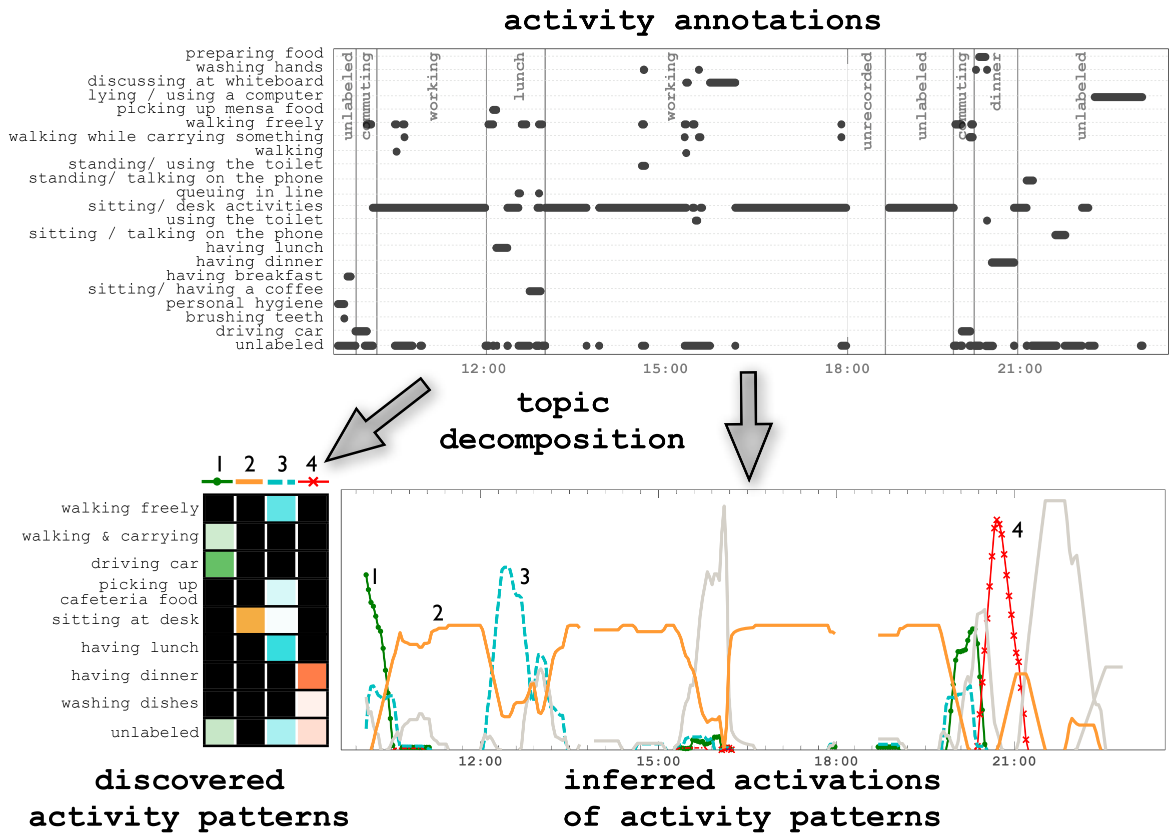


Figure 1. Top: Illustration of our approach on ground truth labels of daily activities. Note that the vertical high-level annotations (commuting, working, etc.) were not given to the algorithm. Lower Left: The matrix shows the contents of four out of ten discovered activity patterns. Lower Right: Inferred activations of the discovered activity patterns during the course of the day (e.g, the pattern in the third column is active during lunch time). Note the high correlation between these activations and the user annotated daily routines in the upper part, suggesting that these activations can be used to model daily routines.

when applied to seven days of ground truth activity labels, including the sequence shown in the upper part of the figure. The columns of the matrix on the lower left represent four out of 10 different activity patterns or *topics* that were automatically identified by the method. Intuitively, each activity has a probability of occurring in the pattern, indicated by the color of the matrix cell. E.g., the third pattern (blue) is most likely to contain the activities *walking freely* and *having lunch*, and also - but slightly less - likely to contain *picking up cafeteria food*, *sitting at desk* and the *unlabeled* class. Similarly, the fourth pattern (red) is likely to contain *having dinner*, *washing dishes* and the *unlabeled* class.

For each of these activity patterns the method is able to tell how much each pattern is *activated* at each point in time. This is shown in the plot on the lower right, in which we plotted the activations of each of the 10 topics for the day shown in the upper part. One can observe that the third pattern (blue) is most active around lunchtime, and that the fourth (red) is active around dinner time. What makes this result remarkable is that no supervision was needed to infer both the activity patterns and their activations over the course of the day. The topic model essentially discovered these activity patterns in a entirely unsupervised way. In this particular case the activations of these activity patterns are highly correlated with the daily routines of the person.

While in this particular example the activity patterns and the daily routines have been discovered in an unsupervised way, they required as input the activity annotations from the user. While this shows the principle applicability of topic models to model daily routines, it is clearly desirable to avoid the time-consuming and error-prone task of manual annotation. Later in the experimental sections we show that topic models can be applied to activity recognition results as well as to sensor data directly. In the latter case no user annotation is required whatsoever and still the discovered daily routines have a high correlation with the user-annotated daily routines.

Topic Models

Topic models stem from the text processing community [8, 4]. They regard a document - e.g. a scientific paper - as a collection of words, discarding all positional information. This is called a "bag-of-words"-representation. As single words capture a substantial amount of information on their own, this simplification has shown to produce good results in applications such as text classification. Assume, for example, an author wants to write a UbiComp paper that covers the three topics “HCI”, “Elderly Care” and “Context-Aware

Computing". Writing this paper then is essentially picking N times a topic from his list of chosen topics and then picking a word appropriate for the topic. Therefore he uses a probability distribution that tells which words are likely to be used for which topic. Different topics might share certain words, which means that both topics assign a high probability to them. This process yields a document in the "bag-of-words" representation with N words.

Most interestingly for the purpose of this paper, topic models allow to infer the inherent topics from an appropriate corpus of documents. E.g. when applied to the corpus of all UbiComp papers one expects that topics such as "HCI", "Elderly Care", and "Context-Aware Computing" (among many other topics) can be discovered automatically without any user annotation or intervention. To illustrate how this can be achieved, we describe the process of writing the documents in a bit more formal way. As mentioned before, the author of document d picks a set of topics. Assuming that he puts different emphasis on the different topics, we model the mixture of topics as (multinomial) probability distribution $p(z|d)$ over topics z. Similarly, the importance of each word for each topic z is also modeled as a (multinomial) probability distribution $p(w|z)$ over words w of a vocabulary. Given these two distributions, we can compute the probability of a word w occurring in document d:

$$p(w|d) = \sum_{z=1}^{T} p(w|z)p(z|d), \quad (1)$$

assuming that there are T topics the documents - e.g. all UbiComp papers - are dealing with. This probability distribution $p(w|d)$ doesn't include any notion of topics any more and in fact can be estimated by simple counting of the words in each document. Having many documents, we observe a data matrix of observed $p(w|d)$ as depicted on the left hand side of the equation in Fig. 2. According to Equation 1 (which is equivalent to the described process of writing the paper), the data matrix can be reconstructed by a matrix product of the word relevances for each topic and a mixture of topics $p(z|d)$ for each document. Estimating the topic model means doing the reverse. The data matrix on the left-hand side is decomposed into the two matrices on the right-hand side, thereby recovering the characteristic words for each topic and the mixture of topics for each document.

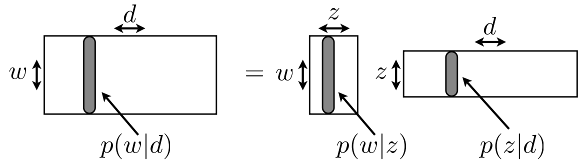


Figure 2. Intuition of topic model decomposition. By introducing an unobserved, latent topic variable z, the observed data matrix of $p(w|d)$ is decomposed into a topic-word matrix of $p(w|z)$ and a document-topic matrix of $p(z|d)$

The described formulation addresses precisely the task we formulated earlier. The data matrix $p(w|d)$ corresponds to the activity data depicted in the upper half of Fig. 1 and the decomposition illustrated in Fig. 2 corresponds to the activity patterns $p(w|z)$ and activations of activity patterns $p(z|d)$ in the lower half. Therefore we propose to discover activity patterns as topic-word distribution and daily routines by topic activation.

In the following experiments we use a particular instantiation of these kind of models - called Latent Dirichlet Allocation (LDA) [4], that extends the described pLSA model to a Bayesian approach by placing a dirichlet prior $p(\theta_d|\alpha)$ with parameter α on the document-topic distributions $p(z|\theta_d)$. Fitting the model is equivalent to finding parameters α for the dirichlet distribution and parameters β for the topic-word distributions $p(w|z,\beta)$ that maximize the likelihood $\mathcal{L}$ of the data for documents $d = 1,\ldots,M$:

$$\mathcal{L}(\alpha,\beta) = \prod_{d=1}^{M} \underbrace{\int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \underbrace{\sum_{z=1}^{T} p(w_n^d|z,\beta)p(z|\theta_d)}_{\text{marginalize over } z} \right) d\theta_d}_{\text{marginalize over topic activations } \theta_d},$$

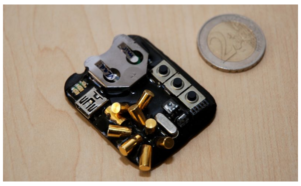
where T is the number of topics and each document d consists of the words w_n^d with $n = 1,\ldots,N_d$. For a more detailed description of the learning process the reader is referred to the original paper by Blei et al. [4]. We also use their implementation, available at [3].

DATASET

To show the effectiveness of the approach, we recorded the daily life of one person over a period of sixteen days. The subject was provided with two wearable sensors, one of which he placed in his right hip pocket, the other on the dominant (right) wrist. The recordings were started in the morning shortly after getting up, and usually ended in the late evening before going to bed. This enabled us to record continuous, non-scripted activities in a natural environment. Due to memory constraints of the sensor platform, the memory had to be emptied after about 4 hrs of recording. A recording of one day typically consists of three such parts, i.e. roughly 12 hrs of data with two gaps in between. In total, our dataset consists of 164 hrs of recordings. Of these, we had to discard 28 hrs due to failures in the sensor hardware.

Sensor Hardware

Fig. 3(a) shows the Porcupine sensor platform [19] which we used to record our set of activities. Besides a 3D accelerometer (ADXL330) and a PIC microcontroller which we used for preprocessing of features, it includes a realtime clock, nine binary tilt switches, a temperature sensor and two light sensors. Data can be stored on 512 kb of flash storage and transferred via a USB connector. In addition, the device features three buttons and three LEDs which can be freely programmed, e.g. for annotation or status display. The platform is small and light enough to be comfortably worn on a wristband (Fig. 3(b)) or slit into the subject's pocket.



(a)

(b)

Figure 3. The wearable sensor platform used for recording activities.

Features

The sensors deliver data at a rate of roughly 100Hz. Due to the memory constraints we subsampled the data by calculating mean and variance over a sliding window of 0.4 seconds (i.e. @ 2.5Hz), and store them along with a timestamp from the realtime clock. This allows to store about four hours of sensor data on the onboard memory of the sensor.

Annotation

To analyze the effectiveness of our approach, we aimed for two different levels of annotations. First, we asked the user to annotate daily routines such as *commuting* or *working*. And second, we also aimed to obtain detailed annotations of the individual activities – at least for part of the data. In total we annotated seven days (84 hrs) in detail, which we used for our experiments. In the experiments reported below we will analyze the recognition of daily routines both with and without these detailed annotations of individual activities. This allows us to show that the approach is not only applicable in supervised settings but also in entirely unsupervised settings.

Finding a good balance between detailed and complete annotations and minimal user disruption is a common problem of studies in activity recognition, especially for long-term studies outside a laboratory. We used a combination of several online and offline annotation methods so that the user had some freedom to choose a method that suited him depending on the situation. Online annotation takes place while the activities are being recorded. We employed three different methods of online annotation, namely experience sampling, a time diary and camera snapshots. During experience sampling, the subject was notified in periodic intervals by an application running on his mobile phone, which presented a set of questions about his current activities. The time diary is a handwritten log in which the subject entered the names, start- and ending times of activities. As a third method, the subject took occasional snapshots with the built-in camera of the mobile phone.

It turned out that for our setting the time diary was the most useful online annotation method, providing detailed information while being far less disrupting than expected. One likely reason for this is that the subject was often working near or at a laptop, which he could use to quickly log activities. Our experience sampling application, while relatively fast and easy to use, tended to miss short events, pose redundant queries and was less precise than the time diary in determining start and ending times of activities. For offline annotation, we visualized the sensor data and aligned it with the annotations from the experience sampling application and the time diary, as well as with the photographs taken by the subject, and had the subject fill in remaining gaps, refine start- and ending times of activities, and also identify and annotate daily routines.

Recorded Activities

Our subject annotated a total of 75 distinct activities and daily routines. For our evaluation, we filtered out activities that occurred only once or for very short durations, and merged similar ones into single classes. Within the individual activities and within the daily routines there is no overlap between annotations, and for both sets we introduced an additional *unlabeled* class, so that in the end each feature is assigned to one activity and one daily routine.

The activity set consists of the following 34 activities, along with the *unlabeled* class (duration in minutes shown in brackets): *sitting / desk activities (3016.9), lying while reading / using computer (196.6), having dinner (125.3), walking freely (123.6), driving car (120.3), having lunch (75.2), discussing at whiteboard (62.6), attending a presentation (48.8), driving bike (46.2), watching a movie (42.5), standing / talking on phone (24.8), walking while carrying something (22.8), walking (22.8), picking up cafeteria food (22.6), sitting / having a coffee (21.8), queuing in line (19.8), personal hygiene (17.2), using the toilet (16.7), fanning barbecue (15.2), washing dishes (12.8), kneeling / doing sth. else (11.6), sitting / talking on phone (8.7), kneeling / making fire for barbecue (8.2), setting the table (8.0), standing / having a coffee (6.7), preparing food (4.6), having breakfast (4.6), brushing teeth (4.3), standing / using the toilet (3.0), standing / talking (2.8), washing hands (2.1), making coffee (1.8), running (1.0), and wiping the whiteboard (0.8)*

Four daily routines (plus the *unlabeled* class) have been annotated that span longer periods of time, typically dozens of minutes to several hours, and which are composed of several distinct activities. The first routine is *commuting (289 min)*, which includes leaving the house and driving to work either by car or by bike, until arriving at the office, and vice versa in the evening. The longest routine is *office work (2814.7 min)*, which mainly comprises desk activities, with occasional interruptions, e.g. when fetching a coffee, visiting an office mate, going to the toilet, attending a meeting, etc. At noon the subject usually went to a nearby cafeteria to have lunch, followed by a stop at a neighboring coffee place. This episode, which usually lasted about an hour per day, is labeled as *lunch routine (391.3 min)*. The last routine is *dinner activities (217.5 min)*, which mostly includes setting the table, having dinner and washing the dishes. As all of the recorded days are weekdays, these four daily routines cover a large percentage of the data, leaving out only some parts in the mornings and evenings.

DISCOVERY OF DAILY ROUTINES BASED ON ACTIVITY RECOGNITION

As discussed earlier, the proposed approach using topic models can be used to model daily routines based on wearable sensors. While the previous section has relied on user generated annotations to discover activity patterns, this section uses supervised activity learning to generate and recognize a vocabulary of activities. Based on the recognized activities topic models are then used to first learn and discover activity pattern which are then in turn used to describe and recognize daily routines. Therefore we first describe how we train a supervised classifier on labeled data, and then use the labels obtained from the classifier as vocabulary for topic estimation. The next section will then describe how the vocabulary can be obtained in an unsupervised way thereby making the approach scalable to large amounts of training data.

Activity Recognition

We evaluated several combinations of features and classifiers on our set of activities. From the acceleration signal we computed several features, including mean, variance, and a number of frequency features, over sliding windows between 0.4 and 4 seconds. As additional feature we used the time of day provided by the realtime clock of the sensor. As classifiers we evaluated SVMs, HMMs and Naive Bayes. They are standard representatives of discriminative and generative classifiers and have been successfully used before for similar tasks (e.g. [16, 10]). Our results are all cross-validated in a leave-one-day-out fashion.

We first compared the three classifiers and different features on a subset of the data spanning two days. It turned out that due to the size of the dataset, SVM and HMM training and classification took significantly longer than Naive Bayes. Due to its time efficiency and since the overall recognition accuracy of Naive Bayes was only marginally lower, we settled for this approach, using as features mean and variance of the 3D-acceleration signal from wrist and pocket motion, plus the time-of-day information from the realtime clock (adding up to a 13-dimensional feature vector). The use of frequency features did not improve the results in our setting, which may be due to the relatively coarse resolution (2.5Hz) of the data.

Overall we achieved an accuracy of 72.7% on the activity dataset. The individual results vary considerably, owing to the diversity of the collected activities. The best five results were obtained for *sitting/ desk activities* (precision 89.5%/ recall 95.4%), *walking freely* (96.2/ 84.2), *standing/ talking on phone* (82.8/ 96.4), *driving bike* (96.7/ 77.1), *having lunch* (76.5/ 97.5) and *personal hygiene* (89.0/ 66.2). A number of activities only occurred on one day, so that the classifiers had no chance of classifying them correctly in our leave-one-day-out crossvalidation protocol. Among these were *kneeling, running, standing while having a coffee, wiping the whiteboard*, and *attending a presentation*. Most of the activities with low recognition scores were either very short (e.g. *washing hands* (precision 3.7%/ recall 3.4%)), so that only little training data was available, or they were confused with other, similar activities (eg. *sitting/ having a coffee* (22.4/ 35.6) was often confused with *desk activities*),

The time-of-day feature enables the classifier to separate activities which share a common motion signature but are performed at different parts of the day – the main improvement in recognition could be observed for the two activities *having lunch* and *having dinner*, for which confusion was virtually eliminated and precision/recall scores improved from 16.8%/ 20.5% to 38.9%/ 47.9% (*having dinner*) and from 62%/ 19% to 97.5%/ 76.5% (*having lunch*). As we used unimodal gaussians to model the Naive Bayes likelihoods, the time-of-day feature did not worsen the results for activities which occur at irregular times during the day – if enough data for such activities exists, then the larger variance usually flattens the likelihood function to a point at which it has little influence on the final posterior. Problems with the time-of-day feature may arise for activities which are not time-dependent but occur only few times in the training data.

Topic Estimation based on Activity Recognition

As a result of our supervised training procedure, we obtain for each data sample a posterior probability for each activity, along with a discrete label that corresponds to the activity with the highest posterior. Our next goal is to discover daily routines from this stream of low-level data, using the framework provided by Latent Dirichlet Allocation.

The main choices one has to make when applying LDA to activity data are the nature and size of the vocabulary, the size of the documents and the number of topics. A simple yet effective way to create documents for the topic models from a stream of activity labels is to use a sliding window of length D over the labels and construct for each window a histogram of label occurrences. In this way each document represents a mixture of activities over a window of time. We found that the outcome of the topic estimation process can be made more robust to noise and misclassifications of the underlying classifier by generating the vocabulary not from the hard assignments of the classifier, but from the soft assignments given by the posterior probabilities for each activity. We achieve this by summing up the posterior probabilities for each activity over the size of one document window, and then generating labels for each activity in proportion to the sum of all posteriors.

Qualitative Results

Figures 4(b) and 4(a) (bottom) show the result of LDA estimation and inference when generating documents over windows of 30 minutes, shifted by 2.5 min at a time. In this example we chose $T = 10$ topics and set the dirichlet prior α to 0.01. The topics in Fig. 4(b) were estimated from six days of data. For each topic z we list all activity labels w with $p(w|z) \geq 0.01$. Fig. 4(a) (bottom) shows the activations of those topics on the day that was left out during training. In each time step we plot the topic activations that correspond to the document covering the preceding 30 minutes.

The first important observation which can be made from the results shown in Fig. 4 is that there are topics that clearly correlate with the daily routines of the subject's day. This can be seen by comparing the topic activations to the daily routines annotated by the subject (Fig. 4(a)). To see how well the estimated topic activations correspond to the mixture of ground truth labels in the respective time window, we also collected the ground truth labels in sliding windows the same size as the documents, i.e. we assigned to each time step the percentage that an activity was 'active' during the last 30 min.

Topics 1 and 2 are both active during office work so that their joint or individual activation is a good indication of office work. In the afternoon topic 6 is activated strongly for a certain period of time, corresponding - on that particular day - to a presentation of a colleague. Topic 6 is a good example of a newly 'discovered' routine – it does not appear in the annotations of the user's daily routines, yet it represents a valid activity pattern that can be modeled and identified. The lunch routine is represented by two topics, namely 3 and 4. As the typical lunch routine is composed of a visit to the cafeteria and the visit of a cafe, topics 3 and 4 have captured the differences in these two "phases" of the lunch routine. Again the activation of either of these topics allows the recognition of the lunch routine. The dinner routine is correlated with the activation of topic 7. The remaining daily routine, *commuting*, is not directly correlated with a single topic but rather with a combination of topics. Both in the evening and in the morning the co-activation of various topics including topics 5, 6 and 3 allow to identify this routine.

Let's now turn to the contents of the topics, i.e. the learned activity labels that have a high probability of being part of a particular topic. As can be seen from Fig. 4(b), the content often represents a meaningful set of activity labels. E.g., the prominent words in topic 3 are having lunch, walking, picking up cafeteria food and queuing in line. Topic 5 is a mixture of driving car, walking, and desk activities and is activated during the *commuting* routine of the subject. Topics 1 and 2 represent desk activities and are active during the *office work* part of the subject's day. Topic 7 contains having dinner and washing dishes, as well as desk activities and driving car, which all correspond to evening and dinner activities of the subject.

Since the accuracy of the underlying classifier that generates the vocabulary is not perfect, there are errors due to misclassifications, some of which are reflected in the contents of topics. E.g., the classifiers for the activities *using the toilet* and *standing / using the toilet* fire relatively often, but only with precision of 18% and 7%, respectively. Partly due to a small amount of training data, they are often confused with similar activities such as *desk activitites* and *standing at the whiteboard*. As a consequence, in the example shown in Fig. 4, their labels are weighted too strong in topics 4 and 6. A more powerful activity recognition algorithm would help to alleviate such problems even though it is expected that significant ambiguities between activities remain. An important and relevant property of topic models is that they are robust to these types of ambiguities.

Evaluation Method

While the plots of the topic activations suggest that the topics are indeed able to discover and model activity patterns and therefore high-level structure in the subject's daily activities, it is not obvious how to quantify the results. We propose two different measures for evaluating the quality of the topic decompositions: correlation and recognition performance. For both measures we use as ground truth the daily routine annotations by the subject. First it should be noted, though, that both methods are not optimal, since LDA is an inherently unsupervised method which is able to discover meaningful structure a user was previously unaware of. Such ability cannot be quantified when evaluating against a predetermined ground truth.

For the correlation measure, we first perform LDA estimation on six of the seven recorded days. We then assign to each activity the topic to which the correlation to the ground truth annotation is highest. Next we perform LDA inference on the seventh day and note for each activity the correlation with its assigned topic. We repeat this in a leave-one-day-out fashion and report the average results for each daily routine. In order to compute recognition performance, we use the topic activation vectors as features for a supervised learning task. More specifically, we first perform LDA estimation and inference on six of the seven days, and then train a nearest neighbor classifier using the obtained topic activation vectors and the daily routine ground truth. We then perform LDA inference on the seventh day and classify each of the resulting activation vectors using nearest neighbor. The results we report are again cross-validated over the seven days of data.

Baseline Results

In order to obtain a baseline for the recognition of routines, we built a supervised classifier using HMMs based on the same features that we use for the LDA-approach, i.e. acceleration features from wrist and pocket sensor, plus time-of-day. We used left-right models and varied the number of states Q, the number of gaussians per state M, as well as the length of the observation sequence O. The cross-validated results for the best parameter-combination that we found ($Q = 5$, $M = 2$, $O = 30min$, shifted by $5min$) are shown in Fig. 5. The *lunch* and *office work* routines can be predicted with high precision and recall. *Lunch* is a short, yet very regular routine, usually taking place between noon and 1pm. *Office work* covers a large part of the day and consists to a large part of sitting activities. In contrast, *dinner* and *commuting* are relatively short routines that occur at relatively irregular times of day, which makes recognition more challenging. This is reflected in the lower recall values. In the remainder of the paper we will use these results as a baseline for the recognition of routines using topic models.

Routine	*Precision*	*Recall*
Dinner	88.6	27.3
Commuting	72.6	31.5
Lunch	84.4	80.7
Office Work	89.2	91.1
Mean	83.7	57.7

Figure 5. Baseline recognition results, using HMMs based on acceleration and time-of-day features.

Quantitative Results

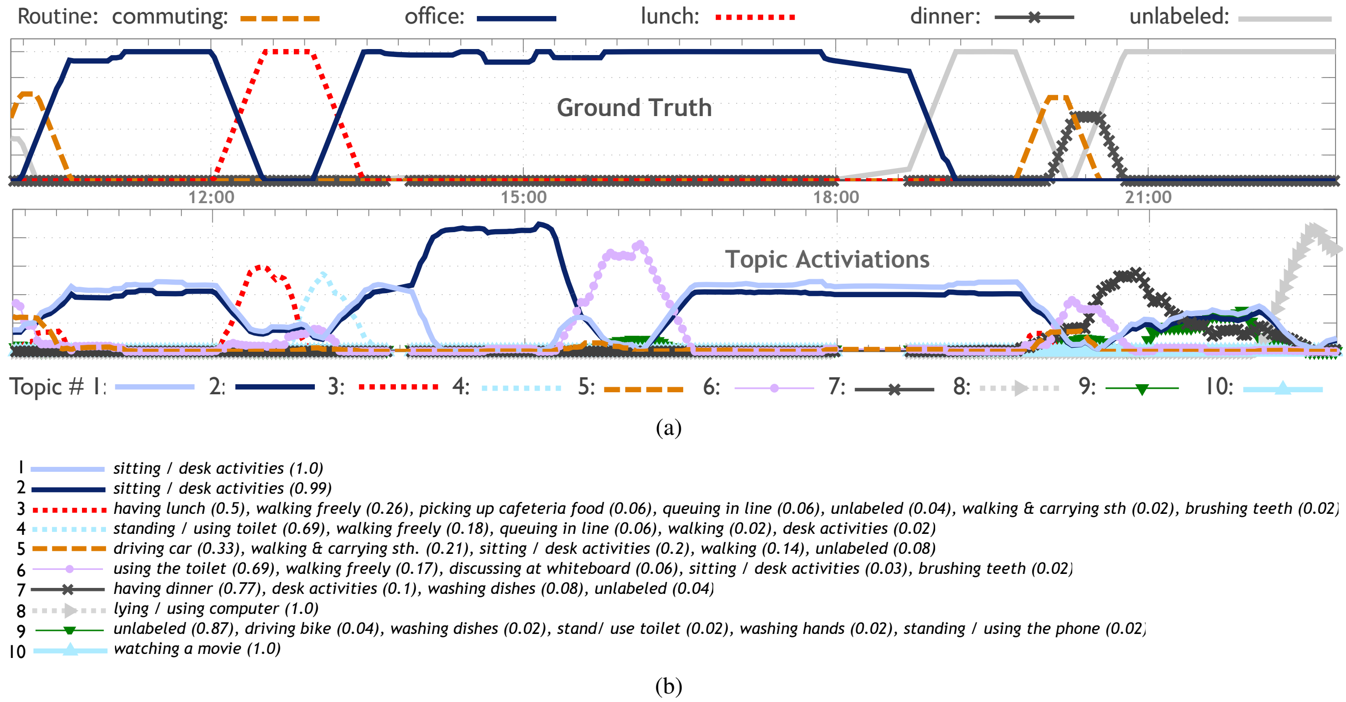


Figure 4. (a) Ground truth and topic activations for one day, based on a vocabulary of learned activity labels. (b) Contents of the ten estimated topics. The numbers in brackets indicate $p(w|z)$, i.e. the probability of the activity label w given the current topic z (labels w with $p(w|z) < 0.02$ are not shown). The distributions were estimated from six days of data. (a) shows the inferred topic activations for the day that was left out during training.

Fig. 6 shows the correlation and recognition results for the best combination of parameters when we used learned activity labels as vocabulary. In this case we used $T = 10$ topics, a document length of 30 min, and soft assignments from class posteriors to generate the words for each document. *Office work* is best correlated and recognized, followed by *lunch*, *commuting* and *dinner*. Comparing to our baseline results (Fig. 5), we can see that the recognition of routines has improved. The values for precision and recall increase throughout, with the exception of precision for *dinner* routine. Overall, the results indicate that the estimated topics relate to high-level structure in the subject's daily routine.

Routine	*Correlation*	*Precision*	*Recall*
Dinner	0.7	75.5	40.2
Commuting	0.6	85.5	51.8
Lunch	0.8	87.0	83.3
Office Work	0.8	96.4	93.7
Mean	0.7	86.1	67.2

Figure 6. Correlation and recognition results when using topics estimated from learned activity labels.

Influence of Parameters

For the daily routines in our data set, correlation with topics dropped noticeably when choosing document windows smaller than 30min. In general our results indicate that choosing document lengths on the order of the average lengths of the routines seems a good strategy. We also found that using more topics may lead to better recognition results when using topics activation vectors as features, but makes (visual) discovery of unknown routines more difficult, as the topic activation plots get more noisy.

UNSUPERVISED LEARNING OF DAILY ROUTINES

In the previous section we showed how topics can be used as a means of inferring high-level structure from a vocabulary of labels representing relatively short-term activities. These labels were learned in a supervised fashion from a stream of sensor data. An advantage of this approach is that the estimated topics carry an inherent meaning, which is expressed by the distribution of labels within each topic. A substantial disadvantage, though, is the amount of annotation effort associated with the supervised learning part. In this section we describe how the vocabulary for the topic estimation can be constructed in an unsupervised fashion and we will show that surprisingly good results can be obtained without any need of tedious and detailed activity annotation.

Clustering of Activity Data

To generate discrete labels from continuous sensor data in an unsupervised fashion we simply use data clustering. This allows to assign to each sample the index of the closest cluster centroid. While this is essentially the basis of our approach, we again found that using soft instead of hard assignments did improve our results. In order to create a vocabulary of size N, we first cluster our feature vectors using K-means clustering with $K = N$. For each feature i we store the distances $d_{1..N}$ to the centroids of each cluster. We then convert these distances to weights $\omega_{1..N}$ with

$$\omega_i = \frac{e^{-\frac{d_i}{\sigma}}}{\sum_{j=1..N} e^{-\frac{d_j}{\sigma}}} \quad (2)$$

Thus smaller distances imply higher weights, and the weights for one feature sum up to one. The parameter σ controls how fast the weights decline for more distant clusters. Empirically we found that setting σ to the standard deviation of all distances worked well. We next use the weights to construct documents of size D in the same fashion as for the supervised case described in the previous section. More specifically, for each cluster i we sum up the weights ω_i over a feature window of length D, and then generate m_i labels for this cluster by multiplying the sum of its weights by the document length D and rounding to the next integer. Since the weights for each feature are a partition of 1, the document will contain at most $\sum m_i = D$ labels.

Results

Fig. 7 shows an example of the result of LDA inference using a vocabulary of 10 cluster labels, together with the daily routine ground truth for this day. The documents were created from sliding windows of 30 min, shifted by 2.5 min at a time. LDA estimation was performed on six of the seven days, and inference on the remaining day. Again one can observe that the topic activations reflect the annotated daily routine structure of the subject's day, even though this time no annotations (neither for activities nor for daily routines) were given at all. Furthermore, there are individual topics whose activation is strongly correlated with the *lunch*, *office work* and *commuting* routines.

Fig. 8 shows correlation and recognition scores for the best combination of parameters when using a vocabulary of cluster labels for topic estimation. In this case we used $T = 10$ topics, a document length of 30 min, and $N = 60$ clusters. Note that low correlation does not necessarily imply bad recognition performance, as can be seen for the *commuting* activity. This is because we compute correlation between individual topics and daily routines, while recognition uses the activations of all topics at each time step. Thus if a daily routine can be characterized by a mixture of topics instead of a single topic, recognition scores may be high even though the best correlation of an individual topic is low.

Comparing the results to the supervised method described in the last section (Fig. 6), one can observe that the mean correlation and precision are lower in the unsupervised case, with about 10% less overall precision and a drop of 0.1 in correlation score. However, overall recall declines only slightly, and the individual recognition scores for *office work* and *commuting* remain high. One likely reason for the drop in precision for the *lunch* and *dinner* routines is that they share many activities and are therefore not separated well by the clustering. As a consequence, recognition of *lunch* drops below our baseline results. However, compared to the baseline, recall for *dinner* and *commuting*, as well as precision for *commuting* are higher, indicating that the approach can compensate for the irregular occurrences of these routines.

Finally, keep in mind that these results are based on predefined ground truth, and thus do not capture the ability of the method to discover previously unknown structure in the data.

Routine	*Correlation*	*Precision*	*Recall*
Dinner	0.6	56.9	40.2
Commuting	0.5	83.5	71.1
Lunch	0.8	73.8	70.2
Office Work	0.6	93.4	81.8
Mean	0.6	76.9	65.8

Figure 8. Correlation and recognition results when using topics estimated from k-means cluster labels.

Discussion

In this section we used clustering as an unsupervised method to generate a vocabulary of discrete labels from a stream of continuous activity data. We used this vocabulary as basis for topic estimation and observed that the estimated topics correlate with daily routine structure in the subject's activities. The main advantage of this approach is that it does not require any labeled training data and yet is able to discover structures that are of relevance to the subject. As the approach is entirely data-driven, we don't rely on any noisy classifier output, and hence there are no 'wrong' words that the topic model has to deal with, as we observed in the supervised case. On the other hand, the contents of the topics, i.e. the distribution over cluster labels, carries no direct meaning for an observer. Such meaning can be established, however, via the additional step of comparing the topic activations to the actual structure of the subject's day, and then identifying topics that correspond to possible daily routines.

CONCLUSION AND OUTLOOK

In this paper we have introduced a novel approach for modeling and discovering daily routines from on-body sensor data. Inspired by machine learning methods from the text processing community, we convert a stream of sensor data into a series of documents consisting of sets of discrete activity labels. These sets are then mined for common topics, i.e. activity patterns, using Latent Dirichlet Allocation. In an evaluation using seven days of real-world activity data, we showed that the discovered activity patterns correspond to high-level behavior of the user and are highly correlated with daily routines such as *commuting*, *office work* or *dinner routine*. The patterns can be based on a learned vocabulary of meaningful activity labels (such as *walking*, *using the phone*, *discussing at whiteboard*, etc.), in which case the discovered patterns are immediately human-readable in that they represent sets of such labels. Learning of labels requires a supervised component, which can be avoided by applying our method directly to unlabeled sensor data using clustering. In this case, the method is fully unsupervised, yet still allows to visualize high-level structure of the data, as well as to identify activity transitions, novelties and anomalies.

Thus, we think that both the (partly) supervised and the unsupervised approach have advantages and limitations, which should be considered in the light of specific application scenarios. Moreover, the approaches need not necessarily ex-

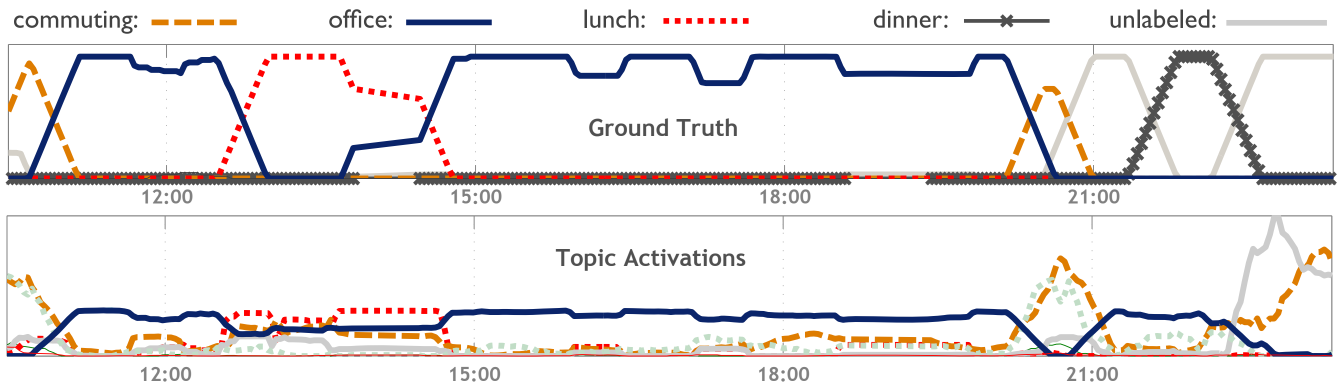


Figure 7. Top: Daily routine ground truth for one day. Bottom: Inferred topic activations, based on a vocabulary of ten cluster labels. Ten topics were estimated from six days of data, and the plot shows the activation of these topics on the day that was left out during training.

clude each other. E.g., the unsupervised approach can help to detect anomalies, but not necessarily tell what exactly happened (e.g. *getting up at night to go to the toilet*, vs. *getting up to sleepwalk*). This could be addressed by the use of semi-supervision, e.g. by presenting the user with a visualization of topic activations such as in Fig. 7, and asking him to label the discovered topics. We intend to address these and other limitations of the unsupervised approach, such as the occasional modeling of noise present in the data, in future work. Other possible extensions would be to use the vector of pattern activations as a high-level feature for more sophisticated classifiers, or to incorporate additional features such as location information.

In conclusion, we believe that our approach is highly appealing for the field of activity recognition, and that so far we have only exploited some of its potential. E.g., as can be seen from the topic activation plots, the probabilistic nature of the approach allows for handling of concurrent and overlapping activities (expressed as co-activation of patterns), and also transitions between activities (e.g. the user is *about to go to lunch*). We consider these properties, together with the ability to decompose routines into their low-level constituents, as a crucial advantage over traditional unsupervised techniques such as clustering.

Acknowledgements. This work has been supported by the European Commission funded project MOBVIS under grant number FP6-511051.

REFERENCES

1. O. Amft, C. Lombriser, T. Stiefmeier, and G. Tröster. Recognition of user activity sequences using distributed event detection. In *Second European Conference on Smart Sensing and Context (EuroSSC)*, October 2007.
2. J. Begole, J. Tang, and R. Hill. Rhythm Modeling, Visualizations, and Applications. *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2003)*, pages 11–20, 2003.
3. D. Blei. C implementation of variational EM for latent Dirichlet allocation (LDA), available at http://www.cs.princeton.edu/ blei/lda-c/, 2006.
4. D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.
5. B. Clarkson and A. Pentland. Unsupervised clustering of ambulatory audio and video. In *icassp*, 1999.
6. N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4):255–268, 2006.
7. R. Hamid, S. Maddi, A. Johnson, A. Bobick, and C. I. I. Essa. Unsupervised discovery and characterization of activities from event-streams. In *UAI*, 2005.
8. T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning Journal*, 42(1):177–197, 2001.
9. E. Horvitz, P. Koch, C. M. Kadie, and A. Jacobs. Coordinate: Probabilistic Forecasting of Presence and Availability. In *Proc. UAI*, pages 224–233. Morgan Kaufmann Publishers, July 2002.
10. T. Huynh, U. Blanke, and B. Schiele. Scalable recognition of daily activities with wearable sensors. In *3rd International Symposium on Location- and Context-Awareness (LoCA)*, pages 50–67, 2007.
11. J. Krumm and E. Horvitz. Predestination: Inferring Destinations from Partial Trajectories. In *Proc. UbiComp*, 2006.
12. L. Liao, D. Fox, and H. Kautz. Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields. *The International Journal of Robotics Research*, 26(1):119, 2007.
13. D. Minnen, T. Starner, I. Essa, and C. Isbell. Discovering characteristic actions from on-body sensor data. In *Proc. ISWC*, October 2006.
14. D. Minnen, T. Starner, J. Ward, P. Lukowicz, and G. Troster. Recognizing and Discovering Human Actions from On-Body Sensor Data. In *Proc. ICME*, pages 1545–1548, 2005.
15. U. Naeem, J. Bigham, and J. Wang. Recognising Activities of Daily Life using Hierarchical Plans. In *EuroSSC*, October 2007.
16. N. Oliver, E. Horvitz, and A. Garg. Layered representations for human activity recognition. *Proc. ICMI*, 2002.
17. D. Patterson, D. Fox, H. Kautz, and M. Philipose. Fine-grained activity recognition by aggregating abstract object usage. In *Proc. ISWC*, pages 44–51, 2005.
18. M. Philipose, K. P. Fishkin, M. Perkowitz, D. J. P. D. Hahnel, D. Fox, and H. Kautz. Inferring Activities from Interactions with Objects. *IEEE Pervasive Computing: Mobile and Ubiquitous Systems*, 3(4):50–57, 2004.
19. K. Van Laerhoven, H. Gellersen, and Y. Malliaris. Long-Term Activity Monitoring with a Wearable Sensor Node. *BSN Workshop*, 2006.
20. J. Zacks and B. Tversky. Event structure in perception and conception. *Psychological Bulletin*, 127(1), 2001.