Towards Recognizing Abstract Activities: An Unsupervised Approach

Albert HEIN and Thomas KIRSTE

Dept. of Computer Science
University of Rostock
18059 Rostock, Germany
{albert.hein, thomas.kirste}@uni-rostock.de

Abstract. The recognition of abstract high-level activities using wearable sensors is an important prerequisite for context aware mobile assistance, especially in AAL and medical care applications. A major difficulty in detecting this type of activities is that different activities often share similar motion patterns. One possible solution is to aggregate these activities from shorter, easier to detect base level actions, but the explicit annotation of these is not trivial and very time consuming. In this paper we introduce a simple clustering based method for the recognition of compound activities at a high level of abstraction using k-Means as an unsupervised learning algorithm. A general problem of these methods is that the resulting cluster affiliations are typically not human readable and some kind of interpretation is needed. To achieve this, we developed a hybrid approach using a generative probabilistic model built on top of the clusterer. We adapted a Hidden Markov Model for mapping the cluster memberships onto high-level activities and sucessfully evaluated the feasibility of this technique using experimental data from two test runs of a home care scenario showing a higher accuracy and robustness than conventional discriminative methods.

Keywords. Activity Recognition, High-Level Activities, Clustering, Probabilistic Models, AAL

1. Introduction

Activity Recognition based on wearable sensors is a growing and fast changing field of research and is widely seen as a major prerequisite for context aware computing and mobile intelligent assistance systems. In our work we focus on assistive technology supporting elderly people during ageing at home, a field of application often referred as *Ambient Assisted Living* (AAL). But as technical aids cannot satisfy all requirements, at some stage people depend on human medical care. An alternative to stationary treatment is the professional ambulant elderly care at home. Especially this kind of service needs accurate documentation of care activities to allow correct accounting for the health insurances. The usual documentation process is to this day still done manually and takes up to 40% of the working time, is error-prone and mostly inaccurate because it does not happen in situ but afterwards. Our research tries to automate this documentation process by recognizing these care activities.

Using inertial measurement units consisting of accelerometers, gyroscopes, magnetometers or combinations of them many research groups already reported successful recognition of simple, often called "base level" activities [1,2,3,4]. Those natural activities like walking, running, sitting, climbing stairs, etc. are characterized by a distinct corresponding, often periodic motion pattern. In most cases simple discriminative pattern recognition approaches, static classifiers like Support Vector Machines or Decision Trees are sufficient for detecting this kind of activities at a surprisingly high rate. Auxiliary generative models built upon the classifiers are mostly used as a kind of temporal smoothing. Depending on the kind of activities to be detected usually only parameters like extracted features and window lengths may vary. Unfortunately the care activities are more complex.

The recognition of higher-level activities still is a current research objective. These are generally more abstract and artificially defined, compound activity sequences with ambiguous motion patterns, where the described methods fail. Simple base-level actions and motions are repeatedly shared between different activities. An example taken from the care scenario would be the big vs. small morning toilet (washing whole body, brushing teeth, shaving vs. only washing the body).

Probabilistic models are capable of handling noisy, unsure and incomplete sensor data and are able to include causal and temporal dependencies and prior knowledge into the decision process which can, besides temporal smoothing, greatly enhance the recognition rate of such complex activities. They are able to represent the compound character at a higher level of abstraction making it possible to compose a compound activity from several easily distinguishable base level actions like building a sentence from single words. If supervised learning is used for the base level activity layer a very detailed annotation of the underlying training material is essential. Each single action building block like step, crouch, turn around, rotate wrist, lean forward, etc. must be tagged. Selecting actions which are actually important or easily distinguishable (e.g. different types of leaning forward) is a non-trivial, extremely time consuming and because of the available sensor data a mostly unrealistic task which must be repeated for every single application domain.

Unsupervised machine learning techniques on the other hand don't need any annotated training data as they aim to find inherent structures. The main disadvantage of these methods for the activity recognition task is that those structures - or clusters - are generally not easily comprehensible as they don't distinctly map to the anticipated activities. Our approach uses generative probabilistic models for learning cluster memberships as model state emissions to map them onto abstract activity classes.

Our key contributions in this paper are: A novel method for recognizing very abstract (arbitrarily defined) compound high-level activities containing shared and repetitive occurring partial base level activities first published in this work. Furthermore we reduced the annotation complexity through the use of unsupervised learning methods needing only coarse annotation of the target activities for the overlying model. We realized the interpretation of cluster affiliations, the inclusion of temporal and causal coherences and of prior knowledge through a probabilistic model which is also capable of handling very different sensor modalities and the automatic determination of model parameters from training data considering individual linear reliability biases. Finally we evaluated the feasibility of our approach regarding recognition accuracy and robustness where it outperforms conventional supervised techniques.
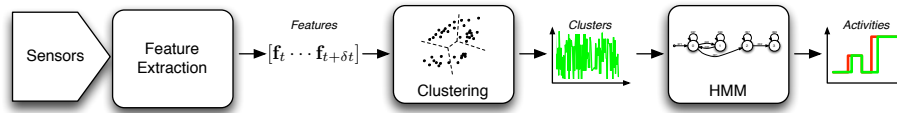
Figure 1. Inference System Overview: For each timestep t a feature vector $\mathbf{f}_t$ is calculated from the raw sensor data. The Clusterer assigns each $\mathbf{f}_t$ to a single cluster which represents a not explicitly named base level activity. The Hidden Markov Model allocates each cluster sighting to the currently most likely abstract activity class.

2. Technical Challenges

While the recognition of basic low-level activities works quite well using classical pattern recognition techniques, abstract and high-level activities are much more problematic and the research at its beginnings. One can illustrate this problem imagining a human observer without special domain knowledge. While simply watching he can intuitively distinguish between base-level actions due to their characteristic motion patterns, but he would not be able to discriminate different abstract and arbitrarily defined activities as he doesn't know about their elementary base-level components.
To summarize some general challenges:

- Interleaved or interrupted activities which are not executed sequentially
- Ambiguities between different activities sharing the same motions or gestures
- Variations in the activity performance between single or multiple subjects and distortions by uninvolved persons
- Different levels of complexity between elementary or compound activities
- Different levels of granularity between coarse motion and fine grained gestures

A layered approach would address most of these problems by representing different levels of abstraction by different model layers. Using discriminative supervised methods for the underlying base-level recognition, as they have proven to be successful at this task, brings out a new difficulty. Choosing relevant and easily distinguishable actions and annotating them is non-trivial and very time consuming as already mentioned in the introduction. Utilising unsupervised clustering methods avoids this problem, but does not produce an interpretable classification result anymore, so some kind of mapping is needed.

3. Algorithms

For detecting abstract compound and high-level activities we are following a hybrid two-level approach reflecting the inherent structure of the given activities. As they are consisting of multiple base level actions corresponding to relatively easy to detect motion patterns, we use an underlying unsupervised discriminative layer for identifying them. As we are not interested in the actual sequence of these low-level actions but in the high-level activity trajectory these basic building blocks must be mapped to abstract activities as we mentioned in the section before. For this task we assign the outputs of the underlying layer as emissions of single states of a generative model (HMM), representing the compound activities. An abstract overview of the system is given in Fig. 1.

3.1. Related Work

The idea of breaking the task of detecting activities into two steps is not new in general. 2005 Lester et. al. [1] presented a related approach for detecting base level activities which used a discriminative layer (boosted decision stumps) and a second layer of HMMs as temporal smoothers. Wang et al. [5] increased the representational power of the probabilistic model by adding a Dynamic Bayesian Network for including temporal characteristics and RFID-based object detection. Our work extends these approaches by using unsupervised learning for the basic actions and one single overlying HMM for modelling and inferring the abstract activities.

As noticed before the explicit labelling of each base level activity is extremely time consuming and the selection of relevant and easy detectable classes is non-trivial, so an unsupervised learning approach is highly desirable. Initial work of Huỳnh et al. [6,7] (using multiple Eigenspaces) and Nguyen et al. [8] (a modified HMM based approach) indicated that an unsupervised discovery of inherent structure in training data from inertal sensors is feasible. Even simple k-Means clustering already showed good results in grouping undefined activities [9]. As in general these inherent structures are not directly corresponding to the appropriate activities. Hence Huỳnh et al. [10] extended the k-Means clustering method with histogram based temporal smoothing and statistical classifiers (Nearest Neighbours and SVM) for assigning them to the appropriate activity, achieving recognition rates about 92% for three high level activities and compared them to a less successful HMM working on the raw sensor data.

3.2. Algorithmic Details

The individual raw sensor data channels are being synchronized and then processed in the feature extraction module. This module calculates 562 different features from half overlapping windows of 1.28s, consisting of frequency domain, statistical, curve, physical, correlation and step detection features (for a detailed explanation see [11]).

In this work we also utilize the k-Means clustering algorithm for autonomous identification and detection of simple base level motion patterns as it showed promising behaviour in prior work [9,10], is easy to implement and resource-friendly, which is important for mobile devices. As a comparison between supervised and unsupervised methods we also evaluate Naive Bayes, the C4.5 Decision Tree and a Support Vector Machine as statistical classifiers which all proved to be successful in terms of activity recognition.

Above this layer the Hidden Markov Model is used for the interpretation of the cluster affiliations. Every hidden state represents one high level activity and emits a cluster membership at a certain probability each timestep. Therefore the model keeps an emission probability table containing multinomial distributions for each state. The individual transition probabilities reflect the likeliness of a state transition and implicitly model the duration of the given activities through self referencing (loops). Together with the prior probabilities this unveils a major problem of HMMs - the huge number of parameters to set which normally is done initially and then refined using an Expectation Maximization algorithm.

In our approach we determine these parameters analytically from the training data in a preprocess. That typically produces an optimal and therefore overfitted model which represents the training examples best but does not generalize on unseen data. For this rea-

son we are using an ad-hoc regularization mechanism inspired by the *Regularized Discriminant Analysis* (see [12], pp. 90-91). Therefore we introduced regularization factors λ_{prior} for prior, λ_{trans} transition and λ_{obs} emission observation probabilities (Eq. 1 to 3 shows how final prior $P(S_t)$, state transition $P(S_t|S_{t-1})$ and observation probabilities $P(O|S_t)$ are obtained from weighted probabilities precalculated from the training data and equally distributed probabilities). This results in weighting factors between $\lambda = 0$ (equally distributed / no training data) and $\lambda = 1$ (perfect fit to training data) which allows a simple and intuitive manual control over the reliability of the model parameters given the individual uncertainty. This way the model can be made resistant against unsure, noisy and incomplete data.

$$P(S_t) = \lambda_{prior} \cdot P(S_t)_{calc} + (1 - \lambda_{prior}) \cdot P(S_t)_{equal} \tag{1}$$

$$P(S_t|S_{t-1}) = \lambda_{trans} \cdot P(S_t|S_{t-1})_{calc} + (1 - \lambda_{trans}) \cdot P(S_t|S_{t-1})_{equal} \tag{2}$$

$$P(O|S_t) = \lambda_{obs} \cdot P(O|S_t)_{calc} + (1 - \lambda_{obs}) \cdot P(O|S_t)_{equal} \tag{3}$$

Another advantage – not regarded in this paper, as the focus lies on IMU[1] tracking, but in prior work – is the possible inclusion of different sensor modalities on various time bases (e.g. event based), particularly RFID object detection (see [13,5,14,15]). As we suspect the state-conditional distributions of the observations to be non-gaussian and multimodal in the general case, we are relying on a sample based modelling approach. Therefore we use the HMM in conjunction with a particle filter [16].

4. Experimental Setup and Sensors

The following experiment was conducted to evaluate the feasibility of our unsupervised approach for recognizing sufficiently realistic health care activities. The chosen repertory only consists of compound activities at a high level of abstraction by trying to consider all of the challenges specified in section 2. As we were interested in a general proof of concept, we were not doing tests out of lab with authentic subjects inside a nursing home at this early stage. It is best practice to carry out initial experiments in a controllable environment under optimal observability. Therefore the test runs have additionally been accompanied by video and audio surveillance to facilitate later manual annotation of the ground truth.

For our setting we roughly rebuilt the floor plan of an apartment consisting of a bedroom, a bathroom, a living room and a kichenette in our laboratory. The test runs were performed by professional care personnel (a geriatric nurse) and a student who helped out as a patient. The scenery was observed by a fisheye and a ceiling mounted dome camera (Example still frames see Fig 2). A general preselection of care activities was given, as this is common for a care plan. The test agenda and the scenario have been developed in close cooperation with a nursing service, which also provided authentic equipment for the tests. We have sampled two runs of an authentic sequence of morning care activities taken from a real person. The activities were directly taken from the service accounting catalogue of the health insurances: "general service" (greeting, fetching newspaper, ...),

---

[1]Inertial Measurement Unit

"big morning toilet" (including washing whole body, brushing teeth), "micturition and defecation", "administration of medications", "bandaging", "preparation of food" and "documentation". We collected 14min (317mb) and 12 minutes (289mb) of raw data. Additionally the experimental environment was equipped with RFID tags to support object interaction detection, which is not in the focus of this paper. To avoid biasing, the test subjects were not involved in planning and setting up the experiment or analyzing the data afterwards in any way. Each subject was instructed to behave as natural as possible and to try to ignore the attached sensors.

We used three SparkFun IMU 6-DOF v3 sensor boards for our data collection. These are equipped with a 3-axis Freescale MMA7260Q accelerometer, 3-axis InvenSense IDG300 gyroscopes and a 2-axis Honeywell HMC1043 magnetometer. The LPC2138 ARM7 microcontroller is also capable of preprocessing the raw data onboard. We used the IMU to sample relative motion and rotation at a rate of 50Hz at a range of 6g to fully capture normal human motion as described in [17]. The raw data was instantaneously transmitted via a class 1 bluetooth link with a max. operating distance of 30 to 100m. Because of the compact size (51x41x23mm) the board could be attached at unobtrusive positiones: at the dominant wrist for recording gestures and object motion, at the chest/upper back and at the hip. These sensor positions have been shown to operate well in the literature and in own prior work [2,3,4].

For our initial tests we were recording raw data with a probably higher number of sensors than actually required for the final appliance, so that it is possible to evaluate single sensor channels or combinations of subsets afterwards on the original sensor data later. All data streams were wirelessly transmitted to a laptop computer where they were immediately formatted, synchronized and saved to disk.

5. Evaluation

Three different aspects were regarded during the evaluation of the clustering/model algorithm. At first the best fitting number of underlying clusters was determined, then the accuracy of the clustering approach was compared to conventional statistical classifiers using the same abstract model. At last the robustness was tested by applying a model trained on the first test run to the second. During this test optimal values for the model

Figure 2. Still frames taken from the surveillance cameras. Fisheye (left) and Dome (right). The test subject is equipped with sensors at the hip and upper back and dominant wrist.

regularization factors λ_{obs} and λ_{trans} were determined. For all experiments the decoding of the HMM was done using a particle filter with 100000 particles.

5.1. Determining the optimal number of clusters

The main idea behind the clustering approach is that all clusters represent single base level actions which are the building blocks of the abstract activities. These actions are not predefined, so the number of clusters/actions is an important parameter, as the k-Means clustering algorithm is not able to find an optimal number of clusters itself but needs a preset value for k. An optimal value will maximize the recognition accuracy of the whole model.

Finding the best value seems to be a typical task for the expectation maximization algorithm. In this case the EM based parameter optimization fails as it leads to massive overfitting. As too high values of k simply allow to memorize specific training examples which produces nearly perfect recognition rates, the EM algorithm simply increases k until nearly each training example is represented by a single cluster. Hence we decided to manually choose a value for k following the best practice of increasing k stepwise and choosing the first local maximum ("kink", see [12], pp. 461-472) after which the ascent of the recognition rate starts to flatten. This was done for both care test runs, resulting in an optimal value of $k = 35$ (Tab. 1). For both test runs the model regularization factors λ_{prior}, λ_{trans} and λ_{obs} were set to 1 which equates a perfectly fitting model with full trust in the sensor data and model probabilities to provide theoretically ideal test conditions as we were only interested in the underlying clustering parameters at this stage.The tests were conducted using a 10-fold stratified cross validation to ensure a realistic estimization of the generalization error. Confusion matrices for $k = 35$ are shown in Tab. 2. Notice that activity 6 ("documentation") was not carried out in the first test run.

							k							
	5	10	15	20	25	30	35	40	50	60	80	100	150	200
care1	65.4	67.2	77.4	88.5	94.0	97.1	**98.2**	93.3	94.2	98.2	98.9	98.2	98.4	98.7
care2	44.4	67.9	84.7	86.6	88.1	89.0	**96.0**	94.3	97.5	97.7	95.5	98.5	97.7	98.5

Table 1. Finding the optimal number of clusters: The table shows the recognition accuracy in % for our model with increasing number of clusters k for both test runs. λ_{prior}, λ_{trans} and λ_{obs} were set to 1 (perfect fit)

care1		estimate					
		0	1	2	3	4	5
	0	**23**	0	0	0	0	0
	1	4	**196**	0	0	0	0
truth	2	0	0	**32**	0	0	0
	3	0	0	1	**33**	0	0
	4	0	0	0	0	**144**	1
	5	0	0	0	0	2	**15**

care2		estimate						
		0	1	2	3	4	5	6
	0	**27**	0	0	0	0	0	0
	1	2	**228**	6	0	0	0	0
	2	0	5	**23**	0	0	0	0
truth	3	0	0	2	**34**	0	0	0
	4	1	0	0	2	**121**	0	0
	5	1	0	0	0	0	**13**	0
	6	0	0	0	0	0	0	**6**

Table 2. Confusion matrices for both test runs at a value of $k = 35$. Different activity classes are *1* "general service" (greeting, fetching newspaper, ...), 1 "big morning toilet" (including washing whole body, brushing teeth), 2 "micturition and defecation", 3 "administration of medications", 4 "bandaging", 5 "preparation of food" and 6 "documentation"

5.2. Comparison of Supervised and Unsupervised Approaches

After estimating the performance of the unsupervised clustering approach, we were interested in how it compares to other conventional methods used in own and related prior work. Therefore we simply replaced the k-Means Clusterer with several relevant supervised learners while keeping the model. The same training data was learned by statistical classifiers regarding the given annotated activity classes. Then the model parameters were set equally to the unsupervised approach, again with λ_{prior}, λ_{trans} and λ_{obs} set to 1, and again cross validated. This proceeding assures direct comparability of the underlying learning algorithms. As statistical classifiers we chose the C4.5 decision tree, a Support Vector Machine (SVM) and Naive Bayes (NB). Table 3 shows the recognition results of all four methods for both test runs. For the k-Means k was set to 35 (see above).

	care1	care2
C4.5	86.9	86.6
Support Vector Machine	85.8	85.6
Naive Bayes	77.2	70.9
k-Means ($k = 35$)	**98.2**	**96.0**

Table 3. Supervised/Unsupervised comparison results: The table shows the specific recognition accuracies in % for both test runs. λ_{prior}, λ_{trans} and λ_{obs} were again set to 1 (perfect fit).

For both test runs C4.5 and SVM showed a comparable recognition accuracy about 96% while Naive Bayes results were settled at 77% / 71% respectively. The best recognition rates were achieved by the embedded clusterer by far with a recognition rate around 98% / 96%.

The detailed results of this comparison on a continuous time trace are illustrated in Fig. 3. The left side shows the outputs of the first layer which are the cluster affiliations (k-Means) or class memberships (statistical classifiers). It can be clearly seen that the $k = 35$ clusters do not correspond to the anticipated activity classes, while the output of the C4.5, SVM and Naive Bayes represents the actual activities although being very noisy. On the right the final outputs of the models are shown in comparison to the ground truth.

5.3. Robustness Experiments

As the recognition accuracy of a model trained and evaluated on a single experiment does not give any information regarding the robustness and generalized performance on completely unseen activity sequences we estimated the general recognition rate with a clusterer/model trained on the first test run and applied it to the second. This method is rather brute force and can just give a hint on the real generalization performance due to the lack of sufficient training data. For getting a more realistic estimation it would be necessary to record a higher number of test runs and evaluate the recognition system using leave-one-out cross validation.

For estimating the achievable performance we had to determine the optimal regularization factors λ_{prior}, λ_{trans} and λ_{obs}, as a model 100% adapted to one test run is unable to explain a differing second one. Factor λ_{prior} was left constant because the prior probabilities in both test cases were the same, so a variation did not have any effect.
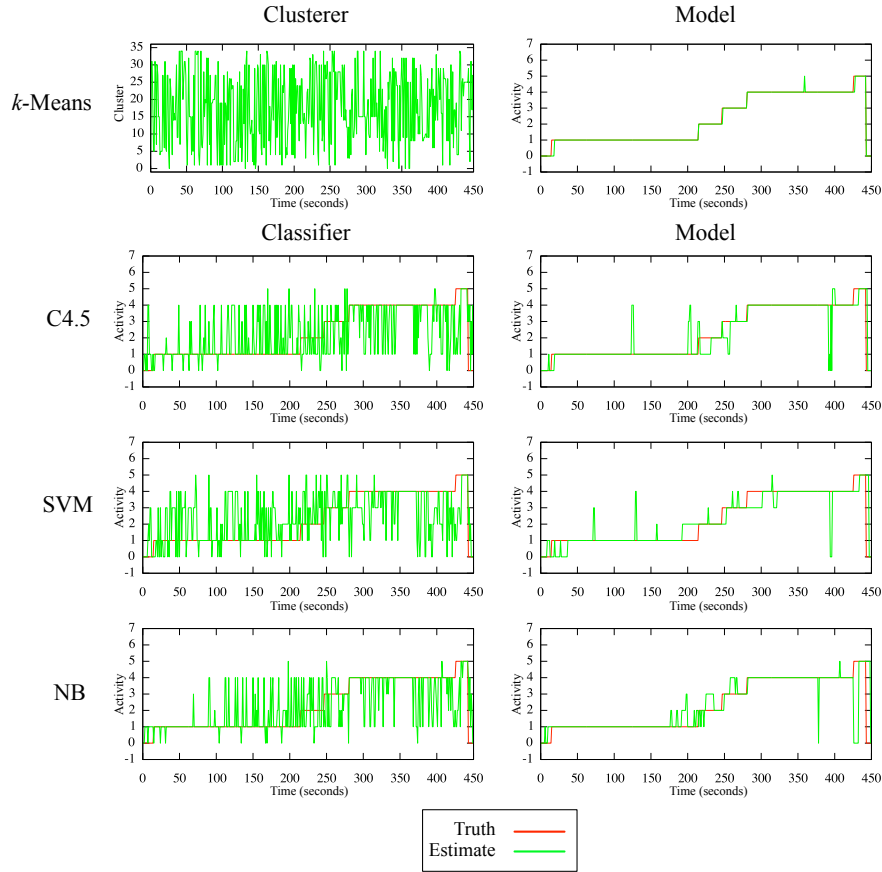
Figure 3. Supervised/Unsupervised comparison results for the first test run: The left side shows the outputs of the first layer (cluster affiliations and class memberships) at a continuous time trace. The diagrams on the right side show the individual activity trajectory recognized by the overlying model compared to the ground truth.

The regularization factors λ_{trans} (representing the trust in the transition probabilities of the model) and λ_{obs} (representing the trust in the cluster assignments) were gradually altered. The size of the steps was decreased in regions where the changes in performance were more significant. $\lambda_{trans} = 1.00$ was not applicable in this test as the model did not allow any temporal changes in the activity trajectory of the test run, which causes the particle filter to make all particles die immediately at the first time the observations don't fit the states anymore.

The actual recognition results are illustrated in Tab. 4 and Fig. 4. The best performance was achieved for $\lambda_{trans} = 0.99$ and $\lambda_{obs} = 0.05$. We were surprised by the relatively high recognition rate of about 75%. As we observed a very high variance in the performance of the test subject we were expecting a very low accuracy. A closer look at the regularization factors shows that recognition results are significantly better when the trust in the model state transition is very high and far away from equally distributed. The optimal trust in the cluster affiliation observation is relatively low and settled around 5%
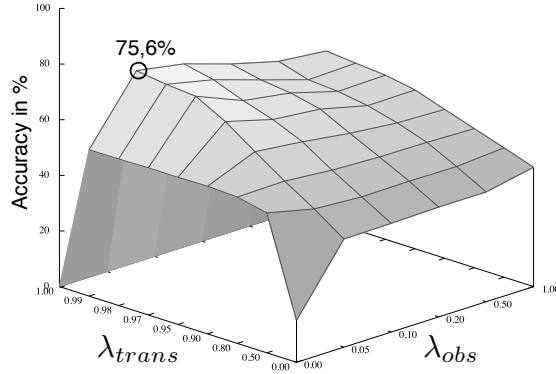
Figure 4. Robustness experiments: The recognition accuracies for our model trained on the first test run and applied on the data of the second in relation to λ_{trans} and λ_{obs}.

		λ_{trans}							
		0.99	0.98	0.97	0.95	0.90	0.80	0.50	0.00
	1.00	60.9	59.9	59.0	57.1	53.5	50.0	46.5	42.9
	0.50	62.6	59.4	58.6	56.9	54.1	50.5	45.2	39.5
λ_{obs}	0.20	67.1	64.8	63.5	61.1	55.8	50.7	45.4	39.5
	0.10	72.6	70.3	66.2	62.8	58.2	51.4	45.2	39.1
	0.05	**75.6**	74.1	73.2	67.5	60.1	52.9	46.3	38.9
	0.00	52.6	52.7	52.9	52.9	53.1	52.6	50.3	15.1

Table 4. Robustness experiments: The detailed recognition accuracies in % for our model trained on the first test run and applied on the data of the second in relation to λ_{trans} and λ_{obs}.

which indicated two things: First, the implicit knowledge saved in the generative model is of superior importance for the robust recognition of abstract activities. Second, the variability in the execution of the two test runs was very high, resulting in highly differentiating cluster affiliation probabilities for the activities of both test runs which again could be expected due to the high variability in execution. Although the trust in cluster observations is relatively low the clustering is an essential step. Letting λ_{obs} converge to zero the recognition rate rapidly falls off with an activity trajectory ending in a flat line or complete chaos respectively.

Besides the high executional variety, the diagram (Fig. 5) and the confusion matrix (Tab. 5) reveal some further problems: First, the test subject did not make any documentation in the first test run, so the model was not trained on this activity and unable to detect it. Therefore it interpreted "documentation" (activity 6) as "general service" (activity 1). Second: Some activities just had a very small number of training examples, which itself already causes some fundamental problems for machine learning algorithms, especially in a high dimensional feature space, regardless the applied learning method.

If we again compare the clustering approach to the supervised methods the clustering method significantly outperforms the conventional methods by more than 40% (Tab. 6).
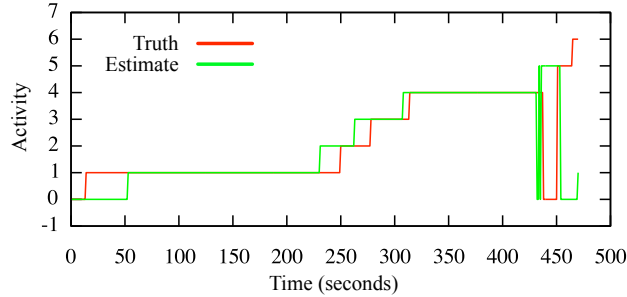
111

Figure 5. Robustness experiments: A continuous time trace of the activities recognized by the model compared to the ground truth. For this detailed view the model was trained on the first test run and applied on the second, parameters were set to $k = 35$, $\lambda_{prior} = 1$ $\lambda_{trans} = 0.99$ and $\lambda_{obs} = 0.05$.

		estimate					
		0	1	2	3	4	5
	0	**14**	0	0	0	0	13
	1	39	**178**	19	0	0	0
	2	0	0	**13**	15	0	0
truth	3	0	0	0	**30**	6	0
	4	3	0	0	0	**118**	3
	5	11	0	0	0	0	**3**
	6	5	1	0	0	0	0

Table 5. Confusion matrix for the model trained on the first test run and applied on the second, parameters were set to $k = 35$, $\lambda_{prior} = 1$ $\lambda_{trans} = 0.99$ and $\lambda_{obs} = 0.05$. Different activity classes are 1 "general service" (greeting, fetching newspaper, ...), 1 "big morning toilet" (including washing whole body, brushing teeth), 2 "micturition and defecation", 3 "administration of medications", 4 "bandaging", 5 "preparation of food" and 6 "documentation".

A surprising exception seems to be the Naive Bayes Classifier with a recognition rate of more than 80%. Unfortunately a closer look unveils the "fraud". The output classes of the classifier caused the HMM to simply detect the two most common activities ("big morning toilet" and "bandaging") which make about 80% of the recorded data and ignore the rest. Hence this approach is completely unusable, although the recognition rate seems exceptionally high.

	care1 → care2
C4.5	29.3
Support Vector Machine	31.0
Naive Bayes	**80.9**
k-Means ($k = 35$)	**75.6**

Table 6. Clusterer/Classifier robustness comparison results: The recognition accuracy in % for the Clusterer/Classifiers/Models trained on the first test run and applied on the data of the second run. ($\lambda_{trans} = 0.99$, $\lambda_{obs} = 0.05$)

112

6. Conclusion

The major goal of this paper was to evaluate the feasibility of a simple clustering based method for the recognition of compound activities at a high level of abstraction. As one of the main difficulties in detecting this type of activities are similar motion patterns shared between multiple activities we proceeded to aggregate these activities from shorter, easier to detect base level actions. As the explicit annotation of these is not trivial and very time consuming we were looking for an unsupervised learning algorithm. A general problem of these methods is that the resulting cluster affiliations are typically not human readable and some kind of interpretation is needed. To achieve this, we developed a hybrid method using a generative probabilistic model built on top of the clusterer. We adapted a Hidden Markov Model for mapping the cluster memberships onto high level activities and evaluated this technique using experimental data from two test runs of a home care scenario. A welcome side effect of this type of model is the ability to also include knowledge about temporal and causal dependencies.

It could be shown that the hybrid k-Means/HMM approach outperforms classical hybrid discriminative/generative methods built upon statistical classifiers significantly in terms of accuracy and robustness. It was found that best overall results could be obtained at a number of about 35 clusters which shows that already about 35 base level actions are able to disambiguate the anticipated high level activities successfully. When using the trained model on a completely unseen test run the recognition accuracy noticably drops from about 96% to 75%, which is still much above random guessing (14%) and above the conventional approaches (31%, Naive Bayes' impressive "bogus" results disregarded). We expect this value to be better in future experiments, as we were relying on very few training data which overemphasized the variability between both test runs. The bad recognition results of the statistical classifiers suggest that conventional methods trying to handle compound activities in a monolithic manner are inappropriate in principle as they do not generalize well.

As k-Means is a very simple algorithm we do not expect it to be the method of choice in general, other clusterers may be even more successful. Probably also supervised learning methods can outperform this simple approach, provided that the appropriate base level actions have been annotated in advance, but as mentioned before this would not be practicable.

We also proposed a simple method for automatically determining the necessary model parameters without time consuming learning under consideration of a few intuitive regularization factors comparable to laplacian correction.

Due to the limited amount of data our results can just be seen as initial steps towards an approach for recognizing abstract high level activities in a manner requiring much less supervision. The next logical step will be collecting a much larger and more realistic out-of-lab dataset recorded during the day to day work inside a nursing home.

We noticed one general disadvantage of the chosen HMM during the evaluation. As the state transition probabilities include both the probability of switching to the next state and the average duration of the current state, the interpretation of this transition matrix is somehow inconvenient and not very handy. Hence we have already started to modify our model to explicitly incorporate state durations disentangled from state switching probabilities. Another thing we are currently investigating is the integration of partial order task models into the modelling process.

Acknowledgements

References

[1] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, "A hybrid discriminative/generative approach for modeling human activities," in *IJCAI* (L. P. Kaelbling and A. Saffiotti, eds.), pp. 766–772, Professional Book Center, 2005.

[2] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, "Activity classification using realistic data from wearable sensors," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 10, no. 1, pp. 119–128, Jan. 2006.

[3] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *AAAI* (M. M. Veloso and S. Kambhampati, eds.), pp. 1541–1546, AAAI Press / The MIT Press, 2005.

[4] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive* (A. Ferscha and F. Mattern, eds.), vol. 3001 of *Lecture Notes in Computer Science*, pp. 1–17, Springer, 2004.

[5] S. Wang, W. Pentney, A.-M. Popescu, T. Choudhury, and M. Philipose, "Common sense based joint training of human activity recognizers," in *IJCAI* (M. M. Veloso, ed.), pp. 2237–2242, 2007.

[6] T. Huynh and B. Schiele, "Unsupervised discovery of structure in activity data using multiple eigenspaces," in *2nd International Workshop on Location- and Context-Awareness (LoCA)*, (Dublin, Ireland), Springer, May 2006.

[7] T. Huynh and B. Schiele, "Towards less supervision in activity recognition form wearable sensors," in *Proceedings of the 10th IEEE International Symposium on Wearable Computing (ISWC)*, (Montreux, Switzerland), October 2006.

[8] A. Nguyen, D. Moore, and I. McCowan, "Unsupervised clustering of free-living human activities using ambulatory accelerometry," *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 4895–4898, 22-26 Aug. 2007.

[9] T. Huynh and B. Schiele, "Analyzing features for activity recognition," in *Proceedings of the 2005 joint conference on Smart objects and ambient intelligence: innovative context-aware services: usages and technologies*, (Grenoble, France), pp. 159–163, ACM Press New York, NY, USA, 2005.

[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)*, 2007.

[11] A. Hein, "Echtzeitfähige merkmalsgewinnung von beschleunigungswerten und klassifikation von zyklischen bewegungen," Master's thesis, University of Rostock, 11 2007.

[12] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*. Springer, August 2001.

[13] D. J. Patterson, D. Fox, H. Kautz, and M. Philipose, "Sporadic state estimation for general activity inference," tech. rep., University of Washington and Intel Research Seattle, July 2004.

[14] D. J. Patterson, D. Fox, H. Kautz, and M. Philipose, "Fine-grained activity recognition by aggregating abstract object usage," *iswc*, vol. 0, pp. 44–51, 2005.

[15] A. Hein and T. Kirste, "Activity recognition for ambient assisted living: Potential and challenges," in *Ambient Assisted Living Ambient Assisted Living Ambient Assisted Living*, pp. 263–268, VDE Verlag, 01 2008.

[16] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing, IEEE Transactions on]*, vol. 50, no. 2, pp. 174–188, Feb 2002.

[17] C. Bouten, K. Koekkoek, M. Verduin, R. Kodde, and J. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *Biomedical Engineering, IEEE Transactions on*, vol. 44, no. 3, pp. 136–147, Mar 1997.

[18] "Landesforschongsschwerpunkt http://marika.lfs-ma.de/," April 2008.