Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence

Feature Learning for Activity Recognition in Ubiquitous Computing

Thomas Plötz, Nils Y. Hammerla, and Patrick Olivier
Culture Lab
School of Computing Science
Newcastle University
Newcastle upon Tyne, UK
{t.ploetz,n.y.hammerla,p.l.olivier}@newcastle.ac.uk

Abstract

Feature extraction for activity recognition in context-aware ubiquitous computing applications is usually a heuristic process, informed by underlying domain knowledge. Relying on such explicit knowledge is problematic when aiming to generalize across different application domains. We investigate the potential of recent machine learning methods for discovering universal features for context-aware applications of activity recognition. We also describe an alternative data representation based on the empirical cumulative distribution function of the raw data, which effectively abstracts from absolute values. Experiments on accelerometer data from four publicly available activity recognition datasets demonstrate the significant potential of our approach to address both contemporary activity recognition tasks and next generation problems such as skill assessment and the detection of novel activities.

1 Introduction

Activity recognition (AR) is a core concern of the ubiquitous computing (ubicomp) community [Atallah and Yang, 2009] and plays a central role in the field's vision of context-aware applications and interaction. In general, sensors, which are either worn on the body and/or embedded into objects and the environment are utilized to capture aspects of movement or a user's behavior. Ideally, by applying signal processing and pattern classification techniques, this sensor data can be automatically analyzed yielding a real-time classification of the activities that users are engaged in.

Activity recognition is a classical (multi-variate) time-series or sequence analysis problem, for which the task is to detect and classify those contiguous portions of sensor data streams that cover activities of interest for the target application. The predominant approach to AR is based on a sliding window procedure, where a fixed length analysis window is shifted along the signal sequence for frame extraction. Consecutive frames usually overlap to some degree but are processed separately. Preprocessing then transforms raw signal data into feature vectors, which are subjected to statistical classifiers that eventually provide activity hypotheses.

As for any pattern recognition task, the keys to successful AR are: (i) appropriately designed feature representations of the sensor data; and (ii) the design of suitable classifiers. The ubicomp literature describes a wide variety of creatively applied classification approaches. By contrast, comparatively little systematic research has addressed the problem of feature design, with almost all previous work using heuristically selected general measures. These features are either calculated in the time domain, calculated on symbolic representations of the sensor data, or spectra based. The lack of systematic research on features has been identified as one of the major shortcomings of current AR systems [Lukowicz *et al.*, 2010]. For example, it is questionable whether the next generation of applications, such as behavioral analysis, or skill assessment can be realized based on the use of heuristically selected features alone. Such problems require quantitative analyzes of the underlying data which are beyond the capabilities of current procedures for discriminating within limited sets of activities and rejecting unknown samples.

The most straightforward approach to feature design is to investigate the nature of the data to be analyzed and to develop a representation that explicitly captures its core characteristics. For ubicomp AR problems, no all-encompassing model exists to afford the expert-driven design of a universal feature representation. However, recent developments in the general machine learning field have the potential to overcome this shortcoming by automatically discovering universal feature representations for such ubicomp sensor data.

We present a general approach to feature extraction and investigate the suitability of feature learning for ubicomp activity recognition tasks. We utilize a learning framework, which automatically discovers suitable feature representations that do not rely on application-specific expert knowledge. We use unsupervised feature learning techniques, namely (variants of) principal component analysis and deep learning, and show how the automatically extracted features outperform standard features across a range of AR applications. Such an automatic feature extraction procedure has important implications for the development future applications since no manual optimization is required. The deep learning approach allows for in-depth analysis of the underlying data since the new representation implicitly highlights the most informative portions of the analyzed data. This is likely to be important for new classes of activity analysis such as skill assessment.

2 State-of-the-Art

A recent survey of preprocessing techniques for AR [Figo *et al.*, 2010] distinguished the principal classes of calculation scheme according to the domain of the preprocessing: (i) time domain; and (ii) the frequency domain. The most widely used feature extraction scheme calculates statistical metrics directly on the raw sensor data, independently for every frame extracted by a sliding window procedure. Commonly used metrics include the mean, standard deviation, energy, entropy, and correlation coefficients. Feature extraction in the frequency domain is usually based on Fourier coefficients calculated for the analysis frames. Huynh and Schiele conducted an experimental evaluation of the capabilities of feature representations, namely statistical metrics and Fourier coefficients [Huynh and Schiele, 2005]. They concluded that Fourier coefficient based representations are more appropriate than statistical metrics.

Whereas the majority of published work utilizes standard features a small number of alternative approaches have been proposed. Recently, time-delay embeddings have been used for activity and gait recognition [Frank *et al.*, 2010]. Time-delay embedding is a technique borrowed from physics, where it is used to describe the state of complex systems by means of phase space analysis. This novel representation of sensor data has proved as significant utility in the analysis of repetitive (i.e periodic or quasi-periodic) activities. However, classifiers based on time-delay embedding representations are less appropriate for non-periodic activities. Another emerging approach is to use discrete domain features and to calculate distance measures on string representations of the sensor data, which has a particular relevance for activity discovery applications (e.g. [Minnen *et al.*, 2006]). However, the quantization of the sensor data required removes detailed information that is important for the in-depth analysis of certain activities of interest.

3 Feature Learning for Activity Recognition

Feature learning is a well-studied approach for static data (e.g., object recognition in computer vision). The goal is to automatically discover meaningful representations of data to be analyzed. Contrary to heuristic feature design, where domain specific expert knowledge is exploited to manually specify features, feature learning seeks to optimize an objective function that captures the appropriateness of the features. Standard approaches include energy minimization [LeCun *et al.*, 2006], manifold learning [Huo *et al.*, 2004], and deep learning using auto-encoders [Hinton, 2007].

We have developed a feature extraction framework for sequential data based on feature learning, which is integrated into a general activity recognition work-flow (Fig. 1). A sliding window procedure extracts overlapping, fixed length frames from continuous sensor data streams, which in our experiments were the x, y, z data values for tri-axial accelerometers (upper left part of Fig. 1). Frames extracted from raw data are used to estimate the parameters of the actual feature learning procedure (see "fex" block in Fig. 1). This feature extractor is then used to transform raw sensor data to be analyzed by the application.

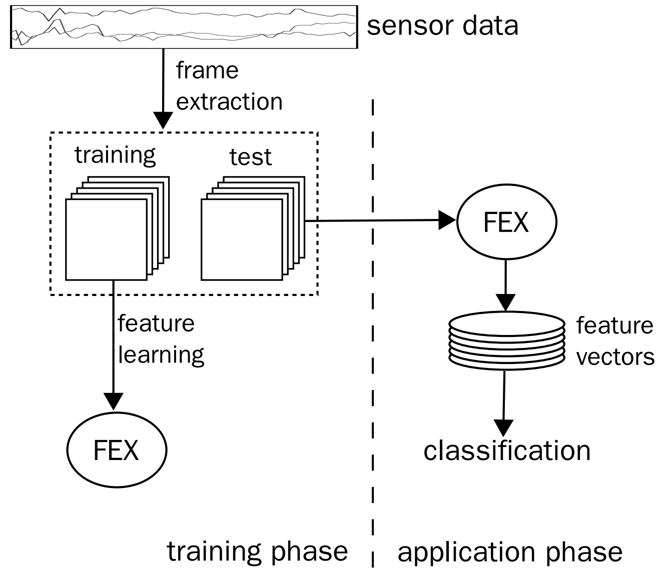

Figure 1: Feature learning for activity recognition – overview.

Our design criteria for feature learning ("fex" in Fig. 1) are as follows:

1. Capable of extracting generally applicable representations – not be limited to specific AR tasks.
2. Must not rely on the availability of ground truth annotations of the training data.
3. Benefits from larger datasets, but not dependent on them.
4. Provides intrinsic information (for sub-frame analysis).
5. Must be computationally feasible and applicable in real-time application contexts.

Given these design requirements we focused on two learning techniques: PCA and auto-encoder based deep learning.

3.1 PCA based Feature Learning

PCA is a well established technique used for decorrelation and dimensionality reduction of data. PCA is a basic form of feature learning since it automatically discovers compact and meaningful representations of raw data without relying on domain specific (or expert) knowledge. The eigenvectors of a sample set's covariance, which correspond to its largest eigenvalues, are utilized to span a lower-dimensional sub-space that concentrates the variance of the original data. The projection of the original data onto the variance-maximizing sub-space serves as a feature representation and can be used either for visualization or fed into a subsequent classifier. Automatic analysis of the eigenvalue spectrum of the sample covariance uncovers the appropriate target-dimensionality of the feature space.

ECDF-based sensor data representation It is well known that PCA performs poorly if the input data are not properly normalized. Unfortunately, blind range normalization often introduces more problems when components relate to completely different aspects of a phenomena; in the context of AR this becomes problematic when large frame-sizes are used. To address this issue we developed an alternative raw data representation based on the empirical cumulative distribution function (ECDF) of the sample data. The idea is to derive a representation of the input data, which is independent of the absolute ranges but preserves structural information. This representation is inspired by approaches used in other application domains of time-series analysis, e.g., bioinformatics [Chou, 1995].

For every frame $\vec{f} = (\vec{x}_i)^T$ we derive the empirical cumulative distribution functions ECDF $\mathcal{E}_i$ of the (whitened) samples $\vec{x}_i$ along each axis $i = \{1, 2, 3\}$ using standard Kaplan-Meier estimation [Cox and Oakes, 1984]. These ECDFs, which monotonically increase within the range of zero to one, describe the probability that the sensor readings are less than or equal to some specific value. By means of a cubic interpolation $\mathcal{C}^{\vec{p}}$ we estimate the values of the inverse of the ECDF $\mathcal{E}_i^{-1}$ at a fixed set of N points $\vec{p} = \{p_1 \dots p_N\}$, which serve as the representation $\vec{x}'_i$ of the sensor data. Using this procedure, sample data are normalized to a common range without destroying inherent structural dependencies (for $i = \{1, 2, 3\}$):

$$\begin{aligned} \vec{x}'_i &= \mathcal{C}^{\vec{p}}\left(\mathcal{E}_i^{-1}(x_i)\right) \in \mathbb{R}^N && (1) \\ \text{with } \mathcal{E}_i &= x_i \mapsto F_{X_i}(x_i) = P(X_i \leq x_i) && (2) \\ \text{and } \mathcal{C}^{\vec{p}} &= \text{cubic interpolation using } \vec{p} = \{p_1 \dots p_N\} \end{aligned}$$

3.2 Deep Learning for Feature Extraction

Autoencoder networks have proved to be a powerful tool for the generic semi-supervised discovery of features [Hinton, 2007]. These aim to learn a lower-dimensional representation of input data, which produces a minimal error when used for reconstructing the original data. As an alternative to PCA based feature extraction for continuous sensor streams we employed deep learning methods for autoencoder based feature learning on sequential data. The desired representation is discovered by means of a feed-forward neural network that consists of one input layer, one output layer and an odd number of hidden layers. Every layer is fully connected to the adjacent layers and a non-linear activation function is used. The objective function during training is the reconstruction of the input data at the output layer. The autoencoder transmits a description of the input-data across each layer of the network. Since the innermost layer of the network has a lower dimensionality, the transmission of a description through this bottleneck can only be achieved as result of a meaningful encoding of the input. This non-linear low-dimensional encoding is hence an automatically learned feature representation.

For robust model training we follow the suggestions given in [Hinton *et al.*, 2006], i.e., we learn the layers of the autoencoder network greedily in a bottom-up procedure, by treating each pair of subsequent layers in the encoder as a Restricted Boltzmann Machine (RBM). An RBM is a fully connected, bipartite, two-layer graphical model, which is able to generatively model data. It trains a set of stochastic binary hidden units which effectively act as low-level feature detectors. One RBM is trained for each pair of subsequent layers by treating the activation probabilities of the feature detectors of one RBM as input-data for the next. Once the stack of RBMs is trained, the generative model is unrolled to obtain our final fully initialized autoencoder network for feature learning.

Different methods exist to model real-valued input units in RBMs. We employ Gaussian visible units for the first level RBM that activate binary, stochastic feature detectors (Gaussian-binary). The subsequent layers can then rely on the common binary-binary RBM. The final layer is a binary-linear RBM, which effectively performs a linear projection.

During training the sample data is processed batch-wise, where each batch ideally comprises samples from all classes in the training-set. Note that the availability of the class information is not mandatory. RBMs can also be trained in a completely unsupervised manner. However, balancing the batches with respect to the distribution of the classes, i.e. performing semi-supervised training, improves the model quality since it removes the potential for artificial biases.

4 Experimental Evaluation

To evaluate the effectiveness of feature learning for AR we conducted a number of experiments using published datasets that compared the proposed approach to state-of-the-art heuristically selected features. Sensor data was analyzed by means of a (previously optimized) sliding window procedure, extracting frames of $n = 64$ contiguous samples, which overlap by $p = 50$ percent. Feature extraction was then performed on a frame-by-frame basis. The focus of our evaluation was on the capabilities of the particular feature representations. Accordingly, we did not focus on classifier optimization but on the features themselves. In accordance with the state-of-the-art in ubicomp AR we selected a standard, instance-based classification approach, Nearest Neighbor (NN), and applied it "as is" to all tasks.

Given ground truth annotations we report the classification accuracy as percentages of correct predictions provided by the NN-classifiers. The experiments were performed as $N =$ 10-fold cross validations (unless mentioned otherwise). Folds were created by randomly choosing samples from the original dataset thereby respecting fold-wise balanced distributions of all classes (i.e. activities to be recognized).

4.1 Datasets

We selected four standard datasets for our evaluation, each of which is described in the literature and is publicly available. All datasets relate to human activities in different contexts and have been recorded using tri-axial accelerometers. Sensors were either worn or embedded into objects that subjects manipulated.

Ambient Kitchen 1.0 (AK) Pham et al. [Pham and Olivier, 2009] describe a dataset in which twenty participants prepared either a sandwich or a salad using sensor-equipped kitchen utensils. Modified Wii-controllers were integrated into the handles of knives, spoons and scoops, serving as a sensing platform for continuous recording of tri-axial acceleration data. In total the dataset comprises almost 4 hours of sensor data, approximately 50% of which cover ten typical food preparation activities. Given the sampling frequency of 40Hz, the sliding window procedure produced almost 55,000 frames.

Darmstadt Daily Routines (DA) In [Huynh *et al.*, 2008] the analysis of activities of daily living (ADL) is addressed by means of worn sensors used to monitor the daily activities of individual subjects in a living lab-like experiment. Two tri-axial accelerometers (wrist-worn and carried in the pocket) recorded movements at 100Hz. Preprocessing and subsampling yields an overall sampling frequency of 2.5Hz. In total more than 24,000 frames were extracted for both the wrist-worn and pocket-carried sensors using our sliding window procedure. Ground truth annotation used 35 activities of different levels of abstraction. Cross-validation experiments were conducted based on class-wise balanced, random selection of frames for creating the folds. We report results only for pocket-sensor experiments, which, as reported in the original publication, yielded significantly better results than those based on the wrist-worn sensor data.

Skoda Mini Checkpoint (Skoda) [Zappi *et al.*, 2008] describe the problem of recognizing activities of assembly-line workers in a car production environment. In the study a worker wore a number of accelerometers while undertaking manual quality checks for correct assembly of parts in newly constructed cars (10 manipulative gestures of interest). We restrict our experiments to a single sensor, which is sufficient to identify all 10 activities (i.e. right arm). In total the dataset comprises 3 hours of recordings from one subject (sampled at 96Hz resulting in 22,000 frames). As a result of the unequal distribution of the samples across the classes we were only able to perform 4-fold cross evaluation.

Opportunity – Preview (Opp) The final dataset relates to a home environment (kitchen) and the analysis of ADL using multiple worn and embedded sensors [Roggen *et al.*, 2010]. Although the activities of multiple subjects, on different days have been recorded, an official excerpt of annotated data for a single subject has recently been released. Our analysis was based on the sensor data recorded by the accelerometer attached to the right arm of the subject. We considered 10 low-level activities of interest plus an *unknown* activity category. The acceleration data were sampled with 64Hz yielding approximately 4,200 frames.

4.2 Features Analyzed: Overview

To analyze the performance of learned features for activity recognition we performed classification experiments that compared the capabilities of state-of-the-art representations of sensor data streams and learned features as already discussed. To allow comparison of the resulting feature representations we ensured that the target dimensionality of each was in approximately the same range. Since we used instance-based classifiers there was no requirement to use *identical* dimensionalities for objective comparisons. This stands in contrast to generative models (such as mixture densities) where small differences in the dimensionality of the underlying data can have a significant impact on the estimation procedure and hence on the capabilities of the models.

Statistical Metrics Probably the most common approach to feature extraction for activity recognition is to use a set of statistical measures to represent frames of contiguous multi-dimensional sensor data. Given the 192-dimensional analysis frames (64×3) provided by our sliding window procedure, we first calculated pitch and roll values. Subsequently, for each source channel (i.e. x, y, z, pitch, and roll) we then calculated mean, standard deviation, energy, and entropy. Together with three correlation coefficients (estimated for all combinations of the x, y, z axes) this yielded a 23-D representation of the raw signal data covered by an analysis frame.

FFT coefficients Characteristic differences in certain activities are apparent from changes in the particular spectra and consequently we can apply frequency transformations to extract feature representations for such classes of activity recognition problems. We performed a channel-wise Fourier analysis on the raw signal data of an analysis frame. Given the resulting spectra we selected the first f coefficients per channel (x, y, z) and concatenated these into a single feature vector. For our experiments we evaluated different choices of f. For our dimensionality range (23-39), differences in classification accuracy were negligible – for succinctness we only report the results for $f = 10$ (target dimensionality of 30).

PCA We performed experiments utilizing PCA-based features where the projection sub-space is spanned by those eigenvectors that correspond to the c = 18, 23, 30, and 39 largest eigenvectors. These selections of c are justified by significant drops in the eigenvalue spectrum of the data and correspond to the selected target dimensionalities of the other approaches investigated. No significant changes in classification accuracy were observed for the four choices of c, hence we present the results for $c = 30$. Experiments were performed both for the raw sensor data and for the ECDF-based representation. Note that kernel PCA based approaches were ruled out for our unsupervised feature extraction approach due to their exorbitant turnaround times during training.

Deep Belief Networks Auto-encoder networks contain a number of free parameters, including the network topology, i.e., the number of internal layers and its dimensionalities. To show the general applicability of the method, the learning parameters and the network layout (one for the raw data, and one for the ECDF-representation) were tuned on the AK dataset via cross-validation and then used as is for the remaining tasks. The optimized network layout consists of a 4-layer model with 1024 units in each hidden layer and 30 units in the top one (192-1024-1024-30). In all experiments, the first layer was trained for 100 epochs while the subsequent layers were trained for 50 epochs. For the DA dataset, which incorporates a large number of classes (35), the distribution of samples in each batch corresponds to that of the training set, while for the other sets each batch is split equally among all classes, holding 10 samples for each.

4.3 Results

Classification accuracy The first set of experiments was devoted to the evaluation of the classification performance as it can be achieved when using the particular feature representations. Fig. 2 presents the results for the four analyzed datasets. Contrasting our results with those already published for these datasets, we found our results to be broadly comparable (accuracies between 74% and 90%). Interestingly, traditional statistical features performed rather poorly on the Skoda and the Opportunity datasets.

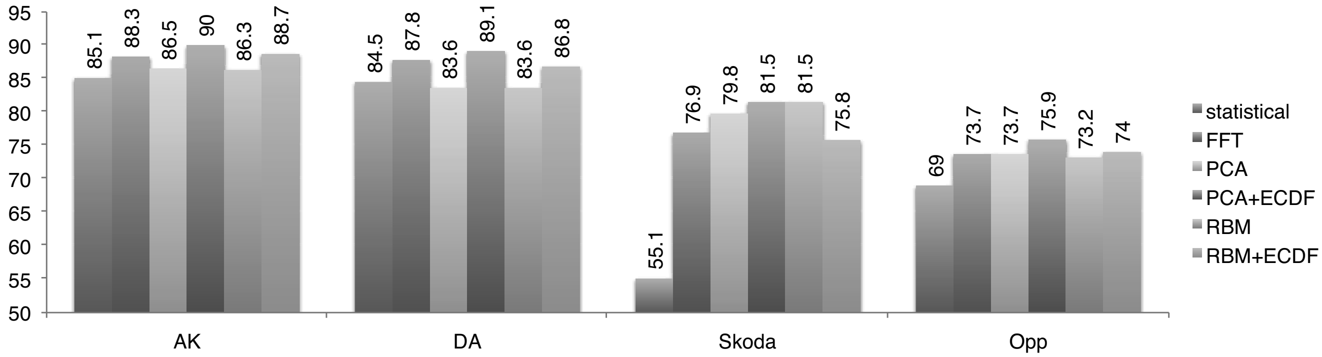

Figure 2: Classification results for experimental evaluation of learned features and heuristically chosen metrics.

Both variants of learned features lead to statistically significant improvements of the classification accuracy (95% confidence) for all datasets analyzed. These improvements on statistical features and FFT based representations are meaningful, especially when we consider that the feature representations have been learned automatically without relying on domain-specific expert knowledge. The results also demonstrate that our feature learning approach greatly benefits from the ECDF-based representation of the input data which yielded significant improvements in classification accuracy for the majority of cases.[1]

In summary, both learning techniques can be used across different AR tasks to discover compact and meaningful feature representations which outperform classical approaches. Features are discovered in an unsupervised manner. For optimization of the deep learning approach prior knowledge about the underlying distributions of the classes is exploited, resulting in a semi-supervised approach.

Influence of Sample Set Size The second set of experiments addressed the sparse data problem. Feature learning relies on the availability of sufficient quantities of sample data. The construction of the projection sub-space for the PCA procedure relies on a statistically robust analysis of the sample set covariance. For small datasets the empirical estimation of covariance matrices can result in singularities, which undermines the sub-space creation. Estimating parameters of the auto-encoders for the second learning approach also relies on a representative sample set. Non-representative sample sets bias the parameter estimation procedure such that the resulting features are not flexible enough to capture unknown data.

We evaluated classification accuracies which can be achieved when the training sets used for estimating the feature extraction procedures are artificially limited. Given the original N-fold cross validation procedures we gradually removed samples from the training set, performed feature learning as before, and ran classification experiments. Fig. 3 illustrates the dependency of the classification results on the amount of sample data available for training the feature extractors. For comparability the x-axis indicates fractions of the original

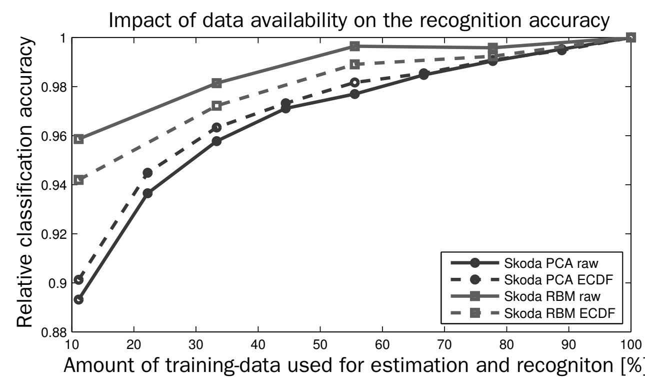

Figure 3: Exemplary evaluation of sparse data problem.

dataset and the y-axis indicates the relative changes in classification accuracy. We ran the evaluation for all four datasets but for the sake of clarity we limit our presentation to the results achieved for the Skoda dataset which is representative of the others (for which similar results were achieved). From the results (Fig. 3) it is clear that the size of the sample set does not substantially influence the capabilities of the resulting classifiers. However, it seems that PCA has a stronger reliance on the quantity of available training-data compared to RBMs. Given the results of the second set of experiments we can conclude that feature learning meets the third design criteria for practical AR applications.

Further Analysis The learned representations can be used for in-depth analysis of the underlying sensor data (the fourth criteria as described in section 3). For example, a frame-wise analysis of the reconstruction error provides insights into the quality of the performed activity. Beyond simple clustering, the default choice for quality assessment of activities, more appropriate metrics can be developed that are potentially the key to quantitative activity analysis.

Once the parameters of the feature learning scheme have been estimated (offline) the extraction of learned features corresponds to simple matrix multiplication. Consequently, the results of feature learning can be applied in online interactive applications (fifth design criteria). For some applications the classifiers might even be implemented on the sensors themselves, which would result in a substantial reduction in data transmission and in practical terms a more responsive system.

[1] The drop in accuracy for RBM+ECDF on Skoda (compared to plain RBM) is reasoned by an overfitting artifact of the unsupervised approach, which could easily be solved by employing the semi-supervised approach as used for the DA experiments.

5 Conclusion

One of the major shortcomings of activity recognition for ubiquitous computing is the lack of systematic approaches to feature extraction. By explicitly addressing this shortcoming we have demonstrated the suitability of feature learning for AR providing the basis for next generation AR applications. We identified practical design criteria for such activity recognition systems with respect to which we developed an activity recognition framework that employs PCA and deep belief networks for feature learning. An alternative representation of the sensor data, based on an estimation of the frame-wise empirical cumulative distribution of the signal, has been developed. The capabilities of feature learning methods were evaluated by means of recognition experiments on four publicly available AR datasets. Automatically estimated features outperformed classic heuristic features for all the analyzed AR tasks we considered. We also demonstrated that feature learning benefits from larger datasets but does not rely on them. The learning approach is computationally feasible and can be applied directly for interactive applications.

Our feature extraction framework has general applicability in ubicomp AR applications, particularly in circumstances where little is known about the target domain. The framework can be used "as is" for activity recognition tasks. Our experimental evaluation provides evidence that feature learning provides reasonable representations, which are immediately usable for further analysis tasks. The deep learning procedure provides sub-frame insights, which is important for a thorough analysis of the captured data.

Based on our findings a number of extensions can also be considered. Although it somewhat circumvents the learning approach we espouse, we could overcome the limitation of current frame-wise analysis procedures that they (typically) treat every sample independently, by explicitly incorporating derivatives into the feature representations. In addition, the linearity assumption could be relaxed during modeling. Nonlinear dependencies within the temporal data could be captured by means of kernel PCA approaches for the sub-space projection procedures.

The methodological key to the next generation of activity recognition lies in the systematic analysis of the analyzed sensor data. Beyond discriminating fixed numbers of certain activities of interest, domains such as behavior monitoring or skill assessment require quantitative classifications of the underlying sequential data streams. Our study represents a starting point for systematic research in such sensor data analysis. Given the promising results of the experimental evaluation, feature learning can be considered as having enormous potential for activity recognition.

References

[Atallah and Yang, 2009] L. Atallah and G. Yang. The use of pervasive sensing for behaviour profiling – a survey. *Pervasive and Mobile Computing*, (5):447 – 464, 2009.

[Chou, 1995] K.C. Chou. A novel approach to predicting protein structural classes in a (20–1)-D amino acid composition space. *Proteins: Structure, Function, and Bioinformatics*, 21(4):319–344, 1995.

[Cox and Oakes, 1984] D.R. Cox and D. Oakes. *Analysis of survival data*. Monographs on statistics and applied probability. Chapman and Hall, 1984.

[Figo *et al.*, 2010] D. Figo, P. Diniz, D. Ferreira, and J. Cardoso. Preprocessing techniques for context recognition from accelerometer data. *Pervasive and Ubiquituous Computing*, pages 645 – 662, 2010.

[Frank *et al.*, 2010] J Frank, S Mannor, and D Precup. Activity and gait recognition with time-delay embeddings. In *Proc. AAAI Conf. on Artificial Intelligence*, 2010.

[Hinton *et al.*, 2006] G.E. Hinton, S. Osindero, and Y.W. Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[Hinton, 2007] G.E. Hinton. To recognize shapes, first learn to generate images. *Progress in brain research*, 165:535–547, 2007.

[Huo *et al.*, 2004] Xiaoming Huo, Xuelei Ni, and Andre K Smith. A survey of manifold-based learning methods. In *Recent Advances in Datamining of Enterprise Data Algorithms and Applications*, pages 691 – 745. 2004.

[Huynh and Schiele, 2005] Tâm Huynh and Bernt Schiele. Analyzing features for activity recognition. In *Proc. Joint Conf. on Smart Objects and Ambient Intelligence*, 2005.

[Huynh *et al.*, 2008] Tâm Huynh, Mario Fritz, and Bernt Schiele. Discovery of activity patterns using topic models. In *Proc. Int. Conf. on Ubiquitous Computing*, 2008.

[LeCun *et al.*, 2006] Y LeCun, S Chopra, and R Hadsell. A tutorial on energy-based learning. In *Predicting Structured Data*. MIT Press, 2006.

[Lukowicz *et al.*, 2010] Paul Lukowicz, Stephen Intille, and Jamie A Ward, editors. *Proc. Int. Workshop on How To Do Good Research In Activity Recognition: Experimental methodology, performance evaluation and reproducibility.*, 2010.

[Minnen *et al.*, 2006] David Minnen, Thad Starner, M Essa, and Charles Isbell. Discovering characteristic actions from on-body sensor data. In *Proc. Int. Symp. on Wearable Comp.*, 2006.

[Pham and Olivier, 2009] Cuong Pham and Patrick Olivier. Slice & dice: Recognizing food preparation activities using embedded accelerometers. In *Proc. Europ. Conf. Ambient Intelligence*, 2009.

[Roggen *et al.*, 2010] D. Roggen, A. Calatroni, M. Rossi, T. Holleczek, K. Förster, G. Tröster, P. Lukowicz, D. Bannach, G. Pirkl, A. Ferscha, J. Doppler, C. Holzmann, M. Kurz, G. Holl, R. Chavarriaga, M. Creatura, and J. del R. Millán. Collecting complex activity data sets in highly rich networked sensor environments. In *7th Int. Conf. Networked Sensing Sys.*, 2010.

[Zappi *et al.*, 2008] Piero Zappi, Clemens Lombriser, Thomas Stiefmeier, Elisabetta Farella, Daniel Roggen, Luca Benini, and Gerhard Tröster. Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection. In *Proc. Europ. Conf. on Wireless Sensor Networks*, 2008.