



# Discovering Behavioural Predispositions in Data to Improve Human Activity Recognition

Maximilian Popko  
maximilian.popko@uni-rostock.de  
University of Rostock  
Rostock, Germany

Stefan Lüdtke  
luedtke@es.uni-mannheim.de  
University of Mannheim  
Mannheim, Germany

Sebastian Bader  
sebastian.bader@uni-rostock.de  
University of Rostock  
Rostock, Germany

Thomas Kirste  
thomas.kirste@uni-rostock.de  
University of Rostock  
Rostock, Germany

## ABSTRACT

The automatic, sensor-based assessment of challenging behavior of persons with dementia is an important task to support the selection of interventions. However, predicting behaviors like apathy and agitation is challenging due to the large inter- and intra-patient variability. Goal of this paper is to improve the recognition performance by making use of the observation that patients tend to show specific behaviors at certain times of the day or week. We propose to identify such segments of similar behavior via clustering the distributions of annotations of the time segments. All time segments within a cluster then consist of similar behaviors and thus indicate a behavioral predisposition (BPD). We utilize BPDs by training a classifier for each BPD. Empirically, we demonstrate that when the BPD per time segment is known, activity recognition performance can be substantially improved.

## CCS CONCEPTS

• **Computing methodologies** → **Cluster analysis; Supervised learning by classification.**

## KEYWORDS

human activity recognition, wearable sensors, machine learning, clustering

## ACM Reference Format:

Maximilian Popko, Sebastian Bader, Stefan Lüdtke, and Thomas Kirste. 2022. Discovering Behavioural Predispositions in Data to Improve Human Activity Recognition. In *iWOAR '22: 7th international Workshop on Sensor-based Activity Recognition and Artificial Intelligence (iWOAR '22), September 19–20, 2022, Rostock, Germany*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/3558884.3558892>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).  
*iWOAR '22, September 19–20, 2022, Rostock, Germany*  
© 2022 Association for Computing Machinery.  
ACM ISBN 978-1-4503-XXXX-X/18/06...\$15.00  
<https://doi.org/10.1145/3558884.3558892>

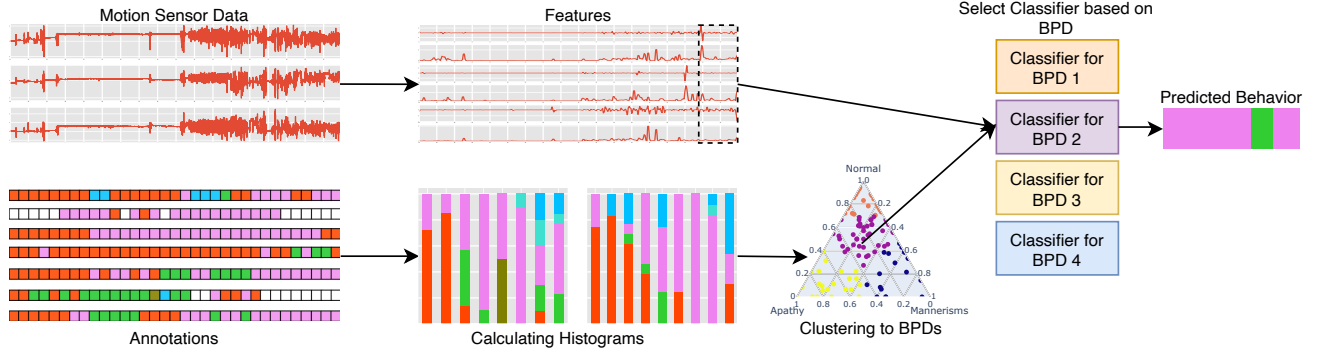
## 1 INTRODUCTION

The recognition of human behavior is important for many domains, like sports, the analysis of manual work processes, or as a component for situation-aware assistants. As another example, sensor-based Human Activity Recognition (HAR) can also be used for the automatic assessment of behavioral symptoms of people with dementia, like apathy or agitation. In contrast to questionnaires like the Cohen-Mansfield Agitation Inventory [4], HAR allows the real-time, objective assessment of symptoms, supporting caregivers in selecting appropriate interventions. However, accurate HAR is challenging due to the high inter- and intra-subject variability. Specifically, the automatic recognition of challenging behavior from wearable sensor data has been of limited success so far [9].

Multiple approaches have been proposed to improve HAR accuracy, by making use of additional external knowledge [10, 12, 13, 15, 18, 20]. In this paper, we investigate how HAR performance can be improved by making use of the fact that there is a tendency of specific subjects to show specific behaviors at certain times of the day or week. For example, sundowning is the increase in restlessness of some people with dementia in the evening [3]. The knowledge of such behavioral predispositions (BPDs) could serve as a useful prior to increase HAR accuracy.

Previous work relied on external knowledge to define BPDs. For example, Lüdtke et al. [13] partitioned warehouse order picking into distinct process steps, the knowledge of which can inform HAR. However, in the context of dementia care, there are no established BPDs that are externally given. The contribution of this paper is to identify such BPDs in a data-driven way: We calculate the distribution of annotations in time segments of a day, and then cluster these histograms. This results in clusters of time frames in which the subjects behaved similarly. Each cluster forms a BPD. Given the BPD, we then train a classifier specifically for data recorded within the BPD, i.e., there is one classifier for each cluster. This classifier is thus biased towards the more frequent classes in the BPD. An overview of this process can be found in figure 1. In summary, this paper provides the following contributions:

- A clustering-based method for the data-driven construction of behavioral predispositions (BPDs)
- An approach for utilizing these BPDs in HAR
- An evaluation of this approach for the sensor-based recognition of behavioral symptom of persons with dementia



**Figure 1: Overview of the concept.** We calculate the distribution of annotations for a given segment of the day (histograms). We then cluster these distributions, i.e. 7-dimensional vectors, that each represents the percentage of the behaviors shown in the time frame. The ternary plot displays the clustering of a subset of annotation labels. Next, we use the cluster label (BPD) to select the classifier (color indicates the corresponding classifier). The classifier predicts the behavior from the features of the motion sensor data.

Empirically, our approach leads to a substantial increase of HAR performance, assuming the BPD for each time segment is given.

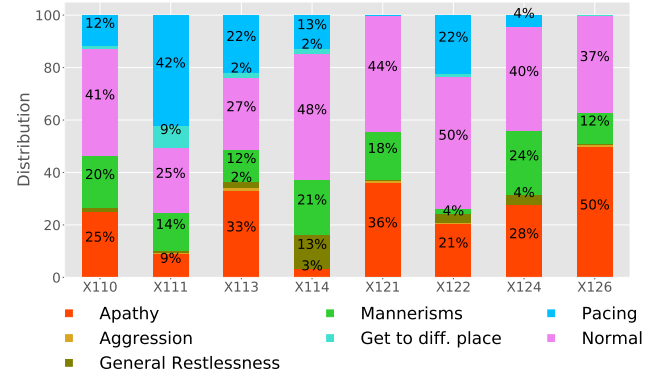
## 2 DATA AND OBSERVATIONS

This section presents the insideDem framework, as well as the data set. We make observations about the distribution of annotations given time frames, which will then lead to the construction of the BPDs and the utilization of them in the methodology section 3.

### 2.1 insideDem Project

Part of the insideDem framework was the recording of the behaviours of dementia diagnosed patients [16]. In two nursing homes, real-time observations and sensor data of a total of 17 residents (11 women and 6 men) with moderate to severe dementia were recorded. The behaviour of persons in one nursing home was also recorded on video. The age range was 73 to 94 years. All persons were under psycho-pharmacological treatment. Dementia medication was taken by 8 residents of the nursing homes. The behaviour of the persons was recorded in real time by an expert in the sitting room of the nursing home. This room was accessible to all residents and their visitors as well as the staff of the nursing home. The behaviour was annotated every five minutes following the annotation scheme which was developed for the insideDem project. This scheme is based on the neuropsychiatric inventory [7] and the Cohens-Mansfield agitation inventory [4]. It consists of seven different behaviours: apathy, general restlessness, mannerisms, pacing, aggression, trying to get to a different place, normal behaviour.

During the day, between 8 a.m. and 6 p.m., the subjects wore two wristbands, which were attached to the dominant hand and the ankle. During the night, a band was worn on the ankle. The sensors recorded accelerations and rotations as well as the sound level of the environment, the light conditions and the sound pressure. In this work, we only use the acceleration sensor data.

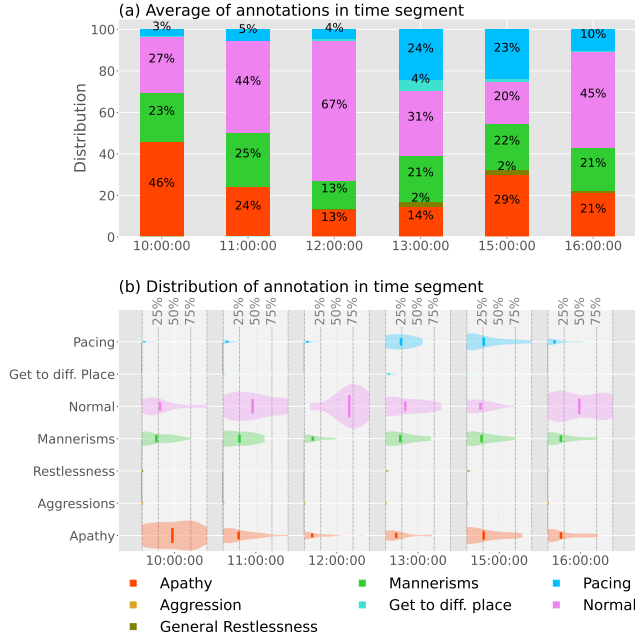


**Figure 2: Distribution of annotations per person over all days of recordings.**

### 2.2 Observations

Figure 2 shows the distribution of annotations for each subject. The behaviour of the eight patients were observed over the course of 29 days and accumulate around 17,000 annotated five-minute intervals. For every subject the behaviours occur in different portions. As example, the patient X111 showed 9% apathy, whereas X126 showed nearly 50% apathy during the time span of 29 days. For X111, pacing was very dominant, while X126 nearly showed no pacing.

In the literature, events for triggering agitated behaviors were observed. The sunset phenomenon as described in [3] may be such a trigger for one BPD. At sunset people with dementia often show more agitated behavior than at other times of the day. Also, the fact of boredom can trigger agitated behavior. If there is nothing planned for an afternoon, dementia diagnosed persons might show agitated behavior because of the boredom [5]. As a result, it is reasonable to look into noticeable time frames of the day which show one person behaving agitated more often.



**Figure 3: Percentual distribution of annotations for one subject and one-hour intervals of the day. (a) shows the average of the annotation and (b) describes the distribution (variance).**

Figure 3(a) now shows the occurrences of the annotations for one-hour segments of the day – averaged over all 29 days. The subject behaves in the morning more apathetic than in the later course of the day. Also at noon, the patient shows more normal behavior. Pacing is mostly happening in the afternoon. Figure 3(b) shows the variances of the annotations between each day. This plot represents the distribution of the proportion each behavior has on the time segment every day. Between 10 and 11 a.m. we saw the patient behaves apathetically to 46% in plot (a). However, in plot (b) we see that the apathy part can differ from 0% to 100% with equal probability from day to day. I.e., there are days in which the person was not apathetic at all from 10 to 11 a.m., and there are days in which the person is 100% apathetic during that hour.

### 3 METHODS

Behavioural predispositions (BPD) describe the tendency to show specific behaviours. In this section, we propose a method to form such BPDs from the annotations of behaviors. In the following, we present a simple model that can use these newly found BPDs to simplify the classification.

#### 3.1 Extracting behavioural predispositions

BPD describe the urge to show specific kinds of behaviors. Knowing such would intuitively mean that the prediction of the current actual behavior becomes easier since it provides an informative prior on the behavior of the subject. Unfortunately an expert model describing BPD for patients with dementia does not exist and thus the aim is to derive them from the data of the insideDem project.

**3.1.1 Time-based BPD.** Following the observations of figure 3 (a), one can argue to use an approach similar to [10] or define that every BPD is one segment of the day. This means from 10 to 11 a.m. the subject always behaves similarly and by knowing the current time, we can use this information to improve the classification. However, this approach has the drawback of high variances of the annotations between each day which lead to BPDs that do not carry a lot of information about the actual behavior.

**3.1.2 Clustering-based BPD.** To overcome the problem of high variances, we propose to find similar distributions globally over all days and time segments instead of grouping them by the time of the day. Let  $p(A_t)$  be the probability distribution of the annotation for the day and time segment  $t$ . Note that  $t$  describes a specific time frame e.g. Jun. 15 10:00–11:00 a.m. This categorical distribution can be obtained by counting the relative frequencies of the annotations in the segment. This counted distribution is called a histogram. The BPD represents the tendency to behave in a certain way. Given a set of histograms, the task is to divide this set into groups, containing similar histograms. With the help of the k-means clustering algorithm, clusters of histograms can be generated which are grouped according to their spatial similarity (Euclidean distance). Each of these clusters describes a behavioral disposition. The histograms of the cluster  $D = d$  therefore represent the distribution  $p(A | D = d)$ .

### 3.2 Classification of Sensor Data

In the scope of this work, we want to show what impact the complete knowledge of the clustered BPD can have on the classification. Therefore, we assume the BPD, i.e. the cluster, as given. Note that this would not be the case in a real world environment. Given a BPD  $d$ , we then have a prior knowledge of the annotation  $p(A | d)$ . To make use of this knowledge, we train a different classifier for each BPD. Thus, each classifier is trained with biased data and is therefore biased to more likely predict the classes that are more present in this BPD. The classifier can be seen as expert for its BPD. The model  $m$  maps sensor data  $s$  and the BPD  $d$  to an annotation  $a$  by choosing the classifier  $c_d$  that was specifically trained for  $d$ . The classifier then maps  $s$  to  $a$ :

$$m(s, d) = c_d(s) = a.$$

## 4 EXPERIMENTAL EVALUATION

The overall goal of the experiments is the evaluation of the impact of the BPD on the HAR performance. We also want to investigate the effects of different strategies for generating BPDs on classification performance. In the following, we describe the experimental design in more detail.

#### 4.1 Experimental Design

To evaluate the impact of the knowledge of differently generated BPDs on HAR performance, we use a factorial design. The factors and levels are presented in table 1. For each subject, a model with a different clustering method, number of clusters (i.e. BPDs), segment length, and classifier is trained. We randomly select 70% of the data for training and 30% for testing the model. Every combination of the levels is evaluated 10 times. Different clusters and segment lengths are used to test the expressiveness of the BPD.

**Table 1: Factors and levels of the experimental design**

Factors	Levels	Description
Clustering	K-means, time-based	Methods for generating the BPDs
#-clusters	1, ..., 20	Number of BPDs
Segment lengths	0, ..., 120 minutes (k-means)	Length of the segment over which the histogram is calculated, only for k-means approach
Classifiers	Majority, naive Bayes, SVM	Classifiers to predict the annotation from sensor data
Subjects	X110, X111, X113, X114, X121, X122, X124, X126	One model for each subject

The segment length plays an important role for the k-means clustering strategy as the length is used to define the size over which the histogram is calculated and then clustered. At a segment length of 30 minutes, we obtain around 400 histograms per subject. We then cluster the histograms of a single subject, resulting in BPDs specifically for this subject.

High number of clusters and small segment lengths yield finer BPDs. In contrast, one cluster means we have no further information about the BPD and thus gives us a baseline of performance. The time-based clustering method describes that a BPD only depends on the time. E.g., having 10 clusters and time-based clustering this method would split the day into 10 parts, and for every part one classifier is trained.

We use three different methods for classifying the sensor data. The majority method functions as a baseline by always outputting the most frequent annotation of the BPD. The naive Bayes (NB) method shows the performance of a simplistic classifier. The NB uses for every BPD a different prior and observation model. Thirdly, a support vector machine (SVM) is trained.

For preprocessing the accelerometer data, features were calculated following the work of [1]. The original data is sampled in 50 Hz. Feature were calculated over a window size of one minute and an overlap of 50%. Statistical features, spatial relations between the three axis and frequency features were used. As measure of accuracy, we use  $F_1$  score.

## 4.2 Results and Discussion

Figure 4 shows the average  $F_1$  scores for each classifier and the strategies used for generating the BPDs. Firstly, we want to investigate only the k-means approach. The SVM shows an increase of  $F_1$  score up to a number of about 10 clusters. From this point on, only smaller segment sizes benefit. For the majority classifier, the  $F_1$  score no longer increases significantly above a number of 6 clusters. The naive Bayes classifier shows a nearly linear increase. Thus, for all classifiers, the  $F_1$  measure increases with a larger number of clusters, as well as a decreasing segment length.

The fact that the majority classifier benefits from this approach comes as no surprise. With an increasing number of clusters and finer segment lengths, the most frequent annotation occupies a higher portion of data within the cluster. Predicting only the most frequent annotation yields an  $F_1$  score of around 0.1 whereas dividing the data by the BPD yields a score of around 0.5 at 20 BPDs and a segment length of 30 minutes. Note that some BPDs only describe one annotation and thus predicting can achieve 100% accuracy by always outputting that one same class. According to this

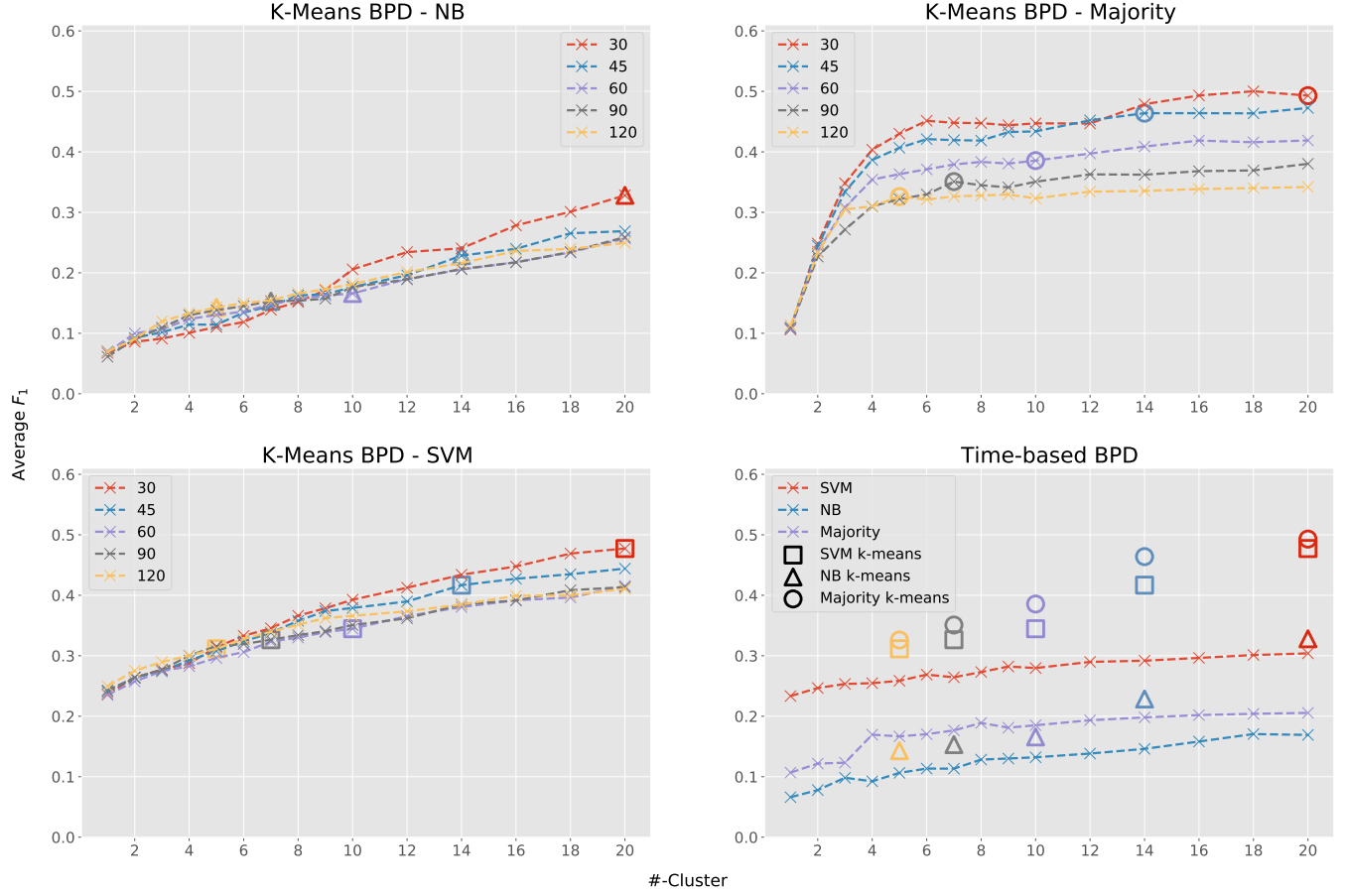
observation, the classification task should become simpler. Given such BPD with only one class, all classifiers used here can benefit and improve with knowledge of the BPD.

However, the  $F_1$  score of the majority classifier proposes that the knowledge of the most frequent annotation in the training data does not give us further advantages when we have more than 6 clusters. Yet the  $F_1$  score of the naive Bayes classifier increases nearly linearly. Also, the SVM has a still increasing  $F_1$  score at 20 BPDs, which underlines the fact that knowing the BPD simplifies the classification task by reducing the complexity of possible behaviors or even ruling them out. Note that the knowledge presented by the BPD converges with the BPD describing the exact underlying histogram, i.e. we have the same number of BPDs (clusters) as different histograms. Further, BPDs, that are created with a segment length of 5 minutes and 7 clusters, describe the original annotations since these were recorded in 5-minute intervals and consist of 7 behaviors.

We now investigate the time-based approach. The time-based approach shows an increase of  $F_1$  score as the number of clusters rises. This can be observed for every classifier evaluated. Although the increase is not high, the knowledge of the clock time can help the classification. Thus dividing the day into segments and training a classifier for every segment each, increases the  $F_1$  score. Comparing this approach with the k-means approach can be done by comparing the points where the number of BPDs and the segment length are approximately equal for both. E.g. splitting the day, where the subjects were recorded, into 20 segments yields a segment length of around 30 minutes and 20 BPDs. These points are denoted by the shapes in figure 4. The plot yields a better performance of the k-means approach for all 5 points and does not depend on the chosen classification method. The inferiority of the time-based approach can be explained through the high variances of annotations among the days (shown in figure 3). The k-means approach, on the other hand, describes the histograms by their similarity and thus achieving less variance within the cluster. Having less variance means that the classifier has an easier task to solve.

Overall, the  $F_1$  score reaches a maximum of around 0.5. This score is produced by using 20 BPDs, a segment length of 30 minutes, and an SVM as a classifier. The same results can be achieved by using the majority method. The SVM outperforms the majority classifier for larger segment lengths.

Figure 5 shows the confusion matrix of the SVMs for the classification with 20 BPDs and a segment length of 30 minutes. We can see that the classes mannerisms, apathy, and normal behavior represent



**Figure 4: Averaged  $F_1$  score for varying number of clusters and segment lengths. For time-based clustering, the segment length is implicitly given through the number of clusters. For k-means clustering, the segment lengths of the histograms are represented by the color. The segment length is in minutes. In the bottom right plot, shapes denote the  $F_1$  score of the classifiers with k-means-based BPDs. The color of the shapes represents the segment length. The five combinations of segment length and number of clusters allow for comparing the k-means-clustering and the time-based method. E.g., the red rectangle allows comparing the performance of the SVM for 20 clusters, corresponding to a segment length of 30 minutes.**

the majority of the annotations. Furthermore, there is great uncertainty between these classes. This can be explained by the similarity of these behaviors in this scenario. Intuitively, normal behavior is similar to apathy, as dementia patients in the nursing home can live more secluded. Mannerisms are difficult to recognize when, for example, they are performed with the non-dominant hand (sensor is worn at the dominant hand). Another uncertainty is between the annotations pacing, getting to a different place, and normal behavior. This can be explained by the fact that these three annotations can be understood as the activity of walking so a misclassification can occur between these classes. The behavior is also annotated in five-minute intervals, but the classification takes place in 30-second intervals. This means that the person may have walked at one point in time, but the interval was annotated differently. Another reason is the exclusive use of the motion sensor on the wrist.

## 5 FUTURE WORK

The evaluation showed that the knowledge of the BPD can yield better classification and thus further investigations are of interest. There are many directions in which future work can be established. In this section, we want to briefly present four topics that can be of great interest for extending this work: generating better BPDs via better clustering strategies, enriching BPDs with more underlying data, improving the classification methods, and constructing BPDs in real environments.

There is the question of whether finding the BPD with k-means clustering is the best choice. Since similar histograms and thus categorical probability distributions are of interest, clustering methods with entropy-based distances, such as the Jensen-Shannon divergence, are of interest. Furthermore, Latent Dirichlet Allocation (LDA) [2] can be investigated as a method for determining behavioral dispositions. The LDA is a generative probabilistic model and is mainly used in applications to assign a topic to documents based



SVM, 20 BPDs, 30 min segments

True Annotation	Apathy	8850	48	1579	1683	48	203
	Restlessness	22	96	41	5	0	26
	Mannerisms	1328	66	4417	1517	22	508
	Normal	1654	7	2465	9310	84	902
	Get to diff. Place	22	0	7	44	24	17
	Pacing	261	11	526	713	122	1988
		Predicted Annotation					
		Apathy	Restlessness	Mannerisms	Normal	Get to diff. Place	Pacing

**Figure 5: Confusion matrix of the predicted behaviors using 20 BPDs and thus 20 SVMs. A segment length of 30 minutes were used to calculate the histograms.**

on their words. In this scenario, a segment of the day can be regarded as a document and its words are the annotations in this segment so that the resulting topics describe the BPD.

Currently, the BPD is only constructed by the annotations. A next step could be to determine the BPD based on the annotation as well as the sensor data. This would lead to BPDs that also hold information about the current motion data and its connection to the annotation. In this work, we created BPDs for a single person. As a next step, the evaluation of global BPDs across the patients is needed.

We also saw that the overall accuracy of the classification was low and thus looking into better options for classification is of interest. In this work, we only presented the approach of training a classifier for each BPD. However, this reduces the number of samples for the classifier. Another approach can be the feeding of the BPD label as input to the classifier. Also, classifiers that can make use of distributions over BPDs are of interest. Methods like mixed-effects neural networks [19] can model BPDs as another feature for activity recognition.

In reality, the BPD is not available and has to be found from sensor data. A good estimation of the BPD is therefore needed to improve the classification. With an increasing number of BPDs, the error of classification will also rise. Thus finding a good number of BPDs and a model to predict them is a crucial point for practical use. This can be done by using hierarchical models which first infer the BPD from the sensor data and then make use of the relations between behavior and BPD. Also, hidden Markov models can be of interest to exploit possible temporal properties of the BPDs.

## 6 RELATED WORK

To improve the classification of sensor data using additional information, firstly, the knowledge has to be modeled and perhaps has to be generated beforehand. Depending on the domain, the kind of knowledge can differ and thus arise different ways of incorporation into human activity recognition. Mostly the improvement of accuracy is the main purpose. However, the simplification of the recognition model can also be of interest. The authors of [12] are aiming to create a model that still maintains the same level of accuracy under less training data than state-of-the-art methods. They use region connection calculus (RCC) [6] to describe the spatial relations between objects and formalize the activities and their causal properties within the domain using the planning domain definition language (PDDL) [8]. A probabilistic concurrent constraint automaton (PCCA) [17] is generated from the PDDL model and reasons about the predicates from the RCC model. The RCC model calculates these predicates from high-level observations. The authors show that their approach can yield similar results in comparison to convolutional neural networks (CNN) but avoid the training data hungry nature of neural networks. On the other hand, if enough training data and further information about the domain is available, the work of Xie et al. [18] aims to improve the accuracy of classification by incorporating prior knowledge in the training process of a neural network. They achieve this goal by modeling rules about the practicability of actions using linear temporal logics (LTL) [14]. To overcome the problem of the multi-interpretable nature of LTL, the author convert them into deterministic finite automata such that they can be embedded as hierarchical graph into the loss calculation of the neural network's training process and achieve learning consistent to the LTL rules and better accuracy at classification. The works by Yordanova et al. [20] and Rueda et al. [15] use computational causal behavioral models (CCBM) [11] to model domain knowledge and reason about the current state of the environment. CCBM also uses PDDL-like rules to describe the domain-specific properties of the activities. Also, the relations between user and object as well as durations of activities can be modeled in CCBM. As an observation model, a decision tree (DT) in [20] and a convolutional neural network in [15] are used to weight the predicted state based on the sensor data. In contrast to these works, the prior knowledge does not need to be embedded directly in the model. Lüdtkke et al. [13] demonstrate how providing external *context information* impacts classification performance. Specifically, they show that providing the current process step of structured activities can improve HAR performance, even when that information is noisy. Even if no external information is available, the authors of [10] made the observation that activities feature different quantities in times of a day. Thus segmenting the day and reasoning about the activities in one segment yields better accuracy than reasoning about the activities of the whole day.

## 7 CONCLUSION

In this paper, we proposed a data-driven method to construct behavioral predispositions and use them as informative prior to improve the performance of sensor-based human activity recognition. We presented a clustering-based approach that can determine time segments that display similar behaviors of a dementia-diagnosed

person. We found such groups by first calculating the distribution of the annotations in the time segment and then clustering these distributions. The resulting clusters explain behavioral predispositions (BPDs). With complete knowledge of the current BPD, we demonstrated empirically that a substantial increase of the  $F_1$  score can be achieved by using classifiers that were specifically trained for each BPD. For practical usage, the investigation of predicting the BPDs from sensor data is of high interest, as the BPD is not given a priori and instead must be inferred.

Furthermore, the presented approach is not only applicable to predicting the behavior of persons with dementia, but also to any continuous HAR task. Whenever the prior for occurrence of activities varies over time, the method described in this paper offers a viable option to improve HAR performance.

## REFERENCES

- [1] Ridwan Alam, Azziza Bankole, Martha Anderson, and John Lach. 2019. Multiple-Instance Learning for Sparse Behavior Modeling from Wearables: Toward Dementia-Related Agitation Prediction. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Berlin, Germany, 1330–1333. <https://doi.org/10.1109/EMBC.2019.8856502>
- [2] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.
- [3] Marco Canevelli, Martina Valletta, Alessandro Trebbastoni, Giuseppe Sarli, Fabrizia D'Antonio, Leonardo Taricotti, Carlo de Lena, and Giuseppe Bruno. 2016. Sundowning in Dementia: Clinical Relevance, Pathophysiological Determinants, and Therapeutic Approaches. *Frontiers in Medicine* 3 (Dec. 2016), 73. <https://doi.org/10.3389/fmed.2016.00073>
- [4] Jiska Cohen-Mansfield. 1996. Assessment of Agitation. *International Psychogeriatrics* 8, 2 (1996), 233–245. <https://doi.org/10.1017/S104161029600261X>
- [5] Jiska Cohen-Mansfield and Perla Werner. 1995. Environmental influences on agitation: An integrative summary of an observational study. *American Journal of Alzheimer's Care and Related Disorders & Research* 10, 1 (Jan. 1995), 32–39. <https://doi.org/10.1177/153331759501000108>
- [6] Anthony G. Cohn, Brandon Bennett, John Gooday, and Nicholas Mark Gotts. 1997. Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. *Geoinformatica* 1, 3 (Oct. 1997), 275–316. <https://doi.org/10.1023/A:1009712514511>
- [7] Jeffrey L. Cummings. 1997. The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. *Neurology* 48, 5 Suppl 6 (1997), 10S–16S.
- [8] M. Fox and D. Long. 2003. PDDL2.1: An Extension to PDDL for Expressing Temporal Planning Domains. *Journal of Artificial Intelligence Research* 20 (Dec. 2003), 61–124. <https://doi.org/10.1613/jair.1129>
- [9] Doreen Goerss, Albert Hein, Sebastian Bader, Margareta Halek, Sven Kernebeck, Andreas Kutschke, Christina Heine, Frank Krueger, Thomas Kirste, and Stefan Teipel. 2020. Automated sensor-based detection of challenging behaviors in advanced stages of dementia in nursing homes. *Alzheimer's & Dementia* 16, 4 (2020), 672–680. <https://doi.org/10.1016/j.jalz.2019.08.193> arXiv:https://alzheimer-journals.onlinelibrary.wiley.com/doi/pdf/10.1016/j.jalz.2019.08.193
- [10] Sozo Inoue, Naonori Ueda, Yasunobu Nohara, and Naoki Nakashima. 2016. Recognizing and Understanding Nursing Activities for a Whole Day with a Big Dataset. *Journal of Information Processing* 24, 6 (2016), 853–866. <https://doi.org/10.2197/ipsjip.24.853>
- [11] Frank Krüger, Alexander Steiniger, Sebastian Bader, and Thomas Kirste. 2012. Evaluating the Robustness of Activity Recognition Using Computational Causal Behavior Models. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (Pittsburgh, Pennsylvania) (UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 1066–1074. <https://doi.org/10.1145/2370216.2370443>
- [12] Sang Uk Lee, Andreas Hofmann, and Brian Williams. 2019. A Model-Based Human Activity Recognition for Human–Robot Collaboration. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Macau, China, 736–743. <https://doi.org/10.1109/IROS40897.2019.8967650>
- [13] Stefan Lüdtkke, Fernando Moya Rueda, Waqas Ahmed, Gernot A. Fink, and Thomas Kirste. 2021. Human Activity Recognition using Attribute-Based Neural Networks and Context Information. In *3rd International Workshop on Deep Learning for Human Activity Recognition*.
- [14] Amir Pnueli. 1977. The Temporal Logic of Programs. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science (SFCS '77)*. IEEE Computer Society, USA, 46–57. <https://doi.org/10.1109/SFCS.1977.32>
- [15] Fernando Moya Rueda, Stefan Lüdtkke, Max Schroder, Kristina Yordanova, Thomas Kirste, and Gernot A. Fink. 2019. Combining Symbolic Reasoning and Deep Learning for Human Activity Recognition. In *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, Kyoto, Japan, 22–27. <https://doi.org/10.1109/PERCOMW.2019.8730792>
- [16] Stefan Teipel, Christina Heine, Albert Hein, Frank Krüger, Andreas Kutschke, Sven Kernebeck, Margareta Halek, Sebastian Bader, and Thomas Kirste. 2017. Multidimensional assessment of challenging behaviors in advanced stages of dementia in nursing homes—The insideDEM framework. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 8, 1 (Jan. 2017), 36–44. <https://doi.org/10.1016/j.dadm.2017.03.006>
- [17] B.C. Williams, M.D. Ingham, S.H. Chung, and P.H. Elliott. 2003. Model-based programming of intelligent embedded systems and robotic space explorers. *Proc. IEEE* 91, 1 (2003), 212–237. <https://doi.org/10.1109/JPROC.2002.805828>
- [18] Yaqi Xie, Fan Zhou, and Harold Soh. 2021. Embedding Symbolic Temporal Knowledge into Deep Sequential Models. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Xi'an, China, 4267–4273. <https://doi.org/10.1109/ICRA48506.2021.9561952>
- [19] Yunyang Xiong, Hyunwoo J. Kim, and Vikas Singh. 2019. Mixed Effects Neural Networks (MeNets) With Applications to Gaze Estimation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Long Beach, CA, USA, 7735–7744. <https://doi.org/10.1109/CVPR.2019.00793>
- [20] Kristina Yordanova, Stefan Lüdtkke, Samuel Whitehouse, Frank Krüger, Adeline Paiement, Majid Mirmehdi, Ian Craddock, and Thomas Kirste. 2019. Analysing Cooking Behaviour in Home Settings: Towards Health Monitoring. *Sensors* 19, 3 (Feb. 2019), 646. <https://doi.org/10.3390/s19030646>