

Aggressive and agitated behavior recognition from accelerometer data using non-negative matrix factorization

Belkacem Chikhaoui¹ · Bing Ye² · Alex Mihailidis²

Received: 9 January 2017 / Accepted: 22 June 2017 / Published online: 8 July 2017
© Springer-Verlag GmbH Germany 2017

Abstract This paper presents a novel approach for aggressive and agitated behavior recognition using accelerometer data. Our approach applies first a noise reduction technique using the moving average filter method. Then, multiple features such as mean, variance, entropy, correlation and covariance are extracted from the filtered acceleration data using a sliding window. Non-negative matrix factorization is then used in order to project the data into a new reduced space that captures the significant structure of the data. The recognition is performed using the rotation forest ensemble method. The proposed approach is validated using extensive experiments on a real dataset collected at Toronto Rehabilitation Institute. We empirically demonstrate that our proposed approach accurately discriminates between behaviors and performs better than several state-of-the-art approaches.

1 Introduction

Behavioral and psychological symptoms of dementia (BPSD) are common and problematic in clinical practice. They represent a significant part of the day-to-day workload of caregivers and care providers (Lawlor 2002).

Challenging behaviors, such as agitation and aggression, are very common in people with dementia and regarded as part of BPSD (Desai and Grossberg 2001). Agitation consists of an unusual state of motor or verbal activity that could be shown by some of the following symptoms such as repetitive walking, wandering, pacing or restlessness, frequent requests for attention or reassurance, frustration, anger or irritability, screaming, cursing, and refusal to allow care to be performed. Whereas aggression is when the behaviors are taken to a more physical point and can be demonstrated by behaviors such as verbal or physical threats, kicking and punching, tearing things, and violent reactions (Mallidou et al. 2013).

These challenging behaviors can cause great suffering for persons with dementia, premature institutionalization, and could result in staggering health care costs, significant loss of quality of life, and a great deal of distress and burden for caregivers (Moore et al. 2013). In addition, Tampi et al. (2011) reported that these challenging behaviors add significantly to the direct and indirect costs of care. For example, according to MS et al. (2002), approximately 30% (\$4115 US) of the total annual cost of a patient with Alzheimer's Disease (AD) (\$14,420 US) is invested in the direct management of BPSD. Therefore, early detection and recognition of these challenging behaviors can help effectively provide better treatment for persons with dementia, which in turn will help reduce caregiver's burden (Desai and Grossberg 2001) and reduce significantly health care costs.

Clinical scales such as the neuropsychiatric inventory (NPI), and the Cohen-Mansfield agitation inventory (CMAI) are the most frequently used methods in pharmacotherapeutic research to monitor BPSD behaviors. These clinical scales are based on direct observation from family caregivers and the care staff to identify challenging

✉ Belkacem Chikhaoui
belkacem.chikhaoui@usherbrooke.ca

Bing Ye
alex.mihailidis@utoronto.ca

Alex Mihailidis
bing.ye@uhn.ca

¹ TELUQ University, Montreal, Canada

² IATSL Laboratory, Toronto Rehab Institute, University of Toronto, Toronto, Canada

behaviors. However, this method is subjective, time consuming and could increase the workload of care staff and caregivers (Desai and Grossberg 2001). Therefore, researchers have focused on developing intelligent systems to automatically monitor and recognize aggression and agitation (Qiu et al. 2007) as not only will technology potentially reduce the manpower and time needed to observe and detect these behaviors (Fook et al. 2007), it may also have the potential to give reliable and consistent results (Hung et al. 2010; Mori et al. 2007; Duong et al. 2005) on predictors of these behaviors.

Much research has been conducted on human behavior recognition (Aggarwal and Cai 1999; Bouziane et al. 2013; Sheng et al. 2015; Zhu et al. 2013; Guo 2011), however, very little work has been done on automatic recognition of agitated and aggressive behaviors in people with dementia. In addition, with the tremendous growth of fitness applications and devices such as smart watches, current fitness devices make physical activities monitoring and tracking less intrusive, which helps in developing practical applications for monitoring and tracking healthy people and more specifically people with dementia. Therefore, the motivations for our current work can be summarized in the following points: (1) the little work on automatic agitation and aggression recognition, (2) the goal of decreasing the suffering of persons with dementia and increasing their quality of life, and (3) the goal of reducing caregivers' burden and related care costs.

In this paper, we propose an effective approach for aggressive and agitated behavior recognition using accelerometer data. Our approach first extracts different features from filtered acceleration data. Then, it applies non-negative matrix factorization technique to project the data into a new reduced space. The recognition is performed using an ensemble learning method based on rotation forests. The combination of non-negative matrix factorization and ensemble learning leads to a significant improvement in the recognition of aggressive and agitated behaviors as compared to the state-of-the-art approaches. The major contributions of this paper can be summarized as follows:

1. This work is, to the best of our knowledge, the first formal study of agitated and aggressive behavior recognition combining acceleration data with non-negative matrix factorization.
2. Combine non-negative matrix factorization and ensemble learning to improve aggressive and agitated behavior recognition.
3. Conduct extensive experiments over a real dataset to validate our proposed approach.

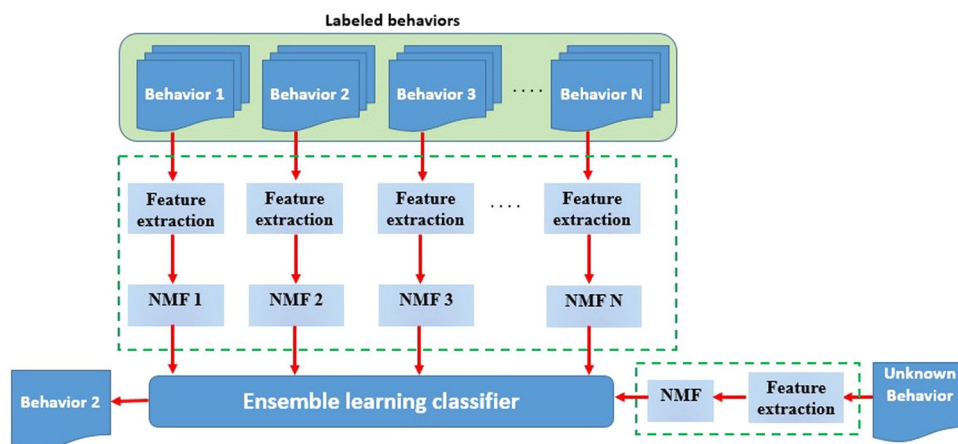
The rest of the paper is organized as follows. First, we give an overview of related work in Sect. 2. Section 3 describes

the proposed approach in terms of features extraction, learning and recognition using rotation forests ensemble method. The results of our experiments on real dataset are presented in Sect. 4. Finally, Sect. 5 presents our conclusions and highlights future work directions.

2 Related work

Much work has been done on daily living activity recognition using accelerometers (Kwapisz et al. 2011; Ravi et al. 2005; Krishnan and Cook 2014; Liu et al. 2009). However, there have been few studies evaluating the application of accelerometers to the measurement of aggressive and agitated behaviors. Mahlberg and Walther (2007) investigated the usefulness of actigraphy as an objective way to measure day-night rhythm disturbances and agitated behaviors in patients with dementia. As a result of their study, the authors concluded that actigraphy could be used in monitoring treatment success in BPSD as it shows correlation between actigraphy and Neuropsychiatric Inventory scores. However, the authors did not study aggressive and agitated behavior recognition using actigraphy. Pan et al. (2013) evaluated the severity of BPSD for vascular dementia (VD) from actigraphy records and compared the results with clinical scores such as NPI and the behavioral pathology in Alzheimer's disease (BEHAVE-AD) rating scale. The authors observed a linear correlation between the changes in activity disturbances plus anxieties and phobias and those of diurnal activity. Moreover, a linear correlation was also observed between the changes in agitation plus irritability scores of the NPI score and the changes in diurnal activity. Tractenberg et al. (2003) compared the effects of treatments such as melatonin on sleep disorders using actigraphic recordings. They found that the actigraphic sleep patterns showed a linear correlation with melatonin. Knuff (2014) investigated the correlation between actigraphy for the measurement of neuropsychiatric symptoms of agitation in older adults with dementia and questionnaire-based measures of NPS, including the Cohen-Mansfield Agitation Inventory and other measures of NPS. The authors found significant positive correlations between overall motor activity as measured by actigraphy mean motor activity (MMA) counts and the CMAI total scores in daytime and evening, while no correlations were observed during nighttime. Moreover, the authors observed that patients with high CMAI scores had higher levels of activity than patients with low CMAI scores. Kim et al. (2013) investigated covariant properties between the symptoms of depressive mood, anxious mood, and fatigue and locomotor activity using an actigraphy. The authors found a positive correlation between depressive mood and locomotor activity. Coronato et al. (2014) proposed a situation-aware system for the detection of stereotyped motion

Fig. 1 Overview of the different steps of our approach



disorders of patients with Autism spectrum disorders. Their proposed system used accelerometer data, whose waveforms, in the case of motion disorders, show clearly identifiable patterns. An artificial neural network (ANN) is used to classify the temporal frames of patient gestures against such patterns and generates an event whenever a temporal frame is classified as a disorder. However, all these studies investigate the usefulness of actigraphy and accelerometers to measure aggressive and agitated behaviors, and no methods have been developed for the automatic recognition of these behaviors.

Other researchers combined acceleration data and physiological data to detect agitation. For instance, Sakr et al. (2010) used bio-physiological measures to detect agitation by monitoring the changes of the heart rate, galvanic skin response and skin temperature of the participants. In another study, Rajasekaran et al. (2011) proposed a wearable device for early detection of anxiety and agitation in people with cognitive impairment. Thomas et al. (2012) proposed a system based on machine learning techniques to segment relevant behavioral episodes from a continuous wearable sensor stream and to classify them into distinct categories of severe behavior such as aggression, disruption, and self-injury. The system was validated using simulated data of episodes of severe behavior acted out by trained specialists, and other daily living activities available datasets. The results from these studies showed accurate detection of disruptive behaviors. However, all these studies looked at physiological data to detect agitated and aggressive behaviors. The difference between these studies and our work remains in the fact that our work focuses only on acceleration data to recognize aggressive and agitated behaviors.

Overall the aforementioned studies investigated the relationships between wearable sensors data and the agitated and aggressive behaviors and no formal approaches have been developed to automatically recognize these behaviors. These points motivate us to propose a new principled approach for agitated and aggressive behavior recognition

from accelerometer data only. Our approach combines non-negative matrix factorization method and ensemble learning classifier for accurate behavior representation and recognition. The discrimination power of non-negative matrix factorization, and the performance of ensemble learning algorithms compared to traditional data mining algorithms will help strengthen our approach and make it effective compared to the existing approaches.

3 Proposed approach

In this section, we describe our approach for aggressive and agitated human behavior recognition in terms of data preprocessing, feature extraction, non-negative matrix factorization and ensemble learning classification. Figure 1 shows an overview of the different steps of our approach. The details of each segment in Fig. 1 are presented in the following sections.

3.1 Data preprocessing

Data collected from accelerometers are often noisy and need to be cleaned before processing. Several filters have been used in the literature to reduce the level of noise in the data such as Kalman filters (Gannot et al. 1998), moving average (Hamed Azami 2012), and low pass filters (Baer et al. 2002). For the sake of simplicity and computational complexity, we choose the simple moving average (Hamed Azami 2012) method.

The simple moving average (SMA) is the average of the values over the last n values, where the last n values help calculating the predicted value as shown in Fig. 1.

$$V_{t+1} = \frac{V_t + V_{t-1} + \dots + V_{t-n+1}}{n}, \quad (1)$$

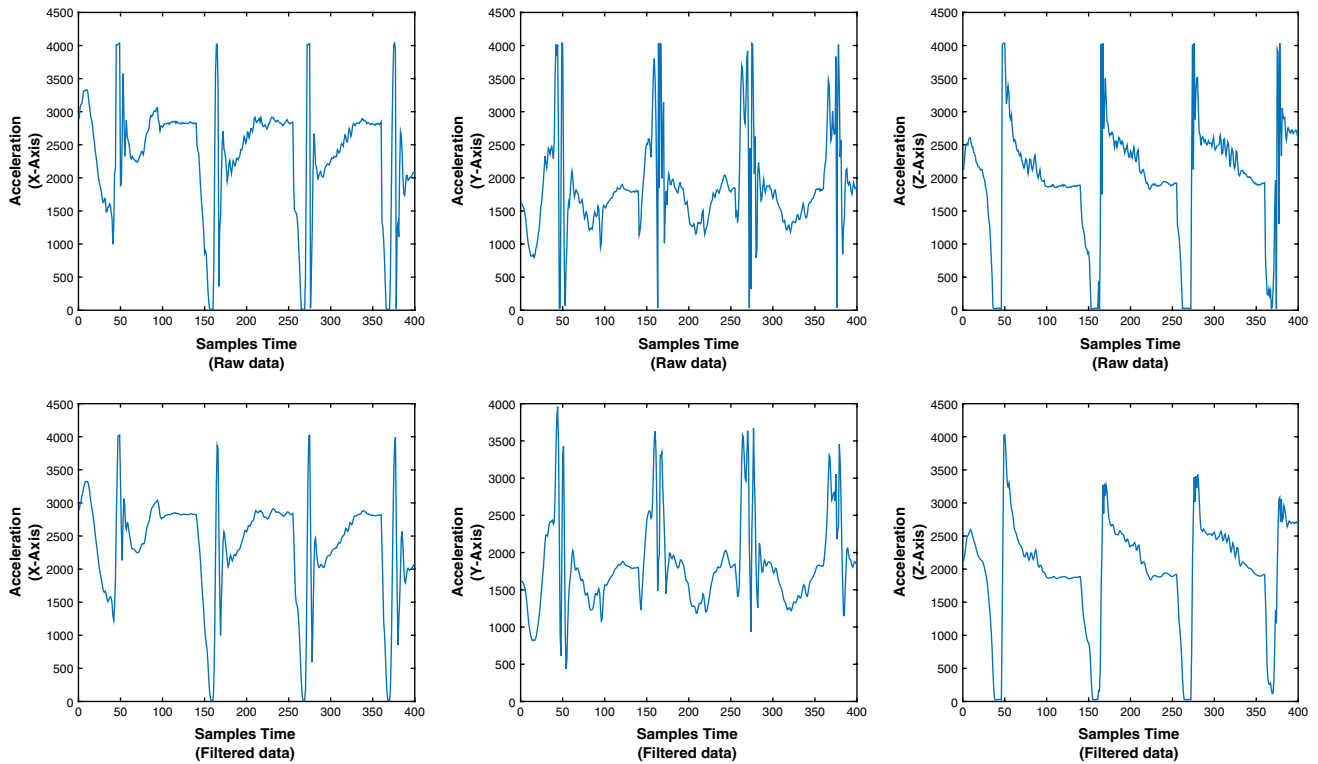


Fig. 2 Example of raw data and filtered data using simple moving average method with $n = 3$

where V_{t+1} is the predicted value and $V_t, V_{t-1}, \dots, V_{t-n+1}$ are the past n values. Figure 2 shows an example of raw acceleration data filtered using the SMA filter.

A window length of $n = 3$ has been experimentally estimated as a good tradeoff between noise reduction in high frequency and signal dynamics preservation in low frequency (Arias-Castro and Donoho 2009; Bruno et al. 2013). Once data is filtered, the next step is to extract features that will be used for the classification step.

3.2 Feature extraction

Features were extracted from the filtered accelerometer data using a window size of $w = 14$ with 50% samples overlapping between consecutive windows. Feature extraction on windows with 50% overlap has demonstrated success in previous work (Bao and Intille 2004). At a sampling frequency of 50 Hz, each window represents data for 0.28 s, which is reasonable given that aggressive actions are usually performed quickly (Stern 2010). The window size of 14 yielded better results as well as many training examples (please see Sect. 5.2 for more details on how to empirically select w). The extracted features are described as follows (Table 1):

The extracted features take into account the description of the three axis when they are taken separately, two

axis conjointly and all the axis together. This will allow to extract rich information about each behavior. Note that some features may have negative values. Therefore, we take the absolute values in order to get features with positive values only. These extracted features will then be used by the non-negative matrix factorization method in order to project the data into a new space. The next section introduces this method and describes how the projection will be done.

3.3 Behavior representation using non-negative matrix factorization

Non-negative matrix factorization (NMF) is a matrix factorization algorithm that finds the positive factorization of a given positive matrix (Lee and Seung 1999, 2000). In NMF, each axis captures the base information of a particular behavior class, and each behavior is represented as an additive combination of the base informations. The class membership of each behavior can be easily determined by finding the base posture (the axis) with which the behavior has the largest projection value. Therefore, the potential of using NMF lies in the discriminative power between the behaviors when projected into the new space. NMF has been successfully applied in different situations such as parts-based representation in human brain (Palmer 1977), learning parts of objects like human faces (Paatero and

Table 1 Statistical features used in our approach

Id	Statistical features	Definition
1	Mean	Mean acceleration for each axis
2	Standard deviation	Standard deviation for each axis
3	25th percentile	25th percentile for each axis Ermes et al. (2008)
4	75th percentile	75th percentile for each axis Ermes et al. (2008)
5	Average absolute difference (AAD)	Average absolute difference between the value of each of the 14 readings within the window and the mean value over those 14 values for each axis Kwapisz et al. (2011), which can be computed as follows: $AAD_X = \frac{1}{w} (\sum_i (X_i - \bar{X}))$, where \bar{X} is the mean value over the 14 readings within the window
6	Average resultant acceleration	Average of the square roots: $\frac{\sum_i \sqrt{x_i^2 + y_i^2 + z_i^2}}{w}$, over the 14 readings within the window
7	Entropy	The entropy of each axis: $E_X = -\sum_i p_i \log_2 p_i$ over the 14 readings within the window
8	Covariance	Covariance between each pair of axis $Cov(X, Y)$, $Cov(X, Z)$ and $Cov(Y, Z)$ computed as follows: $Cov(X, Y) = \frac{1}{w} \sum_{n=1}^w (X_n - \bar{X})(Y_n - \bar{Y})$, where \bar{X} and \bar{Y} are the means for X axis and Y axis values over the 14 readings within the window
9	Pearson correlation	Pearson correlation between each pair of axis $PC(X, Y)$, $PC(X, Z)$ and $PC(Y, Z)$ computed as follows: $PC(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$, where $Cov(X, Y)$ is the covariance between X and Y and σ_X and σ_Y are the standard deviations of X and Y respectively over the 14 readings within the window.

Tapper 1994), face recognition (Li et al. 2001) and document clustering (Xu et al. 2003) among others.

Formally, given a data matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$, NMF consists in factorizing the matrix \mathbf{X} into the non-negative matrix $\mathbf{U} = [u_{ij}] \in \mathbb{R}^{m \times k}$ and the non-negative matrix $\mathbf{V} = [v_{ij}] \in \mathbb{R}^{n \times k}$ as follows:

$$\mathbf{X} \approx \mathbf{UV}^T, \tag{2}$$

by minimizing the following objective function Φ :

$$\Phi = \frac{1}{2} \|\mathbf{X} - \mathbf{UV}^T\|^2 \tag{3}$$

where $\|\cdot\|$ denotes the squared sum of all in the matrix (please see Sect. 5.1 on how to select the rank k). Here, the objective function Φ , which represents the squared Euclidean distance, seeks to minimize the error of the reconstruction of the original matrix \mathbf{X} by the product \mathbf{UV} . The objective function Φ can be rewritten as follows:

$$\begin{aligned} \Phi &= \frac{1}{2} tr((\mathbf{X} - \mathbf{UV}^T)(\mathbf{X} - \mathbf{UV}^T)^T) \\ &= \frac{1}{2} tr(\mathbf{XX}^T - 2\mathbf{XVU}^T + \mathbf{UV}^T\mathbf{VU}^T) \\ &= \frac{1}{2} (tr(\mathbf{XX}^T) - 2tr(\mathbf{XVU}^T) + tr(\mathbf{UV}^T\mathbf{VU}^T)) \end{aligned} \tag{4}$$

here the matrix property $tr(\mathbf{UV}) = tr(\mathbf{VU})$ is used in the derivation steps. Lee and Seung (2000) presented an iterative update algorithm to find a local minimum of the objective function Φ as follows:

$$u_{ij}^{t+1} = u_{ij}^t \frac{(\mathbf{XV})_{ij}}{(\mathbf{UV}^T\mathbf{V})_{ij}} \tag{5}$$

$$v_{ij}^{t+1} = v_{ij}^t \frac{(\mathbf{X}^T\mathbf{U})_{ij}}{(\mathbf{VU}^T\mathbf{U})_{ij}} \tag{6}$$

Lee and Seung (2000) proved that the convergence of the iterations is guaranteed, however, the solution to minimizing the objective function Φ is not unique. If \mathbf{U} and \mathbf{V} are the solutions to Φ , then, \mathbf{UH} and \mathbf{VH}^{-1} will also form a solution for any positive diagonal matrix \mathbf{H} . To this end, a normalization is needed to make the solution unique as follows:

$$u_{ij} = \frac{u_{ij}}{\sqrt{\sum_i u_{ij}^2}} \tag{7}$$

$$v_{ij} = v_{ij} \sqrt{\sum_i u_{ij}^2} \tag{8}$$

Therefore, each data vector \mathbf{x}_i is approximated by a linear combination of the columns of \mathbf{U} , weighted by the components of \mathbf{V} . The non-negative constraints on \mathbf{U} and \mathbf{V} allow additive combinations among different basis. Unlike SVD, no subtraction can occur in NMF. This is the most significant difference between NMF and other matrix factorization algorithms such as SVD, PCA, and vector quantization (VQ) (Cai et al. 2008). For instance, in VQ, each column of \mathbf{V} is constrained to be a unary vector, i.e. one element equal to unity and the remaining elements equal to zero. In PCA the columns of \mathbf{U} are constrained to be orthonormal and the rows of \mathbf{V} to be orthogonal to each other, which is considered as relaxation of the unary property in VQ (Lee and Seung 1999). In contrast, NMF does not allow

negative entries in both matrices \mathbf{U} and \mathbf{V} . The non-negativity property of NMF allows the combination of multiple base information of behavior postures to represent the human behavior.

Note that the output of the NMF method in our approach is the input for the rotation forest ensemble learning classifier as mentioned in the next section.

3.4 Classification using ensemble method

The aim of ensemble methods is to improve the predictive performance of a given model by combining several learning algorithms. It has been proven that conventional classifiers such as random forests, decision trees and SVM are less accurate when compared to ensemble methods (Opitz and Maclin 1999). This motivates us to incorporate ensemble methods in order to build our classification model.

Rotation forest (Rodríguez et al. 2006) is a tree based ensemble method for building classifier ensembles using independently trained decision trees, which means that the base learner classifier in a rotation forest ensemble method is a decision tree. It was found to be more accurate than bagging, AdaBoost and Random Forest ensembles across a collection of benchmark datasets (Kuncheva and Rodríguez 2007). The strength of rotation forests lay in the use of principal component analysis to rotate the original feature axes so that different training sets for learning base classifiers can be formed (Kuncheva and Rodríguez 2007).

Formally, let $\mathbf{x} = [x_1, \dots, x_n]^T$ be a data point described by n features, and let A be an $N \times n$ matrix containing the training example. Let $Y = [y_1, \dots, y_N]^T$ be a vector of class labels for the training data, where y_j takes a value from the class labels $\{w_1, \dots, w_c\}$. Let $D = \{D_1, \dots, D_L\}$ be the ensemble of L classifiers and \mathbf{F} be a feature set. The idea is that all classifiers can be trained in parallel. Therefore, each classifier D_i is trained on a separate training set T_{D_i} to be constructed as follows (Rodríguez et al. 2006):

1. split the feature vector \mathbf{F} into K subsets. The subsets may be disjoint or intersecting.
2. for each of the subsets, select randomly a nonempty subset of classes and then draw a bootstrap sample of objects.
3. run PCA using only the M features in $\mathbf{F}_{i,j}$ and the selected subset of A , where j is the j th subset of features for the training set of classifier D_i . Then, store the obtained coefficients of the principal components $\mathbf{a}_{i,j}^1, \dots, \mathbf{a}_{i,j}^{M_j}$ in a matrix $C_{i,j}$.
4. rearrange the columns of the matrix $C_{i,j}$ in a new matrix B_i^a so that they correspond to the original features in matrix A .
5. the training set for classifier D_i is (AB_i^a, Y) .

6. to classify a new sample \mathbf{x} , we compute the confidence ψ for each class as follows:

$$\psi_j(\mathbf{x}) = \frac{1}{L} \sum_{i=1}^L d_{i,j}(\mathbf{x}B_i^a), \quad j = 1, \dots, c \quad (9)$$

where $d_{i,j}(\mathbf{x}B_i^a)$ is the probability assigned by the classifier D_i indicating that \mathbf{x} comes from class w_j . Therefore, \mathbf{x} will be assigned to the class having the highest confidence value.

Note that rotation forest aims at building accurate and diverse classifiers. Therefore, to maximize the chance of getting high diversity, it is suggested to take disjoint subsets of features. For instance, this can be obtained by taking $M = n/K$, where K is a factor of n . The next section presents the validation of our proposed model. The steps of our approach are presented in Algorithm 1.

Algorithm 1: Classification algorithm using Rotation Forest ensemble method.

```

Input:
- Acceleration data for all behavior instances
- Training set matrix  $X = \emptyset$ 
-  $L$ : the number of classifiers in the ensemble method
-  $K$ : the number of subsets
- the set of class labels  $\{w_1, \dots, w_c\}$ 
Output:
- Class labels for new behavior instances
Training phase
foreach Behavior do
  foreach Behavior instance do
    - Compute the feature set  $\mathbf{F}$ 
    - Add features to matrix temp
    - Take absolute values of matrix  $|temp|$ 
  end
  - Compute matrices  $U$  and  $V$  form matrix temp using Equation 2
  - Add  $X \approx UV^T$  to training set  $X$ 
end
for  $i=1 \dots L$  do
  - Split features  $\mathbf{F}$  into  $K$  subsets:  $\mathbf{F}_{i,j}$  ( $j=1 \dots K$ )
  for  $j=1 \dots K$  do
    - Let  $A_{i,j}$  be the dataset obtained using the features in  $\mathbf{F}_{i,j}$ 
    - Eliminate a random subset of classes from  $A_{i,j}$ 
    - Select a bootstrap sample  $A'_{i,j}$  of size 75% of objects from  $A_{i,j}$ 
    - Run PCA on  $A'_{i,j}$  and store obtained coefficients in a matrix  $C_{i,j}$ 
  end
  - Rearrange the columns of  $C_{i,j}$  in a new matrix  $B_i^a$  so that they match the order of features in matrix  $A$ 
  - Build classifier  $D_i$  using  $AB_i^a$  as a training set
end
Classification phase
for a given  $\mathbf{x}$  do
  - Compute  $d_{i,j}(\mathbf{x}B_i^a)$ 
  - Compute confidence  $\psi_j(\mathbf{x})$  using Equation 6
  - Assign  $\mathbf{x}$  to class having the largest confidence
end

```

4 Validation

We evaluate the performance of our approach on two real human behavior datasets. The first dataset contains aggressive and agitated human behaviors obtained by conducting experiments in Toronto Rehabilitation Institute (TRI), and the second dataset contains normal human behaviors. This dataset is used for comparison purposes with state-of-the-art approaches and it is described in the Sect. 4.3. Each dataset has almost completely distinct sets of actions. We ran our algorithm for subsets of three



Fig. 3 Shimmer sensor with X(left/right), Y(forward/backward) and Z(up/down) axis directions

features ($M = 3$), and ten decision tree classifiers in the ensemble ($L = 10$).

4.1 TRI Dataset

The dataset used in this work is obtained by conducting an experiment in Toronto Rehabilitation Institute-UHN (TRI-UHN). Ten (10) participants, whose ages ranged from 18 to 53 years (6 males and 4 females, 3 among them were left-handed) were involved in this experiment to conduct six (6) aggressive and agitated actions (hitting, pushing, throwing, tearing, kicking and wandering) by wearing a Shimmer¹ accelerometer sensor as shown in Fig. 3. The acceleration data were recorded using the Shimmer connect application² installed on a laptop.

The selected actions have been identified as the most common challenging aggressive and agitated behaviors³ observed from persons with dementia. These behaviors were selected from Cohen-Mansfield Agitation Inventory (CMAI) Scale (Cohen-Mansfield 1991). These behaviors are described as follows:

1. *Hitting* To simulate this behavior, participants were asked to raise one of their hands up and pretend to hit something in front of them.
2. *Pushing* To simulate this behavior, participants were asked to use their both hands at the same time and pretend to push something in front of them.
3. *Throwing* To simulate this behavior, participants were given an object and asked to throw it out as far as possible using one hand. The object is a piece of light foam cut from a camping mattress.
4. *Tearing* To simulate this behavior, participants were given a piece of paper and asked to tear it using both hands.

¹ <http://www.shimmersensing.com>.

² <http://www.shimmersensing.com/shop/shimmer3>.

³ Here we use the terms Behavior and Action interchangeably.

5. *Kicking* To simulate this behavior, participants were asked to raise one of their feet up and pretend to kick something in front of them.
6. *Wandering* To simulate this behavior, participants were asked to look for something that they couldn't find. They were asked to make a step forward and look for something on the ground from side to side and then look up for something from side to side, and then make a step backward and redo the same movements.

Participants were asked to perform the full set of actions using the right side of the body. For instance, hitting and kicking with the right hand and the right foot respectively. Note that two of these actions, pushing and wandering, are not specific to one side of the body. In order to ensure the study is generic and takes into account both left-handed and right-handed people, participants were then requested to repeat the four laterally specific actions, hitting, kicking, throwing and tearing, using the left side of the body. Participants performed all the actions five times. A total of $((10 \text{ (participants)} \times 4 \text{ (behaviors)} \times 5 \text{ (repetitions)} \times 2 \text{ (left hand and right hand)}) + (10 \text{ (participants)} \times 2 \text{ (wandering and pushing)} \times 5 \text{ (repetitions)} \times 1 \text{ (one side of body)})) = 400 + 100 = 500$ behavior instances have been collected in our experiment. A Research Ethics Board (REB) approval was obtained prior to collecting the data. Figure 4 shows an example of skeleton images for each action performed by one participant.

Similarly, Fig. 5 shows an example of acceleration signals data for three actions such as Hitting, Kicking and Wandering performed by one participant.

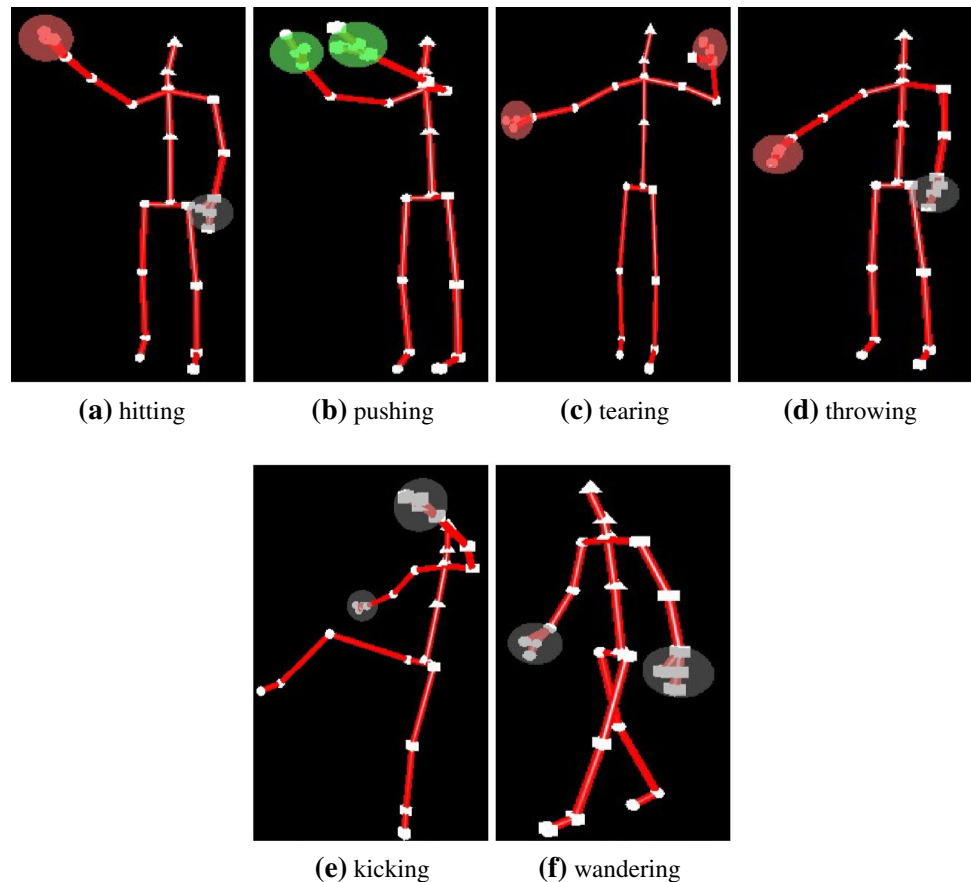
As shown in Table 2, almost all the behaviors, except the Wandering behavior, have small duration which justifies the choice of a small window length in processing the data. Table 3 shows the percentage of instances of each behavior in the right handed and left handed datasets.

As we can observe from Table 3, almost all the behaviors have a similar number of training instances, except for the Wandering behavior, which has more training instances compared to the other behaviors because of its duration period during experiments and the amount of acceleration data collected. Indeed, the bigger is the duration period of a behavior, the larger is the number of training instances.

4.2 Experimental results

We first evaluate the performance of our proposed approach using the TRI dataset. Then, we compare our results to the state-of-the-art methods to demonstrate the superiority and effectiveness of our proposed approach. In our experiments, we used different measures such as accuracy, precision, recall and F-measure to present the results. We experimentally determined the optimal rank of the NMF method

Fig. 4 Example of skeleton images for each action performed by one participant



k that achieved the best classification results. Determining the optimal factorization rank automatically will be considered in our future work.

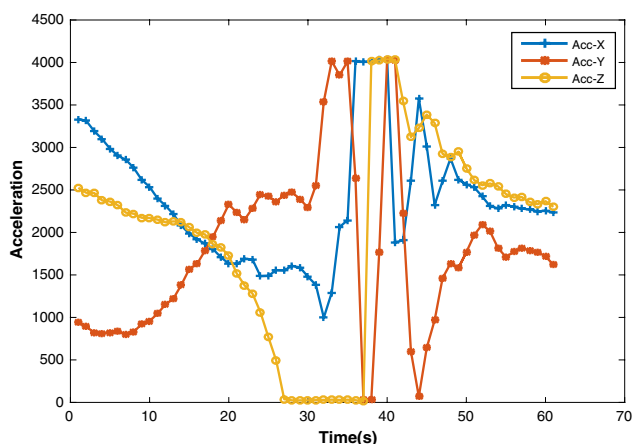
4.2.1 Leave one out cross validation

In this experiment, we used all behavior instances from participants for training and the behavior instances of the remaining participant for testing. We performed the experiment 10 times, excluding one participant at each time. The benefit of such setup is twofold. First, it allows detecting problematic participants and analyzing the sources of some of the classification errors caused by these participants. A problematic participant means his/her behaviors were performed differently compared to other participants. Second, it allows testing the inter-participant generalization of the approach, which constitutes a good indicator about the practicability of our approach. Tables 4 and 5 show respectively the recognition results obtained for the right handed and left handed datasets respectively using the precision, recall and F-measure.

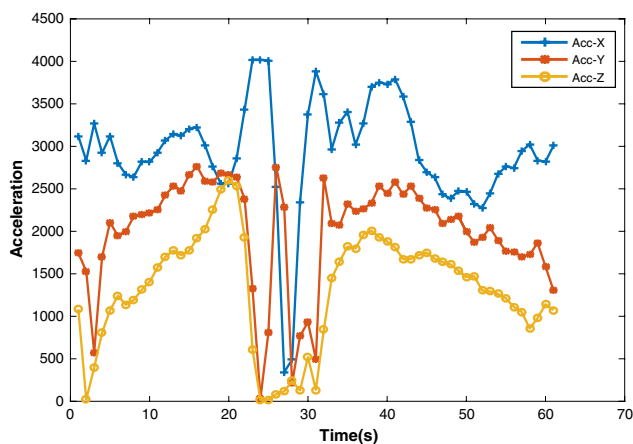
The results obtained using the Right handed dataset are promising compared to those obtained using the Left handed dataset. The good results obtained using the Right

handed dataset can be explained by the fact that the majority of the participants ($n = 7$) were right handed so that behaviors were performed as they normally perform their behaviors. Investigation of the participant errors, in each of the 10 leave one out experiments on the Left handed dataset, revealed that the most problematic behavior instances belonged to participants number 6 and 7. Indeed, by inspecting the behavior classes with high error rate for participant 7, we found that the participant performed the Hitting behavior by rising the hand behind the head and pretend to hit in exactly the same way as the Throwing behavior, while the other participants punch when performing this behavior without rising their hands behind their head as shown in Fig. 6.

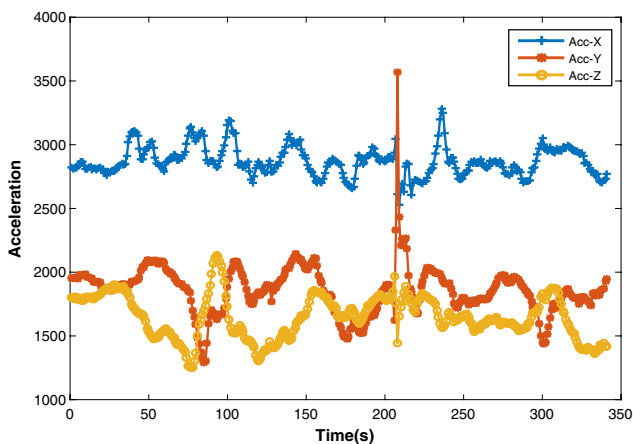
Similarly, participant number 6 performed the throwing behavior with additional movements such as moving left and back while the behavior should be performed only by hands. Moving left and back when performing the throwing behavior created confusions with the Wandering behavior where participants were asked to move forward and backward and left and right. Moreover, the participant performed the tearing behavior by moving the hands forward in the same way as the pushing behavior, and then performed the tearing behavior. This creates a confusion with the Pushing behavior. The variability observed in the ways



(a) hitting



(b) kicking



(c) wandering

Fig. 5 Example of acceleration signals data

participants performed the different behaviors constitutes a good validation setting for our approach. This is demonstrated by the promising results obtained using the Right handed and the Left handed datasets.

Table 2 Average duration of each behavior

Behavior	Average duration (s)
Hitting	2.33
Kicking	2.33
Pushing	2.16
Tearing	3
Throwing	2.83
Wandering	11

Table 3 Percentage of instances of each behavior in each dataset

Behavior	Percentage of instances	
	Right handed dataset	Left handed dataset
Hitting	16.34	16.3
Kicking	16.34	16.31
Pushing	16.34	16.41
Tearing	16.34	16.41
Throwing	16.55	16.41
Wandering	18.06	18.14

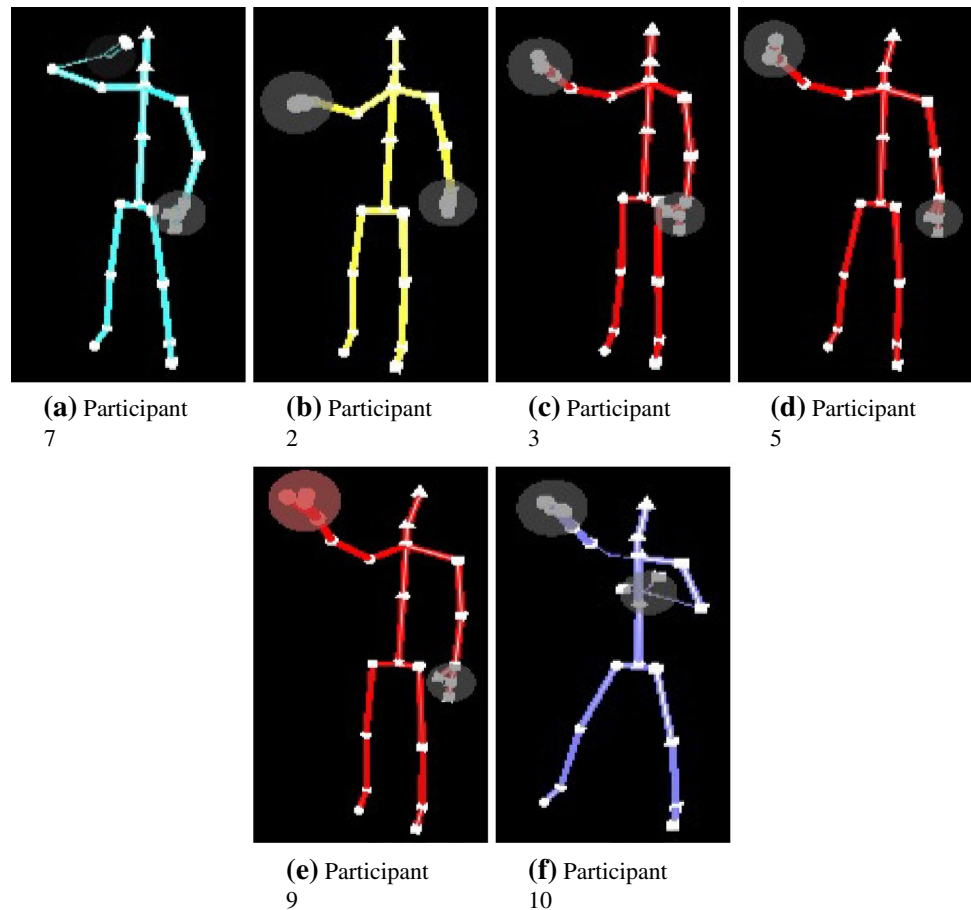
Table 4 Recognition results obtained in the right handed dataset

	Precision (%)	Recall (%)	F-measure (%)
Hitting	100	99.7	99.8
Pushing	100	85.9	92.4
Throwing	87.3	97.3	92.0
Tearing	99.7	91.0	95.2
Kicking	90.0	100	94.7
Wandering	100	99.9	100
Total	97.7	97.4	97.4

Table 5 Recognition results obtained in the left handed dataset

	Precision (%)	Recall (%)	F-measure (%)
Hitting	75.1	98.3	85.1
Pushing	71.2	100	83.2
Throwing	33.4	35.7	34.5
Tearing	81.6	71.5	76.2
Kicking	98.5	71.7	83.0
Wandering	100	100	100
Total	86.9	85.4	85.5

Fig. 6 Example of the Hitting behavior performed by some of the participants



4.3 Comparison with state-of-the-art methods

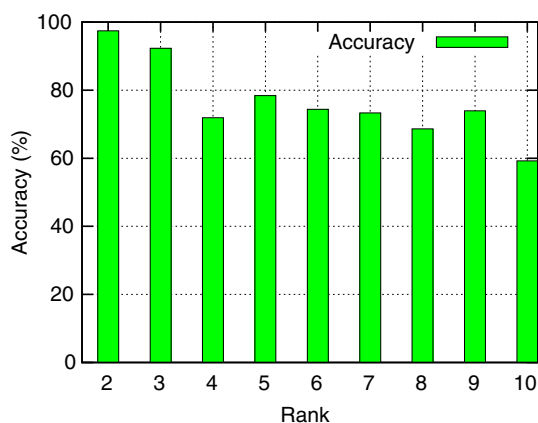
Given that the overall methods proposed in the literature used physiological and acceleration data for aggressive and agitated behavior recognition, and no formal study was proposed to recognize agitated and aggressive behaviors using accelerometer data only, we cannot compare our approach with these methods for the lack of physiological data such as the heart rate. We compared our approach with methods proposed for normal human behavior recognition using acceleration data. The rationale of performing this comparison is that some of the normal human behaviors such as walking, waving hands, clapping hands are fundamentally similar to some of the agitated and aggressive behaviors such as sit down and stand up, clap hands repetitively, wandering and hitting (Masood Manoochehri 2012), which justifies such a comparison. In addition, a comparison with normal human behaviors allows also to validate the genericity of our proposed approach for normal human behavior recognition. We compared our approach with well known approaches for normal behavior recognition in literature. Table 6 summarizes the state-of-the-art approaches and the features and classifiers used for behavior recognition.

The dataset we used for comparison is a human motion dataset (Bruno et al. 2013). The dataset is composed of the recordings of 8 human motions such as climb stairs, descend stairs, getup bed, liedown bed, sitdown chair, standup chair, and walk. Motions were performed by a total of 16 volunteers. The rationale of choosing this dataset is that it contains some actions that are common for people with dementia when they get agitated such as sit down and stand up repetitively (Masood Manoochehri 2012). Figure 7 shows the comparison results obtained for all approaches. We used different experimental settings to compare our approach with the state-of-the-art methods such as 10-fold cross validation, half participant split, 1/3 participant split and 2/3 participant split as recommended by these approaches (Ravi et al. 2005; Bao and Intille 2004; Ermes et al. 2008; Pirttikangas et al. 2006) (Table 7).

The results obtained show clearly the ability of our approach to discriminate between the different behaviors and its superiority compared to the other approaches. As shown in Fig. 7, the only methods that achieve good results also are the method of Bao and Intille (2004) using decision tree classifier with an accuracy greater than 90 %, and the method of Ravi et al. (2005) using the K nearest neighbors classifier with an accuracy greater than 80%.

Table 6 Features and classifiers used by state-of-the-art approaches

Approach	Feature description	Classifier used
Ermes et al. (2008)	Mean, variance, median, skew, kurtosis, 25% percentile, 75% percentile, power of the frequency peak, signal power	Artificial neural network, decision tree
Pirttikangas et al. (2006)	Mean crossing values, correlation between X and Y axis, mean, standard deviation	Multilayer perceptron, K nearest neighbors
Bao and Intille (2004)	Mean, energy, frequency-domain entropy, correlation of acceleration data	Decision tree
Ravi et al. (2005)	Mean, standard deviation, energy, correlation	Decision Tables, Decision Trees, K nearest neighbors, SVM, Naive Bayes
Atallah et al. (2010)	Averaged variance, RMS of signal derivative, mean of signal derivative, average entropy, average cross correlation between each 2 axes, average range, average main frequency of the FFT, total signal Energy, averaged skewness, averaged kurtosis, averaged range of cross covariance, Averaged mean of cross covariance	K nearest neighbors Bayesian classifier

**Fig. 7** Recognition accuracy using different values of NMF rank

An important observation lies in the method of Pirttikangas et al. that employed also a K nearest neighbors classifier, but the results were very low compared to the method of Ravi et al. This can be explained by the set of features employed in each method such as the energy and correlation between X and Z and Y and Z axis features that have not been used in the Pirttikangas et al. method.

5 Discussion

This paper discussed the problem of aggressive and agitated behavior recognition and proposed an effective approach to recognize these behaviors accurately. A non-negative matrix factorization technique combined with an ensemble learning classifier were used to increase the discriminative ability of the extracted features. Experiments were performed on two different datasets: (1) aggressive and agitated behavior dataset, and (2) normal human motion dataset. The recognition results of the proposed approach

were compared with those obtained from four existing state-of-the-art approaches using the normal human motion dataset. The results obtained showed the superiority of our approach over the four state-of-the-art approaches. However, some important choices such as the NMF rank and the sliding window size used in our proposed approach need further explanation.

5.1 NMF rank choice

A critical parameter in NMF is the factorization rank. Choosing the optimal rank for initializing NMF is crucial for the performance of the NMF algorithm. A common way of choosing the rank is to try different values, compute some quality measure of the results, and choose the best value according to this quality measure. In our work, we used the recognition accuracy as a quality measure. Figure 7 shows how the recognition accuracy varies by varying the rank of the NMF technique using the Right handed dataset.

We observe from Fig. 7 that the recognition accuracy is high when the rank of NMF is small (rank = 2 and rank = 3). The accuracy decreases by increasing the value of the NMF rank, which means that the discrimination ability of NMF is higher in low dimensional space. However, the the discrimination ability between the different behaviors decreases when the dimension of the space increases. Besides, performing a NMF factorization with high rank values is time consuming and computationally ineffective. It has been shown that low values of the NMF rank achieved better performance compared to high values (Brunet et al. 2004; Kanagal and Sindhvani 2010). This is also the case in our approach where rank 2 and 3 achieve the best performance. Interestingly though, when the rank of NMF increases the recognition accuracy decreases and stabilises between 70 and 80 %. This is an important observation, which means that when the rank of NMF is between 4

Table 7 Comparison of the recognition accuracy results obtained from the conventional classifiers and our approach

Approach	Classifier	Results							
		10-fold cross validation		half subject split		1/3 split		2/3 split	
		Accuracy	F-measure	Accuracy	F-measure	Accuracy	F-measure	Accuracy	F-measure
Ermes et al. (2008)	Decision tree	76.07	0.76	74.43	0.74	74.29	0.74	74.73	0.74
	Neural network	79.75	0.79	79.95	0.78	78.41	0.77	78.94	0.78
Pirttikangas et al. (2006)	K nearest neighbors	54.15	0.54	53.14	0.52	51.59	0.51	53.84	46.15
	Multilayer perceptron	52.39	0.47	52.35	0.46	52.62	0.48	53.14	0.49
Bao and Intille (2004)	Decision tree	91.12	0.90	91.94	0.90	90.22	0.89	90.68	0.89
Ravi et al. (2005)	Decision Tables	67.79	0.64	64.92	0.58	64.14	0.61	67.21	0.62
	Decision trees	79.34	0.79	76.78	0.76	71.01	0.71	74.73	0.74
	K nearest neighbors	88.44	0.88	86.80	0.86	83.90	0.83	88.27	0.87
	Naive Bayes	72.18	0.70	72.59	0.70	71.92	0.70	71.87	0.69
	SVM	83.12	0.81	80.26	0.77	77.19	0.74	82.40	0.80
Atallah et al. (2010)	K nearest neighbors (K = 5)	71.74	0.71	69.04	0.68	67.78	0.66	70.52	0.69
	K nearest neighbors (K=7)	71.68	0.7	69.09	0.67	67.28	0.65	70.37	0.69
	Bayesian classifier	56.5	0.51	55.72	0.5	56.19	0.51	56.59	0.51
Our approach	Rotation forest	99.31	0.99	98.83	0.98	98.06	0.98	98.89	0.98

Bold values indicate results are better than those obtained by state-of-the-art methods

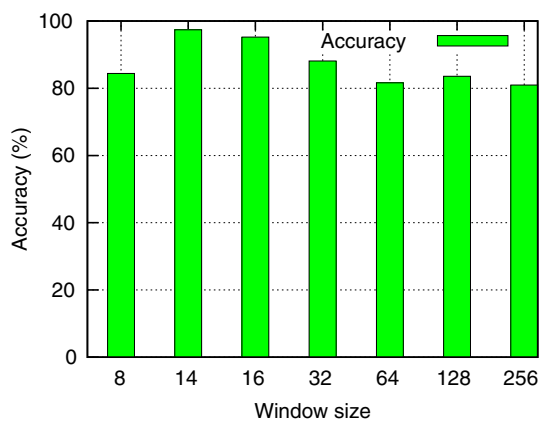
and 10, the projection into the new space does not change the discrimination ability of NMF. This clearly explains the similar accuracies obtained when the rank is greater than 3. This suggests the need for an automatic method that takes into account both the accuracy and the computational complexity in selecting the optimal NMF rank.

5.2 Sliding window size

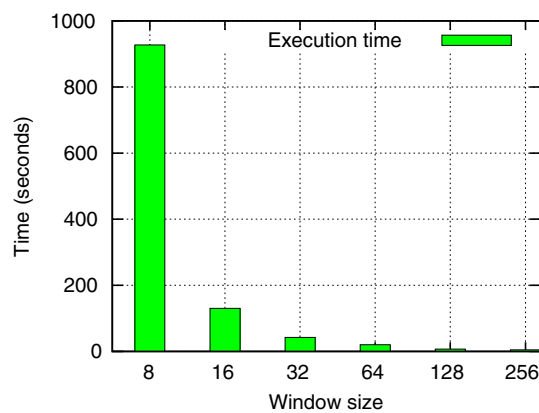
One important parameter in the feature extraction is the size of the sliding window. Indeed, activity classification algorithms typically work with relatively short windows of sensor data in order to improve the classification performance. Short windows generate more training samples and thus increases the performance of the classifier. However, long windows generate less training samples, which can probably decrease the performance of the classifier due to the size of the training data. In addition, in terms of computational complexity, data obtained using short windows requires more computation for training when compared to data obtained using long windows. As a consequence, given the lack of a formal mechanism to automatically choose the optimal window size, a tradeoff between classification performance and computational complexity should be attained. To illustrate how the performance of the classifier decreases when the window size increases, Fig. 8 shows graphically the relation between the window size and the classification accuracy in the Right handed and Left

handed datasets. We used a power of 2 window sizes such as 8, 16, 32, 64, 128, and 256 as used in the literature with 50% overlap. This will help perform a Fast Fourier Transform (FFT) on the data as used by most of the state-of-the-art methods (Ravi et al. 2005; Bao and Intille 2004). As we mentioned previously, the best window size that achieves the best accuracy in our approach was 14. Therefore, we also included this value in graphs.

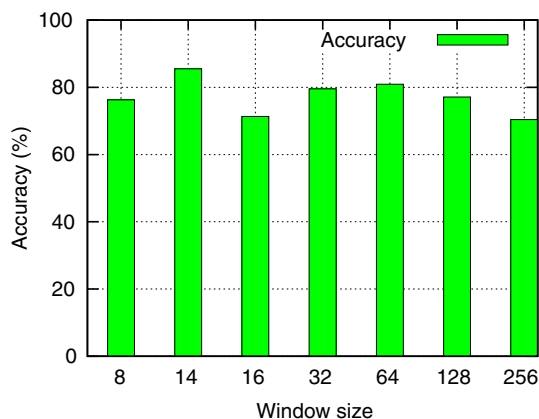
As shown in Fig. 8, the best recognition accuracy was obtained using a sliding window size of 14. The recognition accuracy decreases when the window size increases. In our approach, a short window size performed better due to the duration of each performed behavior which was about 2–3 s except for the Wandering behavior which was about 11 s. However, for long duration activities, this window size may not be the best choice as many confusions may occur between behaviors in addition to the size of the training data. Note that very short window size is not good as shown in Fig. 8. For example, the recognition accuracy was 84.40 in the right handed dataset and 76.28 in the Left handed dataset, which is about 10 % low compared to the accuracy obtained using a window size of 14. Then, the accuracy started decreasing to reach the lowest value with a window size of 256. Note that choosing the window size is data dependent and there is no specific size that works for all datasets. A window with 50% overlap demonstrated success in previous work (Bao and Intille 2004), but no window size was recommended in the literature.



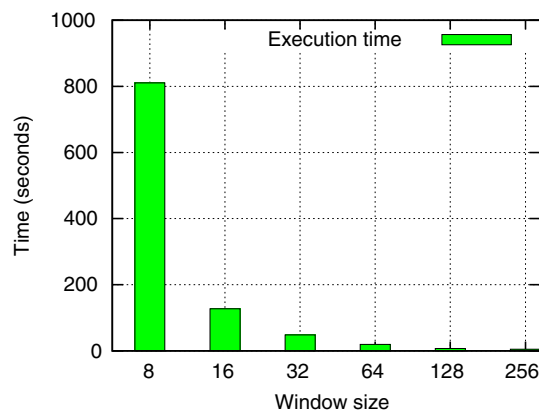
(a) Right handed dataset



(a) Right handed dataset



(b) Left handed dataset



(b) Left handed dataset

Fig. 8 Classification accuracy using different values of window size

Fig. 9 Time execution using different values of window size

5.3 Execution time vs window size

The execution time is an essential part in the development of real time applications. In our work, the execution time depends strongly on the sliding window size. Indeed, a short window size generates more training samples, and consequently more time needed to learn and classify the data. Figure 9 shows the execution time of our approach with different values of window size. A machine with 6 GB of memory and 2.5 GHz processor is used to perform these experiments.

As shown in Fig. 9, more time is needed to learn and classify data generated using short windows such as window size = 8 (927.11 s in the right handed dataset, and 810.39 s in the left handed dataset) and window size = 16 (130.1 s in the right handed dataset, and 127.14 s in the left handed dataset). However, since rotation forest ensemble method can be executed in parallel, and feature extraction can also be performed in parallel, therefore the execution time of our approach can be improved by developing

a parallel version of our approach, which will make our approach practical for real-time applications. This will be considered in our future work.

The findings of this work suggest that automatic recognition of aggressive and agitated behaviors using acceleration data is possible. Although the data was collected from participants in a controlled environment, the good results obtained indicate the benefits of our approach, and constitute a good starting point towards the development of a practical system for aggressive and agitated behavior recognition that can be used for people with dementia. This can be achieved in real settings using accelerometers embedded in wristbands, and by accessing data in real time using bluetooth low energy protocols by deploying our application on smart phones for example. However, to reach our ultimate goal of predicting aggressive and agitated behaviors, it would be interesting in the future to conduct a large scale data collection over a long period of time to collect sufficient data from people with dementia. This will make it possible to analyze the different behavioral patterns associated with aggressive and agitated

behaviors, and to discover some hidden behavioral patterns preceding the occurrence of aggressive and agitated behaviors. These hidden behavioral patterns may be of great importance to uncover the relationships between the different behaviors, which can then be used as predictors of the occurrence of aggressive and agitated behaviors. Consequently, predicting aggressive and agitated behaviors will have a great impact on the management of these behaviors. In fact, predicting the occurrence of these behaviors will allow caregivers and care staff to make early interventions to avoid the occurrence of these behaviors, which will significantly reduce the burden and risks associated with the management of these behaviors.

6 Conclusion

In this paper we have studied the problem of agitated and aggressive behavior recognition. We have proposed an effective approach based on non-negative matrix factorization. Our approach applies first a simple moving average filter to clean the data, then it extracts features from the acceleration signals using a sliding window. After that, a non-negative matrix factorization is used to project the different behaviors into a new space to find a best behavior representation and to increase the discrimination ability of our approach. For classification, we proposed an ensemble method classification based on rotation forest.

We have illustrated the effectiveness and suitability of our approach through extensive experiments on a real agitated and aggressive behavior dataset and common human behavior dataset. The experimental results show the suitability of our approach in representing behaviors and distinguishing between them. In addition, we have also illustrated how our approach outperformed several of the state-of-the-art methods when applied to common human behaviors.

The work we have proposed in this paper constitutes a first step towards the development and deployment of a practical system for the identification of agitated and aggressive behaviors for people with dementia. This in turn, opens new research directions in the ambient assisted living regarding the prediction of the occurrence of agitated and aggressive behaviors in people with dementia, and the issue of big data, specifically with images, videos and audio data, that require efficient and scalable algorithms for processing and management.

References

- Aggarwal J, Cai Q (1999) Human motion analysis: a review. *Comput Vis Image Underst* 73(3):428–440
- Arias-Castro E, Donoho DL (2009) Does median filtering truly preserve edges better than linear filtering? *Ann Stat* 37(3):1172–1206
- Atallah L, Lo B, King R, Yang GZ (2010) Sensor placement for activity detection using wearable accelerometers. In: *Body sensor networks (BSN), 2010 international conference*, pp 24–29
- Azami H, Mohammadi K (2012) An improved signal segmentation using moving average and savitzky-golay filter. *J Signal Inf Process* 3(1):39–44
- Baer T, Moore BC, Kluk K (2002) Effects of low pass filtering on the intelligibility of speech in noise for people with and without dead regions at high frequencies. *J Acoust Soc Am* 112(3):1133–1144
- Bao L, Intille SS (2004) Activity recognition from user-annotated acceleration data. In: *Pervasive computing, second international conference, PERVASIVE 2004, Vienna, Austria, proceedings*, pp 1–17
- Beeri MS, Werner P, Davidson M, Noy S (2002) The cost of behavioral and psychological symptoms of dementia (bpsi) in community dwelling alzheimer's disease patients. *Int J Geriatr Psychiatry* 17(5):403–408
- Bouziane A, Chahir Y, Molina M, Jouen F (2013) Unified framework for human behaviour recognition: an approach using 3d zernike moments. *Neurocomputing* 100: 107–116 (ISSN 0925-2312)
- Brunet JP, Tamayo P, Golub TR, Mesirov JP (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci* 101(12):4164–4169
- Bruno B, Mastrogiovanni F, Sgorbissa A, Vernazza T, Zaccaria R (2013) Analysis of human behavior recognition algorithms based on acceleration data. In: *Robotics and automation (ICRA), 2013 IEEE international conference*, pp 1602–1607
- Cai D, He X, Wu X, Han J (2008) Non-negative matrix factorization on manifold. In: *Data mining, 2008. ICDM '08. Eighth IEEE international conference*, pp 63–72
- Cohen-Mansfield J (1991) Instruction manual for the Cohen-Mansfield agitation inventory (cmai). Research Institute of the Hebrew Home of Greater Washington
- Coronato A, Pietro GD, Paragliola G (2014) A situation-aware system for the detection of motion disorders of patients with autism spectrum disorders. *Expert Syst Appl* 41(17):7868–7877
- Desai AK, Grossberg GT (2001) Recognition and management of behavioral disturbances in dementia. *Prim Care Companion J Clin Psychiatry* 3(3):93
- Duong TV, Bui HH, Phung DQ, Venkatesh S (2005) Activity recognition and abnormality detection with the switching hidden semi-markov model. In: *Computer vision and pattern recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol 1. IEEE, pp 838–845
- Ermes M, Parkka J, Mantjarvi J, Korhonen I (2008) Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Inf Technol Biomed IEEE Trans* 12(1):20–26
- Fook VFS, Thang PV, Mon T, Htwe QQ, Phyo AAP, Jayachandran BJ, Yap P (2007) Automated recognition of complex agitation behavior of demented patient using video camera. In: *9th international conference one-health networking, application and services*, pp 68–73
- Gannot S, Burshtein D, Weinstein E (1998) Iterative and sequential kalman filter-based speech enhancement algorithms. *Speech Audio Process IEEE Trans* 6(4):373–385
- Guo K (2011) Action recognition using log-covariance matrices of silhouette and optical-flow features. PhD thesis
- Hung YX, Chiang CY, Hsu SJ, Chan CT (2010) Abnormality detection for improving elder's daily life independence. In: *Aging friendly technology for health and independence*. Springer, pp 186–194

- Kanagal B, Sindhvani V (2010) Rank selection in low-rank matrix approximations: a study of cross-validation for nmfs. *Reconstruction* 1:1–10
- Kim J, Nakamura T, Kikuchi H, Sasaki T, Yamamoto Y (2013) Co-variation of depressive mood and locomotor dynamics evaluated by ecological momentary assessment in healthy humans. *PLoS One* 8(9)
- Knuff A (2014) Application of actigraphy to the measurement of neuropsychiatric symptoms of agitation in dementia. Master's thesis, Queen's University, Canada
- Krishnan NC, Cook DJ (2014) Activity recognition on streaming sensor data. *Pervasive Mob Comput* 10:138–154 (ISSN 1574-1192)
- Kuncheva LI, Rodríguez JJ (2007) An experimental study on rotation forest ensembles. In: *Proceedings of the 7th international conference on multiple classifier systems, MCS'07*. Springer, Verlag, pp 459–468 (ISBN 978-3-540-72481-0)
- Kwapisz JR, Weiss GM, Moore SA (2011) Activity recognition using cell phone accelerometers. *SIGKDD Explor News* 12(2):74–82
- Lawlor B (2002) Managing behavioural and psychological symptoms in dementia. *Br J Psychiatry* 181(6):463–465
- Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. *Nature* 401(6755):788–791
- Lee DD, Seung HS (2000) Algorithms for non-negative matrix factorization. In: *In NIPS*. MIT Press, Cambridge, pp 556–562
- Li S, Hou X, Zhang H, Cheng Q (2001) Learning spatially localized, parts-based representation. In: *Computer vision and pattern recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE computer society conference*, volume 1, pp 1-207–I-212
- Liu J, Zhong L, Wickramasuriya J, Vasudevan V (2009) uwave: accelerometer-based personalized gesture recognition and its applications. *Pervasive Mob Comput* 5(6):657–675
- Mahlberg R, Walther S (2007) Actigraphy in agitated patients with dementia: monitoring treatment outcomes. *Zeitschrift für Gerontologie und Geriatrie* 40(3):178–184
- Mallidou A, Oliveira N, Borycki E (2013) Behavioural and psychological symptoms of dementia: are there any effective alternative-to-antipsychotics strategies? *OA Fam Med* 1(1):1–6
- Manoochehri M, Huey ED (2012) Diagnosis and management of behavioral issues in front temporal dementia. *Curr Neurol Neurosci Rep* 12(5):528–536
- Moore P, Xhafa F, Barolli L, Thomas A (2013). Monitoring and detection of agitation in dementia: towards real-time and big-data solutions. In: *P2P, parallel, grid, cloud and internet computing (3PGCIC)*, eighth international conference, pp 128–135
- Mori T, Fujii A, Shimosaka M, Noguchi H, Sato T (2007) Typical behavior patterns extraction and anomaly detection algorithm based on accumulated home sensor data. In: *Future generation communication and networking (FGCN 2007)*, vol 2. IEEE, pp 12–18
- Opitz D, Maclin R (1999) Popular ensemble methods: an empirical study. *J Artif Intell Res* 11:169–198
- Paatero P, Tapper U (1994) Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics* 5(2):111–126
- Palmer SE (1977) Hierarchical structure in perceptual representation. *Cognit Psychol* 9(4):441–474
- Pan WD, Yoshida S, Liu Q, Wu CL, Wang J, Zhu J, Cai DF (2013) Quantitative evaluation of severity of behavioral and psychological symptoms of dementia in patients with vascular dementia. *Transl Neurodegener* 2(9):2–7
- Pirttikangas S, Fujinami K, Nakajima T (2006) Feature selection and activity recognition from wearable sensors. In: *Ubiquitous computing systems, third international symposium, UCS 2006, Seoul, Korea, October 11–13, 2006, proceedings*, pp 516–527
- Plötz T, Hammerla NY, Rozga A, Reavis A, Call N, Abowd GD (2012) Automatic assessment of problem behavior in individuals with developmental disabilities. In: *Proceedings of the 2012 ACM conference on ubiquitous computing*, pp 391–400
- Qiu Q, Foo SF, Wai AAP, Pham VT, Maniyeri J, Biswas J, Yap P (2007) Multimodal information fusion for automated recognition of complex agitation behaviors of dementia patients. In: *Information Fusion, 2007 10th international conference. IEEE*, pp 1–8
- Rajasekaran S, Luteran C, Qu H, Riley-Doucet C (2011) A portable autonomous multisensory intervention device (pamid) for early detection of anxiety and agitation in patients with cognitive impairments. In: *Engineering in medicine and biology society, EMBC, 2011 annual international conference of the IEEE*, pp 4733–4736
- Ravi N, Dandekar N, Mysore P, Littman ML (2005) Activity recognition from accelerometer data. *AAAI* 5:1541–1546
- Tractenberg RE, Singer CM, Cummings JL, Thal LJ (2003) The sleep disorders inventory: an instrument for studies of sleep disturbance in persons with Alzheimer's disease. *J Sleep Res* 12(4):331–337
- Rodríguez J, Kuncheva L, Alonso C (2006) Rotation forest: a new classifier ensemble method. *Pattern Anal Mach Intell IEEE Tran* 28(10):1619–1630
- Sakr G, Elhadj I, Huijter HS (2010) Support vector machines to define and detect agitation transition. *Affect Comput IEEE Trans* 1(2):98–108
- Sheng B, Yang W, Sun C (2015) Action recognition using direction-dependent feature pairs and non-negative low rank sparse model. *Neurocomputing* 158:73–80
- Stern TA (2010) *Massachusetts General Hospital handbook of general hospital psychiatry*. Saunders, Elsevier, 6 edition
- Tampi RR, Williamson D, Muralee S, Mittal V, McEnerney N, Thomas J, Cash M, (2011) Behavioral and psychological symptoms of dementia: Part I epidemiology, neurobiology, heritability, and evaluation. *Clin Geriatr* 19(6):2–10
- Xu W, Liu X, Gong Y (2003) Document clustering based on non-negative matrix factorization. In: *Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval*, pp 267–273
- Zhu Y, Chen W, Guo G (2013) Fusing spatiotemporal features and joints for 3d action recognition. In: *Computer vision and pattern recognition workshops (CVPRW), 2013 IEEE conference*, pp 486–491