A hierarchical model for recognizing alarming states in a batteryless sensor alarm intervention for preventing falls in older people

Roberto Luis Shinmoto Torres\textsuperscript{a,\*}, Qinfeng Shi\textsuperscript{b}, Anton van den Hengel\textsuperscript{c}, Damith C. Ranasinghe\textsuperscript{a}

\textsuperscript{a} Auto-ID Lab, School of Computer Science, The University of Adelaide, Adelaide, SA 5005, Australia
\textsuperscript{b} School of Computer Science, The University of Adelaide, SA 5005, Australia
\textsuperscript{c} Australian Centre for Visual Technologies, The University of Adelaide, SA 5005, Australia

\section*{A R T I C L E  I N F O}

\textbf{Article history:}
Received 17 September 2016
Received in revised form 15 February 2017
Accepted 6 April 2017
Available online 10 April 2017

\textbf{Keywords:}
Falls prevention
Hierarchical CRF
Human activity recognition
Bed exit
Chair exit

\section*{A B S T R A C T}
Falls are common among older people, especially in hospitals and nursing homes. The combination of pervasive sensing and statistical learning methods is creating new possibilities for automatic monitoring of activities of hospitalized older people to provide targeted and timely supervision by clinical staff to reduce falls. In this paper we introduce a hierarchical conditional random fields model to predict alarming states (being out of the bed or chair) from a passive wearable embodiment of a sensor worn over garment to provide an intervention mechanism to reduce falls. Our approach predicts alarm states in real time and avoids the use of empirically determined heuristics methods alone or in combination with machine learning based models, or multiple cascaded classifiers for generating alarms from activity prediction streams. Instead, the proposed hierarchical approach predicts alarms based on learned relationships between alarms, sensor information and predicted low-level activities. We evaluate the performance of the approach with 14 healthy older people and 26 hospitalized older patients and demonstrate similar or better performance than machine learning based approaches combined with heuristics based methods.

© 2017 Elsevier B.V. All rights reserved.

\section*{1. Introduction}
Falls among older people in hospitals and nursing homes are a significant problem and a leading cause of unintentional injury-related death for people over 65 in Australia [1] and USA [2]. These are costly events as the hospital expenses alone of a single fall-related injury is estimated at US$35 777 and the direct medical costs of fall injuries in USA is expected to reach US$34.8 billion annually by 2016 (both amounts adjusted for inflation since 2013) [3].

Falls are reported to occur mostly in patients’ rooms (84\%) [4], where unsupervised near bed and chair activities such as bed exits, chair exits and ambulation have been identified as potential activities and locations of falls related injury for older people [4–6]. Current best practice recommendations, which include interventions based on exercise, nutrition, medication and the use of bed or chair exit alarms [7] contribute to falls prevention in hospitals. Nevertheless, falls rates still remain
For example, a 10-year audit on in-hospital falls recorded only at public hospitals in the State of Victoria, Australia, with a population of over 5 million people, determined that falls increased yearly at a rate of about 160 falls per year with more than 21,000 falls recorded in that 10-year period [8]. Recent falls prevention clinical studies in hospital settings [10–12] based on pressure pads to generate an alarm after detecting a bed or chair exit to provide caregivers an opportunity to assist patients found, in general, little to no reduction of falls. One reason may be attributed to “alarm fatigue” [11] due to false alarms. For instance, the long term clinical study of Capezuti et al. [13] showed that approaches using pressure mats alone or in combination with other sensors (infrared beam detectors) were confronted by high false alarm rates (low specificity of 0.3%). This study [13] also found that nursing staff must evaluate the size and body movements of the patients in order to best locate and configure the mat and possibly include other sensors [13], making pressure mat-based approaches more time consuming in practice and potentially cumbersome for nursing staff [14].

Wearable sensors provide new opportunities to monitor patients and develop more effective movement sensor alarm systems for falls prevention [15,16] as continuous motion data can be collected and analyzed in real time. Moreover, as sensors keep reducing in size, more vital medical data can also be gathered from these devices [15,16]. Wearable devices also eliminate privacy violation concerns raised by older people related to the use of intrusive technologies such as cameras [17]. Wearable sensor-based studies have explored the areas of recognition of activities of daily living (ADL) [18–23], falls detection and assessment of the risk or propensity for falls in older people [24–28]. However, the use of this technology in a clinical intervention to prevent falls remains under-explored.

Most studies using body worn sensors for human activity recognition [18–20,22,23,29] used a single or multiple battery powered sensors attached or strapped to the body. This approach can cause discomfort to older people whom have shown a preference for lightweight wearable devices [30,31]; in particular garment-embedded wearable sensors so that they do not need to remember to use them [30]. Moreover, recharging or changing batteries of sensor devices increases the workload of nursing staff as the length of stay of some patients can last several weeks to a month or more. In contrast to the use of battery-powered sensor methods [18–20,22,23,29], we demonstrated in [32], for the first time, the possibility of using a single batteryless, i.e. lightweight, radio frequency identification (RFID) tag with an integrated accelerometer sensor as a bed exit movement sensor alarm with young people.

RFID, as a wireless and unique identification technology, provides the capability to simultaneously monitor multiple patients; moreover, RFID platforms are increasingly being deployed in hospitals for monitoring equipment, patients and personnel [33]. Therefore, integration with existing RFID infrastructure can result in the reduction of operational expenses. Our RFID-based approach for monitoring patients and providing timely alarms to caregivers in a technological intervention to prevent falls is illustrated in Fig. 1. In this proposed intervention, a patient wears the sensor over a loosely worn garment at sternum level, as chest-located sensors [34] are better able to capture upper-body movements to effectively determine postures in older people. The sensor, an accelerometer embedded passive RFID tag in a wearable embodiment (bottom-left of Fig. 1), collects motion information from the patient in a sparse data stream which is processed by our proposed approach for the recognition of alarms of interest associated with high-risk activities, such as getting out of bed, in real time. Hence, once an unsupervised patient performs a given high-risk activity (i.e. exits the bed or chair), our approach recognizes an alarming condition and issues an alarm for hospital staff to quickly access the room to supervise the patient and prevent a fall. To our knowledge, there is no previous literature that uses a batteryless, wireless sensor device with healthy and hospitalized older people for the prevention of falls.

In the context of our clinical intervention, we are interested in discriminating between alarming and no-alarming states of a patient in a hospital setting, where an alarm state corresponds to a bed exit and being out-of-bed or a chair exit and
being out-of-chair whereas a no-alarm encompasses any other state. In our previous studies in [21,32], our approach was to identify posture transitions, for example getting out of bed. The method in [32] was purely heuristics-based with empirically determined threshold levels and produced an output from 20 s segment windows, causing delays in alarm activation and response from caregivers. In [21] we used a two-stage approach where we: (i) predicted an activity label sequence, such as: lying, sitting-on-bed, ambulating from a time series of sensor observations; and (ii) subsequently used a heuristic based empirical method to determine changes from previous and current predicted activity labels to identify alarming activities, for example, sitting-on-bed to ambulating to recognize bed exits. This method [21] achieved better performance than that of [32] when tested on older people. In the present study, our motivation is to formulate the alarm and no-alarm states as high-level activities to be recognized using a hierarchical classifier where the model learns relationships between the alarm and no-alarm states, the current sensor observation and the sequence of predicted low-level activities.

Other studies have also considered the recognition of high-level activities with body worn sensors. For example, walking, running and cycling, in Banos et al. [20] and making coffee, toileting or watching television in Wang et al. [23], were derived from the low-level activities of body or limb movements. In general, methods to determine high-level activities have considered:

- Heuristic methods for the recognition of high level activities [18,32] (see Section 2.1); and
- Multiple stage classifiers [19–23,29] (see Section 2.2).

The use of heuristic methods implies that thresholds or other parameters are empirically determined, i.e. not learned by the model, and calculated or pre-determined for the available testing data [18,23,32]. Often, these parameters must be reevaluated to match the conditions of new testing data. Similarly, for cascaded classifiers [19,20,22], multiple classifiers in sequence are trained where each prediction model requires the evaluation of model parameters (e.g. SVM’s trade-off parameter C and kernel parameters). The use of multiple stages can extend the training time, significantly increasing the time required to find optimal parameters. In addition, most studies do not consider the natural relationships between sequential activities during classification.

Therefore, this article proposes a classification algorithm for learning to recognize in real time high level activities, corresponding to the alarming states of a patient exiting a bed and being out-of-bed, or exiting a chair and being out-of-chair. We expect this model to reduce training time from previous multi-stage classification approaches, reinforce the learning of our alarming states, and eliminate the need to rely on empirically determined heuristic stages where real time inference is required. In this study, our contributions are:

1. We develop a novel hierarchical classifier based on the probabilistic method of conditional random fields (CRF) capable of modeling dependencies of activities in a sequence to learn to predict alarming states instead of using empirically determined heuristic approaches or cascaded classifiers and thus avoiding the use of multiple processing stages. In particular, our hierarchical CRF (HCRF) classifier incorporates the decision function to determine when to generate an alarm by discriminating alarming states—which we refer to as our high-level activities—by constructing relationships between high-level activities, current sensor observation and predicted low-level activities from sensor observations (e.g. sitting, lying, ambulating). Furthermore, inferencing in our HCRF approach is achieved in real time using a sum–product algorithm based method that computes the marginal probabilities for every received sensor data as opposed to inferencing complete sequences as is typical in CRFs.

2. Our learning algorithm utilizes time-domain based information from the accelerometer and RFID data obtained from the RFID infrastructure. This allows the rapid calculation of features in contrast to the use of frequency-domain accelerometer information where data interpolation, given the sparse and irregular reading intervals of our passive sensor data, and transformation to frequency domain are necessary.

3. We evaluate the real-time alarming performance of our approach with cohorts of 14 healthy older people and 26 older hospitalized patients from a geriatrics evaluation and management (GEM) unit since we are interested in reducing falls in hospitals.

2. Related works

This section describes previous methods formulated for human activity recognition. These approaches can be broadly categorized based on their method to recognize activities of interest: (i) heuristics-only approaches; and (ii) multiple-stage approaches, usually consisting of a single or sequence of classifiers followed by a heuristics method for activity recognition. We also describe other studies based on hierarchical approaches using CRFs.

2.1. Heuristics-only approaches

In these methods [18,27,32,35] the activities of interest were transitions such as sitting to standing and standing to sitting [18,27,35] or in-bed to out-of-bed [32]. These approaches used a heuristic decision-tree model based on thresholds where the sensor data has to be interpolated and filtered various times according to the activity of interest; moreover, thresholds and other parameters were empirically determined with the testing population and thus are not learned. None of these methods were implemented in real time as [32], our previous approach evaluated with young adult volunteers,
considered non-overlapping windows of 20 s duration and [18,27,35] used data-loggers and processed each participant’s data in a single batch.

2.2. Multiple stage approaches

These methods [19–23,29,36] usually considered a two-stage solution, where the first stage, a machine learning classifier, predicted low-level activities followed by a second stage for post-processing the prediction stream to determine a label corresponding to a high-level activity. In [21], we considered a two-stage approach where the first stage predicted postures from a sequence of sensor data and the second heuristic-based stage evaluated change of postures from that of being in-bed to out-of-bed to generate an alarm. The studies of Lee et al. [36], Lee et al. [29], Varkey et al. [19], and Wang et al. [23], used cascaded classifier based processes to discriminate between two levels of activities from sensor information, where an identified action, motion or type of movement (e.g. static or dynamic [29], arm moving up or down [19], or stand and walk [36]) is followed by the recognition of the high-level activity of interest (e.g. driving, going upstairs [29]; writing, jogging or walking [19]; and shopping or taking the bus [36]). In Banos et al. [20] multiple binary classifiers were hierarchically ranked and combined, first at individual sensor level and then from multiple sensor sources to recognize an activity being performed.

In the cascaded classifier models of [20,23,36] the high-level classifier waits for the output of the previous level (low-level) classifier in order to perform classification. This can cause some processing delay; for example in [36] the high-level model considers up to 5 min of collected data to make a decision. Other models [19,29] used the sensor data stream for input to the classifiers; however, multiple classification models are trained and used on the high-level classification stage depending on the output of the low-level classifier. In these methods, the high-level activity was determined directly by the high-level classifier [19,29,36]; or a decision stage that decided on the weighted-sum of the outputs of the high-level classifiers where weights were learned; or used a simple heuristics-based score function exceeding a threshold level that was empirically predetermined [23]. Similarly, our approach in [21] used a heuristics method that considered the previous and current predicted activity to issue an alarm.

2.3. CRF-based approaches

Our approach is based on a hierarchical model based on CRFs. Recent approaches using multi-level graphical models, also named hierarchical CRF [37–41] were mostly focused on solving computer vision and imaging problems and did not require real time prediction in their problem formulation as is our case. Moreover, these methods were not intended for sequential time-series data where inference is based on temporal data which is never complete.

Other methods included multilayer approaches such as hidden conditional random fields [42–44] which considers a linear chain model where a layer of variables, in contrast to our approach, is latent and inference of this layer is not required as it is unknown. The study of Chatzis et al. [45] infers labels for two non-latent layers of variables as in our approach. However, these studies [42,45] provide a single classification value for a complete series of observations. As a consequence, these processes [42,45] require the complete set of inputs to perform inference which makes these algorithms unsuitable for our problem which requires a prediction of alarm or no-alarm state for every received sensor observation to meet real time requirements of our application.

In contrast to previous approaches, we develop a method that predicts in real time high level activities in a single classification process without relying on empirically determined parameters or expensive post processing stages. Moreover, all parameters are learned and not determined by testing data. This is the first attempt, to our knowledge, to investigate the use of hierarchical CRFs in the recognition of high-level activities in real time, in our case corresponding to alarming states. We describe the study’s approach and algorithms in the following sections.

3. Description of the proposed intervention

This study’s proposed intervention is illustrated in Fig. 1 where older people use a wearable sensor device called Wearable Wireless Identification and Sensing Platform (W²ISP) attached on top of their clothing and described in detail in Section 4.2. Once the patient exits the bed or chair without supervision, (i) data from upper-body motion; (ii) patient identification; and (iii) radio frequency (RF) information such as RSSI (received signal strength indicator), phase and frequency; are collected via the RFID reader infrastructure, i.e. RFID antennas and readers, and transmitted for high-level activity (alarms) recognition in real time. The focus of this paper is the extraction of features from the received data and the recognition of alarming events using two parallel hierarchical conditional random field classifiers.

The first hierarchical model recognizes the out of bed alarming state (exiting the bed and being out of bed), referred to as bed exit henceforth for simplicity. For example, a bed exit considers a person leaving the bed after sitting or lying on it; however, the change of a person lying to sitting on the bed does not generate an alarm as the person is still on the bed. The second hierarchical model recognizes the out of chair alarm state (exiting a chair and being out of chair), referred to as chair exit for simplicity. The hierarchical models predict an alarm or no-alarm state using an associated confidence level for each
prediction based on the marginal probability of the alarm state. Hence, a high confidence indicates a high likelihood of an occurrence of a high-level activity, corresponding to an alarm signal.

The issued alarm is received by hospital staff, who also wear RFID embedded identification tags and can be uniquely identified by the RFID platform, to access the room and perform necessary intervention to prevent a fall from the identified patient [46]. The RFID name badges allow the system to automatically determine the provision of care to infer when a person is being supervised or being cared for.

4. Data collection

4.1. Participants and environment

In this study, we trialled 14 healthy older volunteers aged between 66 and 86 years old and 26 older patients from the geriatrics ward of an Australian hospital, patients were aged between 71 and 93 years old. The participants had no cognitive impairment, were able to mobilize independently or use a walking aid and signed consent for the trials. The trials for the healthy volunteers were performed in a clinical room which was instrumented to resemble two different clinical settings, called Room1 and Room2 and shown in Fig. 2. The healthy older volunteers were assigned to either clinical setting Room1 or Room2. The setting of Room1, shown in Fig. 2(a), used four RFID reader antennas: one on top of the bed on ceiling level and three at wall level covering the areas a person using the sensor device is likely to walk through, with one of these three antennas facing the chair. In the case of Room2, see Fig. 2(b), two antennas are facing the bed and surrounding areas at ceiling level and one antenna is facing the chair. In the case of the hospitalized patients, the trials were performed in their own hospital rooms. Fig. 2(c) shows a general deployment in a hospital room where three antennas were used: two directed to the area next to the bed and one tilted towards the chair. The chairs were located at either side of the bed. The dimensions for Room3 were not fixed since the experiments were carried out in the rooms occupied by participating hospital patients.

The participants performed a set of predefined activities such as lying on the bed, sitting on the bed, sitting on the chair, walking from bed to chair and chair to bed and walking to and from the door. The participants were informed of the activities to perform and were told to perform these activities as comfortably as possible and no indication was made as to how to do each movement. The duration for the trials for healthy volunteer was of about 90–120 min and about 20–25 min for hospitalized volunteers. Healthy participants performed about 5 trials each and completed on average 2 bed exits and 1 chair exit per trial. Hospitalized participants performed one trial each and completed on average 2–3 chair exits and bed exits, depending on their physical condition. The total number of bed and chair exits performed during trials are shown in Table 1. The data of 3 trials from a single volunteer from the healthy cohort in Room1 and the data of 3 patients from the hospitalized cohort of Room3 were not used due to sensor malfunction and insufficient data collection; hence we use the data of 23 patients in Room3. The data of these trials are available at http://autoidlab.cs.adelaide.edu.au/research/ahr.

4.2. Sensing platform

The sensing technology proposed for our intervention is a batteryless wearable sensor device called \( W^2ISP \) [21,47], see Fig. 1 (bottom left). The \( W^2ISP \) consists of a sensor module embedded in a flexible antenna constructed from C-Foam [47] for comfort of the user and a silver fabric to isolate the device from the person wearing it. The \( W^2ISP \) includes a tri-axial accelerometer (ADXL330) and a 16-bit ultra-low power consumption microcontroller (MSP430F2132). The \( W^2ISP \) powers its components with the energy harvested from the electromagnetic (EM) field produced by off-the-shelf Ultra High Frequency
RFID reader antennas. The accelerometer, in particular, has a minimum full scale range of ±3 g and low power requirement of 180 µA with a supply voltage of 1.8 V. The firmware executing on the microcontroller is an implementation of the ISO-18000-6C air interface protocol [48], so that the sensor can be read by standard commercial off-the-shelf UHF RFID readers. The 10-bit per axis accelerometer data is sampled and embedded in the 96-bit EPC (Electronic Product Code) that also includes a unique identifier [49]. The sensor response is backscattered and subsequently received by an antenna and decoded by an RFID reader.

The RFID reader antennas are powered by an Speedway Revolution reader operating at the regulated Australian RF frequency band of 920–926 MHz and at a maximum regulated power of 1 W. The RFID reader collects information from all antennas and is capable of sending this data in real time via a connection to a local area network to back-end systems for processing.

The passive nature of the device powering also constrains the amount of data collected as sensor readings are produced when the sensor and micro-controller have sufficient power from the RFID reader antennas, this depends on factors such as the distance to the antenna or if there is direct exposure or occlusion to the EM signals from the antennas. Hence, the number of readings obtained from the sensing platform is variable and the sensor data stream is characterized by both noise and sparsity [50].

We are interested in using the three axes acceleration data from the sensor as it contains movement information; previous methods used frequency-domain features (frequency components, energy and entropy) [20,22,51,52], where extracting this information requires processing regularly sampled data or interpolating irregular data. However, we have previously shown in [50] that the combined use of RSSI and acceleration based features in the time domain can improve the performance of a classifier compared to using acceleration and RSSI based features independently, and achieve similar or better performance than features extracted from only acceleration data in time and frequency domain. In addition, extracting frequency domain features produces processing delays; therefore, in this study, we consider time-domain only features from acceleration data and data obtained from the RFID infrastructure, such as RSSI and phase.

We consider the received signal strength indicator (RSSI), which corresponds to the received strength of the signal returned by the sensor. RSSI is an indicator of relative distance to the RFID antenna as a sensor closer to the antenna has higher RSSI readings than a sensor located further away. Previous studies have determined the importance of RSSI for activity recognition [21,32,53]. For example, our study in [21] determined that relative changes in RSSI values can help determine posture variations that may not be noticeable with acceleration information alone.

We can observe the patterns in RSSI and acceleration data in Fig. 3, which shows a bed exit (a)–(b) and chair exit (c)–(d) of a hospitalized person. We can see on all cases that sitting and ambulating postures have scarcer sensor observations than when lying on bed. Fig. 3(a) shows a bed exit, where sitting on bed has very few sensor observations and transferring to ambulating can only be clearly seen in $a_f$ values; however, RSSI values in Fig. 3(b) show that changes for antenna1 can also help determine an exit from the bed. Similarly, Fig. 3(c) shows there is a fast acceleration change when exiting the chair, then acceleration values recover to previous state; however, RSSI values have a clear variation (antenna3) to help discriminate sitting on chair from ambulation. The sine of the body tilting angle (see Section 5.1) follows closely the variations of $a_f$. Other RF information collected is RF phase, which measures the phase angle between a sent RF carrier (frequency channel) and returned signal from the sensor [53].

5. Feature extraction

We are interested in extracting relevant human movement information from the sensor’s raw data as well as the data captured by the RFID infrastructure such as RSSI and phase. We have considered three types of features in this study: (i) instantaneous features; (ii) contextual information features; and (iii) inter-segment features. Since we have exploited information dependent on the number of antennas in the room, we have variable feature vector dimensions $\mathbb{R}^{78}$ for Room1, $\mathbb{R}^{68}$ for Room2 and $\mathbb{R}^{70}$ for Room3.

5.1. Instantaneous features

These features capture information about the actions being performed by a person at a given instant. These features are derived directly from the sensor, RFID infrastructure and user information. The included features are:

- Acceleration readings from three axes: $a_v$, $a_f$, and $a_l$ shown in Fig. 1;
- Sine of body tilting angle in the sagittal plane, $\sin(\theta) = \sin(\arctan(\frac{a_f}{a_v}))$ as in [34];
- Rotational angle yaw = $\arctan(\frac{a_l}{a_f})$;
- Rotational angle roll = $\arctan(\frac{a_l}{a_v})$;
- ID of antenna receiving data ($aID$) [21];
- Received signal strength indicator (RSSI) [21];
- Time difference between observations as in [21];
- Resultant acceleration $^\dagger$, $a_{total} = \sqrt{a_f^2 + a_v^2 + a_l^2}$;
- Gender of the participant.
Fig. 3. Variations of acceleration readings, sine of body tilting angle and received signal strength indicator (RSSI), for a set of activities in the hospital setting in Room 3. (a) Acceleration readings for a bed exit, where readings for Lying on bed are clearly different to those of other activities while readings for Sitting on bed and Sitting on chair are similar. (b) Variations of RSSI showing trend changes for readings captured by specific antennas during transitions between activities. (c) Acceleration readings during a chair exit, also showing similar readings during sitting postures. (d) RSSI variations, mainly from antenna 1, provide trend variation information during changes in posture.

5.2. Contextual information features

These features provide recent temporal context on the action being performed; this is because activities that occurred recently have an impact on the current action as opposed to earlier movements. These are obtained from a fixed time sliding window of 4 s duration. In previous research [54], we found that using this segmentation method produced performance as high as that of other more complex methods for context extraction. The included features are:

- Number of readings per antenna in a segment [54];
- Mutual information from bed and chair area antennas [54] (used antenna 2–antenna 4 in Room 1, antenna 1–antenna 2 in Room 2, and combinations of pairs of antennas in Room 3, due to bed and chair being next to each other);
- ID of antennas receiving higher and lower RSSI from tag responses;
- Displacement in the $a_v$ axis, given by $d_{av} = \int_{t - 4s}^{t} a_v \, dt^2$;
- Mean and standard deviation of three acceleration axes $1$;
- Mean and standard deviation of RSSI from all antennas $1$;
- Pearson correlation between acceleration readings in the time window;
• Total velocity during segment $i$, $v_{i,\text{total}} = \int_{t_{4i}}^{t_{4i+4}} a_{i,\text{total}} dt$;
• Displacement with total acceleration $i$, $d_{i,\text{total}} = \int_{t_{4i}}^{t_{4i+4}} a_{i,\text{total}} dt^2$;
• Standard deviation of variable frequency phase rate (VFPR) $\bar{\sigma}$ as used in [53];
• Standard deviation, median and sum of modulus of constant frequency phase rate (CFPR) $\bar{\sigma}$ as used in [53].

5.3. Inter-segment features

These types of features exploit trends and relationships between two segments. Hence, they provide an insight into the long term variations that are consistent with posture changes.

• Difference of maximum, minimum and median of three acceleration axes from consecutive segments;
• Difference of maximum, minimum and median of sine of body tilting angle from consecutive segments $\dagger$;
• Difference of maximum, minimum and median of RSSI per antenna from consecutive segments.

We performed feature selection using the WEKA data mining tool [55] using the simple classifiers: random forest; the probabilistic models of Bayes network; and logistic regression, to rank the features based on each classifier evaluated with each dataset. Then we eliminated features ranked low across the datasets. In general, we are interested in using or eliminating complete sets of features as we aim to compare all datasets with the same sets of features. For example, we use the values of the 3 acceleration axes in Section 5.1, but we eliminated the mean and standard deviation of these 3 acceleration axes in a segment in Section 5.2 as these were ranked low by the datasets. Features above with $\dagger$ were eliminated for the general evaluation in Section 7.2.

6. Activity recognition

This section provides detail of the formulation of our hierarchical classifiers, describing training, inference and alarm activation methods. From the collected data, let $x_{1:T}$ denote $x_1, x_2, \ldots , x_T$, a sequence of observations of length $T$ associated to two sets of variables $y_{1:T}$ and $h_{1:T}$ that correspond to performed low and high level activities respectively. Low level activities are observable motions such as lie, sit or ambulate, whereas high level activities correspond to an alarm or no-alarm state, which are related to the low level activities. We assume $K$ classes for the low level activities and $H$ classes for the high level activities; hence $y_k \in Y = \{1, \ldots , K\}$ and $h_k \in H = \{1, \ldots , H\}$. The proposed two layer HCRF model is given by the conditional distribution $p(h, y|x)$ and partition function $Z$.

$$Z(x) = \sum_{h_{1:T}} \sum_{y_{1:T}} \exp \left(\sum_{t=1}^{T} \phi_t(h_t, y_t, y_{t-1}, x)\right)$$

We decompose the potential function into a set of smaller potential functions (3) resulting in a loop-free factor graph, which is convenient as is simpler to solve and inference is exact. The factor graph is shown in Fig. 4.

$$\phi_t(h_t, y_t, y_{t-1}, x_t) = \phi_{1,t}(y_t, x_t) + \phi_{2,t}(y_{t-1}, y_t, h_t) + \phi_{3,t}(h_t, x_t).$$

6.1. Training

During training, we aim to maximize the conditional log likelihood $\mathcal{L}$

$$\mathcal{L} = \log p(h_{1:T}, y_{1:T}|x)$$

$$\mathcal{L} = \sum_{t=1}^{T} \phi_{1,t}(y_t, x_t) + \phi_{2,t}(y_{t-1}, y_t, h_t) + \phi_{3,t}(h_t, x_t) - \log Z(x)$$

where $\phi_{1,t}(y_t, x_t) = \langle \theta_1, f_{1,t}(y_t, x_t) \rangle, \phi_{2,t}(y_{t-1}, y_t, h_t) = \langle \theta_2, f_{2,t}(y_{t-1}, y_t, h_t) \rangle$ and $\phi_{3,t}(h_t, x_t) = \langle \theta_3, f_{3,t}(h_t, x_t) \rangle$, here $\theta = \{\theta_1, \theta_2, \theta_3\}$ are parameters to be estimated and $f(.)$ is a boolean value transition or emission feature function. In the case of emission parameters, the trained model assigns a weight parameter for each possible discrete feature value (e.g. for body feature body tilting angle discrete values between $-1$ and $1$ in steps of $0.05$ are considered), for every possible value of variable $y$ in the case of $\theta_1$ and of variable $h$ in the case of $\theta_3$.

An intra-class transition feature function example is $f_{2,t}(y_{t-1} = \text{"Lying"}, y_t = \text{"Sitting on bed"}, h_t = \text{"No Alarm"}) = 1_{y_{t-1}=\text{lying}} \cdot 1_{y_t=\text{sitting on bed}} \cdot 1_{h_t=\text{no alarm}}$, where $1_{\{\}}$ is the indicator function. Finally to exemplify an emission feature function, consider the example of current observation $x_t$ containing a single feature of acceleration $a$, where $a = 0 \text{ g}$, when currently the person is lying, then $f_{3,t}(y_t = \text{"Lying"}, x_t[a] = 0 \text{ g}) = 1_{y_{t-1}=\text{lying}} \cdot 1_{x_t[a]=0 \text{ g}}$. 


tolerance value of 1 \times used in CRF[57], we initialize our model’s parameters to ∼ Graphical model representing our HCRF formulation, indicating inter-class transition potential function Fig. 4.

received variable, yθ in our study we estimate the parameters Giventhat:

emission parameters are calculated similarly. gradients with respect to the inter-class transition parameters φ, the gradient with respect to intra-class transition and emission parameters are calculated similarly.

\[
\frac{\delta \mathcal{L}}{\delta \theta_2} = \sum_{t=1}^{T} f_2(y_{t-1}, y_t, h_t) - \frac{1}{Z(x)} \frac{\delta Z(x)}{\delta \theta_2}.
\]

Given that:

\[
\frac{\delta Z(x)}{\delta \theta_2} = \sum_{h'_1, T} \sum_{y'_{1:T}} \left( \exp \left( \sum_{t=1}^{T} \left( (\theta_1, f_1(y'_t, x)) + (\theta_2, f_2(y'_{t-1}, y'_t, h'_t)) + (\theta_3, f_3(h'_t, x)) \right) \right) \right) \sum_{t=1}^{T} f_2(y'_{t-1}, y'_t, h'_t)
\]

\[
\frac{\delta \mathcal{L}}{\delta \theta_2} = \sum_t f_2(y_{t-1}, y_t, h_t) - \mathbb{E} \left[ \sum_t f_2(y'_{t-1}, y'_t, h'_t) \right]
\]

in our study we estimate the parameters \( \theta = \arg \max \mathcal{L} \) using L-BFGS[56], a quasi-Newton optimization algorithm widely used in CRF[57], we initialize our model’s parameters to ∼0 and stop optimization when the gradient is less than the tolerance value of 1 \times 10^{-10}. Note that in (5) we require calculation of the partition function \( Z \), we have from (1) and (2)

\[
Z(x) = \sum_{h'_1, T} \sum_{y'_{1:T}} \exp \left( \sum_{t=1}^{T} \left( (\theta_1, f_1(y'_t, x)) + (\theta_2, f_2(y'_{t-1}, y'_t, h'_t)) + (\theta_3, f_3(h'_t, x)) \right) \right)
\]

Which can be solved using the sum–product algorithm where messages going forward from the first, \( y_1 \), to the last received variable, \( y_T \), are defined as \( m_{a_1}(y_t) = m_{y_t \theta_2, t+1}(y_t) \). This corresponds to the message propagating from node \( y_t \) to factor \( \phi_{2,t+1} \). Using the sum–product algorithm this message is equivalent to:

\[
m_{a_1}(y_t) = \prod_{j \in N_{y_1} \setminus \{t+1\}} m_{\phi_{j,y_t}}(y_t) \tag{10}
\]

\[
m_{a_t}(y_t) = \sum_{y_{t-1}} \sum_{h_t} \exp \left( \phi_{1,t}(y_t, x) + \phi_{2,t}(y_{t-1}, y_t, h_t) + \phi_{3,t}(h_t, x) \right) m_{a_{t-1}}(y_{t-1}) \tag{11}
\]

where \( N_{y_t \setminus \{t+1\}} \) represents all neighbors of node \( y_t \) with the exception of nodes indexed with \( t + 1 \). From (9) and (11) we have then the expression for the partition function \( Z(x) = \sum_{y_T} m_{a_T}(y'_T) \).

6.2. Inference

We are not interested in the maximum a posteriori (MAP) assignment of labels but in the marginal probabilities for the labels of both variables. However, we separate two different inference processes corresponding to training and validation or testing stages.
During training, we want to obtain the marginal probabilities of variables $h$ and $y$, and to calculate the partition function $Z$. The marginal probability for variables $y_t$ and $h_t$ are given by:

\[
p(y_t | x_{1:t}) = \sum_{y_{t-1}} \sum_{h_{t-1}} p(h, y | x) \tag{12}
\]

\[
p(h_t | x_{1:T}) = \sum_{h_{t-1}} \sum_{y_{t-1}} p(h, y | x) \tag{13}
\]

where $y \sim y_{t-1, t+1}$, similarly for $h \sim h_{t-1}$. As in the case of the partition function, the marginal probability can be solved using the sum–product algorithm. Given that messages from all neighbors to the variable nodes are required, we also calculate messages passing backwards from the end of the sequence given by variable, $y_T$, onto the first element of the chain, $y_1$, and also messages going to the leaves of the chain represented by variables $h_t$. In the case of backward propagation we have:

\[
m_{h_{t-1}}(y_{t-1}) = \sum_{y_{t+1}} \sum_{h_{t+1}} \exp \left( \phi_{1,t-1}(y_{t-1}, y_{t+1}) + \phi_{2,t-1}(h_{t+1}, y_{t+1}) + \phi_{3,t-1}(h_{t+1}, x) \right) m_{h_{t+1}}(y_{t+1}). \tag{14}
\]

Hence the marginal probability for variable $y_t$ in (12) is given by:

\[
p(y_t | x) \propto m_{h_t} y_{t-1} m_{h_{t+1}}(y_{t+1}). \tag{15}
\]

Inference for the cases of testing or validation differs from that of training in that we require real time inference for each received sensor reading as opposed to performing label inference on the complete input sequence. We have previously defined a label inference method using the message propagation from the sum–product algorithm [54]. Given that we only need to infer the current received datum, calculation of the backwards propagation is not necessary to obtain the marginal probabilities of variables $y_t$ and $h_t$ at time $t$, which are given by the expressions:

\[
p(y_t | x_{1:t}) = \sum_{h_{t-1}} \sum_{y_{t-1}} p(h, y | x) \tag{17}
\]

\[
p(h_t | x_{1:T}) = \sum_{h_{t-1}} \sum_{y_{t-1}} p(h, y | x) \tag{18}
\]

6.3. **Alarm activation**

Our HCRF model determines if the current sensor observation corresponds to an alarming state based on considering the confidence level (marginal probability) associated with alarming state predictions i.e. bed exits and chair exits. We consider a confidence model of the form $\mu_c + \gamma_c \sigma_c$, where $\mu_c$ is the mean confidence probability of an alarming state (i.e. bed exit or chair exit), $\sigma_c$ is the standard deviation and $\gamma_c$ is a confidence parameter. Both $\mu_c$ and $\sigma_c$ are determined from the alarming state marginal probabilities of the training data; whereas $\gamma_c$ is a model parameter evaluated during parameter selection (see Section 7.1). Therefore the alarming state for our model at any time $t$ is given by:

\[
\text{alarm}_t \begin{cases} 1 & p(h_t | x_{1:t}) \geq \mu_c + \gamma_c \sigma_c \\ 0 & \text{otherwise.} \end{cases} \tag{19}
\]

This allows us to exploit the confidence level provided by marginal probabilities associated with alarm states to only send alerts for those alarm predictions with high confidence levels. Hence the confidence model parameter allows us to potentially reduce possible false alarms from alarm state predictions where the probability associated with the predicted class labels cannot provide conclusive evidence to generate an alarm.
7. Evaluation

To evaluate our method we consider each room individually as RSSI and phase features depend on the number of antennas and the relative distribution of antennas with respect to the furniture in the room. In addition, people of different demographics perform activities differently, for instance, in contrast to healthy older people, some frail hospitalized patients made several attempts before transitioning out of bed while others preferred to roll out of bed.

In consideration of the actions performed, we determine the sets of labels for the HCRF classifiers to predict as $Y = \{\text{sitting on bed, sitting on chair, lying on bed and ambulating}\}$ and $H = \{\text{alarm or no-alarm}\}$, where the set of labels $H$ corresponds to a bed or chair exit alarm as shown in Fig. 1. The alarming process performance was evaluated with respect to bed and chair exits on all datasets. We evaluate the performance of our hierarchical method and compare it with three baseline methods. The first baseline method uses a multiclass CRF classifier followed by a heuristics based stage to decide an alarm or no-alarm label for each observation. The second method is a multiclass SVM classifier using a one-vs-one approach; and the third method is the multiple-stage classifier of Banos et al. [20], described in Section 2.2. We also demonstrate the use of high-level activity specific feature sets to improve performance with the proposed HCRF classifier; and compare training times required by our HCRF method with the baseline methods to further validate our approach in contrast to using cascaded, multi-stage classifiers.

7.1. Statistical analysis

In this study, true positives (TP) are those correctly identified alarms corresponding to bed or chair exits when: (i) the person is actually exiting or has exited the bed or chair (ground truth); and (ii) the ground truth alarm indicator occurs no more than 5 s after the predicted alarm signal. False positives (FP) are those actions that are incorrectly recognized bed or chair exits (false alarms). False negatives (FN) are those ground truth bed and chair exits that were not recognized by the system (missed).

We evaluate the performance of the proposed system using recall, precision and F-score, the harmonic mean of precision and recall; as these metrics consider the occurrence of errors in relation to TPs. These measurements are defined as: (i) recall (RE) = $TP/(TP + FN)$; (ii) precision (PR) = $TP/(TP + FP)$; and (iii) F-score (FS) = $(2 \times \text{precision} \times \text{recall})/(\text{precision} + \text{recall})$.

We evaluated these metrics using a 10-fold cross validation with around 60% of sequences used for training, and 20% used for testing and validation each. We use a 10-fold cross validation as it allows us to obtain results that are less sensitive to the partitioning of the data. In the datasets from healthy older people cohorts (Room 1 and Room 2), the training, testing and validation subsets contain data from more than one participant, where it is possible that different trials of the same person are distributed in these subsets. However, in the case of older patients (Room 3), given that each patient only performed a single trial due to their frailty, data for training, testing and validation correspond to different patients. Therefore, the testing results are a good indication of the results that can be expected from the HCRF classifier in a real-life deployment. We used the validation sequences for parameter selection i.e. SVM’s C parameter in the range $[2^{-5}, \ldots, 2^{5}]$, regularization parameter for CRF and HCRF in the range $[0, 10^{-5}, 5 \times 10^{-5}, \ldots, 5 \times 10^{-2}, 10^{-1}]$, and confidence parameter $\gamma$ for the alarm activation stage in the range $[0.01, \ldots, 1]$. We chose the parameter that produced highest F-score. We compared results using a two tail $t$-test. A $p$-value ($P < 0.05$) is considered statistically significant.

7.2. Performance comparison

This scenario compares the results of our approach with the three previously mentioned baseline methods. In the case of the first baseline multiclass CRF method, the predicted labels are those of our label set $Y$ and from this output a heuristics based stage sums all previously predicted marginal probabilities in a sliding window of 1 s duration where the first element is the prediction to the last received sensor observation. We then determine the low-level activity with the largest sum in the sliding window as the corresponding label for the last received sensor reading and the corresponding alarming condition. We use this time duration as we have previously determined that the minimum time for a posture transition is about 1.7 s [32]; hence a window size larger than 1.7 s can potentially overlap more than one posture change.

The comparison between our proposed method and the baseline methods is shown in Table 2, for a fair comparison we used the same subset of the features described in Section 5 for all classification models. Comparing results of our HCRF method with the CRF-based heuristics method, for Room 1, the HCRF model has statistically significantly better performance for recall for chair exits ($P = 0.026$) and precision for bed exits ($P = 0.024$), whereas the heuristic method had statistically significantly better performance for recall for bed exit ($P = 0.031$). In the case of Room 2 and Room 3, none of the resulting metrics were statistically significantly different ($P > 0.29$ for Room 2 and $P > 0.10$ for Room 3) compared to our HCRF model. However, in general, the HCRF model provided higher mean performance for bed exits than the heuristics model.

Comparing HCRF results with the SVM multiclass model, for Room 1, bed exit results are higher for the SVM model with statistical significance ($P < 0.028$), whereas chair exits results are not significant ($P > 0.15$). Results for Room 2 and Room 3 indicate an overall higher performance for our HCRF model ($P > 0.06$), except for recall for bed exits in Room 2 where the SVM model has higher performance with significance ($P = 0.014$).

Comparing results with the method of Banos et al. [20], for Room 1 and Room 2, the results are not significantly different ($P > 0.09$ for both rooms). However, our model showed higher mean F-score performance measures for Room 2. In the
The results from the evaluation of the improved and the previous HCRF model are shown in Table 3.

certainly, feature selection can also be employed to determine both chair exits and bed exits. Given that our proposed formulation allows us to influence the learned models by having a specific set of features to describe the high-level activities; we used feature selection methods outlined in Section 5 to craft two feature sets for bed exit alarms and chair exit alarms. Certainly, feature selection can also be employed with the heuristics based baseline model; however, feature selection for the classifier in the baseline method can only improve performance based on the recognition of low-level activities as high-level activities are determined in the heuristic stage. The results from the evaluation of the improved and the previous HCRF model are shown in Table 3.
The improved HCRF increased the performance for chair exits for Room1 and both bed and chair exit for Room2 and Room3. Room1 has an approximate F-score increase of about 7.5% ($P = 0.18$). In the case of bed exits for Room1, the performance did not increase when compared to the use of the original subset of features, hence we use the previous feature set for bed exits. In the case of Room2 and Room3, F-scores for bed exits and chair exits were increased up to 8.5% for both room sets.

7.4. Training time evaluation

We show in Fig. 5 the different total training times, including parameter selection, for each classification method tested in Table 2. Fig. 5 shows that on average our HCRF model takes as much as the CRF-based heuristic model whereas the SVM-based model of [20] takes at most 1400% longer to train than the respective HCRF model. Our model also takes less time to train than the SVM multiclass model for Room1 and Room3; however, for Room2, SVM multiclass achieves the fastest total training time but the model obtained does not translate to better performance. In the case of Room3, which has almost as many sensor observations as in Room2, training times for SVM-based models are almost as long as those for Room1. The training times for SVM-based models are affected by the evaluation of, mainly, two values of $C = \{2^4, 2^5\}$ that together take about 60% and 67% of the total training time for SVM multiclass (in green) and the method of Banos et al. [20] (in blue) respectively.

7.5. Discussion

The proposed HCRF method achieved, in general, similar to higher F-score performance in comparison with those other baseline methods. The HCRF formulation leads to a number of methodological improvements of HCRF compared to our baseline methods: (i) HCRF avoids the use of multi-stage approaches where usually one or more stages are empirically determined heuristic methods with parameters that are not learned, but determined by the user (e.g. size of the sliding window in CRF-based heuristic method); however, parameters for HCRF are learned and determined during validation; (ii) previous studies for human activity recognition with a single body-worn sensor have only considered battery-powered sensor devices, generally testing their approaches with a young population; whereas our research is the first to study the use of wireless, batteryless body-worn sensors in older people and in particular, hospitalized older people; (iii) HCRF was compared with an SVM multiclass classifier and a recent published paper [20] that used a cascaded classification method. HCRF performed better for datasets Room2 and Room3, and HCRF was, in general, faster to train than SVM-based methods as shown in Fig. 5.

However, the overall results are not desirably high enough for a clinical application for the recognition of bed and chair exits. These lower than expected results were driven mainly by the passive nature of the sensor, leading to a lack of observations during ambulation for the three datasets that caused difficulty to appropriately distinguish between sitting and ambulating activities as body trunk postures are similar. In Room1, although the antenna setting covered a broad area around the room, there is a lack of sensor readings when sitting on bed and standing next to the bed as there are few readings in this position and the posture transition of getting out of the bed is not captured. In Room2, which has the highest performance of all datasets, the antennas focused around the bed allow to capture more readings while the person is lying on bed; however, few readings are still captured while sitting on the bed as illumination of the sensor from the antennas are obstructed by the person’s body. In addition, while approaching the chair, older people give their back to the RFID antenna facing the chair and may not face the antenna while sitting on the chair. In Room3, which has the same distribution as Room2 in terms of number
of RFID reader antennas and direction of these antennas, the performance is lower than the other two datasets caused by lack of sensor observations while ambulating and sitting. Moreover, the short distance between bed and chair reduces the number of sensor readings collected while ambulating. Further, body posture of frail patients occlude the sensor as they arch forward when ambulating or sitting.

We can see in Fig. 6 the affect of lack of sensor readings and the similarity of sitting and standing postures on the classifier. Fig. 6(a) and (b) show the confusion matrix from the first layer of the HCRF for Room1 and Room2. Here Ambulating activities were labeled as Sitting-on-bed (≤29%) and Sitting-on-chair (≤14%); these misclassification errors are possible causes for bed exit and chair exit alarm errors as shown in Table 2. These effects are also observed in Room3, shown in Fig. 6(c); in addition, the short distance between the bed and chair cause misclassification between Sitting-on-bed and Sitting-on-chair (≤36%). Moreover, about 80% of Sitting-on-bed labels were incorrectly predicted due to lack of readings while sitting on the bed as ceiling mounted RFID reader antennas were unable to read the sensor when a patient was in a sitting posture and arching forwards. Some patients were lying on bed with raised bed head boards as they read or watched TV during the day and these non-horizontal postures while lying affected Sitting-on-bed label classification.

8. Conclusions

The present study presents a novel hierarchical classification model for a falls prevention intervention that can successfully predict bed and chair exits in real time with better or similar performance than a heuristics based model and SVM-based models, as tested with healthy and hospitalized older people. We have described additional advantages to using our hierarchical classification approach such as the ability to adjust the trained model to the different high level activities of interest and the lower training time required in relation to cascaded classification models. Although the proposed system is not intended to detect falls but to issue an alarm to caregivers when a hospitalized older person is out of the bed or chair, these alarms will allow an attending caregiver to act quickly if a fall has already occurred and prevent consequences arising from a ‘long lie’.

There are however, limitations to be dealt with in future research. For example, the lack of sensor observations obtained from our passive sensor approach. This occurs especially when transitioning between activities, due to the RFID antennas not being able to power the sensor. This inadequate powering of the sensor is caused by the human body occluding the sensor or misalignment of the body worn sensor antenna with the RFID reader antennas. This leads to reduced data while performing some activities (sparse data shown in Fig. 3), affecting the performance from the classifier.

Although we use features, such as RSSI, that depend on the antenna deployment; this is not a limitation in our application context of preventing falls in hospitals as the relative positioning of furniture does not change from room to room. However, changing the settings from, for example, a hospital context to a nursing home context may require gathering new training data and re-training.

The present work has considered a set of activities representative of those performed by patients in a hospital setting to support our wearable falls prevention intervention. As a consequence, the number of activities considered is limited and if extended to monitoring older people outside of a hospital context, such as older people living independently at home, a wider set of activities as well as RFID infrastructure deployments should be considered and evaluated. Future work should consider improving the amount of data collected from the sensor by: (i) exploiting recent lower power consuming sensing components such as the micro-controller and accelerometer to improve the sensor unit by reducing its power consumption; (ii) considering the use of textile integrated antennas for the reducing the size of the sensor unit without decreasing performance [58]; and (iii) improving the illumination of the sensor by RFID antennas and thereby energy harvesting by, for example, changing the location of the sensor from the chest to the shoulder of the person and placing all RFID reader antennas on the ceiling. This deployment strategy is a compromise between collecting more sensor readings and capturing upper body motion information as previously used in [18,34], which we have found is not optimal for a passive sensor deployment. These sensor location changes can reduce occlusions from the participants’ body during activities and ensures the illumination of the sensor from ceiling located antennas.

Other future work in the context of the application should consider the implementation of a randomized control trial (RCT) with a larger cohort of older patients to determine the effectiveness of our approach to reduce the rate of falls in older patients. Moreover, such a long term trial can also indicate the alarming performance of the system over longer periods of
time (during day and night) as opposed to that of short trials, as is the case for our datasets. In addition, we can evaluate the impact of the alarming system on staff and investigate the burden of false alarms to develop a quantitative understanding of ‘alarm fatigue’.

Acknowledgments

This research was supported by grants from the Hospital Research Foundation (THRF) and the Australian Research Council (DP130104614). Ethics approval was obtained from the human research ethics committee of the Queen Elizabeth Hospital (TQEH), South Australia (ethics reference number 2011129). We wish to thank and acknowledge the guidance, direction and facilitation of the studies provided by Prof. Renuka Visvanathan, Director, Aged & Extended Care Services, The Queen Elizabeth Hospital, South Australia, Mr. Stephen Hoskins, and Dr. Shailaja Nair from Aged & Extended Care Services, The Queen Elizabeth Hospital, South Australia.

References


