

Adaptive Probability Scheme for Behaviour Monitoring of the Elderly Using a Specialised Ambient Device

Jonathan Winkley • Ping Jiang

School of Computing, Informatics and Media

University of Bradford

Bradford, UK.

j.j.winkley@bradford.ac.uk • p.jiang@bradford.ac.uk

Abstract A Hidden Markov Model (HMM) modified to work in combination with a Fuzzy System is utilised to determine the current behavioural state of the user from information obtained with specialised hardware. Due to the high dimensionality and not-linearly-separable nature of the Fuzzy System and the sensor data obtained with the hardware which informs the state decision, a new method is devised to update the HMM and replace the initial Fuzzy System such that subsequent state decisions are based on the most recent information. The resultant system first reduces the dimensionality of the original information by using a manifold representation in the high dimension which is unfolded in the lower dimension. The data is then linearly separable in the lower dimension where a simple linear classifier, such as the perceptron used here, is applied to determine the probability of the observations belonging to a state. Experiments using the new system verify its applicability in a real scenario.

Keywords Hidden Markov Model, Fuzzy System, Dimension Reduction, Linear Separation, Elderly Monitoring

1 Introduction

Many behaviour monitoring and anomaly detection techniques have their basic grounding in probabilistic models (Barger et al., 2005, Burgess, 2006, Marques et al., 2011). Behaviours are more likely to be chosen if the observed inputs belong to that state with a higher probability than in others; the correct identification occurs when the observed inputs have optimally defined probabilities of occurring in such behaviours. As with this application, observations in some systems are made up of a combination of observations themselves, resulting in a probability distribution that depends on all members of the input – with observations consisting of a large number of

inputs, the identification of probability values for a state can be a long and complex process. The initial fuzzy method described here greatly simplifies the fusion of inputs to be used as a single belief of a state.

Many techniques have been developed to fulfil the requirement of sensor fusion in the fields of robotics and machine health diagnostics (Sasiadek and Wang, 1999, Aliustaoglu et al., 2009, Hall and Llinas, 2002), using fuzzy logic and incorporating genetic algorithms with neural networks (Sasiadek, 2002). A neural network is ideally suited to the system with a logical or mathematical connection between its inputs and outputs (linear or otherwise). However, the neural network can be flawed if it learns from incomplete or incorrect information, whereas a fuzzy system can draw on human knowledge which is known to be experimentally correct. A fuzzy system also by definition is capable of returning the most probable membership depending on the rule base in the system and therefore gives the best estimate possible for a state.

The observation of a single sensor may have a probability of belonging to many states, but the composition of the observations of many sensors will identify the exact state. This characteristic is true for behaviour monitoring applications which incorporate multiple sensors, such as that of (Zhang et al., 2011) and that described here. The technique uses fuzzy logic to determine a single observation probability from multiple sensors which each have a different probability of belonging to the observation. The advantage of this implementation is that it incorporates a higher element of human knowledge where a system trained on probabilistic data cannot, thus it becomes a more efficient model with the optimality criterion matching that of human reasoning – which in many applications is what is strived for.

The case-driven system documented in Zhou et al (2011) note that the elderly are prone to such problems as

memory loss, resulting in possible error scenarios which could be detected if current sensor observations and their emitting state are compared with normal “user activity patterns”; further acknowledging that to understand user needs any intelligent monitoring system needs to be adaptive and tailor its reactions accordingly. The system hardware and operational inputs are similar to those employed in the system this work is intended for, with wearable nodes gathering direct data and the model inputs being in a high-dimensional vector format. The states (or “cases”) are identified through rules which are generated based on these high-dimension vectors, with experiments returning rule bases containing upwards of 800 rules. Whilst presumably the storage space required is of no consequence in this case, for the application here a more concise approach is desired to reduce complexity in state determining.

Elderly monitoring systems can be categorised to 2 variations: autonomous problem-determining and human problem-determining. Whilst the former category is populated with devices such as that described here and those of Zhou et al (2011) and Avci and Passerini (2012) – requiring only the gathered data to infer a belief regarding the users’ state – the latter category has the need for an element of further human involvement in order to assess the status of a user. Such applications similarly utilise environmentally-located sensors or body-worn nodes (Ferreira and Ambrosio, 2012, Venkatesh et al., 2012) to gather readings relating to the user, before uploading them to some “server” which is accessible by a healthcare professional or some other monitoring service that can identify any issues being faced by the user. These systems have a lower level of processing involved and as such require much less consideration for adaptation to the user, given that storage of the observations in their raw form is usually required and inference of a behaviour or state is made by a human supervisor. These applications could be seen to benefit from the dimension reduction scheme detailed in this work, making it easier in certain cases for the supervisors to identify anomalous observations that may be indicative of a health-status change.

Lee et al. (2008) discuss the problems of data fusion in such health applications, yet their proposed method to solve the issues is only broadly discussed. The method contained herein is a form of fusion as it takes five sensor-obtainable readings and infers a single useful output. The fusion is governed by a human-created rule base, whereas other methods may use mathematical calculations for interpretation of a collection of readings. Whilst being a somewhat deterministic approach in this case, the fuzzy system is widely used in many applications to provide

best-estimates of a state based on multiple input values. Abbod et al. (2001) survey an extensive number of applications within the healthcare field which utilise fuzzy technology: including one which applies fuzzy logic and knowledge bases to determine an Intensive Care Unit (ICU) patient’s physiological state at regular intervals. This type of application exhibits the effectiveness of fuzzy systems in the determining of states from all available data; their use in a healthcare environment (where intelligently reasoned decisions are of utmost importance) go further to proving that in the first instance a fuzzy approach is one which more than adequately approximates that of human reasoning.

After the initial stages of use of the hardware by a consumer, the values within the fuzzy system may in fact be irrelevant or describe them to a lesser degree than the values which are now being obtained through observation with the sensors. In this case, the fuzzy system requires updating or replacing to ensure that the values used to infer a belief are always those which best apply to the user. Whilst several methods have been researched and developed to modify a fuzzy system based on testing and training data (Hong and Lee, 1996, Wang and Mendel, 1992, Shiqian et al., 2000), it was decided that the modification would be approached in a more mathematical manner that would prove more beneficial for both eventual visualisation of states (graphically, for debugging and analysis) and simplicity of use during operation. The fuzzy system works well as a starting point with which to gather the required information, as the human knowledge used for selecting states succeeds in correlating observations which have values in common and can therefore be mathematically identified as similar. The new method however completely replaces the fuzzy system for the determining of initial state probabilities once trained on the data that the system initially obtained, due to the high dimensionality of the data and the complications which arise when updating fuzzy rules autonomously and without any human involvement.

The resultant scheme takes as its input the original high dimension data, passes it through for dimension reduction where it is then sent to a very simple set of classifiers to determine its probability of membership to a state. Once trained, during real time operation the method takes only a few operation cycles to output a value which is then used by the HMM to conclude the most likely state that the user is exhibiting.

The paper is structured as follows. The hardware with which this method is intended to be used is briefly described in the next section. The workings of a standard Hidden Markov Model are presented in section 3 along

with the adapted scheme utilizing the fuzzy system. Section 4 discusses the subsequently implemented dimension reduction scheme with section 5 detailing some results from the simulation of the entire system, before concluding remarks in section 6.

2 Application Hardware

The developed application is a body-worn wireless health care system for the elderly generation, consisting of a sensor-filled mobile phone-type base station (“Verity”) and a direct monitoring device with four sensors that are employed to provide a total of 5 different observation values which help determine the current state of a user. The direct monitoring device, or “Wrote” (*wrist mote*), is worn on the wrist like a watch where the sensors are aligned in their required positions. On this device are two temperature sensors: one contained within the control chip (an ultra-low power Sensium CC981 from Toumaz), and one connected externally and in contact with the user. The internal sensor detects the ambient temperature of the environment, while the external sensor reads the temperature of the user. The chip itself contains an embedded 8051 microprocessor and radio transceiver for communication (based on the Toumaz Nano Sensor Protocol, NSP) with the Verity base station.

The Wrote also contains a 3-axis accelerometer from which can be determined the orientation of the device. The readings give independent values for all 3 axes and can infer the overall acceleration experienced by combining these values. Some detection occurs on the device before the readings are even transmitted to Verity: the Wrote is also capable of detecting sharp spikes in acceleration within milliseconds of occurrence and thereby has the possibility of detecting falls almost as soon as they occur. This scenario also triggers verbal communication to assess the situation and even a dial-out to an external contact if required.

The pulse of the wearer is determined using a piezo pressure sensor housed within the device which is always in contact with the radial artery. Deflections of the skin’s surface during blood flow translate to deflections of the piezos surface; the electrical charge generated by a deflection is translated to a reading of approximate bpm. Although the piezo pressure sensor is not as reliable for bpm accuracy as others incorporating more complex electrocardiogram (ECG) techniques - such as the use of multiple electrodes (Lo et al., 2005, Sopavanit et al., 2009, Mukala et al., 2010, Fulford-Jones et al., 2004) - the brief for the system requires that the device be self-contained

and low-power, so the single piezo strip was identified as the best choice.

To take into account the motion of the user which can significantly affect the reliability of the pulse reading, a Kalman filter is employed with a peak detection algorithm used to extract the most likely “spikes” which indicate a pulse. The peak detection algorithm uses a blind period during which it will ignore any deflection values of the piezo strip; a typical heartbeat is between 1 and 2Hz, so once a significant pulse is detected it is reasonable to say that it is only necessary to begin looking for another pulse after a given time. The blind period is adaptive and depends on the Kalman filter’s estimation of the noise value affecting the pulse rate reading, which is in turn based on the acceleration value returned from the 3-axis accelerometer on the Wrote. Using this approach, the interference from the motion is overcome and provides a reasonable estimate of the user’s pulse rate.

Within the base-station are numerous modules which facilitate the operation of the system. Another Sensium chip is used primarily to receive the 5 observation readings from the Wrote (contact and ambient temperatures, device orientation, acceleration and pulse) but it is also capable of transmitting data back when necessary. The readings are communicated to an on-board control chip from Microchip which processes the data as required using the methods described herein, along with controlling the operation of the system. The base-station also contains an internal temperature sensor and accelerometer which are used to further infer a belief about the state of the user when the readings are compared with those from the sensor device.

Communication with the user is purely through voice recognition and gesture control as there are no interface buttons on the base-station. In the event of a detected “situation” - when readings indicate a problem - the base-station interacts with the user to ascertain if there is a need to call for help or update the detection methods to incorporate a new (previously unseen) state. If the user either does not respond or requests assistance the GSM module on the system engages the installed SIM card and proceeds to call one of the stored contacts. These contacts could be family members, neighbours or an accommodation warden; the emergency services may also be contacted if necessary. The device interacts with the receiver of the call before putting them in verbal contact with the user through the on-board speaker and microphone. Using the GPS module, the contact can also be informed of the user’s location.

The system is intended to be as elder-friendly as possible: with no buttons and simple verbal communication the complexity of using such a device is

intended to be minimal – it is a monitoring system and as such should be no more obtrusive than being visited by a carer.

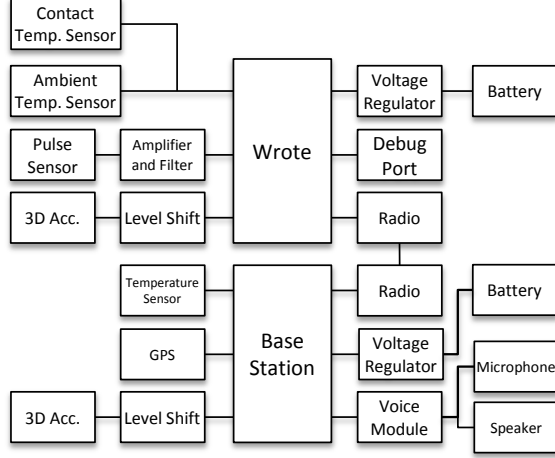


Fig. 1 Basic architecture of the Verity system and interaction of its modules

There are currently 5 distinct basic states which the user can be seen to be in: Sleeping, Sitting, Standing, Walking and Running. These states were identified as being the core states exhibited by a user, given that any other subsequent states such as watching television, cleaning or washing-up can be seen to be variations on these. The system will incorporate a method to detect instances where the behaviour is abnormal when compared to a typical behaviour sequence - such as that implemented by Fine (Fine, 2009), through some form of distance measure - thus identifying possible dangerous events being experienced or the onset of irregularities in a behaviour pattern. This aspect of the implementation is to be addressed in future work however, and is therefore beyond this paper's scope. The state of the user is inferred with the combined Hidden Markov Model and fuzzy system described here. The sensor readings are processed to determine the most likely state outcome at a time instance based on current values and the sequence of states previously observed.

3 Hidden Markov Models

The traditional Hidden Markov Model (HMM) consists of five key components, which will be referred to and explained with notation consistent with Rabiner (1990), for parity. There are N states (S) into which M observations (V) can belong, with probabilities defined by a probability distribution $B = b_j(k)$, where j is the current state and k the observation number. The probability of transitioning from one state to another is an element in a state transition probability distribution matrix which is

defined as $A = a_{ij}$ where i is the current state and j the proceeding state (the state after the transition). The final element of the model is termed π : the initial state distribution, which gives the probability of seeing any state at the first time instance. The elements of the model are defined thus:

$$S = \{S_1, S_2, S_3, \dots, S_N\} \quad (1)$$

$$V = \{V_1, V_2, V_3, \dots, V_M\} \quad (2)$$

$$a_{ij} = P(q_{t+1} = S_j | q_t = S_i) \quad (3)$$

(Note that q_t is the state at time t)

$$b_j(k) = P(V_k \text{ at } t | q_t = S_j), \quad 1 \leq j \leq N \quad (4)$$

$$\pi_i = P(q_1 = S_i), \quad 1 \leq j \leq N \quad (4)$$

The model is denoted in compact form as:

$$\lambda = (A, B, \pi) \quad (5)$$

Within this application the HMM must be able to solve 2 of the basic problems it was developed for (Rabiner, 1990):

1. Given an observation sequence $O = O_1, O_2, \dots, O_T$ and a model λ , how can $P(O|\lambda)$ be efficiently calculated?
2. Given an observation sequence $O = O_1, O_2, \dots, O_T$ and a model λ , how can an optimal state sequence $Q = q_1, q_2, \dots, q_T$ be chosen to best explain the observations?

Therefore, what is the probability of the user producing those readings from the sensors, and what sequence of states must the user exhibit to explain this sequence of observations?

3.2 Calculating sequence probability

For the solution to problems 1 and 2, the Forward-Backward (FB) procedure (Baum and Eagon, 1967) is used:

$$\alpha_t(i) = P(O_1, O_2, \dots, O_T, q_t = S_i | \lambda) \quad (6)$$

I. Initialise:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq j \leq N \quad (7)$$

II. Inductive step:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1 \quad (8)$$

$$1 \leq j \leq N$$

III. Terminate:

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i) \quad (9)$$

Thus by induction can be found the probability of terminating in state S_i at time t having been presented with the observation sequence and the model.

3.3 Calculating an optimal sequence

When the state sequence is to be determined, the probabilities of proceeding states from any point in the sequence to the end of that sequence must be taken into account. For this problem the backward part of the FB procedure is calculated:

$$\beta_t(i) = (O_{t+1}, O_{t+2}, \dots, O_T | q_t = S_i, \lambda) \quad (10)$$

- I. Arbitrary Initialisation (the terminal value is always 1, given its probability of occurrence):

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (11)$$

- II. Inductive step:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (12)$$

$$t = T-1, T-2, \dots, 1,$$

$$1 \leq i \leq N$$

This calculation aids in the finding of an optimal state sequence for the given observations, yet the definition of “optimal” is open to interpretation. A state sequence may consist of states which are most likely at each time step given the observation sequence - regardless of the possibility of the state sequence occurring. It may also be a sequence which logically flows from one state to the next, i.e. takes into account the probability of transitioning from the previous state to the current, along with the observation sequence. Using the FB algorithm the probability of being in a single state at a time, given the observations and model is defined as:

$$\gamma_t(i) = P(q_t = S_i | O, \lambda) \quad (13)$$

Given that $\alpha_t(i)$ accounts for the observation sequence O_t to O_t , and $\beta_t(i)$ accounts for the remaining O_{t+1} to O_T , the above equation can be written in terms of the Forward-Backward variables; the denominator is a normalisation factor which makes the sum of state probabilities total 1:

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N \alpha_t(i) \beta_t(i)} \quad (14)$$

Taking the maximum value of (

(14) therefore gives the individually most likely state at that time:

$$q_t = \operatorname{argmax}_{1 \leq i \leq N} [\gamma_t(i)], \quad 1 \leq t \leq T \quad (15)$$

Another method, the Viterbi Algorithm (Viterbi, 1967), takes into account the likelihood of state transitions in sequence, unlike the previous method. In this property it can be seen to have globally optimised the output, using all available information from within the model. Therefore the resulting state sequence is entirely possible given the observations. However, the algorithm adjusts the entire sequence to match the most likely state at the time. If the next observation most likely belongs to a state which it is unlikely to reach from the current state, the backtracked sequence may change to accommodate it and increase the likelihood of the sequence. What is being determined can be expressed as $P(Q, O | \lambda)$: the probability of seeing the state sequence *and* the observations given the model.

To identify the most likely sequence, a method of back-tracking a maximum probability route is necessary. This is facilitated through use of the array ψ which is populated alongside the probability calculations.

- I. Initialise:

$$\delta_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N \quad (16)$$

$$\psi_1(i) = 0 \quad (17)$$

- II. Recursion Step:

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad (18)$$

$$2 \leq t \leq T$$

$$1 \leq j \leq N$$

$$\psi_t(i) = \operatorname{argmax}_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad (19)$$

$$2 \leq t \leq T$$

$$1 \leq j \leq N$$

- III. Terminate:

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)] \quad (20)$$

$$q_T^* = \operatorname{argmax}_{1 \leq i \leq N} [\delta_T(i)] \quad (21)$$

- IV. The backtracking procedure:

$$q_{t+1}^* = \psi_{t+1}(q_{t+1}^*), \quad (22)$$

$$t = T-1, T-2, \dots, 1$$

The formulas given in this section grant the model the ability to calculate states and observation probabilities at the time specified by t . The HMM parameters used above have been assumed to conform to traditional values for such a model (matrices and standard probability distributions). The next section details the modifications made to enable multi-value observations in an observation sequence, where human linguistic knowledge determines each observation’s membership to the states.

4 Fuzzy Fusion of Inputs

The application of a Fuzzy Inference System (FIS) in HMMs is not uncommon. Kelarestaghi et al. (2001) describe an adjusted HMM which modifies all of its algorithms to utilise variations on fuzzy MIN and MAX operators. For example, the usual Forward algorithm's summing and multiplying induction step now takes values determined by the lower probability – that of transition or seeing the observation in a state. Methods of incorporating additive and non-additive fuzzy systems in HMMs are shown in Verma and Hanmandlu (2009), using fuzzy re-estimations of the Baum-Welch algorithm for the determining of the HMM parameters. For virtual reality training of Bone Marrow Harvesting, de Moraes and dos Santos Machado (2004) discuss a fuzzy approach to return a value of membership of an observation sequence to a state.

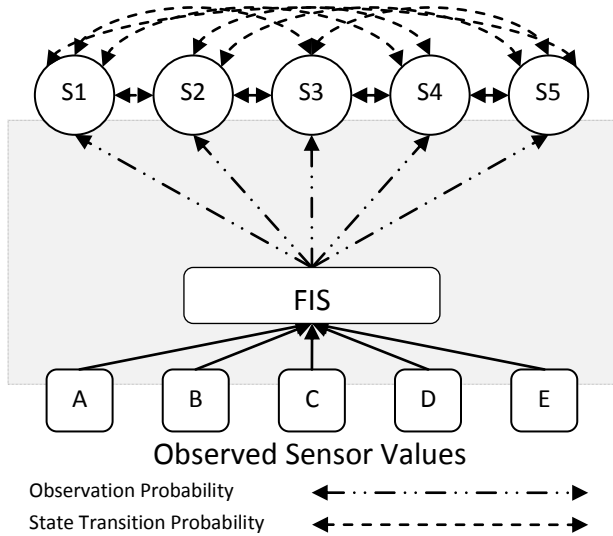


Fig. 2 The FIS interaction within the HMM. The grey shaded box is where in the traditional HMM one would find the possible observations and their output probabilities from the states. In this model, the observations are input to the FIS to obtain the observation probabilities for each state

The usual observation sequence in an HMM is of the form $O = O_1, O_2, \dots, O_T$, where O_i will usually be a single random variable to which a probability of occurrence in a state S_i is assigned $b_i(O_k)$. The problem arises when the observation is a combination of many continuous values, such as that of Verity's multi-sensor system. Assigning probability values to each possible combination of sensor readings is a laborious task, after which there needs to be a method of determining the single probability value for that state. Continuous observations have been considered in HMMs previously (Rabiner, 1990), with a continuous

probability distribution replacing the usual matrix of equation (4).

$$b_j(x(t)) = P(x(t) | q_t = S_j) \quad (23)$$

The FIS is solely applied to the $b_j(k)$ values in place of the usual matrix or distribution, making it a system independent of - yet incorporated into - the HMM. The parameters of the fuzzy system depend on the number of states in the HMM, as it determines the number of linguistic knowledge rules needed which govern the memberships. Rather than the HMM consulting the observation matrix or distribution, it consults the FIS to return the probability of seeing the combination input sensor values in that state.

For each of the Z sensors making up the observation V there should be G membership functions (f) in which a reading from that sensor can belong. For each of the N states there must be at least 1 rule to determine the activation given the sensor readings. The number of membership functions within each sensor's range is determined by a combination of human knowledge of the application and required accuracy, but in this example only a small number of memberships are used for simplicity. Firstly the range of readings possible from the sensor must be split into G individual, human determined, ranges according to the application (e.g. an ambient temperature sensor may return values belonging to 3 ranges of "cold", "normal" or "hot"). For all sensors within the system the formula for their G membership functions can be generalised thus (actual memberships are application specific and can be drawn from experience and testing, or relevant data provided by a healthcare professional in the case of Verity's behaviour monitoring):

$$B_L = \text{lower bound of sensor range}$$

$$Q_{(i)C} = \text{key value of individual range (median)}$$

$$B_U = \text{upper bound of sensor range}$$

$$f(1) = \int_{B_L}^{Q_{(1)C}} (1|x) + \int_{Q_{(1)C}}^{Q_{(2)C}} \left(\frac{Q_{(2)C} - x}{Q_{(2)C} - Q_{(1)C}} \right) x \quad (24)$$

$$f(i) = \int_{Q_{(i-1)C}}^{Q_{(i)C}} \left(\frac{x - Q_{(i-1)C}}{Q_{(i)C} - Q_{(i-1)C}} \right) x + \int_{Q_{(i)C}}^{Q_{(i+1)C}} \left(\frac{Q_{(i+1)C} - x}{Q_{(i+1)C} - Q_{(i)C}} \right) x \quad 1 \leq i \leq (G-1) \quad (25)$$

$$f(G) = \int_{Q_{(G-1)C}}^{Q_{(G)C}} \left(\frac{x - Q_{(G-1)C}}{Q_{(G)C} - Q_{(G-1)C}} \right) x + \int_{Q_{(G)C}}^{B_U} (1|x) \quad (26)$$

For each individual sensor this will give G membership functions which can be descriptively tagged with names such as "low", "medium" and "high". The advantage of using this method to describe a sensor range is that for a

fuzzy definition only a single sensor reading which typifies the membership ($Q_{(ij)C}$) needs to be known as it will become the centre point of the function; for example a membership of “hot” body contact temperature may be centred on 32° but as the system is used increasingly, the membership may be seen to be more relevant when centred at 30° , therefore the function can be shifted yet maintain the same shape characteristics. This inclusion of an element of human reasoning in the system gives a greater degree of accuracy in the estimation of observation probabilities, as the state definitions are themselves based on human knowledge. With the shape of the membership function in place, the key values describing the membership (i.e. its median) can easily be modified from user to user.

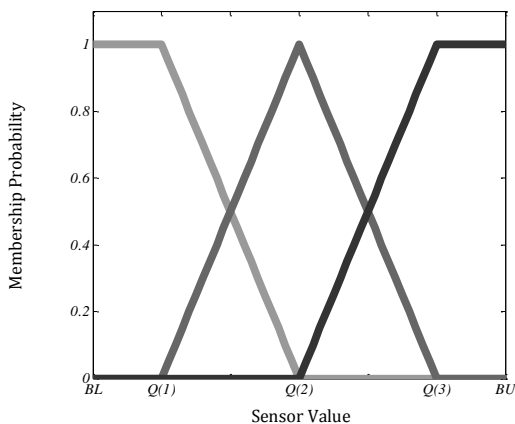


Fig. 3 An example of 3 membership functions for a single sensor

The rules which activate each state are ideally constructed from linguistic rules provided by a physician for each user. It is reasoned that they will have a greater idea of what readings each user should exhibit whilst in the states defined in the model. In this way, they may also be able to view the data obtained by the system and have a greater understanding of the patient’s condition at the time. In the first instance a general rule base can be formed to describe the activation values required for each user’s observable states.

Once the fuzzy rules are linguistically defined for each state, the degree of activation μ_B of each sensor membership μ_N in the FIS is considered using Mamdani’s min-operation method (Mamdani and Assilian, 1975).

$$\mu_B = \mu_1 \wedge \mu_2 \wedge \dots \wedge \mu_n, \quad 1 \leq n \leq Z \quad (27)$$

With Z sensors, $C(i)$, a maximum of Z conditions make up a rule for one state observation probability value $b_j(k)$. Therefore the probability of a sensor producing the corresponding observation value O_i in that state, expressed in HMM terminology:

$$b_j(k) = \min_{1 \leq i \leq Z} [P(O_{ki} | C(i), q_t = S_j)], \quad \begin{aligned} &1 \leq j \leq N \\ &1 \leq t \leq T \\ &1 \leq k \leq M \end{aligned} \quad (28)$$

The HMM’s transition and starting probability matrices can be updated with a method such as the Baum-Welch algorithm (Baum and Eagon, 1967), but as the observation probability matrix has been replaced there is a requirement to develop a more appropriate scheme which allows for updating and adaptation during use.

5 Classification through dimension reduction

With a sufficient set of test data obtained using the fuzzy system, the observation probability determining mechanism can be modified so that the model parameters better suit the user. If the original method remained unmodified, over time the user’s inferred state might differ from the correct one due to changes in transition likelihoods and observation probabilities that are prone to occur as a user’s health changes. Particularly in the case of an elderly user, a period of ill health may significantly alter their general speed of motion or the time required to undertake an activity – retaining the original parameters could result in serious misrepresentation of the user’s daily state sequence.

Whilst numerous methods have been researched and developed to modify a fuzzy system based on testing and training data (Hong and Lee, 1996, Wang and Mendel, 1992, Shiqian et al., 2000), it was decided that to approach the observation probability determining stage in a more mathematical manner would prove more beneficial for both eventual visualisation of states and simplicity of use during operation. The fuzzy system works well as a starting point with which to gather the required information, as the human knowledge used for selecting states succeeds in correlating observations which have values in common and can therefore be mathematically identified as similar.

With a multidimensional data set there is a considerable chance of nonlinearity between those clusters present, however for data sets of a very high dimension it is often difficult to detect the nonlinearity. A number of techniques exist to address the task of linearly separating such data for classification (Pearson, 1901, Cortes and Vapnik, 1995) whilst others attempt to classify in a high dimension using newer variations on well established methods (Wang, 2011, Li et al., 2012); however it is a case of trial and error when it comes to assessing the reliability of the method and its linearly separated data representation.

Similarly, a common approach to identifying and utilising trends in nonlinear data is through use of a back propagation neural network – in this application one might be used with a sigmoid function to provide probability as an output rather than a definitive state classification. Again however, this method would require experimentation in choice of layers and neuron number to provide an adequate result based on the training data and is therefore an inappropriate choice for this particular application (as suggested in the next section).

There is possibility of considerable variation between data from all users. A method must be used which guarantees linear separation, followed by a method with a low enough error rate to provide reasonable estimation of an observation’s belonging to a state. The dimension reduction through curvilinear distances technique of Winkley et al. (2011) was developed to be such a solution: taking the 5 dimension nonlinear data and processing it to be “unfolded” and linearly separated in 2 dimensions, before using the result in a simple classifier to provide probability of membership to a state.

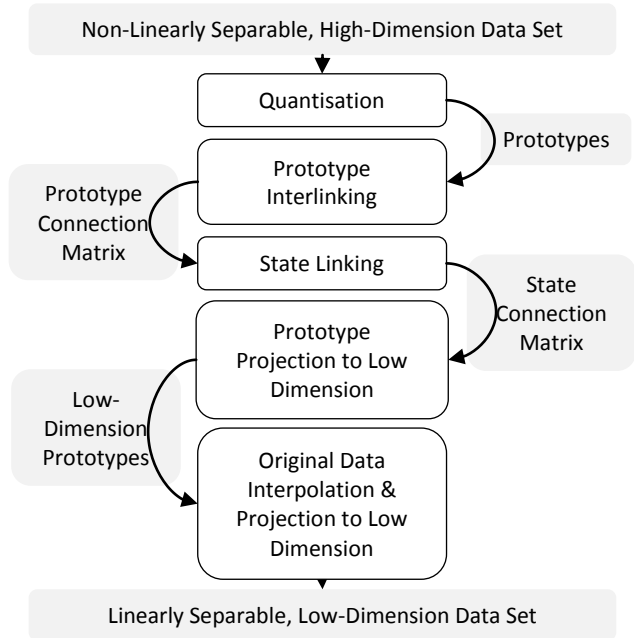


Fig. 4 The 5 step process to reduce the dimensionality of the data set

The scheme is based on the Curvilinear Distance Analysis (CDA) technique proposed by Lee et al. (2004) - which was an extension to the original CCA (Demartines and Herault, 1997) - but modified to allow for nonlinearly connected clusters to be incorporated into the manifold representation and separated in the low dimension. The subsequent classification utilises the dimensionally reduced data to provide a simple matrix of weights, which enables successive data points to be classified almost immediately after being gathered by the sensors and thus

the entire scheme is perfectly suited to a real-time application.

There are 5 steps in the dimension reduction and separation process, all of which are carried out offline with the sensor data obtained through training (Fig 4). Whilst with many data samples the process can take a considerable amount of time, the result greatly optimises the placement of data points in the low dimension so that the general topology from the higher dimension is retained. The cluster data is treated as separate sets in the initial stages before becoming interconnected as a single manifold later on in the process for unfolding. In this way, each set of state data can be accurately modelled to remain true to the overall topology. Firstly, to reduce the magnitude of the calculations in later steps the input data of each state is normalised. This does not affect the weighting of the observation in the final outcome, as the exact value is not required - only a probability of an observation’s occurrence in a state is desired. The normalised data is then quantised to produce a set of prototypes which typify that state’s data. Dynamic vector quantisation is used to select values which best represent the overall cluster, with a tolerable loss value determining how representative the prototypes are: a lower value will result in a larger number of prototypes which closely resemble the overall shape of the cluster. A higher tolerable loss gives fewer prototypes but reduces the computational load of subsequent operations.

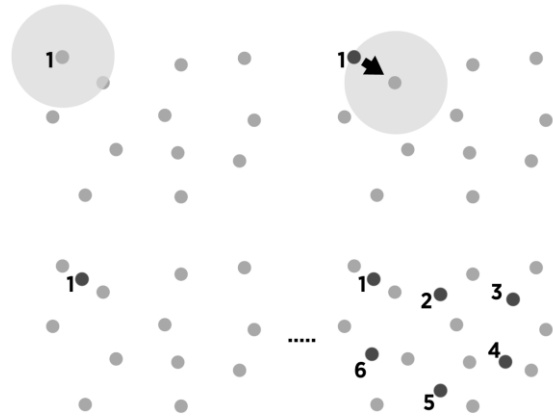


Fig. 5 A representation of the Vector Quantisation process. Point 1 starts as a prototype before moving to represent the average of two points.

From experimentation, the best values for use with Verity’s sensor data have been found to be between 0.05 and 0.2. With a set of 75 observations and a tolerable loss of 0.1, the typical number of prototypes is 32. After completion, the projection error is typically around 10^{-6} with this tolerable loss.

The quantisation process is as follows:
for each state cluster


```

max_distance is 0
for all points in state cluster
    if (distance between 2 points is greater
        than max_distance)
        distance becomes max_distance

radius = max_distance * tolerable_loss
set empty prototype list
prototype_num is 0

while (iteration is acceptable or prototype_num
continues to increase)
    for all data points in cluster
        for all prototypes
            if (data point is not within radius of
                prototype)
                data point becomes a prototype
                prototype_num = prototype_num + 1
            else
                move closest prototype within
                radius by an amount which
                decreases with every iteration

```

The calculated prototypes are put in a list to be processed in subsequent operations. Typically the number of prototypes will be equal to around half the total number of points in that state. Fig 5 illustrates the quantisation.

The next step is a modification to the documented original CDA. Connection of prototypes within the states would originally have been with a k -nearest neighbours or a simple radius-based approach, however it was found that this would sometimes result in an incorrectly structured “web” of neighbourhood linkages (with some “parasitic links”) and some points even being left unconnected. Using an iteration-increasing neighbourhood, the connections between points can be monitored to ensure that no “web” links occur. Employing Dijkstra’s algorithm (1959) also enables the best form of link to be determined as it evaluates the distance between prototypes and retains the connections which arose first (i.e. once already connected by a small neighbourhood, if then included in the larger neighbourhood the smaller connection distance will be retained).

```

neighbourhood is average of 2 smallest Euclidean
distances between prototypes
step is neighbourhood * max Euclidean distance
between prototypes

```

```

while (max distance between all prototypes
determined through Dijkstra is infinite)
    for all prototypes in state cluster
        for all prototypes in state cluster
            if (distance between both prototypes is
                than neighbourhood)
                if (prototypes not already connected in
                    Dijkstra graph)
                    both prototypes linked
            neighbourhood is neighbourhood + step

```

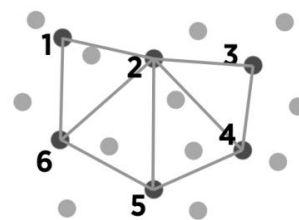


Fig. 6 Connecting the prototypes of a state

The prototype connection process results in a simple graph/matrix which forms the basis of subsequent operations. Fig. 6 provides a graphical representation of the connections within one of the states.

Linking the state clusters together to form a single manifold is the most important step in the linear separation as it is at this point that the maximum distance to project between states is determined. Now that the states are interconnected internally, they must be joined up in a sequence with other states such that their projection resembles the unravelling of a chain, where each link is a single state joined to the next by one linkage. This one linkage is calculated as each cluster is addressed, by measuring the Euclidean distance between all prototypes in one state to the next and the two prototypes which are found to be furthest apart are then joined. In the next state, the furthest prototype from the one linking to the previous is then selected to be that which will connect to the furthest in the next state and so on until all are connected. There is now a matrix containing all prototypes and the distances between those in each state – with some prototypes being linked to other states. Fig 7 shows the resultant connections after the linking operation in a simplified 2D representation.

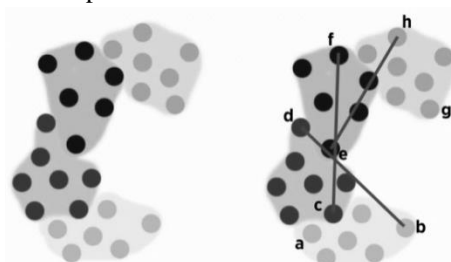


Fig. 7 The furthest prototypes within each state are labelled (a,b), (c,d), (e,f) and (h,g). The two furthest prototypes of two neighbouring states are connected via links bd, cf and eh such that all states are now interconnected through these traversable links

Once all states are connected in a chain as in Fig. 7, the projection of the training set can begin. The projection follows the same process as the original CDA. A matrix of all pairwise curvilinear distances between points is first created by employing Dijkstra’s algorithm on the previous distance matrix. This matrix gives the distance between any two prototypes in the set, given that it is now possible

to traverse a linkage from one state to the next. The projection to the lower dimension attempts to separate prototypes by the values in the distance matrix of the high dimension, but with an acceptable error - given that due to stretching, the distances cannot be recreated exactly. The error function to be minimised matches that of the original CDA and CCA, and the process operates exactly the same as documented in Demartines and Herault (1997), moving prototypes around each other by an amount proportional to the error between their intended placements.

$$E_{CDA} = \sum_{i=1}^n \sum_{j=1}^n (\delta_{i,j}^d - d_{i,j}^p)^2 F(d_{i,j}^p) \quad (30)$$

$\delta_{i,j}^d$ is the calculated curvilinear distance in the higher dimension and $d_{i,j}^p$ is the Euclidean distance in the lower dimension between prototypes i and j . F is a factor which weighs each term of the function and varies between 0 and 1 when its argument increases and decreases respectively. It is intended that this factor enable the CDA method to more importantly reproduce smaller distances over larger ones and so evaluates whether the Euclidean distance between points is small enough to be further optimized in the low dimension. With the training data used for Verity, optimising placement of the prototypes on average takes around 250 iterations for 30 samples of data – a mere 20 or so seconds offline. With greater numbers of samples and prototypes this time can increase, especially if the learning rate decreases with time.

After the prototypes have been placed with an acceptable error in the lower dimension, the original data points are interpolated also. All that is required for this step is to identify the 3 closest prototypes to the data point in the high dimension, and determine the distance of the point to the single closest. This single closest prototype's distance from the other two is then added to the first distance to give an array of 3 distances relating to the data point and the prototypes (Fig. 8). As with the original scheme the error function is minimised, except in the interpolation case it is solely the data point that is radially moved and not the already established prototypes. Interpolation typically takes a few iterations – the aforementioned 30 sample data had its prototypes and original points positioned in the lower dimension in just under 22 seconds with an overall acceptable placement error of 10^{-5} .

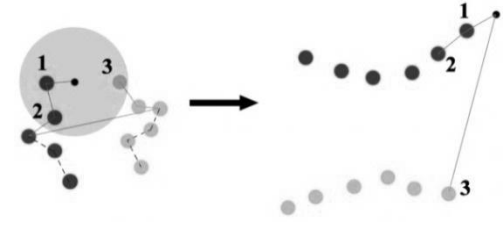


Fig. 8 Selection of 3 closest prototypes to point (smaller, black point; light grey neighbourhood) in higher dimension and the subsequent projections in the lower dimension

Once the data has been dimensionally reduced it is linearly separable and can then be used to train a very simple set of perceptrons such that subsequent data points from the high dimension need only be dimensionally reduced and presented to the perceptrons in order to determine their membership to a state. At their output the perceptrons use a sigmoid function to provide a probability – the closer to the centre of the cluster, the higher the likelihood. Fig. 9 shows the construction of this cascading network of perceptrons.

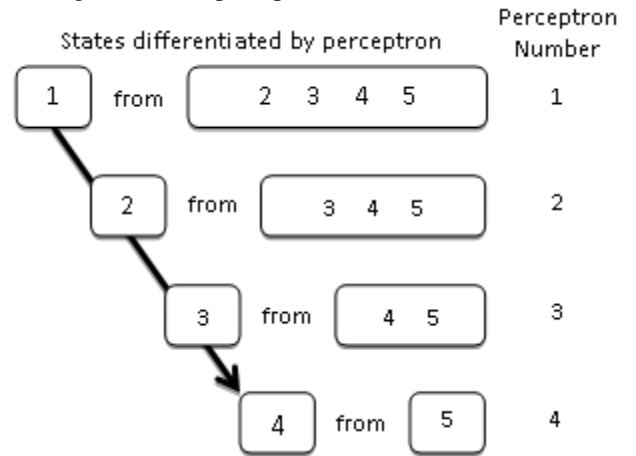


Fig. 9 Architecture of cascading perceptron network

The dimensionally-reduced, linearly-separable version of the data point is presented to the first perceptron, which has been trained to differentiate between state 1 and all others. The output of this perceptron indicates the probability that the point does belongs to state 1. It is then presented to the second perceptron to assess its membership to state 2, and so on. This process continues until the states (and perceptrons) are exhausted, with the final outcome in this example being a membership value to state 5.

The combined curvilinear distance-based reduction technique and perceptron classifier is mathematically proven to result in correct classification of not-linearly separable data points from the high dimension (Winkley et al., 2011).

Once the training data has been used to create a new system with which to generate membership probabilities,

the original fuzzy model of the first stage is mostly discounted from subsequent operations and instead the cascading perceptrons are consulted to return a value of membership to a state. The model is only consulted in the training phase of the dimension reduction, and in queries whereby an unknown data point requires a supervision signal to determine possible membership. In these instances, the FIS is consulted to return a likely state according to the general rules which the user is invited to confirm or suggest an alternative that is then taken as that set of sensor readings' classification.

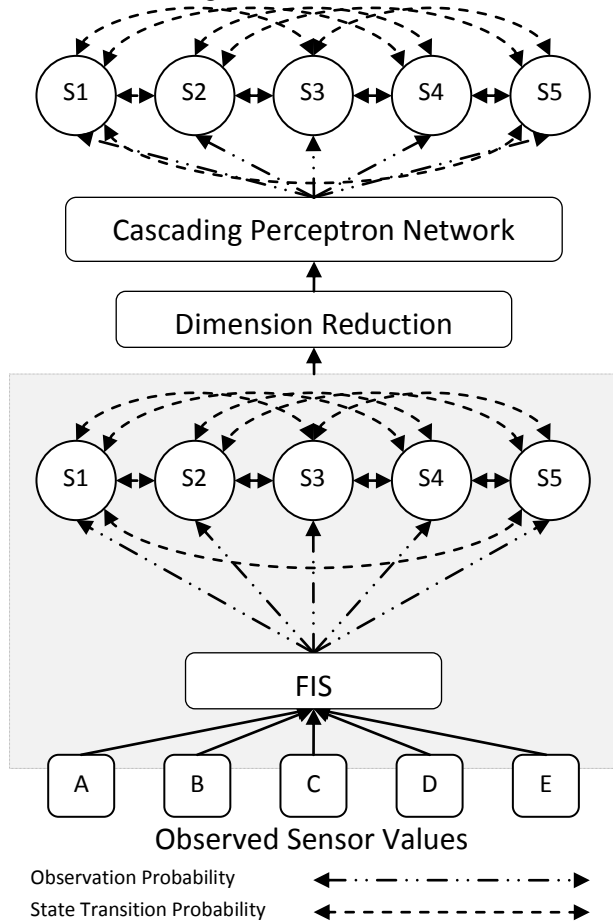


Fig. 10 The new dimension reduction scheme interacting within the HMM. The grey shaded box is the initial FIS implementation, which occurs until sufficient data is received regarding the membership of each observation. Once enough data is available (a cap pre-determined by the programmer), the scheme is dropped and the trained perceptron scheme takes over for future classifications, the observed sensor values being directly submitted to the dimension reduction operation

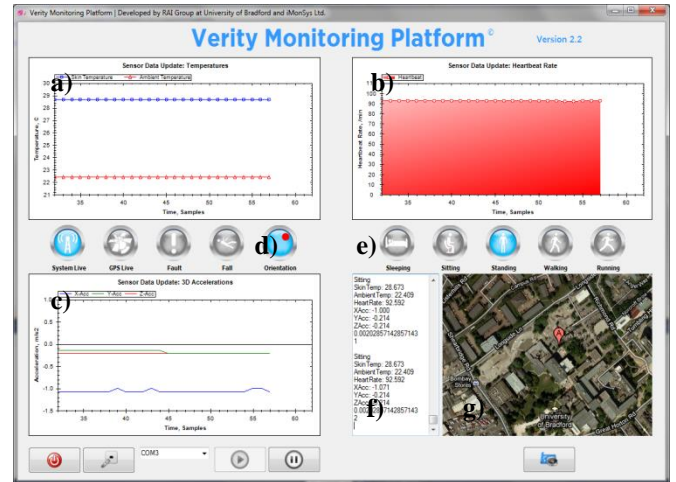


Fig. 11 The Monitoring Platform consists of a) Ambient and Contact Temperatures, b) Pulse Rate, c) Acceleration, d) Orientation, e) State Identification, f) Sensor Debug Window, g) GPS Location. Plus other necessary notifications and buttons for use during monitoring

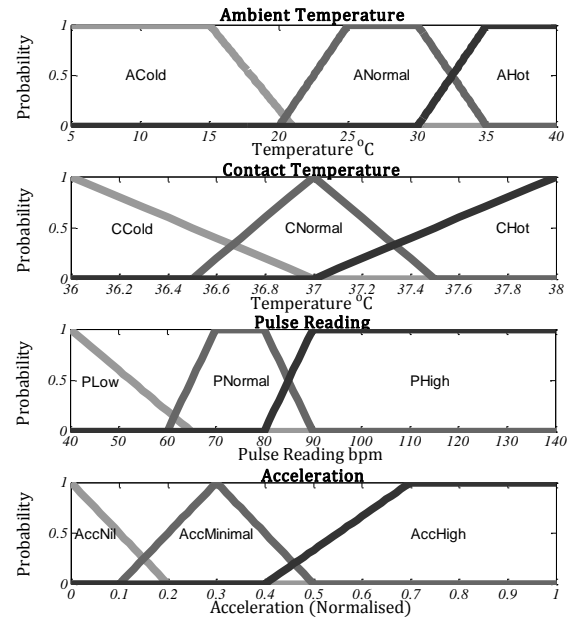


Fig. 12 The test system sensor membership functions

6 Testing

The Verity system and internal software have been tested in a simulated and controlled scenario in order to determine its effectiveness during operation.

Occurring on a desktop computer, the simulation was able to run such that all processes could be viewed in real-time and assessed for suitability of application. The graphical interface used to assess the working device is shown in Fig 11.

Once the sensor memberships were identified as in Fig. 12 the linguistic rules which activate the states were constructed and input into the simulation software:

1. **Sitting:** Ambient Temperature is *Normal*, Contact Temperature is *Normal*, Pulse Reading is *Normal* and Acceleration is *Nil*
2. **Standing:** Ambient Temperature is *not Hot*, Contact Temperature is *Normal*, Pulse Reading is *Normal* and Acceleration is *Nil*
3. **Walking:** Ambient Temperature is *not Hot*, Contact Temperature is *Normal*, Pulse Reading is *not Low* and Acceleration is *Minimal*
4. **Running:** Ambient Temperature is *not Hot*, Contact Temperature is *not Cold*, Pulse Reading is *High* and Acceleration is *High*

The parameters of the model are identified through human knowledge of the situation. Again using human knowledge the transition and initial probabilities were defined (states numbered as in Fig 13).

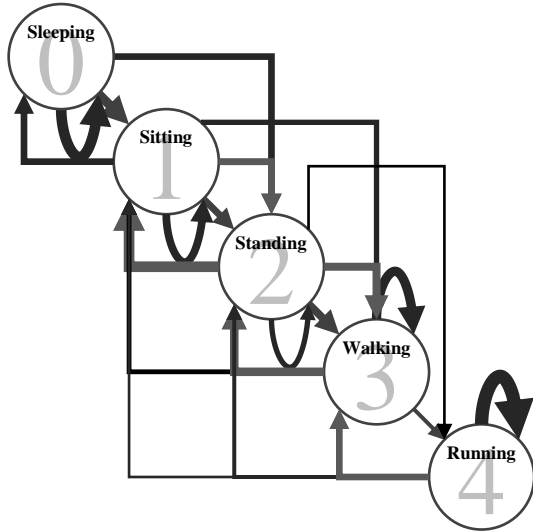


Fig. 13 State transitions. Line weight denotes probability strength

$$a_{ij} = \begin{bmatrix} 0.33 & 0.34 & 0.33 & 0 & 0 \\ 0.25 & 0.35 & 0.3 & 0.1 & 0 \\ 0 & 0.3 & 0.3 & 0.3 & 0.1 \\ 0 & 0.05 & 0.3 & 0.4 & 0.25 \\ 0 & 0 & 0.2 & 0.3 & 0.5 \end{bmatrix} \quad (31)$$

$$\pi = [0.2 \quad 0.3 \quad 0.3 \quad 0.1 \quad 0.1] \quad (32)$$

Once the parameters were all defined, the user wore the Wrote on the wrist while the base station was connected via USB link to the PC. Exhibiting a series of pre-arranged states, the readings which were output from the device are shown in Table 1.

The states progress as generally expected, going from standing to sitting, to standing, to running, to walking and back to running. There is an anomaly in the above data though whereby the determined state is walking (3) during a period of sitting. This is because the above states come

from the fuzzy system, and so is the result of instantaneous decisions based purely on sensor data and not on the time series data. Realistically the anomalous result was corrected by the HMM using both the Forward-Backward and Viterbi algorithms, as they took into account the previous states and the transitions, but as the dimension reduction and classification method is that which is being assessed, the data used must be the raw data from the fuzzy system.

Submitting these values for processing with the CDA-based method resulted in adequate representation in the lower dimension and correct classification using the perceptrons as described above. The graphical representation of the result of the process can be seen in Fig. 14.

Table 1 Data received from the Verity device during simulation

No.	Ambient	Contact	Pulse	Motion	Orientation	State
1	28.699	28.776	76.142	0.000	1	2
2	28.699	28.776	76.142	0.000	0	1
3	28.699	28.818	80.213	0.000	0	1
4	28.699	28.818	80.213	0.000	0	1
5	28.699	28.818	80.213	0.000	0	1
6	28.699	28.838	81.967	0.256	0	3
7	28.699	28.838	80.213	0.170	0	1
8	28.699	28.838	81.967	0.114	1	2
9	28.699	28.838	81.967	0.114	1	2
10	28.699	28.849	81.967	0.115	1	2
11	28.699	28.849	81.967	0.172	1	2
12	28.699	28.849	81.967	0.172	8	2
13	28.699	28.838	81.967	0.598	8	4
14	28.699	28.838	81.967	1.084	8	4
15	28.699	28.849	81.967	1.170	8	4
16	28.699	28.849	81.967	0.827	8	4
17	28.699	28.849	81.967	0.458	8	3
18	28.699	28.828	81.967	0.458	8	3
19	28.699	28.828	81.967	0.458	8	3
20	28.699	28.797	81.967	0.458	8	3
21	28.699	28.797	81.967	0.515	8	3
22	28.699	28.797	81.967	0.516	8	3
23	28.699	28.683	81.967	0.686	8	4
24	28.699	28.683	81.967	0.686	10	4
25	28.699	28.683	81.967	1.627	10	4
26	28.699	28.662	81.967	1.799	10	4
27	28.699	28.704	81.967	2.370	0	4
28	28.699	28.704	81.967	2.828	0	4
29	28.699	28.704	81.967	2.484	0	4
30	28.699	28.704	81.967	2.484	0	4

Table 2 Weights from training perceptrons with Verity test data

State Perceptron	Weight 1	Weight 2	Weight 3
1	-3.3668	-3.7487	-5.9766
2	-6.5034	-6.1405	-37.5824
3	-11.3990	-12.2098	-0.4053

Table 3 Result of using trained weights on unseen data points. The “actual” result is the state the user exhibited, where “result” is the state determined by the system.

Ambient	Contact	Pulse	Motion	Orient.	Actual	Result
28.699	28.838	80.213	0.000	0	1	1
28.699	28.838	76.142	0.170	0	1	1
28.699	28.849	81.967	0.114	8	2	2
28.699	28.797	81.967	0.458	6	3	3
28.699	28.662	80.213	1.799	0	4	4
28.699	28.704	81.967	1.799	10	4	4

The perceptrons were trained on this data in order to produce 4 sets of weights (Table 2) which are used for the classification of subsequent unseen data points (Table 3).

The results of the test show that the method is capable of being used in a real scenario to accurately determine the states exhibited by a user without having to go through the original fuzzy system. The output in the real system is however taken as a value of probability, not a definitive state value. It is only for test purposes that these values appear as classifications.

The method has been tried in comparison with the commonly used feed-forward back-propagation network, where it outperformed in all tests against different networks of differing neuron numbers (ranging from 20 – 40 neurons), going further to prove its effectiveness in its one-pass approach to training without external input.

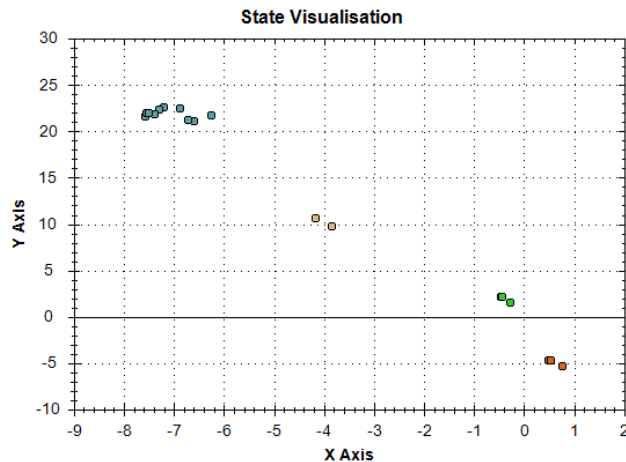


Fig. 14 Result of dimension reduction of the Verity data obtained during simulation. The clusters of points and their different colours show that each state within the data set has been adequately linearly separated and classified correctly as their respective states.

7 Conclusion

What has been shown is that a fuzzy approach can greatly assist when an observation (for use in a probability model such as the HMM) comprises of more than one value. The fusion process and human reasoning aspect provide the

model with greater accuracy in an application which ordinarily relies on human knowledge. The method has been seen to be efficient in the observation probability parameter determining process, especially when data for the model is unavailable for the traditional training methods. The approach provides scope for the future creation of models where there are a greater number of inputs yet the model is still governed by human reasoning.

Whilst other combinational techniques for multiple observations exist, for example Li et al. (2000), the multiple observations are sequences themselves – i.e. their probability of emission from a state in a sequence is determined by the same HMM addressing the overall sequence. The technique described here takes observations with probabilities determined by another model before inclusion in the HMM. It is possible to implement an HMM for each sensor, combining the outcome of the models to obtain a single state belief: the drawback being that each model would not consider the influence each sensor’s reading has on the overall state. Only with the inclusion of the human knowledge of state properties can the probabilities of emission be adequately defined for the combination of readings.

The dimension reduction and classification scheme is intended to take over the state probability determining once there is sufficient data obtained through the FIS, in order to enable further expansion of the data space during use and tailoring to the user. In the experiments, the resultant probability determining scheme performed more than adequately, replacing the FIS and producing results akin to those that would be achieved with the FIS if its rule base was originally describing the user. A key advantage of the replacement is the ability to produce a visualisation of the data space where before it would be impossible due to the high dimensionality of the data. In the use of Verity this becomes somewhat significant, as a healthcare professional may be able to identify from the visualisation that an outlier in the lower dimension is indicative of a possible health issue which may need further investigation. The initial supervision signal of the FIS used to label each of the sensor data observations for the dimensional reduction means that the reliance on the user to specify exactly to what state their current readings belong is reduced, instead being offered a select number of possible options in order to better tailor the model to them.

The information contained within this work primarily describes the first phase of the Verity System’s software implementation. Tests performed throughout the phase show that the direction in which the system is being taken both in the hardware and software design provides great scope and opportunity for real application; future phases

will ultimately build on the improvements identified in the first to produce a device which will become a valuable addition to the ambient healthcare market. The internal control mechanisms are innovative in their design and whilst developed for this specific purpose there is much that can be accomplished by implementing them in other areas, such as industrial control schemes that already employ HMMs and in other data processing tasks where high dimension data is difficult to work with in its raw form.

Evidence of the techniques' capabilities has been given through explanation of their reliable use in a real application. Once the system is deployed to a number of test subjects, it is expected that sufficient data will be available with which to more appropriately tune the initial fuzzy parameters of the model; eventually allowing an "average" set of readings to be calculated from a large population.

Work will be undertaken to ensure the hardware is adequate enough for unsupervised use, and the control scheme will need addressing in terms of deciding which processes can be handled offline by a central hub rather than in real-time on the base station. Currently the dimension reduction operation is the most computationally intensive, so it is expected that this process will occur offline with the resultant parameters for the perceptrons to use in the real-time operation being delivered to the base station upon completion of the process. With the results obtained during the simulation, it is expected that the solution detailed in this paper for the initial probability determining will be easily repeatable in the hardware providing that the input data space continues to conform to the properties of the data obtained in these experiments. However, given the nature of behaviour monitoring and the ever-varying observations seen by sensors due to noise etc., there is possibility for a need to modify the scheme to adapt to more unexpected results during use.

The dimension reduction and classification scheme will again be compared to another method being developed which attempts to classify data directly in the high dimension using a new procedure: primarily to assess which technique is more suitable for this application. Whilst still in the high dimension however, the data is not easily visualised so it is expected that if indeed the new method provides better classification results, this method will still be employed for the purposes of data visualisation.

Ultimately it is foreseen that throughout the next few phases the system will become more honed and suited to its application and the elderly users at whom it is aimed, with the core elements already in place and further

development focussing on bettering the current methods based on the feedback from real use.

References

- Abbod, M F, Von Keyserlingk, D G, Linkens, D A & Mahfouf, M (2001) Survey of utilisation of fuzzy technology in Medicine and Healthcare. *Fuzzy Sets and Systems*, 120, 331-349.
- Aliustaoglu, C, Ertunc, H M & Ocak, H (2009) Tool wear condition monitoring using a sensor fusion model based on fuzzy inference system. *Mechanical Systems and Signal Processing*, 23, 539-546.
- Avci, U and Passerini, A (2012) Improving activity recognition by segmental pattern mining. *Pervasive Computing and Communications Workshops (PERCOM Workshops)*, 2012 IEEE International Conference on, 709-714.
- Barger, T S, Brown, D E & Alwan, M (2005) Health-status monitoring through analysis of behavioural patterns. *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, 35, 22-27.
- Baum, L E & Eagon, J A (1967) An inequality with applications to statistical estimation for probabilistic functions of Markov processes and to a model for ecology. *Bull. Amer. Math. Soc.*, 3, 360-363.
- Burgess, M (2006) Probabilistic anomaly detection in distributed computer networks. *Science of Computer Programming*, 60, 1-26.
- Cortes, C & Vapnik, V (1995) Support-vector networks. *Machine Learning*, 20, 273-297.
- De Moraes, R M & Dos Santos Machado, L (2004) Using fuzzy hidden Markov models for online training evaluation and classification in virtual reality simulators. *International Journal of General Systems*, 33, 281 - 288.
- Demartines, P & Herault, J (1997) Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. *Neural Networks, IEEE Transactions on*, 8, 148-154.
- Dijkstra, E W (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, 1, 269-271.
- Ferreira, L and Ambrosio, P (2012) Towards an interoperable health-assistive environment: The eHealthCom platform. *Biomedical and Health Informatics (BHI)*, 2012 IEEE-EMBS International Conference on. 930-932.
- Fine, B T (2009) Unsupervised anomaly detection with minimal sensing. *Proceedings of the 47th Annual Southeast Regional Conference*, Article 60, 1-5.
- Fulford-Jones, T R F, Gu-Yeon, W & Welsh, M (2004) A portable, low-power, wireless two-lead EKG system.

- Engineering in Medicine and Biology Society., IEMBS '04. 26th Annual International Conference of the IEEE, 1-5 Sept. 2004. 2141-2144.
- Hall, D L & Llinas, J (2002) An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85, 6-23.
- Hong, T P & Lee, C Y (1996) Induction of fuzzy rules and membership functions from training examples. *Fuzzy Sets and Systems*, 84, 33-47.
- Kelarestaghi, M, Slimane, M & Vincent, N (2001) Introduction of fuzzy logic in the Hidden Markov Models.
- Lee, H, Park, K, Lee, B, Choi, J & Elmasri, R (2008) Issues in data fusion for healthcare monitoring. *Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments*. Athens, Greece: ACM.
- Lee, J A, Lendasse, A & Verleysen, M (2004) Nonlinear projection with curvilinear distances: Isomap versus curvilinear distance analysis. *Neurocomputing*, 57, 49-76.
- Li, N, Guo, G D, Chen, L F, Chen, S (2012) Optimal subspace classification method for complex data. *International Journal of Machine Learning and Cybernetics*, 1-9.
- Li, X, Parizeau, M & Plamondon, J (2000) Training Hidden Markov Models with Multiple Observations-A Combinatorial Method. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22, 371-377.
- Lo, B, Thiemjarus, S, King, R & Yang, G (2005) Body Sensor Network - A Wireless Sensor Platform for Pervasive Healthcare Monitoring. *Adjunct Proceedings of the 3rd International conference on Pervasive Computing*. 77-80.
- Mamdani, E H & Assilian, S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, 7, 1-13.
- Marques, V, et al (2011) A dynamic user profiling technique in a AmI environment. *Information and Communication Technologies (WICT), 2011 World Congress on*, 1247-1252.
- Mukala, V, Lakafosis, V, Traille, A & Tentzeris, M M (2010) A novel Zigbee-based low-cost, low-power wireless EKG system. *Microwave Symposium Digest (MTT), 2010 IEEE MTT-S International*, 23-28 May 2010 2010.
- Pearson, K (1901) On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, 559-572.
- Rabiner, L R (1990) A tutorial on hidden Markov models and selected applications in speech recognition. *Readings in speech recognition*. Morgan Kaufmann Publishers Inc.
- Sasiadek, J Z (2002) Sensor fusion. *Annual Reviews in Control*, 26, 203-228.
- Sasiadek, J Z & Wang, Q (1999) Sensor Fusion Based on Fuzzy Kalman Filtering for Autonomous Robot Vehicle. *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*. Detroit, MI, USA.
- Shiqian, W, Meng Joo, E, Maolin, N & Leithead, W E (2000) A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks. *Proceedings of the American Control Conference, 2000*. 2453-2457 vol.4.
- Sopavanit, C, Desudchit, T & Riyamongkol, P (2008) Wireless and wearable EKG device with lossless compression for on-line post-surgery heart monitoring system. *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*, 22-25 Feb. 2009. 1218-1223.
- Venkatesh, V, et al (2012) A secure Ambient Assisted Living (AAL) environment: An implementation view. *Computer Communication and Informatics (ICCCI), 2012 International Conference on*. 1-7.
- Verma, N K & Hanmandlu, M (2009) Additive and Non-Additive Fuzzy Hidden Markov Models. *Fuzzy Systems, IEEE Transactions on*, 18, 40-56.
- Viterbi, A (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *Information Theory, IEEE Transactions on*, 13, 260-269.
- Wang, L X & Mendel, J M (1992) Generating fuzzy rules by learning from examples. *Systems, Man and Cybernetics, IEEE Transactions on*, 22, 1414-1427.
- Wang, L (2011) An improved multiple fuzzy NNC system based on mutual information and fuzzy integral. *International Journal of Machine Learning and Cybernetics*, 2(1): 25-36.
- Winkley, J, Jiang, P & Hossain, A (2011) Dimension Reduction for Linear Separation with Curvilinear Distances. *Soft Computing, Mendel 2011: 17th International Conference on*, Brno, Czech Republic. 515-522.
- Zhang, S, McCullagh, P, Nugent, C, Zheng, H and Baumgarten, M (2011) Optimal model selection for posture recognition in home-based healthcare. *International Journal of Machine Learning and Cybernetics*, 2(1): 1-14.
- Zhou, F, et al (2011) A Case-Driven Ambient Intelligence System for Elderly in-Home Assistance Applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 41 (2), 179-189.