

# 新興模式探勘方法於人類活動辨識之研究

## Emerging Pattern Mining in Human Activity Recognition

林北辰<sup>1</sup>、劉立頌<sup>2</sup>、張晉嘉<sup>3</sup>

<sup>123</sup> 國立中正大學 電機工程學系

<sup>1</sup>amos91100@gmail.com

<sup>2</sup>aliu@ee.ccu.edu.tw

<sup>3</sup>jerry84609@gmail.com

### 摘要

在人類活動辨識活動的研究中，使用慣性感測器當作資料源是一項挑戰，若能在缺乏其他感測器的資訊下也能成功辨識活動，則使用者的活動範圍就不會被侷限在特定的環境中。在活動辨識中利用慣性感測器來辨識低階的動作其準確度比較好，但是考慮複雜活動辨識的研究數量較少，其辨識準確度也不高。而在另一派學者辨識複雜活動使用模式探勘的方法，其資料源會包含環境感測器等其他非慣性感測器，並利用新興模式探勘方法找出一個分段包含類似感測器讀值的不同活動。因觀察到此現象，故本研究利用簡單活動來辨識複雜活動，並利用頻繁模式結合新興模式探勘方法找出活動間具區別度大的模式，減少模式的數量來降低查找模式以及訓練分類模型的計算量。

關鍵字：人類活動辨識、頻繁模式、新興模式、機器學習

### Abstract

In the research of human activity recognition, using inertial sensors as a data source is a challenge. The user's activity range will not be limited to a specific environment if the activity can be successfully identified without additional information from other sensors. The accuracy of using inertial devices to recognize low-level motion activities is well studied, but there are few studies considering the recognition of complex activities and the accuracy is not good. Some of the researcher use pattern mining methods to recognize complex activities and the data source includes environmental sensors and other non-inertial sensors. They adopt emerging pattern mining method to find different activities, which contain similar sensor readings, in one segment. Due to the observation of the phenomenon, we use simple activities to recognize complex activities, and adopt frequent patterns combined with emerging pattern mining methods to find highly different patterns between activities. Eventually, we reduce the amount of calculations for finding patterns and training classification models by using fewer number of patterns.

Keywords: Human activity recognition, frequent pattern, emerging pattern, machine learning

### 1. 前言

近年來遍佈式計算的發展，人們期望更便利的

生活環境，例如智慧家庭提供情境感知系統，能夠自動收集環境中的情境資訊並推理出使用者可能的需求。其中活動感知系統扮演重要的角色，主要分為兩個部分：活動辨識與活動預測。活動辨識目的為辨識使用者當前從事的活動，活動預測目的則為預測使用者接下來可能的活動，而本研究屬於前者。在人類活動辨識的研究中，使用慣性感測器作為資料辨識活動被視為是一項挑戰，在缺乏其他感測器資訊的情況下維持活動辨識的準確度是其目標，目前在辨識移動或手勢等身體動作的低階活動辨識研究，辨識結果能達到90%以上的準確度，在辨識高階的複雜活動中的研究卻不多，辨識的效果也相較不佳。在使用模式探勘方法的人類活動辨識中，許多研究考慮多種感測器當作資料源，利用新興模式方法來找出在一個分段中可能出現具有類似感測器讀值的活動，本研究利用身體動作透過模式探勘的方法，使用模式探勘中的新興模式來找出組成動作相似的活動之間區別度較大的模式，並以此來當作特徵訓練分類器以辨識活動。

### 2. 背景知識與文獻探討

本研究領域為人類活動辨識，是智慧家庭研究領域下的一個子領域，而本章節會介紹一些與本研究有關的背景知識與相關文獻。

#### 2.1. 情境感知系統

情境感知系統是智慧家庭中重要的一部分，是透過從家庭自動化[1]取得環境和使用者中的各種訊號，到提供使用者服務來滿足其需求的過程。透過推理技術將取得的情境資訊，轉換成使用者需求，而這一過程稱為情境感知服務。[17]中將情境分為內部與外部。活動感知系統是其中一個很重要的研究領域，利用外部情境通過使用感知和推理來確定使用者當前的內部活動情境，其中包括了活動辨識以及活動預測兩個主要的子研究。活動辨識目的是要透過所取得的低階資訊推理成高階的資訊，來辨識使用者目前正在從事什麼樣的活動。許多研究在非真實時間上辨識活動準確率已經很好[10]，而感測器技術的蓬勃發展，帶動許多利用感測器讀值在真實時間上辨識活動的研究。

#### 2.2. 人類活動辨識

可以分為簡單活動和複雜活動，簡單活動代表在給定時間內持續執行的活動，複雜活動則是由複

數的簡單活動組成。複雜活動可以有兩種模式，分別是並行以及交錯，並行活動是同時發生簡單活動的集合，交錯活動是在簡單活動之間切換的集合。人類活動識別可以視為典型的模式識別問題[20]，從原始感測器輸入中學習有關人類的活動知識，我們利用活動辨識鏈[2]來表示人類活動辨識的流程。

### 2.3. 活動辨識鏈

人類活動辨識流程會表示為活動辨識鏈，可被分為監督式和非監督式兩種。監督式方法會將擷取的特徵和對應標記作為輸入以訓練分類器，主要流程為資料獲取、資料前處理、資料分段、資料提取、及訓練與分類。要取得用於人類活動辨識資料的類型主要有影像與音訊一派，以及感測器活動辨識一派[14]，感測器可分為環境感測器與穿戴式感測器，本研究使用的慣性感測器屬於穿戴式感測器的一種，常見實物有加速器和陀螺儀。加速器可以監測動態活動、讀出物體沿著某一軸的加速度值，有時間和直角坐標的三軸加速度資訊，這些資料推斷出的情境可以用於長期連續活動的監測及緊急情況，比如跌倒辨識。陀螺儀則是將震動體加到感測器中，以測量在一段時間內的角速度變化。

### 2.4. 資料處理

資料可能因為環境、通訊因素使其包含雜訊，或者需要進一步處理為有用的訊號。資料若有缺失或不平衡，也可以彌補缺失值或是重新取樣來克服[11]。對於感測器生成連續的串流資料，因為資料量與感測器的取樣頻率相比，所以一旦人類活動的持續時間拉長，便會提高使用連續串流資料進行活動分類的難度，且模型在訓練時難以從連續的資料中檢索出重要和有用的訊息，因此需要將資料分段。

資料前處理：通常會由機器學習演算法或是推理方法來實現，但是未來智慧家庭環境會越來越複雜，裝置數量多，所收集到的資料會包含更多的雜訊，這些雜訊原因可能來自電池放電、感測器讀取數值失敗以及網路中訊息丟失。我們可以透過過濾雜訊以留下真實資訊，再使用一些濾波器萃取出需要的資料。

資料分段：分段的資料會包含相關分類的活動訊息，在辨識流程中會需要將這些資料分段並映射到活動標籤，若輸入特徵沒有區分性會造成分類器不能產生有意義的分類結果。活動辨識中，將資料分段的動作被稱為活動分段，需要依賴預先定義的活動知識，或是有先經足夠訓練的模型，主要方法有事件分段和滑動視窗，本研究使用的方式為滑動視窗。滑動視窗是將感測器時間的序列按時間排序，並分為可能重疊的子序列。給定  $n$  個感測器事件  $S = \langle e_1, \dots, e_n \rangle$  的序列，事件視窗識別出一組  $x$  個的視窗  $P = \langle S_1, \dots, S_x \rangle$ ，視窗大小為  $\{w_1, \dots, w_n\}$ ，使得每個  $S_i$  是  $S$  的有序子序列。視窗集合  $P$  是有序、非空以及可能重疊的，且  $S_i$  的總和會是  $S$  的父集合，因此  $S_i$  可以表示為  $\langle e_i, \dots, e_{i+w_i} \rangle$ 。使用滑動視窗的方法是將每個  $S_i$  的最後一個或最近的一個感測器事件映射到活動標籤，滑動視窗感測器事件的順序給予活動進行的資訊，因此有利於映射。由於視窗的有序性，讓實時標記串流資料活動的標記成為有效方

法，且在訓練階段降低了活動辨識的計算複雜度。滑動視窗方法仍然需要做出許多決策，首先需要根據情境資訊的適當性和將要識別的活動類型來確定視窗大小，其次可能需要根據事件與當前情境資訊的相關性在一個視窗內對事件進行加權。

### 2.5 辨識方法

取得前處理過的時序性資料後，就是要擷取並選擇有用的特徵來建構辨識模型。基於慣性感測器的活動辨識，常見的資料型態是串流的三軸加速器和陀螺儀的讀值，根據取樣頻率會有連續且多筆的值對應到一個活動標籤。會透過滑動視窗得到人類身體最直接的資訊，再用機器學習模型進行分類。

在[15]中作者利用加速器訊號透過 L1 與 L2 距離計算相似度來分類。[4]收集了不同真實環境下的資料利用基於視覺和統計分析的方法，將特徵與活動標籤進行比較並找出 6 個特徵，最後使用三種分類器來進行分類。[6]的研究中，透過手機中的加速器和陀螺儀收集到的資料，經過 6 種不同的手工特徵擷取後，比較了數種分類方法。在[9]的研究中，比較了在用 21 種手工特徵擷取並運用數種分類器，以及深度學習中的 ResNet[12]模型自動擷取特徵並分類。表 1 整理了上述每個研究分類結果最好的方法。

表 1. 使用慣性感測器之活動辨識研究

	[15]	[4]	[6]	[9]	
資料類型	慣性感測器	慣性感測器	慣性感測器	慣性感測器	
分段方法	滑動視窗	滑動視窗	滑動視窗	滑動視窗	
特徵擷取	無	手工特徵	手工特徵	ResNet	
分類方法	相似度	決策樹	多層感測器	ResNet	
活動準確度	簡單	67%	86%	93%	96%
	複雜	無	無	50%	無

諸多研究顯示了利用慣性感測器讀值來辨識簡單動作的準確度可以高於 90%，至於複雜活動的研究較少、成效較差，因此本研究將利用簡單活動來辨識複雜活動。另外有一種複雜活動辨識的研究方式，名為模式探勘，目的是從大量資料集中擷取模式與知識，使用到機器學習、統計學、資料庫系統等領域方法。常用於人類活動辨識的方法有頻繁模式探勘以及新興模式探勘，會在第三章進一步介紹。

如表 2，在[13]中作者欲利用頻繁模式探勘方法找出每個活動頻繁出現的模式，並且利用 Allen 的十三種時間關係來表示不同動作之間的時間關係，此方法能有效使模式之間的動作包含更多的關係資訊，但是在模式探勘的步驟與建立分類模型的計算量也會大幅增加。而在[7]中，作者同樣使用了頻繁模式探勘的方法欲找出每個活動的頻繁模式，但是為了解決在一個分段中可能出現的複雜活動，因而使用新興模式以利於區分使用類似感測器的活動。[21]利用感測器讀值序列，而[8]加上了手部動作，同樣都使用了新興模式來找出一個分段或是視窗可能包含不同的活動。

在[7][21][8]中皆使用新興模式來篩選出活動

所包含差異度比較高的感測器資料，而本研究是透過身體動作的序列來辨識活動，因為不同的活動可能會包含類似動作，故選擇結合頻繁模式與新興模式的方法來達到好的分類效果。

表 2. 使用模式探勘之活動辨識研究

	[13]	[7]	[21]	[8]
使用方法	頻繁模式 +時間關係	頻繁模式 +新興模式	新興模式	新興模式
資料類型	動作序列	感測器序列	感測器序列	動作與感測器序列
分段方法	事件分段	事件分段	事件分段	滑動視窗
特徵擷取	自訂支持度	自定義特徵	自定義評分	自定義評分
目的	增加模式中的關係資訊	辨識含其他活動分段的活動	辨識並行和交錯的活動	辨識並行和交錯的活動

### 3 研究方法

本研究中結合兩種不同的模式探勘方法來尋找具有辨識度的特徵，本章會討論研究方法。一開始先將原始資料處理成能夠擷取特徵並且訓練的型態，在訓練階段使用模式探勘方法作為篩選特徵的方法，並對訓練資料做特徵提取，再在測試階段同樣做特徵提取當作新的測試資料來評估分類器。

#### 3.1 資料處理

為找出每個活動的模式，在給定的動作序列下需要將之區分為有序的完整活動，我們會利用資料集中的動作類別標籤所得到活動邊界，將資料分段成數個動作子序列，以及對應的活動類別當作標籤。本小節介紹基於身體動作的活動辨識資料型態，完整的串流活動紀錄會包括時間、事件以及所對應的活動標籤，而本研究將事件視為一個動作。設  $A$  表示包含了  $m$  個動作類別的集合，如公式(1)， $L$  表示包含  $n$  個活動類別標籤的集合，如公式(2)，而一個動作序列  $S$  包含數個有序動作  $\alpha_i$  滿足  $\alpha_i \in A$ 。

$$A = \alpha_1, \alpha_2, \dots, \alpha_m \quad (1)$$

$$L = l_1, l_2, \dots, l_n \quad (2)$$

在模式探勘方法中的事件分段階段，需要將完整的動作序列  $S$  區分為非空的動作子序列  $S_1, \dots, S_k$ ，當  $0 \leq i, j \leq k$  時，滿足  $S_i \cap S_j = \emptyset$ ，每個子序列分別對應不同活動標籤。假設資料取樣頻率是 2hz，則一段 10 秒鐘的動作序列  $A$  如公式(3)所示，序列中 1,2,3 為三個不同的動作編號。原始的動作序列可能會包含連續重複的動作，如果我們將其視為單一的完整動作，並且記錄每個完整動作開始與結束的時間，便可以將動作序列  $A$  簡化為公式(4)。若將時間拉長到一個活動的時間長度，則可以得到一個動作子序列  $A_i$  以及對應的活動標籤  $L_i$ ，而完整的活動序列  $S$  則會包含  $t$  個動作子序列，如公式(5)，每個子序列都會對應一個活動標籤。

$$A = \langle 1,1,1,2,2,2,2,1,1,1,1,3,3,3,3,1,1,1 \rangle \quad (3)$$

$$A = \langle 1,2,1,3,1 \rangle \quad (4)$$

$$S = \langle (A_1, L_1), (A_2, L_2), \dots, (A_t, L_t) \rangle \quad (5)$$

將資料分段成不同的活動後，我們將不同活動

的動作子序列分類成不同資料集，得到  $n$  個包含數筆動作子序列的資料集。最後我們就可以透過模式探勘找出相同活動的模式來做為活動的特徵，在分類階段訓練分類器來辨識活動。接下來的兩個小節會介紹本研究使用的模式探勘方法。

#### 3.2 頻繁模式探勘

頻繁模式探勘是在給定大型資料集中分析與搜索出最頻繁之相關模式的任務[5]，該概念最初是為了挖掘交易資料集而出現，而本研究視每個項目為動作，每筆交易視為動作子序列，將其應用在動作資料集，找出某一活動中常見的模式當作特徵。而頻繁模式定義為出現在資料集中的模式，其頻率不小於使用者指定或系統自動定義的閾值。本節介紹了頻繁模式探勘的基本概念，用於發現資料集中項目集之間有趣的關聯和相關性。

支持度(support):是模式探勘的一個度量方法，它反映了所發現的模式出現的頻繁性。假設  $I$  是一個包含  $m$  個項目的項目集，如公式(6)， $D$  是一組資料集，其中包含的每一筆交易  $T$  是一個非空的項目子集，滿足  $T \subseteq I$ 。令  $X$  為一組項目，滿足  $X \subset I$  且  $X \neq \emptyset$ ，則  $X$  在所有交易  $T$  中出現的總次數則稱為絕對支持度，記作  $\text{count}_D(X)$ ，而  $X$  在  $D$  中的支持度就是  $X$  在所有交易  $T$  中出現的機率總和，如公式(7)，通常又稱其為相對支持度，記作  $\text{sup}_D(X)$ 。

$$I = I_1, I_2, \dots, I_m \quad (6)$$

$$\text{sup}(X) = \frac{\text{count}_D(X)}{|D|} \quad (7)$$

假設有一資料集  $TID$  記錄著交易編號  $T1$  至  $T4$ ，項目清單記錄每一筆交易所包含的項目集，則可以計算出每組項目在這個資料集中的支持度，若有項目  $I1$  出現在交易  $T1$ 、 $T2$ 、 $T3$  中，次數是 3 次，可記為  $\text{count}_{TID}(I1) = 3$ ，佔總交易次數 4 次的比例為 75%，就是項目  $I1$  的支持度，可記為

$$\text{sup}_{TID}(I1) = \frac{\text{count}_{TID}(I1)}{|TID|} = 0.75。$$

頻繁模式：意旨在一個資料集內出現次數頻繁的模式，模式可以是項目集、子序列或是子結構。以數學化的角度來說，如果一個模式的支持度滿足所設定的最小支持度閾值  $\text{min\_sup}$ ，則這個模式就是頻繁模式，如公式(8)，這些閾值可以由使用者設置。延續上一個小節的例子，如果最小支持度閾值設為 50%，則  $I1$  就符合頻繁模式。頻繁模式探勘在挖掘資料集之間的關聯、相關性和許多其他有趣的關係中至關重要，它也有助於資料分類、聚類和其他資料探勘任務。因此，頻繁模式探勘已成為重要的資料探勘任務和研究的重點主題。

$$X \text{ is frequent} \Leftrightarrow \text{sup}(X) \geq \text{min\_sup} \quad (8)$$

$K$ -項目集：有數個項目的集合稱為項目集，一組包含牛奶、麵包的項目可以稱為 2-項目集，判斷  $k$ -項目集為頻繁項目集的方式與上一小節無異，只不過要同時考慮  $k$  個項目同時存在於一筆交易的頻率，頻繁的  $k$ -項目集通常表示為  $L_k$ 。以這種方式可以找到資料集中所有  $k$ -項目的頻繁模式，但是如公式(6)所示，若資料集中的項目集包含  $m$  個項目，則計算所有  $k$ -項目集的支持度所需要的次數如公式

(9)。若  $m$  越大，計算量會呈 2 的指數倍數成長。因此在下一小節會介紹一個資料探勘領域中經典的演算法 Apriori 來解決這個問題。

$$C_1^m + C_2^m + \dots + C_m^m = 2^m - 1 \quad (9)$$

Apriori 演算法:是 R.Agrawal 和 R.Srikant 提出的一種演算法，用於關聯規則挖掘頻繁項目集，採用逐級搜索的迭代方法，使用  $(k-1)$ -項目集探索  $k$ -項目集。先掃描資料集得到每個項目的數量，然後收集滿足最小支持閾值的項目，找到一組頻繁的 1-項目集，記為  $L_1$ ，接下來利用  $L_1$  找出頻繁的 2-項目集，記為  $L_2$ ，依此類推直到找不到更多的  $k$ -項目集，而找出每個  $L_k$  需要對資料集進行一次完整的掃描。

Apriori 有一個重要的特性：一個非空頻繁集合，其子集合必定也為頻繁。根據這個特性可以反向推斷出：若有一個集合為非頻繁，則它的所有超集合都為非頻繁。根據公式(6)(7)，若有一個項目集  $X$  不滿足最小支持度閾值，就算將一個項目  $B$  加到項目集  $X$  中產生新的項目集， $X \cup B$  在資料集中發生的頻率也無法更高。這一個特殊的特性稱為反單調性 (anti-monotone)，其意義是若一個集合無法通過一個測驗，則所有它的超集合也必定無法通過。

**表 3. Apriori 演算法虛擬碼**

Algorithm: Find frequent itemsets using an iterative level-wise approach based on candidate generation.
Input: $D$ , a database of transactions; $min\_sup$ , the minimum support count threshold.
Output: $L$ , frequent itemsets in $D$ .
Method:
(1) $L_1 = \text{find\_frequent\_1-itemsets}(D)$ ;
(2) for( $k=2$ ; $L_{k-1} \neq \emptyset$ ; $k++$ ) do
(3) $C_k = \text{apriori\_gen}(L_{k-1})$ ;
(4) for all transaction $t$ in $D$ do
(5) $C_t = \text{subset}(C_k, t)$ ;
(6) for all candidate $c \in C_t$ do
(7) $c.\text{count}++$ ;
(8) end
(9) $L_k = \{c \in C_k \mid c.\text{count} \geq min\_sup\}$
(10) end
(11) return $L$

### 3.3 新興模式探勘

新興模式可以捕獲資料集之間的重大變化和差異。不同活動之中可能包含相似的動作，在查找頻繁模式時可能受到混淆，我們使用新興模式找出活動之間更具有區別度的模式來解決此問題。根據 3.2.1 小節介紹的支持度，新興模式定義為支持度從一個資料集  $D1$  到另一個資料集  $D2$  顯著增加的項目集，或者不同的類別之間的有用對比。

增長率 (growth rate)：在 [16] 中定義了增長率來找出符合條件的新興模式，假設  $I$  是一個包含  $m$  個項目的項目集，如公式(5)， $D$  是一組資料集，其中包含的每筆交易  $T$  是一個非空的項目子集，滿足  $T \subseteq I$ 。欲找出  $D1$  與  $D2$  中的新興模式，將  $sup_{D1}(X)$  用  $sup_1(X)$  表示，項目集  $X$  從  $D1$  到  $D2$  的增長率記作  $GrowthRate(X)$ ，其定義如公式(10)，假設給定一個增長率的閾值  $p$ ，滿足  $p > 1$ ，如果  $GrowthRate(X) \geq p$ ，則我們稱  $X$  是從  $D1$  到  $D2$  的  $p$ -

新興模式。

$$\begin{cases} sup_1(X) = 0 \text{ and } sup_2(X) = 0, GrowthRate(X) = 0 \\ sup_1(X) = 0 \text{ and } sup_2(X) \neq 0, GrowthRate(X) = \infty \\ otherwise, GrowthRate(X) = \frac{sup_2(X)}{sup_1(X)} \end{cases} \quad (10)$$

dpower：我們希望能夠找出一個活動類別對於所有其他的活動類別差異度最大的模式，因此參考 [ ] 提出的另一種新興模式的評估方法，利用 dpower 來找出新興模式，其定義如公式 (11)，在一個資料集  $D$  中，在類別  $a_i$  中項目集  $X$  出現的次數與每個類別中項目集  $X$  出現次數的比例，乘以在非類別  $a_i$  中非項目集  $X$  出現的次數與每個類別中非項目集  $X$  出現次數的比例。設定最小 dpower 閾值，記作  $min\_power$ ，如果模式具有高於  $min\_power$  的 dpower 值，則將其視為新興模式，而當項目集  $X$  在單一類別中具有較高的支持度，在其餘類別中具有較低的支持度時，則  $X$  是擁有較高 dpower 的模式，而此模式可以高度代表該活動類別。利用頻繁模式與新興模式探勘找出具有區別度的模式後，我們將其當作分類器的特徵變數，將資料對這些特徵作特徵值評估，所得到的特徵向量用以訓練分類器。

$$dpower(X, a_i) = \frac{\text{count}_D(X, a_i)}{\text{count}_D(X)}$$

$$\frac{(|D| - \text{count}_D(X)) - (\text{count}_D(a_i) - \text{count}_D(X, a_i))}{(|D| - \text{count}_D(X))} \quad (11)$$

### 3.4 建模與辨識方法

特徵值評估：本研究使用模式探勘的方法獲取所有活動類別中的模式當作特徵變數，我們將每筆分段後的活動資料對這些特徵變數進行評估，評估的方法為計算每個特徵在每筆活動資料中所出現的比例。假設有一筆活動資料  $A$ ，若要計算特徵變數  $X$  的特徵值，我們計算  $X$  在活動資料  $A$  中出現的次數  $count_A(X)$ ，並除以活動總次數如公式 (12)。將每個特徵都以此計算，最後能得到一組特徵向量，以及對應的活動標籤，我們將之視為一組新的訓練資料，用以訓練分類器。

$$est(X) = \frac{\text{count}_A(X)}{|A|} \quad (12)$$

近鄰演算法 (k-Nearest Neighbor, KNN) [3]：是採用向量空間模型來分類，其概念為透過計算一資料點與其他資料的相似度從高至低排序，選擇前  $n$  個相似度高的資料，並透過投票機制來決定該資料的類別。相似度的計算有很多種，當輸入資料特徵為連續變數時常使用歐氏距離作為距離度量，本研究會使用  $n$  維的歐氏距離當作相似度的測量方法， $n$  是特徵向量的維度，若在實驗中取最相似的前三個樣本點投票選出預測結果，則  $k$  等於 3。近鄰演算法的優點是適合多分類問題，在低維的分類效果很好。缺點則是當維度增大時，因為要遍歷整個資料，會導致計算量龐大，再來是資料類別數量不平衡時效果比較差。

支援向量機 (Support Vector Machine, SVM) [18]：

假設有兩個類別集合要分類，找出一個決策邊界能夠將這兩類集合分開，並且希望這個決策邊界距離兩個類別集合的邊界越遠越好，減少分類器的泛化誤差。基本的支援向量機是一個二元分類器，若要分類多類別的任務時，會訓練一個分類器將一類別與其他類別分開作一對多的分類器，所以當有  $n$  類別需要分類時就訓練  $n$  個一對多的 SVM 來分辨，另外會使用徑向核函數(radial basis function kernel)，將特徵向量映射到高維使其線性可分。支援向量機在解決非線性及高維分類問題時效果好，泛化能力較強。缺點是在樣本數多的時候效率不高，且對缺失資料較敏感。

貝氏分類器(Naïve Bayes Classifier, NB) [19]：是機率模型分類器的一種，主要基於貝氏定理的監督式學習演算法，其認定所有樣本之中的每個特徵都不相關。假設一個條件機率模型有一獨立的類別變數  $C$  包含若干個類別，條件相依於其包含的特徵變數  $F_1, F_2, \dots, F_n$ ，如公式 (13)，分母因為是給定的特徵變數，所以看作是常數，當特徵維度數量  $n$  很大的時候會不利計算。根據單純貝氏分類器假定的條件，每個維度的特徵變數都獨立於其他特徵變數，也就是任兩個特徵變數  $F_i, F_j$  是條件獨立的，因此我們能將每類資料的機率密度函數簡化成此類別在每一維特徵的機率密度函數乘積，如公式(14)，算出一個類別在每一維度的特徵對應的機率密度函數，再將同類別的每個機率密度函數相乘，得到這一類別的機率密度函數。推導出貝氏機率模型後，可以根據規定決策規則來訓練分類器，其中最大後驗機率是一個常見的決策規則，以此對應的分類如公式 (15)。單純貝氏分類器在小規模的資料集中效果表現良好，不需要太多資料也能夠利用先驗知識分類，缺點是因為將每個維度的特徵變數視為相互獨立，若特徵之間相依性高則可能降低分類器的準確度。

$$p(C|F_1, \dots, F_n) = \frac{p(C)p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)} \propto$$

$$p(C)p(F_1, \dots, F_n|C) \dots \propto p(C)p(F_1|C)p(F_2|C, F_1) \dots p(F_n|C, F_1, F_2, \dots, F_{n-1}) \quad (13)$$

$$p(C|F_1, \dots, F_n) \propto p(C) \prod_{i=1}^n p(F_i|C) \quad (14)$$

$$\text{classify}(f_1, \dots, f_n) =$$

$$\arg_c \max p(C = c) \prod_{i=1}^n p(F_i = f_i | C = c) \quad (15)$$

## 4 實作與驗證

本章節總共會進行三個實驗，第一個實驗會觀察使用頻繁模式來辨識活動是否有幫助，並比較在不同支持度閾值下的分類準確度。第二個實驗會比較頻繁模式與加上新興模式對於分類結果的差異，並比較在不同  $d_{power}$  閾值下的分類準確度，第三個實驗是比較 kNN、SVM、NB 三個分類器使用頻繁模式加新興模式方法的分類準確度，並在最後一小節與其他研究討論實驗結果與研究結論。

### 4.1 系統架構與資料集

圖 1 是系統架構流程圖，系統輸入連續的動作

標籤，本研究使用資料集中的身體動作與複雜活動類別，每秒會有 32 筆感測器讀值與對應不同級別的活動標籤，實驗將身體動作標籤與對應的活動標籤當作輸入，並對其進行序列修剪以及資料分段，再將資料分割成訓練集和測試集，訓練時使用模式探勘篩選特徵變數，將得到特徵向量與對應活動標籤當作新的輸入給分類器訓練，測試時同樣利用訓練階段篩選的特徵變數做特徵提取，所得到特徵向量與對應活動標籤當作新的測試資料來評估分類器，詳細的實驗流程會在 4.2 小節說明。

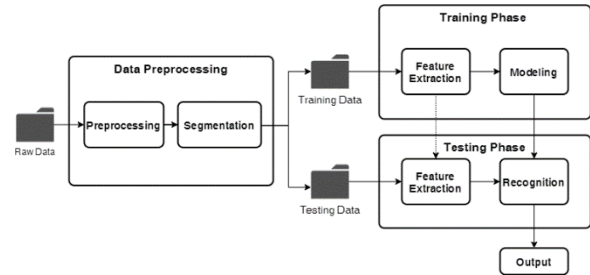


圖 1. 系統架構圖

本實驗使用 UCI 機器學習資料庫中為研究人類活動辨識所提供的 Opportunity 資料集，資料集中包括 4 位受測人員模擬使用者的紀錄，每位使用者執行 5 輪實驗，實驗內容為模擬日常進行一系列固定活動，本研究以左手、右手以及移動等 30 個動作類別，如表 4 來辨識 5 個活動類別，其中個別的流程如下：休息：躺在躺椅上，起身，躺回躺椅；早起：移動到房間，檢查所有在抽屜跟架子的物體都放在對的位子，移動到室外沿著建築物散步；喝咖啡：用咖啡機準備有加糖跟牛奶的咖啡，喝咖啡；吃三明治：使用切麵包機、幾種刀子和盤子，準備水與包含麵包、起司和火腿的三明治，吃三明治；清理：將使用後的物體歸位或是放入洗碗機中，並清理桌面。每個活動會有 20 筆動作子序列，實驗會將資料以每輪模擬日常分為 5 等份作交叉驗證。

表 4.30 種動作類別

Locomotion	Left Arm		Right Arm	
stand	unlock	clean	unlock	clean
walk	stir	bite	stir	bite
sit	lock	cut	lock	cut
lie	close	spread	close	spread
	reach	release	reach	release
	open	move	open	move
	sip		sip	

### 4.2 實驗與討論

我們會先將原始資料以 3.1 小節提到的方法簡化動作序列，接著根據資料集中標籤得到活動的邊界，將每筆動作序列依照活動分段為動作子序列，將不同活動的動作子序列分類成不同活動集，接下來將活動集資料的 80% 當作訓練集，其餘則當作測試集，結束資料處理的步驟。在訓練階段我們以模式探勘作為篩選特徵的方法，利用模式探勘方法篩選出訓練資料中活動的模式當作特徵變數，接著以每個模式在一筆訓練資料中動作的次數計算特徵值作為特徵擷取方法，得到的特徵向量與對應到一個活動標籤當作新的訓練資料，以此資料來訓練分類器。在測試階段，以每

個模式在一筆測試資料中動作的次數計算特徵值，得到的特徵向量與對應到一個活動標籤當作新的測試資料，輸入到分類器評估分類的結果，本研究會使用 python 的 scikit-learn 套件實作分類器。因為本研究是一個多分類的問題，總共五個類別，準確度為輸出的預測類別等於實際類別的次數，除以總分類次數，如公式(16)，m 為樣本數， $f(x_i)$  為預測的類別， $y_i$  為真實類別。

$$Acc: \frac{\sum_{i=1}^m f(x_i)=y_i}{m} \quad (16)$$

### 實驗 1：頻繁模式

本實驗目的是要觀察篩選出活動中常出現的動作當作特徵是否對於辨識活動有幫助，故使用所有的動作與頻繁模式中的動作當作特徵變數。本實驗考慮模式中有一個(1-pattern)或兩個(2-pattern)動作，並且比較在不同的支持度閾值的頻繁模式來觀察辨識結果，而分類器選擇 kNN (k=3)。

表 5. 在不同支持度閾值之 1、2-pattern 準確度

min sup	1-pattern	2-pattern
0.1	0.858	0.825
0.2	0.858	0.825
0.3	0.858	0.842
0.4	0.858	0.842
0.5	0.858	0.85
0.6	0.858	0.85
0.7	0.858	0.817
0.8	0.858	0.817

從表 5 顯示，不同支持度閾值在 1-pattern 的準確度都一樣，推測是因為其模式的數量差異不大，而表 6 顯示，只有在閾值等於 0.8 時多篩選掉 1~2 個模式。而在 2-pattern 時可以觀察出閾值越大或是越小都會降低準確度，可能原因是閾值越大會篩選掉過多的模式，而閾值越小則會沒有頻繁模式的篩選效果，與支持度本身的特性相符合，所以在本實驗觀察到當閾值在 0.5 至 0.6 左右時的篩選結果是最好的，這閾值下的頻繁模式代表出現在同類別活動中至少一半以上。

表 6. 不同支持度閾值的 1-pattern 頻繁模式數量

	1	2	3	4	5
0.1	26	26	26	26	26
0.5	23	23	23	23	23
0.8	23	22	20	21	21

圖 2 顯示，在支持度閾值等於 0.5 時，每個活動對頻繁模式之支持度，可以觀察出同一個頻繁模式可能會同時屬於複數活動，如圖 3，不同活動之頻繁模式相似度顯示出活動之間組成的相似度，其中可以看出休息相對於其他活動的動作組成相似度較低，而喝咖啡與吃三明治兩個活動的相似度最高。圖 2 編號 12 同屬於喝咖啡與吃三明治的頻繁模式，因為一個活動是喝咖啡，一個是喝水，但是編號 9 只屬於喝咖啡的頻繁模式。此觀察結果顯示相似的活動之中還是有區別度高的動作模式，因此在實驗 2 透過新興模式來解決這個問題。

```

*****frequent pattern support*****
pattern relaxing coffee time early morning cleanup sandwich time
0 (stand,) 0.65625 1.0000 1.0000 1.0000 1.0000
1 (walk,) 0.87500 1.0000 1.0000 1.0000 1.0000
2 (sit) 0.59375 NaN NaN 0.6875 1.0000
3 (lie) 1.00000 NaN NaN NaN NaN
4 (left_close,) NaN 1.0000 0.8750 1.0000 0.9375
5 (left_reach,) NaN 1.0000 0.8750 1.0000 1.0000
6 (left_open,) NaN 1.0000 0.8750 0.8125 1.0000
7 (left_release,) NaN 1.0000 0.8125 1.0000 1.0000
8 (left_move,) NaN 1.0000 NaN NaN 1.0000
9 (right_stir,) NaN 1.0000 NaN NaN NaN
10 (right_reach,) NaN 1.0000 1.0000 1.0000 1.0000
11 (right_open,) NaN 0.7500 1.0000 0.8750 0.9375
12 (right_sip,) NaN 0.9375 NaN NaN 1.0000
13 (right_release,) NaN 1.0000 1.0000 1.0000 1.0000
14 (right_move,) NaN 1.0000 NaN NaN 1.0000
15 (right_unlock,) NaN NaN 0.8750 NaN NaN
16 (right_lock,) NaN NaN 0.8125 NaN NaN
17 (right_close,) NaN NaN 0.8750 0.9375 1.0000
18 (right_clean,) NaN NaN NaN 0.9375 NaN
19 (left_bite,) NaN NaN NaN NaN 0.8750
20 (right_bite,) NaN NaN NaN NaN 0.9375
21 (right_cut,) NaN NaN NaN NaN 0.8750
22 (right_spread,) NaN NaN NaN NaN 0.8125
feature number: 23

```

圖 2. 各活動對頻繁模式之 1-pattern 支持度

	relaxing	coffee time	early morning	cleanup	sandwich time
relaxing	1.000000	0.133333	0.142857	0.200000	0.157895
coffee time	0.133333	1.000000	0.562500	0.687500	0.631579
early morning	0.142857	0.562500	1.000000	0.625000	0.500000
cleanup	0.200000	0.687500	0.625000	1.000000	0.684211
sandwich time	0.157895	0.631579	0.500000	0.684211	1.000000

圖 3. 不同活動之頻繁模式相似度

### 實驗 2：結合頻繁模式與新興模式

根據實驗 1 的結果，使用頻繁模式能夠篩選掉在一個活動中出現次數不多的動作，並且支持度閾值在 0.5~0.6 左右的效果最好，然而不同的活動可能有相同的頻繁模式，因此在本實驗目的是為了解決此問題。我們利用新興模式來篩選出活動之間區別度更大的模式當作特徵變數。本實驗考慮模式中有一個(1-pattern)或兩個(2-pattern)動作，根據實驗 1，支持度閾值選擇 0.5，並且比較在不同的 dpower 閾值下的新興模式對於辨識結果的影響，分類器同樣選擇 kNN (k=3)。

表 7. 不同 min dpower 之 1、2-pattern 準確度

min dpower	1-pattern	2-pattern
0.1	0.858	0.85
0.2	0.883	0.85
0.3	0.892	0.858
0.4	0.958	0.858
0.5	0.983	0.867
0.6	0.983	0.867
0.7	0.983	0.867
0.8	0.983	0.867

表 7 顯示，不同 dpower 閾值在 1-pattern 與 2-pattern 的準確度都呈現閾值越高，結果的準確度也越高，當閾值在 0.5 以上時準確度都一樣且最高。如表 8 所示，當 dpower 閾值為 0.1 時沒有篩選掉任何頻繁模式；0.5 時篩選掉 14~15 個頻繁模式；0.8 時又多篩選掉 1~2 個頻繁模式。閾值越小則會沒有新興模式的篩選效果，而越大則篩選出的新興模式在該活動出現的頻率比其他活動高。

表 8. 不同 dpower 閾值的 1-pattern 新興模式數量

	1	2	3	4	5
0.1	23	23	23	23	23
0.5	9	9	8	8	9
0.8	7	7	7	7	7

圖 4 顯示在支持度閾值等於 0.5、dpower 閾值等於 0.5 時，每個活動對新興模式之 dpower 值，可以觀察出同一個新興模式只屬於一個活動，如圖 5 顯示出活動之間組成的相似度，活動彼此之間的模

式相似度為 0。進一步觀察可以看到實驗 1 討論的右手攪拌的動作只屬於喝咖啡的新興模式，此外圖 4 可以看出最能夠代表每個活動的模式，例如編號 0，只有在休息的活動會有此動作。觀察結果顯示利用新興模式確實可以找出各活動差異度大的模式，解決相似的活動之間有相同模式的問題，而 dpower 閾值越大能夠篩選掉更多的模式，剩下來的新興模式則代表著比其他活動更高的出現頻率比例。

```

*****emerging pattern power*****
pattern relaxing coffee time early morning cleanup sandwich time
0 (lie,) 1.0 NaN NaN NaN NaN
1 (right_stir,) NaN 0.941176 NaN NaN NaN
2 (right_unlock,) NaN NaN 0.592019 NaN NaN NaN
3 (right_lock,) NaN NaN 0.541989 NaN NaN NaN
4 (right_clean,) NaN NaN NaN 0.925781 NaN NaN
5 (left_bite,) NaN NaN NaN NaN 0.975610
6 (right_bite,) NaN NaN NaN NaN 0.871184
7 (right_cut,) NaN NaN NaN NaN 0.975610
8 (right_spread,) NaN NaN NaN NaN 0.894599
feature number: 9

```

圖 4. 各活動對新興模式之 1-pattern 的 dpower

```

*****similarity*****
relaxing relaxing coffee time early morning cleanup sandwich time
coffee time 0.0 1.0 0.0 0.0 0.0 0.0
early morning 0.0 0.0 1.0 0.0 0.0 0.0
cleanup 0.0 0.0 0.0 1.0 0.0 0.0
sandwich time 0.0 0.0 0.0 0.0 1.0 1.0

```

圖 5. 不同活動之新興模式相似度

### 實驗 3：不同分類器比較

實驗 2 的結果證實了使用新興模式能夠篩選掉重複屬於不同活動的模式，接下來要觀察新興模式在不同的分類器下是否有相同的效果。以實驗 2 設置的條件，kNN (k=3)、SVM 與 NB 三種分類器使用頻繁模式與搭配新興模式兩種方法作比較。

表 9. 頻繁模式與新興模式於三種分類器之準確度

分類器	頻繁模式		頻繁+新興模式	
	1-pattern	2-pattern	1-pattern	2-pattern
kNN	0.858	0.85	0.983	0.867
SVM	0.842	0.825	0.883	0.842
NB	0.95	0.858	0.925	0.858

實驗結果如表 9，我們可以從中觀察出，不管是頻繁模式還是加上新興模式的方法，1-pattern 的準確度都大於 2-pattern，其原因可能是實驗沒有考慮動作之間的時間關係。在支持度閾值為 0.5、dpower 閾值為 0.5 時，coffee time 活動類別考慮 1-pattern 的新興模式只有 right\_stir，而 2-pattern 時的新興模式如圖 6，可以看到所有 2-pattern 的新興模式都是 right\_stir 與其他動作的組合，在計算特徵值時每個特徵變數會因為都有 right\_stir，反而是依賴另一個動作的特徵值來辨識，但另一個動作不具有區別度，造成 2-pattern 的辨識效果不好。再來是頻繁模式加新興模式的準確度一直都大於只使用頻繁模式，代表使用新興模式所篩選的模式當作特徵變數確實能夠提升辨識準確度。最後是討論不同分類器之間的差異，可以看到 1-pattern 且分類器為 kNN 時，使用頻繁模式加新興模式的數量不多，其準確度是最高的，因為 kNN 在低維的分類效果很好。而 SVM 相較於其他兩個分類器在 1-pattern 和 2-pattern 表現都較差，推測可能是 SVM 需要更多的特徵才能發揮它的分類效果。最後 NB 在 1-pattern 與 2-pattern 的準確度比起其他兩個分類器有著更大的差異，其原因可能是 NB 假設每個特徵變數都是獨立的，而 2-pattern 時每個變數之間存在相依性，與 NB 假設的前提不相符合，也就不能發揮其分類

效果，所以準確度會差比較多。

```

(stand, right_stir)
(walk, right_stir)
(right_stir, left_close)
(right_stir, left_reach)
(right_stir, left_open)
(right_stir, left_release)
(right_stir, left_move)
(right_stir, right_reach)
(right_stir, right_open)
(right_stir, right_sip)
(right_stir, right_release)
(right_stir, right_move)

```

圖 6. coffee time 在 2-pattern 的新興模式

### 4.3. 總結

結合頻繁模式與新興模式的方法對於辨識的準確度有顯著的提升，本研究以支持度閾值 0.5、dpower 閾值 0.5 的辨識結果，與[13]的研究結果比較準確度，如表 10。當考慮 1-pattern 時，在三種不同的分類器下的準確度均比較高，而考慮 2-pattern 時的準確度則都比較低，其中本研究最好的準確度是在 1-pattern 下使用 kNN，而[13]則是在 {1,2}-pattern 下使用 NB。在[13]中使用 Allen 的時間關係將每個動作又分解成不同時間關係的組成，增加了模式的數量以及模式中的時間資訊，但是在找 (k+1)-pattern 時所需要的時間複雜度也會大幅增加，而本研究提出的方法則透過篩選模式來提高辨識準確度，所採用的模式數量更少。

表 10. 與其他研究之比較

分類器	頻繁模式+時間關係		頻繁+新興模式	
	1-pattern	2-pattern	1-pattern	2-pattern
kNN	0.885	0.905	0.983	0.867
SVM	0.875	0.94	0.883	0.842
NB	0.865	0.98	0.925	0.858

在實驗中可以看出使用頻繁模式結合新興模式能篩選出與其他活動差異度更大的模式，在實驗 1 與實驗 2 中我們分別觀察出支持度的閾值太高或太低篩選效果都不好，而 dpower 的閾值則是越高越好，但是高到一定程度後就沒有差異。我們還考慮了一個模式中有不同數量動作，並且使用三種不同的分類器來驗證，而實驗結果顯示結合這兩個模式探動方法都比使用頻繁模式要有更高的準確度。本研究中 1-pattern 使用 kNN 結合兩種模式的分類效果最好，與其他論文比較可以看到我們使用更少的模式數量達到更好的辨識結果。在 2-pattern 的辨識結果效果比較差，可能是模式中的動作間關聯性不大。而其他論文使用 Allen 的 13 種時間關係來表示動作之間的關聯，增加每個模式所擁有的資訊量並擴展特徵維度，但也因此大幅度增加了包括找出頻繁模式以及訓練分類器的計算量。本研究結合兩個模式的方法，能夠找出同一活動頻繁出現的模式，以及判別包含相同模式的相似活動。若不同使用者執行活動差異非常大，或是活動之間相似度越大，則需要減低閾值避免篩選過程中過濾掉過多的模式，但也可能因此影響其他活動的篩選結果。

### 5 結論與未來展望

基於慣性感測器的人類活動辨識研究中，因大部分著重在簡單動作與身體移動的辨識上，而在複雜活動的辨識上的研究較少，其結果也相對較差，

故本研究利用身體動作當作資料來辨識高階的活動。本研究以模式探勘做為篩選特徵的方法，先將原始資料分段成代表不同活動的動作子序列，再利用兩個模式探勘方法篩選出的模式當作特徵變數，以每個模式在一個活動的次數做為特徵值評估方法，得到的特徵向量對應到一個活動標籤，以此資料來訓練分類器。在實驗中比較了使用頻繁模式以及結合新興模式的方法作為特徵篩選，討論在不同支持度與 dpower 閾值的條件下，考慮模式中不同數量的動作，在三個分類器下對於辨識活動準確度的影響。觀察模式的篩選可以看出，只使用頻繁模式所篩選出來的模式，可能是同時屬於不同活動的頻繁模式，而新興模式則具有活動類別之間的獨立性，實驗結果顯示使用頻繁模式加新興模式做為特徵變數比起只使用頻繁模式在不同分類器下都有更好的分類結果。再與其他研究相比，本研究使用了更少的模式當作特徵，並利用 kNN 當作分類器達到 98% 的辨識準確度。

本研究在一個模式中有兩個以上的動作時雖然準確度有提高，效果卻不及一個動作的模式，在未來可嘗試增加動作之間的關聯性如時間關係，將模式的組成細分，再利用模式篩選出有區別度的模式，另外本研究的特徵值評估方法為計算每個動作的出現次數，也可考慮其他因素來改善評估方法。在辨識階段，本研究的分段是將動作序列根據已知的邊界直接分為不同的動作子序列，未來可研究如何針對動作序列找出活動的邊界。另外可加入其他有身體動作與高級別活動的資料集驗證本實驗提出的方法，並且結合慣性感測器讀值辨識動作，與利用動作序列辨識高階活動的兩層架構來辨識活動。

## 6 參考文獻

- [1] Begg, R., & Hassan, R. "Artificial neural networks in smart homes." In *Designing smart homes*, Springer, Berlin, Heidelberg, 2006, 146-164.
- [2] Bulling, A., Blanke, U., & Schiele, B. "A tutorial on human activity recognition using body-worn inertial sensors." *ACM Computing Surveys (CSUR)*, 46(3). 2014, 1-33.
- [3] Collobert, R., & Bengio, S. "Links between perceptrons, MLPs and SVMs." In *Proceedings of the twenty-first international conference on Machine learning*. 2004.
- [4] Dernbach, S., Das, B., Krishnan, N. C., Thomas, B. L., & Cook, D. J. "Simple and complex activity recognition through smart phones." In *2012 eighth international conference on intelligent environments*, IEEE. 2012, 214-221.
- [5] Dong, Guozhu, and Jinyan Li. "Efficient mining of emerging patterns: Discovering trends and differences." *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*. 1999.
- [6] Ferrari, A., Micucci, D., Mobilio, M., & Napolitano, P. "Human activities recognition using accelerometer and gyroscope." In *European Conference on Ambient Intelligence*, Springer, Cham. 2019, 357-362.
- [7] Gu, T., Wu, Z., Tao, X., Pung, H. K., & Lu, J. "epsicar: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition." In *2009 IEEE International Conference on Pervasive Computing and Communications*, IEEE. 2009, 1-9.
- [8] Han, J., Kamber, M., & Pei, J. "6-mining frequent patterns, associations, and correlations: Basic concepts and methods." *Data Mining (Third Edition)*, The Morgan Kaufmann Series in Data Management Systems. 2012, 243-278.
- [9] He, K., Zhang, X., Ren, S., & Sun, J. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, 770-778.
- [10] Krishnan, N., Cook, D. J., & Wemlinger, Z. "Learning a taxonomy of predefined and discovered activity patterns." *Journal of ambient intelligence and smart environments*, 5(6), 2013, 621-637.
- [11] Lin, W. C., Tsai, C. F., Hu, Y. H., & Jhang, J. S. "Clustering-based undersampling in class-imbalanced data." *Information Sciences*, 409. 2017, 17-26.
- [12] Liu, Y., Nie, L., Liu, L., & Rosenblum, D. S. "From action to activity: sensor-based activity recognition." *Neurocomputing*, 181. 2016, 108-115.
- [13] Malazi, H. T., & Davari, M. "Combining emerging patterns with random forest for complex activity recognition in smart homes." *Applied Intelligence*, 48(2). 2018, 315-330.
- [14] Ni, Q., Garcia Hernando, A. B., la Cruz, D., & Pau, I. "The elderly's independent living in smart homes: A characterization of activities and sensing infrastructure survey to facilitate services development." *Sensors*, 15(5). 2015, 11312-11362.
- [15] Parkka, J., Ermes, M., Korpipaa, P., Mantyjarvi, J., Peltola, J., & Korhonen, I. "Activity classification using realistic data from wearable sensors." *IEEE Transactions on information technology in biomedicine*, 10(1). 2006, 119-128.
- [16] Peterson, L. E. "K-nearest neighbor." *Scholarpedia*. 2009.
- [17] Prekop, Paul, and Mark Burnett. "Activities, context and ubiquitous computing." *Computer communications* 26.11. 2003, 1168-1176.
- [18] Rish, I. "An empirical study of the naive Bayes classifier." In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, 2001, 41-46.
- [19] Roggen, D., Calatroni, A., Rossi, M., Holleczeck, T., Förster, K., Tröster, G., ... & Doppler, J. "Collecting complex activity datasets in highly rich networked sensor environments." In *2010 Seventh international conference on networked sensing systems (INSS)*, IEEE. 2010, 233-240.
- [20] Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. "Deep learning for sensor-based activity recognition: A survey." *Pattern Recognition Letters*, 2019, 3-11.
- [21] Wang, L., Gu, T., Tao, X., & Lu, J. "A hierarchical approach to real-time activity recognition in body sensor networks." *Pervasive and Mobile Computing*, 8(1). 2012, 115-130.