

Nhận dạng hoạt động của người bằng điện thoại thông minh

Human Activity Recognition Using Smart Phones

Nguyễn Thăng Ngọc, Phạm Văn Cường

Abstract: In this paper we propose a method and system for recognizing everyday human activities by utilizing the acceleration sensing data from the accelerometer instrumented in smart phones. In our method, human activities are recognized in four steps: data processing, data segmentation, feature extraction, and classification. We rigorously experimented on a dataset consisting of 6 everyday activities collected from 17 users using several machine learning algorithms, including support vector machine, Naïve Bayesian networks, k -Nearest Neighbors, Decision Tree C4.5, Rule Induction, and Neutral networks. The best accuracies are achieved by Decision Tree C4.5 that demonstrates the human activities can be distinguished with 82% precision and 83% recall under the leave-one-subject out evaluation protocol. These results have shown the feasibility of smart phone based real-time activity recognition. In addition, the proposed method based on Decision Tree has been deployed on Samsung smart phones and is able to recognize 6 human activities in real-time.

Keywords: Activity Recognition, Feature Extraction, Mobile Device, Accelerometer, Machine Learning.

I. GIỚI THIỆU

Tự động nhận dạng hoạt động của người (human activity recognition – HAR) đang thu hút được sự quan tâm của cộng đồng nghiên cứu vì có nhiều ứng dụng thực tế như: trợ giúp chăm sóc sức khỏe [1], trợ giúp theo dõi chế độ ăn kiêng (dietary monitoring) [2], ước lượng lượng calorie tiêu thụ ở người (energy

expenditure estimation) [3] v.v.. Mặc dù các nghiên cứu về nhận dạng hoạt động của người đã và đang đạt được những kết quả đáng kể, nhưng phần lớn các nghiên cứu này đã phát triển và triển khai các phương pháp nhận dạng trên máy tính cá nhân. Điều này làm hạn chế giới hạn không gian thực hiện các hoạt động của người (không quá xa để dữ liệu cảm biến có thể kết nối được với máy tính) khiến cho việc thực hiện nhận dạng của người bị giới hạn trong một phòng [12], trong một ngôi nhà [4,15,17,20] hay trong bếp [2,11,13,16]. Trong khi đó khả năng tính toán và số lượng [5] của các điện thoại thông minh ngày càng được nâng cao với sự tích hợp của các bộ vi xử lý đa nhân (multicore processors), bộ nhớ đệm và các bộ cảm biến như cảm biến gia tốc, cảm biến nhiệt, cảm biến định vị v.v.. Những tài nguyên tính toán này có thể sẽ cho phép chúng ta triển khai và thử nghiệm các phương pháp học máy và nhận dạng thích hợp trên điện thoại thông minh. Nếu được như vậy thì các hoạt động hàng ngày của người có thể được nhận dạng mọi nơi, mọi lúc.

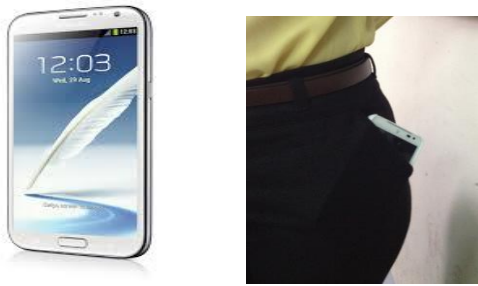
Các nghiên cứu trước đây tiếp cận bài toán nhận dạng hoạt động của người theo ba phương pháp: phân tích hình ảnh hoạt động của người bằng camera hay còn gọi thị giác máy (computer vision), phân tích dữ liệu cảm biến từ các bộ cảm biến được tích hợp vào môi trường hoặc vật dụng (pervasive sensing) và phân tích dữ liệu cảm biến từ các bộ cảm biến được đeo trên người (wearable sensing). Trong cách tiếp cận bằng thị giác máy [11,12], hình ảnh từ camera được tiến hành tiền xử lý, trích chọn các đặc trưng và phân loại. [11] nhận dạng các đối tượng (vật dụng, thức ăn) và đề

xuất một mạng Bayes để suy diễn các hoạt động nấu nướng, trong khi [12] đề xuất một biến thể của mô hình Markov ẩn được huấn luyện từ các đặc trưng (tư thế của người) để suy diễn hoạt động. Việc sử dụng camera để nhận dạng hoạt động thường yêu cầu các hoạt động của người bị giới hạn trong không gian hẹp. Hơn nữa, hiệu suất nhận dạng thường chịu ảnh hưởng bởi các yếu tố từ môi trường như: ánh sáng, hình ảnh bị che khuất v.v.. Trong khi đó thì cách tiếp cận cảm biến tỏa khắp (pervasive sensing) sử dụng các bộ cảm biến được tích hợp vào môi trường [4,15,16,17,18] hoặc tích hợp vào vật dụng [2,13,16] khắc phục được nhược điểm này. Tuy nhiên cách tiếp cận bằng cảm biến tỏa khắp thường yêu cầu sử dụng nhiều bộ cảm biến khiến chi phí khá cao và yêu cầu môi trường phải được thiết lập trước chẳng hạn như phải đặt bộ cảm biến vào môi trường trước khi con người thực hiện các hoạt động. Ngược lại, sử dụng cảm biến đeo trên người [8,9,10,18] có thể nhận dạng hoạt động của người trong phạm vi rộng hơn (như tòa nhà) mà không cần phải sử dụng nhiều bộ cảm biến và thiết lập môi trường trước (pre-setting). Chẳng hạn, một trong những nghiên cứu trước đây của chúng tôi [18] chỉ sử dụng 2 Wii Remotes, một loại thiết bị có chi phí khá thấp (25\$) thường được sử dụng trong các trò chơi tương tác, đeo ở thắt lưng và tay để nhận dạng 14 hoạt động hàng ngày của người với độ chính xác (precision) và độ bao phủ (recall) tới hơn 90%. Khác với điện thoại là thiết bị có thể được mang theo người mọi nơi, mọi lúc, Wii khá cồng kềnh và chưa phù hợp để đeo cả ngày trên người. Hơn nữa, chương trình phân tích dữ liệu cảm biến và thuật toán nhận dạng (mô hình Markov ẩn) trong nghiên cứu [18] được triển khai trên máy tính cá nhân nên cần được cải tiến cho phù hợp khi triển khai trên thiết bị di động.

Một số ít các nghiên cứu trước đây đã nghiên cứu triển khai phương pháp và hệ thống nhận dạng hoạt động của người trên điện thoại thông minh [6,7]. Các nghiên cứu này thường có hai điểm hạn chế là: tập dữ liệu thử nghiệm thường không lớn và các phương pháp nhận dạng đề xuất chưa phù hợp cho nhận dạng các hoạt động của người trong ngữ cảnh thời gian thực. Ví

dụ như Jennifer và các cộng sự [6] đã đề xuất một phương pháp nhận dạng để nhận dạng 6 hoạt động: *đi bộ, chạy tại chỗ, đi lên cầu thang, đi xuống cầu thang, ngồi và đứng*. Các mẫu dữ liệu thu thập từ các điện thoại thông minh Nexus One, HTC Hero và Motorola back-flip được trích chọn các đặc trưng. Những đặc trưng này được sử dụng để huấn luyện và kiểm thử một số bộ phân loại từ Weka [21]. Kết quả cho thấy bộ phân loại perceptron đa lớp (multi-layer perceptron) cho kết quả cao nhất với độ chính xác 91.7%. Hạn chế chính của nghiên cứu này là thời gian huấn luyện và nhận dạng của mạng perceptron đa lớp thường chậm trong các trường hợp số lớp và số lượng nút mạng trong từng lớp tăng. Điều này không phù hợp cho việc nhận dạng hoạt động trong thời gian thực. Một nghiên cứu khác [7] đề xuất thuật toán hỗ trợ máy véc tơ đa lớp (multi-class support vector machine) để nhận dạng 6 hoạt động: *đi bộ, đi lên cầu thang, đi xuống cầu thang, nằm xuống, ngồi và đứng*. Mặc dù đạt được độ chính xác trung bình là 89% nhưng [7] cũng chưa phù hợp cho nhận dạng hoạt động của người trong thời gian thực. Bởi vì hoạt động tự nhiên của người không bị giới hạn chỉ trong 6 hoạt động trên trong khi bất cứ hoạt động nào của người thì sẽ được hệ thống [6,7] phân loại là một trong sáu hoạt động mà họ đã thử nghiệm.

Trong nghiên cứu này, chúng tôi đề xuất một phương pháp nhận dạng hoạt động của người trên điện thoại thông minh được kế thừa từ nghiên cứu trước đây của chúng tôi [18]. Đây là phương pháp có khả năng nhận dạng các hoạt động của người trong thời gian thực bao gồm cả những hoạt động chưa được gán nhãn khi huấn luyện (unknown activities hay other activities). Nhưng khác với [18], chúng tôi đề xuất một tập các đặc trưng mới phù hợp với vị trí và cấu hình của cảm biến gia tốc trong các điện thoại thông minh hiện nay và tiến hành thử nghiệm các đặc trưng này trên 6 thuật toán khác nhau và so sánh các kết quả. Trong đó, thuật toán cho kết quả tốt nhất sẽ được cài đặt (implementation) trên điện thoại thông minh để nhận dạng các hoạt động của người trong thời gian thực.



Hình 1. Điện thoại Samsung Galaxy Note 2 (trái) và vị trí đặt điện thoại (phải)

II. PHƯƠNG PHÁP NHẬN DẠNG HOẠT ĐỘNG CỦA NGƯỜI

Trong phần này chúng tôi trình bày các bước nhận dạng hoạt động của người, bao gồm: *tiền xử lý dữ liệu (pre-processing)*, *phân đoạn dữ liệu (segmentation)*, *trích chọn các đặc trưng (feature extraction)*, và *phân loại (classification)*. Phần này kết thúc bằng trình bày về quá trình *triển khai (implementation)* thuật toán cho nhận dạng hoạt động của người trong thời gian thực.

II.1. Tiền xử lý dữ liệu

Do dữ liệu cảm biến thường chứa nhiễu (noise). Trong trường hợp lý tưởng khi hoạt động ở tần số 100 Hz thì ta sẽ thu được 100 mẫu từ cảm biến gia tốc trong 1 giây. Trên thực tế, vì nhiều lý do ảnh hưởng bởi môi trường hoặc chính thiết bị nên số mẫu thu được có thể ít hơn 100 mẫu trong một giây. Trong trường hợp này chúng tôi sử dụng kỹ thuật nội suy Spline [19] để tái tạo (resampling) các mẫu đã mất do nhiễu. Ngoài ra, giá trị cảm biến có thể biến đổi bất thường (abnormal) quá cao hoặc quá thấp (có thể do va đập hay do chuyển động bất thường của người) thì chúng tôi áp dụng kỹ thuật lọc dải thông cao (high-pass filtering [22]) để loại bỏ giá trị cảm biến *cao* bất thường và kỹ thuật lọc dải thông thấp (low-pass filtering [22]) để loại bỏ giá trị cảm biến *thấp* bất thường.

II.2. Phân đoạn dữ liệu

Dòng dữ liệu cảm biến biến đổi liên tục theo thời gian với tần số 100 mẫu/giây. Để nhận dạng các hoạt động của người trong một khoảng thời gian xác định

thì dòng dữ liệu này phải được phân đoạn thành các *cửa sổ trượt (sliding windows)*. Các nghiên cứu [8, 13, 18] đã chỉ ra rằng kích thước của cửa sổ trượt ảnh hưởng đáng kể đến độ chính xác nhận dạng của thuật toán học máy. Tùy theo ứng dụng mà ta có thể lựa chọn độ dài cửa sổ trượt sao cho phù hợp. Chẳng hạn [8] chọn cửa sổ trượt có độ dài 4,3 giây để nhận dạng các hoạt động như *lau nhà (cleaning)*, *đạp xe (cycling)*... trong khi [13] lựa chọn cửa sổ trượt với độ dài 1,5 giây để nhận dạng các hoạt động nấu nướng như *thái (chopping)*, *trộn thức ăn (stirring)* v.v. Trong nghiên cứu này, chúng tôi lựa chọn cửa sổ trượt có độ dài là 2 giây với thời gian chồng lấn giữa 2 cửa sổ trượt kế tiếp nhau là 1 giây (50% overlapping). Việc lựa chọn này dựa trên kết quả một thực nghiệm kiểm tra chéo 4 lần (4-fold cross validation) trên một tập dữ liệu con (subset) đối với các cửa sổ có kích thước: 1 giây (độ chính xác 73,4%), 1,5 giây (độ chính xác: 75,1%), 2 giây (độ chính xác: 78,3%), và 3 giây (độ chính xác: 77,9%). Như vậy khi hệ thống chạy trong thời gian thực thì kết quả nhận dạng sẽ được trả về sau 1 giây. Việc lựa chọn độ dài cửa sổ như vậy là hợp lý vì trong các hoạt động mà chúng tôi nhận dạng có một số hoạt động diễn ra trong khoảng thời gian tương đối ngắn như: *nhảy (jumping)* và *ngồi xuống (sitting)*. Hơn nữa, với cửa sổ trượt là 2 giây hệ thống nhận dạng sẽ tránh được trì hoãn khi thực hiện nhận dạng các hoạt động trong thời gian thực.

II.3. Trích chọn các đặc trưng

Lựa chọn ra các đặc trưng là công việc quan trọng cần được tiến hành trước khi huấn luyện các mô hình học máy. Với các thuật toán học máy được triển khai trên thiết bị di động thì việc trích chọn các đặc trưng còn bị ràng buộc bởi tài nguyên tính toán hạn chế. Chính vì vậy, trong số các đặc trưng được trích chọn trong nghiên cứu này, có một số đặc trưng được kế thừa từ nghiên cứu [23]. Với mỗi cửa sổ trượt với độ dài 2 giây chứa 200 mẫu dữ liệu (sau pha tiền xử lý), chúng tôi trích chọn ra 6 đặc trưng.

Trước hết, hai tín hiệu bất biến với độ nghiêng (tilt invariant) từ tín hiệu gia tốc 3 chiều tại thời điểm t được tính như sau:

$$\vec{a}_0(t) = [a_{0x}(t), a_{0y}(t), a_{0z}(t)] \quad (1)$$

Vec tơ độ nghiêng $\vec{b}(t)$ được tính bằng trung bình cộng của các giá trị $\vec{a}_0(t)$ trên N mẫu với N là số mẫu trong một cửa sổ trượt ($N=200$).

$$\vec{b}(t) = \frac{1}{N} \sum_{t=1}^N \vec{a}_0(t) \quad (2)$$

Sau đó các góc nghiêng (tilt angles) θ_1 và θ_2 được tính bằng đẳng thức sau:

$$\theta_1 = \arctan\left(\frac{b_y}{b_x}\right) \quad (3)$$

$$\theta_2 = \arctan\left(\frac{b_z}{b_y \sin \theta_1 + b_x \cos \theta_1}\right) \quad (4)$$

Véc tơ gia tốc bù độ nghiêng (tilt compensated acceleration vector) là:

$$\vec{a}_1(t) = \begin{bmatrix} \cos \theta_2 & -\sin \theta_1 \sin \theta_2 & -\cos \theta_1 \sin \theta_2 \\ 0 & \cos \theta_1 & -\sin \theta_1 \\ \sin \theta_2 & \sin \theta_1 \cos \theta_2 & \cos \theta_1 \cos \theta_2 \end{bmatrix} * \vec{a}_0(t) \quad (5)$$

Véc tơ gia tốc trọng trường và bù độ nghiêng tức thời:

$$\vec{a}_2(t) = \left[a_{1x}(t), a_{1y}(t), (a_{1z}(t) - \frac{1}{N} \sum_{t=1}^N a_{1z}(t)) \right] \quad (6)$$

Để loại bỏ các nhiễu do việc xoay (orientation) cảm biến gia tốc (do hướng điện thoại có thể bị thay đổi khi đang thực hiện hoạt động) thì giá trị x và y được xem như là 2 thành phần (thực và ảo) của một số phức. Đặc trưng được trích chọn từ hai tín hiệu gia tốc theo chiều dọc $a_v(t) = a_{2x}(t)$ và cường độ tín hiệu gia tốc theo chiều ngang:

$$a_h(t) = \sqrt{a_{2x}(t)^2 + a_{2y}(t)^2} \quad (7)$$

Hai đặc trưng đầu tiên được trích chọn là:

- Năng lượng gia tốc theo chiều dọc (vertical accelerometer energy: VE):

$$e_v = \int_{t=t_0}^{t_0+T} |a_v| dt \quad (8)$$

- Năng lượng gia tốc theo chiều ngang (horizontal accelerometer energy: HE):

$$e_h = \int_{t=t_0}^{t_0+T} |a_h| dt \quad (9)$$

Bốn đặc trưng còn lại là:

Hjorth chuyển động (Hjorth mobility: HM) [23]: mô tả các đường cong hình dạng (curve) của chuyển động bằng việc tính trung bình các độ dốc tương đối (relative slope). Công thức tính HM như sau:

$$HM = \sqrt{\frac{\text{var}(x(t) \frac{dy}{dt})}{\text{var}(x(t))}} \quad (10)$$

Trong đó, mẫu số dưới căn bậc 2 $\text{var}(x(t))$ là biến trạng (variance) của tín hiệu cảm biến theo chiều x theo thời gian t , còn tử số là đạo hàm bậc 1 của $\text{var}(x(t))$.

Hjorth phức hợp (Hjorth complexity: HC) [23]: là tỉ số giữa đạo hàm bậc nhất của tín hiệu chuyển động và tín hiệu khi đứng yên. Công thức tính HC như sau:

$$HC = \frac{HM(x(t) \frac{dx}{dt})}{HM(x(t))} \quad (11)$$

Các đặc trưng HM và HC được áp dụng cho dữ liệu cảm biến gia tốc theo cả chiều dọc (Vertical Hjorth mobility: VM) và chiều ngang. Do đó, ta có tổng số 6 đặc trưng được trích chọn trong mỗi cửa sổ trượt.

II.4. Các thuật toán phân loại

Các véc tơ đặc trưng sau khi được trích chọn từ các cửa sổ trượt sẽ được sử dụng để huấn luyện các bộ phân loại: máy véc tơ hỗ trợ (support vector machine: SVM), mạng Bayes đơn giản (Naïve Bayes), k láng giềng gần nhất (k-Nearest neighbors), cây quyết định (Decision Tree C4.5), suy diễn dựa trên luật (Rule Induction), mạng nơ ron (Neural network). Trong đó bộ phân loại máy véc tơ hỗ trợ sử dụng trong nghiên cứu này là SVMlib được phát triển bởi Chang et al.

[24] với hàm nhân rbf (radial basic function) và tham số $C=1$. Các bộ phân loại còn lại có sẵn trong thư viện RapidMiner [25].

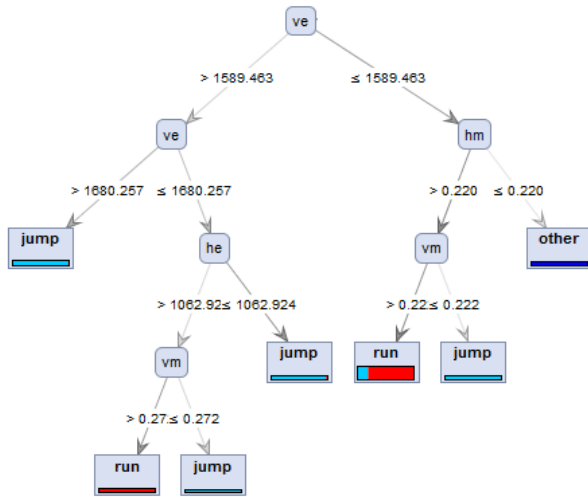
II.5. Cài đặt thuật toán

Sau khi tiến thử nghiệm (chi tiết được trình bày trong mục III), cây quyết định (Decision Tree C4.5) với độ đo lựa chọn đặc trưng là thông tin đạt được (information gain: IG) cho kết quả nhận dạng tốt nhất và được lựa chọn để phát triển cho hệ thống nhận dạng hoạt động của người trong thời gian thực. IG được xác định như sau:

$$Gain(A) = Info(D) - Info_A(D) \quad (12)$$

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (13)$$

$$Info_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} * Info(D_j) \quad (14)$$



Hình 2: Cây quyết định

Trong đó p_i là xác suất của một véc tơ đặc trưng bất kỳ trong tập dữ liệu D có nhãn là hoạt động (lớp) C_i và được ước lượng bằng $|C_{i,D}| / |D|$. Đặc trưng với giá trị đạt được lớn nhất (largest gain) sẽ được sử dụng để phân nhánh (split) cho cây. Trong tập dữ liệu của chúng tôi thì giá trị đạt được của các đặc trưng VE là 1.233, HE là 1.459, VM là 0.917 và HM là 0.197. Do đó, đặc trưng HE được lựa chọn để phân nhánh cho cây quyết định như minh họa trong Hình 2.

Tree

```

ve > 782.095
| vm > 0.228
| | vc > 0.796
| | | ve > 1629.051: jump {jump=25, run=0, walk=0, sit=0, stand=0, other=0}
| | | ve ≤ 1629.051
| | | | hc > 1.189
| | | | | he > 663.951
| | | | | | he > 1413.136: run {jump=0, run=5, walk=0, sit=0, stand=0, other=0}
| | | | | | he ≤ 1413.136
| | | | | | | ve > 1444.058
| | | | | | | | he > 1191.543: run {jump=0, run=3, walk=0, sit=0, stand=0, other=0}
| | | | | | | | he ≤ 1191.543
| | | | | | | | | vc > 0.804: jump {jump=21, run=0, walk=0, sit=0, stand=0, other=0}
| | | | | | | | | vc ≤ 0.804
| | | | | | | | | | ve > 1513.149: jump {jump=1, run=0, walk=0, sit=0, stand=0, other=0}
| | | | | | | | | | ve ≤ 1513.149: run {jump=0, run=2, walk=0, sit=0, stand=0, other=0}
| | | | | | | | | | | ve ≤ 1444.058
| | | | | | | | | | | | hc > 1.334
| | | | | | | | | | | | hc > 1.465: run {jump=0, run=5, walk=0, sit=0, stand=0, other=0}
| | | | | | | | | | | | hc ≤ 1.465: jump {jump=18, run=8, walk=0, sit=0, stand=0, other=0}
| | | | | | | | | | | | hc ≤ 1.334
| | | | | | | | | | | | hc > 1.270: run {jump=0, run=11, walk=0, sit=0, stand=0, other=0}

```

Hình 3: Mô hình phân loại cây quyết định

Thiết kế chi tiết hệ thống nhận dạng hoạt động của người cho điện thoại thông minh được mô tả trong Hình 3. Các giá trị cảm biến gia tốc ba chiều X, Y, Z được lưu trong một vùng đệm (pumping buffer) để tính véc tơ đặc trưng trong thời gian thực. Véc tơ này là đầu vào của mô hình phân loại đã được huấn luyện từ tập dữ liệu thu thập được. Đầu ra của mô đun phân loại cho ra một tin nhắn (text message) là một trong 6 hoạt động: chạy (run), nhảy (jump), đi bộ (walk), ngồi (sit), đứng (stand) và hoạt động chưa có nhãn khác (other). Tin nhắn này là đầu vào của một mô đun thông báo tới người dùng bằng giọng nói (text-to-speech).

III. THỬ NGHIỆM VÀ ĐÁNH GIÁ

Phần này chúng tôi trình bày về các bước tiến hành thử nghiệm và đánh giá phương pháp nhận dạng hoạt động của người: quá trình thu thập dữ liệu, giới thiệu các độ đo đánh giá hệ thống, các kết quả và đánh giá.

III.1. Thu thập dữ liệu

Một chương trình ghi dữ liệu cảm biến (logging) trên điện thoại thông minh đã được nhóm nghiên cứu phát triển để thu thập dữ liệu. Chương trình ghi dữ liệu cảm biến có giao diện khá đơn giản, gồm 1 ô nhập liệu (textbox) cho phép người tham gia thử nghiệm (subject) nhập vào tên hoạt động (ví dụ: jump), sau đó người thử nghiệm gõ vào 1 nút bấm (button) để bắt đầu thực hiện hoạt động và điện thoại sẽ tự động ghi lại tất

cả các dữ liệu cảm biến gia tốc diễn ra trong vòng 2-5 phút tùy theo hoạt động, ví dụ dữ liệu cảm biến của hoạt động nhảy được ghi lại trong 2 phút trong khi dữ liệu đi bộ được ghi lại trong khoảng 5 phút. Người dùng sẽ được thông báo dừng hoạt động khi có 1 tiếng bíp (beep) phát ra từ điện thoại.

17 người với các độ tuổi khác nhau từ 18 đến 55 bao gồm 14 nam và 3 nữ được mời để tham gia thu thập dữ liệu cho nghiên cứu này. Mỗi người được yêu cầu thực hiện 5 hoạt động: *chạy (run)*, *nhảy (jump)*, *đi bộ (walk)*, *ngồi (sit)*, *đứng (stand)* và một hoạt động bất kỳ (*other*) và mỗi hoạt động được yêu cầu thực hiện lặp lại nhiều lần trong khoảng 2-5 phút cho đến khi có tiếng bíp phát ra từ điện thoại. Dữ liệu ghi lại được định dạng như sau: với mỗi mẫu bao gồm 1 nhãn thời gian (timestamp) có cấu trúc *năm-tháng-ngày giờ: phút: giây.mili giây*. Tiếp theo là giá trị cảm biến theo chiều X, chiều Y, và chiều Z. Nhãn thời gian được sử dụng trong bước tiền xử lý dữ liệu trong nghiên cứu này và dùng để đồng bộ với dữ liệu cảm biến khác (trong các nghiên cứu về sau của chúng tôi). Các giá trị cảm biến X, Y, Z được sử dụng để trích chọn đặc trưng như đã mô tả ở phần trên. Hình 5 mô tả trực quan hóa dữ liệu cảm biến gia tốc của một số hoạt động diễn ra trong khoảng 15 giây.

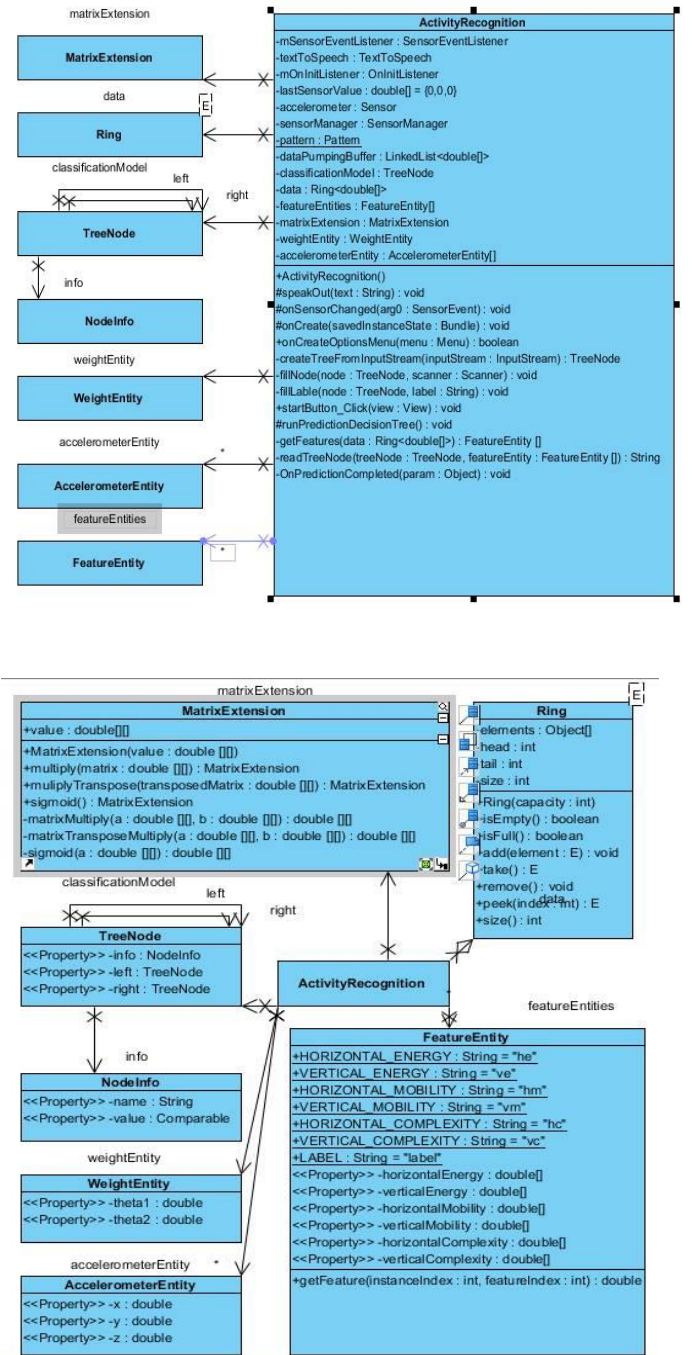
III.2. Các độ đo đánh giá

Trong nghiên cứu này, chúng tôi sử dụng 3 độ đo là: *độ chính xác (precision)*, *độ bao phủ (recall)* và *độ đúng (accuracy)*

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

$$Recall = \frac{TP}{TP + FN} \quad (16)$$

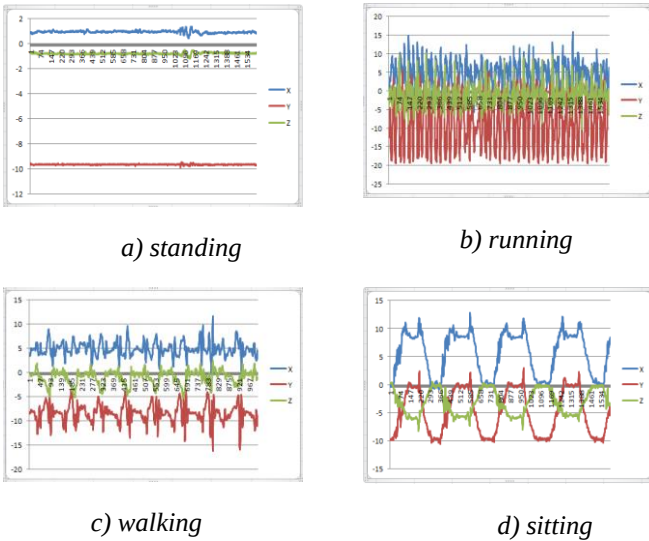
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$



Hình 4: Thiết kế lược đồ lớp (class diagram)

Trong đó, *True Positive (TP)* là tỉ lệ đo số lần hệ thống nhận dạng đúng hoạt động a và số lần thực tế thực hiện hoạt động a; ví dụ hoạt động *chạy* được nhận dạng *đúng* là *chạy*. *True Negative (TN)* là tỉ lệ đo số hệ thống nhận dạng đúng không phải hoạt động a và số lần thực tế không phải a; ví dụ *không phải hoạt*

động chạy được nhận dạng đúng là *không phải chạy*. False Positive (FP) là tỉ lệ đo số lần hệ thống nhận dạng sai hoạt động a và số lần thực tế không phải hoạt động a; ví dụ hệ thống nhận dạng hoạt động là *chạy* nhưng thực tế *không phải là chạy*. False Negative (FN) là tỉ lệ đo số lần hệ thống nhận dạng sai không phải hoạt động a và số lần thực tế là lại hoạt động a; chẳng hạn, thực tế là *chạy* nhưng hệ thống nhận dạng sai là *không chạy*.



Hình 5: Trực quan hóa dữ liệu của một vài hoạt động

III.3. Kết quả và đánh giá

Sau khi trích chọn đặc trưng từ tập dữ liệu thì các véc tơ đặc trưng (feature vectors) được sử dụng để huấn luyện (training) và kiểm thử (testing). Trong nghiên cứu này chúng tôi sử dụng phương pháp đánh giá (evaluation) là *leave-one-subject out*. Với phương pháp này thì chúng tôi sử dụng dữ liệu của 16 người để huấn luyện và dùng dữ liệu của người còn lại để kiểm thử. Quá trình này lặp lại cho đến khi dữ liệu của tất cả 17 người được kiểm thử. Kết quả cuối cùng sẽ là trung bình cộng của tất cả 17 lần kiểm thử. Chú ý rằng dữ liệu huấn luyện không bao gồm dữ liệu của người kiểm thử. Điều này rất quan trọng đối với ứng dụng nhận dạng hoạt động của người bằng các mô hình học máy đã được tiền huấn luyện (pre-trained). Kết quả nhận dạng khi chạy với các thuật toán phân loại khác

nhau được trình bày trong Bảng 1. Kết quả bình quân (của cả 6 hoạt động) được thể hiện dưới dạng *độ đúng* (accuracy) và *độ lệch chuẩn* (standard deviation [26]). Từ bảng 1 ta thấy thuật toán cây quyết định (Decision Tree C4.5) cho kết quả (độ đúng) tốt nhất với trên 80% cho cả 3 trường hợp lựa chọn đặc trưng phân nhánh. Trong đó, đặc trưng phân nhánh information gain cho kết quả cao nhất với độ đúng là 81.06%. Tiếp đến là thuật toán k láng giềng gần nhất (k-NN) cho độ đúng xấp xỉ 81%. Ngược lại, thuật toán Bayes đơn giản (Naïve bayes) cho độ đúng thấp nhất với 71-72%. Có thể do Bayes đơn giản phân loại các đặc trưng dựa trên giả thiết là tất cả các đặc trưng này phải độc lập với nhau. Điều này có thể đôi khi không chắc chắn đúng với bài toán nhận dạng hoạt động của người khi các đặc trưng được trích chọn từ dữ liệu cảm biến gia tốc biến đổi liên tục theo thời gian có thể không hoàn toàn độc lập với nhau.

Thuật toán cây quyết định đã cho kết quả cao nhất và đã được nhóm nghiên cứu chúng tôi lựa chọn để cài đặt (implementation) cho phiên bản nhận dạng hoạt động người thời gian thực trên điện thoại.

Kết quả nhận dạng các hoạt động của thuật toán cây quyết định được trình bày chi tiết trong Bảng 2. Từ Bảng 2 ta thấy các hoạt động *đứng yên* và *ngồi* cho kết quả cao nhất với độ chính xác lớn hơn 93% và độ bao phủ lớn hơn 86%. Khi quan sát dữ liệu thu được từ các hoạt động này sau khi đã trực quan hóa (ví dụ Hình 5) ta thấy các mẫu này phân biệt đáng kể so với các mẫu của những hoạt động khác. Ngược lại, hoạt động chưa gán nhãn (other) cho kết quả thấp nhất với chỉ khoảng trên 53% độ chính xác và 72% độ bao phủ. Do hoạt động chưa gán nhãn được xem như tất cả các hoạt động khác ngoài 5 hoạt động nhảy, đi bộ, ngồi, đứng và chạy nên các mẫu dữ liệu của các hoạt động chưa gán nhãn cũng rất khác nhau và dễ bị phân loại nhầm sang các hoạt động đã gán nhãn.

Độ chính xác và độ bao phủ trung bình là hơn 82% và 83% cho cả 6 loại hoạt động trên. Nếu đem kết quả này so sánh với kết quả của hai nghiên cứu trước đây về nhận dạng hoạt động của người trên bằng điện thoại

thông minh [6,7] là không hợp lý vì [6,7] chưa hướng đến nhận dạng hoạt động trong thời gian thực và trong tập dữ liệu thử nghiệm của họ không bao gồm các hoạt động chưa gán nhãn (other). Điều này dẫn tới các hệ thống [6,7] có thể khó áp dụng cho thực tế vì mọi hoạt động của người phải bị phân loại thành một trong 6 hoạt động đi bộ, chạy tại chỗ, đi lên cầu thang, đi xuống cầu thang, ngồi và đứng. Hơn thế nữa, theo phương pháp đánh giá trong các nghiên cứu [6,7] thì dữ liệu dùng để huấn luyện bao gồm cả dữ liệu kiểm thử. Điều này thiếu khách quan vì trên thực tế là không phải hệ thống nào khi mới cài đặt trên điện thoại thông minh cũng đã có dữ liệu huấn luyện từ người dùng.

Bảng 1: Kết quả nhận dạng (tính bằng độ đúng và độ lệch chuẩn) với các thuật toán phân loại khác nhau.

Thuật toán phân loại	Độ đúng $\pm$ độ lệch chuẩn
SVM (rbf kernel)	80.44% $\pm$ 2.02%
Decision tree (gain ratio)	76.37% $\pm$ 2.05%
Decision tree (information gain)	81.06% $\pm$ 1.59%
Decision tree (gini index)	80.35% $\pm$ 2.53%
K-Nearest Neighbors (KNN)	80.93% $\pm$ 2.91%
Naïve Bayes	71.60% $\pm$ 1.71%
Naïve Bayes (Kernel)	72.47% $\pm$ 1.75%
Rule Induction	78.09% $\pm$ 2.25%
Neural net	79.12% $\pm$ 2.67%

Như phân tích ở trên thuật toán cây quyết định được lựa chọn để phát triển hệ thống. Do đó, các thiết kế chi tiết của hệ thống nhận dạng hoạt động của người trong thời gian thực khi triển khai trên thiết bị di động được mô tả chi tiết trong hình 4.

IV. KẾT LUẬN

Bài báo đã trình bày phương pháp nhận dạng hoạt động của người từ dữ liệu cảm biến gia tốc được tích hợp trong điện thoại thông minh trong thời gian thực. Phương pháp đề xuất đã được đánh giá trên tập dữ liệu gồm 17 người, mỗi người thực hiện 6 hoạt động từ 2-5 phút, bao gồm cả những hoạt động chưa gán nhãn.

Sau khi đánh giá một số thuật toán nhận dạng thì thuật toán cây quyết định (Decision Tree C4.5) cho kết quả tốt nhất với 82% độ chính xác và 83% độ bao phủ bằng phương pháp đánh giá leave-one-subject out. Kết quả này cho thấy tính khả thi của nhận dạng hoạt động của người trong thời gian thực bằng thiết bị di động. Nghiên cứu cải tiến phương pháp để có thể nhận dạng được thêm một số hoạt động hàng ngày như đạp xe, đi cầu thang, nằm v.v.. nhằm mục đích đưa hệ thống được sử dụng như là nền tảng cho một số ứng dụng thực tế như tự động ước lượng năng lượng tiêu thụ hàng ngày (energy expenditure estimation) hay trợ giúp theo dõi sức khỏe của người (health monitoring) sẽ là các nghiên cứu tiếp theo của nhóm tác giả.

Bảng 2: Kết quả nhận dạng của thuật toán cây quyết định.

Hoạt động	Độ chính xác (Precision)	Độ bao phủ (recall)
Hoạt động chưa gán nhãn	53.79%	72.61%
Nhảy (jump)	67.23%	82.35%
Đi bộ (walk)	82.35%	89.94%
Ngồi (sit)	93.42%	86.15%
Đứng (stand)	95.23%	89.46%
Chạy (run)	79.98%	82.69%
Trung bình	82.65%	83.36%

V. LỜI CẢM ƠN

Chúng tôi xin trân thành cảm ơn nhóm sinh viên của Học viện Công nghệ Bưu chính Viễn thông và một số người bạn đã rất nhiệt tình, trách nhiệm và kiên nhẫn tham gia thực hiện các hoạt động cho thử nghiệm của chúng tôi.

TÀI LIỆU THAM KHẢO

- [1] V. OSMANI, S. BALASUBRAMANIAM, D. BOTVICH, "Human Activity Recognition in Pervasive Health-care: Supporting Efficient Remote Collaboration", Journal of Network and Computer Applications (Elsevier), vol. 31, no. 4, pp. 628-655, 2008.

- [2] C.PHAM, D. JACKSON, J. SCHONING, T. BAR., T.PLOETZ, P.OLIVIER “*FoodBoard: Surface Contact Imaging for Food Recognition*”, In Proc. of the 15th ACM International Conference on Pervasive and Ubiquitous Computing (UbiComp), pp. 749-752, 2013.
- [3] F. ALBINALI, S. S. INTILLE, W. HASKELL, M. ROSENBERGER “*Using Wearable Activity Type Detection to Improve Physical Activity Energy Expenditure Estimation*”, In Proc. of the 12th ACM International Conference on Ubiquitous Computing (UbiComp), pp. 311-320, 2010.
- [4] S. S. INTILLE, J. NAWYN, B. LOGAN, G. D. ABOARD “*Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research*”, In Proc. of the ACM International Conference on Computer and Human Factors (CHI) Extended Abstracts, pp. 4763-4766, 2009.
- [5] Digital Trends: <http://www.digitaltrends.com/mobile/mobile-phone-world-population-2014/#!Cs565>
- [6] K. R. JENIFER, G. M. WEISS, S. MOORE “*Activity Recognition Using Cell Phone Accelerometers*”, SIGKDD Explorations, vol. 12, no. 2, pp. 74-82, 2010.
- [7] D. ANGUITA, A. GHIO, L. ONETO, X. PARRA, J. R. ORTIZ, “*Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine*”, In Proc. of the 4th International Workshop on Ambient Assisted Living, pp. 216-223, 2012.
- [8] L. Bao, S. S. INTILLE, “*Activity Recognition from User-Annotated Acceleration Data*”, In Proc. of the 2nd International Conference on Pervasive Computing (Pervasive’2004), pp. 1-17, 2004.
- [9] N. RAVI, N. DANDEKAR, P. MYSORE, M. L. LITTMAN, “*Activity Recognition from Accelerometer Data*”, In Proc. of the 17th Conference on Innovative Applications of Artificial Intelligence (IAAI), pp. 1541-1546, 2005.
- [10] T. HUYNH, U. BLANKE, B. SCHIELE, “*Scalable Recognition of Daily Activities with Wearable Sensors*”, In Proc. of the 3rd International Conference on Location-and Context-Awareness (LoCA), pp. 50-67, 2005.
- [11] J. WU, A OSUNTOGUN, T. CHOUDHURY, M. PHILIPOSE, J. M. REHG, “*A Scalable Approach to Activity Recognition based on Object Use*”, In Proc. of the 11th International Conference on Computer vision (ICCV), pp. 1-8, 2007.
- [12] T. V. DUONG, H. H. BUI, P. Q.DINH, S. VENKATESH, “*Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model*”, In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 838-845, 2005
- [13] C. PHAM, P. OLIVIER, “*Slice&Dice: Recognizing Food Preparation Activities using Embedded Accelerometers*”, In Proc. of European Conference on Ambient Intelligence (AmI), pp. 34—43, 2009.
- [14] M. BUETTNER, R. PRASAD, Ma. PHILIPOSE, D. WETHERER, “*Recognizing Daily Activities with RFID-based Sensors*”, In Proc. of the 11th International Conference on Ubiquitous Computing (UbiComp), pp. 51-60, 2009.
- [15] E. M. TAPIA, S. S. INTILLE, K. LARSON, “*Activity Recognition in the Home Using Simple and Ubiquitous Sensors*”, In Proc. of the 2nd International Conference on Pervasive Computing (Pervasive), pp. 158-175, 2004.
- [16] C.J. HOOPER, A. PRESTON, M. BALAAM, P. SEEDHOUSE, D. JACKSON, C. PHAM, C. LADHA, K. LADHA, T. PLOETZ, P. OLIVIER, “*The French Kitchen: Task-Based Learning in an Instrumented Kitchen*”, In Proc. of the 14th ACM International Conference on Ubiquitous Computing (UbiComp) pp. 193-202, 2012.
- [17] T. V. KASTEREN, A. K. NOULAS, G. ENGLEBIENNE, B. J. A. KROSE, “*Accurate Activity Recognition in a Home Setting*”, In Proc. of the 10th international conference on Ubiquitous Computing (UbiComp), pp. 1-9, 2008.
- [18] C. PHAM, D. N. NGUYEN, P. M. TU, “*A Wearable Sensor based Approach to Real-Time Fall Detection and Fine-Grained Activity Recognition*”, *Journal of Mobile Multimedia*, vol. 9, no. 1&2, pp. 15-26, 2013.
- [19] T. LYCHE, L.L. SCHUMAKER, "On the Convergence of Cubic Interpolating Splines" A. Meir

- (ed.) A. Sharma (ed.) Spline Functions and Approximation Theory, Birkhäuser (1973) pp. 169–189
- [20] T. V. KASTEREN, G. ENGLEBIENNE, B. J. A. KROSE, “Hierarchical Activity Recognition Using Automatically Clustered Actions”, In Proc. of European Conference on Ambient Intelligence (AmI), pp. 82-91, 2011.
- [21] I. WITTEN, E. FRANK, M. A. HALL, “Data Mining: Practical Machine Learning Tools and Techniques” Morgan Kaufmann (2011).
- [22] B. A. SSHENOI, “Introduction to Digital Signal Processing and Filter Design”, John Wiley and Sons, 2006.
- [23] S. I. AHAMED, M. RAHMAN, R. O. SMITH, M. KHAN, “A Feature Extraction Method for Realtime Human Activity Recognition on Cell Phones,” in Proc. of International Symposium on Quality of Life Technologies, 2011.
- [24] SVMlib: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [25] RapidMiner: <http://rapidminer.com/>
- [26] G. SAEED, “Fundamentals of Probability”, Prentice Hall: New Jersey, 2000.
- [27] C. PHAM, C. HOOPER, S. LINDSAY, D. JACKSON, J. SHEARER, J. WAGNER, C. LADHA, K. LADHA, T. PLOETZ, P. OLIVIER, “The Ambient Kitchen: A Pervasive Sensing Environment for Situated Services,” in Proc. of ACM Conference on Designing Interactive Systems (DIS), 2012.
- [23] S. I. AHAMED, M. RAHMAN, R. O. SMITH, M. KHAN, "A Feature Extraction Method for Realtime Human Activity Recognition on Cell Phones," in Proc. of

Ngày nhận bài: 08/01/2015

SƠ LƯỢC VỀ TÁC GIẢ

PHẠM VĂN CƯỜNG



Sinh ngày 2/9/1976 tại Hưng Yên.

Tốt nghiệp ĐH Quốc gia Hà nội năm 1998 ngành CNTT; Thạc sỹ ngành Khoa học Máy tính tại ĐH bang New Mexico, Hoa Kỳ năm 2005; Tiến sỹ

ngành Khoa học Máy tính tại ĐH Newcastle, vương quốc Anh năm 2012.

Hiện đang là giảng viên khoa CNTT 1, Học viện Công nghệ Bưu chính Viễn thông.

Hướng nghiên cứu: nhận dạng hoạt động của người, các phương pháp học máy cho ứng dụng chăm sóc sức khỏe, tính toán di động, tương tác người máy và điện toán tòa khấp.

Mobile: 0944643166;

Email: pcuongcntt@gmail.com

NGUYỄN THẮNG NGỌC



Sinh năm 1991 tại Hải Dương.

Tốt nghiệp ĐH ngành CNTT (chương trình Chất lượng cao), Học viện Công nghệ BCVT.

Hiện đang là kỹ sư nghiên cứu và phát triển của Samsung R&D; Hướng nghiên cứu: nhận dạng hoạt động của người, tính toán

di động.