

Phát hiện hoạt động bất thường của người bằng mạng học sâu tích chập kết hợp mạng bộ nhớ dài ngắn

Nguyễn Tuấn Linh, Nguyễn Văn Thủy, Phạm Văn Cường

Học viện Công nghệ Bưu chính Viễn thông

Tác giả liên hệ: Phạm Văn Cường, cuongpv@ptit.edu.vn

Ngày nhận bài: 17/04/2020, ngày sửa chữa: 24/05/2020

Định danh DOI: 10.32913/mic-ict-research-vn.vvyyy.nx.xyz

Tóm tắt: Bài báo này đề xuất một mô hình học sâu tích chập kết hợp với mạng bộ nhớ dài ngắn (CNN-LSTM) cho bài toán phát hiện các vận động bất thường của người sử dụng cảm biến đeo trên người. Nhờ tận dụng các đặc tính không-thời gian, kiến trúc đề xuất CNN-LSTM đã được thiết kế để tự động học và biểu diễn các đặc trưng hiệu quả trên dữ liệu cảm biến không thuần nhất. Kết quả thử nghiệm trên 4 tập dữ liệu được công bố cho thấy mô hình đề xuất đã cho kết quả cải tiến tốt hơn từ 2% đến 7% F1-score so với các mô hình học máy dựa trên trích xuất đặc trưng thủ công SVM, mô hình học sâu tích chập (CNN) và mô hình mạng bộ nhớ dài ngắn (LSTM).

Từ khóa: cảm biến đeo, cảm biến gia tốc, mạng tích chập, mạng bộ nhớ dài ngắn.

Title: Human Abnormal Activity Detection with Deep Convolutional Long-Short Term Memory Networks

Abstract: This work proposes Deep Convolutional Neural Long-Short Term Networks (CNN-LSTM) to address the problem of human abnormal activity detection using wearable sensors. Our proposed architecture effectively utilizes spatial-temporal characteristics of sensing data for automatically learning and representing features from heterogeneous sensing data. Experimental results have demonstrated that the proposed method has improved from 2% to 7% F1-score better than several shallow and deep models including SVM, CNN and LSTM on 4 published datasets.

Keywords: wearable Sensor, accelerometer, CNN, LSTM

I. ĐẶT VẤN ĐỀ

Phát hiện vận động bất thường của con người là lĩnh vực nhận được nhiều sự quan tâm của cộng đồng nghiên cứu vì đây là lĩnh vực có nhiều ứng dụng trong thực tế như hỗ trợ cho người mất trí nhớ [1], theo dõi người bệnh đột quỵ [2], theo dõi chăm sóc người vận động bất thường [3]v.v... Vận động bất thường được xem là các hoạt động mà con người *không có chủ ý* và thường gây ra những hậu quả xấu đối với chủ thể. Một người bị *ngã* trong khi đang làm việc nhà hoặc một cú *trượt chân* do đường trơn trượt là các ví dụ về vận động bất thường. Những vận động bất thường này khi xảy ra sẽ gây nguy hiểm cho con người (đặc biệt là người cao tuổi). Trong những trường hợp như

vậy, nếu có một hệ thống phát hiện và đưa ra những cảnh báo hoặc tự động kết nối đến người trợ giúp sẽ hạn chế được các rủi ro cũng như giảm thiểu các hậu quả do vận động bất thường đến người.

Hai phương pháp tiếp cận phổ biến để giải quyết bài toán vận động bất thường là: sử dụng cảm biến được tích hợp vào môi trường [6] và cảm biến đeo trên người [4, 5, 22]. Trong cách tiếp cận thứ nhất thì các cảm biến hình ảnh như camera số được thiết đặt để quan sát các hoạt động hàng ngày của người [7] hoặc cảm biến định danh (RFID) được gắn vào trong các vật dụng trong nhà để phát hiện người sử dụng những vật dụng nào, từ đó suy diễn ra các hoạt động hàng ngày và vận động bất thường của người mất

trí nhớ tạm thời [1, 23]. Hạn chế của phương pháp sử dụng camera là có thể gây ra sự xâm lấn không gian riêng tư và việc phát hiện vận động bất thường thường bị giới hạn trong một phạm vi là vùng quan sát được của camera hoặc các cảm biến được tích hợp vào môi trường. Ngược lại, cách tiếp cận thứ hai bằng cảm biến đeo trên người thường không bị giới hạn bởi môi trường, đồng thời cũng giảm thiểu được việc xâm lấn riêng tư. Hơn nữa, với sự phát triển nhanh chóng của các thiết bị điện tử kết nối Internet vạn vật (the Internet of Things) thì các thiết bị đeo ngày càng có sẵn trên thị trường với giá thành rẻ. Chính vì vậy trong nghiên cứu này chúng tôi tiếp cận bài toán phát hiện vận động bất thường theo cách tiếp cận dựa trên cảm biến đeo.

Thời gian gần đây, mặc dù lĩnh vực nghiên cứu này đang đạt được nhiều thành công, tuy nhiên vẫn còn nhiều thách thức cần phải giải quyết để có thể đưa được các hệ thống trên vào ứng dụng thực tế như: làm thế nào một hệ thống phát hiện được các vận động bất thường trong các ngữ cảnh thực tế khác nhau với độ chính xác cao để có thể sử dụng cho các ứng dụng cảnh báo. Trong khi đó, dữ liệu về vận động bất thường thường rất đa dạng, phức tạp và ít có sẵn do các vận động bất thường vô tình xảy ra trong khi thực hiện các hoạt động hàng ngày (bình thường). Điều này dẫn tới khó khăn khi huấn luyện mô hình học máy để đạt được độ chính xác đủ tốt cho việc phát hiện các vận động bất thường. Hơn thế nữa, dữ liệu về vận động bất thường thường mất cân bằng (imbalanced) do tần suất của từng loại vận động bất thường khác nhau một cách tự nhiên.

Trong nghiên cứu này, chúng tôi đề xuất một mô hình mạng học sâu tích chập kết hợp với mạng bộ nhớ dài ngắn có khả năng học từ dữ liệu cảm biến không thuần nhất. Cụ thể hơn, có hai đóng góp chính trong nghiên cứu này:

- Thứ nhất, chúng tôi đề xuất một phương pháp học bằng việc kết mô hình mạng học sâu tích chập (CNN) và mạng bộ nhớ dài ngắn để giải quyết bài toán phát hiện các vận động bất thường từ dữ liệu cảm biến không thuần nhất bao gồm cảm biến gia tốc, cảm biến con quay hồi chuyển và cảm biến từ tính. Trong đó, mô hình CNN đóng vai trò như bộ encoder được huấn luyện để học và biểu diễn các đặc trưng từ nhờ khai thác đặc tính không gian của dữ liệu cảm biến; còn mạng LSTM dùng đóng vai trò bộ

suy diễn (decoder) tận dụng các đặc tính về thời gian của dữ liệu cảm biến.

- Chúng tôi đánh giá phương pháp đề xuất trên một số bộ dữ liệu đã được công bố rộng rãi. Kết quả cho thấy phương pháp đề xuất của chúng tôi hiệu quả hơn so với một số phương pháp truyền thống và phương pháp học sâu khác do chưa tận dụng được hai đặc tính không gian và thời gian của dữ liệu cảm biến.

Nghiên cứu của chúng tôi khác biệt với các nghiên cứu khác ở hai điểm chính. *Thứ nhất* là phương pháp đề xuất đã tận dụng kết hợp được các đặc tính về không-thời gian (Spatial-Temporal Features) từ dữ liệu cảm biến để khai thác việc học và biểu diễn đặc trưng hiệu quả. *Thứ hai* là mô hình đề xuất của chúng tôi chấp nhận đầu vào là dữ liệu cảm biến không thuần nhất đến từ các loại cảm biến khác nhau kết hợp lại để phát hiện các hoạt động bất thường.

II. CÁC NGHIÊN CỨU CÓ LIÊN QUAN

Phát hiện hoạt động bất thường đã và đang thu hút được sự quan tâm của cộng đồng nghiên cứu [11]. Trước đây, phương pháp tiếp cận phát hiện hoạt động bất thường chủ yếu dựa trên các mô hình học máy trong đó học có giám sát [12] được sử dụng phổ biến. Các dữ liệu (mẫu) được gán nhãn để các mô hình có thể học và mô hình được huấn luyện sẽ được đánh giá trên các dữ liệu mới. Do đó, trong trường hợp có các lớp hoạt động bình thường và bất thường, mô hình sẽ học các đặc tính của các điểm dữ liệu này và phân loại chúng là hoạt động bình thường hay bất thường. Bất kỳ điểm dữ liệu nào không phù hợp với lớp hoạt động bình thường sẽ được mô hình phân loại là bất thường [9].

Aran và đồng sự [4] đã đề xuất một phương pháp có thể tự động hoá quan sát và mô hình hoá hoạt động hằng ngày của người cao tuổi, qua đó giúp phát hiện hoạt động bất thường từ dữ liệu thu được bằng cảm biến. Trong phương pháp của họ, sự bất thường liên quan đến các vấn đề về tín hiệu sức khỏe. Với mục đích này, họ đã tạo ra một mô hình không gian xác suất theo thời gian để có thể tóm lược toàn bộ các hoạt động hằng ngày. Họ định nghĩa sự bất thường là những thay đổi đáng kể từ những hoạt động đã được học và được phát hiện, hiệu suất phát hiện được đánh giá bằng phương pháp entropy chéo. Trong nghiên cứu của họ, khi một hoạt động bất thường được phát hiện,

ngay lập tức sẽ có thông báo được gửi đến người chăm sóc.

Ordonez và đồng sự [14] đã thực hiện một phương pháp phát hiện bất thường dựa trên thống kê Bayes, từ đó giúp phát hiện hoạt động bất thường của con người. Phương pháp của họ có khả năng tự động hỗ trợ người già, người khuyết tật sống một mình bằng cách học và dự đoán các hoạt động tiêu chuẩn qua đó cải thiện hiệu suất của hệ thống chăm sóc sức khỏe. Thống kê Bayes được sử dụng để phân tích dữ liệu thu thập được, dự đoán hoạt động dựa trên ba đặc trưng xác suất, bao gồm: xác suất kích hoạt cảm biến (Sensor Activation Likelihood), chuỗi cảm biến (Sensor Sequence Likelihood) và sự kiện cảm biến (Sensor Event Duration Likelihood).

Yahaya và đồng sự [11] đề xuất thuật toán phát hiện đặc trưng mới có tên máy vectơ hỗ trợ một lớp (One-class SVM) giúp phát hiện hoạt động bất thường từ các hoạt động bình thường diễn ra hằng ngày. Sự bất thường trong kiểu nằm ngủ có thể được coi là dấu hiệu của Sự suy giảm nhận thức nhẹ (MCI) ở người cao tuổi hoặc các vấn đề liên quan đến sức khỏe khác. Palaniappan và đồng sự [15] lại đặc biệt quan tâm đến các hoạt động bất thường ở người bằng cách loại trừ tất cả các hoạt động được coi là bình thường. Các tác giả định nghĩa hoạt động bất thường là các hoạt động bất ngờ xảy ra theo một cách ngẫu nhiên. Phương pháp SVM đa lớp được họ sử dụng làm trình phân loại để xác định các hoạt động dưới dạng bảng chuyển trạng thái. Điều này sẽ giúp trình phân loại tránh được các trạng thái không thể đưa ra được (không thể truy cập được) từ trạng thái hiện tại.

Hùng và đồng sự [16] đã đề xuất một phương pháp mới kết hợp SVM và HMM sử dụng một hệ thống các cảm biến thiết lập trong nhà (homecare sensory system). Mạng cảm biến RFID được sử dụng để thu thập các hoạt động hằng ngày của người cao tuổi. Mô hình Markov ẩn (HMM) được sử dụng để học từ dữ liệu được thu thập, trong khi SVM được sử dụng để ước tính liệu hoạt động đó của người cao tuổi có là hoạt động bất thường hay không. Bouchachia và đồng sự [17] lại đề xuất một mô hình RNN để giải quyết các vấn đề về nhận biết hoạt động và phát hiện hoạt động bất thường cho người cao tuổi bị chứng mất trí nhớ.

Mặc dù có một số nghiên cứu phát hiện hoạt động bất thường, tuy nhiên từ các nghiên cứu ở trên vẫn

tồn tại một số điểm hạn chế như: Độ chính xác dự đoán hoạt động bất thường của các phương pháp học nông phụ thuộc khá nhiều kinh nghiệm trích chọn các đặc trưng theo kinh nghiệm chuyên gia. Trong khi đó, một số phương pháp học sâu lại chưa tận dụng đầy đủ đặc tính không-thời gian của dữ liệu cảm biến (đặc biệt là dữ liệu cảm biến không thuần nhất) mà nghiên cứu này tập trung giải quyết.

III. MÔ HÌNH MẠNG HỌC SÂU TÍCH CHẬP KẾT HỢP MẠNG BỘ NHỚ DÀI NGẮN (CNN-LSTM) CHO PHÁT HIỆN VẬN ĐỘNG BẤT THƯỜNG

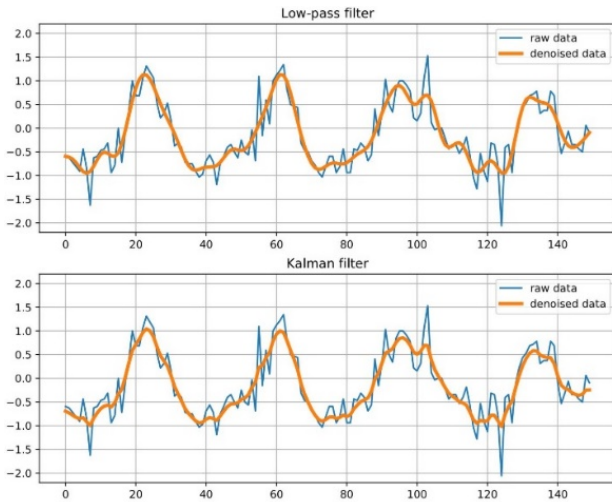
Kiến trúc mạng học sâu tích chập kết hợp mạng bộ nhớ dài ngắn (CNN-LSTM) đề xuất được trình bày trong Hình 2. Dữ liệu cảm biến được tiền xử lý trước khi đưa vào mạng. Kiến trúc mạng bao gồm 3 thành phần chính: tích chập, bộ nhớ dài ngắn và lớp đầu ra. Chi tiết được mô tả dưới đây.

Giả sử $S = S_k, k \in 1, \dots, 3$ tương ứng với 3 loại cảm biến: gia tốc, con quay hồi chuyển, và từ trường. Với cảm biến S_k , nó tạo ra một phép đo theo thời gian, các phép đo có thể được biểu thị bằng đối với ma trận V cho các giá trị đo với $n(k)$ là chiều của vectơ u cho các dấu thời gian (time stamps), $d(k)$ là kích thước cho mỗi phép đo (ví dụ: các phép đo dọc theo trục x, y, z đối với cảm biến), $n(k)$ là số phép đo. Chúng tôi chia các phép đo đầu vào V và u theo thời gian (các cột cho V) để tạo ra một chuỗi các chu kỳ thời gian không chồng lấn với chiều rộng $\tau, W = (V_t^{(k)}, u_t^{(k)})$ trong đó $|W| = T; \tau$ có thể khác nhau đối với các chu kỳ thời gian khác nhau. Để đơn giản chúng tôi giả sử chu kỳ thời gian là cố định. Sau đó, chúng tôi áp dụng biến đổi Fourier cho từng phần tử trong W bởi miền tần số chứa các tần số mẫu cục bộ tốt hơn, độc lập với cách tổ chức dữ liệu chuỗi thời gian trong miền thời gian. Chúng tôi tiến hành sắp xếp các đầu ra thành một $d(k) \times 2f \times T$ tensor $X^{(k)}$ trong đó f là thứ nguyên của miền tần số chứa các cặp pha và tần số cường độ f . Tập hợp các thang đo kết quả cho mỗi cảm biến $X = X^{(k)}$ sẽ là đầu vào của mô hình CNN-LSTM.

1. Lọc và tiền xử lý tín hiệu

Loại bỏ nhiễu: Tín hiệu cảm biến thường chứa nhiều tín hiệu nhiễu, điều này là do môi trường xung quanh

có nhiều vật thể làm bằng kim loại hoặc do bản thân tự cảm biến sinh ra nhiễu. Vì vậy, các tín hiệu thu được cần phải thực hiện lọc bỏ nhiễu. Trong nghiên cứu này, chúng tôi sử dụng bộ lọc thông thấp và bộ lọc Kalman (như minh họa trong Hình 1).



Hình 1. Bộ lọc thông thấp (Low-pass filter) và bộ lọc Kalman.

Đây là những bộ lọc đơn giản, không đòi hỏi quá nhiều tài nguyên tính toán nhưng lại mang hiệu quả cao. Để tránh việc trễ, mỗi chuỗi dữ liệu được đưa qua bộ lọc hai lần, một lần theo chiều thuận và một lần ngược lại.

Tiếp đến chúng tôi căn chỉnh, phân chia các phép đo cảm biến và áp dụng biến đổi Fourier cho mỗi khối cảm biến. Đối với mỗi cảm biến, chúng tôi xếp các đầu ra miền tần số này thành $d(k) \times 2f \times T$ tensor $X^{(k)}$, trong đó $d(k)$ là kích thước đo chiều cảm biến, f là kích thước miền tần số và T là số chu kỳ thời gian.

2. Thành phần mạng tích chập (CNN)

Các lớp tích chập có thể được chia làm hai phần: một mạng con tích chập riêng cho mỗi tensor cảm biến đầu vào $X^{(k)}$ và một mạng con tích chập gộp duy nhất cho đầu ra của K các mạng con tích chập riêng lẻ.

Do cấu trúc của mạng con tích chập riêng cho các cảm biến khác nhau là như nhau nên chúng tôi tập trung vào một mạng con tích chập riêng lẻ với đầu vào $X^{(k)}$. Cần lưu ý rằng $X^{(k)}$ là một $d(k) \times 2f \times T$ tensor, trong đó $d(k)$ cho biết kích thước chiều cảm

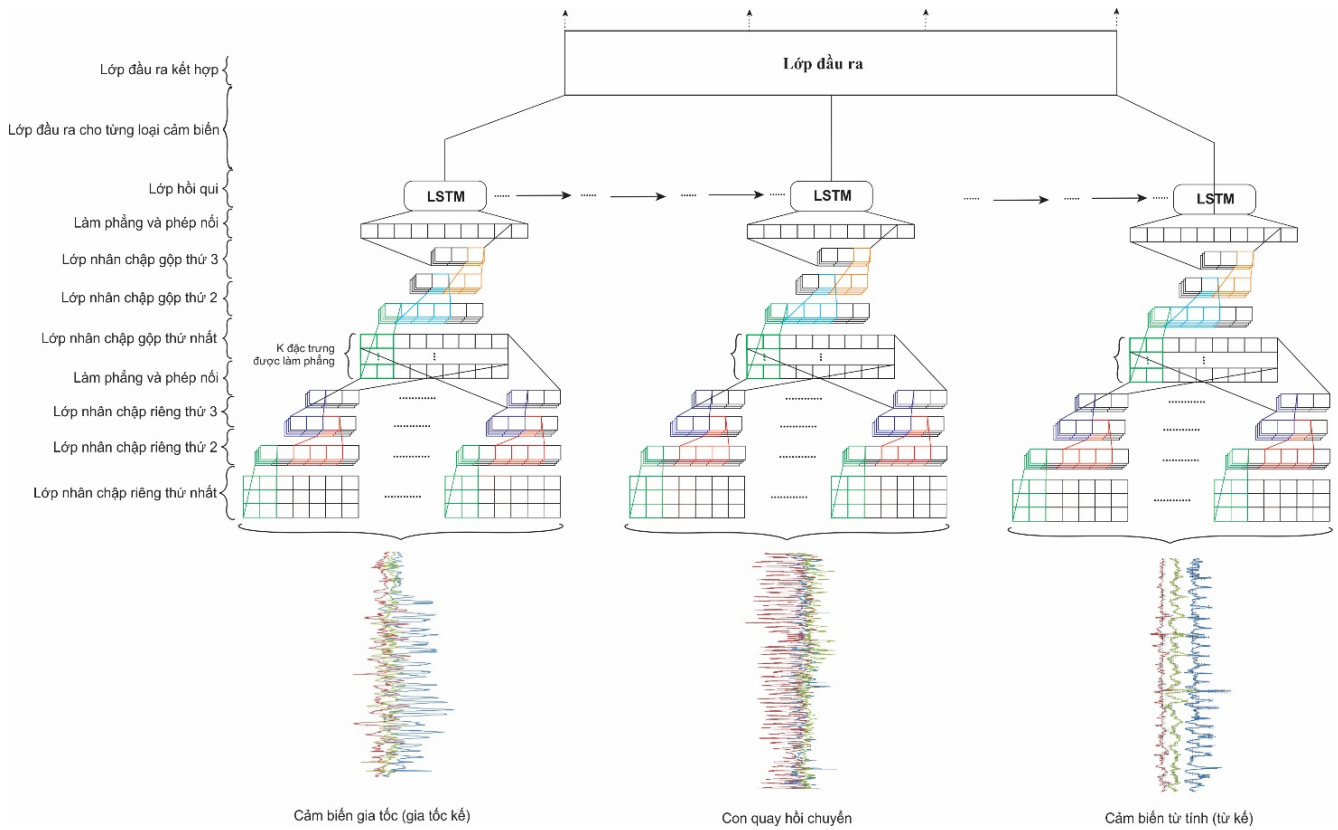
biến, f là kích thước của miền tần số và T là số lượng chu kỳ thời gian. Đối với mỗi chu kỳ thời gian t , ma trận $X_{..t}^{(k)}$ sẽ được đưa vào kiến trúc CNN với ba lớp tích chập. Đặc trưng miền tần số và kích thước số chiều được nhúng trong $X_{..t}^{(k)}$. Miền tần số thường chứa rất nhiều mẫu cục bộ ở một số tần số lân cận. Sự tương tác giữa các phép đo cảm biến thường bao gồm tất cả số chiều. Vì vậy, trước tiên, chúng tôi áp dụng các bộ lọc $2d$ có dạng $(d^{(k)}, cov1)$ cho $X_{..t}^{(k)}$ để học được sự tương tác giữa kích thước số chiều cảm biến và các mẫu cục bộ trong miền tần số với đầu ra $X_{..t}^{(k,1)}$. Tiếp theo, chúng tôi áp dụng các bộ lọc $1d$ với dạng $(1, cov2)$ và $(1, cov3)$ theo thứ bậc để tìm hiểu các mối quan hệ cấp cao hơn của $X_{..t}^{(k,2)}$ và $X_{..t}^{(k,3)}$.

Sau đó, chúng tôi tiến hành làm phẳng ma trận $X_{..t}^{(k,3)}$ thành vectơ $x_{..t}^{(k,3)}$ và ghép tất cả K vectơ $x_{..t}^{(k,3)}$ thành một K dòng ma trận $X_{..t}^{(3)}$ (là đầu vào của mạng con tích chập hợp nhất). Kiến trúc của mạng con tích chập hợp nhất tương tự như mạng con tích chập riêng lẻ. Bộ lọc $2d$ được chúng tôi sử dụng với $(K, cov4)$ để học các tương tác giữa các cảm biến K với đầu ra $X_{..t}^{(4)}$, sau đó bộ lọc $1d$ với $(1, cov5)$ và $(1, cov6)$ được áp dụng ở mức độ nâng cao hơn trên $X_{..t}^{(5)}$, $X_{..t}^{(6)}$.

Đối với mỗi lớp tích chập, CNN-LSTM học với 64 bộ lọc và sử dụng ReLU làm hàm kích hoạt. Ngoài ra, việc chuẩn hoá theo mẻ (batch) được áp dụng để mỗi lớp giảm sự thay đổi đồng biến nội bộ. Chúng tôi tiến hành làm phẳng đầu ra cuối cùng $X_{..t}^{(6)}$ thành vectơ $x_{..t}^{(6)}$. Ghép nối và chiều rộng chu kỳ thời gian $[\tau]$ thành $x_t^{(c)}$ làm đầu vào của các lớp LSTM.

3. Thành phần mạng bộ nhớ dài ngắn (LSTM)

Mạng thần kinh hồi qui (Recurrent Neural Networks-RNN) là những kiến trúc mạnh mẽ có thể giúp tính gần đúng và học các đặc trưng có ý nghĩa trong các chuỗi. Một biến thể của RNN là LSTM có thể lưu trữ được sự phụ thuộc dài hạn giữa các trạng thái (Long-term Dependencies). Trong mô hình đề xuất chúng tôi sử dụng cấu trúc tế bào (cell) xếp chồng lên nhau theo chiều chứa luồng thời gian từ đầu đến cuối (Start to End) của chuỗi dữ liệu thời gian (Time Series). Cấu trúc xếp chồng có thể chạy tăng dần khi có một chu kỳ thời gian mới, giúp xử lý luồng dữ liệu nhanh hơn. Đồng thời chúng tôi áp dụng dropout cho các kết nối giữa các lớp để chuẩn hoá và áp dụng chuẩn hoá theo bó hồi qui (Recurrent Batch Normalization) để giảm sự thay đổi đồng biến nội bộ



Hình 2. Kiến trúc mạng học sâu tích chập kết hợp mạng bộ nhớ dài ngắn (CNN-LSTM)

giữa các bước thời gian (time steps). Đầu vào $x_t^{(c)}$ với $t = 1, \dots, T$ từ những lớp chập trước đó được đưa vào LSTM xếp chồng và tạo đầu ra $x_t^{(r)}$ với $t = 1, \dots, T$ làm đầu vào của lớp đầu ra cuối cùng.

4. Lớp đầu ra

Đầu ra của lớp hồi qui là một chuỗi các vectơ $x_t^{(r)}$ với $t = 1, \dots, T$. Đối với tác vụ định hướng hồi quy (regression-oriented), giá trị của mỗi phần tử trong vectơ $x_t^{(r)}$ nằm trong ± 1 , $x_t^{(r)}$ mã hoá các đại lượng vật lý tại cuối chu kỳ thời gian t . Trong lớp đầu ra, chúng tôi muốn học một từ điển W_{out} (dictionary) W_{out} với một b_{out} (bias) để giải mã $x_t^{(r)}$ thành \hat{y}_t sao cho $\hat{y}_t = W_{out} \cdot x_t^{(r)} + b_{out}$. Do đó, lớp đầu ra là một lớp được kết nối đầy đủ trên đỉnh mỗi chu kỳ với chia sẻ tham số W_{out} và b_{out} .

Đối với tác vụ phân loại, $x_t^{(r)}$ là vectơ đặc trưng tại khoảng thời gian t . Trước tiên, lớp đầu ra cần kết hợp $x_t^{(r)}$ thành một vectơ đặc trưng cố định để xử lý thêm. Đặc trưng trung bình theo thời gian là một lựa chọn. Các phương pháp nâng cao hơn có thể được áp dụng

để tạo ra đặc trưng cuối cùng, ví dụ như mô hình chú ý (attention model) đã minh họa một cách có hiệu quả những tác vụ học quan trọng gần đây. Mô hình chú ý có thể được xem như là việc tính trung bình của các đặc trưng theo thời gian nhưng các trọng số được học bởi các mạng LSTM thông qua ngữ cảnh. Trong nghiên cứu này, chúng tôi vẫn sử dụng các đặc trưng trung bình theo thời gian để tạo ra các đặc trưng cuối cùng $x^r = (\sum_{t=1}^T x_t^{(r)})/T$. Sau đó, chúng tôi đưa x^r và một lớp softmax để tạo ra các loại xác suất dự đoán

IV. THỬ NGHIỆM

1. Tập dữ liệu

Chúng tôi sử dụng 4 tập dữ liệu, bao gồm UTD [14], MobiFall [15], PTITAct [9] và CMDFALL [8]. Chi tiết về mỗi tập dữ liệu như sau:

- UTD [14]: đây là tập dữ liệu được thu thập từ 12 người đeo 2 loại cảm biến là cảm biến gia tốc và con quay hồi chuyển với tần số lấy mẫu là 200Hz. Tập dữ liệu bao gồm 6 hoạt động bình thường và 1

hoạt động bất thường (ngã). Để huấn luyện mô hình CNN-LSTM với bộ dữ liệu này chúng tôi đóng băng (frozen) thành phần dành cho cảm biến từ tính và giảm tần số mẫu (downsampling) xuống còn 100 Hz;

- *MobiFall* [15]: là tập dữ liệu được thu thập từ 15 người để điện thoại thông minh trong túi quần. Dữ liệu cảm biến bao gồm cảm biến gia tốc và con quay hồi chuyển được thu thập với tần số lấy mẫu là 90Hz. Tập dữ liệu bao gồm 9 hoạt động bình thường và 4 hoạt động bất thường là các tư thế ngã khác nhau. Để huấn luyện mô hình CNN-LSTM với bộ dữ liệu này chúng tôi đóng băng (frozen) thành phần dành cho cảm biến từ tính và tái tạo tần số lấy mẫu (upsampling) lên 100 Hz bằng phương pháp GAN cho dữ liệu chuỗi thời gian [18];

- *PTITAct* [9]: là tập dữ liệu được thu thập từ 26 người gắn thiết bị internet vạn vật kết nối (IoT) ở thắt lưng. Thiết bị được tích hợp cảm biến gia tốc, con quay hồi chuyển, và từ kế. Dữ liệu cảm biến được thu thập với tần số lấy mẫu là 50Hz. Tập dữ liệu bao gồm 8 loại vận động bất thường (ngã ở các tư thế khác nhau) và 8 hoạt động bình thường. Trước khi huấn luyện mô hình CNN-LSTM, dữ liệu được upsampling mẫu dữ liệu lên 100 Hz bằng phương pháp GAN cho dữ liệu chuỗi thời gian [18];

- *CMDFall* [8]: là tập dữ liệu khá lớn được thu thập từ 50 người đeo 2 cảm biến tại vị trí cổ tay và thắt lưng. Tập dữ liệu gồm 9 nhân hoạt động bình thường (như đi lại, nằm lên giường, ngồi xuống ghế v.v..) và 11 vận động bất thường (như ngã ngửa, ngã bên trái, đi loạng choạng, trượt chân ...) khác nhau. Do tần số lấy mẫu của tập dữ liệu là 50Hz nên khi thực nghiệm trên tập này, tập dữ liệu được upsampling mẫu dữ liệu lên 100 Hz bằng phương pháp GAN cho dữ liệu chuỗi thời gian [18]; Đây là những tập dữ liệu đã được công bố và được sử dụng khá rộng rãi trong cộng đồng nghiên cứu về phát hiện người ngã và vận động bất thường. Các tập dữ liệu đều có những thử thách như không cân bằng (imbalanced) và có nhiều vận động bất thường khá giống với các hoạt động thường ngày (ngã ra giường vs. ngồi và nằm xuống giường).

2. 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 *điểm cân bằng giữa độ chính xác và độ bao phủ ($F1_{score}$)*:

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1_{score} = \frac{2 * (Precision * Recall)}{Precision + Recall} \quad (3)$$

Trong đó, *True Positive (TP)* là tỉ lệ đo số lần mô hình phát hiện đúng vận động bất thường a và số lần thực tế xảy ra vận động bất thường a ; ví dụ *ngã nghiêng bên phải* được phát hiện đúng là *ngã nghiêng bên phải*. *True Negative (TN)* là tỉ lệ đo số lần mô hình phát hiện đúng không phải vận động bất thường a và số lần thực tế xảy ra không phải vận động bất thường a ; ví dụ *không phải là ngã nghiêng bên phải* được phát hiện đúng không là *ngã nghiêng bên phải*. *False Positive (FP)* là tỉ lệ đo số lần mô hình phát hiện sai vận động bất thường a và số lần thực tế xảy ra không phải vận động bất thường a ; ví dụ *ngã nghiêng bên phải* được phát hiện sai không phải là *ngã nghiêng bên phải*. *False Negative (FN)* là tỉ lệ đo số lần mô hình phát hiện sai không phải vận động bất thường a và số lần thực tế xảy ra vận động bất thường a ; ví dụ *không phải ngã nghiêng bên phải* được phát hiện sai là *ngã nghiêng bên phải*.

3. Các mô hình thử nghiệm (Baselines)

Chúng tôi thực nghiệm với một số mô hình sau:

- *Máy véc tơ hỗ trợ (SVM)*: với các bước tiền xử lý và trích xuất đặc trưng từ dữ liệu cảm biến được tham khảo từ nghiên cứu [9]. Các véc tơ được tính từ các cửa sổ trượt được dùng để huấn luyện mô hình SVM với tham số $C=1$, λ là kết quả của tìm kiếm lưới (grid search) và hàm tích RBF.

- *Mạng CNN* [11]: được hiệu chỉnh để thích hợp với dữ liệu cảm biến [5] của từng tập dữ liệu thử nghiệm như: số lớp tích chập là 3, có 2 lớp max pooling và theo sau là 2 lớp kết hợp đầy đủ (Fully Connected). Số đầu ra của lớp softmax được điều chỉnh bằng số nhân vận động bất thường trên từng tập dữ liệu. Để cải thiện hiệu suất huấn luyện và dự đoán, chúng tôi sử dụng kỹ thuật tối ưu Rectified Adam [19].

- *Mạng LSTM* [16]: được hiệu chỉnh để phù hợp cho các pha huấn luyện và dự đoán trên các tập dữ liệu thử nghiệm. Với đặc tính có thể nhớ thông tin trong một khoảng thời gian dài thì những đặc trưng ở

mức cao trích chọn từ dữ liệu cảm biến được sử dụng hiệu quả tại bước dự đoán.

4. Kết quả và đánh giá

Chúng tôi sử dụng phương pháp kiểm chứng chéo 10 lần. Với phương pháp này, mỗi tập dữ liệu được chia thành 10 phần bằng nhau; 9 phần được lấy ra để huấn luyện và 1 phần được sử dụng để kiểm chứng. Quá trình này lặp lại cho đến khi cả 10 phần được kiểm chứng và kết quả được tính trung bình. Kết quả tổng thể được trình bày trong Bảng I. Trong Bảng

Bảng I
KẾT QUẢ (F1-SCORE) TRÊN 4 TẬP DỮ LIỆU

PP/D.liệu	UTD	MobiFall	PTITAct	CMDFALL
SVM	0.85	0.79	0.87	0.45
CNN	0.94	0.88	0.91	0.82
LSTM	0.92	0.85	0.89	0.80
CNN-LSTM	0.96	0.95	0.93	0.85

I, SVM là bộ phân loại đã từng cho kết quả khá tốt với các đặc trưng được trích chọn thủ công [9]. Tuy nhiên, so với các mô hình học sâu thì SVM thấp hơn đáng kể. Mô hình học sâu CNN với khả năng học các đặc trưng tự động tốt qua các phép tích chập giữa các bộ lọc, đã lựa chọn được các đặc trưng với đặc tính không gian (spatial) rất hiệu quả, đã cho kết quả tốt hơn đáng kể so với SVM. Mô hình LSTM cho kết quả tương đối tốt xấp xỉ với mô hình CNN. Mặc dù học và biểu diễn các đặc trưng không gian chưa phải là điểm mạnh của LSTM, nhưng với khả năng nhớ các thông tin theo chuỗi thời gian trong khoảng thời gian dài cũng giúp LSTM có khả năng dự đoán khá tốt, cạnh tranh được với CNN. Cuối cùng là mô hình đề xuất CNN-LSTM đã cho kết quả cao nhất 96% F1-score trên tập UTD, 95% trên tập MobiFall, 93% trên tập PTITAct, và 85% trên tập CMDFALL. Đây là kết quả cải tiến rất đáng kể so với 3 phương pháp còn lại. Điều này cũng cho thấy mô hình CNN-LSTM hiệu quả hơn nhờ sự kết hợp của việc học và biểu diễn các đặc trưng của dữ liệu theo không-thời gian.

Trong 4 tập dữ liệu kể trên thì tập UTD đơn giản nhất chỉ với 1 vận động bất thường (ngã), tiếp theo tập MobiFall với 4 vận động bất thường. Trong khi đó tập PTITAct và CMDFALL lần lượt là 8 và 11 vận động bất thường. Đặc biệt tập CMDFALL có nhiều vận động bất thường phức tạp hơn các tập dữ liệu khác nên điều này cũng lý giải kết quả các mô hình trên tập

Bảng II

KẾT QUẢ CỦA MÔ HÌNH CNN-LSTM PHÁT HIỆN VẬN ĐỘNG BẤT THƯỜNG TRONG TẬP DỮ LIỆU CMDFALL

Tên hoạt động	Precision	Recall
ngã về phía sau	85.43	79.19
bò trên mặt đất	86.31	84.21
ngã về phía trước	89.56	87.58
ngã về bên trái	87.63	89.14
nằm trên giường và ngã về bên trái	70.42	67.30
nằm trên giường và ngã về bên phải	66.43	68.57
ngã về bên phải	91.62	92.25
ngồi trên ghế và ngã về bên trái	83.26	81.98
ngồi trên ghế và ngã về bên phải	79.12	78.67
nhảy loạn choạng	93.02	92.71
đi loạn choạng	84.25	82.59
Trung bình	86.46%	83.59%

CMDFALL đều thấp hơn các tập dữ liệu khác. Bảng II trình bày kết quả chi tiết phát hiện vận động bất thường của mô hình đề xuất CNN-LSTM thử nghiệm trên tập CMDFALL. Kết quả ở Bảng II cho thấy, CNN-LSTM có thể đạt tới độ chính xác là 86,46% và độ bao phủ 83,59% trên tập dữ liệu CMDFALL. Đây cũng là kết quả tốt nhất so với các phương pháp khác. Một số vận động bất thường rất phức tạp như nằm trên giường và ngã cũng được phát hiện chính xác lên tới 70%. Trong khi đó các tư thế ngã về phía trước, ngã về bên phải, ngã về bên trái v.v... đều được phát hiện với độ chính xác xấp xỉ tới 90%.

V. KẾT LUẬN

Chúng tôi đã đề xuất một mô hình học sâu tích chập kết hợp với mạng bộ nhớ dài ngắn CNN-LSTM để giải quyết bài toán phát hiện các vận động bất thường của người sử dụng cảm biến đeo trên người. Kiến trúc đề xuất CNN-LSTM đã tận dụng được đặc tính không-thời gian của dữ liệu cảm biến để tự động học và biểu diễn các đặc trưng hiệu quả trên dữ liệu cảm biến không thuần nhất. Kết quả thử nghiệm trên 4 tập dữ liệu UTD, MobiFall, PTITAct và CMDFALL cho thấy mô hình đề xuất đã cho kết quả tốt hơn đáng kể so với các mô hình máy véc tơ hỗ trợ (SVM), mô hình học sâu tích chập (CNN) và mô hình mạng bộ nhớ dài ngắn (LSTM). Đặc biệt với độ chính xác lên tới hơn 85% trên bộ dữ liệu CMDFALL cho thấy khả năng phát hiện tốt các vận động bất thường phức tạp. Kết quả này có nhiều tiềm năng cho các ứng dụng hỗ trợ theo dõi người bệnh Parkinson, bệnh về vận động và người cao tuổi.

LỜI CẢM ƠN

Nghiên cứu này được hỗ trợ bởi Quỹ Phát triển Khoa học và Công nghệ Quốc gia (NAFOSTED) với mã số 102.04-2016.23.

TÀI LIỆU THAM KHẢO

- [1] Hoey J, Plotz T, Jackson D, Monk A, Pham C, Olivier P (2011) "Rapid specification and automated generation of prompting systems to assist people with dementia." *Pervasive and Mobile Computing* 7(3):299-318, DOI 10.1016/j.pmcj.2010.11.007
- [2] Gao Y, Long Y, Guan Y, Basu A, Baggaley J, Plotz T (2019) "Towards reliable, automated general movement assessment for perinatal stroke screening in infants using wearable accelerometers." *Proc ACM Interact Mob Wearable Ubiquitous Technol* 3(1):12:1-12:22, DOI 10.1145/3314399
- [3] Khan A, Mellor S, Berlin E, Thompson R, McNaney R, Olivier P, Plotz T (2015) "Beyond activity recognition: Skill assessment from accelerometer data." In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, UbiComp'15*, pp 1155-1166, DOI 10.1145/2750858.2807534
- [4] Pham C., Nguyen ST, Tran QH, Tran S, Vu H, Tran TH, Le TL (2020) "SensCapNet: Deep neural network for non-obtrusive sensing based Human activity recognition." *IEEE Access* 8:86934:86946, DOI 10.1109/ACCESS.2020.2991731
- [5] Pham C, Diep NN, Phuong TM (2017) "E-shoes: Smart shoes for unobtrusive human activity recognition." In: *9th International Conference on Knowledge and Systems Engineering, KSE 2017, Hue, Vietnam, October 19-21, 2017*, pp 269-274, DOI 10.1109/KSE.2017.8119470
- [6] Pavlo D, Feichtenhofer C, Grangier D, Auli M (2019) "3d human pose estimation in video with temporal convolutions and semi-supervised training." In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*
- [7] Sarita C, Mohd AK, Charul (2018) "Multiple anomalous activity detection in videos." In: *Procedia Computer Science* 125 (2018) pp. 336-345.
- [8] Tran TH, Le T, Pham DT, Hoang VN, Khong VM, Tran QT, Nguyen TS, Pham C (2018) "A multi-modal multi-view dataset for human fall analysis and preliminary investigation on modality." pp 1947-1952, DOI 10.1109/ICPR.2018.8546308
- [9] Nguyen, L., Le, A., T., Pham, C.; (2018) "The Internet-of-Things based Fall Detection Using Fusion Feature." In *proc. of the 10th IEEE International Conference on Knowledge Systems Engineering (KSE)*. 129-134
- [10] Ordonez F, Roggen D (2016) "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition." *Sensors* 16(1):115
- [11] Munzner S, Schmidt P, Reiss A, Hanselmann M, Stiefelhagen R, Durichen R (2017) "Cnn-based sensor fusion tech-niques for multimodal human activity recognition." In: *Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pp 158-165
- [12] Guan Y, Plotz T (2017) "Ensembles of deep lstm learners for activity recognition using wearables." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1(2):1-28
- [13] Ignatov A (2018) "Real-time human activity recognition from accelerometer data using convolutional neural net-works." *Applied Soft Computing* 62:915-922
- [14] Dawar N, Kehtarnavaz N. (2018) "A Convolutional Neural Network-Based Sensor Fusion System for Monitoring Transition Movements in Healthcare Applications." In: *proceeding of ICCA* 482-485. 10.1109/ICCA.2018.8444326.
- [15] Vavoulas G, Padiaditis M, Chatzaki C, Spanakis E, Tsiknakis Manolis, (2016) "The MobiFall Dataset: Fall Detection and Classification with a Smartphone." *International Journal of Monitoring and Surveillance Technologies Research*. 2. 44-56. 10.4018/ijm-str.2014010103.
- [16] Liu J, Shahroudy A, Xu D, Wang G (2016) "Spatio-temporal lstm with trust gates for 3d human action recognition." In: *European conference on computer vision, Springer*, pp 816-833
- [17] Chatzaki C, Padiaditis M, Vavoulas G, Tsiknakis M. (2017) "Human Daily Activity and Fall Recognition Using a Smartphone's Acceleration Sensor." 100-118. 10.1007/978-3-319-62704-5-7.
- [18] Jinsung Y, Danial J, Mihaela VDS, (2019) "Time-series Generative Adversarial Networks." In: *proc of 33rd conference on Neural Information Processing Systems (NeurIPS)* pp.1-11.
- [19] Liu L, et al. (2020) "On the variance of the adaptive learning rate and beyond." In *proc. of the international conference on Learning Representation 2020*. <https://arxiv.org/pdf/1908.03265.pdf>
- [20] Hochreiter S, Schmidhuber J (1997) "Long short-term memory." *Neural Computation* 9(8):1735-1780, DOI 10.1162/neco.1997.9.8.1735
- [21] Markham A, Trigoni N (2019) "Selective sensor fusion for neural visual-inertial odometry." In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 10542-10551
- [22] Pham C, Nguyen TTT (2016) "Real-time traffic activity detection using mobile devices." In: *Proceedings of the 10th ACM International Conference on Ubiquitous Information Management and Communication*

(IMCOM) 1-7

[23] Pham VC (2012) "Human activity recognition for pervasive interaction." PhD thesis. Newcastle University

SƠ LƯỢC VỀ CÁC TÁC GIẢ

Nguyễn Tuấn Linh



Tốt nghiệp đại học ngành Công nghệ Thông tin, Đại học Giao thông Vận tải Hà Nội năm 2004. Nhận bằng Thạc Sĩ tại Đại học Thái Nguyên năm 2007.

Hiện là nghiên cứu sinh tại Học viện Công nghệ Bưu chính Viễn Thông.

Lĩnh vực nghiên cứu: kỹ thuật máy tính, điện toán tỏa khắp,

các mô hình học máy và công nghệ cảm biến cho các ứng dụng chăm sóc sức khỏe.

Nguyễn Văn Thủy



Tốt nghiệp đại học năm 1999 ngành Điện tử - Viễn thông, trường Đại học Bách khoa Hà Nội, nhận bằng Thạc sĩ năm 2005 tại ĐH bang New Mexico, Hoa Kỳ, Tiến sĩ năm 2012 tại ĐH Texas at Dallas, Hoa Kỳ.

Hiện công tác tại Học viện Công nghệ Bưu chính Viễn Thông.

Lĩnh vực nghiên cứu: học máy,

lý thuyết thông tin, hệ thống thông tin thông minh, ứng dụng học sâu trong các hệ thống thông tin băng rộng.

Phạm Văn Cường



Tốt nghiệp đại học năm 1998 ngành Công nghệ Thông tin, Đại học Quốc gia Hà Nội, nhận bằng Thạc sĩ năm 2005 tại ĐH bang New Mexico, Hoa Kỳ, Tiến sĩ năm 2012 tại ĐH Newcastle, Vương Quốc Anh.

Hiện là Phó giáo sư, giảng dạy tại khoa CNTT1, Học viện Công nghệ Bưu chính Viễn Thông.

Lĩnh vực nghiên cứu: học máy, điện toán tỏa khắp, tương tác người máy, nhận dạng hoạt động của người, các thuật toán học máy và công nghệ cảm biến cho các ứng dụng chăm sóc sức khỏe, thị giác máy tính, các công nghệ cảm biến, hệ thống nhúng và điều khiển.