NHẬN DẠNG GIỌNG CHỮ CÁI TIẾNG VIỆT SỬ DỤNG DEEP BOLTZMANN MACHINES

Hoàng Thị Thanh Giang, Nguyễn Thị Thúy Hạnh*, Nguyễn Trọng Kương

Khoa Công nghệ thông tin, Học viện Nông nghiệp Việt Nam

Tác giả liên hệ: ntthuyhanh@vnua.edu.vn

Ngày nhận bài: 30.12.2019 Ngày chấp nhận đăng: 26.09.2020

TÓM TẮT

Nhận diện giọng nói là một bài toán thu hút được quan tâm rộng rãi của nhiều nhà nghiên cứu trong lĩnh vực trí tuệ nhân tạo trong những năm gần đây. Chẳng hạn như bài toán xây dựng chương trình để robot có khả năng nhận biết giọng nói của con người, hay các thiết bị có thể hiểu và đối thoại trực tiếp với người cùng nói chuyện. Trong nghiên cứu này, 37 sinh viên của Học viện Nông nghiệp Việt Nam tham gia để thu thập dữ liệu phát âm liên tục 29 chữ cái trong bảng chữ cái tiếng Việt. Qua bước tiền xử lý dữ liệu để trích xuất ra các mẫu âm thanh thuộc tính cho phân lớp, phương pháp nhận dạng chúng tôi sử dụng để nhận diện các mẫu giọng nói là deep Boltzmann machine (DBM), một mạng có khả năng học sâu với kiến trúc nhiều tầng ẩn. Để đánh giá khả năng nhận dạng của phương pháp đề xuất, chúng tôi so sánh DBM với mạng nơron truyền thống (NN) có cùng kiến trúc số tầng ẩn. Kết quả cho thấy khả năng nhận dạng các mẫu âm thanh chữ cái tốt hơn của DBM với khả năng học cho độ chính xác trung bình là 68% trên dữ liệu đào tạo và 51% khi thử với dữ liệu test, trong khi kết quả này của NN là 61% và 48% tương ứng.

Từ khóa: Trí tuệ nhân tạo, học máy, mạng nơron, máy Boltzmann, học sâu.

Speech Recognition of Vietnamese Alphabet using Deep Boltzmann Machines

ABSTRACT

Speech recognition has been attracting many researchers in the field of artificial intelligence recently. For example, the problem of implementing a program for robots to recognize human speech, thereby robots can understand, learn and talk with human. In this study, 37 students from Vietnam National University of Agriculture were involved to acquire speech data of 29 letters in Vietnamese alphabet. The data were preprocessed to extract featured voice chunks for the classification. We then used the deep Boltzmann machine (DBM) as a deep network with stacked hidden layers. To evaluate the proposed method, we compared the learning performance of DBM to a neural network (NN) with the same network structure configuration. The results showed that DBM performed better with accuracies of 68% on the training dataset and 51% on the test dataset, while the respective figures for NN were 61% of training and 48%.

Keywords: Artificial intelligence, machine learning, neural network, Boltzmann machine, deep learning.

1. ĐẶT VẤN ĐỀ

Nhận diện giọng nói là một bài toán thu hút được quan tâm rộng rãi của nhiều nhà nghiên cứu trong lĩnh vực trí tuệ nhân tạo chẳng hạn như bài toán xây dựng chương trình để robot biết nhận biết giọng nói của con người, từ đó phát triển để robot có thể hiểu và đối thoại với người cùng nói chuyện (Kazuhiro & cs., 2010). Hay trong công nghệ giáo dục, việc nhận biết chính xác cách phát âm của một từ cũng là một việc làm cần thiết để trợ giúp cho người bắt đầu học ngôn ngữ đó có thêm nhiều tiện ích trong rèn luyện cách phát âm và nhận biết âm chuẩn. Tuy nhiên ngôn ngữ và giọng nói có yếu tố vùng miền. Vì vậy, để một chương trình máy tính nhận biết được sự đa dạng cách phát âm của một ngôn ngữ thống nhất cũng là một bài toán cần giải quyết khả năng nhận dạng âm và giọng nói mà ở đó độ chính xác phụ thuộc vào khả năng phân lớp với dữ liệu đầy đủ nhất có thể.

Rõ ràng, việc tiếp nhận ngôn ngữ với con người là một quá trình học và lĩnh hội từng bước. Điều này càng thể hiện chi tiết hơn với

việc học một ngoại ngữ nào đó hoặc với trẻ em bắt đầu đi học. Cụ thể, để học nói được hoặc phân biệt từng chữ cái trong một từ thì người học từng bước học cách phát âm của từng chữ cái đó hoặc học cách phát âm cả cụm của một từ. Với một từ điển điện tử thì cách phát âm của một từ, một chữ cái đều lấy cách phát âm ở một vùng nào đó làm chuẩn. Với ngữ giọng khác nhau thì sự phát âm của người học so với một âm chuẩn có sự thay đổi ở mỗi người về âm lượng, ngữ điệu, tần số.

Về ứng dụng nhận biết cách phát âm từ, trong một nghiên cứu gần đây (Samuel & cs., 2018) nhóm tác giả đã nghiên cứu một mô hình mà robot có thể nhận biết cách phát âm của trẻ và đưa ra trợ giúp cho đứa trẻ rèn luyện được kỹ năng nói. Về mặt kỹ thuật, quá trình này gồm việc nhận biết âm thanh giọng nói và nhận biết ngữ nghĩa của ngôn ngữ nhận được.

Với ý tưởng tương tự cho tiếng Việt, một ngôn ngữ có nhiều giọng điệu khác nhau giữa hai miền Bắc và Nam (James & cs., 2010; Hoàng Thị Châu, 1999; Phuong & cs., 2008), mục đích của nghiên cứu này nhằm xây dựng từng bước một chương trình máy tính có thể nhận biết chữ cái thông qua nhiều giọng phát âm khác nhau, dần từng bước phát triển lên nhận biết từ, câu trong tiếng Việt, cũng như phát triển chương trình trợ giúp người học phát âm tiếng Việt trong tương lai.

Việc triển khai các ứng dụng của trí tuệ nhân tạo vào nhận diện các hoạt động của con người đã và đang thu hút rất nhiều nhóm nghiên cứu. Chẳng hạn như nhóm nghiên cứu của Thinh & cs. (2018), hay nghiên cứu của Orken & cs. (2019) cho thấy những nghiên cứu triển khai ứng dụng của thị giác máy tính và học sâu vào nhận diện hoạt động của con người. Các nghiên cứu đó đóng góp thêm vào khả năng ứng dụng đa dạng của trí tuệ nhân tạo trong thực tế.

Gần đây, các phương pháp của học sâu đã chứng tỏ khả năng ứng dụng cao vào các bài toán phân tích dữ liệu lớn và nó đang cuốn hút nhiều quan tâm (Lecun & cs., 2015; Schmidhuber, 2015). Với những thuật toán hiệu quả được trang bị cho việc xây dựng các mạng nhiều tầng, qua đó nâng cao khả năng biểu diễn và nhận biết thuộc tính của dữ liệu thông qua học không giám sát chẳng hạn như CD-k (Hilton, 2012). Trong nghiên cứu trước đây, chúng tôi đã sử dụng phương pháp của học máy để nhận biết các mẫu sóng siêu âm về động mạch vành tim người (Kuong & cs., 2017; 2018a; 2018b). Cụ thể, chúng tôi đã sử dụng DBM trong các nghiên cứu đó. Các kết quả ứng dụng khả năng học của DBM là cơ sở cho chúng tôi sử dụng để giải quyết cho bài toán nhận diện giọng nói trong nghiên cứu này.

2. PHƯƠNG PHÁP NGHIÊN CỨU

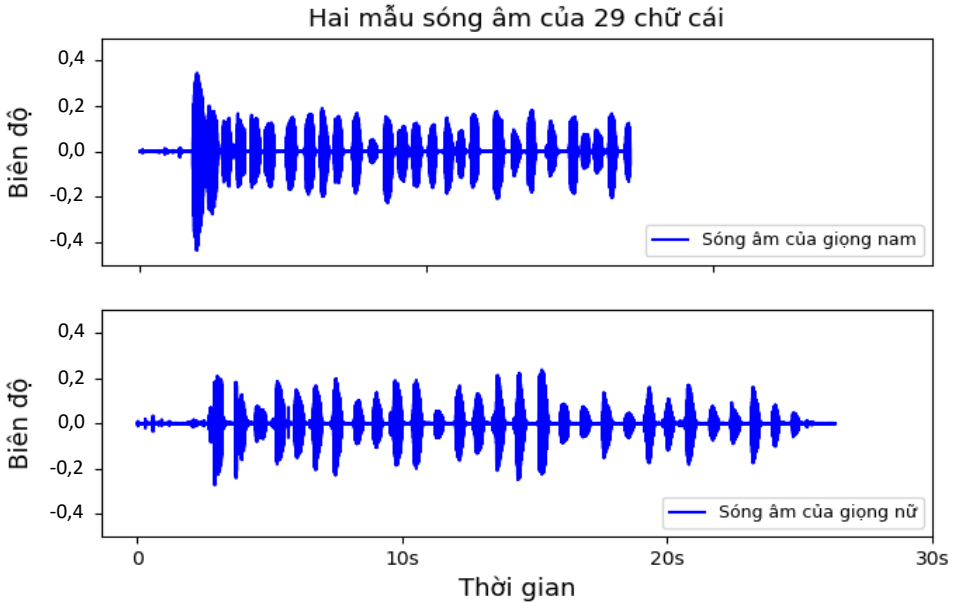
2.1. Thu thập dữ liệu

Dữ liệu được sử dụng trong nghiên cứu này dựa trên sự tham gia của nhóm gồm 37 sinh viên tình nguyện (19 nữ, 18 nam) khoa Công nghệ thông tin, Học viện Nông nghiệp Việt Nam, họ đã đồng ý tham gia cho thu âm giọng nói của mình. Trong khâu thu thập dữ liệu, chúng tôi dựa trên các bước sau:

Bước 1: Thiết kế bảng thu thập dữ liệu bao gồm thông tin về dự án nghiên cứu, mục đích nghiên cứu, các thỏa thuận xin dữ liệu, cam kết sử dụng dữ liệu và thông tin người phụ trách. Mỗi cá nhân tham gia quá trình thu mẫu hoàn toàn được phổ biến các thông tin này và ký thỏa thuận tự nguyện cũng như tinh thần sẵn sàng trợ giúp cho nghiên cứu.

Bước 2: Tìm hiểu thiết bị thu âm thanh. Dựa trên điều kiện vật chất và tìm hiểu các phần mềm thu âm. Cấu hình cách đặt thiết bị thu âm, chẳng hạn như, tần số lấy mẫu, cấu trúc tệp âm thanh thu được. Chúng tôi đi đến sử dụng phần mềm windows recorder, được xem là thuận tiện triển khai với nhóm nghiên cứu.

Bước 3: Lên qui trình và tập huấn lấy mẫu gồm: (i) phổ biến cho người tham gia lấy mẫu về mục đích và cam kết đảm bảo thông tin, (ii) tập huấn cho người phụ trách thu âm về qui trình này, và (iii) tiến hành thu âm sau khi đã liên lạc với các sinh viên tình nguyện.



Hình 6. Biểu diễn sóng âm của hai mẫu âm thanh

Kết quả các mẫu âm thanh thu được là 37 tệp âm thanh của 37 sinh viên tình nguyện, trong đó mỗi tệp là giọng phát âm liên tiếp của 29 chữ cái trong bảng chữ cái tiếng Việt dựa theo từ điển tiếng Việt của Hoàng Phê (2010). Biểu diễn dạng sóng của một tệp âm thanh được mình họa ở hình 1.

2.2. Xử lý và trích xuất đoạn âm thanh thuộc tính

Để tiền xử lý, chuẩn hóa dữ liệu và trích xuất đoạn âm thanh thuộc tính phục vụ cho học và phân lớp (mạng phân lớp được trình bày trong mục 2.3), trước hết tần số lấy mẫu được chúng tôi lấy chuẩn là FS = 22.050 mẫu/giây. Thông thường, dữ liệu chúng tôi thu âm có 2 mức tần số lấy mẫu là 44.100 mẫu/giây và 22.050 mẫu/giây.

Quan sát ở hình 1, dễ thấy rằng dựa vào biên độ dao động cho ta xác định vùng tương ứng với giọng phát âm của một chữ cái nào đó. Khi thiết lập một ngưỡng ngắt của biên độ thì cho phép ta tách các vùng tương ứng với mỗi nhãn là các chữ cái tương ứng, đó là các vùng quan tâm (ROI) cho việc trích ra các đoạn âm thanh thuộc tính phục vụ cho việc phân lớp. Khi thống kê từ dữ liệu chúng tôi có được trung bình khoảng thời gian cho các vùng đó khoảng 0,7 giây, nghĩa là có kích thước bằng 0,7*FS.

Để xác định các vùng âm thanh tương ứng với nhãn là các chữ cái, trước hết chúng tôi dựa vào các điểm đỉnh (peak points) ở đó theo tốc độ phát âm trung bình là 0,7*FS cho mỗi chữ cái nên các điểm đỉnh phải cách nhau tương tự là 0,7*FS. Vùng âm thanh thuộc tính quan tâm sẽ được trích ra xung quanh các điểm đỉnh. Mỗi vùng sóng âm tương ứng với mỗi chữ cái cho thấy biên độ ở vùng đó được dao động mạnh hơn như được thể hiện ở hình 2.

Khi vùng âm thanh được xác định, lấy điểm trung vị chia đôi năng lượng sóng âm của vùng đó làm tâm, đoạn âm thanh thuộc tính có độ dài là 0,7*FS lấy trung vị làm điểm giữa được xác định là đoạn âm thanh thuộc tính tương ứng với mỗi chữ cái. Mô tả sóng âm của một chữ cái và điểm trung vị được thể hiện ở hình 3.

2.3. Mạng deep Boltzmann machine

2.3.1. Mạng restricted Boltzmann machine (RBM) chuẩn

Mạng restricted Boltzmann machine (RBM) là một kiểu mạng nơron học phân bố xác suất của dữ liệu đầu vào, ở đó về kiến trúc, nó sử

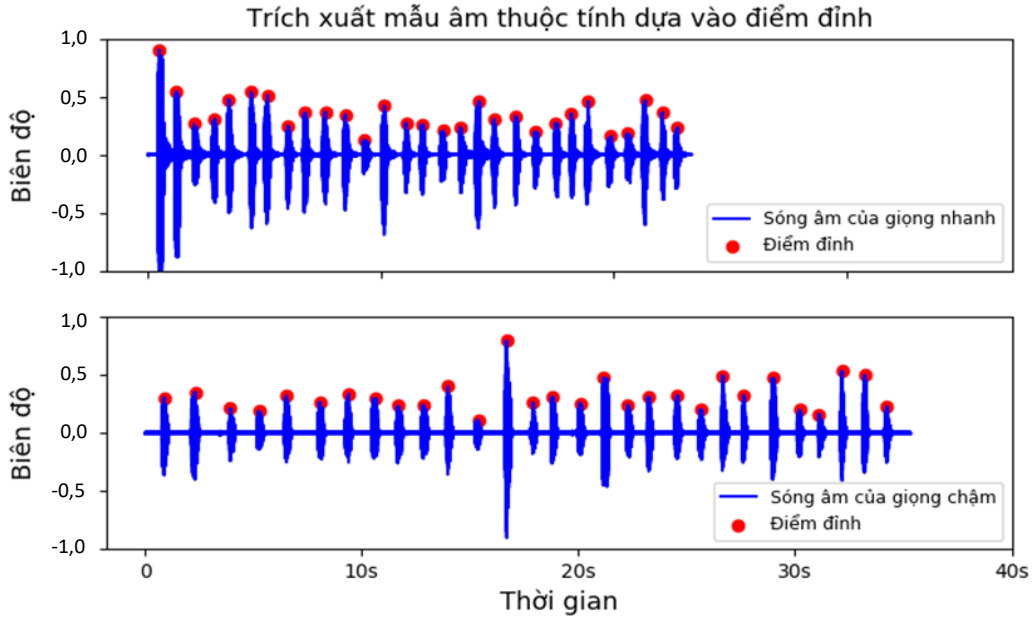
dụng các biến trong tầng ẩn h = (h_1, h_2,..., h_H) để học phân phối của các biến biểu diễn dữ liệu cần học hay dữ liệu input x = (x_1, x_2,..., x_N). Mỗi đơn vị x_i có sự kết nối với trọng số w_{ij} tới mỗi đơn vị h_j. Không có sự kết nối giữa các đơn vị trong cùng tầng ẩn hay cùng tầng dữ liệu. Các trọng số b_i và c_j phản ánh mức độ tác động của mỗi đơn vị x_i và h_j tương ứng trong mạng. Mạng RBM học thông qua việc điều chỉnh hàm năng lượng xác định bởi công thức (1):

$$E(x,h) = -\sum_{i,j} w_{ij} x_i h_j - \sum_i b_i x_i - \sum_j c_j h_j \quad (1)$$

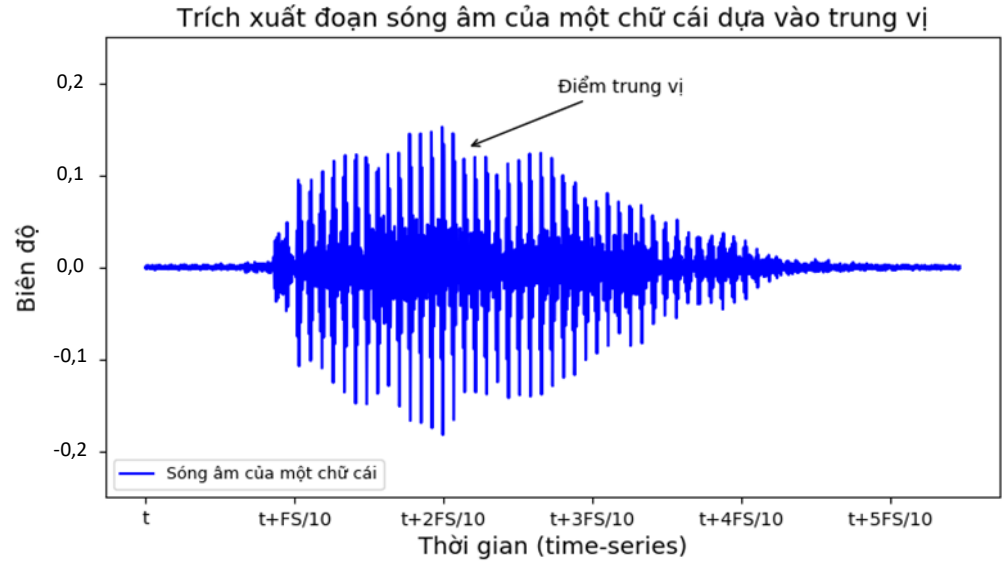
Phân phối đồng thời P(x, h) của và được xác định bởi phương trình (2) sau:

$$P(x,h) = \frac{\exp(-E(x,h))}{Z} \quad (2)$$

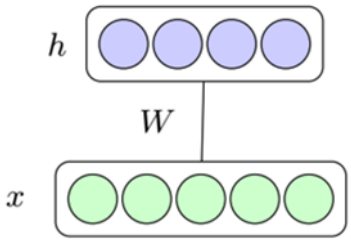
ở đó Z là hằng số chuẩn hóa. Xác suất có điều kiện cho các đơn vị h_j và x_i được xác định dựa theo phân phối Boltzmann bởi (3) và (4):



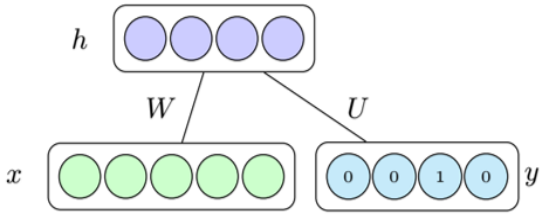
Hình 7. Trích xuất vùng sóng âm tương ứng với nhãn dựa vào các điểm đỉnh



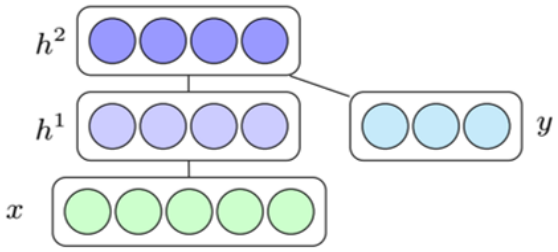
Hình 8. Xác định đoạn sóng âm thuộc tính cho phân lớp dựa vào trung vị



Hình 9. Mô hình RBM chuẩn với tầng input x và tầng ẩn h



Hình 10. Mô hình mạng classRBM



Hình 11. Mô hình mạng DBM

$$P\left(h_j \middle| x\right) = \text{sigm}\left(\sum_i w_{ij} x_i + c_j\right) \quad (3)$$

và

$$P\left(x_i \middle| h\right) = \text{sigm}\left(\sum_j w_{ij} h_j + b_i\right) \quad (4)$$

ở đó $\text{sigm}\left(x\right) = \frac{1}{1+e^{-x}}$ là hàm sigmoid. Mạng RBM chuẩn được mô tả như ở hình 4.

Mạng RBM chuẩn được trang bị thuật toán CD-k (Hilton, 2012), nó cho phép mạng có thể học không giám sát (unsupervised learning) phân phối đồng thời giữa các đơn vị tầng ẩn và tầng input. Ở một nghĩa nào đó, tầng ẩn tham gia vào học phân phối của input và đóng vai trò làm rút ngắn số chiều của tầng input.

2.3.2. Mạng restricted Boltzmann machine phân lớp

Mạng RBM phân lớp (classification restricted Boltzmann machine - classRBM) là một trường hợp mở rộng của RBM bằng cách thêm các đơn vị mã hóa cho nhãn tương ứng với các input (Hugo & cs., 2012). Cụ thể, nếu các input x có nhãn là k trong số K lớp của dữ liệu thì lớp cho nhãn gồm K đơn vị xác định bởi qui tắc "one-hot", hay đơn vị thứ k có giá trị bằng 1 còn lại bằng 0. Khi đó tương tự như mạng RBM chuẩn, hàm năng lượng được cho bởi các phương trình (5) dưới đây:

$$\begin{aligned} E\left(x, h, y_k\right) = &-\sum_{i,j} w_{ij} x_i h_j - \sum_i b_i x_i \\ &- \sum_j c_j h_j - \sum_j U_{kj} x h_j - d_k \end{aligned} \quad (5)$$

ở đó U_{kj}. d_k là các trọng số kết nối với các đơn vị ẩn và trọng số của đơn vị nhãn tương ứng. Không có kết nối giữa các đơn vị nhãn với các đơn vị input. Phân phối đồng thời của các đơn vị được xác định bởi:

$$P\left(x, h, y\right) = \frac{\exp\left(-E\left(x, h, y\right)\right)}{Z} \quad (6)$$

ở đó Z là hằng số chuẩn hóa. Các xác suất có điều kiện được xác định bởi:

$$P\left(h_j \middle| x, y_k\right) = \text{sigm}\left(\sum_i w_{ij} x_i + U_{kj} + c_j\right) \quad (7)$$

và

$$P\left(x_i \middle| h\right) = \text{sigm}\left(\sum_j w_{ij} h_j + b_i\right) \quad (8)$$

$$P\left(y_k \middle| h\right) = \frac{\exp\left(\sum_j U_{kj} h_j + d_k\right)}{\sum_l \exp\left(\sum_j U_{lj} h_j + d_l\right)} \quad (9)$$

Xác suất hậu nghiệm cho việc xác định phân lớp là:

$$P\left(y_k \middle| h\right) = \frac{\exp\left[d_k + \sum_j f\left(\sum_i w_{ij}x_i + U_{kj} + c_j\right)\right]}{\sum_{yl}\exp\left[d_l + \sum_j f\left(\sum_i w_{ij}x_i + U_{lj} + c_j\right)\right]} \quad (10)$$

ở đó f(x) = log(1 + exp(x)) là hàm softplus. Mô hình classRBM được minh họa ở hình 5.

Như vậy khi trang bị thêm tầng nhãn y thì mạng classRBM phục vụ cho việc học có giám sát (supervise learning). ClassRBM đã được chứng tỏ khả năng đào tạo hiệu quả với các thuật toán được trang bị như đã được trình bày bởi Hugo & cs. (2012).

2.3.3. Mạng deep Bolzmann machine và học sâu

Mạng deep Boltzmann machine (DBM) là sự xếp chồng của nhiều RBMs (Lecun & cs., 2015). Với thuật toán hiệu quả CD-k, nó cho phép tầng ẩn h tham gia vào học phân phối của input, đồng thời tầng ẩn lại tham gia như là một input cho tầng ẩn tiếp theo. Đó là cơ sở đẩy mạnh sự phát triển mạng học sâu. Trong nghiên cứu này chúng tôi sử dụng mạng DBM với 2 tầng ẩn ở đó tầng ẩn thứ hai có sự tham gia của mạng classRBM, nghĩa là, việc đào tạo ở mạng thứ 2 là học có giám sát kết hợp với nhãn để nhận diện các đoạn mẫu âm. Mô hình mạng DBM được sử dụng trong nghiên cứu này được mô tả ở hình 6.

2.4. Kết quả phân lớp

Bằng phương pháp trích xuất tự động như được trình bày trong phần 2.2, chúng tôi tiến hành kiểm tra lại và loại bỏ các đoạn có nhiễu âm hoặc các giọng phát âm không thực sự chính xác. Cuối cùng, chúng tôi thu được 817 mẫu âm của 29 chữ cái theo cách phát âm dựa của từ điển của Hoàng Phê (2010). Dữ liệu cho đào tạo (training data) và kiểm tra (test data) được chúng tôi chia ngẫu nhiên theo tỉ lệ 4:1 tương ứng.

Cấu hình cho mạng DBM trong nghiên cứu này là 700 × 150, tức là ở tầng ẩn thứ nhất có 700 nơron tham gia và tầng ẩn thứ 2 có 150 nơron tham gia. Giữa tầng input x và tầng ẩn h^1 là mạng RBM được đào tạo bởi thuật toán CD-1 như giới thiệu bởi Hilton (2012). Giữa tầng ẩn h^1 và tầng ẩn h^2 có sự tham gia của lớp nhãn y hay là mạng classRBM và được đào tạo bởi thuật toán học có giám sát của classRBM được trình bày bởi Hugo & cs. (2012). Các tham số W, U, b, c, d ban đầu được sinh ngẫu nhiên và nhỏ trong giới hạn $[-10^{-3}, 10^{-3}]$.

Để đánh giá khả năng học của DBM, chúng tôi so sánh kết quả của DBM với mạng nơron truyền thống thông thường (NN) trong cùng cấu trúc kích thước của mạng, nghĩa là mạng nơron được sử dụng có 2 tầng ẩn có kích thước lần lượt là 700 và 150, và các tham số của mạng NN cũng được thiết lập tương tự như DBM. Thống kê về độ chính xác của phân lớp ở một trường hợp tốt nhất của DBM và NN trong đào tạo và test được thể hiện ở bảng 1.

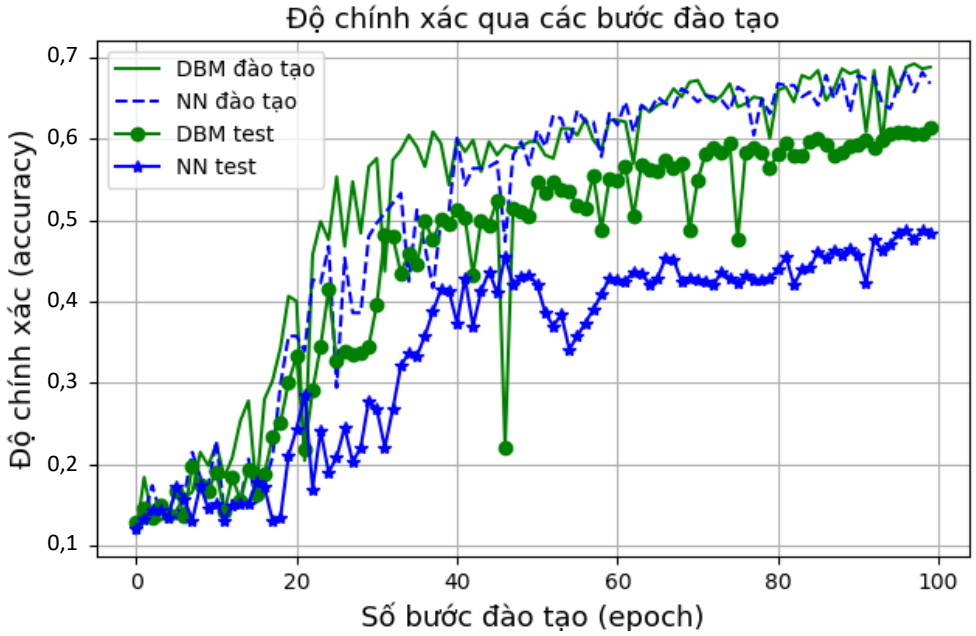
Để thấy được quá trình đào tạo của mạng qua từng bước học và cập nhật tham số, sau mỗi bước độ chính xác của phân lớp được kiểm tra và tính. Việc này được thể hiện ở hình 7. Ở đây chúng tôi tiến hành thực hiện đào tạo mạng ở 100 bước, mỗi bước cập nhật tham số lại thực hiện đánh giá khả năng nhận dạng các mẫu cho học và mẫu cho test lấy kết quả.

3. KẾT LUẬN

Nghiên cứu này đã thu thập được bộ dữ liệu mẫu phát âm bảng chữ cái tiếng Việt từ 37 sinh viên tình nguyện trong Học viện Nông nghiệp Việt Nam. Mặc dù dữ liệu chúng tôi thu được chưa đủ lớn, và chưa thể đầy đủ so với sự đa dạng của giọng phát âm tiếng Việt nói chung, nghiên cứu cũng góp phần bổ sung thêm cơ sở và dữ liệu cần thiết cho các nghiên cứu sâu hơn trong lĩnh vực này.

Bảng 1. Kết quả so sánh độ chính xác trong đào tạo và test của DBM và NN

	DBM	NN
Độ chính xác với dữ liệu đào tạo	68%	61%
Độ chính xác với dữ liệu test	51%	48%



Hình 7. Độ chính xác nhận dạng của DBM và NN qua các bước

Sử dụng mạng DBM trong nghiên cứu này cũng đã chứng tỏ được khả năng nhận dạng mẫu âm thanh và nó cho thấy ứng dụng của học sâu trong nhiều lĩnh vực tính toán nói chung và nhận diện giọng nói nói chung. Cho dù vậy, việc cải thiện khả năng học của mạng DBM cũng cần xem xét hơn nữa sau này. Cấu hình và thiết lập các tham số tối ưu ứng với dữ liệu có được cũng cần có đánh giá thêm.

LỜI CẢM ƠN

Nghiên cứu này được thực hiện từ nguồn kinh phí của đề tài: "Nhận diện chữ cái tiếng Việt qua dữ liệu phát âm của một nhóm sinh viên Học viện Nông nghiệp Việt Nam", mã số: T2019-10-55, cấp bởi Học viện Nông nghiệp Việt Nam. Chúng tôi cũng xin cảm ơn nhóm sinh viên Khoa Công nghệ thông tin đã tham gia tình nguyện trợ giúp cho việc thu thập dữ liệu phục vụ cho nghiên cứu này.

TÀI LIỆU THAM KHẢO

Dhar V. (2015). Data science and prediction. Communications of the ACM, 56 (12): 64-73.

Hilton E.G. (2012). A practical guide to training restricted Boltzmann machines. Lecture Notes in Computer Science, Springer Berlin. 7700: 599-619.

Hoàng Thị Châu (1999). Tiếng Việt trên các miền đất nước (Phương ngữ học). Nhà xuất bản Khoa học Xã hội, Hà nội.

Hoàng Phê (2010). Từ điển tiếng Việt. Nhà xuất bản Đà Nẵng.

Hugo L., Michael M., Razvan P. & Yoshua B. (2012). Learning algorithms for the classification restricted Boltzmann machine. Machine Learning Research. 13(1): 643-669.

James K. (2010). Dialect experience in Vietnamese tone perception. The Journal of the Acoustical Society of America. 127(6): 3749-3757.

Kazuhiro N., Toru T., Hiroshi G.O., Hirofumi N., Yuji H. & Hiroshi T. (2010). Design and implementation of robot audition system HARK - open source software for listening to three simultaneous speakers. Advanced Robotics. 24(5): 739-761.

Kuong N.T., Uchino E. & Suetake N. (2017). IVUS tissue characterization of coronary plaque by classification restricted Boltzmann machine. Journal of Advanced Computational Intelligence and Intelligent Informatics. 21(1): 67-73.

Kuong N.T., Uchino E. & Suetake N. (2018a). Recognition of coronary atherosclerotic plaque tissue on intravascular ultrasound images by using misclassification sensitive training of discriminative restricted boltzmann machine. Journal of Biomimetics, Biomaterials and Biomedical Engineering. 37: 85-93.

Kuong N.T., Uchino E. & Suetake N. (2018b). Coronary plaque classification with accumulative training of deep Boltzmann machines. ICIC Express Letters. 12(9): 881-886.

Lecun Y., Yoshua B. & Hinton E.G. (2015). Deep learning. Nature. 521(7553): 436-444.

Orken M., Nurbapa M., Mussa T., Nurzhamal O., Tolga I.M. & Aigerim Y. (2019). Voice identification using classification algorithms. Intelligent system and computing. Book chapter, InTechOpen.

Phuong P.A., Tao N.Q. & Mai L.C. (2008). An efficient model for isolated Vietnamese handwritten recognition. Proceedings of 2008 international conference on intelligent information hiding and multimedia signal processing. pp. 358-361.

Samuel S., Huili C., Safinah A., Michael K. & Cynthia B. (2018). A social robot system for modeling children's Word pronunciation: socially interactive agents track. Proceedings of the 17th international conference on autonomous agents and multi-agent systems. pp. 1658-1666.

Schmidhuber J. (2015). Deep Learning in neural networks: an overview. Neural Networks. 61: 85-117.

Thinh D.B, Dat T.T., Thuy T.N., Long Q.T. & Van D.N. (2018). Aerial Image Semantic Segmentation using Neural Search Network Architecture. In Proceedings of Multi-Disciplinary International Conference on Artificial Intelligence (MIWAI), Lecture Notes in Artificial Intelligence, Springer.