Tạp chí Tin học và Điều khiển học, T. 19, S. 2 (2003), 131–138

NHẬN DẠNG TỪ CÓ THANH ĐIỆU KHÁC NHAU TRONG TIẾNG VIỆT

ĐẶNG NGỌC ĐỨC[1], LƯƠNG CHI MAI[2]

[1] *Alcatel Network System Vietnam*
[2] *Viện Công nghệ thông tin*

Abstract. Vietnamese is a mono-syllabic tonal language. Recognition of Vietnamese syllables with six different tones is one of problems of Vietnamese automatic recognition systems. In this paper, we present the speech recognition experiments with a Vietnamese speech database of words, which have the same initial and final, but different tones. The database contains 294 sentences of six words "na, ná, nà, nạ, nã, nả" in random order, which are recorded with a man voice in the office environment. Three systems have been developed using Markov Hidden Model (HMM) and HMM/ Neural Network hybrid separately. Three systems are trained with the same set of 214 sentences and then are tested with the same set of 63 sentences, which are independent with previous training set. The experiments show that the hybrid system has best result with word-level accuracy 94.93% and sentence- level accuracy 73.91%.

Tóm tắt. Tiếng Việt là một ngôn ngữ đơn âm và có thanh điệu. Việc nhận dạng các âm tiết tiếng Việt cùng với thanh điệu là một trong các vấn đề của hệ thống nhận dạng tiếng Việt. Bài báo này trình bày quá trình thử nghiệm nhận dạng trên một cơ sở dữ liệu tiếng gồm một tập các từ tiếng Việt giống nhau về âm đầu, âm vần và chỉ khác nhau về thanh điệu. Cơ sở dữ liệu tiếng bao gồm 294 câu, mỗi câu gồm có 6 từ " na, ná, nà, nạ, nã, nả" được sắp xếp theo thứ tự ngẫu nhiên, thu âm do giọng một người đọc trong môi trường văn phòng. Thử nghiệm áp dụng các phương pháp nhận dạng tiếng: mạng nơ ron nhiều lớp, mô hình Markov ẩn (HMM) và hệ thống lai ghép giữa mạng nơ ron và mô hình Markov ẩn(NN-HMM). Các hệ thống nhận dạng được huấn luyện bằng cùng một tập gồm 214 câu, sau đó được tiến hành nhận dạng trên một tập kiểm tra gồm 63 câu, độc lập với các câu đã dùng để huấn luyện trước đó. Kết quả nhận dạng cho thấy hệ thống NN-HMM cho kết quả nhận dạng cao nhất với độ chính xác 94.93% ở mức từ và 73.91% ở mức câu.

1. ĐẶT VẤN ĐỀ

Tiếng Việt được biết đến như là một ngôn ngữ đơn âm, có thanh điệu. Mỗi âm tiết đều có một thanh điệu và thanh điệu đóng vai trò là một âm vị mang tính siêu đoạn. Đó là loại âm vị không có âm đoạn, không độc lập tồn tại, nhưng cũng có chức năng phân biệt nghĩa, nhận diện từ. Đây là đặc điểm riêng của tiếng Việt so với các ngôn ngữ Châu Âu. Một số ngôn ngữ khác như tiếng Hán, tiếng Thái cũng có đặc điểm này như tiếng Việt.

Cấu trúc âm tiết tiếng Việt gồm có hai bậc, trong đó tại bậc 1 có 22 âm đầu, bậc 2 gồm 155 phần vần và 6 thanh điệu [1]. Âm đầu cùng với phần vần có thể kết hợp với 6 thanh điệu khác nhau. Tuy nhiên trên thực tế có một số kết hợp không tồn tại nên chỉ có khoảng 6700 âm tiết trong tiếng Việt. Theo các nhà ngôn ngữ học, mặc dù thanh điệu có ảnh hưởng bao chùm lên toàn bộ âm tiết, nhưng gánh nặng chủ yếu tập trung ở phần vần.

Cho đến nay chưa có nhiều nghiên cứu về nhận dạng thanh điệu trong tiếng Việt cũng như chưa có nhiều nghiên cứu về một hệ thống nhận dạng tiếng Việt hoàn chỉnh với số lượng từ vựng lớn. Việc nghiên cứu ảnh hưởng của thanh điệu trong nhận dạng tiếng Việt là cần

thiết để giúp cho quá trình xây dựng một hệ thống nhận dạng tiếng Việt sau này. 2. Mục đích của thử nghiệm này là dùng các công cụ đã được áp dụng thành công trong bài toán nhận dạng tiếng nói như mạng nơ ron và mô hình Markov để nghiên cứu ảnh hưởng của thanh điệu trong nhận dạng tiếng Việt. Phần còn lại của bài báo được cấu trúc như sau. Phần 2 nêu lại một số kiến thức cơ bản về mô hình Markov ẩn được dùng trong các phần sau. Phần 3 là phần chính của bài báo trình bày các mô hình nhận dạng thử nghiệm dựa trên mô hình Markov ẩn, mô hình kết hợp giữa mạng nơ ron và mô hình Markov ẩn. Các kết quả đánh giá thử nghiệm của từng mô hình cũng được trình bày tiếp sau các mô hình. Cuối cùng là phần kết luận.

MỘT SỐ KHÁI NIỆM CƠ BẢN

2.1. Mô hình Markov ẩn

Một mô hình Markov ẩn (HMM) [5] là một tiến trình ngẫu nhiên kép, trong đó có một tiến trình ẩn chuyển trạng thái theo chuỗi Markov rời rạc và thuần nhất xen kẽ với một tiến trình phát sinh dãy quan sát. Các ký hiệu được sử dụng trong mô hình Markov ẩn là:

N - số trạng thái trong mô hình

M - số ký hiệu quan sát có thể

T - độ dài của dãy quan sát (số ký hiệu trong dãy quan sát)

$\{1, 2, .., N\}$ - tập các trạng thái

q_t - trạng thái của mô hình tại thời điểm t

$V = \{v_1, v_2, ..., v_M\}$ - tập rời rạc các ký hiệu quan sát

$\pi = \{\pi_1, \pi_2, ..., \pi_N\}$ - tập các phân bố xác suất cho trạng thái khởi đầu, π_i là xác suất để trạng thái i được chọn tại thời điểm khởi đầu $t = 1 : \pi_i = P(q_1 = i)$;

$$\begin{cases} \sum_{i=1}^{N} \pi_i = 1 \\ \pi_i \geq 0;\ i = 1, 2, ..., N \end{cases}$$

$A = \{a_{ij}\}$ - ma trận xác suất chuyển với a_{ij} là xác suất để trạng thái j xuất hiện tại thời điểm $t+1$ khi trạng thái i đã xuất hiện tại thời điểm t. Giả thiết rằng a_{ij} là độc lập với thời gian $t : a_{ij} = P(q_{t+1} = j/q_t = i)$,

$$\begin{cases} \sum_{j-1}^{N} a_{ij} = 1;\ i = 1, 2, ..., N \\ a_{ij} \geq 0;\ j = 1, 2, ..., N \end{cases}$$

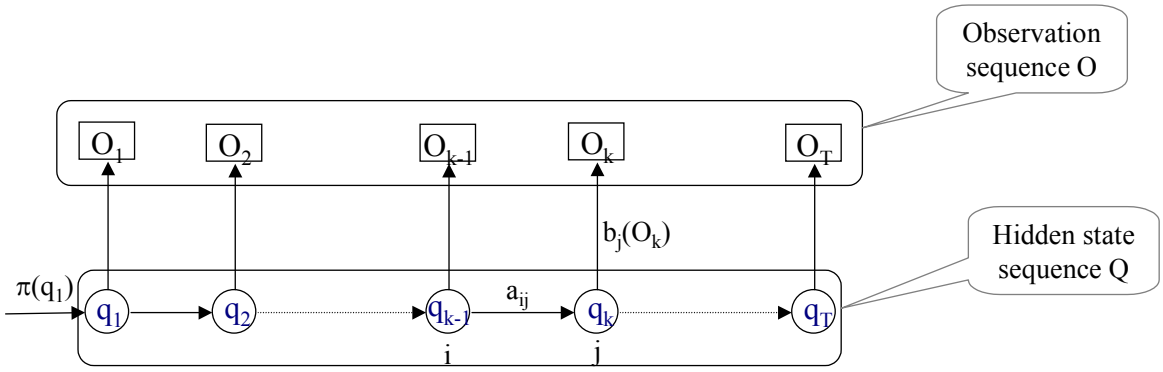
$B = \{b_j(v_k)\}$ - các hàm đo xác suất phát xạ mẫu, $b_j(v_k) = P(v_k$ được phát sinh khi mô hình ở trạng thái $j)$

$$\begin{cases} \sum_{k=1}^{M} b_j(v_k) = 1;\ j = 1, 2, ..., N \\ b_j(v_k) \geq 0;\ j = 1, 2, ..., N;\ k = 1, 2, ..., M \end{cases}$$

O_t biểu thị ký hiệu quan sát tại thời điểm t.

Bộ ba $\lambda = (A, B, \pi)$ được coi là ký pháp gọn của một mô hình Markov ẩn. A, B và π được gọi là bộ tham số (parameters) của mô hình λ. Hoạt động của HMM có thể mô tả như sau: tại thời điểm $t = 1$, mô hình ở trạng thái q_1 nào đó và phát sinh ra một ký hiệu quan sát nhất định O_1, sau đó, tại thời điểm $t = 2$, mô hình chuyển sang trạng thái q_2 và phát sinh ký hiệu quan sát O_2. Cứ tiếp tục như vậy cho đến thời điểm $t = T$, mô hình phát sinh được dãy quan sát $O = (O_1, O_2, ..., O_T)$ bằng dãy trạng thái $Q = (q_1, q_2, ..., q_T)$. Dãy trạng thái Q phụ thuộc vào xác suất chọn trạng thái khởi đầu π_i và xác suất chuyển a_{ij}. Dãy ký hiệu quan sát

$\{O_t\}$ được HMM phát sinh ra phụ thuộc vào dãy trạng thái Q và các hàm đo xác suất phát xạ mẫu $b_j(.)$. Trong trường hợp tập V các ký hiệu quan sát là không gian mẫu không đếm được, các hàm $b_j(.)$ có thể cho bằng hàm mật độ của một phân phối xác suất nào đó.



Hình 1. Mô hình Markov ẩn

2.2. Huấn luyện mô hình Markov ẩn

Bài toán. Với dãy huấn luyện O cần hiệu chỉnh các tham số của mô hình λ để cực đại hoá $P(O/\lambda)$. Ta có:

$$P(O,Q/\lambda) = \pi_{q_1}.b_{q_1}(O_1).a_{q_1q_2}.b_{q_2}(O_2).a_{q_2q_3}...a_{q_{T-1}q_T}.b_{q_T}(O_T)$$

và

$$P(O/\lambda) = \sum_Q P(O,Q|\lambda) = \sum_Q \pi_{q_1} b_{q_1}(O_1) a_{q_1q_2} b_{i_2}(O_2)...a_{q_{T-1}q_T}(O_T)$$

Đặt $\alpha_t(i = P(O_1, O_2, ..., O_t,\ q_t = i/\lambda)$ và $\beta_t(i) = P(O_{t+1}, O_{t+2}, ..., O_T/q_t = i, \lambda),\ 1 \le t \le T$ với giá trị khởi tạo $\lambda_1(i) = \pi_i b_i(O_1)$ và $\beta_T(i) = 1,\ 1 \le i \le N$

Định nghĩa công thức truy hồi $\alpha_{t+1}(j)$ cho tính toán thuận như sau:

$$\alpha_{t+1}(j) = \Big[\sum_{i=1}^{N} \alpha_t(i)a_{ij}\Big] b_j(O_{t+1}) \quad \text{với} \quad t = 1, 2, ..., T-1$$

Tương tự, định nghĩa công thức $\beta_t(i)$ cho tính toán ngược như sau:

$$\beta_t(i) = \Big[\sum_{j=1}^{N} a_{ij} b_j(O_{t+1})\Big] \beta_{t+1}(j) \text{ với } \quad t = T-1, T-2, ..., 1$$

Thuật toán tiến lùi Baum-Welch (Forward-Backward Baum-Welch algorithm):

Bước 1. Xác định:

$$\gamma_t(i) = P(q_t = i/O, \lambda) = \frac{P(q_t = i, O|\lambda)}{P(O|\lambda)} = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)}$$

Bước 2. Xác định: $\xi_t(i,j) = P(q_t = i, q_{t+1} = j/O, \lambda)$

$$= \frac{P(q_t = i, q_{t+1} = j, O|\lambda)}{P(O|\lambda)} = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)}$$

Bước 3. Chỉnh tham số:

$$\overline{\pi}_i = \gamma_1(i); \quad \overline{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}; \quad \overline{b}_j(v_k) = \frac{\sum_{t=1, O_t = v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

Bước 4. Nếu $P(O/\lambda_{\text{mới}}) \leq P(O/\lambda_{\text{cũ}})$ thì kết thúc khác đi quay lại bước 1.

2.3. Nhận dạng Mô hình Markov ẩn

Bài toán: Cho mô hình $\lambda = (A, B, \pi)$ và một dãy quan sát $O = (O_1, O_2, ..., O_T)$. Cần tìm dãy trạng thái $Q = (q_1, q_2, ..., q_T)$ để xác suất $P(O, Q/\lambda)$ đạt cực đại.

Thuật toán Viterbi:

Bước 1. Gọi:

$$f(k, j) = \max_{\{q_t\}_{t=1}^{k}, q_k = j} P(O_1, O_2, ..., O_k, q_1, q_2, ..., q_k|\lambda).$$

Bước 2. Khởi tạo cơ sở quy hoạch động: $f(1, j) = \pi_j b_j(O_1)$.

Bước 3. Tính bảng phương án f bằng công thức truy hồi:

$$f(k, j) = \max_{1 \leq i \leq N} (f(k-1, i).a_{ij}.b_j(O_k))$$

Lưu vết:

$$Trace(k, j) = \arg\max_{1 \leq i \leq N} (f(k-1, i).a_{ij}.b_j(O_k)), (k \geq 2).$$

Bước 4. Truy vết tìm dãy trạng thái tối ưu: $q_T = \arg\max_j f(T, j)$

$$q_t = Trace(t+1, q_t + 1),\ t = T-1, T-2, ..., 1.$$

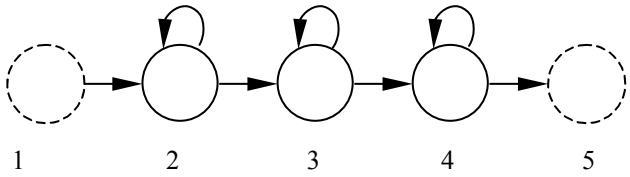
3. THỬ NGHIỆM NHẬN DẠNG THANH ĐIỆU TIẾNG VIỆT

3.1. Môi trường thử nghiệm

Cơ sở dữ liệu dùng cho thực nghiệm bao gồm 294 câu, mỗi câu gồm có 6 từ "na, ná, nà, nạ, nã, nả" được sắp xếp theo thứ tự ngẫu nhiên. Các câu được thu âm trong môi trường trong nhà, do một giọng nam đọc, sử dụng micro thông thường gắn với máy tính, card âm thanh Creative SoundBlaster, tốc độ lấy mẫu 8000Hz, PCM 8 bit mono 8kB/s. Tất cả các câu đều được gán nhãn bằng tay tới mức âm vị với 8 đơn vị nhận dạng: /n/, /as/, /af/, /ar/, /ax/, /aj/, /a/, /.pau/. Thử nghiệm dùng bộ thư viện Tooolkit của Trung tâm nghiên cứu nhận dạng tiếng nói (CSLU - Center of Spoken Language Understanding) do Viện sau đại học Oregon Hoa kỳ phát triển để xây dựng hệ thống nhận dạng dựa mô hình Markov [4] và kết hợp mạng nơ ron với mô hình Markov [2, 3, 6]. Phương pháp nhận dạng dùng bộ thư viện trong bài báo này dựa trên phân tích các khung tín hiệu (frame).

3.2. Thử nghiệm với mô hình Markov ẩn

Mô hình Markov được xây dựng dựa trên bộ thư viện CSLU Tooolkit bao gồm 5 trạng thái (Hình 2). Trong đó có ba trạng thái quan sát (observation state), 1 trạng thái khởi đầu và 1 trạng thái kết thúc [4].



Hình 2. Mô hình Markov ẩn dùng trong thử nghiệm.

Ma trận xác suất chuyển trạng thái trong mô hình được khởi tạo như sau:

$$\begin{matrix} 0.0 & 1 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.6 & 0.4 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.6 & 0.4 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{matrix}$$

Các quan sát O_j chính là vectơ đặc tính gồm 39 thành phần của từng khung tín hiệu. Với mỗi khung tín hiệu 10 ms, tính 13 hệ số cepstral MEL cùng với đạo hàm bậc một, bậc hai của từng hệ số và giá trị của từng hệ số trừ giá trị trung bình. Mô hình HMM monophone độc lập được áp dụng cho từng đơn vị nhận dạng là /n/, /as/, /af/, /ar/, /ax/, /aj/, /a/, /.pau/. Khởi tạo mô hình sử dụng phương pháp lượng tử hoá vectơ (VQ). Mô hình được huấn luyện dựa trên thuật toán EM (expectation/maximization). Trong huấn luyện, mô hình nhúng (embeded model - để nhận dạng các từ na, ná, nà, nạ, nã, nả) dùng để kết hợp các mô hình độc lập nhằm đánh giá lại các tham số dựa trên thuật toán tiến lùi Baum-Welch như đã trình bày trong phần 2.2. Mô hình được huấn luyện bằng 214 câu được gán nhãn bằng tay. Sau khi huấn luyện, sử dụng mô hình để nhận dạng trên một tập thử gồm 63 câu được chọn ngẫu nhiên từ cơ sở dữ liệu 330 câu, các câu dùng để kiểm tra này khác với các câu được dùng trong huấn luyện để đảm bảo tính khách quan. Sau đây là bảng kết quả nhận dạng các từ na, nả, nà, nạ, ná, nã dùng mô hình Markov ẩn. chính xác được chia thành hai mức: mức từ và mức câu.

Bảng 1. Độ chính xác của mô hình Markov ẩn.

Số câu dùng để huấn luyện	Độ chính xác	
	Từ	Câu
214	85%	52%

Mô hình Markov ẩn HMM đã được ứng dụng thành công trong các hệ thống nhận dạng tiếng. Điểm mạnh của HMM là rất phù hợp cho việc biểu diễn một chuỗi đơn vị tiếng nói theo thời gian. Tuy nhiên HMM có đặc điểm là mạnh về mô hình hoá từng loại mẫu nhưng yếu về khả năng phân biệt giữa các loại mẫu. Do đó kết quả nhận dạng của HMM đối với các từ có độ khác biệt ít như na, nả, nà, nạ, ná, nã có độ chính xác không cao (85.50%, 52.53% - Bảng 1).

Bảng 2. Tỷ lệ lỗi giữa các thanh điệu trong nhận dạng băng mô hình Markov ẩn.

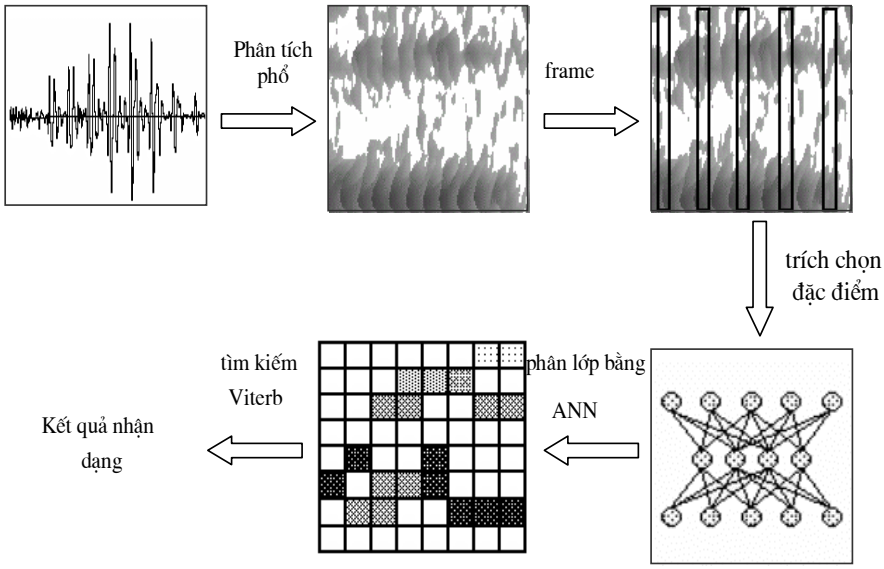
Thanh lỗi	Thanh bị nhận dạng nhầm						Tổng cộng
	Thanh sắc	Thanh huyền	Thanh hỏi	Thanh ngã	Thanh nặng	Thanh không	
Thanh sắc	-	0	0	1	0	1	2
Huyền	1	-	0	0	0	0	1
Hỏi	1	0	-	0	0	0	1
Ngã	5	1	0	-	1	1	8
Nặng	4	2	1	4	-	1	12
Không	1	0	0	0	0	-	1
Tổng cộng	12	3	1	5	1	3	25

Tỷ lệ nhận dạng đối với mức câu khá thấp do tỷ lệ lỗi chèn, xoá nhiễu khá cao (33.59%, 1.08%). Bảng 2 cho thấy số lượng lỗi nhận dạng nhầm giữa các thanh điệu. Kết quả cho thấy tỷ lệ nhận dạng nhầm ở thanh sắc là cao nhất (12 lỗi, 48%) và thanh hỏi và thanh nặng là thấp nhất (1 lỗi, 4%). Thanh dễ bị nhận dạng nhầm với thanh khác là thanh nặng (12 lỗi, 24%) và thanh ngã (8 lỗi, 32%).

3.3. Thử nghiệm với mạng nơ ron kết hợp với mô hình Markov

Quá trình xây dựng các hệ thống nhận dạng được tiến hành dựa trên bộ thư viện CSLU Tooolkit. Thực nghiệm được tiến hành theo phương pháp như sau. Tiếng nói đầu vào được lấy mẫu từng frame khoảng 10ms, hai frame cách nhau 30ms. Sau đó thông tin sẽ được phân tích thành 26 đặc tính bao gồm: 12 hệ số đặc tính ceptral và 12 giá trị đạo hàm của các hệ số này, cộng với mức công suất và giá trị đạo hàm mức công suất. Mỗi khung tín hiệu vào được kết hợp 4 khung phụ cận cách nhau - 60ms, - 30ms, 30ms, 60ms tạo thành một vector 130 đặc tính.

- Bước 1. Xây dựng mạng nơ ron 3 lớp bao gồm 130 nút đầu vào, tư-ơng ứng với vectơ 130 đặc tính của mỗi khung 10ms tín hiệu, 200 nút ẩn và 8 nút đầu ra tương ứng với 8 đơn vị nhận dạng. Huấn luyện mạng này bằng 214 câu được gán nhãn bằng tay. Tập các câu dùng để huấn luyện này cũng là tập được dùng để huấn luyện mô hình Markov trước đó.
- Bước 2. Dùng mạng xây dựng ở bước 1 để gán nhãn tự động các câu đã được gán nhãn bằng tay, sau đó dùng dữ liệu mới này để huấn luyện lại cho mạng đã được huấn luyện ở bước 1.
- Bước 3. Xây dựng mạng kết hợp giữa mạng nơ ron và mô hình Markov ẩn (NN-HMM) sử dụng mạng nơ ron được xây dựng ở bước 2.



Hình 3. Qui trình nhận dạng kết hợp giữa mạng nơ ron và mô hình Markov.

Mạng kết hợp giữa mạng nơ ron và mô hình Markov ẩn sử dụng mạng nơ ron được xây dựng ở bước 3. Khó khăn cơ bản nhất của việc áp dụng mô hình Markov là tính giá trị khởi đầu của tham số. Ý tưởng chính của việc kết hợp này là đầu ra của mạng nơ ron là xác suất của các đơn vị nhận dạng được sử dụng như xác suất phát xạ mẫu của các trạng thái Markov. Tập các xác suất của các khung tín hiệu của một phát âm (utterance) tạo thành ma

trận xác suất, trong đó các cột ma trận là các đơn vị nhận dạng, các hàng là các khung tín hiệu 10ms liên tiếp nhau của phát âm. Thuật toán Viterbi (xem phần 2.3 của bài báo) được áp dụng để nhận biết đường đi tối ưu (từ cần nhận dạng) trên ma trận các xác suất được đưa ra bởi mô hình mạng nơ ron.

Các hệ thống xây dựng ở các bước 1, 2, 3 sau khi đã được huấn luyện cho tiến hành nhận dạng trên cùng một tập dữ liệu thử gồm 63 câu. Tập dữ liệu thử này cũng là tập dùng để kiểm tra đối với mô hình Markov ẩn đã nói ở trên. Sau đây là kết quả nhận dạng với 194 mẫu huấn luyện.

Bảng 3. Độ chính xác nhận dạng dùng mạng nơ ron theo các bước.

Bước 1		Bước 2		Bước 3	
Từ	Câu	Từ	Câu	Từ	Câu
94.57%	71.74%	94.20%	71.74%	94.93%	73.91%

Kết quả ở bước 1 cho độ chính xác khá cao (94.57%,71.74%). Điều này chứng tỏ khả năng phân lớp tốt của mạng nơ ron. So với kết quả nhận dạng của mô hình Markov thì mạng nơ ron có độ chính xác cao hơn rất nhiều (94.57%,71.74%) so với (85.50%, 52.53%). Kết quả nhận dạng của bước 2 có phần giảm sút so với mạng tiến hành ở bước 1 (94.57%, 71.74%) so với (94.20%, 71.74%%). Nguyên nhân của hiện tượng này là do quá trình gán nhãn tự động bằng máy có độ chính xác thấp hơn so với bằng tay. Điều này cũng cho thấy độ chính xác của công việc gán nhẫn có ảnh hưởng đến độ chính xác của quá trình nhận dạng.

Mạng lai ghép giữa mạng nơ ron và mô hình Markov đã được nghiên cứu từ lâu để tận dụng hai ưu điểm của hai phương pháp: khả năng phân biệt lớp của mạng nơ ron và khả năng mô hình hoá cấu trúc thời gian của mô hình Markov. Các thực nghiệm mạng lai ghép NN - HMM trên thế giới cho thấy sự cải thiện đáng kể của hệ thống này so với các hệ thống chỉ dùng mạng nơ ron hay mô hình Markov. Thử nghiệm nhận dạng ở bước 3 cũng cho thấy độ chính xác nhận dạng của hệ thống lai ghép NN - HMM (94.93%, 73.91%) đã được nâng cao so với mạng nơ ron (94.57%, 71.74%) và mô hình Markov (85.50%, 52.53%). Bảng sau cho thấy số lượng lỗi nhận dạng nhầm giữa các thanh.

Bảng 4. Tỷ lệ lỗi của các thanh điệu trong nhận dạng bằng mạng nơ ron tại bước 3.

Thanh lỗi	Thanh bị nhận dạng nhầm						Tổng cộng
	Thanh sắc	Thanh huyền	Thanh hỏi	Thanh ngã	Thanh nặng	Thanh không	
Thanh sắc	-	0	0	0	0	1	1
Huyền	3	-	4	0	0	1	8
Hỏi	0	3	-	0	0	0	3
Ngã	0	0	0	-	0	0	0
Nặng	0	0	1	0	-	0	1
Không	3	1	0	0	0	-	4
Tổng cộng	6	4	5	0	0	2	17

Bảng 4 cho thấy tỷ lệ tỷ lệ nhận dạng nhầm của thanh sắc là khá cao (6 lỗi, 35%), thanh nặng và thanh ngã có tỷ lệ lỗi thấp (0%). Thanh huyền là thanh dễ bị nhận dạng nhầm với các thanh khác nhất (8 lỗi, 47%).

4. KẾT LUẬN

Bài báo này đã trình bày quá trình thực nghiệm nhận dạng một tập gồm các từ tiếng Việt có chung âm đầu, âm vần nhưng khác nhau về thanh điệu "na, ná, nà, nạ, nã, nả". Các phương pháp nhận dạng bao gồm mô hình Markov, mạng nơ ron ba lớp, và hệ thống lai ghép giữa mạng nơ ron và mô hình Markov ẩn NN-HMM. Kết quả nhận dạng cho thấy mạng nơ ron có khả năng phân biệt các thanh điệu tốt hơn mô hình Markov ẩn. Kết quả nhận dạng cho độ chính xác cao nhất với hệ thống lai ghép giữa mạng nơ ron và mô hình Markov ẩn 94.93% ở mức từ và 73.91% ở mức câu. Phân tích tỷ lệ lỗi cho thấy thanh sắc là thanh có tỷ lệ nhận dạng nhầm nhiều nhất (48% đối với HMM và 35% đối với NN-HMM). Thanh nặng là thanh có tỷ lệ lỗi nhận dạng nhầm thấp nhất (1%, đối với HMM và 0%, đối với NN-HMM). Thanh không cũng là thanh ít bị nhận dạng nhầm hơn các thanh khác (12%, đối với HMM và 11%, đối với NN-HMM). Tuy nhiên những kết quả trong bài này chỉ là những kết quả bước đầu, chúng tôi đang tiến hành những thử nghiệm trên cơ sở dữ liệu lớn hơn với các chữ số tiếng Việt được phát âm liên tục. Một trong những nghiên cứu chính tiếp theo là phải xác định được mô hình phiên âm của các âm vị và các từ trong tiếng Việt, song song với các thử nghiệm trên các mô hình nhận dạng, giữa mô hình kết hợp mạng nơ ron và mô hình Markov và bản thân mô hình Markov.

TÀI LIỆU THAM KHẢO

[1] Đỗ Xuân Thảo, Lê Hữu Tỉnh, *Giáo trình tiếng Việt 2*, Nhà xuất bản Giáo dục, 1997.

[2] Hosom JP., Cole R., Cosi P., *Improvement in Neural Network Trainning and Search Technique for Continuous Digit Recognition*, Center for Spoken Language Understanding, Oregon Graduate Institute, 1997.

[3] Hosom JP., Cole R., Fanty M., Schalkwyk J., Yan Y., Wei W., *Training Neural Networks for Speech Recognition*, Center for Spoken Language Understanding (CSLU), Oregon Graduate Institute of Science and Technology February 2, 1999.

[4] J. Schalkwyk , Hosom JP., Ed Kaiser, Khaldom Shobaki, *CSLU-HMM: The CSLU Hidden Markov Markov Modelling Environment*, Center for Spoken Language Understanding (CSLU), Oregon Graduate Institute of Science and Technology, 2000.

[5] Rabiner L., Juang B.H., *Fundamentals of Speech Recognition*, Prentice Hall, 1993.

[6] Yan Y., Fanty M., Cole R., *Speech Recognition using Neural Networks with Forward-Backward Probability Generated Tagets*, Proeedding of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1997.

Nhận bài ngày 20 - 5 - 2002