

SO SÁNH MỘT SỐ BỘ PHÂN LỚP DÙNG CHO NHẬN DẠNG PHƯƠNG NGỮ TIẾNG VIỆT

Nguyễn Hồng Quang², Phạm Ngọc Hưng^{1,2}, Trịnh Văn Loan^{1,2}, Phạm Quốc Hùng¹

¹ Khoa Công nghệ Thông tin, Trường Đại học Sư phạm Kỹ thuật Hưng Yên

² Viện Công nghệ Thông tin và Truyền thông, Trường Đại học Bách khoa Hà Nội

phamngochung@gmail.com, loantv@soict.hust.edu.vn, quangnh@soict.hust.edu.vn, quochungnu@gmail.com

TÓM TẮT— Tiếng Việt là ngôn ngữ có thanh điệu và có nhiều phương ngữ khác nhau. Ảnh hưởng của yếu tố phương ngữ tới các hệ thống nhận dạng tự động tiếng Việt nói là đáng kể. Có nhiều phương pháp khác nhau đã được nghiên cứu và áp dụng cho nhận dạng phương ngữ như GMM, SVM... Bài báo trình bày kết quả thử nghiệm nhận dạng phương ngữ tiếng Việt sử dụng công cụ Weka là tập hợp các thuật giải học máy dùng cho khai phá dữ liệu. Ngữ liệu dùng cho nhận dạng là các giọng Hà Nội, Huế, Thành phố Hồ Chí Minh đại diện cho phương ngữ của ba miền Bắc, Trung, Nam. Các bộ phân lớp SMO, LBK, Jrip, MultilayerPerceptron và PART đã được dùng cho thử nghiệm nhận dạng phương ngữ tiếng Việt. Kết quả thử nghiệm cho thấy tỷ lệ nhận dạng trung bình phương ngữ tiếng Việt cao nhất là 99,5% khi sử dụng bộ phân lớp MultilayerPerceptron. Việc đánh giá ảnh hưởng của tần số cơ bản đến hiệu năng nhận dạng cũng được thực hiện. Chỉ riêng thông tin tần số cơ bản đã cho phép kết quả nhận dạng phương ngữ đạt được 52,2%.

Từ khóa— SVM, nhận dạng phương ngữ, tiếng Việt, Weka, SMO, LBK, Jrip, multilayer perceptron, PART, tần số cơ bản.

I. GIỚI THIỆU

Tiếng Việt là ngôn ngữ có thanh điệu và đa dạng về phương ngữ [1]. Các phương ngữ khác nhau về từ địa phương và phương thức phát âm. Nghiên cứu nhận dạng tự động phương ngữ đóng vai trò quan trọng trong hệ thống nhận dạng tiếng nói cho các ngôn ngữ trong đó có tiếng Việt [2], [3]. Bài báo này sẽ trình bày các thử nghiệm nhận dạng phương ngữ tiếng Việt theo phương thức phát âm mà không phụ thuộc vào nội dung. Điều này cho phép thực hiện hệ thống nhận dạng phương ngữ linh hoạt. Ngữ liệu dùng cho nhận dạng là các giọng Hà Nội, Huế, Thành phố Hồ Chí Minh đại diện cho phương ngữ của ba miền Bắc, Trung, Nam. Có thể thấy rằng, quy luật biến thiên tần số cơ bản F0 đặc trưng cho các thanh điệu tiếng Việt đồng thời quy luật này cũng khác biệt đối với các phương ngữ tiếng Việt khác nhau. Vì vậy, việc khai thác thông tin F0 dùng làm đặc trưng cho hệ thống tự động nhận dạng phương ngữ cũng là yếu tố cần thiết. Trong số các mô hình được sử dụng cho các hệ thống xử lý tiếng nói như định danh ngôn ngữ, định danh người nói... SVM đã được áp dụng khá phổ biến và cho phép các hệ thống này đạt hiệu năng cao [4], [5], [6], [7], [8], [9]. SVM có thể được sử dụng độc lập hoặc kết hợp với GMM để nhận dạng phương ngữ [10], [11], [12], [13], [14], [15], [16]. Các phần tiếp theo của bài báo được tổ chức như sau: phần II trình bày về ngữ liệu dùng cho thử nghiệm nhận dạng phương ngữ tiếng Việt, phần III trình bày các thử nghiệm nhận dạng phương ngữ tiếng Việt sử dụng bộ công cụ Weka. Cuối cùng, phần IV là kết luận.

II. NGỮ LIỆU DÙNG CHO THỬ NGHIỆM

A. Bộ ngữ liệu VDSPEC

Để thực hiện các thử nghiệm, bộ ngữ liệu mới đã được nhóm tác giả tiến hành xây dựng và đặt tên là VDSPEC [17]. Bộ ngữ liệu này không chỉ đặc biệt dành cho nghiên cứu nhận dạng phương ngữ tiếng Việt nói riêng mà còn có thể dùng cho nghiên cứu nhận dạng tiếng Việt nói chung.

Bộ ngữ liệu VDSPEC được ghi âm trực tiếp từ người nói thông qua việc đọc các đoạn văn bản đã được chuẩn bị sẵn. Văn bản này được tổ chức theo các chủ đề khác nhau và cân bằng về thanh điệu (số lượng các từ cho mỗi thanh là xấp xỉ như nhau, khoảng 717 từ). Tiếng nói được ghi âm với tần số lấy mẫu là 16000 Hz, 16 bit cho mỗi mẫu. Độ tuổi của người nói trung bình là 21 tuổi. Ở độ tuổi này, tiếng nói đã ổn định và thể hiện rõ được tiếng địa phương. Mỗi phương ngữ có 50 người nói bao gồm 25 nữ và 25 nam. Giọng Hà Nội được chọn đại diện cho phương ngữ Bắc, giọng Huế cho phương ngữ Trung và giọng Thành phố Hồ Chí Minh đại diện cho phương ngữ Nam. Với mỗi chủ đề, người nói đọc 25 câu, mỗi câu có độ dài ghi âm khoảng 10 giây. Tổng thời gian tiếng nói đã ghi âm của VDSPEC là 45,12 giờ, chiếm dung lượng 4,84 GB bộ nhớ. Số liệu thống kê nội dung ghi âm bộ ngữ liệu VDSPEC được trình bày ở Bảng 1.

Bảng 1. Số liệu thống kê theo phương ngữ của bộ ngữ liệu VDSPEC.

Phương ngữ	Số câu	Thời gian (giờ)
Bắc	6250	14.41
Trung	6250	14.65
Nam	6250	16.06
Tổng	18750	45.12

Thử nghiệm nhận dạng phương ngữ được thực hiện theo cách đánh giá chéo. Với mỗi thử nghiệm, ngữ liệu được chia làm 10 phần. Một trong 10 phần đã chia được chọn cho thử nghiệm. Chín phần còn lại dùng cho huấn luyện. Sau đó, chọn phần tiếp theo để thử nghiệm và tiến hành như vậy với tổng số thử nghiệm cho mỗi phương pháp là 10 lần. Kết quả cuối cùng là trung bình 10 lần thử nghiệm.

B. Trích chọn đặc trưng

Dữ liệu dùng cho huấn luyện và thử nghiệm được xử lý, trích chọn đặc trưng bao gồm 384 hệ số do bộ công cụ OpenSMILE [18] thực hiện. Đây là dữ liệu thống kê của mỗi file ghi âm. Với mỗi file tiếng nói được trích chọn đặc trưng, OpenSMILE sẽ cho ra 384 hệ số. Quá trình tính toán các hệ số này được thực hiện như sau:

Đầu tiên file tiếng nói được phân tách thành một tập các khung với độ dài khung là 25 ms và độ dịch khung là 10ms. Với mỗi khung tiếng nói, tính 16 giá trị đặc trưng:

- Năng lượng khung
- 12 hệ số MFCC (Mel Frequency Cepstral Coefficients)
- Tỷ lệ biến thiên qua trục không (Zero-Crossing Rate)
- Xác suất âm hữu thanh
- Tần số cơ bản

Sau đó tính đạo hàm bậc một theo thời gian của 16 giá trị này, như vậy sẽ thu được 32 tham số. Vì các tham số được tính trên từng khung tiếng nói, như vậy số giá trị cho mỗi tham số chính bằng số khung phân tách được từ file tiếng nói ban đầu. Dựa trên từng tập giá trị của mỗi tham số, tính 12 giá trị thống kê như sau:

- Giá trị nhỏ nhất, giá trị lớn nhất, độ chênh lệch giữa giá trị lớn nhất và giá trị nhỏ nhất, thời điểm đạt giá trị lớn nhất, thời điểm đạt giá trị nhỏ nhất, trung bình số học trên tập giá trị.
- Độ dốc, độ lệch và lỗi trung bình bình phương của xấp xỉ tuyến tính
- Độ lệch chuẩn, skewness (mô men bậc 3), kurtosis (mô men bậc 4)

Như vậy với 1 file tiếng nói, số đặc trưng tính được sẽ bằng $32 \times 12 = 384$ đặc trưng.

III. CÁC THỬ NGHIỆM NHẬN DẠNG PHƯƠNG NGỮ TIẾNG VIỆT

A. Công cụ thử nghiệm

Công cụ dùng thử nghiệm là Weka. Weka gồm tập hợp các thuật giải học máy dùng cho khai phá dữ liệu do Đại học Waikato, New Zealand phát triển [19]. Dữ liệu đầu vào cho huấn luyện và thử nghiệm là file các tham số đặc trưng theo định dạng ARFF (Attribute-Relation File Format) [19].

B. Kết quả đạt được

1. Thử nghiệm sử dụng SMO cho nhận dạng phương ngữ tiếng Việt

SMO là thuật giải tối ưu hóa cực tiểu lần lượt dùng cho phân lớp véc-tơ hỗ trợ (Sequential Minimal Optimization algorithm for support vector classification) [20]. Việc thử nghiệm sử dụng SMO để nhận dạng phương ngữ được thực hiện theo các trường hợp như sau:

a) Thử nghiệm sử dụng SMO với các tham số đặc trưng đầy đủ 384 hệ số

Ngữ liệu phương ngữ dùng cho nhận dạng được chia theo phương pháp đánh giá chéo theo tỷ lệ 1:10. Kết quả thử nghiệm với SMO được cho ở ma trận sai nhầm trên Bảng 2. Số câu nhận dạng đúng phương ngữ và nhận dạng nhầm phương ngữ được thể hiện trong bảng.

Bảng 2. Ma trận sai nhầm đối với thử nghiệm dùng SMO nhận dạng phương ngữ tiếng Việt bao gồm đầy đủ tham số.

Phương ngữ	Bắc	Trung	Nam
Bắc	6041	127	80
Trung	165	6010	73
Nam	62	82	6046

Bảng 2 cho thấy kết quả thử nghiệm có tỷ lệ nhận dạng đúng thấp nhất là 96,2% đối với phương ngữ Trung, cao nhất là 97,7% đối với phương ngữ Nam. Tỷ lệ trung bình của 3 phương ngữ là 96,9%.

b) Thử nghiệm SMO trong trường hợp không có thông tin liên quan trực tiếp F0

Với thử nghiệm này, 12 hệ số đặc trưng cho F0 được loại khỏi tập tham số đặc trưng. Kết quả thử nghiệm được cho trên ma trận sai nhầm ở Bảng 3.

Bảng 3 cho thấy kết quả thử nghiệm có tỷ lệ nhận dạng đúng thấp nhất là 96% đối với phương ngữ Trung, cao nhất là 97,6% đối với phương ngữ Nam. Tỷ lệ trung bình của 3 phương ngữ là 96,7%. Kết quả này thấp hơn so với trường hợp có sử dụng F0 ở thử nghiệm trước.

Bảng 3. Ma trận sai nhầm đối với thử nghiệm dùng SMO nhận dạng phương ngữ tiếng Việt không có tham số liên quan trực tiếp F0.

Phương ngữ	Bắc	Trung	Nam
Bắc	6030	127	91
Trung	170	5998	80
Nam	65	85	6040

c) Thử nghiệm sử dụng SMO trong trường hợp chỉ dùng các tham số liên quan trực tiếp F0

Trong thử nghiệm này, chỉ chọn 12 tham số đặc trưng liên quan trực tiếp F0 để huấn luyện và nhận dạng. Bảng 4 là ma trận sai nhầm cho kết quả đã thử nghiệm.

Bảng 4. Ma trận sai nhầm đối với thử nghiệm dùng SMO nhận dạng phương ngữ tiếng Việt chỉ sử dụng tham số liên quan trực tiếp F0.

Phương ngữ	Bắc	Trung	Nam
Bắc	3650	1517	1081
Trung	2662	2614	972
Nam	1590	1112	3488

Bảng 4 cho thấy tỷ lệ nhận dạng đúng trung bình của 3 phương ngữ là 52,2%. Tỷ lệ nhận dạng đúng thấp nhất là 41,8% đối với phương ngữ Trung, cao nhất là 58,4% cho phương ngữ Bắc. Kết quả này cho thấy thông tin F0 giúp ích tốt cho nhận dạng phương ngữ.

d) Thử nghiệm sử dụng SMO chỉ dùng tham số đặc trưng của MFCC

Trong thử nghiệm này, chỉ chọn 12 tham số liên quan trực tiếp với MFCC. Bảng 5 là ma trận sai nhầm của kết quả thử nghiệm tương ứng.

Bảng 5. Ma trận sai nhầm đối với thử nghiệm dùng SMO nhận dạng phương ngữ tiếng Việt chỉ sử dụng tham số liên quan trực tiếp MFCC.

Phương ngữ	Bắc	Trung	Nam
Bắc	4712	974	562
Trung	1048	4413	787
Nam	404	745	5041

Bảng 5 cho thấy tỷ lệ nhận dạng đúng thấp nhất là 70,6% đối với phương ngữ Trung, cao nhất là 81,4% đối với phương ngữ Nam. Tỷ lệ trung bình là 75,8% cho 3 phương ngữ. Kết quả này cho thấy các tham số đặc trưng cho MFCC cũng đóng vai trò khá quan trọng trong nhận dạng phương ngữ.

2. Thử nghiệm sử dụng IBK cho nhận dạng phương ngữ tiếng Việt

IBK là bộ phân lớp *k* láng giềng gần nhất (Lazy *k*-nearest-neighbor classifier) [19].

Trong trường hợp này, toàn bộ các tham số đã trích chọn đặc trưng gồm 384 hệ số đã được sử dụng. Bảng 6 là ma trận sai nhầm tương ứng.

Bảng 6. Ma trận sai nhầm đối với thử nghiệm nhận dạng phương ngữ tiếng Việt dùng IBK.

Phương ngữ	Bắc	Trung	Nam
Bắc	6203	27	18
Trung	42	6195	11
Nam	51	18	6121

Bảng 6 cho thấy kết quả thử nghiệm có tỷ lệ nhận dạng đúng thấp nhất là 98,9% cho phương ngữ Nam, cao nhất là 99,3% cho phương ngữ Bắc. Trung bình cho 3 phương ngữ là 99,1%.

3. Thử nghiệm nhận dạng phương ngữ tiếng Việt sử dụng MultilayerPerceptron

Với Weka, MultilayerPerceptron là bộ phân lớp sử dụng mạng nơ-ron lan truyền ngược để huấn luyện.

Thử nghiệm này cũng sử dụng toàn bộ các đặc trưng đã được trích chọn. Mạng nơ-ron có cấu hình như sau:

- Lớp đầu vào có 384 nơ-ron tương ứng với 384 đặc trưng tiếng nói.
- Lớp đầu ra có 3 nơ-ron tương ứng với 3 phương ngữ cần nhận dạng.
- Lớp ẩn có số nơ-ron = 194, là trung bình cộng của số nơ-ron của lớp đầu vào và lớp đầu ra. Đây cũng là giá trị cho kết quả nhận dạng tốt nhất trong thử nghiệm. Kết quả được cho trên ma trận sai nhầm ở Bảng 7.

Bảng 7. Ma trận sai nhầm đối với thử nghiệm nhận dạng phương ngữ tiếng Việt dùng MultilayerPerceptron.

Phương ngữ	Bắc	Trung	Nam
Bắc	6221	12	15
Trung	28	6210	10
Nam	18	5	6167

Bảng 7 cho thấy phương ngữ Trung có tỷ lệ nhận dạng đúng thấp nhất là 99,4%, còn phương ngữ Nam có tỷ lệ nhận dạng cao nhất là 99,6%. Trung bình cả 3 phương ngữ có tỷ lệ nhận dạng là 99,5%.

4. Thử nghiệm nhận dạng phương ngữ tiếng Việt sử dụng Jrip

Jrip là thuật giải RIPPER để suy diễn luật một cách hiệu quả và nhanh (RIPPER repeated incremental pruning to produce error reduction) [21].

Toàn bộ các đặc trưng gồm 384 hệ số đã được sử dụng cho thử nghiệm. Bảng 8 là ma trận sai nhầm của trường hợp này.

Bảng 8. Ma trận sai nhầm đối với thử nghiệm nhận dạng phương ngữ tiếng Việt dùng Jrip.

Phương ngữ	Bắc	Trung	Nam
Bắc	5725	264	259
Trung	246	5789	213
Nam	145	128	5917

Từ Bảng 8 có thể thấy tỷ lệ nhận dạng đúng thấp nhất là 91,6% đối với phương ngữ Bắc, cao nhất là 95,6% đối với phương ngữ Nam. Tỷ lệ trung bình của 3 phương ngữ là 93,3%.

5. Thử nghiệm nhận dạng phương ngữ tiếng Việt sử dụng PART

Bộ phân lớp PART được dùng để có các luật từ các cây quyết định riêng phần đã được xây dựng bằng cách sử dụng J4.8. J4.8 là cài đặt mã nguồn mở Java của thuật giải C4.5 và thuật giải này được dùng để tạo cây quyết định do Ross Quinlan phát triển [22].

Thử nghiệm này cũng sử dụng toàn bộ các đặc trưng đã được trích chọn. Bảng 9 là ma trận sai nhầm tương ứng.

Bảng 9. Ma trận sai nhầm đối với thử nghiệm nhận dạng phương ngữ tiếng Việt dùng PART.

Phương ngữ	Bắc	Trung	Nam
Bắc	2720	1536	1992
Trung	2026	3359	863
Nam	943	460	4787

Bảng 9 cho thấy tỷ lệ nhận dạng đúng thấp nhất là 92,5% đối với phương ngữ Bắc, cao nhất là 93,7% đối với phương ngữ Nam. Tỷ lệ trung bình của 3 phương ngữ đạt 93%.

C. Tổng hợp và nhận xét kết quả thử nghiệm nhận dạng

Bảng 10 là kết quả thử nghiệm nhận dạng phương ngữ tiếng Việt sử dụng các phương pháp phân lớp khác nhau đã nêu trên.

Bảng 10. Tỷ lệ nhận dạng đúng đối với các thử nghiệm nhận dạng phương ngữ tiếng Việt.

Phương pháp	Tỷ lệ nhận dạng đúng từng phương ngữ			Trung bình ba phương ngữ
	Bắc	Trung	Nam	
MultilayerPerceptron	99,6%	99,4%	99,6%	99,5%
LBK	99,3%	99,2%	98,9%	99,1%
SMO	96,7%	96,2%	97,7%	96,9%
Jrip	91,6%	92,7%	95,6%	93,3%
PART	92,5%	92,9%	93,7%	93,0%

Trong Bảng 10, các kết quả nhận dạng đúng đã được sắp xếp theo thứ tự từ cao xuống thấp. Bảng này cho thấy phương pháp MultilayerPerceptron cho kết quả nhận dạng đúng cao nhất (99,5%). Còn lại, các phương pháp IBK, SMO, Jrip, PART cũng đều cho kết quả nhận dạng đúng khá cao, trên 93%. Với thử nghiệm dùng phân lớp SMO, tham số F0 đóng vai trò đáng kể khi nhận dạng phương ngữ tiếng Việt. Khi có tham số F0, tỷ lệ nhận dạng đúng được nâng lên. Ngay cả khi chỉ sử dụng tham số F0 cho nhận dạng, kết quả nhận dạng đúng trung bình cho 3 phương ngữ cũng đạt trên 52%. Kết quả này cũng tương đồng với nhận xét F0 đóng vai trò quan trọng trong các ngôn ngữ có thanh điệu mà tiếng Việt là một trường hợp.

IV. KẾT LUẬN

Bài báo đã trình bày các kết quả thử nghiệm nhận dạng cho ba phương ngữ tiếng Việt sử dụng các bộ phân lớp SMO, IBK, MultilayerPerceptron, Jrip, PART. Kết quả cho tỷ lệ nhận dạng đúng đạt khá cao, trung bình đều trên 93%. Đặc biệt, phân lớp dùng MultilayerPerceptron cho kết quả nhận dạng cao nhất là 99,5%. Thử nghiệm cũng góp phần củng cố khẳng định các tham số đặc trưng liên quan trực tiếp đến F0 cải thiện hiệu năng của hệ thống nhận dạng tiếng Việt nói chung. Bộ công cụ Weka sử dụng các tham số đặc trưng do OpenSMILE trích rút cho phép thực hiện nhận dạng phương ngữ tiếng Việt với bộ tham số khá phong phú và tùy biến để nghiên cứu ảnh hưởng của các tham số này đến hiệu quả nhận dạng. Các thử nghiệm đã được thực hiện cho thấy các bộ phân lớp sử dụng trong trường hợp này tỏ ra thích hợp cho các trường hợp nhận dạng các đường bao thông tin trong tiếng nói như định danh phương ngữ, định

danh người nói, định danh ngôn ngữ... Hướng nghiên cứu tiếp theo là kết hợp các bộ phân lớp đã thử nghiệm với các mô hình nhận dạng khác để có thể xây dựng hệ thống nhận dạng tiếng Việt trong đó có nhận dạng phương ngữ đạt hiệu quả tốt.

V. LỜI CẢM ƠN

Bài báo này được thực hiện trong khuôn khổ đề tài nghiên cứu khoa học cấp trường “Nghiên cứu xây dựng hệ thống nhận dạng phương ngữ tiếng Việt sử dụng phương pháp học sâu” của Trường Đại học Bách khoa Hà Nội. Các tác giả chân thành cảm ơn Trường Đại học Bách khoa Hà Nội, Phòng Khoa học Công nghệ, Viện Công nghệ Thông tin và Truyền thông đã hỗ trợ để chúng tôi có thể thực hiện thành công đề tài.

TÀI LIỆU THAM KHẢO

- [1] Hoàng Thị Châu. Phương ngữ học tiếng Việt. NXB Đại học Quốc gia Hà Nội, 2009.
- [2] Phạm Ngọc Hưng, Trịnh Văn Loan, Nguyễn Hồng Quang, Phạm Quốc Hùng. "Nhận dạng phương ngữ tiếng Việt sử dụng mô hình Gauss hỗn hợp", Kỷ yếu Hội nghị Khoa học Công nghệ Quốc gia lần thứ 6 FAIR, 20-21 tháng 6, 2014, ISBN 978-604-913-165-3, pp 449-452, 2014.
- [3] Phạm Ngọc Hưng, Trịnh Văn Loan, Nguyễn Hồng Quang. “Nhận dạng phương ngữ tiếng Việt sử dụng MFCC và tần số cơ bản”. Kỷ yếu Hội nghị Quốc gia lần thứ VIII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR) – Hà Nội, 09-10/7/2015, ISBN: 978-604-913-397-8, trang 523-528, 2015.
- [4] Campbell, W. M., Singer, E., Torres-Carrasquillo, P. A., and Reynolds, D. A., “Language Recognition with Support Vector Machines”. In Proc. Odyssey: The Speaker and Language Recognition Workshop in Toledo, Spain, ISCA, pp. 41-44, 31 May - 3 June 2004.
- [5] Abe, Shigeo. “Support vector machines for pattern classification”. Vol. 53. London: Springer, 2005.
- [6] Shady, Y., and Sharway-Hala H. Zayed. "Speaker independent Arabic speech recognition using support vector machine." 2009.
- [7] Hou, Jue, et al, "Multi-layered features with SVM for Chinese accent identification." Audio Language and Image Processing (ICALIP), 2010 International Conference on. IEEE, 2010.
- [8] Richardson, Fred, and William M. Campbell. "Discriminative keyword selection using support vector machines." Advances in Neural Information Processing Systems. 2008.
- [9] Hanani, Abualsoud, Martin J. Russell, and Michael J. Carey. "Human and computer recognition of regional accents and ethnic groups from British English speech." Computer Speech & Language Journal 27.1 (2013): 59-74.
- [10] Hirschberg, Julia Bell, Fadi Biadisy, and Michael Collins. "Dialect Recognition Using a Phone-GMM-Supervector-Based SVM Kernel." 2010.
- [11] Boril, Hynek, Abhijeet Sangwan, and John HL Hansen. "Arabic Dialect Identification-Is the Secret in the Silence?" and Other Observations." INTERSPEECH. 2012.
- [12] Akbacak, Murat, et al. "Effective Arabic Dialect Classification Using Diverse Phonotactic Models." INTERSPEECH. Vol. 11. 2011.
- [13] Brown G. “Moving towards automatic accent recognition for forensic applications.” INTERSPEECH 2015. Dresden, Germany. 6th Sept, 2015.
- [14] Pedersen, Carol, and Joachim Diederich. "Accent classification using support vector machines." 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007). IEEE, 2007.
- [15] Nour-Eddine, Lachachi, and Adla Abdelkader. "GMM-Based Maghreb Dialect Identification System." JIPS 11.1 (2015): 22-38.
- [16] Biadisy, Fadi, et al. "Discriminative Phonotactics for Dialect Recognition Using Context-Dependent Phone Classifiers." Odyssey. 2010.
- [17] Phạm Ngọc Hưng, Trịnh Văn Loan, Nguyễn Hồng Quang, “Building of corpus for Vietnamese dialect identification”, Journal of Science and Technology Technical Universities, No.109-2015. ISSN 2354-1083, pp.49-55, 2015.
- [18] Eyben, Florian, Martin Wöllmer, and Björn Schuller. "Opensmile: the munich versatile and fast open-source audio feature extractor." Proceedings of the 18th ACM international conference on Multimedia. ACM, 2010.
- [19] Witten, Ian H., and Eibe Frank. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2005.
- [20] John C. Platt, Microsoft Research, jplatt@microsoft.com, Technical Report MSR-TR-98-14, April 21, 1998
- [21] William W. Cohen: Fast Effective Rule Induction. In: Twelfth International Conference on Machine Learning, 115-123, 1995
- [22] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

COMPARISON OF SOME CLASSIFIERS FOR VIETNAMESE DIALECT RECOGNITION

Nguyễn Hồng Quang, Phạm Ngọc Hưng, Trịnh Văn Loan, Phạm Quốc Hùng

ABSTRACT— Vietnamese is a tonal language with many different dialects. The influence of dialectal features on Vietnamese speech recognition systems is significant. There are many different methods which have been studied and applied for dialect recognition such as GMM, SVM ... This paper presents the experimental results of Vietnamese dialect identification using Weka which is a collection of machine learning algorithms for data mining. The corpus used for identification contain the voices of Hanoi, Hue and Ho Chi Minh City considered as the representable voices for Northern, Central and Southern dialects. The classifiers SMO, IBK, Jrip, MultilayerPerceptron, and PART have been used for experiments. Test results showed that the highest average score is 99.5% for MultilayerPerceptron classifier. The influence of fundamental frequency F0 on recognition performance is also evaluated. The score of dialect recognition gets 52.2% using only F0 information.