

## NHẬN DẠNG CHỮ VIẾT: NHỮNG THÀNH TỰU, THÁCH THỨC VÀ HƯỚNG TIẾP CẬN

Nhận bài:

22 – 05 – 2015

Chấp nhận đăng:

25 – 09 – 2015

<http://jshe.ued.udn.vn/>

Phạm Anh Phương

**Tóm tắt:** Trong lĩnh vực nhận dạng thì nhận dạng chữ đang ngày càng có nhiều ứng dụng trong đời sống xã hội. Cho đến nay, bài toán nhận dạng chữ in đã được giải quyết gần như trọn vẹn (sản phẩm FineReader 12.0 của hãng ABBYY có thể nhận dạng chữ in theo 20 ngôn ngữ khác nhau, phần mềm nhận dạng chữ Việt in VnDOCR 4.0 của Viện Công nghệ Thông tin Hà Nội có thể nhận dạng được các tài liệu chứa hình ảnh, bảng và văn bản với độ chính xác trên 98%). Tuy nhiên, trên thế giới cũng như ở Việt Nam thì bài toán nhận dạng chữ viết tay vẫn còn là vấn đề thách thức lớn đối với các nhà nghiên cứu. Bài báo này sẽ tổng hợp những thành quả đạt được và những tồn tại, thách thức hiện nay trong lĩnh vực nhận dạng chữ viết đồng thời nêu lên những hướng tiếp cận mới cho hướng nghiên cứu này.

**Từ khóa:** Nhận dạng chữ in; nhận dạng chữ viết tay; OCR.

### 1. Giới thiệu

Nhận dạng chữ là lĩnh vực được nhiều nhà nghiên cứu trong và ngoài nước quan tâm 000. Cho đến nay, lĩnh vực này cũng đã đạt được nhiều thành tựu lớn lao cả về mặt lý thuyết lẫn ứng dụng thực tế. Lĩnh vực nhận dạng chữ được chia làm hai loại: Nhận dạng chữ in và nhận dạng chữ viết tay.

Đến thời điểm này, công nghệ nhận dạng chữ in đã đạt được những giải pháp tốt để ứng dụng vào các sản phẩm thương mại. Tuy nhiên, nhận dạng chữ viết tay vẫn còn là vấn đề thách thức lớn đối với các nhà nghiên cứu. Nhận dạng chữ viết tay được phân ra làm hai loại: nhận dạng chữ viết tay on-line và nhận dạng chữ viết tay off-line.

Nhận dạng chữ viết tay on-line được thực hiện trên cơ sở lưu lại các thông tin về nét chữ như thứ tự nét viết, hướng và tốc độ của nét trong quá trình viết. Đối với

nhận dạng chữ viết tay off-line, dữ liệu đầu vào là ảnh văn bản nên việc nhận dạng có độ khó cao hơn so với nhận dạng chữ viết tay on-line. Do dữ liệu đầu vào là ảnh văn bản nên nhận dạng chữ viết tay off-line và nhận dạng chữ in còn được gọi chung là **nhận dạng chữ quang học (OCR - Optical Character Recognition)**.

Khó khăn lớn nhất khi nghiên cứu bài toán nhận dạng chữ viết tay là sự biến thiên quá đa dạng trong cách viết của từng người. Điều này gây khó khăn trong việc trích chọn đặc trưng cũng như lựa chọn mô hình nhận dạng. Vì vậy để nghiên cứu về lĩnh vực nhận dạng chữ viết tay, cần phải có một khối lượng kiến thức tương đối rộng liên quan đến nhiều lĩnh vực khác nhau. Sau đây là một số lĩnh vực có liên hệ chặt chẽ đối với nhận dạng chữ viết tay:

Xử lý ảnh (Image Processing): được sử dụng trong các giai đoạn tiền xử lý, tách chữ và trích chọn đặc trưng.

Học máy (Machine Learning): được sử dụng trong giai đoạn huấn luyện và nhận dạng, chẳng hạn như các mạng nơ ron nhân tạo, SVM,...

Lý thuyết nhận dạng (Pattern Recognition): sử dụng các phương pháp luận phân lớp sử dụng trong công đoạn huấn luyện và nhận dạng.

\* Liên hệ tác giả

Phạm Anh Phương

Trường Đại học Sư phạm, Đại học Đà Nẵng

Email: paphuong@yahoo.com

Xác suất thống kê và toán ứng dụng: lý thuyết xác suất đóng vai trò rất quan trọng trong các phương pháp phân lớp thống kê như mô hình Markov ẩn, phương pháp Bayes, k-láng giềng gần nhất, SVM...

Ngôn ngữ học và ngôn ngữ học tính toán (Linguistic and Computational Linguistic): Các kiến thức về ngữ pháp đóng vai trò quan trọng trong công đoạn hậu xử lý, nâng cao độ chính xác cho các hệ thống nhận dạng chữ viết.

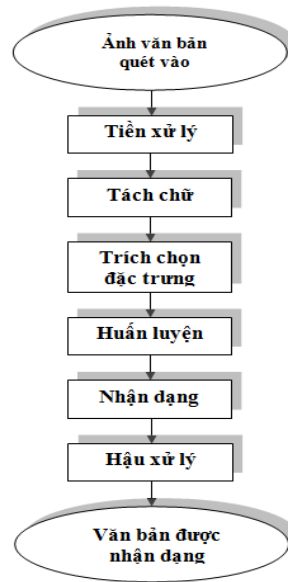
Phần còn lại của bài báo này sẽ được cấu trúc như sau: phần 2 giới thiệu các giai đoạn cơ bản của một hệ nhận dạng chữ viết; phần 3 giới thiệu khái quát một số hướng nghiên cứu về trích chọn đặc trưng; phần 4 trình bày một số phương pháp nhận dạng đang được áp dụng rộng rãi trong các hệ nhận dạng chữ viết; phần 5 thảo luận về tình hình nghiên cứu nhận dạng chữ viết, những tồn tại và thách thức đối với các nhà nghiên cứu. Cuối cùng là phần kết luận với một số hướng nghiên cứu đề xuất.

## 2. Các giai đoạn cơ bản của một hệ nhận dạng chữ viết

Một hệ nhận dạng chữ viết bao gồm năm giai đoạn chính sau đây (Hình 1).

### 2.1. Tiền xử lý

Giai đoạn này góp phần làm tăng độ chính xác phân lớp của hệ thống nhận dạng, tuy nhiên nó cũng làm cho tốc độ nhận dạng của hệ thống chậm lại. Vì vậy, tùy thuộc vào chất lượng ảnh quét vào của từng văn bản cụ thể để chọn một hoặc một vài chức năng trong khối này. Nếu cần ưu tiên tốc độ xử lý và chất lượng của máy quét tốt thì có thể bỏ qua giai đoạn này. Giai đoạn tiền xử lý bao gồm một số chức năng:



Hình 1. Sơ đồ tổng quát của một hệ thống nhận dạng chữ viết

### Nhị phân hóa ảnh



Hình 2. Nhị phân hóa ảnh

### Lọc nhiễu

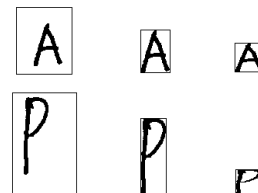


Hình 3. Nhiễu đốm và nhiễu vết

Ảnh khi quét vào thường gặp một số loại nhiễu phổ biến như: nhiễu đốm, nhiễu vết, nhiễu đứt nét... (Hình 3).

### Chuẩn hóa kích thước ảnh

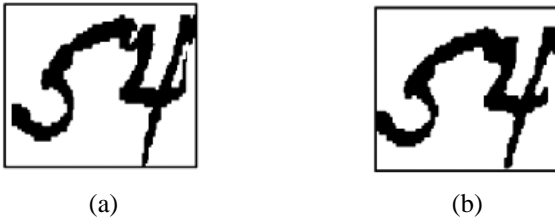
Việc chuẩn hóa kích thước ảnh dựa trên việc xác định trọng tâm ảnh, sau đó xác định khoảng cách lớn nhất từ tâm ảnh đến các cạnh trên, dưới, trái, phải của hình chữ nhật bao quanh ảnh.



Hình 4. Chuẩn hóa các ảnh ký tự "A" và "P" về kích thước cố định

**Làm trơn biên chữ**

Khi lựa chọn đường biên của chữ làm đặc trung để nhận dạng, nếu chất lượng quét ảnh xấu thì các đường biên của chữ sẽ không giữ được dáng điệu trơn tru ban đầu mà hình thành các đường răng cưa giả tạo. Trong các trường hợp này, cần dùng các thuật toán làm trơn biên để khắc phục 0.

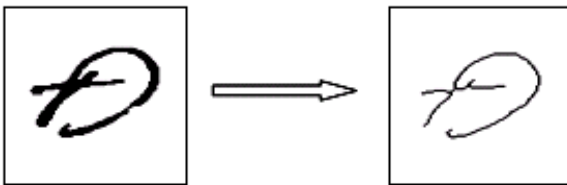


**Hình 5.** (a) Ảnh gốc, (b) Ảnh sau khi được làm trơn biên

**Làm đầy chữ**

Chức năng này được áp dụng với các ký tự bị đứt nét một cách ngẫu nhiên. Ảnh đứt nét gây khó khăn cho việc tách chữ, dễ bị nhầm hai phần liên thông của ký tự thành hai ký tự riêng biệt, tạo nên sai lầm trong quá trình nhận dạng.

**Làm mảnh chữ**

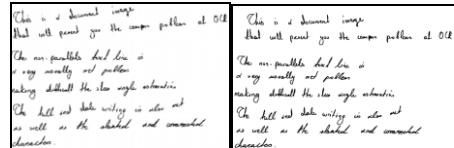


**Hình 6.** Làm mảnh chữ

Đây là một bước quan trọng nhằm phát hiện khung xương của ký tự bằng cách loại bỏ dần các điểm biên ngoài của các nét. Tuy nhiên, quá trình làm mảnh chữ rất nhạy cảm với việc khử nhiễu. Hiện nay có nhiều phương pháp làm mảnh chữ, các thuật toán tìm xương có thể tham khảo ở 0.

**Điều chỉnh độ nghiêng của văn bản**

Do trang tài liệu quét vào không cẩn thận hoặc do sự cố in ấn, các hàng chữ bị lệch so với lề chuẩn một góc  $\alpha$ , điều này gây khó khăn cho công đoạn tách chữ, đôi khi không thể tách được. Trong những trường hợp như vậy, phải tính lại tọa độ điểm ảnh của các chữ bị sai lệch.



**Hình 7.** Hiệu chỉnh độ nghiêng của văn bản

Có nhiều kỹ thuật để điều chỉnh độ nghiêng, kỹ thuật phổ biến nhất dựa trên cơ sở biểu đồ chiếu (projection profile) của ảnh tài liệu; một số kỹ thuật dựa trên cơ sở các phép biến đổi Hough và Fourier; một số kỹ thuật hiệu chỉnh độ nghiêng khác có thể tìm thấy trong 0.

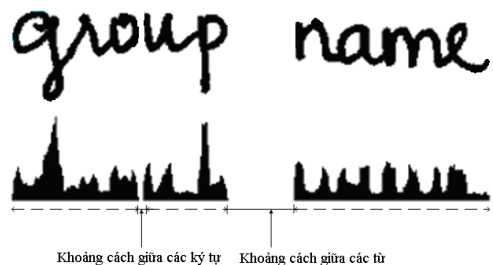
**2.2. Tách chữ**



**Hình 8.** Tách dòng chữ dựa trên histogram theo chiều ngang của khối chữ

Khối này có nhiệm vụ tách từng ký tự ra khỏi văn bản. Chỉ khi nào văn bản được tách và cô lập đúng từng ký tự đơn ra khỏi tổng thể văn bản thì hệ thống mới có thể nhận dạng đúng ký tự đó. Phương pháp tách chữ dùng lược đồ độ sáng được sử dụng khá phổ biến.

Đối với chữ viết tay thì việc tìm đường phân cách giữa các dòng và các ký tự trong văn bản thường rất khó khăn. Khi đó phải xây dựng lược đồ sáng của các dòng chữ, từ đó các đoạn thấp nhất trên lược đồ chính là đường phân cách cần tìm (Hình 8 và 9).



**Hình 9.** Xác định khoảng cách giữa hai kí tự và giữa hai từ dựa trên histogram theo chiều thẳng đứng của dòng chữ

### 2.3. Trích chọn đặc trưng

Mục đích của việc trích chọn đặc trưng là lựa chọn các thuộc tính của các mẫu để xây dựng độ đo về sự khác biệt giữa các lớp mẫu phục vụ trong giai đoạn huấn luyện phân lớp và nhận dạng.

Trích chọn đặc trưng đóng vai trò quan trọng trong một hệ thống nhận dạng. Cho đến nay, đã tồn tại nhiều hướng tiếp cận trích chọn đặc trưng, có thể phân thành các nhóm sau:

Chuẩn hóa ảnh chữ và đối sánh sơ cấp: ảnh chữ cần được chuẩn hóa về kích cỡ, vị trí để có thể đối sánh với các chữ đã được lưu sẵn. Hướng tiếp cận này có thể áp dụng cho việc nhận dạng chữ in với các kiểu chữ cố định, tuy nhiên rất khó áp dụng đối với chữ viết tay.

Biểu diễn ảnh chữ và đối sánh thứ cấp: với các phép biến đổi khác nhau, biểu diễn ảnh ban đầu của chữ được chuyển sang biểu diễn mới ít bị ảnh hưởng bởi nhiễu và tương đối bất biến đối với kích cỡ, vị trí của chữ. Quá trình đối sánh các biểu diễn mới của ảnh được gọi là đối sánh thứ cấp. Nhược điểm của hướng tiếp cận này là độ phức tạp của thuật toán lớn, ảnh hưởng đến tốc độ nhận dạng.

Trích chọn dấu hiệu đặc tả chữ và đối sánh cấu trúc: đây là hướng tiếp cận có nhiều triển vọng để xây dựng các hệ nhận dạng chữ viết tay. Tuy nhiên, các thuật toán trích chọn dấu hiệu đặc tả rất nhạy cảm với nhiễu.

### 2.4. Huấn luyện

Huấn luyện là giai đoạn quan trọng, quyết định đến chất lượng của hệ thống nhận dạng. Giai đoạn này chiếm khá nhiều thời gian, tùy thuộc vào từng phương pháp huấn luyện cũng như số lượng mẫu tham gia huấn luyện. Kết quả sau khi huấn luyện sẽ được lưu lại để phục vụ cho giai đoạn nhận dạng.

### 2.5. Nhận dạng

Giai đoạn nhận dạng riêng từng ký tự là giai đoạn quan trọng nhất, quyết định độ chính xác của hệ thống nhận dạng. Giai đoạn này sử dụng bộ tham số thu được từ giai đoạn huấn luyện để xác định phân lớp cho các mẫu cần nhận dạng. Chất lượng nhận dạng trong giai đoạn này phụ thuộc vào kết quả thu được trong giai đoạn huấn luyện.

### 2.6. Hậu xử lý

Đây là công đoạn cuối cùng của quá trình nhận dạng. Có thể hiểu hậu xử lý là bước ghép nối các ký tự

đã nhận dạng thành các từ, các câu, các đoạn văn nhằm tái hiện lại văn bản đồng thời phát hiện ra các lỗi nhận dạng sai bằng cách kiểm tra chính tả dựa trên cấu trúc và ngữ nghĩa của các từ, các câu hoặc các đoạn văn. Việc phát hiện ra các lỗi, các sai sót trong nhận dạng ở bước này góp phần đáng kể vào việc nâng cao chất lượng nhận dạng 00.

Mô hình ngôn ngữ thống kê N-Grams 0 đã được áp dụng khá thành công trong việc kiểm tra chính tả ở giai đoạn hậu xử lý của các hệ thống nhận dạng chữ viết cũng như các hệ thống nhận dạng tiếng nói. Mục đích của mô hình ngôn ngữ N-Gram là tìm ra xác suất của một từ theo sau một số lượng từ nào đó trong một cụm từ hoặc một câu.

## 3. Các phương pháp trích chọn đặc trưng

Có nhiều phương pháp trích chọn đặc trưng cho ảnh văn bản, nhưng chung quy lại, các phương pháp này có thể gom lại thành ba nhóm chính sau:

### 3.1. Biến đổi toàn cục và khai triển chuỗi

Một tín hiệu liên tục thường chứa nhiều thông tin và có thể sử dụng để làm đặc trưng cho mục đích phân lớp. Các đặc trưng này cũng có thể được trích chọn bằng cách xấp xỉ các tín hiệu liên tục thành các tín hiệu rời rạc. Sau đây là một số phép biến đổi và khai triển chuỗi dùng để biểu diễn ảnh thường được áp dụng trong lĩnh vực nhận dạng chữ: Biến đổi Fourier 0[16] 0, Biến đổi Wavelet 0 0, Phương pháp mô men: Theo phương pháp này, ảnh gốc sẽ được thay thế bằng một tập các đặc trưng vừa đủ để biểu diễn các đối tượng bất biến đối với các phép thay đổi tỷ lệ, tịnh tiến hoặc quay 0, Khai triển Karhunen-Loeve 00.

### 3.2. Đặc trưng thống kê

Các đặc trưng thống kê của ảnh văn bản bảo toàn các kiểu biến đổi đa dạng về hình dáng của chữ. Mặc dù các kiểu đặc trưng này không thể xây dựng lại ảnh gốc, nhưng nó được sử dụng để thu nhỏ số chiều của tập đặc trưng nhằm tăng tốc độ và giảm thiểu độ phức tạp tính toán. Sau đây là một số đặc trưng thống kê thường dùng để biểu diễn ảnh ký tự:

**Phân vùng (zoning):** Trong những năm gần đây, nhiều công trình nghiên cứu trong nước và quốc tế đã áp dụng các đặc trưng này 0000 vào các bài toán OCR.

**Các giao điểm và khoảng cách:** Một đặc trưng thống kê phổ biến là số giao điểm giữa chu tuyến của chữ với một đường thẳng theo một hướng đặc biệt nào đó, các giao điểm này không bị ảnh hưởng bởi việc mất mát điểm ảnh ở biên chữ. Các đặc trưng này từng được G. Vamvakas và các cộng sự áp dụng để phân nhóm sơ bộ các lớp ký tự hệ La Tinh, Hy Lạp 0.

Trương tự, khoảng cách từ biên của khung chứa ảnh tới điểm đen đầu tiên của chu tuyến chữ trên cùng một dòng quét cũng được sử dụng như những đặc trưng thống kê 000. Đặc trưng này có thể mô tả hình dạng khái quát của chữ, tuy nhiên nó cũng rất nhạy cảm với nhiễu và độ nghiêng của chữ.

**Các phép chiếu điểm ảnh:** Các ký tự có thể được biểu diễn bằng cách chiếu các điểm ảnh lên các dòng theo các hướng khác nhau. Các đặc trưng này ít nhạy cảm với nhiễu. Tuy nhiên, để sử dụng tốt lại đặc trưng này thì cần chuẩn hóa độ dày của nét chữ. Đây cũng là loại đặc trưng được nhiều nhà nghiên cứu trong và ngoài nước sử dụng rộng rãi trong các hệ thống OCR 0000.

**Đặc trưng hướng:** Các ký tự bao gồm các nét chữ, các nét này là các đoạn thẳng có hướng, các cung hoặc các đường cong. Hướng của các nét đóng vai trò quan trọng trong việc so sánh sự khác nhau giữa các ký tự. Hướng nét chữ cục bộ của một ký tự có thể được xác định bằng nhiều cách khác nhau: hướng của xương, phân đoạn nét chữ, mã hóa chu tuyến, hướng đạo hàm 0. Hiện nay, các đặc trưng về hướng được áp dụng rộng rãi vì chúng có thể mô tả được hình dáng khái quát của từng ký tự theo sự biến đổi đa dạng của các nét chữ 000.

### 3.3. Đặc trưng hình học và hình thái

Các tính chất cục bộ và toàn cục của các ký tự có thể được biểu diễn bằng các đặc trưng hình học và hình thái. Các loại đặc trưng này có thể phân thành các nhóm sau: Các cấu trúc hình thái 00, Các đại lượng hình học 00, Đồ thị và cây cũng có thể dùng để biểu diễn các từ và các ký tự với một tập các đặc trưng theo một quan hệ phân cấp 0.

Trích chọn đặc trưng hầu hết được thực hiện trên ảnh nhị phân. Tuy nhiên, việc nhị phân hóa ảnh đa cấp xám có thể xóa đi một số thông tin quan trọng của ký tự. Vì vậy, cũng có một số công trình nghiên cứu để trích chọn đặc trưng trực tiếp từ ảnh đa cấp xám 00.

## 4. Các phương pháp nhận dạng chữ viết

Có thể tích hợp theo các hướng tiếp cận sau: Đối sánh mẫu, thống kê, cấu trúc, mô hình Markov ẩn, mạng nơ ron và SVM.

### 4.1. Đối sánh mẫu

Kỹ thuật nhận dạng chữ đơn giản nhất dựa trên cơ sở đối sánh các nguyên mẫu (prototype) để nhận dạng ký tự hoặc từ. Nói chung, toán tử đối sánh xác định mức độ giống nhau giữa hai vectơ (nhóm các điểm, hình dạng, độ cong...) trong một không gian đặc trưng 0.

### 4.2. Phương pháp tiếp cận cấu trúc

Cách tiếp cận theo cấu trúc dựa vào việc mô tả đối tượng nhờ một số khái niệm biểu diễn đối tượng cơ sở trong ngôn ngữ tự nhiên. Một số dạng nguyên thủy thường dùng để mô tả đối tượng như đoạn thẳng, cung,... Mỗi đối tượng được mô tả như một sự kết hợp của các dạng nguyên thủy. Tuy nhiên, vẫn còn nhiều vấn đề liên quan đến nhận dạng cú pháp chưa được giải quyết. Các phương pháp tiếp cận cấu trúc áp dụng cho các bài toán nhận dạng chữ được phát triển theo hai hướng 00:

### 4.3. Các phương pháp thống kê

Hầu hết các kỹ thuật thống kê đều dựa trên cơ sở ba giả thuyết chính sau:

Phân bố của tập đặc trưng là phân bố Gauss hoặc trong trường hợp xấu nhất là phân bố đều.

Có các số liệu thống kê đầy đủ có thể dùng cho mỗi lớp.

Cho tập ảnh  $\{I\}$ , tập ảnh này có thể trích chọn một tập đặc trưng  $\{f_i\} \in F$ ,  $i \in \{1, \dots, n\}$  mà tập đặc trưng này đại diện cho mỗi lớp mẫu riêng biệt.

Sau đây là các hướng tiếp cận thống kê cơ bản được áp dụng trong lĩnh vực nhận dạng chữ:

#### 4.3.1. Nhận dạng phi tham số

Phương pháp này sử dụng để tách các lớp mẫu dọc theo các siêu phẳng được xác định trong một siêu không gian đã cho. Phương pháp phân lớp phi tham số được đánh giá tốt nhất chính là thuật toán phân lớp k-láng giềng gần nhất (k-NN) và thuật toán này được áp dụng rộng rãi trong lĩnh vực nhận dạng chữ viết [19].

#### 4.3.2. Nhận dạng có tham số

Phương pháp này có khả năng thu được một mô hình tham số đối với mỗi ký tự từ các thông tin thích hợp để biểu diễn ký tự, các tham số của mô hình này

dựa trên cơ sở một số xác suất thu được, các ký tự được phân lớp theo một số luật quyết định, chẳng hạn như phương pháp Bayes 0.

#### 4.4. Các phương pháp học máy tiên tiến

##### 4.4.1. Mô hình Markov ẩn

Mô hình Markov ẩn (HMM – Hidden Markov Model) là một mô hình xác suất hữu hạn trạng thái theo kiểu phát sinh tiến trình bằng cách định nghĩa xác suất liên kết trên các chuỗi quan sát. Mỗi chuỗi quan sát được sinh ra bởi một chuỗi các phép chuyển trạng thái, bắt đầu từ trạng thái khởi đầu cho đến trạng thái kết thúc. Tại mỗi trạng thái, một phần tử của chuỗi quan sát được phát sinh ngẫu nhiên trước khi chuyển sang trạng thái tiếp theo. Các trạng thái của HMM được xem là ẩn bên trong mô hình vì tại mỗi thời điểm chỉ nhìn thấy các ký hiệu quan sát, còn các trạng thái khác cũng như sự chuyển đổi trạng thái được vận hành ẩn bên trong mô hình [19].

HMM áp dụng tốt đối với việc nhận dạng chữ viết tay on-line, đặc biệt là nhận dạng chữ viết tay ở mức từ 0.

##### 4.4.2. Mạng nơ ron

Các công trình nghiên cứu về mạng nơ ron để ứng dụng trong lĩnh vực nhận dạng đã được tập hợp, đúc kết trong các sách 0.

Các kiến trúc mạng nơ ron có thể được phân thành hai nhóm chính: mạng truyền thẳng và mạng truyền ngược. Trong các hệ thống nhận dạng chữ, các mạng nơ ron sử dụng phổ biến nhất là mạng SOM (Self Organizing Map) của Kohonen 0 và mạng perceptron đa lớp thuộc nhóm mạng truyền thẳng 0.

Mạng perceptron đa lớp được đề xuất bởi Rosenblatt 0 được nhiều tác giả trong và ngoài nước áp dụng trong các hệ nhận dạng chữ viết tay 00.

Với thuật toán huấn luyện mạng đơn giản nhưng hiệu quả, cùng với những thành công của mô hình này trong các ứng dụng thực tiễn, mạng nơ ron hiện đang là một trong các hướng nghiên cứu của lĩnh vực học máy đang được nhiều nhà nghiên cứu trong và ngoài nước quan tâm 00000.

##### 4.4.3. Máy vectơ tựa

Máy vectơ tựa (SVM – Support Vector Machines) được nghiên cứu từ những năm của thập niên 1960 với những công trình của Vapnik và Lerner (1963), Vapnik và Chervonenkis (1964). Cơ sở của SVM dựa trên nền tảng

của lý thuyết học thống kê và lý thuyết chiều VC (Vapnik Chervonenkis) đã được phát triển qua 3 thập kỷ bởi Vapnik và Chervonenkis. Lý thuyết này bắt đầu có những bước phát triển mạnh mẽ về mặt ứng dụng kể từ những năm cuối của thập niên 1990 (Burges, 1996 [19]; Osumi, 1997 [21] và Platt, 1999 [14]) và từ đó đến nay SVM đã trở thành một công cụ khá mạnh trong nhiều lĩnh vực như: khai phá dữ liệu, nhận dạng chữ viết [9][10]...

Các thuật toán huấn luyện SVM được thực hiện theo ý tưởng sau: tìm siêu phẳng tối ưu trong không gian đặc trưng để cực đại khoảng cách giữa hai lớp mẫu huấn luyện trong bài toán phân lớp nhị phân. Có nhiều thuật toán huấn luyện SVM, các thuật toán chặt khức và thuật toán phân rã 00 hướng tới phân tích bài toán quy hoạch toàn phương (QP - Quadratic Programming) ban đầu thành một dãy các bài toán QP nhỏ hơn. Thuật toán SMO (*Sequential Minimal Optimization*) 0 có thể xem là trường hợp cá biệt của thuật toán phân rã, trong mỗi lần lặp SMO giải một bài toán QP với kích thước là hai bằng giải pháp phân tích, vì vậy không cần phải giải bài toán tối ưu. Các thuật toán này đã được cài đặt trong hầu hết các phần mềm SVM mã nguồn mở hiện nay như SVM<sup>light</sup> 0, LIBSVM [20], SVMToolbox 0 và HeroSvm 0.

#### 4.5. Kết hợp các phương pháp nhận dạng

Các phương pháp phân lớp đã được đề cập ở trên đều có thể áp dụng đối với các hệ nhận dạng chữ viết tay. Mỗi kỹ thuật phân lớp đều có những ưu điểm và nhược điểm riêng. Vấn đề đặt ra là các phương pháp trên có thể kết hợp với nhau theo một cách nào đó để nâng cao hiệu quả nhận dạng hay không? Nhiều công trình nghiên cứu các kiến trúc phân lớp theo ý tưởng kết hợp các kỹ thuật phân lớp đã nêu trên. Các hướng tiếp cận kiến trúc kết hợp các phương pháp để phân lớp có thể chia thành ba nhóm sau: Kiến trúc tuần tự, kiến trúc song song và kiến trúc lai ghép 000.

##### 4.5.1. Kiến trúc tuần tự

Kiến trúc này chuyển kết quả đầu ra của một máy phân lớp thành đầu vào của máy phân lớp tiếp theo 000.

##### 4.5.2. Kiến trúc song song

Kiến trúc này kết nối kết quả của các thuật toán phân lớp độc lập bằng cách sử dụng nhiều chiến lược khác nhau. Trong số các kiến trúc này, tiêu biểu nhất là chiến lược bỏ phiếu 0 và luật quyết định Bayes 0.

### 4.5.3. Kiến trúc lai ghép

Kiến trúc này lai ghép giữa hai kiến trúc tuần tự và song song. Ý tưởng chính là kết hợp các điểm mạnh của cả hai kiến trúc trên và giảm bớt những khó khăn trong việc nhận dạng chữ viết 00.

## 5. Tình hình nghiên cứu về nhận dạng chữ viết

### 5.1. Các nghiên cứu nhận dạng chữ viết trên thế giới

Công nghệ nhận dạng chữ viết tay đã có những bước tiến dài trong các thập kỷ qua. Các phần mềm nhận dạng chữ viết tay on-line của nhiều ngôn ngữ khác nhau đã có mặt trên hầu hết các thiết bị cầm tay PDA (Personal Digital Assistant). Tuy nhiên, việc nhận dạng chữ viết tay off-line cho đến nay vẫn chưa có được giải pháp tổng thể. Các ứng dụng nhận dạng chữ viết tay off-line chỉ giới hạn trong một vài phạm vi hẹp, điển hình như phần mềm nhận dạng các địa chỉ thư ở bưu điện của nhóm nghiên cứu ở trung tâm nghiên cứu về nhận dạng và phân tích văn bản - trường Đại Học Tổng Hợp New York 0, phần mềm nhận dạng chữ viết tay trong lĩnh vực kiểm tra tài khoản ở ngân hàng của nhóm nghiên cứu J.Simon và O.Baret (Laoria/CNRS & ENPC, Paris),...

Kể từ năm 1999, khi Platt đề xuất thuật toán SMO 0 để giải bài toán tối ưu trong kỹ thuật phân lớp SVM thì các nhà nghiên cứu đã tập trung áp dụng phương pháp phân lớp SVM vào các ứng dụng nhận dạng chữ viết tay 00000 hoặc kết hợp SVM với các phương pháp truyền thống khác như mạng nơ ron,... 000.

### 5.2. Các nghiên cứu về nhận dạng chữ viết tiếng Việt

Trong những năm qua, vấn đề nhận dạng chữ viết tay đã và đang được nhiều nhà nghiên cứu trong nước đặc biệt quan tâm về cả hai mặt lý thuyết lẫn triển khai ứng dụng. Tuy nhiên các kết quả nghiên cứu lý thuyết chủ yếu chỉ tập trung vào nhận dạng chữ số hoặc chữ cái tiếng Việt không dấu 00. Chỉ một số ít công trình nghiên cứu đề xuất giải pháp cụ thể cho việc nhận dạng chữ viết tay tiếng Việt, tiêu biểu như: nhận dạng chữ viết tay tiếng Việt on-line 0, nhận dạng chữ viết tay tiếng Việt off-line 0.

Mặt khác, cho đến nay các nghiên cứu ứng dụng nhận dạng chữ viết tay chỉ áp dụng được trong một số phạm vi hẹp. Chẳng hạn như áp dụng vào các ứng dụng

xử lý biểu mẫu tự động từ các biểu mẫu viết tay như các tờ khai, chứng từ, hóa đơn, phiếu đăng ký [1]...

## 6. Kết luận

Nhận dạng chữ viết là lĩnh vực hấp dẫn, có nhiều ứng dụng thiết thực. Tuy nhiên, đây là lĩnh vực khó, đặc biệt là vấn đề nhận dạng chữ viết tay, chưa có công trình nào đề xuất được giải pháp tổng thể. Các nghiên cứu ứng dụng chỉ giới hạn trong những điều kiện cụ thể. Vì vậy, đây là một hướng mở dành cho những người đam mê, quan tâm đến lĩnh vực thị giác máy tính.

## Tài liệu tham khảo

- [1] Hoàng Kiếm, Nguyễn Hồng Sơn, Đào Minh Sơn (2001), “Ứng dụng mạng nơ ron nhân tạo trong hệ thống xử lý biểu mẫu tự động”, Kỷ yếu hội nghị kỷ niệm 25 năm thành lập Viện Công nghệ Thông tin, tr. 560-567.
- [2] Lê Hoài Bắc, Lê Hoàng Thái (2001), “Neural Network & Genetic Algorithm in Application to Handwritten Character Recognition”, Tạp chí Tin học và Điều khiển học, Tập 17, số 4, tr. 57-65.
- [3] Lê Minh Hoàng, Ngô Quốc Tạo, Lương Chi Mai (2001), “Ứng dụng mô hình Markov ẩn trong nhận dạng chữ”, Kỷ yếu hội nghị kỷ niệm 25 năm thành lập Viện Công nghệ Thông tin, tr. 568-577.
- [4] Lương Chi Mai, Nguyễn Hữu Hòa (2001), “Áp dụng mạng nơ ron mờ trong nhận dạng chữ số, chữ viết tay tiếng Việt”, Kỷ yếu hội nghị kỷ niệm 25 năm thành lập Viện Công nghệ Thông tin, tr. 623-631.
- [5] Nguyễn Thị Minh Ánh, Đinh Việt Cường, Ngô Trí Hoài, Nguyễn Việt Hà (2005), “Mô hình liên mạng nơ ron ứng dụng trong nhận dạng ký tự viết tay tiếng Việt”, Kỷ yếu hội thảo Quốc gia: Một số vấn đề chọn lọc của Công nghệ Thông tin, Hải Phòng, tr. 37-46.
- [6] Phạm Anh Phương, Ngô Quốc Tạo, Lương Chi Mai (2008), “Trích chọn đặc trưng wavelet Haar kết hợp với SVM cho việc nhận dạng chữ viết tay tiếng Việt”, Tạp chí Công nghệ Thông tin và Truyền thông, ISSN 0866-7039, kỳ 3, số 20, tr. 36-42.
- [7] Arica N., Yarman-Vural F.T. (2001), “An overview of character recognition focused on off-line handwriting”, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on Volume 31, Issue 2, pp. 216 – 233.
- [8] Christopher J. C. Burges (1998), “A Tutorial on Support Vector Machines for Pattern Recognition”,

- Data Mining and Knowledge Discovery, ISSN: 1384-5810, Vol. 2, No. 2, pp. 121-167.
- [9] Cakmakov D., Gorgevik D. (2005), "Handwritten Digit Recognition Using Classifier Cooperation Schemes", Proceedings of the 2nd Balkan Conference in Informatics, BCI 2005, Ohrid, pp. 23-30.
- [10] Gorgevik D., Cakmakov D. (2004), "An Efficient Three-Stage Classifier for Handwritten Digit Recognition", Proceedings of 17th Int. Conference on Pattern Recognition, ICPR2004, Vol. 4, IEEE Computer Society, Cambridge, UK, pp. 507-510.
- [11] G. Vamvakas, B. Gatos, I. Pratikakis, N. Stamatoopoulos, A. Roniotis and S.J. Perantonis (2007), "Hybrid Off-Line OCR for Isolated Handwritten Greek Characters", The Fourth IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA 2007), ISBN: 978-0-88986-646-1, Innsbruck, Austria, pp. 197-202.
- [12] H. D. Block, B. W. Knight, F. Rosenblatt (1962), "Analysis of A Four Layer Serious Coupled Perceptron", II. Rev. Modern Physics, vol.34, pp.135-152.
- [13] H. J. Kang, S. W. Lee (1999), "Combining Classifiers based on Minimization of a Bayes Error Rates", in Proc. 5th Int. Conf. Document Analysis and Recognition, Bangalore, India, pp.398-401.
- [14] J. Platt (1999), "Fast Training of Support Vector Machines Using Sequential Minimal Optimization", In Advances in Kernel Methods - Support Vector Learning, Cambridge, MA, MIT Press, pp. 185-208.
- [15] J. X. Dong, A. Krzyzak and C. Y. Suen (2003), "A Fast SVM Training Algorithm", International Journal of Pattern Recognition and Artificial Intelligence, vol. 17, no. 3, pp. 367 – 384.
- [16] L. Lam C. Y. Suen (1994), "Increasing Experts for Majority Vote in OCR: Theoretical Considerations and Strategies", in Proc. Int. Workshop Frontiers in Handwriting Recognition, Taiwan, pp. 245-254.
- [17] Mohamed Cheriet, Nawwaf Khurma, Cheng-Lin Liu And Ching Y. Suen (2007), "Character Recognition Systems: A Guide for Students and Practioners", N. Y.: John Wiley & Sons.
- [18] Ngo Quoc Tao, Pham Van Hung (2006), "Online Continues Vietnamese Handwritten Character Recognition based on Microsoft Handwritten Character Recognition Library", IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2006, Singapore, pp. 2024-2026.
- [19] Pham Anh Phuong, Ngo Quoc Tao, Luong Chi Mai (2008), "An Efficient Model for Isolated Vietnamese Handwritten Recognition", The Fourth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE Computer Society, Harbin, China, pp. 358-361.
- [20] Pham Anh Phuong, Ngo Quoc Tao, Luong Chi Mai (2008), "Speeding Up Isolated Vietnamese Handwritten Recognition by Combining SVM and Statistical Features", IJCSSES International Journal of Computer Sciences and Engineering Systems, ISSN 0973-4406, Vol.2, No.4, pp. 243-247.
- [21] J. Platt, N. Cristianini and J. Shawe-Taylor (2000), "Large Margin DAGs for Multiclass Classification", In Advances in Neural Information Processing Systems, volume 2, pp. 547-553.
- [22] R. Collobert and S. Bengio (2001), "Svmtorch: Support Vector Machines for Large-scale Regression Problems", The Journal of Machine Learning Research, Vol. 1, pp 143 – 160.
- [23] R. M. Bozinovic, S. N. Srihari (1989), "Off-line Cursive Script Word Recognition", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.11, no.1, pp.68-83.
- [24] T. Joachims (1998), "Making large-Scale Support Vector Machine Learning Practical", in Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, Cambridge, MA.
- [25] T. Kohonen (1995), "Self Organizing Maps", Springer Series in Information Sciences, vol.30, Berlin.
- [26] V. N. Vapnik (1998), "Statistical Learning Theory", N. Y.: John Wiley & Sons.
- [27] V. Govindaraju, D. Bouchaffra, S. N. Srihari (1999), "Postprocessing of Recognized Strings Using Nonstationary Markovian Models", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.21, no.10, pp. 990-999.
- [28] W. H. Tsai, K.S.Fu (1980), "Attributed Grammar - A Tool for Combining Syntactic and Statistical Approaches to Pattern Recognition", IEEE Trans. System Man and Cybernetics, vol.10, no.12, pp. 873-885.
- [29] Y. Tang, L. T. Tu, J. Liu, S. W. Lee, W. W. Lin, I. S. Shyu (1998), "Off-line Recognition of Chinese Handwriting by Multifeature and Multilevel Classification", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.20, no.5, pp.556-561.
- [30] V. Govindaraju, J. Park, S. N. Srihari (2000), "OCR in A Hierarchical Feature Space", IEEE Trans. Pattern Analysis and Machine Intelligence, vol.22, no.4, pp.400-407.



## OPTICAL CHARACTER RECOGNITION: ACHIEVEMENTS, CHALLENGES AND APPROACHES

**Abstract:** In the field of recognition, Optical Character Recognition (OCR) has had more and more applications in the social life. Up to now, the problem of recognizing printed characters has been almost completely solved (its product ABBYY FineReader 12.0 can recognize printed letters in 20 different languages, the Vietnamese printed character recognition software VnDOCR 4.0 of Ha Noi Institute of Information technology can identify documents containing images, tables and texts with an accuracy level of over 98%). However, in the world as well as in Vietnam, the problem of handwriting recognition still remains a big challenge for researchers. This paper is to present an overview of the achievements, shortcomings and challenges in this field of OCR as well as propose some new approaches for this type of research.

**Key words:** printed character recognition; handwriting recognition; OCR.