

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



CAO XUÂN TRƯỜNG

**NGHIÊN CỨU MẠNG NƠ-RON NHÂN TẠO
ỨNG DỤNG NHẬN DẠNG CHỮ VIẾT TAY ONLINE**

CHUYÊN NGÀNH: KHOA HỌC MÁY TÍNH

MÃ SỐ: 60.48.01.01

LUẬN VĂN THẠC SĨ KỸ THUẬT
NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS. TS TỪ MINH PHƯƠNG

HÀ NỘI - 2019

LỜI CAM ĐOAN

Tôi xin cam đoan các kết quả được trình bày trong luận văn này là công trình nghiên cứu của tôi dưới sự hướng dẫn của GS. TS Từ Minh Phương. Các số liệu, kết quả trong luận văn là hoàn toàn trung thực và chưa được công bố trong bất kỳ công trình nào khác.

Tác giả

Cao Xuân Trường

LỜI CẢM ƠN

Lời đầu tiên, tôi xin phép xin gửi lời cảm ơn sâu sắc đến GS.TS Từ Minh Phương, người trực tiếp hướng dẫn, động viên, khuyến khích tôi trong suốt thời gian nghiên cứu và hoàn thành luận văn cao học.

Xin chân thành cảm ơn quý thầy cô ở Khoa Công nghệ thông tin, quý thầy cô Khoa Sau Đại học, Ban Giám đốc Học viện Công nghệ Bưu chính Viễn thông đã tận tình giảng dạy và tạo điều kiện thuận lợi cho tôi trong quá trình học tập, nghiên cứu.

Cuối cùng, xin cảm ơn cha mẹ và gia đình đã tạo điều kiện thuận lợi cho tôi trong suốt thời gian học tập và thực hiện luận văn tại Học viện.

Xin chân thành cảm ơn!

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC	iii
DANH MỤC HÌNH VẼ	v
DANH MỤC CÁC BẢNG	vii
MỞ ĐẦU	1
Chương 1. TỔNG QUAN VỀ BÀI TOÁN NHẬN DẠNG CHỮ VIẾT TAY ONLINE	2
1.1. Giới thiệu bài toán nhận dạng chữ viết tay online	2
1.1.1. Nhận dạng chữ viết tay online và nhận dạng chữ viết tay offline	2
1.1.2. Các thuộc tính chữ viết tay online	3
1.1.3. Phương pháp thu thập chữ viết tay online	3
1.1.4. Một số vấn đề khi nhận dạng chữ viết tay online	5
1.2. Phương pháp nhận dạng chữ viết tay online	6
1.3. Quy trình phân đoạn chữ viết tay online	7
1.4. Hướng tiếp cận bài toán nhận dạng chữ viết tay online	14
1.5. Ứng dụng bài toán nhận dạng chữ viết tay online	15
1.6. Kết luận chương 1	17
Chương 2. ỨNG DỤNG MẠNG NƠ RON NHÂN TẠO VÀO BÀI TOÁN NHẬN DẠNG CHỮ VIẾT TAY ONLINE	18
2.1. Mạng nơ ron nhân tạo	18
2.1.1. Các khái niệm	18
2.1.2. Đặc trưng của mạng nơ ron	19
2.1.3. Kiến trúc mạng nơ ron nhân tạo	21
2.1.4. Phân loại mạng nơ ron nhân tạo	24
2.1.5. Hàm kích hoạt	26
2.1.6. Phương pháp huấn luyện mạng nơ ron	26

<i>2.1.7. Mạng nơ ron truyền thẳng một lớp ẩn và thuật toán lan truyền ngược.....</i>	<i>29</i>
2.2. Ứng dụng mạng nơ ron vào bài toán nhận dạng chữ viết tay online.....	31
<i>2.2.1. Thu thập dữ liệu.....</i>	<i>31</i>
<i>2.2.2. Mô hình của bài toán nhận dạng chữ viết tay online sử dụng mạng nơ ron nhân tạo</i>	<i>33</i>
<i>2.2.3. Trích chọn đặc trưng và xây dựng vector đầu vào.....</i>	<i>34</i>
<i>2.2.4. Các tham số của mạng nơ ron.</i>	<i>36</i>
<i>2.2.5. Huấn luyện mạng nơ ron.....</i>	<i>38</i>
<i>2.2.6. Nhận dạng chữ viết tay online sử dụng mạng nơ ron</i>	<i>39</i>
2.3. Kết luận chương 2	40
Chương 3. CÀI ĐẶT THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ	41
3.1. Cài đặt thực nghiệm.....	41
<i>3.1.1. Ngôn ngữ lập trình</i>	<i>41</i>
<i>3.1.2. Cấu hình máy</i>	<i>41</i>
<i>3.1.3. Cấu hình mạng nơ ron</i>	<i>41</i>
<i>3.1.4. Giới thiệu chương trình.....</i>	<i>42</i>
3.2. Đánh giá kết quả thực nghiệm	43
<i>3.2.1. Kết quả thực nghiệm.....</i>	<i>43</i>
<i>3.2.2. Đánh giá kết quả.....</i>	<i>44</i>
<i>3.2.3. Đánh giá kết quả.....</i>	<i>46</i>
3.3. Kết luận chương 3	47
KẾT LUẬN	48
1. Các kết quả nghiên cứu của luận văn.....	48
2. Hướng phát triển của luận văn	48
TÀI LIỆU THAM KHẢO.....	49

DANH MỤC HÌNH VẼ

Hình 1.1. Thiết bị CrossPad dùng trong thu nhận chữ viết tay online.....	4
Hình 1.2. Ví dụ về nét trề.....	7
Hình 1.3. Quy trình tách dòng.....	9
Hình 1.4. Các mức độ khó của bài toán phân đoạn trong nhận dạng chữ viết tay	10
Hình 1.5. Độ nghiêng được ước tính bằng hai biểu đồ.....	11
Hình 1.6. Tách dòng văn bản thành các thành phần và điều chỉnh nghiêng ..	12
Hình 1.7. Biểu đồ góc cho dòng văn bản mẫu	12
Hình 1.8. Đường cơ sở và dòng văn bản của một mẫu dòng văn bản.	14
Hình 1.9. Tọa độ của dữ liệu chữ viết tay trong cơ sở dữ liệu UNIPEN.....	15
Hình 1.10. Ứng dụng nhận dạng chữ viết tay trên Google Translate	16
Hình 1.11. Ứng dụng Google Handwriting Input	17
Hình 2.1. Mô hình phi tuyến của một nơ ron.....	21
Hình 2.2. Mô hình phi tuyến thứ hai của một nơ ron	24
Hình 2.3. Mạng nơ ron truyền thẳng nhiều lớp.....	24
Hình 2.4. Mô hình mạng hồi quy	25
Hình 2.5. Các hàm kích hoạt tiêu biểu.....	26
Hình 2.6. Mô hình huấn luyện mạng nơ ron học có giám sát.....	27
Hình 2.7. Mô hình mạng nơ ron học tăng cường.....	28
Hình 2.8. Mạng truyền thẳng một lớp ẩn	29
Hình 2.9. Mô hình thuật toán lan truyền ngược	31
Hình 2.10. Giai đoạn huấn luyện mạng	34
Hình 2.11. Giai đoạn nhận dạng chữ viết tay online	34
Hình 3.1. Giao diện huấn luyện mạng nơ ron.....	42
Hình 3.2. Các giá trị của mạng được lưu lại sau huấn luyện	42
Hình 3.3. Giao diện nhận dạng chữ viết online	43
Hình 3.4. Demo thời gian huấn luyện.....	43

Hình 3.5. Demo minh họa thời gian nhận dạng	44
Hình 3.6. Kết quả nhận dạng khi load file nhiều ký tự.....	46

DANH MỤC CÁC BẢNG

Bảng 2.1. Kết quả thực nghiệm với vector đầu vào khác nhau.....	36
Bảng 3.1. Thống kê kết quả nhận dạng ký tự	44

MỞ ĐẦU

Ngành công nghệ thông tin là một ngành khoa học đang trên đà phát triển mạnh và ứng dụng rộng rãi trên nhiều lĩnh vực. Trong đó nhận dạng chữ viết là một lĩnh vực đem đến nhiều lợi ích thiết thực cho con người. Bài toán nhận dạng chữ viết không chỉ dừng lại ở việc nhận dạng chữ trên giấy mà đã phát triển và mở rộng thành bài toán nhận dạng chữ viết từ các thiết bị điện tử hay còn gọi là nhận dạng chữ viết tay online. Con người từ đó đã có thể viết trực tiếp trên các thiết bị điện tử như smartphone, máy tính bảng hay máy tính xách tay mà không còn phải viết trên giấy như trước đây.

Hiện nay, có nhiều phương pháp khác nhau được áp dụng vào các bài toán nhận dạng. Trong đó việc sử dụng mạng nơ-ron nhân tạo là một trong những phương pháp nhận dạng hiệu quả, có độ chính xác cao. Mạng nơ-ron nhân tạo ra đời xuất phát từ ý tưởng mô phỏng hoạt động của bộ não con người. Kể từ khi ra đời, mạng nơ-ron nhân tạo được ứng dụng rộng rãi trong giải quyết các bài toán về phân loại, dự đoán và nhận dạng bởi khả năng học của nó.

Từ những đặc điểm trên, tôi đã chọn đề tài ***“Nghiên cứu mạng nơ-ron nhân tạo và ứng dụng nhận dạng chữ viết tay online”***. Luận văn này tập trung tìm hiểu phương pháp sử dụng mạng nơ-ron nhân tạo để huấn luyện và nhận dạng chữ viết tay online.

Chương 1. TỔNG QUAN VỀ BÀI TOÁN NHẬN DẠNG CHỮ VIẾT TAY ONLINE

Chương 1 trình bày tổng quan về bài toán nhận dạng chữ viết tay online, về các thuộc tính cơ bản của chữ viết online, các phương pháp nhận dạng chữ viết tay online và các ứng dụng của bài toán này đối với nhiều lĩnh vực khoa học, xã hội.

1.1. Giới thiệu bài toán nhận dạng chữ viết tay online

1.1.1. Nhận dạng chữ viết tay online và nhận dạng chữ viết tay offline

Bài toán nhận dạng chữ viết tay hiện nay đã và đang là bài toán được chú ý đặc biệt và có nhiều ứng dụng thực tế. Nó được nghiên cứu trong khoảng ba bốn thập kỷ gần đây, những vấn đề đặt ra để nghiên cứu rất đa dạng, phụ thuộc vào cách mà chữ viết được mô tả, các dữ liệu được viết ở mức độ nào (ký tự hay câu văn, đoạn văn v.v). Ở mức độ tổng quát nhất bài toán nhận dạng chữ viết tay được chia làm 2 dạng: nhận dạng chữ viết tay online và nhận dạng chữ viết tay offline.

Bài toán nhận dạng chữ viết tay offline được đặt ra để nhận dạng các văn bản viết tay trên giấy bằng bút. Với đặc trưng dữ liệu đầu vào là hình ảnh văn bản viết tay được quét hoặc chụp lại. Sau đó, các thuật toán nhận dạng chữ viết tay sẽ được xây dựng dựa trên các hình ảnh này. Các ứng dụng hiện nay về nhận dạng chữ viết tay offline thường quan tâm tới độ chính xác hơn là việc tối ưu thời gian.

Bài toán nhận dạng chữ viết tay online là bài toán nhận dạng chữ viết tay thu được từ các thiết bị điện tử. Với đặc trưng dữ liệu đầu vào là dãy các điểm thu nhận được trong quá trình con người thực hiện viết trên bề mặt các thiết bị điện tử. Yêu cầu là cần có các thiết bị chuyên dụng như màn hình cảm ứng hay các bảng điện tử để ghi lại quá trình di chuyển của nét bút bao gồm các yếu tố như điểm bắt đầu, điểm kết thúc, các điểm trên mặt phẳng mà nét bút đi qua. Khác với nhận dạng chữ viết tay offline, nhận dạng chữ viết tay online còn cung cấp thêm các thông tin về nét bút, thứ tự các nét được viết mà những điều này trong nhận dạng offline rất khó xác định. Hiện nay việc ứng dụng bài toán nhận dạng chữ viết tay còn rất hạn chế

và vẫn chỉ đang được nghiên cứu sâu rộng với nhiều phương pháp được thực nghiệm như: mô hình mạng nơ ron nhân tạo, mô hình Markov ẩn, v.v.

1.1.2. Các thuộc tính chữ viết tay online

Chữ viết tay online là chữ viết của con người được viết trực tiếp trên màn hình các thiết bị điện tử như điện thoại thông minh, ipad, màn hình máy tính, v.v.

Chữ viết tay online thu được từ các thiết bị điện tử có một số thuộc tính đặc trưng sau [14]:

Pen-up/ pen-down: Đây là các thuộc tính nhị phân cho biết đầu bút có chạm vào màn hình thiết bị điện tử hay không, trong đó pen-up là thuộc tính chỉ nét bút nhấc lên khỏi màn hình và kết thúc một nét, còn pen-down là trạng thái bút chạm màn hình và bắt đầu một nét viết. Một nét được tính từ khi bút đặt xuống để viết đến khi nhấc lên. Các thuộc tính này cho kết quả là số lần nhấc bút hay số nét bút của một chữ được viết.

Speed: Tốc độ viết. Thuộc tính này được thu nhận và tính toán trước khi lấy lại mẫu và sau đó được điều chỉnh cho phù hợp.

X-coordinate: Tọa độ x. Đây là vị trí được lấy sau khi được lọc, có nghĩa là sau khi được trừ đi một đường trung bình từ vị trí thực hay chính là vị trí theo phương ngang của điểm sau khi được chuẩn hóa.

Y-coordinate: Tọa độ y. Tính năng này thể hiện vị trí thẳng đứng của điểm sau khi chuẩn hóa tương tự như tọa độ x.

Writing direction: Hướng viết. Thuộc tính này được cho bởi cosin và sin của góc giữa đoạn thẳng bắt đầu tại điểm đang xét và trục x.

Nghiên cứu các thuộc tính này có vai trò rất quan trọng trong việc xác định các yếu tố đầu vào cho quá trình huấn luyện và nhận dạng.

1.1.3. Phương pháp thu thập chữ viết tay online

**Phương tiện thu thập:*

Khác biệt với nhận dạng chữ viết tay offline, chỉ cần chụp ảnh hay quét dữ liệu đầu vào có thể được ghi nhận để đưa vào xử lý. Việc thu thập dữ liệu chữ viết tay online phải cần các thiết bị chuyên dụng có khả năng ghi nhận các trạng thái của bút như pen up, pen down và tập hợp các điểm mà bút di chuyển qua.

Do dữ liệu lưu trữ trong máy tính phải được rời rạc hóa, nên dãy các điểm thu được sẽ không được lấy liên tục theo thời gian mà được xác định bằng việc lấy mẫu sau những khoảng thời gian xác định. Và có thể xem đường đi của nét bút là các đoạn thẳng nối hai điểm liên tiếp trên tập các điểm mẫu thu được.

Đơn giản nhất có thể dùng các chuyển động của con trỏ chuột của máy tính nhưng dữ liệu thu được thường không chính xác do con người không có thói quen dùng chuột để viết nên các mẫu dữ liệu thu được sẽ khác nhau rất nhiều, gây khó khăn cho bài toán huấn luyện và nhận dạng. Cách lấy dữ liệu thông thường là sử dụng một bảng điện tử và bút chuyên dụng kèm theo có khả năng ghi nhận vị trí các điểm di chuyển của bút từ 80 đến 200 lần một giây. Số điểm lấy được trong một giây càng nhiều thì đường thể hiện của nét bút càng chi tiết, nhưng cũng đồng thời làm tăng khả năng xuất hiện nhiễu trong dữ liệu thu được. Các thiết bị khác như PDA thì sử dụng các màn hình cảm ứng để ghi nhận sự di chuyển của bút hay tay người dùng.



Hình 1.1. Thiết bị CrossPad dùng trong thu nhận chữ viết tay online

**Phương pháp thu thập [1]*

Khi viết, việc nhận ghi dữ liệu có thể được thu thập theo hai cách khác nhau. Với cách thứ nhất, quỹ đạo bút được chuyển đổi thành hình ảnh điểm ảnh và xử lý

với nhận diện ký tự in (OCR) nhận diện như trong nhận dạng off-line. Trong cách thứ hai, quỹ đạo của bút được sử dụng như trình tự của các điểm tọa độ (x, y).

Mỗi phương pháp đều có ưu nhược điểm khác nhau. Tuy nhiên việc nhận dạng online thường được phản ánh rõ qua cách thu thập thứ hai. Phương pháp này có một số đặc điểm quan trọng sau:

- Biết được trình tự viết và có thể được sử dụng trong quá trình nhận dạng chữ viết.
- Thông tin như tốc độ cũng có thể được xem xét như một yếu tố của đầu vào.
- Thông tin số lần nhấc bút được sử dụng hiệu quả để nhận biết số nét vẽ hoặc phân đoạn ký tự viết, v.v.

1.1.4. Một số vấn đề khi nhận dạng chữ viết tay online

Cùng với sự phát triển của Khoa học công nghệ và bùng nổ các thiết bị cảm ứng, nhận dạng chữ online đã và đang chiếm nhiều ưu thế trên nhiều lĩnh vực. Tuy nhiên, trong quá trình nhận dạng chữ viết tay online vẫn gặp phải một số vấn đề bao gồm nhận dạng ký tự, nhận dạng từ, nhận dạng số hay các ký hiệu khác nhau. Cụ thể những khó khăn trong việc nhận dạng do một số yếu tố sau [1]:

- Hạn chế về số lượng người viết. Tỷ lệ nhận dạng đúng cho thiết bị nhận dạng xây dựng đơn phong cách thường cao hơn nhiều so với thiết bị nhận dạng đa phong cách.
- Hạn chế về cách viết. Một người viết có thể viết cùng một ký tự theo nhiều cách khác nhau.
- Khó khăn về ngôn ngữ. Giới hạn về số ký tự nhận dạng, kích thước của từ và nét vẽ.
- Tồn tại nhiều cặp ký tự có hình dạng tương tự, gây khó khăn để nhận biết. Ví dụ các cặp: UV, CL, Oo, I-1, l-1, Z-2, S-5, G-6, Cc, Kk, v.v. Đối với hệ thống nhận dạng dựa trên đặc tính ngữ cảnh sử dụng mới có thể phân biệt được các ký tự này.

- Một vấn đề nữa nảy sinh trong quá trình nhận dạng chữ viết tay online là tồn tại các nét trễ. Bình thường con người có thói quen viết từ trái sang phải, nhưng những nét trễ là những nét không tuân theo quy luật đó, nó được thêm vào các ký tự trước sau khi đã hoàn thành các ký tự sau đó.

Ví dụ: khi viết chữ “th”, nét ngang của chữ “t” thường được viết sau khi đã viết xong chữ “h”, v.v.

1.2. Phương pháp nhận dạng chữ viết tay online

Nhận dạng chữ viết tay online là một chủ đề đã được nghiên cứu chi tiết trong nhiều tài liệu nghiên cứu từ nhiều thập kỷ qua và tồn tại rộng rãi cho tới ngày nay [4], [13]. Hiện tại đã có rất nhiều hướng tiếp cận khác nhau để nhận dạng chữ viết tay online, một số phương pháp học máy thường được áp dụng như: mô hình Markov ẩn, mạng nơ ron nhân tạo hay phương pháp máy vector hỗ trợ (Support Vector Machines - SVM), v.v.

Tuy nhiên có thể chia các phương pháp nhận dạng chữ viết tay online theo hai hướng tiếp cận chính như sau:

- *Over-segment and classify*: Phương pháp phân đoạn và phân loại (phân đoạn và giải mã), phương pháp này được sử dụng trong [13] và [7].

- *Time-sequence interpretation*: Sự biểu diễn trình tự thời gian, cụ thể được thể hiện ở các mô hình Markov ẩn (Hidden Markov Model - HMM) [3], mạng nơ ron thời gian trễ (Time Delay Neural Network - TDNN) [16], và mạng nơ-ron hồi quy (Recurrent Neural Network - RNN) [11], trong đó Long Short Term Memory–LSTM [2] là một mạng cải tiến của RNN để giải quyết vấn đề nhớ các bước dài của RNN, đang nhận được rất nhiều sự chú ý trong học máy hiện nay, v.v.

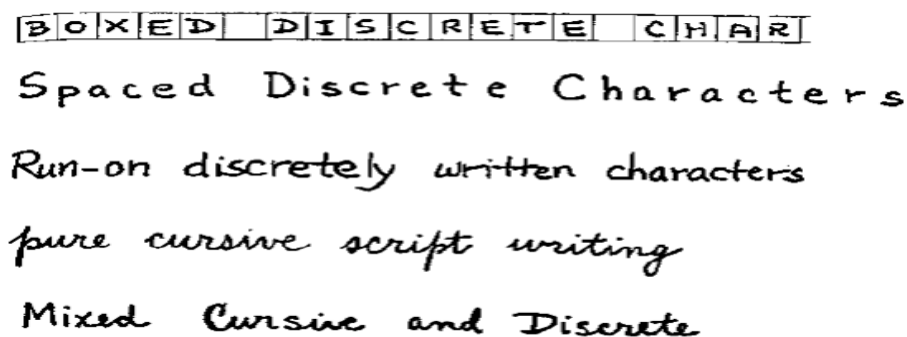
Ngoài việc sử dụng độc lập từng mô hình trên, các nhà khoa học máy tính và các nhà nghiên cứu trên nhiều lĩnh vực ứng dụng AI đã xây dựng các mô hình kết hợp các phương pháp này. Ví dụ: [7] sử dụng phương pháp phân đoạn và giải mã ở cấp độ toàn văn và sử dụng mạng nơ ron thời gian trễ để phân loại các phân đoạn.

Cho tới nay, các nghiên cứu để giải bài toán nhận dạng chữ viết tay online đã đạt được nhiều thành công. Trong lĩnh vực nhận dạng ký tự, các thuật toán học máy và trích chọn đặc trưng cho kết quả khá cao, theo Li và Yeung [12] với phương pháp Lân cận gần nhất sử dụng so sánh mẫu (*Nearest Neighbor using Elastic Matching*) đạt được độ chính xác 97.4% trên 9300 mẫu chữ tiếng Anh, viết thường.

1.3. Quy trình phân đoạn chữ viết tay online

Trong các hệ thống nhận dạng chữ viết tay hoàn chỉnh, yêu cầu được đặt ra là nhận dạng được một trang văn bản hoàn chỉnh. Tuy vậy, trên thực tế, yêu cầu này vẫn đang là một thách thức bởi chưa có một thuật toán nào tối ưu để có thể nhận dạng cả một trang văn bản làm đầu vào. Trước thực tế như vậy, đòi hỏi quá trình nhận dạng phải trải qua các bước phân đoạn. Cụ thể là phân tách các dòng với nhau trong cùng một trang văn bản, tách các từ trên cùng một dòng và phân tích các ký tự riêng biệt của từng từ.

Trong các bài toán nhận dạng chữ viết tay offline, việc phân tích thành các thành phần trong một văn bản được trình bày chi tiết trong tài liệu [13]. Đối với bài toán nhận dạng chữ viết tay online, quá trình này được thực hiện dễ dàng hơn do với mỗi mẫu nhận dạng thu được tọa độ của các nét bút đã có các thông tin về vị trí bắt đầu, vị trí kết thúc, khoảng thời gian liên tiếp giữa các nét vẽ. Trong nhiều trường hợp với các thông tin này cũng đã có thể nhận dạng được các dòng với nhau và các từ có trên một dòng.



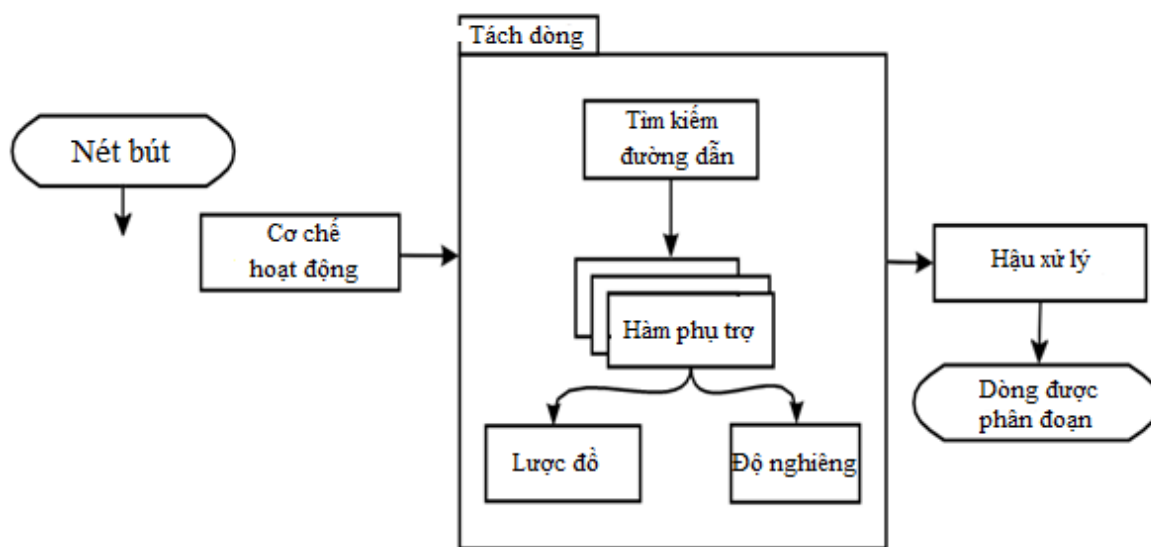
Hình 1.2. Ví dụ về nét trẽ

Quá trình tách dòng:

Phân đoạn dòng là bước tiền đề cần thiết trong gần như hầu hết các hệ thống nhận dạng cho văn bản viết tay. Trong các hệ thống nhận dạng chữ viết tay online, quá trình phân đoạn các dòng văn bản thường được xử lý bằng các thuật toán đơn giản. Ví dụ, nếu chuyển động của bút sang trái và xuống dưới một ngưỡng nhất định, thì giả định là đã bắt đầu một dòng mới. Tuy nhiên, với phương thức chuẩn đoán đơn giản như vậy sẽ thất bại nếu người dùng chuyển đổi giữa các dòng khác nhau. Điều này thường xảy ra khi một chữ cái hoặc một từ bị thiếu được chèn vào sau. Đối với kịch bản như vậy đòi hỏi phải áp dụng một phương pháp hiệu quả hơn.

Đối với nhận dạng offline, các phương thức khá đơn giản có thể được áp dụng nếu khoảng cách giữa các dòng văn bản lân cận liên tiếp đủ lớn. Các phương pháp khác để phát hiện dòng văn bản offline dựa trên các thành phần được kết nối hoặc trên biểu đồ chiều.

Do dữ liệu online có thể chuyển đổi dễ dàng sang định dạng offline, mọi phương thức nhận dạng offline đều có thể được sử dụng để nhận dạng dữ liệu online. Tuy nhiên để có thể phân đoạn được dữ liệu online nên sử dụng thêm các thông tin online thu thập được. Cách tiếp cận hiệu quả là dựa trên việc tìm ra đường đi tối ưu giữa hai dòng văn bản liên tiếp. Đường dẫn tối ưu giữa hai dòng được tìm thấy bằng cách sử dụng lập trình động. Dữ liệu đầu vào cho thuật toán được đề xuất bao gồm dữ liệu online, nhưng tính đến thông tin offline bổ sung. Hầu hết các bước được thực hiện trên phiên bản offline của dữ liệu.

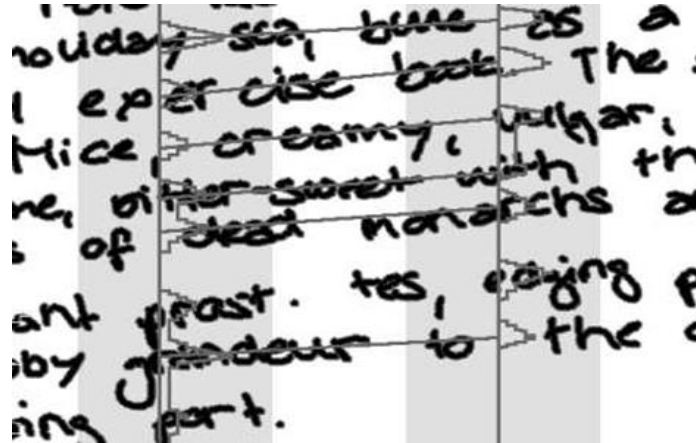


Hình 1.3. Quy trình tách dòng

Hình 1.3 đưa ra mô hình tổng quát về hệ thống phân đoạn dòng văn bản chữ viết tay ở định dạng online. Đầu tiên, một bước tiền xử lý được sử dụng để lọc nhiễu. Một bước tiền xử lý khác là ước lượng ban đầu của điểm bắt đầu của mỗi dòng văn bản. Ước tính này là cần thiết vì các đường dẫn lập trình động sẽ bắt đầu giữa các điểm bắt đầu của mỗi cặp dòng liên tiếp. Điểm bắt đầu của các dòng văn bản được ước tính bằng cách tính toán biểu đồ dọc của các điểm ảnh nền trước ở phần bên trái của tài liệu. Để ngăn chặn bỏ qua các dòng, một điểm bắt đầu mới được tạo ra nếu kích thước khoảng cách giữa hai điểm bắt đầu liên tiếp vượt quá kích thước khoảng cách trung bình. Nó thường không có vấn đề nếu quá nhiều điểm khởi đầu đã được tìm thấy bởi vì điều này thường dẫn đến không gian trống giữa hai đường dẫn, có thể dễ dàng phát hiện và bị xóa.

Cuối cùng, cần ước lượng độ nghiêng của tài liệu cần nhận dạng. Các nghiên cứu đưa ra 02 phương pháp ước lượng sai lệch: Sự suy giảm tuyến tính và phương pháp tiếp cận mới. Trong phương pháp tiếp cận mới, hai biểu đồ của các điểm ảnh tiền cảnh được tính toán. Một biểu đồ nằm ở giữa phía bên trái của vùng văn bản, tức là, nó bao phủ diện tích từ 12,5% đến 37,5% chiều rộng của văn bản; biểu đồ khác nằm ở giữa phía bên tay phải, tức là, nó bao phủ diện tích từ 62,5% đến 87,5%

chiều rộng của văn bản. Tối đa của hai biểu đồ này sau đó được kết hợp bằng thuật toán tìm kiếm tối ưu. Cách tiếp cận được minh họa trong hình 3.4, nơi các khu vực được xem xét các biểu đồ được đánh dấu bằng một nền màu xám. Trong phát hiện dòng văn bản, thử nghiệm phương pháp tiếp cận thứ hai dựa trên biểu đồ thực hiện tốt hơn hồi quy tuyến tính.



Hình 1.4. Các mức độ khó của bài toán phân đoạn trong nhận dạng chữ viết tay

Trong một số trường hợp nhận dạng đơn giản, các ký tự riêng biệt được viết hoàn toàn trong một ô kích thước xác định trước, việc nhận dạng trở nên dễ dàng khi chỉ phải nhận dạng từng ký tự trong từng ô. Bên cạnh đó, việc các ký tự của từng từ cũng được viết tách ra riêng biệt cũng là trường hợp đơn giản cho quá trình nhận dạng.

Vấn đề khó khăn trong quá trình nhận dạng chữ online đó là việc phân đoạn đối với các chữ viết liền nét, một nét bút có thể thực hiện viết cả một hay nhiều ký tự cùng lúc.

Ngoài ra, nét trễ cũng là một khó khăn trong quá trình nhận dạng chữ online, đó là khi người dùng không tuân thủ các quy định viết từ trái qua phải. Nét trễ được thêm vào ký tự bên trái sau khi hoàn thành ký tự bên phải nó.



Hình 1.5. Độ nghiêng được ước tính bằng hai biểu đồ

Tiền xử lý:

Ở bước tách dòng, tiền xử lý các dữ liệu thu được là bước quan trọng trong hệ thống nhận dạng chữ viết tay bởi mỗi người viết có các phong cách viết khác nhau về độ nghiêng, độ xiên, chiều cao, chiều rộng của các ký tự. Phương pháp thường được dùng trong bước tiền xử lý ở một số hệ thống nhận dạng như sau:

- Tiền xử lý trực tiếp:

Trong quá trình thu nhận chữ viết tay, người viết có thể ngồi hoặc đứng. Đặc biệt, trong trường hợp cơ sở dữ liệu online thu được từ bảng viết, trong quá trình viết người viết thường đứng viết, do đó chữ viết tay online thu được thường có một số điểm khác với chữ viết thu được từ việc sử dụng bút viết trên các thiết bị cầm tay. Chữ viết tay trên bảng thường có xu hướng nhỏ dần về phía bên phải, và vị trí chữ phía bên phải thường cao hoặc thấp hơn phía bên trái. Đường cơ sở xác định dòng không phải là một đường thẳng mà có độ gập quăng và xiên lệch. Dòng viết trên bảng cũng không tuân theo một quy tắc cố định về độ nghiêng, mà thường phân thành nhiều phần nhỏ. Quá trình tiền xử lý sẽ tiến hành xử lý trên các khối này một cách riêng biệt. Việc phân tách được thực hiện bằng cách xác định các khoảng trắng trên một dòng. Dòng chữ đó được tách tại vị trí khoảng trắng nếu kích thước của khoảng trắng đó lớn hơn kích thước khoảng trắng trung bình và kích thước của 2 đoạn sau khi được tách phải lớn hơn một ngưỡng cho trước. Ví dụ về tách dòng được thể hiện trong hình sau:

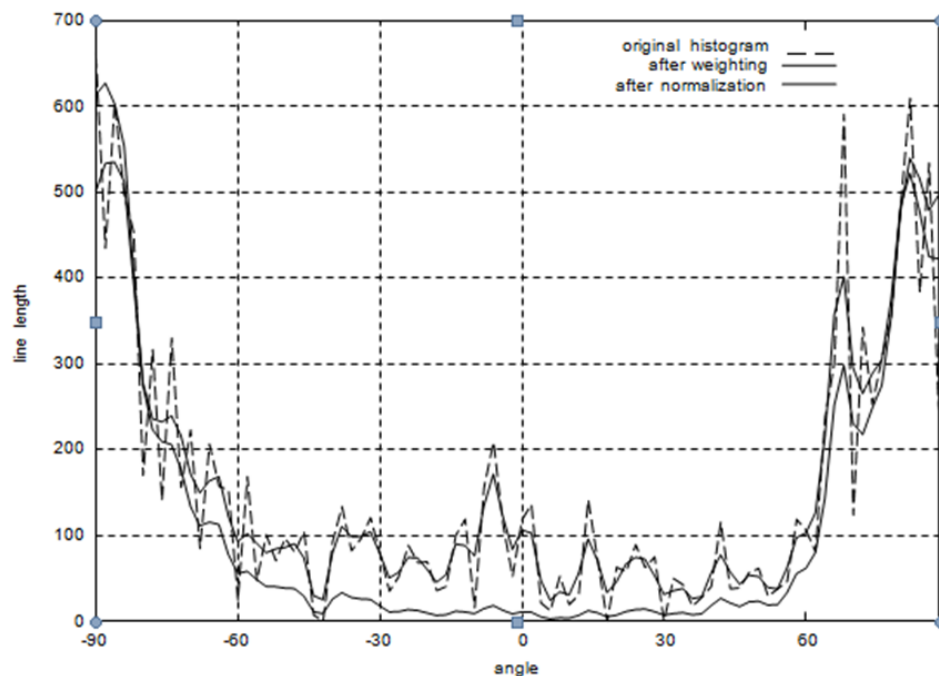
never die. you wish they did.
never die. you wish they did.

Hình 1.6. Tách dòng văn bản thành các thành phần và điều chỉnh nghiêng

- Chuẩn hóa qua các bước:

+ Các phần được hiệu chỉnh đưa về đúng độ nghiêng của chúng. Một hàm hồi quy tuyến tính được tính trên tất cả các điểm trước khi hướng của toàn bộ dòng văn bản được tính dựa theo các tham số hồi quy. Quá trình này được mô tả trong Hình 1.6. Đường màu đen ở dòng chữ phía trên chỉ ra rằng hàm hồi quy tuyến tính được tính cho mỗi phần của dòng chữ.

+ Bước chuẩn hóa về độ xiên được thực hiện dựa vào việc sử dụng lược đồ xám. Ví dụ về lược đồ của dòng được biểu diễn trong Hình 1.7. Phạm vi của lược đồ từ -90° đến 90° với kích thước bước nhảy là 2° .



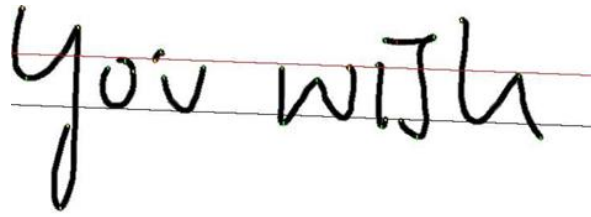
Hình 1.7. Biểu đồ góc cho dòng văn bản mẫu

Những nét trề, như nét gạch ngang của ký tự t hoặc dấu chấm trên ký tự i, là một trong những vấn đề gây khó khăn cho quá trình nhận dạng. Những nét này là nguyên nhân gây ra những biến thể của thông tin thời gian bởi thứ tự viết những nét này của những người viết khác nhau là khác nhau. Do đó những nét trề thường bị xóa khỏi nét viết trong hầu hết các hệ thống nhận dạng chữ viết online. Việc loại bỏ những nét trề này thường sử dụng phương pháp sau: Những nét viết phía trên những vùng đang xét được mà được đưa từ trái qua phải sẽ bị xóa. Để giữ lại được thông tin những nét trề này cho bước nhận dạng, một đặc trưng gọi là Hat feature sẽ được sử dụng trong tập các đặc trưng.

Do chữ viết tay online được thu nhận dựa trên thời gian chứ không phụ thuộc vào vị trí, nên số lượng điểm thu được khác nhau dựa trên tốc độ viết khác nhau. Sự sai khác này cũng cần được loại bỏ để chuẩn hóa dữ liệu. Để chuẩn hóa số lượng điểm, một chuỗi các điểm thu được sẽ được thay bằng một chuỗi mới, trong đó mỗi 2 cặp điểm liên tục trên một quỹ đạo có cùng khoảng cách với nhau.

Bước tiếp theo là tính đường cơ sở và đường corpus. Đường cơ sở tương ứng với đường gốc mà chữ được viết. Đường corpus là đường đi qua các đỉnh của những chữ cái viết thường. Những đường này được sử dụng để chuẩn hóa kích thước của từ. Vị trí lớn nhất và nhỏ nhất của tọa độ y của nét viết được tính để thu được đường cơ sở và đường corpus. Sau đó hai đường hồi quy tuyến tính qua vị trí lớn nhất và nhỏ nhất sẽ được tính với một ràng buộc cụ thể là hai đường kết quả thu được phải có độ dốc bằng nhau. Sau khi đường hồi quy được tính toán, những điểm ít phù hợp nhất sẽ bị loại và một đường hồi quy tuyến tính khác được tính toán, thực hiện. Bước hiệu chỉnh này được thực hiện 2 lần để đưa ra kết quả đường cơ sở và đường corpus. Hình 1.8 là một ví dụ đưa ra đường cơ sở và đường corpus.

Bước cuối cùng của việc chuẩn hóa là chuẩn hóa độ rộng ký tự. Việc này được thực hiện bằng cách điều chỉnh kích thước theo chiều ngang của từ với một tỉ lệ được tính dựa trên số lượng nét theo chiều ngang giữa đường cơ sở và đường corpus.



Hình 1.8. Đường cơ sở và dòng văn bản của một mẫu dòng văn bản.

1.4. Hướng tiếp cận bài toán nhận dạng chữ viết tay online

Đối với bài toán nhận dạng chữ viết tay online có thể tiếp cận theo hai hướng:

Cách tiếp cận thứ nhất: Sử dụng các mô hình học máy để huấn luyện một bộ phân lớp cho toàn bộ các từ có thể từ điển. Tuy nhiên cách tiếp cận này chỉ có thể áp dụng với từ điển có số lượng từ nhỏ (khoảng 100-200). Đối với những từ điển có số lượng từ lớn hơn thì ngay cả việc thu thập dữ liệu cũng gặp khó khăn vì sự đa dạng trong cách thể hiện của mỗi từ.

Cách tiếp cận thứ hai: Tiến hành phân đoạn cả từ rồi sử dụng một bộ phân lớp ở mức ký tự để tìm ra kết quả cuối cùng. Ở cách tiếp cận này chỉ có thể áp dụng nếu đặt ra những ràng buộc nhất định như có khoảng cách rõ ràng giữa các ký tự trong cùng một từ, v.v.

Luận văn tiếp cận theo hướng thứ hai. Tuy nhiên chỉ tập trung vào nhận dạng các ký tự riêng biệt sau khi người viết viết lên các thiết bị điện tử, sau quá trình thu thập, phân tách thành mỗi ký tự riêng biệt có các tọa độ (x,y) về nét bút của mình trên mặt phẳng và được chia thành các SEGMENT hay chính là các nét bút riêng biệt, cụ thể là sử dụng bộ cơ sở dữ liệu UNIPEN (Hình 1.9) để ứng dụng mạng nơ ron vào quá trình huấn luyện và nhận dạng ký tự.

32	.PEN_UP	
33	6769	12858
34	6788	12871
35	6818	12896
36	6832	12915
37	6842	12937
38	6851	12961
39	6777	12954
40	6767	12946
41	6759	12938
42	.PEN_DOWN	
43	6695	12774
44	6689	12754
45	6829	13105
46	6833	13105

Hình 1.9. Tọa độ của dữ liệu chữ viết tay trong cơ sở dữ liệu UNIPEN

1.5. Ứng dụng bài toán nhận dạng chữ viết tay online

Từ giữa những năm 50 của thế kỷ XX, nhận dạng là một lĩnh vực có sức hấp dẫn rất lớn cho việc nghiên cứu và phát triển, thu hút nhiều nhà khoa học tham gia nghiên cứu. Những sản phẩm nhận dạng chữ viết mang tính thương mại bắt đầu xuất hiện vào những năm 60 của thế kỷ XX.

Trong bối cảnh các thiết bị cảm ứng như Smartphone, Ipad, v.v ngày càng trở nên phổ biến, thì bài toán nhận dạng chữ viết online trở nên đặc biệt được quan tâm. Con người có thể viết trực tiếp trên một màn hình tinh thể lỏng (LCD) hiển thị với một cây bút điện tử hay đơn giản là ngón tay. Màn hình là một ma trận nhạy cảm ghi lại sự chuyển động của đầu bút trên bề mặt. Quỹ đạo của các cây bút xuất hiện gần như ngay lập tức trên màn hình. Từ đó, bài toán nhận dạng chữ viết tay online đóng vai trò quan trọng, cho phép máy tính có thể hiểu người viết muốn thực hiện thao tác gì, lệnh gì, v.v.

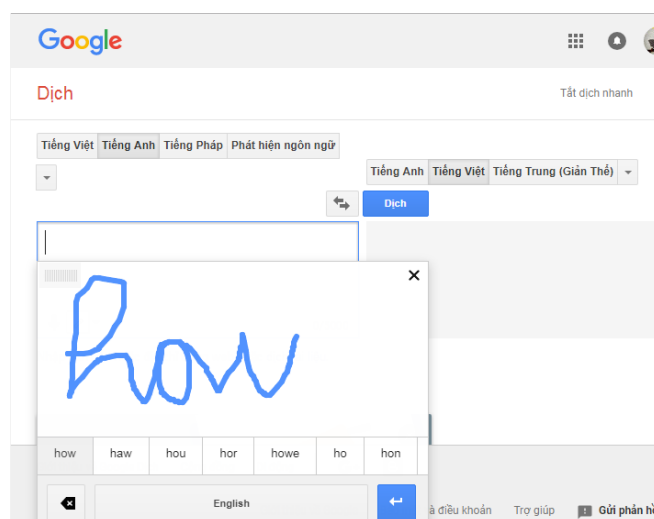
Hiện nay, ứng dụng của bài toán nhận dạng chữ viết tay online là rất lớn. Một ứng dụng quan trọng của nhận dạng chữ viết tay online là sự sao chép của văn bản chữ viết tay cho phép máy tính có thể đọc được văn bản. Rõ ràng văn bản được viết bởi cây bút thì nhanh hơn rất nhiều nếu gõ từ bàn phím, đặc biệt đối với ngôn ngữ như Trung Quốc, Lào hay Thái Lan, v.v. Con người có thể viết trực tiếp các văn bản của mình trên màn hình cảm ứng, máy tính sẽ nhận dạng và đưa ra kết quả

là đoạn văn bản vừa được viết. Bên cạnh đó, ứng dụng của nhận dạng chữ viết tay online còn có thể đưa cho người viết các gợi ý về từ được viết nếu người viết gặp khó khăn khi không thể nhớ chính xác các nét chữ của nhiều ngôn ngữ khác nhau. Đây là một đặc điểm nổi trội của nhận dạng chữ viết tay online.

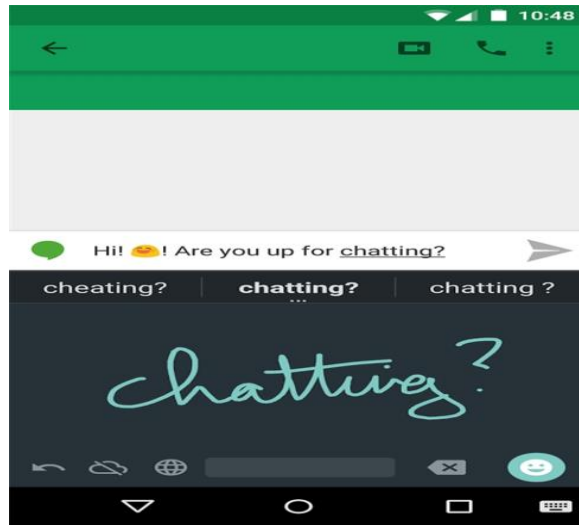
Ngày nay có thể kể tới một số ứng dụng cụ thể của nhận dạng chữ viết tay online như:

- Hệ thống nhận dạng chữ viết tay rời rạc online trên một trạm làm việc của IBM do H.S.M.Beigi, C.C.Tapert, M.Ukeison và C.G.Wolf ở phòng thực hành Watson IBM cài đặt.

- Google đã tích hợp vào dịch vụ Google Translate công cụ nhận dạng chữ viết tay online, điều này có ý nghĩa người dùng có thể viết các ký tự mình muốn lên màn hình thiết bị và Google Translate có thể nhận diện được các ký tự đó. Thậm chí khi người dùng nhìn thấy một từ hoặc một cụm từ của một ngôn ngữ khác mà họ không hiểu thì đơn giản là vào Google translate vẽ nó lên trên điện thoại thông minh hoặc màn hình máy tính bảng của họ, chương trình sẽ nhận diện các từ viết tay và dịch ra. Hiện nay, Google Translate có hỗ trợ nhận dạng chữ viết tay online cho 45 ngôn ngữ. Tuy nhiên tính năng nhận dạng chữ viết tay của Google Translate chỉ có sẵn trên hệ điều hành Android, đây là một điểm hạn chế của hệ thống này.



Hình 1.10. Ứng dụng nhận dạng chữ viết tay trên Google Translate



Hình 1.11. Ứng dụng Google Handwriting Input

- Bên cạnh đó, Google còn tung ra một sản phẩm nhận dạng chữ viết tay khá chính xác là ứng dụng Handwriting Input. Ứng dụng này có khả năng nhận diện chữ viết tay trên Android, hỗ trợ 82 ngôn ngữ. Đặc biệt, ứng dụng còn cho phép nhập các ký tự đặc biệt và biểu tượng Emoji.

1.6. Kết luận chương 1

Chương 1 đã trình bày bài toán nhận dạng chữ viết tay, các đặc trưng của bài toán nhận dạng chữ viết tay online, các thuộc tính cơ bản của chữ viết online, một số phương pháp thu thập và một số vấn đề khó khăn gặp phải khi nhận dạng chữ viết tay online.

Bài toán nhận dạng chữ viết tay online phải trải qua các bước phân đoạn như phân đoạn dòng, từ và ký tự, nhưng trong luận văn chỉ tập trung nghiên cứu về nhận dạng ký tự. Ngoài ra, nội dung chương cũng trình bày các ứng dụng phổ biến của bài toán nhận dạng chữ viết tay online hiện nay.

Chương 1 cũng trình bày tổng quan về một số phương pháp nhận dạng chữ viết online đang được quan tâm và nghiên cứu hiện nay, trong đó mô hình mạng nơ ron nhân tạo kết hợp với các kỹ thuật học sâu là đối tượng nghiên cứu ở chương 2 và là công cụ nhận dạng ở chương 3 của luận văn.

Chương 2. ỨNG DỤNG MẠNG NƠ RON NHÂN TẠO VÀO BÀI TOÁN NHẬN DẠNG CHỮ VIẾT TAY ONLINE

Chương 2 trình bày tổng quan về mạng nơ ron nhân tạo, các đặc trưng cơ bản của mạng, kiến trúc cũng như cách phân loại mạng nơ ron, khái quát về các hàm kích hoạt được sử dụng trong mạng và quá trình huấn luyện và nhận dạng của mạng. Tiếp theo, nội dung chương sẽ trình bày về ứng dụng của mạng nơ ron nhân tạo vào bài toán nhận dạng chữ viết tay online, về quá trình trích chọn đặc trưng từ cơ sở dữ liệu UNIPEN sử dụng mạng nơ ron truyền thẳng một lớp ẩn và thuật toán lan truyền ngược.

2.1. Mạng nơ ron nhân tạo

2.1.1. Các khái niệm

Các nghiên cứu về bộ não người được tiến hành từ hàng nghìn năm nay. Cùng với sự phát triển của khoa học kỹ thuật đặc biệt là những tiến bộ trong ngành điện tử hiện đại, việc con người bắt đầu nghiên cứu mạng nơ ron nhân tạo (Neural Networks -NN) là hoàn toàn tự nhiên, với suy nghĩ làm sao để máy móc có khả năng bắt chước não người.

Trong phần này, sẽ điếm qua lược sử của mạng nơ ron nhân tạo cũng như các khái niệm cơ bản để chúng ta có thể hình dung sự hoạt động của nó.

Sự kiện đầu tiên đánh dấu sự ra đời của mạng nơ ron nhân tạo diễn ra vào năm 1943 khi nhà thần kinh học Warren McCulloch và nhà toán học Walter Pitts công bố bài báo “*A logical calculus of the ideas immanent in Nervous activity*” nhằm mô tả cách thức hoạt động của các nơ ron. Họ cũng tiến hành xây dựng một mạng nơ ron đơn giản bằng các mạch điện. Các nơ ron của họ được xem như là các thiết bị nhị phân với ngưỡng cố định. Kết quả của các mô hình này là các hàm logic đơn giản chẳng hạn như “ $a \text{ OR } b$ ” hay “ $a \text{ AND } b$ ”.

Ứng dụng thực nghiệm đầu tiên của các mạng nơ ron nhân tạo là vào cuối những năm 50 cùng với phát minh của mạng Perceptron (Perceptron Network) và luật học tương ứng bởi Frank Rosenblatt. Mạng này có khả năng nhận dạng các

mẫu. Cùng thời gian đó, Bernard Widrow và Ted Hoff đã đưa ra một thuật toán học mới và sử dụng nó để huấn luyện cho các mạng nơ ron tuyến tính thích nghi, mạng có cấu trúc và chức năng tương tự như mạng của Rosenblatt. Luật học Widrow-Hoff vẫn còn được sử dụng cho đến nay.

Tiếp theo vào những năm 80, việc nghiên cứu về mạng nơ ron phát triển rất mạnh mẽ với việc ra đời của PC. Cùng đó là sự ra đời của thuật toán lan truyền ngược để luyện các mạng nhiều lớp được một số nhà nghiên cứu độc lập tìm ra như David Rumelhart, James McClelland, v.v.

Mạng NN được sử dụng rộng rãi trong giai đoạn 1980 đến 1990, giảm dần sự phổ biến để giải quyết các bài toán trí tuệ nhân tạo vào cuối những năm 1990. Gần đây mạng NN được hồi sinh trở lại bởi hạn chế về mặt kỹ thuật đã được khắc phục, khi mà mạng NN là một thuật toán cần chi phí lớn để tính toán nên cần những máy tính có tốc độ tính toán đủ nhanh mới mang lại hiệu quả.

Hiện nay chưa có một khái niệm tổng quát nào về mạng nơ ron, song phần lớn những người làm việc trong lĩnh vực mạng nơ ron đều có thể đồng ý với khái niệm sau:

“Mạng nơ ron nhân tạo (Artificial Nơ ron Network) gọi tắt là mạng nơ ron (nơ ron network), là mô hình xử lý thông tin phỏng theo cách thức xử lý thông tin của các hệ nơ ron sinh học trong bộ não người. Nó được tạo nên từ một số lượng lớn các phần tử (gọi là phần tử xử lý hay nơ ron) kết nối với nhau thông qua các liên kết (gọi là trọng số liên kết) làm việc như một thể thống nhất để giải quyết một vấn đề cụ thể nào đó.”

2.1.2. Đặc trưng của mạng nơ ron

Một số đặc trưng của mạng nơ ron thu được từ việc mô phỏng trực tiếp bộ não con người như sau:

- *Tính chất phi tuyến:* Đây là tính chất rất quan trọng. Một nơ ron có thể tính toán một cách tuyến tính hay phi tuyến. Một mạng nơ ron được cấu thành bởi sự kết nối các nơ ron phi tuyến thì tự nó sẽ có tính phi tuyến. Hơn nữa, điều đặc biệt là

tính phi tuyến này được phân tán trên toàn mạng. Nó tạo ra các mặt cắt trong không gian dữ liệu không phải là siêu phẳng mà lồi lõm không đều. Từ đó, không gian mẫu có thể được chia thành nhiều miền.

- *Tính chất thích nghi*: Các mạng nơ ron có một khả năng mặc định là biến đổi các trọng số liên kết tùy theo sự thay đổi của môi trường xung quanh. Đặc biệt, một mạng nơ ron đã được tích lũy để hoạt động trong một môi trường xác định có thể được tích lũy lại một cách dễ dàng khi có những thay đổi nhỏ của các điều kiện môi trường hoạt động. Hơn nữa, khi hoạt động trong một môi trường không ổn định (các số liệu thống kê thay đổi theo thời gian), một mạng nơ ron có thể được thiết kế sao cho có khả năng thay đổi các trọng số liên kết của nó theo thời gian thực. Như một quy tắc chung, có thể nói rằng tạo ra một hệ thống càng có khả năng thích nghi thì tính năng của nó sẽ càng mạnh khi hệ thống cần phải hoạt động trong một môi trường không ổn định. Tuy nhiên, tính thích nghi không phải lúc nào cũng đem đến sức mạnh, nó có thể làm điều ngược lại.

- *Tính chất đưa ra lời giải có bằng chứng*: Trong ngữ cảnh phân loại mẫu, một mạng nơ ron có thể được thiết kế để đưa ra thông tin không chỉ về mẫu được phân loại, mà còn về sự tin cậy của quyết định đã được thực hiện. Thông tin này có thể được sử dụng để loại bỏ các mẫu mơ hồ hay nhập nhằng.

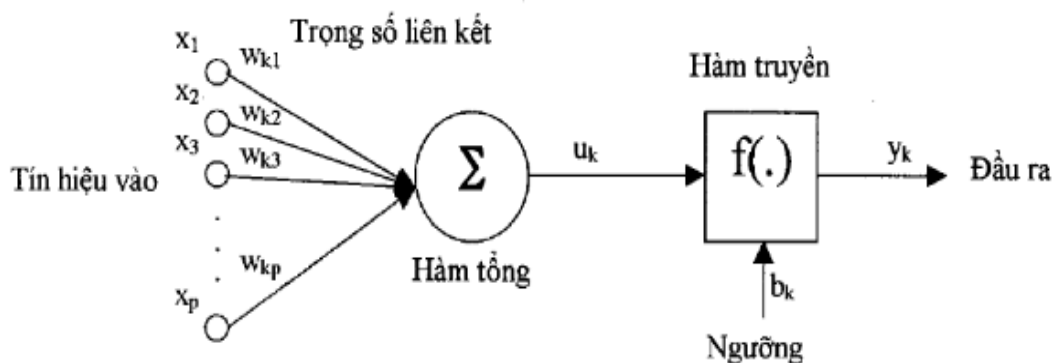
- *Tính chất chấp nhận sai sót*: Não người có khả năng dung thứ lỗi, tức là với những thông tin thiếu chính xác, không đầy đủ mà vẫn có thể xử lý, giải quyết được các vấn đề đặt ra. Mạng nơ ron nhân tạo bắt chước khả năng này của não bộ. Một mạng nơ ron, được cài đặt dưới dạng phần cứng, vốn có khả năng chấp nhận lỗi, hay khả năng *tính toán thô* (không nhạy cảm lỗi), với ý nghĩa là tính năng của nó chỉ *thoái hoá* (chứ không đổ vỡ) khi có những điều kiện hoạt động bất lợi. Ví dụ, nếu một nơ ron hay các liên kết kết nối của nó bị hỏng, việc nhận dạng lại một mẫu được lưu trữ sẽ *suy giảm* về chất lượng. Tuy nhiên, do bản chất phân tán của thông tin lưu trữ trong mạng nơ ron, sự hỏng hóc cũng được trải ra trên toàn mạng. Như vậy, về cơ bản, trong trường hợp này một mạng nơ ron sẽ thể hiện một sự thoái hoá về tính năng hơn là sự đổ vỡ trầm trọng.

- *Tính chất đồng dạng trong phân tích và thiết kế.* Về cơ bản, các mạng nơ ron có tính chất chung như là các bộ xử lý thông tin. Chúng đều được cấu thành từ các nơ ron riêng biệt liên kết với nhau. Liên kết càng phức tạp thì mạng nơ ron càng mạnh. Các mạng tổ hợp có thể được xây dựng bằng cách tích hợp nhiều mạng khác nhau.

2.1.3. Kiến trúc mạng nơ ron nhân tạo

Từ định nghĩa cho thấy đơn vị hình thành nên mạng nơ ron là các nơ ron. Chức năng của mạng nơ ron nhân tạo tương tự như chức năng đơn giản nhất của các nơ ron sinh học đó là tích lũy (ghi nhớ) những tri thức đã học về các sự vật để nhận biết, phân biệt chúng mỗi khi gặp lại. Chức năng này được thực hiện thông qua một chuỗi liên tiếp các quá trình xử lý thông tin của các nơ ron trong mạng.

Về cơ bản, mỗi nơ ron sinh học thực hiện nhiệm vụ của mình thông qua các bước: nhận đầu vào, xử lý đầu vào đó, đưa dữ liệu đã được xử lý ra và liên lạc với các nơ ron khác để gửi dữ liệu ra. Với bản chất là mô phỏng đơn giản của nơ ron sinh học, nơ ron nhân tạo cũng thực hiện nhiệm vụ của mình thông qua các thao tác: nhận đầu vào từ các nơ ron trước nó, xử lý đầu vào bằng cách nhân mỗi đầu vào này với trọng số liên kết tương ứng và tính tổng các tích thu được rồi đưa qua một hàm truyền, sau đó gửi kết quả cuối cùng cho các nơ ron tiếp theo hoặc đưa ra output. Cứ như vậy, các nơ ron hoạt động phối hợp với nhau tạo thành hoạt động chính của mạng nơ ron. Sơ đồ 2.1 chỉ ra mô hình cấu trúc một nơ ron nhân tạo:



Hình 2.1. Mô hình phi tuyến của một nơ ron

Hình 2.1 cho thấy mô hình nơ ron bao gồm các thành phần sau:

- (x_1, x_2, \dots, x_p) với $p \geq 1$: Là các tín hiệu đầu vào của nơ ron. Các tín hiệu này có thể là đầu ra của các nơ ron trước đó hoặc đầu vào ban đầu của mạng và thường được đưa vào dưới dạng một vector p chiều.

- $(w_{k1}, w_{k2}, \dots, w_{kp})$: Là tập các trọng số liên kết của nơ ron k với p đầu vào tương ứng (x_1, x_2, \dots, x_p) . Thông thường, các trọng số này được khởi tạo một cách ngẫu nhiên ở thời điểm khởi tạo mạng và được cập nhật liên tục trong quá trình học của mạng.

- *Hàm tổng* Σ : Là hàm tổng trên một nơ ron, dùng để tính tổng các giá trị kích hoạt lên nơ ron đó. Thông thường, đây là tổng của các tích giữa đầu vào với trọng số liên kết tương ứng của nơ ron.

- u_k là tổng các giá trị kích hoạt lên nơ ron thứ k , giá trị này chính là đầu ra của hàm tổng.

- b_k là ngưỡng (còn gọi là hệ số bias) của nơ ron thứ k , giá trị này được dùng như một thành phần phân ngưỡng trên hàm truyền và cũng được cập nhật liên tục trong quá trình học của mạng.

- $f(.)$ là hàm truyền, còn gọi là hàm kích hoạt (activation function). Hàm truyền được dùng để giới hạn phạm vi đầu ra của mỗi nơ ron. Đối số của hàm là giá trị hàm tổng và ngưỡng b_k . Thông thường, phạm vi đầu ra của mỗi nơ ron được giới hạn trong đoạn $[0, 1]$ hoặc $[-1, 1]$. Như vậy miền giá trị của các hàm kích hoạt cũng là một trong hai đoạn trên. Việc lựa chọn hàm truyền nào cho phù hợp tùy thuộc vào từng bài toán.

- y_k là tín hiệu đầu ra của nơ ron thứ k , mỗi nơ ron thường có một đầu ra.

Dưới dạng công thức toán học, chúng ta có thể mô tả một nơ ron k bằng cặp công thức sau:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (2.1)$$

Và

$$y_k = \varphi(u_k + b_k) \quad (2.2)$$

Hệ số hiệu chỉnh b_k là một tham số ngoài của nơ ron nhân tạo k . Chúng ta có thể thấy được sự có mặt của nó trong công thức (2.2). Một cách tương đương, chúng ta có thể tổ hợp các công thức (2.1) và (2.2) như sau:

$$v_k = \sum_{j=0}^m w_{kj} x_j \quad (2.3)$$

Và

$$y_k = \varphi(v_k) \quad (2.4)$$

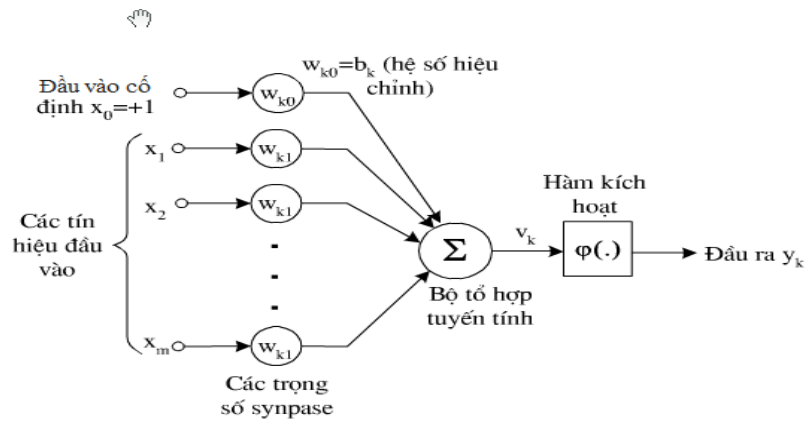
Trong công thức (2.3), chúng ta đã thêm một liên kết mới. Đầu vào của nó là:

$$x_0 = +1 \quad (2.5)$$

Và trọng số của nó là:

$$w_{k0} = b_k \quad (2.6)$$

Như vậy có thể vẽ lại mô hình của nơ ron k như trong Hình 2.2. Trong hình này, nhiệm vụ của hệ số hiệu chỉnh là thực hiện hai việc: (1) thêm một tín hiệu đầu vào cố định là 1 và (2) thêm một trọng số liên kết mới bằng giá trị của hệ số b_k . Mặc dù các mô hình trong Hình 2.1 và Hình 2.2 là khác nhau về hình thức nhưng tương tự về bản chất toán học.



Hình 2.2. Mô hình phi tuyến thứ hai của một nơ ron

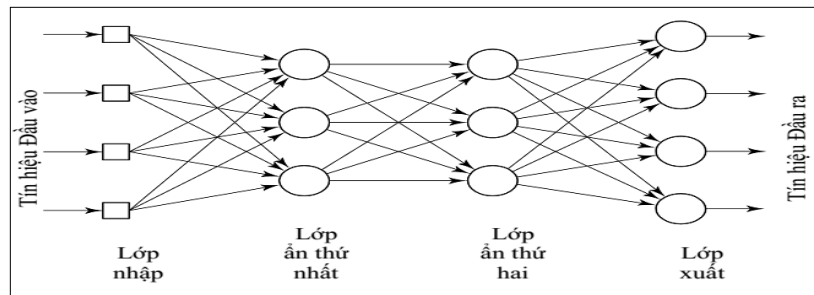
2.1.4. Phân loại mạng nơ ron nhân tạo

Có nhiều cách để phân loại mạng nơ ron nhưng để đơn giản, mạng nơ ron được phân loại theo hình trạng mạng.

Hình trạng của mạng được định nghĩa bởi: số lớp (layers), số đơn vị trên mỗi lớp, và sự liên kết giữa các lớp như thế nào. Các mạng về tổng thể được chia thành hai loại dựa trên cách thức liên kết các đơn vị: Mạng truyền thẳng và mạng hồi quy.

2.1.4.1. Mạng truyền thẳng

Dòng dữ liệu từ đơn vị đầu vào đến đơn vị đầu ra chỉ được truyền thẳng. Việc xử lý dữ liệu có thể mở rộng ra nhiều lớp, nhưng không có các liên kết phản hồi. Nghĩa là, các liên kết mở rộng từ các đơn vị đầu ra tới các đơn vị đầu vào trong cùng một lớp hay các lớp trước đó là không cho phép. Nếu mô hình hoá mạng truyền thẳng bằng một đồ thị, thì nó là một đồ thị có hướng hữu hạn không chu trình, trong đó, mỗi nơ ron là một nút, các liên kết giữa các nơ ron là các cung của đồ thị.



Hình 2.3. Mạng nơ ron truyền thẳng nhiều lớp

- Mạng truyền thẳng có một số đặc điểm sau:

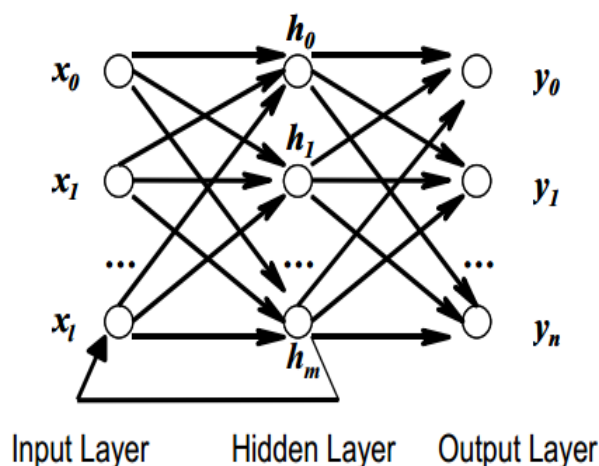
+ Các perceptron được đặt trong các lớp, lớp đầu tiên nhận dữ liệu đầu vào và lớp cuối cùng trả về dữ liệu đầu ra. Lớp ở giữa không có mối liên kết với các thành phần bên ngoài nên còn được gọi là lớp ẩn.

+ Mỗi perceptron trong một lớp có liên kết với perceptron ở lớp tiếp theo, thông tin được truyền thẳng theo một chiều từ lớp này sang lớp kế tiếp.

+ Các perceptron ở cùng một lớp không có liên kết với nhau.

2.1.4.2. Mạng hồi quy

Mạng hồi quy là kiểu kiến trúc mạng không chỉ truyền dữ liệu theo một chiều, mà còn truyền dữ liệu từ bước sau ngược trở lại bước trước. Mạng có các kết nối từ nơ ron đầu ra tới nơ ron đầu vào. Mạng lưu lại các trạng thái trước đó, và trạng thái tiếp theo không chỉ phụ thuộc vào các tín hiệu đầu vào mà còn phụ thuộc vào các trạng thái trước đó của mạng. Kiến trúc mạng này được phát triển từ những năm 1980. Mạng thường được sử dụng để giải quyết các bài toán phân cụm. Một số mạng nơ ron thuộc loại này như mạng Hopfield, Simple competitive network, Adaptive resonance theory network (ART), Kohonen self organizing maps (SOM). Nếu mô hình hoá mạng hồi quy bằng một đồ thị thì nó là một đồ thị có hướng hữu hạn có chu trình.



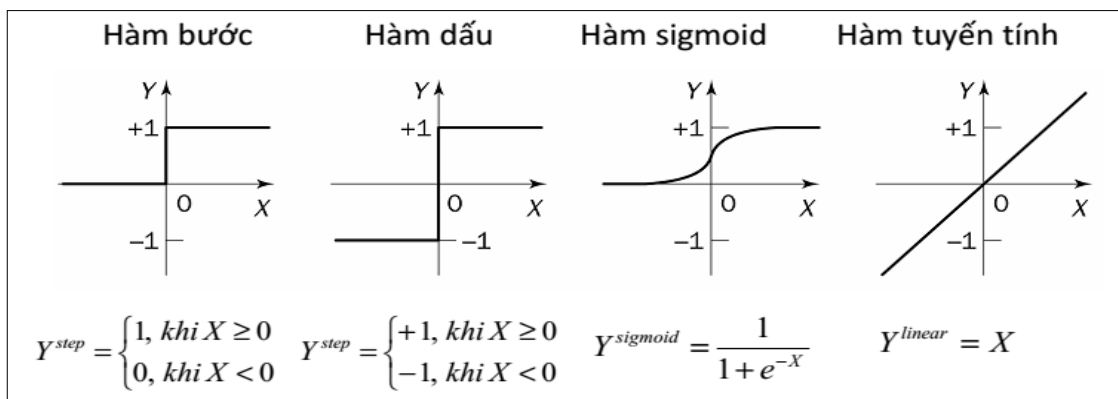
Hình 2.4. Mô hình mạng hồi quy

2.1.5. Hàm kích hoạt

Hàm kích hoạt, ký hiệu bởi $\varphi(v)$, được dùng để xác định đầu ra của mạng nơ ron. Các hàm kích hoạt phải có các đặc điểm sau:

- + Hàm bị chặn trên và chặn dưới
- + Hàm có tính đơn điệu
- + Hàm phải có tính liên tục

Dưới đây là một số kiểu hàm kích hoạt cơ bản thường được sử dụng:



Hình 2.5. Các hàm kích hoạt tiêu biểu

2.1.6. Phương pháp huấn luyện mạng nơ ron

Mạng nơ ron được xây dựng trên việc mô phỏng quá trình xử lý thông tin của bộ não người, do vậy đặc trưng cơ bản của mạng là có khả năng học, khả năng tái tạo lại và xử lý thông tin khi đã học. Ban đầu mạng nơ ron chưa có tri thức, tri thức của mạng nơ ron có được là hình thành qua mỗi lần học. Mạng nơ ron được dạy bằng cách đưa vào đầu vào những kích thích và mạng hình thành những đáp ứng tương ứng, nếu phù hợp với từng loại kích thích thì sẽ được lưu lại. Các kỹ thuật nhằm mục đích hiệu chỉnh các trọng số để mạng có thể nhận biết được quan hệ giữa đầu vào và đích mong muốn đó gọi là học (learning) hay huấn luyện (training). Còn cấu trúc của mạng như số lớp hay cách liên kết giữa các lớp, v.v là cố định trong suốt quá trình huấn luyện.

Có nhiều thuật toán được phát minh ra để tìm ra tập trọng số tối ưu cho các bài toán. Các thuật toán đó có thể chia làm 3 nhóm chính: học có giám sát (có mẫu),

học không có giám sát và học tăng cường. Mỗi kiểu học lại tương ứng với một nhiệm vụ học trừu tượng khác nhau.

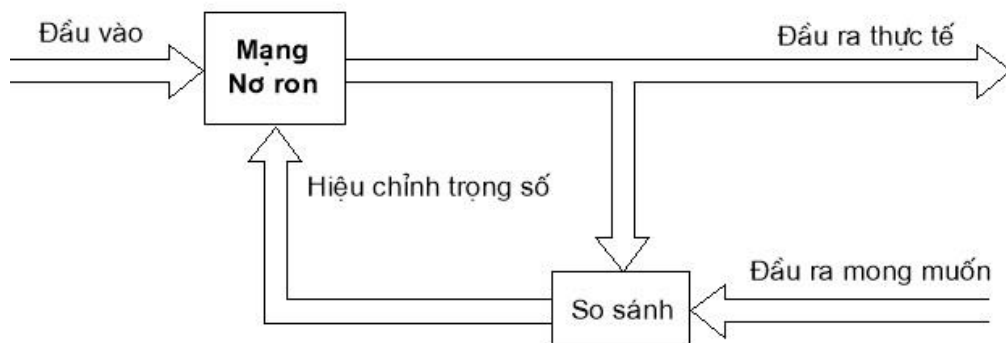
2.1.6.1. Học có giám sát

Phương pháp học có giám sát lần đầu được đề xuất bởi Donald Hebb vào năm 1949 và được ứng dụng chủ yếu cho các bài toán phân lớp, phân loại, nhận dạng. Điển hình cho kỹ thuật này là thuật toán lan truyền ngược (backpropagation). Với phương pháp này, cần đưa vào bộ dữ liệu huấn luyện và kết quả đầu ra mong muốn tương ứng. Thuật toán sẽ tiến hành tính toán trọng số để mỗi vector đầu vào sẽ có kết quả đầu ra đúng với kết quả đầu ra mong muốn. Quá trình học có giám sát cụ thể như sau:

- + Trước tiên cần chuẩn bị một tập dữ liệu huấn luyện được gán nhãn (Có dữ liệu đầu vào và đầu ra mong muốn tương ứng).

- + Lựa chọn một thuật toán phân lớp xây dựng bộ phân lớp và dùng tập dữ liệu huấn luyện để huấn luyện bộ phân lớp. Thuật ngữ học có giám sát được hiểu là học tập dữ liệu đã được gán nhãn trước (các dữ liệu kèm theo nhãn tương ứng này coi như đã được giám sát bởi người thực hiện gán nhãn). Trong quá trình học, trọng số của mạng liên tục được hiệu chỉnh qua mỗi bước lặp để tìm được bộ trọng số thỏa mãn nhất.

- + Sử dụng một tập dữ liệu kiểm tra đã được gán nhãn trước, để kiểm tra tính đúng đắn của mạng.



Hình 2.6. Mô hình huấn luyện mạng nơ ron học có giám sát

2.1.6.2. Học không có giám sát

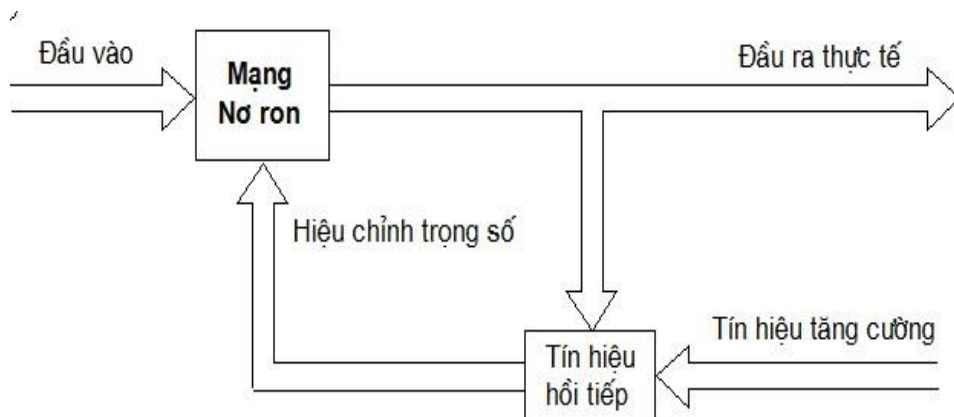
Học không có giám sát là học với tập dữ liệu huấn luyện ban đầu hoàn toàn chưa được gán nhãn. Đây là phương pháp học sử dụng cho lớp bài toán phân cụm. Quá trình huấn luyện của phương pháp học không có giám sát có đặc điểm như sau:

- + Sử dụng bộ dữ liệu huấn luyện chưa được gán nhãn (không có thông tin về giá trị đầu ra mong muốn).
- + Lựa chọn thuật toán học không có giám sát để huấn luyện.
- + Sử dụng phương pháp thử nghiệm để đánh giá chất lượng.

Mạng nơ ron điển hình được huấn luyện bằng phương pháp học không giám sát là bản đồ tự tổ chức (SOM - Self Organizing Map). Với phương pháp này, số lớp phân loại chưa được biết trước và tùy theo tiêu chuẩn đánh giá độ tương tự giữa các mẫu có thể có các lớp phân loại khác nhau.

2.1.6.3. Học tăng cường

Trong luật học có giám sát, các giá trị đầu ra được biết chính xác đối với mỗi đầu vào. Tuy nhiên, trong thực tế có một số trường hợp chỉ biết ít thông tin chi tiết, chẳng hạn mạng chỉ biết rằng giá trị đầu ra thực sự quá cao hay có thể mạng chỉ có được thông tin phản hồi báo rằng đầu ra đúng hay sai. Thuật học dựa trên thông tin đánh giá này được gọi là thuật học củng cố hay học tăng cường, thông tin phản hồi được gọi là tín hiệu tăng cường.



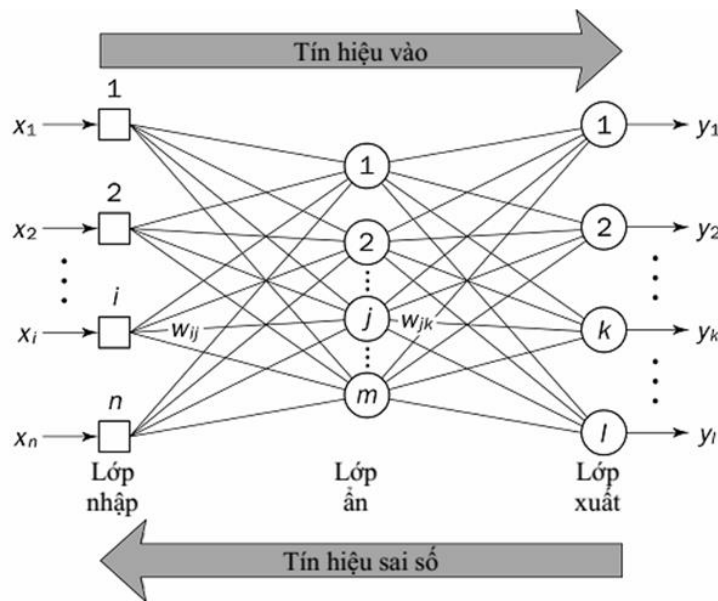
Hình 2.7. Mô hình mạng nơ ron học tăng cường

Các bài toán thường được giải quyết bằng học tăng cường là các bài toán điều khiển, trò chơi và các nhiệm vụ quyết định tuần tự (sequential decision making) khác.

2.1.7. Mạng nơ ron truyền thẳng một lớp ẩn và thuật toán lan truyền ngược

2.1.7.1. Mạng nơ ron truyền thẳng một lớp ẩn

Mạng nơ ron truyền thẳng một lớp ẩn là mạng nơ ron truyền thẳng gồm 3 lớp, trong đó có một lớp đầu vào, một lớp ẩn và một lớp đầu ra. Các nơ ron ở mỗi lớp liên kết với nơ ron ở lớp kế tiếp theo một chiều từ lớp đầu vào đến lớp ẩn và đến nơ ron lớp đầu ra, không có liên kết ngược lại hoặc liên kết giữa các nơ ron trong một lớp. Các nơ ron lớp ẩn hoạt động như bộ phát hiện đặc trưng, đóng vai trò quan trọng trong quá trình vận hành mạng nơ ron.



Hình 2.8. Mạng truyền thẳng một lớp ẩn

Trên thực tế, qua nghiên cứu cũng như áp dụng thực tiễn đã xác định được rằng việc sử dụng một lớp ẩn cho các mạng nơ ron là hợp lý. Việc huấn luyện mạng đối với mạng truyền thẳng một lớp ẩn cũng nhanh hơn và dễ dàng hơn rất nhiều.

2.1.7.2. Thuật toán lan truyền ngược

Phương pháp huấn luyện phổ biến cho mạng nơ ron truyền thẳng một lớp ẩn là thuật toán lan truyền ngược. Thuật toán lan truyền ngược là dạng tổng quát của

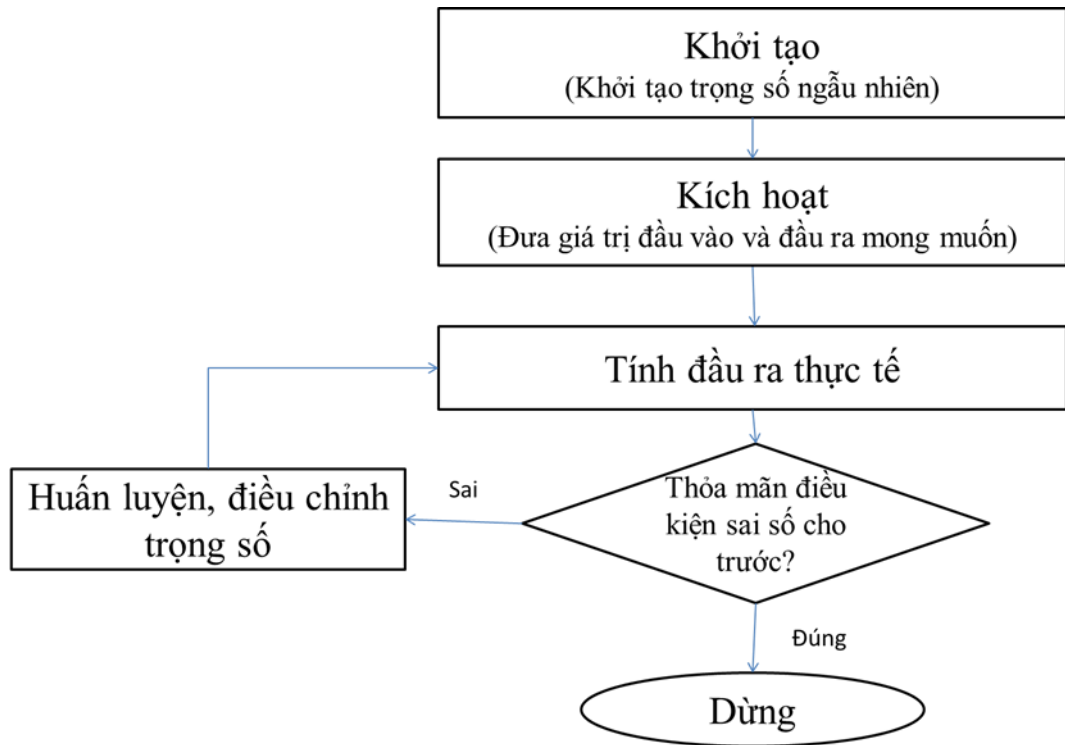
thuật toán trung bình phương tối thiểu (Least Means Square –LMS). Thuật toán này thuộc dạng thuật toán xấp xỉ để tìm các điểm mà tại đó hiệu năng của mạng là tối ưu. Chỉ số tối ưu (performance index) thường được xác định bởi một hàm số của ma trận trọng số và các đầu vào nào đó trong quá trình tìm hiểu bài toán đặt ra.

Huấn luyện mạng nơ ron sử dụng thuật toán Lan truyền ngược gồm hai quá trình: Quá trình truyền tuyến tính và quá trình truyền ngược:

Quá trình truyền tuyến tính: Các vector đầu vào sẽ được cung cấp cho các nơ ron của mạng, tín hiệu sẽ được lan truyền lần lượt qua các lớp mạng. Cuối cùng tính được tập các đầu ra thực sự của mạng. Trong suốt quá trình truyền thẳng, tất cả các trọng số liên kết của mạng được cố định. Từ đó, tìm ra sự khác nhau giữa giá trị thật hàm mẫu mà mạng tính được và kết quả dự đoán của mạng gọi là lỗi (học có giám sát).

Quá trình truyền ngược: Tất cả các trọng số liên kết được hiệu chỉnh theo các luật hiệu chỉnh trọng số. Giá trị lỗi hay gọi là sai số của mạng được tính bằng độ chênh lệch giữa đầu ra thu được với giá trị mong muốn. Giá trị lỗi sẽ được truyền ngược lại lần lượt qua các lớp mạng sao cho quá trình huấn luyện (học) sẽ tìm ra trọng số W_i để lỗi nhỏ nhất tức là làm cho giá trị đầu ra thực sự của mạng càng gần với các giá trị mong muốn càng tốt.

Thuật toán lan truyền ngược có thể được thực hiện thông qua các bước chính như sau:



Hình 2.9. Mô hình thuật toán lan truyền ngược

2.2. Ứng dụng mạng nơ ron vào bài toán nhận dạng chữ viết tay online

2.2.1. Thu thập dữ liệu

Chương trình được thực nghiệm trên bộ dữ liệu UNIPEN.

**Một số đặc điểm về cơ sở dữ liệu UNIPEN [16]:*

Nhận dạng chữ viết tay online giải quyết vấn đề nhận biết các dữ liệu được viết tay từ việc thu thập chúng qua các bộ đệm, cung cấp các thông tin về nét bút cụ thể. Nhận dạng chữ viết tay online từ lâu đã rất hạn chế về số lượng mẫu dữ liệu có sẵn công khai. Để khắc phục cho vấn đề này, dự án UNIPEN được bắt đầu vào năm 1992 theo sáng kiến của Ủy ban kỹ thuật 11 của Hiệp hội quốc tế về nhận dạng mẫu (IAPR). Trong lĩnh vực của nhận dạng chữ viết tay online tồn tại nhiều cơ sở dữ liệu cá nhân. Các cơ sở dữ liệu này tạo thành nguồn tài nguyên tiềm năng giàu có hơn bất kỳ một cơ sở dữ liệu nào, thể hiện ở sự đa dạng của văn bản được nhập vào, điều kiện ghi và các tác giả. Do đó trao đổi dữ liệu là cách tự nhiên để tạo thành một cơ sở dữ liệu lớn đại diện cho các nhiệm vụ mà các nhóm nghiên cứu và phát triển quan tâm. Năm 1993, định dạng UNIPEN được thiết kế, kết hợp các tính năng của

các định dạng nội bộ của một số tổ chức, bao gồm IBM, Apple, Microsoft, Slate (Jot), Hp, v.v.

Định dạng UNIPEN là định dạng ASCII được thiết kế đặc biệt cho dữ liệu được thu thập với bất kỳ thiết bị cảm ứng nhạy cảm, điện trở hay điện từ nào cung cấp thông tin quỹ đạo được viết bằng bút điện từ. Người dùng có thể chuyển đổi định dạng của riêng họ sang hoặc từ định dạng UNIPEN hoặc thu thập dữ liệu trực tiếp từ định dạng đó.

Số lượng tín hiệu tối thiểu là hai: X và Y, tuy nhiên vẫn cho phép nhiều tín hiệu hơn, ví dụ: thông tin về góc hoặc áp lực của bút. Ngược lại với định dạng nhị phân, chẳng hạn như Jot [2], định dạng UNIPEN không được tối ưu hóa để lưu trữ dữ liệu hoặc truyền dữ liệu thời gian thực. Định dạng UNIPEN có các điều khoản cho chú thích dữ liệu về điều kiện ghi, nhà văn, phân đoạn, bố cục dữ liệu, chất lượng dữ liệu, ghi nhãn, và kết quả nhận dạng.

Cơ sở dữ liệu được viết bằng định dạng UNIPEN có thể được kết hợp trong một mục đơn lẻ hoặc được tổ chức thành các thư mục. Định dạng có thể được coi là một chuỗi các tọa độ bút, được ghi nhận bằng các thông tin khác nhau, bao gồm phân đoạn và ghi nhãn. Quỹ đạo bút được mã hóa thành một chuỗi các thành phần. Một thành phần bút xuống là một dấu vết được ghi lại khi bút tiếp xúc với bề mặt của bộ số hóa. Một thành phần bút lên là một dấu vết được ghi lại khi bút nằm gần bộ số hóa mà không chạm vào nó. Các thuộc tính PENUP - PENDOWN chứa tọa độ bút. Cơ sở dữ liệu được chia thành một hoặc nhiều tập dữ liệu bắt đầu bằng START SET. Trong một tập hợp, các thành phần được đánh số ngầm từ 0. Phân đoạn và ghi nhãn được cung cấp bởi: hướng dẫn SEGMENT. Các số kết hợp được sử dụng bởi: SEGMENT để phân định các câu, từ và các chữ. Phân cấp phân đoạn (ví dụ: SENTENCE W lệ phí tự do W) được khai báo với: HIERARCHY. Bởi vì các thành phần được tham chiếu bởi một sự kết hợp độc đáo của tên tập và số thứ tự trong tập hợp đó, có thể tách SEGMENT từ chính dữ liệu.

Tuy nhiên với mức độ nhận dạng ký tự của luận văn đề ra, bộ cơ sở dữ liệu UNIPEN được sử dụng cung cấp các mẫu chứa tọa độ bút và chứa các thông tin mà

người viết đã viết (tức là người viết đã viết những chữ gì, với luận văn là người viết đã viết những ký tự gì) để phục vụ cho quá trình huấn luyện và nhận dạng ký tự.

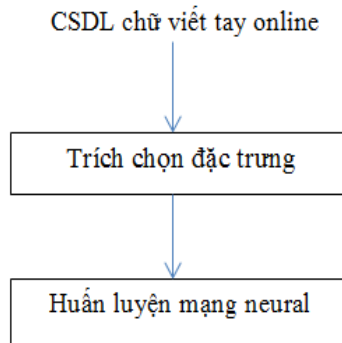
Chương trình thực nghiệm được xây dựng với bộ cơ sở dữ liệu có 50000 mẫu được sử dụng, trong đó có 70% dùng cho bộ Train và 30% cho bộ Test, bao gồm nhiều mẫu viết của nhiều người khác nhau trên thế giới được lựa chọn ngẫu nhiên để đưa vào huấn luyện.

2.2.2. Mô hình của bài toán nhận dạng chữ viết tay online sử dụng mạng nơ ron nhân tạo

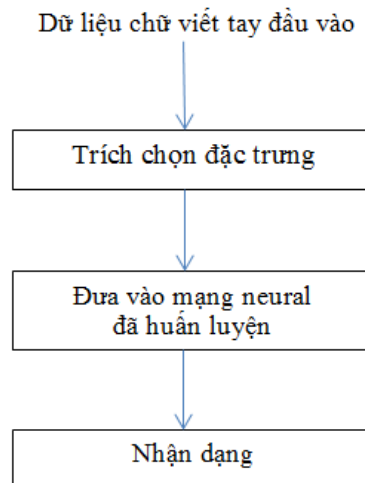
Cho đến nay, việc nhận dạng chữ viết tay vẫn chưa có một giải pháp tổng thể, các ứng dụng của nó chỉ giới hạn trong phạm vi hẹp. Các kết quả chủ yếu về lĩnh vực nhận dạng chữ số và một số hệ chữ cái la tinh, Hy Lạp, Trung Quốc, v.v. Tuy nhiên các kết quả đạt được còn nhiều hạn chế [6]. Các giải pháp tiếp cận để giải bài toán nhận dạng chữ viết tay khá phong phú, một số phương pháp học máy thường được sử dụng như: mô hình Markov ẩn, mạng nơ ron nhân tạo hay phương pháp máy vector tựa. Trong đó, phương pháp mạng nơ ron nhân tạo được đánh giá là phương pháp học máy tiên tiến đang được nghiên cứu cũng như áp dụng rộng rãi.

Với các đặc điểm nổi trội của mạng nơ ron nhân tạo trong báo cáo này, tôi xây dựng mô hình nhận dạng chữ viết tay online dựa trên phương pháp mạng nơ ron nhân tạo, mà trọng tâm là nhận dạng ký tự. Mô hình tổng quát của hệ thống nhận dạng chữ viết tay online sử dụng mạng nơ ron nhân tạo được chia thành hai giai đoạn: Giai đoạn huấn luyện và giai đoạn nhận dạng.

- Giai đoạn huấn luyện gồm các bước chính: Lựa chọn cơ sở dữ liệu huấn luyện. Với từng file cơ sở dữ liệu tiến hành trích chọn đặc trưng rồi đưa vào mạng nơ ron để huấn luyện.



Hình 2.10. Giai đoạn huấn luyện mạng



Hình 2.11. Giai đoạn nhận dạng chữ viết tay online

- Giai đoạn nhận dạng: vẽ/viết ký tự lên màn hình thiết bị nhận dạng, sau đó tiến hành trích chọn đặc trưng. Những đặc trưng này sẽ được đưa vào mạng nơ ron đã huấn luyện ở trên để tiến hành nhận dạng.

2.2.3. Trích chọn đặc trưng và xây dựng vector đầu vào

Trích chọn đặc trưng là quá trình tìm ra các thông tin hữu ích và đặc trưng nhất cho mẫu đầu vào để sử dụng cho quá trình huấn luyện và nhận dạng. Do mỗi ký tự viết tay online, số lượng các điểm trên nét viết là không giống nhau. Trong khi đầu vào của mạng nơ ron yêu cầu phải là một số lượng các giá trị nhất định. Do đó cần tiến hành xử lý để chọn được đầu vào thích hợp đưa vào mạng.

Trong bài toán nhận dạng chữ viết tay online dữ liệu được thao tác là dãy các điểm thì cách trích chọn đặc trưng trực quan nhất là dựa trên sự thay đổi về quan hệ hình học giữa hai điểm hoặc dãy các điểm liên tiếp. Các quan hệ này thường là sự thay đổi theo các chiều tọa độ, hoặc là sự thay đổi về góc tiếp tuyến. Nhưng điểm yếu của cách trích chọn này nằm ở tính cục bộ của nó, chỉ xét đến quan hệ hạn chế giữa các điểm, không thể thể hiện được đặc trưng của toàn bộ nét bút. Trong những cách tiếp cận khác mức cao hơn, đặc trưng được thể hiện bằng hình dạng của các nét chẳng hạn các nét lên, nét xuống, nét nổi thành hình tròn.

Với luận văn, việc trích chọn đặc trưng các mẫu đưa vào được thực hiện dựa trên việc đếm số điểm đen trên mặt phẳng nơi nét bút đi qua, từ đó mạng nơ ron lưu lại các đặc trưng cơ bản này cho những lần trích chọn đặc trưng khác của các mẫu khác với cùng một ký tự được nhận dạng và của các ký tự khác nhau. Qua nhiều vòng lặp trích chọn được nhiều đặc trưng của tất cả các file trong thư mục huấn luyện. Mỗi mẫu được đưa vào huấn luyện là tập hợp các tọa độ (x,y) lưu lại quá trình di chuyển của nét bút trên mặt phẳng các thiết bị điện tử. Với mỗi lần trích chọn sẽ thực hiện việc tìm kiếm các giá trị tọa độ \min_x và \min_y , sau đó tiến hành trừ tất cả các tọa độ của mẫu viết cho các giá trị này để quy các tọa độ về gốc tọa độ $(0,0)$ trong mặt phẳng Oxy. Tiếp tục chia tọa độ \max_x và \max_y cho một giá trị được quy ước trước. Với luận văn được xây dựng thì tọa độ x được chia thành m phần và tọa độ y được chia thành n phần. Sau bước thực hiện này, tiếp tục đếm số điểm có trong từng phần, đây là các đặc trưng thu được sau quá trình trích chọn. Bên cạnh đó, thuộc tính số .SEGMENT cũng có thể được tính toán để biểu diễn số nét bút thực hiện viết mẫu đang được xét này.

Để xác định được vector đầu vào tốt nhất, chương trình đã tiến hành thực nghiệm sử dụng số lượng điểm vector đầu vào khác nhau với các giá trị m và n được lựa chọn ngẫu nhiên và so sánh, đánh giá kết quả giữa các vector này trên cùng một cấu hình của mạng nơ ron.

Bảng 2.1. Kết quả thực nghiệm với vector đầu vào khác nhau

m	n	Số phần tử của vector đầu vào ($m*n$ hoặc $m*n+1$)	Kết quả nhận dạng
4	2	8	40%
4	2	9	45%
5	2	10	58%
5	2	11	82%
3	4	13	76%
3	4	12	73%
7	2	15	64%
7	2	14	62%
4	4	17	54%
4	4	16	53%
8	2	17	49%
8	2	16	45%

Kết quả thực nghiệm thể hiện trong Bảng 2.1 cho thấy vector đầu vào có kích thước 11 phần tử cho kết quả nhận dạng tốt nhất. Do đó luận văn đề xuất sử dụng vector đầu vào gồm 11 phần tử.

Để nâng cao hiệu quả của mạng nơ ron, các phần tử trong vector đầu vào sẽ được chuẩn hóa để nhận giá trị trong khoảng từ 0 đến 1 bằng cách chia các phần tử của vector đầu vào cho tổng các giá trị của chúng.

Bên cạnh các đặc trưng được sử dụng để làm đầu vào cho mạng nơ ron, để phục vụ cho bước huấn luyện mạng, hệ thống sẽ lưu thông tin xem ký tự đưa vào nhận dạng là ký tự nào. Thông tin này sẽ được lưu thành một vector đầu ra tương ứng với các vector đặc trưng đầu vào của mạng nơ ron.

2.2.4. Các tham số của mạng nơ ron.

2.2.4.1. Số phần tử ở các lớp

- Lớp đầu vào:

Số lượng node đầu vào của mạng nơ ron tương ứng với số lượng phần tử của vector đầu vào được trình bày trong vector đầu vào. Số lượng node được lựa chọn là 11.

- Lớp ẩn: Số lượng node ở tầng ẩn được lựa chọn ngẫu nhiên thỏa mãn điều kiện:

+ Số lượng node tầng ẩn phải nằm giữa kích thước của lớp đầu vào và kích thước của lớp đầu ra.

+ Số lượng node ẩn phải bằng $\frac{2}{3}$ kích thước của lớp đầu vào, cộng với kích thước của lớp đầu ra.

+ Số lượng node ẩn phải nhỏ hơn gấp đôi kích thước của lớp đầu vào.

- Lớp đầu ra: Số lượng node của lớp đầu ra tương ứng với số lượng ký tự khác nhau được đưa vào huấn luyện.

2.2.4.2. Hệ số học (Learning Rate)

Learning rate hay hệ số học là một trong những thông số mà điều chỉnh để một mạng nơ ron học nhanh và huấn luyện hiệu quả.

Hãy xem xét một nơ ron đang trải qua quá trình học tập. Giả định rằng trọng số của một liên kết trong mạng một phần được đào tạo là 0,3. Khi mạng được giới thiệu một mẫu huấn luyện mới, thuật toán huấn luyện yêu cầu thay đổi trọng số của nó đến 0,7 để nó có thể học các mẫu mới phù hợp. Nếu cập nhật trọng số ngay lập tức, các mạng nơ ron chắc chắn sẽ học các mẫu mới, nhưng nó có xu hướng quên đi tất cả các mẫu nó đã học trước đó. Điều này là do trọng số hiện tại (0,3) là do kết quả của tất cả việc học mà nó đã trải qua cho đến nay.

Vì vậy, không trực tiếp thay đổi trọng số tới 0,7. Thay vào đó, tăng nó bởi một phần nhỏ (chọn 25%) của sự thay đổi cần thiết. Từ đó, trọng số liên kết được thay đổi thành 0,4 và chuyển sang mẫu đào tạo tiếp theo. Yếu tố này (ở đây là 0,25) được gọi là Learning rate. Căn cứ theo cách này, tất cả các mẫu huấn luyện được

huấn luyện theo thứ tự ngẫu nhiên. Khi chu trình đào tạo được lặp đi lặp lại nhiều lần, cuối cùng mạng nơ ron học tất cả các mẫu có hiệu quả.

Learning rate là một giá trị trong khoảng 0 đến 1. Chọn một giá trị rất gần bằng 0 thì chương trình sẽ đòi hỏi một số lượng lớn các chu trình huấn luyện. Điều này làm cho quá trình huấn luyện rất chậm nhưng chắc chắn. Mặt khác, nếu learning rate lớn, trọng số khác nhau và độ lệch hàm mục tiêu dao động lớn và mạng đạt đến một trạng thái mà việc huấn luyện diễn ra vô ích. Với luận văn, hệ số học được chọn giá trị ngẫu nhiên trong khoảng từ 0 đến 1.

2.2.4.3. Bước đà (Momentum)

Việc sử dụng Momentum (bước đà) có tác dụng giảm các hậu quả hạn chế từ việc sử dụng hệ số học không như muốn, có nghĩa là bước đà làm tăng tốc độ huấn luyện khi hệ số học quá nhỏ khiến quá trình huấn luyện quá chậm hoặc giảm nguy cơ dao động khi sử dụng hệ số học quá lớn.

Về bản chất, trong phần đầu của quá trình huấn luyện, bước đà chỉ có tác dụng giúp quá trình trở nên nhanh hơn. Sau đó khi quá trình huấn luyện gần đạt tới hội tụ.

Nhược điểm của bước đà là giá trị của nó là ngẫu nhiên và phải được sử dụng thử và sai để tìm ra giá trị tốt.

2.2.5. Huấn luyện mạng nơ ron

Mạng nơ ron được sử dụng trong báo cáo luận văn là mạng nơ ron truyền thẳng một lớp ẩn. Trong đó lớp đầu vào và lớp ẩn có số lượng nodes cho phép tùy chỉnh được, để sau quá trình kiểm tra độ chính xác của mạng nơ ron có thể xác định được cấu hình mạng tốt nhất. Lớp đầu ra có số nodes bằng với số lượng người khác nhau cung cấp mẫu chữ viết. Mỗi nơ ron trong một lớp có liên kết trọng số tới mọi nơ ron ở lớp kế tiếp và không có liên kết ngược lại. Các trọng số của mạng ban đầu được khởi tạo ngẫu nhiên và được điều chỉnh khi mạng nơ ron được huấn luyện bởi cơ chế học có giám sát sử dụng thuật toán học lan truyền ngược.

Ở bước huấn luyện, mạng nơ ron sử dụng thuật toán lan truyền ngược backpropagation. Ban đầu trọng số của mạng được đặt ngẫu nhiên cho từng nơ ron. Sau đó vector đầu vào và vector đầu ra tương ứng được đưa vào mạng. Hệ thống thực hiện tính đầu ra thực tế của các nơ ron trong lớp ẩn, so sánh với đầu ra mong muốn (đầu ra tương ứng với vector đầu vào), tính sai số và thực hiện bước huấn luyện trọng số. Mạng nơ ron cập nhật trọng số trong mạng lan truyền ngược bằng cách lan truyền ngược sai số ứng với các nơ ron đầu ra. Sai số này được sử dụng để sửa lỗi trọng số và cập nhật lại trọng số tại các nơ ron đầu ra.

Bước tính sai số và hiệu chỉnh trọng số sẽ lặp lại cho đến khi thỏa mãn điều kiện sai số cho trước. Khi giá trị trung bình bình phương sai số (Mean Squared Error - MSE) này trên tất cả tập huấn luyện nếu đạt giá trị đủ nhỏ, mạng lúc này được xem như đã hội tụ. Điều kiện dừng thứ hai là số lượng vòng lặp tối đa người dùng thiết lập cho mạng. Khi mạng chạy hết số lượng vòng lặp tối đa này mà vẫn chưa hội tụ (chưa thỏa mãn điều kiện sai số cho trước) thì mạng sẽ dừng, không hiệu chỉnh trọng số nữa.

2.2.6. Nhận dạng chữ viết tay online sử dụng mạng nơ ron

Trong giai đoạn nhận dạng, dữ liệu chữ viết tay thu được sẽ được trích chọn đặc trưng. Những đặc trưng này sau đó được đưa vào mạng nơ ron đã được trải qua quá trình huấn luyện để tiến hành nhận dạng chữ. Với mỗi dữ liệu chữ viết đầu vào, kết quả nhận dạng có thể xảy ra một trong bốn trường hợp sau: Nhận dạng đúng ký tự, nhận dạng sau ký tự (kết quả trả về là ký tự khác), không nhận dạng được ký tự do dữ liệu về ký tự chưa có trong cơ sở dữ liệu, không nhận dạng được ký tự mặc dù ký tự này đã có trong cơ sở dữ liệu. Kết quả của bước nhận dạng ký tự sẽ được sử dụng để đánh giá hiệu năng của mạng nơ ron trong nhận dạng chữ viết tay online.

2.3. Kết luận chương 2

Chương 2 trình bày khái quát về mạng nơ ron nhân tạo, về đặc trưng, kiến trúc, các phương pháp huấn luyện mạng và đặc biệt tìm hiểu về mạng nơ ron truyền thẳng một lớp ẩn và thuật toán lan truyền ngược.

Tiếp đó, trình bày 02 mô hình nhận dạng trong quá trình huấn luyện và quá trình nhận dạng chữ. Sau đó là trình bày về quá trình trích chọn các đặc trưng của chữ viết và các tham số đặc trưng trong quá trình huấn luyện và nhận dạng. Quá trình trích chọn đặc trưng được khai thác trên các tọa độ (x,y) trên mặt phẳng mà nét bút đi qua, đồng thời có thêm thông tin của số nét bút.

Luận văn đã xây dựng và tiến hành thực nghiệm và cho kết quả về số lượng tham số đầu vào hợp lý nhất. Cuối cùng, đưa ra cái nhìn khái quát về quá trình huấn luyện và nhận dạng chữ viết online sử dụng mạng nơ ron nhân tạo. Nội dung các chương 1 và chương 2 là tiền đề cho việc xây dựng chương trình thực nghiệm nhận dạng chữ viết tay online ở mức độ nhận dạng ký tự được trình bày ở chương 3 của luận văn.

Chương 3. CÀI ĐẶT THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

Chương 3 giới thiệu về chương trình thực nghiệm nhận dạng chữ viết tay online sử dụng mạng nơ ron tạo. Từ đó đưa ra các đánh giá, nhận xét về hiệu quả của việc sử dụng mạng nơ ron nhân tạo để nhận dạng chữ viết tay online trên bộ cơ sở dữ liệu Unipen với các tham số cấu hình mạng nơ ron khác nhau.

3.1. Cài đặt thực nghiệm

3.1.1. Ngôn ngữ lập trình

Chương trình thực nghiệm nhận dạng chữ viết tay online sử dụng mạng nơ ron được lập trình bằng ngôn ngữ lập trình C# trên Visual Studio 2013.

3.1.2. Cấu hình máy

Cấu hình máy tính được sử dụng để cài đặt và chạy thử nghiệm chương trình như sau:

Bộ vi xử lý: Intel(R) Core(TM) i3-3120M

CPU: 2.50GHz

RAM: 6GB

Hệ điều hành: Windows 7 Ultimate.

3.1.3. Cấu hình mạng nơ ron

Chương trình được thử nghiệm trên mạng nơ ron truyền thẳng 1 lớp ẩn, với thuật toán học là thuật toán lan truyền ngược (backpropagation).

- Số node input: 11.

- Các tham số còn lại gồm: Số node tầng ẩn, số lượng vòng lặp, ngưỡng MSE, momentum (alpha) và learning rate (beta) đều cho phép tùy chỉnh thay đổi để tìm ra cấu hình mạng tốt nhất.

- Alpha (momentum): Nằm trong phạm vi $0 \leq \alpha < 1$.

- Beta (learning rate): Nằm trong phạm vi $0 \leq \beta < 1$.

3.1.4. Giới thiệu chương trình

Chương trình gồm các chức năng chính như sau:

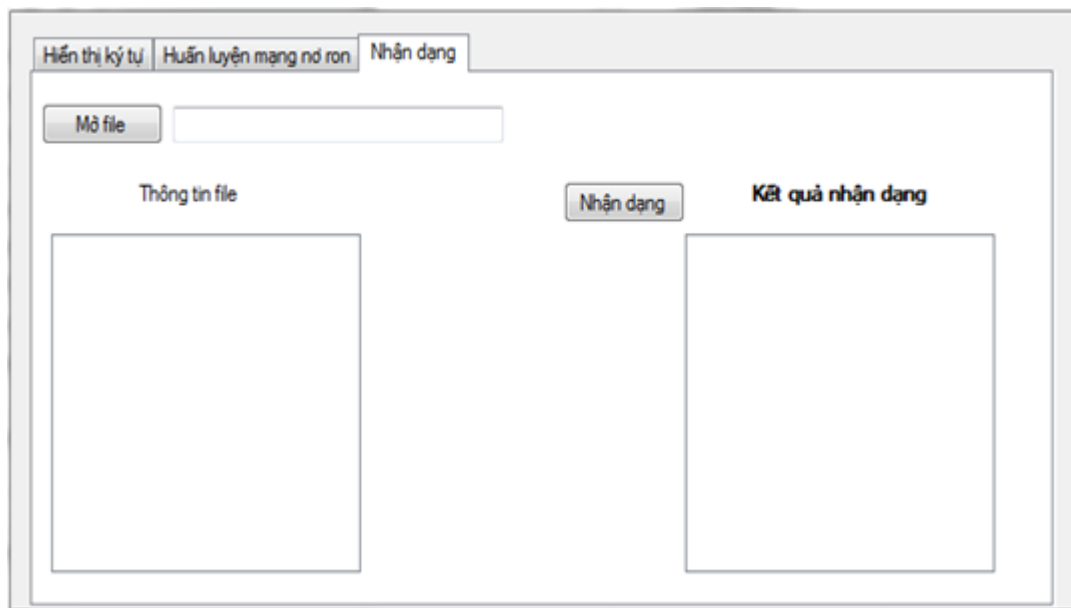
- Huấn luyện mạng nơ ron:

Hình 3.1. Giao diện huấn luyện mạng nơ ron

- Lưu mạng đã huấn luyện:

Hình 3.2. Các giá trị của mạng được lưu lại sau huấn luyện

- Nhận dạng chữ viết trong cơ sở dữ liệu UNIPEN:

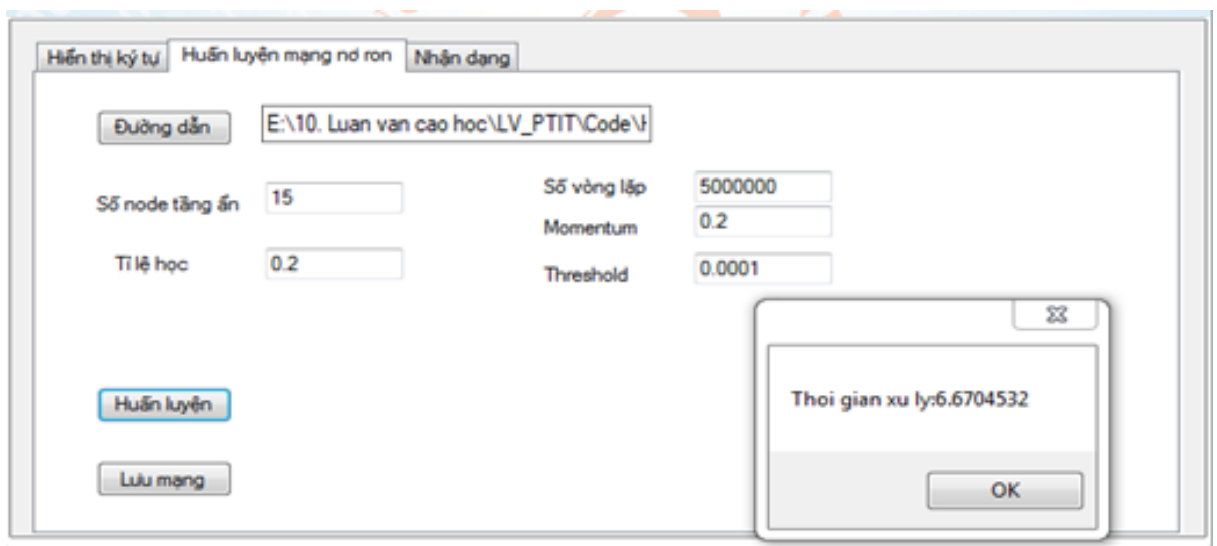


Hình 3.3. Giao diện nhận dạng chữ viết online

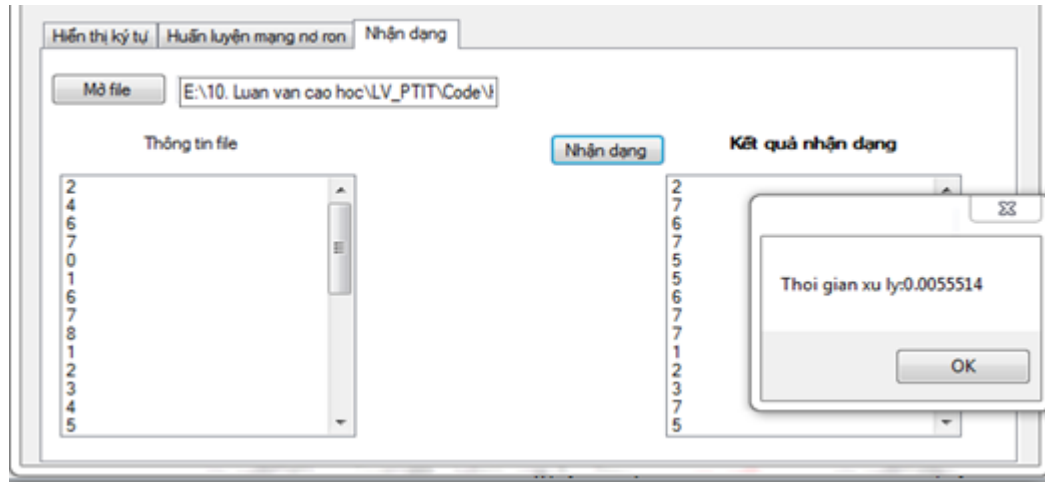
3.2. Đánh giá kết quả thực nghiệm

3.2.1. Kết quả thực nghiệm

Chương trình được xây dựng sử dụng thuật toán lan truyền ngược trong mạng nơ ron nhân tạo. Thời gian trung bình để xử lý một file dữ liệu và trích chọn đặc trưng là 0.15 giây và thời gian huấn luyện trung bình là 7 giây, thời gian nhận dạng trong khoảng xấp xỉ 0.005 giây.



Hình 3.4. Demo thời gian huấn luyện



Hình 3.5. Demo minh họa thời gian nhận dạng

3.2.2. Đánh giá kết quả

Nhận dạng được thực hiện bằng cách, chọn ngẫu nhiên một thư mục trong bộ Train có chứa các mẫu để huấn luyện. Sau khi được huấn luyện, chương trình sẽ sử dụng các mẫu trong bộ Test để kiểm tra kết quả.

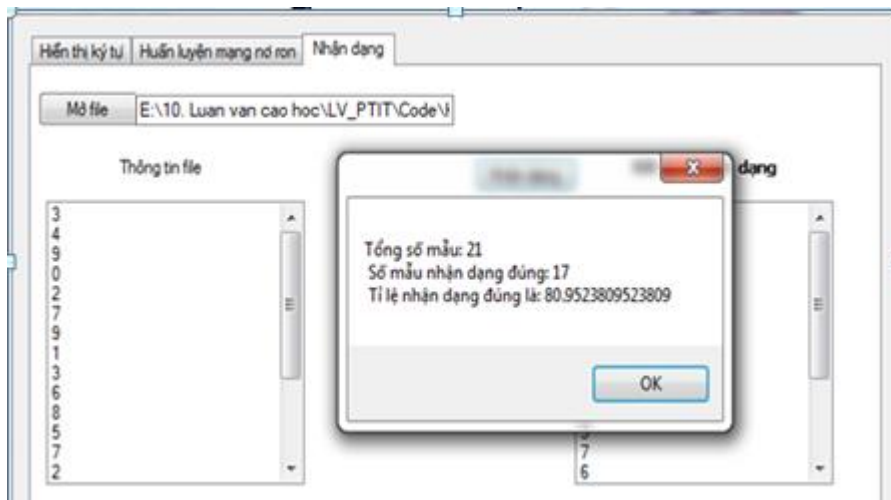
Kết quả nhận dạng cho từng ký tự:

Bảng 3.1. Thống kê kết quả nhận dạng ký tự

Ký tự	Số lượng mẫu để kiểm tra	Kết quả nhận dạng đúng
a	547	0.84
b	269	0.81
c	250	0.95
d	672	0.88
e	582	0.93
f	182	0.82
g	240	0.92
h	250	0.83
i	578	0.73
j	283	0.74
k	238	0.78

l	243	0.64
m	248	0.91
n	237	0.87
o	547	0.92
p	234	0.87
q	256	0.82
r	243	0.37
s	239	0.66
t	245	0.72
u	579	0.66
v	262	0.88
w	178	0.90
x	256	0.84
y	218	0.82
z	154	0.74
0	231	0.76
1	167	0.78
2	342	0.67
3	332	0.68
4	367	0.79
5	453	0.87
6	342	0.85
7	321	0.78
8	423	0.81
9	123	0.77
Tỷ lệ trung bình		0.82

Kết quả nhận dạng khi load file chứa nhiều ký tự khác nhau:



Hình 3.6. Kết quả nhận dạng khi load file nhiều ký tự

Sau khi được thực nghiệm với nhiều cấu hình khác nhau của mạng nơ ron, lựa chọn được cấu hình hợp lý nhất của mạng như sau:

- + Số nơ ron tầng ẩn: 15
- + Momentum: 0.02
- + Tỷ lệ học: 0.2
- + Threshold: 0.0001
- + Số vòng lặp: 5000000

Với cấu hình nhận dạng này, chương trình có thể cho độ chính xác đạt tới 80%.

3.2.3. Đánh giá kết quả

Sau khi thu được các kết quả thực nghiệm, cho thấy tỷ lệ nhận dạng trung bình đạt 80%. Trong đó, phương pháp nhận dạng chữ viết tay sử dụng mạng nơ ron nhân tạo với thuật toán lan truyền ngược, đạt được kết quả tốt ở các ký tự “c”, “d”, “e”, “o”, “p”, “q”, “1”, “3”, “6”, “5”, “9”. Một số ký tự khác như “l” thường nhận nhầm thành số “1”, “u” thành “v”, “v” thành “u” hay các số “4” và “1” cũng hay bị nhận dạng nhầm với nhau v.v.

Kết quả này chưa thực sự tốt do một số nguyên nhân sau:

- Số lượng mẫu đưa vào huấn luyện còn hạn chế, chưa đa dạng. Có những ký tự chưa được huấn luyện dẫn đến khi được đưa vào nhận dạng sai thành ký tự khác hoặc không ra kết quả nhận dạng.

- Các tham số của mạng như: số vòng lặp, tỷ lệ học, v.v chưa hợp lý, chưa đem lại hiệu quả nhận dạng chính xác, thời gian huấn luyện và nhận dạng còn chưa tối ưu.

- Bản chất các ký tự nhận dạng có nhiều điểm giống nhau như: số nét bút vẽ, độ nghiêng, tỷ lệ các nét, v.v khiến quá trình nhận dạng dễ xảy ra nhận dạng nhầm giữa các ký tự này.

- Do phong cách viết đa dạng của người viết. Mỗi ký tự đối với mỗi người trong những hoàn cảnh khác nhau lại cho những cách viết khác nhau. Đôi khi do sự không cẩn thận, viết ẩu, hoặc thiếu nét cũng là những nguyên nhân dẫn đến nhận dạng sai ký tự.

3.3. Kết luận chương 3

Nội dung chương 3 trình bày về quá trình nhận dạng chữ viết online ở mức độ nhận dạng ký tự được xây dựng trong chương trình mô phỏng. Bằng các kiến thức được nghiên cứu và tìm hiểu ở chương 1 và chương 2, chương 3 đã xây dựng thành công phần mềm nhận dạng chữ viết tay online ở cấp nhận dạng ký tự sử dụng mạng nơ ron nhân tạo trên bộ cơ sở dữ liệu UNIPEN.

KẾT LUẬN

1. Các kết quả nghiên cứu của luận văn

Về mặt lý thuyết, luận văn đã trình bày được một số nội dung cơ bản sau:

- Giới thiệu tổng quan về bài toán nhận dạng chữ viết tay, trong đó trọng tâm là bài toán nhận dạng chữ viết tay online, những đặc trưng cơ bản của chữ viết online và các phương pháp thu thập chữ viết tay online.

- Giới thiệu tổng quan về mạng nơ ron, kiến trúc và các phương pháp huấn luyện mạng nơ ron. Cụ thể về mạng nơ ron truyền thẳng một lớp ẩn và thuật toán lan truyền ngược.

Về mặt thực nghiệm:

- Xây dựng được một mạng nơ ron truyền thẳng một lớp ẩn sử dụng thuật toán lan truyền ngược ứng dụng nhận dạng chữ viết tay online ở mức độ nhận dạng ký tự sử dụng cơ sở dữ liệu UNIPEN.

2. Hướng phát triển của luận văn

- Tiếp tục cải thiện và kết hợp nhiều thuật toán để có thể trực tiếp nhận dạng được chữ viết online với các mức độ cao hơn (nhận dạng được chữ viết, câu văn, đoạn văn v.v) với độ chính xác cao hơn.

- Nghiên cứu thêm các thuật toán trích chọn đặc trưng để tỷ lệ nhận dạng đúng cao hơn.

- Xây dựng được phần mềm nhận dạng trực tiếp từ quá trình viết lên các thiết bị điện tử.

TÀI LIỆU THAM KHẢO

Tiếng Việt

- [1] Nguyễn Hoàng Hải (2012), *Hệ thống nhận dạng chữ viết tay in hoa trực tuyến*, Luận văn thạc sĩ khoa học, tr. 14 – 15.

Tiếng Anh

- [2] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber (2009), A novel connectionist system for unconstrained handwriting recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, Tập 31, Số 5, tr. 855-868.
- [3] A. Delaye and E'. Anquetil (2013), HBF49 feature set: A first unified baseline for online symbol recognition, in *Proc. Pattern Recognit*, Tập 46, Số 1, tr. 117–130.
- [4] C. C. Tappert, C. Y. Suen, and T. Wakahara (1990), “The state of the art in online handwriting recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, Tập 12, Số 8, tr. 787–808.
- [5] Daniel Keysers, Thomas Deselaers, Henry A. Rowley, Li-Lun Wang, and Victor Carbune, (2017) Multi-Language Online Handwriting Recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, Tập 39, Số 6, tr. 63–84.
- [6] G. Nagy (2000), Twenty Years of Document Image Analysis in PAMI, *IEEE Trans. Pattern Analysis and Machine Intelligence*, tr. 135-138.
- [7] G. Vamvakas, B. Gatos, I. Pratikakis, N. Stamatopoulos, 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*, tr. 197-202.
- [8] J. A. Pittman (2007), Handwriting recognition: Tablet PC text input, *IEEE Comput.*, Tập 40, Số 9, tr. 49–54.
- [9] J. H. Kim and B. Sin (2014), Online handwriting recognition, in *Proc. Handbook Doc. Image Process. Recognit.*, tr. 887–915.

- [10] J.Makhoul, T. Starner, R. Schwartz, and G. Chou (1994), Online Cursive Handwriting Recognition Using Speech Recognition Methods, *Proc. ICASSP'94*, tr. 125-128.
- [11] J.Schenk and G. Rigoll (2006), Novel hybrid NN/HMM modelling techniques for on-line handwriting recognition, *in Proc. Int. Workshop Frontiers Handwriting Recognit*, tr. 619–623.
- [12] K.-F. Chan, D.-Y. Yeung (1998), Elastic Structural Matching for Online Handwritten Alphanumeric Character Recognition, *Proc. 14th Int. Conf. Pattern Recognition, Brisbane, Australlia*, tr. 1508-1511.
- [13] L.Yaeger, B.Webb, and R.Lyon (1998), Combining nơ ron networks and context-driven search for on-line, printed handwriting recog- nition in the Newton, *in AAAI AI Magazine. Berlin, Germany, Springer*.
- [14] Marcus Liwicki and Horst Bunke (2008), RECOGNITION OF WHITEBOARD NOTES Online, Offline and Combination, tr. 129 -130.
- [15] R. Plamondon and S. N. Srihari (200), On-line and off-line handwriting recognition: A comprehensive survey, *IEEE Trans. Pattern Anal. Mach. Intell.*, Tập 22, Số 1, tr. 63–84.
- [16] S. Jaeger, S. Manke, J. Reichert, and A. Waibel (2001), Online handwrit- ing recognition: The NPen++ recognizer, *Int. J. Doc. Anal. Recog- nit.*, Tập 3, Số 3, tr. 169–180.
- [17] Simon Haykin (2009), Nơ ron networks and learning machines 3rd edition, *Pearson education*, tr. 34 – 37.