

MẠNG NƠ RON LAN TRUYỀN NGƯỢC HƯỚNG CHO NHẬN DẠNG KÝ TỰ

COUNTER - PROPAGATION NETWORK FOR CHARACTER RECOGNITION

TRẦN NGỌC HÀ¹

NGUYỄN THANH THUY²

Abstract. Much effort has been extended in making computer recognize characters automatically. Methods currently widely used for character recognition are mainly involved in pattern matching using image processing techniques. One limitation of such methods is their inability to respond to variations.

This paper presents the use of counter - propagation network based on Kohonen network for character recognition. A first attempt at recognizing 37 symbols has been carried out. Obtained results showed that the system is very robust, stable and still capable to retrieve correctly characters for 25% corrupted images.

1. MỞ ĐẦU

Nhận dạng là lĩnh vực được áp dụng rộng rãi trong thực tế, trong đó nhận dạng chữ là vấn đề được quan tâm ngày càng cao do nhu cầu dịch và xử lý các văn bản một cách tự động. Quá trình xử lý văn bản gồm các bước cơ bản sau:

- Số hoá văn bản bằng scanner, sau đó lưu trữ dưới dạng tệp ảnh
- Xử lý sơ bộ như tăng cường ảnh, tách vùng chữ ra khỏi văn bản, sau đó tách từng ký tự ra dưới dạng các ma trận điểm ảnh.
- Nhận dạng từng ký tự.
- Lưu trữ trở lại dưới dạng text ASCII.

Trong các bước tiến hành ở trên, giai đoạn nhận dạng ký tự quyết định đến tốc độ và độ chính xác của các hệ nhận dạng xử lý văn bản.

1.1 Các phương pháp truyền thống

Một trong những thủ tục quan trọng và tiêu tốn nhiều thời gian của các phương pháp truyền thống là làm mảnh ký tự. Thủ tục này làm giảm độ rộng các nét của ký tự xuống còn một điểm nhằm phát hiện bộ khung xương của ký tự. Độ phức tạp của thuật toán tỷ lệ với độ dày của ký tự. Thủ tục này tạo điều kiện cho việc tìm kiếm các dấu hiệu đặc trưng của ký tự, song nó cũng làm cho một số ký tự trở nên giống nhau hơn. Do đó, gây khó khăn cho các thủ tục ra quyết định sau này. Thủ tục này sẽ không làm việc tốt khi chất lượng quét của scanner tối hoặc bản thân ký tự được lấy từ văn bản mờ có nét đứt. Do đó nó cần được các thủ tục tiền xử lý như làm trơn đường biên, khử nhiễu trợ giúp.

Các phương pháp ra quyết định trong phương pháp nhận dạng truyền thống được cài đặt tĩnh trong chương trình. Các phương pháp này dựa vào việc nghiên cứu trước các đặc trưng của các ký tự như số giao điểm hoặc cấu trúc các ký tự như số điểm kết thúc, số chu trình, số thành

¹ Viện Công nghệ Xạ hiếm, Hà Nội

² Đại học Bách khoa, Hà Nội.

phần liên thông, v. v... Nhìn chung các phương pháp truyền thống đều gặp khó khăn khi chất lượng ảnh sau khi số hoá kém.

1.2 Nhận dạng bằng mạng nơ ron lan truyền ngược hướng.

Mạng nơ ron nói chung và mạng lan truyền ngược hướng nói riêng là sự mô phỏng sinh học bằng máy tính của não người. Nó có khả năng học từ kinh nghiệm hay từ một tập mẫu. Quá trình học của mạng lan truyền ngược hướng là quá trình học có giám sát với một tập mẫu $\{X_s, Y_s\}$ cho trước, ở đây X_s là véc tơ vào (ma trận điểm ảnh của một ký tự) và Y_s là giá trị ASCII của ký tự đó. Thực chất việc học của mạng là biến đổi và ánh xạ tô pô các ký tự xuống một mặt phẳng hai chiều tương ứng với các nơ ron. Sau khi học xong, mạng lan truyền ngược hướng hoạt động như một bảng tra với đầu vào là các véc tơ điểm ảnh của các ký tự. Quá trình học của mạng khá dài song quá trình nhận dạng rất nhanh. Một trong những ưu điểm chính của mạng là không đòi hỏi các quá trình tiền xử lý như làm mảnh, làm trơn đường biên hay khử nhiễu. Ưu điểm này có được nhờ khả năng xử lý các véc tơ vào có một phần bị hỏng (partly corrupted). Ngoài ra mạng có thể có khả năng nhận dạng cả các ngôn ngữ khác nhau, khi đó chỉ cần cung cấp một tập mẫu của ngôn ngữ đó cho quá trình học.

2. MẠNG NƠ RON LAN TRUYỀN NGƯỢC HƯỚNG [1]

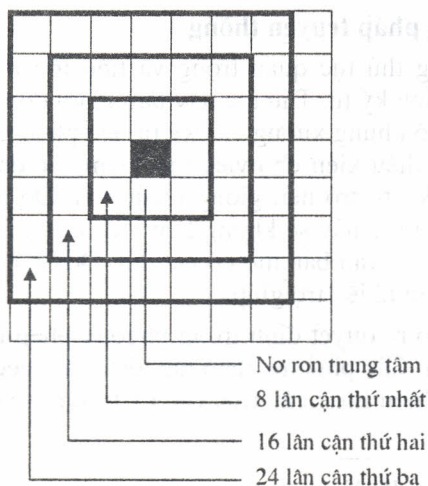
2.1 Mạng Kohonen

Khi xử lý dữ liệu ảnh, thay cho việc xem xét các số liệu đơn lẻ tương ứng với các điểm ảnh ta để ý đến mối quan hệ, vị trí, lân cận của các điểm ảnh, nghĩa là ta có cách nhìn tô pô của số liệu. Nếu ta biết được tô pô của các số liệu, ta có thể giải quyết được vấn đề dữ liệu bị mờ, không đầy đủ. Quá trình học của mạng Kohonen là quá trình tự tổ chức ánh xạ các đặc trưng tô pô của số liệu và không gian hai chiều của các nơ ron [2].

2.1.1 Cấu trúc của mạng Kohonen

Mạng Kohonen có một lớp hoạt động. Các nơ ron được sắp xếp như một ma trận chữ nhật $p \times q$. Trên ma trận nơ ron này chúng ta định nghĩa khái niệm nơ ron lân cận thứ nhất, thứ hai v. v... Hình 1 chỉ ra một ví dụ về lân cận “vuông” của một nơ ron trung tâm.

Để tất cả các nơ ron đều có số lân cận k như nhau, mạng nơ ron phải được “cuộn” các cạnh đối diện với nhau. Cách thiết kế mạng Kohonen như vậy nhằm đưa các véc tơ giống nhau vào vị trí giống nhau của các nơ ron.



Hình 1: Các nơ ron lân cận “vuông” thứ 1, 2, 3...

2.1.2 Quá trình học cạnh tranh

Trong quá trình học cạnh tranh của mạng Kohonen, chỉ có một nơ ron được chọn sau khi một vectơ vào đi qua mạng. Giả thiết có một mạng Kohonen với ma trận trọng số W_{ji} các vectơ có m biến $X(x_1, x_2, \dots, x_m)$. Tiêu chuẩn để chọn các nơ ron trung tâm là:

$$\text{nơ ron trung tâm có } \sum_{i=1}^m (x_i - W_{ji})^2 \rightarrow \min \quad (1).$$

Sau khi nơ ron trung tâm c được chọn, các trọng số W_{ji} của tất cả các nơ ron được điều chỉnh lại theo công thức:

$$W_{ji}^{(new)} = W_{ji}^{(old)} + \eta(t).a(d_c - d_j).(x_i - W_{ji}^{(old)}) \quad (2)$$

Trong đó: $a(d_c - d_j)$ là hàm phụ thuộc topo giảm theo khoảng cách topo $d_c - d_j$ giữa nơ ron j và nơ ron trung tâm c.

$$-\eta(t) = (a_{max} - a_{min}) \frac{t_{max} - t}{t_{max} - 1} + a_{min} \quad (3)$$

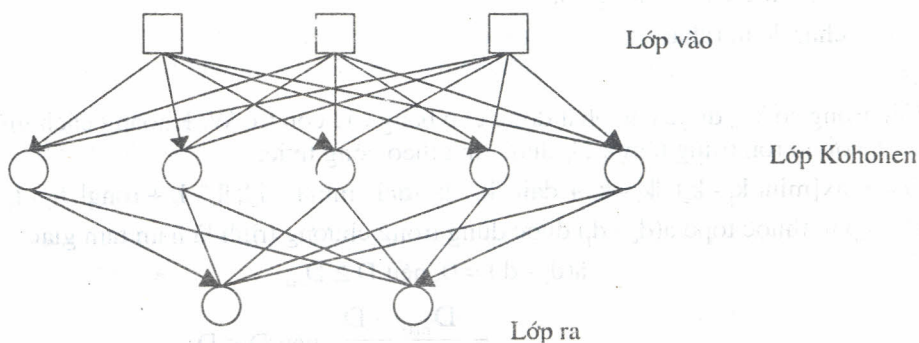
- t - Số đối tượng đã qua mạng
- t_{max} - Tổng số đối tượng qua mạng
- a_{max}, a_{min} - Hằng số học của mạng

Theo phương trình (2) ta dễ nhận thấy dù hiệu $x_i - W_{ji}^{(old)}$ là dương hay âm các trọng số $W_{ji}^{(new)}$ trở nên gần giá trị x_i hơn là các trọng số $W_{ji}^{(old)}$.

2.2 Mạng lan truyền ngược hướng

2.2.1 Cấu trúc mạng lan truyền ngược hướng

Mạng nơ ron lan truyền ngược hướng bao gồm hai lớp hoạt động: lớp Kohonen và lớp ra. Lớp Kohonen được nối đầy đủ với các biến số của vectơ vào và nối đầy đủ với các nơ ron của lớp ra. Số nơ ron của lớp Kohonen bằng số lời giải chúng ta muốn lưu trữ. Số nơ ron của lớp ra bằng số biến của mỗi lời giải. Giá trị của mỗi lời giải được lưu trữ trong các trọng số của các nơ ron trên lớp ra. Hình 2 là một ví dụ về mạng lan truyền ngược hướng lưu trữ 5 lời giải, mỗi vectơ vào có 3 biến và số biến của vectơ ra là 2:



Hình 2: Cấu trúc mạng lan truyền ngược hướng.

2.2.2 Quá trình học cạnh tranh có giám sát

Quá trình học của mạng lan truyền ngược hướng là quá trình học có giám sát. Do đó nó cần có một tập mẫu chuẩn $\{X_s, Y_s\}$. Trong quá trình học véc tơ vào X_s đi vào mạng Kohonen ở

đây diễn ra quá trình học cạnh tranh. Vectơ lời giải Y_s đi vào lớp ra theo hướng ngược lại làm thay đổi các giá trị trọng số của các nơ ron trên lớp ra. Giả thiết chúng ta có mạng lan truyền ngược hướng gồm N nơ ron trên lớp Kohonen và M nơ ron trên lớp ra. W_{ji} là trọng số thứ i của nơ ron thứ j trên lớp Kohonen. C_{ji} là trọng số của nơ ron thứ i trên lớp ra nối với nơ ron thứ j trên lớp Kohonen. Quá trình học của mạng lan truyền ngược hướng bao gồm các bước sau đây:

- Một đối tượng gồm cặp vectơ (X_s, Y_s) được lấy ra từ tập mẫu.
- Vectơ X_s đi vào lớp Kohonen.
- Nơ ron trung tâm được chọn theo phương trình (1).
- Tất cả các trọng số của nơ ron trên lớp Kohonen được điều chỉnh theo phương trình (2).
- Các trọng số của nơ ron trên lớp ra được điều chỉnh theo phương trình:

$$C_{ji}^{(new)} = C_{ji}^{(old)} + \eta(t).a(d_c - d_j).(y_i - C_{ji}^{(old)}) \quad (4)$$

- Quá trình lặp lại đối với đối tượng tiếp theo.

Mỗi lần tất cả các đối tượng mẫu đã đi qua mạng được gọi là một lượt. Thông thường từ vài trăm đến hàng nghìn lượt để mạng ổn định. Khi chọn được các hằng số đặc trưng của quá trình học a_{max} , a_{min} thích hợp, quá trình học của mạng luôn hội tụ.

3. CÀI ĐẶT MẠNG LAN TRUYỀN NGƯỢC HƯỚNG CHO NHẬN DẠNG KÝ TỰ

Một mạng tổng quát cho việc nhận dạng ký tự được cài đặt trên ngôn ngữ C như một lớp (Class) có tên gọi là Netcount. Các tham số của mạng là các biến thành viên còn các chức năng của mạng được thiết kế cho các hàm thành viên. Mạng chỉ có một nơ ron trên lớp ra và có kiểu là ký tự

```

Class Netcount
{protected:
    int dai, rong, N;
    float amax, amin, *W[1600];
    char C[1600];
public:
    Netcount(int, int);
    void hoc(char *, long T);
    char doan (char*);
};
    
```

Các trọng số W_{ji} được cấp phát động cho bảng các con trỏ W . Khoảng cách giữa nơ ron có tọa độ k_j, l_j với nơ ron trung tâm k_c, l_c được tính theo công thức:

$$D = \max[\min(|k_j - k_c|, |k_j - k_c + dai|, |k_j - k_c - dai|), \min(|l_j - l_c|, |l_j - l_c + rong|, |l_j - l_c - rong|)]$$

Hàm phụ thuộc topo $a(d_c - d_j)$ được dùng trong chương trình là hàm tam giác:

$$a(d_c - d_j) = 0, \text{ nếu } D \geq D_{max}$$

$$= \frac{D_{max} - D}{D_{max}}, \text{ nếu } D < D_{max}$$

Trong đó D_{max} là lân cận xa nhất có thể của mạng

$$D_{max} = \max(dai/2, rong/2)+1;$$

Các giá trị Y_s của tập mẫu là các ký tự có giá trị rời rạc nên phương trình (4) được thay đổi bởi $C_{ji} = y_i$ (4')

4. CÁC NGHIÊN CỨU VỀ MẠNG NƠ RON LAN TRUYỀN NGƯỢC HƯỚNG CHO NHẬN DẠNG KÝ TỰ

4.1 Tổ chức số liệu

Tập mẫu được tổ chức trong một tệp số liệu. Các cặp (X_n, Y_n) được viết lần lượt theo từng dòng. Một điều đặt ra phải số thực hoá các vector vào khoảng $[0, 1]$ vì các trọng số của mạng là các số thực. Các nghiên cứu cho thấy việc số thực hoá làm cho mạng có khả năng đoán nhận các ký tự từ ảnh sai lớn hơn. Hơn nữa với việc tổ chức thực hoá chúng ta có thể làm giảm kích thước của vector vào và có khả năng làm việc đối với các ký tự có kích thước ảnh khác nhau. Thực tế chỉ ra các phương pháp số thực hoá khác nhau sẽ ảnh hưởng đến khả năng cực đại mà mạng có thể đoán nhận từ các ảnh sai.

4.2. Cấu trúc và các tham số học

Việc xây dựng mạng là việc gán cho mạng số lượng nơ ron trên lớp Kohonen. Với số lượng nơ ron trên lớp Kohonen càng lớn khả năng đoán nhận các ký tự từ các ảnh có tỷ lệ sai lớn hơn. Tuy nhiên khi tăng số lượng nơ ron, khả năng nhận biết sẽ tiến sát khả năng cực đại mà mạng có thể đoán nhận từ các ảnh sai phụ thuộc vào phương pháp số thực hoá. Chúng ta cũng dễ nhận thấy thời gian học và thời gian đoán nhận, cũng như bộ nhớ của máy tính tăng tỷ lệ thuận với số lượng nơ ron trên lớp Kohonen. Thực tế việc xây dựng mạng là công việc thử nghiệm tăng dần cho đến khi đạt được các chỉ tiêu mong muốn.

Các giá trị trọng số ban đầu thực sự không quan trọng với quá trình học nhưng chúng phải được gán bằng các số ngẫu nhiên từ 0 đến 1.

Các tham số học a_{max} , a_{min} ảnh hưởng không nhiều đến quá trình học nếu chúng thoả mãn các điều kiện như sau:

$a_{max} \in [0.3, 1]$; $a_{min} \in [0, 0.1]$. Với giá trị $a_{max} = 0,5$ và $a_{min} = 0.01$ có thể là giá trị tốt cho quá trình học.

5. NHẬN DẠNG 37 KÝ TỰ SỬ DỤNG MẠNG LAN TRUYỀN NGƯỢC HƯỚNG

Một tập mẫu 37 ký tự từ $A \rightarrow Z$, $0 \rightarrow 9$ và ký tự '<' được tách ra từ tệp ảnh quét bởi scanner có kích thước 32×32 điểm ảnh. Ba thử nghiệm được tiến hành là:

- Không số thực hoá
- Lọc các điểm ảnh bằng mặt nạ 3×3
- Phân mảnh ảnh thành 64 mảnh. Mỗi vùng có số thực = tổng điểm / 16

Bảng 1 thống kê khả năng nhận đúng ký tự từ các ảnh có tỷ lệ sai cực đại của mạng 20×20 nơ ron sau 3000 lượt học.

Bảng 2 thống kê sự phụ thuộc của khả năng nhận dạng các ảnh sai vào kích thước mạng với việc số thực hoá là phân 64 mảnh.

Bảng 1:

| Không số thực hoá | Mặt nạ 3×3 | Phân 64 mảnh |
|-------------------|---------------------|--------------|
| 3% | 15% | 19% |

Bảng 2:

| 10 x 10 | 20 x 20 | 30 x 30 | 40 x 40 |
|---------|---------|---------|---------|
| 13% | 19% | 24% | 25% |

Hình 3. Cho thấy topo của các ký tự của mạng 30 x 30 với phương pháp phân mảnh 64 sau 3000 lần học.

```

0UUU444446666655555SSSSSGGGGG 0 0 0
0OU4444446666633335SSSSSGGGGGOOO
OODD44444666633333SSSS GGGGGOOO
ODDDDDJJJJJJ33333SSSS GGGGGOOO
ODDDDDJJJJJ133333SSSSGGGQQQQQO
ODDDDDJJJJJ11111IIIIIIIIQQQQQQQ
DDDDDDJJJJJ11111II IIIIIQQQQQQQ
DDDDDDJJJJJ11111TI I IIIRQQQQQQ
9999VVVVVJJ11111TIIIIIRRRRRRQ 9
9999VVVVVY11111TIIII RRRRRRR 9
9999VVVVVYYYTIIIIITTTTT RRRRP 9
9999VVVVYYYTIIIIITTTTT ZZZRRRP 9
9999VVVVYYYTIIIIITTTTT ZZZRRRP 9
MMMXXXYYYY77777ZZZZZ2PPPM
MMMXXXYYYY77777ZZZZZ2PPPM
MMMXXXYYYY77777ZZ2222PPP M
MMMXXXYYYY77777ZZ2222PPP M
MMMXXXYYYY77777ZZ2222 HH M
NNNNKKKKKF7777722222H HHH N
NNNNKKKKKF7777722222HH HH N
NNNNKKKK<<FFFFFFE88HHHHH N
NNNNKKKK<<FFFFFFE88HHHHH N
NNNNKK<<<<<CFFFFFFE888HWN N
WWWNK<<<<<CFFFFFFE8888WWW
WWWA<<<<<CCCCLLLLLE888BWWW
WWAAAAA<CCCCLLLLLEBBBW W
UUUAAAAA 6CCCCLLL LEBB 0 0 0 0
0 UUUAAAA66 66666 5 LLLLEBB 0 0 0 0
0UUU4AAA6666666655 55 S EBBB 0 0 0 0
0 UUU444446666666 55555 SS SGGG 0 0 0
    
```

Hình 3: Phân bố topo của các ký tự.

Với việc phân bố của các ký tự ở hình 3 ta dễ nhận thấy mạng đã phát hiện một cách khách quan các đặc trưng tô pô của các ký tự thường được dùng trong các phong pháp nhận dạng cấu trúc truyền thống. Các ký tự có cấu trúc tô pô tương đối giống nhau được sắp xếp đặt gần nhau như các ký tự có điểm kết thúc như nhau {'Z', '2'}, {'5', 'S'}; các ký tự có một chu trình {'O', '0', 'Q', 'R', '9', 'D'}; Các ký tự có hai chu trình {'B', '8'}. Một đặc điểm rất quan trọng là mạng đã phát hiện ra các ký tự có "tiềm năng" giống nhau như các ký tự {'H', 'E', 'W'} rất dễ trở thành có hai chu trình khi ảnh bị sai lớn. Ký tự 'A' khi bị mất góc cuối bên trái có thể trở thành số '4'; ký tự 'U' rất dễ trở thành có chu trình. Ngoài ra mạng đã phát hiện các ký tự có một hay nhiều phần giống nhau khó có khả năng mô tả trong các chương trình nhận dạng truyền thống như mật độ các điểm đen như {'M', 'X', 'A'}, hay nét cong của đường biên ký tự 'G' và 'O'.

KẾT LUẬN:

Từ ví dụ nhận dạng 37 ký tự cho thấy việc nhận dạng ký tự bằng mạng lan truyền ngược hướng có hiệu quả, đơn giản và nhanh hơn các phương pháp truyền thống. Nó có khả năng nhận dạng được các ký tự từ các ảnh có chất lượng tồi với số điểm ảnh sai 25%. Lợi thế chính của mạng loại này xuất phát từ khả năng học các đặc trưng pô tô của các mẫu. Tuy nhiên với một tập mẫu quá lớn việc sử dụng tài nguyên của máy tính sẽ rất lớn.

TÀI LIỆU THAM KHẢO

1. J. Zupan, J. Gasteiger, Neural network for chemist. VCH. 199.
2. T. Kohonen, "Self - Organized Formation of topologically correct Feature Maps" Biol. Cybern 43 (1982) 56