

## TÍCH HỢP KỸ THUẬT MẠNG NƠ RON VÀ GIẢI THUẬT DI TRUYỀN TRONG PHÂN TÍCH DỮ LIỆU

NGUYỄN THANH THÙY, TRẦN NGỌC HÀ

**Abstract.** In this paper we shall investigate an integration of genetic algorithms into defining neural network structure (number of hidden layers, number of neurons in each layer, neural connection weights). This combination showed its effectiveness in an experimentation for chemical data analysis in comparison with using only back-propagation neural network as shown in [4].

### MỞ ĐẦU

Trong những năm gần đây, các mạng nơ ron nhiều lớp lan truyền ngược của sai số ngày càng được sử dụng rộng rãi trong lĩnh vực phân tích số liệu và mô hình hóa [4] nhờ vào khả năng học từ số liệu của chúng. Tuy nhiên, với giải thuật học lan truyền ngược của sai số truyền thống, mà thực chất là lần lượt tối thiểu các hàm giá bằng kỹ thuật ngược hướng gradient, chúng ta gặp một số khó khăn trong việc tìm kiếm các mạng nơ ron như sau:

- Không có khả năng tìm được tập trọng số tối ưu cho một cấu trúc mạng cho trước từ đó dẫn đến hậu quả là không thể xây dựng được một chiến lược học cấu trúc.

- Với các hằng số học, hằng số quán tính cố định, thủ tục học tham số trở nên không hiệu quả.

Trong bài báo này, từ những phân tích các yếu tố ảnh hưởng tới giải thuật lan truyền ngược của sai số truyền thống chúng tôi nghiên cứu và đề ra biện pháp tích hợp giải thuật di truyền với thủ tục học của mạng nơ ron bằng việc xây dựng các hàm giá mới. Giải thuật di truyền được tích hợp với thủ tục học lan truyền ngược của sai số cải tiến tạo thành một giải thuật lai cho quá trình học tham số và được sử dụng quá trình học cấu trúc. Do kết hợp với giải thuật di truyền, giải thuật lan truyền ngược của sai số cải tiến đã sử dụng hằng số học thích nghi và loại bỏ hằng số quán tính trong các phương trình hiệu chỉnh trọng số.

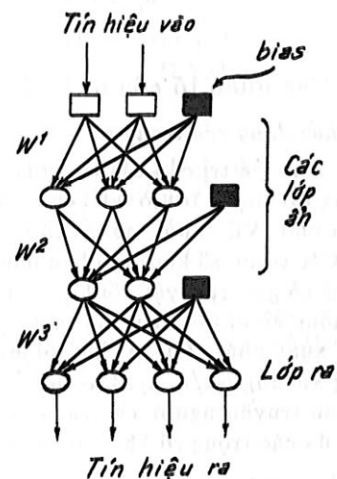
### 1. MẠNG NƠ RON NHIỀU LỚP LAN TRUYỀN NGƯỢC CỦA SAI SỐ

#### 1.1. Cấu trúc

Mạng nơ ron nhiều lớp lan truyền ngược của sai số là một mạng nơ ron feedforward trong đó các lớp của mạng được nối đầy đủ với nhau [5]. Hình 1 là hình ảnh của một mạng lan truyền ngược của sai số với hai lớp ẩn. Đối với một bài toán cụ thể, số nơ ron trên lớp vào bằng số biến của đối tượng vào (véc tơ tín hiệu vào input) và số nơ ron trên lớp ra bằng số biến của véc tơ lời giải (target). Trên mỗi lớp, ngoài lớp ra có thêm một tham số bias. Tín hiệu ra của bias luôn được gán cho là 1.

#### 1.2. Giải thuật học

Thủ tục học của mạng lan truyền ngược là thủ tục học có chỉ đạo; do đó nó cần một tập mẫu  $(X_s, Y_s)$ , với  $X_s$  là véc tơ vào,  $Y_s$  là véc tơ lời giải. Đối với một cặp tín hiệu vào  $X_s$  và lời giải  $Y_s$ , giải thuật học lan truyền ngược tiến hành hai giai đoạn của dòng chảy của số liệu.



Hình 1. Cấu trúc của mạng nơ ron nhiều lớp lan truyền ngược của sai số

- Tín hiệu vào  $X$ , được lan truyền qua mạng từ lớp vào đến lớp ra. Kết quả của việc lan truyền này là sản sinh ra véc tơ tín hiệu ra  $Out^{last}$ .
  - Tín hiệu sai số là kết quả của việc so sánh giữa véc tơ lời giải và véc tơ tín hiệu ra, được lan truyền ngược từ lớp ra tới các lớp phía trên để hiệu chỉnh các trọng số.
- Đối với mỗi cặp tín hiệu này chúng ta xây dựng một hàm giá như sau:

$$E(w) = \frac{1}{2} \sum_{j=1}^n (y_j - Out_j^{last})^2, \quad (1)$$

trong đó  $n$  là số nơ ron trên lớp ra.

Như vậy, với một tập học gồm  $P$  mẫu chúng ta lần lượt xây dựng  $P$  hàm giá như vậy. Việc học của giải thuật thực chất là việc tìm kiếm một tập trọng số trong không gian  $R^M$  ( $M$  là số trọng số có trong mạng) để lần lượt tối thiểu hóa các hàm giá như vậy. Điều đáng chú ý là việc tối thiểu hóa được tiến hành liên tiếp nhau và theo chu kỳ đối với các hàm giá.

Để tối thiểu hóa các hàm giá như vậy, giải thuật lan truyền ngược sử dụng phương pháp giảm gradient (gradient-decent method - deepest decent method). Bằng biểu thức chúng ta có:

$$\Delta w_{ji}^l = -\eta \frac{\partial E}{\partial w_{ji}^l}. \quad (2)$$

Với hàm biến đổi từ tổng tín hiệu Net thành tín hiệu ra Out trong mỗi nơ ron là hàm sigmoidal, chúng ta có công thức cuối cùng của giải thuật lan truyền ngược của sai số là:

$$\Delta w_{ji}^l = \eta \left( \sum_{k=1}^r \delta_k^{l+1} w_{kj}^{l+1} \right) Out_j^l (1 - Out_j^l) Out_j^{l-1} + \mu \Delta w_{ji}^{l(\text{previous})}.$$

Trong đó:

$$\begin{aligned} \delta_j^{last} &= (y_j - Out_j^{last}) Out_j^{last} (1 - Out_j^{last}), \\ \delta_j^l &= \left( \sum_{k=1}^r \delta_k^{l+1} \right) Out_j^l (1 - Out_j^l) \quad \text{với } l = last - 1 \text{ đến } 1. \end{aligned} \quad (3)$$

### 1.3. Các nhân tố của quá trình học

#### a. Khởi động các trọng số

Các giá trị của các trọng số được khởi tạo ban đầu trong mạng lan truyền ngược sai số ảnh hưởng rất mạnh tới lời giải cuối cùng. Các trọng số này thường được khởi tạo bằng những số ngẫu nhiên nhỏ. Việc khởi tạo tất cả các trọng số bằng nhau sẽ làm cho việc học của mạng trở nên không tốt. Các trọng số khởi tạo ban đầu không được quá lớn vì nếu không ngay từ đầu tổng tín hiệu vào Net sẽ có giá trị tuyệt đối lớn và làm cho hàm sigmoidal chỉ có hai giá trị 0 và 1. Điều này làm cho hệ thống sẽ bị tắc ngay tại một cực tiểu cục bộ hoặc tại một vùng bằng phẳng nào đấy gần ngay điểm xuất phát. Giá trị khởi động ban đầu của các trọng số trên lớp thứ  $l$  được chọn ngẫu nhiên trong khoảng  $[-1/n, 1/n]$ , trong đó  $n$  là tổng số trọng số trên lớp  $l$  [5]. Do bản chất của giải thuật học lan truyền ngược của sai số là phương pháp giảm gradient nên việc khởi động các giá trị ban đầu của các trọng số khác nhau sẽ làm cho mạng hội tụ về các cực trị địa phương khác nhau.

#### b. Hằng số học $\eta$

Một nhân tố khác ảnh hưởng tới hiệu lực và độ hội tụ của giải thuật học lan truyền ngược của sai số là hằng số học  $\eta$ . Không có một giá trị duy nhất phù hợp với các bài toán khác nhau. Hằng số học này thường được chọn bằng thực nghiệm cho mỗi bài toán ứng dụng cụ thể bằng phương

pháp thử và sai. Giá trị  $\eta$  lớn làm tăng tốc quá trình hội tụ. Điều này là không có lợi vì thủ tục học sẽ kết thúc rất nhanh tại một cực tiểu cục bộ gần nhất. Nếu giá trị của hằng số học quá nhỏ tốc độ hội tụ của giải thuật lại trở nên rất chậm. Do đó chúng ta cần phải chọn một giá trị thỏa hiệp giữa tốc độ học và việc ngăn chặn hội tụ về các cực tiểu cục bộ. Nói chung giá trị của hằng số nằm trong khoảng  $0,3 + 0,6$  được chứng minh bằng thực nghiệm là khá tốt cho việc lựa chọn ban đầu của quá trình tìm kiếm hằng số học thích hợp.

### c. Các hàm giá

Hàm giá bình phương được định nghĩa theo phương trình (1) không phải là sự lựa chọn duy nhất có thể. Số hạng bình phương sai số  $(y_j - \text{Out}_j)^2$  có thể được thay thế bằng bất cứ hàm có khả năng đạo hàm nào khác  $F(y_j, \text{Out}_j)$ , với điều kiện là các hàm này sẽ đạt cực tiểu khi hai đối số  $y_j$  và  $\text{Out}_j$  bằng nhau. Dựa vào dạng của hàm giá mới chúng ta có thể thu nhận được các phương trình khác nhau cho việc hiệu chỉnh các trọng số. Một điều dễ nhận thấy là chỉ có phương trình hiệu chỉnh các trọng số ở lớp ra là thay đổi theo các dạng hàm giá khác nhau, trong khi đó các phương trình khác của giải thuật lan truyền ngược của sai số vẫn giữ nguyên.

Các hàm giá thường được sử dụng là những hàm dựa trên chuẩn  $L_p$  ( $1 \leq p \leq \infty$ ) [2] bởi sự lợi thế của dạng phương trình toán học đơn giản. Các hàm giá như vậy có dạng là:

$$E = \frac{1}{p} \sum_{j=1}^n |y_j - \text{Out}_j|^p \quad \text{với } 1 \leq p \leq \infty. \quad (4)$$

Trường hợp hàm giá định nghĩa theo phương trình (1) là trường hợp riêng với  $p = 2$  của phương trình (4).

### d. Hằng số quán tính

Tốc độ học của giải thuật lan truyền ngược của sai số có thể rất chậm nếu hằng số học nhỏ và có thể dao động khi hằng số học lớn. Một phương pháp thường dùng cho phép sử dụng hằng số học lớn là thêm thành phần quán tính vào các phương trình hiệu chỉnh các trọng số. Ngoài ra hằng số quán tính  $\mu$  ngăn cản sự thay đổi đột ngột của các trọng số theo hướng khác với hướng mà lời giải đang di chuyển đến. Mặt trái của việc sử dụng thành phần quán tính là chúng ta phải tăng đáng kể bộ nhớ của máy tính gần như gấp đôi để lưu trữ các giá trị hiệu chỉnh ở chu kỳ trước.

## 2. KẾT HỢP GIẢI THUẬT DI TRUYỀN VỚI GIẢI THUẬT LAN TRUYỀN NGƯỢC CỦA SAI SỐ CHO VIỆC HỌC THAM SỐ

### 2.1. Giải thuật di truyền cho việc học tham số

Như chúng ta đều biết, giải thuật di truyền với ba toán tử cơ bản là tái tạo, tạp lai và đột biến có khả năng tìm kiếm toàn cục [1]. Để có thể sử dụng giải thuật di truyền cho việc học tham số, chúng ta cần giải quyết các vấn đề sau:

#### a. Xây dựng hàm giá

Khác với giải thuật lan truyền ngược của sai số chúng ta phải xây dựng một hàm giá duy nhất. Đối với một tập cho trước  $(X_s, Y_s)$ , chúng ta lan truyền lần lượt các véc tơ vào và tích lũy lại sai số thành một sai số tổng thể cho tập học đó:

$$E(w) = \frac{1}{2} \sum_{s=1}^P \sum_{j=1}^n (y_{sj} - \text{Out}_{sj}^{\text{last}})^2. \quad (5)$$

Giải thuật di truyền sẽ tìm kiếm tập trọng số trong không gian  $R^M$  để tối thiểu hàm giá (5).

b. Mã hóa và giải mã các trọng số

Đối với một cấu trúc mạng cho trước chúng ta sắp xếp các trọng số thành một danh sách. Giả sử chúng ta có một mạng nơ ron với  $L$  lớp;  $m, n$  lần lượt là số nơ ron trên lớp vào và lớp ra. Trọng số  $w_{ij}^l$  thứ  $i$  của nơ ron thứ  $j$  trên lớp  $l$  sẽ chiếm vị trí thứ  $k$  trong danh sách theo công thức sau:

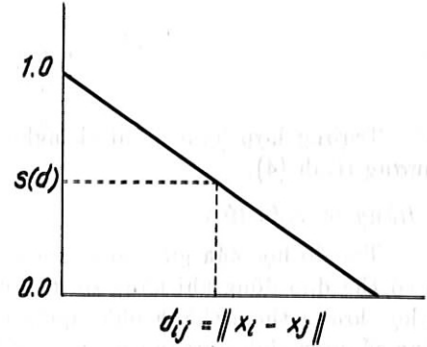
$$k = \left( \sum_{s=1}^{l-1} N(s)(N(s-1) + 1) \right) + (j-1)[N(l-1) + 1] + i. \quad (6)$$

Theo nguyên tắc tập ký tự nhỏ nhất [1] của giải thuật di truyền, chúng ta mã hóa danh sách thành một chuỗi nhị phân như sau: Mỗi trọng số được mã hóa thành một chuỗi có độ dài cố định ví dụ là 20 bit với giá trị  $\max = 10$  và  $\min = -10$ , vì theo các quan sát thực nghiệm các giá trị trọng số trong các mạng thường nằm trong khoảng  $[-10, 10]$ . Các chuỗi 20 bit của các trọng số được nối với nhau tạo thành một chuỗi dài. Chuỗi dài này đại diện cho một tập trọng số của một cấu trúc mạng.

Việc giải mã các trọng số từ một chuỗi dài là công việc cắt chuỗi dài thành các chuỗi con 20 bit. Giả sử giá trị của chuỗi nhị phân được cắt ra là  $x$ , lúc này giá trị của trọng số tương ứng với chuỗi con đó là  $20.x/(2^{21} - 1) - 10$ .

c. Toán tử ở mức quần thể

Ngoài các toán tử của giải thuật di truyền như tái tạo, tập lai và đột biến, chúng tôi đưa thêm một toán tử ví mô hoạt động ở mức quần thể. Toán tử này tính đến sự phân biệt giữa các loài. Sơ đồ mô phỏng sự hình thành các loài trong giải thuật di truyền là hàm chia sẻ do Goldberg và Richardson giới thiệu năm 1987 [2]. Một hàm chia sẻ xác định tính gần cạnh và độ chia sẻ cho mỗi chuỗi trong quần thể. Hình 2 là hình ảnh của một hàm chia sẻ  $s(d(x_i, x_j))$ . Trong đó  $d(x_i, x_j)$  là khoảng cách hamming giữa hai chuỗi  $x_i$  và  $x_j$ . Khi sử dụng toán tử ví mô này độ phù hợp của một chuỗi nào đó được tính lại theo phương trình sau:



Hình 2. Hàm chia sẻ

$$f_s(x_i) = \frac{f(x_i)}{\sum_{j=1}^n s(d(x_i, x_j))}, \quad (7)$$

trong đó  $f(x_i)$  là độ phù hợp ban đầu của chuỗi. Toán tử này hạn chế sự phát triển bùng nổ của một nhóm cá thể trội trong quần thể và gây ra sự hội tụ sớm không mong muốn. Hiệu ứng của hàm chia sẻ là các cá thể được phân bố rải rác đều xung quanh các cực tiểu.

d. Khởi động quần thể đầu tiên

Các trọng số của các cá thể trong quần thể ban đầu được chọn ngẫu nhiên trong khoảng  $[-10, 10]$ . Các trọng số này được khởi động ngẫu nhiên với xác suất cho bởi phân bố  $e^{-|x|}$ . Xác suất phân bố này được rút ra từ quan sát thực nghiệm là lời giải tối ưu có phần lớn các trọng số với giá trị tuyệt đối nhỏ và đồng thời chúng cũng có số ít các trọng số giá trị tuyệt đối lớn. Do đó với phân bố xác suất này cho việc khởi động các trọng số cho phép giải thuật di truyền thăm dò các vùng chứa tất cả các lời giải có thể, đồng thời nó hướng giải thuật vào những vùng hay chứa lời giải nhất.

e. Các tham số của giải thuật di truyền

Các tham số của giải thuật di truyền như xác suất tập lai, xác suất đột biến, số cá thể trong quần thể và số thế hệ được chọn theo phương pháp thử và sai. Các giá trị  $p_{cross} = 0,7$ ,  $p_{mutation} = 0,001$  được coi là các giá trị xuất phát khá tốt. Nếu giá trị  $p_{mutation}$  lớn giải thuật di truyền trở

thành giải thuật tìm kiếm ngẫu nhiên. Số cá thể trong quần thể thường được chọn ở cỡ trung bình 100 đến 200 cá thể cho một quần thể. Số thế hệ cần thiết theo thực nghiệm bằng khoảng 5 - 10 lần tổng trọng số trong mạng.

## 2.2. Ghép nối với giải thuật lan truyền ngược của sai số

Giải thuật di truyền có thể tìm ra vùng quan tâm chứa cực trị toàn cục song nó không có khả năng đạt được cực trị đó. Nói cách khác, giải thuật di truyền không đảm bảo cho sự hội tụ. Trái lại, giải thuật lan truyền ngược của sai số có thể được thực hiện đến bão hòa nhưng khả năng tìm kiếm cực trị toàn cục là có hạn. Do đó, việc kết hợp hai giải thuật trên trong một giải thuật lai là lẽ tự nhiên. Có nhiều cách lai ghép hai giải thuật này song cách đơn giản và hiệu quả là hai giải thuật này được tiến hành nối tiếp nhau.

Đối với một cấu trúc mạng cho trước, chúng ta xuất phát bằng giải thuật di truyền khởi động một quần thể  $N$  chuỗi, mỗi chuỗi là một bản mã của tập trọng số của cấu trúc mạng đã cho. Giải thuật di truyền tiến hành việc tiến hóa quần thể ban đầu. Sau  $G$  thế hệ, 5% cá thể tốt nhất trong  $G$  thế hệ được lưu giữ lại [3]. Như vậy, đầu vào của giải thuật di truyền là một cấu trúc mạng, đầu ra là  $0,05 \cdot N$  tập trọng số. Các tập trọng số này lần lượt được giải thuật lan truyền ngược của sai số luyện đến bão hòa. Sau quá trình luyện bằng giải thuật lan truyền ngược của sai số chúng ta có  $0,05 \cdot N$  tập trọng số mới. Tập trọng số cho kết quả tốt nhất (giá trị hàm giá nhỏ nhất) được giữ lại là kết quả của giải thuật lai di truyền - lan truyền ngược của sai số. Giải thuật lan truyền ngược của sai số trong giải thuật lai này về cơ bản giống như giải thuật lan truyền ngược hướng đã trình bày ở trên và được cải tiến ở một số điểm sau:

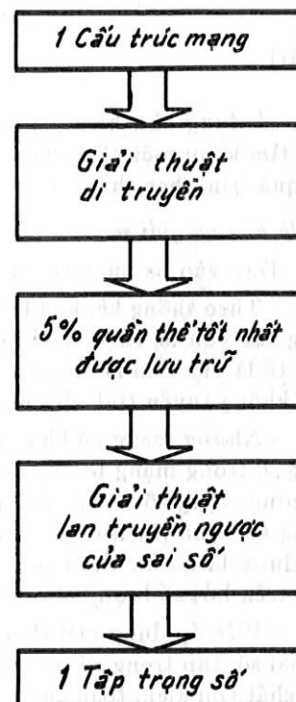
- Chúng ta không cần phải khởi động các giá trị trọng số ban đầu mà tiếp nhận tập trọng số từ giải thuật di truyền như tập trọng số ban đầu.
- Thành phần quán tính trong các phương trình hiệu chỉnh các trọng số bị lược bỏ vì tập trọng số xuất phát đã khá gần lời giải. Tác dụng chống dao động và thay đổi đột ngột của các trọng số theo hướng khác hướng của lời giải của thành phần quán tính trở thành không cần thiết.
- Chúng ta có thể áp dụng hằng số học thích nghi vào trong giải thuật lan truyền ngược của sai số để đảm bảo luôn giảm giá trị của các hàm giá (1). Nói một cách khác làm tăng tốc độ hội tụ của giải thuật.

Việc áp dụng hằng số học thích nghi dựa trên phương pháp trực giác là kiểm tra xem liệu các trọng số mới đã làm giảm hàm giá chưa. Nếu chưa khi đó quá trình hiệu chỉnh các trọng số đã phóng quá xa, lúc này hằng số học  $\eta$  cần phải được giảm đi. Mặt khác nếu vài vòng lặp liên tiếp làm giảm hàm giá, khi đó chúng ta cần phải thử tăng giá trị của hằng số học. Bằng biểu thức, hằng số học cần được cập nhật theo quy tắc sau:

$$\Delta\eta = \begin{cases} +a & \text{nếu } \Delta E \text{ "luôn" nhỏ hơn 0} \\ -b\eta & \text{nếu } \Delta E > 0 \\ 0 & \text{trường hợp còn lại} \end{cases}$$

Trong đó  $\Delta E$  là độ thay đổi của hàm giá;  $a$  và  $b$  là các hằng số dương. Ý nghĩa của việc "luôn" nhỏ hơn 0 được đánh giá dựa trên  $k$  bước lặp gần nhất.

Hình 3 là sơ đồ khối tổng thể của giải thuật lai di truyền - lan truyền ngược của sai số.



Hình 3. Sơ đồ của giải thuật lai

### 3. GIẢI THUẬT DI TRUYỀN CHO VIỆC HỌC CẤU TRÚC

Trong các phần trước chúng ta tập trung chủ yếu vào việc xây dựng một giải thuật lai cho học tham số. Điều này ngầm định là chúng ta đã có một cấu trúc mạng cho trước. Việc áp dụng các giải thuật học chỉ là việc hiệu chỉnh các trọng số sao cho mạng hoạt động như mong muốn. Điều hiển nhiên là với một cấu trúc mạng không phù hợp thì các thủ tục học tham số không thể làm cho mạng hoạt động tốt được. Từ trước việc tìm ra cấu trúc mạng thường là công việc thử và sai hoặc dựa trên các phân tích mang tính thực nghiệm. Trong phần này chúng ta sẽ tập trung và giải thuật giúp cho việc chọn cấu trúc mạng phù hợp nhất đối với một tập học cho trước. Như chúng ta đã biết giải thuật di truyền giúp ích rất nhiều cho việc tiền học tham số nhằm đến các vùng chứa cực trị toàn cục.

Một tiềm năng khác của giải thuật di truyền kết hợp với mạng nơ ron là khả năng tổng hợp các cấu trúc mạng [3]. Ý tưởng cơ bản là xây dựng một giải thuật di truyền tìm kiếm trong không gian các cấu trúc mạng có thể. Để có thể sử dụng giải thuật di truyền cho việc học cấu trúc, chúng ta tiến hành các bước sau:

#### a. Xây dựng hàm giá

Tập học được chia thành hai tập con được gọi là tập luyện và tập kiểm tra. Tỷ lệ giữa hai tập phụ thuộc vào số mẫu trong tập học. Việc chọn, chia tập học thành hai tập con có thể được tiến hành một cách ngẫu nhiên hoặc một cách hệ thống. Giả sử chúng ta chọn tập kiểm tra có  $r$  mẫu từ tập học có  $P$  mẫu. Chúng ta xây dựng hai hàm giá như sau:

$$E_{\text{train}} = \sum_{s=1}^{P-r} \sum_{j=1}^n (y_{s,j} - \text{Out}_{s,j}^{\text{inst}})^2 \quad \text{và} \quad E_{\text{test}} = \sum_{s=1}^r \sum_{j=1}^n (y_{s,j} - \text{Out}_{s,j}^{\text{inst}})^2.$$

Giá trị 
$$E = E_{\text{train}} \left( \frac{P-r}{P} \right) + E_{\text{test}} \left( \frac{r}{P} \right) \quad (8)$$

được sử dụng như hàm giá cho quá trình học cấu trúc. Nói một cách khác là giải thuật di truyền phải tìm kiếm một cấu trúc mạng để tối thiểu hàm giá  $E$ . Trong đó chỉ có tập luyện được sử dụng cho quá trình học tham số.

#### b. Mã hóa và giải mã

Dựa vào ba quan sát trong thực tế là:

- Theo thống kê, hầu hết các báo cáo về sử dụng các mạng nơ ron lan truyền ngược của sai số trong các vấn đề thực tế chỉ gồm có ba lớp trong đó có một lớp ẩn [5]. Việc tìm ra cấu trúc mạng thực tế là việc tìm ra số nơ ron trên lớp ẩn. Thực tế này xuất phát từ khả năng mô phỏng tốt các hàm không tuyến tính của mạng có một lớp ẩn.

- Những mạng có khả năng dự báo cũng như nhắc lại tập đã luyện tốt nhất khi số lượng các trọng số trong mạng là xấp xỉ bằng hoặc nhỏ hơn số lượng các mẫu khác nhau trong tập luyện. Khi số lượng trọng số trong các mạng lớn hơn số các mẫu học khác nhau trong tập học sẽ xảy ra hiện tượng học quá (hiện tượng học tốt, song dự báo lại rất kém). Hiện tượng này hoàn toàn có thể lý giải được là do dư thừa tham số luyện. Do đó với một tập luyện cho trước số lượng trọng số bị chặn trên bởi số lượng mẫu khác nhau có trong tập học.

- Việc áp dụng giải thuật di truyền vào học tham số kết hợp với giải thuật lan truyền ngược của sai số, tập trọng số tốt nhất của một cấu trúc mạng cho trước hầu như luôn được tìm thấy do tính chất tìm kiếm toàn cục của giải thuật di truyền [1]. Kết hợp với việc sử dụng hằng số học thích nghi việc mã hóa các hằng số học trong cấu trúc là không cần thiết.

Từ ba quan sát kể trên, việc mã hóa cấu trúc mạng trong giải thuật di truyền học cấu trúc được tiến hành như sau:

- Từ tập mẫu luyện xác định số đối tượng khác nhau  $M$ , thường là số mẫu có trong tập. Số đối tượng này xác định số trọng số lớn nhất có thể để mạng không bị học quá.

- Xác định số nơ ron lớn nhất có thể trên lớp ẩn theo công thức sau:

$$A \cong \frac{M - n}{m + n + 1} \quad (9)$$

Trong đó  $A$  là số nơ ron trên lớp ẩn,  $m$  là số biến của véc tơ vào,  $n$  là số biến của véc tơ lời giải.

- Các cấu trúc mạng được mã hóa thành chuỗi nhị phân có chiều dài là  $l$  sao cho  $2^{l-1} < A < 2^l$ .

Việc giải mã là việc xây dựng một mạng lan truyền ngược của sai số với một lớp ẩn có số nơ ron bằng giá trị của chuỗi nhị phân.

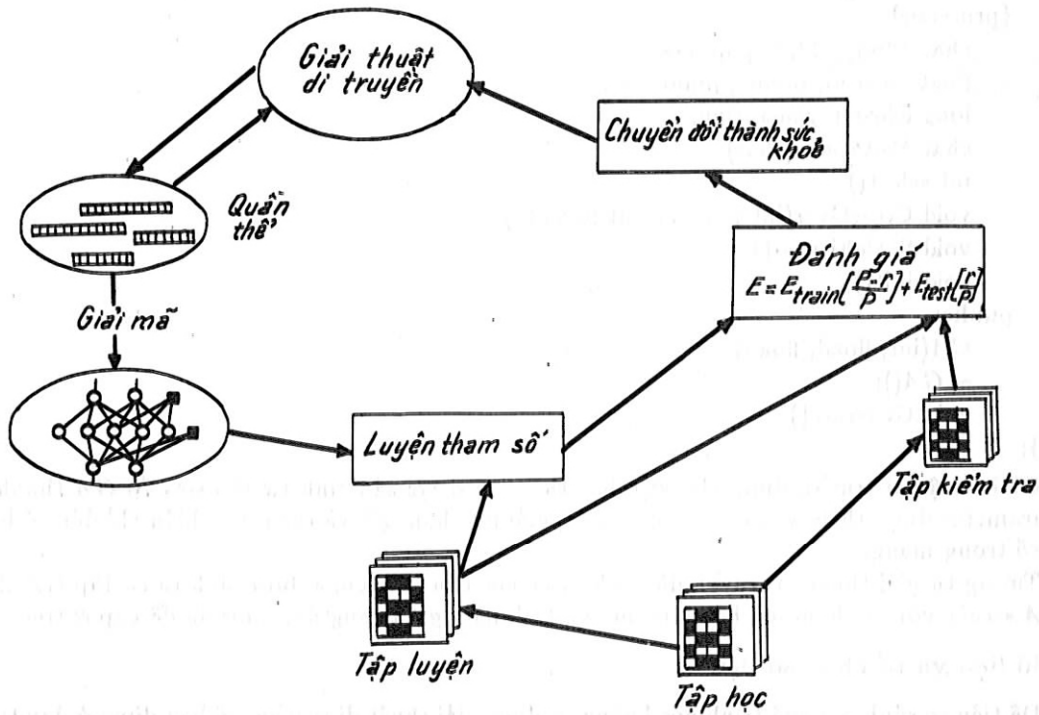
c. Các tham số của giải thuật di truyền

Các tham số của giải thuật di truyền như xác suất tạp lai, xác suất đột biến, số cá thể trong quần thể và số thế hệ được chọn theo phương pháp thử và sai. Để tăng khả năng hội tụ của giải thuật di truyền, xác suất tạp lai được chọn khá lớn và xác suất đột biến nhỏ. Số cá thể trong quần thể thường được chọn ở cỡ trung bình học nhỏ phụ thuộc vào độ dài của chuỗi mã hóa.

d. Khởi động quần thể đầu tiên

Các cá thể được khởi động một cách ngẫu nhiên. Nói cách khác các mạng nơ ron có số lượng nơ ron trên lớp ẩn được chọn với xác suất như nhau trong khoảng từ 2 đến số lượng nơ ron max  $A$ .

Sơ đồ tổng thể cho việc học cấu trúc bao gồm cả giải thuật học tham số như sau:



Hình 4. Chu trình sử dụng giải thuật di truyền cho việc học cấu trúc và tham số

#### 4. CÀI ĐẶT VÀ THỬ NGHIỆM

##### 4.1. Cài đặt

Một mạng lan truyền ngược tổng quát gồm có lớp trong đó có một lớp ẩn được cài đặt trên C như một lớp (class) NetWork. Các tham số của mạng là các biến thành viên được bảo vệ, còn các chức năng của mạng được thiết kế cho các hàm thành viên của lớp.

```

class netWork
{protected:
    int N, n, m;
    float *W, *Out;
    void Lan-truyền();
public:
    NetWork(int, int, int, float *w);
    ~NetWork();
    void Học(char *str, int option);
    void Đoán(float *X, float *Y);
};

```

Các trọng số  $w_i^l$  được cấp phát động cho con trỏ \*W, đối tượng nhận một cấu trúc mạng (cụ thể là số nơ ron trên lớp vào và ra, số nơ ron trên lớp ẩn) và một danh sách các trọng số là kết quả của quá trình học tham số bằng giải thuật di truyền.

Giải thuật di truyền tổng quát cho cả hai thủ tục học cấu trúc và tham số được cài đặt như một lớp sơ cấp GA không có chức năng mã hóa và giải mã. Các toán tử của giải thuật được cài đặt như các hàm thành viên bảo vệ.

```

class GA
{protected:
    char **pop_old, **pop_new;
    float *fitness, pcross, pmutation;
    long lchrom, generation;
    char Mutation (char);
    int select();
    void CrossOver(int parent1, int parent2);
    void ScaleFitness();
    void init();
public:
    GA(int, float, float);
    ~GA();
    void Generate();
};

```

Giải thuật di truyền dùng cho việc học tham số được sản sinh từ lớp sơ cấp GA thành lớp GA-parameter được thêm vào các hàm mã hóa, giải mã, hàm giá và thêm vào biến chỉ đến số lượng trọng số trong mạng.

Tương tự giải thuật di truyền dùng cho việc học cấu trúc cũng được sinh ra từ lớp GA thành lớp GA-struct với các hàm mã hóa, giải mã và tính hàm giá tương ứng như đã đề cập ở trên.

#### 4.2. Số liệu và tổ chức số liệu

Để tiện so sánh với quá trình học không sử dụng giải thuật di truyền, số liệu dùng ở đây tương tự như trong [4] gồm có 49 số liệu. Các số liệu này được tổ chức trong một tệp, các cặp véc tơ vào và lời giải được viết trên từng dòng. Khi dùng trong mạng, các số liệu này cũng được tỷ lệ hóa trong khoảng (0,2, 0,8) [4]. Để phục vụ cho việc học bằng giải thuật di truyền, 5 mẫu được trích ra để dùng cho tập kiểm tra.

#### 4.3. Kết quả và chạy thử nghiệm

Kết quả chạy thử nghiệm cho thấy kết quả cấu trúc mạng được chọn là (2×8×1) khác với kết quả chọn bằng thử nghiệm (2×7×1) như trong [4], nhưng cả hai giá trị nhắc lại và dự báo của mạng tìm ra đều tốt hơn. Điều này dễ hiểu do khả năng tìm kiếm toàn cục của giải thuật di truyền trong



việc học tham số. Giá phải trả cho việc sử dụng giải thuật này là thời gian tính toán gấp 20 lần so với thời gian tính toán cho một lần thử nghiệm được nêu trong [4].

### 5. KẾT LUẬN

Với ví dụ minh họa tương tự như trong [4] cho thấy việc sử dụng giải thuật di truyền trong việc học tham số và cấu trúc sẽ hiệu quả hơn so với việc chỉ sử dụng giải thuật lan truyền ngược của sai số. Một lợi thế cơ bản của giải thuật lai này là làm cho máy tính có thể tự động chọn một cấu trúc mạng phù hợp nhất với một tập số liệu cho trước. Điều này rất có lợi khi chúng ta phải xây dựng một mạng cỡ trung bình và lớn để giải những bài toán thực tế với tập số liệu cực lớn.

### TÀI LIỆU THAM KHẢO

- [1] Davis E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, MA: Addison-Wesley, 1989.
- [2] C. T. Lin, C. S. G. Lee, *Neurofuzzy*, Academic Press, New York, 1993.
- [3] L. R. Medsker, *Hybrid Intelligent Systems*, Kluwer Academic Pub., 1995.
- [4] Trần Ngọc Hà, Nguyễn Thanh Thủy, Mạng nơ ron nhiều lớp lan truyền ngược dùng cho việc mô hình hóa, *Tạp chí Tin học và Điều khiển học* 14 (1) (1998) 26-33.
- [5] J. Zupan, J. Gasteiger, *Neural Networks for Chemists*, VCH Pub., 1993.

Nhận bài ngày 12-8-1998

Nhận lại sau khi sửa ngày 19-1-1999

Nguyễn Thanh Thủy - Trường Đại học Bách khoa Hà Nội.

Trần Ngọc Hà - Viện Công nghệ Xạ hiếm.