Tạp chí Tin học và Điều khiển học, T. 18, S. 3 (2002), 273–278

ỨNG DỤNG THUẬT TOÁN PHÂN TÍCH BIỆT SỐ TUYẾN TÍNH BẰNG GIẢI THUẬT DI TRUYỀN ĐỂ TIẾN HÀNH GIẢI BÀI TOÁN PHÂN LỚP TRONG Y HỌC

NGUYỄN CÔNG PHÚC, DOÃN HỒ LIÊN, CAO VĂN VIÊN, CÙ THU THỦY

Abstract. Linear discriminant analysis method basing on genetic algorithms was used to solve the classifying problem in medical field, through finding out the role of clinical symtomps for taking differential diagnosis between typhoid fever and scrub typhus, as an example. Clinical records of 526 patients suffering from one of those diseases were used as input data for this algorithm. The results showed that the differential value of symptoms can be determined clearly by using this method and could be used in routine medical practicing.

Tóm tắt. Thuật toán phân tích biệt số tuyến tính bằng giải thuật di truyền đã được sử dụng để giải bài toán phân lớp trong y học, thông qua thí dụ: đánh giá vai trò của các triệu chứng lâm sàng trong chẩn đoán phân biệt giữa bệnh thương hàn và bệnh sốt mò. Bệnh án của 526 bệnh nhân bị mắc một trong 2 bệnh trên đã được sử dụng làm dữ liệu đầu vào cho giải thuật. Các kết quả cho thấy giá trị chẩn đoán phân biệt của các triệu chứng lâm sàng có thể được xác định rõ ràng bằng giải thuật này, và các kết quả của bài toán có thể đưa ra ứng dụng được trong thực hành lâm sàng.

1. ĐẶT VẤN ĐỀ

Trong thời gian gần đây, nhiều kỹ thuật chẩn đoán dựa trên những tiến bộ khoa học là nền tảng để giải quyết cho nhiều vấn đề y học. Tuy nhiên, cho tới nay đa phần các tiến bộ này mới chỉ được áp dụng vào khu vực cận lâm sàng. Đặc biệt, ở Việt Nam nói riêng và ở các nước đang phát triển nói chung, thực hành lâm sàng vẫn chủ yếu dựa vào sự đúc rút kinh nghiệm của bản thân người thầy thuốc thông qua những ca bệnh cụ thể.

Việc chẩn đoán phân biệt nhằm phân định rõ trong một trường hợp cụ thể, người bệnh bị bệnh này hay bệnh kia, thường gây lúng túng cho người thầy thuốc, và trong nhiều trường hợp, kết quả cận lâm sàng thường là cơ sở để đưa ra chẩn đoán xác định cuối cùng. Thế nhưng, không phải trong mọi trường hợp chẩn đoán cận lâm sàng luôn luôn thích hợp. Đặc biệt là với những thầy thuốc ở các tuyến cơ sở, xa các khu đô thị lớn, cơ sở vật chất còn thiếu thốn, thì chẩn đoán và điều trị dựa trên cơ sở thực hành lâm sàng là chủ yếu.

Hiện nay, trên thế giới ứng dụng công nghệ thông tin trong y học đang ngày càng trở nên quan trọng. Đã có rất nhiều phương pháp thống kê toán học mới được áp dụng để hỗ trợ cho các chẩn đoán của bác sĩ, trong đó sự hỗ trợ của máy tính chiếm một vị trí đáng kể. Thế nhưng ở Việt Nam, đây còn là một vấn đề hết sức mới mẻ, chưa được nghiên cứu và áp dụng rộng rãi. Các nghiên cứu y học lâm sàng dựa trên một nền số liệu thu thập được, mới chỉ dừng lại ở những con số thống kê và kiểm định tương quan hết sức đơn giản.Việc đưa ra các tiêu chuẩn chẩn đoán phân biệt vẫn chủ yếu dựa vào kinh nghiệm lâm sàng của các thầy thuốc lâu năm - các chuyên gia đầu ngành lâm sàng.

Vì vậy, chúng tôi nghiên cứu đề tài này nhằm mục đích ứng dụng thuật toán phân tích biệt số tuyến tính bằng giải thuật di truyền để hình thành hệ thống trợ giúp các chuyên gia y tế trong việc đưa ra các tiêu chuẩn chẩn đoán phân biệt giữa 2 bệnh có triệu chứng tương tự nhau nhằm hỗ trợ cho công tác thực hành lâm sàng tại các tuyến cơ sở, và kiểm chứng lại hoạt động của hệ thống này bằng số liệu thực tế. Những tiêu chuẩn chẩn đoán phân biệt này sẽ là cơ sở để xây dựng các luật dẫn cho các hệ hỗ trợ chẩn đoán sau này.

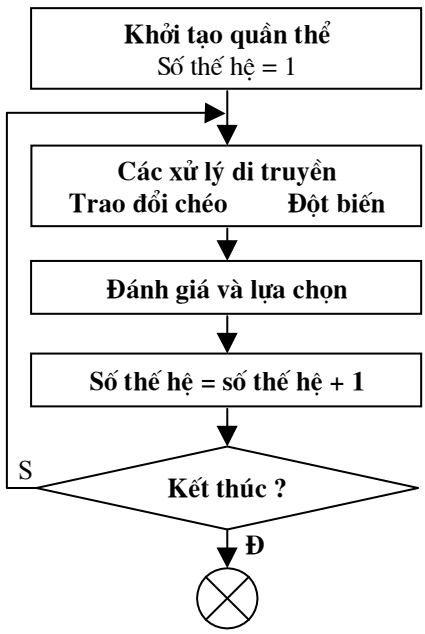
2. PHƯƠNG PHÁP PHÂN TÍCH BIỆT SỐ TUYẾN TÍNH BẰNG GIẢI THUẬT DI TRUYỀN

"*Phân tích biệt số tuyến tính*" (linear discriminant analysis) là phương pháp phân lớp trên cơ sở theo dõi những thuộc tính của các cá thể đã biết trước để xác định hàm phân lớp cho các cá thể (instance) mới. Trong một miền chỉ gồm 2 lớp, kết quả phân tích biệt số tuyến tính là hàm phân lớp có dạng sau:

$$f = a + w_1x_1 + w_2x_2 + ... + w_mx_m$$

trong đó $w_1, w_2, ..., w_m$ là các phần tử của vectơ hệ số w, còn $x_1, x_2, ..., x_m$ là giá trị các thuộc tính của cá thể cần phân lớp. Nếu cá thể cần xét có giá trị hàm phân lớp $f = a + w_i.x_i \leq c$, nó sẽ thuộc lớp thứ nhất, còn trong trường hợp ngược lại, nó sẽ thuộc lớp thứ 2. Các giá trị vector hệ số w và hằng số c có thể thu được thông qua việc phân tích tập các cá thể tương tự đã được phân lớp trước, gọi là *tập mẫu* (training set). Có nhiều phương pháp phân tích tập mẫu, trong đó giải thuật di truyền là một phương pháp mới được áp dụng trong thời gian gần đây. Việc phân lớp trước của tập mẫu này được tiến hành bằng các phương pháp phân lớp khác, mà thông thường là ý kiến của một nhóm chuyên gia.

"*Giải thuật di truyền*" (Genetic Algorithms - GA) là thuật ngữ dùng để chỉ một lớp các thuật toán tìm kiếm trên nguyên tắc mô phỏng quần thể gien tự nhiên để tìm ra giải pháp cần thiết. Giải thuật này đã được John Holland (trường Đại học Tổng hợp Michigan) mô tả từ những năm đầu của thập kỷ 70. Nguyên lý cơ bản của những giải thuật là duy trì một quần thể (population) các bộ gien (nhiễm sắc thể - chromosome) có cấu trúc xác định, mỗi bộ gien sẽ đại diện cho một phương pháp giải quyết vấn đề. Quần thể này được phát triển trên cơ sở cạnh tranh lẫn nhau trong điều kiện biến động có kiểm soát (các toán tử di truyền - genetic operators), bao gồm sự tái tổ hợp (recombination) hay còn gọi là *trao đổi chéo* (crossover), và *đột biến gien* (mutation) của các bộ gien trong quần thể. Những bộ gien có độ phù hợp (fitness) cao hơn sẽ được lựa chọn cho thế hệ sau. Các phép xử lý trao đổi chéo và đột biến gien sẽ góp phần tạo ra những bộ gien mới, làm thay đổi bộ mặt của quần thể. Tuy chỉ có tần số thấp, nhưng với số lượng cá thể lớn và kéo dài qua nhiều thế hệ, những phép xử lý di truyền bảo đảm sẽ không bỏ sót những giải pháp tối ưu không được tạo ra ngay ban đầu. Khi giải thuật kết thúc, người ta đã chứng minh được rằng những bộ gien ưu tú nhất, có mức độ thích nghi cao nhất đại diện cho giải pháp tốt nhất so với yêu cầu của bài toán sẽ được chọn. Về cơ bản, giải thuật di truyền có thể được mô hình hóa như sơ đồ bên.



Những phân tích trên lý thuyết đã cho thấy Giải thuật di truyền có khả năng tích lũy tri thức trong quá trình tìm kiếm thông qua sự truyền lại của bộ gien ưu việt cho thế hệ sau. Đặc biệt giải thuật di truyền cho phép giải những bài toán tối ưu phức tạp một cách nhanh chóng với độ chính xác chấp nhận được mà không cần biết trước phân bố các tham số của đối tượng cần phân lớp. Giải thuật này rất thích hợp với các mô hình để giải bài toán phân lớp trong y tế, nơi mà với mỗi ca bệnh các yếu tố cả bên trong lẫn bên ngoài có ảnh hưởng rất nhiều tới sự xuất hiện của các triệu chứng. Đây cũng là lý do tại sao chúng tôi chọn phương pháp phân tích biệt số tuyến tính bằng giải thuật di truyền trong đề tài nghiên cứu này.

3. THIẾT KẾ CHO GIẢI THUẬT DI TRUYỀN

Nếu coi mỗi bệnh nhân trong mô hình là một thực thể, thì những triệu chứng của bệnh nhân có thể được xem như những thuộc tính của thực thể đó. Giá trị phân lớp của mỗi thuộc tính sẽ bằng với giá trị hệ số tương ứng trong hàm phân lớp. Mục tiêu của bài toán là dựa vào tập các thực thể đã phân lớp trước (tập mẫu), tìm hàm phân lớp f, với vector hệ số w và hằng số c, cho phép xác định một thực thể mới sẽ thuộc về lớp thứ nhất hay lớp thứ hai (như đã trình bày trong Phần 2).

Trong giải thuật di truyền, mỗi phần tử của véctơ hệ số w trong hàm phân lớp f được đại diện bởi một đoạn gien của cá thể. Để tìm được bộ thuộc tính tối thiểu có khả năng phân lớp các cá thể, chúng tôi tiến hành xây dựng sơ đồ gien (genetic schemata) qua nhiều bước. Tại mỗi bước, một gien mới được bổ sung vào sơ đồ gien. Sơ đồ gien này sẽ được đánh giá qua nhiều thế hệ để tìm ra bản gien có độ phù hợp cao nhất. Giải thuật sẽ kết thúc khi bổ sung gien mới không thu được kết quả tốt hơn so với sơ đồ gien cũ.

Chúng tôi chọn ngôn ngữ Delphi 5.0 của Borland (nay là Inprise Inc.) để xây dựng chương trình vì đây là một công cụ lập trình hướng đối tượng 32 bit khá mạnh.

Sau đây là một số vấn đề cơ bản trong thiết kế của giải thuật di truyền.

3.1. Dữ liệu vào

Có 2 loại dữ liệu được cung cấp cho giải thuật:

- *Quần thể gien*: ban đầu bao gồm các bộ gien đại diện cho các vector hệ số của hàm phân lớp. Mỗi gien được mã hóa bằng một số nguyên không dấu 32 bit (kiểu cardinal trong Delphi). Khi đánh giá, giá trị của gien được đem chia cho giá trị tối đa của kiểu nguyên 32 bit (không dấu) để chuyển về giá trị thực trong khoảng [0,1]. Giá trị của các gien ban đầu được lấy ngẫu nhiên.

  Sự hiện diện của các gien trong sơ đồ gien được qui định bởi một thang đánh dấu (mask). Thang đánh dấu là một chuỗi ký tự (string) có độ dài bằng số lượng thuộc tính. Gien sẽ có mặt trong sơ đồ gien nếu ký tự đại diện cho nó trong thang đánh dấu nhận giá trị "1", trong trường hợp ngược lại (ký tự nhận giá trị "0"), gien sẽ không được đánh giá khi lựa chọn.
- *Tập mẫu*: là các thực thể đã được phân lớp trước. Trong nghiên cứu của chúng tôi, tập mẫu là các bệnh nhân trước đó đã được chẩn đoán xác định (thuộc về một trong hai bệnh cần nghiên cứu) bằng các phương pháp khác.

3.2. Kết quả

Yêu cầu của giải thuật là phải tìm ra được vector hệ số w của hàm phân lớp f và hằng số c cho phép xác định một thực thể mới sẽ thuộc lớp thứ nhất (nếu $f = a + w_i.x_i \leq c$) hay lớp thứ hai (nếu $f = a + w_i.x_i > c$) với x là vector thuộc tính của thực thể đang xét. Kết quả này thu được từ bản gien có mức độ phù hợp cao nhất của tất cả các sơ đồ gien có mặt trong giải thuật.

3.3. Các bước của giải thuật

3.3.1. Các toán tử di truyền (genetic operators)

- *Trao đổi chéo*: Phương pháp trao đổi chéo được sử dụng ở đây là phương pháp 2 điểm. Phương pháp này xác định 2 điểm ngẫu nhiên trên cặp nhiễm sắc thể, và đoạn gien tương ứng nằm giữa 2 điểm này sẽ được tráo đổi với nhau. Lựa chọn cá thể trao đổi chéo được tiến hành hoàn toàn ngẫu nhiên: mỗi quần thể có một tỷ lệ trao đổi chéo nhất định được qui định trước. Khi xảy ra trao đổi chéo, mỗi gien sẽ được gán một giá trị ngẫu nhiên (trong khoảng [0,1]), và nếu giá trị này nhỏ hơn tỷ lệ trao đổi chéo, cá thể sẽ được chọn. Các cá thể trao đổi chéo sau đó sẽ bắt cặp ngẫu nhiên để trao đổi chéo.
- *Đột biến*: Tương tự như quá trình lựa chọn cá thể trao đổi chéo, mỗi quần thể đều có một tỷ lệ đột biến qui định trước. Khi xảy ra đột biến mỗi cá thể sẽ được gán một giá trị ngẫu nhiên trong khoảng [0,1] và nếu giá trị này nhỏ hơn tỷ lệ đột biến của quần thể, thì cá thể sẽ được chọn để gây đột biến (đảo giá trị của bit tại vị trí đột biến). Vị trí đột biến trong nhiễm sắc thể được lựa chọn ngẫu nhiên.

3.3.2. Đánh giá (evaluation) và lựa chọn (selection)

Phương pháp đánh giá và lựa chọn được sử dụng trong chương trình như sau:

- *Đánh giá*: Mức độ phù hợp của cá thể được xác định bằng hiệu suất phân lớp của cá thể trên tập mẫu. Với mỗi hàm phân lớp f thu được từ bộ gien của cá thể, khi áp dụng vào tập mẫu sẽ tính được giá trị trọng tâm (centroid) của mỗi lớp. Giá trị hằng số c sẽ được tính bằng trung bình của 2 giá trị trọng tâm. Các kết quả này sẽ được sử dụng để phân lớp cho từng thực thể của tập mẫu, và hiệu suất phân lớp của cá thể được tính bằng tỷ lệ các thực thể của tập mẫu được phân lớp đúng (so với kết quả phân lớp có trước được tiến hành bằng phương pháp khác) sau khi đã loại bỏ tỷ lệ phân lớp đúng ngẫu nhiên.
- *Lựa chọn*: Mỗi cá thể khi hình thành đều được gán một thời gian sống nhất định (tính bằng số thế hệ cá thể có khả năng tồn tại) tỷ lệ với mức độ phù hợp của cá thể. Thời gian sống này sẽ được tính lại cho những cá thể tham gia trao đổi chéo hoặc đột biến. Sau mỗi thế hệ, thời gian sống của tất cả các cá thể đều bị giảm đi 1, và những cá thể nào có thời gian sống bằng 0 sẽ bị loại ra khỏi quần thể.

  Tỷ lệ tái tạo (reproduction rate) sẽ qui định số lượng cá thể được sinh ra trong mỗi thế hệ của quần thể. Việc lựa chọn cá thể tái tạo tiến hành dựa trên thuật toán "*bánh xe roulette*" được D.E. Goldberg mô tả [4]. Theo thuật toán này, mỗi cá thể sẽ chiếm một phần tương ứng với mức độ phù hợp của nó trên vành bánh xe. Khi cần chọn một cá thể, người ta sẽ chọn một điểm ngẫu nhiên trên vành bánh xe và cá thể được chọn là cá thể chứa điểm ngẫu nhiên đó. Thuật toán này bảo đảm những cá thể có độ phù hợp lớn sẽ có khả năng được lựa chọn cao hơn. Tuy nhiên, trong trường hợp các cá thể trong quần thể có độ phù hợp chênh lệch thấp (gần đạt tới điểm hội tụ), các cá thể có kích thước gần như nhau trên vành roulette và thuật toán này không bảo đảm chọn được những cá thể tốt nhất cho thế hệ sau. Vì vậy, chúng tôi áp dụng phương pháp cải tiến là "*phương pháp scaling*". Với phương pháp này, độ phù hợp của mỗi cá thể đều được đem trừ đi độ phù hợp nhỏ nhất của quần thể trước khi đưa vào vành roulette, vì thế mức độ chênh lệch giữa các cá thể sẽ tăng lên.

3.4. Điều kiện kết thúc giải thuật

Giải thuật sẽ kết thúc khi sơ đồ gien mới tạo có mức độ phù hợp không cao hơn mức độ phù hợp của sơ đồ gien hiện có, hoặc toàn bộ số gien có mặt trong giải thuật đã được sử dụng hết.

4. KẾT QUẢ THỬ NGHIỆM

Chúng tôi chọn hai bệnh sốt mò (scrub typhus) và thương hàn (typhoid fever) để thử nghiệm các chương trình xây dựng. Sở dĩ chúng tôi chọn 2 bệnh này vì cả sốt mò và thương hàn đều là những căn nguyên gây sốt kéo dài thường gặp trong những năm gần đây tại Hà Nội nói riêng và các tỉnh phía bắc nói chung. Hai bệnh có những triệu chứng lâm sàng tương tự nhau, và chẩn đoán phân biệt giữa hai bệnh này là việc phải tiến hành thường xuyên trong thực hành lâm sàng.

Số liệu tập mẫu được sử dụng là hồ sơ bệnh án của 2 đề tài nghiên cứu về bệnh thương hàn và bệnh sốt mò được tiến hành trong những năm gần đây: đề tài "*Điều tra tình hình bệnh thương hàn do Salmonella typhi kháng kháng sinh tại Hà Nội và nghiên cứu phác đồ điều trị thương hàn đa kháng hiện nay*" tiến hành trên những bệnh nhân đã được chẩn đoán xác định bệnh thương hàn tại Khoa Truyền nhiễm Bệnh viện Đống Đa, và đề tài "*Nghiên cứu bệnh cảnh lâm sàng, xét nghiệm và điều trị bệnh sốt mò Rickettsia tsutsugamushi trong 4 năm (từ 1997 đến 2000)*" được tiến hành tại Viện Y học lâm sàng các bệnh nhiệt đới năm 2001. Tổng cộng 526 bệnh án thuộc 2 cuộc nghiên cứu nói trên đã được sử dụng, trong đó bệnh sốt mò có 391 bệnh án (74,3%) và thương hàn có 135 bệnh án (25,7%). Trong mỗi bệnh án, 13 triệu chứng được thu thập là: sốt kéo dài, nhiễm trùng nhiễm độc, sốc, gan to, bụng chướng, rối loạn tiêu hoá, phát ban, hạch to, da xung huyết, eschar, nhức đầu, số lượng bạch cầu giảm, yếu tố dịch tễ. Để đánh giá hiệu quả của giải thuật di truyền, chúng tôi so sánh kết quả của chương trình với hàm phân lớp thu được bằng Phương pháp phân tích biệt số tuyến tính Fisher (SPSS for Windows 8.0) trên cùng một tập dữ liệu, và so sánh với thực tế lâm

sàng qua ý kiến của các chuyên gia đầu ngành. Kết quả thu được từ phương pháp phân tích biệt số tuyến tính bằng giải thuật di truyền như sau:

Bảng 1. Kết quả thử nghiệm.

THUỘC TÍNH	HỆ SỐ	THUỘC TÍNH	HỆ SỐ
Nhiễm trùng nhiễm độc	0,6076968619	Eschar	-0,9548554309
Bụng chướng	0,5418880697	Hạch to	-0,7157130976
Rối loạn tiêu hoá	0,1773805710	Xung huyết	-0,6693544257
Số lượng bạch cầu giảm	0,1696534269	Nhức đầu	-0,2530968444

Giá trị trọng tâm (centroid) của các lớp:

Sốt mò = -1,91729332749902; Thương hàn = 0,291548177313886.

So với phương pháp phân tích biệt số tuyến tính Fisher được tiến hành với cùng bộ số liệu bằng chương trình SPSS for Windows 8.0, chúng tôi cũng thu được kết quả tương tự.

Các triệu chứng có mặt trong kết quả mà chúng tôi thu được cũng tương đương với các triệu chứng để nhận ra bệnh thương hàn tại cộng đồng trong mô hình chẩn đoán của Burley (WHO), của Viện Y học lâm sàng các bệnh nhiệt đới [2], và đề tài nghiên cứu của Đỗ Văn Thành [6]. Trong thực tế, lâu nay để xác định hay loại trừ bệnh sốt mò, các chuyên gia thường dựa vào dấu hiệu eschar; và ở những bệnh nhân không có triệu chứng này, các dấu hiệu đào ban, hạch to, xung huyết có tác dụng hướng tới chẩn đoán sốt mò một cách tương đối chắc chắn. Việc chẩn đoán thương hàn thường khó khăn hơn, và trong điều kiện không có những xét nghiệm đặc hiệu thì chỉ có thể hướng tới thương hàn nếu bệnh nhân có biểu hiện nhiều triệu chứng như bụng chướng, rối loạn tiêu hóa, nhiễm trùng nhiễm độc,.... Tất cả những điều này cũng rất gần các hệ số của các triệu chứng mà chúng tôi thu được trên đây.

5. KẾT LUẬN

Trong khuôn khổ của đề tài này, chúng tôi đã bước đầu áp dụng được phương pháp phân tích biệt số tuyến tính dựa trên giải thuật di truyền để giải quyết một bài toán đặt ra trong thực tế là tạo lập mô hình có khả năng trợ giúp các chuyên gia y tế trong việc xây dựng tiêu chuẩn chẩn đoán phân biệt giữa 2 bệnh có triệu chứng tương tự nhau trên lâm sàng, và kiểm tra hoạt động của mô hình này với số liệu thực tế của 2 bệnh thương hàn và sốt mò. Tuy kết quả chưa được chứng minh bằng thực nghiệm, nhưng so sánh với ý kiến nhận xét của những chuyên gia đầu ngành trong lĩnh vực Truyền nhiễm, thì những kết quả của chương trình rất gần với những kinh nghiệm lâm sàng mà các thầy thuốc thu được.

Để tiếp tục hoàn thiện mô hình chẩn đoán phân biệt bằng giải thuật di truyền, các nghiên cứu tiếp theo chúng tôi sẽ tập trung làm rõ một số vấn đề như:

+ Kiểm chứng lại mô hình trên thực nghiệm.
+ Mở rộng cho nhiều bài toán khác trong y tế như vấn đề tiên lượng bệnh, dự báo dịch,...
+ Cải tiến mô hình để chẩn đoán phân biệt nhiều bệnh.
+ Xây dựng mô hình chẩn đoán trong điều kiện không cân bằng (số lượng người mắc một bệnh đang quan tâm chỉ chiếm một phần rất nhỏ trong tổng số người đưa vào nghiên cứu).
+ Xây dựng mô tơ suy diễn mờ trên cơ sở các kết quả thu được để hình thành một hệ thống hỗ trợ chẩn đoán hoàn chỉnh.

TÀI LIỆU THAM KHẢO

[1] Aaron H. Konstam, Linear discriminant analysis using genetic algorithms, *Proceedings of the 1993 ACM/ SIGAPP Symposium on Applied Computing*, Indianapolis (USA), Feb. 14-16, 1993.

[2] Cao Văn Viên, Bệnh thương hàn, Tài liệu đào tạo chuyên ngành Truyền nhiễm, Bệnh viện Bạch Mai, 2001.

[3] Cao Văn Viên, Đỗ Văn Thành, Nhiễm trùng Rickettsia, Tài liệu đào tạo chuyên ngành Truyền nhiễm, Bệnh viện Bạch Mai, 2001.
[4] D. E. Goldberd, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA, 1989.
[5] Dương Đình Thiện *Dịch tễ học lâm sàng*, tập I. NXB Y học, Hà Nội, 1998.
[6] Đỗ Văn Thành, "Nghiên cứu bệnh cảnh lâm sàng, xét nghiệm và điều trị bệnh sốt mò do Rickettsia tsutsugamushi", Luận văn tốt nghiệp Bác sĩ nội trú bệnh viện, 1998, Đại học Y Hà Nội.
[7] John C. Gallagher, "A dynamical systems analysis of the neural basis of behavior in an artificial autonomous agent", Thesis of Doctor of Philosophy, 1998, Case Western Reserve University.
[8] John J. Grefenstette, Introduction: special track on genetic algorithms, *IEEE Expert* **10** (1993).
[9] John J. Grefenstette, Predictive models using fitness distributions of Genetic operators, *Foundation of Genetic Algorithms 3*, Morgan Kauffmann, USA, 1995.
[10] Trần Thị Phi La "Lâm sàng và chẩn đoán phân biệt bệnh thương hàn", Luận án tiến sĩ, 2000, Đại học Y Hà Nội.
[11] *SPSS 8.0 for Windows manual*, SPSS Inc., USA, 1997.

Nhận bài ngày 15 - 12- 2001

Nguyễn Công Phúc - Bệnh viện Đống Đa, Hà Nội
Doãn Hồ Liên - Khoa CNTT, Đại học Bách khoa Hà Nội
Cù Thu Thủy - Trung tâm Tin học, Đại học Tài chính Kế toán Hà Nội
Cao Văn Viên - Viện Y học lâm sàng các bệnh nhiệt đới, Bệnh viện Bạch Mai