

**ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**

LƯƠNG THỊ THU HÀ

**GIẢI THUẬT DI TRUYỀN VÀ ỨNG DỤNG ĐỐI VỚI BÀI TOÁN XÁC
ĐỊNH CÔNG THỨC HỒI QUY TRONG THÍ NGHIỆM HÓA SINH**

**LUẬN VĂN THẠC SĨ
KHOA HỌC MÁY TÍNH**

THÁI NGUYÊN 2015

LỜI CAM ĐOAN

Sau quá trình học tập tại **Trường Đại học công nghệ thông tin & truyền thông**, với những kiến thức lý thuyết và thực hành đã tích lũy được, với việc vận dụng các kiến thức vào thực tế, em đã tự nghiên cứu các tài liệu, các công trình nghiên cứu, đồng thời có sự phân tích, tổng hợp, đúc kết và phát triển để hoàn thành luận văn thạc sĩ của mình.

Em xin cam đoan luận văn này là công trình do bản thân em tự tìm hiểu, nghiên cứu và hoàn thành dưới sự hướng dẫn của thầy giáo **TS. Vũ Vinh Quang**.

Thái Nguyên, tháng 5 năm 2015

Học viên

Lương Thị Thu Hà

LỜI CẢM ƠN

Trong thời gian hai năm của chương trình đào tạo thạc sỹ, trong đó gần một nửa thời gian dành cho các môn học, thời gian còn lại dành cho việc lựa chọn đề tài, giáo viên hướng dẫn, tập trung vào nghiên cứu, viết, chỉnh sửa và hoàn thiện đề tài. Với quỹ thời gian như vậy và với vị trí công việc đang phải đảm nhận, không riêng bản thân em mà hầu hết các sinh viên cao học muốn hoàn thành tốt luận văn của mình trước hết đều phải có sự sắp xếp thời gian hợp lý, có sự tập trung học tập và nghiên cứu với tinh thần nghiêm túc, nỗ lực hết mình; tiếp đến cần có sự ủng hộ về tinh thần, sự giúp đỡ về chuyên môn một trong những điều kiện không thể thiếu quyết định đến việc thành công của đề tài.

Để hoàn thành được đề tài này trước tiên em xin gửi lời cảm ơn đến thầy giáo hướng dẫn **TS. Vũ Vinh Quang**, người đã có những định hướng cho em về nội dung và hướng phát triển của đề tài, người đã có những đóng góp quý báu cho em về những vấn đề chuyên môn của đề tài, giúp em tháo gỡ kịp thời những vướng mắc trong quá trình làm luận văn.

Em xin gửi lời cảm ơn tới các Cán bộ nghiên cứu thuộc Viện Hóa sinh biển thuộc Viện Hàn lâm khoa học và Công nghệ Việt Nam đã cung cấp đầy đủ các số liệu thu được từ các phòng thí nghiệm tại Viện để giúp đỡ Em tiến hành các thí nghiệm thành công.

Em cũng xin cảm ơn các Thầy Cô giáo Trường Đại học Công nghệ thông tin và Truyền thông Thái Nguyên, cũng như bạn bè cùng lớp đã có những ý kiến đóng góp bổ sung cho đề tài luận văn của em. Xin cảm ơn gia đình, người thân cũng như đồng nghiệp luôn quan tâm, ủng hộ hỗ trợ về mặt tinh thần trong suốt thời gian từ khi nhận đề tài đến khi hoàn thiện đề tài này.

Trong nội dung của luận văn chắc chắn còn nhiều thiếu sót. Em rất mong các Thầy Cô cùng bạn bè đóng góp để bản luận văn của Em được hoàn thiện hơn.

Em xin trân trọng cảm ơn.

Thái Nguyên, tháng 5 năm 2015

Học viên

Lương Thị Thu Hà

MỤC LỤC

LỜI MỞ ĐẦU	7
Chương 1 CÁC KIẾN THỨC CƠ BẢN VỀ HÀM HỒI QUY THỰC NGHIỆM	9
1.1. Khái niệm cơ bản về hàm nội suy.....	9
1.1.1 Đa thức nội suy.....	10
1.1.2 Đa thức nội suy Lagrange.....	11
1.1.3 Hàm ghép trơn (Spline)	12
1.1.4 Nội suy bằng hàm hữu tỉ	14
1.2 Bài toán hồi quy	14
1.2.1 Phương pháp bình phương cực tiểu.....	15
1.2.2 Hàm hồi quy tuyến tính	16
1.2.3 Hàm hồi quy bậc 2.....	17
1.2.4 Các phương pháp đưa về dạng tuyến tính	17
1.2.5 Hồi quy nhiều chiều (hồi quy bội).....	18
Chương 2 MỘT SỐ KIẾN THỨC CƠ BẢN VỀ GIẢI THUẬT DI TRUYỀN ..	20
2.1 Các khái niệm cơ bản.....	21
2.1.1 Cá thể, nhiễm sắc thể.....	21
2.1.2 Quần thể.....	21
2.1.3 Chọn lọc.....	21
2.1.4 Lai ghép (Cross-over).....	22
2.1.5 Đột biến (Mutation).....	22
2.1.6 Các tham số của GA	23
2.2 Cơ chế thực hiện của thuật toán di truyền.....	24
2.2.1 Mã hóa	24
2.2.2 Khởi tạo quần thể ban đầu	26
2.2.3 Xác định hàm thích nghi	26
2.2.4 Cơ chế lựa chọn.....	26
2.2.5 Các toán tử di truyền	28
2.3. Thuật toán di truyền kinh điển (GA).....	29
2.3.1 Mã hóa	29

2.3.2 Toán tử chọn lọc.....	30
2.3.3 Toán tử lai ghép.....	31
2.3.4 Toán tử đột biến.....	32
2.4 Thuật toán di truyền mã hoá số thực (RCGA).....	34
2.5. Một số ứng dụng của GA.....	40
Chương 3 BÀI TOÁN MÔ PHỎNG QUÁ TRÌNH CHIẾT XUẤT DUNG MÔI ...	43
3.1. Mô hình bài toán	43
3.2 Xây dựng mô hình GA.....	46
3.2.1 Phương pháp biểu diễn cá thể	46
3.2.2 Xác định hàm thích nghi	47
3.2.3 Các toán tử di truyền	47
3.2.4 Quá trình khởi tạo quần thể.....	48
3.3 Kết quả thực nghiệm	49
KẾT LUẬN.....	54
TÀI LIỆU THAM KHẢO.....	55
PHẦN PHỤ LỤC.....	56

DANH MỤC HÌNH

Hình 2.1 Sơ đồ mô tả GA.....	22
Hình 2.2 Lai ghép CMX.....	38
Hình 2.3 Phân bố của x_j^{ci}	38
Hình 2.4 Toán tử lai ghép SX	39
Hình 3.1 Thiết bị thí nghiệm chiết xuất dung môi	43
Hình 3.2 Biểu đồ biểu diễn giá trị của hàm	51
Hình 3.3 Biểu đồ biểu diễn giá trị của hàm	53

LỜI MỞ ĐẦU

Trong khoa học thực nghiệm, thông qua các kết quả thực nghiệm một vấn đề rất quan trọng là xuất phát từ các bộ số liệu thực nghiệm hay còn gọi là các mốc hồi quy, ta cần phải xác định một quan hệ thống kê giữa các đối tượng sao cho quan hệ này là xấp xỉ tốt nhất ứng với các mốc hồi quy đã xác định. Về mặt toán học việc xác định quan hệ thống kê này thường đưa đến việc xác định các tham số chưa biết thông qua một bài toán cực trị được mô tả bằng phương pháp bình phương tối thiểu và chuyển bài toán về việc giải các hệ phương trình đại số tuyến tính hoặc các hệ phi tuyến tính. Đối với các bài toán này thì khối lượng tính toán là tương đối lớn đối với các hệ đại số tuyến tính còn đối với các hệ phi tuyến thì đại đa số chúng ta không thể xác định được nghiệm của hệ.

Thuật giải di truyền GA (Genetic Algorithm) là một trong những kỹ thuật tìm kiếm lời giải tối ưu đã đáp ứng được yêu cầu của nhiều bài toán và ứng dụng. Điểm mạnh của GA là cho phép xác định lời giải gần tối ưu của các bài toán cực trị thông qua các phép toán lai ghép và chọn lọc các phương án của bài toán với cơ chế hết sức đơn giản nhưng rất hiệu quả. Trong công nghệ thông tin hiện nay, giải thuật GA kết hợp với logic mờ, mạng Nơron đã được ứng dụng nhiều trong lớp các bài toán NP.

Xuất phát từ lý do đó, đề tài đặt vấn đề nghiên cứu về GA và ứng dụng trong việc xác định các công thức hàm hồi quy, ứng dụng vào bài toán xác định công thức gần đúng trong thí nghiệm hóa sinh.

Với những lý do trên, em chọn đề tài: “***Giải thuật di truyền và ứng dụng đối với bài toán xác định công thức hồi quy trong thí nghiệm hóa sinh***” làm luận văn tốt nghiệp.

Nội dung chính của luận văn gồm 3 chương bao gồm:

Chương 1: Trình bày cơ sở toán học trong việc xác định công thức hàm nội suy và hàm hồi quy cùng các thuật toán tương ứng, đây là một lĩnh vực quan trọng của toán học đối với lớp các bài toán thực nghiệm nhằm xây dựng các công thức gần đúng miêu tả mối ràng buộc giữa các số liệu xuất hiện trong các thí

nghiệm tại các phòng thí nghiệm. Các kiến thức này là rất cần thiết làm cơ sở để nghiên cứu các nội dung trong luận văn.

Chương 2: Trình bày các kiến thức cơ bản về giải thuật di truyền, một trong những giải thuật đã và đang được phát triển trong công nghệ thông tin giải quyết các bài toán tối ưu hóa theo tư tưởng quần thể ngẫu nhiên. Thuật toán GA chính là cơ sở để xây dựng thuật toán giải bài toán thực tế được đưa ra trong chương 3.

Chương 3: Nội dung chính của chương 3 trình bày mô hình *bài toán chiết xuất dung môi*, một bài toán quan trọng trong các thí nghiệm về hóa sinh. Trên cơ sở mô hình bài toán, luận văn đã xây dựng thuật toán GA giải quyết bài toán, tiến hành thực nghiệm với số liệu được cung cấp của phòng thí nghiệm tại Viện Hóa sinh biển thuộc Viện Hàn lâm và khoa học Công nghệ Việt Nam. Tiến hành đánh giá và kết luận về mối ràng buộc giữa các số liệu thực nghiệm.

Trong luận văn, các kết quả thực nghiệm được lập trình trên môi trường Matlab version 7.0

Chương 1

CÁC KIẾN THỨC CƠ BẢN VỀ HÀM HỒI QUY THỰC NGHIỆM

Trong chương này, luận văn trình bày một số kiến thức cơ bản về cơ sở toán học trong việc xây dựng các hàm công thức hàm nội suy và hàm hồi quy. Các kiến thức này làm cơ sở trong việc nghiên cứu các chương tiếp sau của luận văn, các kết quả trong chương 1 được tham khảo trong các tài liệu [1, 2, 3, 4, 5]

1.1. Khái niệm cơ bản về hàm nội suy

Chúng ta xét một dạng bài toán xuất phát từ các số liệu thực nghiệm sau:

Cho trước $(n + 1)$ cặp các giá trị thực nghiệm $(x_i, y_i), i = 0, 1, \dots, n$

x_0	x_1	x_2	x_3	x_4	x_n
y_0	y_1	y_2	y_3	y_4	y_n

Các giá trị $(x_i, y_i), i = 0, 1, \dots, n$ được gọi là các mốc nội suy. Cần xác định một hàm số $f(x)$ để sao cho thỏa mãn các điều kiện

$$f(x_i) = y_i, i = 0, 1, \dots, n$$

Tức là đồ thị của hàm $f(x)$ cần đi qua tất cả các mốc nội suy.

Nếu hàm $f(x)$ tồn tại thì hàm số đó được gọi là hàm hồi quy và bài toán xác định $f(x)$ được gọi là bài toán nội suy.

Bài toán này rất có ý nghĩa trong thực tế vì nếu xác định được hàm $f(x)$ thì ta có thể xác định được mọi giá trị của y ứng với mọi $x \in [x_0, x_n]$ - các giá trị đó được gọi là các giá trị nội suy.

Trong toán học, người ta thường xác định dạng hàm $f(x)$ bởi một trong các dạng công thức hàm số sau đây:

- + Đa thức đại số
- + Phân thức đại số
- + Đa thức lượng giác
- + Hàm ghép trơn (Spline)
- + Hàm mũ

Sau đây chúng ta sẽ xét cơ sở toán học của các phương pháp xác định hàm hồi quy trong từng trường hợp cụ thể

1.1.1 Đa thức nội suy

Ta sẽ xác định

$$f(x) = a_0x^n + a_1x^{n-1} + \dots + a_n \quad (1.1)$$

Xuất phát từ điều kiện hàm $f(x)$ cần phải đi qua tất cả các mốc nội suy, dễ thấy rằng các hệ số $a_k, k = 0, 1, \dots, n$ sẽ được xác định thông qua hệ phương trình đại số tuyến tính sau đây:

$$\begin{cases} a_0x_0^n + a_1x_0^{n-1} + \dots + a_n = y_0 \\ a_0x_1^n + a_1x_1^{n-1} + \dots + a_n = y_1 \\ \dots \\ a_0x_n^n + a_1x_n^{n-1} + \dots + a_n = y_n \end{cases} \quad (1.2)$$

Có thể thấy rằng định thức của hệ đại số tuyến tính (1.2) thỏa mãn

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ x_0 & x_1 & \dots & x_n \\ x_0^2 & x_1^2 & \dots & x_n^2 \\ \dots & \dots & \dots & \dots \\ x_0^n & x_1^n & \dots & x_n^n \end{vmatrix} \neq 0 \quad (1.3)$$

Do đó hệ phương trình đại số trên có nghiệm duy nhất, tức là đa thức nội suy luôn luôn tồn tại và duy nhất.

Nhận xét: Để xác định đa thức nội suy theo phương pháp đại số, ta cần phải giải hệ phương trình đại số tuyến tính với $n+1$ ẩn $a_k, k = 0, 1, \dots, n$. Khi đó về mặt toán học, chúng ta cần phải sử dụng các phương pháp giải các hệ phương trình đại số tuyến tính như phương pháp Krame, phương pháp khử Gauss, ... với độ phức tạp tính toán rất cao. Điều này sẽ bất lợi trong việc xác định đa thức nội suy với số mốc nội suy là rất lớn.

Sau đây chúng ta sẽ tìm hiểu các phương pháp xác định đa thức nội suy tránh được việc giải hệ đại số tuyến tính.