

BỘ GIÁO DỤC VÀ ĐÀO TẠO
ĐẠI HỌC ĐÀ NẴNG

PHẠM HỮU LÊ QUỐC PHỤC

NGHIÊN CỨU ỨNG DỤNG
MẠNG NƠ-RON NHÂN TẠO GIẢI QUYẾT
LỚP BÀI TOÁN DỰ ĐOÁN VÀ PHÂN LOẠI

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 60.48.01

TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT

ĐÀ NẴNG – NĂM 2010

Công trình được hoàn thành tại
ĐẠI HỌC ĐÀ NẴNG

Người hướng dẫn khoa học: PGS. TS. Phan Huy Khánh

Phản biện 1: PGS. TS. Lê Văn Sơn

Phản biện 2: TS. Trương Công Tuấn

Luận văn sẽ được bảo vệ tại Hội đồng chấm Luận văn tốt nghiệp thạc sĩ kỹ thuật họp tại Đại học Đà Nẵng vào ngày 15 tháng 10 năm 2010.

Có thể tìm hiểu luận văn tại:

- Trung tâm Thông tin-Học liệu, Đại học Đà Nẵng
- Trung tâm Học liệu, Đại học Đà Nẵng

MỞ ĐẦU

1. Lý do chọn đề tài

Trong bối cảnh ứng dụng công nghệ thông tin ngày càng tăng, dữ liệu phát sinh từ hoạt động quản lý, kinh doanh, sản xuất của các công ty, tổ chức ngày càng nhiều. Các công ty, tổ chức cần phải nhanh chóng đưa ra các quyết định bằng cách xử lý nhiều yếu tố với quy mô và tính phức tạp ngày càng tăng. Để có quyết định chính xác nhất, người quản lý thường thực hiện việc dự đoán hay phân loại vấn đề cần giải quyết trước khi ra quyết định. Ngoài việc dựa trên các yếu tố liên quan trực tiếp đến vấn đề, người ra quyết định còn dựa trên kinh nghiệm bản thân và thông tin có được từ các hoạt động trước đó. Dẫn đến một nhu cầu thực tế là cần có các phương pháp khai phá dữ liệu thu thập được để làm căn cứ ra quyết định.

Trong thực tế cuộc sống, chúng ta bắt gặp nhiều bài toán tương tự như dự đoán thị trường chứng khoán, dự đoán lưu lượng nước, dự đoán lượng gas tiêu, dự đoán năng lực sản xuất, định giá tài sản, đánh giá nhân viên, phân loại khách hàng. Đó là các bài toán thuộc lớp bài toán dự đoán và phân loại, có thể xem là các bài toán cơ bản và có nhiều ứng dụng thực tiễn. Đã có nhiều phương pháp được đưa ra để giải các lớp bài toán trên như phương pháp thống kê, phương pháp hồi quy, cây quyết định, mạng nơ-ron nhân tạo, ...

Trong đó, mạng nơ-ron nhân tạo nhờ khả năng học, nhớ lại và khái quát hóa từ các mẫu dữ liệu huấn luyện, đã trở thành một trong hướng nghiên cứu chính của lĩnh vực trí tuệ nhân tạo. Xét về mặt lý thuyết, mạng nơ-ron nhân tạo tương đối độc lập với bản chất của các quá trình vật lý cần phân loại, dự đoán. Ở mỗi lớp bài toán

trên, đều có các đặc điểm chung khi giải bằng mạng nơ-ron nhân tạo như: thu thập dữ liệu mẫu, tiền xử lý dữ liệu, xác định thông số mạng và huấn luyện. Do đó việc nghiên cứu để tổng quát hóa các bài toán và xây dựng phần mềm ứng dụng mạng nơ-ron nhân tạo có thể dùng cho nhiều bài toán cùng một lớp là hoàn toàn khả thi.

Theo ghi nhận của một nghiên cứu, hơn năm mươi phần trăm các báo cáo nghiên cứu về mạng nơ-ron là sử dụng mạng nơ-ron truyền thẳng và thuật toán lan truyền ngược. Mô hình mạng này được sử dụng rộng rãi bởi vì có khả năng giải được nhiều bài toán ở các lĩnh vực khác nhau: dự đoán, phân loại, mô hình hóa. Mạng nơ-ron nhân tạo truyền thẳng thích hợp để giải quyết bài toán thể hiện mối quan hệ giữa một tập hợp đầu vào và đầu ra biết trước.

Một trong số các trở ngại gặp phải khi ứng dụng mạng nơ-ron nhân tạo cần phải có sự hỗ trợ đầy đủ kiến thức lý thuyết và phương pháp ứng dụng. Trong khi các nghiên cứu về mạng nơ-ron nhân tạo thường ứng dụng vào một bài toán cụ thể, kết quả nghiên cứu khó có khả năng kế thừa, phát triển để ứng dụng rộng rãi cho các bài toán tương tự. Vì vậy việc nghiên cứu chuyên sâu, đầy đủ và mang tính ứng dụng thực tiễn cao là hết sức cần thiết.

2. Mục tiêu và nhiệm vụ chính

Tìm hiểu các đặc trưng của mạng nơ-ron nhân tạo, khả năng và các nguyên tắc để ứng dụng thành công mạng nơ-ron nhân tạo trong thực tế. Nghiên cứu ứng dụng mạng nơ-ron nhân tạo vào lớp bài toán dự đoán và phân loại. Xây dựng phần mềm cho phép người sử dụng mô phỏng và ứng dụng nhanh chóng mạng nơ-ron nhân tạo để giải quyết các bài toán thuộc lớp bài toán phân loại và dự đoán.

3. Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu là lớp bài toán dự đoán và phân loại, sử dụng mạng nơ-ron nhân tạo truyền thẳng huấn luyện bằng thuật toán lan truyền ngược. Phạm vi nghiên cứu là lý thuyết ứng dụng mạng nơ-ron nhân tạo cho bài toán dự đoán và phân loại. Hai kiểu dự đoán được nghiên cứu là (1) dự đoán theo chuỗi thời gian(time-series) trong dự đoán tiêu thụ kem phụ thuộc theo chuỗi thời gian và có tham số ngưỡng; (2) dự đoán mô hình hóa(modeling) trong bài toán xây dựng mô hình định giá giá nhà ở. Với lớp bài toán phân loại tác giả chọn bài toán đánh giá tình trạng tài chính cá nhân.

4. Phương pháp nghiên cứu

Phương pháp nghiên cứu tài liệu: nghiên cứu lý thuyết và ứng dụng mạng nơ-ron nhân tạo trong bài toán dự đoán và phân loại.

Phương pháp thực nghiệm: đi sâu nghiên cứu ứng dụng mạng nơ-ron nhân tạo bắt đầu từ bước chuẩn bị dữ liệu, bao gồm các kỹ thuật cho việc trích chọn đặc trưng, làm sạch dữ liệu, tiền xử lý, kiến trúc mạng, cách huấn luyện và kiểm tra mạng. Thực hiện phân tích ứng dụng mạng nơ-ron vào một số bài toán của mỗi lớp bài toán. Từ các phân tích từng bài toán, tác giả xây dựng thành quy trình, các chỉ dẫn mang tính ứng dụng thực tiễn cao có thể ứng dụng nhanh chóng cho các bài toán tương tự của các lớp bài toán trên.

Xây dựng phần mềm mô phỏng mạng nơ-ron: phân tích, thiết kế phần mềm hướng đối tượng với các tính năng cho phép người sử dụng thực hiện giải các bài toán thực tế bằng mạng nơ-ron nhân tạo. Lập trình phần mềm bằng ngôn ngữ lập trình C++, phần mềm có giao diện trực quan chạy trên hệ điều hành Windows.

5. Ý nghĩa khoa học và thực tiễn của đề tài

Đề tài làm rõ khả năng ứng dụng của mạng nơ-ron trong bài toán dự đoán và phân loại. Cách để xác định bài toán nào thích hợp để giải bằng mạng nơ-ron. Xây dựng thành quy trình với các bước thực hiện cụ thể cho việc giải bài toán dự đoán và phân loại bằng mạng nơ-ron.

6. Bố cục luận văn

Toàn bộ nội dung luận văn được trình bày trong 3 chương:

Chương 1: trình bày tổng quan về mạng nơ-ron, giới thiệu mạng nơ-ron truyền thẳng và thuật toán lan truyền ngược. Trình bày các vấn đề liên quan như sự hội tụ, cách đánh giá kết quả huấn luyện.

Chương 2: trình bày ứng dụng mạng nơ-ron vào một số bài toán cụ thể trong các lớp bài toán dự đoán và phân loại. Đánh giá và nhận xét khả năng ứng dụng mạng nơ-ron trong mỗi lớp bài toán.

Chương 3: trình bày phương pháp tổng quát ứng dụng mạng nơ-ron, bao gồm các nguyên tắc khi sử dụng mạng nơ-ron, cách chọn dữ liệu, trích chọn đặc trưng, cách biểu diễn dữ liệu đầu vào đầu ra cho mạng, một số quy tắc cho việc thiết kế mạng hiệu quả. Xây dựng thành quy trình với các bước thực hiện làm tiền đề cho việc xây dựng phần mềm mô phỏng. Thiết kế phần mềm mô phỏng mạng nơ-ron.

CHƯƠNG 1 - TỔNG QUAN VỀ MẠNG NƠ-RON NHÂN TẠO

1.1 Giới thiệu

1.2 Các thành phần chính của mạng nơ-ron

1.2.1 Mô hình tính toán của nơ-ron(Đơn vị xử lý)

1.2.2 Hàm kết hợp(Summation Function)

1.2.3 Hàm kích hoạt (Active Function)

1. Hàm đồng nhất(Linear function, Identity function)
2. Hàm bước nhị phân(Binary step function)
3. Hàm sigmoid nhị phân(Sigmoid function, logsig)
4. Hàm sigmoid lưỡng cực(Bipolar sigmoid function)

1.3 Phương pháp huấn luyện mạng

1. Học có giám sát(Supervised Learning)
2. Học không có giám sát(Unsupervised Learning)
3. Học tăng cường(Hybrid learning)

1.4 Cấu trúc mạng nơ-ron

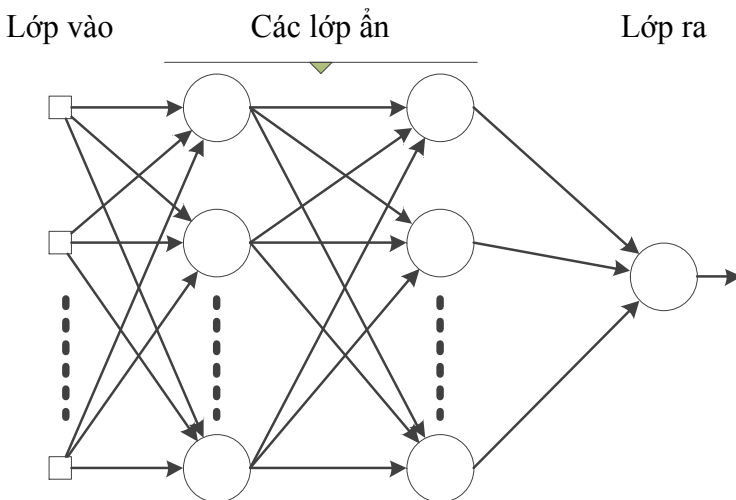
1.5 Mô hình mạng nơ-ron

Mô hình mạng nơ-ron là sự kết hợp giữa cấu trúc mạng, phương pháp huấn luyện và thuật toán huấn luyện. Có nhiều mô hình mạng cho nhiều mục đích sử dụng khác nhau.

1.6 Mạng nơ-ron lan truyền ngược

1.6.1 Kiến trúc

Mạng nơ-ron lan truyền ngược là mạng có cấu trúc truyền thẳng, học có giám sát và sử dụng thuật toán lan truyền ngược. Mạng nơ-ron lan truyền ngược có thể dùng cho các bài toán: phân loại, mô hình hóa, xấp xỉ hàm, dự đoán chuỗi thời gian.



Hình 1.10: Mô hình chi tiết một mạng nơ-ron truyền thẳng

1.6.2 Thuật toán lan truyền ngược

Thuật toán có hai giai đoạn tính toán tách biệt nhau: giai đoạn tiến(forward) và giai đoạn lùi(backward).

Trong giai đoạn tiến, các tín hiệu được tính toán dựa trên cơ sở từ nơ-ron này đến nơ-ron khác theo chiều tiến của mạng. Bắt nguồn từ đầu vào của mẫu dữ liệu, quá trình toán tính cứ tiếp diễn cho tới khi tính được đầu ra của các nơ-ron nằm trong lớp đầu ra của

mạng. Giá trị đầu ra này được so sánh với giá trị đầu ra mong muốn từ đó thu nhận tín hiệu lỗi.

Giai đoạn tính toán lùi bắt đầu tại mức đầu ra bằng cách chuyển tín hiệu lỗi ngược trở lại qua toàn bộ mạng theo từng mức nhằm tính toán gradient cục bộ mỗi nơ-ron. Quá trình này cho phép các trọng số của mạng có thể điều chỉnh.

1.6.3 Các tham số chính

1. Hệ số học(Learning Rate)
2. Bước đà(Momentum)
3. Độ chịu lỗi(Error Tolerance)
4. Đo lường độ chính xác

1.6.4 Sự hội tụ

Nói chung, thuật toán lan truyền ngược không thể xác định được là đã hội tụ hay chưa, như vậy sẽ không có một tiêu chuẩn tuyệt đối nào cho việc dừng thuật toán. Thông thường ta sử dụng độ lỗi trong quá trình huấn luyện làm cơ sở dừng thuật toán. Ngoài ra còn có phương pháp vừa học vừa thực hiện kiểm tra bằng cách tính độ đo lỗi trên tập dữ liệu kiểm tra khác với tập dữ liệu huấn luyện, thuật toán xem là hội tụ khi đạt độ chính xác tới giá trị ngưỡng cho trước.

CHƯƠNG 2 - PHÂN TÍCH BÀI TOÁN DỰ ĐOÁN, PHÂN LOẠI

2.1 Bài toán dự đoán theo dạng mô hình hóa

2.1.1 *Phát biểu bài toán*

Định nghĩa bài toán: ta có các tham số chưa biết α , có thể là một giá trị vô hướng hoặc một vector có chiều dài k ; các biến độc lập X và biến phụ thuộc Y ; cần xây dựng hàm f để xác định Y dựa trên X : $Y \approx f(X, \alpha)$. Bài toán đặt ra là tìm giá trị các tham số chưa biết α dựa trên tập dữ liệu mẫu thu thập được của tập X, Y .

2.1.2 *Ứng dụng mạng nơ-ron trong bài toán dự đoán giá nhà ở*

2.1.2.1 *Yêu cầu*

Công việc của người định giá bất động sản là xem xét căn nhà với các thuộc tính như cũ hay mới, độ rộng, có bao nhiêu phòng ngủ, phòng tắm, ... Tóm lại người định giá sẽ thu thập thông tin về thuộc tính của tài sản, đánh giá mức ảnh hưởng của thuộc tính và tiếp theo là thực hiện ước lượng giá trị của tài sản.

2.1.2.2 *Chọn lựa dữ liệu*

Trong bài toán định giá nhà ở, tác giả lựa chọn các yếu tố sau để phân tích bài toán ở các bước tiếp theo: kiểu nhà, hướng nhà, tuổi nhà, diện tích đất xây dựng, số lượng phòng ngủ, số lượng phòng tắm, số chỗ trong gara, diện tích sân nhà.

2.1.2.3 *Tiền xử lý dữ liệu*

2.1.2.4 *Biểu diễn dữ liệu*

Do đa số mạng nơ-ron chỉ xử lý tín hiệu đầu vào, đầu ra dạng số nằm trong khoảng $[-1, 1]$, trong bước này ra thực hiện biểu diễn dữ liệu để xác định các đầu vào, đầu ra cho mạng. Số lượng đầu vào của mạng sau khi thực hiện bước này thường tăng lên so với số lượng thuộc tính đã chọn ở trên. Theo một nghiên cứu, các giá trị đầu vào không nên biến đổi nằm trong khoảng $[0, 1]$, ta sẽ chọn biến đổi các đầu vào nằm trong khoảng $[-1, 1]$.

2.1.2.5 *Chọn kiến trúc mạng*

Sử dụng mạng nơ-ron lan truyền thẳng và thuật toán học lan truyền ngược. Theo cách biểu diễn ở bước trên, ta có tất cả 21 đầu vào và một đầu ra. Dựa trên kinh nghiệm, ta thiết kế mạng với 1 lớp ẩn có 15 nơ-ron sử dụng hàm kích hoạt sigmoid nhị phân cho tín hiệu ra nằm trong khoảng $[0, 1]$.

2.1.2.6 *Huấn luyện và kiểm tra*

2.1.3 *Nhận xét*

Qua việc phân tích bài toán này cho thấy khả năng áp dụng cho nhiều bài toán khác nhau. Tương tự, có thể dễ dàng xây dựng mô hình mạng nơ-ron cho bài toán dự đoán số lượng sản phẩm dựa trên năng lực sản xuất, hoặc tính giá hợp đồng.

Đối với lớp bài toán này, dữ liệu dự đoán thường là dạng số liên tục trên một miền giá trị, đầu ra nên việc biểu diễn dữ liệu cho dạng đầu ra này theo cách biến tỷ lệ vào khoảng $[0, 1]$. Khi mạng

cho kết quả đầu ra, cần thực hiện việc ngược lại quá trình biến đổi tỷ lệ để cho kết quả theo giá trị ban đầu của bài toán.

Đây là dạng bài toán hồi quy, với dạng bài toán này ra sử dụng cách đo độ chính xác mạng sau huấn luyện theo các công thức.

2.2 Bài toán dự đoán chuỗi thời gian

2.2.1 Phát biểu bài toán

Chuỗi thời gian là một chuỗi các điểm dữ liệu, được đo ở các lần kế tiếp nhau tại các khoảng thời gian thống nhất.

Định nghĩa bài toán: cho một tập gồm n mẫu $\{y(t_1), y(t_2), \dots, y(t_n)\}$ trong một chuỗi thời gian $t_1, t_2, \dots, t_n$, bài toán đặt ra là dự đoán giá trị $y(t_{n+1})$ tại thời điểm t_{n+1} . Giá trị dự đoán của một biến vào thời điểm trong tương lai là dựa trên k giá trị trước đó. Trong một số bài toán, ta có thể dự đoán các giá trị $y(t_{n+2}), y(t_{n+3}), \dots$ cho các thời điểm t_{n+2}, t_{n+3} .

2.2.2 Ứng dụng mạng nơ-ron trong bài toán dự đoán tiêu thụ kem

2.2.2.1 Yêu cầu

Trong quản lý kinh doanh, người quản lý cần dự đoán số lượng sẽ bán được, để kiểm soát tối thiểu chi phí bán hàng, có kế hoạch mua hàng và dự trữ hàng tồn kho phù hợp. Một vấn đề cần quan tâm là khi có một sự kiện khuyến mãi, giảm giá hay quảng bá sản phẩm làm tăng số lượng hàng bán ra, tuy nhiên sự tăng lên này cũng có độ trễ nhất định, ví dụ như quảng cáo vào tuần trước thì tuần tiếp theo số lượng hàng bán mới tăng lên. Như vậy ở bài toán này,

ngoài các yếu tố liên quan đến sản phẩm, nhu cầu người tiêu dùng quyết định số lượng hàng bán được, còn ảnh hưởng của yếu tố thời gian, đó là các sự kiện xảy ra trong quá khứ.

2.2.2.2 *Chọn lựa dữ liệu*

Với bài toán này, dữ liệu chúng ta có sẵn là lịch sử số lượng hàng bán được theo hàng ngày. Giả sử trong CSDL bán hàng các trường dữ liệu: ngày bán, tên mặt hàng, giá bán, số lượng, ngày giao hàng, loại hình quảng cáo trong ngày, nhiệt độ.

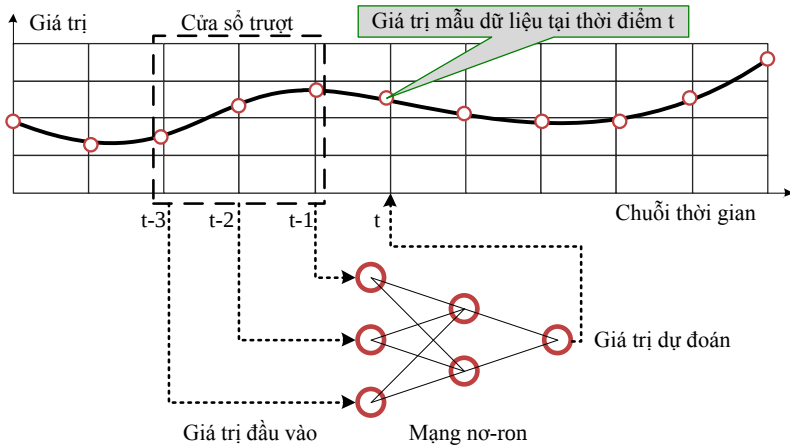
Ta chọn trường dữ liệu ngày bán là biến theo thời gian lấy mẫu. Việc chọn thuộc tính và mẫu dữ liệu hoàn toàn phụ thuộc vào bài toán cần giải. Trong ví dụ này, tác giả lựa chọn các yếu tố sau để phân tích bài toán ở các bước tiếp theo: ngày bán, quảng cáo, nhiệt độ, số lượng hàng bán được.

2.2.2.3 *Tiền xử lý dữ liệu*

2.2.2.4 *Biểu diễn dữ liệu*

2.2.2.5 *Chọn kiến trúc mạng*

Sử dụng mạng nơ-ron lan truyền thẳng và thuật toán học lan truyền ngược. Một dữ liệu đầu vào đưa vào mạng và cho kết quả đầu ra, các trọng số được cập nhật sau khi một mẫu dữ liệu được đưa vào. Xét về mặt tín hiệu, mạng lan truyền thẳng không có “nhớ” mẫu dữ liệu huấn luyện trước. Điều này rõ ràng là không phù hợp cho bài toán dự đoán chuỗi thời gian, khi mà các mẫu dữ liệu tại thời điểm trước dùng để dự đoán cho giá trị tiếp theo. Để giải quyết vấn đề này, thay vì đưa mỗi lần một mẫu dữ liệu vào mạng để học, thì ta đưa $n(n \geq 2)$ mẫu vào cùng một lúc. Kỹ thuật này gọi là kỹ thuật cửa sổ trượt.



Hình 2.2: Dự đoán chuỗi thời gian sử dụng cửa sổ trượt kích thước là 3

Trong bài toán này, ta sẽ dùng thông tin ngữ cảnh và dữ liệu ngày hiện tại cộng với dữ liệu hai ngày trước, cửa sổ trượt có kích thước là 3. Nghĩa là ta cần tất cả 14 đầu vào (5 đầu vào cho ngày thứ nhất, 5 đầu vào cho ngày thứ 2 và 4 đầu vào cho ngày hiện tại) và một đầu ra duy nhất là kết quả dự đoán số lượng cho ngày hiện tại. Sử dụng mạng với một lớp ẩn có 15 nơ-ron.

2.2.2.6 Huấn luyện và kiểm tra

2.2.3 Nhận xét

Mạng lan truyền thẳng không có khả năng nhớ như các mô hình mạng hồi quy, để khắc phục nhược điểm này ta sử dụng kỹ thuật cửa sổ trượt để gửi nhiều mẫu dữ liệu vào cùng một lần. Ngoài ra để tăng khả năng dự đoán, ta phải cung cấp thêm thông tin ngữ cảnh cho mạng. Việc xác định kích thước của cửa sổ trượt cũng là vấn đề khó vì phụ thuộc vào tính chất của bài toán.

Điều kiện kết thúc: cũng như bài toán mô hình hóa ở trên, dùng các công thức đo lường độ chính xác. Công thức thường sử dụng là RMSE.

2.3 Bài toán phân loại

2.3.1 *Phát biểu bài toán*

Định nghĩa bài toán: cho một tập dữ liệu huấn luyện $\{(x_1, y_1), \dots, (x_n, y_n)\}$ để xây dựng một bộ phân loại $h: X \rightarrow Y$ cho phép ánh xạ một đối tượng $x \in X$ đến nhãn phân loại $y \in Y$.

2.3.2 *Ứng dụng mạng nơ-ron trong bài toán đánh giá tình trạng tài chính cá nhân*

2.3.2.1 *Yêu cầu*

Trong lĩnh vực tài chính, cụ thể là hoạt động cho vay của các tổ chức tín dụng, ngân hàng, bảo hiểm; việc đánh giá khả năng tài chính của một khách hàng là một thủ tục của quá trình xét cho vay. Dữ liệu sử dụng được lấy từ CSDL UCI [http://archive.ics.uci.edu/ml/datasets/Statlog+\(German+Credit+Data\)](http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data))

2.3.2.2 *Chọn lựa dữ liệu*

2.3.2.3 *Tiền xử lý dữ liệu*

2.3.2.4 *Biểu diễn dữ liệu*

2.3.2.5 *Chọn kiến trúc mạng*

2.3.2.6 Huấn luyện và kiểm tra

Đối với bài toán phân loại, thì giá trị đầu ra của một nơ-ron sẽ là gần 0 hoặc 1. Thực hiện so sánh giá trị giữa các đầu ra với nhau và chọn giá trị đầu ra lớn nhất để kết luận phân loại.

Điều kiện kết thúc: độ đo mức độ chính xác bài toán phân loại thường căn cứ tỷ lệ nhận dạng đúng trên tập dữ liệu kiểm tra.

2.3.3 Nhận xét

Đối với bài toán phân loại, kiểu dữ liệu đầu ra thường là nhãn phân lớp, được biến đổi theo cách mã hóa nhị phân tạo thành N đầu ra tương ứng với N phân lớp. Kết quả trả về của mạng là dựa trên so sánh các đầu ra và chọn đầu ra có giá trị lớn nhất.

2.4 Tổng kết và so sánh

Bảng 2.13: Bảng so sánh cách ứng dụng mạng nơ-ron cho các lớp bài toán

	Dự đoán theo mô hình hóa	Dự đoán chuỗi thời gian	Phân loại
Phân hoạch dữ liệu	Chọn ngẫu nhiên	Các tập dữ liệu kiểm tra thường chọn từ các mẫu phát sinh mới nhất	Chọn ngẫu nhiên
Kiểu dữ liệu đầu vào	Kiểu ký tự, số, ngày và giờ		
Số lượng đầu vào	Phụ thuộc vào số tham số đầu và cách biểu diễn dữ liệu	Phụ thuộc vào số tham số đầu, cách biểu diễn dữ liệu và cửa sổ trượt	Phụ thuộc vào số tham số đầu và cách biểu diễn dữ liệu

Số lượng đầu ra	Số giá trị cần dự đoán	Số giá trị cần dự đoán, và số thời điểm dự báo trong tương lai($t, t+1, t+2$)	Là số phân lớp cho trước
Kiểu giá trị đầu ra	Thường là kiểu số liên tục	Thường là kiểu số liên tục	Kiểu ký hiệu
Kết quả trả về của mạng	Cần thực hiện biến đổi về miền giá trị ban đầu của biến đầu vào.	Cần thực hiện biến đổi về miền giá trị ban đầu của biến đầu vào.	So sánh các đầu ra để chọn phân lớp
Khởi tạo trọng số	Khởi tạo ngẫu nhiên và có giá trị nhỏ		
Mô hình mạng	Mạng nơ-ron truyền thẳng		
Thuật toán	Thuật toán lan truyền ngược		
Đánh giá kết quả huấn luyện	Dùng các công thức đo độ chính xác	Dùng các công thức đo độ chính xác	Dựa trên tỷ lệ nhận dạng đúng và sai ở tập kiểm tra

CHƯƠNG 3 - XÂY DỰNG ỨNG DỤNG

3.1 Phương pháp ứng dụng mạng nơ-ron

3.1.1 Một số nguyên tắc

3.1.2 Mô hình tổng quan

Phần này sẽ trình bày sơ lược các bước chính khi giải một bài toán bằng mạng nơ-ron: tiền xử lý dữ liệu, chọn lựa biến đầu vào và đầu ra, tổ chức dữ liệu, xây dựng mô hình mạng, đánh giá và sử dụng kết quả.

3.1.3 Chuẩn bị dữ liệu

3.1.3.1 Làm sạch dữ liệu

3.1.3.2 Chọn dữ liệu

Khi đã có dữ liệu, bước tiếp theo là quyết định dữ liệu nào là quan trọng cho bài toán. Quá trình chọn dữ liệu bao gồm hai việc. Đầu tiên là chọn các trường dữ liệu hay tham số nào sẽ là đầu vào của mạng. Thứ hai là chọn các dòng, mẫu dữ liệu.

3.1.3.3 Tiền xử lý

Tiền xử lý dữ liệu là bước tiếp theo sau khi thực hiện làm sạch và chọn lựa dữ liệu. Tiền xử lý có thể là thực hiện lọc dữ liệu, trích chọn các đặc trưng từ dữ liệu và biến đổi vào miền giá trị mà mạng chấp nhận. Việc thực hiện tiền xử lý phụ thuộc rất nhiều vào bản chất dữ liệu và từng bài toán cụ thể.

Bước này đôi khi bao gồm cả việc tạo thêm những trường dữ liệu mới từ một hoặc vài trường dữ liệu ban đầu, đôi khi là thay thế nhiều trường dữ liệu bằng một trường dữ liệu duy nhất.

3.1.3.4 Biểu diễn dữ liệu

Để mạng nơ-ron chúng ta phải chuyển đổi giá trị đầu vào và ra vào khoảng giá trị $[0, 1]$ hoặc $[-1, 1]$. Một nghiên cứu đã chỉ ra rằng, với các dữ liệu đầu vào không nên mã hóa trong khoảng $[0, 1]$.

Mặc dù trong các cơ sở dữ liệu hỗ trợ rất nhiều kiểu dữ liệu, nhưng có thể dễ dàng ánh xạ thành ba kiểu chính: giá trị số liên tục, giá trị số rời rạc, kiểu ký hiệu(nhãn) rời rạc. Riêng kiểu ngày và giờ ta có thể ánh xạ thành giá trị số bằng cách sử dụng hàm thích hợp sẽ trình bày ở phần tiếp theo.

1. Biểu diễn kiểu dữ liệu số
 - a. Biểu diễn kiểu số dạng liên tục.
 - b. Biểu diễn kiểu số dạng số rời rạc
 - Mã hóa dạng 1-of-N.
 - Mã hóa nhị phân.
 - Thermometer
2. Biểu diễn ký hiệu.
3. Biểu diễn kiểu ngày/giờ: nếu thực hiện tiền xử lý kiểu dữ liệu ngày/giờ bằng cách chuyển đổi sang kiểu số, thì chúng ta mất đi thông tin rất quan trọng là tính chu kỳ. Chúng ta cần mã hóa sao cho kiểu ngày giờ giữ được tính chu kỳ.

3.1.3.5 *Tổ chức dữ liệu*

Phân hoạch là quá trình chia dữ liệu thành các tập kiểm định, huấn luyện và kiểm tra. Tập kiểm định được sử dụng để xác định kiến trúc của mạng; tập huấn luyện để luyện mạng cập nhật các trọng số của mạng; tập kiểm tra để kiểm tra hiệu năng của mạng sau khi luyện. Thông thường, việc phân hoạch dữ liệu thực hiện bằng cách chọn ngẫu nhiên theo tỷ lệ huấn luyện, kiểm định và kiểm tra theo tỷ lệ lần lượt 70%, 20% và 10%.

3.1.3.6 *Xác định số lượng dữ liệu*

Ta cần phải ước đoán số lượng dữ liệu cần thiết để có thể sử dụng trong việc xây dựng mạng. Về tổng thể, lượng dữ liệu cần chi phối bởi số các trường hợp cần luyện cho mạng. Bản chất đa chiều của dữ liệu và cách giải quyết mong muốn là các nhân tố chính xác định số các trường hợp cần luyện cho mạng và kéo theo là lượng dữ liệu cần thiết. Số lượng dữ liệu cần để huấn luyện mạng phụ thuộc vào độ phức tạp của bài toán và độ nhiễu dữ liệu có được.

3.1.4 *Thiết kế mạng nơ-ron*

Cần xác định: số lớp ẩn, kích thước của các lớp ẩn, hằng số học, tham số moment, khuôn dạng dữ liệu sẽ đưa vào mạng, dạng hàm kích hoạt, khởi tạo trọng số, tỷ lệ nhiễu mẫu.

3.1.4.1 *Số lớp ẩn*

Vì các mạng có hai lớp ẩn có thể thể hiện các hàm với đáng điệu bất kỳ, nên về mặt lý thuyết không có lý do nào cần sử dụng nhiều hơn hai lớp ẩn. Một số nghiên cứu đã chỉ ra rằng đối với phần lớn các bài toán cụ thể, chỉ cần sử dụng một lớp ẩn cho mạng là đủ.

Việc sử dụng càng nhiều lớp ẩn khiến cho việc huấn luyện trở nên chậm càng lớn.

3.1.4.2 Số nơ-ron trong lớp ẩn

Trong phần lớn các bài toán, không có một cách dễ dàng xác định tối ưu số nơ-ron của một lớp mà không phải luyện mạng sử dụng số các đơn vị trong lớp ẩn khác nhau và dự báo lỗi tổng quát hóa của từng lựa chọn.

3.1.4.3 Lựa chọn hàm kích hoạt

3.1.5 Huấn luyện mạng

3.1.5.1 Khởi tạo trọng số

3.1.5.2 Điều kiện kết thúc huấn luyện

Quá trình huấn luyện dừng khi một trong các điều kiện sau đây xảy ra:

1. Tổng lỗi sau một vòng lặp nhỏ hơn một ngưỡng xác được trước.
2. Tất cả các Δw_{ij} nhỏ hơn hoặc bằng ngưỡng được chỉ định, nghĩa là khi giá trị thay đổi trọng số là nhỏ ta xem như mạng đã hội tụ.
3. Số vòng lặp giới hạn.

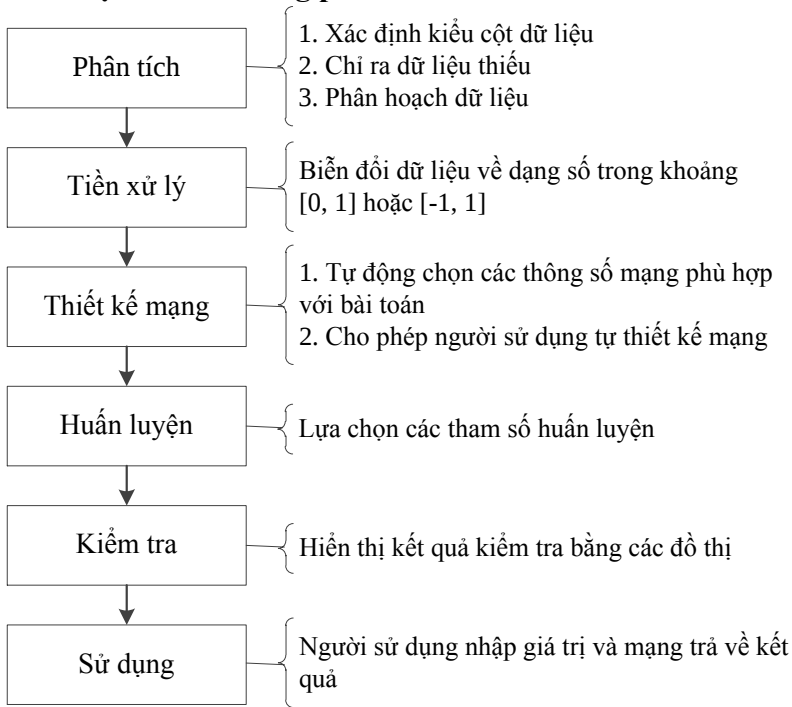
3.1.6 Kiểm tra mạng sau huấn luyện

Để đánh giá mạng có được huấn luyện thành công hay không, cách đánh giá phụ thuộc mô hình mạng và vào bài toán cụ thể. Nếu là bài toán phân loại, thì đánh giá dựa tập dữ liệu kiểm tra,

ta có được tỷ lệ giữa số mẫu phân loại chính xác và so sánh với giá trị ngưỡng sai số cho trước. Nếu là bài toán dự đoán thì đánh giá qua giá trị nhỏ nhất của độ đo độ chính xác.

3.1.7 Hậu xử lý

3.2 Đặc tả chức năng phần mềm



Hình 3.4: Các khối chức năng chính của phần mềm

3.3 Thiết kế phần mềm

Phần mềm được thiết kế theo phương pháp lập trình hướng đối tượng. Phần mềm được lập trình bằng ngôn ngữ C++, sử dụng thư viện giao diện đồ họa wxWidgets.

3.4 Thực nghiệm và đánh giá

Bước 1: Thu thập dữ liệu

Phần mềm cho phép người dùng chọn dữ liệu từ tập tin dạng CSV.

Bước 2: Phân tích

Kết quả ở bước phân tích của phần mềm: phần mềm tự động xác định kiểu dữ liệu cho mỗi cột, tự động phân hoạch các mẫu dữ liệu vào các tập con: huấn luyện, kiểm tra, kiểm định.

Bước 3: Tiền xử lý

Phần mềm giúp chuyển đổi các kiểu dữ liệu về dạng số như với kiểu nhãn phân loại, kiểu ngày, kiểu giờ.

Bước 4: Thiết kế

Phần mềm cho phép người sử dụng thiết kế trực quan mô hình mạng, lựa chọn các tham số mạng như: số lượng lớp ẩn, số nơ-ron trong mỗi lớp ẩn, lựa chọn hàm kích hoạt.

Bước 5: Huấn luyện

Phần mềm cho phép người sử dụng đặt các thông số mạng trước khi thực hiện huấn luyện như hệ số học, số vòng lặp tối đa.

Bước 6: Đánh giá

Phần mềm thực hiện kiểm tra trên tập dữ liệu kiểm tra, biểu diễn mức độ lỗi và biểu đồ giữa giá trị đầu ra của mạng và giá trị đầu ra mong muốn.

Bước 7: Thực thi

Tính năng này cho phép người sử dụng nhập các giá trị đầu vào và phần mềm thực hiện tính toán kết quả trả về.

KẾT LUẬN

Trong đề tài này, tác giả đã thực hiện tìm hiểu tổng quan về mạng nơ-ron, ứng dụng mạng nơ-ron truyền thẳng trong bài toán dự đoán và phân loại. Đề tài tập trung vào việc nghiên cứu xây dựng mô hình tổng quát cho việc ứng dụng mạng nơ-ron cho lớp bài toán dự đoán và phân loại.

1. Kết quả đạt được

Tìm hiểu các khái niệm, các thành phần cơ bản của mạng nơ-ron và các kiểu kiến trúc mạng. Đi sâu phân tích các đặc điểm của mạng nơ-ron, từ đó xác định khả năng ứng dụng của mạng nơ-ron trong bài toán phân loại và dự đoán. Nghiên cứu sâu một kiến trúc mạng kinh điển nhất và được sử dụng rộng rãi là mạng nơ-ron truyền thẳng sử dụng thuật toán lan truyền ngược.

Phân tích ứng dụng mạng nơ-ron truyền thẳng trong lớp bài toán dự đoán và phân loại. Đưa ra được một số nguyên tắc cơ bản cần lưu ý khi giải bài toán bằng mạng nơ-ron. Xây dựng thành một quy trình với các bước thực hiện khá chi tiết để có thể ứng dụng mạng nơ-ron truyền thẳng vào các bài toán khác nhau của các lớp bài toán trên.

Xây dựng phần mềm cung cấp khả năng tổng quát hóa việc sử dụng mạng nơ-ron cho nhiều bài toán khác nhau, cho phép người sử dụng tập hợp dữ liệu mẫu, tiền xử lý đến việc thiết kế mạng và điều chỉnh các thông số cho phù hợp từng bài toán cụ thể. Phần mềm với giao diện trực quan mô phỏng quy trình ứng dụng mạng theo từng bước đã phân tích trong luận văn.

Qua quá trình nghiên cứu lý thuyết và cài đặt phần mềm mô phỏng, tác giả rút ra được một số kết luận như sau:

Mạng nơ-ron với khả năng tổng quát hóa được bài toán, do đó trở thành một công cụ hữu hiệu để có thể giải quyết được nhiều bài toán khác nhau khá hiệu quả. Mạng nơ-ron với nhiều đặc trưng: khả năng học từ dữ liệu, tính thích nghi, chịu lỗi khi dữ liệu không đầy đủ hoặc có nhiễu là những lợi thế của nó so với các phương pháp khác.

Việc ứng dụng mạng nơ-ron sẽ hiệu quả nếu ở các bước đầu tiên như việc lựa chọn mẫu huấn luyện, trích chọn đặc trưng, tiền xử lý được thực hiện tốt. Cần tận dụng các tri thức riêng biệt của bài toán để xây dựng được mô hình mạng thích hợp.

Mạng nơ-ron truyền thẳng và thuật toán lan truyền ngược đã chứng tỏ khả năng rất tốt cho nhiều bài toán phức tạp. Tuy nhiên không có một mô hình chung cho tất cả các bài toán như số lượng nơ-ron và sự hội tụ của mạng. Để có khả năng ứng dụng hiệu quả đòi hỏi phải mất nhiều thời gian để huấn luyện, điều chỉnh các tham số mạng.

2. Một số hướng mở rộng đề tài

Nghiên cứu đầy đủ hơn đặc điểm của từng lớp bài toán để việc ứng dụng mạng nơ-ron trở nên đơn giản và nâng cao hiệu quả. Cải tiến thuật toán lan truyền ngược bằng cách kết hợp các phương pháp khác như thuật giải di truyền, logic mờ. Mở rộng nghiên cứu các kiến trúc mạng và các thuật toán huấn luyện khác, và ứng dụng cho các lớp bài toán khác như phân cụm dữ liệu, phục hồi dữ liệu, ...

Xây dựng các phương pháp biểu diễn dữ liệu cho nhiều kiểu dữ liệu khác nhau như dữ liệu dạng hình ảnh, âm thanh, văn bản,... Phần mềm cần bổ sung nhiều tùy chọn cho phép người sử dụng mô phỏng dễ dàng các mô hình mạng khác nhau.

Trong bối cảnh hiện nay, ứng dụng công nghệ thông tin ngày càng cao và chuyển sang nghiên cứu các hệ thống thông minh, mạng nơ-ron nổi lên như là một giải pháp nhiều hứa hẹn. Mạng nơ-ron thể hiện được nhiều ưu điểm nổi bật so với các phương pháp khác ở khả năng mềm dẻo, linh hoạt và khả năng ứng dụng rộng rãi. Nhưng cũng đòi hỏi sự phức tạp trong việc thiết kế mạng, như không có phương pháp xác định được số lớp ẩn và số nơ-ron bao nhiêu là thích hợp.