

ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

NGUYỄN THỊ THANH HOA

ỨNG DỤNG MẠNG NEURAL TRONG PHÂN LOẠI THƯ RÁC

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kết quả nêu trong luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tác giả luận văn

Nguyễn Thị Thanh Hoa

LỜI CẢM ƠN

Lời đầu tiên em xin gửi lời cảm ơn chân thành đến Các quý thầy cô giáo, Trường Đại học Công nghệ thông tin và Truyền thông đã tận tình giảng dạy, truyền đạt những kiến thức, kinh nghiệm quý báu trong suốt thời gian em theo học tại trường.

Đặc biệt, em xin chân thành cảm ơn thầy giáo **TS. Phạm Thanh Giang** đã rất nhiệt tình và tâm huyết trong việc định hướng và giúp đỡ em hoàn thành luận văn này

Em cũng xin được bày tỏ tình cảm với gia đình, đồng nghiệp, bạn bè đã tạo điều kiện để cá nhân em có thể dành thời gian cho khóa học. Xin chân thành cảm ơn những người bạn lớp cao học CK12B, trong 2 năm qua đã luôn luôn động viên, khích lệ và hỗ trợ em trong quá trình học tập.

Trong quá trình thực hiện Luận văn mặc dù đã cố gắng hết mình, song chắc chắn luận văn của em vẫn còn nhiều thiếu sót. Em rất mong nhận được sự chỉ bảo vào đóng góp tận tình của các thầy cô để luận văn của em được hoàn thiện hơn.

Thái Nguyên, Ngày 15 tháng 4 năm 2015

Tác Giả

Nguyễn Thị Thanh Hoa

ii

MỤC LỤC

| | |
|---|-----------|
| LỜI CAM ĐOAN | i |
| LỜI CẢM ƠN..... | ii |
| MỤC LỤC..... | iii |
| DANH MỤC CÁC BẢNG | iv |
| DANH MỤC CÁC HÌNH | v |
| LỜI MỞ ĐẦU | 1 |
| CHƯƠNG 1: TỔNG QUAN VỀ MẠNG NƠN NHÂN TẠO..... | 3 |
| 1.1. Khái niệm mạng nơon..... | 3 |
| 1.1.1 Giới thiệu | 3 |
| 1.1.1.1. So sánh mạng nơon với máy tính truyền thống..... | 5 |
| 1.1.1.2. Nơon sinh học..... | 6 |
| 1.1.1.3. Chức năng, tổ chức và hoạt động của bộ não con người | 7 |
| 1.1.1.4. Nơon..... | 9 |
| 1.1.2. Mạng Nơon nhân tạo | 10 |
| 1.1.2.1. Nơon nhân tạo..... | 11 |
| 1.1.2.2. Mạng nơon nhân tạo | 15 |
| 1.2. Đặc trưng của mạng nơon | 17 |
| 1.2.1. Tính phi tuyến | 17 |
| 1.2.2. Tính chất tương ứng đầu vào, đầu ra | 17 |
| 1.2.3. Tính chất thích nghi | 17 |
| 1.2.4. Tính chất đưa ra lời giải có bằng chứng | 18 |

iii

| | |
|---|-----------|
| 1.2.5. Tính chất chấp nhận sai sót | 18 |
| 1.2.6. Khả năng cài đặt VLSI (Very-Large-Scale-Intergrated) | 18 |
| 1.2.7. Tính chất đồng dạng trong phân tích và thiết kế..... | 19 |
| 1.3. Phân loại mạng nơron nhân tạo | 19 |
| 1.3.1. Phân loại theo kiểu liên kết nơron..... | 19 |
| 1.3.2. Một số loại mạng nơron điển hình | 20 |
| 1.3.2.1. Mạng dẫn tiến (feedforward) | 20 |
| 1.3.2.2. Mạng quy hồi (recurrent network)..... | 23 |
| 1.4. Xây dựng mạng nơron | 24 |
| 1.5. Huấn luyện mạng nơron..... | 25 |
| 1.5.1. Phương pháp học..... | 25 |
| 1.5.1.1. Học có giám sát | 26 |
| 1.5.1.2. Học không giám sát..... | 26 |
| 1.5.1.3. Học tăng cường | 27 |
| 1.5.2. Thuật toán học | 27 |
| 1.6. Thu thập dữ liệu cho mạng nơron | 28 |
| 1.7. Biểu diễn tri thức cho mạng nơron..... | 30 |
| 1.8. Ứng dụng của mạng nơron. | 33 |
| Kết luận chương 1: | 34 |
| CHƯƠNG 2: BÀI TOÁN PHÂN LOẠI THƯ RÁC | 35 |
| 2.1. Giới thiệu về thư rác | 35 |
| 2.1.1. Khái niệm thư rác | 35 |
| 2.1.2. Các đặc điểm của thư rác | 35 |

| | |
|--|-----------|
| 2.1.3. Phân loại thư rác | 37 |
| 2.1.4. Ảnh hưởng của thư rác | 38 |
| 2.2. Các phương pháp phân loại thư rác | 40 |
| 2.2.1. Phương pháp phân loại theo từ khóa | 40 |
| 2.2.2. Phương pháp phân loại Bayesian | 41 |
| 2.2.3. Phương pháp phân loại SpamAssassin | 46 |
| 2.2.4. Phương pháp phân loại dùng danh sách trắng/đen | 47 |
| 2.2.5. Phương pháp phân loại dựa vào vị trí của bộ lọc | 48 |
| 2.2.6. Phương pháp phân loại dựa trên xác nhận danh tính của người gửi ... | 50 |
| 2.2.7. Phương pháp phân loại thư rác mới dựa trên Mạng xã hội | 53 |
| Kết luận chương 2: | 55 |
| CHƯƠNG 3: ỨNG DỤNG MẠNG NEURAL NGĂN CHẶN SPAM MAIL | |
| 3.1. Xây dựng kiến trúc hệ thống..... | 56 |
| 3.1.1 Phần mềm lọc thư rác POPFile | 56 |
| 3.1.3 Cơ sở dữ liệu POPFile..... | 57 |
| 3.2 Mô hình phân loại thư rác qua Popfile | 61 |
| 3.2.1 Cấu hình kết nối hệ thống mail qua Popfile..... | 61 |
| 3.2.2 Cấu hình Outlook kết nối với Popfile | 62 |
| 3.3 Huấn luyện cho Popfile | 68 |
| 3.4 Kết quả và Đánh giá..... | 68 |
| KẾT LUẬN | 70 |
| TÀI LIỆU THAM KHẢO | 72 |

DANH MỤC CÁC BẢNG

| | |
|---|-----------|
| Bảng 1.1 Một số hàm kích hoạt trong mạng nơron | 14 |
| Bảng 3.1 Kết quả lọc Email trong POPFile sau khi training với 20 Email | 66 |
| Bảng 3.2 Kết quả lọc Email trong POPFile sau khi training với 40 Email | 66 |
| Bảng 3.3 Kết quả lọc Email trong POPFile sau khi training với 80 Email | 66 |

BẢNG DANH MỤC CÁC HÌNH

| TT | TÊN HÌNH | T RANG |
|----|---|--------|
| 1 | Hình 1.1 Mạng nơron sinh học | 6 |
| 2 | Hình 1.2 Sơ đồ nơron sinh học | 9 |
| 3 | Hình 1.3 Mô hình một nơron nhân tạo | 10 |
| 4 | Hình 1.4 Sự tương đương giữa nơron sinh học và nơron nhân tạo [10] | 11 |
| 5 | Hình 1.5 Mô hình phi tuyến thứ hai của một mạng nơron | 13 |
| 6 | Hình 1.6 Sơ đồ đơn giản về một mạng nơron nhân tạo [4] | 15 |
| 7 | Hình 1.7 Mạng truyền thẳng | 19 |
| 8 | Hình 1.8 Mạng quy hồi | 19 |
| 9 | Hình 1.9 Mạng tiến với một mức nơron | 20 |
| 10 | Hình 1.10 Mạng tiến kết nối đầy đủ với một mức ẩn và một mức đầu ra | 21 |
| 11 | Hình 1.11 Mạng hồi quy không có nơron ẩn và không có vòng lặp tự phản hồi | 22 |
| 12 | Hình 1.12 Mạng hồi quy có các nơron ẩn | 22 |
| 13 | Hình 1.13 Sơ đồ đồ thị có hướng đơn giản | 23 |

| | | |
|----|--|-----------|
| 14 | Hình 3.1 Cơ sở dữ liệu Popfile | 51 |
| 15 | Hình 3.2 Mô hình Popfile | 52 |
| 16 | Hình 3.3 Giao diện Popfile | 53 |
| 17 | Hình 3.4 Giao diện chính của phần mềm Microsoft Outlook | 54 |
| 18 | Hình 3.5 Truy cập chức năng vào tài khoản Email | 55 |
| 19 | Hình 3.6 Thiết lập tài khoản Email | 61 |
| 20 | Hình 3.7 Thêm tài khoản Email mới | 62 |
| 21 | Hình 3.8 Sửa tài khoản Email mới | 62 |
| 22 | Hình 3.9 Thay đổi tài khoản Email | 63 |
| 23 | Hình 3.10 Thiết lập thư điện tử | 64 |
| 24 | Hình 3.11 Huấn luyện Email trong POPFile | 65 |