

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

LÃNH LAN HƯƠNG

MẠNG KOHONEN VÀ ỨNG DỤNG TRONG
QUẢN LÝ TRỢ CẤP XÃ HỘI CỦA SINH VIÊN
TRƯỜNG CAO ĐẲNG SƯ PHẠM CAO BẰNG

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2014

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

LÃNH LAN HƯƠNG

**MẠNG KOHONEN VÀ ỨNG DỤNG TRONG
QUẢN LÝ TRỢ CẤP XÃ HỘI CỦA SINH VIÊN
TRƯỜNG CAO ĐẲNG SƯ PHẠM CAO BẰNG**

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

HƯỚNG DẪN KHOA HỌC: PGS. TS LÊ BÁ DŨNG

THÁI NGUYÊN - 2014

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn là công trình nghiên cứu của riêng cá nhân tôi, không sao chép của ai do tôi tự nghiên cứu, đọc, dịch tài liệu, tổng hợp và thực hiện. Nội dung lý thuyết trong trong luận văn tôi có sử dụng một số tài liệu tham khảo như đã trình bày trong phần tài liệu tham khảo. Các số liệu, chương trình phần mềm và những kết quả trong luận văn là trung thực và chưa được công bố trong bất kỳ một công trình nào khác.

Thái Nguyên, tháng 7 năm 2014

Học viên thực hiện

Lãnh Lan Hương

LỜI CẢM ƠN

Lời đầu tiên, em xin gửi lời biết ơn sâu sắc đến PGS.TS Lê Bá Dũng người đã tận tình hướng dẫn, chỉ bảo, giúp đỡ em trong suốt quá trình làm luận văn.

Em cũng xin gửi lời cảm ơn đến các thầy cô giáo trường Đại học Công nghệ thông tin và Truyền thông - Đại học Thái Nguyên, các thầy cô Viện Công nghệ thông tin đã truyền đạt những kiến thức và giúp đỡ em trong suốt quá trình học của mình.

Tôi cũng xin gửi lời cảm ơn tới Ban giám hiệu trường Cao đẳng Sư phạm Cao Bằng đã tạo điều kiện thuận lợi cho tôi tham gia khóa học và trong suốt quá trình hoàn thành luận văn.

Và cuối cùng tôi xin gửi lời cảm ơn tới các đồng nghiệp, gia đình và bạn bè những người đã ủng hộ, động viên tạo mọi điều kiện giúp đỡ để tôi có được kết quả như ngày hôm nay.

Thái Nguyên, tháng 7 năm 2014

Học viên

Lãnh Lan Hương

MỤC LỤC

LỜI CAM ĐOAN.....	i
LỜI CẢM ƠN.....	ii
MỤC LỤC.....	iii
DANH MỤC CÁC KÍ HIỆU, CHỮ CÁI VIẾT TẮT	v
DANH MỤC CÁC BẢNG.....	vi
DANH MỤC CÁC HÌNH.....	vii
MỞ ĐẦU	1
CHƯƠNG 1: TỔNG QUAN VỀ MẠNG KOHONEN VÀ BÀI TOÁN PHÂN CỤM.	2
1.1. Giới thiệu chung về mạng Nơron	2
1.1.1 Lịch sử phát triển của mạng nơron	2
1.1.2. Mạng Nơron sinh học.....	3
1.1.3. Mạng Nơron nhân tạo	5
1.1.4 Ứng dụng.....	14
1.2. Mạng Kohonen và bài toán phân cụm.....	15
1.2.1 Giới thiệu về mạng nơron Kohonen	15
1.2.2 Bài toán phân cụm.....	16
1.3 Các phương pháp phân cụm.....	18
CHƯƠNG 2: SOM VÀ KỸ THUẬT PHÂN CỤM DỮ LIỆU DỰA VÀO SOM...	20
2.1. Thuật toán phân cụm dữ liệu	20
2.2. Một số thuật toán cơ bản trong phân cụm dữ liệu	21
2.2.1 Thuật toán phân cụm phân cấp	21
2.2.2 Thuật toán phân cụm phân hoạch (Thuật toán K-means)	22
2.2.3 Thuật toán phân cụm dựa trên mật độ (Thuật toán DBSCAN)	25
2.2.4 Thuật toán phân cụm dựa trên lưới (Thuật toán STING).....	26
2.2.5 Thuật toán phân cụm dựa trên mô hình (Thuật toán EM).....	27
2.3. Thuật toán phân cụm bằng mạng Kohonen (SOM).....	28
2.3.1 Mạng Kohonen là gì?	28
2.3.2. Cấu trúc mô hình mạng Kohonen	29

2.3.3. Thuật toán phân cụm bằng mạng Kohonen	30
2.4 Một số vấn đề phân cụm với mạng Kohonen	33
2.4.1. Khởi tạo mạng Kohonen (SOM)	33
2.4.2. Huấn luyện mạng Kohonen	34
2.4.3. Tỷ lệ học	35
2.4.4. Cập nhật lại trọng số	36
2.4.5. Xác định neuron chiến thắng	37
2.5. Kohonen (SOM) và công việc phân cụm dữ liệu	38
2.5.1. SOM phân cụm với bản đồ 1 chiều	38
2.5.2. SOM phân cụm với bản đồ 2 chiều	38
2.5.3. Xác định ranh giới các cụm	39
2.5.4. Trực quan mạng	40
2.5.5. Số lượng nhóm khi phân cụm	42
CHƯƠNG 3: ỨNG DỤNG MẠNG SOM TRONG CÔNG VIỆC QUẢN	
LÝ TRỢ CẤP XÃ HỘI CỦA SINH VIÊN TRƯỜNG CAO ĐẲNG SƯ	
PHẠM CAO BẰNG.....	43
3.1. Phát biểu bài toán.....	43
3.2 Khảo sát, đánh giá, thống kê tình hình trợ cấp xã hội của sinh viên trường Cao đẳng Sư phạm Cao Bằng.....	43
3.2.1 Thu thập dữ liệu.....	43
3.2.2. Phân tích dữ liệu.....	45
3.3. Thử nghiệm sử dụng mô hình mạng Kohonen để khảo sát, đánh giá, thống kê tình hình trợ cấp xã hội của học sinh, sinh viên trường Cao đẳng Sư phạm Cao Bằng.....	51
3.3.1. Các chức năng của chương trình	51
3.3.2. Giao diện chương trình	51
3.3.3. Kết quả và phân tích kết quả sau khi huấn luyện mô hình mạng Kohonen. Trực quan mạng U-Matrix.....	53
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	73
TÀI LIỆU THAM KHẢO	74

DANH MỤC CÁC KÍ HIỆU, CHỮ CÁI VIẾT TẮT

HSSV:	Học sinh, sinh viên
CSDL:	Cơ sở dữ liệu
KPDL:	Khai phá dữ liệu
PCDL :	Phân cụm dữ liệu
SOM(Self Organizing Maps):	Mạng nơon tự tổ chức
PE (Processing element):	Phần tử xử lý
BMU(Best - Matching unit):	Đơn vị phù hợp nhất
U-matrix (unified distance matrix):	Ma trận thống nhất khoảng cách
EM (Expectation maximization):	Thuật toán tối đa hóa
STING (STatistical INformation Grid):	Thuật toán thống kê thông tin lưới
DBSCAN (Density Based Spatial Clustering of Applications with Noise):	Phân cụm dữ liệu dựa trên không gian mật độ ứng với nhiễu

DANH MỤC CÁC BẢNG

Bảng 3.1. Dữ liệu thống kê phân cụm theo các lớp trong trường.....	48
Bảng 3.2. Dữ liệu thống kê phân cụm trong 1 lớp.....	50
Bảng 3.3. Kết quả các cụm lớp và tính chất cụm sau khi huấn luyện mạng Kohonen	57
Bảng 3.4. Kết quả các cụm sinh viên và tính chất cụm sau khi huấn luyện mạng Kohonen.....	68

DANH MỤC CÁC HÌNH

Hình 1.1. Cấu tạo mô hình nơron sinh học	4
Hình 1.2. Cấu trúc mô hình một nơron nhân tạo cơ bản.....	5
Hình 1.3. Đồ thị các dạng hàm chuyển	8
Hình 1.4. Mô hình mạng nơron 3 lớp	9
Hình 1.5. Cấu trúc huấn luyện mạng nơron	10
Hình 1.6. Học có giám sát	11
Hình 1.7. Học không có giám sát.....	11
Hình 1.8. Sơ đồ cấu trúc chung của quá trình học	13
Hình 2.1. Các thiết lập để xác định ranh giới các cụm ban đầu	22
Hình 2.2. Tính toán trọng tâm của các cụm mới.....	23
Hình 2.3. Hình dạng các cụm được khám phá bởi thuật toán DBSCAN	25
Hình 2.4. Cấu trúc mạng Kohonen.....	29
Hình 2.5. Phần tử nơron chiến thắng BMU	30
Hình 2.6. Các vùng lân cận	31
Hình 2.9: Hàm tỉ lệ học theo thời gian.....	36
Hình 3.1. Kết quả phân cụm theo các lớp trong trường sử dụng phương pháp trực quan U-Matrix	54
Hình 3.2. Kết quả phân cụm các lớp sử dụng phương pháp trực quan các bản đồ thành phần	59
Hình 3.3 Kết quả phân cụm các nhóm sinh viên trong một lớp sử dụng phương pháp trực quan U-Matrix	66
Hình 3.4. Kết quả phân cụm sinh viên sử dụng phương pháp trực quan các bản đồ thành phần.....	69

MỞ ĐẦU

Thực tế hiện nay tại các trường cao đẳng thuộc khu vực miền núi, đa phần sinh viên là dân tộc thiểu số, có hoàn cảnh khó khăn và được hưởng chính sách trợ cấp xã hội của nhà nước. Hàng năm số lượng sinh viên được tuyển vào các trường ngày càng đông. Công việc quản lý trợ cấp xã hội ở đây chủ yếu thông qua các bảng biểu, thống kê... khó có thể phân tích theo yêu cầu của quản lý công việc một cách nhanh chóng.

Với sự phát triển mạnh mẽ của Công nghệ thông tin và sự ra đời của mạng Nơron đã giúp cho công việc quản lý nói chung và quản lý trợ cấp xã hội cho sinh viên nói riêng có một cái nhìn trực quan. Người quản lý dựa vào đó đưa ra được những đánh giá nhanh chóng, chính xác.

Luận văn với đề tài “**Mạng Kohonen và ứng dụng trong quản lý trợ cấp xã hội của sinh viên trường Cao đẳng Sư phạm Cao Bằng**” khảo sát lĩnh vực KPDL dùng mạng nơron. Luận văn tập trung vào tìm hiểu về mạng nơron Kohonen (SOM) và dùng thuật toán SOM để giải quyết bài toán phân cụm dữ liệu theo mô hình mạng nơron.

Nội dung luận văn gồm có 3 chương:

Chương I. Tổng quan về mạng Kohonen và bài toán phân cụm dữ liệu

Chương II. SOM và kỹ thuật phân cụm dữ liệu dựa vào SOM

Chương III. Ứng dụng mạng Kohonen trong công việc quản lý trợ cấp xã hội của sinh viên trường Cao đẳng Sư phạm Cao Bằng

Do thời gian nghiên cứu và trình độ có hạn, luận văn không tránh khỏi hạn chế và thiếu sót. Em rất mong nhận được sự chỉ bảo, đóng góp ý kiến của các thầy, cô giáo cũng như bạn bè và đồng nghiệp.

Em xin chân thành cảm ơn!

Thái Nguyên, tháng 7 năm 2014

Người viết luận văn

Lãnh Lan Hương