

**ĐẠI HỌC THÁI NGUYÊN**  
**TRƯỜNG ĐẠI HỌC CNTT & TRUYỀN THÔNG**

----- ୨୫୧ -----

**Nguyễn Thế Huy**

**TÌM HIỂU MÔ HÌNH SOM**  
**VÀ ỨNG DỤNG TRONG TƯ VẤN THI ĐẠI HỌC**

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

**Thái Nguyên – 2012**

**ĐẠI HỌC THÁI NGUYÊN**  
**TRƯỜNG ĐẠI HỌC CNTT & TRUYỀN THÔNG**

----- ୨୫ ୧୦ -----

**Nguyễn Thế Huy**

**TÌM HIỂU MÔ HÌNH SOM**  
**VÀ ỨNG DỤNG TRONG TƯ VẤN THI ĐẠI HỌC**

Chuyên ngành: Khoa học máy tính

Mã số: 60.48.01

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

**NGƯỜI HƯỚNG DẪN KHOA HỌC**

**TSKH. Nguyễn Minh Hải**

**Thái Nguyên – 2012**

## LỜI CẢM ƠN

Đầu tiên tôi xin gửi lời cảm ơn chân thành đến thầy TSKH. Nguyễn Minh Hải - Học viện Công nghệ bưu chính viễn thông đã tận tình hướng dẫn, chỉ bảo cho tôi trong suốt quá trình làm luận văn.

Tôi cũng xin gửi lời cảm ơn đến các thầy cô 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 tôi trong suốt quá trình học của mình.

Tôi cũng xin gửi lời cảm ơn tới các đồng nghiệp trong Trung tâm Giáo dục thường xuyên An Dương, gia đình và bạn bè những người đã động viên tạo mọi điều kiện giúp đỡ tôi trong suốt hai năm học.

## **LỜI CAM ĐOAN**

Tôi xin cam đoan toàn bộ nội dung trong luận văn này do tôi tự nghiên cứu, đọc, dịch tài liệu, tổng hợp và thực hiện. Trong luận văn tôi có sử dụng một số tài liệu tham khảo như đã trình bày trong phần tài liệu tham khảo.

Người viết luận văn

Nguyễn Thế Huy

## MỤC LỤC

DANH MỤC CÁC KÍ HIỆU, CHỮ CÁI VIẾT TẮT .....	iv
DANH MỤC CÁC BẢNG.....	v
DANH MỤC CÁC HÌNH VẼ .....	vi
MỞ ĐẦU .....	1
Chương 1: TÌM HIỂU VỀ MẠNG NƠN VÀ MỘT SỐ PHƯƠNG PHÁP PHÂN CỤM.....	3
1.1 Mạng nơon sinh học .....	3
1.1.1 Cấu trúc một nơon sinh học .....	3
1.1.2 Hoạt động của nơon sinh học .....	4
1.2 Mạng nơon nhân tạo.....	4
1.2.1 Cấu trúc và mô hình của một nơon nhân tạo.....	4
1.2.2 Mô hình của mạng nơon nhân tạo .....	7
1.2.3 Mạng nơon một lớp.....	9
1.2.4 Mạng nơon truyền thẳng nhiều lớp .....	10
1.2.5 Mạng Hopfield.....	11
1.3 Các luật học .....	12
1.3.1 Quy tắc học của mạng nơon nhân tạo .....	12
1.3.2 Học có giám sát .....	13
1.3.3 Học không giám sát .....	14
1.3.4 Học tăng cường.....	15
1.4 Một số phương pháp phân cụm dữ liệu.....	15
1.4.1 Phân cụm và các thành phần trong phân cụm dữ liệu .....	15
1.4.2 Phương pháp phân cụm phân cấp.....	16
1.4.3 Phương pháp phân cụm phân hoạch.....	16
1.4.4 Phương pháp phân cụm dựa trên mật độ .....	17
1.4.5 Phân cụm dữ liệu dựa trên lưới .....	17

1.4.6 Phân cụm dữ liệu dựa trên sự ràng buộc .....	17
1.5 Kết luận chương 1 .....	18
Chương 2: PHÂN CỤM DỮ LIỆU SỬ DỤNG MẠNG SOM .....	19
2.1 Thuật toán phân cụm dữ liệu .....	19
2.2 Thuật toán phân cụm tuyến tính không giám sát .....	20
2.2.1 Thuật toán phân cụm K-mean .....	20
2.2.2 Thuật toán phân cụm mờ C-mean .....	22
2.2.3 Thuật toán phân cụm phân cấp .....	24
2.2.4 Thuật toán phân cụm EM (Expectation Maximization) .....	25
2.2.5 Thuật toán phân cụm chất lượng ngưỡng .....	26
2.3 Thuật toán phân cụm phi tuyến tính không giám sát .....	27
2.3.1 Thuật toán phân cụm MST (Minimum spanning tree) .....	27
2.3.2 Thuật toán phân cụm dữ liệu Kernel K-mean. ....	28
2.3.3 Thuật toán phân cụm dựa trên mật độ DBSCAN .....	29
2.4 Mạng nơron Kohonen (SOM) .....	30
2.4.1 Giới thiệu về mạng Kohonen (SOM) .....	30
2.4.2 Cấu trúc của SOM .....	31
2.4.3 Khởi tạo SOM .....	32
2.4.4 Huấn luyện SOM .....	33
2.4.5 Tỷ lệ học .....	34
2.4.6 Hàm lân cận .....	35
2.4.7 Cập nhật trọng số .....	37
2.4.8 Xác định nơron chiến thắng .....	38
2.4.9 Bảo toàn cấu trúc liên kết .....	39
2.5 SOM sử dụng trong phân cụm dữ liệu .....	40
2.5.1 SOM phân cụm với bản đồ một chiều .....	41
2.5.2 SOM phân cụm với bản đồ 2 chiều .....	41

2.5.3 Xác định ranh giới các cụm .....	42
2.5.4 Trực quan mạng .....	43
2.6 Kết luận chương 2 .....	44
<b>Chương 3 : ỨNG DỤNG CỦA MẠNG SOM CHO TRỢ GIÚP HỌC SINH</b>	
<b>THI ĐẠI HỌC.....</b>	<b>46</b>
3.1 Giới thiệu về bài toán trợ giúp học sinh thi đại học .....	46
3.2 Giới thiệu công cụ SOM Toolbox .....	46
3.3 Chương trình thử nghiệm .....	47
3.3.1 Thu thập dữ liệu.....	47
3.3.2 Phân tích tập dữ liệu .....	51
3.3.3 Khởi tạo và huấn luyện SOM .....	51
3.3.4 Kết quả và phân tích dữ liệu sau khi huấn luyện SOM.....	53
3.4 Kết luận chương 3 .....	64
<b>KẾT LUẬN .....</b>	<b>65</b>
<b>TÀI LIỆU THAM KHẢO.....</b>	<b>66</b>

## DANH MỤC CÁC KÍ HIỆU, CHỮ CÁI VIẾT TẮT

SOM (Self Organizing Maps)	Mạng nơron tự tổ chức
ĐH	Đại học
HV	Học viện
PE (Processing element)	Phần tử xử lý
U-matrix (unified distance matrix)	Ma trận thống nhất khoảng cách
EM (Expectation maximization)	Thuật toán tối đa hóa
MST (Minimum spanning tree)	Thuật toán tối thiểu cây mở rộng
BMU (Best – Matching unit)	Đơn vị phù hợp nhất
DBSCAN (Density Based Spatial Clustering of Applications with Noise)	Phân cụm dữ liệu dựa trên không gian mật độ ứng dụng với nhiễu

**DANH MỤC CÁC BẢNG**

Bảng 3.1: Thông tin về một số trường đại học của Việt Nam .....	49
Bảng 3.2: Kết quả các cụm sau khi huấn luyện SOM .....	54

## DANH MỤC CÁC HÌNH VẼ

Hình 1.1: Mô hình nơron sinh học .....	3
Hình 1.2: Mô hình một nơron nhân tạo.....	5
Hình 1.3: Đồ thị các dạng hàm truyền .....	7
Hình 1.4: Mạng nơron ba lớp.....	8
Hình 1.5: Một số dạng mạng nơron .....	10
Hình 1.6: Cấu trúc mạng Hopfield.....	11
Hình 1.7: Học có giám sát.....	14
Hình 1.8: Học không giám sát.....	14
Hình 2.1: Cấu trúc của mạng SOM.....	32
Hình 2.2: Cập nhật BMU và lân cận của nó với mẫu đầu vào $x$ .....	34
Hình 2.3: Hàm tỉ lệ học theo thời gian.....	35
Hình 2.4: Giá trị của hàm lân cận Gaussian(a) và hàm bubble(b) .....	37
Hình 2.5: Bảo toàn cấu trúc liên kết các cụm .....	40
Hình 3.1: Trực quan mạng sử dụng U-matrix.....	54
Hình 3.2: Trực quan các thành phần bản đồ .....	61