ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐH CNTT VÀ TRUYỀN THÔNG

HÀ THỊ THƯ

MẠNG KOHONEN-SOM VÀ ỨNG DỤNG PHÂN CỤM ĐIỂM HỌC SINH THPT

CHUYÊN NGÀNH: KHOA HỌC MÁY TÍNH
MÃ SỐ: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC
HƯỚNG DẪN KHOA HỌC: TS NGUYỄN VĂN TẢO

THÁI NGUYÊN - 2012

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐH CNTT VÀ TRUYỀN THÔNG

HÀ THỊ THƯ

MẠNG KOHONEN-SOM VÀ ỨNG DỤNG PHÂN CỤM ĐIỂM HỌC SINH THPT

CHUYÊN NGÀNH: KHOA HỌC MÁY TÍNH
MÃ SỐ: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC

HƯỚNG DẪN KHOA HỌC
TS NGUYỄN VĂN TẢO

THÁI NGUYÊN - 2012

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn là công trình nghiên cứu của cá nhân tôi, không sao chép của ai. Nội dung lý thuyết trong luận văn có sự tham khảo và sử dụng của một số tài liệu, thông tin được đăng tải trên các tác phẩm, tạp chí và các trang web theo danh mục tài liệu của luận văn. 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, ngày 15 tháng 9 năm 2012

Học viên thực hiện

HÀ THỊ THƯ

MỤC LỤC

DANH MỤC VIẾT TẮT

CSDL	Cơ sở dữ liệu
KPDL	Khai phá dữ liệu
FCM	Fuzzy C-Means
PCDL	Phân cụm dữ liệu

DANH MỤC CÁC HÌNH

MỞ ĐẦU

Sự phát triển mạnh mẽ của Công nghệ nói chung và Công nghệ thông tin nói riêng đã tạo nên nhiều hệ thống thông tin phục vụ việc tự động hoá mọi hoạt động kinh doanh cũng như quản lý trong xã hội. Nhiều hệ quản trị cơ sở dữ liệu mạnh với các công cụ phong phú và thuận tiện đã giúp con người khai thác có hiệu quả các nguồn tài nguyên dữ liệu lớn. Trong đó, khai phá dữ liệu (Data Mining) là quá trình chính trong phát hiện tri thức. Sử dụng các kỹ thuật và các khái niệm của các lĩnh vực đã được nghiên cứu từ trước như học máy, nhận dạng, thống kê, hồi quy, xếp loại, phân nhóm, đồ thị, mạng nơron, mạng Bayes,... được sử dụng để khai phá dữ liệu nhằm phát hiện ra các mẫu mới, tương quan mới, các xu hướng có ý nghĩa.

Luận văn với đề tài **"Mạng Kohonen-SOM và ứng dụng phân cụm điểm học sinh THPT"** khảo sát lĩnh vực KPDL dùng mạng nơron. Luận văn tập trung vào phương pháp học mạng nơron không có giám sát, dùng thuật toán SOM để giải quyết bài toán phân cụm theo mô hình mạng nơron.

Luận văn được thực hiện dưới sự hướng dẫn khoa học của TS. Nguyễn Văn Tảo. Tôi xin chân thành cảm ơn sâu sắc tới Thầy đã chỉ dẫn tận tình giúp tôi có thể hoàn thành bản luận văn này. Tôi cũng vô cùng cảm ơn sự giúp đỡ và động viên khích lệ của người thân trong gia đình tôi, bạn bè và các đồng nghiệp trong trường THPT Ngô Sĩ Liên trong suốt quá trình thực hiện luận văn.

Bắc giang, ngày 15 tháng 9 năm 2012

Hà Thị Thư

CHƯƠNG 1

TỔNG QUAN VỀ MẠNG KOHONEN-SOM

Chương này đề cập các vấn đề sau:

1.1. Sơ lược về mạng neural

1.2. Tổng quan về mạng KOHONEN-SOM

1.3. Cấu trúc của mạng neural KOHONEN

1.4. Thực thi của mạng neural KOHONEN

1.5. Kết luận

1.1. Sơ lược về mạng neural

1.1.1. Lịch sử phát triển

Mạng neural nhân tạo được xây dựng từ những năm 1940 nhằm mô phỏng một số chức năng của bộ não người. Dựa trên quan điểm cho rằng bộ não người là bộ điều khiển. Mạng neural nhân tạo được thiết kế tương tự như neural sinh học sẽ có khả năng giải quyết hàng loạt các bài toán như tính toán tối ưu, điều khiển, công nghệ robot…

Dưới đây là các mốc đáng chú ý trong lịch sử phát triển của mạng neural:

- *Giai đoạn 1*: Có thể tính từ nghiên cứu của William (1890) về tâm lý học với sự liên kết các neural thần kinh. Năm 1940 Mc Culloch và Pitts đã cho biết neural có thể mô hình hoá như thiết bị ngưỡng (Giới hạn) để thực hiện các phép tính logic và mô hình mạng neural của Mc Culloch – Pitts cùng với giải thuật huấn luyện mạng của Hebb ra đời năm 1943.

- *Giai đoạn 2*: vào khoảng gần những năm 1960, một số mô hình neural hoàn thiện hơn đã được đưa ra như: Mô hình Perceptron của Rosenblatt (1958), Adalile của Widrow (1962). Trong đó mô hình Perceptron rất được quan tâm vì nguyên lý đơn giản, nhưng nó cũng có hạn chế vì như Marvin Minsky và Seymour papert của MIT (Massachurehs Insritute of Technology) đã chứng minh nó không dùng được cho các hàm logic phức (1969). Còn Adaline là mô hình tuyến tính, tự