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

TỐNG ĐÌNH TIẾN

TỐI ƯU HÓA CÁC THÔNG SỐ HỆ MỜ SỬ DỤNG PHÂN CỤM DỮ LIỆU TRỪ VÀ GIẢI THUẬT DI TRUYỀN

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN - 2013

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

TỐNG ĐÌNH TIẾN

TỐI ƯU HÓA CÁC THÔNG SỐ HỆ MỜ SỬ DỤNG PHÂN CỤM DỮ LIỆU TRỪ VÀ GIẢI THUẬT DI TRUYỀN

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

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC
P G S . TS LÊ BÁ DŨNG

THÁI NGUYÊN - 2014

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn là kết quả của sự tìm hiểu, nghiên cứu các tài liệu một cách nghiêm túc dưới sự hướng dẫn của PGS. TS Lê Bá Dũng. Nội dung luận văn được phát triển từ ý tưởng, sự sáng tạo của bản thân và kết quả có được hoàn toàn trung thực.

Học viên

Tống Đình Tiến

LỜI CẢM ƠN

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

Em cũng xin được gửi lời biết ơn tới các thầy đã tham gia giảng dạy và chia sẻ những kinh nghiệm quý báu cho tập thể lớp nói chung và cá nhân em nói riêng.

Em cũng xin gửi lời cảm ơn tới Ban giám hiệu trường Đại học Công Nghệ Thông Tin Và Truyền Thông, ban đào tạo sau đại học đã tạo điều kiện thuận lợi cho tôi tham gia khóa học và hoàn thành luận văn.

Cuối cùng tôi xin gửi lời cảm ơn tới gia đình, bạn bè, đồng nghiệp đã luôn ủng hộ, động viên và giúp đỡ để tôi có thể hoàn thành tốt luận văn.

Một lần nữa, em xin chân thành cảm ơn.

Thái Nguyên,15 tháng 3 năm 2014

Học viên

Tống Đình Tiến

MỤC LỤC

DANH MỤC CÁC KÍ HIỆU, CHỮ CÁI VIẾT TẮT

CURE	Clustering Using Representatives
DBSCAN	Density based Spatial Clutering of Application with Noise
STING	STatistical INformation Grid
EM	Expectation Maximization
DENCLUE	Clustering Based on Density Distribution Functions
FCM	Fuzzy C-Means
FSC	Fuzzy Subtractive Clustering
OPTICS	Ordering Points to Identify the Clustering Structure
SC	Subtractive Clustering
GA	Genetic algorithm

DANH MỤC CÁC BẢNG BIỂU

DANH MỤC CÁC HÌNH

MỞ ĐẦU

Trong ngành khoa học máy tính, việc đi tìm kiếm lời giải tối ưu cho các bài toán là vấn đề được các nhà khoa học máy tính đặc biệt rất quan tâm. Mục đích chính của các thuật toán tìm kiếm lời giải là tìm ra lời giải tối ưu nhất cho bài toán trong thời gian nhỏ nhất. Các thuật toán như tìm kiếm không có thông tin, vét cạn (tìm kiếm trên danh sách, trên cây hoặc đồ thị) sử dụng phương pháp đơn giản nhất và trực quan nhất hoặc các thuật toán tìm kiếm có thông tin sử dụng Heurictics để áp dụng các tri thức về cấu trúc của không gian tìm kiếm nhằm giảm thời gian cần thiết cho việc tìm kiếm được sử dụng nhiều nhưng chỉ với không gian tìm kiếm nhỏ và không hiệu quả khi tìm kiếm trong không gian tìm kiếm lớn. Tuy nhiên, trong thực tiễn có rất nhiều bài toán tối ưu với không gian tìm kiếm rất lớn cần phải giải quyết. Vì vậy, việc đòi hỏi thuật giải chất lượng cao và sử dụng kỹ thuật trí tuệ nhân tạo đặc biệt rất cần thiết khi giải quyết các bài toán có không gian tìm kiếm lớn. Giải thuật di truyền (Genetic algorithm) là một trong những kỹ thuật tìm kiếm lời giải tối ưu đã đáp ứng được yêu cầu của nhiều bài toán và ứng dụng.

Thuật giải di truyền đã được phát minh ra để bắt chước quá trình phát triển tự nhiên trong điều kiện quy định sẵn của môi trường. Các đặc điểm của quá trình này đã thu hút sự chú ý của John Holand (ở đại học Michigan) ngay từ những năm 1970. Holand tin rằng sự gắn kết thích hợp trong thuật giải máy tính có thể tạo ra một kỹ thuật giúp giải quyết các vấn đề khó khăn giống như trong tự nhiên đã diễn ra-thông qua quá trình tiến hóa.

Trong thực tế cuộc sống, chúng ta bắt gặp nhiều bài toán như dự đoán thị trường chứng khoán, dự đoán lưu lượng nước, dự đoán lượng ga tiêu thụ, dự đoán năng lực sản xuất, định giá tài sản,…. Đó là các bài toán thuộc lớp bài toán dự đoán và phân lớp, có thể xem là các bài toán cơ bản và có nhiều ứng dụng trong thực tiễn. Đã có nhiều phương pháp được đưa ra để giải quyết các lớp bài toán đó như phương pháp thống kê, cây quyết định, mạng nơron nhân tạo… Việc áp dụng các phương pháp của khai phá dữ liệu (đặc biệt là các phương pháp học máy mạng Nơron kết