ĐẠI HỌC THÁI NGUYÊN

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

NGUYỄN QUANG HUY

NGHIÊN CỨU PHƯƠNG PHÁP NHẬN DẠNG CHỮ VIẾT TAY HẠN CHẾ BẰNG MÔ HÌNH SVM (SUPPORT VECTOR MACHINES)

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

NGUYỄN QUANG HUY

NGHIÊN CỨU PHƯƠNG PHÁP NHẬN DẠNG CHỮ VIẾT TAY HẠN CHẾ BẰNG MÔ HÌNH SVM (SUPPORT VECTOR MACHINES)

Chuyên ngành: KHOA HỌC MÁY TÍNH

Mã số: 60 48 01

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

NGƯỜI HƯỚNG DẪN KHOA HỌC: PGS-TS. NGÔ QUỐC TẠO

Thái Nguyên - 2014

LỜI CÁM ƠN

Để đạt được những kết quả trong quá trình nghiên cứu luận văn, học viên xin chân thành cảm ơn thầy PGS. TS Ngô Quốc Tạo luôn tận tình chỉ bảo, hướng dẫn và giúp đỡ em trong suốt quá trình làm luận văn.

Học viên xin cảm ơn các thầy cô giáo trường Đại học Công nghệ thông tin và Truyền thông đã hướng dẫn và tạo điều kiện cho em trong suốt thời gian học tập tại trường.

Học viên xin chân thành cảm ơn các thầy giáo trong Hội đồng xét duyệt luận văn tốt nghiệp lớp cao học CK11A năm 2014 - Đợt 1 đã nhận xét và góp ý để bài luận văn của em được hoàn thiện hơn.

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn "*Nghiên cứu phương pháp nhận dạng chữ viết tay hạn chế bằng mô hình SVM (Support Vector Machines)"* là do tôi tự nghiên cứu và hoàn thành dưới sự hướng dẫn của PGS-TS. Ngô Quốc Tạo.

Các kết quả đạt được trong quá trình nghiên cứu là hoàn toàn trung thực và khách quan.

Tôi xin chịu trách nhiệm về những lời cam đoan trên.

Thái Nguyên, ngày 05 tháng 05 năm 2014

Người cam đoan

Học viên Nguyễn Quang Huy

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Thuật ngữ, chữ viết tắt	Giải thích
SVM	Support Vector Machine (Máy véc tơ hỗ trợ)
MMH	Maximum Marginal Hyperplane (Siêu phẳng có biên độ lớn nhất)
HMM	Markov Model (Mô hình Markov ẩn)
Kernel	Hàm nhân
MNIST	Bộ mẫu chữ số viết tay NIST - Viện Công nghệ và Tiêu chuẩn Quốc gia Hoa Kỳ (National Institute of Standard and Technology of the United States)
NN	Neuron Network (Mạng nơ ron)
OCR	Optical Character Recognition (nhận dạng chữ quang học)
QP	Quadratic Programing (quy hoạch toàn phương)
USPS	United States Postal service
VC	Vapnik – Chervonenkis

DANH MỤC CÁC HÌNH

MỞ ĐẦU

Biết sử dụng các phương pháp nhận dạng đóng vai trò hết sức quan trọng trong xử lý ảnh, phân tích tài liệu văn bản, đặc biệt là đối với các dạng văn bản viết tay. Hiện nay, nhu cầu cần nhận dạng nội dung văn bản từ các ảnh là rất lớn và thiết thực. Để nâng cao độ tin cậy của các phương pháp phân tích nhận dạng đã có những công trình nghiên cứu theo hướng ứng dụng lớp bài toán đánh giá lựa chọn thông tin để lựa chọn những tổ hợp thông tin chất lượng cao trước khi tiến hành phân tích nhận dạng. Cũng từ đó đề xuất những cách tiếp cận mới giải quyết bài toán nhận dạng trong xử lý số liệu văn bản và thu được kết quả tốt.

Nhận dạng chữ viết và đặc biệt nhận dạng chữ viết tay là bài toán có nhiều ứng dụng thực tế. Máy tính xử lý, nhận dạng các biểu mẫu, phiếu điều tra tự động, bằng cách này ta có thể tiết kiệm được nhiều chi phí về thời gian, công sức cũng như các chi phí khác cho việc nhập dữ liệu.

Ngày nay cùng với sự phát triển về mặt lý thuyết, công nghệ, có rất nhiều hướng đi cho việc giải quyết bài toán nhận dạng chữ viết dựa trên cấu trúc hay cách tiếp cận khác như dùng: logic mờ, giải thuật di truyền, mô hình xác suất thống kê, mô hình Markov ẩn HMM (Hidden Markov Models), mô hình mạng nơron NN (Neural Network Model), mô hình SVM (Support Vector Machine).

Thuật toán phân lớp là yếu tố có vai trò quyết định đến chất lượng của một hệ thống nhận dạng. Các phương pháp nhận dạng truyền thống như đối sánh mẫu, nhận dạng cấu trúc đã được ứng dụng khá phổ biến trong các hệ thống nhận dạng và cũng đã thu được những thành công nhất định. Tuy vậy, với những trường hợp văn bản đầu vào có chất lượng không tốt (nhiễu, đứt nét, dính nét...) thì các thuật toán này tỏ ra không hiệu quả. Để khắc phục điều này, trong những năm gần đây nhiều nhóm nghiên cứu đã sử dụng các thuật toán phân lớp dựa trên mô hình SVM cho các bài toán nhận dạng nói chung và nhận dạng chữ viết

tay nói riêng. Trong luận văn này, học viên xin trình bày thuật toán SVM đối với việc nhận dạng chữ viết tay hạn chế.

MỤC LỤC