

Proceedings of the 13<sup>th</sup> National Conference on  
Fundamental and Applied Information  
Technology Research (FAIR'2020)



NHA TRANG,  
08-09/10/2020

NGHIÊN CỨU CƠ BẢN VÀ ỨNG DỤNG  
CÔNG NGHỆ THÔNG TIN

NHÀ XUẤT BẢN KHOA HỌC  
TỰ NHIÊN VÀ CÔNG NGHỆ



LIÊN HIỆP CÁC HỘI KHOA HỌC VÀ KỸ THUẬT VIỆT NAM

# FAIR

KỶ YẾU HỘI NGHỊ KHOA HỌC CÔNG NGHỆ  
QUỐC GIA LẦN THỨ XIII

## NGHIÊN CỨU CƠ BẢN VÀ ỨNG DỤNG CÔNG NGHỆ THÔNG TIN

NHA TRANG, 08-09/10/2020

Proceedings of the 13<sup>th</sup> National Conference on  
Fundamental and Applied Information Technology Research  
(FAIR'2020)



NHÀ XUẤT BẢN KHOA HỌC TỰ NHIÊN VÀ CÔNG NGHỆ

## MỤC LỤC

1	71	A NEW METHOD OF RNA SECONDARY STRUCTURE PREDICTION BASED ON GENETICS ALGORITHMS, MACHINE LEARNING <i>Duy Bình Đoàn, Minh Tuấn Phạm, Đức Long Đặng</i>	1
2	83	AN APPROACH OF BLE MESH NETWORK FOR SMART HOME APPLICATION <i>Minh Sơn Nguyễn</i>	11
3	59	ÂN TẬP MỤC HỮU ÍCH CAO VÀ PHỔ BIẾN NHẠY CẢM <i>Huyền Triệu Vỹ, Lê Quốc Hải, Trương Ngọc Châu, Lê Quốc Hiếu</i>	16
4	118	CÁC TIÊU CHÍ NGÔN NGỮ TRONG VIỆC XÂY DỰNG KHO NGỮ LIỆU TIẾNG VIỆT <i>Võ Diệp Như, Đinh Điền</i>	26
5	147	CẢI TIẾN ĐỘ CHÍNH XÁC TRA CỨU ẢNH THÔNG QUA HỌC SÂU VÀ HỌC ĐỘ ĐO KHOẢNG CÁCH TỐI ƯU <i>Quỳnh Dao Thi Thủy, Quỳnh Nguyễn Hữu, Tao Ngo Quốc</i>	33
6	127	CHỮ KÝ SỐ TẬP THỂ CHO CÁC NHÓM KÝ <i>Tuan Nguyễn Kim, Duy Hồ Ngọc, Nin Hồ Lê Việt, Long Nguyễn Văn</i>	41
7	69	CHUYỂN TỰ TỰ ĐỘNG TỪ CHỮ NÔM SANG CHỮ QUỐC NGỮ THEO TIẾP CẬN DỊCH MÁY NEURAL <i>Đinh Điền, Trang Minh Chiến, Nguyễn Thị Kim Phượng, Nguyễn Hồng Bửu Long</i>	48
8	185	CLOUD NETWORK MANAGEMENT MODEL BASED ON MOBILE AGENT <i>Nguyễn Minh Phúc, Nguyễn Ái Việt, Trần Quý Nam</i>	56
9	196	CONVERSION MODEL FOR OPEN METEOROLOGICAL DATA <i>Tien Thi Tran, Ngoan Thanh Trieu, Phuoc Minh Ngo</i>	64
10	85	ĐÁNH GIÁ CÁC THUẬT TOÁN RÚT TRÍCH PHỤ HUỘC HÀM: TANE, FUN, FD_MINE VÀ PHỤ THUỘC HÀM PHẢ HỆ TRI THỨC <i>Trịnh Minh Quốc Trung, Nguyễn Đức Thuận</i>	72
11	148	ĐÁNH GIÁ HIỆU NĂNG GIẢI PHÁP CHỐNG TẤN CÔNG LỖ ĐEN TRONG MẠNG CẢM BIẾN KHÔNG DÂY CÓ MẬT ĐỘ NÚT CAO <i>Nguyễn Quốc Cường, Nguyễn Đức Thắng, Trần Thị Bích Phương, Võ Thanh Tú</i>	80
12	129	ĐÁNH GIÁ HIỆU QUẢ CỦA MỘT SỐ MẠNG NƠ RON NHÂN TẠO TRONG PHÁT HIỆN TẤN CÔNG TRÊN TẬP DỮ LIỆU UNSW-NB15 <i>Lê Thị Trang Linh</i>	88
13	161	ĐỀ XUẤT KHUNG KIẾN TRÚC ỨNG DỤNG CHO CHÍNH PHỦ DI ĐỘNG DỰA TRÊN KIẾN TRÚC TỔNG THỂ TẠI VIỆT NAM <i>Phạm Hải Sơn, Lê Hoàng Sơn, Byeong Nam Yoon</i>	93
14	175	DỰ ĐOÁN THỊ TRƯỜNG CỔ PHIẾU BẰNG PHƯƠNG PHÁP HỌC SÂU KHÔNG GIÁM SÁT GENERATIVE ADVERSARIAL NETWORK <i>Bui Thanh Hưng, Nguyễn Minh Lợi</i>	101
15	157	FORECASTING SWEET POTATOES PRICES IN BINH TAN DISTRICT, VINH LONG PROVINCE BY ARTIFICIAL INTELLIGENCE <i>Le Duy Dong, Nguyen Thi Thy Lieu, Vu Thanh Nguyen, Nguyen Thanh Nghiem, Le Kim Nga</i>	107
16	168	GÁN NHÃN NGỮ NGHĨA CHO TỪ ĐIỂN TIẾNG VIỆT DỰA TRÊN WORDNET TIẾNG VIỆT	114

		<i>Yen Le, Thu Tran Thi Anh, Dinh Dien</i>	
17	101	GIẢI PHÁP GIẤU TIN VỚI KHẢ NĂNG NHÚNG CAO DỰA TRÊN KỸ THUẬT ĐỒ CẠNH LOG VÀ SUDOKU <i>Duong Ngoc Van Khanh, Huynh Van Thanh, Vo Thanh C, Son Nguyen Thai</i>	120
18	189	GIẢI PHƯƠNG TRÌNH ĐẠO HÀM RIÊNG DẠNG ELLIPTIC VỚI ĐIỀU KIỆN BIÊN DIRICHLET SỬ DỤNG MẠNG NEURAL DỰA TRÊN THUẬT TOÁN HUẤN LUYỆN MÁY HỌC CỰC ĐỘ <i>Quán Hồ Đắc, Hiếu Huỳnh Trung</i>	128
19	11	GIẤU TIN THUẬN NGHỊCH BẰNG DỊCH CHUYỂN LƯU ĐỒ VÀ CÂY TỨ PHÂN <i>Trần Hoàng Nam, Võ Phước Hưng, Nguyễn Thái Sơn, Huỳnh Văn Thanh</i>	135
20	15	GIẤU TIN THUẬN NGHỊCH TRONG ẢNH JPEG DỰA TRÊN MỞ RỘNG SỰ KHÁC BIỆT CỦA HỆ SỐ QDCT <i>Huỳnh Văn Thanh, Nguyễn Thái Sơn, Võ Phước Hưng, Trần Hoàng Nam, Đỗ Thanh Nghi</i>	143
21	190	HỆ HỖ TRỢ CHẨN ĐOÁN LOÃNG XƯƠNG TRÊN ẢNH X-QUANG CỔ XƯƠNG ĐŨI DỰA TRÊN CHỈ SỐ SINGH <i>Dinh Toan Tran, Thị Quỳnh Như Nguyễn, Minh Hung Le, Anh Thắng Bùi</i>	149
22	156	HỆ THỐNG HỖ TRỢ HỎI ĐÁP THỦ TỤC HÀNH CHÍNH <i>Thanh-Dien Nguyen, The-Phi Pham, Cong-Xuyen Pham, Thanh-Nghi Do</i>	158
23	173	HỆ THỐNG NHẬN DẠNG TÔM <i>Thanh-Tri Trang, The-Phi Pham, Thanh-Nghi Do</i>	165
24	22	HỆ THỐNG TRẢ LỜI TỰ ĐỘNG TƯ VẤN TUYỂN SINH SAU ĐẠI HỌC <i>Trần Thanh Trâm, Trần Thanh Phước, Nguyễn Thị Anh Thư, Nguyễn Thế Hữu, Văn Thế Thành</i>	172
25	32	HUẤN LUYỆN MÔ HÌNH TÓM TẮT TỰ ĐỘNG VĂN BẢN TIẾNG VIỆT TỪ TẬP DỮ LIỆU LỚN <i>Nguyen Ti Hon, Nguyen Thi Ngoc Han, Pham The Phi, Do Thanh Nghi</i>	180
26	178	IVA-EC: GIẢI PHÁP PHÂN TÍCH VIDEO THÔNG MINH DỰA TRÊN ĐIỆN TOÁN BIẾN <i>Đào Vũ Hiệp, Mạc Đình Hiếu, Tống Việt Hùng, Vũ Văn Phán, Hà Quốc Trung</i>	188
27	49	KẾT HỢP ĐẶC TRƯNG DEEP LEARNING TRONG HỖ TRỢ CHẨN ĐOÁN TỰ ĐỘNG BỆNH LÝ VỀ MẮT <i>Phạm Hoàng Oanh, Lê Minh Hưng, Trần Văn Lãng, Dinh Toan Tran</i>	197
28	30	KẾT HỢP ĐẶC TRƯNG SÂU TRONG HỖ TRỢ CHẨN ĐOÁN UNG THƯ VÚ TRÊN NHỮ ẢO ẢO X-QUANG <i>Nguyễn Chí Thanh, Võ Thị Huyền Trang, Lê Minh Hưng, Hoang Le Uyen Thuc</i>	206
29	106	KHẢO SÁT BÀI TOÁN NHẬN DIỆN VÀ ĐO TỐC ĐỘ PHƯƠNG TIỆN THAM GIA GIA THÔNG <i>Trần Hoàng Lộc, Nguyễn Khắc Ngọc Khôi, Phan Đình Duy, Vũ Đức Lung</i>	215
30	169	MAPPING OF RELATIONAL, OBJECT DATABASE TO NOSQL <i>Nguyễn Đình Thuận, Văn Thị Phương Lâm</i>	224
31	1	MÔ HÌNH HÓA HÀNH ĐỘNG CỦA NGƯỜI LÍNH ỨNG DỤNG TRONG MÔ PHÒNG QUÂN SỰ <i>Nguyễn Thị Lan, Nguyễn Trung Kiên, Lê Xuân Thủy</i>	230

32	80	MỘT CÁI TIẾN MỚI CỦA THUẬT TOÁN PHÂN CỤM BÁN GIÁM SÁT MỜ VỚI NHIỀU THAM SỐ MỜ VÀ ỨNG DỤNG TRONG PHÂN ĐOẠN ẢNH NHA KHOA <i>Tuấn Trần Mạnh, Huân Phùng Thế, Hùng Nguyễn Bắc, Sơn Lê Hoàng, Khang Trần Đình, Nga Tu Thu</i>	238
33	13	MỘT ĐỀ XUẤT CÁI TIẾN THUẬT TOÁN JOIN ĐỆ QUI TRÊN TẬP DỮ LIỆU LỚN <i>Trieu Thanh Ngoan, Phan Anh Cang, Phan Thuong Cang</i>	245
34	163	MỘT GIẢI PHÁP NÂNG CAO HIỆU QUẢ GIẢI MÃ CỦA HỆ MẬT ĐA TRỊ VÀ NHẬP NHẮNG MAS <i>Long Thị Lệ, Nguyễn Đình Hân</i>	254
35	90	MỘT LƯỢC ĐỒ ĐỊNH TUYẾN HIỆU QUẢ NĂNG LƯỢNG TRONG MẠNG CẢM BIẾN KHÔNG DÂY ĐA BƯỚC <i>Nguyễn Thi, Phạm Như Việt Anh, Trần Như Đức, Hoàng Trọng Minh</i>	260
36	91	MỘT MÔ HÌNH SUY DIỄN MỜ PHỨC KHÔNG GIAN MỚI CHO BÀI TOÁN PHÁT HIỆN BIẾN ĐỔI CỦA ẢNH VIỄN THĂM <i>Hương Triệu Thu, Giang Nguyễn Long, Sơn Lê Hoàng, Lan Lương Thị Hồng, Tuấn Trần Mạnh, Giang Lê Trung</i>	267
37	115	MỘT MÔ HÌNH TÌM KIẾM ẢNH THEO NGỮ NGHĨA DỰA TRÊN THUẬT TOÁN k-NN VÀ ĐẶC TRƯNG TỬ TỪ <i>Nguyễn Hải Yến, Nguyễn Thị Định, Nguyễn Văn Thịnh, Văn Thế Thành, Lê Mạnh Thạnh</i>	276
38	84	MỘT PHƯƠNG PHÁP GIẢM TẢI TỐI ƯU NĂNG LƯỢNG TRONG HỆ THỐNG ĐIỆN TOÁN BIÊN DI ĐỘNG <i>Hoàng Trọng Minh, Nguyễn Quốc Cường, Dương Thị Lan, Hoàng Thị Thu</i>	288
39	120	MỘT PHƯƠNG PHÁP KẾT HỢP K-MEANS VÀ k-NN CHO BÀI TOÁN TÌM KIẾM VÀ PHÂN TÍCH NGỮ NGHĨA HÌNH ẢNH <i>Huỳnh Thị Châu Lan, Nguyễn Thị Uyên Nhi, Văn Thế Thành, Lê Mạnh Thạnh</i>	295
40	123	MỘT PHƯƠNG PHÁP PHÂN HOẠCH DỮ LIỆU TRÊN CÂY PHÂN CẤP CHO BÀI TOÁN TÌM KIẾM ẢNH THEO NỘI DUNG <i>Nguyễn Thị Thùy Trang, Nguyễn Minh Hải, Nguyễn Văn Thịnh, Văn Thế Thành</i>	306
41	78	MỘT PHƯƠNG PHÁP PHÁT HIỆN ĐƯỜNG BIÊN ẢNH Y TẾ HIỆU QUẢ SỬ DỤNG SUY LUẬN MỜ <i>Hoàng Trọng Minh, Bùi Văn Hậu, Đinh Thị Thu Ngọc, Vũ Như Lân</i>	318
42	172	MỘT SỐ HƯỚNG PHÁT TRIỂN MỚI VỀ HỆ TƯ VẤN MỜ <i>Minh Nguyễn Văn, Giang Nguyễn Long, Sơn Nguyễn Như, Giáp Cù Nguyên, Thông Nguyễn Thọ, Chuân Phạm Minh, Sơn Lê Hoàng</i>	326
43	64	MỘT SỐ HƯỚNG TIẾP CẬN MỚI TRONG PHÂN CỤM BÁN GIÁM SÁT MỜ <i>Huân Phùng Thế, Thái Vũ Đức, Tuấn Trần Mạnh, Thông Phạm Huy, Sơn Lê Hoàng</i>	334
44	177	MỘT SỐ TÍNH CHẤT CỦA MA TRẬN SUP TRÊN KHỐI DỮ LIỆU KHI BỔ SUNG VÀ LOẠI BỎ LỚP ĐỐI TƯỢNG THUẦN NHẤT <i>Trịnh Đình Thắng, Đỗ Thị Lan Anh, Trần Minh Tuyền</i>	342
45	112	MỘT THUẬT TOÁN THUYẾT VẤN SỐ CHO ẢNH Y KHOA BẰNG CÁCH KẾT HỢP DWT-DCT-SVD VÀ PHÉP BIẾN ĐỔI ARNOLD <i>Võ Thành C, Võ Phước Hưng, Dương Ngọc Vân Khanh, Nguyễn Thái Sơn</i>	350
46	136	MOTORBIKE E-MANAGEMENT SYSTEM <i>Tung Nguyen Thanh Xuan, Phuong Vo Thi Luu, Thuan Nguyen Phan Hung</i>	359
47	87	NGĂN CHẶN THÔNG TIN SAI LỆCH NHIỀU CHỦ ĐỀ TRÊN MẠNG XÃ HỘI	369

	TRỰC TUYẾN	
		<i>Dũng Phạm Văn, Trinh Nguyễn Thị Tuyết, Anh Nguyễn Việt</i>
48	81	NGHIÊN CỨU ĐỀ XUẤT CẢI TIẾN THUẬT TOÁN LẬP LỊCH CHO ĐƯỜNG TÀI XUỐNG TRONG MẠNG LONG TERM EVOLUTION VỚI CÁC NODE DI ĐỘNG KẾT HỢP GIỮA MIỀN THỜI GIAN VÀ MIỀN TẦN SỐ <i>Lê Minh Tuấn, Đặng Thanh Hải, Vũ Như Lâm, Lê Hoàng Sơn, Đinh Thu Khánh</i>
49	199	NGHIÊN CỨU KỸ THUẬT HỌC SÂU ỨNG DỤNG VÀO NHẬN DẠNG KÝ TỰ VIẾT TAY CHỮ HOA TIẾNG VIỆT <i>Trần Văn Út Chính, Trần Cao Đệ</i>
50	96	NGHIÊN CỨU VÀ HIỆN THỰC HỆ THỐNG ĐỊNH VỊ TRONG TÒA NHÀ CAO TẦNG SỬ DỤNG CÔNG NGHỆ LORA <i>Ho Thi Mong Trinh Huynh, Le-Huy Trinh, Manh Thao Nguyen</i>
51	158	NGHIÊN CỨU, ĐỀ XUẤT KHUNG KIẾN TRÚC HỆ THỐNG THÔNG TIN TỔNG THỂ CHO CÁC TRƯỜNG ĐẠI HỌC CÔNG LẬP <i>Đào Anh Phương, Lê Quang Minh</i>
52	103	NHẬN DẠNG HÀNH ĐỘNG HÀNG NGÀY CỦA CON NGƯỜI DỰA VÀO CẢM BIẾN GIA TỐC TRÊN ĐIỆN THOẠI THÔNG MINH <i>Lê Hồng Lam, Nguyễn Đức Nhân, Nguyễn Hà Nam, Nguyễn Hoài Nam, Cao Ngọc Ánh, Trần Thủy Bình</i>
53	2	NHẬN DẠNG HÌNH ẢNH THỰC PHẨM BẰNG PHƯƠNG PHÁP DEEP LEARNING <i>Phan Anh Cang, Nguyễn Thanh Hoàng, Trần Hồ Đạt, Nguyễn Văn Hiếu, Phan Thượng Cang</i>
54	193	ONE-MINIUM-ONLY BASIC-SET TRELIS MIN-MAX ALGORITHM FOR NONBINARY LDPC DECODER <i>Phạm Thị Huyền, Đào Tuấn Hùng, Lưu Đức Anh, Nguyễn Anh Tuấn, Phạm Xuân Nghĩa</i>
55	183	PHÂN CỤM BÁN GIÁM SÁT TRONG MẠNG NƠ RON TỐI THIỂU MỜ BẰNG CÁCH GIEO HẠT <i>Vũ Đình Minh, Lê Bá Dũng, Lê Anh Tú, Nguyễn Thanh Sơn</i>
56	39	PHÂN LOẠI HẠCH UNG THƯ PHỐI BẰNG MÔ HÌNH 3D-CNN <i>Trần Văn Quang, Lê Minh Hưng, Dinh Toan Tran, Mau Long Le</i>
57	132	PHÂN LỚP DỮ LIỆU DỰA VÀO PHƯƠNG PHÁP LỰA CHỌN ĐẶC TRƯNG SỬ DỤNG PHỤ THUỘC HÀM XẤP XỈ <i>Phan Anh Phong, Lê Văn Thành, Nguyễn Hải Yến</i>
58	176	PHÂN TÍCH Ý KIẾN NGƯỜI DÙNG THEO KHÍA CẠNH BẰNG PHƯƠNG PHÁP HỌC SÂU KẾT HỢP CNN-LSTM <i>Bui Thanh Hung, Nguyen Quoc Binh</i>
59	16	PHÁT HIỆN BẰNG TRONG TÀI LIỆU DẠNG ẢNH SỬ DỤNG PHƯƠNG PHÁP ĐỊNH VỊ GÓC CORNERNET <i>Dương Phi Long, Võ Duy Nguyên, Nguyễn Thanh Tường Vi, Võ Trung Hiếu, Nguyễn Tân Trần Minh Khang</i>
60	155	PHÁT HIỆN TẤN CÔNG WEB THƯỜNG GẶP DỰA TRÊN HỌC MÁY SỬ DỤNG WEB LOG <i>Hoàng Xuân Dậu, Nguyễn Trọng Hưng</i>
61	151	PHÁT HIỆN VÀ ĐỊNH LƯỢNG TỒN THƯƠNG DO XUẤT HUYẾT NÃO SỬ DỤNG CHỈ SỐ HOUNSFIELD VÀ MẠNG NƠON R-FCN

		<i>Phan Anh Cang, Trần Phan An Trường, Võ Văn Quyền, Phan Thượng Cang</i>	
62	133	PHƯƠNG PHÁP GỢI Ý VIỆC LÀM DỰA TRÊN SỰ TƯƠNG ĐỒNG KỸ NĂNG GIỮA ỨNG VIÊN VÀ CÔNG VIỆC <i>Nguyễn Trần Minh Thu, Trần Thị Trúc Hân, Trần Thị Hồng Nhung</i>	483
63	166	PHƯƠNG PHÁP PHÂN LOẠI NHANH MÃ ĐỘC SỬ DỤNG MẠNG NƠ-RON TÍCH CHẬP <i>Mạc Đình Hiếu, Lê Ngọc Anh, Trần Đức Hùng, Trịnh Văn Hùng, Ngô Minh Phước, Hà Quốc Trung</i>	493
64	139	PHƯƠNG PHÁP SỐ GIẢI BÀI TOÁN BIÊN DIRICHLET CHO PHƯƠNG TRÌNH TAM ĐIỀU HOÀ PHI TUYẾN <i>Đặng Quang Á, Trương Hà Hải, Nguyễn Quốc Hưng, Vũ Vinh Quang</i>	500
65	45	PRIVACY PRESERVING SMS SPAM DETECTION WITH DEEP LEARNING MODELS IN DISTRIBUTED ENVIRONMENT <i>Trần Anh Tú, Huỳnh Văn Nam, Lương Thế Dũng, Đặng Việt Hưng</i>	508
66	77	RÚT GỌN THUỘC TÍNH TRONG BẢNG QUYẾT ĐỊNH THEO TIẾP CẬN TẬP THÔ MỜ TRỰC CẢM <i>Đại Trần Thanh, Giang Nguyễn Long, Châu Hoàng Thị Minh, Ngân Trần Thị</i>	516
67	94	SỬ DỤNG MÔ HÌNH HỌC SÂU TRONG HỖ TRỢ XÁC ĐỊNH VỊ TRÍ CÁC NỐT PHỐI TRÊN ẢNH CẮT LỚP VI TÍNH <i>Giang Sơn Tran, Tom Herbreteau, Chi Cuong Nguyen, Thi Phuong Nghiem, Cuong Do Oanh, Huy Duc Nguyen</i>	524
68	26	SỬ DỤNG MÔ HÌNH LDA-NWF CHO VIỆC TỰ ĐỘNG DÒ TÌM BÁO CÁO LỖI TRÙNG NHAU <i>Nhan Minh Phúc, Nguyễn Thừa Phát Tài, Nguyễn Hoàng Duy Thiện</i>	531
69	188	TÁI TẠO VÀ PHÂN MẢNH CÁC BỘ PHẬN TRÊN MÔ HÌNH 3D TỪ TẬP ẢNH <i>Lê Tiến Mẫu, Nguyễn Tấn Khôi, Romain Raffin</i>	540
70	35	TĂNG CƯỜNG DỮ LIỆU ẢNH CHO BỘ DỮ LIỆU X-QUANG PHỐI SỬ DỤNG GENERATIVE ADVERSARIAL NETWORK <i>Trần Hoàng Phát, Phạm Mạnh Cường, Lê Minh Hưng, Dinh Toan Tran, Tung Hoang</i>	549
71	58	THE IMPROVEMENTS OF SEMANTIC-BASED IMAGE RETRIEVAL USING HIERARCHICAL CLUSTERING TREE <i>Nguyễn Minh Hải, Văn Thế Thành, Trần Văn Lãng</i>	557
72	97	THIẾT KẾ VÀ TỔ CHỨC GIAO THỨC TRUYỀN THÔNG TRONG MẠNG CẢM BIẾN KHÔNG DÂY LORA <i>Nguyễn Văn Thiện, Quách Đức Cường, Trịnh Trọng Chương, Nguyễn Bá Nghiễn</i>	570
73	109	THU THẬP, XỬ LÝ VÀ PHÂN TÍCH DỮ LIỆU MẠNG XÃ HỘI TWITTER BẰNG PHƯƠNG PHÁP NHÚNG THÔNG TIN NÚT <i>Phan Đăng Khoa, Vũ Đức Thi</i>	577
74	141	THUẬT TOÁN DCA CHO MỘT MÔ HÌNH TỐI ƯU DANH MỤC ĐẦU TƯ VỚI RÀNG BUỘC VỀ LỰC LƯỢNG <i>Đức Quỳnh Trần, Thi Lua Nguyen</i>	587
75	194	THUẬT TOÁN HIỆU QUẢ KHAI THÁC MẪU PHỔ BIẾN TRÊN CƠ SỞ DỮ LIỆU ĐỊNH LƯỢNG TĂNG TRƯỞNG <i>Ham Nguyen Duy, Bac Vuong Dinh</i>	595
76	192	TIẾP CẬN HEURISTIC CHO KHAI THÁC TẬP PHỔ BIẾN CÓ CHIỀU DÀI TỐI ĐA	603

		TRÊN DỮ LIỆU GIAO DỊCH CÓ TRỌNG SỐ CỦA ITEM <i>Phan Thành Huân, Lê Hoài Bắc</i>	
77	117	TÌM KIẾM ẢNH THEO NGỮ NGHĨA DỰA TRÊN PHƯƠNG PHÁP GOM CỤM VÀ ONTOLOGY <i>Phan Thị Ngọc Mai, Lê Thị Vĩnh Thanh, Văn Thế Thành, Lê Mạnh Thanh</i>	612
78	79	TÓM TẮT VĂN BẢN BÁO MẠNG ĐIỆN TỬ TIẾNG VIỆT SỬ DỤNG TEXTRANK/VIETNAMESE ONLINE NEWSPAPER SUMMARIZATION BASED ONTEXT RANK <i>Le Thang, Le Quang Minh, Phạm Bảo Sơn</i>	623
79	165	ỨNG DỤNG CHATBOT THÔNG MINH TRONG VIỆC XÁC THỰC HỢP ĐỒNG THÔNG MINH ETHEREUM <i>Tam Bang, Quang Tran, Tho Quan</i>	628
80	130	ỨNG DỤNG GAN TRONG XÂY DỰNG DỮ LIỆU HÌNH ẢNH Y TẾ <i>Doan Thien Minh, Van Dinh Vy Phuong, Phan Manh Thuong, Phạm Công Xuyên</i>	635
81	153	ỨNG DỤNG KỸ THUẬT NHẬN DẠNG TIẾNG NÓI TỰ ĐỘNG VÀO HỖ TRỢ HOẠT ĐỘNG KIỂM SOÁT KHÔNG LƯU <i>Lê Văn Vũ, Hoàng Lê Uyên Thục</i>	641
82	18	ỨNG DỤNG MẠNG PHỨC HỢP TRONG KHAI PHÁ DỮ LIỆU TƯƠNG TÁC NGƯỜI DÙNG <i>Nguyễn Minh Tân, Trần Tiến Dũng</i>	649
83	182	ỨNG DỤNG MÔ HÌNH U-NET PHÁT HIỆN VÙNG BẤT THƯỜNG TRÊN ẢNH MRI NÃO <i>Le Minh Loi, Tran Nguyen Minh Thu, Ho Trong Nguyen, Pham Nguyen Khang</i>	656
84	125	ỨNG DỤNG PHƯƠNG PHÁP GATED RECURRENT UNIT DỰ BÁO NGẮN HẠN DÒNG CHẢY ĐẾN SÔNG VÀM CỎ ĐÔNG <i>Thanh Tung Nguyen, Van Quan Ngo, Thi Thu Huong Nguyen</i>	663
85	21	ỨNG DỤNG WEKA ĐỂ ĐÁNH GIÁ HIỆU SUẤT CÁC PHÂN LOẠI MÁY HỌC CHO BÀI TOÁN ĐỊNH VỊ XE <i>Văn Khánh Hưng, Huỳnh Khả Tú, Nguyễn Quang Phú, Nguyễn Văn Sinh, Lý Tú Nga</i>	671
86	72	ƯỚC LƯỢNG SỐ NGƯỜI TRONG ĐÁM ĐÔNG SỬ DỤNG MẠNG NƠ-RON TÍCH CHẬP <i>Đỗ Phúc Thịnh, Quách Thị Bích Nương, Trần Văn Ninh</i>	677
87	154	URBANEGOAI: A SYSTEM TO MEASURE AIR QUALIFICATION INDEX USING EGOCENTRIC, URBAN SENSING DATA <i>Nguyen Minh Tam, Nguyen Tai Tan Loc, Nguyen Dang Hieu</i>	683
88	146	VỀ BÀI TOÁN XÂY DỰNG BẢNG QUYẾT ĐỊNH NHẤT QUÁN TRONG LÝ THUYẾT TẬP THỒ <i>Vu Duc Thi</i>	690
89	191	VỀ VẤN ĐỀ THỂ HIỆN TẬP PHỤ THUỘC HÀM CỦA KHỐI DỮ LIỆU TRONG MÔ HÌNH DỮ LIỆU DẠNG KHỐI <i>Trịnh Đình Thắng, Trần Minh Tuyền</i>	696
90	50	XÂY DỰNG BỘ NGỮ LIỆU ĐỒNG THAM CHIẾU CHO TIẾNG VIỆT <i>Lê Công Cảnh, Tiêu Vĩnh Phong, Lương An Vinh, Huỳnh Quang Đức</i>	703
91	145	XÂY DỰNG GIẢI PHÁP AN NINH TOÀN DIỆN TRÊN MẠNG IOT VỚI PHƯƠNG THỨC CẢI TIẾN GIAO THỨC DTLS TÍCH HỢP CƠ CHẾ OVERHEARING	710

		<i>Tánh Nguyễn, Ngô Quang Trí, Nguyễn Anh Tuấn, Nguyễn Linh Giang, Nguyễn Ngọc Cương</i>	
92	105	XÂY DỰNG HỆ ĐO XUNG NGẮN TỐC ĐỘ CAO - ỨNG DỤNG KHẢO SÁT MẬT ĐỘ BẦY HẠT TÀI TẠI LỚP TIẾP GIÁP ĐIỆN MÔI - BÁN DẪN CỦA LINH KIỆN MOSFET KÊNH N <i>Le Truong Thinh, Tran Quang Nguyen</i>	718
93	116	XÂY DỰNG HỆ MẬT MÃ ĐƯỜNG CONG ELLIPTIC VỚI KHÓA ĐỐI XỨNG AFFINE ĐỂ MÃ HÓA GIẢI MÃ VĂN BẢN TIẾNG VIỆT <i>Mai Mạnh Trùng, Đỗ Trung Tuấn, Lê Phê Đô, Lê Trung Thực, Đào Thị Phương Anh</i>	724
94	142	XÂY DỰNG HỆ THỐNG THÔNG TIN HỖ TRỢ BỆNH NHÂN TIỂU ĐƯỜNG DỰA TRÊN ONTOLOGY <i>Vỹ Phan Ngọc, Thúy Phạm Thị Thu</i>	733
95	48	XỬ LÝ DỮ LIỆU KHÔNG CÂN BẰNG TRONG BÀI TOÁN DỰ ĐOÁN LỖI PHẦN MỀM <i>Song Toan Le, Thi My Hanh Le, Thanh Binh Nguyen</i>	746
96	34	XỬ LÝ MẬT CÂN BẰNG DỮ LIỆU TRONG PHÂN LOẠI TỔN THƯƠNG DA TRÊN ẢNH SOI DA <i>Võ Minh Thiện, Lê Minh Hưng, Trần Văn Lăng, Kim Tâm Trần</i>	756



## PHÂN LỚP DỮ LIỆU DỰA VÀO PHƯƠNG PHÁP LỰA CHỌN ĐẶC TRƯNG SỬ DỤNG PHỤ THUỘC HÀM XẤP XÌ

Phan Anh Phong, Lê Văn Thành, Nguyễn Hải Yến

Viện Kỹ thuật và Công nghệ, Trường Đại học Vinh

phongpa@gmail.com, thanh.cntt.dhv@gmail.com, nguyenhaiyen1632@gmail.com

**TÓM TẮT:** Lựa chọn đặc trưng là kỹ thuật chọn ra một tập con các đặc trưng phù hợp, liên quan từ tập dữ liệu gốc bằng cách loại bỏ các đặc trưng nhiễu, dư thừa không liên quan nhằm tăng hiệu năng cũng như giảm chi phí trong quá trình khai phá dữ liệu hay học máy. Bài báo này nghiên cứu phương pháp phân lớp dữ liệu dựa vào kỹ thuật lựa chọn đặc trưng với phụ thuộc hàm xấp xỉ và độ đo lỗi  $g_3$ . Một số thử nghiệm phân lớp trên các tập dữ liệu thực tế cho thấy sự phù hợp của hướng nghiên cứu.

**Từ khóa:** Phân lớp dữ liệu, lựa chọn đặc trưng, phụ thuộc hàm xấp xỉ.

### I. GIỚI THIỆU

Lựa chọn đặc trưng là một trong những vấn đề quan trọng trong lĩnh vực khai phá dữ liệu và học máy. Mục đích chính của lựa chọn đặc trưng là tìm ra các đặc trưng hữu ích để mô hình hóa hệ thống, theo đó làm tăng hiệu năng và giảm thời gian thực hiện cho hệ thống. Về bản chất lựa chọn đặc trưng là quá trình tính toán mức độ quan trọng của từng đặc trưng hoặc từng nhóm đặc trưng và sau đó chọn tập con hữu ích nhất trong không gian đặc trưng đó để xây dựng hệ thống [2, 5, 6].

Phân lớp dữ liệu là một bài toán tiêu biểu của khai phá dữ liệu, phần lớn dữ liệu trong bài toán phân lớp khi thu thập được đều có số đặc trưng (thuộc tính) rất nhiều, có thể lên tới hàng chục, hàng trăm, thậm chí là hàng nghìn đặc trưng, chẳng hạn như các bộ dữ liệu trong y tế, trong sinh học,... Ngoài ra, các đặc trưng này có thể có những đặc trưng dư thừa và ít hữu ích cho khai phá dữ liệu và học máy. Việc xây dựng mô hình phân lớp từ dữ liệu như vậy có thể dẫn đến hiệu năng phân lớp thấp cả về tốc độ và khả năng dự báo vì trong dữ liệu có những đặc trưng nhiễu, đặc trưng ít liên quan. Vì vậy, đã có nhiều nghiên cứu cố gắng giải quyết vấn đề này bằng cách sử dụng các kỹ thuật lựa chọn đặc trưng. Hiện nay có 3 cách tiếp cận chính để lựa chọn đặc trưng [6, 7]:

- Mô hình lọc (Filter). Đây là cách lựa chọn đặc trưng sử dụng tính toán trọng số (độ quan trọng), có thể là mối quan hệ giữa các thuộc tính và lớp, sau đó chọn các đặc trưng có trọng số cao hơn một ngưỡng cụ thể. Các thuật toán trong danh mục này bao gồm CfsSubsetEval, độ lợi thông tin và Chi-Square,....

- Mô hình đóng gói (Wrapper). Mô hình này tìm kiếm tập con các đặc trưng tốt bằng cách đánh giá chất lượng của các tập đặc trưng. Việc đánh giá chất lượng thường sử dụng độ chính xác phân lớp của các thuật toán học.

- Mô hình nhúng (Embedded). Mô hình nhúng là sự kết hợp ưu điểm của mô hình lọc và đóng gói bằng cách sử dụng đồng thời tiêu chí đánh giá độc lập và các thuật toán học để đánh giá tập con các đặc trưng, theo đó giúp cải tiến hiệu năng phân lớp.

Khái niệm phụ thuộc hàm được đưa ra bởi Codd có vai trò quan trọng trong lý thuyết cơ sở dữ liệu quan hệ. Các phụ thuộc hàm rất hữu ích trong việc phân tích và thiết kế cơ sở dữ liệu quan hệ. Tuy nhiên, trong thực tế, do có một số giá trị dữ liệu không chính xác hoặc một số ngoại lệ nào đó làm cho các phụ thuộc hàm không thỏa. Sự phụ thuộc tuyệt đối này dường như quá nghiêm ngặt khi ta hình dung tới một quan hệ có hàng nghìn bộ, trong khi đó, chỉ có khoảng vài chục bộ vi phạm phụ thuộc hàm. Điều này làm mất tính chất phụ thuộc vốn có giữa các thuộc tính (đặc trưng) của dữ liệu. Vì vậy, có nhiều nghiên cứu đã mở rộng khái niệm phụ thuộc hàm thành phụ thuộc hàm xấp xỉ, các phụ thuộc này cho phép có một số lượng lỗi nhất định của các bộ dữ liệu đối với phụ thuộc hàm. Các phụ thuộc hàm xấp xỉ không những giúp ta thấy được mối quan hệ tiềm ẩn giữa các thuộc tính mà còn giúp ta thuận tiện hơn trong việc phân tích dữ liệu và đánh giá thông tin [1, 4].

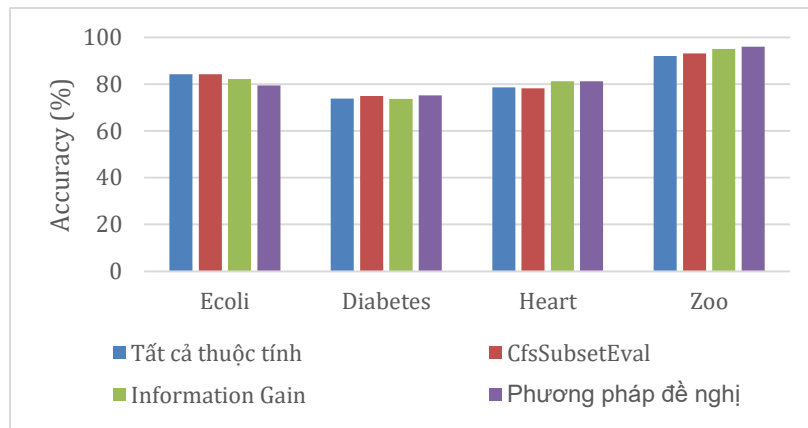
Xuất phát từ cách nhìn đó, bài báo này đề xuất một phương pháp phân lớp dữ liệu dựa vào phụ thuộc hàm xấp xỉ và độ đo lỗi  $g_3$ . Cách tiếp cận sử dụng phụ thuộc hàm và phụ thuộc hàm xấp xỉ để lựa chọn đặc trưng đã được Uncu và Turken nghiên cứu [3]. Tuy nhiên, trong nghiên cứu đó, các tác giả đã đề xuất một thuật toán dựa vào phụ thuộc hàm truyền thống và thuật toán K-láng giềng gần nhất để xác định các biến vào quan trọng cho các hệ thống với thuộc tính vào có miền giá trị liên tục. Ngoài ra để cho thuật toán lựa chọn đặc trưng có khả năng đáp ứng với nhiễu, các phụ thuộc hàm xấp xỉ được sử dụng kết hợp với giá trị độ thuộc để đối phó với tính không chắc chắn trong dữ liệu. Điểm khác biệt của bài báo này với nghiên cứu của Uncu và Turken là sử dụng phụ thuộc hàm xấp xỉ với độ đo lỗi  $g_3$  để lựa chọn đặc trưng cho các bộ phân lớp. Hơn nữa các thử nghiệm được thực hiện trên các bộ dữ liệu thực tế của UCI có độ tin cậy cao trong đánh giá hiệu năng của các phương pháp phân lớp [10].

Bài báo được tổ chức như sau, tiếp theo phần mở đầu, phần II trình bày ngắn gọn các kiến thức liên quan về phụ thuộc hàm, phụ thuộc hàm xấp xỉ và thuật toán khai phá phụ thuộc hàm xấp xỉ mức 1. Phương pháp đề nghị được mô

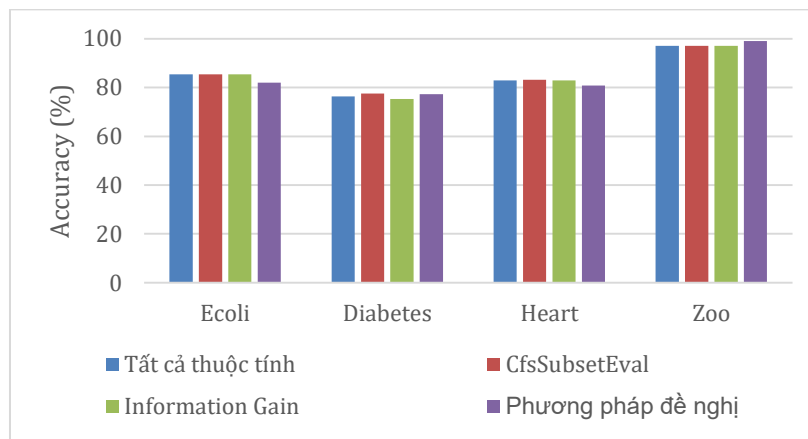
Bảng 8. Hiệu năng của thuật toán phân lớp Naïve Bayes với các phương pháp lựa chọn đặc trưng

Bộ dữ liệu	Tất cả thuộc tính	CfsSubsetEval	Information Gain	Phương pháp đề nghị
Ecoli	<b>85,4167 %</b>	<b>85,4167 % (6)</b>	<b>85,4167 % (4)</b>	81,9429 % (4)
Diabetes	76,3021%	<b>77,48% (4)</b>	75,2604 % (5)	<b>77,2078 % (5)</b>
Heart	82,8383 %	<b>83,1683 % (7)</b>	82,8383 % (6)	80,7634 % (5)
Zoo	97,0297 %	97,0297 % (9)	97,0297 % (9)	<b>99,00 % (9)</b>

Qua thử nghiệm cho thấy phương pháp phân lớp dựa trên lựa chọn thuộc tính sử dụng phụ thuộc hàm xấp xỉ khá phù hợp với các bộ dữ liệu có nhiều thuộc tính rời rạc. Đặc điểm của bộ dữ liệu Ecoli có nhiều thuộc tính liên tục nên phương pháp đề nghị cho hiệu năng phân lớp kém hơn hai phương pháp lựa chọn đặc trưng kia. Điểm yếu này là dễ hiểu vì khái niệm phụ thuộc hàm xấp xỉ đang được định nghĩa dựa trên sự so sánh bằng nhau tuyệt đối giữa các giá trị.



Hình 1. Kết quả về tính chính xác của thuật toán phân lớp cây quyết định với các kỹ thuật lựa chọn đặc trưng



Hình 2. Kết quả về tính chính xác của thuật toán phân lớp Naïve Bayes với các kỹ thuật lựa chọn đặc trưng

### V. KẾT LUẬN

Lựa chọn đặc trưng dựa vào phụ thuộc hàm xấp xỉ là một cách tiếp cận rất tự nhiên để giữ lại những đặc trưng liên quan, theo đó loại bỏ các đặc trưng dư thừa và ít liên quan trong dữ liệu. Kết quả của việc lựa chọn đặc trưng theo phương pháp này là một tập con các đặc trưng từ tập đặc trưng ban đầu nhưng vẫn đảm bảo các tính chất của dữ liệu gốc. Qua kết quả thử nghiệm với 4 bộ dữ liệu tiêu biểu từ UCI đã chứng tỏ được khả năng ứng dụng của phương pháp đề nghị với các bộ dữ liệu có nhiều thuộc tính rời rạc. Với các bộ dữ liệu có nhiều thuộc tính liên tục phương pháp đề xuất có hiệu năng phân lớp chưa tốt, từ đây đặt ra vấn đề cần giải quyết tiếp theo, đó là nghiên cứu phụ thuộc hàm xấp xỉ mở rộng với các độ đo lỗi khác và thử nghiệm với các bộ dữ liệu có số chiều cao.

### TÀI LIỆU THAM KHẢO

[1] J, Atoum, "Approximate Functional Dependencies Mining Using Association Rules Specificity Interestingness Measure", British Journal of Mathematics & Computer Science, 15 (5), 1-10, 2016.

[2] J, Li, K, Cheng, S, Wang, F, Morstatter, R, P, Trevino, J, Tang, & H, Liu, "Feature Selection: A Data Perspective", ACM Computing Surveys, Vol. 50, No. 6, pp. 1-73, 2016,

- [3] O, Uncu, I,B, Turksen, “Two step feature selection: approximate functional dependency approach using membership values”, In Proceeding(s) of FUZZ-IEEE Conference, pp. 1643 - 1648, 2004.
- [4] V, Matos, B, Grasser, “SQL-based Discovery of Exact and Approximate Functional Dependencies”, SIGCSE Bulletin, Vol. 36, No. 4, pp. 58-63, 2004.
- [5] I, Feddaoui, F, Felhi, J, Akaichi, “EXTRACT: new extraction algorithm of association rules from frequent itemsets”, In Proceeding(s) of the IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp. 752-756, 2016.
- [6] O, Villacampa, “Feature Selection and Classification Methods for Decision Making: A Comparative Analysis”, PhD, Thesis, Nova Southeastern University, 2015.
- [7] I, Guyon, and A, Elisseeff, “An Introduction to Feature Extraction, Feature Extraction”, Foundations and Applications, 207(10), 740, 2006.
- [8] C, Giannella, E, Robertson, “On an Information Theoretic Approximation Measure for Functional Dependencies”, Computer Science Department, Indiana University, Bloomington, 2000.
- [9] J, Atoum, “Mining approximate FDs from databases based on minimal cover and equivalent classes”, European Journal of Scientific Research, 2009.
- [10] <https://archive.ics.uci.edu/ml/index.php>.

## **DATA CLASSIFICATION BASED ON FEATURE SELECTION WITH THE APPROXIMATE FUNCTIONAL DEPENDENCE**

**Phan Anh Phong, Le Van Thanh, Nguyen Hai Yen**

***ABSTRACT:** Feature selection is a technique of selecting a subset of relevant and related features from the original dataset by eliminating noise and redundant features in order to increase performance as well as reduce internal costs in the data mining or machine learning process. This paper proposed a new data classification method based on approximate functional dependence and error measure  $g_3$ . The classification experiments on the actual datasets indicate the appropriate of the research direction.*