

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

VŨ ĐỨC ANH

**LỰA CHỌN TAG SNP DỰA VÀO PHƯƠNG PHÁP
TỐI ƯU ĐÀN KIẾN**

LUẬN VĂN THẠC SĨ KHOA HỌC

Thái Nguyên - Năm 2016

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

VŨ ĐỨC ANH

**LỰA CHỌN TAG SNP DỰA VÀO PHƯƠNG PHÁP
TỐI ƯU ĐÀN KIẾN**

Chuyên ngành: Khoa học máy tính

Mã số: 60.48.0101

LUẬN VĂN THẠC SĨ KHOA HỌC

NGƯỜI HƯỚNG DẪN KHOA HỌC

TS. ĐỖ ĐỨC ĐÔNG

Thái Nguyên - Năm 2016

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi, dưới sự chỉ dẫn của TS. Đỗ Đức Đông. Các số liệu, kết quả nêu trong luận văn là trung thực, bảo đảm tính khách quan, luận văn này cho đến nay chưa được bảo vệ tại bất kỳ hội đồng nào và chưa hề được công bố trên bất kỳ phương tiện nào khác. Các tài liệu tham khảo có nguồn gốc xuất xứ rõ ràng.

Tác giả xin chịu trách nhiệm về những lời cam đoan trên.

Thái Nguyên, ngày 20 tháng 8 năm 2016

Tác giả luận văn

Vũ Đức Anh

LỜI CẢM ƠN

Em xin chân thành cảm ơn thầy giáo TS. Đỗ Đức Đông đã trực tiếp giao cho em đề tài, tận tình hướng dẫn và tạo mọi điều kiện cho em hoàn thành luận văn.

Em xin chân thành cảm ơn các thầy cô giáo, các cán bộ nhân viên phòng đào tạo, ban lãnh đạo Trường Đại học Công nghệ thông tin và Truyền thông đã giúp đỡ tạo điều kiện cho em hoàn thành bản luận văn này.

Cuối cùng, em xin chân thành cảm ơn sự quan tâm giúp đỡ của gia đình, bạn bè và tập thể lớp Cao học K13H đã cổ vũ động viên em hoàn thành tốt luận văn của mình.

Thái Nguyên, ngày tháng năm 2016

Học viên Vũ Đức Anh

MỤC LỤC

.....	1
LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC	iii
DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT	v
DANH MỤC CÁC BẢNG	vii
DANH MỤC CÁC HÌNH	viii
MỞ ĐẦU	1
CHƯƠNG 1.	3
TỔNG QUAN VỀ TỐI ƯU ĐÀN KIẾN VÀ BÀI TOÁN LỰA CHỌN TAG SNP	3
1.1. Tìm hiểu về SNP	3
1.1.1. SNP (Single Nucleotide Polymorphisms)	3
1.1.2. Phương pháp xác định SNP.....	5
1.1.3. Tính chất của SNP.....	7
1.1.4. Ứng dụng và triển vọng của nghiên cứu SNP	7
1.2. Bài toán lựa chọn Tag SNPs và các cách tiếp cận hiện nay	11
1.3. Tổng quan về tối ưu đàn kiến	15
1.4. Mục tiêu nghiên cứu của luận văn	21
1.5. Bố cục của luận văn	21
CHƯƠNG 2.	23
MỘT SỐ VẤN ĐỀ TRONG LỰA CHỌN TAG SNP BẰNG PHƯƠNG PHÁP TỐI ƯU ĐÀN KIẾN	23
2.1. Tìm hiểu về bài toán tối ưu tổ hợp tổng quát	23
2.2. Tối ưu đàn kiến	25
2.2.1. Từ kiến tự nhiên đến kiến nhân tạo	25
a/ Kiến tự nhiên	26
b/ Kiến nhân tạo	28
2.2.2. Phương pháp ACO cho bài toán TUTH tổng quát.....	30
a/ Đồ thị cấu trúc	30
b/ Mô tả thuật toán ACO tổng quát	32
2.2.3. Đánh giá ảnh hưởng của các tham số trong thuật toán ACO	35
a/ Thông tin heuristic	35
b/ Số lượng kiến	36

c/ Tham số bay hơi	36
2.3. Bài toán lựa chọn tag SNPs	37
2.4. Phương pháp giải bài toán lựa chọn SNPs bằng thuật toán tối ưu hóa đàn kiến- Thuật toán MACA.....	38
2.4.1. Thuật toán đàn kiến.....	38
2.4.2. Kiến quyết định và cập nhật mùi.....	39
2.4.3. Hiệu chỉnh quy tắc cập nhật mùi – áp dụng quy tắc SMMAS	40
2.4.4. Heuristic	41
2.4.5. Thuật giải MACA	42
CHƯƠNG 3.	45
CHƯƠNG TRÌNH THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ	45
3.1. Mô tả thực nghiệm	45
3.2. Kết quả thực nghiệm và đánh giá.....	46
KẾT LUẬN VÀ HƯỚNG NGHIÊN CỨU TIẾP THEO	50
TÀI LIỆU THAM KHẢO	51
PHỤ LỤC.....	55
A. Mã lệnh cho thuật toán ACA	55
B. Mã lệnh khi sử dụng quy tắc SMASS để cập nhật mùi:	57
C. Mã lệnh cho thuật toán MACA.....	57

DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Kí hiệu và chữ viết tắt	Ý nghĩa
τ_{max}	Cận trên của vết mùi
τ_{min}	Cận dưới của vết mùi
τ_{mid}	Cận giữa của vết mùi
τ_0	Vết mùi được khởi tạo ban đầu
τ_{ij}	Vết mùi trên cạnh
τ_i	Vết mùi trên đỉnh
η_{ij}	Thông tin heuristic trên cạnh
η_j	Thông tin heuristic trên đỉnh
N_c	Số vòng lặp trong thuật toán ACO
N_a	Số kiến sử dụng trong thuật toán ACO
ρ	Tham số bay hơi
3-LAS	Three-Level Ant System (Hệ kiến ba mức)
ACO	Ant Colony Optimization (Tối ưu đàn kiến)
ACS	Ant Colony System (Hệ đàn kiến)
AS	Ant System (Hệ kiến)
G-best	Global-best (Lời giải tốt nhất tính đến thời điểm hiện tại)
I-best	Iteration-best (Lời giải tốt nhất trong bước lặp hiện tại)

MLAS	Multi-level Ant System (Hệ kiến đa mức)
MMAS	Max-Min Ant System (Hệ kiến Max Min)
SMMAS	Smoothed Max-Min Ant System (Hệ kiến Max Min tron)
TSP	Bài toán người du lịch

DANH MỤC CÁC BẢNG

Bảng	Trang
Bảng 2.1. Biểu diễn nhị phân của haplotype và SNP.....	44
Bảng 3.1. Bảng kết quả thực nghiệm khi số lượng haplotype cố định và số lượng SNP thay đổi.....	46
Bảng 3.2. Thời gian chạy thực nghiệm khi số lượng haplotype cố định và số lượng SNP thay đổi.....	47
Bảng 3.3. Bảng kết quả khi số lượng SNP cố định và số lượng haplotype thay đổi.....	48
Bảng 3.4. Thời gian chạy thực nghiệm khi số lượng haplotype cố định và số lượng SNP thay đổi.....	48

DANH MỤC CÁC HÌNH

Hình	Trang
Hình 1.1. Một SNP (Single Nucleotide Polymorphisms)	4
Hình 1.2. Ba bước xây dựng HapMap	12
Hình 1.3. Hoạt động của bầy kiến trong thực tế.....	15
Hình 1.4. Ví dụ về đàn kiến nhân tạo.....	16
Hình 2.1. Thực nghiệm cây cầu đôi.....	27
Hình 2.2. Thí nghiệm bổ sung.....	28
Hình 2.3. Đồ thị cấu trúc tổng quát cho bài toán cực trị hàm.....	32
Hình 2.4. Thuật toán ACO.....	34
Hình 2.5. Tổng quan về thuật toán MACA.....	44