

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

NGUYỄN HÙNG CƯỜNG

**ỨNG DỤNG MỘT SỐ KỸ THUẬT KHAI PHÁ DỮ LIỆU
TRONG XÂY DỰNG MÔ HÌNH DỰ BÁO LƯU LƯỢNG
GIAO THÔNG THEO LOẠI HÌNH PHƯƠNG TIỆN**

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

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

NGUYỄN HÙNG CƯỜNG

**ỨNG DỤNG MỘT SỐ KỸ THUẬT KHAI PHÁ DỮ LIỆU
TRONG XÂY DỰNG MÔ HÌNH DỰ BÁO LƯU LƯỢNG
GIAO THÔNG THEO LOẠI HÌNH PHƯƠNG TIỆN**

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

Tiến sĩ: **Nguyễn Việt Anh**

LỜI CAM ĐOAN

Em xin cam đoan đây là công trình nghiên cứu của riêng em.

Các số liệu, trong luận văn được sử dụng từ nguồn số liệu được điều tra trong nghiên cứu khảo sát mẫu tại các hộ gia đình thực hiện năm 2014. Kết quả nghiên cứu này chưa được sử dụng trong bất kể nghiên cứu nào khác.

Thái Nguyên, tháng 5 năm 2016

TÁC GIẢ

Nguyễn Hùng Cường

MỤC LỤC

DANH MỤC CÁC KÝ HIỆU, CHỮ VIẾT TẮT.....	iii
DANH MỤC CÁC BẢNG	iv
DANH MỤC CÁC HÌNH VẼ	v
LỜI NÓI ĐẦU	vi
CHƯƠNG 1 TÌM HIỂU VỀ KHAI PHÁ DỮ LIỆU VÀ ỨNG DỤNG TRONG XÂY DỰNG MÔ HÌNH DỰ BÁO GIAO THÔNG	1
1.1. Tìm hiểu về khai phá dữ liệu	1
1.1.1. Khái niệm về khai phá dữ liệu	1
1.1.2. Một số kỹ thuật khai phá dữ liệu.....	3
1.1.3. Một số phương pháp khai phá dữ liệu	3
1.1.4. Một số công cụ hỗ trợ khai phá dữ liệu	5
1.2. Ứng dụng khai phá dữ liệu trong dự báo giao thông	5
1.2.1. Khái niệm dự báo	5
1.2.2. Ứng dụng khai phá dữ liệu trong dự báo giao thông	7
1.2.3. Mô hình dự báo giao thông phục vụ quy hoạch.....	8
1.2.4. Mô hình dự báo giao thông 4 bước:.....	8
CHƯƠNG 2 XÂY DỰNG MÔ HÌNH DỰ BÁO LƯU LƯỢNG GIAO THÔNG THEO LOẠI HÌNH PHƯƠNG TIỆN ĐỐI VỚI HÀ NỘI	16
2.1. Xây dựng mô hình dự báo.....	16
2.1.1. Dữ liệu xây dựng mô hình	16
2.1.2. Lý thuyết về mô hình hồi quy tuyến tính.....	17
2.2. Thông tin dữ liệu cần thu thập xây dựng mô hình	18
2.2.1. Thông tin số liệu thống kê.....	18
2.2.2. Thông tin dữ liệu từ điều tra hộ gia đình	19
2.3. Trích chọn thông tin và xây dựng mô hình dự báo	24
2.3.1. Xây dựng mô hình phát sinh, thu hút chuyên đi	24
2.3.1.1. Xây dựng mô hình phát sinh chuyên đi	26
2.3.1.2. Xây dựng mô hình thu hút chuyên đi.....	29

2.3.2. Xây dựng mô hình phân bố chuyển đi	32
2.3.3. Xây dựng mô hình phân chia phương tiện.....	36
CHƯƠNG 3 KIỂM ĐỊNH TRÊN SỐ LIỆU HIỆN TRẠNG VÀ DỰ BÁO ...	43
3.1. Kiểm định trên số liệu hiện trạng.....	43
3.1.1 Kiểm định mô hình phát sinh chuyển đi	43
3.1.2 Kiểm định mô hình thu hút chuyển đi	43
3.1.3 Kiểm định mô hình phân bố chuyển đi	45
3.1.4 Kiểm định mô hình phân chia phương tiện.....	46
3.2. Sử dụng mô hình cho dự báo tương lai.....	48
3.2.1 Dự báo phát sinh, thu hút chuyển đi năm 2020	49
3.2.2 Dự báo phân bố chuyển đi năm 2020	50
3.2.3 Dự báo lưu lượng phương tiện năm 2020	51
KẾT LUẬN	52
KIẾN NGHỊ	53
TÀI LIỆU THAM KHẢO.....	54
PHỤ LỤC	56
Phụ lục 1: Code chương trình trên ngôn ngữ R	56
Phụ lục 2: Mẫu điều tra hộ gia đình.....	64
Phụ lục 3: Kết quả dự báo phân bố chuyển đi năm 2020	77
Phụ lục 4: Kết quả dự báo lưu lượng giao thông phân theo phương tiện năm 2020	79

DANH MỤC CÁC KÝ HIỆU, CHỮ VIẾT TẮT

Từ viết tắt	Tiếng Anh	Tiếng Việt
A	Attraction	Thu hút chuyển đi
Bike	Bicycle	Xe đạp
Bus	Bus	Xe bus công cộng
CSDL	Database	Cơ sở dữ liệu
Car	Car	Xe con
G	Generate	Phát sinh chuyển đi
I	Origin (i)	Vùng đi (i)
Item	Item	Khoản mục
J	Destination (j)	Vùng đến (j)
KPDL	Knowledge Discovery and Data Mining	Khai phá dữ liệu
Moto	Motobike	Xe máy
PCU	Passenger Car Unit	Đơn vị quy đổi về xe con
VOC	Value Of Cost	Chi phí cho sử dụng phương tiện
VOT	Value Of Time	Chi phí thời gian người sử dụng

DANH MỤC CÁC BẢNG

Bảng 2. 1: Thông kê số lượng các chỉ tiêu theo phường.....	18
Bảng 2. 2: Chỉ tiêu về số chuyến đi bình quân trong ngày	20
Bảng 2. 3: chỉ tiêu về số km bình quân phương tiện và chi phí.....	20
Bảng 2. 4: Chỉ tiêu về thu nhập và sở hữu phương tiện.....	20
Bảng 2. 5: Số chuyến đi phân theo mục đích chuyến đi	21
Bảng 2. 6: Số chuyến đi phân theo loại hình phương tiện	23
Bảng 3. 1: Kết quả kiểm định mô hình phát sinh, thu hút	44
Bảng 3. 2: Kết quả kiểm định mô hình phân bổ chuyến đi.....	45
Bảng 3. 3: Kết quả kiểm định mô hình phân chia phương tiện	47
Bảng 3. 4: Dân số, số lao động và học sinh, sinh viên năm 2020.....	48
Bảng 3. 5: Kết quả dự báo số chuyến đi phát sinh và thu hút năm 2020.....	49
Bảng 3. 6: Kết quả phân bổ chuyến đi năm 2020	50
Bảng 3. 7 Tổng hợp dự báo lưu lượng theo loại hình phương tiện năm 2020.....	51

DANH MỤC CÁC HÌNH VẼ

Hình 1. 1: Tiến trình phát hiện tri thức	1
Hình 1. 2: Phát sinh và thu hút chuyển đi	9
Hình 1. 3: Phân bổ chuyển đi	10
Hình 1. 4: Phân chia phương tiện.....	10
Hình 1. 5: Mô phỏng chi phí tăng lên khi khối lượng được phân bổ.....	11
Hình 1. 6: Vận tốc giảm dần khi khối lượng tăng dần	12
Hình 1. 7: Mô hình dự báo 4 bước	13
Hình 2. 1. Mô hình hồi quy tuyến tính.....	17
Hình 2. 2: Đánh giá độ tương quan của các chỉ tiêu đầu vào	26
Hình 2. 3: Thông số đánh giá kết quả phân tích mô hình phát sinh	27
Hình 2. 4: Thông số đánh giá kết quả phân tích mô hình phát sinh khi thêm HSDC	28
Hình 2. 5: So sánh mối tương quan của mô hình phát sinh trước và sau khi thêm HSDC	29
Hình 2. 6: Thông số đánh giá kết quả phân tích mô hình thu hút.....	30
Hình 2. 7: Thông số đánh giá kết quả phân tích mô hình thu hút khi thêm HSDC	31
Hình 2. 8: So sánh mối tương quan của mô hình thu hút trước và sau khi thêm HSDC	31
Hình 2. 9: Mối tương quan số chuyển đi và cự ly vận chuyển	33
Hình 2. 10: Thông số đánh giá, phân tích mô hình phân bổ chuyển đi	34

Hình 2. 11: Thông số đánh giá, phân tích mô hình phân bố chuyến đi khi có biến giả	35
Hình 2. 12: Số liệu thực tế với số liệu lý thuyết khi có và không có biến giả	36
Hình 2. 13: Thông số đánh giá kết quả về độ thỏa dụng đối với xe đạp.....	38
Hình 2. 14: Thông số đánh giá kết quả về độ thỏa dụng đối với xe máy	39
Hình 2. 15: Thông số đánh giá kết quả về độ thỏa dụng đối với xe con.....	40
Hình 2. 16: Thông số đánh giá kết quả về độ thỏa dụng đối với xe bus.....	40

LỜI NÓI ĐẦU

Quy hoạch mạng lưới giao thông có vai trò hết sức quan trọng trong quy hoạch thành phố. Dự báo lưu lượng giao thông là cơ sở cho việc quy hoạch mạng lưới giao thông, nhằm đáp ứng tốt nhất nhu cầu đi lại trong hiện tại và trong tương lai.

Hiện nay để dự báo giao thông người ta thường tiến hành điều tra, khảo sát bằng các biện pháp như phỏng vấn, đếm xe để xây dựng ma trận nhu cầu đi lại trong hiện tại rồi dùng một số kỹ thuật dự báo như hệ số đàn hồi, tốc độ tăng trưởng để dự báo nhu cầu đi lại trong tương lai.

Việc sử dụng hệ số đàn hồi hoặc tốc độ tăng trưởng thường mang tính xu thế chung, có thể không phản ánh chính xác được nhu cầu đi lại cho một

vùng cũng như loại hình phương tiện đi lại của vùng đó đặc biệt là khi cơ cấu chuyển đi thay đổi và hành vi, thói quen sử dụng phương tiện thay đổi.

Nhằm hạn chế bớt sai số trong việc dự báo nhu cầu giao thông, luận văn đề xuất ứng dụng một số kỹ thuật Khai phá dữ liệu trong việc xây dựng mô hình dự báo lưu lượng giao thông theo loại hình phương tiện dựa trên dữ liệu khảo sát thực tế. Dữ liệu khảo sát bao gồm các thông tin về nơi đi, nơi đến; chi phí, thời gian, cự ly, mục đích chuyến đi; đặc điểm nghề nghiệp, tuổi, giới tính của người có chuyến đi v.v..

Nghiên cứu này nhằm xây dựng mô hình dự báo theo hướng tiếp cận hiện đại mà các nước phát triển cũng như các nước đang phát triển hiện đang sử dụng là phương pháp dự báo 4 bước.

Dự báo sẽ là cơ sở hỗ trợ cho việc hoạch định của các nhà quản lý về chính sách đầu tư phát triển cơ sở hạ tầng phù hợp nhu cầu của người dân.

Với ý nghĩa và tầm quan trọng đó, em lựa chọn đề tài “ ***Ứng dụng một số kỹ thuật khai phá dữ liệu trong xây dựng mô hình dự báo lưu lượng giao thông theo loại hình phương tiện*** ” với mục đích tìm hiểu và ứng dụng được một số kỹ thuật khai phá dữ liệu vào bài toán trên từ nguồn số liệu điều tra hộ gia đình được thực hiện năm 2014 tại thành phố Hà Nội, Việt Nam.

Luận văn gồm 3 chương với các nội dung cơ bản sau

Chương 1: Trình bày khái quát về dự báo và ứng dụng khai phá dữ liệu trong xây dựng mô hình dự báo giao thông.

Chương 2: Trình bày các kỹ thuật khai phá dữ liệu trong xây dựng mô hình dự báo lưu lượng giao thông từ bước 1 đến bước 3 trong bài toán dự báo 4 bước để xác định loại hình phương tiện đối với thủ đô Hà Nội.