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

CHU THỊ HẢO

KỸ THUẬT PHÂN CỤM DỮ LIỆU KHÔNG GIAN CÓ RÀNG BUỘC

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

THÁI NGUYÊN, 2017

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

CHU THỊ HẢO

KỸ THUẬT PHÂN CỤM DỮ LIỆU KHÔNG GIAN CÓ RÀNG BUỘC

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: PGS.TS. ĐẶNG VĂN ĐỨC

THÁI NGUYÊN, 2017

MỤC LỤC

DANH MỤC CÁC BẢNG

DANH MỤC CÁC HÌNH

MỞ ĐẦU

Hệ thống thông tin địa lý (GIS) được ứng dụng ngày càng phổ biến, không chỉ trong lĩnh vực giám sát, quản lý, lập kế hoạch về tài nguyên môi trường mà còn trong nhiều bài toán kinh tế xã hội khác. Kết quả là, khối lượng dữ liệu liên quan đến địa lý, còn gọi là dữ liệu không gian thu thập được tăng lên nhanh chóng. Một câu hỏi đặt ra là làm thế nào để tận dụng, khai thác, khám phá, phát hiện những tri thức hữu ích từ kho dữ liệu này?

Khai phá dữ liệu là áp dụng các kỹ thuật và công cụ để trích rút các tri thức có ích từ nguồn dữ liệu về một lĩnh vực nào đó mà ta quan tâm. Khai phá dữ liệu với GIS hay còn gọi là khai phá dữ liệu không gian, mở rộng khai phá dữ liệu trong các CSDL quan hệ, xét thêm các thuộc tính của dữ liệu không gian được phản ánh trong hệ thông tin địa lý, ví dụ khoảng cách (gần kề hay cách xa), điều kiện môi trường tự nhiên hay kinh tế xã hội (rừng núi, đồng bằng, ven biển, đô thị, v.v…).

Các bài toán truyền thống của một hệ thông tin địa lý có thể trả lời các câu hỏi kiểu như:

- Những con phố nào dẫn đến sân bay Tân Sân Nhất ?
- Những căn nhà nào nằm trong vùng quy hoạch mở rộng phố?

Khai phá dữ liệu không gian có thể giúp trả lời cho các câu hỏi dạng:

- Xu hướng của các dòng chảy, các đứt gãy địa tầng ?
- Nên bố trí các trạm tiếp sóng điện thoại di động như thế nào?
- Những vị trí nào là tối ưu để đặt các máy ATM, xăng dầu, nhà hàng,…?

Một trong những bài toán liên quan đến dữ liệu không gian, cụ thể là dữ liệu địa lý có ý nghĩa thực tế cao là bài toán xác định vị trí tối ưu cho việc đặt các cây xăng. Cả nước hiện có 374 tổng đại lý và hơn 14.000 cửa hàng bản lẻ xăng dầu. Để xác định được vị trí đặt các trạm bán lẻ xăng dầu cần

phải tuân theo các quy định của Bộ Công thương, nhất là các quy định về an toàn, phòng chống cháy nổ. Ngoài ra, cây xăng cũng phải đặt ở vị trí thuận lợi cho việc kinh doanh đạt doanh số cao. Hoặc một bài toán khác cũng có ý nghĩa thực tiễn rất lớn đó là xác định vị trí tối ưu để mở một nhà hàng. Hiện nay trên địa bàn thành phố Hà Nội cũng đã có rất nhiều nhà hàng, quán ăn đã được mở ra. Nhưng không phải tất cả các nhà hàng, quán ăn đó đều có thể cho doanh thu tốt. Có khi có nhà hàng mới mở ra được một thời gian ngắn đã phải đóng cửa vì không có khách dẫn đến chủ đầu tư phải chịu thua lỗ nặng. Một trong những nguyên nhân chính dẫn đến thất bại đó là địa điểm kinh doanh chưa hợp lý. Một vị trí tối ưu cho việc mở nhà hàng, quán ăn thì vị trí đó phải thỏa mãn một số yếu tố sau: nằm trong khu vực đông dân cư, gần nhiều cơ quan công sở hay trường học, có khu vực để xe, có quang cảnh xung quanh thoáng mát...các vấn đề này đã được rất nhiều các đề tài nghiên cứu tuy nhiên với những vị trí phức tạp có các ngăn cách con sông hay cây cầu v.v… thì cần phải có những đánh giá chính xác hơn nữa.

Xuất phát từ nhu cầu thực tế đó và do đặc thù, khả năng ứng dụng rất phong phú của kỹ thuật phân cụm dữ liệu trong không gian nên em đã chọn nghiên cứu đề tài kỹ thuật phân cụm dữ liệu không gian có ràng buộc làm luận văn tốt nghiệp cao học.

Trên cơ sở đó cài đặt thử nghiệm một ứng dụng sử dụng kỹ thuật phân cụm dữ liệu không gian, trong đó khai thác thông tin địa lý của các đối tượng để hỗ trợ giải quyết bài toán ví dụ như tìm vị trí tối ưu đặt nhà hàng.

Luận văn được chia thành các chương mục sau

- Chương 1: Tổng quan về khai phá dữ liệu và dữ liệu không gian
- Chương 2: Một số kỹ thuật phân cụm dữ liệu không gian
- Chương 3: Xây dựng chương trình thử nghiệm, kết luận, đánh giá

Luận văn này được hoàn thành dưới sự hướng dẫn tận tình của PGS.TS Đặng Văn Đức, em xin bày tỏ lòng biết ơn chân thành của mình đối với thầy. Em xin chân thành cảm ơn các thầy, cô giáo Viện Công nghệ thông tin,

Trường Đại học Công nghệ thông tin và Truyền thông - Đại học Thái Nguyên đã tham gia giảng dạy, giúp đỡ em trong suốt qúa trình học tập nâng cao trình độ kiến thức. Tuy nhiên vì điều kiện thời gian và khả năng có hạn nên luận văn không thể tránh khỏi những thiếu sót. Em kính mong các thầy cô giáo và các bạn đóng góp ý kiến để đề tài được hoàn thiện hơn