BỘ GIÁO DỤC & ĐÀO TẠO

TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

-------o0o-------

ĐỒ ÁN TỐT NGHIỆP

Ngành công nghệ thông tin

HẢI PHÒNG 2015

BỘ GIÁO DỤC & ĐÀO TẠO

TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

-------o0o-------

TRA CỨU ẢNH DỰA TRÊN NỘI DUNG VỚI PHẢN HỒI LIÊN QUAN SỬ DỤNG MÔ HÌNH HỌC TRÊN ĐỒ THỊ

ĐỒ ÁN TỐT NGHIỆP

Ngành Công nghệ Thông tin

HẢI PHÒNG - 2015

BỘ GIÁO DỤC & ĐÀO TẠO

TR- ỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

-------o0o-------

TRA CỨU ẢNH DỰA TRÊN NỘI DUNG VỚI PHẢN HỒI LIÊN QUAN SỬ DỤNG MÔ HÌNH HỌC TRÊN ĐỒ THỊ

ĐỒ ÁN TỐT NGHIỆP ĐẠI HỌC HỆ CHÍNH QUY

Ngành : Công nghệ Thông tin

Sinh viên thực hiện: PHẠM ANH TOÀN

Giáo viên hướng dẫn: NGÔ TRƯỜNG GIANG

Mã sinh viên : 1112101005

HẢI PHÒNG - 2015

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC DÂN LẬP HẢI PHÒNG

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do – Hạnh phúc
-------o0o-------

NHIỆM VỤ THIẾT KẾ TỐT NGHIỆP

Sinh viên : PHẠM ANH TOÀN **Mã số : 1112101005**

Lớp : CT1501 **Ngành: Công nghệ Thông tin**

Tên đề tài : TRA CỨU ẢNH VỚI PHẢN HỒI LIÊN QUAN SỬ DỤNG MÔ HÌNH HỌC TRÊN ĐỒ THỊ

NHIỆM VỤ ĐỀ TÀI

1. Nội dung và các yêu cầu cần giải quyết trong nhiệm vụ đề tài tốt nghiệp

a. Nội dung:

- Tổng quan về Tra cứu ảnh dựa trên nội dung với phản hồi liên quan
- Tổng quan về mô hình học trên đồ thị.
- Ứng dụng học trên đồ thị cho bài toán tra cứu ảnh.
- Cài đặt chương trình thử nghiệm.

b. Các yêu cầu cần giải quyết

- Hiểu quy trình của một hệ thống tra cứu ảnh dựa trên nội dung, các phương pháp cơ bản trong tra cứu ảnh dựa trên nội dung.
- Hiểu được một số mô hình học dựa trên đồ thị và áp dụng cho cải thiện hiệu quả tra cứu.
- Cài đặt chương trình thử nghiệm

2. Các số liệu cần thiết để thiết kế, tính toán

3. Địa điểm thực tập

CÁN BỘ H- ỚNG DẪN ĐỀ TÀI TỐT NGHIỆP

Ng- ời h- ớng dẫn thứ nhất :

Họ và tên: ……

Học hàm, học vị: ……

Cơ quan công tác: ……..

Nội dung h□ớng dẫn:

……
……
……
……

Ng- ời h- ớng dẫn thứ hai:

Họ và tên : ……

Học hàm, học vị :……

Cơ quan công tác: ……..

Nội dung h□ớng dẫn:

……
……
……
……

Đề tài tốt nghiệp đ- ợc giao ngày 06 tháng 04 năm 2015

Yêu cầu phải hoàn thành tr- ớc ngày 11 tháng 07 năm 2015

Đã nhận nhiệm vụ: Đ.T.T.N	Đã nhận nhiệm vụ: Đ.T.T.N
Sinh viên	Cán bộ h□ớng dẫn Đ.T.T.N

Hải Phòng, ngày…tháng…..năm 2015

HIỆU TR- ỞNG

GS.TS.NG-T Trần Hữu Nghị

PHẦN NHẬN XÉT TÓM TẮT CỦA CÁN BỘ HƯỚNG DẪN

1. Tinh thần thái độ của sinh viên trong quá trình làm đề tài tốt nghiệp:

   .....

   .....

   .....

   .....

   .....

   .....

   .....

   .....

2. Đánh giá chất lượng của đề tài tốt nghiệp (so với nội dung yêu cầu đã đề ra trong nhiệm vụ đề tài tốt nghiệp)

   .....

   .....

   .....

   .....

   .....

   .....

   .....

   .....

3. Cho điểm của cán bộ hướng dẫn:

   (Điểm ghi bằng số và chữ)

.....

.....

Ngày.....tháng.....năm 2015

Cán bộ hướng dẫn chính

(Ký, ghi rõ họ tên)

PHẦN NHẬN XÉT ĐÁNH GIÁ CỦA CÁN BỘ CHẤM PHẢN BIỆN ĐỀ TÀI TỐT NGHIỆP

1. Đánh giá chất lượng đề tài (về các mặt như cơ sở lý luận, thuyết minh chương trình, giá trị thực tế…)

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

2. Cho điểm của cán bộ phản biện

(*Điểm ghi bằng số và chữ*)

.....

.....

Ngày.....tháng.....năm 2015

Cán bộ chấm phản biện

(Ký, ghi rõ họ tên)

LỜI CẢM ƠN

Em xin chân thành cảm ơn Thầy giáo, Thạc sĩ Ngô Trường Giang đã hướng dẫn tận tình chỉ bảo em rất nhiều trong suốt quá trình tìm hiểu nghiên cứu và hoàn thành đồ án này từ lý thuyết đến ứng dụng. Sự hướng dẫn của thầy đã giúp em có thêm kiến thức về lập trình và kiến thức về lĩnh vực xử lý ảnh. Đồng thời, em xin chân thành cám ơn các thầy cô trong khoa Công nghệ thông tin – Trường Đại Học Dân Lập Hải Phòng, cũng như các thầy cô trong trường đã trang bị cho em những kiến thức cơ bản cần thiết trong suốt thời gian học tập tại trường để em hoàn thành tốt đồ án này. Em xin chân thành cảm ơn GS.TS.NGƯT Trần Hữu Nghị, Hiệu trưởng Trường Đại học Dân Lập Hải Phòng, ban giám hiệu nhà trường, khoa Công nghệ thông tin, các phòng ban nhà trường đã tạo điều kiện tốt nhất trong suốt thời gian em học tập và làm tốt nghiệp. Trong quá trình học cũng như trong suốt thời gian làm đồ án tốt nghiệp không tránh khỏi những thiếu sót, em rất mong được sự góp ý quý báu của các thầy cô cũng như tất cả các bạn để kết quả của em được hoàn thiện hơn. Sau cùng, em xin gửi lời cảm ơn đến gia đình, bạn bè đã tạo mọi điều kiện để em xây dựng thành công đồ án này.

Em xin chân thành cảm ơn !

MỤC LỤC

MỘT SỐ TỪ VIẾT TẮT

STT	Từ viết tắt	Mô tả
1	CBIR	Content-Based Image Retrieval
2	EM	Expectation Maximization
3	PCA	Principal Component Analysis
4	RF	Relevance Feedback
5	RGB	Red-Green-Blue
6	SVM	Support Vector Machine
7	TSVM	Transductive Support Vector Machine

MỞ ĐẦU

Với sự phát triển của Internet cũng như các thiết bị ghi và lưu trữ ảnh, kích thước của các tập ảnh số được gia tăng một cách nhanh chóng. Hiệu quả của các công cụ tìm kiếm, tra cứu ảnh được yêu cầu từ rất nhiều lĩnh vực khác nhau bao gồm : trinh sát, thời trang, phòng chống tội phạm, xuất bản, kiến trúc, y tế v.v… Cùng chung mục đích này, rất nhiều các hệ thống tra cứu ảnh đã được phát triển. Có hai nền tảng là : dựa trên văn bản (text-based) và dựa trên nội dung (content-based).

Các phương pháp tiếp cận dựa trên văn bản được sử dụng từ những năm 1970. Trong đó các ảnh được chú thích bởi các mô tả văn bản một cách thủ công, sau đó được sử dụng bởi các hệ thống quản lý cơ sở dữ liệu để thực hiện việc tra cứu ảnh. Có hai nhược điểm cho quá trình tra cứu ảnh dựa trên văn bản. Đầu tiên là yêu cầu về mức lao động đáng kể của con người cho việc chú thích thủ công. Thứ hai là vấn đề chú thích không chính xác do nhận thức chủ quan của con người. Để khắc phục hai nhược điểm trên của hệ thống tra cứu ảnh dựa trên văn bản, khái niệm tra cứu ảnh dựa trên nội dung được giới thiệu vào đầu những năm 1980.

Đồ án trình bày kỹ thuật tra cứu ảnh dựa trên nội dung sử dụng phản hồi có liên quan với mô hình học dựa trên đồ thị, Đồ án bao gồm có 3 phần :

- Chương 1 : Tổng quan về hệ thống tra cứu ảnh dựa trên nội dung với phản hồi liên quan.
- Chương 2 : Mô hình học bán giám sát dựa trên đồ thị.
- Chương 3 : Áp dụng cài đặt chương trình và một số kết quả thực nghiệm.

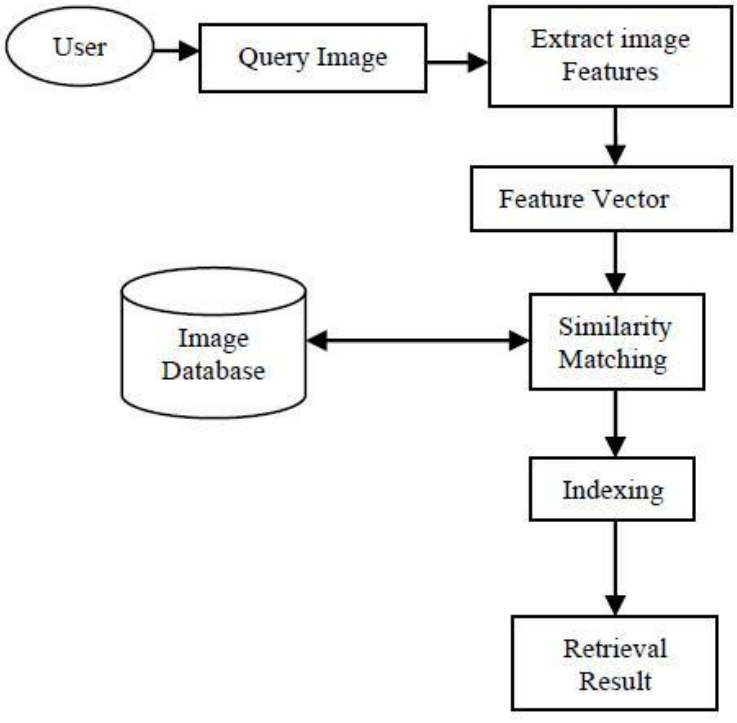
CHƯƠNG 1: Tổng quan về tra cứu ảnh dựa trên nội dung với phản hồi liên quan

1.1 Khái niệm tra cứu ảnh dựa trên nội dung

Một hệ thống CBIR được dùng để tìm kiếm các ảnh số trong một cơ sở dữ liệu lớn và tra cứu những ảnh liên quan dựa trên nội dung thực tế của nó. Nội dung có thể ở dạng các đặc trưng mức thấp hoặc bất kỳ thông tin nào có được từ hình ảnh. Trong CBIR, hình ảnh được trích chọn các đặc trưng mức thấp một cách tự động để biểu diễn nội dung trực quan, sau đó hệ thống sử dụng các véc-tơ đặc trưng để đánh giá độ tương tự giữa các ảnh.

1.2 Những thành phần của một hệ thống tra cứu ảnh dựa trên nội dung

Một hệ thống tra cứu ảnh đòi hỏi các thành phần như trong hình 1-1 [5]. Trong đó có ba thành phần quan trọng nhất trong tra cứu ảnh dựa trên nội dung : trích chọn đặc trưng, đánh chỉ số và giao diện truy vấn cho người dùng.



Hình 1-1: Kiến trúc tổng quan về hệ thống tra cứu ảnh dựa trên nội dung

Các bước tra cứu ảnh trong CBIR thường bao gồm :

- Tiếp nhận truy vấn của người dùng (dưới dạng ảnh hoặc phác thảo).
- Trích chọn đặc trưng của truy vấn và lưu trữ vào cơ sở dữ liệu đặc trưng như là một véc-tơ hoặc không gian đặc trưng.
- So sánh độ tương tự giữa các đặc trưng trong cơ sở dữ liệu với nhau từng đôi một.
- Lập chỉ mục cho các véc-tơ để nâng hiệu quả tra cứu.
- Trả lại kết quả tra cứu cho người dùng.

1.2.1 Các đặc trưng hình ảnh mức thấp

Các đặc trưng của ảnh bao gồm các đặc tính cơ bản và các đặc tính ngữ nghĩa/logic. Các đặc tính cơ bản đó là: màu sắc (color), hình dạng (shape), kết cấu (texture), vị trí không gian (spatial location). Chúng có thể được trích xuất tự động hoặc bán tự động. Đặc tính logic cung cấp mô tả trừu tượng của dữ liệu hình ảnh ở các cấp độ khác nhau. Thông thường, các đặc tính logic được trích chọn bằng tay hoặc bán tự động. Một hoặc nhiều đặc trưng có thể được sử dụng trong ứng dụng cụ thể.

1.2.1.1 Đặc trưng màu sắc

Đặc trưng màu sắc là một trong những đặc trưng được sử dụng phổ biến trong tra cứu ảnh. Màu sắc được định nghĩa trên một không gian màu. Có rất nhiều không gian màu đã được xây dựng sẵn, chúng thường được dùng cho các ứng dụng khác nhau. Những không gian màu gần gũi hơn với nhận thức của con người và được sử dụng rộng rãi trong CBIR bao gồm RGB, LAB, LUV, HSV, HSL ... Vào năm 1999, Gevers và cộng sự đã quan tâm đến các đối tượng lấy từ các điểm quan sát khác nhau và sự chiếu sáng. Theo kết quả, một tập các điểm bất biến đặc trưng màu đã được tính toán. Các bất biến màu được xây dựng trên cơ sở hue, cặp hue-hue, và ba đặc trưng màu được tính

toán từ các mô hình đối xứng. Các đặc trưng màu sắc mặc dù mô tả màu sắc rất hiệu quả nhưng không trực tiếp liên quan đến các ngữ nghĩa mức cao.

1.2.1.2 Đặc trưng kết cấu

Kết cấu không được định nghĩa đầy đủ như là đặc trưng màu sắc, vì thế mà một số hệ thống không sử dụng đặc trưng kết cấu. Tuy nhiên, kết cấu cung cấp các thông tin quan trọng trong việc phân loại ảnh, vì nó mô tả nội dung của nhiều ảnh thực như là: vỏ trái cây, mây, cây, gạch ... Do đó, kết cấu là một đặc trưng quan trọng trong việc định nghĩa ngữ nghĩa mức cao cho mục đích tra cứu ảnh [5]. Các đặc trưng kết cấu thường được sử dụng trong hệ thống tra cứu ảnh bao gồm các đặc trưng phổ, chẳng hạn như các đặc trưng được bao gồm sử dụng lọc Gabor hoặc biến đổi wavelet, thống kê đặc trưng kết cấu trong các cách đo độ thống kê cục bộ, như sáu đặc trưng kết cấu Tamura, và đặc trưng wold được đề xuất bởi Liu và các cộng sự vào năm 1996.

1.2.1.3 Đặc trưng hình dạng

Hình dạng là một khái niệm được định nghĩa khá tốt. Đặc trưng hình dạng của các ứng dụng nói chung bao gồm: tỷ lệ aspect, tuần hoàn, mô tả Fourier, bất biến thời điểm, phân đoạn đường bao liên tiếp [8], v.v.. Đặc trưng hình dạng là đặc trưng ảnh quan trọng, mặc dù chúng chưa được sử dụng rộng rãi trong CBIR như là đặc trưng màu và đặc trưng kết cấu [5]. Đặc trưng hình dạng đã thể hiện tính hữu ích trong nhiều miền ảnh đặc biệt như là các đối tượng nhân tạo. Ảnh màu được sử dụng phổ biến trong nhiều tài liệu, tuy nhiên lại khó khăn để áp dụng đặc trưng hình dạng so với màu sắc và kết cấu do sự thiếu chính xác của phân đoạn. Mặc dù gặp khó khăn, đặc trưng hình dạng vẫn được sử dụng trong một số hệ thống và cho thấy tiềm năng trong RBIR (Region-based image retrieval).

1.2.1.4 Đặc trưng vị trí không gian

Các vùng hoặc đối tượng với thuộc tính màu sắc và kết cấu tương tự có thể được nhận ra một cách dễ dàng bởi ràng buộc không gian [5]. Ví dụ "bầu trời" và "biển" có thể có cùng đặc trưng về màu sắc và kết cấu nhưng lại có vị trí không gian trong ảnh khác nhau. Bầu trời thường xuất hiện ở phía trên của ảnh trong khi biển thường nằm ở dưới cùng. Đặc trưng không gian thường được định nghĩa một cách đơn giản như là "trên, dưới" tùy theo vị trí các vùng trong ảnh.

Mối quan hệ không gian tương đối là quan trọng hơn vị trí không gian tuyệt đối. 2D-string và một số biến thể của nó là cấu trúc chung phổ biến để biểu diễn mối quan hệ về phương hướng giữa các đối tượng như là "trái/phải", "trên/dưới".

1.2.2 Đánh chỉ số

Một vấn đề quan trọng khác trong tra cứu ảnh dựa trên nội dung là đánh chỉ số và tìm kiếm nhanh ảnh dựa trên đặc trưng trực quan. Bởi vì, các véc-tơ đặc trưng của ảnh có xu hướng có số chiều cao và do đó nó không thích hợp cho các cấu trúc đánh chỉ số truyền thống. Việc giảm số chiều thường xuyên được sử dụng trước khi lên kế hoạch đánh chỉ số.

Một trong những công nghệ được sử dụng phổ biến cho việc giảm số chiều là phân tích thành phần chính PCA [5]. Nó là một công nghệ tối ưu trong việc ánh xạ tuyến tính dữ liệu đầu vào một không gian toạ độ, các trục được thẳng hàng để phản ánh các biến thể lớn nhất trong dữ liệu. Hệ thống QBIC sử dụng PCA để làm giảm véc-tơ đặc trưng hình dạng có 20 chiều thành hai hoặc ba chiều. Ngoài công nghệ PCA ra, nhiều nhà nghiên cứu còn sử dụng biến đổi KL để làm giảm số chiều trong không gian đặc trưng. Mặc dù, biến đổi KL có một số thuộc tính hữu dụng như khả năng xác định vị trí hầu hết không gian con quan trọng, các thuộc tính đặc trưng mà quan trọng

đối với việc xác định mô hình tương tự có thể bị phá huỷ trong suốt quá trình giảm các chiều mù. Ngoài hai công nghệ biến đổi PCA và KL, thì mạng nơ-ron cũng là công cụ hữu ích cho việc giảm số chiều đặc trưng.

Sau khi đã giảm số chiều thì dữ liệu đa chiều được đánh chỉ số. Có nhiều phương pháp tiếp cận bao gồm : R-tree, linear quad-trees, K-d-B tree, grid files … Hầu hết các phương pháp này cho hiệu quả hợp lý với không gian có số chiều nhỏ.

1.2.3 Tương tác người dùng

Đối với tra cứu ảnh dựa trên nội dung, người dùng tương tác với các hệ thống tra cứu là rất quan trọng khi các hình thức và thay đổi linh hoạt của truy vấn chỉ có thể thu được bằng cách liên hệ với người sử dụng trong các thủ tục tra cứu. Giao diện người dùng trong các hệ thống tra cứu hình ảnh thông thường bao gồm phần xây dựng truy vấn và phần trình bày kết quả.

1.2.3.1 Xác định truy vấn

Để xác định những loại hình ảnh người sử dụng muốn lấy từ cơ sở dữ liệu thì có thể thực hiện bằng nhiều cách. Và những cách thông thường nhất được sử dụng là: duyệt qua, truy vấn bởi khái niệm, truy vấn bởi bản phác thảo, và truy vấn bởi ví dụ.

Duyệt qua là phương pháp duyệt qua toàn bộ cơ sở dữ liệu theo danh mục các ảnh. Với mục đích này, ảnh trong cơ sở dữ liệu được phân loại thành nhiều mục khác nhau theo ngữ nghĩa hoặc nội dung trực quan. Truy vấn bởi khái niệm là tra cứu ảnh theo mô tả khái niệm liên quan với từng ảnh trong cơ sở dữ liệu [5].

Truy vấn bởi bản phác thảo và truy vấn bởi ví dụ là vẽ ra một bản phác thảo hoặc cung cấp một ảnh ví dụ từ những ảnh với độ tương tự đặc trưng trực quan sẽ được trích chọn từ cơ sở dữ liệu.

Truy vấn bằng cách phác thảo cho phép người sử dụng vẽ một bức phác họa một hình ảnh với một công cụ chỉnh sửa đồ họa cung cấp bởi hệ thống tra cứu hoặc bằng một số phần mềm khác. Truy vấn có thể được hình thành bằng cách vẽ một số đối tượng có tính chất nhất định như màu sắc, kết cấu, hình dạng, kích thước và vị trí. Trong hầu hết các trường hợp, một bản phác thảo thô là đủ, các truy vấn có thể được chọn lọc dựa trên kết quả tra cứu.

Truy vấn bằng ví dụ cho phép người sử dụng xây dựng một truy vấn bằng cách cung cấp một hình ảnh ví dụ. Hệ thống chuyển đổi hình ảnh ví dụ thành một đại diện các đặc trưng nội bộ. Sau đó những hình ảnh được lưu trữ trong cơ sở dữ liệu với các đặc trưng tương tự được tìm kiếm. Truy vấn bằng ví dụ có thể được phân chia thành truy vấn bằng ví dụ bên ngoài, nếu hình ảnh truy vấn không có trong cơ sở dữ liệu, và truy vấn bằng ví dụ bên trong, nếu ngược lại. Đối với truy vấn bằng hình ảnh bên trong, tất cả các mối quan hệ giữa các hình ảnh có thể được tính toán trước. Ưu điểm chính của truy vấn bằng ví dụ là người dùng không cần phải cung cấp một mô tả rõ ràng về mục tiêu, nó được tính toán bởi hệ thống. Nó phù hợp cho các ứng dụng mà mục tiêu là một hình ảnh của cùng một đối tượng, hoặc thiết lập các đối tượng theo các điều kiện xem khác nhau. Hầu hết các hệ thống hiện tại cung cấp các truy vấn hình thức này.

Truy vấn bằng một nhóm ví dụ cho phép người dùng lựa chọn nhiều hình ảnh. Sau đó hệ thống sẽ tìm những hình ảnh phù hợp nhất với đặc điểm chung của nhóm các ví dụ. Bằng cách này, một mục tiêu có thể được xác định chính xác hơn bằng cách xác định các biến thể đặc trưng liên quan và loại bỏ các biến thể không thích hợp trong các truy vấn. Ngoài ra, các thuộc tính của nhóm có thể được chọn lọc bằng cách thêm những mẫu dương. Nhiều hệ thống phát triển gần đây cung cấp truy vấn bằng cả mẫu dương và mẫu âm.

1.2.3.2 Phản hồi liên quan

Khái niệm phản hồi liên quan đã được giới thiệu trong tra cứu ảnh dựa trên nội dung từ khái niệm tra cứu thông tin dựa trên văn bản vào năm 1998 và sau đó đã trở thành một kỹ thuật phổ biến cho CBIR để giảm khoảng cách ngữ nghĩa. Nói chung, phản hồi liên quan nhằm mục đích cải thiện hiệu năng tra cứu với sự tham gia điều chỉnh của người dùng trên kết quả tra cứu.

1.3 Khoảng cách ngữ nghĩa

Trở ngại lớn trong tra cứu ảnh trên nội dung đó là khoảng cách ngữ nghĩa. Con người có xu hướng sử dụng các khái niệm mức cao ví dụ như từ khóa, mô tả bằng văn bản để diễn tả các hình ảnh và đo sự tương tự giữa chúng. Trong khi đó việc trích chọn đặc trưng một cách tự động sử dụng các kỹ thuật thị giác máy hầu hết là các đặc trưng mức thấp (màu sắc, kết cấu, hình dạng, bố cục không gian…). Nói chung không có một mối liên hệ trực tiếp nào giữa các khái niệm mức cao và đặc trưng mức thấp.

Mặc dù đã có rất nhiều thuật toán phức tạp được thiết kế để mô tả các đặc trưng về màu sắc, hình dạng, kết cấu, tuy nhiên những thuật toán này vẫn không thể mô tả đầy đủ ngữ nghĩa của hình ảnh và có nhiều hạn chế khi làm việc với một cơ sở dữ liệu lớn [2]. Thí nghiệm rộng rãi trên hệ thống CBIR cho thấy các nội dung mức thấp thường không mô tả được các khái niệm ngữ nghĩa mức cao trong suy nghĩ của người sử dụng [3]. Do đó, hiệu suất của CBIR vẫn còn xa sự mong đợi của người dùng.

Trong [1] Eakins đã đề cập tới ba cấp độ truy vấn trong CBIR :

- Cấp độ 1 : Tra cứu bằng các đặc trưng nguyên thủy như màu sắc, kết cấu, hình dạng hoặc vị trí không gian của các yếu tố hình ảnh. Điển hình là các truy vấn bằng ví dụ, ‘tìm ảnh giống như thế này’

- Cấp độ 2 : Tra cứu các đối tượng có dạng xác định bởi các đặc trưng gốc và một mức độ suy luận logic. Ví dụ 'tìm ảnh một bông hoa'.

- Cấp độ 3 : Tra cứu bằng các thuộc tính trừu tượng liên quan tới một lượng đáng kể ý nghĩa mức cao về mục đích của đối tượng hoặc miêu tả cảnh vật. Điều này bao gồm tra cứu các sự kiện được đặt tên, các hình ảnh có ý nghĩa về cảm xúc và tinh thần… Ví dụ 'tìm hình ảnh một đám đông vui vẻ'.

Cấp độ 2 và 3 đều ứng với việc tra cứu ngữ nghĩa của hình ảnh. Khoảng giữa cấp độ 1 và cấp độ 2 cũng giống khoảng cách ngữ nghĩa. Cụ thể hơn, sự khác biệt giữa giới hạn khả năng mô tả của đặc trưng hình ảnh mức thấp và sự phong phú về ngữ nghĩa của người dùng được gọi là "khoảng cách ngữ nghĩa".

Để nâng cao hiệu xuất trong CBIR đòi hỏi cần có các phương pháp giảm khoảng cách này. Một trong các phương pháp đó là phản hồi liên quan.

1.4 Kỹ thuật phản hồi liên quan trong CBIR

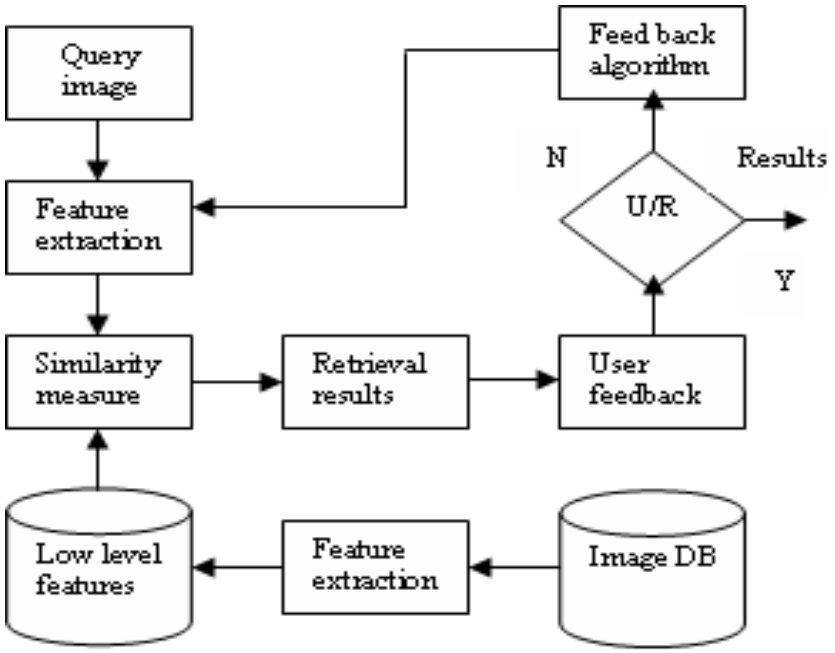
1.4.1 Khái niệm phản hồi liên quan

Nhận thức của con người về độ tương tự của hình ảnh là chủ quan, ngữ nghĩa, và phụ thuộc vào từng nhiệm vụ cụ thể. Mặc dù phương pháp dựa trên nội dung hứa hẹn một hướng đi triển vọng cho tra cứu ảnh, nói chung các kết quả tra cứu dựa trên những điểm tương đồng của các đặc trưng trực quan thuần túy là không nhất thiết có ý nghĩa về nhận thức và ngữ nghĩa. Ngoài ra, mỗi loại đặc trưng trực quan có xu hướng chỉ nắm bắt một khía cạnh của thuộc tính hình ảnh và nó thường khó khăn cho người sử dụng để xác định rõ những khía cạnh khác nhau được kết hợp. Để giải quyết những vấn đề này, tương tác phản hồi liên quan, một kỹ thuật trong hệ thống tìm kiếm thông tin dựa trên văn bản truyền thống, đã được giới thiệu. Với phản hồi liên quan, có thể thiết lập liên kết giữa các khái niệm mức cao và đặc trưng mức thấp. Ý

tưởng chính là sử dụng các mẫu dương và mẫu âm từ người sử dụng để cải thiện hiệu suất hệ thống. Đối với một truy vấn nhất định, đầu tiên hệ thống sẽ trả về một danh sách các hình ảnh được xếp theo một độ tương tự xác định trước. Sau đó, người dùng đánh dấu những hình ảnh có liên quan đến truy vấn (mẫu dương) hoặc không có liên quan (mẫu âm). Hệ thống sẽ chọn lọc kết quả tra cứu dựa trên những phản hồi và trình bày một danh sách mới của hình ảnh cho người dùng. Do đó, vấn đề quan trọng trong phản hồi liên quan là làm thế nào để kết hợp các mẫu dương và mẫu âm để tinh chỉnh các truy vấn và/hoặc điều chỉnh các biện pháp tương tự.

1.4.2 Kiến trúc tổng quan của hệ thống CBIR với phản hồi liên quan

Hình 3-1 cho thấy mô hình tổng quát của một hệ thống tra cứu ảnh từ cơ sở dữ liệu sử dụng phản hồi liên quan



Hình 1-2 : Mô hình tổng quát hệ thống tra cứu ảnh sử dụng phản hồi liên quan

Ý tưởng chính của phản hồi liên quan là chuyển trách nhiệm tìm kiếm xây dựng truy vấn đúng từ người dùng sang hệ thống. Để thực hiện điều này một cách đúng đắn, người dùng phải cung cấp cho hệ thống một số thông tin, để hệ thống có thể thực hiện tốt việc trả lời truy vấn ban đầu.

Việc tìm kiếm ảnh thường dựa trên sự tương tự hơn là so sánh chính xác, và kết quả tra cứu sẽ được đưa ra cho người dùng. Sau đó, người dùng đưa ra các thông tin phản hồi trong một bản mẫu "Các quyết định liên quan" thể hiện thông qua kết quả tra cứu. "Quyết định liên quan" đánh giá kết quả dựa trên ba giá trị. Ba giá trị đó là: liên quan, không liên quan, và không quan tâm. "Liên quan" nghĩa là ảnh có liên quan đến truy vấn của người dùng. "Không liên quan" có nghĩa là ảnh không có liên quan đến truy vấn người dùng. Còn "không quan tâm" nghĩa là người dùng không cho biết bất kỳ điều gì về ảnh. Nếu phản hồi của người dùng là có liên quan, thì vòng lặp phản hồi sẽ tiếp tục hoạt động cho đến khi người dùng hài lòng với kết quả tra cứu. Như hình 2-1 mô tả cấu trúc của hệ thống phản hồi liên quan. Trong hệ thống đó có các khối chính là: cơ sở dữ liệu ảnh, trích chọn đặc trưng, đo độ tương tự, phản hồi từ người dùng, và thuật toán phản hồi.

1.4.2.1 Trích chọn đặc trưng

Trích chọn đặc trưng liên quan đến việc trích chọn các thông tin có ý nghĩa từ ảnh. Vì vậy, nó làm giảm việc lưu trữ cần thiết, và do đó hệ thống sẽ trở nên nhanh hơn và hiệu quả trong CBIR. Khi đặc trưng được trích chọn, chúng sẽ được lưu trữ trong cơ sở dữ liệu để sử dụng trong lần truy vấn sau này. Mức độ mà một máy tính có thể trích chọn thông tin có ích từ ảnh là vấn đề then chốt nhất cho sự tiến bộ của hệ thống diễn giải hình ảnh thông minh. Một trong những ưu điểm lớn nhất của trích chọn đặc trưng là: nó làm giảm đáng kể các thông tin (so với ảnh gốc) để biểu diễn một ảnh cho việc hiểu nội dung của ảnh đó. Hiện nay đã có rất nhiều nghiên cứu lớn về các phương pháp tiếp cận khác nhau để phát hiện nhiều loại đặc trưng trong ảnh. Những đặc trưng này có thể được phân loại như là đặc trưng toàn cục và đặc trưng cục bộ. Các đặc trưng phổ biến nhất mà được sử dụng là màu sắc, kết cấu và hình dạng.

- Đặc trưng toàn cục: Đặc trưng toàn cục phải được tính toán trên toàn bộ ảnh. Ví dụ, mức độ màu xám trung bình, biểu đồ về cường độ hình dạng, v.v… Ưu điểm của việc trích chọn toàn cục là tốc độ nhanh chóng trong cả trích chọn đặc trưng và tính toán độ tương tự. Tuy nhiên, chúng có thể quá nhạy cảm với vị trí và do đó không xác định được các đặc tính trực quan quan trọng. Để tăng cường sự vững mạnh trong biến đổi không gian, chúng ta có thể tìm hiểu trích chọn đặc trưng cục bộ.

- Đặc trưng cục bộ: Trong đặc trưng toàn cục, các đặc trưng được tính toán trên toàn bộ ảnh. Tuy nhiên, đặc trưng toàn cục không thể nắm bắt tất cả các vùng ảnh có đặc điểm khác nhau. Do đó, việc trích chọn các đặc trưng cục bộ của ảnh là cần thiết. Các đặc trưng đó có thể được tính toán trên các kết quả của phân đoạn ảnh và thuật toán phát hiện biên. Vì thế, tất cả chúng đều dựa trên một phần của ảnh với một số tính chất đặc biệt.

- Điểm nổi bật: Trong việc tính toán đặc trưng cục bộ, việc trích chọn đặc trưng ảnh bị giới hạn trong một tập nhỏ các điểm ảnh, đó là những điểm chú ý. Tập các điểm chú ý được gọi là những điểm nổi bật. Những điểm nổi bật là những điểm có dao động lớn trong đặc trưng của vùng lân cận điểm ảnh. Nhiều hệ thống CBIR trích chọn những điểm nổi bật. Năm 2004, Rouhollah và các cộng sự đã định nghĩa điểm nổi bật có mặt trong tra cứu ảnh dựa trên nội dung như là một nhiệm vụ của CBIR, nơi mà người dùng chỉ quan tâm đến một phần của ảnh, và phần còn lại là không liên quan. Ví dụ, chúng ta có thể tham khảo một số đặc trưng cục bộ như là ảnh nguyên bản, đường tròn, đường nét, texel (các phần tử tập trung ở một khu vực kết cấu), hoặc các đặc trưng cục bộ khác, hình dạng của đường nét, v.v…

1.4.2.2 Đo độ tương tự

Trong độ đo tương tự, véc-tơ đặc trưng của ảnh truy vấn và véc-tơ đặc trưng của ảnh trong cơ sở dữ liệu được đối sánh bằng cách sử dụng một thước đo khoảng cách. Các hình ảnh được xếp hạng dựa trên giá trị khoảng cách. Vào năm 2003, Manesh và các cộng sự đã đề xuất phương pháp đo độ tương tự cho việc đối sánh chi tiết các độ đo khác nhau như: Manhattan, weighted mean-variance, Euclidean, Chebychev, Mahanobis, v.v… cho tra cứu kết cấu ảnh với đánh giá thực nghiệm. Họ nhận thấy rằng số liệu khoảng cách Canberra and Bray-Curtis thực hiện tốt hơn các số liệu khoảng cách khác.

1.4.2.3 Phản hồi từ người dùng

Sau khi có kết quả tra cứu, người dùng cung cấp phản hồi về các kết quả liên quan hoặc không liên quan. Nếu kết quả chưa được chấp nhận thì vòng lặp phản hồi sẽ được lặp lại nhiều lần cho đến khi người dùng hài lòng.

1.4.3 Các phương pháp tiếp cận phản hồi liên quan

Trong phương pháp tiếp cận dựa trên thông tin phản hồi liên quan, một hệ thống CBIR học từ thông tin phản hồi được cung cấp bởi người sử dụng. Học trong hệ thống CBIR được phân loại thành học ngắn hạn và học dài hạn. Chọn lọc truy vấn sử dụng thông tin phản hồi liên quan đã đạt được nhiều sự chú ý trong nghiên cứu và phát triển của các hệ thống CBIR. Hầu hết các nghiên cứu đã tập trung vào điều chỉnh truy vấn trong mỗi phiên tra cứu. Điều này thường được gọi là học trong nội bộ truy vấn hoặc học ngắn hạn. Ngược lại, liên truy vấn, còn được gọi là học dài hạn là chiến lược cố gắng để phân tích mối quan hệ giữa các phiên tra cứu hiện tại và quá khứ.

1.4.3.1 Phương pháp học ngắn hạn

Trong học ngắn hạn, chỉ những phản hồi của phiên tìm kiếm hiện tại được sử dụng cho thuật toán học, và các đặc trưng ảnh là nguồn dữ liệu chính.

Thách thức chính trong phương pháp này là tìm sự kết hợp tốt nhất các đặc trưng biểu diễn truy vấn của người dùng. Ví dụ một bộ các đặc trưng tối ưu sẽ bao gồm những đặc trưng mà có thể bắt lấy sự tương tự giữa các mẫu dương hoặc những đặc trưng mà có thể phân biệt các mẫu dương và mẫu âm. Do đó nhiều thuật toán học máy cổ điển được sử dụng trong học ngắn hạn như là SVMs, mô hình học Bayes, boosting và đánh trọng số đặc trưng, phân tích sự khác biệt v.v.. Tuy nhiên, cách tiếp cận học ngắn hạn là nhiệm vụ rất khó bởi vì trước hết kích thước của dữ liệu huấn luyện là nhỏ hơn nhiều so với độ dài không gian đặc trưng, thứ hai là có quá nhiều sự mất cân bằng giữa phản hồi của những người dùng khác nhau. Và cuối cùng quá trình học là trực tuyến sẽ đòi hỏi nhiều thời gian thực hơn.

1.4.3.2 Phương pháp học dài hạn

Phương pháp học dài hạn có thể đạt được độ chính xác tra cứu tốt hơn so với các kỹ thuật RF truyền thống. Có thể sử dụng học tập dài hạn để vượt qua những khó khăn như không có khả năng nắm những ngữ nghĩa hiếm hoi và mất cân bằng giữa các ví dụ phản hồi, và thiếu cơ chế bộ nhớ v.v.. Trên thực tế, khái niệm học dài hạn trong CBIR được thông qua từ công việc của lọc cộng tác. Phương pháp học dài hạn sử dụng các thông tin phản hồi thu thập được từ trước. Nó là một quá trình tích lũy cho việc thu thập thông tin phản hồi nhanh chóng và được lưu trữ trong các hình thức của ma trận. Một ma trận lưu trữ các nhãn được cung cấp bởi người dùng cho mỗi hình ảnh trong mỗi lần lặp. Thông thường kích thước của ma trận lịch sử tìm kiếm là lớn, mô hình thống kê và các phương pháp như phân tích thành phần chính và phân tích ngữ nghĩa tiềm ẩn rất phổ biến trong các phương pháp học tập dài hạn. Tuy nhiên, có những vấn đề trong phương pháp học tập dài hạn.

Những hạn chế của phương pháp học dài hạn :

- Trước hết đây là phương pháp thể hiện sự không phù hợp với những ứng dụng mà hình ảnh thường xuyên được thêm vào hoặc gỡ bỏ. Một

cách tiếp cận tốt hơn là sử dụng mô hình véc-tơ đặc trưng và phân tích mối quan hệ liên truy vấn.

- Thứ hai, là sự thưa thớt của thông tin phản hồi được ghi lại. Chất lượng học dài hạn phụ thuộc rất nhiều vào số lượng người dùng đăng nhập mà hệ thống lưu trữ. Do thiếu các tương tác và cơ sở dữ liệu lớn, nó không phải là dễ dàng để thu thập thông tin đăng nhập một cách đầy đủ.

- Cuối cùng, vấn đề khác là hầu hết các giải pháp học dài hạn chỉ giới thiệu các kiến thức ngữ nghĩa được ghi nhớ cho người sử dụng nhưng thiếu khả năng học tập để dự đoán ngữ nghĩa ẩn trong các mẫu ngữ nghĩa thu được.

1.4.4 Những thách thức trong phản hồi liên quan

Kỹ thuật phản hồi liên quan đã đạt được nhiều tiến bộ vượt bậc từ khi nó được giới thiệu vào năm 2007 bởi Liu và các cộng sự. Các phương pháp mới luôn được đưa ra để khắc phục những nhược điểm tồn tại trong nó. Tuy nhiên, với những nhược điểm nguyên thủy của kỹ thuật phản hồi liên quan trong CBIR thì đến nay vẫn còn phải được các nhà khoa học nghiên cứu thêm. Các hạn chế trong phản hồi liên quan của hệ thống CBIR như sau:

- Không thể trích chọn ngữ nghĩa mức cao: Hầu hết các kỹ thuật RF trong CBIR sẽ rất khó để trích chọn ngữ nghĩa mức cao của ảnh khi chỉ có đặc trưng mức thấp được sử dụng trong RF. Tuy nhiên, cách này vẫn hoạt động tốt trong việc tra cứu thông tin văn bản. Bởi vì, việc tra cứu vẫn được dựa trên từ khoá chứ không phải trên các đặc trưng mức thấp.

- Sự khan hiếm và mất cân bằng các mẫu phản hồi: Mỗi người dùng đều không muốn thao tác nhiều hơn số lần lặp phản hồi để có được kết quả tốt nhất. Vì vậy, số lượng mẫu phản hồi gắn nhãn có được từ người dùng trong một phiên RF là khá nhỏ so với chiều không gian đặc trưng.

Do đó, đối với dữ liệu huấn luyện nhỏ thì hầu hết các thuật toán máy học không thể cho ra kết quả chính xác. Hơn nữa, số lượng mẫu có nhãn tiêu cực thường lớn hơn số lượng mẫu có nhãn tích cực. Các dữ liệu huấn luyện mất cân đối luôn luôn làm cho việc học phân lớp ít đáng tin cậy hơn. Vì thế, đối với các mẫu dữ liệu huấn luyện nhỏ mà đặc biệt là các mẫu tích cực thì hiển nhiên sẽ làm giảm độ chính xác của RF.

- Xử lý thời gian thực: Quá trình học trong RF là trực tuyến và do đó mọi vòng lặp phản hồi bao gồm cả huấn luyện và kiểm tra đều phải thực hiện. Vì thế mà hệ thống sẽ tốn rất nhiều thời gian để xử lý. Có một cách hợp lý để giải quyết vấn đề này là sử dụng phương pháp biểu diễn ảnh và cấu trúc lưu trữ như là một cấu trúc cây phân cấp, v.v…

1.5 Các lĩnh vực ứng dụng của tra cứu ảnh dựa trên nội dung

Ứng dụng của tra cứu ảnh dựa trên nội dung có rất nhiều trong đời sống xã hội, phục vụ cho nhiều mục đích khác nhau, nhằm xác nhận, tra cứu thông tin. Nhờ đó mà giảm bớt công việc của con người, nâng cao hiệu suất làm việc, ví dụ như: Album ảnh số của người dùng, ảnh y khoa, bảo tàng ảnh, tìm kiếm nhãn hiệu, mô tả nội dung video, truy tìm ảnh tội phạm, hệ thống tự nhận biết điều khiển luồng giao thông… Một vài hệ thống lớn đại diện cho các lĩnh vực bao gồm :

- Hệ thống truy vấn ảnh theo nội dung (Query By Image Content) được nghiên cứu và phát triển bởi nhóm nghiên cứu Visual Media Management thuộc công ty IBM, đây là một hệ thống tra cứu ảnh thương mại được phát triển từ rất sớm. Hiện nay, hệ thống này hỗ trợ một vài đo độ tương tự cho ảnh như: trung bình màu sắc, lược đồ màu sắc và kết cấu. Công nghệ sử dụng trong hệ thống bao gồm 2 phần chính là: đánh chỉ số và tìm kiếm. Hơn nữa, hệ thống này còn cung cấp vài cách tiếp cận truy vấn theo đơn đặc trưng, đa đặc trưng và đa giai đoạn.

- Hệ thống Visual SEEK tại trường đại học Columbia. Hệ thống cho phép người dùng nhập vào truy vấn, sử dụng các đặc trưng mức thấp của hình ảnh như: màu sắc, bố cục không gian và kết cấu. Các đặc trưng đó được mô tả theo màu sắc và biến đổi Wavelet dựa trên đặc trưng kết cấu.

- Hệ thống NeTra sử dụng các đặc trưng của ảnh: Màu sắc, hình dạng, kết cấu, vị trí không gian.

- Ngoài ra, còn một số hệ thống khác như: Virage system, Stanford SIMPLICity system, NEC PicHunter system, v.v…

CHƯƠNG 2: Mô hình học bán giám sát dựa trên đồ thị

Một trở ngại lớn trong CBIR đó là khoảng cách ngữ nghĩa giữa các đặc trưng mức thấp và các khái niệm bậc cao. Để giảm khoảng cách này, phản hồi liên quan đã được giới thiệu cho CBIR . Hiện nay, rất nhiều nghiên cứu bắt đầu xem xét phản hồi liên quan là một vấn đề phân loại hoặc học tập. Người dùng đưa vào các mẫu dương hoặc mẫu âm, hệ thống sẽ học tập từ những ví dụ đó để phân chia tất cả dữ liệu thành hai nhóm liên quan hoặc không liên quan. Vì vậy đã có rất nhiều đề án học máy cổ điển có thể áp dụng cho phản hồi liên quan.

2.1 Khái niệm học máy

Học máy là một lĩnh vực nhỏ trong ngành khoa học máy tính, được phát triển từ những nghiên cứu về nhận dạng mẫu và lý thuyết học tập tính toán (computational learning theory) trong trí tuệ nhân tạo.

Học máy tìm hiểu và xây dựng các thuật toán để có thể học tập và đưa ra quyết định trên tập dữ liệu (học từ dữ liệu). Các thuật toán này hoạt động bằng cách xây dựng một mô hình từ ví dụ đầu vào để đưa ra các dự đoán và quyết định, chứ không phải là làm theo chỉ dẫn của một chương trình cố định.

Học máy có liên quan chặt chẽ và thường trùng với thống kê tính toán số liệu; một lĩnh vực chuyên về dự đoán. Nó có mối quan hệ mạnh mẽ với tối ưu hóa, trong đó cung cấp các phương pháp, lý thuyết và ứng dụng của lĩnh vực này. Học máy được sử dụng trong một loạt các nhiệm vụ tính toán thiết kế và lập trình mà rõ ràng các thuật toán dựa trên nguyên tắc là không khả thi. Ví dụ bao gồm các ứng dụng lọc thư rác, nhận dạng ký tự quang học (OCR), công cụ tìm kiếm và thị giác máy tính. Học máy đôi khi được lồng việc khai thác dữ liệu, mặc dù đó là lĩnh vực tập trung nhiều hơn vào phân tích dữ liệu. Học máy và nhận dạng mẫu "có thể được xem như là hai mặt của cùng một lĩnh vực."

Nhiệm vụ học máy thường được chia làm 3 loại chính :

- Học không giám sát : Hệ thống học quan sát một tập các mục chưa gán nhãn, mục đích là để tổ chức các mục này. Nhiệm vụ học bao gồm phân chia các nhóm mục vào các cụm, xác định một outliner để quyết định nếu một mục mới x là khác biệt đáng kể so với các mục trước, giảm số chiều ánh xạ x vào một không gian ít chiều mà vẫn giữ được các thuộc tính nhất định của tập dữ liệu.

- Học có giám sát : Hệ thống học quan sát một tập huấn luyện được gán nhãn bao gồm các cặp (đặc trưng, nhãn), được ký hiệu $\{(x_1, y_1) \dots (x_n, y_n)\}$. Mục tiêu là dự đoán nhãn y cho bất kỳ đầu vào mới có đặc trưng x. Một công việc học có giám sát được gọi là hồi quy nếu y Є R, và là phân loại khi y lấy giá trị trên một tập rời rạc.

- Học tăng cường : Hệ thống học liên tục quan sát trong môi trường x, thể hiện một hành động a và nhận lại một phần thưởng r, mục tiêu là chọn các hành động để làm tối đa phần thưởng trong tương lai.

Một cách phân loại theo nhiệm vụ của học máy phát sinh khi xem xét kết quả đầu ra mong muốn của một hệ thống học máy :

- Trong phân loại, đầu vào được chia thành hai hoặc nhiều nhóm, "người học" phải tạo ra một mô hình để gán dữ liệu đầu vào chưa biết vào một hoặc nhiều nhóm đó. Điều này thường giải quyết bằng việc có giám sát. Lọc thư rác là một ví dụ phân loại, trong đó đầu vào là các thông điệp email và đầu ra là "spam" hoặc "không spam".

- Trong hồi quy cũng là một vấn đề có giám sát, kết quả đầu ra thường là liên tục hơn là rời rạc.

- Trong phân cụm, một tập hợp đầu vào được chia nhóm. Khác với phân loại, các nhóm này là chưa được biết trước. Đây thường là nhiệm vụ của học không giám sát.

- Ước tính mật độ tìm phân phối của đầu vào trên một không gian.

- Giảm thiểu số chiều, đơn giản hóa dữ liệu đầu vào bằng cách ánh xạ chúng đến một không gian ít chiều hơn. Mô hình hóa chủ đề là một vấn đề liên quan, khi chương trình được đưa một danh sách các tài liệu bằng ngôn ngữ con người và nhiệm vụ là tìm ra các tài liệu có cùng một chủ đề.

2.2 Học bán giám sát

Trong tài liệu này học máy tập chung vào nhiệm vụ phân loại, theo truyền thống là một nhiệm vụ của học có giám sát. Để huấn luyện một bộ phân loại cần một tập huấn luyện được gán nhãn. Tuy nhiên việc gán nhãn thường là khó, đắt và chậm để thu thập, bởi vì nó có thể đòi hỏi một bộ chú thích có kinh nghiệm của con người. Ví dụ :

- Giám sát bằng hình ảnh : Việc gán nhãn người một cách thủ công trong một lượng lớn các hình ảnh từ camera giám sát là rất tốn thời gian.

- Nhận dạng giọng nói : Việc viết lại chính xác một giọng nói ở mức âm tiết là hết sức tốn thời gian (400xRT) và yêu cầu chuyên gia trong ngôn ngữ học.

- Phân loại văn bản : Lọc thư rác, phân loại tin nhắn, gợi ý các bài viết trên Internet, rất nhiều công việc cần người dùng gán nhãn cho văn bản ví dụ như “thích” hay “không thích”. Phải đọc và gán nhãn hàng ngàn tài liệu sẽ làm nản chí người dùng.

- Phân tích cú pháp : Để huấn luyện một bộ phân tích cú pháp tốt cần những cặp mẫu câu và cây phân tích cú pháp, việc này đòi hỏi rất nhiều

thời gian để xây dựng bởi những nhà ngôn ngữ học. Các chuyên gia phải mất vài năm để xây dựng các cây phân tích cú pháp cho vài nghìn mẫu câu.

Mặt khác, các dữ liệu không có nhãn thường xuyên có sẵn với số lượng lớn và rất dễ thu thập. Các camera quan sát có thể chạy 24 giờ/ngày, các giọng đọc có thể được ghi âm, các văn bản có thể lấy được trên Internet, các mẫu câu thì có ở khắp nơi … Với cách phân loại truyền thống gặp vấn đề là không thể sử dụng các dữ liệu chưa có nhãn để huấn luyện bộ phân loại.

Câu hỏi được đặt ra là : Cho một tập tương đối nhỏ dữ liệu được gán nhãn $\{(x, y)\}$ và một lượng lớn dữ liệu chưa gán nhãn $\{x\}$, có cách nào để sử dụng cả hai cho việc phân loại? Khái niệm "học bán giám sát" được ra đời từ thực tế là các dữ liệu được sử dụng là giữa học có giám sát và học không giám sát. Học bán giám sát sử dụng cả dữ liệu đã gán nhãn và dữ liệu chưa gán nhãn cho mục đích học tập. Học bán giám sát hứa hẹn độ chính xác cao và lỗ lực chú thích thấp nhất.

Chúng ta có cả một chuỗi các ý tưởng thú vị về cách học tập trên cả hai dữ liệu gán nhãn và không gán nhãn. Đây là một lĩnh vực được phát triển một cách nhanh chóng, trong phần này xin trình bày một cách sơ lược về lịch sử của học bán giám sát.

- Thời gian đầu, việc học bán giám sát giả định rằng có 2 lớp, mỗi lớp có một phân bố Gauss. Giả định dữ liệu đầy đủ lấy được từ một mô hình hỗn hợp. Với một lượng lớn các dữ liệu chưa gán nhãn. các thành phần của mô hình hỗn hợp có thể được xác định với thuật toán Expectation Maximization. Chỉ cần một ví dụ có nhãn cho mỗi thành phần để xác định đầy đủ mô hình hỗn hợp. Mô hình này đã áp dụng thành công cho việc phân loại văn bản.

- Một biến thể khác là tự huấn luyện (self-training) : Một bộ phân loại đầu tiên được đào tạo bằng các dữ liệu có nhãn. Sau đó được dùng để phân loại các dữ liệu chưa có nhãn, những điểm chưa gán nhãn mà chắc chắn nhất cùng với các nhãn được dự đoán của nó được thêm vào tập huấn luyện. Bộ phân loại tiếp tục được huấn luyện như trên. Bộ huấn luyện sử dụng chính dự đoán của nó để tự huấn luyện chính nó.

- Đồng huấn luyện (Co-training) : phương pháp này nhằm giảm sai lầm tăng cường nguy hiểm của tự huấn luyện. Nó giả định các đặc trưng có thể chia thành 2 tập con. Mỗi tập con này đủ để huấn luyện một bộ phân loại tốt. Khởi đầu, 2 bộ phân loại được huấn luyện với các dữ liệu gán nhãn, mỗi bộ trên một tập đặc trưng. Hai bộ phân loại sẽ lặp đi lặp lại việc phân loại dữ liệu chưa có nhãn và dạy bộ phân loại kia bằng dự đoán của nó.

- Với sự phổ biến ngày càng tăng của SVMs, TSVMs nổi lên như là một phần mở rộng của chuẩn SVMs cho học bán giám sát. TSVMs tìm một nhãn cho tất cả các dữ liệu chưa gán nhãn và một siêu phẳng phân cách, với phần lề tối đa đạt được trên cả dữ liệu có nhãn và dữ liệu vừa gán nhãn.

- Gần đây phương pháp học bán giám sát dựa trên đồ thị thu hút được rất nhiều sự chú ý. Phương pháp này bắt đầu với một đồ thị mà các nút là các điểm dữ liệu được gán nhãn và không có nhãn, các cạnh (trọng số) phản ánh sự tương tự giữa các nút. Giả thuyết rằng các nút được nối với nhau bằng một cạnh có trọng số lớn thì sẽ có cùng nhãn, các nhãn có thể lan truyền trong đồ thị.

2.3 Học bán giám sát dựa trên đồ thị

2.3.1 Thuật toán lan truyền nhãn

Trong phần này trình bày thuật toán lan truyền nhãn trên đồ thị, thuật toán này xây dựng vấn đề như một dạng lan truyền trong đồ thị. Nhãn của các nút sẽ lan truyền tới các nút xung quanh theo độ gần của chúng. Trong quá trình đó, chúng ta cố định các nhãn trên các dữ liệu đã gán nhãn, vì vậy các dữ liệu được gán nhãn hoạt động giống như những nguồn đẩy các nhãn tới các dữ liệu chưa gán nhãn.

2.3.1.1 Đặt vấn đề

Giả sử có $\{(x_{1,}, y_1) \dots (x_l, y_l)\}$ là tập dữ liệu được gán nhãn với $y \in (1 \dots C)$ và $\{(\mathrm{x_{l+1,}}, \mathrm{y_{l+1}}) \dots (\mathrm{x_{l+u}}, \mathrm{y_{l+u}})\}$ là tập dữ liệu chưa gán nhãn.

Thông thường $l \ll u$, Đặt $n = l + u$. Chúng ta thường xuyên sử dụng L và U để ký hiệu tương ứng cho dữ liệu gán nhãn và chưa gán nhãn. Giả sử số lượng lớp C là đã biết và các lớp này biểu diễn dữ liệu được gán nhãn. Chúng ta tìm nhãn cho các điểm chưa có nhãn.

Một cách trực quan, với mong muốn các điểm dữ liệu tương đồng có cùng một nhãn. Chúng ta tạo một đồ thị mà các nút là các điểm dữ liệu có nhãn và chưa có nhãn. Cạnh giữa 2 nút i, j thể hiện độ tương tự của chúng. Ban đầu giả sử đồ thị có kết nối đầy đủ với trọng số như sau :

$$w_{ij} = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{\alpha^2}\right) \tag{2.1}$$

2.3.1.2 Thuật toán lan truyền nhãn

Các nhãn được lan truyền qua các cạnh, trọng số cạnh càng lớn thì các nhãn càng dễ lan truyền. Ta định nghĩa một ma trận $n \times n$ xác suất chuyển dịch P

$$P_{ij} = P(i \to j) = \frac{w_{ij}}{\sum_{k=1}^{n} w_{ik}} \tag{2.2}$$

P_{ij} là xác suất của dịch chuyển từ nút i đến j. Khai báo một ma trận nhãn kích thước $l \times C$ Y_L, với hàng i xác định véc-tơ y_i với $i \in L : Y_{ic} = \delta(y_i, c)$. Chúng ta sẽ tính toán nhãn mềm f cho các nút, f là một ma trận $n \times C$, các hàng thể hiện phân bố xác suất của các nhãn. Việc khởi tạo f là không quan trọng, thuật toán lan truyền nhãn được trình bày như sau :

- Bước 1 : $f \leftarrow Pf$
- Bước 2 : Bám vào dữ liệu đã gán nhãn $f_L = Y_L$
- Lặp lại bước 1 cho đến khi f hội tụ

Ở bước 1 các nút lan truyền nhãn của nó tới các nút lân cận. Bước 2 là rất quan trọng, chúng ta muốn cố định nguồn nhãn từ dữ liệu được gán nhãn, vì vậy thay vì để chúng mờ dần đi ta gắn chặt với Y_L.

2.3.1.3 Sự hội tụ

Cần chỉ ra rằng thuật toán hội tụ về một giải pháp đơn giản. Đặt $f = \binom{f_L}{f_U}$. Vì f_L được gắn chặt với Y_L, chúng ta chỉ quan tâm đến f_U. Chia P thành các ma trận con :

$$P = \begin{bmatrix} P_{LL} & P_{LU} \\ P_{UL} & P_{UU} \end{bmatrix} \tag{2.3}$$

Thuật toán trở thành :

$$f_U = P_{UU} f_U + P_{UL} Y_L \tag{2.4}$$

Suy ra :

$$f_U = \lim_{n\to\infty}(P_{UU})^n f_U^0 + \left(\sum_{i=1}^{n}(P_{UU})^{(i-1)}\right)P_{UL}Y_L \quad (2.5)$$

Với f_U^0 là giá trị khởi tạo của f_U. Cần chỉ ra $(P_{UU})^n f_U^0 \to 0$

Vì P là ma trận được chuẩn hóa các hàng và P_{UU} là ma trận con P của nên :

$$\exists\gamma < 1, \sum_{j}^{u}(P_{UU})_{ij} \le \gamma, \forall i = 1 \dots u \quad (2.6)$$

Vì vậy :

$$\sum_{j}(\mathrm{P_{UU}})^{n}{}_{ij} = \sum_{j}\sum_{k}(\mathrm{P_{UU}})^{n-1}{}_{ik}(\mathrm{P_{UU}})_{kj} \quad (2.7)$$

$$= \sum_{k}(\mathrm{P_{UU}})^{n-1}{}_{ik}\sum_{j}(\mathrm{P_{UU}})_{kj} \quad (2.8)$$

$$\le \sum_{k}(\mathrm{P_{UU}})^{n-1}{}_{ik}\gamma \quad (2.9)$$

$$\le \gamma^{n} \quad (2.10)$$

Do đó tổng một hàng của $(P_{UU})^n$ hội tụ về 0 có nghĩa là $(P_{UU})^n f_U^0 \to 0$

Vì vậy việc khởi tạo giá trị cho f_U^0 là không quan trọng. Hiển nhiên

$$f_U = (I - P_{UU})^{-1}P_{UL}Y_L \quad (2.11)$$

là một điểm cố định. Vì vậy nó là điểm cố định và nghiệm duy nhất của thuật toán lặp. Chú ý nghiệm này chỉ tồn tại khi $(I - P_{UU})$ là khả nghịch, điều

kiện này được thỏa mãn khi ở những miền liên thông trong đồ thị có ít nhất một nút được gán nhãn.

2.3.2 Xây dựng đồ thị

- Đồ thị đầy đủ : Mỗi cặp nút phân biệt có một cạnh, hai nút giống nhau sẽ có trọng số cạnh lớn. Ưu điểm của đồ thị đầy đủ là việc học trọng số với một hàm khả vi của trọng số, bất lợi của đồ thị đầy đủ là chi phí tính toán cho đồ thị dày đặc. Theo kinh nghiệm chúng ta thấy rằng đồ thị đầy đủ kém hiệu quả hơn đồ thị thưa.

- Đồ thị thưa : Có thể tạo ra đồ thị kNN hoặc ϵNN. Khi đó mỗi nút chỉ kết nối với một vài nút, như vậy việc tính toán sẽ nhanh chóng hơn và có xu hướng có hiệu quả tốt trong thực nghiệm. Các liên kết giả giữa các nút không giống nhau sẽ được loại bỏ. Với đồ thị thưa các cạnh có thể có trọng số hoặc không có trọng số. Bất lợi của đồ thị thưa là trong việc học trọng số, thay đổi siêu tham số sẽ làm thay đổi các nút láng giếng và làm khó khăn việc tối ưu hóa.

    - kNN : Nút i,j được liên kết bằng một cạnh nếu như i nằm trong k láng giềng gần nhất của j. k là tham số điều chỉnh mật độ của đồ thị. Bán kính láng giềng là khác nhau ở những vùng dữ liệu có mật độ thấp và cao.

    - ϵNN : Nút i,j được liên kết bằng một cạnh nếu như khoảng cách $d(i,j) \leq \epsilon$. $\boldsymbol{\epsilon}$ là tham số điều khiển bán kính láng giềng

2.3.3 Trường ngẫu nhiên Gauss và hàm điều hòa

2.3.3.1 Trường ngẫu nhiên Gauss

Trong phần này bài toán lan truyền nhãn được chính thức hóa bằng cơ sở xác suất. Chiến lược ở đây là định nghĩa một trường ngẫu nhiên liên tục trên đồ thị.

Trước tiên chúng ta định nghĩa một hàm thực trên tập các nút $f : L \cup U \rightarrow R$, f có thể âm hoặc lớn hơn 1. Với mong muốn những điểm chưa gán nhãn giống nhau (xác định bằng trọng số cạnh) sẽ có cùng nhãn. Điều này thúc đẩy lựa chọn một hàm năng lượng bậc 2 :

$$E(f) = \frac{1}{2}\sum_{ij} w_{ij}(f(i) - f(j))^2 \tag{2.12}$$

E đạt giá trị nhỏ nhất khi các hàm không đổi. Khi quan sát một số các dữ liệu được gán nhãn, chúng ta cố định f nhận giá trị trên các dữ liệu được gán nhãn. Chúng ta áp dụng một phân bố xác suất lên hàm f bằng một trường ngẫu nhiên Gauss :

$$p(f) = \frac{1}{Z}e^{-\beta E(f)} \tag{2.13}$$

Với β là tham số "nghịch đảo nhiệt độ" và Z là một hàm phân vùng :

$$Z = \int_{f_L = Y_L} \exp\big(-\beta E(f)\big)\, df \tag{2.14}$$

Chúng ta đang quan tâm đến vấn đề suy luận $p(f_i|Y_L)$ với $i \in U$ hoặc giá trị kỳ vọng $\int_{-\infty}^{\infty} f_i p(f_i|Y_L) df_i$. Phân bố $p(f)$ rất giống với một tiêu chuẩn của trường ngẫu nhiên Markov với các trạng thái rời rạc (the Ising model, or Boltzmann machines (Zhu & Ghahramani, 2002b)). Thực tế sự khác biệt duy nhất là việc nới lỏng cho các trạng thái có giá trị thực. Tuy nhiên việc nới lỏng này lại đơn giản hóa vấn đề suy luận. Bởi vì hàm năng lượng bậc hai, $p(f)$ và $p(f_U|Y_L)$ đều là các phân bố Gauss đa biến. Đây là lý do p được gọi là trường ngẫu nhiên Gauss. Phân phối biên $p(f_i|Y_L)$ cũng là phân bố Gauss đơn biến.

2.3.3.2 Đồ thị Laplace

Chúng ta làm quen với một đại lượng quan trọng trong đồ thị : toán tử Laplace Δ. Đặt D là ma trận đường chéo với $D_{ii} = \sum_j w_{ij}$ là bậc của nút i. Ta có :

$$\Delta = D - W \tag{2.15}$$

Tổ hợp Laplace giúp viết ngắn gọn hàm năng lượng. Chúng ta có thể chỉ ra rằng :

$$\boldsymbol{E(f) = \frac{1}{2}\sum_{i,j}(f(i) - f(j))^2 = f^T \Delta f} \tag{2.16}$$

Trường ngẫu nhiên Gauss được viết là :

$$p(f) = \frac{1}{Z} e^{-\beta f^T \Delta f} \tag{2.17}$$

2.3.3.3 Hàm điều hòa

Có thể chỉ ra rằng hàm f mà làm cực tiểu hóa hàm năng lượng :

$f = argmin_{f_L = Y_L} E(f)$ là một hàm điều hòa. Nó thỏa mãn $\Delta f = 0$ trên các nút chưa gán nhãn và nhận giá trị $f_i = y_i, i \in L$ trên các nút đã gán nhãn. Chúng ta sử dụng h để biểu diễn hàm điều hòa này. Theo tính chất của hàm điều hòa ta có giá trị của h tại các nút chưa gán nhãn :

$$h(i) = \frac{1}{D_{ii}} \sum_{j \sim i} w_{ij} h(j), i \in U \tag{2.18}$$

Vì nguyên lý cực đại của hàm điều hòa (Doyle & Snell, 1984), nên h là duy nhất và thỏa mãn $0 \leq h(i) \leq 1, i \in U$.

Để tìm nghiệm của hàm điều hòa này chúng ta chia ma trận trọng số W (tương tự với ma trận $D, \Delta, \ldots$) thành 4 khối :

$$\begin{bmatrix} W_{LL} & W_{LU} \\ W_{UL} & W_{UU} \end{bmatrix}$$

Nghiệm của $\Delta h = 0$ với $h_L = Y_L$là :

$$h_U = (D_{UU} - W_{UU})^{-1} W_{UL} Y_L \quad (2.19)$$

$$= -(\Delta_{UU})^{-1} \Delta_{UL} Y_L \quad (2.20)$$

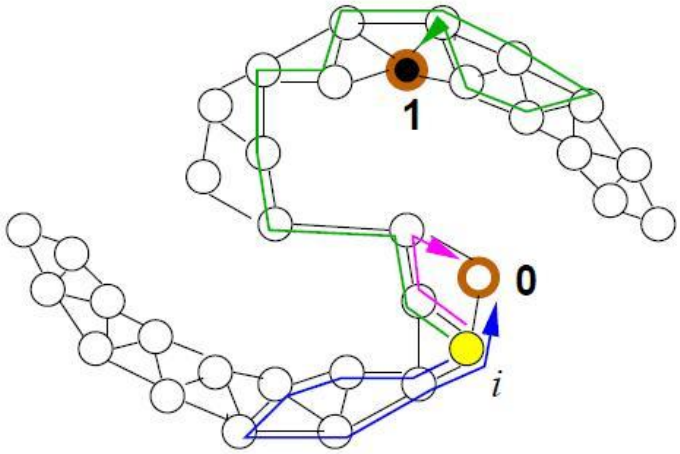
$$= (\boldsymbol{I} - \boldsymbol{P_{UU}})^{-1} \boldsymbol{P_{UL}} \boldsymbol{Y_L} \quad (2.21)$$

Trong biểu thức (2.21) giống với biểu thức (2.11) với $P = D^{-1}W$ là ma trận quá trình chuyển đổi. Bài toán lan truyền nhãn thực tế đã tính toán hàm điều hòa.

2.3.3.4 Giải thích và liên tưởng

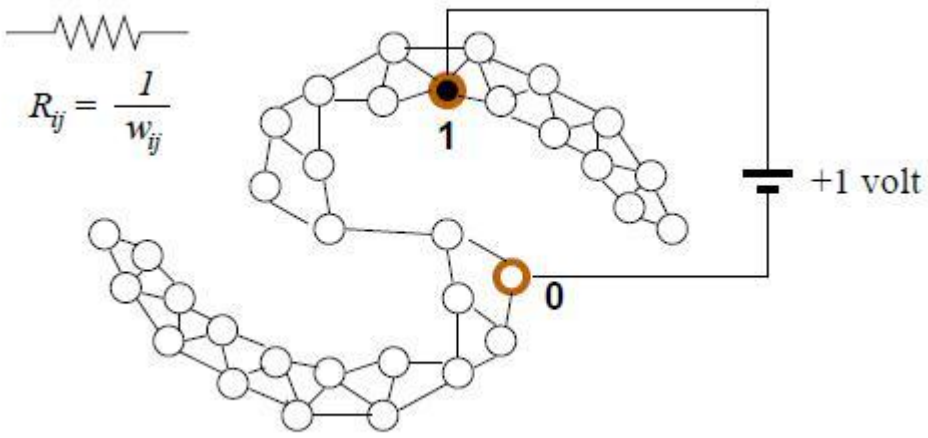
Các hàm điều hòa có thể được xem xét theo một số cách cơ bản khác nhau, và những cách nhìn khác nhau cung cấp một tập các kỹ thuật phong phú và bổ trợ lẫn nhau cho lý luận về cách tiếp cận này đối với vấn đề học bán giám sát.

- Bước ngẫu nhiên : Giả sử có một bước ngẫu nhiên trên đồ thị.Ta bắt đầu từ một nút chưa gán nhãn, di chuyển tới nút lân cận j với xác suất P_{ij} sau mỗi bước. Hàm $h(i)$ chính là xác suất để bước ngẫu nhiên đó xuất phát từ nút i gặp một nút được gán nhãn 1. Ở đây các nút gán nhãn được xem xét như là một "ranh giới hấp thụ" của bước ngẫu nhiên.



Hình 2.1 - Hàm điều hòa và bước ngẫu nhiên trên đồ thị

- Mạng điện tử : Ta có thể xem xét đồ thị như một mạng điện tử, các cạnh của đồ thị có điện trở với độ dẫn điện là W, như vậy điện trở kháng giữa 2 nút i, j là $1/w_{ij}$. Chúng ta nối các nút được gán nhãn dương với một nguồn $+1$ vôn, và các nút được gán nhãn âm với đất. Sau đó hàm h_U chính là kết quả điện thế của mạng điện trên các nút chưa gán nhãn. Hơn nữa hàm h_U sẽ cực tiểu nhiệt lượng thoát ra của mạng điện. Năng lượng nhiệt đó chính là $E(h)$ như trong biểu thức (2.16).



Hình 2.2 - Hàm điều hòa và đồ thị mạng điện tử

2.4 Kết hợp học bán giám sát với học chủ động (Active Learning)

Trong phần này chúng ta sẽ xem qua vấn đề về học chủ động. Chúng ta sẽ kết hợp học bán giám sát và học chủ động một cách tự nhiên và hiệu quả.

Cho đến nay, các dữ liệu có nhãn giả định đã được đưa ra và cố định. Trong thực tế, việc sử dụng học chủ động kết hợp với việc học bán giám sát là có ý nghĩa, có thể cho phép các thuật toán học để chọn các thực thể chưa có nhãn được dán nhãn bởi chuyên gia. Các chuyên gia trả về nhãn, mà sau đó sẽ được sử dụng như (hoặc để làm tăng thêm) tập hợp dữ liệu có nhãn. Nói cách khác, nếu chúng ta phải gắn nhãn một vài trường hợp cho việc học bán giám sát, sẽ hấp dẫn hơn khi để cho các thuật toán học cho chúng ta biết trường hợp cần dán nhãn, chứ không phải là lựa chọn chúng một cách ngẫu nhiên. Chúng ta sẽ giới hạn phạm vi lựa chọn truy vấn đến các tập dữ liệu không có nhãn, một thực tế được gọi là học chủ động hoặc lấy mẫu chọn lọc. Hiện đã có rất nhiều nghiên cứu trong học chủ động. Ví dụ, Tong và Koller (2000) truy vấn chọn để giảm thiểu kích thước không gian phiên bản cho máy vector hỗ trợ; Cohn et al. (1996) giảm thiểu các thành phần phương sai của sai số ước tính tổng quát; Freund et al. (1997) sử dụng một ủy ban phân loại, và truy vấn một điểm bất cứ khi nào các thành viên ủy ban không đồng ý. Hầu hết các phương pháp học tập tích cực không tận dụng hơn nữa của số lượng lớn các dữ liệu không có nhãn khi truy vấn được lựa chọn. Các tác phẩm của McCallum và Nigam (1998b) là một ngoại lệ, EM với dữ liệu không có nhãn được tích hợp vào học tập tích cực. Ngoại lệ khác là (Muslea et al., 2002), trong đó sử dụng một phương pháp học bán giám sát trong quá trình đào tạo. Từ những hoạt động của cộng đồng học máy, có một lượng lớn tài liệu về các chủ đề liên quan chặt chẽ của thiết kế thí nghiệm về thống kê; Chaloner và Verdinelli (1995) đưa ra một cuộc khảo sát về thực nghiệm thiết kế từ một quan điểm Bayesian.

Nền tảng trường ngẫu nhiên Gaussian và hàm điều hòa cho phép một sự kết hợp tự nhiên của học chủ động và học bán giám sát. Nói tóm lại, nền tảng này cho phép ước tính một cách hiệu quả các lỗi tổng quát dự kiến sau khi truy vấn một điểm, dẫn đến một tiêu chí lựa chọn tốt hơn so với một cách ngây chọn điểm có nhãn tối đa sự mơ hồ. Sau đó, một khi các truy vấn được lựa chọn và thêm vào tập dữ liệu được dán nhãn, phân loại có thể được đào tạo sử dụng cả các dữ liệu có nhãn và phần còn lại các dữ liệu chưa có nhãn. Hạn chế tối đa các lỗi tổng quát ước tính lần đầu tiên được đề xuất bởi Roy and McCallum (2001).

Chúng ta thực hiện học tập chủ động với mô hình trường ngẫu nhiên Gauss bằng cách tham lam chọn truy vấn từ các dữ liệu không có nhãn để giảm thiểu các nguy cơ của hàm tối thiểu năng lượng hài hòa. Nguy cơ là lỗi tổng quát ước tính của các phân lớp Bayes, và có thể được tính bằng phương pháp ma trận. Nguy cơ thực sự R(h) của phân loại Bayes dựa trên các hàm điều hòa h là :

$$R(h) = \sum_{i=1}^{n} \sum_{y_i=0,1} [sgn(h_i) \neq y_i] p^*(y_i)$$

Trong đó $sgn(h_i)$ là luật quyết định Bayes với ngưỡng 0.5, $p^*(y_i)$ là phân bố nhãn thực tế chưa biết tại nút i. Chính vì vậy mà $R(h)$ là không thể tính được, để xử lý chúng ta cần thiết phải có một giả định. Bước đầu chúng ta ước tính phân bố chưa biết $p^*(y_i)$ với giá trị kỳ vọng của mô hình Gauss.

$$p^*(y_i = 1) \approx h_i$$

Một cách trực giác, gọi h_i là xác suất tiến đến 1 trong bước ngẫu nhiên trên đồ thị.

Với giả định này ta có thể tính giá trị của nguy cơ ước tính là :

$$R(h) = \sum_{i=1}^{n} [sgn(h_i) \neq 0](1 - h_i) + [sgn(h_i) \neq 1]h_i$$
$$= \sum_{i=1}^{n} \min(1 - h_i, h_i) \quad (2.22)$$

Nếu chúng ta thực hiện học chủ động và truy vấn một nút chưa gán nhãn k chúng ta nhận được câu trả lời y_k(0 hoặc 1). Thêm điểm này vào tập huấn luyện và huấn luyện lại thì trường Gauss và hàm kỳ vọng của nó tất nhiên sẽ thay đổi, chúng ta ký hiệu hàm điều hòa mới là $h^{+(x_k,y_k)}$.

Nguy cơ ước tính sẽ thay đổi :

$$R\left(h^{+(x_k,y_k)}\right) = \sum_{1}^{n} \min(1 - h_i^{+(x_k,y_k)}, h_i^{+(x_k,y_k)})$$

Vì chúng ta không biết câu trả lời y_k sẽ nhận được, một lần nữa chúng ta lại giả định xác suất nhận được câu trả lời $p^*(h_k = 1)$ xấp xỉ bằng h_k. Vì vậy nguy cơ ước tính kỳ vọng sau khi truy vấn nút k là :

$$R\left(h^{+x_k}\right) = (1 - h_k)R\left(h^{+(x_k,0)}\right) + h_k R(h^{+(x_k,1)})$$

Tiêu chí của học chủ động chúng ta sử dụng ở đây là tìm điểm truy vấn k làm tối thiểu giá trị kỳ vọng của nguy cơ ước tính :

$$k = argmin_{k'} R(h^{+x_k}) \quad (2.23)$$

Để thực hiện thủ tục này, chúng ta cần phải tính toán hàm điều hòa $h^{+(x_k,y_k)}$ sau khi thêm (x_k, y_k) vào tập huấn luyện có nhãn hiện thời. Đây là vấn đề đào tạo lại và tính toán chuyên sâu nói chung. Tuy nhiên đối với các trường Gaussian và các hàm điều hòa, có một cách hiệu quả để đào tạo lại. Nhớ lại rằng các giải pháp hàm điều hòa là :

$$h_U = -\Delta_{UU}^{-1} \Delta_{UL} Y_L$$

Giải pháp là gì nếu như gán giá trị y_k cho nút k? Điều này giống như tìm phân bố có điều kiện cho các nút chưa gán nhãn khi biết giá trị y_k . Trong các trường Gauss thì phân bố này trên các dữ liệu chưa gán nhãn là một phân bố chuẩn đa biến $N(h_U, \Delta_{UU}^{-1})$. Một kết quả tiêu chuẩn cho biết giá trị kỳ vọng của điều kiện khi cố định y_k :

$$h_U^{+(x_k,y_k)} = h_U + (y_k - h_k)\frac{\Delta_{UU}^{-1}{}_{.k}}{\Delta_{UU}^{-1}{}_{kk}}$$

Trong đó $\Delta_{UU}^{-1}{}_{.k}$ là cột thứ k của ma trận nghịch đảo Laplace trên dữ liệu chưa gán nhãn, $\Delta_{UU}^{-1}{}_{kk}$ là phần tử thứ k của đường chéo chính cũng trên ma trận đó, cả hai đã được tính toán trước đó khi tính toán hàm điều hòa.

Tóm lại thuật toán học chủ động được biểu diễn trong hình 5.1. Độ phức tạp thời gian để tìm ra truy vấn tốt nhất là $O(n^2)$. Và sau cùng để tính toán hiệu quả, chú ý rằng khi thêm truy vấn và câu trả lời của nó vào L, ở bước lặp tiếp theo cần tính $(\Delta_{UU})_{-k}{}^{-1}$, nghịch đảo của ma trận Laplace trên dữ liệu chưa gán nhãn sau khi bỏ đi hàng/cột ứng với x_k. Thay vì lấy nghịch đảo, có những thuật toán hiệu quả để tính nó từ $(\Delta_{UU})^{-1}$.

Input : L , U , W

While : cần thêm dữ liệu gán nhãn

- Tính h sử dụng (2.4.11)
- Tìm truy vấn k tốt nhất sử dụng (2.5.2)
- Truy vấn x_k và câu trả lời y_k
- Thêm (x_k, y_k) vào L , loại bỏ x_k khỏi U

End

Output : L và h

Hình 2.3 – Thuật toán học chủ động

2.5 Học siêu tham số của đồ thị (Graph Hyperparameter Learning)

Trước đây giả thiết ma trận trọng số W đã được cho và cố định. Trong phần này trình bày sơ lược một số phương pháp để học trọng số từ các dữ liệu có nhãn và chưa có nhãn.

Giả thiết rằng các cạnh của đồ thị có trọng số được tham số hóa bởi các siêu tham số $= \{\sigma_1, \sigma_2, \dots, \sigma_D\}$:

$$w_{ij} = \exp\left(-\sum_{d=1}^{D} \frac{\left(x_{i,d} - x_{j,d}\right)^2}{\sigma_d^2}\right) \tag{2.24}$$

2.5.1 Phương pháp tối đa Evidence

Để học các siêu tham số trong tiến trình Gauss có thể chọn các siêu tham số làm tối đa log likelihood : $\Theta^* = \arg\ max_\Theta \log \mathrm{p}\,(y_L|\Theta).\ \log p(y_L|\Theta)$ được biết như là evidence và thủ tục này gọi là làm tối đa evidence. Có thể giả định một xác suất tiền nghiệm trên Θ và tìm một xác suất hậu nghiệm lớn nhất (MAP) để ước tính :

$$\Theta^* = \arg\max_\Theta \log p(y_L|\Theta) + \log p(\Theta)$$

Evidence có thể là đa mode và thường sử dụng phương pháp gradient để tìm một mode trong không gian siêu tham số. Điều này đòi hỏi tính đạo hàm $\partial \log p(\,y_L|\Theta)/\partial\Theta$. (Được tính cụ thể trong phụ lục D tài liệu tham khảo [6]).

2.5.2 Phương pháp tối thiểu Entropy

Một cách khác, có thể chọn entropy nhãn trung bình như một tiêu chí cho việc học tham số. Việc này chỉ sử dụng hàm điều hòa và không phụ thuộc vào tiến trình Gauss.

Entropy nhãn trung bình $H(h)$ của hàm điều hòa h được định nghĩa là :

$$H(h) = \frac{1}{u}\sum_{l+1}^{n} H_i(h_i) \tag{2.25}$$

Với $H_i(h_i) = -h_i \log h_i - (1-h_i)\log(1-h_i)$ là entropy Shannon của riêng điểm dữ liệu chưa gán nhãn i.

Vì $0 \leq h_i \leq 1, i \in U$ nên entropy nhỏ nghĩa là h_i gần 0 hoặc 1. Điều này cho phép đảm bảo một ma trận trọng số tốt (tương đương với một tập siêu tham số tốt) sẽ cho kết quả chắc chắn hơn trong việc gán nhãn. Tất nhiên có rất nhiều nhãn tùy ý có entropy thấp, và tiêu chí này có thể không hoạt động. Tuy nhiên điều quan trọng cần chỉ ra rằng h được giới hạn trên tập dữ liệu được gán nhãn, và hầu hết các nhãn có entropy thấp là không phù hợp với việc giới hạn này. Thực tế, không gian để có nhãn entropy thấp có thể đạt được bằng hàm điều hòa là nhỏ, và có thể nhờ nó để điều chỉnh các siêu tham số.

Giả sử trọng số đồ thị được tham số hóa như (2.24), sử dụng phương pháp giảm gradient để tìm các siêu tham số δ_d làm tối thiểu H. Gradient được tính là :

$$\frac{\partial H}{\partial \delta_d} = \frac{1}{u} \sum_{i=l+1}^{l+u} \log\left(\frac{1-h_i}{h_i}\right) \frac{\partial h_i}{\partial \delta_d} \tag{2.26}$$

ở đây giá trị $\partial h_i / \partial \delta_d$ có thể lấy ra được từ $\partial h_U / \partial_{\delta_d}$ được cho bởi :

$$\frac{\partial h_U}{\partial \delta_d} = (I - P_{UU})^{-1}\left(\frac{\partial P_{UU}}{\partial \delta_d} h_U + \frac{\partial P_{UL}}{\partial \delta_d} Y_L\right) \tag{2.27}$$

Sử dụng $dX^{-1} = X^{-1}(dX)X^{-1}$. Cả hai $\partial P_{UU}/\partial \delta_d$ và $\partial P_{UL}/\partial \delta_d$ là ma trận con của ma trận $\partial P/\partial \delta_d$.Vì P được tạo bằng cách chuẩn hóa ma trận trọng số W nên ta có :

$$\frac{\partial p_{ij}}{\partial \delta_d} = \frac{\frac{\partial w_{ij}}{\partial \delta_d} - p_{ij} \sum_{k=1}^{l+u} \frac{\partial w_{ik}}{\partial \delta_d}}{\sum_{k=1}^{l+u} w_{ik}} \tag{2.28}$$

Và cuối cùng $\frac{\partial w_{ij}}{\partial \delta_d} = 2 w_{ij} (x_{di} - x_{dj})^2 / \delta_d^3$

CHƯƠNG 3: Áp dụng cài đặt thử nghiệm

3.1 Cài đặt

3.1.1 Nền tảng và ngôn ngữ lập trình

Chương trình được cài đặt trên môi trường Microsoft Visual Studio 2012 với ngôn ngữ C#.

3.1.2 Các thư viện sử dụng

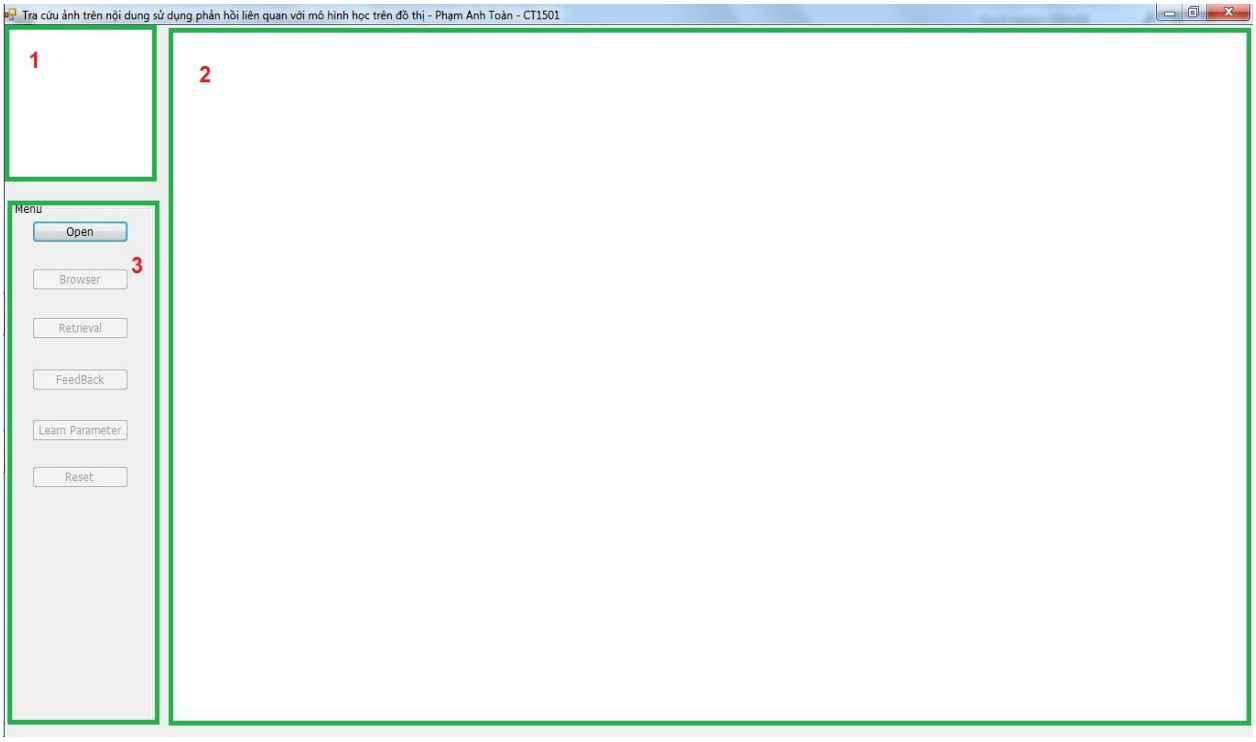
- Để trích chọn đặc trưng ảnh, chương trình sử dụng thư viện FElib.
    - Đặc trưng của ảnh được biểu diễn bởi một vector 809 phần tử:
    - Color histogram, color moments từ phần tử từ 1 đến 81
    - Edge histogram từ 82 đến 118
    - Gabor wavelets transform: các phần tử từ 119 đến 238
    - Local Binary Pattern: các phần tử từ 239 đến 297
    - GIST: các phần tử từ 297 đến 809
- Để hỗ trợ tính toán chương trình sử dụng gói thư viện BLAS/LAPACK.

3.1.3 Cơ sở dữ liệu

Cơ sở dữ liệu bao gồm 2345 ảnh lấy từ cơ sở dữ liệu Corel bao gồm 23 nhóm, mỗi nhóm có khoảng 100 ảnh. Các nhóm này bao gồm nhiều ví dụ từ đơn giản đến phức tạp (về màu sắc và chi tiết).

3.2 Giao diện và các chức năng chính của chương trình

3.2.1 Giao diện chính



Hình 3-1: Giao diện chính chương trình

- Vùng 1 : Hiển thị ảnh truy vấn
- Vùng 2 : Kết quả truy vấn
- Vùng 3 : Menu chức năng chương trình

3.2.2 Các chức năng chính của chương trình

3.2.2.1 Mở ảnh truy vấn và chọn cơ sở dữ liệu truy vấn

- Open : Mở file ảnh truy vấn và trích chọn đặc trưng cho ảnh truy vấn.
- Browser : Mở thư mục cơ sở dữ liệu ảnh

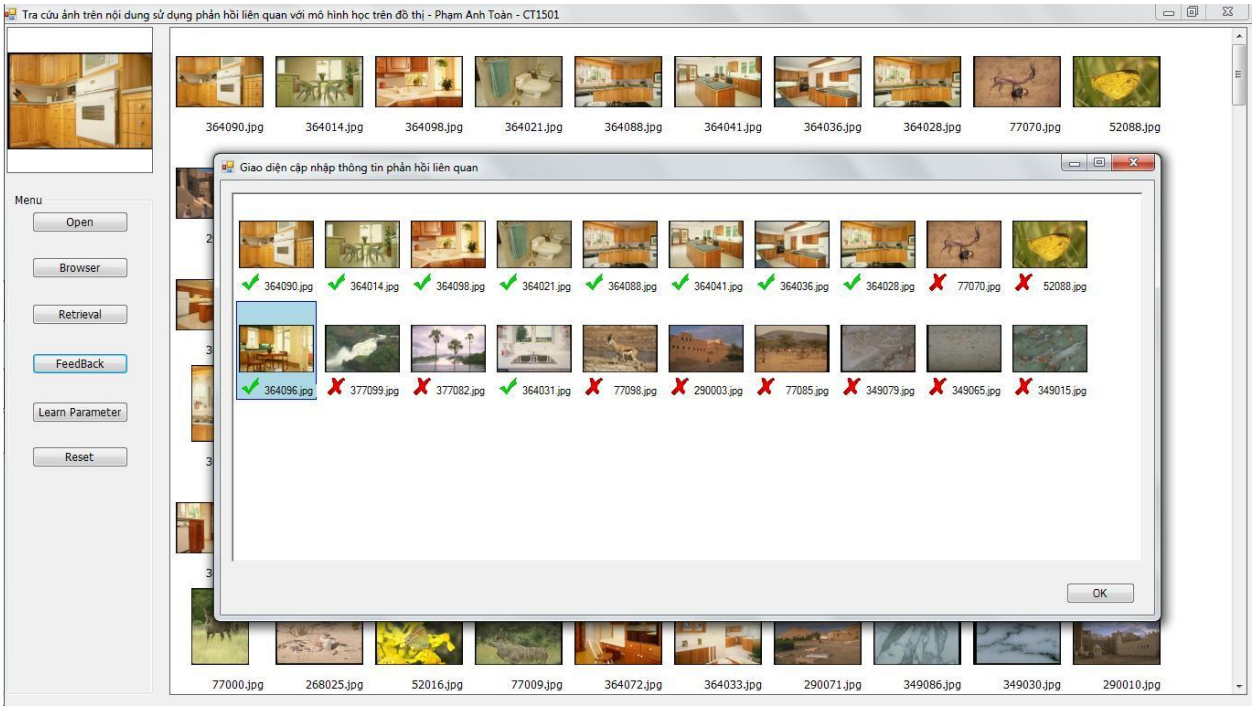
3.2.2.2 Hiển thị kết quả truy vấn

- Retrieval : Hiển thị kết quả tra cứu. Nếu người dùng chưa hài lòng có thể tiếp tục quá trình phản hồi liên quan.

3.2.2.3 Phản hồi liên quan

- FeedBack : Mở giao diện lấy thông tin phản hồi liên quan từ người sử dụng.

- Ban đầu hệ thống đưa cho người dùng 20 ảnh để gán nhãn. Sau đó tại mỗi vòng lập, hệ thống sử dụng thuật toán học chủ động để đưa ra một vài ảnh.



Hình 3-2 : Giao diện lấy thông tin phản hồi liên quan

3.2.2.4 Học tham số cho đồ thị

- Learn Param : Thực hiện quá trình học tham số cho dữ liệu hiện tại.

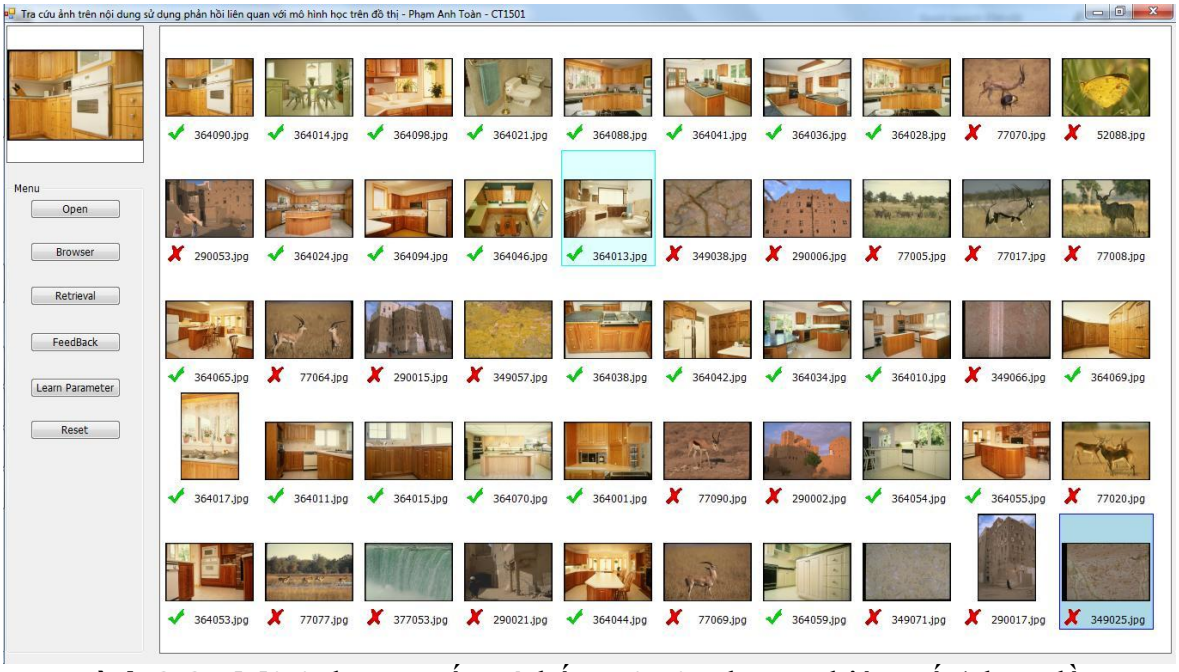
3.2.2.5 Khởi tạo lại quá trình truy vấn.

- Reset : Thiết lập lại quá trình truy vấn. Người dùng có thể chọn ảnh truy vấn khác.

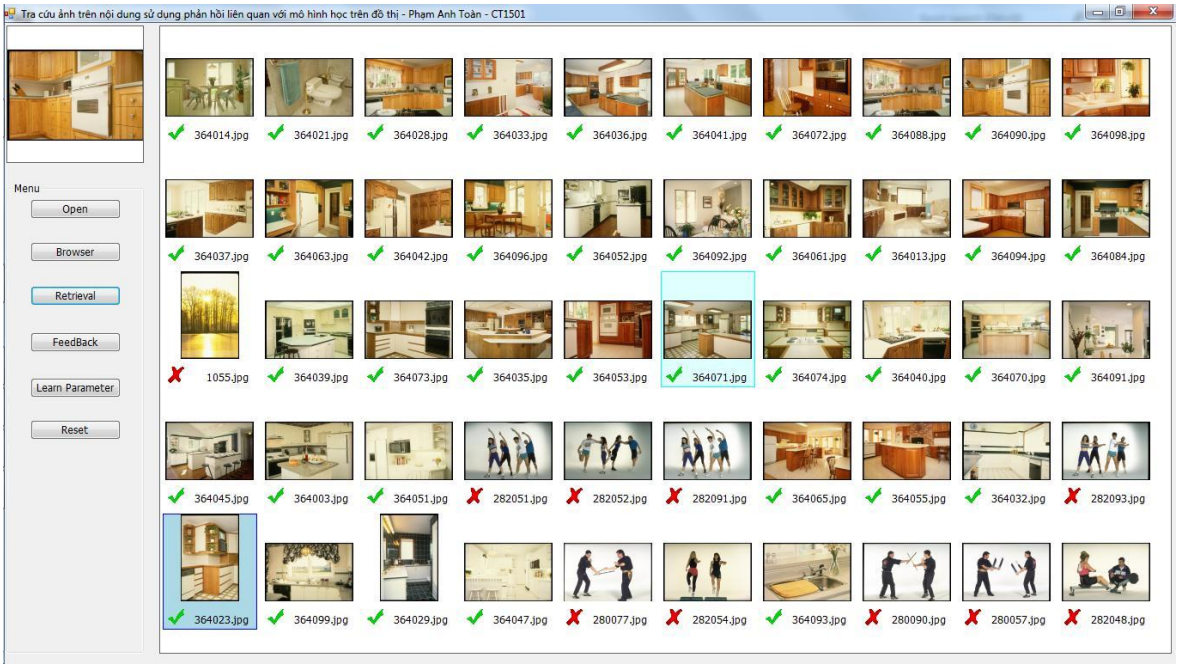
3.3 Một số kết quả thực nghiệm

Tiến hành thử nghiệm với hai ảnh truy vấn khác nhau.

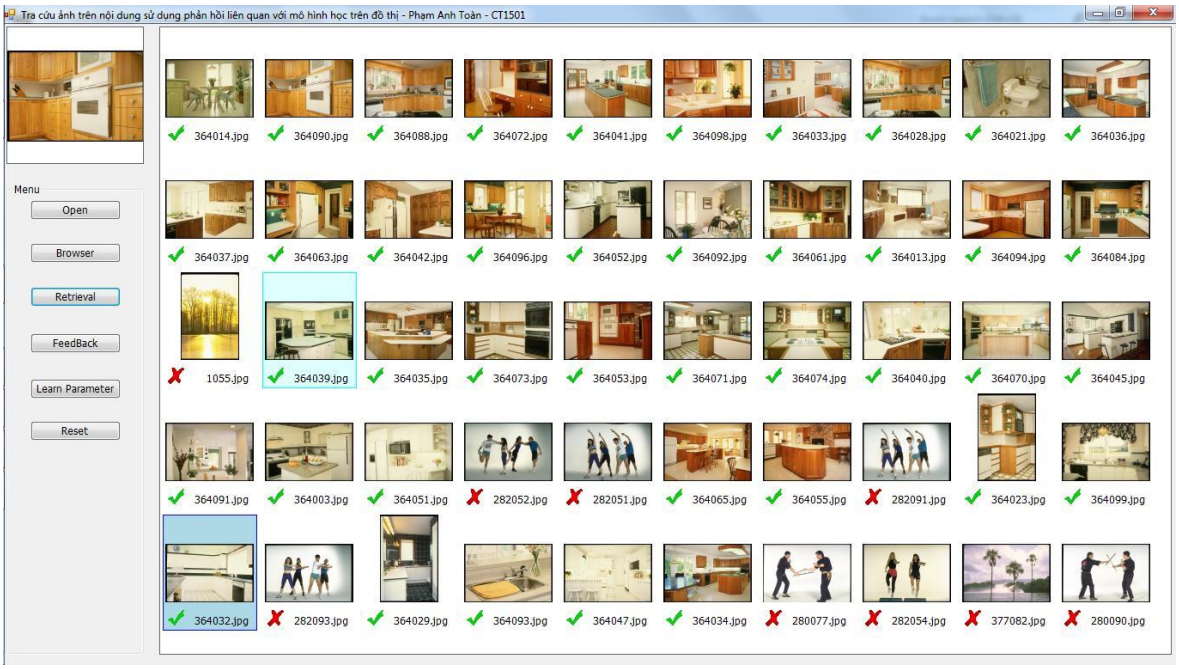
3.3.1 Kết quả thực nghiệm số 1



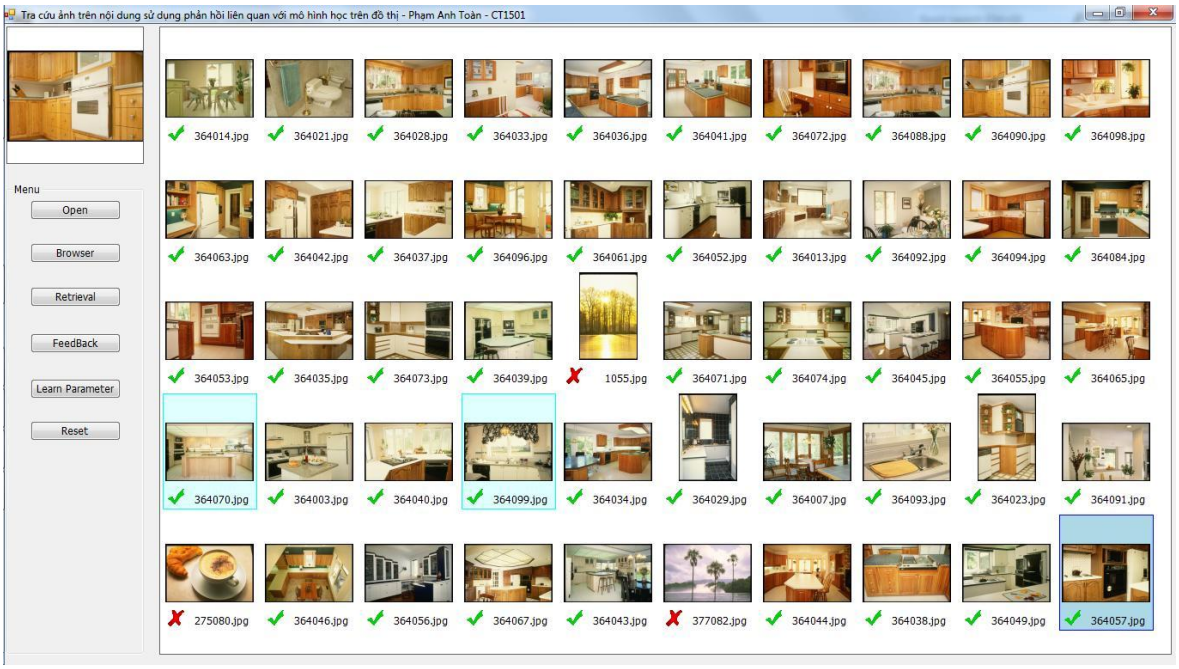
Hình 3-3 : Mở ảnh truy vấn và kết quả của thực nghiệm số 1 ban đầu



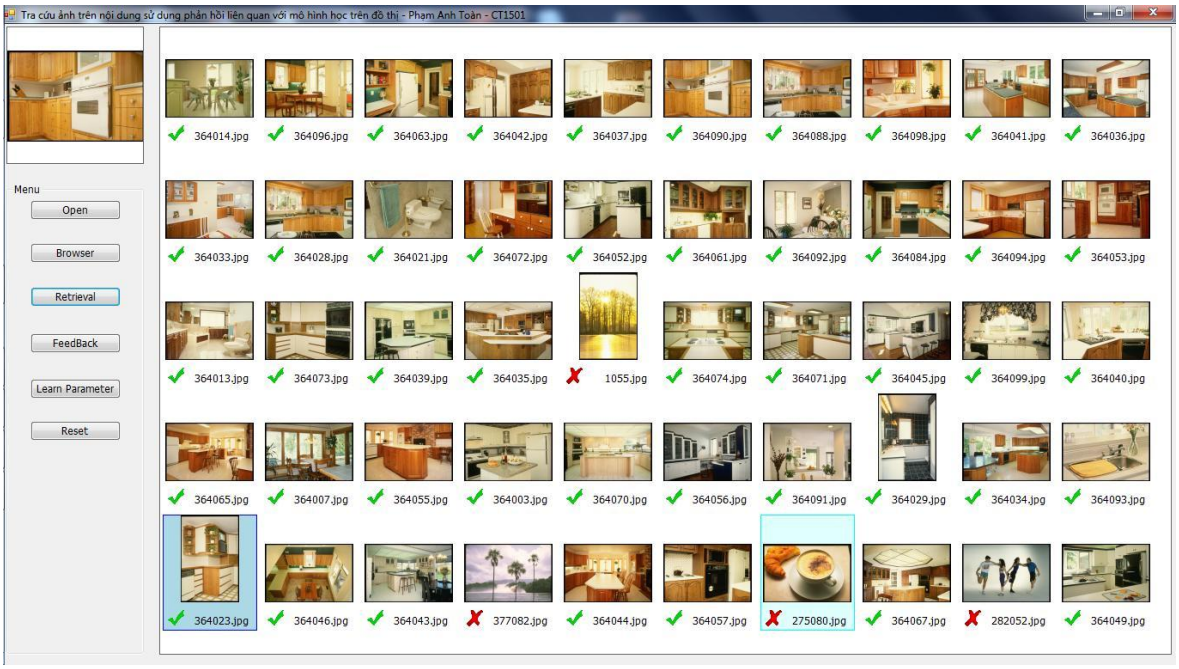
Hình 3-4 : Kết quả của thực nghiệm số 1 sau lần phản hồi thứ nhất



Hình 3-5 : Kết quả của thực nghiệm số 1 sau lần phản hồi thứ 2

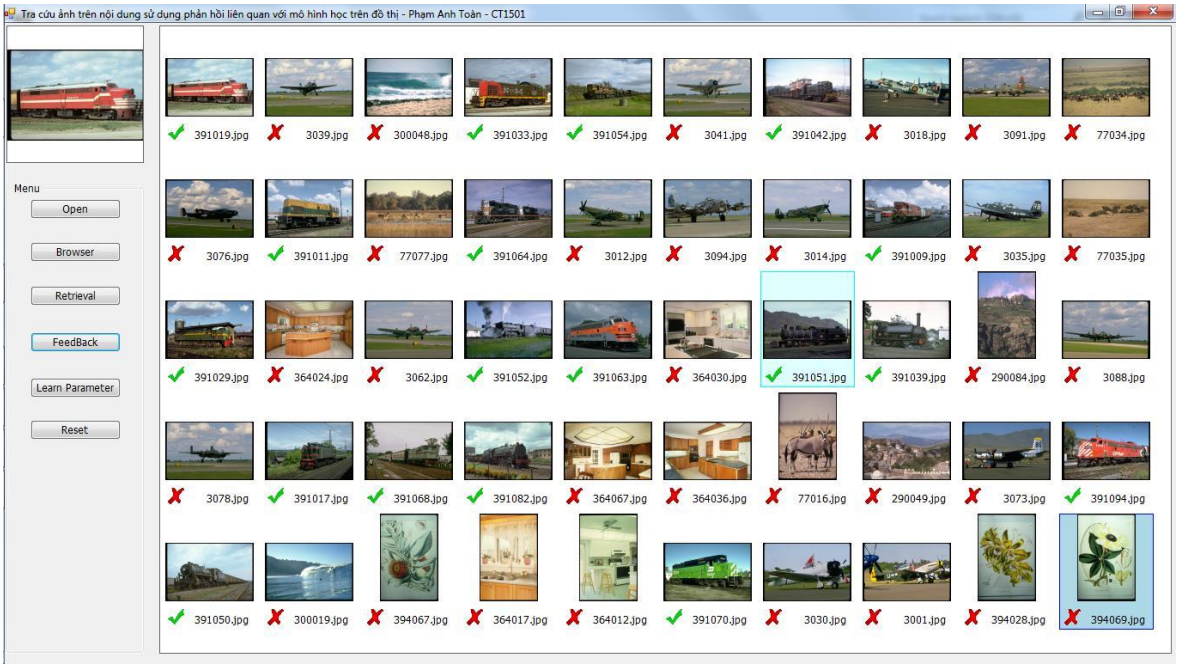


Hình 3-6 : Kết quả của thực nghiệm số 1 sau phản hồi lần 3

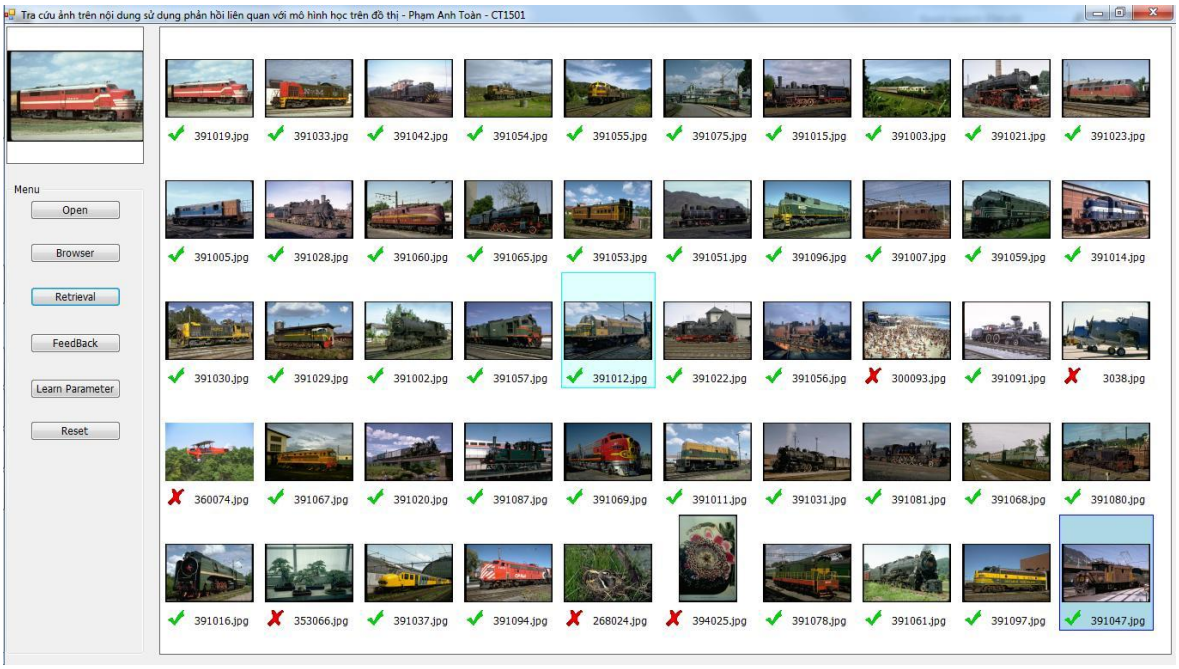


Hình 3-7: Kết quả của thực nghiệm số 1 sau lần phản hồi thứ 4

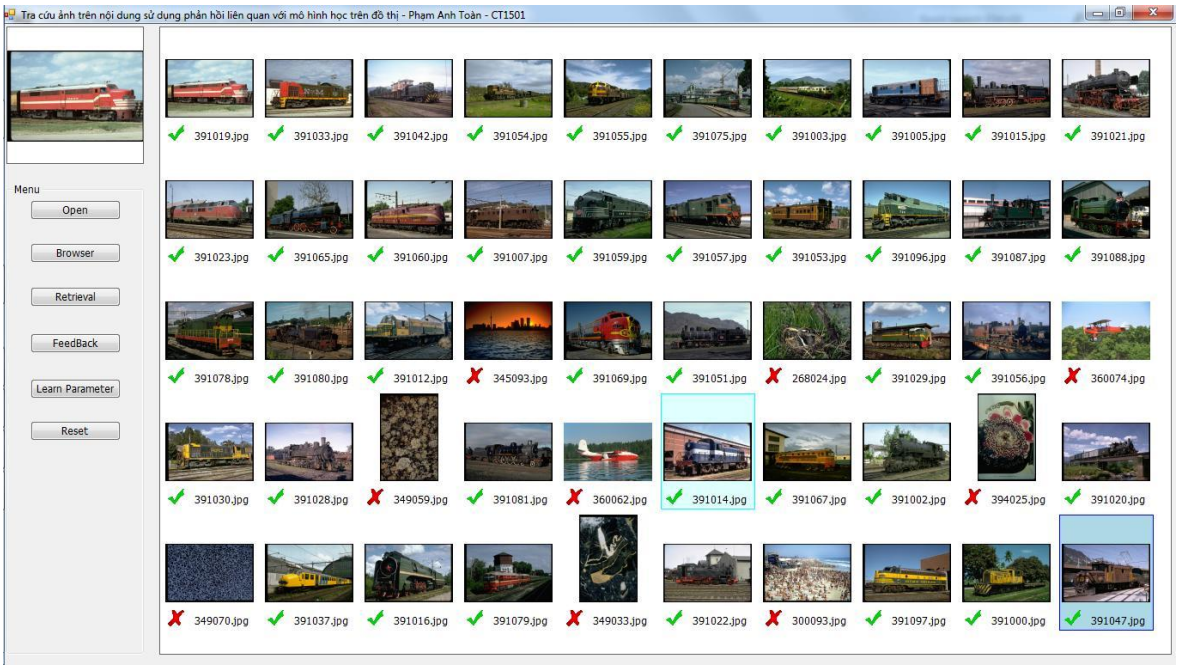
3.3.2 Kết quả thực nghiệm số 2



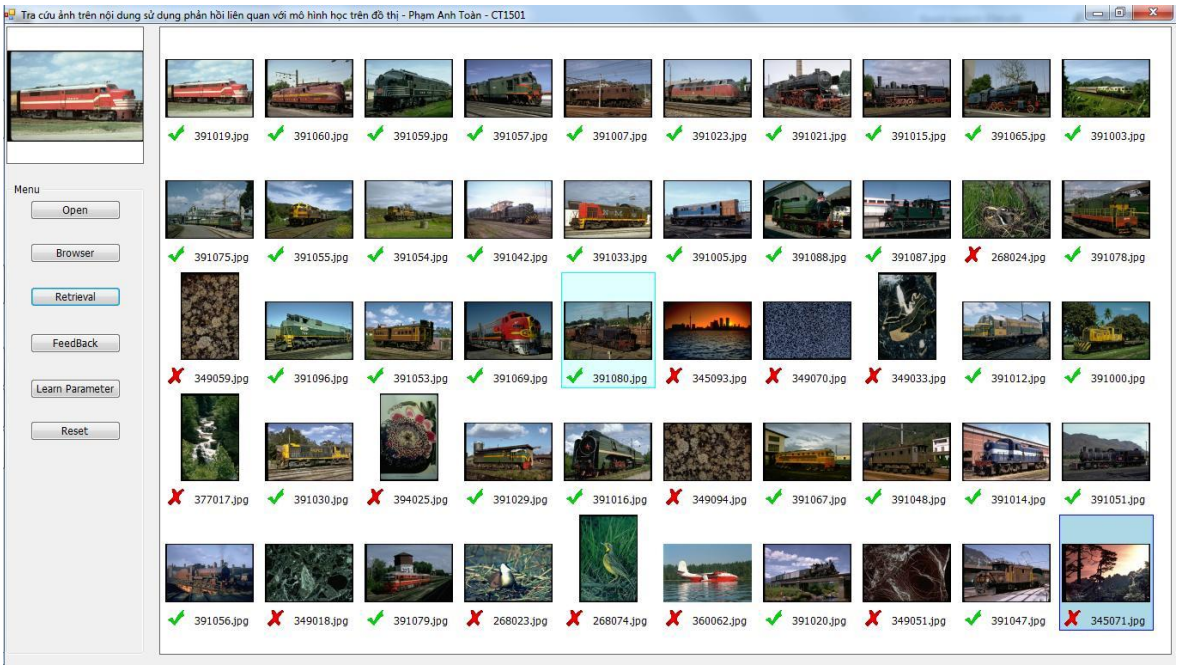
Hình 3-8: Mở ảnh truy vấn và kết quả của thực nghiệm số 2 ban đầu



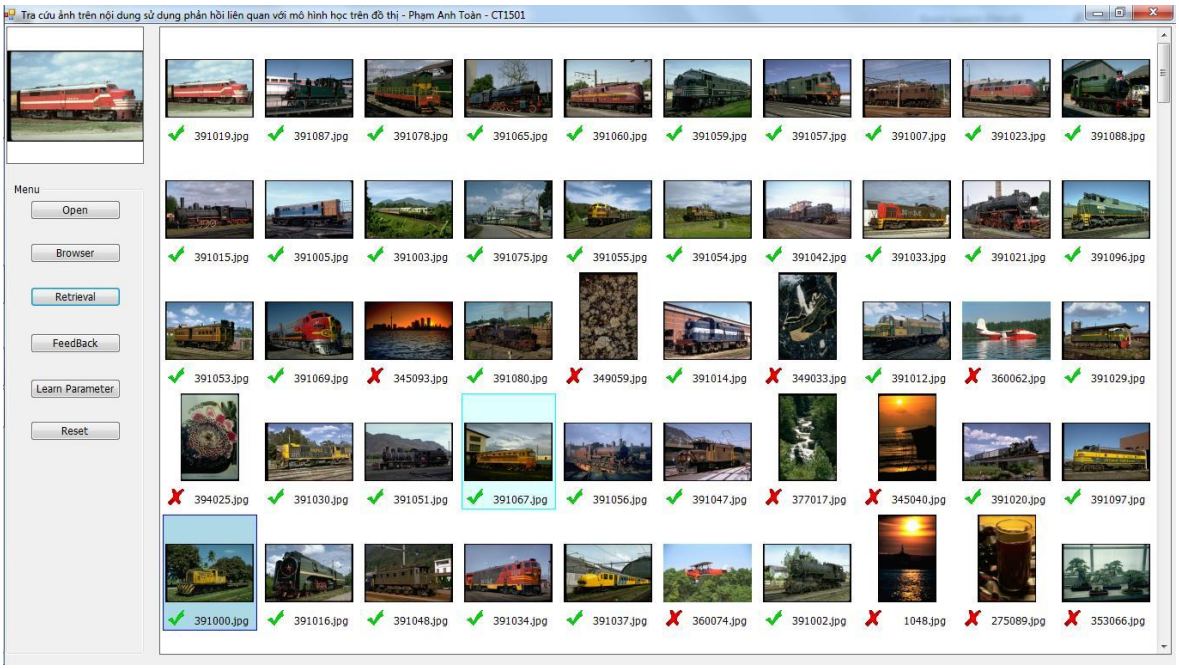
Hình 3-9: Kết quả số 2 sau lần phản hồi thứ nhất



Hình 3-10: Kết quả số 2 sau lần phản hồi thứ 2

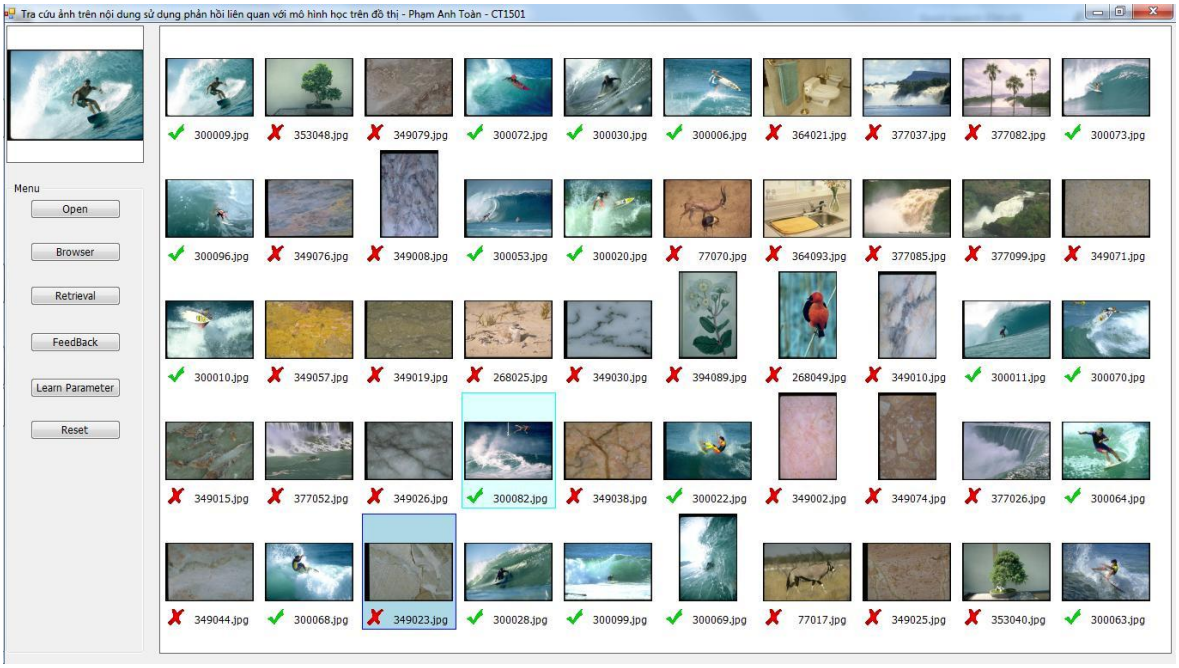


Hình 3-11: Kết quả số 2 sau lần phản hồi thứ 3

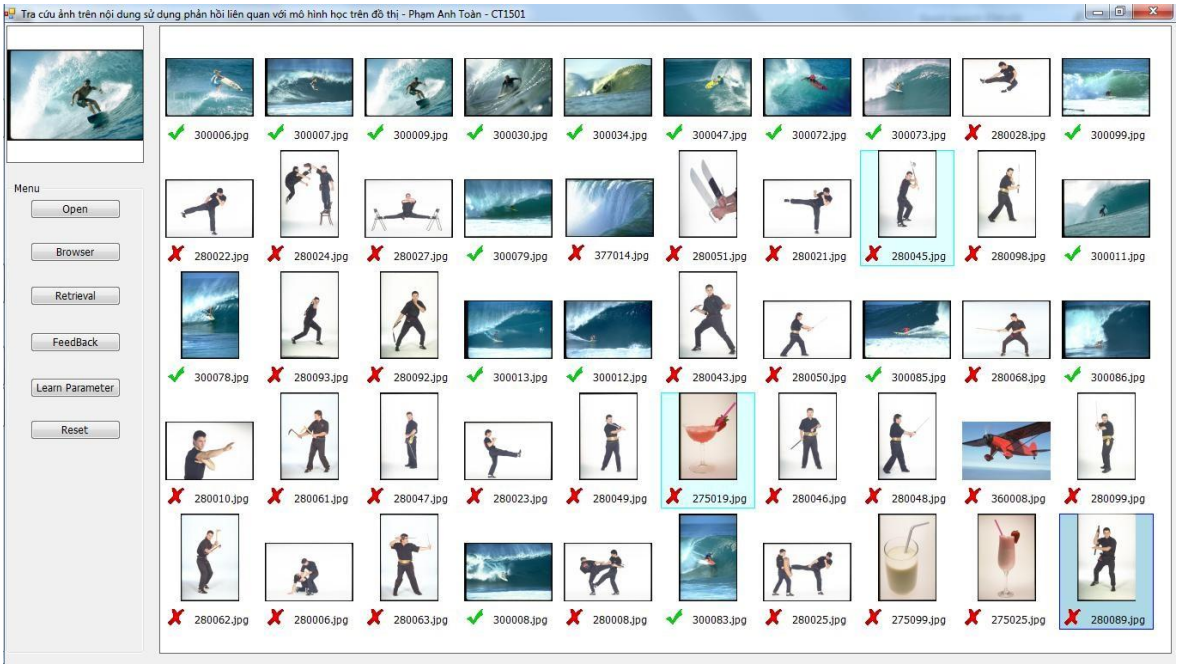


Hình 3-12: Kết quả số 2 sau lần phản hồi thứ 4

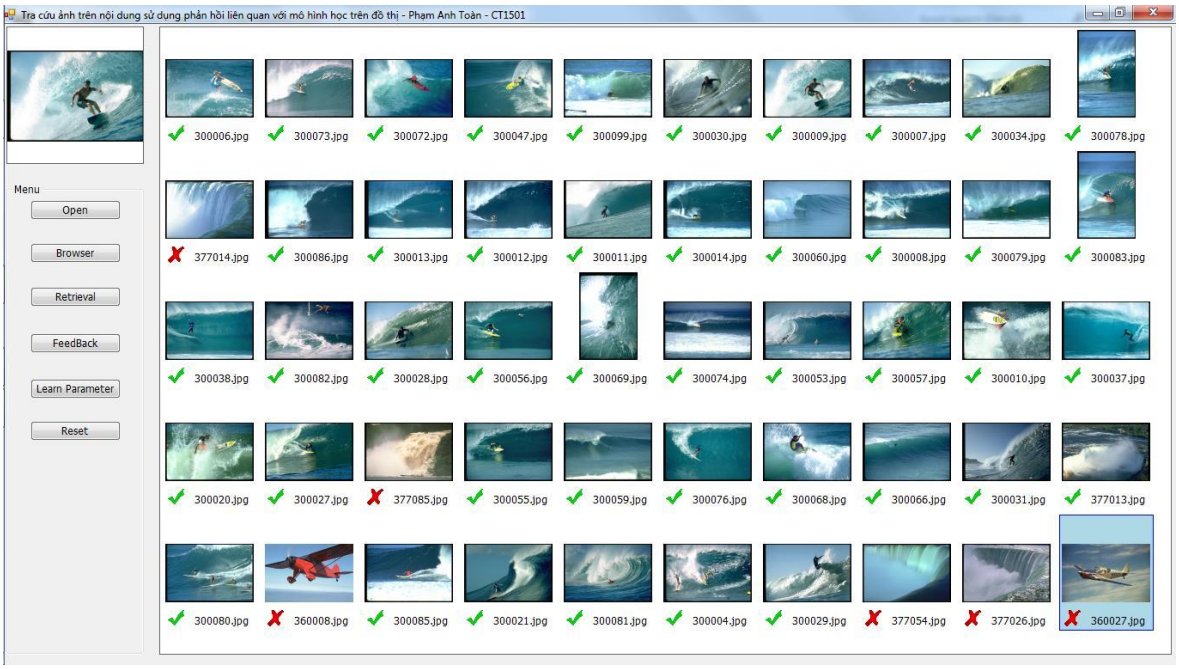
3.3.3 Kết quả thực nghiệm số 3



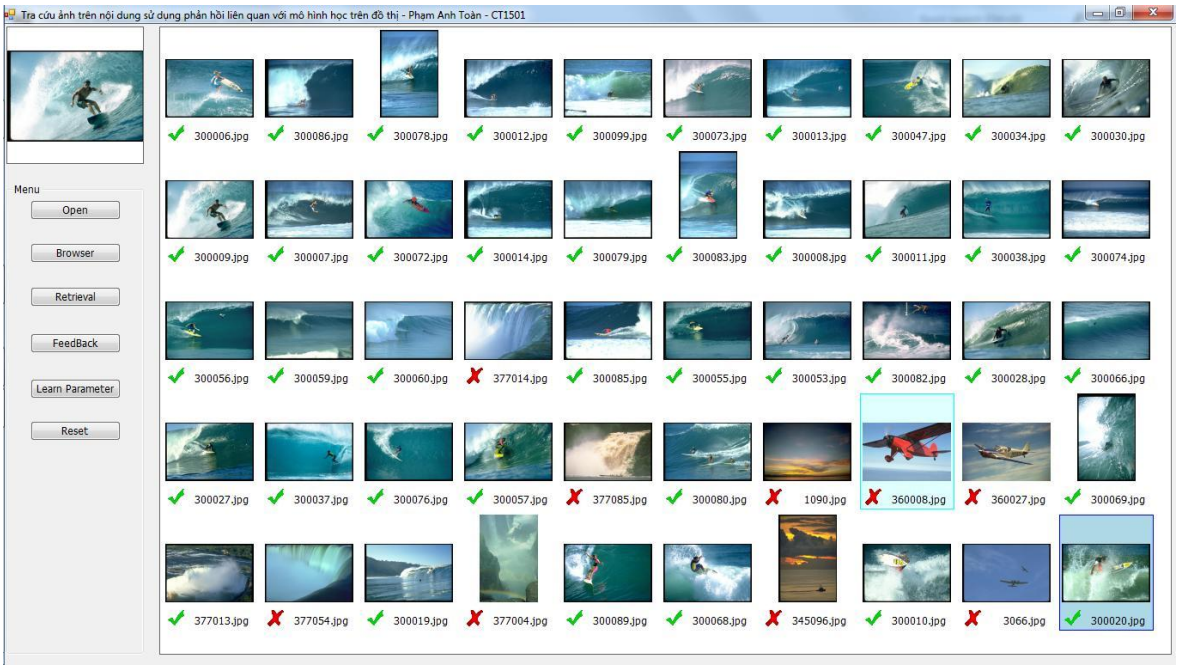
Hình 3-13: Mở ảnh truy vấn và kết quả thực nghiệm số 3 ban đầu



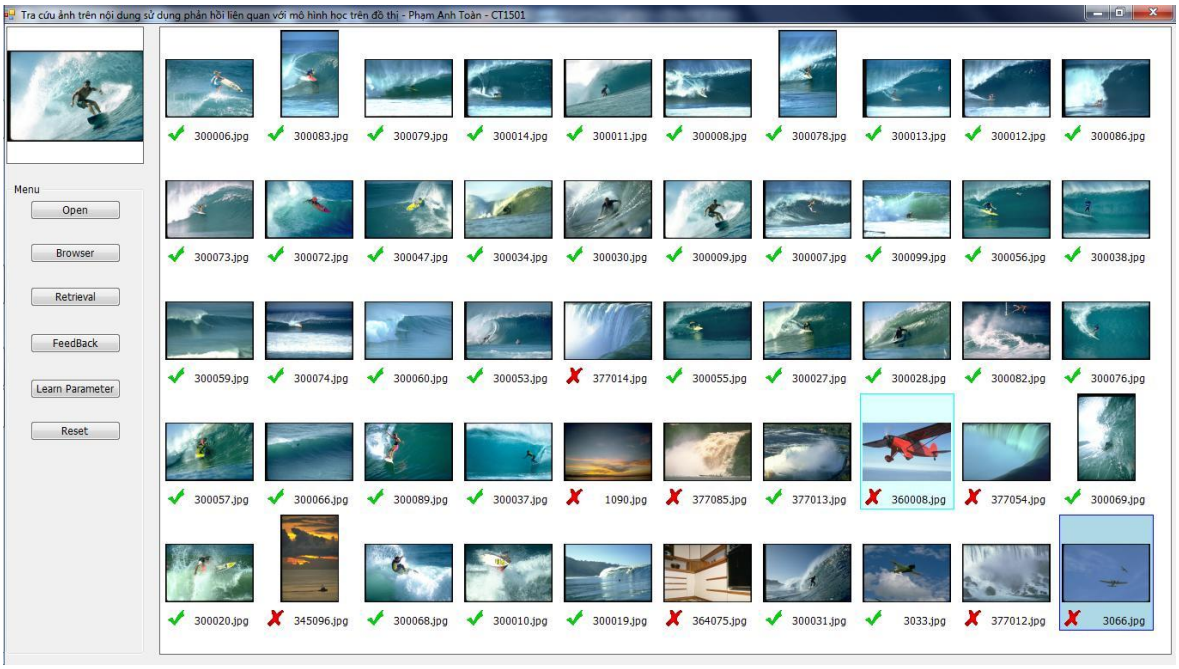
Hình 3-14: Kết quả của thực nghiệm số 3 sau lần phản hồi thứ nhất



Hình 3-15: Kết quả của thực nghiệm số 3 sau lần phản hồi thứ 2



Hình 3-16: Kết quả của thực nghiệm số 3 sau lần phản hồi thứ 3



Hình 3-17: Kết quả của thực nghiệm số 3 sau lần phản hồi thứ 4

KẾT LUẬN

Sau một thời gian tìm hiểu và nghiên cứu đề tài này, em đã đạt được một số kết quả sau:

- Tìm hiểu được cấu trúc của một hệ thống tra cứu ảnh dựa trên nội dung.
- Tìm hiểu được một số phương pháp làm giảm khoảng cách ngữ nghĩa trong tra cứu ảnh dựa trên nội dung.
- Tìm hiểu phương pháp phản hồi liên quan trong tra cứu ảnh.
- Tìm hiểu về một số phương pháp học máy đặc biệt là học bán giám sát dựa trên mô hình đồ thị.
- Xây dựng được chương trình thử nghiệm áp dụng phương pháp phản hồi liên quan sử dụng học bán giám sát trên đồ thị cho tra cứu ảnh dựa trên nội dung.

Tuy nhiên đồ án vẫn còn tồn tại một số vấn đề :

- Phần chương trình cài đặt tính toán còn chậm do cài đặt trong môi trường MS Visual Studio, khả năng của phần cứng có hạn chế.
- Phần cài đặt học siêu tham số chưa cho hiệu quả. Do độ phức tạp tính toán về thời gian của việc tính toán gradient là mn^2.
- Để nâng cao độ chính xác trong tra cứu ảnh cần tiếp tục nghiên cứu về mô hình học bán giám sát.

Em rất mong nhận được sự đóng góp ý kiến từ các Thầy Cô và các bạn để em có thêm kiến thức và kinh nghiệm tiếp tục hoàn thiện nội dung nghiên cứu trong đề tài. Em xin chân thành cảm ơn!

TÀI LIỆU THAM KHẢO

[1] J. Eakins, M. Graham, "Content-based image retrieval", Technical Report, University of Northumbria at Newcastle, 1999.

[2] A. Mojsilovic, B. Rogowitz, Capturing image semantics with low-level descriptors, Proceedings of the ICIP, September 2001, pp. 18–21.

[3] X.S. Zhou, T.S. Huang, CBIR: from low-level features to highlevel semantics, Proceedings of the SPIE, Image and Video Communication and Processing, San Jose, CA, vol. 3974, January 2000, pp. 426–431.

[4] Ying Liu, Dengsheng Zhang, Guojun Lu, Wei-ying Ma, "A survey of content-based image retrieval with high-level semantics," Pattern recognition, volume 40, issue 1, January, 2007, 262-282.

[5] Dr. Fuhui Long, Dr. Hongjiang Zhang and Prof. David Dagan Feng, "Fundamentals of content-based image retrieval", International journal of computer science and information technologies, vol.3 (1), 2012, 3260 – 3263.

[6] Xiaojin Zhu, "Semi-Supervised Learning with Graphs", CMU-LTI-05-192, May 2005.

[7] Pushpa B. PATIL, Manesh B. KOKARE, "Relevance Feedback in Content Based Image Retrieval: A Review", College of Engineering and Technology, Bijapur-586103, India, Institute of Engineering and Technology, Nanded-431606, India.

[8] R. Similar-shape retrieval in shape data management, IEEE Comput. 28 (9) (1995) 57–62Mehrotra, J.E. Gary.

[9] Zhang Xinhua, hyper-parameter learning for graph based semi-supervised learning algorithms, B.Eng., Shanghai Jiao Tong University, China, 2006.