BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

TRẦN VĂN HUY

KẾT HỢP XẾP HẠNG ĐA TẠP VÀ HỌC ĐỘ ĐO TƯƠNG TỰ CHO TRA CỨU ẢNH

TÓM TẮT LUẬN ÁN TIẾN SĨ CÔNG NGHỆ THÔNG TIN

Hà Nội, năm 2025

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

TRẦN VĂN HUY

KẾT HỢP XẾP HẠNG ĐA TẠP VÀ HỌC ĐỘ ĐO TƯƠNG TỰ CHO TRA CỨU ẢNH

Ngành: Công nghệ thông tin
Mã số: 9480201

TÓM TẮT LUẬN ÁN TIẾN SĨ CÔNG NGHỆ THÔNG TIN

NGƯỜI HƯỚNG DẪN KHOA HỌC:
1. TS. NGÔ HOÀNG HUY

2. TS. NGUYỄN VĂN ĐOÀN

Hà Nội, năm 2025

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Trong thập kỷ qua DOMO thống kê dữ liệu của thế giới, cho thấy sự gia tăng đáng kể trong hoạt động trên internet, từ Instagram và X đến Amazon. Từ năm 2013 đến 2024 số người dùng Internet phát triển từ 2.1 tỷ người đến 5.52 tỷ người, cùng với đó số lượng dữ liệu ảnh khổng lồ đã được tải lên internet. Các cơ sở dữ liệu hình ảnh này được sử dụng để cải thiện hiệu suất xử lý thông tin trong các ứng dụng thông minh, phục vụ cho nghiên cứu và cuộc sống hàng ngày.

Kỹ thuật tra cứu ảnh dựa trên nội dung (CBIR) đã được phát triển để tìm kiếm các hình ảnh có liên quan từ cơ sở dữ liệu dựa trên đối tượng hoặc nội dung của hình ảnh đầu vào. Đây là một bài toán được áp dụng rộng rãi trong lĩnh vực thị giác máy tính và mang lại hiệu quả kinh tế trong nhiều ứng dụng, chẳng hạn như: tìm kiếm khuôn mặt, vân tay, hình ảnh y tế, kỹ thuật hình sự, thương mại điện tử và nhiều ứng dụng khác.

Hạn chế của các phương pháp xếp hạng đa tạp hiện tại khi áp dụng cho bài toán tra cứu ảnh dựa trên nội dung:

i. Việc xây dựng đồ thị của các điểm dữ liệu dựa vào đồ thị K-NN là không khả thi với dữ liệu quy mô lớn [15].

ii. Khi thêm ảnh vào cơ sở dữ liệu, phải tính toán lại toàn bộ xếp hạng EMR (như trong EMR, SGR).

iii. Hạn chế trong việc xác định độ tương tự với ảnh ngoài cơ sở dữ liệu.

Trong luận án này, thuật ngữ "xếp hạng đa tạp" là kỹ thuật xếp hạng nhằm khám phá cấu trúc phi tuyến tính của dữ liệu đa tạp và được hiểu là phương pháp xếp hạng các điểm trong CSDL theo thứ tự có liên quan với điểm dữ liệu truy vấn được áp dụng trên tập cơ sở dữ liệu đa tạp.

Để giải quyết một phần các hạn chế ở trên, luận án chọn đề tài – ***Kết hợp xếp hạng đa tạp và học độ đo tương tự cho tra cứu ảnh.***

2. Mục tiêu của luận án

Mục tiêu chung của luận án: Nâng cao hiệu quả tra cứu ảnh dựa trên kết hợp xếp hạng đa tạp và tiếp cận học độ đo tương tự.

Mục tiêu cụ thể của luận án:

Đề xuất được một số giải pháp nâng cao độ chính xác tra cứu ảnh dựa trên nội dung theo tiếp cận xếp hạng đa tạp bao gồm:

Nghiên cứu một số giải pháp nâng cao độ chính xác tra cứu ảnh dựa trên nội dung theo tiếp cận xếp hạng đa tạp bao gồm:

- Nghiên cứu thuật toán xếp hạng đa tạp hiệu quả, nghiên cứu kết hợp nhiều bộ xếp hạng hình ảnh theo đặc trưng mức thấp với xếp hạng của hình ảnh đặc trưng mức cao.

- Nghiên cứu xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR để giải quyết việc ảnh tra cứu nằm ngoài cơ sở dữ liệu.

3. Đối tượng nghiên cứu của luận án

- Các phương pháp hiện tại về Tra cứu ảnh dựa vào nội dung.

- Các kỹ thuật biểu diễn ảnh với đặc trưng mức thấp, đặc trưng véc tơ nhúng, đặc trưng CNN (đặc trưng ảnh được trích rút từ mạng học sâu).

- Các kỹ thuật học máy, học bán giám sát các giá trị xếp hạng EMR.

- Môi trường thực nghiệm, tập dữ liệu ảnh thực nghiệm và phương pháp đánh giá độ chính xác.

4. Phạm vi nghiên cứu

Trong luận án này, phạm vi nghiên cứu bao gồm:

- Nghiên cứu thuật toán xếp hạng đa tạp trong tra cứu ảnh dựa vào nội dung.

- Nghiên cứu phương pháp tổ hợp xếp hạng đa tạp hiệu quả (EMR) kết hợp nhiều bộ xếp hạng hình ảnh theo đặc trưng mức thấp với xếp hạng của hình ảnh đặc trưng mức cao (đặc trưng véc tơ nhúng, đặc trưng CNN).

- Nghiên cứu phương pháp nâng cao hiệu quả tra cứu ảnh bằng cách xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR. Đề xuất kỹ thuật tra cứu EMR Learning để giải quyết việc ảnh tra cứu nằm ngoài cơ sở dữ liệu.

- Trong phạm vi của luận án chỉ tập trung nâng cao chất lượng tra cứu về độ chính xác, các vấn đề về thời gian cho một truy vấn cũng được xem xét ở khía cạnh có thể chấp nhận được.

. 5. Các đóng góp của luận án

Nhằm mục tiêu nâng cao độ chính xác của tra cứu ảnh sử dụng phương học độ đo tương tự, luận án có các đóng góp sau:

(1) Nghiên cứu phương pháp tổ hợp các bộ xếp hạng đa tạp hiệu quả, đề xuất thuật toán CoEMR kết hợp nhiều bộ xếp hạng hình ảnh theo đặc trưng mức thấp, xếp hạng của hình ảnh đặc trưng mức cao [CT1, CT2, CT3, CT4]. Đề xuất phương pháp sử dụng truy vấn nhiều ảnh trên CBIR [CT8].

(2) Nghiên cứu phương pháp nâng cao hiệu quả tra cứu ảnh bằng cách xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR. Đề xuất thuật toán tra cứu EMR Learning để giải quyết việc ảnh tra cứu nằm ngoài cơ sở dữ liệu [CT5, CT6, CT7].

6. Bố cục của luận án

Luận án được tổ chức thành ba chương:

Chương 1: Tra cứu ảnh dựa trên nội dung.

Chương 2: Phương pháp tra cứu ảnh sử dụng thuật toán kết hợp nhiều bộ xếp hạng đa tạp hiệu quả.

Chương 3: Xây dựng độ đo tương tự ảnh theo các giá trị xếp hạng EMR

Cuối cùng, luận án đưa ra một số đề xuất và định hướng nghiên cứu trong tương lai.

Chương 1
TRA CỨU ẢNH DỰA TRÊN NỘI DUNG

1.1. Giới thiệu về tra cứu ảnh dựa vào nội dung

Tra cứu ảnh dựa vào nội dung (CBIR) [31] thu hút rất nhiều sự chú ý từ các nhà nghiên cứu và được sử dụng nhiều trong công nghiệp, thương mại trong những năm qua do nhiều ứng dụng hữu ích của nó. Các thuật toán tra cứu ảnh thường xây dựng các độ đo tương tự toàn cục giữa các vector đặc trưng biểu diễn đối tượng ảnh đối sánh với toàn bộ vector đặc trưng trong CSDL.

1.2. Biểu diễn ảnh bằng vector đặc trưng

Trong chương này luận án trình bày tổng quan về các đặc trưng biểu diễn ảnh

1.2.1. Đặc trưng mức thấp của ảnh

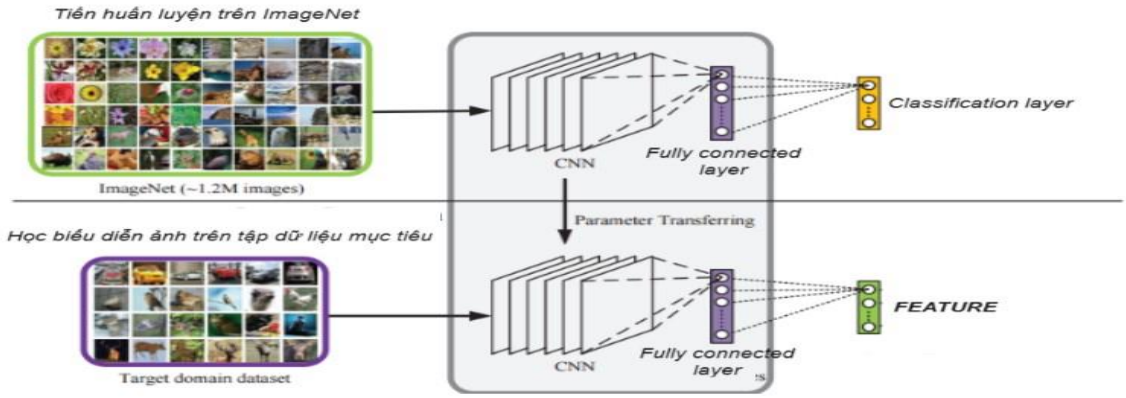
Trình bày tổng quan về các đặc trưng mức thấp được sử dụng trong CBIR nói chung như: Đặc trưng màu sắc; Đặc trưng kết cấu; Đặc trưng hình dạng…

1.2.2. Đặc trưng mức cao của ảnh

Đặc trưng Véc tơ nhúng: Là đặc trưng được trích rút ở lớp fully connected cuối cùng của mạng học sâu

Đặc trưng CNN: Là đặc trưng được trích rút ở tầng cuối (trước tầng phân lớp) của mạng CNN.

Hình 1.. Mô hình trích rút đặc trưng ảnh bằng mô hình học sâu



Hình 1.5 là mô hình thuật toán biểu diễn đặc trưng ảnh được trích rút dựa vào mạng học sâu CNN tiền huấn luyện thu được tập đặc trưng mức cao của ảnh.

1.3. Độ đo khoảng cách, độ đo tương tự

1.3.1. Độ đo khoảng cách, độ đo tương tự

Độ đo tương tự là một trong những phương pháp tốt để máy tính phân biệt được các hình ảnh qua nội dung của chúng. Thông thường hệ thống tra cứu ảnh sẽ truy vấn hình ảnh bằng phương pháp đo tương tự dựa trên các chức năng, việc xác định nó có thể dưới nhiều hình thức như phát hiện biên, màu sắc, vị trí điểm ảnh....

Một số công thức tính độ khoảng cách hay được sử dụng trong CBIR được mô tả như bảng sau:

Bảng 1.2: Mộ số độ đo khoảng cách, độ đo tương tự và công thức

Độ đo	**Công thức tính**
Euclid	$\sqrt{(x-y)^T(x-y)}$
Mahalanobis	$\sqrt{(x-y)^T C^{-1}(x-y)}$; C là psd
Minkowski	$D(x,y)=(\sum_{i=1}^{n}\lvert x_i-y_i\rvert^p)^{\frac{1}{p}}$
Manhattan (Taxicab/City Block) *(Khoảng cách L_1)*	$D(x,y)=\sum_{i=1}^{k}\lvert x_i-y_i\rvert$ (k là số block)
Chebyshev	$D(x,y)=\max_i(\lvert x_i-y_i\rvert)$
Cosine	$D(X,Y)=cos\,\theta=\frac{X.Y}{\lvert\lvert X\rvert\rvert.\lvert\lvert Y\rvert\rvert}$

1.3.2. Học độ đo khoảng cách, độ đo tương tự

Thông thường có 2 phương pháp học độ đo khoảng cách, độ tương tự ảnh

+ Loại học bán giám sát: Gắn trọng số vào các chiều của thành phần dữ liệu đặc trưng ảnh khi tính khoảng cách Euclid thông thường.

Các trọng số được thích ứng thông qua thông tin phản hồi của người sử dụng trong quá trình truy vấn ảnh.

+ Loại với tập dữ liệu ảnh đã gán nhãn (Có thông tin trước về mẫu ảnh tương tự và không tương tự):

Sử dụng các phương pháp học máy như: SVM, EM, KNN, Deep Learning,…

Học độ đo khoảng cách (NCA)

Học độ đo khoảng cách NCA (Neighbourhood Components Analysis) là một kỹ thuật học độ đo khoảng cách được đề xuất bởi Goldberger và cộng sự năm 2004 [125]. NCA không tham số áp dụng cho cả bài toán phân loại và hồi quy. Mục tiêu của NCA là tối ưu hóa trọng số của đặc trưng để tối đa hóa độ chính xác của bài toán bằng cách tìm hàm ánh xạ sao cho các điểm dữ liệu cùng nhãn gần nhau hơn các điểm khác nhãn, giúp giảm sai số phân lớp của k-NN. NCA tối ưu hóa hàm mục tiêu dựa trên sai số phân lớp dự đoán của k-NN trên tập huấn luyện bằng cách tính xác suất các láng giềng cùng nhãn với mỗi điểm dữ liệu và cực đại hóa tổng xác suất.

Học độ đo theo nhóm phân cụm (CMML)

Học độ đo theo nhóm phân cụm (CMML) là một phương pháp hiệu quả để nhiều độ đo khoảng cách dựa trên kết quả phân cụm dữ liệu. CMML phân loại dữ liệu thành các cụm khác biệt và tạo ra một độ đo khoảng cách riêng lẻ cho mỗi cụm. Điều này cho phép mô hình có khả năng học hỏi và áp dụng các độ đo khoảng cách thích hợp với từng loại dữ liệu trong mỗi cụm. Bên cạnh đó, một kỹ thuật điều chỉnh toàn cục được áp dụng nhằm bảo toàn các đặc tính chung của các cụm trong không gian metric đã được học.

1.3.3. Tiếp cận Deep learning cho học độ đo khoảng cách

- Chiến thuật Contrastive (mạng Siamese)

Cho $X_1, X_2 \in I$ là cặp véc tơ đầu vào, và Y là nhãn nhị phân, nếu:

Y = 0 là cặp tương tự

Y = 1 là cặp không tương tự

Khi đó hàm khoảng cách với tham số W được biểu diễn như sau

$$D_W(X_1, X_2) = \|G_W(X_1) - G_W(X_2)\|_2 \quad (1.4)$$

Hàm mất mát được đưa ra như sau

$$L(W, Y, X_1, X_2) = (1 - Y)\frac{1}{2}\left(D_W(X_1, X_2)\right)^2 + (Y)\frac{1}{2}\{max\left(0, m - D_W(X_1, X_2)\right)\}^2 \quad (1.5)$$

, trong đó: m>0 là lề.

Hàm mất mát tổng thể:

$$L = \frac{1}{2N}\sum_{i=1}^{N} L\left(W, Y_i, X_{1,i}, X_{2,i}\right) \quad (1.6)$$

với N là số lượng một lô ảnh huấn luyện.

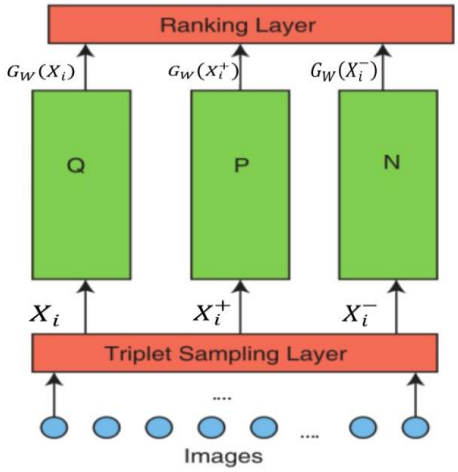
- Chiến thuật Triplet (được cải tiến từ Contrastive)

Tối ưu hóa việc nhúng và sử dụng bộ ba (ảnh neo, ảnh phù hợp và ảnh không phù hợp) để huấn luyện độ đo khoảng cách.

Với một bộ ba $(x_i,\ x_j, x_k)$ sao cho khoảng cách giữa x_i và x_j nhỏ hơn khoảng cách giữa x_i và x_k thì:

$$d_A(x_i; x_j) \leq d_A(x_i; x_k) - m \quad (1.7)$$

, trong đó *m >0* là lề.



Hình 1.10: Mô hình mạng Triplet

Hàm *G(.)* được học và gán sao cho

G(.) < khoảng cách giữa các ảnh tương tự nhau.

Hàm mất mát L được định nghĩa như sau,

khi có bộ ba $T_i = (X_i, X_i^+, X_i^-)$ ta có

$$L(X_i, X_i^+, X_i^-) = \max\{0, m + D_W(X_i, X_i^+) - D_W(X_i, X_i^-)\} \quad (1.8)$$

, trong đó: m>0 là lề,

Hàm mất mát tổng thể:

$$L = \frac{1}{N}\sum_{i=1}^{N} L(X_i, X_i^+, X_i^-) \quad (1.9)$$

1.4. Xếp hạng dựa trên đồ thị với tiếp cận xếp hạng đa tạp

1.4.1 Các phương pháp xếp hạng dựa trên đồ thị

Các mô hình xếp hạng dựa trên đồ thị đã trở nên phổ biến trong nghiên cứu truy vấn ảnh dựa trên nội dung (CBIR) và đã thu hút sự quan tâm đáng kể từ cộng đồng học máy, thị giác máy tính và tìm kiếm thông tin trong thời gian gần đây. Chúng được áp dụng rộng rãi để mô tả mối quan hệ tương đồng/khác biệt giữa các hình ảnh, từ đó tạo ra các bảng xếp hạng hữu ích cho bài toán truy vấn.

Một số mô hình tiêu biểu có thể kể đến như: MR, AGR, SGR, EMR… Các nghiên cứu gần đây [33-35, 39, 63, CT6] đã chứng minh EMR cho kết quả xếp hạng chính xác hơn so với các phương pháp truyền thống khác. Điều này mở ra nhiều triển vọng trong ứng dụng thực tiễn.

1.4.1. Xếp hạng đa tạp cơ bản

Trong số đó, xếp hạng trên cấu trúc dữ liệu đa tạp (Ranking on Data manifold) [33] là một trong những phương pháp đại diện và đã được áp dụng rộng rãi trong các ứng dụng tìm kiếm thông tin và học máy khác nhau.

Việc xây dựng đồ thị biểu diễn các điểm trong cơ sở dữ liệu (CSDL) theo thuật toán xếp hạng đa tạp (MR-Manifold Ranking) đã được đề xuất trong các nghiên cứu [9, 33-35]. Mục tiêu chính của phương pháp MR dựa trên cách tiếp cận đồ thị là sử dụng phương pháp xác định trọng số của mỗi điểm dữ liệu so với các điểm dữ liệu khác dựa trên các thông tin toàn cục và cục bộ biểu diễn bên trong đồ thị

1.4.2. Xếp hạng đa tạp hiệu quả

Để khắc phục hạn chế của xếp hạng đa tạp, trong [4] Bin Xu và các cộng sự đã đề xuất phương pháp xếp hạng đa tạp hiệu quả (Efficient Manifold Ranking - EMR). Phương pháp này tập trung giải quyết hai vấn đề chính:

1- Xây dựng đồ thị neo (Anchor) có khả năng mở rộng thay cho đồ thị *k-NN* truyền thống.

2- Xây dựng mô hình tính toán xếp hạng hiệu quả, có chi phí tính toán nhỏ bằng cách thiết kế hình thức mới của ma trận kề.

Mô hình này có hai pha tách rời: thứ nhất là pha ngoại tuyến (offline) để xây dựng đồ thị điểm neo cho toàn bộ cơ sở dữ liệu; pha thứ hai là pha trực tuyến (online) để xử lý một truy vấn mới.

1.5. Xếp hạng đa tạp hiệu quả và vấn đề tra cứu ảnh

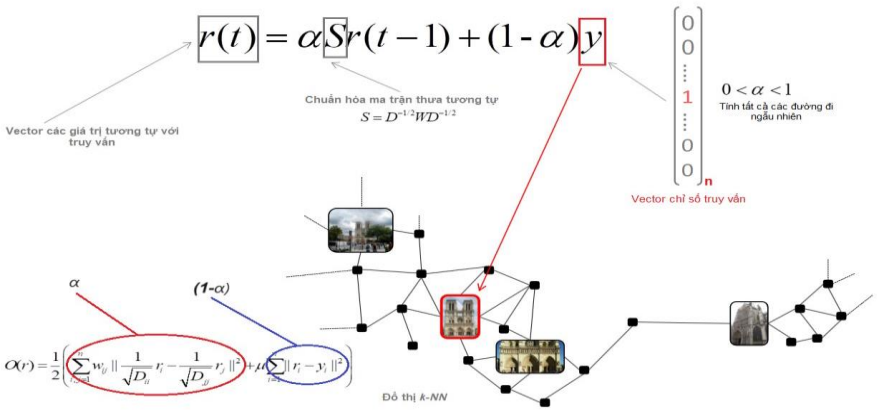
Để khắc phục hạn chế của xếp hạng đa tạp, trong [42] Bin Xu và các cộng sự đã đề xuất phương pháp xếp hạng đa tạp hiệu quả. Phương pháp này tập trung giải quyết hai vấn đề chính:

1- Xây dựng đồ thị neo (Anchor) có khả năng mở rộng thay cho đồ thị *k-NN* truyền thống.

2- Xây dựng mô hình tính toán xếp hạng hiệu quả, có chi phí tính toán nhỏ bằng cách thiết kế hình thức mới của ma trận kề.

Mô hình này có hai pha tách rời: thứ nhất là pha ngoại tuyến (offline) để xây dựng đồ thị điểm neo cho toàn bộ cơ sở dữ liệu; pha thứ hai là pha trực tuyến (online) để xử lý một truy vấn mới.

Tra cứu dùng thuật toán MR trong CBIR:



Hình 1.12. Quá trình tra cứu trong MR với đồ thị *K-NN*

1.6. Phương pháp đánh giá hiệu quả trong CBIR

Độ chính xác là tỉ số giữa các ảnh liên quan được truy vấn với tổng số các ảnh truy vấn, được tính theo công thức: $P_r = \frac{E_Q}{D_Q}$

Độ triệu hồi là tỉ số giữa số ảnh liên quan được truy vấn với toàn bộ số ảnh có liên quan trong CSDL: $R_e = \frac{E_Q}{A_Q}$

Độ chính xác trung bình ARP (Average Retrieval Precision) thường được sử dụng để đánh giá độ chính xác của phương pháp được sử dụng trong CBIR. Hiệu quả truy vấn chung của một hệ thống được đo bằng trung bình tất cả độ chính xác. ARP được tính toán như sau:

$$ARP = average\left(\sum P_i\right) \quad (1.14)$$

Với P_i là độ chính xác của mỗi truy vấn (Độ chính xác trong tra cứu tại điểm dữ liệu thứ *i*). Nó là một độ đo hiệu quả để biểu diễn hiệu suất của hệ thống CBIR. Trong các thực nghiệm ở chương 2 và chương 3, luận án sử dụng độ chính xác trung bình để đánh giá hiệu quả của các phương pháp.

Với trường hợp nhiều truy vấn, độ chính xác trung bình trên tất cả các truy vấn ký hiệu là *MAP* (Mean Average Precision) được định nghĩa như sau:

$$MAP = \frac{1}{Q}\sum_{q=1}^{Q} AP_q \quad (1.15)$$

Với *Q* là tổng số ảnh truy vấn. Công thức (1.26) được dùng để đánh giá hiệu quả tra cứu chung cho một hệ thống CBIR với Q ảnh đưa vào truy vấn.

Giá trị MinP được cải thiện đánh giá hiệu quả phục hồi trong các truy vấn cho kết quả xấu nhất.

$$minP = \min_{1 \le j \le |Q|}\left(\frac{m_j}{N}\right) * 100 \quad (1.16)$$

Giá trị σP nhỏ chứng tỏ rằng độ chính xác là đồng đều và tốt cho tất cả các hình ảnh truy vấn trong Q.

$$\sigma P = \sigma\left\{\frac{m_j}{N} * 100\right\}_{1 \le j \le Q} \quad (1.17)$$

Tổng quan về các độ đo đánh giá hệ thống CBIR có thể tìm thấy trong [85]. Để thể hiện ý nghĩa thống kê, hiệu quả tra cứu được tính trên số lượng các truy vấn. Thông thường, số lượng các truy vấn là từ 100 đến 1000 [86]. Kích thước của tập dữ liệu ảnh là từ 1000 đến 20000 [87]. Với các hệ thống tra cứu quy mô lớn, kích thước tập dữ liệu có thể lên đến 80 triệu ảnh [88].

1.7. Một số CSDL thực nghiệm cho tra cứu ảnh

Bảng 1.3. Các tập dữ liệu ảnh

Tập ảnh	Số lượng ảnh	Số lớp ảnh
Leaf30	11600	116
SIGN300HD	4370	300
Corel30K	31695	306
VGG-60K	60000	500

1.8. Kết luận chương 1

Trong chương 1 luận án đã trình bày về tra cứu ảnh dựa trên nội dung. các độ đo khoảng cách, độ đo

tương tự, tiếp cận học độ đo khoảng cách, độ đo tương tự là thành phần quan trọng trong các hệ thống CBIR.

Cũng ở trong chương này luận án tập trung giới thiệu và phân tích các xếp hạng EMR cho đặc trưng mức thấp, xếp hạng EMR cho đặc trưng mức cao (đặc trưng CNN, đặc trưng véc tơ nhúng), qua đó rút ra một số kết luận về sử dụng EMR cho CBIR như sau:

- Ưu điểm

+ Vừa có tính địa phương: lấy theo lân cận gần nhất

+ Tính toàn cục: có tính lan tỏa giữa các lân cận

+ Đưa ra được độ tương tự giữa ảnh truy vấn và ảnh trong CSDL có hiệu quả khá tốt

+ Nhược điểm

+ Dù đã cải tiến từ MR sang EMR tuy nhiên với hệ thống ảnh lớn thì số anchor lớn -> hệ thống vẫn chậm. Với phép toán ma trận tính toán số lượng anchor lớn thì rất chậm

+ Nếu thêm ảnh mới vào cơ sở dữ liệu thì có khả năng phải xây dựng lại mô hình EMR (tính lại ma trận Z)

+ Chỉ xác định được ảnh truy vấn và ảnh trong cơ sở dữ liệu như vậy sẽ không tính được độ tương tự ảnh của 2 ảnh nằm ngoài cơ sở dữ liệu.

Từ đó, luận án đưa ra 02 vấn đề cần giải quyết.

Vấn đề 1: Xây dựng bộ xếp hạng CSDL ảnh theo ảnh truy vấn dựa trên nhiều bộ xếp hạng EMR.

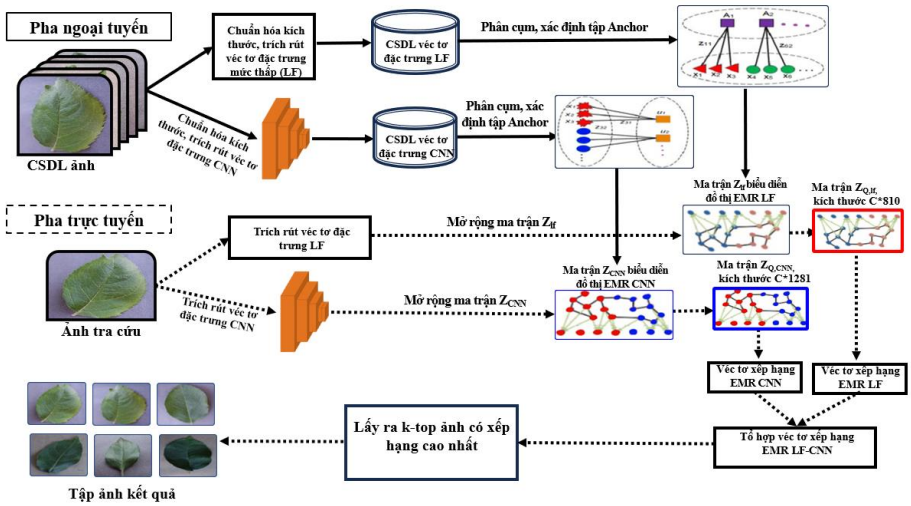
Vấn đề 2: Xây dựng một độ đo tương tự ảnh được học dựa trên kết quả xếp hạng của EMR.

Chương 2 và Chương 3 của luận án này sẽ tập trung giải quyết các vấn đề 1 và vấn đề 2 để nâng cao độ chính xác cho các hệ thống CBIR.

Chương 2
PHƯƠNG PHÁP TRA CỨU ẢNH SỬ DỤNG THUẬT TOÁN KẾT HỢP NHIỀU BỘ XẾP HẠNG ĐA TẠP HIỆU QUẢ

Chương 1 phân tích các kỹ thuật CBIR và xếp hạng đa tạp. Chương 2 đề xuất phương pháp CoEMR để cải thiện độ chính xác CBIR, với thực nghiệm trên các tập dữ liệu như VGG60K, CEDAR. Các phương pháp xếp hạng trên đồ thị như AGR, SGR, MR, EMR đã được áp dụng rộng rãi, nhưng MR hạn chế với cơ sở dữ liệu lớn. Các phương pháp mới như SMR, GMR, FMR kết hợp đặc trưng mức thấp và cao để cải thiện hiệu quả CBIR. Chương này đề xuất kỹ thuật tổ hợp xếp hạng EMR cho đặc trưng mức thấp và cao, xây dựng mô hình CoEMR.

Kiến trúc CBIR đề xuất



Hình 2.1. Mô hình hệ thống đề xuất CBIR với việc sử dụng kết hợp xếp hạng Co*EMR*

Kiến trúc đề xuất hệ thống CBIR trong hình 2.1 bao gồm các bước sau:

+ Pha ngoại tuyến:

+ Pha trực tuyến:

2.1 Tiếp cận kết hợp đặc trưng mức thấp và đặc trưng CNN trong mô hình CoEMR đề xuất

Luận án giới thiệu phương pháp CoEMR mới để cải thiện truy vấn ảnh bằng cách sử dụng đồng thời các đặc trưng mức thấp và đặc trưng CNN. CoEMR áp dụng hai xếp hạng EMR riêng biệt cho từng loại đặc trưng, sau đó kết hợp kết quả để tạo ra thứ hạng mới cho các ảnh. Phương pháp này tận dụng tối đa thông tin từ cả hai bộ đặc trưng, nâng cao độ chính xác của hệ thống truy vấn. Cách tiếp cận giúp phản ánh đúng mức độ tương tự giữa các ảnh một cách toàn diện hơn. Phần này sẽ chi tiết từng bước của phương pháp CoEMR, từ xếp hạng ban đầu đến kết hợp kết quả.

2.2 Phát biểu các ràng buộc cho lớp hàm kết hợp xếp hạng

Đối với một ảnh truy vấn Iq và CSDL E chứa n ảnh, mỗi ảnh I trong E sẽ có hai thứ hạng a và b dựa trên các mô hình xếp hạng EMR khác nhau (EMRLF cho đặc trưng mức thấp và EMRHF cho đặc trưng mức cao). Kết hợp các thứ hạng này tạo ra một thứ hạng mới cho ảnh I so với Iq. Luận án sử dụng hàm kết hợp CB để nhập hai số thực thành một số thực duy nhất, với các yêu cầu: CB nằm trong khoảng

từ min đến max của avà b, và CB không giảm. Các hàm CB bao gồm kết hợp tuyến tính, kết hợp kiểu lựa chọn và kết hợp dạng luỹ thừa bậc lẻ.

Các hàm CB được sử dụng trong luận án bao gồm:

- Kết hợp tuyến tính $CB(r_{1,}r_2) = \{\alpha r_{1,} + \beta r_2\}. \quad (\alpha, \beta > 0 \alpha + \beta = 1)$ (2.1)

- Kết hợp kiểu lựa chọn: $CB(r_{1,}r_2) = \begin{cases} r_2 \\ r_1 \end{cases} \quad if\ r_2 \geq th$ (2.2)

- Kết hợp dạng luỹ thừa bậc lẻ:

$$CB(r_{1,}r_2) = \begin{cases} \sqrt[3]{\frac{\{r_1^3 + r_2^3\}}{2}} \\ \min(r_1, r_2) \end{cases} \quad if\ r_1 + r_2 \geq 0 \tag{2.3}$$

2.3 Kết hợp lựa chọn 2 xếp hạng EMR.

Ý tưởng kết hợp lựa chọn xuất phát từ việc nhận thấy một số đặc trưng CNN có độ phân biệt cao, như vector nhúng của FaceNet cho ảnh khuôn mặt. Wang và cộng sự đã chứng minh hai đặc trưng fc4096a và fc4096b từ AlexNet có tính khái quát tốt và hiệu suất cao trên bộ dữ liệu ImageNet. Tuy nhiên, đặc trưng CNN đôi khi không đạt kết quả mong muốn, do đó kết hợp với đặc trưng mức thấp như màu sắc, hình dạng, và kết cấu có thể cải thiện truy vấn ảnh. Phương pháp kết hợp này giúp tận dụng ưu điểm của cả hai loại đặc trưng, mang lại hệ thống truy vấn hình ảnh hiệu quả hơn. Các bước và kết quả thực nghiệm của phương pháp này sẽ được trình bày chi tiết trong các phần tiếp theo của luận án.

(a) (b)

Hình 2.2. Khi tra cứu ảnh sử dụng đặc trưng CNN đa số cho kết quả tốt (a), tuy nhiên vẫn có những trường hợp ảnh có "cảm quan" tốt nhưng kết quả tra cứu lại kém (b).

2.4 Kết hợp tuyến tính 2 xếp hạng EMR.

Ý tưởng kết hợp tuyến tính xuất phát từ việc nhận thấy đặc trưng mô tả ảnh chữ ký là hạn chế. Đặc trưng mức thấp (HOG, LBP) mô tả tốt các đặc điểm như cạnh, góc, điểm nổi bật, trong khi đặc trưng CNN (SigNet) học được các đặc điểm biên hoặc mẫu hình học. Luận án đề xuất mô hình kết hợp tuyến tính để tận dụng tối đa các đặc trưng mức thấp và mức cao, cải thiện khả năng tổng quát hóa và độ chính xác của mô hình. Phương pháp bao gồm các bước thực hiện cụ thể và kết quả thực nghiệm minh họa cho hiệu quả.

2.5 Kết hợp luỹ thừa bậc lẻ 2 xếp hạng EMR.

Ý tưởng kết hợp luỹ thừa bậc lẻ xuất phát từ việc nhận thấy các đặc trưng ảnh mang thông tin khác nhau: đặc trưng mức thấp mô tả các đặc điểm cơ bản, trong khi đặc trưng CNN (như EfficientNet) học các đặc trưng trừu tượng và phức tạp. Luận án đề xuất mô hình kết hợp luỹ thừa bậc lẻ để tận dụng toàn

bộ các đặc trưng, mang lại hai lợi ích chính: tổng hợp thông tin đa mức, giúp cải thiện nhận diện và hiểu hình ảnh; và tăng tính đa dạng của đặc trưng, giúp mô hình học và phát hiện nhiều loại đặc trưng khác nhau. Phương pháp sẽ được phân tích chi tiết và minh họa bằng các kết quả thực nghiệm trong các phần tiếp theo của luận án.

Thuật toán CoEMR

Giả sử rằng tập dữ liệu ảnh gốc, mỗi ảnh có một vector đặc trưng duy nhất tương ứng và với một ảnh truy vấn I_Q, hệ thống chọn xếp hạng theo hình ảnh đặc trưng mức thấp $r^*_{lf.v.Q,i}$ kết hợp với xếp hạng của hình ảnh đặc trưng mức cao $r^*_{CNN.Q,i}$, trong đó các vector xếp hạng được tính bằng EMR.

Sau đây là thuật toán chi tiết mà chúng tôi đề xuất.

Thuật toán 1. CoEMR (Thuật toán kết hợp các 2 bộ xếp hạng EMR riêng rẽ trên đặc trưng mức thấp và CNN).

Input: $\{I_i\}_{1\le i\le n}$ là tập ảnh gốc, I_Q ảnh truy vấn.

MxN là kích thước ảnh chuẩn hóa cho huấn luyện theo thuật toán EfficientNet.

d: số chiều của một vector đặc trưng ảnh.

C: số lượng điểm neo của thuật toán EMR, tham số a ∈ (0,1) (a ≈ 1),

α, β:trọng số kết hợp tuyến tính xếp hạng của 2 EMR, α, β >0, α + β = 1.

Bước 1 (offline): **Xây dựng đồ thị EMR**

Bước 1a. Huấn luyện tập ảnh E theo mô hình EfficientNet với kích thước chuẩn hóa ảnh đầu vào là M*N, thu nhận được:

1.1. Tham số model W_CNN.

1.2. Chạy từng ảnh $\{I_i\}_{1\le i\le n}$ qua mô hình thu được tập các vector đặc trưng $\{v.E_i\}_{1\le i\le n}$ với chiều d_{CNN}=1280.

1.3. Xác định C điểm neo $\{A_c\}_{1\le c\le C}$ của tập ảnh $\{I_i\}_{1\le i\le n}$ dựa trên một cải tiến của thuật toán FCM, sử dụng các vector đặc trưng CNN $\{v.E_i\}_{1\le i\le n}$

1.4. Xác định ma trận kề W=(w_{ij}) của thuật toán EMR sử dụng các vector đặc trưng CNN $\{v.E_i\}_{1\le i\le n}$.

1.5. Xác định ma trận trọng số Z kích thước C*n của EMR.

Bước 1b. Trích rút đặc trưng mức thấp: Color Moments, LBP, Gabor Wavelets Texture, Edge và GIST, thu được tập véc tơ đặc trưng với số chiều d_{LF}=809. Lặp lại bước 1.3 - 1.5 như ở bước 1a với đặc trưng mức thấp.

Bước 2 (online): Kết hợp xếp hạng 2 EMR

2.1. Chuẩn hóa ảnh I_Q về kích thước M*N, chạy qua mô hình EfficientNet đã xác định ở bước 1.1, thu nhận được vector đặc trưng CNN v.E_Q với d chiều.

2.2. Mở rộng ma trận Z theo EMR [7] (có thể xem công thức (4) ở trên): Sử dụng C giá trị khoảng cách của E_Q với các phần tử neo của A_c, chúng ta thu được ma trận trọng số Z_Q mới có kích thước C*(n+1).

2.3. Đặt $r_Q = \{r_i\}_{1\le i\le n+1}, r_i = 0 \forall i = \overline{1,n}, r_{n+1} = 1.0$. Từ ma trận trọng số Z_Q chúng ta xác định vector thứ hạng $r^*_{CNN.Q,i}$bằng thuật toán EMR [36].

Bước 3: Lặp lại bước 2.1 - 2.3 nhưng với đặc trưng mức thấp lf.E_i ta được giá trị xếp hạng là $r^*_{lf.v.Q,i}$

Bước 4: **Kết hợp vector xếp hạng của 2 EMR:**

- Kết hợp tuyến tính $CB(r_{1,}r_2) = \{\alpha r^*_{lf.v.Q,i} + \beta r^*_{CNN.Q,i}\}_{1\le i\le n}. (\alpha, \beta > 0 \alpha + \beta = 1)$ (2.11)

- Kết hợp kiểu lựa chọn: $CB(r_{1,}r_2) = \begin{cases} r_2 \\ r_1 \quad if\ r_2 \ge th \end{cases}$ (2.12)

- Kết hợp dạng luỹ thừa: $CB(r_{1,}r_2) = \begin{cases} \sqrt[3]{\frac{\{r_1^3 + r_2^3\}}{2}} \\ \min(r_1, r_2) \qquad if\ r_1 + r_2 \ge 0 \end{cases}$ (2.13)

Output: $CB(r_{1,}r_2)$ là thứ hạng tương tự với ảnh I_Q của ảnh I_i trong CSDL ảnh E.

Đánh giá độ phức tạp của thuật toán CoEMR

Thuật toán **CoEMR** chia thành 2 pha: offline, online và được thực hiện bằng cách tổ hợp 2 xếp hạng EMR.

Đối với pha offline, độ phức tạp là:

Phân cụm K-means, xác định $n \times C$ khoảng cách giữa từng tâm cụm và các điểm dữ liệu: *O(C*n*d).*

Tính toán xếp hạng truy vấn: $O(n*C+C^3)$

trong đó C là số cụm, n số mẫu đầu vào, d số chiều véc tơ đặc trưng.

Đối với pha online, độ phức tạp là: *O(C*n*d)*

Vậy thuật toán **CoEMR-S** có độ phức tạp tương đương với độ phức tạp của EMR gốc [42].

Thuật toán CoEMR cung cấp một cách kết hợp giữa đặc trưng mức thấp và đặc trưng CNN một cách hiệu quả, từ đó làm tăng độ chính xác của kết quả truy vấn hình ảnh trong CBIR.

2.6. Thực nghiệm và đánh giá kết quả

2.6.1 Đánh giá hiệu quả của của thuật toán CoEMR

Để đánh giá hiệu quả của thuật toán CoEMR, luận án tiến hành thực nghiệm tổ hợp các xếp hạng EMR cho đặc trưng mức thấp, đặc trưng mức cao trên các bộ dữ liệu theo Bảng 1.1 cụ thể như sau:

Bảng 2.3: Bảng kết quả các phương pháp tổ hợp xếp hạng EMR

TT	Bộ dữ liệu	Bộ đặc trưng 1	Bộ đặc trưng 2	Phương pháp tổ hợp	Kết quả công bố
1	VGG60K	Mức thấp	Véc tơ nhúng	Tuyến tính	CT1, CT2
2	SIGN300HD	Mức thấp	Véc tơ nhúng	Lựa chọn	CT4
3	Leaf30	Mức thấp	CNN	Tuyến tính	CT6

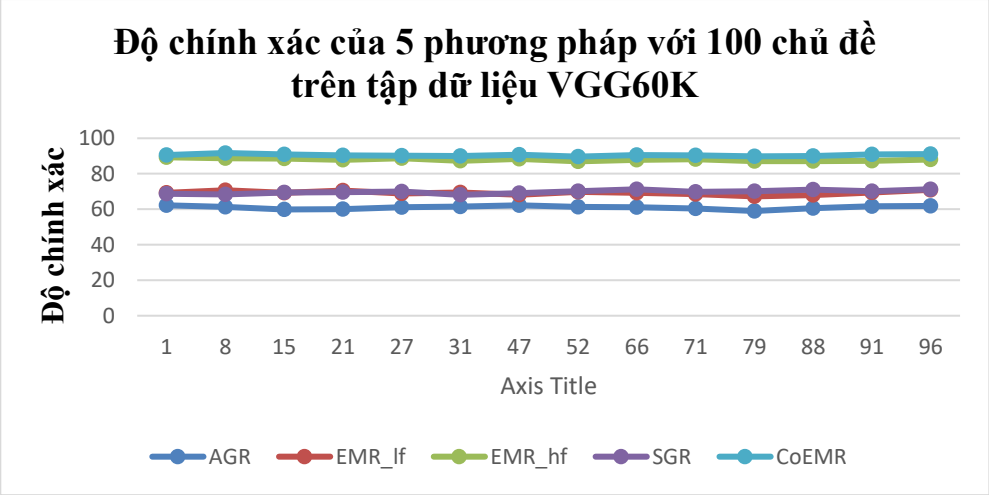
Để đánh giá của phương pháp đề xuất, luận án thực hiện so sánh kết quả khi sử dụng EMR cho xếp hạng đặc trưng mức thấp, sử dụng EMR cho xếp hạng đặc trưng CNN và sử dụng kết hợp xếp hạng theo thuật toán đề xuất CoEMR

Đối với thực nghiệm trên tập dữ liệu VGG60K, luận án chọn ngẫu nhiên 100 chủ đề (trong tổng số 500 chủ đề của , mỗi chủ đề có 120 ảnh) và thực hiện chọn ảnh truy vấn trong tất cả các chủ đề. Kết quả trả về sau tra cứu lần lượt là 10, 20, 30 ảnh. Các tham số được đặt chung cho tất cả các thực nghiệm với bộ dữ liệu VGG60K như sau: tham số a = 0.99, số điểm neo C = 5000, r = 5, nbest cho mỗi lần truy vấn n_b= 120, 20% số lượng mẫu truy vấn lấy ngẫu nhiên. Tham số kết hợp 2 xếp hạng trong CoEMR là $\alpha = 0.3$, $\beta=0.7$.

Bảng 2.6: Độ chính xác của 4 phương pháp ở 10-30 ảnh trả về sau tra cứu trên tập dữ liệu VGG60K

PP tra cứu / Số ảnh trả về	AGR (%)	EMR_lf (%)	EMR_EmbV (%)	SGR (%)	CoEMR (%)
10	80.92	83.54	90.07	90.10	**95.34**
20	61.38	70.61	88.71	68.40	**91.70**
30	56.40	60.10	82.81	61.80	**88.67**

Ở đây, dựa vào bảng có thể dễ nhận thấy rằng một số khuôn mặt "dễ" – rõ ràng, trực diện, tất cả các thuật toán thực hiện tốt, và một số khuôn mặt bị che lấp, lẫn… khó các thuật toán thực hiện cho kết quả tra cứu có độ chính xác không cao.



Hình 2.6. Độ chính xác của 5 phương pháp với 100 chủ đề trên tập dữ liệu VGG60K.

2.7. Kết hợp nhiều truy vấn ảnh trong CBIR

Hầu hết các hệ thống CBIR hiện nay sử dụng một ảnh duy nhất cho mỗi truy vấn, dẫn đến thiếu thông tin và giảm độ chính xác trong tra cứu. Sử dụng đa truy vấn (nhiều ảnh) là một hướng tiếp cận hiệu quả, giúp cải thiện độ chính xác và độ tin cậy bằng cách khai thác nhiều thông tin hơn từ cơ sở dữ liệu và phản ánh tốt hơn ý định của người dùng. Tuy nhiên, việc kết hợp kết quả từ các truy vấn vẫn là một thách thức lớn, đặc biệt với dữ liệu lớn và phức tạp.

Một số phương pháp đã được đề xuất, như học trọng số đặc trưng và xếp hạng song song, nhưng vẫn cần các thuật toán thông minh để tối ưu hóa việc tổng hợp kết quả. Luận án đề xuất phương pháp **Fusion Ranking Multi-Query (FRMQ)**, kết hợp xếp hạng từ hai truy vấn ảnh, giúp nâng cao hiệu quả CBIR. Kết quả thực nghiệm trên tập dữ liệu VGG60K cho thấy độ chính xác được cải thiện rõ rệt khi sử dụng đa truy vấn, đặc biệt với các đặc trưng CNN (EfficientNET) và vector nhúng (FaceNet).

Bảng 2.6: Độ chính xác của phương pháp sử dụng đa truy vấn ở 20 ảnh trả về sau tra cứu trên tập dữ liệu VGG60K

Đặc trưng	Sử dụng 1 ảnh truy vấn	Sử dụng 2 ảnh truy vấn
Đặc trưng mức thấp	70.61%	**76.21%**
Đặc trưng CNN (EfficientNET)	87.15%	**90.46%**
Đặc trưng vector nhúng (FaceNet)	88.71%	**92.67%**

2.8 Kết luận chương 2

Trong chương này, luận án đã trình bày chi tiết phương pháp đề xuất - kết hợp xếp hạng EMR cho đặc trưng mức thấp và mức cao (đặc trưng nhúng, đặc trưng CNN). Phương pháp này mang lại một hướng tiếp cận tổng hợp, giúp khai thác tối đa thông tin từ cả hai loại đặc trưng, khắc phục các hạn chế khi một trong hai loại đặc trưng bị suy giảm độ chính xác xếp hạng trong các cơ sở dữ liệu ảnh thực tế. Bên cạnh đó, trong chương này luận án cũng trình bày đề xuất cũng phương pháp kết hợp xếp hạng cho cả các hệ thống CBIR với nhiều ảnh truy vấn. Kết quả thực nghiệm cho thấy, việc tổng hợp kết quả xếp hạng EMR từ từng ảnh truy vấn giúp nâng cao đáng kể độ chính xác tra cứu ảnh so với hệ thống chỉ sử dụng một ảnh truy vấn.

Cụ thể, đặc trưng mức thấp có ưu điểm ít phụ thuộc vào độ phân giải ảnh nhưng lại nhạy cảm với nhiễu và chất lượng ảnh kém. Trong khi đó, đặc trưng mức cao (vector nhúng từ CNN) có khả năng phân biệt mạnh nhưng bị ảnh hưởng bởi độ phân giải và sự khác biệt trong phân bố thống kê của dữ liệu huấn luyện. Việc kết hợp tuyến tính và phi tuyến giữa hai bộ xếp hạng giúp tận dụng ưu điểm của cả hai loại đặc trưng, nâng cao độ chính xác trong truy vấn ảnh.

Luận án đã đề xuất và đánh giá thực nghiệm các thuật toán CoEMR-L, CoEMR-S và CoEMR-P trên các bài toán tra cứu ảnh dựa trên nội dung (CBIR) với một ảnh truy vấn (CT1-CT4) và nhiều ảnh truy vấn (CT8). Kết quả thực nghiệm trên các tập dữ liệu Leaf30, Sign300HD, Corel30K và VGG60K cho thấy phương pháp kết hợp này mang lại mức cải thiện đáng kể về độ chính xác (tăng khoảng 7% - 15% so với EMR gốc và các phương pháp khác cùng họ).

Luận án cũng đã mở rộng phương pháp EMR-FRMQ kết hợp tuyến tính và phi tuyến trên kết quả xếp hạng của từng ảnh riêng lẻ khi có nhiều ảnh truy vấn. Kết quả cho thấy việc tổng hợp thông tin từ nhiều ảnh truy vấn giúp hệ thống tra cứu ảnh hoạt động ổn định và chính xác hơn, đặc biệt trong các tình huống có độ nhiễu cao hoặc dữ liệu phức tạp.

Thuật toán đề xuất đã được đánh giá qua các chỉ số mAP và minP, cho thấy hiệu suất ổn định trên nhiều tập dữ liệu khác nhau. So sánh với các phương pháp hiện có như AGR, SGR và EMR, kết quả khẳng định CoEMR có độ chính xác cao hơn và độ ổn định tốt hơn, góp phần cải thiện đáng kể hiệu quả của hệ thống tra cứu ảnh.

Ngoài ứng dụng trong CBIR, phương pháp kết hợp xếp hạng EMR còn mở ra nhiều hướng nghiên

cứu tiềm năng trong các lĩnh vực khác như nhận dạng đối tượng, phân loại ảnh và các hệ thống trí tuệ nhân tạo. Điều này không chỉ có giá trị về mặt lý thuyết mà còn có ý nghĩa thực tiễn cao trong việc phát triển các hệ thống tra cứu ảnh thông minh.

Thuật toán kết hợp xếp hạng EMR giữa các bộ đặc trưng mức thấp và mức cao đã chứng minh được hiệu quả trong việc nâng cao độ chính xác tra cứu ảnh, với mức cải thiện từ 7% - 15% so với EMR truyền thống. Đồng thời, luận án cũng mở rộng mô hình xếp hạng cho nhiều ảnh truy vấn, giúp tăng tính ứng dụng thực tế của hệ thống CBIR. Tuy rất mạnh mẽ, EMR vẫn tồn tại những hạn chế, đặc biệt trong việc so sánh độ tương tự giữa các ảnh nằm ngoài cơ sở dữ liệu. Do EMR hoạt động dựa trên cấu trúc đồ thị của tập dữ liệu có sẵn, việc mở rộng ra các ảnh chưa được đưa vào CSDL trở nên khó khăn, làm giảm khả năng tổng quát hóa của mô hình. Hạn chế này gây trở ngại trong nhiều bài toán thực tế như nhận diện khuôn mặt, so khớp chữ ký và các ứng dụng bảo mật, nơi việc đánh giá độ tương tự giữa các ảnh chưa từng xuất hiện trong hệ thống là rất quan trọng.

Để giải quyết vấn đề này luận án cần xây dựng một độ đo tương tự ảnh kế thừa ưu điểm của thuật toán xếp hạng EMR (giá trị xếp hạng sự tương tự cao giữa ảnh truy vấn và các ảnh trong CSDL ảnh được lan truyền theo đồ thị quan hệ kề giữa các ảnh theo cấu trúc k-NN). Thuật toán xây dựng độ tương tự sẽ học các giá trị xếp hạng EMR giữa các cặp ảnh được chọn phù hợp của CSDL ảnh đã cho ở đó một ảnh xem như là ảnh truy vấn, mô hình học khái quát để đưa ra một độ đo tương tự ảnh giữa các cặp ảnh có thể hoàn toàn chưa được bổ sung vào CSDL ảnh đã cho. Chi tiết về cách xây dựng độ đo tương tự ảnh dựa trên các giá trị xếp hạng EMR đề xuất của luận án được trình bày ở chương 3: Xây dựng độ đo tương tự ảnh theo các giá trị xếp hạng EMR.

Chương 3
XÂY DỰNG ĐỘ ĐO TƯƠNG TỰ ẢNH THEO CÁC GIÁ TRỊ XẾP HẠNG EMR

Trong tra cứu ảnh dựa trên nội dung (CBIR), độ đo khoảng cách giữa các đối tượng đóng vai trò quan trọng trong việc xác định mức độ tương đồng giữa chúng. Thay vì sử dụng độ đo Euclid mặc định, phương pháp học độ đo khoảng cách nhằm mục tiêu xây dựng một độ đo phù hợp hơn dựa trên thông tin có trong tập dữ liệu.

Cụ thể, phương pháp học độ đo khoảng cách nhằm tìm ra một ma trận độ đo M sao cho khoảng cách theo ma trận M giữa các đối tượng cùng lớp càng nhỏ, trong khi khoảng cách giữa các đối tượng khác lớp càng lớn. Điều này tương ứng với mục tiêu tối ưu hóa độ phân biệt của các lớp trong không gian đặc trưng.

Phương pháp học độ đo khoảng cách là một khía cạnh quan trọng của học máy, đã và đang được nghiên cứu rộng rãi. Nó đóng vai trò then chốt trong nhiều bài toán như phân loại, xếp hạng, ghép cặp và tra cứu ảnh. Thông thường, các phương pháp này sử dụng các ràng buộc của các cặp hoặc ba đối tượng để học ma trận khoảng cách.

Học độ đo tương tự là phương pháp quan trọng trong lĩnh vực CBIR nhằm xây dựng không gian đặc trưng phù hợp cho bài toán. Đây là một tiếp cận dựa trên giả định rằng càng giống nhau (tương tự) thì độ tương tự càng cao và ngược lại. Các phương pháp học độ đo tương tự nhằm xác định biến đổi sao cho phản ánh đúng mối quan hệ tương tự giữa các vector đặc trưng trong tập dữ liệu, từ đó cải thiện hiệu quả phân loại và tra cứu ảnh.

EMR là một thuật toán phổ biến và mang lại hiệu quả cao trong lĩnh vực truy vấn hình ảnh dựa trên nội dung (CBIR) [36, 42, 127], EMR sử dụng phương trình lan truyền xếp hạng để tính toán vector xếp hạng cho mỗi hình ảnh truy vấn, trong đó thành phần biểu diễn mức độ tương tự giữa hình ảnh truy vấn và hình ảnh thứ trong tập dữ liệu, ngoài ra, EMR còn được sử dụng để đo độ tương tự ngữ nghĩa giữa các đối tượng, như đã được trình bày trong một số nghiên cứu [126]. So với các phương pháp đo khoảng cách thông thường chỉ xét cặp đôi, EMR tính toán độ tương đồng dựa trên cả mối quan hệ lân cận và lan truyền xếp hạng trên toàn bộ đồ thị, phản ánh đầy đủ hơn mối quan hệ giữa các mẫu dữ liệu. Điều này cho thấy tiềm năng rộng lớn của EMR trong việc khai thác các mối liên hệ tiềm ẩn giữa các điểm dữ liệu. Hình 3.1 mô tả 3 hình ảnh thuộc cùng một nhãn trong bộ dữ liệu Corel30K, nếu so sánh bằng “cảm quan” thì khó tìm ra sự tương đồng trong 3 bức ảnh này mặc dù nó có sự tương đồng về mặt ngữ nghĩa được chụp tại cùng một địa điểm ở châu Phi.

16005.jpg 16006.jpg 16007.jpg

Hình 3.1: Ba hình ảnh thuộc bộ dữ liệu Corel30K và được phân loại là có sự tương đồng ngữ nghĩa.

Tuy nhiên, EMR vẫn còn tồn tại một số nhược điểm cần khắc phục. Đầu tiên, EMR chỉ có thể xác định độ tương đồng giữa các hình ảnh trong tập dữ liệu ban đầu, không thể ứng dụng cho các hình ảnh ngoài tập dữ liệu này. Thứ hai, mỗi khi có sự thay đổi hoặc bổ sung dữ liệu mới, EMR phải xây dựng lại toàn bộ đồ thị, dẫn đến việc tính toán phức tạp và tốn kém tài nguyên, đặc biệt là khi làm việc với hệ thống dữ liệu lớn, cụ thể:

- EMR chỉ có thể xác định độ tương đồng giữa các hình ảnh trong tập dữ liệu ban đầu, không thể ứng dụng cho các hình ảnh ngoài tập dữ liệu này.

Trong thực tế, đối với các hệ thống nhận diện hoặc so khớp hình ảnh khuôn mặt hay chữ ký, việc so sánh hai ảnh không nằm trong cơ sở dữ liệu ảnh là một công việc thường xuyên và quan trọng. Các hệ thống này không chỉ cần so sánh các ảnh đã được lưu trữ trước đó mà còn phải xử lý và so sánh các ảnh mới, chưa từng xuất hiện trong cơ sở dữ liệu.

Chẳng hạn, trong một hệ thống nhận diện khuôn mặt, khi một người mới đăng nhập hoặc thực hiện một giao dịch, hệ thống phải so khớp khuôn mặt của người đó với dữ liệu đã có để xác minh danh tính. Tương tự, trong hệ thống xác thực chữ ký, khi một chữ ký mới được cung cấp, hệ thống phải so sánh nó với các mẫu chữ ký đã lưu trữ để xác định tính hợp lệ.

- Mỗi khi thay đổi hoặc bổ sung thêm dữ liệu mới, EMR phải xây dựng lại toàn bộ đồ thị, dẫn đến tính toán phức tạp, nhất là với hệ thống dữ liệu lớn.

Nhằm áp dụng EMR vào bài toán thực tế khi so sánh độ tương tự giữa hai ảnh, chúng ta có thể xem EMR như một độ đo tương tự. Từ đó, luận án đề xuất việc xây dựng mô hình EMR learning, sử dụng kỹ thuật học máy để học trực tiếp từ kết quả xếp hạng của EMR. Mô hình EMR learning có khả năng dự đoán độ tương tự cho các hình ảnh nằm ngoài tập huấn luyện, giúp nâng cao hiệu quả và độ chính xác trong việc tra cứu và nhận dạng ảnh.

Chương này sẽ trình bày chi tiết phương pháp đề xuất, bắt đầu từ việc xây dựng độ đo tương tự EMR, xây dựng tập huấn luyện IC, tập véc tơ đầu vào, đầu ra VDR và các phương pháp học máy hồi

quy. Cuối cùng, các thử nghiệm và đánh giá sẽ được thực hiện để kiểm chứng tính khả thi và hiệu quả của phương pháp đề xuất, từ đó khẳng định tiềm năng ứng dụng của EMR learning trong thực tế.

3.1 Mô hình học xếp hạng EMR

Để xây dựng mô hình học xếp hạng EMR, đầu tiên từ CSDL chứa n ảnh E, sau khi trích chọn đặc trưng ảnh chúng ta thu được CSDL n véc tơ đặc trưng ảnh E (có thể không có nhãn) và xây dựng đồ thị EMR, mô hình EMR learning là một bộ (IC, VDR, S) được khái quát như sau:

(i) $IC \subset \text{ExE} = \{(v_1, v_2) | v_1, v_2 \in E\},\ IC \neq \emptyset.$ (3.1)

Tập IC nói chung cần thỏa mãn yêu cầu sau: $\#\text{IC} \ll (\#E)^2$, nghĩa là số phần tử của IC bé hơn rất nhiều so số cặp ảnh của CSDL ảnh.

(ii) VDR là tập huấn luyện cho mô hình với đầu vào là cặp vector thuộc IC, đầu ra là mức độ tương tự được ước lượng bằng EMR.

(iii) Độ đo tương tự

Độ đo S được xây dựng từ tập VDR bằng một thuật toán học máy, S đo độ tương tự giữa các vector đặc trưng ảnh nằm ngoài CSDL vector đặc trưng ảnh. Do S là học đầu ra của xếp hạng EMR nên độ đo tương tự S của mô hình EMR learning cần thỏa mãn yêu cầu sau:

$S(v_1, v_2) \approx \text{r}[index(v_2)]$ trong đó: $\vec{r}$=EMR(v_1), $v_1, v_2 \in$ IC. (3.2)

3.2 Phương pháp xác định tập IC

Tiếp theo, luận án đề xuất phương pháp lựa chọn tập cặp vector đặc trưng ảnh IC của mô hình EMR learning, cụ thể gồm các bước như sau:

Bước 1: $IC = \emptyset$ (IC rỗng).

Bước 2: Từ cơ sở dữ liệu véc tơ đặc trưng E, chọn ra tập K các vector đặc trưng ảnh ngẫu nhiên (K có thể chọn bằng [n/2]).

Bước 3: Với mỗi vector đặc trưng ảnh $v_1 \in K$, quét toàn bộ CSDL E lấy ra n_b (ví dụ n_b=20) vector $\{v_2, i\}_{1 \le i \le n_b}$ đặc trưng gần nhất với v_1 đo theo xếp hạng của EMR.

Bước 4: Bổ sung $\{v_1, v_2, i\}_{1 \le i \le n_b}$ vào IC.

Kết quả: Ta có tập IC có số cặp vector đặc trưng ảnh là n_b*K.

3.3 Phương pháp xác định tập huấn luyện của EMR learning

Quá trình này bao gồm các bước cụ thể như sau:

Đầu tiên, với mỗi cặp véc tơ đặc trưng $(v_1, v_2) \in IC$ chúng tôi xây dựng một mẫu đầu vào bằng cách tạo ra vector khác biệt X= $\left(|v_{1,i} - v_{2,i}|\right)_{1 \le i \le \dim(v_1)}$, đây là vector chứa các giá trị tuyệt đối của sự chênh lệch giữa các thành phần tương ứng của hai véc tơ đặc trưng

Tiếp theo, giá trị đầu ra được xác định là một số cụ thể, đó là thứ hạng tương tự: $s = r[index(v_2)]$, trong đó: $\vec{r} = \text{EMR}(v_1)$. Điều này có nghĩa là chúng tôi sử dụng kết quả xếp hạng EMR của v_1 để xác định mức độ tương tự của v_2 $so\ với\ v_1$.

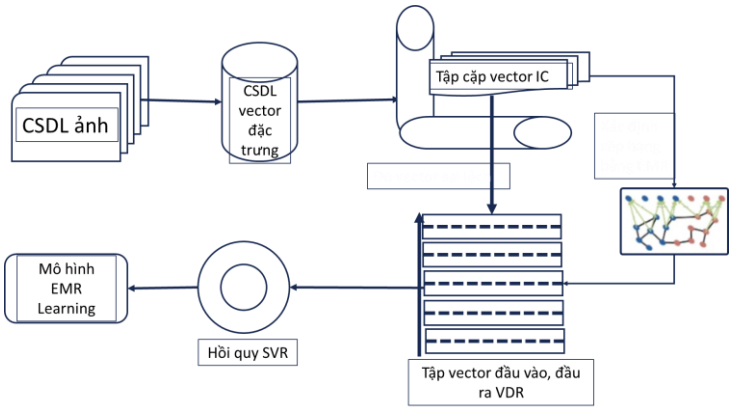
Quá trình này được lặp lại cho toàn bộ các cặp véc tơ trong IC, từ đó chúng tôi thu được tập đầu vào và đầu ra (X, s), mà luận án đặt tên là VDR (vector difference ranking).

Sau khi xây dựng tập VDR, chúng tôi chia tập này thành hai bộ dữ liệu để phục vụ quá trình huấn luyện và kiểm tra mô hình, cụ thể:

- Bộ huấn luyện: 80% dữ liệu được sử dụng để huấn luyện SVR.

- Bộ kiểm tra: 20% còn lại dùng để kiểm tra độ chính xác dự đoán của mô hình.

3.4 Xây dựng độ đo tương tự S của EMR learning dựa trên tiếp cận học máy hồi quy một đầu ra.



Hình 3.2: Các bước xây dựng dữ liệu huấn luyện cho mô hình EMR

Sau đây là thuật toán xây dựng độ đo tương tự dựa trên các giá trị xếp hạng của EMR mà chúng tôi đề xuất.

Thuật toán 3.3. EMR-LSM (Thuật toán xây dựng độ đo tương tự dựa trên các giá trị xếp hạng của EMR).

Input: $\{E_i\}_{1\le i\le n}$ là tập vector đặc trưng ảnh,

C: số lượng điểm neo (anchor points),

s: số phần tử lân cận của một vector đặc trưng,

C: số lượng điểm neo của thuật toán EMR, tham số a ∈ (0,1) (a ≈ 1),

t: tham số tỉ lệ phần trăm số lượng chọn ngẫu nhiên trong tập n phần tử.

Ω_{ML}: thuật toán học máy hồi quy.

Output: độ đo tương tự ảnh Θ được xây dựng bởi thuật toán học máy.

Bước 1: Xây dựng đồ thị của EMR.

1.1. Từ tập vector đặc trưng $\{I_{,i}\}_{1\le i\le n}$.biểu diễn các ảnh trong E.

Gọi *K-means* thu được C điểm neo $\{A_c\}_{1\le c\le C}$ của tập vector $\{IC_i\}_{1\le i\le n}$

1.2. Xác định ma trận trọng số Z kích thước C x n của EMR. Z là ma trận thưa với chỉ C x s phần tử khác 0.

1.3. Từ ma trận thưa Z tính các ma trận kề $W=(w_{ij})$, ma trận đường chéo $D = (D_{ii})$ của đồ thị EMR từ các vectors $\{IC_i\}_{1\le i\le n}$. W là ma trận thưa với chỉ *n x s* số khác 0.0.

Bước 2: Xác định tập cặp vector IC.

Gọi thủ tục 3.1 (IC-S) chúng ta xác định được tập IC.

IC=IC-S

2.2. Xây dựng tập VDR.

Gọi thủ tục 3.2 EMR-TS với IC, chúng ta xác định được tập VDR. Chia VDR thành 2 tập train và test (chia ngẫu nhiên, tỷ lệ chẳng hạn là80:20), chúng ta thu được VDRtrain, VDRtest.

Bước 3: Huấn luyện mô hình học máy hồi quy trên tập VDR bởi thuật toán ΩML sẽ thu được mô hình Θ:

$\forall v_l = (v_{l,in}, v_{l,out})$, $\Theta: v_{l,in} \mapsto v_{l,out}$, ở đây $(v_{l,in}, v_{l,out}) \in VSR_{train}$.

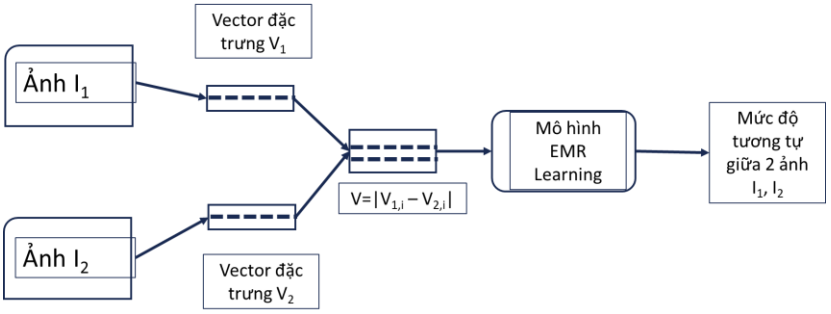
Bước 4: return Θ.

Thuật toán 3.3 EMR-LSM đã mô tả chi tiết việc xây dựng độ đo tương tự S dựa trên các giá trị xếp hạng của EMR, phần tiếp theo luận án sử dụng kết quả từ thuật toán 3.3 - mô hình EMR learning để tính toán độ tương tự ảnh.

Độ phức tạp thuật toán: Độ phức tạp thuật toán của EMR-LSM sẽ tùy thuộc thuật toán hồi quy được sử dụng.

Nếu sử dụng thuật toán hồi quy SVR, do tập VDR có $K*n_b$ phần tử, nên độ phức tạp của EMR-LSM: $O(n*C+C^3)$ (độ phức tạp của EMR) + $O(n*K*n_b*d_v)$, ở đây d_v là số chiều của vector đặc trưng.

3.5 Ước lượng độ tương tự ảnh sử dụng EMR Learning



Hình 3.3: Các bước đánh giá độ tương tự của cặp ảnh bằng EMR Learning

Sau đây là chi tiết thuật toán tính toán độ tương tự ảnh sử dụng EMR Learning.

Thuật toán 3.4. EMR-SM (Thuật toán ước lượng độ tương tự ảnh sử dụng EMR learning).

Input: 2 ảnh I_1, I_2 có thể nằm ngoài dataset ảnh E, nhưng có ngữ nghĩa cùng các ảnh của E ở giai đoạn huấn luyện.

Θ: mô hình học máy hồi quy Θ đã được huấn luyện bằng EMR-LSM.

Output: độ đo tương tự ảnh Θ được xây dựng bởi thuật toán học máy EMR-LSM

Bước 1: Tính đặc trưng kết hợp mức thấp và mức cao của I_1 và I_2.

Trích chọn đặc trưng và biểu diễn các ảnh I_1 và I_2 bằng các đặc trưng là $\{IC_i\}_{1 \le i \le 2}$.

Bước 2: Dự báo độ tương tự của I_1 và I_2.

2.1. Tính vector đầu vào của Θ: $v_{in} = \left(|IC_1[k] - IC_2[k]|\right)_{1 \le k \le d}$.

2.2. Dự báo độ tương tự bởi Θ và trả lại kết quả**:**

return $\Theta(v_{in})$

Tốc độ thực hiện xây dựng mô hình EMR learning phụ thuộc vào tốc độ thực hiện của kết quả dự báo (hồi quy SVR), trong đó SVR được xem là phương pháp học nhanh, hiệu quả dẫn đến việc xây dựng độ tương tự giữa cặp ảnh là phù hợp với bài toán thực tế.

Những đóng góp chính của đề xuất của chúng tôi như sau:

(1) Cung cấp một cách kết hợp giữa đặc trưng mức thấp và đặc trưng học sâu để tăng độ chính xác của kết quả truy xuất hình ảnh trong CBIR.

(2) Chứng minh bằng thực nghiệm trên các tập dữ liệu thực nghiệm VGG60K, Sign300HD, Leaf30 rằng, đo độ tương tự trên các sử dụng EMR-LMS và EMR-MS có hiệu quả.

3.6 Chỉ số đánh giá hiệu quả EMR Learning

Công thức tính toán Chỉ số tương quan Pearson (Correlation Coefficient) cho một mô hình học máy như sau:

$$R_{train} = \frac{\sum_{1 \le i \le \#VDR_{train}} (r_i - \bar{r})(f_i - \bar{f})}{\sqrt{\sum_{1 \le i \le \#VDR_{train}} (r_i - \bar{r})^2} \sqrt{\sum_{1 \le i \le \#VDR_{train}} (f_i - \bar{f})^2}} \quad (3.3)$$

$$R_{test} = \frac{\sum_{1 \le i \le \#VDR_{test}} (r_i - \bar{r})(f_i - \bar{f})}{\sqrt{\sum_{1 \le i \le \#VDR_{test}} (r_i - \bar{r})^2} \sqrt{\sum_{1 \le i \le \#VDR_{test}} (f_i - \bar{f})^2}} \quad (3.4)$$

3.7. Thực nghiệm và các kết quả

3.7.1 Môi trường thực nghiệm và huấn luyện EMR Learning

Để đánh giá hiệu quả của EMR Learning, dữ liệu thử nghiệm được chia thành 2 bộ:

- Bộ huấn luyện: 80% dữ liệu được sử dụng để huấn luyện mô hình EMR Learning.
- Bộ kiểm tra: 20% còn lại dùng để kiểm tra độ chính xác dự đoán của mô hình.

3.7.2 Các tham số và kết quả thực nghiệm mô hình EMR Learning

Kết quả cho thấy trên bộ huấn luyện, chỉ số tương quan đạt 0,94 và 0,92 trên bộ kiểm tra. Đồng thời MAE, RMSE cũng thấp cho thấy độ lệch nhỏ giữa kết quả thực và dự đoán. Điều này chứng tỏ EMR Learning có khả năng học và khái quát hóa tốt, mang lại độ chính xác cao trong việc dự đoán độ tương tự hình ảnh.

Bảng 3.4. Kết quả đánh giá hiệu quả EMR Learning bằng chỉ số tương quan trên các tập dữ liệu

Bộ dữ liệu	Chỉ số đánh giá	
	R_{train}	R_{test}
Sign300HD	0,94	0,92
Leaf30	0.95	0.92
Corel	0.98	0.93

3.8 Học xếp hạng với vấn đề nhận dạng nhãn

Trong trường hợp của các CSDL ảnh có một số giới hạn nhãn của ảnh, chúng ta đối diện với một vấn đề khác biệt so với truy vấn hình ảnh, trong đó đầu vào là hai ảnh và đầu ra là đánh giá về mức độ tương tự giữa chúng.

Đối lập với vấn đề này, là vấn đề nhận dạng nhãn của một ảnh đầu vào theo tập dữ liệu nhãn cho trước.

Tuy cả hai vấn đề liên quan đến hình ảnh, nhưng chúng có sự khác biệt trong cách tiếp cận và giải quyết. Bài toán nhận dạng thường được hiểu chính xác bởi người dùng, không có sự "mờ" như mức độ tương tự giữa hai ảnh trong bài toán truy vấn ảnh.

Tuy nhiên chúng ta có thể vận dụng một phương pháp truy vấn ảnh như EMR cho vấn đề nhận dạng nhãn của ảnh [CT5-CT7].

Trong [63] thì ERR được sử dụng để đánh giá độ chính xác của mô hình AGR, ERR được tính qua chỉ số Acc như sau :

$$ERR(S, N, DS) = 100 - 100 \times Acc(A, N, DS) \quad (3.8)$$

Khi đã cố định N, S và DS chúng ta sẽ chỉ viết gọn là ERR.

Để đánh giá hiệu quả của chỉ số ERR, luận án tiến hành so sánh dựa trên chỉ số ERR là tỷ lệ lỗi trung bình của các phương pháp AGR, EMR và CoEMR trên bộ dữ liệu Leaf30. Kết quả thực nghiệm ở bảng sau cho thấy CoEMR đạt độ chính xác tốt hơn, điều này làm nổi bật tính hiệu quả của việc cải thiện mô hình hóa mối quan hệ trong quá trình truy xuất ảnh.

Bảng 3.5. Bảng so sánh kết quả sử dụng chỉ số ERR

TT	Tham số	AGR	EMR	CoEMR
1.	C=1000, Nbest=100	44,26%	18.89%	**13,63%**
2.	C=1500, Nbest=100	39,43%	19.01%	**13,58%**
3.	C=2000, Nbest=100	41,55%	19.17%	**14,78%**
4.	C=2500, Nbest=100	40,49%	21.34%	**16,29%**
5.	C=3000, Nbest=100	38,33%	22.16%	**16,67%**

3.9. Kết luận Chương 3

Trong chương này, luận án đã đề xuất thuật toán EMR Learning, một phương pháp dựa trên học máy và xếp hạng EMR để xây dựng các độ đo tương tự giữa hai ảnh ngay cả khi chúng nằm ngoài cơ sở dữ liệu ảnh được sử dụng để tổ chức đồ thị xếp hạng của EMR. Phương pháp này được xây dựng dựa trên kết quả xếp hạng đa tạp EMR, kế thừa những ưu điểm của EMR đã được thể hiện trong các kết quả trình bày ở Chương 2, đồng thời khắc phục hạn chế khi so sánh ảnh ngoài CSDL.

Điểm khác biệt cốt lõi của EMR Learning nằm ở việc xây dựng độ đo tương tự ảnh thông qua học bán giám sát từ giá trị xếp hạng EMR, thay vì phải thêm ảnh vào CSDL, xây dựng đồ thị quan hệ kề và tính toán véc tơ xếp hạng như cách tiếp cận truyền thống. Cụ thể, mô hình học máy được sử dụng để mô tả mối quan hệ giữa vector đặc trưng và giá trị xếp hạng, từ đó dự đoán độ tương tự cho các cặp ảnh chưa có trong CSDL.

Thuật toán EMR Learning được triển khai theo một quy trình gồm ba thành phần chính: EMR-TS (lựa chọn tập huấn luyện), EMR-LSM (học mô hình độ tương tự), và EMR-SM (dự báo độ tương tự). Cách tiếp cận này giúp loại bỏ công đoạn xây dựng đồ thị phức tạp, đơn giản hóa thuật toán và tăng khả năng mở rộng cho dữ liệu mới. Đồng thời, phương pháp này cũng nâng cao hiệu quả trong việc ước lượng độ tương tự giữa các ảnh không thuộc tập huấn luyện ban đầu.

Khi xây dựng thuật toán học kết quả xếp hạng của EMR, EMR Learning đã vận dụng hiệu quả các thuật toán hồi quy, bao gồm SVM Regression và Random Forest Regression, trong đó tập nhãn huấn luyện được xây dựng từ cặp dữ liệu gồm vector đặc trưng ảnh và kết quả xếp hạng EMR. Trong quá trình này, CSDL vector đặc trưng có thể được cập nhật liên tục, giúp mô hình thích ứng tốt hơn với dữ liệu mới.

Phương pháp đề xuất đã được đánh giá trên các tập dữ liệu ảnh Leaf30, Sign300HD, Corel30K và VGG60K, với kết quả thực nghiệm cho thấy sự cải thiện đáng kể về độ chính xác, được đo lường qua cả đánh giá trực quan và các chỉ số khách quan. So với các phương pháp học độ đo khoảng cách khác như NCA và CMML, thuật toán EMR Learning cho thấy độ chính xác tăng đáng kể.

Bên cạnh đó, luận án cũng nghiên cứu và ứng dụng kết quả xếp hạng EMR cải tiến vào bài toán nhận dạng nhãn ảnh, theo hướng tiếp cận dựa trên khuôn mẫu (template). Kết quả thu được xác minh tính hiệu quả của phương pháp đề xuất, góp phần vào hướng nghiên cứu trong lĩnh vực tra cứu ảnh dựa trên nội dung, nhận dạng đối tượng và các ứng dụng khác trên các CSDL véc tơ trong lĩnh vực trí tuệ nhân tạo.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Luận án đã trình bày khái quát một số vấn đề cơ bản liên quan đến các bài toán trong CBIR. Trên cơ sở khảo sát và phân tích các nghiên cứu liên quan về tra cứu ảnh, luận án đã tập trung nghiên cứu hai vấn đề cơ bản:

1. Kết hợp 2 xếp hạng EMR (áp dụng cho ảnh được biểu diễn bằng nhiều bộ mô tả).

2. Học độ đo tương tự EMR - EMR learning (đo độ liên quan giữa 2 ảnh có thể nằm trong hoặc ngoài CSDL ảnh).

Với vấn đề 1, luận án đạt được 2 kết quả: (1) Kết quả chính: Đề xuất thuật toán kết hợp tuyến tính và phi tuyến các kết quả xếp hạng riêng rẽ của các bộ đặc trưng mức thấp và của bộ đặc trưng mức cao (đặc trưng nhúng, đặc trưng CNN). (2) Kết quả bổ sung: Đề xuất thuật toán kết hợp tuyến tính và phi tuyến các kết quả xếp hạng theo từng ảnh riêng rẽ trong tra cứu ảnh với nhiều hơn 1 ảnh truy vấn.

Thuật toán kết hợp tuyến tính và phi tuyến các kết quả xếp hạng riêng rẽ của các bộ đặc trưng mức thấp và của bộ đặc trưng mức cao (đặc trưng nhúng, đặc trưng CNN) bao gồm 3 thuật toán CoEMR-L, CoEMR-S, CoEMR-P (cho tra cứu với một ảnh truy vấn) [CT1-CT4], và thuật toán kết hợp tuyến tính và phi tuyến các kết quả xếp hạng theo từng ảnh riêng rẽ trong tra cứu ảnh với nhiều hơn 1 ảnh truy vấn EMR-FRMQ [CT8] (cho tra cứu với nhiều ảnh truy vấn) xếp hạng các véc tơ đặc trưng mức thấp, đặc trưng mức cao (véc tơ nhúng, đặc trưng CNN) sử dụng EMR, kết hợp chúng thành một xếp hạng duy nhất.

Các thuật toán kết hợp xếp hạng của EMR khi áp dụng vào CBIR đã tăng kết đáng kể độ chính xác tra cứu, và đã được minh chứng bằng thực nghiệm trên các tập dữ liệu ảnh lớn thông dụng.

Với vấn đề 2, luận án đạt được 2 kết quả: (1) Kết quả chính: Đề xuất thuật toán xây dựng độ đo tương tự ảnh từ các giá trị xếp hạng EMR theo tiếp cận học bán giám sát. (2) Kết quả bổ sung: Đề xuất thuật toán nhận dạng ảnh theo tiếp cận nhận dạng khuôn mẫu (template) dựa trên kết quả xếp hạng dựa trên EMR cải tiến.

Về chi tiết, luận án đã đề xuất thuật toán xây dựng độ đo tương tự ảnh từ các giá trị xếp hạng EMR theo tiếp cận học bán giám sát - EMR Learning, là thuật toán đo độ tương tự giữa 2 ảnh (có thể nằm ngoài CSDL ảnh được sử dụng để huấn luyện), thuật toán được học dựa trên kết quả xếp hạng EMR và do đó kế thừa được ưu điểm của EMR [CT5-CT7].

Thuật toán EMR learning xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR, vận dụng các kỹ thuật học hồi quy như SVM regression (hoặc có thể thay thế bằng Random forest regression..) để đạt được độ chính xác cao của mô hình học kết quả xếp hạng. Các kết quả thực nghiệm đã chứng tỏ thuật toán EMR Learning đã cải thiện được độ chính xác so với các thuật toán EMR gốc trên cùng bộ đặc trưng. Ngoài ra, luận án đã đề xuất thuật toán nhận dạng ảnh theo tiếp cận nhận dạng khuôn mẫu (template) dựa trên kết quả xếp hạng dựa trên EMR cải tiến. Các kết quả của luận án đã đóng góp vào lĩnh vực tra cứu ảnh dựa trên nội dung bằng cách đề xuất phương pháp kết hợp xếp hạng từ nhiều mô hình EMR và học độ đo tương tự từ giá trị xếp hạng, qua đó nâng cao độ chính xác của hệ thống CBIR.

Trong quá trình thực hiện luận án, chúng tôi (nghiên cứu sinh và tập thể giáo viên hướng dẫn) đã tổng hợp các công trình công bố quan trọng có liên quan trong phạm vi nghiên cứu của luận án, các đề

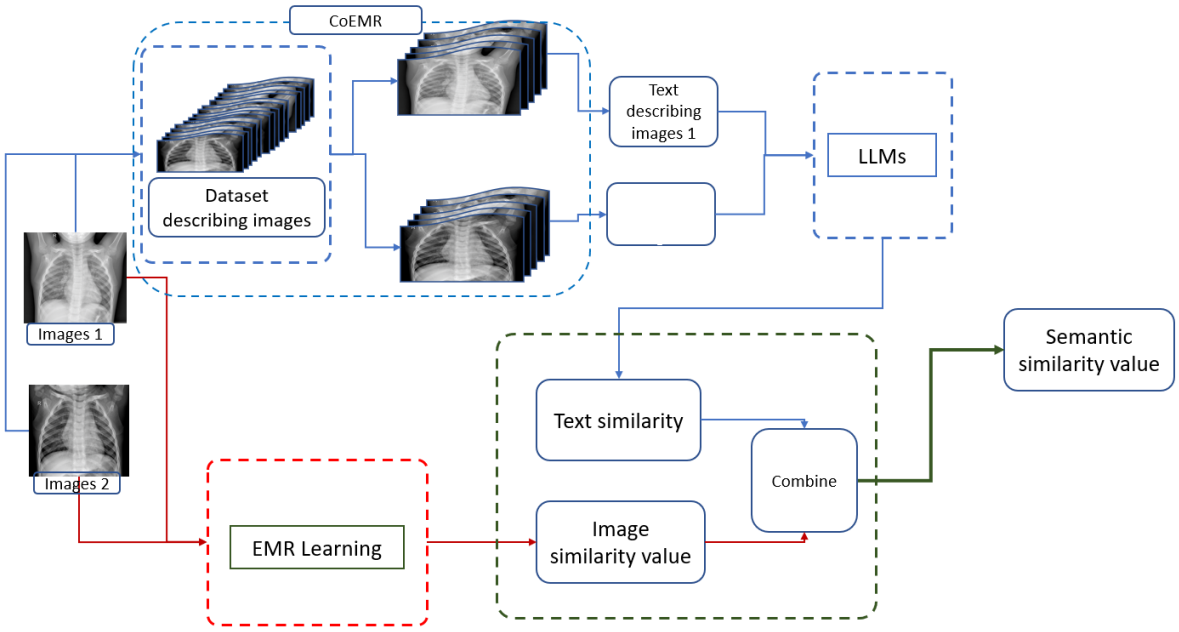
xuất của luận án đã công bố ở các tạp chí/hội thảo Quốc tế/trong nước về xây dựng các vector biểu diễn ảnh cũng như các phương pháp tìm điểm neo và cải tiến thuật toán xếp hạng EMR gốc, và đã kiểm chứng hiệu quả của các thuật toán đề xuất thông qua thực nghiệm.

Ngoài ra, luận án sẽ tiếp tục nghiên cứu và hoàn thiện thuật toán xếp hạng đa tạp, nghiên cứu mở rộng mô hình đa truy vấn, truy cứu thông tin thị giác ở mức ngữ nghĩa cao hơn dựa trên các thuật toán phân đoạn ảnh theo ngữ nghĩa (semantic segmentation).

Hướng nghiên cứu tiếp theo của luận án dự kiến là:

(1) Phát triển bộ dữ liệu huấn luyện gồm bộ ảnh huấn luyện cho bài toán tra cứu ảnh theo nội dung với bộ dữ liệu là các ảnh Y tế, ảnh Nông nghiệp cũng như các loại dữ liệu ảnh phục vụ cho các vấn đề Kinh tế/xã hội…

(2) Nghiên cứu áp dụng kỹ thuật EMR Learning cho các bộ dữ liệu khác nhau và kết hợp mô hình ngôn ngữ lớn (LLMs) tra cứu theo ngữ nghĩa nhằm nâng cao độ chính xác của các hệ thống tra cứu ảnh.



Hình 3.5: Hướng nghiên cứu kết hợp EMR Leaning và các mô hình ngôn ngữ lớn LLMs tính độ tương tự ngữ nghĩa ảnh

DANH MỤC CÁC CÔNG TRÌNH KHOA HỌC CÓ LIÊN QUAN ĐẾN LUẬN ÁN

[CT1] *"Enhancing the performance of manifold ranking in image retrieval using combined rank on low-level features and embedded vectors"*, J. Inf. Hiding Multim. Signal Process. 11(4), 172–186 (2020)

[CT2] *"Nâng cao hiệu năng của đánh hạng đa tạp EMR trong truy vấn hình ảnh sử dụng phương pháp đánh hạng đa đặc trưng ảnh"*, Kỷ yếu Hội nghị Quốc gia lần thứ XV về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), 2022.

[CT3] "*Kết hợp kết quả xếp hạng của nhiều EMR trong truy vấn ảnh theo nội dung*", Kỷ yếu Hội thảo khoa học Đại học Quốc tế Sài Gòn, 2023.

[CT4] *"Xác định độ tương tự của các ảnh chữ ký dựa trên tiếp cận xếp hạng đa tạp"*, Kỷ yếu Hội nghị Quốc gia lần thứ XVI về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), 2023.

[CT5] *"*Fusion methods of multiple EMR rankings in Content-based image retrieval*"* THE 8 TH INTERNATIONAL CONFERENCE ON APPLIED & ENGINEERING PHYSICS (CAEP-8) - 2023.

[CT6] *"Học không giám sát độ đo tương tự trên dữ liệu đa tạp của các bộ mô tả hình ảnh"*, REV-ECIT, 2023.

[CT7] *"*Nonlinear Fusion of multiple Efficient Manifold Rankings in Content-Based Medical Image retrieval*"* International Symposium on Grids & Clouds (ISGC) TAIWAN - 2024.

[CT8] *"Combining EMR rankings for multi-image queries"* International Computer Symposium (ICS) Taiwan. 2024.