

TRA CỨU ẢNH VỚI PHẢN HỒI LIÊN QUAN SỬ DỤNG THUỘC TÍNH THỪA TRONG PHÂN TÍCH PHÂN BIỆT

Đào Thị Thúy Quỳnh

Khoa Công nghệ thông tin 1, Học Viện Công Nghệ Bưu Chính Viễn Thông

Tóm tắt: Trong bài toán tra cứu ảnh theo tiếp cận học máy cho phân lớp, số chiều của không gian đặc trưng thường là lớn trong khi số mẫu huấn luyện là rất nhỏ. Vì thế, giảm chiều dữ liệu là một khâu quan trọng cho bài toán này. Có nhiều cách tiếp cận cho giảm chiều, tuy nhiên không phải cách tiếp cận nào cũng có thể nâng cao hiệu quả tra cứu ảnh. Trong bài báo này, chúng tôi sẽ đề xuất phương pháp tra cứu ảnh tra cứu ảnh với phản hồi liên quan sử dụng thuộc tính thừa dòng trong phân tích phân biệt (Linear Discriminant Analysis for Image Retrieval - LDAIR). Phần thực nghiệm cung cấp các kết quả thực nghiệm để minh chứng độ chính xác của phương pháp đề xuất.

Từ khóa: Tra cứu ảnh dựa vào nội dung, phản hồi liên quan, giảm chiều dữ liệu

I. MỞ ĐẦU

Tra cứu ảnh dựa vào nội dung (CBIR-Content Based Image Retrieval) đã nhận được nhiều sự quan tâm trong thập kỷ qua, do nhu cầu quản lý hiệu quả lượng dữ liệu đa phương tiện. Đo độ tương tự giữa hai ảnh bởi khoảng cách Euclidean trong không gian nhiều chiều thường không hiệu quả bởi vì tồn tại một khoảng trống ngữ nghĩa giữa các đặc trưng trực quan mức thấp và các khái niệm ngữ nghĩa mức cao của ảnh. Để giảm khoảng cách ngữ nghĩa này, hướng tiếp cận là bao gồm kỹ thuật học máy phân lớp vào quá trình tra cứu với phản hồi liên quan. Cơ chế này cho phép, trong danh sách kết quả trả về, người dùng gán các nhãn dương cho những ảnh tương tự ngữ nghĩa với ảnh truy vấn và cũng cho phép người dùng gán các nhãn âm cho các ảnh không tương tự ngữ nghĩa với ảnh truy vấn. Các mẫu phản hồi này được sử dụng làm tập huấn luyện cho kỹ thuật học máy phân lớp. Bởi vì số các mẫu phản hồi thu được là rất nhỏ so với chiều của không gian đặc trưng của dữ liệu ảnh, nó khó có thể thu được mô hình học phân lớp tốt. Trong tình huống này, người ta thường tìm cách, mà giảm chiều dữ liệu ảnh, để thu được

không gian chiều thấp chiều hơn [3], [5], [4], [8]. Trong không gian chiều thấp chiều hơn, các kỹ thuật học máy phân lớp được áp dụng để học các khái niệm ngữ nghĩa mức cao.

Bài toán tra cứu ảnh với phản hồi liên quan sử dụng cách tiếp cận phân lớp được thực hiện như sau. Với danh sách kết quả trả về của hệ thống tra cứu ảnh, người dùng tích chọn một số ảnh là liên quan ngữ nghĩa với ảnh truy vấn (được xem là các mẫu có nhãn dương) và tích một số ảnh khác là không liên quan ngữ nghĩa với ảnh truy vấn (được xem là các mẫu có nhãn âm). Với tập huấn luyện thu được, gồm có hai lớp là âm và dương, một kỹ thuật phân lớp được áp dụng để thu được mô hình phân lớp. Dựa vào mô hình phân lớp này, hệ thống tra cứu ảnh phân hạng các ảnh để được tập kết quả tra cứu. Chúng ta thấy rằng, bài toán tra cứu ảnh với phản hồi liên quan có một số vấn đề sau: (1) số các phản hồi của người dùng thường quá nhỏ so với chiều của không gian đặc trưng (2) số các mẫu phản hồi dương thường ít hơn số các mẫu âm rất nhiều. Trong bài báo này, chúng tôi đề xuất phương pháp tra cứu ảnh có giám sát mới, gọi là Linear Discriminant Analysis for Image Retrieval (LDAIR). Phương pháp LDAIR tận dụng thông tin của ma trận chiều phân biệt để giải quyết các vấn đề phải đối mặt ở trên.

Phần còn lại của bài báo được tổ chức như sau. Trong Phần 2, trình bày chi tiết phương pháp đề xuất. Phần 3 mô tả các thực nghiệm hiệu năng của chúng tôi và thảo luận các kết quả. Cuối cùng, chúng tôi đưa ra kết luận.

II. PHƯƠNG PHÁP TRA CỨU ẢNH ĐỀ XUẤT

Khác biệt so với các phương pháp tra cứu ảnh theo tiếp cận phân lớp đã có, mà phân lớp trên không gian chiều, LDAIR được thiết kế để tra cứu ảnh mà không bị ảnh hưởng bởi vấn đề số mẫu phản hồi nhỏ hơn so với số chiều đặc trưng và sự mất cân bằng giữa mẫu dương và mẫu âm trong tập phản hồi. Phương pháp đề xuất LDAIR có khả năng phù hợp hơn cho tra cứu ảnh với phản hồi liên quan, nơi mà cỡ lớp thường rất nhỏ.

Phương pháp LDAIR có những ưu điểm sau:

LDAIR không đòi hỏi số các mẫu dương phải đủ lớn bởi vì nó có thể cung cấp cơ chế tự động bổ sung mẫu dương vào tập huấn luyện. Nó không đòi hỏi phải huấn

Tác giả liên hệ: Đào Thị Thúy Quỳnh,

Email: quynhdao.ptit@gmail.com

Đến tòa soạn: 22/9/2021, chỉnh sửa: 31/10/2021, chấp nhận đăng: 26/12/2021.

luyện lại mô hình học chiều, mà chỉ cần áp dụng học chiều đã được học trước đó để lấy mẫu dương cho vào tập mẫu.

Cơ chế học ma trận chiều hiệu quả, mà hỗ trợ đồng thời cho hai nhiệm vụ đó là lựa chọn tập đặc trưng quan trọng và bổ sung mẫu huấn luyện dương. Mô hình học chiều chỉ học một lần duy nhất cho mỗi lần lặp phản hồi, sau đó được áp dụng cho cả hai nhiệm vụ học.

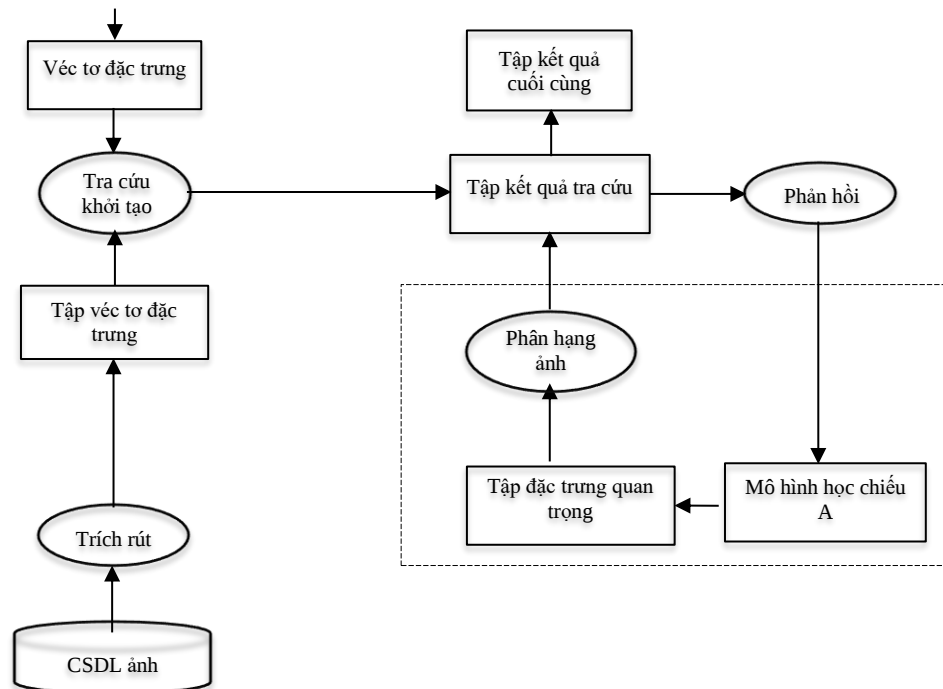
2.1. Mô hình của phương pháp

Trong phần này, chúng tôi trình bày mô hình phương pháp tra cứu ảnh được đề xuất. Từ góc nhìn của luồng dữ liệu, phương pháp tra cứu ảnh đề xuất của chúng tôi được mô tả như Hình 1. Quá trình tra cứu bắt đầu bằng việc trích rút đặc trưng của ảnh truy vấn. Các đặc trưng của các ảnh cơ sở dữ liệu được tiền trích rút theo một thủ tục trích rút đặc trưng được xác định nào đó. Sử dụng các véc tơ đặc trưng này cùng với một độ đo tương tự ảnh nào đó, độ tương tự giữa ảnh truy vấn và các ảnh cơ sở dữ liệu được đánh giá. Sau đó, một tập các ảnh lân cận với ảnh

truy vấn được lựa chọn và tập này được sắp xếp theo thứ tự giảm dần của độ tương tự để được tập kết quả tra cứu.

Người dùng phản hồi trên tập kết quả tra cứu để nhận được tập phản hồi, cũng là tập huấn luyện, trong đó các mẫu giống với ảnh truy vấn được xem là các mẫu dương và các mẫu không giống với ảnh truy vấn được xem là các mẫu âm.

Trên cơ sở tập huấn luyện này, thuật toán học chiều được thực hiện để thu được mô hình học chiều A. Sử dụng mô hình học chiều A trên tập phản hồi và sắp xếp các đặc trưng theo thứ tự giảm dần của độ quan trọng, chúng ta thu được tập đặc trưng quan trọng. Để giải quyết vấn đề cỡ mẫu nhỏ và số mẫu dương ít hơn số mẫu âm trong bài toán CBIR với phản hồi liên quan, mô hình tự động bổ sung mẫu dương thông qua áp dụng mô hình chiều mà vừa được học. Sau đó, phương pháp sẽ thực hiện phân hạng các ảnh để được tập kết quả tra cứu. Khi người dùng thỏa mãn với kết quả thì chúng ta thu được tập kết quả cuối cùng, ngược lại, quá trình này sẽ được lặp lại.



Hình 1. Mô hình phương pháp đề xuất

2.2 Mô hình học chiều lựa chọn tập đặc trưng quan trọng

Trong phần này, đầu tiên chúng tôi giới thiệu ràng buộc ℓ_∞ norm. Tiếp theo, chúng tôi trình bày một thuật toán trích rút đặc trưng mà phục vụ cho mô hình học chiều. Sau đó, chúng tôi trình bày mô hình học cho lựa chọn tập đặc trưng quan trọng mà phù hợp với bài toán tra cứu ảnh cho phản hồi liên quan. Ở đây cũng cần lưu ý rằng, phần này bao hàm các khối: “Học chiều”, “Mô hình học chiều A”, “Sắp xếp đặc trưng” và “Tập đặc trưng quan trọng” trong mô hình trên Hình 1.

Một đặc trưng có thể thuộc loại liên quan, không liên quan hoặc dư thừa. Những đặc trưng không liên quan hoặc dư thừa, được đưa vào xem xét trong quá trình phân lớp, có thể làm giảm hiệu năng phân lớp. Chiều của dữ liệu ảnh thường rất cao và thường bao gồm nhiều đặc

trung dư thừa và không liên quan [7]. Do đó, loại bỏ được các đặc trưng không liên quan và dư thừa sẽ đồng thời giảm thời gian và nâng cao hiệu năng cho phân lớp. Điều này cũng đồng nghĩa với việc tăng độ chính xác và giảm thời gian tra cứu bởi vì quá trình phân lớp là thành phần chính của hệ thống tra cứu ảnh.

Cho $A = [a_1, a_2, \dots, a_d] \in \mathbb{R}^{m \times d}$ biểu thị ma trận chiều. Chiều mẫu x sang không gian chiều thấp chiều được cho bởi $A^T x$, ℓ_∞ norm của A được cho bởi:

$$\|A\|_\infty = \max_{1 \leq i \leq d} \sum_{j=1}^m |a_{ij}| \quad (2.1)$$

Một phương pháp lựa chọn đặc trưng có thể thu được bằng việc cực tiểu chuẩn ℓ_∞ của ma trận chiều như được trình bày trong [2]. Trong đó, ràng buộc chuẩn ℓ_∞ được

sử dụng như một công cụ lựa chọn đặc trưng cho phân lớp.

Trích rút đặc trưng thông qua áp đặt ℓ_∞ lên ma trận chiều của LDA là một cách tiếp cận hiệu quả. Nó có thể lựa chọn và trích rút các đặc trưng phân biệt nhất và các đặc trưng được trích rút có thể giữ được năng lượng chính của dữ liệu gốc. Ma trận chiều được học có khả năng giải thích tốt cho các đặc trưng và có một thuộc tính phân hạng tốt cho lựa chọn các đặc trưng quan trọng nhất. Học mô ma trận chiều được mô tả ngắn gọn ở dưới.

Đầu tiên, chúng ta tính ma trận S_B và S_W

$$S_B = \sum_{i=1}^c n_i (\mu^{(i)} - \bar{x})(\mu^{(i)} - \bar{x})^T \quad (2.2)$$

$$S_W = \frac{1}{n} \sum_{i=1}^c \left(\sum_{j=1}^{n_i} (x_j^{(i)} - \mu^{(i)})(x_j^{(i)} - \mu^{(i)})^T \right) \quad (2.3)$$

Tiếp theo, chúng ta giải bài toán tối ưu sau:

$$\min_A - \frac{A^T S_B A}{A^T S_W A} + \lambda \sum_{j=1}^d \|a_j\|_\infty \quad (2.4)$$

Ở đây $A \in \mathbb{R}^{m \times d}$ ($d < m$) là ma trận chiều phân biệt. S_W và S_B là ma trận khoảng cách giữa kỳ vọng và các mẫu của mỗi lớp khoảng cách giữa các kỳ vọng của các lớp khác nhau. λ là các tham số thỏa hiệp được sử dụng để xác định độ quan trọng của các thuật ngữ liên quan. Phương pháp L-BFGS (Philipp et al., 2016) được sử dụng để giải bài toán tối ưu (2.4).

Phương pháp trích rút đặc trưng theo tiếp cận LDA bị hạn chế đó là nhạy cảm với số chiều được chọn, tức là khi số chiều d là rất nhỏ thì thông tin phân biệt không được bảo toàn. Trong khi đó, bài toán tra cứu ảnh dựa vào phản hồi liên quan là bài toán phân lớp nhị phân ($d < 2$), tức là số chiều thu được trên không gian chiều là rất nhỏ. Điều này làm cho thông tin phân biệt không được bảo toàn. Do đó, độ chính xác phân lớp của các phương pháp theo cách tiếp cận LDA, trên bài toán có cỡ lớp nhỏ, bị suy giảm. Ma trận chiều học được qua bài toán (2.4) ở trên cũng theo tiếp cận LDA, mặc dù nó đã có cải tiến để, trong không gian chiều, mất ít thông tin phân biệt nhất. Tuy nhiên với cỡ lớp là quá nhỏ, thông tin phân biệt vẫn không được đảm bảo. Do đó, với bài toán tra cứu ảnh cho phản hồi liên quan, nếu chúng ta thực hiện theo cách mà chiếu dữ liệu sang không gian chiều và thực hiện phân lớp trên không chiều, chúng ta sẽ gặp phải vấn đề cỡ lớp rất nhỏ như trên.

Để khắc phục vấn đề cỡ lớp rất nhỏ ở trên, chúng tôi đề xuất mô hình học chiều cho lựa chọn tập đặc trưng quan trọng như sau: Đầu tiên chúng tôi học phép chiếu A thông qua cực tiểu $\|A\|_\infty$ như trong bài toán (2.4), sau đó chúng tôi sử dụng mô hình chiếu A để loại đi những đặc trưng dư thừa của không gian gốc để nhận được tập đặc trưng quan trọng. Ở đây cần lưu ý rằng, mô hình học chiều cho ra kết quả là tập các đặc trưng quan trọng trên không gian gốc mà giúp khắc phục vấn đề cỡ lớp nhỏ của bài toán tra cứu ảnh dựa vào nội dung với phản hồi liên quan. Trong mô hình học chiều này, trên tập phản hồi của người dùng, đầu tiên chúng tôi học một phép chiếu A thông qua áp cực tiểu $\|A\|_\infty$ như trong bài toán (2.4).

Thuật toán học chiều cho lựa chọn tập đặc trưng quan trọng được tóm tắt trong Algorithm 1.

Thuật toán 1: Học chiều.

Input: -Ma trận mẫu huấn luyện X , ma trận nhãn Y

- Các tham số λ số đặc trưng quan trọng k

Output: - Ma trận chiều A

- Ma trận đặc trưng quan trọng X_k

Step 1: Tính S_B theo công thức (2.2); Tính và S_W theo công thức (2.3)

Step 2: Giải bài toán tối ưu (2.4) theo (Philipp et al., 2016) để có ma trận chiều A

Step 3: Tính $\|a_i\|_\infty, i = 1, 2, \dots, m$ của A

Step 4: Sắp xếp m dòng của X theo thứ tự giảm dần của $\|a_j\|_\infty$. Xây dựng X_k gồm k dòng trên đỉnh của X .

Step 5: Return A và X_k

2.3. Thuật toán tra cứu

Bài toán tra cứu ảnh với phản hồi liên quan sử dụng phân lớp phải đối mặt với vấn đề: số các phản hồi của người dùng thường quá nhỏ so với chiều của không gian đặc trưng và số các mẫu phản hồi âm thường nhiều hơn số các mẫu dương rất nhiều. Trong Thuật toán học chiều đã giải quyết được vấn đề đặt ra để nâng cao hiệu năng phân lớp. Trong tra cứu ảnh với phản hồi liên quan, tập con đặc trưng trực quan nào, mà có thể biểu thị các thuộc tính ngữ nghĩa của nhóm các mẫu dương và nhóm các mẫu âm trong xây dựng các mô hình phản hồi liên quan hiệu quả (mà gồm phân lớp và phân hạng hiệu quả) là một chủ đề quan trọng. Trong phần này, chúng tôi đề xuất thuật toán tra cứu ảnh chung, mà có thể lựa chọn ra một tập con các đặc trưng hữu ích, gia tăng mẫu phản hồi dương và không bị ảnh hưởng bởi vấn đề cỡ lớp nhỏ.

Thuật toán tra cứu ảnh đề xuất của chúng tôi gọi Thuật toán học chiều trong Bước 2.2 để giảm chiều và thu được tập đặc trưng quan trọng. Bước này giúp giải quyết vấn đề chiều cao và hỗ trợ giải quyết vấn đề các hạn chế đã nêu của bài toán tra cứu ảnh với phản hồi liên quan.

Thuật toán tra cứu ảnh đề xuất được tóm tắt trong Thuật toán 2.

Thuật toán 2: Tra cứu ảnh

Input: F : tập đặc trưng của các ảnh cơ sở dữ liệu,

q : Véc tơ ảnh truy vấn, N : số lượng ảnh trả về

Output: S : Result set

Step 1: Tra cứu ảnh với q để được tập kết quả khởi tạo và lấy N véc tơ ảnh ở top để được tập kết quả I

Step 2: Repeat

Step 2.1: Người dùng phản hồi trên tập I để có tập phản hồi RF

Step 2.2: Thực hiện thuật toán 1 để có ma trận đặc trưng quan trọng X_k

Step 2.3: Thực hiện phân hạng lại tập ảnh

Step 2.4: Lấy N ảnh ở trên đỉnh của danh sách phân hạng trong Step 2.3 làm tập ảnh kết quả S

until (User stops responding);

Step 3. Return S ;

III. ĐÁNH GIÁ THỰC NGHIỆM

Trong phần này, chúng tôi trình bày các thực nghiệm để đánh giá hiệu năng của phương pháp tra cứu ảnh đề xuất của chúng tôi. Thực nghiệm đầu tiên báo cáo so sánh phương pháp của chúng tôi với các phương pháp tra cứu ảnh tiêu biểu. Điều này là để chỉ ra rằng phương pháp đề xuất có độ chính xác cao tổng thể hơn các phương pháp còn lại. Thực nghiệm thứ hai là để nghiên cứu về hiệu quả khi loại bỏ các đặc trưng dư thừa. Thực nghiệm thứ ba là để minh chứng hiệu quả của phương pháp trong việc giải quyết bài toán vấn đề cỡ lớp nhỏ trong tra cứu ảnh với phản hồi liên quan. Thực nghiệm thứ tư kiểm tra thời gian thực hiện trung bình của phương pháp đề xuất.

3.1 Cơ sở dữ liệu ảnh

Corel Image Database

Tập ảnh COREL là một cơ sở dữ liệu ảnh chuyên nghiệp và đã được sử dụng rộng rãi để đánh giá hiệu năng của hệ thống CBIR ([9]; [10]). Do đó, chúng tôi sử dụng tập ảnh này để thực hiện một đánh giá thực nghiệm của phương pháp đề xuất của chúng tôi. Dựa vào ground truth, các ảnh không phù hợp cho tra cứu ảnh trong the Corel photo gallery bị loại đi, chúng ta nhận được một cơ sở dữ liệu ảnh Corel gồm 10,800 ảnh với 80 khái niệm ngữ nghĩa. Trong cơ sở dữ liệu ảnh này, mỗi khái niệm ngữ nghĩa bao gồm khoảng 100 ảnh hoặc hơn, và mỗi ảnh có cỡ 120×80 or 80×120 . Hình 2 chỉ ra một số ảnh mẫu trong cơ sở dữ liệu ảnh Corel. Cơ sở dữ liệu ảnh này sẽ giúp chúng ta đánh giá tự động hiệu năng tra cứu.



Hình 2. Các mẫu trong cơ sở dữ liệu ảnh COREL

Các ảnh trong cơ sở dữ liệu ảnh được biểu diễn bởi tập các đặc trưng mức thấp trong không gian 190 chiều. Như vậy, mỗi ảnh được biểu diễn bởi một véc tơ đặc trưng gồm 190 thành phần. Trong 190 thành phần này, có 102 thành phần cho đặc trưng màu và 88 thành phần cho đặc trưng kết cấu. Trong 102 thành phần cho đặc trưng màu, nó bao gồm 6 thành phần cho color moment, 32 thành phần cho color histogram, và 64 thành phần cho color correlation. Bên cạnh đó, 88 thành phần cho đặc trưng kết cấu được cấu tạo từ 48 thành phần của Gabor và 40 thành phần của Wavelet.

3.2. Thực hiện truy vấn và đánh giá

Trong phần này, chúng tôi sử dụng độ chính xác để đánh giá the hiệu quả của phương pháp tra cứu ảnh. Trong đó, top N ảnh được trả về cho người dùng, và độ chính xác được định nghĩa là tỉ số của số các ảnh liên quan trên N các ảnh được trả về. Trong thực nghiệm, phương pháp đề xuất của chúng tôi sử dụng SVM cho phân lớp. Chúng tôi cũng đánh giá các phương pháp thông qua sử dụng bốn vòng lặp phản hồi, tức là chúng tôi chia cơ sở dữ liệu ảnh thành bốn tập con có cỡ như nhau.

3.3. Thực nghiệm về hiệu năng của phương pháp đề xuất

Trong phần này, để chỉ ra độ chính xác của phương pháp đề xuất, chúng tôi so sánh nó với hai phương pháp khác, mà bao gồm Baseline, DSSA. Baseline là phương pháp tra cứu ảnh mà chỉ sử dụng độ đo khoảng cách Euclidean để đo độ tương tự giữa ảnh truy vấn và ảnh cơ sở dữ liệu. Phương pháp này được sử dụng để chỉ ra tính hiệu quả của phản hồi liên quan trong tra cứu ảnh. Chúng tôi so sánh phương pháp đề xuất với phương pháp DSSA [9] là bởi vì chúng là các phương pháp tra cứu ảnh với phản hồi liên quan truyền thống rất phổ biến. DSSA sử dụng phản hồi liên quan để học một ma trận chiếu phân biệt, mà chiếu dữ liệu từ không gian chiều cao gốc sang không gian chiếu.

Trong thực nghiệm, chúng tôi thiết lập các giá trị của các tham số cho các phương pháp như sau. Với DSSA, số chiều của không gian chiếu là 8 bởi đây là số chiều tối ưu của nó [3].

3.4. Kết quả trên tập Corel

Sau khi tra cứu lần đầu tiên, chúng ta thu được tập kết quả tra cứu khởi tạo gồm 100 ảnh. Người dùng thực hiện phản hồi trên tập 100 ảnh này để thu được tập phản hồi, mà bao gồm các mẫu dương và các mẫu âm. Các phương pháp DSSA và phương pháp đề xuất LDAIR, mà được áp dụng, để phân hạng lại các ảnh trong cơ sở dữ liệu ảnh để nhận được bốn tập kết quả tra cứu tương ứng với bốn phương pháp. Mỗi tập kết quả bao gồm 100 ảnh. Bảng 1 chỉ ra độ chính xác trung bình của ba phương pháp tại top 50 và top 100 ảnh cho lần lặp đầu tiên. Bởi vì trong tra cứu ảnh với phản hồi liên quan, vòng lặp đầu tiên là quan trọng nhất, do đó chúng tôi đưa ra độ chính xác cho vòng lặp đầu tiên. Ở đây cũng cần lưu ý rằng, độ chính xác trung bình là độ chính xác trung bình của tất cả các truy vấn, mà mỗi truy vấn tương ứng với một ảnh

trong cơ sở dữ liệu ảnh. Lấy mỗi trong toàn bộ tập ảnh cơ sở dữ liệu làm ảnh truy vấn rồi tính độ chính xác trung bình của tất cả các truy vấn này sẽ phản ánh đúng hiệu năng của phương pháp hơn bởi lẽ nó bao hàm cả các truy vấn rất xấu (có những truy vấn xấu mà kết quả trả về chỉ có một ảnh liên quan).

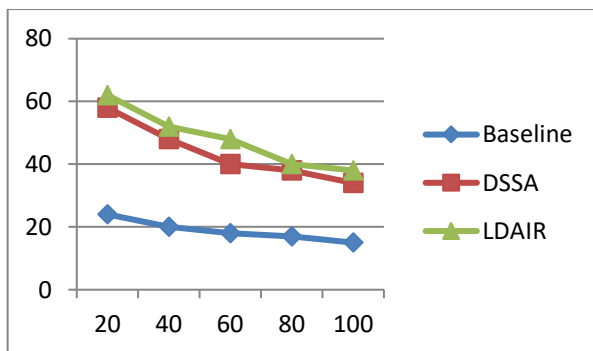
Bảng 1. Độ chính xác trung bình của ba phương pháp trên top 50 và top 100

Average precision	Baseline	DSSA	LDAIR
Top 50	15%	32.08%	38%
Top 100	15.6%	33.5%	39.6%

Các kết quả trên bảng 1 cho chúng ta thấy rằng, trong tất cả các chủ đề, phương pháp Baseline cho độ chính xác thấp nhất. Lý do của việc này là do phương pháp Baseline không có cơ chế học để giảm khoảng trống ngữ nghĩa giữa khái niệm ngữ nghĩa mức cao và đặc trưng trực quan mức thấp, nó chỉ tính khoảng cách Euclide giữa ảnh truy vấn và ảnh cơ sở dữ liệu trên không gian đặc trưng 190 chiều. Độ chính xác của phương pháp đề xuất của chúng tôi là cao nhất trong bốn phương pháp bởi vì nó giải quyết hiệu quả vấn đề giải quyết vấn đề cỡ mẫu nhỏ và bị lệch.

Bảng 1 chỉ ra độ chính xác trung bình của mỗi trong 80 khái niệm ngữ nghĩa tại top 50 và top 100 ảnh trả về cho lần lặp đầu tiên. Ở đây cũng cần lưu ý rằng, độ chính xác của một chủ đề ở đây là độ chính xác trung bình của tất cả các truy vấn, mà mỗi truy vấn tương ứng với một ảnh trong chủ đề đó. Từ Bảng 1, chúng ta nhận thấy, trong số 80 khái niệm ngữ nghĩa, chỉ có 11 mà độ chính xác của phương pháp đề xuất của chúng tôi thấp hơn ba phương pháp còn lại, trong khi độ chính xác của 69 khái niệm ngữ nghĩa của nó (phần in đậm trong Bảng 1) cao hơn các phương pháp còn lại.

Hình 3 chỉ ra the precision-scope curves cho lần lặp phân hồi thứ nhất. Các phương pháp được thực nghiệm trên các scope 20, 40, 60, 80, và 100. Trên tất cả các scope, độ chính xác của phương pháp của chúng tôi cao hơn bốn phương pháp còn lại. Như Hình 3 cho thấy độ chính xác của phương pháp đề xuất LDAIR cao hơn phương pháp DSSA.



Hình 3. So sánh độ chính xác trung bình cả ba phương pháp tại top 20,40,60,80 và 100.

IV. KẾT LUẬN

Trong bài báo này, chúng tôi đề xuất phương pháp tra cứu ảnh LDAIR mới. Bằng việc sử dụng thuộc tính thừa dòng một cách thông minh, phương pháp đề xuất có thể nâng cao độ chính xác tra cứu ảnh cho dù cỡ lớp của tập huấn luyện có thể là rất nhỏ, trong khi đó nó vẫn loại đi được các đặc trưng dư thừa. Không giống như các mô hình của các hệ thống tra cứu ảnh đã có, mà thường gắn với một độ đo tương tự hoặc mô hình học nào đó, phương pháp của chúng tôi đưa ra một mô hình mềm dẻo hơn. Kết quả thực nghiệm trên cơ sở dữ liệu ảnh COREL đã cho thấy rằng hiệu năng của phương pháp đề xuất có thể cải tiến hiệu năng cho bài toán tra cứu ảnh với phản hồi liên quan.

TÀI LIỆU THAM KHẢO

- [1] Andre B, Vercauteren T, Buchner AM, Wallace MB, Ayache N (2012). Learning semantic and visual similarity for endomicroscopy video retrieval. IEEE Transactions on Medical Imaging. 31(6):1276–88.
- [2] Mahdokht et al., (2016) A Linearly-Convergent Stochastic L-BFGS Algorithm, in Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS), Cadiz, Spain. JMLR: W&CP volume 41.
- [3] Huu, Quynh Nguyen, Dung Cu Viet, and Quynh Dao Thi Thuy. "Semantic class discriminant projection for image retrieval with relevance feedback." Multimedia Tools and Applications 80.10 (2021): 15351-15376.
- [4] Liu, Z., Liu, G., Zhang, L., Pu, J., Linear regression classification steered discriminative projection for dimension reduction. Multimed Tools Appl.
- [5] I. T. Jolliffe, Principal Component Analysis, 2nd ed. New-York: Springer-Verlag, 2002.
- [6] L. Zhang, H. Shum and L. Shao, "Discriminative Semantic Subspace Analysis for Relevance Feedback", IEEE Trans. Image Process., vol. 25, no. 3, pp. 1275-1287, 2016. [7] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, second ed. Wiley-Interscience, 2000.
- [8] Z. Lai, Y. Xu, Z. Jin, and D. Zhang, "Human gait recognition via sparse discriminant projection learning," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 10, pp. 1651–1662, Oct. 2014.
- [9] T. Zhang, D. Tao, and J. Yang, "Discriminative locality alignment," in European Conference on Computer Vision, Oct. 2008, pp. 725–738.
- [10] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 7, pp. 1088–1099, July 2006.

IMAGE RETRIEVAL WITH RELEVANCE FEEDBACK USING SPARSE ATTRIBUTE IN DISCRIMINANT FEATURE SELECTION

Abstract: Content-based image retrieval with relevant feedback uses a classification approach that faces some problems. The number of user's feedback samples is too

small compared to the dimensionality of the feature space. Besides, the number of positive feedback samples is much less than the number of negative feedback samples. In this paper, we propose a novel supervised image retrieval method, called Linear Discriminant Analysis for Image Retrieval (LDAIR). The experimental results on the two databases demonstrate that the proposed method obtains competitive precision compared with other content-based image retrieval methods.

Keywords: Content-based image retrieval (CBIR), relevance feedback, dimensionality reduction.



Đào Thị Thúy Quỳnh nhận học vị tiến sĩ Máy tính, chuyên ngành Khoa học máy tính tại Học viện Khoa học và Công nghệ, Viện hàn lâm Khoa học và Công nghệ Việt Nam. Hiện nay, là giảng viên Khoa Công nghệ thông tin 1, Học viện Công nghệ Bưu chính Viễn thông.

Lĩnh vực nghiên cứu: Trí tuệ nhân tạo, học máy, xử lý ảnh, tra cứu ảnh dựa vào nội dung.

Email: quynhdao.ptit@gmail.com