

HỌC ĐỘ ĐO KHOẢNG CÁCH GIỮA CÁC BỘ ĐẶC TRƯNG ẢNH KẾT HỢP THÔNG TIN PHẢN HỒI LIÊN QUAN TRONG TRA CỨU ẢNH DỰA TRÊN NỘI DUNG

Hoàng Xuân Trung¹, Trần Công Hùng², Ngô Hoàng Huy³, Đào Văn Tuyết²,
Đoàn Văn Hòa⁴, Trần Văn Huy⁵, Ngô Nguyên Khôi⁶

¹Đại học Kinh doanh và Công nghệ Hà Nội

²Trường Đại học Quốc tế Sài Gòn

³Trường Đại học CMC

⁴Viện Khoa học và Công nghệ Quân sự

⁵Trường Đại học Hồng Đức

⁶Đại học Conestoga College- Cambridge, ON, CA

trungvniit@gmail.com, tranconghung@siu.edu.vn, nhuhuy@cmc-u.edu.vn, tuyetdv@gmail.com,
doanvanhoa@gmail.com, tranlehuy@hdu.edu.vn, ngonguyenkhai98@gmail.com

TÓM TẮT: Phản hồi liên quan đã được sử dụng rộng rãi để thu hẹp khoảng cách ngữ nghĩa giữa đặc trưng mức thấp và các khái niệm mức cao nhằm cải thiện hiệu năng của tra cứu ảnh dựa trên nội dung. Trong bài báo này, chúng tôi đề xuất độ đo tương tự ảnh bằng việc kết hợp các giá trị tương tự của các bộ đặc trưng khác nhau dựa trên tích phân mờ. Chúng tôi so sánh phương pháp đề xuất với một số phương pháp khác và hiệu quả của nó đã được chứng minh bởi các thực nghiệm mở rộng trên tập dữ liệu Corel Photo Gallery, VGG2K và Leaf2k.

Từ khóa: Tra cứu ảnh theo nội dung, Phản hồi liên quan, Tích phân mờ, Tối ưu hàm mục tiêu, Gradient, Học độ đo khoảng cách.

I. GIỚI THIỆU

Trong bối cảnh cơ sở dữ liệu hình ảnh ngày càng phổ biến và lượng ảnh tăng đột ngột, việc xây dựng hệ thống tìm kiếm và xếp hạng ảnh trở nên cấp thiết. Các công cụ tìm kiếm như Google Lens, ByImageSearch đang thúc đẩy thương mại điện tử toàn cầu.

Phương pháp tra cứu ảnh dựa trên nội dung (CBIR) sử dụng các bộ mô tả ảnh để biểu diễn các đặc trưng cơ bản như màu sắc, kết cấu và hình dạng. Tuy có nhiều cách để xây dựng độ đo khoảng cách, độ đo tương tự, nhưng việc tạo ra độ đo tương tự thích hợp thường gặp khó khăn do khái niệm "tương tự" trong ngữ cảnh ảnh thường mơ hồ... Nếu chúng ta tiếp cận theo hướng dữ liệu đa tập, chúng ta có thể phát triển độ đo bất biến địa phương [11]. Thực tế cho thấy mọi tập hợp vector bất kỳ đều nằm trên một đa tạp cụ thể (Phụ lục A). Mặc dù đã có nhiều thuật toán phức tạp để mô tả đặc trưng mức thấp của ảnh và tính toán độ tương tự giữa chúng [4, 7-10], khả năng đo khoảng cách ngữ nghĩa giữa đặc trưng mức thấp và mức cao vẫn còn là thách thức.

Phản hồi liên quan (RF) đóng góp vào việc tăng cường hiệu suất của CBIR bằng cách sử dụng thông tin phản hồi từ người dùng để điều chỉnh trọng số đặc trưng. Thay vì tự động, điều này tạo ra độ đo tương tự phản ánh quan điểm cụ thể của người dùng và tối ưu hiệu năng tra cứu ảnh.

Chúng tôi đề xuất phương pháp học độ đo khoảng cách dựa trên tích phân mờ [2, 8]. Thông tin phản hồi người dùng dùng để học trọng số đặc trưng, áp dụng như hàm mật độ mờ trong tích phân mờ để đo khoảng cách tổng thể. Điều này cải thiện hiệu suất tra cứu ảnh dựa trên nội dung với ít vòng lặp phản hồi.

Phần còn lại của bài báo được tổ chức như sau. Mục II, giới thiệu tóm tắt khái niệm liên quan về độ đo khoảng cách có trọng số và đề xuất xây dựng độ đo tương tự dựa trên thông tin phản hồi từ người dùng và thuật toán tính toán độ đo khoảng cách phi tuyến tính. Các kết quả thực nghiệm được đưa ra và phân tích trong Mục III. Kết luận và hướng nghiên cứu tương lai được trình bày trong Mục IV.

II. NGHIÊN CỨU LIÊN QUAN VÀ PHƯƠNG PHÁP ĐỀ XUẤT

A. Độ đo khoảng cách có trọng số

Cho tập n ảnh $I = \{I_i\}_{i=1}^n \subseteq R^m$ với $I_i = \{I_{i,j}\}_{1 \leq j \leq n, 1 \leq i \leq T}$ trong đó mỗi $I_{i,t}$ là một vector đặc trưng d_i chiều, ở đây mỗi bộ đặc trưng thứ t biểu diễn về màu sắc, kết cấu hoặc hình dạng và I_q^+, I_q^- là 2 tập ảnh liên quan và không liên quan về ngữ nghĩa với ảnh tra cứu $I_q = \{I_{q,j}\}_{1 \leq j \leq n, 1 \leq i \leq T}$ được gán nhãn từ thông tin phản hồi của người dùng. Gọi

$D_t(I_{q,t}, I_{i,t})$ là giá trị khoảng cách giữa ảnh I_q và ảnh I_i đối với bộ đặc trưng thứ t được đo bởi một độ đo khoảng cách D_t nào đó.

Độ đo khoảng cách toàn cục giữa 2 ảnh (ảnh tra cứu và ảnh trong cơ sở dữ liệu) cần phải được lựa chọn dựa trên mô hình kết hợp độ đo khoảng cách D_t giữa các bộ đặc trưng t ($t=1...T$) mà người dùng mong muốn. Mô hình kết hợp phổ biến nhất là trung bình trọng số với các trọng số độ liên quan $0 \leq w_t, t = \overline{1, T}, \sum_{t=1}^T w_t = 1$ được gán cho mỗi đặc trưng thứ t .

$$D(I_q, I_i) \stackrel{\text{def}}{=} \sum_{t=1}^T w_t D_t(I_{q,t}, I_{i,t}) \quad (1.1)$$

Tuy nhiên, như đã đề cập ở trên, mô hình này coi các đặc trưng ảnh là độc lập mà không xét tới sự tương tác giữa các đặc trưng phụ thuộc vào ảnh tra cứu I_q . Do vậy, kết quả sau khi kết hợp không phản ánh đúng mong muốn của người dùng về độ tương tự giữa ảnh trong tập dữ liệu và ảnh đầu vào I_q với các ảnh I_i khác nhau [2, 8].

Trong bài báo này, độ đo khoảng cách D kết hợp các độ đo khoảng cách Euclid riêng rẽ giữa các bộ đặc trưng, được xây dựng dựa trên tích phân mờ rời rạc [2, 3, 8]. Ở đó độ đo mờ sử dụng trong tích phân mờ được xét như là độ liên quan của các tổ hợp đặc trưng tùy ý. Độ đo mờ cho phép gán độ quan trọng không những cho từng đặc trưng riêng lẻ mà còn cho tất cả các nhóm có thể của các đặc trưng, và do vậy, mô hình kết hợp trở lên linh hoạt hơn so với cách kết hợp các độ đo khoảng cách riêng rẽ có trọng số.

B. Phân hồi liên quan và xếp hạng theo EMR

Phân hồi liên quan là một phương pháp hỗ trợ cho việc học bán giám sát dữ liệu ảnh trong truy hồi ảnh dựa trên nội dung. Phân hồi tự động được thực hiện bằng cách kéo tự động N mẫu ảnh (chẳng hạn $N = 20$) có độ tương tự gần nhất với ảnh tra cứu sau mỗi lần xếp hạng. EMR (Efficient Manifold Ranking) [7] là một thuật toán tính toán việc xếp hạng nhanh và hiệu quả dựa trên việc xây dựng một đồ thị điểm neo trên toàn tập dữ liệu. EMR có thể được sử dụng như một bộ lọc để rút gọn cơ sở dữ liệu ảnh khi duyệt tập ảnh có ngữ nghĩa liên quan với ảnh đầu vào do thuật toán này có thể lọc các mẫu ảnh quá khác biệt về ngữ nghĩa, nói cách khác EMR ít khi để lọt những ảnh có độ tương đồng ngữ nghĩa với ảnh tra cứu.

C. Độ đo khoảng cách toàn cục đề xuất

Đề xuất tích phân mờ để xây dựng độ đo khoảng cách giữa hai ảnh I_1 và I_2 , là một hướng tiếp cận đa chiều và linh hoạt, mang lại nhiều lợi ích hơn so với các phương pháp truyền thống. Phương pháp này tận dụng sự tích hợp thông tin từ mọi đặc trưng (màu sắc, hình dạng, kết cấu,...), tạo ra một độ đo toàn diện về sự tương tự giữa hai ảnh.

Tích phân mờ không chỉ linh hoạt trong việc tương tác giữa các đặc trưng mà còn mô hình hóa được động thái không tuyến tính và phi tuyến, cùng với khả năng quyết định trọng số một cách linh hoạt. Kết quả đưa ra một phương pháp hiệu quả và chính xác trong việc đánh giá sự tương tự giữa ảnh, và đồng thời giữ được độ linh hoạt để áp dụng trong nhiều ngữ cảnh phân tích hình ảnh khác nhau.

Độ đo khoảng cách toàn bộ D giữa hai ảnh I_q và ảnh I_i được định nghĩa như sau:

$$C_g(I_q, I_i) \stackrel{\text{def}}{=} \sum_{k=1}^T (D_{t_k}(I_{q,t_k}, I_{i,t_k}) - D_{t_{k-1}}(I_{q,t_{k-1}}, I_{i,t_{k-1}})) * g(A_k), \quad (1.2)$$

ở đây, $\{D_t(I_{q,t}, I_{i,t})\}_{t=1}^T$ là các độ đo khoảng cách riêng rẽ đã chuẩn hóa giữa các đặc trưng bộ thứ t , $D_t(I_{q,t}, I_{i,t}) \in [0, 1]$ và $\{t_k\}_{1 \leq k \leq T}$ là hoán vị $\{1, 2, \dots, T\}$ sao cho:

$$D_{t_0}(I_{q,t_0}, I_{i,t_0}) \stackrel{\text{def}}{=} 0, \quad \forall k = \overline{1, T}: D_{t_1}(I_{q,t_1}, I_{i,t_1}) \leq \dots \leq D_{t_T}(I_{q,t_T}, I_{i,t_T}), \quad (1.3)$$

$A_k \stackrel{\text{def}}{=} \{t_k, t_{k+1}, \dots, t_T\}$ biểu diễn chỉ số các bộ đặc trưng, và $g(A_k)$ là giá trị độ đo fuzzy của tập A_k .

Bài toán học độ đo khoảng cách có trọng số dựa trên tích phân mờ được quy về xác định độ đo mờ g từ các tập dữ liệu I_q để phân lớp I_q^+, I_q^- từ các giá trị khoảng cách $\{C_g(I_q, I_i)\}_{i=1}^n$ sao cho: $C_g(I_q, I_i) \approx y_i$, ở đây,

$$y_i = \begin{cases} \min_{1 \leq t \leq T} D_t(I_{q,t}, I_{i,t}) : I_i \in I_q^+ \\ \max_{1 \leq t \leq T} D_t(I_{q,t}, I_{i,t}) : I_i \in I_q^- \end{cases} \quad (1.4)$$

Để xác định được độ đo mờ g của tích phân mờ, hàm mục tiêu bậc 2 sau được xây dựng:

$$E_q(g) = \sum_{i=1}^N \left(C_g(I_q, I_i) - y_i \right)^2 \rightarrow \min, \quad (1.5)$$

ở đây $g = \{g_A\}_{A \subseteq \{1,2,\dots,T\}}$, $g_\emptyset = 0, g_{\{1,2,\dots,T\}} = 1, g_A \leq g_B \forall A \subset B \subseteq \{1,2,\dots,T\}$

Khi xác định các trọng số (giá trị độ đo mờ) của tích phân mờ, thuật toán sử dụng phương pháp gradient để tìm hướng tối ưu nhanh nhất tiến gần tới giá trị tối ưu của hàm mục tiêu. Di chuyển theo phương gradient âm giúp giảm giá trị của hàm mục tiêu. Ngoài ra, tiếp cận gradient descent cho phép điều chỉnh tốc độ học (learning rate), ảnh hưởng đến tốc độ và ổn định của quá trình tối ưu hóa. Điều này mang lại tính linh hoạt và khả năng kiểm soát trong việc điều chỉnh quá trình học, đồng thời tránh rơi vào các điểm cực tiểu cục bộ trong quá trình lặp tối ưu hóa độ dốc của hàm.

Sử dụng phương pháp Gradient để giải hàm mục tiêu 1.5 (xác định 2^T-2 giá trị g_A), và công thức chỉnh lập theo từng $I \in NB(I_q) \cap (I_q^+ \cup I_q^-)$ là:

$$\begin{aligned} \Delta g_{I_k I_{k+1} \dots I_T} &\stackrel{\text{def}}{=} -\eta * E_q(g), \\ g_{I_k I_{k+1} \dots I_T} &= g_{I_k I_{k+1} \dots I_T} + \Delta g_{I_k I_{k+1} \dots I_T} \end{aligned} \quad (1.6)$$

ở đó $\{I_k\}_{1 \leq k \leq T}$ là hoán vị $\{1,2,\dots,T\}$ sao cho: $D_{I_1}(I_{q,I_1}, I_{i,I_1}) \leq \dots \leq D_{I_T}(I_{q,I_T}, I_{i,I_T})$, và quy ước $D_{I_0}(I_{q,I_0}, I_{i,I_0}) = 0$.

D. Thuật toán đề xuất

Thuật toán 1 trình bày cách giải và tính toán để tìm được độ đo $g(A_k)$ là giá trị độ đo fuzzy của tập chỉ số các bộ đặc trưng. Đầu vào là tập dữ liệu ảnh với các bộ đặc trưng mức thấp và đầu ra là bộ trọng số g được tính toán dựa trên từng ảnh tra cứu q bằng cách giải hàm mục tiêu $E_q(g)$ theo công thức 1.5.

Thuật toán 1. Độ đo khoảng cách phi tuyến

Input: $I = \{I_{i,t}\}_{1 \leq i \leq n, 1 \leq t \leq T}$: cơ sở dữ liệu T bộ đặc trưng của các ảnh.

T bộ vector đặc trưng mức thấp,

N : số ảnh tương tự nhất với ảnh đầu vào để hiệu chỉnh,

R_F : số lần lặp phân hồi,

$Q = \{I_{i,t}\}_{1 \leq q \leq |Q|, 1 \leq t \leq T}$ là tập ảnh điều chỉnh các giá trị độ đo mờ,

$NB(Q) = \bigcup_{I_q \in Q} NB(I_q)$, ở đây $NB(I_q) = \{I_{i,t}\}_{1 \leq i \leq n, 1 \leq t \leq T} \subset I$ là N ảnh thuộc tập ảnh I mà gần nhất về đặc trưng mức thấp so với ảnh I_q .

Output: Các bộ trọng số $G = \{g_q\}_{q \in Q}$, ở đây

$$g_q = \{g_{q,A}\}_{A \subseteq \{1,2,\dots,T\}}, g_{q,\emptyset} = 0, g_{q,\{1,2,\dots,T\}} = 1, g_{q,A} \leq g_{q,B} \forall A \subset B \subseteq \{1,2,\dots,T\}.$$

Khởi đầu. $G = \emptyset$.

Lặp. For each $I_q \in Q$:

Bước 1. Khởi tạo, $\forall t = \overline{1, T} : g_{q,t} = 0; g_{q,\{1\}} = \frac{1}{T} \forall t = \overline{1, T}, g_{q,A} = \frac{\#A}{T} \forall A \subseteq \{1,2,\dots,T\}$.

Bước 2. For each $I_i \in NB(I_q) \setminus I_q$,

2.1. For $t = \overline{1, T}$ tính khoảng cách cho bộ t : $D_{i,t} = D_t(I_{Q,t}, I_{i,t}) \hat{I} [0, 1], t = \overline{1, T}$.

2.2. Sắp xếp theo thứ tự tăng dần: $D_{i,t_k} = \text{sort}\{D_{i,t}\}_{t \in T}$ với bộ chỉ số $\{t_k\}_{1 \leq k \leq T}$ tương ứng.

Bước 3. For $l = \overline{1, R_F}$.

3.1. Tính khoảng cách C_g theo công thức tích phân mờ: $C_{g,l} = \sum_{k=1}^T (D_{i,t_k} - D_{i,t_{k+1}}) * g_{q,\{t_k, t_{k+1}, \dots, t_T\}}^{(l)}$

3.2. Xác định N giá trị nhỏ nhất của tập $\{C_{g,l}\}_{1 \leq l \leq R_F}$, tương ứng là $\{C_{g,j}\}_{1 \leq j \leq N}$

3.3. For $j = \overline{1, N}$, tính

$$y_{i_j} = \begin{cases} \min_{1 \leq t \leq T} D_t(I_{q,t}, I_{i_j,t}) : I_{i_j} \in NB(I_q) \cap I_q^+ \\ \max_{1 \leq t \leq T} D_t(I_{q,t}, I_{i_j,t}) : I_{i_j} \in NB(I_q) \cap I_q^- \end{cases}$$

3.4. Hiệu chỉnh Gradient: $\Delta g_{q,A}^{(l)} = \eta \frac{\partial J_{q,g}}{\partial g_{q,A}}$, $g_{q,A}^{(l+1)} = \min\{\max\{g_{q,A}^{(l)} + \Delta g_{q,A}^{(l)}, 0\}, 1\}$

3.5. Hiệu chỉnh g_A để đảm bảo ràng buộc đơn điệu của g : $g_A \leq g_B \forall A \subset B \subseteq \{1, 2, \dots, T\}$, for $k = \overline{1, T}$,

for each $A \in 2^{\{1, 2, \dots, T\}}$, sao cho $\#A = k$.

$$g_{q,A} = \max_{B \subset A, \#B = \#A - 1} \{g_{q,A}, g_{q,B}\}$$

3.6. Tính hàm mục tiêu $E_q(g)$ và kiểm tra điều kiện dừng:

Nếu $|E_q^{(l+1)} - E_q^{(l)}| \leq th$ thì dừng vòng lặp.

Bước 4. $g_q = \{g_{q,A}\}_{A \subseteq \{1, 2, \dots, T\}}$, $G = G \cup \{g_q\}$.

Kết thúc: return G .

Độ phức tạp của Thuật toán 1 là độ phức tạp của thủ tục xác định tập $NB(Q)$ và độ phức tạp vòng lặp phân hồi để hiệu chỉnh các trọng số tích phân mờ theo phương pháp gradient. Độ phức tạp của thuật toán sẽ phụ thuộc vào việc xác định tập $NB(Q)$ và số chiều của các bộ vector đặc trưng biểu diễn ảnh (khi tính toán khoảng cách Euclid theo từng vector đặc trưng, chiều của các vector ảnh hưởng đến tốc độ tính toán). Chúng ta xét các trường hợp để xác định tập $NB(Q)$ như sau:

Trường hợp 1: Xác định $NB(Q)$ bằng phép duyệt vét cạn tập n ảnh I , độ phức tạp của thuật toán là $O(|Q|*n*T*dmax) + O(RF*|Q|*2T)$, ở đây n là số lượng ảnh, $|Q|$ là số lượng ảnh để học trọng số mờ, T là số bộ đặc trưng biểu diễn ảnh, $dmax$ là số lớn nhất các chiều vector đặc trưng theo T bộ đặc trưng, RF là số lần phân hồi để học các trọng số tích phân mờ. Do T thường nhỏ (trong bài báo sử dụng T là 5), nên độ phức tạp của thuật toán đề xuất là $O(|Q|*n*dmax) + O(RF*|Q|)$.

Trường hợp 2: Nếu sử dụng thuật toán ANN như Faiss của Facebook [12,13] để xác định tập $NB(Q)$ thì độ phức tạp của thuật toán đề xuất là $O(|Q|*(N + dmax*\log(n))) + O(RF*|Q|)$, trong đó N là số lượng láng giềng được tìm kiếm cho mỗi ảnh của tập Q .

Trường hợp 3: Nếu sử dụng thuật toán xếp hạng đa tập EMR để xác định $NB(Q)$ thì độ phức tạp của thuật toán đề xuất là $O(n*C*dmax) + O(C^3) + O(RF*|Q|)$, trong đó C là số lượng điểm neo của đồ thị quan hệ kề giữa các vector đặc trưng khi xây dựng EMR.

III. THỰC NGHIỆM

A. Bộ dữ liệu hình ảnh thực nghiệm

Trong nghiên cứu này, chúng tôi trình bày một thuật toán mới được đề xuất và kiểm nghiệm trên ba bộ dữ liệu khác nhau, bao gồm bộ dữ liệu Oliva với 2688 ảnh (DS1), bộ dữ liệu gồm 2400 khuôn mặt được trích chọn từ tập dữ liệu VGG60K (DS2), và bộ dữ liệu lá cây Leaf2k gồm 20 lớp, mỗi lớp chứa 100 ảnh, được lựa chọn từ bộ dữ liệu bệnh

cây trồng (DS3). Việc ước tính khoảng cách giữa các ảnh khuôn mặt hoặc lá cây đối mặt với một thách thức phức tạp, bởi vì những hình ảnh này cùng thuộc một loại đối tượng và chỉ được phân biệt dựa trên các chi tiết trong hình ảnh, bao gồm cả nền và tư thế của chúng, đồng thời còn phải đối mặt với nhiều biến thể có thể xuất hiện.



Hình 1. Minh họa bộ dữ liệu thực nghiệm (a) DS1, (b) DS2 và (c) DS3

B. Trích chọn đặc trưng

Trong các thử nghiệm này chúng tôi đã trích chọn 5 đặc trưng mức thấp (low feature - LF) gồm: Color Moments, LBP, Gabor Wavelets Texture, Edge và GIST để mô tả một hình ảnh. Tất cả các đặc trưng ảnh được chuẩn hóa sao cho mỗi thành phần vector của mỗi ảnh nằm trong phạm vi $[-1, 1]$ và sau đó được nối thành một vector với số chiều là $d_f = 809$ (xem [6]).

C. Công thức đánh giá và các tham số thực nghiệm

1. Công thức đánh giá

Đối với mỗi hình ảnh tra cứu $I_q \in Q$, chúng tôi chọn $N = 290$ với DS1, $N = 120$ với DS2 và $N = 200$ với DS3 là số lượng hình ảnh trong một lớp có giá trị tương tự cao nhất. Giá trị chính xác là tỷ lệ trung bình giữa số lượng hình ảnh có liên quan trong hình ảnh N được truy xuất bởi sự giống nhau của từng hình ảnh q . Gọi tập hợp các phần tử có liên quan đến tra cứu $I_q \in Q$ là $\{d_1, d_2, \dots, d_{m_j}\}$, độ chính xác cho tất cả các tra cứu được tính như sau:

$$mAP = \left(\frac{1}{|Q|} \sum_{j=1}^{|Q|} \frac{m_j}{N} \right) * 100 \quad (1.6)$$

2. Thiết lập thực nghiệm

Quá trình học độ đo khoảng cách đề xuất trong bài báo được thực hiện với hệ số $T = 5$ tương ứng với năm loại đặc trưng mức thấp đã chọn. Tập Q được xây dựng từ 20% số mẫu ảnh I_q có trong từng bộ ảnh dữ liệu, số lần phân hồi $R_F = 1$. Với mỗi ảnh $I_q \in Q$, EMR được sử dụng như là một bộ lọc để tạo ra $NB(I_q)$ gồm 500 ảnh có mức xếp hạng cao nhất trong một bộ dữ liệu tương ứng và với từng I_q . Các tham số của EMR được cài đặt như sau $\alpha = 0,99$ và số điểm neo với bộ dữ liệu Oliva là $C = 150$, với bộ dữ liệu VGG2K là $C = 500$ và với bộ dữ liệu Leaf2k là $C = 200$.

Chúng tôi tiếp tục chạy Thuật toán 1 trên 500 ảnh đã được lọc bởi EMR để tính toán độ đo khoảng cách phi tuyến tính g thích hợp cho từng I_q và lưu g vào cơ sở dữ liệu.

Quá trình các bước kiểm tra độ chính xác của tra cứu ảnh dựa trên nội dung được thực hiện với một ảnh mới là ảnh đầu vào. EMR được sử dụng để xếp hạng với ảnh này trên toàn bộ tập dữ liệu và thu được 500 ảnh có mức hạng cao nhất. Chương trình tìm ảnh gần nhất với ảnh đầu vào trong tập ảnh tra cứu Q theo khoảng cách giữa các vector đặc trưng 809 chiều và lấy ra bộ trọng số g tương ứng. Chúng trình tiếp tục tính toán khoảng cách tương tự giữa ảnh đầu vào với mỗi ảnh trong bộ 500 ảnh trên, sử dụng bộ hệ số g , và cuối cùng xếp hạng các ảnh dựa trên những khoảng cách này.

D. Kết quả thực nghiệm và luận giải

Việc đánh giá hiệu quả truy xuất hình ảnh của thuật toán đề xuất trong bài báo này được tiến hành thông qua Thực nghiệm 1 và Thực nghiệm 2. Các thử nghiệm được tiến hành với 2 bộ dữ liệu DS1 và DS2, khi sử dụng các độ đo khoảng cách Euclid, NCA và phương pháp đề xuất.

Thực nghiệm 1. Sử dụng độ đo Euclid và DML trên tập ảnh Oliva 2688

Chúng tôi chia thực nghiệm này thành hai trường hợp: 1a - sử dụng khoảng cách Euclid và NCA cho từng đặc trưng riêng biệt; 1b - sử dụng cho bộ 5 đặc trưng mức thấp không trọng số. Thực nghiệm 1a đạt độ chính xác cao nhất là 45,66% đối với DML (NCA) trên đặc trưng GIST, trong khi sử dụng độ đo Euclid chỉ đạt 39,92%. Trong thực nghiệm 1b, độ chính xác với DML (NCA) là 44,05% cao hơn 2,76% so với việc sử dụng độ đo Euclid. Áp dụng độ đo khoảng cách phi tuyến tính đề xuất, chúng tôi thu được kết quả khá cao là 83,38%. Các kết quả thu được từ thực nghiệm chỉ ra rằng việc sử dụng các tính toán áp dụng thuật toán học đề xuất mang lại kết quả tốt hơn so với việc sử dụng độ đo khoảng cách Euclid trên từng đặc trưng và trên bộ 5 đặc trưng không trọng số.

Bảng 1a. Áp dụng độ đo khoảng cách NCA cho từng đặc trưng mức thấp riêng rẽ

Phương pháp Đặc trưng ảnh	Euclid	DML (NCA)
GIST	39,92%	45,66%
Color	25,06%	28,91%
Edge	23,35%	24,20%
Lbp	33,72%	41,57%
Gabor	29,19%	33,62%

Bảng 1b. Áp dụng cho bộ 5 đặc trưng

Phương pháp Đặc trưng ảnh	Euclid	DML (NCA)	Phương pháp đề xuất
Đặc trưng mức thấp	41,29%	44,05%	83,38%

Một thực tế là, với các độ đo khoảng cách Euclid và NCA, mọi ảnh trong tập cơ sở dữ liệu được coi là như nhau khi so sánh với ảnh I_q . Những ảnh có độ tương tự cao và cả những ảnh khác biệt cao cũng được lựa chọn để tham gia vào quá trình huấn luyện. Điều này làm giảm độ chính xác vì thuật toán cần tính toán để phân biệt các trường hợp khó xác định về độ tương tự ngữ nghĩa thay vì quan tâm đến các ảnh có độ tương tự ngữ nghĩa rất thấp. Với phương pháp của chúng tôi đề xuất, EMR được dùng như là một bộ lọc ban đầu, loại bỏ những ảnh có độ tương tự ngữ nghĩa thấp và sử dụng hữu hạn số lượng ảnh (cụ thể 500 ảnh) để tham gia vào quá trình học và tính toán bộ hệ số g phù hợp với ảnh I_q . Cuối cùng, chúng tôi kiểm tra độ chính xác tra cứu bằng cách tìm g phù hợp với ảnh đầu vào và xếp hạng thêm một lần nữa. Độ chính xác được cải thiện đáng kể (khoảng 39%) sau hai pha gồm lọc và xếp hạng.

Thực nghiệm 2. Sử dụng độ đo Euclid và DML trên tập ảnh VGG2k 2400 ảnh

Tương tự Thực nghiệm 1, chúng tôi chia cũng thực nghiệm này thành hai trường hợp: 2a sử dụng khoảng cách Euclid và NCA cho từng đặc trưng riêng biệt; 2b sử dụng cho bộ 5 đặc trưng mức thấp không trọng số. Thực nghiệm 2a đạt độ chính xác cao nhất là 19,81% đối với DML (NCA) trên đặc trưng Color, trong khi sử dụng độ đo Euclid chỉ đạt 15,02%. Trong thực nghiệm 2b, độ chính xác với DML (NCA) là 18,75% cao hơn 3,66% so với việc sử dụng độ đo Euclid. Áp dụng độ đo đề xuất tính toán được từ Thuật toán 1, chúng tôi thu được kết quả khá cao là 54,48%.

Bảng 2a. Áp dụng độ đo khoảng cách NCA cho từng đặc trưng mức thấp riêng rẽ

Phương pháp Đặc trưng ảnh	Euclid	DML (NCA)
GIST	17,03%	17,25%
Color	15,02%	19,81%

Edge	7,60%	7,66%
Lbp	10,44%	11,16%
Gabor	8,87%	9,22%

Bảng 2b. Áp dụng cho bộ 5 đặc trưng

Phương pháp Đặc trưng ảnh	Euclid	DML (NCA)	Phương pháp đề xuất
Đặc trưng mức thấp	15,09%	18,75%	54,48%

Tương tự như những luận giải ở phần Thực nghiệm 1, việc chọn lọc dữ liệu huấn luyện (lọc và loại bỏ những ảnh có độ tương tự thấp) và áp dụng độ đo khoảng cách đề xuất g đem lại độ chính xác tra cứu ảnh cao cả trên cả bộ dữ liệu VGG2K.

Thực nghiệm 3. Sử dụng độ đo Euclid và DML trên tập ảnh Leaf2k 2000 ảnh

Tương tự Thực nghiệm 1, 2 chúng tôi chia cũng thực nghiệm này thành hai trường hợp: 3a sử dụng khoảng cách Euclid và NCA cho từng đặc trưng riêng biệt; 3b sử dụng cho bộ 5 đặc trưng mức thấp không trọng số. Thực nghiệm 3a đạt độ chính xác cao nhất là 25,52% đối với DML (NCA) trên đặc trưng GIST, trong khi sử dụng độ đo Euclid chỉ đạt 22,18%. Trong thực nghiệm 3b, độ chính xác với DML (NCA) là 25,97% cao hơn 3,82% so với việc sử dụng độ đo Euclid. Áp dụng độ đo đề xuất tính toán được từ Thuật toán 1, chúng tôi thu được kết quả khá cao là 75,56%.

Bảng 3a. Áp dụng độ đo khoảng cách NCA cho từng đặc trưng mức thấp riêng rẽ

Phương pháp Đặc trưng ảnh	Euclid	DML (NCA)
GIST	22,18%	25,52%
Color	20,43%	23,18%
Edge	11,55%	12,87%
Lbp	15,42%	16,96%
Gabor	13,56%	15,29%

Bảng 3b. Áp dụng cho bộ 5 đặc trưng

Phương pháp Đặc trưng ảnh	Euclid	DML (NCA)	Phương pháp đề xuất
Đặc trưng mức thấp	22,15%	25,97%	75,56%

IV. KẾT LUẬN

Bài báo trình bày một phương pháp hiệu quả để học độ đo khoảng cách giữa các ảnh dựa trên hai pha xử lý là lọc ảnh có độ khác biệt cao với ảnh tra cứu và hiệu chỉnh kết hợp các bộ đặc trưng tương ứng với từng ảnh. Thuật toán đề xuất tính toán độ đo khoảng cách phi tuyến tính của ảnh, sử dụng tích phân mờ và dựa trên phản hồi liên quan. Các thực nghiệm đã chứng minh hiệu năng của tra cứu hình ảnh dựa trên nội dung tăng lên đáng kể khi sử dụng độ đo đề xuất. Trong thời gian tới, chúng tôi sẽ tiếp tục hướng nghiên cứu trên đề học và xác định độ đo phi tuyến tính $g(A_k)$ cho ảnh bất kỳ.

TÀI LIỆU THAM KHẢO

[1] Juan Luis Suárez-Díaz, Salvador García, Francisco Herrera, “A Tutorial on distance metric learning: mathematical foundations, algorithms, experimental analysis, prospects and challenges (with appendices on mathematical background and detailed algorithms explanation),” *arXiv:1812.05944v3*, 2020.

[2] Hichem Frigui, “Interactive image retrieval using fuzzy sets,” *Pattern Recognition Letters*, vol. 22, no. 9, July 2001, pp. 1021-1031.

[3] Anath Rau Krishnan, Maznah Mat Kasim, Engku Muhammad Nazri Engku Abu Bakar, “A Short Survey on the Usage of Choquet Integral and its Associated Fuzzy Measure in Multiple Attribute Analysis,” *Procedia Computer Science*, vol. 59, 2015, pp. 427-434.

[4] HOI, Steven; LIU, Wei; and Chang, Shih-Fu, “Semi-supervised distance metric learning for collaborative image retrieval,” *CVPR 2008: IEEE Conference on Computer Vision and Pattern Recognition*, 23-28 June 2008, Anchorage, AK: Proceedings. 1-7. Research Collection School of Information Systems.

[5] De Vazelhes, W., Carey, C. J., Tang, Y., Vauquier, N., & Bellet, A., “Metric-learn: Metric Learning Algorithms in Python,” *J. Mach. Learn. Res.*, 21, 138-1, 2020.

[6] Suárez, J. L., García, S., & Herrera, F., “pyDML: A Python Library for Distance Metric Learning,” *J. Mach. Learn. Res.*, 21, 96-1, 2020.

- [7] B. Xu, J. Bu, C. Chen, C. Wang, D. Cai and X. He, EMR, "A scalable graph-based ranking model for content-based image retrieval," *IEEE Transactions on knowledge and data engineering*, vol. 27, no. 1, pp. 102-114, 2013.
- [8] N. T. Giang, N. Q. Tao, N. D. Dung and N. H. Huy, "Learning interaction measure with relevance feedback in image retrieval," *Journal of Computer Science and Cybernetics*, vol. 32, no. 2, pp. 113-131, 2016.
- [9] H. X. Trung, D. V. Tuyet, N. H. Huy, S. Ablameyko, N. Q. Cuong and H. V. Quy, "A Novel Non-Gaussian Feature Normalization Method and its Application in Content Based Image Retrieval," *Nonlinear Phenomena in Complex Systems*, vol. 22, no. 1, pp. 1-17, 2019.
- [10] Q. H. Van, H. N. Hoang, T. D. Van, S. Ablameyko, H. T. Van, D. P. T. Kim and M. Ablameyko, "A modified Efficient Manifold Ranking Algorithm for Large Database Image Retrieval," *Nonlinear Phenomena in Complex Systems*, 2020.
- [11] Rassoul Hajizadeh, A. Aghagolzadeh, M. Ezoji, "Local distances preserving based manifold learning," *Expert Systems with Applications*, vol. 139, 2020.
- [12] W. S. T. a. P. L. Zhao, "SONG: Approximate nearest neighbor search on GPU," *2020 IEEE 36th International Conference on Data Engineering (ICDE). IEEE*, 2020.
- [13] P. a. S. A. Arulmozhi, "A comparative study of hash based approximate nearest neighbor learning and its application in image retrieval," *Artificial Intelligence Review*, 52 (2019), pp. 323-355, 2019.

Phụ lục A

Bổ đề 1: Mọi tập hữu hạn n vector m chiều, mọi số nguyên dương $d: 1 \leq d \leq \left\lceil \frac{m}{2} \right\rceil$ và $M = \{x_i = (x_{i,j})_{j=1}^n\} \subset R^m (m > 1)$ thỏa mãn

$$x_{i,k} \neq x_{j,k} \forall i \neq j, \forall 1 \leq k \leq d \quad (*) \text{ thì tồn tại một đa tạp } d \text{ chiều } D \subset R^m : x_i \in D \forall i = \overline{1, n}.$$

Chúng tôi xin dẫn bổ đề về tính tổng quát này do ông Hoàng Văn Quý cung cấp, chúng tôi xin trân trọng cảm ơn ông Hoàng Văn Quý – Khoa CNTT&TT, Trường Đại học Hồng Đức đã cho phép trích dẫn kết quả nghiên cứu chưa được công bố về dữ liệu đa tạp.

DISTANCE METRIC LEARNING BETWEEN IMAGE FEATURES COMBINED WITH RELATED FEEDBACK IN CONTENT-BASED IMAGE RETRIEVAL

Hoang Xuan Trung, **Tran Cong Hung**, Ngo Hoang Huy, Dao Van Tuyet,
Doan Van Hoa, Tran Van Huy, Ngo Nguyen Khoi

Abstract: Relevance feedback has been widely used to bridge the semantic gap between low-level features and high-level concepts in order to improve the performance of content-based image retrieval. In this paper, we propose a measure of image similarity by combining similarity values of different feature sets based on the Fuzzy integral. We compare the proposed method with some other methods and its effectiveness has been proven by extensive experiments on Corel Photo Gallery, VGG2K and Leaf2k dataset.

Keywords: Content-based image retrieval (CBIR), Relevant feedback, Fuzzy integral, Optimizing objective function, Gradient, Distance metric learning (DML).