

BỘ CÔNG THƯƠNG  
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

HOÀNG VĂN QUÝ

NGHIÊN CỨU CẢI TIẾN THUẬT TOÁN XẾP HẠNG  
ĐA TẠP TRONG TRA CỨU ẢNH

LUẬN ÁN TIẾN SĨ CÔNG NGHỆ THÔNG TIN

Hà Nội, năm 2023

BỘ CÔNG THƯƠNG  
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

HOÀNG VĂN QUÝ

NGHIÊN CỨU CẢI TIẾN THUẬT TOÁN XẾP HẠNG  
ĐA TẬP TRONG TRA CỨU ẢNH

Ngành: Công nghệ thông tin

Mã số: 9480201

LUẬN ÁN TIẾN SĨ CÔNG NGHỆ THÔNG TIN

NGƯỜI HƯỚNG DẪN KHOA HỌC:

- TS. NGÔ HOÀNG HUY
- TS. NGUYỄN THẾ CƯỜNG

Hà Nội, năm 2023

## MỞ ĐẦU

### 1. Tính cấp thiết của đề tài

Với sự phát triển của công nghệ Internet và sự phổ biến của các thiết bị kỹ thuật số, việc chụp ảnh hoặc trích xuất đối tượng trong ảnh mà chúng ta quan tâm đã trở nên dễ dàng và tiện lợi. Thực tế là số lượng ảnh được tạo ra hàng ngày trong cuộc sống của chúng ta là rất lớn. 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 vào 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. Chưa khai thác tốt tính đa biểu diễn của ảnh bằng nhiều bộ đặc trưng. Khi kết hợp nhiều bộ đặc trưng, chiều vector biểu diễn ảnh có thể rất cao dẫn đến khó khăn trong tính toán khoảng cách và xác định điểm neo (như trong EMR, SSG).

iii. Lựa chọn các điểm neo chất lượng bằng các thuật toán phân cụm dựa trên tâm (như Fuzzy C-Means, phân cụm Gaussian Mixture Model) trên các tập dữ liệu lớn có số chiều vector cao rất khó khăn, trong khi phân cụm K-means thì không biểu diễn được trường hợp một vector đặc trưng có nhiều hơn một điểm neo đại diện nó.

iv. Không kết hợp được với thuật toán lân cận xấp xỉ (ANN) để giảm việc tính toán trực tiếp tất cả các khoảng cách giữa cơ sở dữ liệu vector đặc trưng ảnh và tập các điểm neo đại diện. Khi kết hợp với kỹ thuật ANN, việc thay thế phân cụm K-means bởi các thuật toán phân cụm dựa trên xác định tâm cụm như FCM mới trở nên khả thi trong quá trình ngoại tuyến (offline) xây dựng các đồ thị quan hệ kề.

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 các hạn chế của xếp hạng đa tạp trong tra cứu ảnh dựa vào nội dung, luận án chọn đề tài: *Nghiên cứu cải tiến thuật toán xếp hạng đa tạp trong tra cứu ảnh.*

### 2. Mục tiêu của luận án

**Mục tiêu chung của luận án:** Nâng cao độ chính xác của tra cứu ảnh dựa trên cải tiến thuật toán xếp hạng đa tạp.

#### **Mục tiêu cụ thể của luận án:**

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

- Nghiên cứu cải tiến nội tại của thuật toán xếp hạng đa tạp hiệu quả EMR với phương pháp chọn điểm neo mới.

- Kết hợp đặc trưng mức thấp và đặc trưng mức cao trong biểu diễn ảnh để nâng cao độ chính xác trong tra cứu ảnh.

### 3. Đối tượng nghiên cứu của luận án

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

- Phương pháp xếp hạng đa tạp trong tra cứu ảnh dựa vào nội dung, các kỹ thuật và những thách thức trong xếp hạng đa tạp.

- Các kỹ thuật biểu diễn ảnh với đặc trưng mức thấp, đặc trưng CNN và kết hợp đặc trưng mức thấp và đặc trưng CNN (đặc trưng ảnh được trích rút từ mạng học sâu).
- Kỹ thuật lập trình song song cho phân cụm FCM cải tiến và xếp hạng đa tạp trên tập dữ liệu lớn có số chiều rất cao.
- 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 hiệu quả trong tra cứu ảnh.
- Nghiên cứu phương pháp cải tiến thuật toán xếp hạng đa tạp hiệu quả (EMR) thông qua việc tìm điểm neo bằng thuật toán phân cụm mờ FCM cải tiến.
- Nghiên cứu trích rút đặc trưng ảnh từ mạng CNN tiền huấn luyện, kết hợp đặc trưng mức thấp và đặc trưng CNN trong biểu diễn ảnh với số chiều rất cao.
- 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

Với mục tiêu nâng cao độ chính xác của việc tra cứu ảnh bằng phương pháp xếp hạng đa tạp, luận án đạt được hai đóng góp chính:

Thứ nhất, cải tiến thuật toán xếp hạng đa tạp EMR bằng cách sử dụng thuật toán phân cụm mờ *lvdc-FCM* để tìm điểm neo. Trên cơ sở đó, đề xuất thuật toán *EMR-(lvdc-FCM)* áp dụng hiệu quả trong CBIR [CT1, CT2, CT4].

Thứ hai, nâng cao hiệu quả tra cứu ảnh bằng phương pháp kết hợp đặc trưng cấp thấp và cấp cao trích xuất từ mạng CNN. Chi tiết đóng góp bao gồm: (i) Đề xuất phương pháp tinh chỉnh mạng CNN để trích rút đặc trưng trước khi kết hợp với đặc trưng mức thấp; (ii) Đề xuất thuật toán HD-EMR dựa trên các điểm neo ước lượng bằng FCM cải tiến kết hợp ANN (thuật toán *LDM-FCM*) để tra cứu ảnh không cần sử dụng các phương pháp giảm chiều dữ liệu [CT3, CT5, CT6].

Ngoài ra, luận án còn đưa ra bổ đề mới về tính tổng quát của dữ liệu đa tạp trên các tập hữu hạn vector, cung cấp cơ sở lý luận cho ứng dụng xếp hạng đa tạp trong tra cứu ảnh theo nội dung. Trong quá trình nghiên cứu, luận án đề xuất sử dụng kiến trúc GPU để thực hiện hiệu quả thuật toán phân cụm *LDM-FCM*.

#### 6. Bố cục của luận án

Luận án được tổ chức thành ba chương:

Chương 1: Tra cứu ảnh dựa vào nội dung và xếp hạng đa tạp trong CBIR.

Chương 2: Nâng cao độ chính xác tra cứu ảnh với xếp hạng đa tạp cải tiến.

Chương 3: Nâng cao hiệu quả tra cứu ảnh sử dụng đặc trưng kết hợp mức thấp và mức cao

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

## Chương 1

# TRA CỨU ẢNH DỰA VÀO NỘI DUNG VÀ XẾP HẠNG ĐA TẬP TRONG CBIR

## 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.

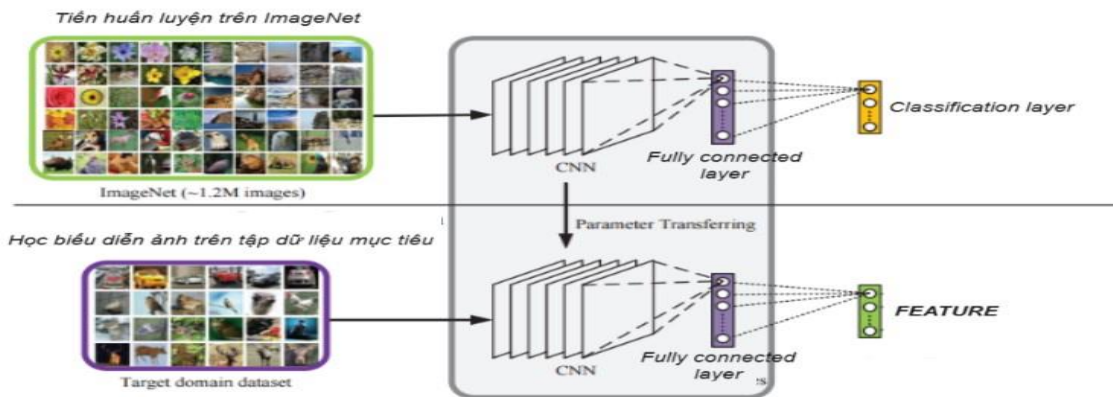
## 1.2. Biểu diễn ảnh bằng vector đặc trưng

### 1.2.1. Đặc trưng mức thấp của ảnh

Trong chương này luận án 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 CNN:** Là đặc trưng được trích rút ở tầng cuối (trước tầng phân lớp) của mạng CNN.



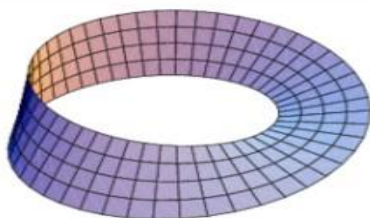
**Hình 1.5. Mô hình trích rút đặc trưng ảnh bằng mô hình học sâu [53]**

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.

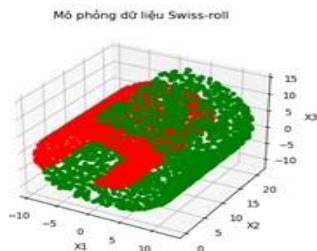
## 1.3. Dữ liệu đa tạp

### 1.3.1. Các khái niệm đa tạp

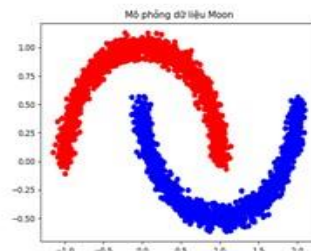
Khái niệm đa tạp được hiểu như sau [55, 56]: Một đa tạp tôpô  $n$  chiều là một không gian tôpô mà mỗi điểm có lân cận đồng phôi với tập con mở của  $\mathbb{R}^n$ , nói một cách khác, là không gian tôpô tách được với mỗi điểm của nó có một lân cận đồng phôi với một tập mở trong không gian Euclide  $n$  chiều. Như vậy có thể hiểu: Đa tạp chính là khái niệm toán học mở rộng của đường và mặt.



Dải Mobius là một đa tạp hai chiều



Cuộn giấy Thụy Sĩ (Swiss-roll) là đa tạp 2 chiều



2-moon là đa tạp 2 chiều

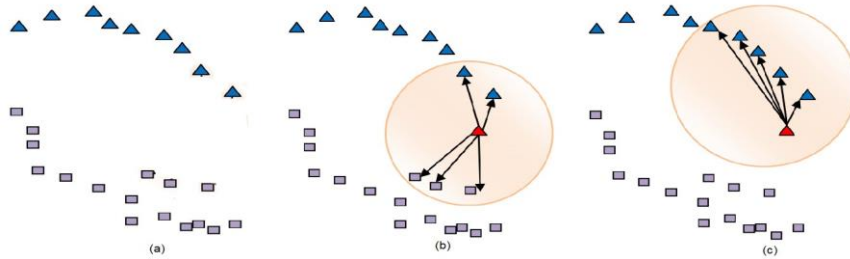
**Hình 1.6. Mô phỏng các không gian đa tạp [56]**

**1.3.2. Dữ liệu đa tạp**

Các hình ảnh "có ý nghĩa" thường tạo thành một đa tạp trong không gian của tất cả các hình ảnh: Những hình ảnh này không phân bố một cách ngẫu nhiên trong không gian đa chiều mà thay vào đó, chúng tạo thành một cấu trúc được gọi là đa tạp.

**1.4. Xếp hạng đa tạp**

Rất nhiều dữ liệu thực tế được biểu diễn một cách phù hợp trong không gian cấu trúc đa tạp cục bộ hơn là trong những không gian khác. Hình 1.9(a) biểu diễn tập hợp các điểm mẫu dữ liệu phân bố theo cấu trúc 2-moon.



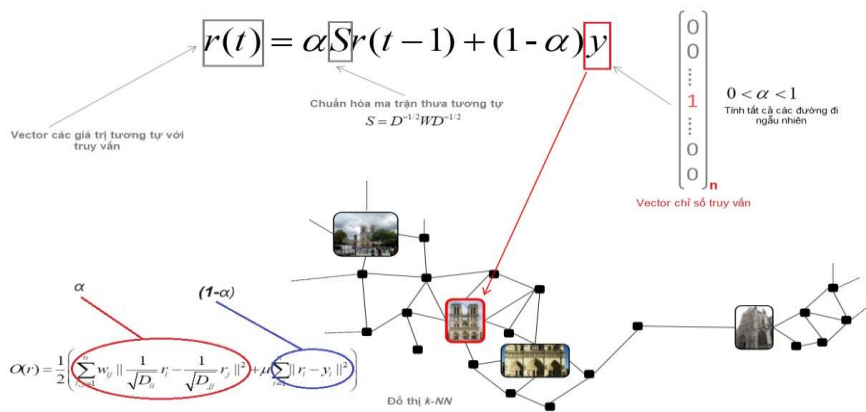
**Hình 1.9. Mô tả xếp hạng trên tập dữ liệu 2-Moon**

**1.4.1. Xếp hạng đa tạp cơ bản**

**1.4.2. Xếp hạng đa tạp hiệu quả**

**1.5. Tra cứu ảnh với xếp hạng đa tạp**

Tra cứu dùng thuật toán MR trong CBIR:



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

**1.6. Độ đo tương tự và tìm kiếm xấp xỉ láng giềng gần nhất**

**1.6.1. Độ đo tương tự**

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

**Bảng 1.2. Một số độ đo khoảng cách và công thức**

Độ đo	Công thức tính
Manhattan (Taxicab/City Block) (Khoảng cách $L_1$ )	$d(x, y) = \sum_{i=1}^k  x_i - y_i $ (k là số block)
Chebyshev	$d(x, y) = \max_i ( x_i - y_i )$
Minkowski	$d(x, y) = (\sum_{i=1}^D  x_i - y_i ^p)^{\frac{1}{p}}$
Cosine	$d(X, Y) = 1 - \cos \theta = 1 - \frac{X \cdot Y}{\ X\  \cdot \ Y\ }$

### 1.6.2. Tìm kiếm xấp xỉ láng giềng gần nhất (ANN)

Trong các hệ thống CBIR, thay vì đối sánh ảnh truy vấn với toàn bộ CSDL, việc đối sánh chỉ thực hiện trên một tập con nhất định (tập các tâm cụm chẳng hạn) nên việc tìm kiếm theo ANN có thể không đưa ra kết quả chính xác 100% như tìm kiếm chính xác, mà sẽ trả về kết quả gần đúng nhất trong thời gian ngắn nhất. Điều này là do thuật toán tìm kiếm ANN làm việc bằng cách xác định một tập hợp các điểm gần nhất và không đảm bảo tìm kiếm chính xác nhất. Tuy nhiên, với một bộ chỉ mục tốt hoặc một thuật toán gom cụm tốt và một thuật toán tìm kiếm phù hợp, kết quả của tìm kiếm ANN vẫn có thể rất chính xác và hiệu quả trong nhiều trường hợp.

### 1.7. 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

$$Pr = \frac{E_q}{D_q}$$

vấn, được tính theo công thức:

Độ triệu hồi là tỉ số giữa số ảnh liên quan được truy vấn với toàn bộ số ảnh có liên

$$Re = \frac{E_q}{A_q}$$

quan trong CSDL:

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:

$$ARP = average\left(\sum P_i\right) \quad (1.25)$$

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:

$$MAP = \frac{1}{Q} \sum_{q=1}^Q AP_q \quad (1.26)$$

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.

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].

### 1.8. 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
Logo-2K+	22725	303
Corel30K	31695	306
VGGFACE2-S	60000	500

### 1.9. Kết luận chương 1

Trong chương 1 luận án đã trình bày tổng quan về tra cứu ảnh dựa vào nội dung với xếp hạng đa tạp. Bên cạnh đó luận án phân tích về dữ liệu đa tạp trong tra cứu ảnh. Từ các phân tích và đánh giá về xếp hạng đa tạp trong tra cứu ảnh luận án đã phát hiện và đề xuất,

chứng minh Bổ đề 1 là đảm bảo toán học cho các nghiên cứu về tra cứu ảnh theo hướng tiếp cận dữ liệu đa tạp như học đa tạp và xếp hạng đa tạp.

Cũng ở trong chương này luận án tập trung giới thiệu và phân tích các thuật toán xếp hạng đa tạp (MR, EMR) trong CBIR, qua đó rút ra một số hạn chế của các thuật toán trên như sau:

- Các hệ thống CBIR sử dụng theo hướng tiếp cận xếp hạng đa tạp thường mới sử dụng các đặc trưng mức thấp dẫn đến độ chính xác chưa cao.

- Việc triển khai toán xếp hạng đa tạp như MR không khả thi áp trên các tập CSDL lớn với số chiều vector đặc trưng ảnh rất cao (với các hạng tầng tính toán không chuyên dụng - chẳng hạn các máy tính không hỗ trợ GPU).

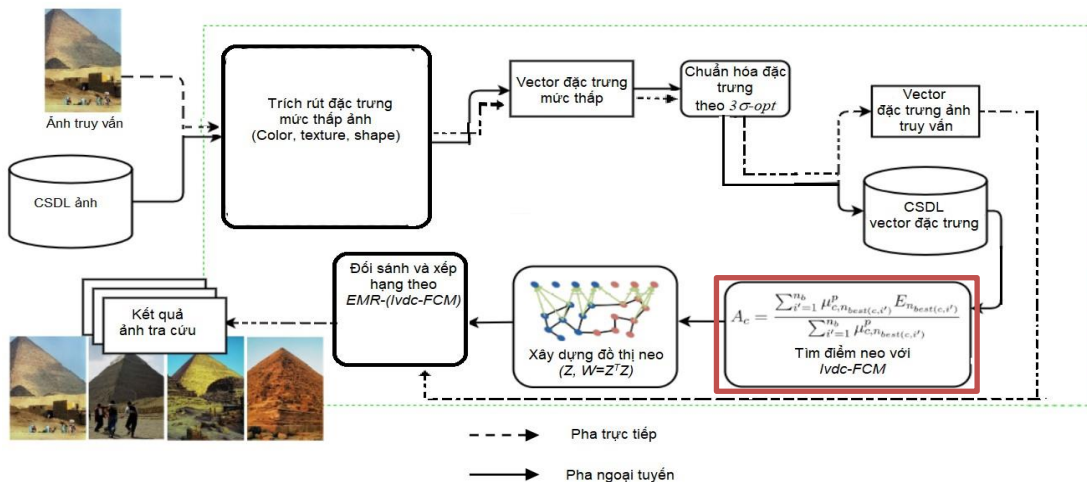
- Để xếp hạng được các CSDL ảnh lớn, các thuật toán xếp hạng đa tạp cải tiến như EMR, SGR, AGR... thường sử dụng phương pháp tìm các điểm neo (mỗi điểm đại diện cho một nhóm vector) được chọn bằng phương pháp phân cụm K-means và chưa có nghiên cứu nào đề xuất phương pháp tìm điểm neo mới hiệu quả hơn.

Chương 2 và Chương 3 của luận án này sẽ tập trung cải tiến thuật toán EMR để nâng cao độ chính xác cho các hệ thống CBIR.

## Chương 2 NÂNG CAO ĐỘ CHÍNH XÁC TRA CỨU ẢNH VỚI XẾP HẠNG ĐA TẬP CẢI TIẾN

### 2.1. Giới thiệu

### 2.2. Kiến trúc CBIR đề xuất



**Hình 2.1. Mô hình hệ thống đề xuất CBIR với đặc trưng mức thấp sử dụng EMR-(lvdc-FCM)**

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

+ *Pha ngoại tuyến:*

+ *Pha trực tuyến:*

### 2.1. Xây dựng đồ thị neo trong thuật toán EMR

Bước đầu tiên của việc xây dựng cấu trúc đồ thị neo bao gồm việc xây dựng  $C$  điểm neo đại diện cho các điểm trong CSDL với số chiều vector được bảo toàn. Đây có thể là  $C$  điểm được lấy mẫu ngẫu nhiên các điểm từ không gian dữ liệu đầu vào  $X \in \mathbb{R}^{n \times D}$  hoặc được xây dựng theo những cách phức tạp hơn như phân cụm. Trong bước này, luận án sử dụng

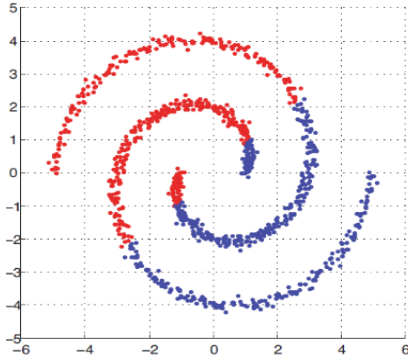


tâm cụm thu được từ thuật toán *lvdc-FCM* (thuật toán 2.1) làm điểm neo. Thuật toán *lvdc-FCM* là một phương pháp phân cụm không giám sát mềm, tạo ra  $C$  tâm cụm ở đó mỗi điểm dữ liệu có thể có nhiều hơn một tâm cụm đại diện. Từ thuật toán 2.1 thu được tập tâm cụm  $\{A_c\}$  và ta gán  $U = \{A_c\}_{2 \leq c \leq C}$ .

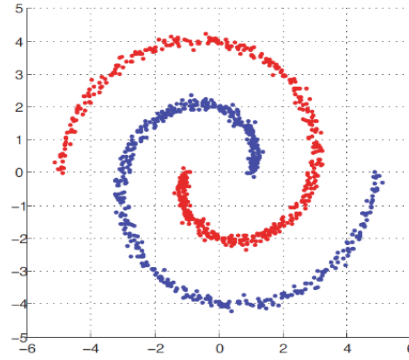
### 2.2.1. Phương pháp tìm điểm neo bằng thuật toán *lvdc-FCM*

Trong các nghiên cứu về EMR [15, 30, 29, 70], phương pháp tìm điểm neo ở bước một luôn được thực hiện thông qua chọn ngẫu nhiên trong tập dữ liệu gốc hoặc dùng thuật toán phân cụm K-means. Do hạn chế của K-means là tìm ra tập tâm cụm có dạng hình cầu và mỗi điểm dữ liệu chỉ thuộc vào 1 tâm cụm duy nhất. Như vậy nếu dữ liệu phân bố theo dạng 2-moon (hình 2.2) hoặc trong trường hợp các điểm dữ liệu có nhiều hơn một tâm cụm đại diện thì thuật toán phân cụm K-means chưa xử lý tốt.

Từ các ý tưởng trên luận án đề xuất phương pháp tìm điểm neo bằng thuật toán phân cụm dựa trên cải tiến thuật toán FCM (được gọi lại *lvdc-FCM*).



Kết quả phân cụm theo K-means



Kết quả mong muốn

**Hình 2.2: Kết quả phân cụm K-means và kết quả mong muốn trên tập dữ liệu phân bố theo 2-moon**

---

### Thuật toán 2.1. *lvdc-FCM* (Xác định các điểm neo bằng FCM cải tiến)

---

**Input:**  $LF = \{LF_i\}_{1 \leq i \leq n}$  dữ liệu đặc trưng mức thấp, hằng số  $p > 1$ ,  $C$  số cụm mong muốn

( $C$  có thể rất lớn từ 10% đến 20% của số các ảnh trong CSDL),  $D = \dim(LF_i)$ ,  $i = \overline{1, n}$ ,  $n_b$  là số  $n_{best}$  các điểm gần một tâm cụm  $A_c, c = \overline{1, C}$ ,  $\mu_\epsilon > 0$  và  $L_{max}$  là số vòng lặp tối đa.

**Output:** Tập các tâm cụm (điểm neo)  $\{A_c\}_{1 \leq c \leq C}$ .

**Bước 1:**

1.1 Khởi tạo tập tâm cụm (\*)

1.2: Tính  $J_0(A, \mu)$  theo công thức (2.6) (Sử dụng *FCM* gốc giải hàm mục tiêu này)

**Bước 2:** Lặp  $l = \overline{1, L}$ :

2.1: Tính ma trận độ thuộc  $\mu = [\mu_{ci}]_{C \times n}$  theo công thức (2.10).

2.2: Chuẩn hóa các độ thuộc  $\{\mu_{c,i}\}_{1 \leq c \leq C, 1 \leq i \leq n}$  theo ràng buộc (2.2) và bằng công thức (2.12)

2.3: Tính lại tâm các cụm  $\{A_c\}_{1 \leq c \leq C}$  theo công thức (2.11).

2.3: Tính  $J_l(A, \mu)$  theo công thức (2.6).

2.4. Ra khỏi vòng lặp nếu  $J_l(A, \mu) \approx J_{l-1}(A, \mu)$  (sai số dưới ngưỡng) hoặc  $l > L_{max}$ .

**Bước 3:** Trả về  $\{A_c\}_{1 \leq c \leq C}$ .

Với việc đề xuất các công thức cập nhật ma trận độ thuộc theo (2.10) và hiệu chỉnh tâm cụm theo công thức (2.11) đảm bảo thuật toán cải tiến *lvdc-FCM* có thể thực hiện khởi tạo bằng nhiều cách theo (\*) với số lượng tập tâm cụm có thể rất lớn.

Độ phức tạp của thuật toán 2.1 là:

Bước khởi tạo  $C$  tâm cụm bằng thuật toán K-means là:  $O(n * C * D)$

Xác định  $n \times C$  khoảng cách giữa từng tâm cụm và các điểm dữ liệu:

$O(L * n * C * D)$

Xác định  $n_b$  láng giềng từng tâm cụm:  $O(I)$

Độ phức tạp của phần ước lượng ma trận độ thuộc:  $O(L * C * n_b)$

Độ phức tạp của phần cập nhật tâm cụm:  $O(L * C * n_b * D)$

Do vậy độ phức tạp toàn bộ thuật toán là:  $O(n * C * D) + O(L * n * C * D) + O(L * C * n_b) + O(L * C * n_b * D)$  ở đây  $D$  là số chiều vector,  $L$  là số vòng lặp tối đa,  $C$  là số tâm cụm,  $n_b$  là số láng giềng gần nhất của một tâm cụm  $A_c$ .

### 2.2.2. Tìm các điểm neo gần nhất của một điểm dữ liệu

Bước thứ hai của việc xây dựng đồ thị neo là tìm các điểm neo gần nhất cho mỗi điểm dữ liệu đặc trưng mức thấp  $LF_i$ . Điều này có thể dễ dàng được thực hiện bằng cách sử dụng các độ đo khoảng cách  $d(LF_i, A_c)$  đã được chuẩn hóa như Euclide, Mahalanobis giữa mỗi điểm dữ liệu và tâm cụm.

---

#### Thuật toán 2.2. $S\_Neighbors(LF, U)$

(Xác định  $s$  điểm neo gần nhất của mỗi điểm dữ liệu)

---

**Input:** Các điểm  $LF = \{LF_i\}_{1 \leq i \leq n} \in \mathbb{R}^D$ , các điểm neo  $U = \{A_c\}_{1 \leq c \leq C}$ .

**Output:**  $s$  điểm neo gần nhất cho mỗi điểm dữ liệu  $LF_i$ .

```

1:   for  $i = 1 : n$  do
2:       for  $j = 1 : C$  do
3:            $dist[j] = d(LF_i, A_c)$ 
4:            $anchor\_neighbor[i] = argpartition(dist, s)$ 
5:   return  $s = anchor\_neighbor$ 

```

ở đây:  $d(LF_i, A_c)$  là khoảng cách  $L_2$  từ điểm  $LF_i$  đến điểm neo  $A_c$  và lưu vào mảng  $dist[j]$ ;

$argpartition$ : Hàm này nhận vào một mảng và một số nguyên  $s$ , trả về mảng gồm  $s$  chỉ số của các phần tử nhỏ nhất trong mảng đầu vào.

Kết quả trả về là  $s$  điểm neo gần nhất của mỗi điểm dữ liệu  $LF_i$  nhất, ta ký hiệu là  $A_{[s]}$ .

### 2.2.3. Tính trọng số hồi quy giữa điểm neo và điểm dữ liệu

Xây dựng ma trận trọng số  $Z \in \mathbb{R}^{C \times n}$  biểu diễn mối quan hệ giữa các vector đặc trưng mức thấp trong  $LF = \{LF_i\}_{1 \leq i \leq n} \in \mathbb{R}^{n \times D}$  và các điểm neo trong  $A = \{A_c\}_{1 \leq c \leq C}$ . Hàm giá trị  $r(LF) | r: LF \rightarrow \mathbb{R}$  được ước lượng cho mỗi điểm dữ liệu với các điểm neo như là một trung bình trọng số của các nhãn trên điểm neo.

$$r(LF_i) = \sum_{k=1}^C z_{ki} r(A_k), \quad i = \overline{1, n} \quad (2.13)$$

$$\text{Với điều kiện} \quad \begin{cases} \sum_{k=1}^C z_{ki} = 1 \\ z_{ki} \geq 0 \end{cases} \quad (2.14)$$

Do vậy khi đã tìm thấy  $s$  điểm neo lân cận của từng điểm dữ liệu, các trọng số hồi quy giữa điểm neo và các điểm dữ liệu được xây dựng bằng cách giải quyết bài toán tối ưu hóa sau:

$$\min_{z_i} \varepsilon(z_i) = \frac{1}{2} \left\| x_i - \sum_{s=1}^{|NN(LF_i)|} A_{s \in NN(LF_i)} z_{is} \right\|^2 \quad (2.15)$$

$$\text{Với ràng buộc } \sum_s z_{is} = 1, z_i \geq 0 \quad (2.16)$$

Trong đó  $NN(LF_i)$  là chỉ số của các điểm neo gần nhất của  $LF_i$ .

Để tính vector trọng số hồi quy  $z_{ki}$  của mỗi điểm dữ liệu với các điểm neo gần nhất, luận án sử dụng phương pháp hồi quy hàm nhân *Nadaraya-Watson* để gán các trọng số theo công thức (2.17):

$$z_{ki} = \begin{cases} \frac{K_{d_s} \left( \frac{d \|LF_i, A_k\|}{d_s} \right)}{\sum_{k \in N_b(i,s)} K_{d_s} \left( \frac{d \|LF_i, A_k\|}{d_s} \right)} & \forall k \in N_b(i,s); \\ 0 & \forall k \notin N_b(i,s) \end{cases} \quad (2.17)$$

Để gán trọng số cho  $z_{ki}$  ta sử dụng hàm nhân *Epanechnikov* là phương trình bậc 2:

$$K(t) = \begin{cases} \frac{3}{4}(1-t^2) & \text{if } |t| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.18)$$

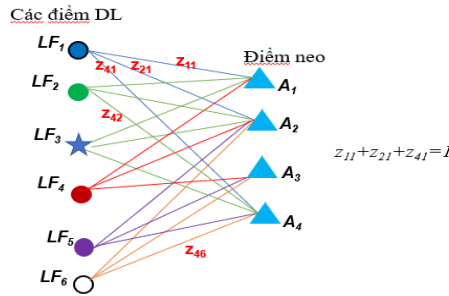
Và các thông số như sau:

$$\begin{aligned} N_b(i,s) &= nbest(\|LF_i, A_l\|; s), \\ d_s &= \max_{l \in N_b(i,s)} \{d(LF_i, A_l)\} \end{aligned} \quad (2.19)$$

Với  $nbest(i)$  là vector thứ  $i$  trong  $nbest$  vector gần  $A_c$  nhất và  $d(LF_i, A_l)$ - khoảng cách  $L_2$  giữa hai vector có cùng kích thước.

Với  $k$  là chỉ số của điểm neo cụ thể khi đang xét đến mối quan hệ giữa một điểm dữ liệu  $x_i$  và điểm neo đó  $A_k$  trong quá trình tính toán trọng số  $z_{ki}$ .  $l$  là chỉ số dùng để lặp qua tất cả các điểm neo trong quá trình tính tổng mẫu số của phân số khi tính  $z_{ki}$ . Mẫu số này là tổng trọng số của tất cả các điểm neo đối với điểm dữ liệu cụ thể  $x_i$ , được dùng để chuẩn hóa trọng số  $z_{ki}$  sao cho tổng trọng số đối với mỗi  $x_i$  bằng 1.

Hình 2.3 minh họa cách các điểm dữ liệu được kết nối với các điểm neo, trong đó các cạnh là trọng số hồi quy.



**Hình 2.3. Biểu diễn đồ thị Neo của điểm dữ liệu  $LF_i$  và điểm neo  $A_k$  với  $z_{ki}$  là trọng số hồi quy biểu diễn mối quan hệ giữa dữ liệu và điểm neo ( $s=3$ )**

#### 2.2.4. Xây dựng ma trận hồi quy

Trong bước này ta xây dựng ma trận hồi quy  $Z$  như sau:

$$Z_{i \langle k \rangle} = z_{ki} \quad \forall i \in \{1, \dots, n\}, 1 \leq k \leq s; \quad (2.20)$$

Với  $\langle k \rangle$  là các chỉ số của  $s$  điểm neo gần điểm dữ liệu  $i$  nhất (được tìm ở thuật toán 2.2) và ma trận  $Z$  có chiều  $(C \times n)$ .

Việc xây dựng ma trận hồi quy được thực hiện như trong Thuật toán 2.3.

**Thuật toán 2.3.** *Regression\_Weights(LF, U, s)*

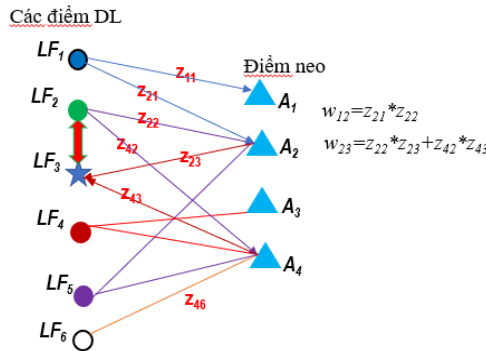
*Xây dựng Ma trận hồi quy*

- 1:  $\langle k \rangle_{i \in \{1, \dots, n\}} = S\_Neighbors(LF, U)$
- 2: for  $i = 1 : n$  do
- 3: Tính  $z_{ki}$  theo các công thức (2.17), (2.18), (2.19) và theo ràng buộc (2.16)
- 4:  $Z_{\langle k \rangle, i} = z_{ki}$
- 5: return  $Z$

Mỗi phần tử  $z_{ki}$  của ma trận  $Z$  biểu diễn trọng số mà điểm dữ liệu  $LF_i$  gán cho điểm neo  $A_k$  ( $A_k \in A_{[s]}$  gần  $LF_i$  nhất), và trọng số này được mịn hóa thông qua hàm nhân hồi quy. Điều này cho phép  $Z$  nắm bắt được mối quan hệ phi tuyến giữa các điểm dữ liệu và các điểm neo.

Như vậy ma trận trọng số hồi quy  $Z$  là ma trận “đo mức độ” quan hệ giữa các điểm dữ liệu và điểm neo; *Ma trận trọng số hồi quy  $Z$  là ma trận thưa, các phần tử khác 0 tương ứng trong ma trận biểu thị mối quan hệ giữa các điểm dữ liệu và các điểm neo.*

### 2.2.5. Xây dựng ma trận kề



Hình 2.4. Ma trận kề  $W$  biểu diễn mối quan hệ giữa các điểm dữ liệu

### 2.3. Kỹ thuật xếp hạng trong EMR

$$EMR(r; Q) = \frac{1}{2} \left( \sum_{1 \leq i, j \leq n+1} w_{ij} \left\| \frac{r_i}{\sqrt{D_{ii}}} - \frac{r_j}{\sqrt{D_{jj}}} \right\|^2 + \mu \sum_{i=1}^{n+1} \|r_i - r_{0,i}\|^2 \right) \rightarrow \min \quad (2.23)$$

Để giải hàm mục tiêu trên, trong xếp hạng đa tạp thường lập giải phương trình:

$$r(t+1) = \alpha S r(t) + (1-\alpha)y \quad (2.24)$$

Hoặc giải bằng phương trình đóng  $r^* = (I - \alpha S)^{-1} y$  (2.25)

$r$  là vector điểm số xếp hạng.

Trong các pha tra cứu, thông thường các điểm truy vấn có thể có hoặc không có trong CSDL. Và để thuận tiện trong tính toán, trong các trường hợp ta gán  $I_{n+1} = Q$ ,  $r_{0,n+1} = 1$ ,  $r_{0,i} = 0$ ,  $i = 1, n$

a) Với trường hợp điểm truy vấn có trong CSDL:

Việc tính toán xếp hạng được thực hiện theo phương trình (2.25): Khi đó  $S$  biến đổi theo  $Z$  như sau:  $S = D^{-1/2} W D^{-1/2} = D^{-1/2} Z^T Z D^{-1/2}$  (thay  $W$  bằng công thức (2.21)). Và từ công thức (2.26) tính  $D$  ta có thể thực hiện việc biến đổi như sau:

$$D_{ii} = \sum_{j=1}^{n+1} z_i^T z_j = z_i^T v \quad (2.28)$$

b) Khi điểm truy vấn nằm ngoài CSDL: Thông thường trong các hệ thống tra cứu, điểm tra cứu thường không có trong CSDL. Do vậy với một truy vấn mới thì ta thực hiện tính  $s$  điểm neo gần kề với điểm truy vấn (theo thuật toán 2.2), sau đó gán trọng số giữa điểm truy vấn và  $s$  điểm neo gần kề của nó và hình thành cột mới  $Z_{C \times (n+1)}$  gọi là  $z_t$

Khi đó ta có:  $D_t = z_t^T v$ ,  $h_t = z_t D_t^{-1/2}$  với  $h_t$  là cột mới của ma trận  $H$ . Để tính nhanh công thức xếp hạng dạng đóng (2.27), ta chỉ quan tâm đến việc tính toán nghịch đảo:

$$B = \left( HH^T - \frac{1}{\alpha} I_C \right)^{-1} = \left( \sum_{i=1}^n h_i h_i^T - \frac{1}{\alpha} I_C \right)^{-1} \quad (2.30)$$

Và  $B'$  là của cột  $h_t$  mới thêm vào được tính như sau:

$$B' = \left( \sum_{i=1}^n h_i h_i^T + h_t h_t^T - \frac{1}{\alpha} I_C \right)^{-1} \approx B \quad (2.31)$$

Khi  $n$  rất lớn và  $h_t$  rất thưa; ta có thể xem xét ma trận  $B$  như nghịch đảo của một ma trận hiệp phương sai. Phương trình (2.31) chỉ ra việc thêm một điểm đơn lẻ sẽ không ảnh hưởng đến ma trận hiệp phương sai của một cơ sở dữ liệu lớn. Nói cách khác, việc tính toán của  $B$  có thể được thực hiện ở giai đoạn ngoại tuyến.

Khi đó vector khởi tạo  $y_t$  ban đầu như sau:

$$y_t = \begin{bmatrix} 0_n \\ 1 \end{bmatrix} \quad (2.32)$$

Với  $0_n$  là vector  $0$  có độ dài  $n$ .

Và công thức xếp hạng dạng đóng (2.27) được viết lại như sau:

$$r^{(n+1) \times 1} = \left( I_{n+1} - \begin{bmatrix} H^T B \\ h_t^T B \end{bmatrix} \begin{bmatrix} H & h_t \end{bmatrix} \right) \begin{bmatrix} 0_n \\ 1 \end{bmatrix} \quad (2.33)$$

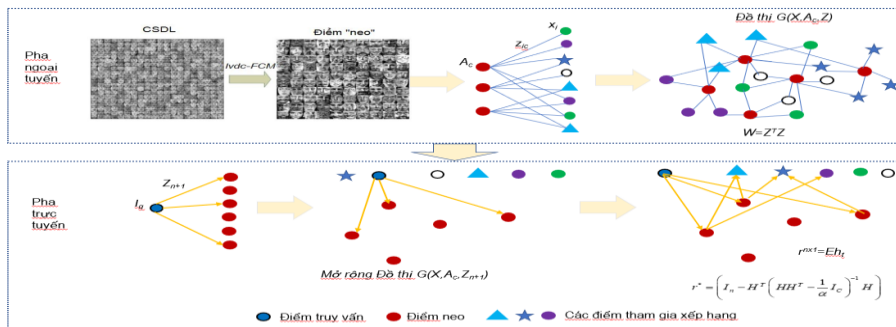
Trong công thức (2.33) ta tập trung vào việc xếp hạng  $n$  phần tử đầu của  $r$  và được tính:

$$r^{n \times 1} = -H^T B h_t = E h_t \quad (2.34)$$

Với ma trận  $E = -H^T B$  được tính ở pha ngoại tuyến. Do vậy việc tính toán xếp hạng theo công thức (2.34) sẽ thực hiện được theo thời gian thực, đảm bảo cho việc tính toán khi điểm truy vấn không có trong CSDL.

#### 2.4. Thuật toán EMR với *lvdc-FCM* trong CBIR

Để khắc phục hạn chế của EMR gốc luận án đề xuất phương pháp mới để xác định các điểm neo và thực hiện áp dụng trong CBIR đề xuất như hình 2.6 mô tả quy tìm điểm neo theo *lvdc-FCM* và xếp hạng.



**Hình 2.6. Xây dựng đồ thị EMR với *lvdc-FCM***

Thuật toán xếp hạng đa tập EMR gốc được thay đổi lại khi sử dụng tập các điểm neo xác định bằng thuật toán *lvdc-FCM* như sau:

**Thuật toán 2.4. EMR-(lvdc-FCM)** (*Thuật toán xếp hạng EMR sử dụng lvdc-FCM để xác định các Anchor point*).

**Input:**  $\{LF_i\}_{1 \leq i \leq n}$  dữ liệu đặc trưng mức thấp.

$LF_Q$ : đặc trưng mức thấp của ảnh truy vấn.

$C$ : số lượng điểm neo của thuật toán EMR, tham số  $a \in (0,1)$  ( $a \approx 1$ ),

$D = \dim(LF_i), i = \overline{1, n}$

**Output:**  $r = \{r_i\}_{1 \leq i \leq n}, r_i \in [0,1], \forall i = \overline{1, n}$  là thứ hạng tương tự với ảnh  $Q$  của ảnh  $I_i$  trong CSDL ảnh  $E$ .

**Pha offline:**

**Bước 1:** Chuẩn hóa  $\{LF_i\}_{1 \leq i \leq n}$  theo thuật toán  $3\sigma\text{-opt}$  [93]

**Bước 2:** Tìm tập điểm neo bằng thuật toán *lvdc-FCM* (*thuật toán 2.1*)

**Bước 3:** Tìm  $s$  điểm neo gần một điểm dữ liệu theo thuật toán (2.2)

**Bước 4:** Tính ma trận hồi quy  $Z$  theo thuật toán (2.3) (\*)

**Bước 5:** Tính  $W$  với  $W = Z^T Z$

**Bước 6:** Chuẩn hóa  $S = H^T H$

(với  $H = ZD^{-\frac{1}{2}}$ ) và tính  $D$  theo công thức (2.28)

**Pha online:**

**Bước 7:** Mở rộng Tính  $Z_Q = Z_{n+1}$  theo thuật toán (2.3) (\*\*)

**Bước 8:** Đặt  $r_Q = \{r_i\}_{1 \leq i \leq n+1}, r_i = 0 \forall i = \overline{1, n}, r_{n+1} = 1.0$

Tính toán xếp hạng theo công thức (2.26) (\*\*\*)

**Bước 9:** trả về  $r = \{r_{Q,i}^*\}_{1 \leq i \leq n}$ .

## 2.5. Thử nghiệm và đánh giá kết quả

### 2.5.1. Đánh giá hiệu quả của của thuật toán *lvdc-FCM* đề xuất

Để đánh giá tính hiệu quả của *lvdc-FCM* đề xuất luận án dùng chỉ số đánh giá dành cho các thuật toán phân cụm như: Silhouette Score, Davies-Bouldin Score, Calinski-Harabasz Score, Xie-Beni index.

Chỉ số càng thấp thì phân cụm càng tốt. Chỉ số Xie-Beni index được tính bởi công thức sau:

$$XB_{index} = \frac{\sum_{j=1}^n \sum_{i=1}^C u_{ij}^2 \|x_j - A_i\|^2}{n \cdot \min_{i \neq k} \|A_i - A_k\|^2}$$

Thực hiện đánh giá trên tập dữ liệu 20D với 3000 điểm, số cụm 30, nbets = 100; số vòng lặp lớn nhất là 150;

**Bảng 2.1. Kết quả các chỉ số đánh giá trên tập dữ liệu nhân tạo 20D**

Thuật toán	Silhouette Score	Davies-Bouldin Score	Calinski-Harabasz Score	Xie-Beni index
K-means	0.45217	<b>1.012239001369301</b>	19915	
FCM	0.44329	1.0178483330467862	19914.74451	0.00212
<i>lvdc-FCM</i>	<b>0.48825</b>	1.0278284360498895	<b>19932.72572</b>	<b>0.00149</b>

Chỉ số Xie-Beni index đo có sự tham gia của 3 thành phần gồm tập dữ liệu, tập tâm cụm và ma trận độ thuộc (không có ở K-means) nên chỉ số sánh với K-means ở chỉ số này. Với bảng kết quả này các chỉ số của *lvdc-FCM*, FCM đều tốt hơn so với K-means ở mức độ phân biệt các tâm cụm.

Chỉ số IFVIndex (Index of Fuzzy Validity) là một phép đo được sử dụng để đánh giá chất lượng của phân cụm mờ (fuzzy clustering) trong khả năng phân chia các nhóm và tính mờ của phân cụm. IFVIndex cung cấp một số liệu đánh giá về sự tách biệt và sự tập trung của các nhóm phân cụm, dựa trên các giá trị tập hợp đồng nhất (fuzziness) và giá trị tập trung của các điểm dữ liệu trong mỗi nhóm. Giá trị độ đo IFVIndex được tính bởi phương trình sau:

$$IFV_{index} = \frac{1}{C} \sum_{j=1}^C \left\{ \frac{1}{N} \sum_{k=1}^N u_{kj}^2 \left[ \log_2 C - \frac{1}{N} \sum_{k=1}^N \log_2 u_{kj} \right]^2 \right\} \times \frac{SD_{\max}}{\sigma_D}$$

Trong đó:  $SD_{\max} = \max_{k \neq j} \|A_k - A_j\|^2$  và  $\sigma_D = \frac{1}{C} \sum_{j=1}^C \left( \frac{1}{N} \sum_{k=1}^N \|X_k - A_j\|^2 \right)$

Giá trị IFVIndex cài đặt trong thực nghiệm tỉ lệ thuận với với hiệu năng của các thuật toán.

**Bảng 2.2. Chỉ số Xie-Beni index và chỉ số IFVIndex trên tập dữ liệu 20D với 3000 điểm, số cụm 30, số vòng lặp 150.**

Thuật toán	Xie-Beni index	IFVIndex
FCM	0.00212	11.704905011975184
<i>lvdc-FCM</i>	<b>0.00149</b>	<b>12.583643380361265</b>

Với tập dữ liệu đặc trưng mức thấp được trích chọn từ CSDL ảnh (tập Corel30K, tập logo-2K+, tập VGGFACE2-S), luận án thực hiện đánh giá chỉ số Xie-Beni index để so sánh FCM gốc và *lvdc-FCM*. Bảng kết quả như sau:

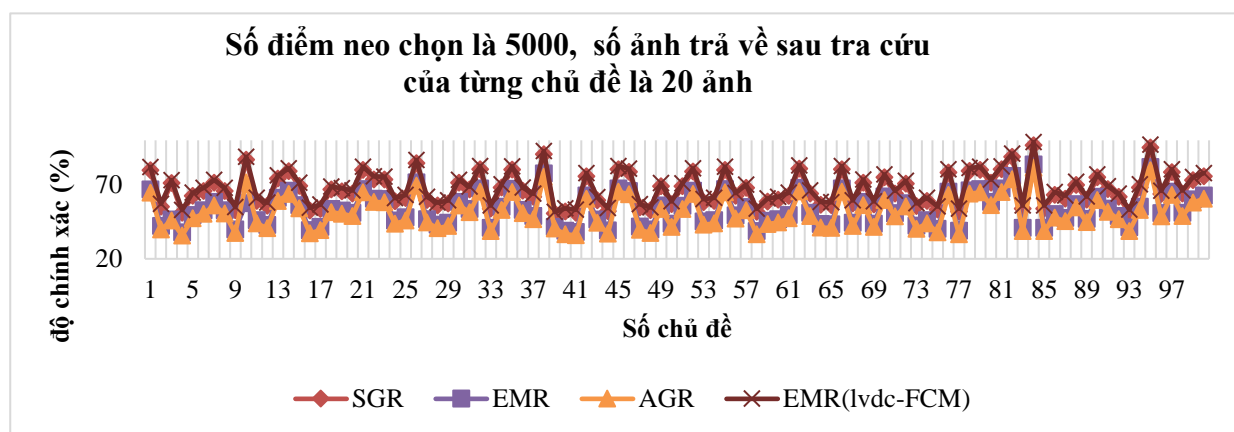
**Bảng 2.3. Chỉ số Xie-Beni index và chỉ số IFVIndex trên tập dữ liệu LF809 của tập dữ liệu Corel30K số cụm 5000 và số vòng lặp 100**

Thuật toán	Xie-Beni index	IFVIndex
FCM	0.00412	118.2531745545215
<i>lvdc-FCM</i>	<b>0.00249</b>	<b>119.1000254356127</b>

Từ kết quả thực nghiệm ở các bảng 2.2 và bảng 2.3 cho thấy các chỉ số *Xie-Beni index* và *IFVIndex* trên các tập dữ liệu khác nhau (tập 20D- 3000 điểm và tập dữ liệu đặc trưng mức thấp 809D của Corel30K) của thuật toán đề xuất *lvdc-FCM* cho ta tập tâm cụm phân biệt rõ hơn thuật toán FCM gốc.

### 2.5.2. Đánh giá hiệu năng tra cứu ảnh với EMR-(*lvdc-FCM*)

Đối với thực nghiệm này, luận án chọn ngẫu nhiên 100 chủ đề (trong tổng số 303 chủ đề của tập dữ liệu Corel30K, mỗi chủ đề có 100 ảnh) và thực hiện chọn ảnh truy vấn trong tất cả các chủ đề. Kết quả trả về sau tra cứu là 20 ảnh. Từ kết quả hình 2.12 ta thấy phương pháp AGR cho độ chính xác thấp nhất. Cách lựa chọn điểm neo trong AGR là chọn ngẫu nhiên tập điểm đại diện trong CSDL làm điểm neo. Đối với 2 phương pháp SGR và EMR, phương pháp lựa chọn điểm neo là dùng thuật toán phân cụm K-means. Cũng trong hình này, độ chính xác trung bình của phương pháp đề xuất là cao nhất so với các phương pháp còn lại.



**Hình 2.12. Độ chính xác của 4 phương pháp ở 20 ảnh trả về sau tra cứu trên tập dữ liệu COREL30K**

Ở đây, dựa vào biểu đồ có thể dễ nhận thấy rằng một số chủ đề dễ, tất cả các thuật toán thực hiện tốt, và một số chủ đề khó các thuật toán thực hiện cho kết quả tra cứu có độ chính xác không cao.

**Bảng 2.4. Bảng kết quả tra cứu ảnh trên các tập dữ liệu với các tập điểm neo lớn và độ chính xác trung bình cho từng bộ điểm neo**

Số điểm neo C	Logo-2k+		Corel30K		VGGFACE2-S	
	EMR gốc	EMR-(lvdc-FCM)	EMR gốc	EMR-(lvdc-FCM)	EMR gốc	EMR-(lvdc-FCM)
7000	57.5%	65.7%	63.2 %	68.4 %	<b>67.5%</b>	73.3%
8000	56.3%	<b>68.8%</b>	64.5 %	69.5 %	66.2%	74.5%
9000	<b>56.7%</b>	65.5%	65.6 %	69.9 %	66.6%	74.6%
10000	55.1%	64.6%	<b>66.3 %</b>	<b>70.8 %</b>	66.5%	<b>74.7%</b>

Từ số liệu các thực nghiệm, việc lựa chọn phương pháp tìm điểm neo và số lượng điểm neo ảnh hưởng trực tiếp đến kết quả tra cứu. Thông thường số lượng điểm neo được sử dụng bằng khoảng 5% đến 15% số lượng mẫu của CSDL.

**Bảng 2.5. Thời gian thực thi khi tra cứu một truy vấn (điểm neo 5000, ảnh trả về 20) trên tập Corel30K**

Phương pháp	Thời gian (s)
AGR	0.293
EMR	0.309
SGR	0.235
EMR-(lvdc-FCM)	0.317

## 2.6. Kết luận chương 2

Trong chương này luận án đã trình bày phương pháp tìm điểm neo mới dựa trên thuật toán phân cụm FCM cải tiến (gọi là thuật toán *lvdc-FCM*) trong điều kiện: số vector dữ liệu đặc trưng ảnh và số cụm lớn.

Với thuật toán đề xuất, luận án đã tiến hành thực nghiệm đánh giá hiệu quả của thuật toán *lvdc-FCM* so với các thuật toán FCM gốc và K-means trên tập dữ liệu nhân tạo được hình thành từ các tập *swiss\_roll*, tập *moons*, tập *s\_curve*, tập *friedman3* và tập *gaussian\_quantiles*. Với các chỉ số đánh giá thông dụng như Silhouette Score, Davies-Bouldin Score, Calinski-Harabasz Score, Xie-Beni, thuật toán đề xuất đều cho kết quả đánh giá tốt hơn các thuật toán FCM gốc và K-means.



Trên cơ sở thuật toán phân cụm *lvdc-FCM*, luận án trình bày thuật toán cải tiến của EMR, gọi là EMR-(*lvdc-FCM*) với phương pháp tìm các điểm neo bằng thuật toán đề xuất *lvdc-FCM* thay thế cho thuật toán K-means và áp dụng trong CBIR để nâng cao độ chính xác trong tra cứu [CT1, CT2, CT4]. Vận dụng thuật toán *lvdc-FCM*, thuật toán EMR-(*lvdc-FCM*) đã xây dựng đồ thị xếp hạng đa tạp với số lượng điểm neo rất lớn, khai thác ưu điểm của phép phân cụm *lvdc-FCM* so với K-means trong tra cứu ảnh. Thuật toán đề xuất được thực nghiệm tra cứu trên ba tập dữ liệu ảnh: Logo-2k+, Corel30K, VGGFACE2-S cho độ chính xác được cải thiện rõ rệt như đã chỉ ra ở các thực nghiệm đánh giá bằng trực quan và bằng chỉ số đánh giá khách quan. Kết quả thực nghiệm được so sánh với các thuật toán EMR, SGR, AGR cho độ chính xác của phương pháp đề xuất đã được cải thiện và đáng tin cậy. Sự cải thiện về độ chính xác tăng đáng kể (khoảng 7% - 15%) so với thuật toán gốc và các thuật toán khác cùng họ.

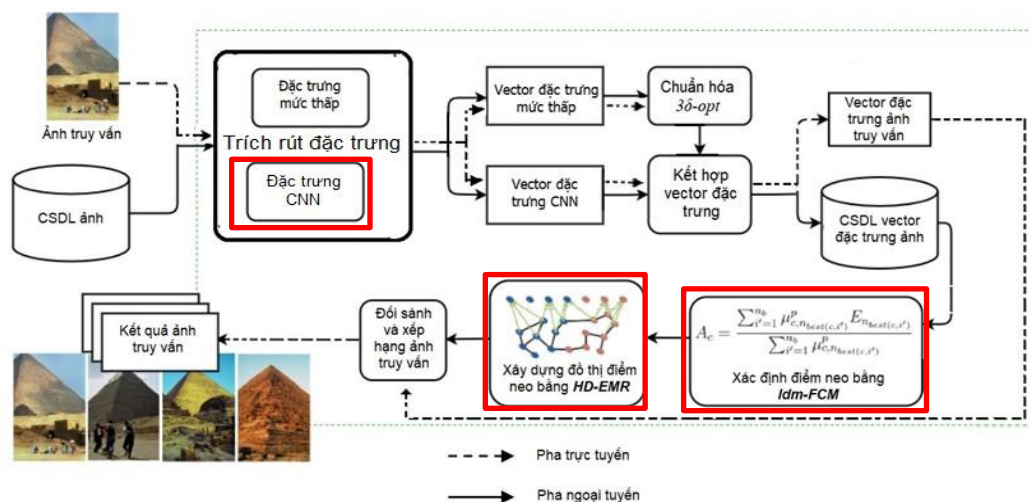
### Chương 3

## NÂNG CAO HIỆU QUẢ TRA CỨU ẢNH THEO TIẾP CẬN KẾT HỢP ĐẶC TRƯNG MỨC THẤP VÀ ĐẶC TRƯNG MỨC CAO

### 3.1. Giới thiệu

### 3.2. Hệ thống CBIR đề xuất

Mô hình tổng quát cho các đề xuất của luận án trong **hình 3.1** được mô tả như sau:



**Hình 3.1. Hệ thống đề xuất**

### 3.3. Biểu diễn ảnh bằng kết hợp đặc trưng mức cao từ EfficientNetB7+ và đặc trưng mức thấp

#### 3.3.1. Trích rút đặc trưng mức cao của ảnh sử dụng CNN

Đầu ra của mạng CNN có thể là một hoặc nhiều vector đặc trưng được tạo ra bằng cách cho ảnh đầu vào chạy xuyên qua mạng với các trọng số đã được huấn luyện. Tùy thuộc vào từng bài toán khác nhau, chúng ta có thể sử dụng các vector đặc trưng đó theo nhiều cách khác nhau. Chẳng hạn, có thể sử dụng một bộ phân lớp cơ bản để phân lớp các vector đặc trưng, hoặc có thể gắn một tầng phân lớp (chẳng hạn dùng hàm SoftMax) vào cuối của mô hình mạng để dự đoán nhãn của các vector đặc trưng.

### 3.3.2. Mạng EfficientNet

Mạng EfficientNet là kiến trúc mạng nơ-ron học sâu được đề xuất tại nghiên cứu. Mạng EfficientNet được thiết kế nhằm tối ưu hoá đồng thời hiệu suất và tài nguyên tính toán với mục tiêu đạt được kết quả tốt nhất với kích thước mạng tối ưu nhất.

EfficientNet sử dụng một số kỹ thuật quan trọng để tăng hiệu quả của mạng sử dụng phương pháp tinh chỉnh tỉ lệ để tự động điều chỉnh độ rộng, độ sâu và độ phân giải thông qua một tham số trung gian nhằm giúp mô hình linh hoạt và mềm dẻo với các kích thước đầu vào khác nhau mà không làm gia tăng quá độ phức tạp của mô hình.

### 3.3.3. Kỹ thuật đề xuất tinh chỉnh mạng efficientNetB7

- a) Kỹ thuật transfer learning
- b) Sử dụng mô hình cơ sở EfficientNetB7
- c) Cải tiến mô hình cơ sở EfficientNetB7

**Thuật toán 3.1:** Trích rút đặc trưng ảnh với mạng CNN tiền huấn luyện

**Input:**

- Tập ảnh cần trích rút đặc trưng:  $I=(I_1, I_2, \dots, I_n)$
- Mô hình tiền huấn luyện  $\Omega = (Model\ EfficientNetB7)$

**Output:** Vector đặc trưng biểu diễn ảnh  $HF=(HF_1, HF_2, \dots, HF_n)$

1. Model ← LoadModel( $\Omega$ )
2.  $HF \leftarrow \phi$
3. for  $i=1, \dots, n$  do
  - 3.1  $HF_i \leftarrow \text{ExtractFeature}(I_i, Model);$
  - 3.2  $HF = HF \cup HF_i$
4. Return  $HF$

### 3.4. Kết hợp đặc trưng

**Thuật toán 3.2: Thuật toán trích rút và kết hợp các đặc trưng**

**Thuật toán 3.1:** Kết hợp đặc trưng:  $LF\text{-}DFE(E)$  { $LF$ -đặc trưng mức thấp,  $DFE$ -đặc trưng nhúng học sâu}

**Input:** Tập dữ liệu ảnh  $E = \{E_i\}_{1 \leq i \leq n}$

$\Omega$ : Mô hình tiền huấn luyện ( $\Omega \in \{EfficientNet, EfficientNetB7+ \dots\}$ )

**Output:** Tập vector đặc trưng kết hợp  $\{IC_i\}_{1 \leq i \leq n}$

**Bước 1:** Trích rút đặc trưng mức thấp mô tả màu, kết cấu và hình dạng của ảnh:

**Bước 2:** Trích rút đặc trưng CNN:

**Bước 3:** Kết nối đặc trưng mức thấp  $LF$  và đặc trưng mức cao thu được tập vector đặc trưng chiều rất cao biểu diễn CSDL ảnh  $E$ :

**Bước 4:** Trả về  $\{IC_i\}_{1 \leq i \leq n}, \Omega_{ft}$ .

### 3.5. CBIR với kỹ thuật xếp hạng EMR dựa vào GPU-platform

Trong phần này, việc sử dụng phân cụm FCM với tập dữ liệu vector lớn, số chiều rất cao và số cụm rất lớn để xác định các điểm neo cho đồ thị của EMR đã được thực hiện.

Phần này tập trung vào việc áp dụng tính toán trên GPU trong giai đoạn ngoại tuyến để xác định các điểm neo cho đồ thị của EMR.

### 3.5.1. Tối ưu hóa thuật toán tìm điểm neo dựa trên GPU-platform

- Trong *LDM-FCM* đề xuất kỹ thuật tối ưu bước hiệu chỉnh ma trận độ thuộc của thuật toán *lvdc-FCM* khi xử lý các vector dữ liệu cần phân cụm có số chiều rất cao bằng tiếp cận của các thuật toán *ANN*.

- Ngoài ra để tối ưu phép hiệu chỉnh ma trận độ thuộc, ma trận độ thuộc được biểu diễn thành ma trận thưa hai chiều để giảm tài nguyên lưu trữ và có thể chuyển lên bộ nhớ GPU-RAM để thực hiện tính toán song song.

Chuyển các ma trận thưa *DIS* và *POS* lên GPU-RAM thu được hai ma trận  $DIS_{gpu}$  và  $POS_{gpu}$ :

$$DIS_{c,i,gpu} \leftarrow DIS_{c,cpu} = \{D_{i_{c,1}}, D_{i_{c,2}}, \dots, D_{i_{c,n_b}}\}_{1 \leq c \leq C} \quad (3.1)$$

$$POS_{c,gpu} \leftarrow POS_{c,cpu} = \{i_{c,1}, i_{c,2}, \dots, i_{c,n_b}\}_{1 \leq c \leq C}$$

Biểu diễn ma trận độ thuộc là ma trận thưa (kích thước  $C \times n_b$ ), được tính toán song song trên GPU theo công thức sau:

$$\mu_{c,i,gpu} = \begin{cases} \frac{1}{\sum_{\substack{1 \leq c' \leq C \\ i \in POS_{c',gpu}}} \left( \frac{DIS_{c,i,gpu}}{DIS_{c',i,gpu}} \right)^{\frac{2}{p-1}}} & \text{for } i \in POS_{c,gpu} \\ 0.0 & \text{for } i \notin POS_{c,gpu} \end{cases} \quad (3.2)$$

$$\sum_{i \in POS_{c,gpu}} \mu_{c,i,gpu} = 1 \quad (3.3)$$

Và chuẩn hóa để đảm bảo

Tính giá trị hàm mục tiêu của FCM trên GPU theo công thức:

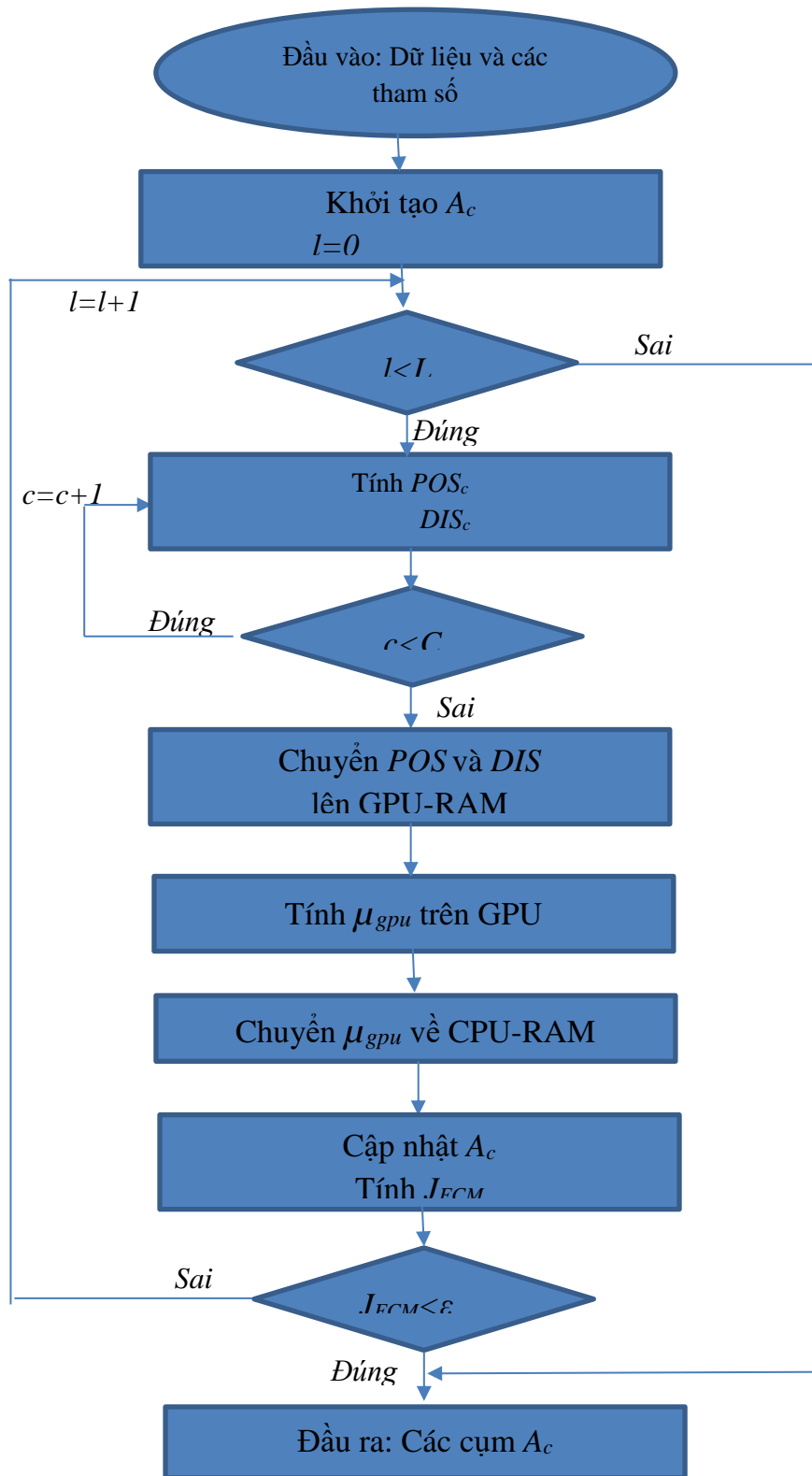
$$J(A, \mu) = \sum_{c=1}^C \sum_{i \in POS_{c,gpu}} \mu_{c,i,gpu}^p DIS_{c,i,gpu}^p \quad (3.4)$$

Kiểm tra điều kiện dừng. Nếu không thỏa thì cập nhật tập tâm cụm  $A_c$ , bước này được thực hiện trên CPU:

Chuyển ma trận độ thuộc  $\mu_{c,i}$  về CPU-RAM:  $\mu_{c,i} \leftarrow \mu_{c,i,gpu}$  và cập nhật tập tâm cụm trên toàn bộ tập dữ liệu với công thức (3.5) được tính toán ở CPU.

$$A_c = \frac{\sum_{i \in POS_c} \mu_{c,i}^p * IC_i}{\sum_{i \in POS_c} \mu_{c,i}^p} \quad (3.5)$$

Sơ đồ thuật toán được thể hiện như hình 3.9:



**Hình 3.9.** Sơ đồ thuật toán LDM-FCM

### 3.5.2. Thuật toán LDM-FCM

#### Thuật toán 3.2: LDM-FCM (Local data manifold Fuzzy C-means)

**Input:** Tập vector đặc trưng kết hợp  $\{IC_i\}_{1 \leq i \leq n}$  của dataset  $n$  ảnh  $E$ .  
Hằng số  $p > 1$  ( $p = 2$ );  $C$ - số cụm mong muốn

$D = \dim(IC)$ <sup>1</sup>;  $n_b$  số lớn nhất các vector  $IC_i$  thuộc về một cụm  $\{A_c\}_{1 \leq c \leq C}$ <sup>2</sup>  
 $\varepsilon > 0$  là hằng số đủ nhỏ;  $L_{max}$  là số vòng lặp tối đa.

$IDX_E(D)$  là bộ chỉ mục của  $\{IC_i\}_{1 \leq i \leq n}$ <sup>3</sup>

Tham số GPU platform:

+ *block size* =  $\{blk_r, blk_c\}$ : tham số block 2D.

+ *thread size* =  $\{thrd_r, thrd_c\}$ : số thread 2D trên mỗi khối.

**Output:** Tập các tâm cụm  $A = \{A_c\}_{1 \leq c \leq C}$

Các bước trong thuật toán được thực hiện theo thứ tự:

**Bước 1:** Khởi tạo tập tâm cụm  $\{A_c\}$  với số cụm mong muốn  $C$ .

**Bước 2:** Lặp theo  $l: l = 1, L_{max}$ ,

**Bước 3:** Copy dữ liệu từ GPU-RAM về CPU-RAM:  $\mu_{c,i} \leftarrow \mu_{c,i, gpu}$ .

**Bước 4:** Trả về  $\{A_c\}_{1 \leq c \leq C}$ .

Thuật toán *LDM-FCM* đề xuất là một cải tiến so với FCM gốc với hiệu suất cao và là kỹ thuật mới được áp dụng cho việc phân cụm dữ liệu:

- Trong thuật toán đã tích hợp với kỹ thuật ANN để tăng tốc độ xác định các vector gần một tâm cụm.

- Tính toán nhanh các ma trận liên quan bằng cách biểu diễn chúng dưới dạng ma trận thưa 2 chiều để giảm tải nguyên lưu trữ và sử dụng công nghệ lập trình đa luồng trên GPU để tăng tốc độ tính toán.

Đầu ra của thuật toán *LDM-FCM* là các tâm của các cụm dữ liệu. Trong phạm vi ứng dụng cho thuật toán xếp hạng EMR trong CBIR, các vector tâm cụm này được gán cho các điểm neo để xây dựng đồ thị HD-EMR, chi tiết được trình bày trong phần tiếp theo.

### 3.5.3. Xây dựng đồ thị HD-EMR với đặc trưng chiều rất cao

Việc xây dựng đồ thị HD-EMR trên toàn bộ CSDL vector đặc trưng chiều cao được thực hiện với các bước phân tích và thực hiện qua các bước như sau:

**Bước 1:** Tìm điểm neo bằng thuật toán *LDM-FCM* (thuật toán 3.2)

**Bước 2:** Tìm  $s$  điểm neo gần nhất với mỗi điểm dữ liệu  $x_i$ :

---

**Thuật toán 3.3.**  $s\_Neighbors(IC_i, A_c)$

(Xác định  $s$  điểm neo gần nhất của mỗi điểm dữ liệu)

---

**Input:** Các điểm  $IC = \{IC_i\}_{1 \leq i \leq n} \in R^D$ , các điểm neo  $A = \{A_c\}_{2 \leq c \leq C}$ .

**Output:**  $s$  điểm neo gần nhất cho mỗi điểm dữ liệu  $IC_i$ .

1: Lập chỉ mục cho  $C$  điểm neo:  $IDX_{A_c} = IndexFlatL2(D, A_c)$

2: Tìm  $s$  điểm neo gần một điểm dữ liệu:

for  $i \leftarrow 1$  to  $n$  do

$ANNSearch(\{A_c, IDX_{A_c}\}, s, IC_i)$ <sup>4</sup>

3: return  $s$

---

<sup>1</sup> Số chiều vector của các đặc trưng  $\{IC_i\}_{1 \leq i \leq n}$ ,  $D$  có thể rất lớn.

<sup>2</sup> Giả định một cụm tâm  $A_c$  có thể chỉ gồm  $n_b$  các vector gần tâm cụm  $A_c$  nhất theo khoảng cách  $L_2$ .

<sup>3</sup> Được xác định bởi thuật toán *IndexFlatL2* của thư viện Faiss. Sử dụng kỹ thuật tìm kiếm ANN đã có và sử dụng khoảng cách  $L_2$

<sup>4</sup> xác định được các chỉ số  $c_{i,1}, \dots, c_{i,s}$  là  $s$  số nguyên khác nhau thuộc  $\{1, 2, \dots, C\}$  sao cho  $s$  điểm neo  $A_c$  gần  $IC_i$  nhất theo độ đo khoảng cách  $L_2$ .

**Bước 3:** Tính ma trận trọng số hồi quy  $Z$ : Sau khi tìm được  $s$  điểm neo gần nhất của một điểm dữ liệu ta tính ma trận  $Z$  là ma trận hồi quy biểu diễn mối quan hệ giữa điểm neo và điểm dữ liệu theo công thức:

**Bước 4:** Tính ma trận kê và chuẩn hóa ma trận chuyên tiếp: Ma trận kê biểu diễn mối quan hệ giữa các điểm dữ liệu  $W$  được tính theo công thức (2.21):  $W=Z^T Z$  và ma trận

chuyên tiếp  $S = D^{-1/2} W D^{1/2}$  với  $D_{ii} = \sum_{j=1}^n w_{ij} = \sum_{j=1}^n z_i^T z_j$  ;

**Bước 5:** Xử lý một truy vấn:

Giả sử CSDL ảnh được biểu diễn bởi CSDL vector đặc trưng  $IC = \{IC_i\}_{1 \leq i \leq n}$  là vector đặc trưng biểu diễn ảnh kết hợp giữa mức thấp và mức cao.

a) Ảnh truy vấn  $Q$  có trong CSDL ảnh  $E$ .

**Nhận xét:**

b) Ảnh truy vấn  $Q$  không có trong CSDL ảnh  $E$

Với  $FE = \{IC_i\}_{1 \leq i \leq n}$  - tập CSDL vector đặc trưng  $d$  chiều của CSDL ảnh  $E$  có  $n$  ảnh. Khi  $Q \notin E$ , ta đặt  $E_{n+1} = Q$  có biểu diễn ảnh là vector đặc trưng  $IC_{n+1}$ , ký hiệu  $r_{0_{n+1}} = 1$ ,  $r_{0_i} = 0$ ,  $\forall i = 1, n$  và các ma trận liên quan như sau:

$$\begin{aligned} z_{n+1} &= (z_{c,n+1})_{1 \leq c \leq C}, & D_{n+1} &= z_{n+1} v \\ h_{n+1} &= z_{n+1} D_{n+1}^{-1/2}, & H_{new} &= H \oplus h_{n+1} \quad \text{với } v = \sum_{j=1}^n z_j \end{aligned}$$

Do vậy chúng ta chỉ tính toán ma trận  $D$  mà không cần tính ma trận kê  $W$

Các xếp hạng của CSDL vector đặc trưng  $\{IC_i\}_{1 \leq i \leq n}$  với vector truy vấn  $IC_Q$  được tính theo công thức:

$$r_Q = (r_i)_{1 \leq i \leq n} = \left\{ (1-\alpha) \alpha H^T B \right\} h_{n+1} \quad (3.8)$$

Với  $B = (I_c - \alpha H H^T)^{-1}$  và  $I_c$  là ma trận đơn vị  $C \times C$ .

Ta chứng minh công thức (3.8):

Ta biến đổi như sau:

$$\begin{aligned} r &= \alpha H^T (I_c - \alpha H H^T)^{-1} h_{n+1} = \alpha H^T \sum_{t=0}^{+\infty} (\alpha H H^T)^t h_{n+1} \Rightarrow r_{t+1,Q} = \alpha H^T \sum_{i=0}^{t+1} (\alpha H H^T)^i h_{n+1} \\ &= \alpha H^T \left\{ I_c + \sum_{i=1}^{t+1} (\alpha H H^T)^i \right\} h_{n+1} = \alpha H^T h_{n+1} + \alpha H^T \left\{ \sum_{i=1}^{t+1} (\alpha H H^T)^i \right\} h_{n+1} \\ &= \alpha H^T h_{n+1} + \alpha H^T \left\{ (\alpha H H^T) \sum_{i=0}^t (\alpha H H^T)^i \right\} h_{n+1} \\ &= \alpha H^T h_{n+1} + \alpha H^T \left\{ (H(\alpha H^T)) \sum_{i=0}^t (\alpha H H^T)^i \right\} h_{n+1} \\ &= \alpha H^T h_{n+1} + \alpha H^T H \left\{ (\alpha H^T) \sum_{i=0}^t (\alpha H H^T)^i \right\} h_{n+1} \\ &= \alpha H^T h_{n+1} + \alpha (H^T H) \left\{ (\alpha H^T) \sum_{i=0}^t (\alpha H H^T)^i \right\} h_{n+1} \\ &= \alpha H^T h_{n+1} + \alpha (H^T H) r_{t,Q} = \alpha S r_{t,Q} + \alpha H^T h_{n+1}. \end{aligned}$$

Vậy  $r_Q = \lim_{t \rightarrow \infty} r_{t,k}$  với  $r_{t+1,Q} = \alpha H^T h_{n+1} + (\alpha H^T H) r_{t,Q} = \alpha S r_{t,Q} + \alpha H^T h_{n+1}$ .

Khi đó, các ma trận tham gia vào tính toán xếp hạng các vector trong CSDL với vector truy vấn của ảnh  $Q$  có các tính chất như sau:

$z_{n+1}$  là vector thưa  $C$  chiều với chỉ có  $s$  thành phần khác 0;

$h_{n+1}$  là vector thưa  $n$  chiều và chỉ có  $C$  thành phần khác 0;

$h_{n+1} h_{n+1}^T$  là ma trận cỡ  $C \times C$  và thưa;

$Z$  là ma trận cỡ  $C \times n$ , và có  $s \times n$  thành phần khác 0. Các phần tử khác 0 tương ứng trong ma trận biểu thị mối quan hệ giữa các điểm dữ liệu và các điểm neo;

$H$  là ma trận thưa cỡ  $C \times n$  và có  $s \times n$  thành phần khác 0;

$HH^T$  là ma trận cỡ  $C \times C$ , là ma trận thưa do các hệ số này bằng 0 trên tập cặp chỉ số điểm neo mà các điểm neo này không cùng thuộc một  $s$ -best của bất kỳ vector  $IC_i$  nào. Thật vậy, ta có:

$$\forall (c_1, c_2) \in \{1 \leq c_1, c_2 \leq C \mid \neg \exists i: 1 \leq i \leq n \wedge A_{c_1} \in N_b(i; s) \wedge A_{c_2} \in N_b(i; s)\},$$

$$HH^T[c_1, c_2] = \sum_{i=1}^n H[c_1, i] H[c_2, i] = 0$$

Với ma trận  $(1-\alpha)\alpha H^T B = (1-\alpha)\alpha H^T (I_C - \alpha HH^T)^{-1}$  có thể tính sẵn ở bước offline.

### 3.5.4. Thuật toán HD-EMR trong CBIR

Từ phân tích và đánh giá trên, luận án đề xuất thuật toán HD-EMR như sau:

**Thuật toán 3.3:** HD-EMR (High Dimesional-EMR)

**Input:** Tập CSDL vector đặc trưng ảnh  $IC = \{IC_i\}_{1 \leq i \leq n}$ , đặc trưng ảnh truy vấn  $IC_Q$

$\Omega$ : Mô hình học sâu đào tạo trước ( $\Omega \in \{Resnet; VGG, EfficientNet, \dots\}$ ).

$C$ : số lượng điểm neo (số cụm thu được ở thuật toán LDM-FCM).

$s$ : số phần tử điểm neo lân cận của một vector đặc trưng

$\alpha \in [0, 1)$

**Output:**  $r = \{r_i\}_{1 \leq i \leq n}$ ,  $r_i \in [0, 1] \forall i = \overline{1, n}$  là thứ hạng tương tự với ảnh  $Q$  lấy ra trong cơ sở dữ liệu ảnh  $E$

Các bước của thuật toán bao gồm:

**Bước 1:** (offline) (có thể dùng GPU-platform)

**Bước 2:** (online):

**Bước 3:** Trả về vector xếp hạng  $\{r_Q\}$ .

## 3.6. Thực nghiệm và các kết quả

### 3.6.1. Môi trường thực nghiệm và huấn luyện mạng EfficientNetB7+

Việc sử dụng mạng EfficientNetB7+ [CT6] cho phép tận dụng được các thông tin đa dạng và phong phú từ các tập dữ liệu Logo-2K+, Corel30K và VGGFACE2S. Điều này giúp cải thiện khả năng trích rút đặc trưng từ các hình ảnh và tăng tính đa dạng của bộ vector đặc trưng biểu diễn ảnh.

**Bảng 3.5. Bảng các đặc trưng sử dụng trong thực nghiệm**

Các mô tả	Kiểu đặc trưng	Số chiều vector đặc trưng
GCM	Color	81
LBP	Texture	59
GWT	Texture	120
EDH	Shape	37
GIST	Shape	512

Trích chọn từ mạng pretrain Resnet50	CNN	2048
Trích chọn từ mạng pretrain VGG19	CNN	4096
Trích chọn từ mạng pretrain EfficientNetB7	CNN	2560
Trích chọn từ mạng pretrain EfficientNetB7+	CNN	2560
Kết hợp đặc trưng (B7+ và LF)	CNN+LF	3369

### 3.6.2. Các tham số và kết quả thực nghiệm hệ thống đề xuất HD-EMR

Hình thông CBIR được đề xuất như ở hình 3.1 được thực hiện qua các thực nghiệm sau:

Các thông số thực nghiệm như sau:

a. Sử dụng các đặc trưng mức cao (đặc trưng CNN) được trích rút từ mô hình mạng được huấn luyện từ trước như: EfficientNetB7, VGG19, Resnet50 và mạng hiệu chỉnh đề xuất EfficientNetB7+. Số chiều của các vector đặc trưng thu được như bảng 3.6.

b. Tham số cho *LDM-FCM* (thực hiện trên môi trường Google-Colab)

c. Tham số trong thuật toán HD-EMR:

$\alpha = 0.99$ ;  $n_b = 350$ ;  $s = 5$  (số điểm neo lân cận gần nhất của một điểm dữ liệu);

Số điểm neo  $p = 4000$  đến  $7000$  điểm neo

**Bảng 3.6. Bảng kết quả tra cứu trên 3 tập dữ liệu với số ảnh trả về là 20**

CSDL	Logo-2K+	VGGFACE2-S	Corel30K
Số điểm neo	(C=5500)	(C=7000)	(C=5000)
VGG19	76.35%	77.78%	65.29%
Resnet50	79.25%	80.23%	68.91%
(B7)	80.20%	81.08%	68.98%
(B7+)	84.90%	86.40%	71.80%
(B7+ và LF)	<b>89.60%</b>	<b>93.50%</b>	<b>82.20%</b>

Từ kết quả thực nghiệm, khi ta dùng mạng trích rút đặc trưng EfficientNetB7+ (gọi tắt là B7+) cho kết quả trong tra cứu tốt hơn so với các mạng tiền huấn luyện (độ chính xác cao hơn các mạng còn lại khoảng 5%-10%). Khi tra cứu với đặc trưng kết hợp cho độ chính xác cao hơn khoảng 5%-7% so với khi chỉ sử dụng mình đặc trưng CNN.

### 3.7. Kết luận chương 3

Trong chương này luận án đã đề xuất các cải tiến nâng cao độ chính xác của hệ thống xếp hạng EMR trong đó các ảnh được biểu diễn ảnh bằng các đặc trưng mức cao (đặc trưng CNN) kết hợp đặc trưng mức thấp trong CBIR thông qua giải quyết hai vấn đề quan trọng sau:

Vấn đề 1: Trích rút đặc trưng mức cao phù hợp với nội dung ảnh từ các mạng CNN đã được huấn luyện và tinh chỉnh mạng CNN để trích rút đặc trưng ảnh của các CSDL riêng biệt. Trong phần này luận án đã tinh chỉnh và đề xuất mạng EfficientNetB7+ là mạng dùng để trích rút đặc trưng mức cao của ảnh (đặc trưng CNN) cho hệ thống CBIR đề xuất. Các thực nghiệm đánh giá độ chính xác kết quả tra cứu ảnh đã được cải thiện khi sử dụng đặc trưng mức cao được trích rút từ mạng EfficientNetB7+.

Vấn đề 2: Đề xuất thuật toán *LDM-FCM* tìm tập điểm neo hiệu quả trên tập dữ liệu vector chiều rất cao. Trong phần này luận án đã xử lý được hai yếu tố then chốt của đồ thị neo là số điểm neo rất lớn và số chiều vector cũng rất cao cùng đồng thời xuất hiện.

Với thuật toán LDM-FCM tìm, luận án cũng đề xuất thuật toán HD-EMR, là một cải tiến thuật toán xếp hạng EMR cho các ảnh được biểu diễn bằng đặc trưng kết hợp mức thấp và mức cao (đặc trưng CNN). Đồ thị của HD-EMR được xây dựng từ các điểm neo thu



được từ thuật toán phân cụm LDM-FCM khi áp dụng lên CSDL vector đặc trưng ảnh. Ngoài ra trong thuật toán HD-EMR, một lần nữa thuật toán tìm kiếm vector ANN lại được áp dụng, nhưng chỉ mục dữ liệu được đánh trên tập dữ liệu các điểm neo để cập nhật lại các giá trị mới sẽ được bổ sung vào ma trận  $Z$  của đồ thị EMR trong quá trình tra cứu ảnh.

Thuật toán HD-EMR trong CBIR đề xuất cho thấy độ chính xác của hệ thống đạt kết quả cao hơn 5-10% so với EMR-(*lvdc-FCM*) và EMR.

## 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, luận án đã tập trung nghiên cứu các vấn đề cơ bản trong tra cứu ảnh tiếp cận dữ liệu đa tạp, các kết quả đạt được của luận án được tóm tắt như sau:

**Kết quả chính 1:** Cải tiến thuật toán xếp hạng đa tạp EMR bằng cách sử dụng thuật toán phân cụm mờ *lvdc-FCM* để tìm điểm neo. Trên cơ sở đó, đề xuất thuật toán *EMR-(lvdc-FCM)* áp dụng hiệu quả trong CBIR.

**Kết quả chính 2:** Nâng cao hiệu quả tra cứu ảnh bằng phương pháp kết hợp đặc trưng cấp thấp và cấp cao trích xuất từ mạng CNN. Chi tiết đóng góp bao gồm: (i) Đề xuất phương pháp tinh chỉnh mạng CNN để trích xuất đặc trưng trước khi kết hợp với đặc trưng mức thấp; (ii) Đề xuất thuật toán HD-EMR dựa trên các điểm neo ước lượng bằng FCM cải tiến kết hợp ANN (thuật toán *LDM-FCM*) để tra cứu ảnh không cần sử dụng các phương pháp giảm chiều dữ liệu.

Ngoài ra, luận án còn đưa ra bổ đề mới về tính tổng quát của dữ liệu đa tạp trên các tập hữu hạn vector, cung cấp cơ sở lý luận cho ứng dụng xếp hạng đa tạp trong tra cứu ảnh theo nội dung. Trong quá trình nghiên cứu, luận án đề xuất sử dụng kiến trúc GPU để thực hiện hiệu quả thuật toán phân cụm *LDM-FCM*.

Luận án 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 kiếm neo và cải tiến thuật toán xếp hạng EMR gốc, đã kiểm chứng hiệu quả của các thuật toán đề xuất thông qua thực nghiệm.

Hướng nghiên cứu tiếp theo của luận án:

(1) Phát triển CSDL ảnh cho bài toán tra cứu ảnh theo nội dung với các chủ đề về ảnh Y tế, ảnh Nông nghiệp cũng như các loại CSDL ảnh phục vụ cho các vấn đề Kinh tế/xã hội...

(2) Nghiên cứu các thuật toán trích rút đặc trưng ảnh bằng các mạng học sâu (deep learning) và phân cụm dữ liệu ảnh bằng các mạng neural đồ thị (Graph Neural Network), các mạng học sâu theo tiếp cận phân đoạn ngữ nghĩa nhằm nâng cao độ chính xác của các hệ thống tra cứu ảnh theo hướng khía cạnh nổi bật (salient aspects).

(3) 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 ngữ nghĩa (semantic segmentation) trên các tập CSDL lớn.

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

- [CT1] **Quy Hoang Van**, Huy Tran Van, Huy Ngo Hoang, Tuyet Dao Van, Sergey Ablameyko, “*A modified Efficient Manifold Ranking Algorithm for Large Database Image Retrieval*”, *Nonlinear Phenomena in Complex Systems*, vol. 23, no. 1 (2020), pp. 79 - 89
- [CT2] **Hoàng Văn Quý**, Ngô Hoàng Huy, Nguyễn Thế Cường, “*Một cải tiến thuật toán FCM để phân cụm dữ liệu lớn và ứng dụng cho truy vấn ảnh*”, Tạp chí nghiên cứu khoa học và công nghệ số 69, tháng 10 năm 2020.
- [CT3] **Quy Hoang-Van**, Cong Nguyen-Dinh, “*Detecting Leftover Food and the Shrimp for Estimating of the Shrimp Body Length Based on CNN*”, May 2021, In book: *Intelligent Systems and Networks* (pp.239-247).
- [CT4] **Hoàng Văn Quý**, Nguyễn Thế Cường, Ngô Hoàng Huy, Trần Văn Huy, Hoàng Xuân Trung, Nguyễn Văn Đoàn, Nguyễn Văn Quyên, “*Một phương pháp tra cứu ảnh hiệu quả kết hợp đặc trưng mức thấp và đặc trưng học sâu nhúng với đánh hạng ta đạp*”, Kỷ yếu Hội nghị Quốc gia lần thứ XXV về điện tử, truyền thông và công nghệ thông tin, 2022, trang 100-106, ISBN 978-604-80-7468-5.
- [CT5] **Van Quy, H.**, Dzung, P.T.K., Huy, N.H., Van Huy, T. (2023). “*Improved EfficientNet Network for Efficient Manifold Ranking-Based Image Retrieval*”. *Intelligent Systems and Networks. ICISN 2023. Lecture Notes in Networks and Systems*, vol 752. Springer, Singapore. [https://doi.org/10.1007/978-981-99-4725-6\\_79](https://doi.org/10.1007/978-981-99-4725-6_79).
- [CT6] **Quy Van Hoang**, Ngo Huy Hoang, Dzung Thi Kim Pham, and Tuyet Dao Van. " *Efficient content-based image retrieval based on anchor point selection in manifold ranking with combined CNN and low-level features*", *Nonlinear Phenomena in Complex Systems*, Vol.26 (366-384), No.4, 2023.