BỘ GIÁO DỤC VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

……..…*……**

CÙ VIỆT DŨNG

NÂNG CAO ĐỘ CHÍNH XÁC CỦA TRA CỨU ẢNH THEO NỘI DUNG DỰA TRÊN TIẾP CẬN HỌC ĐA TẠP TỪ THÔNG TIN PHẢN HỒI CỦA NGƯỜI DÙNG

Chuyên ngành: Khoa học máy tính

Mã số: 9 48 01 01

TÓM TẮT LUẬN ÁN TIẾN SĨ NGÀNH MÁY TÍNH

Hà Nội – 2023

Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Người hướng dẫn khoa học 1: PGS. TS. Nguyễn Hữu Quỳnh
Người hướng dẫn khoa học 2: PGS. TS. Ngô Quốc Tạo

Phản biện 1:
Phản biện 2:

Luận án sẽ được bảo vệ trước Hội đồng chấm luận án tiến sĩ, họp tại Học viện Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi … giờ ..', ngày … tháng … năm 202

Có thể tìm hiểu luận án tại:

- Thư viện Học viện Khoa học và Công nghệ
- Thư viện Quốc gia Việt Nam

LỜI MỞ ĐẦU

1. Tính cấp thiết của luận án

Tra cứu ảnh dựa vào nội dung (Content base image retrieval - CBIR) đã thu hút nhiều sự quan tâm trong những thập kỷ qua. Nó là thách thức to lớn do khoảng trống giữa các đặc trưng mức thấp và các khái niệm ngữ nghĩa mức cao. Để thu hẹp khoảng trống này, phản hồi liên quan (Relevant feedback - RF) được giới thiệu như một công cụ mạnh để tăng cường hiệu năng của CBIR. Chúng ta thấy rằng, bài toán tra cứu ảnh với phản hồi liên quan có một số vấn đề sau: (1) chỉ khám phá các cấu trúc Euclide toàn cục, hoặc chỉ xem xét cấu trúc cục bộ của các mẫu trong cùng một lân cận; (2) số lượng mẫu thu được từ phản hồi của người dùng thường nhỏ và mất cân bằng giữa hai lớp dương và lớp âm; (3) Chưa quan tâm đến nhiều khía cạnh khác nhau của đối tượng dữ liệu ảnh. Do đó, độ chính xác của các phương pháp tra cứu ảnh sử dụng học máy cho phản hồi thường kém hiệu quả.

Do vậy, việc đề xuất phương pháp tra cứu ảnh hiệu quả để giải quyết các hạn chế trên là một nhu cầu cần thiết, chính vì thế mà luận án chọn đề tài "Nâng cao độ chính xác của tra cứu ảnh theo nội dung dựa trên tiếp cận học đa tạp từ thông tin phản hồi của người dùng".

2. Mục tiêu của luận án

Mục tiêu chung của luận án: Nâng cao độ chính xác của tra cứu ảnh dựa trên học đa tạp để giảm chiều từ thông tin phản hồi của người dùng.

Mục tiêu cụ thể của luận án: Đề xuất được một số kỹ thuật tra cứu ảnh để nâng cao độ chính xác tra cứu ảnh bao gồm:

-Đề xuất phương pháp tìm ma trận chiếu tối ưu theo tiếp cận học đa tạp.

-Đề xuất phương pháp tự động bổ sung mẫu dương vào tập huấn luyện, giải quyết vấn đề mất cân bằng của tập huấn luyện. Đồng thời tận dụng các khía cạnh khác nhau của đối tượng để tạo ra một bộ phân lớp mạnh.

3. Các đóng góp của luận án

Luận án có các đóng góp sau:

(1) Đề xuất phương pháp tìm ma trận chiếu tối ưu theo tiếp cận học đa tạp [CT5]. Phương pháp này xem xét cấu trúc cục bộ của các mẫu dương và âm thuộc hai lân cận khác nhau để học một phép chiếu mà dữ liệu có thể phân biệt trên không gian chiếu, dẫn đến cải tiến độ chính xác cho tra cứu ảnh.

(2) Đề xuất phương pháp tự động bổ sung các mẫu dương vào tập huấn luyện để giải quyết vấn đề mất cân bằng tập huấn luyện [CT4]. Phương pháp này có thể: (a) bổ sung một số mẫu dương vào tập huấn luyện; (b) tận dụng các khía cạnh khác nhau của đối tượng để tạo ra một bộ phân lớp mạnh

4. Bố cục của luận án

Luận án này được bố cục thành ba chương:

Chương 1 giới thiệu tổng quan về tra cứu ảnh dựa vào nội dung.

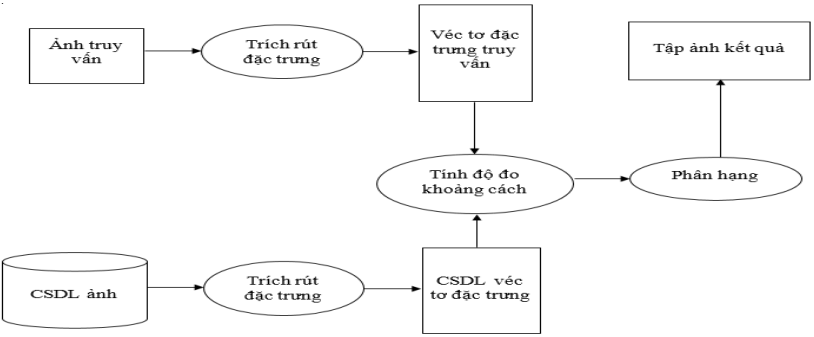
Chương 2 mô tả phương pháp tìm ma trận chiếu tối ưu theo tiếp cận học đa tạp trong tra cứu ảnh, gọi là chiếu phân biệt lớp ngữ nghĩa cho tra cứu ảnh (SCDPIR - Semantic class discriminant projection for image retrieval).

Chương 3 trình bày phương pháp cân bằng tập mẫu phản hồi và kết hợp tra cứu ảnh đa khía cạnh. Cuối cùng, luận án đưa ra một số kết luận và định hướng nghiên cứu trong tương lai

CHƯƠNG 1. TỔNG QUAN VỀ TRA CỨU ẢNH DỰA VÀO NỘI DUNG

1.1. Giới thiệu về tra cứu ảnh

Nhiệm vụ của hệ thống CBIR sử dụng nội dung trực quan được trích rút tự động thành các đặc trưng nhiều chiều và tìm ra một số hình ảnh tương tự với hình ảnh truy vấn trong cơ sở dữ liệu lớn.

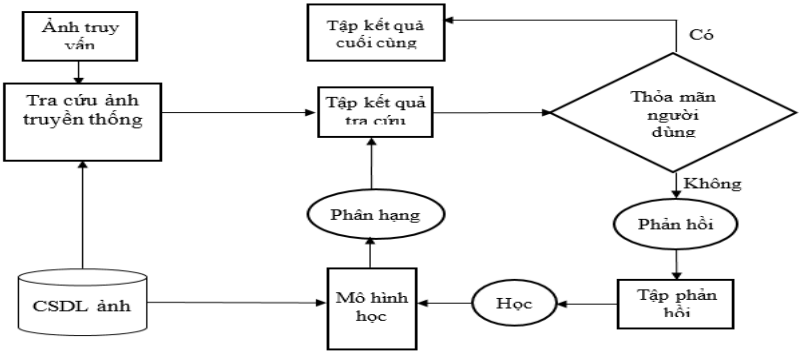


Hình 1.1. Sơ đồ tra cứu ảnh dựa vào nội dung truyền thống

1.2. Giới thiệu về phản hồi liên quan

1.2.1. Cơ chế phản hồi liên quan

Trong CBIR thường đưa người dùng vào mỗi vòng lặp tra cứu, cơ chế này được gọi là "phản hồi liên quan" (relevant feedback - RF).



Hình 1.6. Sơ đồ tra cứu ảnh với phản hồi liên quan

1.2.2. Học đa tạp trong tra cứu ảnh dựa vào nội dung

Việc học đa tạp với mục tiêu là tạo ra một không gian con nơi các ảnh liên quan được chiếu gần nhau trong khi các ảnh không liên quan được chiếu cách xa nhau bằng cách học cấu trúc cục bộ được hình thành bởi lân cận của ảnh truy vấn và ảnh được phản hồi. Điều này đạt được bằng cách nhúng ảnh truy vấn và tập ảnh phản hồi như tập điểm dữ liệu (các nút) trong đồ thị k láng giềng gần nhất có trọng số. Ánh xạ tối ưu được tìm thấy dựa trên ma trận trọng số trên mỗi cạnh, sao cho các điểm lân cận trong đồ thị được ánh xạ với nhau bằng cách tối thiểu hàm chi phí. Mỗi ảnh cơ sở dữ liệu sau đó cũng được ánh xạ sang không gian chiếu mới, thu được kết quả tra cứu mới là tập hàng xóm gần nhất với ảnh truy vấn. Sau mỗi vòng phản hồi, cấu trúc cục bộ của không gian đa tạp lại được học lại.

1.2.3. Rà soát một số nghiên cứu liên quan

Ban đầu, cách tiếp cận tra cứu ảnh với RF giả thiết rằng, tồn tại của một điểm truy vấn lý tưởng mà nếu tìm thấy được sẽ cho kết quả phù hợp với mong muốn của người dùng. Cách tiếp cận này được gọi là "dịch chuyển điểm truy vấn" (QPM - Query Point Movement). Trong RF, các mẫu do người dùng cung cấp thường rất nhỏ so với chiều của đặc trưng, do đó chúng ta phải giải quyết bài toán gọi là "lời nguyền về số chiều - curseof dimensionality". Khi số chiều đặc trưng quá lớn so với số lượng mẫu trong tập huấn luyện, các mô hình học máy có thể rơi vào tình trạng quá khớp. Để giải quyết vấn đề này, một số tác giả đề xuất các kỹ thuật giảm chiều như phân tích thành phần chính (PCA- Principal Components Analysis) [53, 54] và phân tích phân biệt tuyến tính (LDA - Linear Discriminant Analysis) [55]. Trong những năm gần đây, có nhiều thuật toán học đa tạp để giảm chiều đã được đề xuất để khám phá cấu trúc đa tạp. Có thể kể đến một số phương pháp

đa tạp như Locality Preserving Projections, Augmented Relation Embedding, Maximum Margin Projection, Locally Linear Embedding và Laplacian Eigenmaps. Tuy nhiên, các phương pháp này chỉ thực hiện được với các điểm dữ liệu trong tập huấn luyện, và nó không đưa ra rõ ràng phép chiếu có thể thực hiện cho các điểm dữ liệu kiểm tra mới. Bên cạnh đó, các phương pháp này chỉ xem xét tính chất hình học trong một lớp, trong khi bỏ qua mối liên hệ của các mẫu từ các lớp khác nhau. Mặt khác, các phương pháp thường không quan tâm đến các ảnh thuộc lân cận khác nhau mặc dù chúng có thể vẫn liên quan với truy vấn. Do đó, các phương pháp tra cứu ảnh này thường có hiệu quả hạn chế

1.3. Lý thuyết liên quan đến luận án

Trong phần này, trình bày tổng quan ngắn gọn về lý thuyết đồ thị, độ đo khoảng cách và máy véc tơ hỗ trợ, nhân Radial Basis Function và sử dụng nó làm cơ sở cho cơ chế phân hạng cho pha phản hồi trong hệ thống đề xuất được giới thiệu trong các chương sau.

1.4. Đánh giá độ chính xác CBIR

1.4.1. Độ chính xác và độ chính xác trung bình

Để đánh giá hiệu quả của các hệ thống CBIR, độ chính xác được sử dụng. Độ chính xác (precision) là tỷ lệ của số lượng ảnh liên quan với ảnh truy vấn và số lượng tất cả ảnh được hiển thị hàng đầu trả về gọi là phạm vi (scope) cụ thể K, thường được gọi là P@K.

Hiệu quả chính xác tra cứu chung của một hệ thống được đo bằng trung bình tất cả độ chính xác. AP được tính toán như sau:

$$\mathbf{AP} = \frac{\sum_{i=1}^{N} precision(i)}{\mathrm{N}} \qquad (1.1)$$

Với $precision(i)$ là độ chính xác của mỗi truy vấn và N là số lượng ảnh được đưa lần lượt làm ảnh truy vấn.

1.4.2. Một số tập ảnh dữ liệu dùng cho tra cứu ảnh dựa vào nội dung

Tên tập dữ liệu	Số chủ đề	Số ảnh
COREL	80	10800
SIMPLIcity	10	1000
Oxford	11	5062
Caltech 101	101	8742

1.4.3. Kịch bản phản hồi liên quan trong thực nghiệm

Trong hệ thống tra cứu ảnh thực tế, một ảnh truy vấn thường không có trong cơ sở dữ liệu ảnh do đó luận án sử dụng bốn phần kiểm chứng chéo để đánh giá các thuật toán.

Việc lựa chọn thông tin phản hồi được mô phỏng tự động dựa trên thông tin từ tập tin cây nền. Với mỗi truy vấn được gửi, hệ thống tra cứu và phân hạng các ảnh trong cơ sở dữ liệu. Tập kết quả khởi tạo gồm K ảnh hàng đầu sau khi phân hạng được lựa chọn làm các ảnh phản hồi. Người dùng tương tác với hệ thống thông qua đánh dấu trong tập kết quả tra cứu khởi tạo các ảnh có cùng chủ đề (cùng khái niệm) với ảnh truy vấn làm ảnh liên quan (mẫu phản hồi dương) và những ảnh còn lại không đánh dấu làm ảnh không liên quan (mẫu phản hồi âm) và lấy thêm K/2 ảnh tiếp theo được xếp hạng ngay sau tập kết quả tra cứu khởi tạo làm mẫu chưa được gán nhãn

1.5. Kết luận chương 1.

Trong chương 1, luận án đã trình bày lý thuyết tổng quan về một hệ thống tra cứu ảnh dựa vào nội dung và phản hồi liên quan. Bên cạnh đó, cũng phân tích một số phương pháp phản hồi liên quan nhằm giảm khoảng trống ngữ nghĩa. Qua đó, phân tích, đánh giá ưu nhược điểm một số phương pháp CBIR hiện có để đề xuất một số phương pháp nhằm giải quyết những hạn chế đã phân tích.

CHƯƠNG 2. PHƯƠNG PHÁP HỌC CHIẾU PHÂN BIỆT LỚP NGỮ NGHĨA CHO TRA CỨU ẢNH VỚI PHẢN HỒI LIÊN QUAN

Trong chương 2 này, luận án sẽ đề xuất phương pháp học chiếu phân biệt lớp ngữ nghĩa cho giảm chiều trong tra cứu ảnh [CT5] để giải quyết hạn chế: số chiều của đặc trưng thường cao hơn rất nhiều so với số mẫu trong tập phản hồi và các mẫu nằm ở hai không gian con (hai lân cận) khác nhau chưa được xét đến.

2.1. Giới thiệu

Các hình ảnh trong CBIR được thể hiện bằng vectơ đặc trưng thường có kích thước rất cao từ hàng chục đến hàng trăm trong hầu hết các trường hợp nên gặp phải vấn đề "curse of dimensionality". Các phương pháp giảm chiều có thể được áp dụng để giải quyết vấn đề đó bằng cách chiếu các điểm không gian chiều cao sang một không gian khác chiều thấp hơn. Các phương pháp không giám sát xử lý dữ liệu không có nhãn, bao gồm phân tích thành phần chính (PCA), chiếu bảo toàn cục bộ (LPP), nhúng tuyến tính cục bộ (LLE), nhúng bảo toàn lân cận (Neighborhood Preserving Embedding - NPE), và Supervised Isomap (S-Isomap). Các phương pháp học có giám sát tiêu biểu gồm phân tích phân biệt tuyến tính (LDA), chiếu bảo toàn cục bộ tối ưu có giám sát (Supervised Optimal Locality Preserving Projection - SoLPP), phân tích lề Fisher (Marginal Fisher Analysis - MFA), nhúng láng giềng phân biệt (discriminant neighborhood embedding - DNE), chiếu phân biệt phân lớp hồi quy tuyến tính (Linear Regression Classification Steered Discriminative Projection - LRCDP), và nhúng đồ thị bảo toàn phân biệt toàn cục và cục bộ (Discriminative Globality And Locality Preserving Graph Embedding - DGLPGE). Các phương pháp bán giám sát tiêu biểu bao gồm nhúng quan hệ gia tăng

(Augmented Relation Embedding - ARE), chiếu cực đại lề (Maximum margin projection - MMP), và phân tích phân biệt bán giám sát (Semisupervised Discriminant Analysis - SDA)

Các phương pháp kể trên chỉ quan tâm đến nén và tách biệt các điểm thuộc cùng một lân cận mà bỏ qua việc nén và tách biệt các điểm khác lân cận, tức là không đảm bảo các điểm liên quan ngữ nghĩa mà ở các lân cận khác nhau là gần ảnh truy vấn trong không gian con chiều thấp hơn. Bên cạnh đó, các phương pháp nêu trên chỉ thực hiện được với các điểm dữ liệu trong tập huấn luyện, và nó không đưa ra rõ ràng phép chiếu có thể thực hiện cho các điểm thử mới. Do đó, chúng không hiệu quả cho tra cứu ảnh.

Để khắc phục vấn đề trên, luận án đề xuất một phương pháp học chiếu phân biệt lớp ngữ nghĩa (Semantic Class Discriminant Projection - SCDP) [CT5]. Trong SCDP, có thể bảo toàn trung thực cấu trúc cục bộ của các điểm dữ liệu trong không gian đặc trưng trực quan nhiều chiều gốc, quan tâm đến cả điểm khác lân cận và tìm một ma trận chiếu tốt cho chúng.

2.2. Nghiên cứu liên quan

Trong phần này, rà soát ngắn gọn DNE, ARE, MMP, và DAG-DNE, chúng là cơ sở cho phương pháp đề xuất.

2.3. Đề xuất phương pháp học chiếu phân biệt lớp ngữ nghĩa trên dữ liệu đa tạp

Xây dựng hàm mục tiêu

Cho một tập $\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N$ trong $\mathbb{R}^n$, tìm một ma trận biến đổi $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d)$ mà ánh xạ N điểm này thành một tập $\mathbf{y}_1, \mathbf{y}_2, \dots \mathbf{y}_N$ trong $\mathbb{R}^d$ ($d \ll n$) sao cho $\mathbf{y}_i$ biểu diễn $\mathbf{x}_i$, ở đây $\mathbf{y}_i = \mathbf{U}^T\mathbf{x}_i$

Cho $\mathbb{Q} \subset \mathbb{R}^n$ là một không gian đặc trưng ảnh n chiều, và σ: $\mathbb{Q} \times \mathbb{Q} \to \mathbb{R}$ là một hàm khoảng cách nào đó. Cho ma trận $\mathbf{X} =$

$[\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N] \in \mathbb{R}^{n \times N}$ biểu diễn N ảnh trong tập ảnh và N điểm dữ liệu $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_N\}$ được lấy mẫu từ đa tạp con cơ bản M. Giả sử rằng chúng ta có N_1 điểm được gán nhãn, và N_2 điểm còn lại là chưa có nhãn, ở đây $N_1 + N_2 = N$. Để mô hình cấu trúc hình học cục bộ của M, đầu tiên chúng ta xây dựng một đồ thị quan hệ đặc trưng G^F. Với mỗi điểm dữ liệu $\mathbf{x}_\mathrm{i}$, chúng ta tìm k lân cận gần nhất của nó và đặt một cạnh giữa $\mathbf{x}_\mathrm{i}$ và các lân cận của nó thu được ma trận $\mathbf{W}^{\boldsymbol{F}} \in \mathbb{R}^{N \times N}$, được xác định như sau:

$$w_{ij}^F = \begin{cases} e^{-\frac{\rho^2(\mathbf{x}_i, \mathbf{x}_j)}{\tau}}, & \text{nếu } \mathbf{x}_i \in k - NN(\mathbf{x}_j) \\ & \text{hoặc } \mathbf{x}_j \in k - NN(\mathbf{x}_i) \\ 0, & \text{ngược lại;} \end{cases} \quad (2.1)$$

ở đây $\rho^2(\mathbf{x}_i, \mathbf{x}_j)$ là độ đo khoảng cách Euclide (L_2), τ là một số vô hướng dương nào đó, và $k - NN$ là ký hiệu cho k lân cận cận gần nhất.

Với phản hồi liên quan, tôi sử dụng **IR** để biểu thị tập các ảnh không liên quan đến ảnh truy vấn, **R** gồm các ảnh liên quan đến ảnh truy vấn và tập **UL** gồm các ảnh chưa có nhãn. Để khám phá cả thông tin phân biệt và hình học của đa tạp dữ liệu, xây dựng hai đồ thị quan hệ tương tự liên quan G^R và không tương tự G^{IR}.

Các ma trận trọng số $\mathbf{W}^{\boldsymbol{R}} \in \mathbb{R}^{N \times N}$ và $\mathbf{W}^{\boldsymbol{IR}} \in \mathbb{R}^{N \times N}$ của G^R và G^{IR} tương ứng được định nghĩa như sau:

$$w_{ij}^R = \begin{cases} \alpha, \text{nếu } (w_{ij}^F > 0 \wedge w_{ij}^F \leq 1) \wedge (\mathbf{x}_i \in \mathbf{R} \wedge \mathbf{x}_j \in \mathbf{R}) \\ 1, \text{nếu } (w_{ij}^F > 0 \wedge w_{ij}^F \leq 1) \wedge (\mathbf{x}_i \in \mathbf{UL} \wedge \mathbf{x}_j \in \mathbf{UL}) \\ 0, \text{ngược lại;} \end{cases} \quad (2.2)$$

$$w_{ij}^{IR} = \begin{cases} 1, \text{nếu } (w_{ij}^F > 0 \wedge w_{ij}^F \leq 1) \wedge (\mathbf{x}_i \in \mathbf{R} \wedge \mathbf{x}_j \in \mathbf{IR}) \\ \quad \text{hoặc } (w_{ij}^F > 0 \wedge w_{ij}^F \leq 1) \wedge (\mathbf{x}_i \in \mathbf{IR} \wedge \mathbf{x}_j \in \mathbf{R}) \\ 0, \text{ ngược lại;} \end{cases} \quad (2.3)$$

Trong (2.2), khi hai ảnh i và j thuộc cùng một lân cận và cùng nhãn dương, chúng nên nhận một giá trị trọng số cao α.

Chúng ta xác định ma trận $\mathbf{S\_S} \in \mathbb{R}^{N\times N}$ lưu trữ thông tin giống nhau về ngữ nghĩa liên quan với truy vấn giữa hai mẫu $\mathbf{x}_i$ và $\mathbf{x}_j$ (lưu ý rằng hai mẫu $\mathbf{x}_i$ và $\mathbf{x}_j$ không cần thiết thuộc cùng một lân cận):

$$s\_s_{ij} = \begin{cases} 1, \text{nếu } \mathbf{x}_i \in \mathrm{R} \wedge \mathbf{x}_j \in \mathrm{R} \\ 0, \text{ngược lại;} \end{cases} \quad (2.4)$$

Cho $\mathbf{U}$ là một chiếu mà ánh xạ một mẫu x_i trong không gian gốc thành một mẫu tương ứng $\mathbf{y}_i$ trong một không gian chiều thấp hơn.

$$\mathbf{y}_i = \mathbf{U}^T\mathbf{x}_i \quad (2.5)$$

Hiển nhiên trong lân cận cục bộ của một mẫu $\mathbf{x}_i$, trung bình của các mẫu thuộc cùng lân cận và cùng nhãn được tính như sau:

$$\mathbf{m}_i = \sum_j \mathbf{x}_j w_{ij}^R \quad (2.6)$$

Sau khi chiếu, trung bình của các mẫu thuộc cùng lân cận và cùng nhãn có thể được tính từ (2.6) và (2.7)

$$\mathbf{m}_i^{(\mathrm{y})} = \sum_j \mathbf{y}_j w_{ij}^R \quad (2.7)$$

Một tiêu chuẩn cho chọn một ánh xạ tốt là tối ưu hai hàm mục tiêu dưới các ràng buộc thích hợp.

$$\min_{\mathbf{U}} \sum_{ij} (\left\|\mathbf{y}_i - \mathbf{y}_j\right\|^2 w_{ij}^R + \left\|\mathbf{m}_i^{(\mathrm{y})} - \mathbf{m}_j^{(\mathrm{y})}\right\|^2 s\_s_{ij}) \quad (2.8)$$

$$max_{\boldsymbol{U}} \sum_{ij} (\left\|\boldsymbol{y}_i - \boldsymbol{y}_j\right\|^2 w_{ij}^{IR} + \left\|\boldsymbol{m}_i^{(\boldsymbol{y})} - \boldsymbol{m}_j^{(\boldsymbol{y})}\right\|^2 (1 - s\_s_{ij})) \quad (2.9)$$

Phép chiếu tối ưu

Bài toán (2.8) được viết lại như sau:

$$arg\ min_{\mathbf{U}^{\mathrm{T}}\mathbf{U}=\mathbf{I}}\ trace(\mathbf{U}^T\mathbf{C}\mathbf{U}) \quad (2.10)$$

trong đó $\mathbf{C} = \mathbf{C_x} + \mathbf{C_m}$ tương ứng trong $\boldsymbol{s\_s}$

Bài toán tối ưu (2.9) có thể viết lại như sau:

$$arg\ max_{\mathbf{U}^T\mathbf{U}=\mathbf{I}}\ trace(\mathbf{U}^T\mathbf{B}\mathbf{U}) \quad (2.11)$$

trong đó $\mathbf{B} = \mathbf{B_x} + \boldsymbol{B_m}$, tương ứng trong $\boldsymbol{w^{IR}}$

Từ hàm mục tiêu (2.11) và (2.12), vấn đề tìm phép chiếu $\mathbf{y} = \mathbf{U}^T\mathbf{x}$ sẽ được đưa về bài toán tối ưu sau:

$$\mathbf{U} = arg\ max_{\mathbf{U}} \frac{trace(\mathbf{U}^T\mathbf{B}\mathbf{U})}{trace(\mathbf{U}^T\mathbf{C}\mathbf{U})} \quad (2.12)$$

Vậy ma trận $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d)$ gồm d véc tơ lớn nhất tương ứng với các trị riêng $\wedge = diag\ (\lambda_1, \lambda_{2,} \dots, \lambda_d)$ của ma trận $(\mathbf{C}^{-1}.\mathbf{B})$ với điều kiện $\mathbf{C}$ khả nghịch.

Do đó, để nhúng một ảnh truy vấn $\mathbf{q}^{(x)} \in \mathbb{Q}$, chúng ta ánh xạ nó vào đa tạp bởi $\mathbf{q}^{(y)} = \mathbf{U}^{\boldsymbol{T}}\mathbf{q}^{(x)}$. Tìm các điểm lân cận của $\mathbf{q}^{(y)}$ sử dụng khoảng cách Euclide, và phân hạng ở đỉnh trong danh sách trả về.

Thuật toán 2.1. Thuật toán chiếu phân biệt lớp ngữ nghĩa (SCDP).
Input: $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N\} \in \mathbb{R}^n$ gồm N ảnh với R, IR, UL ⊂X,
R : tập ảnh có nhãn dương, IR: tập ảnh có nhãn âm, UL: tập ảnh không có nhãn, d: số chiều không gian chiếu và k, α: các tham số.
Output: Ma trận chiếu $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d)$

Bước 1: $w_{ij}^F \leftarrow \begin{cases} e^{-\frac{\sigma^2(\mathbf{x}_i, \mathbf{x}_j)}{\tau}}, & \text{nếu } \mathbf{x}_i \in k - NN(\mathbf{x}_j) \\ & \text{hoặc } \mathbf{x}_j \in k - NN(\mathbf{x}_i) \\ 0, & \text{ngược lại;} \end{cases}$

Bước 2:

$$w_{ij}^R \leftarrow \begin{cases} \alpha, \text{nếu } (w_{ij}^F > 0 \wedge w_{ij}^F \le 1) \wedge (\mathbf{x}_i \in \mathbf{R} \wedge \mathbf{x}_j \in \mathbf{R}) \\ 1, \text{nếu } (w_{ij}^F > 0 \wedge w_{ij}^F \le 1) \wedge (\mathbf{x}_i \in \mathbf{UL} \wedge \mathbf{x}_j \in \mathbf{UL}) \\ 0, \text{ngược lại;} \end{cases}$$

$$w_{ij}^{IR} \leftarrow \begin{cases} 1, \text{nếu } (w_{ij}^F > 0 \wedge w_{ij}^F \le 1) \wedge (\mathbf{x}_i \in \mathbf{R} \wedge \mathbf{x}_j \in \mathbf{IR}) \\ \quad \text{hoặc } (w_{ij}^F > 0 \wedge w_{ij}^F \le 1) \wedge (\mathbf{x}_i \in \mathbf{IR} \wedge \mathbf{x}_j \in \mathbf{R}) \\ 0, \text{ ngược lại;} \end{cases}$$

$$s\_s_{ij} \leftarrow \begin{cases} 1, \text{ if } \mathbf{x}_i \in \mathbf{R} \wedge \mathbf{x}_j \in \mathbf{R} \\ 0, \text{ ngược lại;} \end{cases}$$

Bước 3:

$\mathbf{B} \leftarrow (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T + (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T$ với $\mathbf{x}_i, \mathbf{x}_j \in w_{ij}^{IR}$ và $\mathbf{m}_i = \sum_j \mathbf{x}_j w_{ij}^R$

$\mathbf{C} \leftarrow (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T + (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T$ với $\mathbf{x}_i, \mathbf{x}_j \in w_{ij}^R$ và $\mathbf{m}_i = \sum_j \mathbf{x}_j w_{ij}^R$

Bước 4: $\mathbf{U} = arg\ max_{\mathbf{U}} \frac{trace(\mathbf{U}^T \mathbf{B} \mathbf{U})}{trace(\mathbf{U}^T \mathbf{C} \mathbf{U})}$ với $(\mathbf{U}^T \mathbf{C} \mathbf{U}) = \mathbf{I}$

$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d)$ với mỗi cột là véc tơ riêng tương ứng với các trị riêng $\lambda_1 > \lambda_2 > \cdots > \lambda_d$.

Độ phức tạp của thuật toán SCDP là $O((n+d)n^2$ trong đó n là số đặc trưng, d là số chiều trong không gian chiếu

2.4. Tra cứu ảnh với học chiếu phân biệt lớp ngữ nghĩa

Thuật toán 2.2. Tra cứu ảnh với học chiếu phân biệt lớp ngữ nghĩa (SCDPIR).

Input: **DB**: Tập ảnh dữ liệu, **q:** Ảnh truy vấn khởi tạo, ***N:*** Số lượng ảnh trả về tại mỗi lần lặp, d: số chiều không gian chiếu

Output: **S*:*** Tập ảnh kết quả

Bước 1: **X** ←Retrieval-Init(**q**, **DB**, *N*>;

Bước 2: **Repeat**

Bước 2.1: **IR**←***Feedback***(**X**, −1);

Bước 2.2 **R**←***Feedback***(**X**, 1)*;*

Bước 2.3 **UL**←**X** − (**IR** ∪ **R**)

Bước 2.4 $\mathbf{U} \leftarrow \boldsymbol{SCDP}(\mathbf{X}, \mathbf{R}, \mathbf{IR}, d, k, \alpha)$;

Bước2.5 $\mathbf{DB}^{(\mathbf{y})} \leftarrow \boldsymbol{Mapping}(\mathbf{DB}, \mathbf{U})$;

$\mathbf{q}^{(\mathbf{y})} \leftarrow \boldsymbol{Mapping}(\mathbf{q}, \mathbf{U})$

Bước 2.6 $\mathbf{S} \leftarrow$ Retrieval< $\mathbf{q}^{(\mathbf{y})}$, $\mathbf{DB}^{(\mathbf{y})}$, N>;

until (Người dùng dừng phản hồi);

Bước 3. **Return S**;

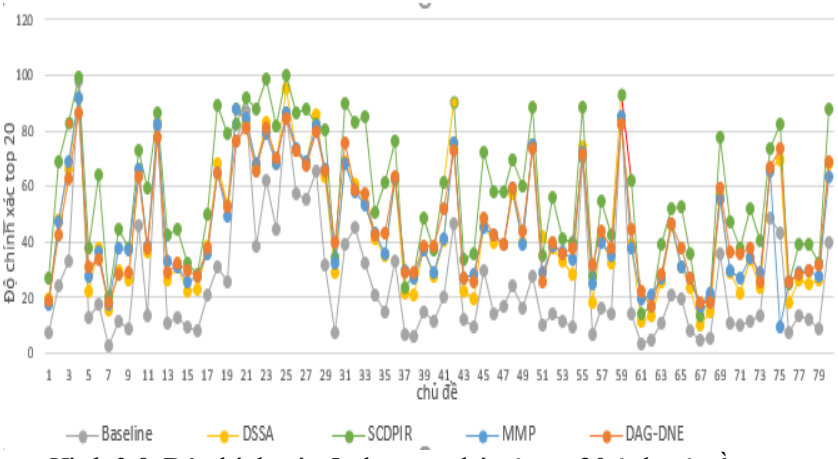
Độ phức tạp là $O(l + (\mathrm{n} + \mathrm{d})\mathrm{n}^2)$ trong đó l là số ảnh, n là số chiều của không gian đặc trưng gốc và d là số chiều của không gian chiếu.

2.5. Đánh giá hiệu năng tra cứu ảnh với học chiếu phân biệt lớp ngữ nghĩa u năng tra cứu ảnh

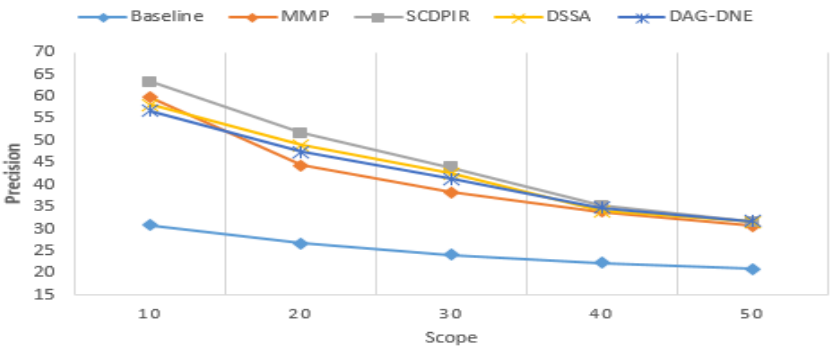
2.5.1. Độ chính xác tra cứu ảnh

So sánh độ chính xác của thuật toán tra cứu ảnh đề xuất với baseline, MMP, DSSA và DAG-DNE dùng tham số k=12, $\alpha = 50$.

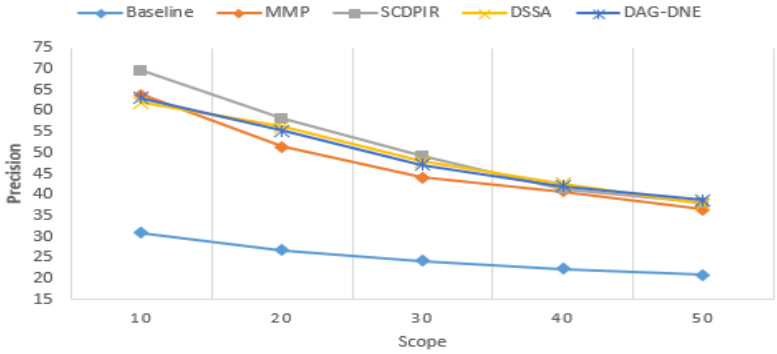
Kết quả của tập ảnh dữ liệu Corel



Hình 2.8. Độ chính xác 5 phương pháp ở top 20 ảnh trả về



a) lần lặp phản hồi thứ nhất



(b) lần lặp phản hồi thứ hai

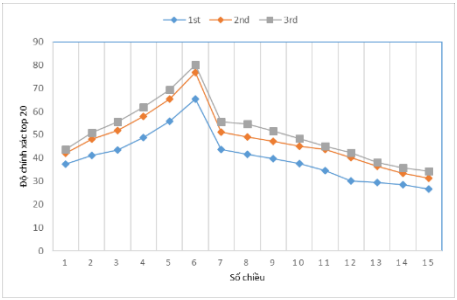
Hình 2.9. Các đường cong precision-scope trung bình của các thuật toán khác nhau cho hai lần lặp đầu tiên.

Kết quả của tập dữ liệu ảnh SIMPLIcity

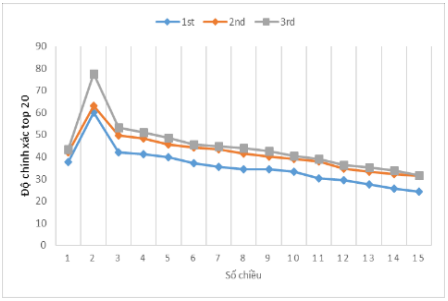
Với tập Corel 10K8 cho ta thấy hiệu năng của phương pháp đề xuất đã cải thiện đáng kể, nhưng để trực quan hóa phép chiếu phương pháp đề xuất tập Corel không tối ưu vì số lượng ảnh quá nhiều. Do đó trong phần này, các thực nghiệm được thực hiện trên tập dữ liệu ảnh

SIMPLIcity có 1000 ảnh để trình bày việc trực quan hóa kết quả của bốn phương pháp MMP, DSSA, DAG-DNE và SCDPIR

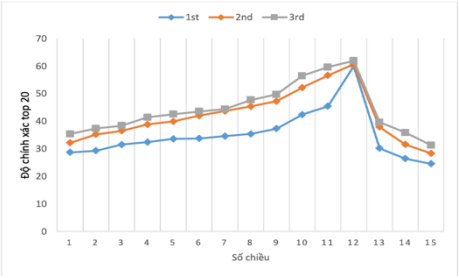
2.5.2. Chiều của không gian chiếu phân biệt lớp ngữ nghĩa



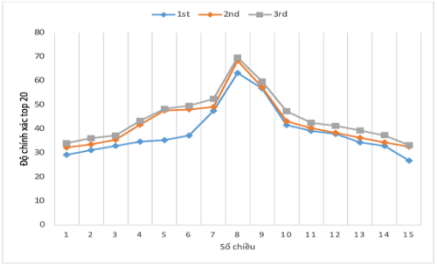
(a) Chiều của không gian (SCDP)



(b) Chiều của không gian (MMP)



(c) Chiều không gian(DAG-DNE)



(d) Chiều không gian (DSSA)

Hình 2.11. Hiệu năng của bốn phương pháp theo số chiều

Chúng ta thấy rằng hiệu năng của MMP luôn nhận được hiệu năng tốt nhất tại hai chiều (Hình 2.11 (b)), hiệu năng của SCDP luôn có hiệu năng tốt nhất tại sáu chiều (Hình 2.11 (a)), DSSA đạt hiệu năng tốt nhất tại số chiều rất lớn là 8 chiều (Hình 2.11 (d)), và DAG-DNE đạt hiệu năng tốt nhất tại số chiều rất lớn là 12 chiều (Hình 2.11 (c)). Như vậy, số chiều chiếu tối ưu của SCDPIR cao hơn của MMP nhưng thấp hơn của DAG-DNE và DSSA. Nhưng, hiệu suất của SCDPIR cao hơn nhiều so với MMP khi nó ở số chiều tương đối thấp và điều này có thể

chấp nhận được trong các ứng dụng thực tế. Ngoài ra, với thuật toán DAG-DNE, hiệu năng đạt được tốt nhất với số chiều tương đối lớn cao và nó sẽ bị vấn đề quá khớp khi áp dụng trong các ứng dụng tại thế giới thực

2.6. Kết luận chương 2

Trong chương này, luận án trình bày phương pháp SCDP có thể khám phá được cấu trúc phi tuyến của dữ liệu trên không gian gốc để tìm được ma trận chiếu. Bên cạnh đó, trong chương 2 đã đánh giá thực nghiệm trên hai tập dữ liệu Corel 10K8 và SIMPLIcity đã thể hiện độ chính xác của phương pháp đề xuất đã được cải thiện và đáng tin cậy

CHƯƠNG 3. CÂN BẰNG TẬP MẪU PHẢN HỒI VÀ KẾT HỢP TRA CỨU ẢNH ĐA KHÍA CẠNH

3.1. Giới thiệu

Các bài toán phản hồi liên quan rất khác so với bài toán phân lớp truyền thống bởi vì các phản hồi được cung cấp bởi người dùng thường bị giới hạn trong các hệ thống tra cứu ảnh thực. Do đó, các phương pháp học mẫu nhỏ là hứa hẹn cho RF. Tuy nhiên, hầu hết các cách tiếp cận không quan tâm đến những ảnh chưa được gán nhãn dương hoặc âm dù chúng rất là hữu ích cho quá trình học phản hồi hay giảm chiều để nâng cao độ chính xác tra cứu. Bên cạnh đó, chúng còn bỏ qua sự cân bằng số mẫu dương và âm trong tập phản hồi.

Trong chương 3, đề xuất một phương pháp cân bằng tập mẫu phản hồi và kết hợp tra cứu ảnh đa khía cạnh (CIR) [CT4] thực hiện (a) bổ sung một số mẫu dương nhằm xây dựng tập mẫu cân bằng (BSFG - balanced sample feedback based on the graph); (b) tận dụng thông tin hình học trong việc giảm chiều hiệu quả (SCDP) (đã trình bày trong chương 2); (c) tận dụng các khía cạnh của đối tượng để xây dựng bộ phân lớp mạnh (CMAC).

3.2. Kỹ thuật cân bằng tập mẫu phản hồi sử dụng học bán giám sát đồ thị

Cho đồ thị lân cận gần nhất $\mathrm{G} = (\mathrm{X}, \mathrm{S})$ là một đồ thị vô hướng với tập đỉnh $\mathrm{X} = \{\mathrm{x}_1, \mathrm{x}_2, \dots \mathrm{x}_N\} \in R^n$. N đỉnh (ảnh) này là kết quả của việc thực hiện truy vấn trước đấy.

Giả sử rằng đồ thị G được đánh trọng số, tức là mỗi cạnh giữa hai đỉnh x_i và x_j mang một trọng số không âm $\mathrm{s}_{ij} \geq 0$. Ma trận kề có trọng số của đồ thị là ma trận $S = (s_{ij})_{i,j=1,\dots N}$.

Gọi $kNN(\mathrm{x}_i)$ là k lân cận gần nhất của điểm x_i. Nếu $\mathrm{x}_i \in kNN(\mathrm{x}_j)$ (hoặc $\mathrm{x}_j \in kNN(\mathrm{x}_i)$), $\mathrm{s}_{ij} = 1$. Ngược lại, $\mathrm{s}_{ij} = 0$. Do G là vô hướng chúng ta yêu cầu $\mathrm{s}_{ij} = \mathrm{s}_{ji}$.

Giả sử có m điểm đã được người dùng gán nhãn (bao gồm cả ảnh truy vấn gốc) $LX = \{\mathrm{x}_1, \mathrm{x}_2, \dots \mathrm{x}_m\} \in R^n$ và $N - m$ điểm chưa được người dùng gán nhãn $UX = \{\mathrm{x}_{N-m+1}, \mathrm{x}_{N-m+2}, \dots \mathrm{x}_{N-m}\} \in R^n$. Để phục cho việc xác định điểm x_i, nơi mà lớp dương có mật độ cao xung quanh điểm đó, xây dựng đồ thị G^{label}.

Đồ thị G^{label} có các đỉnh giống như các đỉnh của đồ thị G và có ma trận trọng số S^{label}. Cho $label(\mathrm{x}_i)$ là nhãn của điểm x_i (nhãn này hoặc là liên quan hoặc là không liên quan). Với mỗi điểm x_i, tập $kNN^{label}(\mathrm{x}_i)$ bao gồm các điểm lân cận của x_i mà có cùng nhãn với x_i hoặc chưa có nhãn. Lý do của việc này là chúng ta xem những điểm đủ gần với x_i dường như là có liên quan đến x_i. Cụ thể:

$$kNN^{label}(\mathrm{x}_i) = \{\mathrm{x} |\ label(\mathrm{x}) == label(\mathrm{x}_i)\ hoặc\ \mathrm{x}\ \in UX\} \quad (3.1)$$

Chúng ta xác định S^{label} là ma trận trọng số của G^{label} như sau:

$$s_{ij}{}^{label} = \begin{cases} \beta\text{, nếu } label(\mathrm{x}_i) == label(\mathrm{x}_j) \\ 1\text{, nếu } \mathrm{x}_i \text{ và } \mathrm{x}_j\ \in UX \text{ nhưng } \mathrm{x}_i\ \in kNN^{label}(\mathrm{x}_j) \\ \qquad\qquad hoặc\ \mathrm{x}_j\ \in kNN^{label}(\mathrm{x}_i) \\ 0\text{, ngược lại} \end{cases} \quad (3.2)$$

Trong (3.2), giá trị β cao hàm ý hai ảnh có cùng nhãn và do đó có cùng ngữ nghĩa.

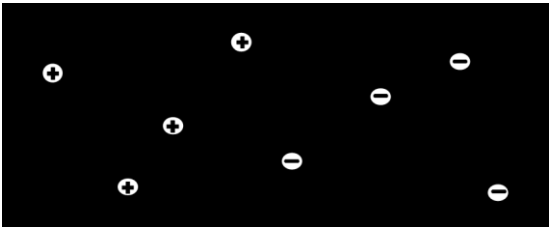
Trên đồ thị G^{label}, bậc của đỉnh $\mathrm{x}_i \in X$ được xác định bằng:

$$d_i{}^{label} = \sum_{j=1}^{N} s_{ij}{}^{label} \quad (3.3)$$

Với mỗi điểm chưa được gán nhãn $\mathrm{x}_i \in UX$, tìm điểm có bậc $d_i{}^{label}$ cao nhất trong số các điểm thuộc lân cận $kNN^{label}(\mathrm{x}_i)$ và lấy nhãn của điểm đó làm nhãn tạm thời của x_i. Cụ thể: Nhãn tạm thời của x_i sẽ được gán là nhãn của x^* với x^* được xác định như sau:

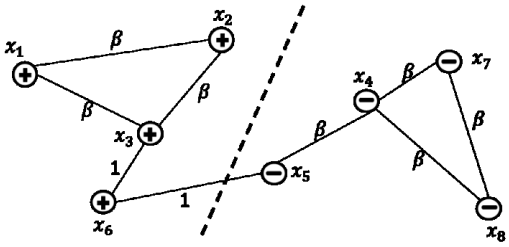
$$x^* = \underset{x_j \in kNN^{label}(x_i)}{\operatorname{argmax}} (d_j^{label}) \quad (3.4)$$

Thủ tục xác định nhãn tạm thời này được minh họa trên Hình 3.5.



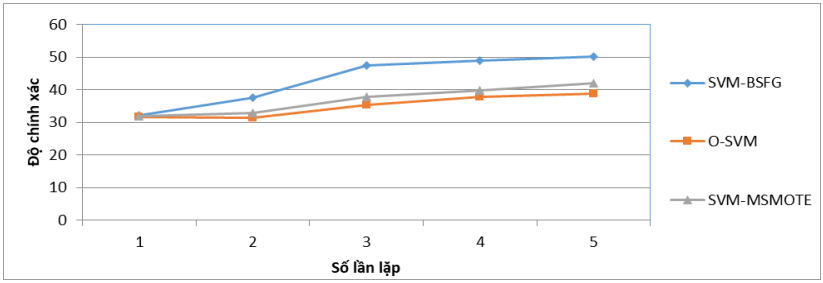
Hình 3.5. Minh họa xác định nhãn tạm thời

Ý tưởng để xác định nhãn cuối cùng của một điểm x_i như sau. Đầu tiên, phân hoạch đồ thị theo Ncut thành hai lớp: lớp âm và lớp dương. Sau đó, kiểm tra xem điểm x_i thuộc lớp nào



Hình 3.6. Đồ thị G^{label} được phân chia theo tiêu chí Ncut.

Hiệu năng của BSFG



Hình 3.7. Độ chính xác của ba phương pháp O-SVM, SVM-MSMOTE, và SVM-BSFG.

3.3. Kỹ thuật kết hợp các bộ phân lớp theo khía cạnh

Vấn đề cân bằng mẫu đã giải quyết được thông qua học bán giám sát dựa vào đồ thị. Tuy nhiên, nó chưa khám phá được thuộc tính thống kê cho phân lớp dữ liệu. Với nhận định rằng, không có một bộ phân lớp nào có thể biểu diễn được tất cả các khía cạnh hữu ích của dữ liệu đầu vào. Với các khía cạnh khác nhau của một mẫu đang xét, các bộ phân lớp này có thể được huấn luyện độc lập trên tập mẫu theo khía cạnh đã có nhãn. Các bộ phân lớp con có thể được tổ hợp thành một bộ phân lớp mạnh theo kỹ thuật bầu cử đa số. Trong luận án, một khía cạnh được xác định là một đặc trưng: màu, hình dạng hoặc kết cấu. Bài toán được phát biểu thành thuật toán tổ hợp các bộ phân lớp theo khía cạnh (Combine Multiple Aspect Classiers - CMAC).

Thuật toán 3.2 Thuật toán kết hợp bộ phân lớp theo khía cạnh (CMAC)

Input: reduced_Aspect$_i$, i = 1,…, k **:** Các tập mẫu theo khía cạnh đã giảm chiều:

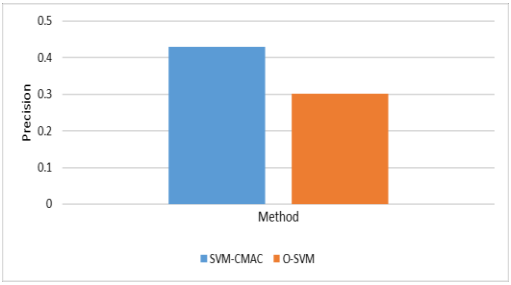
Output: β: Bộ phân lớp được kết hợp:

Bước 1: For i=1, ..., k

$$C^i \leftarrow \text{Aspect Classifiers (reduced\_Aspect}_i\text{)};$$

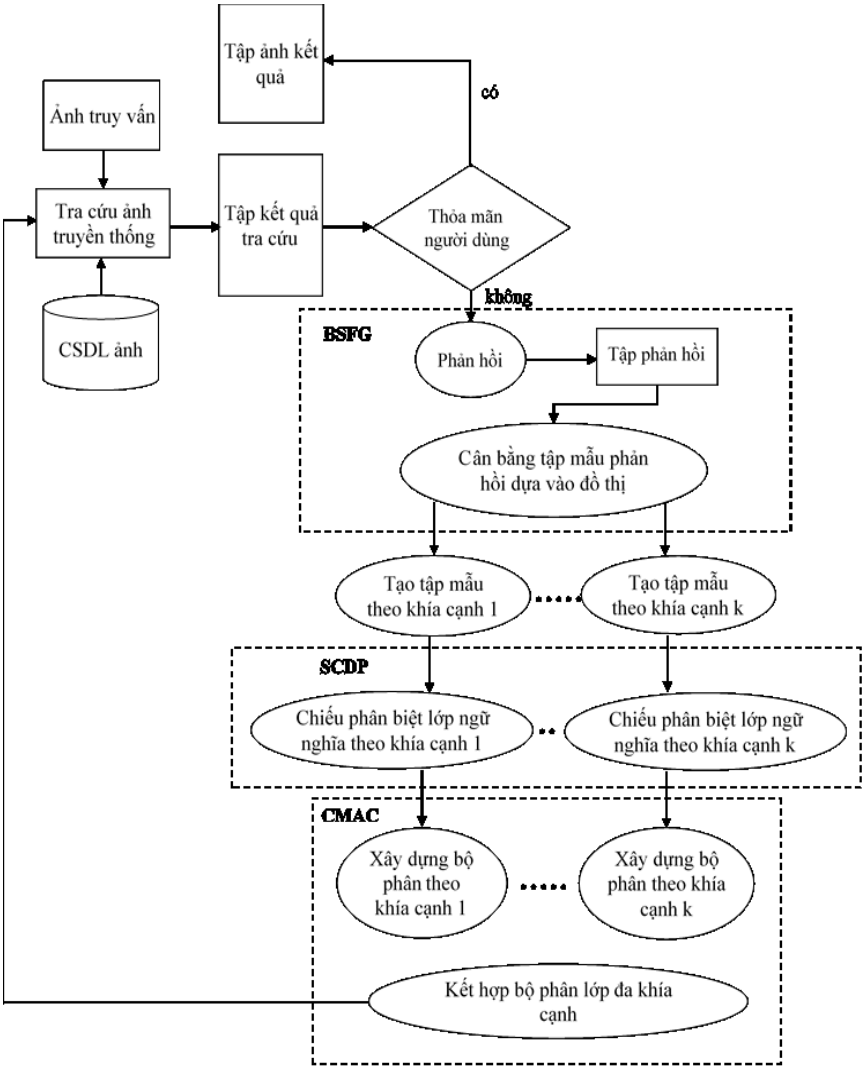
Bước 2: $\beta(x) = \underset{y\in\{-1,\ 1\}}{\operatorname{argmax}} \sum_b \delta_{sgn\left(C^i(x)\right),y}$

Hiệu năng của CMAC



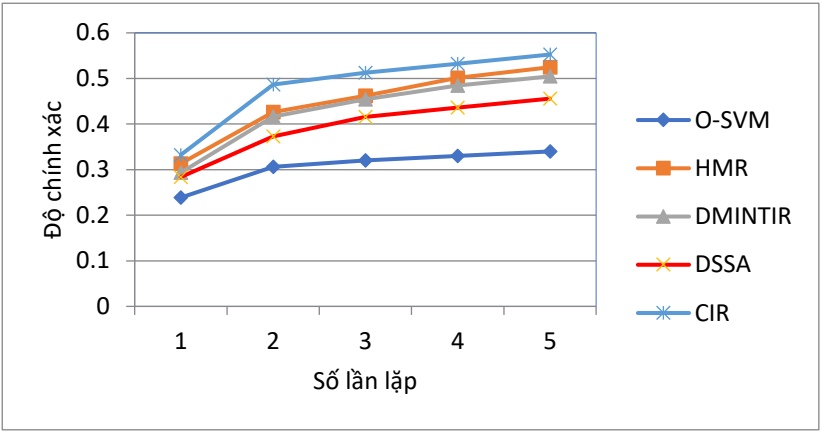
Hình 3.8. Độ chính xác của O-SVM và SVM-CMAC

3.4. Phương pháp tra cứu ảnh kết hợp chiếu phân biệt lớp ngữ nghĩa đa khía cạnh.



Hình 3.9. Sơ đồ tra phương pháp tra cứu ảnh kết hợp chiếu phân biệt lớp ngữ nghĩa đa khía cạnh

3.5. Đánh giá độ chính xác của phương pháp CIR



Hình 3.10. Độ chính xác của năm phương pháp

3.6. Kết luận chương 3

Trong luận án đã đề xuất phương pháp CIR nâng cao độ chính xác của hệ thống tra cứu sử dụng RF có thể: (1) tận dụng được thông tin của các mẫu chưa có nhãn; (2) khai thác được cấu trúc phi tuyến của dữ liệu đa tạp và (3) tận dụng được các khía cạnh hữu ích khác nhau của đối tượng. Các kết quả thực nghiệm trên tập dữ liệu ảnh ảnh Corel đã chỉ ra rằng phương pháp đề xuất đã cải tiến đáng kể độ chính xác tra cứu

KẾT LUẬN

Độ chính xác của một hệ thống tra cứu ảnh dựa vào nội dung đã và đang được cộng đồng nghiên cứu quan tâm cải tiến. Nhiều phương pháp đã được đề xuất trong thời gian qua. Tuy nhiên, sự chênh lệch giữa đặc trưng mức thấp của ảnh và cảm nhận trực quan từ người dùng về nội dung ảnh làm cho độ chính xác của hệ thống tra cứu ảnh vẫn còn khoảng cách với nhu cầu của người dùng. Các đóng góp chính trong luận án này cũng theo định hướng sử dụng cơ chế phản hồi liên quan để thu hẹp sự chênh lệch khoảng cách này.

Luận án đã có các đóng góp sau:

(1) Đề xuất phương pháp tìm ma trận chiếu tối ưu theo tiếp cận học đa tạp [CT5]. Phương pháp này xem xét cấu trúc cục bộ của các mẫu dương và âm thuộc hai lân cận khác nhau để học một phép chiếu mà dữ liệu có thể phân biệt trên không gian chiếu, dẫn đến cải tiến độ chính xác cho tra cứu ảnh.

(2) Đề xuất phương pháp tự động bổ sung các mẫu dương vào tập huấn luyện để giải quyết vấn đề mất cân bằng tập huấn luyện [CT4]. Phương pháp này có thể: (a) bổ sung một số mẫu dương vào tập huấn luyện; (b) tận dụng các khía cạnh khác nhau của đối tượng để tạo ra một bộ phân lớp mạnh

Một số vấn đề cần được nghiên cứu tiếp trong tương lai:

- Nghiên cứu mạng nơ ron tích chập để nâng cao độ chính xác tra cứu trên tập ảnh lớn hơn.
- Nghiên cứu áp dụng cơ chế băm sâu để nâng cao tốc độ tra cứu.
- Từng bước tiến đến việc đưa hệ thống vào áp dụng một số lĩnh vực trong cuộc sống.

NHỮNG ĐÓNG GÓP MỚI CỦA LUẬN ÁN

Nhằm mục tiêu nâng cao độ chính xác của tra cứu ảnh sử dụng học máy để giảm chiều từ thông tin phản hồi của người dùng, luận án có các đóng góp sau:

(1) Đề xuất phương pháp tìm ma trận chiếu tối ưu theo tiếp cận học đa tạp [CT5]. Phương pháp này xem xét cấu trúc cục bộ của các mẫu dương và âm thuộc hai lân cận khác nhau để tìm phép chiếu, đảm bảo tính phân biệt trên không gian chiếu, đồng thời cải tiến độ chính xác tra cứu ảnh.

(2) Đề xuất phương pháp tự động bổ sung mẫu dương vào tập huấn luyện, giải quyết vấn đề mất cân bằng của tập huấn luyện [CT4]. Phương pháp này bổ sung các mẫu dương vào tập huấn luyện đồng thời tận dụng các khía cạnh khác nhau của đối tượng để tạo ra một bộ phân lớp mạnh.

DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ

Trong nước:

[CT1] Cù Việt Dũng, Nguyễn Hữu Quỳnh, An Hồng Sơn, Đào Thị Thúy Quỳnh, Cải tiến tra cứu ảnh thông qua kết hợp các bộ phân lớp không gian con ngẫu nhiên, *Kỷ yếu Hội nghị KHCN Quốc gia lần thứ XII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin*, **2018,** 72-78

[CT2] Cù Việt Dũng, Nguyễn Hữu Quỳnh, Ngô Quốc Tạo, Trần Thị Minh Thu, Một phương pháp tra cứu ảnh học biểu diễn và học đa tạp cho giảm chiều với thông tin từ người dùng, *Kỷ yếu Hội nghị KHCN Quốc gia lần thứ XII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin*, **2019,** 307-314

[CT3] Cù Việt Dũng, An Hồng Sơn, Nguyễn Hữu Quỳnh, Ngô Quốc Tạo, Đào Thị Thúy Quỳnh, Phương pháp học bán giám sát dựa vào đồ thị xây dựng tập mẫu cân bằng cho tra cứu ảnh, *Kỷ yếu Hội nghị KHCN Quốc gia lần thứ XII về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin*, **2021,** 143-149

Quốc tế:

[CT4] Nguyen Huu Quynh, Cu Viet Dung, Dao Thi Thuy Quynh, Ngo Quoc Tao, Phuong Van Canh, Graph-based semisupervised and manifold learning for image retrieval with SVM-based relevant feedback, *Journal of Intelligent & Fuzzy Systems(SCIE,IF=1.637)*, **2019**, *37*, 711–722

[CT5] Nguyen Huu Quynh, Cu Viet Dung, Dao Thi Thuy Quynh, (2021), Semantic class discriminant projection for image retrieval with relevance feedback. *Multimedia Tools and Applications (SCIE, IF = 2.313, Q1),* **2021,** *80*, 15351–15376