



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0042067

(51)^{2020.01} G06K 9/00

(13) B

(21) 1-2021-07930

(22) 09/12/2021

(45) 25/12/2024 441

(43) 26/06/2023 423

(73) CÔNG TY CỔ PHẦN NGHIÊN CỨU VÀ ỨNG DỤNG TRÍ TUỆ NHÂN TẠO
VINAI (VN)

Tòa nhà văn phòng Symphony, Đường Chu Huy Mân, KĐT sinh thái Vinhomes
Riverside, Phường Phúc Lợi, Quận Long Biên, Thành phố Hà Nội, Việt Nam

(72) Nguyễn Xuân Bắc (VN); Bùi Đức Toàn (VN).

(74) Công ty TNHH Tâm nhìn và Liên danh (VISION & ASSOCIATES CO.LTD.)

(54) PHƯƠNG PHÁP PHÂN NHÓM BẰNG CÁCH SỬ DỤNG MÔ HÌNH BỘ MÃ
HÓA-BỘ GIẢI MÃ DỰA TRÊN CƠ CHẾ CHÚ Ý VÀ VẬT GHI DỪNG ĐỂ
NHẬN DẠNG ẢNH

(57) Sáng chế đề cập đến phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý và vật ghi. Phương pháp này trích xuất các đặc trưng ảnh, phân nhóm để tạo thành các nhóm vector đặc trưng ảnh, và dựa vào điểm số tương đồng cosin giữa các vector đặc trưng ảnh để sắp xếp mỗi nhóm vector đặc trưng ảnh thành chuỗi các vector đặc trưng ảnh. Chuỗi các vector đặc trưng ảnh này có thể bao gồm các vector mã hóa khoảng cách cosin được ghép nối với các vector đặc trưng ảnh tương ứng và được sử dụng làm chuỗi dữ liệu đầu vào trong các mô hình mạng nơ-ron mã hóa và giải mã để tạo ra chuỗi dữ liệu đầu ra từ chuỗi dữ liệu đầu vào. Chuỗi dữ liệu đầu ra là chuỗi nhị phân có giá trị ở một vị trí là 1 hoặc 0 biểu thị ảnh tương ứng với vị trí đó thuộc cùng nhóm hoặc khác nhóm so với ảnh trung tâm của nhóm.

S101: trích xuất các đặc trưng ảnh từ cơ sở dữ liệu ảnh có chứa nhiều ảnh để thu được bộ dữ liệu đặc trưng ảnh gồm có các vector đặc trưng



S102: phân nhóm các vector đặc trưng ảnh được lấy mẫu từ bộ dữ liệu đặc trưng ảnh nêu trên thành các nhóm đặc trưng ảnh dựa trên điểm số tương đồng cosin



S103: tạo ra chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin có chứa thông tin mã hoá khoảng cách cosin và các vector đặc trưng ảnh



S104: sử dụng chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin để làm chuỗi dữ liệu đầu vào của mạng nơron mã hoá



S105: giải mã, bởi mạng nơron giải mã, để tạo ra chuỗi dữ liệu đầu ra, trong đó dữ liệu đầu vào của mạng nơron giải mã chính là dữ liệu đầu ra của mạng nơron mã hóa

Lĩnh vực kỹ thuật được đề cập

Sáng chế thuộc lĩnh vực trí tuệ nhân tạo và đề cập đến phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý, và vật ghi có chứa chương trình máy tính để thực hiện phương pháp này. Cụ thể hơn, phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý tạo ra chuỗi dữ liệu đầu vào từ các nhóm đặc trưng ảnh sử dụng thông tin về điểm số tương đồng cosin, để giải mã thành chuỗi dữ liệu đầu ra thông qua các mạng nơron mã hoá và giải mã, trong đó mỗi vị trí trong chuỗi dữ liệu đầu ra có thể tương ứng với một ảnh và dựa vào giá trị tại vị trí trong chuỗi dữ liệu đầu ra để nhận dạng hoặc phân loại ảnh.

Tình trạng kỹ thuật của sáng chế

Kỹ thuật nhận dạng ảnh, mà cụ thể hơn là phân loại (hoặc phân nhóm) ảnh khuôn mặt người hoặc ảnh phong cảnh (có thể được gọi chung là phân loại trực quan hoá) đã và đang dành được sự quan tâm rất lớn trong lĩnh vực học máy. Các giải pháp là phân loại ảnh khuôn mặt người hoặc ảnh phong cảnh có thể được chia thành ba nhóm chính bao gồm các phương pháp phân loại trực quan hoá học không giám sát, các phương pháp phân loại trực quan hoá học nửa giám sát (semi-supervised), và các phương pháp phân loại trực quan hoá học có giám sát.

Dựa trên các đặc trưng của dữ liệu trực quan hoá là dễ thu thập, do đó trong thực tế có thể tiếp cận được các kho dữ liệu khổng lồ các ảnh trực quan hoá. Tuy nhiên, việc khai thác thông tin từ các ảnh trực quan hoá này có thể nói là tương đối khó khăn, chẳng hạn như trong việc ghi chú thích (ví dụ trích xuất đặc trưng trong ảnh để thể hiện thành các thông tin gắn với ảnh, nhận diện ảnh, phân loại ảnh, phân nhóm ảnh, hoặc tương tự), lý do là có nhiều yếu tố phức tạp có thể ảnh

hưởng tới ảnh trực quan hoá, ví dụ độ sáng, các tư thế chụp, chẳng hạn phụ thuộc theo các hoàn cảnh chụp ảnh thực tế. Do đó việc nghiên cứu, đề xuất và đưa ra các mô hình được tham số hoá thực hiện việc khai thác thông tin từ các ảnh trực quan hoá, ví dụ phân loại các ảnh trực quan hoá, có hiệu quả được nâng cao cơ bản là quan trọng và cần thiết.

Một trong số các mô hình được biết đến rộng rãi là các mạng GCN (Graph Convolutional Network – Mạng tích chập đồ thị), giải quyết tốt vấn đề phân loại trực quan hoá theo cách học không giám sát. Các mạng GCN sử dụng cùng các ý tưởng tương đồng trong lý thuyết đồ thị phổ để thiết kế các bộ lọc được tham số hoá để thích hợp trong các mạng CNN (Convolutional Neural Network – Mạng nơron tích chập), và đã được thể hiện là một trong số các phương pháp hiệu quả nhất trong việc giải quyết phân loại các mẫu phức tạp. Một số ví dụ về các mạng GCN đã được bộc lộ trong các bài báo có tiêu đề là “Learning to cluster faces via confidence and connectivity estimation”, của tác giả Lei Yang và các cộng sự, được đăng tải tại biên bản lưu hội nghị IEEE về thị giác máy tính và hội thảo nhận dạng mẫu, 2020; và bài báo có tiêu đề là “Learning to cluster faces on an affinity graph”, của tác giả Lei Yang và các cộng sự, được đăng tải tại biên bản lưu hội nghị IEEE về thị giác máy tính và hội thảo nhận dạng mẫu, 2019.

Nói chung, các mạng GCN hướng tới việc tạo ra các đồ thị có tính tương tự (affinity graph), sử dụng các vector đặc trưng ảnh được lấy mẫu trong cơ sở dữ liệu ảnh trực quan hoá để làm các đỉnh, trong đó các đỉnh lân cận được nối với nhau dựa trên điểm số tương đồng cosin (cosine similarity) giữa các vector đặc trưng ảnh. Đồ thị có tính tương tự nêu trên thường là đồ thị ở quy mô rộng, có thể chứa hàng triệu đỉnh đồ thị, và do đó các mạng GCN được cho là có khối lượng tính toán lớn, và có yêu cầu cao về sử dụng bộ nhớ. Thêm vào đó, các mạng này khá nhạy cảm đối với các mẫu nặng và nhiễu.

Do đó, có nhu cầu về giải pháp được cải tiến liên quan đến kỹ thuật nhận dạng ảnh, có thể giảm thiểu yêu cầu về sử dụng bộ nhớ và khối lượng tính toán, có thể đạt được hiệu quả tốt ngay cả với các mẫu nặng và nhiễu.

Bản chất kỹ thuật của sáng chế

Mục đích của sáng chế là đề xuất phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý, có thể khắc phục được một hoặc một số vấn đề nêu trên.

Mục đích khác của sáng chế là đề xuất phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý, cơ bản là có thể giảm yêu cầu về sử dụng bộ nhớ và khối lượng tính toán, cơ bản là có thể đạt được hiệu quả tốt ngay cả với các mẫu nặng và nhiều.

Cần hiểu rằng, sáng chế không bị giới hạn ở các mục đích được nêu ra trên đây, ngoài các mục đích này sáng chế còn có thể bao gồm các mục đích khác nữa mà sẽ hiển nhiên đối với người có hiểu biết trung bình về lĩnh vực kỹ thuật tương ứng, được nêu rõ hoặc hàm chứa trong các nội dung mô tả dưới đây.

Để đạt được một hoặc một số mục đích nêu trên, sáng chế đề xuất phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý, phương pháp này bao gồm các bước:

trích xuất các đặc trưng ảnh từ cơ sở dữ liệu ảnh X có chứa nhiều ảnh x_i , bởi mô hình trích xuất đặc trưng ảnh, để thu được bộ dữ liệu đặc trưng ảnh gồm có các vector đặc trưng ảnh f_i , trong đó mỗi vector đặc trưng ảnh f_i tương ứng với một ảnh x_i trong cơ sở dữ liệu ảnh X đã nêu;

phân nhóm các vector đặc trưng ảnh được lấy mẫu từ bộ dữ liệu đặc trưng ảnh nêu trên thành các nhóm đặc trưng ảnh C_i dựa trên điểm số tương đồng cosin (cosine similarity) $s_{i,j}$ giữa các vector đặc trưng ảnh f_i và f_j , trong đó mỗi nhóm đặc trưng ảnh C_i có một vector đặc trưng ảnh trung tâm;

sắp xếp các vector đặc trưng ảnh f_i trong mỗi nhóm đặc trưng ảnh C_i thành chuỗi các vector đặc trưng ảnh S_i theo thứ tự tăng dần hoặc giảm dần dựa trên điểm số tương đồng cosin $s_{i,j}$ của các vector đặc trưng ảnh so với vector đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu;

tạo ra các vector mã hoá khoảng cách cosin e_t từ các thành phần là các điểm số tương đồng cosin $s_{i,j}$, trong đó mỗi vector mã hoá khoảng cách cosin e_t tương ứng với vector đặc trưng ảnh f_t trong nhóm đặc trưng ảnh C_i và các điểm số tương đồng cosin $s_{i,j}$ tạo thành vector mã hoá khoảng cách cosin e_t này là điểm số tương đồng cosin giữa các vector đặc trưng ảnh với vector đặc trưng ảnh f_t trong cùng nhóm đặc trưng ảnh C_i đã nêu;

ghép nối vector đặc trưng ảnh f_t với vector mã hoá khoảng cách cosin e_t tương ứng để tạo thành vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_t^* tương ứng;

tạo ra chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* tương ứng với mỗi nhóm đặc trưng ảnh C_i , trong đó các thành phần của chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* này là các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_t^* tương ứng với các vector đặc trưng ảnh f_t thuộc nhóm đặc trưng ảnh C_i tương ứng, và các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_t^* được sắp xếp theo thứ tự tăng dần hoặc giảm dần dựa trên điểm số tương đồng cosin $s_{i,j}$ của các vector đặc trưng ảnh so với vector đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu;

sử dụng chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* để làm chuỗi dữ liệu đầu vào của mạng nơron mã hoá (encoder neural network), trong đó mạng nơron mã hoá này được tạo cấu hình để tạo ra biểu diễn được mã hoá tương ứng của mỗi trong số các đầu vào trong chuỗi dữ liệu đầu vào nhờ sử dụng cơ chế chú ý (attention mechanism) thể hiện được độ chú ý trong các biểu diễn được mã hoá của chuỗi dữ liệu đầu vào;

giải mã, bởi mạng nơron giải mã (decoder neural network), để tạo ra chuỗi dữ liệu đầu ra, trong đó mạng nơron giải mã này được tạo cấu hình để tiếp nhận các biểu diễn được mã hoá là dữ liệu đầu vào để giải mã thành chuỗi dữ liệu đầu ra.

Theo một phương án, bước mạng nơron mã hoá tạo ra biểu diễn được mã hoá tương ứng của mỗi trong số các đầu vào trong chuỗi dữ liệu đầu vào nhờ sử dụng cơ chế chú ý thể hiện được độ chú ý trong các biểu diễn được mã hoá của chuỗi dữ liệu đầu vào bao gồm các bước:

chiếu chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* vào ít nhất một không gian con, trong đó với mỗi không gian con, tiến hành các hoạt động:

xác định ma trận huấn luyện được thứ nhất, thứ hai và thứ ba;

chiếu chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* vào siêu không gian thứ nhất, thứ hai và thứ ba để lần lượt tạo ra các đặc trưng siêu không gian thứ nhất, thứ hai và thứ ba dựa vào ma trận huấn luyện được thứ nhất, thứ hai và thứ ba;

tính toán các điểm số độ chú ý $r_{i,j}$ giữa vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* và các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_j^* trong chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* dựa trên đặc trưng siêu không gian thứ nhất của vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* và các đặc trưng siêu không gian thứ hai của các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_j^* ; và

tạo ra đầu ra không gian con của vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* bằng cách tính tổng có trọng số của các đặc trưng siêu không gian thứ ba của các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_j^* , trong đó các trọng số được gán cho các đặc trưng siêu không gian thứ ba là các điểm số độ chú ý $r_{i,j}$ tương ứng;

biến đổi tuyến tính kết quả ghép nối các đầu ra không gian con của chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* tương ứng với mỗi không gian con để thu được đầu ra độ chú ý; và

tạo ra các biểu diễn được mã hoá dựa trên đầu ra độ chú ý.

Tốt hơn là, chuỗi dữ liệu đầu ra của mạng nơron giải mã là chuỗi nhị phân y_i có độ dài tương ứng với độ dài của chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* , khi giá trị tại vị trí thứ t của chuỗi nhị phân y_i có giá trị là 1 sẽ biểu thị vector đặc trưng ảnh thứ t có cùng nhãn với vector đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu, và khi giá trị tại vị trí thứ t của chuỗi nhị phân y_i có giá trị là 0 sẽ biểu thị vector đặc trưng ảnh thứ t không có cùng nhãn với vector đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu.

Vector mã hoá khoảng cách cosin e_t được xác định thông qua biểu thức sau:

$$e_t = \{s_{t,i}\}_{i=1}^k$$

trong đó $s_{t,i}$ là điểm số tương đồng cosin giữa các vector đặc trưng ảnh thứ i với vector đặc trưng ảnh thứ t .

Vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_t^* được xác định thông qua biểu thức sau:

$$f_t^* = \text{concat}(f_t, e_t)$$

trong đó concat là hàm ghép nối hai vector thành một vector.

Mạng nơron mã hoá và mạng nơron giải mã nêu trên được huấn luyện sử dụng hàm mục tiêu được xác định thông qua biểu thức sau:

$$\mathcal{L}_i(\hat{y}_i, y_i) = - \sum_{t=1}^k [y_i^t \times \log(\sigma(\hat{y}_i^t)) + (1 - y_i^t) \times \log(1 - \sigma(\hat{y}_i^t))]$$

trong đó σ là hàm xích ma.

Tốt hơn là, mô hình trích xuất đặc trưng ảnh nêu trên được huấn luyện sử dụng hai tập dữ liệu gồm có tập dữ liệu được gán nhãn D_L , và tập dữ liệu không được gán nhãn D_U .

Theo một khía cạnh khác, sáng chế đề xuất vật ghi có chứa chương trình máy tính bao gồm các lệnh mà khi được thực thi sẽ làm cho máy tính thực hiện phương pháp phân nhóm ảnh nêu trên.

Mô tả vắn tắt các hình vẽ

Hình 1 là lưu đồ thể hiện phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý theo một phương án ưu tiên của sáng chế;

Hình 2 là lược đồ thể hiện cách thức sắp xếp lại nhóm các đặc trưng ảnh thành chuỗi theo một phương án ưu tiên của sáng chế;

Hình 3 là ảnh chụp màn hình thể hiện đoạn chương trình thể hiện bộ sắp xếp G; và

Hình 4 là sơ đồ khối thể hiện phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý theo một phương án ưu tiên của sáng chế.

Mô tả chi tiết sáng chế

Dưới đây, các ưu điểm, hiệu quả và bản chất của sáng chế có thể được hiểu rõ hơn thông qua việc mô tả chi tiết các phương án ưu tiên thực hiện có dựa vào các hình vẽ kèm theo. Tuy nhiên, cần hiểu rằng các phương án này chỉ được mô tả với mục đích làm ví dụ giúp cho việc hiểu rõ hơn về bản chất và các ưu điểm của sáng chế, mà không giới hạn phạm vi của sáng chế theo các phương án được mô tả này.

Nhìn chung, như được mô tả dưới đây phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý hướng tới việc phân loại các ảnh hoặc dữ liệu trực quan hoá, ví dụ nhận dạng xem các ảnh có thuộc

cùng nhóm hoặc không thuộc cùng nhóm, cụ thể hơn các ảnh chụp mặt người có ở vị trí cùng tư thế chụp, các ảnh phong cảnh có phải thuộc cùng nhóm chụp cảnh hồ nước, lâu đài cổ, v.v., chẳng hạn. Tuy nhiên, cần hiểu rằng các kỹ thuật hoặc nguyên lý theo sáng chế không chỉ giới hạn ở việc phân loại ảnh hoặc dữ liệu trực quan hoá, mà có thể được ứng dụng trong các ứng dụng nhận dạng ảnh đa dạng khác nhau, chẳng hạn như ghi chú hoặc gán nhãn cho các ảnh hoặc dữ liệu trực quan hoá.

Trên Hình 1 thể hiện phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý theo một phương án ưu tiên của sáng chế.

Như được thể hiện trên hình vẽ, phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý theo phương án ưu tiên này bao gồm các bước được mô tả dưới đây:

Bước S101: trích xuất các đặc trưng ảnh từ cơ sở dữ liệu ảnh để thu được các vector đặc trưng ảnh.

Ở đây, để thuận tiện cho việc mô tả và biểu diễn, cơ sở dữ liệu ảnh được gọi là cơ sở dữ liệu ảnh X có chứa nhiều ảnh x_i , các vector đặc trưng ảnh được gọi là các vector đặc trưng ảnh f_i , mỗi vector đặc trưng ảnh f_i tương ứng với một ảnh x_i trong cơ sở dữ liệu ảnh X đã nêu.

Trong bước này, việc trích xuất các đặc trưng ảnh từ cơ sở dữ liệu ảnh X có chứa nhiều ảnh x_i có thể được thực hiện thông qua mô hình trích xuất đặc trưng ảnh M. Sử dụng mô hình trích xuất đặc trưng ảnh M, đầu vào ảnh x_i thuộc cơ sở dữ liệu ảnh (ví dụ có kích thước là $h \times w \times 3$, trong đó h biểu thị chiều cao và w biểu thị chiều rộng của ảnh) được đưa vào mô hình trích xuất đặc trưng ảnh M để trích xuất hoặc thu được các đặc trưng trực quan hoá.

Theo một ví dụ cụ thể, các đặc trưng trực quan hoá là các vector đặc trưng ảnh f_i có kích thước là $1 \times d$, trong đó d là số đặc trưng của mỗi ảnh được trích

xuất bởi mô hình trích xuất đặc trưng ảnh M . Để thuận tiện, các vector đặc trưng ảnh f_i có thể được biểu diễn là $f_i = M(x_i)$.

Nói chung, các mô hình trích xuất đặc trưng ảnh là đã biết, và được sử dụng phổ biến, ví dụ các mô hình mạng CNN, chẳng hạn. Việc mô tả chi tiết các mô hình trích xuất đặc trưng ảnh này được dự định lược bỏ để tập trung cho các nội dung được cho là quan trọng hơn của sáng chế.

Theo một phương án ưu tiên, việc sử dụng mô hình trích xuất đặc trưng ảnh M được tối đa hoá về hiệu quả nhờ sử dụng hai tập dữ liệu gồm có tập dữ liệu được gán nhãn D_L , và tập dữ liệu không được gán nhãn D_U trong quá trình huấn luyện cho mô hình trích xuất đặc trưng ảnh M . Trước tiên, mô hình trích xuất đặc trưng ảnh M được huấn luyện sử dụng tập dữ liệu được gán nhãn D_L theo cách học có giám sát thông thường. Sau đó, mô hình trích xuất đặc trưng ảnh M sau khi được huấn luyện bởi tập dữ liệu được gán nhãn D_L này được sử dụng để khởi động các mẫu huấn luyện được trích xuất trong tập dữ liệu không được gán nhãn D_U . Quá trình huấn luyện này có thể tương ứng với việc học nửa giám sát, trong đó tập dữ liệu không được gán nhãn D_U lớn hơn rất nhiều so với tập dữ liệu được gán nhãn D_L .

Bước S102: phân nhóm các vector đặc trưng ảnh được lấy mẫu từ bộ dữ liệu đặc trưng ảnh nêu trên thành các nhóm đặc trưng ảnh dựa trên điểm số tương đồng cosin.

Ở đây, để thuận tiện cho việc mô tả và biểu diễn, nhóm đặc trưng ảnh được gọi là nhóm đặc trưng ảnh C_i , điểm số tương đồng cosin được gọi là điểm số tương đồng cosin $s_{i,j}$, điểm số tương đồng cosin là điểm số tương đồng cosin $s_{i,j}$ giữa các vector đặc trưng ảnh f_i và f_j .

Điểm số tương đồng cosin giữa hai vector là đặc trưng toán học đã biết giữa hai vector. Ví dụ điểm số tương đồng cosin giữa hai đỉnh v_i, v_j của đồ thị có tính tương tự được biểu diễn bởi ma trận liên kề W , là cosin của góc giữa hai vector

gồm có một vector được biểu thị bởi hàng thứ i và hàng thứ j của ma trận liên kết W , gọi là W^i, W^j .

Điểm số tương đồng cosin được xác định như sau: $\sigma_{ij} = \frac{W^i \cdot W^j}{\|W^i\| \|W^j\|}$.

Theo một phương án ưu tiên, việc phân nhóm các vector đặc trưng ảnh nêu trên được thực hiện nhờ sử dụng giải thuật k-lân cận gần nhất (k-nearest neighbors) dựa trên điểm số tương đồng cosin, gọi là mô hình k-lân cận gần nhất K.

Mỗi nhóm đặc trưng ảnh C_i có một vector đặc trưng ảnh trung tâm f_i , và có thể được biểu diễn là $C_i = K(f_i, F, k)$, trong đó $F = M(X)$ là tập hợp các đặc trưng được trích xuất từ cơ sở dữ liệu ảnh X , và k là số lân cận gần nhất.

Các nhóm đặc trưng ảnh C_i tạo thành tập các nhóm đặc trưng ảnh C có thể được biểu diễn là $C = \{C_i\}_{i=1}^N$.

Bước S103: tạo ra chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin có chứa thông tin mã hoá khoảng cách cosin và các vector đặc trưng ảnh.

Theo một phương án ưu tiên, chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin có chứa thông tin mã hoá khoảng cách cosin và các vector đặc trưng ảnh là chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* tương ứng với mỗi nhóm đặc trưng ảnh C_i , trong đó các thành phần của chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* này là các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* tương ứng với các vector đặc trưng ảnh f_i thuộc nhóm đặc trưng ảnh C_i tương ứng, và thứ tự các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* dựa trên điểm số tương đồng cosin $s_{i,j}$ tăng dần hoặc giảm dần của các vector đặc trưng ảnh so với vector đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu.

Để tạo ra các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* , trước tiên các vector mã hoá khoảng cách cosin e_i được tạo thành từ các thành

phần là các điểm số tương đồng cosin $s_{i,j}$, trong đó mỗi vector mã hoá khoảng cách cosin e_t tương ứng với một nhóm đặc trưng ảnh C_i và các điểm số tương đồng cosin $s_{i,j}$ tạo thành vector mã hoá khoảng cách cosin e_t này tương ứng là điểm số tương đồng cosin giữa các vector đặc trưng ảnh với vector đặc trưng ảnh f_t trong cùng nhóm đặc trưng ảnh C_i đã nêu. Sau đó, ghép nối vector đặc trưng ảnh f_t với vector mã hoá khoảng cách cosin e_t tương ứng để tạo thành vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_t^* tương ứng.

Theo một phương án ưu tiên, chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* được tạo ra nhờ:

sắp xếp các vector đặc trưng ảnh f_i trong mỗi nhóm đặc trưng ảnh C_i thành chuỗi các vector đặc trưng ảnh S_i theo thứ tự tăng dần hoặc giảm dần dựa trên điểm số tương đồng cosin $s_{i,j}$ của các vector đặc trưng ảnh so với vector đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu, và

ghép nối vector đặc trưng ảnh f_t với vector mã hoá khoảng cách cosin e_t tương ứng, tại mỗi vị trí của vector đặc trưng ảnh thứ t trong chuỗi các vector đặc trưng ảnh S_i , để tạo thành vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_t^* để tạo thành chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* đã nêu.

Cần hiểu rằng, sáng chế không bị giới hạn ở phương án ưu tiên này, mà các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* có thể được tạo ra mà không cần tạo ra chuỗi các vector đặc trưng ảnh S_i , ví dụ các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_t^* có thể được tạo ra trước, và sau đó được sắp xếp thành chuỗi để tạo thành chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* , chẳng hạn.

Theo một phương án ưu tiên, như được thể hiện trên Hình 2, chuỗi các vector đặc trưng ảnh S_i được tạo ra từ nhóm đặc trưng ảnh C_i có chứa các vector đặc trưng ảnh f_i là trung tâm thông qua việc sắp xếp bởi bộ sắp xếp G , biểu diễn là $S_i = G(C_i)$.

Theo phương án ưu tiên này, các điểm số tương đồng cosin $s_{i,j}$ giữa các vector đặc trưng ảnh f_j trong nhóm với vector đặc trưng ảnh f_i là trung tâm của nhóm được tính, và các vector đặc trưng ảnh f_j sẽ được sắp xếp theo thứ tự giảm dần theo điểm số tương đồng cosin để tạo thành chuỗi.

Để cung cấp thông tin rõ ràng hơn, ảnh chụp màn hình đoạn chương trình thể hiện bộ sắp xếp G được thể hiện trên Hình 3 để tham khảo.

Theo một phương án ưu tiên, vector mã hoá khoảng cách cosin e_t được xác định thông qua biểu thức sau:

$$e_t = \{s_{t,i}\}_{i=1}^k$$

trong đó $s_{t,i}$ là điểm số tương đồng cosin giữa các vector đặc trưng ảnh thứ i với vector đặc trưng ảnh thứ t .

Vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_t^* được xác định thông qua biểu thức sau:

$$f_t^* = \text{concat}(f_t, e_t)$$

trong đó concat là hàm ghép nối hai vector thành một vector.

Bước S104: sử dụng chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin để làm chuỗi dữ liệu đầu vào của mạng nơron mã hoá.

Ở bước này, mạng nơron mã hoá được tạo cấu hình để tạo ra biểu diễn được mã hoá tương ứng của mỗi trong số các đầu vào trong chuỗi dữ liệu đầu vào nhờ sử dụng cơ chế chú ý thể hiện được độ chú ý trong các biểu diễn được mã hoá của chuỗi dữ liệu đầu vào. Để thu được độ chú ý trong các biểu diễn được mã hoá của chuỗi dữ liệu đầu vào, chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* được chiếu vào ít nhất một không gian con. Với mỗi không gian con, ma trận huấn luyện được thứ nhất, thứ hai và thứ ba được xác định. Ma trận huấn luyện được thứ nhất, thứ hai và thứ ba có thể được gọi lần lượt là ma trận truy vấn $W^Q \in \mathbb{R}^{d \times d}$, ma trận khóa $W^K \in \mathbb{R}^{d \times d}$, ma trận giá trị $W^V \in \mathbb{R}^{d \times d}$. Sau đó, chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* được

chiều vào siêu không gian thứ nhất, thứ hai và thứ ba (lần lượt được gọi là siêu không gian truy vấn, khóa và giá trị) để lần lượt tạo ra các đặc trưng siêu không gian thứ nhất, thứ hai và thứ ba (lần lượt được gọi là các đặc trưng siêu không gian truy vấn Q, khóa K và giá trị V) dựa vào ma trận huấn luyện được thứ nhất, thứ hai và thứ ba theo công thức:

$$Q = S_i * W^Q, Q \in R^{d \times d'}$$

$$K = S_i * W^K, K \in R^{d \times d'}$$

$$V = S_i * W^V, V \in R^{d \times d'}$$

Sau đó, các điểm số độ chú ý r_{ij} giữa vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* và các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_j^* trong chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* được tính toán dựa trên đặc trưng siêu không gian thứ nhất của vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* và các đặc trưng siêu không gian thứ hai của các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_j^* theo công thức:

$$r_{i,j} = \frac{e^{\frac{1}{\sqrt{d'}} Q_i K_j^T}}{\sum_{j=1}^k e^{\frac{1}{\sqrt{d'}} Q_i K_j^T}}$$

Đầu ra không gian con Z_i của vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* được tạo ra bằng cách tính tổng có trọng số của các đặc trưng siêu không gian thứ ba V_j của các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_j^* , trong đó các trọng số được gán cho các đặc trưng siêu không gian thứ ba là các điểm số độ chú ý r_{ij} tương ứng:

$$Z_i = \sum_{j=1}^k r_{i,j} \cdot V_j.$$

$$Z = Att(Q, K, V) = \{Z_i\}_{i=1}^k$$

Gọi số không gian con là m , thì chiều đặc trưng của mỗi không gian con là $d' = \frac{d}{m}$. Đầu ra không gian con của chuỗi các vector đặc trưng ảnh có thông tin mã

hoá khoảng cách cosin S_i^* tương ứng với mỗi không gian con $Z_{s,i}$ được tính như sau:

$$Z_{s,i} = Att(Q_{s,i}, K_{s,i}, V_{s,i}), 1 \leq i \leq m$$

Kết quả ghép nối các đầu ra không gian con của chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* tương ứng với mỗi không gian con được biến đổi tuyến tính để thu được đầu ra độ chú ý Z_M :

$$Z_M = concat(Z_{s,1}, \dots, Z_{s,m}) \cdot W^M$$

Trong đó W^M là ma trận trọng số bổ sung.

Mạng nơron mã hoá có thể tạo ra các biểu diễn được mã hoá dựa trên đầu ra độ chú ý Z_M . Theo một phương án của sáng chế, mạng nơron mã hoá bao gồm mạng tiếp thuận theo điểm (Point-wise Feed Forward Network - FFN) để tiếp nhận đầu ra độ chú ý Z_M .

Bước S105: giải mã, bởi mạng nơron giải mã, để tạo ra chuỗi dữ liệu đầu ra, trong đó dữ liệu đầu vào của mạng nơron giải mã chính là dữ liệu đầu ra của mạng nơron mã hoá.

Ở đây, dữ liệu đầu vào của mạng nơron giải mã hay dữ liệu đầu ra của mạng nơron mã hoá là các biểu diễn được mã hoá.

Theo một phương án ưu tiên, chuỗi dữ liệu đầu ra của mạng nơron giải mã là chuỗi nhị phân y_i có độ dài tương ứng với độ dài của chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* , khi giá trị tại vị trí thứ t của chuỗi nhị phân y_i có giá trị là 1 sẽ biểu thị vector đặc trưng ảnh thứ t có cùng nhãn với vector đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu, và khi giá trị tại vị trí thứ t của chuỗi nhị phân y_i có giá trị là 0 sẽ biểu thị vector đặc trưng ảnh thứ t không có cùng nhãn với vector đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu.

Mạng nơron mã hoá và mạng nơron giải mã nêu trên có thể được huấn luyện sử dụng hàm mục tiêu được xác định thông qua biểu thức sau:

$$\mathcal{L}_i(\hat{y}_i, y_i) = - \sum_{t=1}^k [y_i^t \times \log(\sigma(\hat{y}_i^t)) + (1 - y_i^t) \times \log(1 - \sigma(\hat{y}_i^t))]$$

trong đó σ là hàm xích ma.

Nói chung, các mạng nơron mã hoá và giải mã nêu trên là đã biết và có thể được ứng dụng tương tự như các mạng nơron mã hoá và giải mã sử dụng trong các mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý. Một ví dụ về mô hình dạng này được đề xuất trong bài báo có tiêu đề là “Attention is all you need”, của tác giả Ashish Vaswani và các cộng sự. Một ví dụ khác về mô hình dạng này được đề xuất trong các tài liệu patent Mỹ số US 10452978 B2, US 10719764 B2, US 10839259 B2, US 10956819 B2. Toàn bộ nội dung của các tài liệu này được dự định đưa vào đây để viện dẫn và có thể được kết hợp vào giải pháp được đề xuất theo sáng chế theo cách thức đã biết bất kỳ.

Các đặc trưng và nguyên lý hoạt động của các mạng nơron mã hoá và giải mã theo sáng chế là hoàn toàn tương tự với các mạng nơron mã hoá và giải mã được đề xuất hoặc sử dụng trong bài báo và các tài liệu patent đã nêu. Do đó các mô tả chi tiết về các mạng nơron mã hoá và giải mã này được dự định lược bỏ để tập trung cho các nội dung được cho là quan trọng hơn của sáng chế.

Như được thể hiện trên Hình 4, phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý được minh hoạ thông qua các bước từ S201 đến S205, sẽ được mô tả chi tiết hơn dưới đây.

Ở bước S201, nhóm đặc trưng ảnh C_i (bao gồm các vector đặc trưng ảnh f_i có kích thước là $1 \times d$) được sắp xếp lại thành chuỗi các vector đặc trưng ảnh S_i .

Tiếp theo, ở bước S202, tại mỗi vị trí trong chuỗi các vector đặc trưng ảnh S_i , vector mã hoá khoảng cách cosin e_i (có kích thước là $1 \times k$) sẽ được ghép nối với vector đặc trưng ảnh f_i tương ứng để tạo thành chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* , trong đó mỗi thành phần của chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* là một vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* có kích thước là $1 \times$

($k+d$), mà là kết quả ghép nối của vector mã hoá khoảng cách cosin e_i và vector đặc trưng ảnh f_i tương ứng.

Ở bước S203, chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* được sử dụng để làm chuỗi dữ liệu đầu vào của mạng nơron mã hoá để tạo ra biểu diễn được mã hoá tương ứng của mỗi trong số các đầu vào trong chuỗi dữ liệu đầu vào nhờ sử dụng cơ chế chú ý (self attention) thể hiện được độ chú ý trong các biểu diễn được mã hoá của chuỗi dữ liệu đầu vào tương ứng với vị trí trong chuỗi.

Tiếp theo, ở bước S204, các biểu diễn được mã hoá được tạo ra ở bước S203 nêu trên được sử dụng làm dữ liệu đầu vào của mạng nơron giải mã, để tạo ra chuỗi dữ liệu đầu ra y_i là chuỗi nhị phân.

Cuối cùng, ở bước S205, chuỗi dữ liệu đầu ra y_i được kết hợp để tạo thành nhóm đặc trưng ảnh đầu ra được nhận dạng.

Đối với phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý được mô tả trên đây, mô hình nhận dạng ảnh mà sử dụng phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý được đề xuất theo sáng chế có thể được hiểu là nhóm các mô hình để thực hiện việc nhận dạng ảnh (ví dụ phân loại ảnh) bao gồm mô hình trích xuất đặc trưng ảnh, mạng nơron mã hoá, mạng nơron giải mã, và các thành phần liên quan để thực hiện chức năng nhận dạng ảnh, ví dụ bộ sắp xếp G để thực hiện việc sắp xếp, bộ mã hoá khoảng cách cosin để thực hiện mã hoá khoảng cách cosin, bộ nhớ và bộ tính toán, chẳng hạn.

Trên đây, sáng chế đã được mô tả chi tiết theo các phương án ưu tiên thực hiện. Hiển nhiên là, người có hiểu biết trung bình về lĩnh vực kỹ thuật này có thể dễ dàng tạo ra các thay đổi và cải biến dựa vào các phương án được mô tả. Do đó, các thay đổi và cải biến này được coi là không nằm ngoài phạm vi của sáng chế như được xác định trong yêu cầu bảo hộ kèm theo.

YÊU CẦU BẢO HỘ

1. Phương pháp phân nhóm bằng cách sử dụng mô hình bộ mã hóa-bộ giải mã dựa trên cơ chế chú ý, phương pháp này bao gồm các bước:

trích xuất các đặc trưng ảnh từ cơ sở dữ liệu ảnh X có chứa nhiều ảnh x_i , bởi mô hình trích xuất đặc trưng ảnh, để thu được bộ dữ liệu đặc trưng ảnh gồm có các vectơ đặc trưng ảnh f_i , trong đó mỗi vectơ đặc trưng ảnh f_i tương ứng với một ảnh x_i trong cơ sở dữ liệu ảnh X đã nêu;

phân nhóm các vectơ đặc trưng ảnh được lấy mẫu từ bộ dữ liệu đặc trưng ảnh nêu trên thành các nhóm đặc trưng ảnh C_i dựa trên điểm số tương đồng cosin $s_{i,j}$ giữa các vectơ đặc trưng ảnh f_i và f_j , trong đó mỗi nhóm đặc trưng ảnh C_i có một vectơ đặc trưng ảnh trung tâm;

sắp xếp các vectơ đặc trưng ảnh f_i trong mỗi nhóm đặc trưng ảnh C_i thành chuỗi các vectơ đặc trưng ảnh S_i theo thứ tự tăng dần hoặc giảm dần dựa trên điểm số tương đồng cosin $s_{i,j}$ của các vectơ đặc trưng ảnh so với vectơ đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu;

tạo ra các vectơ mã hoá khoảng cách cosin e_i từ các thành phần là các điểm số tương đồng cosin $s_{i,j}$, trong đó mỗi vectơ mã hoá khoảng cách cosin e_i tương ứng với vectơ đặc trưng ảnh f_i trong nhóm đặc trưng ảnh C_i và các điểm số tương đồng cosin $s_{i,j}$ tạo thành vectơ mã hoá khoảng cách cosin e_i này là điểm số tương đồng cosin giữa các vectơ đặc trưng ảnh với vectơ đặc trưng ảnh f_i trong cùng nhóm đặc trưng ảnh C_i đã nêu;

ghép nối vectơ đặc trưng ảnh f_i với vectơ mã hoá khoảng cách cosin e_i tương ứng để tạo thành vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* tương ứng;

tạo ra chuỗi các vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* tương ứng với mỗi nhóm đặc trưng ảnh C_i , trong đó các thành phần của chuỗi các vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* này là các vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* tương ứng với các

vector đặc trưng ảnh f_i thuộc nhóm đặc trưng ảnh C_i tương ứng, và các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* được sắp xếp theo thứ tự tăng dần hoặc giảm dần dựa trên điểm số tương đồng cosin $s_{i,j}$ của các vector đặc trưng ảnh so với vector đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu;

sử dụng chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* để làm chuỗi dữ liệu đầu vào của mạng nơron mã hoá, trong đó mạng nơron mã hoá này được tạo cấu hình để tạo ra biểu diễn được mã hoá tương ứng của mỗi trong số các đầu vào trong chuỗi dữ liệu đầu vào nhờ sử dụng cơ chế chú ý thể hiện được độ chú ý trong các biểu diễn được mã hoá của chuỗi dữ liệu đầu vào;

giải mã, bởi mạng nơron giải mã, để tạo ra chuỗi dữ liệu đầu ra, trong đó mạng nơron giải mã này được tạo cấu hình để tiếp nhận các biểu diễn được mã hoá là dữ liệu đầu vào để giải mã thành chuỗi dữ liệu đầu ra.

2. Phương pháp theo điểm 1, trong đó bước mạng nơron mã hoá tạo ra biểu diễn được mã hoá tương ứng của mỗi trong số các đầu vào trong chuỗi dữ liệu đầu vào nhờ sử dụng cơ chế chú ý thể hiện được độ chú ý trong các biểu diễn được mã hoá của chuỗi dữ liệu đầu vào bao gồm các bước:

chiếu chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* vào ít nhất một không gian con, trong đó với mỗi không gian con, tiến hành các hoạt động:

xác định ma trận huấn luyện được thứ nhất, thứ hai và thứ ba;

chiếu chuỗi các vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* vào siêu không gian thứ nhất, thứ hai và thứ ba để lần lượt tạo ra các đặc trưng siêu không gian thứ nhất, thứ hai và thứ ba dựa vào ma trận huấn luyện được thứ nhất, thứ hai và thứ ba;

tính toán các điểm số độ chú ý $r_{i,j}$ giữa vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* và các vector đặc trưng ảnh có thông

tin mã hoá khoảng cách cosin f_j^* trong chuỗi các vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* dựa trên đặc trưng siêu không gian thứ nhất của vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* và các đặc trưng siêu không gian thứ hai của các vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_j^* ; và

tạo ra đầu ra không gian con của vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_i^* bằng cách tính tổng có trọng số của các đặc trưng siêu không gian thứ ba của các vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_j^* , trong đó các trọng số được gán cho các đặc trưng siêu không gian thứ ba là các điểm số độ chú ý r_{ij} tương ứng;

biến đổi tuyến tính kết quả ghép nối các đầu ra không gian con của chuỗi các vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* tương ứng với mỗi không gian con để thu được đầu ra độ chú ý; và

tạo ra các biểu diễn được mã hoá dựa trên đầu ra độ chú ý.

3. Phương pháp theo điểm 1, trong đó chuỗi dữ liệu đầu ra của mạng nơron giải mã là chuỗi nhị phân y_i có độ dài tương ứng với độ dài của chuỗi các vectơ đặc trưng ảnh có thông tin mã hoá khoảng cách cosin S_i^* , khi giá trị tại vị trí thứ t của chuỗi nhị phân y_i có giá trị là 1 sẽ biểu thị vectơ đặc trưng ảnh thứ t có cùng nhãn với vectơ đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu, và khi giá trị tại vị trí thứ t của chuỗi nhị phân y_i có giá trị là 0 sẽ biểu thị vectơ đặc trưng ảnh thứ t không có cùng nhãn với vectơ đặc trưng ảnh trung tâm trong cùng nhóm đặc trưng ảnh C_i đã nêu.

4. Phương pháp theo điểm 1, trong đó:

vectơ mã hoá khoảng cách cosin e_t được xác định thông qua biểu thức sau:

$$e_t = \{s_{t,i}\}_{i=1}^k$$

trong đó $s_{t,i}$ là điểm số tương đồng cosin giữa các vector đặc trưng ảnh thứ i với vector đặc trưng ảnh thứ t ,

vector đặc trưng ảnh có thông tin mã hoá khoảng cách cosin f_t^* được xác định thông qua biểu thức sau:

$$f_t^* = \text{concat}(f_t, e_t)$$

trong đó concat là hàm ghép nối hai vector thành một vector.

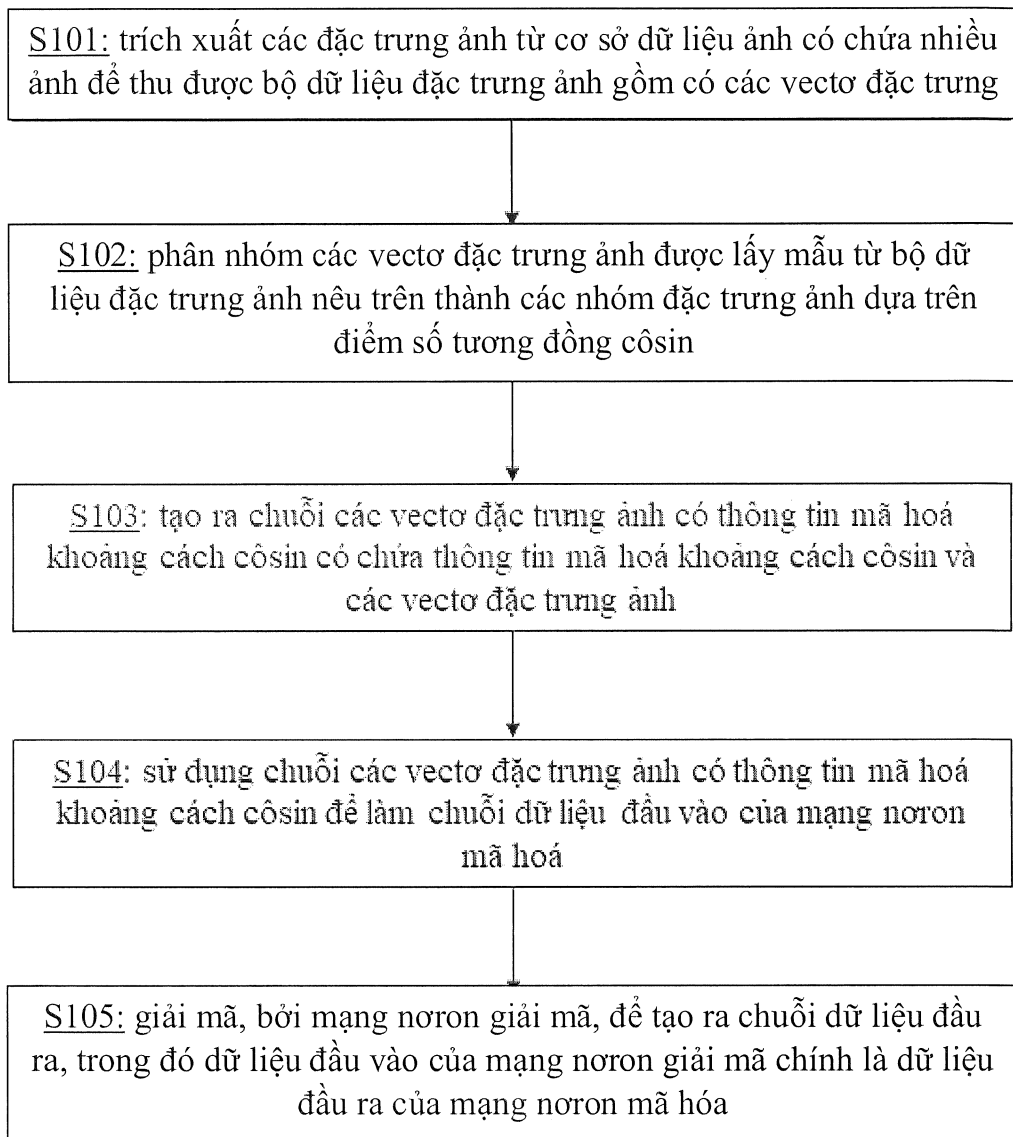
5. Phương pháp theo điểm 1, trong đó mạng nơron mã hoá và mạng nơron giải mã nêu trên được huấn luyện sử dụng hàm mục tiêu được xác định thông qua biểu thức sau:

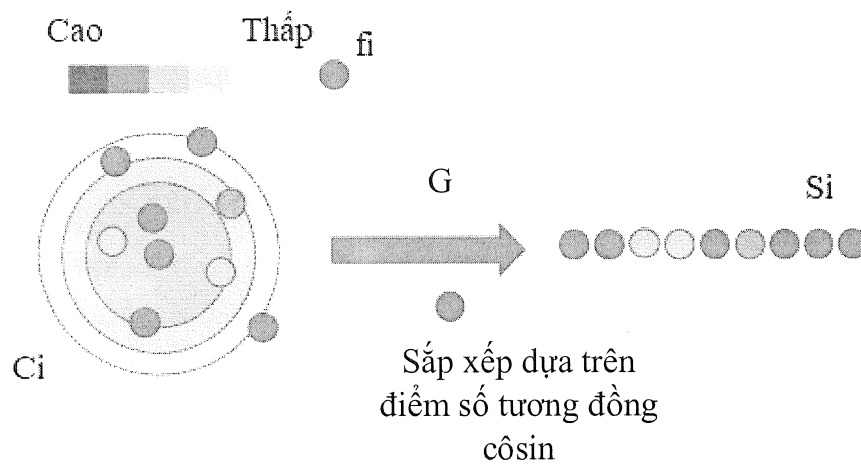
$$\mathcal{L}_i(\hat{y}_i, y_i) = - \sum_{t=1}^k [y_i^t \times \log(\sigma(\hat{y}_i^t)) + (1 - y_i^t) \times \log(1 - \sigma(\hat{y}_i^t))]$$

trong đó σ là hàm xích ma.

6. Phương pháp theo điểm 1, trong đó mô hình trích xuất đặc trưng ảnh nêu trên được huấn luyện sử dụng hai tập dữ liệu gồm có tập dữ liệu được gán nhãn D_L , và tập dữ liệu không được gán nhãn D_U .

7. Vật ghi có chứa chương trình máy tính bao gồm các lệnh mà khi được thực thi sẽ làm cho máy tính thực hiện phương pháp theo điểm 1.

**Hình 1**



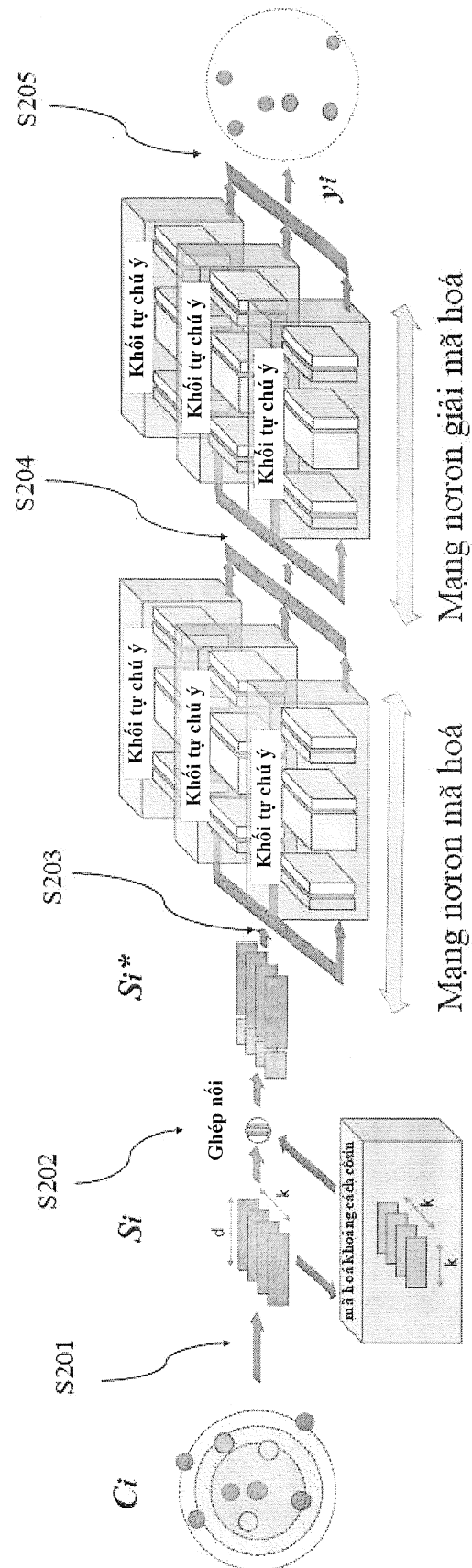
Hình 2

```

1: Input: Input image  $x_i$ , Visual classifier  $\mathcal{M}$ , data set  $\mathbf{X}$ ,
   length of visual sequence  $k$ , and step size  $\Delta_r$ 
2: Output: Visual Sequence  $\mathbf{S}_i$ 
3:  $\mathbf{S}_i = \emptyset; \mathbf{f}_i = \mathcal{M}(x_i)$ 
4:  $radius = 0$ 
5: while  $|\mathbf{S}_i| \leq k$  do
6:    $S = \{x_j \in \mathbf{X} | radius \leq s_{ij} \leq radius + \Delta_r\}$  where
      $s_{i,j} = \text{dist}(\mathbf{f}_i, \mathcal{M}(x_j))$ 
7:   if  $S = \emptyset$  then
8:     break
9:   end if
10:   $ind \leftarrow (\text{argsort}(S))_v := |\{u \in \{1, \dots, |S|\} | x_u \in S,$ 
      $s_{i,u} \leq s_{i,v}\}|$ 
11:  for all  $u$  in  $ind$  do
12:     $\mathbf{S}_i = \mathbf{S}_i + [\mathcal{M}(x_u)], x_u \in S$ 
13:  end for
14:   $radius = radius + \Delta_r$ 
15: end while
16: return  $\mathbf{S}_i$ 

```

Hình 3



Hình 4