



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

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

(11)



1-0024129

(51)⁷ G06F 17/30

(13) B

(21) 1-2014-02947

(22) 06/02/2013

(86) PCT/US2013/024881 06/02/2013

(87) WO2013/119626 15/08/2013

(30) 61/595,502 06/02/2012 US

(45) 25/06/2020 387

(43) 25/12/2014 321A

(73) Interdigital Madison Patent Holdings (FR)

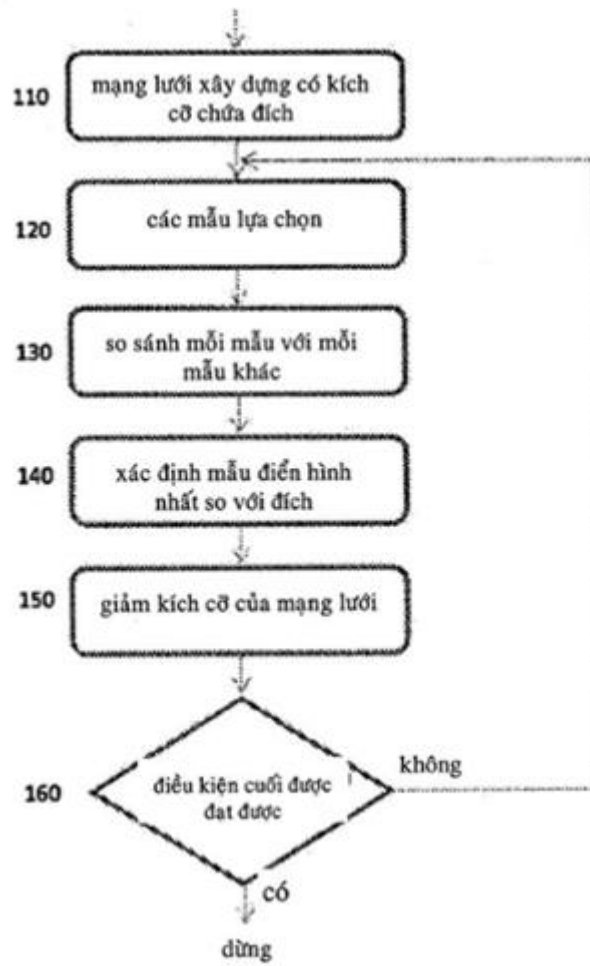
3 rue du Colonel Moll, 75017 Paris, France.

(72) MASSOULIE, Laurent (FR); IOANNIDIS, Efstratios (GR).

(74) Công ty Luật TNHH T&G (TGVN)

(54) PHƯƠNG PHÁP VÀ THIẾT BỊ TÌM KIẾM NỘI DUNG TRONG CƠ SỞ DỮ LIỆU

(57) Sáng chế đề cập đến phương pháp và thiết bị tìm kiếm nội dung trong cơ sở dữ liệu. Trong việc tìm kiếm nội dung tương tác thông qua phép so sánh, việc tìm kiếm đối với đối tượng đích trong cơ sở dữ liệu được thực hiện bằng cách tìm đối tượng tương tự nhất so với đích từ danh sách ít các đối tượng. Danh sách đối tượng mới sau đó được trình bày dựa trên sự lựa chọn trước đây. Quá trình này được lặp lại cho đến khi đích được bao gồm trong danh sách được trình bày, trong đó điểm kết thúc tìm kiếm. Giải pháp cho vấn đề tìm kiếm nội dung tương tác được đề xuất theo kịch bản yêu cầu không đồng nhất, trong đó các đối tượng đích được chọn từ xác suất phân bố không đồng nhất. Được xét đến rằng các đối tượng được nhúng trong không gian ma trận kép trong đó quan sát được hoàn toàn đối với thuật toán tìm kiếm. Dựa trên những giả định này, phép so sánh hiệu quả dựa trên phương pháp tìm kiếm được đề xuất mà chi phí của nó trong mục của số câu hỏi có thể được giới hạn bằng hằng số kép của phép lồng c , và entropi phân bố yêu cầu, H . Cụ thể hơn, sáng chế thể hiện rằng chi phí tìm kiếm trung bình chia tỷ lệ $C_F = O(c^5 H)$ mà cải thiện giới hạn đã được biết đến trước đây và là bậc tối ưu đối với hằng số c .



Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến việc tìm kiếm nội dung tương tác thông qua phép so sánh.

Tình trạng kỹ thuật của sáng chế

Việc tìm kiếm nội dung thông qua phép so sánh là trường hợp đặc biệt của tìm kiếm lân cận gần nhất (nearest neighbor search - NNS). Các nguyên lý được mô tả ở đây mở rộng công việc trước đó bằng cách xem xét chương trình NNS dùng cho đối tượng được lồng trong không gian ma trận. Cũng giả định rằng việc lồng có kích thước thực nhỏ- một giả định mà được chứng minh bởi nhiều nghiên cứu thực tế. Công việc trước đó xem xét sự điều hành mạng lưới, cấu trúc dữ liệu tiên định dùng cho sự hỗ trợ NNS trong không gian ma trận kép. Kỹ thuật tương tự cũng đã được xem xét dùng cho đối tượng được lồng trong không gian thỏa mãn đặc tính đóng gói hình cầu đã biết, trong khi công việc khác dựa trên ma trận bị thu hẹp tăng. Tất cả giả định nêu trên có sự kết nối đến hằng số kép được xem xét ở đây. Trong tất cả công việc trước đây, nhu cầu trên các đối tượng đích được giả thiết là đồng nhất.

NNS với việc truy cập vào oracle so sánh đã được nghiên cứu trước đây. Ưu điểm đáng kể của các nghiên cứu trước đây mà giả định rằng các đối tượng là được nhúng trước vào không gian ma trận được loại bỏ; thay vì yêu cầu tính tương tự giữa các đối tượng được thỏa mãn bằng khoảng ma trận, các công việc trước đây chỉ xem xét rằng hai đối tượng bất kỳ có thể được phân loại theo tính tương tự của chúng với đích bất kỳ bằng oracle so sánh. Tuy nhiên, các công việc này cũng xét đến yêu cầu đồng nhất, nên các nguyên lý ở đây là phần mở rộng tìm kiếm bằng phép so sánh với tính không đồng nhất. Về khía cạnh này, sự phân bố yêu cầu không đồng nhất là điểm khởi đầu cho các nguyên lý ở đây. Theo các giả định rằng tồn tại không gian ma trận và thuật toán tìm kiếm nhận biết được nó, các nguyên lý theo sáng chế làm gia tăng chi phí tìm kiếm trung bình. Vấn đề chính của các công việc trước đây là ở chỗ sự tiếp cận của chúng là không nhớ, tức là, nó không sử dụng các so sánh trước, trong khi đó sáng chế giải quyết vấn đề này bằng cách sử dụng cấu trúc dữ liệu mạng lưới.

Phép so sánh cặp đôi giữa các ảnh đã được đề xuất trước đây. Sau đó, nó được

mở rộng đến ngữ cảnh để tìm kiếm nội dung. Sử dụng oracle so sánh không chỉ giới hạn ở tra cứu/tìm kiếm nội dung. Thang biểu phân loại của cá nhân có xu hướng biến động nhiều. Ngoài ra, các thang biểu phân loại có thể khác nhau giữa mọi người. Vì các lý do này, tự nhiên hơn để sử dụng phép so sánh cặp đôi làm cơ sở cho các hệ thống khuyến cáo. Các ưu điểm của phương pháp này và các thách thức làm thế nào để hệ thống vận hành đã được mô tả.

Bản chất kỹ thuật của sáng chế

Các hạn chế này và những hạn chế khác và nhược điểm của kỹ thuật hiện nay được giải quyết theo các nguyên lý mà đề xuất phương pháp tìm kiếm nội dung tương tác thông qua phép so sánh.

Theo một khía cạnh, sáng chế đề xuất phương pháp tìm kiếm nội dung trong cơ sở dữ liệu. Phương pháp này bao gồm các bước xây dựng mạng lưới có kích cỡ chứa đích, chọn nhiều mẫu, so sánh mỗi mẫu với mọi mẫu khác, và xác định mẫu gần nhất với đích. Phương pháp còn bao gồm các bước làm giảm kích cỡ của mạng lưới đến kích cỡ nhỏ hơn mà chứa đích. Phương pháp còn bao gồm bước lặp lại việc chọn, so sánh, xác định, và các bước làm giảm cho đến khi kích cỡ của mạng lưới đủ nhỏ để định vị đích.

Theo khía cạnh khác, sáng chế đề xuất thiết bị tìm kiếm nội dung trong cơ sở dữ liệu. Thiết bị này bao gồm máy tính mà thực hiện các bước bao gồm phương pháp được mô tả ở đây. Máy tính có thể bao gồm mạch xây dựng mạng lưới có kích cỡ chứa đích. Máy tính cũng có thể bao gồm mạch chọn các mẫu, và mạch so sánh này vận hành trên các mẫu.

Máy tính cũng bao gồm mạch xác định mà tìm mẫu gần nhất với đích và mạch làm giảm kích cỡ của mạng lưới thành kích cỡ nhỏ hơn mà chứa đích. Máy tính cũng bao gồm mạch điều khiển dẫn đến mạch xây dựng mạng lưới, mạch chọn các mẫu, mạch so sánh, mạch xác định, và mạch làm giảm kích cỡ của mạng lưới để lặp lại sự vận hành của chúng nếu điều kiện cuối đã không đạt được.

Các khía cạnh này và các khía cạnh khác, các đặc điểm và ưu điểm của các sáng chế sẽ trở nên rõ ràng dựa vào phần mô tả chi tiết sau đây của các phương án

được nêu làm ví dụ, mà được thể hiện kết hợp với các hình vẽ kèm theo.

Mô tả chi tiết hình vẽ

Fig.1 thể hiện một phương án của phương pháp thực hiện tìm kiếm nội dung theo sáng chế.

Fig.2 thể hiện thiết bị thực hiện tìm kiếm nội dung theo sáng chế.

Fig.3 thể hiện phương án minh họa các thành phần bao gồm thiết bị của Fig.2.

Mô tả chi tiết sáng chế

Sáng chế đề cập đến phương pháp và thiết bị tìm kiếm nội dung tương tác thông qua phép so sánh. Phương pháp này được giới hạn "tương tác" bởi vì có các bậc được lặp lại tương tác với kết quả của giai đoạn nêu trước. Phương pháp điều hành thông qua cơ sở dữ liệu của các đối tượng (ví dụ, các đối tượng, hình ảnh, phim ảnh, bài viết, v.v) có một số đặc tính có thể đo bằng cách sử dụng phép so sánh. Cụ thể, phương pháp xác định, từ hai đối tượng cùng một lúc, đối tượng gần nhất với đích (ví dụ, hình ảnh hoặc phim ảnh hoặc bài viết, v.v). Gần với đích, tức là khoảng cách, có thể được đo bằng một số cách, như hiệu tuyệt đối, tổng của các hiệu tuyệt đối, v.v. Dựa trên sự lựa chọn, phương pháp chọn cặp mới của các đối tượng, và quá trình được lặp lại ở các giai đoạn tương tự cho đến khi cặp đối tượng chứa đích được mong muốn. Trong mỗi giai đoạn, danh sách nhỏ của các đối tượng được trình bày dùng để so sánh. Một đối tượng trong số danh sách được chọn làm đối tượng gần nhất với đích; danh sách đối tượng mới sau đó được trình bày dựa trên sự lựa chọn trước đây. Quy trình này tiếp tục cho đến khi đích được bao gồm trong danh sách được trình bày, khi đó điểm đích tìm thấy và kết thúc tìm kiếm điểm này.

Theo phương án thay thế, quy trình có thể được lặp lại cho một số phép lặp, hoặc cho đến khi đối tượng được chọn trong khoảng cách giới hạn của đích được mong muốn. Ngoài ra, phương án thay thế có thể được sử dụng để định vị đích trong mạng lưới sau khi mạng lưới được thu hẹp để tất cả các đối tượng của nó trong phạm vi khoảng cách giới hạn của đích.

Phương pháp này yêu cầu:

1) Nhúng ma trận của các đối tượng, tức là, sự biểu diễn các đối tượng trong không gian ma trận mô tả các đặc điểm chúng. Ví dụ, điều này có thể là giá trị điểm ảnh của đối tượng hình ảnh. Khoảng cách trong không gian ma trận này thỏa mãn được như thế nào "tương tự" hoặc "gần" các đối tượng là.

2) Kết quả của phép so sánh ở mỗi giai đoạn cho thấy các đối tượng gần nhất với đích

Ở mỗi giai đoạn, phương pháp tạo ra cặp mới của đối tượng để đề xuất làm các khả năng đích.

Các đối tượng được đề xuất có thể được sử dụng trong phiên bản tiếp theo phương pháp, hoặc nếu chúng chứa đích hoặc gần giống với đích được mong muốn, việc tìm kiếm có thể được dừng.

Trong các thuật ngữ đơn giản, phương pháp xây dựng cây tổ chức các đối tượng trong hệ thứ cấp. Các nút trong cây này mà nằm trong cùng mức "che phủ" vùng có kích thước tương ứng của không gian ma trận trong đó các đối tượng được biểu diễn. Phương pháp tiến hành bằng cách đề xuất các cặp của các đối tượng trong lớp thứ nhất của cây: nhận dạng đối tượng nào trong các đối tượng ở mức của cây gần nhất với đích giới hạn sự lựa chọn của các đối tượng mà nằm dưới đối tượng này trong hệ thứ cấp. Sau đó, phương pháp tiến hành một cách đệ quy bằng cách đề xuất các cặp đối tượng trong số con của nút này.

Phương pháp được đề xuất có các thuộc tính sau đây:

1) Phương pháp này tìm đối tượng ngoài vùng tìm kiếm ngoài nhanh, trong vài cặp được đề xuất.

2) Đảm bảo rằng nó tạo ra công việc với nhu cầu không đồng nhất: nghĩa là, nó hoạt động ngay cả khi một số các đối tượng giống hơn được chọn hơn các đối tượng khác.

So với công việc trước trong lĩnh vực này, phương pháp hiện nay có sự đảm bảo tốt hơn, mà phương pháp này tìm kiếm các đối tượng nhanh hơn. Phương pháp hiện nay yêu cầu hiểu biết về không gian ma trận tổng thể, trong khi phương pháp trước đây yêu cầu hiểu biết về bậc của khoảng cách giữa các đối tượng và đích, mặc

dù không phải là giá trị bằng số chính xác của những khoảng cách này. Phương pháp hiện nay không yêu cầu hiểu biết về khả năng đối tượng có thể được chọn, trong khi các phương pháp trước đây thực hiện. Phương pháp hiện nay cũng thực hiện thuật toán khác nhau về cơ bản so với công việc trước trong lĩnh vực này.

Kiểu điều hướng tương tác này, cũng đã biết đến như cách tìm kiếm thăm dò, có nhiều ứng dụng độ bền lâu. Một ví dụ về sự vận hành thông qua cơ sở dữ liệu về hình ảnh của con người được chụp trong môi trường không kiểm soát được, như cơ sở dữ liệu Flickr hoặc Picasa. Phương pháp tự động hóa có thể bị lỗi để khai các đặc điểm quan trọng từ hình ảnh này. Ngoài ra, trong nhiều trường hợp thực tế, các hình ảnh mà trình bày tập mô tả mức độ thấp tương tự (như các đặc điểm chọn lọc) có thể có nội dung ngữ nghĩa rất khác nhau và sự mô tả mức độ cao, và do đó được nhận thấy tính khác biệt bởi người sử dụng.

Ngược lại, sự tìm kiếm nhân lực đối với cá nhân cụ thể có thể chọn một cách dễ dàng từ danh sách hình ảnh gần giống với người mà cần hướng đến. Về mặt hình thức, hành vi của người sử dụng có thể được mô hình hóa bằng cách gọi tên là oracle so sánh. Cụ thể, giả định rằng cơ sở dữ liệu của hình ảnh được trình bày lại bằng tập hợp N được cung cấp với khoảng ma trận d . Ma trận này thỏa mãn "khoảng cách" hoặc "tính không tương tự" giữa các ảnh của người khác nhau. Oracle/ người có đích cụ thể $t \in N$ trong ý nghĩ, và có thể trả lời các câu hỏi của loại sau đây: "Giữa hai đối tượng x và y trong N , trong đó đối tượng điển hình nhất so với t dưới ma trận d ?"

Do đó, mục đích của việc tìm kiếm nội dung tương tác thông qua phép so sánh là tìm dãy số của các cặp được đề xuất về các đối tượng đến oracle/con người mà dẫn đến đối tượng đích với một số thách thức như có thể thực hiện được.

Sáng chế được mô tả ở đây xem xét vấn đề theo kịch bản yêu cầu không đồng nhất, trong đó đối tượng đích $t \in N$ được lấy mẫu từ sự phân bố xác suất μ . Trong trường hợp này, việc tìm kiếm nội dung tương tác thông qua phép so sánh có mối quan hệ chặt chẽ với chương trình "hai mươi câu hỏi trò chơi" cổ điển. Cụ thể, oracle thành viên là oracle mà có thể trả lời các câu hỏi dạng sau đây: "Đưa ra tập con $A \subset N$, làm t thuộc về A ?"

Điều này cũng đã biết rằng để tìm đích t đều cần nhập ít nhất các câu hỏi $H(\mu)$, trên trung bình, đến oracle thành viên, trong đó $H(\mu)$ là entropi của μ . Ngoài ra, tồn tại thuật toán (lập trình Huffman) mà tìm ra đích duy nhất các câu hỏi $H(\mu) + 1$ trên trung bình.

Việc tìm kiếm nội dung thông qua phép so sánh khởi hành từ thiết lập nêu trên ở giả định rằng cơ sở dữ liệu N được cấp với ma trận d . Oracle thành viên mạnh hơn so với oracle so sánh như, nếu khoảng cách ma trận d đã biết, so sánh các câu hỏi có thể được mô phỏng thông qua các câu hỏi thành viên. Ngược lại, oracle thành viên khó thực hiện trong thực tế: trừ khi A có thể được thực hiện một cách ngắn gọn, người sử dụng sẽ trả lời câu hỏi thành viên trong thời gian dài trong $|A|$. Điều này trái ngược với oracle so sánh, trong đó các câu hỏi có thể được đưa ra ở thời gian không đổi. Tóm lại, nghiên cứu của chúng tôi là tìm kiếm thông qua phép so sánh tìm kiếm đặc tính tương tự giới hạn đến các thiết lập cổ điển (a) dùng cho oracle mà dễ dàng hơn để thực thi và (b) theo điều giả định bổ sung về cấu trúc của cơ sở dữ liệu (có nghĩa là, nó được cấp với khoảng ma trận).

Bằng trực giác, thực hiện việc tìm kiếm dùng cho đối tượng thông qua phép so sánh sẽ phụ thuộc không chỉ dựa trên entropi về sự phân bố đích, nhưng cũng dựa trên cấu trúc của tập đích N , như được mô tả bằng ma trận d . Cụ thể, nó đã được thiết lập mà các câu hỏi $\Omega(cH(\mu))$ là cần thiết, trong sự kỳ vọng, để định vị đích bằng cách sử dụng oracle so sánh, trong đó c được gọi tên là hằng số kép của ma trận d . Ngoài ra, sơ đồ tồn tại mà định vị đích trong các câu hỏi $O(c^3 H \log(1/\mu^*))$, trong kỳ vọng, trong đó $\mu^* = \min_{x \in N} \mu(x)$. Theo các nguyên lý ở đây, sự cải thiện trên giới hạn trước thu được bằng cách đề xuất thuật toán mà định vị đích bằng các câu hỏi $O(c^5 H(\mu))$, trong sự kỳ vọng.

Định nghĩa và ký hiệu

Xem xét tập các đối tượng N , trong đó $|N| = n$. Chúng tôi giả định rằng tồn tại không gian ma trận (M, d) , trong đó $d(x, y)$ là khoảng cách giữa $x, y \in M$, sao cho các đối tượng trong N được nhúng trong (M, d) : tức là, tồn tại ánh xạ một một từ N đến tập con của M .

Các đối tượng trong N có thể biểu diễn, ví dụ, hình ảnh trong cơ sở dữ liệu. Phép nhúng ma trận có thể được thông qua làm ánh xạ của cơ sở dữ liệu nhập mục vào tập đặc điểm (ví dụ, tuổi của người được mô tả, tóc của cô ấy và màu mắt, v.v.). Khoảng cách giữa hai đối tượng sau đó sẽ thu nhập như thế nào "tương tự" hai đối tượng là các đặc điểm w.r.t. Trong các đối tượng sau đây, một số ký hiệu sẽ được viết là $N \subset M$, lưu trữ trong ý nghĩ mà có thể khác nhau giữa các đối tượng vật lý (hình ảnh) và phép lồng của chúng (thuộc tính này mô tả chúng).

A. Oracle so sánh

Oracle so sánh là oracle mà, đưa ra hai đối tượng x, y và đích t , trả lời đối tượng điển hình nhất so với t . Cụ thể,

$$\text{Oracle}(x, y, t) = \begin{cases} x & \text{if } d(x, t) < d(y, t), \\ y & \text{if } d(x, t) > d(y, t), \\ x \text{ hoặc } y & \text{if } d(x, t) = d(y, t), \end{cases} \quad (1)$$

Quan sát rằng nếu $x = \text{Oracle}(x, y, t)$ thì $d(x, t) \leq d(y, t)$; tuy nhiên điều này không nhất thiết phải có ngụ ý rằng $d(x, t) < d(y, t)$.

Điều này quan trọng cần lưu ý ở đây rằng mặc dù nó viết $\text{Oracle}(x, y, t)$ để ứng xuất rằng câu hỏi luôn xảy ra liên quan đến một số đích t , để thực hành đích kín và duy nhất biết đến bởi oracle. Theo một cách khác, theo "oracle như con người" tương tự, người sử dụng có đích trong ý nghĩ và sử dụng nó để so sánh hai đối tượng, nhưng không bao giờ để lộ nó cho đến khi thực tế được trình bày với nó.

B. Yêu cầu, Entropi và hằng số kép

Sự phân bố xác suất μ trên tập đối tượng ở N trong đó có thể được gọi là sự yêu cầu. Có nghĩa là, μ sẽ là hàm số không âm sao cho $\sum_{t \in N} \mu(t) = 1$. Nói chung, sự yêu cầu có thể là không đồng nhất như $\mu(t)$ có thể biến đổi đích khác biệt chéo. Phân bố đích μ sẽ phát vai trò quan trọng theo sự phân tích sau. Cụ thể, hai khối lượng mà tác động thực hiện việc tìm kiếm theo sơ đồ được mô tả sẽ có entropi và hằng số kép của sự phân bố đích. Hai khái niệm này được xác định công thức dưới đây.

Entropi của μ được xác định như

$$H(\mu) = \sum_{x \in \text{supp}(\mu)} \mu(x) \log \frac{1}{\mu(x)}, \quad (2)$$

Trong đó $\text{supp}(\mu)$ hỗ trợ μ . Entropi cực đại của μ được xác định như

$$H_{\max}(\mu) = \max_{x \in \text{supp}(\mu)} \log \frac{1}{\mu(x)}. \quad (3)$$

Đưa ra đối tượng $x \in N$, hình tròn kín có bán kính hình tròn x $R \geq 0$ được biểu thị bằng

$$B_x(R) = \{y \in M : d(x,y) \leq R\} \quad (4)$$

Đưa ra tập hợp $A \subset N$ cho

$$\mu(A) = \sum_{x \in A} \mu(x).$$

Hằng số kép $c(\mu)$ của sự phân bố μ được xác định nhỏ nhất là $c > 0$ trong đó

$$\mu(B_x(2R)) \leq c \cdot \mu(B_x(R)), \quad (5)$$

cho bất kỳ $x \in \text{supp}(\mu)$ và bất kỳ $R \geq 0$. Ngoài ra, nó có thể nói rằng μ là hai lần c nếu $c(\mu) = c$.

Chú ý rằng, ngược lại với entropi $H(\mu)$, hằng số kép $c(\mu)$ phụ thuộc vào cấu trúc liên kết của $\text{supp}(\mu)$, được xác định bằng lồng N trong không gian ma trận (μ, d) .

Bảng I

TÓM TẮT KÝ HIỆU

N	Tập đối tượng
(μ, d)	Không gian ma trận
$d(x, y)$	Khoảng cách giữa $x, y \in M$
μ	Sự phân bố nhu cầu
$H(\mu)$	Entropi của μ
$H_{\max}(\mu)$	Entropi cực đại của μ
$B_x(r)$	Bán kính hình tròn r được chỉnh tâm tại x
$c(\mu)$	Hằng số kép của μ

Trong tính toán vấn đề, ký hiệu của công việc trước đây trong lĩnh vực này

được cho phép. Việc truy cập được đưa ra đối với oracle so sánh, đã mong muốn điều hành thông qua N cho đến khi đối tượng đích được tìm thấy. Cụ thể, việc tìm kiếm nội dung mong muốn được xác định như sau. Cho t là đối tượng đích và s một số đối tượng mà thỏa mãn làm điểm khởi đầu. Thuật toán tìm kiếm nội dung mong muốn đề xuất đối tượng w và yêu cầu oracle để chọn, giữa s và w , đối tượng gần nhất với đích t , tức là, nó gọi ra $\text{oracle}(s, w, t)$. Quá trình này được lặp lại cho đến khi oracle trả lời một ký hiệu nào đó không phải là s , tức là, đối tượng được đề xuất là "giống hơn" so với đích t . Khi điều này xảy ra, nói theo đề xuất của một số w' , nếu $w' \neq t$, việc tìm kiếm nội dung mong muốn lặp lại quá trình tương tự hiện tại từ w' . Nếu ở bất kỳ điểm đối tượng được đề xuất là t , kết thúc quá trình.

Cụ thể, cho x_k, y_k là cặp k -th của đối tượng được nhập vào oracle: x_k là đối tượng hiện tại, mà tìm kiếm nội dung mong muốn cố gắng nâng cao, và y_k là đối tượng được đề xuất, được nhập vào oracle dùng để so sánh với x_k . Cho

$$o_k = \text{Oracle}(x_k, y_k, t) \in \{x_k, y_k\}$$

là hệ số thỏa mãn của oracle, và xác định

$$\mathcal{H}_k = \{(x_i, y_i, o_i)\}_{i=1}^k, \quad k = 1, 2, \dots$$

trình tự của các dữ liệu nhập k thứ nhất đưa ra oracle, cũng như hệ số thỏa mãn thu được. $\mathcal{H}_k$ là "lịch sử" của việc tìm kiếm nội dung lên đến và bao gồm dữ liệu k -th đến oracle.

Đối tượng khởi đầu luôn là một trong hai đối tượng thứ nhất được nhập vào oracle, tức là $x_1 = s$. Ngoài ra, để tìm kiếm nội dung mong muốn,

$$x_{k+1} = o_k, \quad k = 1, 2, \dots$$

tức là, đối tượng hiện tại luôn gần nhất với đích trong số các đối tượng được nhập cho đến nay.

Ngược lại, sự lựa chọn của đối tượng được đề xuất y_{k+1} sẽ được xác định nhờ lịch sử $\mathcal{H}_k$ và đối tượng x_k . Cụ thể, đưa ra $\mathcal{H}_k$ và đối tượng hiện tại x_k tồn tại ánh xạ $(\mathcal{H}_k, x_k) \rightarrow F \{ \mathcal{H}_k, x_k \} \in N$ sao cho

$$y_{k+1} = F(\mathcal{H}_k, x_k), \quad k = 0, 1, \dots,$$

trong đó, $x_0 = s \in N$ (đối tượng khởi đầu) và $H_0 = \emptyset$ {tức là, trước bất kỳ sự so sánh xảy ra, không có lịch sử}.

Ảnh xạ F được gọi là chính sách lựa chọn để tìm kiếm nội dung mong muốn. Nói chung, nếu chính sách lựa chọn cho phép được ngẫu nhiên hóa; trong trường hợp này, đối tượng được trả lời bởi $F\{H_k, X_k\}$ sẽ là biến ngẫu nhiên, mà phân bố

$$\Pr(F(H_k, X_k)=W), W \in N \quad (6)$$

Được xác định đầy đủ bằng (H_k, X_k) . Quan sát rằng F phụ thuộc vào đích t chỉ gián tiếp, thông qua H_k và X_k ; điều này phù hợp với giả định rằng t chỉ "được bộc lộ" khi nó được định vị cuối cùng.

Chính sách lựa chọn được cho là không nhớ nếu nó phụ thuộc vào x_k nhưng không dựa trên lịch sử H_k . Có nghĩa là, phân bố tương tự khi $x_k = x \in N$, không phụ thuộc vào phép so sánh được thực hiện trước khi đến x_k .

Giả định rằng khi $x_k = t$, kết thúc kết quả tìm kiếm (tức là, con người phát hiện ra rằng điều này thực sự là đích), mục đích mong muốn là lựa chọn F để con số truy cập vào oracle là được cực tiểu hóa. Cụ thể, đưa ra đích t và chính sách lựa chọn F , chi phí tìm kiếm được xác định:

$$C_F(t) = \inf\{k: x_k = t\}$$

Con số đề xuất vào oracle cho đến khi t được tìm thấy. Đây là biến ngẫu nhiên, là F như được ngẫu nhiên hóa; cho $E[C_F(t)]$ là kỳ vọng của nó. Việc tìm kiếm nội dung thông qua chương trình so sánh sau đó được xác định như sau:

VIỆC TÌM KIẾM NỘI DUNG THÔNG QUA PHÉP SO SÁNH (CONTENT REARCH THROUGH COMPARISONS-CSTC): Đưa ra phép nhúng của N vào (M, d) và phân bố yêu cầu (μ, d) , lựa chọn F để giảm tối thiểu chi phí tìm kiếm được kỳ vọng

$$\bar{C}_F = \sum_{t \in N} \mu(t) E[C_F(t)].$$

Chú ý rằng, khi F được ngẫu nhiên hóa, biến tự do trong vấn đề tối ưu hóa nêu trên là sự phân bố.

CẬN DƯỚI VÀ THUẬT TOÁN KHÔNG NHỚ

Cận dưới dựa vào con số được kỳ vọng của các câu hỏi mà con số này cần nhập vào oracle so sánh để định vị đích t đã được thiết lập trước đây theo sáng chế.

Định lý 1. Đối với bất kỳ số nguyên K và D , tồn tại không gian ma trận (M, d) và phép đo đích μ bằng entropi $H(\mu) = K \log(D)$ và hằng số kép $c(\mu) = D$ sao cho chi phí tìm kiếm trung bình của bất kỳ chính sách lựa chọn F thỏa mãn

$$\bar{C}_F \geq H(\mu) \frac{c(\mu) - 1}{2 \log(c(\mu))}. \quad (7)$$

Điều quan tâm là chính sách lựa chọn không nhớ đơn giản thỏa mãn giới hạn trên nghĩa là trong phạm vi chỉ số $O(c^2(\mu)H_{\max}(\mu))$ của giới hạn này.

Thuật toán algorit 1 .Tìm kiếm nội dung không nhớ

Dữ liệu nhập: oracle($\cdot, \cdot, i$), yêu cầu phân bố μ , đối tượng khởi đầu s . Dữ liệu xuất: đích t . 1: $x \leftarrow s$

2: While $x \neq t$ do

3: Mẫu $y \in N$ từ sự phân bố xác suất

$$\Pr_x(y) \propto \frac{\mu(y)}{\mu(B_x(d(x, y)))}. \quad (8)$$

5: end while

Định lý 2. Chi phí tìm kiếm được kỳ vọng của thuật toán 1 được giới hạn bởi $C \leq 6c^3(\mu) \cdot H(\mu) \cdot H_{\max}(\mu)$

Có một số quan sát quan tâm được thực hiện về thuật toán 1. Bắt đầu, chính sách lựa chọn không nhớ có các đặc tính hấp dẫn sau. Đối với hai đối tượng y, z mà có khoảng cách tương tự từ x , nếu $\mu(y) > \mu(z)$ thì y có xác suất cao hơn được đề xuất. Khi hai đối tượng y, z tương đương với đích, nếu $d(y, x) < d(z, x)$ thì y có xác suất cao hơn được đề xuất. Sự phân bố (8) do đó cả hai đối tượng gần với x cũng như đối với các đối tượng mà tương đương với đích.

Ngoài ra, để thực hiện chính sách được phác thảo trong thuật toán 1, điều này được tính toán rằng, ở mỗi x , y ngẫu nhiên là được lấy mẫu từ sự phân bố (8). Điều này giả định rằng sự phân bố μ và phép lồng μ (hoặc ma trận khoảng d) là được biết trước.

Tuy nhiên, điều này thực tế rằng thuật toán 1 có thể được thực hiện thậm chí nếu chỉ mỗi quan hệ thứ tự giữa các đối tượng, chứ không phải là khoảng cách thực tế giữa các đích của chúng là đã biết. Điều này là quan trọng, sau đó có thể thu được bằng duy nhất dữ liệu oracle so sánh. Cụ thể, tất cả các mối quan hệ sắp xếp thứ tự này có thể được bộc lộ bằng cách yêu cầu oracle $|N|$ hỏi ngoại tuyến {ví dụ, trong pha huấn luyện}.

Như được chú ý, chỉ số khác biệt chính giữa giới hạn trên trong định lý 2 và giới hạn trên trong định lý 1 của bậc là $c^3 H_{\max}$. Kết quả tiếp theo, xuất hiện trong phần tiếp theo khử bỏ tecbi $H_{\max}$ ở chi phí phụ thuộc vào sự đo kích thước kép thông qua tecbi $O(c^5)$.

THUẬT TOÁN DỰA TRÊN MẠNG LƯỚI $\in$

Đích trong phần này là để thiết lập so sánh dựa trên tìm kiếm có thể cạnh tranh trong việc nhận dạng đối tượng đích $t \in \mathcal{N}$ ban đầu được lấy mẫu theo xác suất phân bố μ trong một vài bước C_F mà giá trị trung bình C_F nghiệm được

$$\bar{C}_F \leq H(\mu) c^*(\mu)$$

Dùng cho một số số mũ k cố định được đồng nhất. Cuối cùng, một số kết quả trung gian được thiết lập.

A. Mạng lưới $\in$

Mạng lưới $\in$ được xác định như sau:

Định nghĩa 1. mạng lưới $\in$ của tập con $A \subset \mathcal{N}$ là tập hợp tối đa các điểm $\{x_1, \dots, x_k\}$ của A sao cho $i \neq j, d(x_i, x_j) > \in$.

Để xây dựng mạng lưới $\in$, cần phải truy cập vào không gian ma trận dưới và khoảng d giữa bất kỳ hai điểm. Quá trình xây dựng mạng lưới có thể diễn ra ở dạng mong muốn trong thời gian $o(K|A|)$, trong đó K là kích cỡ của mạng lưới $\in$. Có trong thuật toán hiệu quả thực tế mà có thể xây dựng mạng lưới như vậy.

Bổ đề 1. Xác định hình tròn $B_x(R) \subset \mathcal{N}$, và số nguyên $t > 0$, bất kỳ mạng lưới $(R/2^t)$ $\{x_1, \dots, x_k\}$ của $B_x(R)$ sao cho:

$$B_x(R) \subset \bigcup_{i=1}^k B_{x_i}(R/2^t), \quad (9)$$

và đối với tất cả $i \neq j$

$$B_{x_i}(R/2^{\ell+1}) \cap B_{x_j}(R/2^{\ell+1}) = \emptyset. \quad (10)$$

Ngoài ra, bản số k của bất kỳ mạng lưới $\{R2^l\}$ này hầu hết là c^{l+3} .

Chứng minh: Nếu (9) không thỏa mãn, thì tồn tại y ở $B_X(R)$ sao cho $d\{y, x_i\} > R/2^l$ cho tất cả $i=1, \dots, k$. Điều này mâu thuẫn với mức tối đa của $\{x_i, \dots, x_k\}$.

Cho tất cả $i \neq j$, bất kỳ điểm z trong sự tương giao $B_{x_i}(R/2^{\ell+1}) \cap B_{x_j}(R/2^{\ell+1})$ sao cho

$$d(x_i, x_j) \leq d(x_i, z) + d(x_j, z) \leq 2R/2^{\ell+1} = R/2^{\ell}.$$

Điều này mâu thuẫn với tính chất rằng $d\{x_i, x_j\} > R/2^l$, do đó sự tương giao $B_{x_i}(R/2^{\ell+1}) \cap B_{x_j}(R/2^{\ell+1})$ là tập rỗng cần thiết.

Cuối cùng, tính chất (10) có hệ quả là

$$\mu(\cup_{i=1}^k B_{x_i}(R/2^{\ell+1})) = \sum_{i=1}^k \mu(B_{x_i}(R/2^{\ell+1})).$$

Ngược lại, việc áp dụng 1 + 2 lần thực tế rằng μ là hai lần c , sau đó cho tất cả $i = 1, \dots, k$,

$$\begin{aligned} \mu B_{x_i}(R/2^{\ell+1}) &\geq c^{-\ell-2} \mu B_{x_i}(2R) \\ &\geq c^{-\ell-2} \mu B_X(R), \end{aligned}$$

bởi vì thực tế rằng $B_X(R) \subset B_{x_i}(2R)$, theo từ $x_i \in B_X(R)$. Đến kết luận, lưu ý rằng

$$\cup_{i=1}^k B_{x_i}(R/2^{\ell+1}) \subset B_X(2R).$$

thì

$$\begin{aligned} c\mu(B_X(R)) &\geq \mu(B_X(2R)) \\ &\geq \mu(\cup_{i=1}^k B_{x_i}(R/2^{\ell+1})) \\ &\geq kc^{-\ell-2} \mu(B_X(R)). \end{aligned}$$

Giới hạn trên $k \leq c^{l+3}$ theo một cách trực tiếp.

Theo điều cần thiết hiện tại:

Bổ đề 2. Cho $\delta \in (0, 1)$ kiểm nghiệm $\delta > 1/3$. Cho hình tròn $B_X(R)$ sao cho tồn tại $a, y \in N$, cho $d(x, y) = R$ và $\mu(\{y\}) > 0$. Sau đó thỏa mãn sau. Cho $p > 0$ sao cho $p < \min(\delta, (1 - \delta y^2)R)$, và cho $f > 0$ là số nguyên dương sao cho

$$2^t \left(\frac{R}{2} - \frac{\rho}{1-\delta} \right) > R \frac{2-\delta}{1-\delta}. \quad (11)$$

Sau đó cho bất kỳ $z \in B_X(R)$, z có

$$\mu \left(B_x \left(\frac{\rho}{1-\delta} \right) \right) \leq (1 - c^{-t}) \mu \left(B_x \left(\frac{R}{1-\delta} \right) \right). \quad (12)$$

Chúng minh: Cho $z \in B_X(R)$ được cố định. Cho $B' := B_z(\frac{\rho}{1-\delta})$. Lưu ý rằng theo giả định rằng $\rho < \delta R$, điều này cho phép rằng B' được bao gồm trong hình tròn $B := B_x(\frac{R}{1-\delta})$.

Bằng cách giả định, tồn tại $y \in N$ sao cho $d(x, y) = R$ và $\mu(\{y\}) > 0$. Do đó $d(x, z)$ hoặc $d(y, z)$ được giới hạn dưới bởi $R/2$: thật vậy, theo bất đẳng thức tam giác,

$$d(x, y) = R \leq d(x, z) + d(y, z).$$

Giả định đầu tiên là $d(x, z) \geq R/2$. Theo lại bất đẳng thức tam giác, cho bất kỳ $z' \in B'$, z' có

$$d(x, z) \leq d(x, z') + d(z, z')$$

Sao cho

$$d(x, z') \geq \frac{R}{2} - \frac{\rho}{1-\delta}.$$

Chú ý rằng giới hạn dưới $R/2 - \rho/(1-\delta)$ là dương theo sự giả định $\rho < (1-\delta)2R$. Có nghĩa là, cho bất kỳ $\alpha > 0$, hình tròn B' là rời từ hình tròn B'' được xác định như

$$B'' := B_x \left(\frac{R}{2} - \frac{\rho}{(1-\delta)} - \alpha \right).$$

Điều này yêu cầu rằng

$$\mu(B'') < \mu \leq \mu(B) - \mu(B') \quad (13)$$

Cho l là số nguyên kiểm nghiệm (11), l sao cho, đối với một số α dương không đủ lớn,

$$2^t \left(\frac{R}{2} - \frac{\rho}{1-\delta} - \alpha \right) \geq \frac{R}{1-\delta}.$$

Điều này yêu cầu rằng

$$\mu(B) \leq \mu \left(B_x \left(2^t \left(\frac{R}{2} - \frac{\rho}{1-\delta} - \alpha \right) \right) \right).$$

Việc áp dụng 1 lần tính chất hai lần c của μ , bất đẳng thức này còn có hệ quả là

$$\mu(B) \leq c^t \mu(B'').$$

Được kết hợp với (13), bất đẳng thức cuối cùng này dẫn đến

$$\mu(B') \leq (1 - c^{-t})\mu(B),$$

mà giới hạn được mong muốn (12).

Tiếp theo giả định rằng $d(x, z) < R/2$, sao cho $d(y, z) \geq R/2$ cần thiết. Bây giờ cho bất kỳ $z' \in B'$, theo bất đẳng thức tam giác z' có

$$d(y, z) \leq d(y, z') + d(z, z'),$$

để xác định B''' hiện tại là

$$B''' := B_y \left(\frac{R}{2} - \frac{\rho}{(1-\delta)} - \alpha \right).$$

Đối với một số $a > 0$ nhỏ một cách tùy ý, hai hình tròn B' và B''' là rời nhau. Lưu ý còn B' này chứa B , vì đối với bất kỳ $z''' \in B'''$, z''' có

$$d(x, z''') \leq d(x, y) + d(y, z''') \leq R + R/2,$$

và giả định $\delta > 1/3$ đảm bảo rằng $(3/2)R \leq R/(1-\delta)$ là bán kính của B .

Tương tự với (13) do đó chúng ta có

$$\mu(B''') \leq \mu(B) - \mu(B').$$

Cho 1 là số nguyên dương kiểm nghiệm (11). Áp dụng bất đẳng thức tam giác có hệ quả là phép lồng

$$B \subset B_t \left(2^t \left(\frac{R}{2} - \frac{\rho}{1-\delta} - \alpha \right) \right)$$

phải thỏa mãn cho $a > 0$ không đủ lớn. Thật vậy, đối với bất kỳ điểm $x' \in B$, x' có

$$d(y, x') \leq R + \frac{R}{1-\delta} = R \frac{2-\delta}{1-\delta},$$

và tính chất (11) đảm bảo rằng x' trong hình tròn $B_y(2^l(R/2 - \rho(1-\delta) - \alpha))$ tương ứng.

Cuối cùng, bằng cách sử dụng 1 lần tính chất hai lần c của μ cho phép thiết lập $\mu(B) \leq c^l \mu(B''')$; được kết hợp với (13), điều này như trong trường hợp trước để tính chất mong muốn (12).

Chú ý 1. Đối với $R > 0$ đưa ra, giả định theo bổ đề 2 được xác nhận nếu giả định có $p = R/4$, $\delta = 1/3 + \epsilon$ để đủ nhỏ $\epsilon > 0$, và $l = 5$. Thật vậy, điều kiện $p < \min(\delta, (1 - \delta)2)R$ bởi vì thỏa mãn $1/4 < 1/3$. Viết lại $(1 - \delta)^{-1} = (3/2) + \epsilon'$ đối với một số ϵ' dương nhỏ tùy ý, điều kiện (11) đọc sau khi đơn giản hóa bằng R :

$$2^l (1/2 - (1/4)(3/2 + \epsilon')) > 1 + 3/2 + \epsilon',$$

Mà được xác nhận rõ ràng đối với $l = 5$ và $\epsilon' > 0$ đủ nhỏ.

B. Thuật toán và giới hạn dưới

Thuật toán 2. Việc tìm kiếm nội dung mạng lưới $\in$

Dữ liệu nhập: Oracle($\cdot, f$) yêu cầu phân bố μ , đối tượng khởi đầu s , phép lồng (M, d) .

Dữ liệu xuất: đích t . 1 : Khởi chạy $x_0 \leftarrow s$.

2: Bán kính tìm kiếm R_0 ban đầu theo $R_0 := \sup_{y \in Nd} d(x_0, y)$.

3: $j \leftarrow 0$.

4: While $x_j \neq t$ do

5: Xây dựng mạng lưới $(\frac{R_j}{4})$

6: Bằng cách sử dụng oracle so sánh, tìm đối tượng điển hình nhất x_{j+1} so với đích t trong số các điểm trong mạng lưới $(\frac{R_j}{4})$ và x_j .

7: Cập nhật bán kính tìm kiếm

$$R_{j+1} = \inf \{ R : \mu(B_{x_{j+1}}(R)) = \mu(B_{x_{j+1}}(R_j/4)) \}.$$

8: $j \leftarrow j+1$

9: end while

Thuật toán được đề xuất theo sáng chế dựa trên mạng lưới $\in$ có thể được tìm thấy ở thuật toán 2. Tóm lại, chiến lược tìm kiếm được xem xét tiến hành trong các giai đoạn. Các giai đoạn này được ký hiệu là $j = 1, \dots, S$. Khi bắt đầu giai đoạn j , mẫu điển hình nhất hiện nay đưa ra được ký hiệu là x_j , và bán kính hiện nay tìm kiếm là R_j mà sao cho quan điểm về sự lựa chọn thực hiện trong giai đoạn trước, đích tìm kiếm là cần thiết trong phạm vi hình tròn $B_j := B_{x_j}(R_j)$. Điều này còn được đặt tại mỗi giai đoạn

j , bán kính tìm kiếm R_j sao cho tồn tại điểm $y_j \in N$ sao cho $\mu(\{y_j\}) > 0$ và $d(x_j, y_j) = R_j$, tức là, sự phân bố nhu cầu μ đặt một số lượng trên biên của B_j .

Giai đoạn thứ nhất được khởi chạy bằng cách chọn lọc ứng cử ban đầu bất kỳ $X_1 \in N$. Bán kính tìm kiếm ban đầu tương ứng sau đó được xác định như $R_1 := \sup_{y \in \text{supp}(\mu)} d(x_1, y)$. Do đó, bằng cách xây dựng hình tròn ban đầu B_1 thật vậy có lượng khác không tại biên của nó.

Tìm kiếm trong khi giai đoạn j bất kỳ tiến hành như sau. Trung tâm x_j tìm kiếm hiện tại được hoàn thành bằng các điểm bổ sung của B_j để tạo ra mạng lưới p_j của B_j , trong đó $p = R_j/4$. Sau khi đối tượng so sánh được thực hiện giữa lựa chọn cuối cùng và mỗi điểm của mạng lưới mà phân biệt với x_j . Kết thúc những so sánh, cho x là lựa chọn cuối cùng của người sử dụng. Rõ ràng, sự lựa chọn này là trong số các điểm của mạng lưới, mà gần nhất với đích của việc tìm kiếm.

Vì (quan điểm của bổ đề 1) kết hợp hình tròn được chỉnh tâm tại các điểm của mạng lưới, và có bán kính p_j , bao gồm toàn bộ cơ sở tìm kiếm hiện tại B_j , điều này cho phép rằng cần thiết đích phải nằm trong hình tròn $B_{x_j}(p_j)$.

Một công đoạn cuối cùng cần thiết để định rõ như thế nào giai đoạn tiếp theo $j + 1$ được khởi chạy. Trung tâm để tìm kiếm ở giai đoạn $j + 1$ sẽ được thiết lập để $x_{j+1} := x'_j$. Điều này đã biết rằng đích nằm trong phạm vi $B_{x_{j+1}}(p_j)$. Sau khi định rõ bán kính tìm kiếm R_{j+1} là R nhỏ nhất sao cho $\mu(B_{x_{j+1}}(R)) = \mu(B_{x_{j+1}}(p_j))$. Do đó cần thiết, $R_{j+1} < p_j$, và ngoài ra mức tối thiểu của R_{j+1} có hệ quả là phép đo μ đặt một số lượng trên biên của hình tròn tìm kiếm kết quả S_{j+1} . Như, phương pháp này đã thực sự đảm bảo bằng cách xây dựng ở bất kỳ giai đoạn j (a) đích nằm trong hình tròn hiện tại B_j và (b) hình tròn chứa đối tượng của lượng khác không tại biên của nó.

Số lượng câu hỏi nhập vào oracle có thể được giới hạn bằng thuật toán 2.

Thuật toán 2 là thuật toán mong muốn mà sử dụng lịch sử tìm kiếm để đề xuất các đối tượng mới. Một phương án của phương pháp 100 theo sáng chế được thể hiện trong Fig.1. Phương pháp bao gồm bước 110 xây dựng mạng lưới theo kích cỡ đã biết. Mạng lưới này được xây dựng một cách đảm bảo để chứa đích (nghĩ về nó như quả bóng chứa điểm bên trong). Phương pháp còn bao gồm bước 120 để lựa chọn một vài

mẫu và cũng bao gồm bước 130 dùng để so sánh các mẫu với một mẫu khác. Mẫu gần giống với đích chọn lựa trong bước 140 và sau đó mạng lưới khác có kích cỡ nhỏ hơn (tức là, hình tròn nhỏ hơn) được xây dựng lại trong bước 150 vòng tròn đối tượng này. Phương pháp phải đảm bảo rằng đích được chứa trong mạng lưới. Quá trình này được lặp lại cho đến khi điều kiện cuối đạt được trong bước 160, như đích định vị. Nếu điều kiện cuối đã đạt được, đích có thể được đặt trong phạm vi mạng lưới và phương pháp dừng lại. Nếu điều kiện cuối chưa đạt được, phương pháp khôi phục trở lại bước 120 và chọn các mẫu có kích cỡ mạng lưới nhỏ hơn.

Một phương án về thiết bị 200 thực hiện việc tìm kiếm nội dung được thể hiện trong Fig.2. Thiết bị này bao gồm máy tính mà thực hiện phương pháp 100.

Một phương án chi tiết về thiết bị 200 dùng để tìm kiếm nội dung được thể hiện trong Fig.3. Thiết bị này bao gồm mạch xây dựng mạng lưới 210. Mạng lưới này được xây dựng một cách đảm bảo để chứa đích. Thiết bị còn bao gồm mạch lựa chọn mẫu 220. Thiết bị cũng bao gồm mạch so sánh 230. Mạch so sánh 230 có thể so sánh các mẫu trong các cặp, hoặc tất cả cùng một lúc, phụ thuộc vào nguồn tài nguyên và/hoặc thời gian sẵn có. Thiết bị cũng bao gồm mạch xác định 240. Mạch xác định 240 xác định các mẫu gần nhất với đích. Sự xác định có thể được thực hiện trong một hoặc nhiều cách khác nhau, như sự khác biệt tuyệt đối, v.v. Thiết bị còn bao gồm mạch giảm mạng lưới 250. Mạch giảm mạng lưới 250 phải đảm bảo rằng đích vẫn được chứa trong mạng lưới, trong khi làm giảm kích cỡ của mạng lưới. Quá trình này được lặp lại cho đến khi điều kiện cuối đạt được. Thiết bị cũng bao gồm mạch điều khiển 260 mà được sử dụng để kiểm soát hệ vận hành của các thành phần khác nhau và, cụ thể, kiểm soát con số của phép lặp mà các thành phần thực hiện giảm mạng lưới so với điều kiện cuối, mà được giám sát bởi mạch điều khiển.

Điều kiện cuối có thể là một điều kiện hoặc sự kết hợp của các điều kiện. Ví dụ, một điều kiện có thể là mạng lưới đủ nhỏ để định vị đích. Điều kiện có thể khác là kích cỡ của mạng lưới nằm trong giá trị ngưỡng. Điều kiện có thể khác là mạch trong phương pháp 100 được thực hiện tần số được xác định trước. Điều kiện có thể khác là đích của chính nó chọn lựa khi xác định mẫu gần nhất với đích.

Theo phương án khác, kích cỡ của mạng lưới có thể bị giảm bằng cách thực

hiện được lắp lại hệ vận hành của mạch cho đến khi mạng lưới bị giảm, và sau đó phương án thay thế có thể được sử dụng để định vị thực sự đích trong phạm vi giảm kích cỡ mạng lưới. Phương án này có thể được sử dụng, ví dụ, khi nó tính toán hiệu quả hơn để thực hiện lựa chọn cuối cùng bằng phương pháp thay thế chứ không phải là thực hiện phép lắp nhiều hơn của mạch.

Định lý 3. Chi phí tìm kiếm được kỳ vọng của thuật toán 2 có thể được giới hạn bởi

$$\bar{C}_{\mathcal{F}} \leq (c^5 - 1) \left(1 + \frac{H(\mu)}{\log(1/(1 - c^{-5}))} \right). \quad (14)$$

Tại mỗi giai đoạn j một so sánh được thực hiện giữa lựa chọn cuối cùng và mỗi điểm của mạng lưới p_j mà phân biệt với x . Kích cỡ của mạng lưới p_j này, theo bổ đề 1, tại hầu hết c^5 . Do đó, tại hầu hết so sánh nhị phân $c^5 - 1$ là cần thiết ở mỗi giai đoạn.

Ký hiệu lại bằng x'_j lựa chọn cuối cùng ở giai đoạn j . Cũng ký hiệu bằng $\pi_j := \mu(B_{x'_j}(R(1 - \delta)))$ lượng đặt bằng phép đo μ trên vùng tìm kiếm B_j , sau khi tăng bán kính của nó bằng chỉ số $1/(1 - \delta)$, trong đó $\delta = 1/3 + \epsilon$, cho một số ϵ nhỏ chọn lựa như trong chú ý 1. Giờ theo bổ đề 2 và chú ý 1 rằng cần thiết,

$$\mu(B_{x'_j}(\rho_j/(1 - \delta))) \leq (1 - c^{-5})\pi_j.$$

Cũng chú ý rằng, giới hạn, bởi bổ đề 2 và đối số cảm ứng, nó đảm bảo rằng ở mỗi giai đoạn j tìm kiếm

$$\pi_j = \mu(B_{x'_j}(R_j/(1 - \delta))) \leq (1 - c^{-5})^{j-1}.$$

Sau đó đặt điều kiện trên thành phần đích $z \in N$. Xem xét xác suất của nó $\mu(\{z\})$ và giới hạn trước trên xác suất của vùng tìm kiếm sau giai đoạn j , rõ ràng tìm kiếm sẽ được hoàn thành sau khi giai đoạn j được đề xuất

$$(1 - c^{-5})^{j-1} \leq \mu(\{z\}),$$

Hoặc tương đương, được đề xuất

$$j \geq 1 + \frac{\log(1/\mu(\{z\}))}{\log(1/(1 - c^{-5}))}.$$

Số trung bình của giai đoạn, S , sau khi được giới hạn dưới bằng

$$\begin{aligned}\bar{S} &\leq \sum_{z \in \mathcal{N}} \mu(\{z\}) \left(1 + \frac{\log(1/\mu(\{z\}))}{\log(1/(1-c^{-5}))}\right) \\ &= 1 + \frac{H(\mu)}{\log(1/(1-c^{-5}))}.\end{aligned}$$

Chú ý là, trong giai đoạn, tại hầu hết so sánh $c^5 - 1$ được thực hiện, giới hạn trên (14) sau.

Điều này chú ý rằng định lý 3 đưa ra giới hạn trên mà giới hạn dưới so khớp (7), đến sự không khớp về số mũ của hằng số kép c . Ngược lại, với thuật toán 1, mà có thể được thực hiện duy nhất bằng cách sử dụng mối quan hệ sắp xếp thứ tự giữa các đối tượng chứ không phải là khoảng cách chính xác, thuật toán 2 thật sự yêu cầu hiểu biết đầy đủ về không gian ma trận dưới. Quan trọng, thuật toán 2 không yêu cầu hiểu biết về phân bố đích μ . Tất cả các bước trong thuật toán (và, cụ thể, sự co của hình tròn B_j đảm bảo nó có lượng khác không tại biên) có thể được thực hiện miễn là $\text{supp}(\mu)$ hỗ trợ đã biết.

Kết luận

Các nguyên lý được mô tả ở đây đề xuất giải pháp về vấn đề của việc tìm kiếm nội dung thông qua phép so sánh (CSTC) theo nhu cầu không đồng nhất, giao điểm biểu diễn về cấu trúc liên kết và entropi của phân bố đích. Chiến lược tìm kiếm được xem xét ở thuật toán 2 được tính dựa trên quá trình xây dựng mạng lưới $\in$ ở giai đoạn khác nhau của việc tìm kiếm, mà cấp thiết truy cập vào thông tin chi tiết về hình học của không gian tìm kiếm (M, d) , nhưng không có thông tin về sự phân bố nhu cầu μ .

Một hoặc nhiều hệ thống xử lý có các đặc điểm đặc biệt và các khía cạnh về các phương án được ưu tiên hiện nay theo sáng chế đã được đề xuất. Tuy nhiên, các đặc điểm và các khía cạnh được mô tả hệ thống xử lý cũng có thể được ứng dụng cho hệ thống xử lý khác. Ví dụ, các hệ thống xử lý và các đặc điểm có thể được sử dụng ở thuộc tính của hệ thống thiết bị hình ảnh khác. Hệ thống xử lý và các đặc điểm không cần phải được sử dụng theo tiêu chuẩn.

Tham chiếu trong bản mô tả đến "một phương án" hoặc "phương án" hoặc "một hệ thống xử lý" hoặc "hệ thống xử lý" của sáng chế, cũng như sự biến đổi khác của chúng, có nghĩa là đặc điểm đặc biệt, cấu trúc, đặc trưng, và được nêu được mô tả

trong sự kết hợp với phương án được bao gồm trong ít nhất một phương án của sáng chế. Do đó, xuất hiện mệnh đề "trong một phương án" hoặc "trong phương án" hoặc "trong một hệ thống xử lý" hoặc "trong hệ thống xử lý", cũng như bất kỳ sự biến đổi khác, xuất hiện ở những nơi khác nhau thông qua bản mô tả không cần thiết tất cả đề cập đến phương án tương tự.

Hệ thống xử lý được mô tả ở đây có thể được thực hiện trong, ví dụ, phương pháp hoặc quá trình, thiết bị, chương trình phần mềm, dòng dữ liệu, hoặc tín hiệu. Thậm chí nếu chỉ được thảo luận theo sáng chế về dạng đơn của hệ thống xử lý (ví dụ, được thảo luận chỉ như phương pháp), hệ thống xử lý về các đặc điểm được thảo luận cũng có thể được thực hiện trong các dạng khác (ví dụ, thiết bị hoặc chương trình phần mềm máy tính). Thiết bị có thể được thực hiện trong, ví dụ, phần cứng thích hợp, phần mềm, và phần sụn. Phương pháp có thể được thực hiện trong, ví dụ, thiết bị như, ví dụ, bộ xử lý, trong đó đề cập đến các thiết bị xử lý nói chung, bao gồm, ví dụ, máy tính, bộ vi xử lý, mạch được cài đặt sẵn, hoặc thiết bị logic lập trình. Bộ xử lý cũng bao gồm các thiết bị truyền thông, như, ví dụ, máy tính, điện thoại di động, máy hỗ trợ cá nhân/di động kỹ thuật số ("portable/personal digital assistants -PDAs"), và các thiết bị khác mà liên lạc dễ dàng về thông tin giữa người sử dụng cuối cùng.

Hệ thống xử lý của các quá trình khác nhau và các đặc điểm được mô tả ở đây có thể được biểu hiện trong một loạt các thiết bị hoặc các ứng dụng khác nhau. Ví dụ thiết bị này bao gồm máy chủ mạng lưới, máy tính xách tay, máy tính cá nhân, điện thoại di động, PDA, và các thiết bị truyền thông khác. Nên được rõ ràng, thiết bị có thể được di chuyển và thậm chí được cài đặt trong phương tiện chuyển động.

Ngoài ra, phương pháp có thể được thực hiện theo sự hướng dẫn được thực hiện bởi bộ xử lý, và sự hướng dẫn này (và/hoặc giá trị dữ liệu được tạo ra bằng hệ thống xử lý) có thể được lưu trên bộ xử lý trung bình có thể đọc được như, ví dụ, mạch được cài đặt sẵn, sóng mang phần mềm hoặc thiết bị lưu trữ khác như, ví dụ, đĩa cứng, đĩa compac, bộ nhớ truy cập ngẫu nhiên ("random access memory -RAM"), hoặc bộ nhớ chỉ đọc ("read-only memory -ROM"). Sự hướng dẫn có thể tạo ra chương trình ứng dụng hữu hình được biểu hiện trên bộ xử lý trung bình có thể đọc được. Sự hướng dẫn có thể là, ví dụ, trong phần cứng, phần sụn, phần mềm, hoặc sự kết hợp. Sự hướng dẫn

có thể được tìm thấy trong, ví dụ, hệ thống điều hành, ứng dụng riêng biệt, hoặc kết hợp cả hai. Bộ xử lý có thể là được đặc trưng, do đó, như, ví dụ, cả hai thiết bị cấu hình để thực hiện quá trình và thiết bị này bao gồm bộ xử lý trung bình có thể đọc được (như thiết bị lưu trữ) có sự hướng dẫn để thực hiện quá trình. Ngoài ra, bộ xử lý trung bình có thể đọc được có thể lưu trữ, ngoài ra đến hoặc đặt sự hướng dẫn, giá trị dữ liệu được tạo ra bằng hệ thống xử lý.

Như sẽ được nêu rõ bởi người có hiểu biết trung bình trong lĩnh vực này, hệ thống xử lý có thể sử dụng tất cả hoặc phần của phương pháp được mô tả ở đây. Hệ thống xử lý có thể bao gồm, ví dụ, sự hướng dẫn dùng để thực hiện phương pháp, hoặc dữ liệu được tạo ra bởi một phương án được mô tả.

Một số hệ thống xử lý đã được mô tả. Tuy nhiên, sẽ được hiểu rằng sự biến đổi khác nhau có thể được thực hiện. Ví dụ, các thành phần của các hệ thống xử lý khác nhau có thể được kết hợp, được bổ sung, được biến đổi, hoặc được loại bỏ để tạo ra hệ thống xử lý khác. Ngoài ra, người có hiểu biết trung bình trong lĩnh vực này sẽ hiểu được rằng cấu trúc và quy trình khác có thể được thay thế được bộc lộ và hệ thống xử lý thu được sẽ thực hiện ít nhất hầu như cùng chức năng, trong ít nhất hầu như cùng cách, thu được ít nhất hầu như cùng (các) kết quả như các phương án được bộc lộ. Do đó, các phương án này và các phương án khác được dự định bởi sáng chế và nằm trong phạm vi của các điểm yêu cầu bảo hộ kèm theo.

YÊU CẦU BẢO HỘ

1. Phương pháp tìm kiếm nội dung trong cơ sở dữ liệu, phương pháp này bao gồm các bước:

xây dựng mạng lưới có kích cỡ, mạng lưới này chứa đích, trong đó mạng lưới này bao gồm tập dữ liệu trong cơ sở dữ liệu có bán kính tra cứu, bằng cách bao gồm dữ liệu với đó các khoảng cách đến các dữ liệu khác bất kỳ ít nhất là trị số nhỏ nhất;

chọn nhiều mẫu, trong đó mỗi mẫu là một mẫu dữ liệu trong cơ sở dữ liệu;

so sánh mỗi mẫu với mọi mẫu khác, bằng cách sử dụng chuẩn đo khoảng cách;

xác định mẫu gần nhất với đích dựa trên sự so sánh này;

làm giảm kích cỡ của mạng lưới thành kích cỡ nhỏ hơn, mạng lưới kích thước được làm giảm chứa đích bằng cách loại bỏ dữ liệu mà bên ngoài bán kính của sự so sánh gần nhất;

lặp lại các bước chọn, so sánh, xác định, và làm giảm đối với ít nhất hai phép lặp cho đến khi kích cỡ của mạng lưới nằm trong giá trị ngưỡng.

2. Phương pháp theo điểm 1, trong đó bước lặp lại được thực hiện cho một số phép lặp được xác định trước.

3. Phương pháp theo điểm 1, trong đó đích được định vị bằng phương pháp tìm kiếm thay thế sau khi mạng lưới trở nên đủ nhỏ.

4. Thiết bị tìm kiếm nội dung trong cơ sở dữ liệu, bao gồm:

mạch xây dựng mạng lưới có kích cỡ, mạng lưới này chứa đích, trong đó mạng lưới này bao gồm tập dữ liệu trong cơ sở dữ liệu có bán kính tra cứu bằng cách bao gồm dữ liệu với đó các khoảng cách đến các dữ liệu khác bất kỳ ít nhất là trị số nhỏ nhất;

mạch chọn nhiều mẫu, trong đó mỗi mẫu là một mẫu dữ liệu trong cơ sở dữ liệu;

mạch so sánh mà vận hành trên các mẫu bằng cách sử dụng chuẩn đo khoảng cách;

mạch xác định mà tìm mẫu gần nhất với đích dựa trên sự so sánh;

mạch giảm kích cỡ của mạng lưới đến kích cỡ nhỏ hơn, mạng lưới được làm giảm kích cỡ này chứa đích bằng cách loại bỏ dữ liệu mà bên ngoài bán kính của sự so sánh gần nhất; và

mạch điều khiển làm cho mạch xây dựng, mạch lựa chọn, mạch so sánh, mạch xác định, và mạch làm giảm và để lặp lại sự vận hành của chúng đối với ít nhất hai phép lặp cho đến khi kích cỡ của mạng lưới cuối cùng nằm trong giá trị ngưỡng.

5. Thiết bị theo điểm 4, trong đó mạch điều khiển này làm cho mạch xây dựng, mạch lựa chọn, mạch so sánh, mạch xác định, và mạch này làm giảm để lặp lại sự vận hành của chúng cho đến khi kích cỡ của mạng lưới cuối cùng nằm trong giá trị ngưỡng.

6. Thiết bị theo điểm 4, trong đó mạch điều khiển làm cho đích được định vị bằng phương pháp tìm kiếm thay thế sau khi mạng lưới trở nên đủ nhỏ.

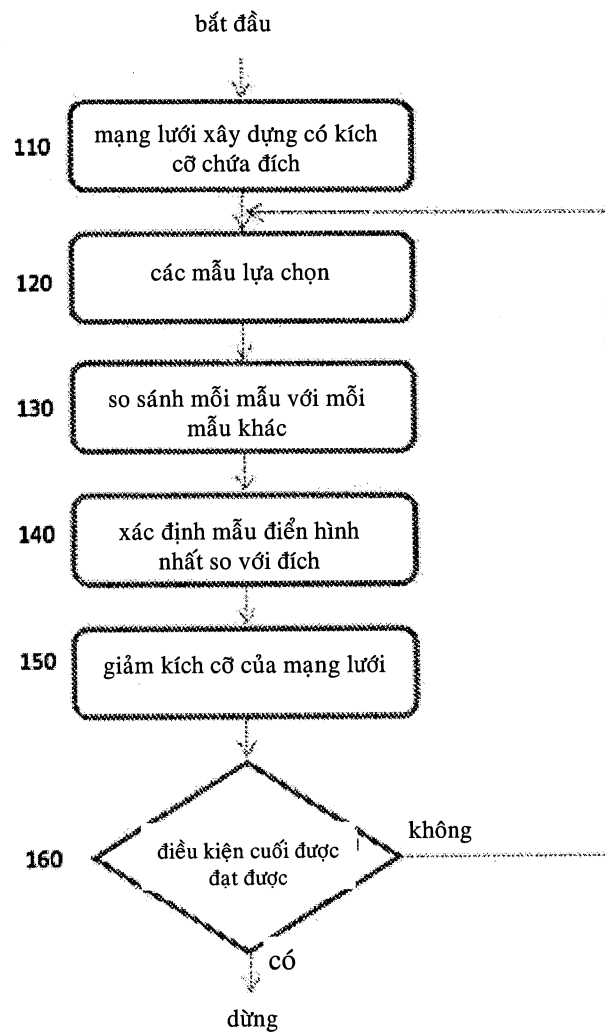


Fig. 1

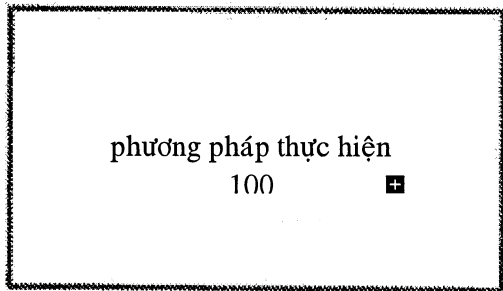


Fig. 2

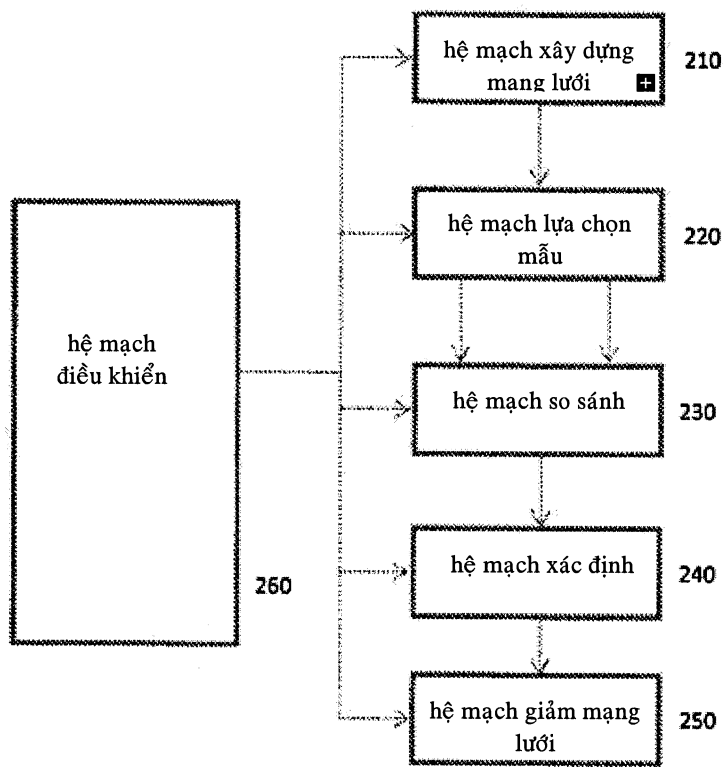


Fig. 3