

DỮ LIỆU ĐA PHƯƠNG TIỆN VÀ MIRS

Hệ cơ sở dữ liệu đa phương tiện

HK1, 2023 - 2024

Dữ liệu văn bản

- Chứa thông tin chủ đạo
- **Input:** bàn phím, các chương trình nhận dạng âm thanh và ký tự, dữ liệu lưu trên đĩa, phụ đề phim...



Comment au temps du bon roy Artus il
estoit ung tres expert Nigromancien que
on appelloit Merlin.

Dữ liệu văn bản

PHÂN RÃ MA TRẬN VỚI YẾU TỐ THỜI GIAN TRONG HỆ THỐNG GỢI Ý

Lê Ngọc Quỳnh, Nguyễn Hữu Hòa và Nguyễn Thái Nghe

Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ

Thông tin chung:

Ngày nhận bài: 15/09/2017

Ngày nhận bài sửa: 10/10/2017

Ngày duyệt đăng: 20/10/2017

Title:

Matrix and tensor factorization with temporal effect in recommender systems

Từ khóa:

Hệ thống gợi ý, làm trơn hàm mũ, phân rã ma trận, phân rã nhân tử

Keywords:

Exponential smoothing, matrix factorization, recommender systems, tensor factorization

ABSTRACT

This paper proposes the construction of a recommender system to predict users' preferences based on matrix factorization techniques. Because of the changes of users' preferences time by time, to achieve more accurate result, exponential smoothing is integrated into the matrix factorization model by utilizing tensor factorization. This usage aims at exploiting and taking advantage of information about the time and the order of users' giving feedbacks. The model is tested relied on the datasets in suggestion and evaluation using the root mean squared error. The experimental results demonstrate fairly good performance of the proposed method.

TÓM TẮT

Bài viết này đề xuất một giải pháp dự đoán sở thích của người dùng dựa trên kỹ thuật phân rã ma trận (Matrix Factorization – MF) có tích hợp yếu tố thời gian trong hệ thống gợi ý (Recommender Systems – RS). Do sở thích của người dùng có thể thay đổi theo thời gian, để kết quả gợi ý có độ chính xác cao hơn chúng tôi đề xuất tích hợp phương pháp dự báo san bằng hàm mũ (Exponential Smoothing – ES) vào mô hình Tensor Factorization với mục tiêu khai thác và tận dụng được các thông tin về thời gian cũng như trình tự (sequence) mà người dùng đã đưa ra phản hồi. Thực nghiệm ban đầu trên các tập dữ liệu chuẩn trong lĩnh vực gợi ý và đánh giá bằng độ đo RMSE (Root Mean Squared Error) đã cho thấy hướng nền tảng nêu cho kết quả rất khả quan

- Kích thước lưu trữ: không đáng kể so với các dữ liệu đa phương tiện khác
- Định dạng: đa dạng
 - Văn bản thường (file ASCII)
 - Văn bản có cấu trúc, được định dạng (màu sắc, độ bóng...): html, xml, rtf, word, mã nguồn, latex, pdf...

Nén dữ liệu

- Nén không làm mất mát thông tin
- Trên thực tế
 - Tần suất xuất hiện của các ký tự khác nhau
- Một số kỹ thuật
 - Huffman coding
 - Run-length Coding
 - Lempel Zip-Welch Coding (LZW)

Nén dữ liệu: Huffman coding

- Dựa trên tần suất xuất hiện của ký tự (item)
- Từ nào xuất hiện nhiều thì dùng ít bit để mã hóa và ngược lại
- Mã cho mỗi ký tự được lưu trong Code book tương ứng cho mỗi tài liệu

Ký tự	Tần suất	Mã
F	0.8	1
T	0.16	01
X	0.02	001
Z	0.02	000

Nén dữ liệu: Huffman coding

Ví dụ: Giả sử ta có chuỗi

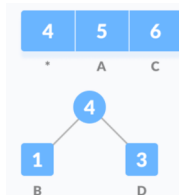
B C A A D D D C C A C A C

B1 Tần số xuất hiện của mỗi ký tự **B: 1, C: 6, A: 5, D: 3**

B2 Sắp xếp theo tần suất **B: 1, D: 3, A: 5, C:6**

B3 Mỗi ký tự được xếp là nút lá

B4 Tạo một nút rỗng z. Gán nút lá có tần số thấp nhất cho nút lá trái và tần số thấp nhì cho nút lá phải. Giá trị z bằng tổng giá trị 2 nút lá

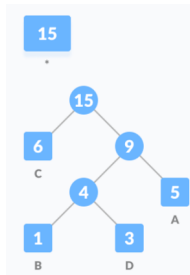
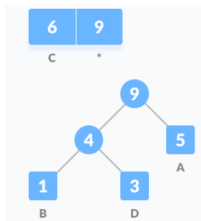


Nén dữ liệu: Huffman coding

B5 Bỏ 02 giá trị tần số thấp nhất khỏi hàng đợi và thêm giá trị nút tạm

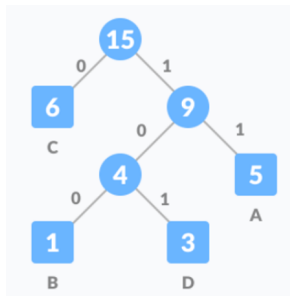
B6 Thêm nút z vào cây

B7 Lặp lại các bước 3 – 5 cho tất cả ký tự



Nén dữ liệu: Huffman coding

B8 Với mỗi nút không là nút lá, gán 0 cho nút lá trái và 1 cho nút lá phải



Ký tự	Tần số	Code	Kích thước
A	5	11	$5 * 2 = 10$
B	1	100	$1 * 3 = 3$
C	6	0	$6 * 1 = 6$
D	3	101	$3 * 3 = 9$
4 * 8 bits = 32 bits		09 bits	28 bits

- Chưa nén: $8 \times 15 = 120$
- Nén: $32 + 9 + 28 = 69$

Nén dữ liệu: Run-length Coding

- Dựa trên số lần lặp liên tục của các ký tự
- Ví dụ: xét đoạn văn bản eeeeeeebtntnnnnnnn
- Kết quả: **@e8bt@n7**

Với @ là ký tự đặc biệt để chỉ rằng đang mã hóa

Nén dữ liệu: LZW

- Dựa trên tần suất lặp của các cụm ký tự
- Xây dựng từ điển cho các cụm ký tự
 - Giả sử tài liệu có 10.000 ký tự
 - Chia thành 2.000 cụm ký tự
 - Trong đó có 500 cụm khác nhau
 - Nếu không nén: $10.000 \times 8 \text{ bit} = 80.000 \text{ bit}$
 - LZW: $2.000 \times 9 \text{ bit} = 18.000 \text{ bit}$
 - Để biểu diễn được 500 token cần 9 bit

Token	Chuỗi ký tự
...	...

Dữ liệu đồ họa và ảnh động

Đồ họa

- Pixel-based graphic: xử lý giống ảnh số
- Vector-based graphic:
 - Mô hình được định nghĩa trước
 - Không gian lưu trữ thấp
 - Dễ lấy nội dung

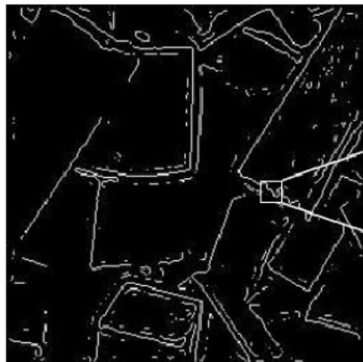
Ảnh động

- Pixel-based: giống video
- Vector-based: giống vector-based graphic nhưng có thêm thông số về thời gian

Ảnh số

- Ảnh số là một chuỗi các điểm ảnh để biểu diễn 1 vùng sẽ được hiển thị trên màn hình của người sử dụng
- **Input:** camera, scanner, output của các chương trình mô phỏng hay các phần mềm tạo và xử lý ảnh
- Định dạng: jpg, png, bmp, tiff,...

Ảnh số: Ảnh trắng đen



1	1	0	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	1	1	0
0	0	0	0	0	1

Mức nhị phân – 1 bit:

- 0 - đen
- 1 - trắng

Ảnh số: Ảnh trắng đen



230	229	232	234	235	232	148
237	236	236	234	233	234	152
255	255	255	251	230	236	161
99	90	67	37	94	247	130
222	152	255	129	129	246	132
154	199	255	150	189	241	147
216	132	162	163	170	239	122

Mức xám – 8 bit:

- 0 - đen
- 255 - trắng

Ảnh số: Ảnh màu



49	55	56	57	52	53
58	60	60	58	55	57
58	58	54	53	55	56
83	78	72	69	68	69
88	91	91	84	83	82
69	76	83	78	76	75
61	69	73	78	76	76

Red

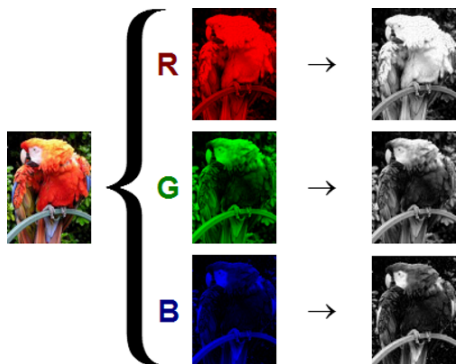
64	76	82	79	78	78
93	93	91	91	86	86
88	82	88	90	88	89
125	119	113	108	111	110
137	136	132	128	126	120
105	108	114	114	118	113
96	103	112	108	111	107

Green

66	80	77	80	87	77
81	93	96	99	86	85
83	83	91	94	92	88
135	128	126	112	107	106
141	129	129	117	115	101
95	99	109	108	112	109
84	93	107	101	105	102

Blue

Ảnh số: Ảnh màu



- Mỗi điểm ảnh có 3 thành phần: R, G, B
- 1 ảnh số = 3 ma trận giá trị số biểu diễn cường độ R, G, B
- Các hệ tọa độ màu: RGB, YUV, HSV

Kỹ thuật nén

Dựa trên

- Dư thừa trong ảnh (lặp giá trị)
- Dư thừa theo khả năng nhận thức (nhìn thấy) của con người
 - Ví dụ: 02 điểm lân cận ở 02 hàng liền nhau thì có giá trị gần giống nhau mà con người khó phân biệt
- Với các mẫu khác nhau, độ nhạy cảm về nhận thức là khác nhau

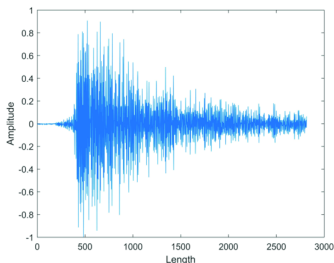
Mục tiêu: Giữ lại các thông tin quan trọng và loại bỏ các thông tin khác

Kỹ thuật nén

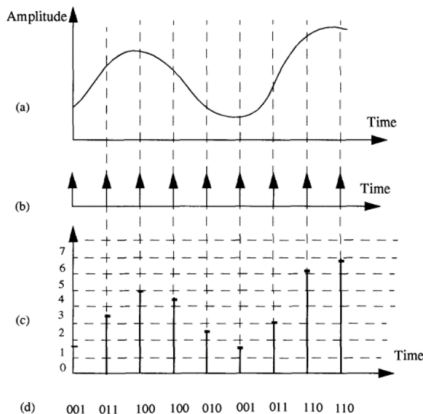
- Spatial Subsampling
- Prediction Coding
- Transform Coding
- Vector Quantization (tương tự LZW)
- Fractal Image Coding
- Chuẩn nén JPEG
- ...

Dữ liệu âm thanh

- Tiếng động, tiếng nói (văn bản đi kèm), nhạc, phim, các chương trình dịch tự động từ văn bản
- Tín hiệu âm thanh là tín hiệu tương tự và liên tục
- Input: microphone → số hóa và lưu trữ
- Không gian lưu trữ lớn
- Thường được nén để giảm kích thước (mp3, acc, ogg...)



ADC (Analog-to-Digital Converter)



- Lấy mẫu (sampling)
- Lượng tử hóa (quantization)
- Mã hóa (encoding)

Kỹ thuật nén

- Lượng tử phi tuyến
- Predictive coding
- Chuẩn nén audio: MPEG - Audio

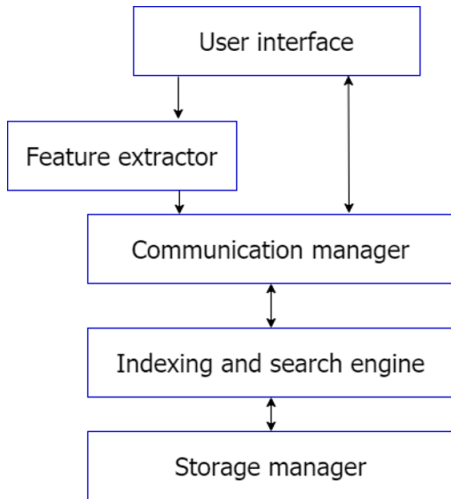
Video

- Gồm mỗi chuỗi các khung hình (frames) (25, 30, 50 frames/s)
- Input: video camera \rightarrow số hóa
- Định dạng: đa dạng (mp4, avi, ...)
- Không gian lưu trữ: tốn nhất
- Dữ liệu phải được nén

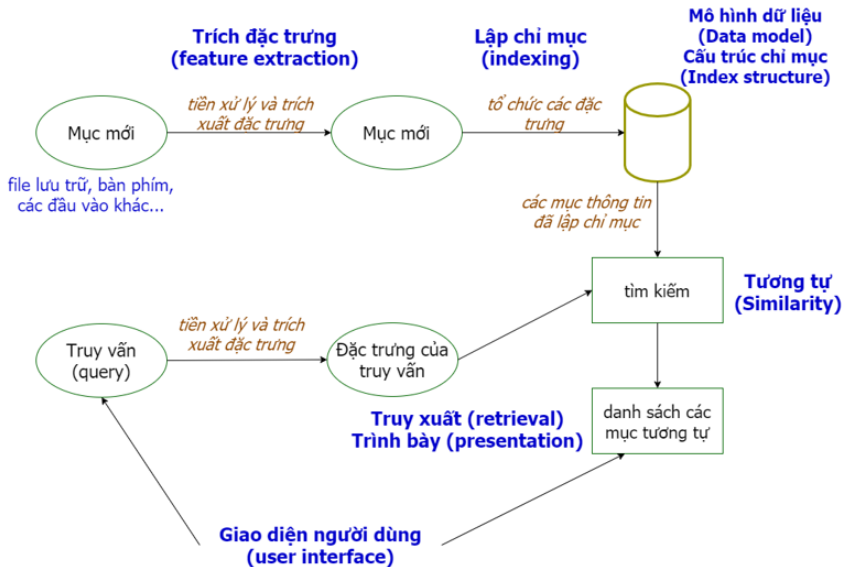
Video

- Frame rate đủ lớn để thấy truyền tải sự chuyển động liên tục (≥ 25 frames/s)
- Băng thông để truyền tải tín hiệu: frame rate càng lớn thì băng thông càng lớn (25 frames/s được sử dụng)
- Tốc độ làm tươi màn hình hiển thị (≥ 50 lần/s)
- Nén:
 - Dựa vào dư thừa về mặt không gian (giữa các điểm ảnh) (spatial redundancy)
 - Dựa vào dư thừa về mặt thời gian (giữa các khung hình) (temporal redundancy)

Kiến trúc MIRS



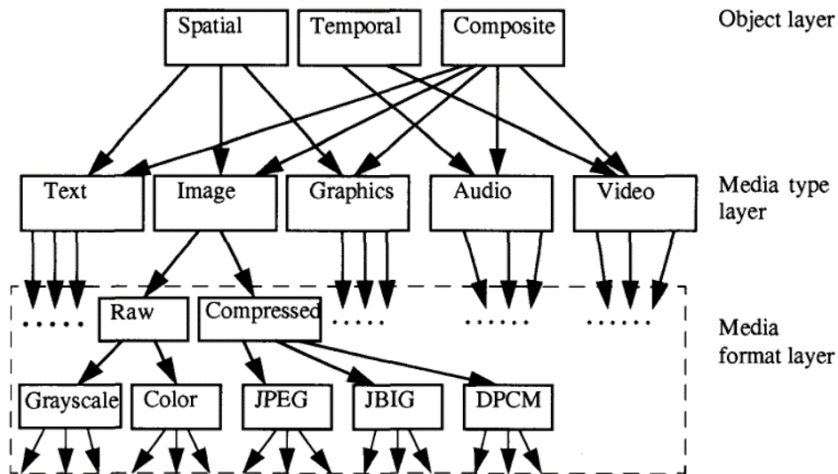
Các hoạt động chính



Mô hình dữ liệu

- Data model xác định
 - Cách thông tin tổ chức và lưu trữ
 - Những loại truy vấn được hỗ trợ
- Mô hình dữ liệu nên:
 - Có thể mở rộng
 - Có thể đại diện cho loại phương tiện cơ bản, các mối quan hệ thời gian và không gian
 - Linh hoạt → các mục có thể được xác định, truy vấn, tìm kiếm ở các mức độ trừu tượng khác nhau
 - Cho phép lưu trữ và tìm kiếm hiệu quả

Mô hình dữ liệu đa phương tiện tổng quát



Mô hình dữ liệu đa phương tiện tổng quát

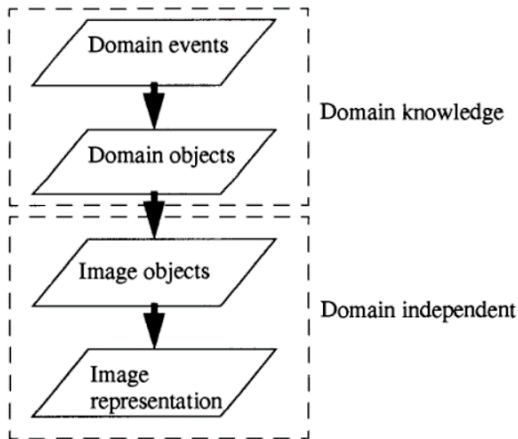
- Dựa trên hướng đối tượng
- Nhiều lớp
 - Lớp đối tượng (Object layer)
 - Lớp kiểu đa phương tiện (Media type layer)
 - Lớp định dạng phương tiện (Media format layer)

Một số vấn đề khác

- Mỗi lớp
 - Không được thiết kế hoàn chỉnh
 - Không có tiêu chuẩn chung
 - Phần lớn MIRS: cho ứng dụng cụ thể
 - Giới hạn số lượng đặc trưng
 - Giới hạn số lượng loại dữ liệu
- Mô hình dữ liệu đặc biệt cho mỗi ứng dụng
- VIMSYS: ảnh + video
 - Mô hình video tổng quát
 - Cấu trúc giản đồ hình ảnh

Ví dụ: VIMSYS

Visual Information Management System



Giao diện người dùng

Yêu cầu của User Interface

- Thêm các mục cơ sở dữ liệu dễ dàng
- Nhập truy vấn dễ dàng và hiệu quả
- Thể hiện kết quả truy vấn cho người dùng hiệu quả
- Thân thiện người dùng

→ cho phép người dùng

- Chỉ định các loại input khác nhau
- Tạo các đối tượng đa phương tiện
- Chỉ định các loại thuộc tính được trích xuất và lập chỉ mục

Giao diện người dùng - Hỗ trợ truy vấn

Truy vấn đa phương tiện

- Đa dạng (diverse)
- Không rõ/mờ (fuzzy)

Các công cụ

- Tìm kiếm
 - Bảng từ khóa, tham số
 - Bảng ví dụ
- Duyệt: bắt đầu với
 - Một truy vấn rất lớn
 - Dựa trên tổ chức CSDL
 - Chọn mục ngẫu nhiên

- Sàng lọc truy vấn: phản hồi (feedback)

Giao diện người dùng - Trình bày kết quả

- Trình bày tất cả các loại phương tiện, mối quan hệ thời gian + không gian, QoS
- Cách trích xuất và trình bày thông tin cần thiết để duyệt: đoạn âm thanh dài, video dài, hình ảnh lớn.
- Thời gian phản hồi phải ngắn (thời gian hệ thống con giao tiếp + thời gian tìm kiếm CSDL)
- Xác nhận phản hồi về mức độ liên quan và sàng lọc truy vấn

Trích xuất đặc trưng (feature extraction)

- Xác định hiệu quả truy xuất
- Yêu cầu
 - đầy đủ nhất có thể để trình bày nội dung của các mục thông tin
 - được trình bày và lưu trữ nhỏ gọn
 - Việc tính toán khoảng cách giữa các tính năng phải hiệu quả, nếu không, thời gian phản hồi của hệ thống sẽ dài

Trích xuất đặc trưng (feature extraction)

Mức độ đặc trưng	Ví dụ	Kỹ thuật xử lý
Metadata (siêu dữ liệu)	Tên tác giả, ngày, tựa...	DBMS
Chú thích văn bản (Text annotation)	Từ khóa mô tả nội dung: đẹp, buồn...	IR
Mức độ thấp (low-level)	Audio: phân phối tần số... Hình: phân phối màu, kết cấu, hình dạng,...	CBR
Mức độ cao (high-level)	Nhận dạng và diễn dịch cho con người	CBR

Lập chỉ mục (indexing)

Một đối tượng \sim nhiều đặc trưng

Một đặc trưng \sim nhiều tham số

Lập chỉ mục trong MIRS nên:

- Phân cấp
- Diễn ra ở nhiều cấp độ
 - Phân loại ứng dụng
 - Các cấp độ khác nhau của các đặc trưng
 - Mối quan hệ không gian và thời gian giữa các đối tượng

Đo mức độ tương tự (similarity)

- Mức tương tự: được tính dựa trên đặc trưng đã trích xuất
- Mức độ liên quan của các kết quả truy xuất: được đánh giá bởi con người (chủ quan và phụ thuộc vào ngữ cảnh)
- Các giá trị tương tự được tính phải phù hợp với đánh giá của con người?
 - Đặc trưng được sử dụng?
 - Biện pháp đánh giá tương tự được sử dụng?

QoS (Quality-of-Service)

Dữ liệu đa phương tiện yêu cầu:

- Băng thông lớn
- Không gian lưu trữ lớn và tốc độ truyền cao
- Giới hạn độ trễ và rung
- Đồng bộ hóa không gian và thời gian

Các thành phần chính:

- Các host (bao gồm client và server) dưới sự kiểm soát của hệ điều hành
- Bộ quản lý lưu trữ
- Hệ thống truyền thông tải hoặc truyền thông