Ứng Dụng Đặc Trưng Đa Phương Thức Trong Hệ Khuyến Nghị Sách Tiếng Việt Trong Thương Mại Điện Tử

Lê Xuân Bình[1,2] **Thái Minh Lâm**[1,2] **Mã Kim Phát**[1,2]
[1]Khoa Khoa học và Kỹ thuật Thông tin, Trường Đại học Công nghệ Thông tin,
Thành phố Hồ Chí Minh, Việt Nam
[2]Đại học Quốc gia Thành phố Hồ Chí Minh, Việt Nam
{22520131, 22520745, 22521071}@gm.uit.edu.vn

Tóm tắt nội dung

1 Giới thiệu

Sự bùng nổ của thương mại điện tử đã thay đổi căn bản hành vi mua sách: người dùng ngày nay ra quyết định dựa trên nhiều yếu tố như hình ảnh bìa, tóm tắt nội dung và đánh giá trực tuyến (Zhang et al., 2019). Mặc dù các hệ thống khuyến nghị truyền thống hoạt động hiệu quả trên dữ liệu tương tác, chúng thường gặp hạn chế với dữ liệu thưa và chưa khai thác triệt để thông tin ngữ nghĩa từ dữ liệu phi cấu trúc (Su and Khoshgoftaar, 2009). Các nghiên cứu hiện có về đa phương thức thường chỉ xem hình ảnh hay văn bản là đặc trưng bổ sung đơn giản, thiếu các đánh giá sâu sắc về hiệu quả thực sự của việc kết hợp các phương thức này (Deldjoo et al., 2020). Đặc biệt, sự thiếu hụt một bộ dữ liệu chuẩn hóa chứa đầy đủ thông tin văn bản và hình ảnh cho sách tiếng Việt đang là rào cản lớn cho nghiên cứu trong nước.

Nghiên cứu này tập trung giải quyết các thách thức trên thông qua ba đóng góp chính:

- Xây dựng bộ dữ liệu: Chúng tôi giới thiệu bộ dữ liệu sách tiếng Việt đầu tiên tích hợp đầy đủ đặc trưng đa phương thức.
- Đánh giá thực nghiệm: Chúng tôi triển khai và so sánh các mô hình SOTA để làm rõ tác động của việc tích hợp đặc trưng đa phương thức so với các phương pháp đơn phương thức.
- Triển khai hệ thống: Chúng tôi xây dựng một hệ thống khuyến nghị hoàn chỉnh, tối ưu hóa cho khả năng mở rộng và tích hợp công nghệ dữ liệu lớn, phù hợp với yêu cầu thực tế của thương mại điện tử.

2 Các công trình liên quan

Về hệ khuyến nghị sách: Các bộ dữ liệu chuẩn trong lĩnh vực khuyến nghị sách đã có sự phát triển từ bộ dữ liệu thưa Book-Crossing (Ziegler and et al., 2005) đến các tài nguyên quy mô lớn như Amazon Books (McAuley and et al., 2015) và Goodreads (Wan and McAuley, 2018). Mặc dù các bộ dữ liệu này cung cấp lịch sử tương tác người dùng phong phú, chúng chủ yếu tập trung vào các siêu dữ liệu có cấu trúc. Xu hướng chuyển dịch gần đây hướng tới tích hợp đa phương thức được thể hiện rõ trong nghiên cứu của (Spillo and et al., 2025), qua việc bổ sung các tín hiệu văn bản và hình ảnh cho bộ dữ liệu DBbook. Tuy nhiên, các tài nguyên này hầu như chỉ dành cho tiếng Anh. Cho đến nay, vẫn chưa có một bộ dữ liệu đa phương thức chuẩn hóa, quy mô lớn nào được xây dựng riêng cho thị trường sách Việt Nam.

Các hướng tiếp cận khuyến nghị cốt lõi: Các nghiên cứu hiện hành phân loại chiến lược khuyến nghị thành ba trụ cột chính: (1) Lọc cộng tác (Collaborative Filtering - CF) và các biến thể nơ-ron (NCF), vốn mô hình hóa các tương tác tiềm ẩn giữa người dùng và mục tiêu (Su and Khoshgoftaar, 2009; He and et al., 2017); (2) Lọc dựa trên nội dung (Content-Based Filtering - CBF), tận dụng sự tương đồng của siêu dữ liệu (Lops et al., 2011); và (3) Khuyến nghị dựa trên tri thức (Knowledge-aware Recommendation), sử dụng Đồ thị tri thức nhằm giảm thiểu sự thưa thớt dữ liệu (Guo et al., 2022). Dù có hiệu quả mạnh mẽ, các mô hình này thường coi dữ liệu phi cấu trúc chỉ là thông tin phụ trợ.

Kết hợp đa phương thức và Hiệu quả theo miền: Các hệ thống khuyến nghị đa phương thức (MRS) thường sử dụng các đặc trưng văn bản và hình ảnh thô để bổ trợ cho ma trận tương tác (He and McAuley, 2016; Deldjoo et al., 2020). Tuy nhiên, hiện có một lỗ hổng quan trọng trong việc đánh giá tầm quan trọng của từng phương thức đối với các miền cụ thể (Wei et al., 2024). Nhiều mô hình hiện nay bỏ qua thực tế rằng khả năng dự đoán của một phương thức (ví dụ: phần tóm tắt sách so với hình ảnh bìa) thay đổi đáng kể tùy thuộc vào

bối cảnh. Nghiên cứu của chúng tôi giải quyết vấn đề này bằng cách phân tích hiệu quả cụ thể của các đặc trưng hình ảnh và văn bản tiếng Việt trong một khung khuyến nghị thống nhất.

3 Bộ dữ liệu

3.1 Thu thập dữ liệu

Chúng tôi đã xây dựng một bộ dữ liệu mới bằng cách thu thập dữ liệu từ Tiki.vn[1], một nền tảng thương mại điện tử nổi bật về sách tại Việt Nam. Quá trình trích xuất dữ liệu được thực hiện thông qua phương pháp lai: sử dụng BeautifulSoup[2] để truy xuất nhanh các siêu dữ liệu và đánh giá của người dùng, và Selenium[3] để thu thập các phần mô tả văn bản động. Dữ liệu thô bao gồm tiêu đề sách, hình ảnh bìa, mô tả văn bản, giá cả và lịch sử đánh giá của người dùng.

3.2 Tiền xử lý và Tăng cường dữ liệu

Dữ liệu thô (Hình 1) đã trải qua quá trình làm sạch kỹ lưỡng, bao gồm chuẩn hóa Unicode, sửa lỗi chính tả cho văn bản tiếng Việt và loại bỏ các nhiễu không liên quan như hashtag, thẻ HTML và biểu tượng cảm xúc (emoji). Để giải quyết vấn đề dữ liệu thưa thớt và duy trì sự đa dạng của dữ liệu trong quá trình lọc k-core ($k = 5$), chúng tôi đã thực hiện tăng cường dữ liệu: đối với các mục có ít hơn k lượt đánh giá, chúng tôi tạo ra các đánh giá tổng hợp và gán chúng cho các mã định danh khách hàng (customer ID) giả định với số lượt tương tác ban đầu bằng 1. Chiến lược này giúp ngăn chặn việc mất đi các item ít được đánh giá và duy trì mật độ cấu trúc của ma trận tương tác.

Đối với tác vụ khuyến nghị đa phương thức, chúng tôi áp dụng tiêu chí lọc như sau để đảm bảo mô hình hoạt động tốt: (1) Loại bỏ bất kỳ bản ghi sách nào thiếu hình ảnh bìa hoặc mô tả văn bản. (2) Lọc Core-k: Áp dụng bộ lọc $k = 3$ để đảm bảo mỗi người dùng và mục tiêu còn lại đều có ít nhất ba tương tác liên quan.

3.3 Phân tích dữ liệu và Thống kê

Bộ dữ liệu sau khi làm sạch bao gồm 1.115 cuốn sách và 26.050 lượt đánh giá. Các thuộc tính chi tiết được cung cấp trong phần Phụ lục.

Chúng tôi đã thực hiện phân tích để hiểu rõ các phân phối cơ bản của dữ liệu (Hình 2). Phân phối của các xếp hạng trung bình cho thấy một sự lệch dương đáng kể, với sự tập trung cao của các mức điểm từ 4.0 đến 5.0. Điều này cho thấy tâm thế ủng hộ chung của khách hàng, đòi hỏi chiến lược khuyến nghị phải có khả năng phân biệt hiệu quả giữa các mục đều có xếp hạng cao. Ngoài ra, phân tích mối quan hệ giữa giá sách (theo thang log) và xếp hạng trung bình cho thấy không có sự tương quan rõ rệt. Kết quả này gợi ý rằng giá cả không phải là yếu tố quyết định chính đến sự hài lòng của khách hàng trong bộ dữ liệu này, từ đó củng cố việc chúng tôi tập trung vào nội dung, ảnh bìa và sở thích người dùng thay vì các yếu tố về giá cả.

4 Khung thực nghiệm

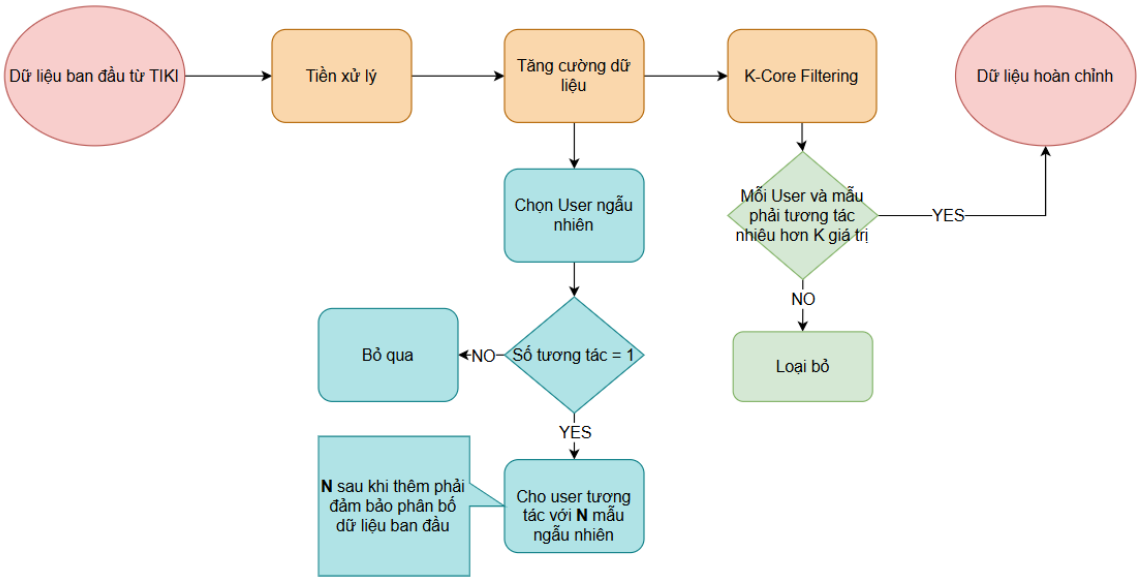
Cách tiếp cận phân chia dữ liệu có sự khác biệt tùy thuộc vào phương pháp thực hiện. Cụ thể, đối với hướng tiếp cận lọc cộng tác (Collaborative Filtering) và khuyến nghị dựa trên nội dung (Content-based Filtering), chúng tôi áp dụng chiến lược Leave-one-out. Theo đó, toàn bộ lịch sử tương tác giữa người dùng và sản phẩm được sắp xếp theo thời gian; sản phẩm cuối cùng trong danh sách tương tác của mỗi người dùng sẽ được đưa vào tập đánh giá (test set), trong khi tất cả các tương tác trước đó được sử dụng làm tập huấn luyện (training set). Đối với hướng tiếp cận sử dụng các đặc trưng đa phương thức, để tối ưu hóa quá trình huấn luyện và tinh chỉnh siêu tham số cho các mô hình học sâu phức tạp, chúng tôi phân chia bộ dữ liệu thành ba phần riêng biệt: huấn luyện (training set), phát triển (dev set) và kiểm thử (test set). Cụ thể, chúng tôi mở rộng chiến lược Leave-one-out bằng cách trích xuất hai sản phẩm cuối cùng của mỗi người dùng: sản phẩm cuối cùng được đưa vào tập kiểm thử, sản phẩm ngay trước đó được sử dụng làm tập phát triển để phục vụ việc lựa chọn mô hình tối ưu, và toàn bộ các sản phẩm còn lại cấu thành nên tập huấn luyện. Phương pháp này đảm bảo rằng mô hình không chỉ được đánh giá trên dữ liệu chưa biết mà còn được hiệu chỉnh một cách khách quan trước khi đưa ra kết quả cuối cùng. Chúng tôi sử dụng các mô hình tiền huấn luyện để trích xuất đặc trưng cho từng phương thức trong Bảng 1

Bảng 1: Encoder for text & image

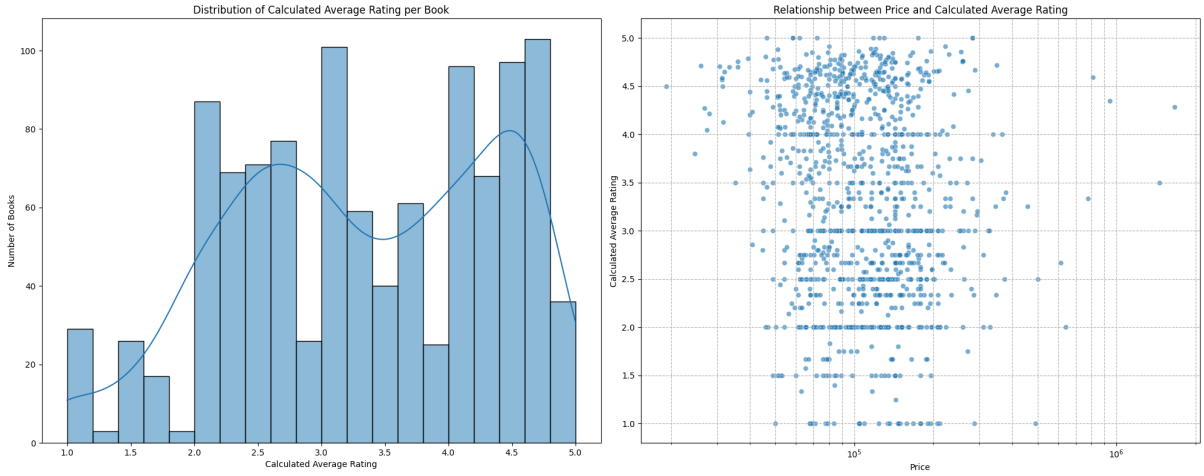
Text Encoder	ViSoBERT
Image Encoder	ViT, ResNet-152

Sau đó, chúng tôi áp dụng MMRec (Zhou, 2023) như một khung thực nghiệm nhằm huấn luyện và đánh giá các mô hình khuyến nghị đa phương thức.

[1] TIKI
[2] beautifulsoup4
[3] Selenium



Hình 1: Chiến lượt xử lý và tăng cường dữ liệu



Hình 2: Phân phối đánh giá trung bình (Trái) và Mối quan hệ giữa giá cả và đánh giá (Phải).

Quy trình tổng quát gồm: đóng gói dữ liệu, thiết lập cấu hình, huấn luyện mô hình và đánh giá hiệu suất.

5 Thực nghiệm và Đánh giá kết quả

5.1 Độ đo đánh giá

Để đánh giá hiệu năng của các mô hình khuyến nghị, chúng tôi sử dụng bốn độ đo tiêu chuẩn tại các ngưỡng $k \in \{5, 10\}$. Gọi R_u là tập hợp k mục được khuyến nghị hàng đầu cho người dùng u, và T_u là tập hợp các mục dữ liệu thực tế (các mục mà người dùng đã thực sự tương tác).

Precision@k và Recall@k Precision đo lường tỷ lệ các mục được khuyến nghị là có liên quan, trong khi Recall đo lường tỷ lệ các mục có liên quan trong dữ liệu thực tế đã được hệ thống khuyến nghị thành công:

$$Precision@k = \frac{|R_u \cap T_u|}{k} \tag{1}$$

$$Recall@k = \frac{|R_u \cap T_u|}{|T_u|} \tag{2}$$

Mean Average Precision (mAP@k) mAP là độ đo xem xét đến thứ hạng của các mục có liên quan trong danh sách. Trước tiên, chúng tôi tính toán Độ chính xác trung bình (Average Precision - AP) cho một người dùng cụ thể, trong đó $rel(i)$ là một biến chỉ báo nhị phân có giá trị bằng 1 nếu mục tại vị trí

thứ i có liên quan và bằng 0 nếu ngược lại:

$$AP@k = \frac{1}{\min(|T_u|, k)} \sum_{i=1}^{k} (Precision@i \times rel(i)) \tag{3}$$

Sau đó, mAP được tính bằng giá trị trung bình của $AP@k$ trên tất cả người dùng U trong tập kiểm thử.

Normalized Discounted Cumulative Gain (NDCG@k) NDCG đánh giá chất lượng xếp hạng bằng cách áp dụng hình phạt đối với các mục liên quan nằm ở vị trí thấp trong danh sách thông qua cơ chế suy giảm logarit. Độ đo DCG (Discounted Cumulative Gain) được định nghĩa như sau:

$$DCG@k = \sum_{i=1}^{k} \frac{2^{rel(i)} - 1}{\log_2(i+1)} \tag{4}$$

Giá trị NDCG thu được bằng cách chuẩn hóa DCG theo IDCG (Ideal DCG) — đây là giá trị DCG tối đa đạt được khi danh sách được xếp hạng một cách hoàn hảo:

$$NDCG@k = \frac{DCG@k}{IDCG@k} \tag{5}$$

5.2 Kết quả thực nghiệm

Bảng 2 và 3 tóm tắt hiệu năng của các mô hình khuyến nghị được đánh giá dựa trên tất cả các độ đo đã nêu.

Tất cả các kiến trúc đa phương thức (VBPR, LATTICE, FREEDOM, MMGCN, SLMRec) đều đạt kết quả vượt trội so với baseline BPR (mô hình chỉ dựa trên tương tác thuần túy). Điều này khẳng định rằng việc tích hợp thông tin hình ảnh và văn bản giúp giảm bớt vấn đề dữ liệu thưa thớt (sparsity) và cải thiện độ chính xác của hệ thống gợi ý. Trong số các mô hình được thử nghiệm, MMGCN đạt hiệu năng cao nhất trên hầu hết các chỉ số. Cụ thể, kiến trúc **MMGCN** với thông tin Text (Type) đạt R@5 = 0.0424 và N@10 = 0.0315, cao gấp khoảng 4 lần so với mô hình BPR. Kết quả này cho thấy khả năng của mạng đồ thị (GNN) trong việc lan truyền thông tin đa phương thức qua cấu trúc người dùng.

Tuy nhiên, có thể thấy được việc lựa chọn **Encoder** đóng vai trò then chốt đến kết quả, cụ thể như sau:

- Đối với **Textual Encoder**, ViSoBERT cho thấy sự ổn định và hiệu quả rất cao khi xử lý dữ liệu văn bản tiếng Việt. Đặc biệt, thông tin Text (Type) (thể loại sản phẩm) thường mang lại kết quả tốt hơn Text (Desc) (mô tả chi tiết). Điều này có thể giải thích do thông tin thể loại có tính cô đọng cao, giúp mô hình dễ dàng học được các đặc trưng phân loại rõ rệt.
- Đối với **Visual Encoder**, ViT (CLS) đạt kết quả tốt hơn so với ResNet-152. Ví dụ, trong mô hình MMGCN, cấu hình sử dụng ViT đạt R@10 = 0.0566 (cao nhất trong bảng), chứng tỏ kiến trúc Self-attention của Transformer giúp bắt được các đặc trưng hình ảnh quan trọng tốt hơn so với các lớp tích chập (CNN) truyền thống.

Để đánh giá khách quan hiệu quả của các mô hình đa phương thức hiện đại, chúng tôi tiến hành so sánh với các nhóm phương pháp truyền thống bao gồm các phương pháp như: Lọc cộng tác dựa trên người dùng (User-Based Collaborative Filtering) và Khuyến nghị dựa trên nội dung (Description-Based và Typebook-Based). Các phương pháp tiếp cận truyền thống này đạt kết quả rất thấp. User-based (Cosine) chỉ đạt Rec@5 = 0.0057 và NDCG@10 = 0.0065. Trong khi đó, mô hình MMGCN (Text-Type) đạt Rec@5 = 0.0424 và NDCG@10 = 0.0315. Mô hình đề xuất cải thiện hiệu suất gấp ~ 7.4 lần về khả năng gợi ý chính xác (Recall) và gấp ~ 4.8 lần về khả năng xếp hạng (NDCG).

Cuối cùng, việc kết hợp đồng thời cả văn bản và hình ảnh (Text + Image) không phải lúc nào cũng cho kết quả cao nhất ở mọi chỉ số so với việc dùng đơn lẻ một phương thức mạnh (như Description Type)

6 Triển khai hệ thống

Sau khi thực nghiệm, chúng tôi chọn ra mô hình có kết quả tốt nhất để xây dựng hệ thống khuyến nghị.

Hệ thống được triển khai giả định trên nền tảng Dữ Liệu Lớn được minh hoạ trong Hình 3 và 4. Kiến trúc sử dụng Streamlit cho giao diện dùng, giao tiếp thông qua FastAPI để xử lý các yêu cầu không bộ. Luồng dữ liệu được điều phối thông qua Apache Kafka, đảm bảo khả năng chịu lỗi và băng thông cao. Công nghệ xử lý cốt lõi là Apache Spark (Structured Streaming), tích hợp mô hình đề đa phương thức được đào tạo từ trước để thực hiện suy luận theo thời gian thực. Đặc biệt, để dễ dàng triển khai các môi trường và nền tảng khác nhau chúng tôi đóng gói cả hệ thống bằng Docker để đảm bảo triển khai liền mạch và tính nhất quán.

Bảng 2: Kết quả thực nghiệm của các phương pháp tiếp cận truyền thống.

	Pre@5	**Rec@5**	**mAP@5**	**NDCG@5**	**Pre@10**	**Rec@10**	**mAP@10**	**NDCG@10**
User-based (Cosine)	0.0011	0.0057	0.0027	0.0034	0.0016	0.0156	0.0039	0.0065
User-based (Pearson)	0.0006	0.0028	0.0008	0.0013	0.0004	0.0042	0.0011	0.0018
Content-based (Cosine)	0.0011	0.0057	0.0026	0.0034	0.0017	0.0170	0.0041	0.0070
Content-based (Pearson)	0.0011	0.0057	0.0026	0.0034	0.0017	0.0170	0.0041	0.0070
Typebook-based (Cosine)	**0.0028**	0.0141	**0.0052**	**0.0074**	**0.0018**	**0.0183**	**0.0058**	**0.0088**
Typebook-based (Pearson)	0.0019	**0.0099**	0.0022	0.0039	0.0014	0.0141	0.0027	0.0053

Bảng 3: Kết quả thực nghiệm của các mô hình đa phương thức dựa trên các Encoder khác nhau.

Model	**Modality**	**Encoder**	**P@5**	**R@5**	**mAP@5**	**N@5**	**P@10**	**R@10**	**mAP@10**	**N@10**
BPR	-	-	0.002	0.0099	0.0043	0.0057	0.0014	0.0141	0.0049	0.0071
VBPR	Text (Desc)	ViSoBERT	0.0023	0.0113	0.0035	0.0054	0.0018	0.0184	0.0046	0.0078
	Text (Type)	ViSoBERT	0.0031	0.0156	0.0058	0.0081	0.0027	0.0269	0.0072	0.0116
	Image	ViT (CLS)	0.0025	0.0127	0.0099	0.0106	0.0023	0.0226	0.0114	0.014
	Image	ResNet-152	0.0028	0.0141	0.007	0.0088	0.0021	0.0212	0.0078	0.011
	Text + Image	ViSoBERT + Res152	0.0034	0.017	0.0081	0.0102	0.0023	0.0226	0.0088	0.012
LATTICE	Text (Desc)	ViSoBERT	0.0025	0.0127	0.0031	0.0054	0.0024	0.024	0.0045	0.009
	Text (Type)	VisoBERT	0.0031	0.0156	0.0086	0.0103	0.0023	0.0226	0.0096	0.0124
	Image	ViT (CLS)	0.0028	0.0141	0.0054	0.0075	0.0028	0.0283	0.007	0.0118
	Image	ResNet-152	0.0014	0.0071	0.0031	0.0041	0.0013	0.0127	0.0038	0.0058
	Text + Image	ViSoBERT + ViT (CLS)	0.0037	0.0184	0.0067	0.0095	0.0028	0.0283	0.0078	0.0125
FREEDOM	Text (Desc)	ViSoBERT	0.0025	0.0127	0.0074	0.0086	0.0021	0.0212	0.0085	0.0114
	Text (Type)	VisoBERT	0.0054	0.0269	0.0113	0.015	0.0035	0.0354	0.0124	0.0178
	Image	ViT (CLS)	0.0028	0.0141	0.007	0.0087	0.002	0.0198	0.0076	0.0105
	Image	ResNet-152	0.0006	0.0028	0.0017	0.002	0.001	0.0099	0.0025	0.0041
	Text + Image	ViSoBERT + ViT (CLS)	0.0048	0.024	0.0111	0.0142	0.0031	0.0311	0.012	0.0165
MMGCN	Text (Desc)	ViSoBERT	0.0076	0.0382	0.023	0.0267	0.0048	0.0481	0.0243	0.0299
	Text (Type)	VisoBERT	**0.0085**	**0.0424**	**0.0236**	**0.0282**	0.0052	0.0523	**0.025**	**0.0315**
	Image	ViT (CLS)	0.0082	0.041	0.0133	0.0201	**0.0057**	**0.0566**	0.0156	0.0254
	Image	ResNet-152	0.0076	0.0382	0.0121	0.0185	0.004	0.0396	0.0124	0.019
	Text + Image	ViSoBERT + ViT (CLS)	0.0079	0.0396	0.0192	0.0243	0.0045	0.0453	0.0198	0.026
SLMRec	Text (Desc) + Image	ViSoBERT + ViT (CLS)	0.0059	0.0297	0.0128	0.0169	0.005	0.0495	0.0155	0.0234
	Text (Type) + Image	ViSoBERT + ViT (CLS)	0.0062	0.0311	0.0177	0.021	0.005	0.0495	0.0201	0.0269
	Text (Desc) + Image	ViSoBERT + ResNet-152 (CLS)	0.0071	0.0354	0.0164	0.021	0.0055	0.0552	0.0191	0.0275
	Text (Type) + Image	ViSoBERT + ResNet-152 (CLS)	0.0059	0.0297	0.0194	0.022	0.0038	0.0382	0.0205	0.0247

Giao diện của hệ thống được minh họa trong Hình 3

7 Hạn chế

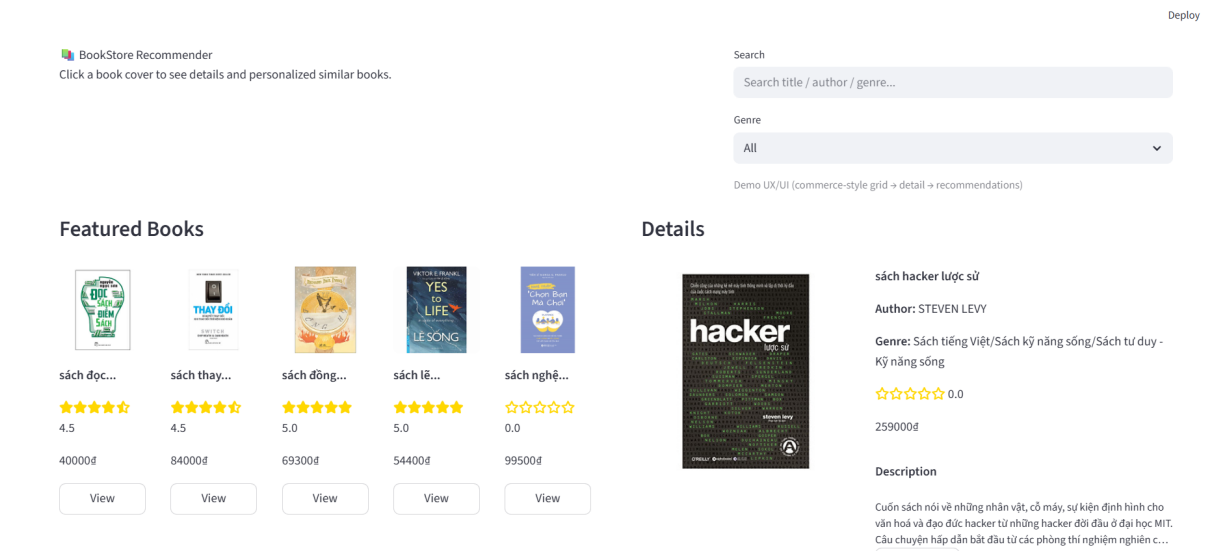
Số lượng mẫu sách vẫn còn khá hạn chế so với ứng dụng thực tiễn, điều này dẫn đến ma trận user-item bị thưa khiến cho các phương pháp tiếp cận theo hướng này còn hạn chế.

8 Kết luận

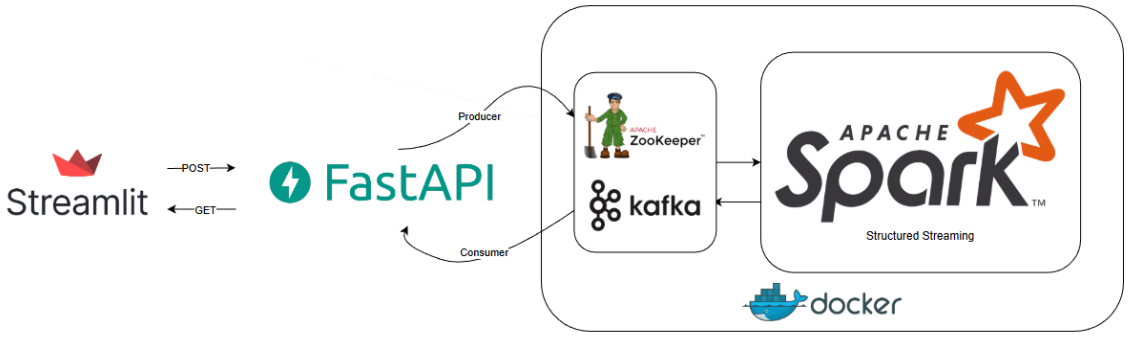
Mô hình khuyến nghị đa phương thức giải quyết được vấn đề dữ liệu bị thưa, là khuyết điểm của các mô hình khuyến nghị truyền thống như lọc cộng tác và khuyến nghị dựa trên nội dung. Mặc dù chúng tôi đã sử dụng embedding tiên tiến - VisoBERT cho phương pháp truyền thống, tuy nhiên các backbone của các mô hình đa phương thức tận dụng tốt hơn dẫn đến việc cho ra kết quả vượt trội. Bên cạnh đó, các phương pháp truyền thống chỉ tính toán tương đồng cục bộ giữa hai thực thể. Ngược lại, kiến trúc dựa trên đồ thị như MMGCN, LATTICE cho phép thông tin từ một người dùng hoặc sản phẩm lan truyền qua nhiều lớp, giúp tìm ra những mối liên hệ ẩn mà các phương pháp truyền thống không thể phát hiện.

Chúng tôi cũng thành công trong việc triển khai mô hình khuyến nghị tiên tiến trên bối cảnh dữ liệu lớn, tận dụng các công nghệ hiện đại nhằm có thể là luồng triển khai tham khảo khi thực hiện trên bối cảnh dữ liệu lớn thực tế.

Cuối cùng, điều cần nhấn mạnh là hiệu năng của các mô hình khuyến nghị đa phương thức phần lớn dựa vào các Encoder mà chúng ta lựa chọn. Cho nên, tùy vào mục đích khuyến nghị chúng ta có thể sử dụng các Encoder phù hợp như huấn luyện sẵn hoặc fine-tune theo nhu cầu của bài toán.



Hình 3: Giao diện hệ thống ứng dụng khuyến nghị sách



Hình 4: Kiến trúc pipeline xử lý dữ liệu và tích hợp mô hình khuyến nghị thời gian thực.

9 Lời cảm ơn

A Phụ lục

References

Yashar Deldjoo, Markus Schedl, Paolo Cremonesi, and Gabriella Pasi. 2020. Recommender systems leveraging multimedia content. *ACM Computing Surveys*, 53(5).

Qingyu Guo, Fuzhen Zhuang, Chuan Qin, and et al. 2022. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering*.

Ruining He and Julian McAuley. 2016. Vbpr: Visual-aware personalized ranking from implicit feedback. In *Proc. of AAAI*.

Xiangnan He and et al. 2017. Neural collaborative filtering. In *Proc. of WWW*.

Pasquale Lops, Marco de Gemmis, and Giovanni Semeraro. 2011. Content-based recommender systems: State of the art and trends. In *Recommender Systems Handbook*.

Julian McAuley and et al. 2015. Image-based recommendations on styles and substitutes. In *Proc. of SIGIR*.

Giuseppe Spillo and et al. 2025. See the movie, hear the song, read the book: Extending movielens-1m, last.fm-2k, and dbbook with multimodal data. In *Proc. of RecSys '25*.

Xiaoyuan Su and Taghi M. Khoshgoftaar. 2009. A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009.

Mengting Wan and Julian McAuley. 2018. Item recommendation on monotone implicit feedback. In *Proc. of RecSys*.

Bảng 4: Bảng thông tin sách

Thuộc tính	**Mô tả**
product_id	ID của sách
product_name	Tên của sách
authors	Tên tác giả
price	Giá bán của sách
seller_id	ID của người bán sách
seller_type	Phân loại người bán sách (OFFICIAL_STORE, TRUSTED_STORE)
rating_average	Điểm đánh giá trung bình của sách
review_count	Số lượng đánh giá của sách
order_count	Số lượng sách đã bán
url	Liên kết sản phẩm
image	Liên kết ảnh bìa sản phẩm
description	Mô tả nội dung sách
type_book	Thể loại sách
product_index	Cấp phát động tăng dần của ID sách

Bảng 5: Bảng tương tác người dùng với sách

Thuộc tính	**Mô tả**
customer_id	ID của người dùng
product_id	ID của sách được tương tác
rating	Điểm đánh giá của người dùng dành cho sách
content	Nội dung bình luận của người dùng
customer_index	Cấp phát động tăng dần của ID người dùng
product_index	Cấp phát động tăng dần của ID sách

Yinwei Wei, Xiang Wang, Liqiang Nie, and et al. 2024. Multi-modal recommender systems: A survey. *ACM Computing Surveys*.

Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2019. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1):1–38.

Xin Zhou. 2023. Mmrec: Simplifying multimodal recommendation. *Proceedings of the 5th ACM International Conference on Multimedia in Asia Workshops*.

Cai-Nicolas Ziegler and et al. 2005. Improving recommendation lists through topic diversification. In *Proc. of WWW*.