



GIẢI THƯỞNG ĐỀ TÀI MÔN HỌC XUẤT SẮC
UEH500 2023

TÊN CÔNG TRÌNH DỰ THI

MSĐT: BIT16

UEH-LBR*: ỨNG DỤNG GỢI Ý SÁCH THEO NHU CẦU SINH VIÊN DỰA TRÊN GIẢI THUẬT DI TRUYỀN

*UEH LIBRARY BOOK RECOMMENDER

01. GIỚI THIỆU ĐỀ TÀI

ĐIỀU GÌ ĐANG XẢY RA VỚI THƯ VIỆN ĐIỆN TỬ THÔNG MINH CỦA UEH?

Sinh viên đang không quá quan tâm? – Hay do nền tảng này quá khó để tiếp cận?

Master's Theses

Giải pháp nâng cao chất lượng dịch vụ tại VUS Bình Dương

Author(s): Nguyễn Thị Hoài Thư Advisor: Assoc. Prof. Dr. Bùi Thanh Tráng
(2023)

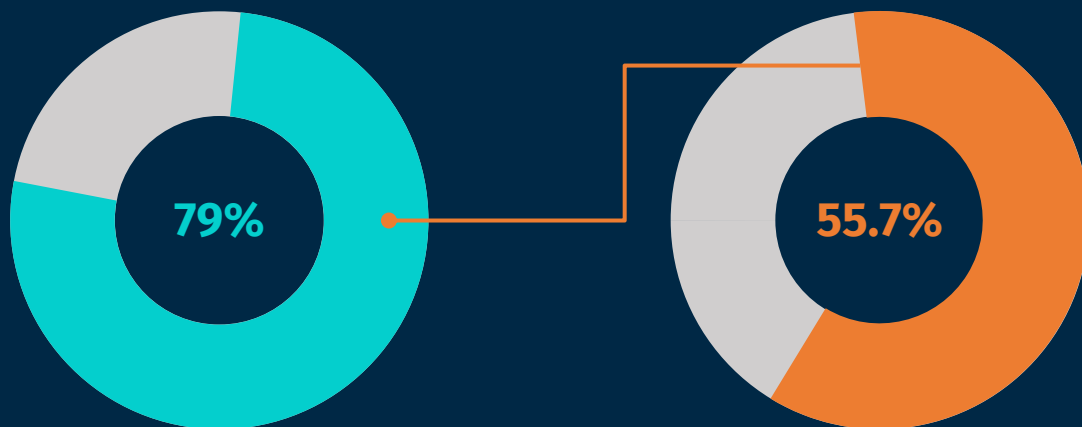
0

Văn bản pháp Luật cho môn Luật kinh doanh

Author(s): Võ Phước Long (2020)

35,848

MỨC ĐỘ QUAN TÂM VỚI SÁCH ĐIỆN TỬ



Mức quan tâm cao

Khó khăn
trong việc tiếp cận

MỨC ĐỘ QUAN TÂM VỀ TÍNH NĂNG ĐỀ XUẤT SÁCH



Mức quan tâm cao

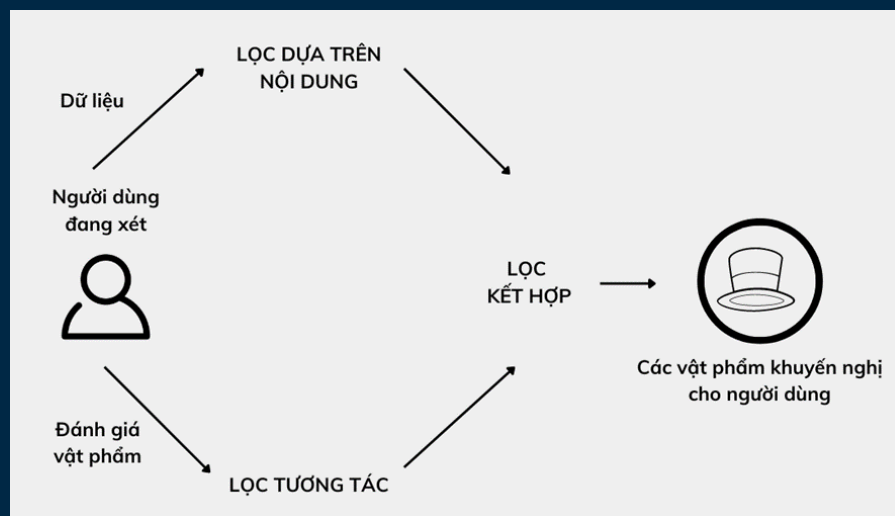
Mong muốn
bổ sung tính năng

HỆ THỐNG KHUYẾN NGHỊ

Hệ thống Khuyến nghị (Recommender System) sử dụng công nghệ AI để thực hiện phân tích và hiểu khối dữ liệu cá nhân, từ đó, đưa ra các dự đoán, gợi ý đề xuất phù hợp với sở thích của người dùng tại thời điểm bất kỳ trên các ứng dụng và nền tảng trực tuyến.

- Lọc Cộng tác (Collaborative Filtering)
- Lọc dựa trên Nội dung (Content-Based Filtering)

→ Nghiên cứu ứng dụng: Lọc Kết hợp (Hybrid Filtering)



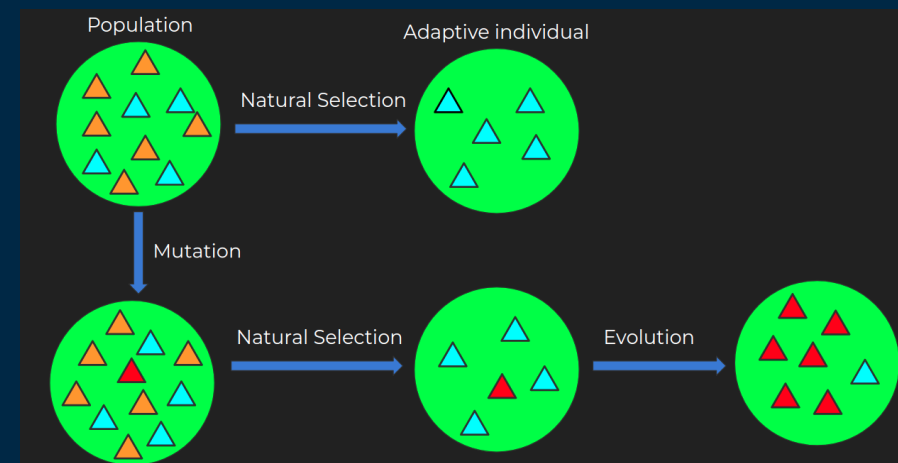
Minh họa Lọc Kết hợp

GIẢI THUẬT DI TRUYỀN

Giải thuật Di truyền (GA) là một phương pháp tìm kiếm tối ưu ngẫu nhiên dựa trên nguyên lý tiến hóa tự nhiên, bao gồm di truyền, đột biến, chọn lọc tự nhiên và trao đổi chéo.

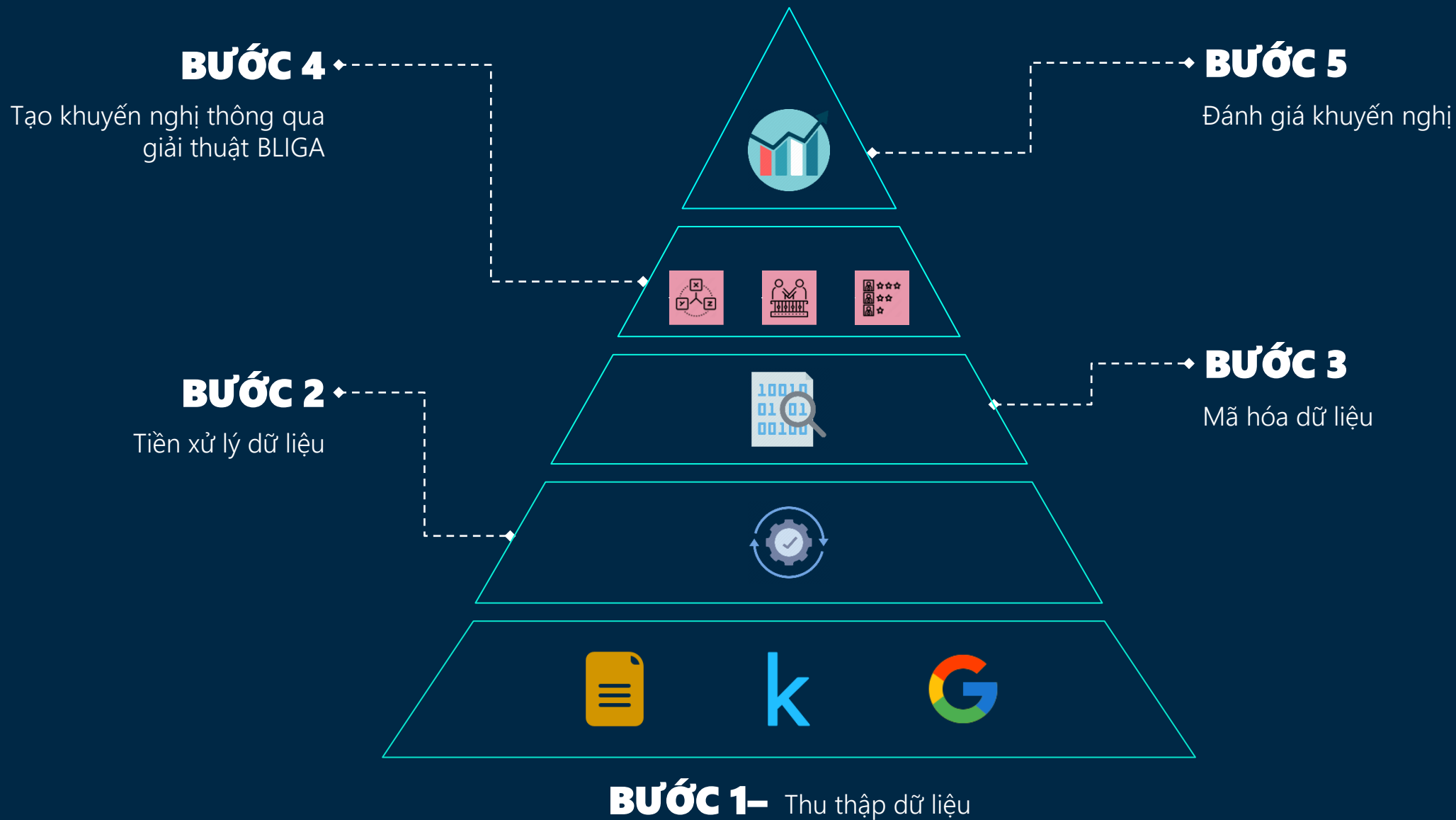
GA giải quyết các bài toán tối ưu hóa bằng cách sử dụng các quá trình cơ bản: lai tạo (Crossover), đột biến (Mutation) và chọn lọc (Selection) cho các cá thể trong quần thể.

Mô tả từng thành phần và mối liên hệ giữa những toán tử di truyền:



Minh họa các bước thực hiện giải thuật di truyền

02. MÔ HÌNH NGHIÊN CỨU





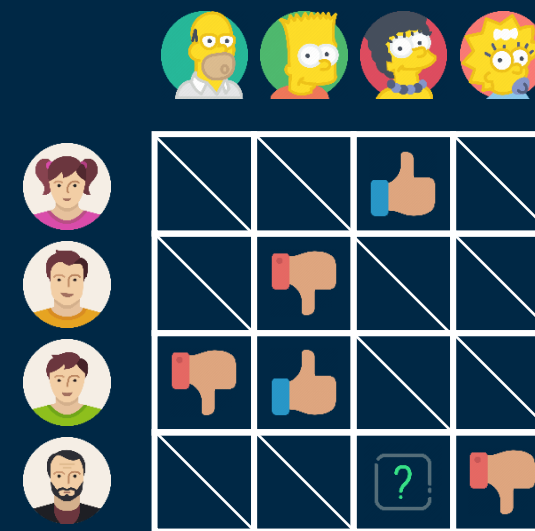
01. Cold Start

Khởi động nguội



02. Data Sparsity

Thưa thớt dữ liệu



BLIGA sử dụng các giải pháp **ĐA TIÊU CHÍ**, **THÔNG TIN NGŨ NGHĨA** và **DỮ LIỆU LỊCH SỬ**, giúp tạo ra các khuyến nghị mang tính chính xác cao ngay cả khi có ít thông tin về người dùng. Từ đó cho phép hệ thống đưa ra các đề xuất cho người dùng mới hoặc các mặt hàng mới.

03. PHƯƠNG PHÁP NGHIÊN CỨU

BLIGA TẠO KHUYẾN NGHỊ DỰA TRÊN BA TIÊU CHÍ LỘC



01. TƯƠNG QUAN CAO VỀ ĐẶC ĐIỂM NGŨ NGHĨA

Danh sách sách có độ tương quan cao:



02. HÀNG XÓM PHẢI HÀI LÒNG VỚI CÁC VẬT PHẨM THUỘC BLI

		5
		4
		3
		5
		3

		5

Người dùng có cùng sở thích như AU, phải hài lòng với một số vật phẩm đó:



03. HÀNG XÓM THUỘC VỀ AU CÓ ĐÁNH GIÁ DỰ ĐOÁN CAO

Sử dụng công thức tính tổng trọng số đã điều chỉnh của Resnick để dự đoán đánh giá của vật phẩm:

$$P_{AU,i} = \overline{r_{AU}} + \frac{\sum_{u \in U+} (r_i^u - \overline{r_u}) \times P_{sim}(AU,u)}{\sum_{u \in U+} P_{sim}(AU,u)}$$

		5
		4

		4
		3

		3
		5

Tiếp đến, BLIGA sẽ tính tổng dự đoán đánh giá của danh sách gợi ý đang xét và cho ra đề xuất có trọng số cao nhất:

$$predictSatRating(ind, AU) = \sum_{i \in ind} P_{AU,i}$$



DÂN SỐ = 10



TIÊU CHÍ DỪNG

5

1. KHỞI TẠO CÁ THỂ

1				5.52
2				1.72
3				5.54
...				
10				3.72

Sử dụng hệ số Jaccard để đo lường mối tương quan ngữ nghĩa giữa các vật phẩm:

$$SemSimI(p, q) = \frac{F_{11}}{F_{10} + F_{01} + F_{11}}$$

$$SemCorrRating(ind) = \sum_{p, q \in ind} SemSimI(p, q)$$

LỰA CHỌN

1			
3			
...			

LAI TẠO

3. ỨNG DỤNG TOÁN TỬ DI TRUYỀN

2. ỨNG DỤNG TIÊU CHÍ LỌC THỨ 1-
TÌM CÁ THỂ CÓ ĐỘ TƯƠNG QUAN NGỮ NGHĨA CAO

DÂN SỐ = 10



TIÊU CHÍ DỪNG

5

4. TIÊU CHÍ LỌC THỨ 2- HÀNG XÓM PHẢI HÀI LÒNG VỚI DANH SÁCH THUỘC BLI



XÉT

1

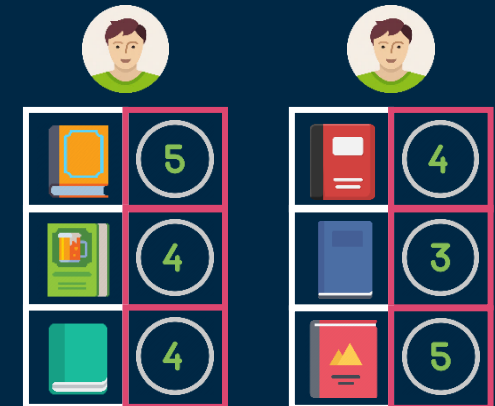
Lần chạy
đầu tiên

5

Số lần
cần phải chạy
(Tiêu chí dừng)

ĐÁP ỨNG ĐỦ

6. ÁP DỤNG TIÊU CHÍ LỌC THỨ 3- DỰ ĐOÁN ĐÁNH GIÁ:



bestMem ...

ĐỀ XUẤT

Sự dụng tích của phép đo tương quan
Pearson và Jaccard nhị phân để đo trọng số:

$$Psim(AU, u) = \frac{\sum_{i \in I} (r_i^{AU} - \bar{r}_{AU})(r_i^u - \bar{r}_u)}{\sqrt{\sum_{i \in I} (r_i^{AU} - \bar{r}_{AU})^2} \sqrt{\sum_{i \in I} (r_i^u - \bar{r}_u)^2}}$$

$$Jaccard(AU, u) = \frac{|\vec{V}_{AU} \cap \vec{V}_u|}{|\vec{V}_{AU} \cup \vec{V}_u|}$$

$$SatRating(ind) = \sum_{u \in U} Psim(AU, u) \times Jaccard(AU, u)$$



01. THÔNG TIN DỮ LIỆU ĐÁNH GIÁ

	Bộ dữ liệu Google	Bộ dữ liệu Kaggle	Bộ dữ liệu UEH
Số người dùng	100	250	49
Số sách	200	145	10
Số thể loại	348	145	10
Số đánh giá có	6636	11665	282
Thang đánh giá	1-5	1-5	1-5
Số đánh giá thiếu	13364	24585	208
Độ thưa thớt của dữ liệu	66.82%	67.82%	42.45%

VẤN ĐỀ

Quy mô dữ liệu còn khá nhỏ

Độ phân phối không đều

Độ thưa thớt của dữ liệu tương đối cao



02. KẾT QUẢ ĐÁNH GIÁ

Tiêu chí đánh giá	Bộ dữ liệu Google	Bộ dữ liệu Kaggle	Bộ dữ liệu UEH
Ngưỡng	3	3	3
MAE	0.2802	0.2763	0.3180
Precision	0.9090	0.8376	0.9770
Recall	0.8396	0.9189	0.8765
F1-Score	0.8729	0.8764	0.9240

MAE:

Đo lường trung bình chênh lệch tuyệt đối giữa các dự đoán và giá trị thực tế cho tất cả các vật phẩm được đánh giá bởi một người dùng

PRECISION:

Đo lường tỷ lệ các mục phù hợp được dự đoán chính xác (True Positive) trên tổng số các vật phẩm phù hợp được dự đoán

RECALL:

Đo lường tỷ lệ mục phù hợp được dự đoán chính xác (True Positive) trên tổng số các vật phẩm phù hợp thực tế

F1 – SCORE:

Trung bình điều hòa của Precision và Recall, cung cấp chỉ số duy nhất cân bằng giữa chúng

LỰA CHỌN HƯỚNG ỨNG DỤNG



Giải thuật:

BLIGA



Giải quyết:



HẠN CHẾ

Yêu cầu đối với dữ liệu đầu vào (dữ liệu huấn luyện):



Tính tương thích với bối cảnh Thư viện điện tử UEH:



HƯỚNG MỞ RỘNG ĐỀ TÀI

Tăng tính chính xác của kết quả dự đoán:

BIDIRECTIONAL ENCODER REPRESENTATIONS from TRANSFORMERS (BERT)
& **GIẢI THUẬT VỀ MẠNG NƠ-RON**

Giảm độ phức tạp:

STOCHASTIC GRADIENT DESCENT (SGD)

Khắc phục vấn đề không thể giải thích kết quả của BLIGA:

LOCAL INTERPRETABLE MODEL-AGNOSTIC EXPLANATIONS (LIME)
& **SHapley ADDITIVE exPLANATIONS (SHAP)**

Tăng tính chính xác cho dữ liệu đầu vào:

ÁP DỤNG CÔNG NGHỆ XỬ LÝ NNTN (NLP)