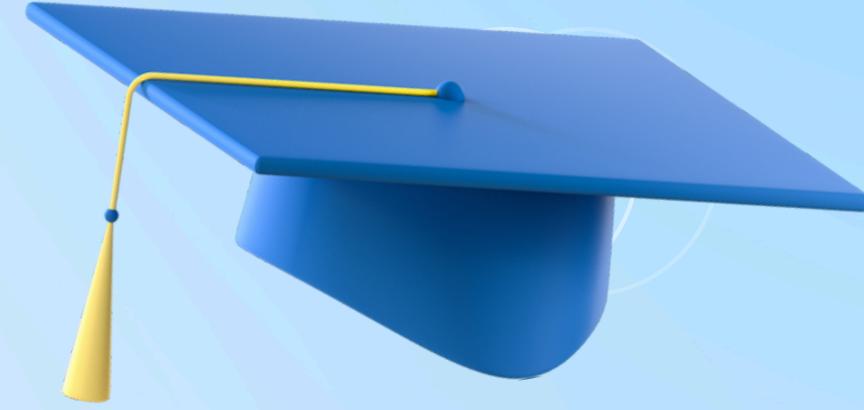




TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN



KHÓA LUẬN TỐT NGHIỆP

**ĐỀ TÀI: MỘT PHƯƠNG PHÁP CẢI TIẾN MÔ HÌNH
NHÂN TỐ ẨN CHO BÀI TOÁN TƯ VẤN NHÓM**

GV Hướng dẫn: Tiết Gia Hồng

Sinh viên thực hiện:

NGUYỄN LÊ DUY
LÊ PHAN DUY TÙNG



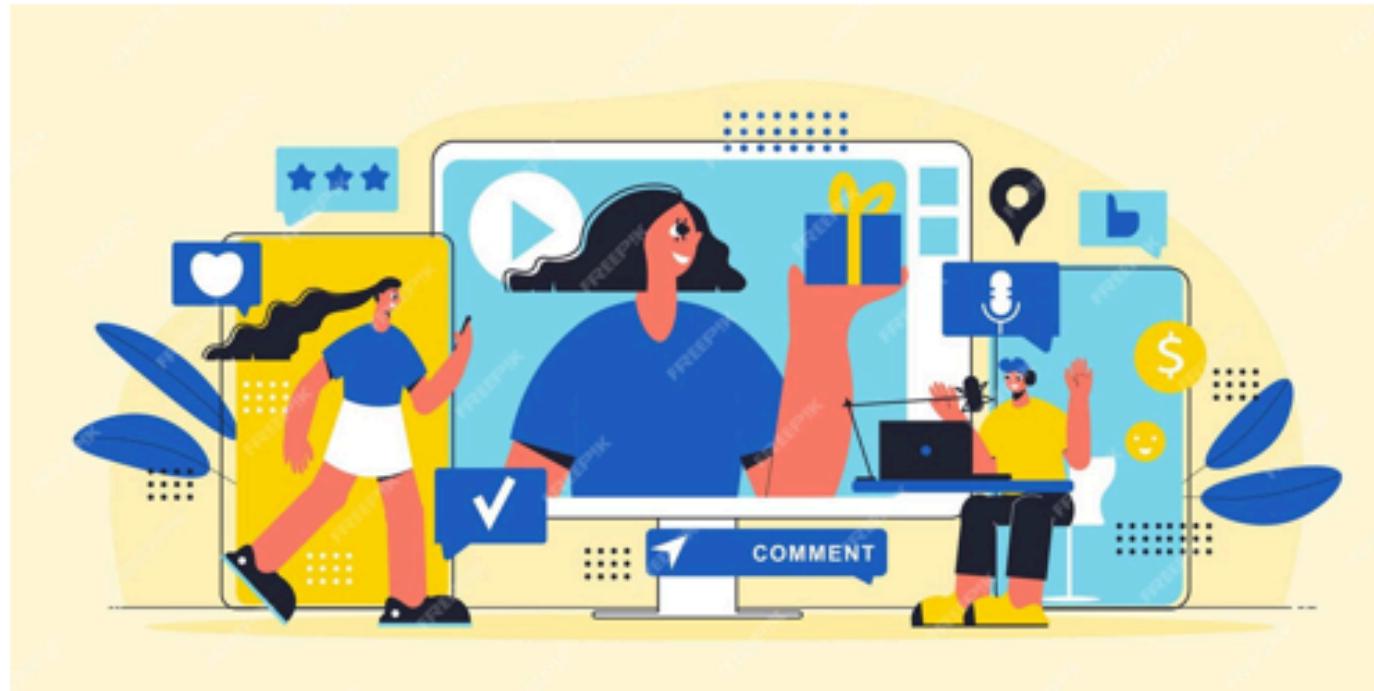
NỘI DUNG TRÌNH BÀY



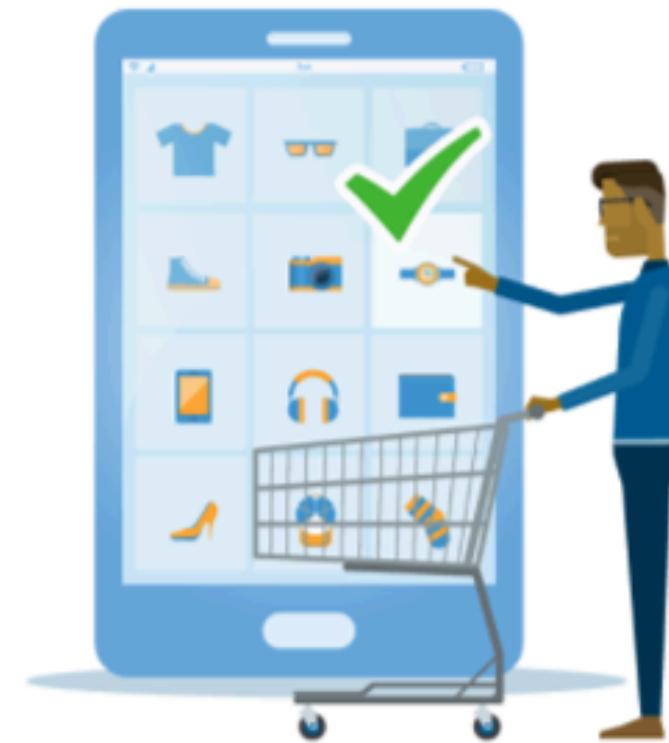
- Giới thiệu
- Các nghiên cứu liên quan
- Hướng tiếp cận
- Kết quả thực nghiệm
- Kết luận

GIỚI THIỆU

BÀI TOÁN TƯ VẤN



Người dùng bị quá tải thông tin



Giúp người dùng tìm thấy sản phẩm phù hợp nhanh chóng



Recommender System



GIỚI THIỆU

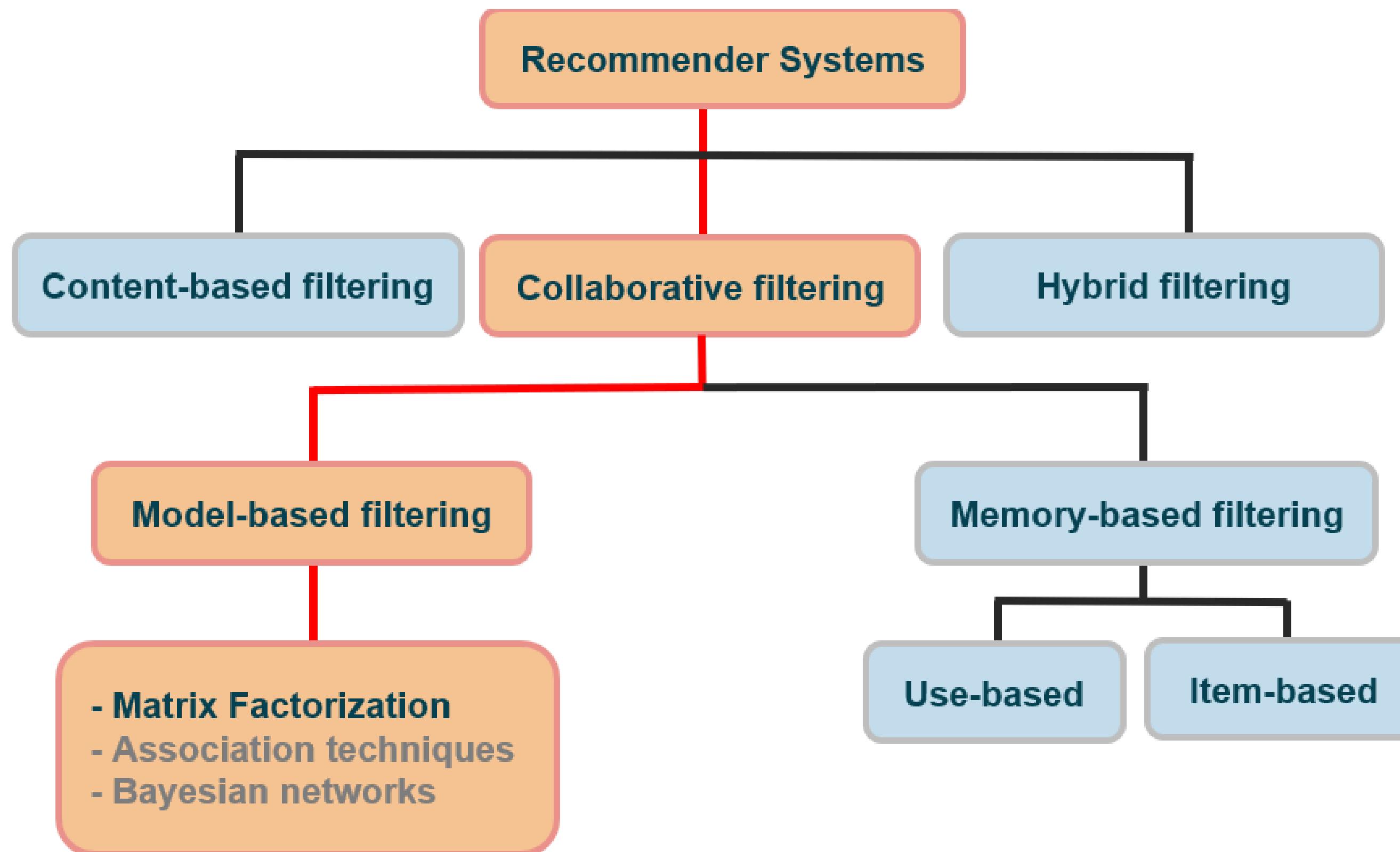
BÀI TOÁN TƯ VẤN

Ma trận rating user-item: $R[m \times n]$

	n sản phẩm i				
	i_1	i_2	...	i_n	
u_1	1	*	...	4	rating của các sản phẩm đã được trải nghiệm
u_2	3	5	...	*	các sản phẩm chưa được trải nghiệm
u_3	*	2	...	5	
...	
u_m	1	*	...	2	

⇒ Dự đoán sở thích của người dùng trên các sản phẩm chưa trải nghiệm

CÁC NGHIÊN CỨU LIÊN QUAN



CÁC NGHIÊN CỨU LIÊN QUAN

Mô hình nhân tố ẩn

Ma trận item-factor
V (s×n)

	i_1	i_2	i_3	i_4	i_5
f_1	-0.31	0.56	0.14	0.12	0.42
f_2	0.17	0.72	-0.46	0.83	-0.37

	f_1	f_2
u_1	0.95	-0.39
u_2	-0.52	0.18
u_3	0.13	0.4
u_4	0.36	-0.73

Ma trận user-factor **H (m×s)**

	i_1	i_2	i_3	i_4	i_5
u_1	4	*	3	5	1
u_2	5	*	*	4	*
u_3	2	*	1	3	*
u_4	*	*	3	*	2

Ma trận đánh giá user-item **R(m×n)**

	i_1	i_2	i_3	i_4	i_5
<i>Item-bias</i>	-0.32	0.46	0.13	-0.11	0.44

Bias sản phẩm p_i

	<i>User-bias</i>
u_1	0.16
u_2	-0.41
u_3	-0.52
u_4	0.79

Bias người dùng O_u

Tối ưu hàm mục tiêu:

$$\min_{H, V, o_u, p_i} \frac{1}{2} * \sum_{\{(u,i) | u=1..m \Lambda i=1..n \Lambda r_{u,i} \neq *\}} (r_{u,i} - \hat{r}_{u,i})^2$$

CÁC NGHIÊN CỨU LIÊN QUAN

Mô hình nhân tố ẩn

- Hàm mục tiêu

$$\min_{H, V, o_u, p_i} \frac{1}{2} * \sum_{\{(u,i) \mid u = 1 \dots m \wedge i = 1 \dots n \wedge r_{u,i} \neq *\}} (r_{u,i} - o_u - p_i - \mu - \mathbf{H}_{u,:} * \mathbf{V}_{i,:}^T)^2 + \frac{\lambda}{2} \|\mathbf{H}\|^2 + \frac{\lambda}{2} \|\mathbf{V}\|^2 + \frac{\lambda}{2} \sum_{u=1}^m {o_u}^2 + \frac{\lambda}{2} \sum_{i=1}^n {p_i}^2$$

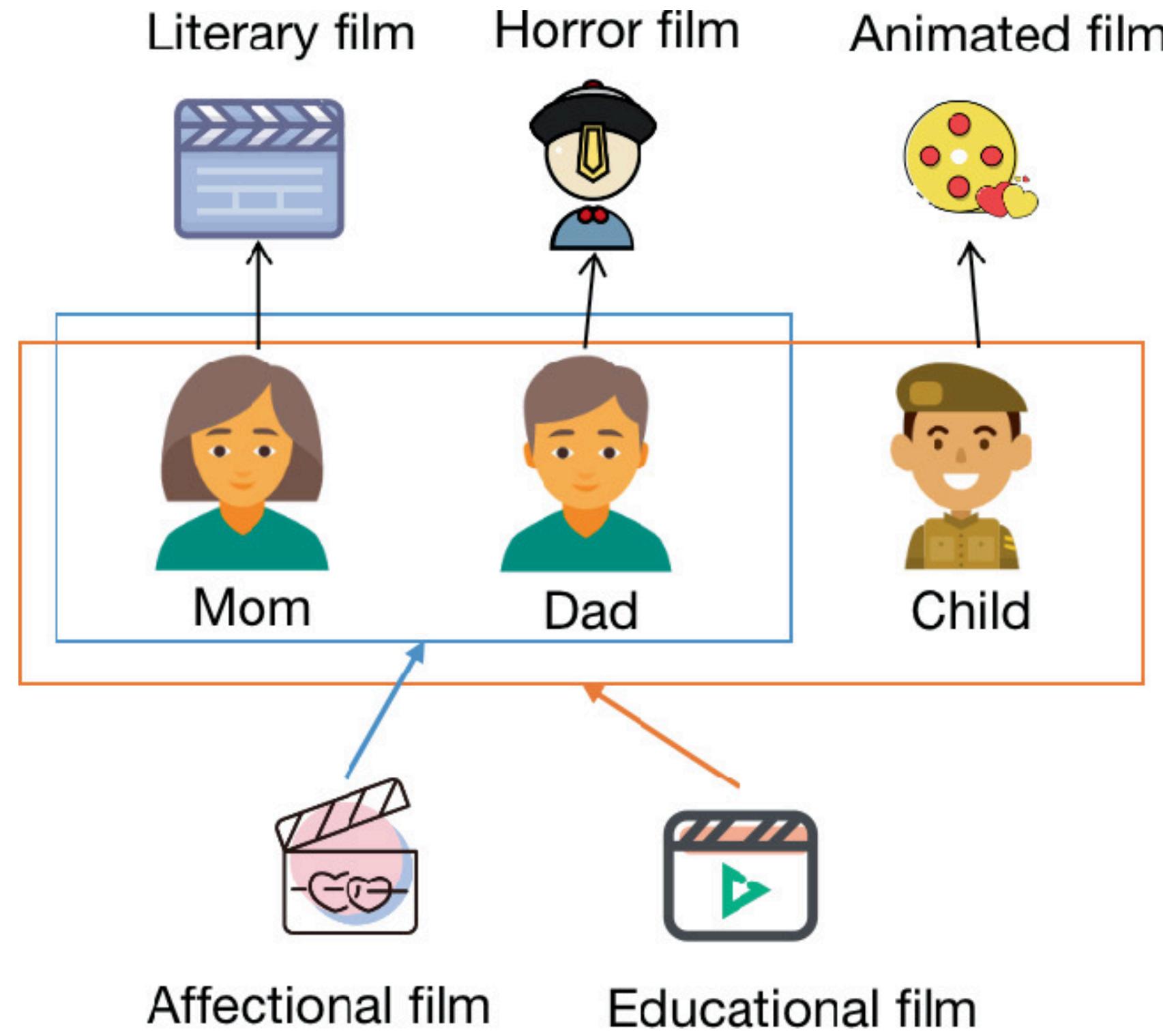
⇒ Sử dụng thuật toán Stochastic Gradient Descent (SGD)

- Dự đoán đánh giá khuyết: $\hat{r}_{u,i} \approx o_u + p_i + \mu + \mathbf{H}_{u,:} * \mathbf{V}_{i,:}^T$



CÁC NGHIÊN CỨU LIÊN QUAN

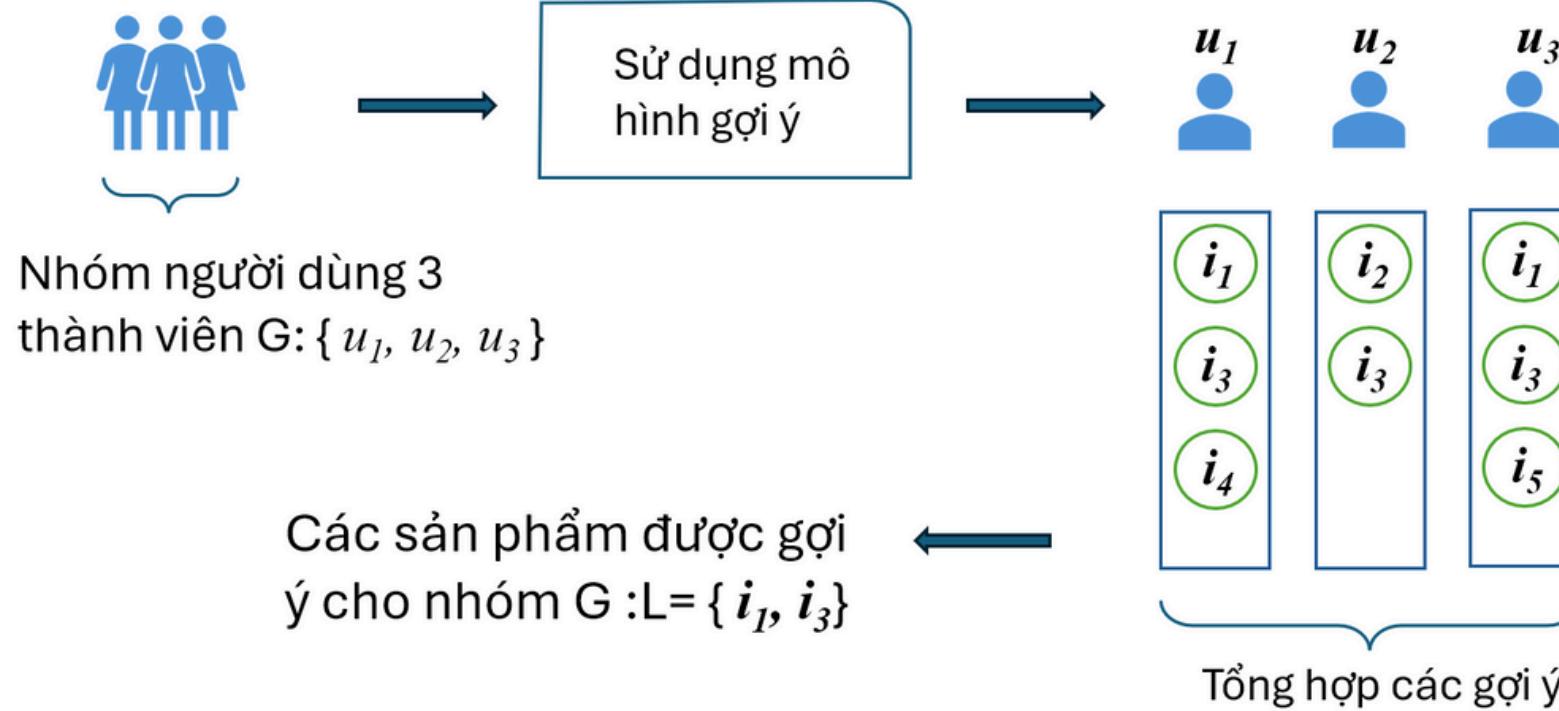
Xu hướng tư vấn nhóm



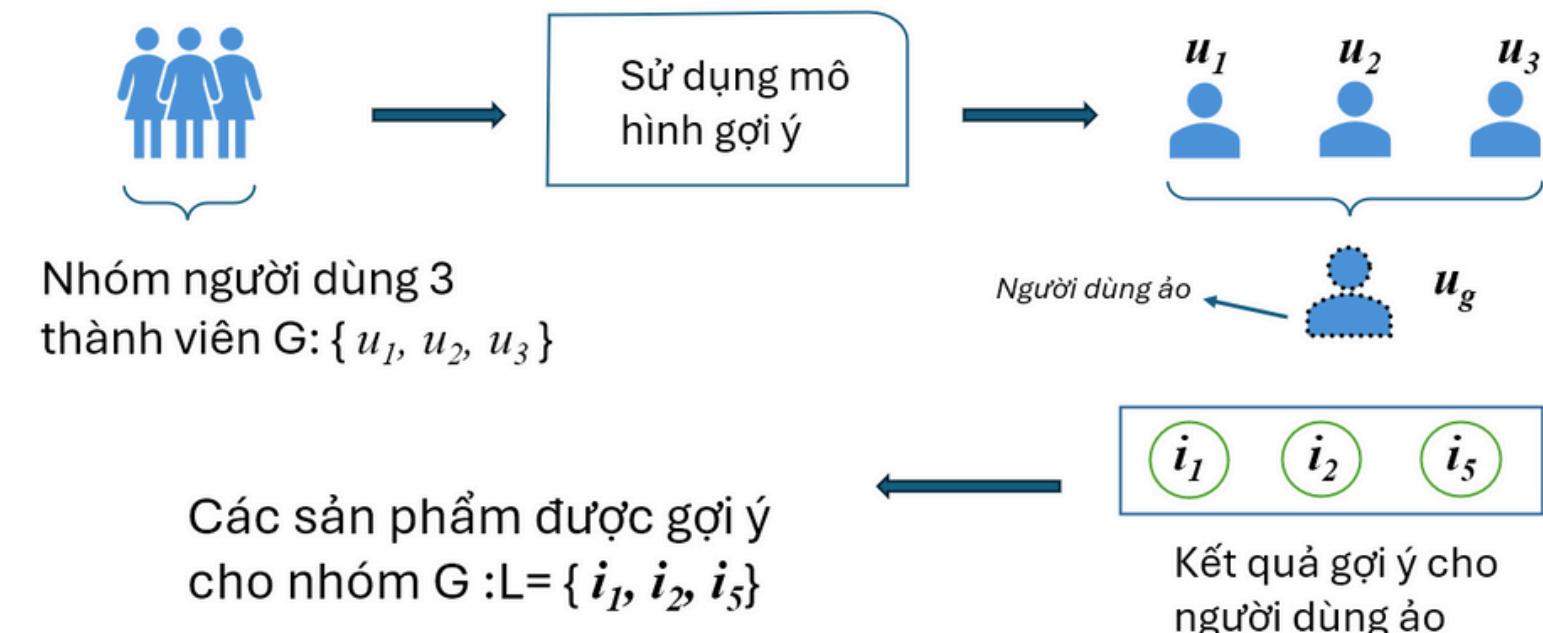
CÁC NGHIÊN CỨU LIÊN QUAN

Các chiến lược gợi ý nhóm

Tổng hợp gợi ý



Tổng hợp profile



CÁC NGHIÊN CỨU LIÊN QUAN



Các phương pháp gợi ý nhóm

	Rated by All group Members (RAM)	Rated by at Least one group Member (RLM)	Averaging Observed-Filled Ratings of All Members & detail Weights (AOFRAM&W)
Ý tưởng	tổng hợp các sản phẩm được đánh giá bởi tất cả thành viên nhóm	tổng hợp các sản phẩm được đánh giá bởi bất kì thành viên nhóm	Điền các đánh giá khuyết của người dùng trong nhóm trước khi tổng hợp
Ưu điểm	Đảm bảo sở thích tất cả thành viên	Tận dụng hết các đánh giá quan sát được	Cân bằng sở thích của tất cả thành viên, tận dụng các đánh giá quan sát và điền khuyết
Khuyết điểm	không tận dụng hết đánh giá quan sát được	không đảm bảo sở thích của các thành viên trong nhóm	Tầm ảnh hưởng của người dùng được xác định dựa trên số lượng đánh giá

CÁC NGHIÊN CỨU LIÊN QUAN



Savvy saver - Full review provided by  Staples

 **Savvy saver** - Apr 4, 2013

Pros: Cheap; Works well

Lovely purchase

Helpful?



assetydoc - Full review provided by  Staples

 **The best deal on pencils** - Oct 10, 2013

Pros: Pricewise the Best; Glides Smoothly; Comfortable Grip; Great Lead Capacity; Erases Easily

Truly, I've bought these pencils for years. They are the best quality for the best price

Helpful?

⇒ Tích hợp thêm text review để cải thiện độ chính xác của dự đoán

CÁC NGHIÊN CỨU LIÊN QUAN

Ưu điểm:

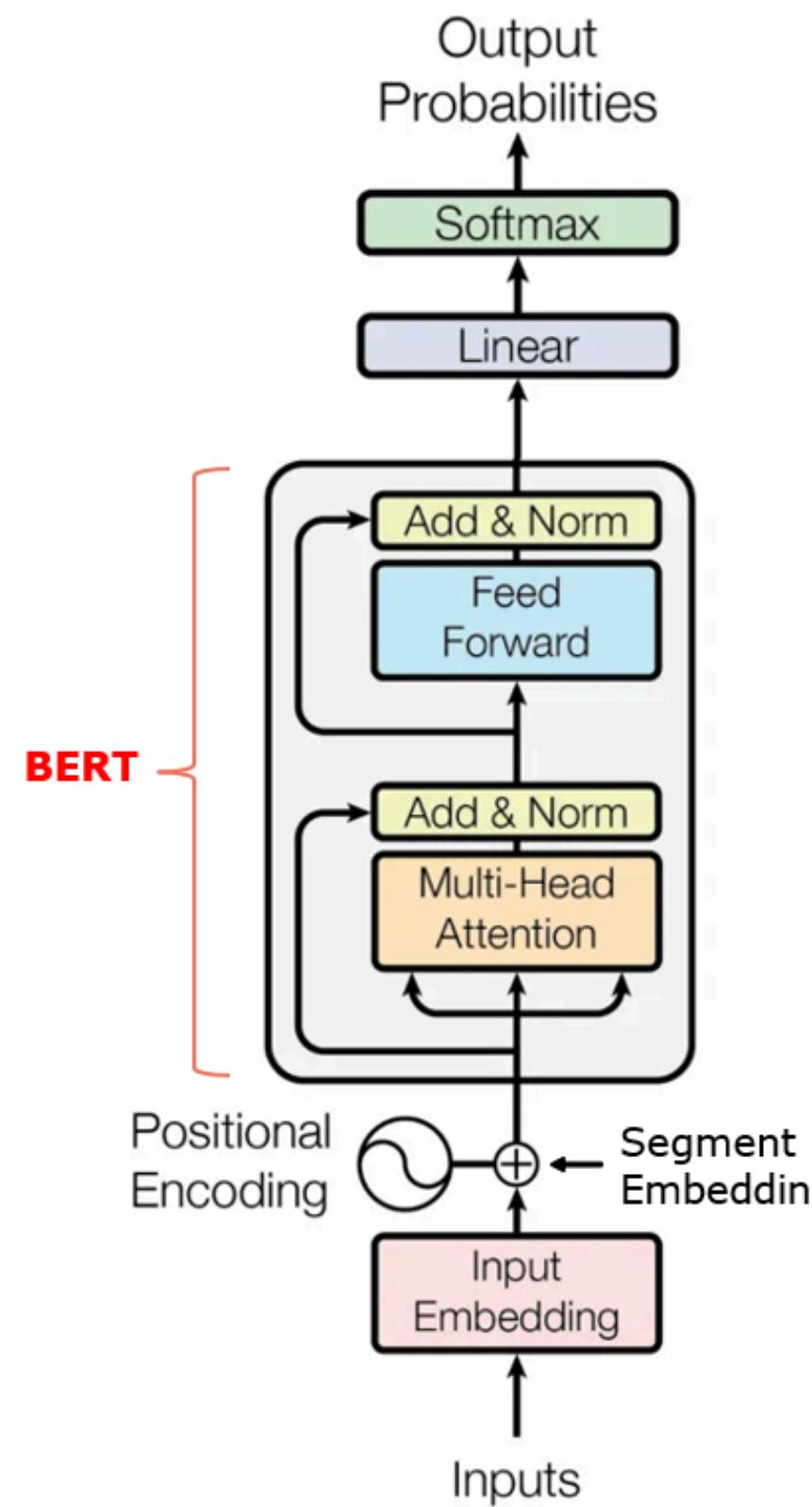
- Khám phá được sở thích của người dùng và xu hướng của nhóm người dùng
- Độ chính xác có thể được đảm bảo khi dữ liệu quan sát đủ lớn

Nhược điểm:

- Dễ dẫn đến sai sót nếu dữ liệu đầu vào không tốt

Mục tiêu: Tích hợp nguồn dữ liệu đánh giá bằng văn bản (text review) của người dùng để tăng độ chính xác cho ma trận đánh giá user-item

HƯỚNG TIẾP CẬN



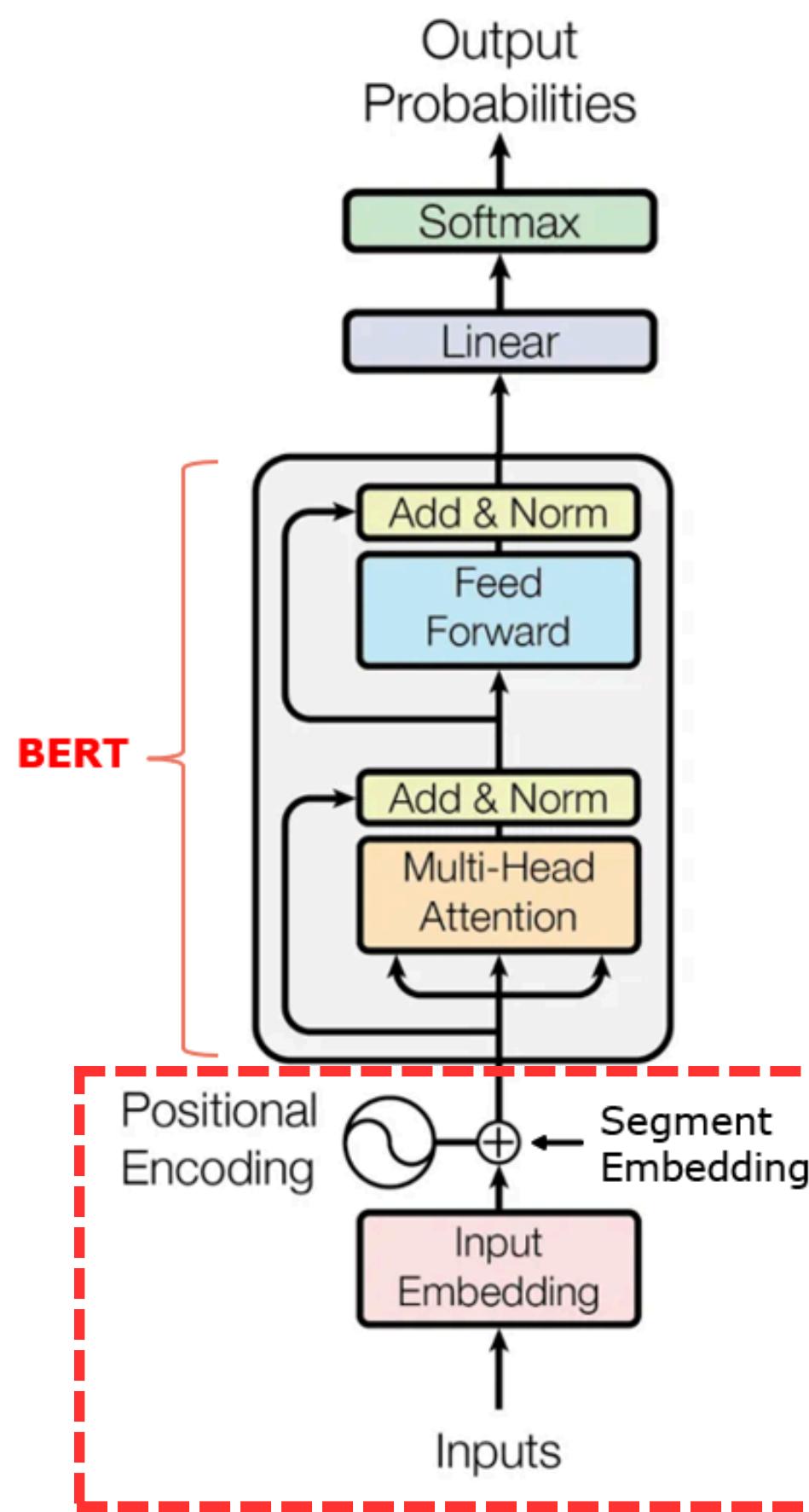
Tổng quan về mô hình BERT

- Bidirectional Encoder Representation from Transformer (BERT) dựa trên kiến trúc của mô hình Transformer
- Có 2 biến thể chính: BERT Base và BERT Large
- BERT sử dụng cơ chế self-attention hai chiều(bidirectional self-attention) để nắm bắt ngữ cảnh của từ dựa trên toàn bộ câu.



HƯỚNG TIẾP CẬN

Input cho mô hình BERT



Original sentence (tokens)	[CLS]	I	love	this	book	[SEP]
	3552.566	9980.851	6666.314	7512.261	5463.142	3571.487
	2745.925	8373.997	6239.623	8207.994	8669.221	9007.898

	1070.708	8752.749	4611.106	6827.572	9521.112	9664.859
	1652.976	4445.452	1937.651	3222.745	9338.361	1971.318
	+	+	+	+	+	+
Position Embedding (vector of size 768).	POS(0, 0)	POS(1, 0)	POS(2, 0)	POS(3, 0)	POS(4, 0)	POS(5, 0)
	POS(0, 1)	POS(1, 1)	POS(2, 1)	POS(3, 1)	POS(4, 1)	POS(5, 1)

	POS(0, 766)	POS(1, 766)	POS(2, 766)	POS(3, 766)	POS(4, 766)	POS(5, 766)
	POS(0, 767)	POS(1, 767)	POS(2, 767)	POS(3, 767)	POS(4, 767)	POS(5, 767)
	+	+	+	+	+	+
Segment Embedding (vector of size 768).	SEG(0, 0)					
	SEG(0, 1)					

	SEG(0, 766)					
	SEG(0, 767)					
	=	=	=	=	=	=
Encoder Input (vector of size 768)	420.386	7909.878	6167.866	7480.045	4497.961	3687.495
	4562.843	8386.358	1013.103	845.160	1034.689	7394.715

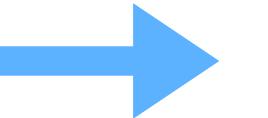
	7395.997	9878.506	2487.140	7411.603	5240.469	1362.285
	5830.822	6096.133	7675.256	1092.178	9843.646	40.205

HƯỚNG TIẾP CẬN

Single-head attention

[CLS]	420.386	4562.843	7395.997	5830.822
I	7909.878	8386.358	9878.506	6096.133
love	6167.866	1013.103	2487.140	7675.256
this	7480.045	845.160	7411.603	1092.178
book	4497.961	1034.689	5240.469	9843.646
[SEP]	3687.495	7394.715	1362.285	40.205

Input matrix X (6, 768)

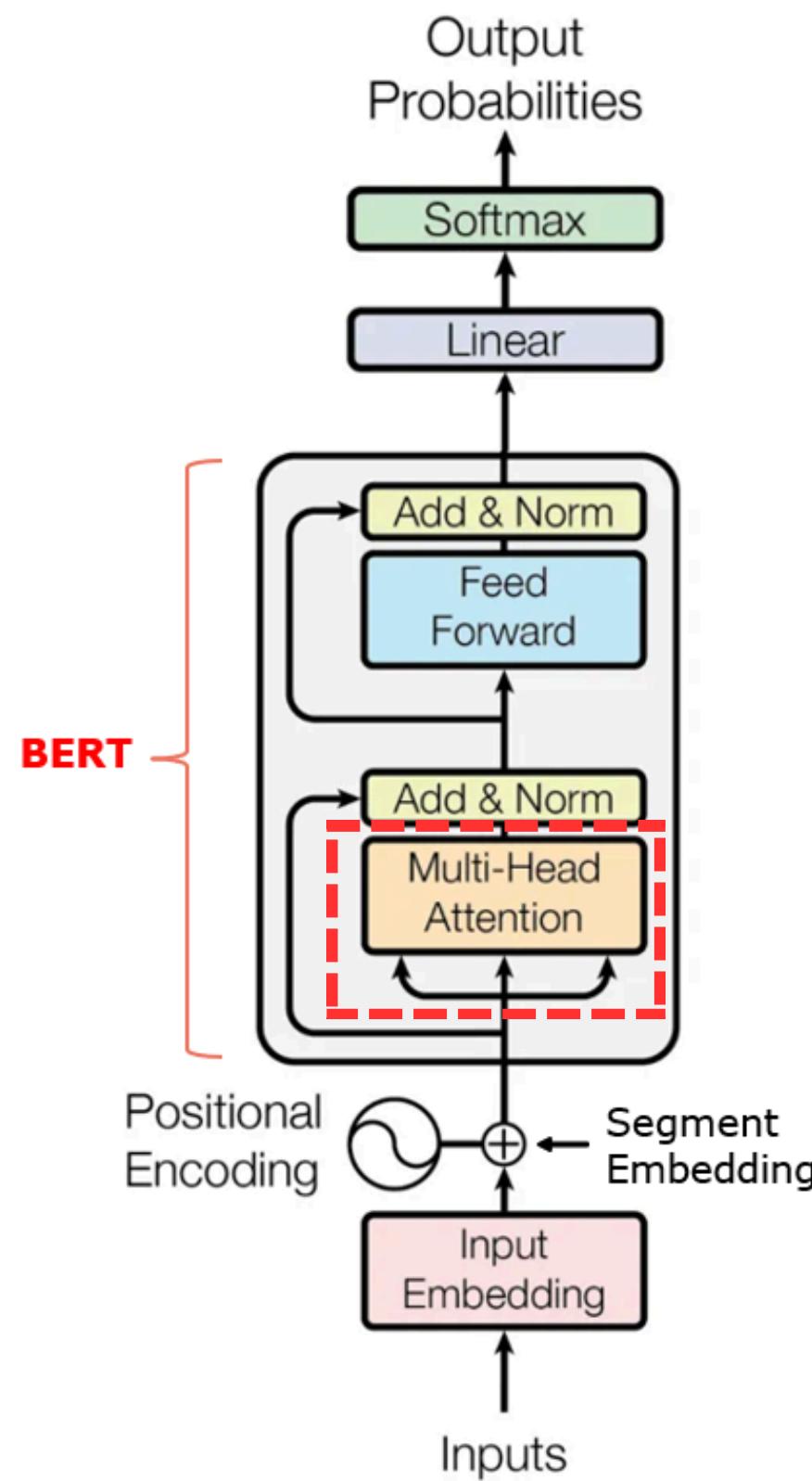


[CLS]	820.386	369.83	195.929	8930.2
I	8909.38	9388.484	4578.56	96.138
love	467.66	120.135	3447.140	5767.56
this	480.45	8885.49	741.809	1098.578
book	447.1	1834.304	728.559	6843.698
[SEP]	987.45	95.15	1965.85	640.25

Attention matrix Z₁ (6, 64)

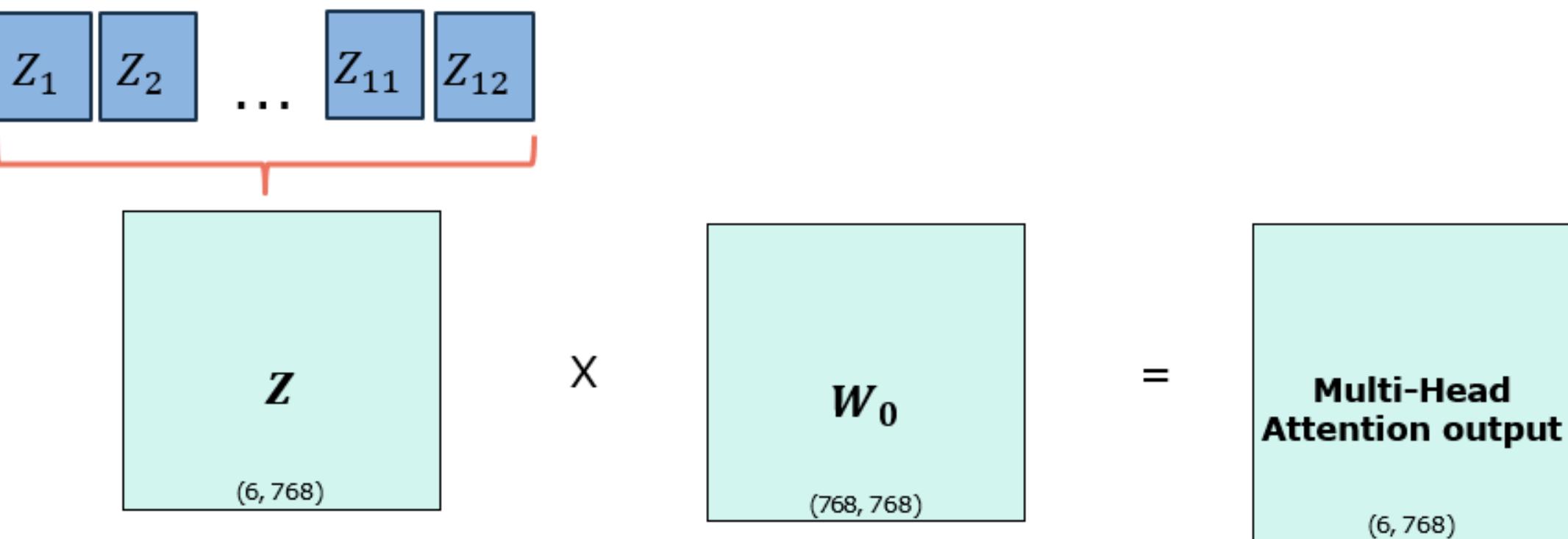
Each row of the attention matrix represents the embedding of the sequence:
It captures not only the meaning of each token, not only its position, but also
the interaction of each token with all the other tokens

HƯỚNG TIẾP CẬN

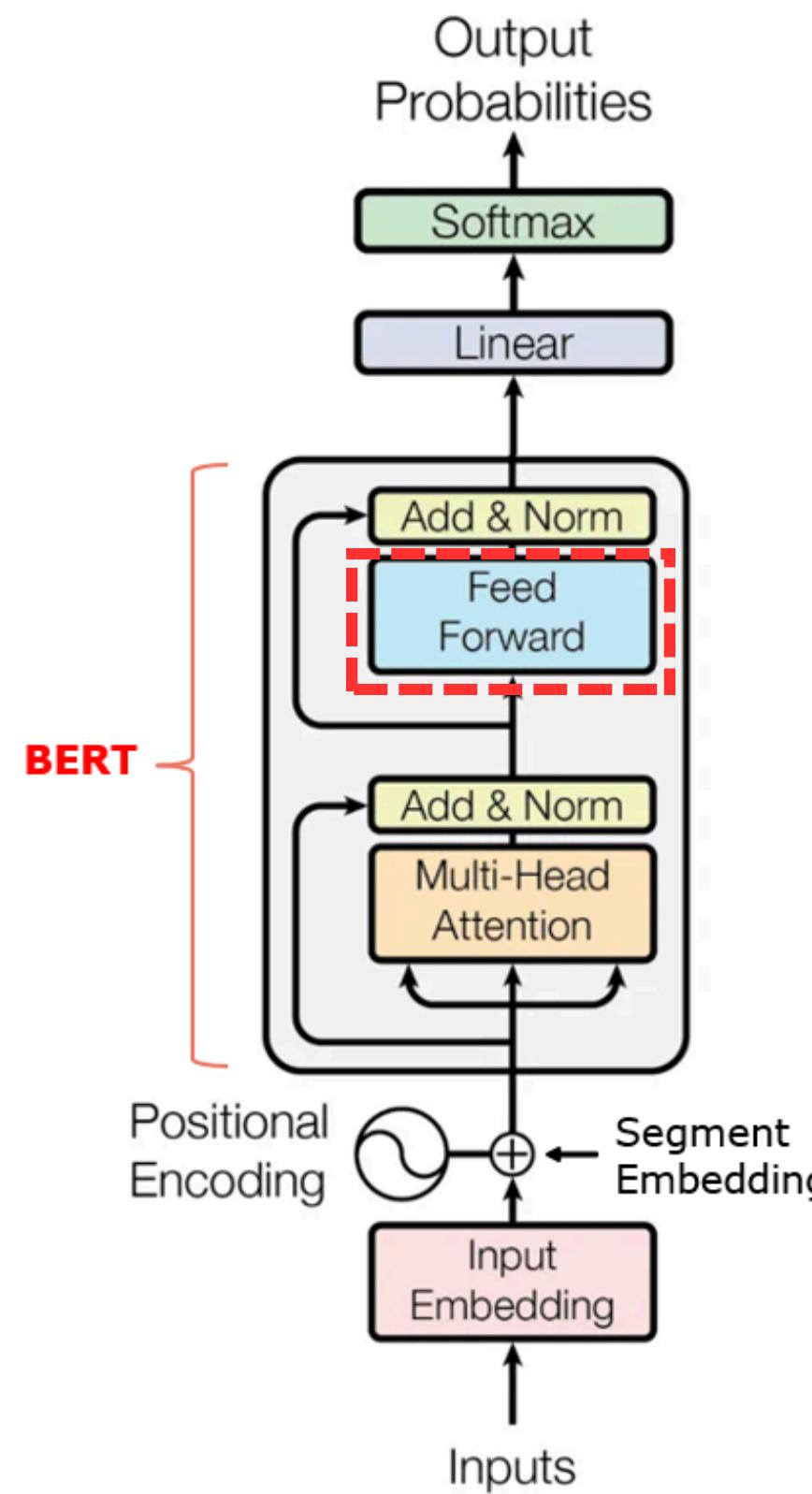


Multi-head attention

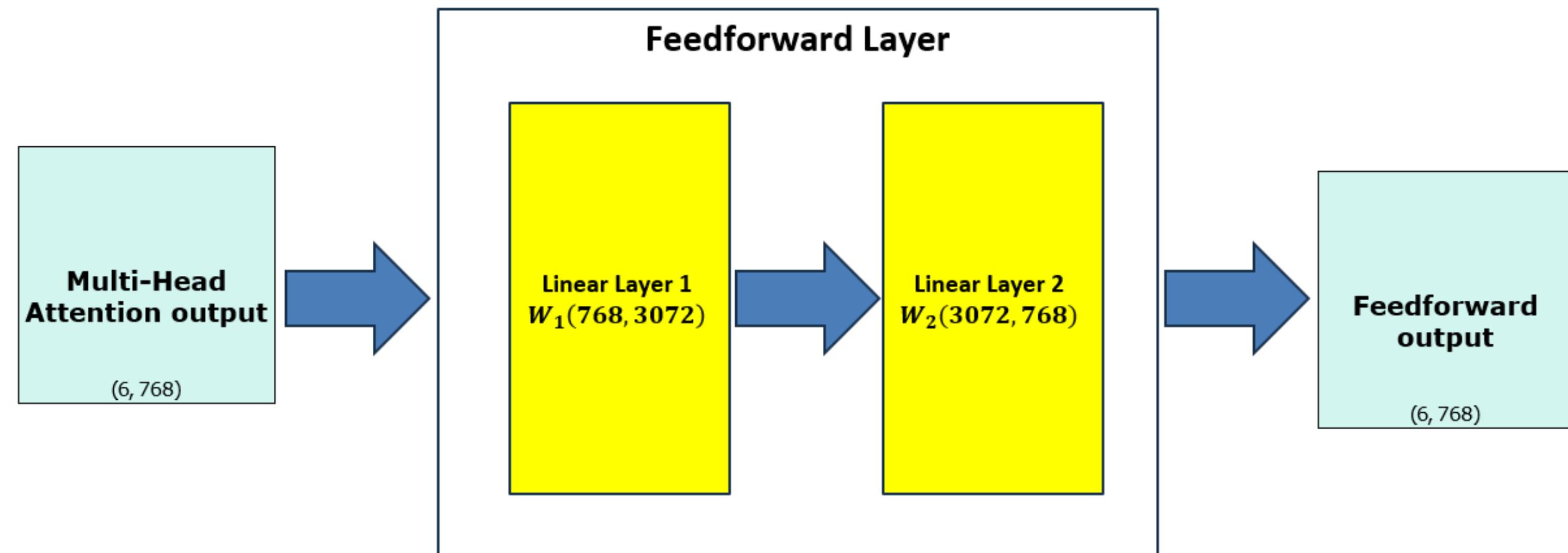
$$\text{Multi - head attention} = \text{Concatenate}(Z_1, Z_2, \dots, Z_{12})W_0$$



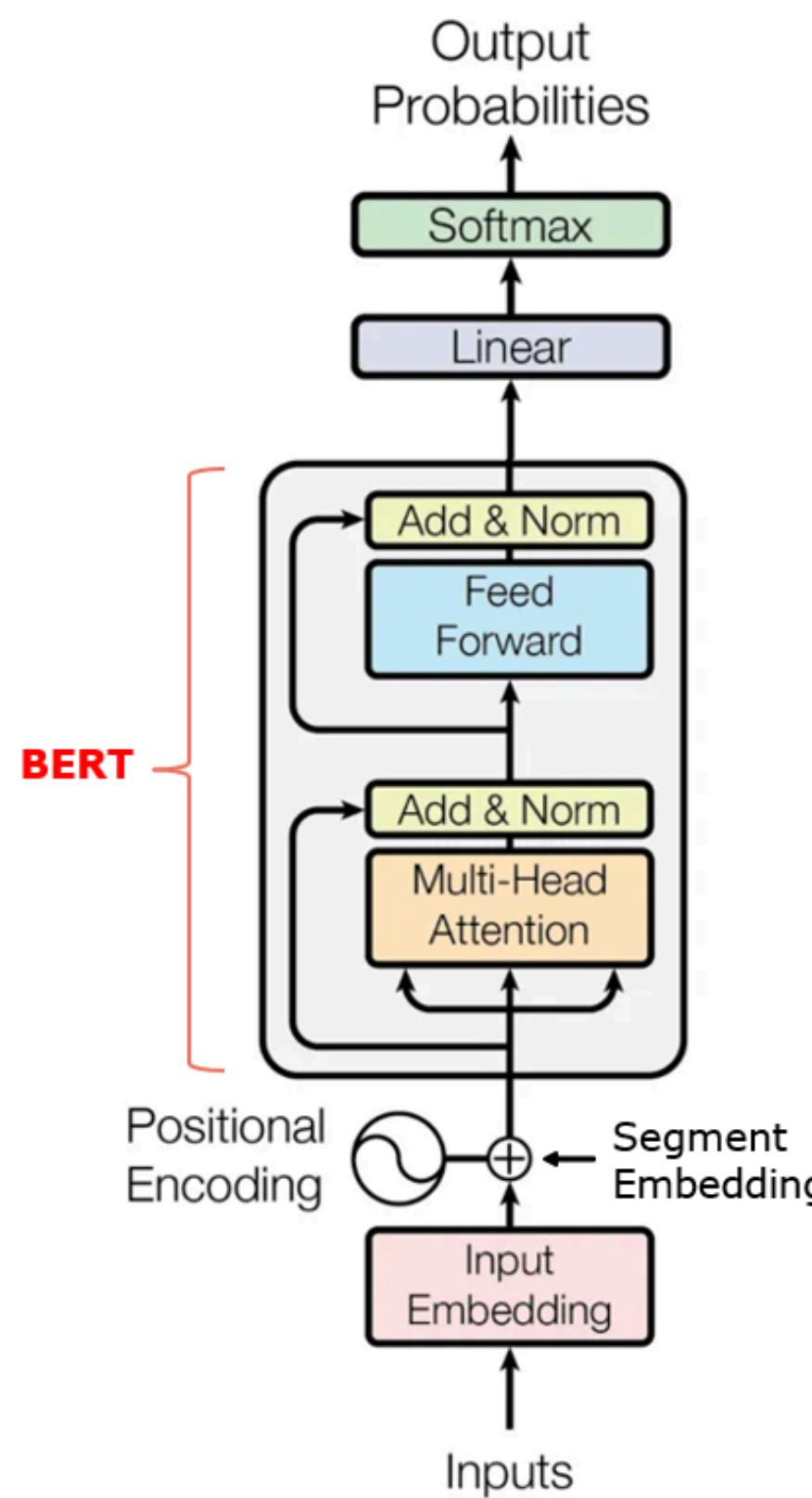
HƯỚNG TIẾP CẬN



Feedforward layer



HƯỚNG TIẾP CẬN



Output của mô hình BERT

Original sentence
(tokens)

[CLS] I love this book [SEP]

BERT

Encoder output
(vector of size 768)

789	909.88	167.866	74.045	497.961	367.4
843	86.3	13.103	45.16	10.689	734.75
...
997	987.5	487.140	741	52.469	162.285
830.822	96	775.256	1092	43.646	40.205

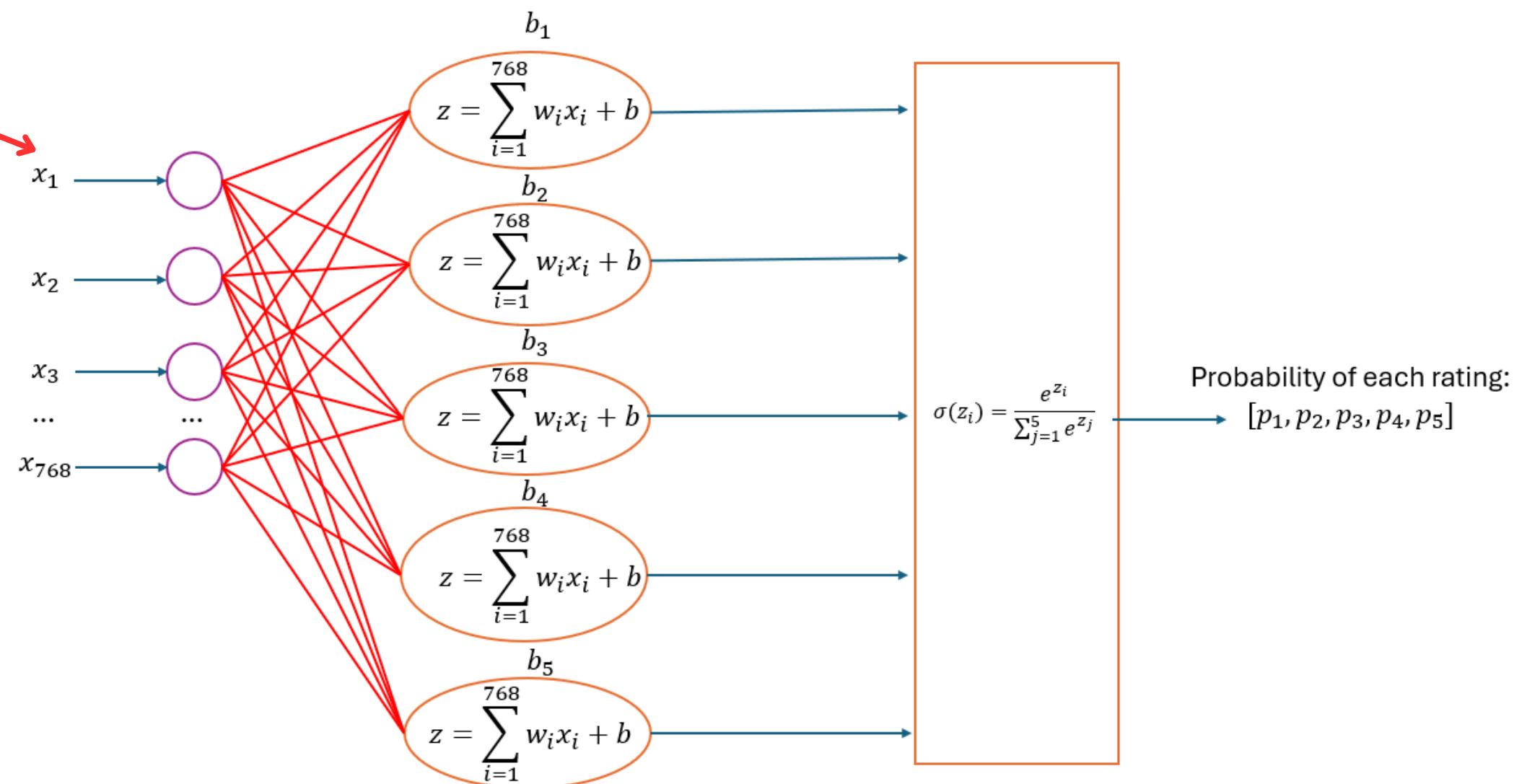
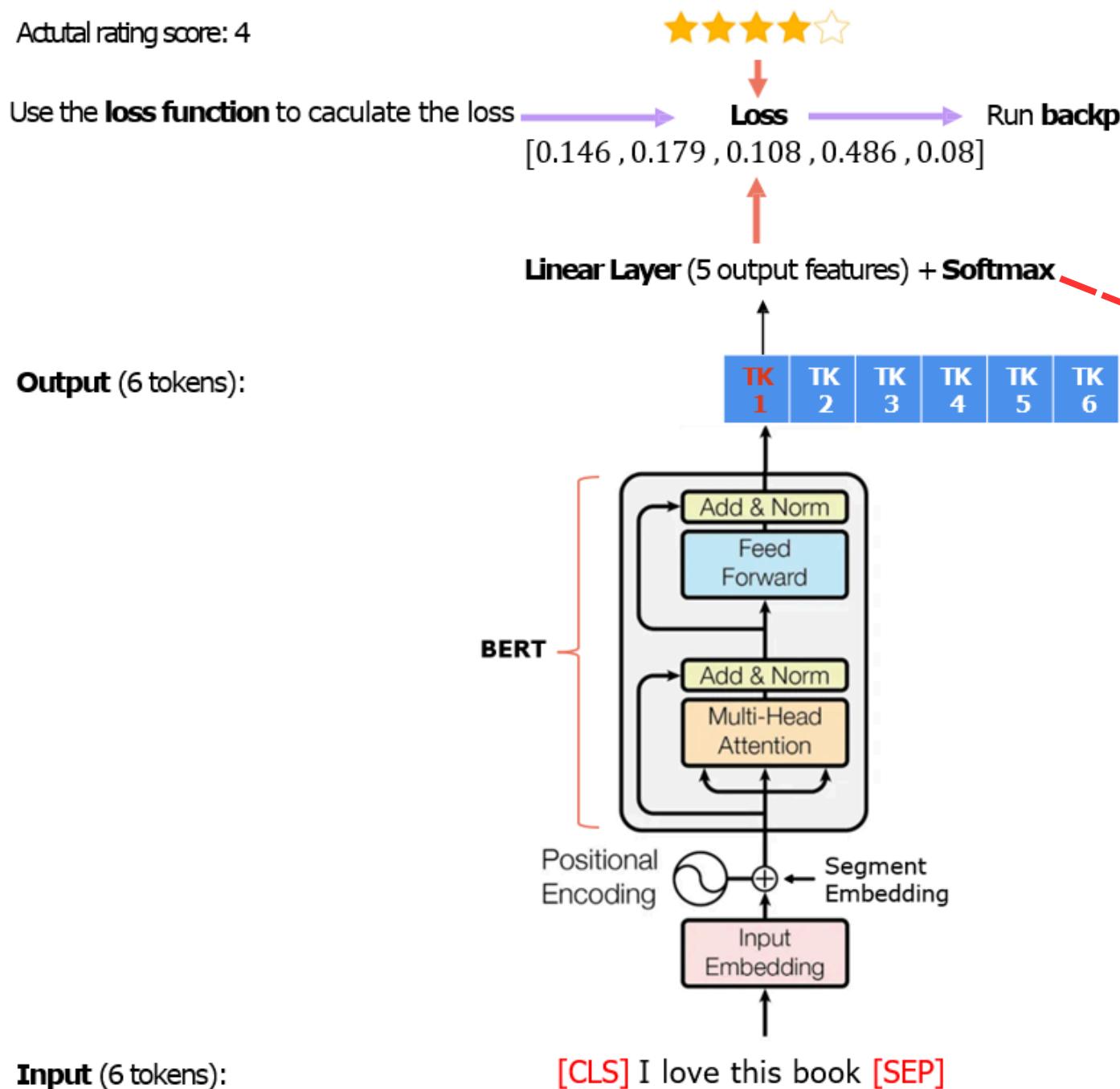


HƯỚNG TIẾP CẬN

Tinh chỉnh mô hình BERT cho bài toán phân loại văn bản

Actual rating score: 4

Use the **loss function** to calculate the loss → **Loss** → Run **backpropagation** to update the weights



Input (6 tokens):

[CLS] I love this book [SEP]

HƯỚNG TIẾP CẬN

Chiến lược tổng hợp profile AOFRAM&W

	i_1	i_2	i_3	i_4	i_5	Number of ratings (k_u)
u_1	4	*	3	5	*	3
u_2	5	*	*	*	*	1
u_3	2	*	1	*	*	2

$$c_{u,i} = \begin{cases} 1 - |r_{u,i} - o_u - p_i - \mu - \mathbf{H}_{u,:} * \mathbf{V}_{i,:}^T|^{[0,1]} & \text{if } r_{u,i} \neq * \\ 1 - |o_u + p_i + \mu - \mathbf{H}_{u,:} * \mathbf{V}_{i,:}^T|^{[0,1]} & \text{if } r_{u,i} = * \end{cases}$$



	i_1	$k_u^{[\beta,1]}$	$c_{u,i}^{[\beta,1]}$	$w_{u,i}$
u_1	4	0.7	1	0.7
u_2	5	0.4	0.65	0.26
u_3	2	1	0.4	0.4

	i_3	$k_u^{[\beta,1]}$	$c_{u,i}^{[\beta,1]}$	$w_{u,i}$
u_1	3	0.7	1	0.7
u_2	3.23	0.4	0.4	0.16
u_3	1	1	0.63	0.63

$$w_{u,i} = c_{u,i}^{[\beta,1]} * k_u^{[\beta,1]}$$

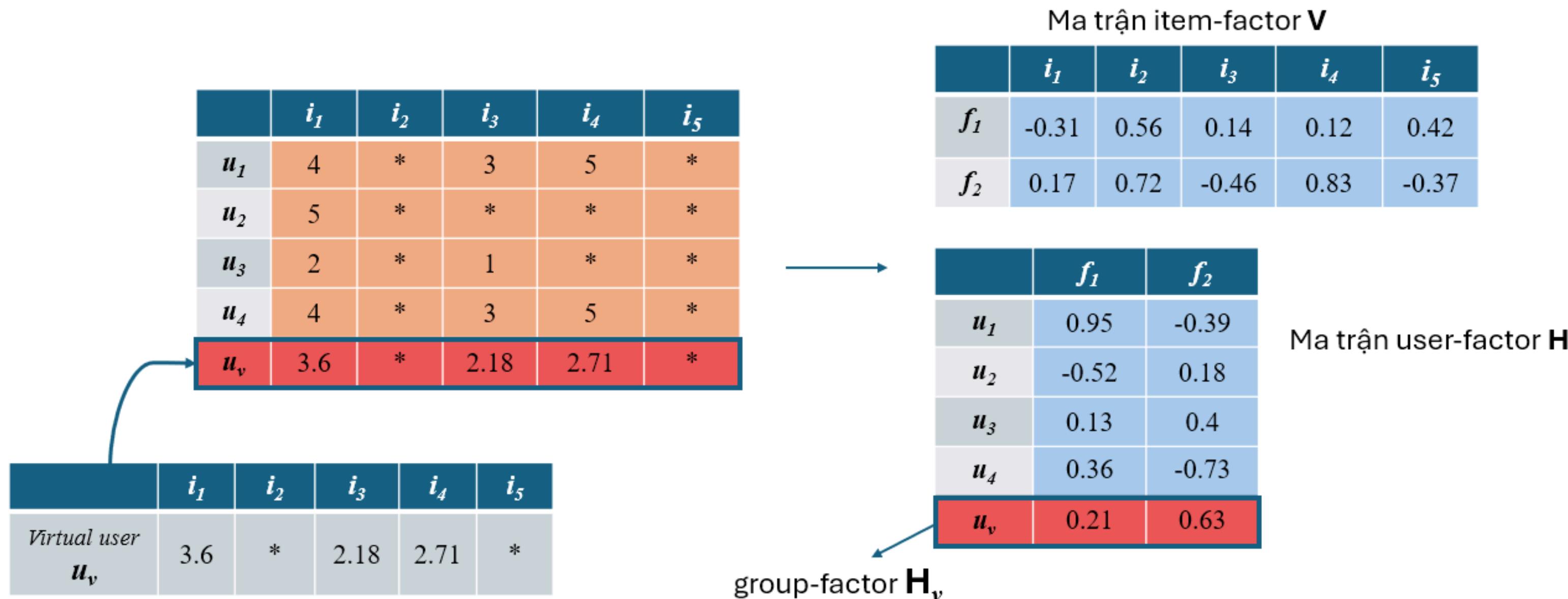


	i_1	i_2	i_3	i_4	i_5
$Virtual user v$	3.6	*	2.18	2.71	*

$$\frac{3 \times 0.7 + 3.23 \times 0.16 + 1 \times 0.63}{0.7 + 0.16 + 0.63} = 2.18$$

HƯỚNG TIẾP CẬN

Mô hình nhân tố ẩn cho bài toán tư vấn nhóm



Hàm mục tiêu:

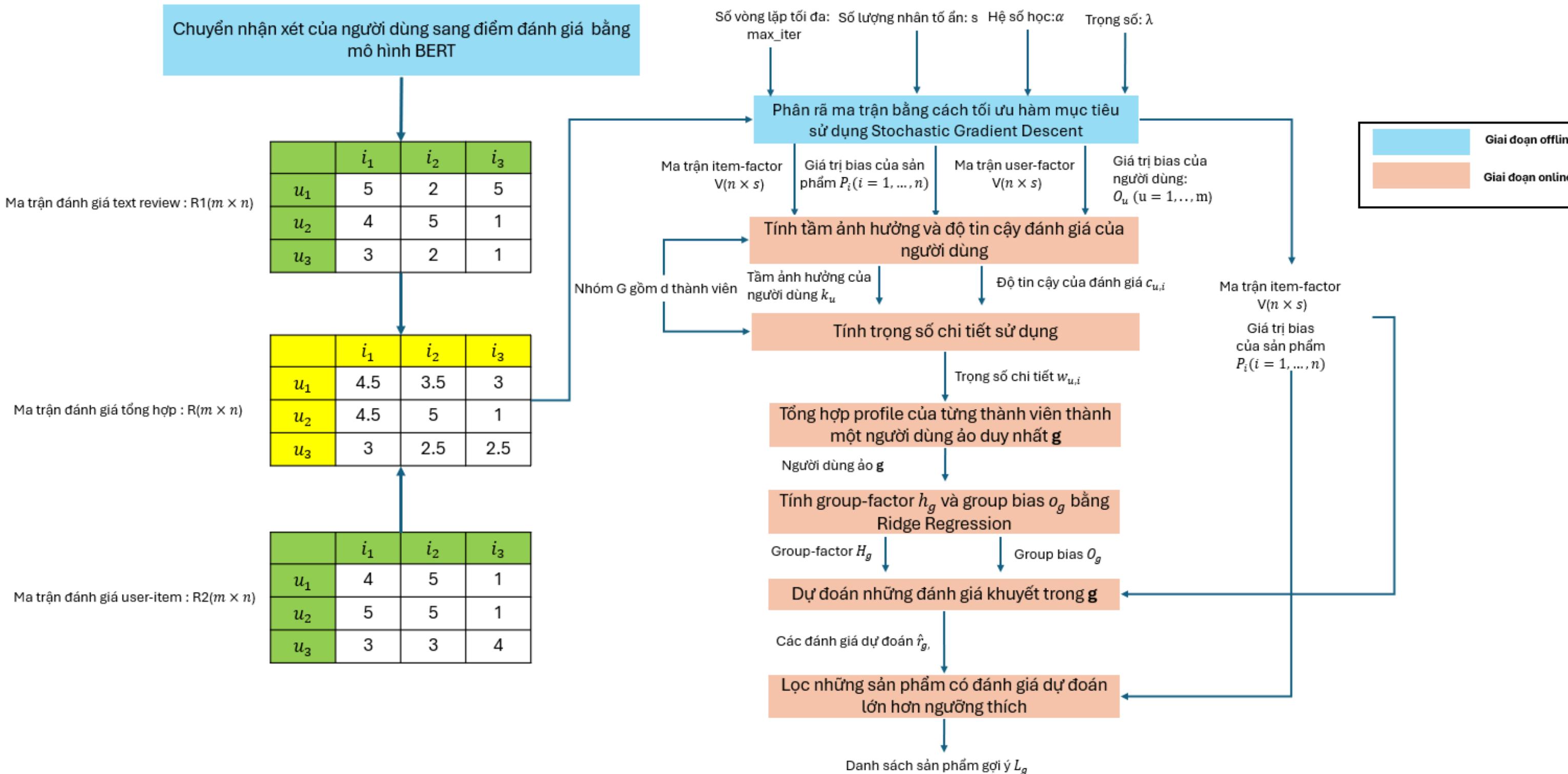
$$\min_{\mathbf{H}_v, o_v} \frac{1}{2} \times \sum_{\{i \mid i=1 \dots n \wedge r_{v,i} \neq *\}} (r_{v,i} - o_v - p_i - \mu - \mathbf{H}_{v,:} \times \mathbf{V}_{i,:}^T)^2 + \frac{\lambda}{2} \left\| \mathbf{H}_{v,:} \right\|^2 + \frac{\lambda}{2} o_v^2$$

Dự đoán:

$$r_{g,i} \approx o_g + p_i + \mu + \mathbf{h}_{g,:} * \mathbf{V}_{i,:}^T$$

HƯỚNG TIẾP CẬN

Mô hình được tích hợp nhận xét người dùng



THỰC NGHIỆM



Bộ dữ liệu thực nghiệm được trích xuất từ Amazon dataset 2018.

- **Digital Music** (phiên bản 5-core) chứa
 - 169.781 đánh giá có kèm theo nhận xét
 - 16.566 người dùng và 11.797 sản phẩm
- **Musical Instrument** (phiên bản 5-core)
 - 231.392 đánh giá có kèm theo nhận xét
 - 27.530 người dùng và 10.620 sản phẩm
- Mỗi người dùng, sản phẩm trong tập đều có ít nhất 5 đánh giá.

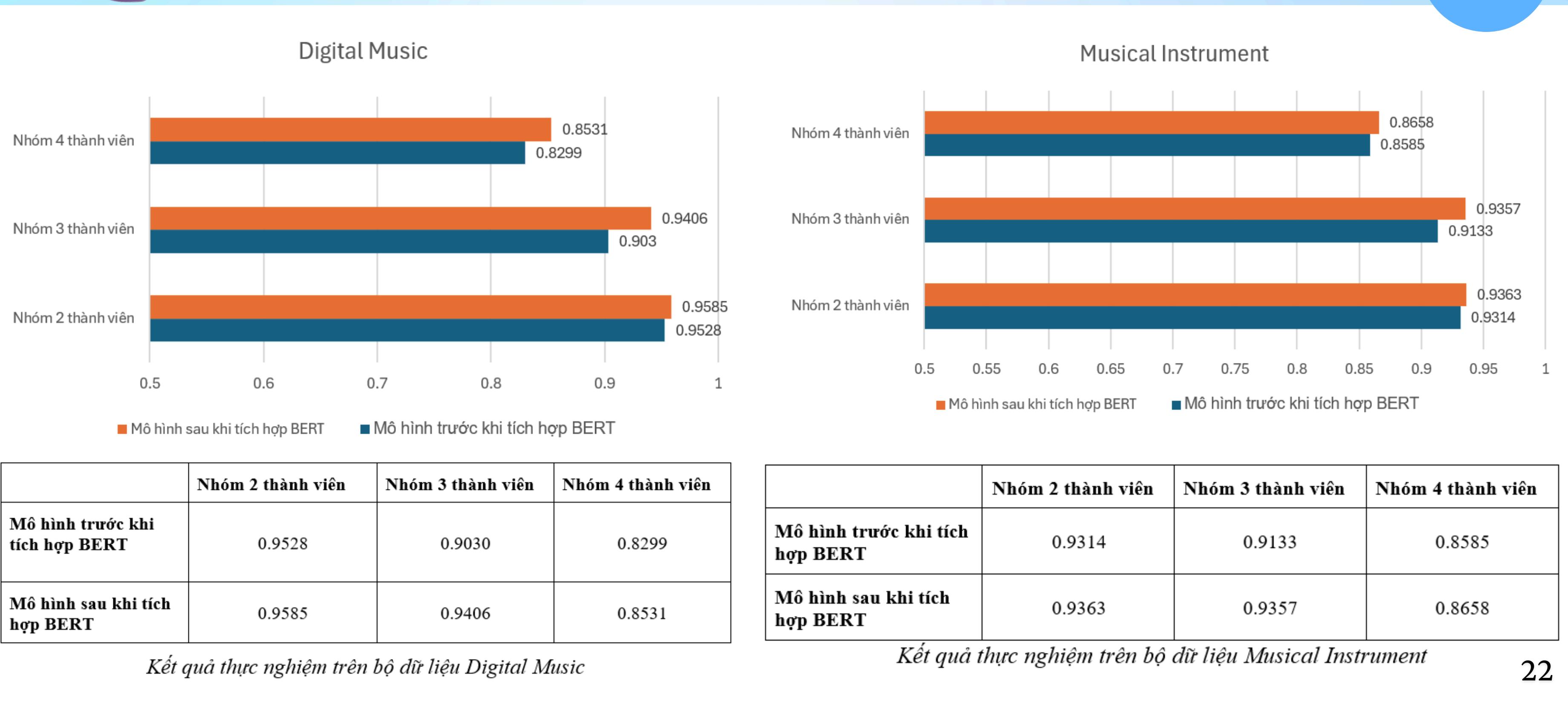
Độ đo đánh giá : F-Score

$$precision_c = \frac{|T \cap C|}{|C|} \quad recall_c = \frac{|T \cap C|}{|T|}$$

$$F_{score} = \frac{2 \times precision_c \times recall_c}{precision_c + recall_c}$$



KẾT QUẢ THỰC NGHIỆM



KẾT LUẬN

Đóng góp của đề tài

- Tích hợp thành công rating được tạo ra từ nhận xét của người dùng vào mô hình nhân tố ẩn.
- Triển khai thành công giải pháp đã đề xuất cho bài toán tư vấn nhóm.

Hướng phát triển

- Cải thiện độ chính xác của mô hình nhân tố ẩn bằng cách tích hợp thêm thông tin từ mạng xã hội
- *Cải tiến phương pháp tổng hợp profile bằng số lượng tương tác của người dùng với từng danh mục sản phẩm



THANK YOU

FOR YOUR ATTENTION



HƯỚNG TIẾP CẬN



Tinh chỉnh mô hình BERT cho bài toán phân loại văn bản
Bộ dữ liệu huấn luyện được trích xuất từ cuộc thi **Yelp Dataset Challenge 2015**.

- Bộ dữ liệu huấn luyện:
 - 650,000 đánh giá(rating) kèm theo nhận xét (text review)
 - Với mỗi mức đánh giá 1-5 sẽ có 130.000 mẫu
- Bộ dữ liệu kiểm tra:
 - 50,000 đánh giá(rating) có kèm theo nhận xét (text review)
 - Với mỗi mức đánh giá 1-5 sẽ có 10.000 mẫu

Hàm mất mát được sử dụng là hàm Cross-entropy loss:

$$CE\ Loss = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k y_{ij} \log (p_{ij})$$

