

# **MINER: Multi-Interest Matching Network for NewsRecommendation**

**DS300.011**

**GVHD: Huỳnh Văn Tín**

**THÀNH VIÊN NHÓM 11**

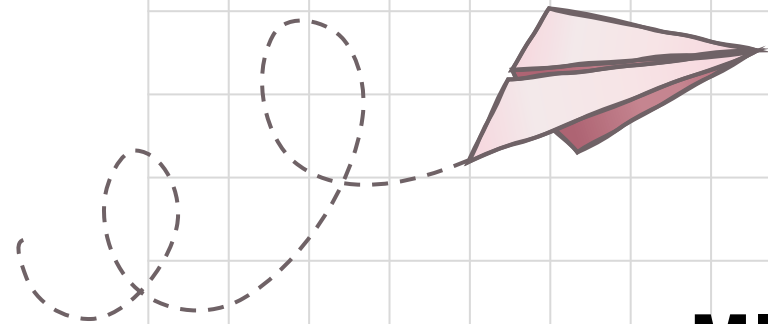
20521609 – Nguyễn Hoàng Minh

20521998 – Nguyễn Thiện Thuật

20521614 – Tạ Nhật Minh

**Link paper:**

<https://aclanthology.org/2022.findings-acl.29.pdf>



# 1. Introduction

## MINER: Multi-Interest Matching Network for News Recommendation

Các đóng góp của bài báo:

- Đề xuất mô hình **poly attention** để trích xuất **nhiều vector sở thích** cho mỗi người dùng.
- Đề xuất **category-aware attention weighting strategy** nhằm nắm bắt tín hiệu sở thích rõ ràng dựa trên thông tin danh mục tin tức.
- MINER đạt **kết quả tốt nhất trên tập dữ liệu MIND** và đứng đầu bảng xếp hạng chính thức vào tháng 9 năm 2021.



## 2. Related work

### Phương pháp gợi ý truyền thống:

- Đặc trưng được biểu diễn dưới dạng danh mục và được đại diện bằng ID.
- Sử dụng các mạng học sâu để tích hợp cả shallow và deep network hoặc deep & cross sub-network để mô hình hóa các đặc trưng người dùng.
- Tập trung vào vấn đề gợi ý tuần tự, nhằm nắm bắt hành vi tuần tự của người dùng thông qua mô hình chuỗi như RNN, CNN và mạng self-attention.

### Gợi ý tin tức dựa trên mạng nơron:

- Một số phương pháp sử dụng cơ chế attention để đại diện cho tin tức và người dùng.
- Sử dụng BERT tiền huấn luyện để mã hóa tin tức.
- Hầu hết các phương pháp học một vector duy nhất đại diện cho người dùng.



# 3. Task representation

<i>Historical Clicked News</i>		
1	Finance	Man who inherited 6 figures shares advice he'd give his younger self.
2	Sports	Foles will start for Jaguars over Minshew after bye week.
3	Sports	Pete Carroll takes swipe at Patriots over their strict culture.
4	Food	The best Trader Joe's desserts of all time.
5	Politics	Senate to try to override Trump emergency declaration veto Thursday.
6	Sports	NFL had no choice but to send a clear message with Garrett punishment.
7	Sports	Umpire Jeff Nelson leaves game with concussion after being hit by foul balls.
8	Food	Wendy's is turning 50 years old, and is gifting us free food through 2020.

<i>Recommended by NRMS+BERT</i>	
Sports	NFL week 8 power rankings: old-school football rules the day.
Sports	Patriots wanted a test. Now, they need some answers.
Politics	40 conservative groups sign ethics complaint against Pelosi.

<i>Recommended by MINER</i>	
Sports	Patriots wanted a test. Now, they need some answers.
Food	<b>National Dessert Day: Where to get free dessert at Wendy's.</b>
Health	Simple diet changes helped this guy lose 75 pounds in 9 months.

Figure 5: Case study on top 3 news recommended by *NRMS+BERT* and *MINER* in a sampled impression. The news actually clicked by the user is highlighted in blue.

# 4. Background

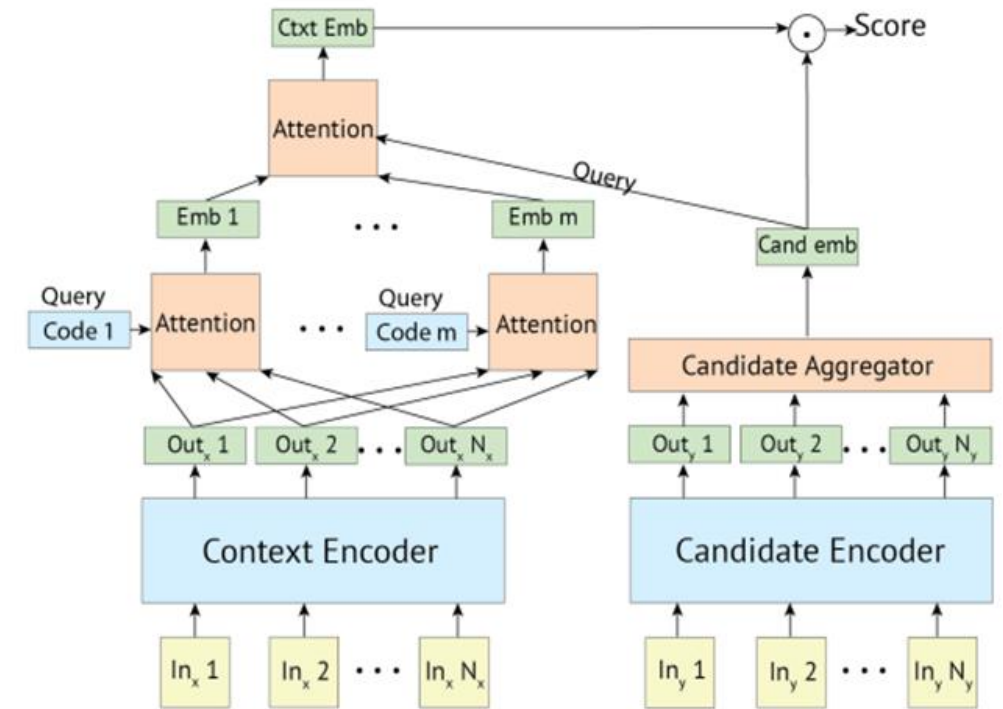
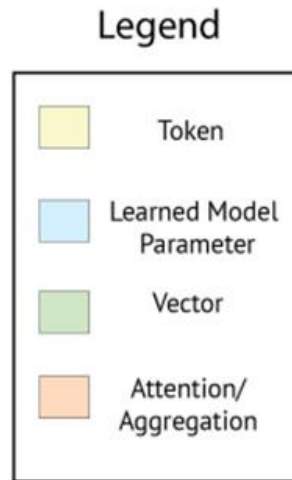
**Test on 2 task:** Sentence selection in dialogue + Article search in IR

- **m** context codes ( $c_1, \dots, c_m$ )
- **m** vectors ( $y_{ctx}^1 \dots y_{ctx}^m$ )
- Where  $c_i$  extracts representation  $y_{ctx}^i$ .  
That is, we obtain  $y_{ctx}^i$  using:

$$y_{ctx}^i = \sum_j w_j^{c_i} h_j \quad \text{where} \quad (w_1^{c_i}, \dots, w_N^{c_i}) = \text{softmax}(c_i \cdot h_1, \dots, c_i \cdot h_N)$$

- Finally, given our **m** global context features, we attend over them using  $y_{candi}$  as the query:

$$y_{ctx} = \sum_i w_i y_{ctx}^i \quad \text{where} \quad (w_1, \dots, w_m) = \text{softmax}(y_{candi} \cdot y_{ctx}^1, \dots, y_{candi} \cdot y_{ctx}^m)$$



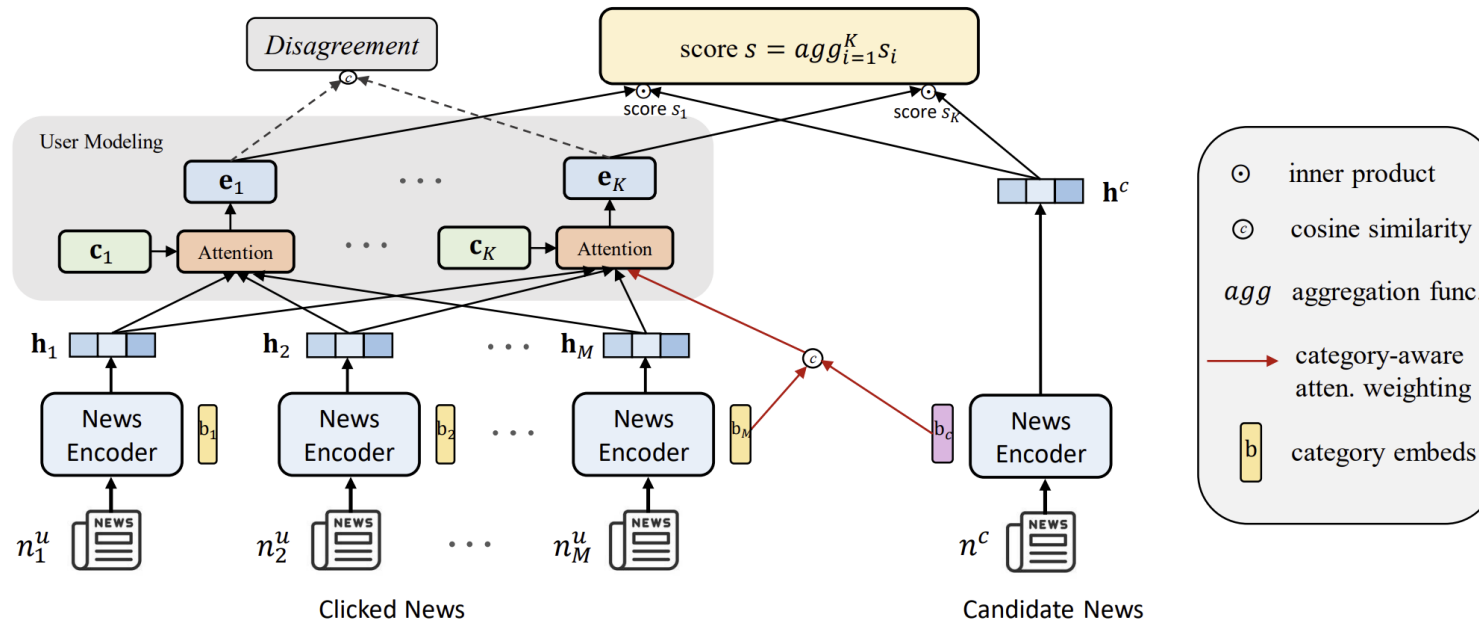
(c) Poly-encoder

Link paper:

<https://arxiv.org/pdf/1905.01969.pdf>

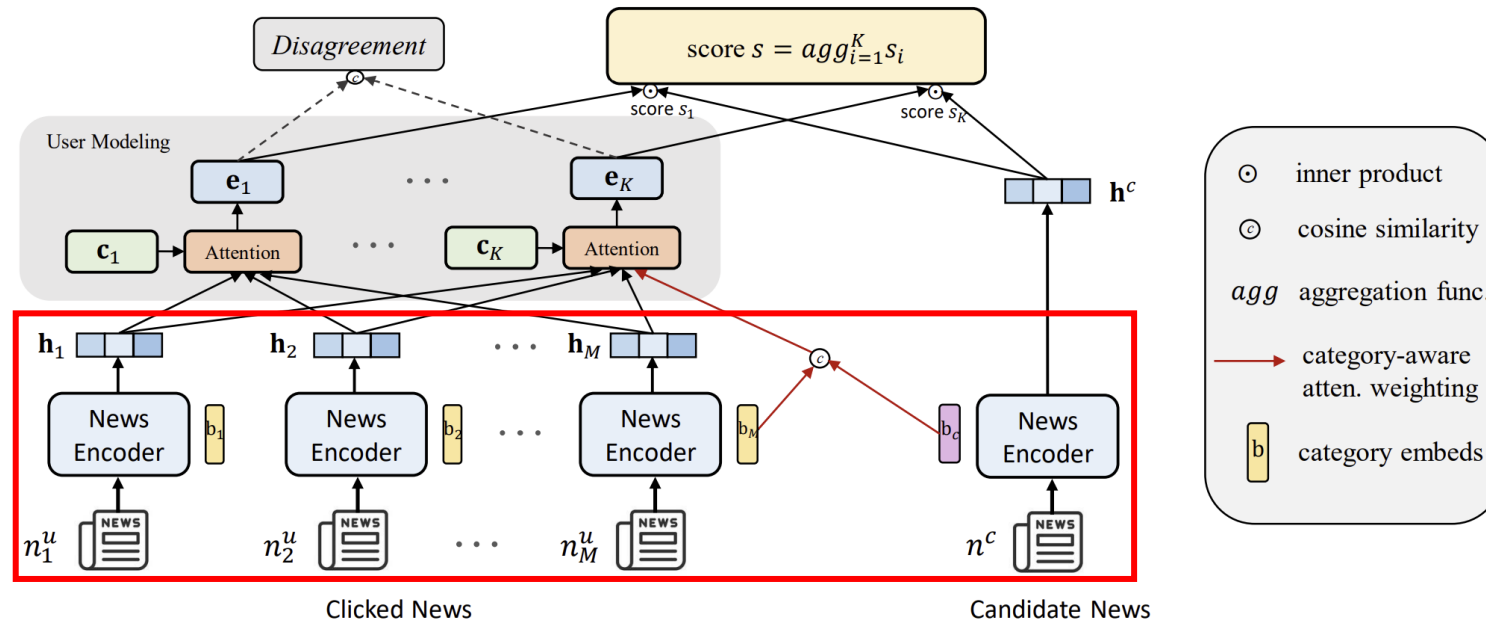
# 5. MINER

## Overall framework + Problem Formulation



- Với user  $u$  và tin tức ứng cử viên  $n^c$ . Mục tiêu là tính toán điểm sở thích (interest score  $s$ ) – Thể hiện sự quan tâm của user đối với tin tức  $n^c$ .
- User  $u$  bao gồm một danh sách các tin tức lịch sử được nhấp vào  $N^u = [n_1^u, n_2^u, \dots, n_M^u]$ , với  $M$  là số lượng tin tức người dùng  $u$  đã nhấp. Mỗi tin tức  $n$  sẽ có **title T** và **category ct**.
- Danh sách tin tức ứng cử viên  $N^c$  sẽ được sắp hạng dựa trên điểm sở thích và topK sẽ được đề xuất cho user  $u$ .

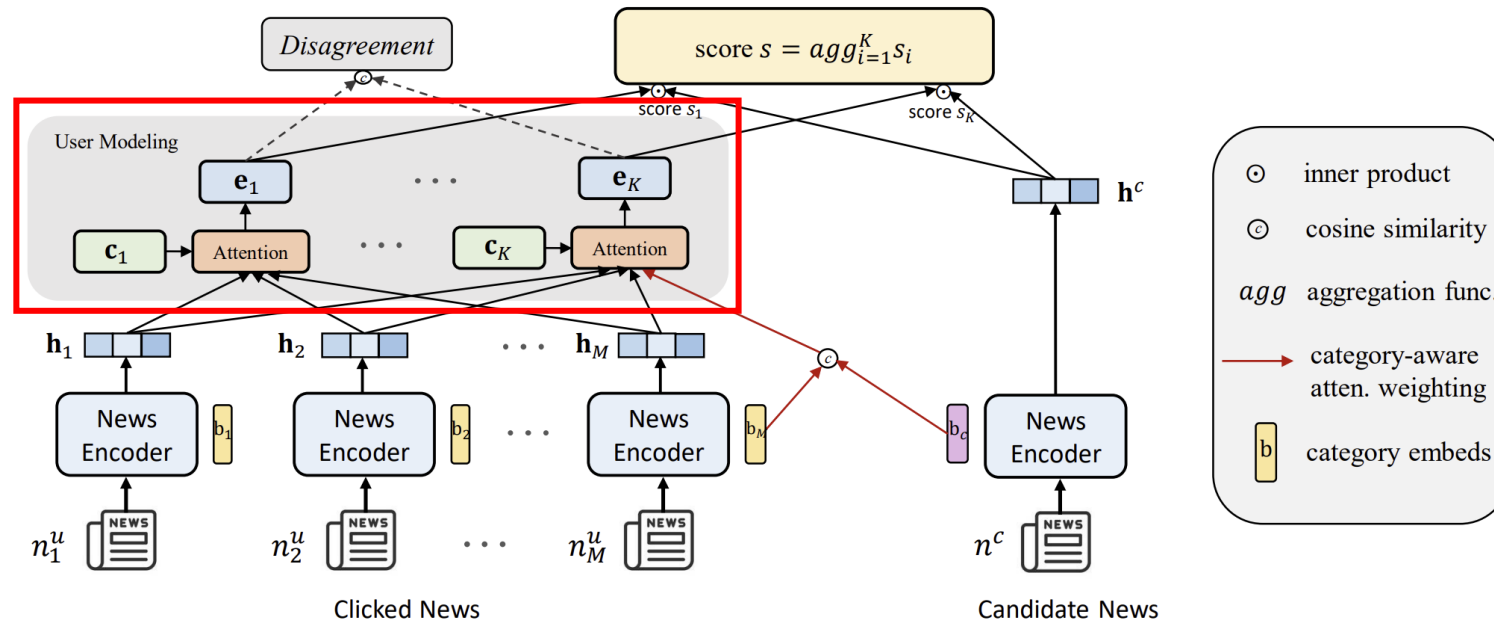
# 5. MINER News Encoder



- Sử dụng **pre-trained BERT** cho news encoder, giúp nắm bắt ngữ nghĩa của tin tức.
- Sử dụng **[CLS] token** là embedding  $h$  của tin tức.
- Người dùng  $u$  và Tin tức ứng cử viên  $n^c$  sẽ được encode thành  $H^u = [h_1, h_2, \dots, h_M]$  và  $h_c$ .
- Trong Ablation Study, dùng **shallow word embeddings** và **self-attention networks** để thay thế BERT.

# 5. MINER

## Multi-Interest User Modeling

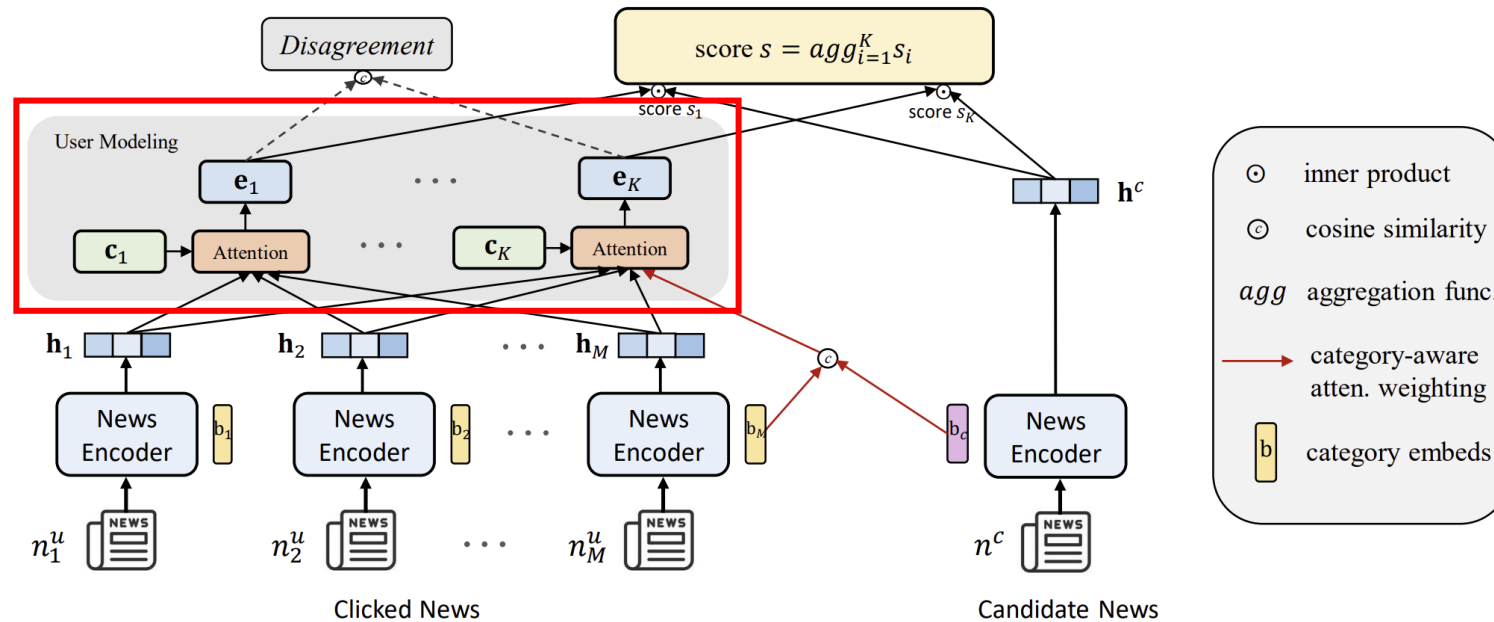


- **Đầu vào:** Danh sách các embedding tin tức người dùng đã nhấn.
- **Đầu ra:** Danh sách vector  $e_i$  thể hiện được sở thích về tin tức mà người dùng  $u$  quan tâm.
- Họ đề xuất nhiều vector biểu diễn để thể hiện sở thích của từng người dùng  $u$   $E^u = [e_1, e_2, \dots, e_K]$  Với mỗi  $e_i$  biểu diễn mỗi sở thích riêng biệt của người dùng.



# 5. MINER

## Multi-Interest User Modeling



- **A poly attention scheme:** (lấy ý tưởng từ *Poly-Encoder*) Dùng để trích xuất **K** Vector sở thích Thông qua **K** additive attentions.

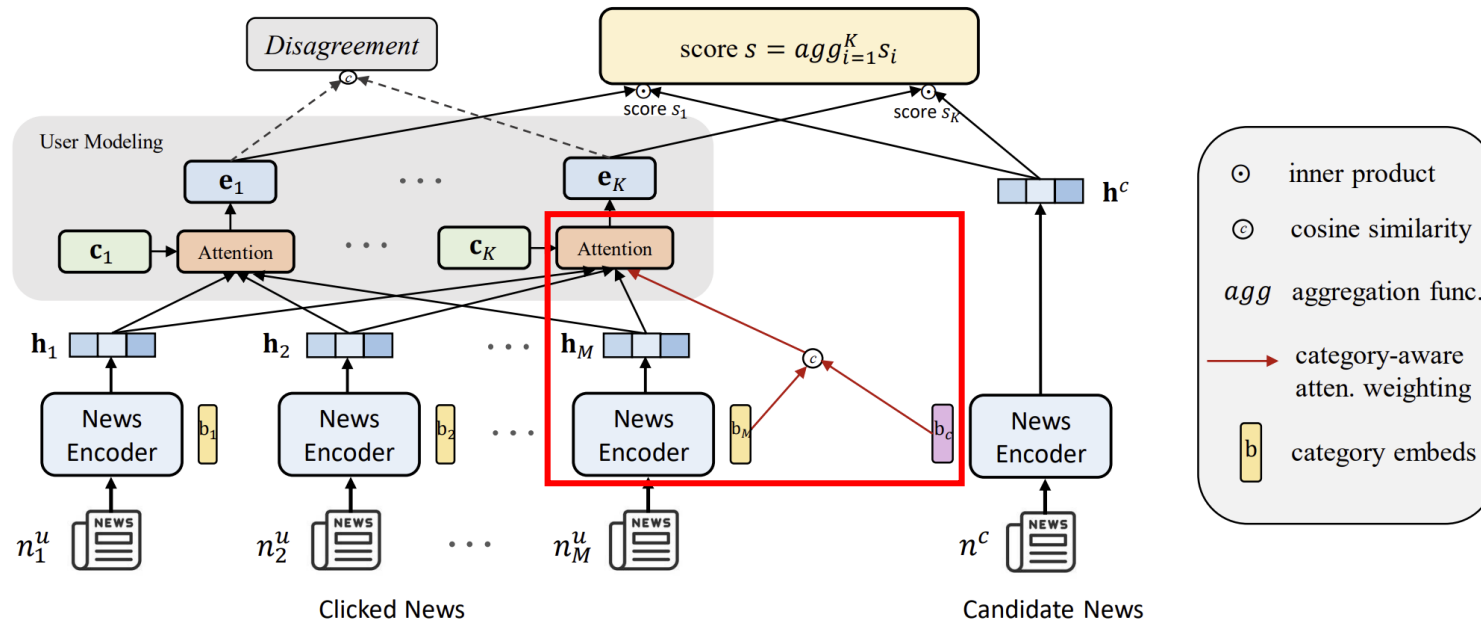
$$e_i = \sum_{j=1}^M w_j^{c_i} h_j, \quad w_j^{c_i} = \text{softmax}(\phi_h^{c_i}(h_j)), \quad (1)$$

$$\phi_h^{c_i}(h_j) = c_i^T \tanh(W^h h_j), \quad (2)$$

- Với  $c_i$  và  $W^h$  là 2 tham số huấn luyện. Vector biểu diễn sở thích ko có ý nghĩa rõ ràng.

# 5. MINER

## Category-aware Attention Weighting

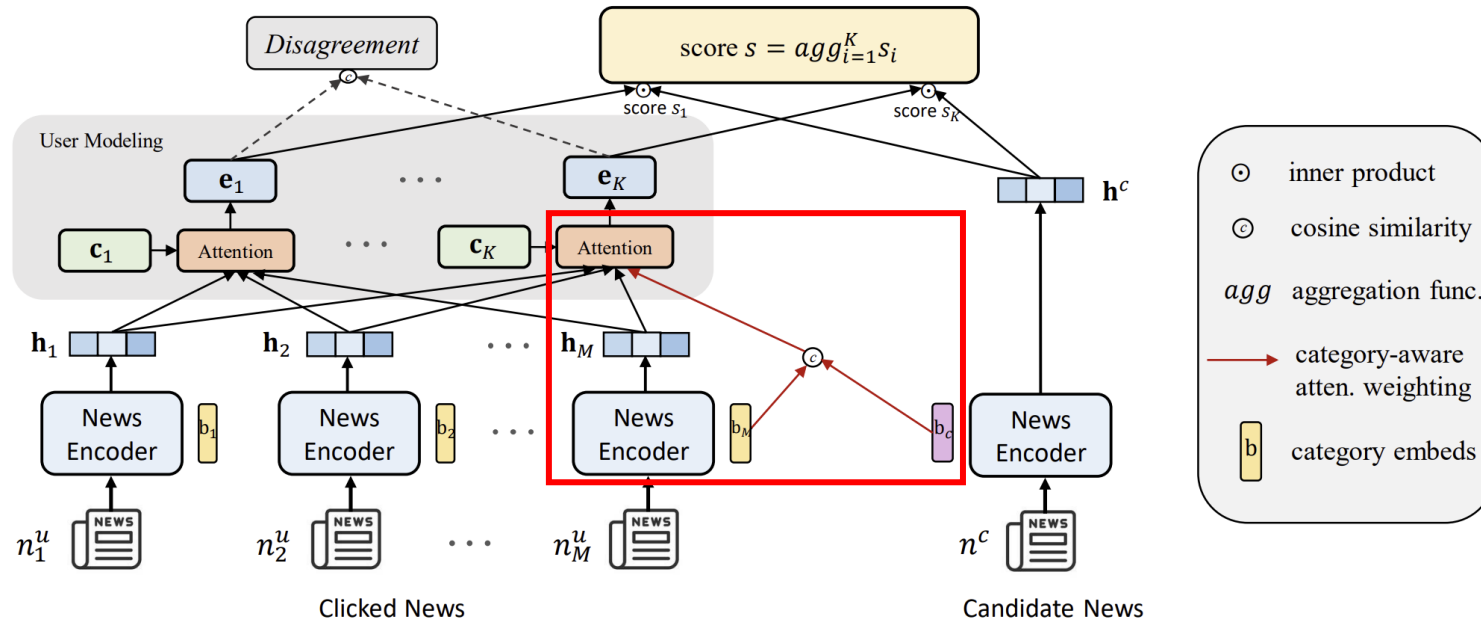


- The category information (e.g., Sports, Health) có thể ảnh hưởng đến sự quan tâm của người dùng.
- Họ đề xuất **Category-aware Attention Weighting strategy** để đánh giá tin tức theo mức độ tương đồng giữa các thể loại tin tức cũ và thể loại tin tức ứng cử viên. (*cùng thể loại thì độ tương đồng cao*)

$$w_j^{c_i} = \text{softmax}(\phi_h^{c_i}(\mathbf{h}_j) + \lambda \cos(\mathbf{b}_j, \mathbf{b}_c)), \quad (5)$$

# 5. MINER

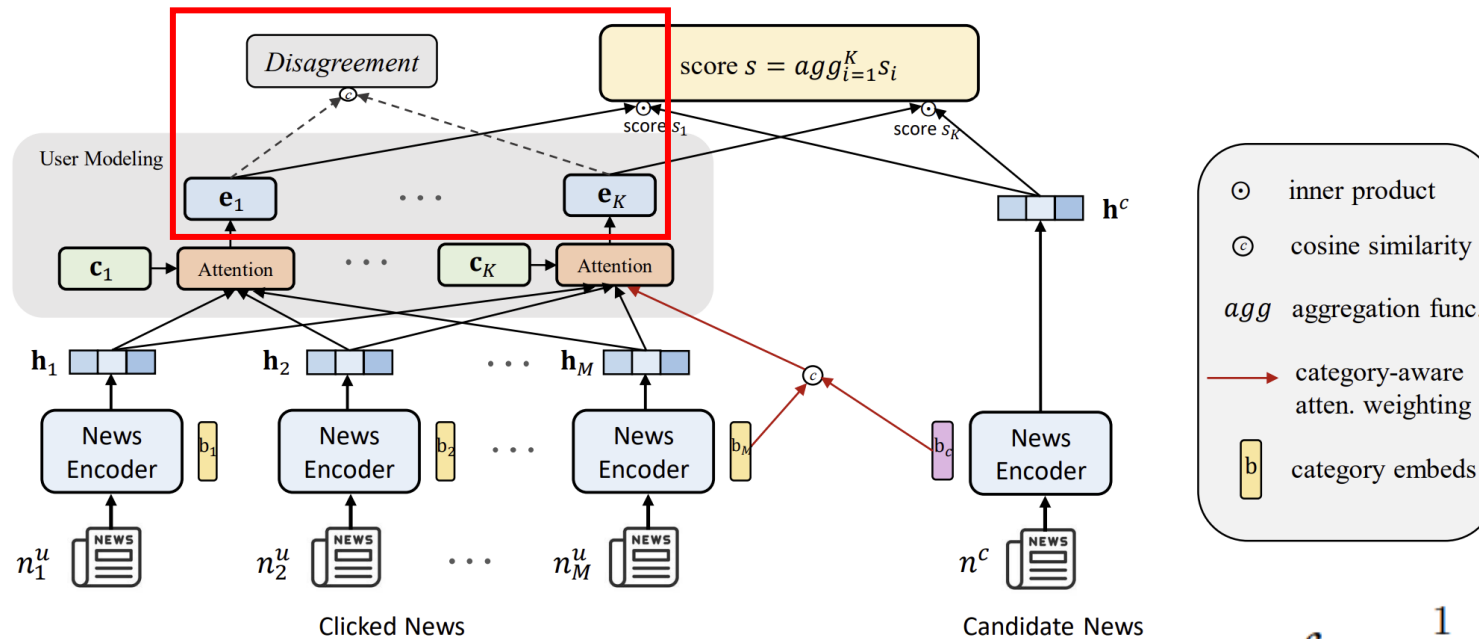
## Category-aware Attention Weighting



- Sử dụng **Pre-trained Glove** để word embedding các category words (e.g., Sports, Health). Bổ sung vào công thức (1) với bias term.
- Với  $\cos(\cdot)$  là cosine similarity giữa 2 category embeddings và  $\lambda$  là tham số học.

# 5. MINER

## Disagreement Regularization

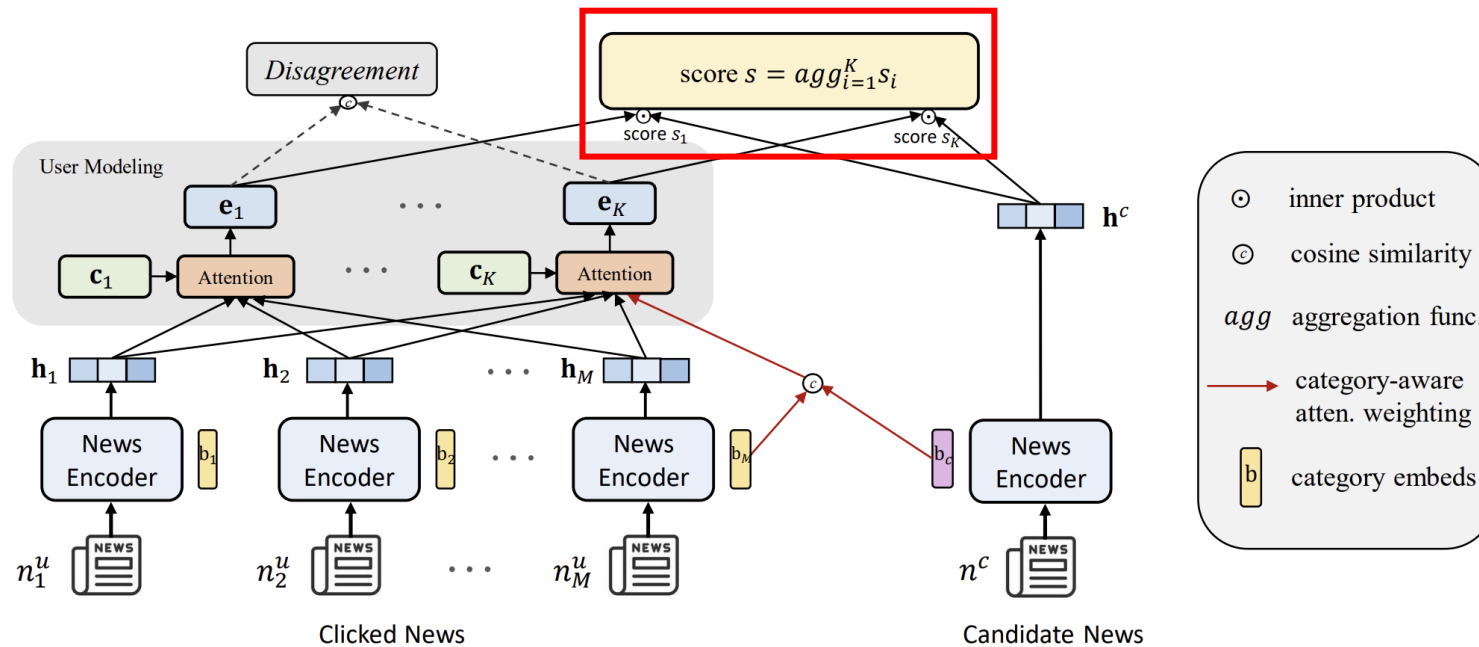


- **Mục đích:** làm cho các biểu diễn sở thích  $e_i$  được trích xuất trở nên đa dạng hơn.
- Lấy ý tưởng từ bài (Multi-head attention with disagreement regularization), làm tăng khoảng cách giữa các vector  $e_i$  trong quá trình training.
- Tính thông qua cosine similarity giữa 2 vector  $e_i$ . Giảm tối thiểu trung bình cosine (maximize khoảng cách).
- Với  $K$  là số lượng vector sở thích.

$$\mathcal{L}_D = \frac{1}{K^2} \sum_{i=1}^K \sum_{j=1}^K \frac{\mathbf{e}_i^\top \mathbf{e}_j}{\|\mathbf{e}_i\| \|\mathbf{e}_j\|}, \quad (3)$$

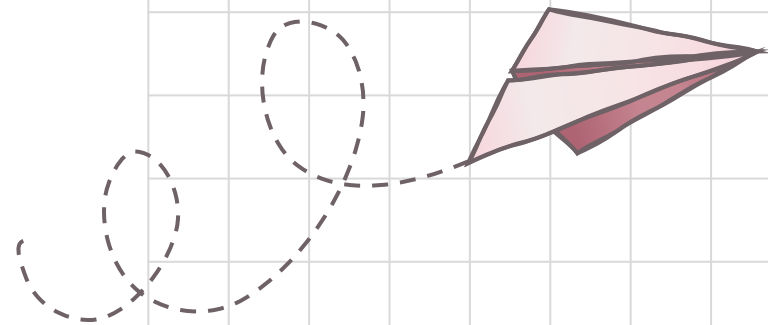
# 5. MINER

## Click Predictor



- Tính điểm tương đồng (matching score) giữa tin tức ứng của viên với người dùng theo công thức:

$$s_i = \mathbf{e}_i^\top \mathbf{h}^c. \quad (4)$$



# 5. MINER

## Click Predictor

- Các cách họ đề xuất để tính **click score** cuối cùng:
  - **MINER-max**:  $s = \max_{i=1}^K s_i$ .
  - **MINER-mean**:  $s = \text{mean}_{i=1}^K s_i$ .
  - **MINER-weighted**: (Deep knowledge-aware network for news recommendation)

$$s = \sum_{i=1}^K w_i s_i,$$
$$w_i = \text{softmax}(\mathbf{e}_i^\top \text{gelu}(\mathbf{W}^e \mathbf{h}^c)),$$

○ Với  $\mathbf{W}^e$  là tham số huấn luyện.



# 5. MINER

## Model Training

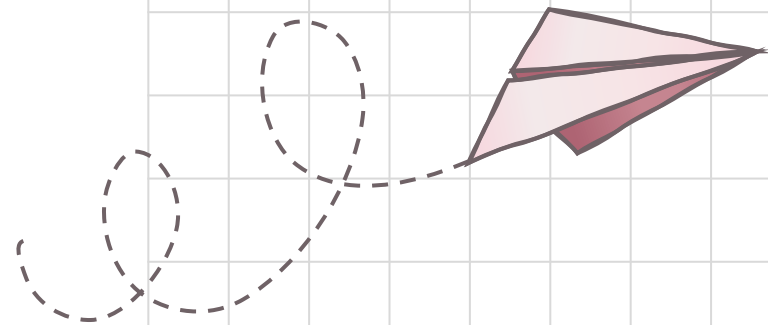
- Sử dụng **Loss NCE** để **tính lỗi** dự đoán.

$$\mathcal{L}_{NCE} = - \sum_{i=1}^{|\mathcal{D}|} \log \frac{\exp(s_i^+)}{\exp(s_i^+) + \sum_{j=1}^L \exp(s_i^j)}. \quad (6)$$

- Trong dataset training  $\mathcal{D}$ , mỗi tin tức đã nhấn là positive sample  $\mathbf{n}_i^+$
  - Trong dataset training  $\mathcal{D}$ , chọn ngẫu nhiên  $L$  tin tức chưa được nhấn (negative sample  $[\mathbf{n}_i^1, \dots, \mathbf{n}_i^L]$ ).
  - Dự đoán click scores: Tập  $s_i^+$  và negative sample  $[s_i^1, \dots, s_i^L]$ .
  - Tính lỗi giữa tập dự đoán và tập đã chọn.
- Kết hợp với **Loss ở (3)**, Hàm **Loss để huấn luyện** là:

$$\mathcal{L} = \mathcal{L}_{NCE} + \beta * \mathcal{L}_D, \quad (7)$$

- Với  $\beta$  là hyper-parameter và bằng **0.8** dựa trên hiệu suất của tập huấn luyện.



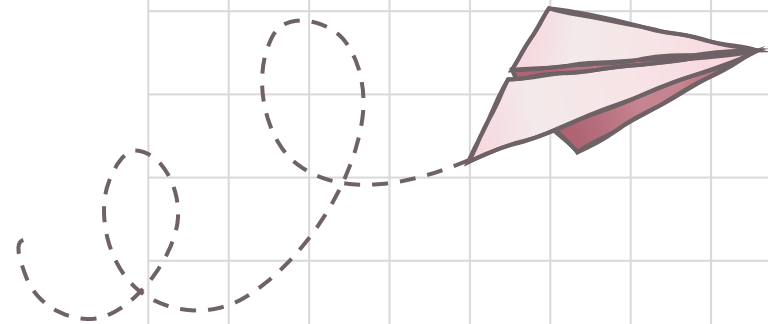
# 6. Experiments

## Setup – Dataset

- Sử dụng bộ dữ liệu gợi ý tin tức thực tế MIND được thu thập từ nhật ký hành vi người dùng trên Microsoft News.
- Gồm hai phiên bản: MIND-large và MIND-small:
  - MIND-large: Hơn 15 triệu nhật ký từ 1 triệu người dùng.
  - MIND-small: Lấy ngẫu nhiên 50.000 người dùng từ MIND-large.
- Mỗi record ghi lại tin tức đã nhấp và không được nhấp mà được hiển thị cho người dùng, cùng với lịch sử hành vi nhấp tin tức của người dùng.
- Bộ dữ liệu MIND cũng chứa nhãn danh mục của mỗi tin tức.

	MIND-small	MIND-large
# News	65,238	161,013
# Categories	18	20
# Impressions	230,117	15,777,377
# Clicks	347,727	24,155,470





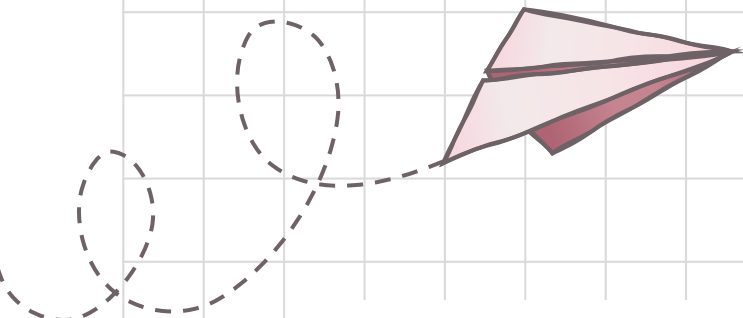
# 6. Experiments

## Setup – Setting

- Sử dụng 50 tin tức đã nháp gần đây nhất của người dùng để học đại diện người dùng.
- Chỉ sử dụng tiêu đề tin tức và giới hạn độ dài tối đa là 20.
- Sử dụng mô hình **bert-base-uncased** làm mô hình được tiền huấn luyện để khởi tạo bộ mã hóa tin tức.
- Số lượng **context codes** K được đặt là 32.
- Kích thước của các vector mã ngữ cảnh là 200.
- **Category Embedding** được khởi tạo bằng vector **Glove** có kích thước 300 và được cố định trong quá trình huấn luyện.
- Tỷ lệ lấy mẫu tiêu cực L được đặt là 4.
- Huấn luyện trong 5 **epoch** và **batch\_size** là 128.
- Sử dụng Adam làm thuật toán tối ưu hóa.
- Đánh giá hiệu suất sử dụng bốn chỉ số xếp hạng: **AUC**, **MRR**, **nDCG@5**, và **nDCG@10**.

# 6. Experiments

## Main Result



#	Methods	MIND-small				MIND-large			
		AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
1	LibFM	59.74	26.33	27.95	34.29	61.85	29.45	31.45	37.13
2	DeepFM	59.89	26.21	27.74	34.06	61.87	29.30	31.35	37.05
3	DKN	62.90	28.37	30.99	37.41	64.07	30.42	32.92	38.66
4	NPA	64.65	30.01	33.14	39.47	65.92	32.07	34.72	40.37
5	NAML	66.12	31.53	34.88	41.09	66.46	32.75	35.66	41.40
6	LSTUR	65.87	30.78	33.95	40.15	67.08	32.36	35.15	40.93
7	NRMS	65.63	30.96	34.13	40.52	67.66	33.25	36.28	41.98
8	HieRec <sup>†</sup>	67.95	32.87	36.36	42.53	69.03	33.89	37.08	43.01
9	LSTUR+BERT <sup>‡</sup>	68.28	32.58	35.99	42.32	69.49	34.72	37.97	43.70
10	NRMS+BERT <sup>‡</sup>	<u>68.60</u>	<u>32.97</u>	<u>36.55</u>	<u>42.78</u>	69.50	34.75	37.99	43.72
11	UNBERT <sup>§</sup>	67.62	31.72	34.75	41.02	<u>70.68</u>	<u>35.68</u>	<u>39.13</u>	<u>44.78</u>
12	MINER- <i>max</i>	67.39	32.37	35.93	42.11	69.97	35.03	38.37	44.05
13	MINER- <i>mean</i>	69.49	33.44	37.37	43.53	71.37	36.06	39.56	45.21
14	MINER- <i>weighted</i>	<b>69.61</b>	<b>33.97</b>	<b>37.62</b>	<b>43.90</b>	<b>71.51</b>	<b>36.18</b>	<b>39.72</b>	<b>45.34</b>

# 6. Experiments

## Ablation Study

Kiểm tra ảnh hưởng của **stacked BERT encoder** bằng cách thay bằng **word embeddings cơ bản**.

- **SOTA non-BERT model HieRec**

**Kết quả:**

- BERT có vai trò **quan trọng** trong MINER khi hiệu suất **giảm đáng kể** nếu sử dụng word embeddings cơ bản.
- Mô hình biến thể này vẫn có thể vượt trội hơn **SOTA non-BERT model HieRec**.

Loại bỏ **disagreement regularization** và **category-aware attention**:

**Kết quả:**

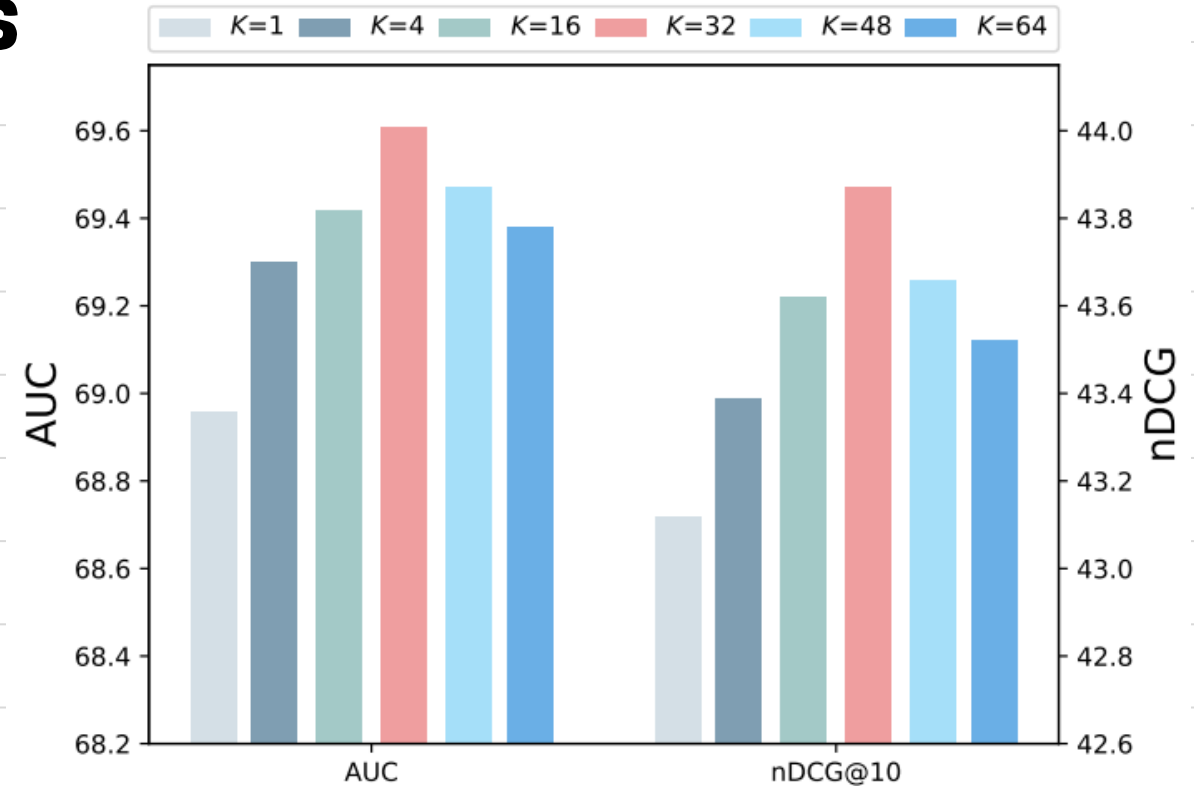
- Lợi ích của việc đa dạng hóa các interest vectors và tích hợp category.
- Sự **suy giảm hiệu suất** khi loại bỏ **category-aware attention** **lớn hơn** disagreement regularization  
⇒ tính quan trọng của **category signals** trong MINER.

Model	AUC	MRR	nDCG@10
HieRec (Qi et al., 2021)	67.95	32.87	42.53
MINER w/o BERT	68.07	32.93	42.62
w/o disagreement	67.42	32.38	42.12
w/o category	67.13	32.06	41.73
MINER with BERT	69.61	33.97	43.90
w/o disagreement	69.49	33.46	43.56
w/o category	69.38	33.60	43.60

# 6. Experiments

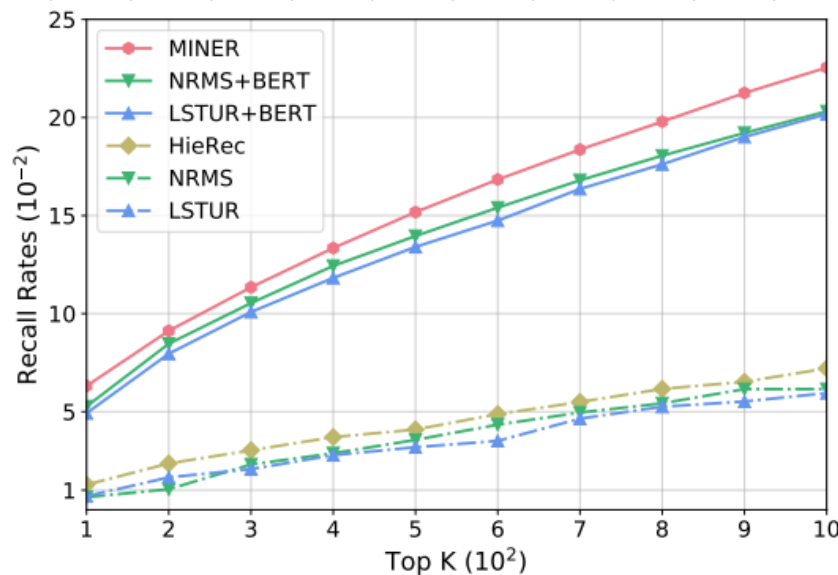
## Number of Interest Vectors

- Hiệu suất **lúc đầu tăng** và **sau đó giảm**.
- Kết quả tốt nhất khi **K = 32**
- K quá lớn → nhiều tham số dư thừa → ảnh hưởng xấu.
- Khi K = 1, AUC vẫn đạt kết quả tốt là **68,92**.



# 6. Experiments

## Effectiveness on News Recall



#	Methods	MIND-small				MIND-large			
		AUC	MRR	nDCG@5	nDCG@10	AUC	MRR	nDCG@5	nDCG@10
1	LibFM	59.74	26.33	27.95	34.29	61.85	29.45	31.45	37.13
2	DeepFM	59.89	26.21	27.74	34.06	61.87	29.30	31.35	37.05
3	DKN	62.90	28.37	30.99	37.41	64.07	30.42	32.92	38.66
4	NPA	64.65	30.01	33.14	39.47	65.92	32.07	34.72	40.37
5	NAML	66.12	31.53	34.88	41.09	66.46	32.75	35.66	41.40
6	LSTUR	65.87	30.78	33.95	40.15	67.08	32.36	35.15	40.93
7	NRMS	65.63	30.96	34.13	40.52	67.66	33.25	36.28	41.98
8	HieRec <sup>†</sup>	67.95	32.87	36.36	42.53	69.03	33.89	37.08	43.01
9	LSTUR+BERT <sup>‡</sup>	68.28	32.58	35.99	42.32	69.49	34.72	37.97	43.70
10	NRMS+BERT <sup>‡</sup>	68.60	32.97	36.55	42.78	69.50	34.75	37.99	43.72
11	UNBERT <sup>§</sup>	67.62	31.72	34.75	41.02	70.68	35.68	39.13	44.78
12	MINER-max	67.39	32.37	35.93	42.11	69.97	35.03	38.37	44.05
13	MINER-mean	69.49	33.44	37.37	43.53	71.37	36.06	39.56	45.21
14	MINER-weighted	69.61	33.97	37.62	43.90	71.51	36.18	39.72	45.34

MINER đạt được hiệu quả cao.

Kết quả từ thí nghiệm này cũng tương đồng với các kết quả trong bảng



# 6. Experiments

## Case Study and Visualization

<i>Historical Clicked News</i>		
1	Finance	Man who inherited 6 figures shares advice he'd give his younger self.
2	Sports	Foles will start for Jaguars over Minshew after bye week.
3	Sports	Pete Carroll takes swipe at Patriots over their strict culture.
4	Food	The best Trader Joe's desserts of all time.
5	Politics	Senate to try to override Trump emergency declaration veto Thursday.
6	Sports	NFL had no choice but to send a clear message with Garrett punishment.
7	Sports	Umpire Jeff Nelson leaves game with concussion after being hit by foul balls.
8	Food	Wendy's is turning 50 years old, and is gifting us free food through 2020.

<i>Recommended by NRMS+BERT</i>	
Sports	NFL week 8 power rankings: old-school football rules the day.
Sports	Patriots wanted a test. Now, they need some answers.
Politics	40 conservative groups sign ethics complaint against Pelosi.

<i>Recommended by MINER</i>	
Sports	Patriots wanted a test. Now, they need some answers.
Food	<b>National Dessert Day: Where to get free dessert at Wendy's.</b>
Health	Simple diet changes helped this guy lose 75 pounds in 9 months.

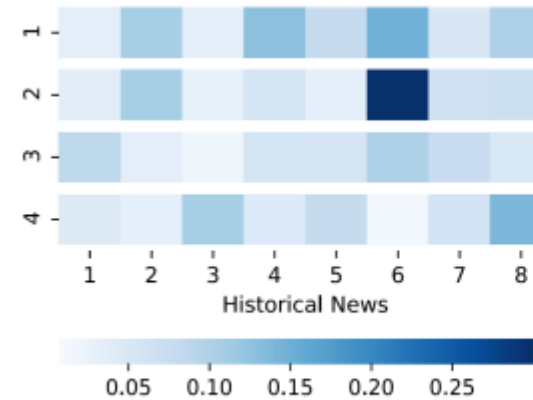
- Sports news on the top.
- NRMS+BERT → single user → khó nắm bắt **interest** với những **user** khác.
- MINER → nhiều khía cạnh → phát hiện **user** cũng quan tâm.
- **Health news** liên quan đến **Sports and Food**.



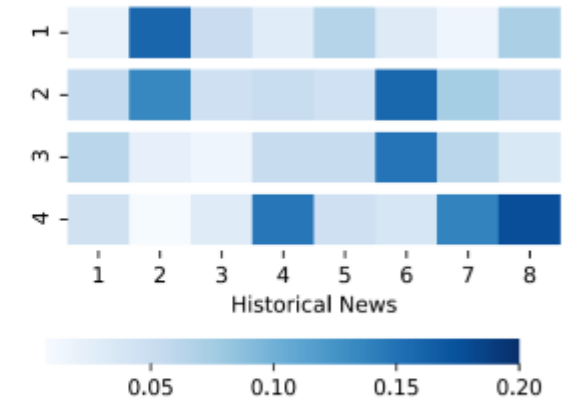
# 6. Experiments

## Case Study and Visualization

Historical Clicked News		
1	Finance	Man who inherited 6 figures shares advice he'd give his younger self.
2	Sports	Foles will start for Jaguars over Minshew after bye week.
3	Sports	Pete Carroll takes swipe at Patriots over their strict culture.
4	Food	The best Trader Joe's desserts of all time.
5	Politics	Senate to try to override Trump emergency declaration veto Thursday.
6	Sports	NFL had no choice but to send a clear message with Garrett punishment.
7	Sports	Umpire Jeff Nelson leaves game with concussion after being hit by foul balls.
8	Food	Wendy's is turning 50 years old, and is gifting us free food through 2020.



(a) w/o disagreement



(b) with disagreement

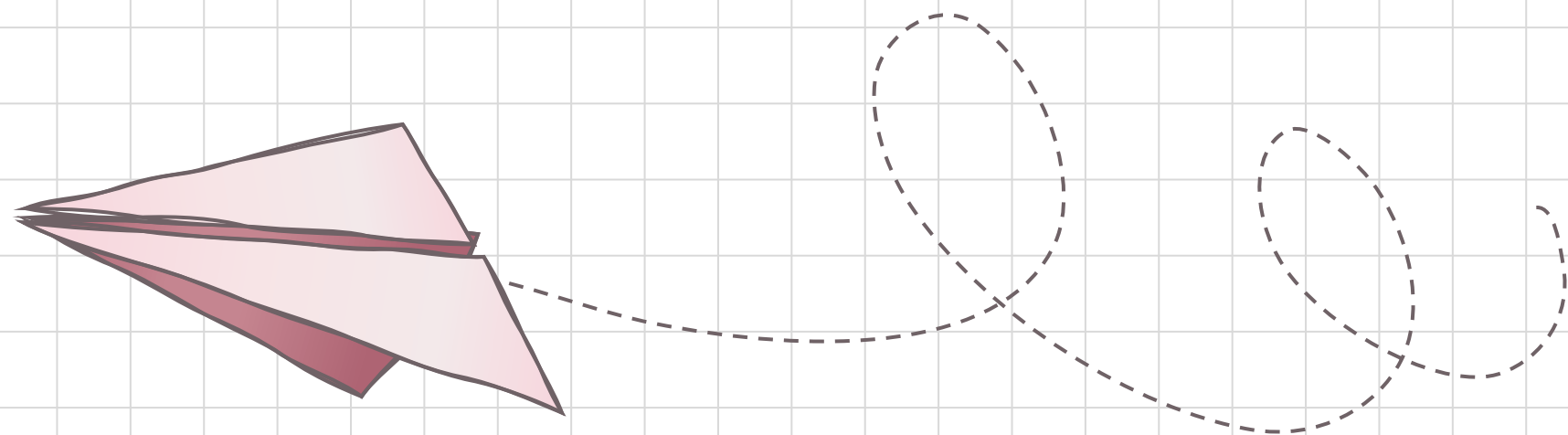
- Huấn luyện MINER với 4 vector **interest** (tức là  $K = 4$ ) và trực quan hóa các trọng số chú ý trước và sau khi áp dụng **disagreement**.
- **Trước:** Interest thứ 2 và news thứ 6 là Sports news.
- **Sau:** Interest thứ 4 và news thứ 8 là Food news.

# Conclusion

- Đề xuất phương pháp MINER để nắm bắt sở thích người dùng từ lịch sử.
- Đề xuất **sơ đồ poly attention** để tìm hiểu **multiple user interest vectors** thông qua **soft attention**.
- Đề xuất **disagreement regularization** để cải thiện poly **attention**, nhằm để **interests vectors** đa dạng.
- Thực hiện nghiên cứu trường hợp và trực quan hóa kết quả.



Q&A!



# Thank you for listening!

