

NAML-LSTUR: A combined model with Attentive Multi-View Learning and Long- and Short-term User Representations for News Recommendation

Nguyễn Hoàng Minh, Nguyễn Thiện Thuật, Tạ Nhật Minh
{20521609, 20521998, 20521614}@gm.uit.edu.vn

Trường Đại học Công nghệ Thông tin, Đại học Quốc gia Thành phố Hồ Chí Minh

TÓM TẮT

Trong đề tài này, chúng tôi đề xuất mô hình khuyến nghị (**NAML-LSTUR**) cho bài toán gợi ý tin tức (News Recommendation) với sự kết hợp của hai mô hình state-of-the-art trong lĩnh vực. Kết quả thực nghiệm cho thấy mô hình đề xuất của chúng tôi cao hơn phần lớn các mô hình state-of-the-art hiện tại.

MÔ TẢ BÀI TOÁN

NAML + LSTUR			
Category	Sub-Category	Title	Abstract
tv	humor	Chrissy Teigen Scars John Lege...	The model says that scaring ...
news	newsus	Disney's Hulu is raising prices for...	Hulu's live TV bundle for cord...
movies	movienews	13 Reasons Why's Christian Na...	13 Reasons Why's Christian N...
tv	tv-celebrity	Jane Fonda Avoids Fifth Arrest...	This week, the actress was...
tv	tv-celebrity	Joshua Jackson and Jodie Turner...	Joshua Jackson and Jodie ...

Category	Sub-Category	Title	Abstract
movies	movies-celebrity	Jason Momoa Teases 'Way ...	The actor also gushed about...
lifestyle	lifestylebuzz	This Artist Reimagined Disney...	Princess Ariel as Pennywise?
foodanddrink	foodnews	Jennifer Lawrence Hired A Food...	Kinda mad I wasn't invited, TBH.
movies	movies-celebrity	Bruce Willis brought Demi Moore...	Demi wasn't sure how her ex...
tv	tv-celebrity	Pamela Anderson gets backlash...	Pamela Anderson is under fire...
news	elections-2020-us	The Woman Who Flipped Off ...	The cyclist who lost her job...
lifestyle	lifestyledidyouknow	Behind-the-scenes facts about...	The iconic children's show is...
news	newsus	Famed Hollywood Boulevard...	He was the Walk of Fame...
autos	autosenthusiasts	State Trooper Stops Banana Car...	Is it a crime to be this...

Recommendation model

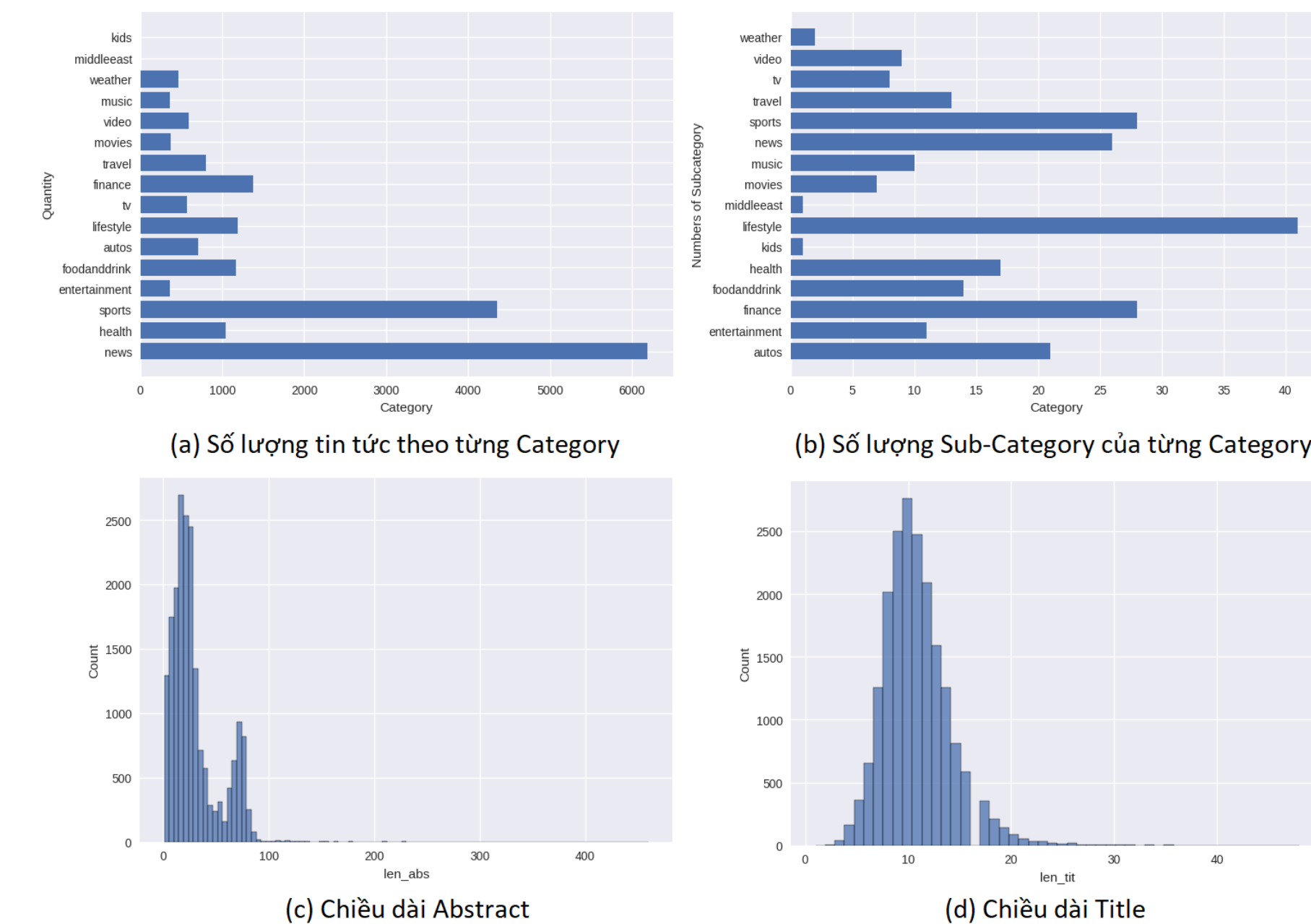


BỘ DỮ LIỆU

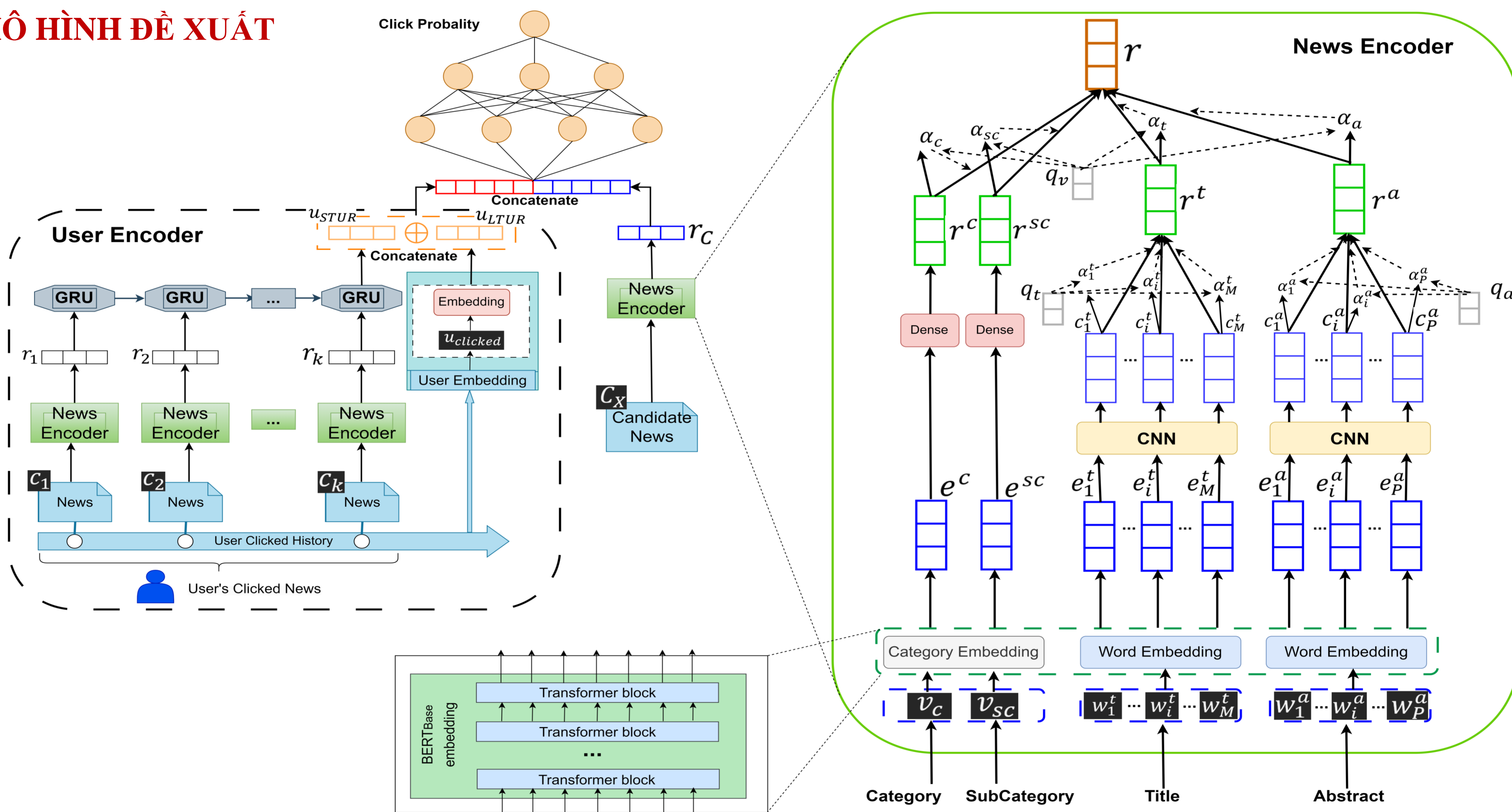
MINDtiny (Source: Author): Được trích xuất từ bộ dữ liệu MINDsmall (tỉ lệ: 1:15)

- 5000 behaviors in the training set.
- 2500 behaviors in the test set.

#News 19557
#Category 16
#Sub-Category 23
#User 4834
Avg. Abstract len. 30.02
Avg. Title len. 10.88



MÔ HÌNH ĐỀ XUẤT



TÀI LIỆU THAM KHẢO

- [1] Wu, Fangzhao, et al. "Mind: A large-scale dataset for news recommendation." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.
- [2] Wu, Chuhuan, et al. "Neural news recommendation with attentive multi-view learning." arXiv preprint arXiv:1907.05576 (2019).
- [3] An, Mingxiao, et al. "Neural news recommendation with long- and short-term user representations." Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019.

THỬ NGHIỆM

Hàm dự đoán và Hàm tính lỗi:

- Hàm predict được sử dụng để tính xác suất phù hợp của **Vector đặc trưng của user** (u) với **Vector tin tức ứng viên** (r_c):

$$\hat{y} = u^T r_c$$

- Hàm tính lỗi được tính toán dựa trên **Pseudo-rank score** và **Cross-entropy** để tính lỗi giữa **Tin tức được người dùng nhấn** (\hat{y}_i^+) và **K Tin tức không được người dùng nhấn** ($[\hat{y}_1^- + \hat{y}_2^- + \dots + \hat{y}_K^-]$):

$$p_i = \frac{\exp(\hat{y}_i^+)}{\exp(\hat{y}_i^+) + \sum_{j=1}^K \exp(\hat{y}_{i,j}^-)}; \mathcal{L} = -\sum_{i \in S} \log(p_i)$$

Các độ đo sử dụng:

- Area under curve (AUC)
- Mean reciprocal rank: (MRR)

$$MRR = \frac{1}{|C_{test}|} \sum_{s \in C_{test}} \frac{1}{rank(s)}$$

- Normalized discounted cumulative gain (NDCG@K)

$$NDCG@K = Z_K \sum_{i=1}^K \frac{2^{r(i)} - 1}{\log_2(i + 1)}$$

BẢNG KẾT QUẢ ĐÁNH GIÁ CÁC MÔ HÌNH

Model	AUC	MRR	nDCG@5	nDCG@10
DKN	0.5854	0.2659	0.2909	0.3518
NAML	0.6471	0.3119	0.3465	0.4028
LSTUR	0.5372	0.2369	0.2459	0.3099
TANR	0.5778	0.2527	0.2759	0.336
Ours:				
NAML+LSTUR	0.6195	0.2528	0.2675	0.3356
NAML+LSTUR-nn	0.5459	0.2371	0.2512	0.3070
NAML-nn	0.4956	0.2040	0.2133	0.2743

KẾT LUẬN

Kết quả đạt được:

- Đề xuất mô hình mới (NAML-LSTUR) cho bài toán NewsR.
- Đề xuất các mô hình biến thể khác của NAML-LSTUR.
- MINDtiny cho việc huấn luyện bài toán NewsR với ít tài nguyên.

Hướng phát triển:

- Hoàn thành huấn luyện mô hình với định hướng BERT-base.
- Tìm giải pháp cải thiện "hàm dự đoán tin tức ứng viên" hiệu quả hơn.