DS204.N21

DS002

Ex-MINER

Extension on MINER with Improved Encoder for News Recommendation

Danh sách thành viên

20521609 – Nguyễn Hoàng Minh

20521614 – Tạ Nhật Minh

20521998 – Nguyễn Thiện Thuật

Nhóm 12

24hDeadline

NỘI DUNG

01 Background

02 Ex-MINER

03 Tổng kết

THÔNG TIN PAPER

BACKGROUND

01

Paper 01

ACL Anthology FAQ Corrections Submissions Search...

MINER: Multi-Interest Matching Network for News Recommendation

Jian Li, Jieming Zhu, Qiwei Bi, Guohao Cai, Lifeng Shang, Zhenhua Dong, Xin Jiang, Qun Liu

Abstract

Personalized news recommendation is an essential technique to help users find interested news. Accurately matching user's interests and candidate news is the key to news recommendation. Most existing methods learn a single user embedding from user's historical behaviors to represent the reading interest. However, user interest is usually diverse and may not be adequately modeled by a single user embedding. In this paper, we propose a poly attention scheme to learn multiple interest vectors for each user, which encodes the different aspects of user interest. We further propose a disagreement regularization to make the learned interests vectors more diverse. Moreover, we design a category-aware attention weighting strategy that incorporates the news category information as explicit interest signals into the attention mechanism. Extensive experiments on the MIND news recommendation benchmark demonstrate that our approach significantly outperforms existing state-of-the-art methods.

Anthology ID: 2022.findings-acl.29
Volume: Findings of the Association for Computational Linguistics: ACL 2022
Month: May
Year: 2022
Address: Dublin, Ireland

Paper 02

MIND: A Large-scale Dataset for News Recommendation

Fangzhao Wu, Ying Qiao, Jiun-Hung Chen, Chuhan Wu, Tao Qi, Jianxun Lian, Danyang Liu, Xing Xie, Jianfeng Gao, Winnie Wu, Ming Zhou

Abstract

News recommendation is an important technique for personalized news service. Compared with product and movie recommendations which have been comprehensively studied, the research on news recommendation is much more limited, mainly due to the lack of a high-quality benchmark dataset. In this paper, we present a large-scale dataset named MIND for news recommendation. Constructed from the user click logs of Microsoft News, MIND contains 1 million users and more than 160k English news articles, each of which has rich textual content such as title, abstract and body. We demonstrate MIND a good testbed for news recommendation through a comparative study of several state-of-the-art news recommendation methods which are originally developed on different proprietary datasets. Our results show the performance of news recommendation highly relies on the quality of news content understanding and user interest modeling. Many natural language processing techniques such as effective text representation methods and pre-trained language models can effectively improve the performance of news recommendation. The MIND dataset will be available at https://msnews.github.io.

PDF Cite Search Code Video

Anthology ID: 2020.acl-main.331
Volume: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics
Month: July
Year: 2020
Address: Online
Venue: ACL

MÔ TẢ TASK

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.

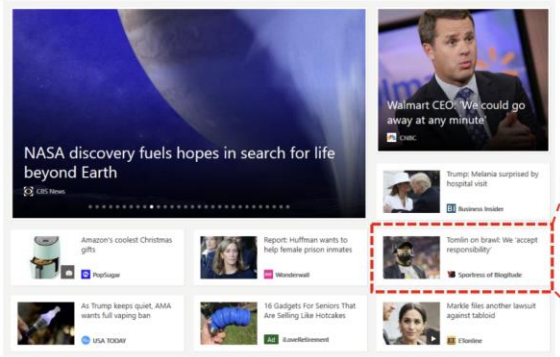
Recommended by NRMS+BERT	
Sports	NFL week 8 power rankings: old-school football rules the day.
Sports	Patriots wanted a test. Now, they need some answers.
Politics	40 conservative groups sign ethics complaint against Pelosi.

Recommended by MINER	
Sports	Patriots wanted a test. Now, they need some answers.
Food	**National Dessert Day: Where to get free dessert at Wendy's.**
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.

MIND

Dataset

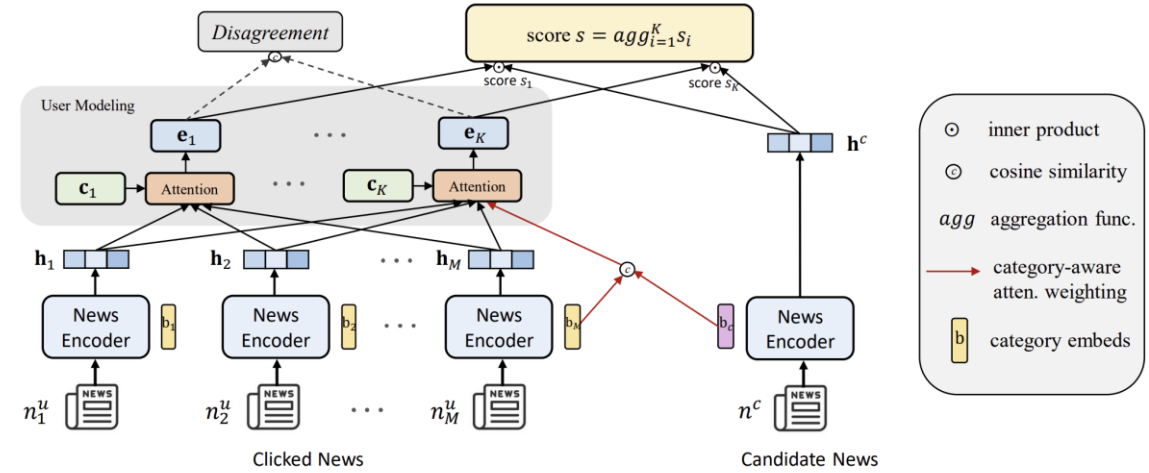


Title	Mike Tomlin: Steelers 'accept responsibility' for role in brawl with Browns
Category	Sports
Abstract	Mike Tomlin has admitted that the Pittsburgh Steelers played a role in the brawl with the Cleveland Browns last week, and on Tuesday he accepted responsibility for it on behalf of the organization.
Body	Tomlin opened his weekly news conference by addressing the issue head on. "It was ugly," said Tomlin, who had refused to take any questions about the incident directly after the game, per Brooke Pryor of ESPN. "It was ugly for the game of football. I think all of us that are involved in the game, particularly at this level, ...

(a) An example Microsoft News homepage

(b) Texts in an example news article

Figure 1: An example homepage of Microsoft News and an example news article on it.



MINER

Kiến trúc

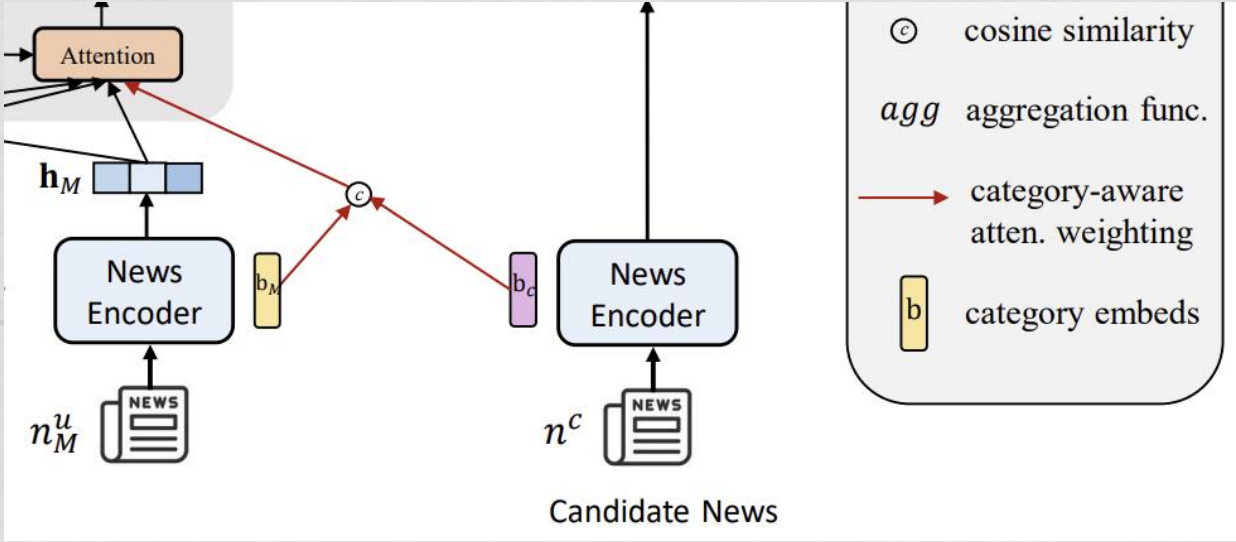
NHỮNG VẤN ĐỀ TRONG KIẾN TRÚC CỦA MINER

1. News Encoder

- **Chỉ sử dụng mỗi [CLS] để làm vec-tơ đặt trưng.**
- **Chưa thử nghiệm với các mô hình pretrain LM mới nhất** *(Chỉ thực hiện với các mô hình shallow word embeddings (GLOVE, Faxtext, W2V...) và BERT base, non-BERT: HieRec).*

2017). In this paper, we adopt the pre-trained BERT (Devlin et al., 2019) as news encoder, which can effectively capture the deep semantics of news texts. BERT has been successfully applied in various text ranking problems (Khattab and Zaharia, 2020; Karpukhin et al., 2020). Specifically, we feed tokenized news text into BERT model and use the output of `[CLS]` token as the news embedding $\mathbf{h}$. Thus the user u and candidate news n^c are encoded as $\mathbf{H}^u = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_M]$ and $\mathbf{h}^c$, respectively.

2. Category-aware Attention Weighting



Chưa khai thác hết thông tin khác từ bộ dữ liệu như SubCategory và Title.
(những thuộc tính này cũng quan trọng ảnh hưởng tới việc người dùng có xem hay không)

3. Cách chọn dữ liệu huấn luyện của MINER

Dùng Title cho News Encoder *(không phù hợp)*

Settings Following previous work (Wu et al., 2019b; Qi et al., 2021), we utilize users' most recent 50 clicked news to learn user representations. We only use news title for the experiments in this paper and the maximum length is set to 20.[5] The *bert-base-uncased* is used as the pre-trained model to initialize news encoders. The number of context

⇛ Đề xuất mô hình mới phát triển mở rộng mô hình MINER có tên là Ex-MINER cải thiện encoder từ văn bảng.

02 EX - MINER

Ex-MINER: Extension on MINER with Improved Encoder for News Recommendation

GIỚI THIỆU ĐỀ TÀI

- Ex-MINER được đề xuất nhằm giải quyết 3 vấn đề mà nhóm nhận thấy MINER vẫn còn hạn chế
- **Mục tiêu:** cải thiện việc trích xuất đặt trưng từ văn bảng hiệu quả hơn mô hình MINER
- **Bộ dữ liệu sử dụng để so sánh:** MINDsmall

GIẢI PHÁP PHÁT TRIỂN

Fine-tuning News Encoder

- **Thực nghiệm:** Dùng full feature BERT (chứ ko chỉ CLS), ALBERT, XLM-RoBERTa, GPT2.
- MINER dùng Title => **Ex-MINER dùng Abstract**

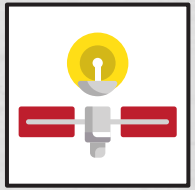
Phát triển Category-aware Attention Weighting

- MINER dùng Category => **Ex-MINER dùng Category + SubCategory**
- **Ex-MINER tích hợp Title với Category:**
  - Lấy Category Cộng (add) với Title và lấy trung bình của vec-tơ đó (tỉ lệ 2:1)
  - Lấy Category Concat với Title.

Dùng các kỹ thuật để trích xuất nội dung Abstract tốt hơn

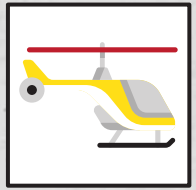
- Pretrain text summarize (BART,...), NER, POS,...

03 Tổng kết



Tình hình hiện tại

- Hoàn thành nghiên cứu về bài báo MINER.
- Đề xuất được những giải pháp cải tiến mô hình MINER.
- Trong giai đoạn nghiên cứu source code của bài báo.



Khó khăn

- Khó khăn trong việc đọc hiểu code của bài báo.
- Mô hình có thể sử dụng lượng lớn tài nguyên cho việc tính toán *(Vì mô hình phải tính qua từng tin tức ứng với mỗi user).*

THANK YOU!

Ex-MINER: Extension on MINER with Improved Encoder for News Recommendation