

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

Nguyễn Hoàng Minh^{1,2} Nguyễn Thiện Thuật^{1,2} Tạ Nhật Minh^{1,2}

¹ Trường Đại học Công Nghệ Thông Tin, Hồ Chí Minh, Việt Nam

² Đại học Quốc Gia, TP. Hồ Chí Minh, Việt Nam
{20521609, 20521998, 20521614}@gm.uit.edu.vn

Abstract

Đề xuất tin tức là một kỹ thuật rất cần thiết để giúp người dùng có thể dễ dàng tìm thấy các tin tức mà họ quan tâm và giảm bớt tình trạng quá tải thông tin. Việc học các tin tức và đặc điểm của người dùng là rất quan trọng cho hệ thống đề xuất tin tức. Các phương pháp đề xuất tin tức hiện có thường học tập các đặc điểm tin tức dựa trên thông tin tin tức đơn lẻ (ví dụ: tiêu đề, loại tin tức...), nó có thể không đủ để hiểu toàn bộ nội dung tin tức đó. Đồng thời, người dùng thường có cả sở thích dài hạn và sở thích ngắn hạn. Tuy nhiên, các phương pháp đề xuất tin tức hiện tại thường học các sở thích đơn lẻ của người dùng và điều này có thể không đủ. Trong đề tài này, chúng tôi đề xuất mô hình khuyến nghị có tên là NAML-LSTUR cho bài toán gợi ý tin tức với sự kết hợp của hai mô hình state-of-the-art trong lĩnh vực là mô hình NAML và LSTUR. Mục tiêu của mô hình này là nhằm giúp cho hệ thống học tốt nhất các thông tin của tin tức một cách toàn diện (ví dụ: tiêu đề, tóm tắt tin tức, loại tin tức...) và học được thói quen của người dùng cả ngắn hạn và dài hạn. Thêm vào đó, mô hình của chúng tôi cũng đề xuất sử dụng word embedding theo định hướng BERT-base nhằm trích xuất tốt hơn đặc điểm của tin tức. Kết quả thực nghiệm trên tập dữ liệu MINDtiny cho thấy mô hình đề xuất của chúng tôi cao hơn phần lớn các mô hình state-of-the-art hiện tại.

Từ khóa: News recommendation, NAML-LSTUR, Attentive Multi-View Learning, Long- and Short-term User Representations

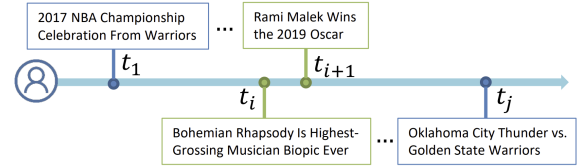
I. Introduction

Các nền tảng tin tức trực tuyến như Google News và Microsoft News đã trở thành một phương thức phổ biến để người dùng cập nhật tin tức mới nhất trong ngày. Số lượng lớn tin tức được tạo ra mỗi ngày khiến người dùng khó tìm thấy những tin tức mà họ quan tâm. Chính vì vậy, để giảm bớt tình trạng quá tải thông tin và cải thiện trải nghiệm người dùng, hệ thống đề xuất tin tức đã trở thành một phần rất cần thiết của các nền tảng này.

Tiêu đề của tin tức đầu tiên trong Hình 1 là chính xác và quan trọng để đại diện cho tin tức này, trong khi tiêu đề của tin tức thứ hai ngắn, mơ hồ và ít thông tin hơn. Bên cạnh đó, các từ khác nhau trong cùng một tin tức có thể có tầm quan trọng khác nhau, các từ được đánh dấu đỏ đại diện cho các từ quan trọng trong tiêu đề và nội dung tin tức. Trong Hình 2, nếu người dùng là người hâm mộ “Golden

Category	Sports	Entertainment
Title	Astros improve outfield, agree to 2-year deal with Brantley	The best games of 2018
Body	Outfielder Michael Brantley agreed to a two-year, \$32 million contract with Houston, sources familiar with the deal told Yahoo Sports, bringing his steady left-handed bat to the top of an Astros ...	The Best Games of 2018 Superheroes, super-dads, and Super Mario parties brought the joy in 2018 to millions of players who ate up the year's impressive achievements in gaming...

Hình 1: Ví dụ về 2 tin tức của một người dùng đã xem.



Hình 2: Ví dụ về sở thích dài và ngắn hạn về việc đọc tin tức.

State Warriors”, người dùng này có thể có xu hướng đọc nhiều tin tức bóng rổ về đội NBA này trong vài năm. Chúng tôi gọi loại sở thích này của người dùng là sở thích dài hạn. Ngoài ra, nhiều sở thích của người dùng có thể phát triển theo thời gian và có thể được tác động bởi các bối cảnh cụ thể hoặc nhu cầu tạm thời, như trong Hình 2, việc duyệt tin tức về bộ phim “Bohemian Rhapsody” khiến người dùng đọc một số tin tức liên quan như “Rami Malek Giành giải Oscar 2019” vì “Rami Malek” là một diễn viên quan trọng trong bộ phim này, mặc dù người dùng này có thể chưa bao giờ đọc tin tức về “Rami Malek” trước đây. Chúng tôi gọi loại sở thích này của người dùng là sở thích ngắn hạn. Do đó, cả sở thích của người dùng dài hạn và ngắn hạn đều quan trọng đối với việc đề xuất tin tức và việc phân biệt sở thích của người dùng dài hạn với sở thích ngắn hạn có thể giúp tìm ra được các sở thích của người dùng được chính xác hơn.

Việc kết hợp chính xác giữa sở thích của người dùng và tin tức ứng cử viên là chìa khóa để đề xuất tin tức tốt nhất đến người dùng. Các phương pháp hiện tại thường tìm cách trích xuất ra một vectơ sở thích của người dùng bằng cách tổng hợp các tin tức đã đọc trước đó thông qua các mô hình deep learning, sau đó so khớp nó với các vectơ tin tức ứng cử viên để tìm ra xác suất tin tức đó có phù hợp với người dùng hay không. Ví dụ các tác giả [1] đã đề xuất một phương

pháp sử dụng mạng nơ-ron, CNN và GLOVE để học các đặc trưng thông tin của người dùng và tin tức bằng cách khai thác các loại thông tin khác nhau trong tin tức. Các tác giả [2] đề xuất trích xuất các đặc trưng sở thích của người dùng cả dài hạn và ngắn hạn thông qua mạng nơ-ron và GRU, các tác giả [3] đã đề xuất sử dụng knowledge graph representation và KCNN để trích xuất đặc trưng của tin tức và Attention Net để thể trích xuất đặc trưng của người dùng thông qua các tin tức họ đã xem. Tuy nhiên, các mô hình này vẫn chỉ giải quyết tốt nhất một phần nhiệm vụ nhất định, những tác vụ còn lại không hiệu quả bằng các mô hình khác. Nhận thấy điều này, chúng tôi đề xuất một mô hình có thể kết hợp các điểm mạnh từ các phần của mô hình khác với nhau với mục đích sử dụng tốt nhất điểm mạnh của chúng cho từng nhiệm vụ trong mô hình đề xuất.

Trong đề tài này, chúng tôi đề xuất “A combined model with Attentive Multi-view Learning and Long- and Short-term User Representations for News Recommendation: **NAML-LSTUR**”, là một mô hình kết hợp của hai mô hình state-of-the-art của bài toán là NAML và LSTUR, có nhiệm vụ học tập nội dung từ tin tức thông qua nhiều thông tin khác nhau của tin tức (giống NAML) và học các đặc điểm của người dùng thông qua hai đặc trưng: sở thích ngắn hạn (thông qua các tin tức mà người dùng đã đọc) và sở thích dài hạn (thông qua lịch sử người dùng đã nhấn) theo thời gian. Ngoài ra chúng tôi cũng lấy cảm hứng từ mô hình DKN cho việc xây dựng hàm dự đoán tin tức ứng viên thông qua mạng nơ-ron. Hơn thế nữa, nghiên cứu của chúng tôi cũng đề xuất sử dụng word embedding theo định hướng BERT-base nhằm tăng mức độ hiệu quả của việc trích xuất đặc trưng từ tin tức.

Những đóng góp chính của nghiên cứu này có thể được tóm tắt như sau:

- Đề xuất mô hình mới có tên là NAML-LSTUR cho bài toán gợi ý tin tức.
- Đề xuất các mô hình biến thể khác của NAML-LSTUR.
- Bộ dữ liệu MINDtiny cho việc huấn luyện bài toán đề xuất tin tức với ít tài nguyên.

II. Related work

Bài toán đề xuất tin tức (News recommendation)

Mục đích của các hệ thống đề xuất tin tức là tìm các bài báo mà người dùng quan tâm thông qua các tin tức ứng cử viên [4]. Có hai mục tiêu quan trọng trong đề xuất tin tức: cách biểu diễn các tin tức và cách mô hình hóa mức độ quan tâm của người dùng đối với tin tức từ các hành vi trước đây của họ [5]. Đề xuất tin tức có mối quan hệ mật thiết và đa dạng với xử lý ngôn ngữ tự nhiên. Bởi vì, thứ nhất, tin tức là một dạng văn bản phổ biến và các kỹ thuật mô hình hóa văn bản như CNN và Transformer có thể được áp dụng để thể hiện đặc trưng của tin tức. Thứ hai, việc học các đặc trưng sở thích của người dùng từ các tin tức họ đã xem tương đồng với việc học các đặc trưng từ tin tức từ các câu. Do đó, khuyến nghị tin tức đã thu hút sự chú ý ngày càng tăng trong cộng đồng xử lý ngôn ngữ tự nhiên [2].

Các phương pháp đề xuất truyền thống

Các phương pháp đề xuất tin tức truyền thống thường dựa vào feature engineering để thể hiện các bài báo và sự quan

tâm của người dùng. Trong các hệ thống đề xuất, hầu hết các đặc trưng được biểu thị dưới dạng ID (ví dụ: itemID, userID,...), dẫn đến nhiều nghiên cứu tập trung vào việc mô hình hóa các tương tác của người dùng. Ví dụ: các mô hình LibFM [6] trích xuất đặc trưng các tương tác của người dùng thông qua inner products theo cặp. DeepFM [7] tiếp tục cải tiến bằng cách tích hợp shallow and deep networks. Các mô hình DCN [8] có các tương tác thông qua deep and cross sub-networks. Tuy nhiên, các phương pháp này dựa vào kỹ thuật trích xuất đặc trưng, điều này cần một lượng lớn kiến thức chuyên ngành và bộ nhớ máy tính. Ngoài ra, các phương pháp này không thể nắm bắt ngữ cảnh và nội dung tin tức (điều này rất quan trọng để hiểu tin tức và biểu diễn đặc trưng của người dùng).

Trong khi hầu hết các nghiên cứu biểu diễn người dùng thông qua một vectơ embedding, thì tác giả [9] đề xuất một phương pháp a capsule routing để trích xuất nhiều vectơ sở thích của người dùng. Tuy nhiên, mô hình này được thiết kế đặc biệt để phù hợp cho đề xuất trên sàn thương mại điện tử. Trong khi đó, chúng tôi đặt mục tiêu tìm cách thể hiện sở thích của người dùng thông qua lịch sử xem tin tức của họ theo thời gian.

Neural News Recommendation

Đối với đề xuất tin tức, các phương pháp dựa trên ID truyền thống thường gặp phải vấn đề nên bắt đầu từ đâu vì các bài báo cập nhật rất nhanh [10]. Do đó, nhiều phương pháp đã bắt đầu nghiên cứu sử dụng neural networks để tự động tìm hiểu và so khớp tin tức với các đặc trưng của người dùng [5]. Họ bắt đầu chú ý nhiều hơn đến vấn đề đề xuất tuần tự, nhằm mục đích nắm bắt hành vi tuần tự của người dùng thông qua mô hình tuần tự như là RNN, CNN và self-attention. Ví dụ, [5] đề xuất học các đặc trưng tin tức từ nội dung chính của tin tức bằng cách sử dụng denoising auto-encoders. Họ còn đề xuất học các đặc trưng của người dùng từ tin tức đã xem bằng mạng GRU. [3] đã đề xuất học các đặc trưng tin tức từ các tiêu đề của tin tức thông qua knowledge-aware CNN, kiến trúc này kết hợp thông tin trong knowledge graphs. Tuy nhiên, các phương pháp này chỉ có thể khai thác một loại thông tin trong tin tức duy nhất (title, abstract,...), điều này có thể không đủ để tìm hiểu các biểu diễn chính xác của tin tức và người dùng. Khác với các phương pháp này, cách tiếp cận của chúng tôi có thể học các đặc trưng của tin tức và người dùng bằng cách kết hợp các loại thông tin tin tức khác nhau như tiêu đề, nội dung và danh mục chủ đề thông qua phần user encode của mô hình NAML.

Các cơ chế Attention cũng đã được áp dụng rộng rãi trong đề xuất tin tức để học các tin tức và sở thích của người dùng, chẳng hạn như personalized attention networks [11] và multi-head self-attentions [12]. Những nghiên cứu gần đây cũng đã áp dụng pretrained BERT [13] để mã hóa tin tức do tính ưu việt của nó trong việc hiểu văn bản. Tuy nhiên, hầu hết các phương pháp học cách biểu diễn đặc trưng người dùng theo từng tin tức đơn lẻ, điều này có thể không mô hình hóa đầy đủ mối liên kết giữa các tin tức mà người dùng đã xem. Ngược lại, trong mô hình của chúng tôi đề xuất sử dụng user coder của LSTUR để học cách biểu diễn sở thích của người dùng thông qua mối liên kết nội dung giữa các bài báo theo thời gian.

III. MINDtiny dataset

1. MIND dataset

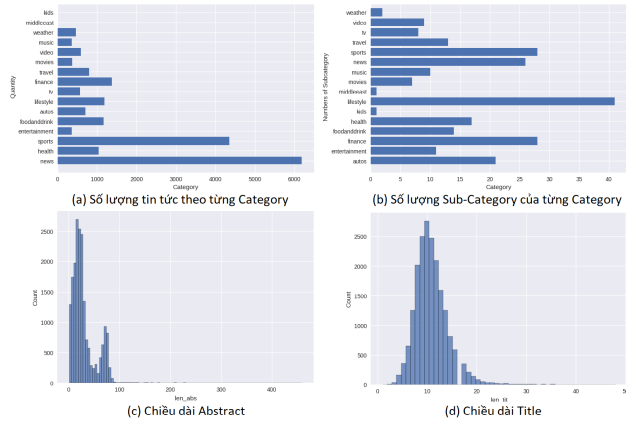
MIND (Microsoft InnerEye NLP Dataset) là bộ dữ liệu siêu lớn chứa dữ liệu liên quan đến tin tức và người dùng, được tạo ra bởi Microsoft. MIND gồm hai tập dữ liệu "Hành vi của người dùng" (User Behavior - UB) và "Bình luận" (News Comments - NC) với tổng cộng hơn 2 triệu tin tức, hơn 152.000 người dùng, hơn 100 triệu lượt xem bài và hơn 1,3 triệu bình luận từ nhiều thể loại báo khác nhau. Do đó, MIND là tài nguyên rất quý giá trong lĩnh vực Natural Language Processing (NLP), đặc biệt về chủ đề đánh giá và tương tác của người dùng với các nội dung trên web.

2. Xây dựng MINDtiny

Với kích thước dữ liệu khổng lồ, MIND có thể không phù hợp cho việc sử dụng trên những thiết bị có tài nguyên thấp. Vì vậy, để đáp ứng nhu cầu sử dụng trên những thiết bị với tài nguyên hạn chế, nhóm nghiên cứu đã tạo ra một bộ dữ liệu nhỏ có tên là MINDtiny. MINDtiny được tạo ra bằng cách lấy mẫu ngẫu nhiên một phần tỷ lệ nhỏ của dữ liệu MINDsmall, với kích thước nhỏ hơn với tỉ lệ 1:15, giúp cho các nhà nghiên cứu về NLP tiết kiệm thời gian khi đào tạo và kiểm thử các mô hình và thuật toán.

3. Phân tích bộ dữ liệu MINDtiny

Thông tin chi tiết về tập dữ liệu MINDtiny được tổng hợp trong Bảng 1 và Hình 3. Tập dữ liệu này chứa 4.844 người dùng và 19.557 tin tức. Có 16 thể loại và 23 thể loại chi tiết.



Hình 3: Thống kê chính của bộ dữ liệu MINDtiny.

#News	19557
#Category	16
#Sub-Category	23
#User	4844
Avg. Abstract len.	30.02
Avg. Title len.	10.88

Bảng 1: Thống kê chi tiết của bộ dữ liệu MINDtiny.

Bộ dữ liệu MINDtiny tập trung vào các tin tức thuộc thể loại sport khoảng 4300 tin tức và new trên 6000 tin tức. Số

lượng Sub-Category của từng thể loại (Category) phân tán ở nhiều giá trị. Độ dài của Abstract tập trung vào hai độ dài chính từ 0 - 40 word với số lượng khoảng 1500 - 2500 Abstract và từ 60 - 80 word với số lượng khoảng 400 - 900 Abstract. Độ dài của Title giao động từ 5 - 15 word với số lượng khoảng 1000 - 2500 Title. Có thể thấy tiêu đề tin tức thường rất ngắn, trung bình khoảng 10,88 từ. Trong khi đó, tóm tắt của tin tức dài hơn và có thể chứa nhiều thông tin phong phú hơn. Do đó, việc tích hợp các thông tin tin tức khác nhau như tiêu đề và tóm tắt có thể cải thiện hiệu suất.

IV. Our approach

Trong phần này, chúng tôi đề xuất mô hình NAML-LSTUR cho bài toán đề xuất tin tức. Có ba mô-đun chính: Mã hóa tin tức sử dụng news encoder của mô hình NAML, mã hóa người dùng sử dụng user encoder của mô hình LSTUR và một hàm dự đoán khả năng click để dự đoán xác suất người dùng duyệt một bài báo ứng cử viên lấy ý tưởng từ mô hình DKN. Kiến trúc tổng thể của phương pháp đề xuất của chúng tôi được thể hiện ở Hình 4.

1. Mã hóa tin tức

Để học cách biểu diễn các tin tức từ nhiều nguồn thông tin của tin tức khác nhau như tiêu đề, nội dung và chủ đề hay các chủ đề phụ, Mô-đun News Encoder đã được chúng tôi sử dụng. Với mỗi loại thông tin sẽ có các đặc điểm riêng, chúng tôi sử dụng một attentive multi-view learning framework để học cách biểu diễn thông tin một cách thống nhất bằng cách coi mỗi loại thông tin như một tầm nhìn cụ thể về tin tức. Bộ mã hóa tin tức bao gồm bốn thành phần chính, được chúng tôi mô tả trong Hình 4.

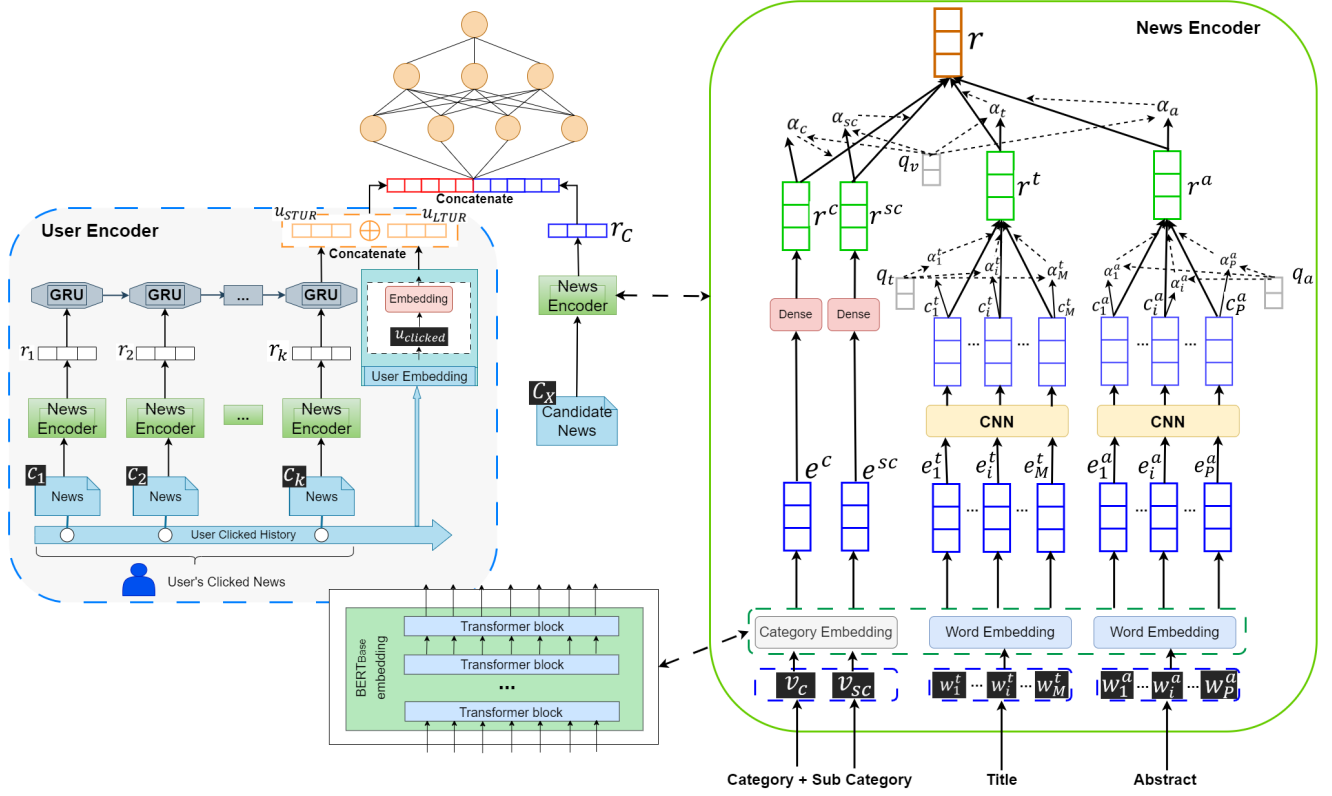
Thành phần đầu tiên của news encoder là bộ mã hóa tiêu đề (title encoder) gồm 3 lớp: lớp nhúng từ, lớp trung gian và lớp mã hóa, để học cách biểu diễn tin tức dựa trên tiêu đề. Lớp nhúng từ chuyển đổi chuỗi từ trong tiêu đề tin tức thành chuỗi các vector ngữ nghĩa có số chiều thấp, thông qua một bảng tra cứu nhúng từ. Ký hiệu cho chuỗi từ trong tiêu đề tin tức là $[w_1^t, w_2^t, \dots, w_M^t]$, với M là độ dài tiêu đề. Chúng tôi đề xuất sử dụng BERT-base để thực hiện tác vụ này.

Mạng convolutional neural network (CNN) là lớp thứ hai của bộ mã hóa tiêu đề, CNN được sử dụng để học các biểu diễn từ dựa theo ngữ cảnh của chúng ở tiêu đề tin tức. Biểu diễn ngữ cảnh của từ thứ i là c_i^t , được tính toán bởi công thức:

$$c_i^t = \text{ReLU}(F_t + e_{(i-k):(i+k)}^t + b_t) \quad (1)$$

Với $e_{(i-k):(i+k)}^t$ là quá trình ghép các embedding từ vị trí $(i-k)$ đến $(i+k)$, $F_t \in R^{N_f + (2k+1)^D}$ và $b_t \in R^{N_f}$ là thông số kernel và bias của các bộ lọc CNN với $2k+1$ là window size và N_f là số bộ lọc và hàm kích hoạt là hàm ReLU. Đầu ra của lớp này là chuỗi các biểu diễn từ theo ngữ cảnh $[c_1^t, c_2^t, \dots, c_M^t]$.

Word-level attention network là lớp thứ ba của bộ mã hóa. Đối với mỗi từ trong cùng một tiêu đề thì sẽ có tính thông tin khác nhau, vì vậy việc xác định những từ quan trọng trong tiêu đề sẽ giúp việc học các biểu diễn tin tức có tính thông tin cao hơn. Chúng tôi đã sử dụng word-level attention network



Hình 4: Kiến trúc mô hình NAML-LSTUR.

để thực hiện tác vụ này. Trọng số chú ý của từ thứ i trong tiêu đề là α_i^t , được tính theo công thức:

$$a_i^t = q_t^T \tanh(V_t \times c_i^t + v_t) \quad (2)$$

$$\alpha_i^t = \frac{\exp(a_i^t)}{\sum_{j=1}^M \exp(a_j^t)} \quad (3)$$

Với V_t và v_t là các tham số chiều hình, q_t đại diện cho vectơ truy vấn. $r^t = \sum_{j=1}^M \alpha_j^t c_j^t$ là kết quả cuối cùng của bộ mã hóa tiêu đề (Title Encoder).

Thành phần thứ hai trong News Encoder là bộ mã hóa tóm tắt (Abstract Encoder), được dùng để thực hiện tác vụ biểu diễn tin tức từ phần tóm tắt của chúng. Bộ mã hóa tóm tắt có kiến trúc tương tự với bộ mã hóa tiêu đề với ba lớp.

Lớp thứ nhất là lớp nhúng từ, tương tự với bộ mã hóa tiêu đề, BERT-base được sử dụng cho tác vụ này. Chuỗi từ của một phần tóm tắt được biểu diễn bởi $w_1^abs, w_2^abs, \dots, w_P^abs$ với P là chiều dài của phần tóm tắt. Lớp nhúng từ có nhiệm vụ chuyển đổi nó thành một chuỗi vectơ từ $[e_1^abs, e_2^abs, \dots, e_P^abs]$. Lớp thứ hai là mạng CNN, có nhiệm vụ lấy chuỗi vectơ đầu vào và học cách biểu ngữ cảnh từ đó bằng cách bắt chước các ngữ cảnh cục bộ. Lớp cuối cùng là một mạng chú ý. Trọng số chú ý của từ thứ i trong tóm tắt là α_i^a , được tính theo công thức:

$$a_i^abs = q_{abs}^T \tanh(V_{abs} \times c_i^abs + v_{abs}) \quad (4)$$

$$\alpha_i^abs = \frac{\exp(a_i^abs)}{\sum_{j=1}^P \exp(a_j^abs)} \quad (5)$$

với V_{abs} và v_{abs} là các tham số chiều hình, q_{abs} đại diện cho vectơ truy vấn.

Sự tổng hợp các biểu diễn ngữ cảnh của các từ với trọng số là trọng số chú ý, $r^abs = \sum_{j=1}^P \alpha_j^abs c_j^abs$ là kết quả cuối cùng của bộ mã hóa tóm tắt.

Thành phần cuối cùng của News Encoder là bộ mã hóa thể loại (Category Encoder). Bộ mã hóa này được sử dụng với mục đích biểu diễn các tin tức từ các thể loại tương ứng. Ở đây chúng tôi kết hợp thông tin từ các thể loại và nhóm thể loại phụ để học cách biểu diễn tin tức. Đầu vào là các ID của thể loại v_c và thể loại phụ v_{sc} , với 2 lớp là lớp nhúng thể loại (Category Embedding – sử dụng BERT-base) và lớp mật độ (Dense), các thể loại tin tức sẽ được chuyển đổi và biểu diễn bằng cách chuyển đổi Category Embedding thông qua:

$$r^c = \text{ReLU}(V_c \times \exp(c) + v_c) \quad (6)$$

$$r^{sc} = \text{ReLU}(V_{sc} \times \exp(sc) + v_{sc}) \quad (7)$$

Với V_c, v_c, V_{sc}, v_{sc} là các tham số trong lớp mật độ.

Thành phần cuối trong bộ mã hóa tin tức là mô-đun attentive pooling. Chúng tôi chọn một mạng view-level attention để mô hình hóa tính thông tin từ các loại thông tin khác nhau cho tác vụ biểu diễn các tin tức. Lấy $\alpha_t, \alpha_{abs}, \alpha_c, \alpha_{sc}$ lần lượt là các trọng số chú ý của tiêu đề, phần tóm tắt, thể

loại và thể loại phụ. Trọng số chú ý của tiêu đề được tính theo công thức:

$$a_t = q_v^T \tanh(U_v \times r^t + u_v) \quad (8)$$

$$\alpha_t = \frac{\exp(a_t)}{\exp(a_t) + \exp(a_{abs}) + \exp(a_c) + \exp(a_{sc})} \quad (9)$$

với U_v và u_v là các tham số chiều, q_v là vectơ truy vấn chú ý.

Biểu diễn tin tức thống nhất cuối cùng mà bộ mã hóa tin tức học được là tổng của các biểu diễn tin tức từ các góc nhìn khác nhau được trọng số bằng trọng số chú ý của chúng:

$$r = \alpha_c r_c + \alpha_{sc} r_{sc} + \alpha_t r_t + \alpha_{abs} r_{abs} \quad (10)$$

2. Mã hóa người dùng

Bộ mã hóa thông tin người dùng được dùng để học cách mà người dùng thường xuyên đọc các tin tức từ lịch sử của họ. Nó được tạo ra từ hai bộ phận, gồm một mô hình ngắn hạn (STUR) để lưu trữ ngắn hạn sở thích của người dùng và một mô hình dài hạn (LTUR) để lưu trữ những sở dài hạn của họ. Chúng tôi sẽ trình bày chi tiết hơn về từng bộ phận trên.

2.1. Short-Term User Representation

Các người dùng có thể có nhu cầu ngắn hạn khi đọc các tin tức, có thể bị ảnh hưởng bởi ngữ cảnh cụ thể hoặc yêu cầu thông tin tạm thời. Vì vậy, chúng tôi chọn học các biểu diễn ngắn hạn từ người dùng thông qua thời gian gần đây để hiểu được sở thích tạm thời của họ. Sử dụng mạng gated recurrent networks (GRU) để hiểu các mẫu đọc tin tức liên tục. Chuỗi duyệt tin tức từ người dùng được sắp xếp theo thời gian dưới dạng tăng dần là $C = \{c_1, c_2, \dots, c_k\}$ với k là độ dài của chuỗi này. Bộ mã hóa tin tức giúp chúng tôi có được các biểu diễn của các bài báo đã duyệt qua $\{e_1, e_2, \dots, e_k\}$. Các biểu diễn cho người dùng ngắn hạn được biểu diễn như sau:

$$r_t = \sigma(W_r[h_{t-1}, e_t]) \quad (11)$$

$$z_t = \sigma(W_z[h_{t-1}, e_t]) \quad (12)$$

$$h_t = z_t \cdot h_t + (1 - z_t) \cdot \tanh(W_h[r_t \cdot h_{t-1}, e_t]) \quad (13)$$

Với σ là hàm sigmoid, là tích phân tử của vectơ, W_r, W_z, W_h , là các tham số của mạng GRU. Biểu diễn người dùng ngắn hạn là trạng thái ẩn cuối cùng của mạng GRU $u_s = h_k$.

2.2. Long-Term User Representations

Bên cạnh sở thích ngắn hạn tạm thời, người dùng cũng có thể có sở thích dài hạn trong việc đọc tin tức trực tuyến. Vì vậy, chúng tôi tiếp tục sử dụng việc học biểu diễn dài hạn của người dùng. Các biểu diễn dài hạn được học từ quá trình nhúng ID của người dùng. Với u là ID của người dùng, W_u là bảng tra cứu các biểu diễn dài hạn, $u_l = W_u[u]$.

2.3. Long- and Short-Term User Representation

Trong đề tài này, chúng tôi sử dụng kết hợp cả phương pháp biểu diễn người dùng ngắn hạn và dài hạn. Cả hai được kết hợp để tạo ra biểu diễn người dùng thống nhất như Hình 4. Chúng tôi nối các biểu diễn người dùng dài hạn và ngắn hạn với nhau để tạo thành biểu diễn cuối cùng của người dùng, mô hình được thể hiện tại User Encoder trong Hình 4.

3. Dự đoán khả năng click của người dùng

3.1. Dot Product

Mô hình dự đoán khả năng nhấp chuột được sử dụng để dự đoán xác suất người dùng duyệt tin tức ứng viên dựa trên các biểu diễn của họ. Ký hiệu biểu diễn của một tin tức ứng viên là D_c và biểu diễn của người dùng u là u . Theo tác giả [5], điểm xác suất nhấp \hat{y} được tính bằng tích vô hướng của vectơ biểu diễn của người dùng u và tin tức ứng viên D_c , $\hat{y} = u^T \times r_c$. Chúng tôi đã nghiên cứu và thấy rằng, tích vô hướng không chỉ là phương pháp có hiệu quả thời gian tốt nhất mà còn là phương pháp có hiệu quả tính toán tốt nhất, điều này đồng quan điểm với tác giả [5].

3.2. Neural nơ-ron network

Các lịch sử được nhấp có thể tác động khác nhau đối với tin tức đề xuất. Để mô tả sự đa dạng về sở thích của người dùng, chúng tôi sử dụng mạng chú ý (attention network) để mô hình hóa các tác động khác nhau của tin tức được nhấp của người dùng đối với tin tức đề xuất. Mạng chú ý được minh họa trong phần trên cùng của Hình 4. Cụ thể, đối với news được nhấp chọn t_k^i của user i và news đề xuất t_j , trước tiên nối các embedding, sau đó áp dụng một mạng nơ-ron H để tính toán trọng số tác động chuẩn hóa bằng hàm softmax:

$$s_{t_k^i, t_j} = \frac{\exp(H(e(t_k^i), e(t_j)))}{\sum_{k=1}^{N_i} \exp(H(e(t_k^i), e(t_j)))} \quad (14)$$

Mạng attention H nhận các embedding từ news và user làm đầu vào và đầu ra trọng số tác động. Embedding của user i đối với news đề xuất t_j được tính toán bằng tổng trọng số của những news được nhấp của user đó. Cuối cùng, với embedding của user i $e(i)$ và của news đề xuất t_j $e(t_j)$, xác suất user i nhấp vào news t_j được dự đoán bởi một mạng nơ-ron G:

$$p_{i, t_j} = G\left(\sum_{k=1}^{N_i} s_{t_k^i, t_j} e(t_k^i)\right) \quad (15)$$

4. Huấn luyện mô hình

Hàm tính lỗi

Theo nghiên cứu trước đó [1], chúng tôi cũng sử dụng hàm NCE loss để huấn luyện. Ứng với mỗi tin tức đã được nhấn bởi người dùng được coi là mẫu dương tính (\hat{y}_i^+), chúng tôi lấy mẫu ngẫu nhiên K tin tức không được người dùng này nhấn vào được coi là mẫu âm tính ($\hat{y}_1^-; \hat{y}_2^-; \dots; \hat{y}_K^-$). Sau đó chúng tôi dự đoán xác suất người dùng đó với tập tin tức (S) có cả mẫu dương tính và mẫu âm tính. Bằng cách này, chúng tôi sử dụng Pseudo-rank score đã được nêu ở nghiên cứu [14] sử dụng cho việc tính lỗi dự đoán từ mô hình, công thức Pseudo-rank score được phát biểu như sau:

$$p_i = \frac{\exp(\hat{y}_i^+)}{\exp(\hat{y}_i^+) + \sum_{j=1}^K \exp(\hat{y}_{i,j}^-)} \quad (16)$$

Sau đó, hàm NCE loss dựa trên Pseudo-rank score và Cross-entropy trên tập dữ liệu S, được biểu diễn như sau:

$$L = - \sum_{i \in S} \log(p_i) \quad (17)$$

Chiến lược huấn luyện

Để tiến hành thử nghiệm mô hình NAML-LSTUR, mô hình của chúng tôi chia làm 3 mô hình thử nghiệm tương ứng với ba biến thể của mô hình NAML-LSTUR để tiến hành so sánh với các mô hình khác trong bài toán:

- NAML-LSTUR: Đối với mô hình này, Chúng tôi sử dụng Mã hóa tin tức của NAML, Mã hóa người dùng của LSTUR và Hàm dự đoán khả năng click là Dot Product.
- NAML-LSTUR-nn: Mã hóa tin tức và Mã hóa người dùng tương tự như mô hình trên nhưng Hàm dự đoán khả năng click chúng tôi sử dụng mạng nơ-ron.
- NAML-nn: Mã hóa tin tức và Mã hóa người dùng chúng tôi sử dụng kiến trúc mô hình NAML và Hàm dự đoán khả năng click chúng tôi sử dụng mạng nơ-ron.

V. Experiments

1. Cài đặt thử nghiệm

1.1. Dữ liệu

Chúng tôi tiến hành đánh giá phương pháp của mình trên bộ dữ liệu MIND. Bộ dữ liệu gồm hai phiên bản: MIND-large chứa hơn 15 triệu nhật ký hiển thị được tạo từ 1 triệu người dùng và MIND-small lấy mẫu ngẫu nhiên 50.000 người dùng. Ở đây chúng tôi trích xuất từ bộ MIND-small để có được MIND-tiny chứa 5000 nhật ký được tạo ra từ 4838 người dùng để phù hợp với quá trình huấn luyện của chúng tôi. Một nhật ký hiển thị ghi lại tin tức đã được click và không được click mà được hiển thị cho một người dùng tại thời điểm nhất định lịch sử click tin tức của họ trước khi nhận được nhật ký này. Các thống kê được ghi lại ở Bảng 1.

1.2. Cài đặt

Chúng tôi sử dụng 50 tin tức được người dùng nhập vào gần đây để học các biểu diễn người dùng. Chúng tôi sử dụng tiêu đề tin tức cho các thí nghiệm trong bài báo này và độ dài tối đa được đặt là 20, đối với phần tóm tắt tin tức thì độ dài tối đa là 50 và số mã ngữ cảnh K là 2. Mạng CNN được sử dụng để học biểu diễn các tin tức với số bộ lọc là 300 và kích thước cửa sổ là 3. Các nhúng thể loại được khởi tạo bởi các vectơ GloVe 100 chiều và được giữ nguyên trong quá trình huấn luyện. Tỷ lệ negative sampling L được đặt là 2 trong quá trình huấn luyện, mỗi tin tức tích cực sẽ được ghép cặp với 2 tin tức tiêu cực. Chúng tôi thực hiện huấn luyện trong 2 epoch với kích thước lô là 128 với tốc độ học là e^{-4} . Chúng tôi sử dụng Adam (Kingma và Ba, 2015) làm thuật toán tối ưu hóa. Tương tự các nghiên cứu trước đó chúng tôi sử dụng bốn chỉ số xếp hạng là AUC, MRR, nDCG@5 và nDCG@10, để đánh giá hiệu suất của mô hình.

2. Các độ đo đánh giá hiệu suất mô hình

2.1. AUC

AUC (Area Under the ROC Curve) là một phương pháp đánh giá hiệu quả của mô hình phân loại. ROC (Receiver Operating Characteristic) là một đường cong biểu diễn hiệu suất của một mô hình phân loại trong việc phân biệt giữa hai lớp dữ liệu khác nhau. AUC là diện tích nằm dưới đường cong ROC và có giá trị nằm trong khoảng từ 0 đến 1.

2.2. MRR

MRR thường được sử dụng để đánh giá vị trí của ứng cử viên chính xác đầu tiên, MRR score được tính toán theo công thức sau:

$$MRR = \frac{1}{|C_{test}|} \sum_{s \in C_{test}} \frac{1}{rank(s)} \quad (18)$$

2.3. nDCG@K

NDCK được sử dụng rộng rãi trong các hệ thống khuyến nghị. Nó xem xét vị trí xếp hạng của ứng cử viên và mong muốn các ứng cử viên có liên quan cao xuất hiện sớm hơn trong danh sách top k. nDCG@k được tính toán như sau:

$$nDCG@k = Z_K \sum_{i=1}^k \frac{2^{r(i)} - 1}{\log_2(i + 1)} \quad (19)$$

Với $r(i)$ là điểm đánh giá của tin tức viên I trong danh sách đã xếp hạng và Z_K là một hằng số chuẩn hóa nhằm giúp cho danh sách đã được xếp hạng đạt nDCG bằng 1.

3. Các mô hình so sánh

DKN

Deep Knowledge-Aware Network (DKN) là một hệ thống tích hợp knowledge graph vào gợi ý tin tức. DKN là một hệ thống khuyến nghị dựa trên nội dung và dùng để dự đoán tỉ lệ nhấp chuột của người dùng. Thành phần chính của DKN là KCNN (Knowledge-aware Convolutional Neural Network) là một mạng neural tích chập được liên kết đa kênh và chuyên sâu về từ vựng. KCNN coi các từ và thực thể là nhiều kênh và giữ mối quan hệ rõ ràng để căn chỉnh chúng trong quá trình tích chập. DKN còn sử dụng mô-đun attention để tổng hợp lịch sử hoạt động người dùng đối với tin tức ứng cử viên.

NAML

Neural News Recommendation with Attentive Multi-View Learning (NAML) là một phương pháp được đề xuất cho hệ khuyến nghị tin tức sử dụng mô hình mạng neural có khả năng học và biểu diễn các thông tin của tin tức và người dùng từ các loại thông tin tin tức khác nhau. Nhân tố trung tâm của phương pháp là hai bộ mã hóa tin tức và mã hóa người dùng. Mô hình attentive multi-view để học các biểu diễn thống nhất của tin tức từ tiêu đề, nội dung và thể loại bằng cách xem xét chúng như các quan điểm khác nhau của tin tức trong bộ mã hóa tin tức. Tại bộ mã hóa người dùng, áp dụng cơ chế attention để chọn tin tức có thông tin hữu ích cho quá trình học biểu diễn người dùng.

LSTUR

Long- and Short-term User Representations (LSTUR) là một phương pháp khuyến nghị tin tức dựa trên mạng neural có thể học được cả cách biểu diễn người dùng ngắn và dài hạn. Sử dụng mạng attention để chọn các từ quan trọng trong quá trình học biểu diễn các tin tức từ tiêu đề và thể loại cho bộ mã hóa tin tức. Trong bộ mã hóa người dùng, đề xuất biểu diễn người dùng dài hạn thông qua nhúng IDS của họ, đồng thời biểu diễn các người dùng ngắn hạn từ các lịch sử người dùng thông qua mạng GRU.

TANR

Topic-Aware News Representation (TANR) là một phương pháp khuyến nghị tin tức dựa trên mạng neural để học biểu diễn tin tức liên quan đến chủ đề. Trong bộ mã hóa tin tức, học các khái niệm của tin tức thông qua tiêu đề bằng mạng CNN, sử dụng mạng attention để chọn các từ quan trọng. Ngoài ra, đề xuất học biểu diễn tin tức liên quan đến chủ đề bằng cách huấn luyện bộ mã hóa tin tức chung cho một nhiệm vụ phân loại chủ đề phụ trợ. Trong bộ mã hóa người dùng, mạng attention được dùng để chọn các tin tức quan trọng để học khái niệm người dùng.

4. Thử nghiệm

Kết quả thử nghiệm

Sau khi thực nghiệm trên bộ dữ liệu MINDtiny, chúng tôi thu được kết quả như bảng 2. Theo như kết quả, các mô hình do chúng tôi đề xuất có những sự khác biệt khá rõ ràng so với các mô hình được so sánh.

Bảng 2: So sánh kết quả đánh giá trên dữ liệu thử nghiệm.

Model	AUC	MRR	nDCG@5/10
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	0.2759/0.3360
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

Đầu tiên, mô hình NAML-LSTUR cho kết quả tốt nhất trong thử nghiệm của chúng tôi, chỉ xếp vị trí thứ 2 sau mô hình NAML (AUC = 0.62). Mô hình NAML-LSTUR-nn đứng ở vị trí thứ 3 so với các mô hình state-of-the-art được chúng tôi sử dụng để so sánh. NAML-nn cho kết quả xấu nhất. Về độ đo MRR và nDCG@K, NAML-LSTUR có phần thấp hơn mô hình DKN.

Thứ hai, về kiến trúc mô hình, NAML và NAML-LSTUR có điểm tương đồng là sử dụng tóm tắt của tin tức dẫn đến hiệu suất mô hình tốt hơn các mô hình khác, điều này có thể có sự tác động của tóm tắt tin tức đến mô hình đề xuất.

Thứ ba, về hàm dự đoán khả năng click của người dùng, trong ba mô hình đề xuất của chúng tôi, mô hình sử dụng mạng nơ-ron đã làm giảm hiệu quả đáng kể hiệu suất đề xuất. Các nghiên cứu trước cũng chứng minh được hàm sử dụng dot product cho kết quả tốt hơn hàm sử dụng nơ-ron.

Thứ tư, xét về mô hình NAML và NAML-LSTUR, ta có thể nhận thấy rằng có thể phương pháp trích xuất đặc trưng người dùng của mô hình NAML hiệu quả hơn so với phương pháp của mô hình LSTUR.

Tóm lại, từ kết quả cho thấy, ta có thể kết luận mô hình có hàm dự đoán sử dụng mạng nơ-ron cho kém hiệu quả hơn dot product, đồng thời nó cũng làm giảm hiệu suất của mô hình NAML. Nhìn chung mô hình tốt nhất của chúng tôi cao hơn phần lớn các mô hình state-of-the-art hiện tại mà chúng tôi sử dụng để so sánh (sau NAML).

Nghiên cứu diễn hình và trực quan hóa kết quả

Chúng tôi thực hiện gợi ý tin tức cho một trường hợp cụ thể của người dùng trong bộ dữ liệu kiểm thử. Trong hình 5, chúng tôi hiển thị lịch sử tin tức người dùng đã xem gần nhất với số mẫu là 9, từ cao xuống thấp thể hiện cho thời gian người ấy đã xem (cao nhất là người dùng mới xem gần đây). Trong hình 6, chúng tôi cũng hiển thị 5 đề xuất hàng đầu tin

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...

Hình 5: Lịch sử của người dùng.

tức từ mô hình NAML-LSTUR. Có thể thấy, Mô hình tập trung vào đề xuất các thể loại liên quan với movies và tv, thứ mà được người dùng quan tâm phần lớn trong lịch sử. Dòng có chữ màu xanh là tin tức thực sự mà người dùng đã nhấp.

NAML + LSTUR			
Category	Sub-Category	Title	Abstract
tv	humor	Chrissy Teigen Scares 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 ...

Hình 6: Tin tức được đề xuất cho người dùng.

VI. Conclusion

1. Kết quả đạt được

Trong đề tài này, chúng tôi đã đề xuất mô hình mới có tên NAML-LSTUR cho bài toán Đề xuất tin tức với sự kết hợp các đặt điểm của hai mô hình NAML và LSTUR từ nghiên cứu trước, kết quả cho thấy cao hơn phần lớn các mô hình hiện tại. Ngoài ra, chúng tôi cũng đề xuất trích xuất văn bản theo định hướng BERT-base nhằm tăng hiệu suất hơn so với các phương pháp truyền thống. Thêm vào đó, chúng tôi cũng đề xuất sử dụng mạng nơ-ron lấy ý tưởng từ mô hình DKN để xây dựng hàm dự đoán click của người dùng. Chúng tôi cũng đề xuất các mô hình biến thể khác của NAML-LSTUR là NAML-LSTUR-nn và NAML-nn. Chúng tôi cũng đã xây dựng bộ dữ liệu MINDtiny cho việc huấn luyện bài toán đề xuất tin tức với các hệ thống có tài nguyên máy tính hạn chế.

2. Hướng phát triển

Trong tương lai, đề tài có thể phát triển bằng cách hoàn thành huấn luyện mô hình với định hướng BERT-base và tiến hành so sánh mức độ ảnh hưởng. Tìm giải pháp cải thiện hàm dự đoán click của người dùng hiệu quả hơn. Thực nghiệm các nghiên cứu cắt bỏ để đo lường mức độ tác động của các phần của kiến trúc và các phần của nội dung tin tức.

Tài liệu

- [1] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, and X. Xie, “Neural news recommendation with attentive multi-view learning,” *arXiv preprint arXiv:1907.05576*, 2019.
- [2] M. An, F. Wu, C. Wu, K. Zhang, Z. Liu, and X. Xie, “Neural news recommendation with long-and short-term user representations,” in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 336–345, 2019.
- [3] H. Wang, F. Zhang, X. Xie, and M. Guo, “Dkn: Deep knowledge-aware network for news recommendation,” in *Proceedings of the 2018 world wide web conference*, pp. 1835–1844, 2018.
- [4] A. S. Das, M. Datar, A. Garg, and S. Rajaram, “Google news personalization: scalable online collaborative filtering,” in *Proceedings of the 16th international conference on World Wide Web*, pp. 271–280, 2007.
- [5] S. Okura, Y. Tagami, S. Ono, and A. Tajima, “Embedding-based news recommendation for millions of users,” in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1933–1942, 2017.
- [6] S. Rendle, “Factorization machines with libfm,” *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 3, no. 3, pp. 1–22, 2012.
- [7] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, “Deepfm: a factorization-machine based neural network for ctr prediction,” *arXiv preprint arXiv:1703.04247*, 2017.
- [8] R. Wang, B. Fu, G. Fu, and M. Wang, “Deep & cross network for ad click predictions,” in *Proceedings of the ADKDD’17*, pp. 1–7, 2017.
- [9] C. Li, Z. Liu, M. Wu, Y. Xu, H. Zhao, P. Huang, G. Kang, Q. Chen, W. Li, and D. L. Lee, “Multi-interest network with dynamic routing for recommendation at tmall,” in *Proceedings of the 28th ACM international conference on information and knowledge management*, pp. 2615–2623, 2019.
- [10] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, *et al.*, “Mind: A large-scale dataset for news recommendation,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 3597–3606, 2020.
- [11] C. Wu, F. Wu, M. An, J. Huang, Y. Huang, and X. Xie, “Npa: neural news recommendation with personalized attention,” in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2576–2584, 2019.
- [12] C. Wu, F. Wu, S. Ge, T. Qi, Y. Huang, and X. Xie, “Neural news recommendation with multi-head self-attention,” in *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pp. 6389–6394, 2019.
- [13] C. Wu, F. Wu, T. Qi, and Y. Huang, “Empowering news recommendation with pre-trained language models,” in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1652–1656, 2021.
- [14] F. Qi, Y. Yang, J. Yi, Z. Cheng, Z. Liu, and M. Sun, “Quoter: A benchmark of quote recommendation for writing,” *arXiv preprint arXiv:2202.13145*, 2022.