TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Deepika Uttam Sambrekar [223897IVSM]

CONTRASTIVE LEARNING BASED NEWS RECOMMENDATION

Master's Thesis

Supervisor: Dr. Roshni Chakraborty
Ph.D.

Co-supervisor: Dr. Gert Kanter
Ph.D.

Tallinn 2025

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Deepika Uttam Sambrekar [223897IVSM]

KONTRASTÕPPEL PÕHINEV UUDISTE SOOVITUSSÜSTEEM

Magistritöö

Juhendaja: Dr. Roshni Chakraborty
Ph.D.

Kaasjuhendaja: Dr. Gert Kanter
Ph.D.

Tallinn 2025

Author’s Declaration of Originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Deepika Uttam Sambrekar

02.01.2025

Abstract

Presently, news recommendations have become a widely utilized channel for users to access news articles that match their interests. Daily vast number of news articles are generated, and it becomes difficult to recommend news articles that align with users' preferences. However, to capture effective news recommendation that has rich textual context and accurately match them with users' interests, we introduce ConNewsRec, a personalized news recommendation using an inventive deep learning technique, showcasing two significant findings: analyzing users' News Topical Representations based on the topics in their recent reading history may yield superior results and secondly, employing a contrastive learning module to incorporate news article titles could potentially be more efficient than directly combining them. Through extensive experiments conducted on the MIND-Small dataset, we validate the effectiveness of our ConNewsRec model.

The thesis is written in english and is 50 pages long, including 6 chapters, 7 figures and 3 tables.

Annotatsioon
Kontrastõppel Põhinev Uudiste Soovitussüsteem

Praegu on uudiste soovitused muutunud laialdaselt kasutatavaks kanaliks, mille kaudu kasutajad pääsevad ligi nende huvidele vastavatele uudisteartiklitele. Iga päev luuakse tohutul hulgal uudisteartikleid ja kasutajate eelistustele vastavaid uudisteartikleid on raske soovitada. Kuid selleks, et jäädvustada tõhusaid uudistesoovitusi, millel on artiklite rikkalik tekstikontekst, ning sobitada need täpselt kasutajate huvide ja asjakohaste teemadega. Tutvustame ConNewsReci, isikupärastatud uudistesoovitust, mis kasutab leidlikku süvaõppe tehnikat, tutvustades kahte olulist järeldust: Esialgu võib kasutajate uudiste aktuaalsete esituste analüüsimine nende hiljutise lugemisajaloo teemade põhjal anda suurepäraseid tulemusi. Teiseks võib kontrastiivse õppemooduli kasutamine uudisteartiklite pealkirjade ja kokkuvõtete lisamiseks olla tõhusam kui nende otsene kombineerimine. MIND-Small andmestikuga läbi viidud ulatuslike katsete abil kinnitame oma ConNewsReci mudeli tõhusust.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 50 leheküljel, 6 peatükki, 7 joonist, 3 tabelit.

List of Abbreviations and Terms

NR	News Recommendation
CF	Collaborative Filtering
BERT	Bidirectional Encoder Representations from Transformers
MLM	Masked Language Model
GRU	Gated Recurrent Unit
CL	Contrastive Learning
ICF	Item-based Collaborative Filterin
PLM	Pre-trained Language Model
NLP	Natural Language Processing
CNN	Convolutional Neural Networks
ZSSD	Zero-shot Stance Detection
MIND	Microsoft News Dataset
AUC	Area Under the Curve
MMR	Mean Reciprocal Rank
nDCG	Normalized Discounted Cumulative Gain
s_{kc}	Interest Scores
n_c	Candidate Article
u_k	User k
u_1	User 1
u_2	User 2
V	Set of News Articles
l	Total number of Clicked Articles
i	i^{th} Element
T	Set of Titles
W_{v_k}	Sequence of Words from the title of article v_k
w_o	o^{th} word in the title
H_{v_k}	BERT's Embeddings for W_{v_k}
M_H	Multi Headed Attention Mechanism
L_R	LeakyReLU
N_E	News Encoder
U_E	User Encoder
o_w	Total Number of BERT's Embeddings
M	Attention Embeddings of News Encoder

E_{u_k}	User's Topical News Representation u_k
R	Customized News Representation
r_v	Final News Representation
Z	Attention Embeddings of User Encoder
O_i	Customized User Representation
u_r	User Representation
r^c	Candidate News Representation
$\hat{y}$	Click probability
$P(removal)$	Probability of Removal
$P(random)$	Random Probability
$MaxPop$	Maximum Popularity
Pop	Popularity
$AvgPop$	Average Popularity
T_{rem}	Remaining article's titles
T_s	Remaining titles after dropout and shuffle
n	Number of Negative Samples for $Loss_{CL}$
$\|S\|$	Size of the Training Set
n_s	Number of Negative Examples for $Loss_{REC}$
τ	Temperature
V_1^n, V_2^n, b_1^n	Parameters for News Representation
q_2, V_2^u, b_2^u	Parameters for User Representation
$V_1^l, V_2^l, b_1^l, b_2^l$	Parameters for Contrastive Learning
q_c, V^c, b^c	Parameters for click prediction.
$\tilde{\alpha}_{ij}^n$	Attention Weight in the News Encoder module
ϱ_i^n	Intermediate Attention Score for News Encoder
$\tilde{\varphi}_{ij}^u$	Attention Weight in User Encoder module
$\tilde{\kappa}_i^u$	Intermediate Attention Score for User Encoder

Table of Contents

List of Figures

List of Tables

1. Introduction

Disseminating information to the masses at a large scale has always remained an important task. The invention of *printing press* by *Johannes Gutenberg* revolutionized news dissemination through newspapers, which has remained a major component of news media. However, in the 21^{st} century, the era of digitization has transformed the news consumption behaviour. All leading newspapers across the globe have reported a sharp drop in their print circulation. With the advent of smartphones, proliferation of internet and massive increase in the usage of online social networks in the daily lives, news media is witnessing a major disruption where digital news contents are gaining much popularity and ordinary citizens are collectively becoming part of the information dissemination process.

A plethora of recent studies and surveys on news consumption behavior highlights that the majority of the population get the daily news feeds from news websites or online social networks, like *Twitter*, *Reddit* and *Facebook* rather than the printed offline newspapers [1], [2]. Nowadays, a majority of the population get the daily news feeds from news websites or online social networks. More specifically, these studies indicate that approximately half of the news consumers from Europe use *Facebook* and *Twitter* as a source of daily news[1]. Similarly, in USA, the fraction of users who rely on social media platforms rather than the newspapers has doubled in the last few years. These platforms provide readers unbridled access to continuous stream of news information throughout the day and their attention is not bounded by the contents of any specific printed newspaper at hand. Therefore, it becomes important for the news media sources to both focus on the contents as well as the interest of the readers to drive their popularity among the readers.

Studies based on theories like *Uses and Gratification theory*[2] [3], [4] have shown that readers choose news media sources in a way that fulfill their specific communication needs. Thus, readers switch across different news media platforms based on their interests and their gratification needs. Therefore, these impertinent changes in news consumption demands news media sources to understand the pulse of the readers and adapt themselves according to the trends [5]. Additionally, the high volume of news generated makes it difficult for readers to identify the specific news that interest them. Therefore, news media agencies provide personalized reading recommendations to users to retain audience and provide better news reading experience for the users [6]. Although a plethora of recommender systems exist for different applications, they are not directly applicable for

[1]https://www.statista.com/statistics/718019/social-media-news-source/
[2]https://www.communicationtheory.org/uses-and-gratification-theory/

news recommendation (NR) [7], [8]. The reason being the news recommendation varies significantly from other domains of recommendations, such as, it requires to consider the temporal relevance in recommendation unlike movie recommendation, occurrence of breaking news based events, dynamic change in user interests.

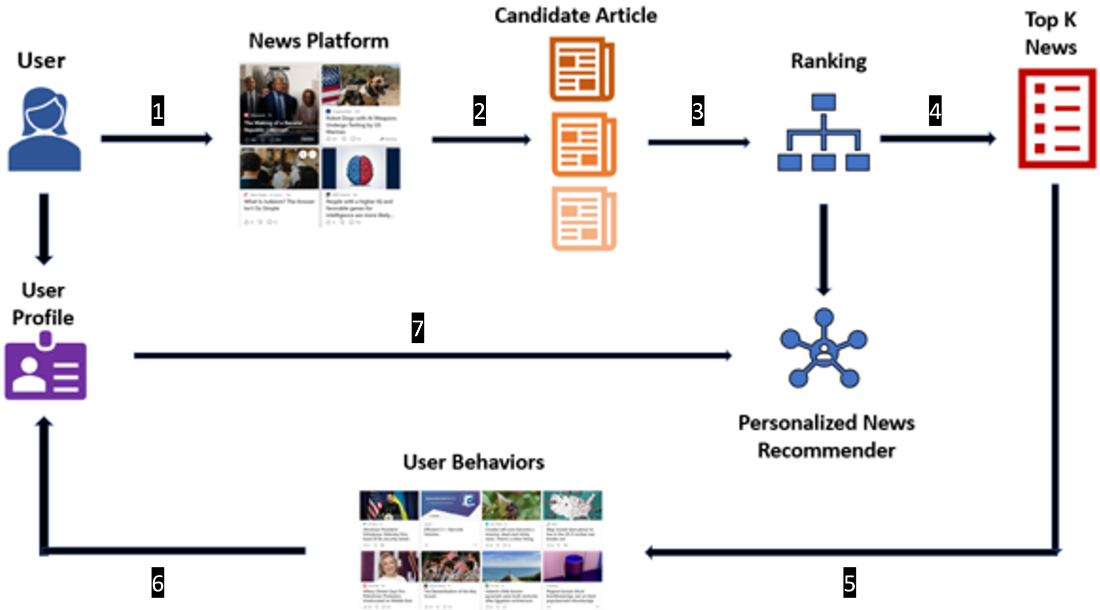

Figure 1. *General visualization of News Recommendation Systems.*

News recommendation could be visualized as shown in Figure 1[3], where an user is provided with a selected subset of news articles from all the news articles available on that day by a news recommender. A news recommender selects a specific subset of news articles for a user on the basis of the user's news consumption and her/his topical interests. Further, given the subset of news articles at a given timestamp, the user chooses to read one or few of these news articles which implicitly aids the news recommender to update the user's news consumption and her/his topical interests. Therefore, a news recommender intends to learn from the implicit user behavior to provide her/him the most suitable subset of news articles at a particular timestamp. The main challenges for news recommendation include understanding of the user interests on the basis of the reading behavior which requires identifying the implicit factors given a news article, such as, topic or entities and the explicit factors, such as the news event. Furthermore, identifying the relevant and specific news articles from the huge number of news articles available is extremely difficult due to high data sparsity. Additionally, news recommender systems need to provide a balanced representation of the news articles through diversity, recency, novelty and serendipity [9].

Existing research works on news recommendation differ on the basis of their identifying representative of user interests from users' news consumption and further, matching with

[3]images of "News Platform" and "User Behaviors" were obtained from https://www.msn.com/en-in

available news to identify the most relevant and specific news articles given a user [10], [11]. To achieve this, various techniques are employed, such as Collaborative Filtering (CF) [12], Content-Based Filtering (CBF) [13], Knowledge-Based Approaches [14], and Neural Network (NN) based [15] methods, each with distinct strengths and limitations. Collaborative filtering forecast user interests by studying the preferences of similar users, but faces challenges such as the cold start problem, scalability issues, and data sparsity [16], [17], [18]. Content-based filtering, makes news recommendations based on how well an article fits with the user's past reading. This method limits exposure to new or diverse topics by restricting it to contents that is similar to what the user has already interacted with [19], [20]. In a knowledge-based approach, the system often aligns understanding of both the content and the user's interests to match articles with user needs. This approach finds it difficult to adjust to new or specialized subjects that are not well-represented in the existing knowledge base [21], [22].

Currently, neural news recommendation techniques focus on developing various Neural Network(NN) architectures to encode user and news representations effectively [23], [24]. Early approaches, such as Neural News Recommendation with Attentive Multi-View Learning (NAML) [25], Neural News Recommendation with Multi-Head Self-Attention (NRMS) [26], advanced the field by introducing attention mechanisms to enhance representation learning. Convolutional Neural Networks (CNNs) [27], set the groundwork by extracting features from news content [28]. News representations using a denoising autoencoder utilize a GRU [29], [30] network to learn user representations from the news they peruse. More recent techniques, such as BERT-based models [31] improve news recommendations by using pre-trained transformers to capture contextual relationships in text, addressing the limitations of earlier methods. User-News Matching BERT for News Recommendation (UNBERT) [32], uses pre-trained language models to align news embeddings and users for personalized recommendation, including temporal and contextual behavior, which may limit its ability to capture dynamic and multi-faceted user interests. Prompt4NR [33], a novel framework for news recommendation, reorganizes the task as a [MASK] prediction problem using diverse prompt templates and multi-prompt ensembling, resulting in higher performance in experiments. Using Prompt4NR can increase the complexity due to its dependency on multi-prompt ensembling, highlighting the need for more adaptive and efficient solutions.

The Neural Networks and the other state-of-art techniques, mentioned above, however face the primary limitations of cold start problem and data sparsity. These remain a significant challenge, as the new users or news articles have little to no historical interaction data, which can provide incorrect suggestions. Data sparsity makes the problem worse by limiting the number of users' news articles interactions, making personalization less effective by likely introducing biases into the suggestions. These systems also struggle in capturing complex user preferences, managing sparse or partial interaction data, and maintaining scalability in the absence of substantial labeled datasets.

In contrast to conventional techniques, contrastive learning (CL) has shown strong performance in both supervised [34] and unsupervised [35] settings. Contrastive learning is a technique that trains models by increasing the similarity between related pairs of data (positive samples) while minimizing the similarity between unrelated pairs (negative samples) [36]. In Contrastive News Recommendations based on Curriculum Learning (CNRCL) [37], news articles that are similar to the candidate news are treated as positive examples, while negative examples are unclicked news articles selected based on their similarity to user preferences. CNRCL combines curriculum learning for user-specific negative sampling with contrastive learning. In the work Improving News Recommendation with Channel-Wise Dynamic Representations and Contrastive User Modeling [38], Wang et. al propose the MCCM model which enhances news recommendation by leveraging channel-wise dynamic convolution for multiperspective news features and frequency-aware contrastive learning to capture key user behaviors while reducing noise. In this, positive examples are users' clicked history with dropout (category-based replacement), while negative examples are clicked histories from other users. In Multi-Interest Extraction joint with Contrastive Learning (MIECL) [39], positive examples are considered to be the pair of user's interest-level representation and its corresponding interest prototype while, negative examples are consider as pair of the user's interest-level representation and unrelated interest prototypes. MIECL employs multiple interest prototypes and a graph-based user encoder to capture unique and context-rich user representations, enabling more detailed and diverse user modeling. KGCL [40] uses knowledge graphs and contrastive learning to manage data noise, improving user preference modeling and handling ambiguous data effectively. Contrastive learning enhances the quality of news recommendations, explicitly modeling the differences between news articles and user preferences for various news articles even when interaction data is sparse or nonexistent.

Contrastive learning can overcome the problems of cold start and data sparsity issues by learning more reliable and generalizable representations for both users and news articles. Contrastive learning effectively captures the nuanced trends in user behavior and content, reducing the need for labeled data for training. This makes the system more efficient, even in situations of sparse or dynamic data. Furthermore, contrastive learning reduces high computational complexity by optimizing resource utilization. However, existing contrastive learning-based techniques face several limitations. For instance, CNRCL depends on user-specific negative sampling and the added complexity of curriculum learning restricts its scalability. Although MCCM is effective at capturing multi-perspective news aspects, it has difficulty with extreme noise in datasets and has significant computational requirements. Similarly, MIECL faces challenges in managing multiple prototypes and graph-based encoders, along with difficulties in distinguishing overlapping user interests. Since KGCL relies on knowledge graphs, its relevance is restricted in fields with noisy or incomplete graph data. In order to improve user and content representation learning, contrastive learning-based approaches, which emphasize differentiating between similar and dissimilar interactions have become a viable method in news recommendation.

In this thesis, we propose a contrastive learning based framework, *ConNewsRec*, which is a novel contrastive learning based framework for news recommendation. *ConNewsRec* comprises of three phases, i.e., *News Encoder*, *User Encoder* and *Contrastive Learning Module*. *News Encoder* intelligently integrates both the explicit features of a news article from the news text-based representation and the implicit features that is news topic based representation. *News Encoder* is a multi-head attention-based approach coupled with BERT to generate the news embedding and *User Encoder* generates the representation of every user on the basis of their news reading behavior. Therefore, *News Encoder* and *User Encoder* can effectively capture both the explicit user news topical choices along with the implicit news article semantics through the topic based understanding. Additionally, we propose a novel *Contrastive Learning Module* that can handle the high data sparsity, variance in user profiles along with diversity and novelty.

Validation of *ConNewsRec* has been done on the MIND-Small dataset which comprises of $94,000$ news articles and $50,000$ users. The organization of the thesis is as follows: We discuss the existing research works in Section 2.1 and *ConNewsRec* in Section 3. We provide details of the experimental settings and results in Section 4. We finally summarize in Section 6.

2. Related Work

The exponential growth of online news consumption coupled with the demand for personalized content delivery has brought considerable attention to the field of news recommendation systems in recent years. This section provides an overview of the research and methodologies in the new recommendation systems, ranging from traditional to more advanced approaches.

2.1 Traditional News Recommender Approaches

Traditional recommendation systems are commonly classified into three primary categories, such as, Content-based filtering, Collaborative filtering and Knowledege-based techniques.

Content-based Filtering: User profiles and item descriptions provide information about the preferences and past selections of the users that inform content-based filtering recommendations [20]. A content-based recommender constructs a profile of the user's preferences based on the features present in items the user has previously rated. Contextual data significantly enhances recommendations across various fields. To further enhance the effectiveness of these recommendations, it is crucial to address issues such as redundant context, information overload, and data redundancy [41]. Content-based recommendation systems typically include mechanisms for describing items available for recommendation. They also involve creating a user profile outlining their preferences, and comparing items to the user profile to ascertain suitable recommendations [19]. Content-based algorithms prioritize recommending items that show similarity to those that the user had previously favored. This technique relies on the analysis of specific features or metadata associated with items to comprehend their content [16].

Collaborative Filtering(CF): This technique is one of the most well-known, extensively applied, and well-established approaches in the field of recommender systems. This CF technique is on the notion that users who favor similar articles are likely to be drawn to similar content; this suggests articles based on an analysis of past preferences and behaviors of users with similar tastes [17]. This approach analyzes both individual user data and collective community data to understand user preferences and make recommendations [18], [42]. Consider for example, if *UserA* that is u_1 and *UserB* that is u_2 have previously read similar articles, and u_1 recently read an article that u_2 hasn't seen yet, the recommendation would be to suggest this article to u_2 as well. News articles are usually represented by their unique IDs in many collaborative filtering techniques. However, on many news websites, new articles are constantly being published, and older ones quickly disappear. As a result, using article IDs for representation frequently results in serious problems with cold starts and low performance [43]. Using Item-based collaborative filtering (ICF) in user preference modeling preference modeling facilitates the implementation of online personalization [44]. An algorithm for collaborative filtering based on user input has been created specifically for the MapReduce program framework and deployed on the Hadoop platform [45]. Predictions are derived from the construction of an underlying model of user preferences [46].

Knowledge-based techniques: This approach, weighs user preferences and article content to make recommendations utilizing explicit knowledge and predetermined rules. Knowledge-aware recommender systems use information gathered from domain-specific descriptions or knowledge graphs to uncover patterns and extra information not found in an item's features [22]. Enhancing conventional methods of information retrieval and recommendation by incorporating outside data from knowledge bases has been suggested as a potential remedy for some of the limitations encountered in recommender systems within the news domain. Knowledge graphs [21], which are directed, labeled heterogeneous graphs, depict real-world topics and their relationships [47]. Deep Knowledge-Aware Network for News Recommendation (DKN) [48], combining the textual content of news and external knowledge such as entities in news articles to enhance the recommendation. Conventional news recommendation methods tend to overlook the large number of knowledge entities and common sense information found in news articles [14]. These methods provide recommendations that consider the nuances of the content as well as user preferences by utilizing pre-established rules and domain expertise.

2.2 Neural Network based News Recommendation

Traditional identification-based approaches for news recommendation frequently encounter challenges due to the quick update of news articles, a phenomenon commonly known as the cold-start problem [23]. Many studies have demonstrated the strong text modeling capabilities of pre-trained language models (PLMs), which have allowed them to make major advancements in the field of NLP [24]. In contrast to conventional models, which are typically trained directly on labeled data for particular tasks. Deep learning-based news recommendation systems use neural networks to simulate complicated interactions between users and news articles, hence improving suggestion accuracy and personalization. They can encode text information that is applicable to all tasks through this process before fine-tuning for specific tasks. News recommendation shares a strong connection with natural language processing (NLP). News articles are a global form of textual data, and serve as the backbone of this domain. Leveraging advanced text modeling techniques like Convolutional Neural Networks (CNN) and Transformer architectures offers seamless means to represent and comprehend the nuanced content encapsulated within news articles [49]. News representations from news titles through a knowledge-aware CNN network have been developed to integrate information from knowledge graphs [14]. Utilizing a CNN to encode news content complemented by a Gated Recurrent Unit (GRU) network to capture user preferences based on their historical interactions. By using multi-view attention across different aspects like title, topic category, and entities, the User-as-Graph approach represents individual news items. Furthermore, it enables the modeling of relationships between user behaviors by representing each user as a customized heterogeneous graph [50].

Recently, the focus has shifted to using sentence or document encoders that can generate contextual token representations. Using unlabeled text data, these encoders are first pre-trained and then refined for particular supervised tasks [51]. Contemporary NLP systems with pre-trained word embeddings. Compared to embedding that are trained from scratch, these embedding constitute a significant advancement. Pre-trained word embedding are typically created to achieve a variety of goals, such as language modeling tasks that require the user to distinguish between words that are correct and incorrect in left and righ hand contexts [52]. Integrating knowledge graph data from news topics and leverage pre-trained BERT models have advanced text comprehension capabilities, However, the majority of these methods only learn one user embedding, potentially falling short in accurately representing the diverse range of user interests [49], [53], [54]. Some methods also incorporate knowledge graph information from news topics. Yet, most methods learn single user embedding which may not adequately model the diverse user interests [32].

2.2.1 Contrastive Learning

Contrastive learning is a self-supervised technique for learning effective representations by contrasting positive and negative pairs of data samples, aiming to bring similar samples closer in the embedding space while pushing dissimilar samples further apart [55], [56], [57], [58], [36], [59], [60]. In the context of news recommendation, the objective of contrastive learning is to learn significant representations of the data, creating an embedding space without relying on explicit labels. This involves bringing the representations of similar patterns, such as news articles that users have engaged with or shown interest in (positive examples), closer together in the embedding space. At the same time, it distances the representations of dissimilar patterns, articles that users have not interacted with or expressed interest in (negative examples). This approach enables the system to better understand user preferences and increase the quality of recommendations.

PerCoNet [61] presents a deep persona-aware network using cross-view contrastive learning, which improves personalized news recommendation systems by incorporating explicit user personas into both the news encoder and the user encoder for more detailed and personalized recommendations. The CNRCL model [37] uses curriculum learning for news recommendations by controlling the negative sampling process based on user interests, making it more aligned with individual preferences compared to standard methods. MIECL [39] uses multiple interest prototypes and a user encoder to learn unique user representations for each prototype. Additionally, a graph-enhanced user encoder is employed to improve user representations by including contextual information under each interest background, thereby improving the granularity and diversity of user modeling. The KGCL [40] framework handles data noise in recommendation systems through the use of knowledge graphs and contrastive learning. It enhances how user preferences are learned by incorporating additional self-supervised signals and effectively handling unclear or ambiguous data. The contrastive learning framework has also been used to obtain robust sentence representations, such as in the context of the Stance Detection task [62]. Wu et al. [63] utilized contrastive learning to develop noise-resistant sentence representations by applying various sentence-level augmentation techniques, including span deletion, substitution, and reordering. Some recommendation systems have started incorporating contrastive learning into their machine-learning workflows. DHCN [64] uses contrastive learning along with hypergraphs to model higher-order relationships for better node representations. RAP [65] uses the contrastive learning mechanism to enhance the accuracy of sequence denoising processes. A hierarchical contrastive learning strategy was proposed by Liang et al. to improve the performance of Zero-shot Stance Detection (ZSSD) [66]. This approach seeks to capture associations between various stance labels as well as between target-specific and target-invariant features.
This demonstrates contrastive learning's adaptability across domains, effectively capturing complex relationships and enabling robust, noise-resistant representations in news recommendation systems. Current contrastive learning techniques provide advantages like, creating strong, noise-resistant representations, effectively capturing dynamic user preferences, and reducing reliance on labeled data. Methods like CNRCL and MIECL help improve personalization and understanding of user behavior. However, they still have limitations. These approaches rely excessively on good augmentations, struggle to handle new users or articles (cold-start problem), and require a significant amount of computational power, which makes them difficult to utilize in real-time systems. This shows the need for better contrastive learning techniques that are more efficient, adaptable, and capable of capturing both context and user behavior effectively.

In the context of news recommendation, the presence of rich textual elements such as abstracts and titles offers a great chance for contrastive learning. These features can be treated as two separate perspectives of each news article, making them naturally suitable for contrast learning approaches. To leverage this, our ConNewsRec model introduces a novel contrastive learning module, with the goal of enhancing news and user encoder modeling abilities. This improvement leads to an overall improvement in news recommendation tasks' performance. This enhancement contributes to an overall performance boost in news recommendation.

3. Methodology

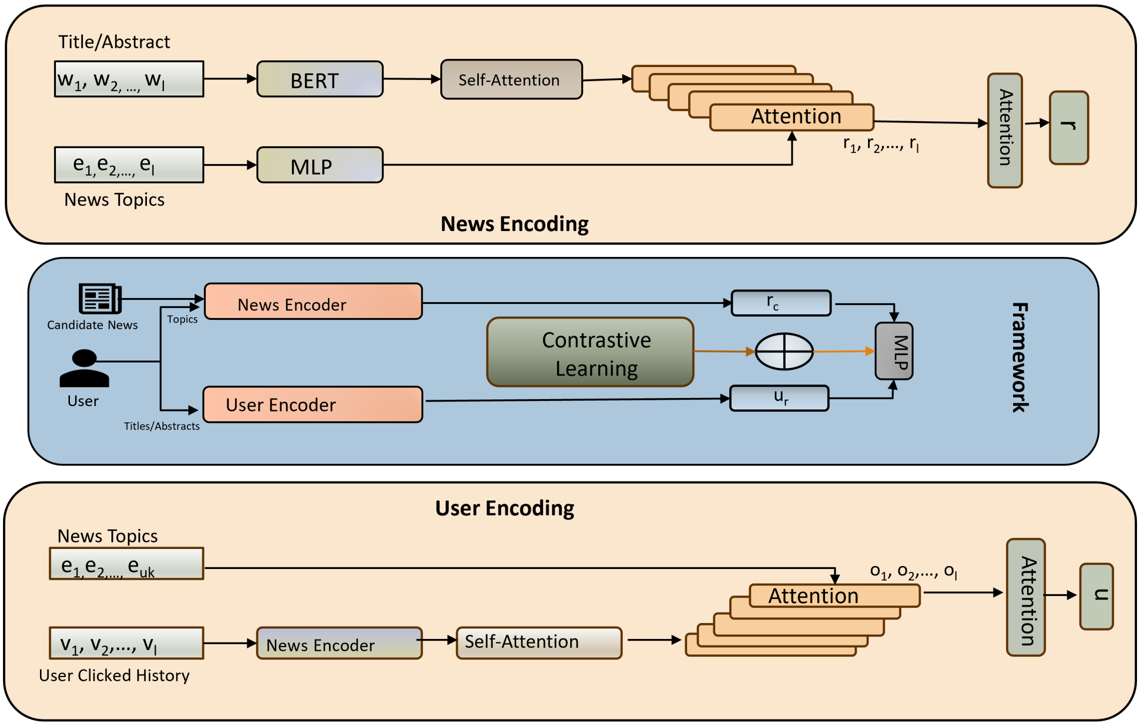

Figure 2. *The ConNewsRec model architecture*

In this Section, we discuss the problem statement formally followed by the proposed model, *ConNewsRec* which is a contrastive learning based news recommendation.

3.1 Problem Statement

In *ConNewsRec*, given that a user u_k's news reading behavior comprises of past-clicked news articles, $V = \{v_1, v_2, ..., v_l\}$ where, l is the total number of clicked news articles of the user u_k, the objective is to predict an interest score s_{kc}, for a candidate news article n_c, that the user u_k has not yet read. The interest score s_{kc} ranging from 0 to 1 quantifies the level of interest the user u_k has in the candidate news article. Each news article, v_i in the set V, is associated with a title t_i, and a set of topics e_v, $(v = 1, ...v_t)$, which together form a representation for the particular news article v_i.

In Figure 2, we show an overview of the proposed model architecture *ConNewsRec* comprises of three modules: a) News Encoder, b) User Encoder, and c) Contrastive Learning Module. In News Encoder, we generate a representative embedding of a news article v_i, by encoding its textual content, such as the title and abstract, along with the topical relevance captured by a multi-headed attention mechanism. The user encoder generates a representative embedding for a user u_k by integrating his topical interests using the 20 most recent news articles from the user's history and the set of representations, e_v, of the user u_k's clicked articles. Lastly, we elaborate on contrastive learning to amplify comprehension of the connections between users and news articles. These modules collectively predict the likelihood of recommending a candidate news article, n_c, to the user u_k, based on his/her interest score s_{kc}. Each of these components is elaborated in detail in the following sections.

3.2 News Encoder

News encoder intends to integrate the relevant information from a news article to generate comprehensive news article embedding. In order to capture the relevant and representative information from each news article, we explore both the explicit information through news article title and abstract and implicit information through the topical information of the news article. While the explicit information can effectively aid in representation of the news article content, the implicit information aids in identifying user interests through news topics irrespective of the news article title and abstract. The news encoder is performed to learn the news representation between news titles and user topics, by capturing the essential features and meaning of content present in the news articles and the interests represented by the user's topics. In this thesis, we employ a multi-headed attention mechanism coupled with BERT embedding to generate news article embedding. BERT employs bidirectional encoding that establishes a better semantic understanding of news articles and its ability to capture context and nuances of the content [31].

3.2.1 News Representation

BERT receives the titles of the news articles V, as input. Each of the titles can be represented as a sequence of words, i.e. $W_{v_k} = \{w_1, w_2, ..., w_o\}$, where w_i, is the i^{th} word in the title-sequence of v_k and o signifies the total number of words in the title of the particular article v_k in V. The subscript v_k in W_{v_k} is used to denote that the sequence of words corresponds to the news article v_k. We then process the title of the news article v_k by adding special [CLS] tokens at the beginning of every title and [SEP] tokens at the end of the title. This processed input is then passed through the BERT model. The two dense embedding layers will be employed to transform the outputs from BERT into a sequence of embedding, followed by the application of a LeakyReLU [67] (L_R) activation function denoted as $H_{v_k} = \{h_1, h_2, ..., h_{o_w}\}$, where h_i represents the embedding corresponding to the i^{th} word in the input sequence. It is important to note that the size of H_{v_k} (o_w) is taken to be less than the size of the sequence W_{v_k} when used in Equation (3.1).

$$h_i = V_1^n * L_R(V_2^n * BERT(w_i) + b_2^n) + b_1^n, \tag{3.1}$$

where, V_1^n, V_2^n, b_1^n and b_2^n are the learnable weights and biases are used in the model. V_1^n and V_2^n collectively govern the transformations applied to the output of BERT embeddings through two dense layers. b_1^n and b_2^n adjust the output of the first dense layer by adding a bias to each neuron's activation. To further capture the interactive semantics, each matrix H_{v_k} is transformed into another representation M through the application of a multi-head-self attention mechanism (M_H). The attention mechanism captures different aspects of the input data, such as news articles and user preferences. The multi-head attention mechanism produces attention weights, which make it easier to calculate a similarity score between each news article and the user's preferences. Through the integration of these attention weights across different heads, the system takes into account multiple aspects of both the articles and the user's preferences, leading to recommendations that are more precise and personalized to individual preferences. This allows multiple attention mechanisms to weigh the significance of different elements within H_{v_k}, encapsulating the complex interconnections and associations among terms to generate the sequence of processed titles, M as shown in Equation. (3.2).

$$M = M_H(H_{v_k}) \tag{3.2}$$

where, $M = [m_1, m_2, ..., m_{o_w}]$ is the sequence of the transformed title vectors for the article v_k processed from the attention mechanism.

3.2.2 News Topical Representation

In order to capture the implicit news information through the news topics of users as shown in Figure 3, we identify the distinguishing topics of each news article. There are several mechanisms to identify news topics, such as through topic modelling approaches like latent dirichlet allocation [68], keyphrases detection approaches RAKE [69], KEYBERT [70], etc. In this thesis, we focus on identifying the topics of user-clicked articles, as it can provide an understanding of the user's collective interests across different news articles.

The mapping of user preferences through topics is unambiguous in comparison to topic modeling. For example, a pair of users, say u_1 and u_2, are interested in news topics, such as, u_1 enjoys reading articles about *sports* and *music*, particularly articles about *Selena Gomez* and the game of *cricket*. On the other hand, u_2 is more interested in *business* and *music*, specifically musical instruments like the *guitar* and topics related to starting a business, like *startup strategies*. Although u_1 and u_2 enjoy reading articles on *music*, u_1 focuses more on reading *persona of singers*, while u_2 is more interested in *musical*

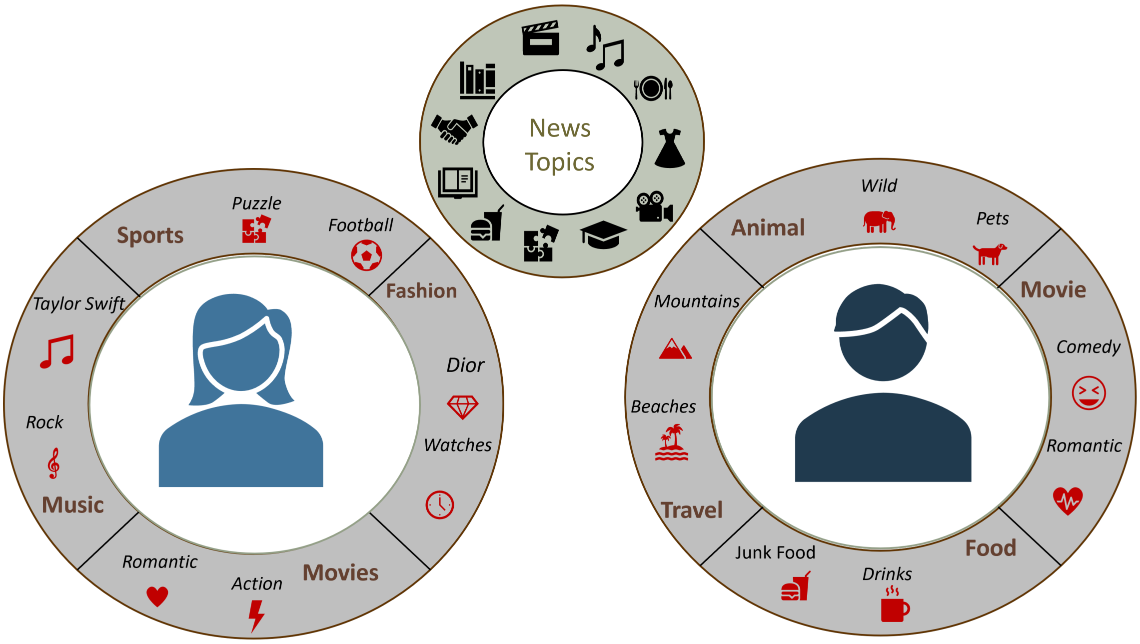

Figure 3. *An illustration of different Users' News Topical Representation.*

instruments. This distinction highlights how users within the same category can have varied interests in specific topics within that category. We propose the implementation of a News Topical Representation approach below, to address this diversity of interests and provide topic-based news article suggestions. This approach aims to recognize the context and content of news articles enabling us to recommend articles aligned with the user's interests.

The user's news topical representation E_{u_k} is constructed as a collection of topics to explicitly characterize the user's u_k news reading interests. User's news topical representation comprises numerous topics as $E_{u_k} = \{e_1, e_2, ..., e_l\}$, where each e_i represents the i^{th} topic and l signifies the maximum number of topics within the user's news topical representation. The user persona reflects the collective interests of the user, aggregated from the topics extracted from the news articles they have recently clicked. Topics within user's topical representation E_{u_k} can provide valuable additional context to aggregate the previously processed content, M. Repeated attention is applied to each combination of topics and the components of M, thereby improving the representation from M into R.

$$\varepsilon_i = L_R(V_3^n \times e_i + b_3^n), \tag{3.3}$$

$$\tilde{\alpha}_{ij}^n = \frac{exp(\varepsilon_i^T Q^n m_j)}{\sum_{k=1}^{o_w} exp(\varepsilon_i^T Q^n m_k)}, \tag{3.4}$$

$$r_i = \sum_{j=1}^{o_w} \tilde{\alpha}_{ij}^n m_j, \tag{3.5}$$

In this context, V_n^3, b_n^3, Q_n are learnable model weights and biases, adjusted during the training process. Here, V_n^3 a weight matrix that applies a linear transformation to the topics vector e_v. The bias b_n^3 shifts the result of the linear transformation to help the network better model the non-linear application of LeakyReLU (L_R) [67]. $R = \{r_1, r_2, ..., r_{e_l}\}$ is a customized semantically enhanced representation of the news titles t_i and the topics e_{u_k}. News articles are added together and weighted according to their individual attention weights to get each r_i. The attention weight between the i^{th} topic and the j^{th} news article's title is denoted as $\tilde{\alpha}_{ij}^n m_j$, indicating the relative importance of each news article to the topic. The softmax function in Equation (3.4) is used to calculate these weights by multiplying the topic and news representation vectors by their dot product Equation (3.5)

To produce the final representation r_v for the news article, we employ a hyperbolic tangent function followed by a softmax normalization, as is shown from Equations (3.6)-(3.8).

$$\rho_i^n = q_n^T * tanh(V_4^n \times r_i + b_4^n), \tag{3.6}$$

$$\varrho_i^n = \frac{exp(\rho_i^n)}{\sum_{j=1}^{e_l} exp(\rho_i^n)}, \tag{3.7}$$

$$r_v = \sum_{i=1}^{e_l} \varrho_i^n r_i, \tag{3.8}$$

Here q_n^T, V_4^n, and b_4^n are again the learnable weights and biases for news representation in Equation (3.6). For each r_v, the intermediate attention score, denoted as ϱ_i^n in Equation (3.7), is computed. The result is then passed through the hyperbolic tangent ($tanh$), which is an activation function that compresses the values into the range $[-1, 1]$. The vector is multiplied by this output using the dot product. The r_v is a weighted sum of every individual article's title in Equation (3.8).

3.3 User Encoder

User's long-term stable reading behavior can be captured from their history. Correspondingly, the topical representations of the news articles clicked by the users are also considered. This is done because a user's short-term interests are usually reflected in the recently clicked news articles, which capture temporary trends or momentary deviations from their long-standing preferences. For enhanced personalization, the user encoder can dynamically characterize a user's interests by integrating their long-term stable preferences with their recently clicked news articles. Based on previous works [61], we have considered 20 news articles from each user's news consumption history based on their chronological reading order.

For each user u_k, from the set of read news articles $V = \{v_1, v_2, ..., v_{n_u}\}$ from the user's history, the titles of the news articles are tokenized with $BertTokenizer$ using specialized tokens such as [SEP] and [CLS]. n_u is the number of articles taken for each user. The now tokenized titles (T_{u_k}), for the user u_k along with the user's topical representations E_{u_k} are supplied to the news encoder (N_E). The news encoded pair (T_{u_k}, E_{u_k}) is now passed through a multi-headed attention mechanism (M_H) to produce a rich semantic representation z_{u_k} for each user u_k.

$$z_{u_k} = M_H(N_E(T_{u_k}, E_{u_k})), \tag{3.9}$$

We form $Z = [z_1, z_2, ..., z_{u_k}]$, an array of vectors (each corresponding to a user), which has contextualized embeddings of the news titles and the users' topical representations after passing through the multi-head attention mechanism in Equation (3.9). A customized representation is then generated using attention weights (φ^u_{ij}) which are calculated from the enriched embeddings z_j and the topical representations e_i. The attention weights in Equation (3.10) are then normalized using the softmax function as shown in Equation (3.11).

$$\varphi^u_{ij} = q_1^T \times L_R(V_1^u \times (e_i \bigoplus z_j) + b_1^u), \tag{3.10}$$

$$\tilde{\varphi}^u_{ij} = \frac{exp(\varphi^u_{ij})}{\sum_{k=1}^{n_u} exp(\varphi^u_{ik})} \tag{3.11}$$

$$o_i = \sum_{j=1}^{n_u} \tilde{\varphi}^u_{ij} zj, \tag{3.12}$$

The parameters q_1, V_1^u, and b_1^u are variables within the model that can be adjusted through learning in Equation (3.10).

The O encompasses elements, $O = [o_1, o_2, ..., o_{u_k}]$, for each user u_k, which collectively represent various facets of the model's functionality and are modified during the learning process.

Using Equations (3.13)-(3.15) the o_i's are further condensed to produce the user representation u_r.

$$\kappa_i^u = q_2^T \times tanh(V_2^u \times o_i + b_2^u), \tag{3.13}$$

$$\tilde{\kappa}_i^u = \frac{exp(\kappa_i^u)}{\sum_{j=1}^{u_k} exp(\kappa_i^u)} \tag{3.14}$$

$$u_r = \sum_{i=1}^{u_k} \tilde{\kappa}_i^u o_i \tag{3.15}$$

Here q_2, V_2^u, b_2^u are learnable weights and biases for the representation o_i in Equation (3.13). The κ_i^u in Equation (3.13), is an attention weight calculated by multiplying the transposed parameter vector q_2^T, and the hyperbolic tangent of a linear transformation of the representation o_i. The final user representation is obtained by summing all the weighted topics, where each entity o_i is multiplied by its corresponding weight $\tilde{\kappa}_i^u$, obtained by normalizing κ_i^u with a softmax function. As a result, relevant topics will influence the final user representation u_r in Equation (3.15).

3.4 Click Prediction

The Click Predictor module predicts the probability that a user clicks on a candidate news article n_c. The calculation of the click probability stems from the News Encoder and the User Encoder modules described above. In particular, it starts by gathering the titles T_c's of a candidate news articles n_c's, along with the user's news topical representation E_{u_k}, and inputing them into the news encoder to obtain the news representation r_{v_c}. The user representation u_r, is obtained once the user encoder receives the titles of the user's read articles $\{v_1, v_2, ..., v_{u_k}\}$ from his/her history along with the topical representations. Once these two representations (the news representation, r_{v_c}, and the user representation u_r) are available to the system, the vectors are then concatenated into a single vector. This is then fed into a multi-layer perceptron (MLP) layer to compute the click probability $\hat{y}$. The procedure entails evaluating the combined impact of these variables to determine the probability that a user will click on a specific news article, as in Equation (3.16).

$$\hat{y} = sigmoid(q_c^T \times L_R(V^c \times (u_r \bigoplus r_{v_c}) + b^c)), \tag{3.16}$$

Here q_c, V^c, b^c are learnable weights and biases to be used for click prediction. q_c is a weight vector applied to the output of the LeakyReLU (L_R) function [67]. V^c is a weight matrix applied to the concatenated vector $(u_r \bigoplus r_{v_c})$ and b^c is a bias vector added to the linear transformation.

3.5 Contrastive Learning Module

In this section, we propose a novel contrastive learning module to identify user's news preferences and news consumption behavior. The core idea behind the proposed module is to ensure robustness irrespective of the high data sparsity in news recommendation datasets. For example, there is a high variance in news consumption behavior across users and additionally, there is high data sparsity as most of the users read very few news articles (as shown in Subsection 4.1). Furthermore, news recommender approaches often face issues with cold start problem. To mitigate the above mentioned challenges, our proposed module modifies the news articles selection probability such that articles of different popularity are included in the learning process. This minimizes the effect of sparsity and increases the recommendation system's robustness. Traditionally, contrastive learning minimizes the relationship between dissimilar users (negative pairs) while maximizing the similarity between comparable users (positive pairs). In this method, the similarity is weighted by the popularity of the news articles. We modify the selection probability of each news article according to its popularity in order to prevent the model from overfitting to highly popular articles, which could prevent the learned representations.

For instance, consider a scenario where a user, fascinated with sports and technology encounters two news articles: one about an Oscar-winning movie and another about a trending sports analytics app. Although the movie article might be appealing to a broader spectrum of audience, it does not align with the user's preferences; in contrast, the sports analytics app article is a perfect match. In a contrastive learning setup, the sports-app article would be treated as a positive example for this user and the Oscar-winning movie article as a negative example. By pulling a user's preference profile closer to relevant contents and pushing irrelevant contents away, the framework learns to recommend articles that truly match a user's personal interests. This example illustrates how a contrastive learning module distinguishes between articles that are popular from a general perspective and those that truly reflect a user's personal preferences. In this thesis, we adjust the contrastive learning module further by formulating a selection probability condition among the news articles based on their popularity. By implementing the probability condition, aptly titled as the *probability of removal* or simply $P(removal)$, our contrastive learning module creates a dynamic where news articles with $P(removal)$ less than a certain threshold value are considered to be more representative or informative for the framework.

$$P(removal) = P(random) \times \left(\frac{MaxPop - Pop}{MaxPop - AvgPop} \right) \tag{3.17}$$

Shown in Equation (3.17) is the “probability of removal", $P(removal)$ formalism, where each news article is first given an initial random probability ($P(random)$), which indicates its fundamental possibility of being chosen without taking into account its popularity. To factor in user engagement, we identify the highest popularity score ($MaxPop$) achieved by any news article v_i on a given topic. This score signifies the article with the greatest user interest within that topic. Furthermore, the mean popularity score ($AvgPop$) for every article on the subject is used to compute the average user popularity. The difference between an article's popularity (Pop) and its greatest popularity ($MaxPop$) is first calculated in order to fine-tune the initial selection probability. Then, by dividing this difference by the range between $MaxPop$ and $AvgPop$, the average popularity is normalized. We adjust the initial random probability ($P(random)$) of *0.6* by multiplying it by this normalized value (for $P(random)$, we also take subsequent values of $0.2, 0.4$ and 0.8). Higher than average popularity articles will have lower normalized values, which will lead to a lower adjusted likelihood. In contrast, articles with popularity closer to the average will have larger normalized values, resulting in an increased adjusted probability. This approach ensures that articles with popularity closer to average will be removed less frequently than those with higher popularity, which would be removed more frequently.

Following this, for the selection process, a removal condition is implemented. Specifically, if the probability of removal is greater than a predetermined threshold of 0.5 i.e. $P(removal) > 0.5$, the article is removed from the set of titles T of a particular user's history. By applying this condition, the model focuses more on articles with popularity scores closer to the average, thereby promoting a more balanced and specific understanding of user interests. The remaining titles, (T_{rem}), effectively enhances the model's capability to encode both news articles and user preferences.

In particular, considering a user u_k, let's consider a user's recent reading activity, is the *titles* of the news articles they have engaged are represented as $T = \{t_1, t_2, \ldots, t_{u_k}\}$. Initially, we start by randomly dropping out and then shuffling the remaining titles to obtain a subset T_s from T. T_{rem} for each user u_k, which is the set of titles remaining after applying the probability of removal condition, is also sent to the user encoder (U_E). Subsequently, we generate u^R and u^s executing the following operation:

$$u^R = U_E(T_{rem}, E_u), \tag{3.18}$$

and

$$u^s = U_E(T_s, E_u) \tag{3.19}$$

After passing through the user encoder, in Equation 3.18 and Equation 3.19, an MLP layer followed by the LeakyRelu (L_R) activation function will transform u^R and u^s into u_{c_l} and $u_{c_l}^+$ respectively.

$$u_{c_l} = L_R(V_1^l \times L_R(V_2^l \times u^R + b_2^l) + b_1^l) \tag{3.20}$$

and

$$u_{c_l}^+ = L_R(V_1^l \times L_R(V_2^l \times u^s + b_2^l) + b_1^l) \tag{3.21}$$

This transformation is associated with the learnable model parameters $V_1^l, V_2^l, b_1^l, b_2^l$. For the same user u_j, $u_{c_l}^+$ in Equation 3.21 can be interpreted as the positive example with respect to u_{c_l} in Equation 3.20.

For every user u_j, a temperature hyperparameter τ and n other users $u_i, (i \neq j)$, are present, from whom negative examples u_i^- can be sampled for contrastive learning, which can be formulated as follows:

$$Loss_{CL} = -log\left(\frac{exp(\frac{u_{c_l}^T u_{c_l}^+}{\tau})}{exp(\frac{u_{c_l}^T u_{c_l}^+}{\tau}) + \sum_{i=1}^{n_b} exp(\frac{u_{c_l}^T u_i^-}{\tau})}\right) \tag{3.22}$$

3.6 Model Training

For the primary task of news recommendation, the model is trained using a negative sampling strategy. For a particular user, all news articles that were clicked in, are treated as positive examples, while the rest of the news articles as negative. So, for a user u_k, this is used to distinguish between the news articles that were interacted with and those that did not. We create mini-batches S where for every positive example v_i we randomly select n_S negative examples. This is to ensure that there is a balanced representation of both positive and negative instances during training. Consequently, the main *recommendation-loss* ($Loss_{REC}$), can be written as in Equation (3.23).

$$Loss_{REC} = -\sum_{i=1}^{|S|} log\frac{\hat{yi}}{\hat{yi} + \sum_{j=1}^{n_S}\hat{yj}}, \tag{3.23}$$

In this n_S, denotes the number of negative examples for each positive example and $|S|$ is the size of the training set S.
In our approach, the contrastive learning task can be trained in parallel with the main news

recommendation task. This enables the joint learning of both the tasks. Thus, the combined loss function of our model encapsulates the effects of both the contrastive learning and the news recommendation tasks and is given by Equation (3.24):

$$Loss = Loss_{REC} + \lambda \times Loss_{CL} \tag{3.24}$$

Here, λ is treated as a hyper-parameter which denotes the weight given to $Loss_{CL}$ with respect to $Loss_{REC}$, where $Loss_{REC}$ pertains to the loss due to the primary objective of news recommendation, while $Loss_{CL}$ is the loss associated with the secondary task of contrastive learning.

4. Experiments

In this Section, we present the dataset, evaluation metrics, baselines hyperparameters details, ablation study and baselines, next.

4.1 Dataset

Table 1. *Dataset Summary*

Dataset	Training Dataset	Testing Dataset
#USERS	#50000	#50000
#News Articles	#51282	#42416
#News Articles per user	#20	#20
#topics Per User	#5	#5

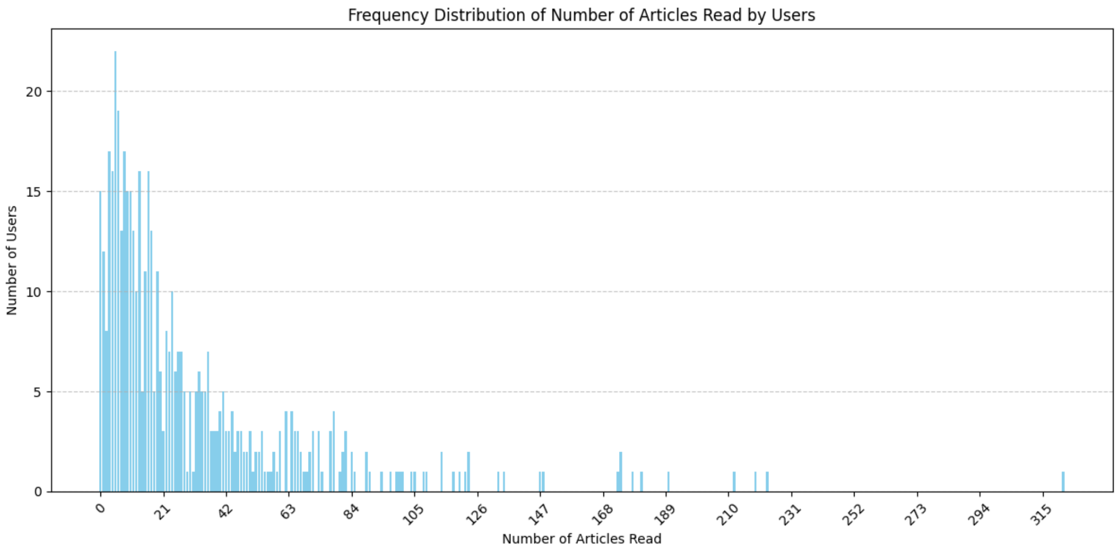

Figure 4. *Frequency distribution of number of Articles read by users.*

We experimented on the Mind-Small dataset [5] which is a subset of the Microsoft News Dataset (MIND), comprised of $51,282$ and $42,416$ English news articles in the training and validation datasets respectively, as shown in Table 1. This dataset provides a rich source of data for researching user behavior and preferences. It contains over 1 million user interactions with news articles. Every news article comprises of rich textual attributes, such as topic/category, subcategory, title, abstract, title topics, and abstract topics. User behavior is represented through impressions, that is, records that log both the news articles

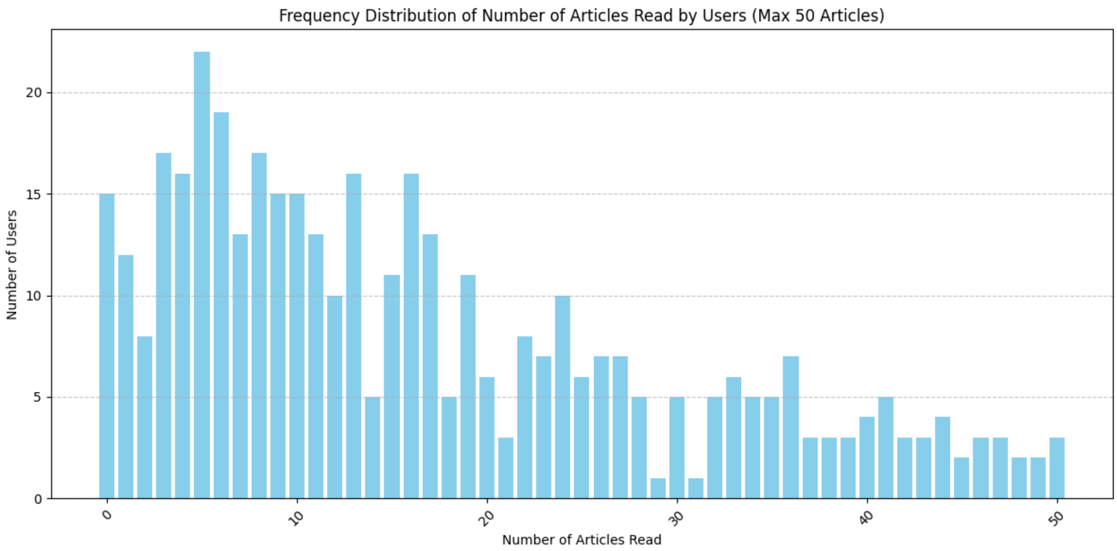

Figure 5. *Frequency distribution of number of article-interactions by users (Max 50 Articles)*

a user has clicked on and those they have not, along with their past click activities leading up to the current impression. The MIND-Small dataset comprises approximately 1 million impressions from 50,000 users from validation, and 50,000 users from testing.

To analyze user engagement within the dataset, Figure 4 shows a histogram illustrating the frequency distribution of the total number of articles read by users. The x-axis indicates the total number of news articles that each user has read, while the y-axis indicates the total number of users. The diversity of engagement patterns demonstrates right-skew, with approximately 65% of users reading fewer than 20 articles. Only 10% of users have read more than 50 articles, indicating a wide gap in engagement levels among users. Since only 10% of users have read more than 50 articles, Figure 5 shows the same frequency distribution of user read articles for a maximum of 50 news articles for better understanding. The right-skewed distribution can now be seen to have some notable peaks at 5, 15, and 25 articles. Figure 5 points out quite elaborately that the frequency of users drops quite considerably for the number of news articles more than 20 [61], with only around 5 or fewer users reading more than 20 i.e. up to 30 or 40 articles. This was one of the primary reasons why we chose to limit the maximum number of news articles considered per user to be 20 (as already mentioned in Table 1). Figure 6 depicts the spread of user involvement across different content topics in the *Skewed Region*, that is, for users who read up to around 20 news articles. The x-axis lists the content categories, and the y-axis indicates the frequency of user engagement. The findings show that the news category dominates user activity. The sports and lifestyle categories follow while other categories, such as movies, music, and finance, show significantly lower engagement. This indicates that the main interest of people in this area appears to be news material then sports and

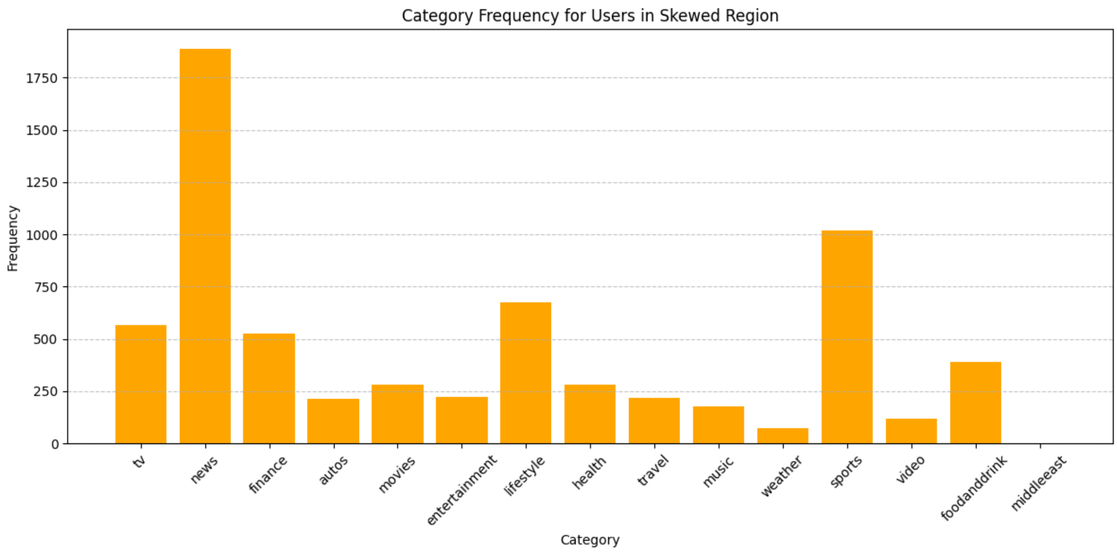

Figure 6. *Category frequency for users in Skewed Region.*

lifestyle. Figure 7 demonstrates that once again, the news is the most frequently engaged category by users in the "Tail Region", or the users who usually read more than 20 articles, significantly outpacing all other categories. Sports and lifestyle are thereafter, although they receive far less attention than news. While areas like music, cars, and movies generate little involvement, other categories like TV, health, money, and food and drink exhibit considerable interest. Users in this area too are mostly interested in news, while they do occasionally pay attention to lifestyle and sports-related information.

Data Preprocessing: In preprocessing stage first, raw data, including news articles and user interaction logs, is mapped and organized into dictionaries and lists, with indexes for news IDs, titles, abstracts, entities, and user histories. Textual data, such as news titles and abstracts, is tokenized using *BertTokenizer* to standardize and prepare it for embedding generation using BERT. Titles are padded to have a maximum of 20 tokens each to generate uniform tensors. Topical representations for each user are padded to have 5 tokens each. Tokens are then transformed into high-dimensional embeddings by BERT, which capture the text's semantic meaning. Attention masks are used to identify which tokens should be considered by the model and which should be ignored, such as padding tokens. Finally, the processed data, including embeddings and attention masks, is organized as tensors. PyTorch was employed for managing tensors. This step prepares the data for direct input into the model.

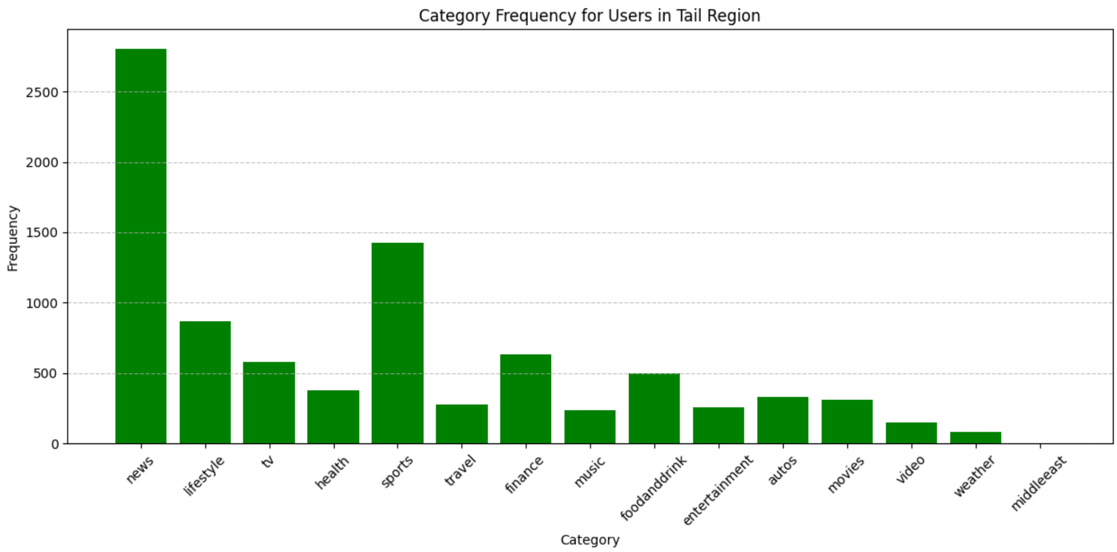

Figure 7. *Category frequency for users in Tail Region.*

4.2 Baselines

In this subsection, we present several state-of-the-art approaches to abstractive summarization, which serve as the baseline methods for our comparisons as follows:

1. **UNBERT** : User-News matching BERT (UNBERT) [32] for news recommendation, is modeled to tackle the cold-start problem by integrating user and news representations both at Word-level (WLM) and News-level (NLM). UNBERT uses these masking strategies to enhance textual understanding and improve recommendation accuracy even for new or infrequent users or items.
2. **MINS** : Multi-interest news sequence (MINS): A GRU-based network [71] generates a multi-interest session representation following a parallel-interest network identifies and routes possible news interests. Additionally, a news encoder with multi-head self-attentions ensures accurate news representation.
3. **CupMar** : CupMar [72] is a deep neural network for news recommendation that integrates user context based on recent and long-term preferences with textual analysis of news articles. The model includes a News Encoder that uses dense layers and self-attention to combine different news aspects, and a User-Profile Encoder that uses GRU-based extractors to extract user preferences from historical data.
4. **GAINRec** : Global trAnsition graph attentIon Network-based news Recommendation model (GAINRec) [73], a model that uses a global transition graph and attention mechanisms to combine aggregate behavior patterns and customized preferences the model's efficacy is demonstrated on real-world datasets.
5. **MINER** : Multi-Interest Matching Network for News Recommendation (MINER)

[28], captures diverse user interests from historical reading behaviors instead of relying on a single user embedding. Several user interest vectors, each representing a distinct element of interest, are learned by MINER using a poly attention approach. Created a category-aware attention method to re-weight historical news based on category similarity, and we added disagreement regularization to the attention mechanism.

6. **CNRCL** : The Contrastive News Recommendations based on Curriculum Learning (CNRCL) [37] for individualized news suggestions. Utilizing curriculum learning, CNRCL customizes negative sampling based on the interests of its users. To improve the relevancy of news suggestions, it involves contrastive learning as an additional task to overcome the sparsity of user click data.

4.3 Evaluation Metrics

We discuss the metrics used to evaluate the performance of ConNewsRec with existing research works as :

- **Area Under the Curve (AUC)**: AUC is a widely use metric for news recommendation systems. The AUC score value ranges from 0 to 1. A higher AUC score indicates a better recommendation accuracy reflecting the model's capability to predict based on users' interests.
- **Mean Reciprocal Rank (MRR)**: MRR is again a commonly used metric in news recommendation systems as here the ranking of news articles matters. The MRR ranking gives a higher importance to items appearing earlier. A higher MRR score indicates that the recommendation model is consistently placing the most relevant news articles towards the top of the recommendation list, thereby demonstrating its effectiveness.
- **Normalized Discounted Cumulative Gain (nDCG)**: nDCG is used for evaluating the ranking quality of the lists generated by news recommendation models. The metric used in this thesis are the nDCG@5 and the nDCG@10, both of which focus on the top 5 and 10 recommendations respectively, emphasizing the importance of presenting highly relevant news articles to a user, early in the list, to maximize satisfaction.

4.4 Hyperparameters

During our experiments, we set the batch size to 36. We applied a dropout probability is 0.1, and utilized the Adam optimizer with a learning rate of 2×10^{-5}. As the number of news articles is small ConNewsRec performance suffers from inadequate user information. The maximum length allowed for a news title is 20, if there are too many news articles, personas might contain noisy objects, which would lower performance. Selecting the top 5 topics from each news article appeared to be the optimal choice. For MIND-Small, we use the pre-trained *Bert Base* model ('bert-base-uncased') with 12 layers, 768 hidden units and 12 attention heads [31]. The values assigned to the hyperparameters λ is 1. We conduct experiments to analyze the impacts of ConNewsRec's on different learning rate. Initially, λ serves as the balancing weight, controlling the ratio of the multi-task loss in Equation 3.24). Competitive results are observed when λ= 1. We then investigate how topical representation construction is affected by the quantity of topics that are extracted from each news article.

4.5 Ablation Study

In this section, we outline an ablation study conducted on our ConNewsRec to compare with different variants, discussed as follows:

- **ConNewsRec-noCL:** We eliminated the user's news topical representation from ConNewsRec emphasizes the importance of incorporating the explicit topical representation in enriching news and user representations for personalized recommendations.
- **ConNewsRec-noAttn:** We omitted the multi-head attention that enhance the representation learning of news articles and user preferences by capturing various aspects of the information.
- **ConNewsRec-R1:** We evaluated this variant of ConNewsRec that employs a random probability ($P(random)$) is set to 0.2. During the training phase, this parameter adds controlled randomness.
- **ConNewsRec-R2:** We evaluated a variant of the ConNewsRec model in which a random probability ($P(random)$) is set to 0.4.
- **ConNewsRec-R3:** We evaluated this variant of ConNewsRec that employs a random probability ($P(random)$) is set to 0.8.

5. Results

In this section, we present the results, including comparisons with baseline models and findings from the ablation studies.

5.1 Comparison with Baselines

Table 2. *An evaluation of our ConNewsRec technique against the existing baselines on the MIND-Small Dataset*

Approach	AUC	MRR	nDCG@5	nDCG@10
ConNewsRec	0.5077	0.2122	0.2859	0.3668
MINER [28]	0.6961	0.3397	0.3762	0.4390
UNBERT [32]	0.6762	0.3172	0.3475	0.4102
MINS [71]	0.6710	0.3171	0.3525	0.4150
CupMar [72]	0.6415	0.2961	0.3289	0.3902
GAINRec [73]	0.6817	0.3287	0.3647	0.4257
CNRCL [37]	0.6935	0.3420	0.3811	0.4420

Table 2 compares the performance of the our proposed ConNewsRec technique with several existing baseline methods on the MIND-Small Dataset.The evaluation metrics include AUC (Area Under the Curve), MRR (Mean Reciprocal Rank), nDCG@5, and nDCG@10, which collectively assess the effectiveness and ranking quality of the models in recommending news articles.

The evaluation of our proposed ConNewsRec model against existing baselines, namely UNBERT, MINS, CupMar, GAINRec, MINER, and CNRCL, highlights some notable differences across four key performance metrics. While ConNewsRec achieves an AUC score of 0.5077, MRR of 0.2122, nDCG@5 of 0.2859, and nDCG@10 of 0.3668, its current implementation does not yet outperform state-of-the-art baseline models like CNRCL or MINER. The UNBERT model achieves an AUC of 0.6762, MRR of 0.3172, nDCG@5 of 0.3475, and nDCG@10 of 0.4102, the MINS model achieves an AUC of 0.6710 and higher nDCG scores of 0.3525 (nDCG@5) and 0.4150 (nDCG@10). The best-performing model, CNRCL, achieves the highest scores with an AUC of 0.6935, MRR of 0.3420, nDCG@5 of 0.3811, and nDCG@10 of 0.4420, which is approximately 20.5% higher than ConNewsRec. Other baselines such as MINER and UNBERT also demonstrate higher performance, with MINER achieving an AUC that is 37.1% higher and an nDCG@10 score that is 19.6% higher than ConNewsRec, demonstrating their ability to capture complicated user behaviors which measures the model's ability to differentiate between relevant and irrelevant news articles.

While ConNewsRec underperforms compared to these baselines, it highlights areas for potential improvements, such as improved integration of content semantics and dynamic user modeling. This analysis underscores the need for further enhancements to improve the model's ranking and recommendation capabilities.

5.2 Ablation Study Results

Table 3. Ablation study results comparing various modifications of the ConNewsRec model

Approach	AUC	MRR	nDCG@5	nDCG@10
ConNewsRec(P 0.6)	0.5077	0.2122	0.2859	0.3668
ConNewsRec-noCL	0.5021	0.1782	0.2585	0.3231
ConNewsRec-noAttn	0.5017	0.1909	0.2606	0.3348
ConNewsRec-R1(P 0.2))	0.5002	0.1780	0.2590	0.322128
ConNewsRec-R2(p 0.4)	0.4988	0.1795	0.2555	0.3239
ConNewsRec-R3(P 0.8)	0.5005	0.1846	0.2581	0.3289

Table 3 shows the ablation study results and illustrates the impact of various components and adjustments on the ConNewsRec model's performance across four metrics: AUC, MRR, nDCG@5, and nDCG@10. The baseline configuration (ConNewsRec(P 0.6)) has the best performance, with an AUC of 0.5077, MRR of 0.2122, nDCG@5 of 0.2859, and nDCG@10 of 0.3668. The ConNewsRec model's (ConNewsRec-noCL) performance, excluding the user's news topical representation from the ConNewsRec model results in a substantial decline in performance across all metrics, with the AUC decreasing from 0.5077 to 0.5021, the MRR dropping from 0.2122 to 0.1782 and nDCG@5, nDCG@10 from 0.2859 to 0.2585, 0.3668 to 0.3231 respectively. This highlights the importance of explicit topical representation in enhancing user and news embeddings for better recommendations. Excluding the multi-head attention mechanism (ConNewsRec-noAttn), also slightly reduces the model's effectiveness, as reflected by a decrease in AUC from 0.5077 to 0.5017, MRR from 0.2122 to 0.1909 and a reduction in nDCG@10 from 0.3668 to 0.3348. This emphasizes how multi-head attention is crucial for gathering many informational facets and improving the representation learning of news articles and user preferences.

The study also explores the effect of ConNewsRec model's performance is affected when random probability ($P(random)$) is varied during training (ConNewsRec-R1, ConNewsRec-R2, ConNewsRec-R3). Models with random probability ($P(random)$) that is ConNewsRec-R1, with ($P(random) = 0.2$), demonstrate the lowest performance across all metrics, with an AUC of 0.5002, MRR of 0.1780, and nDCG@10 of 0.3221, indicating that low randomness reduces the model's ability to generalize. Increasing randomness to ($P(random) = 0.4$) that is ConNewsRec-R2 yields slightly better results, with an AUC of 0.4988 and MRR of 0.1795, while increasing $P(random)$ to 0.8 (ConNewsRec-R3) enhances performance even more, with an AUC of 0.5005 and MRR of 0.1846. However, this indicates that the baseline configuration with $P(random)$ value of 0.6 achieves the most optimal balance for the model.

These findings suggest that controlled randomness during training can influence the model's ability to generalize and learn effective representations, with higher randomness potentially leading to more robust recommendations. Overall, the ablation study highlights the importance of explicit topical representation and multi-head attention in improving the performance of the ConNewsRec model, along with the importance of controlled randomization in making strong recommendations. These findings highlight potential areas for further optimization and refinement.

6. Summary

In this thesis, we have developed the ConNewsRec (Contrastive Learning-based News Recommendation) model, a novel approach to personalized news recommendations that leverages topical news representations and contrastive learning. The model's architecture was designed to improve the semantic similarity between users' preferences and news articles with a multi-head attention mechanism using the users' topical representation, built from his clicked news articles and contrastive learning. Integrating news representation and contrastive learning within a unified framework was done not only to improve the explainability of the recommendation process but also to allow for more precise modeling of user interests. The user encoder in this model is designed to dynamically characterize the user's interests by integrating their long-term stable preferences from the user's history with their recently clicked news articles. Furthermore, contrastive learning was introduced to enhance the discriminative power of the model, empowering it to differentiate more effectively between closely related news articles. Although, our novel approach and present experiments demonstrated that the ConNewsRec underperforms when compared to the existing state-of-the-art methods notably in terms of accuracy, relevance, and personalization of news recommendations. However, the discoveries from this thesis underline the potential of combining topical news representation with contrastive learning to enhance news recommender systems.

Future studies might concentrate on improving topic extraction quality as it directly impacts the accuracy of user profiles. The efficiency and performance of the model could be improved by optimizing data utilization through the implementation of an active learning framework, in which the model requests labels for the most informative data samples. These solutions could improve the robustness and general applicability of the ConNewsRec model in personalized news recommendation systems, while also mitigating some of its current drawbacks.

References

[1] Janette Lehmann et al. "Transient news crowds in social media". In: *Seventh International AAAI Conference on Weblogs and Social Media*. 2013.

[2] Sinan Aral and Michael Zhao. "Social Media Sharing and Online News Consumption". In: *Available at SSRN 3328864* (2019).

[3] Anita Whiting and David Williams. "Why people use social media: a uses and gratifications approach". In: *Qualitative Market Research: An International Journal* 16.4 (2013), pp. 362–369.

[4] Rebecca Dolan et al. "Social media engagement behavior: A framework for engaging customers through social media content". In: *European Journal of Marketing* (2019).

[5] Chuhan Wu et al. "Personalized news recommendation: Methods and challenges". In: *ACM Transactions on Information Systems* 41.1 (2023), pp. 1–50.

[6] Miaomiao Li and Licheng Wang. "A survey on personalized news recommendation technology". In: *IEEE Access* 7 (2019), pp. 145861–145879.

[7] Eliza Mitova et al. "News recommender systems: A programmatic research review". In: *Annals of the International Communication Association* 47.1 (2023), pp. 84–113.

[8] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. "News recommender systems–Survey and roads ahead". In: *Information Processing & Management* 54.6 (2018), pp. 1203–1227.

[9] Chong Feng et al. "News recommendation systems-accomplishments, challenges & future directions". In: *IEEE Access* 8 (2020), pp. 16702–16725.

[10] Maria Panteli et al. "Recommendation Systems for News Articles at the BBC". In: *INRA@RecSys*. 2019. URL: `https://api.semanticscholar.org/CorpusID:212411370`.

[11] Nabila Amir et al. "On the current state of deep learning for news recommendation". In: *Artificial Intelligence Review* 56.2 (2023), pp. 1101–1144.

[12] Florent Garcin et al. "Personalized News Recommendation Based on Collaborative Filtering". In: *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*. Vol. 1. 2012, pp. 437–441. DOI: `10.1109/WI-IAT.2012.95`.

[13] Kevin Joseph and Hui Jiang. "Content based News Recommendation via Shortest Entity Distance over Knowledge Graphs". In: *Companion Proceedings of The 2019 World Wide Web Conference*. WWW '19. San Francisco, USA: Association for Computing Machinery, 2019, pp. 690–699. ISBN: 9781450366755. DOI: 10.1145/3308560.3317703. URL: https://doi.org/10.1145/3308560.3317703.

[14] Hongwei Wang et al. *DKN: Deep Knowledge-Aware Network for News Recommendation*. 2018. arXiv: 1801.08284 [stat.ML].

[15] Keunchan Park, Jisoo Lee, and Jaeho Choi. "Deep Neural Networks for News Recommendations". In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. CIKM '17. Singapore, Singapore: Association for Computing Machinery, 2017, pp. 2255–2258. ISBN: 9781450349185. DOI: 10.1145/3132847.3133154. URL: https://doi.org/10.1145/3132847.3133154.

[16] G Geetha et al. "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System". In: *Journal of Physics: Conference Series* 1000.1 (Apr. 2018), p. 012101. DOI: 10.1088/1742-6596/1000/1/012101. URL: https://dx.doi.org/10.1088/1742-6596/1000/1/012101.

[17] D. Jannach et al. *Recommender Systems: An Introduction*. Cambridge University Press, 2010. ISBN: 9781139492591. URL: https://books.google.ee/books?id=eygTJBd_U2cC.

[18] Abhinandan Das et al. "Google news personalization: scalable online collaborative filtering". In: *The Web Conference*. 2007. URL: https://api.semanticscholar.org/CorpusID:207163129.

[19] Michal Kompan and Maria Bielikova. "Content-Based News Recommendation". In: vol. 61. Sept. 2010, pp. 61–72. ISBN: 978-3-642-15207-8. DOI: 10.1007/978-3-642-15208-5_6.

[20] Lihong Li et al. "A contextual-bandit approach to personalized news article recommendation". In: *The Web Conference*. 2010. URL: https://api.semanticscholar.org/CorpusID:207178795.

[21] Jiameng Bai et al. "KSR: Knowledge-based Sequential News Recommendation System". In: *IOP Conference Series: Materials Science and Engineering* 799.1 (Mar. 2020), p. 012042. DOI: 10.1088/1757-899X/799/1/012042. URL: https://dx.doi.org/10.1088/1757-899X/799/1/012042.

[22] Yong Xu et al. "NRKM: News Recommendation Based on Knowledge Graph with Multi-View Learning". In: *Proceedings of the 2022 3rd International Conference on Control, Robotics and Intelligent System*. CCRIS '22. Virtual Event, China: Association for Computing Machinery, 2022, pp. 123–127. ISBN: 9781450396851. DOI: 10.1145/3562007.3562030. URL: https://doi.org/10.1145/3562007.3562030.

[23] Fangzhao Wu et al. "Mind: A large-scale dataset for news recommendation". In: *Proceedings of the 58th annual meeting of the association for computational linguistics*. 2020, pp. 3597–3606.

[24] Hangbo Bao et al. "Unilmv2: Pseudo-masked language models for unified language model pre-training". In: *International conference on machine learning*. PMLR. 2020, pp. 642–652.

[25] Chuhan Wu et al. *Neural News Recommendation with Attentive Multi-View Learning*. 2019. arXiv: 1907.05576 [cs.CL]. URL: https://arxiv.org/abs/1907.05576.

[26] Chuhan Wu et al. "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)*. Ed. by Kentaro Inui et al. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 6389–6394. DOI: 10.18653/v1/D19-1671. URL: https://aclanthology.org/D19-1671.

[27] Taeho Kim et al. "Is It Enough Just Looking at the Title? Leveraging Body Text To Enrich Title Words Towards Accurate News Recommendation". In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. CIKM '22. Atlanta, GA, USA: Association for Computing Machinery, 2022, pp. 4138–4142. ISBN: 9781450392365. DOI: 10.1145/3511808.3557619. URL: https://doi.org/10.1145/3511808.3557619.

[28] Jian Li et al. "MINER: Multi-Interest Matching Network for News Recommendation". In: *Findings of the Association for Computational Linguistics: ACL 2022*. Ed. by Smaranda Muresan, Preslav Nakov, and Aline Villavicencio. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 343–352. DOI: 10.18653/v1/2022.findings-acl.29. URL: https://aclanthology.org/2022.findings-acl.29.

[29] Shumpei Okura et al. "Embedding-based News Recommendation for Millions of Users". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '17. Halifax, NS, Canada: Association for Computing Machinery, 2017, pp. 1933–1942. ISBN: 9781450348874.

DOI: 10.1145/3097983.3098108. URL: https://doi.org/10.1145/3097983.3098108.

[30] Mingxiao An et al. "Neural News Recommendation with Long- and Short-term User Representations". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Ed. by Anna Korhonen, David Traum, and Lluís Màrquez. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 336–345. DOI: 10.18653/v1/P19-1033. URL: https://aclanthology.org/P19-1033.

[31] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. arXiv: 1810.04805 [cs.CL].

[32] Qi Zhang et al. "UNBERT: User-News Matching BERT for News Recommendation". In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. Ed. by Zhi-Hua Zhou. Main Track. International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, pp. 3356–3362. DOI: 10.24963/ijcai.2021/462. URL: https://doi.org/10.24963/ijcai.2021/462.

[33] Zizhuo Zhang and Bang Wang. "Prompt Learning for News Recommendation". In: *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '23. ACM, July 2023. DOI: 10.1145/3539618.3591752. URL: http://dx.doi.org/10.1145/3539618.3591752.

[34] Kaiming He et al. *Momentum Contrast for Unsupervised Visual Representation Learning*. 2020. arXiv: 1911.05722 [cs.CV]. URL: https://arxiv.org/abs/1911.05722.

[35] Prannay Khosla et al. *Supervised Contrastive Learning*. 2021. arXiv: 2004.11362 [cs.LG]. URL: https://arxiv.org/abs/2004.11362.

[36] Tiening Sun et al. "Rumor detection on social media with graph adversarial contrastive learning". In: *Proceedings of the ACM Web Conference 2022*. 2022, pp. 2789–2797.

[37] Xingran Zhou et al. *A Contrastive News Recommendation Framework based on Curriculum Learning*. Feb. 2024. DOI: 10.21203/rs.3.rs-3968976/v1.

[38] Jingkun Wang et al. "Improving News Recommendation with Channel-Wise Dynamic Representations and Contrastive User Modeling". In: *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*. WSDM '23. Singapore, Singapore: Association for Computing Machinery, 2023, pp. 562–570. ISBN: 9781450394079. DOI: 10.1145/3539597.3570447. URL: https://doi.org/10.1145/3539597.3570447.

[39] Shicheng Wang et al. “Multi-interest Extraction Joint with Contrastive Learning for News Recommendation”. In: *Machine Learning and Knowledge Discovery in Databases*. Ed. by Massih-Reza Amini et al. Cham: Springer International Publishing, 2023, pp. 606–621. ISBN: 978-3-031-26387-3.

[40] Yuhao Yang et al. “Knowledge Graph Contrastive Learning for Recommendation”. In: *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, July 2022, pp. 1434–1443. DOI: 10.1145/3477495.3532009. URL: http://dx.doi.org/10.1145/3477495.3532009.

[41] Lei Chen and Meimei Xia. “A context-aware recommendation approach based on feature selection”. In: *Applied Intelligence* 51 (2021), pp. 865–875.

[42] Paul Resnick et al. “Grouplens: An open architecture for collaborative filtering of netnews”. In: *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. 1994, pp. 175–186.

[43] J. Ben Schafer et al. “Collaborative Filtering Recommender Systems”. In: *The Adaptive Web: Methods and Strategies of Web Personalization*. Ed. by Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 291–324. ISBN: 978-3-540-72079-9. DOI: 10.1007/978-3-540-72079-9_9. URL: https://doi.org/10.1007/978-3-540-72079-9_9.

[44] Feng Xue et al. “Deep Item-based Collaborative Filtering for Top-N Recommendation”. In: *ACM Trans. Inf. Syst.* 37.3 (Apr. 2019). ISSN: 1046-8188. DOI: 10.1145/3314578. URL: https://doi.org/10.1145/3314578.

[45] Zhi-Dan Zhao and Ming-sheng Shang. “User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop”. In: *2010 Third International Conference on Knowledge Discovery and Data Mining*. 2010, pp. 478–481. DOI: 10.1109/WKDD.2010.54.

[46] John S. Breese, David Heckerman, and Carl Kadie. “Empirical analysis of predictive algorithms for collaborative filtering”. In: *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*. UAI’98. Madison, Wisconsin: Morgan Kaufmann Publishers Inc., 1998, pp. 43–52. ISBN: 155860555X.

[47] Amir Bakarov. “A survey of word embeddings evaluation methods”. In: *arXiv preprint arXiv:1801.09536* (2018).

[48] Hongwei Wang et al. *DKN: Deep Knowledge-Aware Network for News Recommendation*. 2018. arXiv: 1801.08284 [stat.ML]. URL: https://arxiv.org/abs/1801.08284.

[49] Fangzhao Wu et al. "MIND: A Large-scale Dataset for News Recommendation". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Ed. by Dan Jurafsky et al. Online: Association for Computational Linguistics, July 2020, pp. 3597–3606. DOI: 10.18653/v1/2020.acl-main.331. URL: https://aclanthology.org/2020.acl-main.331.

[50] Chuhan Wu et al. "NPA: Neural News Recommendation with Personalized Attention". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*. KDD '19. ACM, July 2019. DOI: 10.1145/3292500.3330665. URL: http://dx.doi.org/10.1145/3292500.3330665.

[51] Jacob Devlin et al. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. 2019. arXiv: 1810.04805 [cs.CL].

[52] Chuhan Wu et al. *Empowering News Recommendation with Pre-trained Language Models*. 2021. arXiv: 2104.07413 [cs.IR].

[53] Qi Zhang et al. "UNBERT: User-News Matching BERT for News Recommendation". In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. Ed. by Zhi-Hua Zhou. Main Track. International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, pp. 3356–3362. DOI: 10.24963/ijcai.2021/462. URL: https://doi.org/10.24963/ijcai.2021/462.

[54] Paul Covington, Jay Adams, and Emre Sargin. "Deep neural networks for youtube recommendations". In: *Proceedings of the 10th ACM conference on recommender systems*. 2016, pp. 191–198.

[55] Thomas Kipf, Elise van der Pol, and Max Welling. *Contrastive Learning of Structured World Models*. 2020. arXiv: 1911.12247 [stat.ML].

[56] Ching-Yun Ko et al. *Revisiting Contrastive Learning through the Lens of Neighborhood Component Analysis: an Integrated Framework*. 2022. arXiv: 2112.04468 [cs.LG].

[57] Jizong Peng et al. *Self-Paced Contrastive Learning for Semi-supervised Medical Image Segmentation with Meta-labels*. 2021. arXiv: 2107.13741 [cs.CV].

[58] Senthil Purushwalkam and Abhinav Gupta. *Demystifying Contrastive Self-Supervised Learning: Invariances, Augmentations and Dataset Biases*. 2020. arXiv: 2007.13916 [cs.CV].

[59] Yonglong Tian et al. *What Makes for Good Views for Contrastive Learning?* 2020. arXiv: 2005.10243 [cs.CV].

[60] Mohammadreza Zolfaghari et al. *CrossCLR: Cross-modal Contrastive Learning For Multi-modal Video Representations*. 2021. arXiv: 2109.14910 [cs.CV].

[61] Rui Liu et al. *PerCoNet: News Recommendation with Explicit Persona and Contrastive Learning*. 2023. arXiv: 2304.07923 [cs.IR].

[62] Udhaya Kumar Rajendran and Amine Trabelsi. "Exploration of Contrastive Learning Strategies toward more Robust Stance Detection". In: *Proceedings of the 13th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*. Ed. by Jeremy Barnes, Orphée De Clercq, and Roman Klinger. Toronto, Canada: Association for Computational Linguistics, July 2023, pp. 431–440. DOI: 10.18653/v1/2023.wassa-1.37. URL: https://aclanthology.org/2023.wassa-1.37.

[63] Zhuofeng Wu et al. *CLEAR: Contrastive Learning for Sentence Representation*. 2020. arXiv: 2012.15466 [cs.CL].

[64] Xin Xia et al. *Self-Supervised Hypergraph Convolutional Networks for Session-based Recommendation*. 2022. arXiv: 2012.06852 [cs.IR].

[65] Tiening Sun et al. "Rumor Detection on Social Media with Graph Adversarial Contrastive Learning". In: *Proceedings of the ACM Web Conference 2022*. WWW '22. <conf-loc>, <city>Virtual Event, Lyon</city>, <country>France</country>, </conf-loc>: Association for Computing Machinery, 2022, pp. 2789–2797. ISBN: 9781450390965. DOI: 10.1145/3485447.3511999. URL: https://doi.org/10.1145/3485447.3511999.

[66] Jiaying Zou et al. "Zero-Shot Stance Detection via Sentiment-Stance Contrastive Learning". In: *2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI)*. 2022, pp. 251–258. DOI: 10.1109/ICTAI56018.2022.00044.

[67] *Rectifier (neural networks)*. URL: https://en.wikipedia.org/w/index.php?title=Rectifier_(neural_networks)&oldid=1261563230.

[68] David M Blei, Andrew Y Ng, and Michael I Jordan. "Latent dirichlet allocation". In: *Journal of machine Learning research* 3.Jan (2003), pp. 993–1022.

[69] Inggrid Resmi Benita and ZKA Baizal. "News Recommender System Based on User Log History Using Rapid Automatic Keyword Extraction". In: *JURNAL MEDIA INFORMATIKA BUDIDARMA* 6.4 (2022), pp. 2341–2345.

[70] Peitian Zhang et al. "GateFormer: Speeding Up News Feed Recommendation with Input Gated Transformers". In: *arXiv preprint arXiv:2201.04406* (2022).

[71] Rongyao Wang and Wenpeng Lu. *Modeling Multi-interest News Sequence for News Recommendation*. 2022. arXiv: 2207.07331 [cs.IR]. URL: https://arxiv.org/abs/2207.07331.

[72] Dai Hoang Tran et al. "CupMar: A deep learning model for personalized news recommendation based on contextual user-profile and multi-aspect article representation". In: *World Wide Web* 26 (2022), pp. 713–732. URL: https://api.semanticscholar.org/CorpusID:248704896.

[73] Qing Meng et al. "Recognize News Transition from Collective Behavior for News Recommendation". In: *ACM Trans. Inf. Syst.* 41.4 (Apr. 2023). ISSN: 1046-8188. DOI: 10.1145/3578362. URL: https://doi.org/10.1145/3578362.

Appendix 1 – Non-Exclusive License for Reproduction and Publication of a Graduation Thesis[1]

I, Deepika Uttam Sambrekar

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis "Contrastive Learning based News Recommendation", supervised by Dr. Roshni Chakraborty and Dr. Gert Kanter
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

02.01.2025

[1]The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.