



News recommender system: a review of recent progress, challenges, and opportunities

Shaina Raza¹  · Chen Ding¹

Accepted: 9 July 2021 / Published online: 21 July 2021
© Crown 2021

Abstract

Nowadays, more and more news readers read news online where they have access to millions of news articles from multiple sources. In order to help users find the right and relevant content, news recommender systems (NRS) are developed to relieve the information overload problem and suggest news items that might be of interest for the news readers. In this paper, we highlight the major challenges faced by the NRS and identify the possible solutions from the state-of-the-art. Our discussion is divided into two parts. In the first part, we present an overview of the recommendation solutions, datasets, evaluation criteria beyond accuracy and recommendation platforms being used in the NRS. We also talk about two popular classes of models that have been successfully used in recent years. In the second part, we focus on the deep neural networks as solutions to build the NRS. Different from previous surveys, we study the effects of news recommendations on user behaviors and try to suggest possible remedies to mitigate those effects. By providing the state-of-the-art knowledge, this survey can help researchers and professional practitioners have a better understanding of the recent developments in news recommendation algorithms. In addition, this survey sheds light on the potential new directions.

Keywords News · Recommender system · Beyond-accuracy · Evaluation measures · Datasets · User behavior · Deep learning

1 Introduction

With the advancement in interactive communication technology, the internet has become a major source of news due to its 24/7 availability, instant updating and free distribution. According to a report by Pew Research Center Journalism in 2018,¹ roughly nine-in-ten adults (93%) in US tend to read news online (either mobile or

¹ <https://www.journalism.org/fact-sheet/digital-news>.

✉ Shaina Raza
shaina.raza@ryerson.ca

Chen Ding
cding@ryerson.ca

¹ Ryerson University, Toronto, Canada

desktop) through digital newspapers, social media, news apps, etc. Despite such an advancement in technology, the studies have shown that online media does not define significantly different criteria for newsworthiness (Shoemaker 2006) than printed media. One reason for this could be the lack of prescribed procedures to offer a wide variety of news in a timely manner and the inability of the system to model user behaviors in a better way. Therefore, there is a need to move towards tools and techniques such as recommender systems (Adomavicius and Tuzhilin 2005) to provide news updates tailored to readers' information needs.

Many news sources and agencies such as CNN, BBC, New York Times, The Washington Post provide anytime, anywhere access to news readers so that they browse through latest news using online portals. To attract higher volume of traffic to their websites, these online portals are increasingly adopting recommender systems to improve user experience on their sites. The term 'user experience' may have different interpretations in a recommendation domain, such as usability, usefulness, effectiveness or satisfactory interaction with the system (Konstan and Riedl 2012; Knijnenburg et al. 2012). The task of recommending appropriate and relevant news stories to news readers is challenging. The reason is that the news domain is faced with certain challenges that are different from those of other application domains of recommender systems.

Among these unique challenges, timeliness is one of the most important challenges. It takes into account factors such as very short duration of news stories, their recency, popularity, trends, and a high magnitude of news stories arriving every second. Another important challenge in news domain is the highly dynamic user behavior. News readers may have long-term or short-term preferences that evolve over time, either gradually or abruptly. Recently, there is a considerable amount of manipulation taking place with the news content. For example, deceptive information is disseminated to the public in the form of false news and propaganda (Helberger 2019). This has given rise to an uprising challenge in terms of quality control of the news content.

As mobile technologies and applications become more prevalent in people's lives, news feeds from news aggregators (such as Google, Yahoo) and social media (such as Facebook and tweets) have taken over how people discover the news content. Once a news portal's recommendation functionality is installed, news feeds can be algorithmically tailored for each user. Personalization is a useful feature of NRS since it gives news based on the preferences and interests of a news reader. However, overly personalized news stories limit readers' exposure to different types of news. At the individual level, a news reader may get bored of reading similar types of news stories all the time. Over-personalization may also affect a reader's behavior in the long run, causing them to avoid counter-attitudinal (attitude that contradicts one's own beliefs) information (viewpoints, opinions) (Helberger 2019). This type of behavior, at the societal level, poses a threat to democracy in the form of people's denial of opposing viewpoints.

Too much personalization in an NRS is often the result of recommendation approaches that place too much emphasis on prediction accuracy. These typical accuracy-centric approaches may fail to consider other aspects of subjective user experiences (such as choice satisfaction, perceived system effectiveness, better recommendations, and exposure to different points of view) when evaluating the recommendation quality. When developing a good NRS, one must consider the beyond-accuracy aspects to evaluate the quality of news recommendations.

1.1 Previous surveys and challenges discussed

In addition to the NRS-related papers, we also reviewed the previous surveys to see what they had covered. The challenges addressed in the literature often correspond to what is being investigated in the research during that time. For example, in classical NRS surveys (Borges and Lorena 2010; Karwa 2015; Dwivedi and Arya 2016), issues such as personalization, accuracy, cold-start problem, and scalability have been discussed. In some later NRS surveys (Karimi et al. 2018; Chakraborty et al. 2019), the new issues addressed (in addition to those covered in previous surveys) are beyond-accuracy aspects. Recently, the NRS surveys (Li and Wang 2019; Feng et al. 2020; Qin and Lu 2020) have covered topics such as cold start, news content and feature engineering, and changing user preferences. The challenges discussed by each of these surveys are listed in Table 1.

Each of the preceding surveys revealed a few issues related to the news recommendation problem. However, their discussions are mostly from the perspective of computer scientists, ignoring the effects of news recommendation on user behaviors. Also, in the past few years, deep learning has become a popular option for building recommender systems, but they were not included in these surveys. Below we list the major differences between our paper and the previous surveys on the NRS.

1. In the previous surveys, the common challenges related to the news domain were considered. In addition to these common challenges, such as timeliness and user modelling, we discuss new challenges such as content quality and the effects of news recommendations on user behaviours. We provide an overview of the state-of-the-art research that addresses these new challenges.
2. We focus on the most popular recommendation models that are successfully used to build the NRS, with a special emphasis on deep learning-based models due to a lack of coverage on this topic in previous surveys.
3. The impact of news recommendations on user behaviours is a growing concern in the news industry. Although this issue has been raised by online journalism (Möller et al. 2018; Helberger 2019), we believe that it is also related to the discipline of computer science and information systems. Thus, different from previous surveys, we discuss changes in user behaviours that come in effect after recommendations. We also discuss possible remedies from computer science, psychology and journalism that do exist but have not been fully applied in recommender systems to mitigate those post-algorithmic news recommendation effects. In the discussion section, we also offer our own ideas of possible remedy approaches.

1.2 Searching strategy, scope and research trends

In this survey, we have defined a searching strategy, scope, research goals and objectives to classify the literature. We take a neutral stance while reviewing the papers to avoid any risk of bias in the included studies. We identify and select the following collections of bibliographies: ACM Digital Library, SpringerLink, IEEE Xplore and Elsevier, to find the pertinent literature. Besides those bibliographies, we use the following scholarly search engines: GoogleScholar, DBLP, CiteSeerX, MS Academic Search, Web of Science, ScienceDirect and ResearchGate, to find the related papers. We also browse the conference proceedings and journal transactions to look for the titles and abstracts to find more papers

which might have been initially skipped in the earlier search. We specify mid-year 2012 as the starting date and early 2021 as the closing date for our literature review. Besides this specified time frame, we also include a few classical and a few latest publications because of their relevancy to the topic.

We use the Boolean search query (“News”) AND (“Recommender System” OR “Recommendation System” OR “Recommendations”) OR (“Deep Learning”) AND (“News Recommendations” OR “News Recommenders”) to search the bibliographies with the following inclusion criteria: (i) papers written in English and (ii) relevancy and usefulness to the topic. Processing all the papers strictly was not practical. As a result, we decided to include only journal and conference papers, excluding grey literature, workshop presentations, and papers that report abstracts or presentation slides. Out of around 156 papers from the data extraction process, we finalize around 126 papers, out of which 92 are the manuscripts that proposed or designed an NRS, 8 are the survey papers and 26 are those that help us study the nature of the news domain. Out of the last group of papers, some articles are from journalism, general recommendation and information filtering domains. Figure 1 shows the approximate number of NRS papers considered in this time frame in a per-year basis.

The figure clearly shows the increasing amount of research and demand for NRS in the field of recommender systems. The increase in the trendline in the later years is credited to the CLEF NEWSREEL Challenge (Brodthorn and Hopfgartner 2014) as well as the emergence and development of deep learning based recommender systems. The CLEF NEWSREEL platform (a campaign-style evaluation lab) was designed to encourage researchers to develop novel recommenders for news domains, so we see a clear rise in the number of publications during the years 2015–2017. Although it ended in 2018, due to the continuation pattern, we still see many papers in 2018. The effect of its ending is reflected in 2019, hence we see less work in 2019. Since 2016, there is a gradual increase of papers on deep learning-based recommender systems, both in the general domain and in the news domain. A higher number of publications in the year 2020 is possibly credited to the benchmark dataset MIND (from Microsoft). This trend is expected to continue in the year 2021, when the MIND dataset is released for the news recommendation challenge.²

The primary goal of this survey paper is to highlight the most pressing challenges in the NRS that affect user behaviors at various stages of the news recommendation life cycle (before, during, and after).

The rest of the paper is organized as follows. In Sect. 2, we highlight the characteristics of the news domain. In Sect. 3, we present an overview of research on the NRS. In Sect. 4, we describe the conventional algorithmic solutions for addressing the major challenges in the NRS. In Sect. 5, we focus on the deep learning-based solutions to the NRS. In Sect. 6, we explain the effects of news algorithms on user behaviors. We discuss the research implications and future work in this field in Sect. 7. Finally, we conclude the survey in Sect. 8. A mind map diagram is given in Fig. 2 that shows the evolution of this survey.

² <https://msnews.github.io/competition.html>.

Table 1 Challenges discussed in different NRS surveys

Survey	Challenges discussed
Borges and Lorena (2010)	Accuracy
Karwa (2015)	Cold-start, data sparsity, recency, implicit user feedback, changing interests of users, scalability, unstructured content
Dwivedi and Arya (2016)	Data sparsity, changing users' interests, news content, recommendation techniques
Karimi et al. (2018)	Recommendation paradigms, user modeling, cold start, data sparsity, recency, beyond-accuracy measures, scalability
Chakraborty et al. (2019)	Recency, relevance, diversity, accuracy, recommendation techniques
Li and Wang (2019)	Recency, popularity, massive and unstructured data
Qin and Lu (2020)	News content feature engineering
Feng et al. (2020)	Cold start, explicit user feedbacks, changing users' interests

2 Characteristics of news domain

Before reviewing the challenges specific to NRS, we first highlight the characteristics that distinguish the news domain from other application domains of recommender systems such as recommending movies, music, books, restaurants or such.

Average Consumption Time Typically, the duration of consuming a news story (time taken by a user to read a news article) is measured in terms of the article length which on average is under 200 words. According to a report by PEW research center,³ stories under 250 words require readers an average of 43 s in terms of the engagement time, whereas stories whose word count exceeds 5,000 engage people for at least 270 s (4.5 min). Compared to this, a movie is typically 90~120 min long, a music item on average is between 3 and 5 min long and a book may take even longer duration.

Lifespan of News Items News items typically have shorter shelf-lives as they expire quite soon (maybe minutes, hours or barely a few days) compared to other products such as music, books, movies that may span several days, weeks, months or even years. Also, the gap between a news item's release time and time of reviews (comments) on news sites or social media sites is minimal (second, minute, hour or so) compared to other products.

Catalog Size of News Items News stories tend to flood the system within a very short span of time, for example at the rate of thousands of incoming news items per hour. On the other hand, the catalog size of music or movie services may range typically in hundreds or thousands, but these items stay for longer time periods.

Expected Request-Response Rate Timely delivery of the news content is vital and considered as a unique characteristic in the news domain. The requests for news items on a news aggregator site is sometimes greater than 100/sec and the expected response should preferably be sent within 100 ms in order to provide news in real-time (Kille et al. 2017).

Sequential Consumption News items are often consumed in a sequence where a reader may want to be updated about different news stories at a time. The difference between the sequential consumption of music items and news items is that in the former

³ <https://www.journalism.org/2016/05/05/2-analysis-suggests-some-readers-willingness-to-dig-into-long-form-news-on-cellphones/>.

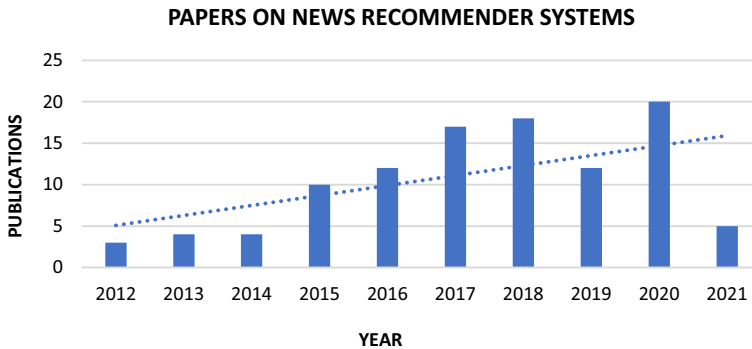


Fig. 1 Number of papers on NRS per year from mid-year 2012 till early year 2021

case the items are often repeated more than once within a sequence (Schedl et al. 2018), whereas in the latter case, the readers want to be updated with different or ongoing stories rather than the repeated stories (Park et al. 2017a).

Diversity A user usually consumes one music or movie genre at a time and occasionally switches to a different genre when in a different mood or situation. On the other hand, the diversity in news domain is crucial not only to keep readers engaged during the online reading process but also to expose readers to counter-attitudinal behavior (Raza and Ding 2020). Diversity in news media is a key principle for a democratic society (Helberger 2019).

Consumption Behavior News items are often consumed anonymously and mostly without explicit user profiles (Doychev et al. 2015; Sottocornola et al. 2018). Though this problem can be mitigated by considering implicit signals like click patterns, reading time spent on an item, browsing and navigational patterns (Ilievski and Roy 2013; Trevisiol et al. 2014), these implicit signals may sometimes be wrongly interpreted as an indicator of user's appreciation or interests. For example, longer reading time could be because of the user's fatigue or idle time and may not be an indicator of the user interest (Ma et al. 2016).

Privacy Concern Online media consumption has also resulted in the threat to users' privacy through excessive analyses on readers' data (Desarkar and Shinde 2014).

Reading Context Reading context is highly evolving, time-ordered and social, and is specific to the news domain (Raza and Ding 2020). The most widely used contexts in NRS are location (Asikin and Wörndl 2014) and time (Park et al. 2017b). Lommatzsch et al. (Lommatzsch et al. 2017) evaluated users' dynamics with respect to the context of time and day of the week. According to their findings, there are more visitors on news portals during working days than on weekend. In addition to time and location, a reader's context may relate to some latest event or trending news, weather or even some personality trait (mood, interest). For example, during the Olympic games, people who are usually not interested in sports news may want to get update on the latest results of some games.

Impact of Social Media Social media has greatly influenced the way news stories are searched and gathered (Cucchiarelli et al. 2018). Readers like to learn more about a news story by tracking its impact on social media. The dialogue, duration, public reactions and outcomes of a news story on social media may also help the journalists to determine which issues need further attention.

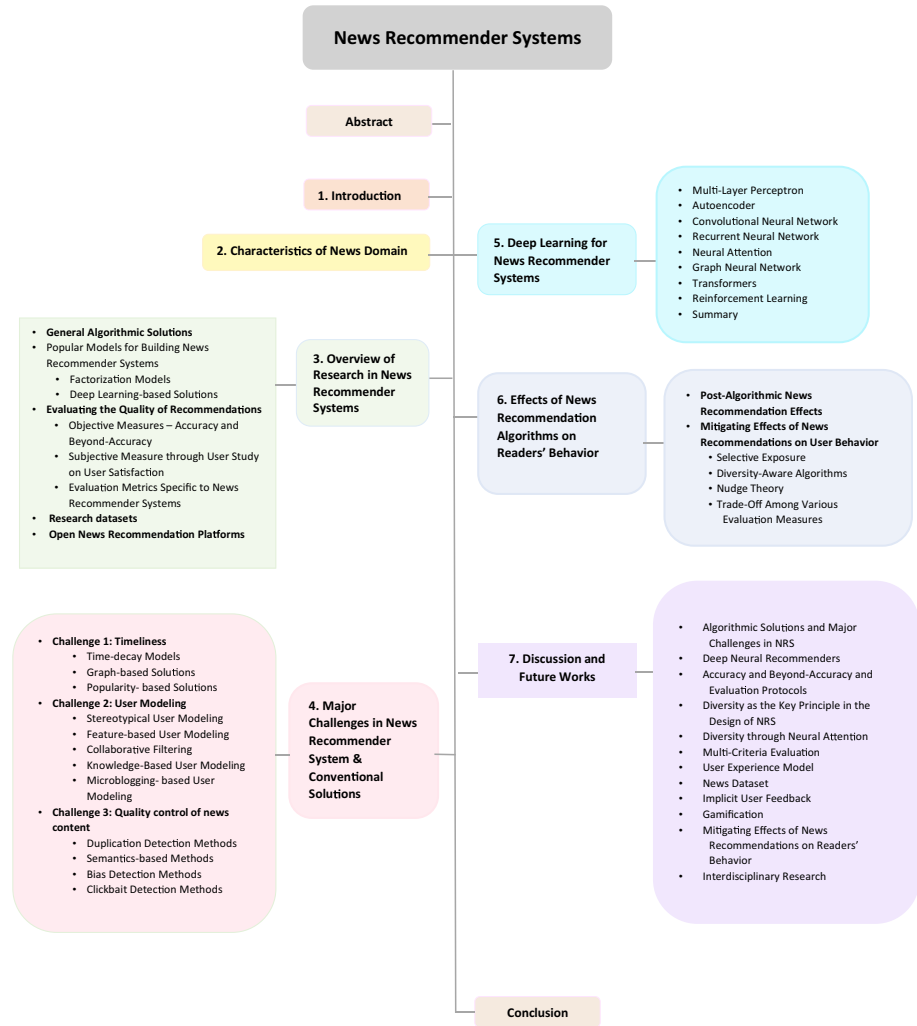


Fig. 2 Mind map diagram of the survey paper

Emotions Emotion commands attention and creates feelings in a reader for the event/character. A music or movie item intuitively evokes emotions in users, which in turn affect their preferences. Emotions are increasingly driving the news consumption behavior and they are both a challenge to the quality of what is produced, and also a chance for the NRS to further reinvent itself (Beckett and Deuze 2016).

Biases News items are initially consumed for information purpose; however, biases can be invoked through presenting news in different styles and tones (Helberger 2019). A good news story should be one that offers details to the readers so that they can make their own judgement and forge an emotional connection with a character/event.

Multimodal News Information In today's information age, the Web is critical for disseminating information and news. Social media, in particular, can easily notify users of global events and has grown in popularity as a big source of news. These news articles

often use multiple modalities, such as texts, videos, podcasts, to convey information more effectively. When it is in the text format, it can be delivered in different languages. Most of the research work today focus on the text-based news articles in one language, without considering the complications brought by multiple modalities and languages, since it is challenging to quantify the cross-modal and cross-language entity representations in today's news domain. Due to a lack of active research on recommending news in the non-textual format and in multiple languages, in this survey, we only review the papers on recommending text-based news in one language. However, we do recognize the need to have more research on multi-modal and multi-language news recommendation.

3 Overview of research in news recommender systems

We present an overview of the NRS research in this section. In Sect. 4, we present the major challenges for the NRS and some conventional solutions to address them. In Sect. 5, we present the deep learning-based NRS.

3.1 General algorithmic solutions

The traditional algorithms used in recommender systems can be classified as: collaborative filtering (CF), content-based filtering (CBF) and hybrid approaches (Adomavicius and Tuzhilin 2005). There are two important things required to build any recommender systems, i.e., the content of the users and items, and their underlying interactions. A CBF algorithm builds a recommender by comparing the user-profile and item-profile based on the content of a shared attribute space. Contrary to this, the CF approach is content-free where the features of items are often not known in prior. CF exploits user behaviors in terms of ratings, history and interactions on items.

While these traditional recommendation algorithms can be applied to the news domain, their performance may not be good. There are various scenarios that we need to consider, such as the dynamics of the news environment, relevance of the news items and users' interests that are highly context dependent. Though CF can be used to address the problem of dynamic content generation of news items, it requires a sufficient amount of users' interactions (stored as histories) to make recommendations. By the time NRS manages to collect enough consumption data from users, the value of news content is decayed, thus making recommendations obsolete. The CBF, on the contrary, can address users' evolving interests by always updating the user profiles with the latest news they have read (Wang et al. 2018b). However, CBF cannot handle the large number of temporary and anonymous users that are common in an NRS. Also, the statistical methods to compute the similarity between user-item profiles in a CBF, may fail to capture the semantics and the contexts in news data. To remedy the pitfalls of both CF and CBF in the NRS, researchers and designers propose hybrid solutions to news recommendations by combining these two types of algorithms. In the past few years, researchers also began to focus on the context (situation such as time, location, mood, etc.) as the additional information to improve the quality of news recommendations. An analysis of 79 (out of 92) papers on NRS in our survey is shown in Fig. 3.

The statistics in Fig. 3 show that CBF is the most used algorithm to solve the problem of news recommendations. Since CBF methods are primarily based on the content metadata to produce recommendations, it is much easier for the researchers and developers to

develop an NRS. The hybrid comes up as the second most popular choice to build an NRS. The CF comes as the next popular choice (also the least popular among the three).

3.2 Popular models for building news recommender systems

Many models have been used in the past to build an NRS. One of the most popular and successful classes of models for the NRS is the latent factor model, especially factorization methods. In recent years, the deep learning-based solutions have come up as an emerging branch of recommender systems. We consider them as the other most popular class of models successfully used for the NRS. These models are briefly covered below.

3.2.1 Factorization models

Factorization method is a class of algorithms used in recommender systems that work by decomposing the user-item interaction matrix into a product of lower dimensional matrices. Here we discuss the factorization models used in the NRS research.

3.2.1.1 Matrix factorization (MF) Matrix factorization is one of the most popular recommendation algorithms that got its first recognition in the Netflix competition (Koren et al. 2009). Matrix factorization can be used to discover the latent features that exhibit in the interactions between two different types of entities (e.g., users and items). In a recent NRS (Raza and Ding 2019), the MF is extended to include the news-related information and to model the temporal dynamics in readers' behaviors. This work introduces a novel predictor to include various temporal effects in the MF model, including time bias, user bias, and item bias. These added biases tend to capture much of the observed signals, especially the temporal dynamics.

3.2.1.2 Non-negative matrix factorization (NMF) NMF, like the MF method, is a decomposition technique in which the matrix R is split into the product of two matrices U and V . However, unlike MF, the NMF has the property that none of the three matrices R , U , and V have any negative elements. Typically, there are many missing user-item interactions in an NRS, resulting in very sparse matrices. In such situations, the NMF models usually perform better than the original MF. This is due to the default functionality of the NMF algorithm in addressing the missing-value assumption (Gillis 2020). However, Singular Value Decomposition (SVD)-based MF may produce better results if the ratings matrix is not too sparse.

In a related NRS (Yan et al. 2012), news-related information is included into the NMF model, where NMF is used for clustering news documents and topic discovery. In another paper (Shu et al. 2019), the NMF is used to learn latent space embeddings from news content and user-news interactions.

3.2.1.3 Tensor factorization (TF) TF extends the MF model by introducing the latent vectors with additional dimension(s). TF-based recommender systems go beyond the limitations of MF techniques by considering additional information about users and items, which results in more accurate recommendations (Frolov and Oseledets 2017). So, TF methods are useful in NRS scenarios where we need to consider more contextual recommendations, such as time, location, and social interactions. However, including too many dimensions may result in costly computations.

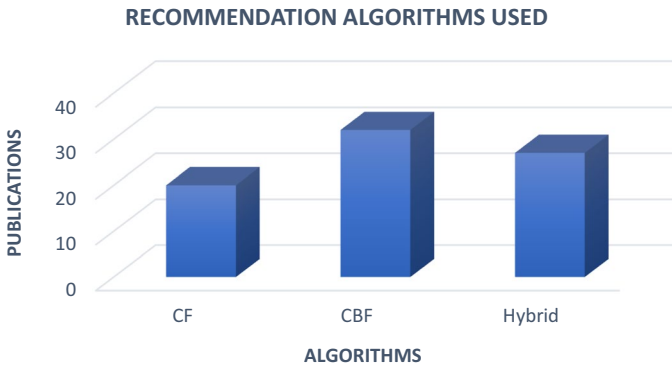


Fig. 3 Algorithms used in NRS

In a related NRS (Wang et al. 2015), TF is used to include the contextualized information related to news items and news readers into the recommendation model.

3.2.1.4 Probabilistic matrix factorization (PMF) PMF (Mnih and Salakhutdinov 2007) is a type of MF models with Gaussian observation noise. PMF is a variation of MF methods that takes its intuition from the Bayesian learning for parameter estimation. This model scales linearly with the number of observations and performs well on large, sparse, and highly imbalanced datasets like those found in the news domain.

In a social recommender system, PMF is used to combine social network structure and user-item rating matrix (Ma et al. 2008). The same idea is used in an NRS (Lin et al. 2012), which incorporates news content, user interactions, and social network information into a PMF model to address the data sparsity issue.

3.2.1.5 Bayesian personalized ranking (BPR) A general limitation of the traditional item prediction methods (e.g., MF methods) is that they are not optimized for ranking the items (e.g., news items). The BPR optimization uses pairs of items to produce more personalized rankings for each user. The MF models can also be used with the BPR to provide users with a personalized and ranked list of items (Rendle et al. 2012).

In a related NRS (Xia et al. 2014) based on a Bayesian model, the readers are recommended with the latest news stories by calculating the joint probability of the news. In another NRS (Gharahighehi and Vens 2019), an extension of BPR is proposed that uses the user consumption levels to recommend news topics to the readers.

3.2.1.6 Generalized linear modeling (GLM) A CF method is often formulated for the prediction of unobserved ratings in a large and mostly empty rating matrix. Though not strictly an MF methodology, both methods (MF and GLM) have their origins in latent factor models. The GLM (McCullagh 2019) can also be used together in conjunction with the CF, where it can use its probabilistic modelling to factorize a high-dimensional rating matrix. In a recent NRS (Raza and Ding 2020), the knowledge is transferred from a high-dimensional news domain that is factorized using GLM into a CF model. The CF model is then used to predict and recommend news items for the users.

3.2.1.7 Neural extensions Much of the recent research in recommender systems is based on creating the neural extensions of these successful latent factor methods discussed above. For example, Neural Network Matrix Factorization (NNMF) (Dziugaite and Roy 2015) replaces the inner product in the PMF formulation with a neural network, and is able to learn an appropriate nonlinear function of user and item latent variables. Neural Collaborative Filtering (NCF) (He et al. 2017) extends the CF model, and the Deep Matrix Factorization (DFM) (Xue et al. 2017) extends the traditional MF model to map the users and items into a common low-dimensional space with non-linear projections. These models continue to inspire NRS researchers, resulting in several useful news recommendation models.

3.2.2 Deep learning-based solutions

The deep learning based NRS began to evolve in the later years, i.e., since 2016 (Karatzoglou et al. 2016). In our survey, we found more than 30 papers published since 2017 that use deep neural networks to solve the news recommendation problem. The rising popularity of these methods shows that deep learning will become the most popular methods in the near future to work in this domain. The general statistics of deep learning based NRS are shown in Fig. 4.

As shown in Fig. 4, the approach to deep learning is increasingly being employed to develop NRS solutions with every coming year. The number is lower in 2021, mainly because it is only the middle of 2021 at the time of writing of this paper and many papers are not yet published or posted online. We will go over the deep learning-based models to news recommendations in Sect. 5.

3.3 Evaluating the quality of recommendations

We categorize the evaluation measures in NRS into two types: objective measures—accuracy and beyond-accuracy, and subjective measures through the user study on user satisfaction. Below we review measures under each category and how they are used in different research work. The definitions for the actual evaluation metrics that have been used in NRS so far and which categories they belong to are given in Table 2.

3.3.1 Objective measures: accuracy and beyond-accuracy

The goal of a recommender system is to predict how likely users would enjoy the unknown items based on what the system has known about them. Therefore, much of the early work in recommender systems focused on providing recommendations to the users according to their preferences. These recommender systems have been evaluated according to accuracy metrics that measure the algorithm performance by comparing its prediction against a known user rating of an item (Herlocker et al. 2004; Gunawardana and Shani 2009). However, such accuracy-centric evaluations cannot answer the question about if users are satisfied with the recommendations. For example, Amazon claimed to generate an additional 10% to 30% of its revenue in 2015 from the sale of diverse (non-personalized) items (Srihari 2015). This kind of insufficiency has shifted some researchers' focus to different goals for a recommender system, which can address other aspects beyond accuracy. Generally, recommending everything related to users' preferences would result in good accuracy. However, for news consumption, though

accuracy is important, other factors are equally crucial to satisfy users' needs. Below we discuss the beyond-accuracy aspects in NRS.

3.3.1.1 Diversity Diversity measures the degree of 'dissimilarity' among the recommended items. It is mostly implemented through re-ranking of the recommendation lists. Some well-known metrics are: Intra-List Similarity (ILS) (similarity between any two lists of recommended items); temporal or Lathia's diversity (in the sequence of recommendation lists over time); normalized diversity; and other measures as discussed by Kunaver and Porl (Kunaver and Porl 2017). The traditional pairwise diversity ILS remains a popular metric to evaluate diversity in NRS (Li and Li 2013; Gu et al. 2014; Maksai et al. 2015; Raza and Ding 2020). The ILS can be computed among the items, topics, categories, tags or even sentiments (tone) (Helberger 2019) in an NRS. Since the typical ILS method is computed for each individual user, it is a computationally expensive process for an NRS where there are millions of users and items. Thus, it requires more research to consider various aspects, such as level of diversification, scalability issues in an NRS.

3.3.1.2 Coverage Coverage represents the percentage of distinct items/users/ratings that a recommender system can recommend. Popular interpretations of coverage include item coverage (percentage of items), user coverage (percentage of users), catalog coverage (percentage of recommended user-item pairs) and interaction coverage (rating predictions) over potential items, users, user-item pairs, or ratings respectively (Han and Yamana 2017). Coverage in an NRS is treated no differently than that in other recommendation domains. It is mostly used to determine the coverage of items in the news domain (De Francisci Morales et al. 2012; Maksai et al. 2015). In some cases, coverage is defined as the measure of number of users' visits on the website during different times to determine the topical coverage [11]. The research on coverage for NRS is still very limited and discusses item coverage mostly. It is important to have more research on coverage since this aspect is related not only to the recommended items but also to the whole NRS.

3.3.1.3 Novelty Novelty determines how different or unknown a recommendation is to what has been previously recommended to a user (Vargas and Castells 2011). Silveira et al. (2019) defined novelty at three levels: a user has never heard of the item in his life (life-level), item is unknown to the user as per his consumption history (system level), and finally the non-redundant item in the recommendation list (recommendation level). Introducing novelty is more challenging in the NRS because almost everything that is happening inside the news domain is novel. In its simplest form, novelty is defined as the inverse of popularity or the ratio of unknown items in the top-N recommended list of news items (Garcin and Faltings 2013; Gu et al. 2014; Maksai et al. 2015; Saranya and Sudha Sadasivam 2017; Raza and Ding 2020). So far, the inclusion of novelty in NRS is limited to the item level only. Novelty should also be covered in terms of overall content, events and uniqueness of news stories to the users.

3.3.1.4 Serendipity Serendipity is a composite concept that includes various aspects such as the degree of relevance (usefulness), novelty (new) and unexpectedness (surprise) (Kotkov et al. 2016). Serendipity is different from novelty. An item is novel if a user is not familiar with or has not consumed or forgotten about the item, whereas an item is serendipitous if the user does not expect or would not have discovered this item but found it fortunate and interesting to have it recommended to him. For example, if the user is recommended

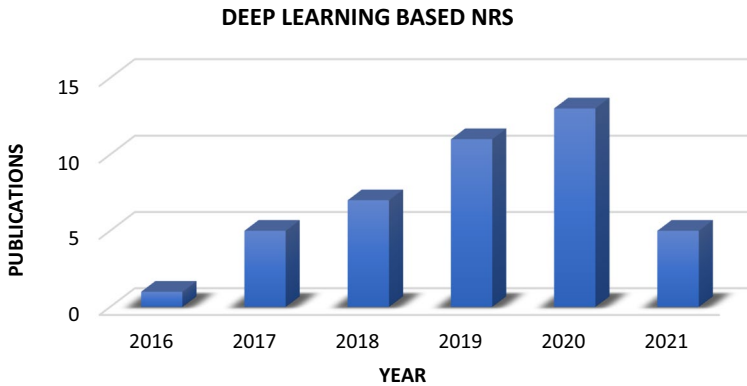


Fig. 4 Deep learning algorithms used in NRS from 2016 till early year 2021

a news story that he has never heard of, this news story is novel to him but not serendipitous if he is not interested in that topic. On the contrary, if the user finds this news story interesting enough to change his attitude on that news category or topic, this news item is a serendipitous item (Asikin and Wörndl 2014). In one NRS (Maksai et al. 2015), the serendipity aspect is defined as being composed of accuracy, novelty and diversity. In a few other NRS (Jenders et al. 2015; Cucchiarelli et al. 2018), serendipity is defined in terms of news topics that are both semantically related and are unexpected. The literature shows only limited research on the serendipity in the NRS. One reason for this could be that serendipity is a composite aspect with many combinatory definitions, which makes it difficult for the researchers to evaluate.

3.3.2 Subjective measures through user study on user satisfaction

User experience is a subjective term, with different meanings and interpretations. It is affected by many factors during different stages of recommendations, i.e., before, during and after the recommendations are made. For example, recommending something trending or related to the user's context (e.g., demographics) during the sign-up process increases the user's loyalty to the system. Similarly, proactively recommending some news stories at the side pane during the normal reading process may persuade users to stay in the system longer. If a recommender system can include these features, it may increase the user's trust with the system.

In recommender systems, the user experience is usually evaluated through three prominent ways: (i) by carrying out user studies where the subjects are given certain questionnaires during different stages of recommendations (Konstan and Riedl 2012), (ii) by combining study on longitudinally logged data with the questionnaire-based user study (Nguyen et al. 2014), and (iii) by addressing other evaluation measures such as combining accuracy and beyond-accuracy measures in certain ways (Maksai et al. 2015).

The user experience framework by Knijnenburg et al. (2012) consists of six components: objective system aspects (algorithm, presentation, interface and additional features of a recommender system); user experience (choices, evaluations of the system by the user); perception or subjective system aspects (user's evaluation of the objective

Table 2 Evaluation Metrics accuracy (acc): beyond-accuracy (beyond-acc)

Metric	Description	Type
Accuracy	Number of correct predictions over total predictions	acc
Precision (prec)	Proportion of relevant items over total recommendations	acc
Recall (rec)	Proportion of relevant items over total relevant items	acc
F1-score (F1)	Weighted average of the precision and recall	acc
Mean Reciprocal Rank (tMRR)	Multiplicative inverse of rank of the first correct item	ranking, acc
Mean Average Precision (MAP)	The average precisions across all relevant queries	ranking, acc
Rank	Percentile-ranking within the ordered list items	ranking, acc
Cumulative Rating	Total relevance of all documents above a rank position in top recommended items	ranking, acc
Success	Current item that is in a set of recommended items	ranking, acc
Novelty (nov)	Ratio of unseen items over recommended items	beyond-acc
Serendipity (seren)	Measure of unexpectedness with relevance (relevance is 1 if recommended item is interacted with otherwise 0)	beyond-acc
Coverage (cov)	Percent of items that the model is able to recommend	beyond-acc
Diversity (div)	Degree of dissimilar recommendations either at system level (aggregate diversity) or for each user (individual diversity)	beyond-acc
Hit Rate (HR)	Ratio of hits in ranked items over the number of users	acc
Log-loss	Probability of a prediction input between 0 and 1	acc
Average Reciprocal Hit Rate (ARHR)	A hit is inversely relative to its position in top recommendations	acc
Root-mean-square error (RMSE)	Difference between predicted and actual rating	acc
Click-through rate (CTR)	The likelihood of a news item that will be clicked	acc
Discounted Cumulative Gain (DCG)	Gain of an item as per to its position in recommendation list	acc
Area under curve (AUC)	ROC curve plots recall against fallout (false positive rate)	acc
Customer Satisfaction Index (CSI)	The satisfaction degree of user on the recommendations	beyond-acc
Personalized (NRS-specific)	Current item that is in a set of recommended list but is not among the popular items (Garcin et al. 2013)	relevancy, acc
Saliency (NRS-specific)	A function of entities' frequency in news articles, with a decay factor (Cucchiarelli et al. 2018)	beyond-acc
Future-Impact (NRS-specific)	Evaluate a tradeoff between recency and relevancy (Chakraborty et al. 2019)	beyond-acc
Tradeoff (NRS-specific)	Evaluate a tradeoff between high accuracy and reasonable diversity (Raza and Ding 2020)	beyond-acc
Senti (NRS-specific)	Evaluate sentiment diversity of recommendations, motivated by MRR and hit ratio (Wu et al. 2020a)	beyond-acc

aspects); situational (different contexts such as social, trust, choice goal) and personal characteristics (gender, location) as external features; and objective interaction (observable behavior such as browsing, viewing, signing-in, rating, consuming).

A few small-scaled frameworks for user experience are also proposed specifically for the NRS. The framework in one NRS (Asikin and Wörndl 2014) considers only three factors (i.e., appropriate, like, surprising) to evaluate user behaviors. The other framework (Constantinides and Dowell 2018) considers six factors, i.e., reading frequency, reading time, time of day, reading style, browsing strategy and location (context) to evaluate the user experience. However, these are all implicit factors and can only be used as indicators of user experience. They are not the direct measures for user experience.

The researchers in prior work in NRS have associated user satisfaction with objective measures. They assumed that user experience is a global phenomenon for all users, so they use one measure for all. For some researchers, it can be measured through accuracy (Nguyen et al. 2014; Viana and Soares 2016; Su et al. 2016). They demonstrated that higher ratings provide more pleasing and satisfactory experience to the users. For some other researchers, the user experience is more related to beyond-accuracy aspects. For example, a few authors (Asikin and Wörndl 2014; Jenders et al. 2015) claim that increasing serendipity in the NRS would yield higher user satisfaction. Some associate user experience with a higher degree of novelty (Saranya and Sudha Sadasivam 2017).

We show a distribution of accuracy and beyond-accuracy metrics used in NRS papers in Fig. 5.

The statistics in Fig. 5 show that accuracy is the most widely used evaluation measure in the NRS. The researchers also put some efforts to introduce diversity in news recommendations. There is very limited work in novelty, coverage and the most important aspect, i.e., user experience in NRS. In general, the quality metrics used in the NRS research are more or less same as those used in general recommender systems. However, a few evaluation metrics are designed specifically for the NRS, which are discussed next.

3.3.3 Evaluation metrics specific to news recommender systems

3.3.3.1 Personalized Garcin et al. (2013) propose a personalized @ k metric, which removes k popular items from the recommendation list to produce a smaller set of recommendations. The goal is to eliminate the popularity bias that occurs when data is collected from websites that automatically recommend the most popular items.

3.3.3.2 Saliency Cucchiarelli et al. (2018) propose a saliency metric. The saliency of entities (named entities) is calculated as a function of their frequency in news articles, with a decay factor based on the distance of the positional index of the first occurrence in the text. The idea of this metric is inspired by the news-specific discourse structure, which tends to provide brief summaries of the most important facts and entities in the first paragraphs.

3.3.3.3 Future-impact Chakraborty et al. (2019) propose a future-impact metric that trade-offs between recency (age of a news story after it is published) and importance (relevance). A news story with a higher future-impact score is thought to be a high-impact story, and vice versa. Usually, the news stories that are recently published are given the highest future-impact scores.

3.3.3.4 Tradeoff Raza and Ding (2020) propose a tradeoff metric for balancing high accuracy (precision, recall measures) with reasonable diversity (diversity and novelty aspects). The assumption is that higher accuracy leads to better personalization and thus improves readers' experiences with the NRS. Reasonable diversity, on the other hand, helps readers get diversified news so that they don't get bored reading the same news stories over and over. This metric is designed to keep readers engaged in the reading process while also recommending diverse news to them.

3.3.3.5 Senti Wu et al. (2020a) propose a metric Senti (from word sentiment) to evaluate the sentiment diversity of news recommendations. This metric normalizes MRR and hit ratio scores. The Senti is positive if the top-ranked news has the same sentiment orientation as the overall sentiment, and it is higher if the sentiments are stronger.

We have also considered three standard evaluation protocols while classifying the literature for evaluation measures in Fig. 6. These evaluation protocols refer to the experimental settings in which we measure the quality of recommendations and include offline experimentation/simulation, online experimentation (A/B or real-time tests) and user studies (Gunawardana and Shani 2009).

As can be seen in Fig. 6, there are 13 papers using online evaluation, 58 using offline evaluation, and 9 using the user study. The offline evaluation protocol is the most widely used in the NRS research. One reason for this could be that online evaluation and the user study are often considered as an expensive approach in real-time settings in an NRS.

3.4 Research datasets

Since the objects to recommend in the news domain are mostly text documents, the news datasets mainly consist of textual data. There are different types of datasets that we can consider: (i) publicly available datasets for non-commercial and research purposes, (ii) proprietary datasets, (iii) crawled datasets, or (iv) synthetic datasets created with simulated (anonymized or hidden or added) values. The details of a few datasets such as Plista, Adressa, Yahoo, Outbrain, and a few open source frameworks have been given by Karimi et al. (2018), so here we just give a brief overview of them. However, we also discuss some datasets that are either new or less discussed, such as Yahoo news, Hacker news, BuzzFeed and some fake news datasets.

3.4.1 Plista

Plista is a dataset developed by Plista (an advertising company) and Technische Universität Berlin to promote research in NRS (Kille et al. 2013). It consists of logs from 13 German news portals collected from June 2013. It also contains millions of impressions (articles views) and some time-related information. This dataset is accessible upon request for research purposes.

3.4.2 Adressa

Adressa (Gulla et al. 2017) is a publicly available benchmark dataset developed by Adressavisen (a local newspaper in Norway) and Norwegian University of Science and Technology (NTNU). Like Plista, Adressa does not have explicit ratings, but different from Plista, it includes reading time in addition to reading counts.

3.4.3 Yahoo webscope

Yahoo Webscope⁴ is a reference library that provides datasets for non-commercial users such as academics and scientists. Yahoo provides benchmark datasets for news as well. These datasets are: R6A—*Yahoo! Front Page Today Module User Click Log Dataset*, R6B—*Yahoo! Front Page Today Module User Click Log Dataset*, R11—*Yahoo News Video dataset*, L33—*Yahoo News Ranked Multi-label Corpus* and L32—*The Yahoo News Annotated Comments Corpus*. Among these datasets, the two news datasets (R6A and R6B) with ratings and news category information provided by Yahoo! Front Page Today are of importance for researchers to evaluate their recommendation algorithms. These two datasets consist of the timestamp information and explicit ratings, which makes them a favorite option for developing and evaluating CF solutions. However, one limitation of these datasets is that news items are represented by their features where actual content of the news stories are anonymized without any additional information. It might be difficult to make recommendations in the absence of any information on the stories. These datasets are also available upon request for research purposes.

3.4.4 Hacker news

Hacker News,⁵ run by YCombinator,⁶ is a popular social news website. It is widely known among the people in the IT industry where they can share news, demonstrate their projects, ask questions, post jobs and comment on news stories as a community. Hacker News provides a big dataset under the MIT License since its launch in 2006. This dataset is also available as a public dataset through Google BigQuery⁷ (a RESTFUL web service providing exploratory analysis of massive datasets in conjunction with Google storage). This dataset consists of news stories from various sources, which may be useful for the researchers working on news recommendations. However, the texts and the comments on news do not go through the censor process and may include profanity. Hacker News does not take responsibility for what the authors have written.

3.4.5 BuzzFeed news

BuzzFeed⁸ is a company that provides news and entertainment content on digital media. They publish data related to fake news, social media and various news patterns. They have released some datasets and made them available on GitHub.⁹ These datasets are useful for researchers working on fake news investigating rumors, misinformation and detecting factual claims. However, one limitation is that these datasets are particularly for fake news detection and may not be a proper source for building a personalized NRS.

⁴ <https://webscope.sandbox.yahoo.com/>.

⁵ <https://github.com/HackerNews/API>.

⁶ <https://news.ycombinator.com/>.

⁷ <https://cloud.google.com/bigquery/what-is-bigquery>.

⁸ <https://www.buzzfeed.com/>.

⁹ <https://github.com/BuzzFeedNews/everything>.

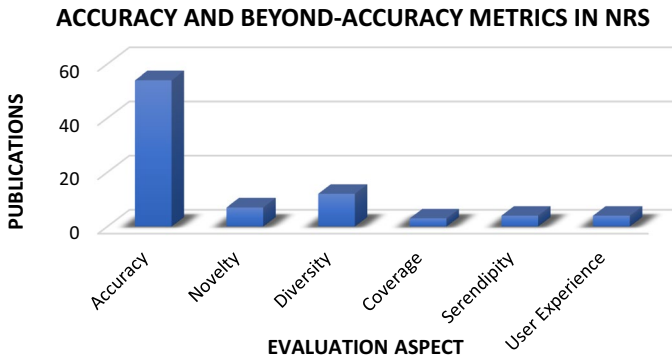


Fig. 5 Distribution of accuracy and beyond-accuracy aspects in NRS

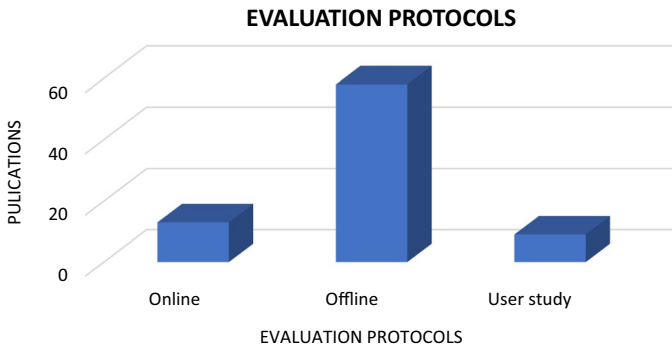


Fig. 6 Distribution of evaluation approaches used in NRS

3.4.6 Microsoft news dataset (MIND)

The MIND dataset is a large-scale benchmark dataset (Wu et al. 2020b) for news recommendation research. MIND contains about 160 k English news articles, and more than 15 million impression logs generated by 1 million users. Every news article is identified by rich textual content including title, abstract, body, category, and entities. The impression log contains the click events, non-clicked events and historical news click behaviors of a user. MIND-small is a small version of the original MIND dataset, which consists of 50,000 users and their behavior logs. The users are anonymized. Both versions of the dataset can be accessed online.¹⁰

3.4.7 Fake news datasets

Fake news has become a serious problem for spreading rumors and misinformation, and consequently made negative impact on politics, regional stableness, and sometimes even

¹⁰ <https://msnews.github.io/>.

people's daily life, especially during the US election and the pandemic period. Because of this, many fake news datasets are made accessible for open research in recent years. Though they are not directly related to the NRS research, they are useful for fake news detection, which could be a crucial step before making recommendations. A few prominent ones are listed here: BS Detector,¹¹ Credbank-data,¹² BuzzFace,¹³ MisInfoText,¹⁴ NewsTrust,¹⁵ SFU Opinion and Comments Corpus,¹⁶ NELA-GT-2018 (Nørregaard et al. 2019), NELA-GT-2019 (Horne 2020), NELA-GT-2020 (Gruppi et al. 2021), Fakeddit (Nakamura et al. 2019), FakeNewsNet (Shu et al. 2018), NYtimes covid-19-data¹⁷ and LIAR (Wang 2017) datasets.

3.4.8 Other datasets

There are some classical news datasets such as Reuters Corpora¹⁸ and 20 Newsgroups¹⁹ that are used for news categorization. Some of the recent ones include Amazon news datasets at Fast.ai²⁰ and Global Database of Events, Language and Tone (GDELT)²¹ that can be used for text categorization and detailed analyses on news and user data. Some news-related datasets²² have recently been made public by the Huggingface library (a library for Transformer models).

In Fig. 7, we show the distribution of datasets used in previous NRS research.

As can be seen in Fig. 7, 62 papers are using private (mostly crawled) datasets, 16 papers are using public datasets.

Most of the time, the researchers prefer to build their own news recommendation datasets for two important reasons: lack of publicly available datasets and unique requirements on certain types of data for their research. In that, they crawl news from different news publishers. These datasets are usually proprietary to the organization who created them. There are also synthetic datasets that are domain dependent and are created by taking data from some benchmark datasets and enriching them by including related information and interactions either artificially or in a semi-autonomous way.

We have also included the information in Tables 3 and 4.

3.5 Open news recommendation platforms

Over the last few years, there are several libraries that have been developed for recommendations. A few prominent ones are discussed briefly here.

¹¹ <https://github.com/thiagovas/bs-detector-dataset>.

¹² <http://compsocial.github.io/CREDBANK-data/>.

¹³ <https://github.com/gstantia/BuzzFace>.

¹⁴ <https://github.com/sfu-discourse-lab/MisInfoText>.

¹⁵ <http://resources.mpi-inf.mpg.de/impact/credibilityanalysis/data.tar.gz>.

¹⁶ <https://github.com/sfu-discourse-lab/SOCC>.

¹⁷ <https://github.com/nytimes/covid-19-data>.

¹⁸ <https://trec.nist.gov/data/reuters/reuters.html>.

¹⁹ <http://qwone.com/~jason/20Newsgroups/>.

²⁰ <https://course.fast.ai/datasets>.

²¹ <https://www.gdeltproject.org/data.html>.

²² <https://huggingface.co/datasets?search=news>.

MIND (Wu et al. 2020b) is a recent news benchmark dataset. The contributors of the dataset provided an environment in the form of a competitive event and Leaderboard²³ for researchers to work on the news recommendation problem. In conjunction with The Web Conference 2021,²⁴ the contributors also offered an International Workshop on News Recommendation and Intelligence. This incentive requested research and technical report articles on many aspects of news recommendations.

Apache Mahout is a distributed machine learning library implemented in Java and contains some CF algorithms. This framework is available both for academic and commercial use to work with real-world news data (Beck et al. 2017).

Idomaar (Scriminaci et al. 2016) is a benchmark framework that enables efficient reproducible evaluation of recommendation algorithms in real-world settings. Unlike other frameworks implemented in Java, Python or C++, it is implemented as web service, which offers flexibility in the programming languages.

StreamingRec (Jugovac et al. 2018) is written in Java and offers a variety of pre-built news recommendation algorithms for implementation and comparative evaluations. It simulates real-world news recommendation scenarios.

CLEF NEWSREEL and Open Recommendation Platform (ORP) CLEF NEWSREEL platform was designed to encourage researchers to develop novel recommenders using the Plista dataset and evaluate them in real time through ORP. ORP consists of distributed systems where recommendation providers and consumers interact over a standardized protocol to deliver recommendations. The researchers used CLEF NEWSREEL for online evaluation as well as the replay-based (simulation or offline) evaluation (Domann and Lommatzsch 2017; Kumar et al. 2017). It also includes the Idomaar framework, the Plista dataset and offers a few online algorithms and data analysis techniques.

Among these frameworks, Idomaar and Apache Mahout frameworks were developed for general recommender systems, whereas CLEF NEWSREEL, streamingRec and MIND were designed specifically for NRS. CLEF NEWSREEL is obsolete now. The MIND platform is still active.

4 Major challenges in news recommender systems and conventional solutions

In this section, we discuss the major challenges of the NRS and their solutions. A few challenges such as cold-start, data sparsity have been reviewed in the previous survey (Karimi et al. 2018). They are common to the general recommenders too. So, we decide to skip them in this survey. We include two challenges (timeliness and user modeling) that have been discussed before, but we try to provide some new insights and perspectives in our discussion. We also identify the news content quality as an emerging challenge, which is not discussed before.

Here, we provide a categorization of the conventional solutions from the state-of-the-art to address these major challenges. We use the term “conventional” to refer to the non-neural solutions, and we leave the discussion of the deep-learning-based solutions to Sect. 5.

²³ <https://msnews.github.io/>.

²⁴ <https://www.2021.thewebconf.org/program/workshops/>.

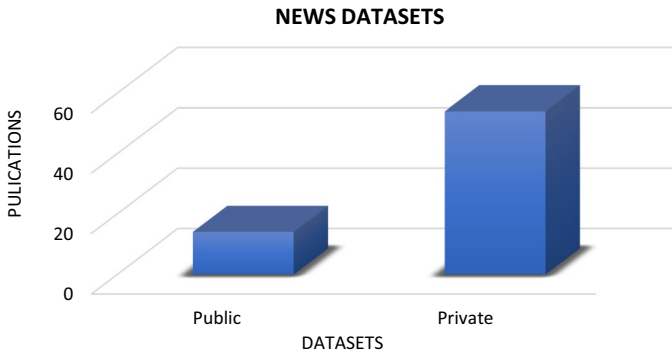


Fig. 7 Distribution of datasets in NRS

4.1 Challenge 1: timeliness

The earlier an event is reported, the more newsworthy it becomes. According to the working notes of CLEF NEWSREEL challenge (Brodt and Hopfgartner 2014), a well-formed recommendation must respond to a request within a given time frame (100 ms). It requires faster, real-time processing and much more computations to make recommendations for a large number of news articles that are found in the news domain. Popularity, recency, freshness, trends, uniqueness, and low latency are the characteristics that should be factored into an NRS to give timely suggestions.

Solutions Several conventional techniques used in general recommenders have been applied to address the challenge of timeliness in NRS. These models are discussed below.

4.1.1 Time-decay models

Recommendation algorithms designed to give more weight to recent items with sensitivity to time are called time-decay models (Ding and Li 2005; Xia et al. 2010). The term ‘time-decay’ refers to the decline in terms of value of data over time. To be able to accommodate the time-decay effect of news items, it is important to build an effective short-term preference model that can predict the recent news items to readers.

A naïve and popular time-decay model is to use sliding/timing windows. A timing window in a time-decay model considers only recent news items or rating data, with older data being discarded or weighted less (De Francisci Morales et al. 2012). In the literature, there are various reports about the sizes and weights of timing windows. Some authors (Fortuna et al. 2015; Okura et al. 2017; Sottocornola et al. 2018) state that the timing windows should not be of fixed size (large or small) and should be adaptive. In general, a larger timing window leads to concept drifts (target variables change their values with time) (Muralidhar et al. 2015; Sottocornola et al. 2018) and a smaller one would not have sufficient data to build a short-term preference model (Sottocornola et al. 2018).

4.1.2 Graph-based solutions

The second group of algorithms is graph-based that models the sequential reading process in an NRS. Graph-based recommendation models represent the relationship between users

Table 3 Algorithms (alg.), challenge, solution, dataset, evaluation (eval) metric: accuracy (acc), beyond-accuracy (beyond-acc), and evaluation protocol (protocol) for NRS papers

Paper	Alg	Challenge	Solution	Dataset	Eval. Metric	Protocol
Park et al. (2009)	CBF	Content quality	Bias detection	Private—Google news	F1-score (acc)	Offline, User Study
Xia et al. (2010)	CF	Timeliness	Time-decay	Private—Alibaba.com	CSI (beyond-acc), prec (acc)	Offline
De Francisci Morales et al. (2012)	CBF	Timeliness User modeling	Time-decay Microblog	Private—Twitter, Yahoo	MRR (acc), cov (beyond-acc)	Offline
Agarwal et al. (2013)	CBF	User modeling	Ontology	Private—RSS news feed	prec, rec (acc)	Offline
Prawesh and Padmanabhan (2012)	CBF	Timeliness	Popularity	—	—	—
Boutet et al. (2013)	CF	User modeling Content quality	Feature-based Bias detection	Private—Arxiv,Digg	prec, rec (acc)	Offline, User Study
Ilievski and Roy (2013)	CBF	User modeling	Knowledge-based	Private—German news papers	Rank(acc)	Offline
Li and Li (2013)	CBF, CF	User modeling	Feature-based	Private—not mentioned	prec, rec, F1-score, NDCG (acc), div (beyond-acc)	Offline
Jonnalagedda and Gauch (2013)	CBF, CF	Timeliness	Popularity-based	Private—Twitter, CNN and BBC	Accuracy (acc)	Offline, User Study
Garcin et al. (2013)	CBF	Timeliness	Graph-based	Private—Tribune de Geneve and 24heures.ch	Personalized (acc), nov (beyond-acc)	Offline
Gu et al. (2014)	CBF	User modeling	Microblog	Private—news.sina.com	F1, prec, rec (acc), div, nov (beyond-acc)	Offline
Asikin and Wörrdl (2014)	CBF	User modeling	Stereotypical	Private	seren (beyond-acc)	User study
Trevisiol et al. (2014)	CBF	Timeliness	Graph-based	Private—Facebook, Twitter, Reddit	prec, MRR (acc)	Offline
Oh et al. (2014)	CBF	User modeling	Feature-based	Private—Korean news	HR (acc)	Offline
Muralidhar et al. (2015)	CF, CBF	Timeliness User modeling	Time-decay Feature-based	Private—Washington Post	HR (acc)	Offline
Xiao et al. (2015)	CF	Timeliness, User modeling	Sequential Collaborative	Private Chinese news	prec, rec, F1 (acc)	Offline

Table 3 (continued)

Paper	Alg	Challenge	Solution	Dataset	Eval. Metric	Protocol
Maksai et al. (2015)	CBF	Timeliness	Popularity	Private—Swissinfo, Yahoo FrontPage, lePoint.fr	div, nov, seren, cov (beyond-acc), RMSE, CTR (acc)	Online, Offline
Jenders et al. (2015)	CBF	Content quality	Feature-based	Private—New York Times	seren (beyond-acc)	User Study
Garrido et al. (2015)	CBF	User modeling	Stereotypical	Private—heraldo.es	–	Offline
Lu et al. (2015)	CBF, CF	User modeling	Feature-based, Collaborative	Private—Bing Now news	Accuracy	Offline
Doychev et al. (2015)	CBF, CF	Timeliness	Popularity	Plista	prec, CTR (acc)	Online, Offline
Ma et al. (2016)	CBF, CF	User modeling	Stereotypical, Collaborative	Private—Bing Now news	MAP, MRR, CTR (acc)	Offline
Jonnalagedda et al. (2016)	CBF	User modeling	Microblog	Private—Twitter	NDCG (acc)	Offline
Chakraborty et al. (2017)	CBF	User modeling	Clickbait removal	Private—BuzzFeed & Wikinews	prec, rec, F1 (acc)	Offline, Online
Viana and Soares (2016)	CBF, CF	User modeling	Feature-based	–	–	User Study, Simulation
Rizos et al. (2016)	CBF	User Modeling	Feature-based, Microblog	Private—Reddit	Accuracy	Offline
Okura et al. (2016)	CBF	Content quality	Duplicate detection	Private—Yahoo Front-Page	CTR (acc)	Online
Guan et al. (2017)	CBF, CF	Timeliness	Deep neural network	Private—news.sohu.com	prec, rec, F1 (acc)	Offline
Robindro et al. (2017)	CBF	User modeling	Stereotypical	BBC news	NDCG (acc)	Offline
Khattar et al. (2017)	CF	Timeliness	Session-based, Knowledge-based	Private—Veooz.com news	MAP, Hit rate, NDCG (acc)	Offline
Okura et al. (2017)	CBF	Content quality	Duplicate detection, Time decay	Private—Yahoo Front-Page	AUC, MRR, NDCG, CTR (Online) (acc)	Online, Offline
Kumar et al. (2017)	CBF	Content quality	Knowledge-based	Plista	HR, NDCG (acc)	Online, Offline
Cucchiarelli et al. (2018)	CBF	User modeling	Microblogging	Private—Twitter, Wikipedia	Saliency, seren (beyond-acc), MAP (acc)	Online, User Study

Table 3 (continued)

Paper	Alg	Challenge	Solution	Dataset	Eval. Metric	Protocol
Constantinides and Dowell (2018)	CBF, CF	User modeling	Stereotyping	Private—Habito.com News	Accuracy	Offline, User Study
Jugovac et al. (2018)	CF	User modeling Timeliness	Feature-based Session-based	Outbrain.com, Plista	MRR, F1-Score (acc)	Online, Offline
de Souza Pereira Moreira (2018)	CBF CF	Timeliness	Session-based	Adressa, Globo.com	Rec, NDCG (acc)	Offline
Sottocornola et al. (2018)	CBF CF	Timeliness	Time-decay, Session-based	Private—not mentioned	prec (acc)	Offline
Lin et al. (2012)	CBF	User Modeling	PMF	Private—not mentioned	Accuracy	Offline
Yan et al. (2012)	CBF	User Modeling	NMF	Tweet data. News data crawled from sogou.com	Accuracy	Offline
Xia et al. (2014)	CF	Content Quality	BPR	Private—not mentioned	Accuracy	Offline
Wang et al. (2015)	CBF	User Modeling	TF	Not mentioned	—	—
Shu et al. (2019)	CBF	User Modeling	NMF	BuzzFeed, PolitiFact	Accuracy	Offline
Gharahighehi and Vens (2019)	CBF	User Modeling	BPR	Private- not mentioned	Accuracy	Offline
Raza and Ding (2019)	CBF	User Modeling, Timeliness	MF (baseline predictors)	NYTimes—crawled	prec,rec (acc)	Offline
Raza and Ding (2020a)	CBF	User Modeling	GLM with CF	NYTimes—crawled	prec, rec (acc), div, nov (beyond-acc)	Offline

Table 4 Algorithms (alg.), DL mechanism, dataset, evaluation (eval) metric: accuracy (acc), beyond-accuracy (beyond-acc), and evaluation protocol (protocol) for NRS papers

Paper	Alg	DL mechanism	Dataset	Eval. Metric	Protocol
Song et al. (2016)	CBF	RNNs (LSTM) for timeliness in user modeling for short-term preferences. Sem2Vec (BoW) for news representation	Private—crawled, not mentioned	prec, rec, F1, AUC, MAP, MRR (acc)	Offline
Okura et al. (2017)	CBF	Denoising AE for news representations and RNNs (LSTM) for timeliness in user modeling for short-term preferences	Private -Yahoo Japan news	AUC, MRR, DCG, CTR (acc)	Offline
Wang et al. (2017)	CBF + CF	CNNs for news representations; Attention for user modeling	Private—crawled	AUC, prec, rec, F1 (acc)	Offline, online
Kumar et al. (2017)	CF	Doc2Vec for news representation, RNNs (Bidirectional LSTM) for timeliness in user modeling for short-term preferences	Private—crawled	HR, NDCG (acc)	Offline
Lian et al. (2018)	CBF	CNNs for news representations, Attention for user modeling	Private—Bing	AUC, Log loss (acc)	Offline
Wang et al. (2018a)	CBF	Knowledge graph with CNNs for news representations, Attention for user modeling	Private—Bing news logs	CTR (acc)	Offline
Zheng et al. (2018)	CBF	RL, Deep Q-Network for news and user representations	Private—not given	CTR, prec, NDCG, (acc), div (beyond-acc)	Offline
Yu et al. (2018)	CBF	CNNs for news text, AE for images representation, MLP to recommend	Private—frnews.ifeng.com.ca	Accuracy	Offline
de Souza Pereira Moreira (2018)	CBF, CF	Word2Vec + CNN for news representations, RNNs (GRUs) for timeliness in user modeling for short-term preferences	Globo.com, Plista, Addressa	HR, MRR, Accuracy (acc); cov, nov, div (beyond-acc)	Online and Offline

Table 4 (continued)

Paper	Alg	DL mechanism	Dataset	Eval. Metric	Protocol
Zhang et al. (2019)	CBF	CNNs for news representations, RNNs (GRUs) for timeliness in user modeling for short-term preferences	Addressa	prec, rec, AUC (acc)	Offline
Zhu et al. (2019)	CBF	CNNs for news representations, RNNs (GRUs) for timeliness in user modeling for short-term preferences	Addressa	prec, rec, AUC (acc)	Offline
Cao et al. (2017)	CBF	Stacked AE	MovieLens (not news)	prec, rec, AUC (acc)	Offline
An et al. (2019)	CBF, CF	CNNs for news representations, RNNs (GRUs) for timeliness in user modeling for short-term preferences	MIND	AUC, NDCG, MRR (acc)	Offline
Wu et al. (2019a)	CBF	CNNs for news representations, Attention for user modeling	MIND	AUC, NDCG, MRR (acc)	Offline
Wu et al. (2019b)	CBF, CF	CNNs for news representations, Attention for user modeling	MIND	AUC, NDCG, MRR (acc)	Offline
Wu et al. (2019c), Ge et al. (2020)	CBF, CF	Transformer for news representations, GNNs for user modeling	MIND	AUC, NDCG, MRR (acc)	Offline
Wu et al. (2020a)	CBF, CF	Attention for news and user representations, MSE loss in news module for diversity	MIND	AUC, NDCG, MRR (acc), Senti (beyond-acc)	Offline
Lee et al. (2020)	CBF, CF	GNNs for news and user representations	–	–	–
Yang et al. (2020)	CBF, CF	Knowledge Graphs + CNNs for news; Attention for user modeling	Private-Wikidata, Weibo	prec, rec, F1, AUC (acc)	Offline
Wu et al. (2021)	CBF, CF	BERT for news representations; Attention for user representation	MIND	AUC, NDCG, MRR (acc), Senti(beyond-acc)	Offline

and items using links (weighted or unweighted). These models are also used to predict the next-news items by modeling the sequential dependencies over user-item interactions.

Some representative models include: (i) Context Trees that provide news recommendations to anonymous readers based on their news browsing patterns (Garcin et al. 2013; Maksai et al. 2015), (ii) Browse-Graphs to model sequential patterns from the readers' consumption histories (Trevisiol et al. 2014), and (iii) Markov decision process that models the sequential reading process in an NRS (Khattar et al. 2017). These traditional models are intuitive solutions to model the sequential dependencies among user-item interactions. However, due to an increasing number of states, these models may fail to capture the complex patterns from a large amount of data, as in the news domain.

4.1.3 Popularity-based solutions

The third group of models in the NRS is the popularity-based models. They are based on the popularity of news items in terms of clickthrough rate, or social ties on social network sites. A traditional method of including popularity in an NRS is to simply count the total number of visits on news articles (Doychev et al. 2015). However, calculating popularity based on top-N articles is prone to amplification (popularity bias or temporal bias), which is caused by exclusively selecting top-N articles while overlooking the good (N + 1)th candidate article (s). In this case, some good articles are unfairly penalized during the hard cut-off, despite the fact that the differences between these articles and the top-N recommendations are negligible. This issue can be mitigated if recommendations are generated probabilistically with feedback loops in which an article's likelihood of being chosen is proportional to its current popularity (count) (Prawesh and Padmanabhan 2012). News stories can also be ranked according to their popularity in the popular micro-blogging sites like Twitter (Jonnalagedda et al. 2016). In some NRS, the trends are also used to determine the popularity of news items (Chakraborty et al. 2019).

Although popularity-based models are easy to implement, it does not ensure that all popular news is credible and truly popular. According to a report by nbcnews,²⁵ false news stories are more popular, and they are 70% more likely to be retweeted than true stories.

Overall, traditional timeliness models may be limited in their ability to address dynamic user behaviors in an NRS.

4.2 Challenge 2: user modeling

Typically, users' preferences are modeled in two ways: explicit feedbacks and implicit feedbacks (Knijnenburg et al. 2012). Explicit feedback data is quantifiable, e.g., the rating of movies by users on Netflix or products on Amazon or news items on Flipboard. Often in an NRS, it happens that a user may read the whole news article but does not explicitly specify the rating. In this case, we consider implicit feedbacks that act as a proxy for a user's interest. Examples of implicit feedback data include clicks on links, browsing history, reading time spent and percentage (5%, 50% or 75%) of scrolling a news story.

²⁵ <https://www.nbcnews.com/health/health-news/fake-news-lies-spread-faster-social-media-truth-does-n854896>.

In an NRS, we need to consider several aspects of user modelling, such as anonymous news readers, profiling information for registered users, passive news consumption, negative implicit feedbacks, and relevance of readers' intents.

Solutions We review the pertinent literature to find out different user modeling techniques used in the NRS. These models are discussed below.

4.2.1 Stereotypical user modeling

The first approach is stereotypical user modeling. A stereotype is a collection of characteristics that frequently co-occur in people (Rich 1979). In this approach, a user is assigned to a class of users, and predictions about users' preferences are inferred from prior information about the class. When we do not have the complete background knowledge about a user, we can use this modelling technique. Well-known stereotypes in the NRS are based on geolocation (Asikin and Wörndl 2014; Garrido et al. 2015; Robindro et al. 2017) and on users' habits (Constantinides and Dowell 2018).

Though stereotyping allows users to be classified into different groups, there are two issues with stereotyping in the NRS: (i) there is no way to learn a completely new stereotype, and (ii) too much stereotyping may result in segregated user groups or filter bubbles among like-minded users.

4.2.2 Feature-based user modeling

The second approach is feature-based user modeling. A news article's content typically consists of features such as categories, headlines, sources, and topics. These features are extracted using statistical text representation methods such as bag-of-words (BoW), TFIDF, Hashing, and Word2vec. If the content of a news story is similar to the one that the user has previously read, it is recommended to the user. A general limitation of these traditional methods is that they do not consider the semantic (meaning in text) and contexts (situation in which a reader interacts with the news) while making news recommendations.

A user interest profile typically consists of long-term interests that can be captured from keywords extracted from a user's previous readings (Oh et al. 2014) or from his/her implicit feedback information (Muralidhar et al. 2015). Because users' preferences in a news domain are quite volatile and many users are anonymous, it is difficult to have complete profiling information using these statistical methods. These traditional methods are also limited to capture the time-ordered dependencies in readers' preferences.

4.2.3 Collaborative filtering

User modeling based on users' interactions, i.e., the collaborative filtering approach, to make recommendations does not require analysis of item features. These methods collect interests from similar users and store them as histories. However, if the temporal distinction in user preferences is not preserved, an NRS may be unable to effectively predict the next news article based on similar user preferences. This necessitates that an NRS incorporate

the time sequence of user behaviors into the traditional CF approach (Xiao et al. 2015; Khattar et al. 2017; Raza and Ding 2019).

4.2.4 Knowledge-based user modeling

Knowledge-based user modeling approach is often used to apply semantics (Khattar et al. 2017), ontologies (Agarwal et al. 2013) or other contexts (situation in which a user is currently in) to model users' preferences (Wang et al. 2018b). In a few NRS, OWL ontologies based on IPTC²⁶ (International Press Telecommunication Council) standards (Agarwal et al. 2013) and free knowledge bases such as Wikipedia or Microsoft Satori (Wang et al. 2018a) are used to build rich content profiles. These models allow for the reuse of domain knowledge but creating a new knowledge base may be expensive.

4.2.5 Microblogging-based user modeling

Microblogging user modelling makes use of social media platforms (such as Twitter) to model users' preferences and provide them with personalized and trending news services. There are numerous examples in the literature where users' interest profiles were inferred from microblogs (De Francisci Morales et al. 2012; Gu et al. 2014; Jonnalagedda et al. 2016). Although microblogging provides rich user interaction data, additional measures are required to assess the quality of such content (Kang et al. 2015). For example, communications and discussions in microblogs in comparison to curated news stories are usually not much trustworthy (Kang et al. 2015; Cucchiarelli et al. 2018).

In general, traditional methods for user modelling in an NRS are not very successful. In an NRS, user modelling should include not only users' histories, but also their short-term, seasonal, diversified, and sequential interests.

4.3 Challenge 3: quality control of the news content

With the majority of news media moving online, the initial difficulty for the research community was figuring out how to efficiently handle and evaluate the massive volume of unstructured information (most internet news is in textual format) in real time. Big data technology (e.g., Spark, Hadoop and cloud technology) has partially resolved the efficiency and scalability issue, while the latest development in the NLP field (e.g., the embedding-based and deep learning models) has partially resolved the feature engineering issue. The new and unsolved challenge is the quality control of the news content.

The researchers from social science usually do two types of content analyses in the news domain: quantitative and qualitative (Hamborg et al. 2019). To evaluate the quality of the news content, qualitative analysis usually requires the gold-standard test (human interpretation), which is a time-consuming task. Quantitative analysis determines the frequency of specific words or phrases in news articles, as well as other statistical features of news, such as the number of articles published on a news topic, the number of words per story, the placement of the news story on the website, and so on. In comparison to social science, quality control of the news domain is a new and understudied research topic in computer science.

²⁶ <https://iptc.org/>.

The issues with the content quality in the limited research can be summarized as: duplication, lack of semantics, spamming and biases in the news items.

Duplication Similar content appears at multiple locations (URLs) from different news sources (Doychev et al. 2015; Okura et al. 2017; Robindro et al. 2017). This can affect the ranking of news articles and is likely to bore the readers with repeated recommendations.

Lack of semantics Multiple jargons and slangs with missing semantics can often be found in news stories (Mohallick and Özgöbek 2017). They are hard to interpret using available NLP libraries.

Spamming Clickbaits (catchy news headlines) are used to trick news readers so that they click heavily on the news sites (Chakraborty et al. 2016). It is difficult to extract the hidden meanings from the clickbait that is used to manipulate readers. Even when these semantics are captured, the tactics of the used spamming techniques may later be modified.

Biases The style in which the news stories are written and the tone in which they are presented reflect the biases of the publishers, authors and the media group (Kang et al. 2015). The hyper-partisan bias (bias from publishers) is a major issue in today's news.

Solutions We reviewed the pertinent literature to find out how different authors addressed quality control issues in the NRS. These methods are discussed below.

4.3.1 Duplication detection methods

The traditional statistical methods such as TF-IDF or BoW techniques based on content features are used to recommend similar news articles to the target user (Doychev et al. 2015). But similar news articles are often repeated in the sense that they refer to the same news story presented in different ways from different publishers. Some duplication detection methods are discussed in the previous NRS research. For example, in one NRS (Okura et al. 2016), a threshold was used to filter out repetitive news articles (with similarity greater than a pre-defined maximum value). Another NRS (Robindro et al. 2017) addresses repetitive recommendations by clustering similar articles (using k-means) and then selecting a representative from each cluster. These traditional clustering-based methods are incapable of generating content embeddings for a large number of news articles and detecting duplication.

4.3.2 Semantics-based methods

To improve the quality of news recommendations, a few authors addressed the lack of semantics in NRS. For example, semantics about news stories are captured from the news structure metadata (taxonomy) in one NRS (Ilievski and Roy 2013). While this method (Ilievski and Roy 2013) focuses on higher-level semantics, we find it falls short of providing a complete representation of semantics from news bodies, titles, and so on. In another NRS (Khattar et al. 2017), the ontology is used to introduce semantic similarity among news articles. Another NRS uses the concepts and named entities from Wikipedia pages to capture the semantics of news articles (Cucchiarelli et al. 2018). Other issues that are not addressed by these methods include changing ontologies, scalability, and multilingualism.

4.3.3 Bias detection methods

These methods can detect bias in news articles. Sentiment analysis techniques have been used in a few NRS to detect sentiment bearing words from news text (Ilievski and Roy

2013; Wang and Wu 2015; Khattar et al. 2017; Cucchiarelli et al. 2018). The exploration–exploitation principle is used in another NRS to reduce bias in news articles (Boutet et al. 2013). These bias detection methods are limited, and more research is needed to detect the level of bias and to mitigate with the new tactics of introducing biases in the data.

4.3.4 Clickbait detection methods

There is limited work in the NRS to address the clickbait (catchy deceptive headlines) problem. In one NRS, the clickbait can be distinguished from the regular news headlines through a classification method (Chakraborty et al. 2016). The method is trained on a clickbait dataset gathered from a few domains that publish a large number of clickbait articles: ‘BuzzFeed’, ‘Upworthy’, ‘ViralNova’, ‘Scoopwhoop’, and ‘ViralStories’. The proposed classifier, then, identifies the clickbait headlines based on linguistic and syntactic nuances that appear more frequently in clickbait headlines.

Because the tactics of clickbait creators change over time, a typical classification model trained on a specific dataset from a specific time may suffer from data and concept drifts. As a result, to keep up with changing tactics, a classification model may need to be trained on a regular basis. Furthermore, the semantics and hidden patterns from the clickbait data should also be included in the classification models.

These papers and conventional solutions are summarized in Table 3. As can be seen in Table 3, the challenge that is addressed the most in NRS is user modeling, followed by the timeliness. The work on content quality is marginal and it needs more attention in NRS.

5 Deep learning models for news recommender systems

In this section, we cover Deep Learning (DL)-based solutions that have been widely applied in the NRS research in recent years. Many of the challenges that an NRS faces are seen to be addressed using these DL models. These methods build user models in different ways than the traditional recommendation models, and they deal with timeliness and other NRS-related issues in a more advanced way. There are certain advantages of DL that makes it a preferable approach in the NRS over some conventional solutions, which are discussed below.

The first advantage of DL is its strength when dealing with the content-based recommendation. It is inevitable for a typical CBF method to handle massive amount of data that is also multimodal (text/audio/video). For instance, when dealing with the textual data (news story, reviews, comments, tweets, etc.), images or videos, the deep neural methods like CNNs/RNNs (An et al. 2019) or language models like BERT (Devlin et al. 2018) are indispensable in the representation learning (feature learning) tasks.

The second significant advantage of DL is its ability to learn multiple interactions between the users and items. The DL-based NRS (de Souza Pereira Moreira 2018) also demonstrates sufficient performance gains over traditional CF methods (Xiao et al. 2015) in learning rich user-item interactions from the news data.

The third strength of DL is in sequential modeling. The sequential modeling task is an important approach for mining the temporal dynamics (changing user behavior over time) and session-based news recommendation tasks. Compared to this, the traditional CBF or

CF are often built on the static datasets, where there is no consideration of the temporal or sequential factors.

The fourth strength of DL methods is in dealing with the cold-start and data sparsity issues of conventional recommendation methods. The cold-start and data sparsity problem in conventional NRS is the result of insufficient rating information. The DL can extract useful features from the news and user data, which improves the estimation of user and item profiles and, as a result, improves the recommendation accuracy.

Next, we discuss the DL-based models for news recommendations.

5.1 Multi-layer perceptron (MLP)

MLP is a feed-forward neural network in which there are multiple hidden layers between the input and output layer. In a recommender system, the MLP can be used to add non-linear transformations on top of a typical MF, to learn rich user-item interactions. For example, the NCF (He et al. 2017) uses the non-linearity of MLP to learn the user-item interactions in the CF model. The MLP is also used in a few NRS (Song et al. 2016; Yu et al. 2018) to learn useful representations from data.

Overall, MLP is a simple and efficient model that is used to create neural extensions of MF based models.

5.2 Autoencoder (AE)

AE is a neural network that learns to copy its input to the output in an unsupervised manner. It has an internal (hidden) layer that describes a code to represent input, and it is made up of two primary components: an encoder to map the input into code, and a decoder to map the code to reconstruct input. In a recommender system, the AE and its variants are often used to learn hidden patterns to reconstruct users' ratings from their historical interactions (Wu et al. 2016). AE methods are also used to compress a dataset into a lower-dimensional feature subspace while preserving most of the relevant information.

Denosing auto-encoders (a type of AE) are used in an NRS to create news article representations (Okura et al. 2017). In another NRS (Cao et al. 2017), the stacked AE are used to extract the low dimensional features from a sparse rating matrix.

Overall, the AE are effective at learning useful representations from news data (news content and user feedback) in a low-dimensional space.

5.3 Convolutional neural network (CNN)

CNN is a feed-forward neural network with convolutional layers and pooling operations and have achieved great success in the field of computer vision, particularly for automated diagnosis in medicine (Göçeri 2020a, b). A CNN typically has two sets of layers: (i) convolution layers for generating local features from the data; and (ii) pooling (or sub-sampling) layers to select only representative local features (i.e., features with the highest score via activation functions) from the previous (convolution) layer. Compared to MLP networks, the CNNs have fewer parameters and perform faster (He et al. 2018).

A CNN can extract useful features from the news data by using convolution operations (also known as kernels or filters) at varying levels of granularity, thus eliminate the need for manual feature engineering (Yu et al. 2018). They are frequently used to extract local

text features from news headlines (Wang et al. 2018a; An et al. 2019; Wu et al. 2019a) or from entire news bodies (Zhu et al. 2019). The knowledge gained from these news representations is then used to make recommendations by computing the similarity between the candidate and the clicked news (Wang et al. 2018a; Zhu et al. 2019).

Overall, CNNs are useful methods for representing multimodal (text, audio, video) features from the news data.

5.4 Recurrent neural network (RNN)

RNN models are used to model variable-length sequence data. In a recommender system, the RNNs are often used to model sequential dependencies in the rating data and for session-based recommendation tasks (Hidasi et al. 2016). Two well-known variants of RNNs are Long short-term memory (LSTM) and Gated recurrent unit (GRU). The key difference between the two is that GRU does not need memory units as in LSTM, so, GRUs are faster to train. However, it is easier to learn longer sequences with LSTM.

The GRUs have been used in a few NRS to learn short-term users' preferences from the interaction histories (Okura et al. 2017; An et al. 2019; Zhang et al. 2019). The results demonstrated a significant improvement over the traditional temporal models, with slightly better performance of GRUs over LSTM (Okura et al. 2017).

Song et al. (Song et al. 2016) propose to learn user's short-term preferences using unidirectional LSTM (Song et al. 2016). The unidirectional LSTM only preserves information of the past. The improvement over unidirectional LSTM is made by replacing it with the bidirectional LSTM in another NRS (Kumar et al. 2017). Bidirectional LSTM runs the user input sequence in two ways, one from the past to the future (forward pass) and one from the future to the past (backward pass).

In a few recent NRS (de Souza Pereira Moreira 2018; An et al. 2019; Wu et al.), the GRUs are successfully used to learn users' short-term preferences. Some NRS (Zhu et al. 2019) also use LSTM networks for spotting users' preferences in shorter time periods. One NRS (Wu et al. 2019a) also adds the neural attention (Vaswani et al. 2017) on each state of RNN to get rich sequential features during different clicking time.

Overall, RNNs are useful for modeling session-based and sequence-based recommendation tasks. These models can also be used to incorporate additional news-related information during different temporal steps (An et al. 2019).

5.5 Neural attention

The neural attention (Vaswani et al. 2017) is based on the idea that a model pays attention to a specific part when processing a huge amount of information. Neural attention has gained remarkable success in a variety of machine learning applications, including language modelling, image captioning, and text classification. The attention mechanism is also employed in recommender systems to filter out noisy content and to select the most representative items.

In some NRS (Wang et al. 2018a; An et al. 2019; Wu et al. 2019a), the attention is used at the word-level to learn informative words from the news content. The attention is also applied at the news-level to model the informativeness of different kinds of news information for learning useful news representations (Wu et al. 2019b). For example, if news headline is more important than the other pieces of news (news body, topic, taxonomy), then it should be weighed more. Because the informativeness of the same words and news

may differ amongst users, the idea of personalized attention network is applied in another NRS (Wu et al. 2019a). The personalized attention network uses the embedding of the user information as the query vector of word and news-level attention networks; and attends to significant words and news in different ways based on user preferences.

The attention mechanism are useful to learn news and user representations from the neural networks and is the backbone of Transformer models (Devlin et al. 2018).

5.6 Graph neural network (GNN)

Recently, the GNN models (Scarselli et al. 2008) have gained increasing popularity in a variety of domains, including social networking, recommendation systems, search engines and related. GNN is a type of neural network that operates directly on the structure of the graph. Essentially, each node in the graph is linked to a label, and the task is to predict the label. GNNs are used for classification tasks, such as text classification, labelling sequences, machine translation, and for prediction tasks.

GNNs are used for the recommendation tasks in a few recent NRS (Wu et al. 2019c; Lee et al. 2020; Sheu and Li 2020; Yang et al. 2020; Ge et al. 2020). GNewsRec (Ge et al. 2020) is a GNN-based news recommendation system that constructs a reader-news-topic graph to learn the embeddings from the news features and reader's clicks. Both representations (news and reader) are then used to determine the click probability of the candidate news to be recommended next.

Overall, the GNNs are promising models that yield outstanding outcomes when paired with the attention mechanism (Wu et al. 2020c).

5.7 Transformers

Transformer model, introduced in the neural attention paper (Vaswani et al. 2017), has achieved state-of-the-art performance in NLP tasks. Transformers are intended for handling the sequential data in the same manner as RNNs. However, in contrast to RNNs, the transformers do not require the sequential data to be processed in order (one after other). Instead, the transformers process sequential data in parallel. The main crux of a Transformer is the self-attention layer. The self-attention looks at an input sequence and decides at each step which other parts of the sequence are important.

The idea of Transformers is taken from the transfer learning, where a big language model is usually trained on billions of words, and the knowledge from the big model is transferred to similar smaller NLP tasks. For example, the Google BERT (Devlin et al. 2018) model is pre-trained on a large corpus of unlabeled text including the entire Wikipedia and Toronto Book Corpus, and is used to train other models on downstream NLP tasks, for the purpose of making better predictions. Well-known Transformer models are BERT, BART, ALBERT, GPT-2, RoBERTa, and other listed here.²⁷

The deep bidirectional self-attention BERT is employed to model the sequences in user's behavior for the click prediction task in a recommender system (Sun et al. 2019). A recent NRS (Wu et al. 2021) is built on the same idea to use the BERT for the task of news recommendations.

²⁷ <https://huggingface.co/transformers/>.

5.8 Reinforcement learning (RL)

Deep RL methods are based on trial-and-error paradigm and have demonstrated human-level performance across various domains such as games, robotics, finances and even recommenders (Francois-Lavet et al. 2018). RL consists of five components (agents, environments, states, actions, and rewards) to get knowledge from the raw data. Deep Q-Learning (DQN) is a RL strategy that, given a current state, helps to find the maximum expected future reward of an action. The DQN structure has been applied in an NRS (Zheng et al. 2018) to model the dynamics in users' preferences and that of news content. The RL models can also be used to define the best sequence of decisions through the interaction with the news environment and observation of rewards (clicks).

5.9 Summary

The DL methods have proven to be very successful in building the NRS and appear to have the great potential to be further used in the future. Despite the success of these methods, one limitation is noted. It is that the current NRS research (including DL-based models) focus too much on the accuracy of the models to provide recommendations to the users. The aspects beyond accuracy such as novelty, serendipity, diversity and a composite user model are not very much covered in these approaches. These deep learning solutions and the challenges they have addressed for the NRS are summarized in Table 4.

As can be seen in Table 4, the user modeling is the most widely addressed challenge in the DL-based NRS. The challenge of timeliness is also addressed in these models. Usually, the sessions-based recommendation tasks are used to model timeliness in users' short-term preferences. These sessions are created in a chronological order of item click events or the publication time of the news items. There is not much work seen in terms of addressing the content quality challenge in these methods. Among all the DL methods, the CNNs and RNNs are popular choices for article and user representations. The attention mechanism can be found in the most recent DL papers. GNNs (with addition of attention) and the Transformers (primarily based on the neural attention) are also used in some recent papers. The accuracy metric and offline protocol remain the popular evaluation methods used in the DL-based NRS. There are also other useful DL models that are not seen in the NRS work lately, which we will discuss in Sect. 7 (Discussion).

6 Effects of news recommendation algorithms on readers' behavior

News organizations such as BBC, New York Times, The Guardian, and such, have worked hard to provide more personalized news stories to readers via their websites and applications. These recommendations are tailored to readers' preferences based on the topics of interest they have indicated in their profiles or, in some cases, the content they have recently consumed. It is a great accomplishment to provide readers with everything that

truly reflects their interest. However, relying solely on machine learning algorithms, as in recommender systems, is not without risk. They are thought to have a negative impact on the production of news (fake news, exaggerated news, racism, persecution, stereotypes, and so on), readers' psychology, consumption behaviors, and overall user experiences with an NRS.

Although these negative impacts are being recognized in the literature of computer science, there is only a limited amount of work (Nguyen et al. 2014; Allcott and Gentzkow 2017; Möller et al. 2018; Helberger 2019) that briefly touch the issue (post-algorithmic effects of news recommendations on readers' behavior). This issue has been widely discussed in other disciplines such as information science and mass communication, where they blame news recommendation algorithm developers for making poor design choices.

The birth of social media, fake news, and polarized political media groups has been blamed for the effects of news recommendations on user behaviors (Allcott and Gentzkow 2017). Some authors (Beam 2014; Quattrociocchi et al. 2016; Anspach 2017) see social media interference in news media as a threat to democracy. For example, Quattrociocchi et al. (2016) conduct a study on the Facebook group user engagement data to determine whether echo chambers exist on social media. According to their findings, social network users create like-minded echo chambers on certain issues, limiting their exposure to counter-attitudinal behavior.

According to our findings, this topic has a high social relevance in a variety of disciplines, including computer science, journalism, political science, and economics. We gathered some statistics from Pew Research Center²⁸ reports, which conducted extensive surveys on these issues. Following these steps, we identify the major effects on users' behaviors. We also discuss the possible mitigation strategies in this section.

6.1 Post-algorithmic news recommendation effects

Filter bubble corresponds to intellectual isolation caused by personalized searches or algorithms to selectively assume the information an individual wants to see (Pariser 2011).

Echo chamber refers to an information bubble around a user, where the user is only exposed to articles that reinforce their existing beliefs (Flaxman et al. 2016).

Polarization refers to the divergent views on policy (politics, religion, beliefs) into ideological extremes (Dandekar et al. 2013). The frequent interactions between like-minded persons result in polarization.

Fragmentation of the public sphere refers to the disintegration of the shared public sphere into smaller publics where the citizens in those spheres become are less aware of outside issues (Helberger 2019).

Dehumanization refers to the control of human judgement through predictive modelling without readers knowing how it is done. All human decisions are overtaken by artificially generated logic (Page et al. 2018).

Biased assimilation refers to biases in readers caused by algorithms. Users begin to process new information in a biased manner, which eventually reinforces them to critically examine disconfirming evidences (Dandekar et al. 2013).

Denial to Counter-attitudinal behavior Counter-attitudinal behavior is defined as behavior that does not align with one's points of view but is valued and regarded as a high level of exposure to different points of view. (Beam 2014). Denial to counter-attitudinal behavior is an issue caused by filter bubble or echo chambers.

²⁸ <http://www.pewresearch.org/>.

Reinforced digital gate-keeping refers to the selection and extraction of all news through digital gates (recommenders) with no human judgement (Möller et al. 2018).

Deep Fakes refers to media created by artificial neural networks that takes a person in an existing image/video and replace with someone else, for example,²⁹ the deep fake of Obama public announcement, and Donald Trump speaking informally. Deepfakes are created using social media, which has resulted in fake news and other conspiracy theories.

6.2 Mitigating effects of news recommendations on user behavior

We have reviewed the state-of-the-art solutions to mitigate the effects of news algorithms on readers' behavior. First, we discuss the solutions from the state-of-the-art NRS papers and then we discuss other solutions in Sect. 7.

6.2.1 Selective exposure

Selective exposure research is taken from Festinger's cognitive dissonance theory (Festinger 1962), a discipline of psychology that states people prefer to view information that supports their own perspectives (Hart et al. 2009). According to this theory, dissonant information (information that does not match with the user attitude) increases uncertainty and discomfort in a user. As a result, the user may read information that is pro-attitudinal (congruent with user behavior) and try to avoid information that is counter-attitudinal (conflicting with his perspectives). However, empirical research (Brundidge 2010) in selective exposure indicates that readers may also want to select and read different news stories to gain knowledge for both pro- and counter-attitudinal information.

Garret (2009), for example, demonstrates through a user study that during the elections days, people tend to search online news about their favorite candidates. Further, the same participants also went on to search online news for the opposing candidates and read about their perspectives. This finding contradicts the pro-attitudinal user behavior.

Beam (2014) demonstrates through a user study that during selective exposure, users only select news stories that match with their own preferences. While doing so, they may be presented with news stories that contradict with their own beliefs; in this case, they may still want to read them so that they can form their own opinion on a particular issue. Flaxman et al. (2016) support this, demonstrating through a large-scale user study that selective exposure during online news consumption exposes readers to information that does not always align with their political beliefs.

Another group of researchers (Flaxman et al. 2016; Newman et al. 2018) believe that social media users are much more likely to encounter sources they would not normally encounter, exposing them to opposing viewpoints. Flaxman et al. (2016) conduct a user study and analyze the web browsing histories of 50,000 US citizens who regularly read online news. The results demonstrate that the usage of social networks and search engines exposes users to counter-attitudinal information. The Reuters Report 2018 (Newman et al. 2018) also presents a user study and affirms the previous research that social media plays a role in increasing users' exposure to news.

Dandekar et al. (2013) use DeGroot's graphical model of opinion to address the polarization problem in the news domain, in which individuals update their opinions based on

²⁹ <https://www.creativebloq.com/features/deepfake-examples>.

a weighted averaging of their current opinions and those of their neighbors. Herlberger (2019) also proposes a democratic recommender system providing news readers with a diverse mix of news recommendations.

Overall, more research is needed to include selective exposure in the design of an NRS.

6.2.2 Diversity-aware algorithms

These algorithms take diversity into account at various stages of the recommendation process, such as during the re-ranking process (after recommendations are generated) or the optimization phase (the recommendation process). A recommendation algorithm is typically programmed to promote exposure to unpopular items (long tail items) during the re-ranking process. During the optimization phase, the recommendation algorithm is tailored so that diversity, as well as the (built-in) accuracy objective, are included in the recommendation process. The news topics, writing styles, tags, perspectives, contexts, and ideologies are some of the factors that are considered to be diversified in an NRS (Resnick et al. 2013; DiFranzo and Gloria-Garcia 2017; Möller et al. 2018; Helberger 2019). Möller et al. (2018) also propose to incorporate the diversity in an NRS as a democratic function identifiable in news articles, subjects, tones, styles of writing and political content.

In an earlier NRS (Rao et al. 2013), the news recommendation list is expanded by using the news taxonomy information to find relevant news items from encyclopedia websites. In another NRS (Zheng et al. 2018), the multi-arm bandit methods with exploration–exploitation optimization is used to tradeoff between accuracy and diversity. In a recent NRS (Raza and Ding 2020a), the diversity is included through the use of regularization (Ridge regression for accuracy and Lasso regression for diversity) during the optimization phase. The diversity is, then, balanced with high accuracy in the model.

An aspect is a collection of attributes, components, or services that can be used to categorize information. The aspects can diversify the news recommendations by providing readers with different perspectives on a news topic. In one NRS (Park et al. 2009), the news events are classified based on various aspects (topics) and then users are provided with different perspectives on news. Although there has been little work in aspect-level presentation in an NRS, it can be very useful to classify or cluster news articles based on other aspects (styles, tags, categories, sentiments) to make recommendations.

6.2.3 Nudge theory

This refers to giving subtle nudge (touch or push) in the form of small design changes that encourage users to make other choices in their general interest (van der Heijden and Kosters 2015). Nudging is a behavior change strategy that motivates people to achieve goals, and it can influence the behavior of news readers.

There have been some cases in real-world, where algorithms have been manipulated to steer readers toward fake news. For example, YouTube was manipulated consistently alongside the Guardian news to nudge readers towards sensational and fake news during US elections 2016 (DiFranzo and Gloria-Garcia 2017). Recently, the news websites are used in conjunction with the social media add-ons to spread the anti-vaccination misinformation and the rumor that incorrectly compared the number of registered voters in 2018 to the number of votes cast in US Elections 2020.³⁰ The implications of such news are seen

³⁰ <https://archive.is/OXJ60>.

in the anti-vaccine movements preventing the global fight against the COVID-19, or in the post-election unrest.

Despite these negative examples, nudges can be extremely beneficial when used transparently and ethically. Algorithms can be programmed to guide users toward more politically balanced news consumption and exposure. Resnick et al. (2013) design an interface (a browser add-on) that nudges the users to select more news rather than just relying on the algorithmic recommendations. There is also some work that demonstrates the general design and architecture of a smart nudge in a recommender system (Karlsen and Andersen 2019). Algotransparency³¹ is also an information group that informs citizens on how people are nudged from an initially neutral search on YouTube to the progressively biased information throughout each subsequent phase of the recommendation cycle.

Overall, nudges, if used correctly, can help users make wise choices through selective exposure. However, it is difficult to observe users' behaviors excessively during nudging.

6.2.4 Trade-off among various evaluation measures

Maksai et al. (2015) quantify the trade-off between different metrics such as accuracy-coverage, accuracy-diversity, accuracy-serendipity, diversity-serendipity to test the performance of their recommendation algorithms. The results show that accuracy, when combined with beyond-accuracy measures, improves user behavior within an NRS.

Concerns about the potential negative consequences of personalization in the NRS have grown in recent years (Haim et al. 2018). Personalization is often the result of recommendations that align highly with users' preferences. Usually, a high accuracy results in higher personalization in a recommender system. However, we believe that personalization should not be completely ignored; otherwise, users may lose interest in an NRS where everything that is recommended is different or diverse. In fact, as demonstrated in a recent study, personalization can be balanced with reasonable diversity in an NRS (Raza and Ding 2020a).

Chakraborty et al. (2019) also take a closer look to balance three metrics: recency, importance (or popularity) and diversity in an NRS. In that, they propose a future-impact metric that takes the popularity signals from crowd-sourced information and the personalized information from the past news data to predict the impact of news stories for a news reader.

Overall, there has been little research in NRS that balances the built-in accuracy aspect with various aspects of quality evaluation and beyond-accuracy aspects.

6.2.5 Summary

In the NRS, there is only a small amount of work that considers these factors (such as diversity, selective exposure, nudges, and aspects) in its design. The absence of such methods results in news recommendations that are entirely driven by the algorithmic logic of recommendation models or by the motivations of stakeholders (political figures, trading factors, etc.). The limited work addressing these issues is summarized in Table 5. There are also some suggestions that we discuss further in Sect. 7.

³¹ <https://algotransparency.org/>.

Table 5 Post-algorithmic challenges in the NRS and the solutions

Paper	Challenge	Algorithmic Cause	Effect on readers' behavior	Solution
Dandekar et al. (2013)	Polarization	Narrowed readers' exposure	Causes denial to others' viewpoint	Degroot's graphical model of opinion
Resnick et al. (2013)	Filter bubble	Over personalization	Creates filter bubbles, polarization	Selective exposure, diversity, nudge theory
Beam (2014)	Counter-attitudinal	Over personalization	Affects readers' acceptance to opposing viewpoints	Selective exposure
Li et al. (2014)	Filter bubble	Over personalization	Readers get bored of old news	Budgeted maximum coverage
Maksai et al. (2015)	Filter bubble	Over personalization	Readers get bored of similar news stories	Trade-off among various evaluations
Flaxman et al. (2016)	Filter bubble, Echo chambers	Ideological segregation	Affects voters and functioning democracies	Selective exposure
Allcott and Gentzkow (2017)	Filter bubble, Echo chambers	Algorithmic fake news	Separates readers from contradictory perspectives	Selective exposure
DiFranzo and Gloria-Garcia (2017)	Filter bubble,			
Echo chambers	Social media spread fake news	Causes readers' denial to others' opinions	Diversity-aware methods and nudges	Diversity in re-ranking
Möller et al. (2018)	Filter bubble	Accuracy-centric algorithms	Extreme opinions, misinterpreted facts	Diversity in re-ranking
Helberger (2019)	Filter bubble	Over personalization	Affects negatively on the democracy	Exposure diversity, ban manipulative practices
Zheng et al. (2018)	Filter bubble	Over personalization	Readers get bored of similar news stories	Dueling Bandit Gradient Descent for diversity
Chakraborty et al. (2019)	Filter bubble	Over personalization	Readers tend to get bored of similar news stories	Future-impact metric

7 Discussion on research implications and future work

In this section, we highlight our major findings in this survey and discuss the research implications and future work.

7.1 Algorithmic solutions and major challenges in NRS

Our review of selected publications revealed that NRS research is gradually gaining attention over time. One reason for this increase is the high conversion rate of traditional news media users to online news readers. This growth has provided researchers with numerous research opportunities to develop solutions to the unique challenges of news domain. Due to the rapid advancement of various DL methods, there has been a recent evolution in NRS research.

As discussed in Sect. 4, the traditional recommendation algorithms are not enough for building an NRS and can only partially address the challenges in the NRS. It requires a lot of modifications, extensions and variations on the standard recommendation approaches to meet the needs of news readers. The latent factor models as discussed in Sect. 3.2 and the DL models in Sect. 5, are two major classes of successful models to address the challenges faced by the NRS. The DL models, in particular, continue to be used in the recent research.

7.2 Deep neural recommenders

We draw a classification of successful models used in the NRS in Sect. 5. This information can be useful to researchers in this field, especially new researchers, to gain some knowledge and understand the guidelines on how to choose a suitable model or framework for building an NRS. For example, the Restricted Boltzmann machines (RBMs) (Salakhutdinov et al. 2007) with only two layers can be used to extract features from large news datasets using low-rank representations. The Deep Belief Network (Hu et al. 2014), a multi-layer learning architecture with a stack of RBMs, can be used to extract useful features from the news content.

There are also other DL models that can be applied in an NRS. For example, the Generative Adversarial Network (GAN) (Goodfellow et al. 2014) consisting of two competing (adversarial) neural networks (a discriminator and a generator) that run in against each other to generate new synthetic samples of data that can pass for real data. For example, GANS can be used to generate new data for an NRS with the same statistics as the training set.

There are a variety of neural networks that may be combined to produce models that are both powerful and expressive. For example, the CNNs can be used to learn feature representations from the news content and RNNs can be used for sequential user modeling. Combining AE and RNNs can capture the sequential information (through RNN) from the item content while using the lower-dimensional feature representations (through AE). These models can also be integrated with the neural attention to pick useful news recommendations.

Transfer learning can also be used to address the data sparsity problem of the NRS by transferring the knowledge from large pre-trained models to the problem of news recommendation. However, the challenge here is that the pre-trained model should be based on the news dataset. Otherwise, the noise and outliers from other unrelated datasets can be transferred into the news recommendations.

Despite the significant advances in DL theory, these methods are not without flaws. For example, DL methods demand much more data and require much more parameter adjusting than standard methods. Also, these models behave like blackboxes, providing limited interpretability (due to hidden layers, weights and activation functions) and little explainability (explanation for the internal working) in the recommendation tasks.

7.3 Accuracy and beyond-accuracy aspects, and evaluation protocols

We shed some light on accuracy and beyond-accuracy aspects in this survey. Accuracy is important but the quality of news recommendations cannot be improved without considering beyond-accuracy aspects. As shown in Fig. 5 and Tables 3 and 4, the research in beyond-accuracy aspects in NRS is limited and seems to appear trivially in recent years.

There has been some limited work in NRS that has used online evaluation and user study techniques to test. However, as seen in Tables 3 and 4, as well as Fig. 6, the offline option is the most popular model evaluation protocol. Usually, the online evaluations are costly for large-scale news data could be one reason for this. One future research direction is to test these NRS models in real-world settings either by reducing the computational cost of these methods through techniques such as quantization, compression and pruning methods (Kitaev et al. 2020); or by working to manage more computational resources for in real-time experimental setup.

7.4 Diversity as the key principle in the design of NRS

In the state-of-the-art of NRS, there is little work on the diversity aspect. Diversification in an NRS is necessary not only to keep readers engaged in the reading process, but also to keep readers from becoming trapped in filter bubbles. In order to understand why and how much diversity can be included in an NRS, academics and designers should collaborate with news organizations and social media platforms. The architecture of the news or social media website, the incorporation of the nudge theory, selective exposure, and the detection of fake news are all key aspects that should be considered while developing an NRS.

7.5 Diversity through neural attention

The neural attention can be successfully used to introduce diversity in session-based recommender systems (Nema et al. 2018). Generally, the diversity is intrinsically reflected in the users' short-term interests (Wang et al. 2018a). Under normal conditions, the attention mechanism can be used to sum the weights of the hidden layer to generate the representation vector. The issue with this approach is that if there are repeated actions in a session, the recommendations that are generated for those sessions are also similar. Therefore, it is

critical to respond to users' idiosyncratic clicks during different time intervals to include diversity. A scaling weight can be assigned to the query vectors in the attention mechanism. The idea is to dampen the importance of repeated clicks and to give some weight to non-repeated users' actions using attention. So far, there is not much work in the NRS that considers including diversity in attention-based models.

7.6 Multi-criteria evaluation

There are also other aspects for evaluation that are unexplored in an NRS, such as trustworthiness (level of user trust on system), preserving privacy, efficiency (ease of the search and accessibility of the information), robustness (ability to make relevant predictions in the presence of noisy data), as well as the trade-offs among various aspects. Including these aspects in an NRS could enhance the user experience.

7.7 User experience model

There is no benchmark to evaluate user experiences in a recommender system. Also, the existing user-modeling evaluation frameworks (Konstan and Riedl 2012; Knijnenburg et al. 2012) for other recommendation domains are too expensive for an NRS. The evaluations in these frameworks are through user studies or experiments only, which is not practical in an NRS with real-time constraints. It is also a challenging task to adapt these models for the news domain. Another issue with these frameworks is that they rely only on the user study, and they do not consider any accuracy and beyond-accuracy aspects. Nonetheless, without these basic metrics, it could not offer a complete picture of the user experience. There is a need for a benchmark user modeling framework in the NRS to evaluate the experience of news readers. Such framework is not only required to provide better or enjoyable experience to the readers (as in other recommender domains), but also vital for an NRS to play its democratic, liberal and deliberative role in the community.

7.8 News dataset

Our findings from Sect. 4 reveal that there are very few datasets in the NRS. Many of the datasets shown in Tables 2 and 3 and Fig. 7 are privately owned, having been created to meet the immediate research needs of the problem to be solved. There should be more challenges, such as CLEF NEWSREEL or MIND Leaderboard, to encourage researchers to design better NRS in real-time constraints.

7.9 Implicit user feedback

In an NRS, we often need implicit ratings to infer latent information from enriched user interactions. However, it could be tricky sometimes to decide whether the implicit feedback is positive or negative. For example, time spent on news articles should not always be considered as user's engagement during news reading because it could be the idle time

(Agarwal and Singhal 2014). Skips from the readers are often considered as indicator of user's interest in different topics, but it could be because of repetitive news stories that force the user to skip those to find new news items (Ma et al. 2016). It is not clearly mentioned in the literature that how to find out which specific property of the system makes the recommendation uninteresting to the user. If we could devise some way to differentiate between positive and negative preferences, we can improve the quality of recommendations based on positive preferences and avoid suggesting news items to the users if these result in negative or neutral preferences.

7.10 Gamification

The gamification means the use of game design elements in other applications where there is no gaming context (Chou 2019). The purpose of gamification is to motivate and promote user activities. The idea of gamification has not been used in the NRS. But it can be similar to Google Guides in Google Maps. In that, the NRS can assign rewards, in the form of points, badges, avatars, leaderboards etc., to the readers based on their explicit interactions with the system. This can be a useful tool to improve user engagement and to overcome cold-start problem in the NRS.

7.11 Mitigating effects of news recommendations on readers' behavior

The effect of news recommendations on user behavior is one of the most overlooked area in the research of recommender systems. This topic has not attracted enough attention in the computer science field before the emergence of grave issues like fake news, deep fakes, yellow journalism (exaggerating facts or spreading rumors), ideological segregation and extremism in society due to the media war. By highlighting these problems related to the effects of news recommendations on readers' behavior in Sect. 6, we have presented the new research opportunities for the academic scholars to work along this direction.

So far, the solutions based on selective exposure, diversity-aware algorithms and suggestion on banning manipulative practices are not enough for two reasons: (i) they are only demonstrated on small scale experimental setup, (ii) they are based on avoidance to these techniques, which are insufficient to detect and prevent such effects from the system. The researchers in this field need to find other ways (either algorithmic or heuristic) to prevent, detect and break down those effects (filter bubbles, echo chambers) if they are prevailing. There are a few suggestions that might be useful to mitigate the effects of news recommendations. These are given below:

- *Transparency* The design of news recommendation algorithms should give a much clearer view of the world as it is, not as the user wants it to be. It is no more a hidden fact that search engines such as Google use many dimensions of our online and offline behavior to determine the links that we are most likely to click from a given search.³² In the battle to keep the news readers engaged all the time, the news recommendation algorithms are being designed in a similar way as these search engines. However, we argue that to reduce the post-algorithmic effect, we should re-design these algorithms so that they allow users to indicate their interests and then find relevant (novel, recent,

³² <https://fortune.com/2017/06/28/gmail-google-account-ads-privacy-concerns-home-settings-policy/>.

important) content from diversified sources accordingly. This is much similar to introducing selective exposure and motivated information processing in the NRS.

- *Going incognito* Going incognito (private mode) in browser turns off history tracking, hides cookies and logs the users out from social media sites like Google and Facebook. These social network sites transmit information about users to other websites and create echo chambers around users. In this way, the news browser is depersonalized, and a news reader receives news stories from different sites and perspectives that they would otherwise not see without incognito.
- *Rules and regulations of recommender system's objectivity* User information is highly exposed during the profiling phase in the recommender systems. Although there are rules and regulations such as General Data Protection Regulation (GDPR), to protect the misuse of personal information from companies and public institutes. But when it comes to recommender systems, none of the solutions comply with these regulations. The researchers and designers of NRS need to follow these rules and regulations, not only for privacy-preserving, but to make NRS a reliable system.

7.12 Interdisciplinary research

There is a need for interdisciplinary research where the expertise from both social science and computer science can be combined. The researchers may utilize the recent advancements in text analysis, representation learning and attention-based models to address the challenges specific to the news domain.

This section can only provide a partial list of some of the challenges, research directions, future opportunities and issues in the NRS. We would like to have this survey to serve as a doorway to a wealthy source of open research problems that make NRS a productive and interesting research area to work on.

8 Conclusion

NRS has been increasingly used in recent years to provide better suggestions to end users so that they can consume online news from various sources. There are many unique challenges associated with the NRS, most of which are inherited from the news domain. Out of these challenges, the issues related to timeliness, evolving readers' preferences over dynamically generated news, quality of news content and the effects of news recommendations on users' behavior are prominent ones. The general recommendation algorithms are insufficient to provide news recommendations since they need to be modified, varied or extended to a large extent. Recently, the DL-based solutions have addressed much of those limitations of conventional recommenders. Accuracy is considered as a standard evaluation measure to assess the quality of a recommender system. However, beyond accuracy, other aspects such as diversity, coverage, novelty, serendipity are also important to provide better user experience in an NRS. Datasets, open recommendation platforms and evaluation protocols together play a role in developing recommendation solutions in the news domain. We have covered them in this survey so that the readers can get an insight into the current research practices and may start to help develop them. Different from other survey papers,

we also discuss about the effects of news recommendations on readers' behavior in this survey. Lastly, though this survey is centered around the NRS, the knowledge and insights gained from the findings of this survey can also be used to build recommender solutions for other application domains.

Acknowledgements This work is partially sponsored by Natural Science and Engineering Research Council of Canada (Grant 2020-04760).

References

- Adomavicius G, Kwon YO (2008) Overcoming accuracy-diversity tradeoff in recommender systems: a variance-based approach
- Adomavicius G, Tuzhilin A (2005) Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans Knowl Data Eng* 17:734–749
- Agarwal S, Singhal A (2014) Handling skewed results in news recommendations by focused analysis of semantic user profiles. *IEEE*, pp 74–79
- Agarwal S, Singhal A, Bedi P (2013) IPTC based ontological representation of educational news RSS feeds. In: *Proceedings of the Third International Conference on Trends in Information, Telecommunication and Computing*. Springer, pp 353–359
- Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. *J Econ Perspect* 31:211–236
- An M, Wu F, Wu C, et al. (2019) Neural News Recommendation with Long- and Short-term User Representations. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Florence, Italy, pp 336–345
- Anspach NM (2017) The new personal influence: how our facebook friends influence the news we read. *Polit Commun* 34:590–606
- Asikin YA, Wörndl W (2014) Stories around you: location-based serendipitous recommendation of news articles. In: *UMAP Workshops*. Citeseer
- Baldwin R (2014) From regulation to behaviour change: giving nudge the third degree: giving nudge the third degree. *Modern Law Rev* 77:831–857
- Beam MA (2014) Automating the news: how personalized news recommender system design choices impact news reception. *Commun Res* 41:1019–1041
- Beck PD, Blaser M, Michalke A, Lommatzsch A (2017) A system for online news recommendations in real-time with apache mahout. In: *CLEF (Working Notes)*
- Beckett C, Deuze M (2016) On the role of emotion in the future of journalism. *Soc Media Soc* 2:2056305116662395
- Borges HL, Lorena AC (2010) A survey on recommender systems for news data. In: *Szczerbicki E, Nguyen NT, Kacprzyk J (eds) Smart information and knowledge management*. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 129–151
- Boutet A, Frey D, Guerraoui R, et al. (2013) WHATSUP: a decentralized instant news recommender. In: *2013 IEEE 27th International Symposium on Parallel and Distributed Processing*. pp 741–752
- Brodth T, Hopfgartner F (2014) Shedding light on a living lab: The CLEF NEWSREEL open recommendation platform. In: *Proceedings of the 5th Information Interaction in Context Symposium*. ACM, NY, USA, pp 223–226
- Brundridge J (2010) Encountering “difference” in the contemporary public sphere: the contribution of the Internet to the heterogeneity of political discussion networks. *J Commun* 60:680–700
- Cao S, Yang N, Liu Z (2017) Online news recommender based on stacked auto-encoder. In: *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*. pp 721–726
- Chakraborty A, Ghosh S, Ganguly N, Gummadi KP (2019) Optimizing the recency-relevance-diversity trade-offs in non-personalized news recommendations. *Inf Retr J* 22:447–475
- Chakraborty A, Paranjape B, Kakarla S, Ganguly N (2016) Stop clickbait: Detecting and preventing clickbaits in online news media. In: *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. pp 9–16
- Chakraborty A, Ghosh S, Ganguly N, Gummadi KP (2017) Optimizing the recency-relevancy trade-off in online news recommendations. In: *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp 837–846

- Chou Y (2019) Actionable gamification: beyond points, badges, and leaderboards. Packt Publishing Ltd, Birmingham
- Constantinides M, Dowell J (2018) A framework for interaction-driven user modeling of mobile news reading behaviour. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization - UMAP '18. ACM Press, Singapore, Singapore, pp 33–41
- Cucchiarelli A, Morbidoni C, Stilo G, Velardi P (2018) What to write and why: a recommender for news media. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing. ACM, NY, USA, pp 1321–1330
- Dandekar P, Goel A, Lee DT (2013) Biased assimilation, homophily, and the dynamics of polarization. *Proc Natl Acad Sci* 110:5791–5796
- De Francisci Morales G, Gionis A, Lucchese C (2012) From chatter to headlines: harnessing the real-time web for personalized news recommendation. In: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. ACM, NY, USA, pp 153–162
- de Souza Pereira Moreira G (2018) CHAMELEON: a deep learning meta-architecture for news recommender systems. In: Proceedings of the 12th ACM Conference on Recommender Systems - RecSys '18. ACM Press, Vancouver, British Columbia, Canada, pp 578–583
- Desarkar MS, Shinde N (2014) Diversification in news recommendation for privacy concerned users. In: 2014 International Conference on Data Science and Advanced Analytics (DSAA). pp 135–141
- Devlin J, Chang M-W, Lee K, Toutanova K (2018) Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805
- DiFranzo D, Gloria-Garcia K (2017) Filter bubbles and fake news. *XRDS* 23:32–35
- Ding Y, Li X (2005) Time weight collaborative filtering. In: Proceedings of the 14th ACM international conference on Information and knowledge management - CIKM '05. ACM Press, Bremen, Germany, p 485
- Domann J, Lommatzsch A (2017) A highly available real-time news recommender based on apache spark. In: Jones GJF, Lawless S, Gonzalo J, et al. (eds) *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Springer International Publishing, pp 161–172
- Doychev D, Rafter A, Lawlor A, Smyth B (2015) News recommenders: real-time, real-life experiences. In: Ricci F, Bontcheva K, Conlan O, Lawless S (eds) *User modelling, adaptation and personalization*. Springer International Publishing, New York, pp 337–342
- Dwivedi SK, Arya C (2016) A survey of news recommendation approaches. In: 2016 International Conference on ICT in Business Industry Government (ICTBIG). pp 1–6
- Dziugaite GK, Roy DM (2015) Neural network matrix factorization. arXiv preprint arXiv:1511.06443
- Feng C, Khan M, Rahman AU, Ahmad A (2020) News recommendation systems-accomplishments, challenges & future directions. *IEEE Access* 8:16702–16725
- Festinger L (1962) *A theory of cognitive dissonance*. Stanford University Press, California
- Flaxman S, Goel S, Rao JM (2016) Filter bubbles, echo chambers, and online news consumption. *Public Opin Q* 80:298–320
- Fortuna B, Moore P, Grobelnik M (2015) Interpreting news recommendation models. In: *WWW'15 Companion: Proceedings of the 24th International Conference On World Wide Web*. pp 891–892
- Francois-Lavet V, Henderson P, Islam R et al (2018) An introduction to deep reinforcement learning. *FNT Mach Learn* 11:219–354
- Frolov E, Oseledets I (2017) Tensor methods and recommender systems. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 7:e1201.
- Garcin F, Faltings B (2013) PEN recsys: a personalized news recommender systems framework. ACM Press, New York
- Garcin F, Dimitrakakis C, Faltings B (2013) Personalized news recommendation with context trees. In: Proceedings of the 7th ACM conference on Recommender systems - RecSys '13 105–112.
- Garrett RK (2009) Politically motivated reinforcement seeking: reframing the selective exposure debate. *J Commun* 59:676–699
- Garrido AL, Buey MG, Ilarri S, et al (2015) KGNR: A knowledge-based geographical news recommender. In: 2015 IEEE 13th International Symposium on Intelligent Systems and Informatics (SISY). pp 195–198
- Ge S, Wu C, Wu F, et al (2020) Graph enhanced representation learning for news recommendation. In: Proceedings of the Web Conference 2020 2863–2869. h
- Gharahighchi A, Vens C (2019) Extended bayesian personalized ranking based on consumption behavior. *Artificial intelligence and machine learning*. Springer, New York, pp 152–164
- Gillis N (2020) Nonnegative matrix factorization. SIAM, New Delhi

- Göçeri E (2020a) Convolutional neural network based desktop applications to classify dermatological diseases. In: 2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS). IEEE, pp 138–143
- Göçeri E (2020b) Impact of deep learning and smartphone technologies in dermatology: automated diagnosis. In: 2020 Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, pp 1–6
- Goodfellow I, Pouget-Abadie J, Mirza M, et al (2014) Generative adversarial nets. In: Advances in neural information processing systems. pp 2672–2680
- Gruppi M, Horne BD, Adalı S (2021) NELA-GT-2020: a large multi-labelled news dataset for the study of misinformation in news articles. arXiv preprint arXiv:210204567
- Gu W, Dong S, Zeng Z, He J (2014) An effective news recommendation method for microblog user. *Sci World J* 2014:1–14
- Guan X, Peng Q, Li Y, Zhu Z (2017) Hierarchical neural network for online news popularity prediction. In: 2017 Chinese Automation Congress (CAC). pp 3005–3009
- Gulla JA, Zhang L, Liu P, et al (2017) The addressa dataset for news recommendation. In: Proceedings of the International Conference on Web Intelligence. ACM, NY, USA, pp 1042–1048
- Gunawardana A, Shani G (2009) A survey of accuracy evaluation metrics of recommendation tasks. *J Mach Learn Res* 10:2935–2962
- Haim M, Graefe A, Brosius H-B (2018) Burst of the filter bubble?: effects of personalization on the diversity of *Google News*. *Digit J* 6:330–343
- Hamborg F, Donnay K, Gipp B (2019) Automated identification of media bias in news articles: an interdisciplinary literature review. *Int J Digit Libr* 20:391–415
- Han J, Yamana H (2017) A Survey on recommendation methods beyond accuracy. *IEICE Trans Inf Syst* 100:2931–2944
- Hart W, Albarracín D, Eagly AH et al (2009) Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychol Bull* 135:555
- He X, Liao L, Zhang H, et al (2017) Neural Collaborative Filtering. In: Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp 173–182
- He X, Du X, Wang X, et al (2018) Outer product-based neural collaborative filtering. arXiv preprint arXiv:180803912
- Helberger N (2019) On the democratic role of news recommenders. *Digit J* 7(8):993–1012
- Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems. *ACM Trans Inf Syst* 22:5–53
- Hidasi B, Karatzoglou A, Baltrunas L, Tikk D (2016) Session-based recommendations with recurrent neural networks. arXiv:151106939 [cs]
- Horne B (2020) NELA-GT-2019
- Hu L, Cao J, Xu G, et al (2014) Deep modeling of group preferences for group-based recommendation. In: Twenty-Eighth AAAI Conference on Artificial Intelligence
- Ilievski I, Roy S (2013) Personalized news recommendation based on implicit feedback. In: Proceedings of the 2013 international news recommender systems workshop and challenge. ACM, pp 10–15
- Jenders M, Lindhauer T, Kasneci G et al (2015) A serendipity model for news recommendation. In: Hölldobler S, Peñaloza R, Rudolph S (eds) KI 2015: Advances in artificial intelligence. Springer International Publishing, Cham, pp 111–123
- Jonnalagedda N, Gauch S (2013) Personalized news recommendation using twitter. In: 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). pp 21–25
- Jonnalagedda N, Gauch S, Labille K, Alfarhood S (2016) Incorporating popularity in a personalized news recommender system. *Peer J Comput Sci* 2:e63
- Jugovac M, Jannach D, Karimi M (2018) Streamingrec: A Framework for Benchmarking Stream-based News Recommenders. In: Proceedings of the 12th ACM Conference on Recommender Systems. ACM, NY, USA, pp 269–273
- Kang B, Hollerer T, O'Donovan J (2015) The full story: Automatic detection of unique news content in microblogs. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). pp 1192–1199
- Karatzoglou A, Hidasi B, Tikk D, et al (2016) RecSys' 16 Workshop on Deep Learning for Recommender Systems (DLRS). In: Proceedings of the 10th ACM Conference on Recommender Systems. pp 415–416
- Karimi M, Jannach D, Jugovac M (2018) News recommender systems—Survey and roads ahead. *Inf Process Manag* 54:1203–1227

- Karlsen R, Andersen A (2019) Recommendations with a nudge. *Technologies* 7:45
- Karwa BD (2015) A survey on various techniques of personalized news recommendation system. *Int J Sci Adv Res Technol (IJSART)* 1:7
- Khattar D, Kumar V, Varma V (2017) Leveraging moderate user data for news recommendation. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW). pp 757–760
- Kille B, Lommatzsch A, Hopfgartner F, et al (2017) CLEF 2017 NewsREEL overview: offline and online evaluation of stream-based news recommender systems. In: *Proceedings of the CEUR Workshop, 2017*, pp. 1–14
- Kille B, Hopfgartner F, Brodt T, Heintz T (2013) The plista dataset. In: *Proceedings of the 2013 International News Recommender Systems Workshop and Challenge*. ACM, NY, USA, pp 16–23
- Kitaev N, Kaiser Ł, Levskaya A (2020) Reformer: the efficient transformer. arXiv:200104451 [cs, stat]
- Knijnenburg BP, Willemsen MC, Gantner Z et al (2012) Explaining the user experience of recommender systems. *User Model User-Adap Inter* 22:441–504
- Konstan JA, Riedl J (2012) Recommender systems: from algorithms to user experience. *User Model User-Adap Inter* 22:101–123
- Koren Y, Bell R, Volinsky C (2009) Matrix factorization techniques for recommender systems. *Computer* 42:30–37
- Kotkov D, Wang S, Veijalainen J (2016) A survey of serendipity in recommender systems. *Knowl-Based Syst* 111:180–192
- Kumar V, Khattar D, Gupta S, et al (2017) Deep neural architecture for news recommendation. In: *CLEF (Working Notes)*
- Kunaver M, Porl T (2017) Diversity in recommender systems a survey. *Know-Based Syst* 123:154–162
- Lee D, Oh B, Seo S, Lee K-H (2020) News Recommendation with Topic-Enriched Knowledge Graphs. In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. pp 695–704
- Li M, Wang L (2019) A survey on personalized news recommendation technology. *IEEE Access* 7:145861–145879
- Li L, Zheng L, Yang F, Li T (2014) Modeling and broadening temporal user interest in personalized news recommendation. *Expert Syst Appl* 41:3168–3177
- Li L, Li T (2013) News recommendation via hypergraph learning: encapsulation of user behavior and news content. In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*. ACM, NY, USA, pp 305–314
- Lian J, Zhang F, Xie X, Sun G (2018) Towards better representation learning for personalized news recommendation: a multi-channel deep fusion approach. In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, pp 3805–3811
- Lin C, Xie R, Li L, et al (2012) PRemISE: personalized news recommendation via implicit social experts. In: *Proceedings of the 21st ACM international conference on Information and knowledge management - CIKM '12*. ACM Press, Maui, Hawaii, USA, p 1607
- Lommatzsch A, Kille B, Albayrak S (2017) Incorporating context and trends in news recommender systems. In: *Proceedings of the International Conference on Web Intelligence*. ACM, NY, USA, pp 1062–1068
- Lu Z, Dou Z, Lian J, et al (2015) Content-based collaborative filtering for news topic recommendation. In: *Twenty-Ninth AAAI Conference on Artificial Intelligence*
- Ma H, Yang H, Lyu MR, King I (2008) Sorec: social recommendation using probabilistic matrix factorization. In: *Proceedings of the 17th ACM conference on Information and knowledge management*. pp 931–940
- Ma H, Liu X, Shen Z (2016) User fatigue in online news recommendation. In: *WWW*
- Maksai A, Garcin F, Faltings B (2015) Predicting online performance of news recommender systems through richer evaluation metrics. In: *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, NY, USA, pp 179–186
- McCullagh P (2019) Generalized linear models
- Mnih A, Salakhutdinov RR (2007) Probabilistic matrix factorization. *Adv Neural Inf Process Syst* 20:1257–1264
- Mohallick I, Özgöbek Ö (2017) Exploring privacy concerns in news recommender systems. In: *Proceedings of the International Conference on Web Intelligence*. ACM, NY, USA, pp 1054–1061
- Möller J, Trilling D, Helberger N, van Es B (2018) Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Inf Commun Soc* 21:959–977

- Muralidhar N, Rangwala H, Han ES (2015) Recommending temporally relevant news content from implicit feedback data. In: 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI). pp 689–696
- Nakamura K, Levy S, Wang WY (2019) r/fakeddit: a new multimodal benchmark dataset for fine-grained fake news detection. arXiv preprint arXiv:191103854
- Nema P, Khapra M, Laha A, Ravindran B (2018) Diversity driven attention model for query-based abstractive summarization. arXiv:170408300 [cs]
- Newman N, Fletcher R, Levy DA, Nielsen RK (2018) The Reuters institute digital news report 2018. Reuters Institute for the Study of Journalism
- Nguyen TT, Hui P-M, Harper FM, et al (2014) Exploring the filter bubble: the effect of using recommender systems on content diversity. In: Proceedings of the 23rd international conference on World wide web - WWW '14. ACM Press, Seoul, Korea, pp 677–686
- Nørregaard J, Horne BD, Adali S (2019) NELA-GT-2018
- Oh KJ, Lee WJ, Lim CG, Choi HJ (2014) Personalized news recommendation using classified keywords to capture user preference. In: 16th International Conference on Advanced Communication Technology. pp 1283–1287
- Okura S, Tagami Y, Tajima A (2016) Article de-duplication using distributed representations. In: Proceedings of the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp 87–88
- Okura S, Tagami Y, Ono S, Tajima A (2017) 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. ACM Press, Halifax, NS, Canada, pp 1933–1942
- Page X, Wisniewski P, Knijnenburg BP, Namara M (2018) Social media's have-nots: an era of social disenfranchisement. Internet Res 28:1253–1274
- Pariser E (2011) The filter bubble: How the new personalized web is changing what we read and how we think. Penguin
- Park Y, Oh J, Yu H (2017b) RecTime: real-time recommender system for online broadcasting. Inf Sci 409–410:1–16
- Park S, Kang S, Chung S, Song J (2009) NewsCube: delivering multiple aspects of news to mitigate media bias. In: Proceedings of the 27th international conference on Human factors in computing systems - CHI 09. ACM Press, Boston, MA, USA, p 443
- Park K, Lee J, Choi J (2017a) deep neural networks for news recommendations. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, NY, USA, pp 2255–2258
- Pravesh S, Padmanabhan B (2012) Probabilistic news recommender systems with feedback. In: Proceedings of the Sixth ACM Conference on Recommender Systems. ACM, NY, USA, pp 257–260
- Qin J, Lu P (2020) Application of news features in news recommendation methods: a survey. In: International Conference of Pioneering Computer Scientists, Engineers and Educators. Springer, pp 113–125
- Quattrociocchi W, Scala A, Sunstein CR (2016) Echo chambers on facebook. Available at SSRN 2795110
- Rao J, Jia A, Feng Y, Zhao D (2013) Taxonomy based personalized news recommendation: novelty and diversity. In: International Conference on Web Information Systems Engineering. Springer, pp 209–218
- Raza S, Ding C (2019) News recommender system considering temporal dynamics and news taxonomy. In: 2019 IEEE International Conference on Big Data (Big Data). IEEE, pp 920–929
- Raza S, Ding C (2020) A regularized model to trade-off between accuracy and diversity in a news recommender System. In: Proceedings of the 2019 IEEE International Conference on Big Data, pp. 551–560
- Rendle S, Freudenthaler C, Gantner Z, Schmidt-Thieme L (2012) BPR: bayesian personalized ranking from implicit feedback. arXiv preprint arXiv:12052618
- Resnick P, Garrett RK, Kriplean T, et al (2013) Bursting Your (Filter) Bubble: strategies for promoting diverse exposure. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion. ACM, NY, USA, pp 95–100
- Rich E (1979) User modeling via stereotypes. Cogn Sci 3:329–354
- Rizos G, Papadopoulos S, Kompatsiaris Y (2016) Predicting news popularity by mining online discussions. In: Proceedings of the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, pp 737–742
- Robindro K, Nilakanta K, Naorem D, Singh NG (2017) An unsupervised content based news personalization using geolocation information. In: 2017 International Conference on Computing, Communication and Automation (ICCCA). pp 128–132

- Salakhutdinov R, Mnih A, Hinton G (2007) Restricted Boltzmann machines for collaborative filtering. In: Proceedings of the 24th international conference on Machine learning - ICML '07. ACM Press, Corvallis, Oregon, pp 791–798
- Saranya KG, Sudha Sadasivam G (2017) Personalized news article recommendation with novelty using collaborative filtering based rough set theory. *Mobile Netw Appl* 22:719–729
- Scarselli F, Gori M, Tsoi AC et al (2008) The graph neural network model. *IEEE Trans Neural Netw* 20:61–80
- Schedl M, Zamani H, Chen C-W et al (2018) Current challenges and visions in music recommender systems research. *Int J Multimed Inf Retr* 7:95–116
- Scriminaci M, Lommatzsch A, Kille B, et al (2016) Idomaar: a framework for multi-dimensional benchmarking of recommender algorithms
- Sheu H-S, Li S (2020) Context-aware graph embedding for session-based news recommendation. In: Fourteenth ACM conference on recommender systems. pp 657–662
- Shoemaker PJ (2006) News and newsworthiness: a commentary. *Communications* 31:105–111
- Shu K, Mahudeswaran D, Wang S, et al (2018) FakeNewsNet: a data repository with news content, social context and dynamic information for studying fake news on social media. arXiv preprint arXiv:180901286
- Shu K, Wang S, Liu H (2019) Beyond news contents: the role of social context for fake news detection. In: Proceedings of the twelfth ACM international conference on web search and data mining. pp 312–320
- Silveira T, Zhang M, Lin X et al (2019) How good your recommender system is? A survey on evaluations in recommendation. *Int J Mach Learn Cyber* 10:813–831
- Song Y, Elkahky AM, He X (2016) Multi-rate deep learning for temporal recommendation. In: Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval - SIGIR '16. ACM Press, Pisa, Italy, pp 909–912
- Sottocornola G, Symeonidis P, Zanker M (2018) Session-based news recommendations. In: Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18. ACM Press, Lyon, France, pp 1395–1399
- Srihari R (2015) Amazon and the age of personalised marketing
- Su X, Özgöbek Ö, Gulla JA, et al (2016) Interactive mobile news recommender system: a preliminary study of usability factors. In: 2016 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP). pp 71–76
- Sun F, Liu J, Wu J, et al (2019) BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. pp 1441–1450
- Trevisiol M, Aiello LM, Schifanella R, Jaimes A (2014) Cold-start news recommendation with domain-dependent browse graph. In: Proceedings of the 8th ACM Conference on Recommender systems - RecSys '14. ACM Press, Foster City, Silicon Valley, California, USA, pp 81–88
- van der Heijden J, Kusters M (2015) From mechanism to virtue: evaluating nudge-theory. *Social Science Research Network*, Rochester, NY
- Vargas S, Castells P (2011) Rank and relevance in novelty and diversity metrics for recommender systems. In: Proceedings of the Fifth ACM Conference on Recommender Systems. ACM, NY, USA, pp 109–116
- Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. *Adv Neural Inf Process Syst* 5998–6008
- Viana P, Soares M (2016) A hybrid recommendation system for news in a mobile environment. In: Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics. ACM, NY, USA, p 3:1–3:9
- Wang Z, Hahn K, Kim Y et al (2018b) A news-topic recommender system based on keywords extraction. *Multimed Tools Appl* 77:4339–4353
- Wang F, Wu Y (2015) Sentiment-bearing new words mining: exploiting emoticons and latent polarities. In: Gelbukh A (ed) *Computational linguistics and intelligent text processing*. Springer International Publishing, New York, pp 166–179
- Wang S, Zou B, Li C, et al (2015) CROWN: a context-aware recommender for web news. In: 2015 IEEE 31st International Conference on Data Engineering. IEEE, pp 1420–1423
- Wang X, Yu L, Ren K, et al (2017) Dynamic attention deep model for article recommendation by learning human editors' demonstration. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17. ACM Press, Halifax, NS, Canada, pp 2051–2059

- Wang H, Zhang F, Xie X, Guo M (2018a) DKN: deep knowledge-aware network for news recommendation. In: Proceedings of the 2018 World Wide Web Conference. International World Wide Web Conferences Steering Committee, pp 1835–1844
- Wang WY (2017) “liar, liar pants on fire”: a new benchmark dataset for fake news detection. arXiv preprint arXiv:170500648
- Wu Y, DuBois C, Zheng AX, Ester M (2016) Collaborative denoising auto-encoders for top-n recommender systems. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining - WSDM '16. ACM Press, San Francisco, California, USA, pp 153–162
- Wu C, Wu F, An M, et al (2019a) NPA: neural news recommendation with personalized attention. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19. ACM Press, Anchorage, AK, USA, pp 2576–2584
- Wu C, Wu F, Ge S, et al (2019b) 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). Association for Computational Linguistics, Hong Kong, China, pp 6388–6393
- Wu C, Wu F, Qi T, et al (2019c) Reviews meet graphs: enhancing user and item representations for recommendation with hierarchical attentive graph neural network. 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). Association for Computational Linguistics, Hong Kong, China, pp 4883–4892
- Wu C, Wu F, Qi T, Huang Y (2020a) SentiRec: sentiment diversity-aware neural news recommendation. In: Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing, pp 44–53
- Wu F, Qiao Y, Chen J-H, et al (2020b) Mind: a large-scale dataset for news recommendation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp 3597–3606
- Wu Z, Pan S, Chen F, et al (2020c) A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*
- Wu C, Wu F, Yu Y, et al (2021) NewsBERT: distilling pre-trained language model for intelligent news application. arXiv preprint arXiv:210204887
- Wu F, Qiao Y, Chen J-H, et al MIND: a large-scale dataset for news recommendation. 10
- Xia C, Jiang X, Liu S, et al (2010) Dynamic item-based recommendation algorithm with time decay. In: 2010 Sixth International Conference on Natural Computation, pp 242–247
- Xia Z, Xu S, Liu N, Zhao Z (2014) Hot news recommendation system from heterogeneous websites based on bayesian model. *Sci World J*
- Xiao Y, Ai P, Hsu C et al (2015) Time-ordered collaborative filtering for news recommendation. *China Commun* 12:53–62
- Xue H-J, Dai X, Zhang J, et al (2017) Deep matrix factorization models for recommender systems. In: IJCAI. Melbourne, Australia, pp 3203–3209
- Yan X, Guo J, Liu S, et al (2012) Clustering short text using ncut-weighted non-negative matrix factorization. In: Proceedings of the 21st ACM international conference on Information and knowledge management, pp 2259–2262
- Yang J, Wan J, Wang Y, Mao Y (2020) Social network-based news recommendation with knowledge graph. In: 2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA). IEEE, pp 1255–1260
- Yu B, Shao J, Cheng Q, et al (2018) Multi-source news recommender system based on convolutional neural networks. In: Proceedings of the 3rd International Conference on Intelligent Information Processing. ACM, pp 17–23
- Zhang L, Liu P, Gulla JA (2019) Dynamic attention-integrated neural network for session-based news recommendation. *Mach Learn* 108:1851–1875
- Zheng G, Zhang F, Zheng Z, et al (2018) DRN: a deep reinforcement learning framework for news recommendation. In: Proceedings of the 2018 World Wide Web Conference. International World Wide Web Conferences Steering Committee, pp 167–176
- Zhu Q, Zhou X, Song Z et al (2019) DAN: deep attention neural network for news recommendation. *Proc AAAI Conf Artif Intell* 33:5973–5980