

Research Progress of News Recommendation Methods

Jing Qin

¹Northeastern University, Shenyang 110169, China.

Correspondence: annyproj@126.com

Abstract

Due to researchers' aim to study personalized recommendations for different business fields, the summary of recommendation methods in specific fields is of practical significance. News recommendation systems were the earliest research field regarding recommendation systems, and were also the earliest recommendation field to apply the collaborative filtering method. In addition, news is real-time and rich in content, which makes news recommendation methods more challenging than in other fields. Thus, this paper summarizes the research progress regarding news recommendation methods. From 2018 to 2020, developed news recommendation methods were mainly deep learning-based, attention-based, and knowledge graphs-based. As of 2020, there are many news recommendation methods that combine attention mechanisms and knowledge graphs. However, these methods were all developed based on basic methods (the collaborative filtering method, the content-based recommendation method, and a mixed recommendation method combining the two). In order to allow researchers to have a detailed understanding of the development process of news recommendation methods, the news recommendation methods surveyed in this paper, which cover nearly 10 years, are divided into three categories according to the abovementioned basic methods. Firstly, the paper introduces the basic ideas of each category of methods and then summarizes the recommendation methods that are combined with other methods based on each category of methods and according to the time sequence of research results. Finally, this paper also summarizes the challenges confronting news recommendation systems.

1. Introduction

A recommendation system (RS) [1] is a tool that actively pushes services onto users, recommending user preference correlated information to users through the method of "mind-reading", and continuously updating recommended content according to changes to users' preference. A RS not only saves time for users to acquire information but also allows users to explore more potential interests. At the same time, RSs are also considered the most effective means of information exposure. News websites are important channels for people to capture information strictly related to their work and daily life. It is essential to provide the needed information promptly and accurately to the users. Taking the "Yahoo NEWS" website as an example, if a user reads a piece of news titled "South Americans marvel at total solar eclipse," the webpage will automatically show content related to the solar eclipse.



Figure 1: Recommendations after browsing a news item in Yahoo News.

The demonstration in Figure 1 only shows the recommended news content after non-registered users browsed a piece of news. If the users are registered to the website or are logged on to the news website on their mobile devices, they can get more news recommendations satisfying their personal preferences.

Due to different requirements in various business fields, researchers have proposed different recommendation methods for each type of field (e.g., e-commerce [2], travel [3], social networks [4], news [5]). Unlike other business fields, a news portal website has strong real-time performance, fast update speed, and a large quantity of content. Each news item contains a large amount of text content. News recommendation systems first used the user-based collaborative filtering (UserCF) method [6], and then began using content-based (CB) recommendation methods [7], model-based collaborative filtering recommendation methods (e.g., matrix factorization method) [8], and hybrid CB and CF methods (e.g., recommendation methods based on deep learning [9], attention-based methods [10], knowledge graph (KG)-based methods [11]).

To provide researchers with an extensive overview of news recommendation systems in recent years, this review outlines research published from 2011 to 2020, including some representative correlative papers about news recommendation methods. Unlike other reviews of news recommendation methods [5,12-15], this paper summarizes the research progress of news recommendation methods step by step, starting from the basic ideas of the most basic news recommendation methods and following the timeline of the development of recommendation methods.

The summary presented in this paper is structured into six steps: (1) the characteristics of news and the general news recommendation frameworks, (2) the basic idea of CF methods and their research progress regarding news recommendation methods, (3) the basic idea of content-based recommendation methods and the research progress in applying them to news recommendation methods, (4) the basic idea of combined CB and CF recommendation methods and the research progress in applying them to news recommendation methods, (5) evaluation metrics in news recommendation methods, and (6) the research prospects of news recommendation methods.

2. The Characteristics of News and the General News recommendation Framework

2.1 The characteristics of news

Compared with the items recommended by other recommendation systems, such as those for e-commerce, travel, and videos, news is mainly characterized by strong timeliness, rich news content, and proximity.

(1) Strong Timeliness

News articles are a report of recent facts that are of social significance and arouse public interest. Timeliness is reflected in two aspects: recent occurrence and new content. Users are likely to pay more attention to current affairs news items.

(2) Rich News Content

News has a fixed structure, which generally includes five parts: title, lead, subject, background, and conclusion. To facilitate retrieval, editors of some news websites add keywords to news articles. In addition, each news content contains a large amount of topic information, such as sports, medical treatment, and music. The background in each news item also implies the information presented in the news. Therefore, the mining of news content is an important task of news recommendation methods.

(3) Proximity

News items are close to users geographically or in terms of thoughts and interests. The closer the location of the incident is to the user's location, and the more concerned the user is, the higher is the value of the news to the user. The closer the matter is to the vital interests of users, and the more attention users give to it, the higher is the value of news to users.

2.2 The general news recommendation framework

In recent years, many recommendation methods have paid more attention to the analysis of news content due to the large amount of information contained in news content. When selecting news data analysis, not only is the news data itself on the website considered, but the use of auxiliary data (e.g., cross-domain data, knowledge base) to analyze news content is considered, as shown in Figure 2.

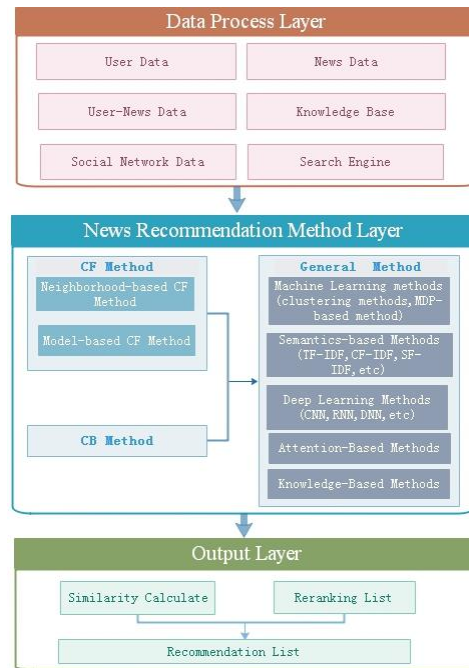


Figure 2: The general news recommendation framework.

The news recommendation framework is mainly divided into three parts: data processing layer, news recommendation model layer and output layer. The data processing layer contains the extracted and correlated data and word embedding process. The general data sources include user data (e.g., user registration information), news data (e.g., data on news websites), user browsing information data (e.g., user comments, browsing records), social network data (e.g., user friend data on social networks), knowledge base (e.g., Wikipedia, Freebase, Baidu Baike), and search engine tools (e.g., Bing search, Google search). The news recommendation model layer mainly uses semantic model, topic model, machine learning model, the deep learning model, and so on. In practical application, one or more models are combined to realize the task of news recommendation. The output layer generates the final recommendation list to the target user by calculating the similarity and re-ranking according to the prediction results of the model. The news vector representation and the user vector representation are also obtained from the learning-based model.

3. The Basic Idea and Progress of CF-Based News Recommendation Methods

This section will introduce the basic principle and recommendation process of the CF method, and then summarize the research progress regarding the CF method in the context of news recommendation.

3.1 The Basic Idea of CF Recommendation Methods

Goldberg [1] raised the concept of CF (i.e., people collaborating to help one another perform filtering by reading their reactions to documents they read) in 1992, and the idea of CF is applied in the mail filtering system (Tapestry). Reference [6] proposed user-based collaborative filtering (UserCF) method, which was applied in Netnews, and this was the first news recommendation system. The UserCF method is mainly used to predict users' rating of unread news in [6]. The UserCF method's rating prediction consists of three steps: (1) generating a user-item matrix using user rating data, (2) calculating the user similarity by Pearson correlation coefficients (Equation 1), and (3) predicting the user's rating for unread news by user similarity.

$$sim(u,v) = \frac{\sum_{i \in I} (R_{u,i} - \bar{R}_u)(R_{v,i} - \bar{R}_v)}{\sqrt{\sum_{i \in I} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{i \in I} (R_{v,i} - \bar{R}_v)^2}} \quad (1)$$

where $R_{u,i}$ denotes the rating of user u for item i , $R_{v,i}$ denotes the rating of user v for item i , \bar{R}_u is the average rating of the u -th users, and \bar{R}_v is the average rating of the v -th users.

Reference [16] proposed the item-based collaborative filtering (ItemCF) method. The ItemCF method is similar to UserCF, but ItemCF calculates the similarity between items. Reference [16] used three similarity computation methods (pearson correlation coefficients (Equation 2)), cosine similarity (Equation 3), and adjusted cosine similarity (Equation 4)).

$$sim(i,j) = cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| \|\vec{j}\|} \quad (2)$$

where "." denotes the dot-product of the two vectors (\vec{i} and \vec{j}). The adjusted cosine similarity formula is shown below.

$$sim(i,j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_u)^2}} \quad (3)$$

where \bar{R}_u denotes the average of the u -th user's ratings. Pearson correlation similarity is given by

$$sim(i,j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2} \sqrt{\sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}} \quad (4)$$

where $R_{u,i}$ denotes the rating of user u for item i , and \bar{R}_i is the average rating of the i -th items. $R_{u,j}$ denotes the rating of user u for item j , and \bar{R}_j is the average rating of the j -th items.

Three kinds of similarity calculation method were compared, and the results show the mean absolute error (MAE) metric of adjusted cosine-based similarity is the lowest in [16]. However, it is still necessary to choose the similarity calculation method according to the specific dataset.

CF methods are divided into memory-based CF methods and model-based CF methods [17]. UserCF belongs to memory-based CF, and ItemCF belongs to model-based CF [14]. Besides ItemCF, model-based CF includes many methods, such as latent factor model (LFM) and machine learning-based methods. UserCF and ItemCF also are considered as neighborhood-based method [18]. The purpose of neighborhood-based methods is to calculate the relationship between the user or item [18]. Reference [18] also considers the LFM as another important model-based CF method. The purpose of the LFM is to connect users' interests and items through latent factors.

Reference [18] proposed a matrix factorization (MF) based recommendation method, and MF belongs to the model-based CF methods. The basic idea of the MF model is to decompose the user-item rating matrix into user latent factor vector and item latent factor vector in same latent factor space of dimensionality. Assume \mathbf{R}^f denotes the latent factor space of dimensionality f , and \mathbf{q}_i denotes the factor vector of item i , and \mathbf{p}_u is the factor vector of user u , and $\mathbf{q}_i, \mathbf{p}_u \in \mathbf{R}^f$. The basic MF model is shown in Equation (5)

$$\hat{r}_{ui} = \mathbf{q}_i^T \mathbf{p}_u \quad (5)$$

where, \hat{r}_{ui} denotes user u 's rating of item i . The challenge of the MF model is to obtain factor vectors $(\mathbf{q}_i, \mathbf{p}_u)$ by a learning method [18] (e.g., (stochastic gradient descent, SGD), (alternating least squares, ALS)).

The main advantages and disadvantages of collaborative filtering (CF) methods are shown in Table 1.

Table 1. The main advantages and disadvantages of CF methods

| Method | Main Advantages | Main Disadvantages |
|--------|--------------------------------------|--|
| UserCF | 1. Fit for fewer users | 1. Does not consider the context of the item |
| | 2. Results are diverse | 2. Cold-start problem |
| ItemCF | 1. Fit for fewer items | 3. Results lack interpretability |
| | 2. Results are interpretable | 1. Does not consider the context of the item |
| MF | 1. Score prediction is more accurate | 2. Results lack diversity |
| | 2. Low space complexity | 3. Cold-start problem |
| | | 1 Does not consider the context of the item |
| | | 2. Cold-start problem |
| | | 3. Results lack interpretability |

As can be seen from the table, each CF method has its advantages, but in the news field, due to the sparse data and fast news update speed, the direct use of the CF method for realizing news recommendation can no longer meet the needs of modern news recommendation systems. Thus, in recent years, many improved methods were developed based on CF methods, especially using model-based CF methods.

3.2 Progress of CF-Based News Recommendation Methods

CF methods can recommend items for users only according to the user's scoring data. Due to the scarcity of users' scoring data in the context of news, only using CF methods will lead to low recommendation accuracy. Thus, in recent years, in order to improve the recommendation accuracy of CF methods, researchers have proposed a large number of hybrid methods based on CF methods. The particular methods include clustering-based methods [8, 19, 20, 21] and the Markov decision process (MDP)-based method [22].

3.2.1 Combining with clustering-based methods

The clustering method combined with CF includes news clustering and user clustering. The clustering method for news also takes into account the representation of news content. In order to introduce the integrity of the method, all news recommendation methods related to clustering methods are introduced in this section.

Reference [8] used two kinds of CF methods (memory-based CF method and model-based CF method). The memory-based CF method uses items' co-visitation count. The model-based CF method uses clustering-based method ((probabilistic latent semantic indexing, PLSI) [23] and MinHash). The results of user clustering and the statistical results of news are stored in the user table (UT) and Story Table (ST) respectively. The UT and ST are stored in the Bigtable infrastructure [24]. The recommendation system contains three components: the news statics server (NSS), the news personalization server (NPS) and the news front end (NFE). The NSS is used for updating statistics in the ST. The NPS is used for generating news articles recommendations. The NFE is responsible for obtaining user requests (recommend requests and update statics requests), and feeding back the results of user requests. When the NFE receives the recommend request, the process of recommendation is as shown in Figure 3.

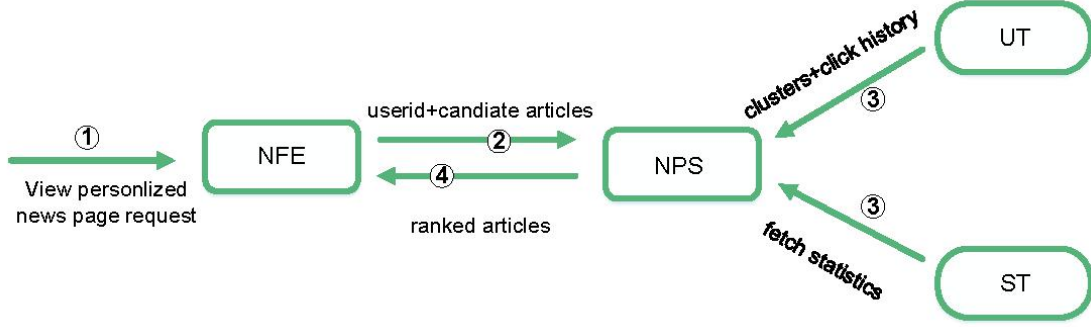


Figure 3: The process of recommendation (the NFE receives the recommend request) [8].

Reference [19] proposed the scalable two-stage personalized news recommendation system (SCENE). The recommendation framework of SCENE contains three components (news articles clustering, user profile construction, and personalized news recommendation). The news articles clustering component using locality sensitive hashing (LSH) [25] and hierarchical clustering to generate news groups. The user profile construction component combines accessed news content, similar access patterns and preferred name entities according to the user's history. The personalized news recommendation component compares the similarities between each small news group and the user's accessed news content, then recommends top-k news items to target users. Reference [19] compared latent dirichlet allocation (LDA) [26] and PLSI for topic detection. LDA is suitable for a small news corpus, and PLSI is suitable for a larger news corpus.

Reference [20] proposed a combine CF recommendation method with the user and news article cluster method based on WordNet-enabled k-means (W-kmeans) [27]. W-kmeans [27] uses the WordNet [28] database to cluster news articles in the user session. Reference [20] proposed a recommendation architecture containing three components (session identification, clustering user sessions, and recommendation stage). The advantage of the proposed method is the use of word hypernyms with WordNet and clustering users by session instead of user's browsing history. The proposed method is better than the CF methods of latent semantic CF, and neighbor-based CF.

Reference [21] proposed cluster-based CF recommendation methods. The recommendation methods focused on the news trends, and user preferences immediately changed. The recommendation architecture consists of three components (user modeling, clustering and evaluation, and personalized list generation). The user modeling component is used to calculate the user's vector representation using the word2vec [29] model and the user's vector representation is updated by users clicking on a news article. The clustering and evaluation model is used for user k-means clustering and calculate the click through rate (CTR). The personalized list generation model is used to generate a personalized news article list in real-time according to the user's request. In order to obtain fresh news articles, the paper proposed time decay function (TDF) and user time decay function (UTDF). The TDF calculates the scores of time decay according to the article's date of publication, and the UTDF calculates the scores of time decay according to the user's last access time. The study adopted mean average precision (MAP) and normalized discounted cumulative gain (NDCG) metrics [30] to evaluate the proposed method in offline experiments and online experiments. In the offline experiment, the performance of the proposed method was better than the compared clustering

method (MinHash, Non-negative Matrix Factorization (NMF)) in the paper. In the online experiment, the performance of the UTDF was better than the TDF in CTR metric.

A summary of cluster-based CF methods is shown in Table 2.

Table 2. A summary of cluster-based CF news recommendation methods

| Ref. | Methods | Main Contribution |
|------|---------------------------------|---|
| [8] | MinHash, PLSI | 1. fit for the massive scale of the datasets 2. support clustering over dynamic datasets |
| [19] | LSH, Hierarchical clustering | 1. use the name entities and news recency 2. support clustering on newly-published news articles |
| [20] | w-kmeans | 1. consider the word hypernyms with WordNet 2. clustering on session data |
| [21] | k-means | 1. fit for the massive scale of the datasets 2. fit for immediate news recommendation |

The combination of the CF method and the clustering method enhances the recommendation performance. However, there are some shortcomings to this combination, as follows: (1) the timeliness of news is not considered in clustering; (2) the context information of users is not considered; (3) It is hard to use clustering method for data with uneven categories; (4) the diversity of recommended news is only considered by the authors of [19], while the diversity is not considered in other methods.

3.2.2 Combining with the MDP-based method

Reference [22] proposed an ItemCF with MDP method. The MDP method suitable for recommendation of news articles. Unlike ItemCF, the proposed method measures similarity of news items by semantic similarity [31]. The MDP uses four tuple representation: $\langle \mathcal{S}, \mathcal{A}, \mathbf{R}_{wd}, \mathbf{TP} \rangle$, where \mathcal{S} denotes a set of states (i.e., each read news article as a state), \mathcal{A} denotes a set of actions, \mathbf{R}_{wd} denotes the reward function, and \mathbf{TP} denotes the transition probability between states. This study defines three reward functions as follows:

$$Rwd_1(A_i, n_j, n_k) = \begin{cases} \frac{dt(n_k) - dt(n_j)}{N}, & trans_i(n_j, n_k) > 0 \\ 0, & otherwise \end{cases} \quad (6)$$

where, $trans_i(n_j, n_k)$ denotes the number of times $i-1$ articles were read from n_j to n_k ; $dt(n_k)$ and $dt(n_j)$ denotes the discovery time for news article n_k and n_j respectively.

$$Rwd_2(A_i, n_j, n_k) = \begin{cases} sim(n_j, n_k), & trans_i(n_j, n_k) > 0 \\ 0, & otherwise \end{cases} \quad (7)$$

where, $sim(n_j, n_k)$ denotes the semantic similarity measure [31].

$$Rwd_3(A_i, n_j, n_k) = \alpha * Rwd_1(A_i, n_j, n_k) + (1 - \alpha) * Rwd_2(A_i, n_j, n_k) \quad (8)$$

where, α is a hyperparameter (α is 0.7 is best in this experiment).

In this experiment, the proposed MDP methods are better than the other compared method (RegSVD [32], bigram [33], KNN [16]). Also, the performance of the **Rwd₃** was better than the **Rwd₁** and **Rwd₂**. However, when applying the MDP method, it is only related to the preamble state selected by the user, ignoring the complete history of the user, and at the same time, it does not consider the diversity of recommendation results.

4. The Basic Idea and Progress of News Recommendation Methods Based on CB

The content-based recommendation method was the important recommendation method in recommendation system research. According to currently retrieved information, news recommendation system Newsweeder [7] was the first to use CB recommendation methods in 1995. CB recommendation methods are divided into text content statistics-based methods, semantics statistics-based methods, and methods use a mixture of both methods. There are fewer recommendation systems that only use semantics statistics-based methods. Text content statistics-based news recommendation systems include News Dude [34], Newsjunkie [35], YourNews [36]. News recommendation systems based on the above mixed methods include OntoSeek [37], News@hand [38], etc.

4.1 The Basic Idea of CB-based methods

Text content statistics-based news recommendation methods often use vector space model (VSM) [39] to express keywords in every document. According to the news previously selected by users, this method can obtain users' unread news and similar news through calculating similarities between previously read news and unread news, and generate a recommendation list according to the result of similarity calculation. G.Salton [39] proposed the VSM expressing method in the 1960s. When news content uses VSM representation, every document is represented by n-dimension' space vectors, i.e., every dimension vector corresponds to a word in the given document's collection word list. Every document's content represents the vector, which is expressed by the keyword's weight. The weight refers to the relevance between the document and the keyword.

The conventional methods of calculating keyword weight include term frequency-inverse document frequency (TF-IDF) [40] and improved TF-IDF methods. The method of TF-IDF also considers inverse document frequency (IDF) [41] while it is calculating term frequency (TF). IDF [41] means the word frequently appears in one document but seldom appears in other documents.

In content-based news recommendation methods, the first step is to profile news data and user data; then to calculate the similarity between the profiles and the news previously selected by the user with the methods of cosine similarity, semantic similarity, hybrid similarity, etc.; and finally to recommend similar top-k news to users. When profiling news and user data, the news characteristics and user characteristics, which contain a large quantity of text, can be expressed as a character vector by the methods of TF-IDF, one-hot, word2vec, doc2vec [42], etc. The general recommendation process is as shown in Figure 4.

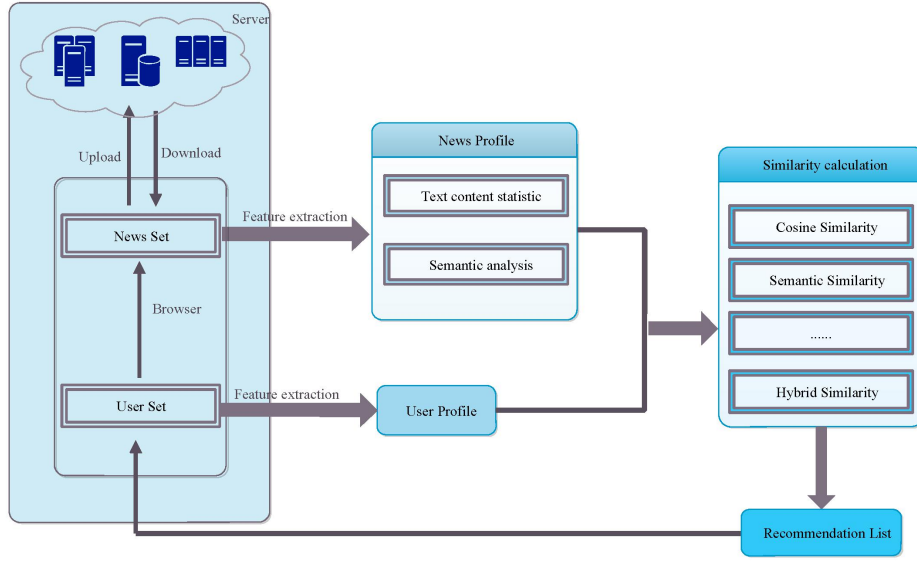


Figure 4: Recommendation process of CB news recommendation methods.

As is shown in Figure 4, CB news recommendation methods consist of two parts. One profiles the news and user, and the other calculates similarity after profiling. Researchers have also researched and improved news recommendation methods for these two aspects.

4.2 The Progress of CB News Recommendation Methods

Regarding content-based news recommendation methods, the research results in recent years have mainly focused on the VSM-based weight calculate methods and social network-based method.

4.2.1 VSM-Based Weight Calculate Methods

The core issue of studies on semantics-based news recommendation methods in recent years is to calculate weight value in VSM [39] expression. On the basis of the commonly used TF-IDF [40] method, many methods have been proposed, such as concept frequency-inverse document frequency (CF-IDF) [43], CF-IDF+ [44], synset frequency-inverse Document Frequency (SF-IDF) [45], Bing-SF-IDF [46], SF-IDF+ [47], Bing-SF-IDF+ [48], Bing-CF-IDF+ [49].

The CF-IDF method uses the concept in news content of replacing the characteristic word calculating weight, and the formula for calculating the CF-IDF weight is as follows.

$$cf - idf_{i,j} = cf_{i,j} \times idf_i \quad (9)$$

where $cf_{i,j}$ denotes the concept frequency, and idf_i denotes the inverse document concept frequency, calculated by the following formulas:

$$cf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (10)$$

where $n_{i,j}$ denotes the number of occurrences of a concept C_i in document D_j , and $n_{k,j}$ denotes the total number of occurrences of all concepts in D_j .

$$idf_i = \log \frac{|D|}{|\{d:c_i \in d\}|} \quad (11)$$

where $|D|$ denotes the total number of documents, and d denotes the number of documents in which the concept c_i appears.

The CF-IDF+ method adds the weight of the semantic relationship between concepts on the CF-IDF method basis. The formula for calculating the CF-IDF+ weight is as follows.

$$cf - idf_{i,j} = cf_{i,j} \times idf_i \times \omega_r \quad (12)$$

where ω_r denotes the weight of the semantic relationship. The CF-IDF+ method identifies three different weights (superclasses, subclasses, and domain relationships).

The Bing CF-IDF+ method adds page counts gathered by Web search engines for specific named entities. The Bing-CF-IDF+ method is a weighed combination of the Bing and the CF-IDF+ similarity measures, and the formula is as follows.

$$sim_{Bing-CF-IDF+}(u,d) = \alpha \times sim_{Bing}(u,d) + (1 - \alpha) \times sim_{CF-IDF+}(u,d) \quad (13)$$

where u denotes the profile of the user u , and d denotes the unread news item d . $sim_{Bing}(u,d)$ denotes Bing similarity measures, and $sim_{CF-IDF+}(u,d)$ denotes **CF-IDF+** similarity measures. $sim_{CF-IDF+}(u,d)$ uses the cosine similarity measure, normalized using a min-max scaling between 0 and 1. For Bing similarity measures, the formula is as follows.

$$sim_{Bing}(u,d) = \frac{\sum_{(u,d) \in V} \log \frac{\frac{p(u,d)}{N}}{\frac{p(u)}{N} \times \frac{p(d)}{N}}}{|V|} \quad (14)$$

where V denotes the set of named entity pairs (cartesian product of named entity set in user profile and named entity set in unread news items), N denotes the total number of web pages, $p(u,d)$ denotes the Bing page count for pair (u,d) , $p(u)$ denotes the page counts for named entities u , and $p(d)$ denotes the page counts for named entities d .

The SF-IDF method is also similar to TF-IDF, as it uses the WordNet synonym sets. The SF-IDF formula is as follows.

$$sf - idf_{i,j} = sf_{i,j} \times idf_i \quad (15)$$

where $sf_{i,j}$ denotes the synset frequency, idf_i denotes the inverse document synset frequency.

SF-IDF+ method adds the weight of semantic relationship between synset on SF-IDF method basis. The formula for calculating the SF-IDF+ weight is as follows.

$$sf - idf_{i,j} = sf_{i,j} \times idf_i \times \omega_r \quad (16)$$

where ω_r denotes the weight of the semantic relationship related synset.

The Bing SF-IDF method adds page counts gathered by Web search engines for specific named entities. The Bing-SF-IDF method is a weighed combination of the Bing and the SF-IDF similarity measures, and the formula is as follows.

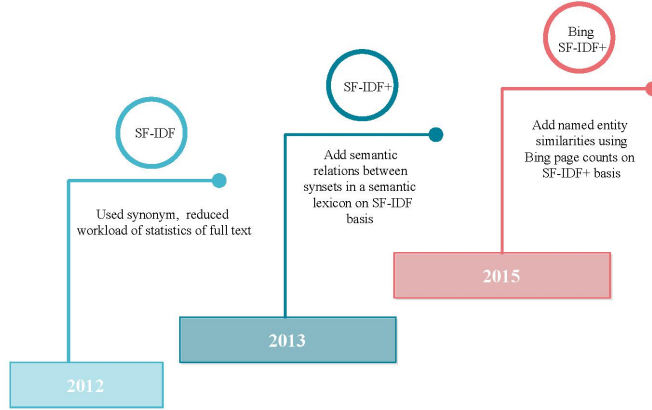
$$sim_{Bing-SF-IDF}(u,d) = \alpha \times sim_{Bing}(u,d) + (1 - \alpha) \times sim_{SF-IDF}(u,d) \quad (17)$$

where u denotes the profile of the user u , and d denotes the unread news item d . $sim_{Bing}(u,d)$ denotes Bing similarity measures, and $sim_{SF-IDF}(u,d)$ denotes SF-IDF similarity measures. $sim_{SF-IDF}(u,d)$ uses the cosine similarity measure, normalized using a min-max scaling between 0 and 1. The Bing-SF-IDF+ method is a weighed combination of the Bing and the SF-IDF+ similarity measures, and the formula as follows.

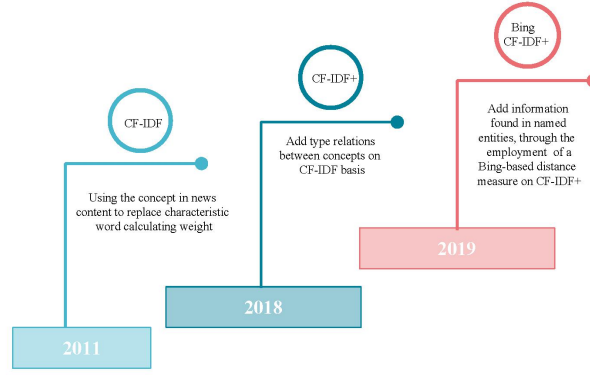
$$sim_{Bing-SF-IDF+}(u,d) = \alpha \times sim_{Bing}(u,d) + (1 - \alpha) \times sim_{SF-IDF+}(u,d) \quad (18)$$

The development of these methods is shown in the Figure 5, in which, Figure. 5a depicts an improved process for the CF-IDF, CF-IDF+, and Bing-CF-IDF+ methods. Figure 5b depicts an improved process for the SF-IDF, SF-IDF+, and Bing-SF-IDF+ methods.

Reference [48] made a comparison of some methods, such as Bing-SF-IDF+, SF-IDF, and TF-IDF. Bing-SF-IDF+ is better than the SF-IDF and TF-IDF methods in terms of F1-Score and Kappa statistics [50]. Reference [49] made a comparison of some methods, such as Bing-CF-IDF+, TF-IDF, CF-IDF and CF-IDF+, which shows that Bing-CF-IDF+ is better than TF-IDF, CF-IDF and CF-IDF+ in terms of F1-score and Kappa statistic.



a



b

Figure 5: The improved process of the CF-IDF, CF-IDF+, Bing CF-IDF+ (a), The improved process of the SF-IDF, SF-IDF+, Bing-SF-IDF+ (b)

However, these methods also have their drawbacks, such as not deleting stop words, not considering the position information of particular words in the article, not considering synonyms, not considering the relationships between words, and a high dependency on lexicons. In addition, the diversity of recommended results has not been considered.

4.2.2 Combining with Social Networks-Based Methods

Hot news is usually related to hot topics on social networks. Twitter is a popular social platform where you can get user location information [51] and topic information from Twitter via its API (<https://api.twitter.com/>). Therefore, in the research of news recommendation methods, the information on Twitter is used to improve the accuracy of news recommendation. Use the user's Twitter information in [52, 53] to model the user profile and then recommend news articles for the user. There are three ways to model users using Twitter data (hashtag-based, topic-based, and entity-based). To verify the effectiveness of using Twitter data to recommend news articles, the authors constructed a data set by Twitter information and news articles. Then verify the influence of three different user modeling methods on the recommendation effect in the data set. Experiments show that the entity-based method has the best effect, but in the short-term recommendation, the topic-based method is better than the entity-based method.

Reference [54] creates a social news service. The social news service is a content-based method. Also, this method uses two content indexes (one from Twitter and one from RSS feeds). Reference [55] uses the popularity of topics on Twitter to sort news. Reference [51] uses the user's location information to find location-related topic information from Twitter. Reference [56] proposes fuzzy logic method for predicting users' interest. According to the user's interest category, virus news articles and their categories are analyzed by mining social media (Facebook and Twitter).

For the convenience of readers, the contents related to the methods of combining with social networks are summarized as follows (see Table 3).

Table 3. A summary of the social networks-based methods news recommendation methods.

| Ref. | Time | Main Methods | Social networks |
|------|------|-------------------------------------|-------------------|
| [52] | 2011 | Topic Model | Twitter |
| [54] | 2011 | Lucene indexes, TF-IDF | Twitter |
| [53] | 2013 | Topic Model | Twitter |
| [55] | 2013 | SOLR (in Apache Lucene project) | Twitter |
| [51] | 2018 | fuzzy clustering, naïve Bayes model | Twitter |
| [56] | 2020 | fuzzy logic method | Twitter, Facebook |

In addition, reference [57] proposed a neural social collaborative ranking (NSCR) approach, which seamlessly links users' information in the information domain with that in social networks. Regarding the information domain, it is used to strengthen users' and items' embedded learning. Regarding the social network domain, it is used to spread learning to non-bridge users through bridge users. Even though the NSCR method has not been tested on news datasets, it can still be applied to news recommendations.

In summary, the use of social networks can alleviate the problem of cold-start in news recommendations and recommend news according to popular trends in the news. However, the auxiliary information provided by social networks for news content analysis is limited, and cannot accurately analyze the entities in the news and the relationships between entities. In addition, in the above-mentioned news recommendation method based on social networks, the diversity of news recommendations and the domain knowledge related to the news are not considered.

5 The Progress of News Recommendation Methods Based on Combine CF and CB

CF method will bring about the problem of cold-start, especially when the data is sparse, the problem of cold-start is more serious. CB method can alleviate the problem of cold-start by analyzing news content. But in the CB method, the recommended results lack diversity (too similar to the history of users). In the news recommendation methods mixed with CB and CF methods, the CB method is used for the representation of news content, and CF methods are used to calculate the similarity of users or news or MF-based modelling. For example, the News Recommendation System "Fab" introduced in 1997 [58].

In the methods mixing CF and CB, the main research results in recent years have focused on the methods of assisting recommendations in combination with search engine-based query methods, with the timeliness of news mining methods, with deep learning, attention mech, and KG. The CF method used in combination with deep learning, attention mechanism, and KG is a model-based method. The KG and attention-based methods were both applied to news recommendation research in 2018, and are also widely used at present.

5.1 Combining with search engine

The search engine is a way of enriching the news content and calculating the news similarity. References [59, 60] use a search engine (enrich the news's content) combined with CF methods. Reference [59] used methods of content-based collaborative filtering (CCF) and extended CCF (CCF+), which calculate the similarities in two document collections with the help of the Fisher Kernel function [61], and enrich search text' implication by using the Bing

search engine. Through the experiment, the root mean square error (RMSE) metric performs better than Bayesian probabilistic matrix factorization (BPMF) method, and an effective content-based filtering method (Content). Reference used user profile as a search query for the search engine. Reference [60] proposed that using a search engine brings three advantages (short response time, fast processing of new content, and limited storage requirements).

Although search engines can quickly retrieve news content, there are still some drawbacks in the mining of news content, such as a lack of event analysis, entity analysis, and synonym analysis in news content. In addition, the reliability of the results found in search engines cannot be guaranteed.

5.2 Combining with the timeliness of news

Because of the timeliness of news content, the news life cycle is given attention in news recommendation methods.

Reference [62] proposed a recommendation method of sports news. The RS has two components, one is CB-based component, and the other is CF-based component. In content-based component, only news articles in three days are recommended, and in the CF-based component, the older news articles are recommended. When analyzing the news article, 5 to 10 keywords can be extracted to represent the news article. The evaluation of results is good in KPI (Key Performance Indicators) and SUS (System Usability Scale).

Reference [63] also proposed a hybrid CB and CF method, namely FeRe. In the CB part of the method, not only the news content but also the popularity and trends of the news are considered. For example, different types of news have different life cycles. Therefore, users have different preferences for different types of news. Some news users do not care when the news occurs but pay more attention to the content. In the CF part of the method, candidate news predictions are scored by calculating the similarity between users. Experiments showed that the recall rate of the FeRe method is better than those of either the CB method or CF method alone.

Reference [64] proposed a session-based recommendation method, using the CF method and the CB method to calculate the similarity of news articles. The CF method uses the session-item matrix, in which the similarity of items is calculated as follows:

$$simCF(t_p, i, j) = \frac{\sum_{s \in s_{t_p, i} \cap s_{t_p, j}} (r_{s, i} \cdot r_{s, j})}{\sqrt{\sum_{s \in s_{t_p, i}} (r_{s, i})^2} \sqrt{\sum_{s \in s_{t_p, j}} (r_{s, j})^2}} \quad (19)$$

where t_p denotes the time period p , each t_p time period represents a session, $s_{t_p, i}$ denotes the session set of item i in the specified time period p , and $s_{t_p, j}$ denotes the session set of item j in the specified time period p . $r_{s, i}$ and $r_{s, j}$ respectively indicate whether items i and j are accessed in the session; if accessed, the result is 1, otherwise it is 0.

The content-based method uses the article category to calculate similarity, as follows:

$$simCB(t_p, i, j) = \frac{\sum_{c \in \mathcal{C}_{t_p, i} \cap \mathcal{C}_{t_p, j}} (f_{c, i} \cdot f_{c, j})}{\sqrt{\sum_{c \in \mathcal{C}_{t_p, i}} (f_{c, i})^2} \sqrt{\sum_{c \in \mathcal{C}_{t_p, j}} (f_{c, j})^2}} \quad (20)$$

where \mathcal{C} denotes the set of topic categories, $f_{c, i}$ and $f_{c, j}$ respectively denote the probability that articles i and j belong to topic category c , and $\mathcal{C}_{t_p, i}$ and $\mathcal{C}_{t_p, j}$ respectively denote the topic category to which article i and j belong in time period t .

After combining Equations (19) and (20), the formula is as follows:

$$simCBCF(t_p, i, j) = \alpha * simCB(t_p, i, j) + (1 - \alpha) * simCF(t_p, i, j) \quad (21)$$

where α represents the weight value, and 0, 0.5, and 1 were used in the experiment. The experimental results showed that the precision rate was the highest at $\alpha = 1$ and the lowest at $\alpha = 0$. However, when the timepoint value was small, i.e., $\alpha = 0.5$, it was similar to the precision rate achieved at $\alpha = 1$. Compared with the method based on the popularity baseline method, the average accuracy of the CBCF method was higher.

Some researchers also consider news popularity as a recommendation factor [65,66]. For example, reference [65] proposed an online recommendation method that combines trends and short-term user habits, takes the life of items into account, and divides news items into three classes (very popular, medium popular, unpopular). Reference [66] proposed a based on most-popular strategies recommend methods online and offline in the CLEF NewsREEL challenge.

In summary, although the accuracy of the recommended results has improved in terms of news popularity and news life cycle, the following problems still exist: (1) the acquisition of users' complete preferences is ignored. For example, only the preferences in the current session are obtained [64], while in other methods, only the long-term preferences of users are obtained while ignoring the short-term preferences of users; (2) the connections between news content are ignored; and (3) the relationship between popular news and news that is important to users is ignored.

5.3 Combining with Deep Learning Methods

Based on the idea of MF, reference [67] firstly proposed neural collaborative filtering method. In 2017, deep learning methods were used in the news recommendation domain for the first time [9]. Compared with machine learning methods, deep learning methods' advantages are stronger learning ability and generalization ability. Since 2017, CF-based news recommendation methods used recurrent neural network (RNN) and convolutional neural network (CNN) models. The RNN model contains bidirectional RNN (Bi-RNN) and Long Short-Term Memory networks (LSTM), this model was used for modeling sequence data. The CNN model contains 1-D CNN, 2-D CNN, and 3-D CNN. The 1-D CNN similar to RNN, it mainly used for modeling sequence data. The 2-D CNN mainly used for modeling picture data, and the 3-D CNN mainly used for modeling video data. CNN also used for modeling the vector representation of news items in news recommendation methods. Reference [9] used two different RNN models: one RNN model was used for modeling based on session data, and the other RNN model was used for modeling based on session and user click history. The study also used a CNN model to obtain users' long-term interest preferences.

Reference [68] proposed a 3-D CNN-based news recommendation method. This model contains three parts (3-D tensor, 3-D convolutions, and score aggregation). The 3-D tensor part uses a three-dimensional structure to denote the similarity tensor. The proposed recommendation framework as follows (see Figure 6).

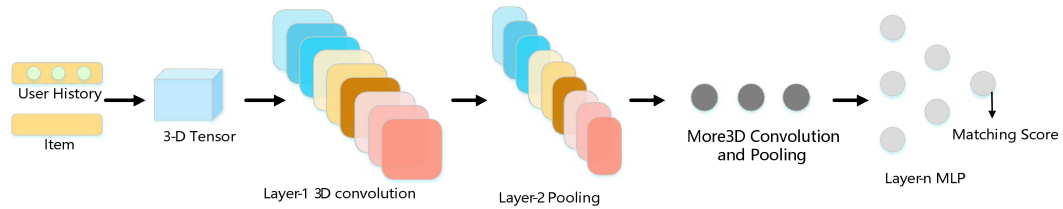


Figure 6: 3-D CNN-based news recommendation framework [68].

Reference [69] also used the 3-D CNN model. The proposed model architecture includes User-History Component and Item Component components. The User-History Component used the 3-D CNN model, and the item Component used the 2-D CNN model. Reference [70] not only uses user browsing information, but also uses the content information of news, and obtains article-embedding through the news information that users browsed. The recommendation model proposed in this paper includes two components (user profile component and article representation component). The user profile component is used to generate vector representation from news headlines and contents through doc2vec [42] embedding. The article representation component uses a method similar to MF, and uses a neural network to establish a model for calculating similarity between user preferences and candidate sets. The logistic function is directly used to generate the recommendation list of the target user during similarity calculation. The model architecture is shown in Figure 7.

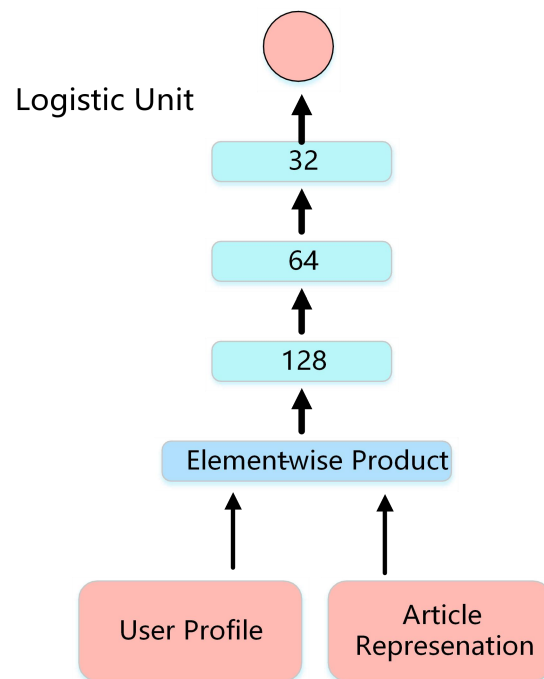


Figure 7: Neural Network Architecture [70].

Reference [71] proposes the hybrid recurrent attention machine (HRAM) recommendation model, which also consists of two components. The first component uses the generalized matrix factorization method to predict the user will give the news articles. The second component is the user history component, which is implemented by an attention-based recurrent network. The model architecture is shown in Figure 8.

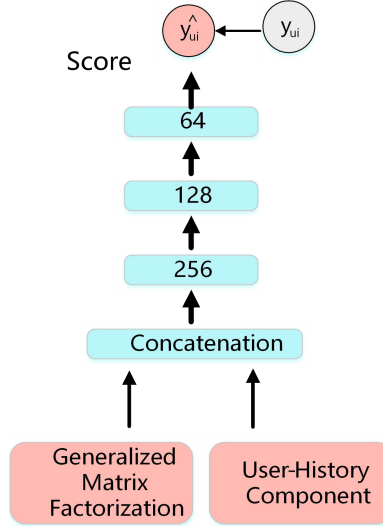


Figure 8: Hybrid Recommendation Attention Machine (HRAM) [71].

In the above-mentioned deep learning method, the traditional ID-based matrix decomposition method is replaced by the point multiplication method based on user characteristics and news characteristics, and a multi-layer neural network is added to make the model result more accurate than the traditional MF method. However, these methods still have some deficiencies: (1) they do not consider the timeliness of news (2) they lack diversity in recommendation results (3) they ignore the fine-grained representation of news content (4) they do not consider the upper and lower information of users.

5.4 Combining with Attention-Based Methods

Attention mechanisms are mainly used to mine the hot spots of users' interest preferences and obtain keywords from news articles. The recommendation methods based on attention mechanisms include one or more attention layers. In the news recommendation model, the news profile aspect uses a CNN, and the user profile aspect uses an RNN and then adds a different attention mechanism layer. The attention mechanism in the news profile aspect extracts the critical words, entities, events, etc. The attention mechanism in the user profile aspect extracts the hotspots of the user's internet.

Reference [10] proposed the deep fusion model (DFM), which can be applied to item retrieval and item re-ranking. The DFM includes representation learning and an attention mechanism. In contrast to cold-start users who use knowledge from other domains, other users directly use the news domain. The model leverages an attention mechanism to alleviate the factors mentioned above and let the model dynamically combine knowledge from different domains. The proposed model performed better than compared models (logistic regression (LR), gradient boosting decision trees (GBDT), factorization machine (FM), SVDFeature, DeepFM, YoutubeRank [72], and DeepWide [73]) in terms of area under the curve (AUC) [74] and Logloss.

Reference [75] proposed the deep attention neural network (DAN) model, which uses the two attention-based models. The attention-based RNN (ARNN) model is used to find the sequential features of users' clicks, and the attention-based neural network (ANN) model is used to discover features of user's interest. The DAN model includes three models (PCNN, ANN, and ARNN). The PCNN is used as a news representation extractor (combining CNN-based title embedding and the CNN-based entity of news embedding on the PCNN model), the ANN is used to extract user interest, and the ARNN is used as a sequential information extractor (with the LSTM model). Notably, the DAN model enriched the presentation of news articles compared to other methods (LibFM, DSSM [76], DeepWide, DeepFM, DMF, DKN [11]) on Adressa dataset [77] in the study.

Reference [78] proposed a neural approach with personalized attention (NPA) model. The NPA model contains three modules (news encoder module, user encoder module, and click prediction module). The news encoder module aims to express the news. The news encoder module contains three sub-modules (word embedding, contextual representations, and mining the import words using attention mechanisms). The user encoder module is used to learn the representations of users from the representations of their clicked news. The click predictor module is used to predict target users' click scores on each candidate news item. The proposed method performed better than compared methods (LibFM; CNN; DSSM; WideDeep; DeepFM; DFM; DKN) in terms of AUC, mean reciprocal rank (MRR), and NDCG metrics.

Reference [79] proposed a topic-aware news encoder, and a user encoder in the news encoder used a CNN and an attention network for representations of news titles. The news encoder contains three layers (word embedding expression layer, CNN network layer, and attention mechanism layer). The word embedding expression layer converts a news title from a word sequence into a vector sequence. The CNN network layer uses the CNN to learn contextual word representations by capturing local contexts. The attention mechanism layer selects important words in news titles to learn informative news representations. The user encoder uses the user's browser history and the attention network as a representation of the user. Also, this paper added a news topic classification model to predict the topic of news. The proposed method performed better than the compared methods (LibFM; CNN; DSSM; WideDeep; DeepFM; DFM (Deep Fusion Model); DKN; GRU) in terms of the AUC, MRR, and NDCG metrics.

The attention mechanism generates news vector representations more comprehensively than other content-based methods and can represent news in a fine-grained manner based on the word level and can better tap users' interests. However, the abovementioned news recommendation method based on attention mechanism still has the following problems: (1) ignoring the loss of information between each attention layer; (2) ignoring the differences between long-term and short-term preferences of users; (3) not considering the relationship between entities in news, which affects the accuracy of news recommendations; (4) not considering the popularity and timeliness of news; (5) not considering the context information of users, which affects the reading experience of users. Similarly, the diversity of recommendation results is not considered in the attention mechanism method.

5.5 Combining with Knowledge Graph-Based Methods

KG was proposed by Google in 2012. It was first applied to the field of news recommendation in 2018 and has also been a research hotspot in recent years [11]. KG is a

method that represents knowledge by using a graph model. It is divided into general KG (FreeBase [80], Yago [81]), and domain KG (News Graph [82]). Many methods of using KG also combine attention mechanisms [82,83,84], thus being better suited for news vector representation.

Reference [11] proposes the deep knowledge-aware network (DKN) framework that brings a knowledge graph into news recommendation and adds an attention module. The DKN framework is a content-based recommendation for CTR prediction, fit for highly time-sensitive news. This knowledge graph consists of the news content with a relevant entity. The DKN includes a critical component, namely knowledge-aware convolutional neural networks (KCNN), to fuse the word-level and knowledge-level representations of news and generate a knowledge-aware embedding vector. Experiment results indicate that the values of F-Measure and Area Under the Curve (AUC) were superior to those of other models compared (LibFM, a feature-based factorization model [85]; knowledge powered convolutional neural Network, KPCNN [86]; deep structured semantic model, DSSM; DeepWide; DeepFM [87]; YouTubeNet [72]; and deep matrix factorization model, DMF [88]).

Reference [89] proposed a novel self-attention-based model. Compared with DKN model, the model uses word-level description instead of topic-level description. The attention model contains four-level self-attention modules (word-level, item-level, user-level, and multi-head attention). The word-level model transformed the words in news articles to word-based representations. The item-level model fused the representation among words, entities, and contexts. The user-level model was used to construct user-embedding. The multi-head attention model was used to concatenate history and candidate news, and then calculated the CTR via a fully connected layer. The method proposed is better than the compared methods (LR, DNN, DeepFM, DeepWide, DKN, Popularity) for the AUC metric and NDCG metric.

Reference [82] proposed News Graph; it is the first time the knowledge graph was used in the news field to realize news recommendation tasks. News Graph is based on Microsoft Service Network (MSN) news data, and collaborative relations are added to the original KG, including Same News Triples, Same User Triples, and Same Session Triples. To verify the effectiveness of News Graph, four tasks were verified in the experiment (predicting the user's click on news items, news classification, news popularity, and the matching degree between the locations of news items and the interest of local users). The experimental results showed that News Graph performed better than the original KG in the above tasks in accuracy, AUC, and F1-score metrics.

Reference [90] proposed the SED (short distance over knowledge graphs) method, which consists of three steps: (1) finding the named entities involved in the news in FreeBase, and then finding the nodes connected by these entities; (2) generating subgraphs according to the nodes found in the first step; and (3) finding similar news in the subgraph by calculating the shortest distance between entities. In addition, CNRec, a public dataset for KG research was proposed by the authors of [90] (<https://github.com/kevinj22/CNRec>). The SED method can alleviate the cold-start problem, and the combination of the SED method with the traditional TF-IDF and doc2vec methods can also improve the recommendation performance.

Reference [83] proposed the topic-enriched knowledge graph recommendation system (TEKGR), which improves upon the existing KG methods by mining external knowledge in news entities and related topics. The TEKGR framework is divided into three layers (using

KG to model news headlines with three coding methods, obtaining users' interest expressions through attention mechanisms, and using the dot production method to calculate users' scores on candidate news). Experiments showed that TEKGR is superior to comparative methods (DKN, DeepFM, LibFM, LSTUR [91]) in AUC and F1-score metrics.

Reference [84] proposed a graph enhanced representation learning (GERL) approach, which is a mixture of KG and attention mechanisms. GERL not only models users and news but also considers the neighbors of users and news, thus increasing the learning ability between users and news. In addition, the transformer [92] model is applied when generating news vector representations. Experiments showed that the GERL method is superior to DKN, LibFM, DSSM, and other methods according to AUC, MRR, and NDCG metrics.

Reference [93] proposed a node2vec-based [94] economic news recommendation model. This method is mainly used to recommend economic news related to stocks for users, and the timeliness of news is considered in the model. In this paper, a securities KG was established, which includes stock, industry knowledge, related concepts, and other contents. Experiments showed that the model is superior to the traditional CF and CB recommendation methods, and can recommend personalized economic news for users in real-time in practical applications.

The KG can compensate for the cold-start problem in the recommendation system, as well as digging up more potential news information. Therefore, the accuracy of recommendation results is improved, but the abovementioned news recommendation method based on KG still has the following shortcoming: (1) the continuity of news events is not considered; (2) there are too many named entities and link nodes in the KG, and important entities and links are not considered, which makes the method more suitable for recommendation applications in the online environment; (3) the creation of KG does not consider cross-language domain knowledge, and KG cannot be directly used to complete cross-language recommendations; (4) in the field of news, there are not many domain KGs. For example, a KG specifically used in news recommendation was created by the authors of [82].

6 Evaluation Metrics in News Recommendation Methods

The evaluation of methods by research papers in recent years mainly concentrated on the precision of recommendation methods, the effectiveness of recommendation list ranking, the effectiveness of real-time recommendation and the application effect of recommendation. In these metrics, except for root mean squared error (RMSE) and mean absolute deviation (MAE), the larger the value of the other metrics, the better the recommendation effect.

The evaluation metrics of the recommendation's precision [95] include precision, recall, hit rate (HR), F-Measure (or F-Score), accuracy, RMSE, and MAE. Precision indicates the probability of users' interest in the news listed in the recommendation list. Recall indicates the probability of news that users are interested in to be recommended. HR indicates the total number of correct news recommendation items in each user's top-k recommendation list. It is an evaluation metric of recall rate used in top-k recommendation methods. F-Measure (or F-Score) metric comprehensively considers precision and recall. The accuracy metric indicates the ratio between numbers of correct recommended news items and that of all news samples. RMSE and MAE are used to forecast the score that users give for news, for example, the MF-based recommendation method.

The evaluation metrics of recommendation list ranking effectiveness includes normalized discounted cumulative gain (NDCG), mean average precision (MAP), and area under the curve (AUC) [74,96]. NDCG is used to indicate the relevance between users and recommended news items. MAP is similar to NDCG, but in the MAP metric, the relevance between users and recommended news items is a binary correlation. AUC measures the probability that a recommendation system ranks a positive user-item interaction higher than negative ones.

The evaluation metrics of the practical application of real-time recommendation includes KPIs (Key Performance Indicators) and SUS (System Usability Scale) in [62]. KPIs metric used to evaluate if the recommendation system is in good shape or not. SUS metric used to evaluate the usability of the recommendation system.

The evaluation metrics of the application effect of recommendation include the click through rate (CTR) metric, Kappa statistic metric [50], and mean reciprocal rank (MRR) metric [74]. The CTR metric is used to evaluate the number of hits on the news recommended. The Kappa statistic is a metric to measure whether the proposed classifications made are better than those made by a random guess. The MRR metric is used to evaluate the advantages and disadvantages of a recommendation list.

7 Prospects for News Recommendation Methods

It can be seen that in the research of the above-mentioned news recommendation methods, researchers mainly focus on personalized recommendation and recommendation accuracy. However, there are few studies on user fatigue, non-personalized news recommendation, privacy protection, news diversity, and other aspects, which are still important issues in future news recommendation research.

7.1 User Fatigue in News Recommendation

The news recommended for users in the news recommendation system is news similar to that in the user's browsing history, which makes users tired when reading news, thus affecting the user's reading experience. However, the factor of user fatigue is rarely considered in news recommendation methods. In [97], user fatigue is defined as the loss of interest of users in repeatedly recommended items, and CTR of news items is used to measure user fatigue. However, in [97], the recommendation method only considers the user's gender, age, and CTR of news items, and does not consider the semantic relationship of news items, users' long-term and short-term interests, etc. How to measure the factors of users' reading fatigue is also a problem worth studying. In addition, the diversity and serendipity aspects of relieving users' reading fatigue can be further studied in news recommendation methods.

7.2 Privacy Protection in News Recommendations

Most news recommendation methods are based on the historical behavior of users. Thus, these methods ignore the privacy protection of users' behaviors. Reference [98] pointed out that there are relatively few recommendation methods based on privacy protection in the field of news recommendation, but it is also a problem worth studying in the field of news. In 2020, [99] proposed a news recommendation method based on local differential privacy (LDP) [100]. The purpose of this method is to save the user's behavior only in the user's device, instead of uploading the user data to the server. The portrait of the user is not a portrait of a user but a group portrait of the user, thus making up for the impact of the lack of user information on the accuracy of the portrait of the user.

7.3 Non-personalized in News Recommendations

Reference [101] proposed a non-personalized news recommendation method based on future-impact. The purpose of this method is to provide users with news items that will become hot spots (news items that users will pay more attention to) in the future. The non-personalized news recommendation method can not only alleviate the cold start problem, but also reduce the complexity of the recommendation method in online news recommendation. However, non-personalized recommendation does not consider users' preferences and also affects the experience of some users. Thus, the combination of non-personalized recommendation and personalized recommendation methods is also a research direction for future news recommendation methods.

7.4 Recommendation of Long-Tail News

Long-tail news refers to news that is rarely recommended. Due to users rarely browsing long-tail news, long-tail news browsing records are few. In the recommendation model, long-tail news rarely becomes recommended news for users. However, long-tail news is also valuable and needs to be mined by the recommendation system and recommend to the target users. Due to the timeliness of news, it is hard to mine long-tail news that meets users' preferences in a timely and accurate manner. Therefore, in future news recommendation methods, the recommendation model of long-tail news is an important research direction, which can be modeled using KG or attention mechanisms.

7.5 Recommendation of High-Quality News

In the news recommendation method, recommend results only consider the accuracy rate and do not consider the news' quality. High-quality news refers to the authenticity, refinement, and integrity of the content. There are usually many news reports on the same event, and it is difficult to find the best news to recommend to users. Taking current affairs news as an example, high-quality news is refined and authoritative. Mining high-quality news content is closely related to the choice of news content analysis methods. In the future, the recommendation of high-quality news will consider narrowing the scope of candidate news (i.e., analyzing news content after classifying news according to events and removing fake news), analyzing news content on social media, and analyzing news content with expert knowledge.

8. Conclusions

News recommendation methods are a challenging field. This review summarized the improvement of the CF method, the CB method, and the hybrid method in the domain of news recommendation. This paper mainly summarized the research progress regarding the cluster-based CF methods and the MDP-based method within the CF method. This paper also depicted the research progress regarding semantics-based methods and social networks-based methods within the CB method. Then, this paper summarized the hybrid methods, including methods used in combination with search engine-based query methods, with the timeliness of news mining methods, with deep learning methods, attention mechanism-based methods, and the KG method. Finally, this paper discussed the challenges facing news recommendation methods.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this paper.

Funding Statement

This research received no external funding.

Acknowledgments

The author would like to thank the authors of all the references.

References

- [1] Goldberg, D.; Nichols, D.A.; Oki, B.M.; Terry, D.B. Using Collaborative Filtering to Weave an InformationTapestry. *Commun. ACM* **1992**, *35*, 61–70. doi:10.1145/138859.138867.
- [2] Agichtein, E.; Hakkani-Tür, D.; Kallumadi, S.; Malmasi, S. ConvERSe'20: The WSDM 2020 Workshop on Conversational Systems for E-Commerce Recommendations and Search. *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining*, Houston, TX, USA, February 3-7, 2020, 2020, pp. 897–898. doi:10.1145/3336191.3371882.
- [3] Renjith, S.; Sreekumar, A.; Jathavedan, M. An extensive study on the evolution of context-aware personalized travel recommender systems. *Inf. Process. Manag.* **2020**, *57*. doi:10.1016/j.ipm.2019.102078.
- [4] Campana, M.G.; Delmastro, F. Recommender Systems for Online and Mobile Social Networks: A survey. *Online Soc. Networks Media* **2017**, *3-4*, 75–97. doi:10.1016/j.osnem.2017.10.005.
- [5] Karimi, M.; Jannach, D.; Jugovac, M. News recommender systems - Survey and roads ahead. *Inf. Process. Manage.* **2018**, *54*, 1203–1227. doi:10.1016/j.ipm.2018.04.008.
- [6] Resnick, P.; Iacovou, N.; Suchak, M.; Bergstrom, P.; Riedl, J. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. *CSCW '94, Proceedings of the Conference on Computer Supported Cooperative Work*, Chapel Hill, NC, USA, October 22-26, 1994, 1994, pp. 175–186. doi:10.1145/192844.192905.
- [7] Lang, K. NewsWeeder: Learning to Filter Netnews. *Machine Learning, Proceedings of the Twelfth International Conference on Machine Learning*, Tahoe City, California, USA, July 9-12, 1995, pp.331–339. doi:10.1016/b978-1-55860-377-6.50048-7.
- [8] Das, A.S.; Datar, M.; Garg, A.; Rajaram, S. Google News Personalization: Scalable Online Collaborative Filtering. *Proceedings of the 16th International Conference on World Wide Web; ACM: New York, NY, USA, 2007; WWW '07*, pp. 271–280. doi:10.1145/1242572.1242610.
- [9] Park, K.; Lee, J.; Choi, J. Deep Neural Networks for News Recommendations. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 -10, 2017, 2017*, pp. 2255–2258. doi:10.1145/3132847.3133154.
- [10] Lian, J.; Zhang, F.; Xie, X.; Sun, G. Towards Better Representation Learning for Personalized News Recommendation: a Multi-Channel Deep Fusion Approach. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden., 2018*, pp. 3805–3811. doi:10.24963/ijcai.2018/529.
- [11] Wang, H.; Zhang, F.; Xie, X.; Guo, M. DKN: Deep Knowledge-Aware Network for News Recommendation. *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018, 2018*, pp. 1835–1844. doi:10.1145/3178876.3186175.
- [12] Borges, H.L.; Lorena, A.C.A Survey on Recommender Systems for News Data. In *Smart Information and Knowledge Management: Advances, Challenges, and Critical Issues*; 2010; pp. 129–151. doi:10.1007/978-3-642-04584-4_6.
- [13] Harandi, M.; Gulla, J.A. Survey of User Profiling in News Recommender Systems. *Proceedings of the 3rd International Workshop on News Recommendation and Analytics (INRA 2015) co-located with 9th ACM Conference on Recommender Systems (RecSys 2015)*, Vienna, Austria, September 20, 2015., 2015, pp. 20–26.
- [14] Li, M.; Wang, L. A Survey on Personalized News Recommendation Technology. *IEEE Access* **2019**, *7*, 145861–145879. doi:10.1109/ACCESS.2019.2944927.
- [15] Feng, C.; Khan, M.; Rahman, A.U.; Ahmad, A. News Recommendation Systems - Accomplishments, Challenges & Future Directions. *IEEE Access* **2020**, *8*, 16702–16725. doi:10.1109/ACCESS.2020.2967792.
- [16] Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. Item-based Collaborative Filtering Recommendation Algorithms. *Proceedings of the 10th International Conference on World Wide Web; ACM: New York, NY, USA, 2001; WWW '01*, pp. 285–295. doi:10.1145/371920.372071.
- [17] Breese, J.S.; Heckerman, D.; Kadie, C.M. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. *UAI '98: Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, University of Wisconsin Business School, Madison, Wisconsin, USA, July 24-26, 1998, 1998, pp. 43–52.
- [18] Koren, Y.; Bell, R.M.; Volinsky, C. Matrix Factorization Techniques for Recommender Systems. *IEEE Computer* **2009**, *42*, 30–37. doi:10.1109/MC.2009.263.
- [19] Li, L.; Wang, D.; Li, T.; Knox, D.; Padmanabhan, B. SCENE: a scalable two-stage personalized news recommendation system. *Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2011, Beijing, China, July 25-29, 2011, 2011*, pp. 125–134. doi:10.1145/2009916.2009937.
- [20] Bouras, C.; Tsogkas, V. Improving news articles recommendations via user clustering. *Int. J. Machine Learning & Cybernetics* **2017**, *8*, 223–237. doi:10.1007/s13042-014-0316-3.
- [21] Yoneda, T.; Kozawa, S.; Osone, K.; Koide, Y.; Abe, Y.; Seki, Y. Algorithms and System Architecture for Immediate Personalized News Recommendations. *CoRR* **2019**, abs/1909.01005, [1909.01005].

- [22] Khattar, D.; Kumar, V.; Varma, V. Leveraging Moderate User Data for News Recommendation. 2017 IEEE International Conference on Data Mining Workshops, ICDM Workshops 2017, New Orleans, LA, USA, November 18-21, 2017, 2017, pp. 757–760. doi:10.1109/ICDMW.2017.104.
- [23] Hofmann, T. Latent semantic models for collaborative filtering. *ACM Trans. Inf. Syst.* 2004, 22, 89–115. doi:10.1145/963770.963774.
- [24] Chang, F.; Dean, J.; Ghemawat, S.; Hsieh, W.C.; Wallach, D.A.; Burrows, M.; Chandra, T.; Fikes, A.; Gruber, R.E. Bigtable: A Distributed Storage System for Structured Data. *ACM Trans. Comput. Syst.* 2008, 26, 4:1–4:26. doi:10.1145/1365815.1365816.
- [25] Gionis, A.; Indyk, P.; Motwani, R. Similarity Search in High Dimensions via Hashing. *VLDB'99, Proceedings of 25th International Conference on Very Large Data Bases*, September 7-10, 1999, Edinburgh, Scotland, UK, 1999, pp. 518–529.
- [26] Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 2003, 3, 993–1022.
- [27] Bouras, C.; Tsogkas, V. W-kmeans: Clustering News Articles Using WordNet. *Knowledge-Based and Intelligent Information and Engineering Systems - 14th International Conference, KES 2010, Cardiff, UK, September 8-10, 2010, Proceedings, Part III*, 2010, pp. 379–388. doi:10.1007/978-3-642-15393-8_43.
- [28] Lin, D. WordNet: An Electronic Lexical Database. *Computational Linguistics* 1999, 25, 292–296.
- [29] Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26*; Burges, C.J.C.; Bottou, L.; Welling, M.; Ghahramani, Z.; Weinberger, K.Q., Eds.; Curran Associates, Inc., 2013; pp.3111–3119.
- [30] Wang, Y.; Wang, L.; Li, Y.; He, D.; Liu, T. A Theoretical Analysis of NDCG Type Ranking Measures. *COLT 2013 - The 26th Annual Conference on Learning Theory*, June 12-14, 2013, Princeton University, NJ, USA, 2013, pp. 25–54.
- [31] IJntema, W.; Goossen, F.; Frasincar, F.; Hogenboom, F. Ontology-based news recommendation. *Proceedings of the 2010 EDBT/ICDT Workshops*, Lausanne, Switzerland, March 22-26, 2010, 2010. doi:10.1145/1754239.1754257.
- [32] Paterek, A. Improving regularized singular value decomposition for collaborative filtering. *Proceedings of Kdd Cup & Workshop 2007*.
- [33] Garcin, F.; Zhou, K.; Faltings, B.; Schickel, V. Personalized News Recommendation Based on Collaborative Filtering. *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, 2012, Vol. 1, pp. 437–441.
- [34] Billsus, D.; Pazzani, M.J. A Personal News Agent That Talks, Learns and Explains. *Proceedings of the Third Annual Conference on Autonomous Agents, AGENTS 1999, Seattle, WA, USA, May 1-5, 1999*, pp. 268–275. doi:10.1145/301136.301208.
- [35] Gabrilovich, E.; Dumais, S.T.; Horvitz, E. Newsjunkie: providing personalized newsfeeds via analysis of information novelty. *Proceedings of the 13th international conference on World Wide Web, WWW 2004, New York, NY, USA, May 17-20, 2004*, pp. 482–490. doi:10.1145/988672.988738.
- [36] Ahn, J.; Brusilovsky, P.; Grady, J.; He, D.; Syn, S.Y. Open user profiles for adaptive news systems: help or harm? *Proceedings of the 16th International Conference on World Wide Web, WWW 2007, Banff, Alberta, Canada, May 8-12, 2007*, pp. 11–20. doi:10.1145/1242572.1242575.
- [37] Guarino, N.; Masolo, C.; Vetere, G. OntoSeek: Content-Based Access to the Web. *Intelligent Systems & Their Applications IEEE* 1999, 14, 70–80.
- [38] Cantador, I.; Bellogin, A.; Castells, P. News@hand: A Semantic Web Approach to Recommending News. *Adaptive Hypermedia and Adaptive Web-Based Systems, 5th International Conference, AH 2008, Hannover, Germany, July 29 - August 1, 2008*, Proceedings, 2008, pp. 279–283. doi:10.1007/978-3-540-70987-9_34.
- [39] Salton, G.; Wong, A.; Yang, C. A Vector Space Model for Automatic Indexing. *Commun. ACM* 1975, 18, 613–620. doi:10.1145/361219.361220.
- [40] Salton, G.; Buckley, C. Term-Weighting Approaches in Automatic Text Retrieval. *Inf. Process. Manage.* 1988, 24, 513–523. doi:10.1016/0306-4573(88)90021-0.
- [41] Jones, K.S. Jones, K.S.: A Statistical Interpretation of Term Specificity and its Application in Retrieval. *Journal of Documentation* 28(1), 11-21. *Journal of Documentation* 1972, 28, 11–21.
- [42] Le, Q.; Mikolov, T. Distributed Representations of Sentences and Documents. *Proceedings of the 31st International Conference on Machine Learning*; Xing, E.P.; Jebara, T., Eds.; PMLR: Beijing, China, 2014; Vol. 32, *Proceedings of Machine Learning Research*, pp. 1188–1196.
- [43] Goossen, F.; IJntema, W.; Frasincar, F.; Hogenboom, F.; Kaymak, U. News personalization using the CF-IDF semantic recommender. *Proceedings of the International Conference on Web Intelligence, Mining and Semantics, WIMS 2011, Sogndal, Norway, May 25 - 27, 2011*, p. 10. doi:10.1145/1988688.1988701.
- [44] de Koning, E.; Hogenboom, F.; Frasincar, F. News Recommendation with CF-IDF+. *Advanced Information Systems Engineering - 30th International Conference, CAiSE 2018, Tallinn, Estonia, June 11-15, 2018, Proceedings*, 2018, pp. 170–184. doi:10.1007/978-3-319-91563-0_11.
- [45] Capelle, M.; Frasincar, F.; Moerland, M.; Hogenboom, F. Semantics-based news recommendation. *2nd International Conference on Web Intelligence, Mining and Semantics, WIMS '12, Craiova, Romania, June 6-8, 2012*, pp. 27:1–27:9. doi:10.1145/2254129.2254163.
- [46] Hogenboom, F.; Capelle, M.; Moerland, M. News Recommendation Using Semantics with the Bing-SF-IDF Approach. *Advances in Conceptual Modeling - ER 2013 Workshops, LSAWM, MoBiD, RIGiM, SeCoGIS, WISM, DaSeM, SCME, and PhD Symposium*, Hong Kong, China, November 11-13, 2013, *Revised Selected Papers*, 2013, pp. 160–169. doi:10.1007/978-3-319-14139-8_18.
- [47] Moerland, M.; Hogenboom, F.; Capelle, M.; Frasincar, F. Semantics-based news recommendation with SF-IDF+. *3rd International Conference on Web Intelligence, Mining and Semantics, WIMS '13, Madrid, Spain, June 12-14, 2013*, p. 22. doi:10.1145/2479787.2479795.
- [48] Capelle, M.; Moerland, M.; Hogenboom, F.; Frasincar, F.; Vandic, D. Bing-SF-IDF+: a hybrid semantics-driven news

- recommender. Proceedings of the 30th Annual ACM Symposium on Applied Computing, Salamanca, Spain, April 13-17, 2015, 2015, pp. 732–739. doi:10.1145/2695664.2695700.
- [49] Brocken, E.; Hartveld, A.; de Koning, E.; van Noort, T.; Hogenboom, F.; Frasincar, F.; Robal, T. Bing-CF-IDF+: A Semantics-Driven News Recommender System. *Advanced Information Systems Engineering*; Giorgini, P.; Weber, B., Eds.; Springer International Publishing: Cham, 2019; pp. 32–47.
- [50] Cohen, J. A Coefficient of Agreement for Nominal Scales. *Psychological Bulletin* 1960, 20, 37–46. doi:10.1177/0013164460020000104.
- [51] Tiwari, S.; Pangtey, M.S.; Kumar, S. Location Aware Personalized News Recommender System Based on Twitter Popularity. *Computational Science and Its Applications – ICCSA*; 2018; pp. 650–658.
- [52] Abel, F.; Gao, Q.; Houben, G.; Tao, K. Analyzing User Modeling on Twitter for Personalized News Recommendations. *User Modeling, Adaption and Personalization - 19th International Conference, UMAP 2011, Girona, Spain, July 11-15, 2011. Proceedings*, 2011, pp. 1–12. doi:10.1007/978-3-642-22362-4_1.
- [53] Abel, F.; Gao, Q.; Houben, G.; Tao, K. Twitter-Based User Modeling for News Recommendations. *IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013*, pp. 2962–2966.
- [54] Phelan, O.; McCarthy, K.; Bennett, M.; Smyth, B. Terms of a Feather: Content-Based News Recommendation and Discovery Using Twitter. *Advances in Information Retrieval - 33rd European Conference on IR Research, ECIR 2011, Dublin, Ireland, April 18-21, 2011. Proceedings*, 2011, pp. 448–459. doi:10.1007/978-3-642-20161-5_44.
- [55] Jonnalagedda, N.; Gauch, S. Personalized News Recommendation Using Twitter. *2013 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, Atlanta, Georgia, USA, 17-20 November 2013, Workshop Proceedings*, 2013, pp. 21–25. doi:10.1109/WI-IAT.2013.144.
- [56] Manoharan, S.; Senthilkumar, R. An Intelligent Fuzzy Rule-Based Personalized News Recommendation Using Social Media Mining. *Comput. Intell. Neurosci.* 2020, 2020, 3791541:1–3791541. doi:10.1155/2020/3791541.
- [57] Wang, X.; He, X.; Nie, L.; Chua, T. Item Silk Road: Recommending Items from Information Domains to Social Users. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*, pp. 185–194. doi:10.1145/3077136.3080771.
- [58] Balabanovic, M.; Shoham, Y. Content-Based, Collaborative Recommendation. *Commun. ACM* 1997, 40, 66–72. doi:10.1145/245108.245124.
- [59] Lu, Z.; Dou, Z.; Lian, J.; Xie, X.; Yang, Q. Content-Based Collaborative Filtering for News Topic Recommendation. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA., 2015*, pp. 217–223.
- [60] Pessemier, T.D.; Leroux, S.; Vanhecke, K.; Martens, L. Combining Collaborative Filtering and Search Engine into Hybrid News Recommendations. *Proceedings of the 3rd International Workshop on News Recommendation and Analytics (INRA 2015) co-located with 9th ACM Conference on Recommender Systems (RecSys 2015), Vienna, Austria, September 20, 2015., pp. 14–19.*
- [61] Jaakkola, T.; Haussler, D. Exploiting Generative Models in Discriminative Classifiers. In *Advances in Neural Information Processing Systems 11*; Kearns, M.J.; Solla, S.A.; Cohn, D.A., Eds.; MIT Press, 1999; pp. 487–493.
- [62] Lenhart, P.; Herzog, D. Combining Content-based and Collaborative Filtering for Personalized Sports News Recommendations. *Proceedings of the 3rd Workshop on New Trends in Content-Based Recommender Systems co-located with ACM Conference on Recommender Systems (RecSys 2016), Boston, MA, USA, September 16, 2016., pp. 3–10.*
- [63] Lv, P.; Meng, X.; Zhang, Y. FeRe: Exploiting influence of multi-dimensional features resided in news domain for recommendation. *Inf. Process. Manage.* 2017, 53, 1215–1241. doi:10.1016/j.ipm.2017.04.008.
- [64] Sottocornola, G.; Symeonidis, P.; Zanker, M. Session-based News Recommendations. *Companion of the Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon, France, April 23-27, 2018*, pp. 1395–1399. doi:10.1145/3184558.3191582.
- [65] Lommatzsch, A.; Kille, B.; Albayrak, S. Incorporating context and trends in news recommender systems. *Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, August 23-26, 2017*, pp. 1062–1068. doi:10.1145/3106426.3109433.
- [66] Lommatzsch, A.; Johannes, N.; Meiners, J.; Helmers, L.; Domann, J. Recommender Ensembles for News Articles based on Most-Popular Strategies. *Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum, Évora, Portugal, 5-8 September, 2016., 2016*, pp. 657–668.
- [67] He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; & Chua, T.-S. Neural Collaborative Filtering. *Proceedings of the 26th International Conference on World Wide Web - WWW '17*, 173–182.
- [68] Kumar, V.; Khattar, D.; Gupta, S.; Varma, V. Word Semantics Based 3-D Convolutional Neural Networks for News Recommendation. *2017 IEEE International Conference on Data Mining Workshops (ICDMW), 2017*, pp. 761–764.
- [69] Khattar, D.; Kumar, V.; Varma, V.; Gupta, M. Weave&Rec: A Word Embedding based 3-D Convolutional Network for News Recommendation. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018, 2018*, pp. 1855–1858. doi:10.1145/3269206.3269307.
- [70] Khattar, D.; Kumar, V.; Gupta, M.; Varma, V. Neural Content-Collaborative Filtering for News Recommendation. *Proceedings of the Second International Workshop on Recent Trends in News Information Retrieval co-located with 40th European Conference on Information Retrieval (ECIR 2018), Grenoble, France, March 26, 2018., 2018*, pp. 45–50.
- [71] Khattar, D.; Kumar, V.; Varma, V.; Gupta, M. HRAM: A Hybrid Recurrent Attention Machine for News Recommendation. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018, 2018*, pp. 1619–1622. doi:10.1145/3269206.3269311.
- [72] Covington, P.; Adams, J.; Sargin, E. Deep Neural Networks for YouTube Recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016*, pp. 191–198. doi:10.1145/2959100.2959190.

- [73] Cheng, H.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Anderson, G.; Corrado, G.; Chai, W.; Ispir, M.; Anil, R.; Haque, Z.; Hong, L.; Jain, V.; Liu, X.; Shah, H. Wide & Deep Learning for Recommender Systems. Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, DLRS@RecSys 2016, Boston, MA, USA, September 15, 2016, pp. 7–10. doi:10.1145/2988450.2988454.
- [74] Bellogín A., Said A. Recommender Systems Evaluation. In: Alhajj R., Rokne J. (eds) Encyclopedia of Social Network Analysis and Mining. Springer, New York, NY, 2018. https://doi.org/10.1007/978-1-4939-7131-2_110162
- [75] Zhu, Q.; Zhou, X.; Song, Z.; Tan, J.; Guo, L. DAN: Deep Attention Neural Network for News Recommendation. Proceedings of the AAAI Conference on Artificial Intelligence 2019, 33, 5973–5980. doi:10.1609/aaai.v33i01.33015973.
- [76] Huang, P.; He, X.; Gao, J.; Deng, L.; Acero, A.; Heck, L.P. Learning deep structured semantic models for web search using clickthrough data. 22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013, 2013, pp. 2333–2338. doi:10.1145/2505515.2505665.
- [77] Gulla, J.A.; Zhang, L.; Liu, P.; Özgöbek, Ö.; Su, X. The Adressa dataset for news recommendation. Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, August 23-26, 2017, pp. 1042–1048. doi:10.1145/3106426.3109436.
- [78] Wu, C.; Wu, F.; An, M.; Huang, J.; Huang, Y.; Xie, X. NPA: Neural News Recommendation with Personalized Attention. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019., pp. 2576–2584. doi:10.1145/3292500.3330665.
- [79] Wu, C.; Wu, F.; An, M.; Huang, Y.; Xie, X. Neural News Recommendation with Topic-Aware News Representation. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Association for Computational Linguistics: Florence, Italy, 2019; pp. 1154–1159. doi:10.18653/v1/P19-1110.
- [80] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD international conference on Management of data (SIGMOD '08). Association for Computing Machinery, New York, NY, USA, 1247–1250. <https://doi.org/10.1145/1376616.1376746>
- [81] abian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In Proceedings of the 16th international conference on World Wide Web (WWW '07). Association for Computing Machinery, New York, NY, USA, 697–706. <https://doi.org/10.1145/1242572.1242667>
- [82] Liu, D.; Bai, T.; Lian, J.; Zhao, X.; Sun, G.; Wen, J.; Xie, X. News Graph: An Enhanced Knowledge Graph for News Recommendation. Proceedings of the Second Workshop on Knowledge-aware and Conversational Recommender Systems, co-located with 28th ACM International Conference on Information and Knowledge Management, KaRS@CIKM 2019, Beijing, China, November 7, 2019, pp. 1–7.
- [83] Lee, D., & Lee, K. News Recommendation with Topic-Enriched Knowledge Graphs. 695–704.
- [84] Ge, S., Wu, C., Wu, F., Qi, T., & Huang, Y. Graph Enhanced Representation Learning for News Recommendation. 2863–2869.
- [85] Rendle, S. Factorization Machines with libFM. ACM Trans. Intell. Syst. Technol. 2012, 3, 57:1–57:22. doi:10.1145/2168752.2168771.
- [86] Wang, J.; Wang, Z.; Zhang, D.; Yan, J. Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, 2017, pp. 2915–2921. doi:10.24963/ijcai.2017/406.
- [87] Guo, H.; Tang, R.; Ye, Y.; Li, Z.; He, X. DeepFM: A Factorization-Machine based Neural Network for CTR Prediction. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, pp. 1725–1731. doi:10.24963/ijcai.2017/239.
- [88] Xue, H.; Dai, X.; Zhang, J.; Huang, S.; Chen, J. Deep Matrix Factorization Models for Recommender Systems. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, pp. 3203–3209. doi:10.24963/ijcai.2017/447.
- [89] Gao, J.; Xin, X.; Liu, J.; Wang, R.; Lu, J.; Li, B.; Fan, X.; Guo, P. Fine-Grained Deep Knowledge-Aware Network for News Recommendation with Self-Attention. 2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018, Santiago, Chile, December 3-6, 2018, 2018, pp. 81–88. doi:10.1109/WI.2018.0-104.
- [90] Joseph, K., & Jiang, H. Content based news recommendation via shortest entity distance over knowledge graphs. The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019, 690–699.
- [91] Mingxiao An, Fangzhao Wu, Chuhan Wu, Kun Zhang, Zheng Liu, and Xing Xie. 2019. Neural news recommendation with long- and short-term user representations. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 336–345.
- [92] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In NIPS. 5998–6008.
- [93] Ren, J., Long, J., & Xu, Z. Financial news recommendation based on graph embeddings. Decision Support Systems, 125(January), 113115.
- [94] A. Grover, J. Leskovec. node2vec: Scalable feature learning for networks, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 855–864.
- [95] Asela Gunawardana, Guy Shani. A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. J. Mach. Learn. Res. 10: 2935-2962 (2009)
- [96] Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 1997, 30, 1145–1159. doi:10.1016/S0031-3203(96)00142-2.
- [97] Ma, H.; Liu, X.; Shen, Z. User Fatigue in Online News Recommendation. Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016, pp. 1363–1372. doi:10.1145/2872427.2874813.
- [98] Mohallick, I.; Özgöbek, Ö. Exploring privacy concerns in news recommender systems. Proceedings of the International

- Conference on Web Intelligence, Leipzig, Germany, August 23-26, 2017, pp. 1054–1061. doi:10.1145/3106426.3109435.
- [99] Qi, T.; Wu, F.; Wu, C.; Huang, Y.; Xie, X. cNews Recommendation Model Training via Federated Learning. CoRR 2020, abs/2003.09592, [2003.09592].
- [100] Ren, X.; Yu, C.; Yu, W.; Yang, S.; Yang, X.; McCann, J.A.; Yu, P.S. LoPub: High-Dimensional Crowdsourced Data Publication With Local Differential Privacy. IEEE Trans. Information Forensics and Security 2018, 13, 2151–2166. doi:10.1109/TIFS.2018.2812146.
- [101] Chakraborty, A.; Ghosh, S.; Ganguly, N.; Gummadi, K.P. Optimizing the recency-relevance-diversity trade-offs in non-personalized news recommendations. Inf.Repr.J. 2019, 22, 447–475. doi:10.1007/s10791-019-09351-2.