

News Recommender Systems - Survey and Roads Ahead¹

Mozhgan Karimi^a, Dietmar Jannach^{b2}, Michael Jugovac^c

^aDepartment of Mathematics and Computer Science, University of Antwerp, Antwerp, Belgium

^bDepartment of Applied Informatics, AAU Klagenfurt, 9020 Klagenfurt, Austria

^cDepartment of Computer Science, TU Dortmund, 44221 Dortmund, Germany

mozhgan.karimi@tu-dortmund.de, dietmar.jannach@aau.at, michael.jugovac@tu-dortmund.de

Abstract

More and more people read the news online, e.g., by visiting the websites of their favorite newspapers or by navigating the sites of news aggregators. However, the abundance of news information that is published online every day through different channels can make it challenging for readers to locate the content they are interested in. The goal of News Recommender Systems (NRS) is to make reading suggestions to users in a personalized way. Due to their practical relevance, a variety of technical approaches to build such systems have been proposed over the last two decades. In this work, we review the state-of-the-art of designing and evaluating news recommender systems over the last ten years. One main goal of the work is to analyze which particular challenges of news recommendation (e.g., short item life times and recency aspects) have been well explored and which areas still require more work. Furthermore, in contrast to previous surveys, the paper specifically discusses methodological questions and today's academic practice of evaluating and comparing different algorithmic news recommendation approaches based on accuracy measures.

Keywords: News Recommender Systems, Survey, Evaluation

1. Introduction

The newspaper industry has experienced a substantial transformation during the last twenty years. Today, readers can find various sources of news online, e.g., on the web presences of traditional newspaper companies, on digital-only news sites, or on news aggregation platforms provided, for example, by Google³ or Yahoo!⁴. Additionally, the digital form of information delivery allows publishers to distribute new or updated content in real-time, leading to an increased speed of publication. The availability of the various (often free) online news sources has led to a constant increase of users of such platforms [138]. At the same time, however, the abundance of available information and the constant update cycle make it increasingly challenging for readers to keep track of news that are most relevant to them.

Recommender Systems have shown to be a valuable tool to help users in such situations of information overload [84]. The main tasks of such systems are typically to filter incoming streams of information according to the users'

¹<https://doi.org/10.1016/j.ipm.2018.04.008>, This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.0 Generic License.

²Corresponding author

³<https://news.google.com/>

⁴<https://www.yahoo.com/news/>

preferences or to point them to additional items of interest in the context of a given object. During the past decades, significant advances in recommendation technology have been made. Recommenders have been successfully applied in a variety of domains, and the recommendable objects include movies, books, travel and tourism services, research articles, search queries, and many more [9, 15, 144].

News recommendation in general represents another application domain in which several of the known techniques for building automated recommendations can be applied. In fact, there have been a number of reported instances where such systems are being used in live environments to deliver news recommendations on popular websites (see, e.g., [32, 39, 93, 115, 118]). However, news recommendation problems often have certain characteristics that are either not present at all or that are at least less pronounced in other domains. For example, in contrast to other domains like movie recommendation, the relevance of news items can change very rapidly (i.e., decline or re-increase due to recent real-life events) [5, 141] and the “item churn” is generally high [39]. In fact, since news web sites are often continuously updated, some articles can be superseded by a “breaking news” article on the same topic several times during a single day, which might require constant updates to the recommendation models. Another typical challenge in the news domain is that a user’s interest can dynamically change, depending on different contextual factors like the time of the day, the features of the user’s device (e.g., mobile phone vs. desktop), or the user’s current location [20, 90, 125].

Due to the high practical relevance of the news recommendation problem and its specific challenges, a considerable number of research works has been published on this topic in particular within the last ten years. Many of these works propose novel algorithmic approaches to generate personalized recommendations. These algorithms are typically evaluated using offline experimental designs and existing log data. In recent years, however, recommender systems research in a variety of domains has shown that it is also important to evaluate recommendations in a more user- or system utility-oriented way [48, 78, 83, 96, 151]. These developments have also led to a number of alternative forms of assessing the quality of the recommendations, e.g., in terms of diversity [80, 161, 172].

With this work, we consider these recent developments and provide an overview of what has been achieved in the last ten years with respect to the different challenges in the news recommendation domain. Based in this analysis, we identify existing research gaps and potential areas for future research. In contrast to some previous overview works like [14, 109, 141], we however focus not only on the underlying algorithmic approaches used to create the recommendations, but also on questions related to the empirical evaluation and the user perception of such systems. To provide a starting point for future research, we finally also report insights from a set of experiments, which highlight the importance of short-term model updates when standard accuracy measures are applied in the evaluation.

The paper is organized as follows. Next, in Section 2, we briefly summarize how we selected existing research works for consideration in our survey. Afterwards, in Section 3, we discuss the various challenges of news recommendation in more detail and review existing algorithmic approaches to deal with these challenges. Then, in Section 4, we provide a survey of today’s academic practice of benchmarking and evaluating different technical approaches. Our paper ends with an outlook on possible directions for future research in Section 5.

2. Survey Method and Research Scope

To develop the survey part of the paper, we investigated, in a structured way, more than 140 research articles that appeared between 2005 and 2016 in relevant computer science and information systems publication outlets. Figure 1 shows how many NRS papers were considered in this time frame per year. The trend of the graph indicates that research interest in the topic of news recommendation has grown steadily over the years to the point that news has become an important sub-topic in the RS research field.

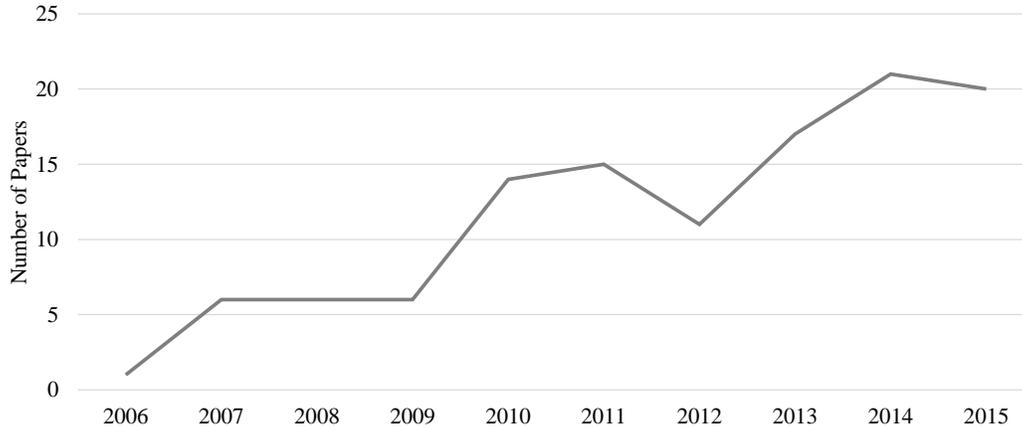


Figure 1: Number of NRS papers per year. As the survey only covered 2016 until August, this year was excluded.

Regarding the scope of our review, we focus on papers that describe approaches for “classical” news recommendation scenarios, e.g., on news aggregation sites. Similar technical approaches can in many cases be applied for “news feed filtering” problems, for example on Facebook or Twitter, where the content of *social* news feeds are personalized to the users. These scenarios feature many of the same challenges, e.g., an emphasis on freshness and an overwhelming number of new items every day. However, news feed filtering on social networks has some unique aspects that are different from news recommendation. For example, the recommendable items are not necessarily news items but can be some currently trending viral videos that are not related to a recent event in the real world. Also, on news sites, information about a visitor’s social network friends and general interest profile is in many cases non-existent. However, this information oftentimes plays a central role in news feed personalization.

The identification of papers for our survey was done according to the following strategy. We first scanned the proceedings and volumes of a set of relevant conference series (RecSys, WWW, SIGIR, and KDD) and journals (Expert Systems with Applications and Knowledge-Based Systems) for articles that fall into the above-described scope. Additionally, we used the keywords “news recommendation” and “news recommender” to search for papers in Google Scholar. Using the resulting set of articles as a starting point, we followed the references of the retrieved articles to find additional papers on the topic. The survey methodology is therefore not that of a traditional “systematic review” in the sense of, e.g., [94], where pre-defined search queries and inclusion and exclusion criteria are used. Nonetheless, as we followed a defined search strategy, had a defined research scope, and used pre-structured forms to

classify the papers along different dimensions, we are confident that the risk of the introduction of a researcher bias into the survey is low.

A few other survey works on the topic of news recommendation already exist. When looking at these existing works, we find that some challenges of news recommendation that we identify in our work were also summarized by Özgöbek et al. [141]. The authors of this work however only consider a small number of (twelve) papers and do not consider evaluation aspects in their survey. Li et al. [109] also discuss challenges of news recommendation, but do not discuss existing approaches to deal with these issues in depth. The survey paper by Borges and Lorena [14], finally, discusses news recommendation mostly from a quite general perspective and focuses the analysis on different aspects of six specific papers.

3. News Recommendation – Challenges and Algorithmic Approaches

Before we discuss the challenges of the domain in more detail along with algorithmic approaches that were proposed in the literature, we will briefly review which types of algorithms are generally used for the news recommendation problem.

3.1. Recommendation Paradigms in News Recommendation

Recommender systems are often classified into four main categories [84, 108]: collaborative filtering (CF), content-based filtering (CB), knowledge-based techniques (KB), and hybrid approaches. In academic settings, collaborative filtering is the most common approach in the recommender systems literature according to the survey presented in [85], while all other approaches are much less frequently employed. The general dominance of collaborative filtering is in some sense not surprising because this method, whose recommendations are based on the “wisdom of the crowd”, is domain-independent and does not require detailed (and domain-specific) information about the recommendable items. Furthermore, a number of public benchmark data sets, e.g., from MovieLens⁵, are available, which has been a driving factor for the academic community.

In the news recommendation domain, however, things are different. An analysis of the 112 (of 144) papers in our survey that propose one or more recommendation algorithms reveals the following.⁶ Content-based filtering, which is, roughly speaking, based on analyzing a reader’s past documents of interest and recommending “more of the same”, was used as an underlying paradigm in 59 of the analyzed papers. Approaches, like the Google News personalization system [39], that rely on collaborative filtering and on interest patterns in a community without considering the content of a news article are only proposed in 19 of the 112 papers. However, 45 of the papers use a hybrid approach and combine content-based filtering and collaborative filtering. This means that although collaborative filtering alone is

⁵<https://grouplens.org/datasets/movielens/>

⁶The complete categorization of all surveyed recommendation approaches from the literature based on, e.g., the algorithm paradigm, used data sets, and evaluation metrics, can be found in the appendix in Table A.3. Note also that papers can be attributed to multiple algorithm paradigms in case they propose more than one approach.

not the method of choice for most researchers, relying on patterns in the community behavior in addition to content-based user profiles is often considered to be promising. Knowledge-based techniques, which are based on explicit domain knowledge to map user preferences to article features, were not applied in any of the analyzed papers, neither in isolation nor in combination with other techniques. This is, however, not very surprising, since such approaches are generally only employed in high-involvement product domains with a long life cycle.

Since the recommendable objects in the news domain are text documents, which can be automatically analyzed with standard techniques from Information Retrieval (IR), it is in fact not too surprising that many researchers rely on content-based techniques. However, in general, content-based techniques are considered not to be very accurate in offline experiments when using IR measures like precision and recall [80]. As in other domains, the main dilemma of academic research – and to some extent also for recommendation service providers – is that it is not always clear to what extent high accuracy in offline experiments translates into online success [60]. Only very few studies are available that compare the offline and online performance of recommendation algorithms, e.g., [10] in the field of research paper recommendation. In the news domain, a comparative offline/online experiment was recently conducted by Garcin et al. [55], who also found that offline performance is not necessarily a predictor of online success. In their case, a popularity-based method performed best offline, but, in the end, showed the worst results in an online test, placing far behind a more sophisticated algorithm.

As we will discuss later, the effectiveness of *pure* collaborative filtering methods might be overestimated in offline experiments, which calls for the design of novel hybrid approaches that combine multiple sources of preference signals and which are able to balance prediction accuracy with other quality factors like novelty or diversity.

3.2. Challenges of User Modeling for News Recommendation

To be able to personalize the reading suggestions, recommender systems have to create and maintain user profiles that capture each user’s reading preferences over time. How these user profiles are built and which information is collected and used is usually related to the chosen recommendation paradigm. Similar to other application domains of recommender systems, one can in general try to stimulate users to provide explicit preference information, e.g., in the form of “like” or “dislike” statements, or monitor and interpret the user’s past behavior (implicit feedback) over time. Typical implicit feedback signals in the news domain include reading an article, sharing it, printing it, or commenting on it. On a more fine-grained level, it is also possible to record dwelling times or mouse movements as interest indicators.

Generally, different types of information can be used to construct the user profiles. Many papers rely mainly on the content of the news articles inspected by a user to infer the interest profiles. Some works, in addition, consider features of the users and the community they belong to to create the profiles. Other approaches finally use various types of information in parallel.

An example of a system that relies mostly on content information is the Athena news recommendation system [74], where the user profile is based on the keywords of the articles that were read by a user. New or not yet viewed articles are then ranked according to the similarity of their content with the aggregated user profile. A similar approach was

used by Rao et al. [153], who employ user profiles that have a “bag-of-concepts” format with DBPedia as a knowledge base in the background. Standard distance measures like cosine similarity or the Jaccard coefficient can then be used to assess the relevance of a news article for a given profile. Similar to these approaches, Chu and Park [31] rely on the content of the read articles for profile construction, and additionally consider demographic aspects to assess the relevance of new articles for a user.

Another work that considers demographic information as a key factor for user profiling is [103]. In this approach, the users were first segmented according to their demographics and article access patterns. The relevance of a new article for a given user was then calculated based on various factors, including the estimated interest of the user’s segment in the topic of the article or the number of users in the segment who have already read the article. Instead of clustering the users, Li et al. [112] applied a graph-based model based on which the logical relationship and similarities between a newly posted article and user comments about the article on a social media site were calculated. This topic-based model was then used to retrieve other relevant news articles.

The basis for the user modeling approach in [5] is a weighted term vector, which is generated by considering the topics of the articles that were read by a user. Different from other approaches, however, the authors propose to maintain a short-term and a long-term model, where the short-term model only considers the 20 most recently read articles of each user. Besides being able to capture long-term and short-term interests for recommending, the topic-based approach of Ahn et al. [5] was designed in a way that users can be put in control of their recommendations. For each recommended article, the relevant topics were displayed and users could then update their profile accordingly when they did not agree with the system’s assumptions about the preferences.

An example of a work that solely relies on click behavior and does not consider article content is presented by Das et al. [39], who describe the inner workings of the Google News Personalization system at that time. Similar to Lee and Park [103], they clustered the user community, this time however based mainly on the past article reading behavior. To predict the relevance of an item, they then used both a long-term collaborative filtering model and a short-term model based on article co-visitations.

In a later paper, Liu et al. [115] present an alternative approach that was also implemented for the Google News service. In contrast to the previous approach, their newer model takes both content information as well as the reading patterns in the community into account. Their proposed Bayesian framework tries to remove the community bias from the user’s consumption behavior in each topic to extract the user’s “genuine” long-term interests. Furthermore, it is also designed to take short-term changes in the user’s interests into account. To predict the relevance of new articles, the proposed approach finally also considers the user’s location and the reading behavior of the user community in the past hour.

Generally, considering long-term and short-term preferences and balancing their importance represent an important problem in news recommendation [72, 110]; see also the discussion later in Section 3.4 on recency aspects. One of the key design questions is whether two separate models should be maintained (as, e.g., in [6]) or an integrated model with a time decay factor (as in [111]). A comparison of different strategies of considering the long-term and short-term

preferences was done by Huy et al. [72], who observed that a combination of both models in a hybrid approach led to the best results.

3.3. Cold-Start Problems and Data Sparsity Issues in News Recommendation

Cold-start situations are a common problem in many application domains of recommender systems. *User cold-start* refers to situations where little is known about the preferences of the user for which a recommendation is requested; *item cold-start* means that a new recommendable item (i.e., an article) was added that was not viewed or rated by many people yet. The problem of general data *sparsity* is often connected to this problem and in particular relevant for collaborative filtering approaches.

3.3.1. The Permanent Cold-Start Problem

In principle, many of the technical and domain-independent approaches that were proposed over the years to deal with cold-start situations can be applied in the news recommendation context. One general approach is to try to incorporate additional information (e.g., about the user’s context) into the recommendation process when little is known about the current user. In the news recommendation domain, for example, works exist that consider the user’s location [49, 132, 166], the time of the day [40, 49, 132], or demographic information [103]. Other examples of contextual information explored in the literature include the website (or domain) that the user has visited before landing on the news page or the type of the device used to access the news [40, 168].

Instead of considering only the limited information of the user’s history, Lin et al. [113] propose to construct an implicit social network of other users with similar reading behavior. The authors then suggest to determine a set of “experts” whose opinion is used to make recommendations for users for which only a short reading history is available. To identify these experts, they considered other users who have read the same articles in a given time frame, and they subsequently involved these users more heavily in the factorization process if they have a high probability of being a good proxy for a cold-start user. Along the same lines, Zheng et al. [186] propose to enrich the profiles of cold-start users with the profiles of a group of similar other users based on the general news topics they seem to be interested in.

An alternative to incorporating additional aspects related to the user is to consider certain features of the news articles themselves when assessing their relevance for a cold-start user, e.g., the freshness of the article or its general popularity. Also, one can default to a non-personalized strategy in case of too little information. The “YourNews” system [5], for example, first shows a list of recently published articles and starts the personalization process after the first article has been read. In another approach, Montes-García et al. [132] consider the geographical proximity of the news event to the user and the *credibility of a news source* as factors that can determine the relevance of an item to a cold-start user. In their case, the credibility of the news source was determined manually by a team of journalists. The context-tree approach by Garcin et al. [53] also takes the recency of the news items into account, but in addition factors the item’s popularity into the recommendation process. By design, their method is also suitable for one-time or first-time users, which are not uncommon for news platforms. Finally, Tavakolifard et al. [166] propose to use external data sources – in their case Twitter – to estimate an item’s current popularity more precisely and correspondingly obtain a better picture of its potential relevance for the current user.

For the *item cold-start* problem, adopting a content-based recommendation strategy already solves a major part of the problem. In content-based approaches, items are generally recommended based on the past content-wise preferences of individual users, which are usually determined by their past navigation or item rating behavior. Therefore, the required information about a news article can be extracted easily, e.g., from its keywords, making it instantly recommendable without the need for a detailed click history. Different alternative technical approaches are possible to infer the user preferences. Li et al. [108], for example, propose a hybrid approach to construct the user profiles which considers factors such as similarities between access patterns for different news items. In addition, their approach utilizes item properties that are specific to news, like the news content itself, as well as the user’s preferences for *named entities* appearing in the news item.

3.3.2. Data Sparsity Issues

When applying collaborative filtering approaches, the recommendation problem is often considered as a matrix completion task, with users and items as rows and columns, respectively, and matrix entries representing the relevance of items for users. In many domains, including the news domain, this matrix can be extremely sparse and only a tiny fraction of the matrix entries are known. Finding a sufficient number of similar users, for example, in a nearest-neighbors collaborative filtering approach, can therefore be very challenging [143]. Correspondingly, a number of approaches have been proposed in the context of news recommender systems to deal with the sparsity problem, e.g., by incorporating additional sources of knowledge, predominantly for collaborative filtering, but also for other approaches.

Mannens et al. [127], for example, propose a recommendation method that mitigates sparsity by complementing binary consumption values in the matrix with “potential consumption” values between 0 and 1 based on a collaborative filtering algorithm, which is then re-executed until the matrix is dense enough. Furthermore, to reduce the uncertainty introduced by this probabilistic approach, they also suggest to post-filter the news articles based on Linked Data sources. In contrast, Morales et al. [133] propose to use information from social network sites to deal with the sparsity problem. In their work, they analyzed the posts of users and their social friends on Twitter and generated a mapping between tweets and news articles based on the entities in the articles.

Leveraging information about the relationships between users is also the idea of the work presented in [114] and [113], which has already been mentioned in Section 3.3.1. In contrast to other approaches, the authors create an *implicit* virtual network, based on the users’ access logs. In their approach, which combines content information and collaborative filtering, the idea is to identify (two) expert users who are considered as influencers for users without a sufficiently large consumption history. These suspected influence relationships are then used to make the rating matrix denser.

Instead of using social network information, Lu et al. [120] propose to combine different forms of user behavior, context, and content features of the news items to find similar users in sparse-data situations. Specifically, they consider the user’s browsing, commenting, and publishing behavior as implicit feedback signals.

An even richer set of preference signals was combined in the News@hand system presented in [11, 24, 25]. The authors use ontologies and Semantic Web technology to represent the knowledge about users and items and utilize a

variety of preference indicators, such as page clicks, rating, and comments. To deal with the sparsity problem, they then propose a graph-based method to spread the information via the semantic relations of the network of concepts.

Li et al. [105] address the data sparsity issue by framing the recommendation task as a so called *contextual bandit* problem. In such a problem formulation [101], every recommendation task is represented as a choice between k possible “arms” of a multi-armed bandit. The recommendation strategy makes this choice only based on a context vector, for which Li et al. [105] chose a compact representation of users and articles (e.g., based on demographic information, location, and news categories). Afterwards, a reward is observed based on the user reaction (click). The advantage of contextual bandit approaches is that they can be used to balance exploration of the item space and exploitation of previous knowledge even in sparse data situations. Generally, one of several well-studied optimization strategies can be used to optimize the expected reward, i.e., in this case, the overall number of clicks. To evaluate the usefulness of their approach, the authors tested it on data sets with various sparsity levels.

Generally, when considering recommendation as a matrix completion problem, different forms of rating imputation and other forms of advanced statistical methods can be applied. Agarwal et al. [3], for example, analyze the correlations between different post-read actions (such as sharing, commenting, printing, or emailing article links) and are thus able to interchange the information across actions types to derive additional data points to feed into the recommendation process.

Overall, a number of techniques for dealing with the data sparsity issue in the news domain were presented over the years. Since data sparsity is such a central problem in news recommendation, many papers evaluate their algorithmic proposals – even when they are not explicitly focusing on data augmentation – on data sets of different densities to analyze their behavior in such situations.

3.4. *The Importance of Considering the Recency of News Articles*

Whatever algorithm is used for making personalized reading suggestions, the recency or “freshness” of an article will impact its relevance for the readers in many cases. Clearly, there are recommendation scenarios where a reader might be interested in older news as well, e.g., when investigating the development of a story over time or when looking for articles related to the currently read one. On typical landing pages of news sites, however, the recency of the articles is typically one important ranking criterion besides other factors such as the estimated general attractiveness of an article.

Technically, the recency of an article can be considered in the recommendation process at three stages. We can filter assumedly outdated news before computing relevance predictions or an item ranking (*pre-filtering*); we can incorporate the recency factor into the algorithms themselves (*recency modeling*); or we can filter or downrank articles after the main ranking process (*post-filtering*). However, one key design question in this context is how we balance the possible trade-off between article freshness and assumed relevance for the individual user.

Examples of works that apply a pre-filtering strategy are [39, 41], and [159]. In one of the news recommendation approaches by Desarkar and Shinde [41], for example, the authors selected the 100 most recent articles from the user’s preferred publisher websites and considered only these articles in the subsequent ranking process. Das et al. [39], on

the other hand, considered recency as one of several factors for item pre-filtering; other factors being, for example, the user’s language preferences or user-defined topics of interest.

In recency modeling approaches, the freshness of an article is considered simultaneously in conjunction with different other features of the item or the user. In principle, such an approach has the advantage that the different factors can more easily be balanced in an integrated way compared to when assumedly outdated items are filtered in a separate process and with separate heuristics. Examples of recency modeling approaches include [12, 148], and [182]. In the work by Pon et al. [148], freshness is considered along with a “multiple topic tracking” technique for users with several interests to train a news classifier. To this end, users are represented by a number of vectors instead of just a single average Term Frequency-Inverse Document Frequency (TF-IDF) vector, which is commonly used to weight the importance of keywords. In this way, short-term topic interest can be accounted for by giving preference to user profile vectors that have the highest similarity to the recently consumed news articles. Yeung and Yang [182] propose an approach that relies on a variety of features to predict the relevance of news articles (e.g., user features, item features, usage patterns, and context factors). The freshness of an article is considered as one of the item features and more recent articles receive a higher weight in the ranking process. Similarly, Bieliková et al. [12] integrated the publication time of an article as one of the factors in their tree-based model for computing the relevance of the articles.

Post-filtering based on article recency was applied, for example, in [108, 186], and [129]. In the method by Zheng et al. [186], news articles are first ranked according to their estimated relevance based on the interest groups the user belongs to. In a second step, the ranking is adjusted based on item freshness and popularity. A similar two-stage approach was adopted by Li et al. [108], which was already explained in detail in Section 3.3.1. A decay-based method has been proposed by Medo et al. [129]. In their work, they tested different score decay strategies for old news (no decay, medium decay, strong decay) and evaluated the impact of these strategies on the performance of the news recommender system.

Generally, to what extent older articles should be filtered can depend on different factors. First, in case of cold-start users with a very short reading history, we can resort to focusing the recommendations on articles that were recently added and which are relatively popular [49, 53, 166]. Such non-personalized recommendations might however lead to lower user satisfaction. Filtering out older articles can also be problematic if there are many readers who are interested in investigating developing stories and in reading older news articles on the same topic. Finally, whether or not it is good to focus strongly on the latest news in the recommendations can even depend on the individual user and his or her current context and goals. One might be interested to catch up with the latest news in the morning during the week, but be more interested in longer (and perhaps older) stories when visiting the news site on the weekend.

In any case, an open question seems to be how to select a threshold to filter old news or how to set a decay factor. When reviewing the literature, a number of different strategies and threshold values are used and it is not clear if an approach that is working well in one setting will also be suitable for another use case. In many cases, details about the specific settings are not reported in the research papers, which makes these works hard to reproduce.

Table 1 gives an overview of approaches that consider item recency as a relevant factor. In the last two columns, we show which thresholds were applied in case they were reported and which time windows were used to decrease

the relevance of an article based on its recency. We can observe significant differences. Some works only focus on articles of the last six hours [49], whereas others consider articles of the last twenty days [136]. Some authors add a decay penalty for every second since the article was published [38], whereas others do this only for each hour [183]. In many cases, however, this detailed information is unfortunately not reported at all.

Table 1: Overview of news recommendation approaches that consider recency as a factor. The first column categorizes the papers according to their recency strategy. The next two columns list the paper’s author(s) and its publication year. The fourth column shows the employed recommendation strategy (CB = Content-based filtering, CF = Collaborative filtering, H = Hybrid). The fifth column lists the data sets used in the paper. The last two columns show (a) how old articles can be at most to be considered recent and (b) the granularity of applying a decay factor, e.g., for every hour after the initial publication.

Appr.	Author(s)	Year	Strategy	Data Set(s)	Maximum Age	Decay Interval
Pre-Filtering	Das et al. [39]	2007	CF	Google News	–	–
	Frasincar et al. [50]	2009	CB	Hermes News Portal	1 year	user-defined
	Saranya and Sadhasivam [159]	2012	H	Indian news websites	–	–
	Desarkar and Shinde [41]	2014	CB	Plista data set	–	–
	Lenhart and Herzog [104]	2016	CB, H	Sport1.de	3 days	–
Recency Modeling	Lee and Park [103]	2007	H	Korean mobile service	24 hours	second
	Pon et al. [148]	2007	CB	Yahoo! News, Reuters	–	–
	Chu and Park [31]	2009	H	Yahoo! Front Page	–	–
	Fortuna et al. [49]	2010	CB	–	6 hours	hour
	Li et al. [112]	2010	H	Digg, Reuters	–	–
	Yeung et al. [183]	2010	H	–	24 hours	hour
	Yeung and Yang [182]	2010	H	–	24 hours	hour
	Gershman et al. [58]	2011	CB	–	–	hour
	Guy et al. [65]	2011	H	Lotus Connections	–	–
	Li et al. [110]	2011	CB	Several news websites	–	–
	Bieliková et al. [12]	2012	CB	SME.SK	–	–
	Morales et al. [133]	2012	H	Yahoo! News / Toolbar, Twitter	2 days	–
	Wen et al. [176]	2012	H	CNN, CBC, Yahoo! News	–	hour
	Garcin et al. [53]	2013	CB, CF, H	Tribune de Genève, 24 Heures	–	–
	Montes-García et al. [132]	2013	H	Manually created data set	–	–
	Tavakolifard et al. [166]	2013	H	PolarisMedia, Twitter	–	–
	Darvishy et al. [38]	2015	CF	Several news websites	–	second
	Kliegr and Kuchař [95]	2015	CF	Plista data set	–	second
	Muralidhar et al. [136]	2015	CB, CF, H	Washington Post	20 days	day
	Lak et al. [100]	2016	CB	–	–	–
Post-Filtering	Medo et al. [129]	2009	CF	–	–	–
	Li et al. [108]	2011	H	Several news websites	–	minute
	Zheng et al. [186]	2013	H	Several news websites	–	–

3.5. Considering Quality Factors Beyond Prediction Accuracy

The main optimization goal of researchers in academia is to accurately predict the relevance of a news item for a user. Most commonly, classification or ranking accuracy measures are used for that purpose (see Section 4 for more details). However, in practice, predicting the relevance of an item is in many cases not enough. If, for example, a user is interested in politics and has shown strong interest in articles about an ongoing presidential election in the past, recommending more articles about this topic is probably a good choice. However, recommending *solely* articles about the election, or *solely* about politics, might be too monotonous for users and would probably not lead to high user engagement in the future. In case of a news aggregation site, it is furthermore important that the recommended news are not too similar to each other. Presenting three articles from three different sources about, e.g., the same plane accident might be of little value for users.

Therefore, one additional challenge, besides accurately predicting whether an article is relevant for a user or not, is to take additional quality factors into account. In the recommender systems literature, *diversity*, *novelty*, and *serendipity* are often considered as such quality factors that have to be balanced with prediction accuracy (see, e.g., [29]).

3.5.1. Diversity

Users of news recommender systems can be interested in a variety of topics. A recommender system should therefore be able to address these varied tastes and generate diversified recommendation lists [171]. Empirical research suggests that increasing the diversity of the recommendations can lead to a better quality perception by users [48, 151, 188].

An example for a work that considers diversity in the news recommendation domain is presented in [108], which we already mentioned earlier in the context of recency filtering and cold-start techniques. In the proposed two-stage approach, articles were clustered in the first phase, and the recommendations were personalized and adapted, e.g., with respect to recency, in the second phase. In both stages, the authors propose to add some level of diversification. According to the authors, the first-level clustering strategy implicitly introduces some level of *topic* diversity, while the second stage implements a greedy approach that explicitly minimizes the similarity among items in a recommendation set. In the latter case, the diversity was measured by computing the average pairwise (dis-)similarity of documents, and an experimental evaluation showed that their approach was beneficial in terms of both accuracy and diversity when compared to the works presented in [31, 39, 105, 115].

Later on, Li and Li [107] followed an alternative and more advanced approach to diversify news recommendations in the context of a learning-to-rank optimization scheme. In their work, they propose a framework that relies on a hypergraph to model the relations between different “media objects”, which included users, news articles, topics, and named entities. Using the same experimental setup as in their previous work [108], the authors were able to demonstrate further improvements both in terms of accuracy and diversity.

Typically, there is a trade-off between accuracy, e.g., in terms of classification accuracy, and diversity, e.g., in terms of intra-list similarity [188]. To address this issue, Desarkar and Shinde [41] analyzed two possible approaches of balancing these goals in the news domain and tested them for a special privacy-targeted scenario where no past

interaction history is available for the users and the relevance of each candidate item is consequently estimated only by its popularity. In their case, diversification is described as a bi-criteria optimization problem where both the relevance and the dissimilarity between the news objects should be high.

Outside the domain of news recommendation, a number of works have been published over the past ten years on the topic of diversifying recommendations [29]. Besides the question of how to balance diversity with accuracy, researchers investigated *to what extent* the level of diversification should be determined by the preferences of the individual user [162] or how diversity preferences can depend on time aspects [102]. In the news domain, these aspects have not been investigated much yet, even though some of these aspects – like the consideration of the time of the day when recommending [40] – might have an effect on the user’s short-term diversity preferences.

3.5.2. Novelty

Novelty, as a quality criterion for recommendations, was defined in terms of the *non-obviousness* of the item suggestions by Herlocker et al. [68]. Informally, novel items are, according to their definition, those that the user has not seen yet, but which are relevant to him or her. An alternative definition was provided by Vargas and Castells [172], who defined novelty as the inverse general popularity of an item. The assumption here is that less popular items are more likely to be unknown to users and recommending long-tail items will, in general, lead to higher novelty levels. Several ways of numerically quantifying the degree of novelty are possible, including alternative ways of considering popularity information [187] or the *distance* of a recommendable item to the user’s profile [137, 154].

Like diversity and accuracy, novelty and accuracy aspects can represent a trade-off. In their work on news recommendation, Garcin et al. [53] relied on the definition by Herlocker et al. [68] and tested different configurations of weighting the two factors to achieve both high accuracy and high novelty. Rao et al. [154] followed an approach where the novelty of an item is defined by its distance to the user’s profile. Specifically, they compared the distance of the concepts appearing in the user profile and the news article, where the taxonomy was constructed from encyclopedic knowledge. The evaluation results of their user study show that this approach cannot only be used to increase accuracy but also reduce the perceived redundancy among news articles.

Generally, as for diversity, the “optimal” degree of novelty can depend on the user’s current situation and context, as discussed recently by Kapoor et al. [89] in the context of the music domain. Determining the right balance of recommending (a) things that are most likely relevant for the user and (b) recommending items that help the user discover new things remains challenging. Limited research on these aspects exists in the news recommendation domain. As discussed in Section 3.3.1, the news domain has the special characteristic that most of the recommendable items are usually new and unknown to the readers. In terms of some of the definitions from above, all recommendations are novel. Therefore, novelty can probably not be determined on the basis of the individual item itself, but rather in terms of the content the news item is about, e.g., based on whether or not the user already knows about the real-world event the news item is reporting on.

3.5.3. Serendipity

Serendipity is sometimes also mentioned as a desirable quality characteristic of recommendations. Herlocker et al. [68] describe serendipitous recommendations as those that help users find surprising and interesting items that they might not have noticed otherwise. Ge et al. [57], based on an earlier approach by Murakami et al. [135], defined serendipitous recommendations as items that are not only novel but also positively surprising for the user and proposed a generic metric based on the concepts of *unexpectedness* and *usefulness*. Zhang et al. [185] finally proposed a related conceptualization in the context of a music recommendation application in which the level of serendipity is determined based on the distance between the recommendations and the user's *expected content*.

To our knowledge, only limited research on serendipity aspects of the news recommendation problem exist so far. In principle, one can apply techniques based on Latent Dirichlet Allocation (LDA) models – as proposed by Zhang et al. [185] for music recommendation – to the news domain and focus on recommending news items that are content-wise dissimilar to the user's typical latent topics. However, whether or not this will lead to higher levels of user satisfaction in this domain, still has to be explored.

Overall, one key question today is how to quantify the level of serendipity of a recommendation list and how such serendipity measures would be related to existing novelty measures. Again, adding serendipitous items to the recommendations comes with the risk of decreasing the average relevance of the recommendation lists. Furthermore, Kotkov et al. [99] recently discussed the problem that today in many (general RS) research data sets used for offline experimentation not enough information about the user's context (e.g., time, location, or mood) is available. The user's context can however have a significant impact on the user's perception of serendipitous item recommendations. In our view, more research is therefore required also outside the field of news recommendations to better understand the concept of serendipity in recommender systems and how to measure it.

3.6. Scalability Issues

Scalability is the last of the main challenges of news recommendation that we discuss in this section. Large news websites like Google News can have hundreds of millions of users per month. At the same time, the number of articles that can be searched and recommended can be huge as well, with thousands of new items appearing every day.⁷

A common approach to deal with these huge amounts of data is to apply clustering techniques of different types. Clustering can in many cases help to speed up the computations; it can however also lead to compromises with respect to the accuracy of the relevance predictions. In addition to such algorithmic approaches that, e.g., allow us to do the computations on a more coarse-grained level, researchers often propose to rely on distributed computing mechanisms and to execute the calculations on multiple servers in parallel.

The application of various forms of clustering techniques for different aspects of the news recommendation problem has been proposed, e.g., in [39, 120, 129], and [174]. In some of these cases, the main goal is to search for similar users (neighbors) only within a smaller cluster of users without the need to scan the entire database. In the case of

⁷Google News archives news articles after 30 days, meaning they will only appear in archive search results.

the Google News system, as described in [39], the authors use a combination of (MinHash) clustering and distributed computing based on the MapReduce framework to make the approach scalable. Lu et al. [120] also combine MapReduce and a clustering technique to increase the scalability. Specifically, they use a “Jaccard–Kmeans” based clustering technique where the similarity of users is computed in multiple dimensions, e.g., based on their past behavior and the content the users preferred in the past. Other examples of works on news recommendation that rely on distributed computing include [156] and [159].

Besides clustering the users, one can also try to cluster the available news articles and group them based on their features and the user’s preferences. In [108], for example, the authors propose a two-stage recommendation approach in which they apply different clustering techniques to create a hierarchy of news clusters. This hierarchy can then be efficiently traversed from top to bottom to find a set of news articles similar to the user’s reading interest.

Medo et al. [129] adopt a different approach to achieve scalability in the context of a news platform where users can post news items. In their method, they construct a local neighborhood network of users based on the similarity of their rating behavior (likes/dislikes). In contrast to traditional k-nearest-neighbor (kNN) approaches, they do not store the whole neighborhood, but only the most similar neighbors for each user. Once a user posts a new article on the system, it is propagated along the directed edges of the graph to their “followers”, who can spread it further by liking it. The final ranking of the news items is then determined by a score, which is based on the like/dislike statements of the target user’s neighbors. In a follow-up work, Wei et al. [174], expand the method by adapting it to take the users’ reading patterns into account. Additionally, they propose different stochastic solutions to the problem of updating the neighborhood that improve the scalability of their approach further, e.g., by randomly selecting new neighbors.

Many of above-mentioned approaches describe how to achieve scalability on a theoretical level or give an account of how real-world systems implement large-scale operations, as in the case of [39]. What sets the news domain apart from other domains that are investigated in recommender systems research, however, are the live news recommendation challenges organized by Plista, which we will explain in detail in Section 4.2. Such challenges give researchers the opportunity to test the scalability of their approaches “in the wild”. An example is the work by Doychev et al. [45], whose content-based recommender system addresses the requirement of the Plista challenge to respond quickly to a large amount of recommendation requests using a variety of state-of-the-art data processing and management technologies. These range from a two-tiered load balancing stage (Nginx, Tornado), to query execution using a non-relational database and full-text search engine (MongoDB, ElasticSearch), to a recommendation cache based on an in-memory database (Redis).⁸ In practice, this results in average response times of about only 3 ms.

In conclusion, the amount of literature on this issue shows that the scalability problem has been addressed to a certain degree by the news recommender research community. Nonetheless, because of the constant increase in the amount of available online news articles, scalability remains one of the key problems in this domain, and, thus, more sophisticated solutions and further scalability-oriented evaluations are necessary.

⁸<https://www.nginx.com>, <http://www.tornadoweb.org>, <https://www.mongodb.com>, <https://www.elastic.co>, <https://redis.io>

3.7. Discussion

The review of recent works in this section shows that during the last ten years substantial efforts went into the development of new techniques to deal with the particular challenges of news recommendation. Figure 2 shows the distribution of topics and challenges that were addressed in the reviewed works. The diagram shows that the largest number of papers was devoted to the problems of user profiling, which is a common challenge for all recommender systems domains. Additionally, a number of papers focus on the specific problems of data sparsity and user and item cold-start. This is in fact expected, given that news recommendation systems face a *continuous item cold-start* situation, and personalization techniques that work well, e.g., for movie recommendation, do not always work well under such circumstances. Also, it is not very surprising that recency aspects received a lot of interest, due to the unique short-livedness of the items in the domain.

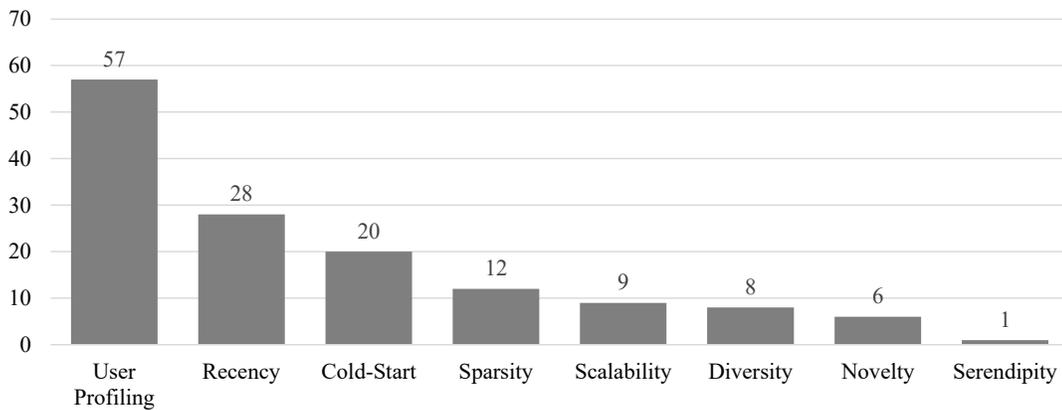


Figure 2: Distribution of research challenges (other than accuracy) addressed by news recommendation papers published in the last decade. Each paper can fall into multiple categories.

Beginning around the year 2011, we observe more and more papers on beyond-accuracy quality aspects like diversity, novelty, and serendipity, i.e., topics that were not strongly in the focus of researchers in earlier years. This recent trend can be considered as a positive development, since it has already been shown in other domains in which recommenders have been applied that focusing solely on accuracy measures to compare algorithms can be insufficient and potentially misleading [80, 128]. Nonetheless, more work is needed to better understand how to balance the sometimes competing goals of achieving high prediction accuracy and, for example, list diversity in parallel.

Looking at the different challenges discussed in this section so far, we can identify a number of additional areas in which more work is still required. In the context of user modeling, for example, a number of papers differentiate between long-term preferences and short-term interests. In our view, it is however not fully clear yet how these aspects should be balanced, how we can deal with an interest drift over longer periods of time, or how we should deal with *exceptional* events and specific contextual conditions, which induce a probably only short-lived interest at the user's side.

Furthermore, in terms of user modeling, today's news sites also provide only limited ways for users to explicitly inform the recommender about their preferences or to give the system feedback on individual recommendations. More research is therefore required to put users into control of their recommendations. Today, Amazon.com implements a small set of such features on their website and provides users with explanations for the recommendations. Many users however seem to make limited use of these functionalities [88].

From the data perspective, more and more types of information are becoming accessible regarding the user's current context, e.g., the current location, weather conditions, other people nearby etc. Also, continuously new data sources like DBpedia and other sources containing structured information, e.g., in terms of ontologies, become available. Some works exist which try to leverage structured information about named entities or other concepts mentioned in the news articles to find similar articles. Still, we see a lot of remaining potential to build even better news recommenders that rely on such additional data sources.

As discussed in the section on aspects of item freshness, there is a number of further open questions. Looking at existing works, it is not fully clear yet how quickly older news should be discarded in the recommendation process. In fact, this aspect might even depend on the category of the news items. Some breaking news article might be outdated with the next incoming article on the same topic. An article about a sports event might not be very relevant anymore after the weekend. In contrast, an in-depth news report on a certain topic might be an interesting read even after months. In addition, whether the focus should be more on recent articles or not could also depend on the preferences of the individual user and the time of the day. Some first steps toward a better understanding of these aspects have been taken, but in our view a substantial amount of further investigation is still required.

4. Evaluating and Benchmarking News Recommender Systems

Having reviewed the general challenges and current algorithmic approaches in news recommendation, we will discuss methodological questions related to the evaluation of news recommendation approaches in this section. To this end, we will review today's academic practice of evaluating and benchmarking different algorithms and then reflect on possible limitations in this context caused by the availability of public evaluation data sets. Finally, we will compare open source frameworks that can be used for news recommendation evaluation.

4.1. Evaluation Approaches

Recommender systems are typically evaluated in one of the following three approaches: through offline experimentation and simulation based on historical data, through laboratory studies, or through A/B (field) tests on real-world websites. Sometimes, parts of the theoretic properties of the proposed algorithms can be demonstrated using a formal proof. Figure 3 shows the distribution of the different evaluation approaches for the examined papers.

Our analysis shows that offline experimentation is the predominant way of evaluating and comparing different algorithms in the field of news recommendation. User studies are done to a much smaller extent and field tests (A/B studies) are still comparably rare, even though we observe more of such papers in the recent past. Comparable analyses

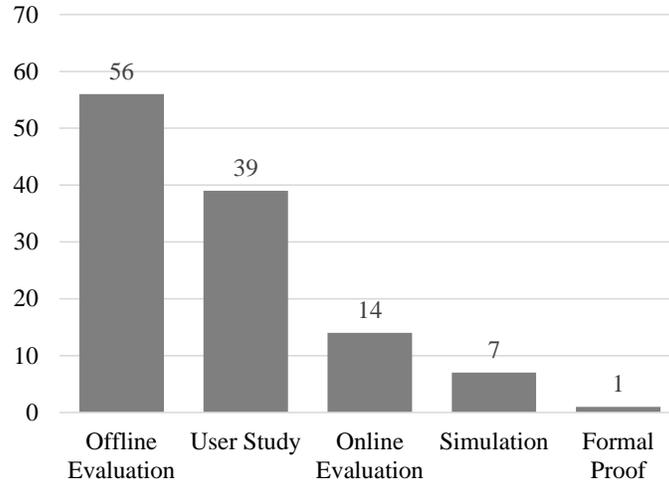


Figure 3: Distribution of evaluation approaches used in news recommendation papers published in the last decade. Each paper can fall into multiple categories.

have been conducted by Jannach et al. [85] on recommender systems in general and by Beel et al. [9] in the domain of research article recommendation, which both revealed a similar distribution of evaluation methods.

Figure 4 shows which measures were used by researchers to quantify the performance and characteristics of the evaluated approaches. The results show that most of the works focus on accuracy measures of one type or another, including information retrieval measures like precision and recall, rank-based measures like *Mean Reciprocal Rank* or *Normalized Discounted Cumulative Gain*, or prediction measures like the *Root Mean Square Error*.

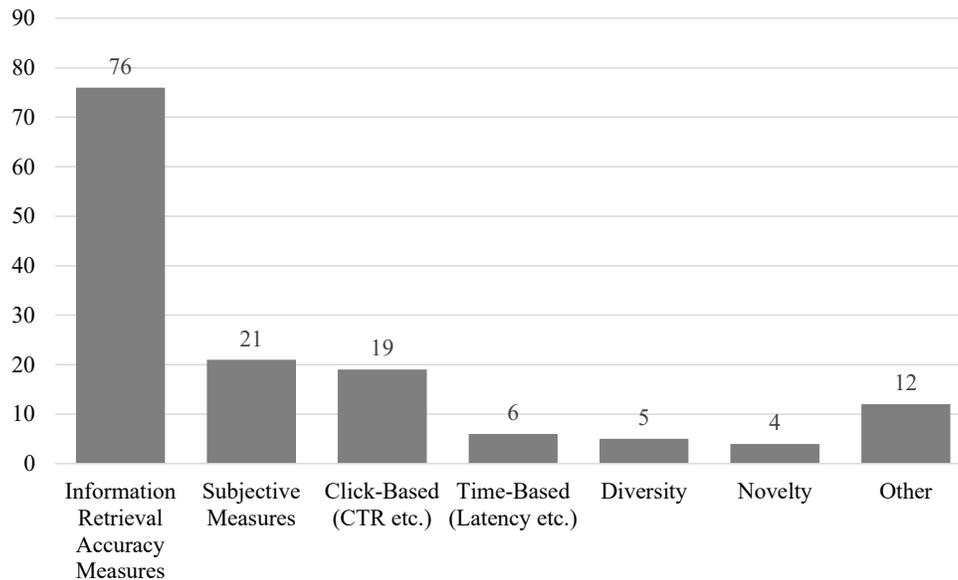


Figure 4: Distribution of evaluation measures/metrics used in news recommendation papers published in the last decade. Each paper can fall into multiple categories.

Among the 112 surveyed papers that introduce one or more new algorithms, only 19 used click-based metrics (such as the Click-Through-Rate) as an evaluation measure, even though such metrics are typically considered as success measures in practice, e.g., in the advertising industry. Measuring click rates, however, requires access to past navigation data and/or to a real-world system, which might explain the infrequent usage of these measures. An even smaller number of papers (about 10 out of 112) also considered aspects like diversity, novelty, or serendipity in their evaluations. Scalability, even though an important problem in practice, was also in the focus of only a few research works. Note in this context that while some papers take aspects like diversity or novelty into consideration in the design process of their algorithm, they do not explicitly quantify any improvements w.r.t. these aspects with standard metrics in their experimental evaluation. A complete list of challenges addressed by each paper and the metrics that were used in the evaluation process can be found in the appendix in Table A.3.

4.2. Research Data Sets

The types of evaluations that can be done in academic research often depend on the characteristics of publicly available data sets, e.g., their size or the amount of user and item information. In the following, we will review a number of such data sets used in the literature and discuss their limitations in terms of what kind of research can be done with them. A list of the papers that present or analyze news recommendation data sets can be found in the appendix in Table A.4.

Yahoo!’s Data Sets. A number of research works in news recommendation are based on different data sets provided by Yahoo! (see, e.g., [124, 126, 139]). One prominent example is the “Yahoo! Front Page Today Module (R6A)” data set, which contains information about news articles that were displayed on the “Featured Tab” on Yahoo!’s landing page and (a fraction of) the click data for these articles.⁹ The data was collected during the first ten days of May 2009 and contains over 45 million user visits. For each visit, the timestamp and the displayed article ID is available, as well as a six-dimensional feature vector that was extracted from the item-user interactions via a specific form of regression analysis [32].

Later on, an alternative data set taken from the same website was published (named R6B), which contains data collected during fourteen days and comprising over 130 anonymized user features like age, gender, extracted behavioral variables etc. The two Front Page data sets were specifically constructed in a way so that an unbiased evaluation of recent explore-exploit based techniques (using contextual bandits) is possible [31, 106].

In 2016, Yahoo! released a huge “News Feed” data set containing interactions of about 20 million users collected on different Yahoo! websites over the period of three months in February 2015. Again, different user and item features are part of this data set, which contained about 110 billion lines. Currently, however, the data is no longer available, perhaps due to issues related to the privacy of users.¹⁰

While very valuable for researchers, one of the limitations of the still available “Front Page” data sets is that the data collection period is very short (i.e., at most 14 days), making it impossible to conduct research on personalization

⁹<https://webscope.sandbox.yahoo.com/>

¹⁰<https://motherboard.vice.com/read/yahoos-gigantic-anonymized-user-dataset-isnt-all-that-anonymous>

based on long-term user models. Also, given that the data was sampled only during one specific period, certain events that happened exactly during this period might have lead to certain biases in the data.

SmartMedia Adressa Data Set. Recently, a click log data set with approximately 20 million page visits from a Norwegian news portal as well as a sub-sample with 2.7 million clicks (referred to as “light version”) were released.¹¹ The data set was collected during ten weeks in the first quarter of 2017 and contains the click events of about 2 million users and about 13 thousand articles (light version: one week, 15 thousand users, one thousand articles). However, according to Gulla et al. [64], the clicks exhibit a strong concentration bias on only small subset of the items. In addition to the click logs, the data set contains some contextual information about the users such as geographical location, active time (time spent reading an article), and session boundaries. Furthermore, subscribed users who pay for access to the “pluss” category of articles can be identified; they make up 10.2 % of the user base. This subscription information could enable future investigations that focus on the willingness to pay for news content and the effect of a paid subscription on the consumption and interest patterns of users. Finally, even though the public data set contains some basic information about the news articles themselves in the form of keywords, the article content is only available upon request.

Outbrain Data Set. In the context of the 2016/2017 Outbrain Click Prediction challenge hosted on Kaggle¹², participants were asked to predict clicks on targeted content ads. The data set that was used for this challenge contains roughly 2 billion page visits for about 560 US publishers generated by a user base of approximately 700 million visitors during two weeks in June 2016. Even though the challenge focused on ad click prediction, the main click log of page visits, which contains the unique item and user IDs, timestamps, geolocations, etc., can be used for news recommendation evaluation as well. In contrast to other news recommendation data sets, however, the content of the news articles is not released. Instead only keywords, categories, and entities are provided as article metadata in the form of IDs. In fact, since even the publisher can only be identified by ID, researchers cannot be sure which type of news (or other textual content) the respective publishers provide. Therefore, while the possibilities to evaluate content-based recommendation schemes on this data set are rather limited, the massive amount of data offers a great opportunity to test more complex machine learning models.

Proprietary and Non-Public Data Sets. Unfortunately, much research in the field of news recommendation is based on proprietary and non-public data sets. This in particular holds for research works like [39, 55, 93], which investigate the relationship between offline performance (measured in terms of accuracy metrics) and online success (measured in terms of click-through-rates). Often, additional application-specific ways of evaluating the performance of the algorithms in the live setting are used in these papers. In some cases, like in [55], the used real-world data sets were again collected only during a short period of time and contain a comparably small number of user interactions.

¹¹<http://reclab.idi.ntnu.no/dataset/>

¹²<https://www.kaggle.com/c/outbrain-click-prediction/data>

In some cases, researchers who use their own proprietary data sets in addition demonstrate the effectiveness of their methods on non-news data sets (e.g., from MovieLens as done in [39]). Such data sets are however often very different in terms of their characteristics (e.g., with respect to their sparsity or the fact that there exists no constant item cold-start situation), and it therefore remains unclear whether or not evaluations on such non-news data sets are truly representative of the news domain [142].

The Plista Data Set and the NewsREEL Challenge. In 2013, the *Plista*¹³ data set was released [90] for research purposes. Plista is a German company that provides content and advertisement recommendations for a large number of websites. The data set contains different types of activities by news editors (CREATE, UPDATE) and news readers (IMPRESSION, CLICK) recorded for about one dozen web portals during June 2013. The data set is comparably large and contains about 80 million article impressions and about 1 million clicks on recommended items. A small number of features of the users and the news items are provided as well. The various challenges of news recommendations mentioned in previous sections – e.g., data sparsity, recency biases, very few recorded interactions for many users – become once again apparent when analyzing the data set.

In addition to making the data set publicly available, Plista holds regular contests (the CLEF-NewsREEL challenge¹⁴) in which researchers can test their recommendation algorithms in a live environment, which is a unique opportunity in the entire research field. During the contest, participants receive recommendation requests, which they have to answer within a limited time frame. If the recommendations are generally considered “recommendable”, e.g., because they are not too old, they are served to real users, and feedback is given to the participating researchers in case their recommendation was clicked. One main insight of these contests is that simply recommending those items that have been published within the last few minutes and which have been popular in this time frame seems to be a hard-to-beat approach when the click-through-rate is taken as a success measure [121]. A recent analysis of the performance of other approaches that took part in the NewsREEL challenge can be found in [46].

4.3. Open Source News Evaluation Frameworks

As discussed above, the availability and quality of public news-related data sets can be a problem when evaluating news recommendation algorithms. However, even if researchers have access to a data set that fulfills their requirements, they need to compare their own algorithm implementation with other well-known approaches in a reproducible way to demonstrate the viability of their strategy. To this end, a small number of open source frameworks are available that are either specifically targeted towards evaluating news recommendation algorithms or can be adapted to this purpose.

In theory, any recommender system evaluation framework, such as the popular LensKit¹⁵ or LibRec¹⁶ libraries, could be used to benchmark news recommendation algorithms. However, as discussed earlier, one unique aspect of

¹³<https://www.plista.com/>

¹⁴<http://www.clef-newsreel.org>

¹⁵<http://lenskit.org/>

¹⁶<https://www.librec.net/>

the news domain is the importance of article freshness. Since most general purpose RS evaluation frameworks use cross-validation or sliding window schemes, the freshness problem cannot be treated adequately. However, a few frameworks exist that implement a “streaming” evaluation scheme, which aims to simulate real-life recommendation scenarios more closely. In such an evaluation scheme, clicks in the data set are processed in a chronological way, and after every click in the data set, algorithms have to generate a recommendation list that is then compared with the actual next clicks of the user.

Table 2 shows a summary of the above-mentioned streaming evaluation frameworks for news recommendation algorithms. Of the four available frameworks, only one is designed explicitly for the news domain and, as such, includes special-purpose algorithms for news recommendation. The variety of algorithms in the IR & E framework, however, is rather low, with only a few baselines and one contextual bandit approach. In contrast, the general-purpose frameworks FLURS and ALPENGLow offer a number of more complex algorithm implementations, such as incremental learning versions of kNN and matrix factorization. However, as discussed earlier, these general-purpose collaborative filtering strategies can often not deal with new items quickly enough. IDOMAAR does not provide any algorithms of its own except for a few simple baselines. The idea behind this framework is instead to offer a standardized evaluation environment. In contrast to the other frameworks, where algorithms are implemented as drop-in classes in Python or C++ respectively, in IDOMAAR each recommendation strategy has to be implemented as a web service endpoint to provide maximum separation between implementation and evaluation and to offer flexibility in the programming language, at the possible cost of efficiency. Finally, STREAMINGREC, an open-source news recommendation framework developed by the authors of this paper, offers a range of baseline approaches and more advanced algorithms in an extensible Java environment. Similar to Idomaar, StreamingRec’s evaluation procedure aims to simulate real-life click behavior in the news domain, but without the overhead of implementing algorithms as web-service endpoints.

In our view, in particular IDOMAAR and the more recent STREAMINGREC framework represent the most valuable starting points for researchers. IDOMAAR has the advantage of defining language-agnostic interfaces. The framework does, however, not include any pre-implemented algorithms. STREAMINGREC, on the other hand, while being tied to the Java language, offers a wide range of pre-implemented news recommendation algorithms that can be used as baselines in comparative evaluations. Many of these algorithms are also capable of immediately considering new articles without costly model retraining.

4.4. Discussion

Research on the topic of news recommendation was historically often subject to a number of constraints and limitations with respect to the evaluation methodology. Up to the very recent past, only very few public data sets containing user-item interactions for news specific applications were available. While some of the data sets are comparably large,

¹⁷<https://github.com/umeshksingla/news-recommend-ire>

¹⁸<https://github.com/takuti/flurs>

¹⁹<https://github.com/crowdrec/idomaar>

²⁰<https://github.com/rpalovics/Alpenglow>

²¹<https://github.com/mjugo/StreamingRec>

Table 2: Open source frameworks for news article recommendation

Name	Language	Algorithm implementation	Available algorithms	Available metrics	News specific
IR & E ¹⁷	Python	Python class	Baselines, Contextual Bandit	Hitrate, NDCG, MAP	Yes
FluRS ¹⁸	Python	Python class	Baselines, kNN, BPR, MF, FM, Matrix Sketching	Precision, Recall, MAP, AUC, MRR, MPR, NDCG	No
Idomaar ¹⁹ [160]	Ruby, Python	Web service endpoint	Baselines	Precision, Recall, MAP, ARP, MRR, NDCG, Coverage, Shannon entropy, Category diversity	No
Alpenglow ²⁰ [52]	C++, Python	C++ class	Baselines, Item-item, kNN, MF, Hybrids	Precision, Recall, NDCG	No
StreamingRec ²¹	Java	Java class	Baselines, Lucene, Co-Occurence, Seq. Patterns, kNN, BPR	Precision, Recall, F1, MRR, Nb. of rec's, Efficiency	Yes

many only cover the user interactions of a few consecutive days, which makes learning of long-term preference models more or less impossible.

A further practical problem in this context is that researchers often use subsamples of the larger data sets to test their often computationally complex methods. Since these subsamples or the specific procedures to create them are usually not publicly shared, the reproducibility of the research results often remains limited. In some cases, researchers resort to data sets from other domains (e.g., movies) to test their news recommendation methods. Such approaches however come with a number of limitations as discussed in [142]. In particular, the aspects of recency and general popularity play a much less pronounced role in other domains, which makes it unclear to what extent algorithms that work well for other domains are suitable for news recommendation.

To our knowledge, only one data set in the news recommendation domain is available which contains explicit feedback on news articles.²² The YOW data set²³, which is described in detail by Özgöbek et al. [142], contains ratings for about 380 articles, which were collected from 21 paid users over a period of four weeks. The data set is therefore comparably small and in addition does not contain information about the news articles themselves, which makes experimentation with content-based or hybrid approaches infeasible.

From a methodological perspective, classical IR accuracy measures like precision or recall dominate the research landscape. These measures are often, like in the CLEF-NewsREEL challenge, based on click-through data. While

²²Non-public rating data sets for news recommendation were for example used by Cantador et al. [24] and Ji et al. [86].

²³<https://users.soe.ucsc.edu/~yiz/papers/data/YOWStudy/>

these measures are therefore based on real-world user interactions, a number of limitations remain. First, it is not fully clear today whether being able to successfully predict clicks in an offline experiments will translate into long-term business success in a deployed application [55], which can be a general problem of offline evaluations [10, 78]. Past live studies in the domain of mobile game recommendation show for example that recommending comparably popular items leads to high click-through-rates in an online shop but not to the strongest effect in sales [79]. In addition, the click-through-rate might only be an indicator of the user’s attention, but not necessarily a sign of genuine interest in the topic and thus proof of the recommender system’s success. A second potential issue in the context of measuring in offline experiments is that high values for precision and recall can in different domains be achieved by recommending mostly popular items [80]. In some cases, and depending on the specific variant of the measures, popularity-based baseline methods are often hard to beat by more complex methods [36]. In practice, however, users can easily recognize that the recommendations of such a popularity-based method are not based on the currently viewed news article and not tailored to their personal viewing history, which might lead to limited acceptance of the recommendations. This intuition is also supported by the experiments in Garcin et al. [55], where a popularity-based method was the winner in the offline evaluation, but performed poorly in an online scenario.

Finally, a number of protocols variations can be applied in offline experiments. The simplest approach is to apply a time-agnostic cross-validation protocol, which is often done in the literature to evaluate recommenders, e.g., in the movie domain. However, splitting up user interactions into training and test sets without considering the consumption order disregards the importance of item freshness in the news domain. Therefore, the obtained results might be not very representative of an algorithm’s true performance. Another approach is to split the data using one of several time-aware evaluation protocols (see [20] for an overview), and use several consecutive days for training and the last one for testing. The outcomes of such an evaluation might depend on the specific protocol variant, which, as a result, might limit the comparability of the results of different researchers. Furthermore, the results of the 2017 CLEF-NewsREEL challenge indicate [122] that being able to very quickly react to incoming news articles, e.g., within a few minutes, is very important to achieve high click-through rates. It therefore remains unclear how effective complex recommendation algorithms are that can only be re-trained overnight.

5. Conclusions and Future Directions

Our review has shown that news recommendation is an active topic of research and that in recent years significant advances have been made in different directions. In this concluding section, we summarize some key challenges (as discussed in more detail in Sections 3.7 and Section 4.4) and sketch possible directions for future work.

In our analysis of *algorithmic* approaches in Section 3, we found that content-based methods are quite frequently used in the academic literature, i.e., in almost half of the papers. The observations from our experimental evaluation and from news recommendation challenges in the real world, however, suggest that relying solely on content-based techniques can be insufficient. Factors like general article popularity and recency are highly important in the domain and collaborative-content-based hybrid techniques, as used in the field study by Kirshenbaum et al. [93], are therefore the method of choice when it comes to optimizing IR accuracy measures or click-through rates. However, in this

context, more research is needed to better understand how factors like diversity, novelty, or popularity contribute to the quality perception of users, as focusing only on accuracy measures might lead to mostly non-personalized recommendations and little discovery. Better personalization is however only possible if more is known about the preferences of the individual users. It is therefore important to further investigate the relative importance of long-term and short-term (contextual) preference models. Research in that direction is however still hampered by the lack of data sets that contain longer user interaction histories.

From an algorithmic perspective, questions regarding scalability were addressed in a number of the reviewed papers. Nowadays, the tools for collecting and storing big amounts of interaction data seem to have reached a high enough maturity level to support basic data processing in real-time. However, training and continuously updating the machine learning models (as done in other domains, e.g., in [81, 82]), remains a challenge, in particular when the models are based on deep learning strategies. In the news domain, new articles that are possibly relevant for a larger number of users can appear several times a day, and these articles have to be immediately considered by the recommendation methods. Otherwise, the performance of such models can be limited. This calls for more research on hybrid methods that are, for example, based on sophisticated long-term models and computationally efficient short-term adaptation heuristics.

With respect to the *data* perspective, today’s research data sets comprise a number of user features and item features, and quite recently, two large news recommendation data sets have been made publicly available. These new data sets represent a promising starting point for further *reproducible* research into scalability, sparsity, and user modeling aspects of news recommendation. Additionally, researchers are nowadays able to test their approaches on a variety of data set and publishers, which can give further insight into the generalizability of newly proposed algorithms; a crucial aspect of RS design, which has not yet been given enough attention in the literature. At the same time, with the continuous development of the Social Web, the Semantic Web, and the Internet of Things (IoT), a variety of additional data sources will continue to become available in the future. These data sources, for example, include information about the user’s behavior on Social Web platforms, additional information about the relationships between concepts and named entities in the news articles, or contextual information about the user’s situation derived from smartphone sensors or future IoT devices. A number of research avenues, therefore, exist on how to leverage these possibly huge amounts of additional types of information in the news recommendation process.

Looking at *methodological aspects*, we found that IR measures and click-through rates are the prevalent evaluation measures in the academic literature. Whether or not optimizing algorithms for such measures leads to the best *value* for the different stakeholders (consumer, news provider, news aggregator) has not been investigated to a large extent in the recommender systems literature in general. More research and more public data sets that include explicit information about the business value are therefore required to understand, e.g., the economic perspective of news recommenders (“RECO-nomics”) [78] and how to balance the interests of the involved stakeholders. In this context, extended co-operations with news publishers or third-party recommendation providers are also required to better understand the business value of news recommendations.

Appendix A. Tables

Table A.3: Overview of all surveyed papers that propose one or more news recommendation approaches. The first two columns show the paper’s author(s) and its publication year. The third column lists the paradigm(s) (i.e., the recommendation strategy) employed by the algorithm(s) in the paper. The fourth column names the challenges that the paper addresses. Note that we do not mention “Accuracy” explicitly as a challenge, since nearly every paper addresses this issue, and that we leave out challenges that are only addressed by very few papers, such as “Manipulation Resistance” or “Accessibility”. The fifth column lists the data sources, in case they have been disclosed. Column six shows the evaluation metrics used in the paper and column seven shows the type(s) of evaluation approach(es). Note that in some cases where unconventional or redundant metrics were used we lists them under the term “Other”. The abbreviations used in the table are the following: Str. = Recommendation Strategy, CB = Content-Based Filtering, CF = Collaborative Filtering, H = Hybrid; Abs. Nb. of Clicks = Absolute Number of Clicks, AUC = Area Under the Curve, CTR = Click Trough Rate, MAE = Mean Absolute Error, MAP = Mean Average Precision, MRR = Mean Reciprocal Rank, NCDG = Normalized Discounted Cumulative Gain, ROC = Receiver Operating Characteristic, Subj. Measures = Subjective Measures, TF-IDF = Term Frequency-Inverse Document Frequency; Sim. = Simulation.

Author(s)	Year	Str.	Addressed Challenges	Data Source	Eval. Metric	Eval. Type
Ha [66]	2006	CB	–	Unknown	TF-IDF, Other	Offline
Ahn et al. [5]	2007	CB	Cold-Start, User Profiling	Unknown	Precision, Recall, Subj. Measures	User Study
Bogers and van den Bosch [13]	2007	CB	–	Reuters	MAP, Kendall’s τ	User Study
Das et al. [39]	2007	CF	Recency, Scalability, User Profiling	Google News	Precision, Recall, Custom CTR	Offline, Online
Lee and Park [103]	2007	H	Cold-Start, Recency, User Profiling	Korean Mobile News	Abs. Nb. of Clicks, Other	Online
Pon et al. [148]	2007	CB	Recency	Reuters, Yahoo! News	F-Score, Precision, Recall	User Study
Sobecki and Szczepański [164]	2007	CF	–	Polish WikiNews Portal	Custom Rating Prediction Error	User Study
Bellogín et al. [11]	2008	CB, H	Sparsity, User Profiling	News@hand	Subj. Measures	User Study
Cantador et al. [22]	2008	CB, H	User Profiling	News@hand	Subj. Measures	User Study
Cantador et al. [25]	2008	H	Sparsity, User Profiling	News@hand	Subj. Measures	User Study
Chen et al. [30]	2008	CB	–	SOHU RSS Content Center	Hit Rate, Other	User Study
Cantador and Castells [23]	2009	CB, H	Cold-Start, Sparsity, User Profiling	News@hand	Subj. Measures	User Study
Chu and Park [31]	2009	H	Cold-Start, Recency, User Profiling	Yahoo! Front Page	CTR	Offline

Author(s)	Year	Str.	Addressed Challenges	Data Source	Eval. Metric	Eval. Type
Chu et al. [32]	2009	H	User Profiling	Yahoo! Front Page Today Module	CTR	Offline, Online
Frasincar et al. [50]	2009	CB	Recency, User Profiling	Hermes News Portal	Subj. Measures	User Study
Medo et al. [129]	2009	CF	Recency, Scalability	Unknown	Custom Accuracy	Sim.
Cleger-Tamayo et al. [34]	2010	CF	–	Daily News Ahora	Kullback-Leibler Divergence, Spearman's ρ	Offline
Di Massa et al. [42]	2010	H	User Profiling	Unknown	Subj. Measures	User Study
Fortuna et al. [49]	2010	CB	Cold-Start, Recency, User Profiling	Unknown	Precision	Online
IJntema et al. [74]	2010	CB	User Profiling	Athena	Accuracy, Precision, Recall, Specificity	User Study
Kompan and Bieliková [98]	2010	CB	–	SME.SK	F-Score, Precision, Recall	Offline
Li et al. [105]	2010	CB	Sparsity, User Profiling	Yahoo! Front Page	CTR	Offline
Li et al. [112]	2010	H	Novelty, Recency, User Profiling	Digg, Reuters	MAP, Custom Novelty	Offline
Liu et al. [115]	2010	H	Cold-Start, User Profiling	Google News	CTR	Online
Nishitarumizu et al. [139]	2010	CB	User Profiling	Yahoo! News	F-Score, Precision, Recall	Offline
Phelan et al. [146]	2010	CB, H	User Profiling	RSS feeds, Twitter	CTR, Subj. Measures	User Study
Wang et al. [173]	2010	H	Novelty	Digg, Reuters	MAP, Custom Novelty	Offline
Yeung and Yang [182]	2010	H	Recency, User Profiling	Unknown	Unknown	User Study
Yeung et al. [183]	2010	H	Recency, User Profiling	Unknown	Unknown	User Study
Wei et al. [174]	2011	CF	Scalability	Unknown	Other	Sim.
Agarwal et al. [2]	2011	CF	Cold-Start	Yahoo! News	AUC, Precision, ROC, Other	Offline
Cantador et al. [24]	2011	CB, H	Cold-Start, Sparsity	News@hand	Precision, RMSE	User Study
Drury et al. [47]	2011	CB	User Profiling	Unknown	Accuracy, Recall	Offline

Author(s)	Year	Str.	Addressed Challenges	Data Source	Eval. Metric	Eval. Type
Frasincar et al. [51]	2011	CB	User Profiling	Athena	Accuracy, F-Score, Precision, Recall, Specificity	User Study
Gershman et al. [58]	2011	CB	Recency	Unknown	MAP, NDCG	Offline
Goossen et al. [61]	2011	CB	User Profiling	Athena	Accuracy, F-Score, Precision, Recall, Specificity, ROC, Kappa Statistic	User Study
Guy et al. [65]	2011	H	Recency, User Profiling	Lotus Connections	Subj. Measures, Other	User Study
Li et al. [108]	2011	H	Cold-Start, Diversity, Recency, Scalability, User Profiling	Unknown	F-Score, Precision, Recall, Subj. Measures, Custom Diversity, Other	Offline, User Study
Li et al. [110]	2011	CB	Recency, User Profiling	Unknown	F-Score, Precision, Recall	Offline
Lv et al. [124]	2011	CB	Novelty	Yahoo! News	NDCG, Subj. Measures	Offline, User Study
Pon et al. [147]	2011	CB	User Profiling	Digg, Tagger, Yahoo! News	F-Score, Precision, Recall, Other	Offline
Prawesh and Padmanabhan [149]	2011	CB	–	Unknown	Other	Sim.
Wu et al. [179]	2011	CB	Novelty, User Profiling	163 News Website	Precision, Recall	Offline
Agarwal et al. [3]	2012	CF	Sparsity	Unknown	MAP, Precision, Recall	Offline
Bieliková et al. [12]	2012	CB	Recency	SME.SK	F-Score, Precision, Recall	Sim.
Capelle et al. [26]	2012	CB	User Profiling	Ceryx	Accuracy, Precision, Recall, Specificity	User Study
Cleger-Tamayo et al. [33]	2012	CB	User Profiling	Daily News Ahora	NDCG, Precision	Offline
Kirshenbaum et al. [93]	2012	H	User Profiling	Forbes.com	CTR, Abs. Nb. of Clicks	Online
Lin et al. [114]	2012	H	Cold-Start, Sparsity	Unknown	RMSE	Offline

Author(s)	Year	Str.	Addressed Challenges	Data Source	Eval. Metric	Eval. Type
Morales et al. [133]	2012	H	Recency, Sparsity	Twitter, Yahoo! News, Yahoo! Toolbar	Coverage, MRR, Precision	Offline
Prawesh and Padmanabhan [150]	2012	CB	–	Unknown	Other	Formal Proof
Saranya and Sadhasivam [159]	2012	H	Recency, Scalability, User Profiling	Unknown	Subj. Measures	User Study
Shmueli et al. [163]	2012	H	–	Newsvine, Yahoo! News	AUC	Offline
Wen et al. [176]	2012	H	Recency, User Profiling	CBC, CNN, Yahoo! News	Subj. Measures	User Study
Boutet et al. [17]	2013	CF	–	Arxiv, Digg, Custom	F-Score, Precision, Recall, Other	Offline, Sim.
Capelle et al. [27]	2013	CB	User Profiling	Ceryx	Accuracy, F-Score, Precision, Specificity	User Study
Garcin et al. [53]	2013	CB, CF, H	Cold-Start, Novelty, Recency, User Profiling	Tribune de Genève, 24 Heures	Custom Novelty, Other	Offline
Ilievski and Roy [75]	2013	CB	User Profiling	Unknown	MAP	Offline
Li and Li [107]	2013	H	Cold-Start, Diversity	Unknown	F-Score, NDCG, Precision, Recall, Set Diversity	Offline
Mannens et al. [127]	2013	H	Sparsity	The Vlaamse Radio-en Televisieomroep	Precision, Recall	Offline
Moerland et al. [131]	2013	CB	User Profiling	Ceryx	F-Score, Precision, Recall, Specificity	User Study
Montes-García et al. [132]	2013	H	Cold-Start, Recency, User Profiling	Unknown	Precision, Recall	User Study
Rao et al. [153]	2013	CB	User Profiling	New York Times, Sina News	F-Score, Precision, Recall	Offline
Rao et al. [154]	2013	CB	Diversity, Novelty, User Profiling	Sina News	F-Score, Precision, Recall, Custom Novelty	User Study
Said et al. [157]	2013	CB	–	Plista	CTR, Other	Online

Author(s)	Year	Str.	Addressed Challenges	Data Source	Eval. Metric	Eval. Type
Son et al. [165]	2013	CB	User Profiling	New York Times	Kappa Statistic, NDCG	Offline
Tavakolifard et al. [166]	2013	H	Cold-Start, Recency, User Profiling	PolarisMedia, Twitter	–	No Eval.
Werner and Cruz [177]	2013	CB	–	–	–	No Eval.
Zheng et al. [186]	2013	H	Cold-Start, Diversity, Recency	Unknown	F-Score, Set Diversity	Offline
Adnan et al. [1]	2014	CB	–	bdnews24.com	Subj. Measures	User Study
Agarwal and Singhal [4]	2014	H	User Profiling	Unknown	F-Score, Precision, Recall	Offline
Asikin and Wörndl [7]	2014	CB	Serendipity	Jaring-Ide	Subj. Measures	User Study
Bouras and Tsogkas [16]	2014	H	User Profiling	Unknown	F-Score, MAE, Other	Offline
Boutet et al. [18]	2014	CF	–	Unknown	F-Score, Precision, Recall, Other	Sim., User Study
Cui and Shi [37]	2014	CB	–	Sogou Labs	F-Score, Precision, Recall	Offline
Desarkar and Shinde [41]	2014	CB	Diversity, Recency	Plista	Custom Click Rate, Recall, Subj. Measures	Offline
Doychev et al. [45]	2014	CB	Scalability	Plista	CTR, Precision	Offline, Online
Go et al. [59]	2014	H	–	Unknown	Subj. Measures	User Study
Jiang and Hong [87]	2014	H	User Profiling	Xiamen University News Portal	F-Score, Precision, Recall, Other	Offline
Li et al. [111]	2014	H	User Profiling	Unknown	F-Score, Precision, Recall	Offline
Lin et al. [113]	2014	H	Cold-Start, Sparsity	Unknown	F-Score, Other	Offline
Lommatzsch [117]	2014	H	–	Plista	CTR, Precision	Online
Lu et al. [120]	2014	H	Scalability, Sparsity	ChouTi	Accuracy, Recall	Offline
Meguebli et al. [130]	2014	CB	User Profiling	ALJazeera, CNN, Independent, Telegraph	NDCG, Precision	Offline

Author(s)	Year	Str.	Addressed Challenges	Data Source	Eval. Metric	Eval. Type
Moreno et al. [134]	2014	CB	User Profiling	Guardian	Other	Sim.
Oh et al. [140]	2014	CB	User Profiling	Unknown	Hit Rate	Offline
Sadhasivam et al. [156]	2014	H	Scalability, User Profiling	Unknown	Subj. Measures	User Study
Trevisiol et al. [168]	2014	CF	Cold-Start	Yahoo! News	MRR, Precision	Offline
Vanchinathan et al. [170]	2014	CF	–	Yahoo! News	Abs. Nb. of Clicks, Other	Offline
Xia et al. [181]	2014	CB	–	Unknown	Accuracy	User Study
Ardissono et al. [6]	2015	H	User Profiling	Il Fatto Quotidiano	Subj. Measures	User Study
Bansal et al. [8]	2015	H	Cold-Start	ArsTechnica Gadgets, AT-Science, DailyMail	AUC, Hit Rate	Offline
Capelle et al. [28]	2015	CB	User Profiling	Ceryx	Accuracy, F-Score, Precision, Recall, Specificity, Kappa Statistic	User Study
Darvishy et al. [38]	2015	CF	Recency	Unknown	F-Score, Precision, Recall	Offline
De Pessemier et al. [40]	2015	CB	Cold-Start, User Profiling	Stream Store News Service	Subj. Measures	User Study
Dong et al. [44]	2015	H	–	CaiXin	F-Score, Precision, Recall	Offline
Doychev et al. [46]	2015	CB	–	Plista	CTR, Other	Online
Garrido et al. [56]	2015	CB	User Profiling	–	–	No Eval.
Gu et al. [62]	2016	CB	Diversity	Sina News	F-Score, NDCG, Custom Diversity	Offline
Kliegr and Kuchař [95]	2015	CF	Recency	Plista	CTR, Other	Offline, Online
Lommatzsch and Albayrak [118]	2015	CB, CF, H	Cold-Start	Plista	CTR, Precision	Online
Muralidhar et al. [136]	2015	CB, CF, H	Recency	Washington Post	Hit Rate, Other	Offline
Pessemier et al. [145]	2015	H	–	–	–	No Eval.

Author(s)	Year	Str.	Addressed Challenges	Data Source	Eval. Metric	Eval. Type
Ren et al. [155]	2015	CB	User Profiling	Hermes News Portal	F-Score, Precision, Recall	Offline
Vanchinathan et al. [169]	2015	CF	Diversity	Yahoo! Front Page Today Module	CTR, Custom Diversity	Offline
Yingyuan et al. [184]	2015	CF	–	Unknown	F-Score, Precision, Recall, Other	Offline
Domann et al. [43]	2016	–	Scalability	Plista	CTR, Other	Online
Huy et al. [72]	2016	CB	User Profiling	Dantri.com.vn, Vietnam-Net.vn, VNExpress.vn	F-Score, Precision, Recall	User Study
Ji et al. [86]	2016	CF	Sparsity	XMUNews	RMSE, Other	Offline
Lak et al. [100]	2016	CB	Recency	Unknown	Accuracy	Offline
Lenhart and Herzog [104]	2016	CB, H	Diversity, Recency	Sport1.de	Custom Click Rate, Subj. Measures	Online, User Study
Ma et al. [125]	2016	CB	User Profiling	Bing Now	MAP, MRR	Offline

Table A.4: Overview of the surveyed news recommender systems papers that do not propose a new recommendation algorithm

Cat.	Author(s)	Year	Description
Survey	Borges and Lorena [14]	2010	Survey with a focus on six specific news recommender systems
	Li et al. [109]	2011	Discussion of various NRS issues and additional experimental analysis based on approaches previously proposed in [108]
	Özgöbek et al. [141]	2014	Discussion of NRS challenges and countermeasures from the literature
	Özgöbek et al. [142]	2014	Survey of existing data sets and proposal of the SmartMedia data set
	Harandi and Gulla [67]	2015	Survey on user profiling
Plista/Challenges	Kille et al. [90]	2013	Introduction of the Plista data set
	Said et al. [158]	2013	Analysis of click data gathered via the Plista API
	Brodt and Hopfgartner [19]	2014	Introduction of the Plista living lab platform
	Hopfgartner et al. [71]	2014	Introduction of the living lab of CLEF '14
	Kille et al. [92]	2015	Presentation the CLEF '15 NewsREEL challenge
Hopfgartner et al. [70]	2016	Analysis of the CLEF '15 NewsREEL challenge	
Research Framework	Cantador et al. [21]	2008	Presentation of a semantic-based news recommendation framework
	Garcin and Faltings [54]	2013	Presentation of a personalized news recommendation framework
	Ingvaldsen et al. [76]	2015	Presentation of the user interface of the SmartMedia NRS framework
	Ingvaldsen et al. [77]	2015	Presentation of contextual features of the SmartMedia NRS framework
	Kille et al. [91]	2015	Description of the Plista platform and analysis of Plista data
	Lommatzsch and Werner [119]	2015	Proposal of an offline evaluation framework, tested on algorithms previously proposed in [118]
Constantinides and Dowell [35]	2016	Proposal of an approach to user interface personalization	
Classification/Clustering	Qiu et al. [152]	2009	Proposal of a topic detection method for news item clustering
	Wei et al. [175]	2009	Proposal of a one-class news article classifier
	Wu et al. [180]	2010	Proposal of a clustering approach based on affinity propagation and LDA
	Ho et al. [69]	2012	Proposal of an event recommender that extracts data from news articles
	Werner et al. [178]	2014	Proposal of a multi-label classifier for economic news based on ontologies
	Iglesias et al. [73]	2015	Proposal of an approach to classify news article into categories based on their content
Other	Knobloch-Westerwick et al. [97]	2005	Analysis of the effect of displaying popularity indicators in news item lists
	Tintarev and Masthoff [167]	2006	Comparison of two similarity measurement methods
	Li et al. [106]	2011	Proposal of an unbiased evaluation method for contextual bandit approaches
	Luostarinen and Kohonen [123]	2013	Comparison of three standard NRS approaches and two feature extraction methods on a custom data set
	Garcin et al. [55]	2014	Comparison of the offline and online performance of an NRS approach previously proposed in [53]
	Liu et al. [116]	2015	Proposal of an approach to predict the number of comments on a news article based its content and social shares
	Maksai et al. [126]	2015	Proposal of a model to predict the online performance of an NRS by means of offline metrics
	Gulla et al. [63]	2016	Analysis of news article life spans and the effect of Facebook on these

References

- [1] M. N. M. Adnan, M. R. Chowdury, I. Taz, T. Ahmed, R. M. Rahman, Content based news recommendation system based on fuzzy logic, in: Proceedings of the 2014 International Conference on Informatics, Electronics & Vision (ICIEV '14), 1–6, 2014.
- [2] D. Agarwal, B.-C. Chen, B. Pang, Personalized Recommendation of User Comments via Factor Models, in: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP '11), 571–582, 2011.
- [3] D. Agarwal, B.-C. Chen, X. Wang, Multi-faceted Ranking of News Articles Using Post-read Actions, in: Proceedings of the 21st International Conference on Information and Knowledge Management (CIKM '12), 694–703, 2012.
- [4] S. Agarwal, A. Singhal, Handling skewed results in news recommendations by focused analysis of semantic user profiles, in: Proceedings of the 2014 International Conference on Optimization, Reliability, and Information Technology (ICROIT '14), 74–79, 2014.
- [5] J.-W. Ahn, P. Brusilovsky, J. Grady, D. He, S. Y. Syn, Open User Profiles for Adaptive News Systems: Help or Harm?, in: Proceedings of the 16th International Conference on World Wide Web (WWW '07), 11–20, 2007.
- [6] L. Ardissono, G. Petrone, F. Vigiatiuro, News Recommender Based on Rich Feedback, in: Proceedings of the 21st Conference on User Modeling, Adaptation and Personalization (UMAP '15), 331–336, 2015.
- [7] Y. A. Asikin, W. Wörndl, Stories Around You: Location-based Serendipitous Recommendation of News Articles, in: Extended Proceedings of the 22nd Conference on User Modeling, Adaptation, and Personalization (UMAP '14), 2014.
- [8] T. Bansal, M. Das, C. Bhattacharyya, Content Driven User Profiling for Comment-Worthy Recommendations of News and Blog Articles, in: Proceedings of the 9th Conference on Recommender Systems (RecSys '15), 195–202, 2015.
- [9] J. Beel, B. Gipp, S. Langer, C. Breiterger, Research-paper recommender systems: a literature survey, *International Journal on Digital Libraries* 17 (4) (2016) 305–338.
- [10] J. Beel, S. Langer, A Comparison of Offline Evaluations, Online Evaluations, and User Studies in the Context of Research-Paper Recommender Systems, in: Proceedings of the 19th International Conference on Theory and Practice of Digital Libraries (TPDL '15), 153–168, 2015.
- [11] A. Bellogín, I. Cantador, P. Castells, A. Ortigosa, Discovering relevant preferences in a personalised recommender system using machine learning techniques, in: Proceedings of the 2008 ECML-PKDD Workshop on Preference Learning (ECML-PKDD '08), 2008.

- [12] M. Bieliková, M. Kompan, D. Zeleník, Effective hierarchical vector-based news representation for personalized recommendation, *Computer Science and Information Systems* 9 (1) (2012) 303–322.
- [13] T. Bogers, A. van den Bosch, Comparing and Evaluating Information Retrieval Algorithms for News Recommendation, in: *Proceedings of the 2007 Conference on Recommender Systems (RecSys '07)*, 141–144, 2007.
- [14] H. L. Borges, A. C. Lorena, A Survey on Recommender Systems for News Data, in: *Smart Information and Knowledge Management*, vol. 260 of *Studies in Computational Intelligence*, Springer, 129–151, 2010.
- [15] J. Borrs, A. Moreno, A. Valls, Intelligent tourism recommender systems: A survey, *Expert Systems with Applications* 41 (16) (2014) 7370 – 7389.
- [16] C. Bouras, V. Tsogkas, Improving news articles recommendations via user clustering, *International Journal of Machine Learning and Cybernetics* (2014) 1–15.
- [17] A. Boutet, D. Frey, R. Guerraoui, A. Jegou, A.-M. Kermarrec, WHATSUP: A Decentralized Instant News Recommender, in: *Proceedings of the 27th International Symposium on Parallel Distributed Processing (IPDPS '13)*, 741–752, 2013.
- [18] A. Boutet, D. Frey, R. Guerraoui, A. Jégou, A.-M. Kermarrec, Privacy-Preserving Distributed Collaborative Filtering, in: *Networked Systems*, vol. 8593 of *Lecture Notes in Computer Science*, Springer, 169–184, 2014.
- [19] T. Brodt, F. Hopfgartner, Shedding Light on a Living Lab: The CLEF NEWSREEL Open Recommendation Platform, in: *Proceedings of the 5th Information Interaction in Context Symposium (IIiX '14)*, 223–226, 2014.
- [20] P. G. Campos, F. Díez, I. Cantador, Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols, *User Modeling and User-Adapted Interaction* 24 (1) (2013) 67–119.
- [21] I. Cantador, A. Bellogín, P. Castells, News@hand: A Semantic Web Approach to Recommending News, in: *Proceedings of the 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH '08)*, 279–283, 2008.
- [22] I. Cantador, A. Bellogín, P. Castells, Ontology-Based Personalised and Context-Aware Recommendations of News Items, in: *Proceedings of the 2008 International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT '08)*, 562–565, 2008.
- [23] I. Cantador, P. Castells, Semantic contextualisation in a news recommender system, in: *Proceedings of the 2009 Workshop on Context-Aware Recommender Systems (CARS '09)*, 2009.
- [24] I. Cantador, P. Castells, A. Bellogín, An enhanced semantic layer for hybrid recommender systems: Application to news recommendation, *International Journal on Semantic Web and Information Systems* 7 (1) (2011) 44–78.
- [25] I. Cantador, M. Szomszor, H. Alani, M. Fernández, P. Castells, Enriching Ontological User Profiles with Tagging History for Multi-Domain Recommendations, in: *Proceedings of the 1st International Workshop on Collective Semantics: Collective Intelligence & the Semantic Web (CISWeb '08)*, 2008.

- [26] M. Capelle, F. Frasincar, M. Moerland, F. Hogenboom, Semantics-based News Recommendation, in: Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics (WIMS '12), 27:1–27:9, 2012.
- [27] M. Capelle, F. Hogenboom, A. Hogenboom, F. Frasincar, Semantic News Recommendation Using Wordnet and Bing Similarities, in: Proceedings of the 28th Annual Symposium on Applied Computing (SAC '13), 296–302, 2013.
- [28] M. Capelle, M. Moerland, F. Hogenboom, F. Frasincar, D. Vadic, Bing-SF-IDF+: A Hybrid Semantics-driven News Recommender, in: Proceedings of the 30th Annual Symposium on Applied Computing (SAC '15), 732–739, 2015.
- [29] P. Castells, N. J. Hurley, S. Vargas, Novelty and Diversity in Recommender Systems, in: F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (Eds.), Recommender Systems Handbook, chap. 26, Springer, 881–918, 2015.
- [30] W. Chen, L.-j. Zhang, C. Wang, C. Chen, J.-j. Bu, Pervasive Web News Recommendation for Visually Impaired People, in: Proceedings of the 2008 International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT '08), 119–122, 2008.
- [31] W. Chu, S.-T. Park, Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models, in: Proceedings of the 18th International Conference on World Wide Web (WWW '09), 691–700, 2009.
- [32] W. Chu, S.-T. Park, T. Beaupre, N. Motgi, A. Phadke, S. Chakraborty, J. Zachariah, A Case Study of Behavior-driven Conjoint Analysis on Yahoo! Front Page Today Module, in: Proceedings of the 15th International Conference on Knowledge Discovery and Data Mining (KDD '09), 1097–1104, 2009.
- [33] S. Cleger-Tamayo, J. M. Fernández-Luna, J. F. Huete, Top-N news recommendations in digital newspapers, Knowledge-Based Systems 27 (2012) 180 – 189.
- [34] S. Cleger-Tamayo, J. M. Fernández-Luna, J. F. Huete, R. Pérez-Vázquez, J. C. Rodríguez Cano, A Proposal for News Recommendation Based on Clustering Techniques, in: Trends in Applied Intelligent Systems, vol. 6098 of Lecture Notes in Computer Science, Springer, 478–487, 2010.
- [35] M. Constantinides, J. Dowell, User Interface Personalization in News Apps, in: Extended Proceedings of the 24th Conference on User Modeling, Adaptation and Personalisation (UMAP '16), 2016.
- [36] P. Cremonesi, Y. Koren, R. Turrin, Performance of Recommender Algorithms on Top-n Recommendation Tasks, in: Proceedings of the 4th Conference on Recommender Systems (RecSys '10), 39–46, 2010.
- [37] L. Cui, Y. Shi, A Method based on One-class SVM for News Recommendation, in: Proceedings of the 2nd International Conference on Information Technology and Quantitative Management (ITQM '14), 281 – 290, 2014.

- [38] A. Darvishy, H. Ibrahim, A. Mustapha, F. Sidi, New Attributes for Neighborhood-based Collaborative Filtering in News Recommendation, *Journal of Emerging Technologies in Web Intelligence* 7 (1) (2015) 13–19.
- [39] A. S. Das, M. Datar, A. Garg, S. Rajaram, Google News Personalization: Scalable Online Collaborative Filtering, in: *Proceedings of the 16th International Conference on World Wide Web (WWW '07)*, 271–280, 2007.
- [40] T. De Pessemier, C. Courtois, K. Vanhecke, K. Van Damme, L. Martens, L. De Marez, A user-centric evaluation of context-aware recommendations for a mobile news service, *Multimedia Tools and Applications* 75 (6) (2015) 3323–3351.
- [41] M. Desarkar, N. Shinde, Diversification in news recommendation for privacy concerned users, in: *Proceedings of the 2014 International Conference on Data Science and Advanced Analytics (DSAA '14)*, 135–141, 2014.
- [42] R. Di Massa, M. Montagnuolo, A. Messina, Implicit News Recommendation Based on User Interest Models and Multimodal Content Analysis, in: *Proceedings of the 3rd International Workshop on Automated Information Extraction in Media Production (AIEMPro '10)*, 33–38, 2010.
- [43] J. Domann, J. Meiners, L. Helmers, A. Lommatzsch, Real-time news recommendations using Apache Spark, in: *Working Notes of the 7th Conference and Labs of the Evaluation Forum (CLEF '16)*, 2016.
- [44] H. Dong, J. Zhu, Y. Tang, C. Xu, R. Ding, L. Chen, UBS: A Novel News Recommendation System Based on User Behavior Sequence, in: *Proceedings of the 8th International Conference on Knowledge Science, Engineering and Management (KSEM '15)*, 738–750, 2015.
- [45] D. Doychev, A. Lawlor, R. Rafter, B. Smyth, An Analysis of Recommender Algorithms for Online News, in: *Working Notes for the 2014 Conference and Labs of the Evaluation Forum (CLEF '14)*, 825–836, 2014.
- [46] D. Doychev, R. Rafter, A. Lawlor, B. Smyth, News recommenders: Real-time, real-life experiences, in: *Proceedings of the 23rd International Conference on User Modeling, Adaptation and Personalization (UMAP '15)*, 337–342, 2015.
- [47] B. Drury, J. J. Almeida, M. H. M. Morais, Magellan: An adaptive ontology driven “breaking financial news” recommender, in: *Proceedings of the 6th Iberian Conference on Information Systems and Technologies (CISTI '11)*, 1–6, 2011.
- [48] M. D. Ekstrand, F. M. Harper, M. C. Willemsen, J. A. Konstan, User Perception of Differences in Recommender Algorithms, in: *Proceedings of the 8th Conference on Recommender Systems (RecSys '14)*, 161–168, 2014.
- [49] B. Fortuna, C. Fortuna, D. Mladenović, Real-Time News Recommender System, in: *Machine Learning and Knowledge Discovery in Databases*, vol. 6323 of *Lecture Notes in Computer Science*, Springer, 583–586, 2010.
- [50] F. Frasinca, J. Borsje, L. Levering, A semantic web-based approach for building personalized news services, *International Journal of E-Business Research (IJEER)* 5 (3) (2009) 35–53.

- [51] F. Frasincar, W. IJntema, F. Goossen, F. Hogenboom, A semantic approach for news recommendation, in: M. E. Zorrilla, J.-N. Mazn, scar Ferrndez, I. Garrigs, F. Daniel, J. Trujillo (Eds.), *Business Intelligence Applications and the Web: Models, Systems and Technologies*, chap. 5, IGI Global, 2011.
- [52] E. Frigó, R. Pálovics, D. Kelen, L. Kocsis, A. A. Benczúr, *Alpenglow: Open Source Recommender Framework with Time-aware Learning and Evaluation*, in: *Poster Proceedings the 11th Conference on Recommender Systems (RecSys '17)*, 2017.
- [53] F. Garcin, C. Dimitrakakis, B. Faltings, *Personalized news recommendation with context trees*, in: *Proceedings of the 7th Conference on Recommender Systems (RecSys '13)*, 105–112, 2013.
- [54] F. Garcin, B. Faltings, *PEN Recsys: A Personalized News Recommender Systems Framework*, in: *Proceedings of the 2013 International News Recommender Systems Workshop and Challenge at RecSys '13*, 3–9, 2013.
- [55] F. Garcin, B. Faltings, O. Donatsch, A. Alazzawi, C. Bruttin, A. Huber, *Offline and online evaluation of news recommender systems at swissinfo.ch*, in: *Proceedings of the 8th Conference on Recommender Systems (RecSys '14)*, 169–176, 2014.
- [56] A. L. Garrido, M. G. Buey, S. Ilarri, I. Fürstner, L. Szedmina, *KGNR: A knowledge-based geographical news recommender*, in: *Proceedings of the 13th International Symposium on Intelligent Systems and Informatics (SISY '15)*, 195–198, 2015.
- [57] M. Ge, C. Delgado-Battenfeld, D. Jannach, *Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity*, in: *Proceedings of the 4th Conference on Recommender Systems (RecSys '10)*, 257–260, 2010.
- [58] A. Gershman, T. Wolfe, E. Fink, J. G. Carbonell, *News personalization using support vector machines*, in: *Proceedings of the Workshop on Enriching Information Retrieval at SIGIR '11*, 2011.
- [59] E. Go, E. H. Jung, M. Wu, *The effects of source cues on online news perception*, *Computers in Human Behavior* 38 (2014) 358 – 367.
- [60] C. A. Gomez-Uribe, N. Hunt, *The Netflix Recommender System: Algorithms, Business Value, and Innovation*, *Transactions on Management Information Systems* 6 (4) (2015) 13:1–13:19.
- [61] F. Goossen, W. IJntema, F. Frasincar, F. Hogenboom, U. Kaymak, *News Personalization Using the CF-IDF Semantic Recommender*, in: *Proceedings of the 2011 International Conference on Web Intelligence, Mining and Semantics (WIMS '11)*, 10:1–10:12, 2011.
- [62] W. Gu, S. Dong, M. Chen, *Personalized news recommendation based on articles chain building*, *Neural Computing and Applications* 27 (5) (2016) 1263–1272.

- [63] J. A. Gulla, C. Marco, A. D. Fidjestøl, X. Su, J. E. Ingvaldsen, Ö. Özgöbek, The Intricacies of Time in News Recommendation, in: *Extended Proceedings of the 24th Conference on User Modeling, Adaptation and Personalisation (UMAP '16)*, 2016.
- [64] J. A. Gulla, L. Zhang, P. Liu, O. Özgöbek, X. Su, The Adressa Dataset for News Recommendation, in: *Proceedings of the 2017 International Conference on Web Intelligence (WI '17)*, 1042–1048, 2017.
- [65] I. Guy, I. Ronen, A. Raviv, Personalized activity streams: Sifting through the “river of news”, in: *Proceedings of the 5th Conference on Recommender Systems (RecSys '11)*, 181–188, 2011.
- [66] S. H. Ha, Digital content recommender on the Internet, *IEEE Intelligent Systems* 21 (2) (2006) 70–77.
- [67] M. Harandi, J. A. Gulla, Survey of User Profiling in News Recommender Systems, in: *Proceedings of the 3rd International Workshop on News Recommendation and Analytics (INRA '15) at RecSys '15*, 20–26, 2015.
- [68] J. L. Herlocker, J. A. Konstan, L. G. Terveen, J. T. Riedl, Evaluating Collaborative Filtering Recommender Systems, *Transactions on Information Systems* 22 (1) (2004) 5–53.
- [69] S.-S. Ho, M. Lieberman, P. Wang, H. Samet, Mining Future Spatiotemporal Events and Their Sentiment from Online News Articles for Location-aware Recommendation System, in: *Proceedings of the 1st International Workshop on Mobile Geographic Information Systems (MobiGIS '12) at SIGSPATIAL '12*, 25–32, 2012.
- [70] F. Hopfgartner, T. Brodt, J. Seiler, B. Kille, A. Lommatzsch, M. Larson, R. Turrin, A. Serény, Benchmarking News Recommendations: The CLEF NewsREEL Use Case, *SIGIR Forum* 49 (2) (2016) 129–136.
- [71] F. Hopfgartner, B. Kille, A. Lommatzsch, T. Plumbaum, T. Brodt, T. Heintz, Benchmarking News Recommendations in a Living Lab, in: *Information Access Evaluation. Multilinguality, Multimodality, and Interaction*, vol. 8685 of *Lecture Notes in Computer Science*, Springer, 250–267, 2014.
- [72] N. T. Huy, V. A. Nguyen, et al., Implicit Feedback Mechanism to Manage User Profile Applied in Vietnamese News Recommender System, *International Journal of Computer and Communication Engineering* 5 (4) (2016) 276.
- [73] J. A. Iglesias, A. Tiemblo, A. Ledezma, A. Sanchis, Web news mining in an evolving framework, *Information Fusion* 28 (2015) 90–98.
- [74] W. IJntema, F. Goossen, F. Frasinca, F. Hogenboom, Ontology-based News Recommendation, in: *Workshop Proceedings of the 13th International Conference on Extending Database Technology (EDBT '10)*, 16:1–16:6, 2010.
- [75] I. Ilievski, S. Roy, Personalized News Recommendation Based on Implicit Feedback, in: *Proceedings of the 2013 International News Recommender Systems Workshop and Challenge at RecSys '13*, 10–15, 2013.

- [76] J. E. Ingvaldsen, J. A. Gulla, Ö. Özgöbek, User Controlled News Recommendations, in: Proceedings of the Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS '15) at RecSys '15, 45–48, 2015.
- [77] J. E. Ingvaldsen, Ö. Özgöbek, J. A. Gulla, Context-Aware User-Driven News Recommendation, in: Proceedings of the 3rd International Workshop on News Recommendation and Analytics (INRA '15) at RecSys '15, 33–36, 2015.
- [78] D. Jannach, G. Adomavicius, Recommendations With A Purpose, in: Proceedings of the 10th Conference on Recommender Systems (RecSys '16), 7–10, 2016.
- [79] D. Jannach, K. Hegelich, A case study on the effectiveness of recommendations in the mobile internet, in: Proceedings of the 2009 Conference on Recommender Systems (RecSys '09), 205–208, 2009.
- [80] D. Jannach, L. Lerche, I. Kamehkhosh, M. Jugovac, What recommenders recommend - An analysis of recommendation biases and possible countermeasures, *User Modeling and User-Adapted Interaction* 25 (5) (2015) 427–491.
- [81] D. Jannach, M. Ludewig, When Recurrent Neural Networks Meet the Neighborhood for Session-Based Recommendation, in: Proceedings of the 11th Conference on Recommender Systems (RecSys' 17), 306–310, 2017.
- [82] D. Jannach, M. Ludewig, L. Lerche, Session-based item recommendation in e-commerce: on short-term intents, reminders, trends and discounts, *User Modeling and User-Adapted Interaction* 27 (3) (2017) 351–392.
- [83] D. Jannach, P. Resnick, A. Tuzhilin, M. Zanker, Recommender Systems — Beyond Matrix Completion, *Communications of the ACM* 59 (11) (2016) 94–102.
- [84] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, *Recommender Systems: An Introduction*, Cambridge University Press, 2010.
- [85] D. Jannach, M. Zanker, M. Ge, M. Gröning, Recommender Systems in Computer Science and Information Systems - A Landscape of Research, in: Proceedings of the 13th International Conference on E-Commerce and Web Technologies (EC-WEB '12), 76–87, 2012.
- [86] Y. Ji, W. Hong, Y. Shangguan, H. Wang, J. Ma, Regularized singular value decomposition in news recommendation system, in: Proceedings of the 11th International Conference on Computer Science Education (ICCSE '16), 621–626, 2016.
- [87] S. Jiang, W. Hong, A vertical news recommendation system: CCNS, in: Proceedings of the 9th International Conference on Computer Science Education (ICCSE '14), 1105–1114, 2014.
- [88] M. Jugovac, D. Jannach, Interacting with Recommenders Overview and Research Directions, *Transactions on Interactive Intelligent Systems* 7 (3) (2017) 10:1–10:46.

- [89] K. Kapoor, V. Kumar, L. Terveen, J. A. Konstan, P. Schrater, “I Like to Explore Sometimes”: Adapting to Dynamic User Novelty Preferences, in: *Proceedings of the 9th Conference on Recommender Systems (RecSys ’15)*, 19–26, 2015.
- [90] B. Kille, F. Hopfgartner, T. Brodt, T. Heintz, The Plista Dataset, in: *Proceedings of the 2013 International News Recommender Systems Workshop and Challenge at RecSys ’13*, 16–23, 2013.
- [91] B. Kille, A. Lommatzsch, T. Brodt, News Recommendation in Real-Time, in: F. Hopfgartner (Ed.), *Smart Information Systems: Computational Intelligence for Real-Life Applications, Advances in Computer Vision and Pattern Recognition*, Springer, 149–180, 2015.
- [92] B. Kille, A. Lommatzsch, R. Turrin, A. Serény, M. Larson, T. Brodt, J. Seiler, F. Hopfgartner, Stream-Based Recommendations: Online and Offline Evaluation as a Service, in: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, vol. 9283 of *Lecture Notes in Computer Science*, Springer, 497–517, 2015.
- [93] E. Kirshenbaum, G. Forman, M. Dugan, A Live Comparison of Methods for Personalized Article Recommendation at Forbes.com, in: *Proceedings of the 2012 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD ’12)*, 51–66, 2012.
- [94] B. Kitchenham, P. Brereton, A systematic review of systematic review process research in software engineering, *Information and Software Technology* 55 (12) (2013) 2049 – 2075.
- [95] T. Kliegr, J. Kuchař, Benchmark of Rule-Based Classifiers in the News Recommendation Task, in: *Proceedings of the 2015 Conference and Labs of the Evaluation Forum (CLEF ’15)*, 130–141, 2015.
- [96] B. P. Knijnenburg, M. C. Willemsen, Z. Gantner, H. Soncu, C. Newell, Explaining the User Experience of Recommender Systems, *User Modeling and User-Adapted Interaction* 22 (4-5) (2012) 441–504.
- [97] S. Knobloch-Westerwick, N. Sharma, D. L. Hansen, S. Alter, Impact of Popularity Indications on Readers’ Selective Exposure to Online News, *Journal of Broadcasting & Electronic Media* 49 (3) (2005) 296–313.
- [98] M. Kompan, M. Bieliková, Content-Based News Recommendation, in: *Proceedings of the 11th International Conference on E-Commerce and Web Technologies (EC-Web ’10)*, 61–72, 2010.
- [99] D. Kotkov, J. Veijalainen, S. Wang, Challenges of Serendipity in Recommender Systems, in: *Proceedings of the 12th International Conference on Web Information Systems and Technologies (WISE ’16)*, 251–256, 2016.
- [100] P. Lak, C. Babaoglu, A. B. Bener, P. Pralat, News Article Position Recommendation Based on the Analysis of Article’s Content - Time Matters, in: *Proceedings of the 3rd Workshop on New Trends in Content-Based Recommender Systems (CBRecSys ’16) at RecSys ’16*, 11–14, 2016.
- [101] J. Langford, T. Zhang, The Epoch-Greedy Algorithm for Multi-armed Bandits with Side Information, in: *Proceedings of the 21st Annual Conference on Neural Information Processing Systems (NIPS ’07)*, 817–824, 2007.

- [102] N. Lathia, S. Hailes, L. Capra, X. Amatriain, Temporal Diversity in Recommender Systems, in: Proceedings of the 33rd International Conference on Research and Development in Information Retrieval (SIGIR '10), 210–217, 2010.
- [103] H. Lee, S. J. Park, MONERS: A news recommender for the mobile web, *Expert Systems with Applications* 32 (1) (2007) 143–150.
- [104] P. Lenhart, D. Herzog, Combining Content-based and Collaborative Filtering for Personalized Sports News Recommendations, in: Proceedings of the 3rd Workshop on New Trends in Content-Based Recommender Systems (CBRecSys '16) at RecSys '16, 2016.
- [105] L. Li, W. Chu, J. Langford, R. E. Schapire, A Contextual-bandit Approach to Personalized News Article Recommendation, in: Proceedings of the 19th International Conference on World Wide Web (WWW '10), 661–670, 2010.
- [106] L. Li, W. Chu, J. Langford, X. Wang, Unbiased Offline Evaluation of Contextual-bandit-based News Article Recommendation Algorithms, in: Proceedings of the 4th International Conference on Web Search and Data Mining (WSDM '11), 297–306, 2011.
- [107] L. Li, T. Li, News recommendation via hypergraph learning: encapsulation of user behavior and news content, in: Proceedings of the 6th International Conference on Web Search and Data Mining (WSDM '13), 305–314, 2013.
- [108] L. Li, D. Wang, T. Li, D. Knox, B. Padmanabhan, SCENE: a scalable two-stage personalized news recommendation system, in: Proceedings of the 34th International Conference on Research and Development in Information Retrieval (SIGIR '11), 125–134, 2011.
- [109] L. Li, D.-D. Wang, S.-Z. Zhu, T. Li, Personalized News Recommendation: A Review and an Experimental Investigation, *Journal of Computer Science and Technology* 26 (5) (2011) 754–766.
- [110] L. Li, L. Zheng, T. Li, LOGO: A Long-short User Interest Integration in Personalized News Recommendation, in: Proceedings of the 5th Conference on Recommender Systems (RecSys '11), 317–320, 2011.
- [111] L. Li, L. Zheng, F. Yang, T. Li, Modeling and broadening temporal user interest in personalized news recommendation, *Expert Systems with Applications* 41 (7) (2014) 3168 – 3177.
- [112] Q. Li, J. Wang, Y. P. Chen, Z. Lin, User comments for news recommendation in forum-based social media, *Information Sciences* 180 (24) (2010) 4929 – 4939.
- [113] C. Lin, R. Xie, X. Guan, L. Li, T. Li, Personalized news recommendation via implicit social experts, *Information Sciences* 254 (2014) 1–18.

- [114] C. Lin, R. Xie, L. Li, Z. Huang, T. Li, PRemiSE: Personalized News Recommendation via Implicit Social Experts, in: Proceedings of the 21st International Conference on Information and Knowledge Management (CIKM '12), 1607–1611, 2012.
- [115] J. Liu, P. Dolan, E. R. Pedersen, Personalized News Recommendation Based on Click Behavior, in: Proceedings of the 15th International Conference on Intelligent User Interfaces (IUI '10), 31–40, 2010.
- [116] Q. Liu, M. Zhou, X. Zhao, Understanding News 2.0: A framework for explaining the number of comments from readers on online news, *Information & Management* 52 (7) (2015) 764–776.
- [117] A. Lommatzsch, Real-Time News Recommendation Using Context-Aware Ensembles, in: Proceedings of the 36th European Conference on IR Research (ECIR '14), 51–62, 2014.
- [118] A. Lommatzsch, S. Albayrak, Real-time Recommendations for User-item Streams, in: Proceedings of the 30th Annual Symposium on Applied Computing (SAC '15), 1039–1046, 2015.
- [119] A. Lommatzsch, S. Werner, Optimizing and Evaluating Stream-Based News Recommendation Algorithms, in: *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, vol. 9283 of *Lecture Notes in Computer Science*, Springer, 376–388, 2015.
- [120] M. Lu, Z. Qin, Y. Cao, Z. Liu, M. Wang, Scalable news recommendation using multi-dimensional similarity and Jaccard-Kmeans clustering, *Journal of Systems and Software* 95 (2014) 242–251.
- [121] C. Ludmann, H.-J. Appelrath, Lessons Learned from Using a Data Stream Management System for Real-time Recommendation of Popular News Articles based on Real User Streams, in: Proceedings of the Workshop on Large Scale Recommender Systems (LSRS '16) at RecSys '16, 2016.
- [122] C. A. Ludmann, Recommending News Articles in the CLEF News Recommendation Evaluation Lab with the Data Stream Management System Odysseus, in: Working Notes for the 2017 Conference and Labs of the Evaluation Forum (CLEF '17), 2017.
- [123] T. Luostarinen, O. Kohonen, Using Topic Models in Content-Based News Recommender Systems, in: Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA '13), 251–474, 2013.
- [124] Y. Lv, T. Moon, P. Kolari, Z. Zheng, X. Wang, Y. Chang, Learning to Model Relatedness for News Recommendation, in: Proceedings of the 20th International Conference on World Wide Web (WWW '11), 57–66, 2011.
- [125] H. Ma, X. Liu, Z. Shen, User Fatigue in Online News Recommendation, in: Proceedings of the 25th International Conference on World Wide Web (WWW '16), 1363–1372, 2016.
- [126] A. Maksai, F. Garcin, B. Faltings, Predicting Online Performance of News Recommender Systems Through Richer Evaluation Metrics, in: Proceedings of the 9th Conference on Recommender Systems (RecSys '15), 179–186, 2015.

- [127] E. Mannens, S. Coppens, T. De Pessemier, H. Dacquin, D. Van Deursen, R. De Sutter, R. Van de Walle, Automatic news recommendations via aggregated profiling, *Multimedia tools and applications* 63 (2) (2013) 407–425.
- [128] S. M. McNee, J. Riedl, J. A. Konstan, Being accurate is not enough: How accuracy metrics have hurt recommender systems, in: *Proceedings of the 2006 Conference on Human Factors in Computing Systems (CHI '06)*, 1097–1101, 2006.
- [129] M. Medo, Y.-C. Zhang, T. Zhou, Adaptive model for recommendation of news, *EPL (Europhysics Letters)* 88 (3).
- [130] Y. Meguebli, M. Kacimi, B.-I. Doan, F. Popineau, Building Rich User Profiles for Personalized News Recommendation, in: *Workshop Proceedings of the 22nd Conference on User Modeling, Adaptation, and Personalization (UMAP '14)*, 2014.
- [131] M. Moerland, F. Hogenboom, M. Capelle, F. Frasincar, Semantics-based News Recommendation with SF-IDF+, in: *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics (WIMS '13)*, 22:1–22:8, 2013.
- [132] A. Montes-García, J. M. Álvarez-Rodríguez, J. E. Labra-Gayo, M. Martínez-Merino, Towards a journalist-based news recommendation system: The Wesomender approach, *Expert Systems with Applications* 40 (17) (2013) 6735 – 6741.
- [133] G. D. F. Morales, A. Gionis, C. Lucchese, From chatter to headlines: harnessing the real-time web for personalized news recommendation, in: *Proceedings of the 5th International Conference on Web Search and Web Data Mining (WSDM '12)*, 153–162, 2012.
- [134] A. Moreno, L. Marin, D. Isern, D. Perelló, Dynamic Learning of Keyword-Based Preferences for News Recommendation, in: *Proceedings of the 2014 International Joint Conferences on Web Intelligence and Intelligent Agent Technologies (WI-IAT '14)*, 347–354, 2014.
- [135] T. Murakami, K. Mori, R. Orihara, Metrics for Evaluating the Serendipity of Recommendation Lists, in: *Proceedings of the 2007 Annual Conference of the Japanese Society for Artificial Intelligence (JSAI '07)*, 40–46, 2007.
- [136] N. Muralidhar, H. Rangwala, E.-H. S. Han, Recommending Temporally Relevant News Content from Implicit Feedback Data, in: *Proceedings of the 27th International Conference on Tools with Artificial Intelligence (IC-TAI '15)*, 689–696, 2015.
- [137] M. Nakatsuji, Y. Fujiwara, A. Tanaka, T. Uchiyama, K. Fujimura, T. Ishida, Classical Music for Rock Fans?: Novel Recommendations for Expanding User Interests, in: *Proceedings of the 19th International Conference on Information and Knowledge Management (CIKM '10)*, 949–958, 2010.

- [138] N. Newman, R. Fletcher, D. Levy, R. Nielsen, The Reuters Institute Digital News Report 2016, Tech. Rep., Reuters Institute for the Study of Journalism, 2016.
- [139] A. Nishitarumizu, T. Itokawa, T. Kitasuka, M. Aritsugi, Improving a News Recommendation System in Adapting to Interests of a User with Storage of a Constant Size, in: Proceedings of the 12th International Asia-Pacific Web Conference (APWEB '10), 109–115, 2010.
- [140] K.-J. Oh, W.-J. Lee, C.-G. Lim, H.-J. Choi, Personalized news recommendation using classified keywords to capture user preference, in: Proceedings of the 16th International Conference on Advanced Communication Technology (ICACT '14), 1283–1287, 2014.
- [141] Ö. Özgöbek, J. A. Gulla, R. C. Erdur, A survey on challenges and methods in news recommendation, in: Proceedings of the 10th International Conference on Web Information Systems and Technologies (WEBIST '14), 278–285, 2014.
- [142] O. Özgöbek, N. Shabib, J. A. Gulla, Data sets and news recommendation, in: Proceedings of the News Recommendation and Analytics Workshop (NRA '14) at UMAP '14, 2014.
- [143] M. Papagelis, D. Plexousakis, T. Kutsuras, Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences, in: Proceedings of the 3rd International Conference on Trust Management (iTrust '05), 224–239, 2005.
- [144] D. H. Park, H. K. Kim, I. Y. Choi, J. K. Kim, A literature review and classification of recommender systems research, *Expert Systems with Applications* 39 (11) (2012) 10059 – 10072.
- [145] T. D. Pessemier, S. Leroux, K. Vanhecke, L. Martens, Combining Collaborative Filtering and Search Engine into Hybrid News Recommendations, in: Proceedings of the 3rd International Workshop on News Recommendation and Analytics (INRA '15) at RecSys '15, 14–19, 2015.
- [146] O. Phelan, K. McCarthy, B. Smyth, Buzzer-Online Real-Time Topical News Article and Source Recommender, in: *Artificial Intelligence and Cognitive Science*, vol. 6206 of *Lecture Notes in Computer Science*, Springer, 251–261, 2010.
- [147] R. Pon, A. Cárdenas, D. Buttler, T. Critchlow, Measuring the interestingness of articles in a limited user environment, *Information Processing & Management* 47 (1) (2011) 97 – 116.
- [148] R. K. Pon, A. F. Cárdenas, D. Buttler, T. Critchlow, Tracking Multiple Topics for Finding Interesting Articles, in: Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining (KDD '07), 560–569, 2007.
- [149] S. Prawesh, B. Padmanabhan, The "Top N" News Recommender: Count Distortion and Manipulation Resistance, in: Proceedings of the 5th Conference on Recommender Systems (RecSys '11), 237–244, 2011.

- [150] S. Prawesh, B. Padmanabhan, Probabilistic news recommender systems with feedback, in: Proceedings of the 6th Conference on Recommender Systems (RecSys '12), 257–260, 2012.
- [151] P. Pu, L. Chen, R. Hu, A User-centric Evaluation Framework for Recommender Systems, in: Proceedings of the 5th Conference on Recommender Systems (RecSys '11), 157–164, 2011.
- [152] J. Qiu, L. Liao, P. Li, News Recommender System Based on Topic Detection and Tracking, in: Rough Sets and Knowledge Technology, vol. 5589 of *Lecture Notes in Computer Science*, Springer, 690–697, 2009.
- [153] J. Rao, A. Jia, Y. Feng, D. Zhao, Personalized News Recommendation Using Ontologies Harvested from the Web, in: Web-Age Information Management, vol. 7923 of *Lecture Notes in Computer Science*, Springer, 781–787, 2013.
- [154] J. Rao, A. Jia, Y. Feng, D. Zhao, Taxonomy Based Personalized News Recommendation: Novelty and Diversity, in: Proceedings of the 14th International Conference on Web Information Systems Engineering (WISE '13), 209–218, 2013.
- [155] R. Ren, L. Zhang, L. Cui, B. Deng, Y. Shi, Personalized Financial News Recommendation Algorithm Based on Ontology, in: Proceedings of the 3rd International Conference on Information Technology and Quantitative Management (ITQM '15), 843 – 851, 2015.
- [156] G. Sadhasivam, K. Saranya, E. Praveen, Personalisation of News Recommendation Using Genetic Algorithm, in: Proceedings of the 3rd International Conference on Eco-Friendly Computing and Communication Systems (ICECCS '14), 23–28, 2014.
- [157] A. Said, A. Bellogin, A. de Vries, News Recommendation in the Wild: CWIs Recommendation Algorithms in the NRS Challenge, in: Proceedings of the 2013 International News Recommender Systems Workshop and Challenge at RecSys '13, 2013.
- [158] A. Said, J. Lin, A. Bellogin, A. de Vries, A Month in the Life of a Production News Recommender System, in: Proceedings of the 2013 Workshop on Living Labs for Information Retrieval Evaluation (LivingLab@CIKM '13), 7–10, 2013.
- [159] K. G. Saranya, G. S. Sadhasivam, A Personalized Online News Recommendation System, *International Journal of Computer Applications* 57 (18) (2012) 6–14.
- [160] M. Scriminaci, A. Lommatzsch, B. Kille, F. Hopfgartner, M. Larson, D. Malagoli, A. Serény, T. Plumbaum, Idomaar: A Framework for Multi-dimensional Benchmarking of Recommender Algorithms, in: Poster Proceedings of the 10th Conference on Recommender Systems (RecSys '16), 2016.
- [161] G. Shani, A. Gunawardana, Evaluating Recommendation Systems, in: F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (Eds.), *Recommender Systems Handbook*, chap. 8, Springer, 257–297, 2011.

- [162] Y. Shi, X. Zhao, J. Wang, M. Larson, A. Hanjalic, Adaptive Diversification of Recommendation Results via Latent Factor Portfolio, in: Proceedings of the 35th International Conference on Research and Development in Information Retrieval (SIGIR '12), 175–184, 2012.
- [163] E. Shmueli, A. Kagian, Y. Koren, R. Lempel, Care to Comment?: Recommendations for Commenting on News Stories, in: Proceedings of the 21st International Conference on World Wide Web (WWW '12), 429–438, 2012.
- [164] J. Sobocki, L. Szczepański, Wiki-news interface agent based on AIS methods, in: Proceedings of the 1st International Symposium on Agent and Multi-Agent Systems: Technologies and Applications (KES-AMSTA '07), 258–266, 2007.
- [165] J. W. Son, A. Kim, S. Park, A location-based news article recommendation with explicit localized semantic analysis, in: Proceedings of the 36th International Conference on Research and Development in Information Retrieval (SIGIR '13), 293–302, 2013.
- [166] M. Tavakolifard, J. A. Gulla, K. C. Almeroth, J. E. Ingvaldesn, G. Nygreen, E. Berg, Tailored News in the Palm of Your Hand: A Multi-perspective Transparent Approach to News Recommendation, in: Proceedings of the 22nd International Conference on World Wide Web (WWW '13), 305–308, 2013.
- [167] N. Tintarev, J. Masthoff, Similarity for news recommender systems, in: Proceedings of the Workshop on Recommender Systems and Intelligent User Interfaces at AH '06, 2006.
- [168] M. Trevisiol, L. M. Aiello, R. Schifanella, A. Jaimes, Cold-start news recommendation with domain-dependent browse graph, in: Proceedings of the 8th Conference on Recommender Systems (RecSys '14), 81–88, 2014.
- [169] H. P. Vanchinathan, A. Marfurt, C.-A. Robelin, D. Kossmann, A. Krause, Discovering Valuable Items from Massive Data, in: Proceedings of the 21th International Conference on Knowledge Discovery and Data Mining (KDD '15), 1195–1204, 2015.
- [170] H. P. Vanchinathan, I. Nikolic, F. De Bona, A. Krause, Explore-exploit in top-N Recommender Systems via Gaussian Processes, in: Proceedings of the 8th Conference on Recommender Systems (RecSys '14), ACM, New York, NY, USA, 225–232, 2014.
- [171] S. Vargas, L. Baltrunas, A. Karatzoglou, P. Castells, Coverage, Redundancy and Size-awareness in Genre Diversity for Recommender Systems, in: Proceedings of the 8th Conference on Recommender Systems (RecSys '14), 209–216, 2014.
- [172] S. Vargas, P. Castells, Rank and Relevance in Novelty and Diversity Metrics for Recommender Systems, in: Proceedings of the 5th Conference on Recommender Systems (RecSys '11), 109–116, 2011.
- [173] J. Wang, Q. Li, Y. P. Chen, J. Liu, C. Zhang, Z. Lin, News Recommendation in Forum-Based Social Media, in: Proceedings of the 24th Conference on Artificial Intelligence (AAAI '10), 1449–1454, 2010.

- [174] D. Wei, T. Zhou, G. Cimini, P. Wu, W. Liu, Y.-C. Zhang, Effective mechanism for social recommendation of news, *Physica A: Statistical Mechanics and its Applications* 390 (11) (2011) 2117–2126.
- [175] Z. Wei, J. Xun, X. Wang, One-class classification based finance news story recommendation, *Computational Information Systems* 5 (6) (2009) 1625–1631.
- [176] H. Wen, L. Fang, L. Guan, A hybrid approach for personalized recommendation of news on the Web, *Expert Systems with Applications* 39 (5) (2012) 5806–5814.
- [177] D. Werner, C. Cruz, Precision Difference Management Using a Common Sub-Vector to Extend the Extended VSM Method, in: *Proceedings of the 2013 International Conference on Computational Science (ICCS '13)*, 1179–1188, 2013.
- [178] D. Werner, N. Silva, C. Cruz, A. Bertaux, Using DL-reasoner for hierarchical multilabel classification applied to economical e-news, in: *Proceedings of the 2014 Conference on Science and Information (SAI '14)*, 313–320, 2014.
- [179] X. Wu, F. Xie, G. Wu, W. Ding, Personalized news filtering and summarization on the web, in: *Proceedings of the 23rd International Conference on Tools with Artificial Intelligence (ICTAI '11)*, 414–421, 2011.
- [180] Y. Wu, Y. Ding, X. Wang, J. Xu, Topic based automatic news recommendation using topic model and affinity propagation, in: *Proceedings of the 2010 International Conference on Machine Learning and Cybernetics (ICMLC '10)*, 1299–1304, 2010.
- [181] Z. Xia, S. Xu, N. Liu, Z. Zhao, Hot news recommendation system from heterogeneous websites based on bayesian model, *The Scientific World Journal* 2014.
- [182] K. F. Yeung, Y. Yang, A Proactive Personalized Mobile News Recommendation System, in: *Proceedings of the 2010 International Conference on Developments in E-Systems Engineering (DeSE '10)*, 207–212, 2010.
- [183] K. F. Yeung, Y. Yang, D. Ndzi, Context-Aware News Recommender in Mobile Hybrid P2P Network, in: *Proceedings of the 2nd International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN '10)*, 54–59, 2010.
- [184] X. Yingyuan, A. Pengqiang, H. Ching-hsien, W. Hongya, J. Xu, Time-ordered collaborative filtering for news recommendation, *China Communications* 12 (12) (2015) 53–62.
- [185] Y. C. Zhang, D. O. Séaghdha, D. Quercia, T. Jambor, Auralist: Introducing Serendipity into Music Recommendation, in: *Proceedings of the 5th International Conference on Web Search and Data Mining (WSDM '12)*, 13–22, 2012.
- [186] L. Zheng, L. Li, W. Hong, T. Li, PENETRATE: Personalized news recommendation using ensemble hierarchical clustering, *Expert Systems with Applications* 40 (6) (2013) 2127–2136.

- [187] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, Y.-C. Zhang, Solving the apparent diversity-accuracy dilemma of recommender systems, *Proceedings of the National Academy of Sciences* 107 (10) (2010) 4511–4515.
- [188] C.-N. Ziegler, S. M. McNee, J. A. Konstan, G. Lausen, Improving recommendation lists through topic diversification, in: *Proceedings of the 14th International Conference on World Wide Web (WWW '05)*, 22–32, 2005.